Pedestrian Attribute Recognition Based on Deep Learning

Peng Wang*, Qikun Wang, Shengfeng Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*E-mail: *2324365941@qq.com*

www.tust.edu.cn

Abstract

This paper studied pedestrian attribute recognition based on deep learning, for its importance in the fields of smart city construction. Firstly, the research status of pedestrian attribute recognition and common deep learning models was introduced. Secondly, considering the accuracy decline problem and gradient problem of the neural network, the residual network was used as the main body of the neural network model. Thirdly, the model was trained to classify multiple person attributes through two data sets, Market-1501 and DukeMTMC-reID. Finally, the pedestrian attribute recognition model was tested, and good results were obtained.

Keywords: Neural networks, Pedestrian attributes, ResNet50

1. Introduction

The contribution of attribute recognition technology in medical, security, intelligent furniture and other fields has attracted more and more attention. After the introduction of deep learning algorithms in the computer field, the computer has realized the processing and application of massive information through continuous learning. The person attribute recognition based on deep learning extracts the feature information of a known pedestrian photo through the convolution and pooling network model, and classifies it to obtain several attributes about the person. The acquisition of these attributes brings important practical applications to the fields of smart city construction and military security.

In this paper, we first introduce the research background and current research status of pedestrian attribute recognition. In the second chapter, the theoretical basis of deep learning will be introduced, and the attribute recognition of pedestrians will be mainly studied and verified by relevant experiments. In the third chapter, the full text is summarized, discussed and analyzed.

2. Methods and results of pedestrian attribute recognition

In the field of computer vision, deep learning has become a basic tool. In this study, a large number of deep learning concepts will be used, and deep learning neural networks will be used to identify pedestrian attributes. Finally, results will be obtained in experiments..

2.1. The emergence of artificial neural networks

T Landahl et al. first proposed artificial neural network, which is a network model built by imitating the connection of nerve cells in the nervous system of animals. As shown in Fig. 1, each circle represents neurons and arrows represent the direction of signal transmission, which is similar to the synapses in nerve cells to transmit signals.

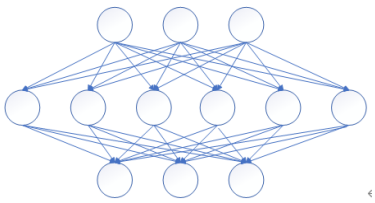


Fig. 1 Artificial neural network structure diagram

As the basic part of the neural network model, the calculation formula of neurons is as follows:

$$y = f_{active}(\sum_{i=1}^{n} w_i * x_i + b) \quad (1)$$

Where, xi represents the ith input of the neuron, wi is the weight parameter of the input, which can be changed, n represents a total of n neuron inputs, and b is the bias value. Here is called the activation function, a neuron multiplies the input with a weight and then adds a bias value, which is output to the next layer through the activation function.

In the deep learning network, the weight value and bias value are the parameters that need to be learned. Through continuous learning, the gap between the output value and the real value becomes smaller and smaller.

2.2. Principle analysis of convolution layer

As shown in Fig. 2, the function of the convolution layer is to use multiple convolution operations to extract multi-channel eigenvalues from the output of the upper layer, and then send these eigenvalues as outputs to the next layer. The convolution layer reduces the parameters to be trained in the process of layer by layer extraction. In a given image, the local pixel information is converted to the corresponding information of the output image after weighted addition. The part of the convolution layer used for convolution is called the convolution kernel, its size can be defined, and the parameters in it are the parameters we want to train.

Other parameters we define in the convolution kernel determine the effect of image data extraction, including fill, number of output channels, number of input channels, and step size. The size of the convolution image is:

$$H_{out} = \left[\frac{H_{in}+2\times padding[0]-dilation[0]\times(kernelsize[0]-1)-1}{stride[0]} + 1\right] \quad (2)$$

$$W_{out} = \left[\frac{W_{in}+2\times padding[1]-dilation[1]\times(kernelsize[1]-1)-1}{stride[1]} + 1\right] \quad (3)$$

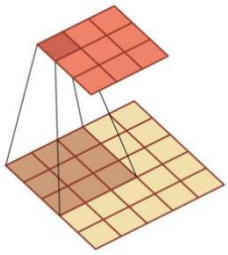


Fig. 2. Schematic diagram of convolution operation of 3*3 convolution kernel

2.3. Principles and advantages of ResNet network

With the more and more extensive application of deep learning in computer vision, the depth of neural network is also deepening. However, it is found that with the deepening of network depth, the results obtained are not getting better and better, and the problem of gradient disappearance is becoming more and more serious. Therefore, He et al proposed residual convolutional neural network. The residual block structure is added to the network structure, and the short circuit design is added next to the convolutional layer to solve the problem of the gradient disappearing and the accuracy of the training set.

ResNet is divided into 18, 34, 50, 101 and 152 layers, which differ in the number of layers with 4 groups of convolutional layers in the middle.

2.4. Meaning of loss function

The loss function is an operation function used to measure the difference between the predicted value f(x) of the model and the real value Y. It is a non-negative real value function, usually expressed by L(Y, f(x)). The smaller the loss function, the better the robustness of the model. We use is called the two-dimensional cross-entropy loss function. This function is a binary classification function, BCELoss is the binary loss function between the target value and the predicted value, the formula is:

$$l_n = -w_n * [y_n * log\, x_n + (1 - y_n) * log(1 - x_n)] \quad (4)$$

Where Wn represents the matrix of weights, Xn represents the prediction matrix of function output, and Yn represents the target matrix.

2.5. Introduction and application of data sets

In recent years, thanks to the exploration of many researchers on pedestrian attribute recognition, there are many open source data sets that can be used, and we will mainly use the following two data sets.

First, the first data set is Mark-1501 data set.

3-1(a)Picture under camera a

3-1(b)Pictures in other cameras

Fig. 3 Markrt-1501 data set

As shown in Fig. 3, the Mark-1501 dataset was originally used for gender reidentification research, and was filmed on the Tsinghua campus using six cameras with different viewing angles. Lin et al. [1]. annotated attributes for each person in this dataset. In this dataset, a total of 1501 pedestrians and 32,668 character rectangles are included, as shown in Table 1. In so many images, there are two parts: the training set and the test set.

Table 1 Labeling probabilities of some data attributes for Mark-1501

label	probability	label	probability
Young	0.0186	Adult	0.2130
Old	0.0107	Bag	0.2463
Handbag	0.1145	Downblue	0.1638
Downbrown	0.0919	Downgreen	0.0186
Downpink	0.0386	Downwhite	0.0772
Downyellow	0.0133	Upblue	0.0613
upgreen	0.0746	Uppurple	0.0399
upred	0.3901	Upyellow	0.3901
clothes	0.3901	Up	0.9481
hair	0.3262	gender	0.4261

Then, the second data set is DukeMTMC-reID.

In the training set, there were 12,936 images, including 751 pedestrians, with an average of 17.2 images per pedestrian. In the test set, there were 750 pedestrians with 19,732 images, an average of 26.3 images per person.

As shown in Fig. 4, the DukemtMC-Reid dataset is a subset of the DukeMTMC dataset. On the campus of Duke University, eight cameras were used to capture the video stored in the form of a pedestrian border box, each frame of which was manually marked by someone. In the video, images are captured every 120 frames, and the resulting images make up the data set. There were 1,404 pedestrians in the dataset, most of whom were captured by two or more cameras, and 36,411 images made up the DukeMTMC-reID dataset. In this data set, the training set consists of seven hundred and two images of people in rows randomly selected, and the remaining images are used as the test set. It is worth noting that there are 408 people caught by only one camera, and they are included as interference items in the data set. Lin et al. labeled a total of 23 attributes in this dataset, including gender, whether to wear boots, whether to wear a hat, whether to have a backpack, shoe color, seven lower body clothing colors and eight upper body clothing colors.

4-1(a)Picture under camera a

4-2(b)Pictures in other cameras

Fig.4.DukeMTMC-reID

2.6. Evaluation index design

The number of correct identifications for each semantic attribute and the number of all samples is calculated, and the quotient of these two numbers is evaluated as the accuracy of each attribute identification. The average accuracy of all attributes can indicate the extent of the model effect. The formula used is as follows:

$$acc_i = \frac{T_i}{N} \tag{5}$$

$$acc_{mean} = \frac{\sum_{i=1}^{c} acc_i}{c} \tag{6}$$

In addition, due to the great imbalance of pedestrian attribute sample data, the negative sample of a certain

attribute may occupy more than 90%. Therefore, a balanced indicator is also needed, and the average accuracy rate (AP) can balance the identification accuracy of positive and negative samples. The mAP of the average accuracy of each attribute can show the superiority of the algorithm on unbalanced data sets. For the ith attribute, TPi represents the number of positive samples correctly predicted, Pi is the total number of positive samples, TNi represents the number of negative samples correctly predicted, and Ni is the total number of negative samples, then:

$$AP_i = \frac{\frac{TP_i}{P_i}+\frac{TN_i}{N_i}}{2} \tag{7}$$

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{c} \tag{8}$$

For the research of pedestrian attribute recognition, we should not only pay attention to the accuracy of individual attributes, but also pay attention to how many attributes in each image can be successfully recognized. This index mainly statistics the accuracy (acc.), accuracy (prec.), recall rate (Recc.) and F1 values of attribute recognition at the sample level.

Accuracy: In both positive and negative cases, the proportion of the predicted correct number to the total number is expressed by the formula:

$$ACC = \frac{TP+TN}{TP+FP+FN+TN} \tag{9}$$

Precision is relative to the prediction results of positive examples. The accuracy of the predicted positive examples is evaluated by the proportion of the real positive examples in the predicted positive examples. The formula is as follows :

$$precision = \frac{TP}{TP+FP} \tag{10}$$

The recall rate is judged according to the actual sample. Its main purpose is to judge the proportion of the predicted positive example in the actual positive example, and can be expressed by the formula :

$$recall = \frac{TP}{TP+FN} \tag{11}$$

The F1 score is the harmonic mean of the correct rate and the recall rate, defined as :

$$F1 = \frac{2*precision*recall}{precision+recall} \tag{12}$$

2.7. experimental result analysis

The conclusion on Mark-1501 is the average accuracy is 0.9631; the average F1 score is 0.6492. The following figures (Fig. 5 and Fig. 6) show the recognition results of some attributes.

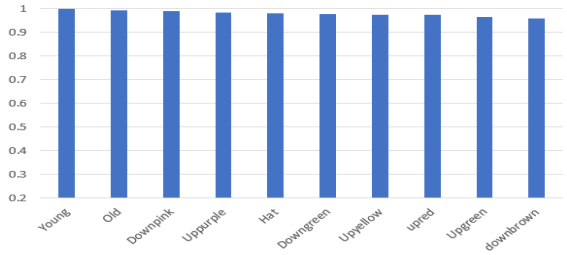


Fig.5. Accuracy of some attributes

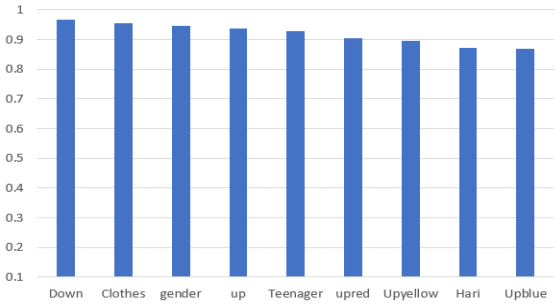


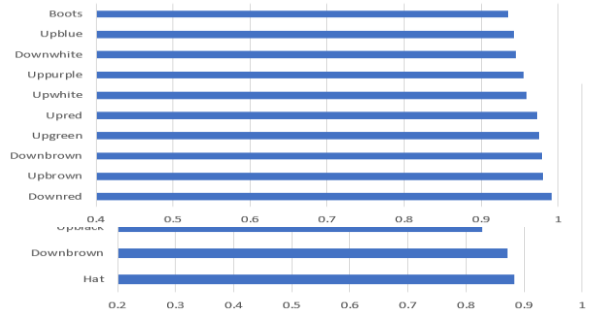
Fig.6. Precision of some attributes

On the DukeMTMC-reid dataset, we concluded that the average accuracy was 0.9152 and the average F1 score was 0.5739.The following figure (Fig.7 and Fig.8)shows the recognition results of some attributes.As shown in the figure, figure 7 shows partial accuracy, and figure 8 shows partial precision.

Fig. 7 Accuracy of some attribute

Fig. 8 Precision of some attributes

3. Conclusion



In this paper, nonlinear functions are introduced into a four-dimensional conservative chaotic system to generate multiple scrolls. After the introduction of nonlinear function, the equilibrium point of the system changes from a fixed point to a set of equilibrium points. The system carries out basic characteristic analysis, and discusses its divergence, equilibrium point and whether the energy is

conservative. The equilibrium points obtained by introducing one-dimensional nonlinear function are divided into two categories. For its Lyapunov exponent analysis, after introducing the sine function without multiple angles, The Lyapunov exponents of the system equations show similar periodic characteristics to the sine function, and the Lyapunov exponents obtained by changing the initial values are very different. With the change of initial value, the phase diagrams obtained are also different, and the number of vortex attractors formed is also different, which verifies the multi stability.

Then, the nonlinear function is extended, two nonlinear functions are introduced, and the system with two nonlinear functions is further analyzed. The obtained phase diagram changes from one-dimensional to two-dimensional scroll attractor, and the relationship between the number and arrangement of scroll and the threshold width is obtained. By changing the initial value, the phase diagram with different internal distribution but the same number of scroll is obtained.

References

1. LIN Y, ZHENG L, ZHENG Z, et al. Improving Person Re-identification by Attribute and Identity Learning. arXiv:1703.07220 [cs.CV], 2017.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

.

Mr. Shengfeng Wang

In 2023, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China.He is pursuing a master's degree in engineering from Tianjin University of Science and Technology.

Mr. Qikun Wang

In 1996, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China. He is a senior engineer, research direction is HVAC, building technology.