

Robust Vehicle Detection Framework Using YOLO12 with Advanced Augmentations for Adverse Weather Conditions

Huy Hoang Nguyen[1], Huong Diem Thi Nguyen[2], and Vinh Dinh Nguyen[1(✉)]

[1] Department of Information Technology, FPT University, Can Tho, Vietnam
huynhce182126@fpt.edu.vn, vinhnd29@fe.edu.vn
[2] IVS Joint Stock Company, Can Tho, Vietnam

Abstract. Accurate vehicle detection is an enormous challenge for intelligent transportation systems in adverse weather conditions. The occlusions, poor visibility and data imbalance across vehicle classes have caused these problems. Therefore, we present a strong object detection pipeline based on YOLO12, the latest YOLO cycle of technology, coupled with targeted data augmentation strategies. It is expected to increase detection performance in these complex events. The pipeline includes geometric and chromatic augmentations to reproduce spatial distortions or variations in illumination conditions, improving model robustness under a wide range of real-world scenarios. A CutMix-based augmentation strategy is also proposed that minimises severe imbalances among classes by artificially boosting minority instances of a minority class while maintaining spatial coherence. After running extensive experiments on the DAWN dataset, which includes vehicle images in the presence of snow, fog, rain, and sandstorm, our training pipeline driven by the augmentation algorithm substantially improves detection accuracy. The proposed method is found to obtain an mAP50 of 0.979 and mAP50–95 of 0.901 on the augmented dataset, significantly outperforming the baseline scores of 0.652 and 0.405, respectively. These results underline a fundamental role for the augmentation strategies when designed carefully, to augment the precision and generalization of object detection models that are deployed in hostile environmental settings.

Keywords: Object Detection · YOLO12 · Data Augmentation · CutMix · Class Imbalance · Vehicle Detection · Adverse Weather · Intelligent Transportation Systems

1 Introduction

The challenge of reliably detecting vehicles in severe weather continues to be an issue for both Intelligent Transport Systems and Autonomous Driving Applications. Weather events including; fog, rain, snow, and sand will all cause

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16768, pp. 264–276, 2027.
https://doi.org/10.1007/978-3-032-30488-9_17

reduced visual clarity, distorted appearances of objects and lower fidelity of detection [1]. All of these issues are critical to several highly safety-related applications including; Automated Traffic Monitoring, Collision Avoidance and Road Infrastructure Management. While recent advancements in Deep Learning have made significant improvements in detection accuracy using Convolutional Neural Networks (CNNs), Transformer Based Architectures and other variants, there continue to be requirements for developing stronger Model Architectures and Data Preparation Procedures to produce consistent detection performance under various weather conditions. Among the more recently developed Models, the YOLO (You Only Look Once) Family has demonstrated a number of desirable attributes; including a single stage approach to Object Detection that produces a good trade-off between Speed and Accuracy. These advantages were further enhanced by the introduction of the Area Attention and R-ELAN Modules into the YOLO12 Model that improves the ability of the Model to extract relevant features from Images in Complicated Conditions [2]. Advances in Architecture continue but the most pressing issue related to Vehicle Datasets is Extreme Class Imbalance within those datasets. This imbalance causes a natural bias toward prediction of Majority Classes [3] and is exacerbated when large numbers of different types of Vehicles exist within the dataset due to their inherently skewed distributions. Data Augmentation has shown significant benefits over the past few years to enhance the generalization and robustness of models under varying lighting conditions, viewing angles and Environments. Both geometric (rotation, flip, shear) and chromatic (brightness, hue, saturation) augmentations can simulate a range of imaging variations present in Real World Imaging, thereby enhancing Model Resilience [4]. Additionally, CutMix was introduced as a way to combine regions and labels to generate New Training Samples and represents one of the latest methods of Data Augmentation that has shown to be superior at addressing class imbalance [5].

Here, we present a solid vehicle detection framework using YOLO12 as the benchmark and various geometric, chromatic, and CutMix augmentations to boost detection performance in different weather states. Our contributions are summarized as follows:

- We create an augmentation-based training pipeline, including YOLO12 with geometric, chromatic, and CutMix augmentations, to increase vehicle detection robustness under adverse weather.
- For the specific minority vehicle classes, we propose a targeted CutMix approach to alleviate severe class imbalance to improve detection accuracy across all categories.
- We offer detailed evaluations of the DAWN dataset and show improvement in both mAP and generalization capability over non-augmented baselines.

The rest of this report is divided into the following sections: In Sect. 2, we review the existing work on detecting vehicles under difficult weather conditions. We then describe our proposed methodology, which includes a description of the data set, the various forms of augmentation applied to that data set, and the architecture used for our deep neural network models in Sect. 3. Experimental

results are shown in Sect. 4, where we evaluate how different methods to balance classes affect performance. Finally, in Sect. 5, we conclude the paper and provide a summary of possible future areas of study.

2 Related Work

The vehicle detection can go down quickly in rain, fog, snow, or sand, which contributes to an ongoing problem in robust vehicle detection in such environments for autonomous vehicles and smart traffic systems. In recent years, the emphasis has been on improving the YOLO-based framework by adapting the architecture and the image processing methods in order to attain high precision and reliability in these environments.

Rjoub et al. (2021) combined federated learning with YOLO CNN to model objects in snowy landscapes to yield superior results to centralized and gossip algorithms on the CADC dataset, suggesting the potential of federated YOLO to be deployed under severe weather conditions [6]. Wu et al. (2021) proposed a YOLO-based system for vehicle classification and counting, with a high accuracy in motorcycle, small, and large vehicle classification on a wide variety of road designs, contributing to traffic flow analysis applications [7]. Boppana et al. (2022) presented a comparative research of YOLO variations for vehicle detection in extreme weather using the AAU Rain Snow dataset. Their results demonstrate YOLOv3's advantage over the other variants when it comes to precision and mAP and thus sets that it is a baseline for adverse weather detection [8]. Liu et al. (2022) have proposed Image-Adaptive YOLO (IA-YOLO), which includes a differentiable image processing module integrated with YOLOv3 and significantly advanced detection on fog and low-light areas [9]. Ding et al. (2023) also recently proposed CF-YOLO by incorporating Cross Fusion block in YOLOv5s trained on their RSOD real-snow dataset. Their model was able to better suppress snow distortion. Their system was able to handle snow-induced distortions better than state-of-the-art techniques [10].

Recently, Tian and Kim (2023) enhanced detection of vehicle detection by combining weather classification and Faster R-CNN and Dark Channel Prior preprocessing which improved fog and rain image detection over fog and rain images [11]. Panja et al. (2024) evaluated YOLOv8 models using an augmentation technique on the DAWN dataset. They discovered that for improved mAP, horizontal flip augmentation increased the output, and YOLOv8m performed at best 71.2%, and that the data augmentation can have advantages in enhancing the robustness of the detection [12]. Aloufi et al. (2024) proposed a multi-objective model combining weather classification, SRGAN enhancement, and the YOLOv5 for simultaneous weather recognition and object detection; they achieved a mAP rate of 74.6% under severe weather conditions [13]. Ogino et al. (2025) introduced an ERUP-YOLO that utilizes differentiable Bezier and kernel-based preprocessing filters in YOLOv3 and improved detection robustness while detecting fog or low light using adaptive image enhancement [14].

Lastly, Huang and Lin (2025) presented a domain adaptation method combining adversarial and semi-supervised models and using YOLO to enhance the

performance of detection across illumination scenarios and weather domains [15]. These studies are collectively illustrating the ongoing development of YOLO-based techniques for the adverse weather vehicle detection where the results of domain adaptation, image enhancement, and integrated frameworks are used to validate the effectiveness of these approaches. Yet achieving real-time generalization across a variety of unseen weather scenarios and the computational efficiency for the model remain challenges (Table 1).

Table 1. Summary of Recent Advances in Vehicle Detection under Adverse Weather Conditions.

Year	Authors	Dataset	Method	Performance	Limitations
2021	Rjoub et al. [6]	CADC	Federated YOLO CNN	Improved over centralized	Limited to snowy conditions
2021	Wu et al. [7]	Traffic surveillance images	YOLO-based vehicle counting	Accurate counts in intersections	Focused only on classification and counting
2022	Boppana et al. [8]	AAU Rain Snow	YOLOv3 vs other YOLOs	YOLOv3 best precision, mAP	No enhancement methods explored
2022	Liu et al. [9]	Foggy and low-light images	Image-Adaptive YOLO (IA-YOLO)	Improved detection in adverse weather	Requires end-to-end joint training
2023	Ding et al. [10]	RSOD (real snow)	CF-YOLO with Cross Fusion block	Outperformed SOTA in snow	Limited to snowy conditions
2023	Tian and Kim [11]	Various weather images	Weather classification + Faster R-CNN + DCP	Improved detection in fog/rain	Complex multi-stage pipeline
2024	Panja et al. [12]	DAWN	YOLOv8n/s/m + flip augmentation	mAP up to 71.2% (YOLOv8m)	Tested only flip augmentation
2024	Aloufi et al. [13]	DAWN (augmented)	Multi-objective model (weather classification + SRGAN + YOLOv5)	mAP 74.6%	High computational complexity
2025	Ogino et al. [14]	Adverse weather datasets	ERUP-YOLO (BPW + KBL filters + YOLOv3)	Superior detection in fog, low-light	Requires differentiable filter design
2025	Huang and Lin [15]	Public + private datasets	Adversarial + semi-supervised YOLO	Outperformed prior methods across domains	Needs unlabeled data for semi-supervised training

3 Methodology

3.1 Data Collection

We analyzed the DAWN dataset [16], a free vehicle dataset for object detection in harsh weather conditions, i.e. fog, rain, sandstorm, and snow, at different intensities, to assess the model robustness. DAWN consists of six annotated classes: bicycle, bus, car, motorcycle, person, and truck. Every picture has bounding boxes and class labels, and this allows to measure detection and classification of vehicles under challenging conditions. We divide our data-set into training (70%), validation (20%) and testing (10%) sets. All of our augmentation methods were used solely on our training sample so that we avoid a case of "data leak". None of our synthetic samples will be included in either of our validation or test sets. Only for our training phase have we artificially created all of our synthetic samples via CutMix and all of the rest of our augmentation techniques (Fig. 1).



Fig. 1. Example images from DAWN showing vehicle classes under different weather conditions.

3.2 Data Preprocessing and Augmentation

Comprehensive data preprocessing and augmentation strategies were applied to improve YOLO12's performance and robustness across diverse weather scenarios.

Data Preprocessing: All images from the data set are resized to a pixel resolution of $640 \cdot 640$, so that they can be processed by the YOLO12 network and to keep annotations consistent. The scaling of corresponding bounding-box coordinates was done proportionally to prevent loss of spatial accuracy when the image is resized. By making this conversion, it allows for the use of fixed-size mini-batches and helps to ensure that each model reaches a state of convergence during training.

Data Augmentation: To increase a models ability to generalize as well as to be more robust in varying environments, there are 2 primary ways that data can be augmented: geometrically (spatial) or chromatically (color).

1. Geometric Augmentations: These types of augmentations add an element of variance to the location of objects. This allows the model to find features that do not rely on object orientation or position as shown in Fig. 2.

- **Horizontally Flipping the Image** This type of flip is used to create left right images to show how the model will see left and right lanes when it looks at them from different locations.
- **Rotating Images** The images are rotated by random amounts within $\pm 10°$ to account for camera tilt variations.
- **Shearing Images** Images are subjected to horizontal and vertical shear of $\pm 5°$ to allow the model to take into consideration the distortion that comes with looking down a road.

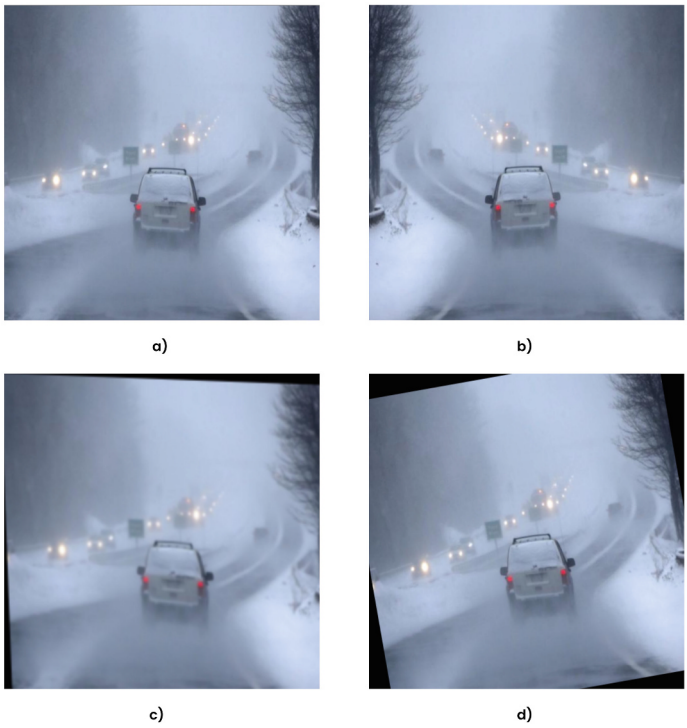

Fig. 2. Geometric augmentations: (a) original, (b) flipped, (c) sheared, (d) rotated.

2. Chromatic Augmentations. Chromatic augmentation represents random lighting conditions and environmental variations (weather), etc., as well as the effects of image degradation due to sensor limitations (such as blurring) as illustrated in Fig. 3. The chromatic augmentation techniques include:

1. Hue, Saturation, Brightness: Representing environmental changes through a random adjustment to hue, saturation and brightness.

2. Blur: Simulating defocusing, rain, fog, etc. Using a Gaussian filter.

3. Noise Addition: Adding white Gaussian noise that models sensor noise and compression artifacts.

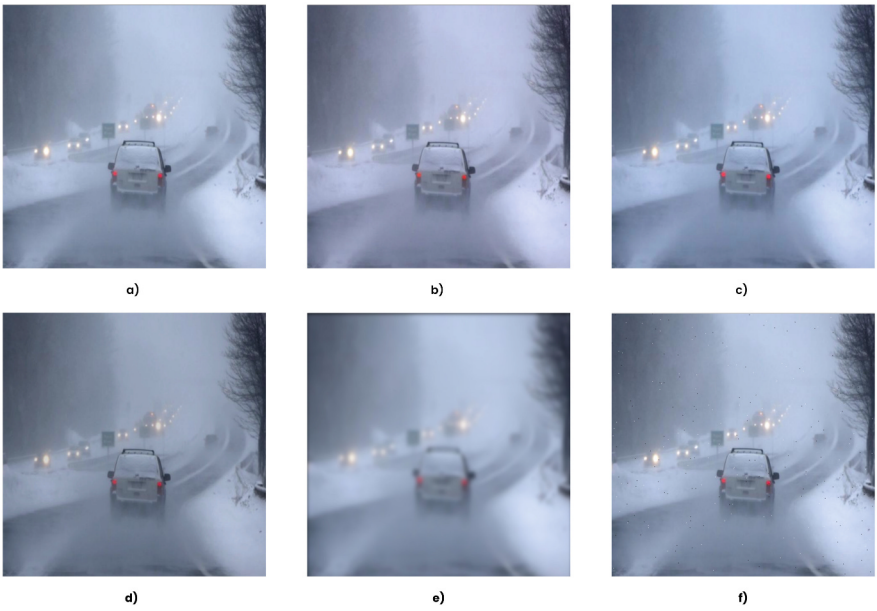

Fig. 3. Chromatic augmentations: (a) original, (b) hue adjusted, (c) saturation adjusted, (d) brightness adjusted, (e) blurred, (f) noise added.

3. A Method to Increase Minority Class Representation Using Cutmix. To solve the problem of large class imbalances in our data set, we employed an augmentation technique based on cut mix. We used this method specifically for those classes where there is too little data (i.e., the minority classes). For example, if one image contains vehicles from a majority class and another image includes vehicles from a minority class, we would use the cut mix method to extract patches (bounding boxes) of vehicles belonging to the minority class and paste them into a new target image that contained vehicles belonging to some other class than the minority class.

We ensured that all of the images were first normalized to have the same size (640×640) so as to ensure consistent resizing prior to the application of the augmentation method. Also, when placing the patches of the minority class object onto a target image that already had objects drawn on it (to preserve annotations), care was taken not to place the patch over an area of an object that was already annotated. In addition, the number of times we applied each iteration of the augmentation was limited by both the ratio at which the minority classes were represented in relation to the majority classes after each iteration, or by reaching an augmentation limit.

By applying targeted cut mix augmentations we were able to increase the amount of data available for each minority class, and therefore enable the models

to learn more discriminant features for these classes. However, given the degree of difference in representation in our original data set, we could never achieve perfect distribution among classes.

3.3 Model Architecture

YOLO12. YOLO12 [2], developed in 2025, is chosen for the base detection because of its good trade-off between high accuracy and real-time performance. As a single-stage detector, it maintains the efficiency of the YOLO family while introducing a number of important improvements. More precisely, Area Attention was proposed by YOLO12, which segments feature maps into regions for self-attention while limiting the computational cost. Moreover, it utilizes the R-ELAN architecture to leverage residual connections and bottleneck modules for feature aggregation, thus improving the stability of the training. Also, the model integrates optimized attention mechanisms, i.e., FlashAttention implementation, elimination of positional encoding, reduced MLP ratios as well as attention depth, and the incorporation of a 7×7 separable convolution that implicitly encodes positional information. YOLO12 covers multiple capabilities, such as segmentation and classification, pose estimation, and oriented object detection in addition to detection. In addition to its parameter-efficient structure, it allows flexible deployment for a wide spectrum of environments including edge devices and cloud systems. Table 2 shows YOLO12's performance benchmarks, highlighting its suitability for robust real-time vehicle detection.

Table 2. Performance comparison of YOLO12 model variants on COCO validation dataset.

Model	Size (pixels)	mAP_{50-95}	Speed T4 TensorRT (ms)	Params (M)	FLOPs (B)	Comparison (mAP/Speed)
YOLO12n	640	40.6	1.64	2.6	6.5	+2.1% / -9% (vs. YOLOv10n)
YOLO12s	640	48.0	2.61	9.3	21.4	+0.1% / +42% (vs. RT-DETRv2)
YOLO12m	640	52.5	4.86	20.2	67.5	+1.0% / -3% (vs. YOLO11m)
YOLO12l	640	53.7	6.77	26.4	88.9	+0.4% / -8% (vs. YOLO11l)
YOLO12x	640	55.2	11.79	59.1	199.0	+0.6% / -4% (vs. YOLO11x)

Proposed Approach. YOLO12 was paired with a fully-integrated preprocessing and augmentation methodology that improved overall detection performance across all of the numerous types of adverse weather conditions present within the DAWN dataset (as illustrated in Fig. 4). A number of steps were taken to preprocess each image from this dataset. Each image was resized into a dimension of 640×640. Corresponding to these new dimensions were the respective bounding boxes for vehicles that appeared within the original images. As a means of improving the models' ability to handle variability due to class imbalance, and thereby also increasing their robustness; several augmentations were implemented on the images. Examples include: geometric-based augmentations (i.e. flip, rotate, shear); color-based augmentations (i.e. hue/saturation/brightness variations; blur/noise additions); and an example of cutmix-based augmentations. Next, based upon parameters specifically chosen to be optimized for adverse weather conditions, YOLO12 was trained using the previously described data. Lastly, the resulting trained model produced real-time output in the form of bounding boxes and corresponding class probability values associated with detecting vehicles.

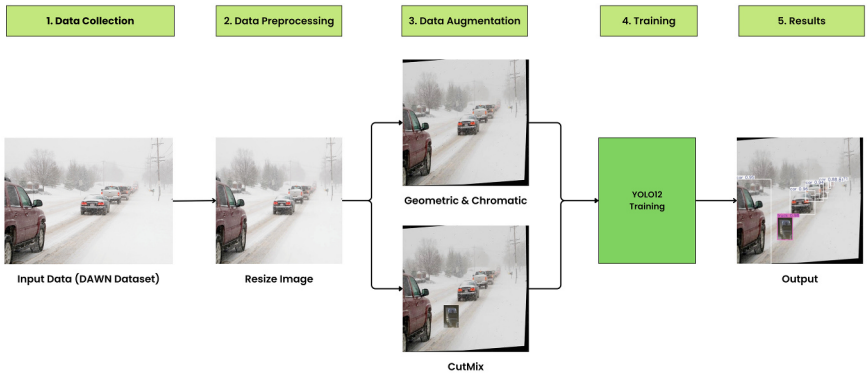

Fig. 4. Pipeline: data preprocessing, augmentations, CutMix, YOLO12 training, and vehicle detection results.

This pipeline combines advanced augmentations with YOLO12 to ensure robust and efficient vehicle detection in real-world traffic monitoring.

3.4 Implementation Details

Model training was performed by the Ultralytics YOLO architecture including the key configurations on the following parameters. Maximum training epochs were 200 epochs, 16 batch size, 640×640 pixels of input resolution. AMP was activated for faster training with less memory utilization. At that point, an initial learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005 were

employed in a stable manner during the learning process to achieve a consistent generalizable convergence. Checkpoints were retained every 10 epochs and early stopping applied if the validation performance failed to improve over 30 consecutive epochs. These configurations aimed at maximizing detection and classification performance under different weather conditions while optimizing training efficiency.

3.5 Evaluation Metrics

To measure the overall detection capabilities of the model, the authors employed four common quantitative measurement methods; Precision (P) Recall (R), Mean Average Precision (mAP50) and Mean Average Precision 50-95(mAP50-95). The models' Precision was measured as the ratio of True Positive (TP) results that are accurate detections in comparison to TP + False Positives (FP) or total number of correct predictions. Recall measured how well the model was able to correctly predict every object within the dataset as "TP / (TP+FN)" or the amount of total relevant objects detected. The mAP50 average precision metric calculates the Average Precision for each class in the dataset when using an Intersection Over Union (IoU) of 0.5 and then averages the sum of those values as mAP50. This allows for better understanding of how well a given model can detect objects at different levels. Additionally, the mAP50-95 average precision metric evaluates models on the same criteria but also takes into account the model's ability to localize objects through use of a range of IoU threshold settings from 0.5 to 0.95. Using this method allows researchers to obtain a broader view of how well a model performs in terms of both detection and localization.

4 Experiment Results

4.1 Model Training Results

Table 3 show the results for YOLO12 in 3 different trainings: (1) the original data set. (2) with the added use of both color and geometry augmentations. (3) with a combination of color, geometry, and CutMix as additional augmentations. Experiments were conducted using PyTorch and Ultralytics YOLO framework.

Table 3. Performance comparison of YOLO12 under different augmentation settings.

Dataset	Precision	Recall	mAP50	mAP50-95
Original	0.763	0.562	0.652	0.405
Geometric & Chromatic	0.94	0.88	0.953	0.787
Geometric & Chromatic + CutMix	0.968	0.956	0.979	0.901

YOLO12 trained using a combination of geometric and chromatic augmentations (with cutmix), was trained with the best mAP values (mAP50=0.979, mAP50-95 =0.901) of all 3 training methods.

4.2 Discussion

Geometric and chromatic augmentations increased baseline by .301 (from 0.652 to 0.953) while adding CutMix further improved detection rate of 0.979 (and an increase of .038 for map50-95: 0.901). As a result, this demonstrates that it can effectively address class imbalance issues. Baseline model fared poorly on minority classes such as bus and bicycle; however, using CutMix to create more representative images of those classes leads to better overall traffic detection rates. These results suggest that geometric and chromatic augments are important for improving generalization; in contrast, CutMix is essential for balanced detection in real-world traffic systems.

5 Conclusion

This paper has suggested an approach to build up a resilient vehicle classifier in all different types of weather condition, which is by combining cutmix based class balance into yolo12. The experimental results demonstrated that the use of these three different methods to augment the dataset significantly improved the average precision at 50 percent confidence (mAP50) of 65.2% for original data to 97.9% for augmented datasets, as well as mAP50-95 of 40.5% to 90.1%. These results have also shown improved detection accuracy and robustness for the model when deployed in extreme weather conditions. Additionally, since this was able to provide a much better distribution among classes, it provided a better per-class performance than the baseline model.

Overall, the proposed method has good promise for deployment on edge devices in real time in intelligent transportation system or in autonomous vehicles where there are needs for reliable detections in the presence of bad weather. In future research we would like to look into incorporating new approaches for domain adaptation as well as optimize the various versions of yolo12 for best possible edge device deployment.

References

1. Hassaballah, M., Kenk, M.A., Muhammad, K., Minaee, S.: Vehicle detection and tracking in adverse weather using a deep learning framework. IEEE Trans. Intell. Transp. Syst. **22**(7), 4230–4242 (2020)
2. Ultralytics. Ultralytics yolo12 (2024). https://docs.ultralytics.com/vi/models/yolo12/. Accessed: 04-July-2025
3. Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. **106**, 249–259 (2018)
4. Howard, A.G.: Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
5. Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., Yoo, Y.: Cutmix: regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6023–6032 (2019)

6. Rjoub, G., Wahab, O. A., Bentahar, J., Bataineh, A.S.: Improving autonomous vehicles safety in snow weather using federated YOLO CNN learning. In International Conference on Mobile Web and Intelligent Information Systems, pp. 121–134. Cham: Springer International Publishing (2021)
7. Wu, J. D., Chen, B. Y., Shyr, W. J., Shih, F.Y.: Vehicle Classification and Counting System Using YOLO Object Detection Technology. Traitement du Signal, **38**(4) (2021)
8. Boppana, U. M., Mustapha, A., Jacob, K., Deivanayagampillai, N.: Comparative analysis of single-stage yolo algorithms for vehicle detection under extreme weather conditions. In: IOT with Smart Systems: Proceedings of ICTIS 2021, Volume 2, pp. 637–645. Springer Singapore (2022)
9. Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., Zhang, L.: Image-adaptive YOLO for object detection in adverse weather conditions. In Proceedings of the AAAI Conference on Artificial Intelligence (vol. 36, No. 2, pp. 1792–1800) (2022)
10. Ding, Q., et al.: CF-YOLO: cross fusion yolo for object detection in adverse weather with a high-quality real snow dataset. IEEE Trans. Intell. Transp. Syst. **24**(10), 10749–10759 (2023)
11. Tian, E., Kim, J.: Improved vehicle detection using weather classification and faster R-CNN with Dark Channel Prior. Electronics **12**(14), 3022 (2023)
12. Panja, E., Hendry, H., Dewi, C.: YOLOv8 analysis for vehicle classification under various image conditions. Sci. J. Inform. **11**(1), 127–138 (2024)
13. Aloufi, N., Alnori, A., Basuhail, A.: Enhancing autonomous vehicle perception in adverse weather: a multi objectives model for integrated weather classification and object detection. Electronics **13**(15), 3063 (2024)
14. Ogino, Y., Shoji, Y., Toizumi, T., Ito, A.: ERUP-YOLO: Enhancing Object Detection Robustness for Adverse Weather Condition by Unified Image-Adaptive Processing. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 8597–8605. IEEE (2025)
15. Huang, Y.C., Lin, H.Y.: All-Weather Vehicle Detection and Classification with Adversarial and Semi-Supervised Learning. In: International Conference on Pattern Recognition, pp. 330–345. Springer, Cham (2025)
16. KENK, Mourad, DAWN, Mendeley Data, V3 (2020). https://doi.org/10.17632/766ygrbt8y.3