Application of Convolutional Neural Network Technology in Vehicle Parking Management

Yihao Cheng*
Xi'an University of Science &Technology, Xi'an, Shaanxi 710054, China
569335645@qq.com

Guohua Yue
Xi'an University of Science &Technology, Xi'an, Shaanxi 710054, China
yueguohua@163.com

ABSTRACT

Most of the existing intelligent parking management systems guide vehicles to park based on available spaces and do not differentiate between vehicle types. However, with the increase in the number of electric vehicles, there is a new need to efficiently allocate parking spaces with charging posts and those without charging posts to the required vehicle types. For this reason, we need an intelligent parking management system that can identify the vehicle type and determine whether the vehicle parking is legal or not. To achieve this, we need to separate stationary vehicles from moving vehicles in the parking lot to track whether the moving vehicles are parked in accordance with the parking area designated by the parking lot. First, we use the YOLO network to try to identify the vehicles in the parking lot. Then, we use inter-frame difference method to separate the moving vehicles and mark them. Finally, we determine whether these moving vehicles enter the illegal area or not. After the vehicles are legally parked, the moving vehicles become stationary vehicles and are no longer in our attention. Therefore, we do not display their markers anymore. If the vehicle enters an illegal area where entry is prohibited, the system displays a warning message. With this system, we can individually determine whether a moving vehicle is intruding into an illegal area, while ignoring stationary vehicles.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Computer vision**; • **Computer vision tasks**; • **Scene anomaly detection**;

KEYWORDS

YOLO neural network, Convolution neural network, Inter-frame difference method, Vehicle identification, Parking management

ACM Reference Format:
Yihao Cheng* and Guohua Yue. 2021. Application of Convolutional Neural Network Technology in Vehicle Parking Management. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3487075.3487191

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487191

1 INTRODUCTION

Computer use various imaging hardware systems instead of the human eyes to present specific images of objects as input and use computer systems instead of the human brain to process and interpret the images for further interpretation and analysis. On this whole, the computer can directly perceive and learn from the external world and has certain environmental adaptation capabilities to perform certain tasks with some degree of intelligence based on visual feedback. That is what we called computer vision [1].

Intelligent surveillance is one of the application aspects of computer vision [2]. It has applications in object recognition, boundary-crossing recognition, trajectory tracking, lost object recognition, license plate recognition, speed measurement, traffic statistics, and anomalous behavior recognition. Overall, surveillance systems can be divided into two generations. Older surveillance systems use videotape or hard disk to store surveillance footage, the system itself cannot process images and cannot be accessed remotely. Newer intelligent surveillance systems can be connected to the Internet and can perform some processing of surveillance images to assist the monitor in making decisions [3].

Convolutional neural network is a concept that was introduced in the 1980s. It is built after the mechanism of biological visual perception. With the evolution of computing devices, convolutional neural networks have been rapidly developed and applied to the field of computer vision. Many efficient and interesting network structures have been proposed, such as residual networks and VGG networks.

In 2000, Tan's team achieved rough recognition of vehicles based on their shape and size [4]. In 2010, Burnos and his team used fuzzy set and data fusion to further improve the correct rate of vehicle identification to 95% [5]. This was followed by Arrospide's team in 2013, which used gradient histograms to increase the correct recognition rate to 96% [6]. As support vector machines, Alexnet, and other multilayers convolutional neural networks continue to gain popularity, car recognition rates are becoming higher and higher. Among them, we are most interested in YOLO neural networks [7].

Inspired by the above paper, we combine YOLO network with inter-frame difference method to build a surveillance system for parking lots [8]. The cameras installed in the parking lot input the image of the interior of the parking lot, and the YOLO network identifies the vehicles in the image, and then uses the inter-frame difference method to filter out the vehicles in motion that we need to pay attention to, and judge their behavior, to assist the parking lot manager in the purpose of vehicle parking management.

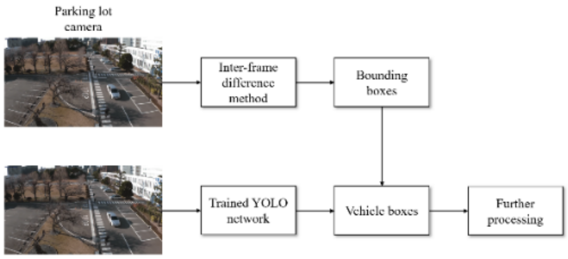


Figure 1: Parking Image Processing Process.

2 MODEL AND ALGORITHM

In this paper, the complete process of the parking lot image is shown in Figure 1. The camera reads the parking lot frames and passes them to the YOLO network to identify vehicles. The inter-frame difference method filters out vehicles that are moving from them, and the system detects the position of moving vehicles and determines whether the movement of these vehicles is legal.

2.1 Inter-frame Difference Method

When we need to separate the motion target, we want to distinguish the target from the background. This is called background detection. There are several methods for background detection, such as inter-frame difference method, background subtraction, and average background method. The parking lot is characterized by a constantly changing background as vehicles are parked. The overall light and darkness of the parking lot will also change under natural lighting and weather change conditions. The background subtraction method requires manual updating of the background, and the average background method cannot adapt to the natural lighting environment of the parking lot. In summary, we select the inter-frame difference method as the background detection in this paper.

2.1.1 Background Detection. The principle of inter-frame difference method for detecting moving objects is to superimpose two frames of continuous image information or several frames of continuous image information at equal positions and then do the difference processing, and the result is the general edge of the extracted moving target. This process of separating the moving target from the background is called background detection.

To exclude interference due to camera shake or weather changes, we need to do some pre-processing of the images before doing differential analysis.

Before performing morphological filtering, we advanced median filtering with binarization processing. Median filtering eliminates the more isolated noise. For the binarization process, we use the following equation.

$$V(i,\ j) = \begin{cases} 255, & D(i,\ j) \geq Threshold \\ 0, & D(i,\ j) < Threshold \end{cases} \quad (1)$$

In the formula, D(i, j) represents the grayscale value of the corresponding point at the (i, j) position, and V(i, j) is the value of the pixel at the (i, j) point after thresholding, which is 0 or 255. Here, 0 is set as the point of the background image, 255 is set as the point of the foreground image, and Threshold refers to the pre-set threshold value before binarization. We can see that the selection of the threshold value is very critical. It plays a key role in the accuracy of foreground and background segmentation.

In this paper, we will use the global threshold method. That is, the threshold value will remain a definite and constant value throughout the program processing. The advantage of this method is that it is computationally simple and suitable where there is a significant gap between the background image and the foreground image.

After performing these two steps, we have extracted the background and foreground and are ready to work on morphological filtering.

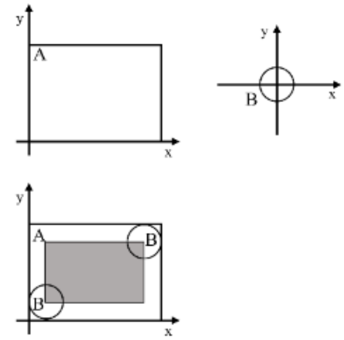


Figure 2: Corrosion Operation.

2.1.2 Morphological Filtering. We have separated the foreground from the background. However, median filtering can only eliminate smaller individual noise, not larger blocks of noise, which may cause a whole foreground to be split into several pieces or a situation where there are voids in the foreground. To avoid the impact of these problems on later operations, we need to morphologically filter the images.

The most commonly used morphological filtering methods are erosion and expansion.

The corrosion operation is shown in Figure 2 below.

As we can see, A is the object to be processed and B is the structure element. We make B scan the whole image, and when B is completely contained by A, the outermost center point of B forms a contour line. The A inside this contour line is retained and the outer A is discarded. This operation is called erosion. B is usually circular or rectangular.

The Erosion operation filters out fine lines in the image, giving it a smoother outline. But at the same time, it also shrinks the target body area, so we need a subsequent expansion operation. The principle of the expansion operation is shown in Figure 3 below.

As shown in the Figure 3, A is the object to be processed and B is the structure element.

The expansion operation can be regarded as a pairwise operation of the corrosion operation. We let B scan the entire image. We keep the center point of B at this moment as long as B intersects with A. Thus, the outermost centroid of B will form a contour line containing A. We consider the whole range inside the contour line as A. That is called the expansion operation.

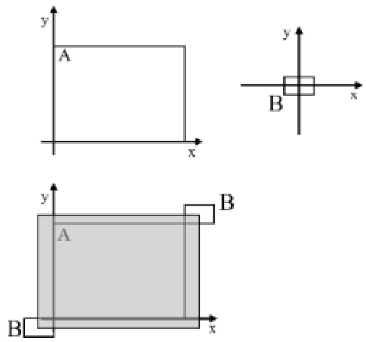


Figure 3: Expansion Operation.

By first eroding and then expanding, which has been called the opening, we can remove the noise from the image. By first expanding and then eroding, we can remove the holes inside the target, which has been called the closing. In this paper, we first corrode and then expand the target image. After this, we use bounding boxes to mark out consecutive regions with an area larger than a certain threshold, which are the motion targets we detect.

2.2 YOLO Network

The raw images of the parking lot captured by the camera are fed into a pre-trained neural network for recognition. We compare the vehicle location information recognized by the neural network with the moving target location information obtained by the inter-frame difference method, and then we can filter out the moving vehicles. We use the YOLO network for vehicle recognition.

2.2.1 The Structure of YOLO Network. Many algorithms have been developed for image classification. For example, the two-stage algorithm of R-CNN class with high accuracy. Another class is the high-speed one-stage algorithm similar to YOLO. YOLO stands for "You Only Look Once", which means only one CNN operation is needed. YOLO has undergone many developments, and in the third generation, the residual structure was added, allowing the network to be built deeper.
The sliding window is very time-consuming. To be able to perform parallel computations and reduce the computation time, the neural network can employ a convolutional layer instead of a fully connected layer. By convolutional operations instead of sliding windows, the results are stored in the resulting smaller matrix, and each cell is equivalent to one computation of the sliding window.
In YOLO network, this idea is taken one step further by directly segmenting the original image into small non-overlapping squares and then generating the feature map by convolution. Each element of the feature map corresponds to a small square of the original image.
Specifically, YOLO partitions the input image into an S×S grid. Each cell is responsible for detecting targets whose center point is within the grid. Each cell detects B bounding boxes and the respective confidence level of the bounding boxes. By multiplying the confidence of the bounding boxes by the IOU (intersection over union), we can obtain the class-specific confidence scores of each bounding box. By performing the same operation for each grid, we get a lot of bounding boxes. Specifically, the calculation is as follows:

$$\Pr\left(class_i|object\right) * \Pr\left(object\right) * IOU_{pred}^{truth} = \Pr\left(class_i\right) * IOU_{pred}^{truth} \tag{2}$$

To eliminate our unwanted bounding boxes, we use non-maximum suppression (NMS). It solves the problem of the same target being detected multiple times.
First, we eliminate all the bounding boxes with a confidence scores less than the set threshold. Secondly, we select the bounding box with the highest confidence score and calculate the IOU value between this box and other bounding boxes in turn, if it is greater than a certain threshold (overlap is too high), then we eliminate the boxes with too high overlap. This process is repeated for the remaining bounding boxes until all detected boxes are processed.
After this work is done, we can draw the remaining bounding boxes on the screen and the recognition work is successful.
Overall, the process of object recognition by YOLO neural network is as follows:
Input: Original image

1) Divide the image into an S×S grid
2) Recognition of all grids with convolutional operations instead of sliding windows
3) For each obtained bounding box, the confidence score of the bounding box is obtained by multiplying its probability of containing the object *Pr(object)* by the IOU between the bounding box and the true boundary, and then by the conditional probability *Pr($class_i$|object)* at the confidence level of each bounding box to get the confidence score as formula (2)
4) Eliminate redundant bounding boxes with NMS
5) Draw the bounding box, recognition is complete

In summary, all aspects of the YOLO network are well suited to our needs.

2.2.2 YOLO Training. To train the YOLO network, we use Stanford University's cars database [9]. It contains many different vehicle types and is well labeled. This time we only let the neural network determine the location of the vehicle, not the specific model. In subsequent applications, we can use this database to make more detailed judgments, such as the make and specific model of a vehicle. For example, a parking management system can assign parking spaces to vehicles according to their energy type, allowing for the most efficient use of charging posts.
The initial learning rate is 0.0032. The dataset is divided into 20 batches. The mean-square error functions are used as loss functions. The specific formula is shown in the following equation. Where, coordError, iouError and classError represent coordinate error between predicted and calibrated data, IOU error and classification error, respectively.

$$loss = \sum_{i=0}^{S^2} coordError + iouError + classError \tag{3}$$

The main structure of the YOLO neural network is shown in Figure 4. When we calculate the error function, we also need to make corrections for each calculation term. We multiply the coordError

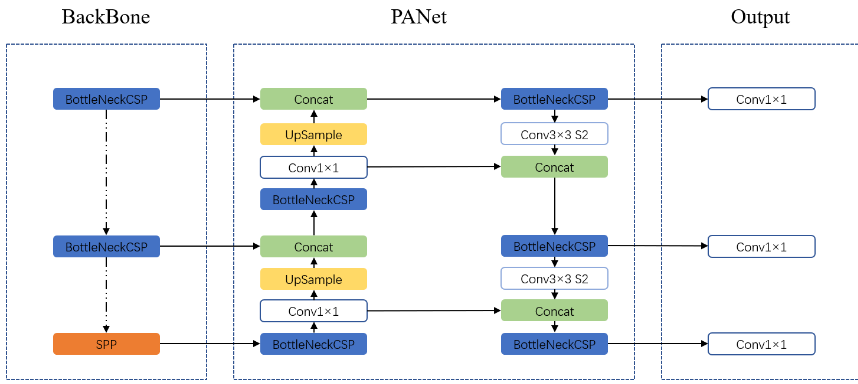


Figure 4: Overview of the Main Structure of the YOLO Network.

by 5 to boost the contribution of the coordinate error to the loss function, and we multiply the iouError by 0.5 to correct for the effect of the error of the grid containing the objects in the calculation of the gradient of the network parameters.

3 RESULTS AND ANALYSIS

The inter-frame difference method has a high requirement on the frame rate of the video. Therefore, we must ensure that the speed of the neural network in recognizing vehicles can keep up with the speed of the video playback. The YOLO network is well suited to our needs. On a GPU device, YOLO can recognize objects at a speed of 45 frames per second, which exceeds the recording frame rate of most cameras and can be applied to real-world scenarios.

In this study, we use a video resolution of 1280×720 and a frame rate of 30. It contains moving vehicles, stationary vehicles, several parking areas, and pedestrians.

We have delineated in the video the areas where entry is prohibited, the areas where parking is allowed and tried to separate moving vehicles from stationary vehicles for the purpose of aiding parking lot management.

3.1 Experiment

(1) Firstly, we feed the video into the inter-frame difference program [10]. According to the characteristics of the parking lot image, we adjust the size of the kernel and threshold of the inter-frame differencing, as well as the size of the erosion and expansion.

According to the characteristics of slow running vehicles inside the parking lot and little change in the background, we set the size of the core to 7×7 and find the inter-frame difference every 5 frames. When the inter-frame difference area is larger than 1000, we mark it as a moving object. This threshold is to exclude unexpected disturbances like leaves being blown by the wind and similar.

The inter-frame difference method returns us a four-dimensional vector (x, y, w, h), where x and y are the starting points of the upper-left corner of the bounding box, w is the width of the bounding box, and h is the height of the bounding box. Based on this vector, we can then mark the moving object on the screen.

(2) Then, we feed the video into the YOLO neural network. This network is a pre-trained vehicle recognition network that can identify vehicles within the image range.

Similar to the inter-frame difference method, the YOLO neural network also returns a four-dimensional vector. The first two terms of this vector are the upper left corner of the bounding box and the last two terms are the lower right corner of the bounding box.

(3) Once we obtain the bounding box of the inter-frame difference method and the bounding box of the YOLO network, we can compare the positions of these two bounding boxes. A vehicle bounding box is selected, and the IOU value between it and the moving object bounding box is calculated in turn. When it is less than a threshold value, it means that there is a vehicle but no moving object at that place, we remove this bounding box. This threshold value should be taken between 0.4 and 0.8. Because the bounding box delineated by the interframe difference method often includes the shadow of the vehicle, the area will be much larger than the area of the vehicle itself. The program will make judgments on all vehicle bounding boxes in turn. Eventually we will filter out all moving vehicles and eliminate stationary vehicles.

(4) After filtering out all the moving vehicles, all we have to do is to make a judgment on their location. Beforehand, we have delineated a no-driving area on the video screen. If a moving vehicle enters this area, its bounding box should turn warning red and a warning message should be displayed. This step can be done by coordinate detection. When all these steps have been performed, the results are displayed on the video.

3.2 Results Analysis

3.2.1 Analysis of Inter-Frame Difference Method. Figure 5(a) shows the original video footage. Figure 5(b) shows the results of the inter-frame differencing method. Figure 5(c) shows the processing of the inter-frame difference method.

We can see from the pictures that there is a lot of noise in the processing due to camera shake and wind. Although we have filtered out a lot of interference through the minimum area threshold, this still has a possible impact on the program.

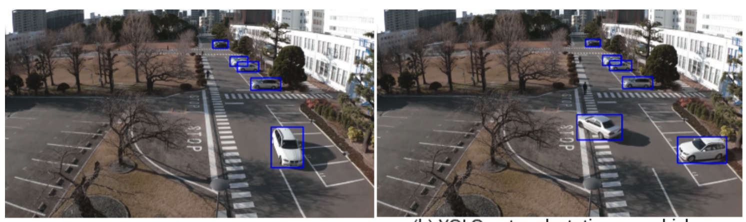
Figure 6(a) and 6(b) shows the results of the enhanced inter-frame difference method. We can see that the noise of the interframe

(a) Original video footage (b) Inter-frame difference method (c) Unenhanced processing screen

Figure 5: Original Screen and Unenhanced Inter-Frame Difference Method.

(a) Inter-frame difference method (b) Enhanced processing screen

Figure 6: Enhanced Inter-Frame Difference Method.



(a) YOLO network motion vehicle recognition (b) YOLO network stationary vehicle recognition

Figure 7: YOLO Neural Network Recognition.

(a) Motion vehicle recognition (b) Stationary targets are no longer marked

Figure 8: YOLO Network Combined with Inter-Frame Difference Method.

difference method is greatly reduced after the erosion and expansion treatment. The white areas in Figure 6(b) are large moving targets such as pedestrians and vehicles, and the noise due to camera shake and tree shake has been completely eliminated.

3.2.2 Analysis of YOLO Network. Figure 7(a) and 7(b) show the recognition results of the YOLO network. Figure 7(a) contains a vehicle in motion. In figure 7(b), this car is parked in a parking space. The program still recognized it without any problem. At the same time, a vehicle pulls into the shot below.

3.2.3 Analysis of Mobile Vehicle Identification Program. Figure 8(a) and 8(b) shows the results of the inter-frame difference method combined with the YOLO network. In Figure 8(a) we can see that

(a) Added a no-entry zone (b) Warning displayed when vehicle enters

Figure 9: Warning for Vehicles Entering No-Entry Areas.

stationary vehicles that have finished parking in the distance have been excluded from the marker. The program only marks the vehicles that are coming towards us. And in Figure 8(b), when the approaching vehicle enters the parking space and finishes parking, the program no longer marks it either, but moves on to mark another vehicle that is moving in the screen.

Figure 9(a) and 9(b) shows the delineation and warning of the no-entry zone. As shown in Figure 9(a), we have delineated an area in the video as a no-entry zone. If the vehicle drives into the area, then its marker box will change color and a warning message will be displayed on the screen. We can see that all is well with the marker box of the car that is approaching in the distance of the image. This is because it has not yet entered the no-entry zone. In Figure 9(b), we can see the car driving into a no-entry zone. Its marker box turns purple and a warning message is displayed in the center of the box. When it drives away from the no-entry zone, the marker box will return to normal and the warning message will disappear.

4 CONCLUSIONS

4.1 Conclusion

To assist in parking lot vehicle management, we combine YOLO neural network with inter-frame difference method. This program is able to separate moving vehicles from stationary vehicles and recognize and detect moving vehicles separately, which is suitable for environments such as parking lots where there are a large number of stationary vehicles as background. After designating a no-entry zone, the program is also able to issue warnings about entry into the no-entry zone to assist administrators in detecting anomalies. If the classification of vehicle types is refined so that no-entry areas are compared to vehicle tags, the division of parking areas for different vehicle types can be achieved. The program achieves our goal of assisting in parking management.

4.2 Future Work

We plan to write a user-interactive graphical interface in which the user can choose to manually mark the no-parking zones. The marking is done by double click on the selected point on the monitoring screen and the inside of the polygon formed by several points is the no-parking zone. Alternatively, the user can select a time period. During this time period, the system will automatically identify parking boxes and mark them as no-parking zones. For example, parking is prohibited from 22:00 to 6:00. Any parking during this period will trigger an alarm. Further, the administrator should be able to edit the labels of the no-parking zones. The system should be able to filter the vehicles entering the area based on the tags, warning about illegal entries and not reacting to legitimate ones. We hope that this system can be applied not only to parking management but also to more environments with a large number of stationary targets and several moving targets and strict requirements on the target location, such as laboratory hazardous drug management or warehouse management.

REFERENCES

[1] Hiroshi Ishikawa, Cheng-Lin Liu, Tomas Pajdla, Jianbo Shi(2021). Computer Vision – ACCV 2020[M].:2021-02-26.

[2] Hasmat Malik, Nuzhat Fatema, Jafar A. Alzubi(2021). AI and Machine Learning Paradigms for Health Monitoring System[M].:2021-02-26.

[3] Sagaya Selvaraj A.,Anusha S(2021). RFID Enabled Smart Data Analysis in a Smart Warehouse Monitoring System using IoT[J]. Journal of Physics: Conference Series, 1717(1).

[4] Tan T N, Baker K D(2000). Efficient image gradient based vehicle localization[J]. IEEE Transactions on Image Processing, 9(8):1343-1356.

[5] Piotr B(2010). Alternative Automatic Vehicle Classification Mmethod[J]. Metrology and Measurement Systems, 2010, 17(3):323-333.

[6] Arróspide Jon, Salgado L, Camplani M(2013). Image-based on-road vehicle20 detection using cost-effective Histograms of Oriented Gradients[J]. Journal of Visual Communication and Image Representation, 24(7):1182-1190.

[7] Joseph Redmon, Ali Farhadi. YOLOv3: An Incremental Improvement[J]. arXiv preprint arXiv: 1804.02767

[8] Liang, R., Yan, L., Gao, P., Qian, X., Zhang, Z., Sun, H (2010). 'Aviation video moving-target detection with inter-frame difference'. Image and Signal Processing (CISP), Yantai, China, pp. 1494–1497

[9] Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei(2013). 3D Object Representations for Fine-Grained Categorization. 4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013.

[10] PanasonicBusiness: 4K Camera in use at Panasonic Office in Japan [EB/OL]. [2015-6-23]. https://www.youtube.com/watch?v=wqlO5i3N-FU