Convolutional Neural Networks (CNNs) for Image Recognition and Detection

Aparna Mohan

Department of Computer Science and Engineering, Jawaharlal College of Engineering and Technology, Mangalam, Lakkidi, Ottappalam, Palakkad, Kerala, India

Abstract **– Nowadays, face recognition is widely uses in many security based applications. Even mobile phones and other such gadgets consider face as one of the most secure biometric application. Deep learning based models are used for face recognition. Deep features are obtained by using several convolutional and pooling layers to extract features from input images.**

Keywords **– Convolutional neural network (CNN), face recognition, LBP, face detection, texture classification**

I. INTRODUCTION

Convolutional neural network is widely used in pattern and image recognition. Because they have number of advantages compared to other techniques. Neurons exchange messages between each other. A neural network is a system of interconnected artificial “neurons”. Each connection have its own numeric weights, tuned during the training process. In each of the network there are multiple layers of feature detecting “neurons”. Each of these layers consists of many neurons that respond to different combination of input. Human face consists of complex multidimensional meaningful visual stimuli. Therefore developing a computational model for human face recognition is a difficult one. Hence convolutional neural networks (CNNs) have been established as a powerfull class of neural network models among all others for image recognition problems.

In today’s era images and videos have become ubiquitous over the internet. This leads to the development of algorithms that analyze the semantic content of various applications. For the better understanding of image content, statement of the art, results on image recognition, segmentation, detection and retrieval the convolutional neural network have been considered as an effective class among other neural network models.

A. Architecture

While training CNN’s the input to the network is a fixed size of 224x224 RGB image. Before training the input data the image is preprocessed, the only preprocessing done for this is subtracting the mean RGB value from each pixel. Then the preprocessed image passed through a stack of convolutional layers along with convolution filters of size 1x1. In ConvNet training spatial pooling is carried out by five max pooling layers. The max pooling layers follow some of the convolutional layers but do not follow all. Max pooling requires 2x2 pixel window to perform training with a stride 2.

The convolutional layers have different depth in different architectures. These layers are followed by three Fully Connected (FC) layers. The first two layers consist of 4096 channels and the third layer contains 1000 channels. The final layer is known as soft-max layer. This Conv layers also consists of some hidden layers, which are equipped with rectification non linearity. No networks in CNN contain Local Response Normalisation (LRN).

B. Training

While training image, let S be the smallest side of an isotropically trained image. From this image the input image is cropped and the crop size is fixed to 224x224. The value of S can take any value not less than 224, where S=224 then the crop will capture the entire image statistics by spanning the smallest side of a training image. When S>>224 the crop will correspond to a small part of the image, containing a small object or an object part. There are mainly two approaches for setting the training scale S. the first step is to fix S, corresponding to single scale training. The image content within the sampled crops represents multiscale image statistics. The second step is done by setting S as multi scale training, where each training image is individually rescaled by random sampling from a certain range [Smin, Smax]. But the objects in the image will be of different size, so it is beneficial to take this into account during training. Training set augmentation can also be seen by scale jittering. This is done where a single model is trained to recognize objects over a wide range of scales.

C. Testing

At the time of testing, the trained ConvNet and the input image is passed to the testing phase. The testing process is classified as follows, the image is isotropically rescaled to a predefined smallest image. This is referred to as the test scale. Then the network is densely fed over the rescaled test image. Then the fully connected layers are first converted to convolutional layers. This results into a fully convolutional net. This fully convolutional net is then applied to the uncropped image. The output will be a class score map that consists of number of channels equal to the number of classes and a variable spatial resolution. This channels will be completely depends up on the input image size. In order to obtain fixed sized vector the class score map is spatially averaged or sum pooled. Here there is no need to sample multiple crops at test time as the fully convolutional network is applied over the whole image. The accuracy of testing can be slightly improved by using large set of crops and also it results in a finer sampling of the input image when compared with fully convolutional net.

Since there are different convolution boundary conditions, the multi crop evaluation is complementary to dense evaluation. In multi crop evaluation applying a ConvNet to a crop results in to feature maps padded with zeros. In the case of dense evaluation the padding of same crop will comes from the neighbouring parts of an image because of convolution and spatial pooling. This will increases the overall network fields and the context will be captured.

D. Local Binary Pattern

A particular case of texture spectrum model and a type of visual descriptor that is used for classification is known as local binary pattern (LBP). An LBP feature vector will divide the examined window into cells. For each pixel in a cell it compares the center pixel value with the 8 neighboring pixel value either clockwise or counter clockwise. If the centre pixel value is greater than the neighboring pixel value then replace it with 0 otherwise replace it with 1. This value is then converted in to decimal value for more convenience. Then compute the histogram over the cell according to the frequency of each number. This histogram is the LBP feature vector of the image. For more convenience normalize the histogram and then concatenate the normalized histograms of all cells. This will give the feature vector for the entire window.

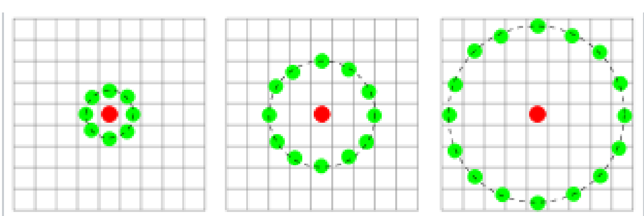

Fig 1 LBP Operator

LBP is one of the methods in texture classification. The LBP codes obtained from the image are collected in the form of a histogram. The classification is then performed by computing and measuring the histogram similarities. The LBP operator denote pixels of an image by using 3x3 neighbourhood. This 3x3 neighbourhood is also known as matrix. In the matrix form rach of the pixel consists a value and this value can be varied depending upon the image and pixel quality.

II. CONCLUSIONS

In convolutional neural network, especially for large scale image classification the representation depth is beneficial for the classification accuracy. The performance of LBP operator along with convolutional neural network for face detection is studied. CNN architectures are capable of learning powerful features and robust to details of the connectivity of the architectures. The method presented in this review article contributes a method for face recognition based on convolutional neural network and local binary pattern by capturing the necessary facial charecteristics.

ACKNOWLEDGMENT

I gratefully acknowledge my guide for the support to complete this work and Wikipedia for the correct information source.

REFERENCES

[1]. Karen Simonyan and Andrew Zisserman (2015). Very deep convolutional networks for large-scale image recognition: Visual Geometry Group, Department of Engineering Science, University of Oxford.

[2]. Timo Ahonen, Abdenour Hadid, and Matti Pietikainen (2004). Face Recognition with Local Binary Patterns: Machine Vision Group, Infotech Oulu.

[3]. Snehal Humne and Prachi Sorte (2018). A Review on Face Recognition using Local Binary Pattern Algorithm: M.E. Dept. of I.T., R.M.D. School of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India Asst. Professor, Dept. of I.T., R.M.D. School of Engineering, Savitribai Phule Pune University, Pune, Maharashtra,India.

[4]. Rui Shao, Xiangyuan Lan, Pong C. Yuen (2017). Deep Convolutional Dynamic Texture Learning with Adaptive Channel-discriminability for 3D Mask Face Anti-spoofing: Department of Computer Science, Hong Kong Baptist University.

[5]. Abhilasha A Patil, Lakshmi Maka, Abhilasha A Patil and Lakshmi Maka (2015). User Recognition Based on Face using Local Binary Pattern (LBP) with Artificial Neural Network (ANN) : International Journal of Ethics in Engineering & Management Education.