11TH INTERNATIONAL SCIENTIFIC CONFERENCE
ON DEFENSIVE TECHNOLOGIES
OTEH 2024
Serbia, Tara, 9-11 October 2024

COMPARATIVE ANALYSIS OF YOLO ALGORITHMS FOR AIRCRAFT DETECTION IN REMOTE SENSING IMAGES

TOUATI ADLI
University of Defence Belgrade, Military Academy, , adlitouati94@gmail.com

DIMITRIJE BUJAKOVIĆ
University of Defence Belgrade, Military Academy, dimitrijebujakovic@gmail.com

BOBAN BONDŽULIĆ
University of Defence Belgrade, Military Academy bondzulici@yahoo.com

MOHAMMED ZOUAOUI LAIDOUNI
University of Defence Belgrade, Military Academy, mohammedz.laidouni@gmail.com

MILENKO ANDRIĆ
University of Defence Belgrade, Military Academy, andricsmilenko@gmail.com

Abstract: *Accurate object detection in remote sensing images, particularly for military aircraft, is imperative for strategic decision-making. However, this task encounters numerous challenges, including diverse aircraft categories, their variable sizes and geometries, complex backgrounds, environmental factors (weather conditions, lighting variations, shadows) and sensor parameters. This paper presents a comparative study of three state-of-the-art detection algorithms, including YOLOv5, YOLOv7, and YOLOv8, and it aims to explore the strengths and limitations of each algorithm in addressing the above challenges. Considering the specific problem of military aircraft detection, the experimental results were conducted on the Military Aircraft Recognition dataset (MAR20). The obtained results demonstrate that YOLOv7 outperforms other algorithms in terms of detection performance, achieving a global mean average precision (mAP) and ($mAP_{0.5}$) by 67.7%, and 90.3% respectively.*

Keywords: *Aircraft detection; remote sensing images; YOLO*.

1. INTRODUCTION

Remote sensing target detection is a vital task in both civilian and military applications, including environmental monitoring, disaster management, urban planning, military surveillance, etc. [1]. The main objective of this task is to identify and categorize objects of interest in remote sensing images [2]. In the military field, aircraft pose significant strategic menaces, making their identification and localization critical factors that influence decision-making on the battlefield [2]. The aircraft targets detection remains a difficult task due to the several challenges it encounters [3-6], including, top-down view, aircraft types similarity, small and densely scattered objects, and target-background resemblance. Furthermore, a various external factor, including weather conditions, lighting, shadows, sensor parameters, etc., can significantly impact the accuracy of aircraft detection.

Object detection task in remote sensing image is carried out using traditional and deep learning techniques. Traditional approaches rely on handcrafted features and machine learning algorithms. In recent years, deep learning techniques have gained a significant attention, and become the predominant methods for classification and target detection tasks [5]. Particularly, the appearance of convolutional neural networks (CNNs) [1]. The power of deep learning methods lies in automatically extracting and learning higher-level semantic features and providing a stronger feature representation capability unlike traditional machine learning methods that are based on handcrafted features [1].

Target detection algorithms can be divided into two main groups [1, 5, 7, 8]: single-stage target detection algorithms, and two-stage target detection algorithms. The single-stage algorithms, including You Look Only Once [9] (YOLO), Fully Convolutional One-Stage Object Detection (FCOS) [10], Single Shot MultiBox Detector (SSD) [11], Deconvolutional Single Shot Detector (DSSD) [12] and RetinaNet [13] are designed to detect a target within images by combining regression and classification tasks [5]. These models generally use a single feedforward step in order to create bounding boxes and class probabilities directly from entire image. The two stage target detection algorithms such as regions with convolutional neural networks (R-CNN) [14], Fast R-CNN [15], Faster R-CNN [16], Cascade R-CNN [17] and spatial pyramid pooling network (SPP-Net) [18] are based on generated collection of regions of interest (RoIs) from

Corresponding author: T. Adli;
E-mail: adlitouati94@gmail.com
DOI: 10.5937/OTEH24059A

image, to identify and classify the objects based on these regions [19]. Two-stage target detection algorithms show superior detection accuracy and slower detection speed [1, 5, 6]. In contrast, single-stage algorithms are generally faster and sacrifice detection accuracy, making them useful for applications where detection speed is critical.

This paper presents a comparative study for aircraft detection in remote sensing images of three state-of-the-art single-stage target detection algorithms, including YOLOv5 [20], YOLOv7 [21] and YOLOv8 [22]. The experimental results were conducted on the Military Aircraft Recognition (MAR20) [2] dataset, which the largest publicly available dataset for military aircraft recognition in remote sensing images. This paper is structured as follows: Section 2 provides an overview of YOLOv5, YOLOv7, and YOLOv8 algorithms. Following that, the obtained experimental results were presented and discussed in Sections 3 and 4, and in the last section is given study conclusion.

2. OVERVIEW OF THE YOLO ALGORITHMS

2.1. YOLOv5

YOLOv5 [20], introduced in 2020, is a real-time, one stage detection algorithm that prioritizes detection speed over accuracy through the utilization of a more compact network architecture. This strategy achieves a high efficiency in detection task [23, 24]. YOLOv5 encompasses five variants, namely, YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. The YOLOv5 network architecture comprises four main components, which are: Input, Backbone, Neck and Prediction.

Before being fed into the network, the input image undergoes preprocessing to improve detection accuracy. The Backbone part extracts feature from input images, while the Neck component combines features from various detection layers of the Backbone using the PANet [25] strategy. The Prediction part employs a complete intersection over union (CIoU) loss function to improves the speed and accuracy of detection. Additionally, it utilizes non-maximum suppression (NMS) [26] to detect multiple and occluded objects in the input image.

2.2. YOLOv7

The YOLOv7 [21] represents a cutting-edge one-stage real-time object detection algorithm. It demonstrates a favorable compromise between detection speed and accuracy, surpassing the previous YOLO series in accuracy and speed. An outstanding novelty within YOLOv7 is the introduction of the E-ELAN methodology, which effectively guides computational blocks to capture a diverse range of features. This approach integrates convolutional layers followed by detection layers to predict bounding boxes and object classes, using methods like expansion and shuffling to improve learning dynamics while preserving gradient flow integrity.

2.3. YOLOv8

YOLOv8 [22], is a real-time object detection algorithm, building upon the success of its predecessors, while integrating new features and improvements to make it faster, more accurate, and easier to use. YOLOv8 can be used in many tasks, including object detection, tracking, segmentation, image classification, and posture estimation. YOLOv8 comes in five variants, YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x.

3. EXPERIMENTS

In this section, we first present the dataset used in the experiments, called MAR20 dataset. Following this, it is provided an explanation of the used evaluation metrics and implementation details. The experiments primarily involve a comparative study of three detection algorithms, including YOLOv5m, YOLOv7 and YOLOv8m.

3.1. MAR20 Dataset

MAR20 [2] dataset is the largest public dataset for remote sensing image military aircraft recognition. It comprises 3842 high-resolution remote sensing images encompassing 22341 instances of 20 distinct aircraft types, labeled from A1 to A20 (refer to Fig. 1). These images were collected using Google Earth in 60 military airports. The majority of images in the dataset have a resolution of 800x800 pixels. Two annotation methods are given: horizontal bounding box and oriented bounding box for each target instance. Table 1 illustrates the distribution of instances for each aircraft type in the dataset.

Figure 1. The twenty aircraft types in MAR20 dataset.

In our experiments, the original training set, comprising 1331 images (7870 instances), was divided into training and validation subsets, encompassing 90% (1198 images) and 10% (133 images), respectively, and the whole test set was used for testing. The horizontal bounding box annotations were utilized throughout these experiments.

Table 1. Distribution of instances in train and test sets for each aircraft type.

Aircraft	Train	Test	All	Aircraft	Train	Test	All
A1	554	1092	1646	A11	151	356	507
A2	736	993	1729	A12	240	462	702
A3	316	860	1176	A13	561	1091	1652
A4	179	463	642	A14	659	1119	1778
A5	435	827	1262	A15	200	418	618
A6	107	334	441	A16	785	1847	2632
A7	191	489	680	A17	439	958	1397
A8	365	579	944	A18	160	148	308
A9	475	611	1086	A19	519	717	1236
A10	368	556	924	A20	430	551	981

3.2. Evaluation Metrics

For the evaluation of the object detection algorithms, we employ commonly used evaluation metrics, including mean average precision (mAP) averaged for intersection over union IoU∊{0.5, 0.55, 0.6, …, 0.95} (COCO's standard metric), mean average precision ($mAP_{0.5}$) at IoU=0.5 (PASCAL VOC's metric) [16], precision (P), and recall (R).

The IoU is computed by determining the overlapping area between the predicted region (A) and the actual ground truth (B):

$$IoU = \frac{A \cap B}{A \cap B} \tag{1}$$

where A∩B indicate the intersection area of the predicted and ground truth bounding boxes and A∪B is their union.

The expressions for P and R are defined as follows:

$$\mathrm{P} = \frac{TP}{TP + FP}, \ \mathrm{R} = \frac{TP}{TP + FN} \tag{2}$$

where, *TP* (true positives) refers to the number of positive samples correctly identified, *FP* (false positives) represents the number of negative samples erroneously identified as positive samples, and *FN* (false negatives) denotes the number of positive samples were incorrectly identified as negative samples.

The formulas of $mAP_{0.5}$ and mAP, are given by:

$$\mathrm{mAP}_{0.5} = \frac{1}{N}\sum_{i=1}^{N} AP_i,\ IoU = 0.5 \tag{3}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i,\ IoU = 0.5 : 0.05 : 0.95 \tag{4}$$

where N is the number of categories, and AP_i is the average precision of each *i*-th category of objects. It represents the area under precision-recall curve and it can be calculated by:

$$AP = \int_0^1 P(R)dR \tag{5}$$

3.3. Implementation Details

The experimental platform runs on Windows 11, within an Intel® Core™ i7-12650H/16G @2.30GHz, equipped with an NVIDIA RTX 3050Ti/4GB. Python 3.9 and Pytorch framework (version 2.0.1) were used for implementation of algorithms, and CUDA 11.8 was utilized to accelerate the calculation.

In the training phase the Stochastic Gradient Descent (SGD) algorithm was utilized to optimize the loss function. Specifically, a momentum of 0.937 and a weight decay coefficient of 0.0005 were used. To enhance training stability, a warmup strategy was implemented. The initial learning rate was set at 0.01, with a batch size of 2 and the number of epochs was set at 300. The resolution of input image was maintained at 640x640 pixels.

3.4. Training results

Fig. 2 illustrates the variations of performance metrics, including $mAP_{0.5}$ (Fig. 2 (a)), mAP (Fig. 2 (b)), recall (Fig. 2 (c)) and precision (Fig. 2 (d)) of YOLOv5m, YOLOv7 and YOLOv8m. These metrics measures the performance of object detection algorithms. The higher values indicative of improved accuracy in detection capabilities. YOLOv8m shows better training speed and efficiency compared to YOLOv5m and YOLOv7. However, YOLOv7 outperforms the others in terms of overall performance, particularly in $mAP_{0.5}$ and recall.

4. EXPERIMENTAL RESULTS

A comparative evaluation was conducted between the YOLOv5m, YOLOv7, and YOLOv8m algorithms on the MAR20 dataset under the same configurations. Table 2 illustrates the comparative performance metrics for the three detection algorithms. These findings clearly demonstrate that YOLOv7 surpasses YOLOv5m and YOLOv8m across almost all metrics (precision, recall, $mAP_{0.5}$, and mAP).

Table 2. Aircraft detection results

Algorithm	P (%)	R (%)	$mAP_{0.5}$ (%)	mAP (%)	Inference time (ms)
YOLOv5m	82.2	82.2	86.9	64.5	**17.3**
YOLOv7	**85.6**	**84.2**	**90.3**	**67.7**	19.3
YOLOv8m	83.4	78.5	86.1	65.2	24.1

The YOLOv7 achieves $mAP_{0.5}$ and mAP scores of 90.3% and 67.7% respectively, outperforming YOLOv5m by 3.4% and 3.2% respectively, and YOLOv8m by 4.2% and 2.5%.

Furthermore, YOLOv7 exhibits exceptional precision in distinguishing aircraft features, reducing recognition errors, and attaining a recall rate of 84.2%, which is 2%

and 5.7% higher than YOLOv5m and YOLOv8m, respectively.

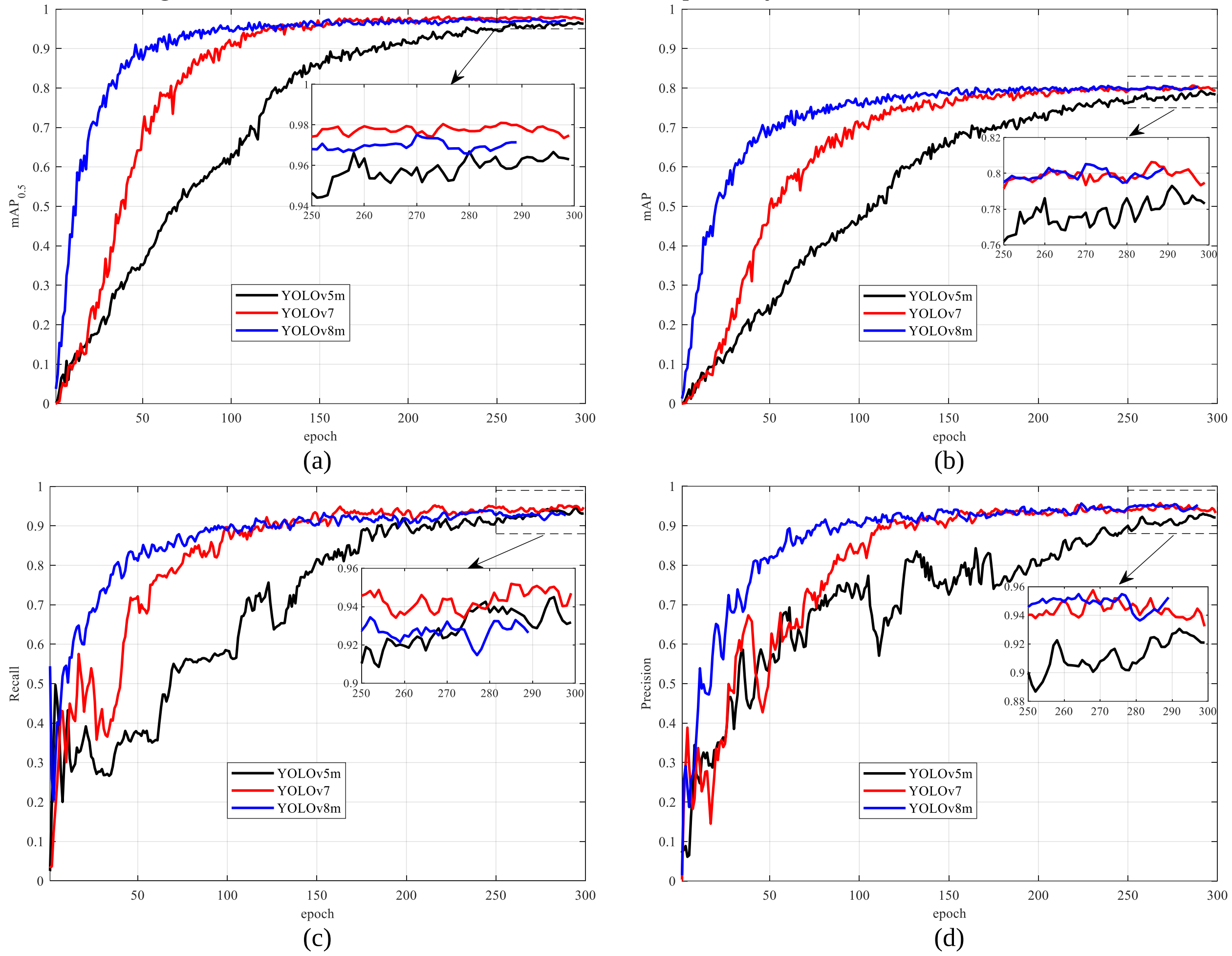

Figure 2. Performance curve: (a) $mAP_{0.5}$, (b) mAP, (c) recall, and (d) precision.

While the YOLOv5m network architecture is comparatively smaller, resulting in a reduced inference time of 17.3 ms against 19.3 ms for YOLOv7 and 24.1 ms for YOLOv8m.

4.1. Detection Results By-Class

For a more comprehensive comparative analysis, Table 3 presents the detection performance metrics, including mAP, $mAP_{0.5}$, precision, and recall by classes, for the YOLOv5m, YOLOv7, and YOLOv8m detection algorithms. This detailed breakdown offers a clearer view of the strengths and limitations of each algorithm.

The analysis of the Table 3 shows that YOLOv7 algorithm demonstrates superior overall performance compared to YOLOv5m and YOLOv8m. It achieves the mAP exceeds or closely approaches 67%, except for aircraft types A5 (52.8%), A15 (40.8%), A13 (60.3.4%), A19 (62.1%) and A20 (62.7%), which exhibit lower detection rates. The observed low detection rate observed for aircraft types A15 and A13 across all algorithms can be attributed to their high similarity with other aircraft types (such as A5 and A16), as well as their relatively smaller size compared to other types. This similarity reduces the discriminative capacity of the features extracted by the network, making it challenging to differentiate and detect these aircraft types.

Despite the lower instances number of aircraft A18 compared to other aircraft types, in addition the high inter-class similarity with aircraft A4, the YOLOv7 demonstrates high detection performance compared to other algorithms. It achieves a mAP of 64.8.5%, $mAP_{0.5}$ of 80.4%, precision of 87%, and recall of 70.3%.

Fig. 3 presents detection results using YOLOv5m, YOLOv7 and YOLOv8m algorithms on the same images extracted from MAR20 test set, including four cases: successful detection, Fig. 3 (a), small object detection, Fig. 3 (b), false alarm, Fig. 3 (c), and target-background resemblance Fig. 3 (d).

In Fig. 3 (a) it is worth noting that all algorithms successfully detected and recognized correctly the aircraft type A5 instances, where YOLOv7 demonstrate the best overall confidence value.

Fig. 3 (b) shows the detection results of small objects. YOLOv5m and YOLOv8m misidentified the aircraft A5 instances and identified them as A15 (the images marked with red boxes contain misidentification), due to the high inter-class similarity between them. However, YOLOv7 successfully detects and identifies all instances with a high confidence value. In Fig. 3 (c) YOLOv7 successfully detect and identify all instances with better confidence value. However, YOLOv5 and YOLOv8m algorithms show false alarm detection.

Table 3. Detection results of YOLOv5m, YOLOv7 and YOLOv8m algorithms on various classes of MAR20 dataset.

	YOLOv5m				YOLOv7				YOLOv8m			
	P (%)	R (%)	$mAP_{0.5}$ (%)	mAP (%)	P (%)	R (%)	$mAP_{0.5}$ (%)	mAP (%)	P (%)	R (%)	$mAP_{0.5}$ (%)	mAP (%)
A1	**75.1**	76.6	82.1	65.0	75.0	**84.1**	**85.5**	**68.7**	71.5	82.0	81.0	65.1
A2	98.6	84.8	95.5	69.1	**98.9**	**89.5**	**98.3**	**71.4**	98.3	77.6	92.4	67.9
A3	89.6	87.7	95.6	67.6	91.5	**91.7**	**97.2**	**70.0**	**93.4**	84.3	95.4	68.1
A4	76.2	**89.2**	87.4	65.8	83.4	83.6	**91.7**	**71.1**	**85.0**	84.4	89.7	67.0
A5	66.7	79.9	75.2	49.4	64.6	**83.1**	**80.8**	**52.8**	**66.9**	80.4	76.8	51.0
A6	93.7	**89.2**	94.9	71.7	96.2	86.2	**95.1**	**74.0**	**96.3**	84.1	94.3	72.6
A7	81.3	92.6	93.5	74.2	90.7	**94.9**	**97.6**	**78.7**	**90.8**	91.3	96.7	78.4
A8	86.1	94.6	94.4	73.6	85.8	**94.8**	**96.1**	**74.4**	**87.0**	91.4	94.1	73.5
A9	**89.5**	79.9	91.7	67.3	94.8	**85.8**	**94.2**	**69.6**	86.8	75.4	87.9	65.3
A10	**95.2**	93.9	98.1	70.4	93.0	**95.9**	**98.4**	**71.6**	92.8	91.2	96.3	70.4
A11	67.5	74.2	72.2	55.3	**72.9**	**85.2**	**85.7**	**67.3**	72.6	81.5	82.6	65.4
A12	83.9	85.7	92.6	70.9	**86.7**	**86.1**	**93.8**	**72.0**	83.7	78.0	89.4	69.1
A13	73.5	66.9	78.2	55.0	81.6	**74.6**	**85.6**	**60.3**	**84.4**	58.7	79.8	57.9
A14	92.9	83.8	93.6	71.5	**93.1**	**85.4**	**94.8**	**72.6**	92.2	79.3	93.1	71.4
A15	53.0	**59.1**	58.3	38.0	**59.9**	56.5	**63.5**	**40.8**	53.5	49.3	51.0	33.9
A16	**98.0**	**80.7**	**95.3**	71.6	97.2	78.1	95.1	72.0	97.3	71.9	94.4	**72.9**
A17	96.5	95.0	97.7	76.6	95.7	**95.1**	**97.9**	**77.9**	**97.2**	91.6	96.7	76.6
A18	72.3	**78.4**	76.2	59.5	**87.0**	70.3	**80.4**	**64.8**	63.5	69.3	66.5	54.7
A19	73.3	70.9	80.2	58.6	75.1	**79.1**	**84.7**	**62.1**	**76.6**	69.5	81.6	61.2
A20	81.5	80.9	84.8	59.5	**89.4**	**84.6**	**90.5**	**62.7**	78.0	78.8	82.2	61.2
All	82.2	82.2	86.9	64.5	**85.6**	**84.2**	**90.3**	**67.7**	83.4	78.5	86.1	65.2

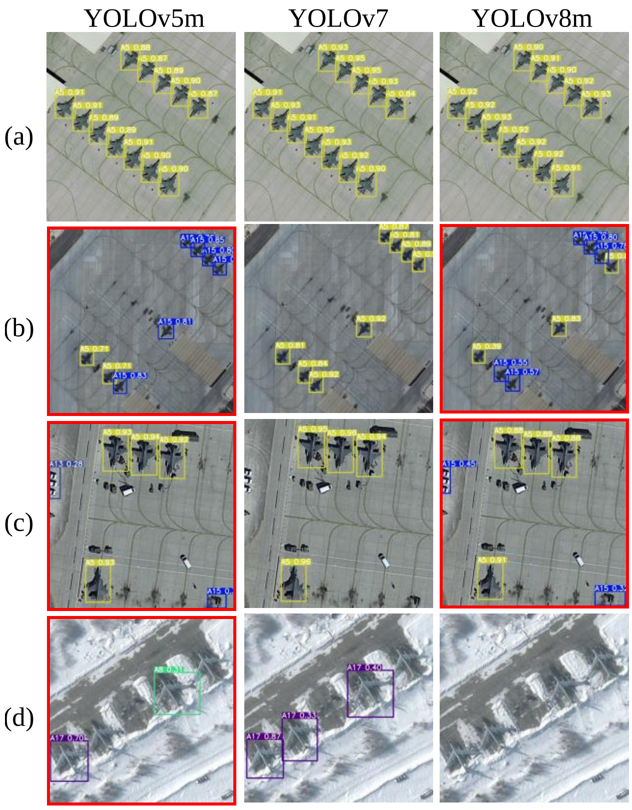

Figure 3. Detection results of YOLOv5m, YOLOv7, YOLOv8m on the same images (a, b, c, and d) extracted from test set of MAR20 dataset.

Detection results in Fig. 3 (d) presents the case of target-background resemblance, YOLOv5m shows misidentification and cannot detect all instances, and YOLOv8 failed to detect all target instances within image. YOLOv7 demonstrate a high discrimination performance and successfully detects almost all instances within image outperforming YOLOv5m and YOLOv8m.

5. CONCLUSION

A comparative study on aircraft detection in remote sensing images was conducted to evaluate three state-of-the-art detection algorithms, YOLOv5, YOLOv7, and YOLOv8, using MAR20 dataset. The experimental results demonstrate that YOLOv7 exhibits a superior overall detection performance compared to YOLOv5, and YOLOv8 algorithms. YOLOv7 achieves $mAP_{0.5}$ and mAP scores of 90.3% and 67.7% respectively, which are higher than YOLOv5m by 3.4% and 3.2% respectively, and higher than YOLOv8m by 4.2% and 2.5%. Moreover, YOLOv7 demonstrate better distinguishing between aircraft features, thus reducing recognition errors with a recall rate of 84.2%. Additionally shows a faster learning capacity compared to other algorithms in situation of lower aircraft instances number.

References

[1] Li Z, Yuan J, Li G, Wang H, Li X, Li D, Wang X. RSI-YOLO: object detection method for remote sensing images based on improved YOLO. Sensors. 2023; 23(14): 6414. https://doi.org/10.3390/s23146414.

[2] Yu W, Cheng G, Wang M, Yao Y, Xie X, Yao X, Han J. MAR20: a benchmark for military aircraft recognition in remote sensing images. Journal of Remote Sensing (Chinese). 2022. https://doi.org/10.11834/jrs.20222139.

[3] Zuo J, Xu, G, Fu K, Sun X, Sun H. Aircraft type recognition based on segmentation with deep convolutional neural networks. IEEE Geoscience

and Remote Sensing Letters. 2018; 15(2): 282-286. https://doi.org/10.1109/LGRS.2017.2786232.

[4] Zhang L, Zhang Y. Airport detection and aircraft recognition based on two-layer saliency model in high spatial resolution remote-sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2017; 10(4): 1511-1524. https://doi.org/10.1109/JSTARS.2016.2620900.

[5] Zhou F, Deng H, Xu Q, Lan X. CNTR-YOLO: improved YOLOv5 based on ConvNext and transformer for aircraft detection in remote sensing images. Electronics. 2023; 12(12): 2671. https://doi.org/10.3390/electronics12122671.

[6] Liu Z, Gao Y, Du Q, Chen M, Lv W. YOLO-Extract: improved YOLOv5 for aircraft object detection in remote sensing images. IEEE Access. 2023; 11: 1742-1751. https://doi.org/10.1109/ACCESS.2023.3233964.

[7] Cao X, Zhang Y, Lang S, Gong Y. Swin-transformer-based YOLOv5 for small-object detection in remote sensing images. Sensors. 2023; 23(7): 3634. https://doi.org/10.3390/s23073634.

[8] Ma J, Wang X, Xu C, Ling J. SF-YOLOv5: improved YOLOv5 with swin transformer and fusion-concat method for multi-uav detection. Measurement and Control. 2023; 56(7-8): 1436-1445. https://doi.org/10.1177/00202940231164126.

[9] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; Las Vegas, NV, USA; June 27-30,2016. p. 779-788. https://doi.org/10.1109/CVPR.2016.91.

[10] Tian Z, Shen C, Chen H, He T. FCOS: fully convolutional one-stage object detection. IEEE/CVF International Conference on Computer Vision (ICCV). 2019; Seoul, Korea (South); October 27-02 November, 2019. p. 9626-9635. https://doi.org/10.1109/ICCV.2019.00972.

[11] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: single shot multibox detector. European Conference on Computer Vision (ECCV). 2016; 14th European Conference, Amsterdam, The Netherlands; October 11-14, 2016. p. 21-37. https://doi.org/10.1007/978-3-319-46448-0_2.

[12] Fu CY, Liu W, Ranga A, Tyagi A, Berg AC. DSSD: deconvolutional single shot detector. arXiv:1701.06659v1 [Preprint]. 2017 [accessed 19 October 2023]: [11 p.]. Available from: https://doi.org/10.48550/arXiv.1701.06659.

[13] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. IEEE International Conference on Computer Vision (ICCV). 2017; Venice, Italy; October 22-29 2017. p. 2999-3007. https://doi.org/10.1109/ICCV.2017.324.

[14] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014; Columbus, OH, USA; June 23-28; 2014. p. 580-587. https://doi.org/10.1109/CVPR.2014.81.

[15] Girshick R. Fast R-CNN. IEEE International Conference on Computer Vision (ICCV). 2015; Santiago, Chile; December 07-13, 2015. p. 1440-1448. https://doi.org/10.1109/ICCV.2015.169.

[16] 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. https://doi.org/10.1109/TPAMI.2016.2577031.

[17] Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018; Salt Lake City, UT, USA; 18-23 June, 2018. p. 6154-6162. https://doi.org/10.1109/CVPR.2018.00644.

[18] Purkait P, Zhao C, Zach C. SPP-Net: deep absolute pose regression with synthetic views. British Machine Vision Conference (BMVC). 2018. Available from: https://doi.org/10.48550/arXiv.1712.03452.

[19] Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learn. Syst. 2019; 30(11): 3212-3232. https://doi.org/10.1109/TNNLS.2018.2876865.

[20] Jocher G. YOLOv5. Available online. 2023. https://github.com/ultralytics/yolov5. accessed 10 September 2023.

[21] Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2023; Vancouver, BC, Canada, June 17-24, 2023. p. 7464-7475. https://doi.org/10.1109/CVPR52729.2023.00721.

[22] Jocher G, Chaurasia A, Qiu J. YOLOv8. Available online. 2023, https://github.com/ultralytics/ultralytics. accessed 25 September 2023.

[23] Bai C, Bai X, Wu K. A review: remote sensing image object detection algorithm based on deep learning. Electronics. 2023; 12(24):4902. https://doi.org/10.3390/electronics12244902.

[24] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015; 37(9): 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824.

[25] Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018; Salt Lake City, UT, USA, June 18-23, 2018. p. 8759-8768. https://doi.org/10.1109/CVPR.2018.00913.

[26] Neubeck A, Van Gool L. Efficient non-maximum suppression. International Conference on Pattern Recognition (ICPR'06). 2006; Hong Kong, China; August 20-24, 2006; 3: p. 850-855. https://doi.org/10.1109/ICPR.2006.479.