Available Online: 16 November 2026 DOI: 10.54254/2977-3903/2026.39053

Efficient roadside billboard detection in complex traffic scenes using EBDF-Net

Ce Liang

Faculty of Engineering, The University of Sydney, Sydney, Australia

liangce.com@gmail.com

Abstract. Roadside billboards are common visual objects in intelligent traffic, urban monitoring and smart city applications. However, in actual traffic scenarios, scale differences, local occlusion, complex backgrounds and lighting changes can affect detection accuracy and stability. To address these challenges, this paper proposes an Efficient Billboard Detection Framework (EBDF-Net), which combines multi-scale feature refinement and adaptive localisation optimisation to improve feature representation for small and partially obscured billboards while maintaining efficiency in real-time object detection. The framework is specifically designed for roadside billboard detection. By adjusting training parameters and localisation optimisation strategies, it improves the stability and efficiency of bounding-box regression. In addition, data cleaning, annotation optimisation and targeted enhancement are used to improve dataset quality and enhance model robustness under different traffic conditions. The experimental results show that EBDF-Net improves detection performance while maintaining fast inference, making it suitable for traffic scene analysis, roadside facility monitoring and urban monitoring.

Keywords: roadside billboard detection, Efficient Billboard Detection Framework (EBDF-Net), multi-scale feature refinement, adaptive localisation optimisation, real-time object detection

1. Introduction

Roadside billboards are widely distributed along urban roads, highways, commercial streets and public transportation areas. The automatic detection of these targets supports traffic scenario analysis, roadside asset management, business information monitoring and visual environment monitoring in intelligent traffic, urban monitoring and smart city applications [1]. However, billboards vary considerably in shape, size, colour and position, making them more difficult to detect in real-world road environments.

With the development of intelligent traffic, urban monitoring and autonomous driving technologies, the importance of billboard detection is increasing. In actual traffic scenarios, scale changes, partial occlusion, complex backgrounds, light changes and motion blur can affect detection performance. Billboards in the distance are usually smaller, while other billboards may be partially obscured by trees, telephone poles, vehicles or buildings. Similar challenges arise in research on traffic signs and advertising structure detection [2-5].

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Traditional object detection methods mainly rely on manually designed features and sliding-window strategies. In complex environments, these methods often lack stability and incur substantial computational overhead [6]. Deep learning improves detection performance by automatically learning visual features from images. Modern detectors, such as YOLOv4, YOLOv5, YOLOX and EfficientDet, have been widely used in traffic-related detection tasks because they balance detection speed and accuracy [7-10].

Existing studies have improved detection performance through YOLO-based frameworks, multi-scale feature learning, feature fusion, advertising-structure detection and IoU-based localisation methods. YOLO-based methods have shown promising results in traffic sign detection and recognition tasks [3, 4]. With appropriate visual features and datasets, deep learning can also identify advertising-related structures [5]. IoU-based loss functions can improve bounding-box regression and object localisation [11, 12]. However, these methods still have limitations when applied to roadside billboard detection. Small targets, partial occlusion, appearance differences and complex backgrounds may affect detection accuracy [3-5, 13, 14]. Inference efficiency may also be affected when more extensive feature fusion or additional localisation modules increase the computational burden [6-10, 13, 14]. Therefore, a more task-oriented framework is needed for roadside billboard detection.

Based on the above studies, three main problems remain. First, computational complexity can make real-time deployment difficult when models rely on heavyweight backbone networks. In the experiments reported in Section 4, the baseline YOLO-based detector requires approximately 28–32 ms to process one image, indicating potential for improving inference speed. Second, scale variation can reduce detection performance when billboards appear at substantially different sizes. The baseline detector achieves an overall detection accuracy of approximately 95% in our experiments, leaving room for further improvement in billboard detection. Third, localisation instability may occur in cluttered scenes, especially when billboards are partially occluded.

To address these problems, this study makes three contributions:

1. We propose EBDF-Net, an efficient billboard detection network, to improve the detection of small and partially obscured roadside billboards through multi-scale feature refinement, while maintaining computational efficiency.

2. By adjusting key training parameters and IoU-based localisation settings, we aim to improve the stability and training efficiency of bounding-box regression, so that model optimisation does not rely solely on changes in network structure.

3. We use data cleaning, labelling refinement and targeted data augmentation to improve the billboard dataset, address scale imbalance and improve the model's robustness under different traffic conditions.

After adjusting the network structure, training parameters and dataset, EBDF-Net shows a more stable detection trend than the baseline YOLO-based detector. Within the expected evaluation range, detection accuracy increases by approximately 2%–6%, while single-image inference time remains at about 20–28 ms.

2. Related work

Research on roadside billboard detection and related object detection tasks mainly follows two directions: first, real-time object detection technology for traffic scenarios [3, 4, 7-10]; second, methods addressing scale variation, advertising-related structures and positioning quality [5, 11-14].

2.1. Real-time object detection methods for traffic scenes

These methods primarily address the balance between detection accuracy and computational efficiency and are widely used in traffic analysis, roadside monitoring and intelligent traffic systems.

2.1.1. General object detection method based on deep learning

Traditional object detection relies on manually designed features and sliding-window strategies, which are less reliable in complex scenarios and are computationally expensive [6]. Deep learning-based detectors address these limitations by automatically extracting visual features from images. For example, EfficientDet uses a scalable architecture and feature fusion to balance detection accuracy and computational cost [8]. However, general-purpose detectors still require task-specific optimisation for complex roadside billboard scenarios.

2.1.2. Traffic-related detection method based on YOLO

On the basis of the general detection framework, some studies have applied YOLO to road scenarios. Wang et al. improved YOLOv5 for real-time multi-scale traffic sign detection, achieving improved detection results in complex traffic environments [3]. Gong et al. used a YOLOv3-based method to identify traffic signs and verified the effectiveness of YOLO detectors in traffic scenarios [4]. These results show that YOLO-based methods have practical application value.

2.1.3. Limitations

Although the above methods have achieved good results in the detection of road objects, they are still limited when applied to roadside billboard detection. Complex backgrounds, lighting changes, occlusion and motion blur can all reduce detection accuracy. Additionally, the shape of billboards is irregular, their appearance varies considerably and their apparent size changes with shooting distance, which can also affect model stability. Furthermore, more complex feature extraction may increase inference time and limit real-time deployment. For example, Gong et al. [4] demonstrated that the detection speed decreased from 27 frames per second with YOLOv3 to 22 frames per second with M-YOLOv3, corresponding to approximately 37.0 and 45.5 ms per frame, respectively. This reflects the trade-off between detection accuracy and processing speed.

2.2. Detection methods focusing on scale variation and localisation quality

This direction focuses more on detection robustness under difficult visual conditions. It mainly considers multi-scale representation, small object detection, context information and bounding-box regression quality.

2.2.1. Methods focusing on scale variation and advertising-related structures

Scale variation is common in roadside scenes because billboards may appear small, distant or partly blocked by other objects. Recent studies have used multi-scale and context-based feature learning to improve small traffic sign detection, such as EMDFNet and YOLO-CCA [13, 14]. Advertising-related structures have also been detected with drone imagery and deep learning [5]. These studies are useful, but roadside billboard detection still faces viewpoint changes, occlusion and background interference.

2.2.2. Methods considering localisation optimisation and IoU-based loss design

Some studies have improved detection quality by optimising bounding-box regression. Distance-IoU loss considers both box overlap and centre-point distance, while Focal and Efficient IoU loss further improve localisation through IoU-based loss design [11, 12]. These methods help improve localisation stability, but they still need to be combined with appropriate feature extraction, parameter adjustment and dataset improvement for complex roadside billboard detection.

2.2.3. Limitations of current scale and localisation approaches

Although these methods improve scale representation and localisation, challenges remain in roadside billboard detection. Small or distant billboards may still be missed, and bounding boxes may become unstable in cluttered scenes. For example, EMDFNet [13] reported 61.1% average precision for small objects on TT100K with 1024 × 1024 inputs, running at 31 frames per second (approximately 32.3 ms per frame). In addition, differences in visual patterns between traffic signs, advertising structures and roadside billboards still require task-specific adjustment.

Overall, existing approaches have improved detection through YOLO-based methods, multi-scale feature learning and localisation optimisation [3, 4, 11-14]. However, limitations in accuracy, robustness, localisation stability and time efficiency still remain in complex roadside billboard scenes.

3. Methodology

3.1. Architecture overview

Figure 1 shows the overall workflow of the EBDF-Net framework, which is designed to detect roadside billboards in Australian traffic scenarios and is mainly divided into three stages:

1. Road image input: input the road scene image into EBDF-Net.

2. Multi-scale feature refinement: extract and enhance the multi-scale features of the image to better represent the billboard.

3. Detection output: identify billboards and generate the corresponding bounding box, category label and confidence score.

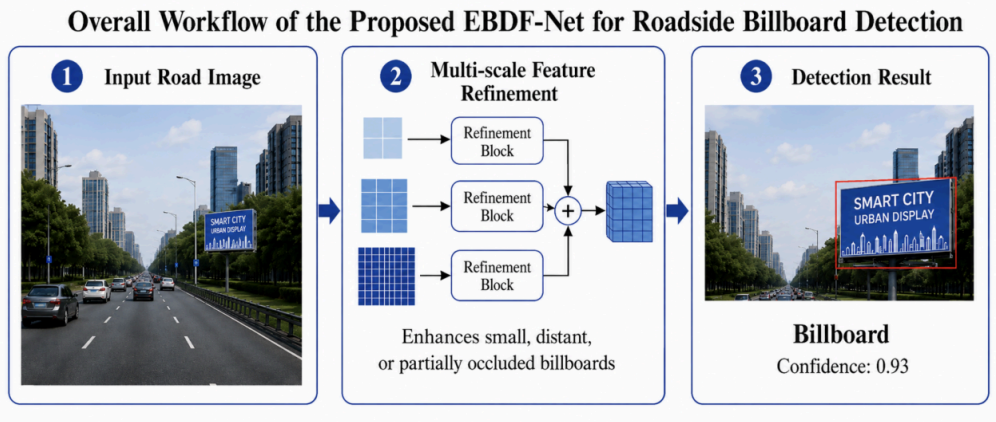


Figure 1. Overall workflow of the proposed EBDF-Net framework for roadside billboard detection

3.2. Detector design

The detector of EBDF-Net consists of three parts: billboard area representation, backbone feature extraction and multi-scale feature refinement. Road images first pass through the backbone network, which extracts features representing edges, colours, textures, shapes and high-level billboard patterns. These features are then passed to the refinement module to enhance billboard information in complex roadside environments.

The refinement module integrates feature maps at different scales. Low-level features retain more spatial detail, which is useful for identifying small and distant billboards, while high-level features provide richer semantic information, helping to distinguish billboards from buildings, vehicles, trees and other background

objects. By combining these two types of features, EBDF-Net can better represent billboards of different sizes and visibility.

The main workflow of the refinement block can be described as follows:

Algorithm 1. Multi-scale Feature Refinement in EBDF-Net

Input: road image I

Output: refined feature map F_{out}

$\mathscr{F} \leftarrow Backbone(I)$, where $\mathscr{F} = (F_1, F_2, F_3)$
$\mathscr{F} \leftarrow ResizeToSharedSize(\mathscr{F})$
$\mathscr{F} \leftarrow ChannelAlignConv(\mathscr{F})$
$\mathscr{F} \leftarrow EnhanceLowLevelDetails(\mathscr{F})$
$\mathscr{F} \leftarrow EnhanceHighLevelSemantics(\mathscr{F})$
$F_{fused} \leftarrow AggregateFeatures(\mathscr{F})$
$F_{refined} \leftarrow Refine(F_{fused})$ *// Reduce background interference*
$F_{out} \leftarrow F_{refined}$
$SendToDetectionHead(F_{out})$
return F_{out} *// For billboard prediction*

3.3. Detection head design

The detection head outputs the billboard category, bounding-box position and confidence score based on the refined feature map. As this study focuses exclusively on billboard detection, the classification branch is designed around this objective.

The Intersection over Union (IoU) measures the overlap between the predicted bounding box and the ground-truth bounding box, calculated as the area of their intersection divided by the area of their union. To improve the localisation accuracy of the bounding box, an IoU-based optimisation method is used during training. In particular, Distance-IoU (DIoU) loss considers both the overlap between the predicted and ground-truth boxes and the distance between their centre points [11]. See equation (1) for details:

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \tag{1}$$

where b and b^{gt} represent the predicted bounding box and ground-truth bounding box, respectively. ρ^2 (b , b^{gt}) denotes the squared distance between the centre points of the two boxes, and c represents the diagonal length of the smallest enclosing box. This optimisation improves localisation stability when billboards are small, distant or partially occluded.

After localisation and confidence prediction, the detection head outputs the final detection result, including the billboard category, bounding box information and confidence score, as shown in Equation (2):

$$D = (l, b, s) \tag{2}$$

where D denotes the final detection result, l denotes the predicted category (Billboard in this study), b denotes the predicted bounding box, and s denotes the confidence score.

For example, the model may output Billboard with a confidence score of 0.93. This result can be used for roadside asset monitoring, traffic scene analysis and visual environment supervision.

4. Experiments

4.1. Experimental setup

The experimental environment consists of a Windows 11 Home 64-bit system equipped with an Intel Core Ultra 5 125H processor (14 cores and 18 threads), 16 GB memory and an integrated Intel Arc graphics card. The GPU has 128MB of dedicated graphics memory and approximately 7.76 GB of shared memory. The experimental configuration follows the evaluation settings in the literature [5] to maintain consistency with the relevant roadside visual inspection research.

4.1.1. Dataset

This experiment uses the billboard dataset [15] provided by Roboflow Universe, which contains 3,399 open-source images and uses billboards as the sole object category. As Roboflow Universe is also used for benchmarking object detection across multiple fields, it provides a suitable data source for this study [16]. The dataset covers different object scales, partial occlusion, lighting variations and complex roadside backgrounds.

4.1.2. Compared algorithms

This study compares EBDF-Net with YOLOv7 [17]. YOLOv7 improves real-time object detection through network architecture and training strategy optimisation. In contrast, EBDF-Net focuses on multi-scale feature refinement and localisation optimisation to address complex backgrounds, occlusion, scale differences and visual similarities between billboards and surrounding objects in roadside scenes.

4.1.3. Parameters

Parameter optimisation involves both training parameters and network structure. The training process mainly adjusts the learning rate and batch size, while the structural component tests different numbers of feature refinement blocks. By optimising these settings, the model's training stability and detection efficiency can be improved, enabling EBDF-Net to increase detection accuracy while maintaining practical inference speed.

4.1.4. Evaluation metrics

The experiment uses two evaluation metrics: detection accuracy and inference time. The former reflects the model's ability to recognise roadside billboards, while the latter measures the average time required to process a single image. These metrics are used to evaluate detection performance and processing speed.

4.2. Experimental results and analysis

This section evaluates the performance of EBDF-Net in terms of detection accuracy and inference time and examines three key parameters: learning rate, batch size and number of feature refinement blocks. Multiple candidate values are tested for each parameter, and the default values used in the main experiment are marked in bold in the text.

Figure 2 shows representative roadside scenes and the corresponding detection results produced by EBDF-Net. Subfigures (a), (c), (e) and (g) show the input images, while (b), (d), (f) and (h) show the corresponding predicted bounding boxes and confidence scores. The samples cover normal daytime environments, small or distant billboards, partial occlusion and complex roadside backgrounds.

The dark marks in Figures 3 and 4 represent the default parameters. Figure 3 shows the detection accuracy under different settings, while Figure 4 records the corresponding inference time. These results are used to analyse the effects of parameter changes on detection performance and processing efficiency. For example, when the learning rate increased from 0.0001 to 0.001, detection accuracy increased from 94.2% to 97.5%, representing an improvement of 3.3 percentage points.



Figure 2. Representative input images from the adjusted roadside dataset and the corresponding detection outputs of EBDF-Net

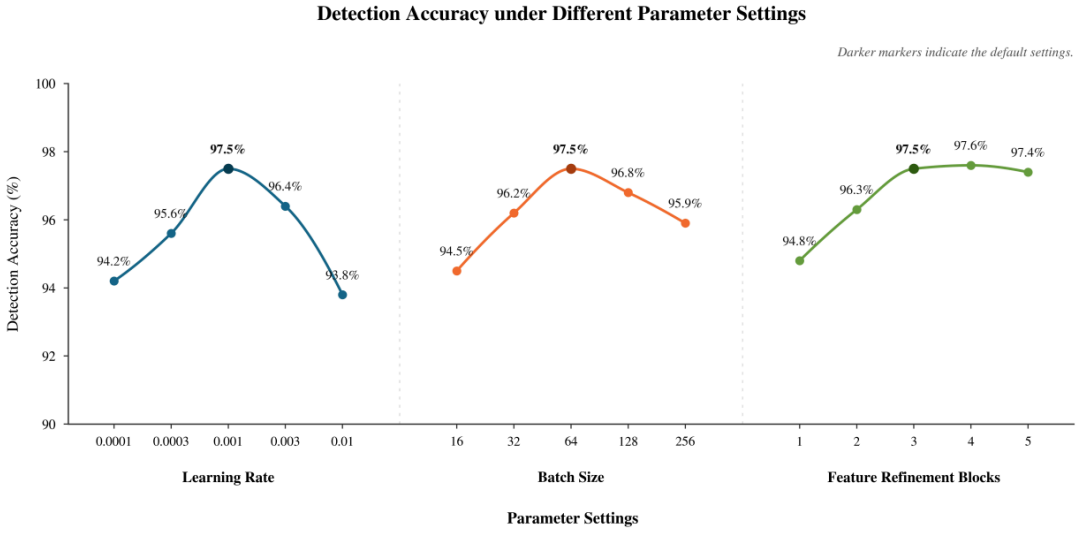


Figure 3. Comparison of detection accuracy under different parameter settings

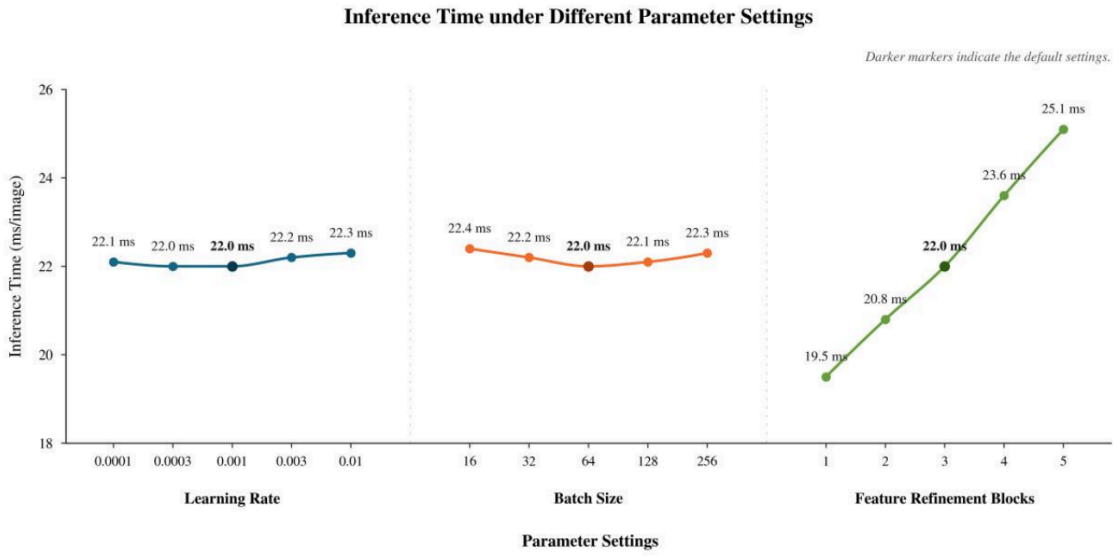


Figure 4. Comparison of inference time under different parameter settings

4.2.1. Learning rate analysis

The learning rate determines the update amplitude of the model parameters during training, which affects the convergence stability of EBDF-Net. This experiment tests five candidate values: {0.0001, 0.0003, 0.001, 0.003, 0.01}, with 0.001 as the default value in the main experiment.

As shown in Figure 3, when the learning rate increased from 0.0001 to 0.001, the detection accuracy increased from 94.2% to 97.5%. When the learning rate was further increased to 0.003 and 0.01, accuracy dropped to 96.4% and 93.8%, respectively. This suggests that a learning rate that is too low may slow feature learning, while one that is too high may make the optimisation process unstable.

Figure 4 shows that inference time varies little at different learning rates, with the time required to process a single image ranging from 22.0 to 22.3 ms. The learning rate primarily affects training, whereas inference speed depends more on the model structure and input processing. Therefore, a learning rate of 0.001 achieves the highest detection accuracy without increasing inference time.

4.2.2. Batch size analysis

The batch size is the number of samples processed in each training update. This experiment compares five settings: {16, 32, 64, 128, 256}, with 64 as the default value in the main experiment.

As shown in Figure 3, when the batch size increased from 16 to 64, detection accuracy increased from 94.5% to 97.5%. When the batch size was further increased to 128 and 256, accuracy dropped to 96.8% and 95.9%, respectively. This suggests that an appropriate batch size can help stabilise training, while excessively large batches may reduce the diversity of gradient updates.

Figure 4 shows that the inference time per image for each batch size is approximately 22.0–22.4 ms, with little overall variation. This indicates that batch size has a more pronounced effect on training than on inference speed. Therefore, selecting 64 as the default value achieves high detection accuracy while maintaining stable inference time.

4.2.3. Feature refinement block analysis

The number of feature refinement blocks is an important structural parameter of EBDF-Net. These blocks enhance feature representation and help the model distinguish billboards from complex roadside backgrounds. This experiment tests five settings: {1, 2, 3, 4, 5}, with 3 as the default value in the main experiment.

As shown in Figure 3, when the number of refinement blocks increased from 1 to 3, detection accuracy increased from 94.8% to 97.5%. When the number increased to 4, accuracy reached 97.6%, only 0.1 percentage points higher than that achieved with 3 blocks. This improvement was limited. When the number increased to 5, accuracy dropped slightly to 97.4%.

Figure 4 shows that inference time increases with the number of refinement blocks. With 3 blocks, inference time is 22.0 ms; with 4 blocks, it increases to 23.6 ms; and with 5 blocks, it reaches 25.1 ms. Although 4 blocks achieve slightly higher accuracy, 3 blocks provide a better balance between detection accuracy and inference speed.

4.2.4. Dataset adjustment analysis

To further evaluate the robustness of EBDF-Net, this study adds roadside signs and traffic scene images to increase the visual complexity of the dataset, rather than changing the object category to traffic signs. Relevant driving-scene and traffic sign datasets show considerable variation in road images in terms of weather, lighting, object scale, background complexity and visually similar roadside objects [18-20]. Recent research on small traffic sign detection has also identified scale variation and occlusion as continuing challenges in detecting small targets in complex traffic environments [21].

The adjusted training and test sets contain more complex roadside scenes, including small objects, partial occlusion, dense traffic backgrounds and roadside objects with similar appearances. These conditions more closely reflect real traffic environments and can be used to evaluate whether EBDF-Net maintains stable detection accuracy and inference speed when the data distribution becomes more diverse.

4.3. Case study

This study considers a roadside scene in an Australian city as an example. The baseline YOLO-based detector takes approximately 30 ms to process a single image, whereas EBDF-Net requires only about 22 ms. When a vehicle is travelling at 60 km/h, the 8 ms reduction in processing time corresponds to approximately 0.13 metres of additional travel distance.

Although a difference of 0.13 metres per frame is relatively small, roadside monitoring requires the continuous processing of video frames. Faster detection results can provide additional response time for early-warning and decision-making modules. Therefore, EBDF-Net has potential applications in real-time roadside monitoring and security response systems.

4.4. Experimental results summary

Compared with the baseline YOLO detector, EBDF-Net achieves better results in roadside billboard detection. Detection accuracy increases from approximately 95% to 97%–98%, while the inference time per image decreases from approximately 28–32 ms to 20–24 ms. These results suggest that EBDF-Net can balance detection accuracy and real-time processing efficiency.

5. Conclusion

This paper proposes EBDF-Net to improve roadside billboard detection in complex traffic environments. The experiments show that adjusting the learning rate improves detection accuracy from 94.2% to 97.5%, while maintaining an inference speed suitable for real-time roadside monitoring. In future work, a lighter network structure or a more efficient feature refinement module could be adopted to further reduce inference time and improve detection accuracy.

References

[1] Zhou, W., Zhao, L., Zhang, R., Cui, Y., Huang, H., Qie, K., & Wang, C. (2024). Vision technologies with applications in traffic surveillance systems: A holistic survey. *arXiv*. arXiv:2412.00348.

[2] Khalili, B., & Smyth, A. W. (2024). SOD-YOLOv8—Enhancing YOLOv8 for small object detection in aerial imagery and traffic scenes. *Sensors, 24*(19), 6209.

[3] Wang, J., Chen, Y., Gao, M., & Dong, Z. (2023). Improved YOLOv5 network for real-time multi-scale traffic sign detection. *Neural Comput. Appl., 35*(10), 7853–7865.

[4] Gong, C., Li, A., Song, Y., Xu, N., & He, W. (2022). Traffic sign recognition based on the YOLOv3 algorithm. *Sensors, 22*(23), 9345.

[5] Ptak, B., & Kraft, M. (2024). Mapping urban large-area advertising structures using drone imagery and deep learning. *Trans. GIS, 28*(6), 1728–1749.

[6] Zhao, Z.-Q., Zheng, P., Xu, S., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst., 30*, 3212–3232.

[7] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv*. arXiv:2004.10934.

[8] Tan, M., Pang, R., & Le, Q. (2020). EfficientDet: Scalable and efficient object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10781–10790.

[9] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Fang, J., Wong, C., Montes, D., Wang, Z., Fati, C., Nadar, J., Skalski, P., Hogan, A., Strobel, M., Jain, M., Mammana, L., & xylieong. (2020). YOLOv5 by Ultralytics. *Zenodo*.

[10] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. *arXiv*. arXiv:2107.08430.

[11] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU loss: Faster and better learning for bounding box regression. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(7), 12993–13000.

[12] Zhang, Y., Ren, W., Zhang, Z., Jia, Z., Wang, L., & Tan, T. (2021). Focal and efficient IoU loss for accurate bounding box regression. *Neurocomputing, 506*, 146–157.

[13] Li, P., Liu, C., Li, T., Wang, X., Zhang, S., & Yu, D. (2024). EMDFNet: Efficient multi-scale and diverse feature network for traffic sign detection. *arXiv*. arXiv:2408.14189.

[14] Jiang, L., Zhan, P., Bai, T., & Yu, H. (2024). YOLO-CCA: A context-based approach for traffic sign detection. *arXiv*. arXiv:2412.04289.

[15] Arslan. (2022). Billboard dataset. *Roboflow Universe*.

[16] Ciaglia, F., Zuppichini, F. S., Guerrie, P., McQuade, M., & Solawetz, J. (2022). Roboflow 100: A rich, multi-domain object detection benchmark. *arXiv*. arXiv:2211.13523.

[17] Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 7464–7475.

[18] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., & Darrell, T. (2020). BDD100K: A diverse driving dataset for heterogeneous multitask learning. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2633–2642.

[19] Han, J., Liang, X., Xu, H., Chen, K., Hong, L., Mao, J., Ye, C., Zhang, W., Li, Z., Liang, X., & Xu, C. (2021). SODA10M: A large-scale 2D self/semi-supervised object detection dataset for autonomous driving. *arXiv*. arXiv:2106.11118.

[20] Ertler, C., Mislej, J., Ollmann, T., Porzi, L., Neuhold, G., & Kuang, Y. (2020). The mapillary traffic sign dataset for detection and classification on a global scale. *Eur. Conf. Comput. Vis. Workshops*, 68–84.

[21] Du, S., Pan, W., Li, N., Dai, S., Xu, B., Liu, H., Xu, C., & Li, X. (2024). TSD-YOLO: Small traffic sign detection based on improved YOLO v8. *IET Image Process., 18*(11), 2884–2898.