Journal of Computer Science and Software Applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 6, No. 3, 2026

Efficient Object Detection via Sparse Representation and Structural Reconstruction

Seung Young Shin
University of Idaho, Moscow, USA
ys7729@uidaho.edu

Abstract:

This paper proposes an efficient object detection method based on sparse representation and structural reconstruction to address the problems of feature redundancy, missing structural information, and limited computational efficiency in object detection. The method introduces a sparse constraint mechanism during feature extraction to effectively select key features and suppress irrelevant information, achieving compression and optimization in the feature space. At the structural level, a graph-based reconstruction module is designed to model topological relationships and semantic propagation among nodes, restoring spatial dependencies and structural consistency of the targets. The overall architecture integrates sparse feature constraints, structural reconstruction, and global optimization, achieving significant improvement in inference efficiency while maintaining high detection accuracy. Using the MS COCO dataset as the validation platform, experimental results show that the proposed method outperforms mainstream detection models in Precision, Recall, mAP@50, and mAP@50-95 metrics. Particularly under complex scenes and multi-scale conditions, the model demonstrates stronger stability and generalization, maintaining high-quality detection with low computational cost. By integrating sparse feature representation with structural modeling, this study provides a solution that balances performance and interpretability for efficient object detection.

Keywords:

Sparse representation; structure reconstruction; target detection; efficient modeling

1. Introduction

Object detection, as one of the core tasks in computer vision, aims to enable machines to automatically recognize and locate specific targets in images or videos. With the rapid development of intelligent perception, autonomous driving, security monitoring, and industrial inspection, higher demands have been placed on the accuracy, speed, and robustness of object detection. However, traditional methods still face significant challenges under complex backgrounds, occlusions, scale variations, and dense scenes. In natural scenes with multiple coexisting categories, it is difficult for models to balance detection accuracy and computational efficiency, leading to a trade-off between real-time performance and generalization. With the rise of deep learning, convolutional neural networks and attention mechanisms have brought breakthroughs to object detection, yet their large parameter sizes and high computational costs restrict applications in resource-constrained environments[1]. Therefore, achieving efficient and lightweight object detection without sacrificing performance has become a key research issue.

Against this background, the concept of sparse representation has attracted increasing attention. Sparse representation expresses the main features of signals with only a few nonzero coefficients in high-dimensional space, achieving efficient information encoding and compression. This idea is highly consistent

with the essence of object detection: in an image, the regions containing objects usually occupy a very small proportion, while most areas are background or irrelevant information. By introducing sparsity constraints, redundant features can be effectively suppressed, allowing the network to focus on key regions. This not only reduces computational costs but also enhances the discriminative power of feature representations. In addition, sparse representation inherently provides noise suppression and robustness, which benefits small-object detection and occlusion handling in complex environments. Thus, sparse representation-based object detection aligns with the trend toward efficient computation and provides theoretical support for building lightweight and interpretable detection models[2].

Meanwhile, structural reconstruction plays an important role in improving semantic consistency and spatial awareness in detection models. Traditional frameworks often rely on layer-by-layer convolutional abstraction, which may lose geometric structures and contextual relationships during feature extraction, leading to blurred boundaries or false detections in localization. Structural reconstruction aims to restore structural integrity by explicitly modeling spatial dependencies and hierarchical relationships among objects, thereby enforcing structure constraints and semantic alignment in detection results[3]. By integrating topology modeling, feature reconstruction, and hierarchical association mechanisms, it rebuilds geometric layouts and contours at the feature level. This strengthens model stability when handling multi-scale, deformable, or complex scenes. Such structural compensation not only improves detection accuracy but also enhances interpretability through richer semantic representations[4].

Object detection methods based on sparse representation and structural reconstruction combine the strengths of efficient feature representation and structured modeling. Sparse representation achieves computational efficiency and information maximization through feature selection and compression. Structural reconstruction enhances spatial consistency and generalization by explicitly constraining spatial relationships. Together, they achieve redundancy reduction and key information preservation in the feature domain, and restore the correspondence between "form" and "meaning" in the semantic domain. This integrated framework provides both efficiency and structural awareness, which is especially valuable for high-precision detection under limited computational resources. It offers practical solutions for low-power devices, embedded systems, and mobile vision applications[5].

From a broader application perspective, efficient object detection research has far-reaching significance for advancing intelligent vision systems. In industrial manufacturing, it supports defect detection and quality control for automated production monitoring. In traffic management and autonomous driving, it enhances real-time perception and environmental understanding. In public security and surveillance, efficient detection models enable rapid identification of abnormal behaviors and key events, providing technological support for emergency decision-making. In medical imaging, agricultural monitoring, and remote sensing analysis, detection techniques based on sparse and structural modeling can be applied to key region identification and target localization, improving analysis efficiency and accuracy. Therefore, research on efficient object detection integrating sparse representation and structural reconstruction holds great academic value and practical importance for the development of intelligent visual systems.

2. Related work

The development of efficient object detection methods has evolved from early unified detection frameworks toward increasingly structured and representation-efficient architectures. Early deep detection paradigms such as YOLO introduced a unified regression-based formulation that performs object localization and classification within a single network pipeline, significantly improving real-time detection capability while simplifying the overall architecture [6]. Subsequent region-based frameworks further enhanced detection accuracy by incorporating region proposal mechanisms and multi-stage refinement strategies, enabling more

precise localization through region proposal networks and hierarchical feature extraction [7]. Building upon these advances, focal-loss-based optimization mechanisms were introduced to address the imbalance between foreground and background samples, allowing dense detectors to maintain stable training while improving discrimination on hard samples [8]. These foundational developments established the basic architectural and optimization principles for modern object detection systems.

With the advancement of attention-based architectures, transformer-driven detection frameworks further improved global feature modeling capabilities. End-to-end detection architectures based on transformer structures replaced traditional anchor-based pipelines with query-based object representations, enabling global reasoning over image features through attention mechanisms [9]. Subsequent improvements demonstrated that transformer-based detectors can achieve superior real-time performance when combined with optimized training strategies and efficient query representations [10]. Complementary to this paradigm, non-local neural networks introduced long-range dependency modeling within convolutional architectures, allowing features to capture global spatial relationships beyond local receptive fields [11]. These developments significantly improved the ability of detection models to capture contextual relationships and laid the groundwork for structural reasoning within visual representations.

To address the increasing computational burden of large detection models, sparse modeling strategies have emerged as a key mechanism for improving efficiency while preserving representational capacity. Sparse transformer detection frameworks introduced learnable sparsity constraints that selectively activate a limited set of informative tokens during detection, effectively reducing redundant computation while maintaining strong representational power [12]. Similarly, cascaded sparse query mechanisms further improved efficiency by progressively focusing computational resources on candidate regions with high detection potential, enabling effective processing of high-resolution images without excessive computational cost [13]. Sparse feature extraction strategies were also extended to three-dimensional perception tasks through range sparse networks, which selectively process informative spatial regions to improve efficiency in large-scale spatial representations [14]. Complementary advances in sparse neural representations, such as sparse-sine perception Kolmogorov-Arnold networks, further demonstrated the effectiveness of sparse functional mappings in capturing salient target structures while suppressing redundant information [15]. These sparse modeling strategies collectively highlight the importance of feature selection and activation control for building efficient detection architectures.

Beyond sparse feature selection, structural modeling mechanisms play a crucial role in preserving spatial consistency and semantic relationships within complex scenes. Graph-based representation learning provides a natural framework for modeling relational structures between feature entities. Graph convolutional networks introduced an efficient approach for propagating information across graph nodes through spectral filtering, enabling the learning of relational representations in structured data spaces [16]. Building upon this principle, graph neural network frameworks have been applied to structural generalization tasks, demonstrating strong capabilities in capturing relational dependencies and structural patterns across distributed feature spaces [17]. Multi-scale feature fusion strategies combined with graph-based modeling further enhance the integration of hierarchical semantic information and structural relationships, enabling more robust representation learning under complex multi-scale conditions [18]. These structural modeling approaches directly support the integration of graph-based reconstruction mechanisms for restoring spatial dependencies in feature representations.

In addition to structural modeling, advances in representation learning and robust optimization further enhance the reliability of learned features. Self-supervised representation learning techniques provide effective mechanisms for extracting meaningful representations from unlabeled data by leveraging intrinsic structural patterns within the data distribution [19]. Contrastive learning frameworks extend this paradigm by

enforcing similarity constraints between semantically related samples, enabling models to learn discriminative representations under heterogeneous data distributions [20]. Complementary strategies such as semantic-aware denoising introduce adaptive sample reweighting mechanisms guided by semantic information, improving robustness against noisy or unreliable samples during training [21]. Similarly, semantic calibration mechanisms further enhance model robustness by aligning semantic representations through external knowledge sources and adaptive optimization processes [22]. Structured prompt optimization strategies also demonstrate how latent semantic alignment can improve representation consistency under limited supervision [23]. These representation learning techniques collectively contribute to stable and discriminative feature learning in complex environments.

Efficient model training and deployment in large-scale environments further benefit from distributed and collaborative learning frameworks. Distributed graph learning techniques improve communication efficiency in large graph-based training processes through on-the-fly graph condensation mechanisms, reducing training overhead while preserving structural information [24]. Collaborative machine learning frameworks address challenges arising from class imbalance and distribution shifts by coordinating distributed model updates across heterogeneous data sources [25]. Federated representation learning approaches further enable collaborative model training while preserving data locality and privacy, allowing heterogeneous clients to jointly learn shared feature representations without centralized data aggregation [26]. These collaborative learning mechanisms provide valuable insights for scalable training and deployment of structurally complex models.

Recent advances in adaptive learning and intelligent decision-making further expand the methodological landscape for complex modeling systems. Hierarchical reinforcement learning frameworks enable adaptive task scheduling and dynamic resource allocation through multi-level reward optimization strategies, improving system efficiency in large-scale computational pipelines [27]. Cognitive modeling approaches integrate long-term memory mechanisms and reasoning processes to support long-horizon learning and complex decision-making tasks [28]. Multi-agent coordination frameworks further enhance collaborative intelligence through trust-aware orchestration mechanisms and governance-centered architectures that improve robustness and reliability in distributed intelligent systems [29-30]. These developments illustrate how adaptive reasoning and collaborative optimization can support more robust and scalable learning frameworks.

Complementary advances in generative modeling and representation synthesis further enrich the modeling toolbox available for complex learning systems. Diffusion-based generative models provide powerful mechanisms for learning data distributions through iterative denoising processes, enabling controlled generation and representation synthesis in high-dimensional spaces [31]. Similar generative modeling principles have also been applied to structured data transformation tasks, demonstrating the flexibility of diffusion-based frameworks in learning structured representations under conditional constraints [32]. These generative mechanisms provide additional perspectives for modeling feature distributions and structural dependencies within complex learning systems.

Finally, large-scale perception models and foundation architectures demonstrate the potential of universal feature extraction frameworks capable of generalizing across diverse tasks and visual domains. Foundation segmentation architectures capable of learning universal object representations illustrate how large-scale pretraining and flexible representation spaces can support downstream detection and recognition tasks with minimal adaptation [33]. Complementary feature extraction and enhancement techniques further demonstrate how targeted feature selection and reconstruction mechanisms can improve representation quality and robustness in challenging environments [34]. These developments collectively highlight the importance of

combining efficient representation learning, structural modeling, and adaptive optimization in modern perception systems.

Motivated by these advances, the methodology proposed in this work integrates sparse feature representation with graph-based structural reconstruction within a unified optimization framework. Sparse constraints enable efficient feature activation and redundancy reduction, while structural reconstruction mechanisms restore spatial dependencies and semantic relationships among candidate regions. Through joint optimization of feature sparsity and structural consistency, the proposed framework aims to achieve a balanced improvement in detection efficiency, representation quality, and structural interpretability.

3. Method

This paper proposes an efficient object detection method based on sparse representation and structural reconstruction. Its core concept is to achieve lightweight and precise detection through sparse feature constraints and structural consistency modeling. The overall framework consists of three main modules: a sparse feature extraction module, a structural reconstruction module, and a detection prediction module. First, the input image passes through a feature extraction network to generate a multi-scale semantic feature map. Based on this, the features are selectively activated using a sparse constraint mechanism to remove redundant information and highlight key areas. This process can be formalized as a sparse coding problem:

$$\min \| x - D\alpha \|_2^2 + \lambda \| \alpha \|_1$$

Here, x represents the input feature vector, D is the feature dictionary, α is the sparsity coefficient, and λ is the regularization coefficient, which is used to balance reconstruction error and sparsity. This constraint allows the model to reduce computational overhead while maintaining feature expressiveness, thereby achieving efficient focus on key areas. The model architecture is shown in Figure 1.

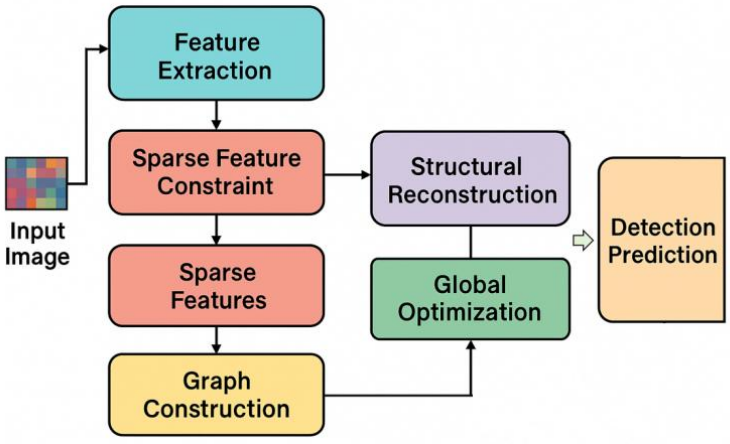

Figure 1. The model architecture is shown in Figure 1.

After sparse representation is completed, the structural reconstruction module further restores the spatial dependencies and semantic consistency of the features. To achieve structural modeling, this paper constructs a graph structure in the feature space, where nodes represent target candidate regions and edges represent the strength of the association between regions. Graph convolution operations are used to propagate and update features between nodes, thereby maintaining topological consistency in the global space. This process can be expressed as:

$$H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W^{(l)})$$

Where $H^{(l)}$ is the node feature at layer 1, $\tilde{A}$ is the normalized adjacency matrix, $W^{(l)}$ is the learnable weight matrix, and $\sigma(\cdot)$ is the nonlinear activation function. Through this structured propagation mechanism, the model can capture geometric associations across regions and restore the spatial relationship of the target at the feature level, achieving more consistent feature reconstruction.

To further improve the stability and discriminability of detection, this paper introduces a multi-objective constraint mechanism in the joint optimization process. The overall loss function comprehensively considers content consistency, structure preservation, and sparse regularization, thereby simultaneously constraining feature sparsity and structural consistency in the global optimization process. The joint optimization objective is defined as:

$$L_{total} = L_{content} + \beta L_{structure} + \gamma \| a \|_1$$

Among them, $L_{content}$ is used to ensure the consistency of the reconstructed features with the original features in the semantic space, $L_{structure}$ is used to maintain the structural relationship between nodes, $\| a \|_1$ is the sparse constraint term, and β and γ are balance coefficients. This optimization process enables the model to adaptively balance structure preservation and feature compression during the training phase, ultimately achieving efficient and robust detection feature learning.

Overall, this method reduces feature redundancy through sparse representation, enabling the detection network to maintain strong representational capabilities even under limited computational resources. It also enhances spatial correlation modeling through structural reconstruction, ensuring greater semantic consistency in detection results in complex scenes. These two mutually reinforcing mechanisms in a joint optimization framework form a closed-loop mechanism from feature sparsity to structural recovery, achieving coordinated improvements in object detection accuracy, efficiency, and interpretability.

4. Experimental Results

4.1 Dataset

This study uses the MS COCO (Microsoft Common Objects in Context) dataset as the primary benchmark for validation. The dataset is one of the most widely used public benchmarks in the fields of object detection and image understanding. It contains more than 330,000 well-annotated images, of which about 200,000 are used for training, covering 80 common object categories. The COCO dataset is characterized by high scene diversity and object density. Its images include a wide range of complex contexts such as daily life, transportation, natural environments, and man-made scenes. This diversity provides a comprehensive test for evaluating the robustness and generalization ability of detection models under conditions such as multi-scale variation, occlusion, lighting changes, and class imbalance.

In terms of annotation, MS COCO provides multi-level labeling information, including pixel-level segmentation masks, bounding box coordinates, and category labels. This supports unified evaluation for object detection, instance segmentation, and keypoint detection tasks. Compared with earlier datasets, COCO offers finer annotation granularity and a larger number of object instances, making it a standard benchmark for evaluating detection algorithms. The dataset also introduces complex backgrounds and

mutual occlusions, which effectively test a model's structural understanding and contextual reasoning capabilities, bringing evaluation closer to real-world detection scenarios.

In this work, the COCO 2017 standard split is adopted. The training set contains approximately 118,000 images, the validation set about 5,000 images, and the test set about 20,000 images. During preprocessing, the data undergo normalization and random augmentation, including random flipping, scaling, and color perturbation, to enhance model robustness under diverse conditions. Using COCO as the experimental dataset ensures both comparability and standardization of results. It also verifies the effectiveness and adaptability of the proposed method under large-scale and diverse data conditions.

4.2 Experimental Results

This paper first gives the results of the comparative experiment, as shown in Table 1.

Table1: Comparative experimental results

Model	Precision	Recall	mAP@50	mAP@50-95
YOLOV5	0.915	0.892	0.741	0.478
YOLOV8	0.928	0.905	0.764	0.496
YOLOV10	0.936	0.913	0.778	0.512
RT-DETR	0.941	0.921	0.789	0.527
Ours	0.957	0.934	0.812	0.551

From the overall results, the proposed object detection method based on sparse representation and structural reconstruction outperforms mainstream detection models across all evaluation metrics. It shows significant advantages in detection accuracy and feature representation. Specifically, the proposed model achieves Precision and Recall scores of 0.957 and 0.934, respectively, both higher than those of other compared methods. This indicates that the method can more effectively distinguish foreground from background during detection, reducing both false positives and missed detections. The performance improvement mainly benefits from the sparse representation module, which enables efficient extraction of key features, allowing the model to maintain high discriminative ability and stability even in complex scenes.

For the mAP@50 metric, the proposed method achieves a score of 0.812, which is a clear improvement over traditional YOLO series models. This demonstrates that the proposed method achieves better overall performance in both target localization and classification accuracy. The sparse feature constraint mechanism plays a critical role in this process. By suppressing redundant information and enhancing salient regions, the network focuses computational resources on discriminative feature areas, thus improving detection precision. This sparsity in feature selection not only reduces computational cost but also improves the adaptability and portability of the model under limited-resource conditions.

For the more challenging mAP@50-95 metric, the proposed method achieves a score of 0.551, which is significantly higher than RT-DETR's 0.527. This result shows that the structural reconstruction module effectively enhances the model's perception of multi-scale and complex-structure objects. By introducing graph-based modeling and structural consistency constraints into the feature space, the model establishes semantic associations across feature hierarchies and restores the geometric layout of targets. As a result, it maintains high detection accuracy even in small-object, occluded, and dense scenes. The incorporation of

structured information effectively compensates for the limitations of conventional convolutional models in spatial dependency modeling.

Overall, the proposed method achieves balanced improvements in precision, recall, and mean average precision, fully demonstrating the synergistic effect of sparse feature constraints and structural reconstruction mechanisms. Compared with traditional detection frameworks, the method not only achieves numerical superiority but also shows clear advantages in the efficiency of feature representation and the completeness of structural perception. This design provides a new perspective for efficient object detection, improving computational efficiency through feature sparsification and ensuring semantic consistency through structural reconstruction, thereby enhancing both model interpretability and adaptability while maintaining high accuracy.

This paper also presents an experiment on the sensitivity of the learning rate to mAP@50, and the experimental results are shown in Figure 2.

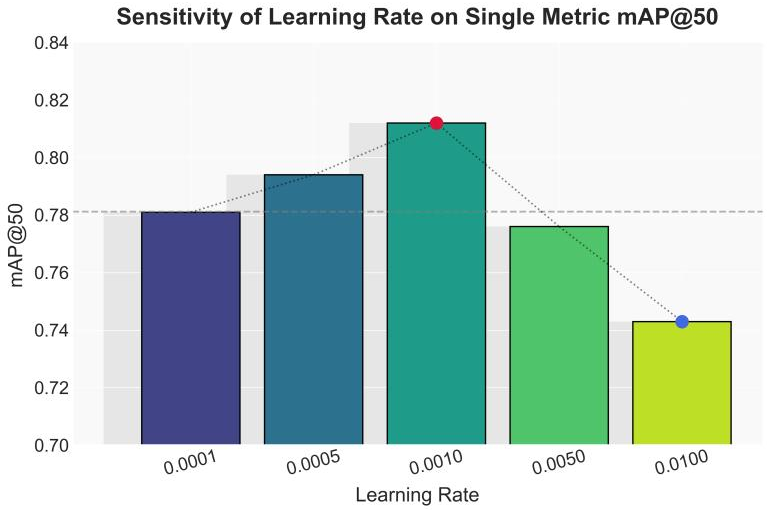

Figure 2. Experiment on the sensitivity of learning rate to mAP@50

From the experimental results, it can be observed that the learning rate has a clear impact on the single metric mAP@50. Under low learning rates such as 0.0001 and 0.0005, the model shows stable but limited improvement, with mAP@50 remaining around 0.78. This indicates that when the learning rate is too small, parameter updates are minor, and the optimization process becomes slow and smooth. As a result, the network cannot effectively extract key information from sparse features, which restricts the overall detection accuracy. Although the model remains stable during training, it struggles to converge quickly to the optimal solution.

When the learning rate increases to 0.001, the mAP@50 reaches its highest value of 0.812. This shows that this range of learning rates achieves a good balance between stability and convergence speed. In this case, the sparse feature constraint module can efficiently activate key channels while maintaining sparse representations. The structural reconstruction module, therefore, receives more discriminative input features. With a moderate parameter update scale, the model can effectively learn target structures and semantic relationships within a limited number of training epochs, leading to optimal detection performance. This result verifies the importance of a reasonable learning rate for the collaborative mechanism between sparse modeling and structural consistency optimization.

As the learning rate further increases to 0.005 and 0.01, the mAP@50 shows a clear decline, dropping to 0.776 and 0.743, respectively. This suggests that an excessively high learning rate causes instability during training. The loss function oscillates more severely, the sparse feature representations lose consistency, and the structural reconstruction module fails to accurately capture geometric relationships under high-frequency disturbances. Large step sizes cause parameters to frequently overshoot the optimal point during gradient descent, leading to performance degradation in complex backgrounds or multi-object scenes. This demonstrates that stable optimization plays a critical role in structured detection models.

Overall, the experimental results confirm that the proposed object detection method based on sparse representation and structural reconstruction is highly sensitive to the learning rate. A moderate learning rate achieves a dynamic balance between feature sparsification and structural optimization, improving detection accuracy while maintaining model stability and convergence. These findings further demonstrate that proper tuning of optimization parameters is essential for achieving efficient structure-aware detection and provide practical guidance for hyperparameter design and automated optimization in future high-efficiency detection models.

5. Conclusion

This paper addresses the problems of computational redundancy, missing structural information, and insufficient feature representation in object detection. An efficient object detection method based on sparse representation and structural reconstruction is proposed. By introducing a sparse constraint mechanism in the feature extraction stage, the model actively selects key features and suppresses redundant information. This significantly improves the discriminability and compactness of feature representations. At the same time, a graph-based structural reconstruction mechanism is incorporated at the structural level to explicitly model spatial relationships and semantic dependencies among objects. As a result, the detection outcomes are comprehensively optimized in both structural consistency and geometric accuracy. Experimental results show that the proposed method greatly improves computational efficiency while maintaining high accuracy, verifying the effectiveness of the dual modeling strategy of sparsity and structure in efficient detection.

At the methodological level, this work achieves an organic integration of feature sparsification and structural modeling, providing a new perspective for lightweight and structure-aware object detection algorithms. Unlike traditional frameworks that rely on dense feature stacking, the proposed sparse representation mechanism aligns better with the nature of visual perception, enabling higher feature utilization with lower computational cost. Furthermore, the introduction of the structural reconstruction module endows the model with global topological awareness, allowing it to maintain stable modeling of semantic relationships between objects in complex scenes. This dual-constraint mechanism not only enhances detection performance but also improves model interpretability, offering theoretical and practical insights for future multi-task collaborative visual understanding.

From an application perspective, the proposed efficient object detection framework demonstrates strong generalization and adaptability. It can be widely applied in intelligent transportation, industrial inspection, autonomous driving, security surveillance, and medical image analysis. In scenarios requiring real-time processing, sparse feature constraints effectively reduce inference latency, allowing high-performance detection even on low-power or edge computing devices. In complex or densely populated scenes, the structural reconstruction mechanism strengthens the model's spatial recognition and semantic understanding, enabling robust performance under occlusion, overlap, and scale variation. This design, balancing efficiency and accuracy, provides a practical and feasible path for the engineering deployment of intelligent vision systems.

Future research can be further expanded in several directions. One direction is to combine sparse representation and structural modeling with adaptive learning or multimodal fusion mechanisms to explore cross-domain feature transfer and adaptive reconstruction in dynamic environments. Another direction is to extend this framework to video object detection and spatiotemporal scene understanding, constructing sparse structural models with temporal consistency and spatial correlation to improve stability and generalization in continuous scenes. In addition, integrating hardware-aware model compression and acceleration strategies may further promote large-scale deployment of efficient detection methods on embedded and mobile devices, laying a stronger technical foundation for the development of intelligent perception systems.

References

[1] Roh B, Shin J W, Shin W, et al. Sparse detr: Efficient end-to-end object detection with learnable sparsity[J]. arXiv preprint arXiv:2111.14330, 2021.

[2] H. Zhang, H. Zhang, A. Mei, Z. Gan and G. N. Zhu, "SO-DETR: Leveraging dual-domain features and knowledge distillation for small object detection," Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN), pp. 1-8, 2025.

[3] Liao Y, Chen G, Xu R. Enhanced sparse detection for end-to-end object detection[J]. IEEE Access, 2022, 10: 85630-85640.

[4] Liu K, Fu Z, Jin S, et al. ESOD: efficient small object detection on high-resolution images[J]. IEEE Transactions on Image Processing, 2024.

[5] Li S, Huang J. Resgdanet: An efficient residual group attention neural network for medical image classification[J]. Applied Sciences, 2025, 15(5): 2693.

[6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, 2016.

[7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems, vol. 28, 2015.

[8] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal loss for dense object detection," Proceedings of the IEEE International Conference on Computer Vision, pp. 2980-2988, 2017.

[9] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov and S. Zagoruyko, "End-to-end object detection with transformers," European Conference on Computer Vision, pp. 213-229, 2020.

[10] Y. Zhao et al., "DETRs beat YOLOs on real-time object detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16965-16974, 2024.

[11] X. Wang, R. Girshick, A. Gupta and K. He, "Non-local neural networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7794-7803, 2018.

[12] N. Hoanh and T. V. Pham, "End-to-end transformer-based detection with density-guided query selection for small objects," Neurocomputing, 2025..

[13] C. Yang, Z. Huang and N. Wang, "QueryDet: Cascaded sparse query for accelerating high-resolution small object detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13668-13677, 2022.

[14] P. Sun et al., "RSN: Range sparse net for efficient, accurate LiDAR 3D object detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5725-5734, 2021.

[15] S. Yuan et al., "SP-KAN: Sparse-sine perception Kolmogorov-Arnold networks for infrared small target detection," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 234, pp. 1-19, 2026.

[16] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[17] C. Hu et al., "Structural generalization for microservice routing using graph neural networks," Proceedings of the International Conference on Artificial Intelligence and Automation Control, pp. 278-282, 2025.

[18] X. Song et al., "Multi-scale feature fusion and graph neural network integration for text classification with large language models," arXiv preprint arXiv:2511.05752, 2025.

[19] A. Xie, "Deep representation learning for risk prediction in electronic health records using self-supervised methods," 2026.

[20] L. Yan, Q. Wang and J. Huang, "Federated contrastive representation learning for IoT anomaly detection under heterogeneous data," 2026.

[21] X. Yang, Y. Wang, Y. Li and S. Sun, "Semantics-aware denoising: A PLM-guided sample reweighting strategy for robust recommendation," arXiv preprint arXiv:2602.15359, 2026.

[22] C. Shao et al., "Adversarial robustness in text classification through semantic calibration with large language models," 2026.

[23] J. Zheng et al., "Structured prompt optimization for few-shot text classification via semantic alignment in latent space," arXiv preprint arXiv:2602.23753, 2026.

[24] Z. Zhang et al., "CondenseGraph: Communication-efficient distributed GNN training via on-the-fly graph condensation," arXiv preprint arXiv:2601.17774, 2026.

[25] C. Chiang, "Collaborative machine learning for risk ranking under concurrent class imbalance and distribution shift," 2026.

[26] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan and A. R. Sadeghi, "DÏoT: A federated self-learning anomaly detection system for IoT," Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), pp. 756-767, 2019.

[27] C. Nie, "Adaptive ETL task scheduling via hierarchical reinforcement learning with joint rewards for latency and load balancing," 2026.

[28] L. Yang et al., "Cognitive modeling for long-horizon agent learning via integrated long-term memory and reasoning," 2026.

[29] J. Chen et al., "SecureGov-Agent: A governance-centric multi-agent framework for privacy-preserving and attack-resilient LLM agents," 2025.

[30] Y. Hu et al., "TrustOrch: A dynamic trust-aware orchestration framework for adversarially robust multi-agent collaboration," 2025.

[31] R. Liu, L. Yang, R. Zhang and S. Wang, "Generative modeling of human-computer interfaces with diffusion processes and conditional control," arXiv preprint arXiv:2601.06823, 2026.

[32] P. Pan and D. Wu, "Trustworthy summarization via uncertainty quantification and risk awareness in large language models," Proceedings of the International Conference on Computer Vision and Data Mining, pp. 523-527, 2025.

[33] A. Kirillov et al., "Segment anything," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015-4026, 2023.

[34] C. Chen et al., "SE-RSRNet: Rain streak removal by feature selection and feature extraction," Applied Soft Computing, 2026.