



Research article

An airport apron ground service surveillance algorithm based on improved YOLO network

Yaxi Xu¹, Yi Liu², Ke Shi^{3,*}, Xin Wang^{4,*}, Yi Li⁴ and Jizong Chen⁴

¹ School of Economics and Management, Civil Aviation Flight University of China, Guanghan 618307, China

² Department of Big Data and Artificial Intelligence, Civil Aviation Management Institute of China, Beijing 100102, China

³ School of Civil Aviation Supervisor Training, Civil Aviation Flight University of China, Guanghan 618307, China

⁴ School of Computer Science, Civil Aviation Flight University of China, Guanghan 618307, China

* **Correspondence:** Email: cafucsk@163.com, cafucwx@163.com.

Abstract: To assure operational safety in the airport apron area and track the process of ground service, it is necessary to analyze key targets and their activities in the airport apron surveillance videos. This research shows an activity identification algorithm for ground service objects in an airport apron area and proposes an improved YOLOv5 algorithm to increase the precision of small object detection by introducing an SPD-Conv (spath-to-depth-Conv) block in YOLOv5's backbone layer. The improved algorithm can efficiently extract the information features of small-sized objects, medium-sized objects, and moving objects in large scenes, and it achieves effective detection of activities of ground service in the apron area. The experimental results show that the detection average precision of all objects is more than 90%, and the whole class mean average precision (mAP) is 98.7%. At the same time, the original model was converted to TensorRT and OpenVINO format models, which increased the inference efficiency of the GPU and CPU by 55.3 and 137.1%, respectively.

Keywords: airport apron; object detection; service activity recognition; intelligent video surveillance

1. Introduction

In recent years, China's civil aviation industry has experienced rapid development. The CAAC (Civil Aviation Administration of China) has put forward the "Plan of the Civil Aviation Development of China". The plan aims to take full advantage of airport and other civil aviation construction facilities, combining them with big data and new generation information technology, such as artificial intelligence. The goal is to transition from a large civil aviation country to a strong one, with a focus on building a smart civil aviation system. Smart travel is a crucial component of this technology system. The "Plan of the Civil Aviation Development of China" places great emphasis on how to achieve this. The development of artificial intelligence technology and enhanced computability have led to increased interest in intelligent monitoring of video data streams using machine learning and deep learning algorithms across various fields. The use of digital and intelligent technology to improve digital perception, data decision-making, and lean management capabilities of civil aviation is a topic worthy of in-depth study in the field. The "Roadmap for the Construction of Intelligent Civil Aviation" issued by the CAAC in 2022 proposes focusing on the construction of four types of airports; strengthening the service capacity of airport flights, passengers, and freight transportation; promoting the development of airport operational synergies, operational intelligence, and digital infrastructure construction; and enhancing the supportability of airports, service levels, and operational efficiency. Using intelligent algorithms to automatically identify the nodes of ground protection from airport monitoring videos and track the chronological status of protection operations is an effective approach to improving the lean management of flight protection. This is also an important direction for developing smart airports.

The process of ground service during flight transit refers to the taxiing of the aircraft to the parking space after it enters the harbor. The process begins with setting the wheel block and ends with its removal after a series of services have been performed on the aircraft. The flight transit ground monitoring video contains various airport scene objects, such as aircraft, staff, tractor-trailers, etc. The airport involves various ground service activities, including wheel blocking, baggage loading and unloading, refueling, loading of flight food, etc. These activities affect the allocation of airport personnel, the scheduling of airport resources, and passengers' travel routes. Additionally, they reflect the economic and operational efficiency of the airport, as well as the passengers' experience. These factors reflect the airport's economic efficiency, operational effectiveness, and passenger experience.

With the recovery of civil aviation traffic after the epidemic and the trend of increasing year by year [1], the increase in the number of outbound flights leads to the increase in the security pressure of airport flights over the station. Some airports have reached capacity, and their existing facilities cannot meet development needs. As a result, the issue of inadequate flight protection management at airports has become more prominent. The airspace resources at busy airports and on busy air routes are limited, resulting in restricted flight operations and increased delays. This has impacted the overall performance of airport facilities and systems, hindering their full utilization. Flight transit security delays account for 15.45% of all delays at large hub airports in China. In developed countries like those in Europe and the United States, flight transit security delays are the second largest cause of airport delays, accounting for 25% of all delays. Flight grounds have a direct impact on the punctuality of flights, the operational efficiency of airports, and the overall passenger experience of airport services. Currently, the data for each protection operation node is recorded manually, resulting in excessive manual intervention and inadequate automation. As a result, completion of each protection operation

link and resource scheduling cannot be obtained in a timely manner. This leads to inappropriate ground protection operation process connection and time planning, reducing the efficiency of flight transit protection and resulting in the wastage of protection resources and increased operation costs. Flight delays result in economic losses for airlines and cause passenger anxiety. Ground service in the apron area has a significant impact on airport operational efficiency. Inefficient ground service often leads to flight delays. The efficient use of limited transit time and airport security resources to reduce flight delays caused by ground service is a crucial issue to be addressed in current airport operations and security. This can help improve the punctuality of transit flights.

This paper proposes a new algorithm for monitoring ground service processes in airport apron areas using service activity recognition. We analyze how to divide the monitoring video of the airport's apron area, including the docking of the aircraft food delivery truck, the docking of the aircraft pushback tug, and the opening/closing of the cargo door. An improved, optimized small object detection algorithm is used to protect key actions throughout the process and extract the iconic behavioral actions of aircraft. The flowchart of the algorithm in this paper is shown in Figure 1. In this study, we first collected the actual surveillance video of the ground service process and annotated the object detection dataset. Second, we divided the dataset into the training dataset and the test dataset. Lastly, we improved the YOLOv5 algorithm network and trained the improved algorithm to get the object detection model.

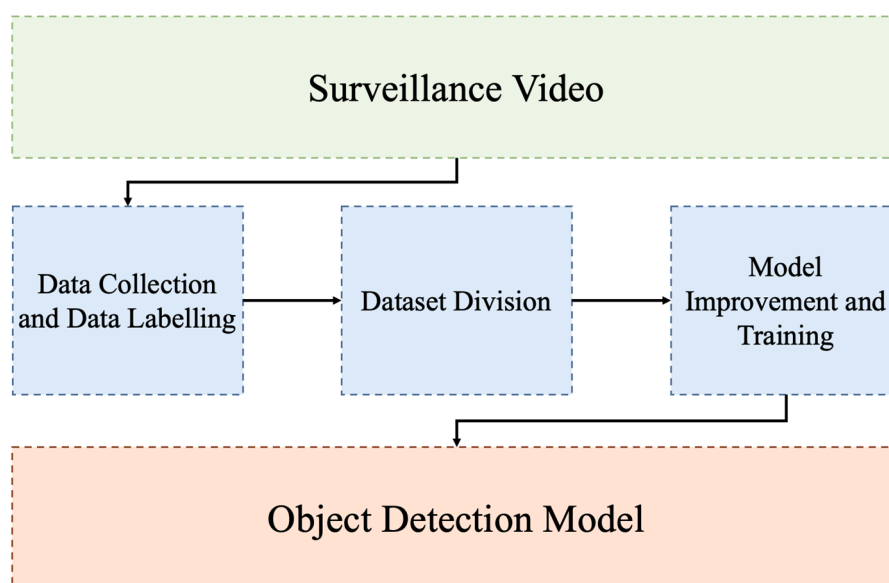


Figure 1. The process of the whole algorithm.

The information extracted from the monitoring video can effectively support and guarantee the deployment and allocation of resources at the airport. The main contributions of this paper and the differences from other papers are shown below:

- In comparison to similar studies, this paper shifts the focus of image labelling from the objects present in the field, such as people and types of cars, to the specific activities delineated by the ground service nodes. This approach avoids the consumption of computational resources and the loss of computational time caused by the need for secondary analysis of the detected objects in similar studies

to identify the process node. At the same time, it is more closely aligned with the method of ground service node delineation outlined in the CAAC's regulatory documents.

- This paper presents an improved algorithm network based on YOLOv5s for efficient and precise extraction of information related to ground service nodes in the apron area. The algorithm is specifically designed for small object detection, making it more suitable for this task.

2. Related works

Detection algorithms are commonly categorized into two types: two-stage algorithms and single-stage algorithms. Both types take a set of images and corresponding object labels as input and output a set of bounding boxes. The two-stage object detection algorithm divides the detection problem into a stage of generating candidate regions and a stage of obtaining the final bounding box based on the secondary correction of the candidate regions. Although it is more precise than the single-stage algorithm, it is also slower. Single-stage object detection algorithms require only one feature extraction to achieve object detection, providing a speed advantage over two-stage algorithms. With a suitable dataset, their precision performance is comparable to that of two-stage algorithms. YOLO (You Only Look Once) [2] uses the entire graph as input to the network and global features to improve object detection, all within the convolutional neural network (CNN). To perform the object detection task, the CNN directly regresses the location of the bounding box and its class. CornerNet [3] detects a pair of corners of the bounding box and groups them to form the final detected bounding box. FCOS (Fully Convolutional One-Stage) [4] predicts the bounding box by using all the points in the real bounding box. The industry prefers single-stage object detection algorithms over two-stage object detection algorithms because they are fast and precise. Among them, the YOLO series of algorithms are particularly popular object detection algorithms due to their lightness and efficiency. For example, Liu et al. [5] detected leather defects based on YOLOv5 and had some success. Choi et al. [6] proposed a new approach called Gaussian YOLOv3, which is suitable for autonomous driving applications. Li et al. [7] suggested using YOLOv3 for detecting foreign object debris on airport runways, which is of significant importance for ensuring flight safety. Shi et al. [8] combined a channel attention mechanism and YOLOv5 to finish small object detection, enabling wide-range detection of birds during aircraft flight in airport airspace. Cai et al. [9] proposed real-time detection of surface cracks on bridges based on YOLOv3.

Surveillance video contains a wide variety of scene information, such as the subject of surveillance and its activities. With the iteration of computer hardware and the rapid development of artificial intelligence technology, more and more researchers are adopting computer vision technology to analyze surveillance images and videos, and use the information obtained from the analysis method to finish the tasks about resource scheduling and safety support for airports, governmental and corporate organizations, roads, and other areas [10,11]. Therefore, the development of intelligent monitoring technology for specific scenarios has become the current research direction of many scholars and application engineers. For example, Cai et al. [12] combined deep learning and traffic surveillance to design a vehicle counting and classification model for intelligent transportation systems (ITS). Kumar et al. [13] used computer vision technology to detect human movement activity based on thermal surveillance video. Raza et al. [14] proposed a novel system based on machine learning and image processing to provide an efficient surveillance system for pedestrian detection and tracking at night.

Table 1. Research progress on airport apron area analysis algorithm.

Related work	Description	Methods	Feature
James et al. [11]	A comprehensive airport apron scene automatic visual monitoring system.	The integration of scene tracking, motion detection, object tracking, data fusion, and video event recognition methods enables scene understanding and video event recognition, thereby improving the effectiveness of airport apron monitoring.	The system uses bottom up feature tracking methods, such as the Kanade Lucas Tomasi (KLT) algorithm. It also includes a data fusion module and dynamic context. Cognitive vision techniques based on spatio-temporal reasoning are employed.
Sabine et al. [15]	A comprehensive framework for assessing the overall risk of airport ground operations, taking into account the actions of all relevant stakeholders.	A process model for surface operations was developed, which implemented triangulation. A reference dataset was used to determine the database. A macroscopic scenario tool was introduced to support SMS change management, training and education, and safety communication functionalities.	Developing and implementing an effective safety management system.
Lu et al. [16]	Improve aircraft turnaround service processes by analyzing and monitoring airport gate activity.	Applying machine learning and computer vision techniques to airport camera video surveillance data processing.	Detects a variety of gate events/activities including aircraft arrivals, aircraft departures, and aircraft taxi starts. Real time monitoring mode.
Ying et al. [17]	A comprehensive analysis and evaluation have been conducted on the personnel and equipment involved in ramp control operations, the environmental factors encountered during ramp control operations, and the existing ramp control procedures.	The analytic hierarchy process (AHP) and expert rating method have been utilized to establish an evaluation method suitable for assessing the operational support capabilities of civil aviation	The proposed method can provide guidance and reference for ensuring the orderly operation of ramp control.

Continued on next page

Related work	Description	Methods	Feature
Van Phat et al. [18]	Predicting aircraft transit times based on vision algorithms for monitoring turnaround process analysis.	Monitoring of aircraft turnaround process (object detection, object tracking, activity detection, and push-back prediction) using a video analytics framework based on convolutional neural networks.	Video analysis framework. Aircraft model identification. Real time activity detection and display.
Gorkow [19]	Aircraft turnaround management application based on computer vision.	Deep learning based object detection method.	Detection of overflight activities and objects.
Wang et al. [20]	Improvement of aircraft ground service efficiency by tracking ground service equipment.	Real time high precision tracking device consisting of a kinematic (RTK) unit and a heading unit.	Real time tracking of ground service equipment, collision detection, and scheduling of ground service equipment.
Wang et al. [21]	Real time Onboard Positioning and Heading System (ROPHS) design to obtain real time status of various types of ground service equipment (e.g., multi car logistics trains) through equipment.	ROPHS hardware with geometry-based recursive algorithms to compute tractor-trailer position. Model based tracking algorithms to compute position and real time velocity.	Real time status, including the exact position and speed of any point on the multi car logistics train, as well as the variable number of traction vehicles and the order in which they are connected at any given time of travel.
Yıldız et al. [22]	Turnaround control system for automatic detection and monitoring of airport ground service movements. & Deep learning and computer vision based approach for airport ground service detection and tracking.	Deep learning and computer vision based approach for airport ground service detection and tracking.	High precision target tracking and service detection. Real time data processing/analysis. Time-stamped cycles for ground service.
Ours	Automatic monitoring, analysis and alerting of airport ground service status based on airport transit surveillance video.	Detection and tracking of airport ground services based on big data vision methods.	Real time data processing and analysis. High precision status labeling.

Airport apron area analysis and safety problems are key research issues in the civil aviation industry. The application of intelligent technology in the airport apron area has become a hot research direction in recent years. It mainly improves the airport apron area operating efficiency through assessing hardware equipment and facilities parameters, collating and analyzing text-based data, and monitoring video data and other various data. This paper will explore these research directions, as shown in Table 1.

In addition to the papers described in the table above, Thai et al. [23] used convolutional neural networks for airport-airside surveillance. Liu et al. [24] used CNN to identify small objects in airport clearances, such as unmanned aerial vehicles (UAVs) and bird flocks. These researchers studied the aircraft turnaround process in monitoring videos of airport apron areas for different objects. However, in the existing studies on airport apron ground service surveillance algorithms, most researchers tend to divide the algorithm into two parts or add some preprocessing methods to the algorithm to optimize the analysis result. For example, the algorithm is divided into the object detection stage and the activity analysis stage of the detection object, and some researchers choose to divide the regions in the video to limit the scope of detection to obtain more effective detection results and active analysis objects. For example, Van Phat et al. [18] analyzed the running time of transit aircraft activities and predicted aircraft launching time by detecting targets such as aircraft, fuel trucks, fuel lines, and baggage carriers. They also analyzed boarding time by considering the position and speed of the aircraft, and by detecting targets such as aircraft and passenger elevator trucks. Gorkow M. et al. [19] determined critical conditions by detecting targets such as aircraft and passenger elevator cars. For example, when there are at least two people on the stairs, it is considered that boarding has started.

The “Technical Specification for Airport Collaborative Decision Making System” [25] document, proposed by the Civil Aviation Administration of China and released on February 18, 2022, identifies forty-five nodes during the ground service period. These include aircraft arrival, cargo door open/close, and refuel, among others. Compared to directly detecting the object and its state (such as aircraft, refueling truck, corridor bridge, etc.), the document proposed by CAAC suggests specific activities to divide the nodes of ground service, such as opening and closing passenger doors, cargo load, and more. Therefore, to differentiate from currently commonly used annotation methods, this paper chooses to annotate the activities rather than just annotate the main objects when annotating the surveillance video dataset. Compared with the common methods, this annotation method is more suitable for the division of processes in documents, on the one hand. On the other hand, it can save the computing resources of the whole process because it does not need to conduct a secondary analysis of activities.



Figure 2. Sample image of airport apron area surveillance dataset.

3. Data processing

3.1. Data collection

This paper uses actual surveillance video of ground service process from multiple high throughput airports in China. The experiment includes diverse datasets to ensure the reliability and usability of the collected monitoring data, such as different weather conditions (sunny/rainy) and light conditions (day/night/dusk/early morning). The collected dataset cases are shown in Figure 2. A dataset of 12 flights was collected experimentally. As the apron surveillance video involves a lengthy waiting process, there may be a significant number of duplicated images in the saved images after frame extraction. To address this, we filtered out the duplicated images by randomly selecting one frame every 5 seconds. The filtered images were then grouped together after disruption and used for testing. The dataset consisted of a total of 15,118 images, with 12,105 used for training and 3013 for testing. The resolution of the images is 1920×1080 . The distribution and production of the dataset is shown in Figure 3.

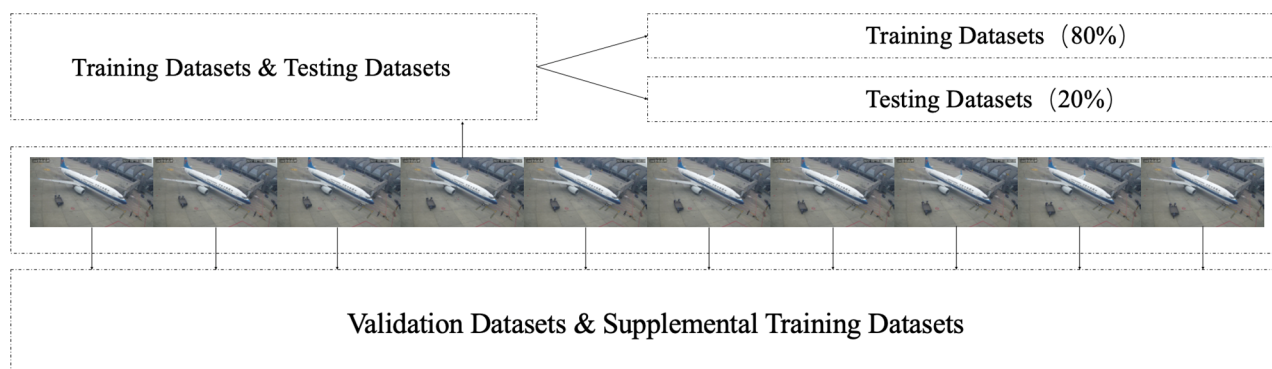


Figure 3. Data collection processing.

3.2. Data annotation approach based on service activity

Regarding labeling methods, most researchers use the target labeling method, which can cause issues when analyzed by algorithms in the future. This is because the appearance of the main target in the surveillance video does not necessarily mean that the corresponding service activity has begun. For example, when a tank refueller appears around the aircraft or in the work area but the fuel line is not yet connected to the aircraft, the refueling process has not truly started. Therefore, this type of labelling method based on targets typically requires numerous algorithms to determine the primary monitoring area or to identify the target in the observation area visually. Subsequently, other judgement algorithms are used for further division, which inevitably results in a significant waste of computational resources, either directly or indirectly. To address the aforementioned issues, this paper classifies ground service activities based on the judgement conditions of different nodes during the ground service period. Table 2 below shows the 16 categories that can be directly determined by the activity detection results. Figure 4 illustrates the dataset's labelling cases. Figure 5 shows the cases of each category.

Table 2. Activity categories and judgment characteristics table. (The label name in the figure is the label name, not the exclusive noun of this category.)

Category number	Label name	Category description
1	Bridge Connected	Boarding bridges connect with passenger door
2	Cargo Back Door Open	Cargo back door (external) opens/closes 45 degrees
3	Cargo Front Door Open	Cargo front door (external) opens/closes 45 degrees
4	Passenger Door Open	Passenger door opens/closes 45 degrees
5	Catering Truck Connected	Aircraft catering vehicle raised to visibility
6	Cargo Loading	Lifting platform vehicle raises to level with hatch door
7	Push Back Tractor Connected A	Tow bar overlaps wheels
8	Push Back Tractor Connected B	Clamp overlaps wheels
9	Baggage Loading	Front section of self-propelled conveyer-belt loader raises closely to hatch door
10	Fueling A	Fuel line connection
11	Cone	Cone is placed
12	Airplane	Aircraft head appears
13	Cargo Back Door Open B	Cargo back door (inner) opens/closes 45 degrees
14	Cargo Front Door Open B	Cargo front door (inner) opens/closes 45 degrees
15	Baggage Loading B	Baggage towing tractor raises to level with hatch
16	Fueling B	Fuel stepladder connection

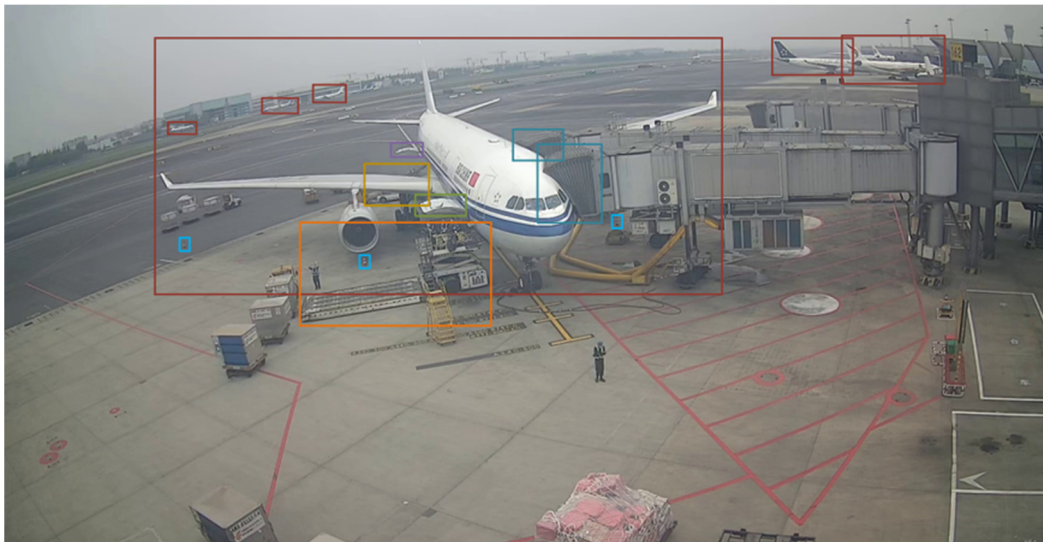


Figure 4. Annotation sample of the dataset.

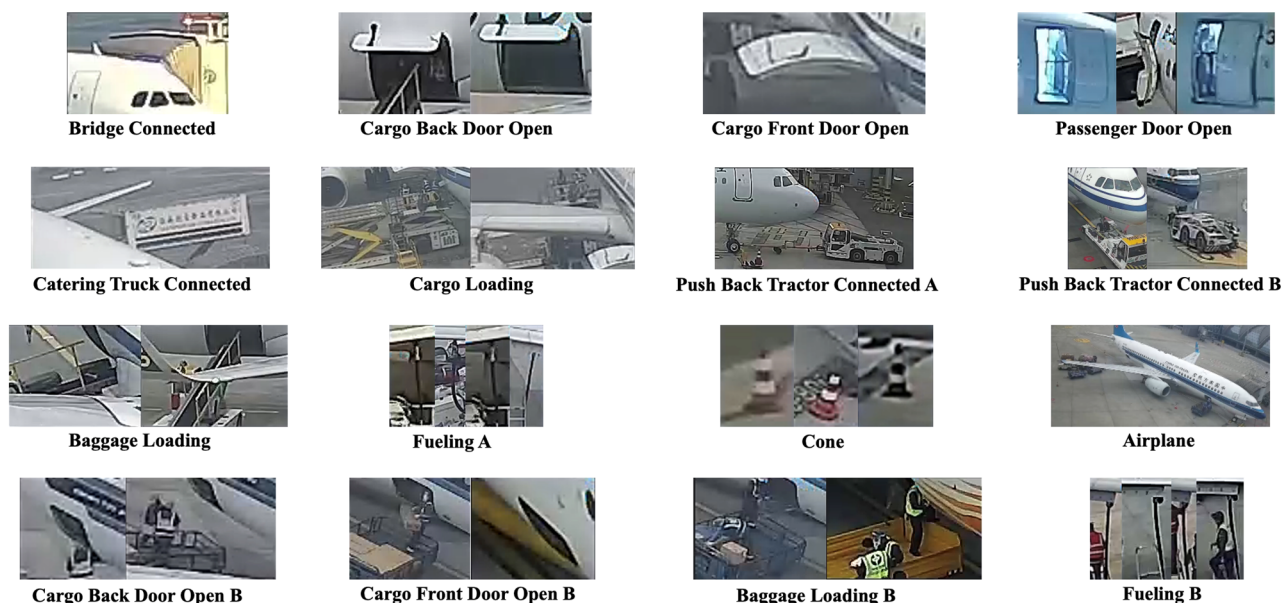


Figure 5. Cases of airport for each category.

4. Improved YOLOv5 algorithm

The algorithms used for apron area surveillance analysis are typically deployed on surveillance cameras or cloud computing servers for high-volume computation. Therefore, the model must be simple to use, fast, and precise. It should also support acceleration schemes such as ONNX or TensorRT as much as possible. In summary, this paper's benchmark model selection of single-stage object detection algorithms in the YOLOv5 series is based on its precision, speed, and hardware/cloud computing server deployment needs for object detection. The YOLOv5 models, including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, come in four different sizes. As the model deepens, the detection precision improves, but the detection speed decreases. To achieve a balance between detection precision and speed, this paper adopts the YOLOv5s network as the basic network. The network structure of YOLOv5s is divided into five parts: input, backbone, neck layer, detection head, and output. The backbone layer extracts features, the neck layer fuses features, and the head layer analyzes and predicts types.

4.1. Small object detection improvements based on SPD-Conv

One drawback of current convolutional neural network designs is the loss of information and increased difficulty in learning feature representations when using stride convolution and pooling layers. When the object's size has high resolution and a certain size, there may be redundant information. In traditional convolutional neural networks, these objects would still be detected even if some feature information is lost. The model can still be well trained and learn features through the remaining information. However, if the target is a small object, there will be minimal redundant information. Using this approach may cause the model to lose its ability to learn features.

There are challenges in monitoring in the airport apron area, which involves processing a large amount of visual information. The main targets are small objects and rich in detailed information such

as cones, objects with connected states, etc. Therefore, the effectiveness of the airport field monitoring video is highly dependent on the ability to capture details completely. To improve the precision of detection results, it is important to minimize information loss. In this network, a non-stride convolution (SPD-Conv) [26] is introduced to the YOLOv5 architecture.

The structure of SPD-Conv is shown in Figure 6. For an input image of $S \times S \times C_1$, it is divided into several sub-blocks, and the size of the sub-block is $\frac{S}{stride} \times \frac{S}{stride} \times C_1$. After concatenating sub-blocks along channels to get the final feature map, its size is $\frac{S}{stride} \times \frac{S}{stride} \times C_1^2$. The SPD-Conv structure can reduce the spatial resolution of the input feature map and increase the number of channels without losing spatial information, providing a richer feature representation for networks. This approach addresses the issues of loss of fine grained information and low efficiency of feature learning that arise from the use of stride convolution and pooling layers in the CNN architecture.

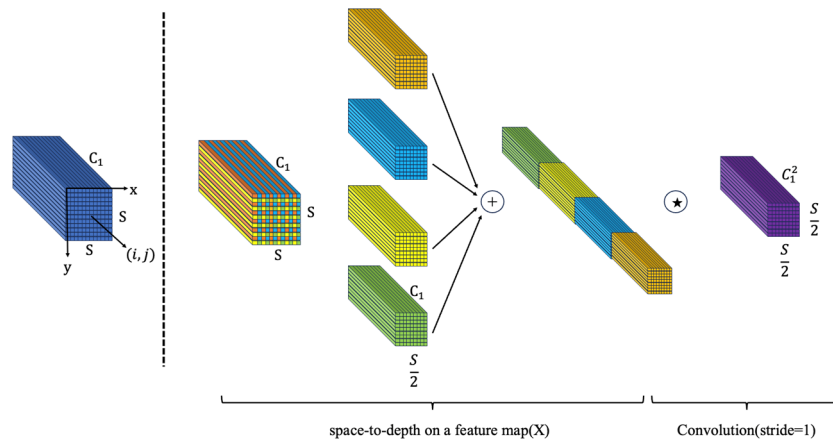


Figure 6. SPD-Conv structure when scale = 2.

Figure 7 below shows the overall network architecture of the approach proposed in this paper, which includes basic YOLOv5s network architecture and the SPD-Conv module for small object detection.

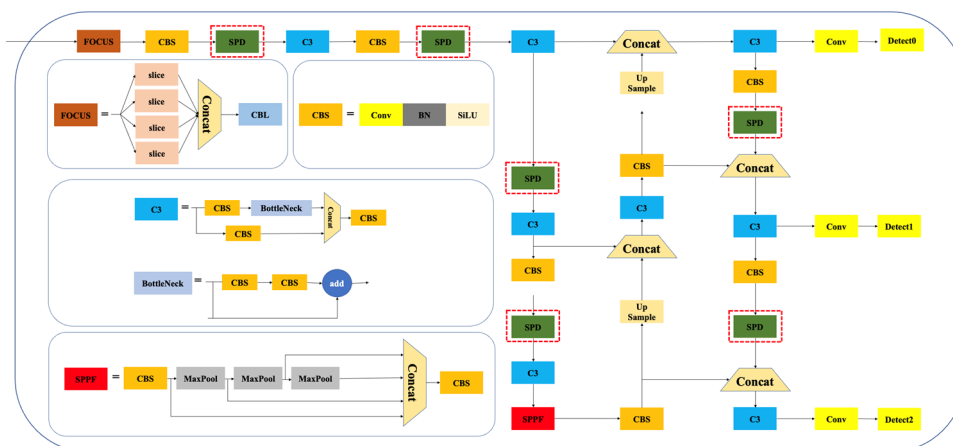


Figure 7. Algorithm architecture figure of YOLOv5s-SPD-Conv.

4.2. CPU and GPU accelerated inference solution

Once the model has been trained and tested, it must be deployed on hardware equipment and facilities, such as cloud computing platforms or development boards. However, the equipment and facilities used in the actual production environment are often less powerful than those used in training models in terms of computational performance. Many enterprises and deep learning framework researchers have proposed high performance deep learning inference optimizers or inference libraries. Examples include TensorRT by NVIDIA, OpenVINO by Intel, and ONNX by Microsoft and Facebook. These optimization solutions offer targeted optimization for different types of graphic image processing algorithms, thereby expanding the range of applications for arithmetic hardware such as CPU/GPU and related accelerators. Using TensorRT as an example, this optimization tool is particularly effective for convolutional and anti-convolutional layers with many channels. This is due to the fact that GPUs are capable of performing intensive operations in parallel, resulting in higher efficiency when processing large matrix operations.

5. Experiments

5.1. Environments

Our running environment for experimentation entails the following factors: The CPU of this experiment was an Intel (R) Core (TM) i9-10900X. The memory was 96GB. The GPU used an Nvidia GeForce RTX 3090. The environment for running was Win 11 + Python 3.8 + PyTorch 1.10.0.

5.2. Evaluation indicators

The algorithm for object detection uses average precision (AP) as an evaluation metric. Mean average precision (mAP) is the average of all classes' AP values. It is then divided into two more precision metrics based on IoU (Intersection over Union): AP50 and AP50:95. The metrics are calculated as shown below:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$mAP = \frac{\sum_{i=1}^n (Recall_{x_2} - Recall_{x_1}) \times Max_{precision}}{n} \quad (3)$$

5.3. Experiment and results

The training parameters are the following. The batch size is 64, the epoch value is 300, the weight decay is 0.0005, the learning rate is 0.01, and the optimizer is SGD.

Table 3 compares the overall results of each object between the original YOLOv5s network and the YOLOv5s network with the SPD-Conv block introduced to improve the detection precision of small objects.

Table 3. Single category detection precision comparison table.

Label name	YOLOv5s	YOLOv5s-SPD-Conv
Bridge Connected	99.5	99.5
Cargo Back Door Open	99.5	99.5
Cargo Front Door Open	99.5	99.5
Passenger Door Open	99.5	99.5
Catering Truck Connected	99.5	99.5
Cargo Loading	99.5	99.2 (-0.3%)
Push Back Tractor Connected A	84.2	92.4 (+8.2%)
Push Back Tractor Connected B	99.4	98.8 (-0.6%)
Baggage Loading	99.5	99.5
Fueling A	98.9	99.4 (+0.5%)
Cone	97.5	97.9 (+0.4%)
Airplane	96.6	96.7 (+0.1%)
Cargo Back Door Open B	99.5	99.5
Cargo Front Door Open B	99.5	99.5
Baggage Loading B	99.5	99.5
Fueling B	99.5	99.5

The PT models of YOLOv5s-SPD-Conv were converted into TensorRT, ONNX, OpenVINO, and TorchScript formats to improve model inference efficiency and facilitate deployment in production environments. Table 4 displays the inference efficiency of each model. Table 5 compares the mean average precision of our method and some common object detection algorithms. Figure 8 illustrates the schematic diagram of the model detection results.

Table 4. Comparison table of inference efficiency of hardware-accelerated inference schemes.

Approach	CPU (ms)	GPU (ms)
Origin (PT)	181.4	16.0
TorchScript	333.8	-
ONNX	155.9	15.9
OpenVINO	76.5 (+137.1%)	-
TensorRT	-	10.3 (+55.3%)

Table 5. Detection result comparison table.

Method/Metrics	mean average precision (AP50/AP50:95)	Speed (ms)
Faster R-CNN ResNet50 [27]	98.3/92.9	45.8
SSD300 [28]	98.7/85.7	20.0
Efficientdet [29]	98.2/79.6	56.9
Cascade-Mask-R-CNN [30]	98.6/ 94.7	115.0
YOLOv5s	98.2/84.8	16.0
YOLOv5s-SPD-Conv (Ours)	98.7/90.0	16.0 (origin) / 10.3 (Tensor RT)



Figure 8. The schematic figure of detection algorithm result. (The category name in the figure is the label name, not the exclusive noun of this category.)

Table 3 shows that the original YOLOv5s is less effective in detecting small objects or objects with more tiny details. For example, the precision values of Push Back Tractor Connected A and Cone are particularly low. However, the detection precision is higher for the door open/close situation, bridge connect, and baggage load, which involve medium or large size objects or objects with more obvious features. This is likely due to the similarities between easy-to-distinguishable categories, while some of the key features of the confusable categories show less differences. As detailed in Figure 9, the similarities between some of the easily distinguish categories are demonstrated in Figure 9(a), using Cargo Back Door Open and Fueling B as examples. In Figure 9(b), Push Back Tractor Connected A and Push Back Tractor Connected B are used as examples to demonstrate the differences between some of the easily confused categories.

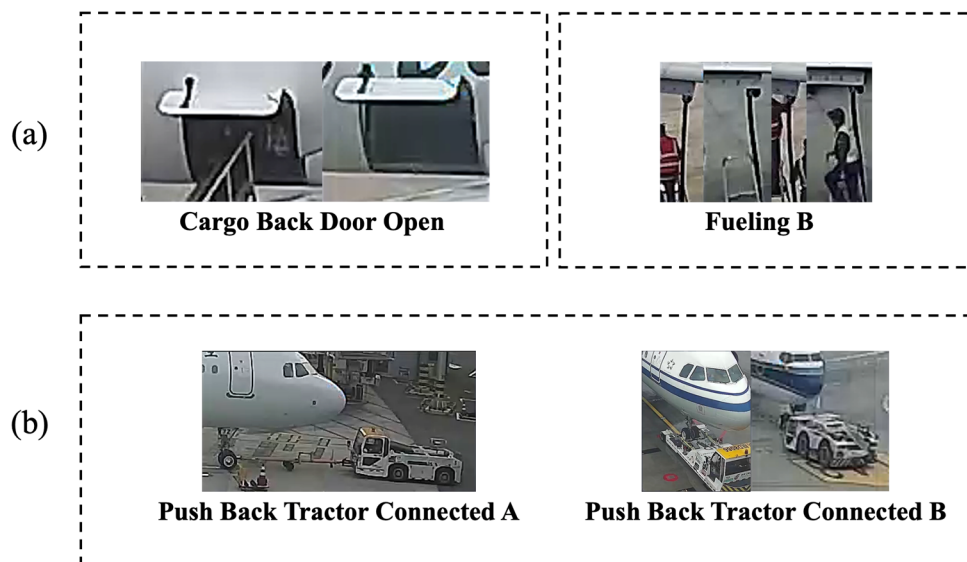


Figure 9. Similarities between the easily distinguishable category and micro-differences between the easily confusable categories.

After adding the SPD-Conv algorithm to improve the YOLOv5s algorithm, the overall precision has improved significantly, despite a slight decrease in precision for some objects (possibly due to training randomness). As shown in Figure 10, by merging the feature information, which will be sent to the detection head, with the original inputting image to construct a visual heatmap, it can be clearly observed that after adding the SPD-Conv module, the redundant and irrelevant information in the image has been reduced. (In the heatmap, the color closer to red means the algorithm will pay more attention to this area, and the detected objects are also likely to be in this area). The improved YOLOv5-SPD-Conv algorithm's precision for all detection classes is now above 90%, which effectively addresses the issue of the original network's lack of precision in recognizing small or detail-lacking objects.

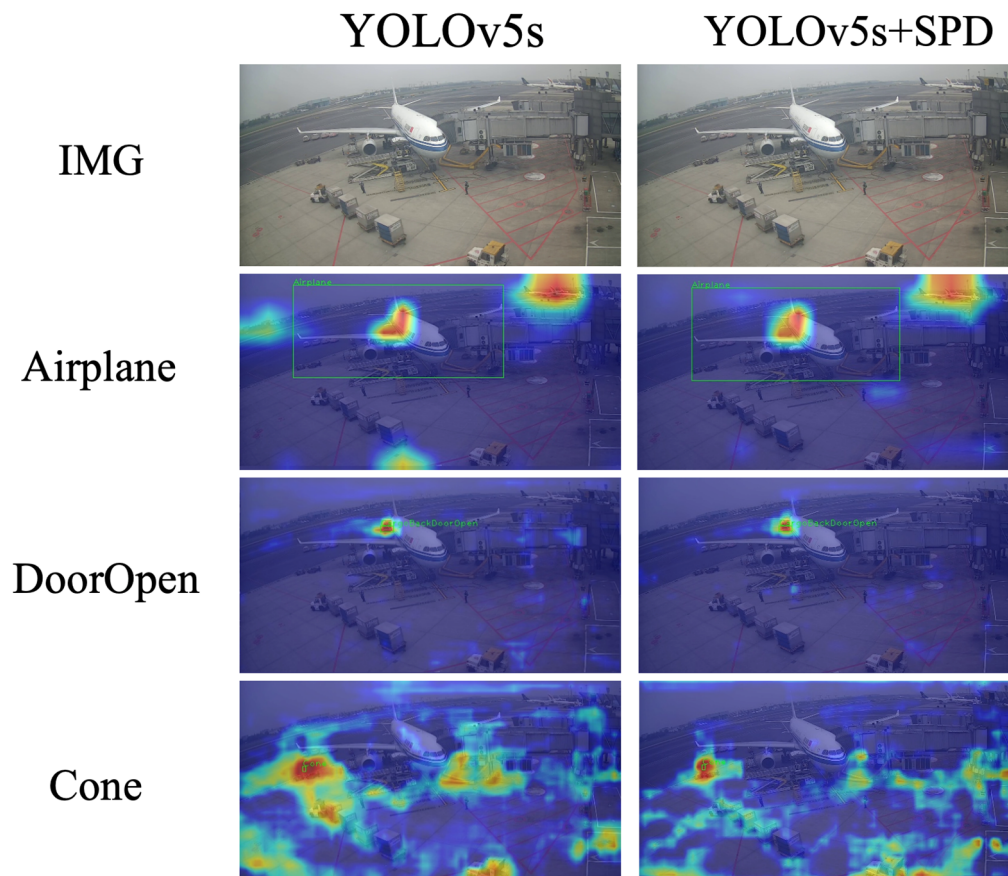


Figure 10. Visualized heatmaps of categories at different scales.

Table 4 shows that the introduction of hardware acceleration architecture improves model inference efficiency by 137.1% on CPU and 55.3% on GPU, making it more suitable for deployment in a product environment.

Table 5 shows that the algorithm proposed in this paper is more suitable for airport apron area service activity detection than the detection current advanced object algorithms, considering the comprehensive computational efficiency and computational precision. At the same time, compared to the original YOLOv5s framework, the improved algorithm proposed in this paper has improved the precision of detection while ensuring the speed of the overall structure does not decrease.

6. Conclusions

This paper demonstrates our research on using computer vision technologies for detecting ground service activities from airport apron surveillance videos. The algorithm recognizes the operational information of the key monitoring objects and activities involved in each ground service process in the monitoring video. Unlike the commonly used object specific detection algorithms, the algorithm in this paper can be more relevant to the production application of detecting the activity of the airport apron surveillance videos. At the same time the algorithm proposed in this paper requires no delineation of the area of activity and secondary analyses of the results of the object detection algorithm. So, the algorithm proposed in this paper is more efficient in terms of computational efficiency. The mean average precision result of the proposed approach is 98.7%, and the approach can be deployed on CPU and GPU devices with higher efficiency. The approach provides effective information for optimizing resources at civil aviation airports, such as personnel, equipment, and facilities.

There could be future improvements to the algorithm. As shown in Figure 11(a), the categories in different surveillance scenarios have similarities in outline, shape, size, etc., which implies that there is some correlation between different categories. This means the study of algorithms with few-shot object detection and transfer learning is also a good direction. This direction can be a good solution when the airports are in the phase of renovation and expansion or new construction and cannot provide sufficient datasets for training models. On the other hand, as shown in Figure 11(b), there is some similarity in the categories from scene to scene though. More than that, however, there are differences in the appearances of different airlines' equipment, differences in lighting, differences in working positions, etc. These cause the precision of the model to be extremely affected by scenes of the surveillance. For example, if the scenes of the surveillance video appear in both the training dataset and the testing dataset, then the precision of this surveillance will be improved. This may be due to the large variability in airport construction, as well as the constraints of the size of the surveillance range and the requirements of the job. It is difficult to show all the poses of each target in the surveillance video, which can lead to a more homogeneous dataset and may result in overfitting and lower robustness during the experiment. How to overcome these differences to improve model precision or trying to explore zero-sample learning of the model are good directions for research.

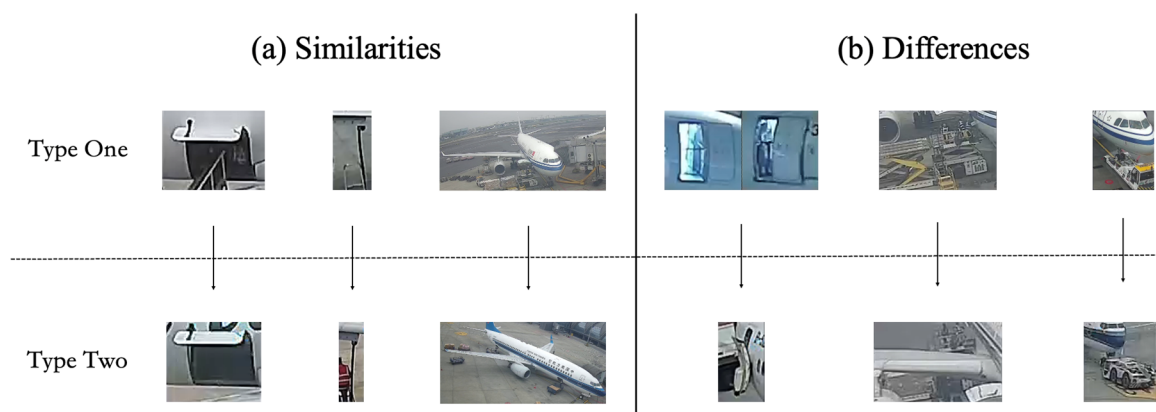


Figure 11. The similarities and differences in categories.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

This project is financially supported by the Fundamental Research Funds for the Central Universities (NO. ZJ2022-008), the R&D Program of Key Laboratory of Flight Techniques and Flight Safety, CAAC (No. FZ2022ZZ01), the Safety Science R&D Program of CAAC (No. MHAQ2023031).

Conflict of interest

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

References

1. Civil Aviation Administration of China, CAAC Issues 2022 Statistical Bulletin of Civil Airport Production in China, 2023. Available from: https://www.caac.gov.cn/XXGK/XXGK/TJSJ/202303/t20230317_217609.html.
2. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788. <https://doi.org/10.1109/CVPR.2016.91>
3. H. Law, J. Deng, Cornernet: Detecting objects as paired keypoints, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 734–750. <https://doi.org/10.1007/s11263-019-01204-1>
4. Z. Tian, C. Shen, H. Chen, T. He, Fcos: Fully convolutional one-stage object detection, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, (2019), 9626–9636. <https://doi.org/10.1109/ICCV.2019.00972>
5. J. Liu, M. Wang, X. Xie, Y. Song, L. Xu, Leather defect detection algorithm based on improved YOLOv5, *Comput. Eng.*, **49** (2023), 240–249. <https://doi.org/10.19678/j.issn.1000-3428.0064587>
6. J. Choi, D. Chun, H. Kim, H. J. Lee, Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 502–511. <https://doi.org/10.1109/ICCV.2019.00059>
7. P. Li, H. Li, Research on fod detection for airport runway based on yolov3, in *2020 39th Chinese Control Conference (CCC)*, IEEE, (2020), 7096–7099. <https://doi.org/10.23919/CCC50068.2020.9188724>
8. X. Shi, J. Hu, X. Lei, S. Xu, Detection of flying birds in airport monitoring based on improved YOLOv5, in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, IEEE, (2021), 1446–1451. <https://doi.org/10.1109/ICSP51882.2021.9408797>
9. F. H. Cai, Y. X. Zhang, J. Huang, A bridge surface crack detection algorithm based on YOLOv3 and attention mechanism, *Pattern Recognit. Artif. Intell.*, **33** (2020), 926–933. <https://doi.org/10.16451/j.cnki.issn1003-6059.202010007>

10. Y. M. Shi, K. B. Jia, The Research and implementation of moving object detecting and tracking in intelligence video monitor system, in *2011 International Conference on Multimedia and Signal Processing*, IEEE, **2** (2011), 105–108. <https://doi.org/10.1109/CMSP.2011.111>
11. J. Ferryman, M. Borg, D. Thirde, F. Fusier, V. Valentin, F. Brémond, et al., Automated scene understanding for airport aprons, in *AI 2005: Advances in Artificial Intelligence. AI 2005. Lecture Notes in Computer Science*, Springer, Berlin, **3809** (2005), 593–603. https://doi.org/10.1007/11589990_62
12. N. Cai, G. He, Multi-cloud resource scheduling intelligent system with endogenous security, *Electron. Res. Arch.*, **32** (2024), 1380–1405. <https://doi.org/10.3934/era.2024064>
13. M. Kumar, S. Ray, D. K. Yadav, Moving human detection and tracking from thermal video through intelligent surveillance system for smart applications, *Multimedia Tools Appl.*, **82** (2023), 39551–39570. <https://doi.org/10.1007/s11042-022-13515-6>
14. A. Raza, S. A. Chelloug, M. H. Alatiyyah, A. Jalal, J. Park, Multiple pedestrian detection and tracking in night vision surveillance systems, *CMC-Comput. Mater. Continua*, **75** (2023), 3275–3289. <http://doi.org/10.32604/cmc.2023.029719>
15. S. Wilke, A. Majumdar, W. Y. Ochieng, Airport surface operations: A holistic framework for operations modeling and risk management, *Saf. Sci.*, **63** (2014), 18–33. <https://doi.org/10.1016/j.ssci.2013.10.015>
16. H. L. Lu, S. Vaddi, V. Cheng, J. Tsai, Airport gate operation monitoring using computer vision techniques, in *16th AIAA Aviation Technology, Integration, and Operations Conference*, (2016), 3912. <https://doi.org/10.2514/6.2016-3912>
17. Y. Zou, Q. Ying, R. Liu, M. Rong, Research on evaluation method for operation support capability of airport apron, in *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, IEEE, (2020), 1043–1047. <https://doi.org/10.1109/ICCASIT50869.2020.9368541>
18. T. V. Phat, S. Alam, N. Lilith, P. N. Tran, B. T. Nguyen, Aircraft push-back prediction and turnaround monitoring by vision-based object detection and activity identification, in *Proc. 10th SESAR Innov. Days.*, 2020.
19. M. Gorkow, *Aircraft Turnaround Management Using Computer Vision*, 2020. Available from: <https://medium.com/@michaelgorkow/aircraft-turnaround-management-using-computer-vision-4bec29838c08>.
20. S. Wang, Y. Che, H. Zhao, A. Lim, Accurate tracking, collision detection, and optimal scheduling of airport ground support equipment, *IEEE Internet Things J.*, **8** (2020), 572–584. <https://doi.org/10.1109/JIOT.2020.3004874>
21. S. Wang, C. Li, A. Lim, ROPHS: Determine real-time status of a multi-carriage logistics train at airport, *IEEE Trans. Intell. Transp. Syst.*, **23** (2021), 6347–6356. <https://doi.org/10.1109/TITS.2021.3055838>
22. S. Yıldız, O. Aydemir, A. Memiş, S. Varlı, A turnaround control system to automatically detect and monitor the time stamps of ground service actions in airports: a deep learning and computer vision based approach, *Eng. Appl. Artif. Intell.*, **114** (2022), 105032. <https://doi.org/10.1016/j.engappai.2022.105032>
23. P. Thai, S. Alam, N. Lilith, B. T. Nguyen, A computer vision framework using Convolutional Neural Networks for airport-airside surveillance, *Transp. Res. Part C Emerging Technol.*, **137** (2022), 103590. <https://doi.org/10.1016/j.trc.2022.103590>

24. S. Liu, R. Wu, J. Qu, Y. Li, HDA-Net: hybrid convolutional neural networks for small objects recognition at airports, *IEEE Trans. Instrum. Meas.*, **71** (2022), 1–14. <https://ieeexplore.ieee.org/abstract/document/9939036>
25. *MH/T 6125—2022*, Technical specifications for airport collaborative decision making system, 2022. Available from: <https://www.caac.gov.cn/XXGK/XXGK/BZGF/HYBZ/202202/P020220228396026654632.pdf>.
26. R. Sunkara, T. Luo, No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Cham, Springer Nature Switzerland, (2022), 443–459. https://doi.org/10.1007/978-3-031-26409-2_27
27. S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Adv. Neural Inf. Process. Syst.*, (2015), 28.
28. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, et al., Ssd: Single shot multibox detector, in *Computer Vision—ECCV 2016. ECCV 2016. Lecture Notes in Computer Science()*, Springer, Cham, **9905** (2016), 21–37 https://doi.org/10.1007/978-3-319-46448-0_2
29. M. Tan, R. Pang, Q. V. Le, Efficientdet: Scalable and efficient object detection, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 10781–10790.
30. K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, et al., MMDetection: Open mmlab detection toolbox and benchmark, preprint, arXiv:1906.07155. <https://doi.org/10.48550/arXiv.1906.07155>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>).