ENHANCING LOW-LIGHT IMAGE QUALITY FOR OBJECT DETECTION USING SUPER RESOLUTION

Yusuf Demir & Ender Orman
Department of Computer Engineering
Hacettepe University, Ankara, Türkiye
{yusuf_demir, enderorman}@hacettepe.edu.tr

ABSTRACT

This study explores how deep learning-based super-resolution (SR) methods can improve both the visual quality and object detection performance of low-light images. We evaluate several SR models, starting from a simple SRCNN and adding components like residual blocks, Squeeze-and-Excitation (SE), and Convolutional Block Attention Module (CBAM). Our most advanced model, YASRNet, combines spatial and channel attention with efficient upsampling.

To test the impact, we apply a YOLOv8 detector to three types of images: low-light inputs, super-resolved outputs, and high-light ground truths. While advanced SR models improve perceptual metrics like SSIM and PSNR, they do not always help detection. In some cases, models with higher SSIM perform worse in detecting objects than simpler ones. This shows that traditional SR metrics do not always match detection performance under low-light conditions.

Our results show a clear trade-off between making images look better and helping detectors perform better. This highlights the need for SR models that focus on preserving edges, not just improving visual appearance.

1 INTRODUCTION

Object detection systems often show lower performance on low-light or low-resolution images. These conditions are common in real-world applications such as surveillance, autonomous vehicles, and mobile devices, where sensor quality and lighting are limited. Improving image quality before detection can help the system perform more reliably.

Super-resolution (SR) methods aim to reconstruct high-quality images from low-resolution inputs. Deep learning-based SR, especially with Convolutional Neural Networks (CNNs), has shown strong results. One of the earliest models, SRCNN, learns a direct mapping between low- and high-resolution images. Later improvements include deeper layers, attention blocks, and advanced upsampling techniques.

In this study, we apply several SR models to low-light images and evaluate how they affect object detection. Instead of focusing only on visual quality metrics like SSIM or PSNR, we measure the change in detection results using a YOLOv8 model. Our findings show that higher SSIM does not always lead to better detection. In fact, simpler SR models that better preserve edge structures often help YOLO more than visually refined outputs. This highlights the need for task-aware SR design when used with detection systems.

2 RELATED WORK

Early super-resolution (SR) methods, such as bicubic interpolation, often produced blurry results and lacked the ability to recover fine details. The introduction of deep learning, particularly convolutional neural networks (CNNs), brought significant improvements. SRCNN (1) was among the first to use CNNs for SR, learning mappings between low- and high-resolution images. Subsequent models like VDSR (2) and EDSR (3) introduced deeper architectures and residual learning, enhancing reconstruction quality. ESRGAN (4) further improved perceptual quality by incorporating adversarial loss and refined residual structures.

While these models excel in enhancing visual quality, their impact on downstream tasks like object detection is not always positive. Shermeyer and Van Etten (5) demonstrated that applying SR to satellite imagery improved detection performance only under certain conditions, with diminishing returns at higher resolutions. Similarly, Haris et al. (6) proposed a task-driven SR approach, integrating detection loss into the SR training process, leading to better detection outcomes compared to traditional SR methods.

In low-light conditions, object detection becomes particularly challenging due to noise and poor illumination. Noh et al. (7) emphasized the importance of precise supervision in feature-level SR for small object detection, highlighting that naive SR can sometimes degrade detection performance. Yin et al. (8) introduced PE-YOLO, combining a pyramid enhancement network with YOLOv3, achieving improved detection in dark environments by enhancing image details and suppressing noise.

These studies underscore that while SR can enhance image quality, its benefits for object detection, especially under challenging conditions like low light, depend on task-specific adaptations. Our work builds upon these insights, evaluating various SR architectures and their effects on YOLOv8 detection performance in low-light environments, emphasizing the need for edge-preserving techniques tailored for detection tasks.

3 Dataset and Preprocessing

The images used for super-resolution training were obtained from the LOL dataset (9). All images were used with their original sizes to preserve spatial detail, which is important for object detection performance using YOLOv8. The images were normalized to the $[0, 1]$ range by converting them to tensors.

4 Method

This study examines the effect of image super-resolution on object detection performance in low-light conditions. The proposed pipeline consists of two main stages: an image enhancement module that restores visual quality through super-resolution, and an object detection module that evaluates the impact of enhancement on detection accuracy.

4.1 Pipeline Overview

In the proposed pipeline, low-light images are first processed by a super-resolution model to generate enhanced versions. These enhanced images are subsequently passed to a YOLO-based object detector. To assess the effectiveness of super-resolution as a preprocessing step, detection performance on enhanced images is compared to that on the original low-light inputs.

4.2 Super-Resolution Models

We evaluate four different convolutional architectures for super-resolution. The first is the classical SRCNN, which serves as a simple and interpretable baseline. EnhancedSRCNN extends this by adding residual blocks to improve feature learning and representation capacity. EnhancedSRCNNv2 introduces Squeeze-and-Excitation (SE) attention mechanisms to selectively emphasize informative channels. YASRNet further integrates spatial and channel attention via CBAM and uses pixel shuffle for efficient upsampling. Additionally, we propose an edge-aware training approach for SRCNN using a hybrid loss based on MSE and Laplacian-filtered edge differences, aiming to preserve object boundaries that are important for detection tasks.

4.3 Baseline SRCNN Model

Our baseline super-resolution model follows the original Super-Resolution Convolutional Neural Network (SRCNN), one of the earliest deep learning methods for single-image super-resolution. It consists of three convolutional layers with ReLU activations and a skip connection from input to output.

The architecture uses large kernels for feature extraction and progressively reconstructs the output. A residual connection helps the model learn finer details by focusing on high-frequency components.

Although lightweight and fast to train, SRCNN offered limited benefits in downstream object detection. While perceptual quality improved slightly, detection performance remained close to that of the original low-light inputs, highlighting the challenge of bridging perceptual gains and detection utility.

4.4 Enhanced SRCNN Model

The Enhanced SRCNN model extends the baseline architecture by adding depth through residual blocks. These additions improve the network's capacity to capture more complex structures and finer details.

The model starts with a 9×9 convolutional layer followed by a ReLU activation, similar to the baseline. It then includes two residual blocks, each consisting of two 3×3 convolutional layers with skip connections. This design promotes better gradient flow and enables the network to focus on high-frequency content.

While this deeper architecture led to a small improvement in SSIM, object detection performance slightly decreased compared to the baseline. This suggests that enhanced perceptual quality does not always translate into better detection results, especially under low-light conditions.

4.5 Enhanced SRCNN v2 Model

The Enhanced SRCNN v2 model expands upon earlier SRCNN variants by combining residual learning with channel attention mechanisms. It consists of four residual blocks and two Squeeze-and-Excitation (SE) blocks, allowing the network to focus on informative channels and recover finer textures.

This architecture significantly increased SSIM scores compared to simpler models, showing clear improvement in perceptual quality. However, object detection performance degraded when using its outputs with YOLOv8. This suggests that improving SSIM alone does not guarantee better detection results.

SSIM primarily reflects perceived image quality and smoothness, but object detectors rely on sharp edges and localized structural features. The attention and deep residual components may introduce smoothing that reduces contrast at object boundaries, negatively impacting detection.

Similar findings were reported by (5), who showed that super-resolution improved image quality but not always detection accuracy in satellite imagery. Our results confirm this gap, emphasizing that SR models used for detection tasks must prioritize edge preservation rather than perceptual enhancement alone.

Finally, a 5×5 convolutional layer reconstructs the super-resolved output image from the refined feature maps. By combining deep residual learning and attention mechanisms, the Enhanced SR-CNN v2 model is capable of generating high-quality high-resolution images with improved detail restoration and texture sharpness.

4.6 YASRNet Model

The YASRNet (Yet Another Super-Resolution Network) architecture is a high-capacity model designed to perform single-image super-resolution by combining deep residual learning, hybrid attention mechanisms, and efficient upsampling strategies. This model enhances fine detail restoration by integrating both channel and spatial attention modules, based on the Convolutional Block Attention Module (CBAM) (10).

The network begins with an entry module composed of a 9×9 convolutional layer with 64 filters followed by a ReLU activation. This wide kernel captures extensive contextual information from the low-resolution input image.

The core feature extraction is performed through two stages. Each stage consists of two residual blocks followed by a CBAM attention block. The residual blocks enable the network to learn more complex transformations while preserving information across layers through skip connections. Each CBAM block applies attention in two sequential steps:

- **Channel Attention:** Global average pooling and max pooling are applied along the spatial dimensions to generate two descriptors. These are passed through shared multilayer perceptrons (implemented as 1×1 convolutions with a reduction ratio) and combined via element-wise addition, followed by a sigmoid activation to produce channel-wise weights.
- **Spatial Attention:** Average and max pooling are applied across the channel dimension, and the resulting descriptors are concatenated and passed through a 7×7 convolution, followed by a sigmoid activation to yield spatial attention weights.

This hybrid attention allows the model to adaptively focus on the most informative feature maps and spatial regions.

For upsampling, the network uses a convolutional layer to increase the number of channels to $4\times$ the original count, followed by a `PixelShuffle` operation to upscale the spatial resolution by a factor of 2. This method avoids checkerboard artifacts often associated with transposed convolutions and is efficient in memory and computation.

Finally, the output layer is a 5×5 convolution that maps the refined high-resolution feature map to the output image. Overall, YASRNet effectively combines attention-driven enhancement and upsampling techniques to produce high-quality super-resolved outputs.

4.7 Object Detection

We evaluated the effect of super-resolution on object detection using a fixed-weight YOLOv8l model. To ensure fair comparison, no fine-tuning was performed. A confidence threshold of 0.8 was applied to filter out low-confidence predictions.

Detections were analyzed across three input types: original low-light images, super-resolved outputs, and high-light ground truth images. For each, we measured the number of detected objects and the average confidence score per image.

This filtering helped reduce noise and allowed us to better assess how image enhancement affects detection performance in low-light conditions.

5 Experimental Results

This section presents the performance evaluation of the baseline SRCNN and the proposed Enhanced SRCNN models. We assess image enhancement quality using both visual comparisons and quantitative metrics, including PSNR and SSIM.

In addition to image quality metrics, we evaluate the impact of super-resolution on object detection performance. Object detection is performed on high-light images, low-light images, and super-resolved outputs produced by the Enhanced SRCNN model. This comparison allows us to assess how well the enhanced images support downstream tasks such as object detection.

5.1 Quantitative Evaluation of Super-Resolution Performance

To assess the reconstruction quality of each model, we compute two standard image quality metrics over the test set: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

PSNR measures the pixel-wise similarity between the predicted image and the ground truth, with higher values indicating better fidelity. SSIM, on the other hand, evaluates perceptual similarity by comparing luminance, contrast, and structural information, and is more aligned with human visual perception.

As summarized in Table 1, the best-performing models are **YASRNet** and **EnhancedSRCNNv2**, achieving PSNR scores of 22.19 and 20.64 dB, and SSIM scores of 0.8367 and 0.7958, respectively. These models incorporate deeper feature extraction, residual connections, attention mechanisms (SE or CBAM), and efficient upsampling (PixelShuffle), all of which contribute to preserving fine textures and structural details.

In contrast, the baseline **SRCNN** model achieved a PSNR of 17.26 dB and SSIM of 0.7298. Variants such as **SRCNN_Edge** and **SRCNN_ssim**, which introduced hybrid loss functions incorporating edge or perceptual similarity, showed slight improvements in SSIM but limited gains in PSNR.

These results demonstrate that modern enhancements—particularly attention and deeper residual learning—lead to notable improvements in perceptual quality, even if pixel-wise gains are modest.

Model	Epochs	PSNR (dB)	SSIM
YASRNet	30	22.19	0.8367
EnhancedSRCNNv2	30	20.64	0.7958
EnhancedSRCNN	30	17.37	0.7456
SRCNN	30	17.26	0.7298
SRCNN_Edge	30	16.79	0.7128

Table 1: PSNR and SSIM scores for each model after 30 training epochs.

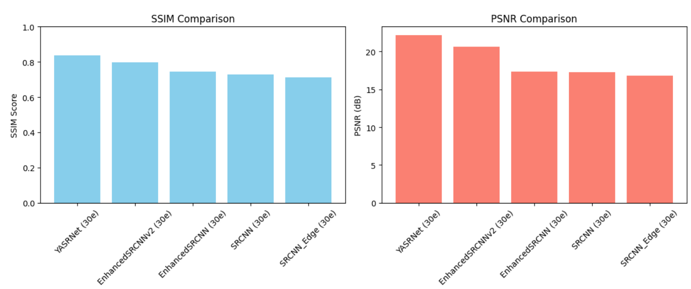
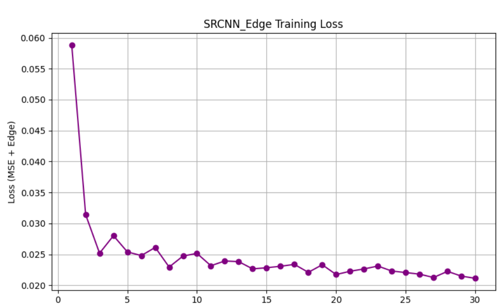


Figure 1: Comparison of average PSNR and SSIM scores between SRCNN and Enhanced SRCNN.

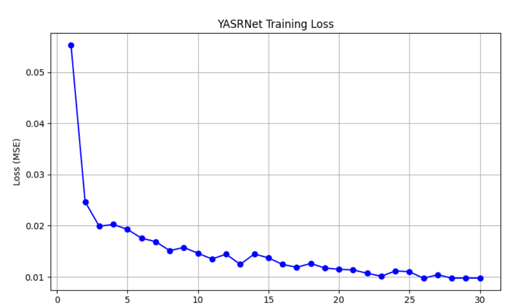
5.2 Training Loss Analysis

To illustrate the training behavior of different models, we include two representative training loss curves. Figure 2 shows the loss trend for **SRCNN_Edge**, which incorporates an edge-aware loss function, and for **YASRNet**, which uses a deeper architecture with attention mechanisms.

Both models demonstrate stable convergence over 30 epochs, with variations reflecting differences in architectural complexity and loss formulation.



(a) SRCNN_Edge Loss Curve



(b) YASRNet Loss Curve

Figure 2: Training loss curves over 30 epochs for SRCNN_Edge and YASRNet models.

5.3 Runtime Environment and Performance

All experiments were conducted using a system equipped with **2× NVIDIA T4 GPUs**, each with 16 GB of VRAM. The models were trained using PyTorch with mixed precision disabled, and training was performed sequentially per model to ensure consistent resource allocation.

Table 2 summarizes the total training time (in seconds) required by each model over 30 epochs.

Model	**Training Time (s)**
SRCNN	1057.36
EnhancedSRCNN	1646.68
EnhancedSRCNNv2	2506.08
YASRNet	4819.63
SRCNN_Edge	1063.20

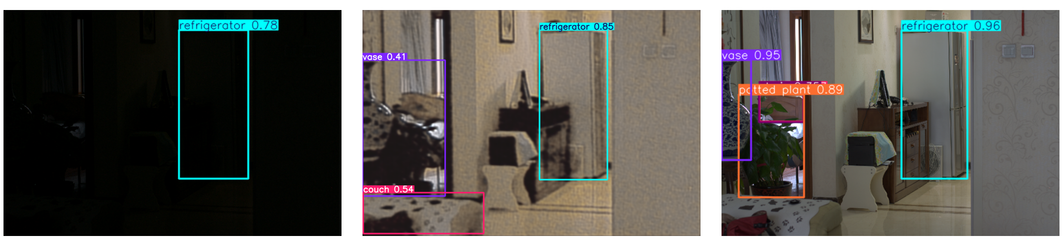
Table 2: Total training time (in seconds) for each model using 2× NVIDIA T4 GPUs.

5.4 Object Detection Performance After Super-Resolution Enhancement

To assess the impact of super-resolution on downstream tasks, object detection was performed using a YOLOv8l model. Among the evaluated super-resolution models, **YASRNet** was selected to generate enhanced images due to its superior reconstruction quality.

Object detection was conducted on three image types: low-light inputs, YASRNet-enhanced outputs, and high-light ground truth images. This evaluation aims to determine whether super-resolution can enhance object visibility and detectability in degraded lighting conditions.

Good Detection Example

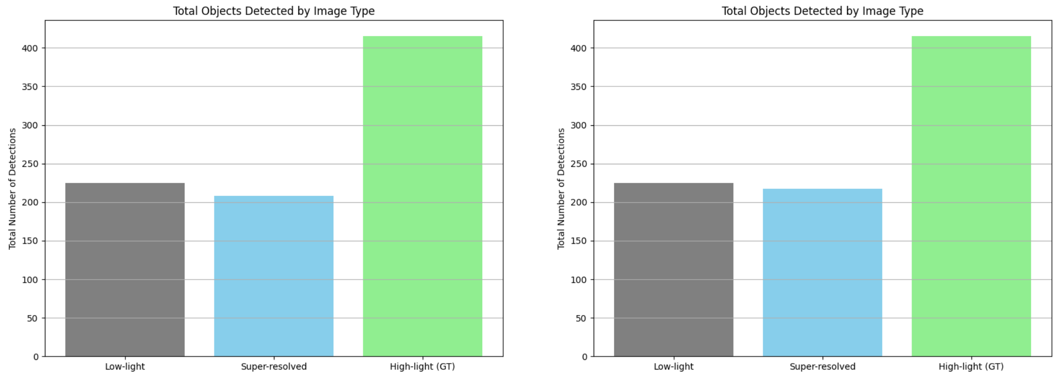


(a) Low-light Input (b) Super-Resolved (YASRNet) (c) High-light Ground Truth

Poor Detection Example

(d) Low-light Input (e) Super-Resolved (YASRNet) (f) High-light Ground Truth

Figure 3: YOLOv8l detection examples: top row shows improved results after super-resolution; bottom row shows limited effect.



(a) Detection counts: YASRNet

(b) Detection counts: SRCNN_Edge

Figure 4: Total number of objects detected by YOLOv8l for YASRNet and SRCNN_Edge super-resolved images.

6 Discussions and Conclusions

The results of this project demonstrate that while super-resolution techniques can significantly improve perceptual image quality, they do not always lead to better performance in downstream tasks such as object detection. Our enhanced models, including those incorporating residual and attention mechanisms, achieved noticeably higher SSIM scores compared to the baseline SRCNN, indicating better preservation of structural details at a perceptual level.

However, these improvements did not consistently translate to better object detection performance when evaluated using the YOLOv8l model. In some cases, the super-resolution process—even with enhanced architectures—introduced smoothing effects that diminished important edge and texture features. These features are critical for object detection, as YOLO relies heavily on localized patterns and sharp boundaries to accurately classify and localize objects.

This finding highlights a key challenge in applying super-resolution as a preprocessing step: improving visual quality for human perception does not always align with the requirements of machine perception. In fact, models trained to optimize for perceptual metrics may inadvertently remove fine-grained details essential for detection networks.

Overall, the project successfully explored multiple super-resolution architectures and their perceptual impact, but also revealed important trade-offs when integrating SR into a broader computer vision pipeline. Future work may explore loss functions or training objectives that balance both human-centric quality metrics and task-specific feature preservation for models like YOLO.

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *European Conference on Computer Vision (ECCV)*, 2014, pp. 184–199. [Online]. Available: https://arxiv.org/pdf/1501.00092.pdf

[2] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/papers/Kim_Accurate_Image_Super-Resolution_CVPR_2016_paper.pdf

[3] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf

[4] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, “ESRGAN: Enhanced super-resolution generative adversarial networks,” in *Eu-*

ropean Conference on Computer Vision Workshops (ECCVW), 2018. [Online]. Available: https://openaccess.thecvf.com/content_ECCVW_2018/papers/11133/Wang_ESRGAN_Enhanced_Super-Resolution_Generative_Adversarial_Networks_ECCVW_2018_paper.pdf

[5] J. Shermeyer and A. V. Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1432–1441. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2019/papers/EarthVision/Shermeyer_The_Effects_of_Super-Resolution_on_Object_Detection_Performance_in_Satellite_CVPRW_2019_paper.pdf

[6] M. Haris, G. Shakhnarovich, and N. Ukita, "Task-driven super resolution: Object detection in low-resolution images," in *arXiv preprint*, 2018. [Online]. Available: https://arxiv.org/abs/1803.11316

[7] J. Noh, W. Bae, W. Lee, J. Seo, and G. Kim, "Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 9725–9734. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/papers/Noh_Better_to_Follow_Follow_to_Be_Better_Towards_Precise_Supervision_ICCV_2019_paper.pdf

[8] X. Yin, Z. Yu, Z. Fei, W. Lv, and X. Gao, "Pe-yolo: Pyramid enhancement network for dark object detection," in *Proceedings of the 2023 International Conference on Computer Vision (ICCV)*, 2023. [Online]. Available: https://arxiv.org/abs/2307.10953

[9] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference (BMVC)*, 2018. [Online]. Available: https://arxiv.org/pdf/1808.04560

[10] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19. [Online]. Available: https://arxiv.org/abs/1807.06521