ISSN 2355-3286

Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection Using Aerial Images

Indri Purwita Sary[1], Edmun Ucok Armin[2], Safrian Andromeda[3]
[1,2,3] Electrical Engineering, Universitas Singaperbangsa Karawang, Indonesia
[1] indripurwitasary@gmail.com, [2] edmun.ucok.armin@ft.unsika.ac.id, [3] safrian@ft.unsika.ac.id

Accepted on 10 June 2023
Approved on 30 June 2023

***Abstract*— The development of UAV technology has reached the stage of implementing artificial intelligence, control, and sensing. The use of cameras as UAV data inputs is being employed to ensure flight safety, search for missing persons, and disaster evacuation. Human detection using cameras while flying is the focus of this article. The application of human detection in pedestrian areas using aerial image data is used as the dataset in the deep learning input process. The architectures discussed in this study are YOLOv5 and YOLOv8. The precision, recall, and F1-score values are used as comparisons to evaluate the performance of these architectures. When both architecture performances are applied, YOLOv8 outperforms YOLOv5. The performance of the YOLOv8 model is greater than the YOLOv5 model for Precision, and F1-Score, the difference in the value of each performance is 2.82%, and 0.98%. As for the recall performance value, YOLOv5 is greater than the YOLOv8 model with a difference of 0.54%.**

***Index Terms*— Aerial Image; YOLOv5; YOLOv8; Precision; Recall; F1-Score.**

I. INTRODUCTION

Research on Unmanned Aerial Vehicles (UAVs) has made significant advancements recently. The latest progress in UAV development involves the implementation of artificial intelligence, control, and sensing technologies [1]. UAVs have been utilized by humans to aid in various tasks such as work assistance, surveillance, rescue operations, and security. Detecting humans from camera inputs on UAVs is a crucial task to ensure flight safety Furthermore, the utilization of UAVs in human search missions for safety purposes has also been developed [2]. An example of such utilization is security monitoring, where UAVs equipped with cameras are deployed for aerial surveillance to monitor environments or detect incidents [3]. In addition, detecting humans from UAV images is also used in search and rescue missions for locating and aiding missing or injured individuals in restricted areas [4]. Visual surveillance using UAV platforms has become fascinating. The majority of research works on visual data captured by UAVs are primarily focused on object detection and tracking tasks [5].

Introduction commonly contains the background, purpose of the research, problem identification, and research methodology conducted by the authors which been describe implicitly. Except for Introduction and Conclusion, other chapter's title must be explicitly represent the content of the chapter.

Research related to human object detection using cameras has been extensively conducted in recent years. Various situations are utilized to capture human images. A study on detecting non-rigid small-sized individuals at low altitudes using the VisDrone2019 dataset has been conducted by Xiang Qing Zhang et al., 2022[2]. This research utilized human images captured by cameras mounted on UAVs. The study focuses on improving object detection in complex backgrounds and poor lighting conditions. The DCLANet method employed in this research demonstrates the capability to detect non-rigid small-sized human objects in aerial images taken at low altitudes.

Garbage problems can be overcome by utilizing technology, one of the efforts that continues to be carried out, one of which is by developing robot technology to be implemented as a garbage cleaning tool [7], [8]. In this study, an internet-based garbage collection robot of things will be developed which can be controlled via a smartphone, making it easier to clean up garbage and becoming one of the efforts to maintain environmental cleanliness during the Covid-19 pandemicResearch on the use of YOLOv3 for tracking walking individuals and automatically capturing frontal photos was also conducted by Qifeng Shen in 2018 [6]. The data used in this study involved images captured by UAVs. The methods employed in this research included person detection and recognition using artificial neural networks within YOLOv3, Locality-constrained Linear Coding (LLC) method for face detection and localization, and vision-based UAV control. The results of the study demonstrated that the proposed methods were effective and practical for

tracking individuals and capturing frontal face images using UAVs.

Object detection of humans using aerial image data has also been conducted by Charalampos Symeonidis in 2022. The dataset used in this research is called AUTH-Persons. The dataset consists of videos of human crowds from an aerial perspective. The evaluation method for human object detection is referred to as NMS (Non-Maximum Suppression). This paper describes the dataset, its structure, and the methods used to evaluate the performance of three human detectors: Single Shot Detector (SSD), YOLOv3, and YOLOv4-tiny [1]. The results show that YOLOv3-512 (DarkNet53) achieves the best performance in terms of average precision (AP) at the intersection over union (IoU) thresholds of 0.5 and 0.95. The results also indicate that Seq2Seq-NMS outperforms other NMS methods in terms of AP at the 0.5 and 0.95 IoU thresholds. However, the paper notes that the shift in visual data distribution between training and testing samples can disproportionately negatively affect DNN-based NMS methods that exploit appearance-based features compared to those that do not.

Research on object detection and classification with Unmanned Aerial Vehicles (UAVs) using machine learning algorithms has been conducted. This study compares the performance of two architecture versions, namely YOLOv3 and YOLOv5, for object detection in UAV images. In 2022, Teddy Surya Gunawan utilized data preparation, model training, and implementation methods. The research results demonstrate that the YOLOv5 architecture outperforms YOLOv3 in object detection and classification in UAV images [7].

Research on human detection using two camera data, namely thermal imaging camera and images from a UAV camera, was conducted by Jewel Kate D in 2022. The data collected was processed through a neural network to identify whether the object is a human, and the YOLOv5 algorithm was used to classify the detected objects by the drone. This research utilized the CoCo dataset for the training process and evaluated the accuracy of the device by determining the number of humans detected by the device at various distances and comparing it with the actual number of humans. The research results demonstrate that the device is capable of detecting humans and achieving a low error percentage [8].

Jun-Hwa Kim has implemented YOLOv8 in his research. The YOLOv8 architecture is used to detect drones and birds. This architecture is employed to differentiate between drones and birds. Testing of aerial bird and drone images was conducted using 77 videos. The training results were then used to evaluate 30 video images over 93 epochs. The evaluation metric used was Average Precision (AP) for each test video, with detections considered correct if the Intersection over Union (IoU) with the ground truth box was above 0.5. The frame per second (FPS) of the P2 layer and Multi-Scale Image Fusion (MSIF) with the YOLO-V8-M model were 45.7 fps and 17.6 fps, respectively, for image sizes of 640 and 1280 [9].

Detecting human objects is commonly used as input in various UAV missions. Therefore, the author chose human detection as the object of interest in aerial images. The YOLO architecture has multiple versions, and YOLOv8 is the latest architecture to be used in the testing. Based on previous research, the YOLOv5 architecture has been tested for human detection. Hence, the performance of the YOLOv5 and YOLOv8 architectures will be compared in detecting humans in aerial images.

II. METHODS

The method used in this research is deep learning using the YOLOv5 and YOLOv8 architectures. Deep learning will be applied to human images in pedestrian scenarios. This study focuses on comparing the performance of both architectures.

A. YOLOv5

There have been changes to the standard architecture of model arrangement in YOLOv5. The model arrangement is now divided into three components: backbone, neck, and head. The backbone of YOLOv5 is Darknet 53. Darknet 53 is a new network architecture that focuses on feature extraction characterized by small filter windows and residual connections [10]. The neck of YOLOv5 acts as a connector between the backbone and the head. The neck of YOLOv5 functions to gather and refine the features extracted by the backbone, with a focus on enhancing spatial and semantic information at various scales [11]. The head of YOLOv5 consists of three branches, each predicting features at different scales. Each head produces bounding boxes, class probabilities, and confidence scores. Finally, the network uses Non-maximum Suppression (NMS) to filter overlapping bounding boxes [11].

B. YOLOv8

YOLOv8 is the latest version of the object detection model architecture, succeeding YOLOv5. YOLOv8 introduces improvements in the form of a new neural network architecture [11]. Two neural networks are implemented, namely the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN), along with a new labeling tool that simplifies the annotation process. This labeling tool contains several useful features, such as automatic labeling, shortcut labeling, and customizable hotkeys. The combination of these features makes it easier to annotate images for training the model.

FPN works by gradually reducing the spatial resolution of the input image while increasing the number of feature channels. This results in the creation

of a feature map that is capable of detecting objects at different scales and resolutions. On the other hand, the PAN architecture can combine features from different network levels through skip connections. Consequently, the network can capture features more effectively at various scales and resolutions, which is crucial for accurately detecting objects of different sizes and shapes [12].

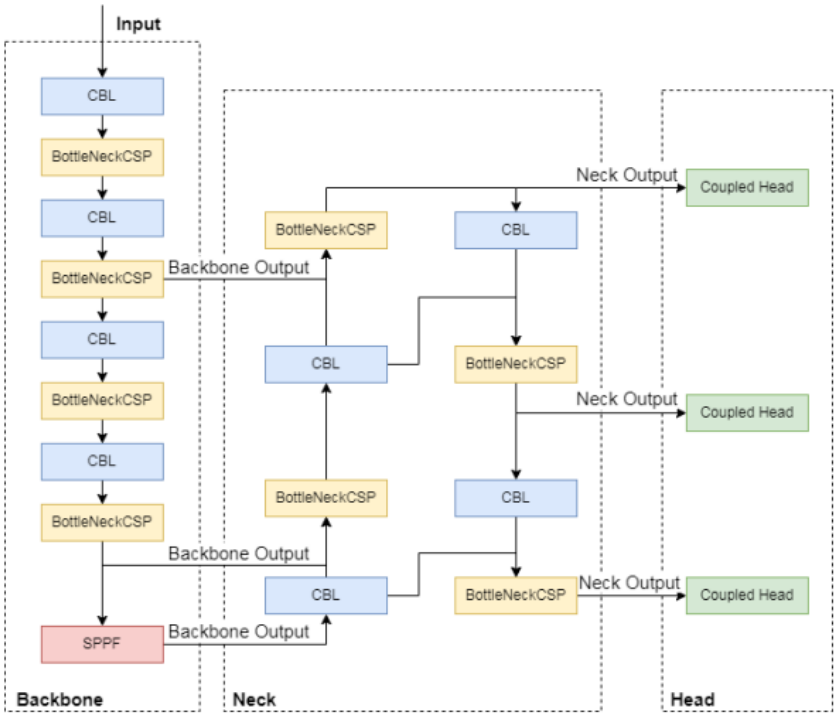

Figure 1. The structure of YOLOv5 [12]

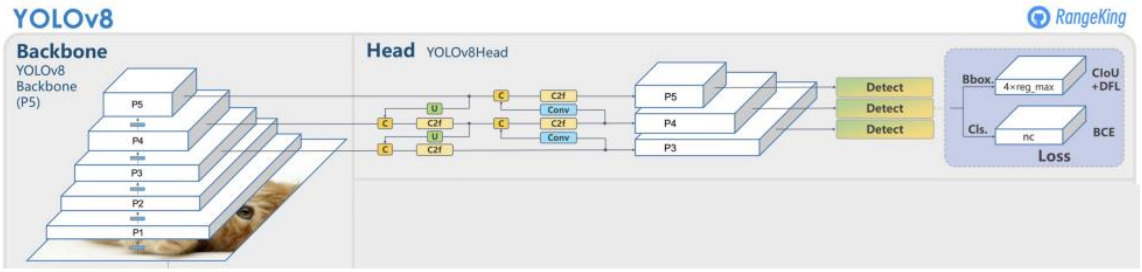

Figure 2. YOLOv8 Architecture[10]

C. *Evaluation Metrics*

Confusion matrix is used to evaluate the performance of a machine learning model. The confusion matrix is a matrix that displays the predictions of the actual classification and the predicted classification [13]. There are four classifications in the confusion matrix, namely True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) derived from actual and predicted values. The definitions of the confusion matrix are shown in Table 4, where TP (True Positive) is the number of positive samples correctly classified; TN (True Negative) is the number of negative samples correctly classified; FP (False Positive) is the number of negative samples wrongly classified as positive; FN (False Negative) is the number of positive samples wrongly classified as negative [14]. An illustration of the confusion matrix can be seen in Figure 2. The model's performance can be calculated using precision, recall, and F1-score derived from the confusion matrix.

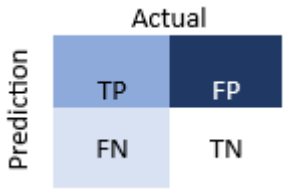

Figure 3. Confusion Matrix

Precision is the ratio of TP to the total number of predicted positive data. In the denominator, there is the variable FP as the divisor. This can be written using Equation 1[15].

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (1)$$

On the other hand, recall is defined as the ratio of TP to the total number of actually positive instances. The denominator includes FN as the divisor, and it can be written using Equation 2[15].

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

When recall is very high, precision will be very low, and vice versa. There is a trade-off relationship between precision and recall. This trade-off relationship implies that the sum of these two variables equals 1. The harmonization of the average between precision and recall is called the F1-score. Based on Equation 3[15], the best value for the F1-score is 1.0, while the worst value is 0.0.

$$\text{F1} = \frac{2\times\text{precision}\times\text{recall}}{\text{precision+recall}} \quad (3)$$

III. Results and Discussion

The aerial image data used in this paper is the pedestrian dataset from Roboflow. The dataset from https://universe.roboflow.com/edmundpub/pedestrian aerial. Our research used google collab for training and validate the model with 100th epoch. 828 aerial images were used in training the YOLOv5 and YOLOv8 for human detection. The models were validated using 233 aerial images in human detection.

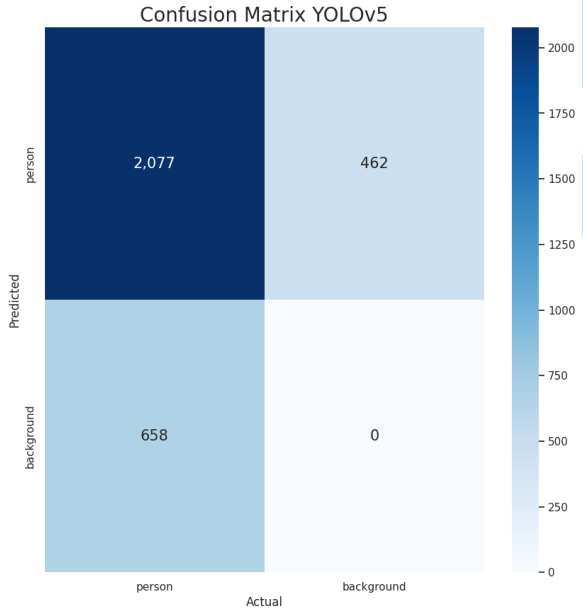

Figure 4. Confusion Matrix YOLOv5

The confusion matrix of YOLOv5 at the last epoch can be seen in Figure 4. The values of TP 2077, FP 463, and FN 658. The number of each value become from detected the human in aerial image from validation stage. Based on Eq.1, the precision value at the last epoch is 0.8177 or 81.77%. Recall value based on Equation 2 is 0.7594 or 75.41%. F1-score results obtained with Eq.3 are 0.7876 or 78.76%. These performance results are summarized in Table 1. While the precision, recall, and F1-Score values for each epoch are illustrated in Fig.6.

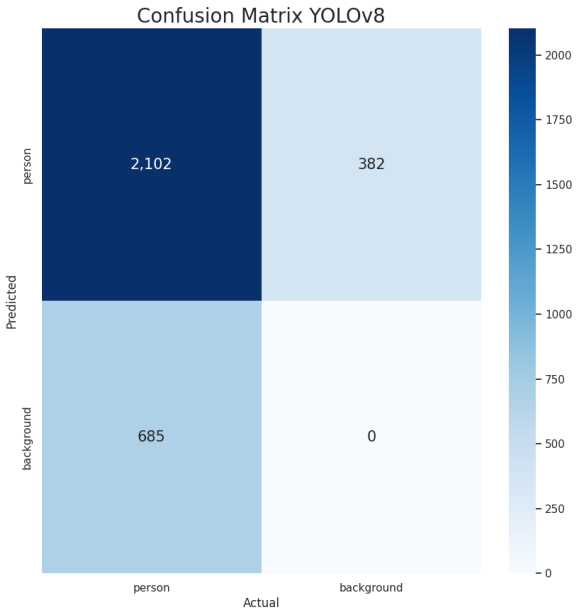

Figure 5. Confusion Matrix YOLOv8

The confusion matrix on YOLOv8 at the last epoch can be seen in Figure 5. TP values are 2102, FP 382, and FN 685. Based on Equation 1, the precission value at the last epoch is 0.8462 or 84.62%. Recall value based on Equation 2 is 0.7540 or 75.40%. F1-score results obtained with Equation 3 are 0.7998 or 79.98%. these performance results are summarized in Table 1. While the precision, recall, and F1-Score values for each epoch are illustrated in Fig 7.

TABLE I. YOLOV5 AND YOLOV8 MODEL PERFORMANCE

Architecture	Precision	Recall	F1-Score
YOLOv5	0.8180	0.7594	0.7876
YOLOv8	0.8462	0.7540	0.7974

The performance values of the YOLOv5 and YOLOv8 models are summarized in Table 1. The YOLOv8 model is 0.0282 or 2.82% larger than YOLOv5 for the precision value. The difference in the recall value of the YOLOv5 model is greater than that of YOLOv8 which is 0.0054 or 0.54%. The F1-score value of the YOLOv8 model is greater than that of YOLOv5 by 0.0098 or 0.98%.

The performance results of precision, Recall, and F1-score in graphical form can be seen in Fig.6 and Fig.7. The performance values for the YOLOv5 and YOLOv8 model graphs are known to increase until the last epoch. From the characteristics of the graphs in Figure 7 and Figure 8, it can be seen that the training process does not occur overfitting and underfitting. This shows that the learning process of deep learning has run without any problems.

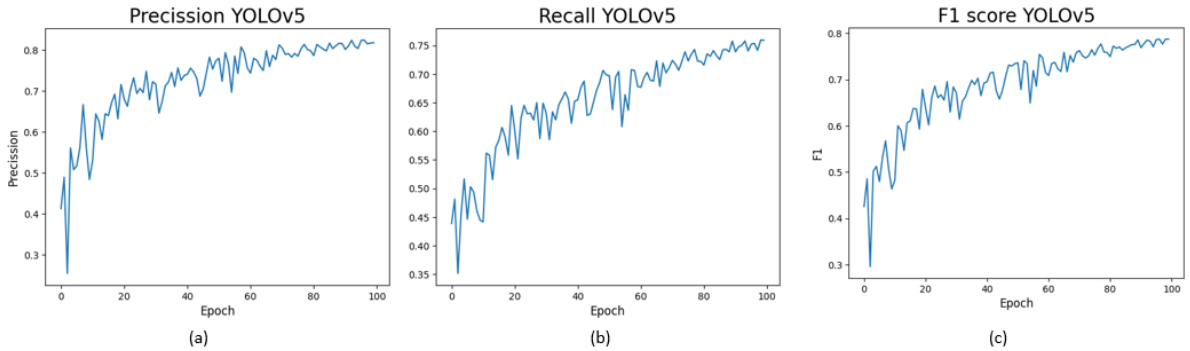

Figure 6. YOLOv5 Performance Results

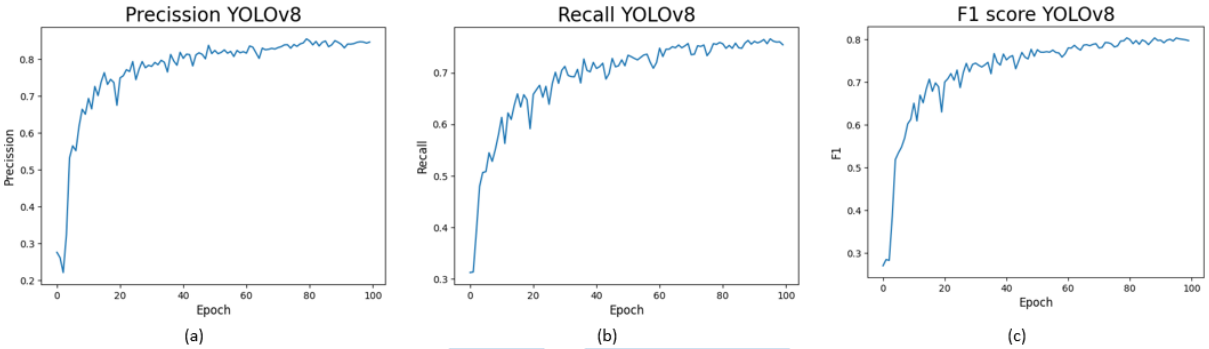

Figure 7. YOLOv8 Performance Results

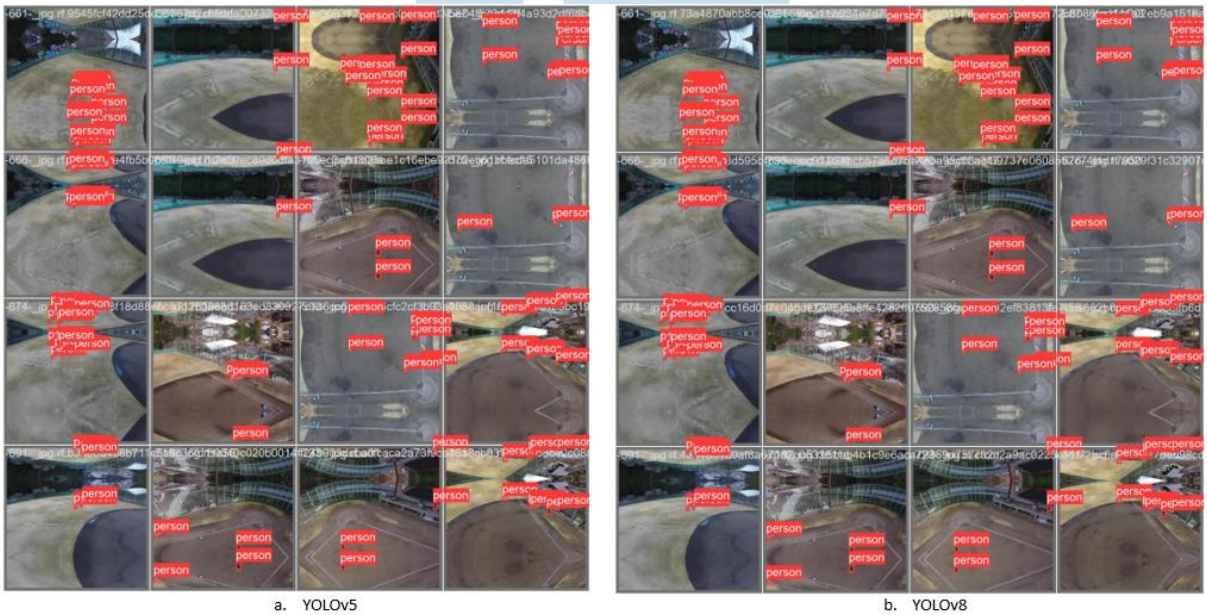

Figure 8. Human Detection Results

The results of aerial image training that has been applied to the YOLOv5 and YOLOv8 models can be seen in Figure 8. The person label in Figure 8 is the result of human detection. Both images show that the YOLOv5 and YOLOv8 models have successfully detected humans.

IV. CONCLUSIONS

Based on the research results discussed, it is known that the YOLOv5 and YOLOv8 models have successfully detected humans in aerial images. There are differences in performance values in human detection. The performance value of the YOLOv8 model is greater than the YOLOv5 model for precision and F1-score, the difference in the value of each performance is 2.82%, and 0.98%. As for the recall performance value, YOLOv5 is greater than the YOLOv8 model with a difference of 0.54%.

REFERENCES

[1] C. Symeonidis, I. Mademlis, I. Pitas, and N. Nikolaidis, "AUTH-PERSONS: A DATASET FOR DETECTING HUMANS IN CROWDS FROM AERIAL VIEWS," in Proceedings - International Conference on Image Processing, ICIP, IEEE Computer Society, 2022, pp. 596–600. doi: 10.1109/ICIP46576.2022.9897612.

[2] X. Zhang, Y. Feng, S. Zhang, N. Wang, and S. Mei, "Finding Nonrigid Tiny Person With Densely Cropped and Local Attention Object Detector Networks in Low-Altitude Aerial Images," IEEE J Sel Top Appl Earth Obs Remote Sens, vol. 15, pp. 4371–4385, 2022, doi: 10.1109/JSTARS.2022.3175498.

[3] H. Ugochi Dike, Q. Wu, Y. Zhou, and G. Liang, "Unmanned Aerial Vehicle (UAV) Based Running Person Detection from a Real-Time Moving Camera," in Proceedings of the 2018 IEEE International Conference on Robotics and Biomimetics, IEEE, 2018, pp. 2273–2278.

[4] A. G. Popa, L. Ichim, and D. Popescu, "Real-time person detection from UAV images using performant neural networks," in 2022 14th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ECAI54874.2022.9847477.

[5] S. Zhang et al., "Person Re-Identification in Aerial Imagery," IEEE Trans Multimedia, vol. 23, pp. 281–291, 2021, doi: 10.1109/TMM.2020.2977528.

[6] Q. Shen, L. Jiang, and H. Xiong, "Person Tracking and Frontal Face Capture with UAV," in 2018 18th IEEE International Conference on Communication Technology, IEEE, 2018, pp. 1412–1416.

[7] T. S. Gunawan, I. M. M. Ismail, M. Kartiwi, and N. Ismail, "Performance Comparison of Various YOLO Architectures on Object Detection of UAV Images," in 8th IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 257–261. doi: 10.1109/ICSIMA55652.2022.9928938.

[8] J. K. D. Lagman, A. B. Evangelista, and C. C. Paglinawan, "Unmanned Aerial Vehicle with Human Detection and People Counter Using YOLO v5 and Thermal Camera for Search Operations," in 2022 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2022 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 113–118. doi: 10.1109/I2CACIS54679.2022.9815490.

[9] J.-H. Kim, N. Kim, and C. S. Won, "High-Speed Drone Detection Based On Yolo-V8," in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Jun. 2023, pp. 1–2. doi: 10.1109/ICASSP49357.2023.10095516.

[10] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-Time Flying Object Detection with YOLOv8," ArXiv, vol. 1, no. 2305.09972, May 2023, [Online]. Available: http://arxiv.org/abs/2305.09972

[11] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond," ACM Comput Surv, Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.00501

[12] H. Liang, J. Chen, W. Xie, X. Yu, and W. Wu, "Defect detection of injection-molded parts based on improved-YOLOv5," in Journal of Physics: Conference Series, Institute of Physics, 2022. doi: 10.1088/1742-6596/2390/1/012049.

[13] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," Commun Stat Theory Methods, vol. 49, no. 9, pp. 2080–2093, May 2020, doi: 10.1080/03610926.2019.1568485.

[14] Y. Zhang, T. Zuo, L. Fang, J. Li, and Z. Xing, "An Improved MAHAKIL Oversampling Method for Imbalanced Dataset Classification," IEEE Access, vol. 9, pp. 16030–16040, 2021, doi: 10.1109/ACCESS.2020.3047741.

[15] Z. Ning, X. Wu, J. Yang, and Y. Yang, "MT-YOLOv5: Mobile terminal table detection model based on YOLOv5," in Journal of Physics: Conference Series, IOP Publishing Ltd, Jul. 2021. doi: 10.1088/1742-6596/1978/1/012010.