Automatic Detection of Tiny Drainage Outlets and Ventilations on Flat Rooftops from Aerial Imagery

Lukas Arzoumanidis[1], Weilian Li*[1,3], Julius Knechtel[2], Yudha Kosmayadi[1], Youness Dehbi[1]

[1] Computational Methods Lab, HafenCity University Hamburg, Germany – {firstname.lastname}@hcu-hamburg.de
[2] Institute of Geodesy and Geoinformation, University of Bonn, Germany – {lastname}@igg.uni-bonn.de
[3] Faculty of Geosciences and Engineering, Southwest Jiaotong University, China

Keywords: tiny object detection, drainage outlet, ventilation, flat rooftop inspection, as-built BIM, coarse-to-fine visual recognition

Abstract

Flat rooftops on residential and industrial buildings house critical drainage and ventilation systems, which play essential roles in channeling water away from structures and preventing moisture accumulation. These utilities are vital for maintaining the structural integrity of rooftops, safeguarding against water pooling and moisture buildup that could otherwise lead to damage or even collapse, particularly during extreme weather events. However, current inspection and maintenance practices for these systems are predominantly manual, making them time-consuming, labor-intensive, and sometimes hazardous. This paper presents an automated approach to detecting drainage outlets and ventilation systems on flat rooftops, using a custom-labeled dataset of high-resolution aerial imagery. We evaluated two different object detection methods, with FCOS (Fully Convolutional One-Stage Object Detection) outperforming Faster R-CNN in identifying these small utilities. The outcomes pave the way for new applications, as detected utilities can act as sparse data points that trigger constraint-based reasoning processes for estimating hidden utility networks in as-built Building Information Modeling (BIM) contexts. Embedding these identified objects into GIS or BIM models represents an initial step towards coarse-to-fine visual recognition, enabling customized semantic mission planning for autonomous exploration and inspection using Unmanned Aerial Vehicles (UAVs). The labeled dataset used in this study is publicly available by following this link `https://zenodo.org/records/14040571`.

1. Introduction

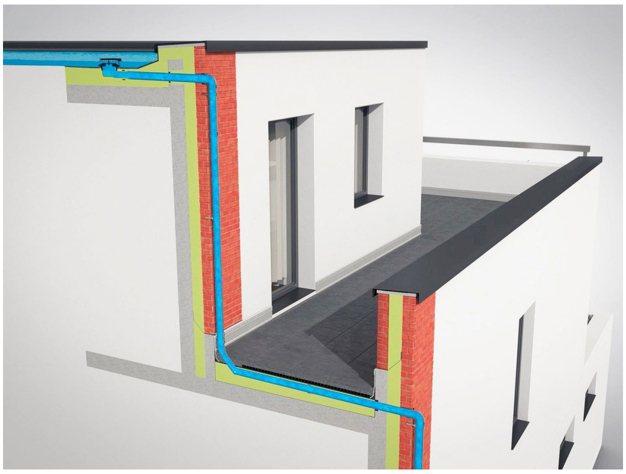
In recent years, the frequency of extreme precipitation events increased drastically (Myhre et al., 2019; Li et al., 2024). However, buildings, especially high-rise buildings with flat rooftops, sometimes referred to as *parapet roofs*, can take substantial damage from large amounts of rain- or snowfall as they have to carry the additional weight of the accumulated precipitation mass. It can be roughly estimated that a water accumulation of just 10 cm corresponds to a weight of approximately 100 kg/m^2. For a flat rooftop with an area of 200 m^2, this means an additional load of 20 tons. To overcome this disadvantage, drainage outlets are used for flat rooftops such as illustrated in Figure 1a, where the associated pipes lead into the rainwater drainage system as can be seen in Figure 1b. Effective drainage on such roofs prevents water from accumulating, which otherwise leads to puddling and stagnant water. This standing water adds stress to roofing materials, gradually weakening them and potentially causing structural damage over time, e.g, through mold and mildew or cracks as well as leakages. In this context, roof drainage systems aim to divert water away from the structure to prevent issues from water pooling on the rooftop or near the foundation. Since warehouses and industrial buildings in industrial areas often have low-slope roofs that don't facilitate natural runoff as well as pitched roofs, built-in drainage system are mostly a crucial alternative.

Beyond water drainage, rooftops are simultaneously accommodated with roof ventilations which can be likewise observed from on the rooftops as can be seen on the most right of Figure 1a. They have the purposes to reduce moisture buildup in the attic or roof cavity, which can occur from everyday activities, such as cooking or showering and are critical as they prevent mold, mildew, and wood rot, which could damage the structural integrity and insulation of walls and roof elements.

However, once drainage outlets or ventilations are clogged and get blocked or damaged, e.g., by foliage or other debris, rainwater or snowfall can accumulate, leading to roof leaks or in the worst case to a collapse of the entire roof due to the inability of water to drain and the buildup of moisture from indoor usage below. Therefore, the regular inspection of drainage outlets and ventilations is an essential task for rooftop maintenance as it helps to diagnose damage or clogging in an early state and thereby reduces the threat posed by large precipitation accumulation. Currently, roof drainage outlets are primarily inspected manually, which is tremendously time-consuming and labor-intensive for large facilities and buildings. In addition, inspecting drainage outlets of flat roofs, e.g., on high-rise buildings, poses significant safety risks for the inspectors. With the availability of low cost drones and high-resolution aerial imagery, derived by planes or satellites, automated inspections have attracted increasing attention and are now applied in various fields Rakha and Gorodetsky (2020), such as power lines Li et al. (2023), bridges Mandirola et al. (2022) and rain gutter Dehbi et al. (2020).

In this paper, we lay the groundwork for fully automated rooftop inspection and exploration, eliminating the need for a human pilot. A crucial component of this approach is the automatic detection of drainage outlets and ventilation systems, essential for effective rooftop inspection, monitoring, and mapping. These elements, classified as small objects, occupy only a few pixels in aerial images, making their detection particularly challenging. Consequently, remote detection of rooftop drainage outlets and ventilation systems represents a foundational step in a coarse-to-fine detection and modeling work-

(a) drainage outlets for rainwater and ventilation for flat rooftops
Hagmans GmbH©[1]



(b) in-wall drainage pipes for parapet roof and stepped roof surfaces
(Behr, 2021)

Figure 1. Examples for visible and hidden rooftop infrastructure.

flow, ultimately enriching as-built GIS or BIM models. These models enable informed, automated mission planning for subsequent close-range in-situ exploration and inspection. Additionally, once these small rooftop features are accurately identified and segmented, they offer valuable data for inferring hidden utility networks within buildings, as illustrated in Figure 1b. The spatial arrangement of drainage and ventilation features on rooftops can further trigger a reasoning process about the internal layout and function use of rooms below. For instance, detected ventilation systems are often correlated with bathrooms or kitchens in the apartments beneath.

In this context, deep learning-based approaches have proven particularly effective Rabbi et al. (2020). A key advantage of deep learning is that it eliminates the need for manual feature engineering, which would be challenging in our case due to the minimal pixel footprint of each object. However, a significant drawback of deep learning is its requirement for a large corpus of labeled training data for each object detection task. To address this, we annotated drainage outlets and ventilation systems across 740 rooftop images, each representing a unique rooftop. This dataset will be made publicly available alongside this work under the CC BY 4.0 (Creative Commons Attribution 4.0) license.

This work proposes an automatic detection approach of small and tiny objects on rooftops using deep learning and discusses its potential enrichment for model-image registration applied in autonomous Unmanned Aerial Vehicle (UAV) navigation and task execution. For this task, we apply two deep learning-based methods, namely Faster R-CNN (Ren et al., 2017) and FCOS (Tian et al., 2019), both of which have proven effective for object detection.

The remainder of this paper is structured as follows: Section 2 provides an overview over the related work. Section 3 describes the dataset used in the study. Section 4 gives insights into the introduced approach. Section 5 discusses the experimental results. Section 6 concludes and summarizes the paper and provides an outlook for future research.

[1] `https://hagmans-gmbh.de/flachdach_entwaesserung.html`

2. Related Work

While extensive research has been conducted on the detection and segmentation of rooftops for many years, the focus has predominantly been on general rooftop identification and classification, e.g., using stochastic processes like Markov Random Fields (Katartzis and Sahli, 2008) or Support Vector Machines (Baluyan et al., 2013; Mohajeri et al., 2018). Additionally, research has also explored rooftop reconstruction from 3D point clouds through filters that are based on prior knowledge represented by density distributions or informed model sampling (Dehbi et al., 2019, 2021). Methods for detecting rooftops as large, singular structures based on RGB aerial imagery or 3D point clouds are well-established, utilizing various machine learning and computer vision techniques, such as deep learning (Buyukdemircioglu et al., 2021). Besides the general rooftop identification also the detection and segmentation of larger objects located on rooftops, such as photovoltaic panels, has been substantially researched (Castello et al., 2019; Wang et al., 2023; Soujanya et al., 2024). However, the detection of tiny objects which are also situated on these rooftops, such as small installations and facilities, or equipment, remains underexplored.

To enhance small object detection in aerial imagery, Rabbi et al. (2020) proposed combining an edge-enhanced super-resolution Generative Adversarial Networks (EESRGAN) with the detection network Faster Regional Convolutional Neural Network (Faster R-CNN) in an end-to-end framework. This approach aims to improve the detection performance of small objects in aerial imagery. Another approach that aims to improve the accurate detection of small objects in aerial imagery proposes an anchor-free detector named FE-CenterNet. In their work, they proposed a feature enhancement module (FEM) which is formed of a feature aggregation structure (FAS) and an attention generation structure (AGS) (Shi et al., 2022).

To assess damage on flat metal rooftops, Abdullah et al. (2022) applied deep learning techniques to detect faulty or defective bolts, which can serve as indicators of potential rooftop damage. Similarly, Hezaveh et al. (2017) utilized a Convolutional Neural Network to identify and evaluate small areas affected by hail impact, allowing them to infer the extent of damage to the rooftop.

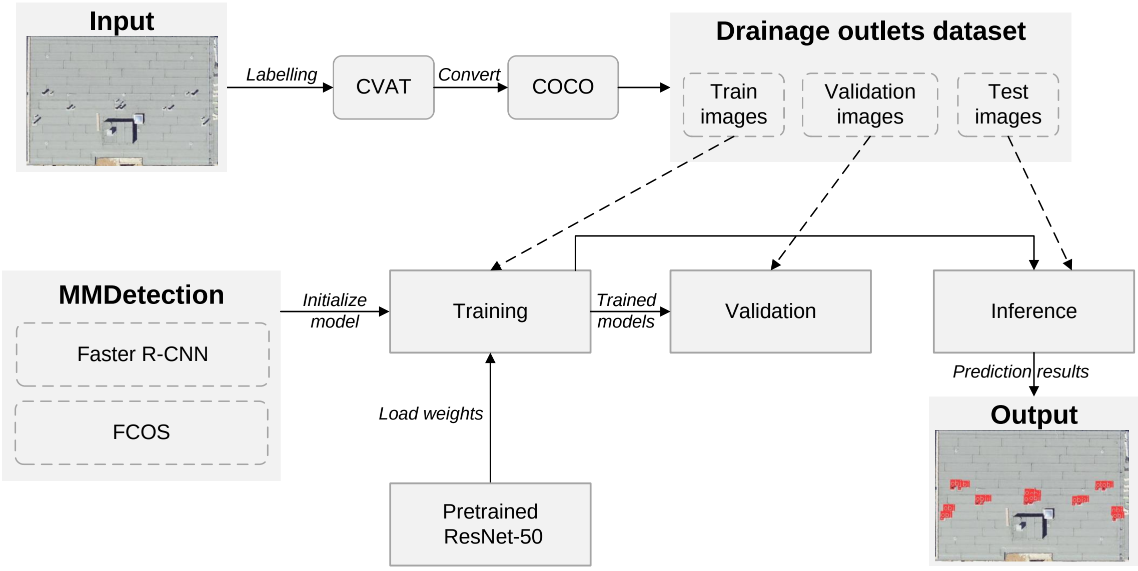


Figure 2. Overview of our approach for the automatic detection of drainage outlets on flat rooftops.

In an effort to detect improperly positioned equipment on flat rooftops – equipment that may obstruct proper water drainage and complicate maintenance services – dos Santos et al. (2023) recently explored the application of several deep learning approaches for detecting objects on rooftops. Their study compared various techniques for identifying installations, such as condensers, circular antennas, ventilation pipes, and kitchen exhaust chimneys, using aerial images captured by unmanned aerial vehicles (UAVs). The experimental results demonstrated that the best-performing deep learning model for this task was Faster R-CNN. Similarly, the recent work of Mayer et al. (2023) proposed the use of UAVs equipped with thermographic sensors to detect thermal bridges on rooftops. They highlighted that identifying such thermal leakages and implementing subsequent improvements could significantly enhance energy efficiency of existing buildings.

Several of the previously mentioned approaches have partially addressed the detection of small rooftop utilities, which is a focus of this work. However, most of these methods rely on close-range remote sensing for identification using UAVs, which simplifies the task since the objects are relatively larger. In contrast, we assert that mapping such small objects as a preliminary step enhances 3D building models by incorporating additional semantic knowledge about rooftops, thereby facilitating an automatic coarse-to-fine visual recognition process. At the core of this paper, we initiate the first step in addressing the identification of target objects based on orthophotos. These objects represent a negligible portion of the underlying roof, making them prone to being mistaken for white noise, which complicates their detection. This gap in the literature underscores the necessity for specialized approaches tailored to the unique challenges of detecting tiny objects on rooftops, including their inaccessibility, small size, and often occluded nature. Once these objects have been identified remotely, a customized mission plan can be developed for autonomous UAV navigation and task execution to facilitate drone-based on-site inspections.

Currently, a variety of datasets are employed for developing methods in tiny object detection. These datasets can generally be categorized into three primary domains: traffic signal detection, pedestrian detection, and aerial image detection (Wei et al., 2024). Examples of aerial image datasets include DOTA[2] and DIOR[3]. However, since such datasets focus on common object categories like vehicles, bridges, and storage tanks they lack annotated information for object categories such as drainage outlets and rooftop ventilations, which are essential for addressing our problem at hand.

Additionally, thermal sensors could be employed to detect tiny infrastructure, such as rooftop ventilations, since they typically emit vaporized air at a temperature different from the surrounding environment and structures. This temperature difference makes them detectable by thermal cameras. To train a detection model using images from thermographic sensors, a recently published dataset containing visible-thermal tiny objects has been made available (Ying et al., 2024). However, this dataset includes labels for objects such as roads or bridges which may render it unsuitable for our specific problem. This is due to the fact that the dataset's modality differs significantly from that of the unseen data expected during deployment in our specific use case.

3. Dataset

To the best of our knowledge, no existing dataset is specifically designed for detecting tiny drainage and sewage systems on rooftops. Therefore, as previously mentioned, we have created and annotated our own dataset to address this specific task. The original high-resolution aerial imagery are orthophotos with a ground sampling distance of 7.5 cm, provided by the Office for Land Management and Geoinformation of the City of Bonn, Germany [4]. Our dataset was created through manual annotation using the Computer Vision Annotation Tool (CVAT) [5] and comprises 740 image pairs. Each pair consists of a rooftop image and a corresponding annotated mask indicating the drainage outlets and ventilations, as illustrated in Figure 3. Since rooftops vary in size, we aimed to create image pairs that capture a single rooftop per image without overlaps or cutoffs.

[2] https://captain-whu.github.io/DOTA/index.html
[3] http://www.escience.cn/people/gongcheng/DIOR.html
[4] https://stadtplan.bonn.de
[5] https://www.cvat.ai

Consequently, the dimensions of each image pair differ. The dataset was then split randomly into 80% for training, 10% for validation, and 10% for testing.

To take profit from available object detection libraries like MM-Detection, we converted the dataset from its original pairwise format into annotations of the Common Objects in Context (COCO) (Lin et al., 2014) format for object detection as illustrated in Figure 3. This format is widely recognized in object detection tasks and is structured as a single JSON file containing labels, metadata, and detailed information about the images and object locations as can be seen in Figure 3. As menitioned, the dataset is available at the following link: `https://zenodo.org/records/14040571`. In addition to the COCO-formatted dataset, we provide the dataset in its original format to support various machine learning tasks, such as semantic segmentation and panoptic segmentation, as well as to accommodate different data-loading requirements for diverse deep learning models.

4. Methodology

This section presents our framework, detailing the object detection models applied and providing an overview of the metrics used to assess the quality of the results.

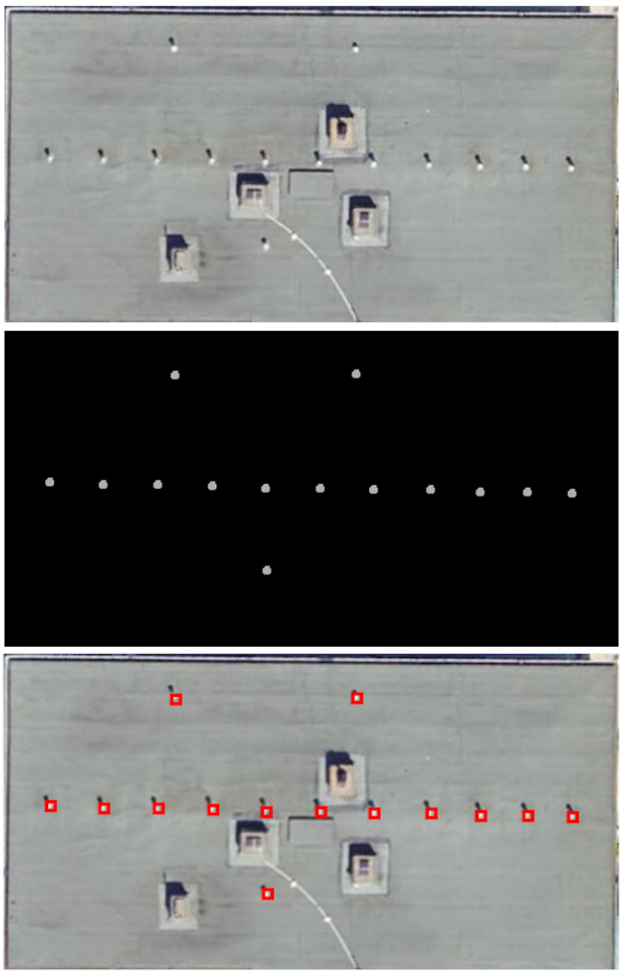


Figure 3. (Top) Exemplary image illustrating a plain flat rooftop, (Middle) illustrates manually hand-annotated rooftop ventilations, (Bottom) illustrates annotated ventilations for COCO.

After converting our training data into the COCO object detection annotation format, we train two architecturally distinct deep learning models to detect drainage outlets and ventilation points simultaneously, as shown in Figure 2. Notably, drainage outlets and ventilation systems are grouped under a single category, with no information available on their distribution. As a result, both object classes cannot be distinguished during detection, and their individual performance cannot be evaluated. For our experiments, OpenMMLab's MMDetection, a library that offers a framework within which object detection modules are provided, evaluated and visualized is applied Chen et al. (2019). We choose this library due to its large variety of object detection model implementations and its application in other approaches for the detection of rooftop objects, such as dos Santos et al. (2023) and Mayer et al. (2023). For our experiments, we chose Faster R-CNN and FCOS for the task of detecting drainage outlets and ventilation systems in high-resolution aerial imagery.

In this work, we have chosen to focus solely on deep learning-based models for the object detection task that are not overly resource-intensive. The goal is to perform inference later on mid-flight on a drone to automatically detect objects that are useful for navigation-related tasks, such as coarse-to-fine visual recognition, enabling customized semantic mission planning for autonomous UAV inspections. Given the computing performance and energy constraints of low-powered devices like drones, which limit the feasibility of running deep learning models, we aim to test only those with a low parameter count.

Faster R-CNN is an anchor-based object detection approach that consists of a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, enabling nearly computationally cost-free region proposals. An RPN is able to simultaneously predict object bounds and objectness scores at each position which supports the underlying convolutional neural network where to focus on when predicting the type of object (Ren et al., 2017). In contrary, FCOS (Fully Convolutional One-Stage Object Detection) follows the strategy of an anchor-free object detection approach which avoids any computational overhead related to anchor boxes and, hence, avoids computational costs for regional proposals at all (Tian et al., 2019).

During training, both models maximize the probability of correct detections through their loss functions by outputting confidence scores for each object in the scene. For validation during and after training, we use the mean Average Precision (mAP). Mean AP is a common evaluation metric used for measuring the performance of a model for object detection and information retrieval tasks. In our work, we use the mAP to measure the accuracy of bounding box predictions against their annotated ground-truth. The following equation describes the mAP formally:

$$\mathrm{mAP} = \frac{1}{n} \sum_{i=1}^{k=n} \mathrm{AP}_i, \tag{1}$$

where n is equal to the number of classes and AP_i is the average precision of class i. However, to calculate the mAP we need to calculate the Intersection over Union (IoU) as well as Precision and Recall metrics. The IoU describes the ratio between the area of overlap to the area of union between the predicted and the ground-truth bounding box. Precision is usually referred to as the ratio of true positives against the total predictions made while Recall is defined as the ratio of actually correctly detected true positives against all true positives. However, to control the quality of predictions an IoU threshold is applied beforehand to filter out true positives which do not have enough overlapping with the ground-truth bounding box. The mentioned AP_i in Equation 1 is defined as the area below the Precision-Recall (PR) curve for each object class. The PR curve is generated by sorting the predictions by their confidence score and calculating Precision and Recall at each confidence score threshold accordingly. Finally, the AP values are calculated for each class and the mAP is calculated as the mean precision across various IoU thresholds. In the implementation of MMDetection which is used in this work, the mAP calculates the IoU thresholds for all values lying in $[0.5, 0.95]$. Thus, mAP_{50} is the precision at a single threshold of IoU = 0.5 and mAP_{75} is the precision at a single threshold of IoU = 0.75 accordingly. Unfortunately, the off-the-shelf implementation of MMDetection used in this work does not calculate or return additional evaluation metrics after inference. For this reason, we limited ourselves to using mAP, which is commonly applied in the context of object detection.

Furthermore, we not only compare both approaches in their original implementations but also evaluate them against a recently proposed parameter optimization method called Slicing Aided Hyper Inference (SAHI), which is designed to fine-tune general object detection techniques for detecting small objects (Akyon et al., 2022). By incorporating SAHI during the inference stage, our approach aims to highlight potential benefits for detecting tiny objects, such as drainage outlets or ventilation systems on flat rooftops.

5. Experimental Results

This section presents and discusses our experiments and provides the parameter values have been used for the sake of replication. The experiments were conducted on an NVIDIA RTX 2000 Ada with 8GB of VRAM. MMDetection 3.3.0 with Python 3.10, CUDA 12.2, and Pytorch 2.1.0 have been used to train two models, specifically Faster R-CNN and FCOS. To further refine our trained models for small object detection, we applied, as previously mentioned, SAHI to optimize model performance at the inference stage. For each experiment, the network was trained for 100 epochs with a learning rate of 0.0025 and Stochastic Gradient Descent (SGD) as optimizer. With the mentioned setup and parameter configuration, we trained two models in order to perform the task of drainage outlet and ventilation detection on flat rooftops accordingly.

Our experimental results highlight the differences in performance between Faster R-CNN and FCOS during training and testing. Overall, both models exhibit comparable performance trends during training, as illustrated by their respective mAP, mAP_{50}, and mAP_{75} curves in Figure 4. Notably, the FCOS model consistently outperforms Faster R-CNN during training, particularly in mAP_{50}, where FCOS achieves a peak score of 0.65 between the 40th and 50th epochs – approximately 40% higher than Faster R-CNN's corresponding values.

Model	mAP	mAP_{50}	mAP_{75}
FCOS	**0.207**	**0.652**	**0.075**
FCOS + SAHI	0.193	0.646	0.058
Faster R-CNN	0.067	0.241	0.006
Faster R-CNN + SAHI	0.065	0.232	0.007

Table 1. Bounding box mean Average Precision with different IoU thresholds on the test dataset.

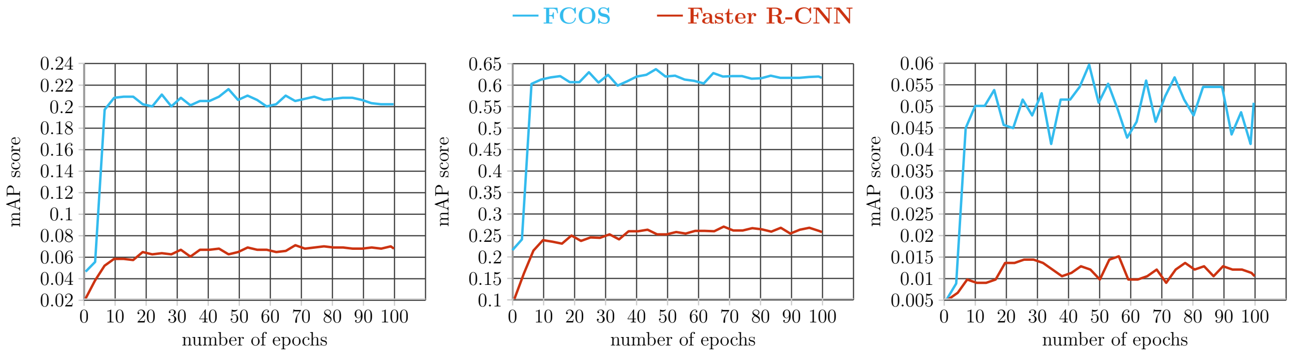
According to Figure 4a, the average mAP scores for FCOS range between 0.2 and 0.22, significantly exceeding Faster R-CNN's scores, which remain between 0.06 and 0.07. Additionally, both models show a marked drop in performance at higher IoU thresholds, as evidenced by mAP_{75} being consistently lower than mAP_{50} as can be seen in Figures 4b and 4c. This disparity suggests that both models struggle with detection precision under stricter matching conditions, where predictions must more closely align with ground-truth boxes to be classified as correct detections. This conclusion is further substantiated by the considerably lower mAP_{75} scores relative to the overall average mAP scores for both models.

The average mAP, mAP_{50}, and mAP_{75} scores during inference closely mirror the performance observed during validation. As shown in Table 1, both FCOS and Faster R-CNN achieve higher performance at lower IoU thresholds, as evidenced by the relative scores of mAP_{50} and mAP_{75}. Similarly, both models demonstrate higher average mAP scores compared to mAP_{75} scores, reinforcing the trends observed during training.

The experimental results presented in Figures 5 and 6 illustrate the performance differences between FCOS and Faster R-CNN in detecting drainage outlets and ventilations on previously unseen test images. Notably, Faster R-CNN shows a marked tendency to over-detect target objects as the confidence threshold decreases, evident in the increased detection of drainage outlets and ventilations on the left side of Figures 6c and 6b. Conversely, as the confidence threshold becomes more stringent, the detection rate of the desired objects declines, as shown in the upper section of Figure 6a.

With a relatively low confidence threshold of 0.2, FCOS successfully detects nearly all target objects within the unseen test image without over-detecting targets, as shown in Figure 5a, indicating a higher detection performance compared to Faster R-CNN. This finding aligns with the training results previously described and suggests that both models may encounter challenges in learning effective feature representations for small objects, likely due to the inherent difficulties posed by their size. As mentioned in Section 2, FCOS is an anchor-free model meaning it does not rely on predefined anchor boxes. On smaller scales this architectural design can have an advantage for tiny objects because it avoids the limitations of anchors that may not be tuned for tiny scales. Therefore, Faster R-CNN which is an anchor-based method might struggle with detecting tiny objects.

To further enhance our models' ability to detect tiny objects, we applied SAHI at the inference stage for both trained models, Faster R-CNN and FCOS. The confidence threshold was set to 0.1, the slice height and width to 400, and the overlap height and width ratio to 0.3, closely aligning with the recommended hyperparameter values from Akyon et al. (2022). Consistent with the evaluation of both models in their original form, FCOS + SAHI exhibited significantly higher mAP, mAP50, and mAP75 scores compared to Faster R-CNN. However, the combination of both models with SAHI resulted in slightly lower overall mAP scores compared to the respective original models, as shown in Table 1. This is a noteworthy observation, as SAHI is specifically designed to enhance tiny object detection in existing models. One approach to understanding why SAHI fails to improve model accuracy is Gradient-weighted Class Activation Mapping (Grad-CAM), as proposed by Selvaraju et al. (2017). Grad-CAM visualizes the regions of the input image that the models focus on, both with and without SAHI. By comparing



(a) mAP (b) mAP_{50} (c) mAP_{75}

Figure 4. Validation mAP scores calculated on the predicted and ground-truth bounding box during training.

(a) inference on test image (b) ground-truth of test image (c) annotated mask of test image

Figure 5. FCOS inference on test image with a confidence threshold of 0.2.

(a) inference with confidence threshold 0.3 (b) inference with confidence threshold 0.6 (c) inference with confidence threshold 0.9

Figure 6. Faster R-CNN inference test image results for different confidence thresholds.

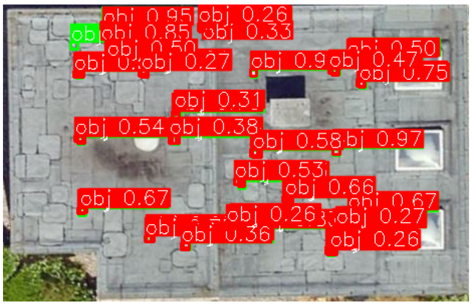
these regions, a more insightful analysis of the training process can be performed. Attention-Guided Learning implements this by designing loss functions that are aligned with activation maps or by leveraging knowledge of the activated regions as a regularization term that penalizes activations outside the target regions (Li et al., 2018).

Overall, when selecting FCOS as model our approach proves to be effective in addressing the challenging task of identifying tiny drainage outlets and ventilation utilities on rooftops. The detected utilities can be visually highlighted with bounding boxes, and their locations are recorded in COCO format. Given that the provided test data is georeferenced in a coordinate reference system the global position of the detected utilities is available. This information allows the utilities to be integrated as semantic information into existing 3D building models such as CityGML or as-built BIM. As an example, such semantic knowledge could enable a maintenance drone to perform model-image registration, thereby enhancing navigation and task execution, similar to the approach demonstrated in Dehbi et al. (2020). Enhancing the navigation and localization of UAVs plays a critical role in the mission planing process for an autonomous task execution, e.g., exploration or inspection.

Furthermore, inspired by the concepts in Dehbi et al. (2022), the detected objects can serve as sparse observations that initiate a reasoning process to estimate hidden utility networks, such as in-wall piping, in existing buildings.

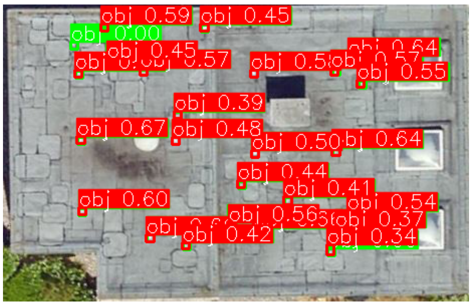
6. Conclusion and Outlook

In this work, we employ two architecturally distinct approaches for object detection of drainage outlets and ventilation structures on flat rooftops. For the automatic identification of



(a) Faster R-CNN + SAHI

(b) Faster R-CNN + SAHI



(c) FCOS + SAHI

(d) FCOS + SAHI

Figure 7. Faster R-CNN + SAHI and FCOS + SAHI results, including confidence scores. Ground truth is shown as green bounding boxes with a score of 0. When the model successfully detects an object, its red bounding box overlaps with the green ground truth bounding box.

these tiny objects, we manually annotated a dataset containing 740 high-resolution aerial images of rooftops. With a ground sampling distance of 7.5 cm, the task of detecting these targets is challenging due to the extremely small size of the objects at this resolution.

Our results demonstrate that FCOS detects target objects reliably, achieving an mAP_{50} of 65.2%. In contrast, Faster R-CNN was less effective, with an mAP_{50} of only 24.1%, indicating lower reliability in target detection. To further improve the mAP score and achieve a higher number of accurate detections, we will evaluate the performance of other light weight and actively developed deep learning-based object detection models, such as YOLO. Additionally, further research will focus on the explainability of deep learning-based object detection on our dataset, e.g., using Class Activation Mapping (CAM).

As mentioned, the proposed object detection approach for tiny rooftop utilities enhances 3D building models by incorporating additional semantic knowledge serving as a preliminary step towards developing knowledge-driven localization methods for drone navigation tasks, such as on-site inspections. Besides, the identified utilities will serve as sparse observations to trigger a constraint-based reasoning for the enrichment of BIM models by the underlying as-built state. Further, the identified utilities could serve as basis to schedule a mission planing for an automatic exploration using autonomous UAVs. Both investigation directions will be subject of upcoming research topics.

7. Acknowledgment

The authors would like to thank the Office for Land Management and Geoinformation of the City of Bonn for providing high-resolution orthophotos. We also extend our gratitude to Guessous Energie Building GmbH for their valuable insights and discussions, and to Gizem Sen for her outstanding support in annotating the dataset.

References

Abdullah, M., Ahmad, M., Kim, Y.-G., Han, D., 2022. Small objects and faults detection on corrugated metal roof using drone equipped with deep learning. *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 1–4.

Akyon, F. C., Altinuc, S. O., Temizel, A., 2022. Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection. *2022 IEEE International Conference on Image Processing (ICIP)*, 966-970.

Baluyan, H., Joshi, B., Al Hinai, A., Woon, W. L., 2013. Novel approach for rooftop detection using support vector machine. *International Scholarly Research Notices*, 2013(1), 819768.

Behr, C., 2021. Attikaentwässerung - Notausgang für Starkregen. *Das Fachmedium der TGA-Branche*, 05.

Buyukdemircioglu, M., Can, R., Kocaman, S., 2021. Deep Learning based roof type classification using very high resolution aerial imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2021, 55–60.

Castello, R., Roquette, S., Esguerra, M., Guerra, A., Scartezzini, J.-L., 2019. Deep learning in the built environment: automatic detection of rooftop solar panels using Convolutional Neural Networks. *Journal of Physics: Conference Series*, 1343(1), 012034.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C. C., Lin, D., 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv:1906.07155*.

Dehbi, Y., Henn, A., Gröger, G., Stroh, V., Plümer, L., 2021. Robust and fast reconstruction of complex roofs with active sampling from 3D point clouds. *Transactions in GIS*, 25(1), 112–133.

Dehbi, Y., Klingbeil, L., Plümer, L., 2020. UAV mission planning for automatic exploration and semantic mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 521–526.

Dehbi, Y., Knechtel, J., Niedermann, B., Haunert, J.-H., 2022. Incremental constraint-based reasoning for estimating as-built electric line routing in buildings. *Automation in Construction*, 143, 104571.

Dehbi, Y., Koppers, S., Plümer, L., 2019. Probability density based classification and reconstruction of roof structures from 3D point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W16, 177–184.

dos Santos, L. M. A., Zanoni, V. A. G., Bedin, E., Pistori, H., 2023. Deep learning applied to equipment detection on flat roofs in images captured by UAV. *Case Studies in Construction Materials*, 18, e01917.

Hezaveh, M. M., Kanan, C., Salvaggio, C., 2017. Roof damage assessment using deep learning. *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 6403–6408.

Katartzis, A., Sahli, H., 2008. A stochastic framework for the identification of building rooftops using a single remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1), 259-271.

Li, K., Wu, Z., Peng, K.-C., Ernst, J., Fu, Y., 2018. Tell me where to look: Guided attention inference network. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9215–9223.

Li, W., Zhu, J., Pirasteh, S., Zhu, Q., Guo, Y., Luo, L., Dehbi, Y., 2024. A 3D virtual geographic environment for flood representation towards risk communication. *International Journal of Applied Earth Observation and Geoinformation*, 128, 103757.

Li, Z., Zhang, Y., Wu, H., Suzuki, S., Namiki, A., Wang, W., 2023. Design and application of a UAV autonomous inspection system for high-voltage power transmission lines. *Remote Sensing*, 15(3), 865.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft COCO: Common objects in context. D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (eds), *Computer Vision – ECCV 2014*, Springer International Publishing, 740–755.

Mandirola, M., Casarotti, C., Peloso, S., Lanese, I., Brunesi, E., Senaldi, I., 2022. Use of UAS for damage inspection and assessment of bridge infrastructures. *International Journal of Disaster Risk Reduction*, 72, 102824.

Mayer, Z., Kahn, J., Hou, Y., Götz, M., Volk, R., Schultmann, F., 2023. Deep learning approaches to building rooftop thermal bridge detection from aerial images. *Automation in Construction*, 146, 104690.

Mohajeri, N., Assouline, D., Guiboud, B., Bill, A., Gudmundsson, A., Scartezzini, J.-L., 2018. A city-scale roof shape classification using machine learning for solar energy applications. *Renewable Energy*, 121, 81-93.

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M., Stohl, A., 2019. Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific Reports*, 9(1), 16063.

Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., Chao, D., 2020. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sensing*, 12(9).

Rakha, T., Gorodetsky, A., 2020. Unmanned aerial system (UAS) applications in the built environment: Towards automated building inspection procedures using drones. *Unmanned Aerial Remote Sensing*, 239–260.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. *2017 IEEE International Conference on Computer Vision (ICCV)*, 618–626.

Shi, T., Gong, J., Hu, J., Zhi, X., Zhang, W., Zhang, Y., Zhang, P., Bao, G., 2022. Feature-enhanced centerNet for small object detection in remote sensing images. *Remote Sensing*, 14(21).

Soujanya, T., Alsalami, Z., Srinath, S., Sengupta, J., Das, A., 2024. Rooftop photovoltaic panel segmentation using improved mask region-based convolutional neural network. *2024 Second International Conference on Data Science and Information System (ICDSIS)*, 1–4.

Tian, Z., Shen, C., Chen, H., He, T., 2019. Fcos: Fully convolutional one-stage object detection. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9626–9635.

Wang, J., Chen, X., Shi, W., Jiang, W., Zhang, X., Hua, L., Liu, J., Sui, H., 2023. Rooftop PV Segmenter: A size-aware network for segmenting rooftop photovoltaic systems from high-resolution imagery. *Remote Sensing*, 15(21).

Wei, W., Cheng, Y., He, J., Zhu, X., 2024. A review of small object detection based on deep learning. *Neural Computing and Applications*, 36(12), 6283-6303.

Ying, X., Xiao, C., Li, R., He, X., Li, B., Li, Z., Wang, Y., Hu, M., Xu, Q., Lin, Z., Li, M., Zhou, S., An, W., Sheng, W., Liu, L., 2024. Visible-thermal tiny object detection: A benchmark dataset and baselines.