



# Design and Implementation of Intelligent Inspection and Alarm Flight System for Epidemic Prevention

Jiwei Fan , Xiaogang Yang \*, Ruitao Lu , Xueli Xie and Weipeng Li

Department of Automation, Rocket Force University of Engineering, Xi'an 710000, China; fjw19900619@163.com (J.F.); lrt19880220@163.com (R.L.); 18509242741@163.com (X.X.); williamli\_pro@163.com (W.L.)

\* Correspondence: doctoryxg@163.com

**Abstract:** Unmanned aerial vehicles (UAV) and related technologies have played an active role in the prevention and control of novel coronaviruses at home and abroad, especially in epidemic prevention, surveillance, and elimination. However, the existing UAVs have a single function, limited processing capacity, and poor interaction. To overcome these shortcomings, we designed an intelligent anti-epidemic patrol detection and warning flight system, which integrates UAV autonomous navigation, deep learning, intelligent voice, and other technologies. Based on the convolution neural network and deep learning technology, the system possesses a crowd density detection method and a face mask detection method, which can detect the position of dense crowds. Intelligent voice alarm technology was used to achieve an intelligent alarm system for abnormal situations, such as crowd-gathering areas and people without masks, and to carry out intelligent dissemination of epidemic prevention policies, which provides a powerful technical means for epidemic prevention and delaying their spread. To verify the superiority and feasibility of the system, high-precision online analysis was carried out for the crowd in the inspection area, and pedestrians' faces were detected on the ground to identify whether they were wearing a mask. The experimental results show that the mean absolute error (MAE) of the crowd density detection was less than 8.4, and the mean average precision (mAP) of face mask detection was 61.42%. The system can provide convenient and accurate evaluation information for decision-makers and meets the requirements of real-time and accurate detection.

**Keywords:** unmanned aerial vehicle; epidemic prevention and control; crowd density detection; face mask detection



**Citation:** Fan, J.; Yang, X.; Lu, R.; Xie, X.; Li, W. Design and Implementation of Intelligent Inspection and Alarm Flight System for Epidemic Prevention. *Drones* **2021**, *5*, 68. <https://doi.org/10.3390/drones5030068>

Academic Editor:

Pablo Rodríguez-Gonzálvez

Received: 7 June 2021

Accepted: 25 July 2021

Published: 27 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The current novel coronavirus pneumonia epidemic is raging around the world. As of 1 February 2021, the number of infections worldwide has exceeded 100 million, and the cumulative death toll has exceeded 2 million. Epidemiological investigations have shown that the novel coronavirus mainly spreads through respiratory droplet transmission and contact transmission. Most of the cases can be traced to close contacts with confirmed cases [1]. Therefore, if we can avoid crowd gatherings and if we can monitor personal contact and promptly remind the public to wear masks, we can effectively control and prevent the spread of the epidemic [2]. In the control and prevention of the novel coronavirus pneumonia epidemic, UAV high-altitude inspections, as an effective means of reducing the risk of contact and making up for the shortage of personnel for epidemic prevention and control, have become a powerful tool in the fight against the epidemic. UAVs have the characteristics of flexible maneuverability, fast inspections, and high work efficiency, and have gradually formed an all-round three-dimensional inspection pattern of “air inspections–ground monitoring–communications, command, and control”, which plays an important role in improving the epidemic prevention and control systems and mechanisms, and in improving the efficiency of the national public health emergency management systems [3].

Due to the strong interpersonal transmission characteristics of the novel coronavirus, inspectors cannot fully visit the scene to understand and grasp the situation when conducting epidemic prevention and control work. Therefore, many places use UAVs to conduct aerial inspections of key areas and use wireless image transmission equipment to transmit the inspections back to the ground command in real-time to help staff monitor and supervise the situation. The headquarters rely on real-time returns of on-site images to form accurate judgments and analyses of the on-site situation, thereby realizing the integrated three-dimensional patrol and defense mode from the air and ground. However, due to the high cost of manpower monitoring and it being prone to flaws, computer vision technology needs to be introduced to assist in epidemic patrol monitoring [4–6]. In recent years, the combination of computer vision technology and UAV technology has become increasingly common, which has lifted the fundamental technical limitations for UAVs to deal with perception problems and secondary developments and applications [7–10]. In response to the actual requirements of epidemic prevention and control, the use of remote-control systems, airborne infrared temperature measurement systems, situational awareness, and other technologies has gradually replaced human inspection work. Currently, the application areas of UAVs for epidemic prevention and control mainly include safety inspections, disinfection sprays, thermal sensing, temperature measurement, and prevention and control propaganda.

Based on the analysis of the functional requirements of the UAV for epidemic prevention and control, we designed an intelligent flight system for epidemic prevention inspection, detection, and alarm. The system uses advanced technologies such as neural networks and artificial intelligence and can automatically capture gathered crowds and independently recognize the wearing of face masks and intelligent voice prompts. Compared with manual epidemic inspection and control, the intelligent epidemic inspection, detection, and alarm flight system has a larger inspection area, higher mobility, and lower cost. At the same time, it also avoids direct contact between personnel, reduces the probability of mutual infection, and reduces the risk of epidemic transmission. It is a powerful tool for the implementation of anti-epidemic information publicity, dense crowd situation awareness, and other tasks. The main contributions of the system are as follows:

- (1) Based on the quadrotor UAV, an intelligent inspection and warning flight system for epidemic prevention was designed;
- (2) Based on a convolutional neural network, a dense crowd image analysis and personnel number estimation technology was used to estimate and analyze the crowd in the inspection area online, which provides convenient and accurate evaluation information for decision makers;
- (3) Face mask detection methods based on deep learning were used to detect the face of pedestrians on the ground and identify whether they were wearing masks;
- (4) Based on intelligent voice warning technology, the system can avoid personal contact when reminding, dissuading, and publicizing the policy regarding face masks to ensure strong promotion of epidemic prevention and control.

The structure of this paper is as follows: Section 2 introduces the related work. Section 3 describes the research methods of this paper. Section 4 provides the experimental results and analysis of the proposed algorithm. Section 5 is the conclusion.

## 2. Related Work

During the period of epidemic prevention and control, wearing masks has been one of the most effective ways to prevent COVID-19 from spreading [11]. In public places, the supervision of mask-wearing is still a key factor in epidemic prevention and control. Therefore, mask-wearing detection is a core task. In many areas, UAVs are used for patrol inspection and epidemic prevention propaganda, but they are still limited to relying solely on human monitoring and control. If UAVs can be used to estimate crowd density and detect the wearing of masks, they can save human resources, and also analyze, prevent, and deal with emergencies effectively. At present, the commonly used target detection methods

of UAVs include the optical flow method [12], the matching method, the background difference method [13], and the frame difference method [14], as well as the method based on deep convolutional neural networks, which has performed well in recent years. The optical flow method can achieve better target detection when the camera is moving. However, the optical flow method is generally limited by many assumptions, and noise, light sources and changes in shadows easily affect the detection results. Before target detection, the matching method needs a manual design of the features of the target, a construction of the template containing the features of the moving target, and detection of the target by translating the target template on the image to be matched and calculating the similarity matching value of different regions. Common target detection methods of the matching method include the Haar feature + AdaBoost algorithm [15], the Hog feature + support vector machine (SVM) algorithm [16,17], and the Deformable Part Model (DPM) algorithm [18]. However, the characteristics of moving objects cannot easily be designed by hand. Manual design needs a lot of prior knowledge and design experience, which has certain limitations. When the moving target is deformed or the angle of view is changed, it is easy to cause the template feature to not represent the target well, which leads to the failure of the matching method. When the background model is known, the background subtraction method can accurately detect the moving target and obtain the target position. However, in practical applications, the establishment of the background model and changes in the external dynamic scene are very sensitive, so when the update of the background model cannot adapt to the change in the scene, it will seriously affect the detection results. When the camera is fixed, the frame difference method can better detect the moving objects in the environment. However, when the moving speed of the target is very slow or stationary, the target is not easily detected and is prone to misjudgment. Compared with the traditional methods, the target detection method based on a deep convolutional neural network can be widely used in complex scenes, such as different densities, target scale changes, different perspectives, and target occlusion [19–22]. Deep convolutional neural networks used for target detection do not need manually designed target features; they need enough pictures and corresponding target tags [23,24]. When the target in the picture needs to be recognized, the weight file learned is loaded into the network, and the forward reasoning of the network is used to output whether the picture contains the target that needs to be recognized. According to the different methods of network candidate region selection, the deep learning target detection methods were developed from the Region-Convolutional Neural Networks (R-CNN) [25–27] series algorithm based on the candidate region of the YOLO [28–30] series and the single-shot multibox detector (SSD) [31] series algorithm based on regression; its detection accuracy and efficiency have been improved to varying degrees. The characteristics of common target detection algorithms of UAVs are shown in Table 1.

**Table 1.** Characteristics of common target detection algorithms.

Algorithm	Characteristic
Optical flow algorithm	It can detect the moving target and is susceptible to interference by noise, light source, and other factors
Matching algorithm	It needs a lot of prior knowledge and design experience, and the feature calculation is large
Background subtraction algorithm	When the detection scene is fixed, it is easy to implement and sensitive to changes in external dynamic scene
Frame difference algorithm	When the moving speed of the target is very slow or stationary, it is easy to produce misjudgments
Deep learning algorithm	The target dataset is needed for training, so the detection accuracy is high and the real-time performance is good

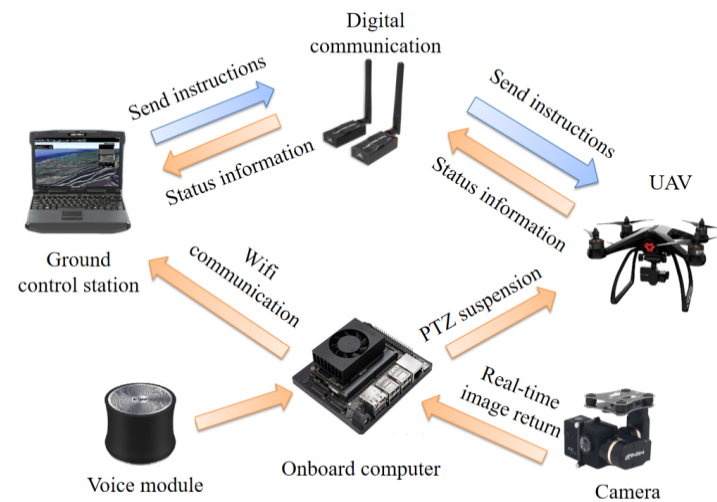
In recent years, with the development of deep learning neural networks, the complexity of target detection algorithms is increasing, and the model is becoming larger. This brings great challenges for the computing ability of hardware, and also places higher requirements on the processor in terms of operation speed, reliability, and integration. The main arm architecture processor at present is not fast enough to meet the real-time requirements when a large number of image operations are carried out. When analyzing a large amount of unstructured data to build and train the deep learning model, CPUs do not have strong processing abilities, and even multi-core CPUs cannot easily perform deep learning, which greatly limits the application of deep learning. NVIDIA Jetson is an embedded system, designed by NVIDIA, for the new generation of autonomous machines. It is an embedded modular supercomputer specially designed for deep learning neural networks, which provides a perfect single-board operation environment for digital image processing and various machine learning and artificial intelligence algorithms. It has been widely used in UAVs, intelligent robot systems, mobile medical imaging, and other fields [32–34]. For high-density crowd aggregation analysis, a popular method is to generate a heat map of the crowd, and then the crowd count becomes the integral calculation of the heat map. The pedestrian density and aggregation degree per square meter can be calculated at the same time. Compared with the target detection scheme, the thermal graph can more effectively reflect the distribution of a dense crowd, because it is difficult to detect dense targets and very small targets in object detection. The thermal graph generation is a peer-to-peer (P2P) scheme, which can more effectively analyze crowd density by predicting each pixel. Most face-detection methods are improved and optimized based on the deep learning object detection algorithm; that is, the general object detection model is configured specifically to adapt to the face detection task. In many object detection models, the SSD algorithm is the most commonly used face detection algorithm [35,36]. Compared with the ordinary face detection model, which only judges whether there is a face or not, face mask detection is divided into two categories: face with mask and face without mask. To better realize the potential of UAVs in the field of epidemic prevention and control, emergency response, and other fields, and to achieve effective prevention and control of an epidemic situation and reduce the risk of cross-infection among first-line staff, accelerating the application of UAVs has been suggested, learning from the application mode and the development experience of UAVs in other fields, and continuously carrying out scientific and technological empowerment and precise implementation. This allows first-line staff to make greater contributions to national security and public security on the premise of ensuring their security.

### 3. Our Approach

During the development of the intelligent epidemic prevention patrol detection and alarm flight system, the use of the environment and habits of the system were fully considered, which allows for strong system reliability. Additionally, the operability and completeness of the system were fully considered. The workflow of the system is shown in Figure 1. Firstly, the operator delimits the patrol area in the ground station, and the ground station plans the flight path according to the delimited area and sends the plan to the flight platform to execute the planned path. During the flight, the UAV detects the crowd density on the ground. If a crowd is gathered, the UAV plays the epidemic prevention policy propaganda, persuading the crowd to disperse. At the same time, the UAV comes close to observe the wearing of masks. If there are people who do not wear masks in the crowd, the UAV will remind them by voice to wear a mask. Finally, the inspection log and captured images are fed back to the ground station to provide materials for personnel contact tracing.



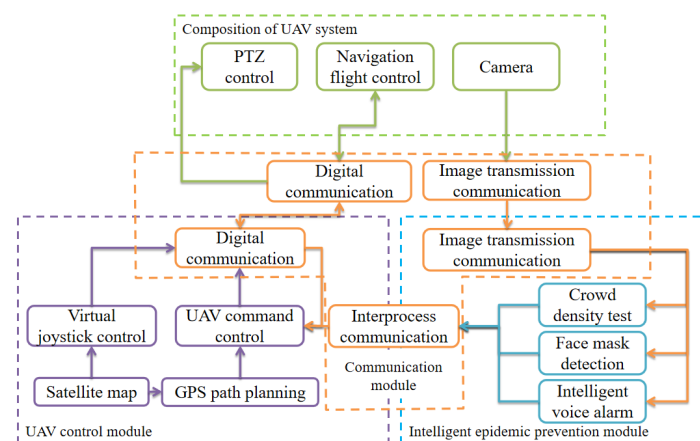




**Figure 2.** System overall structure chart.

### 3.2. Software Design

The software architecture flow of an intelligent epidemic prevention patrol detection alarm flight system is shown in Figure 3. The software part mainly includes the UAV control module, an intelligent epidemic prevention module, and a multi-process information communication module. The UAV control module is mainly used for UAV position and attitude control, patrol area planning, and mission communication. The intelligent epidemic prevention module includes crowd density detection, face mask detection, intelligent voice broadcast, and other functional modules, in which the crowd density detection module is used for the UAV to detect and distinguish the ground crowd aggregation; the face mask detection and recognition module is used for monitoring and warning the wearing of masks in the public environment; the intelligent voice broadcast module is responsible for the broadcast of epidemic prevention policy documents, crowd dispersal calls, and mask-wearing prompts. The multi-process information communication module is mainly responsible for scheduling the timing coordination and information sharing of each software process in the system, creating and closing processes, and other coordination operations to prevent process blocking, causing software crashes. The biggest advantage of the architecture is its flexibility; that is, the system can expand new functional modules at any time according to its actual needs by opening up a new subprocess supporting shared memory.



**Figure 3.** Software architecture.

### 3.2.1. UAV Navigation Control Module

This system can control the flight of UAV by a remote controller or by the UAV control module. The UAV control module, also known as a ground control station, is the command center of the whole UAV system. Its main functions include mission planning, UAV position monitoring, and route map display. Mission planning mainly includes processing mission information, planning inspection area, and calibrating flight routes. The UAV position monitoring and route map display parts are convenient for operators to monitor the UAV and track status in real time. The UAV remains in contact with the ground control station through the wireless data link during the mission. In the case of special circumstances, the ground control station needs to perform navigation control so that the UAV can fly according to the safest route.

### 3.2.2. Crowd Density Detection Module

High-density crowd aggregation can easily lead to the spread of an epidemic in a large area. The analysis of high-density crowds is conducive to the real-time separation and control of the crowd and the prevention of accidents. Currently, a popular method is to generate a heat map of the crowd, and then the crowd count becomes the integral calculation of the heat map. The pedestrian density and concentration per square meter can also be calculated. The system used the Scale Aggregation Network (SANet) as the baseline algorithm for crowd density detection, and the network structure is shown in Figure 4 [37]. Based on the innovation design paradigm, a multi-scale aggregation feature extraction encoder was constructed to improve the expression ability and scale diversity. The decoder is composed of convolution and a transpose convolution, and multi-scale features were fused to generate a density map of the same size as the input image. The strategy of combining the Euclidean loss function and the local consistency loss function was used to overcome the ambiguity of the generated graph, caused by the assumption that the pixels are independent of each other. Local consistency was calculated by Structural Similarity (SSIM) to measure the consistency between the generated and real density maps [38].

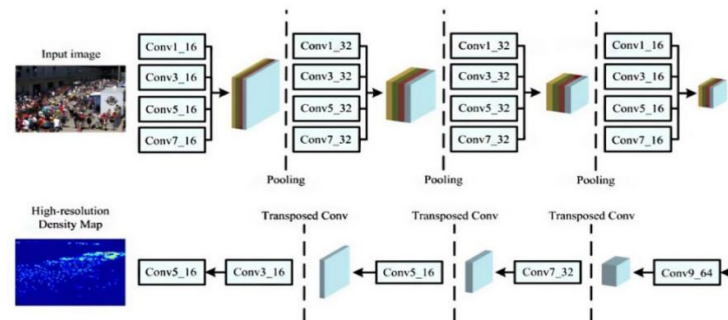


Figure 4. Structure diagram of crowd density detection model.

Although the crowd density estimation method takes spatial information into account, most of the output density maps have low resolution and lose a lot of detail. To generate a high-resolution density map, DME was used as a decoder. DME is composed of a series of convolutions and transposed convolutions. Four convolutions were used to improve the details of the feature map, step by step. Three transposed convolutions were used to repair the spatial resolution, and each transposed product doubled the size of the feature image. Regarding loss function design, SSIM and Euclidean distance were combined. SSIM was used to measure the consistency/similarity between the estimated density map and the real value, and three local statistical values: mean, variance, and covariance, were calculated. The SSIM range was  $-1$  to  $1$ , and the SSIM value was  $1$  when two images were the same. SSIM used the following method of calculation [39]:

$$\text{SSIM} = \frac{(2\mu_F\mu_Y + C_1)(2\sigma_{FY} + C_2)}{(\mu_F^2 + \mu_Y^2 + C_1)(\sigma_F^2 + \sigma_Y^2 + C_2)} \quad (1)$$

where  $\mu_F$  and  $\mu_Y$  are means,  $\sigma_F$  and  $\sigma_Y$  are variances,  $\sigma_{FY}$  is covariance,  $C_1$  and  $C_2$  are constants. The loss function of local consistency is as follows:

$$L_c = 1 - \frac{1}{N} \sum_x \text{SSIM}(x) \quad (2)$$

### 3.2.3. Face Mask Detection Module

In the new epidemic situation, reminding the public to wear masks is an effective means of maintaining public health and safety. This paper describes a system of judging whether people are wearing masks that has been improved and optimized based on the SSD algorithm. The SSD algorithm combines the anchor mechanism of faster R-CNN and the regression idea of YoLo, and improves the speed and accuracy [40–42]. The multi-scale convolution feature map was used to predict the object region, and a series of discrete and multi-scale default frame coordinates were output. The small convolution kernel was used to predict the coordinates of bounding boxes and the confidence of each category. We used the open-source lightweight face mask detection model as the baseline. The model is designed based on SSD architecture. The input image size was  $260 \times 260$ . The backbone network had eight layers, with a total of 28 convolution layers. Among them, the top eight convolution layers were the backbone maps; that is, the feature extraction layers, and the bottom 20 layers were the positioning and classification layers. The tagging information of all faces was read from the AIZOO open-source face mask dataset, the height-to-width ratio of each face was calculated, and the distribution histogram of the face aspect ratio was obtained. The normalized face aspect ratio was between 1 and 2.5. Therefore, according to the data distribution, we set the width-to-height ratio of the anchor in five positioning layers to 1, 0.62, and 0.42. The configuration information of the five positioning layers is shown in Table 2.

**Table 2.** Anchor parameter configuration of network location layer.

Convolution Layer	Feature Map Size	Anchor Size	Anchor Aspect Ratio
first floor	$33 \times 33$	0.04, 0.056	1, 0.62, 0.42
second floor	$17 \times 17$	0.08, 0.11	
third layer	$9 \times 9$	0.16, 0.22	
fourth floor	$5 \times 5$	0.32, 0.45	
fifth floor	$3 \times 3$	0.64, 0.72	

Moreover, to adapt the model to the UAV visual angle image, we used DJ-Innovations (DJI) air2 UAV to collect information about 300 UAV face mask detection datasets and labeled the faces with labeling software. On the model parameters trained by the AIZOO dataset, a low learning rate was set for transfer training. The subsequent processing mainly depended on the non-maximum suppression (NMS) method. We used a single class of NMS; that is, faces with a mask and faces without a mask, to perform NMS together to improve the speed. The loss function is defined as the weighted sum of location loss (LOC) and confidence loss (CONF):

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (3)$$

where  $N$  is the number of positive samples of the a priori box,  $c$  is the predictive value of category confidence,  $l$  is the position prediction value of the corresponding bounding box of the prior box, while  $g$  is the position parameter of the ground truth and the weight



coefficient  $\alpha$  is set to 1. For position error, the smooth L1 loss was adopted, which is defined as follows:

$$L_{loc}(x, l, g) = \sum_{i \in Pos}^N \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k smooth_{L1}(l_i^m - \hat{g}_j^m) \quad (4)$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (5)$$

where  $l_i^m$  is the offset between the prediction box and the prior box,  $\hat{g}_j^m$  represents the offset value between the actual prediction box and the prior box,  $Pos$  is the set of positive samples,  $(cx, cy)$  is the center coordinate, and  $w$  and  $h$  represent the width and height of the box. For the confidence error, softmax loss was used, which is defined as follows:

$$L_{conf}(x, c) = - \sum_{i \in Pos}^N x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg}^N \log(\hat{c}_i^o) \quad (6)$$

$$\text{where } \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$$

#### 4. Experimental Results and Discussion

To comprehensively evaluate the feasibility and superiority of the intelligent anti-epidemic patrol detection and alarm flight system, this paper carried out a UAV crowd density flight test and face mask detection test. The UAV physical display is shown in Figure 5.



**Figure 5.** UAV physical chart.

##### 4.1. Crowd Density Test

To comprehensively evaluate the crowd density detection method, we tested and analyzed it in the ShanghaiTech dataset and the real scenes. The ShanghaiTech dataset is composed of two sub-datasets: ShanghaiTech-A dataset is mainly dense crowds, including 300 training images and 182 test images, and the image size is not fixed. The minimum number of people in the image is 33, and the maximum number is 3139, with an average of 501. The ShanghaiTech-B dataset mainly collects images of relatively few people, and its image size is in pixels. The ShanghaiTech-B dataset mainly collects images of relatively few people, and its image resolution is  $768 \times 1024$  pixels. The dataset contains 400 training

images and 316 test images. There are at least nine targets and, at most, 578 targets in the images, with an average of 123 targets in each image. The dataset and real test results are shown in Figures 6 and 7. From the subjective effect, the crowd density detection algorithm used in this paper can correctly reflect the distribution of the crowd. In terms of quantitative indicators, the mean absolute error (MAE) and mean square error (MSE) of the crowd density method in Shanghai crowd-counting datasets A and B was better than other methods, reaching a higher level. The results of the airborne test on UAV showed that the algorithm can accurately detect the location of ground crowds. The performance comparison results of the different detection algorithms for crowd density are shown in Table 3.

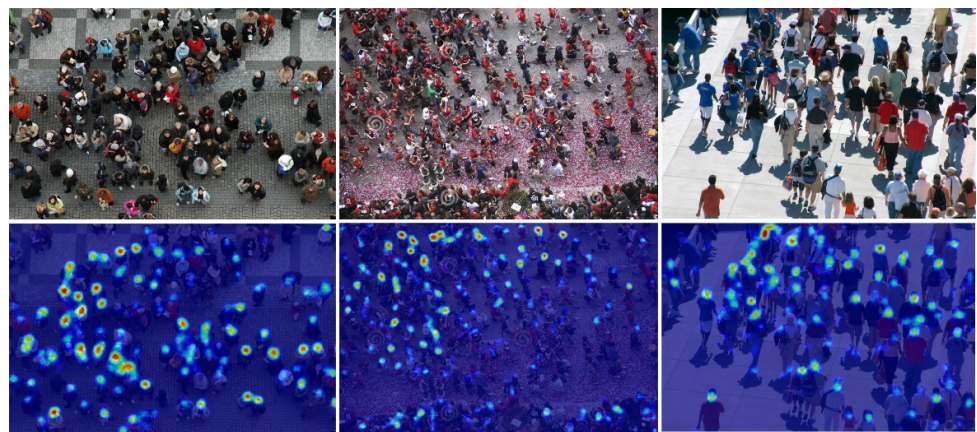


Figure 6. Test results of the crowd density test dataset.

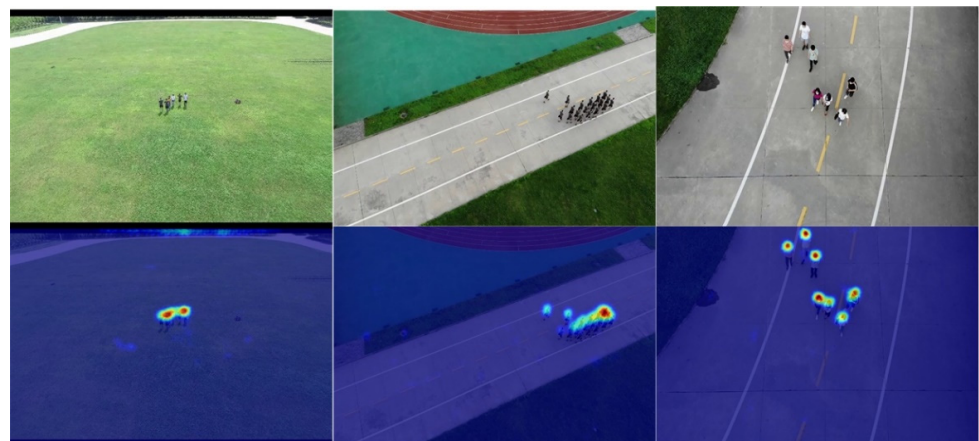


Figure 7. Real test results of crowd density detection.

Table 3. Performance comparison of crowd density detection algorithms.

Method	Part A		Part B	
	MAE	MSE	MAE	MSE
MCNN	110.2	173.2	26.4	41.3
Cascaded-MTL	101.3	152.4	20.1	31.1
Switch-CNN	90.4	135	21.6	33.4
CP-CNN	73.6	106.4	20.1	30.1
CSRNet	68.2	115	10.6	16
SANet	67	104.5	8.4	13.6

#### 4.2. Face Mask Detection Test

To comprehensively evaluate the face mask detection methods, we tested them in datasets and real scenes, respectively. On the constructed UAV face mask detection dataset, the performances of lightweight object detection models such as YOLOv3 tiny, Yolo nano, and PVA-Net were tested. The evaluation indexes are mean average precision (mAP) and frames per second (FPS). The statistical results are shown in Table 4. It can be seen that the algorithm used in this paper has high mAP, FPS, precision, and recall.

**Table 4.** Performance comparison of face mask detection algorithms in dataset experiments.

Method	mAP/%	FPS	Precision/%	Recall/%
YOLOv3-tiny	58.47	20.1	68.2	30.2
YOLO Nano	54.85	22.7	66.3	35.2
PVA-Net	51.67	23.7	67.5	36.5
Our	61.42	22.5	70.2	30

To test the performance of the face mask detection algorithm in real scenes, the experiment was carried out on the UAV with Xavier NX. By reading the ground face images collected by a Pan-Tilt-Zoom (PTZ) camera, the wearing of masks was detected. When there was a face without a mask, the face was marked with a red box, and the face with a mask was marked with a green box. The test scenes of the experimental line were a way, a campus playground, a subway station exit, etc. The UAV reads the ground face images collected by the PTZ camera to detect the wearing of masks. When there was a face not wearing a mask, the face was marked with red, and faces correctly wearing a mask were marked with a green box. When there were special circumstances, such as pedestrians with their heads down or wearing masks improperly, the face was marked red, as if they were not wearing masks. As can be seen from Figure 8, the algorithm of this project still had high accuracy in the test on real scenes.



**Figure 8.** Cont.





Figure 8. Real test results of crowd density detection.

## 5. Conclusions

To carry out intelligent and efficient inspection of the epidemic area, this paper designed an intelligent inspection and alarm flight system for epidemic prevention. The system was a small UAV platform with comprehensive functions and high integration. It combined UAV autonomous navigation, deep learning, and other technologies, and can be used for crowd-gathering monitoring and alarm, monitoring the wearing of masks among pedestrians, and epidemic prevention policies propaganda. It can also safely and efficiently complete the task of intelligent epidemic prevention and control detection and alarm. The experimental results show that the system has good practicability, which can maximize epidemic prevention and control. In the future, the platform can also be equipped with small infrared temperature-measuring equipment and adopt a multi-machine cooperation mode to realize an omni-directional inspection of people's body temperature outdoors, to ensure that abnormalities are detected as early as possible to block the spread of the epidemic.

**Author Contributions:** Conceptualization, X.Y., J.F. and R.L.; Methodology, J.F. and R.L.; Software, X.X.; Investigation, W.L. and J.F.; Resources, X.X.; Writing—original draft preparation, J.F. and R.L.; Writing—review and editing, X.Y., J.F. and W.L.; Visualization, J.F.; Supervision, J.F. and X.X.; Project administration, X.Y.; Funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 61806209, in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JQ-490, in part by the Aeronautical Science Fund under Grant 201851U8012. (Corresponding author: Xiaogang Yang).

**Data Availability Statement:** The datasets used or analysed during the current study are available from the corresponding author on reasonable request.

**Acknowledgments:** The authors are grateful to Qingge Li for her help with the preparation of figures in this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhan, J.; Liu, Q.S.; Sun, Z.; Zhou, Q.; Hu, L.; Qu, G.; Zhang, J.; Zhao, B.; Jiang, G. Environmental impacts on the transmission and evolution of COVID-19 combining the knowledge of pathogenic respiratory coronaviruses. *Environ. Pollut.* **2020**, *267*, 115621. [[CrossRef](#)] [[PubMed](#)]
2. Adhikari, S.P.; Meng, S.; Wu, Y.J.; Mao, Y.P.; Ye, R.X.; Wang, Q.Z.; Sun, C.; Sylvia, S.; Rozelle, S.; Raat, H.; et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: A scoping review. *Infect. Dis. Poverty* **2020**, *9*, 1–12. [[CrossRef](#)] [[PubMed](#)]
3. Liu, Y.; Liu, Z.; Shi, J.; Wu, G.; Chen, C. Optimization of Base Location and Patrol Routes for Unmanned Aerial Vehicles in Border Intelligence, Surveillance, and Reconnaissance. *J. Adv. Transp.* **2019**, *2019*, 9063232. [[CrossRef](#)]
4. Choi, H.; Geeves, M.; Alsalam, B.; Gonzalez, F. Open source computer-vision based guidance system for UAVs on-board decision making. *IEEE Aerosp. Conf. Proc.* **2016**, *2016*, 1–5. [[CrossRef](#)]
5. Guan, X.; Huang, J.; Tang, T. Robot vision application on embedded vision implementation with digital signal processor. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 17. [[CrossRef](#)]
6. Khuc, T.; Nguyen, T.A.; Dao, H.; Catbas, F.N. Swaying displacement measurement for structural monitoring using computer vision and an unmanned aerial vehicle. *Meas. J. Int. Meas. Confed.* **2020**, *159*, 107769. [[CrossRef](#)]
7. Barisic, A.; Car, M.; Bogdan, S. Vision-based system for a real-time detection and following of UAV. *Int. Work. Res. Educ. Dev. Unmanned. Aer. Syst. RED-UAS* **2019**, *2019*, 156–159. [[CrossRef](#)]
8. Yang, X.; Liao, L.; Yang, Q.; Sun, B.; Xi, J. Limited-energy output formation for multiagent systems with intermittent interactions. *J. Frankl. Inst.* **2021**. [[CrossRef](#)]
9. Xi, J.; Wang, C.; Yang, X.; Yang, B. Limited-budget output consensus for descriptor multiagent systems with energy constraints. *IEEE Trans. Cybern.* **2020**, *50*, 4585–4598. [[CrossRef](#)]
10. Xi, J.; Wang, L.; Zheng, J.; Yang, X. Energy-constraint formation for multiagent systems with switching interaction topologies. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2020**, *67*, 2442–2454. [[CrossRef](#)]
11. Ağalar, C.; Öztürk Engin, D. Protective measures for covid-19 for healthcare providers and laboratory personnel. *Turkish J. Med. Sci.* **2020**, *50*, 578–584. [[CrossRef](#)]
12. Wu, Q.; Zhou, Y.; Wu, X.; Liang, G.; Ou, Y.; Sun, T. A Real-time Running Detection System for UAV Imagery based on Optical Flow and Deep Convolutional Networks. *IET Intell. Transp. Syst.* **2020**, *14*, 278–287. [[CrossRef](#)]
13. Athilingam, R.; Kumar, K.S.; Thillainayagi, R.; Hameedha, N.A. Moving Target Detection Using Adaptive Background Segmentation Technique for UAV based Aerial Surveillance. *J. Sci. Ind. Res.* **2014**, *73*, 247–250.
14. Briese, C.; Seel, A.; Andert, F. Vision-based detection of non-cooperative UAVs using frame differencing and temporal filter. In Proceedings of the 2018 International Conference on Unmanned Aircraft Systems, Dallas, TX, USA, 12–15 June 2018; pp. 606–613.
15. Xu, X.; Duan, H.; Guo, Y.; Deng, Y. A Cascade Adaboost and CNN Algorithm for Drogue Detection in UAV Autonomous Aerial Refueling. *Neurocomputing* **2020**, *408*, 121–134. [[CrossRef](#)]
16. Xia, L.; Zhang, R.; Chen, L.; Huang, Y.; Xu, G.; Wen, Y.; Yi, T. Monitor Cotton Budding Using SVM and UAV Images. *Appl. Sci.* **2019**, *9*, 4312. [[CrossRef](#)]
17. Shao, Y.; Mei, Y.; Chu, H.; Chang, Z.; He, Y.; Zhan, H. Using infrared HOG-based pedestrian detection for outdoor autonomous searching UAV with embedded system. In Proceedings of the Ninth International Conference on Graphic and Image Processing, Chengdu, China, 12–14 December 2018.
18. Qiu, S.; Wen, G.; Fan, Y. Occluded Object Detection in High-Resolution Remote Sensing Images Using Partial Configuration Object Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1909–1925. [[CrossRef](#)]
19. Sindagi, V.A.; Patel, V.M. A Survey of Recent Advances in CNN-based Single Image Crowd Counting and Density Estimation. *Pattern Recognit. Lett.* **2017**, *107*, 3–16. [[CrossRef](#)]
20. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)]
21. Lu, R.; Yang, X.; Li, W.; Fan, J.; Li, D.; Jing, X. Robust Infrared Small Target Detection via Multidirectional Derivative-Based Weighted Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2020**. [[CrossRef](#)]



22. Lu, R.; Yang, X.; Jing, X.; Chen, L.; Fan, J.; Li, W.; Li, D. Infrared Small Target Detection Based on Local Hypergraph Dissimilarity Measure. *IEEE Geosci. Remote. Sens. Lett.* **2020**. [\[CrossRef\]](#)
23. Fang, L.; Zhiwei, W.; Anzhe, Y.; Xiao, H. Multi-Scale Feature Fusion Based Adaptive Object Detection for UAV. *Acta Opt. Sin.* **2020**, *40*, 127–136. [\[CrossRef\]](#)
24. Lai, Y.C.; Huang, Z.Y. Detection of a Moving UAV Based on Deep Learning-Based Distance Estimation. *Remote. Sens.* **2020**, *12*, 3035. [\[CrossRef\]](#)
25. Avola, D.; Cinque, L.; Diko, A.; Fagioli, A.; Foresti, G.L.; Mecca, A.; Pannone, D.; Piciarelli, C. MS-Faster R-CNN: Multi-Stream Backbone for Improved Faster R-CNN Object Detection and Aerial Tracking from UAV Images. *Remote Sens.* **2021**, *13*, 1670. [\[CrossRef\]](#)
26. Yingjie, L.; Fengbao, Y.; Peng, H. Parallel FPN Algorithm Based on Cascade R-CNN for Object Detection from UAV Aerial Images. *Laser Optoelectron. Prog.* **2020**, *57*, 201505. [\[CrossRef\]](#)
27. Yang, Y.; Gong, H.; Wang, X.; Sun, P. Aerial target tracking algorithm based on faster R-CNN combined with frame differencing. *Aerospace* **2017**, *4*, 32. [\[CrossRef\]](#)
28. Boudjit, K.; Ramzan, N. Human detection based on deep learning YOLO-v2 for real-time UAV applications. *J. Exp. Theor. Artif. Intell.* **2021**, 1–18. [\[CrossRef\]](#)
29. Sadykova, D.; Pernebayeva, D.; Bagheri, M.; James, A. IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV Imaging. *IEEE Trans. Power Deliv.* **2019**, *35*, 1599–1601. [\[CrossRef\]](#)
30. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [\[CrossRef\]](#)
31. Ding, L.; Xu, X.; Cao, Y.; Zhai, G.; Yang, F.; Qian, L. Detection and tracking of infrared small target by jointly using SSD and pipeline filter. *Digit. Signal Process.* **2020**, *110*, 1–9. [\[CrossRef\]](#)
32. Li, G.; Ren, P.; Lyu, X.; Zhang, H. Real-Time Top-View People Counting Based on a Kinect and NVIDIA Jetson TK1 Integrated Platform. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 15 December 2016; IEEE: Piscataway, NJ, USA, 2017; pp. 468–473.
33. Wood, S.U.; Rouat, J. Real-time Speech Enhancement with GCC-NMF: Demonstration on the Raspberry Pi and NVIDIA Jetson. In Proceedings of the Interspeech 2017 Show and Tell Demonstrations, Stockholm, Sweden, 20–24 August 2017; pp. 2048–2049.
34. Mittal, S. A Survey on optimized implementation of deep learning models on the NVIDIA Jetson platform. *J. Syst. Archit.* **2019**, *97*, 428–442. [\[CrossRef\]](#)
35. Jang, Y.; Gunes, H.; Patras, I. Registration-free Face-SSD: Single shot analysis of smiles, facial attributes, and affect in the wild. *Comput. Vis. Image Underst.* **2019**, *182*, 17–29. [\[CrossRef\]](#)
36. Hwang, C.L.; Wang, D.S.; Weng, F.C.; Lai, S.L. Interactions between Specific Human and Omnidirectional Mobile Robot Using Deep Learning Approach: SSD-FN-KCF. *IEEE Access* **2020**, *8*, 41186–41200. [\[CrossRef\]](#)
37. Fan, H.; Ling, H. SANet: Structure-Aware Network for Visual Tracking. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Work, Honolulu, HI, USA, 21–26 July 2017; pp. 2217–2224. [\[CrossRef\]](#)
38. Channappayya, S.S.; Bovik, A.C.; Heath, R.W. Rate bounds on SSIM index of quantized images. *IEEE Trans. Image Process.* **2008**, *17*, 1624–1639. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Fei, L.; Yan, L.; Chen, C.; Ye, Z.; Zhou, J. OSSIM: An Object-Based Multiview Stereo Algorithm Using SSIM Index Matching Cost. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6937–6949. [\[CrossRef\]](#)
40. Sun, C.; Ai, Y.; Wang, S.; Zhang, W. Mask-guided SSD for small-object detection. *Appl. Intell.* **2020**, *51*, 3311–3322. [\[CrossRef\]](#)
41. Hu, K.; Lu, F.; Lu, M.; Deng, Z.; Liu, Y. A Marine Object Detection Algorithm Based on SSD and Feature Enhancement. *Complexity* **2020**, *2020*, 5476142. [\[CrossRef\]](#)
42. Liu, X.; Li, Y.; Shuang, F.; Gao, F.; Zhou, X.; Chen, X. Issd: Improved ssd for insulator and spacer online detection based on uav system. *Sensors* **2020**, *20*, 6961. [\[CrossRef\]](#) [\[PubMed\]](#)