International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

A comparative study on Convolutional Neural Network and Viola-Jones Algrithm for faster Face Recognition

Rubeena M M[1], Dr.Manish T I[2]

[1] M.Tech scholar, Dept. Of Computer Science and Engineering, Adi Shankara Institute of Engineering and Technology, Kalady, Kerala, India
[2]Professor, Dept. Of Computer Science and Engineering, Adi Shankara Institute of Engineering and Technology, Kalady, Kerala, India

---

Abstract - *Facial recognition system is a technique to verify or identify a face from digital image or video. This research field has wide applications from the need of Automatic recognition's, Surveillance systems, Design of human-computer interface etc. These researches implicate knowledge and researchers from areas such as Pattern recognition, Neuroscience, Computer vision, machine learning, Psychology, and Image processing, etc. There are various computer algorithms that are relevant in the discipline of face recognition, but this paper focus on two of the most popular methods: Convolutional Neural Networks and the Viola-Jones algorithm. In this paper, we'll go through general ideas and structures of face recognition, important issues and finally give a comparison of algorithms.*

Key Words: Face Recognition, Viola-Jones, Convolutional Neural Network (CNN).

1. INTRODUCTION

Face detection is a commodity that is prevalent in our lives but also something many technology users have probably never thought about in depth. Face recognition is the application of computer technology to identify faces in digital images. In face recognition using images, a picture given or taken from a digital camera, we would desire if there is any person inside the picture, where his/her face locates at, and who he/she is. The face recognition method is generally separated into three stages: Face Detection, Feature Extraction and Face Recognition.

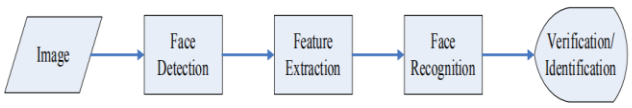

Fig -1: Face Recognition Basic Architecture.

1.1 Face Detection

The foremost goal of this step is to decide whether or not human faces seem with inside the given image, and to discover those faces. The expected outputs of this step have tracts consists of each face in the input image. For making the face recognition system further more robust and easy to design face, besides serving because the pre-processing for face recognition, face detection will be used for region-of-hobby detection, re-targeting, video and image.

1.2 Feature Extraction

The human-face patches are extracted after the face is detected from images. There are some disadvantages for using these patches directly, first, each patch contains over 1000 pixels, which are too enormous to build a strong recognition system. Second, the face patches may taken from different camera alignments, in distinct illuminations, with various facial expressions or pose and may cause occlusion and clutter of face. To overcome these, feature extraction is done for information packing, dimension reduction, salience extraction, and noise cleaning. Then a face patch is transformed into a vector with fixed dimension or a set of reliable points and their corresponding locations.

1.3 Face Recognition

The last step is to recognize the identities of the faces that is detected. For automatic recognition, a face database is required to build. There will be several images of each person in the database, their features are extracted and stored in the database. When an input image is introduced, firstly the face is detected then features are extracted and compare these features with the features of faces stored in the database. The two applications of face recognition are identification and verification. Face identification termed as from the given face image, the system tell who he / she is or the most credible identification. And face verification termed as from the given a face image and a guess of the identification, the system to tell true or false about the guess.

Face detection algorithms are classified into four categories: Knowledge-Based, Feature Invariant, Template Matching, and the Appearance-Based method. Two of the most widely incorporated face detection methods at the moment are the Viola-Jones algorithm and Convolutional Neural Networks.

2. VIOLA-JONES ALGORITHM

Viola-Jones method is one of the Appearance-based method. Fast processing speed and high accuracy of face detection are important factors in face recognition which is accomplished well using this algorithm. In this algorithm, when an image is introduced, then the algorithm compares the already defined eyes distance with the image given. Then with the matching eyes distance and pupil distance, the eyes in the image will be detected.

The way that the Viola-Jones algorithm actually works is through the execution of four main steps: Haar-like Features, Integral Image, Adaboost Training, and Attentional cascade. In Haar-like features step, the image break up into multiple sets of two adjacent rectangles located at any scale and position within an image by using a set of rectangular digital image features. The Viola Jones algorithm then classifies the image using three types of features, square features, three-square features, and four-square features. The value of these features was the difference between black and white regions. These rectangles are then applied to the image that has been opened within the program. In each sub-window image, the total number of Feature Haar was very large, much larger if compared to the number of pixels. For faster classification, the learning process should eliminate the majority of features available, and focus on a small set of necessary features. When the Haar-like features are applied to the image, there is first one main region that is being examined, the area from the forehead to the eyes. In that region of space, the region of the eyes is darker than the region of the nose and cheeks. so if the Haar-like features can be matched then the image can go to the next step with the Adaboost training.

The Adaboost training define the region from the eyes to the cheeks. That is by employing a learning algorithm that is used to teach the program to look over a set of possible areas and then choose the areas with the best features resulting in a reduced image with more defined regions. After the features are selected they are put through the Adaboost learning algorithm to narrow down the number of features that are looked at and then passed on to the cascading stage. The Adaboost training makes sure that the region that is being examined is as precise as possible in order to help obtain the best accuracy. AdaBoost steered to form face templates.

The last step is the Attentional cascade. During this step, the program is trying to maintain the smallest error percentage rate as possible. In order to do this, the attentional cascade's main focus is to eliminate as many false positives as possible. The Viola-Jones method is the combination of four general keys: Haar like Feature, Integral Image, Adaboost learning and Cascade classifier. Haar like Feature is the contrast of the number of pixels from the area inside the rectangle. Haar like Feature values are obtained from the difference in the number of dark area pixel values by the number of bright area pixels:

$$F(Haar) = \sum F_{white} - \sum F_{black}$$

Fig -2: Haar-like Feature Equation.

Integral image was a technique for calculating the feature value quickly by changing the value of each pixel into a new image representation. The calculation of the value of a feature could be done quickly by computing the integral image value at four points as shown in Figure 3. If the integral value of image of point 1 was A, point 2 was A + B, point 3 was A + C, and at the point 4 was A + B + C + D, then the number of pixels in region D could be known by 4 + 1 (2 + 3).

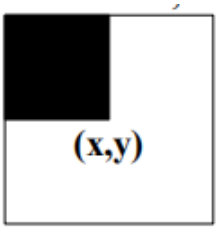

Fig -3: Integral Image (x,y).

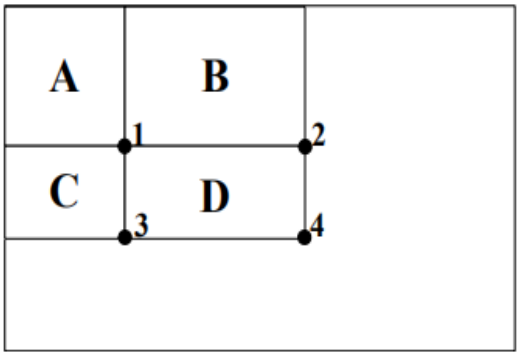

Fig -4: Integral Image Score Count.

The Adaboost learning algorithm, used to improve classification performance with simple learning to combine many weak classifier into one powerful classifier. The Cascade classifier was a method to combine complex classifier in a multilevel structure that could increase the speed of object detection by focusing on only possible imagery areas. The pattern of filters on the cascade was resolved by the weight given by AdaBoost algorithm. The largest weighted filter was placed in the first process, aiming to erase the non-face image area as quickly as possible. The last stage was showed the object of the sample image that has been detected face or not face.

Decreasing the number of possible faces not only helps decrease false positives, it also helps to increase the number of correct detections. This approach was used to set up a face detection system which is relatively 15 faster than any previous approach. Viola-Jones algorithm with definite threshold value provide the result with fast detection rate, high accuracy and the average detection rate is 97.41\%. Thus this method was the fast and

accurate method in image processing. Face Recognition was considered to reduce the shortcomings of the fingerprint system and help users to perform faster attendance. The Viola-Jones method, despite it has a high accuracy rate, does demand longer computation time. In contrast, CNNs have addressed computation time concerns and have successfully improved upon them.

3. CONVOLUTIONAL NEURAL NETWORK

Conceptually Convolutional neural networks (CNNs) work in an akin way to the Viola-Jones method. In this method Face recognition system is achieved using Deep Learning's sub-field that is Convolutional Neural Network (CNN). CNN is a multi-layer network trained to execute a precise task using classification. In CNN architecture there are mainly four types of layers Convolutional Layer, ReLu, Pooling Layer, and Fully-Connected Layer.

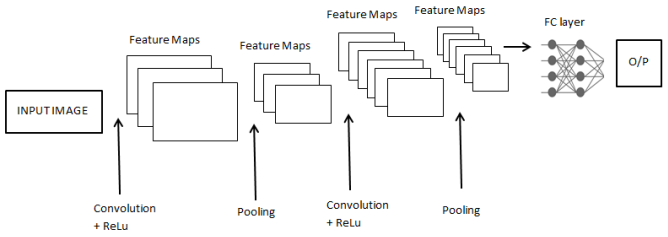

Fig -5: Convolutional Neural Network Architecture.

Convolutional layer is where most of the computational heavy lift. The convolution step is to train image to recognize faces, when a new image is introduced into the system the convolution step uses its previous knowledge to test where the faces might be. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel, to form an output matrix called feature map. The computation is as follows, multiplying each pixel in the feature by the value in the corresponding pixel in the image. The answers are then divided by the total number of pixels in the image. Every matching pixel returns a value of 1 so whatever else is -1. In CNN the image volume is transformed into an output volume. There are three hyper-parameters that restrain the size of the output volume: depth, stride and zero padding. The depth of output volume is the number of filters that we are using which corresponds to the depth or dimension of input image. Stride represents the number of pixels that shifts over in the input matrix. When the stride is 1 then the filters move one pixel at a time and when the stride is 2 then the filters jump 2 pixels at a time . Thus yield smaller output volumes spatially. Zero padding allows to regulate the spatial size of the output volumes. That is, the picture is padded with zeros, so that it fits or drop the part of the image where the filter did not fit. This process is called as valid padding, which keeps only credible part of the image.

After the convolution step completes once, it must repeat itself multiple times until it has the desirable number of possible locations narrowed down. Then the pooling phase reduces the image down enough to keep all feasible features but still narrow the window of viable locations of the face. There are three types of pooling Max Pooling, Average Pooling and Sum Pooling. Max pooling extract the largest element from the revised feature map. The average pooling also act same as max pooling, by taking the largest element. Sum pooling takes the sum of all the elements in the feature map. The ReLU stands for Rectified Linear Unit for a non-linear operation which gives an output as $f(x) = \max(0,x)$. Thus all negative features from the image were changed to zeros which is mostly a step to keep the CNN mathematically sane by keeping the CNN from approaching infinity. Fully connected layer decides whether or not the final selection of features is correct. So the matrix is flattened into vector and lead into a fully connected layer like a neural network. For every value that is output by the previous steps it gets a vote as to whether or not the value matches up with the image value.

CNN can involuntarily gain features to capture complex visual variations by holding a enormous amount of training data and its testing phase can be simply correlate on GPU cores for acceleration. The accuracy achieved by using CNN is higher than other methods because of less error. The accuracy of CNN output can be increased by using more different images and performing more iterations while training the model.

4. COMPARISON

While there exist similarities between the Viola-Jones algorithm and Convolutional Neural Networks, they do both differ in a few areas. The Viola-Jones algorithm is not capable to detect faces in various positions. If the face is not presented in a front-facing position with proper lighting, the accuracy of the results drops dramatically. While both methods work in a series of steps, the steps in the Viola-Jones method are set whereas the steps in CNNs are much less structured. The Viola-Jones algorithm encounter problems in positions range like side view and for low lighting because their Haar-like features do not map very well to varying positions. By contrast, CNNs have the ability to detect faces in different positions including side views and different lighting scenarios, and therefore CNNs are much more diverse in how they correctly handle their input. Convolutional Neural Networks require multiple pass-throughs and therefore store large amounts of information, CNNs require much more space locally than the Viola-Jones algorithm.

5. DISCUSS

As previously stated, the use of CNNs requires a larger storage capacity to run than does the Viola-Jones algorithm. Because of this, the places where CNNs can be implemented correctly become limited. Therefore, while CNNs are faster and much more reliable in terms of accuracy, the Viola-Jones algorithm is still widely used today. The large memory of CNN can be solved using R-CNN, that is Region-based-CNN. Lets say if CNN have feature dimension of 4096, then if we use R-CNN then the feature dimension will be 2000. So computation using R-CNN is more efficient than CNN, since less memory is used. The main disadvantage of R-CNN is that, we can't train the whole image in one go. That is, need to train every part of the image independently. Hence the CNN algorithm is more reliable for Face recognition system. Even-though CNN algorithm alone shows better accuracy, yet Combination of CNN algorithm with other techniques shows more promising results.

6. CONCLUSION

When looking at the Viola-Jones algorithm and Convolutional Neural Networks for face detection, it is difficult to say which one is best overall. Both methods have strengths and weaknesses when it comes to certain areas of face detection. While CNNs are much faster, they do require more memory space and are therefore are more expensive to implement. The Viola Jones algorithm is preceeding algorithm yet least memory requirement of it is implemented much more likely. The CNN can be used as a merger with other efficient algorithms can provide a combo-algorithm which will enhance the technology.

REFERENCES

[1] Paul Viola, Michael Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137–154, 2004.

[2] Rudolfo Rizki Damanik, Delima Sitanggang, Hendra Pasaribu, Hendrik Siagian, Frisman Gulo, "An Application Of Viola Jones Method For Face Recognition For Absence Process Efficiency", IOP Conf. Series: Journal of Physics: Conf. Series 1007 (2018) 012013, doi :10.1088/1742-6596/1007/1/012013.

[3] Danai Triantafyllidou, Anastasios Tefas, "Face Detection Based On Deep Convolutional Neural Networks Exploiting Incremental Facial Part Learning", 23rd International Conference on Pattern Recognition (ICPR), December 4-8, 2016.

[4] Mehul K Dabhi, Bhavna K Pancholi, "Face Detection System Based on Viola - Jones Algorithm", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Volume 5 Issue 4, April 2016.

[5] Wei-Lun Chao, "Face Recognition", GICE, National Taiwan University, 2007.

[6] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, Gang Hua, "A Convolutional Neural Network Cascade for Face Detection", June 23, 2015.

[7] Asifullah Khan, Anabia Sohail, Umme Zahoora, Aqsa Saeed Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks", Artificial Intelligence Review, 2019, DOI: https://doi.org/10.1007/s10462-020-09825-6.

[8] Daniel Saez Trigueros, Li Meng, Margaret Hartnett, "Face Recognition: From Traditional to Deep Learning Methods", arXiv:1811.00116v1 [cs.CV] , 31 Oct 2018.

[9] Patrik Kamencay, Miroslav Benco, Tomas Mizdos, Roman Radil, "A New Method for Face Recognition Using Convolutional Neural Network", Digital Image Processing And Computer Graphics, Volume: 15, Number: 4 ,2017.

[10] Monali Nitin Chaudhari, Gayatri Ramrakhiani, Mrinal Deshmukh, Rakshita Parvatikar, "Face Detection using Viola Jones Algorithm and Neural Networks", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.

[11] Atiqur Rahman Ahad, Tanmoy Paul, Ummul Afia Shammi, Syoji Kobashi, "A Study on Face Detection Using Viola-Jones Algorithm for Various Backgrounds, Angels and Distances", Biomedical Soft Computing and Human Sciences, Vol.233, No.1, pp.27-366, May -2018.

[12] Vikram K, S.Padmavathi, "Facial Parts Detection Using Viola Jones Algorithm", International Conference on Advanced Computing and Communication Systems (ICACCS -2015), Jan. 06 – 07, 2017.