The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-6-2025
ISPRS, EARSeL & DGPF Joint Istanbul Workshop "Topographic Mapping from Space" dedicated to Dr. Karsten Jacobsen's 80th Birthday
29–31 January 2025, Istanbul, Türkiye

Tiny Object Detection in Super-Resolved Sentinel-2 Imagery

Christian Ayala[1], Juan Francisco Amieva[1], Pablo Vega[1], Roland Perko[2], Sebastian Aleksandrowicz[3]

[1] Tracasa Instrumental - Calle Cabárceno 6, 31621 Sarriguren, Spain - {cayala, jfamieva, pvega}@itracasa.es
[2] JOANNEUM RESEARCH - Leonhardstrasse 59, 8010 Graz, Austria - roland.perko@joanneum.at
[3] Space Research Centre of the Polish Academy of Sciences - Bartycka 18A, 00-716 Wasaw, Poland - saleksandrowicz@cbk.waw.pl

Keywords: Super-resolution, Object detection, Object density estimation, Data fusion, Sentinel-2.

Abstract

The detection of tiny objects in satellite imagery is a critical task with wide-ranging applications, including environmental monitoring, urban planning, disaster response, and the surveillance of critical transport infrastructure. Sentinel-2 satellite data, characterized by providing rich spectral information at a moderate spatial resolution (10–60m), poses significant challenges for the identification of small-scale features due to limited spatial detail and the effects of mixed pixels. This study investigates the potential of super-resolution techniques to enhance Sentinel-2 imagery for improved tiny object detection. A dataset was meticulously annotated to identify aircraft across diverse areas of interest, enabling rigorous evaluation using advanced methodologies. Detection was performed using a hybrid approach that combines a YOLOv8-based object detector and a vision-transformer-based object density estimator. The fusion of these complementary methods significantly reduces false positives, resulting in improvements in precision and F1 score. The findings underscore that super-resolved Sentinel-2 imagery offers a viable and cost-effective solution for detecting tiny objects, particularly in scenarios where access to high-resolution imagery is restricted or economically prohibitive.

1. Introduction

Globally and freely available satellite images are currently limited by their ground sampling distance (GSD), e.g. Sentinel-2 (S2) with 10-60m. Regarding object detection, the question arises of the spatial limit of object sizes where reliable detection is still possible. (Kaur and Singh, 2023) indicates small object size as one of the main problems of object detection. In this work, the limit is pushed by enhancing the resolution of the input data and by combining two complementary methods for object detection and object density estimation.

Main aim. The frame for this study is the development of a system for continuous monitoring of critical transport infrastructure. It exploits the continuous acquisition and availability of S2 imagery to detect tiny objects and proactively produce reports of activity. For this study, the tiny objects selected are aircraft. These are objects with sizes close to the GSD of a S2 image and appear in various sizes, shapes, and colors.

2. State of the Art

In the following, we give an overview of related work for all necessary subtopics for tiny object detection in S2 imagery.

2.1 Super-resolution

In recent years, an increase in computational capabilities and the development of Deep Neural Network architectures have pushed the boundaries of super-resolution (SR) techniques, enabling the generation of high-quality high-resolution images with improved accuracy and efficiency. In general, the classification of SR methods depends on the number of input images. The first category includes single image super-resolution (SISR) methods, and the second comprises multi-frame super-resolution (MISR) methods.

Among the SISR methods, the most promising are those based on machine learning techniques, especially deep learning (DL). DL-based methods can be divided into two groups: convolutional architecture-based (CNN) methods and methods based on Generative Adversarial Networks (GANs) (Zhou and Feng, 2019). GAN-based methods tend to generate high-frequency noise (Park et al., 2018), making them less useful for satellite-based image applications. For these reasons, a significant number of SR approaches based on CNNs have emerged, primarily focusing on the Red, Green, Blue, and Near Infrared bands.

The first SRCNN, which directly learns an end-to-end mapping between low- and high-resolution images, was proposed by (Dong et al., 2014). It was later redesigned and called FSRCNN (Dong et al., 2016). Since then, many other CNN-based frameworks have emerged, including ResNet (He et al., 2015a), DRNN (Kim et al., 2016b), and VDSR (Kim et al., 2016a).

A great deal of research has been conducted, demonstrating the possibility of super-resolving S2 images. For example (Galar et al., 2019) applied Enhanced Deep Residual Network and used RapidEye imagery as high-resolution reference images to obtain 5m S2. It has been later upgraded to quadruple the resolution to 2.5m (Galar et al., 2020). Some works increase the resolution even higher, like by a factor of 8 in (Wolters et al., 2023), however then objects get hallucinated. (Lanaras et al., 2018) proposed a solution to SR 20m bands of S2 using 10m bands.

2.2 Object detection

Traditional object detection relies of three stage computation phases: selection of region, extraction of features and classification (Kaur and Singh, 2023). All three steps have their drawbacks as for example necessity to inspect entire image with sliding multi-scale window or to select manually appropriate features. These drawbacks may be time consuming and computationally demanding.

Object detection algorithms based on deep learning offer a possible solution to limitations of traditional methods. Deep learning-based object detection algorithms can be categorized into two primary types. The first category covers region based detectors also known as two-stage detectors. They work in a two-stage approach: first, regions of interest are proposed and second, objects are localized and classified within those regions (Kaur and Singh, 2023). First CNN of this type was Region-Based CNN (R-CNN) (Girshick et al., 2014) and was successfully applied for small objects detection by (Chen et al., 2017). Although the method is highly accurate the main drawback is its speed. Consequently, numerous follow up R-CNN algorithms have been proposed such as Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2017) or SPP-Net (He et al., 2015b).

The second category of object detection methods are one-stage detectors. These perform both localization and classification in a single pass through the network. In this category family of You Only Look Once (YOLO) (Redmon et al., 2016, Redmon and Farhadi, 2018, Bochkovskiy et al., 2020) object detectors plays a significant role, along with RetinaNet (Lin et al., 2020). In general, single-shot detectors are considered less accurate, especially when it comes to small objects (Yin et al., 2020), but at the same time significantly faster than two-stage detectors.

2.3 Density estimation

Object density estimation emerged as a research topic in computer vision in cases when objects are very tiny or strongly overlap, such that detecting each instance is very difficult or even impossible. The main concept is to estimate the object density function whose integral over any image region holds the count of objects within this region. Classical applications are cell counting in medical imagery or person counting in terrestrial images which were initially solved by regression-based machine learning (Lempitsky and Zisserman, 2010). With the rise of deep learning, CNN-based methods were employed (cf. the review in (Perko et al., 2021) and (Li et al., 2018)), followed by vision transformers (Liang et al., 2022). The latter also estimates the center point of each object and its confidence in addition to the density value. Even though there is a lot of development within the field of computer vision, the paradigm is only little applied in remote sensing, which is addressed in (Rodriguez and Wegner, 2019, Perko et al., 2022).

2.4 Data fusion

For fusing the results of object detection and density estimation several methods exist following different goals. One option is to use the coarse density estimate to limit the search space for object detection yielding speed-up (also called context priming) (Zou et al., 2023). Another option is to combine the confidences in a probabilistic manner to re-weight the detections based on the according (local) density (Perko and Leonardis, 2010). As a result, false positives are down-weighted while true positives are up-weighted. Another version is to use the thresholded density maps to reject all detections outside this binary mask. In this case the density is used as a spatial filter to remove detections (cf. (Perko and Leonardis, 2007)).

3. Data Set

For the proof of concept, a total of 5,378 aircraft were annotated within 118 areas of interest. Figure 1 illustrates the distribution of aircraft sizes in our dataset. The sizes range from extra-small and small (up to 12 m or 16 m in length) to large and extra-large (up to 50 m in length or beyond).

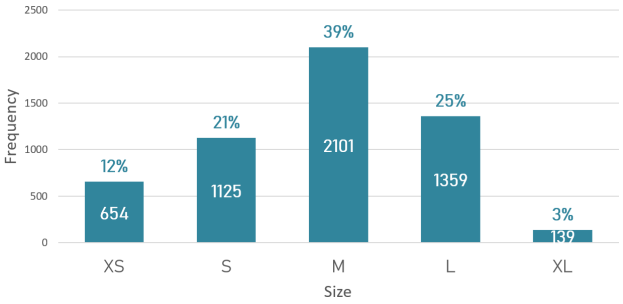


Figure 1. Distribution of aircraft sizes.

Ensuring annotation is reliable required not only the involvement of staff specialized in the matter, but also the use of very high-resolution (VHR) imagery. These are image with spatial resolution under 1 meter. The detection using low-resolution assets can be challenging, and the identification of specific models of aircraft is only possible on the largest aircraft if using S2 super-resolved imagery.

The use of VHR data introduced a new requirement. For annotation to be consistent to the ground truth of both super-resolved and VHR imagery must be the same. This means that both images should be acquired at the same time, thus ensuring the reality represented in the high-resolution image is equal to the low-resolution image. A total of 536 archived VHR satellite images were acquired on areas of interest such as airports and harbors. Figure 2 shows the differences in acquisition time between high- and low-resolution couples across the dataset. Additionally, Figure 3 represents an example of a pair of low and high-resolution images acquired at the exact same moment, which enables accurate labeling of the low-resolution dataset.

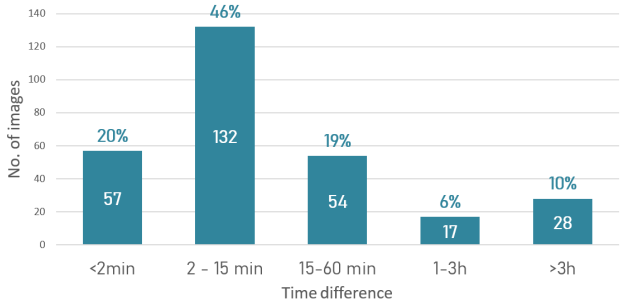


Figure 2. Differences in acquisition time between low- and high-resolution couples across the dataset.

The provision of imagery is limited to existences in the archives. A thorough archive search work was applied to ensure that acquisition times were as close as possible. As a result, up to 85% of imagery was acquired within one hour difference.

4. Methodology

The proposed fully automatic workflow for tiny object detection consists of the steps described in this section and is illustrated in Figure 4.

4.1 Super-resolution

The limited spatial resolution of S2 images makes it challenging to detect objects like aircraft, particularly smaller ones.

(a) High-resolution (b) Low-resolution

Figure 3. Both low- and high-resolution images must represent the same ground truth.

Therefore, the first step in the proposed workflow involves enhancing image sharpness using a super-resolution deep learning model for the $10m$ bands (red - B04, green - B03, blue - B02, near infrared - B08). For this purpose, an Enhanced Deep Residual Network (EDSR) (Lim et al., 2017) was trained, considering a reference high-resolution sensor as the ground truth, as proposed in (Galar et al., 2020). Specifically, we have opted for PlanetScope, as it is as similar as possible to S2 (in terms of spectral bands), while providing images at a greater spatial resolution ($3.125m$ GSD). To tailor this model for defense use cases, only images of military bases have been considered. As a result, the super-resolution model increases the spatial resolution of S2 images by a factor of 4 to $2.5m$ GSD, making it possible to accurately detect and identify military assets.

4.2 Object detection

Taking the super-resolved S2 images as input, an object detection model has been used to detect aircraft. In this regard, the You Only Look Once (YOLO) (Redmon et al., 2016) family of object detection models has been considered due to their good tradeoff between speed and accuracy. Instead of relying on sliding window approaches as traditional methods do, YOLO treats object detection as a single regression problem, predicting bounding boxes and class probabilities for objects directly from the input image in a single forward pass.

Among the key innovations of the YOLOv8 architecture are the use of spatial attention mechanisms, which help the model focus on relevant parts of the image, leading to more precise object localization, and a novel feature fusion module that effectively combines high-level semantic features with low-level spatial information, improving detection accuracy for small objects (Sohan et al., 2024). Both characteristics are of paramount importance for detecting aircraft in satellite imagery, as objects are often located in specific parts of the image following particular patterns and tend to vary in size, with the majority being small.

To address the challenge of small object detection, we utilized the Slicing Aided Hyper Inference (SAHI) method (Akyon et al., 2022), a generic slicing-aided fine-tuning pipeline that can be applied to any existing object detector. This approach enhances the small object detection performance of any current object detector without requiring fine-tuning, leveraging slicing-aided inference. By dividing input images into overlapping patches, this method effectively increases the relative pixel area of small objects in the images fed into the network, facilitating their detection.

4.3 Density estimation

Within this work, we present a custom-tailored vision transformer-based variant of (Liang et al., 2022) to estimate the object density, confidence, and location. Here features are extracted by a CNN encoder, which are then fed to a transformer encoder-decoder module with prediction heads (cf. Figure 5). The initial framework is designed for terrestrial 8-bit RGB images, such that we first quantize the super-resolved S2 16-bit reflectance also considering appropriate nodata handling. Second, bands are selected where red, green, and near infrared performed best. Finally, the model was pretrained on the NWPU-Crowd data set (Wang et al., 2020) followed by a transfer learning to our custom S2-based dataset (epochs 7500, learning rate 1e-4, crop size 256, batch size 16, queries 700).

Object density heatmaps are generated by applying Gaussian filtering on the individual detections (cf. Figure 4).

4.4 Data fusion

Although the spatial resolution of the super-resolved S2 images is four times greater than that of the original images, challenges for object detection systems may still arise depending on the size of the target objects. As noted in Section 3, the objects in our dataset vary in size from $107m^2$ to $3272m^2$. Consequently, some objects may occupy only a few pixels (e.g., 1 or 2 S2 pixels). Due to the presence of such small objects, object detection models may experience significant performance limitations. Specifically, they may either fail to identify many objects or overcompensate by detecting spurious ones, resulting in a high number of false positives.

In our research, we primarily address the latter issue by integrating the outputs of a density estimation approach as a spatial filtering step. The idea is to use the object density heatmaps generated by this approach to remove false predictions from the object detection model, thereby reducing the number of false positives while preserving most of the true positives, thus, following the paradigm in (Perko and Leonardis, 2007). Figure 4 illustrates the processing flow described above.

5. Results

Table 1 compares the performance of the object detection model alone (baseline) with the proposed data fusion approach, which filters out false positives using heatmaps generated by the density estimation model. The results show that this approach significantly reduces false detections (from 487 to 186, a ≈62% reduction) while only slightly decreasing true positives (from 507 to 481, a ≈5% reduction), thereby increasing precision (from 0.51 to 0.72, a ≈41% improvement).

This analysis can be further broken down by object size to understand the impact of the proposed method across different object scales:

- *XS*: Both methods achieve a low recall of 0.04, detecting only 3 of the 78 true objects, highlighting the significant challenge of identifying such small targets. However, the baseline method achieves a precision of 0.75 due to one false positive, while the proposed method eliminates all false positives, achieving a perfect precision of 1.00. Despite this improvement in precision, the F1 score remains

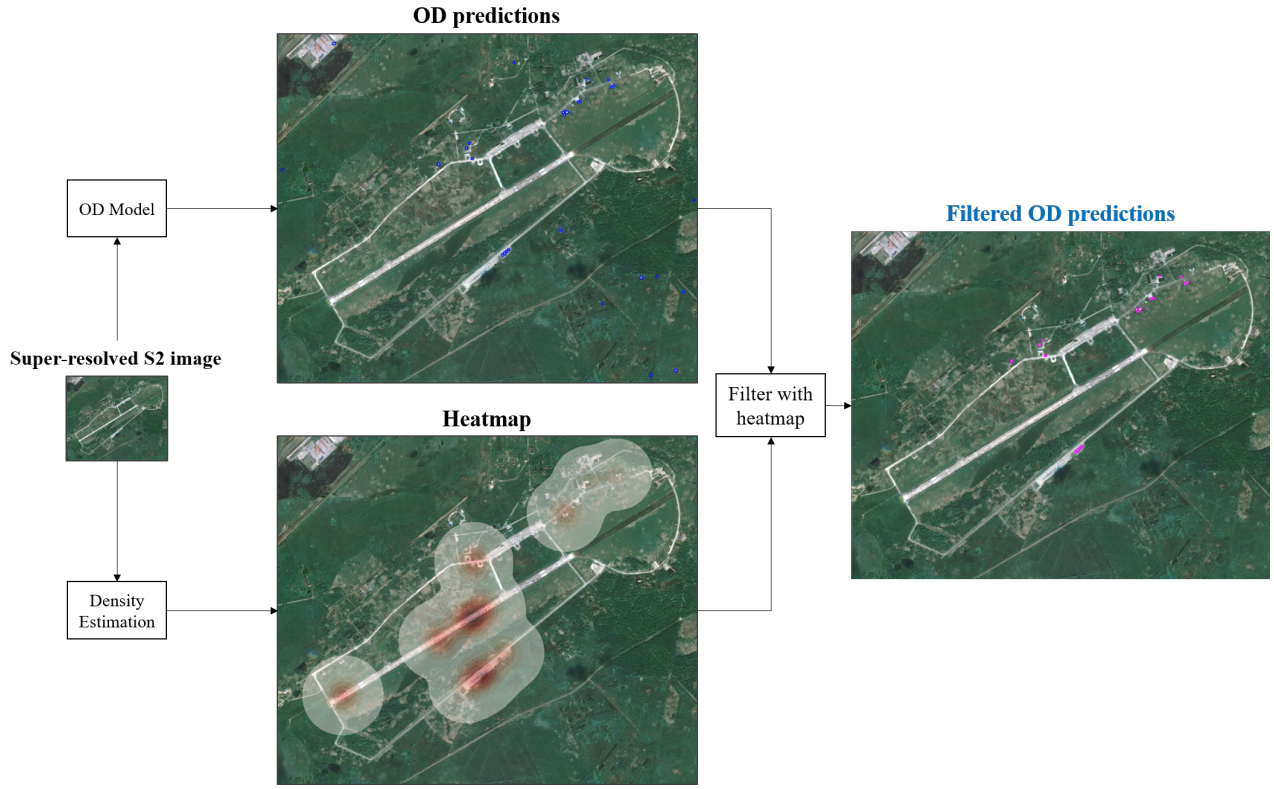


Figure 4. Proposed processing flow that filters out false predictions made by the object detection (OD) model using the object density heatmap produced by the density estimation approach.

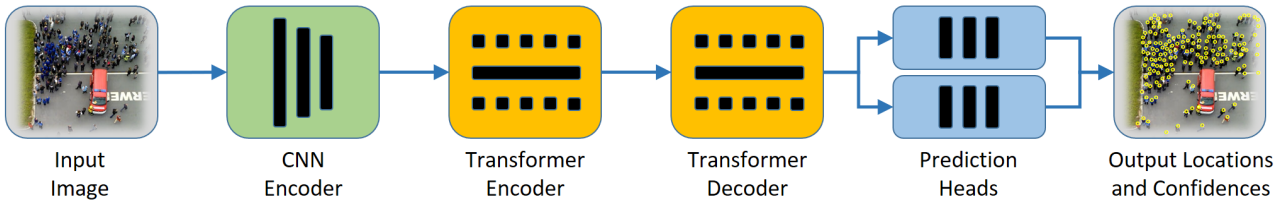


Figure 5. Deep learning architecture initially designed for crowd counting and localization, based on CNN encoder, Transformer encoder, Transformer Decoder, and prediction heads (cf. (Liang et al., 2022)). This design predicts the location of objects in the form of 2D points, together with confidences and the overall object count.

Experiment	Size	TP	FP	FN	Precision	Recall	F1
Baseline	XS	3	1	75	0.75	0.04	0.07
	S	46	183	165	0.20	0.22	0.21
	M	225	254	165	0.47	0.58	0.52
	L	202	40	30	0.83	0.87	0.85
	XL	31	10	0	0.76	1.00	0.86
	Overall	**507**	487	**435**	0.51	**0.54**	0.53
Proposed	XS	3	0	75	1.00	0.04	0.07
	S	46	58	165	0.44	0.22	0.29
	M	211	104	179	0.67	0.54	0.60
	L	192	20	40	0.91	0.83	0.86
	XL	29	3	2	0.91	0.94	0.92
	Overall	481	**186**	461	**0.72**	0.51	**0.73**

Table 1. Results obtained for the test set.

constant at 0.07 for both methods, as the low recall dominates the overall performance. This suggests that while the proposed method is effective in avoiding false detections, it does not enhance the model's ability to detect extra-small objects.

- *S*: For small objects, the proposed method demonstrates a substantial reduction in false positives (from 183 to 58), leading to a noticeable improvement in precision (from 0.20 to 0.44, a 120% increase). However, recall remains constant at 0.22, as the number of true positives remains unchanged. This trade-off highlights the effectiveness of the proposed method in filtering out false positives for small objects while maintaining detection rates.

- *M*: Medium-sized objects show a more balanced improvement. The proposed approach reduces false positives significantly (from 254 to 104) while maintaining a relatively high recall (0.54 vs. 0.58 for the baseline). Precision sees a major boost (from 0.47 to 0.67, a ≈43% improvement), and the F1 score rises from 0.52 to 0.60, reflecting better overall performance.

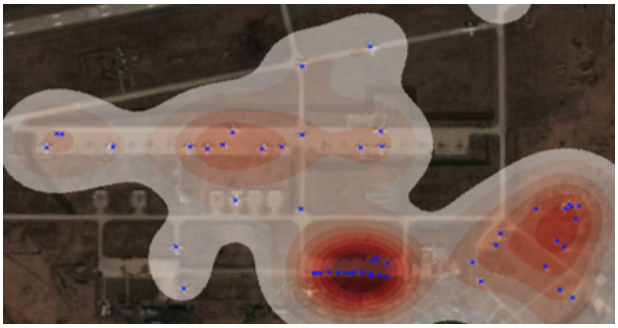
- *L*: Large objects are well-detected by both methods, but the proposed approach further enhances performance. It achieves a higher precision (0.91 vs. 0.83) with fewer false positives (20 vs. 40) and a slightly lower recall (0.83 vs. 0.87). This results in an improved F1 score (0.86 vs. 0.85), showcasing the method's ability to optimize performance for this size category.

- *XL*: For extra-large objects, both methods perform exceptionally well, but the proposed method demonstrates superior precision (0.91 vs. 0.76) and a marginally lower recall (0.94 vs. 1.00). This leads to a higher F1 score (0.92 vs. 0.86). The small number of false positives (3 vs. 10) highlights the proposed method's accuracy in detecting large, easily distinguishable objects.

In summary, the proposed approach excels particularly for medium to extra-large objects, where precision improvements are the most significant. For smaller objects (XS and S), challenges persist, as reflected by relatively low F1 scores. However, the ability to reduce false positives across all size categories, especially for small and medium objects, makes the proposed method highly effective overall.

Figure 6 shows the object detection and density estimation results for a given airbase. By looking at this figure, one can draw similar conclusions to those derived from the quantitative analysis.

(a) Object Detection results



(b) Density Estimation results

Figure 6. Object detection and density estimation results for the same airbase.

Following this approach, object detection of critical transport infrastructure with S2 imagery becomes feasible, which initially seemed unfeasible due to the high number of false predictions. This opens up new possibilities for monitoring critical transport infrastructure using freely available satellite imagery, with the associated advantages of obtaining a global track of activity in strategic areas of interest. Moreover, our method integrates the benefits of super-resolution in the downstream algorithms, enabling acceptable performance in the detection of small objects.

6. Conclusions and Future Research

The results demonstrate the efficacy of the proposed data fusion approach in improving the precision of object detection by significantly reducing false positives across all object sizes, especially for small and medium objects. This improvement comes with only a minor trade-off in recall, indicating the robustness of the method. For larger objects, where the baseline already performed well, the proposed method further refined the precision while maintaining high recall, underscoring its scalability across different detection challenges. However, the analysis also highlights persistent challenges in detecting extra-small objects, where neither the baseline nor the proposed method achieved significant improvements in recall. These findings emphasize the need for tailored solutions to address size-specific detection difficulties in real-world applications.

Future research should focus on enhancing the recall for smaller objects without compromising precision. Potential strategies include employing higher-resolution inputs, designing multi-scale detection architectures, or integrating contextual and temporal information from adjacent observations. Additionally, exploring advanced data augmentation techniques, such as leveraging synthetic or adversarial samples, could improve model robustness. Another promising research line lies in incorporating self-supervised or semi-supervised learning approaches to better utilize unlabeled data, which is often abundant but underutilized. Furthermore, developing unified evaluation metrics that consider the trade-offs between precision, recall, and application-specific priorities could provide a more comprehensive understanding of model performance and guide further innovations in the field. Lastly, we plan to extend this work by incorporating other object categories, such as helicopters and vessels, to provide a deeper understanding of the model's effectiveness across different scenarios.

Disclaimer

Funded by the European Union under Grant Agreement N. 101103622. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [DEFIS]. Neither the European Union nor the granting authority can be held responsible for them.

References

Akyon, F. C., Onur Altinuc, S., Temizel, A., 2022. Slicing aided hyper inference and fine-tuning for small object detection. *IEEE International Conference on Image Processing*, 966–970.

Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M., 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934 [cs, eess].

Chen, C., Liu, M.-Y., Tuzel, O., Xiao, J., 2017. R-CNN for small object detection. *Asian Conference on Computer Vision*, Springer, 214–230.

Dong, C., Loy, C. C., He, K., Tang, X., 2014. Learning a deep convolutional network for image super-resolution. *European Conference on Computer Vision*, 184–199.

Dong, C., Loy, C. C., Tang, X., 2016. Accelerating the super-resolution convolutional neural network. arXiv:1608.00367 [cs.CV].

Galar, M., Sesma, R., Ayala, C., Albizua, L., Aranda, C., 2020. Super-resolution of Sentinel-2 images using convolutional neural networks and real ground truth data. *Remote Sensing*, 12(18), 2941.

Galar, M., Sesma, R., Ayala, C., Aranda, C., 2019. Super-resolution for Sentinel-2 images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W16, 95-102.

Girshick, R., 2015. Fast R-CNN. *IEEE International Conference on Computer Vision*, 1440–1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

He, K., Zhang, X., Ren, S., Sun, J., 2015a. Deep residual learning for image recognition. arXiv:1512.03385 [cs.CV].

He, K., Zhang, X., Ren, S., Sun, J., 2015b. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1904–1916.

Kaur, R., Singh, S., 2023. A comprehensive review of object detection with deep learning. *Digital Signal Processing*, 132, 103812.

Kim, J., Lee, J. K., Lee, K. M., 2016a. Accurate image super-resolution using very deep convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 1646–1654.

Kim, J., Lee, J. K., Lee, K. M., 2016b. Deeply-recursive convolutional network for image super-resolution. arXiv:1511.04491 [cs.CV].

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 305-319.

Lempitsky, V., Zisserman, A., 2010. Learning to count objects in images. *Advances in Neural Information Processing Systems*, 1324–1332.

Li, Y., Zhang, X., Chen, D., 2018. CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes. *IEEE Conference on Computer Vision and Pattern Recognition*, 1091–1100.

Liang, D., Xu, W., Bai, X., 2022. An end-to-end transformer model for crowd localization. *European Conference on Computer Vision*, Springer, 38–54.

Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 136–144.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2020. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 318–327.

Park, S.-J., Son, H., Cho, S., Hong, K.-S., Lee, S., 2018. Srfeat: Single image super-resolution with feature discrimination. *European Conference on Computer Vision*, Springer, 455–471.

Perko, R., Klopschitz, M., Almer, A., Roth, P. M., 2021. Critical Aspects of Person Counting and Density Estimation. *Journal of Imaging*, 7(2).

Perko, R., Leonardis, A., 2007. Context awareness for object detection. *Workshop of the Austrian Association for Pattern Recognition*, 65–72.

Perko, R., Leonardis, A., 2010. A framework for visual-context-aware object detection in still images. *Computer Vision and Image Understanding, Special Issue on Multi-Camera and Multi-Modal Sensor Fusion*, 114(6), 700-711.

Perko, R., Mustafić, S., Almer, A., Roth, P. M., 2022. Counting everything in remote sensing – The need for benchmarks. *IEEE International Geoscience and Remote Sensing Symposium*, 5369–5372.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, real-time object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.

Redmon, J., Farhadi, A., 2018. YOLOv3: An incremental improvement. arXiv:1804.02767 [cs].

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*, 39(6), 1137–1149.

Rodriguez, A. C., Wegner, J. D., 2019. Counting the uncountable: Deep semantic density estimation from space. *German Conference on Pattern Recognition*, Springer, 351–362.

Sohan, M., Sai Ram, T., Rami Reddy, C. V., 2024. A review on YOLOv8 and its advancements. *Data Intelligence and Cognitive Informatics*, 529–545.

Wang, Q., Gao, J., Lin, W., Li, X., 2020. NWPU-Crowd: A large-scale benchmark for crowd counting and localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6), 2141–2149.

Wolters, P., Bastani, F., Kembhavi, A., 2023. Zooming out on zooming in: Advancing super-resolution for remote sensing. arXiv:2311.18082 [cs.CV].

Yin, S., Li, H., Teng, L., 2020. Airport detection based on improved faster RCNN in large scale remote sensing images. *Sensing and Imaging*, 21(1), 49.

Zhou, L., Feng, S., 2019. A review of deep learning for single image super-resolution. *International Conference on Intelligent Informatics and Biomedical Sciences)*, 139–142. ISSN: 2189-8723.

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3), 257–276.