APPLIED TO HIERARCHICAL NAVIGATION BASED ON DENSITY DETECTION AND PEDESTRIAN FLOW DETECTION

Wei-Zhi Zhang and Wei-Liang Lin

National Chung Hsing University, Taichung City, Taiwan, ROC

ABSTRACT

This article is based on the YOLOv4[1] object detection method, and builds a crowd density detector suitable for this research content through a rigorous crowd training data set, so as to obtain density information. It also uses the FairMOT[2]-based object tracking method to identify and obtain through each frame Information on the direction of people flow for each object.

Hierarchical navigation is proposed, which is divided into wide area and narrow area. The wide area considers the density and the flow of people as the path planning, while the narrow area tends to avoid obstacles in the field, and implement this in the Unity virtual environment, and implement hierarchical navigation in the National Chung Hsing University internship store.

1. INTRODUCTION

Now is the era of AI and it spreads to various fields, and longer deep learning is beginning to be widely used. Its specific tasks include object detection, semantic segmentation, pose estimation, object tracking, and behavior recognition and autonomous driving. The AI analysis of the objects in the scene by vision technology is one of the most important research issues at present. Through the expansion of the powerful fitting capabilities of deep learning, the accuracy of each task is improved.

Service and upgrade interaction through the customer guidance system is an intelligent technology that has gradually emerged in shopping malls and supermarkets. The basic realization method is to use the monitors in the shopping malls to determine the crowding degree of the crowd and the direction of the flow of people in the screen by the algorithm. This information is used as the evaluation basis for path planning, and the Intelligent Autonomous Mover will lead to the customer's designated product area, as shown in Figure 1.

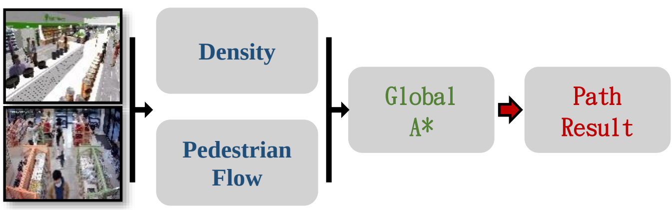

Fig.1 Customer guidance algorithm flow

2. RELATED WORK

Nowadays, the development of computer vision has become more and more mature, and the technology of object detection and object tracking is also widely used in many fields, such as traffic flow calculation, self-driving car research and development, factory defect detection.

This section will introduce the selection of YOLOv4[1] and FairMOT[2] as the basis for the detection of density and flow of people, and the hierarchical navigation algorithm designed in this article will be used to achieve path planning results in a virtual environment.

2.1. YOLOv4[1]

This article chooses YOLOv4[1] because it has a high-accuracy, high-real-time network architecture, and only requires a GPU to train quickly, and will introduce several important Module innovations, as shown in Figure 2.

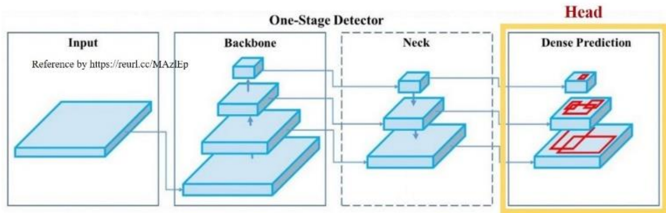

Fig.2 Object detector

CSPNet is an innovation of Backbone. Its main purpose is to enable the network architecture to obtain richer gradient fusion information and reduce the amount of calculation. The method is to first divide the feature map of the Base Layer into two parts, and then one part undergoes several operations: 1.Transition, 2.Concatenation, 3.Transition, and then merge this operation with another part, as shown in Figure 3.

YOLOv4[1] adds a bottom-up feature pyramid behind the FPN layer, and it contains two PAN structures. This combined operation improves the feature extraction capability. The FPN layer presents fine semantic features from top to bottom, and conveys delicate spatial features from bottom to top, as shown in Figure 4.

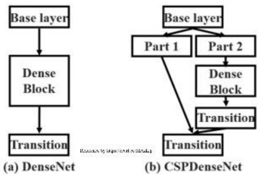

Fig.3 Backbone's innovation

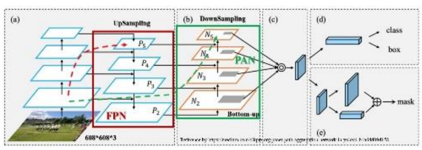

Fig.4 Neck's innovation

2.2. FairMOT[2]

The Two-stage method has achieved good results on multiple public data sets, but in actual deployment they usually face the problem of weak scalability, especially when there are a large number of people in the image. Such as supermarkets and other densely populated In the scenario, the cost of extracting Re-ID features for each person through a network alone will increase linearly, resulting in a larger delay time.

The focus of FairMOT's[2] research is to combine Object Detection and Re-ID algorithms in the same network architecture to improve inference performance, as shown in Figure 5.

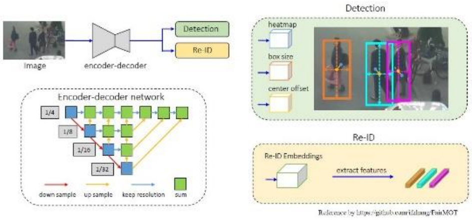

Fig.5 FairMOT[2] network architecture

The main innovation of FairMOT[2] is that Anchor will cause ambiguity of the target object, so Anchor-Free method should be used instead, as shown in Figure 6.

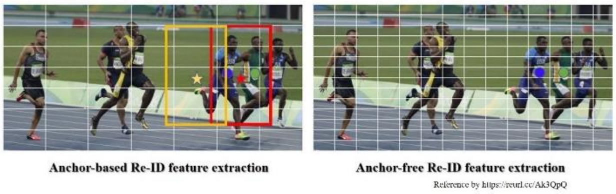

Fig.6 Anchor-based vs Anchor-free

2.3. A* search algorithm

Pathfinding is a common basic artificial intelligence program in role-playing related video games. When the player clicks on a certain point on the map, the character on the screen will automatically cross the obstacle and reach the destination, which is the purpose of the path search algorithm. Of course, self-propelled vehicle navigation and GPS satellite navigation software for vehicles will also need to use A* or other different forms of path search algorithms, as shown in Figure 7.

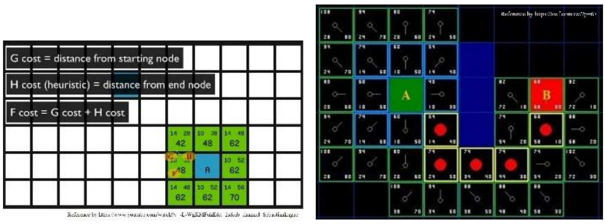

Fig.7 A search algorithm*

2.4. Visualization of path planning

Unity3D is a fully integrated professional game engine that allows users to easily create interactive content such as 3D games and scene simulations. However, not only the scenes are very realistic, but also C# can be used to compile scripts to implement many existing algorithms and complex object operations, as shown in Figure 8.

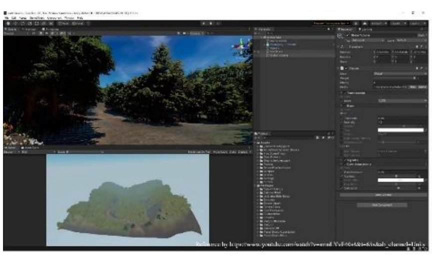

Fig.8 Unity platform

3. EXPRIMENT

This section will introduce the algorithms selected in the previous chapters and design them as tools required for customer guidance, such as density detectors, pedestrian flow detectors, and wide-area A* navigation.

3.1. Density detector

The pre-trained model of YOLOv4's[1] object detection method is based on the results trained by COCO Dataset, so it is not suitable for the density detector in this article. We must use a rigorous crowd training data set to create a suitable the density detector used in this study. For the above reasons, we choose CrowdHuman Dataset to train the object detection model, and use the 416x416 image size and eliminate objects that are too small to reduce the training time and improve the accuracy of the model.

We will select a street view surrounded by high-density people as the test result of the density detector, and manually count the number of people in this picture, we can see that the number of people in this picture is about 113 people, and the result predicted by the density detector is 109 people, and the test accuracy rate is 109/113 = 96%, as shown in Figure 9.

Fig.9 Ground True and Prediction

The category is set as the prediction frame of the head area and the whole body area, the results of Training Loss ≈ 19 and Validation mAP(%) ≈ 80% are obtained after training, as shown in Figure 10.

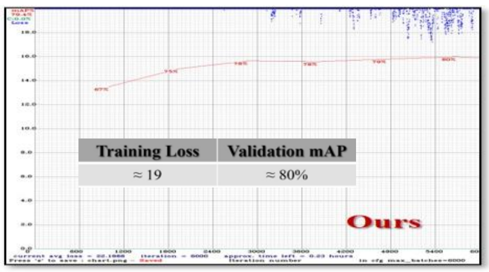

Fig.10 Training loss and mAP

3.2. Pedestrian detector

The center point obtained by the object detection in the object tracking method associated with the Re-ID algorithm is stored as a sequence with the center point of each pedestrian in each frame of the video, and it will be executed until the end of the video. That is, the displacement of pedestrians has a time dimension, as shown in Figure 11.

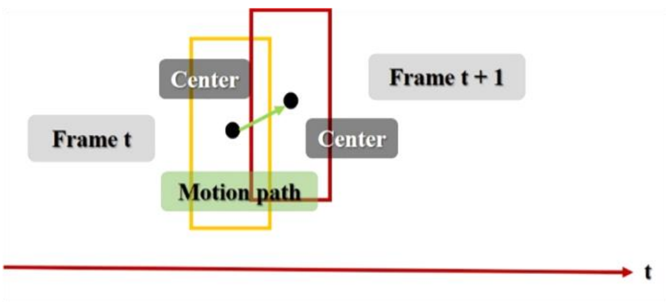

Fig.11 Displacement detection

3.3. Hierarchical navigation

The setting of wide-area navigation. This wide-area navigation will have several important points. A larger grid is used to establish the search area. This is because the wide area does not need to consider the location of obstacles, so that no fine grid is used to show the obstacles. , And a larger grid can also cover a larger part of the density and the area where the flow of people corresponds to the actual field, as shown in Figure 12.

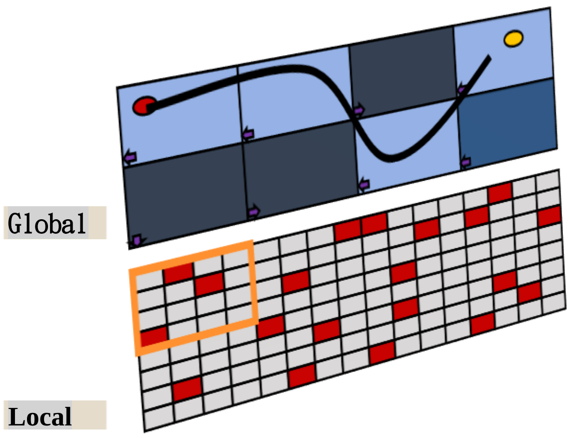

Fig.12 Hierarchical navigation

4. RESULT

Based on the monitor screen as input, we will get information on the flow of people and density, and then use the wide-area navigation algorithm to get the path results, as shown in Figure 13.

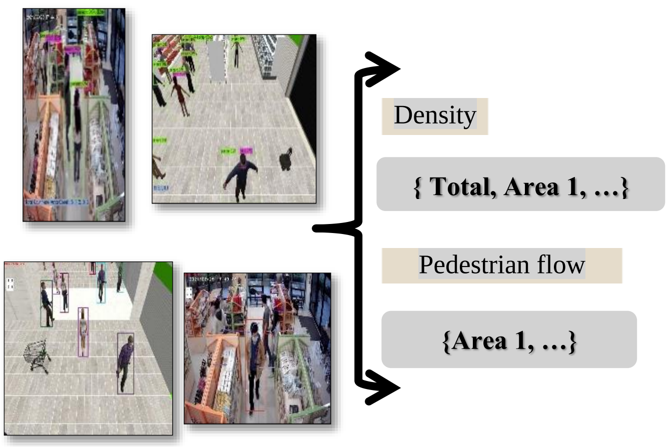

Fig.13 Density and Pedestrian flow

Since the original A* Pathfinding is based on the principle of the shortest path, it means that the evaluation of the path search is based on choosing a lower F value, and tends to avoid obstacles in the field, such as product display areas, people, carts, etc. . However, when there is a crowded area in the field, the vehicle will not choose to avoid the area with many people, but will only choose the route with the least cost to reach the destination, as shown in Figure 14.

*Fig.14 Original A**

Based on the route planning result of the wide-area navigation algorithm, input this into the wide-area navigation algorithm and multiply it by the weight according to the density and the value of the flow of people to estimate the result of the route planning to avoid the multi-person area through the algorithm. In addition, combining wide-area navigation and narrow-area navigation, this hierarchical navigation can not only avoid areas surrounded by people based on density and flow of people, but also further avoid all obstacles in the field through narrow areas, as shown in figure 15.

*Fig.15 Hierarchical A**

5. CONCLUSION

According to the research results of this article, through the monitor, Intelligent Autonomous Mover and algorithm design, not only can complete customer guidance without disturbing other customers, but also obtain corresponding personnel flow information for the mall to analyze the store operation.

6. REFERENCES

[1] Alexey Bochkovskiy and Chien-Yao Wang and Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv preprint arXiv:2004.10934, https://arxiv.org/pdf/2004.10934.pdf, 2020.

[2] Yifu Zhang and Chunyu Wang and Xinggang Wang and Wenjun Zeng and Wenyu Liu, “FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking,” arXiv preprint arXiv:2004.01888, 2020.

[3] N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645-3649, doi: 10.1109/ICIP.2017.8296962.

[4] Shaoqing Ren and Kaiming He and Ross Girshick and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” arXiv preprint arXiv:1506.01497, https://arxiv.org/pdf/1506.01497.pdf, 2016.