Cross-Domain Fine-Grained Classification: A Review

Stefan Wolf

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
stefan.wolf@kit.edu

Abstract

Fine-grained classification is an interesting but challenging task due to the high amount of data needed to achieve a high accuracy. However, the high specificity of the classes makes it difficult to collect a large amount of samples. Thus, the use of cross-domain learning is an interesting aspect since an abundant amount of data exists for some domains like web images exists. In this review, the current works of cross-domain fine-grained classification are summarized and potential areas for future work are highlighted. Even though first works exist, the variety of methods is still small and interesting cross-domain settings are rarely considered. Thus, the field of cross-domain fine-grained classification provides a large room for future research.

1 Introduction

Fine-grained image classification is a task which has gained attention in recent years due to the application of convolutional neural networks achieving promising results. Another reason for the gained attention is that the applications of fine-grained classification are manifold. Gebru et al. [12] predict the model of vehicles in order to visually estimate the income of regions in the US. Similar approaches could be applied on parking areas of supermarkets to estimate the income of customers. Usuyama et al. [30] apply fine-grained image classification to visually identify pills in order to reduce the risk of medication errors.

Compared to regular image classification, fine-grained classification induces a lower inter-class variance with often only small details distinguishing classes while view and appearance of images can be highly different within classes resulting in a high intra-class variance. Thus, it is a more difficult task than coarse-grained image classification. Additionally, the high specificity of classes limits the availability of images and increases the required knowledge of annotators. Another aspect making the application of fine-grained classification difficult is that the classes of datasets are usually very specific to a certain region and a certain timeframe. For example, the distribution of cars differs heavily for different regions like Europe and Asia and new car models are constantly introduced rendering available datasets quickly outdated.

However, this can be compensated by crawling images from the web or creating synthetic images by rendering 3D models of cars or pills. While these approaches can create a large amount of labeled data, they often induce a large domain gap since images from the web are usually more polished than images in real-world applications like surveillance and synthetic images do not reach the realism of camera images. However, domain adaptation can support the learning process to enable the use of data from different domains than the targeted domains. Since these image sources are also useful for normal image classification, a broad range of literature is dedicated towards domain adaption for coarse-grained classification as summarized by Wang and Deng [32].

However, cross-domain fine-grained classification introduces new challenges like the small inter-class variance compared to the large inter-domain variance which requires careful consideration during adaptation [41]. Moreover, the high specificity of fine-grained classes brings up cross-domain scenarios which are uncommon for regular image classification like a supervised partially zero-shot scenario. In this case, some classes in the target domain do not have images at all while all other classes have abundant images with annotations available in the target domain [30]. In contrast, classic domain adaptation usually assumes the availability of all classes in the target domain, even though all data in the target domain is unlabeled [9].

Since existing surveys about fine-grained classification [35, 14] or cross-domain classification [32] do not consider the works combining both fields, this survey gives an overview over the existing works about cross-domain fine-grained classification.

The remainder of this work is structured as follows: section 2 summarizes the literature targeting fine-grained classification while section 3 shortly describes existing works regarding cross-domain classification. Section 4 gives a more detailed view on the intersection of both research areas combining cross-domain and fine-grained classification. Section 5 summarizes the content of this work and highlights research aspects which are interesting for future work.

2 Fine-Grained Classification

Since fine-grained classification shares many of its challenges with usual image classification as mostly investigated on the ImageNet [6] dataset, current deep learning models proven on ImageNet have been established as a good starting point for fine-grained classification [31]. However, the high intra-class variance compared to the low inter-class variance of fine-grained classification tasks motivate the exploration of adaptations. In this regard, multiple authors have found ways to improve the performance of deep-learning models when they are exposed to fine-grained classification tasks. The development branches can be roughly categorized as part-based models, bilinear CNNs, multi-task learning, hierarchical classification, metric learning, temporal classification and webly-supervised approaches.

Part-based models. While localizing discriminating parts has not been widely adopted since CNNs are used for classification, fine-grained classification is an area where part-based classification can still be advantageous. A reason is that the distinguishing regions might not be automatically determined during training because of the datasets being too small for the problem at hand. Thus, multiple authors have approached the integration of part-based classification schemes with CNNs to improve the accuracy [37, 8]. Huang et al. [15] use a part-based model to provide an interpretation of the classification to the user. However, hand-engineered part models suffer from being not optimal for classification and requiring a high annotation effort. Thus, Simon and Rodner [25] propose an unsupervised part-based model using feature map activations to find discriminating parts.

Bilinear CNNs. Following the idea of part-based models that localization of discriminating parts and creating features should be separated, Lin, RoyChowdhury, and Maji [22] propose a two-stream architecture consisting of two CNNs that combines the final feature maps of both networks with a bilinear module. The bilinear module calculates the outer product of both feature vectors for each pixel of the feature map. The expectation of the authors is that one network locates discriminating parts while the other network extracts discriminating features. Due to the high computational demands of the high dimensional outer product of both feature maps, adaptations have been proposed to reduce its dimension [10, 19]. Yu et al. [40] extend the bilinear CNN scheme by applying cross-layer bilinear pooling, i.e., they pool bilinear features between layers of the network instead of only pooling bilinear features after the last layer.

Multi-task learning. In multi-task learning, an auxiliary task is additionally solved to the main task with the auxiliary task being related to the main task. Due to the relation, it is expected that the auxiliary task supports solving the main task. Depending on the method, the solving of the auxiliary task might be either limited to the training process [3] or might also be performed during inference [26, 23]. To enhance the quality of fine-grained class predictions, Sochor, Herout, and Havel [26] feed automatically extracted 3D bounding boxes of the cars as additional information to the network in order to support the network regularizing the perspective. Chen, Liu, and Yu [3] propose a similar approach by predicting the viewpoint of the image as an auxiliary task. Providing knowledge about the viewpoint during training supports the network in coping with the large intra-class variance due to the high variation regarding viewpoints. Lin et al. [23] propose an approach that fits a 3D model on the image, exploits the localization of object parts to extract features at constant locations, and uses theses features to perform a classification. Due to repetitively applying this scheme, the correct 3D model for the vehicle model is chosen from a database resulting in a higher classification accuracy.

Hierarchical classification. Fine-grained classes are usually part of a more complex class hierarchy. For example, while fine-grained bird classification is commonly done on the level of species, each species is part of a certain genus which is part of a certain family of birds. Cars are often classified on the level of the year a certain iteration of a model was presented. However, this is part of a hierarchy containing the model and the manufacturer. In the case of cars, a second hierarchy can be built by assigning each model to a certain type of car like van or sedan. Hierarchical classification can be seen as a special form of multi-task learning since the classification of coarse-grained categories is used as an auxiliary task to improve the accuracy of the fine-grained classification. Huo et al. [16] exploit these hierarchies by training multiple layers of the hierarchy in a round-robin manner. Buzzelli and Segantin [2] train cascaded classifiers to reduce the number of classes per classifier.

Metric learning. CNNs for classification usually apply a linear layer combined with a softmax activation on the last feature layer to generate the output probability distribution. During training, the backpropagation algorithm is applied which mostly finds an adequate embedding for the final feature layer that tends to minimize intra-class variance and maximize inter-class variance. However, for fine-grained classification tasks, an explicit loss formulation that increases distance between classes and decreases the distance between features of the same class is commonly applied as alternative (using a kNN-classifier for inference) or additional to a softmax-based classifier [27, 18, 39]. Particularly for a high amount of classes as common in fine-grained classification, metric learning proved advantageous [39].

Temporal classification. Object classification is mostly performed on still images. While a single image is usually sufficient for coarse-grained classification, the discriminating parts for fine-grained classification are often only visible in specific perspectives. Thus, Zhu et al. [44] and Alsahafi et al. [1]

investigated fine-grained classification on videos. Zhu et al. [44] regard the redundant information in videos as the main challenge and propose an approach that combines the feature maps from multiple images and processes the feature maps in an iterative manner. In each step, redundant information from previous steps is suppressed while discriminating parts are attended to. Alsahafi et al. [1] use an object detector to extract accurate crops of the vehicle for each image and combine the per-image classification results by averaging.

Webly-supervised. The amount of samples per class is scarce for fine-grained datasets compared to coarse-grained datasets like ImageNet [6] due to the specificity of classes and the difficulty of labeling. Therefore, multiple authors use community provided image collections in the web which have labels available like, e.g., Flickr. With the class name as query large amounts of data can be gathered [36, 7].

Other works. Some works have investigated methods which have not yet brought up a new branch of research or they consider aspects uncommonly explored. Touvron et al. [28] propose a method called Grafit that enables fine-grained classification based on a training dataset only containing coarse-grained labels. This is achieved by combining an instance loss and a kNN loss in order to learn a fine-grained feature embedding. Cui et al. [5] propose a new training scheme for long-tailed class distributions with a small number of frequent classes and a large number of rare classes. Moreover, the authors propose a domain similarity metric to find a good dataset for pre-training a network prior to training it on the target dataset. Zhang et al. [42] follow an ensemble strategy by training multiple expert networks with the maximization of the Kullback-Leibler divergence between the output probability distributions of each classifier as additional optimization target.

3 Cross-Domain Classification

In a cross-domain classification setting, at least two domains are involved called source and target with the evaluation being performed on data of the target domain. While abundant data is available for the source domain, the availability of data in the target domain is limited in some form. Mostly the limitation is in form of missing labels. In this case, the adaptation for the target domain is called unsupervised domain adaptation. More settings in the context of fine-grained classification are described in section 4. Formally, cross-domain classification can be described with two sets of images X and labels Y called (X_s, Y_s) and (X_t, Y_t) for source and target, respectively. In a domain adaptation setting the distributions P of the image samples between both domains differ: $\mathrm{P}(X_s) \neq \mathrm{P}(X_t)$. However, the classification task is kept the same: $\mathrm{P}(Y_s|X_s) = \mathrm{P}(Y_t|X_t)$. Following the taxonomy of Wang and Deng [32], methods for domain adaptation can be categorized in discrepancy-based, adversarial-based, and reconstruction-based methods.

Discrepancy-based domain adaptation. Methods based on discrepancy use a certain criterion for fine-tuning a deep learning model to optimize it for the target domain. One type of criteria are class criteria which base the fine-tuning process on class labels [29]. Pseudo labels can be generated if no labels are available in the target domain [43]. Methods that use a statistic criterion minimize the distance between the statistical distributions of both domains, e.g., with a Kullback-Leibler divergence [46]. An architectural criterion optimizes the architecture of deep learning models to generate more domain-invariant features. Such an architectural improvement is adaptive batch normalization [21]. A geometric criterion is another type which has been used for aligning domains [4].

Adversarial-based domain adaptation. Adversarial approaches try to confuse the main network regarding the domain. They can be categorized in generative and non-generative methods. Generative methods generate a transformed input sample that contains the same content as the source sample with an appearance matching target samples [24]. Non-generative approaches reduce the domain gap in the feature space. A domain classifier is added and connected to the network via a gradient reversal layer that leads to the features being adjusted towards values most unsuitable for domain classification [9].

Reconstruction-based domain adaptation. The aim of these approaches is the reconstruction of samples from the source or target domain with the aim to achieve a domain-invariant representation of samples. The application of a combination of an encoder and a decoder with the decoder trying to reconstruct the sample from the features produced by the encoder is one approach [13]. Another approach is to use adversarial networks in the form of a Cycle-GAN that transforms a sample from one domain to the other and afterwards, reconstructs the original sample from the transformed sample [45].

4 Cross-Domain Fine-Grained Classification

In this section, an overview of existing works regarding the intersection of fine-grained classification and cross-domain classification is given. Cross-domain fine-grained classification is significantly more difficult than either of the tasks of cross-domain classification or fine-grained classification since domain adaptation and fine-grained classification are contradictive in terms of feature adjustment. While fine-grained classification requires the features to capture fine details in the image to cope with the low inter-class variance, domain adaptation is drastically changing the features in order to reduce the high inter-domain variance [33, 41].

The works are categorized by the domain adaptation setting type they apply, i.e., unsupervised, semi-supervised or supervised partially zero-shot. The different types are visualized in Figure 4.1 and an overview of the different approaches is given in Table 4.1.

Unsupervised domain adaptation. The task of unsupervised domain adaptation assumes a source domain with a large labeled dataset and a target domain with a large unlabeled dataset while both datasets include samples for all categories. The first to explore such a setting with fine-grained categories are Gebru, Hoffman, and Fei-Fei [11]. The authors exploit auxiliary attributes commonly available in fine-grained datasets by adding an additional classification head per attribute and applying an attribute consistency loss forcing the predicted attributes to match the main classification category, e.g., the body type like

Authors	Setting	Domains
Gebru et al. [11]	Unsupervised Semi-supervised	Vehicles: marketing shots, GSV
Wang et al. [33]	Unsupervised	Vehicles: marketing shots, GSV
Wang et al. [34]	Unsupervised Semi-supervised	Vehicles: marketing shots, GSV Retail products: studio images, supermarket shelves, web images
Yu, Jiang, and Li [41]	Unsupervised	Vehicles: marketing shots, GSV
Li et al. [20]	Semi-supervised	Vehicles: marketing shots, GSV
Usuyama et al. [30]	Supervised partially zero-shot	Pills: reference images, consumer images

Table 4.1: Overview of cross-domain fine-grained methods with the cross-domain setting considered and the domains between a domain adaptation was investigated. GSV stands for Google Street View.

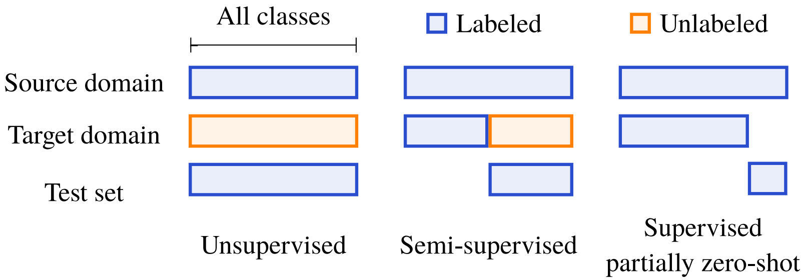

Figure 4.1: Types of settings in cross-domain fine-grained classification. The bars illustrate the classes. In an unsupervised setting, a large amount of unlabeled target samples is available for all classes. In contrast to an unsupervised setting, a part of the classes have labeled target samples available in a semi-supervised setting. The evaluation is only performed on the part of classes which have no labeled samples. In a supervised partially zero-shot setting, for a small part of the classes no target samples are available at all while this part is used for evaluation.

sedan or van matching the concrete model for vehicle classification. Since the number of training samples per attribute category is higher than per main category, attribute prediction is more stable across domains which results in a more accurate prediction of the main category due to the attribute consistency loss. The use of auxiliary attributes is additionally applied to a domain confusion loss as proposed by Tzeng et al. [29]. The approach is evaluated for vehicle classification with an adaptation from web-scraped marketing shots to Google Street View (GSV) images. Wang et al. [33] propose a quite similar approach that uses coarse-grained labels to initially train the network on an easier task that has a higher inter-class variance and is less prone to features being deteriorated during domain adaptation. The distribution of coarse-grained labels is extended to the dimension of the fine-grained labels enabling a progressive adaptation from coarse-grained to fine-grained training based on curriculum learning while using adversarial adaptation [9] to simultaneously adjust the features to the target domain. The authors also evaluate their approach on the previously mentioned vehicle classification setting with an adaptation from marketing shots to GSV images. The approach proposed by Wang et al. [34] also employs adversarial domain alignment to reduce the domain gap in the feature space. Additionally, a self-attention module is proposed that identifies class-discriminating regions and applies a part-wise classification with a result fusion step. Additional to also evaluating the marketing shots to GSV images adaptation scenario with vehicle classification, the authors propose a new setting with fine-grained classification of retail products in three domains, i.e., professional studio images, images of supermarket shelves and web images. Yu, Jiang, and Li [41] propose a method targeting fine-grained domain adaptation by maintaining quality of class-separating features during the adaptation process. This is achieved by employing domain-specific class labels with the domains being swapped after a pre-training phase resulting in domain confusion while keeping the class-separating characteristics of the features intact. Again, the vehicle classification setting adaption from marketing shots to GSV images is evaluated.

Semi-supervised domain adaptation. In the setting of semi-supervised domain adaptation, a part of the samples from the target domain are labeled. In the case of fine-grained classification, the subsets of labeled and unlabeled samples are split by classes [11]. Gebru, Hoffman, and Fei-Fei [11] and Wang et al. [34] also evaluate their approaches explained above in a semi-supervised setting. While Gebru, Hoffman, and Fei-Fei [11] employ a cross entropy loss for the labeled samples in the target domain, Wang et al. [34] propose a contrastive loss for category-level alignment using the labeled target samples. Li et al. [20] propose the integration of a residual correction block before the final classification layer which is trained to minimize the difference between the distributions of features of the source and target domain. The decision to incorporate the residual correction block is based on the insight that early features are domain and task invariant compared to late features [38]. The authors evaluate their method in a setting adapting fine-grained vehicle classification from marketing shots to GSV images.

Supervised partially zero-shot domain adaptation. Usuyama et al. [30] are the first to propose a fine-grained domain adaptation setting different to unsupervised or semi-supervised, i.e., compared to a semi-supervised scenario, they prohibit the use of the unlabeled samples. Thus, a part of the classes has no samples available in the target domain at all making the setting more difficult. This scenario has a high practical importance since the high specificity of fine-grained classes makes it difficult to collect samples for certain classes or to ensure that a certain class is in a set of samples that has been randomly collected. The evaluation in this scenario is done on the classes that have no samples in the target domain. Such a setting is called supervised partially zero-shot following Ishii, Takenouchi, and Sugiyama [17]. Usuyama et al. [30] propose a new fine-grained dataset called ePillID containing images of pills in two domains, i.e., reference images with a specified viewpoint and lighting and a masked background and consumer images which have a greater intra-class variance. The authors evaluate a baseline approach using metric learning in a cross-domain setting.

5 Conclusion

In this work, a review of fine-grained, cross-domain, and cross-domain fine-grained classification was given. Cross-domain learning is particularly interesting for fine-grained classification due to the specificity of fine-grained classes making it highly difficult to collect abundant data for all classes. However, the challenges of fine-grained classification, i.e., a high intra-class variance and a low inter-class variance, exacerbate domain adaptation which additionally has to cope with a high inter-domain variance. Multiple approaches have been proposed to address these problems with traditional domain adaptation methods like a domain confusion loss or adversarial learning as starting point. The additional learning of auxiliary attributes or coarse-grained labels has shown to be advantageous in cross-domain scenarios and is a promising prospect for future research. The three distinguished settings for cross-domain fine-grained classification have a major impact on the applicability of the approaches. Thus, all three settings should be investigated in order to give practitioners a broad range of available approaches to solve problems with the data at hand. Particularly, the supervised partially zero-shot setting has not yet been widely explored while in practice a guarantee that images are available for all classes in the target domain is hard to provide because of the high specificity of fine-grained classes. Another interesting area for future research might be the use of synthetic data to enlarge the dataset.

References

[1] Yousef Alsahafi et al. “CarVideos: A Novel Dataset for Fine-Grained Car Classification in Videos”. In: *16th International Conference on Information Technology-New Generations (ITNG 2019)*. 2019.

[2] Marco Buzzelli and Luca Segantin. “Revisiting the CompCars Dataset for Hierarchical Car Classification: New Annotations, Experiments, and Results”. In: *Sensors* 21.2 (2021).

[3] Qianqiu Chen, Wei Liu, and Xiaoxia Yu. “A Viewpoint Aware Multi-Task Learning Framework for Fine-Grained Vehicle Recognition”. In: *IEEE Access* 8 (2020), pp. 171912–171923.

[4] Sumit Chopra, Suhrid Balakrishnan, and Raghuraman Gopalan. “Dlid: Deep learning for domain adaptation by interpolating between domains”. In: *ICML workshop on challenges in representation learning*. Vol. 2. 6. 2013.

[5] Yin Cui et al. “Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2018.

[6] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009.

[7] Haodong Duan et al. “Omni-Sourced Webly-Supervised Learning for Video Recognition”. In: *Computer Vision – ECCV 2020*. 2020.

[8] Jie Fang et al. “Fine-Grained Vehicle Model Recognition Using A Coarse-to-Fine Convolutional Neural Network Architecture”. In: *IEEE Transactions on Intelligent Transportation Systems* 18.7 (2017), pp. 1782–1792.

[9] Yaroslav Ganin et al. “Domain-Adversarial Training of Neural Networks”. In: *Journal of Machine Learning Research* 17.59 (2016), pp. 1–35.

[10] Yang Gao et al. “Compact Bilinear Pooling”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[11] Timnit Gebru, Judy Hoffman, and Li Fei-Fei. “Fine-Grained Recognition in the Wild: A Multi-Task Domain Adaptation Approach”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Oct. 2017.

[12] Timnit Gebru et al. “Fine-Grained Car Detection for Visual Census Estimation”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 31.1 (Feb. 2017).

[13] Muhammad Ghifary et al. “Domain Generalization for Object Recognition With Multi-Task Autoencoders”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Dec. 2015.

[14] Chun-feng GUO et al. “A Survey of Fine-Grained Image Classification Based on Deep Learning”. In: *DEStech Transactions on Computer Science and Engineering* ica (2019).

[15] Shaoli Huang et al. “Part-Stacked CNN for Fine-Grained Visual Categorization”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[16] Yuqi Huo et al. “Coarse-to-Fine Grained Classification”. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR’19. 2019.

[17] Masato Ishii, Takashi Takenouchi, and Masashi Sugiyama. “Partially Zero-shot Domain Adaptation from Incomplete Target Data with Missing Classes”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Mar. 2020.

[18] Alper Kayabasi, Kaan Karaman, and Ibrahim Batuhan Akkaya. “Comparison of distance metric learning methods against label noise for fine-grained recognition”. In: *Automatic Target Recognition XXXI*. Vol. 11729. 2021.

[19] Shu Kong and Charless Fowlkes. “Low-Rank Bilinear Pooling for Fine-Grained Classification”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017.

[20] Shuang Li et al. “Deep Residual Correction Network for Partial Domain Adaptation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.7 (2021), pp. 2329–2344.

[21] Yanghao Li et al. “Adaptive Batch Normalization for practical domain adaptation”. In: *Pattern Recognition* 80 (2018), pp. 109–117.

[22] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. “Bilinear CNN Models for Fine-Grained Visual Recognition”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Dec. 2015.

[23] Yen-Liang Lin et al. “Jointly Optimizing 3D Model Fitting and Fine-Grained Classification”. In: *Computer Vision – ECCV 2014*. 2014.

[24] Ming-Yu Liu and Oncel Tuzel. “Coupled Generative Adversarial Networks”. In: *Advances in Neural Information Processing Systems*. Vol. 29. 2016.

[25] Marcel Simon and Erik Rodner. “Neural Activation Constellations: Unsupervised Part Model Discovery With Convolutional Networks”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Dec. 2015.

[26] Jakub Sochor, Adam Herout, and Jiri Havel. “BoxCars: 3D Boxes as CNN Input for Improved Fine-Grained Vehicle Recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016.

[27] Kihyuk Sohn. “Improved Deep Metric Learning with Multi-class N-pair Loss Objective”. In: *Advances in Neural Information Processing Systems*. Vol. 29. 2016.

[28] Hugo Touvron et al. “Grafit: Learning Fine-Grained Image Representations With Coarse Labels”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2021.

[29] Eric Tzeng et al. “Simultaneous Deep Transfer Across Domains and Tasks”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Dec. 2015.

[30] Naoto Usuyama et al. “ePillID Dataset: A Low-Shot Fine-Grained Benchmark for Pill Identification”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2020.

[31] Krassimir Valev et al. “A systematic evaluation of recent deep learning architectures for fine-grained vehicle classification”. In: *Pattern Recognition and Tracking XXIX*. Vol. 10649. 2018.

[32] Mei Wang and Weihong Deng. “Deep visual domain adaptation: A survey”. In: *Neurocomputing* 312 (2018), pp. 135–153.

[33] Sinan Wang et al. “Progressive Adversarial Networks for Fine-Grained Domain Adaptation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2020.

[34] Yimu Wang et al. “An Adversarial Domain Adaptation Network for Cross-Domain Fine-Grained Recognition”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Mar. 2020.

[35] Xiu-Shen Wei et al. “Fine-Grained Image Analysis with Deep Learning: A Survey”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[36] Zhe Xu et al. “Webly-Supervised Fine-Grained Visual Categorization via Deep Domain Adaptation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.5 (2018), pp. 1100–1113.

[37] Hantao Yao et al. “Coarse-to-Fine Description for Fine-Grained Visual Categorization”. In: *IEEE Transactions on Image Processing* 25.10 (2016), pp. 4858–4872.

[38] Jason Yosinski et al. “How transferable are features in deep neural networks?” In: *Advances in Neural Information Processing Systems*. Vol. 27. 2014.

[39] Baosheng Yu et al. “Correcting the Triplet Selection Bias for Triplet Loss”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Sept. 2018.

[40] Chaojian Yu et al. “Hierarchical Bilinear Pooling for Fine-Grained Visual Recognition”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Sept. 2018.

[41] Han Yu, Rong Jiang, and Aiping Li. “Striking a Balance in Unsupervised Fine-Grained Domain Adaptation Using Adversarial Learning”. In: *Knowledge Science, Engineering and Management*. 2020.

[42] Lianbo Zhang et al. “Learning a Mixture of Granularity-Specific Experts for Fine-Grained Categorization”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2019.

[43] Xu Zhang et al. *Deep Transfer Network: Unsupervised Domain Adaptation*. arXiv:1503.00591 [cs.CV]. Mar. 2015.

[44] Chen Zhu et al. “Fine-grained Video Categorization with Redundancy Reduction Attention”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Sept. 2018.

[45] Jun-Yan Zhu et al. “Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks”. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Oct. 2017.

[46] Fuzhen Zhuang et al. “Supervised Representation Learning: Transfer Learning with Deep Autoencoders”. In: *Proceedings of the 24th International Conference on Artificial Intelligence*. IJCAI’15. 2015.