Research on Facial Recognition of Drones Based on YOLO Visual Model

Shuai Fang[1,*], Yunong Li[2], Hanshuo Zhang[3]

[1] Detroit Green Technology Institute, Hubei University of Technology, Wuhan, 430000, China
[2] Institute of Intelligent Engineering, Shenyang City University, Shenyang, 110000, China
[3] Suzhou Science&Technology Town Foreign Language High School, Suzhou, 215000, China
*a900082@correo.umm.edu.mx

Abstract. In today's society, with the vigorous development of various undertakings, drones have been widely applied in numerous industry fields. This trend has driven the rapid evolution of drone systems to adapt to diverse application scenarios. However, existing technologies and systems have some minor issues that need to be addressed. By introducing advanced facial recognition technology, we aim to endow related devices with the ability to accurately identify, interpret and execute instructions, thereby effectively liberating labor. Therefore, based on the YOLOv8s visual model and using the WIDERFACE dataset as the training set, we have developed a system composed of four parts: data input layer, feature extraction backbone, feature fusion neck, and detection output head. In the feature extraction backbone part, we have conducted research on facial recognition and applied it to drones. Through relevant research, the experimental results show that the model still demonstrates excellent facial recognition capabilities in various backgrounds and complex lighting conditions, achieving high accuracy levels in many scenarios.

Keywords：YOLOV8S, Face Recognition, Drones

1 Introduction

With the development of society and the widespread use of drones in various industries, drone systems are rapidly evolving to adapt to various application fields. With the rapid development of technologies such as artificial intelligence, big data, cloud computing, and 5G communication, unmanned aerial vehicle systems have become more intelligent and autonomous [1]. At the same time, with the development of computer technology, the application of information technology has penetrated into various aspects of life. In some specific scenarios, we need to analyze the unique characteristics of humans. On the market, methods such as convolutional neural networks, recurrent neural networks, and long short-term memory networks have basically achieved face recognition, but each method has some problems, and the recognition accuracy of each method cannot meet the standard for identifying

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_87

people. We hope to use facial recognition to enable relevant equipment to accurately recognize, interpret, and execute commands, thereby freeing up labor.

So far, relevant research on facial recognition methods mainly includes deep learning based facial detection methods, such as the robust and accurate MTCNN algorithm, which belongs to the cascaded CNN method; Method based on single-stage series: The fast YOLO algorithm based on two-stage can better recognize faces, but the speed is not ideal. The most representative is the Fast RCNN algorithm [2]. Moreover, with the application of deep convolutional neural networks, models such as ArcFace MagFace and AdaFace have significantly improved the ability to recognize faces in different scenarios [3-5]. In 2015, Deep ID algorithms such as VGGNet, which share similarities with traditional and well-developed convolutional neural networks, emerged, followed by the continuous emergence of DeepID2, DeepID2+, and DeepID3 over time. But the learning level of these methods is not deep enough, and their recognition rate is not easy to improve. Moreover, the detection speed and training time cannot meet the needs of today's society. Later on, Google also developed a facial recognition model based on convolutional neural system - FaceNet model, which achieved end-to-end matching by classifying the features extracted from the initial image through deep learning without processing [2]. However, it could not effectively handle images affected by external factors to ensure recognition accuracy.

Regarding the description of the current situation in the above background, many methodological issues have arisen. Therefore, we propose the following questions: Can we find a method to use drones instead of humans for in-depth reconnaissance and face recognition to achieve application. Can we ensure that external influences can be eliminated in any environment to better recognize faces and improve accuracy.

For the significance, there are several points as follows: Firstly, a model called YOLOv8s is proposed that can be applied to drones, is easy to learn, and can form a good learning outcome as much as possible. Secondly, modules like C2f can be added to improve performance and reduce learning time [3]. By applying the above methods to improve the relevant facial recognition system, it becomes more practical and provides a new idea for the improvement of facial recognition technology in the future.

2 Methods and materials

2.1 Overall Technical Proposal

For this project, we adopted YOLO's method to achieve more accurate facial recognition. We adopt the most basic and traditional detection method for YOLO, as shown in Figure 1. Overall, it requires several major steps to achieve, including network structure design, image segmentation and prediction, loss function calculation, and non maximum suppression. Among them, image segmentation and prediction, loss function calculation, and non maximum suppression are key.

In image partitioning and prediction, grid partitioning refers to dividing the input image into grids. If the center of the target falls within a certain grid, that grid is responsible for detecting the target. For category prediction: In addition to bounding box prediction, each grid also needs to predict the category of the target [6]. The network will output the probability of each possible category in each grid, and by analyzing these probabilities, it can determine which category the target belongs to. And for confidence prediction, each bounding box will also have a confidence score to indicate whether the target is included in the bounding box and the accuracy of the prediction. The calculation of confidence score is usually based on factors such as the degree of overlap between the bounding box and the real target box, as well as the accuracy of category prediction.

In the calculation of loss functions, coordinate loss, confidence loss, and category loss are mainly calculated using mean square error, binary cross entropy loss, and multi class cross entropy loss. In non maximum suppression, preliminary screening is performed based on the confidence threshold, filtering out bounding boxes with confidence scores below the threshold and retaining bounding boxes with higher confidence. For the retained bounding boxes, calculate their overlap, sort them in descending order of confidence score, and then traverse each bounding box in sequence. For the current bounding box, other bounding boxes with an overlap greater than a certain threshold will be suppressed, that is, these bounding boxes will be deleted, and only the bounding boxes with the highest confidence and overlap not exceeding the threshold will be retained as the final detection result.

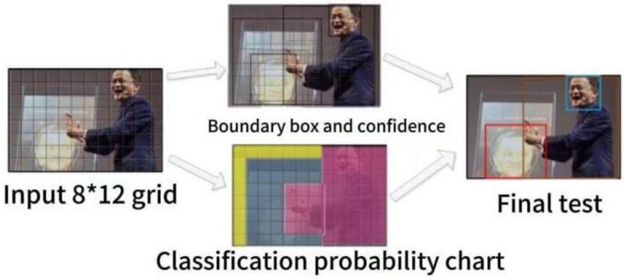

Fig. 1. YOLO detection method [2]

Figure 2 shows the YOLO network architecture model. In this model, in order to improve the accuracy of face detection, we first filter based on the confidence scores of different categories of each bounding box, removing those bounding boxes with lower confidence [7]. Next, non maximum suppression processing will be applied to the remaining bounding boxes. This process can effectively reduce redundant bounding boxes and ultimately accurately locate the correct face detection box in the image. Through these steps, we not only improved the efficiency of detection, but also significantly enhanced the reliability of the detection results [2].

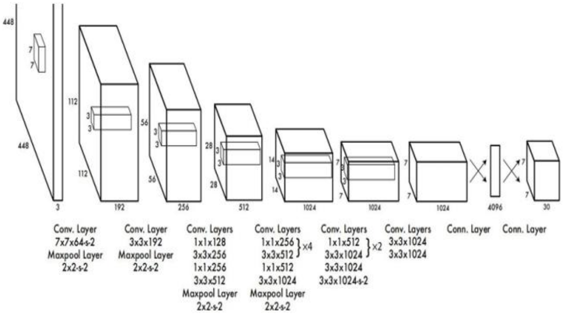

Fig. 2. YOLO Network Structure Model [2]

2.2 Principle of Experimental Simulation Design Process

We use the WIDERFACE dataset as the training set for the YOLO model. The WIDER FACE dataset was constructed by The Chinese University of Hong Kong and is currently one of the most authoritative face detection datasets in the industry. The simulation design process is shown in Figure 3.

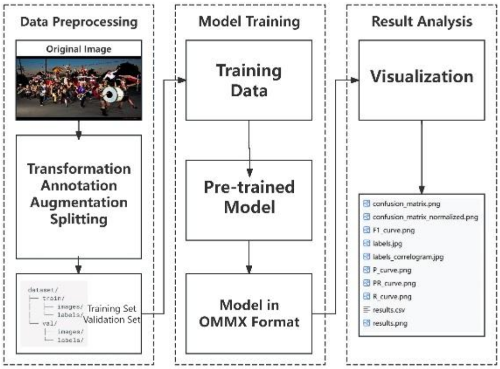

Fig. 3. Schematic diagram of simulation design process

2.3 Introduction to Methods of Application

The overall architecture of the system is shown in Figure 4, mainly composed of four parts: data input layer, feature extraction backbone, feature fusion neck, and detection output head. In the feature extraction backbone, Conv module, C2f module, and SPPF module are used to construct a feature extractor, which is responsible for capturing the salient features of the target from the input data. The feature fusion neck is composed of C2f module, upsampling component, and feature stacking unit. Its main function is to integrate information from multi-scale feature maps to obtain richer feature representations. After feature fusion, the information will be transmitted to the detection output head, which is composed of multiple convolutional layers and is responsible for performing target category determination and position prediction tasks [8].

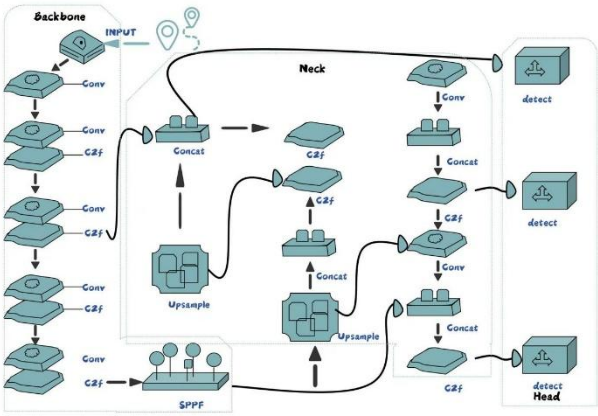

Fig. 4. Self made YOLOv8s Lightweight Structure Diagram

2.4 Feasibility of Evaluation Methods

This method has a high feasibility because our material can be trained using some original images in open source databases online to form a better model for implementation, and some pre trained models in the past have references that can be modified through reference code for optimization, achieving better results[6].Moreover, because this method can continuously improve the recognition of faces by drones, through the improvements of previous generations, we can find the patterns inside, use these details to guide us, and thus come up with improvement directions and ideas, which can improve the recognition accuracy of the model on the basis of previous generations[9]. Meanwhile, the use of this method is quite suitable for our level, providing us with operational space.

3 Results and discussion

3.1 Simulation results

The confusion matrix for figure 5 displays the classification results of the model in the "face" and "background" categories. The model accurately predicted 15331 "faces", but also mistakenly identified 2393 "backgrounds" as "faces".

The significance of confusion matrix lies in its ability to help us intuitively understand the classification performance of the model and the specific situation of misclassification. Through these data, we can optimize the model in a targeted manner, reducing the occurrence of misjudgments and omissions.

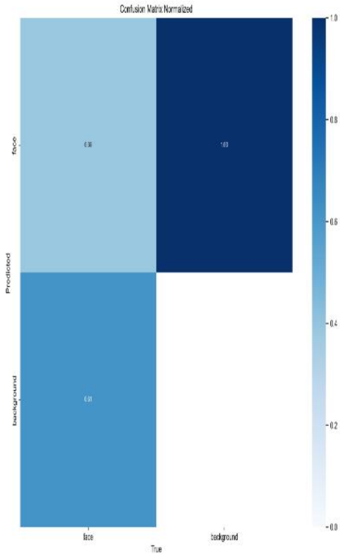

Fig. 5. Results of confusion matrix confusion matrix

Figure 6 shows the performance of the model at different confidence thresholds using the F1 curve. The F1 score is an indicator that comprehensively considers both accuracy and recall. When the confidence threshold is high, the F1 score also increases accordingly. This means that when the model requires a high level of confidence in the recognition results, its accuracy improves, but at the same time, it also increases the occurrence of false negative examples. Therefore, how to balance precision and recall remains a key issue that we need to address.

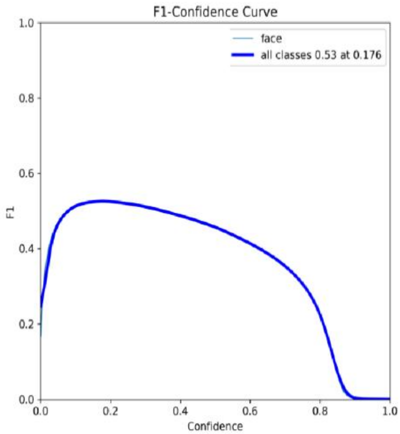

Fig. 6. F1 Curve

Figure 7 shows the change in accuracy as the confidence threshold increases. A higher confidence level improves the accuracy, indicating that the model can more accurately identify positive samples.

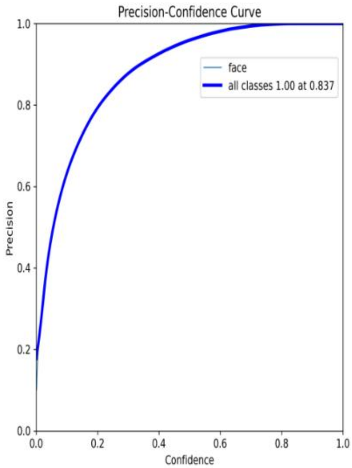

Fig. 7. Accuracy Confidence Curve

Figure 8 shows the change in recall rate as the confidence threshold increases. As the confidence level increases, the recall rate decreases, and the model ignores some low confidence positive samples

In summary, these two graphs illustrate how the accuracy and recall of the model vary at different confidence levels. The accuracy curve shows that with increased confidence, the model can more accurately identify positive samples. The recall curve reveals that as the confidence level increases, the model's recognition of low confidence positive samples decreases, resulting in some faces being ignored.

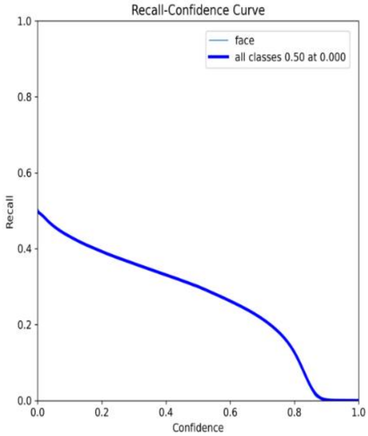

Fig. 8. Recall Confidence Curve

Figure 9 shows the relationship between precision and recall [10]. Ideally, the curve should be as close to the upper right corner as possible, indicating that high recall rates also maintain high precision.

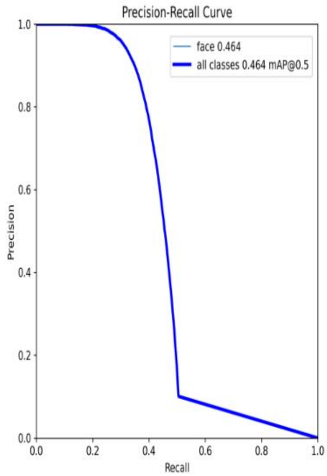

Fig. 9. Accuracy recall curve

3.2 Evaluation Results

Figure 10 shows various loss functions and evaluation metrics during the training process. Through these curves, we can observe whether the training converges. Through these evaluation curves, we can gain a more comprehensive understanding of the performance of the YOLOv8s model in facial recognition tasks and provide data support for future research and improvement.

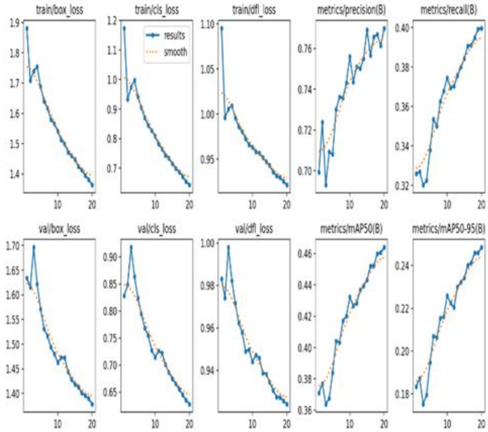

Fig. 10. Loss and evaluation metrics during the training process

4 Conclusion

This study achieved certain experimental results through a drone face recognition system based on the YOLOv8s model. The experiment shows that the model can effectively recognize faces under different backgrounds and lighting conditions, and exhibits high accuracy in many cases. However, despite some progress, the model still has some issues, such as partial detail omissions caused by low recall rates, as well as misjudgments and missed detections in complex backgrounds. These issues pose challenges to the accuracy of recognition and the reliability of the system. Through the display of evaluation indicators such as confusion matrix, F1 curve, accuracy and recall curves, we further realize that the performance of the model still needs to be optimized under certain conditions.

Although the current research results provide some support for the application of YOLO model in facial recognition, there is still room for improvement in practical applications. Especially in terms of accuracy and recall of facial recognition, further improvement is needed. Future research will focus on optimizing the balance between recall and accuracy, improving the model's recognition ability in complex backgrounds, and reducing the interference of background noise on recognition results. In addition, due to the significant impact of external factors such as lighting changes and shooting angles on recognition performance, more advanced image enhancement techniques and data preprocessing methods need to be explored in the future to improve the robustness and adaptability of the model.

In summary, although we have obtained a relatively good result for the YOLO model, we have also identified some issues during our research. For example, the recall rate of this model is relatively low, which may result in some details being overlooked and affect recognition accuracy. Secondly, there may be misjudgments or omissions due to the complex background of some images. In future research, corresponding improvements will be made to the code and model based on feasibility to reduce the impact of external factors such as images.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Chenglin C.: 'Design and implementation of a vision-based human pose command recognition system for UAVs'. MSc thesis, National University of Defense Technology, 2019.
2. Zhi X.: 'Face detection and recognition from small UAV perspectives'. MSc thesis, Changchun University of Science and Technology, 2022.
3. Jiankang D., Jinghui G., Niannan X., et al.: 'ArcFace: Additive angular margin loss for deep face recognition', Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, Long Beach, USA, June 2019, pp. 4690–4699.
4. Qingyan M., Shengcai Z., Zhen H., et al.: 'MagFace: A universal representation for face recognition and quality assessment', Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, Nashville, USA, June 2021, pp. 14225–14234.
5. Mingyang K., Anil K.J., Xiaoming L.: 'AdaFace: Quality adaptive margin for face recognition', Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2022.
6. Zhenxiong F., Chunhua D.: 'Face recognition algorithm based on MAML online meta-learning', Comput. Technol. Dev., 2025, pp. 1–10.
7. Jiangdong L., Boyan M., Fen Z.: 'Object detection system based on deep learning under edge computing environment', Autom. Instrum., 2022, (4), pp. 85–88.
8. Fei Z., Jian W., Yuesong Z.: 'Lightweight infrared small target detection under oblique perspective based on YOLOv5', Infrared Technol., 2025, 47, (2), pp. 217–225.
9. Danqing X., Yiquan W.: 'Progress of deep learning algorithms for optical remote sensing image target detection', J. Remote Sens., 2024, 28, (12), pp. 3045–3073.
10. Zihan L., Xiaoling X.: 'Lightweight face recognition algorithm based on improved YOLOv8s', Inf. Technol. Inf., 2024, (12), pp. 201–204.