AI and multi-spectral imaging:

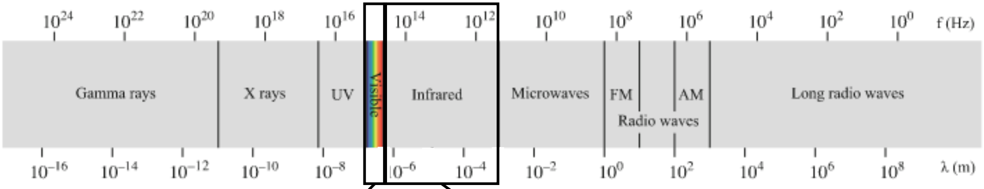
Implementing a deep learning model for the segmentation of common thermal urban features to assist in the automation of infrastructure-related maintenance

Elena Vollmer, Leon Klug, Rebekka Volk, Frank Schultmann
Institute for Industrial Production (IIP), Karlsruhe Institute of Technology (KIT)

Motivation

Figure: Electromagnetic spectrum [Gade et al. (2014)]

← Increasing frequency (f)



Increasing wavelength (λ) →

How can we easily process and analyze an RGB-T and UAV-based dataset with a common deep learning model for (heat-related) urban infrastructure maintenance?

RGB = Red Green Blue = visible
T = Thermal
UAV = Unmanned Aerial Vehicle

Case Study

- Support of energy supply system monitoring: District heating networks

leakage in underground pipe ⟶ heated medium escapes ⟶ surroundings & surface experience ↑T ⟶ anomalous hotspot in thermal images

- Identify common (thermal) features in urban settings
  - Classify false alarms while searching for leakages

⟶ Multi-class semantic segmentation problem

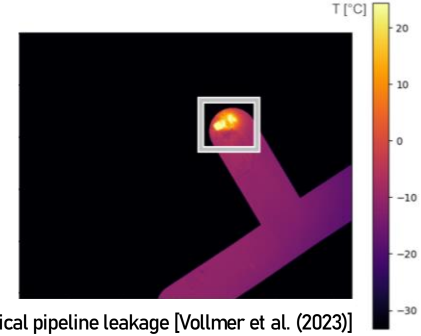


Figure A: Critical pipeline leakage [Vollmer et al. (2023)]

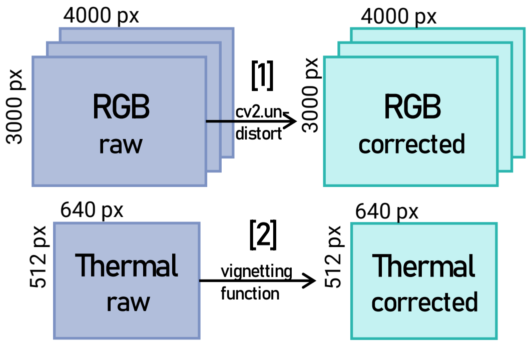
Figure B: Utilized UAV and Controller System

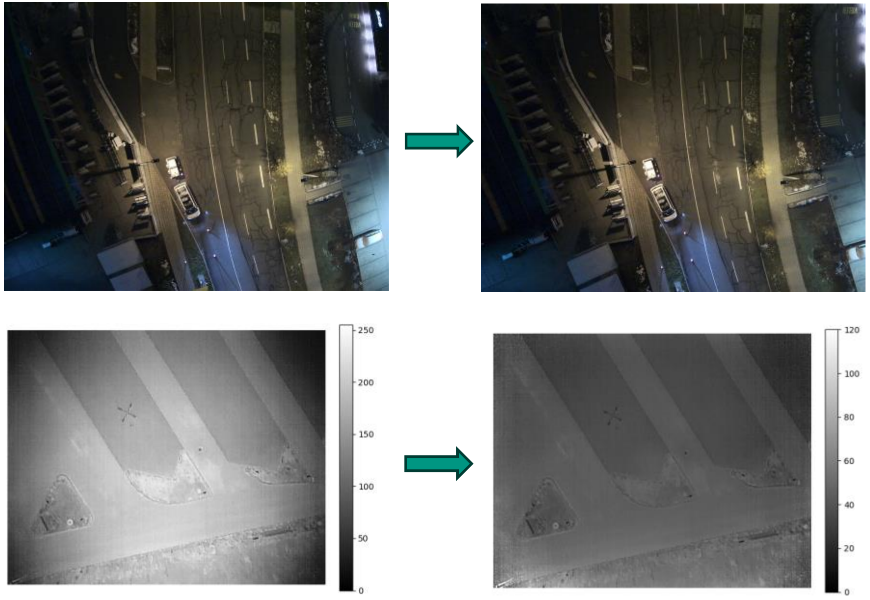
Data:

- 793 images from two urban areas (Munich & Karlsruhe, Germany)
  - Dual camera[1]: RGB + TIR
  - UAV[2]-based: 90° pitch (facing down), 60m flight height
  - Night-time flights

[1] Zenmuse XT2 camera with a FLIR Tau 2 thermal sensor
[2] DJI M600 and DJI M300 UAVs

Case Study: Data Processing and Training Pipeline



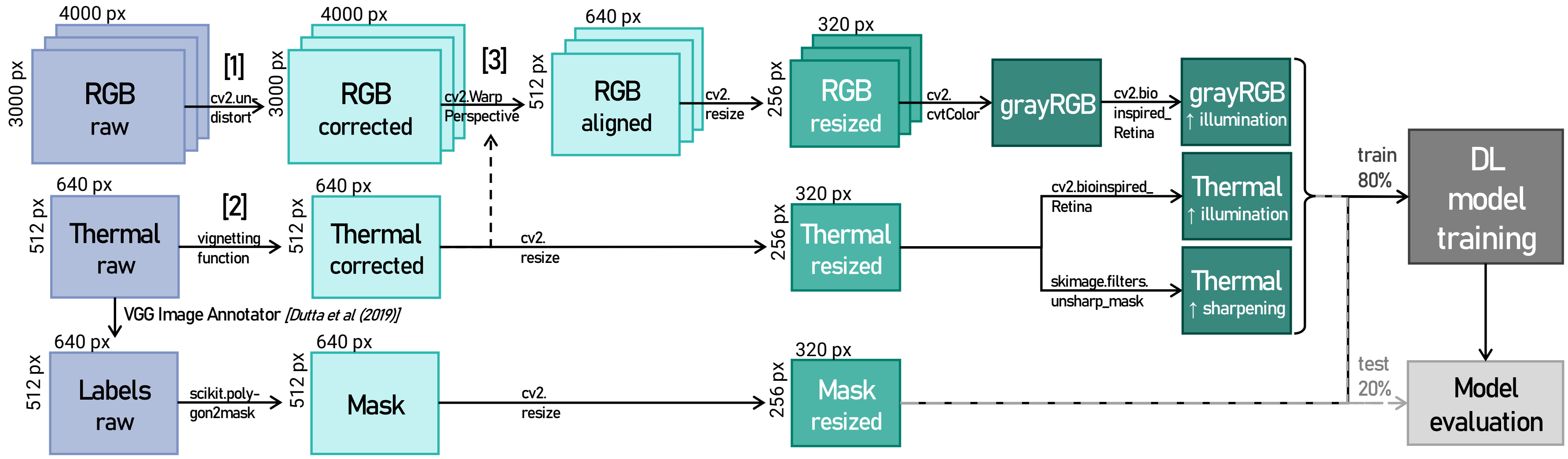


[1] Fish-eye distortion removal with camera calibration *[Hou et al. 2021, Mayer et al. 2023a]*

[2] Vignetting effect removal with approximated radial polynomial function *[Bal et al. 2023]*

Image sources: IIP, KIT

Case Study: Data Processing and Training Pipeline



[1] Fish-eye distortion removal with camera calibration *[Hou et al. 2021, Mayer et al. 2023a]*

[2] Vignetting effect removal with approximated radial polynomial function *[Bal et al. 2023]*

[3] Alignment with homography matrix estimated using matching feature points *[Hou et al. 2021, Mayer et al 2023a]*

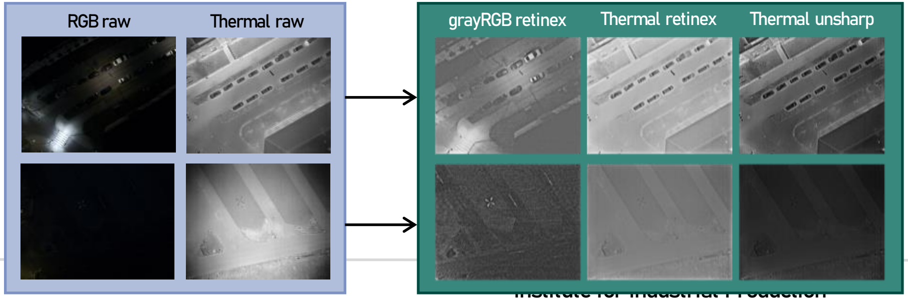
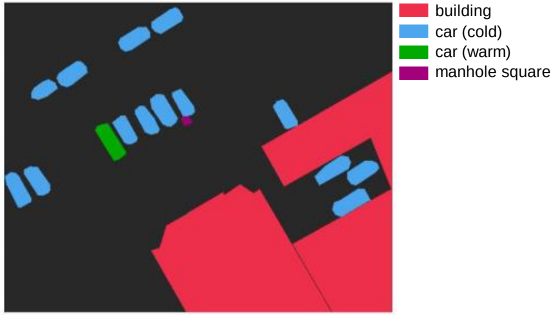


Image sources: IIP, KIT

Case Study: Annotations

- Annotation of 8001 common urban feature classes:
  - 18% buildings, 45% cars (warm, cold), 19% manholes round (warm, cold), 4.5% manholes square (warm, cold), 3% people, 9.5% streetlamps (warm, cold), and 1% miscellaneous warm objects
- Concatenation into classes:
  - Class imbalance pronounced

Class	# Annotations	# Pixels (*10^3)
Background	-	37 063.96
Building	1404	9 087.95
Car (cold)	2531	601.90
Car (warm)	1034	325.60
Manhole round	1536	50.51
Manhole square	358	12.79
Miscellaneous	81	8.38
Person	275	7.64
Street Lamp	782	27.18



Model Selection

- Multi-class semantic segmentation problem

→ U-Net

  - Most widely used in remote sensing *[Lv et al. 2023]*
  - Among most popular for urban feature segmentation *[Neupane et al. 2021, Ulku et al. 2020]*
  - Proficient at multispectral satellite image analysis *[Iglovikov et al. 2017]*
  - Various toolboxes, such as „segmentation_models“ *[Iakubovskii 2019]*

Architecture

  - Encoder-decoder structure for semantic segmentation and small datasets *[Ronneberger et al. (2015)]*
  - Transfer learning with ImageNet pretrained weights to compensate small dataset

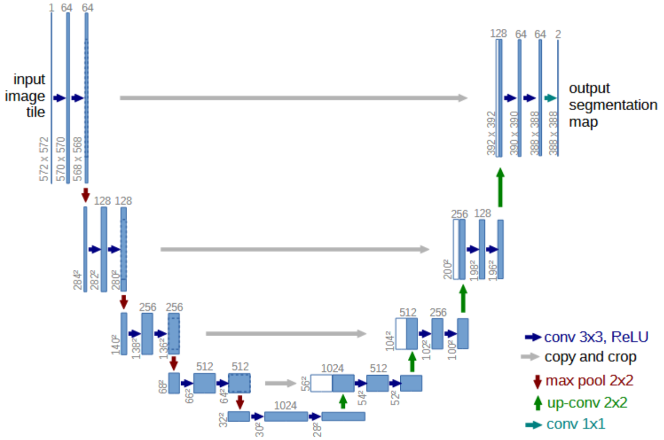


Figure: U-Net model architecture [Ronneberger et al. (2015)]

Institute for Industrial Production
Karlsruhe Institute of Technology

Evaluation Metrics

Karlsruhe Institute of Technology

- Common semantic segmentation metrics, specifically for imbalanced data

Accuracy	Balanced Accuracy	Mean Intersection over Union (IoU)	Weighted Mean IoU	Weighted F1-Score
▪ Percentage of correctly classified pixels out of all pixels	▪ Averaged percentage of correctly classified pixels per class i ▪ Check accuracy consistency over all categories	▪ Averaged similarity of predicted A and labelled B areas of a class i ▪ Check correctness of segmentation form and position	▪ Averaged similarity of predicted A and labelled B areas, weighted by class i prevalence ▪ Considers more common classes	▪ Averaged harmonic mean of Precision and Recall, weighted by class i prevalence ▪ Considers more common classes
$A = \frac{TP + TN}{TP + FP + TN + FN}$	$bA = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i}$	$mIoU = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert A_i \cap B_i\rvert}{\lvert A_i \cup B_i\rvert}$	$wmIoU = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i \frac{\lvert A_i \cap B_i\rvert}{\lvert A_i \cup B_i\rvert}$	$F_1 = \frac{1}{\sum_{i=1}^{n} w_i}\sum_{i=1}^{n} w_i \frac{2 * TP_i}{2 * TP_i + FP_i + FN_i}$

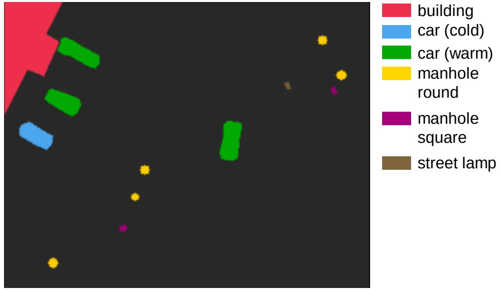
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative
w = weighting factor
(number of true class instances)

Ablation Study A: Backbone

- For comparison: Models trained for 25 epochs with a batch size of 8

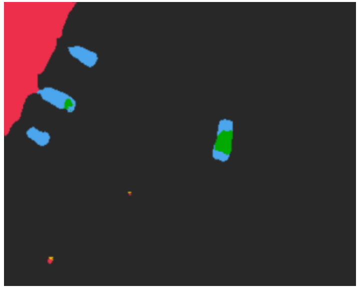
Backbone[1]	Accuracy	Balanced Accuracy	MeanIoU	Weighted MeanIoU	Weighted F1 Score
ResNet101	0.92867	0.36805	0.31952	0.88485	0.93026
ResNet152	0.93740	0.40942	0.35679	0.89603	0.93679
SeNet154	0.94460	0.33254	0.30553	0.90220	0.93845

- Deeper architectures better suited
- SeNet154 vs. ResNet152:
  - Model size: 1,46 GB vs. 0,79 GB
  - Prediction time: 2000ms vs. 798ms



Annotation mask / Ground truth

Prediction with ResNet152



Prediction with SeNet154

→ ResNet better at identifying under-represented classes

[1] All backbones are pretrained on the "ImageNet Large Scale Visual Recognition Challenge 2012" dataset

Ablation Study B: Loss Function & Hyperparameters

- Cross-entropy (CE) based loss most common in remote sensing *[Neupane et al. 2021]*
- Modified variant for class imbalance *[Lin et al, 2018]*
  → Sigmoid Focal CE
  - Works well for U-Net-based model for satellite imagery *[Dong et al, 2019]*
  - γ: focusing factor for attention on difficult-to-learn instances (default: 2)
  - α: weighting factor for dealing with imbalance (default: 0.25)

→ Higher LR favours underrepresented classes

Exp No.	Parameter					Accuracy	Balanced Accuracy	Mean IoU	Weighted Mean IoU	Weighted F1 Score
	α	γ	LR	EP	BA					
I	0.25	2	10^{-3}	25	8	0.93740	0.40942	0.35679	0.89603	0.93679
II	0.25	2	$2 \cdot 10^{-2}$	25	8	0.87732	0.41337	0.30651	0.81577	0.88135
III	0.25	2,5	$5 \cdot 10^{-4}$	30	8	0.94773	0.45254	0.40282	0.90747	0.94254
IV	0.25	2	10^{-3}	25	14	0.79478	0.37890	0.27814	0.73518	0.81487
V	0.25	2	10^{-3}	25	11	0.93218	0.40951	0.35875	0.88327	0.92669
VI	0.25	2	10^{-3}	25	6	0.92877	0.41599	0.34853	0.88084	0.92453
VII	0.25	2,5	$5 \cdot 10^{-4}$	30	12	0.93927	0.43621	0.37996	0.89538	0.93476
VIII	0.3	3	10^{-3}	25	8	0.93972	0.43982	0.38167	0.89677	0.93651
IX	0.35	3	10^{-3}	30	9	0.93551	0.44818	0.39308	0.88578	0.92763
X	0.3	3	10^{-3}	35	8	0.94782	0.53389	0.44056	0.91183	0.94708
XI	0.5	3	10^{-3}	35	8	0.94776	0.47693	0.41880	0.90851	0.94368
XII	0.3	4	10^{-3}	30	8	0.90352	0.44891	0.36107	0.85377	0.90928
XIII	0.3	3	$5 \cdot 10^{-4}$	35	8	0.95421	0.52678	0.43399	0.92057	0.95192

Legend: alpha = α / gamma = γ / learning rate = LR / epochs = EP / batch size = BA

Key Take-Aways

- Feature engineering (data processing) helps adapt to acquisition circumstances (lighting conditions, etc)
- Best model combination:

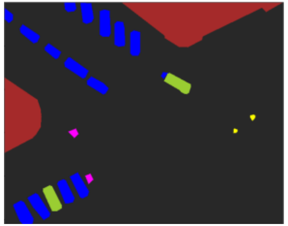
Exp No.	Parameter					Accuracy	Balanced Accuracy	Mean IoU	Weighted Mean IoU	Weighted F1 Score
	α	γ	LR	EP	BA					
X	0,3	3	10^{-3}	35	8	0.94782	0.53389	0.44056	0.91183	0.94708

→ Focal cross entropy loss is useful for RGB+T multispectral data

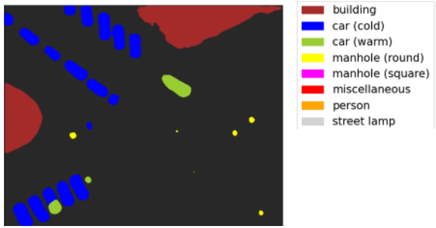
→ Higher learning rate favours balanced accuracy and mean IoU

→ ResNet152 more adept at identifying underrepresented classes than SeNet154

Annotation mask / Ground Truth



Model prediction



- Limitation: Segmentation of underrepresented classes require improvement

→ Data augmentation, increase annotation amounts, different models, feature engineering

Thank you for your attention. Any questions?

elena.vollmer@kit.edu

Karlsruhe Institute of Technology (KIT)
Kaiserstraße 12
76131 Karlsruhe
www.kit.edu

Institute for Industrial Production (IIP)
Hertzstraße 16 - Building 06.33
76187 Karlsruhe

References

- Beyerer, J., Ruf, M. and Herrmann, C. (2018). CNN-based thermal infrared person detection by domain adaptation. In: *Defense + Security*. 8. https://doi.org/10.1117/12.2304400
- Bal, A. and Palus, H. (2023) Image Vignetting Correction Using a Deformable Radial Polynomial Model. In: *Sensors* (Basel, Switzerland) 23 (3). https://doi.org/10.3390/s23031157
- Dutta, A., Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2276—2279, MM '19, Association for Computing Machinery, ISBN 9781450368896, https://doi.org/10.1145/3343031.3350535
- Friman, O., Follo, P., Ahlberg, J., Sjokvist, S. (2014). Methods for Large-Scale Monitoring of District Heating Systems Using Airborne Thermography. In: *IEEE Transactions on Geoscience and Remote Sensing*, 52(8): 5175–82. https://doi.org/10.1109/TGRS.2013.2287238
- Gade R. and Moeslund, T. (2014). Thermal Cameras and Applications: A Survey. In: *Machine Vision and Applications*, 25(1):245–262. ISSN 0932-8092. https://doi.org/10.1007/s00138-013-0570-5
- He, Y et al. (2021). Infrared machine vision and infrared thermography with deep learning: A review. In: *Infrared Physics & Technology*. 116. 103754. https://doi.org/10.1016/j.infrared.2021.103754
- Hou, Y., Volk, R., Chen, M., Soibelman, L. (2021). Fusing tie points' rgb and thermal information for mapping large areas based on aerial images: A study of fusion performance under different flight configurations and experimental conditions. In: *Automation in Construction* 124, 103554. https://doi.org/10.1016/j.autcon.2021.103554
- Iakubovskii, P. (2019). Segmentation Models. https://github.com/qubvel/segmentation_models
- Iglovikov, V., Mushinskiy, S., Osin, V. (2017). Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition, http://arxiv.org/abs/1706.06169
- Hossain, K., Villebro, F., and Forchhammer, S. (2020) UAV Image Analysis for Leakage Detection in District Heating Systems using Machine Learning. Pattern Recognition Letters, 140:158–164. ISSN 01678655. https://doi.org/10.1016/j.patrec.2020.05.024
- Kütük, Z., Algan, G. (2022). Semantic Segmentation for Thermal Images: A Comparative Survey. https://doi.org/10.48550/arXiv.2205.13278
- Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P. (2018). Focal Loss for Dense Object Detection. In: *IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/10.48550/arXiv.1708.02002
- Lv, J., Shen, Q., Lv, M., Li, Y., Shi, L., Zhang, P. (2023). Deep learning-based semantic segmentation of remote sensing images: a review. In: *Frontiers in Ecology and Evolution* 11, ISSN: 2296-701X, URL: https://www.frontiersin.org/articles/10.3389/fevo.2023.1201125

References

- Mayer, Z., Kahn, J., Götz, M., Hou, Y., Beiersdörfer, T., Blumenröhr, N., Volk, R., Streit, A., Schultmann, F. (2023a). Thermal Bridges on Building Rooftops. In: *Scientific Data* 10(1), 268, ISSN 2052-4463, https://doi.org/10.1038/s41597-023-02140-z
- Mayer, Z., Kahn, J., Hou, Y., Götz, M., Volk, R., Schultmann, F. (2023b). Deep learning approaches to building rooftop thermal bridge detection from aerial images. In: *Automation in Construction,* 146, p. 104690. Elsevier BV. https://doi.org/10.1016/j.autcon.2022.104690
- Neupane, B., Horanont, T., Aryal, J. (2021). Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis. In: *Remote Sensing,* 13(4), 808, ISSN 2072-4292, https://doi.org/10.3390/rs13040808
- Ronneberger, O., Fischer, P., Brox, T. (2015): U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI).* pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Song, K., Zhao, Y., Huang, L., Yan, Y., Meng, Q. (2023). RGB-T image analysis technology and application: A survey. In: *Engineering Applications of Artificial Intelligence* 120, 105919, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2023.105919
- Tu, Z., Ma, Y., Li, Z., Li, C., Xu, J., Liu, Y. (2023). RGBT Salient Object Detection: A Large-Scale Dataset and Benchmark. In: *IEEE Transactions on Multimedia*, 25: 4163-4176, 2023, https://doi.org/10.1109/TMM.2022.3171688
- Ulku, I., Barmpoutis, P., Stathaki, T., Akagunduz, E. (2019). Comparison of single channel indices for U-Net based segmentation of vegetation in satellite images. In: *Twelfth International Conference on Machine Vision (ICMV)*, vol. 11433, p. 1143319, International Society for Optics and Photonics, SPIE (2020), https://doi.org/10.1117/12.2556374
- Vollmer, E., Volk, R., Schultmann, F. (2023). Automatic analysis of UAS-based thermal images to detect leakages in district heating systems. In: *International Journal of Remote Sensing*, ISSN 0143-1161, https://doi.org/10.1080/01431161.2023.2242586
- Wang, L., Wang, J., Liu, Z., Zhu, J., & Qin, F. (2022). Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. In: *The Crop Journal,* 10(5): 1435–1451. Elsevier BV. https://doi.org/10.1016/j.cj.2022.01.009
- Wang, T., Kim, G., Kim, M. and Jang, J. (2023). Contrast Enhancement-Based Preprocessing Process to Improve Deep Learning Object Task Performance and Results. In: *Applied Sciences,* 13. 10760. https://doi.org/10.3390/app131910760