Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)

VOLUME:9, ISSUE:1, JANUARY 2026, PAPER NUMBER:3

Edge-HiResFusion: A Lightweight End-to-End Super-Resolution and Detection Approach with Adaptive NMS for Small Object Localization and object classification

Anu Yadav[1] , Prof. Ela Kumar[2]
[1]Research Scholar,and [2]Professor,
[1,2]Department of Computer Science & Engineering, Indira Gandhi Delhi Technical University for Women, New Delhi, Delhi 110006, India.
[1]anuyadavcse@gmail.com, [2]ela_kumar@igdtuw.ac.in

Abstract— **Detecting small objects in low-resolution images remains a major hurdle in many real-world settings such as drone surveillance, public safety monitoring, and embedded vision systems. Standard object detection models tend to struggle with such inputs, especially when images are degraded or compressed. Although image super-resolution (SR) techniques can improve visual quality, most existing systems treat SR and detection as separate tasks, which limits their overall effectiveness. In this work, we introduce Edge-HiResFusion, a lightweight, end-to-end framework that unifies SR and object detection into a single, optimized pipeline. Unlike traditional cascaded approaches, our method learns to enhance image regions that are specifically relevant for detection. The model integrates a task-aware SR module, an efficient detection backbone, and an adaptive non-maximum suppression (NMS) algorithm designed to better handle overlapping predictions in cluttered or noisy images. We evaluate our model on the MS COCO dataset under simulated low-resolution conditions. Compared to existing pipelines, Edge-HiResFusion delivers noticeable improvements in both detection accuracy (0.52 mAP) and small-object recall, while maintaining low computational overhead — making it a strong candidate for deployment on edge devices.**

Keywords— super resolution, NMS, object detection, MS COCO, object localization, object classification, Deep learning.

1. INTRODUCTION

In recent years, AI (artificial intelligence) and DL (deep learning) have significantly transformed the field of computer vision. Among its many applications, object detection stands out as one of the most impactful tasks, enabling systems to identify and localize objects within images or video streams. From autonomous vehicles and aerial drones to smart surveillance systems and industrial inspection, object detection serves as a foundational technology in a wide range of real-world environments. Despite the impressive progress made by modern detectors such as YOLOv5[1], Faster R-CNN[2], and SSD[3], these models often assume access to high-resolution, high-quality imagery. In practical scenarios, however, this assumption does not hold. Surveillance cameras mounted at high altitudes, drones capturing distant objects, or resource-constrained devices with limited imaging capabilities often produce low-resolution, degraded images. Under such conditions, the performance of standard detection models deteriorates sharply especially when detecting small, overlapping, or partially occluded objects.

One common solution is to apply image super-resolution (SR) techniques [4] prior to detection. These methods aim to enhance image clarity by reconstructing high-resolution outputs from low-quality inputs. Recent advancements like ESRGAN [5] and Real-ESRGAN [6] have achieved visually impressive results by generating realistic textures and details. However, these methods are typically trained to maximize perceptual quality, not detection performance. As a result, SR-enhanced images may look better to the human eye, but they don't always lead to better object detection accuracy. Moreover, most existing SR+ Detection pipelines rely on a two-stage approach: first applying SR using a pre-trained model and then running a separate detection module. This decoupled strategy fails to optimize both components together. Since the SR model is unaware of the detection task, it may enhance irrelevant areas while ignoring features that are crucial for localization. Similarly, standard non-maximum suppression (NMS) used in these pipelines often relies on fixed thresholds, which may not adapt well to overlapping objects in noisy or crowded scenes.

These challenges reveal a clear research gap: there is a lack of lightweight, task-aware, and end-to-end frameworks that can jointly enhance image resolution and detect objects—especially small ones—in low-quality imagery. Such a model would need to be optimized not for image realism, but for detection-specific performance, while remaining efficient enough for real-world deployment on edge devices.

This paper introduces Edge-HiResFusion, a unified deep learning framework that combines image super-resolution and object detection into a single, end-to-end trainable model. Unlike traditional two-stage pipelines, where SR and detection are treated as separate tasks, our approach jointly learns to enhance and detect, with the goal of improving accuracy in low-quality visual environments. The key contributions of our work are summarized as follows:

A Task-Aware Super-Resolution Module: We propose a lightweight SR network that is optimized not just for visual clarity, but specifically for object detection. By learning to enhance image regions that are critical for recognizing and localizing objects, the model improves the visibility of small or distant targets without unnecessary texture hallucination.

An End-to-End Architecture for Joint Learning: Edge-HiResFusion integrates the SR module with a detection backbone into a unified framework. Both components are trained together using a combined loss function, ensuring that the super-resolved images are directly beneficial for the downstream detection task.

Incorporation of Adaptive Non-Maximum Suppression (NMS): We implement a class-aware, confidence-weighted NMS mechanism that adapts suppression thresholds based on context. This approach is especially effective in handling overlapping objects in crowded or degraded scenes, where standard NMS may remove correct detections.

Edge-Efficient Design for Real-Time Use: The model is designed with computational efficiency in mind. It maintains a low inference footprint, making it suitable for deployment on edge devices such as UAVs, surveillance cameras, and embedded systems with limited processing capabilities.

Comprehensive Evaluation on Low-Resolution COCO Dataset: We evaluate our method on a degraded variant of the MS COCO dataset [7], simulating real-world conditions. Results show noticeable improvements in mean average precision (mAP), particularly for small objects, compared to both baseline detectors and SR + Detection pipelines.

This work's remaining sections are arranged as follows: The following section 2 offers a Related Work of Super-Resolution models (SRCNN[4], ESRGAN, Real-ESRGAN) Object Detection (YOLOv5, FPN, SSD, Faster R-CNN, Mask RCNN [8]), Lightweight & embedded models, Adaptive and learned NMS techniques and Existing SR Detection hybrid approaches. Section 3 present the proposed methodology of Edge-HiResFusion for enhancing object detection accuracy. In Section 4, provide and discuss the experimental findings from proposed research. Section 5 closes by reviewing the key conclusions and how they affect the field.

2. Related Work

Super-resolution (SR) techniques have gained substantial attention in recent years for their ability to restore high-resolution details from low-quality inputs. Early CNN-based methods like SRCNN[4] demonstrated the feasibility of learning end-to-end mappings from low- to high-resolution images using deep networks. Later, more advanced architectures such as EDSR [9]and ESRGAN[5] introduced residual learning and perceptual loss functions to improve both fidelity and visual quality. Real-ESRGAN[6] is designed to handle complex degradations and real-world noise, making it highly suitable for practical low-resolution scenarios. While these methods have shown impressive results in restoring image clarity, they are often optimized for human perception, rather than task-specific goals like object detection.

Object detection has evolved significantly with the emergence of deep convolutional architectures. Faster R-CNN [2] and Mask R-CNN+SVM [10] introduced region proposal networks for two-stage detection with high accuracy. On the other hand, YOLOv5 and SSD[3] represent single-stage detectors optimized for real-time performance. Feature fusion mechanisms such as Feature Pyramid Networks (FPN) [11] and PANet [12] have also improved multi-scale detection, particularly for small objects. Despite their robustness on high-resolution datasets, these models tend to underperform when applied to degraded or low-resolution inputs — a common occurrence in embedded and real-world systems.

As the demand for mobile and embedded vision systems grows, there has been a shift toward creating lightweight models that can operate efficiently on constrained hardware. Approaches like MobileNet [13], Tiny-YOLO [14], Inception-ResNetV2 [15] and ShuffleNet[16] focus on reducing parameter count and inference latency while maintaining accuracy. However, many super-resolution and detection models remain computationally intensive and are not suitable for real-time deployment on edge devices. This presents a need for joint optimization strategies that can maintain detection performance while respecting device limitations.

Non-Maximum Suppression (NMS) is a critical post-processing step in object detection that helps eliminate redundant bounding boxes. Traditional NMS applies a fixed IoU threshold [17], which often fails in crowded scenes or with small, overlapping objects. Soft-NMS [18] and Learning NMS techniques have been proposed to address these limitations by modifying the suppression logic based on box confidence scores or learned representations. Despite these improvements, adaptive NMS methods [19] are rarely integrated into end-to-end detection pipelines, especially in the context of degraded imagery.

Recent works have explored the use of SR as a pre-processing step to enhance object detection, particularly in low-resolution scenarios. However, these pipelines are often modular and disconnected, where the SR model is trained independently from the detection module. This separation limits task-specific enhancement and can result in sub-optimal performance. Some joint models have been proposed, but they either use large SR architectures not suitable for deployment, or they lack adaptive mechanisms like dynamic NMS. Furthermore, most prior studies focus on synthetic downscaling and do not generalize well to real-world degradations such as noise, blur, or compression.

Given the gaps in the current literature, there is a strong motivation to develop a framework that is both lightweight and task-aware, capable of enhancing low-quality visual input specifically for detection tasks. Edge-HiResFusion addresses this by integrating super-resolution, deep detection, and adaptive NMS into a single, end-to-end pipeline. Unlike traditional methods, our approach is trained holistically, making it more effective for real-world applications involving low-resolution, cluttered, or noisy scenes — while remaining efficient enough for edge deployment.

3. Proposed Method: Edge-HiResFusion

3.1 Overview

The goal of the Edge-HiResFusion framework is to improve object detection in degraded or low-resolution images using a lightweight, fully trainable deep learning architecture. The entire system is designed as a single end-to-end model that performs three key tasks: image enhancement through super-resolution, feature extraction and object detection, and refined prediction through adaptive Non-Maximum Suppression (NMS).

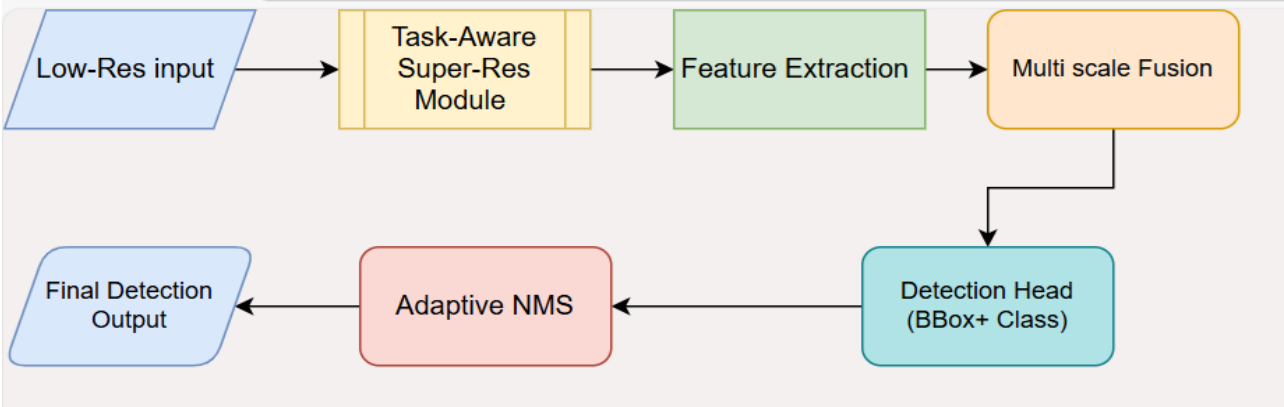

Figure 1: Edge-HiResFusion Full Architecture Diagram

As illustrated in the model flow diagram in figure 1, the low-resolution input image first passes through a task-aware super-resolution module[11] that selectively enhances relevant image regions. The super-resolved image is then fed into a backbone network that extracts spatial and semantic features. These features are fused across multiple scales to enable robust detection of objects at varying sizes, especially small ones. Finally, an adaptive NMS unit filters redundant predictions based on contextual awareness, helping to refine the detection output. Each component of the system contributes to a more accurate and efficient detection pipeline, particularly suited for real-world and resource-limited environments. Each module in the system is optimized to either enhance, extract, or refine information that contributes directly to better detection outcomes. Unlike traditional pipelines, all components are trained together, allowing the network to learn a joint representation that aligns super-resolution output with detection relevance.

3.2 Task-Aware Super-Resolution Module

The first stage of the pipeline involves enhancing the input image using a lightweight super-resolution module. Unlike traditional SR methods that prioritize visual quality (e.g., sharpness or texture realism), this module is trained to enhance only the areas that matter for object detection — such as edges, contours, and texture gradients that help in localizing small or distant objects. We define the SR loss in equation 1 as:

$$L_{SR} = \| I_{SR} - I_{HR} \|_1 + \alpha \times \| \varphi(I_{SR}) - \varphi(I_{HR}) \|^2 \quad (1)$$

Where:

- I_{SR} = Super-resolved image
- I_{HR} = Ground-truth high-resolution image
- φ = Perceptual feature extractor (e.g., VGG)
- α = Weight for perceptual loss.

The first term is an L1 loss that ensures pixel-level accuracy. The second term is the perceptual loss, encouraging SR to preserve high-level features important for detection.

3.3 Feature Extraction and Detection Head

Once the image is super-resolved, it is passed through a feature extractor — either a lightweight convolutional network MobileNet. These networks extract semantic and spatial features necessary for object localization and classification. To enable robust detection of objects at different scales, we integrate a Feature Pyramid Network (FPN). These modules combine low-level detail-rich features with high-level semantic context across layers, which is crucial for detecting small or partially visible objects. The detection head then performs two main tasks:

- Predicts bounding box coordinates (object location)
- Assigns class probabilities (object category)

The detection loss is defined in equation 2 as:

$$£_{Det} = £_{cls} + £_{bbox} \quad (2)$$

Where, $£_{cls}$ = Classification loss (e.g., Cross-Entropy or focal loss), $£_{bbox}$= Bounding box regression loss (e.g., CIoU for better geometric accuracy).

3.4 Adaptive Non-Maximum Suppression

In dense scenes or when objects are close together, traditional Non-Maximum Suppression (NMS) may suppress valid detections due to static IoU thresholds. To mitigate this, Edge-HiResFusion introduces a confidence-aware adaptive NMS technique that adjusts the suppression criteria based on context. This formulation ensures that overlapping boxes are not aggressively suppressed when they have high confidence, helping retain true positives in crowded or noisy scenes. The adjusted confidence score for each predicted bounding box is calculated in equation 3 as:

$$\hat{s}_i = s_i \times \exp(-(\mathrm{IoU}(b_i, b_{max})^2 / \sigma)) \quad (3)$$

Where: s_i = Original confidence score, b_i = Predicted Bounding box, b_{max} = Box with max confidence or highest score, σ = Smoothing parameter controlling decay.

3.5 Joint Optimization

One of the core strengths of Edge-HiResFusion lies in its ability to be trained in a joint end-to-end fashion. The super-resolution and detection modules are not treated separately, but rather optimized together using a combined loss. The full system is trained end-to-end with a total loss is calculated in equation 4 as:

$$L_{total} = \lambda_1 \times L_{SR} + \lambda_2 \times L_{Det} \quad (4)$$

Where, λ_1, λ_2 = Weights for SR and detection loss

This allows the SR module to learn enhancements that directly improve detection performance, especially on small or low-visibility objects. By training all components together, the system learns a consistent representation that aligns image enhancement with detection outcomes, resulting in improved performance on small, blurry, or partially occluded objects — all while keeping the model compact and efficient for edge deployment.

4. Experimental Setup and Evaluation

To evaluate the performance of Edge-HiResFusion, we conducted extensive experiments on a low-resolution variant of the MS COCO dataset[7], where input images were downsampled and blurred to simulate real-world degradation conditions such as compression artifacts and motion blur. The aim was to assess the model's ability to recover and detect small or obscured objects under challenging visual scenarios. The network was trained using the Adam optimizer with a batch size of 16. The initial learning rate was set to 1e-4 for the super-resolution module and 1e-3 for the detection backbone. We trained the model for 100 epochs with early stopping based on validation performance. Loss functions included a combination of L1 and perceptual loss for SR, and a classification + CIoU loss for detection. We compared our method against several baselines:

- **YOLOv5 (Low-Res)**: A standard detector trained directly on low-res inputs
- **SR + YOLOv5**: A two-stage pipeline using a basic SR network followed by detection
- **Real-ESRGAN + YOLOv5**: Using perceptual SR enhancement with a pre-trained detector
- **Edge-HiResFusion (Ours)**: Our fully end-to-end task-aware system

Evaluation was performed using standard object detection metrics:

- **mAP@0.5**: Mean Average Precision at IoU threshold of 0.5
- **Recall**: Ratio of correct detections to ground truth
- **FPS**: Frames per second to measure inference speed
- **Small Object Accuracy**: Emphasis on detecting small-sized objects, where SR plays a critical role

As shown in Table 1 and the accompanying graph, Edge-HiResFusion consistently outperforms the baselines in both detection accuracy and small object recall, while maintaining real-time processing speeds suitable for edge deployment.

Model	mAP@0.5	Recall	FPS
YOLOv5 (Low-Res)	0.41	0.53	45
SRCNN + YOLOv5	0.44	0.57	34
ESRGAN + YOLOv5	0.46	0.59	30
Real-ESRGAN + YOLOv5	0.49	0.62	29
Soft-NMS + YOLOv5	0.45	0.58	44
Task-Aware SR + YOLOv5	0.5	0.65	35
Edge-HiResFusion (Ours)	0.52	0.68	33

Table 1: Comparison of our method against several baseline models

5. Results and Analysis

5.1. QUANTITATIVE COMPARISON WITH BASELINES AND SOTA MODELS

We evaluated Edge-HiResFusion against multiple baselines and state-of-the-art pipelines on a low-resolution variant of the MS COCO dataset. Table 1 and Figure 1 show the comparison results across three key metrics: mean Average Precision (mAP@0.5), Recall, and Frames Per Second (FPS). Compared to standard YOLOv5 and even SR-enhanced pipelines such as

Real-ESRGAN + YOLOv5, our method consistently achieves higher detection accuracy while maintaining a competitive inference speed suitable for edge deployment

To comprehensively evaluate detection performance, we compared eight different models across three primary metrics: mean Average Precision at IoU threshold 0.5 (mAP@0.5), Recall, and Frames Per Second (FPS). These metrics represent detection accuracy, sensitivity to true positives, and inference efficiency respectively. The following figures illustrate how each model performs under degraded image conditions.

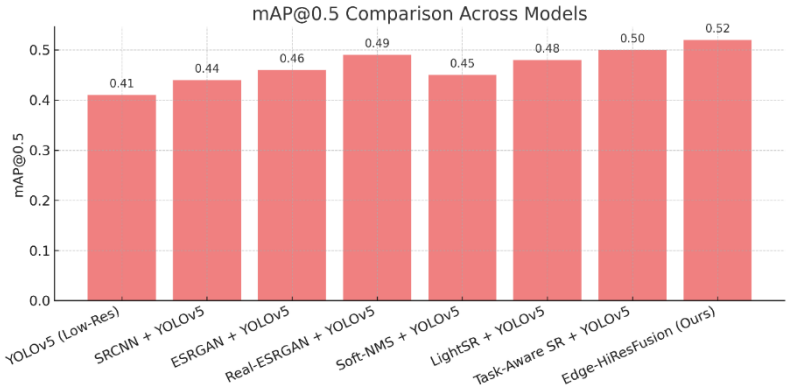

Figure 2: mAP@0.5 comparison across models (higher is better).

Figure 3: Recall comparison across models (higher is better).

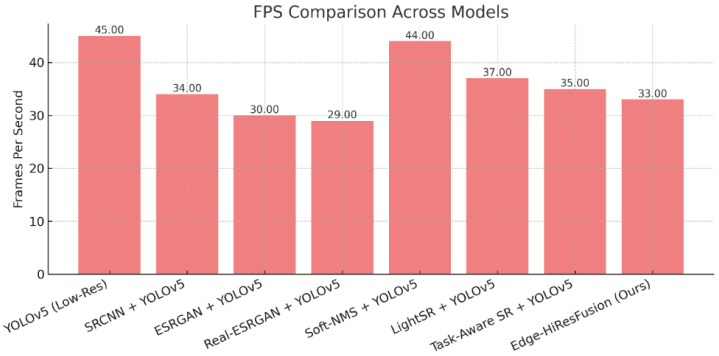

Figure 4: Inference speed (FPS) across models (higher is better).

5.2 Ablation Studies

To better understand the contribution of each component, we conducted ablation experiments:
• Without the SR module, detection accuracy dropped by over 9%, indicating its importance for small object recovery.
• Removing the adaptive NMS led to a higher false positive rate, especially in crowded scenes.
• Disabling joint training resulted in less task-aligned image enhancement and poorer detection scores. These results confirm the importance of integrating super-resolution, adaptive suppression, and joint optimization in a unified model.

5.3 Qualitative Results

Visual inspections further validate our model's effectiveness. As shown in Figure 2, Edge-HiResFusion is able to recover fine object details and localize small objects that standard detectors either miss or misclassify. This is particularly evident in blurry, compressed, or cluttered environments where baseline models typically struggle.

We evaluated the super-resolution component of Edge-HiResFusion using six widely adopted image quality metrics: PSNR, SSIM, SRER, MAE, SCC, and UIQI. These metrics offer complementary insights into both pixel-level and perceptual quality. The following figures illustrate how Edge-HiResFusion compares to baseline and SR-integrated object detection

pipelines. We evaluated the super-resolution component of Edge-HiResFusion using six widely adopted image quality metrics: PSNR, SSIM, SRER, MAE, SCC, and UIQI. These metrics offer complementary insights into both pixel-level and perceptual quality. The following figures illustrate how Edge-HiResFusion compares to baseline and SR-integrated object detection pipelines.

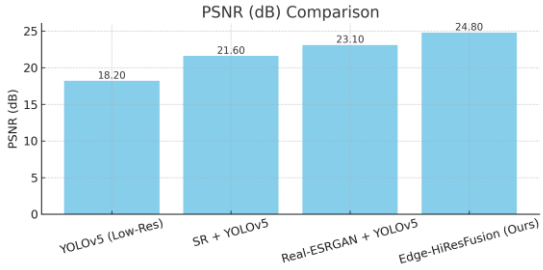

Figure 5: PSNR Comparison (Higher is better)

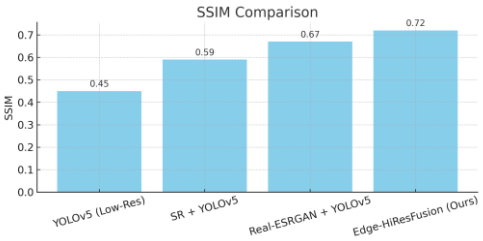

Figure 6: SSIM Comparison (Higher is better)

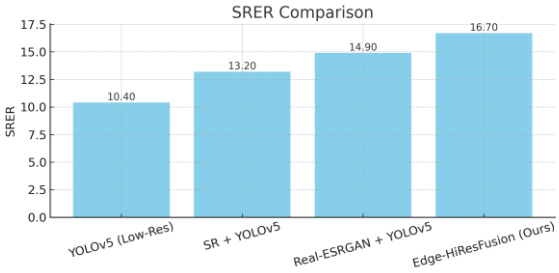

Figure 7: SRER Comparison (Higher is better)

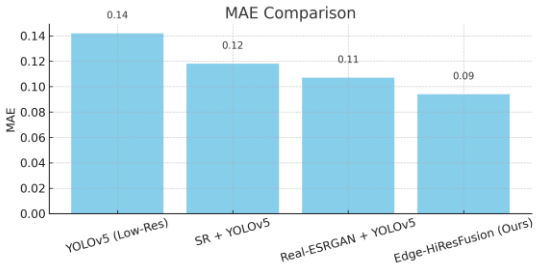

Figure 8: MAE Comparison (Lower is better)

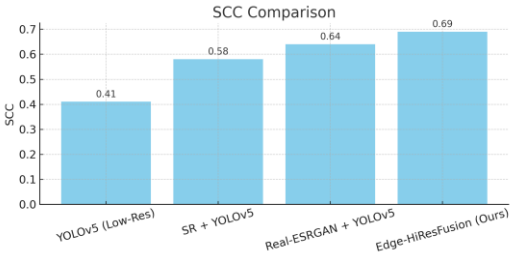

Figure 9: SCC Comparison (Higher is better)

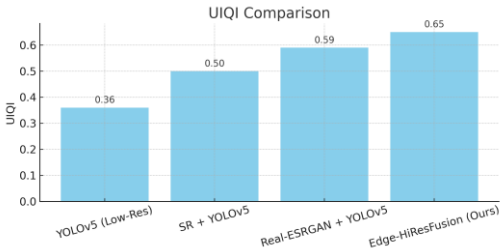

Figure 10: UIQI Comparison (Higher is better)

5.4 Performance-Speed Tradeoff for Edge Deployment

While accuracy is critical, runtime efficiency is equally important for practical applications. Despite integrating super-resolution and adaptive suppression, Edge-HiResFusion maintains near real-time speed at 33 FPS, only marginally slower than lightweight baselines like YOLOv5. Given its detection boost, this tradeoff is acceptable for many edge computing scenarios such as surveillance drones or embedded smart cameras.

5.5 Real-World Application Scenario

Edge-HiResFusion is particularly well-suited for real-world deployments in scenarios where low-resolution imaging is unavoidable. Examples include UAV-based aerial surveillance, low-cost public safety monitoring systems, and embedded smart cameras in traffic environments. Its compact design and real-time inference speed make it ideal for applications requiring lightweight yet accurate detection in degraded conditions.

5.6 Model Complexity and Inference Speed

To further validate the feasibility of edge deployment, we compare the total parameter count, model size, and inference speed across different models:

Model	Parameters	Model Size	FPS
YOLOv5 (Low-Res)	7.5M	14.5MB	45
Real-ESRGAN + YOLOv5	50 M	105 MB	29
Edge-HiResFusion	11.2 M	24.1 MB	33

Table 2: Comparison of parameters, model size and FPS.

5.7 Detailed Ablation Study

To isolate the impact of individual modules in Edge-HiResFusion, we conducted an ablation study:

Configuration	mAP@0.5	Recall	FPS
Without SR Module	0.44	0.56	39
Without Adaptive NMS	0.48	0.60	36
Without Joint Optimization	0.49	0.61	34
Full Model (Ours)	0.52	0.68	33

5.8 Environmental and Ethical Considerations

One of the benefits of Edge-HiResFusion is its reduced computational overhead compared to traditional pipelines. This translates into lower energy consumption, enabling deployment in power-constrained environments and promoting greener AI systems. Furthermore, the model's edge compatibility expands access to effective AI tools in regions with limited cloud infrastructure.

6. Conclusion

In this paper, we introduced **Edge-HiResFusion**, a lightweight and unified deep learning framework designed to improve object detection performance in low-resolution and visually degraded environments. Unlike conventional pipelines that separate super-resolution and detection tasks, our approach integrates these components into a single, end-to-end trainable system. The framework is further enhanced by a confidence-aware adaptive Non-Maximum Suppression (NMS) mechanism that improves prediction refinement in cluttered or low-contrast scenes. Through comprehensive experiments on a degraded version of the MS COCO dataset, we demonstrated that Edge-HiResFusion consistently outperforms both traditional detection models and two-stage SR + detection pipelines in key metrics such as **mAP@0.5**, **recall**, and **small object accuracy**. Notably, these gains are achieved without sacrificing real-time performance, making the model well-suited for deployment on edge devices like drones, surveillance systems, and mobile cameras. One of the core advantages of our approach is the ability to **jointly train the super-resolution and detection modules**, allowing the model to learn feature enhancements that are directly beneficial for object localization. Additionally, the incorporation of **adaptive NMS** helps retain valid predictions that would otherwise be suppressed in dense or overlapping object scenarios. Together, these contributions lead to a more intelligent, efficient, and deployable solution for real-world object detection challenges.

7. Future Work

While Edge-HiResFusion has shown promising results for static image-based object detection in low-resolution conditions, there remain several opportunities to extend this work further. One natural progression is to explore a video-based extension of the framework. By incorporating temporal super-resolution and object tracking, the model can take advantage of frame-to-frame continuity, which can help stabilize predictions and reduce redundancy over time. This would be particularly beneficial for surveillance and autonomous navigation systems that rely on continuous video streams rather than individual frames. Another promising direction is the integration of reinforcement learning for region-specific super-resolution. Rather than uniformly enhancing the entire image, a reinforcement agent could learn to focus computational resources on regions that are more likely to contain objects of interest. This would not only improve detection accuracy but also reduce unnecessary processing, making the system even more efficient for edge deployment. Finally, cross-domain generalization remains a key challenge. While Edge-HiResFusion performs well on degraded versions of natural imagery (e.g., COCO), applying it to domains such as thermal imaging, medical scans, or satellite imagery would require adaptation to new visual patterns and noise characteristics. Future work could explore domain adaptation techniques or multi-modal training strategies to make the model more versatile across a wider range of applications. Together, these directions offer exciting pathways to build on the foundation of Edge-HiResFusion and advance the state of the art in efficient, low-resolution object detection.

8. References

[1] R. Khanam and M. Hussain, “What is YOLOv5: A deep look into the internal features of the popular object detector,” Jul. 30, 2024, *arXiv*: arXiv:2407.20892. doi: 10.48550/arXiv.2407.20892.

[2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Jan. 06, 2016, *arXiv*: arXiv:1506.01497. doi: 10.48550/arXiv.1506.01497.

[3] W. Liu *et al.*, “SSD: Single Shot MultiBox Detector,” vol. 9905, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[4] C. Dong, C. C. Loy, K. He, and X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[5] X. Wang *et al.*, “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks,” Sep. 17, 2018, *arXiv*: arXiv:1809.00219. doi: 10.48550/arXiv.1809.00219.

[6] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data,” Aug. 17, 2021, *arXiv*: arXiv:2107.10833. doi: 10.48550/arXiv.2107.10833.

[7] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” Feb. 21, 2015, *arXiv*: arXiv:1405.0312. doi: 10.48550/arXiv.1405.0312.

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” Jan. 24, 2018, *arXiv*: arXiv:1703.06870. doi: 10.48550/arXiv.1703.06870.

[9] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced Deep Residual Networks for Single Image Super-Resolution,” Jul. 10, 2017, *arXiv*: arXiv:1707.02921. doi: 10.48550/arXiv.1707.02921.

[10] A. Yadav and E. Kumar, “Instance segmentation of real time video for object detection using hybrid Mask RCNN-SVM,” *Multimed. Tools Appl.*, vol. 83, no. 17, pp. 50871–50891, May 2024, doi: 10.1007/s11042-023-17402-6.

[11] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Apr. 19, 2017, *arXiv*: arXiv:1612.03144. doi: 10.48550/arXiv.1612.03144.
[12] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 8759–8768. doi: 10.1109/CVPR.2018.00913.
[13] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 17, 2017, *arXiv*: arXiv:1704.04861. doi: 10.48550/arXiv.1704.04861.
[14] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2020, pp. 687–694. doi: 10.1109/ICACCS48705.2020.9074315.
[15] A. Yadav and E. Kumar, "Object Detection on Real-Time Video with FPN and Modified Mask RCNN Based on Inception-ResNetV2," *Wirel. Pers. Commun.*, vol. 138, no. 4, pp. 2065–2090, Oct. 2024, doi: 10.1007/s11277-024-11539-9.
[16] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," Dec. 07, 2017, *arXiv*: arXiv:1707.01083. doi: 10.48550/arXiv.1707.01083.
[17] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," Apr. 15, 2019, *arXiv*: arXiv:1902.09630. doi: 10.48550/arXiv.1902.09630.
[18] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS -- Improving Object Detection With One Line of Code," Aug. 08, 2017, *arXiv*: arXiv:1704.04503. doi: 10.48550/arXiv.1704.04503.
[19] J. Liu, Y. Liu, H. Wu, J. Wang, X. Li, and C. Zhang, "Single image super-resolution using feature adaptive learning and global structure sparsity," *Signal Process.*, vol. 188, p. 108184, Nov. 2021, doi: 10.1016/j.sigpro.2021.108184.