

Research on Dense Crowd Area Detection Method Based on Improved YOLOv5 and Improved DBSCAN Clustering Algorithm

Guchang Yuan¹, Zhonghua Ma^{1,2*}

¹Tianjin University of Technology and Education, Tianjin, China

²Lvliang Vocational and Technical College, Lvliang, China

Email: *mazh@tute.edu.cn

How to cite this paper: Yuan, G.C. and Ma, Z.H. (2024) Research on Dense Crowd Area Detection Method Based on Improved YOLOv5 and Improved DBSCAN Clustering Algorithm. *Journal of Applied Mathematics and Physics*, 12, 4206-4212.

<https://doi.org/10.4236/jamp.2024.1212259>

Received: November 21, 2024

Accepted: December 24, 2024

Published: December 27, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In modern society, dense crowd detection technology is particularly important due to the frequent occurrence of crowd scenes such as stations, shopping malls, and event sites, which are often accompanied by safety risks, like stampede accidents. Although many studies have made progress in estimating population density, the ability to accurately identify dense areas in multi-scale scenarios still needs to be improved. To solve this problem, this paper proposed an improved multi-scale dense crowd detection method based on YOLOv5 and improved the DBSCAN clustering algorithm to identify densely crowded areas. Experiments show that the improved multi-scale dense crowd detection method can identify target crowds at multiple scales, and the accuracy of its detection results is around 70%. In addition, by calculating the crowd density under the same scale conditions and visualising the dense areas, we were able to solve the problem of dividing the crowded areas and visualise the dense areas more accurately. These improvements enhanced the applicability and reliability of the model in practical applications and provided strong technical support for security monitoring and management.

Keywords

Dense Crowd Detection, YOLOv5, Multi-Scale Detection, DBSCAN Clustering

1. Introduction

In modern society, there are dense crowds in various scenes such as stations, shopping malls, and event sites. Dense crowds can bring safety risks, such as stampede accidents. In order to effectively prevent and control the occurrence of such risks,

a dense crowd detection algorithm based on computer vision technology and real-time capture and analysis of the state of the crowd in the area through the surveillance cameras installed on the site came into being. Through real-time monitoring, real-time processing and analysis of video streams provide strong support for security management. Rafid [1] *et al.* focused on the impact of using the LAB color space on the performance of NASNetMobile for dense crowd detection through the fine-tuning process. Liu [2] and others designed an intelligent campus security system using image detection and IoT communication technology for crowd detection and hazardous action detection. Arefin [3] *et al.* emphasized the importance of crowd density estimation for safety and public services, leading to the development of an artificial intelligence project for this purpose. Fujii [4] *et al.*, used the CrowdMAC framework to address incomplete crowd density maps, achieving robust crowd density prediction. Li [5] *et al.* introduced LLM Count, which enhances the performance of fixed crowd detection in various application scenarios using large language models. These studies have collectively advanced the technology for detecting dense crowds in different fields, but they still cannot accurately identify dense areas within dense crowds under multi-scale conditions.

In order to solve the problem of multi-scale dense crowd area identification in dense crowd scenes, a multi-scale dense population detection and dense area identification algorithm through an improved clustering algorithm was proposed. On a large scale, the YOLOv5 detection method was used to detect sparse crowd targets in the whole scene. On a small scale, the YOLOv5 detection method was used again to detect targets in dense populations after image segmentation. The results of the two scales are combined by the Non-Maximum Suppression (NMS) method, which complements each other and removes redundant results. Experiments show that compared with the YOLOv5 detection method alone, the proposed multi-scale detection method has a significant improvement in detection accuracy, and finally, the function of automatic identification of densely populated areas is realized by improving the DBSCAN clustering algorithm.

2. Background and Related Work

2.1. Improved Dense Crowd Detection Algorithm for YOLOv5

The YOLOv5 algorithm is not accurate in multi-scale object detection when dealing with complex scenarios such as dense and small targets. To address the multi-scale issue and find a balance between detection speed and accuracy, this paper proposes an improved YOLOv5 dense crowd detection algorithm. The improved YOLOv5 dense crowd detection algorithm is a multi-scale detection method from near and far, which has the ability to perceive targets at different scales and is suitable for various complex scenarios.

Figure 1 is a flowchart of the improved YOLOv5 algorithm. First, the input image is directly fed into the YOLOv5 network for rough detection, which is mainly to detect large target groups at close range. Then, in order to identify the long-distance population, the original image was divided into 9 sub-blocks for

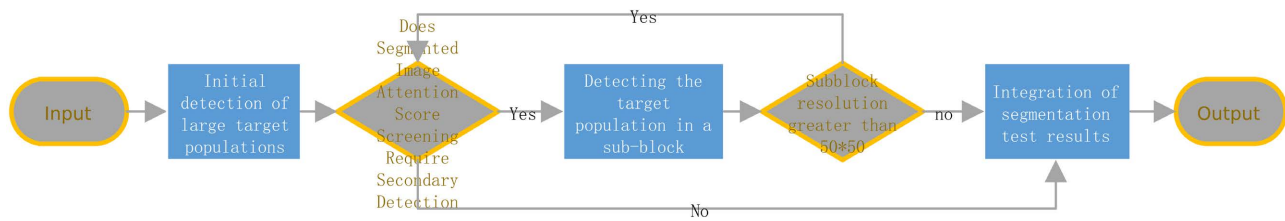


Figure 1. Flowchart of the improved YOLOv5 algorithm.

initial segmentation detection. The sub-blocks that need to be detected are screened through the attention score and sent to the YOLOv5 network for detection, and the attention score of the sub-blocks needs to be greater than 0.55. Equation (1) is the formula for the attention score, where edge density is the ratio of the total number of edge pixels in the image to the total number of pixels in the image, and colour contrast is calculated by calculating the variance of the colour image crop over the three channels of RGB.

$$\text{score} = 0.1 * \text{edge density} + 0.9 * \text{color contrast} \quad (1)$$

In this way, medium-sized populations can be detected. Duplicate detection frames are then removed using the Non-Maximal Suppression (NMS) algorithm to ensure that only one detection frame is retained for each target. In this case, when the resolution of the sub-block is greater than 50×50 , each sub-block obtained by the previous partition is divided into 9 smaller sub-blocks again. Similarly, the sub-blocks that need to be detected are screened by the attention score and sent to the YOLOv5 network for detection. This stage focuses on the detection of smaller targets, again using the NMS algorithm to remove duplicate detection frames to ensure the accuracy of small target detection, until the resolution of the sub-block is less than 50×50 stopping the segmentation. Finally, all the segmentation detection results are restored to the original image to obtain the final detection results. By integrating the segmentation block detection results, various targets in the image can be comprehensively covered to improve the overall detection effect.

Based on the coco128 dataset, this class is deleted except for the class 0 (person), and the hyperparameters are adjusted for training with 90 iterations. In **Figure 2** a is the original figure, and b, c, and d are the corresponding model detection results under the condition of confidence equal to 0.1. In **Figure 2**, YOLOv10 detects 10 people, while YOLOv5 detects 20 people, and the improved YOLOv5 detects more than 60 people, we can find that the detection effect of YOLOv10 is not as good as that of YOLOv5, which is due to the fact that YOLOv10 is suitable for detecting smaller objects or distant objects, and is not suitable for detecting the case of a large change in scale. The improved YOLOv5 detection effect proposed in this paper is significantly better than YOLOv5 and YOLOv10. In **Figure 5**, the total number of people in the original figure is around 80, and from the point of view of the accuracy rate of detecting the number of people, the accuracy rate of the detection results of YOLOv10 is about 10% lower, and the accuracy rate of the detection results of YOLOv5 is about 20% lower, and the accuracy rate



Figure 2. Comparison of crowd target recognition between YOLOv10, YOLOv5 and improved YOLOv5.

of the detection results of the improved YOLOv5 is high between 70% and 80%. This verifies that our proposed improved YOLOv5 algorithm is suitable for detecting situations where the scale of the crowd varies greatly, and the improved YOLOv5 algorithm also lays the foundation for detecting aggregated areas in dense crowds.

2.2. Improved DBSCAN Clustering Algorithm for Dense Crowd Area Detection

With the continuous development of technology, the detection and management of densely crowd areas have become increasingly important. To better address scenarios with dense crowds, we have applied the improved YOLOv5 dense crowd detection algorithm proposed in the previous section to video stream crowd detection to identify dense areas within. First, when the input data is a video stream, we apply the improved YOLOv5 dense crowd detection algorithm to each frame of the video. For the detection results of each frame image, we calculate the center point coordinates of each detection box and store them in the variable detection information, which will be used as the basis for subsequent cluster analysis.

We will improve the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method to find dense areas by clustering these centre points. DBSCAN is a density-based clustering algorithm, which is effective in discovering clusters of arbitrary shapes and has good robustness to noisy data. DBSCAN is a density-based clustering algorithm that can effectively find clusters of arbitrary shapes and is robust to noisy data. For the clustering results obtained using

DBSCAN, we will identify the dense area. To do this, we first need to define what a dense area is. Intuitively, if a crowd forms a compact mass in space without obvious gaps, it can be defined as a dense area. However, from a mathematical point of view, an area is considered dense if the spacing between people in an area is very small and the distance between people is less than a specific value (e.g. 0.5 m). On this basis a dense area is further defined by setting a specific threshold of people, e.g. an area with more than a certain number of people per square metre (e.g. 5 people/m²) is defined as a dense area [6], and the density of the crowd is determined by calculating the ratio of the number of people within a specific range to the area of that range. Further, it can be defined in terms of relative density, with areas with significantly higher population density compared to surrounding areas considered densely populated.

Using these understandings of dense areas, this paper devises a definition in the field of dense crowd detection that conforms to what is known as a dense area. First of all, the variable detection info found in the centre point coordinates through the formula (2) into the same scale conditions of the coordinates, the meaning of the formula (2) is to picture box under the bottom edge of the centre point for the new coordinates of the origin, in the distance from the new coordinates of the origin of the base of the proportional enlargement, and finally restore the origin of the coordinates to get the coordinates under the same scale conditions of the coordinates of the x, y .

$$x = \left(x_0 - \frac{W}{2}\right) * \left(1 + \left|\frac{x_0 - \frac{W}{2}}{W}\right|\right) + W, \quad y = 2H - (H - y_0) * \left(1 + \frac{H - y_0}{H}\right) \quad (2)$$

DBSCAN is used to obtain a series of classes for the points represented by these newly obtained coordinates, and the ratio of the number of people in the range of these classes to the area of the range is used to determine the density of the crowd, at this time, if the density of the crowd in a certain area exceeds the set threshold, then the area is called a dense area, and finally, we will reduce these dense areas to the real picture.

In order to calculate the crowd density and reduce the dense areas to the real picture, we will improve the DBSCAN, that is, on the basis of the previous step of using the DBSCAN to obtain a series of classes, we will perform the following operations to find a minimum outer rectangle for all the points in the clustered classes, so as to obtain a minimum rectangular region that can cover all the points. The following is the method for determining the boundaries of the minimum rectangle: for all the points in these classes, use equation (3) to calculate.

$$x_{i,\min} = \min(x_j), \quad y_{i,\min} = \min(y_j), \quad x_{i,\max} = \max(x_j), \quad y_{i,\max} = \max(y_j) \quad (3)$$

where are the coordinates of all points, and j is traversed from 1 all the way to the population size n_i . The area of the smallest rectangle is then calculated using Equation (4):

$$Area_i = (x_{i,max} - x_{i,min}) \times (y_{i,max} - y_{i,min}) \quad (4)$$

Then, the number of people density is calculated using Equation (5): assuming that the number of people detected n_i is distributed in this rectangle, then the number of people density D_i in the area can be expressed as:

$$D_i = \frac{n_i}{Area_i} \quad (5)$$

After calculating the population density at the same scale, a area that is defined as a dense area must meet the population density of the area that exceeds the set threshold $D_{threshold}$, and these areas are redrawn together to create a complete dense area, as shown in Equation (6):

$$Dease\ Area = \begin{cases} True, & D_i > D_{threshold} \\ False, & otherwise \end{cases} \quad (6)$$

Finally, the dense region is restored in the real picture to get the dense region of the most initial picture.

This method not only improves the flexibility and accuracy of crowd density estimation, but also helps us better understand the distribution of people in the image. In practice, dense areas need to be given special attention and appropriate control. Finally, we will detect and cluster each frame in the video stream and connect the results into a smooth video stream. In this way, we can intuitively observe the distribution of the population and the changes in the dense area, which provide strong support for crowd management and safety assurance.

3. Results

A video of students crossing a footbridge was randomly collected for the experiment, and **Figure 3** shows the results of the experiment by collecting dense regions from the video. The blue detection boxes in **Figure 3** are the crowd targets we detected, and these targets are used to find the dense regions using the improved clustering method and are plotted with red boxes. The experimental results demonstrate that the improved method can accurately detect the dense regions of the crowd and successfully capture the crowd's aggregation characteristics in the video. In addition, it can be observed from **Figure 3** that the distribution of dense

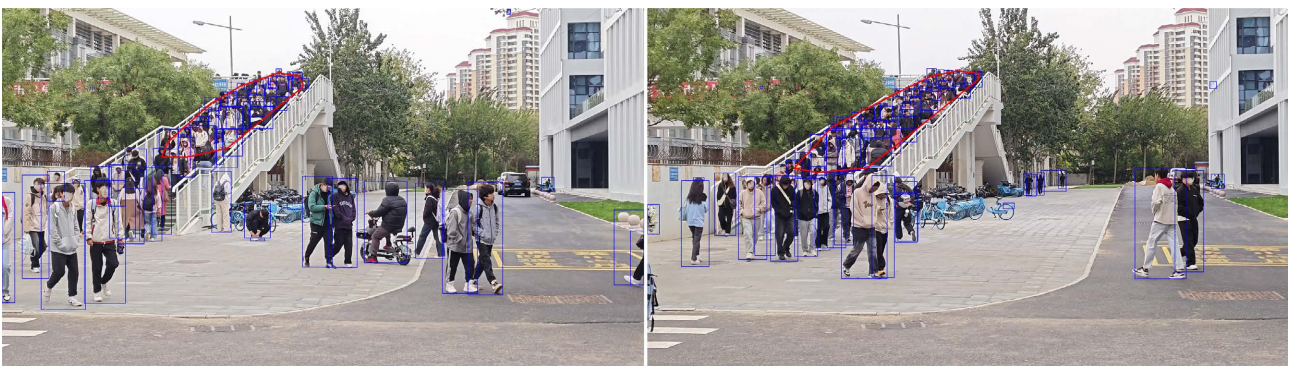


Figure 3. Dense area recognition experiments in the collection video.

regions in different images is well adapted: the shapes and sizes of the dense regions are highly consistent with the distribution of the crowd. This indicates that our method is not only applicable to target clustering in static scenes, but also able to respond dynamically to the changes of dense regions in videos. Meanwhile, compared with traditional clustering methods (e.g., DBSCAN, K-Means), our improved clustering method is able to deal with dynamic crowd density and dense regions under multi-scale conditions, while traditional clustering methods can only identify crowded regions and compute their crowd density for crowded targets of the same scale.

4. Conclusion

This study not only improved the performance of the YOLOv5 algorithm under multi-scale conditions but also successfully implemented an effective method for crowd density assessment and dense area recognition. This method can identify dense areas in surveillance scenarios, providing a powerful technical tool for public safety management. This achievement will greatly enhance the monitoring and response capabilities in the field of public safety, especially in high-risk, densely populated situations.

Funding

This paper is partially supported by Fundamental Research Program of Shanxi Province (Grant No. 202303021211245).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Saputra, D.B.A., Sari, C.A. and Rachmawanto, E.H. (2024) Jasmine Flower Classification with CNN Architectures: A Comparative Study of Nasnetmobile, VGG16, and Xception in Agricultural Technology. *Advance Sustainable Science Engineering and Technology*, **6**, 0240409. <https://doi.org/10.26877/asset.v6i4.790>
- [2] Liu, Y.C., Liu, X.X. and Gao, Q. (2024) Image Detection Based on the Grey Wolf Algorithm Optimization and Its Application in Campus Security. *Proceedings of Fourth International Conference on Computer Vision and Data Mining (ICCVDM 2023)*, Changchun, 20-22 October 2023.
- [3] Arefin, M., Wadud, Md.A.H. and Rahman, A. (2024) Indoor and Outdoor Crowd Density Level Estimation with Video Analysis through Machine Learning Models. <https://arxiv.org/abs/2405.07419>
- [4] Fujii, R., Hachiuma, R. and Saito, H. (2024) CrowdMAC: Masked Crowd Density Completion for Robust Crowd Density Forecasting. <https://arxiv.org/abs/2407.14725>
- [5] Li, B.Y., Ding, S.Y., Ma, D., Wu, Y.X., Liao, H.J. and Hu, K.Y. (2024) LLMCount: Enhancing Stationary mmWave Detection with Multimodal-LLM. <https://arxiv.org/abs/2409.16209>
- [6] Zhang, J.J., Shi, Z.G. and Li, J.C. (2018) Research Status and Trends of Population Counting and Crowd Density Estimation Technologies. *Computer Engineering & Science*, **40**, 282-291.