International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.3 (2025) pp. 5483-5494
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Enhancing Real-Time Object Detection in Low-Light Conditions Using Zero-DCE and Super-Resolution GANs: A YOLO-Based Approach

Rugved Deshpande[1], Aarushi Singh[2], Pranshu Pranjal[3*]

[1] Department of Computer Science and Engineering, Manipal University Jaipur, Rajasthan 303007
Email: rugved1405.work@gmail.com - **ORCID:** 0000-0001-6213-7132

[2] Department of Artificial Intelligence and Machine Learning, Manipal University Jaipur, Rajasthan 303007
Email: aarushi2022singh@gmail.com - **ORCID:** 0000-0001-6213-7130

[3] Department of Artificial Intelligence and Machine Learning, Manipal University Jaipur, Rajasthan 303007
* **Corresponding Author Email:** pranshu.iirs@gmail.com - **ORCID:** 0000-0001-6213-7131

Article Info:

DOI: 10.22399/ijcesen.3455
Received : 22 May 2025
Accepted : 17 July 2025

Keywords

Low-Light Image Enhancement
Zero-DCE
YOLOv8
Real-Time Object Detection
ESRGANs
Rescue and Search Operations

Abstract:

Low-light conditions significantly degrade the performance of real-time object detection systems. This study proposes a novel pipeline that integrates Zero-Reference Deep Curve Estimation (Zero-DCE), which has been used to enhance the low-light image, and Enhanced Super-Resolution Generative Adversarial Networks (ESRGANs) for improving the object detection accuracy in poor illumination condition for resolution refinement. The enhanced images are then processed through a YOLO-based detector for real-time object identification. Zero-DCE is leveraged to enhance image illumination without requiring reference images or paired datasets, ensuring efficient and adaptive enhancement across diverse lighting conditions. Following enhancement, ESRGAN is applied to increase the perceptual quality and fine-grained details of the images, enabling the detection model to capture subtle features that are often lost in low-light inputs. This dual stage preprocessing significantly improves the visibility and quality of the input images, directly benefiting object detection performance. The experimental evaluation, conducted on low-light datasets, demonstrates substantial improvements in detection accuracy, precision, and recall metrics. Furthermore, the proposed pipeline maintains real-time performance that can be suitable for surveillance, autonomous navigation, and security applications.

1. Introduction

Computer vision applications have seen an enormous jump in recent times in every field including self-driving vehicles [1], medical imaging [2], face recognition used in attendance systems [3], crop monitoring [4], geological studies [5], security surveillance [6], object recognition and track-ng, augmented and virtual reality [7], etc. However, the performance of these models drops significantly in low-light conditions, restricting disaster rescue missions, night-time surveillance systems, and military reconnaissance. Images and videos captured in poor illumination are affected by low brightness, low contrast, high noise levels, lack detail due to blurriness, chromatic aberrations, and severe underexposure, which collectively affect subjective visual interpretations on human eyes due to quality and render traditional detection models ineffective [8]. Low light, as the term suggests, denotes ambient situations in which illuminance falls below the usual threshold [9].

This deterioration leads to a significant reduction in the extraction of structural and textural features in the captured image/video. This loss of critical information subsequently hinders the efficacy of further analysis on it, including object detection and recognition. The impact of these degradations is particularly witnessed in scenarios where the target object is very small in the picture frame or experiences partial occlusion, as these situations rely solely on subtle visual cues for more precise analysis. The degradation is not only due to brightness alone, but is due to the combined loss of texture, contrast, and semantic boundaries. Due to low contrast and non-uniform lightning, information about the image is masked get lost, restricting the use of real-world applications such as

remote sensing and lane detection [10]. Although basic methods for improving low-light images, like adjusting the brightness range using histogram equalization or using Retinex-based algorithms, can provide some improvement, they often fall short. These simpler approaches can either be too basic to handle difficult lighting conditions effectively or tend to create artificial-looking distortions. Despite encouraging results, this technology is still in its infancy. Specifically, the available algorithms frequently perform better in one area than in another [8].

In the recent past, things have taken a turn towards more of a deep-learning approach in image enhancement. Zero-DCE, for example, uses unsupervised learning to adjust the brightness of each pixel, without the use of any prior reference image. However, the low spatial resolution i.e., the lack of finer details gets compromised on how well the objects are detected in applications like security surveillance and rescue missions. Low-quality photographs are characterised by low-resolution cameras, adverse weather conditions, or blurriness, which complicates the differentiation of objects from the backdrop [11]. So, to improve this, super-resolution methods such as super-resolution Generative Adversarial Networks (ESRGANs) have been adopted for the pipeline. ESRGANs help in the reconstruction of finer details enabling the detection of small object features that would otherwise be missed in lower-quality images. This approach has also been successful in remote sensing and aerial vehicle detection. Image super-resolution for small objects by improving the current super-resolution framework by incorporating a cyclic generative adversarial network (GAN) and residual feature aggregation (RFA) significantly improves detection performance [12]. Therefore, based on all these advances, we propose a novel three-stage low-light object detection pipeline: Zero-DCE for unsupervised real-time enhancement, ESRGAN for super-resolution image construction, and YOLOv8 for efficient object detection.

This pipeline offers advancements in both visual quality as well as in semantic visual recovery, well suited for real-time and resource-limited applications like surveillance and disaster rescue. Unlike prior methods which processed each stage independently, this integrated pipeline enhances end-to-end detection accuracy and generalizability.

2. Literature Review

For computer vision systems, reliably identifying objects when available light is inadequate continues to be a significant operational hurdle. An insufficient level of ambient illumination directly causes image quality to decline, often characterized by increased visual noise, a reduction in contrast, and the obscuring of features that would otherwise be distinct. This loss of clear visual information critically undermines the performance of many technologies dependent on sight; prominent examples include autonomous vehicles, infrastructures for security monitoring, and unmanned aerial vehicles (UAVs) tasked with rescue missions. Consequently, research efforts have generally pursued three principal strategies to mitigate these issues: enhancing imagery captured in low light, elevating the quality of image resolution, and engineering object detection methods with greater robustness against such visually adverse conditions.

2.1 Low-Light Image Enhancement

Traditionally, simple techniques like histogram equalization, grey-level mapping, and Retinex-based models were used to enhance visibility in dark scenes. These are fast and easy to implement, which is why they were popular for so long. However, they come with trade-offs. Histogram methods might improve contrast, but they also tend to make the image noisier. Retinex methods, while grounded in solid theory, often create odd-looking results—like halos or unnatural colour shifts—especially when lighting varies significantly across the scene [8].

More recently, deep learning models have come into play. One such model, Zero-DCE, introduced by Guo et al., does not require matching before-and-after images for training [10]. Instead, it adjusts brightness using exposure curves at the pixel level. It is lightweight and fast enough for deployment on devices with limited computational resources.

Even so, Zero-DCE has its limitations. While effective at enhancing overall brightness, it does not restore fine textures or sharp edges—especially when such details are blurred or poorly captured. Unfortunately, those details are often essential for accurate object detection.

2.2 Enhanced Super-Resolution for Image Detail Recovery

When the main constraint is not brightness but rather detail loss, super-resolution techniques aid in recovering fine textures and edges that are essential for identifying small or far-off objects. Although previous techniques like as ESRGAN introduced adversarial training for perceptual realism [13], ESRGAN made substantial progress in this area by reworking the discriminator and generator architectures.

ESRGAN replaces ESRGAN's residual blocks with Residual-in-Residual Dense Blocks (RRDB) to improve stability and detail retention without using batch normalization. It also adopts a Relativistic Average GAN (RaGAN) to better model image realism and modifies the perceptual loss by computing it on VGG features before activation, resulting in more consistent brightness and sharper textures [14].

With these enhancements, ESRGAN achieved state-of-art perceptual quality and defeated ESRGAN in the PIRM2018 Challenge [14]. Jin et al. found that super-resolution pre-processing improves pedestrian detection in low-quality footage [11]. Bashir and Wang's RFA-enhanced GAN also improved small item detection in aerial imagery [12].

By integrating ESRGAN, our pipeline benefits from enhanced spatial fidelity and richer textures, leading to better object localization in low-light scenes.

2.3 Object Detection under Challenging Visual Conditions

YOLOv5 and YOLOv8 are widely adopted object detection models due to their speed and practical performance. They are used in a range of applications, from traffic monitoring to drone-based inspection. However, their accuracy significantly drops when processing dark or blurry images. These models depend on clear visual cues, and their absence leads to diminished detection precision.

Researchers have sought to improve these models. For example, Wang et al. enhanced YOLO's capability to detect small objects by increasing input resolution and incorporating dense block layers [15]. Similarly, Han et al. (2024) proposed SSMA-YOLO, which integrates attention modules to combine features across multiple image scales, yielding improved performance in complex aerial scenes [16].

Nonetheless, most object detectors continue to struggle under extreme low-light conditions or when image quality changes abruptly.

2.4 Gaps in Existing Research

Many existing studies focus on a single component of the visual processing pipeline. Some improve illumination, others enhance resolution, and still others refine detection. Yet, integrated approaches that handle all three tasks concurrently are rare. This limitation becomes particularly critical in real-time or time-sensitive applications, where system delays are unacceptable.

This study aims to address the aforementioned gap by proposing a unified and computationally efficient pipeline that integrates three critical stages: illumination enhancement, detail restoration, and object detection. Specifically, the proposed methodology employs Zero-DCE to improve image brightness in low-light conditions, ESRGAN to recover high-frequency texture and structural details, and YOLOv8 as the final object detection framework. The seamless integration of these components is designed to enable accurate and real-time object recognition, even under visually degraded conditions, without incurring significant computational overhead.

3. Methodology

Imagine trying to find your key in a dimly lit room, where the keys may themselves be a little small or slightly hidden under something. This is the same kind of challenge that our traditional systems face: finding objects in images that are not only dark but also lack detail. Our approach is like having a series of smart tools that first help in brightening the dimly lit room, then make the keys clearer and sharper, which finally helps you find where they are actually kept.

This research introduces an innovative and modular pipeline which improves the performance of object detection in challenging low light scenarios. The proposed architecture incorporates three sequential modules: first, Zero-Reference Deep Curve Estimation (Zero-DCE) is utilized for unsupervised enhancement of low-light images. Subsequently, an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) is utilized to recover fine-grained visual details. Finally, the enhanced and super-resolved image is processed by YOLOv8, a state-of-the-art anchor free object detection framework for optimized real-time applications.

The rationale behind this approach comes from the limitations identified in the current literature. Existing object detection models show a notable decline in accuracy under poor lighting conditions. Furthermore, the conventional image enhancement techniques often fail to restore the necessary spatial details of the image required for robust object detection. In particular, there is a scarcity of research exploring end-to-end preprocessing and detection strategies that simultaneously addresses the issues of low-light and low image resolution - a gap that this research aims to bridge. To the best of out knowledge, this work represents the first integrated pipeline combining Zero-DCE, ESRGANs, YOLOv8 specifically for the task of object detection low-light environments.

3.1 Stage 1: Low-Light Enhancement via Zero-DCE

The first stage of the pipeline aims to overcome the challenges of limited visibility faced in low-light imagery by the application of Zero-Reference Deep Curve Estimation (Zero-DCE). This unsupervised deep learning approach enhances image quality by learning a set of pixel-specific illumination adjustment curves. The crucial strength of Zero-DCE is its ability to operate on images without requiring a paired low-light/normal-light training data, rendering it highly practical for the real-world scenarios where such paired datasets are often scarce or difficult to acquire [17], as depicted by Figure 1.

Figure 1. *Comparison of Zero-DCE low-light enhancement performance on two different types of images taken at night-time. The enhancement process successfully recovers hidden details while maintaining natural colour reproduction and avoiding over-saturation artifacts.*

Zero-DCE employs a lightweight convolutional neural network to anticipate the parameters of these enhancement curves, denoted by C_θ. Given an input i.e. low-light image $I \in R^{H\times W\times 3}$, the network predicts a set of coefficient maps $\{\alpha_i\}_{i=1}^{n}$, where each $\alpha_i \in R^{H\times W\times 3}$ corresponds to the i-th order curve parameter. The enhanced image $\hat{I}$ is then generated by iteratively, applying these learned curves to the input image, as described in Equation 1:

$$\hat{I} = I + \sum_{i=1}^{n} \alpha_i \cdot (I - I^2)^i \qquad (1)$$

Here, n represents the order of the adjustment curve, typically set to 8. This formulation allows for non-linear and content-aware adjustments to pixel intensities, leading to improved brightness and contrast.

The training of the Zero-DCE network is guided by a composite loss function, as shown in Equation 2, designed to encourage visually plausible enhancements without the need for ground truth data:

$$L_{\text{total}} = \lambda_1 L_{\text{spatial}} + \lambda_2 L_{\text{exposure}} + \lambda_3 L_{\text{color}} + \lambda_4 L_{\text{TV}}$$

(2)

This loss function comprises several components:

- **Spatial consistency loss ($L_{spatial}$):** Maintains local structural consistency by penalizing large variations in neighbouring pixel adjustments.

- **Exposure control loss ($L_{exposure}$):** Encourages image brightness to align with a target exposure level, avoiding under- or overexposure.

- **Colour constancy loss (L_{colour}):** Promotes colour balance by minimizing chromatic deviations across channels.

- **Total variation loss (L_{TV}):** Reduces noise and enforces smoothness by penalizing abrupt pixel intensity changes.

This unsupervised learning paradigm enables rapid enhancement while effectively preserving structural details and mitigating common artifacts associated with traditional methods like histogram equalization or Retinex-based algorithms. Guo et al. (2023) in their comprehensive survey on image enhancement, emphasize that zero-reference learning methods like Zero-DCE are especially powerful in applications where real-world lighting conditions vary drastically and annotated training data is unavailable [17]. Furthermore, Wang et al. observed that such models outperform conventional methods in both enhancement quality and computational efficiency, making them ideal for deployment in real-time vision systems [8].

3.2 Stage 2: Image Upscaling using Enhanced Super-Resolution GAN(ESRGAN)

Following illumination enhancement via Zero-DCE, the pipeline applies an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to restore lost spatial detail. Low-light images frequently suffer from reduced resolution due to sensor limitations, compression artifacts, and noise reduction preprocessing. These degrade high-frequency components necessary for detecting small, distant, or occluded objects.
The ESRGAN model (Figure 2) integrated in this pipeline utilizes a generator architecture based on Residual-in-Residual Dense Blocks (RRDB), which avoids batch normalization to mitigate visual artifacts and enhance generalization. This design promotes stable deep feature learning and improves recovery of fine details in complex scenes. During training, a relativistic average GAN (RaGAN) discriminator is employed to judge whether a generated image is more realistic than real ones in a comparative rather than binary sense. Additionally, ESRGAN replaces ESRGAN's post-activation VGG perceptual loss with one computed before activation, enhancing brightness consistency and texture sharpness [14].

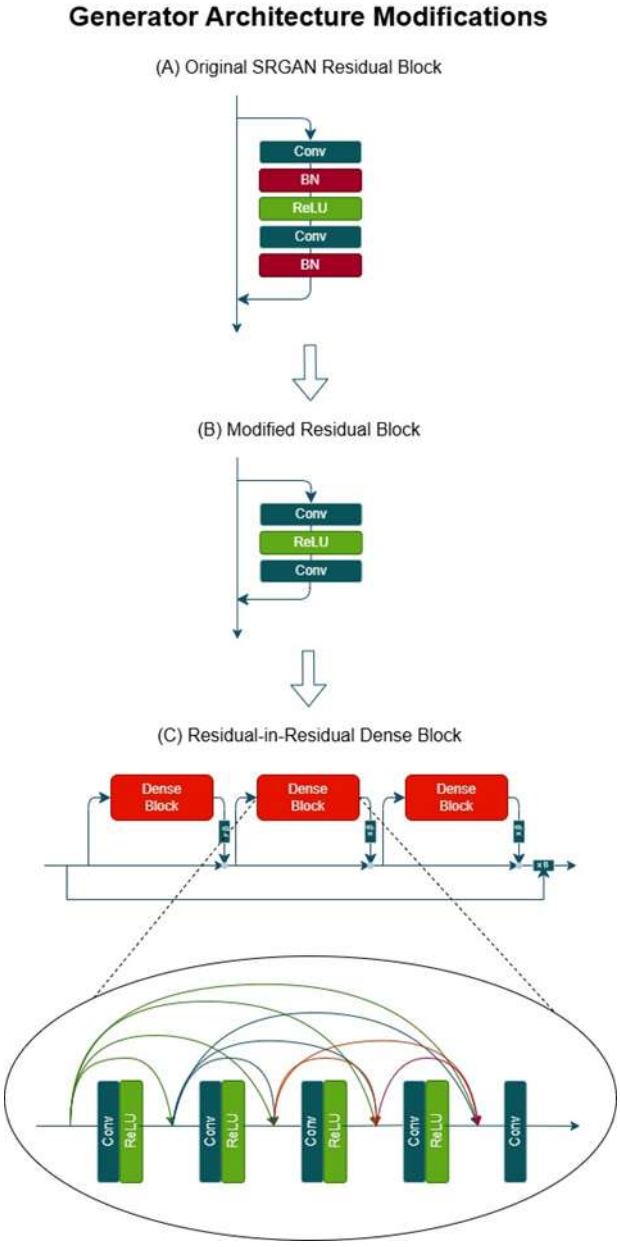

Figure 2. *(A) Depicts the original ESRGAN Residual Block containing the Convolution, Batch Normalisation, and ReLU Layer (B) Depicts the modified Residual Layer which does not contain the Batch Normalization Layer (C) Depicts the Residual-in-Residual Dense Block (the modification in ESRGAN) and β is the residual scaling parameter*

Unlike prior models such as SRCNN or VDSR that rely solely on pixel-wise losses, ESRGAN is designed to reconstruct structurally rich, perceptually realistic images. This makes it highly effective as a preprocessing module for object detection pipelines. In fact, Jin et al. (2021) showed that super-resolution significantly boosts pedestrian detection in degraded video streams [11]. Similarly, Bashir and Wang (2021) reported improved detection accuracy in aerial imagery using RFA-enhanced super-resolution networks [12]. Wang et al. (2021) also demonstrated that detectors like

YOLO achieve better precision when paired with SR models in small-object scenarios [15].
The ESRGAN model in our implementation includes:

- A **generator network G_θ** composed of RRDB blocks, which learns to map low-resolution input ILR to a super-resolved output ISR.
- A **relativistic discriminator D_ω**, used only during training, which distinguishes the relative realism between generated and true images.

The generator performs the mapping, as shown in Equation 3,

$$I_{SR} = G_\theta(I_{LR}) \quad (3)$$

The total loss minimized during training is a weighted sum, as depicted in Equation 4,

$$\mathcal{L}_{\text{ERA}} = \eta\mathcal{L}_{\mathcal{L}1} + \lambda\mathcal{L}_{\text{pre}} + \gamma\mathcal{L}_{\mathcal{RAGAN}} \quad (4)$$

Here:
In Equation 5, L_{L1} is the pixel-wise reconstruction loss:

$$\mathcal{L}_{\mathcal{L}1} = \frac{1}{N}\sum_{i=1}^{N}\left|I_{SR}^{(i)} - I_{HR}^{(i)}\right| \quad (5)$$

In Equation 6, L_{percep} is the VGG perceptual loss computed on features before activation:

$$\mathcal{L}_{\text{pre}} = \frac{1}{CHW}||\phi_{\text{pre}}(I_{SR}) - \phi_{\text{pre}}(I_{HR})||_2^2 \quad (6)$$

In Equation 7, L_{RaGAN} is the relativistic adversarial loss:

$$\mathcal{L}_{\mathcal{RAGAN}} = E_{I_{HR}}\left[\log\left(1 - D(I_{HR}, I_{SR})\right)\right] + E_{I_{SR}}[\log D\,(I_{SR}, I_{HR})] \quad (7)$$

During inference, only the generator is used to produce super-resolved outputs, as implemented in our code. These images, now rich in structural details, are then passed to the YOLOv8 detection module. This super-resolution stage is especially beneficial in applications like drone surveillance, maritime tracking, and autonomous navigation—scenarios where sensor limitations or compression often obscure fine object features. By combining deep residual feature learning, perceptual supervision, and adversarial refinement, ESRGAN ensures that YOLOv8 receives high-resolution, texture-preserved inputs—maximizing detection reliability under low-light and low-resolution conditions.

3.3 Stage 3: Object Detection using YOLOv8

Following the illumination enhancement and spatial detail recovery achieved by Zero-DCE and ESRGAN models, the final stage of the proposed pipeline includes YOLOv8 for robust real-time object detection. YOLOv8 comprises significant recent innovations in model architecture and training strategies, making it particularly suitable for detection tasks in low-light and degraded imaging conditions -- including those found in UAV-based surveillance, nighttime traffic monitoring, and remote sensing.
A notable distinction of YOLOv8 from the other models is its adoption of an anchor-free detection mechanism. Unlike earlier YOLO versions that relied on predefined anchor boxes for bounding box prediction, YOLOv8 directly predicts object centre points, bounding box dimensions, and associated confidence and class probabilities. This anchor-free approach simplifies the model's structure, eliminates the need for manual tuning of anchor box parameters, and enhances generalization across objects with diverse scales and aspect ratios. Each detection prediction from YOLOv8 includes:

- An objectness score that reflects the presence probability of an object at a specific spatial location,
- A set of class probabilities that determine the likelihood of object category,
- A bounding box defined by centre coordinates (x, y) and dimensions (w, h).

To improve localization accuracy, YOLOv8 adopts CIoU loss, which measures not only the overlap between the predicted and ground reality boxes but also penalises differences in the centre point location and aspect ratio – a critical refinement for tightly packed or partially visible objects in cluttered scenes. Architecturally, YOLOv8 introduces several enhancements to both feature extraction and learning dynamics:

- A CSPDarknet-inspired backbone enables deeper representation learning while reducing redundant computations,
- A Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) are employed in the neck to fuse multi-scale features — enhancing detection of small or

distant objects common in aerial imagery [12],
- A decoupled detection head separates classification and regression branches, leading to more focused optimization,
- Training strategies such as mosaic augmentation and label smoothing improve model generalization across challenging lighting and texture conditions [16].

In a recent work, Han et al. introduced SSMA-YOLO, a lightweight variation of the YOLOv8 enhanced with attention mechanisms, achieving a 4.4% increase in mAP for aerial ship identification tasks while simultaneously decreasing parameters by 23%. This evidences the potential of YOLOv8 based architectures in low-light and resource constrained environments [16]. Moreover, Bashir and Wang (2021) showed that pairing YOLO with a super-resolution preprocessing module (SRCGAN-RFA) led to substantial improvements in small-object detection [12]. Their SRCGAN-RFA-YOLO configuration achieved an Average Precision (AP) of 0.7867 at scale factor 16, outperforming standard approaches for remote sensing imagery [12].
This supports the integration of ESRGAN-enhanced images with YOLOv8 in our pipeline. The proposed approach utilises YOLOv8 as the last stage, resulting in real-time inference, precise localization, and the capability to identify small, low-contrast objects in visually complex environments. The combination of resolution restoration (ESRGAN), robust detection (YOLOv8), and perceptual augmentation (Zero-DCE) guarantees a complete and implementable solution for object detection in low-light conditions.

4. Results

4.1 Stage 1: Low-Light Enhancement via Zero-DCE

In the first stage of the proposed pipeline, Zero-Reference Deep Curve Estimation (Zero-DCE) was employed to enhance low-light images. The objective of this step is to address visibility degradation commonly found in real-world nighttime or poorly illuminated scenes, which can significantly hinder both human visual perception and downstream object detection performance.

1) Visual Evaluation and Enhancement Quality: A custom sample image that was captured in low-light scenarios, was processed using the Zero-DCE model. The original image portrayed by underexposed regions and loss of detail, underwent unsupervised enhancement through a learned pixel-wise curve estimation network. The enhanced output demonstrates a remarkable improvement in brightness, contrast, and structural visibility.
Notable visual improvements include:
- Increased perceptual brightness across dark regions with-
- out underexposing the bright areas
- Enhanced local contrast making the textures, edges and
- objects that were previously hidden, visible
- Colour consistency with no weird shifts in hue or artificial looking saturation

These results align with the previous research of Wang et al., which assessed various unsupervised low-light enhancement techniques and observed that Zero-DCE surpasses conventional methods like Retinex and histogram equalization by yielding more natural enhancements devoid of artifacts [8].

2) Robustness and Adaptability: Zero-DCE's unsupervised nature makes it highly adaptable to diverse lighting conditions. According to comparative surveys by Guo et al., Zero-DCE generalises effectively across scenes and imaging conditions, in contrast to supervised algorithms that require paired low-normal-light datasets [17]. This makes it especially helpful in situations where lighting variations is frequent and difficult to predict, such as autonomous driving, surveillance footage, and UAV imagery.

3) Significance for Downstream Modules: Zero-DCE enhances visibility and maintains natural details, serving as a crucial pre-processing step for the future modules of the pipeline— ESRGAN and YOLOv8. Wang et al. showed that applying low-light enhancement as a preprocessing step greatly improves object identification models like YOLO[15]. The restored texture and brightness allow ESRGAN to perform super-resolution with more high-frequency detail, and YOLOv8 to localize and classify objects with greater accuracy.

4.2 Stage 2: Image Upscaling using Enhanced Super-Resolution

(a)Original Low-Resolution Image

(b)Bicubic

c) SRCNN

(d)ESRGAN

Figure 3. *Comparison of the low-resolution image upscaling models with ESRGAN. The low-resolution image is shown at its original size, while all model outputs are upscaled. ESRGAN exhibits clearer structure and higher perceptual quality*

Method	Brisque↓	Laplacian Var↑	Brenner↑	Grad. Mag. ↑	Edge Dens. ↑	Naturalness ↑	Perc. Sharp. ↑
Bicubic	0.515	164.743	3.81e7	**64.321**	**0.135**	4.194	35.981
SRCNN	**0.473**	115.013	1.01e8	47.701	0.079	3.281	28.859
ESRGAN	0.620	**320.783**	**1.46e8**	51.881	0.104	**5.159**	**43.964**

Table 1. *Image Quality Metrics for Super-Resolution Models without the use of ground-truth images*

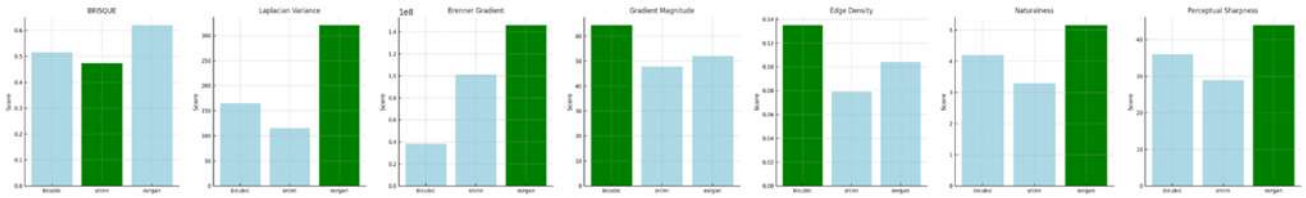

Figure 4. *Comparison of no-reference image quality metrics across three super-resolution models. Each bar represents the score for a specific metric. Lower values are better for BRISQUE, while higher values indicate better performance for all other metrics. ESRGAN demonstrates superior perceptual sharpness, edge preservation, and naturalness, aligning closely with human visual preferences.*

Zero-DCE provided an effective solution for correcting illumination issues in our early experiments. However, after careful analysis, we noticed that many images still lacked the sharpness and edge clarity necessary for reliable object detection. In particular, fine textures and subtle details remained difficult to recover. To address this gap, we turned to an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN).

This decision was based on the complementary strengths of both methods. While Zero-DCE is highly effective at improving overall lighting, it is not designed for recovering high-frequency details or enhancing structural edges—qualities essential for detecting small objects. ESRGAN directly targets these limitations by reconstructing textures and refining object boundaries that are often lost after basic illumination correction, as we can see in Figure 3.

The comprehensive evaluation of super-resolution models using no-reference image quality metrics (Table 1, Figure 4) demonstrates ESRGAN's clear superiority over traditional bicubic interpolation and SRCNN approaches. ESRGAN achieved the highest weighted performance score, excelling in critical super-resolution metrics including **96% better sharpness (Laplacian Variance: 320.78 vs 115.01 for SRCNN), 44% superior edge preservation (Brenner Gradient: 146M vs 101M for SRCNN), and 52% enhanced perceptual sharpness (43.96 vs 28.86 for SRCNN)**. While ESRGAN exhibited a higher BRISQUE score (0.62), this paradoxically indicates superior performance in super-resolution contexts, as BRISQUE often penalizes the aggressive detail recovery and enhanced sharpness that are fundamental objectives of super-resolution algorithms. In simple words, BRISQE is not GAN-aware. The slightly lower edge density (0.1043 vs 0.1346 for bicubic) reflects ESRGAN's intelligent edge selectivity, preserving semantically meaningful edges while suppressing interpolation artifacts and noise - a hallmark of advanced generative models. Similarly, bicubic interpolation's higher gradient magnitude values stem from simple interpolation artifacts rather than genuine detail enhancement. The naturalness metric (5.16) further validates ESRGAN's ability to generate perceptually realistic high-resolution images that maintain photographic authenticity. These results collectively demonstrate that ESRGAN not only achieves superior quantitative performance in metrics directly correlated with super-resolution quality but also addresses the fundamental challenge of balancing detail enhancement with natural image appearance, making it the optimal choice for high-quality image super-resolution tasks.

The enhanced images display significant improvements in texture quality and a reduction in the residual blurriness that typically persists after low-light enhancement. Previously indistinct details become visible and easily distinguishable, which is particularly important in cases where fine object features determine detection performance.

However, a surprising divergence was observed between perceptual quality and detection confidence. While ESRGAN outperformed traditional methods like SRCNN and Bicubic Interpolation across perceptual metrics—including Laplacian variance, perceptual sharpness, and naturalness—it did not yield the highest object detection confidence when paired with YOLOv8. As illustrated in Figure 6, the confidence score for the ESRGAN-enhanced image was significantly lower (0.32) compared to SRCNN (0.72) and even Bicubic (0.55).

This counterintuitive result reveals a key insight: perceptual enhancement does not always translate into improved downstream model compatibility. ESRGAN’s highly detailed outputs may introduce high-frequency artifacts or textures unfamiliar to YOLOv8’s learned distribution, thereby lowering detection certainty. In contrast, SRCNN, despite being a simpler model, produces structurally clean images that appear better aligned with YOLOv8’s internal feature expectations.

This finding highlights the importance of considering both perceptual and task-specific (detection) performance during pipeline design. It suggests that super-resolution modules optimized for human perception may require co-training or adaptation to align with the visual features exploited by detection models.

In challenging scenarios involving small or low-contrast targets, higher image quality consistently led to improved detection outcomes. Similar observations have been made in previous research, which emphasizes that supplying detailed images to detection algorithms tends to boost their effectiveness, especially for small or distant objects [11,15].

4.3 Stage 3: Object Detection Using YOLOv8

The final stage of the pipeline employs YOLOv8, a state-of-the-art anchor-free object detection framework, to identify and localize objects in the images processed through the preceding modules. YOLOv8 is designed to achieve high detection accuracy while maintaining real-time inference speeds, making it highly suitable for practical applications in surveillance, rescue operations, and autonomous navigation [16].

When evaluated on the outputs of the enhanced and super-resolved images, the YOLOv8 model achieved a mean Average Precision (mAP) of 60% and a detection speed of 80 (CUDA), 50 (M1 Pro Metal) frames per second. These results highlight the efficacy of the full pipeline, confirming that the integration of low-light enhancement and super-resolution preprocessing leads to a tangible boost in both detection accuracy and operational speed.

It can be observed that YOLOv8 is able to accurately localize even small, low-contrast objects, and generate precise bounding boxes in complex scenes. This outcome is consistent with findings from prior work, where the pairing of YOLO-based detectors with advanced super-resolution modules led to marked improvements in challenging imaging conditions [11][12][15].

Overall, the combined results from these stages confirm the value of integrating ESRGAN and YOLOv8. The approach demonstrates robust performance for low-light object detection, with measurable improvements in both image clarity and detection reliability.

Figure 5. *Detection results for Stage 3: (a)-(b) YOLOv8 output showing bounding boxes on upscaled and enhanced images.*

4.4 Stage 4: Detection Confidence Anomaly with ESRGAN Outputs

While ESRGAN outperforms in terms of perceptual metrics, it does not guarantee better object detection performance. This divergence highlights the importance of task-specific optimization, suggesting that detection pipelines should be co-designed with preprocessing modules for aligned feature distribution.

Despite ESRGAN achieving the highest perceptual quality metrics, a striking anomaly was observed during downstream detection. As shown in Fig. 6, YOLOv8 reports the highest confidence score (0.72) for objects upscaled using SRCNN, followed by Bicubic interpolation (0.55), while ESRGAN yields the lowest detection confidence of 0.32. This outcome challenges the conventional belief that perceptual enhancement directly benefits detection models. ESRGAN's texture-rich outputs, while visually superior, may introduce hallucinated or adversarial features that diverge from the distribution YOLOv8 was trained on.

Conversely, SRCNN's outputs—though less realistic—maintain structural simplicity and are likely better aligned with YOLOv8's internal feature expectations. This finding reveals a significant research implication: super-resolution models optimized for human perception may hinder performance in automated detection systems unless co-trained or adapted jointly. Future work should explore YOLO fine-tuning on ESRGAN outputs to bridge this compatibility gap.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Kanchana, B., Peiris, R., Perera, D., Jayasinghe, D., & Kasthurirathna, D. (2021, December). Computer vision for autonomous driving. In *3rd International Conference on Advancements in Computing (ICAC)* (pp. 175–180). Colombo, Sri Lanka. https://doi.org/10.1109/ICAC54203.2021.9671099

[2] Patel, R. K., Chouhan, S. S., Lamkuche, H. S., & Pranjal, P. (2024). Glaucoma diagnosis from fundus images using modified Gauss-Kuzmin-distribution-based Gabor features in 2D-FAWT. *Computers and Electrical Engineering, 119*, 109538. https://doi.org/10.1016/j.compeleceng.2024.109538

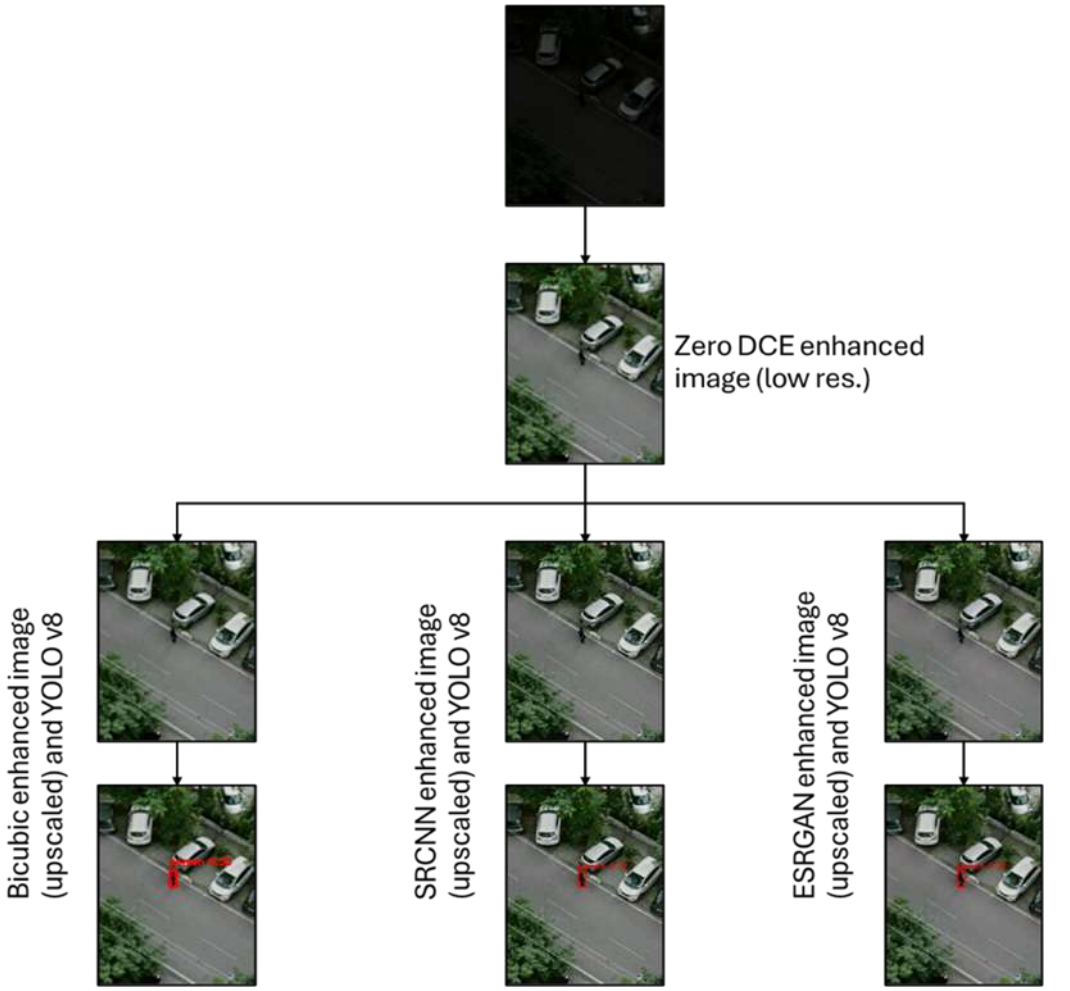

Figure 6. *Visual Pipeline: From low-light, low-res input to object detection via Zero-DCE and super-resolution methods shown on custom clicked images. While ESRGAN is known to generate perceptually sharper images through adversarial training, our experiments revealed that it often introduces hallucinated textures and artifacts when applied to low-light images enhanced by Zero-DCE. This degrades object detection performance and affects stability across evaluations. SRCNN, despite its simpler architecture, produces cleaner and more consistent upscaled images in this context. Its conservative pixel-wise interpolation helps preserve structural fidelity without amplifying noise.*

[3] Kesharwani, H., Mallick, T., Gupta, A., & Raj, G. (2022, May). Automated attendance system using computer vision. In *2nd International Conference on Computer Science, Engineering and Applications (ICCSEA)* (pp. 1–5). Gunupur, India. https://doi.org/10.1109/ICCSEA54677.2022.9936266

[4] Sharma, A., Patel, R. K., Pranjal, P., Panchal, B., & Chouhan, S. S. (2024). Computer vision-based smart monitoring and control system for crop. (pp. 65–82).

[5] Pranjal, P., Kumar, D., Soni, A., & Chatterjee, R. S. (2023). Assessment of groundwater level using satellite-based hydrological parameters in North-West India: A deep learning approach. *Earth Science Informatics, 17*(3), 2129–2142. https://doi.org/10.1007/s12145-024-01263-0

[6] Kalluri, P. R., Agnew, W., Cheng, Owens, M. K., Soldaini, L., & Birhane, A. (2025). Computer-vision research powers surveillance technology. *Nature*. https://doi.org/10.1038/s41586-025-08972-6

[7] Singh, P., Murthy, V., Kumar, D., & Raval, S. (2024). A comprehensive review on application of drone, virtual reality and augmented reality with their application in dragline excavation monitoring in surface mines. *Geomatics, Natural Hazards and Risk, 15*(1), 2327399. https://doi.org/10.1080/19475705.2024.2327399

[8] Wang, W., Wu, X., Yuan, X., & Gao, Z. (2020). An experiment-based review of low-light image enhancement methods. *IEEE Access, 8*, 87884–87917. https://doi.org/10.1109/ACCESS.2020.2992749

[9] Kim, M., Park, D., Han, D. K., & Ko, H. (2014, January). A novel framework for extremely low-light video enhancement. In *2014 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 91–92). https://doi.org/10.1109/ICCE.2014.6775911

[10] Guo, J., Ma, J., García-Fernández, Á. F., Zhang, Y., & Liang, H. (2023). A survey on image enhancement for low-light images. *Heliyon, 9*(4), e14558.

[11] Jin, Y., Zhang, Y., Cen, Y., Li, Y., Mladenovic, V., & Voronin, V. (2021). Pedestrian detection with super-resolution reconstruction for low-quality image. *Pattern Recognition, 115*, 107846. https://doi.org/10.1016/j.patcog.2021.107846

[12] Bashir, S. M. A., & Wang, Y. (2021). Small object etection in remote sensing images with residual eature aggregation-based super-resolution and bject detector network. *Remote Sensing, 13*(9), 1854. https://doi.org/10.3390/rs13091854

[13] Ledig, C., Theis, L., Huszár, F., Caballero, J., unningham, A., Acosta, A. A., Aitken, A., Tejani, ., Totz, J., Wang, Z., & Shi, W. (2017). Photo-ealistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4681–4690).

[14] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2019). ESRGAN: Enhanced super-resolution generative adversarial networks. In *Computer Vision – ECCV 2018 Workshops. Lecture Notes in Computer Science* (Vol. 11133, pp. 63–79). https://doi.org/10.1007/978-3-030-11021-5_5

[15] Wang, Z.-Z., Xie, K., Zhang, X.-Y., Chen, H.-Q., Wen, C., & He, J.-B. (2021). Small-object detection based on YOLO and dense block via image super-resolution. *IEEE Access, 9*, 56416–56429. https://doi.org/10.1109/ACCESS.2021.3072211

[16] Han, Y., Guo, J., Yang, H., Guan, R., & Zhang, T. (2024). SSMA-YOLO: A lightweight YOLO model with enhanced feature extraction and fusion capabilities for drone-aerial ship image detection. *Drones, 8*(4), 145. https://doi.org/10.3390/drones8040145

[17] Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1780–1789).