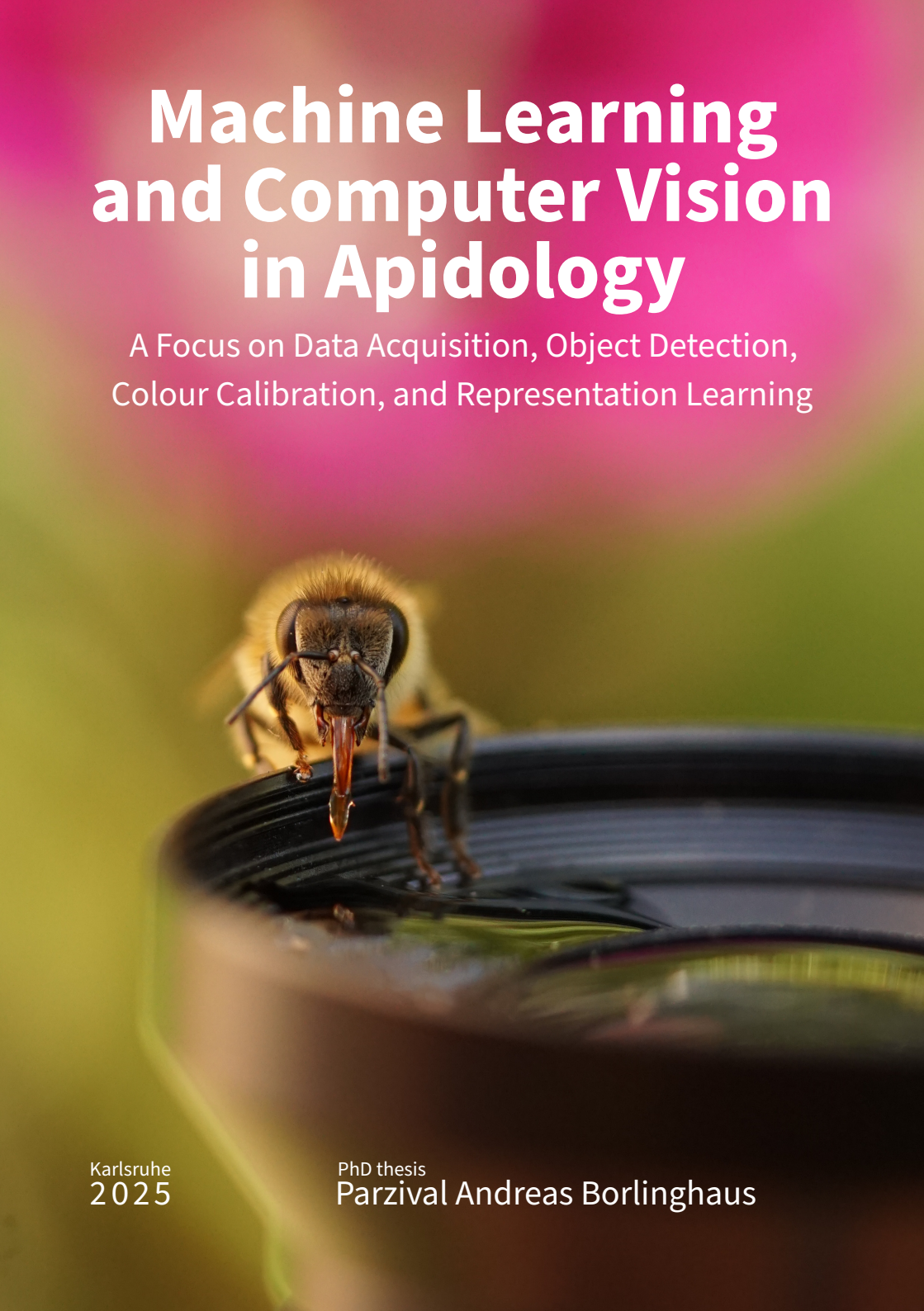
Machine Learning and Computer Vision in Apidology

A Focus on Data Acquisition, Object Detection, Colour Calibration, and Representation Learning



Karlsruhe
2025

PhD thesis
Parzival Andreas Borlinghaus

Machine Learning and Computer Vision in Apidology

A Focus on Data Acquisition, Object Detection, Colour Calibration, and Representation Learning

Zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften von der KIT-Fakultät für Wirtschaftswissenschaften des Karlsruher Instituts für Technologie (KIT) genehmigte Dissertation.

Karlsruhe
2025

PhD thesis
Parzival Andreas Borlinghaus

Bibliographic information

Title	Machine Learning and Computer Vision in Apidology: A Focus on Data Acquisition, Object Detection, Colour Calibration, and Representation Learning
Author	Parzival Andreas Borlinghaus
Institute	Institute for Operations Research, Karlsruhe Institute of Technology
Supervision	Prof. Dr. Oliver Grothe, Institute for Operations Research, Karlsruhe Institute of Technology
Co-Supervision	PD i.R. Dr. Peter Rosenkranz, State Institute of Bee-Research, University of Hohenheim
Date of the oral examination	February 4, 2025

Abstract

We are in the midst of the Fourth Industrial Revolution, and so are the bees. The advent of new and more affordable sensors, along with advanced data processing methods, has introduced versatile applications in apidology. Ironically, these technologies are now being employed to better understand the impacts of previous industrial revolutions, including the consequences of changes in land use, climatic factors, and the spread of invasive species facilitated by global trade routes. The shift towards digitized apidology has become crucial in response to the dramatic decline in insect populations worldwide. This dissertation embraces this shift by leveraging and advancing the capabilities of machine learning in computer vision to enhance our understanding and protection of pollinators.

The first set of contributions focus on improving video data acquisition around honey bee hives (*Apis mellifera*) and evaluating devices for quantifying bee mortality for regulatory purposes, thus making plant protection products safer for non-target organisms. A study on the re-identification of bumblebees (*Bombus terrestris*) demonstrates how future behavioural studies can be conducted without markers, non-invasively and purely visual. Additionally, techniques for the automatic calibration of pollen colours on smartphone images and biodiversity assessment are presented, enabling landscape monitoring through citizen scientists. Finally, we present a groundbreaking honey bee brood monitoring system that can monitor *Varroa destructor* reproduction in cells, and supports breeding for varroa-resistant traits.

Keywords: Machine Learning, Computer Vision, Apidology, Honey Bees, Pollen Analysis, Brood Analysis

Photo credits: Macro shots of pollinators in and around Karlsruhe were generously provided by Lilith Schumann.

Contents

Acronyms

CIE	International Commission on Illumination
CMC	Cumulative Matching Characteristics
CNN	Convolutional Neural Network
CV	Computer Vision
EFSA	European Food Safety Authority
FC	Fully Connected Layer
GMM	Gaussian Mixture Model
mAP	Mean Average Precision
ML	Machine Learning
MOTA	Multiple Object Tracking Accuracy
PCA	Principal Component Analysis
re-ID	Re-Identification
SGD	Stochastic Gradient Descent
VSH	Varroa-sensitive hygiene

Danksagung

Dass ich aus allen Pfaden, welche mir nach dem Studium offen standen, diesen gewählt habe, verdanke ich meinem Doktorvater, Prof. Oliver Grothe. Für diesen initialen Funken, Deinen Rat und die beständige Unterstützung bin ich dir zutiefst dankbar.

Ein ebenso herzlicher Dank gilt Richard, der mich von Beginn an in die Kunst des Publizierens eingeweiht hat. Wann immer ich mit meinem Latein am Ende war, konntest du mir die Welt der *Apidae* einen Flügelschlag näherbringen.

Meinen Kolleginnen und Kollegen am Lehrstuhl danke ich für die inspirierende Begleitung während der vergangenen Jahre. Besonders erwähnen möchte ich Jakob, der jede neue Netzwerkarchitektur und jeden Algorithmus aufmerksam angehört, hinterfragt und mit mir bis in die Tiefe diskutiert hat. Es bleibt bedauerlich, dass von den vielen Entwicklungen nur wenige Eingang in diese Arbeit gefunden haben.

Ich möchte Stephan danken, der mir stets ein paar Jahre voraus war. Seine akademischen Fußstapfen dienten mir als Orientierung, und seine Arbeitsmoral wirkte verführerisch wie *Boletus edulis*. Der Austausch über Flieger und Flugzeuge hat meine eigene Flugbahn stets stabilisiert.

Im PKV habe ich Ausgleich gefunden, ein erfrischendes Verhältnis zum Risiko entwickelt und Hochstapler fürs Leben gefunden. Ein ganz besonderer Dank gilt meiner Partnerin - du bist die beste Spotterin, die ich mir wünschen könnte.

Meiner Familie, insbesondere meinen Eltern, danke ich von Herzen für ihren Rückhalt. Von ihnen weiß ich, dass es manches zu entdecken gibt, wenn man den Zweck, der den Dingen innewohnt, hinterfragt und neu bestimmt.

Schließlich möchte ich der Deutschen Bundesstiftung Umwelt (DBU) für die ideelle und materielle Unterstützung während meiner Promotion danken.

1 Introduction

'In the garden of technology, bees either pollinate dreams or harvest nightmares.'

HiveHelp.ai language model on opportunities yet to be shaped, 2024.

For longer than the headlines about AI innovations, we have been confronted with news about insect decline. These headlines appear not only in newspapers but also on federal and even European election posters. Since the pivotal study by Hallmann et al. [100] in 2017, insect decline has become a widely discussed issue. The authors of the study reveal that the biomass of flying insects in German nature reserves has decreased by more than three-quarters over a period of 27 years. The causes of this decline are diverse, ranging from pesticide use and habitat loss to climate change and disease [21].

Among the more than one million described insect species, the honey bee (*Apis mellifera*) stands out. This is not only due to the many opportunities to encounter one of the million colonies or one of the 130 000 beekeepers in Germany [109], but also because the honey bee plays a key role in researching insect decline. Thanks to its widespread presence and availability, it serves as a model organism for understanding insect populations and the threats they face.

Besides shared challenges, *A. mellifera* also faces species-specific threats. In this context, the Varroa mite – a parasitic invader – must be mentioned, as it remains the bees' greatest foe [208, 241]. However, the next invasive enemy of honey bees is already at our doorstep. Specifically, the Asian hornet (*Vespa velutina*) has crossed the border from France and was first spotted in Germany

2014, only few kilometres away from KIT. Ten years later, I observe Asian hornets weekly preying at my bee colonies in Karlsruhe.

Thanks to the growing awareness of insects' importance to our ecosystem, research interest is high. Karl von Frisch was awarded the Nobel Prize in Physiology or Medicine in 1973 for his groundbreaking work observing and explaining animal behaviour, particularly that of bees. Citing from his laudatio he did not attempt to 'understand it from his own experiences, from his own way of thinking, feeling and acting' [142], but instead through 'means of scientific methods, by systematic observation and by experimentation' [142]. Today, 51 years after the laudatio was hold, there are still many questions awaiting the engagement of curious minds. However, the tools to conduct systematic observation and experimentation have drastically evolved. While von Frisch had to mark individual bees with coloured dots in his experiments to decode the 'language of bees', today, hundreds of bees can be distinguished automatically using QR codes and machine learning [53].

The pioneers of the first convolutional neural networks, which have shaped modern computer vision more than anything else, could hardly have imagined that their innovations would attract significant interest in fields as distant as apidology. Yet this is exactly where this work makes its mark – and it is far from an isolated case. Today, computer vision and machine learning are not only being used to study bee behaviour with QR codes, but also to study Varroa infestations with cameras placed inside [251] and outside [18] the hive, monitor intruders [156], identify individual fruit flies [170] and use wing patterns to classify species [197]. Thousands of bees can now be tracked simultaneously on honeycombs, and individual bees can even be followed in the field [195]. Remarkably, even bee language, once painstakingly observed by von Frisch, can now be decoded and projected onto a map using modern technology [253].

And still, the potential of machine learning in apidology is far from exhausted. Only recently, 750 beekeepers from 24 countries manually analysed the colour diversity in 18 000 corbicular pollen samples of honey bees, dedicating an estimated 6 000 hours of voluntary labour [41] – time that could have been saved through automation. In a notable study from last year, honey bee brood cells were deliberately infested with both live and dead parasites to better understand which cues trigger worker bees to inspect brood cells [229]. An automated, non-invasive brood monitoring device could have allowed for continuos observation, thereby offering a time dimension to the study.

Although computer vision has gained popularity in tackling the challenge of accurate bee counting, its precision remains insufficient for real-world applications [181].

In this work, we bridge these gaps by exploring various ways machine learning and computer vision can be used to improve existing methods, both in terms of time efficiency and quality, while opening up new possibilities around the hive: for counting, re-identification, biodiversity assessment and brood monitoring. The thesis is divided into two parts. The first part consists of chapters on computer vision (Chapter 2) and apidology (Chapter 3) and introduces the reader to the key principles of both fields. The second part presents the research contributions that are described in detail below.

Part two opens with the observation that the successful application of computer vision depends on high-quality image data; however, the literature demonstrates that obtaining such data is often a considerable challenge. Based on own experiments and results of others, Chapter 4 shows how a device can be constructed to capture high quality images of bees in front of the hive. The resulting data set is unprecedented in quantity and quality. More than 150 000 centre-cropped images of drones, nectar and pollen foragers have been made freely available to the community.

As highlighted in the introduction, insect mortality is a critical issue – but how many insects did not return from their foraging flight… *exactly*? For more than a century, accurately counting bees at the hive entrance to assess colony losses due to natural or anthropogenic factors has been a persistent challenge [181]. In the absence of precise data, mortality rates have to be estimated, with severe consequences for ecotoxicology and risk assessment of plant protection products [78]. In Chapter 5, we propose a protocol for validating bee counting systems, as the lack of reliable data on the accuracy of these systems currently hinders progress in this area.

While the aforementioned QR codes provide an elegant solution for distinguishing larger numbers of bees, their application remains labour intensive. In Chapter 6, we investigate whether individual recognition can be achieved on the basis of visual features alone by studying larger, more variable and less numerous bumblebees (*Bombus terrestris*). We find that the best performing model can identify two thirds of individuals on the first attempt, thereby providing future studies with an automated alternative to manual marking, allowing for larger studies and longer study periods.

In Chapter 7, we investigate whether bees, owing to their flower constancy, can serve not only in their invaluable role as pollinators but also as biodiversity sensors. By collecting pollen, they carry data that imaging sensors can use to visually assess the biodiversity of their foraging landscape. This concept was developed into an app that enables beekeepers to use their smartphone camera to estimate the biodiversity around their apiaries, and thereby assessing the site's nutritional value [40]. This was achieved through the successful use of everyday objects for colour calibration, addressing the challenge of making colours from smartphone images comparable across different devices, locations, and times. The app provides an automated, objective, and reproducible method for determining corbicular pollen colours, benefiting both scientific research and beekeeping practices.

Building on this app for pollen detection and colour extraction, we conducted the first major investigation into the natural colour variation within pollen types (Chapter 8). Using automated colour extraction and Gaussian Mixture Models for data fusion, we successfully clustered pollen loads based on pollen type probabilities from laboratory analysis and colours extracted from 85 531 individual pellets. This enabled us to quantify the type-specific colour dispersion in the form of covariance matrices for half of the major European pollen types.

In Chapter 9, we introduce a device that represents a significant advance in honey bee brood monitoring, and in Varroa research in particular. This low-cost system uses, for the first time, a contact image sensor to regularly capture non-invasive images of all activity within the brood cells. Among other pathogens, the system detects Varroa infestation and allows assessment of reproductive success, offering exceptional potential for breeding Varroa-resistant bees, evaluating acaricide treatments and conducting ecotoxicological studies in general.

Chapters 5 to 9 are based on Borlinghaus et al. [35], Borlinghaus, Tausch, and Rettenberger [36], Borlinghaus, Jung, and Odemer [31], Borlinghaus, Odemer, and Tausch [34], and Borlinghaus, Gülzow, and Odemer [30]. A summary is provided in Table 1.1 on the next page.

Returning to the initial metaphor, this research aims to plant seeds of innovation in the 'garden of technology', nurturing new devices and methods that will, we hope, yield further valuable insights in the future.

2024	Chapter 9 and Chapter 8	
–	*An Insightful View: In-Hive Flatbed Scanners for Non-invasive Long-Term Behaviour and Disease Monitoring of Honey Bee Brood Combs* [30] – Authored by P. Borlinghaus, J. M. Gülzow and R. Odemer. – Submitted to Smart Agricultural Technology (Q1, CS 4.2, IF 6.3). – Presented by P. Borlinghaus at Southern Ontario Bee Researchers' Symposium (BeeCon) in York, CA (2024).	
–	*Natural colour dispersion of corbicular pollen limits colour-based classification* [34] – Authored by P. Borlinghaus, F. Tausch and R. Odemer. – Publ. in ISPRS Open Journal of Photogrammetry and Remote Sensing (CS 5.1).	
2023	**Chapter 7 and Chapter 6**	**Citations**
–	*Introducing Pollenyzer: An App for Automatic Determination of Colour Diversity for Corbicular Pollen Loads* [31] – Authored by P. Borlinghaus, J. Jung and R. Odemer. – Published in Smart Agricultural Technology (Q1, CS 4.2, IF 6.3). – Presented by R. Odemer at Hazards of Pesticides to Bees: Bee Protection Group 15th International Symposium. International Commission for Plant-Pollinator Relationships (ICPPR) in York, UK (2022). – Presented by P. Borlinghaus at 9th European Congress of Apidology (EurBee) in Belgrade, SRB (2022). – Presented by P. Borlinghaus at 19th COLOSS Conf. in Bled, SI (2023).	4
–	*A Purely Visual Re-ID Approach for Bumblebees (Bombus Terrestris)* [36] – Authored by P. Borlinghaus, F. Tausch and L. Rettenberger. – Published in Smart Agricultural Technology (Q1, CS 4.2, IF 6.3). – Presented by F. Tausch at the IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR) in New Orleans, US (2022). – Presented in extended form by P. Borlinghaus at the 16th German Probability and Statistic Days (GPSD) in Essen, DE (2023).	10
2022	**Chapter 5**	**Citations**
–	*Honey Bee Counter Evaluation – Introducing a Novel Protocol for Measuring Daily Loss Accuracy* [35] – Authored by P. Borlinghaus, R. Odemer, F. Tausch and O. Grothe. – Publ. in Computers and Electronics in Agriculture (Q1, CS 15.3, IF 7.7).	15

Table 1.1: Overview of the incorporated publications. Where available, citation counts are sourced from Google Scholar, CiteScores (CS) and Impact Factors (IF) from journal websites (Elsevier), and Journal Quartiles (Q) from Scimago. Data accurate as of September 19, 2024.

Part I

Technological and Biological Foundations

2 Computer Vision

Computer vision (CV) is a well-established discipline within computer science that has recently experienced significant advancements thanks to machine learning (ML) techniques. Its goal is to make the visual world accessible to computers [262]. The field focuses on developing algorithms and methods that enable computers to extract meaningful information from images and videos [223]. Computer vision has been successfully applied across various domains, including medical image analysis [144] and agriculture [115].

In this chapter, we will provide an overview of the key concepts relevant to this work, beginning with colours and colour spaces, followed by an introduction to Convolutional Neural Networks (CNNs) and their application in ResNet and U-Net architectures for image classification and segmentation. To conclude the chapter, we will explore how representation and metric learning can be used to generate descriptions of individuals that allow for their re-identification.

2.1 Colours and Colour Spaces

Colour is a sensory perception triggered by radiation from light sources that reaches our eyes. This radiation stimulates the photoreceptor cells in the eye, known as cones, which create a colour sensation that is processed by our brain [154]. It is important to note that colour is not merely a physical property but, above all, a subjective sensory experience [200]. The human eye contains three types of cones, each sensitive to different wavelengths of light: short, medium, and long wavelengths. The perception of colour arises from the combination of stimuli from these three cones [134]. These combinations can be mathematically represented in what is known as a colour space.

The most well-known colour space is the RGB colour space, which represents colours as combinations of the three primary colours: red, green, and blue. However, RGB values represent colours relative to a specific device or display, and without standardization [130]. The same RGB values can appear differently on various devices, meaning they may be perceived as different colours. To represent absolute colours, colour spaces such as the CIE XYZ colour space have been developed by the Commission internationale de l'éclairage (CIE), based on a standard illuminant and a standard observer, which simulate average human vision under specific lighting conditions. The CIE XYZ colour space is designed to be device-independent and represents colours based on human colour perception in three dimensions, each defined by the visual response to the stimulus under laboratory conditions [130].

For this work, the CIELAB colour space holds special importance. It is designed to be perceptually *uniform*, meaning that the distances between colours can be meaningfully compared and correspond to the differences perceived by the human eye [154]. In this space, a colour is described by L^*, a^*, and b^* coordinates. The luminance component, L^*, ranges from 0 (black) to 100 (white), while the colour axes a^* range from -128 (green) to $+127$ (red) and b^* from -128 (blue) to $+127$ (yellow) [164]. Its special design makes the CIELAB colour space an essential tool for quantifying differences in colour perception.

To give RGB data a colorimetric interpretation, a calibration is needed [130]. For this, colour calibration cards are used. These cards contain colour patches with known values under a standard illuminant and observer. A digital image of the card is taken and used to establish a correspondence between the RGB values of the camera and the known colours of the patches. This correspondence is used to fit a transformation to the target colour space, which can be used to calibrate all images that were taken under equal conditions [140]. If the data are within the same linear colour space, or converted accordingly, the transformation from the camera to the target colour space can be represented by a 3×3 matrix, essentially a rotation and scaling of the camera's colour space [140]. When needed, a larger 3×4 transformation matrix can be employed to enable translation, allowing for the adjustment of the camera's black level.

Due to the perceptual uniformity of the CIELAB colour space and the orthogonality of its axes, differences between two colours can be described by the

CIEDE76 Colour difference	Interpretation
$0 \leq \Delta E < 1$	Observer does not notice the difference.
$1 \leq \Delta E < 2$	Only experienced observer can notice the difference.
$2 \leq \Delta E < 3.5$	Inexperienced observer also notices the difference.
$3.5 \leq \Delta E < 5$	Clear difference in colour is noticed.
$5 \leq \Delta E$	Observer notices two different colours.

Table 2.1: Colour differences as seen by a standard observer [164].

Euclidean distance. This distance is referred to as ΔE [189] and is defined as:

$$\Delta E\left(L_1^*, a_1^*, b_1^*,\ L_2^*, a_2^*, b_2^*\right) = \sqrt{\left(L_2^* - L_1^*\right)^2 + \left(a_2^* - a_1^*\right)^2 + \left(b_2^* - b_1^*\right)^2} \tag{2.1}$$

In experiments by Mokrzycki and Tatol [164], the distances were quantified and made accessible for interpretation. The results are summarised in Table 2.1.

Figure 2.1 illustrates ΔE colour differences for six common European pollen colours. Please note that due to the limitations of display and printing, the absolute colour values may not be accurately represented. However, it is evident that as the ΔE value increases, the colour differences from the original colour become more pronounced.

2.2 Convolutional Neural Networks and Their Building Blocks

Convolutional Neural Networks (CNNs) have a long history in the field of machine learning, with their origins dating back to the late 1980s. One of the early milestones was the work by LeCun et al. [138], who applied backpropagation to handwritten zip code recognition, creating one of the first practical CNNs. However, these early networks were relatively small and limited in depth. The real breakthrough came in 2012 when Krizhevsky, Sutskever, and Hinton [128] introduced a much larger and deeper CNN with five CNN layers, three Fully Connected Layers (FC) and a total of 60 million parameters that was trained on a massive dataset of over a million images. This model, known

Figure 2.1: For six common European pollen colours [32] (centre), exemplary real pollen loads are displayed with colour differences ranging from $\Delta E \approx 2$ to $\Delta E \approx 32$. As the ΔE value increases, the colour differences from the original shade become more pronounced.

as AlexNet, revolutionised the field by significantly improving the state-of-the-art in image recognition. Following this, the trend continued with even deeper architectures, such as the 19-layer VGG network [225], which further pushed the boundaries of CNNs in terms of depth and performance.

In addition to convolutional layers, the basic building blocks of a CNN include pooling layers, activation functions, and fully connected (FC) layers.

A convolutional layer applies a set of filters to the input feature map. Figure 2.2 illustrates such isolated convolution operation with RGB input (left), a single kernel (centre) and a resulting single-channel feature map (right). A kernel of shape $3 \times 3 \times C$ is applied to a close-up image of two pollen loads. At each position, the resulting value in the output feature map is a weighted sum of the input values within the kernels receptive field and the kernel's trainable parameters. In practice, multiple independent kernels learn different aspects of low-level features that serve as the foundation for subsequent layers. However, before a feature map is passed to the next layer, activation functions, such as the Rectified Linear Unit ReLU $(x) = \max(0, x)$ [171] or the Sigmoid function $\sigma(x) = 1/(1 + e^{-x})$ are applied to the layer's output. Activation functions can introduce non-linearity into the network, enabling it to learn more complex functions [139].

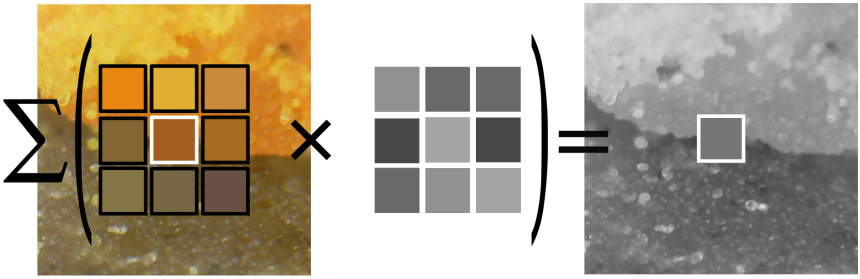


Figure 2.2: An isolated convolution operation with RGB input and a single output channel is illustrated. The operation involves an element-wise multiplication between corresponding elements of both $K \times K \times C$ tensors, followed by a summation of the resulting values. H and W denote the spatial dimensions, $C = 3$ the number of input channels, and $K = 3$ the kernel size.

Fully connected layers can be understood as a special case of CNNs, as they are equivalent to convolutional layers with kernel size $H \times W \times C$, where H and W are the spatial dimensions of the layer's input. As the name implies, FC layers connect every neuron in one layer to every neuron in the next layer, allowing the network to learn patterns without restrictions on their spatial location. It is evident that convolutional layers with typical kernel sizes of 3×3 to 7×7 reduce the number of parameters dramatically, compared to their $H \times W$ sized fully connected counterpart. However, this comes at the expense of the receptive field, i.e. the amount of pixels in the input image that can influence a layer's output. This is nevertheless a favourable trade-off, as lightweight convolutional layers can be cheaply stacked and combined with pooling layers.

Pooling layers downsample the feature maps generated by the convolutional layers, thus reducing the spatial dimensions of the data and therefore increasing the receptive field of the following convolution.

During training, the network's output $\hat{y}$ must be compared with the ground truth labels y to update the network's weights. Depending on the problem at hand, different loss functions $L(y, \hat{y})$ are commonly used. For classification problems (and segmentation can be considered pixel-wise classification), the

(multi-class) cross-entropy loss is the method of choice [207]. It is defined as:

$$L(\hat{y}_c, y_c) = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \ , \tag{2.2}$$

where $\hat{y}_c$ is the predicted probability that the input y belongs to class $c \in \{1, 2, \ldots, C\}$. Similarly, y_c indicates the true class affiliation. Based on the calculated loss, optimizers such as Stochastic Gradient Descent (SGD) [203] or Adam [120] push the network's weights in the direction of a better solution. The SGD algorithm is applied by feeding a minibatch of training data through the network, comparing the results to the desired output, calculating the gradients of the loss function with respect to each weight in the network and updating the weights accordingly [139]. SGD is considered stochastic because each iteration over a minibatch provides only a noisy estimate of the gradient calculated over the entire dataset [139].

One of the primary strengths of CNNs lies in their ability to automatically learn hierarchical features from data. In the early layers, CNNs capture low-level features such as edges and textures, while the deeper layers progressively extract more abstract, high-level features such as motifs and arrangements thereof [139]. This hierarchical feature learning enables CNNs to recognize complex patterns and relationships within the data, making them highly effective as feature extractors or backbones for various downstream tasks. In the following sections, we will introduce ResNet-18, a commonly used feature extractor for classification tasks, and U-Net, a widely adopted architecture for image segmentation tasks.

2.2.1 Residual Network Architecture

The previously described deepening of CNNs cannot be extended indefinitely, as the training becomes unstable, a problem known as the exploding and vanishing gradient. If the gradients become too small, the training stagnates; if they become too large, they become imprecise and can harm the training process [101, 102]. This leads to the counter intuitive observation that a deeper network is not necessarily as good as or better than its shallower counterpart. This is surprising because the additional layers could, in theory, learn to adopt a result-neutral state. From the fact that deeper networks perform worse in their experiments, even though the additional layers could

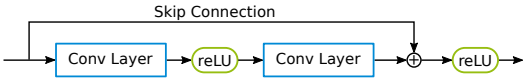


Figure 2.3: A skip connection surrounds both layers of the residual block. Adapted from He et al. [102].

theoretically learn at least an identity mapping, He et al. [102] conclude that 'the solvers might have difficulties in approximating identity mappings by multiple nonlinear layers.'

The authors propose a structure called a residual block, in which pairs of layers are surrounded by a skip connection, as illustrated in Figure 2.3. This design ensures that adding depth to the network does not impair learning. Instead, it simplifies the process of learning the identity mapping, which, as the authors demonstrate, is not only easier but also highly beneficial. The skip connection bypasses two convolutional layers, directly adding the input to the output without introducing any additional parameters. This shortcut allows the layers to focus solely on learning the residuals relative to the input [102]. By adopting this approach, very deep networks can be trained without encountering the instability that typically arises in such cases. The ResNet architectures studied by the authors range from 18 to 152 layers, with experiments even conducted on networks as deep as 1202 layers. All ResNets follow a similar architecture, as depicted in Figure 2.4 for ResNet-18.

The relatively compact ResNet-18 [102] consists of an initial convolutional layer, followed by four groups of two residual blocks each. Each group shares the same spatial dimension. The spatial dimensions are reduced by max pooling after the initial layer and through striding in the first convolutional layers of groups 2, 3, and 4. A final average pooling layer is followed by a fully connected layer, making the network suitable for classification tasks. As identity connections cannot work for different spatial and channel dimensions, 1×1 convolutions are used to match the dimensions where necessary. When used as a backbone for other tasks, the final FC layer is removed, and the output of the last convolutional layer serves as a learned representation of the input image.

Despite the relatively shallow depth of 18 layers, which theoretically allows training without skip connections, their use accelerates the training process and justifies their inclusion [102].

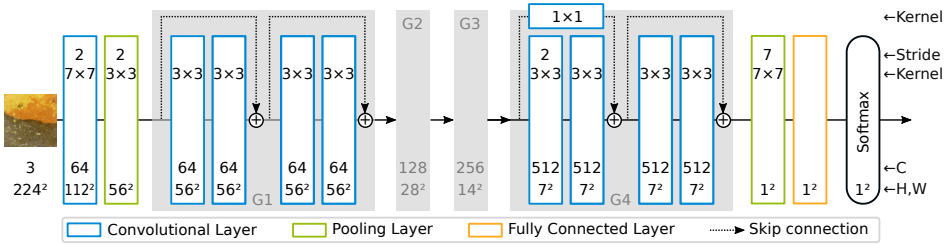


Figure 2.4: The ResNet-18 architecture [102] consists of 18 trainable layers, organised into 4 groups (G1 to G4), each containing 2 ResNet blocks that incorporate identity mappings. To match the input and output dimensions between these groups, 1×1 convolutions with a stride of 2 are employed. Throughout the network, pooling layers and strided convolutions gradually reduce the spatial dimensions from 224×224 at the input to 1×1 at the output.

2.2.2 U-Net Architecture

The U-Net can be understood as an advanced form of the autoencoder architecture. Autoencoders are networks designed to reconstruct their input as accurately as possible. They consist of an encoder, which transforms the input into a compact representation, and a decoder, which maps this representation back to the input. Hinton and Salakhutdinov [104] describe the encoder part of the architecture as a 'nonlinear generalization of PCA' [104], with its output often referred to as the bottleneck. Autoencoders can effectively reconstruct the input but are not designed to classify it. This highlights a significant conceptual difference, as the labels and inputs are identical during autoencoder training.

The U-Net, introduced by Ronneberger, Fischer, and Brox [207], is a specialised type of autoencoder network developed for image segmentation. In this context, the label is the corresponding segmentation mask for the input image. While skip connections in an autoencoder would interfere with compression, in U-Nets they significantly contribute to its success.

As shown in Figure 2.5, a U-Net consists of a contracting path (left) and an expanding path (right) that exchange information through so-called skip connections (dashed lines) [207].

The contracting path in the U-Net architecture follows a standard convolutional network structure, applying two 3×3 convolutions (without padding),

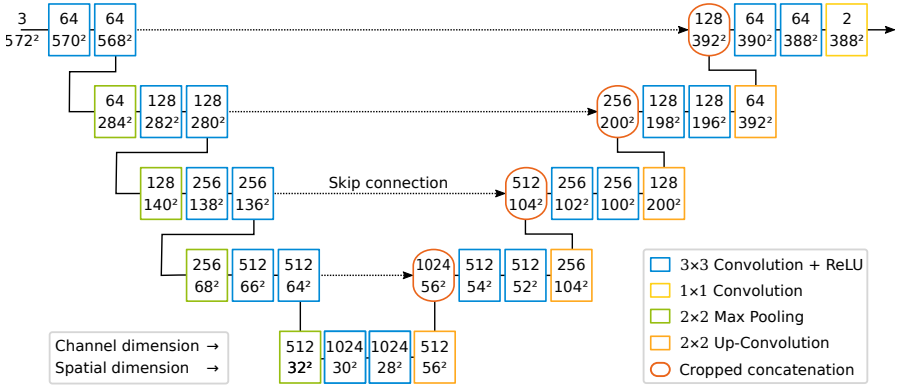


Figure 2.5: Original U-Net architecture, adapted from Ronneberger, Fischer, and Brox [207]. The architecture consists of a contracting path (on the left) and an expanding path (on the right), which are interconnected by skip connections (dashed lines).

each followed by a ReLU activation function, and a 2×2 max pooling operation with a stride of 2 for downsampling.

The expanding path involves an up-convolution layer [260] that doubles the spatial dimension and halves the number of channels. The cropped concatenation, highlighted in red in Figure 2.5, performs a centre crop on the larger feature map before concatenating along the channel dimension. The final layer, one of a total of 23 convolutional layers, uses a 1×1 kernel to map each 64-component feature vector to the target classes [207].

In the expanding path, the network has access to both the encoded representation of the entire image from the bottleneck and the features of the input image at the corresponding resolution. The combination of these two sources of information enables the network to perform image-to-image translation tasks, such as segmentation, with high resolution [207]. Since segmentation involves pixel-wise classification, a pixel-wise distribution over the classes can be obtained using softmax activation on the output feature maps. During training, cross-entropy loss is applied [207].

A notable feature of the U-Net is its complete absence of fully connected (FC) layers, which qualifies it as a Fully Convolutional Network [147]. This allows it to process images of variable size, as the input image size is not constrained by the fixed dimensions of an FC layer [207].

2.3 Object Re-Identification

The computer vision task of re-identification (re-ID) involves associating objects across different images based on visual attributes. Naturally, interest in re-ID of humans is particularly high, driving many developments in the field [94]. However, its utility has also been recognised in wildlife observation, for example, to analyse camera trap footage [216].

Re-identification can be divided into two cases: when all individuals are known (closed-set re-ID), and when this is not the case (open-set re-ID). Both scenarios involve a probe image and a gallery set. For closed-set re-ID, the goal is to determine which identity in the gallery matches the probe. In open-set re-ID, it must also be determined *if* the individual is known to the gallery, which typically requires applying a threshold to the similarity metric [54].

In the following, we will introduce the key concepts of re-ID. First, we will discuss representation learning, focusing on the triplet loss, a widely-used method for training embedding networks. These networks extract descriptive features that can be compared using standard metrics to identify similar individuals. If the features are considered fixed, an alternative approach is to learn a suitable metric to achieve the desired outcome; this is termed metric learning and will be the second concept we explore. Finally, two approaches are presented to assess the quality of re-ID systems.

2.3.1 Representation Learning

To identify an object, a description is needed to compute similarities between the probe image of interest and the gallery set. This description is referred to as the representation or embedding of image $\mathbf{X}$ and needs to be learned [139]. The dimensionality of embedding vector $\mathbf{w} \in \mathbb{R}^d$ depends on the complexity of the application. Once objects are embedded, (dis)similarity can be determined, for example, by general-purpose metrics like the Euclidean distance

$$d_{\mathrm{L}_2}(\mathbf{w}_1, \mathbf{w}_2) = \|\mathbf{w}_1 - \mathbf{w}_2\|_2 \,. \tag{2.3}$$

When the dimensionality is considerable, the embedding space becomes exceedingly large and thus sparsely populated. In such instances, the Euclidean distance is not always a reliable measure. This phenomenon, known as the

'Curse of Dimensionality' [13, 1], can be mitigated by utilising metrics less vulnerable to sparsity, such as cosine distance [12, 11]:

$$d_{\cos}(\mathbf{w}_1, \mathbf{w}_2) = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\|\|\mathbf{w}_2\|} . \tag{2.4}$$

Given such metric, the goal is to train an embedding network that infers embeddings that yield high similarities for identical individuals. A naive approach is to train and modify an auxiliary classification network. As demonstrated by Taigman et al. [235], a CNN backbone can be used as a feature extractor and complemented by a FC for classification. During this auxiliary training, the FC layer learns to predict classes (identities) from the feature maps generated by the feature extractor. However, this method can only determine identities known during training and for which sufficient training data was available. Since the feature extractor has learned to generate feature maps that allow easy determination of class affiliation, the trick is to discard the FC layer during inference and interpret the output of the feature extractor as the desired image embedding. These embeddings can be used to determine similarities between probe and gallery images, regardless of whether they were in the training dataset or not. The disadvantages of this approach include the lack of assurance that the obtained representations generalize well and that the embeddings of the same identity typically scatter more than necessary [218].

One way to avoid the inefficiencies and indirect features associated with the auxiliary classification task is to adopt the contrastive learning para-digm, specifically utilizing the triplet loss. The precise mechanism of this approach is detailed in the following section.

2.3.2 Triplet-Loss

The triplet loss was first adapted for re-ID by Schroff, Kalenichenko, and Philbin [218] and aims to assign semantic locations in the embedding space to unseen identities. By enforcing a distance margin α between embeddings of different identities, the triplet-loss allows to train the re-ID model without detour through a cut-off auxiliary classification network.

The starting point is again a feature extractor network that computes an embedding from an input image. In each training step, three images are selected:

an anchor image, a corresponding image of the same identity (positive image), and an image of a different identity (negative image). These triplets give the name to the presented loss [218].

The (squared) distance between the embedding of $\mathbf{X}^a$ (anchor image) and $\mathbf{X}^p$ (positive image) should be small, but large for $\mathbf{X}^a$ and $\mathbf{X}^n$ (negative image). The exact distance d is of secondary importance, as long as the conditions $d(\mathbf{w}^a, \mathbf{w}^p) \leq \alpha$ and $d(\mathbf{w}^a, \mathbf{w}^n) > \alpha$ are satisfied for the triplet's embeddings $\mathbf{w}^a$, $\mathbf{w}^p$ and $\mathbf{w}^n$. By further enforcing $\|\mathbf{w}\|_2 = 1$, all embeddings are constrained to lie on a hypersphere [218].

Note that margin α provides the network with the flexibility not to project all images of the same identity to a single point in the embedding space, thereby allowing them to 'live on a manifold' [218]. This is less complex to learn and sufficient for re-ID systems [254]. Mathematically speaking, the feature extractor is optimized to find embeddings such that the following triplet constraint holds for all possible triplets $\mathcal{T}$ [218]:

$$\|\mathbf{w}_i^a - \mathbf{w}_i^p\|_2^2 + \alpha < \|\mathbf{w}_i^a - \mathbf{w}_i^n\|_2^2, \ \forall(\mathbf{w}_i^a, \mathbf{w}_i^p, \mathbf{w}_i^n) \in \mathcal{T} \,. \tag{2.5}$$

The loss to be minimised is then given by

$$L_{\text{triplet}} = \sum_i^N \left[\|\mathbf{w}_i^a - \mathbf{w}_i^p\|_2^2 - \|\mathbf{w}_i^a - \mathbf{w}_i^n\|_2^2 + \alpha\right]_+ , \tag{2.6}$$

where N is the cardinality of $\mathcal{T}$ [218].

In practice, it is neither possible nor desirable to use all triplets for training. Triplets that already satisfy the condition in eq. 2.5 provide no learning signal for the feature extractor and slow down the training process. Conversely, for a given anchor $\mathbf{X}^a$, images $\mathbf{X}^p$ and $\mathbf{X}^n$ have the greatest learning potential if they most significantly violate the constraints:

$$\underset{\mathbf{X}_i^p}{\operatorname{argmax}} \|\mathbf{w}_i^a - \mathbf{w}_i^p\|_2^2 \text{ (hard positive)} \tag{2.7}$$

and

$$\underset{\mathbf{X}_i^n}{\operatorname{argmin}} \|\mathbf{w}_i^a - \mathbf{w}_i^n\|_2^2 \text{ (hard negative)} \tag{2.8}$$

respectively [218]. The downside of training with hard triplets is twofold. First, identifying the hard triplets is computationally expensive. Second, there is a risk that the training process may be dominated by mislabelled or poorly imaged samples, which can adversely affect the model's performance [218]. To address these issues, the authors chose to sample triplets not from the entire dataset but from within a minibatch. To ensure sufficient diversity in the minibatch, the batch size is in the order of thousands. Enforcing a minimum number of images per class ensures that both positive and negative samples are included. While training on all available positive samples, only hard negative samples were considered, which was found to be sufficient [218].

2.3.3 Metric Learning

Previously, we explored how representation learning can be used to find embeddings that align with a given metric. In this section, we consider the reverse scenario: how to fit a metric that aligns well with a given set of features.

Let $\mathbf{w} \in \mathbb{R}^d$ be a hand-crafted numerical representation for an image, where each entry describes a specific characteristic of the depicted object. Let $d(\bullet)$ denote a distance metric, which is a pairwise function that measures the (dis)similarity between objects [12]. The goal of metric learning is to fit a metric that makes optimal use of the given embeddings and is tailored to the specific problem at hand, such as re-identification. The downside of choosing $d_{\cos}$ or d_{L_2} is that all features encoded in the components of $\mathbf{w}$ equally impact the similarity $d.(\mathbf{w}_1, \mathbf{w}_2)$. Not accounting for the varying importance of different features may lead to suboptimal performance in tasks where some features are more relevant than others [11]. For example, if the goal is to compare bees based on their pollen-carrying trait, the feature describing pollen should be heavily weighted in the similarity metric. Conversely, when determining if two images depict the same bee, the pollen feature should be disregarded, as a single bee can be observed with or without pollen. Note the difference from the previous examples in sections 2.3.1 and 2.3.2, where $\mathbf{w}$ consisted of *learned* features rather than hand-crafted features. In representation learning, the learning algorithm implicitly weights features through adjustments to the feature extractor, making explicit weighting on the part of the metric unnecessary. In metric learning, however, the metric is fit to the features rather than the features to the metric.

Here, we focus on linear metrics that are learned using a weakly supervised learning paradigm. This approach relies on equivalence constraints, meaning it only knows whether pairs or triplets are similar or dissimilar, without any class label information [11]. A significant advantage is that equivalence constraints can sometimes be automatically gathered, for instance, through object tracking [126].

We start with the Mahalanobis distance [153], which measures the distance between a point and a distribution in units of standard deviations:

$$d_{\Sigma^{-1}}(\mathbf{w}, \mathbf{w}') = \sqrt{(\mathbf{w} - \mathbf{w}')^T \Sigma^{-1} (\mathbf{w} - \mathbf{w}')}\,, \tag{2.9}$$

where $\mathbf{w}$ and $\mathbf{w}'$ are embedding vectors from the same distribution with covariance matrix Σ [11]. We then replace the inverse of the distribution's covariance matrix with learnable parameters $\mathbf{M}$, and yield the generalised distance defined as

$$d_{\mathbf{M}}(\mathbf{w}, \mathbf{w}') = \sqrt{(\mathbf{w} - \mathbf{w}')^T \mathbf{M} (\mathbf{w} - \mathbf{w}')}\,, \tag{2.10}$$

where the positive semidefinite matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$ weights the features [11]. If $\mathbf{M}$ is equal to the identity matrix, one obtains the Euclidean distance. By expressing $\mathbf{M}$ through $\mathbf{L}^T\mathbf{L}$, where $\mathbf{L} \in \mathbb{R}^{k \times d}$ and k is the rank of $\mathbf{M}$, Bellet, Habrard, and Sebban [11] show that the Mahalanobis distance corresponds to the Euclidean distance after the data has been linearly transformed by $\mathbf{L}$:

$$\begin{aligned} d_{\mathbf{M}}(\mathbf{w}, \mathbf{w}') &= \sqrt{(\mathbf{w} - \mathbf{w}')^T \mathbf{M} (\mathbf{w} - \mathbf{w}')} \\ &= \sqrt{(\mathbf{w} - \mathbf{w}')^T \mathbf{L}^T \mathbf{L} (\mathbf{w} - \mathbf{w}')} \\ &= \sqrt{(\mathbf{L}\mathbf{w} - \mathbf{L}\mathbf{w}')^T (\mathbf{L}\mathbf{w} - \mathbf{L}\mathbf{w}')} \end{aligned}$$

There are various ways to learn $d_{\mathbf{M}}$ [126, 11]. However, they generally involve learning the quadratic form $d^2_{\mathbf{M}}$ rather than $d_{\mathbf{M}}$ given in eq. 2.10. In the following, we will present the method proposed by Köstinger et al. [126]. The authors model the dissimilarity between two embeddings as the *likelihood* of them being dissimilar. Statistically, the decision on whether a pair is dissimilar can be determined using a likelihood ratio test. In this test, the probability of a pair being dissimilar is compared to the probability of it being similar.

A pair $(\mathbf{w}_i, \mathbf{w}_j)$ is considered similar if the null hypothesis H_0 (dissimilar) is rejected. This occurs when $\delta(\mathbf{w}_i, \mathbf{w}_j)$ becomes small:

$$\delta(\mathbf{w}_i, \mathbf{w}_j) = \log\left(\frac{p(\mathbf{w}_i, \mathbf{w}_j|H_0)}{p(\mathbf{w}_i, \mathbf{w}_j|H_1)}\right). \tag{2.11}$$

By setting $\mathbf{w}_{ij} = \mathbf{w}_i - \mathbf{w}_j$, the authors make the metric independent of the locality in feature space and obtain:

$$\delta(\mathbf{w}_{ij}) = \log\left(\frac{p(\mathbf{w}_{ij}|H_0)}{p(\mathbf{w}_{ij}|H_1)}\right) = \log\left(\frac{f(\mathbf{w}_{ij}|\theta_0)}{f(\mathbf{w}_{ij}|\theta_1)}\right), \tag{2.12}$$

where $f(\mathbf{w}_{ij}|\theta_\cdot)$ is a probability density function indicating whether a pair $(\mathbf{w}_i, \mathbf{w}_j)$ is dissimilar, given the respective hypothesis. Since $\mathbf{w}_{ij} = -\mathbf{w}_{ji}$ (symmetry), the authors then assume a zero-centred Gaussian structure of the difference space with $\theta_1 = (\mathbf{0}, \Sigma_{\mathrm{eq(i,j)=1}})$ and $\theta_0 = (\mathbf{0}, \Sigma_{\mathrm{eq(i,j)=0}})$:

$$\delta(\mathbf{w}_{ij}) = \log\left(\frac{\left(2\pi \det\left(\Sigma_{\mathrm{eq}(i,j)=0}\right)\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{w}_{ij}^T \Sigma^{-1}_{\mathrm{eq}(i,j)=0} \mathbf{w}_{ij}\right)}{\left(2\pi \det\left(\Sigma_{\mathrm{eq}(i,j)=1}\right)\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{w}_{ij}^T \Sigma^{-1}_{\mathrm{eq}(i,j)=1} \mathbf{w}_{ij}\right)}\right), \tag{2.13}$$

where $\mathrm{eq}(i,j) = 1$ indicates the existence of a equivalence constraint for pair (i,j) and $\mathrm{eq}(i,j) = 0$ indicates the absence of such a constraint. The unscaled covariance matrices for the cases where the pairs are similar or dissimilar are calculated as follows:

$$\Sigma_{\mathrm{eq}(i,j)=1} = \sum_{\mathrm{eq}(i,j)=1} \mathbf{w}_{ij}\mathbf{w}_{ij}^T \tag{2.14}$$

and

$$\Sigma_{\mathrm{eq}(i,j)=0} = \sum_{\mathrm{eq}(i,j)=0} \mathbf{w}_{ij}\mathbf{w}_{ij}^T. \tag{2.15}$$

When applying logarithm rules to eq. 2.13, dropping constant terms, and rearranging, one obtains:

$$\delta\left(\mathbf{w}_{ij}\right) = \mathbf{w}_{ij}^{T}\Sigma_{\mathrm{eq}(i,j)=1}^{-1}\mathbf{w}_{ij} - \mathbf{w}_{ij}^{T}\Sigma_{\mathrm{eq}(i,j)=0}^{-1}\mathbf{w}_{ij} \tag{2.16}$$

$$= \mathbf{w}_{ij}^{T}\left(\Sigma_{\mathrm{eq}(i,j)=1}^{-1} - \Sigma_{\mathrm{eq}(i,j)=0}^{-1}\right)\mathbf{w}_{ij}\,, \tag{2.17}$$

which is the quadratic form of eq. 2.10.

If we estimate $\tilde{\mathbf{M}} = \left(\Sigma_{\mathrm{eq}(i,j)=1}^{-1} - \Sigma_{\mathrm{eq}(i,j)=0}^{-1}\right)$ and ensure $\tilde{\mathbf{M}}$ is positive semidefinite through spectrum clipping, we finally obtain $\hat{\mathbf{M}}$ and thus the learned metric that determines the similarity between pairs of embeddings [126]:

$$d_{\hat{\mathbf{M}}}^{2}\left(\mathbf{w}_i - \mathbf{w}_j\right) = \left(\mathbf{w}_i - \mathbf{w}_j\right)^{T}\hat{\mathbf{M}}\left(\mathbf{w}_i - \mathbf{w}_j\right)\,. \tag{2.18}$$

In this section, we have seen how equivalence constraints can be used to fit a metric that appropriately weights a set of fixed features.

2.3.4 Evaluation of Re-Identification Algorithms

It is not clear how to evaluate the performance of re-ID algorithms, as different approaches may focus on different aspects of the task. For example, a system designed to find the top-k most similar images may perform poorly in a scenario where the user is only interested in the top-ranked image. To address this, the re-ID community has developed several evaluation metrics, including the Cumulative Matching Characteristics curve (CMC) [263] and mean average precision (mAP) [263]. Both are briefly introduced below.

The Cumulative Matching Characteristics curve is an easy-to-interpret method for evaluating re-ID algorithms. In the context of computer vision, the term CMC-k represents the probability that a query identity (probe) will appear within the top-k of the ranked candidate list [263, 256]. The candidate list is generated by the re-ID algorithm and contains images from the gallery set, ranked in accordance with the matching probability assigned by the algorithm.

The CMC calculations take into account only the first match in the gallery, regardless of the number of true matches. This makes it impossible to assess the recall ability of the algorithm. While this is an acceptable approach if the

user's main concern is to find a match at the top of the ranking, it is otherwise undesirable [263].

The mean Average Precision metric (mAP) is capable of differentiating between algorithms that have the initial correct match in the same position, by additionally considering subsequent matches and their respective positions [256]. The mAP is calculated by averaging the area under the Precision-Recall curve for each query, as detailed in reference [263].

3 Apidology and Precision Beekeeping

This chapter centres on the honey bee (*Apis mellifera*), which is the main subject of this work. Bumblebees (*Bombus terrestris*), in contrast, are of secondary importance and are only addressed within the context of re-identification in Chapter 6.

Consequently, Section 3.1 highlights the importance of research on pollinators with an emphasis on honey bees. Section 3.2 discusses the biology of honey bees, focusing on aspects relevant to the applications and technologies examined, including behavioural traits and pathogens. Finally, Section 3.3 covers technological interventions designed to monitor and support bee populations, referred to as 'Precision Beekeeping' in the literature [98, 258].

3.1 Relevance of Pollinators and Their Decline

Pollinators, particularly bees, are crucial to both ecological health and agricultural productivity. They facilitate the reproduction of many plant species by transferring pollen, which significantly contributes to biodiversity and ecosystem stability. It is estimated that 87.5 % of all flowering plant are animal pollinated [184].

According to the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services [21], the volume of crops dependent on pollinators has tripled over the past fifty years, with between 5 % and 8 % of global crop production now reliant on them. This contribution is crucial, as it directly influences the availability and diversity of food products such as fruits, vegetables, and nuts. Estimates suggest that the total loss of pollinators could lead to a short-term reduction of 1 –2 % in global GDP due to potential decreases in crop yields [143].

Despite their importance, pollinators are experiencing significant declines worldwide [17, 172, 179]. Regional assessments reveal alarming threats to insect pollinators. The European Red List indicates that 9 % of European bee species are threatened with extinction, and 37 % are in decline [179]. A widely cited study by Hallmann et al. [100] reports a 76 % decrease in the biomass of flying insects in German nature reserves between 1989 and 2016.

Several factors contribute to the decline of pollinators. Habitat loss due to intensive agriculture and urbanisation is a primary driver, alongside pollution from synthetic pesticides and fertilisers [214]. Biological threats, such as pathogens and introduced species, also play a critical role. For instance, Varroa mites have severely impacted honey bee populations [208]. Similarly, bumblebees face threats from parasites and fungal diseases [21].

Climate change further exacerbates these problems by altering habitats and affecting the availability of food sources for pollinators. The interaction of these factors results in a decline in pollinator populations, which has direct consequences for agricultural productivity and global food supply.

The honey bee plays a special role in the study of the causes of pollinator decline. Although it is not itself affected by the decline – in fact, FAO data shows that global honey bee populations have nearly doubled since 1960 due to the large number of managed hives – it serves as a well-studied and widely available model organism for researchers [172]. For example, regulatory measures for pollinator protection are often defined based on their impact on honey bees. Such read-across is justified with additional safety margins and provide a pragmatic solution to effectively protect more vulnerable pollinators as well [78].

To better understand this model organism in the context of this work, the following section provides a short introduction to the biology of honey bees.

3.2 Honey Bee Biology and Threats

Karl von Frisch once remarked, ‘A farmer can keep a single cow, a dog, or even a chicken if he wants, but he cannot keep a single bee.’ [87], highlighting the complex social structure of honey bee colonies that fascinates many. This section introduces the biology of the honey bee (*Apis mellifera*), drawing on von Frisch’s foundational work ‘Aus dem Leben der Bienen’ [87].

In the wild, honey bees naturally inhabit cavities, but managed honey bee colonies are typically kept in hives. One type of hive being used in this work is the Dadant magazine hive, which measures 509×509 mm externally. Inside, there are up to 12 frames fitted with wax foundation sheets, which the bees build into combs. In Chapter 9, we will show how such a frame can be replaced by a commercial flatbed scanner, allowing precise observation of brood cells. The hives are usually made of wood with a metal roof. The brood chamber is at the bottom and is stacked with one or more honey supers during the nectar flow. These hives are designed to provide optimal conditions for the bees while allowing for easy inspection and management.

Honey bees produce the wax they need to build brood and honeycombs through wax glands. The hexagonal shape of the comb cells provides stability and, compared to other shapes that also avoid gaps, it has the advantage of requiring the least perimeter for the same area, thus minimising the amount of building material needed. Wax usage is further reduced by building cell walls with a thickness of just 0.1 mm. In Chapter 9, we will explore how bees can also accept plastic foundations as a basis for their combs.

Once a brood comb is fully constructed, the queen begins laying eggs. During the peak of colony growth in the spring, she can lay up to 1 500 eggs per day, which is roughly equivalent to her own body weight. Naturally, the number of bees that perish daily must be of similar order. In Chapter 5, we will theoretically explore how to accurately determine the mortality rate within a bee colony. The worker bees prepare the cells for egg-laying, with worker brood cells measuring 5.25 ± 0.5 mm, slightly smaller than the 6.4 ± 0.5 mm drone cells [151]. Based on the size and need, the queen decides whether to lay unfertilised eggs, which develop into the larger male drones, or fertilised eggs, which produce new worker bees. In Chapter 9, we utilize this behaviour to attract and observe drone brood on the aforementioned flatbed scanner.

Three days after an egg is laid, it develops into a white larva, which is then fed by the worker bees. After six days, the larval stage is complete, and the cell is sealed. The larva pupates, and after 12 more days, a young worker bee emerges. The development of a queen is slightly shorter, taking 16 days, while drone development (see Figure 3.1) takes a bit longer at 24 days and is detailed in connection with pathogens. A newly emerged queen must first embark on a mating flight, during which she mates with a few dozen drones. The sperm she collects during this flight will be used for the rest of her life to fertilize eggs.

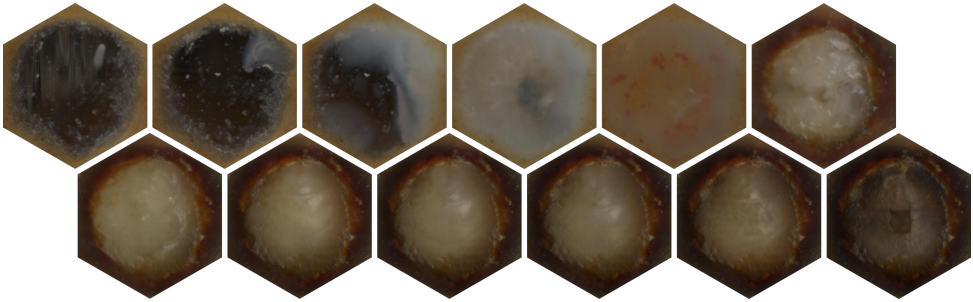


Figure 3.1: Starting from day 2, the cell scans show 24 days of drone brood development at 48-hour intervals, as seen from below. Capping is expected to begin around day ten.

Honey bees rely on sugar-rich nectar for energy and protein-rich pollen for growth, both of which are collected on foraging flights, to which a separate section is devoted. In late summer, when the honey flow ceases, bees may begin robbing weaker colonies of their honey stores. This behaviour can significantly weaken already vulnerable colonies. It is often triggered by the scent of honey, which can become particularly enticing if beekeepers leave feeding frames exposed. In Chapter 5, we will artificially induce robbing in an uninhabited hive. This method introduces a controlled scenario which allows us to evaluate the accuracy of bee counters.

The lifespan of worker bees depends on their level of activity. Spring bees, which are heavily involved in construction and care, live for only about three weeks. Summer bees have a lifespan of about 40 days, while winter bees survive the entire winter period, which lasts 6-7 months. A queen, on the other hand, can live for 4-5 years. At its peak, the population of a honey bee colony exceeds 40 000 individuals, including a single queen and, in the early summer months, a few hundred drones. As winter approaches, colony size shrinks to about a quarter.

3.2.1 Corbicular Pollen and Foraging Behaviour

Corbicular pollen loads (or pollen baskets) measure approximately 3.5 mm in length and 2 mm in width [124]. Pollen foragers gather pollen from flowers, forming it into packets that adhere to their hind legs with the help of spe-

cialised hairs. Upon returning to the hive, these pollen packets are deposited in separate cells from the nectar.

Pollen is a vital protein source for the developing brood, making it indispensable for the colony's growth [40]. A diverse pollen diet is crucial for bees, as the nutritional composition of pollen can vary significantly. For example, the protein content of cypress (*Cupressus arizonica*) is 2.4 %, whereas that of primrose (*Dodecatheon clevelandii*) is as high as 61.7 % [211].

While foraging, bees inadvertently transfer pollen between flowers, thus playing a crucial role in plant pollination. Though many insects contribute to pollination, honey bees are particularly effective due to their large population size, wide foraging range, and the fact that they continue to collect beyond their immediate needs, preparing for future scarcity [87, 124].

The median foraging distance for a honey bee is around 1.7 km, and multiple flights are performed per day [249]. In spring and summer bees collect nectar, which is then processed, dehydrated, and stored as honey. This honey, enriched with enzymes by the bees' processing, serves as a long-lasting food reserve, essential for surviving the winter months when resources are scarce.

Analysing 114 publications, Keller, Fluri, and Imdorf [117] found for Europe that the majority of pollen collected came from just a few plant species. The five most common pollen sources, namely maize, clover, dandelion, plantain and rape, accounted on average for more than 60 % of the total pollen collection. Interestingly, the authors observed that hives in the same apiary often collect similar pollen but in different proportions [117]. Honey bees exhibit a remarkable behaviour known as floral constancy, where they consistently collect pollen from the same flower species until the source is depleted. This behaviour, which optimizes foraging efficiency, was thoroughly described by Knoll [124]. Research by Newstrom-Lloyd, Raine, and Li [174] further highlights this phenomenon, showing that honey bees, when observed foraging, carry pollen from a different species only 5 % of the time (see also Percival [186]).

A unique aspect of honey bee foraging is their ability to communicate resource locations to nest mates through the waggle dance [89]. This social communication allows the colony to coordinate and optimize foraging efforts, unlike other pollinators such as bumblebees, which must individually explore and evaluate new foraging sites. As Visscher and Seeley [249] explain,

'through recruitment communication, a honey bee colony operates as an information center in which the reconnaissance of foragers is pooled and processed to focus the colony's efforts on the best forage sites found about the nest.' [249] This collective intelligence allows honey bees to make informed decisions, further enhancing the efficiency of their foraging strategy.

The strong association between bees and specific flowers is driven by colour and scent, leading to the formation of pure pollen packets [124]. In Chapter 7, we leverage this behaviour to assess biodiversity within the foraging area by analysing the colour of pollen loads.

The following two sections will discuss the main pathogens affecting honey bees: chalkbrood and Varroa.

3.2.2 Chalkbrood

Chalkbrood is a widespread infection caused by the fungus *Ascosphaera apis*, which only affects honey bee brood [4]. The disease is easily identified by the larvae, which are surrounded by a white fungus. After some time, the larvae dry out, leaving a hard, chalky mass, hence the name chalkbrood [103].

Although fatal to individual larvae, the disease is usually not lethal to the colony as a whole [4]. Adult bees are not susceptible, but they and beekeepers can spread the fungus within and between colonies. According to Aronstein and Murray [4], *Ascosphaera apis* can infect worker, drone and queen brood. In case of moderate infection, the infected brood will be detected and removed from the hive by the workers, where it can be found during colony inspections. This hygienic behaviour is the most important factor in keeping the disease under control [4].

3.2.3 Varroa

The 1-2 millimetre parasitic mite, *Varroa destructor*, has dictated the actions of 600 000 European beekeepers since the 1970s. Without their intervention, most of the 21 million managed honey bee colonies would inevitably face collapse [208, 86].

The delicate balance between the Varroa mite and its host was disrupted when colonies of the Western honey bee, *Apis mellifera*, were introduced to

East Asia in the mid-20th century. Originally, the parasite thrived on the Eastern honey bee, *Apis cerana*, reproducing solely in the seasonally limited drone brood. However, after switching hosts, the mite acquired the ability to also reproduce within the worker brood of the Western honey bee. This shift destabilised the equilibrium, and the parasite began decimating its new host in large numbers [86, 62, 208, 259].

Similar to a tick feeding on human blood, the mite feeds on the fat body of honey bees [194]. Its bite transmits various diseases and weakens the brood by reproducing within sealed cells [241]. The result has been significant colony losses and extensive damage. This issue is now global, affecting all continents except Antarctica and minor isolated places [167, 97, 175, 25]. According to the New South Wales government's situation report[1], the first varroa mites in Australia were detected by Bee Biosecurity Officers at the Port of Newcastle in 2022, despite strict biosecurity measures aimed at preventing their incursion. In August 2024, the first mite was found outside the initial containment district, highlighting its rapid spread. There are several reasons why the ectoparasitic mite is (one of) the most damaging enemies of the Western honey bee. This includes the imbalance between host and parasite, the rapid and worldwide spread and the fact that most bee colonies would collapse without intervention of the beekeeper [208, 110].

3.2.3.1 Life Cycle

The life cycle of the Varroa mite is closely linked to both adult bees and the developing larvae in brood cells. A mite spends part of its life on adult bees, which it uses to move within and between hives (phoretic phase, or more accurately, the dispersal phase [241]), and another part of its life on brood (reproductive phase) [208].

During the dispersal phase, the female Varroa mite attaches itself to an adult bee, usually clinging to the bee's thorax or abdomen. During this phase, the mite feeds on the bee's fat body tissue [194] and also uses the bee as a means of transport. The mite relies primarily on chemical cues to locate a host, although it can also detect differences in brightness and vibration [208].

[1] *NSW Government Australia*, `https://www.dpi.nsw.gov.au/emergencies/biosecurity/current-situation/varroa-mite-emergency-response`. Accessed September 13, 2024.

The reproductive phase begins when the mite enters a brood cell containing a larva in its final stage of development. This occurs approximately 15-20 hours before the cell is capped in worker brood and 40-50 hours before capping in drone brood [27]. The mite hides at the bottom of the cell in the larval food, possibly as a strategy to avoid detection by hygienic bees [108].

Once the cell is capped, the mite becomes active again. About five hours after the larva has consumed enough food to release the mite, the reproductive cycle begins. 60-70 hours after capping, the foundress lays her first unfertilised egg that is placed under the cell cap to avoid injury or entrapment by the pupating larva [73]. Subsequent fertilised eggs are laid at intervals of about 30 hours, progressively lower down the cell wall [264]. In total, a mother mite can produce up to five daughters in worker cells and up to six in the longer-capped drone cells [91].

Female offspring takes about 5.8 days to develop, while male offspring takes on day longer. Each offspring goes through two nymphal stages: the protonymph and the deutonymph. Both stages include an immobile phase, during which the mite moults, and a mobile phase, where it actively feeds and moves within the cell [73].

The newly hatched mites feed at a feeding site created by the mother mite, typically located in the lower quarter of the cell. This feeding site is established by a bite into the larva, which serves as the primary feeding point for both the mother and her offspring. Interestingly, the feeding site is often located near the faecal accumulation site, where the mites frequently gather and mate [74, 73]. This otherwise hidden behaviour was captured in detail using the scanner method described in Chapter 9.

Under natural conditions, a mother mite usually undergoes about two reproductive cycles within capped brood cells [86, 157]. However, in artificial conditions, up to seven reproductive cycles have been observed [212]. This reproductive capability, coupled with the mite's ability to exploit both worker and drone brood, makes Varroa destructor a 'formidable foe' [241] of the Western honey bees.

3.2.3.2 Monitoring

The following section introduces both manual and automated methods for monitoring and measuring Varroa mite infestation in honey bee colonies.

Figure 3.2: Systematically opened drone brood in search for Varroa mites.

Generally, the more invasive a method, the more accurate the measurements. However, invasive methods cannot be repeated frequently, or at all in some extreme cases. The monitoring methods described are sorted by invasiveness and grouped in phoretic, reproductive and combined mite assessment.

One of the simplest and most popular methods among beekeepers for estimating natural mite infestation involves placing a diagnostic board beneath the hive. Over several days, the number of dead mites collected on the board is counted, providing an indirect estimate of the colony's mite load. It is crucial to ensure the board is secure from both bees and ants, as ants may remove the mites, leading to inaccurate results [65, 38]. Using an oil-soaked cloth can prevent this and yield more reliable data [71]. The results must be interpreted in relation to the colony size, as the damaging threshold depends on the number of bees in the colony. Although this method does not provide an exact count of mites, it offers valuable insight into the infestation level.

A more direct method for assessing the infestation level of adult bees requires a sample of nursing bees to be dusted with a few teaspoons of powdered sugar. This causes the mites to detach and enables them to be counted without causing significant harm to the bees [150]. Alternatively, live bees can be submerged in alcohol, which kills the bees and dislodges the mites, allowing them to be sieved and counted [71].

Manually opening capped brood cells, as shown in Figure 3.2, enables the assessment of Varroa infestation within the brood. In addition to detecting mites and their offspring, the presence of faeces in the cells serves as a reliable indicator of mite infestation [71].

The accuracy of these methods is influenced by the overall mite abundance since practical sample sizes may not accurately capture rare events. To obtain

reliable results, over two percent of the brood (or bees) should be infested [71]. Given that only a small fraction is typically infested, the literature recommends a combined inspection of more than 200 brood cells and 200 adult bees for robust population size estimates [38].

By determining both the relative infestation levels of adult bees and brood, as well as the total number of adult bees and brood cells, it is possible to estimate the absolute number of mites in the colony.

Without the need for extrapolation, the absolute infestation level of a colony can be determined by combining an effective acaricide treatment with the previously described diagnostic plate method. This approach, while accurate, disrupts the continuity of the experiment [71]. For an even more definitive, albeit highly destructive, method, the entire colony can be frozen, thereby killing all the bees. The mites can then be washed off with alcohol and counted [71].

In addition to these traditional methods, computer vision techniques have been applied to count phoretic mites on adult honey bees [18, 23, 221, 180]. However, these methods only detect *visible* mites on the exterior of the bees. Their practical relevance is limited because they do not account for mites hidden beneath the bee's sclerites [194]. Relying solely on visual inspection for Varroa infestation on adult bees is cautioned against, as this approach may not accurately reflect the true mite load [110]. While computer vision can identify a significant proportion of phoretic mites, its application in real-world scenarios overlooks concealed mites, which are crucial for effective management strategies.

In contrast, in Chapter 9, we propose using contact image sensors placed beneath the brood cells to detect mites during the larval and pupal stages. This approach offers two key advantages. First, Varroa mites are confined to the brood cell for several weeks, increasing the likelihood of detection. Second, Varroa mites exhibit sexual dimorphism [107], meaning that male and female offspring differ significantly in form and colour, allowing for the estimation of both sex and age. This method holds promise for a more detailed understanding of infestation levels and the reproductive process, particularly in longitudinal studies that were previously difficult to achieve.

3.2.3.3 Control

Due to drifting between colonies, robbing behaviour, and collapsing colonies, hives that are not isolated are constantly exposed to reinfestation. For this reason, the primary goal of Varroa control is not the complete elimination of the parasite, but rather to keep the infestation level below the damage threshold [110]. There is no 'single best method' [246] for achieving this goal; instead, a situationally dependent combination of strategies is recommended [246].

In a survey of 2 238 UK beekeepers, 21 % reported that they were not treating their colonies for Varroa. Among the majority that treated their colonies, approximately 84 % used chemical methods [245]. Among non-chemical treatments drone brood removal [55] was the most widely practised, accounting for 5.4 % of beekeeping practices. This method involves inserting empty frames into the hive, which the colony readily uses in spring to rear larger drone brood. Since Varroa mites have a preference for drone brood, the infestation in these cells is particularly high. By waiting until the brood is capped, the frame can be removed along with both the brood and the mites. The colony continues to maintain enough drones to avoid long-term harm, while effectively delaying the development of the Varroa population [55].

Previous chemical treatments have shown to cause resistance in Varroa mites [113, 206, 202, 242], and those still in use can also pose risks to the bees [246]. If the fear of the mites were not greater than the concern about side effects [246], treatments would have been discontinued long ago. The widespread use of oxalic acid [245] does not penetrate sealed brood cells, which limits the timing of its application. Formic acid, while effective, can damage the colony if not dosed correctly [246]. Thermal treatments, such as temporarily heating brood frames, are difficult to implement. The temperature range that is harmful to mites but tolerable for bee brood is very narrow, and the energy requirements make field application challenging [246].

In conclusion, beekeepers have managed to reduce Varroa pressure through various measures, but a definitive solution or ideal treatment remains elusive. One promising approach could be the development of self-help strategies, such as breeding bees with enhanced grooming, hygienic behaviour, and Varroa-sensitive hygiene traits. These behaviours could enable bees to protect themselves against the parasite through social immune defence [241].

3.2.3.4 Hygienic Behaviour and Varroa Sensitive Hygiene

The behaviour of worker bees in locating and addressing problematic capped cells is known as hygienic behaviour [166]. Specifically, when worker bees selectively remove larvae from cells infested with Varroa mites, this is referred to as Varroa-sensitive hygiene (VSH) [166, 241]. To assess general hygienic behaviour traits, two common techniques are employed: the pin-killing test and the freeze-kill brood assay [137]. These methods involve deliberately killing capped brood, either by piercing or freezing, and then measuring the percentage of cells cleaned out by worker bees on the following day [166].

Testing for the more specific VSH behaviour in the field involves prohibitively complex procedures, including manual infestation and observation of individual cells [241, 166]. Alternatively, genetic analysis has been used to evaluate colonies for this trait [22]. Although genetic analyses are increasingly common, they remain accessible primarily to specialists.

While breeding is a promising approach and beneficial traits can be identified in the genetic makeup of bees, measuring the success of breeding programmes remains challenging.

3.3 Hive Monitoring

The trend towards digitalisation in beekeeping, known as precision beekeeping, has already led to the development and application of various monitoring tools. These include systems for tracking swarming behaviour [193], humidity [93], temperature [127], flight activity [61, 159], pollen intake [176, 6], and phoretic mites at the hive entrance [23, 220] or within the hive [162]. Additionally, patterns of in-hive social behaviour have been studied using observation hives [92, 119].

These advancements utilise a range of sensor technologies, including acoustic sensors [42], visual sensors [19], weight sensors [161], chemical sensors, RFID tags [61], light barriers [118, 190, 187], and vibrational sensors [14].

In a review by Bilik et al. [19], 50 automated beehive monitoring applications focusing on computer vision were presented. The advantage of camera-based approaches lies in their minimal invasiveness and the potential to perform multiple detections simultaneously, such as identifying castes, monitoring

activity, measuring pollen influx, and detecting Varroa infestations [156]. Danieli et al. [66] evaluate precision beekeeping practices, highlighting both their strengths and weaknesses. Challenges such as real-time data analysis and the availability of communication and power supply still need to be addressed. While basic information like temperature, humidity, and weight is important, these metrics require data processing and fusion to effectively support beekeepers and become truly valuable [66]. Such integrated insights could potentially surpass the knowledge gained from routine on-site inspections in the future [66].

3.4 Honey Bees as Biosensors

Some studies monitor bee colonies not for the sake of the colony itself, but to use the bees directly or indirectly as biosensors.

Since honey bees can be conditioned to respond to chemical cues, Bromenshenk et al. [44] suggest their use as biosensors for detecting chemicals and have demonstrated their effectiveness in landmine detection [43].

Milla et al. [163] highlight the importance of ground vegetation surveys, noting that climate change and other factors are causing species extinctions and changes in plant phenology and distribution. However, biomonitoring is both time-consuming and expensive. Therefore, the authors tested whether ground vegetation surveys could be 'conducted' or 'assisted' by bees. Bees were considered as biosensors that were 'read out' by DNA metabarcoding of the collected pollen. It was shown that more taxa were found by DNA metabarcoding than by human surveys in the field. This approach, substituting metabarcoding with computer vision, has also been adopted by others [240, 31] and was automated in Chapter 8.

Part II

Contributions in Apidology

4 Data Acquisition

The application of computer vision to beehives is not entirely new; however, a significant obstacle remains the lack of well-designed, high-resolution datasets [19]. In the following, we present insights gained from working with publicly available datasets over the last few years and, as a preface to the practical part of this work, we introduce a new system specifically designed to produce high quality datasets for computer vision research in apidology.

Examining image data captured at the entrance of beehives reveals several limitations in current acquisition methods, including inconsistent illumination, blurriness, shadows, resolution, and colour fidelity (see Figure 4.1). We will evaluate the factors that contribute to better image quality and propose improvements for a more sophisticated camera setup.

When comparing the different images, it becomes apparent that a key distinguishing factor in the setup is the use of a transparent barrier that prevents the bees from flying. The authors of Figure 4.1b and 4.1e opted not to implement this measure, which can lead to variations along the z-axis, resulting in defocused images. Additionally, bees in flight are more prone to motion blur and can easily be mistaken for their shadows.

Another advantageous design choice is the implementation of active lighting. Active lighting helps to minimize shadows (Figure 4.1d) and reduces brightness gradients caused by incident light (Figure 4.1a). Adequate illumination is crucial for achieving a high frame rate, or more specifically, a short shutter speed. A short shutter speed reduces motion blur from rapid movements and bees in flight. If the objective is to analyse image colours, such as identifying pollen colours, consumer-grade cameras are inadequate due to automatic colour corrections, white balancing, and other uncontrollable image enhancements. Therefore, maintaining consistent lighting conditions and image acquisition is vital, as well as ensuring that the image sensor is properly calibrated to the existing lighting environment.

Figure 4.1: Image material from five hive observation cameras. From (a) to (e) as published in [205], [131], [53], [35], and [136].

The higher the initial quality of a dataset, the less severe the impact of degradation during data adaptation and augmentation. In the following sections, we outline the hardware and software considerations that lead to a setup specifically tailored to the needs of computer vision research in beehives.

4.1 Hardware Considerations

Figure 4.2 presents the setup of the proposed camera system. For Camera (1), a Basler acA3088-57uc industrial camera was selected, featuring an IMX178 sensor with 6MP resolution and a maximum frame rate of 59 fps. The image was cropped directly on the camera to 2736×1024 pixels to minimise unnecessary data transfer and reduce the need for software-based post-processing. The lens is an 8 mm RICOH lens with manual aperture and focus control. A lens-specific vignetting correction was applied internally by the camera. Short shutter speeds of 1.3 ms ensure sharp images, even when bees (2) are in motion.

By restricting the height of the passage in the hive, not only can the camera focus be precisely adjusted, but the bees also move more slowly, which can simplify downstream tasks. To limit the height of the passage into the hive, we selected museum glass (3), which offers excellent properties in terms of reflection and light transmission.

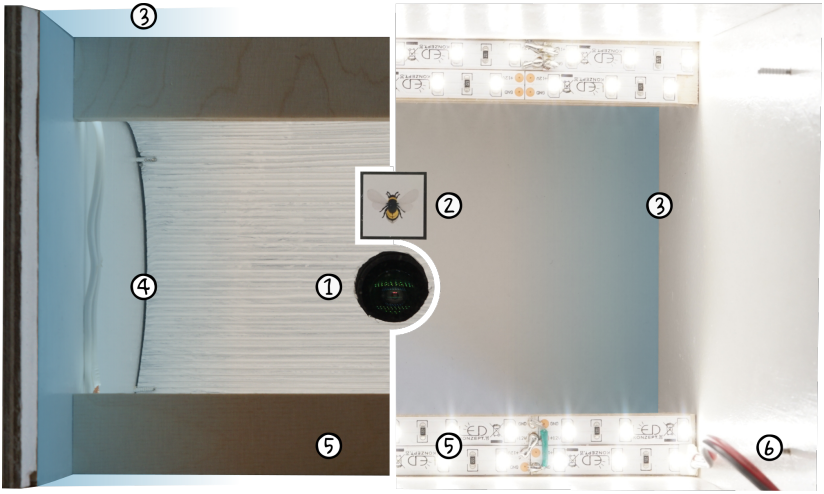


Figure 4.2: Left: Hardware setup from the bee's perspective. Right: Hardware setup from the camera's perspective. (1) Camera lens, (2) Object of study, (3) Sketched glass, (4) Diffuser, (5) LED and LED cover, (6) LED power supply.

For even lighting, a diffuser (4) made of matte white corrugated cardboard spans the LED strips (5, right) in a parabolic shape. The LED cover (5, left) ensures that only indirect, diffused light illuminates the scene. The light itself comes from LEDs with a colour rendering index of over 95 %, powered by a 12 V power supply (6).

The camera housing shown in Figure 4.3 is made of 1 cm thick plywood protected with waterproof clear varnish. The modular design of the exterior (left) and interior (right) allows easy access and maintenance. A fan cools the camera, which can become hot when exposed to intense sunlight. The passage from the flight board to the observation area is lined with black Molton fabric and angled to prevent direct stray light. As bees come into contact with the glass, daily cleaning is recommended to prevent the transparent surface from becoming cloudy.

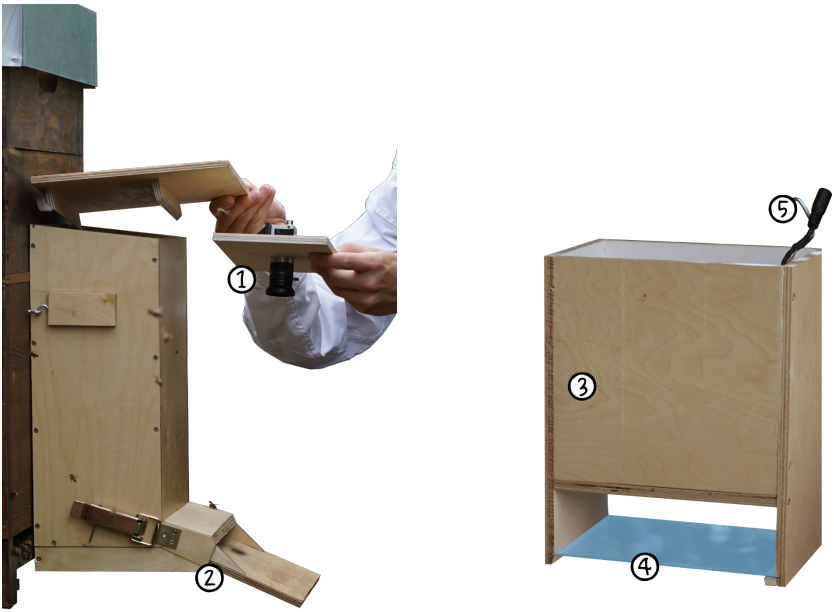


Figure 4.3: (Left) Camera (1) is placed inside the housing and connected to a controller that is placed in an empty super (not shown). The flight board (2) is angled to prevent direct light from reaching the scene. (Right) The interior of the housing (a) includes a glass height restrictor (4) and the power supply for the LEDs and fan (5). When used, the entire structure (b) is placed inside of (a).

4.2 Software Considerations

The camera generates vast amounts of data that must be efficiently encoded in real-time, making the use of a GPU essential. To ensure maximum image quality, the compression level was set so that virtually no compression artefacts are noticeable. The data is stored in the YCbCr422 pixel format as H.264 encoded videos, each 30 seconds long. Splitting the videos into short sequences facilitates handling and makes the system less prone to errors.

Figure 4.4: A full 2736 × 1024 pixel frame acquired by the camera, demonstrating uniform illumination and no vignetting.

4.3 Results

This section outlines how the device was used to assemble a new dataset, intended to supersede or complement previous ones.

To obtain images of individual bees, single specimens were confined within the camera's field of view for 5 minutes. Figure 4.4 shows the entire field of view, demonstrating uniform illumination across the area with no interference from daylight. Simple background subtraction was used to identify and crop the bee from the image, as illustrated in Figure 4.5. To ensure consistent quality, bees at the edges were excluded. The dataset, including trajectory information, caste details, and pollen presence, has been made available to the community in full resolution [28]. A total of 76 worker and 11 drone specimens were filmed, yielding 157 529 cropped and centred 512 × 512 pixel images organised into 3 430 trajectories.

(a) **(b)**

Figure 4.5: Exemplary cropped and centred images from the dataset, measuring 512 × 512 pixel: (a) Image of a drone. (b) Image of a forager bee walking upside down along the glass.

5.

Parts of this chapter are based on

HONEY BEE COUNTER EVALUATION – INTRODUCING A NOVEL PROTOCOL FOR MEASURING DAILY LOSS ACCURACY [35]

Parzival Borlinghaus[1], Richard Odemer[2], Frederic Tausch[3], Katharina Schmidt[3], Oliver Grothe[1]

[1]Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, Institute for Bee Protection, Braunschweig, Germany

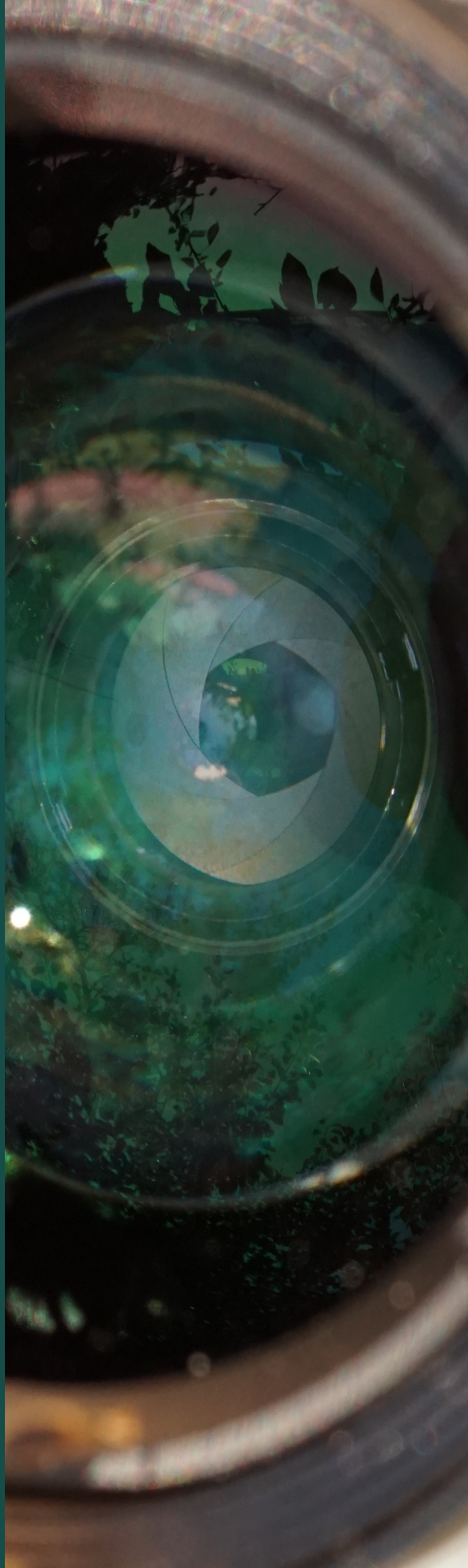
[2]apic.ai GmbH, Karlsruhe, Germany

CRediT

Conceptualisation: PB, RO, FT; Data curation: PB, RO, FT; Formal analysis: PB; Funding acquisition: KS; Investigation: PB, RO; Methodology: PB; Project administration: PB, KS; Resources: RO, FT, KS; Software: PB, FT; Supervision: OG; Validation: PB; Visualisation: PB; Writing - original draft: PB; Writing - review & editing: RO, FT, KS, OG.

Publisher

Computers and Electronics in Agriculture published by Elsevier B.V., vol. 197, p. 10657, Jun. 2022, doi: 10.1016/j.compag.2022.106957.



5.1 Introduction

Bee pollinators are not only of great importance for ecosystems [184] but also make a major economic contribution to us through their ecosystem services [21]. This pollinator group, however, has been suffering from a decline for quite some time, affecting already 37 % of European bee species [17, 172, 179] and, posing a threat to global food security and nutrition. Honey bees have not been affected by this decline, and the number of bee colonies worldwide has almost doubled since they were recorded in the 1960s [81]. Regardless, they represent an essential part of pollination ecology and, in part, have become an important model organism for investigating the drivers of bee decline [172].

Because pesticides are suspected of being one of these drivers, the European Food Safety Authority (EFSA) was tasked with developing more stringent guidelines for the risk assessment of pesticides [76, 78]. Since '[t]he viability of each colony, the pollination services it provides, and its yield of hive products all depend on the colony's strength and, in particular, on the number of individuals it contains' [78], EFSA has formulated specific protection goals aligned to colony size and mortality. Thus, characteristics worth protecting can be collectively assessed by evaluating daily bee losses. However, high-quality data on background mortality, i.e. the natural daily loss of (forager) bees are scarce.

To monitor even small changes in bee mortality, tools are needed that can accurately count bees entering and leaving the hive. For the vast majority of electronic bee counters, determining daily losses is, thus, of primary interest [181]. Since foragers usually fly only during the day and return to the hive at dawn, observation intervals of 24 hours are well suited and suggested by EFSA as a reference [78].

The ability to accurately count bees is not only important from a regulatory perspective, but also relevant to scientists, beekeepers and other ecotoxicological issues. Beekeepers are provided, for example, with the means to evaluate the quality of sites, to observe swarm behaviour and colony development. If abnormalities occur, counting data helps to find the underlying causes. Pham-Delègue et al. [188] call for sublethal effects, such as sensory impairment and associated effects on foraging behaviour, to be routinely recorded in addition to lethal effects of pesticides and considered in risk assessment, e.g., through the use of newly developed bee counting systems (see Chmiel et al. [60] on

sublethal effects). Schuhmann et al. [219] point out that ecotoxicological studies of individual plant protection products in the laboratory can be rather artificial, as field studies often involve mixtures of pesticides and their interactions are not well understood. However, field studies take place under more realistic conditions and lethal effects can be accurately determined with bee counters. Bermig et al. [15] consider the determination of activity and background mortality in a control population a good way to quantify both sublethal and long-term effects.

The EFSA guidance document contains more than a hundred, rather unreliable, publications on background mortality, many of which could have been greatly simplified by the use of automated bee counters [78]. Accurate determination of bee losses thus goes hand in hand with accurate determination of bee activity, which also has wide-ranging applications. For example, Danka and Beaman [67] used electric bee counters to compare the pollination performance of different bee species based on the number of flights recorded.

In the following, we will review current approaches to bee counting (Section 5.2) and bee counter evaluation (Section 5.3), highlighting the difficulties of meaningful assessment. From these insights we will derive a novel protocol for determining the accuracy of daily losses (Section 5.4). Subsequently, a case study is carried out to assess the practicality of this protocol (Section 5.5).

5.2 Review of Bee Counters

Over the last hundred years, several technical devices have been tested on beehives, which today form the basis of precision beekeeping [257]. Scales, thermometers, hygrometers, anemometers, microphones, radar, photoelectric sensors, capacitive sensors, and RFID transmitters have been used to monitor bee health (for a description of some methods, see Zacepins et al. [258], Marchal et al. [155], and Odemer [181]). Within precision beekeeping, these technologies are divided into three categories: (1) Sensors that collect data at the apiary level, (2) at the colony level and (3) at the level of single individuals [258]. Bumanis et al. [46] describe how, even with sensors of different categories, multi-modal data from bee-related sources can be merged through data fusion.

As our focus is on daily loss measurements, sensors belonging to the third category are of particular interest. Two of the first devices registered counts based on the weight of the bees, however, technical development allowed the implementation of novel methods. From (infrared) photoelectric sensors [201, 233, 145, 56, 111, 209, 77, 190, 228, 118] to capacitive sensors [49, 209, 15] and camera-based methods [48, 58, 59, 236, 244, 152, 132, 177, 114, 169, 238, 133], many ideas have emerged. A rather unique approach to tracking bee movements is presented by Souza Cunha et al. [227], who leverage the Doppler effect through radar measurements.

In recent years, the trend has been towards camera-based methods [181]. The reasons are lower maintenance, greater potential, and the possibility of less invasive data collection. Camera-based methods do not require special tunnels that could get dirty or blocked. Further potential lies in the additional classification of tracked objects (e.g., workers, drones, and intruders) [156], classification of tracked behaviour (e.g., thermoregulation and guarding) [127], and detection of parasites [222] and corbicular pollen loads [226, 156, 6, 178]. A comprehensive historical overview of the development of automatic bee counters was presented by Odemer [181].

5.3 Review of Bee Counter Evaluation Methods

Despite a large number of counting systems, Odemer [181] criticizes the lack of standardised methods for determining the accuracy of bee counters. His review of 38 different bee counters shows that 63 % of them were not validated at all, while of those that included information on precision, 17 % did not include information on the methodology used for validation. Without proper evaluation of counting systems, comparison of different counters is not possible and interpretation of counting results is limited. Therefore, Odemer calls for a validation standard for bee counters [181].

As an example of why validation is important, Figure 5.1 illustrates three hypothetical measurement results concerning the evaluation of bee counters. Figure 5.1A shows two measurement curves of faulty bee counters (blue lines). One gives only zero values (dashed line), the other measures noise (solid line). Figure 5.1B shows the trace of a systematically biased bee counter (blue line). In this case, the counter misses a portion of the returning bees. The systematic bias results in an unrealistically high daily loss indicated by

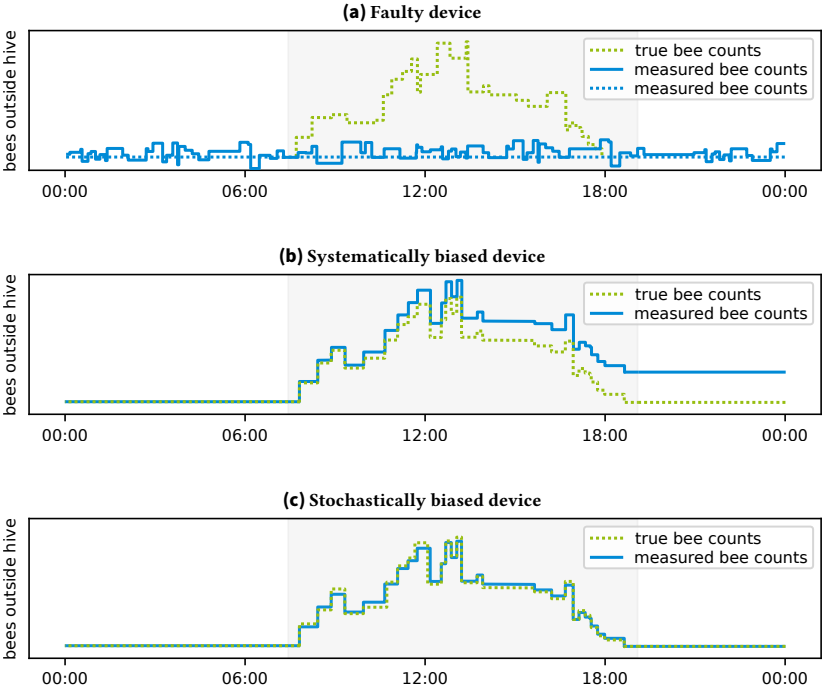


Figure 5.1: Example illustrations of measurement time series' of malfunctioning bee counters. The true but unknown measurement series is illustrated by green dotted lines. The grey area marks the period when flight activity is expected between dusk and dawn. (a) Measurement series of two faulty counting devices, one measuring zero-values (dashed), one measuring noise. (b) Measurement series of a systematically biased counting device showing accumulating errors. (c) Measurement series of stochastically biased counting device.

the high number of bees that are supposedly still outside the hive at the end of the day. Figure 5.1C shows the trace of a stochastically biased counting system (blue line). The counting system provides values that are sometimes too large and sometimes too small, with no preferred direction.

The evaluation of a bee counter aims at excluding malfunctioning counters (Figure 5.1A), quantifying possible biases (Figure 5.1B, 5.1C), and determining the accuracy of daily loss measurements. The procedure in cases (B) and (C) is not straightforward because the true 24-h bee count is neither known nor

determinable: it is impossible to obtain the true bee count by manually counting the number of incoming and outgoing bees in real-time [181]. Bermig et al. [15] noted that manual counting of bees in short video sequences was only possible at 0.3 times the speed and had to be done separately for each of the 24 entrance and exit tubes of the system. This indication is also consistent with the author's experience that it requires about four hours to correctly capture all the bees in a one-minute video at a very high bee density and 40 frames per second. Meikle and Holst's assessment that 'the use of human observation, while probably accurate, clearly limits the time in which the hive can be observed due to fatigue' [160] is shared by many other researchers [209, 238, 15].

Despite these difficulties, there are approaches to partial evaluation of bee counters.

5.3.1 Evaluation by Observation

Evaluation by manual observation is the most widely used evaluation technique [181] and has been used in numerous studies [48, 58, 111, 177, 227, 15, 133, 198, 152, 244]. Experimenters note short periods, rarely longer than 3 minutes, and compare their observations with their measurements. Odemer [181] describes various ways to make such comparisons. The annotation is time-consuming and the benefit is small. A brief example illustrates the problem. For 60-second samples, the smallest detectable error is one bee per minute. Therefore, a counting system is either error-free (to the best of our knowledge, no such system exists today) or has a sampling error of at least one bee per minute.

At this point, there is no reason to consider this error as stochastic, but we must assume that the errors add up and do not level off. It follows that the error in a 24-hour time interval could be (up to) 1 440 bees. The EFSA [79] states the general natural background mortality rate at 3.75 % after reviewing the literature. To determine a 10 % increase in mortality, it is necessary to distinguish the loss of 375 bees from 413 bees in a colony of 10 000 bees. Therefore, the resolution of such an evaluation method cannot meet the requirements.

This problem does not arise with the robbers test described below because the need to annotate the data is conveniently avoided. However, the sample

evaluation has a key advantage. Unlike the robbers test, a bad bee counter will always give bad results, while a good counter will always give good results.

5.3.2 Evaluation by Robbers Test

BeeSCAN stands out among such counting systems whose publication contains precise information on evaluation. Ten years back, BeeSCAN was the most widely used bee counter, which, with regular maintenance, was considered useful [258, 215]. It was first introduced and further developed by Struye [232, 231]. In 32 parallel entrance and exit tubes, infrared light barriers register the direction of movement of the bees and are thus able to determine the activity in both directions. The construction is considered to be maintenance-intensive, as the tunnels can quickly become dirty [215, 15, 145].

To evaluate the BeeSCAN bee counter reliably and realistically, Struye [232] described a novel experimental setup, which the author calls the robbers test. In this test, the counting device replaces the bottom of an empty hive containing only a food source (robber hive). This food attracts robbers from nearby colonies that can only reach the food by passing through the counter and vice versa. At the end of the day, the difference of outgoing and incoming bees must be zero as bees will abandon the food source before sunset (corrected for the number of dead bees remaining in the hive) [232, 15, 181]. Although the underlying count events can be in the order of hundreds of thousands, the robbers test provides only one data point per day, which is the deviation from zero.

While it is true that a good bee counter should achieve a difference close to zero, the opposite conclusion is false. A poor counter may produce a good result (see Figure 5.1A). An extreme example is a counter that gives only zero counts and thus achieves the best possible result in the robbers test. Beyond this edge case, however, the robbers test also fails in more realistic situations. For example, a perfect counter is indistinguishable from a counter with arbitrarily strong symmetric error (Figure 5.1C).

Another problem not named by Struye [232] is the observation that the movement patterns of the bees change due to the setup of the robbers test. The bees move through the sensor passages more goal-oriented, and long

dwell times in the passages are rather the exception [209]. This behavioural variation tends to underestimate the error rate, as long dwell times generally complicate the counting problem [145, 58]. In addition, no activity is expected at the feeder in the darkness that would occur under normal conditions [15, 181]. Because recording in darkness is particularly difficult for camera-based bee counters, the robbers test for such systems may further underestimate the error.

5.3.3 Evaluation by Literature Comparison

Evaluation methods based on comparison with literature or expertise work differently. For example, Ngo et al. [177] and Gonsior et al. [95] evaluated their counters through ecotoxicological studies. For this purpose, a group of bee colonies was exposed to insecticides known to have negative effects on the flight behaviour and homing ability of bees. The authors demonstrated that their counter was able to detect a significant group difference compared to untreated colonies and concluded that the bee counter was functional.

Other methods include the correlation of bee counts with temperature, humidity, and solar radiation [201, 145, 56, 111, 177]. These factors are known to affect flight activity, and therefore the observation of strong correlations demonstrates the principle operability of bee counters. Liu, Leonard, and Feddes [145], Chen et al. [56], and Jiang et al. [111] also compared the measured data with extreme weather events and found, for example, a decrease in flight activity associated with rain. Rickli et al. [201] correlated the measured activity with different colony sizes. A positive correlation at this point is consistent with expectation and can again be interpreted as an indication of a functioning bee counter.

All methods that attempt to confirm correlations known from the literature help to identify faulty bee counters. However, they are not suitable for quantifying the accuracy of daily loss measurements.

5.3.4 Rationale for a New Evaluation Protocol

At the current state of understanding, two things are noteworthy. First, the challenge of determining daily loss in an automated manner with accuracy suitable for regulatory risk assessment has not yet been solved. Second, there

Key features	Evaluation by observation	Robbers test	Literature comparison
Find malfunctioning bee counters by ensuring a minimum correlation between measurements and actual flight counts.	yes	no	to some extent
Distinguish between stochastic and systemic errors.	no	yes	to some extent
Quantify daily loss uncertainties.	no	to some extent	no

Table 5.1: Key features of different evaluation approaches. None of the methods is suitable for determining the uncertainties of daily loss measurements.

is a lack of sufficient evaluation of existing systems [201, 181]. Marchal et al. [155] attribute the difficulty of bee counting to the large number of similar insects that must be observed noninvasively under turbulent conditions.

A large number of counting systems, the variety of sensors used, and the need for an accurate bee counting system require a robust evaluation protocol to determine daily losses. None of the evaluation procedures found in the literature can rule out malfunctioning bee counters and quantify the remaining errors as daily loss uncertainties.

As exemplified in Figure 5.1a, Table 5.1 shows that the robbers test alone is not capable of detecting malfunctioning counters. However, the advantage of the test is that it reveals errors that build up over long periods (Figure 5.1B). In particular, the conclusion that a good robbers test result indicates a well-working bee counter is incorrect. In contrast, evaluation by observation is a more accurate and reliable version of literature comparison. Both methods indicate the degree of correlation between measurements and actual flight counts. However, their results are not directly transferable to the accuracy of daily losses.

The following protocol outlines a method for estimating the Mean Absolute Error of daily bee loss measurements for various types of bee counters.

5.4 A Novel Evaluation Protocol for Bee Counters

Various studies have shown that the accuracy of a bee counter depends to some extent on external factors. One important factor affecting the performance of all bee counters is the activity of the bees themselves. Counters that use tunnels for separation, for example, are unable to distinguish bees that are moving closely together [145, 15]. Crowding of bees causes Spangler [228] to recommend his bee counter only for very small colonies, and Struye [232] to guarantee correct measurements only when bees are no closer than one millimetre. Such factors may vary from system to system. Campbell, Mummert, and Sukthankar [48] report that shade, debris, and intense bee movement negatively affect the performance of their system. Jiang et al. [111] reported reflections and strong nervousness of the bees as additional sources of error.

To ensure that a counting system works well in all relevant situations (in the following termed *scenario*), it is useful to determine such performance factors. For each scenario, an evaluation by observation is proposed (see Section 5.3.1). If it has been determined that only small deviations occur for all scenarios, the next step is to test whether these small deviations add up over one day. For this purpose, several robbers tests are performed (see Section 5.3.2). A large deviation from the target value of zero indicates either a very large stochastic error or a (small) systematic bias of the counting system. However, larger errors could already be excluded due to the sufficiently good performance in the evaluation by observation. Any systematic biases need to be further investigated and corrected.

Without loss of generality, a camera-based bee counter, denoted as f, can be modelled as a function that maps 24-hour video data v from application in the field to an estimated daily loss of bees, $\hat{L}_v$. While v denotes data from regular bee counting scenarios, r denotes data obtained from a robbers test. We define the daily loss as the difference of outgoing bees and incoming bees. With L_v being the true and unknown loss and e_v the bee counter's error, we can write:

$$f(v) = \hat{L}_v = L_v + e_v \ . \tag{5.1}$$

Note, the value of interest, the bee counter's error e_v, cannot be determined without knowing the true loss L_v. Naturally, in field mortality measurements, L_v is also unknown. However, when analysing video data r within the framework of a robbers test, the situation is different, as the true loss is known to

be $L_r = 0$ (see Section 5.3.2). Therefore, for robbers test data, the following holds:

$$e_r = \hat{L}_r - L_r = \hat{L}_r - 0 \ . \tag{5.2}$$

This enables us to calculate the mean absolute error $\bar{e}_R$ using the results $\hat{L}_{r^{(i)}}$ of all $i \in \{1, \ldots, N\}$ robbers tests:

$$\bar{e}_R = \frac{1}{N} \sum_{i=1}^{N} \text{abs}\left(\hat{L}_{r^{(i)}}\right) \ . \tag{5.3}$$

However, $\bar{e}_R$ is not applicable outside of robbers test trials [209], which clearly limits this approach. Furthermore, it overlooks the performance factors discussed earlier in this section, such as the observation that bee counters perform better on rainy days than on days with active bee flight.

Instead, we propose a function $\hat{e}(d_v)$ that estimates the deviation from the true loss L_v while accounting for the difficulty d_v of the video data v. This difficulty metric d_v is based on performance factors, making it equally applicable to both robbers test data and regular field data. A key advantage of using d_v in $\hat{e}(d_v)$ is that it allows us to use robbers test data to fit the function.

This concept is implemented through a nine-step procedure detailed below. Following this, a case study offers additional guidance for practical implementation.

Step 1: Performance factors

An operator expert for the bee counter being evaluated determines the set of relevant factors $\mathcal{F}$ that affect the performance of the counter. For practicality, it must be ensured that each performance factor can be determined robustly and automatically. This means that for non-camera-based bee counters, without further ado, there is only one performance factor at hand, the activity determined with the counter itself. Caution is required as in this case the counter itself is involved in the evaluation. Apart from this particularity, there are no differences for other bee counting approaches.

Example: Based on experience with the counter at hand, an expert knows that a large number of bees and infrared lighting during the nighttime negatively impact counting accuracy and identifies

$$\mathcal{F} = \{\text{bee activity, illumination}\} \tag{5.4}$$

as performance factors.

Step 2: Deriving scenarios

All (reasonable) combinations of performance factors are compiled and a set of scenarios $\mathcal{S}$ is derived.

Example: 'Few bees move slowly under infrared lighting' is considered one of the scenarios.

Step 3: Scenario annotation

For each scenario, corresponding video samples must be collected and manually labelled. The annotations are considered as ground truth in the next step.

Example: For each scenario, 10, 20, or 30-second videos are annotated.

Step 4: Scenario evaluation-by-observation

In this step, the performance of the counter in each scenario is assessed using a suitable quality measure m. To do this, all annotated videos are processed with the counting system f and compared against the ground truth. However, the performance metric m is open and depends on the type of device.

The most intuitive performance measure is the deviation of the registered difference between outgoing and incoming bees $\hat{L}_s$ and their true values L_s,

$$m(s) = \hat{L}_s - L_s \ , \tag{5.5}$$

where s is the scenario in question.

However, due to the short duration, some scenarios contain bees but no counting events. A metric based solely on counting events is bound to be poorly resolved, especially for short duration, and does not take advantage of the hundreds of annotations underlying the count data. It is advisable to use metrics that can handle all types of scenarios. For example, the Multiple Object Tracking Accuracy m_{MOTA} [16] is determined by the number of false-positive detections, false-negative detections, and id switches. Since much more information is included, more accurate performance estimations are obtained. It is suitable for any video-based counting device and any scenario that contains at least one annotated object.

Counters for which no information is available other than the number of counting events (e.g. counters that solely use photoelectric or capacitive sensors) must fall back on the deviation of outgoing and incoming bees given in Equation 5.5.

Step 5: Scenario difficulty rating

All scenarios are rated by a difficulty function $d : \boldsymbol{\mathcal{S}} \rightarrow [0, 1]$ on a scale from zero (easiest) to one (most difficult). The scenario that achieved the best result in the previous step is assigned difficulty level zero, and the scenario with the worst result is assigned difficulty level one. Therefore, the difficulty for a scenario $s \in \boldsymbol{\mathcal{S}}$ is defined by:

$$d(s) = \frac{m(s_{\mathrm{best}}) - m(s)}{m(s_{\mathrm{best}}) - m(s_{\mathrm{worst}})} \,, \tag{5.6}$$

which rates each scenario on a scale from the easiest to the most difficult, preserving their relative differences. This transformation serves to make the results easier to interpret and independent from the metric chosen in the previous step.

Step 6: Robbers tests

Robbers tests are carried out during several days and, if possible, with different colonies [181]. Since flight activity in a robbers test scenario is usually lower than in full-sized colonies, a further reduction in bee traffic should be avoided

where possible. This means that the food source in the robber hive should be highly attractive (e.g., honeycombs) and should last for at least several days to generate as much flight activity as possible. In addition, weather conditions must stimulate bees to forage, and bee density should ideally be high within 2 km of the robber hive. If the trials are conducted in a tunnel tent to prevent the spread of robbing to neighbouring colonies or to prevent disease, the experiments should be conducted with strong colonies with more than 20 000 bees.

Step 7: Difficulty of robbers tests' measurement conditions

To predict a bee counter's error $\hat{e}_{d_v}$ based on the difficulty of the (video) data r obtained from a robbers test, the data r is split into small time intervals. Each interval is then matched to the most similar scenario, using the performance factors identified in step 1. Assuming the scenarios are of equal length, the difficulty of a robbers test is computed as the average difficulty of the scenarios it includes ($\boldsymbol{\mathcal{S}}_r$):

$$d_r = \frac{1}{|\boldsymbol{\mathcal{S}}_r|} \sum_{s \in \boldsymbol{\mathcal{S}}_r} d(s) \,, \tag{5.7}$$

where $|\boldsymbol{\mathcal{S}}_r|$ is the number of scenarios in the robbers test. By replacing robbers test video data r with arbitrary bee counting video data v, the difficulty d_v is defined similarly to Equation 5.7.

Step 8: Modelling accuracy based on difficulty

The bee counters estimated deviation from the true loss measurement, $\hat{e}(d_v)$, is modelled using an appropriate model class (e.g. linear or polynomial models) and data r obtained from robbers tests. There should be a positive correlation between the difficulty d_r and the models error prediction, $\hat{e}(d_r)$.

Step 9 (optional): Plausibility check

It might be useful to additionally compare the measured mortality with the mortality expected from the literature or to show correlations with light

intensity or temperature. Rosenquist [209] wrote: 'Long-term observations over several weeks under regular conditions do not provide accurate reference values. Nevertheless, they were very important for functional verification'. [209]

5.5 Case Study

The following case study illustrates the application of the evaluation protocol to a commercially available camera-based bee counter. It serves as a starting point for implementing the protocol in practice.

5.5.1 Materials and Methods

The range of commercial bee monitoring solutions capable of detecting daily losses is limited. US-based Keltronix Inc. offers Eyesonhives[1], a camera-based monitoring solution. However, as the system determines the level of flight activity in front of the hive without direction, it is unable to determine numbers for incoming and outgoing bees.

Similarly, BeeScanning[2], BeeAndme[3] and Arnia[4] offer different tools and sensors, that also do not allow direct conclusions to be drawn about daily losses. BeeCheck[5], which was developed for the specific purpose of counting bees by the Federal Research Centre for Cultivated Plants (JKI), has not yet reached product status. Hiverize[6] provides building instructions for monitoring bees, but these are limited to weather data and weight measurements. Of particular note is the BeeSCAN[7] system, but development stopped more than 20 years ago [232]. In addition, Beehivemonitoring[8] offers a non-camera-based module for counting bees.

[1] *Eyesonhives*, https://www.eyesonhives.com/. Accessed April 1, 2022.
[2] *BeeScanning*, https://beescanning.com/. Accessed April 1, 2022.
[3] *BeeAndme*, http://beeandme.com/. Accessed April 1, 2022.
[4] *Arnia*, https://www.arnia.co. Accessed April 1, 2022.
[5] *BeeCheck*, https://www.dbu.de/projekt_57031643/01_db_2409.htm. Accessed April 1, 2022.
[6] *Hiverize*, https://hiverize.org. Accessed April 1, 2022.
[7] *BeeSCAN*, http://users.telenet.be/lowland/. Accessed April 1, 2022.
[8] *Beehivemonitoring*, https://beehivemonitoring.com/. Accessed April 1, 2022.

Finally, the monitoring technology of the company apic.ai in Karlsruhe, Germany, was chosen because it is a self-designed, state-of-the-art, and commercially available bee counter. It is not a simple counting device, since among other things the corbicular pollen loads can also be quantified visually. Regardless, the system will be referred to as a bee counter. It consists of a camera unit attached to the entrance of the hives and the software BRAT (Bee Recognition and Tracking), which analyses the collected image data. The camera is the Raspberry Pi Camera (2.1) with an IMX219 sensor and a resolution of 0.168 millimetres per pixel. An Nvidia Jetson Nano acts as the controller, which is battery-buffered and powered by a solar cell. True-colour LEDs with a high CRI value serve as the light source.

All bees pass through the camera's field of view as they enter and leave the hive. In the camera unit, a plastic glass prevents the bees from crawling over each other, but grouping is still possible. The camera unit is shown in Figure 5.2. The hardware is designed for autonomous use in the field. The locally stored data is transferred to the cloud at the end of the study, where bee localization, tracking, and counting are conducted. The exact evaluation procedure is described below and follows exactly the protocol proposed in Section 5.4.

In addition, the 'Hohenheimer Einfachbeute' with Zander frames and queen-right colonies of different and local *A. mellifera* subspecies were used. Bees had shown no visual signs of diseases and were maintained by a professional beekeeper. To attract enough robber bees, we used honeycombs with stored honey (approx. 2 kg per comb).

5.5.2 Procedure and Results

Step 1: Performance factors

Three performance factors based on experience and literature were considered. These were (1) the number of bees in the field of view of the camera, (2) the type of lighting (infrared/white light), and (3) the degree of crowding [228, 145, 15]. Looking more closely at video samples, the total number of bees was less relevant and is already included in their degree of crowding. Since it is assumed that infrared light has no effect on bee behaviour [7], the use of infrared light is generally recommended between dusk and dawn. However, in this case study, the counting device was not attached to the beehive itself,

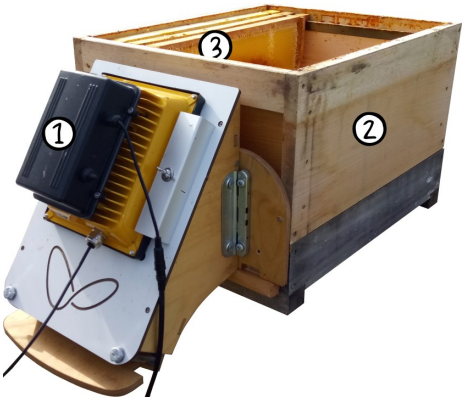


Figure 5.2: The camera unit of the apic.ai bee counter in use. The system (1) is mounted at the entrance of the hive (2). The hive is unoccupied and contains filled honeycombs (3) as bait for the robbing test. *Image credits: R. Odemer.*

but an 'empty' robber hive (Figure 5.2). Therefore, for convenience, it was decided not to switch from white light to infrared light since the bees were not in the robber hive during the night hours. This also means that lighting did not have to be considered as a factor in the evaluation.

The remaining crowdedness factor $\mathcal{F} = \{\text{crowdedness}\}$ was determined as presented in Figure 5.3. The crowdedness score ranges from zero to one and expresses the proportion of the image area covered by densely crowded bees. The choice of performance factors was then confirmed, as a strong and significant negative correlation was found between the quality measure chosen in step (4) and the crowdedness factor described here. It can therefore be assumed that no relevant performance factors were overlooked in this case study.

Step 2: Deriving scenarios

Since only one relevant performance factor was found, the compilation of the scenarios was not very complex. A total of $|\mathcal{S}| = 59$ video clips with very different occupancy were selected. Each clip was ten seconds long and had a frame rate of 40 fps. Figure 5.4 shows example frames from three scenarios that differ greatly in their crowdedness score. Note that the crowdedness

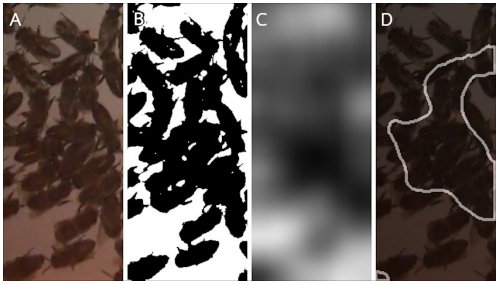


Figure 5.3: Determination of the crowdedness factor in input images. From left to right: (A) A Section of the raw camera image. (B) Foreground-background separation with simple thresholds. (C) A (Gaussian) blur version from image (B), revealing areas of densely crowded bees. (D) Result of thresholding for image (C). The contours show the portion of the input image where densely crowded bees were located. The crowdedness value indicates the proportion of the highlighted contour to the total area. In the present case, the crowdedness is approximately 34 %.

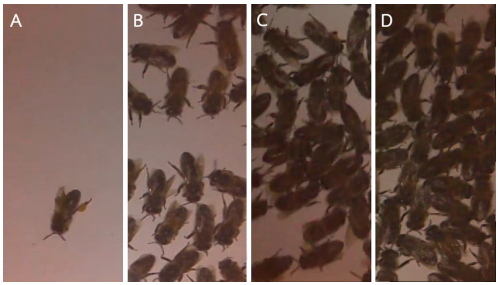


Figure 5.4: Sample images from three 10-second scenarios with varying degrees of crowding (A: 0 %, B: 4 %, C: 34 %, D: 91 %).

score is only indirectly related to the number of bees. For the performance of the bee counter, it is more important how close the bees are. While counting the bees in (A) and (B) does not cause any problems for man and machine, the same is much more error-prone in (C) and (D).

Step 3: Scenario annotation

All 59 scenarios were annotated. The centre of each bee was carefully marked on each clip in each frame to obtain the bees' trajectories. With a large

number of bees, annotation is extremely time-consuming, taking up to 40 minutes per ten-second video clip.

The labels must be double-checked, as errors due to operator fatigue cannot be ruled out. It was found that even small annotation mistakes have strong effects on further analyses (e.g., a video with hardly any bees was wrongly classified as the most difficult because a single bee was forgotten to be annotated). This must be avoided at all events.

Step 4: Scenario evaluation-by-observation

The general functionality of the bee counter was demonstrated using the previously annotated scenarios. The Multiple Object Tracking Accuracy (MOTA) metric [16] was employed to assess the quality of the detection and tracking algorithms. It takes into account not only the interrupted paths of the tracker (#id-switches) but also errors of the detector (#false-positive, #false-negative). For scenario $s \in \boldsymbol{S}$ it is defined as:

$$m_{\mathrm{MOTA}}(s) = 1 - \frac{\#\text{false-positives} + \#\text{false-negatives} + \#\text{id-switches}}{\#\text{labelled objects}}. \quad (5.8)$$

Since some scenarios ($N = 9$) contained empty video clips (#labelled objects is zero), no MOTA score could be calculated in these cases. However, since there were no false-positive detections in all these cases, the quality measure was manually set to 100 %.

Empirical MOTA scores ranged from 94.01 % to 100 % (Median = 1.0, IQR = 0.0003), indicating a well-performing system. The vast majority of scenarios did not pose a challenge to the system studied, and only cases of extreme overcrowding resulted in errors (see Figure 5.4 right). In the most extreme scenario, 102 individuals and a total of 29 846 positions were annotated in a 10-second video sequence. In this case, the number of errors was 17.52 per trajectory. A correlation analysis showed that the MOTA scores of the scenarios and the crowdedness factors were highly correlated (Pearson's $r = -0.8953$, $p < 0.0001$).

Despite the good result, it cannot yet be ruled out whether the remaining errors accumulate over the day (i.e. are systematic) or not. At this point,

information from the robbers test is missing to be able to make statements about longer periods.

Step 5: Scenario difficulty rating

The performance score MOTA of each scenario is transformed according to the protocol to represent the difficulty. That is, the scenario with the best score (100 %) in the previous step is assigned difficulty zero, and the scenario with the worst score (94.01 %) is assigned difficulty one. Thus, following Equation 5.6, the difficulty $d(s)$ for scenario $s \in \boldsymbol{S}$ is calculated using

$$d(s) = \frac{m_{\text{MOTA}}(s_{\text{best}}) - m_{\text{MOTA}}(s)}{m_{\text{MOTA}}(s_{\text{best}}) - m_{\text{MOTA}}(s_{\text{worst}})} \,. \tag{5.9}$$

Step 6: Robbers tests

The robbers tests were carried out in Braunschweig, Germany in the period from September 23, 2021 to October 14, 2021. Due to bad weather, the opening of boxes to restock honeycombs, maintenance of the system, and missing data, most days had to be discarded, as even the shortest periods without data can have fatal implications for the evaluation. The remaining four complete measuring days October 1, 2021 (D1) to October 4, 2021 (D4) are shown in Figure 5.5, starting one day after the robber hive was refilled with three full honeycombs. Count data for this period was made publicly available in an online repository [35]. A typical activity pattern of entries and exits can be observed, which is strongly related to the diurnal pattern [64]. While in occupied hives activity can usually be measured at night, too [64], in robber hives there is neither thermoregulation nor guard bees or regular activity after dark as the box is empty.

From previous experiences, we knew that the general flight activity in robbers tests is lower than in a full-sized colony. For this reason, the experimental design of [15] was modified and tunnels were omitted. At the site, about 60 colonies were situated in the flight radius, which could potentially visit the robber hive. As a result, the bees sometimes fought violently at the hive entrance. These behavioural changes, however, had a positive effect on crowding and complicated the measurement conditions, contradicting

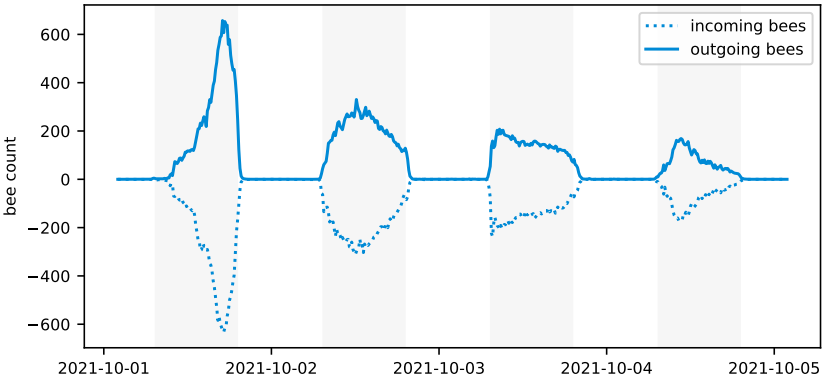


Figure 5.5: Bidirectional bee activity aggregated in ten-minute intervals as determined by the bee counter. For example, 6 573 bees left the robber hive between 5:00 p.m. and 5:10 p.m. on October 1, 2021. The first day was preceded by the renewal of the food source. The amount of forage (and the bees' interest) steadily decreased over the next four days. The shaded intervals mark the periods between dusk and dawn.

	r = D1	r = D2	r = D3	r = D4
e_r	−1871	−828	−113	+56
d_r ± SD	0.13 ± 0.31	0.08 ± 0.10	0.04 ± 0.08	0.02 ± 0.05

Table 5.2: For each robbers test, denoted D1 through D4, the difference between outgoing and incoming bees, e_r, is presented. Negative values indicate that more incoming bees were counted, while positive values indicate that more outgoing bees were counted. An ideal bee counter would yield $e_r = \pm 0$. As anticipated, e_r shows strong correlation with the difficulty of the robbers tests, d_r.

Rosenquist's [209] observation that bees move in and out of the hive quickly rather than dwell at the entrance in the robbers test.

Subtracting incoming bees from outgoing bees yields robbers test errors from −1871 to +58 bees (see Table 5.2). These errors are net errors, as the deviations tend to balance out in both directions throughout the trial.

Step 7: Difficulty of robbers tests' measurement conditions

To determine the difficulty of the measurement conditions of an entire robbers test, d_r, the video data were divided into 30-second time intervals. For each

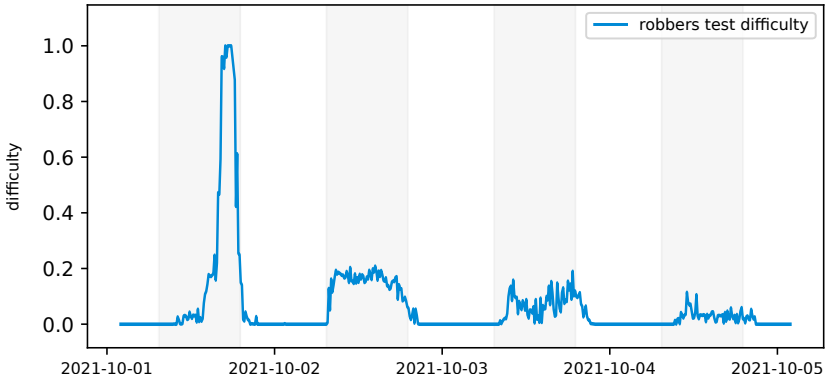


Figure 5.6: To find out the difficulty of the robbers test's measurement condition, the robbers test video data was divided into 30-second intervals. For each such interval, the most similar scenario was determined based on the measurable performance factors, and its difficulty was assigned. The average difficulty across all 30-second intervals of a robbers test trial resulted in the daily difficulty.

video sequence the 'crowding' performance factor was calculated using the image segmentation procedure shown in Figure 5.3. These factors were used to assign each video sequence the most similar scenario, and the scenarios' difficulties $d(s)$ were used to calculate d_r as shown in Equation 5.7.

Figure 5.6 shows the mean difficulty in 10-minute time frames for the duration of the trial. Note the parallels to the activity in Figure 5.5. A day consisting of only the most difficult scenario would have a daily difficulty of one. The difficulties for each robbers test, d_r, are shown in Table 5.2.

Step 8: Modelling accuracy based on difficulty

Selecting an appropriate model for $\hat{e}(d_v)$ is difficult due to the small number of robbers tests ($N = 4$). Since the data points offer little guidance, the conclusion from the previous steps is used. In step (4), a strong linear correlation was observed between the difficulty of the scenarios and the MOTA performance. A correlation coefficient of $r = -0.8953$ suggests that $\hat{e}(d_v) \sim d_v$ also follows a linear relationship that can be modelled as:

$$\hat{e}(d_v) = \beta_1 d_v + \eta \ . \tag{5.10}$$

This model is fitted using the robbers test data given in Table 5.2. For the values of the exogenous variable d_v, d_r can be plugged in directly. However, the values for the endogenous variable, the standard deviation of the measurement, must first be calculated. The standard deviation of a single measurement with a known expected value of zero can be determined for each $r \in \mathrm{D1}, \mathrm{D2}, \mathrm{D3}, \mathrm{D4}$ as follows:

$$\sigma\left(e_{d_r}\right) = \sqrt{\frac{1}{1}\sum_{i=1}^{1}\left(\hat{e}_{d_r}^{(i)} - 0\right)^2} \tag{5.11}$$

$$= \mathrm{abs}\left(e_{d_r}\right) \ . \tag{5.12}$$

Note, standard deviation, rather than variance, is modelled because it has the unit 'bees', as do the errors underlying the MOTA score. An intercept is omitted, as we assume no errors in the easiest scenario.

Linear regression using python package `statsmodels` at version 0.12.2 yields $\hat{\beta}_1 = 12\,673$ ($N = 4$, $r^2 = 0.95$). The fitted model $\hat{e}(d_v)$ with its estimated coefficient $\hat{\beta}_1$ can now be used to determine the standard deviation to the daily loss measurements of difficulty d_v.

Step 9: Plausibility check

In this context, reference should be made to the study by Gonsior et al. [95]. There, a previous model of the bee counter was assessed as part of an ecotoxicological trial. In an experiment, four bee colonies were fed over ten days with a neonicotinoid-spiked sugar solution, known to have sublethal effects on the flight behaviour of bees. When comparing the number of foraging flights with those of four untreated control colonies, significant group differences were found. The result demonstrates that the system was able to show the expected effect and represents a general plausibility test.

5.5.3 Discussion of Case Study

Using scenarios is a logical consequence of the observation that the performance of bee counters fluctuates based on environmental factors. One main factor was found to have a strong effect on bee counter performance. A total of 59 scenarios with varying degrees of 'crowding' were drawn and carefully

labelled. Since all scenarios achieved MOTA scores greater than 94.01 %, evaluation by observation confirmed the basic functionality of the bee counter. The strong correlation between scenarios' MOTA scores and difficulties indicates that the crowdedness performance factor is highly suitable for inferring the accuracy of the bee counter. Although the MOTA score requires special treatment of scenarios for which the number of labelled objects equals zero, the advantage outweighs the disadvantage. Instead of relying on sparse count events, the metric incorporates thousands of true positives, false positives, false negatives, and ID switches from detection and tracking.

This promising pre-evaluation justified the implementation of several robbers tests. Despite the long experimental period of 21 days, most of the robbers tests had to be rejected. Reasons for this included the late season and the resulting increase in rainy and cold days, but also the maintenance and restocking of the feed. The robbers test also shows the importance of the operating time of the bee counter. If the system had been down for ten minutes during the most active period on D1, which would correspond to an operating time of 99.31 %, 6 573 outgoing bees and a similar number of incoming bees could not have been recorded. The impact on daily loss measurements, which is expected to be a fraction of this number, is significant. Thus, not including days with missing data is important, but reduces the number of valid trials.

The utilization of the original robbers test described by Struye [232], in which bees from multiple colonies rob the same robber hive, could solve the problem of consistent and goal-oriented bee behaviour described by Rosenquist [209]. The robber experiments contain moments of extreme crowding and thus simulate a more realistic flight behaviour at the robber hive.

A total of four robbers tests were available for further analysis and showed that despite the high MOTA values of the scenarios, the measurement of the 24-hour intervals is prone to error. This is reflected in a mean absolute deviation of 717 (±846) bees compared to a target value of zero. Due to limited data availability, it was not possible to make a reliable statement whether the bee counter unilaterally counts more bees or not. The observation of a bias has been reported by several researchers for other bee counters and would lead to skewed counts, distorting the loss measurements [177, 15]. Consequently, the reason for a preferential direction should be investigated and eliminated. The data further corroborate findings from other authors that on days with higher flight traffic, the error of the counter increases [181].

Since bees react to visible light, it is reasonable to use infrared light that is invisible to bees at night [7]. Nevertheless, the robbers tests were performed here exclusively with white light and consequently, the performance factor 'lighting' was omitted. This had no consequences for the robbers test since no bees are expected at the feeder during the infrared hours anyway. This also means that the influence of the factor could not have been determined at all. However, for conducting experiments in the realistic scenario, the lighting factor is potentially important, since under certain circumstances bees may stay near the entrance after dark [181]. One way out would be to conduct additional robbers testing experiments with all-day infrared lighting. However, this would have required modifications to the system that were beyond the scope of this case study. Therefore, the regression model only applies to the trials that do not require infrared light.

Due to the small number of valid robbers tests, the model had to be derived from considerations and could not be checked directly on the data. The results are plausible and consistent with expectations.

To determine whether the tested bee counter is suitable for regulatory risk assessment and the implementation under Good Laboratory Practice, the error of the counter must be comprehensible and considered. Assuming that the robbers tests have a similar level of difficulty to that of full-sized bee colonies, the expected error in difficult cases may exceed 1 500 bees. Thus, at the stage of development (September 2021), it is not possible to determine accurate daily losses or bee mortality as defined by regulatory requirements. The estimated absolute error of the system under field conditions is too large to detect small differences in colony population dynamics. In addition to improving the bee counter, it would be possible to simplify the measurement conditions. For example, limiting the number of colonies to smaller sizes (see Spangler [228]) or by making structural changes to the flight board that forms the camera's field of view to reduce crowding.

5.6 Discussion of Protocol

To date, there is no method to determine daily honey bee loss or background mortality with the accuracy required by the 2013 EFSA bee guidance document [78]. Existing bee counters are not sufficiently suitable for this purpose,

especially because a standardised evaluation protocol has not been available [181].

Although there were approaches to evaluate bee counters, the results did not allow conclusions on the accuracy in daily loss measurements. It was argued that (1) evaluation procedures based only on sample evaluation do not reveal bias or provide an indication of the accuracy of daily loss measurements and that (2) evaluation procedures based only on robbers tests cannot distinguish with confidence between inoperative and operative counters and do not provide accuracy measurements under realistic conditions.

A combination of the two approaches could, however, solve these problems. Hence, the presented evaluation protocol was tested on a commercial counting device. Here it became obvious that requirements for accurate loss measurement are high. Even minimal deviations accumulate quickly because the observed quantity (activity) is two orders of magnitude larger than the quantity of interest (loss).

Apart from that, small inaccuracies in the determination of individual bee movements can have serious consequences. In Struye's BeeSCAN, the activity was about a hundred times the measured daily loss [232]. Given this factor and the relatively small reference size, effects such as those in Rickli et al. [201] and Bermig et al. [15] can quickly occur. There, promising bee counters led to unrealistic results in the determination of daily losses. The first study reported losses that exceeded expected results by a factor of five, while the second reported a gain of over 14 000 bees in just one day. In both cases, small deviations, not measurable in short periods, added up over the day due to the sheer volume of entries and exits.

To detect field-relevant changes in daily losses, bee counters must have no more than 1 error per 1 000 entries or exits, depending on colony size and requirements. These numbers assume 105 000 flights, a colony size of 30 000 bees, and a natural mortality rate of 3.75 % [79].

No generalised threshold has been established to indicate whether or not the pre-evaluation has failed and a redesign is necessary. Instead, it is recommended to report the entire evaluation process. Whether a bee counter is useful or not depends on the intended application. The result of the evaluation shows which errors are to be expected under which measuring conditions. Whether these errors are acceptable or not is ultimately decided by the potential user: If there is a bee counter which has shown daily loss a error of ±100

bees, the inaccuracy corresponds to about 10 % of the expected background mortality in the previous example.

Future designs of counters must be technically sound and capable of operating efficiently and autonomously under field conditions [181]. In addition, it is necessary to generate validated data with a standardised protocol that meets scientific requirements and allows accurate conclusions to be drawn about the daily loss of foragers. Without this standardization, no progress in this field will be possible. However, with the technological advancements that exist today and will exist in the future, such standardization should be readily implementable.

5.7 Conclusion

High-quality data on honey bee background mortality are currently unavailable due to a lack of methodology to generate them. With the here presented evaluation for daily loss measurements, a protocol was introduced that should be suitable for determining the accuracy of electronic bee counters under field conditions in a standardised way. The protocol combines existing approaches into a new, harmonised method that can be performed regardless of how the bee counter operates. The thorough evaluation is time-consuming but only needs to be done once for a bee counter system. The work thus makes innovations in practice measurable and creates the basis for comparability of bee counting systems, enabling faster progression of the sector. Hence, it should be possible to advance the field by developing counters that meet or even exceed scientific and regulatory requirements.

6.

Parts of this chapter are based on

A PURELY VISUAL RE-ID APPROACH FOR BUMBLEBEES (*BOMBUS TERRESTRIS*) [36]

Parzival Borlinghaus[1], Frederic Tausch[2], Luca Rettenberger[3]

[1]Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]apic.ai GmbH, Karlsruhe, Germany

[3]Institute for Automation and Informatics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

CRediT

Conceptualisation: PB, FT, LR; Data curation: FT; Formal analysis: PB, FT, LR; Investigation: FT; Methodology: PB, FT; Project administration: PB; Resources: FT; Software: FT; Validation: PB; Visualisation: PB, FT; Writing - original draft: PB; Writing - review & editing: FT, LR.

Publisher

Smart Agricultural Technology published by Elsevier B.V., vol. 3, p. 100135, Feb. 2023, doi: 10.1016/j.atech.2022.100135.

6.1 Introduction

In computer vision, animal re-identification (re-ID) refers to the task of identifying individuals from a a group of the same species. Machine-assisted re-ID from image data, such as that obtained via camera traps, enables the study of behavioural patterns with minimal effort [216].

Automated animal re-ID has been successfully demonstrated for mammals in many cases [85, 45, 125, 252], including apes, whales, and elephants. However, identification is not limited to mammals; promising examples for the related task of insect classification [196, 217, 69] and re-ID already exist. Arbuckle et al. [2] developed the Automatic Bee Identification System (ABIS), which can reliably identify bee species based on wing patterns. From this, Francoy et al. [83] showed that these patterns also allow the identification of individual Africanized honey bees. By utilizing self-organised maps and CNNs respectively. Kastberger, Radloff, and Kranner [116] and Murali et al. [170] added giant honey bees (*Apis laboriosa*) and fruit flies (*Drosophila*) to the list of successfully re-identified insects. All aforementioned insect classification and identification approaches analysed the wing patterns that are formed from vein junctions. As special image acquisition is required, such approaches are not suitable for application on living animals in the wild. Therefore, a recently published approach to re-identify honey bees by their abdomen stands out.

Using the abdomen as a re-ID feature, Chan et al. [53] made honey bee re-ID applicable in camera traps. In their work, the authors emphasize the importance of time-invariant features, that is, features that do not change over the time bees spend foraging. As honey bees store a significant amount of nectar in their honey stomach and carry pollen on their hind legs, changes in their appearance between leaving and returning are to be expected. Therefore, to compile the training data, it was not sufficient to film and track individual bees entering the hive for a short moment (short-term dataset). For this reason, Chan et al. [53] extended the training data with a long-term dataset. By gluing tags on the thorax of 181 bees, they were able to distinguish the bees over time. Since the authors solely used the bee's abdomen as a re-ID feature, the glued-on tag did not leak information to the re-ID system. As expected, re-identification based on the unmarked abdomen was significantly worse in all evaluation setups over long periods. The evaluation showed that, at best, the correct bee could be identified among ten other bees (so-called distractors) in 62 percent of the cases [53]. When replicating the study, Roth

[210] found that increasing the number of distractors to 50 approximately halved the reported CMC-1 score. To test the accuracy of the approach in the real world, the number of distractors would need to be increased by a factor of a thousand. It is easy to imagine that honey bee re-ID cannot yield useful results in practice. Since neural networks trained with a triplet loss function well for the re-ID of other animals, the issue rather lies in the fact that the abdomen does not provide sufficient discrimination among bees. The similarity between the worker bees studied is intensified by the fact that they share the same mother. Additionally, different curled abdomens do also greatly influence the likelihood of correct re-identification [210].

Surprisingly, to our knowledge, there are no attempts to transfer visual animal re-ID to the bumblebee domain, especially since it is known that bumblebees differ much more in body shape and live in much smaller colonies than honey bees [96]. This is despite the fact that bumblebees, as important pollinators, have been the subject of repeated research. Mola and Williams [165] review methods for studying the movements of bumblebees, including QR codes, paint stains or RFID chips, all of which being particular important for ecotoxicological studies as required for pesticide testing [78]. However, tagging individuals is extremely laborious and only possible for a small number of individuals [237]. Furthermore, it cannot be ruled out that tagging individuals may affect the observed behaviour. For these reasons, it is desirable to replace the existing procedures with a purely visual re-identification.

This paper comprises two contributions. (1) We investigate to what extent 99 bumblebees from two colonies are re-identifiable on visual material and (2) which features are of decisive importance. To address these research questions, a Convolution Neuronal Network (CNN) based model, a simple body shape model and a random baseline model were benchmarked using the publicly available Bumblebee Re-ID Dataset [239].

6.2 Materials and Methods

For the purpose of determining behavioural patterns, entomologists are interested in identifying bumblebees as they re-enter the hive [96, 9]. In terms of computer vision, the underlying problem can thus be characterised as a bumblebee retrieval problem across the same camera. In practice, a bumblebee would be captured by the camera when leaving the hive and some

images of this individual would be stored together with an ID in the so-called gallery for later retrieval. Images of returning bumblebees (probe images) are then compared against the gallery. There may be several images of different bumblebees in the gallery at this time. Since bumblebees are registered when they fly out, it can be assumed that in a real application all incoming bumblebees are already known to the system, which classifies this work as closed re-ID [54].

The re-ID task is commonly divided into object detection, tracking and ID retrieval [263]. While the computer vision community already provides solutions for the first two tasks [247, 156, 238], this work focuses on the ID retrieval, i.e. algorithms that output a ranked candidate list of potential matches for any probe image. In the following, it is explained how the corresponding bumblebee dataset is composed. For an explanation of the evaluation of re-ID models, namely the concept of Cumulative Matching Characteristics (CMC) and Mean Average Precision (mAP), refer to section 2.3.4 on page 26.

6.2.1 Dataset

The dataset was previously published by Tausch et al. [239]. The images were filtered and finally show 99 individuals from two commercial bumblebee hives with more than 22 images available per individual. The dataset was split randomly into a test and training dataset, keeping the ratio of male and female individuals constant.

In both the test and training datasets, $g = 17$ images were selected per individual as gallery images and $p = 5$ images were selected as probe images. To minimise temporal features in the image sequence of a bumblebee, gallery images were randomly selected from the first $(n - g - p)/2 + g$ images and probe images from the remaining ones, where n denotes the number of available images. Training was performed on the first $(n-g-p)/2+g$ images of the training dataset. There are a minimum of 22 images per individual ($M = 69 \pm 65.3$). Using the naming convention of Chan et al. [53], the given dataset can be characterised as a short-term dataset.

6.2.2 Re-ID Models

Three bumblebee re-ID models were developed and compared. The BumbleNet model extracts discriminative features from raw pixel data of cropped and centred bumblebees. It was hypothesised that such a model could find unwanted but useful information on background pixels. In such scenario, the model would overfit to the highly correlated images in the dataset. Such external cues could be dirt, reflections, stains or other illumination patterns, all being intensified by the poor image quality (see Figure 6.1). This is problematic for all kinds of short-term re-ID datasets and hinders generalization to field data. To shed light on this hypothesis, two BumbleNet variants and the non-pixel-based BodyShape model were fit. While the former receive masked input images from a segmentation network, the latter receives a pre-processed list of shape descriptors. Both approaches aim to strip potentially misleading information. Finally, a random baseline model was added to serve as a benchmark.

6.2.2.1 BumbleNet Model

FaceNet is a face recognition system that employs convolutional neural networks (CNNs) to map images of human faces to a dense feature space that is descriptive of the person depicted [218]. The feature vector of an image is referred to as its embedding. The authors succeeded to derive a function that maps the faces of the same individual to embedding vectors that are as similar as possible while being dissimilar to other individuals (see Section 2.3.1 on representation learning). Consequently, the embedding can be employed to re-identify a particular individual. By substituting human faces with bumblebees, we obtain BumbleNet.

BumbleNet is trained with triplet equivalence constraints, which are enforced by the triplet loss, as detailed in Section 2.3.2.

Three variants of the BumbleNet were trained in total, differing solely in terms of the pre-processing of the training data. This approach makes the network more robust for real-world applications and provides insights into the discriminative power of different features.

The network is able to discern that the position of wings and legs is not a reliable indicator for re-identification due to the availability of a large num-

Figure 6.1: Foreground segmentation using a U-Net. (a) Input image, (b) labelled body contour covering an area of 138.5 mm^2, (c) predicted body contour covering 120.2 mm^2. Note the intended absence of legs and wings.

ber of images of each individual. However, short-term training data cannot account for the possibility that bumblebees may leave the hive without pollen and subsequently return carrying pollen, thereby significantly altering their appearance in the field. By segmenting the body, this potential pitfall can be eliminated, albeit at the cost of lost detail. To quantify the extent of information loss associated with the segmentation, the performance is presented with (BumbleNetSeg) and without (BumbleNetRaw) segmentation.

Given that not only the background but also extremities such as wings and legs must be removed, it is not sufficient to calculate a binary mask by thresholding or background subtraction. Instead, a U-Net architecture similar to that presented by Ronneberger, Fischer, and Brox [207] was trained to segment the bumblebee's body by removing legs and wings (refer to Section 2.2.2 for architectural details). Although this entailed the annotation of hundreds of bumble bees, it was crucial to remove potentially time-variant features that would assist the network in re-identifying individuals in the given short-term dataset but would be ineffective in real-world applications. Figure 6.1 illustrates the input, label, and output of the segmentation network. Given the average body size of bumblebees, which is approximately 95 mm^2 [96], the image depicts a particularly large specimen. The average discrepancy between the computed and labelled bumblebee size is approximately 18 mm^2.

Furthermore, in order to ascertain which features are decisive, a third variant of the BumbleNet model was trained. When training the BumbleNetSegScaled model, scaling was added to the list of possible input augmentations. The application of random scaling between 80 % and 130 % of the initial dimensions prevents the model from utilising the individual's size as a feature.

BumbleNet variants	Model characteristics
Raw	The training data is neither scaled nor segmented. This model has the greatest amount of pixel information available. However, the model also has the greatest potential to extract features that are unrelated to the task at hand.
Seg	The training data underwent preprocessing, whereby the bumblebee's body was segmented. This removed noise (and potentially pertinent information) from the input image, thereby enhancing the network's robustness for real-world applications.
SegScaled	In addition to the aforementioned segmentation, the input data was randomly scaled. This prevents the model from utilising body size as a feature, thereby forcing it to learn other characteristics.

Table 6.1: Comparison of the three BumbleNet variants.

A comparison of the models with and without the scaled inputs allows for the determination of the relevance of this feature. Augmentations that do not affect the bumblebee shape are employed in all three cases, including slight colour jittering, Gaussian noise, contrast variations and rotation. A comparison of the models can be found in Table 6.1.

To obtain the desired embedding, the triplet loss is employed [218]. The triplet loss minimizes the distance between the embeddings of an anchor image and a positive image, while maximizing the distance to a negative image. The positive image shows the same bumblebee as the anchor image while the negative image is guaranteed to show a different individual. Figure 6.2 illustrates the triplet loss in practice. To obtain the embeddings we use the ResNet-18 [102] architecture as backbone and an embedding size of 128, which both have proven to be a sensible choice in a conducted hyperparameters search (refer to Section 2.2.1 for details on ResNet-18).

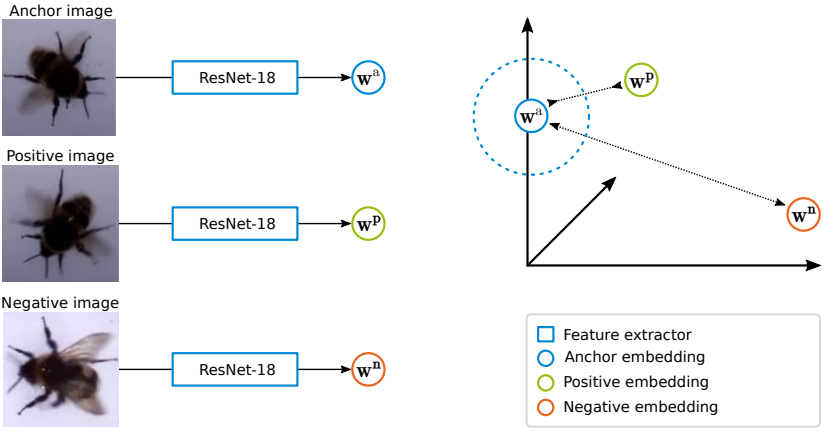


Figure 6.2: Illustration of triplet loss. (Left) Anchor, positive, and negative images are embedded using the ResNet-18 feature extractor. (Right) In the embedding space, the anchor and positive images should be closer to each other than a specified margin (indicated by the dashed circle), while the negative image should be positioned far away from both. During training, the model learns to optimise embeddings by minimising the anchor-positive distance and maximising the anchor-negative distance.

6.2.2.2 Random Baseline Model

To show that a model has learned important appearance descriptors, it must show better results than a random baseline model. This role is accomplished by a model that generates randomly ordered lists of candidates. The model's performance is then calculated as the average mAP and CMC score of all possible candidate lists.

6.2.2.3 BodyShape Model

The BodyShape model consists of a feature extractor and a learned metric. After applying the same bumblebee segmentation as for models BumbleNet-Seg and BumbleNetSegScaled, a rotated bounding box and a rotated ellipse was fitted to the segmented body. This allowed the extraction of several features that form the input for the BodyShape model (segmentation area, segmentation perimeter, enclosing rotated bounding box width, enclosing rotated bounding box height, enclosing rotated ellipse width and enclosing

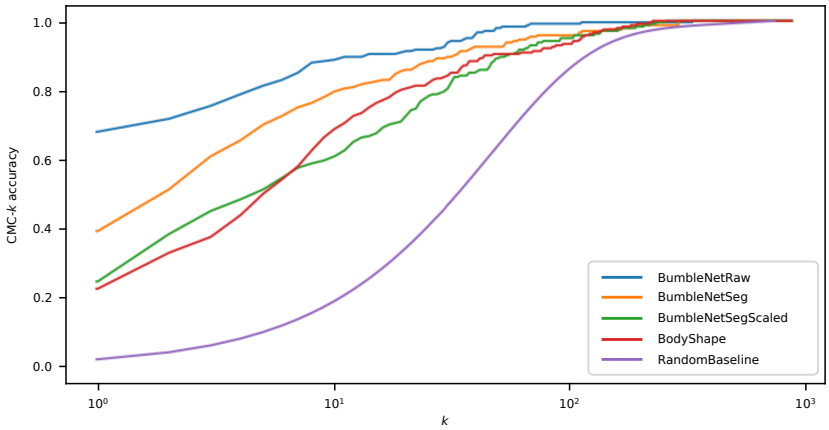


Figure 6.3: For $k = 1, \ldots, 833$ (gallery size) the graph shows the probability that a probe image is among the top-k elements of the ranked retrieved results (CMC-k).

rotated ellipse height). In order to perform a re-identification based on these 7 properties, a distance metric was learned that outputs small distances for equal IDs and large distances for different IDs. For a detailed explanation of metric learning, refer to Section 2.3.3.

6.3 Results

Figure 6.3 shows the CMC-k accuracy for all models, while Table 6.2 summarises key performance metrics. With a CMC-1 score of 0.6792 BumbleNetRaw is the model that gives the best results on the test data. Segmenting the training data lowers the CMC-1 score by 42 %. As expected, the CMC score improves for higher values of k. For CMC-10, the accuracy already exceeds 88 % (BumbleNetRaw).

If visual features are not used at all (BodyShape model), the performance drops by almost 64 % compared to the best model. If the body size feature is omitted, which was implemented by randomly scaling the training data for the BumbleNetSegScaled model, the CMC-1 score drops by almost 66 %. A comparison with the random baseline, however, shows that all models

Models	CMC-1	CMC-5	CMC-10	mAP
B.N.Raw	**0.6792**	**0.8125**	**0.8875**	**0.3509**
B.N.Seg	0.3916	0.7	0.7958	0.2446
B.N.SegScaled	0.2458	0.5125	0.6083	0.1295
BodyShape	0.2291	0.4958	0.6958	0.1876
RandomBaseline	0.0208	0.0999	0.1898	0.0292

Table 6.2: Comparison of models performance with respect to CMC-1, CMC-5, CMC-10 and mAP score, ordered by CMC-1.

succeeded in re-identifying bumblebees based on distinct characteristics and outperformed random guesses by far (10 to 32 times).

A look at the mAP scores in Table 6.2 does not reveal any new information: The mAP scores of the models are broadly consistent with the CMC scores, so there is nothing to suggest that considerations of the model's recall capability should influence model choice.

6.4 Discussion and Conclusion

In relation to research question (1), to what extent bumblebees are re-identifiable on visual material, we have been able to show promising results. The best model is able to correctly assign two thirds of the individuals at first go (BumbleNetRaw). Since the model may have used some undesired information, the performance could be worse in reality. A safe lower bound for the model performance is the BumbleNetSeg with a CMC-1 score of 39 %. However, the retrieval accuracy can be further improved by statistically modelling the probability of presence of single individuals. This would allow the gallery size to be adjusted dynamically to those individuals that are assumed to be outside the hive.

It was already known that bumblebees show a remarkable diversity in their body size. This finding was confirmed by comparing the BumbleNetSeg and BumbleNetSegScaled models, as omitting the size information led to a sharp drop in the CMC-1 score. However, it was also found that other features play a role, still allowing a fifth of the bees to be correctly identified on the first attempt (see BumbleNetSegScaled, Table 6.2). The BodyShape

model outperforms the non-BodyShape model (BumbleNetSegScaled) by a small margin, showing that an individual's size is indeed the most important feature.

Due to the low contrast and resolution of the images, other features could become more important as image quality increases. Therefore, a dataset with more detailed and higher-quality images, including bumblebees carrying pollen, would be desirable. Enhanced data could be obtained using a camera device such as the one proposed in Chapter 4. Furthermore, a new dataset should integrate long-term and short-term re-ID data as demonstrated by Chan et al. [53]. Long-term data would enable the transfer of models, such as BumbleNetRaw, into applications with greater confidence, as these models become more robust against short-term features.

Finally, a bumblebee re-ID is possible on the given dataset. A real-world application would benefit from clever boosting through dynamic galleries and statistical modelling of presence probabilities. Also, relying solely on variations in body shape misses the potential of other visual features.

7.

Parts of this chapter are based on

INTRODUCING POLLENYZER: AN APP FOR AUTOMATIC DETERMINATION OF COLOUR DIVERSITY FOR CORBICULAR POLLEN LOADS [31]

Parzival Borlinghaus[1], Jakob Jung[1], Richard Odemer[2]

[1]Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, Institute for Bee Protection, Braunschweig, Germany

CRediT

Conceptualisation: PB; Formal analysis: PB; Investigation: PB, JJ; Methodology: PB; Project administration: PB; Resources: RO; Software: PB, JJ; Writing - original draft: PB, Writing - review & editing: JJ, RO.

Publisher

Smart Agricultural Technology published by Elsevier B.V., vol. 5, p. 100263, Jun. 2023, doi: 10.1016/j.atech.2023.100263.

7.1 Introduction

Pollen plays a crucial role in the nutrition of honey bees as it is the sole source of proteins and fats [40]. It is also an essential component for brood care and serves as a reliable indicator of the quality of a foraging location. Furthermore, a varied diet of pollen helps to increase the strength of colonies, which is particularly important for their survival during the winter months [40].

7.1.1 Chromatic Pollen Assessment

Given the importance of a diverse pollen supply, the question arises how pollen diversity can be determined. Methods of laboratory analysis include the analysis of pollen grains in honey, bee bread or pollen from pollen traps [41].

An alternative, which is less costly and requires neither expert knowledge nor laboratories, is an approximation of local and temporal pollen diversity using the colour diversity of pollen from pollen traps. Pollen traps force bees to pass through a grid when entering the hive. The grid is designed in a way that the bees lose their pollen load when passing the pollen trap, which accumulates in a container underneath. The pollen can then be manually sorted by colour to determine colour diversity, a process known as (manual) chromatic assessment.

Colour diversity, in turn, can be used to indirectly estimate species diversity in flowering plants, an approach carried out by Conti et al. [63] for a particularly species-rich area in Italy. At intervals of 3 weeks, a total of 19 pollen samples were collected from three hives using pollen traps from April to September for 3 days each. In order to exclude different colour perceptions, a 10 g aliquot was subjected to a manual chromatic assessment by the very same expert. A self-made colour chart with 30 typical pollen colours served as a reference for sorting. The actual pollen diversity was then examined in the laboratory. Under these conditions, the researchers figured that each new colour found corresponded to 1.52 new species in the sample.

Conti et al. [63] also investigated whether a diversity index (Shannon) obtained by chromatic assessment was significantly different from that based on laboratory analyses. Contradictory results were obtained, as there were significant differences in a direct comparison of two differently determined

diversity indices, but not when looking across all samples. Kirk [123] summarises his long-term observations as follows: 'Does colour diversity reflect phylogenetic diversity (...)? The broad phylogenetic association with pollen colour found here indicates that colour diversity will also broadly reflect phylogenetic diversity.' [123]

Besides the different colour perception of humans [5] and the influence of the light situation on the observed colours [123], the limitations are due to the properties of the pollen: It is known that pollen can change its colour depending on the degree of dryness [122], that different species can produce the same pollen colours [105] and, conversely, different pollen colours can be assigned to the same plant species in rare cases [122]. These effects result in fewer colours being found in chromatic assessment than species in laboratory analysis [63, 41].

7.1.2 Citizen Science Investigation on Pollen

Assuming the connection of pollen diversity and chromatic diversity, the spatial and temporal pollen (colour) diversity for many European countries was surveyed in 2014 and 2015 in a large-scale citizen science study named 'Citizen Science Investigation on Pollen' (CSI Pollen) [230]. Here, with the help of 750 beekeepers (citizen scientists), almost 18 000 pollen samples were collected and undergone manual chromatic assessment [41].

It is evident that a study that relies on hundreds of volunteers has to accept some limitations. For example, the citizen scientists could not be expected to determine the exact size of the colour clusters. Instead, beekeepers were asked to categorise pollen colours according to their frequency as very rare, rare or abundant. A simplification that was necessary to facilitate participation.

The study design also made it inevitable that different people were used in the visual evaluation. In the context of differences in human colour perception [5], this poses risks that seemed unavoidable at the time. The evaluation of the CSI Pollen study showed that the most significant random effect was attributed to the beekeeper, meaning that different beekeepers would likely find different numbers of pollen colours in the same pollen sample. This is not surprising, since each of the 18 000 samples contained up to 20 g or 2 000 pollen, which made the manual chromatic assessment very tedious. Dimou and Thrasyvoulou [72] report that sorting a 20 g sample takes an

average of 84 min, resulting in a theoretical total workload of up to 33.4 hours per participant.

Depending on patience and concentration, it was up to the participants to sort the pollen carefully (i.e. to form many and therefore homogeneous clusters) or to work fast (i.e. to form fewer and more heterogeneous clusters). The authors, unlike Conti et al. [63] and Kirk [123], decided not to provide the participating beekeepers with fixed reference colour charts. Without a reference, the decision on whether to classify colours as different (or not) was left exclusively to the beekeeper.

Practical reasons against reference colours are the difficulties in providing suitable high-quality prints (c. f. Kirk [123, 121]) and possible problems when pollen cannot be clearly assigned to one of two reference colours. Hence, no colour name could be assigned to the manually sorted pollen in the CSI study. This information and an exact cluster size could not be obtained for practical reasons and could not be included in the study's scope.

Nevertheless, the CSI pollen study is of great impact. Through the involvement of citizen scientists, the project managed to expand across Europe, making it 'much larger than individual researchers could ever achieve' [41]. While this holds true, further automation is the logical consequence of this proven idea.

7.1.3 Contributions

In this work, we seek to objectify, accelerate and extend the manual chromatic assessment carried out before. By presenting an app that allows quantifying pollen load from pollen traps and determining their colour diversity objectively and automatically, the aforementioned shortcomings of the chromatic assessment can be solved. Compared to the human based chromatic assessment approach [63, 41], the proposed solution allows calibrated cluster colours to be determined, making the colours comparable across time and place. Calibrated pollen colours allow the determination of conversion factors from colour diversity to species diversity at the level of individual colours, as previously suggested as future work by Conti et al. [63]. This means that samples that show rare colours are also likely to contain fewer species in a laboratory analysis and vice versa.

The automated chromatic assessment offers the possibility to use the general abundance of pollen colours in the field for the prediction of local species diversity, which is expected to strengthen the correlation and to improve the prediction accuracy. In addition, the coarse colour frequency categories from Brodschneider et al. [41] can be replaced by real count values, allowing the calculation of popular biodiversity indices that require knowledge of cluster sizes.

By making less of an imposition on citizen scientists, it is likely that even more participants can be reached for even longer periods of time with fewer dropouts. Digital support for the beekeeper is also called precision beekeeping and has already been used successfully elsewhere [35, 36].

In this way, the app provides beekeepers with important information about the nutritional status of their colonies, while scientists benefit from aggregated information about local and temporal biodiversity. The web app is freely available on all devices.

7.2 Materials and Methods

The app named 'Pollenyzer' allows users, i.e., beekeepers, to analyse the colour diversity of pollen samples. At the same time, it invites to enrich each pollen analysis with scientifically relevant metadata. The software underlying Pollenyzer is open source and publicly available.[1] The automatic chromatic assessment includes five steps. (1) A pollen trap is mounted and pollen is collected for a typical period of one to three days. (2) The collected pollen is placed on an A5 paper (210 mm × 297 mm) and photographed. (3) The resulting image is uploaded and processed via the app. Optionally, users can provide metadata such as time and location. (4) The user is presented with visualisations of the results. (5) Optionally, users can post-process the automatic clustering of pollen colours according to their own perception.

The app is implemented as a progressive web app that can be accessed and, if desired, installed on all devices via the browser using standard web technologies.[2] In addition to the frontend, the app consists of three backend

[1] *App Repository*, `https://github.com/pollenyzer/beesypollen`. Accessed March 13, 2023.
[2] *Pollenyzer Web Interface*, `https://pollenyzer.github.io`. Accessed March 13, 2023.

components, the implementation of which is explained below. Subsequently, we present a study that investigated whether image processing of pollen samples can achieve the same study results as manual chromatic assessment.

7.2.1 Pollen Load Detection

The literature already contains examples of image processing of photos showing pollen loads. For example, Chica and Campoy [57] and Salazar-González et al. [213] used conventional image processing methods to separate the pollen from background pixels. However, with both methods it was not possible (and necessary) to separate individual pollen loads. In other words, instead of the positions of individual pollen loads, only the area occupied by foreground pixels could be discerned, a disadvantage that can be remedied with today's AI-supported methods.

Convolutional Neural Networks (CNNs) have been successfully used for the last decade and are considered state of the art in the field of image processing [129]. In this work, a CNN, more specifically a U-Net [207], was trained to detect pollen on A5 paper.

Instead of the original U-Net architecture described in Section 2.2.2, a smaller variant was used here. We chose all convolutional layers to have the same kernel size of (3×3), followed by a batch normalisation layer and a rectified linear activation. Two such layers form a convolutional block followed by a max-pooling layer with stride size of (2, 2). The contracting path consists of four such convolutional blocks with 16, 32, 64, and 128 features per convolutional layer. The contracting path is similar in structure, but uses deconvolution layers instead of max-pooling. As is typical for U-Nets, the convolutional blocks in the contracting path are connected to their corresponding counterparts in the expanding path through so-called skip connections. Since the task of detecting pollen is not overly complicated, the size of the network is comparably small containing less than two million trainable parameters.

U-Nets were originally designed for image segmentation, but have already been successfully used for object detection [37]. In the present case, the network's output is a single layer segmentation map showing the networks confidence that a pixel is a pollen's midpoint. Therefore, to obtain a list of

pollen coordinates, U-Nets require additional blob detection on the segmentation map at inference time. Here, blob detection is implemented as a search for local maxima on the thresholded segmentation map.

To simplify the pollen detection, the net was trained to detect pollen on white A5 paper. The paper with known dimensions served as a size reference and allows to automatically scale input images according to the expected pollen dimensions. The network was trained on images of shape $96 \times 96 \times 3$ and binary masks as labels. Figure 7.1 shows five training input images (top), labels (middle) and output segmentation maps (bottom). Since detections at the image borders were especially hard and error-prone, a workaround was implemented. During training, cross-entropy loss was applied only to the 48×48 pixel centre of the image, ignoring potential errors that arise in border regions. In Figure 7.1c one special characteristic of the masks can be noticed. Looking closely, one can see that overlapping neighbouring pollen (white circles) were separated by additional black borders. This was to encourage the network to draw clear boundaries between detected pollen on the segmentation map, thereby facilitating subsequent blob detection. The advantage of this method is evident in the same figure (bottom). Instead of manipulating the label mask, Ronneberger, Fischer, and Brox [207] achieved the same goal by introducing an additional weighting label for edge separation.

7.2.2 Dataset and Annotation Scheme

The pollen came partly from a commercial pollen mix, partly from collected pollen over the course of an entire season and represented a high colour diversity. In total there were 19 996 corbicular pollen loads in the training data set and 2 629 pollen in the test data set. The pollen were taken from a total of 75 images with a resolution of up to 70 MP.

The annotation was done under certain guidelines. (1) If stacked, only pollen mainly located in the uppermost layer and (2) only intact pollen parts large enough to extract robust pollen colours were annotated, excluding pollen fragments. Figure 7.1a illustrates the problem annotators faced when the transition from pollen to pollen fragments was seamless. Thus, it was hardly possible to maintain a very strict annotation style across the entire dataset, as in some cases the correct annotation was ambiguous. This inconsistency is reflected in the fact that borderline cases (see Figure 7.1a) are assigned a mediocre confidence on the segmentation map and are included or excluded

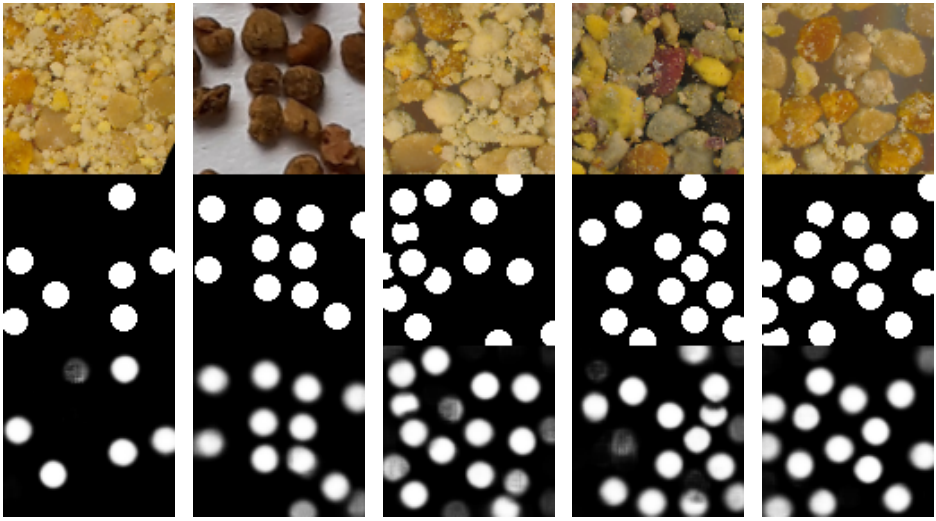


Figure 7.1: Five training samples showing the input image data (top), the annotations (middle) and the neural networks output layer (bottom). Pure white pixels reflect a one hundred percent confidence of the neural network that there is a pollen centre at the pixel's location.

from the results by the choice of confidence threshold. The annotation could be sped up significantly as no real pollen segmentations were done. The much faster annotation of the pollen midpoints was sufficient to generate mask labels afterwards.

A total of 550 000 image parts from the annotated dataset were randomly cropped to the shape 96×96 and augmented. A non-overlapping 50 000 of which have been used for validation purposes. Augmentation included slight scaling, shearing, random rotation, slight alteration of contrast, brightness and colour, and various blurring algorithms. This data set of over half a million images, masks and pollen coordinates was published as part of this work [29].

7.2.3 Training and Validation

Training was performed on a single Nvidia RTX 3090 GPU with a batch size of 64. The RMSProp optimiser was used in combination with a binary cross-entropy loss. The model has 1.9×10^6 parameters and training was stopped after 2 hours as increasing overfitting was observed.

Due to the seamless transition from pollen to pollen parts, not only labelling but also the determination of performance metrics has become more difficult. As the labels were sometimes ambiguous or inaccurate, the calculation of intersection over union (IoU), a typical metric for recognition and localisation problems, is not well suited. In this case, the IoU would mainly quantify the inconsistency of the annotators rather than the quality of the detections. Therefore, pollen detection was performed on separate test images. An annotator then assessed whether a detection fulfilled the previously mentioned criteria. In this way, false-positive, false-negative, and true-positive detections were determined, and more appropriate metrics such as precision, recall, and their harmonic mean were calculated and reported.

7.2.4 Colour Extraction

After the pollens' midpoints have been located, their colours were determined. To obtain a robust colour value, the colour values of multiple pixels needed to be taken into account. However, the naive approach of averaging the pixels in a certain radius around the pollen centre led to poor results. This was due to the observation that pollen was usually composed of two colours, a primary colour and a secondary colour. The secondary colour (or shade) is inherent to the three-dimensionality of the pollen and occurred in any typical lighting situation. For this reason, two alternative approaches were tested to separate the primary and secondary colour: the Gaussian Mixture Model (GMM) and the k-means algorithm. While the GMM assumes that the colours in the sample are normally distributed around two mean values, k-means makes no assumptions about the distribution.

The k-means algorithm is not robust to outliers, but it is much faster, often produces similar results and is therefore preferable in all cases where users are waiting for a timely response. Separating the secondary colour provides additional reliability in cases where the detected midpoints are not perfectly

centred. In such cases the largest cluster's centroid was considered the primary colour.

7.2.5 Colour Calibration

Due to the distributed nature of citizen science studies, we can not expect beekeepers to adhere to specific experimental constraints like matching imaging devices or accurately reproducible illumination. Hence, we expected images to be taken under vastly different lighting conditions and with varying spectral sensor sensitivities of the capturing devices. As a result, different, so-called device-dependent RGB values are assigned to the same pollen colour on different images. This not only impairs the comparability of clustering results between images of distinct participants but also between photographs of a single beekeeper, as there exists no known relation between the respective device-dependent colour spaces caused by differences in illumination and sensor sensitivities. In order to overcome this issue, our goal was to find translations of these colour values from device-dependent colour spaces to a standardised device-independent colour space. This process is commonly referred to as colour calibration [140]. Refer to Section 2.1 for a detailed introduction to colours, calibration and colour spaces.

A widely used method for colour calibration is the use of colour calibration cards. Such cards are of well-monitored production quality, correspondingly expensive and consist of a certain number of colour patches for which reference colour values are known in an absolute colour space based on the CIE Standard Observer. If placed within an image, they allow for a comparison between the observed colour of each patch and the respective reference colour. Thereby, a relation between the device-dependent colour space and the reference colour space can be inferred. Using this relationship, image colours can be transformed from device-dependent RGB values to an absolute colour space, i.e. CIEXYZ, CIELAB or sRGB. In case of a citizen science study, however, each participant would have to obtain their individual colour card, which would be a significant expense. This places an additional burden on the motivation of voluntary participation in such studies. Hence, a suitable and severely less expensive alternative needed to be found.

Examples are provided in the literature, where established colour checkers were substituted with custom devices that fulfil the requirements of a specific domain. Bautista, Hashimoto, and Yagi [8] constructed custom colour patches

small enough to be used in whole slide scanning, while Zhang, Nie, and Zhao [261] designed a colour card with reference colours that are particularly prevalent in diagnostic imaging of tongues. Salazar-González et al. [213] chose to use designated photo equipment under laboratory conditions to ensure pollen colour fastness in a food context. Chica and Campoy [57] circumvented the calibration by using a special computer vision device to detect non-local pollen loads by colour. Our primary constraints, however, were acquisition cost and unrestricted availability of the device to participants. We also wanted to avoid the expensive production and distribution of self-made colour cards. Hence, the object we were searching for should already be a commonplace item. Moreover, the replacement needed to comprise a collection of standardised colours suited for colour calibration.

An item that met these requirements can be found on the packaging of Kellogg's (TM) products: the PrintSpec colour strip by Mellow Colour (TM), placed on the bottom of most Kellogg's cereal products. It is designed to monitor colour consistency of printers in packaging plants and consists of 24 individual colour patches and is presented in Figure 7.2. Participants may easily obtain their copies from a wide array of grocery stores at a significantly lower price than conventional colour checkers. There are a number of factors that can influence the appearance of colours during the printing process of such packaging. These include the type and colour of the substrate and the specifics of the printing method used. For this reason, we obtained the reference colour values for the PrintSpec colour strip from the corresponding MediaStandard Print - a guideline for standardised printing processes - on which the Kellogg's printing process is based.

To allow colour calibration, the user can roughly cut out the PrintSpec colour strip and place it anywhere in the image. Similarly to the pollen detection procedure in section 7.2.1, we employed a U-Net for the automatic detection of PrintSpec colour strips in images. Once the the PrintSpec was identified by the network, we performed a perspective correction, extracted the colour patches and used these colours in conjunction with the known CIELAB ($L^*a^*b^*$) reference colours for the given packaging material to perform colour calibration. For this task, we employed the respective algorithms implemented in the colour correction module of the `opencv-contrib` (v4.6.0.66) python package. These allowed us to find a linear transformation from the linearised device-dependent colour space to some absolute colour space - in our case linear sRGB - that minimised the CIEDE2000 colour distance between reference colours and transformed image colours. Finally, the colour correction,

represented by a colour correction matrix, was performed by simple matrix multiplication.

The CIEDE2000 colour difference equation is the latest evolution of the CIE76 colour distance that was first proposed in 1976 and essentially describes the Euclidean distance in CIELAB colour space (see Luo, Cui, and Rigg [149] on the development of the CIEDE2000 formula and Robertson [204] on the CIE76 colour distance). Both formulae calculate distances between colours in the CIELAB colour space and as such aim to represent the human perception of colour. In experiments by Mokrzycki and Tatol [164] the following was found: CIE76 differences of $0 < \Delta E \leq 1$ are not noticed by a human observer, $1 < \Delta E \leq 2$ are noticed only by experienced observers, $2 < \Delta E \leq 3.5$ are also noticed by inexperienced observers, $3.5 < \Delta E \leq 5$ are noticed as a clear differences in colour while values above 5 are perceived as distinct colours. More details can be found in Section 2.1.

To answer the question to what extent an almost free colour correction with everyday objects can replace a professional colour correction, we tested our method on 10 pictures taken at intervals of one hour over the course of a day. This test setup was intended to cover the natural light changes in our recommended photo setup using indirect daylight. As shown in Figure 7.2, these images contained a professional Calibr8 ColorChart SG XS, a PrintSpec colour strip taken from a common packaging and a print of 16 circular patches of representative pollen colours that were taken from a well-printed book cover. We performed colour calibration with both the ColorChart and the PrintSpec colour strip (our method) and compared the results, which are shown in Section 2.1.

7.2.6 Pollen Clustering

The clustering of the extracted pollen colours is an essential part of the automatic chromatic assessment and can be performed with [63] and without [41] reference colours (cluster centroids). Due to a lack of suitable reference colours, a hierarchical clustering of the pollen colours was performed and flattened at a inter-centroid distance of 20. With the calibration being optional, the interpretation of this threshold cannot be more precise than that it has been chosen to roughly mimic the clustering of people in a typical daylight situation.

Figure 7.2: Test image containing (from top to bottom): the Calibr8 ColorChart SG XS, a cutout of the PrintSpec colour strip (our method), and a print of 16 representative pollen colours arranged in a semicircle.

It is known that people perceive colours differently and take different care when sorting pollen, suggesting that humans do not serve well as a universal benchmarks. Therefore, in order to meet the expectations of individual users, the possibility is offered to manually adjust the predicted number of clusters. Although the threshold may seem arbitrary, it produces consistent results if a colour calibration strip is provided.

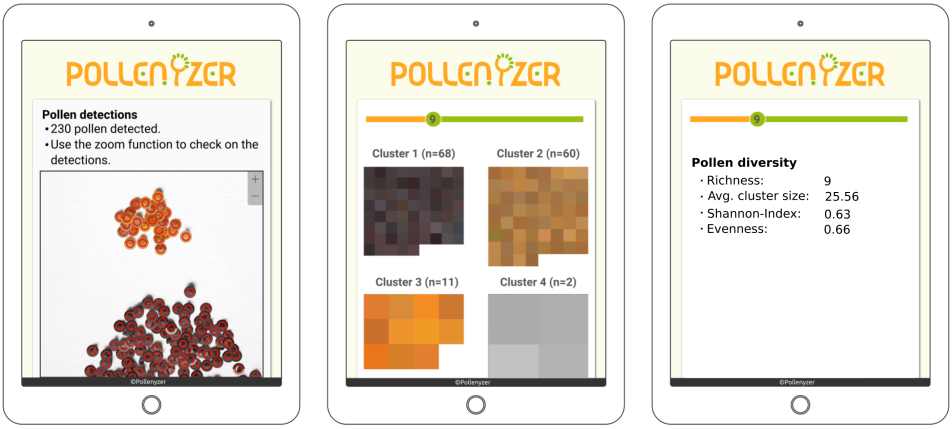


Figure 7.3: Illustration of the app's user interface. Left: Red circles on the input image highlight detected pollen. Middle: All extracted pollen colours are displayed sorted by cluster. A slider allows to change the number of clusters and thus their homogeneity. Right: Pollen diversity indices are calculated according to the number of clusters.

7.2.7 User Interface

The user interface allows interaction with the app on all devices. There are two simple user flows for data submission and data analysis. The first flow consists of an explainer, a page with the purpose of collecting metadata and the possibility to take or upload a photo. The type of metadata inquired is based on the CSI pollen survey mentioned earlier and includes information on location, landscape and collection duration. Namely, the user is asked to provide the coordinates of the apiary, the day of opening the trap, the number of days that the trap was closed and to choose the most appropriate habitat type [39]. After receiving the server's response, results are processed and displayed to the user as shown in Figure 7.3. All detected pollen loads are first highlighted on the input image so that users can convince themselves of the detection accuracy (left image). The next display illustrates all extracted pollen colours for each cluster found. A slider allows to increase or decrease the number of clusters if the colours seem too heterogeneous or homogeneous. Depending on the number of clusters, the calculated biodiversity indices are displayed. In future versions it will be possible to export the results and display chronological data.

7.2.8 Reproduction Study

To make sure that the app serves its purpose, the study of Conti et al. [63] already presented at the beginning was reproduced. There, a model was set up to investigate the linear relationship between the number of pollen colours observed and the number of palynological types determined using light microscopy. For both the dependent variable (palynological types count) and the independent variable (colour count), we only considered types or colours that accounted for more than one percent of the sample.

Pollen was collected from two full-sized honey bee colonies using commercial pollen traps (Heinrich Holtermann KG, Brockel, Germany). The traps were placed at the entrance of the hives and activated once a week for 24 hours. Sampling took place from April to September 2022 in Almke near Wolfsburg, Germany. A total of 54 samples were collected, of which 35 samples were palynologically analysed at the Expert Centre for Bees and Beekeeping, Mayen, Germany. The results of the analysis and a sample preparation protocol can be found in the supplementary of Borlinghaus, Jung, and Odemer [31], the image material was uploaded to a data repository [32]. Mix-ups occurred in two samples and another two samples had to be discarded due to the low quantity, so that a total of 31 pollen samples were available for further evaluation with the Pollenyzer app.

Each pollen sample was placed on a white A5 paper together with a printer test strip and photographed with a smartphone camera (Samsung Galaxy S8) in indirect sunlight at a north-facing window. For a high quality reference, the pollen was also scanned at a resolution of 1200 dpi. Both, photographs and scans are made available [32]. The Pollenyzer app was used to detect the pollen pellets, extract calibrated colours and form clusters. Analogous to Conti et al. [63], the number of palynological types s, was modelled based on the number of colours c identified in the sample:

$$s = \beta_0 + \beta_1 c + \epsilon \ . \tag{7.1}$$

The linear model was estimated from the 31 samples and the strength of the linear relationship was compared with the original study.

true positives	false-positives	false-negatives	precision	recall	F_1
1924	24	88	98.77 %	95.63 %	97.17 %

Table 7.1: Six metrics demonstrate the app's capabilities in pollen detection.

7.3 Results

The following section summarises the evaluation results of three app functions, namely detection, colour extraction and colour calibration. The results of the reproduction study are listed at the end of this section.

7.3.1 Pollen Detection

On test images that were not used for training or validation, 1 924 pollen were correctly detected. A correct detection is characterised by both annotation guidelines being met. Therefore, the detected pollen loads must be placed mainly in the foreground and its size must allow robust colour extraction. The latter is not the case, for example, if a detection deviates from the pollen centre or pollen fragments of small size have been detected as pollen and consequently no colours can be extracted. Table 7.1 shows that in 24 cases false positive and in 88 further cases false negative detections were reported. Overall, the precision was 98.77 %, the recall (or sensitivity) was 95.63 % and the harmonic mean of both values (F_1) was 97.17 %. It was observed that both false negative and false positive errors are more likely to occur with dark pollen, as there is a risk of confusion with shadows.

7.3.2 Colour Extraction

Figure 7.4 shows exemplary the extracted pollen colours from various pollen detections. The area between grey circles is filled with the extracted primary colour, leaving the inner circle untouched for reference. The extracted pollen colour reproduces the pollen colour so well that it is hardly possible to distinguish it from the pixels in the inner circle. Note the slight shading within the inner circles, that could be isolated by extracting the primary colour, making

the colour ring appear slightly brighter and purer. This visual examination shows that the chosen method serves its purpose well.

Figure 7.4: The illustration shows detected pollen loads and their extracted colours. The area between the grey circles is filled with the primary colour that was determined with k-means in the vicinity of the detected pollen centres.

7.3.3 Colour Calibration

The aim of colour calibration was to ensure that a subject always had the same colours after calibration, regardless of the lighting conditions and hardware used to capture an image. In our particular case, there was the additional objective of achieving the same quality during the calibration process as when using professional equipment.

To test the extent to which both objectives can be fulfilled by our method, a series of images were taken covering typical lighting scenarios. Following the procedure of Kirk [123], we recommend taking the pollen images in indirect daylight, ideally at a north-facing window. This will prevent unnecessary colour variation and give more consistent results. Adhering to this guideline, ten images were taken with the same mobile phone camera at one-hour intervals over the course of a day. The colours recorded in these images changed according to the position of the sun. Each image included a professional

CIE76			CIEDE2000		
Mean	SD	Max	Mean	SD	Max
5.97	6.44	25.20	3.63	3.85	14.42

Table 7.2: CIE76 and CIEDE2000 colour distances between corresponding colours before colour calibration.

colour chart, a PrintSpec colour strip and a high quality print of 16 different typical pollen colours (see Figure 7.2).

To determine the colour differences caused by light and camera before any calibration, we calculated the average pairwise distance of all 16 pollen colours to their nine corresponding pollen colours in each of the other nine images. The results are shown in Table 7.2 and distances are reported for both distance metrics CIE76 and CIEDE2000. With average distances of 5.97 (±6.44) for CIE76 and 3.63 (±3.85) for CIEDE2000, equivalent colour patches can be clearly perceived as two distinct colours by a human observer [164]. If photographs are taken in unfavourable conditions, such as with a flash or candlelight, even greater variations in colour can be expected and the need for calibration becomes even more urgent.

As expected, the colour distances between different images decreased after colour calibration (see Table 7.3). For the Calibr8 colour checker, the average pairwise colour distances decreased to 3.56 (±3.08) for CIE76 and 2.31 (±2.08) for CIEDE2000. Using our method, the average pairwise colour distances decreased to 3.37 (±2.76) and 2.03 (±1.52) respectively.

However, these figures alone do not show whether our method can replace a professional colour checker, as they only indicate that colours become more consistent across images after calibration, but do not reveal anything about

	CIE76			CIEDE2000		
Calibration Method	Mean	SD	Max	Mean	SD	Max
Calibr8	3.56	3.08	19.15	2.31	2.08	12.58
PrintSpec	3.37	2.76	18.83	2.03	1.52	8.02

Table 7.3: CIE76 and CIEDE2000 colour distances between corresponding colours after colour calibration for both methods.

CIE76			CIEDE2000		
Mean	SD	Max	Mean	SD	Max
4.95	1.68	9.63	3.96	1.37	9.50

Table 7.4: CIE76 and CIEDE2000 colour distances between calibration results of the Calibr8 ColorChart and the PrintSpec colour strip.

the absolute nature of these colours. Therefore, we also calculated the average colour distance between the calibration results of the PrintSpec and Calibr8 colour checker for each image. The smaller the distance, the more similar the results of the test strip will be to common practice. The results are reported in Table 7.4. We could observe an average distance of 4.95 (±1.68) for CIE76 and 3.96 (±1.37) for CIEDE2000.

7.3.4 Reproduction Study

Figure 7.5 shows the calibrated colours of a total of 40 918 pollen loads detected on 31 samples. A dimensional reduction was carried out to represent the three-dimensional colour space. The contour lines provide information about the frequency of the individual pollen colours. The dark areas around the purple hue most likely indicate the presence of Phacelia (*Phacelia tanacetifolia*), for which this colour is typical. As expected, many different shades of yellow are represented in high abundance. Shades of red are also prominent and, in contrast to light grey, much more common.

Following Conti et al. [63], a linear regression was conducted to explain the dependent variable of palynological types s, by the independent variable, pollen colours c. For the samples analysed, the fitted regression model is given by $s = 1.65 + 0.39c$, with $F_{1,29} = 24.10$, $p < .000$ and $R^2 = 0.45$. Figure 7.6 shows the distribution of the data alongside the regression line.

7.4 Discussion

At the beginning of this work, it was argued that automating the chromatic assessment of pollen from pollen traps has the potential to be more objective,

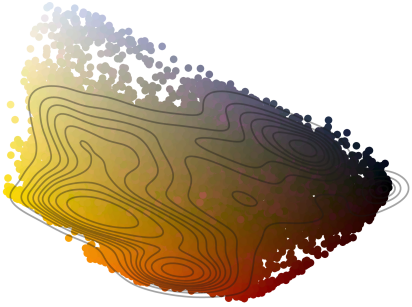


Figure 7.5: The pollen colour spectrum collected from two bee colonies at a site in Almke near Wolfsburg, Germany over the period of one bee year. The colours were calibrated using print test strips and were taken from a mobile phone camera. The contour lines mark the frequency of the total of 40 918 pollen colours shown.

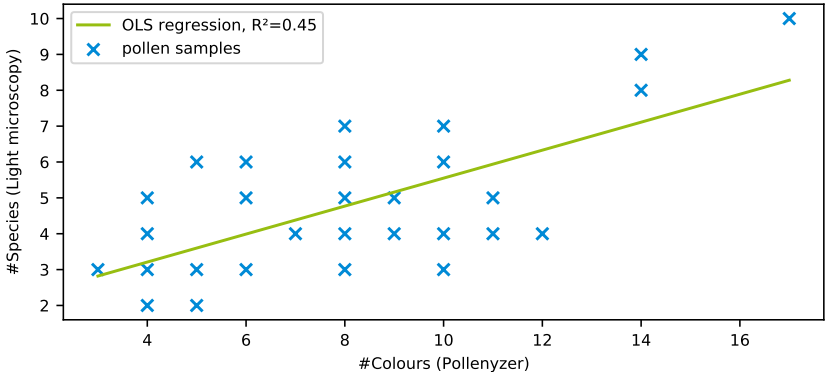


Figure 7.6: The regression line shows the correlation of pollen colours and species in a pollen sample ($N = 31$). The coefficient of determination $R^2 = 0.45$ is almost identical to the original study ($R^2 = 0.44$) and shows that manual and automatic chromatic assessment are equally suitable to estimate the number of species from the observed pollen colours of a pollen sample.

reproducible, faster and more comprehensive than the status quo. For example, determining the size of pollen clusters allows more accurate prediction of colour diversity, the faster procedure reduces the labour input of study participants and can thus increase willingness to participate, and the uniform application of an algorithm eliminates the subjectivity of the human factor

in assessing colour similarity. All this motivates the implementation of the automatic chromatic assessment in the form of the Pollenyzer app.

The evaluation of the app has shown that pollen loads, if photographed on a white A5 paper, can be detected and counted almost without errors. The data set used to train the pollen detector was published along with this work, allowing others to further improve its accuracy [29]. In addition, the subsequent colour extraction allows the pollen colour to be reproduced in a particularly 'pure' way through the pixel-by-pixel classification of the colours into primary and secondary colours. This procedure also makes the app more robust against imprecise detection of pollen midpoints, as outliers up to half of the pixels can be effectively eliminated.

Until now, one of the obstacles to automating chromatic assessment has been the colour inconsistency of pollen images from mobile phones. We have succeeded in solving this problem with simple means. The use of an everyday object, namely a print test strip, makes both a photograph under studio conditions or the purchase of an expensive, professional colour chart redundant.

Surprisingly, compared to a professional colour chart our method achieves slightly better colour consistency, though imperceptible, between images of the same pollen colours in varying lighting conditions. It produces smaller average distances, shows slightly less variation and has a smaller maximum deviation than the professional counterpart. The unlikely fact that a print test strip leads to better results than a professional colour chart might be due to the slightly more central position of the print test strip in the photo. For both models, the average colour differences after calibration was less pronounced, but still noticeable for inexperienced observers [164].

In addition to the individually achieved colour consistency, the models' interchangeability was also studied. As this requires the comparison of two calibrated colorimetric values which, as noted above, are not 100 % consistent themselves, this comparison is limited. All the more surprising is the small CIE76 difference of 4.95 between the two methods and the comparatively small scatter. The calibration results of both methods are, on average, perceived as clearly different but not as a different colours [164].

The low variation of 1.68 indicates a tendency towards systematic error, which could be explained by unwanted differences in the target colour spaces. Although the PrintSpec test strip is used to monitor printing processes and

therefore warrants for consistent colours by design, industry tolerances exist and must be taken into account. A deviation from the target colours of the print test strip within the respective tolerances could explain the observed shift between the PrintSpec target colour space and the Calibr8 target colour space.

We have therefore shown that, in a typical test setup for this application, a PrintSpec colour strip can provide comparable colour calibration results to the common practice. It should be emphasised that perfect colour consistency could not be achieved in either case and that the differences between the two methods are of the order of the general inaccuracy of colour calibration. In terms of human perception, an inexperienced observer would still see differences between the calibrated colours on average, but would no longer perceive them as different colours. Given the biological variability of pollen colours and the inherent tolerances of cluster algorithms, it can be claimed that the method serves its purpose.

When clustering pollen colours, using reference colours (i.e., fixed cluster centroids) instead of hierarchical clustering would enable a simpler and more informative representation in time series data. However, this procedure would not only require the curation of a suitable number of calibrated reference colours, but also make calibration for any input image mandatory. When curating, it is important to note that the selected colours should not only have similar distances to each other, but that the distances should also incorporate tolerances that reflect the expected colour degradation due to ambient light, sensor capabilities and calibration. The number of reference colours and their similarity should adapt to the precision available today. If better and better cameras become available in the future, the number of reference colours can be higher and their colours more similar. Whereas until now species diversity has been estimated by colour diversity, the introduction of distinct reference colours would allow species diversity to be estimated more accurately by the occurrence of specific pollen colours, as proposed by Conti et al. [63]. This notion is based on the following argument: At a hypothetical apiary, ten plant species are available. Few of them produce rare pollen colours (e.g. purple in Germany), many of them produce common colours (e.g. yellow in Germany). Given a pollen sample from this site that shows only one rare colour (e.g. purple), one can already assume that even a laboratory analysis will reveal only a few or even a single palynological types, because it would be very unlikely to find two species at the same location that produce a rare pollen colour. Exactly the opposite is true for a sample that shows a single

common colour (e.g. yellow). Now, one can expect that many palynological types will be found under the microscope. Thus, observing a rare colour should have a small impact on the prediction of species diversity, whereas observing a common colour should have a large impact. However, up to now, only the number and proportion of colours have been taken into account but not which ones.

Yet there are also disadvantages to the reference-colour approach and reasons that ultimately lead to the choice of implementing a hierarchical clustering. Firstly, it should be noted that reference colours can only show their full strength in the context of colour calibration, which is currently offered optionally and is not required by users. Secondly, when bees bring in pollen of a species that falls exactly on a colour boundary, each pollen of this species will be randomly assigned to one cluster or the other according to its natural colour dispersion. If such a case occurs, biodiversity indices based on it will be unrealistically high. Thirdly, the compilation of pollen colours requires a large database of calibrated pollen colours, preferably acquired through the Pollenyzer app itself. However, the compilation of such a database will only begin with the release of the Pollenyzer app and will not be available in the short term. For these reasons, the implementation of a pollen clustering based on reference colours is considered future work. By then, a hierarchical clustering of the pollen was performed and flattened at a threshold that approximately mimics the clustering of humans.

The ultimate goal of estimating biodiversity based on observed colour diversity is subject to known limitations, regardless of the method used. These include the fact that different plant species can produce very similar pollen colours and, in rare cases, the same plant species can produce different pollen colours [123]. Moreover, pollen colours can change with fluctuating humidity and composition [105]. Despite these general limitations, a linear relationship between colours and species with an R^2 of 0.44 has been established by Conti et al. [63]. On the one hand, replacing manual colour assessment with automatic assessment introduces two additional sources of error: the need for pollen detection and colour calibration. We have shown that pollen detection is extremely accurate, even with overlapping and damaged pollen. However, colour calibration is more difficult to evaluate as PrintSpec colour strips are subject to limited but unknown colour variations due to printing. In our test case, calibration with the PrintSpec colour strip yielded slightly different but more consistent colours than the professional alternative at a fraction of the cost.

On the other hand, automating colour assessment also eliminates an important source of error: the human factor. Differences in colour perception, patience and thoroughness lead to a high degree of subjectivity in pollen sorting. As shown in the CSI pollen study, the random factor 'beekeeper' is the most important factor influencing the number of pollen colours found.

The application of the app on real samples was conducted to show whether it also yields plausible results in practice. For this purpose, the study on the possibilities of (manual) chromatic assessment by Conti et al. [63] was reproduced. When comparing the results, it is noticeable that both the intercept (1.65 vs. 3.49) and the coefficient (0.39 vs. 1.52) are lower for Pollenyzer. The reason for this is partly due to general differences in the environment of the apiaries and partly due to the fact that Conti et al. [63] used about 72 hour collection intervals and the samples used here were collected in only 24 hours. Shorter collection intervals lead to a generally lower number of species and correspondingly lower intercept. As the regression's slope depends on factors we could not reproduce, i.e, location and season, their comparison is of little use. However, the coefficients of determination (R^2) show that the manual and the chromatic evaluation are equally suitable for determining the species diversity in a pollen sample (0.45 vs. 0.44) and only differ in the time required.

Therefore, the increased objectivity and the enormous time savings make automatic assessment an attractive alternative, despite the potential introduction of new sources of error. Nevertheless, the accuracy of the method should not be overestimated. This also means that the temptation to choose cluster thresholds lower than justified by the natural variability of pollen colour, lighting and colour extraction should be resisted.

7.5 Conclusion

Biodiversity monitoring is a task that attracts the attention of researchers, policy makers and other stakeholders worldwide. With the Pollenyzer app, beekeepers get access to an automated tool that can determine pollen colour, and thus plant diversity, in the landscapes where their bees forage. In other words: We can see the landscape through the eyes of bees and identify what is beneficial to them. The app is able to calibrate images and make them comparable across time and place. Databases can be created with metadata

and georeferences to map plant diversity in a standardised way. Since we have shown that automatic pollen colour identification can facilitate and replace all aspects of manual colour identification, future citizen science projects to create pollen databases are easily possible. Nationwide projects and eventually PAN European Networks could provide an important monitoring tool for the development of bee-relevant plant diversity in the EU.

8.

Parts of this chapter are based on

NATURAL COLOR DISPERSION OF CORBICULAR POLLEN LIMITS COLOR-BASED CLASSIFICATION [34]

Parzival Borlinghaus[1], Frederic Tausch[2], Richard Odemer[3]

[1]Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]apic.ai GmbH, Karlsruhe, Germany

[3]Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, Institute for Bee Protection, Braunschweig, Germany

CRediT

Conceptualisation: PB, FT; Data curation: PB, FT, RO; Formal analysis: PB; Investigation: PB; Methodology: PB; Project administration: PB; Resources: RO; Software: PB; Visualisation: PB; Writing - original draft: PB; Writing - review & editing: FT, RO.

Publisher

ISPRS Open Journal of Photogrammetry and Remote Sensing published by Elsevier B.V. on behalf of International Society of Photogrammetry and Remote Sensing (isprs), vol. 12, p. 100063, Apr. 2024, doi: 10.1016/j.ophoto.2024.100063.

8.1 Introduction

Pollen collection by honey bees (*Apis mellifera* L.) is of multiple ecological and agricultural importance [184]. At its core, pollination is a vital ecosystem service with immense economic value that supports agricultural productivity and biodiversity conservation efforts worldwide. The importance of this service is particularly pronounced given agriculture's increasing dependence on pollinators, a trend underscored by the Biodiversity and Ecosystem Services (IPBES) [21]. Recent findings by Casas Restrepo et al. [52] have highlighted the potential of pollen-based regional differentiation and offer new insights into the geographical classification of apicultural products. Consequently, the analysis of pollen is emerging not only as a means to delineate honey varieties, but also as a promising way to monitor biodiversity dynamics in different ecosystems [163], facilitating informed decision-making in land conservation and management.

In the field of pollen analysis, a variety of methods have been developed to recognize plant taxa within samples, each offering unique advantages and insights [51, 135]. DNA metabarcoding, for example, has revolutionised our understanding of floral diversity by revealing a broader range of plant taxa in the vicinity of beehives compared to traditional manual surveys [163]. Similarly, studies utilizing amino acid composition analysis have successfully delineated pollen samples based on temporal and spatial variation, providing valuable temporal insights into ecosystem dynamics [3]. In addition, the integration of spectroscopic techniques, including Fourier transform near-, mid- and Raman spectroscopy (ATR-FTIR), together with colour analysis and microscopy has enabled precise discrimination of pollen origins [24, 234], demonstrating the versatility of modern analytical approaches in elucidating botanical origins.

Despite the methodological advancements, the practical application of pollen analysis often defaults to the technically less complex yet gold standard method of light microscopy [51, 148, 250]. This technique relies on the morphological examination of individual pollen grains, which are subsequently classified into pollen types utilizing established pollen atlases [80]. However, the limitations inherent in this method become evident when attempting to identify plant taxa at the species level due to morphological similarities among pollen grains. Nonetheless, the designation of pollen types provides valuable insights into taxonomic diversity, with naming conventions reflecting vary-

ing degrees of taxonomic resolution [68]. It is important to understand in the context of this work that one type of pollen can be produced by several species. Notably, the concept of floral constancy, wherein honey bees tend to collect pollen pellets of the same botanical origin during a single foraging flight [124, 174, 186], has facilitated chromatic assessments of pollen pellets, offering a less labour-intensive yet economically viable means of assessing floral diversity [41, 63].

In previous studies we have shown that results from tedious manual chromatic assessment can be reproduced equally well in a fraction of time using the Pollenyzer app, a photograph of a pollen sample and everyday objects for colour calibration [31]. Although accurate colour-based botanical classification remains problematic [122], existing pollen colour books provide valuable guidelines for classification that are supplemented by spatial and seasonal considerations [105, 122]. However, the inherent variability of pollen colours poses a significant challenge for standardised classification systems, especially the natural dispersion of pollen colours and its impact on classification accuracy have not yet been sufficiently investigated. In particular, the quantification of natural colour dispersion requires the compilation of a dataset containing large amounts of pollen of the same origin, as the limited data from hand-picked corbicular pollen collections prove to be insufficient [105, 122].

To overcome this challenge, we combined images of pollen trap contents with standard palynological laboratory analyses. Our approach requires that the pooled results from the laboratory can be linked to the individual pollen colours derived from the image data, i.e. that the gap between the pollen type information at the sample level and the colour information at the pollen load level is bridged. Therefore, we used Gaussian Mixture Models (GMMs) to assign the most likely pollen type from the laboratory results to individual pollen loads. After cleaning the data, we were able to assign 62 354 calibrated pollen colours from 86 samples to one of 30 pollen types, including 14 of the 31 major European types [117]. Subsequently, 253 chromatic means and covariance matrices describing the colour dispersion of the occurrence of each pollen type were determined from the data. Finally, the mean distance to the centre of the distribution was calculated, averaged for each pollen type and interpreted to quantify the type-specific dispersion.

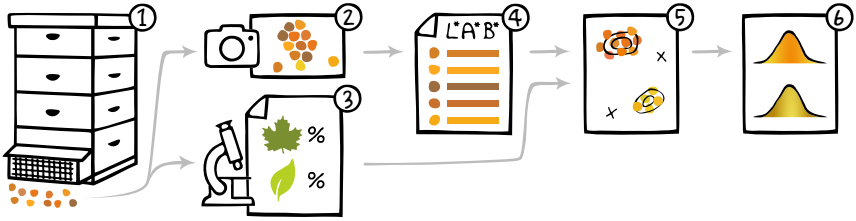


Figure 8.1: Over periods of 24-48h, pollen samples were collected in pollen traps (1). The pollen samples were then digitised (2) and the pollen types identified (3). Using the Pollenyzer software, the calibrated pollen colours were extracted from the photos (4). Based on prior information (3) and pollen colours (4) a GMM was fitted (5) and the distributions' parameters reported (6).

8.2 Materials and Methods

Figure 8.1 illustrates the methodological approach of this study in 6 steps. In step (1), pollen pellets were collected using pollen traps, the contents of which we refer to as pollen samples. Each sample was digitised (2) before the pollen colours were calibrated and extracted from the image in step (4) using the Pollenyzer app introduced in the previous chapter. In addition, the composition of the sample was palynologically analysed in the laboratory (3). Both pieces of information were combined in a Gaussian Mixture Model (5) to determine the type for each pollen load. Finally, in step (6), the type-specific colour dispersion was determined by interpreting the covariance matrices estimated by the GMMs.

8.2.1 Data Collection

In this study, four datasets provided a total of 107 pollen samples, 86 of which were suitable for further analysis and are listed in Table 8.1. Each sample originated from a single full-sized honey bee colony (*A. mellifera*) and was collected over a 24-hour period with an entrance pollen trap [51]. For logistical reasons, the pollen collection for dataset D was extended to a duration of 48 hours.

For datasets A, B, and C, the pollen pellets were imaged in a specially designed light dome setup [240]. This setup was tailored for this particular application. To ensure accurate colour measurements, the camera used was calibrated with

Site	Samples	Collection	Location	Colonies	Image acquisition
A	27	24h, Apr-Sep ’21	Wolfsburg	4	Light dome
B	9	24h, May-Aug ’21	Mayen	4	Light dome
C	40	24h, unknown	3 Federal states	10	Light dome
D	31	48h, Apr-Sep ’22	Wolfsburg	2	Scanner

Table 8.1: Dataset Overview: Samples were collected at various locations across Germany between 2021 and 2022.

a Calibr8 colour chart, resulting in CIELAB colour values for the subsequent analyses (refer to Section 2.1 for details on the CIELAB colour space).

Dataset D was scanned with a commercial flatbed scanner and calibrated with the same Calibr8 colour calibration card. Because scanning was done after palynological analysis, some pollen were missing from the images compared to their microscopic equivalents. Subsequently, the Pollenyzer software was used to detect the pollen on all images, extract the pollen colours and to obtain calibrated pollen colours in CIELAB colour space [31].

8.2.2 Palynological Analyses

All samples were palynologically analysed by light microscopy and expert judgment at the Expert Center for Bees and Beekeeping, Mayen, Germany (for detailed methodological protocol, see data repository in Borlinghaus, Odemer, and Tausch [33]). In brief, each sample was subjected to morphological analysis, which included routine homogenization of the corbicular pollen. Approximately 100 individual pollen grains were then extracted from the mixture and subjected to expert classification to determine the relative pollen type abundance. It is important to point out that although experts can occasionally identify taxa to species level, this is generally not feasible. In such cases, only the pollen type representing a group of often closely related plant taxa is determined. In cases where pollen types are named after a species (e.g., *Centaurea cyanus*), current knowledge indicates that the pollen type (identified by its morphology) is produced exclusively by a single botanical origin. A less precise identification is highlighted by an abbreviated type suffix (e.g. *Taraxacum* T.), indicating that some, but not all, taxa of the given

rank share the same morphology. When pollen grains from a whole family cannot be distinguished morphologically, the type is designated by the family name [68].

In total, the pollen colours from the Pollenyzer app and the proportion per pollen type determined in the laboratory were available for each of the 86 samples. Henceforth, the individual samples were labelled with dataset and sample designation, e.g. sample D03.

8.2.3 Colour Dispersion Model

A Gaussian Mixture Model was adapted to find the missing link between pollen colours of individual pollen loads and pollen type probabilities from laboratory analyses (step 5, Figure 8.1). The exact procedure is explained below.

8.2.3.1 Gaussian Mixture Model

When data consists of normally distributed components, a Gaussian Mixture Model (GMM) can be constructed by combining several (multivariate) Gaussian distributions, each of which has a certain weight related to the component's proportion of the overall data. The parameters of a GMM, the weights, mean vectors and covariance matrices, can be fitted by the Expectation-Maximization (EM) algorithm [70]. The EM algorithm ensures incremental improvements but no global optimum. The iterative procedure is stopped when convergence is reached. Similar to the well-known k-means algorithm [146], the EM algorithm requires an initial parametrization for the weights, means and covariances. When fitting a GMM, it is assumed that the underlying data is a mixture of well-separated Gaussian distributions and that the number of components is known [199].

8.2.3.2 Visualization of Pollen Colours in Two Dimensions

Pollen colours were handled as trichromatic $L^*a^*b^*$ colour values. Only for visualization, the colour dimensions were reduced to two using Principal Component Analysis (PCA). The PCA was fitted once on all available 85 531 pollen colours. To make the visualizations easy to compare, the same

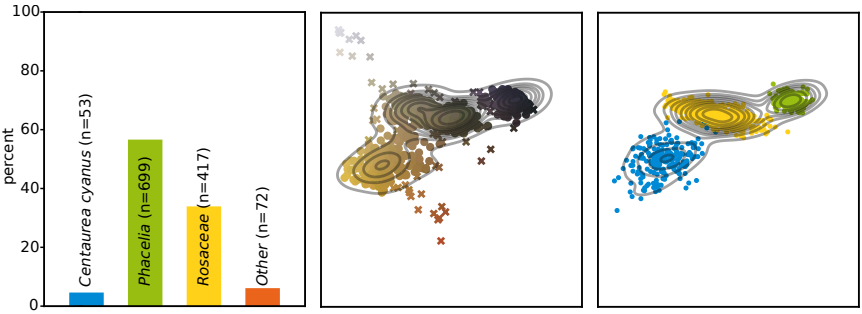


Figure 8.2: (Left) Corresponding palynological assessment. (Middle) Exemplary pollen colour scatter plot of sample D17. Outliers are plotted as ×. (Right) Fitted GMM model.

transformations (and axis limits) were used for all dimensionally reduced $L^*a^*b^*$ pollen colours, both here and in the data repository [33].

8.2.3.3 Cluster Annotation

To make the fitting of the GMM more robust, the initial chromatic mean values were provided by annotation. Figure 8.2 exemplifies how the cluster mean values were annotated with the corresponding pollen type. On the left, the result of the laboratory analysis is shown, which lists three pollen types that account for more than 5 % of the total. In the middle, all colours extracted from sample D17 are shown as a two-dimensional scatter plot. Contour lines provide information about the distribution's density and reveal three distinct clusters. Through prior knowledge, literature and logical combination, the individual clusters were assigned to the three major pollen types: (1) *Phacelia* is known for its distinct dark purple colour [122]. (2) According the laboratory analysis, pollen type Rosaceae is much more common than *Centaurea cyanus* and must therefore produce grey pollen. (3) The remaining pollen type, *C. cyanus*, was identified to species level under the microscope and could therefore be verified as light brown pollen using a pollen atlas. Alternatively, it could simply be assumed to match the remaining cluster. The annotation was simplified by the fact that often only a few, large clusters were found. The laboratory results showed that on average 2.15 pollen types made up an abundance of more than 10 % per sample, which is corroborated by literature [3, 20].

8.2.3.4 Model Assumptions

It is plausible to assume that pollen colours from the same botanical origin follow a normal distribution around a chromatic mean. Given the presence of different pollen types, it is reasonable to expect that the data within a sample is distributed as a mixture of three-dimensional Gaussian distributions.

The number of components were known as they were taken from the palynological analysis. Pollen types with a share of less than 5 % were not considered and added to the share of outliers. Further adjustments to the number of components were made based on visual inspection considering the ease of cluster separation. Samples had to be excluded if the modes of the colour distribution could not be easily assigned to pollen types, i.e. the underlying mixtures were not well-separated. For the remaining samples, the GMM assumption of separability was deemed satisfied.

Considering that a single sample contains only very few abundant species, it is unlikely that two of them have the same pollen type (recall that two species can be so morphologically similar that they may be assigned the same pollen type). It is therefore plausible to assume that in a single sample each pollen type comes from a single floral origin. If this assumption is violated, two cases can occur. Either both species produce the same colours or different ones. The first case can be ignored as it does not change the interpretation of the results and in the second case, the annotation of the pollen types is ambiguous, the sample stands out and can be discarded.

8.2.3.5 Steps to determine the natural pollen colour distribution

Annotation step: The samples were annotated as described above. Samples that could not be accurately annotated were discarded.

EM-initialization: The EM algorithm was used to determine the chromatic mean and the colour dispersion for each annotated pollen type of each sample. Gaussian distributions were initialised with the corresponding annotated cluster centroid. As a starting point for the (full) 3×3 covariance matrix, all pollen in a sample were temporarily assigned to the nearest (annotated) cluster centroid and the covariance matrix was calculated for each such cluster. The weights of the Gaussian distributions corresponded to the known proportion of each pollen type.

EM-algorithm: In each iteration, the mean vectors and covariances were updated. Only the weight vector was treated as known prior information and was kept constant.

Outlier removal: Since GMMs are considered sensitive to outliers, the EM algorithm was run repeatedly and 20 % of the outliers were removed in each of five runs. The number of outliers was calculated beforehand by multiplying the proportion of unidentified pollen types and non-annotated pollen types by the number of pollen in the sample. In each outlier-removal step, those pollen colours were removed that had the lowest probability of originating from any of the fitted Gaussian distributions.

Fitting: After removing the outliers, a final fitting was performed for each sample. The estimated mean vectors and covariance matrices of each pollen type's distribution were filed. Figure 8.2 shows the fitted Gaussian distributions on the right. The number of plotted elements per component corresponds to the proportion of the pollen type in the palynological assessment multiplied by the number of pollen pellets in the sample.

Preparation of results: For most pollen types, several Gaussian distributions have been fitted, i.e. several mean vectors and covariance matrices were available as they occurred in multiple samples. Each of these distributions belongs to the same pollen type, but not necessarily to the same floral origin. To make the covariance accessible to interpretation, the mean Euclidean distance to the mode of the distribution was calculated. The result, a scalar, is henceforth referred to as the standard deviation of the distribution. Such a Euclidean distance in CIELAB colour space is usually denoted as CIE76 and is a measure for colour differences [204]. For each pollen type, the median and inter-quartile range of the standard deviation of the colour distributions were reported. We used results from Mokrzycki and Tatol [164] for interpretation (refer to Table 2.1 in Section 2.1 for a summary).

8.3 Results

On 86 images of pollen samples, a total of 253 pollen types were annotated, 30 of which were different. For each sample, a GMM was fitted and the covariance was converted into standard deviations that quantify the expected colour dispersion from the chromatic mean. Table 8.2 shows the median

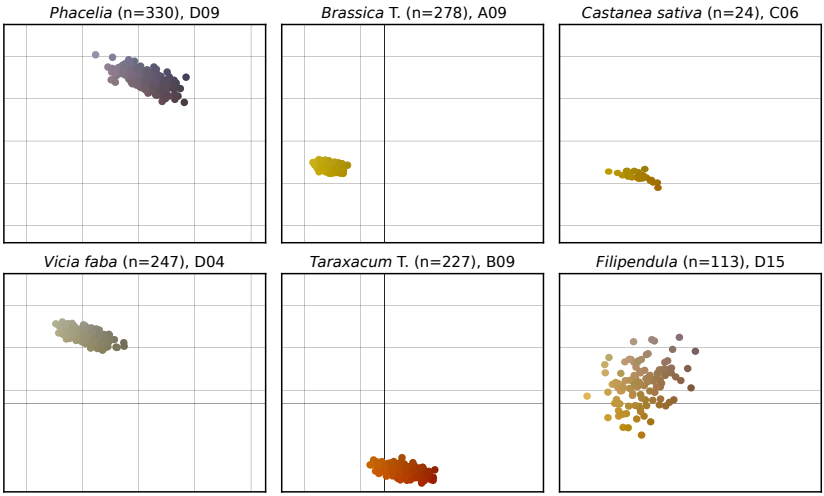


Figure 8.3: Colour dispersion for six exemplary pollen types.

and inter-quartile range of the standard deviation of the colours for each pollen type, as well as the colour means for each occurrence. Figure 8.3 shows selected colour distributions for *Phacelia*, *Brassica* T., *Castanea sativa*, *Vicia faba*, *Taraxacum* T. and *Filipendula*. *Brassica* T. had the lowest colour dispersion (2.94) whereas *Filipendula* had the highest (10.25).

Pollen type	Colour dispersion median (IQR)	Chromatic means
Acer	4.72 (3.84-6.21)	
Achillea T.	4.52 (4.24-4.89)	
Aesculus	9.39 (8.71-10.39)	
Anemone T.	6.19 (5.58-6.79)	
Artemisia T.	5.99 (4.74-7.25)	
Asparagus	4.67	
Balsaminaceae	4.85 (4.38-5.22)	
Brassica T.	2.94 (2.46-5.08)	+20
Buddleja	4.09	
Castanea sativa	5.31	
Centaurea cyanus	5.47 (4.57-6.16)	
Chenopodium	5.19 (4.41-5.69)	
Cornus T.	3.75 (3.32-4.80)	
Filipendula	10.25 (7.84-12.96)	
Hedera	7.80 (7.49-8.91)	
Hydrangeaceae	5.91 (5.87-5.96)	
Parthenocissus	5.53 (5.08-8.03)	
Phacelia	6.67 (4.66-7.43)	+12
Plantaginaceae	5.29 (4.68-7.86)	+5
Poaceae	7.27	
Potentilla	7.51 (6.47-8.55)	
Prunus T., *Pyrus* T., *Rubus* T.	6.39 (5.32-7.30)	+13
Ranunculaceae	6.60 (6.20-7.00)	
Rosaceae	7.03 (6.61-8.50)	+9
Salix	10.22 (7.62-10.37)	+7
Sinapis T.	4.42 (3.73-7.25)	
Taraxacum T.	4.80 (3.58-5.75)	+9
Trifolium pratense	8.50 (8.38-9.09)	
Trifolium repens	6.93 (6.56-7.98)	
Vicia faba	4.82 (4.62-4.98)	

Table 8.2: Chromatic mean values and the estimated deviation therefrom based on the fitted GMM parameters. The latter is measured as the CIE76 colour difference. IQR values are omitted for pollen types with a single observation. Humans perceive colour dissimilarities greater than 5 as different colours.

The data repository contains comprehensive results for individual samples, which would go beyond the scope of the presentation in this work [33]. There, the fitted GMM parameters in JSON format alongside visualizations similar to Figure 8.2 are available in the 'pollen samples' folder. Also scatter plots of the fitted Gaussian distributions for each pollen type are provided in the 'pollen types' folder (Figure 8.3).

Taking the assigned pollen type predictions of the GMM models as ground truth, a pollen classification can be performed that simulates a real-world use case ignoring the class probabilities from the laboratory analysis. Therefore, for all 62 354 available pollen colours that were not previously labelled as outliers, all 30 pollen types were assumed to be equally likely. Classification was performed by selecting the most likely colour distribution (out of 253) and assigning their corresponding pollen type as the resulting class.

The axes of the confusion matrix in Figure 8.4 were arranged by colour, which results in a more compact representation of confusion. As shown on the vertical axis, class sizes range from 22 (*Asparagus*) to 8 309 (*Prunus* T., *Pyrus* T., *Rubus* T.). The diagonal values represent the one-versus-rest classification accuracy which ranges from 0 % (*Anemone* T., *Castanea sativa*) to 99 % (*Phacelia*). Considering all data points, the correct classification rate is 67 % (n = 62 354 , classes = 30).

8.4 Discussion

The availability of food resources in a landscape is discussed as an important factor in the health of honey bees and other pollinators. Quantifying this availability, and thus biodiversity, may be useful for designing specific insect pollinator support programs. Currently, there is a lack of methods that can automatically detect available food resources in relevant landscapes. Identifying plant species by pollen colour could facilitate the evaluation of such support programs. However, it is not known if (automatic) chromatic assessment of pollen is sufficient for more than an estimate of species diversity, as identification has proven to be even more difficult than previously known.

In our study, we collected and analysed 86 pollen samples from beehives in four German states. Using light microscopy, we identified the pollen types and catalogued the results in a database, which was supplemented by

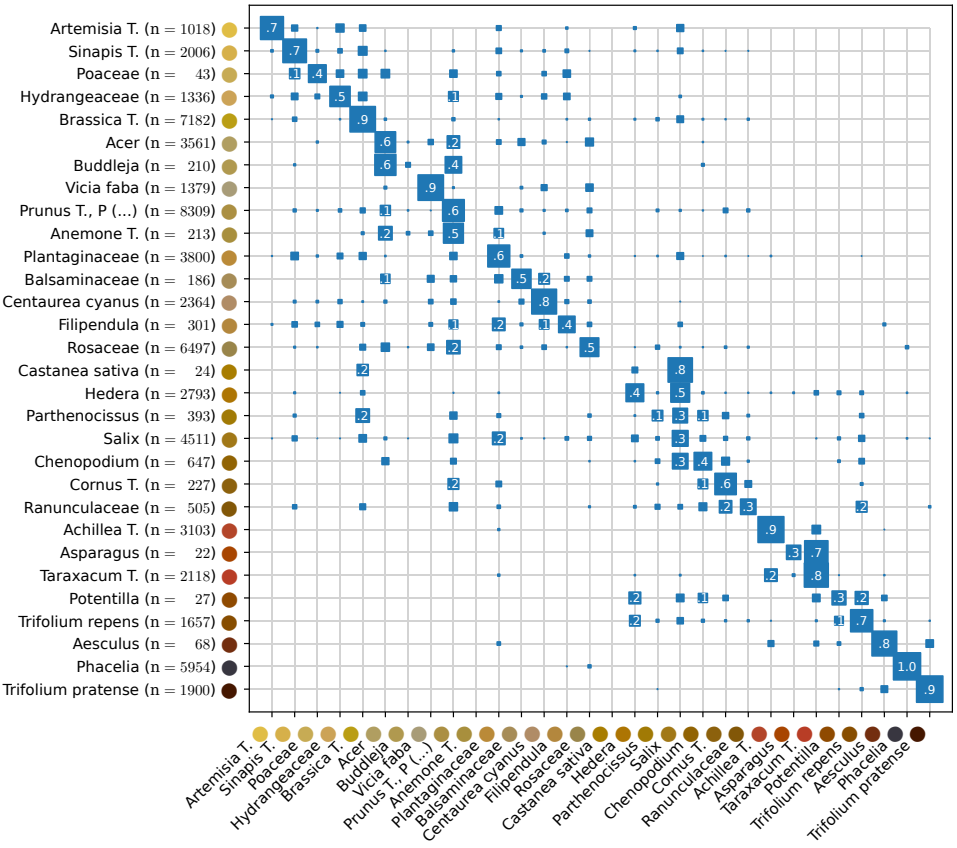


Figure 8.4: Confusion matrix for pollen type classification. Similar colours are likely to be confused. Types are ordered by colour, class sizes are given in brackets on the y-axis. Values greater than 0.1 are annotated and rounded to one decimal place. The true classes are shown on the vertical axis, the predicted classes are plotted horizontally. The values are row-wise normalised. Note how small class sizes led to unstable GMMs that made larger, similarly coloured classes more likely than true classes (e.g. *Castanea sativa*).

calibrated pollen colours. The subsequent adaptation of Gaussian Mixture Models enabled us to interpret the distribution parameters. We were able to show that the expected colour deviation from the chromatic mean was greater than five for two thirds of the species and pollen types in the dataset (6.10 ± 1.82). Thereby, a value of five marks the limit above which humans

perceive colours as completely different [164]. It highlights the complex nature of pollen colours, which exhibit a multitude of nuances even within a single species. Consequently, the effort to categorize pollen colours under a single colour proves insufficient, as the natural dispersion already spans several colours on average.

The measured colours of the pollen type that was both highly abundant and had the lowest dispersion was *Brassica* T. However, the median standard deviation was still 2.94 (an 'inexperienced observer [...] notices the difference' [164]), which leads us to conclude that the classification of individual pollen pellets cannot be successful without additional resources, e.g. in the form of extensive expertise. The finding that plant species produce (or obtain through typical collection practices and storage) pollen colours of an unexpectedly high colour spectrum limits the use of colour books for species identification when using pollen traps as collection method. The reason for this is not primarily that pollen of different origins looks the same, but that pollen of the same origin does not look the same. This result is based on 85 531 corbicular pollen in 86 samples containing 14 of the 31 major European pollen types [117], distributed over several sites and a two-year sampling period in Germany.

While we are aware of the potential limitations of our GMM model, confirmation using samples that allow clear colour assignment supports our conclusions (e.g., A09, A11, A12, A20, A21, A23, C14, C15, C16, C20, C21, C30, C35, C36, C37, D08, D18, D30). By additionally reporting the median and IQR of all colour deviations of a pollen type, outliers were largely excluded from the interpretation, which corroborates our conclusion. As mentioned above, pollen is known to change colour, for example due to drying [51]. We therefore argue that colour dispersion would be lower if the colours were determined immediately, ideally already in the field [121]. While this is true in principle, the collection period of 24-48 hours and storage in the freezer make the results of this work more applicable from a practical point of view since this is common practice [51, 50].

The confusion matrix in Figure 8.4 showed instabilities when estimating a GMM with small class sizes. Against intuition, two types had a correct classification rate of zero. Given the identity of training and test data, one would expect that these two normal distributions would overfit to their corresponding pollen colours. Due to the few data points, the data did not always allow a robust mean estimation leading to slightly higher matching probabilities

with wrong, similar looking types. On the contrary, the confusion matrix also showed that some pollen types can be easily distinguished. For example, *Brassica* T, *Vicia faba* and *Phacelia* produce distinct yellow, grey and dark purple colours respectively. In general, the accuracy of classification always depends on the number of classes and similarity. In another dataset with many dark pollen colours, *Phacelia* could be much more difficult to recognize. However, for Central Europe, especially Germany, we consider the explanatory power of the confusion matrix to be acceptable due to the diversity of collection sites and duration. The same caution applies to the generalization of the overall correct classification rate of 67 %, which is also closely linked to our dataset. It would be desirable to validate the classifier on a test split at the sample level. However, this would have required that both sets contained the same species, as the species determine the pollen colour. Since we only had pollen type class labels, this approach deemed unfeasible. Convinced that within-sample splitting does not test the generality of the results, we did not perform a test split. We claim that this is permissible, as the classifier is not used in an inferential but a purely descriptive way and is therefore inverted in its typical application.

In contrast to *Brassica* T., which we described as a relatively homogeneous pollen type with little dispersion, the dispersion in *Filipendula*, for example, was strikingly large (see Figure 8.3 for examples). We attribute this to the fact that in the samples containing Filipendula (D13, D14, D15, D16), the proportion of Filipendula pollen, each below 5 %, was too low to allow for a robust GMM estimation. Additionally, each sample also contained *Sinapis* pollen, which is very similar and present in larger proportions, which may have violated the assumption of well-separated Gaussian distributions. Another source of overestimated colour dispersion could be samples where different species produce the same pollen type with a similar colour, which may have been mistakenly approximated by the same Gaussian distribution..

When comparing the photographed pollen with the results of our melissopalynological laboratory analysis, it became clear that a laboratory analysis is far from an ideal method. On the one hand, only a limited number of pollen grains are examined under the microscope, which inevitably leads to different proportions per pollen type depending on the subsample size. On the other hand, humans themselves also appear to be a source of error. For example, in the selection of pollen grains examined and the level of rigour. Obvious discrepancies between laboratory findings and actual pollen type occurrence even led to the exclusion of samples in our study.

Recent technological advances in pollen research could provide a more reliable identification. Dunker et al. [75] combined multispectral imaging flow cytometry and deep learning to rapidly and accurately identify species, quantify pollen grains, and extract features of recent pollen. In addition, metabarcoding methods could also be considered to improve the reliability, cost-effectiveness, and processing speed of pollen analyses [51].

For example, we found a notable discrepancy in sample C07. The laboratory tests showed that 22.67 % *Phacelia* pollen was present in this sample. In fact, only 8.18 % corbicular *Phacelia* pollen was visible in the photo, which could be reliably identified due to its distinct colour (see Table 8.2). Laboratory tests carried out according to the common standard [148, 250] can therefore sometimes deviate considerably from reality. With an estimated 10 700 pollen grains morphologically analysed for this study, errors are to be expected and the question is raised whether sampling of pollen grains is sufficiently standardised to obtain comparable data.

In our previous studies [31], we achieved an overall precision of 98.77 % for automatic pollen detection, indicating low error rates. However, it is important to point out a limitation we identified: The algorithm occasionally failed to detect dark-coloured pollen, mistaking it for shadows in the image [31]. This systematic overlooking of darker shades could lead to a bias in our analyses. Further limitations result from the use of GMMs. The method assumes that pollen colours of the same botanical origin follow a normal distribution around a chromatic mean and that the dataset as a whole is distributed like a mixture of three-dimensional Gaussian distributions. Such assumptions may not hold true in all cases and could lead to inaccuracies in colour determination. Furthermore, the annotation step relies on prior knowledge, literature, and logical combination to assign pollen clusters to specific types. This process introduces subjectivity and potential errors, particularly in cases where clusters are not clearly separable or when multiple species produce similar colours. This must be considered when interpreting the results.

While we argue that today individual pollen cannot be reliably classified by colour because of colour dispersion within species, this might be possible with a sample of many pollen of the same origin. A larger number of pollen would allow estimation of both a robust chromatic mean and the covariance matrix. Because the structure of the covariance matrix varies, it could be used to identify pollen type, which has not yet been addressed. We therefore propose to test this hypothesis in future work. Clearly, the knowledge of the

flowering periods and the location allows great restrictions of the possible candidates and can simplify the classification. Even without a database on flowering periods, similar results could be achieved by taking into account the probability of co-occurrence of individual types.

8.5 Conclusion

Our study, using real-life pollen samples collected from pollen traps, revealed a strikingly wide variation in the colour dispersion among pollen pellets presumed to originate from the same botanical source. This variance is of such magnitude that to the human eye, the average colour dispersion appears equivalent to the range between distinctly dissimilar colours within each of the analysed cases. Given this substantial variability, our findings strongly caution against relying solely on colour as a reliable criterion for the classification of individual pollen pellets. In light of our findings, it is evident that there is a critical need for further research and exploration in the field of pollen classification. Technologically more complex, but also more accurate methods than colour determination, exist. For example combined multispectral imaging flow cytometry and deep learning, as well as molecular techniques, may enable more robust replication and higher sample throughput needed to compensate for the observed shortcomings. Ultimately, this would finally allow to assess pollen diversity in order to create a European pollen map, as suggested in previous research.

9.

Parts of this chapter are based on

AN INSIGHTFUL VIEW: IN-HIVE FLATBED SCANNERS FOR NON-INVASIVE LONG-TERM BEHAVIOR AND DISEASE MONITORING OF HONEY BEE COMBS [30]

Parzival Borlinghaus[1], Jörg Marvin Gülzow, Richard Odemer[2]

[1]Institute for Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated Plants, Institute for Bee Protection, Braunschweig, Germany

CRediT

Conceptualisation: PB, JMG, RO; Data curation: PB, MG; Formal analysis: PB; Funding acquisition: PB; Investigation: PB; Methodology: PB, JMG; Project administration: PB; Resources: PB, JMG; Software: PB, JMG; Supervision: PB; Validation: PB, JMG; Visualisation PB; Writing - original draft: PB; Writing - review & editing: JMG, RO.

Submitted to

Smart Agricultural Technology.

9.1 Introduction

The health and functioning of a honey bee colony are intricately tied to the condition of its combs. Each comb comprises thousands of cells, utilized for storing nectar and pollen or for rearing brood. However, these cells also provide habitats for various pathogens and pests, including the chalkbrood fungus (*Ascosphaera apis*) and *Varroa destructor* mites, which complicate colony management and health.

Chalkbrood disease, caused by *A. apis*, infects larvae through ingested spores, leading to fungal mycelia growth and the eventual death of the larvae. Infected larvae transform into hard, chalk-like 'mummies', posing a significant challenge to colony health [112]. The disease manifests when larvae are chilled around the capping time of the cells, a condition difficult to replicate accurately in research environments [192, 82]. Despite attempts to control chalkbrood by selecting colonies with efficient hygienic behaviour, current methodologies often fall short. They either fail to simulate natural conditions accurately or involve invasive techniques that disrupt colony health.

Similarly, Varroa mites represent a major threat to honey bee colonies. The mites' life cycle is closely tied to the brood, with their reproduction rates often varying significantly between laboratory and field conditions [191, 173, 71]. Effective monitoring of Varroa infestations has traditionally relied on labour-intensive and destructive methods, such as manually examining hundreds of cells and pupae [27, 141]. These methods can weaken the colony and provide skewed results due to the disturbance caused by cell removal.

Hygienic behaviour in *Apis mellifera* bees, including detecting, uncapping, and removing diseased or dead brood, is a crucial component of colony health management. Varroa-sensitive hygiene (VSH) specifically targets Varroa mites by interrupting their reproductive cycle, which can significantly reduce mite populations and improve colony resilience [229]. Despite the benefits of VSH behaviour, the ability to study and monitor these behaviours effectively remains limited [166, 241]. Existing methods either do not accurately replicate natural conditions or are invasive, undermining their utility for continuous and non-destructive observation [27, 248, 71].

To address these limitations, we propose an innovative method that integrates a thin flatbed scanner into a brood frame, allowing it to be placed inside a hive. This setup features a 3D-printed, wax-coated mesh foundation on the

scanner glass to support brood care. Connected to a Raspberry Pi computer, the scanner captures images of cells at regular intervals, enabling continuous observation of the bee brood lifecycle (eggs, larvae, pupae) and the presence of pathogens, nectar, and pollen.

This study aims to develop and explore a non-invasive method for monitoring Varroa mites and chalkbrood in honey bee colonies. We hypothesize that the integrated scanner method will:

1. Facilitate continuous, non-destructive monitoring of Varroa infestations and chalkbrood infections.
2. Provide more accurate and detailed data on Varroa and chalkbrood dynamics compared to traditional invasive techniques.
3. Enable observation of hygienic behaviours, such as VSH and general brood removal, without disrupting the colony.

This approach preserves the colony's integrity while offering detailed, continuous monitoring. By integrating technology with traditional beekeeping practices, our research addresses significant gaps in the field and provides a valuable tool for enhancing hive health management. This method represents a new perspective on monitoring bee health and pathogen dynamics, potentially leading to more effective strategies for managing honey bee colonies.

9.2 Materials and Methods

In the following, the hardware setup, foundation design, pilot study, image acquisition, and data management are described.

9.2.1 Hardware Setup

Scanners designed for A4 paper (210 × 297 mm) fit well into the standard frames of many beehives, requiring no additional protective measures and making installation straightforward. For this study, a CanoScan LiDE 210 consumer scanner was selected due to its wide availability, cost-effectiveness, and thin profile (39 mm, including the lid). It produces high-resolution images up to 2400 dpi, although higher resolutions slow scanning speed. For operation in the hive, the lid was removed, and the scanner was integrated into

a Dadant frame (see Figure 9.1a and 9.1b). A small cut-out in the brood-box facilitated cable management. The scanner is powered via USB, requiring only a single cable for power and data transmission.

The control unit, a Raspberry Pi 4 Model B, was housed in an empty honey super on top of the hive to protect it from moisture and from the bees. Since a mains connection was available, energy considerations did not influence the choice of the control unit. The Linux-based operating system enabled scheduled tasks to be executed automatically at specific times or intervals. These cron jobs effectively reduce the required software to a single scan image command provided by the `sane-utils` software package.

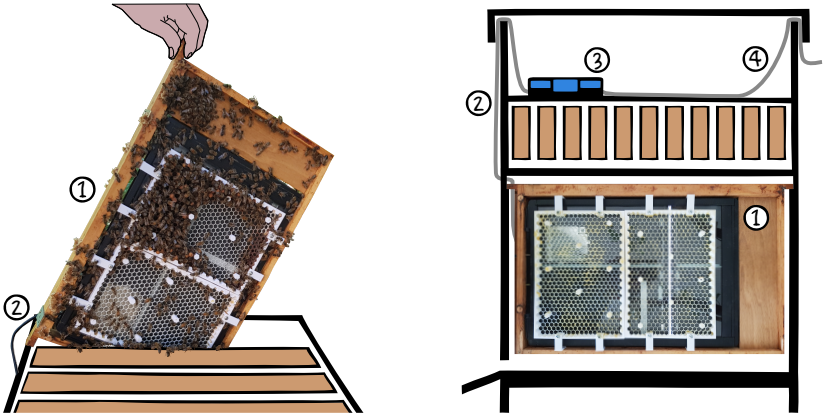


Figure 9.1: (a) The scanner (1) is inspected. Note the white plastic foundation, which was coated with a thin layer of wax before its initial insertion. (b) The control unit (3) is connected to a mains power supply (4) and placed in an empty honey super on top of the hive. A single USB cable (2) transfers both power and data between the controller and the scanning device.

9.2.2 Foundation Design

To facilitate the construction of transparent-floor cells on a scanning surface, we used a 3D-printed scaffold honeycomb structure that securely clipped onto the scanner. This scaffold provided a stable base and support for the bees, allowing them to build their cells in a controlled and observable environment. Figure 9.1(a) shows the brood frame with the built-in scanner and

the white plastic foundation. The choice of cell diameter influences whether worker or drone brood is laid. The mesh, designed with OpenSCAD software (https://openscad.org), had an inner cell diameter of 6.9 mm (drone brood) and was made of polylactic acid (PLA). The mesh height was 1.4 mm, which the bees continued to extend. Following standard practice [255], we coated the plastic mesh with a thin layer of liquid wax to enhance its acceptance by the bees.

9.2.3 Pilot Study

The study aimed to demonstrate the feasibility of using a flatbed scanner within a bee colony. Image data was collected at 30-minute intervals over three months, starting from March 13, 2024 to July 10, 2024. The study used a queen-right honey bee colony located in Karlsruhe, Germany, housed in a Dadant brood box.

Given the significant role of drone brood in the reproduction of *V. destructor* and the high probability of finding infested cells (97 %, see Odemer et al. [183]), we targeted this brood type using drone cell foundations on the scanner. In western honey bees, drone brood infestation is approximately ten times higher than that of worker brood [90, 26, 10]. This preference is attributed to several factors: drone development extends by two additional days, providing mites with more time to reproduce [26]; the pre-capping period during which drone brood attracts mites is two to three times longer than that for worker brood [27]; drone brood is more frequently visited by nurse bees, increasing the likelihood of mite transfer [47]; and drone larvae produce higher levels of kairomones, which attract mites [243].

Following common practice, the drone frame was first placed on the outer edge of the hive. After one week, the frame was repositioned closer to the centre of the brood nest to improve acceptance. However, the choice of brood type can vary depending on the study; in some experimental setups, a foundation with worker or mixed brood might be more suitable.

9.2.4 Image Acquisition and Annotation

Initially, images were manually triggered until automation began three days after the first egg was laid (May 8, 2024). A custom annotation tool was

developed to (1) extract individual cells (768 × 768 pixels) from the large source images (20 464 × 28 110 pixels) and (2) facilitate rapid iteration and labelling of these clipped images. Labels included egg laying, bee bread, adult Varroa (including position), Varroa offspring (including position), brood removal, hatched bee, and chalkbrood onset.

9.2.5 Data Management

At 2400 dpi, image acquisition takes around 10 minutes per scan, producing images of 30-40 MB each, amounting to approximately 1.6 GB of data per day. A WiFi connection was used to transfer data from the Raspberry Pi's SD card, preventing storage issues on the device. Given that images of this size are time-consuming to load, centre-cropped cells were automatically extracted and organized into folders, allowing for quick navigation and inspection. Efficient data management strategies are essential to handle the large volume of information generated during the continuous monitoring period, especially when multiple devices are used in parallel.

9.3 Results

The study's outcome is presented with a focus on the device's general applicability, possibilities for brood development, pathogen and parasite observation, and the monitoring of bee bread and stored pollen.

9.3.1 Hive Environment and Behavioural Observations

Unfavourable weather at the study's start hindered comb building. After moving the scanner closer to the brood nest, bees began drawing out the foundation, with the first eggs observed on May 5, 2024. The wax layer was observed to be repurposed for comb building, expanding over time but covering only about one-third of the scanner surface. Figure 9.2a shows the state of the drawn out portion of the comb at the end of the trial while Figure 9.2b shows the occupation and infestation of each cell spatially aligned.

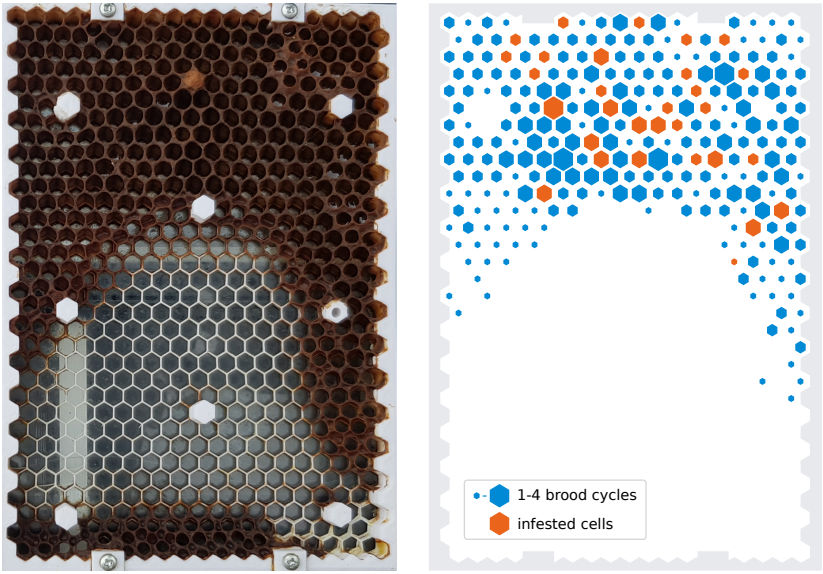


Figure 9.2: (a) The drawn out portion of the scanner at the end of the 3-month period. (b) The total number of brood (hatched and removed) per cell, with the size indicating the brood count ranging from 1 to 4. Cells where Varroa mites were detected at any point are highlighted in red. No cell experienced more than one invasion.

9.3.2 Image Data Collection and Annotation

During the study period, a total of 2 819 images were recorded and manually labelled for each of the 419 maintained cells. Larvae and pupae generally show minimal movement during most stages of their development, resulting in minimal changes between images. Consequently, significant events, such as the appearance of Varroa mites, the development of chalkbrood, or the removal of brood, are well detectable even in rapid iteration through all the 1 181 161 relevant cell images. Animations of particularly notable cells are available in the supplementary material of the published study.

Due to its size of over 100 GB, the raw image data will be made public only upon reasonable request. The annotations, however, can be found in JSON format in the supplementary material.

9.3.3 Brood Development

A total of 511 eggs were laid, with approximately 58.32 % of these being removed before hatching. Figure 9.3 illustrates the amount of brood maintained on the scanner surface and the factors contributing to its decline. Initially, a large number of cells became available simultaneously, indicating the queen's visits to the comb and resulting in a significant peak in egg-laying (Figure 9.3, top). Subsequently, only individually vacated cells were filled, leading to smaller peaks. As the season progressed, the queen reduced her laying activity and new brood was rarely observed.

Approximately 24 days after major egg-laying events, significant hatching events followed (Figure 9.3, middle), which later became more scattered. Figure 9.3 (bottom) shows both, strategic brood removals, with over 30 drones removed per day, and individual removals targeting specific cells.

When synchronizing brood development, as shown in Figure 9.4 (bottom), two periods of high brood removal become evident: one before and one after cell capping at day ten. Most of the brood was removed between day 5 and 7 (39 %) and between day 10 and 12 (35 %).

9.3.4 Pathogen and Parasite Observation

A total of 2 larva were infected by chalkbrood and 30 cells were invaded by Varroa mites. Figure 9.6a shows a cell bottom covered with the white mycelium of the fungus *A. apis*, which causes chalkbrood. The first signs of mycelium appeared on June 2, 2024 at midnight, roughly five hours before the image in Figure 9.6a was captured. Thirty minutes later, the bees began removing the infected brood.

Varroa mites were regularly first spotted swimming (and being trapped) in larval food. Figure 9.6b,c showcase such an incident, which is further discussed in Ifantidis [108], lasting until the larvae consumed enough food to release the parasite. Since mites can move within the cells, they (including their offspring) are not always present at the cell bottom. Figure 9.6d shows the foundress mite (red circle) and her offspring (white circle) passing by the image sensor.

Figure 9.5 shows the visibility patterns of adult Varroa mites and their offspring across all 30 infested brood cycles. Most adult mites were initially

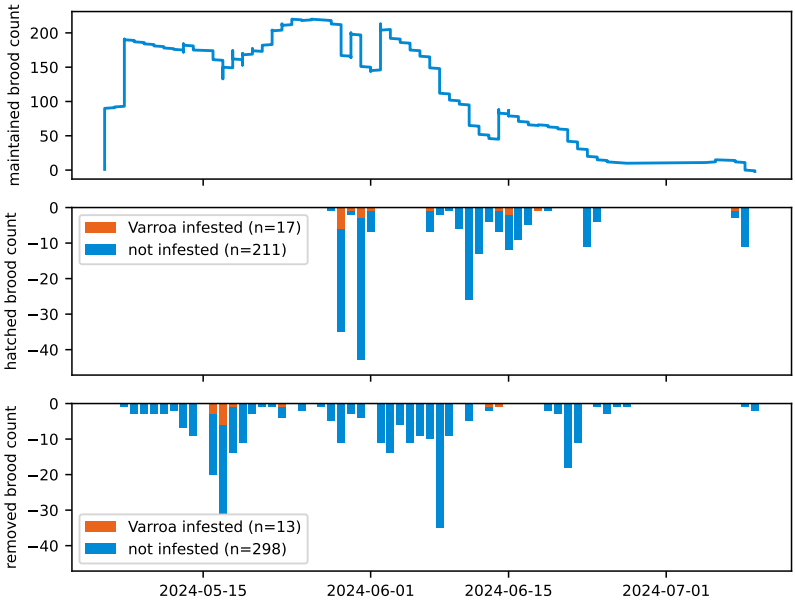


Figure 9.3: (Top) The number of (Varroa infested) brood that was maintained at the given point in time. The number increases when new eggs were detected and decreases due to hatched or removed brood. (Middle) The daily number of (Varroa infested) hatched brood. (Bottom) The daily number of (Varroa infested) brood removals.

observed before cell capping, often trapped in larval food (around day 10). Offspring were consistently first detected by day 16 at the latest.

In contrast to the trapped mites in Figure 9.6b,c, Figure 9.6d show adult and hatched mites that can move freely. Note the white dots beneath the mother mite, which are excrements.

9.3.5 Beebread and Stored Pollen

Throughout the entire trial, empty cells were used to store nectar, a well-known and desired practice to ensure that the food supply is kept close to the brood nest. However, it was surprising that nectar was rarely stored for more than a few hours before being emptied or refilled. Towards the end

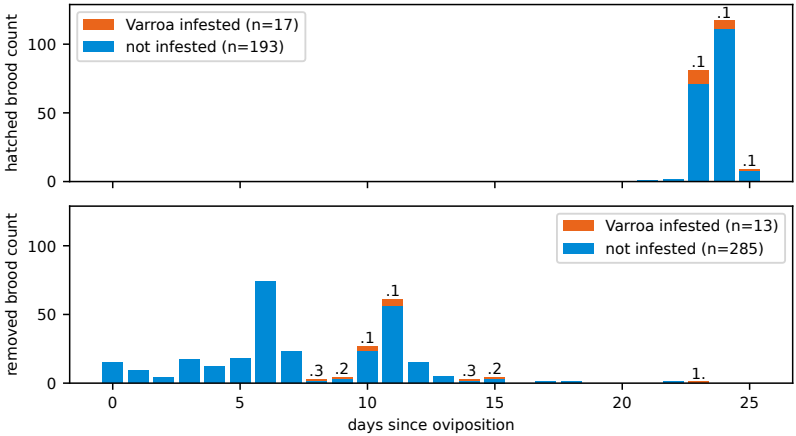


Figure 9.4: (Top) The number of Varroa-infested drones that hatched at a given age. (Bottom) The number of Varroa-infested brood that was removed at a given age. The dashed line marks the expected day of cell capping. Text annotations indicate the (rounded) fraction of infested brood.

of the drone season, drone cells became increasingly superfluous and were repurposed for minor storage of beebread. Figure 9.6f,g illustrates the storage of nectar and beebread.

9.4 Discussion

Current methodologies for brood cell observations in honey bee research are invasive and unsuitable for continuous monitoring, which disturbs the study subjects and limits the practicality of long-term research [173, 71]. This study aimed to address these limitations by developing a non-invasive, continuous monitoring technique using a modified flatbed scanner. Filling this gap is crucial for advancing our understanding of honey bee health, particularly in studying larvae, pupae, and Varroa mite infestations. By providing a method for real-time, detailed observations, this study aims to enhance field research accuracy and potentially transform beekeeping practices through better pest and pathogen management.

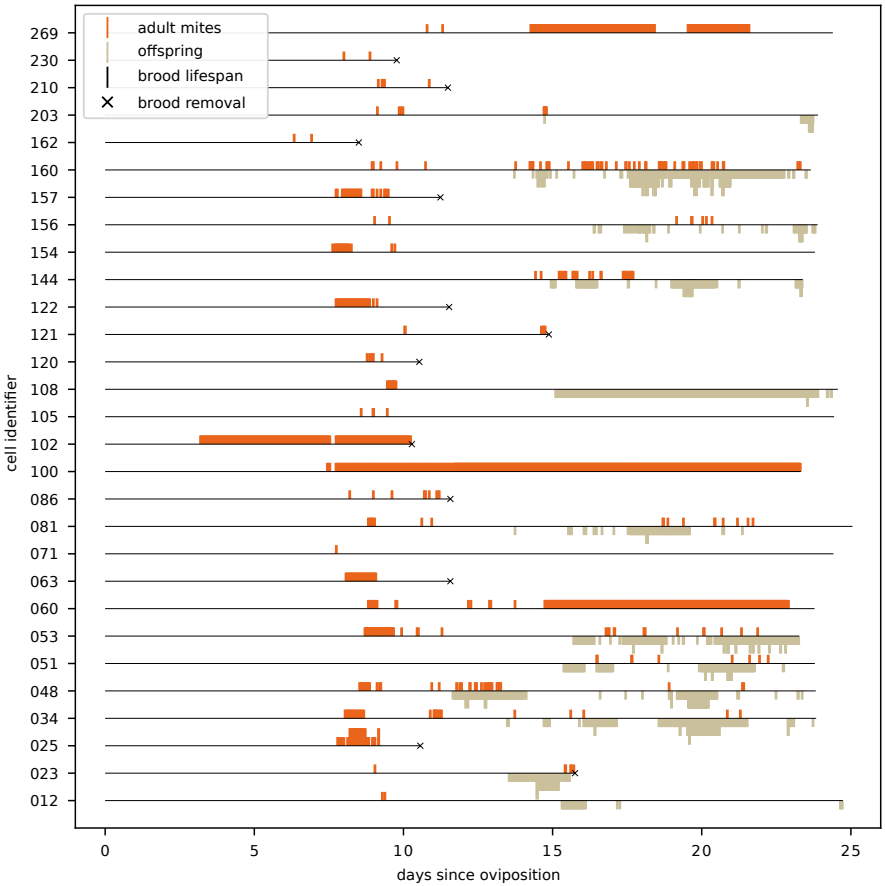


Figure 9.5: For all infested cells (y-axis), black lines depict the brood's lifespan, while grey and red bars indicate the periods when parasites were visible to the annotator. Multiple detections at the same time are stacked; for instance, in cell 160, up to three offspring were observed simultaneously, and in cell 025, two adult mites were visible at the same time. Brood removals are marked with ×-symbols.

A key result of this study was the successful detection and monitoring of Varroa mites and *A. apis* in vivo, which enabled the achievement of unprecedented sample intervals of minutes without additional costs. These

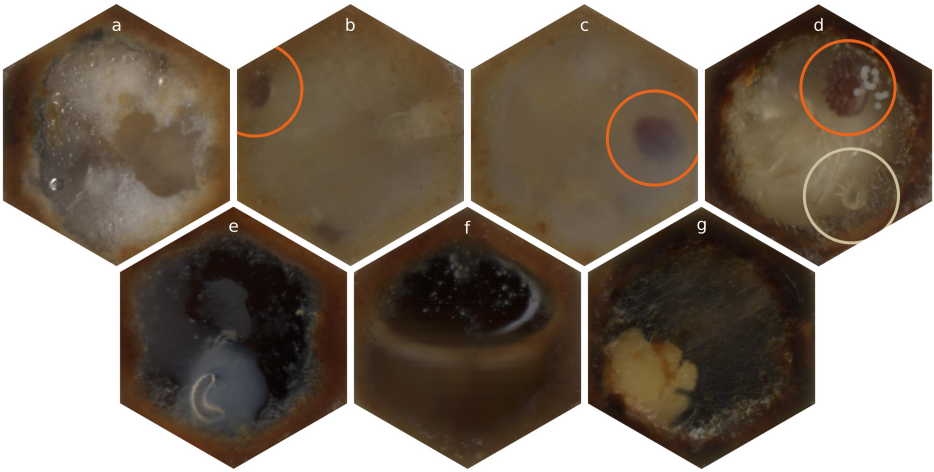


Figure 9.6: (a) Chalk brood fungus evolves in a drone brood cell. Image was taken hours before the brood was removed and the cell was cleaned. (b, c) Adult mites (red circle) trapped in larval food. For an in-depth explanation of this pattern refer to [108]. (d) Foundress with offspring (white circle). White dots indicate the defecation of the mother mite, which is typically piled. (e) A recently hatched larva. (f) Nectar and (g) pollen stored in drone cells.

observations align with literature indicating that mites often hide at the cell bottom and exhibit movement patterns related to their developmental stages and the host larva [73].

To assess whether Varroa mites remain hidden from the sensor for extended periods, it is essential to analyse their visibility patterns. We observed frequent Varroa detections in most infested cells, enabling a robust analysis, particularly in cells where foundresses successfully reproduced. Rare detections of Varroa mites were mainly observed in cells that were cleaned out early, leaving little opportunity for detection.

The first observation of offspring frequently coincided with the host larva's moult on days 14/15, which may have prevented earlier visits to the cell bottom. Continuous visibility of the offspring aligned with their immobile developmental stages, during which some mites remained at the cell bottom. Extended periods of adult mite visibility were associated with atypical behaviour. For example, the foundress in cell 60 fell into a state of agony and ceased activity; the mite in cell 100 died in the larval food, never being freed

or removed; and the mite in cell 102 entered unusually early, remaining in the food for an extended period before being removed along with the brood. The mite in cell 269 was alive but inactive for no apparent reason.

We conclude that it is highly unlikely for our method to miss infested cells with successful mite reproduction. While the likelihood of missing infestations increases when the observation period is shortened due to early brood removal, we estimate this risk to be minimal, as adult mites are often trapped in the larval food early on, making them clearly visible to the scanner.

Our method allowed for the simultaneous observation of up to 2 500 brood cells, thereby significantly increasing the sample size in comparison to traditional methods [27, 84, 182]. The high-resolution images captured were of exceptional clarity, displaying even the smallest mite offspring and early-stage fungal infections. This aligns well with existing literature which emphasises the necessity for detailed, real-time data in order to gain an accurate understanding of pest and pathogen dynamics within honey bee hives [141].

This continuous monitoring approach is in accordance with previous research that has advocated for innovative techniques to improve infestation measurements and pathogen detection accuracy [99]. The advancement is particularly significant in light of the challenges highlighted by Lefebre et al. [141], where achieving sufficient single infested cells for reliable mite-non reproduction, reduced mite reproduction, and decreased mite reproduction calculations proved difficult. This underscores the need for more efficient monitoring methods. Furthermore, our method corroborates the findings of McGruddy et al. [158], who demonstrated the efficacy of RNA interference (RNAi) in reducing mite reproductive success. Our approach provides a complementary method for observing and validating these biological effects in real time.

Furthermore, Siefert et al. [224] applied video recording techniques to examine honey bee brood care behaviour in the context of chronic neonicotinoid exposure, thereby identifying notable discrepancies in nursing behaviour and larval development. Our scanner-based technique refines this approach by offering non-invasive, high-frequency monitoring, thereby expanding the scope of our ability to study these critical behaviours and their environmental stressors in greater detail in full sized colonies.

The findings of our study offer valuable insights into the natural behaviour of honey bees and their interactions with pests and pathogens. For example, the humid season facilitated the development of *A. apis*, which resulted in

the formation of chalkbrood. Our continuous monitoring system was able to detect this early, providing an advantage over traditional weekly inspections. This finding is consistent with the literature, which indicates that less severe chalkbrood infections are often overlooked by beekeepers due to the rapid removal of mummified brood by worker bees [103].

Our scanner-based technique offers advantages over previous methods, such as macro video recording setups that required special lighting and limited the number of observable cells [224]. Unlike visual inspection techniques that focus on mites visible on the exterior of bees [18, 23, 221, 180], our method detects mites within the brood cells. Therefore, it is not affected by mites hidden beneath the bee's sclerites [194], thus providing a more comprehensive assessment. [110].

Moreover, the device's capacity to monitor brood cell removal events revealed a notable occurrence of mass brood removal, which was likely attributable to external factors such as disease, food scarcity, or space constraints affecting all drone brood in a similar manner. This observation is of great importance for breeding programs that aim to promote VSH traits, as it allows for the differentiation between mass removal events and selective hygienic behaviour directed towards infested cells [185]. For example, a brood removal on June 14, 2024 stands out as it was the only removal on that day, targeting a single Varroa-infested cell among an average of 69 brood cells. This makes it highly unlikely to have occurred by chance (1.44 %; see Figure 9.3 bottom).

Within a cell, two spots are crucial for the mite's reproduction: the feeding spot where the foundress bites the larvae and her offspring nourish themselves, and the defecation spot where the male and female offspring mate. The bite can be made visible using chemicals but cannot be seen with the naked eye, making it undetectable in the acquired image data. The mother mite defecates at the same spot, forming a pile of white excrement. Infestation can be quickly verified by searching for such excrement on the walls when infested cells are opened. Unexpectedly, one foundress preferred the floor over the cell's walls for defecation (see Figure 9.6). Since all members of the mite family visit this spot frequently, this observation is particularly interesting, and the corresponding time-lapse video is included in the supplementary material.

Our data showed that worker bees selectively removed Varroa-infested cells, thereby corroborating previous findings on VSH behaviour. This result is consistent with the observations of Sprau, Traynor, and Rosenkranz [229],

who demonstrated the complexity of the cues utilized by honey bees to detect and remove mites from brood cells, including movement and odour. Their findings highlight the importance of comprehensive and ongoing observation to elucidate the subtle mechanisms underlying VSH behaviours, supporting the value of our non-invasive imaging methodology. Furthermore, these insights align with the recent work of Morin and Giovenazzo [168], who identified a range of traits, including mite-non reproduction, recapping activity, and hygienic behaviour, as key predictors of mite infestation levels. The intricate and dynamic nature of these behaviours underscores the need for non-invasive, comprehensive, and continuous monitoring approaches to accurately assess and enhance these traits in breeding programs.

Despite its promising findings, this study has limitations. The sample size and scope were constrained by the available resources, which may affect the generalizability of the results. We observed that bees took approximately three weeks to draw out cells and lay eggs, a delay influenced by factors such as weather conditions and materials used. Visual inspection at the opened hive during scanning showed no signs of disturbance whatsoever. Also, no signs of agitation were heard outside the hive at night when the sensor was triggered and the apiary was in complete silence. This suggests that the light and sound emitted by scanning in the hive environment does not cause disruptions. While the scanner did not cause immediate disturbance, potential long-term effects on colony and pathogens require further investigation to ensure the method's suitability for diverse environments and larger-scale applications. The narrow depth of field and the temporal resolution of the scanner, although suitable for our purposes, might limit its applicability in other research contexts. Moreover, the substantial data size generated by high-resolution images necessitates efficient data management strategies, which could be challenging in larger-scale implementations.

Cleaning of the glass during the experiment was neither possible nor required. However, we did observe the accumulation of dirt, especially in unused cells. Cells that are regularly used for brood or nectar storage are effectively cleaned by worker bees and by the larva's food consumption, both enhancing the transparency. Almost no cell degraded in a way that made observation impossible over a period of three months. In the supplementary material, we included a time-lapse video of the cell that contained the highest number of eggs, illustrating the accumulation of debris and the bees' cleaning behaviour.

Future studies should aim to scale this monitoring technique to larger bee populations and diverse environmental conditions, to validate its effectiveness across various beekeeping scenarios. A critical aspect of this research involves examining the long-term impacts of continuous scanning on colony health and behaviour, including factors such as scanner light, electronics, and noise. Moreover, exploring how this technology can be applied in breeding programs to select for VSH traits could lead to notable advancements in managing Varroa mite infestations.

Moreover, exploring how this technology can be applied in breeding programs to select for VSH traits could lead to notable advancements in managing Varroa mite infestations. As successfully done elsewhere in apidology [35, 31, 34, 36], we plan to use the massive amount of collected and annotated data to automate the image data analysis utilizing artificial neural networks. Developing a suitable computer vision model would make brood cell scanning more accessible and practical for widespread application. Large-scale studies building on this proof of concept could potentially lead to the replacement of all brood combs and open new avenues for addressing exciting research questions in the future.

9.5 Conclusion

In conclusion, this study presents a novel, non-invasive method for continuous monitoring of honey bee brood cells, addressing a significant gap in current research methodologies. By enabling detailed, real-time observations of Varroa mites and *A. apis*, this scanner-based technique has the potential to revolutionize the study of honey bee health and pest management. The major contribution of this study lies in its ability to enhance the accuracy and practicality of field research, paving the way for improved beekeeping practices and more resilient honey bee populations. This advancement aligns with the overarching goal of mitigating the impacts of pests and pathogens on honey bee colonies, ultimately supporting the sustainability of apiculture and agricultural ecosystems.

10 Conclusion

> *'But let younger investigators be warned (...), as they strive impatiently to publish their results after long years of frustration. Let them test their findings doubly and trebly before they regard any interpretation as certain. For often nature reaches her goal by another path (...)'*
>
> *Karl von Frisch [88], 1971.*

This thesis has demonstrated various ways in which machine learning in computer vision can be utilised to explore a wide array of topics in apidology. The groundwork for this was established in Chapter 4, where a device was developed to capture high-quality footage of bees within their natural environment. The resulting dataset is substantial and offers significant potential for future research endeavours.

Innovation continues to flourish in addressing a century-old challenge: the accurate determination of bee mortality in the field, which is crucial for ecotoxicological studies and species conservation. Despite the ongoing exploration of new technologies – most recently, camera-based systems – the quality of bee counters has stagnated. This work identified the primary issue as the difficulty in evaluating these counters, leading to the development of a standardised assessment protocol, detailed in Chapter 5, and tested using a commercial camera-based system. Now that further progress has become measurable, there is renewed hope for genuine innovation and advancement

In Chapter 6, representation learning was applied to the visual features of bumblebees, processed through a ResNet-18 backbone, enabling the re-

identification of individual bumblebees. This deep learning approach demonstrated clear advantages over hand-crafted features that were combined with metric learning.

In Chapter 7, an app was developed that, for the first time, can calibrate the colour of photographed pollen samples using everyday objects with known colours. This user-friendly approach allows beekeepers to easily assess the quality of their apiaries based on colour-derived biodiversity indices. On the other hand, scientists can use the app to conduct large-scale citizen science studies, combining biodiversity data with metadata to create a comprehensive monitoring system.

In a further step, the thesis investigated the extent to which the floral origin of pollen could be inferred from the calibrated pollen colours determined by the aforementioned app. It was revealed that, in addition to the known challenges – such as pollen colour similarity between different plant species – there is also significant dissimilarity within the same species. This novel insight was made possible through the first large-scale combination of morphological laboratory analysis with calibrated pollen colours, facilitated by adapted Gaussian Mixture Models.

In the final contribution, we tested the use of contact image sensors inside the hive to monitor bee brood. During the three-month pilot study, it became evident that both chalkbrood and Varroa infestations could be reliably detected. The innovative use of this imaging technique opens up groundbreaking possibilities in apidology. For example, breeders of Varroa-resistant honey bee traits could benefit from simple, cost-effective, and regular monitoring of infestation levels, allowing them to quantify breeding success. This offers hope that future bee populations may become self-sufficient in defending against their greatest threat.

In conclusion, this thesis has confirmed what other authors have already predicted: ‘deep learning and computer vision will transform entomology’ [106]. For my part, I am firmly convinced that in the future we will see more and more digital innovations around the beehive. Building on this work, a forthcoming research project is already in the planning to further develop the brood monitoring device into an off-grid capable and fully automated computer vision system.

Bibliography

[1] Charu C. Aggarwal, Alexander Hinneburg, and Daniel A. Keim. 'On the Surprising Behavior of Distance Metrics in High Dimensional Space'. In: *Database Theory — ICDT 2001*. Ed. by Jan Van Den Bussche and Victor Vianu. Red. by Gerhard Goos, Juris Hartmanis, and Jan Van Leeuwen. Vol. 1973. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 420–434. ISBN: 978-3-540-41456-8 978-3-540-44503-6. DOI: 10.1007/3-540-44503-X_27.

[2] Tom Arbuckle, Stefan Schröder, Volker Steinhage, and Dieter Wittmann. 'Biodiversity Informatics in Action: Identification and Monitoring of Bee Species using ABIS'. In: *Proc. 15th Int. Symp. Informatics for Environmental Protection*. 2002.

[3] Ana M. Ares, Laura Toribio, Jesús A. Tapia, Amelia V. González-Porto, Mariano Higes, Raquel Martín-Hernández, and José Bernal. 'Differentiation of bee pollen samples according to the apiary of origin and harvesting period based on their amino acid content'. In: *Food Bioscience* 50 (2022), p. 102092. ISSN: 2212-4292. DOI: 10.1016/j.fbio.2022.102092.

[4] K. A. Aronstein and K. D. Murray. 'Chalkbrood disease in honey bees'. In: *Journal of Invertebrate Pathology* 103 (2010), S20–S29. ISSN: 0022-2011. DOI: 10.1016/j.jip.2009.06.018.

[5] Yuta Asano, Mark D. Fairchild, Laurent Blondé, and Patrick Morvan. 'Color matching experiment for highlighting interobserver variability'. In: *Color Research & Application* 41.5 (2016), pp. 530–539. DOI: 10.1002/col.21975.

[6] Z. Babic, R. Pilipovic, V. Risojevic, and G. Mirjanic. 'Pollen Bearing Honey Bee Detection in Hive Entrance Video Recorded by Remote Embedded System for Pollination Monitoring'. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* III-7 (2016), pp. 51–57. ISSN: 2194-9050. DOI: 10.5194/isprs-annals-III-7-51-2016.

[7] Roy J. Barker. 'Is There Any Evidence that Honeybees are Attracted by Infrared?' In: *Bee World* 53.2 (1972), pp. 66–68. ISSN: 0005-772X, 2376-7618. DOI: 10.1080/0005772X.1972.11097407.

[8] Pinky A. Bautista, Noriaki Hashimoto, and Yukako Yagi. 'Color standardization in whole slide imaging using a color calibration slide'. In: *Journal of Pathology Informatics* 5 (2014), p. 4. ISSN: 2229-5089. DOI: 10.4103/2153-3539.126153.

[9] Matthias A. Becher, Grace Twiston-Davies, Tim D. Penny, Dave Goulson, Ellen L. Rotheray, and Juliet L. Osborne. 'Bumble-BEEHAVE: A systems model for exploring multifactorial causes of bumblebee decline at individual, colony, population and community level'. In: *Journal of Applied Ecology* 55.6 (2018), pp. 2790–2801. ISSN: 1365-2664. DOI: 10.1111/1365-2664.13165.

[10] Joop Beetsma, Willem Jan Boot, and Johan Calis. 'Invasion behaviour of Varroa jacobsoni Oud.: from bees into brood cells'. In: *Apidologie* 30.2 (1999), pp. 125–140. ISSN: 0044-8435. DOI: 10.1051/apido:19990204.

[11] Aurelien Bellet, Amaury Habrard, and Marc Sebban. *Metric Learning*. Morgan & Claypool Publishers, 2015. 153 pp. ISBN: 978-1-62705-366-2.

[12] Aurélien Bellet, Amaury Habrard, and Marc Sebban. *A Survey on Metric Learning for Feature Vectors and Structured Data*. 2014. DOI: 10.48550/arXiv.1306.6709.

[13] R. Bellman. *Dynamic Programming*. Rand Corporation research study. Princeton University Press, 1957. ISBN: 978-0-691-07951-6.

[14] Martin Bencsik, Yves Le Conte, Maritza Reyes, Maryline Pioz, David Whittaker, Didier Crauser, Noa Simon Delso, and Michael I. Newton. 'Honeybee Colony Vibrational Measurements to Highlight the Brood Cycle'. In: *PLOS ONE* 10.11 (2015). Publisher: Public Library of Science, e0141926. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0141926.

[15] Sven Bermig, Richard Odemer, Alina J. Gombert, Malte Frommberger, Ralf Rosenquist, and Jens Pistorius. 'Experimental validation of an electronic counting device to determine flight activity of honey bees (Apis mellifera L.)' In: *Journal of Cultivated Plants* (2020), pp. 132–140. DOI: 10.5073/JFK.2020.05.03.

[16] Keni Bernardin, Alexander Elbs, and Rainer Stiefelhagen. 'Multiple Object Tracking Performance Metrics and Evaluation in a Smart Room Environment'. In: *Sixth IEEE International Workshop on Visual Surveillance, in conjunction with ECCV*. Vol. 90. Citeseer, 2006.

[17] J. C. Biesmeijer et al. 'Parallel Declines in Pollinators and Insect-Pollinated Plants in Britain and the Netherlands'. In: *Science* 313.5785 (2006), pp. 351–354. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1127863.

[18] Simon Bilik, Lukas Kratochvila, Adam Ligocki, Ondrej Bostik, Tomas Zemcik, Matous Hybl, Karel Horak, and Ludek Zalud. 'Visual Diagnosis of the Varroa Destructor Parasitic Mite in Honeybees Using Object Detector Techniques'. In: *Sensors* 21.8 (2021), p. 2764. ISSN: 1424-8220. DOI: 10.3390/s21082764.

[19] Simon Bilik, Tomas Zemcik, Lukas Kratochvila, Dominik Ricanek, Miloslav Richter, Sebastian Zambanini, and Karel Horak. 'Machine learning and computer vision techniques in continuous beehive monitoring applications: A survey'. In: *Computers and Electronics in Agriculture* 217 (2024), p. 108560. ISSN: 0168-1699. DOI: 10.1016/j.compag.2023.108560.

[20] Aycan Bilisk Tosunoglu, Ibrahim Cakmak, Gulsah Saatcioglu, Adem Bicakci, and Hulusi Malyer. 'Spectrum of Pollen Collected by Honeybees in Bursa Lowland Area in High Season'. In: *Bee Science* 8.4 (2008), pp. 143–148.

[21] Intergovernmental Science-Policy Platform on Biodiversity {and} Ecosystem Services (IPBES). 'Assessment Report on Pollinators, Pollination and Food Production'. In: *Zenodo* (2016). DOI: 10.5281/zenodo.3402857.

[22] Miriam Bixby, Kathy Baylis, Shelley E. Hoover, Rob W. Currie, Andony P. Melathopoulos, Stephen F. Pernal, Leonard J. Foster, and M. Marta Guarna. 'A Bio-Economic Case Study of Canadian Honey Bee (Hymenoptera: Apidae) Colonies: Marker-Assisted Selection (MAS) in Queen Breeding Affects Beekeeper Profits'. In: *Journal of Economic Entomology* 110.3 (2017), pp. 816–825. ISSN: 0022-0493, 1938-291X. DOI: 10.1093/jee/tox077.

[23] Kim Bjerge, Carsten Eie Frigaard, Peter Høgh Mikkelsen, Thomas Holm Nielsen, Michael Misbih, and Per Kryger. 'A computer vision system to monitor the infestation level of Varroa destructor in a honeybee colony'. In: *Computers and Electronics in Agriculture* 164 (2019), p. 104898. ISSN: 01681699. DOI: 10.1016/j.compag.2019.104898.

[24] Roman Bleha, Tetiana V. Shevtsova, Martina Živčáková, Anna Korbářová, Martina Ježková, Ivan Saloň, Ján Brindza, and Andriy Synytsya. 'Spectroscopic Discrimination of Bee Pollen by Composition, Color, and Botanical Origin'. In: *Foods* 10.8 (2021). Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, p. 1682. ISSN: 2304-8158. DOI: 10.3390/foods10081682.

[25] Humberto Boncristiani, James D. Ellis, Tomas Bustamante, Jason Graham, Cameron Jack, Chase B. Kimmel, Ashley Mortensen, and Daniel R. Schmehl. 'World Honey Bee Health: The Global Distribution of Western Honey Bee (*Apis mellifera* L.) Pests and Pathogens'. In: *Bee World* 98.1 (2021), pp. 2–6. ISSN: 0005-772X, 2376-7618. DOI: 10.1080/0005772X.2020.1800330.

[26] W. J. Boot, J. Schoenmaker, J. N. M. Calis, and J. Beetsma. 'Invasion of Varroa jacobsoni into drone brood cells of the honey bee, Apis mellifera'. In: *Apidologie* 26.2 (1995), pp. 109–118. ISSN: 0044-8435. DOI: 10.1051/apido:19950204.

[27] Willem J. Boot, Johan N. M. Calis, and Joop Beetsma. 'Differential periods of Varroa mite invasion into worker and drone cells of honey bees'. In: *Experimental & Applied Acarology* 16.4 (1992), pp. 295–301. ISSN: 1572-9702. DOI: 10.1007/BF01218571.

[28] Parzival Borlinghaus. *High Quality Honey Bee Dataset*. Publisher: figshare. 2024. DOI: 10.6084/m9.figshare.26999512.v1.

[29] Parzival Borlinghaus. *Pollen detection dataset*. Publisher: figshare. 2022. DOI: 10.6084/m9.figshare.21679949.v1.

[30] Parzival Borlinghaus, Jörg Marvin Gülzow, and Richard Odemer. 'An Insightful View: In-Hive Flatbed Scanners for Non-invasive Long-Term Behaviour and Disease Monitoring of Honey Bee Brood Combs'. Paper submitted for publication. 2024.

[31] Parzival Borlinghaus, Jakob Jung, and Richard Odemer. 'Introducing Pollenyzer: An App for Automatic Determination of Colour Diversity for Corbicular Pollen Loads'. In: *Smart Agricultural Technology* 5 (2023), p. 100263. ISSN: 2772-3755. DOI: 10.1016/j.atech.2023.100263.

[32] Parzival Borlinghaus and Richard Odemer. *Pollenyzer dataset - honey bee pollen samples with scans, photos and laboratory analysis*. Publisher: figshare. 2023. DOI: 10.6084/m9.figshare.23065052.

[33] Parzival Borlinghaus, Richard Odemer, and Frederic Tausch. *Color dispersion of corbicular pollen loads - Supplementary Material*. Publisher: figshare. 2024. DOI: 10.6084/m9.figshare.25434733.

[34] Parzival Borlinghaus, Richard Odemer, and Frederic Tausch. 'Natural color dispersion of corbicular pollen limits color-based classification'. In: *Open Journal of Photogrammetry and Remote Sensing* (2024).

[35] Parzival Borlinghaus, Richard Odemer, Frederic Tausch, Katharina Schmidt, and Oliver Grothe. 'Honey bee counter evaluation – Introducing a novel protocol for measuring daily loss accuracy'. In: *Computers and Electronics in Agriculture* 197 (2022), p. 106957. ISSN: 01681699. DOI: 10.1016/j.compag.2022.106957.

[36] Parzival Borlinghaus, Frederic Tausch, and Luca Rettenberger. 'A Purely Visual Re-ID Approach for Bumblebees (Bombus terrestris)'. In: *Smart Agricultural Technology* 3 (2023), p. 100135. ISSN: 2772-3755. DOI: 10.1016/j.atech.2022.100135.

[37] Katarzyna Bozek, Laetitia Hebert, Alexander S. Mikheyev, and Greg J. Stephens. 'Towards Dense Object Tracking in a 2D Honeybee Hive'. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT: IEEE, 2018, pp. 4185–4193. ISBN: 978-1-5386-6420-9. DOI: 10.1109/CVPR.2018.00440.

[38] Manuela R. Branco, Neil A. C. Kidd, and Robert S. Pickard. 'A comparative evaluation of sampling methods for Varroa destructor (Acari: Varroidae) population estimation'. In: *Apidologie* 37.4 (2006), pp. 452–461. ISSN: 0044-8435, 1297-9678. DOI: 10.1051/apido:2006010.

[39] Robert Brodschneider. *C.S.I. Pollen: Anleitung zur Untersuchung der für Honigbienen erhältlichen Pollendiversität.* Accessed October 8, 2024. 2014. URL: bienenstand.at/wp-content/uploads/2014/08/CSIPollen%C3%96.pdf.

[40] Robert Brodschneider and Karl Crailsheim. 'Nutrition and health in honey bees'. In: *Apidologie* 41.3 (2010), pp. 278–294. DOI: 10.1051/apido/2010012.

[41] Robert Brodschneider et al. 'CSI Pollen: Diversity of Honey Bee Collected Pollen Studied by Citizen Scientists'. In: *Insects* 12.11 (2021), p. 987. ISSN: 2075-4450. DOI: 10.3390/insects12110987.

[42] Jerry Bromenshenk. 'Honey Bee Acoustic Recording and Analysis System for Monitoring Hive Health'. Pat.

[43] Jerry Bromenshenk et al. 'Can Honey Bees Assist in Area Reduction and Landmine Detection?' In: *The Journal of Conventional Weapons Destruction* 7.3 (2003). ISSN: 1533-9440.

[44] Jerry J. Bromenshenk, Colin B. Henderson, Robert A. Seccomb, Phillip M. Welch, Scott E. Debnam, and David R. Firth. 'Bees as Biosensors: Chemosensory Ability, Honey Bee Monitoring Systems, and Emergent Sensor Technologies Derived from the Pollinator Syndrome'. In: *Biosensors* 5.4 (2015), pp. 678–711. ISSN: 2079-6374. DOI: 10.3390/bios5040678.

[45] C. Brust, T. Burghardt, M. Groenenberg, Christoph Käding, H. Kühl, Marie L. Manguette, and Joachim Denzler. 'Towards Automated Visual Monitoring of Individual Gorillas in the Wild'. In: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (2017). DOI: 10.1109/ICCVW.2017.333.

[46] Nikolajs Bumanis, Olvija Komasilova, Vitalijs Komasilovs, Armands Kviesis, and Aleksejs Zacepins. 'Application of Data Layering in Precision Beekeeping: The Concept'. In: *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE, 2020, pp. 1–6.

[47] Nicholas W. Calderone and L. P. S. Kuenen. 'Differential tending of worker and drone larvae of the honey bee, Apis mellifera, during the 60 hours prior to cell capping'. In: *Apidologie* 34.6 (2003), pp. 543–552. ISSN: 0044-8435, 1297-9678. DOI: 10.1051/apido:2003054.

[48] Jason Campbell, Lily Mummert, and Rahul Sukthankar. 'Video Monitoring of Honey Bee Colonies at the Hive Entrance'. In: *Visual observation & analysis of animal & insect behavior, ICPR* 8 (2008), pp. 1–4.

[49] Jennifer M Campbell, Douglas C Dahn, and Daniel A J Ryan. 'Capacitance-based sensor for monitoring bees passing through a tunnel'. In: *Measurement Science and Technology* 16.12 (2005), pp. 2503–2510. ISSN: 0957-0233, 1361-6501. DOI: 10.1088/0957-0233/16/12/015.

[50] Maria G. Campos, Stefan Bogdanov, Ligia Bicudo De Almeida-Muradian, Teresa Szczesna, Yanina Mancebo, Christian Frigerio, and Francisco Ferreira. 'Pollen composition and standardisation of analytical methods'. In: *Journal of Apicultural Research* 47.2 (2008), pp. 154–161. ISSN: 0021-8839, 2078-6913. DOI: 10.1080/00218839.2008.11101443.

[51] Maria G. Campos et al. 'Standard methods for pollen research'. In: *Journal of Apicultural Research* 60.4 (2021), pp. 1–109. ISSN: 0021-8839, 2078-6913. DOI: 10.1080/00218839.2021.1948240.

[52] Luis Carlos Casas Restrepo, Ilver Eduardo Gutierrez Alabat, Guillermo Salamanca Grosso, and Francisco de Assis Ribeiro dos Santos. 'Markers for the spatial and temporal differentiation of bee pollen harvested by Apis mellifera L. in the Eastern Andes of Colombia'. In: *Journal of Apicultural Research* 62.3 (2023), pp. 556–569. ISSN: 0021-8839. DOI: 10.1080/00218839.2021.1916188.

[53] Jeffrey Chan, Hector Carrión, Rémi Mégret, José Rivera, and Tugrul Giray. 'Honeybee Re-identification in Video: New Datasets and Impact of Self-supervision:' in: *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. 17th International Conference on Computer Vision Theory and Applications. SCITEPRESS - Science and Technology Publications, 2022, pp. 517–525. ISBN: 978-989-758-555-5. DOI: 10.5220/0010843100003124.

[54] Solene Chan-Lang, Quoc-Cuong Pham, and Catherine Achard. 'Closed and Open-World Person Re-Identification and Verification'. In: *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA). 2017, pp. 1–8. DOI: 10.1109/DICTA.2017.8227416.

[55] Jean-Daniel Charrière, Anton Imdorf, Boris Bachofen, and Anna Tschan. 'The removal of capped drone brood: an effective means of reducing the infestation of varroa in honey bee colonies'. In: *Bee World* 84.3 (2003), pp. 117–124. ISSN: 0005-772X, 2376-7618. DOI: 10.1080/0005772X.2003.11099587.

[56] Wei-Sheng Chen, Chien-Hao Wang, Joe-Air Jiang, and En-Cheng Yang. 'Development of a monitoring system for honeybee activities'. In: *2015 9th International Conference on Sensing Technology (ICST)*. 2015 9th International Conference on Sensing Technology (ICST). Auckland, New Zealand: IEEE, 2015, pp. 745–750. ISBN: 978-1-4799-6314-0. DOI: 10.1109/ICSensT.2015.7438495.

[57] Manuel Chica and Pascual Campoy. 'Discernment of bee pollen loads using computer vision and one-class classification techniques'. In: *Journal of Food Engineering* 112.1 (2012), pp. 50–59. ISSN: 02608774. DOI: 10.1016/j.jfoodeng.2012.03.028.

[58] Guillaume Chiron, Petra Gomez-Krämer, and Michel Ménard. 'Detecting and tracking honeybees in 3D at the beehive entrance using stereo vision'. In: *EURASIP Journal on Image and Video Processing* 2013.1 (2013), p. 59. ISSN: 1687-5281. DOI: 10.1186/1687-5281-2013-59.

[59] Guillaume Chiron, Petra Gomez-Krämer, and Michel Ménard. 'Outdoor 3D Acquisition System for Small and Fast Targets: Application to Honeybee Monitoring at the Beehive Entrance:' in: *Proceedings of GEODIFF 2013*. GEODIFF Workshop. Barcelona, Spain: SciTePress - Science, 2013, pp. 10–19. ISBN: 978-989-8565-49-5. DOI: 10.5220/0004346300100019.

[60] John A. Chmiel, Brendan A. Daisley, Andrew P. Pitek, Graham J. Thompson, and Gregor Reid. 'Understanding the Effects of Sublethal Pesticide Exposure on Honey Bees: A Role for Probiotics as Mediators of Environmental Stress'. In: *Frontiers in Ecology and Evolution* 8 (2020), p. 22. ISSN: 2296-701X. DOI: 10.3389/fevo.2020.00022.

[61] Théotime Colin, Ryan J. Warren, Stephen R. Quarrell, Geoff R. Allen, and Andrew B. Barron. 'Evaluating the foraging performance of individual honey bees in different environments with automated field RFID systems'. In: *Ecosphere* 13.5 (2022), e4088. ISSN: 2150-8925. DOI: 10.1002/ecs2.4088.

[62] Yves Le Conte, Marion Ellis, and Wolfgang Ritter. 'Varroa mites and honey bee health: can Varroa explain part of the colony losses?' In: *Apidologie* 41.3 (2010), pp. 353–363. ISSN: 0044-8435, 1297-9678. DOI: 10.1051/apido/2010017.

[63] Ida Conti, Piotr Medrzycki, Francesca V. Grillenzoni, Francesca Corvucci, Simone Tosi, Valeria Malagnini, Martina Spinella, and Mauro G. Mariotti. 'Floral Diversity of Pollen Collected by Honey Bees (Apis mellifera L.) – Validation of the Chromatic Assessment Method'. In: *Journal of Apicultural Science* 60.2 (2016), pp. 209–220. ISSN: 2299-4831. DOI: 10.1515/jas-2016-0028.

[64] K. Crailsheim, N. Hrassnigg, and A. Stabentheiner. 'Diurnal behavioural differences in forager and nurse honey bees (Apis mellifera carnica Pollm)'. In: *Apidologie* 27.4 (1996), pp. 235–244. ISSN: 0044-8435. DOI: 10.1051/apido:19960406.

[65] Benjamin Dainat, Rolf Kuhn, Daniel Cherix, and Peter Neumann. 'A scientific note on the ant pitfall for quantitative diagnosis of Varroa destructor'. In: *Apidologie* 42.6 (2011), pp. 740–742. ISSN: 1297-9678. DOI: 10.1007/s13592-011-0071-3.

[66] Pier Paolo Danieli, Nicola Francesco Addeo, Filippo Lazzari, Federico Manganello, and Fulvia Bovera. 'Precision Beekeeping Systems: State of the Art, Pros and Cons, and Their Application as Tools for Advancing the Beekeeping Sector'. In: *Animals* 14 (2023), p. 70. DOI: 10.3390/ani14010070.

[67] Robert G Danka and Lorraine D Beaman. 'Flight Activity of USDA–ARS Russian Honey Bees (Hymenoptera: Apidae) During Pollination of Lowbush Blueberries in Maine'. In: *Journal of Economic Entomology* 100.2 (2007), p. 6.

[68] Pim De Klerk and Hans Joosten. 'The difference between pollen types and plant taxa: a plea for clarity and scientific freedom'. In: *Quaternary Science Journal* 56.3 (2007), pp. 162–171. ISSN: 2199-9090. DOI: 10.3285/eg.56.3.02.

[69] Dario De Nart, Cecilia Costa, Gennaro Di Prisco, and Emanuele Carpana. 'Image recognition using convolutional neural networks for classification of honey bee subspecies'. In: *Apidologie* 53.1 (2022), p. 5. ISSN: 1297-9678. DOI: 10.1007/s13592-022-00918-5.

[70] A. P. Dempster, N. M. Laird, and D. B. Rubin. 'Maximum Likelihood from Incomplete Data via the EM Algorithm'. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1 (1977), pp. 1–38.

[71] Vincent Dietemann et al. 'Standard methods for varroa research'. In: *Journal of Apicultural Research* (2013). DOI: 10.3896/IBRA.1.52.1.09.

[72] Maria Dimou and Andreas Thrasyvoulou. 'Efficient use of pollen traps to determine the pollen flora used by honey bees'. In: *Journal of Apicultural Research* (2006), pp. 42–46. ISSN: 00218839. DOI: 10.3896/IBRA.1.45.1.10.

[73] Gérard Donzé and Patrick M. Guerin. 'Behavioral attributes and parental care of Varroa mites parasitizing honeybee brood'. In: *Behavioral Ecology and Sociobiology* 34.5 (1994), pp. 305–319. ISSN: 1432-0762. DOI: 10.1007/BF00197001.

[74] Gérard Donzé and Patrick M. Guerin. 'Time-activity budgets and space structuring by the different life stages ofVarroa jacobsoni in capped brood of the honey bee,Apis mellifera'. In: *Journal of Insect Behavior* 10.3 (1997), pp. 371–393. ISSN: 1572-8889. DOI: 10.1007/BF02765605.

[75] Susanne Dunker, Elena Motivans, Demetra Rakosy, David Boho, Patrick Mäder, Thomas Hornick, and Tiffany M. Knight. 'Pollen analysis using multispectral imaging flow cytometry and deep learning'. In: *New Phytologist* 229.1 (2021), pp. 593–606. ISSN: 0028-646X, 1469-8137. DOI: 10.1111/nph.16882.

[76] EFSA Panel on Plant Protection Products and their Residues (PPR). 'Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera, Bombus spp. and solitary bees)'. In: *EFSA Journal* 10.5 (2012), p. 2668. ISSN: 1831-4732. DOI: 10.2903/j.efsa.2012.2668.

[77] E. H. Erickson, H. H. Miller, and D. J. Sikkema. 'A Method of Separating and Monitoring Honeybee Flight Activity at the Hive Entrance'. In: *Journal of Apicultural Research* 14.3 (1975), pp. 119–125. ISSN: 0021-8839, 2078-6913. DOI: 10.1080/00218839.1975.11099814.

[78] European Food Safety Authority. 'Guidance on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees)'. In: *EFSA Journal* 11.7 (2013). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2013.3295.

[79] European Food Safety Authority (EFSA), Alessio Ippolito, Monica del Aguila, Elisa Aiassa, Irene Muñoz Guajardo, Franco Maria Neri, Fernando Alvarez, Olaf Mosbach-Schulz, and Csaba Szentes. 'Review of the evidence on bee background mortality'. In: *EFSA Supporting Publications* (2020). ISSN: 23978325. DOI: 10.2903/sp.efsa.2020.EN-1880.

[80] Knut Fægri, Johannes Iversen, Peter Emil Kaland, and Knut Krzywinski. *Textbook of pollen analysis*. IV ed. Caldwell, NJ: Blackburn Press, 2007. 328 pp. ISBN: 978-1-930665-01-9.

[81] FAOSTAT. *Food and Agriculture Organization of the United Nations. (2021). Stocks of beehives world.* URL: http://www.fao.org/faostat/en/#home (visited on 01/14/2021).

[82] J. M. Flores, J. A. Ruiz, J. M. Ruz, F. Puerta, M. Bustos, F. Padilla, and F. Campano. 'Effect of temperature and humidity of sealed brood on chalkbrood development under controlled conditions'. In: *Apidologie* 27.4 (1996), pp. 185–192. ISSN: 0044-8435. DOI: 10.1051/apido:19960401.

[83] Tiago Mauricio Francoy, Dieter Wittmann, Martin Drauschke, Stefan Müller, Volker Steinhage, Marcela A. F. Bezerra-Laure, David De Jong, and Lionel Segui Gonçalves. 'Identification of Africanized honey bees through wing morphometrics: two fast and efficient procedures'. In: *Apidologie* 39.5 (2008), pp. 488–494. ISSN: 0044-8435, 1297-9678. DOI: 10.1051/apido:2008028.

[84] Eva Frey, Richard Odemer, Thomas Blum, and Peter Rosenkranz. 'Activation and interruption of the reproduction of *Varroa destructor* is triggered by host signals (*Apis mellifera*)'. In: *Journal of Invertebrate Pathology* 113.1 (2013), pp. 56–62. ISSN: 0022-2011. DOI: 10.1016/j.jip.2013.01.007.

[85] Alexander Freytag, Erik Rodner, Marcel Simon, Alexander Loos, Hjalmar S. Kühl, and Joachim Denzler. 'Chimpanzee Faces in the Wild: Log-Euclidean CNNs for Predicting Identities and Attributes of Primates'. In: *Pattern Recognition*. Ed. by Bodo Rosenhahn and Bjoern Andres. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 51–63. ISBN: 978-3-319-45886-1. DOI: 10.1007/978-3-319-45886-1_5.

[86] Ingemar Fries, Anton Imdorf, and Peter Rosenkranz. 'Survival of mite infested (Varroa destructor) honey bee (Apis mellifera) colonies in a Nordic climate'. In: *Apidologie* 37.5 (2006), pp. 564–570. ISSN: 0044-8435, 1297-9678. DOI: 10.1051/apido:2006031.

[87] Karl von Frisch. *Aus Dem Leben der Bienen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993. ISBN: 978-3-540-56763-9 978-3-642-61235-0. DOI: 10.1007/978-3-642-61235-0.

[88] Karl von Frisch. *Bees: Their Vision, Chemical Senses, and Language*. Cornell University Press, 1971. 159 pp. ISBN: 978-0-8014-7176-6.

[89] Karl von Frisch. *The Dancing Bees: An Account of the Life and Senses of the Honey Bee*. Vienna: Springer Vienna, 1954. ISBN: 978-3-7091-4549-4 978-3-7091-4697-2. DOI: 10.1007/978-3-7091-4697-2.

[90] S. Fuchs. 'Preference for drone brood cells by Varroa jacobsoni Oud in colonies of Apis mellifera carnica'. In: *Apidologie* 21.3 (1990), pp. 193–199. ISSN: 0044-8435. DOI: 10.1051/apido:19900304.

[91] Claudia Garrido and Peter Rosenkranz. 'The reproductive program of female Varroa destructor mites is triggered by its host, Apis mellifera'. In: *Experimental & Applied Acarology* 31.3 (2003), pp. 269–273. ISSN: 1572-9702. DOI: 10.1023/B:APPA.0000010386.10686.9f.

[92] Tim Gernat, Vikyath D. Rao, Martin Middendorf, Harry Dankowicz, Nigel Goldenfeld, and Gene E. Robinson. 'Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks'. In: *Proceedings of the National Academy of Sciences* 115.7 (2018), pp. 1433–1438. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1713568115.

[93] Sergio Gil-Lebrero, Francisco Quiles-Latorre, Manuel Ortiz-López, Víctor Sánchez-Ruiz, Victoria Gámiz-López, and Juan Luna-Rodríguez. 'Honey Bee Colonies Remote Monitoring System'. In: *Sensors* 17.12 (2016), p. 55. ISSN: 1424-8220. DOI: 10.3390/s17010055.

[94] Shaogang Gong and Tao Xiang. 'Person Re-identification'. In: *Visual Analysis of Behaviour: From Pixels to Semantics*. Ed. by Shaogang Gong and Tao Xiang. London: Springer, 2011, pp. 301–313. ISBN: 978-0-85729-670-2. DOI: 10.1007/978-0-85729-670-2_14.

[95] Gundula Gonsior, Frederic Tausch, Katharina Schmidt, and Silvio Knäbe. 'Impact of an Oomen feeding with a neonicotinoid on daily activity and colony development of honeybees assessed with an AI based monitoring device'. In: *Julius-Kühn-Archiv* 465 (2020). Number: 465, pp. 25–29. ISSN: 2199-921X. DOI: 10.5073/jka.2020.465.006.

[96] Dave Goulson. *Bumblebees: Their Behaviour and Ecology*. Oxford University Press, 2003. 260 pp. ISBN: 978-0-19-852607-0.

[97] Ernesto Guzmán-Novoa, Leslie Eccles, Yireli Calvete, Janine Mcgowan, Paul G. Kelly, and Adriana Correa-Benítez. 'Varroa destructor is the main culprit for the death and reduced populations of overwintered honey bee (Apis mellifera) colonies in Ontario, Canada'. In: *Apidologie* 41.4 (2010), pp. 443–450. ISSN: 1297-9678. DOI: 10.1051/apido/2009076.

[98] Hugo Hadjur, Doreid Ammar, and Laurent Lefèvre. 'Toward an intelligent and efficient beehive: A survey of precision beekeeping systems and services'. In: *Computers and Electronics in Agriculture* 192 (2022), p. 106604. ISSN: 0168-1699. DOI: 10.1016/j.compag.2021.106604.

[99] Harriet Hall, Martin Bencsik, and Michael Newton. 'Automated, non-invasive Varroa mite detection by vibrational measurements of gait combined with machine learning'. In: *Scientific Reports* 13.1 (2023), p. 10202. ISSN: 2045-2322. DOI: 10.1038/s41598-023-36810-0.

[100] Caspar A. Hallmann et al. 'More than 75 percent decline over 27 years in total flying insect biomass in protected areas'. In: *PLOS ONE* 12.10 (2017). Ed. by Eric Gordon Lamb. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0185809.

[101] Boris Hanin. 'Which Neural Net Architectures Give Rise to Exploding and Vanishing Gradients?' In: *32nd Conference on Neural Information Processing Systems*. NeurIPS. Vol. 31. Montréal, Canada, 2018.

[102] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 'Deep Residual Learning for Image Recognition'. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[103] L. A. F. Heath. 'Development of Chalk Brood in a Honeybee Colony: A Review'. In: *Bee World* 63.3 (1982), pp. 119–130. ISSN: 0005-772X, 2376-7618. DOI: 10.1080/0005772X.1982.11097876.

[104] G. E. Hinton and R. R. Salakhutdinov. 'Reducing the Dimensionality of Data with Neural Networks'. In: *Science* 313.5786 (2006). Publisher: American Association for the Advancement of Science, pp. 504–507. DOI: 10.1126/science.1127647.

[105] Dorothy Hodges. *The pollen loads of the honeybee. A guide o their identification by colour and form*. London: Bee Research Association, 1952. ISBN: 0-900149-50-7.

[106] Toke T. Høye et al. 'Deep learning and computer vision will transform entomology'. In: *Proceedings of the National Academy of Sciences* 118.2 (2021). Publisher: National Academy of Sciences Section: Perspective. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2002545117.

[107] M. D. Ifantidis. 'Ontogenesis of the Mite Varroa Jacobsoni in Worker and Drone Honeybee Brood Cells'. In: *Journal of Apicultural Research* 22.3 (1983), pp. 200–206. ISSN: 0021-8839. DOI: 10.1080/00218839.1983.11100588.

[108] Michael D. Ifantidis. 'Some Aspects of the Process Varroa Jacobsoni Mite Entrance Into Honey bee (Apis Mellifera) Brood Cells'. In: *Apidologie* 19.4 (1988), pp. 387–396. ISSN: 0044-8435. DOI: 10.1051/apido:19880406.

[109] Deutscher Imkerbund. *Imkerei in Deutschland - D.I.B. Mitgliederstatistik*. https://deutscherimkerbund.de/. 2023.

[110] Cameron J Jack and James D Ellis. 'Integrated Pest Management Control of Varroa destructor (Acari: Varroidae), the Most Damaging Pest of (Apis mellifera L. (Hymenoptera: Apidae)) Colonies'. In: *Journal of Insect Science* 21.5 (2021), p. 6. ISSN: 1536-2442. DOI: 10.1093/jisesa/ieab058.

[111] Joe-Air Jiang, Chien-Hao Wang, Chi-Hui Chen, Min-Sheng Liao, Yu-Li Su, Wei-Sheng Chen, Chien-Peng Huang, En-Cheng Yang, and Cheng-Long Chuang. 'A WSN-based automatic monitoring system for the foraging behavior of honey bees and environmental factors of beehives'. In: *Computers and Electronics in Agriculture* 123 (2016), pp. 304–318. ISSN: 01681699. DOI: 10.1016/j.compag.2016.03.003.

[112] Rebecca N Johnson, M Tauheed Zaman, Meredith M Decelle, Adam J Siegel, David R Tarpy, Eli C Siegel, and Philip T Starks. 'Multiple micro-organisms in chalkbrood mummies: evidence and implications'. In: *Journal of Apicultural Research* 44.1 (2005), pp. 29–32. ISSN: 0021-8839. DOI: 10.1080/00218839.2005.11101143.

[113] Reed M. Johnson, Lizette Dahlgren, Blair D. Siegfried, and Marion D. Ellis. 'Acaricide, Fungicide and Drug Interactions in Honey Bees (Apis mellifera)'. In: *PLOS ONE* 8.1 (2013). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0054092.

[114] Sanket Kachole, Gordon Hunter, and Olga Duran. 'A Computer Vision Approach to Monitoring the Activity and Well-Being of Honeybees'. In: *Intelligent Environments*. Ambient Intelligence and Smart Environments 28 (2020), pp. 152–161. DOI: 10.3233/AISE200036.

[115] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. 'Deep learning in agriculture: A survey'. In: *Computers and Electronics in Agriculture* 147 (2018), pp. 70–90. ISSN: 0168-1699. DOI: 10.1016/j.compag.2018.02.016.

[116] Gerald Kastberger, Sarah Radloff, and Gerhard Kranner. 'Individuality of wing patterning in Giant honey bees (Apis laboriosa)'. In: *Apidologie* 34.3 (2003), pp. 311–318. DOI: 10.1051/apido:2003020.

[117] Irene Keller, Peter Fluri, and Anton Imdorf. 'Pollen nutrition and colony development in honey bees: part 1'. In: *Bee World* 86.1 (2005), pp. 3–10. ISSN: 0005-772X. DOI: 10.1080/0005772X.2005.11099641.

[118] William B. Kerfoot. 'A Photoelectric Activity Recorder for Studies of Insect Behavior'. In: *Journal of the Kansas Entomological Society* 39.4 (1966), pp. 629–633.

[119] Toshifumi Kimura, Mizue Ohashi, Ryuichi Okada, and Hidetoshi Ikeno. 'A new approach for the simultaneous tracking of multiple honeybees for analysis of hive behavior'. In: *Apidologie* 42.5 (2011), pp. 607–617. ISSN: 0044-8435, 1297-9678. DOI: 10.1007/s13592-011-0060-6.

[120] Diederik P. Kingma and Jimmy Ba. 'Adam: A Method for Stochastic Optimization'. In: International Conference on Learning Representations (ICLR). 2015. arXiv: 1412.6980[cs].

[121] William D J Kirk. 'Recording the Colours of Pollen Loads'. In: *Bee World* 75.4 (1994), pp. 169–180. ISSN: 0005-772X. DOI: 10.1080/0005772X.1994.11099224.

[122] William D. J. Kirk. *A Colour Guide To Pollen Loads Of The Honey Bee*. 2nd ed. Cardiff: International Bee Research Association, 2006. 54 pp. ISBN: 0-86098-248-3.

[123] William D. J. Kirk. 'The Colours of Pollen Available to Honey Bees through the Year'. In: *Bee World* 95.3 (2018), pp. 74–77. ISSN: 0005-772X, 2376-7618. DOI: 10.1080/0005772X.2018.1449280.

[124] Fritz Knoll. *Die Biologie der Blüte*. Vol. 57. Verständliche Wissenschaft. Berlin, Heidelberg: Springer Berlin Heidelberg, 1956. ISBN: 978-3-642-86221-2 978-3-642-86220-5. DOI: 10.1007/978-3-642-86220-5.

[125] Matthias Körschens, Björn Barz, and Joachim Denzler. *Towards Automatic Identification of Elephants in the Wild*. 2018.

[126] Martin Köstinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof. 'Large scale metric learning from equivalence constraints'. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. ISSN: 1063-6919. 2012, pp. 2288–2295. DOI: 10.1109/CVPR.2012.6247939.

[127] Douglas S. Kridi, Carlos Giovanni N. de Carvalho, and Danielo G. Gomes. 'Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection'. In: *Computers and Electronics in Agriculture* 127 (2016), pp. 221–235. ISSN: 01681699. DOI: 10.1016/j.compag.2016.05.013.

[128] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 'ImageNet Classification with Deep Convolutional Neural Networks'. In: *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc., 2012.

[129] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 'ImageNet classification with deep convolutional neural networks'. In: *Communications of the ACM* 60.6 (2017), pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386.

[130] Rolf G. Kuehni. *Color: An Introduction to Practice and Principles*. John Wiley & Sons, 2012. 308 pp. ISBN: 978-1-118-17384-8.

[131] Vladimir Kulyukin. 'Audio, Image, Video, and Weather Datasets for Continuous Electronic Beehive Monitoring'. In: *Applied Sciences* 11.10 (2021). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 4632. ISSN: 2076-3417. DOI: 10.3390/app11104632.

[132] Vladimir Kulyukin and Sarbajit Mukherjee. 'On Video Analysis of Omnidirectional Bee Traffic: Counting Bee Motions with Motion Detection and Image Classification'. In: *Applied Sciences* 9.18 (2019), p. 3743. ISSN: 2076-3417. DOI: 10.3390/app9183743.

[133] Vladimir Kulyukin, Sarbajit Mukherjee, Angela Minichiello, and Tadd Truscott. 'BeePIV: A Method to Measure Apis Mellifera Traffic with Particle Image Velocimetry in Videos'. In: *Applied Sciences* 11.5 (2021), p. 2276. ISSN: 2076-3417. DOI: 10.3390/app11052276.

[134] Ashis Kumar Samanta, ed. *Colorimetry*. IntechOpen, 2022. ISBN: 978-1-83962-940-2 978-1-83962-941-9.

[135] Ujjwal Layek, Nandita Das, Arijit Kundu, and Prakash Karmakar. 'Methods Employed in the Determining Nectar and Pollen Sources for Bees: A Review of the Global Scenario'. In: *Annals of the Entomological Society of America* 115.6 (2022), pp. 417–426. ISSN: 0013-8746. DOI: 10.1093/aesa/saac013.

[136] Thi-Nhung Le, Duc-Ngoc Tran, Thi-Thu-Hong Phan, Hong-Thai Pham, Thi-Lan Le, and Hai Vu. 'A Robust Multiple Honeybee Tracking Method from Videos Captured at Beehive Entrance'. In: *2023 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*. ISSN: 2770-6850. 2023, pp. 1–6. DOI: 10.1109/MAPR59823.2023.10289105.

[137] Gil Leclercq, Tjeerd Blacquière, Nicolas Gengler, and Frédéric Francis. 'Hygienic removal of freeze-killed brood does not predict Varroa-resistance traits in unselected stocks'. In: *Journal of Apicultural Research* 57.2 (2018), pp. 292–299. ISSN: 0021-8839. DOI: 10.1080/00218839.2018.1426350.

[138] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 'Backpropagation Applied to Handwritten Zip Code Recognition'. In: *Neural Computation* 1.4 (1989), pp. 541–551. ISSN: 0899-7667. DOI: 10.1162/neco.1989.1.4.541.

[139] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 'Deep learning'. In: *Nature* 521.7553 (2015). Publisher: Nature Publishing Group, pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539.

[140] Hsien-Che Lee. *Introduction to Color Imaging Science*. Cambridge: Cambridge University Press, 2005. ISBN: 978-0-521-84388-1. DOI: 10.1017/CBO9780511614392.017.

[141] Regis Lefebre, David Claeys Bouuaert, Emma Bossuyt, Lina De Smet, Marleen Brunain, Ellen Danneels, and Dirk C. de Graaf. 'Comprehensive Approach to Phenotype Varroa destructor Reproduction in Honey Bee Drone Brood and Its Correlation with Decreased Mite Reproduction (DMR)'. In: *Insects* 15.6 (2024), p. 397. ISSN: 2075-4450. DOI: 10.3390/insects15060397.

[142] Jan Lindsten. *Nobel Lectures, Physiology or Medicine 1971-1980*. www.nobelprize.org/prizes/medicine/1973/ceremony-speech/. Singapore: World Scientific Publishing Co, 1992.

[143] Christian Lippert, Arndt Feuerbacher, and Manuel Narjes. 'Revisiting the economic valuation of agricultural losses due to large-scale changes in pollinator populations'. In: *Ecological Economics* 180 (2021), p. 106860. ISSN: 09218009. DOI: 10.1016/j.ecolecon.2020.106860.

[144] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. 'A survey on deep learning in medical image analysis'. In: *Medical Image Analysis* 42 (2017), pp. 60–88. ISSN: 1361-8415. DOI: 10.1016/j.media.2017.07.005.

[145] C. Liu, J. J. Leonard, and J. J. Feddes. 'Automated Monitoring of Flight Activity at a Beehive Entrance using Infrared Light Sensors'. In: *Journal of Apicultural Research* 29.1 (1990), pp. 20–27. ISSN: 0021-8839, 2078-6913. DOI: 10.1080/00218839.1990.11101193.

[146] S. Lloyd. 'Least squares quantization in PCM'. In: *IEEE Transactions on Information Theory* 28.2 (1982). Conference Name: IEEE Transactions on Information Theory, pp. 129–137. ISSN: 1557-9654. DOI: 10.1109/TIT.1982.1056489.

[147] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 'Fully Convolutional Networks for Semantic Segmentation'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. CVPR. 2015.

[148] J. Louveaux, Anna Maurizio, and G. Vorwohl. 'Methods of Melissopalynology'. In: *Bee World* 59.4 (1978), pp. 139–157. ISSN: 0005-772X. DOI: 10.1080/0005772X.1978.11097714.

[149] M. R. Luo, G. Cui, and B. Rigg. 'The development of the CIE 2000 colour-difference formula: CIEDE2000'. In: *Color Research & Application* 26.5 (2001), pp. 340–350. ISSN: 1520-6378. DOI: 10.1002/col.1049.

[150] P A Macedo, J Wu, and M D Ellis. 'Using inert dusts to detect and assess varroa infestations in honey bee colonies'. In: *Journal of Apicultural Research* 41.1 (2002), pp. 3–7. ISSN: 0021-8839. DOI: 10.1080/00218839.2002.11101062.

[151] Matías Maggi, Natalia Damiani, Sergio Ruffinengo, David De Jong, Judith Principal, and Martín Eguaras. 'Brood cell size of Apis mellifera modifies the reproductive behavior of Varroa destructor'. In: *Experimental and Applied Acarology* 50.3 (2010), pp. 269–279. ISSN: 1572-9702. DOI: 10.1007/s10493-009-9314-7.

[152] Baptiste Magnier, Gaëtan Ekszterowicz, Joseph Laurent, Matthias Rival, and François Pfister. 'Bee Hive Traffic Monitoring by Tracking Bee Flight Paths:' in: *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. International Conference on Computer Vision Theory and Applications. Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 563–571. ISBN: 978-989-758-290-5. DOI: 10.5220/0006628205630571.

[153] P. C. Mahalanobis. 'On the Generalized Distance in Statistics'. In: *Proceedings of the National Institute of Sciences of India* 2.1 (1936).

[154] Daniel Malacara. *Color vision and colorimetry: theory and applications*. In collab. with Society of Photo-optical Instrumentation Engineers. 2nd ed. OCLC: ocn696100092. Bellingham, Wash: SPIE, 2011. 175 pp. ISBN: 978-0-8194-8397-3.

[155] Paul Marchal, Alexis Buatois, Stéphane Kraus, Simon Klein, Tamara Gomez-Moracho, and Mathieu Lihoreau. 'Automated monitoring of bee behaviour using connected hives: towards a computational apidology'. In: *Apidologie* 51.3 (2020), pp. 356–368.

[156] Julian Marstaller, Frederic Tausch, and Simon Stock. 'DeepBees - Building and Scaling Convolutional Neuronal Nets For Fast and Large-Scale Visual Monitoring of Bee Hives'. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. Seoul, Korea (South): IEEE, 2019, pp. 271–278. ISBN: 978-1-72815-023-9. DOI: 10.1109/ICCVW.2019.00036.

[157] S J Martin and D Kemp. 'Average number of reproductive cycles performed by Varroa jacobsoni in honey bee (Apis mellifera) colonies'. In: *Journal of Apicultural Research* 36.3 (1997), pp. 113–123. ISSN: 0021-8839. DOI: 10.1080/00218839.1997.11100937.

[158] Rose A. McGruddy, Zoe E. Smeele, Brian Manley, James D. Masucci, John Haywood, and Philip J. Lester. 'RNA interference as a next-generation control method for suppressing Varroa destructor reproduction in honey bee (Apis mellifera) hives'. In: *Pest Management Science* 80.9 (2024), pp. 4770–4778. ISSN: 1526-4998. DOI: 10.1002/ps.8193.

[159] Rémi Mégret, Ivan F. Rodriguez, Isada Claudio Ford, Edgar Acuña, Jose L. Agosto-Rivera, and Tugrul Giray. 'LabelBee: a web platform for large-scale semi-automated analysis of honeybee behavior from video'. In: *Proceedings of the Conference on Artificial Intelligence for Data Discovery and Reuse*. AIDR '19: Artificial Intelligence for Data Discovery and Reuse 2019. Pittsburgh Pennsylvania: ACM, 2019, pp. 1–4. ISBN: 978-1-4503-7184-1. DOI: 10.1145/3359115.3359120.

[160] W. G. Meikle and N. Holst. 'Application of continuous monitoring of honeybee colonies'. In: *Apidologie* 46.1 (2015), pp. 10–22. ISSN: 0044-8435, 1297-9678. DOI: 10.1007/s13592-014-0298-x.

[161] William G. Meikle, Niels Holst, Théotime Colin, Milagra Weiss, Mark J. Carroll, Quinn S. McFrederick, and Andrew B. Barron. 'Using within-day hive weight changes to measure environmental effects on honey bee colonies'. In: *PLOS ONE* 13.5 (2018). Ed. by Adrian G. Dyer, e0197589. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0197589.

[162] Massimiliano Micheli, Simone Pasinetti, Matteo Lancini, and Gabriele Coffetti. 'Development of a monitoring system to assess honeybee colony health'. In: *2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*. 2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor). 2022, pp. 234–238. DOI: 10.1109/MetroAgriFor55389.2022.9964541.

[163] Liz Milla, Alexander Schmidt-Lebuhn, Jessica Bovill, and Francisco Encinas-Viso. 'Monitoring of honey bee floral resources with pollen DNA metabarcoding as a complementary tool to vegetation surveys'. In: *Ecological Solutions and Evidence* 3.1 (2022). ISSN: 2688-8319. DOI: 10.1002/2688-8319.12120.

[164] W. S. Mokrzycki and M. Tatol. 'Colour difference Delta E - a survey'. In: *Machine Graphics & Vision International Journal* 20.4 (2011), pp. 383–411. ISSN: 1230-0535.

[165] John M. Mola and Neal M. Williams. 'A review of methods for the study of bumble bee movement'. In: *Apidologie* 50.4 (2019), pp. 497–514. ISSN: 1297-9678. DOI: 10.1007/s13592-019-00662-3.

[166] Fanny Mondet, Alexis Beaurepaire, Alison McAfee, Barbara Locke, Cédric Alaux, Solene Blanchard, Bob Danka, and Yves Le Conte. 'Honey bee survival mechanisms against the parasite *Varroa destructor*: a systematic review of phenotypic and genomic research efforts'. In: *International Journal for Parasitology* 50.6 (2020), pp. 433–447. ISSN: 0020-7519. DOI: 10.1016/j.ijpara.2020.03.005.

[167] Linde Morawetz, Hemma Köglberger, Antonia Griesbacher, Irmgard Derakhshifar, Karl Crailsheim, Robert Brodschneider, and Rudolf Moosbeckhofer. 'Health status of honey bee colonies (Apis mellifera) and disease-related risk factors for colony losses in Austria'. In: *PLoS ONE* 14.7 (2019). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0219293.

[168] Marie-Lou Morin and Pierre Giovenazzo. 'Mite non-reproduction, recapping behavior, and hygienic behavior (freeze-kill method) linked to Varroa destructor infestation levels in selected Apis mellifera colonies'. In: *Journal of Veterinary Diagnostic Investigation* 35.6 (2023), pp. 655–663. ISSN: 1040-6387. DOI: 10.1177/10406387231172141.

[169] Sarbajit Mukherjee and Vladimir Kulyukin. 'Application of Digital Particle Image Velocimetry to Insect Motion: Measurement of Incoming, Outgoing, and Lateral Honeybee Traffic'. In: *Applied Sciences* 10.6 (2020), p. 2042. ISSN: 2076-3417. DOI: 10.3390/app10062042.

[170] Nihal Murali, Jon Schneider, Joel Levine, and Graham Taylor. 'Classification and Re-Identification of Fruit Fly Individuals Across Days With Convolutional Neural Networks'. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa Village, HI, USA: IEEE, 2019, pp. 570–578. ISBN: 978-1-72811-975-5. DOI: 10.1109/WACV.2019.00066.

[171] Vinod Nair and Geoffrey E Hinton. 'Rectified Linear Units Improve Restricted Boltzmann Machines'. In: *Proceedings of the 27th international conference on machine learning*. ICML. 2010, pp. 807–814.

[172] National Research Council (U.S.) and Committee on the Status of Pollinators in North America, eds. *Status of pollinators in North America*. OCLC: 891355949. Washington, D.C: National Academy of Sciences, 2007. ISBN: 978-0-309-10289-6 978-0-309-66381-6.

[173] F. Nazzi and N. Milani. 'A technique for reproduction of Varroa jacobsoni Oud under laboratory conditions'. In: *Apidologie* 25.6 (1994), pp. 579–584. ISSN: 0044-8435. DOI: 10.1051/apido:19940608.

[174] Linda Newstrom-Lloyd, Ian Raine, and Xun Li. 'The power of pollen profiles for planting trees for bees'. In: *Trees for Bees* (2017), p. 12.

[175] Beatrice Tchuidjang Nganso et al. *Managed honey bee colony losses and causes during the active beekeeping season 2022/2023 in nine Sub-Saharan African countries*. Preprint. 2024. DOI: 10.1101/2024.04.30.591982.

[176] Thi Nha Ngo, Dan Jeric Arcega Rustia, En-Cheng Yang, and Ta-Te Lin. 'Automated monitoring and analyses of honey bee pollen foraging behavior using a deep learning-based imaging system'. In: *Computers and Electronics in Agriculture* 187 (2021), p. 106239. ISSN: 01681699. DOI: 10.1016/j.compag.2021.106239.

[177] Thi Nha Ngo, Kung-Chin Wu, En-Cheng Yang, and Ta-Te Lin. 'A real-time imaging system for multiple honey bee tracking and activity monitoring'. In: *Computers and Electronics in Agriculture* 163 (2019), p. 104841. ISSN: 01681699. DOI: 10.1016/j.compag.2019.05.050.

[178] Dinh-Tu Nguyen, Thi-Nhung Le, Thi-Huong Phung, Duc-Manh Nguyen, Hong-Quan Nguyen, Hong-Thai Pham, Thi-Thu-Hong Phan, Hai Vu, and Thi-Lan Le. 'Improving pollen-bearing honey bee detection from videos captured at hive entrance by combining deep learning and handling imbalance techniques'. In: *Ecological Informatics* 82 (2024), p. 102744. ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2024.102744.

[179] A Nieto et al. *European red list of bees*. Luxembourg: Publication Office of the European Union, 2014. ISBN: 978-92-79-44512-5.

[180] Alicia Noriega-Escamilla, César J. Camacho-Bello, Rosa M. Ortega-Mendoza, José H. Arroyo-Núñez, and Lucia Gutiérrez-Lazcano. 'Varroa Destructor Classification Using Legendre–Fourier Moments with Different Color Spaces'. In: *Journal of Imaging* 9.7 (2023), p. 144. ISSN: 2313-433X. DOI: 10.3390/jimaging9070144.

[181] Richard Odemer. 'Approaches, challenges and recent advances in automated bee counting devices: A review'. In: *Annals of Applied Biology* (2021), aab.12727. ISSN: 0003-4746, 1744-7348. DOI: 10.1111/aab.12727.

[182] Richard Odemer. 'Reproductive capacity of *varroa destructor* in four different honey bee subspecies'. In: *Saudi Journal of Biological Sciences* 27.1 (2020), pp. 247–250. DOI: 10.1016/j.sjbs.2019.09.002.

[183] Richard Odemer, Franziska Odemer, Gerhard Liebig, and Doris de Craigher. 'Temporal increase of Varroa mites in trap frames used for drone brood removal during the honey bee season'. In: *Journal of Applied Entomology* 146.9 (2022), pp. 1207–1211. ISSN: 1439-0418. DOI: 10.1111/jen.13046.

[184] Jeff Ollerton, Rachael Winfree, and Sam Tarrant. 'How many flowering plants are pollinated by animals?' In: *Oikos* 120.3 (2011), pp. 321–326. ISSN: 1600-0706. DOI: 10.1111/j.1600-0706.2010.18644.x.

[185] Delphine Panziera, Frank van Langevelde, and Tjeerd Blacquière. 'Varroa sensitive hygiene contributes to naturally selected varroa resistance in honey bees'. In: *Journal of Apicultural Research* 56.5 (2017), pp. 635–642. ISSN: 0021-8839. DOI: 10.1080/00218839.2017.1351860.

[186] Mary Percival. 'Pollen Collection By Apis Mellifera'. In: *New Phytologist* 46.1 (1947), pp. 142–165. DOI: 10.1111/j.1469-8137.1947.tb05076.x.

[187] Uroš Pešović, Siniša RanĎić, and Zoran Stamenković. 'Design and implementation of hardware platform for monitoring honeybee activity'. In: *Proceedings of 4th International Conference on Electrical, Electronics and Computing Engineering-Serbia*. 2017.

[188] Minh-Hà Pham-Delègue, Axel Decourtye, Laure Kaiser, and James Devillers. 'Behavioural methods to assess the effects of pesticides on honey bees'. In: *Apidologie* 33.5 (2002), pp. 425–432. ISSN: 0044-8435, 1297-9678. DOI: 10.1051/apido:2002033.

[189] Phil Green. *Fundamentals and Applications of Colour Engineering*. Wiley-SID Series in Display Technology. Hoboken, NJ: Wiley, 2024. ISBN: 978-1-119-82718-4.

[190] R. S. Pickard and D. Hepworth. 'A method for electronically monitoring the ambulatory activity of honeybees under dark conditions'. In: *Behavior Research Methods & Instrumentation* 11.4 (1979), pp. 433–436. ISSN: 1554-351X, 1554-3528. DOI: 10.3758/BF03205697.

[191] Vincent Piou and Angélique Vétillard. 'Varroa destructor rearing in laboratory conditions: importance of foundress survival in doubly infested cells and reproduction of laboratory-born females'. In: *Apidologie* 51.6 (2020), pp. 968–983. ISSN: 1297-9678. DOI: 10.1007/s13592-020-00775-0.

[192] F. Puerta, J. M. Flores, M. Bustos, F. Padilla, and F. Campano. 'Chalkbrood development in honeybee brood under controlled conditions'. In: *Apidologie* 25.6 (1994), pp. 540–546. ISSN: 0044-8435. DOI: 10.1051/apido:19940604.

[193] Michael-Thomas Ramsey, Martin Bencsik, Michael Ian Newton, Maritza Reyes, Maryline Pioz, Didier Crauser, Noa Simon Delso, and Yves Le Conte. 'The prediction of swarming in honeybee colonies using vibrational spectra'. In: *Scientific Reports* 10.1 (2020). Publisher: Nature Publishing Group, p. 9798. ISSN: 2045-2322. DOI: 10.1038/s41598-020-66115-5.

[194] Samuel D. Ramsey et al. 'Varroa destructor feeds primarily on honey bee fat body tissue and not hemolymph'. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.5 (2019), pp. 1792–1801. ISSN: 1091-6490. DOI: 10.1073/pnas.1818371116.

[195] Malika Nisal Ratnayake, Adrian G. Dyer, and Alan Dorin. 'Tracking individual honeybees among wildflower clusters with computer vision-facilitated pollinator monitoring'. In: *PLOS ONE* 16.2 (2021). Publisher: Public Library of Science, e0239504. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0239504.

[196] Allan Rodrigues Rebelo, Joao M. G. Fagundes, Luciano A. Digiampietri, Tiago M. Francoy, and Helton Hideraldo Bíscaro. 'A fully automatic classification of bee species from wing images'. In: *Apidologie* 52.6 (2021), pp. 1060–1074. ISSN: 1297-9678. DOI: 10.1007/s13592-021-00887-1.

[197] Allan Rodrigues Rebelo, Joao Marcos Garcia Fagundes, Luciano Antonio Digiampietri, and Helton Hideraldo Bíscaro. 'Methods for Automatic Image-Based Classification of Winged Insects Using Computational Techniques: A Systematic Literature Review'. In: *XVI Brazilian Symposium on Information Systems*. SBSI'20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–8. ISBN: 978-1-4503-8873-3. DOI: 10.1145/3411564.3411641.

[198] Sai Kiran Reka. 'A Vision-Based Bee Counting Algorithm for Electronic Monitoring of Langsthroth Beehives'. Master thesis. Logan, Utah: Utah State University, 2016.

[199] Douglas Reynolds. 'Gaussian Mixture Models'. In: *Encyclopedia of Biometrics*. Ed. by Stan Z. Li and Anil Jain. Boston, MA: Springer US, 2009, pp. 659–663. ISBN: 978-0-387-73003-5. DOI: 10.1007/978-0-387-73003-5_196.

[200] Manfred Richter. *Einführung in die Farbmetrik*. Vol. 2608. Walter de Gruyter, 2011.

[201] M. Rickli, G. Bühlmann, L. Gerig, H. Herren, H. J. Schürch, W. Zeier, and A. Imdorf. 'Zur Anwendung eines elektronischen Bienenzählgerätes am Flugloch eines Bienenvolkes'. In: *Apidologie* 20.4 (1989), pp. 305–315. ISSN: 0044-8435. DOI: 10.1051/apido:19890403.

[202] Frank D. Rinkevich. 'Detection of amitraz resistance and reduced treatment efficacy in the Varroa Mite, Varroa destructor, within commercial beekeeping operations'. In: *PLoS ONE* 15.1 (2020), e0227264. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0227264.

[203] Herbert Robbins and Sutton Monro. 'A Stochastic Approximation Method'. In: *The Annals of Mathematical Statistics* 22.3 (1951). Publisher: Institute of Mathematical Statistics, pp. 400–407. ISSN: 0003-4851.

[204] Alan R. Robertson. 'The CIE 1976 Color-Difference Formulae'. In: *Color Research & Application* 2.1 (1977), pp. 7–11. ISSN: 1520-6378. DOI: 10.1002/j.1520-6378.1977.tb00104.x.

[205] Ivan F. Rodriguez, Remi Megret, Edgar Acuna, Jose L. Agosto-Rivera, and Tugrul Giray. 'Recognition of Pollen-Bearing Bees from Video Using Convolutional Neural Network'. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, NV: IEEE, 2018, pp. 314–322. ISBN: 978-1-5386-4886-5. DOI: 10.1109/WACV.2018.00041.

[206] Sóstenes R Rodríguez-Dehaibes, Gabriel Otero-Colina, Violeta Pardio Sedas, and Juan A Villanueva Jiménez. 'Resistance to amitraz and flumethrin in Varroa destructor populations from Veracruz, Mexico'. In: *Journal of Apicultural Research* 44.3 (2005), pp. 124–125. ISSN: 0021-8839. DOI: 10.1080/00218839.2005.11101162.

[207] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 'U-net: Convolutional networks for biomedical image segmentation'. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[208] Peter Rosenkranz, Pia Aumeier, and Bettina Ziegelmann. 'Biology and control of *Varroa destructor*'. In: *Journal of Invertebrate Pathology* 103 (2010), S96–S119. ISSN: 0022-2011. DOI: 10.1016/j.jip.2009.07.016.

[209] Ralf Rosenquist. *Sensorsystem zur Erfassung und Einschätzung von Einflüssen auf die Gesundheit, Bestäubungsleistung und Vitalität von Bienenvölkern*. Abschlussbericht. GERO Meßsysteme GmbH, 2019.

[210] C Roth. 'Effectiveness of Bee Reidentification using their Abdomen - A Reproduction and Analysis'. Seminar Thesis. Karlsruhe Institute of Technology, 2024.

[211] T. H. Roulston and J. H. Cane. 'Pollen nutritional content and digestibility for animals'. In: *Plant Systematics and Evolution* 222.1 (2000), pp. 187–209. ISSN: 1615-6110. DOI: 10.1007/BF00984102.

[212] A. De Ruijter. 'Reproduction of Varroa mites during successive brood cycles of the honeybee'. In: *European Research on Varroatosis Control*. Num Pages: 5. CRC Press, 1988. ISBN: 978-1-00-321146-4.

[213] Claudia Y. Salazar-González, Francisco J. Rodríguez-Pulido, Anass Terrab, Consuelo Díaz-Moreno, Carlos A. Fuenmayor, and Francisco J. Heredia. 'Analysis of Multifloral Bee Pollen Pellets by Advanced Digital Imaging Applied to Functional Food Ingredients'. In: *Plant Foods for Human Nutrition* 73.4 (2018), pp. 328–335. ISSN: 0921-9668, 1573-9104. DOI: 10.1007/s11130-018-0695-9.

[214] Francisco Sánchez-Bayo and Kris A.G. Wyckhuys. 'Worldwide decline of the entomofauna: A review of its drivers'. In: *Biological Conservation* 232 (2019), pp. 8–27. ISSN: 00063207. DOI: 10.1016/j.biocon.2019.01.020.

[215] Ricarda Scheiner et al. 'Standard methods for behavioural studies of Apis mellifera'. In: *Journal of Apicultural Research* 52.4 (2013), p. 158. DOI: 10.3896/IBRA.1.52.4.04.

[216] Stefan Schneider, Graham W. Taylor, Stefan Linquist, and Stefan C. Kremer. 'Past, present and future approaches using computer vision for animal reidentification from camera trap data'. In: *Methods in Ecology and Evolution* 10.4 (2019). Ed. by Robert B. O'Hara, pp. 461–470. ISSN: 2041-210X, 2041-210X. DOI: 10.1111/2041-210X.13133.

[217] Stefan Schroder, Dieter Wittmann, Wilhelm Drescher, Volker Roth, Volker Steinhage, and Armin B Cremers. 'The New Key to Bees: Automated Identification by Image Analysis of Wings'. In: (2002), p. 9.

[218] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 'Facenet: A unified embedding for face recognition and clustering'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. CVPR. 2015, pp. 815–823.

[219] Antonia Schuhmann, Anna Paulina Schmid, Sarah Manzer, Janna Schulte, and Ricarda Scheiner. 'Interaction of Insecticides and Fungicides in Bees'. In: *Frontiers in Insect Science* 1 (2022). ISSN: 2673-8600. DOI: https://doi.org/10.3389/finsc.2021.808335.

[220] Stefan Schurischuster and Martin Kampel. 'Image-based Classification of Honeybees'. In: *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA). ISSN: 2154-512X. 2020, pp. 1–6. DOI: 10.1109/IPTA50016.2020.9286673.

[221] Stefan Schurischuster, Beatriz Remeseiro, Petia Radeva, and Martin Kampel. 'A Preliminary Study of Image Analysis for Parasite Detection on Honey Bees'. In: *Image Analysis and Recognition*. Ed. by Aurélio Campilho, Fakhri Karray, and Bart ter Haar Romeny. Cham: Springer International Publishing, 2018, pp. 465–473. ISBN: 978-3-319-93000-8. DOI: 10.1007/978-3-319-93000-8_52.

[222] Stefan Schurischuster, Beatriz Remeseiro, Petia Radeva, and Martin Kampel. 'A Preliminary Study of Image Analysis for Parasite Detection on Honey Bees'. In: *Image Analysis and Recognition*. Ed. by Aurélio Campilho, Fakhri Karray, and Bart ter Haar Romeny. Vol. 10882. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 465–473. ISBN: 978-3-319-92999-6 978-3-319-93000-8. DOI: 10.1007/978-3-319-93000-8_52.

[223] Linda G Shapiro and George C Stockman. *Computer vision*. Upper Saddle River, NJ: Prentice Hall, 2001. ISBN: 0-13-030796-3.

[224] Paul Siefert, Rudra Hota, Visvanathan Ramesh, and Bernd Grünewald. 'Chronic within-hive video recordings detect altered nursing behaviour and retarded larval development of neonicotinoid treated honey bees'. In: *Scientific Reports* 10.1 (2020), p. 8727. ISSN: 2045-2322. DOI: 10.1038/s41598-020-65425-y.

[225] Karen Simonyan and Andrew Zisserman. 'Very Deep Convolutional Networks for Large-Scale Image Recognition'. In: *3rd International Conference on Learning Representations (ICLR 2015), Computational and Biological Learning Society*. 2015. arXiv: 1409.1556[cs].

[226] Tomyslav Sledevič. 'The Application of Convolutional Neural Network for Pollen Bearing Bee Classification'. In: *2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*. 2018. DOI: 10.1109/AIEEE.2018.8592464.

[227] A.E. Souza Cunha, J. Rose, J. Prior, H.M. Aumann, N.W. Emanetoglu, and F.A. Drummond. 'A novel non-invasive radar to monitor honey bee colony health'. In: *Computers and Electronics in Agriculture* 170 (2020), p. 105241. ISSN: 01681699. DOI: 10.1016/j.compag.2020.105241.

[228] Hayward G. Spangler. 'Photoelectrical Counting of Outgoing and Incoming Honey Bees'. In: *Journal of Economic Entomology* 62.5 (1969), pp. 1183–1184. ISSN: 1938-291X, 0022-0493. DOI: 10.1093/jee/62.5.1183.

[229] Lina Sprau, Kirsten Traynor, and Peter Rosenkranz. 'Honey bees (Apis mellifera) preselected for Varroa sensitive hygiene discriminate between live and dead Varroa destructor and inanimate objects'. In: *Scientific Reports* 13.1 (2023), p. 10340. ISSN: 2045-2322. DOI: 10.1038/s41598-023-37356-x.

[230] Jozef Sjef van der Steen and Robert Brodschneider. 'Public Participation In Bee Science: C.S.I. Pollen'. In: *Bee World* 91.1 (2014), pp. 25–27. ISSN: 0005-772X, 2376-7618. DOI: 10.1080/0005772X.2014.11417585.

[231] M. H. Struye. *BeeSCAN Handbuch*. 1994.

[232] M. H. Struye. 'Possibilities and limitations of monitoring the flight activity of honeybees by means of BeeSCAN bee counters'. In: *COLLOQUES-INRA* (2001), pp. 269–278.

[233] M. H. Struye, H. J. Mortier, G. Arnold, C. Miniggio, and R. Borneck. 'Microprocessor-controlled monitoring of honeybee flight activity at the hive entrance'. In: *Apidologie* 25.4 (1994), pp. 384–395. ISSN: 0044-8435. DOI: 10.1051/apido:19940405.

[234] Agata Swiatly-Blaszkiewicz, Dagmara Pietkiewicz, Jan Matysiak, Barbara Czech-Szczapa, Katarzyna Cichocka, and Bogumiła Kupcewicz. 'Rapid and Accurate Approach for Honeybee Pollen Analysis Using ED-XRF and FTIR Spectroscopy'. In: *Molecules* 26.19 (2021), p. 6024. ISSN: 1420-3049. DOI: 10.3390/molecules26196024.

[235] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. 'DeepFace: Closing the Gap to Human-Level Performance in Face Verification'. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014 IEEE Conference on Computer Vision and Pattern Recognition. ISSN: 1063-6919. 2014, pp. 1701–1708. DOI: 10.1109/CVPR.2014.220.

[236] Rahman Tashakkori and Ahmad Ghadiri. 'Image processing for honey bee hive health monitoring'. In: *SoutheastCon 2015*. SoutheastCon 2015. Fort Lauderdale, FL, USA: IEEE, 2015, pp. 1–7. ISBN: 978-1-4673-7300-5. DOI: 10.1109/SECON.2015.7133029.

[237] Frederic Tausch. 'Visual Re-Identification of Bumblebees Returning from Foraging Trips'. Bachelor's Thesis. Karlsruhe: Karlsruhe Institute of Technology, 2020.

[238] Frederic Tausch, Katharina Schmidt, and Matthias Diehl. 'Current achievements and future developments of a novel AI based visual monitoring of beehives in ecotoxicology and for the monitoring of landscape structures'. In: *Hazards of pesticides to bees : 14th International Symposium of the ICP-PR Bee Protection Group, October 23 - 25, 2019 Bern, Switzerland - Proceedings* - (2020). Conference Name: 14th International Symposium of the ICP-PR Bee Protection Group; Bern, Switzerland; 2019.10.23-25 ISBN: 9783955470951, pp. 124–129. DOI: 10.5073/jka.2020.465.064.

[239] Frederic Tausch, Simon Stock, Julian Fricke, and Olaf Klein. 'Bumblebee Re-Identification Dataset'. In: 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW). Snowmass Village, CO, USA: IEEE, 2020, pp. 35–37. ISBN: 978-1-72817-162-3. DOI: 10.1109/WACVW50321.2020.9096909.

[240] Frederic Tausch, Jan Wagner, and Simon Klaus. 'Pollinators as Data Collectors: Estimating Floral Diversity with Bees and Computer Vision'. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023, pp. 643–650.

[241] Kirsten S. Traynor, Fanny Mondet, Joachim R. de Miranda, Maeva Techer, Vienna Kowallik, Melissa A. Y. Oddie, Panuwan Chantawannakul, and Alison McAfee. '*Varroa destructor*: A Complex Parasite, Crippling Honey Bees Worldwide'. In: *Trends in Parasitology* 36.7 (2020), pp. 592–606. ISSN: 1471-4922. DOI: 10.1016/j.pt.2020.04.004.

[242] Jérôme Trouiller. 'Monitoring Varroa jacobsoni resistance to pyrethroids in western Europe'. In: *Apidologie* 29.6 (1998), pp. 537–546. ISSN: 0044-8435. DOI: 10.1051/apido:19980606.

[243] Jerome Trouiller, Gerard Arnold, Bertrand Chappe, Yves Le Conte, and Claudine Masson. 'Semiochemical basis of infestation of honey bee brood byVarroa jacobsoni'. In: *Journal of Chemical Ecology* 18.11 (1992), pp. 2041–2053. ISSN: 1573-1561. DOI: 10.1007/BF00981926.

[244] Gang Jun Tu, Mikkel Kragh Hansen, Per Kryger, and Peter Ahrendt. 'Automatic behaviour analysis system for honeybees using computer vision'. In: *Computers and Electronics in Agriculture* 122 (2016), pp. 10–18. ISSN: 01681699. DOI: 10.1016/j.compag.2016.01.011.

[245] Alexandra Valentine and Stephen J. Martin. 'A survey of UK beekeeper's Varroa treatment habits'. In: *PloS One* 18.2 (2023), e0281130. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0281130.

[246] Jozef Van Der Steen and Flemming Vejsnæs. 'Varroa Control: A Brief Overview of Available Methods'. In: *Bee World* 98.2 (2021), pp. 50–56. ISSN: 0005-772X, 2376-7618. DOI: 10.1080/0005772X.2021.1896196.

[247] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. 'The inaturalist species classification and detection dataset'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 8769–8778.

[248] Caroline Vilarem, Vincent Piou, Fanny Vogelweith, and Angélique Vétillard. 'Varroa destructor from the Laboratory to the Field: Control, Biocontrol and IPM Perspectives—A Review'. In: *Insects* 12.9 (2021), p. 800. ISSN: 2075-4450. DOI: 10.3390/insects12090800.

[249] P. Kirk Visscher and Thomas D. Seeley. 'Foraging Strategy of Honeybee Colonies in a Temperate Deciduous Forest'. In: *Ecology* 63.6 (1982), p. 1790. ISSN: 00129658. DOI: 10.2307/1940121.

[250] Werner Von Der Ohe, Livia Persano Oddo, Maria Lucia Piana, Monique Morlot, and Peter Martin. 'Harmonized methods of melissopalynology'. In: *Apidologie* 35 (Suppl. 1 2004), S18–S25. ISSN: 0044-8435, 1297-9678. DOI: 10.1051/apido:2004050.

[251] George Voudiotis, Anna Moraiti, and Sotirios Kontogiannis. 'Deep Learning Beehive Monitoring System for Early Detection of the Varroa Mite'. In: *Signals* 3.3 (2022), pp. 506–523. ISSN: 2624-6120. DOI: 10.3390/signals3030030.

[252] W. Wang, R. A. Solovyev, A. L. Stempkovsky, D. V. Telpukhov, and A. A. Volkov. 'Method for Whale Re-identification Based on Siamese Nets and Adversarial Training'. In: *Optical Memory and Neural Networks* 29.2 (2020), pp. 118–132. ISSN: 1060-992X, 1934-7898. DOI: 10.3103/S1060992X20020058.

[253] Fernando Wario, Benjamin Wild, Margaret J. Couvillon, Raúl Rojas, and Tim Landgraf. 'Automatic methods for long-term tracking and the detection and decoding of communication dances in honeybees'. In: *Frontiers in Ecology and Evolution* 3 (2015). ISSN: 2296-701X. DOI: 10.3389/fevo.2015.00103.

[254] Kilian Q Weinberger, John Blitzer, and Lawrence K Saul. 'Distance Metric Learning for Large Margin Nearest Neighbor Classification'. In: *Advances in neural information processing systems* 18 (2005).

[255] K. Weiss. 'Experiences with Plastic Combs and Foundation'. In: *Bee World* (1983). Publisher: Taylor & Francis. ISSN: 0005-772X.

[256] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. 'Deep learning for person re-identification: A survey and outlook'. In: *IEEE transactions on pattern analysis and machine intelligence* 44.6 (2021). Publisher: IEEE, pp. 2872–2893.

[257] A. Zacepins, E. Stalidzans, and J. Meitalovs. 'Application of information technologies in precision apiculture'. In: *Proceedings of the 13th International Conference on Precision Agriculture (ICPA 2012)*. 2012.

[258] Aleksejs Zacepins, Valters Brusbardis, Jurijs Meitalovs, and Egils Stalidzans. 'Challenges in the development of Precision Beekeeping'. In: *Biosystems Engineering* 130 (2015), pp. 60–71. ISSN: 1537-5110. DOI: 10.1016/j.biosystemseng.2014.12.001.

[259] Romée van der Zee, Alison Gray, Lennard Pisa, and Theo de Rijk. 'An Observational Study of Honey Bee Colony Winter Losses and Their Association with Varroa destructor, Neonicotinoids and Other Risk Factors'. In: *PLOS ONE* 10.7 (2015). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0131611.

[260] Matthew D. Zeiler, Graham W. Taylor, and Rob Fergus. 'Adaptive deconvolutional networks for mid and high level feature learning'. In: *2011 International Conference on Computer Vision*. 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011, pp. 2018–2025. ISBN: 978-1-4577-1102-2 978-1-4577-1101-5 978-1-4577-1100-8. DOI: 10.1109/ICCV.2011.6126474.

[261] Bob Zhang, Wei Nie, and Shuping Zhao. 'A novel Color Rendition Chart for digital tongue image calibration'. In: *Color Research & Application* 43.5 (2018), pp. 749–759. ISSN: 1520-6378. DOI: 10.1002/col.22234.

[262] Yu-Jin Zhang. *3D Computer Vision: Foundations and Advanced Methodologies*. Singapore: Springer Nature Singapore, 2024. ISBN: 978-981-19760-2-5 978-981-19760-3-2. DOI: 10.1007/978-981-19-7603-2.

[263] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 'Scalable person re-identification: A benchmark'. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1116–1124.

[264] Bettina Ziegelmann, Anne Lindenmayer, Johannes Steidle, and Peter Rosenkranz. 'The mating behavior of Varroa destructor is triggered by a female sex pheromone'. In: *Apidologie* 44.3 (2013), pp. 314–323. ISSN: 1297-9678. DOI: 10.1007/s13592-012-0182-5.

Declaration of AI-Assisted Technologies in the Writing Process

During the preparation of this work, the author used DeepL (DeepL SE, Cologne, Germany) and ChatGPT-3.5 & 4.0 (OpenAI, San Francisco, CA, USA) in order to translate and refine text and to structure content. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published work. The quote in Section 1 was generated by the named language model.