Implementation and Comparison of Convolutional Neural Network (CNN) and Local Binary Patterns Histogram (LBPH)

Xinyu Liang, Shangrui Xie, Baohui Zhu
Master of Science in Information Systems
College of Engineering, Northeastern University
Boston, MA 02115

April 26, 2019

ABSTRACT

As a computer vision technology, face recognition has been widely used in medicine, security, education and many other fields. Two kinds of facial recognition algorithm are applied in this project, CNN and LBPH. Each of them has different advantages. In this paper, the image encoded by the local binary matrix is obtained and used as input to train the network. With the data set of 'Labeled Faces in the Wild Home', the two algorithms were results in accuracy of 90% and 59% respectively.

***K*eywords** LBPH · CNNs · facial recognition

1 Introduction

Face recognition is a typical research topic in pattern recognition, image analysis, and understanding. Face recognition [1] not only has important theoretical value but also has important practical value in practice, such as case detection, identity recognition, and other fields. However, due to the changes in image acquisition conditions and the changes in the face itself, different photos of individuals are quite different, sometimes even completely different, and the results may be mistaken for photos of different people, which brings difficulties to face recognition. Therefore, improving the accuracy of the face recognition system is one of the important goals of researchers.

Face feature extraction is very important to face recognition system so finding an effective feature to describe a face is the key to face recognition. Principal Component Analysis (PCA) [2], Linear Discriminant Analysis (LDA) and others in face recognition describe the face image from a global perspective, which can better extract the global features of the image but is susceptible to the influence of light and location. Local binary mode (LBP) was first proposed by Ojala [3], which is not sensitive to light and can well extract local texture features of images. In recent years, LBP has been widely used in texture classification, face image analysis, and other fields.

2 Methods

2.1 Train Local Binary Patterns Histograms Model

LBP is a kind of texture description operator. Its basic principle is to select the center pixel of the image as the threshold value and get a string of binary codes by comparing the pixel values in the field. The local texture features of the image are described by this binary code. The specific calculation method is that pixels greater than or equal to the threshold are denoted as 1; Otherwise, it's 0. [4] As shown in Fig. 1, this binary code is read in turn and then converted to a decimal number, which serves as the LBP value of the center point.

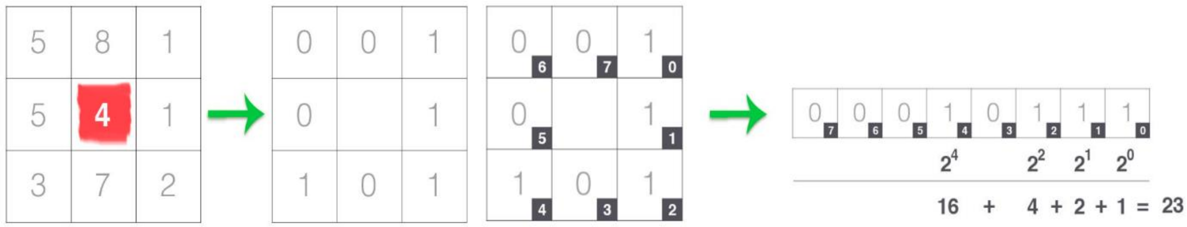


Fig. 1. LBP operator

LBPH (Local Binary Patterns Histograms) is a Local Binary coding histogram. Firstly, each pixel is taken as the center to judge the relationship between the gray value of surrounding pixels and its Binary coding, so as to obtain the LBP coding image of the whole image. Then the LBP image is divided into gradx*grady areas, and the LBP coded histogram of each region is obtained, and then the LBP coded histogram of the whole image is obtained. [5] The purpose of face recognition is achieved by comparing the LBP coded histogram of different face images. Its advantage is that it will not be affected by illumination, scaling, rotation, and translation.

Local Binary Pattern Histogram (LBPH) is a simple but very efficient algorithm to extract important information in the images, which based on Local Binary Pattern (LBP). Same as LBP, the LBPH uses the center pixels of the extracted field as a threshold. And with this threshold, we can fetch the points on a circle, decided by a given radius, and set them into 0 and 1, which based on the value of the threshold. And then we list the value of each point into an array in clockwise order, and we regard the list of number in the array as a binary number so that we can get a decimal number based on that. [6] The process is shown in figure 2. Using the decimal number to replace the whole area, we can get an image with fewer pixels but without losing important features. We can get the LBP code of each pixel in the image from the extracted binary number and get a local binary mode encoded image from the original image which is shown in figure 4.

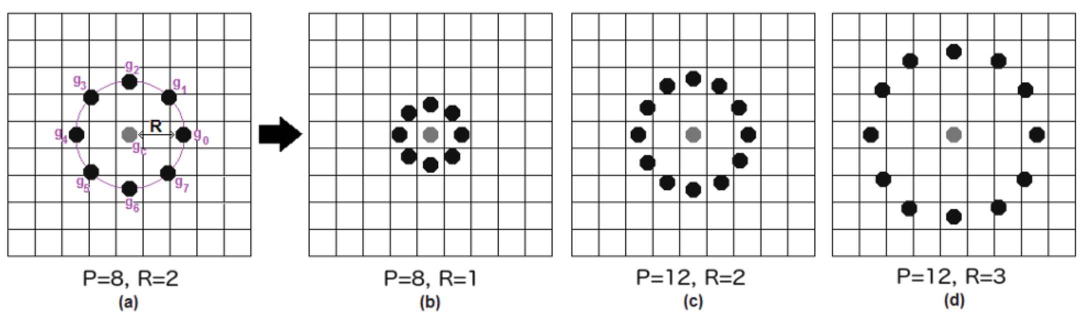


Fig. 2 LBPH principal image

(a) When the radius is 2, the 8 points on the circle are used to represent the whole area. (b)When the radius is 1, got the nearby 8 points to represent the area. (c) When the radius is 2, got the 3 points on the middle of each edge to represent the area. (d) When the radius is 3, got the point as the graph shows.

Fig. 3 LBPH data extraction

In fig. 3, it shows the first 3 steps of data extraction. We could concluded that the main feature of the graph remains no matter how the graph has been extracted.

2.1.2 Code Reference

1. https://towardsdatascience.com/face-recognition-how-lbph-works-90ec258c3d6b

2. https://en.wikipedia.org/wiki/Local_binary_patterns

Above are the main references, got the idea of LBPH and the way to build the algorithm based on the python.

3. https://docs.opencv.org/2.4.13.7/modules/contrib/doc/facerec/facerec_tutorial.html

A project of the OpenCV model, there is the LBPH code in C in this page. The python code of this project is based on the logic of this page and did improvement.

4. https://www.pyimagesearch.com/2015/12/07/local-binary-patterns-with-python-opencv/

A webpage of the LBPH in python, but the code is based on the link under 3, this one is only for the reference.

2.2 Convolutional Neural Networks Model

2.2.1 Train Convolutional Neural Networks Model

The convolutional neural network model is a kind of supervised learning network model. The basic idea is to construct several convolutional layers, pooling layers and full connection layers, and obtain representative feature information through data training, so as to classify and predict samples.

Convolutional neural networks avoid explicit feature sampling and implicitly learn from training data. As a result, the convolutional neural network is obviously different from other classifiers based on neural network, and the feature extraction function is integrated into multi-layer perceptron through structural reorganization and weight reduction. [7] It can directly process grayscale images and can be directly used to process image-based classification. The convolution network has the following advantages in image processing compared with the general neural network: a) the input image and the network topology can be well matched; b) feature extraction and pattern classification are carried out at the same time and are generated during training; c) weight sharing can reduce the training parameters of the network and make the structure of the neural network simpler and more adaptable.

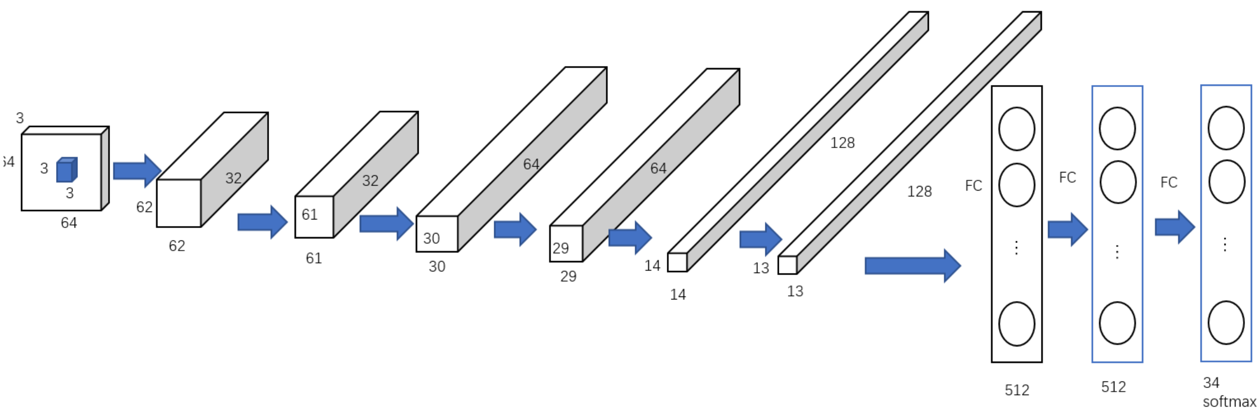


Fig.4 CNN model realization

Fig.4 is the structure of the convolutional neural network proposed in this paper. It has many layers, including convolution layers, max-pooling layers, dropout layers, fully connected layers, ReLU layer, Batch Normalization layer and output layer. The size of the input image is 64*64 pixels and RGB (red, green and blue) colors. The experiment steps are as follows.

According to the following figures, face images become more and more abstract because of losing features. However, each channel will retain some important information about facial features and give up irrelated information, such as background and clothes. After many layers, the model will get almost all important features and use them for classifying.

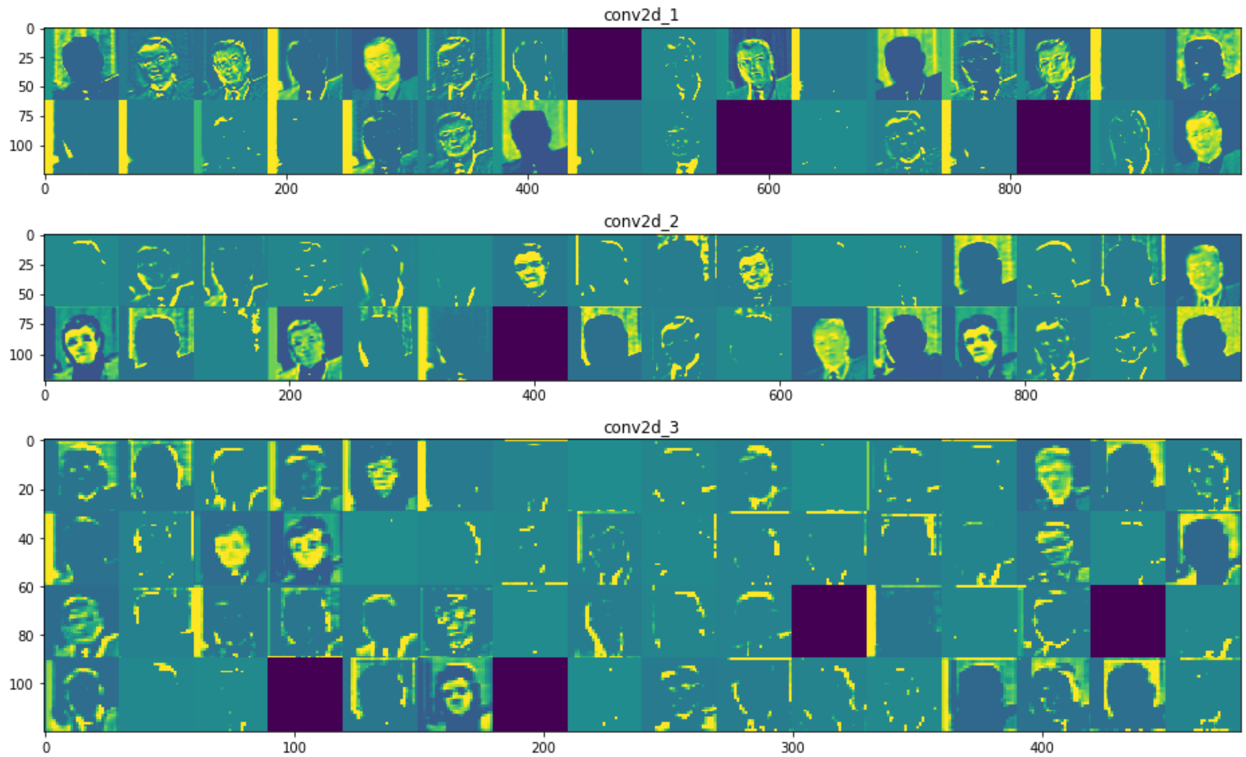


Fig.5 Visualization of conv2d layers

Conv2D layer means each pixel will do convolution with convolution kernel, so a two-dimensional matrix will be transformed to another two-dimensional matrix. During this process, several input features will be converted to a new feature, which usually is more important and iconic. ReLU layer is defined as "y=max (0, x)" mathematically and it is the most common activation function in CNN. The max-pooling function will take the max value of a matrix and create a new matrix with the values. This process will reduce the computational cost and will maintain the original features of the picture to some extent.

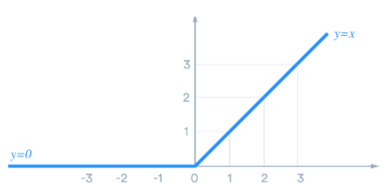


Fig.6 ReLU

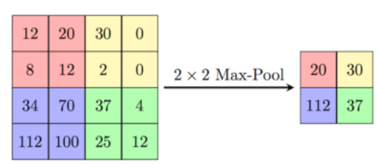


Fig.7 Max-Pooling

Dropout layer is usually used to improve the neural network by reducing over-fit. It is usually used on fully connection layers or after max-pooling layers. Batch Normalization layer will normalize the inputs of each layer, so a model is able to be trained with a higher learning rate and also helps reduce overfitting by regularization. While it is used together with dropout, the model will train better. Fully connection layers connect every neuron from two layers which could be used for classifying.

2.2.2 Code

1. https://keras.io/examples/cifar10_cnn/

The Keras document example of a deep CNN model. A new CNN model has been created based on this link and the parameters are also been adjusted in order to get a higher accuracy score.

2. https://github.com/seathiefwang/FaceRecognition-tensorflow

A project of face recognition which is used to classify the writer and others. The image preprocess has been used to accommodate to this project.

3. https://github.com/fchollet/deep-learning-with-python-notebooks/blob/master/5.4-visualizing-what-convnets-learn.ipynb

A tutorial of CNN layers visualization. Some of the code has been used in this project to draw CNN layers images.

4. Other

The code for training images-set selection (a person with more than 30 images) and the code to train model by using GPU.

3 Results

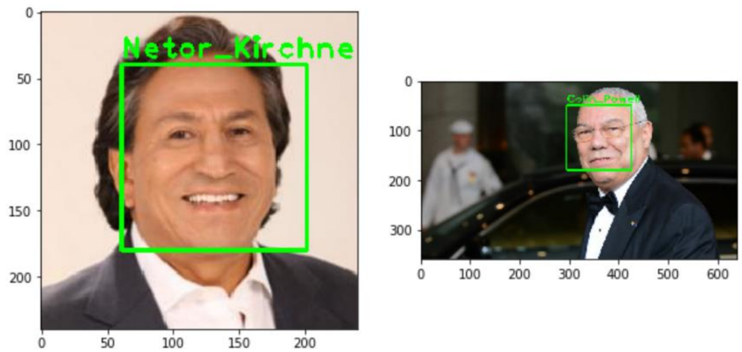
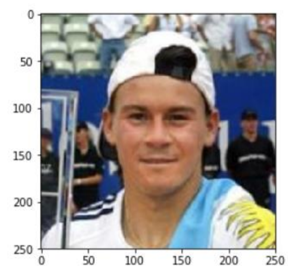


Fig.8 Example Prediction of LBPH




```
result: [[4.0370212e-03 7.8032826e-07 4.4045702e-02 7.5837893e-06 2.2742615e-05
  3.9950196e-06 8.1902081e-06 2.3784563e-10 5.5442506e-06 1.0758594e-06
  3.2100621e-01 6.0325509e-01 1.0417468e-08 7.5813309e-06 1.1734722e-06
  6.8127670e-11 7.7276200e-05 6.2944478e-06 2.5066044e-13 2.6305694e-02
  2.4370522e-07 2.2602784e-07 1.2161671e-04 3.9733202e-08 9.7043859e-04
  4.3276636e-09 3.3645719e-07 6.9365327e-05 4.4544613e-05 2.6657704e-08
  2.2876372e-09 1.0700264e-07 3.3854253e-09 1.0270230e-06]]

This person is:
['Guillermo_Coria']
```

Fig.9 Example Prediction of CNN

After train the model of CNN and LBPH, we use the test data to obtain the accuracy of the trained model. At the end, the accuracy of LBPH is nearly 59% and CNN is nearly 90%.

4 Discussion

According to Fig.9, Fig.10 and Fig.11, three different photos of a same person were predicted as three different people. Both CNN model and LBPH model are not able predict images of Guillermo Coria correctly. The reason why this happens is because these images have many different features which cannot be recognized by the models. For example, the mouth and hair are different among three images. The mouth in Fig.10 must be highlighted as another significant feature. To be honest, it is difficult for a real person to distinguish these photos immediately.

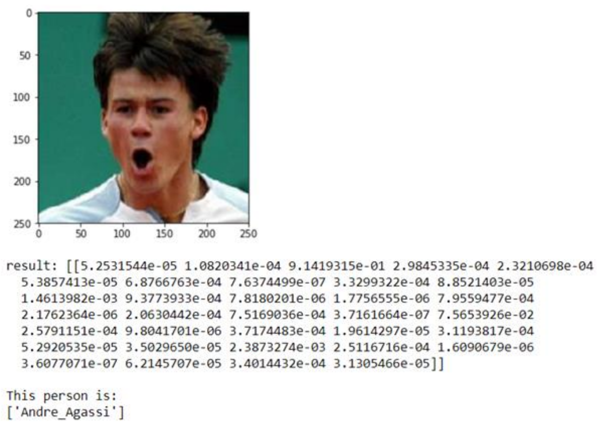


Fig. 10 Prediction: Guillermo Coria

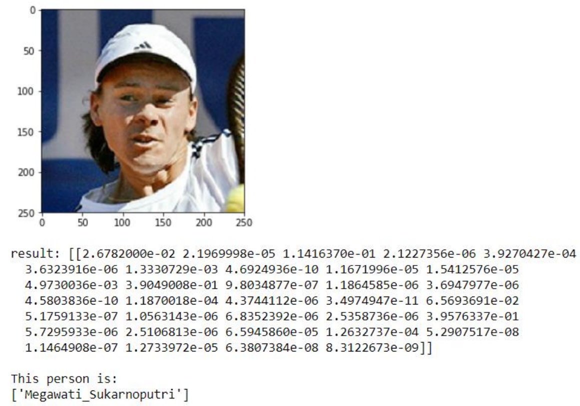


Fig. 11 Prediction: Guillermo Coria

However, after observing them carefully, we can also notice that these images have some similar features, like ears, eyes and nose. These features will help model to recognize them as same person, which are ignored by CNN model. From Fig.5, we can see the visualization of conv2d layers and get the reason why it happens. With running layers steps by steps, features of images are highlighted while losing other features, which includes nose and ears. It is hard to find nose or ears features image in the third conv2d layer.

In order to further improve the accuracy of face recognition and avoid losing features information, we can try to add the LBPH algorithm into the pooling layer. This method could make the original image information lose less, preserve more meaningful features and make prediction result stronger. Ke, Pengfei et al. [8] The accuracy of LBPH combined with CNN can reach 97.65%. On this basis, if combined with BN or other algorithms, the accuracy can even reach 100%. Thus, although the LBPH algorithm has a low accuracy when used alone, it can achieve better results when combined.

Training CNN model cost much more time than LBPH model, and it is worth because CNN model works much better than LBPH model. But we noticed that LBPH model can retain images features better, especially some inconspicuous but important features. CNN model will focus on important features and classify them accurately but ignore other features.

5 What We Are Doing Now

From the research above, we discussed how the CNN and LBPH work on the face recognition and we also mentioned how to implement the CNN model and LBPH algorithm. And we get some hints from the principles of CNN and LBPH that we can use the LBPH algorithm to replace the Max-Pooling layer in the CNN model. Because Max-Pooling only can get the max number in the pooling area and use that to replace the area. But when you put some noise in the graph, the max-pooling would work very bad on it, for it collecting a lot of noise points rather than the main features. And if we use LBPH to replace the max-pooling, we will get a small image that main features remain in it, because of the advantages of LBPH.

Here, in the "keras layer lbph.ipynb" file, you can view our code of LBPH layer. Unfortunately, we have not finished the code, because we meet a problem on extract information of images from the Tensors transformed by the layer. We will try to solve the problem and finish the code by communicating with the professor in the future.

6 References

[1] G. M. Zafaruddin, and H. S. Fadewar, "Face Recognition: A holistic approach review". In Contemporary Computing and Informatics (IC3I), 2014 International Conference on. IEEE, November 2014, pp. 175-178.

[2] Turk, Matthew & Pentland, Alex. (1991). Eigenfaces for Recognition. Journal of cognitive neuroscience. 3. 71-86.

[3] Zhao, Guoying & Pietikäinen, Matti. (2007). Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. IEEE transactions on pattern analysis and machine intelligence. 29. 915-28.

[4] IT. Ojala, M. Pietikainen, D. Harwood. A comparative study of texture measures with classification based on feature distributions [J].Patterncognition,1996,29(1):51-59.

[5] R. Yann, M. Sebastien. Face authentication using Adapted local binary pattern histograms[C]. In: the 9th European Conference on Computer Vision, 2006, 4:321-332

[6] Arel I,Rose D C? Karnowski T P. Deep machine learning-A new frontier in artificial intelligence research [Research Frontier][J]. Computational Intelligence Magazine, IEEE, 2010, 5(4): 13-18

[7] Hou, L., K. Singh, D. Samaras, T. M. Kurc, Y. Gao, R. J. Seidman, and J. H. Saltz. Automatic Histopathology Image Analysis with CNNs. In 2016 New York Scientific Data Summit (NYSDS), 16, 2016. doi:10.1109/NYSDS.2016.7747812.

[8] Ke, Pengfei & Cai, Maoguo & Wang, Hanmo & Chen, Jialong. (2018). A novel face recognition algorithm based on the combination of LBP and CNN. 539-543.