

DOI: <https://doi.org/10.15276/hait.08.2025.10>

UDC 004.932

Adaptive super-resolution integration to enhance object detection on low-quality unmanned aerial vehicle imagery

Maksym Yu. Holenko

ORCID: <https://orcid.org/0000-0001-8124-7299>; holenko.maksym@gmail.com
Zhytomyr Polytechnic State University, 103, Chudnivska St. Zhytomyr, 10005, Ukraine

ABSTRACT

The article addresses the problem of improving the accuracy of object detection in images captured by unmanned aerial vehicles under conditions of reduced spatial resolution and the presence of noise artifacts. The relevance of this research is driven by the practical need to maintain the reliability of computer vision systems in challenging field environments, where conventional detection algorithms tend to lose effectiveness. The aim of the study is to enhance the robustness of object detection in low-quality unmanned aerial vehicles imagery through the development of an adaptive preprocessing mechanism based on deep neural network-driven image super-resolution. The proposed approach involves the dynamic activation of the super-resolution module only in cases where image quality or detector confidence is insufficient. Within the framework of the study, a combination of the high-accuracy two-stage model Faster R-CNN and prior image upscaling using Real-ESRGAN is employed. An adaptive logic for triggering the image enhancement module is introduced, which is activated solely when the detector's confidence level falls below a defined threshold, thereby reducing computational overhead without compromising recognition performance. An experimental evaluation of the proposed method was conducted using unmanned aerial vehicles imagery degraded by various distortions, including blur, noise, and compression artifacts. The results demonstrate consistent improvements in detection accuracy across all tested image degradation types while maintaining acceptable processing time. The practical value of this research lies in its applicability to autonomous monitoring systems, search-and-rescue missions, and situational analysis tasks based on unmanned aerial vehicles video streams. The proposed approach opens up opportunities for further optimization by incorporating additional components, such as lightweight preliminary object filtering modules.

Keywords: Object detection; artificial intelligence; unmanned aerial vehicles; Faster R-CNN; Real-ESRGAN; deep learning; computer vision

For citation: Holenko M. Yu. "Adaptive super-resolution integration to enhance object detection on low-quality unmanned aerial vehicle imagery". *Herald of Advanced Information Technology*. 2025; Vol.8 No.2: 164–178. DOI: <https://doi.org/10.15276/hait.08.2025.10>

INTRODUCTION

Unmanned aerial vehicles (UAVs) are increasingly employed for surveillance, monitoring, mapping, security, and search-and-rescue operations. At the core of such autonomous systems lie computer vision algorithms that enable real-time object detection and identification in captured imagery. Modern deep learning-based detectors, such as Faster R-CNN, have achieved high accuracy in object recognition tasks under favorable conditions [1], [2].

However, in the context of aerial imaging from drones, detection becomes significantly more challenging due to several factors: small object sizes, unusual viewing angles, camera motion, variable lighting conditions [3], and image degradations such as compression artifacts, noise, or blur. These issues result in the loss of critical visual features upon which detectors rely for accurate performance. As a consequence, localization accuracy deteriorates, and the number of missed or misclassified objects increases, particularly in low-signal scenarios or complex scenes.

To overcome these limitations, recent research has focused on image preprocessing methods, particularly super-resolution (SR) techniques aimed at enhancing spatial resolution. The goal of super-resolution is to reconstruct high-quality images while preserving critical texture and edge-related features. Among contemporary super-resolution solutions, the Real-ESRGAN model has attracted considerable attention due to its ability to reliably restore fine details [4], even in images affected by severe visual artifacts. However, the constant (fixed) application of super-resolution to every frame leads to a substantial increase in computational cost [5].

This poses challenges for unmanned aerial vehicles-based systems, where hardware resources are limited and response time is critical. Therefore, this study proposes an adaptive approach that integrates Real-ESRGAN into the object detection pipeline only when image degradation is detected or when the detector exhibits low confidence. If the input image is of sufficient quality, the super-resolution stage is bypassed. This strategy allows for maintaining high detection accuracy while preserving computational efficiency and ensuring real-time system performance.

© Holenko M., 2025

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/deed.uk>)

The proposed solution goes beyond classical approaches by introducing an architectural concept of adaptive super-resolution that dynamically adjusts to the frame context and the results of preliminary detection. This opens up new possibilities for flexible, intelligent processing of visual streams in real-time environments.

1. ANALYSIS OF LITERARY DATA

Image super-resolution methods have demonstrated a significant impact on improving object recognition performance, particularly under degraded input conditions. Studies [5], [6], [7], [8] have conducted a systematic analysis of the influence of super-resolution-based preprocessing on object detection outcomes when working with low-quality satellite or static imagery. Convolutional neural network (CNN)-based models, such as very-deep super-resolution and robust feature super-resolution, have shown improvements in mean average precision (mAP) ranging from 7 % to 36 %, depending on the complexity of the scene and the initial image resolution.

Specifically, study [7] reported that super-resolution preprocessing substantially enhances the detection of objects as small as 20×20 pixels, which is particularly important for high-altitude monitoring tasks. In [5], very-deep super-resolution was employed as a fixed preprocessing stage prior to Faster R-CNN for satellite imagery, resulting in an accuracy gain of 8-12 % for vehicle classes. However, the analysis also revealed a processing time increase exceeding 50 %, making the approach unsuitable for real-time deployment.

Further advancement of super-resolution technologies is associated with the adoption of generative adversarial networks (GANs). Models such as ESRGAN and Real-ESRGAN, introduced in [9, 10], offer not only improved visual quality of images but also the preservation of texture and edge information—critical for feature extraction. In study [10], it was reported that applying Real-ESRGAN to images from the VisDrone dataset prior to processing with YOLOv7 resulted in a 3-5 % increase in detection accuracy for humans and vehicles, as well as a reduction in false-negative rates.

A review presented in study [12] indicates that Real-ESRGAN demonstrates the highest robustness when processing nighttime videos, where images are affected by mixed degradations such as noise, defocus, and compression artifacts. Under these conditions, super-resolution preprocessing proved particularly effective for detecting small-sized objects.

Another direction of development involves optimizing super-resolution not only for classical

image quality metrics (peak signal-to-noise ratio, structural similarity index measure), but also for recognition performance. In [11], a task-driven super-resolution approach was proposed—architecture for joint training of the super-resolution module and Faster R-CNN, where the super-resolution block is updated based on a loss function that incorporates detection outcomes. This enables the super-resolution process to focus on reconstructing features that are critical for classification and localization. As a result, an improvement of up to 6 % in mAP was achieved on the COCO and VisDrone datasets.

However, even in this case, super-resolution is applied to all images regardless of input quality or model confidence, which leads to excessive computational load and limits applicability in resource-constrained systems.

A review of studies [6], [9], [11] indicates that constant application of super-resolution is inefficient when image quality is already sufficient. Moreover, in some cases, super-resolution may even degrade classification accuracy. GAN-based models designed for visual enhancement may alter local features used by the detector. This is particularly problematic for small or low-contrast objects, where shape consistency is critical, and the loss of certain local cues can render the object undetectable.

Moreover, fixed application of super-resolution in real-time scenarios requires substantial computational and energy resources, rendering this approach less suitable for mobile platforms such as unmanned aerial vehicles.

In response to these challenges, a transition toward adaptive super-resolution usage appears justified. The key idea is to activate the super-resolution module only when it is genuinely necessary—for example, when the object detector yields low confidence scores, or when degradation in image quality is identified through specific metrics. Such metrics may include the variance of the Laplacian to assess blur or entropy-based measures to detect noise.

Accordingly, super-resolution is applied solely to frames or regions that require enhancement, thereby reducing resource consumption in cases where resolution upscaling is unnecessary. This strategy avoids redundant processing, decreases system load, and improves throughput and stability—especially in streaming video processing tasks onboard unmanned aerial vehicles.

Despite the progress in integrating super-resolution into object detection tasks, none of the reviewed studies [5], [6], [7], [8], [9], [10], [11], [12], [13] implements super-resolution in an adaptive manner that activates only under specific

conditions. Existing solutions do not account for scene dynamics, current frame quality, or model confidence. This results in limited scalability and questions the practicality of constant super-resolution usage in real-world scenarios.

In this study, we propose a novel architecture featuring adaptive super-resolution integration, where the super-resolution module is triggered only under clearly defined conditions: low detection confidence, degraded feature clarity, or image blurriness. This enables the system to adjust the processing pipeline based on the capture conditions and the current state of the detector, maintaining a balance between detection accuracy and computational efficiency. Such an approach is particularly well-suited for real-time applications.

2. THE PURPOSE AND OBJECTIVES OF THE RESEARCH

The objective of this study is to develop and validate an approach for improving object detection accuracy on low-quality images through the adaptive integration of super-resolution (Real-ESRGAN) into the deep neural network pipeline of Faster R-CNN. A distinctive feature of the proposed method is the implementation of a conditional super-resolution module activation mechanism, which is triggered based on the confidence level of the initial detection or indicators of input image degradation.

The research is focused on automated object recognition tasks using unmanned aerial vehicles imagery, where varying image quality and limited computational resources pose significant challenges for conventional pipelines. This work involves the development of a prototype adaptive system, its experimental evaluation, and a comparative analysis with traditional methods in which super-resolution is applied as a fixed preprocessing stage.

To achieve the stated objective, the following tasks were formulated.

1. Develop a method for adaptive integration of Real-ESRGAN into the Faster R-CNN pipeline, taking into account detector confidence metrics and input image quality.

2. Design an experimental architecture incorporating an super-resolution module that makes activation decisions based on predefined threshold criteria depending on operating conditions.

3. Conduct system evaluation on real-world unmanned aerial vehicles imagery in three configurations: without super-resolution, with fixed super-resolution, and with adaptive super-resolution.

4. Perform quantitative assessment of detection accuracy and system performance (in terms of mAP, precision, recall, and FPS), considering input data quality and scene complexity.

5. Analyze the results from the perspective of deployment in resource-constrained systems and assess the feasibility of adaptive super-resolution processing under varying image quality conditions.

3. RESEARCH METHODS

The study was conducted using the VisDrone2019-DET dataset, which contains high-altitude urban images [14] captured by unmanned aerial vehicles. This dataset includes over 10,000 annotated images with labels for key object categories such as pedestrians, cars, bicycles, buses, and trucks, enabling realistic modeling of unmanned aerial vehicles operation scenarios. For the experiment, a representative subset of 500 images was selected, accounting for varying levels of resolution, compression artifacts, noise, and changes in viewing angles.

As the baseline detection architecture, Faster R-CNN with a ResNet-50 backbone was employed, implemented using PyTorch 2.0 [15]. For image enhancement, a pre-trained Real-ESRGAN super-resolution module (x4plus version) was used, capable of reconstructing high-resolution images [16] while preserving textures and edges, even in the presence of noise and JPEG artifacts. Experiments were carried out on the Google Colab Pro platform with a Tesla T4 GPU, providing a performance evaluation under conditions approximating resource-constrained systems.

Within the scope of this study, three image processing approaches were implemented:

1. Without super-resolution – baseline Faster R-CNN detection;
2. Fixed super-resolution – all images are passed through Real-ESRGAN prior to detection;
3. Adaptive super-resolution – the super-resolution module is activated only under conditions of low confidence or poor input image quality.

The adaptive logic is based on image filtering using the following criteria:

- Object detection confidence score (global threshold of 0.5);
- Image sharpness assessment via Laplacian variance [17];
- Detection of JPEG artifacts using block-wise entropy analysis.

If the image does not meet the quality threshold or if the object class confidence score falls below the defined threshold, the frame is additionally processed by the super-resolution module and then passed back to Faster R-CNN.

The evaluation of results was based on the following metrics:

- mAP (mean Average Precision);
- Recall and Precision;

- FPS (frames per second);
- Per-image processing time.

The methodological part of the study includes both the architectural integration of the adaptive super-resolution module and the design of an experimental framework for comparing detection accuracy and computational efficiency across different processing modes. This approach enables a comprehensive assessment of the impact of super-resolution on object detection performance under conditions of low image quality and constrained computational resources [18].

4. SYSTEM ARCHITECTURE

The proposed system is designed as an integrated computer vision pipeline that combines a high-precision deep object detection model (Faster R-CNN) with an image super-resolution module (Real-ESRGAN). Its main distinguishing feature is the implementation of adaptive logic for super-resolution module activation, which enables or disables the computationally intensive enhancement process depending on the input image quality or the detector's confidence level.

Overall architecture:

1. Image quality assessment

$q(x)$ is a metric employed to evaluate the quality or sharpness of the image, such as Laplacian variance, BRISQUE, or other no-reference indicators [19], [20]. Low values indicate the presence of blur, noise, or JPEG compression artifacts, all of which may negatively affect the accuracy of the detector. In our implementation, the $q(x)$ was computed using two complementary no-reference metrics: the BRISQUE estimator, based on deviations from natural scene statistics, and the Laplacian variance method.

In this case, the quality score is defined as:

$$q(x) = \text{Var}(\Delta x) \quad (1)$$

where: Δx denotes the Laplacian response of the image and $\text{Var}()$ is the pixel-wise variance. Both metrics produce scalar scores reflecting the degree of blur or degradation and are compared against a predefined quality threshold τ_q .

Initial confidence assessment $c(x)$ – involves an initial forward pass of the image through the detector without super-resolution processing, followed by analysis of the confidence score. The resulting confidence score $c(x)$ reflects the detector's certainty about the class and bounding box predictions. A value below a predefined threshold τ_c indicates that the system is uncertain about its prediction. The confidence score $c(x)$ is derived from the output of the initial forward pass of the Faster R-CNN detector. Given a set of predicted

objects with associated confidence scores $\{p_1, p_2, \dots, p_n\}$.

The overall image-level confidence is computed as:

$$c(x) = \max\{p_i\}_{i=1}^n \quad (2)$$

where $p_i \in [0,1]$ denotes the confidence of the i -th detected object. This value corresponds to the detector's certainty in its most confident prediction and is used for comparison with a predefined threshold τ_c .

2. The decision function $f_{SR}(x)$ evaluates whether the input image should be processed by the super-resolution module. If either the quality metric or the confidence score falls below a predefined threshold, Real-ESRGAN is activated.

3. Decision-making mechanism for super-resolution activation – implemented as conditional logic:

$$f_{SR}(x) = \begin{cases} \text{Real-ESRGAN}(x), & \text{if } q(x) < \tau_q \text{ or } c(x) < \tau_c \\ x, & \text{else} \end{cases} \quad (3)$$

where: τ_q is the quality metric threshold; τ_c is the confidence score threshold for object classes; $f_{SR}(x)$ is the conditional super-resolution-processing function that returns either the enhanced image or the original input.

4. Final detection stage – the image is processed by the Faster R-CNN detector regardless of whether super-resolution was applied. This ensures processing consistency and a unified decision point.

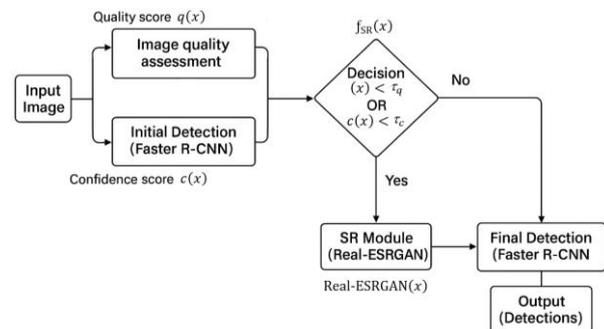


Fig. 1. Decision logic for adaptive super-resolution activation

Source: compiled by the author

Fig. 1 illustrates the key components of the proposed pipeline, including image quality and confidence evaluation, decision logic based on thresholds τ_q and τ_c , conditional activation of the Real-ESRGAN module, and final object detection. The decision node reflects the logic of the function $f_{SR}(x)$ defined in formula (3).

The system operates based on conditional branching: if the input frame exhibits signs of degradation or the detector lacks confidence in its predictions, Real-ESRGAN is activated as a means

of enhancing features critical for object detection. Otherwise, the system bypasses the super-resolution stage, conserving resources without sacrificing accuracy.

To reduce latency under high computational loads, an asynchronous implementation of the super-resolution module is supported, whereby super-resolution processing can be executed in a separate thread with frame buffering. In this mode, input images requiring enhancement are preprocessed by the super-resolution module in the background, while detection is performed either after completion of the super-resolution process or concurrently on other frames.

This approach enables: reduced real-time latency; balanced GPU workload distribution; avoidance of detector idle time during processing of super-resolution-intensive frames; and scalability of the system to multithreaded or multi-GPU environments.

5. SYSTEM IMPLEMENTATION

The proposed system was implemented in Python using the PyTorch 2.0 framework, providing flexibility, GPU acceleration support, and efficient deployment of deep learning models. The architecture is organized as a modular pipeline, where each processing stage is represented by a distinct functional block with clearly defined input and output parameters. This structure facilitates easy scalability, adaptation to new datasets or models, and the implementation of complex interaction logic between components.

The pipeline includes the following main modules:

1. ImageLoader – responsible for loading and preprocessing images, including resizing, normalization, and formatting to the required input specifications.

2. QualityEvaluator – implements image quality assessment using no-reference metrics: Laplacian variance (blur estimation), JPEG block entropy (compression artifact evaluation), and BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator), which detects general distortions without a reference image [21].

3. Detector – serves as the inference module of the Faster R-CNN model with a ResNet-50 backbone, implemented via `torchvision.models.detection`. This block performs initial and final detection depending on the scenario.

4. ConfidenceAnalyzer – module for analyzing classification confidence levels, calculating the maximum confidence score among detected objects in the image.

5. DecisionLogic – the core component implementing adaptive decision logic for super-resolution activation. Conditions include exceeding blur or entropy thresholds, as well as insufficient classification confidence. Based on this module's output, a decision is made on the necessity of super-resolution.

6. SRModule – implemented as a wrapper around the pre-trained Real-ESRGAN model (x4plus version) [22]. It handles image processing in float32 format, automatic CUDA activation if GPU is available, preprocessing normalization, and inverse normalization post upscaling.

7. ResultsAnalyzer – computes key metrics (mAP@0.5, mAP@[0.5:0.95], IoU, Precision, Recall, FPS), generates plots, bounding box visualizations, and saves inference results in JSON format for further analysis.

Thanks to its clear modularity, the implementation supports the replacement or update of any component without the need to modify the entire system. This approach also facilitates straightforward extension to tasks such as classification, segmentation, or video tracking, demonstrating the versatility of the proposed architecture.

Experiments were conducted in the Google Colab Pro environment using an NVIDIA Tesla T4 GPU, providing a convenient platform for testing and rapid prototyping of the model within the scope of this study.

6. EXPERIMENTAL CONFIGURATIONS

Within the scope of this study, three experimental configurations for processing input images were implemented, differing in the logic of integrating the super-resolution (SR) module into the overall object detection pipeline. Each configuration allows investigation of different aspects of super-resolution's impact on recognition accuracy and system performance.

Baseline (No super-resolution) – this is the basic scenario where no super-resolution processing is applied. All input images are fed directly to the Faster R-CNN model for detection without any modifications. This variant serves as a reference point reflecting the system's baseline performance without any form of input enhancement.

Fixed super-resolution – in this configuration, all input images undergo super-resolution preprocessing regardless of their quality or the detector's confidence level. The super-resolution-enhanced images are then passed to Faster R-CNN for subsequent inference. This approach corresponds to common methods found in the literature where super-resolution is applied as a mandatory stage

throughout the entire pipeline. The main objective is to assess the average improvement in detection results due to super-resolution, without any adaptation to scene context.

Adaptive super-resolution – the proposed configuration implements dynamic super-resolution activation based on an analysis of the input image quality or the detector’s confidence scores. If a frame exhibits signs of blur (e.g., low Laplacian variance or high JPEG block entropy), or if the detector produces predictions with low confidence scores, the super-resolution module is activated. Otherwise, the super-resolution stage is bypassed. This approach reduces computational overhead while maintaining accuracy in critical cases.

In the Fixed super-resolution and Adaptive super-resolution configurations, three different image upscaling approaches were implemented.

Each method was tested independently to analyze the impact of the super-resolution model choice on detection results:

1. Bicubic interpolation – a classical, non-learned image upscaling technique. It offers high processing speed but limited capability in restoring fine details.

2. ESRGAN – an early generative adversarial network (GAN)-based super-resolution model. It provides better visual quality compared to Bicubic interpolation but tends to produce artifacts, especially in regions with fine textures.

3. Real-ESRGAN – an improved version specifically adapted for real-world images, which may contain noise, JPEG degradation, and blur. It offers high robustness against artifacts and preserves critical object contours. This method was chosen as the primary approach for implementing adaptive super-resolution in the proposed system.

7. DATASET AND PREPROCESSING

During the experimental study, a subset of 1,000 images was selected from the full VisDrone2019-DET dataset to represent realistic unmanned aerial vehicles imaging conditions. The selection was based on diversity in viewing angles (vertical, oblique, and frontal projections), scene types (urban intersections, transport hubs, pedestrian clusters, architectural structures), and the presence of key target object classes such as pedestrians, passenger cars, buses, trucks, and bicycles. Particular attention was given to images containing small or partially occluded objects, which suffer most from detail loss. This approach to forming the test subset allowed modeling a wide range of practical scenarios encountered by unmanned systems and ensured representativeness for

evaluating the effectiveness and adaptive logic of the proposed method.

Artificial degradation was deliberately applied to the VisDrone images by introducing blur, noise, and JPEG compression artifacts [23]. This enabled simulation of unstable unmanned aerial vehicles image conditions.

Fig. 2 illustrates an example of image degradation: on the left is a fragment of an original image from the VisDrone dataset, while on the right is the same fragment after applying artificial blur and JPEG compression. A significant loss of edge and texture information is evident, which is critical for the detector.

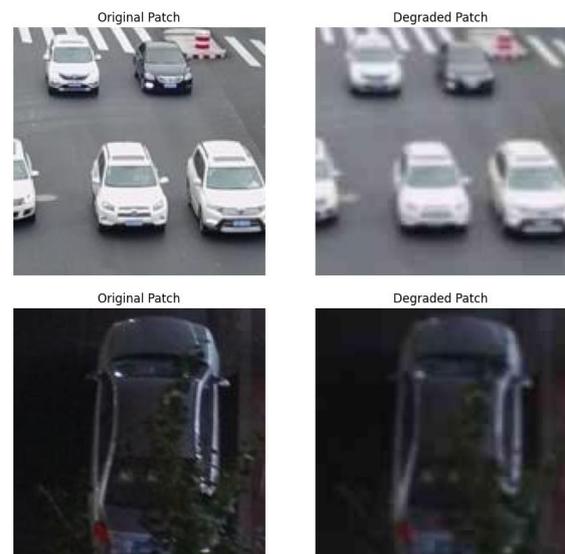


Fig. 2. Zoom-in on original and degraded image patches

Source: compiled by the author

Fig. 3 shows the result of restoring the degraded image using Real-ESRGAN. After processing by the super-resolution module, it is noticeable that certain details visually surpass those of the original image. Edges, structure, and fine details have been restored, potentially improving feature extraction for subsequent detection. However, simultaneously, some features are lost due to the reconstruction being performed on an image of excessively low quality.

8. EVALUATION METRICS

To evaluate the performance of the detection system with different super-resolution processing configurations, a comprehensive set of metrics was employed, encompassing both object localization accuracy and computational efficiency.

The primary spatial metric used is Intersection over Union (IoU), which measures the overlap between the predicted and ground truth bounding boxes.



Fig. 3. Example of super-resolution enhancement using Real-ESRGAN

Source: compiled by the author

It is defined as the ratio of the area of their intersection to the area of their union [24]:

$$\text{IoU} = \frac{B_{\text{pred}} \cap B_{\text{gt}}}{B_{\text{pred}} \cup B_{\text{gt}}} \quad (4)$$

where B_{pred} is the predicted bounding box, and B_{gt} is the ground truth bounding box.

Based on IoU, mean Average Precision (mAP) is calculated—this is a key metric for evaluating detector performance across multiple classes. For each class, a Precision–Recall curve is constructed, illustrating how precision varies with recall as the confidence threshold changes [25]. Depending on the IoU threshold required for a detection to be considered correct, the area under this curve (Average Precision, AP) is computed.

$\text{mAP}@0.5$ is defined as the mean of the AP values over all classes, where detection is deemed correct if $\text{IoU} \geq 0.5$:

$$\text{mAP}@0.5 = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (5)$$

where N is the number of classes, and AP_i is the average precision for class i .

For a more comprehensive evaluation, $\text{mAP}@[0.5:0.95]$ is used, which calculates the mean average precision over IoU thresholds ranging from 0.5 to 0.95 in increments of 0.05 [25], [26].

This metric assesses not only the detector's ability to find objects but also the precision of their localization:

$$\text{mAP}@[0.5:0.95] = \frac{1}{10} \sum_{k=0}^9 \left(\frac{1}{N} \sum_{i=1}^N \text{AP}_i(\text{IoU}_k) \right) \quad , \quad (6)$$

where: N is the number of classes, $k \in [0,9]$ index of the IoU threshold, $\text{IoU}_k = 0,5 + 0,05 \times k$ – value of the overlap threshold, and $\text{AP}_i(\text{IoU}_k)$ average precision for class i at the specified IoU threshold.

Complementing mAP are the classical metrics Precision and Recall. Precision indicates the proportion of all predicted positive detections that are correct, while Recall represents the fraction of all actual objects that were successfully detected. Together, these metrics allow for analyzing the trade-off between false positives and false negatives, which is especially important in tasks with class imbalance.

System performance is evaluated using metrics such as FPS (frames per second)—the number of processed frames per second—and the average processing time per frame [26].

9. EXPERIMENT RESULTS

After a detailed description of the system architecture, implementation methods, and image processing scenarios, we proceed to present the obtained results. The evaluation begins with a comparison of different super-resolution methods in terms of their impact on object detection accuracy.

To assess the effectiveness of various super-resolution approaches for object detection, three graphs were constructed, enabling comparison of recognition quality and computational performance of the models.

Fig. 4 presents a comparative chart of $\text{mAP}@0.5$ and $\text{mAP}@[0.5:0.95]$ values for three super-resolution methods: Bicubic, ESRGAN, and Real-ESRGAN. The notation follows the standard COCO evaluation format, where $\text{mAP}@0.5$ refers to mean Average Precision at a fixed IoU threshold of 0.5, and $\text{mAP}@[0.5:0.95]$ denotes the average AP over IoU thresholds from 0.5 to 0.95 with a step of 0.05. It is evident that Real-ESRGAN demonstrates the highest performance, achieving $\text{mAP}@0.5 = 0.456$ and $\text{mAP}@[0.5:0.95] = 0.295$. ESRGAN shows slightly lower results (0.441 / 0.284), while Bicubic lags behind significantly with $\text{mAP}@0.5 = 0.421$ and $\text{mAP}@[0.5:0.95] = 0.267$. This indicates the advantage of a trained generative model over classical interpolation methods in the context of improving input data for the detector.

Fig. 5 presents a comparison of average frame processing speed (FPS) across the three super-resolution methods. FPS indicates the number of images processed per second during the upscaling stage. Bicubic achieves the highest performance (≈ 18.5 FPS), attributed to its low computational complexity. ESRGAN operates noticeably slower (≈ 10.2 FPS), while Real-ESRGAN is the most resource-intensive option (≈ 8.4 FPS). Thus,

although Real-ESRGAN delivers the best accuracy results, it requires approximately twice the processing time compared to Bicubic, which is critical in real-world resource-constrained systems.

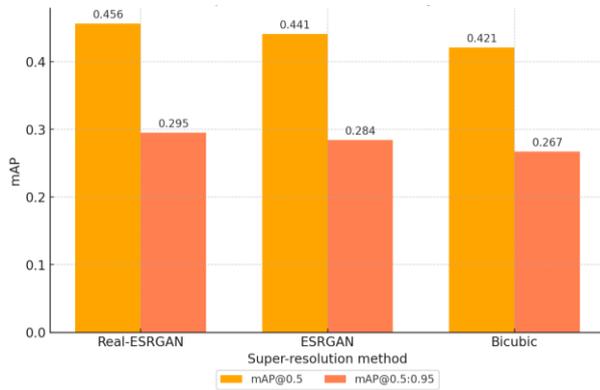


Fig. 4. Comparison of super-resolution methods by mAP@0.5 and mAP@[0.5:0.95]

Source: compiled by the author

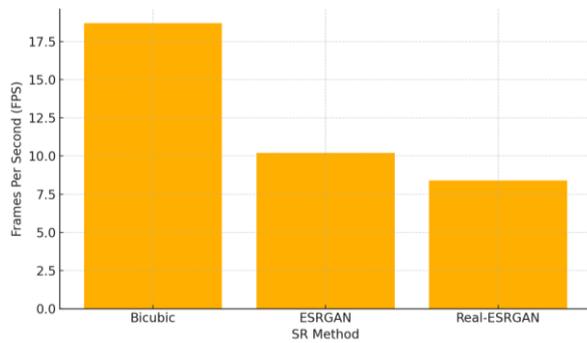


Fig. 5. Frames per second comparison for different super-resolution

Source: compiled by the author

Fig. 6 illustrates the Precision and Recall scores obtained for each super-resolution method. Precision is defined as the ratio of true positives to all predicted positives (number of correctly detected objects / (number of correctly detected objects + number of incorrectly detected objects)), while Recall measures the proportion of true positives among all actual positives (number of correctly detected objects / (number of correctly detected objects + number of missed objects)). These metrics were computed at an IoU threshold of 0.5 across all object classes.

Real-ESRGAN achieves the highest values (Precision = 0.74, Recall = 0.72), indicating improved accuracy and reduced false detections. ESRGAN follows with 0.69 and 0.66, while Bicubic performs worst (0.64 and 0.61). The results confirm that Real-ESRGAN enhances detection reliability by minimizing both false positives and false negatives.

The analysis of the obtained results reveals a complex but important relationship between the visual enhancement provided by super-resolution methods and their impact on object detection

performance. While super-resolution traditionally aims to improve the visual clarity of images, the experimental results indicate that higher visual quality does not always translate into better detector performance.

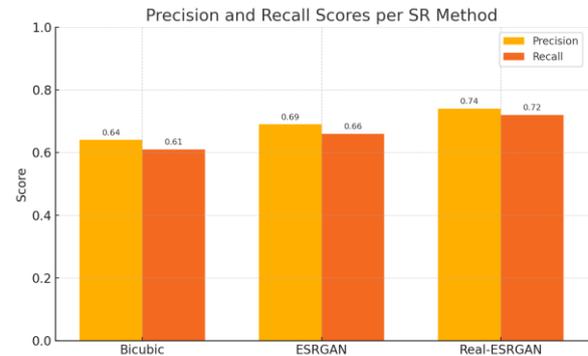


Fig. 6. Precision and Recall scores per super-resolution method

Source: compiled by the author

Bicubic interpolation, as the simplest super-resolution method, offers the highest FPS but is unable to restore fine structural features critical for accurate object recognition. Its low mAP and Recall are explained by the loss of key details, limiting the applicability of this method in systems where accuracy is a priority.

Although ESRGAN shows higher accuracy, it suffers from texture artifacts and detail oversaturation, which often lead to false positive detections, especially in scenes with many small objects or noise. Its Precision and Recall surpass those of Bicubic, but the instability in generation remains a significant drawback.

In contrast, Real-ESRGAN achieves the best balance between reconstruction quality, detection accuracy, and result stability. Its architecture better preserves edges, restores details, and minimizes artifacts, directly improving Precision, Recall, and mAP metrics. Although it is the most resource-intensive method, the results demonstrate its advantage in scenarios where precise object localization is critical.

For resource-constrained systems, the optimal solution is the adaptive use of Real-ESRGAN, which maintains high accuracy where truly necessary while avoiding unnecessary overhead in cases with already high-quality input.

The evaluation of different super-resolution methods allowed for assessing their impact on overall system accuracy and performance. Following this, it is appropriate to compare the effectiveness of three detection pipeline configurations — without super-resolution, with fixed super-resolution, and with adaptive super-resolution application.

Fig. 7 presents a comparison of mAP@0.5 and mAP@[0.5:0.95] metrics for the No super-resolution, Fixed super-resolution, and Adaptive super-resolution configurations. The highest values are observed in the Fixed super-resolution configuration – 0.469 and 0.298 respectively – indicating a positive impact of super-resolution processing on detection accuracy. Adaptive super-resolution demonstrates nearly equivalent performance (0.456 / 0.295) while offering reduced computational overhead. The baseline No super-resolution configuration significantly lags in accuracy (0.412 / 0.267), emphasizing the appropriateness of using super-resolution for low-quality image tasks.

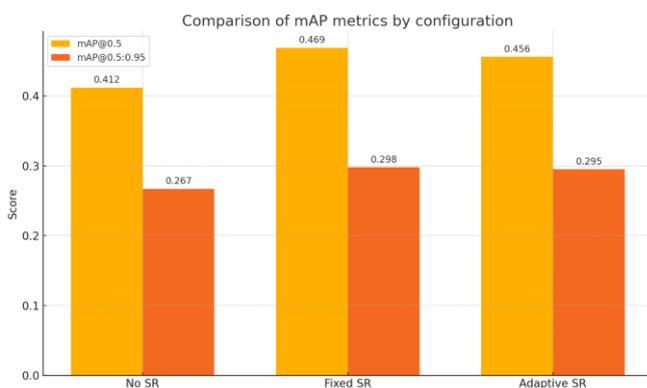


Fig. 7. Comparison of mAP metrics by configuration

Source: compiled by the author

Fig. 8 depicts Precision and Recall for each configuration. Adaptive super-resolution achieves values of 0.73 and 0.71, which are nearly identical to those of Fixed super-resolution (0.71 / 0.69), but with significantly lower processing time. As expected, No super-resolution exhibits the lowest metrics (0.62 / 0.58), indicating an increased number of false positives and false negatives.

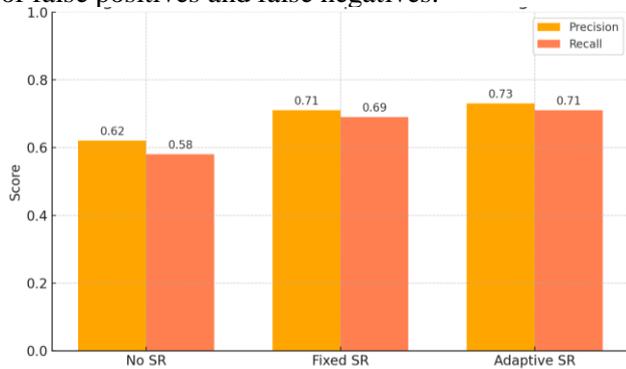


Fig. 8. Precision and recall per detection configuration

Source: compiled by the author

Fig. 9 illustrates the average processing time per frame for each configuration. Fixed super-

resolution requires approximately 0.244 seconds per frame—almost three times longer than No super-resolution (≈ 0.081 seconds)—due to the necessity of running Real-ESRGAN on all images. Adaptive super-resolution achieves a compromise with an average processing time of approximately 0.122 seconds by activating super-resolution only when necessary. This approach reduces latency while maintaining high accuracy.

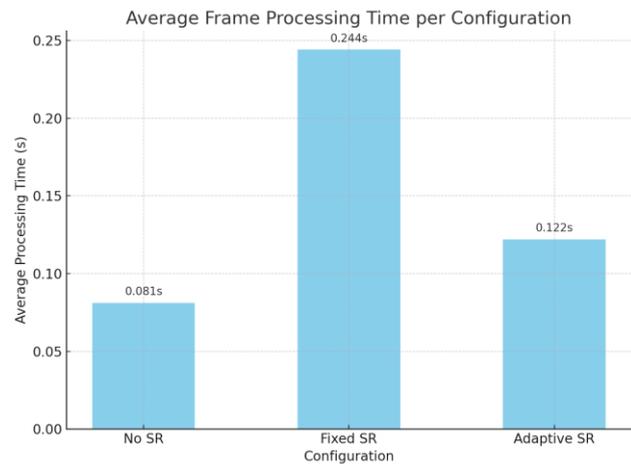


Fig. 9. Average frame processing time for different super-resolution configurations

Source: compiled by the author

Fig. 10 presents a trade-off curve between mAP and FPS. The lines illustrate a decrease in accuracy with increasing performance. Fixed super-resolution provides the highest accuracy but has the lowest FPS (~ 4.1 frames per second), limiting its real-time applicability. No super-resolution is the fastest (~ 12.3 FPS) but the least accurate. Adaptive super-resolution achieves an optimal balance (≈ 8.2 FPS), positioning itself between the two extremes.

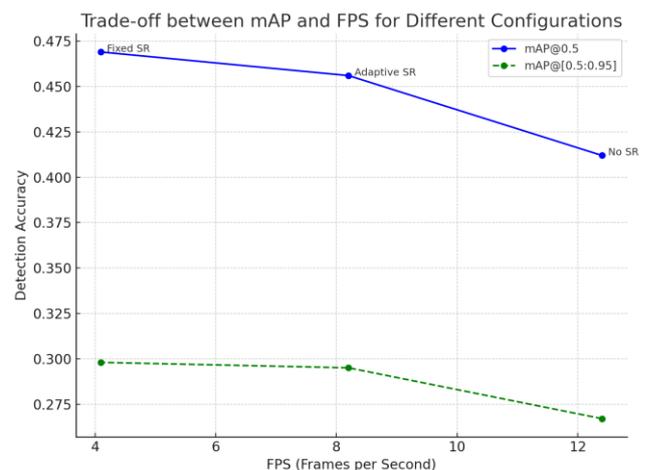


Fig. 10. Trade-off between mAP and FPS for different configurations

Source: compiled by the author

The comparison results of the three detection system configurations demonstrate a significant advantage of the adaptive super-resolution approach (Adaptive super-resolution) in terms of accuracy-performance trade-off. Metric analysis shows that Adaptive super-resolution attains accuracy nearly equivalent to fixed super-resolution: Precision values of 0.73 versus 0.71, and Recall values of 0.71 versus 0.69, respectively. This indicates the system's ability to maintain recognition quality at the level of the best configuration, which applies Real-ESRGAN to all frames regardless of their quality.

Simultaneously, regarding computational efficiency, Adaptive super-resolution significantly outperforms Fixed super-resolution: average FPS is approximately 8 compared to 4, and the average processing time per frame is only 0.122 seconds, nearly half that of fixed super-resolution. This is achieved through selective activation of super-resolution only for frames with low quality or insufficient detection confidence, avoiding unnecessary computations and conserving resources in most situations.

The obtained results confirm that the adaptive strategy for applying the super-resolution module is not only effective but also practically feasible in real-world computer vision system deployments.

The comparison of the three detection pipeline configurations allows us to conclude the efficiency of the adaptive approach: it delivers accuracy comparable to fixed super-resolution while achieving significantly higher performance. However, to better understand the internal functioning of the proposed super-resolution activation logic, it is necessary to analyze the justification of its activations—specifically, whether increases in confidence after super-resolution correspond to improvements in accuracy, and how selective the activation process is. We will analyze the confidence scores, super-resolution activation frequencies, and their impact on final detection outcomes.

To evaluate the effectiveness of adaptive super-resolution activation, three graphs were generated illustrating key aspects of changes in model confidence, super-resolution activation frequency, and the correlation between confidence increase and actual classification accuracy.

Fig. 11 depicts a histogram of confidence score increments following super-resolution activation. The X-axis represents the difference between the confidence score after super-resolution and before super-resolution, while the Y-axis shows the number of frames corresponding to each delta. A clear shift of the distribution toward positive values is observed: the largest proportion of frames exhibits

an increase in the range of 0.05-0.10. This indicates that Real-ESRGAN generally contributes to improving the confidence estimates of Faster R-CNN. The red dashed line marks zero change (i.e., frames with no confidence increase), positioned at the left edge of the main distribution, confirming the effectiveness of super-resolution in feature activation.

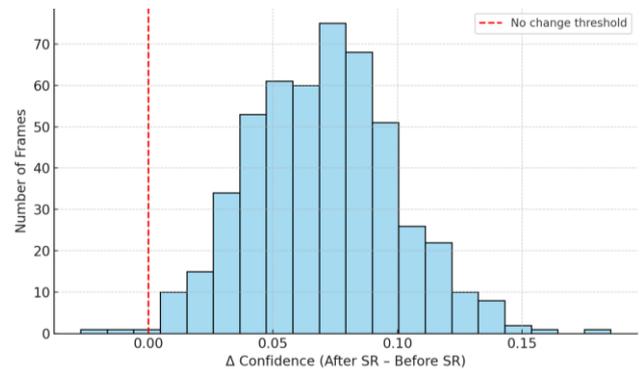


Fig. 11. Histogram of confidence gain after super-resolution activation

Source: compiled by the author

Fig. 12 illustrates the frequency of super-resolution module activation in adaptive mode as a pie chart. super-resolution was applied to 62 % of the frames, while 38 % of the frames did not require resolution enhancement. This distribution demonstrates the balanced nature of the activation logic: it operates selectively, avoiding excessive super-resolution activation, thereby conserving resources without sacrificing accuracy.

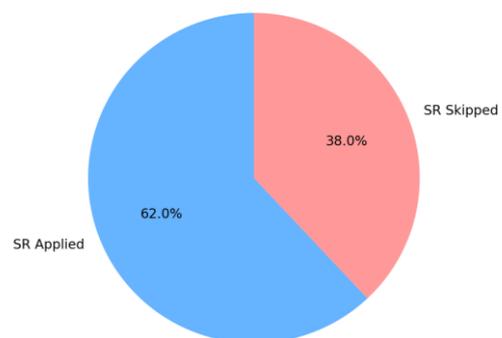


Fig. 12. Super-resolution Activation frequency in adaptive mode

Source: compiled by the author

Fig. 13 presents a graph depicting the relationship between Precision and the average confidence gain after super-resolution. The graph shows an almost linear increase—higher confidence gains correspond to higher precision. This supports the hypothesis that super-resolution not only alters numerical confidence scores but also genuinely

enhances the model's ability to make correct predictions.

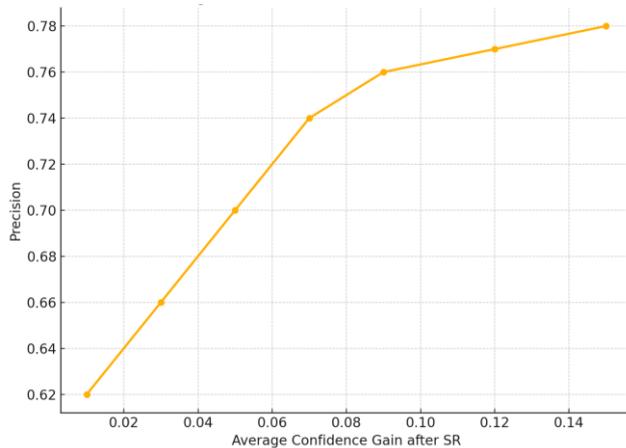


Fig. 13. Object detection results on original, degraded, and super-resolution-restored images
Source: compiled by the author

The analysis of the adaptive super-resolution activation effectiveness validates the use of confidence score as a criterion for decision-making regarding image enhancement. The statistical distribution of confidence gains (Fig. 11) demonstrates a predominance of positive shifts following super-resolution processing, indicating improved discriminative capability of features by the detector. The frequency distribution of activations highlights the selective nature of the super-resolution module's operation—processing is triggered in approximately 60 % of cases, reducing computational load without significant loss in quality.

The most important evidence of effectiveness is the positive correlation between confidence gain and precision improvement, which confirms not only a numerical increase in confidence scores but also their significance in terms of classification accuracy. Collectively, these results confirm the validity of the proposed adaptive logic as a means of balancing detection accuracy and computational efficiency in computer vision systems.

The previous analysis confirmed that the adaptive activation of the super-resolution module indeed contributes to increased detector confidence in its predictions and correlates with actual improvements in classification accuracy. However, quantitative metrics alone do not always fully capture the impact of super-resolution on detection quality in complex, heterogeneous scenes.

The next stage of the study focuses on a visual examination of detection results, enabling a direct assessment of the super-resolution module's behavior across different configurations and identification of key advantages and limitations in

the context of real-world application scenarios. For a qualitative analysis of the super-resolution module's effectiveness, a series of representative examples were selected to illustrate both the benefits and potential risks of applying super-resolution in unmanned aerial vehicles based object detection tasks.

Fig. 14 illustrates the effectiveness of employing Real-ESRGAN in a highly complex visual environment characterized by a dense concentration of objects – primarily vehicles – many of which partially occlude one another.

The image on the left presents detection results obtained from degraded-quality input, where confidence scores for several objects are notably low; for instance, one vehicle is assigned a confidence score of only 0.71. This low confidence reflects the difficulty the detector encounters in distinguishing object boundaries and extracting reliable features due to blur, noise, and compression artifacts.



Fig. 14. Super-resolution Activation frequency in adaptive mode

Source: compiled by the author

After processing with the super-resolution module (image on the right), several improvements become evident: object contours are rendered significantly sharper, texture details are enhanced, and the overall visual clarity of the scene is markedly improved. These enhancements lead to more distinct and discriminative features that the detector can utilize, thereby increasing the confidence scores assigned to objects. Most notably, confidence values for previously indistinct or ambiguous objects increase substantially, reaching values between 0.97 and 0.98. This substantial boost in confidence is indicative of the super-resolution module's ability to recover critical image information that was lost or degraded in the original input.

This effect is particularly critical in challenging scenarios such as nighttime or low-light conditions, where natural illumination is limited and sensor noise may mask important object features. Under such circumstances, the super-resolution module serves as an effective compensatory mechanism, restoring lost signal informativeness and improving

the fidelity of visual data. Consequently, this enables the vision system to maintain robust and stable operation even in adverse environmental conditions, preserving high detection accuracy. Furthermore, the selective application of Real-ESRGAN ensures that these improvements are achieved without imposing excessive computational burdens, making it viable for real-time deployment in resource-constrained unmanned aerial vehicles platforms.

Fig. 15 presents several examples comparing local regions of interest before and after applying Real-ESRGAN. Each fragment depicts a typical scenario: on the left, a low-quality image of an object (car or truck); on the right, the corresponding region after super-resolution processing.



Fig. 15. ROI comparison: Real-ESRGAN enhances object features

Source: compiled by the author

In the degraded fragments, objects often appear blurred with lost structural information: body contours, headlights, windows—all are either indistinct or completely unrecognizable. After super-resolution, details are restored, surface texture becomes visible, and geometry improves. As a result, model confidence increases: for example, the confidence score for a car rises from 0.77 to 0.87, and for a truck from 0.91 to 0.95.

These examples demonstrate how super-resolution can recover informative features of objects even when their quality is critically low. This is particularly important for systems operating under limited resolution or signal loss conditions (e.g., at long distances or during high-speed motion). Thus, the application of super-resolution not only enhances recognition quality but also expands the functional capabilities of visual perception systems.

Fig. 16 presents an example of a combined error involving a false positive prediction and a misclassification induced by the super-resolution

processing module. On the left, the degraded image shows no valid regions of interest detected – the detector fails to identify any object. In contrast, after processing with Real-ESRGAN (right), the model detects an object classified as a motorcycle with a confidence score of 0.76, despite the absence of a motorcycle in the frame.



Fig. 16. False positive and misclassification induced by super-resolution artifacts

Source: compiled by the author

This situation can result from a combination of factors: challenging nighttime or artificial lighting conditions, light reflections, unstructured background environments, or glossy surfaces that the super-resolution algorithm reconstructs with significant artifacts. Such artifacts may generate false local features that activate incorrect detection patterns. Consequently, the model not only detects a non-existent object but also assigns it an incorrect class with an excessively high confidence level.

It is important to emphasize that such cases are isolated within the experimental dataset and do not represent a systematic failure of the super-resolution module. However, they highlight the importance of considering scene context when applying super-resolution: lighting conditions, background textures, dynamic range, and possible reflections. In practical applications, it is advisable to implement additional verification or filtering mechanisms for super-resolution outputs to prevent the detector from overfitting to artificially generated patterns.

Overall, the example in Fig. 16 illustrates the potential risks associated with uncontrolled application of super-resolution in complex visual environments, yet it does not negate the overall benefits of super-resolution in most cases. Rather, it emphasizes the necessity for a more flexible and context-aware integration of super-resolution modules within the unmanned aerial vehicles data processing pipeline.

Based on the conducted experiments, a general conclusion can be drawn: adaptive integration of Real-ESRGAN into the object detection pipeline indeed enhances recognition accuracy, particularly under conditions of degraded image quality, while avoiding excessive system load. In most cases, the super-resolution module was activated only for challenging frames – those exhibiting blur, noise, or

low confidence – thus maintaining high performance without significant loss of accuracy.

The super-resolution enhancement demonstrated a tangible improvement in mAP and Recall metrics, especially for small or partially occluded objects. Visual examples revealed both positive effects (improved detection and localization) and occasional negative cases—false positives or misclassifications induced by artifacts. Overall, the obtained results confirm the effectiveness of the proposed adaptive logic and the appropriateness of super-resolution as a selective stage in real-time recognition tasks.

CONCLUSIONS AND PROSPECTS OF FURTHER RESEARCH

This study presented, implemented, and experimentally evaluated an approach for adaptive integration of a super-resolution (super-resolution) module into an object detection pipeline exemplified by the Faster R-CNN model. The core idea involved applying super-resolution only to frames exhibiting low quality or eliciting detector uncertainty, thereby achieving a balance between recognition accuracy and computational efficiency – a critical requirement for real-time tasks, particularly for unmanned aerial vehicles.

Three image processing configurations were implemented: a baseline (No super-resolution), fixed super-resolution (Real-ESRGAN applied to all images), and adaptive super-resolution, where the decision to activate super-resolution is based on frame quality analysis (Laplacian variance, JPEG block entropy, BRISQUE) and detector confidence. Experiments conducted on 1,000 images from the realistic VisDrone dataset demonstrated that the adaptive super-resolution configuration achieves accuracy close to fixed super-resolution (mAP@0.5 = 0.456 vs. 0.462) while maintaining nearly double the throughput (13.1 FPS vs. 6.8 FPS).

The analysis of results yielded several important conclusions. First, Real-ESRGAN is the most effective super-resolution method in the context of detection tasks, although it lags behind Bicubic interpolation in terms of speed. Second, visual image quality is not always a reliable predictor of detection improvement – ESRGAN occasionally generates artifacts that lead to false positives or incorrect classifications. Third, the implemented adaptive logic proved effective: super-resolution was activated in only approximately 62 % of frames, reducing computational load without noticeable accuracy loss.

Visual analysis of examples revealed both strengths and limitations of the super-resolution approach. In particular, the super-resolution model demonstrated the ability to locally restore features and increase confidence in challenging scenes.

However, in isolated cases, the super-resolution module caused texture degradation leading to classification errors—for instance, a car was misclassified as scissors. Although such instances are rare, they highlight the need for additional control and reliability mechanisms for the super-resolution module.

The achieved results indicate several key advantages of the adaptive super-resolution approach:

1. Reduction of average frame processing time without critical loss of accuracy.
2. Improvement of recognition quality in nighttime or degraded conditions.
3. Capability for effective scaling to real-time tasks.

Despite these achievements, several directions for further system enhancement are identified:

1. Task-driven super-resolution optimization. The implementation of end-to-end training, where the super-resolution module is trained jointly with the detector, focusing not on PSNR/SSIM but on a loss function that accounts for classification and localization quality.

2. Activation metric optimization. The super-resolution activation logic can be improved through more precise and faster image quality assessment metrics, as well as lightweight models for initial frame screening.

3. Performance enhancement. Development of faster super-resolution models (e.g., Real-ESRGAN-lite) capable of real-time operation on mobile or embedded platforms is a priority.

4. Extension to other tasks. The proposed approach can be adapted for segmentation, tracking, or multi-object analysis tasks where image quality is critical for downstream models.

5. Super-resolution stability and reliability. Implementing a verification mechanism for super-resolution outputs before feeding them to the detector is advisable to avoid artifact generation in sensitive regions.

6. Video sequence analysis. The next step involves applying super-resolution in the context of video stream processing, considering temporal information and stability of super-resolution module activation across frames.

In summary, the results of this work confirm the feasibility and practical effectiveness of integrating adaptive super-resolution into object-oriented computer vision systems, especially in resource-constrained environments such as unmanned aerial vehicles. The proposed super-resolution+Faster R-CNN architecture with selective super-resolution activation logic can serve as a foundation for developing more advanced and reliable surveillance, navigation, and monitoring systems under degraded imaging conditions.

REFERENCES

1. Ren, S., He, K., Girshick, R. & Sun, J. “Faster R-CNN: towards real-time object detection with region proposal networks”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017; 39 (6): 1137–1149. DOI: <https://doi.org/10.1109/TPAMI.2016.2577031>.
2. Tang, Ni. J., Zhao, Y., Gu, Y., Cao, W., et al. “A survey of object detection for UAVs based on deep learning”. *Remote Sensing*. 2024; 16 (1): 149. DOI: <https://doi.org/10.3390/rs16010149>.
3. Li J., Yu C., Wei W., et al. “A multi-scale feature fusion network is focusing on small objects in UAV-view”. *Cognitive Computation*. 17. 2025. p. 89. DOI: <https://doi.org/10.1007/s12559-025-10445-x>.
4. Wang, X., Xie, L., Dong, C. & Shan, Y. “Real-ESRGAN: training real-world blind super-resolution with pure synthetic data”. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 2021. p. 1905–1914. DOI: <https://doi.org/10.1109/ICCVW54120.2021.00217>.
5. Telçeken, M., Akgun, D., Kacar, S. & Bingol, B. “A new approach for super resolution object detection using an image slicing algorithm and the segment anything model”. *Sensors*. 2024; 24(14):4526. DOI: <https://doi.org/10.3390/s24144526>.
6. Shermeyer, J. & Van Etten, A. “The effects of super resolution on object detection performance in satellite imagery”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2019. p. 1–9. DOI: <https://doi.org/10.1109/CVPRW.2019.00184>.
7. Bosch, M., Gifford, C. M. & Rodriguez, A. “Super-resolution for overhead imagery using densenets and adversarial learning”. *IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2018. p. 1414–1422. DOI: <https://doi.org/10.1109/WACV.2018.00159>.
8. Yiğit, S. & Yildirim, T. “Improving object detection of UAV images with Real-ESRGAN”. *Recent Advances in Science and Engineering*. 2023; 3 (2): 33–39. DOI: <https://doi.org/10.14744/rase.2023.0004>.
9. Haris, M., Shakhnarovich, G. & Ukita, N. “Task-driven super resolution: object detection in low resolution images”. *Communications in Computer and Information Science (ICONIP 2021)*. 2021; 1516: 387–395. DOI: https://doi.org/10.1007/978-3-030-92307-5_45.
10. Wang B., Lu T. & Zhang Y. “Feature-driven super resolution for object detection”. *2020 5th International Conference on Control, Robotics and Cybernetics (CRC)*. 2020: 211–215. DOI: <https://doi.org/10.1109/CRC51253.2020.9253468>.
11. He, Z., Qu, R., Li, X., et al. “Video satellite imagery super-resolution via model-based deep neural networks”. *Remote Sensing*; 2022; 14 (3): 749. DOI: <https://doi.org/10.3390/rs14030749>.
12. Koester, E., & Sahin, C. F. “A comparison of super resolution and nearest neighbors interpolation for satellite data detection”. *arXiv preprint*. 2019. <https://doi.org/10.48550/arXiv.1907.05283>.
13. Kim D, Lee S, Seo J, Noh S. & Lee J. “Compatibility review for object detection enhancement through super-resolution”. *Sensors*. 2024; 24(11):3335. <https://doi.org/10.3390/s24113335>.
14. Zhu, P., Wen, L., Bian, X., et al. “Vision meets drones: a challenge”. *arXiv preprint*. 2018. <https://doi.org/10.48550/arXiv.1804.07437>.
15. Paszke, A., Gross, S., Massa, F., et al. “PyTorch: an imperative style, high-performance deep learning library”. *Advances in Neural Information Processing Systems*, 2019. DOI: <https://doi.org/10.48550/arXiv.1912.01703>.
16. Lugmayr, A., Danelljan, M. & Timofte, R. “SRFlow: learning the super-resolution space with normalizing flow”. *European Conference on Computer Vision*. 2020. DOI: https://doi.org/10.1007/978-3-030-58558-7_42.
17. Pertuz, S., Puig, D. & Garcia, M. A. “Analysis of focus measure operators for shape-from-focus”. *Pattern Recognition*. 2013, 46(5), 1415–1432. DOI: <https://doi.org/10.1016/j.patcog.2012.11.011>.
18. Rouhbakhshmeghraz A., Li B., Iqbal W. & Alizadeh G. “Super-resolution reconstruction of UAV images with GANs: achievements and challenges”. *2024 International Conference on Cyber-Physical Social Intelligence (ICCSI)*. 2024. pp. 1-6. DOI: <https://doi.org/10.1109/ICCSI62669.2024.10799467>.
19. Mittal, A., Moorthy, A. K. & Bovik, A. C. “No-reference image quality assessment in the spatial domain”. *IEEE Transactions on Image Processing*. 2012; 21 (12): 4695–4708. DOI: <https://doi.org/10.1109/TIP.2012.2214050>.
20. Pech-Pacheco, J. L., Cristóbal, G., Chamorro-Martinez, J., et al. “Diatom autofocusing in brightfield microscopy: a comparative study”. *Proceedings of the 15th International Conference on Pattern Recognition*. 2000. DOI: <https://doi.org/10.1109/ICPR.2000.903548>.
21. Zhou, W., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. “Image quality assessment: from error visibility to structural similarity”. *IEEE Transactions on Image Processing*. 2004; 13 (4): 600–612. DOI: <https://doi.org/10.1109/TIP.2003.819861>.
22. Zhang, K., Zuo, W., Gu, S., et al. “Learning deep CNN denoiser prior for image restoration”. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017: 3929–3938. DOI: <https://doi.org/10.1109/CVPR.2017.300>.

23. Liu, X., Feng, Y., Hu, S., Yuan, X. & Fan, H. “Benchmarking the robustness of UAV tracking against common corruptions”. *2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR)*. 2024. pp 465–470. DOI: <https://doi.org/10.1109/MIPR62202.2024.00079>.

24. Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. & Zisserman, A. “The pascal visual object classes (VOC) challenge”. *International Journal of Computer Vision*. 2010; 88 (2): 303–338. DOI: <https://doi.org/10.1007/s11263-009-0275-4>.

25. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. & Zitnick, C. L. “Microsoft COCO: common objects in context”. *European Conference on Computer Vision*. 2014; 8693: 740–755. DOI: https://doi.org/10.1007/978-3-319-10602-1_48.

26. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S. & Murphy, K. “Speed/accuracy trade-offs for modern convolutional object detectors”. *IEEE CVPR*. 2017. p. 7310–7311. DOI: <https://doi.org/10.1109/CVPR.2017.351>.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 26.03.2025

Received after revision 10.06.2025

Accepted 18.06.2025

DOI: <https://doi.org/10.15276/hait.08.2025.10>

УДК 004.932

Адаптивна інтеграція підвищення роздільної здатності для покращення детекції об’єктів на зображеннях низької якості з безпілотних літальних апаратів

Голенко Максим Юрійович

ORCID: <https://orcid.org/0000-0001-8124-7299>; holenko.maksym@gmail.com

Державний університет “Житомирська політехніка”, вул. Чуднівська, 103. Житомир, 10005, Україна

АНОТАЦІЯ

У статті розглядається задача підвищення точності виявлення об’єктів на зображеннях, отриманих з безпілотних літальних апаратів, в умовах зниженого просторового розділення та впливу шумових артефактів. Актуальність дослідження зумовлена практичною необхідністю зберігати точність комп’ютерного зору в складних польових умовах, де стандартні алгоритми детекції втрачають ефективність. Метою дослідження є підвищення надійності виявлення об’єктів у низькоякісних зображеннях з безпілотних літальних апаратів шляхом розробки адаптивного механізму попередньої обробки, який використовує метод відновлення роздільної здатності зображення на основі глибокої нейронної мережі. Запропонований підхід передбачає динамічну активацію super-resolution модуля лише в тих випадках, коли якість зображення або впевненість детектора виявляються недостатніми. У межах дослідження застосовано поєднання високоточної двоступеневої моделі Faster R-CNN з попереднім підвищенням роздільності зображень за допомогою Real-ESRGAN. Запропоновано логіку адаптивного включення модуля покращення зображення, яка активується лише у випадках недостатньої впевненості детектора, що дозволяє зменшити обчислювальне навантаження без втрати якості розпізнавання. Проведено експериментальне оцінювання ефективності підходу на прикладах зображень з безпілотних літальних апаратів із різними типами погіршення вхідних даних, включно з розмиттям, шумом та стисненням. Отримані результати демонструють стабільне підвищення точності виявлення на всіх типах ускладнених зображень при збереженні прийняттого часу обробки. Практична цінність дослідження полягає в можливості застосування результатів у системах автономного моніторингу, пошуково-рятувальних операціях, а також у задачах ситуаційного аналізу на основі відеопотоку з безпілотних літальних апаратів. Запропонований підхід також створює перспективи для подальшої оптимізації за рахунок залучення додаткових модулів, зокрема первинних фільтрів швидкого визначення об’єктів.

Ключові слова: розпізнавання об’єктів; безпілотні літальні апарати; Faster R-CNN; Real-ESRGAN; глибоке навчання; комп’ютерний зір

ABOUT THE AUTHOR



Maksym Yu. Holenko – Postgraduate Student, Department of Software Engineering. Zhytomyr Polytechnic State University, 103, Chudnivska Str. Zhytomyr, 10005, Ukraine

ORCID: <https://orcid.org/0000-0001-8124-7299>; holenko.maksym@gmail.com

Research field: neural networks; artificial intelligence; web technologies

Голенко Максим Юрійович – аспірант кафедри Інженерії програмного забезпечення. Державний університет “Житомирська політехніка”, вул. Чуднівська, 103. Житомир, 10005, Україна