Information technology for detecting cars on aerial imaging using a modified YOLO-OBB architecture

Volodymyr Hnatushenko[†], Vita Kashtan[*†] and Oleksii Kazymyrenko[†]

Dnipro University of Technology, Dmytra Yavornytskoho Ave 19, Dnipro, 49005, Ukraine

Abstract

This paper presents an information technology for car detection on aerial imagery using a modified YOLOv11 architecture with support for oriented bounding boxes (OBB), enabling improved accuracy in vehicle detection within complex scenes on high-resolution images. The core YOLOv11-OBB architecture integrates modules: Freq-SpaFEM, SPPF, DAGLF, AMSF2, and RAFA—designed to optimize feature extraction, refinement, and aggregation, thereby enhancing the quality of vehicle detection in high spatial resolution imagery. Experimental studies demonstrated a significant performance improvement of the proposed approach over existing object detection models according to key metrics: mAP@50 at 96.8%, Precision at 99.5%, Recall at 96.8%, and OBB IoU at 98.5%. The developed information technology can be applied in urban planning for infrastructure analysis and design, and in traffic management systems for monitoring vehicle flows to support informed decision-making processes.

Keywords

car detection, aerial imagery, YOLOv11-OBB, oriented bounding boxes, feature extraction, high-resolution images, object detection, deep learning

1. Introduction

Object detection is one of the most challenging tasks in the field of computer vision, as it involves both object classification and accurate determination of their spatial position in an image (detection). Given the rapid development of digital image processing technologies, artificial intelligence, and deep learning in particular, object detection methods have undergone significant growth, which has opened up new prospects for their application in high-precision automated systems. This task is particularly relevant in the context of processing images obtained from aerial remote sensing platforms, including uncrewed aerial vehicles (UAVs). Aerial photographs generated by UAVs are characterized by high spatial resolution, wide coverage bandwidth, and flexible imaging parameters, including platform height, sensor tilt angle, and choice of scene coverage geometry. These characteristics make UAV imagery effective in applied tasks such as automated monitoring of transport infrastructure, urban planning, emergency response, etc.

Despite the numerous advantages of aerial imagery obtained with the help of UAVs, their use in object detection tasks is accompanied by several technical difficulties. In particular, the ortho-perspective characteristic of aerial imagery ensures that objects are invariant to rotation, but at the same time reduces the expressiveness of texture and morphological features, which makes it challenging to identify and recognize objects. In addition, the scale characteristics of objects depend on the platform's flight altitude, optical properties of the sensor, and other parameters, which lead to variability in the size of target objects and their high spatial resolution. It makes it

MoDaST 2025: Modern Data Science Technologies Doctoral Consortium, June, 15, 2025, Lviv, Ukraine

[*] Corresponding author.

[†] These authors contributed equally.

vvgnat@ukr.net (V. Hnatushenko); vitalionkaa@gmail.com (V. Kashtan); kazymyrenko.o.v@nmu.one (O. Kazymyrenko)

0000-0003-3140-3788 (V. Hnatushenko); 0000-0002-0395-5895 (V. Kashtan); 0000-0001-5506-6128 (O. Kazymyrenko)

difficult to form stable feature descriptors in cases of small-scale or crowded objects. Another factor that reduces detection efficiency is the complex and heterogeneous structure of the background. One image can combine urbanized areas, natural landscapes, transportation elements, etc., which causes a high level of background noise and increases the risk of false positives during classification. Atmospheric phenomena also have a significant impact, such as cloudiness, fog, shadowy areas, or uneven lighting, which affect the quality of data analysis. They can distort the visual characteristics of objects, reducing the reliability of automated object identification [1, 2].

Vehicles, in particular cars, are of practical interest among the objects of aerial imagery, as they are of practical importance for the implementation of intelligent monitoring systems, traffic flow modeling, urban planning, and improving the efficiency of security technologies. Detecting cars on UAVs is complicated by their small size (usually 15-30 pixels), high density in transportation hubs, similarity of color and texture features to the environment (roads, parking lots), and a variety of shapes and orientations. These factors significantly complicate automated vehicle identification, which requires the development of models capable of providing accurate localization and classification of vehicles in conditions of complex backgrounds and large-scale variability.

2. Related Works

Methods for detecting vehicles in aerial images are conventionally divided into: multi-scale object detection, tiny object detection, dense object detection, and deep learning.

One of the most extensively studied areas is the detection of multi-scale objects, which is particularly relevant in vehicle detection tasks, where object size largely depends on imaging altitude, camera focal length, and other image parameters. Architectures with multi-level feature construction are widely used to address the issue of scale variability. One of the most common solutions is the Feature Pyramid Network (FPN), proposed in [3], which implements a hierarchical top-down feature construction approach with lateral connections. This method enables the generation of feature maps at different hierarchical levels, providing an adequate representation of objects at various scales and improving the accuracy of small object detection without significantly increasing computational cost. However, the study [4] highlights several limitations of the conventional FPN. In particular, the basic feature fusion approach does not account for contextual alignment, which complicates the formation of high-quality descriptors and may result in the loss of critical spatial details. Additionally, the use of traditional upsampling when merging multi-scale features can cause mismatches between spatial representations at different levels. To overcome these limitations, the authors proposed an improved approach, the Adaptive Feature Pyramid Network, which allows for more precise integration of spatial information from different levels of the feature hierarchy. Another direction in enhancing multi-scale detection involves improving classical one-stage detectors. In [5], a modification of the base Single Shot MultiBox Detector (SSD) architecture called Feature Fusion SSD (FSSD) was introduced, aimed at improving accuracy with minimal reduction in processing speed. The key innovation of FSSD lies in the introduction of a lightweight and efficient feature fusion module, where features from different network levels and scales are merged into a unified representation. This fused representation is then downsampled to form a new feature pyramid, which is passed to multiple detection blocks for prediction. This approach enhances contextual information representation and improves small object detection performance without significant computational overhead.

Tiny object detection methods are particularly relevant for the analysis of high-resolution images, where objects, such as vehicles, may occupy only a few dozen pixels. One simple yet effective solution is the adaptation of the training strategy, specifically by lowering the Intersection over Union (IoU) threshold when assigning positive and negative samples for anchor regions in Region Proposal Networks (RPNs). It allows for more flexible training sample generation, which improves the localization of small objects. In [6], a fundamentally new approach to object representation was proposed, known as RepPoints. Instead of representing an object with a fixed rectangular bounding box, RepPoints describe it as a set of learnable sampling points that

simultaneously serve for both localization and semantic representation. These points are automatically trained to spatially cover the object and capture its most significant local regions. RepPoints eliminate the need for the traditional anchor box mechanism. In [7], a practical and straightforward approach to dynamic label assignment was introduced, which takes into account the current training state based on model predictions. The method incorporates the predicted IoU scores into the positive sample selection process, enabling the selection of the most relevant examples with better alignment to ground truth objects. This approach helps reduce the discrepancy between classification and localization scores and enhances the quality of bounding boxes.

Dense object detection methods are aimed at addressing the challenges associated with high object density and compactness in images. Such conditions are often observed in aerial photographs, where objects, for example, ships in harbors, airplanes in aircraft graveyards, or vehicles in parking lots and on roads are located close to one another. In [8], the authors proposed the SCRDet++ method to improve the detection of densely and compactly arranged objects in aerial imagery. The core of the approach lies in introducing an instance-level denoising mechanism in the feature map, which enhances the detection accuracy of small and crowded objects. To address problems related to object orientation changes, a constant IoU coefficient was added to the loss function, resolving issues of angular periodicity and boundary interchangeability. In [9], a Constraint-based Fine-tuning Approach (CFA) was proposed to reduce catastrophic forgetting in incremental learning tasks without increasing the model's parametric complexity. The method adapts Average Gradient Episodic Memory (A-GEM) to the generalized few-shot object detection (G-FSOD) scenario by introducing additional constraints on gradient search. It enables the formulation of a new gradient update rule that ensures more effective knowledge alignment between base and novel classes.

Deep learning-based methods [10], particularly convolutional neural networks (CNNs), have significantly improved the effectiveness of object detection in images over the past decade [11, 12]. Two-stage architectures such as R-FCN, Faster R-CNN, and MS-Faster R-CNN [13, 14] have gained particular popularity for achieving state-of-the-art (SOTA) accuracy by first generating region proposals and then refining object boundaries and classes. In contrast, one-stage models, particularly the YOLO series [15–18], perform detection without the proposal generation stage, which provides significantly higher processing speed. These models are better suited for real-time applications but may fall short of two-stage methods in terms of accuracy when detecting small or densely clustered objects.

Despite the rapid development of object detection methods based on convolutional neural networks (CNNs), there are still limitations that significantly affect their effectiveness in aerial image object detection tasks. In particular, although modern CNN architectures exhibit rotation invariance at the feature level, traditional approaches still predominantly operate with horizontal bounding boxes (HBBs), which are unable to accurately describe oriented objects typical for aerial imagery (e.g., ships, airplanes, large vehicles). In such cases, the use of oriented bounding boxes (OBBs) [11, 19] is more appropriate, as they allow for more precise localization of objects with complex geometries, take into account their orientation, and improve the separation of densely packed instances. The implementation of OBBs effectively forms a separate task, oriented object detection, which requires modifications to both detector architectures and loss functions [20]. Another issue is the input image size constraints inherent to many existing models [21]. For instance, Faster R-CNN typically resizes the shorter image side to 600 pixels, while YOLO operates with a fixed input size of 608×608 pixels. At the same time, the typical dimensions of images from remote sensing datasets, such as ITCVD or DOTA, far exceed these limitations, necessitating additional solutions such as image tiling or scalable architectures.

This work aims to develop an information technology for vehicle detection in aerial images using a modified YOLOv11 architecture with support for oriented bounding boxes (OBB), which enables improved detection accuracy of vehicles in complex scenes within high spatial resolution imagery.

3. Proposed technology

The proposed information technology is presented in Figure 1 and is functionally structured into five key stages. Each of these stages is an integral part of the overall data processing workflow and the model development and validation process, ensuring a systematic approach to solving the task of vehicle detection in aerial images.

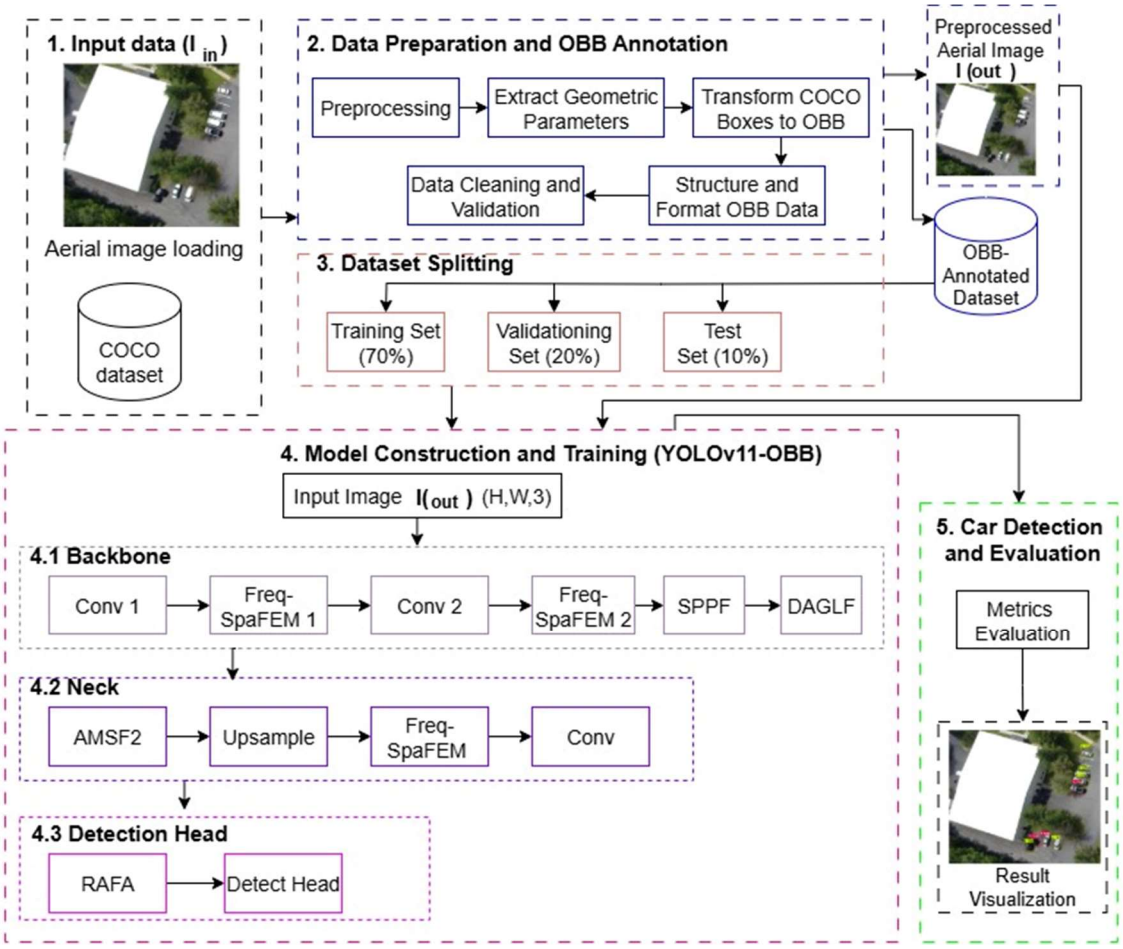

Figure 1: Diagram of the proposed information technology.

At the first step of the technology implementation, the input aerial image is loaded, and the COCO dataset [22] is used as a basic dataset with vehicle markings. Input to the information technology is an aerial image of the following resolution:

$$I \in \mathbb{R}^{H \times W \times 3}, \tag{1}$$

where H is image height, W is image width, and 3 is the number of channels (RGB).

The second step involves data preparation and OBB annotation. First, the uploaded aerial images are pre-processed [23]: normalization, resizing, contrast-limited adaptive histogram equalization (CLAHE), and color correction. To ensure the stability of neural network training, the primary data undergoes a normalization step:

$$I'_{(x,y,c)} = \frac{I_{(x,y,c)} - \mu_c}{\sigma_c}, \tag{2}$$

where $I_{(x,y,c)}$ is the intensity of a pixel in the coordinate (x,y) in channel c, μ_c, σ_c are the mean value and standard deviation of pixel values in the channel, I' is the normalized image.

To unify the input data, images are scaled to a fixed size of 640×640 pixels. Padding is used to preserve the original scene geometry to avoid distortion of the aspect ratio. Next, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance the local contrast of the image, which is especially important for aerial images where objects may have low contrast with the background due to lighting conditions or shadows. Unlike global histogram equalization, CLAHE processes the image in small, non-overlapping regions (tiles) and then uses bilinear interpolation to smooth the boundaries between tiles. Let $p(k)$ be the normalized histogram of pixel intensities for each tile, where k is the intensity level. The intensity transformation from I_{in} to I_{out} is performed using the cumulative distribution function (CDF) $C(k)$:

$$I_{out} = C(I_{in}) = \sum_{j=0}^{I_{in}} p(j). \quad (3)$$

After preprocessing, the extraction of geometric parameters is performed. For each object, an oriented bounding box is calculated:

$$B_{OBB} = (x_c, y_c, w, h, \theta), \quad (4)$$

where (x_c, y_c) are the coordinates of the box center; w,h are the width and height of the box; θ is the rotation angle of the box relative to the x-axis x.

It allows transitioning from HBB to OBB using a transformation matrix:

$$\begin{bmatrix} x_i' \\ y_i' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_i & - & x_c \\ y_i & - & y_c \end{bmatrix} + \begin{bmatrix} x_c \\ y_c \end{bmatrix}. \quad (5)$$

Next, objects are filtered based on area and overlap, incorrect labels are cleaned, and the data is saved in a format compatible with oriented bounding box processing. Finally, all OBB data are organized and formatted to create a unified OBB-annotated dataset ready for model training. The outputs of this stage are the preprocessed aerial image and the structured OBB-annotated dataset.

At the third step, data splitting is performed. The oriented bounding box dataset, consisting of 2,758 aerial images, was logically divided into three mutually independent subsets to ensure objective evaluation of the model's performance and prevent bias in the results. The training set covers 70% of the total data volume, comprising 1,931 images, and is used directly to optimize the neural network weights during training. The validation subset contains 552 images (20%) and is intended for monitoring intermediate results and tuning model hyperparameters to reduce the risk of overfitting. The test set, formed from the remaining 10% (275 images), is used exclusively at the final stage to assess the model's generalization ability independently. It allows objective determination of its performance on previously unseen data.

The fourth step is the construction of the YOLOv11-OBB model. This step is the core of the proposed information technology implementation. The model takes the image I_{out} as input. The architecture is conventionally divided into three functional components: the backbone, the neck, and the detection head. The backbone performs hierarchical feature extraction from the image after the second stage, using a sequence of convolutional layers (Conv1, Conv2) integrated with Frequency-Spatial Feature Extraction Modules Freq-SpaFEM1 and Freq-SpaFEM2. Freq-SpaFEM simultaneously extracts local spatial and global frequency characteristics, which are essential when processing data with complex textures and small objects typical of aerial images. The backbone concludes with SPPF (Spatial Pyramid Pooling Fast) [24] and DAGLF (Dual Attention Global Local Fusion) layers. SPPF enables aggregation of contextual features at different scales. At the same time, DAGLF allows simultaneous focus on local details and the global scene context, enhancing the semantic expressiveness of features and reducing the impact of background noise.

The Neck component is an intermediate link between the feature extraction block (Backbone) and the detection head, enabling the fusion of features from different network depth levels. It includes the AMSF2 module (Adaptive Multi-Scale Feature Fusion), which enhances generalization for objects of various sizes. The component also incorporates Upsample layers (to align spatial

dimensions with higher network levels), Freq-SpaFEM (to strengthen spatial-frequency consistency of feature maps), and Conv layers (for further feature refinement before passing them to the detection block).

The detection head is the final component of the architecture responsible for directly generating predictions for object detection and classification. It contains the RAFA module (Region-Aware Feature Aggregation), which enables the model to distinguish objects from the background better and consider the context around each object, as well as the Detect Head, which produces the final OBB predictions, including coordinates, dimensions, rotation angles, and forms the output oriented bounding boxes.

The final step of information technology includes the practical application of the trained model for vehicle detection and evaluation of its performance. A comprehensive analysis of the results is conducted using standard evaluation metrics [25] for object detection quality (Precision, Recall, mAP, OBB IoU) to confirm the effectiveness of the proposed technology. Additionally, detection results are visualized to demonstrate the model's ability to localize and classify vehicles in aerial images accurately.

4. Experiment

Experiments were conducted to study the effectiveness of the proposed information technology for detecting cars on aerial photographs, the results of which allowed us to formulate reasonable conclusions about the efficacy of the model.

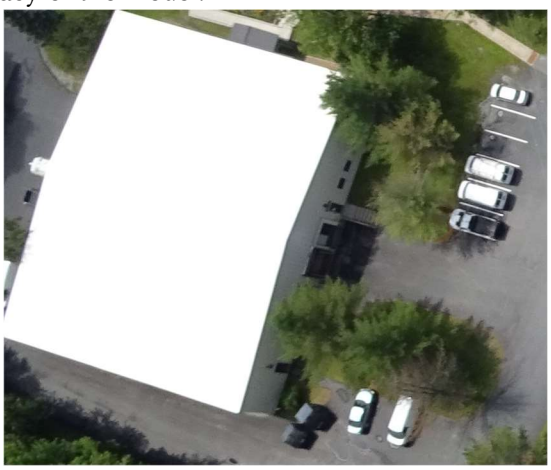

Figure 2: Aerial image after preprocessing.

Figure 2 shows a test aerial image after the preprocessing stage, which includes Contrast Limited Adaptive Histogram Equalization (CLAHE). Visual analysis demonstrates that applying CLAHE effectively enhanced the local contrast of the image, making objects, such as vehicles and infrastructure elements, more clearly visible against the background. It is especially noticeable in areas with low distinction between objects and shadows or in shaded regions. Simultaneously, white balance correction minimized color distortions during aerial imaging, providing a more natural and representative color range.

The developed YOLOv11-OBB model was trained to optimize its ability to detect vehicles in aerial images using oriented bounding boxes accurately. The training process was conducted on a training dataset consisting of 1,931 images, with a validation set of 552 images used to monitor model performance during training and promptly detect signs of overfitting. The model was trained for 300 epochs. The choice of this number of epochs was based on analyzing the dynamics of loss curves and performance metrics on the validation set, which allowed achieving performance stabilization and error minimization (Fig. 3).

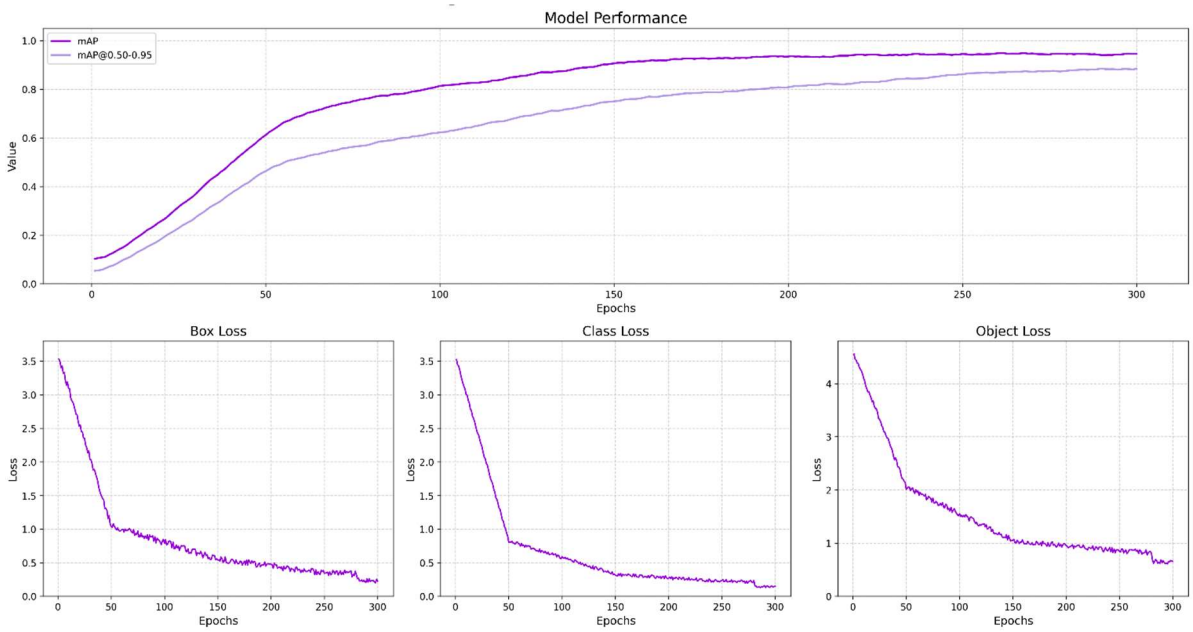

Figure 3: Model training performance and loss curves.

During training, the model was evaluated using the following metrics: mean Average Precision (mAP), mAP@0.50–0.95 (COCO metric), and loss curves. In the initial stages of training, the mAP value increases rapidly, reaching over 0.85 by the 300th epoch, indicating high localization and classification accuracy. The mAP@0.50–0.95 value steadily increases, reaching approximately 0.85 by the end of training, which confirms the model's effective generalization capability. The Box Loss curve (bounding box loss) shows a sharp decrease during the first 50 epochs and then stabilizes at around 0.25, indicating reliable spatial localization. The Class Loss curve (classification loss) similarly decreases early in training and stabilizes at approximately 0.15, reflecting effective learning of class-specific features. Objectness Loss also stabilizes at around 0.1, demonstrating high sensitivity to object presence.

The loss curves for the training and validation sets (train/val_box_loss, train/val_cls_loss, train/val_dfl_loss) show a synchronized decrease in loss functions across both subsets without significant divergence, indicating the absence of overfitting and effective generalization (Fig. 4). The precision and recall metrics (precision(B), recall(B), mAP50(B), mAP50–95(B)) demonstrate steady improvement in both accuracy and recall, with values exceeding 0.85 at the final stage of training. In particular, the mAP50 and mAP50–95 curves on the validation set show consistent growth, confirming the reliability and generalization capability of the trained model.

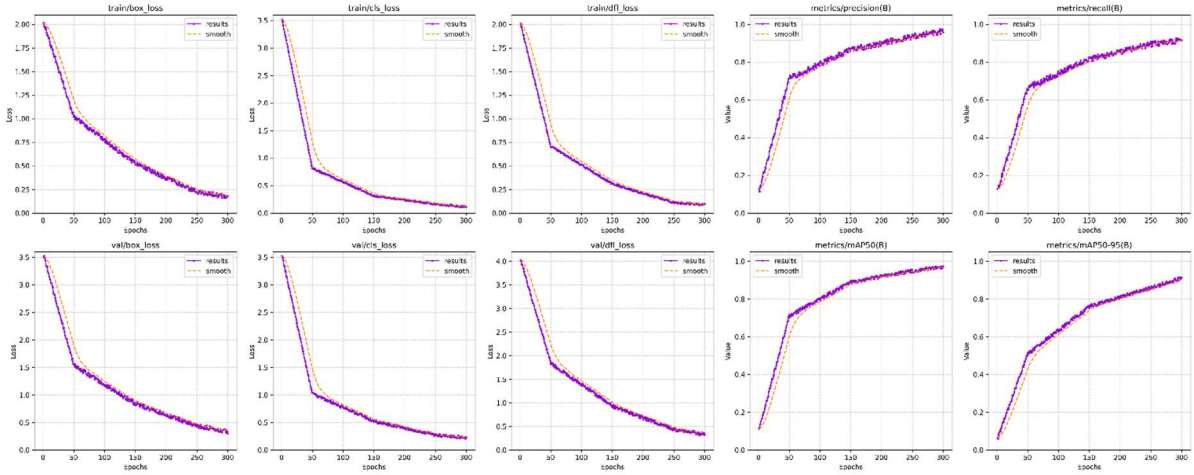

Figure 4: Detailed training and validation metrics.

5. Results and discussion

To visually demonstrate the effectiveness of the proposed information technology for vehicle detection in aerial images (Fig. 5), a visual comparison was conducted between its results and those of existing models on images from the test set, with the predicted bounding boxes overlaid.

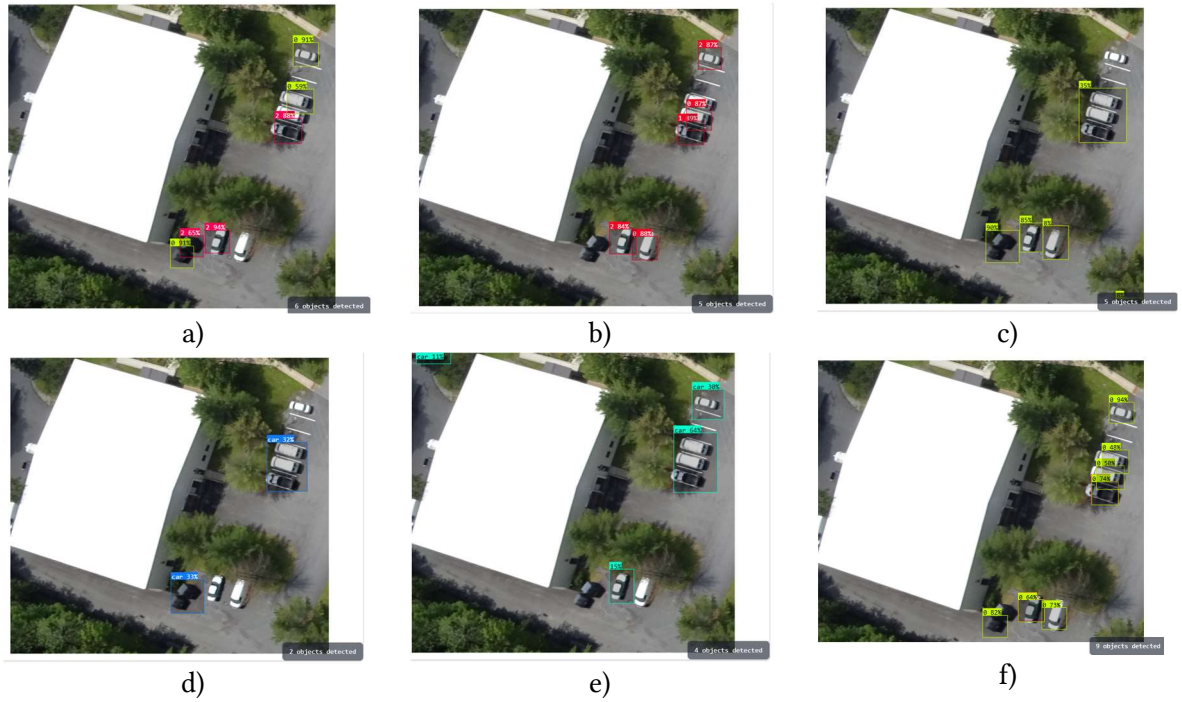

Figure 5: A comparison of the visual results of car detection by different models: a) YOLOv11; b) YOLOv8; c) YOLOv7; d) YOLOv5; e) CNN; f) a proposed model.

Many false negatives are observed in the image fragments obtained using earlier-generation CNN models, such as YOLOv5 (Fig. 5d, Fig. 5e)—some vehicles remain undetected. Additionally, there are false positives and inaccurate bounding box dimensions. The detected boxes exhibit distorted geometric properties, indicating a limited ability of such models to isolate small objects, especially those with arbitrary orientations, precisely. The results obtained using YOLOv7 / YOLOv8 models (Fig. 5b, Fig. 5c) show noticeably better performance: the number of detected vehicles increases, and the boxes more accurately correspond to the objects' positions. However, using only axis-aligned rectangles leads to less effective coverage of angled or elongated objects. In particular, there are cases of excessive overlap with adjacent objects or unnecessary inclusion of background. The visual results of the newer YOLOv11 model (Fig. 5a) indicate a further improvement in detection accuracy compared to YOLOv7/YOLOv8. Nonetheless, in cases where objects are significantly oriented for the image axis, spatial inaccuracies still occur due to the mismatch between the bounding box geometry and the actual shape of the vehicle. The results of the proposed YOLOv11-OBB model (Fig. 5f) demonstrate the highest detection accuracy for vehicles in aerial images. All cars in the image were successfully identified, and the bounding boxes' shape aligns with the objects' actual geometry. OBB minimizes unnecessary background coverage and reduces the likelihood of overlap with adjacent objects, which is particularly critical when vehicles are densely packed.

For the quantitative analysis of the effectiveness of the developed information technology for vehicle detection in aerial images, a comparison with existing models was conducted using the following evaluation metrics (Fig. 6): mean Average Precision at a 50% intersection-over-union threshold (mAP@50), classification Precision, Recall, and mean spatial overlap accuracy for oriented bounding boxes (OBB IoU).

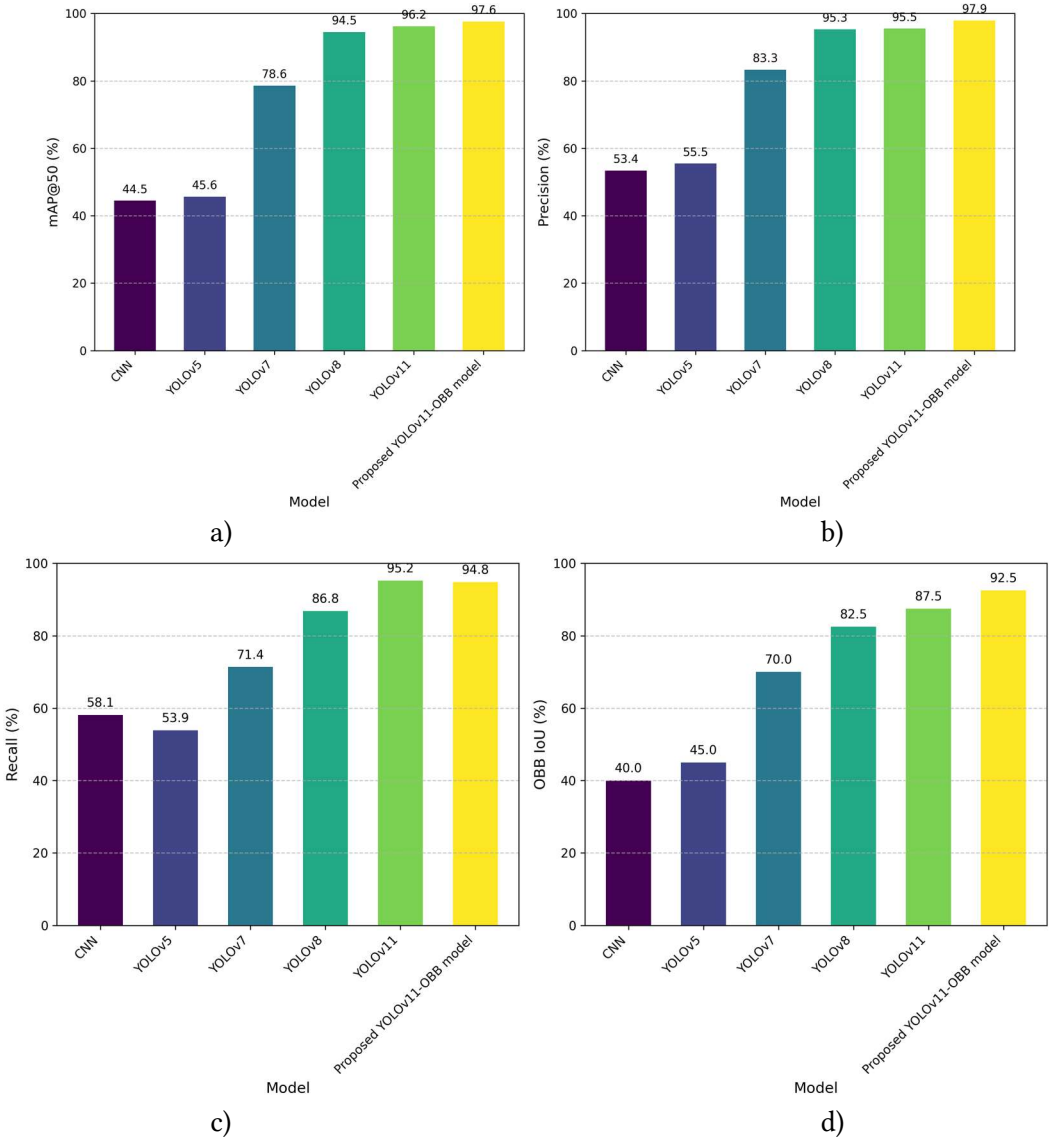

Figure 6: Performance metrics comparison of object detection models: a) mAP@50 (%); b) Precision (%); c) Recall (%); d) OBB IoU (%).

The proposed YOLOv11-OBB model demonstrates the highest mean Average Precision (mAP@50) of 96.8%, surpassing the results of other compared models (CNN – 44.5%, YOLOv5 – 45.6%, YOLOv7 – 78.6%, YOLOv8 – 94.5%, YOLOv11 – 95.8%). This result indicates the model's strong ability to detect vehicles with sufficient overlap accurately, achieved through enhanced architecture and oriented bounding boxes. The Precision metric, which reflects the model's ability to minimize false positives, reaches 99.5% for the proposed model, significantly outperforming (CNN – 53.4%, YOLOv5 – 55.5%, YOLOv7 – 83.3%, YOLOv8 – 93.7%, YOLOv11 – 95.9%). The Recall metric demonstrates a value of 96.8% for YOLOv11-OBB, the highest among the compared models (CNN – 58.1%, YOLOv5 – 53.9%, YOLOv7 – 71.4%, YOLOv8 – 86.8%, YOLOv11 – 94.2%), indicating high sensitivity in detecting vehicles with minimal missed detections. The localization quality for oriented bounding boxes, measured by OBB IoU, reaches 92.5% for YOLOv11-OBB, significantly higher than CNN – 40.0%, YOLOv5 – 45.0%, YOLOv7 – 70.0%, YOLOv8 – 82.5%, YOLOv11 – 87.0%. This high OBB IoU score confirms the effectiveness of oriented bounding boxes and specialized architectural components (Freq-SpaFEM, DAGLF, AMSF2, RAFA), contributing to accurate feature extraction and detection of objects with complex geometries. It is especially critical in densely packed vehicles, overlaps, or arbitrary orientations typical in aerial imagery.

To evaluate each model's ability to classify and identify objects, confusion matrices correctly [26] were constructed for the test dataset (Fig. 7). The resulting matrices allow for assessing the

number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for two classes: "Car" and "Non-car" (background/other objects).

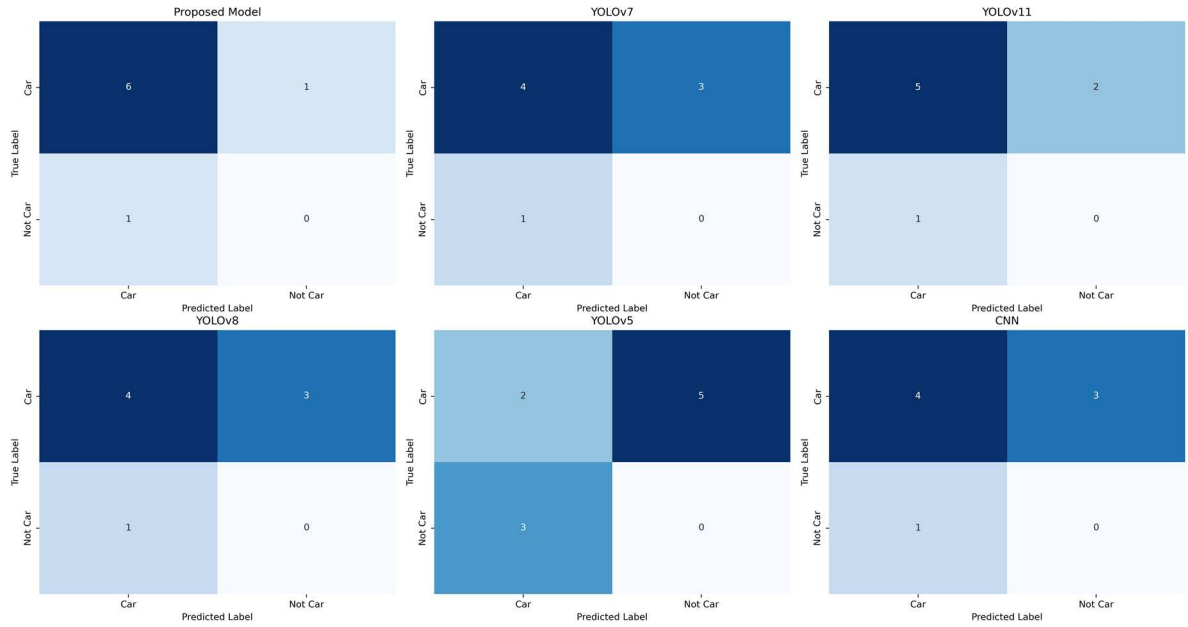

Figure 7: Confusion Matrices for object detection models.

The proposed model (YOLOv11-OBB) exhibits only one false positive case (an object incorrectly classified as a car) and one false negative case (a vehicle that was not detected), indicating high accuracy with a minimal number of errors. YOLOv7 and YOLOv8 detect fewer true cars and produce more false positive detections, reflecting lower precision than the proposed model. YOLOv11 detects more cars than YOLOv7 and has fewer false positives; however, it still underperforms relative to the proposed model regarding correctly identified vehicles and exhibits twice as many false positives. YOLOv5 shows significantly poorer results, detecting only two true cars, with the highest number of false positive detections and a substantial count of missed vehicles. The performance of the baseline CNN is comparable to that of YOLOv7 and YOLOv8 according to this confusion matrix.

6. Conclusions

A car detection information technology based on the modified YOLOv11 architecture supporting oriented bounding boxes has been developed and experimentally evaluated. The quantitative and visual analysis results demonstrate the proposed technology's effectiveness. Specifically, the modified YOLOv11-OBB model achieved the highest performance among the compared counterparts across key metrics: mAP@50 (96.8%), Precision (99.5%), Recall (96.8%), and OBB IoU (98.5%). The significant outperformance of other models, including the standard YOLOv11, confirms the feasibility and advantages of integrating oriented bounding boxes and optimized architectural components such as Freq-SpaFEM, DAGLF, AMSF2, and RAFA. These modifications enabled the model to effectively address challenges related to small object sizes, arbitrary orientations, and complex backgrounds, which are typical for aerial imagery.

The analysis of the confusion matrices further confirmed the high accuracy of the proposed YOLOv11-OBB model, as it achieved the highest number of accurate positive detections while simultaneously minimizing both false positives and false negatives. This result demonstrates the model's ability to identify vehicles accurately using high-resolution imagery. Compared to other models, YOLOv11-OBB significantly reduces the number of errors, which is particularly important for practical applications such as traffic monitoring and urban planning.

Acknowledgements

The study was conducted as part of the international educational project “Safe Artificial Intelligence: The European Legal Dimension” [101176092, a joint project of Dnipro University of Technology, Erasmus+ Jean Monnet Foundation, and the European Education and Culture Executive Agency (EACEA)]. Support from the European Commission for the publication of this work does not imply endorsement of its content, which solely reflects the views and opinions of the authors, and the Commission cannot be held responsible for any use that may be made of the information contained therein. 

Declaration on Generative AI

The authors used Grammarly to check the grammar.

References

[1] B. Wang, B. Xu, A feature fusion deep-projection convolution neural network for vehicle detection in aerial images, PLOS ONE 16.5 (2021) e0250782. doi:10.1371/journal.pone.0250782.

[2] V. V. Hnatushenko, P. I. Kogut, M. V. Uvarov, On Optimal 2-D Domain Segmentation Problem via Piecewise Smooth Approximation of Selective Target Mappings, J. Optim., Differ. Equ. Their Appl. 27.2 (2019) 60. doi:10.15421/141908.

[3] C. Wang, C. Zhong. "Adaptive Feature Pyramid Networks for Object Detection". IEEE Access 9 (2021) 107024–107032. doi:10.1109/access.2021.3100369.

[4] K. Sunil, S. Beniwal, Sentiment Analysis: A Tool for Mining Opinions and Emotions, SSRN Electron. J. (2020). doi:10.2139/ssrn.3746951.

[5] Li, Zuoxin, Fuqiang Zhou. "FSSD: Feature Fusion Single Shot Multibox Detector". ArXiv abs/1712.00960 (2017): n. pag.

[6] Z. Yang, S. Liu, H. Hu, L. Wang, S. Lin. "RepPoints: Point Set Representation for Object Detection". IEEE/CVF International Conference on Computer Vision (ICCV), 2019. doi:10.1109/iccv.2019.00975.

[7] T. Zhang, B. Luo, A. Sharda, G. Wang, Dynamic Label Assignment for Object Detection by Combining Predicted IoUs and Anchor IoUs, J. Imaging 8.7 (2022) 193. doi:10.3390/jimaging8070193.

[8] X. Yang, J. Yan, W. Liao, X. Yang, J. Tang, T. He. "SCRDet++: Detecting Small, Cluttered and Rotated Objects via Instance-Level Feature Denoising and Rotation Loss Smoothing". IEEE Trans. Pattern Anal. Mach. Intell. (2022) 1. doi:10.1109/tpami.2022.3166956.

[9] K. Guirguis, A. Hendawy, G. Eskandar, M. Abdelsamad, M. Kayser, J. Beyerer. "CFA: Constraint-based Finetuning Approach for Generalized Few-Shot Object Detection". IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (2022). doi:10.1109/cvprw56347.2022.00449.

[10] V. Kashtan, V. Hnatushenko, Deep Learning Technology for Automatic Burned Area Extraction Using Satellite High Spatial Resolution Images, in: Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making, Springer International Publishing, Cham, 2022, pp. 664–685. doi:10.1007/978-3-031-16203-9_37.

[11] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren. "Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression". Proc. AAAI Conf. Artif. Intell. 34.07 (2020) 12993–13000. doi:10.1609/aaai.v34i07.6999.

[12] V. Y. Kashtan, V. V. Hnatushenko, Automated building damage detection on digital imagery using machine learning, Nauk. Visnyk Natsionalnoho Hirnychoho Universytetu No. 6 (2023) 134–140. doi:10.33271/nvngu/2023-6/134.

[13] Z. Li, Y. Chen, G. Yu, Y. Deng. "R-FCN++: Towards Accurate Region-Based Fully Convolutional Networks for Object Detection". Proc. AAAI Conf. Artif. Intell. 32.1 (2018). doi:10.1609/aaai.v32i1.12265.
[14] D. Avola, L. Cinque, A. Diko, A. Fagioli, G. L. Foresti, A. Mecca, D. Pannone, C. Piciarelli, MS-Faster R-CNN: Multi-Stream Backbone for Improved Faster R-CNN Object Detection and Aerial Tracking from UAV Images, Remote Sens. 13.9 (2021) 1670. doi:10.3390/rs13091670.
[15] A. Bochkovskiy, C. Wang, H. Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection". ArXiv abs/2004.10934 (2020): n. pag.
[16] K. Li, X. Zheng, J. Bi, G. Zhang, Y. Cui, T. Lei, RMVAD-YOLO: A Robust Multi-View Aircraft Detection Model for Imbalanced and Similar Classes, Remote Sens. 17.6 (2025) 1001. doi:10.3390/rs17061001.
[17] S. Patil, S. Waghule, S. Waje, P. Pawar, S. Domb, Efficient Object Detection with YOLO: A Comprehensive Guide, Int. J. Adv. Res. Sci., Commun. Technol. (2024) 519–531. doi:10.48175/ijarsct-18483.
[18] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, L. Zhang. "Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions". Proc. AAAI Conf. Artif. Intell. 36.2 (2022) 1792–1800. doi:10.1609/aaai.v36i2.20072.
[19] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, L. Zhang. "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images". IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2018). doi:10.1109/cvpr.2018.00418.
[20] J. Ding, N. Xue, Y. Long, G.-S. Xia, Q. Lu. "Learning RoI Transformer for Oriented Object Detection in Aerial Images". IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2019). doi:10.1109/cvpr.2019.00296.
[21] V. Hnatushenko, P. Kogut, M. Uvarov, On Flexible Co-registration of Optical and SAR Satellite Images, in: Advances in Intelligent Systems and Computing, Springer International Publishing, Cham, 2020, pp. 515–534. doi:10.1007/978-3-030-54215-3_33.
[22] The Common Objects in Context-stuff. URL: https://paperswithcode.com/dataset/coco-stuff.
[23] G. Tariku, I. Ghiglieno, A. Simonetto, F. Gentilin, S. Armiraglio, G. Gilioli, I. Serina, Advanced Image Preprocessing and Integrated Modeling for UAV Plant Image Classification, Drones 8.11 (2024) 645. doi:10.3390/drones8110645.
[24] K. Xia, Z. Lv, C. Zhou, G. Gu, Z. Zhao, K. Liu, Z. Li, Mixed Receptive Fields Augmented YOLO with Multi-Path Spatial Pyramid Pooling for Steel Surface Defect Detection, Sensors 23.11 (2023) 5114. doi:10.3390/s23115114.
[25] V. Hnatushenko, D. Mozgovoy, V. Vasyliev. "Accuracy evaluation of automated object recognition using multispectral aerial images and neural network". Tenth International Conference on Digital Image Processing (ICDIP 2018), Shanghai, China, SPIE, 2018, p. 72. doi: 10.1117/12.2502905.
[26] A. Vijayakumar, S. Vairavasundaram, YOLO-based Object Detection Models: A Review and its Applications, Multimed. Tools Appl. (2024). doi:10.1007/s11042-024-18872-y.