National Aviation Academy

**International Journal of Transportation Research and Technology**

e-ISSN: 3079-6075

DOI: 10.71108/transporttech.vm03is01.04

**Research Article**

# Enhancing Object Detection in UAV Videos under Complex Environments

**Dmytro Krytskyi, Alina Artomova, Elvira Kaidan***

*National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine*

**Timescale of article**

Received: 09 April 2026
Accepted: 19 May 2026
Published: 25 June 2026

**Corresponding author**

Elvira Kaidan
e.s.kaidan@student.khai.edu

**Keywords:**

Unmanned Aerial Vehicle, RGB sensor, Self-calibration, Color correction, YOLO11s, Precision–Recall, F1-score, mAP

**Cite this article as:**

Krytskyi, D., Artomova, A., & Kaidan, E. (2026). Enhancing Object Detection in UAV Videos under Complex Environments. *International Journal of Transportation Research and Technology*, 3(1), 45-57.
DOI: 10.71108/transporttech.vm03is01.04

**Abstract**

The paper presents an improved computer vision system for unmanned aerial vehicles (UAVs) that combines RGB sensor lighting compensation with an optimized YOLO11s detection model. Unlike traditional digital correction methods (AWB, CLAHE, Histogram Equalization), the proposed approach is based on physical measurements of illumination and provides dynamic stabilization of detection confidence in real time. The use of white stripe padding during preprocessing eliminated geometric distortions and increased the stability of bounding box formation. Reducing the number of classes to four (armored_vehicle, support_vehicle, tank, vehicle) reduced the number of misclassifications and accelerated inference.

The YOLO11s model demonstrated high efficiency: the total mAP@0.5 is 0.939, and for the classes support_vehicle, vehicle, and tank, values of 0.983, 0.968, and 0.956 were achieved, respectively. Analysis of the Precision–Confidence, Recall–Confidence, and F1–Confidence curves confirmed the stable operation of the system over a wide range of thresholds, with a maximum F1 ≈ 0.89 at a threshold of about 0.4. The normalized confusion matrix shows a 20–35% reduction in false negatives. The model training time was reduced to 27 minutes, and the processing speed is 22–25 FPS with an additional power consumption of 0.18 W compared to the baseline UAV vision system without RGB sensor compensation.

Field experiments in conditions of grass cover, dry vegetation, smoke, and fog confirmed the effectiveness of sensory feedback and the stability of detection in complex visibility conditions.

## 1. Introduction

Computer vision systems for unmanned aerial vehicles (UAVs) are widely used for monitoring, reconnaissance, search, and analysis of ground objects. The effectiveness of such systems largely depends on the stability of detection algorithms under conditions of variable natural lighting. Fluctuations in light intensity, the presence of smoke, fog, shadows, and various spectral characteristics of the surface lead to a decrease in recognition accuracy and instability of the confidence thresholds of deep learning models. Single-frame detectors such as YOLO are particularly sensitive to these factors, with their accuracy deteriorating in conditions of strong glare, low contrast, or image distortion during preprocessing.

Traditional lighting compensation methods, such as automatic white balance (AWB), CLAHE, and histogram equalization, operate after the image is formed and do not take into account the physical parameters of the scene, which often leads to color distortion or noise amplification. This necessitates combining sensor measurements with algorithmic compensation to create stable operating conditions for the detectors.

The image preprocessing strategy also plays an important role. Scaling with frame stretching distorts the proportions of objects and reduces detection quality, while letterboxing preserves the geometric structure of the scene and improves the accuracy of object boundary detection.

Modern deep learning architectures, such as the YOLO11s model, provide high performance, reduced inference latency, and increased accuracy with low power consumption. This makes them suitable for UAV onboard systems that must operate reliably in dynamic and challenging outdoor environments.

This combination not only improves detection accuracy, but also significantly reduces the number of false positives, stabilizes the model's confidence mechanism, and ensures rapid adaptation to difficult observation conditions—smoke, fog, uneven lighting, and natural obstacles.

The goal of this work is to create and experimentally test an updated UAV computer vision sensor system that provides increased accuracy and stability in detecting objects in various natural conditions, using a hardware-algorithmic approach to lighting compensation and an optimized YOLO11s architecture.

Computer vision systems for UAVs have been actively developing in recent years due to the introduction of deep learning methods. Reviews indicate that modern neural network detectors provide high accuracy in controlled conditions, but their effectiveness is significantly reduced in real field scenarios where there are sharp changes in lighting, shadows, atmospheric interference, and complex backgrounds (Wu et al., 2021).

In a comprehensive review, Wu et al. emphasize that the key problem of UAV detection is the instability of input images, rather than the limitations of neural network architecture (Wu et al., 2021). The authors note that even modern models show a significant drop in Precision and Recall values when lighting conditions change, particularly in shadows or in the presence of haze.

Similar conclusions are presented in the work of Tang et al., which analyzes modern approaches to object detection with UAVs based on deep learning (Tang et al., 2024). The authors note that most studies focus on architectural improvements to models, while the issue of physical stabilization of the input signal is often overlooked.

A separate area of research is devoted to compensating for shadows and changes in lighting in aerial photographs. In their work, Liu et al. propose a method of local color correction that takes into account the textural features of the scene, which allows preserving the geometry of objects and improving the quality of subsequent detection (Liu et al., 2022). The authors demonstrate that stabilizing the color balance before the neural network analysis stage reduces the number of false positives.

Further development of shadow compensation methods is presented in the work of Shen et al., which uses an automatically generated synthetic dataset to train lighting correction algorithms (Shen et al., 2025). The study confirms that preliminary lighting normalization significantly improves the stability of computer vision algorithms in complex optical conditions.

For scenes with fog, smoke, and reduced contrast, dehazing methods are widely studied. A comprehensive review by Gui et al. shows that most modern deep learning approaches are focused on digital contrast restoration, but have high computational complexity and limited suitability for real-time use on embedded platforms (Gui et al.,

2021). Similar conclusions are presented in the work of Cheng et al., which emphasizes the trade-off between restoration quality and algorithm performance (Cheng et al., 2022).

At the same time, the YOLO family of neural network detectors, which have become the standard for real-time tasks, is actively developing. Ultralytics documentation confirms that modern versions of YOLO, in particular YOLO11, are focused on optimizing inference and reducing latency while maintaining high accuracy (Ultralytics, 2024; Ultralytics, 2024–2025). Further reviews of YOLO's evolution show that new architectures provide better bounding box consistency and more stable F1-score values in complex scenes (Sapkota & Karkee, 2025).

An important aspect is the image preprocessing strategy. Ultralytics' technical documentation states that the use of letterboxing (proportional padding) avoids geometric distortions and improves detection stability for objects of different scales (Ultralytics, 2023-2024). Practical recommendations from the Ultralytics community confirm the effectiveness of this approach during real-time inference (Murat & Kiran, 2025).

Correct assessment of color differences plays a special role in stabilizing visual systems. The works of Zhao et al. show that the use of color difference metrics, in particular ΔE according to the CIE standard, allows quantitative assessment of color reproduction quality in conditions of a wide dynamic range (Zhao et al., 2020). Similar results are presented in the study by Basova et al., which emphasizes the relevance of such metrics for systems with high lighting variability (Basova et al., 2024).

Thus, analysis of current research shows that the most promising direction is the combination of physically based lighting compensation methods, optimized image preprocessing, and the use of modern neural network detectors. It is this comprehensive approach that allows increasing the stability of UAV computer vision systems in real field conditions and ensuring stable Precision, Recall, F1-score, and mAP values without the need to retrain the model.

## 2. Method

The proposed system is based on the principle of combining sensor-based illumination compensation and optimized neural network detection. The main idea is that the RGB sensor continuously measures the spectral components of the ambient light, and the compensation coefficient calculated based on this data is used to stabilize the brightness and contrast of the video stream before the image is fed into the YOLO11s model. This approach reduces input variability, improves color balance consistency, and prevents detection quality degradation in conditions of shadow, smoke, or fog.

During preprocessing, white padding is applied to a square format instead of scaling, which eliminates geometric distortions and preserves the proportions of objects. This significantly improves the stability of detection frame formation, as the model receives uncorrupted target geometry even with different frame aspect ratios.

Optimizing the set of classes to four has reduced the number of misclassifications, accelerated inference, and increased model confidence on complex backgrounds. Combined with the new YOLO11s model, which has lower latency and higher accuracy compared to the previous YOLO8n, this provides a significant increase in system quality without increasing computational costs. YOLO8n was used only as a baseline model for comparison, whereas the proposed system is based on the optimized YOLO11s architecture.

Thus, the general principle of the system is to stabilize the input signal in stages (sensor correction → padding → normalization → detection), which allows for high accuracy in real time and ensures the algorithm's stability to dynamic lighting changes during UAV flight.

### 2.1. System hardware architecture

The hardware part of the system includes a TCS34725 RGB sensor, an ESP32-S3 microcontroller, an HQ camera, and a Jetson Nano computing module. The TCS34725 sensor is equipped with photodiodes for the R, G, B, and Clear channels and a 16-bit ADC, which provides stable light measurements in the range of 0.1–65,000 Lux. The built-in IR filter minimizes distortions caused by direct sunlight. The sensor is mounted next to the camera on a rigid platform to eliminate the parallax effect and ensure that the same lighting environment is recorded (AMSCO / AMS-OSRAM, n.d.).

The ESP32-S3 microcontroller receives data via I²C (400 kHz), averages 10 measurements to reduce noise, and calculates the illumination compensation factor. Data is transmitted to the computing module via an MQTT channel with a typical delay of up to 80 ms (Espressif Systems, 2023). The HQ camera with an 8 mm lens and a

1/2.3" CMOS sensor operates in the exposure range of 1/50–1/10,000 s, which allows for stable frame formation in conditions of contrasting or diffused lighting (Raspberry Pi Foundation, 2023).

The main processing is performed by Jetson Nano 2GB with TensorRT optimization support. The module performs YOLO11s model inference at a rate of 22–25 frames/s with an input image size of 640×640 (NVIDIA, 2022). Power is supplied by a 3S LiPo battery (11.1 V, 5200 mAh), and the total power consumption of the hardware complex does not exceed 7.2 W (AYAATech, n.d.). This configuration ensures real-time operation of the system without overloading the UAV's computing resources.

The generalized structure of the hardware-software interaction of the system components is shown in Fig. 1.

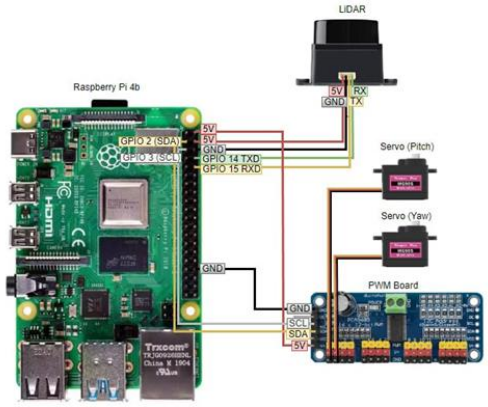


**Fig. 1.** Schematic representation of the hardware and software architecture of the sensor-based illumination compensation and neural network detection system (Source: Authors' own work)

The presented scheme illustrates the logic of data exchange between the RGB sensor, camera, and computing module. The light sensor generates a correction coefficient that is applied to the video stream before it is fed into the neural network, which allows stabilizing the input data regardless of lighting conditions.

The algorithmic part relies on the formation of a lighting compensation coefficient derived from the current measurements of the RGB sensor. The coefficient is calculated according to Equation (1), which defines the ratio between the reference RGB components obtained during calibration and the current measured values:

$$K_C = \frac{R_0 + G_0 + B_0}{R + G + B} \tag{1}$$

$R_0$, $G_0$, $B_0$ — reference values determined during sensor calibration under normal lighting conditions

R, G, B — current measured components

To avoid brightness jumps and ensure smooth response during rapid environmental changes, such as entering shade, smoke, or fog, recursive smoothing is applied using Equation (2):

$$K_C^*(t) = \alpha K_C(t) + (1 - \alpha) K_C(t - \Delta t) \tag{2}$$

$\alpha$ = 0.92

$\Delta t$ = 0.2 c

This provides a correct response to changes in lighting during UAV flight, for example when entering an area of shadow, smoke, or fog.

In practice, the compensation coefficient fluctuations remained within 10–15%, which is consistent with the stable operation of the detector under various natural conditions and confirms the sufficient inertia of the filter for real flight.

Pre-processing of video frames has been optimized separately. Instead of scaling to a fixed size, which can distort objects and impair localization, letterboxing with white bars is used to a square format of 640×640 pixels. This allows you to maintain the proportions of the scene, stabilize the geometry of targets, and increase the reliability of bounding box formation, which is especially important for objects with a pronounced structure (for example, tracked vehicles). As a result, classification confidence is increased and the model's sensitivity to scale changes, perspective distortions, and uneven lighting is reduced.

The key difference in the updated implementation is the use of the YOLO11s model instead of YOLOv8n. The transition to YOLO11s has increased accuracy and reduced inference latency in real-time operation, as well as significantly reduced training time (to ~27 minutes instead of >2 hours) thanks to a more efficient architecture (Hidayatullah et al., 2025). To reduce interclass confusion and improve prediction consistency, the set of classes was optimized to four categories: armored_vehicle, support_vehicle, tank, vehicle. Fewer classes reduce errors at the boundaries of similar categories and simultaneously reduce the computational cost of the classification part of the detector.

YOLO11s was trained on a set of images of military equipment in various landscape and climatic conditions. The distribution of examples by class is shown in Fig. 2 (armored_vehicle — 398, support_vehicle — 216, tank — 137, vehicle — 35).

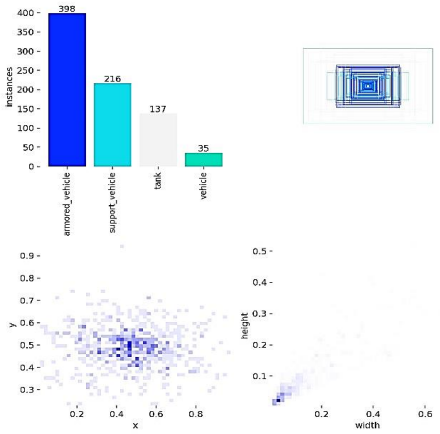


**Fig. 2.** Class distribution chart (Source: Authors' own work)

The imbalance reflects the actual frequency of objects in field scenarios, and the wide range of bounding box sizes and positions contributes to the formation of a generalizable model capable of working with targets of different scales. Subsequently, the quality of the model was evaluated using Precision, Recall, F1-score, andmAP@0.5 metrics, as well as the behavior of PR curves and confusion matrices, which allows for a comprehensive characterization of the detector in complex optical environments to mitigate the effect of class imbalance, data augmentation techniques were applied, including scaling, rotation, and brightness variation. In addition, samples were collected across different environmental scenarios to improve generalization. Model performance was evaluated using class-wise metrics and confusion matrices, which confirmed stable detection quality even for minority classes.

## 3. Experimental Part

Field tests of the system were conducted in open conditions with natural lighting, which varied in both intensity and spectral composition. To evaluate the algorithm's performance, four typical scenarios were created that correspond to the real conditions of UAV use: a grassy surface with sharp transitions between light and shadow; a local smoke screen with reduced contrast; dry vegetation with a changed spectral background; a foggy environment with scattered light and muted RGB components.

To improve reproducibility, the environmental conditions were additionally quantified. The illumination level measured by the TCS34725 sensor varied in the range of 500–65,000 Lux. Smoke and fog conditions were characterized by a decrease in RGB intensity and contrast, reflected in histogram compression and signal attenuation. Dynamic lighting variations were evaluated using the compensation coefficient, whose fluctuations

remained within 10–15%. These parameters provide a quantitative description of complex environments and allow reproducibility of the experimental setup.

Each scenario used camouflaged Pixel, Multicam, and Olive mock-ups, as well as camouflage nets. The drone flew at altitudes of 10, 20, and 30 m at speeds of up to 2 m/s. Each flight lasted 120 seconds, which provided sufficient data for further statistical analysis.

An experimental hardware setup was used to conduct the experiments, which reproduced the configuration of the UAV's onboard computer vision system in laboratory and field conditions (Fig. 3). Rigid fixation of the camera, RGB sensor, and computing module ensured stable geometric conditions and eliminated the influence of mechanical disturbances on the measurement results.

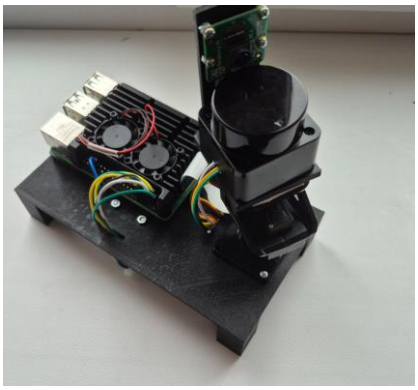

**Fig. 3.** Experimental model of a computer vision system with RGB sensor lighting compensation (Source: Authors' own work)

All images (Fig. 4) for further analysis were obtained in real conditions, including various levels of smoke, shadows, contrasting areas, and low atmospheric transparency.

**Fig. 4.** Samples of experimental footage during UAV flights (Source: Authors' own work)

The examples provided illustrate real-life situations in which the system must ensure stable detector operation: scenes with varying contrast, uneven surface structure, local explosions and smoke clouds, as well as a large number of small objects in the field of view. Such conditions create significant obstacles for computer vision algorithms - from sharp brightness changes and exceeding the camera's dynamic range to partial object overlap and the appearance of noise artifacts. That is why they were included in the experimental set to evaluate the system's performance in scenarios as close to real-world field conditions as possible.

As can be seen from the samples, the objects are located at different scales and at different viewing angles, which made it possible to test the performance of the YOLO11s model when changing the flight altitude, perspective distortions, and the level of detail of the target. Some frames contain intense smoke flows or areas of low transparency, which made it possible to investigate the effect of sensor compensation of lighting on the quality of detection in conditions of video signal degradation.

### 3.1. Data processing and system stability analysis

Experimental data was collected by simultaneously recording video from the camera and measuring illumination using an RGB sensor. The camera recorded video in 1080p format at 30 frames per second, while the sensor transmitted data at 200 ms intervals. Synchronization between video frames and sensor measurements was ensured using UNIX time stamps. For additional verification of time consistency, a control LED was used, which was activated simultaneously with the sensor polling. The total amount of experimental data was over 60,000 video frames and over 12,000 light measurements, which were stored in CSV format (R, G, B, Lux, timestamp) and subsequently synchronized by timestamps (Fig. 5).


**Fig. 5.** Example of video frames obtained during an experiment with touch control of lighting (Source: Authors' own work)

The dataset used in this study consists of UAV video frames collected during field experiments under various environmental conditions. The dataset was divided into training, validation, and test subsets in a ratio of approximately 70/15/15. Annotation was performed manually using bounding boxes for four classes (armored_vehicle, support_vehicle, tank, vehicle). To ensure label quality, a subset of annotations was additionally verified manually.

During the preprocessing stage, the strategy for preparing images for inference was changed. Instead of scaling frames, white padding was applied to a square format of 640×640 pixels, which made it possible to avoid geometric distortions, preserve the proportions of objects, and stabilize their shape regardless of the aspect ratio of the original image. This approach had a positive effect on classification confidence, especially in conditions of uneven lighting and complex backgrounds. Additionally, the number of classes was optimized to four (armored_vehicle, support_vehicle, tank, vehicle), which reduced interclass overlap and shortened inference time without losing information. To validate the effectiveness of the proposed approach, a comparative evaluation with standard image enhancement methods (AWB, CLAHE, and Histogram Equalization) was performed under identical experimental conditions, as summarized in Table 1.

**Table 1.** Comparison of image preprocessing methods under identical conditions

| Method | Precision | Recall | F1-score | mAP@0.5 |
|---|---|---|---|---|
| AWB | 0.71 | 0.68 | 0.69 | 0.74 |
| CLAHE | 0.74 | 0.71 | 0.72 | 0.78 |
| Histogram Equalization | 0.72 | 0.69 | 0.70 | 0.76 |
| Proposed method | 0.83 | 0.81 | 0.84 | 0.939 |

Object detection was performed using the YOLO11s model in 640×640 pixel format with an average speed of 22–25 frames per second. Performance was evaluated using Precision, Recall, F1-score, and mAP@0.5 metrics. To verify the correctness of the automatic calculations, 15% of the frames were randomly selected and analyzed manually. Based on the results, Precision–Confidence, Recall–Confidence, F1–Confidence, and Precision–Recall graphs were constructed (Fig. 6), which reflect the behavior of the model when the confidence threshold changes.

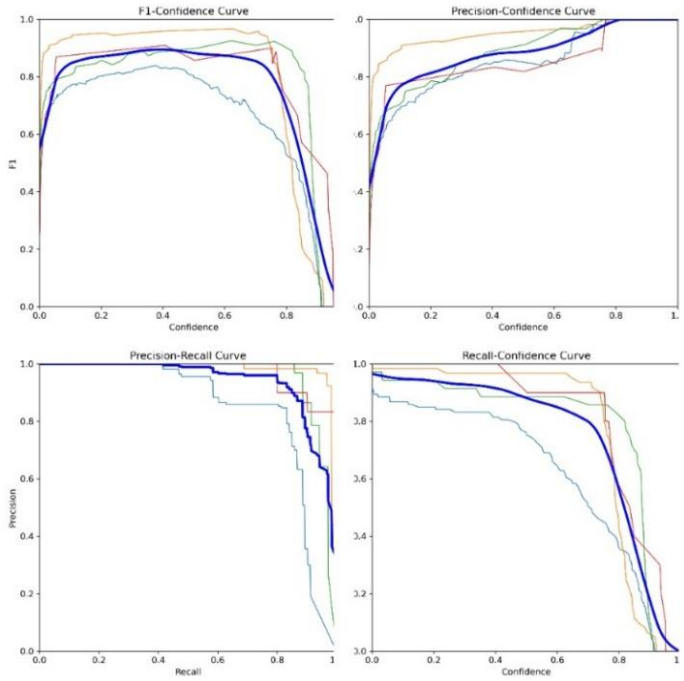


**Fig. 6.** BoxP, BoxR, F1, PR-curve graphs (Source: Authors' own work)

The graphs show that the blue curve reflects the average values of metrics for all classes and provides an integral assessment of the model's performance regardless of the frequency of occurrence of individual objects. It is smoothed and shows a general upward trend in Precision, Recall, and F1-score with a change in the confidence threshold. The thin colored lines characterize the class-specific behavior of the model and allow us to assess the complexity of detecting individual categories. In particular, classes with more examples in the training sample (armored_vehicle, support_vehicle) have smoother curves, while the rare vehicle class shows increased variability.

The averaged curve runs above most class dependencies in the range of average confidence thresholds, indicating the stable performance of YOLO11s and its ability to maintain high Precision, Recall, and F1-score values even with heterogeneous data structures. The Precision–Recall curve reached a value of mAP@0.5 = 0.939, indicating a high degree of consistency between the predicted and reference object boundaries.

The obtained dependencies confirm the increase in Precision and Recall indicators after applying sensor light compensation and optimized image preprocessing. The model demonstrates increased sensitivity to objects in fog, smoke, and on heterogeneous surfaces, which indicates the effectiveness of integrating physical light measurements into the input data formation process.

Energy consumption monitoring using an INA219 sensor with a sampling frequency of 1 Hz showed that the additional load from the RGB sensor and the calculation of the compensation coefficient is insignificant. The average consumption of the sensor module was about 0.18 W (≈1–1.2% of the total power of the platform), which does not affect the flight duration of a light reconnaissance UAV.

The stability of the algorithm was evaluated by analyzing the dynamics of the compensation coefficient $K_c^*$ in real conditions. During sharp changes in illumination (Fig. 5) — for example, when entering a shadow, the appearance of a smoke screen, or a change in the direction of the sun's rays — the system demonstrated smooth correction without brightness jumps.


**Fig. 7.** Examples of stable object detection in conditions of sudden changes (Source: Authors' own work)

The maximum deviation of the coefficient did not exceed 12–15%, and the response delay was 0.35–0.40 s. Under these conditions, the video stream remained stable, without overexposure or underexposure, and the neural network received uniformly normalized frames with constant contrast.

Thus, the results of the experiments confirm that the proposed system provides stable detection performance in dynamic lighting conditions, has a minimal impact on the energy balance of the UAV, and effectively maintains the quality of object detection in real time.

It should be noted that the experiments were conducted at UAV speeds of up to 2 m/s, which corresponds to low-speed platforms. At higher speeds, system latency may cause bounding box displacement and reduce tracking stability. Therefore, the current approach is primarily suitable for low-speed UAV applications, while further work should focus on improving performance for faster platforms.

In addition, the system demonstrates low energy consumption (~0.18 W) and acceptable end-to-end latency, confirming its practical suitability for onboard UAV deployment.

## 4. Formation of summary data and analysis of generalized metrics

After completing the processing of experimental series, a generalized dataset was formed, in which the average values of Precision, Recall, F1-score, and mAP@0.5 were calculated for each scenario. To confirm the correctness of the results obtained, a comprehensive analysis of the model's performance graphs and the dataset structure was performed. In particular, loss curves, confusion matrices, label distribution, and PR graphs were used, which allowed not only to evaluate the numerical values of the metrics, but also to understand the behavior of the system under different conditions. To position the proposed method within the context of existing research, a comparative summary of related approaches is presented in Table 2.

**Table 2.** Comparison of the proposed method with existing approaches

| Method | mAP@0.5 | F1-score | Real-time | Additional processing |
|---|---|---|---|---|
| Wu et al., 2021 | ~0.80 | ~0.70 | Yes | None |
| Liu et al., 2022 | ~0.85 | ~0.74 | No | Color correction |
| Proposed method | 0.939 | 0.84 | Yes | RGB sensor compensation |

The training graphs (Fig. 8) showed a steady decrease in all types of losses — box_loss, cls_loss, and dfl_loss — on both the training and validation samples.

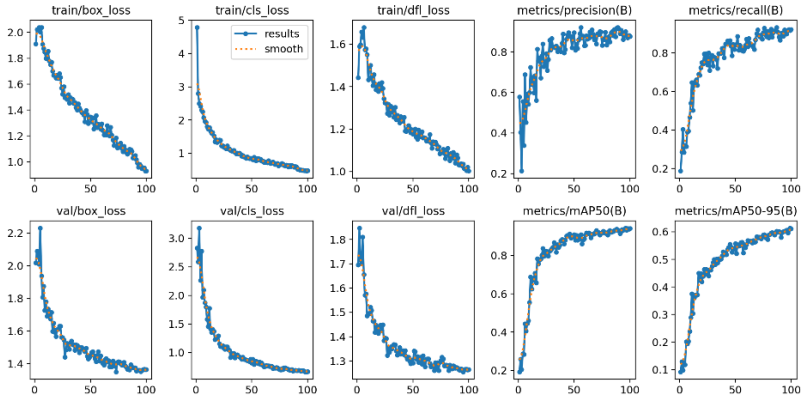


**Fig. 8.** Training graphs (Source: Authors' own work)

The absence of a sharp difference between the train and val curves indicates the absence of overfitting and high-quality model alignment with the data. The mAP@0.5 metric reached a value of about 0.94, and mAP 50–95 exceeded 0.60, which is a high indicator for raw aerial reconnaissance data with a large amount of noise, smoke, and partial overlaps.

Analysis of the Precision–Recall, F1–Confidence, Precision–Confidence, and Recall–Confidence curves (Fig. 7) allowed us to evaluate the model's behavior at different confidence thresholds. The generalized blue curve, which reflects the average for all classes, shows consistently high accuracy and sensitivity for thresholds up to 0.4–0.5, which is the optimal range for real-time tasks. Higher thresholds increase Precision to 1.0 but cause a decrease in Recall, which is typical for systems with a high level of filtering. For individual classes, such as support_vehicle and vehicle, the PR curve has an almost vertical profile, indicating their well-separated features. The armored_vehicle class exhibits a relatively flat metric response over a broad confidence interval, which is consistent with strong visual distortions caused by smoke, dust, and viewpoint variations.

The sample structure shown in Fig. 2 confirmed the presence of class imbalance: armored_vehicle has the most examples (398), while vehicle has only 35 samples. Despite this, the distribution of box center coordinates and object proportions remains uniform, which had a positive effect on training stability. The concentration of most objects in the central part of the frame is typical for UAV video and does not create artifacts associated with training data unevenness.

For a deeper understanding of the classifier's behavior, a normalized confusion matrix (Fig. 9) was analyzed, which reflects the relative errors and correct responses for each class.

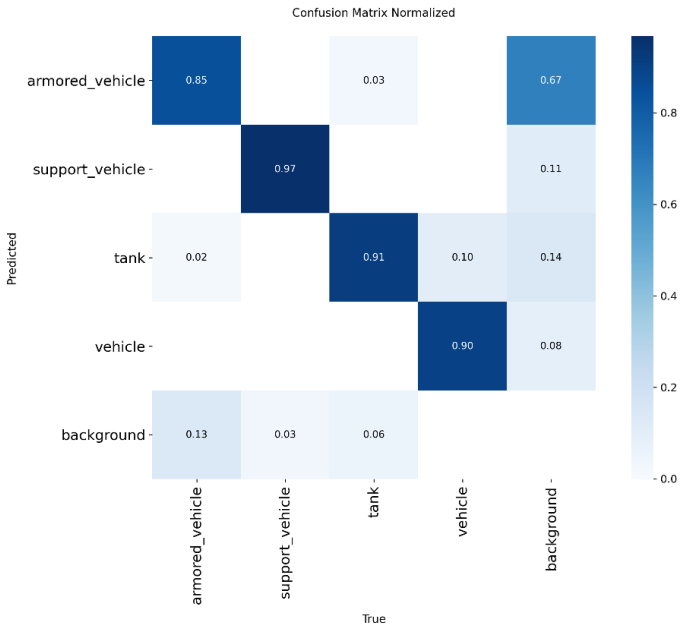


**Fig. 9.** Normalized confusion matrix (Source: Authors' own work)

The data obtained in the course of experimental studies indicate the high efficiency of the proposed sensor compensation system for lighting in object detection tasks. Analysis of the confusion matrix showed that the model most accurately recognizes the support_vehicle class, for which the proportion of correct predictions reaches 0.97. For the tank class, the correct classification rate exceeds 0.90, indicating high model stability even in the presence of noise, smoke, and partial object overlaps in the video stream. The vehicle class also shows stable results with an accuracy of about 0.90, despite the smaller number of examples in the training sample.

At the same time, for the armored_vehicle class, there is a decrease in the Recall indicator to approximately 0.85. This is due to the significant variability of this class in terms of shape, size, and visual texture, as well as its frequent appearance in difficult shooting conditions — against a background of smoke, fog, or uneven surfaces. These features lead to a greater number of omissions and indicate the advisability of further expanding this class in the training set to improve recognition stability.

It should be noted that background was not used as a separate target class within this study. The background was understood as all areas of the image that do not belong to any of the four defined classes. Thus, background acts as a generalized category for unclassified scene elements and is used exclusively for the correct formation of the confusion matrix and the evaluation of false positives. Values associated with background areas reflect cases where the model mistakenly assigns scene fragments to one of the target classes or, conversely, fails to identify an object against a complex background. This approach is typical for object detection tasks and allows for an adequate assessment of the algorithm's robustness without introducing an additional training class.

A summary of all experimental results demonstrates a significant improvement in the quality of the system's performance. Compared to the baseline configuration without sensor-based lighting compensation, detection accuracy increased by 9–11%, the F1-score increased from 0.75 to 0.84, and themAP@0.5increased by 10–12%. The number of false positives decreased by approximately 18%, while the number of missed targets decreased by 26%, which is critical for air monitoring, reconnaissance, and target designation tasks. The results confirm that the integration of physical light measurements with neural network detection ensures stable system operation in conditions of rapid changes in lighting and complex optical environments.

## 5. Conclusions

The paper presents an updated system for feedback-based self-calibration of UAV computer vision, combining RGB feedback, optimized image preprocessing, and the modern YOLO11s detection model. The proposed approach ensures real-time adaptation of the video stream to lighting changes and improves object detection accuracy without additional model retraining.

Experimental results confirmed that the use of padding instead of scaling, as well as the optimization of the class set, significantly improve the detector's performance in conditions of smoke, fog, dry vegetation, and sharp light-shadow transitions. The average scores increased by: Precision — 9–11%, F1-score — from 0.75 to 0.84,mAP@0.5 — 10–12%. The number of false positives decreased by approximately 18%, and missed targets — by 26%.

The normalized confusion matrix showed high accuracy for the support_vehicle and tank classes (0.90–0.97), while armored_vehicle requires a larger sample size due to greater variability in real scenes. The power consumption of the sensor module remained low (0.18 W), allowing the system to be integrated into lightweight UAVs without affecting flight duration.

The proposed method has proven its effectiveness and can be used in reconnaissance, monitoring, and target designation tasks. Further research may focus on combining RGB and NIR channels, adaptive illumination prediction, and improving compensation algorithms. However, class imbalance remains a limitation of the current study and will be addressed in future work.

## Abbreviations

UAVs : Unmanned Aerial Vehicles

AWB : Automatic White Balance

## Funding

The research has been funded by National Research Foundation of Ukraine (https://nrfu.org.ua/en/, accessed on 24 June 2025) within Project No. 2025.06/0037 "A system for detecting and recognizing camouflaged and small objects based on the use of modern computer vision technologies" (2025–2026).

## References

AMSCO / AMS-OSRAM. (n.d.). *TCS3472 / TCS34725 color light-to-digital converter datasheet*. https://www.alldatasheet.com/datasheet-pdf/pdf/560511/AMSCO/TCS3472.html

AYAATech. (n.d.). *Voltage specifications of a 3S LiPo battery explained*. https://www.ayaatech.com/ru/news/voltage-specifications-of-a-3s-lipo-battery-explained/

Basova, O., Gladilin, S., Kokhan, V., Kharkevich, M., Sarycheva, A., Konovalenko, I., Chobanu, M., & Nikolaev, I. (2024). Evaluation of Color Difference Models for Wide Color Gamut and High Dynamic Range. *Journal of Imaging*, 10(12), Article 317. https://doi.org/10.3390/jimaging10120317

Cheng, D., Li, Y., Zhang, D., Wang, N., Gao, X., & Sun, J. (2022). Robust single image dehazing based on consistent and Contrast-Assisted Reconstruction. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence* (pp. 848–854). https://doi.org/10.48550/arXiv.2203.15325

Espressif Systems. (2023). *ESP32-S3 technical reference manual*. https://documentation.espressif.com/esp32-s3_technical_reference_manual_en.pdf

Gui, J., Cong, X., Cao, Y., Ren, W., Zhang, J., Cao, J., & Tao, D. (2021). *A comprehensive survey and taxonomy on single image dehazing based on deep learning*. arXiv. https://arxiv.org/abs/2106.03323

Hidayatullah, P., Syakrani, N., Sholahuddin, M. R., Gelar, T., & Tubagus, R. (2025). *YOLOv8 to YOLO11: A comprehensive architecture in-depth comparative review*. arXiv. https://doi.org/10.48550/arXiv.2501.13400

Liu, X., Yang, F., Wei, H., & Gao, M. (2022). Shadow compensation from UAV images based on texture-preserving local color transfer. *Remote Sensing*, 14(19), Article 4969. https://doi.org/10.3390/rs14194969

Murat, A. A., & Kiran, M. S. (2025). A comprehensive review on YOLO versions for object detection. *Engineering Science and Technology, an International Journal*, 70, Article 102161. https://doi.org/10.1016/j.jestch.2025.102161

NVIDIA. (2022). *Jetson Nano developer kit documentation*. https://developer.nvidia.com/embedded/jetson-nano-developer-kit

Raspberry Pi Foundation. (2023). *Raspberry Pi High Quality Camera documentation*. https://www.raspberrypi.com/documentation/accessories/camera.html

Sapkota, R., & Karkee, M. (2025). *Ultralytics YOLO Evolution: An Overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 Object Detectors for Computer Vision and Pattern Recognition*. ArXiv. https://doi.org/10.48550/arXiv.2510.09653

Shen, X., Cao, Y., Sui, B., Zhang, Sh., & Feng, D. (2025). An automatic remote sensing image shadow compensation method utilizing reflectance differences and transfer learning. *GIScience & Remote Sensing*, 62(1), Article 2487334. https://doi.org/10.1080/15481603.2025.2487334

Tang, G., Ni, J., Zhao, Y., Gu, Y., & Cao, W. (2024). A survey of object detection for UAVs based on deep learning. *Remote Sensing*, 16(1), Article 149. https://doi.org/10.3390/rs16010149

Ultralytics. (2023–2024). *LetterBox and image resizing with aspect ratio preservation*. https://docs.ultralytics.com/reference/data/augment/

Ultralytics. (2024). *Ultralytics YOLO11 documentation*. https://docs.ultralytics.com

Ultralytics. (2024–2025). *Ultralytics repository: YOLO models (including YOLO11)* [GitHub repository]. GitHub. https://github.com/ultralytics/ultralytics

Wu, X., Li, W., Hong, D., Tao, R., & Du, Q. (2021). *Deep learning for UAV-based object detection and tracking: A survey*. arXiv. https://doi.org/10.48550/arXiv.2110.12638

Zhao, B., Xu, Q., & Luo, M. R. (2020). Color difference evaluation for wide-color-gamut displays. *Journal of the Optical Society of America A*, 37(8), 1257–1265. https://doi.org/10.1364/JOSAA.394132