Exploring The Efficiency of Resnet and Densenet in Gender Recognition

Yuan Jiang*

School of Information and Intelligence Science and Technology, Hunan Agriculture University, Changsha, Hunan, 410128, China

* Corresponding Author Email: yzb@hunau.net

Abstract. With the development of face recognition and image classification. There are many scenarios where gender classification techniques need to be applied. For example, the stores can recommend different products according to the gender of customers. There are several methods of gender classification, one of them is the convolutional neural network (CNN), which shows its promising potential in image classification. With other relevant variables and parameters being the same, this work tries to compare the Resnet with the Densenet to find a model that is suitable for specific gender classification datasets. The classification efficiency of the model is examined from the three directions of running prediction loss. This paper finds that in the case of small-scale data sets, the classification performanceof the Densenet model ourperforms that of the Resnet model, and the number of parameters used by the Densenet is significantly less than that of the Resnet. When a high-precision classification or recognition model is needed, Densenet is preferred.

Keywords: Gender classification, deep learning, Densenet, Resnet.

1. Introduction

Face analysis attracts researchers' attention. Some surveys are written for face recognition [1] and on face detection [2, 3]. Facial. With the maturity of deep learning technology, face recognition technology has been applied to more and more scenarios, such as human-computer interaction, Targeted advertising, identity verification and census. At the same time, gender recognition is an important part of face recognition, which can help face recognition to complete face recognition work more efficiently and quickly. Because if the ratio of men and women is equal and the gender is identified in advance, the query volume can be reduced by about half on average, which can greatly improve the efficiency of face recognition. Automatic gender classification based on face images is quite useful. It could be used in many human-computer interaction applications. Therefore, accurate facial feature extraction and gender recognition are becoming more and more important. This paper will prove the more efficient model of both the Resnet [4] and the Densenet [5] of CNN model in the current gender classification model. There is a lot of information on human faces, such as eyebrow bones and beards, which are important information for distinguishing male and female. In this experiment, this information is used as features for extraction, and use Resnet model and Densenet model of CNN for classification. Compared the Resnet model and the Densenet to discuss their accuracy and speed for gender classification to determine a more efficient model for use. The comparison is based on the image datasets called Gender Classification Datasets that from the Kaggle [6]. For the traditional model, the Resnet model is used, which detector by B. Moghaddam, its performance is better than other traditional classifier models (linear, quadratic, Fisher linear discriminant, etc.). For the experimental results, the accuracy of it is also more stable, which also helps to compare with Densenet model of convolutional neural networks.

2. Method

2.1. Dataset

The image data of this experiment comes from the gender classification datasets in Kaggle [6]. It includes images of human faces with both genders. There are about 23,000 images for training and about 5,500 images for validation. In addition, 10% of the training will be selected set as the test set. The selected image data does not need data cleaning, because the authors of this dataset have already completed this step. Some examples are shown in Figure 1.

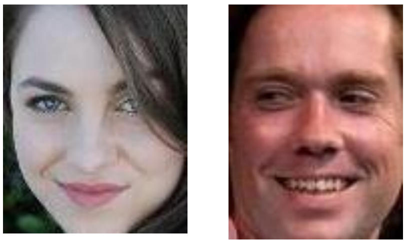

Fig. 1 Example images of the dataset.

2.2. Pre-processing

2.2.1 Grayscale adjustment

The final image obtained by face image processing is usually a binary image, and because of the difference in location, equipment, lighting, etc., the quality of the collected color face image is different, so the image should be processed in a unified grayscale to smooth out these differences. In practical applications, grayscale adjustment can be carried out by using commonly used methods such as average method, histogram transformation method, power transformation method, and logarithmic transformation method. The purpose of Grayscale Adjustment is to improve the quality images [7], and to make them clearer. Image gray-scale transformation is a fundamental operation in image enhancement technology. However, only using the gray processing method to improve the accuracy is not significant enough. In this paper, image filtering and normalization are also added to the data preprocessing, aiming to improve the classification accuracy as little as possible while affecting the classification speed.

2.2.2 Image filtering

In practical applications, face collection is carried out on site, and the quality of face images will be affected by various noises from all aspects, such as a large number of electromagnetic signals in the surrounding environment, digital image transmission by electromagnetic signal interference and other channels affected, thus affecting the quality of face images [8]. To guarantee the quality of the face image, the pre-preprocessing will reduce the noise of the image to reduce the impact of noise on the subsequent situation. There are many principles and methods of noise removal treatment, and common ones in applications include mean filtering, median filtering, etc. Median filtering algorithms are now commonly used to preprocess face images. The experiment uses bilateral filtering. And the reason why is that the traditional median filtering and Gaussian filtering do not preserve the edge, this paper presents an image filtering method that can preserve the edge. Bi-lateral refers to the fact that this filtering method considers two aspects of information: the information lies in spatial domain and the pixels' intensity information. In the implementation, only the filtering function of the Gaussian filter in the spatial domain is multiplied by the pixel 's closeness, which is very simple.

$$h(x) = k_d^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi)\, c(\xi, x) d\xi \quad (1)$$

The front of the integral number is the normalization factor. Here, this work considers weighting all the pixels. C and s are closeness and similarity functions, x represents the required point, and f(x) denotes the intensity of the point. f(x)->h(x) is the image before and after filtering. The bilateral filtering in paper called "Bilateral Filtering for Gray and Color Images" has more advantages than Gaussian filtering and ordinary filtering in maintaining edge-preserving smoothing [9].

2.2.3 Image size normalization

When face recognition technology performs simple face training, it will encounter different image pixel sizes in the face library, which requires normalization of the size of the image before face recognition comparison. In use, more common size normalization algorithms are required, such as bilinear interpolation algorithm, nearest neighbor interpolation algorithm and cubic convolution algorithm for processing applications.

2.3. Model

2.3.1 Perliminary knowledge

CNN is widely used in machine learning, which components are three folds, including the convolutional layer, maxpooling layer and fully connected layer.

The convolutional layer mainly extracts features, which hyperparameters include kernel size, step length and filling. Jointly, they decide the shape of the feature map output from the convolution layer. The size of the kernel could be set to a value smaller than input. If the kernel is large, it could beneficial to extract complex spatial correlations and hence more semantic features. The step size determines the location to conduct convolution. If the step size is set to 1, all elements in the entire feature map will be scanned. It the step size equals n, then n-1 points will be omitted between two scans. With the calculation of the convolution in a network, the feature map size will gradually be decreasing.

The downsampling could lower computationally cost, but it will not decrease the performance. When the data is convoluted, the dimension is getting higher. After continuous multiple convolutions, many parameters will be generated, which will not only increasing the computational burden, but also easily cause overfitting. Therefore, a pooling layer is usually applied after the convolution to compress data, reduce dimension, and reduce the number of weights.

Fully connected layer could be used to classify images. After multiplying convolutions and poolings, some fully connected layers are connected as the end of the network. The fully connected layer gathers global information for chassificaiton. To improve the nonlinearity of CNN, activation functions such as the ReLU will be adapted after each fully connected operation. The SoftMax layer is always placed at the end of a neural network to transfer output feature to probabilities.

2.3.2 Resnet18

The Resnet is based on blocks. Each block is made up of a shortcut and some convolutions. The shortcut passes and combines the information in the middle layers. Doing so will not generate additional weights, and ensure that the performance of the network is stable. Most importantly, it alleviates gradient diminishing to ensure the trainable of very deep network. Moreover, to guarantee the same input and output dimensions, 1x1 convolution dimension up or down to match can be added. The paper [4] explains that Resnet can effectively solve the problems of vanishing gradients, exploding gradients, and network degradation. In Resnet, the output of layer l can be expressed by the following formula.

$$x_{\ell} = H_{\ell}(X_{\ell-1}) + X_{\ell-1} \tag{2}$$

2.3.3 Densenet

DenseNet takes a more densely connected approach, a dense CNN that propagates forward by densely connecting each layer to the subsequent layers. It is designed to make sure that information at multiple levels could be combined freely. In DenseNet, a single layer is connected to many other layers, so there will be L (L + 1) / 2 connections. In each layer, feature map in the previous layers is used as input, and its own feature mapping is also used as input for all the subsequent layers.The paper [5] describes four salient features of densenet. 1.that mitigate gradient vanishing 2. Enhanced dissemination of features across networks. 3.Feature reuse implemented and enhanced. 4.Effectively reduce weights [10]. In DenseNet, the current layer is densely connected to all subsequent layers, and the output of the first layer can be denoted as:

$$x_{\ell} = H_{\ell}([X_0, X1, \dots, X_{\ell}]) \tag{3}$$

The [X0, X1, ..., Xl-1] represents the feature concatenation of tensors from the 0- (l-1) layers.

3. Result

This paper compares the performance of Densenet and Resnet in gender classification by comparing the number of parameters, classification accuracy and classification loss. The results are shown in Table 1. Params is the number of parameters, Train Acc and Validation Acc are classification accuracy, and Loss is classification loss.

Table 1. Performances of the Densenet and Resnet.

Method	Depth	Params	Train Acc	Train Loss	Validation Acc	Validation Loss
Resnet18	110	1.7M	95.59%	12.21%	96.75%	6.64%
Densenet(k=12)	40	1.0M	96.38%	11.51%	97.23%	5.73%
Resnet18	164	10.2M	96.21%	11.89%	96.96%	6.55%
Densenet(k=12)	100	7.0M	96.41%	11.32%	97.35%	5.68%

According to the data, Densenet obtains higher classification accuracy and lower classification loss with fewer parameters than Resnet in this dataset scenario, and shows better performance in gender classification in this case.

4. Discussion

The final performances of this paper are the same as the hypothesis. In this data set environment, Densenet has more advantages in gender classification than Resnet, because Desnet can retain more features in the cross-layer concatenate operation, which is more accurate than Resnet. Through experiments, it is verified that Densenet can be applied to image recognition in life. It can play an ideal effect in work that requires accuracy, such as in medical image recognition. Previous works uses Densenet for the identification of Predicting COVID-19 Using CT Image, and has achieved good results. However, compared with Desnet, Resnet has better performance in classification speed. Therefore, Resnet is more suitable for large-scale classification and less demanding accuracy. But most life situations require the most accurate identification and classification methods, so Densenet should have better prospects and more areas of use.

5. Conclusion

The purpose of this study is to find out the network models that apply to sex classification. It has examined that Resnet and Densenet 's performance in sex classification accuracy and discussed the applicability of both in real life applications. The experimental results show that Resnet and Densenet have good performance in the accuracy of gender classification. The accuracy is more than 95 %, but the classification loss is more than 5 %. In the experiment, Densenet can extract more features, Resnet has faster classification speed, but in the case of fewer parameters, Densenet classification is more accurate and less classification loss. Therefore, Densenet performs better than Resnet in small-scale gender classification. The study highlights the importance of exploring the differences in performance between Resnet and Densenet, uncovering the different advantages of Resnet and Densenet. Theoretical, this paper fills the gap in the comparison of classification models in gender classification, and provides a reference for readers to select network models according to the needs of the experiment, so that readers can choose between models according to their needs. In practice, it provides insights for experiments that need better models to achieve high precision. This study considers several limitations, 1. The experimental data set is not large enough to match the actual industrial data scale. 2. There are too few Resnet models explored to cover resnet50 and so on. Increased data size, increased model types, and more comparative data to choose from in post-experiment.

References

[1] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: A literature survey. ACM computing surveys (CSUR). 2003, 35(4):399-458.

[2] Hjelmås E, Low BK. Face detection: A survey. Computer vision and image understanding. 2001, 83(3):236-74.

[3] Yang MH, Kriegman DJ, Ahuja N. Detecting faces in images: A survey. IEEE Transactions on pattern analysis and machine intelligence. 2002, 24(1):34-58.

[4] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016, 770-778.

[5] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017, 4700-4708.

[6] Ashutosh Chauhan. Gender Classification Dataset. URL: https://www.kaggle.com/datasets/cashutosh/gender-classification-dataset. 2019.

[7] Wang J, Lee S. Data Augmentation Methods Applying Grayscale Images for Convolutional Neural Networks in Machine Vision. Applied Sciences. 2021, 11(15):6721.

[8] Liu S, Tian G, Xu Y. A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. Neurocomputing. 2019, 338:191-206.

[9] Tomasi C, Manduchi R. Bilateral filtering for gray and color images. InSixth international conference on computer vision, 1998, 839-846.

[10] Zhang K, Guo Y, Wang X, Yuan J, Ding Q. Multiple features reweight densenet for image classification. IEEE Access. 2019 Jan 14; 7:9872-80.