The use of Network Science in the gait recognition

Mathematical Modelling and Machine Learning

Ruilin He 121108423

November 22, 2021

Contents

Abstract.....3

1. Introduction.....4

2. What is gait recognition?.....5

3. How gait recognition works.....7

4. Basic concepts of Convolutional Neural Networks.....8

4.1 Padding.....8

4.2 Stride.....9

4.3 Local receptive fields.....9

4.4 Shared weights.....9

4.5 Pooling.....10

5. Basic structure of convolutional neural network.....10

5.1 Input layer.....11

5.2 Convolutional calculation layer.....14

5.3 Rectified Linear Unit layer.....16

5.4 Pooling layer.....17

5.5 Fully-Connected layer.....18

6. Operation of Convolutional Neural Network.....19

7. Compared with deep neural network (DNN).....20

8.Different model of CNN in gait recognition.....23

8.1 R-CNN.....23

8.2 Fast R-CNN.....25

8.3 Faster R-CNN.....27

8.4 Mask R-CNN.....28

8.5 3D CNN.....29

9. Advantages and disadvantages of gait recognition.....32

Reference.....35

Abstract

This report mainly shows the use of convolutional neural networks in gait recognition and explains what is the basic principle of gait recognition and the operation of gait recognition. Then discussed some basic concepts and structures of convolutional neural networks. Besides, compared with deep neural networks about processing images and mentioned some improved models based on convolutional neural networks for gait recognition, finally discussed gait recognition the advantages and disadvantages.

Key Words: convolutional neural network; gait recognition; feature extraction; detection;

1. Introduction

When it comes to recognition systems in artificial intelligence, it is a popular choice to use facial recognition and fingerprint recognition. With the advent of modern science, an innovative and unique tool different from other identification systems, gait recognition, has gradually attracted people's attention. It is realized that gait recognition could play an important role in recognition work due to its long-distance recognition characteristics. Although gait recognition is still not widely used commercially, this powerful recognition system has grown rapidly and has great potential. This essay will explain how gait recognition works and discuss the advantages and disadvantages of gait recognition from different perspectives.

2. What is gait recognition?

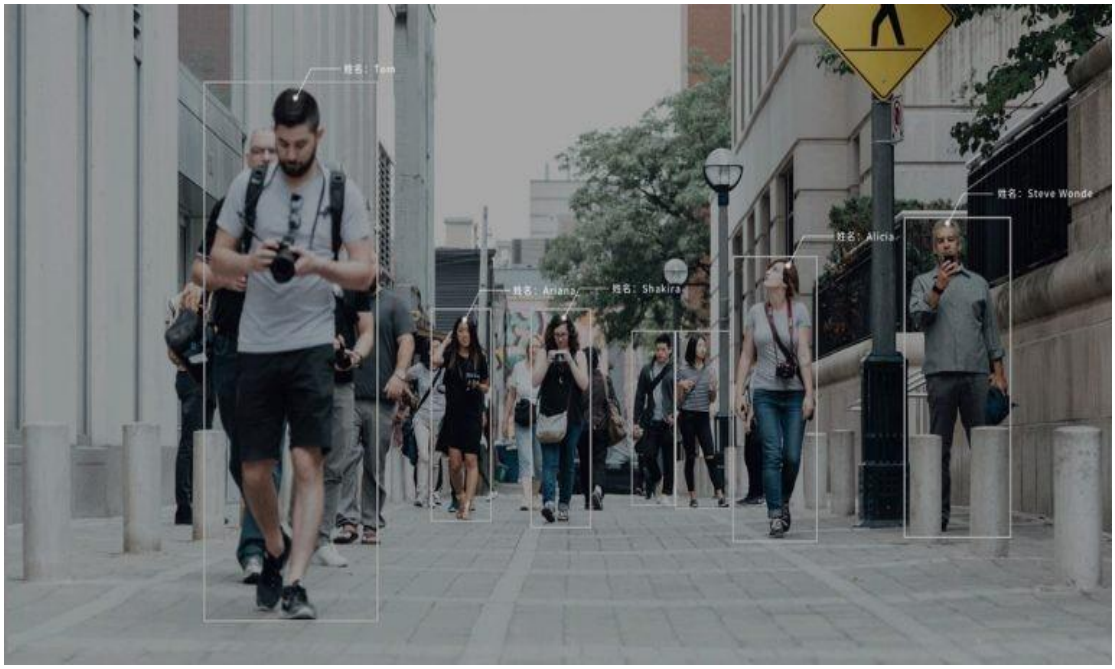
Before starting the discussion, it is necessary to understand what gait recognition is. It is common knowledge that different people have similar ways of walking, depending on their height, weight, length of arms and legs. However, in fact, different body types, head shapes, muscle distribution, and bone sizes could make people have different gaits. Based on these characteristics, as long as a large number of tests are performed and enough data is collected, the computer can use deep learning to observe the different characteristics of people's walking for identity recognition. In other words, the motion features of different human postures are the main features that the gait recognition system needs to extract.

Gait recognition is a brand-new technology that has not been commercialized. Compared with other recognition technologies, gait recognition has the advantages of not requiring physical contact and being able to recognize at a distance. Therefore, in the case of video recognition of the target population, it has more potential than face recognition.

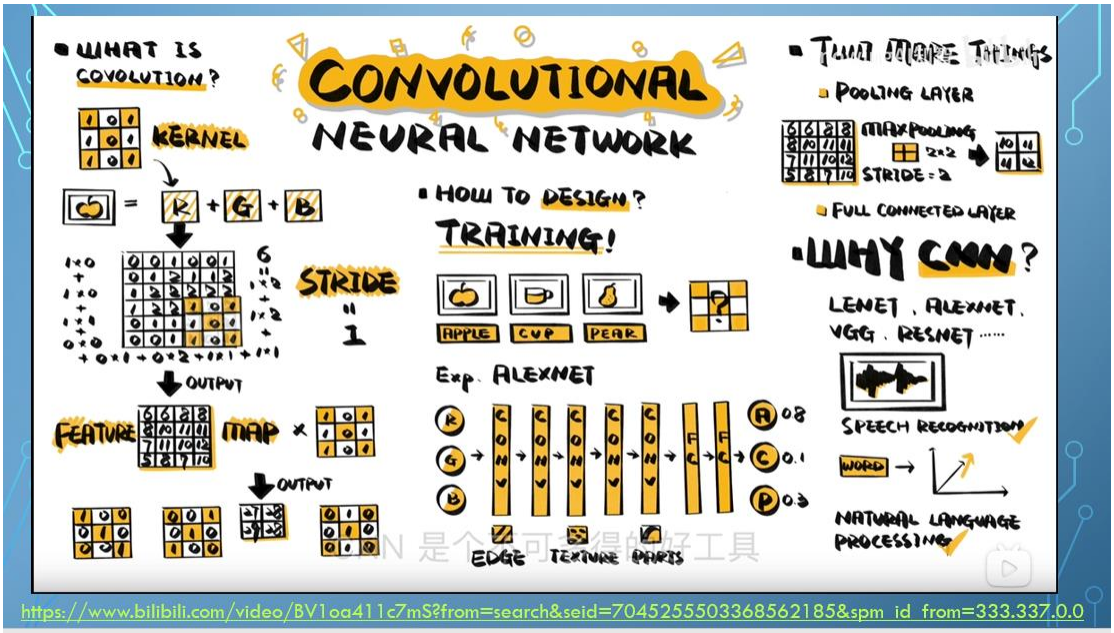
What is more, humans are good at gait recognition. They can distinguish familiar people based on walking posture within a certain distance. Gait recognition is based on this concept to allow machines to learn to recognize as humans. However, due to the brunch of data in sequence images, the computational complexity of gait recognition is relatively high, and it is hard to process. Although the gait recognition has made some progress, it is worth considering whether this technology could be successfully applied in practice, or it would still take several years of development.

3. How gait recognition works

The principle of gait recognition is mainly like this. First, the gait recognition will obtain data through video cameras, motion sensors and other related equipment to establish the database and analyze, extract the characteristics of the gait, and finally perform the recognition. For example, analyze the pedestrians on the sidewalk, first inspect the pedestrians, and then segment the background of the pedestrians, leaving the characteristics of the pedestrians walking, comparing them with the asynchronous state in the database, and finally analyzing the data to identify the pedestrians' identity. In the work of gait recognition, it is believed that the use of convolutional neural networks for processing is the most efficient method. In addition, to further understand the operation of gait recognition, it is necessary to understand the convolutional neural network.



4. Basic concepts of Convolutional Neural Networks



4.1 Padding

Two shortcomings will arise when performing convolution operations. The first disadvantage is that every time people do a convolution operation, the image will shrink from 6×6 to 4×4. After people do it a few times, the image will become small and might be reduced to 1×1. The second disadvantage is that the edge information of the image plays a small role, and much information about the edge position of the image will be lost in the convolution. After many convolution operations, the image will be reduced, so those pixels in the corner participate in the convolution calculation less, which means that some information about the edge position of the image is lost, so the missing information is reduced by using

padding.

4.2 Stride

During convolution, not only padding is needed to increase the information, but also a part of the information needs to be compressed by setting the stride. The purpose of setting the stride is to reduce the number of input parameters and reduce the amount of calculation.

4.3 Local receptive fields

In the CNN model, the feature vector about the location in the feature map is calculated based on the input of the fixed area of the previous layer. In this case, this area is the local receptive field of this location, and images outside the local receptive field will not affect the feature vector of the map layer. Its size is the size of the filter, so it can be considered that this size is the depth of the input image.

4.4 Shared weights

Weight sharing is often used to control the number of parameters in convolutional neural networks, so the shared weight is the value of the shared filter. It is worth noting that neurons of different depths will not

share the same weight. If in a convolution kernel, each receptive field uses a different weight value, the weight value of the convolution kernel is different. If in this case, the number of parameters will be huge, which will make the calculation process long and complicated. Weight sharing means that if a feature is useful in calculating a position, it is also useful when it is calculating a different position.

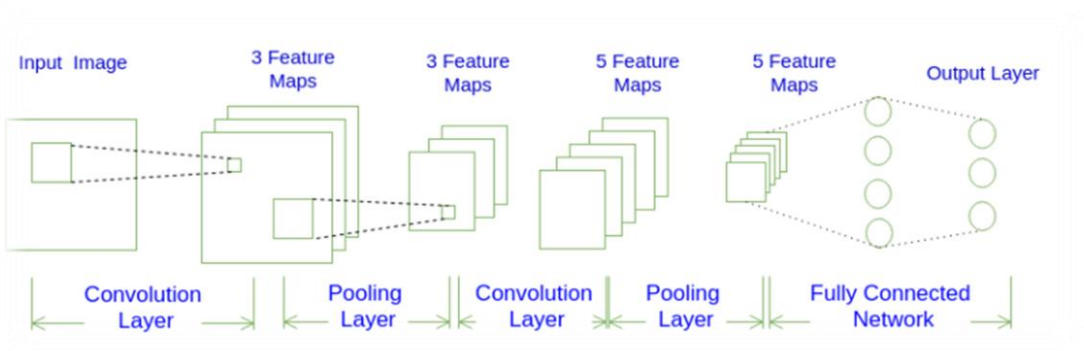
4.5 Pooling

In a convolutional neural network, adjacent pixels may have similar values. In this case, the values of the output pixels are resemblance. Therefore, in the output of the convolutional layer, only the key information needs to be retained, and most of the information can be deleted. In addition, a simple way is to use pooling to reduce the value of the output value by reducing the size of the input.

5. Basic structure of convolutional neural network

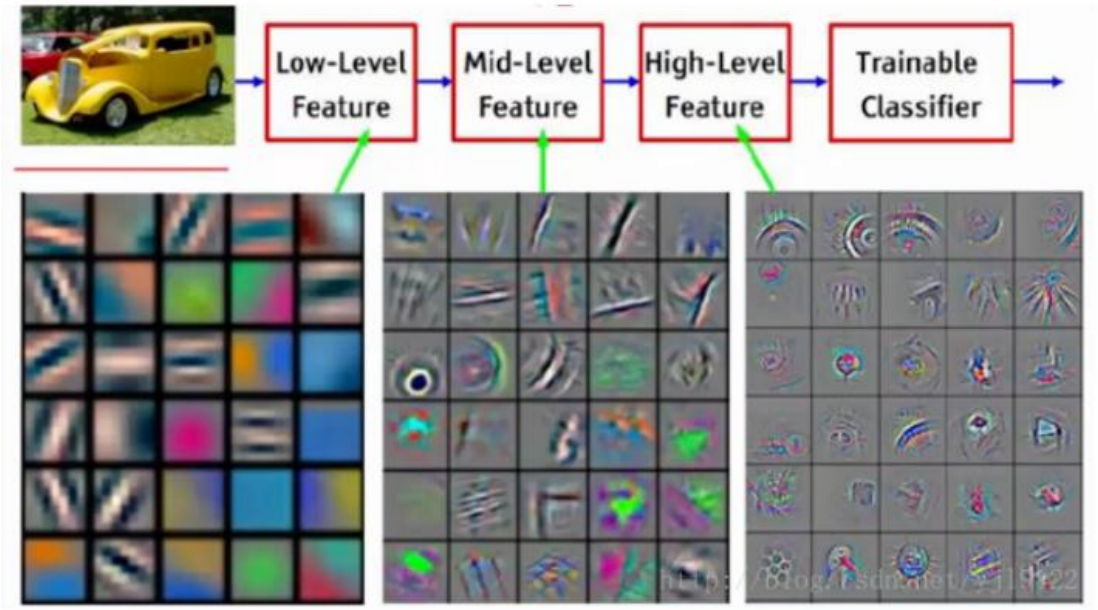
In a convolutional neural network, it is generally divided into 5 layers, Input layer, Convolutional calculation layer, Rectified Linear Unit layer,

Pooling layer, and Fully-Connected layer. Data enters these five layers in order for related operations. Each layer has a different role corresponding to different data processing, according to these five layers for related processing, the desired data can be obtained.



5.1 Input layer

In this layer, the data needs to be preprocessed, this is because if the input data units are different, the training time of the neural network may become longer and it will become difficult to converge. Therefore, before entering the next layer, the usual operation is to grayscale or normalize the data.



What is graying? The picture is generally stored through a three-dimensional matrix, the size of the matrix is (width, height, 3). In here, width represents the width of the image, height represents the height of the image, and 3 represents red, green, and blue three different colors. It is believed that a picture is formed by superimposing different degrees of red, green, and blue. Since RGB cannot reflect the shape characteristics of the image, it is only the color adjustment, and CNN just needs to extract the shape characteristics of the image. In this case, the three-channel picture can be turned into one channel.

Graying methods in the analysis:

1. Component method

The brightness of the three components in the input image is used as the

gray value of the three gray images, and then a gray image is selected for use according to actual application needs.

2. Maximum value: the maximum value of the three-component brightness in the input image is used as the gray value of the gray image.

3. Weighted average method: Average the three-component brightness in the input image to obtain a grayscale image for subsequent operations.

Besides, the sigmoid function is used as the activation function in neural networks. The function value of the sigmoid function is between [0, 1]. When inputting data that is far greater than 1, such as 5 or 10, After the value of the sigmoid function might be very close or even equal, it will not have the expected training effect. Normalizing the data can better solve this problem. And normalization can make the neural network converge faster.

normalization methods in the analysis:

1. Min-max standardization:

This method refers to linear transformation of the input data, and the corresponding result value is placed in zero and one. Max is the maximum value of the data, and min is the minimum value of the data. However, it is worth noting that if new data input is added, max and min may change and

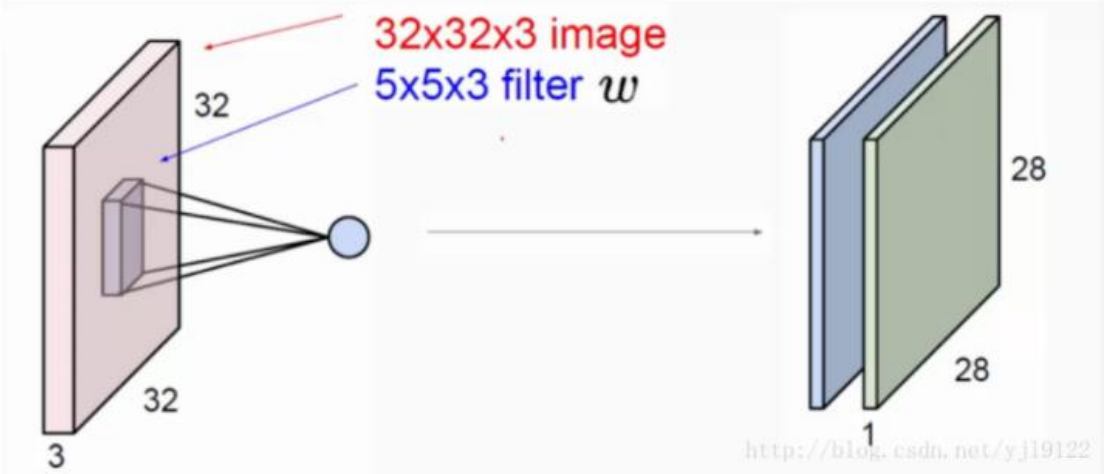
need to be redefined again.

2. Z-score standardization method

This method requires the mean and standard deviation of the original input data to standardize the data. After the data is processed, it will become following the standard normal distribution, the mean is 0 and the standard deviation is 1. In addition, if the input data range is too large, the impact on the pattern classification will also increase. On the contrary, the impact of a small data range may be small.

5.2 Convolutional calculation layer

Convolutional layer is the key to convolutional neural network, and the main task in this layer is to use a filter for feature extraction. In this layer, the previously mentioned concepts such as padding and stride will appear here.



Convolutional layer function:

1. In the convolutional layer, different filters make up the parameters of the convolutional layer. Due to the convenience of calculation, the width and height of the filter will be as small as possible, but the depth is consistent with the input data. It can be considered that this model allows the filter to be activated when it sees some related types of features. For example, the boundary and color of the upper layer in different directions, and the model can even be the shape of the upper layer.

2. At the same time, it can be called the output in the model, it will only observe a part of the input data, and the parameters of all nearby outputs are the same.

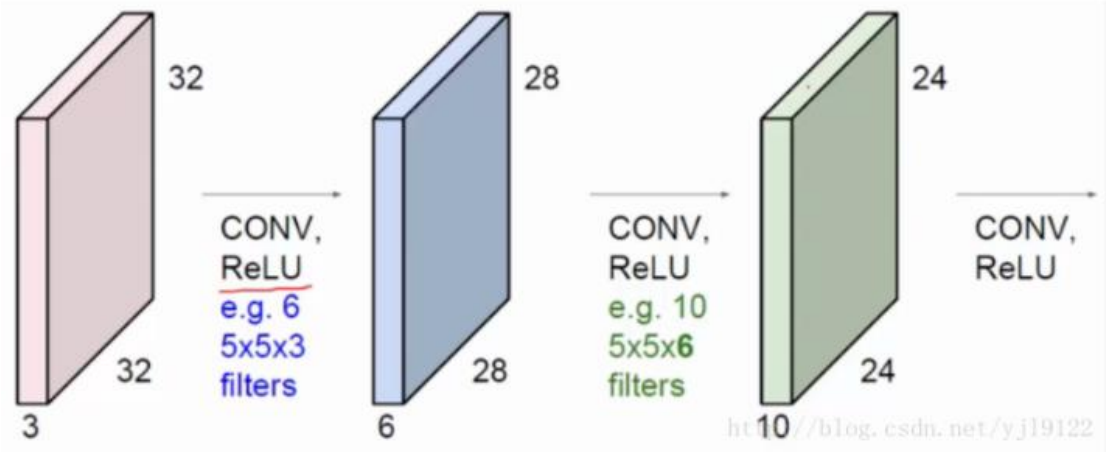
3. There is another way to optimize the model by reducing the number of parameters. Since convolution can share weights between different parameters, it can reduce the number of parameters and prevent overfitting caused by too many parameters.

In addition, a feature of the CNN network is that the ratio of the first few layers of the convolutional layer will be small, but the calculation ratio is large, but the fully connected layer is the opposite. It is worth noting that

most CNN networks have this feature.

5.3 Rectified Linear Unit layer

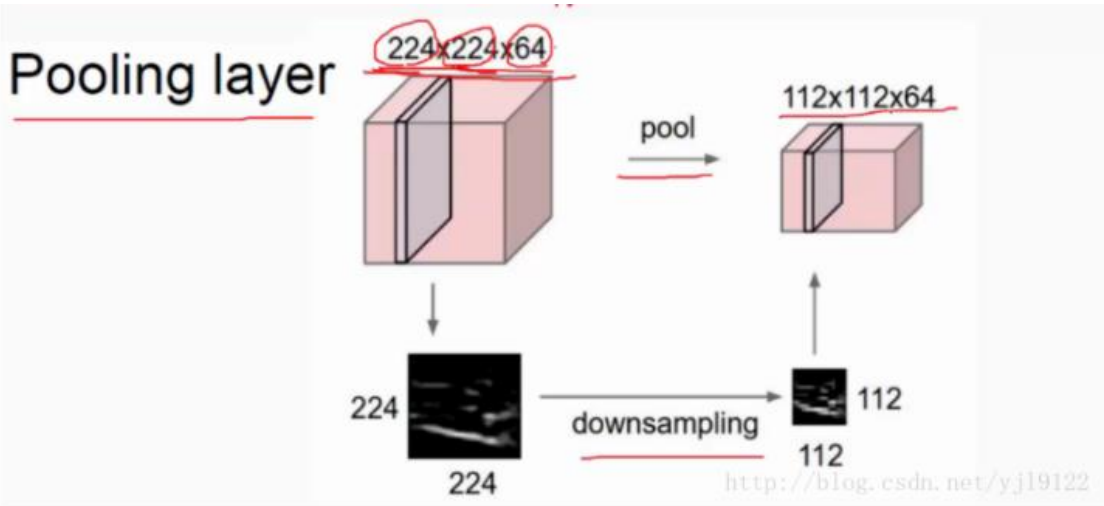
The upper-level operations are linear, but it is necessary to consider nonlinear operations in some practical applications. Therefore, a non-linear relationship can be constructed in the calculation results, usually called an activation function. According to the principles of biology, the next neuron will be activated only when the signal is transmitted before it is greater than the threshold of the neuron in a biological neuron. The activation functions often used in CNN networks are sigmoid function, tanh function, RELU function. In this layer, the selection function is used to map the result of the convolutional layer. In recent years, the activation function does not choose the sigmoid or tanh function, but the rectified linear unit layer function, which is characterized by fast convergence and simple gradient calculation, but it is more fragile. In this case, the expressive power of the network is strong enough, it can be approximated to almost any function so that the neural network can be applied to many nonlinear models.



5.4 Pooling layer

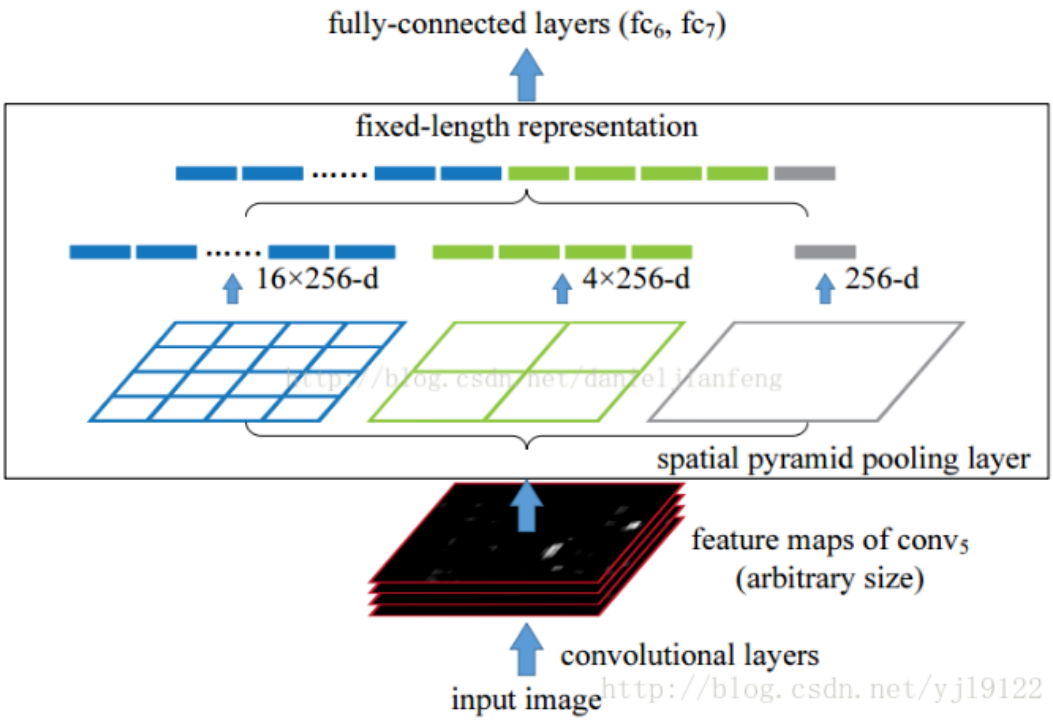
After passing through the ReLU layer, the data will enter the pooling layer for processing. Generally, two methods of maximum pooling and average pooling are used. For example, maximum pooling uses a 3×3 matrix for scanning input data. In this case, the maximum value of the 3×3 area will be taken. Besides, average pooling is similar to maximum pooling, that is, 3 times the average of 3 regions will be taken. It is worth noting that the input data of the pooling layer is generally the output of the convolutional layer. The use of pooling reduces the dimensionality of each feature map, but it can maintain important information. It reduces the consumption of computing resources and can effectively control over-fitting. The pooling layer exists in the middle of the continuous convolutional layer and is responsible for compressing the amount of data and parameters. The pooling layer has no parameters, compressing data is the main task of this

layer.



5.5 Fully-Connected layer

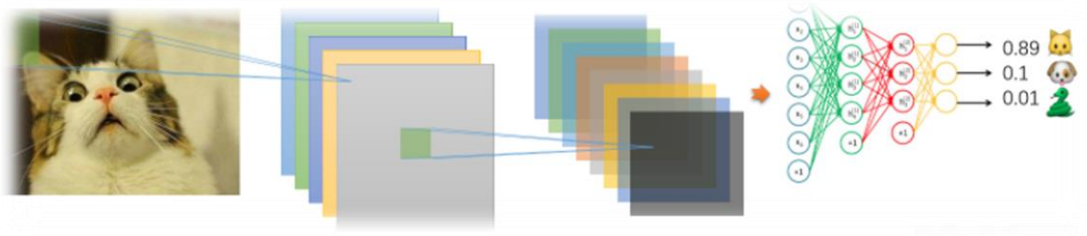
Finally, after multiple processing, the data will pass to his fully connected layer, which will expand the feature map obtained by the last layer of convolution into a one-dimensional vector to provide input options for the model. In other words, it is the inductive integration of features that can better perform the final classification or regression.



6. Operation of Convolutional Neural Network

The original data such as a picture will become a matrix filled with RGB values after being input. The convolution kernel is placed in this matrix at a certain distance. In calculations, this is called stride. Then, the convolution kernel and the image are aligned, multiplied, and then added together to output a feature map. In addition, setting different convolution kernels will output different feature maps. Finally, by combining the extracted features, you can get the probability that the picture is a certain item. In gait recognition, the machine processes the gait recognition by

analyzing human gait features, namely, performing related motion detection, background segmentation, and feature extraction on the gait motion in the image. Secondly, continue to process the data, store the relevant gait characteristics of the person in the database, and wait for the call. Finally, compare the gait characteristics of the target person with the gait characteristics of the gait database. If it matches the gait characteristics of the database successfully, the information of the person in the database will be output. If the information does not match, the machine will continue to collect and store the gait in the database.



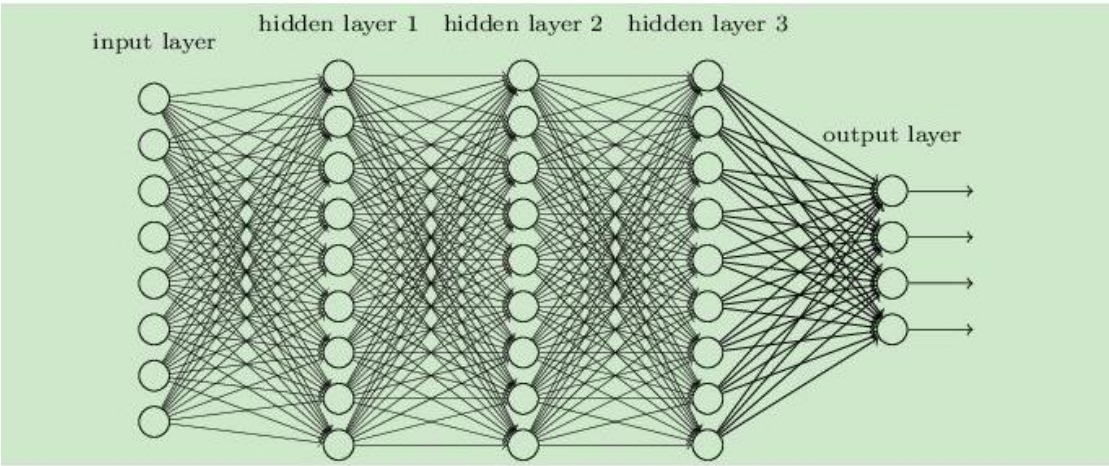
7. Compared with deep neural network (DNN)

In DNN, all neurons in the upper layer have a corresponding neuron in the lower layer, which will lead to many unnecessary parameters. Too many unnecessary parameters not only slow down the calculation speed but also lead to overfitting. Imagine that the input is an image with low pixels, which will generate lots of weights, and the calculation of many weights

will make the learning effect of the network worse. If people use the DNN network to analyze the image not only may there be overfitting, but also the possibility of falling into the local optimal solution becomes greater. Therefore, it is often considered to use a more efficient CNN network in image recognition. For the task of gait recognition, each pixel has a relatively close connection with its surrounding pixels, the connection may be small. For images, if each feature has a corresponding next layer when the network is operating, the possibility of generating a large number of weights will be increased, and their values will be small, which shows their influence is small, it is worthless to the recognition.

In CNN, due to the characteristics of local connections, the neurons in the latter layer will only connect to some key neurons, and it is not necessary to connect all the neurons in the previous layer. In this case, many unnecessary parameters could be reduced and overfitting can be avoided. At the same time, Weight sharing allows different connections to share the same weight, instead of each connection having a different weight, thereby reducing many unnecessary parameters and increasing the efficiency of learning. It is worth noting that the convolutional neural network can obtain better results by retaining some important parameters as much as possible and reducing unnecessary parameters.

However, it does not mean that the DNN model is less efficient than the CNN model, but people are more likely to choose the CNN model in the direction of recognition. The CNN model also has shortcomings in image processing. Since the pooling layer will lose valuable information, a large bunch of data is needed to support the analysis of the CNN model. Moreover, it is worth noting that the use of a gradient descent algorithm can reduce the error so that the output of the training converges to a local minimum instead of the global minimum. Therefore, in gait recognition, we can make some related improvements to the CNN model to make the model more suitable for recognition and analysis, they all have outstanding performance in the application of gait recognition.

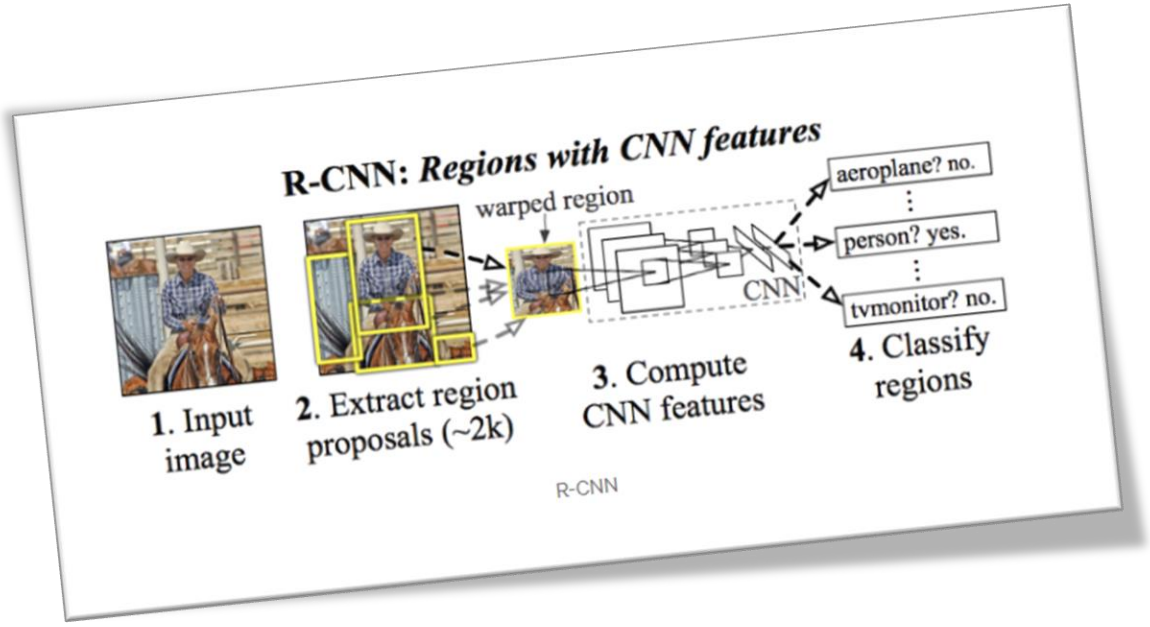


8.Different model of CNN in gait recognition

In order to make the analyzed data more accurate, it is feasible to make improvements based on the CNN model. In gait recognition, many improved CNN models are used, these models are more accurate and worthy of reference. The following are some excellent improved CNN models.

8.1 R-CNN

R-CNN is another improvement of convolutional neural networks in practical applications. Based on the convolutional neural network, a method called Region Proposal is added to analyze the problem. This algorithm can be summarized in three steps:



(1) Selection of candidate regions

Region Proposal is a method of extracting regions. Firstly, it checks the existing small regions, merges the two most probable regions, repeats this step continuously until the image can be merged into one candidate region, and finally outputs a candidate region. In addition, standardizing the extracted target image according to the suggestion. What is more, input of CNN can be considered as acquiring a region of the target image and then performing feature extraction or convolution operation. Finally, it is standardized according to the proposed extraction target image and used as CNN data input.

(2) CNN feature extraction

Like ordinary convolutional neural networks, convolution or pooling are performed according to the input to get the output. In other words, the feature image is output through convolution and pooling in this operation.

(3) Perform classification and boundary regression

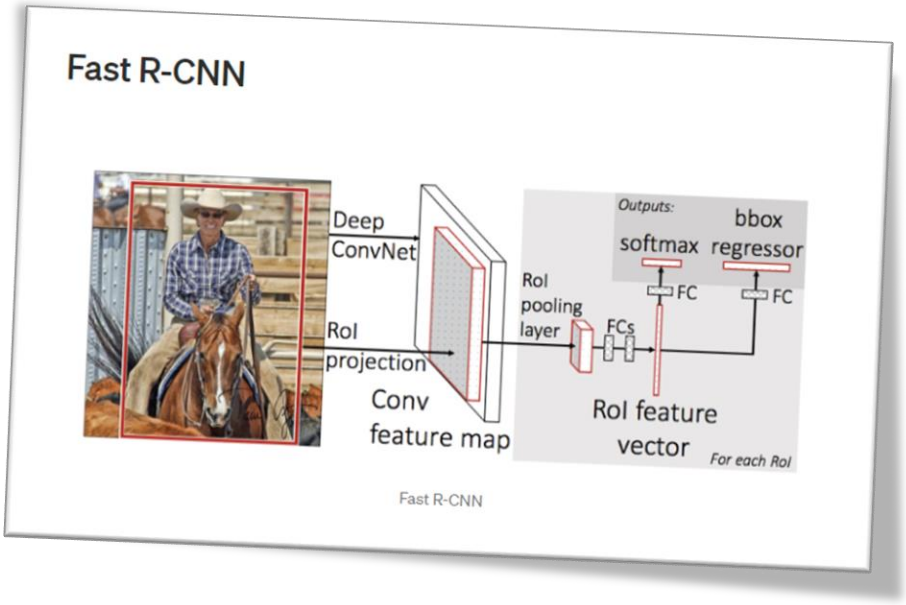
Classify the results output in the previous step to obtain accurate regional information. In this case, multiple detections can be reduced, and a higher accuracy target can be obtained.

Indeed, RCNN has made achievements in detection. However, a series of

problems such as low efficiency and long waiting time still make this model not widely used. R-CNN needs to extract the image corresponding to the candidate area in advance before it can be used. It is worth noting that a large amount of disk space will be occupied using this method. For the traditional CNN model, the input map needs to fix the size of the image. The deformation of the image during the normalization process will cause the image size to change. Therefore, a different region proposal needs to be input each time for the calculation. In this case, it will lead to repeated extraction of the same feature too many times, which will lead to many useless calculations. Therefore, to better solve the problem of image recognition, Fast R-CNN appeared.

8.2 Fast R-CNN

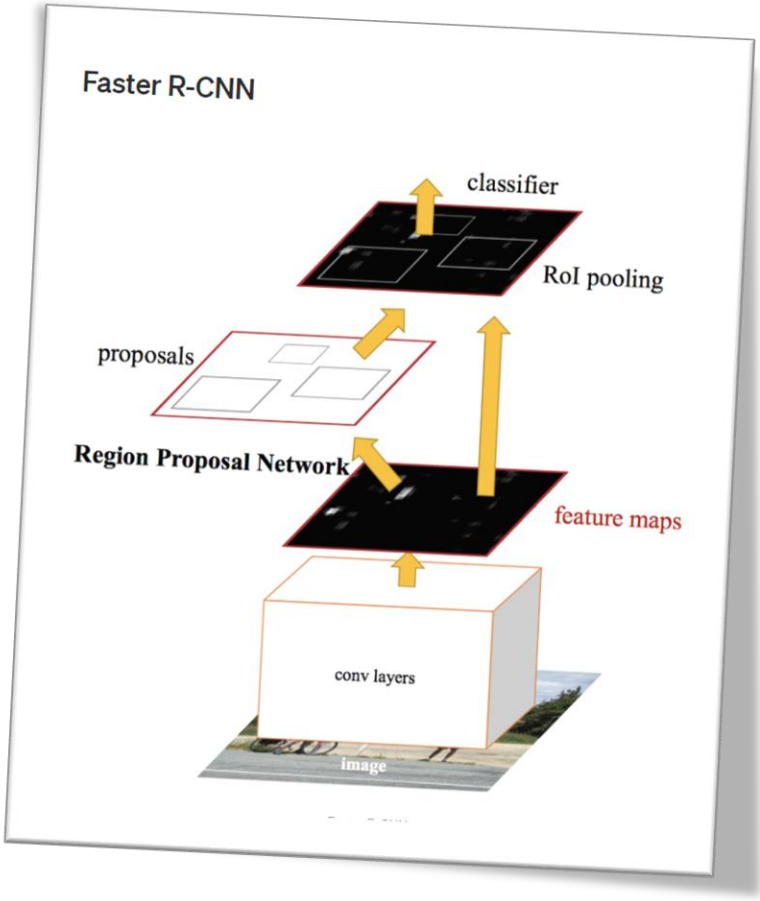
Fast R-CNN can be said to be an improved model for R-CNN problems. Compared with the previous R-CNN, Fast R-CNN has improved speed during testing and improved speed during training. And increase the space required for data training.



In R-CNN, its most obvious shortcoming is that the input image needs to be segmented and before the convolution operation is converted to a size suitable for the operation. It is considered that this should not happen for detection. Due to this situation, it is unavoidable that the image will lose some features, which will have an impact on the feature selection. Based on this situation, the founder of Fast R-CNN made some improvements to R-CNN. The difference between Fast R-CNN and R-CNN is that Fast R-CNN has no restrictions on all inputs. Therefore, in the ROI Pooling layer, each input ROI area on the size feature map will be represented according to the features of a fixed dimension, to ensure that further operation can be performed.

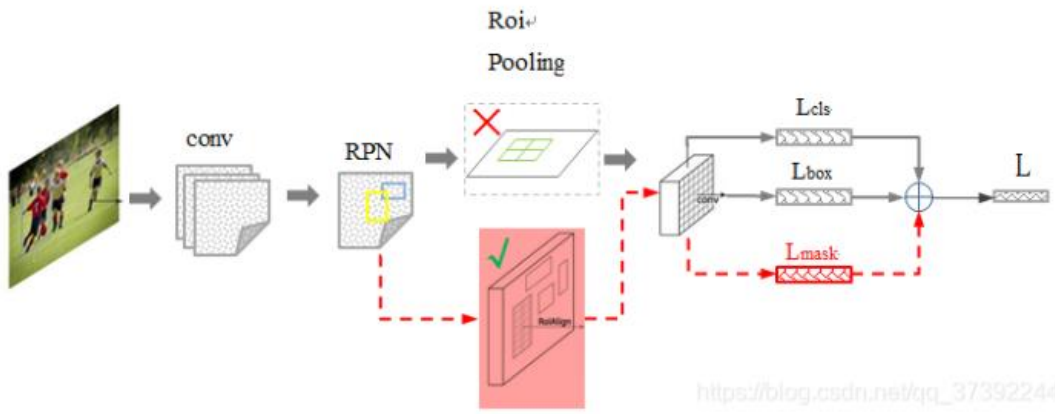
8.3 Faster R-CNN

From R-CNN to Fast R-CNN, and then to Faster R-CNN, it is worth noting that there are improvements in inefficiency. The biggest difference between Faster R-CNN and RCNN and Faster R-CNN is that several steps required for data detection are all completed by deep neural networks. Besides, it could run on the GPU, which significantly improves the efficiency of the operation. Faster R-CNN is composed of two models responsible for different functions, namely the RPN model and the Fast R-CNN model. It is worth considering that multiple models can be used in gait recognition or other graphics-related recognition to improve data analysis.



8.4 Mask R-CNN

Mask R-CNN can be obtained through some changes in Faster R-CNN. In Faster R-CNN, because the input of the fully connected layer needs to unify the size of the output result, but the ROI generated by the RPN network is inconsistent in size, there is an ROI Pool layer for processing ROIs of different sizes into a uniform size for output. Finally, the output result of the ROI pooling layer is used as the input result of the fully connected layer.



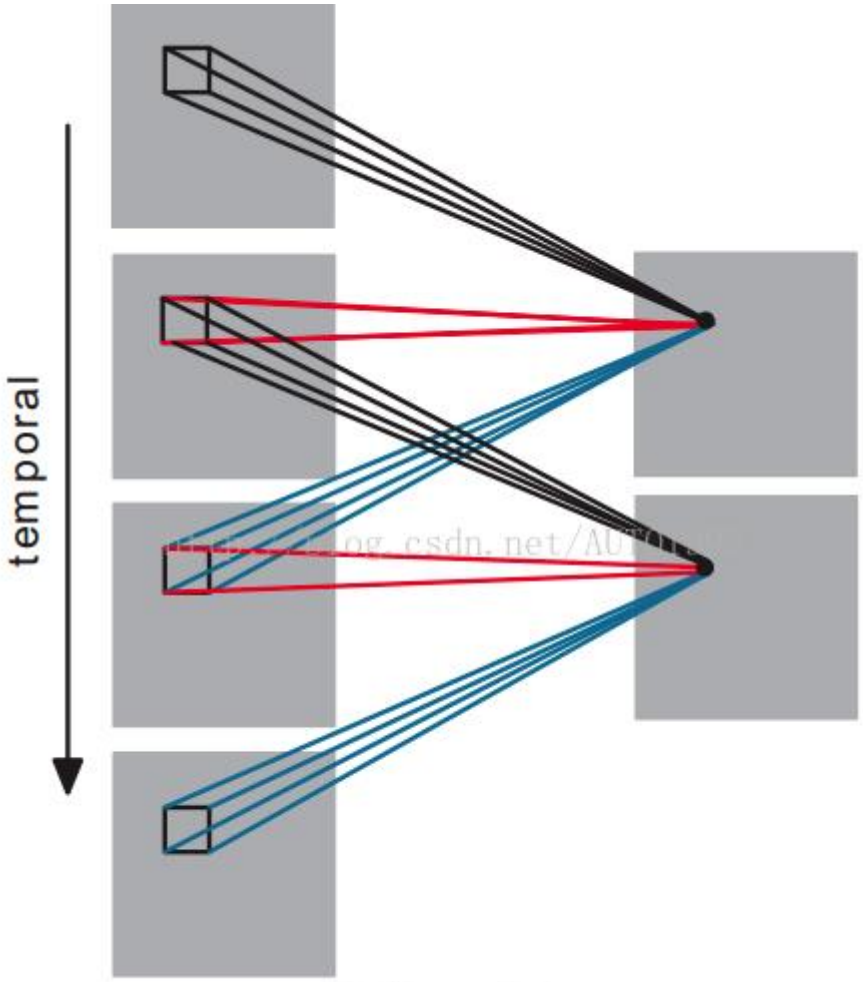
However, this operation of the ROI Pool cannot be segmented, because the position correspondence between the input and output ROI pixels is inconsistent. Since Mask R-CNN wants to expand on Faster R-CNN and implement instance segmentation, a new layer, ROI-Align is proposed to replace the ROI Pool layer in the original Faster R-CNN model.

8.5 3D CNN

3D CNN is improved based on 2D CNN. Since 2D CNN cannot effectively capture timing information, we use 3D CNN so that the timing information in the video can be used. Since humans have timing information when walking, this model is commonly used in gait recognition. By adding the time dimension to the input data of the neural network, the neural network can simultaneously extract temporal and spatial features for video processing and gait recognition. Regarding the time dimension as the third

dimension, convolution is performed on consecutive images, and a cube is formed by stacking multiple consecutive frames and using a 3D convolution kernel. In this case, each feature map will be connected to multiple consecutive frames of input data in the previous layer to capture motion information. For instance, convolving three consecutive frames with a three-dimensional convolution kernel can be understood as convolving three images with three different two-dimensional convolution kernels, and adding the convolution results. Through this processing, the network acquires a certain relationship of continuous time.

(a) 2D convolution



(b) 3D convolution

9. Advantages and disadvantages of gait recognition

The chief reason why gait recognition has developed dramatically is that gait recognition has long-distance recognition. Compared to facial recognition and fingerprint recognition, it is realized that other recognition tools can only be tested at close distances, and gait recognition has the advantages of long-distance recognition. Due to the feature of long-distance recognition, gait recognition can recognize at a distance and react in advance. For example, in some important places like military bases or bank vaults, if there are other unidentified persons appearing tens of meters away, the system might issue an alarm in advance, so that there will be enough time for a defense.

Another factor driving to the development of gait recognition is that gait recognition would not be affected by cover objects. Sunglasses or masks do not prevent gait recognition for identification. It is worth considering that gait recognition can be used in criminal tracking. In this case, some

criminals who attempt to conceal themselves will be recorded on the camera, and their identities can then be identified by gait recognition.

Besides, gait recognition technology can even be used as a special tool for athletes to adjust their posture. By analyzing the gait of champion athletes, we can analyze some techniques that are conducive to improving running performance. Other athletes can imitate the running posture of excellent athletes to get better results.

Nevertheless, it is worth noting that the success rate of gait recognition still needs to be improved. Camera angle, weather conditions, uneven roads and even clothing lighting will affect the accuracy, which makes it difficult for gait recognition to play its role and cannot be successfully recognized.

Indeed, gait recognition is a key tool for dealing with problems, but gait recognition can also be a source of problems especially if it is not managed well. It is worth considering how to use gait recognition without infringing on the privacy of others because gait recognition does not require people's cooperation.

In conclusion, as the only technology with long-distance recognition ability in the current recognition technology, gait recognition could be widely used

in various recognition applications. It is worth noting that this innovative technology may infringe human privacy based on its unique ability. Therefore, it is important to find a balance between the use of gait recognition technology and the protection of human privacy rights, or it could be used in some public places instead of private facilities. Moreover, its correct recognition ability still needs to be strengthened, which is also the main reason why the gait recognition technology has not been commercialized for the time being. In this case, it is considered that combining face recognition, iris recognition and gait recognition to create a more accurate recognition technology.

Reference

Doushagao(2019)'CNN-Convolutional Neural Network Input Layer', 30 July
Available at: https://blog.csdn.net/qq_38646027/article/details/97786102/ (Accessed: October 21, 2021)

Chanxin001（2019）'Understand the role of each layer of the convolutional neural network in detail', 17 January
Available at: https://blog.csdn.net/woai8339/article/details/86523967/ (Accessed: October 21, 2021)

W_bird(2018)' Convolutional neural network-convolutional layer, pooling layer meaning', 10 March Available at: https://blog.csdn.net/qq_30979017/article/details/79506593/(Accessed: October 21, 2021)

Zhu_Lydia(2019)' Understanding of channels, shared weights, feature mapping, etc. in neural networks and CNNs', 15 March Available at: https://blog.csdn.net/zhu_Lydia / (Accessed: October 21, 2021)

Yizhen(2018)' Deeplearning.ai Convolutional Neural Network', 14 December Available at: https://zhuanlan.zhihu.com/p/52145860/ (Accessed: October 21, 2021)

yaqiLYU(2019)'Local receptive field of convolutional neural network', 29 January Available at: https://zhuanlan.zhihu.com/p/44106492/ (Accessed: October 21, 2021)

Yeluzichengxuyuan(2017)' [Neural Network and Deep Learning] Notes', 18 September Available at: https://www.cnblogs.com/yeluzi/p/7521781.html /(Accessed: October 21, 2021)

RecFaces(2021)' Gait Recognition System: Deep Dive into This Future Tech', 18 June Available at: https://recfaces.com/articles/what-is-gait-recognition/ (Accessed: October 21, 2021)

Gentelyang(2018)' RCNN, Fast RCNN, Faster RCNN finishing summary', 27 May, 2018 Available at: https://blog.csdn.net/gentelyang/article/details/80469553/(Accessed: November 22, 2021)

echo_hao(2019)'Interpretation of Mask R-CNN Principle', 27 March,2019 Available at: https://blog.csdn.net/qq_37392244/article/details/88844681/(Accessed: November 22, 2021)

Lavi_qq_2910138025(2018) 'Understand 3D CNN and 3D convolution', 18 November, 2018 Available at: https://blog.csdn.net/liuweiyuxiang/article/details/84202352/(Accessed: November 22, 2021)

Bobby0322(2019)' What is a convolutional neural network', 21 February, 2019 Available at: https://www.jianshu.com/p/1ea2949c0056/(Accessed: November 22, 2021)

Piupiurui(2020)'One-dimensional convolution (1D-CNN), two-dimensional convolution (2D-CNN),three-dimensional convolution (3D-CNN)', 5 June,2020 Available at: https://blog.csdn.net/yizhishuixiong/article/details/106566730/(Accessed: November 22, 2021)

marleylee（2018）'https://blog.csdn.net/marleylee/article/details/81294291', Convolutional neural network-input layer, convolutional layer, activation function, pooling layer, fully connected layer, July 30, 2018 Available at: https://blog.csdn.net/marleylee/article/details/81294291