

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 8, August 2018

A Survey on Segregation of Mixture of Images into Different Folders using YOLO and CNN

Riddhi Zunjarrao ¹

B.E Student, Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India¹

ABSTRACT: We present our system to sort out the images into different folders according to the face of the person found out in the input image. Face detection is an unavoidable step in the system where face is to recognize. This paper proposes detection of face with the help of You Look Only Once (YOLO) and face recognition by using Convolutional Neural Networks(CNN.) Faces are detected in one regression problem with YOLO, so it is computationally faster and cheaper. Once the faces are detected, they are processed through a face recognition algorithm. Convolutional Neural Networks are used to correctly classify the face into the germane folders. For the classification, the faces detected through YOLO are passed through various convolutional layer to correctly recognize face by using various feature maps.

KEYWORDS: Convolutional Neural Network, You Look Only Once, Face recognition.

I. INTRODUCTION

Advancement in technology has been taking place in a very rapid manner, the department of human computer interaction as well advanced quite a lot. Classify images have also been quite trending in recent years. Also facial recognition has been booming field of research. Facial recognition is recognizing the face of some germane person according to given trained model, by taking in consideration some attributes of facial features. Facial recognition has been widely used in various fields, for security reasons, for tagging people on Social media. Recently Facebook has also developed a feature where we would be able to obtain information about where our photo is used even if we haven't been tagged in the same. The very same advancement would be very useful for us.

However, our system deals with mixture of photos given to the model and then by considering the training obtained by the model, it will properly classify the images into various folders. This model will be helpful for segregating the tons of images that our to be classified according to the person given in that photo. This model on the small scale could be used for separating images in the folder which contains all the family pictures, or say suppose for partitioning the photos at a wedding function.

The proposed system works in two stages 1) Face detection 2) Face recognition

Firstly, the input image will be worked upon with a face detection algorithm, the detection algorithm will put a bounding box around the objects detected as people. The object detection algorithm, You Look Only Once (YOLO) [3] will be first used to detect the people in given input image. In the past years, many face detection algorithms [2] have been used, it could be either feature based or image based. Feature based algorithms detect face using facial features like edges, motion, point distribution models and many other. And the image based algorithm for face detection involve neural network, statistical approach. However YOLO out performs all of them. In terms of real-time performance, with the support of a GPU, YOLO-PC re-trains a deep convolutional neural network to detect face at more than 40 fps (frames per second)[9].

Later, when the face is correctly detected in the input image, it will be passed through a facial recognition algorithm. The facial recognition algorithm will then be able to classify the image into the respective person's name. Here we will use Convolutional neural network (CNN) algorithm [5]. CNNs use a variation of multilayer perceptrons designed to require minimal pre-processing. Relatively less pre-processing is required by the CNN as compared to the various alternative image classification algorithms.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 8, August 2018

II. PROPOSED METHODOLOGY



The system proposed by this paper will work in three phases:

- 1) Face Detection using You Look Only Once(YOLO)
- 2) Face recognition using Convolutional Neural Network(CNN)
- 3) Classification using Softmax

2.1 YOLO Algorithm

The YOLO[1] looks upon the input image as a single regression problem. Straight from image pixels to the bounding box is viewed as single problem. YOLO scans the image in one go.

1. YOLO divides the image into $S \times S$ grids.
2. The bounding boxes are responsible for predicting upto 5 bounding boxes.
3. For each bounding box the class is predicted along with its prediction score.
4. The number of the bounding boxes we get will be $13 \times 13 = 169$, and each grid producing upto 5 bounding boxes. Therefore a total of $169 \times 5 = 845$ bounding boxes in total
5. Most of the bounding boxes will have very low confidence score, and in the end only the bounding boxes crossing a certain threshold value would be included. For face detection threshold value of 30 sounds good enough.
6. Each bounding in Yolo predicts 5 parameters : x, y, w, h and confidence.
The (x, y) coordinates represent the center of the box relative to the bounds of the grid cell. The width and height are predicted relative to the whole image. Finally the confidence prediction represents the IOU between the predicted box and any ground truth box
7. YOLO algorithm has a convolutional network used : convolutional 3×3 kernel and a max pooling of 2×2 kernel.

2.2 Convolutional Neural Network

Convolutional neural networks [5] were introduced to deal with all loopholes of simple neural networks. CNNs are feed-forward networks, they have the ability of extracting topological properties from the unprocessed input image without any pre-processing needed. Therefore, these networks integrate feature selection into the training process [3]. Furthermore, CNNs are able to recognize patterns with regard to robustness to distortions, extreme variability. Also simple geometric transformations like translation, scaling, squeezing, rotation, noise and stroke width [6, 7]. CNN has a structure which includes convolutional layer, max pooling layer, flattening and fully connected layer.

2.2.1 Convolutional layer

Convolutional layer is the layer which applies a convolutional operation to the input image. Basically this where activation function is performed, between the input image and the feature detector. The feature detector is a 3×3 matrix, the convolutional layer compresses the original input image. The value of the stride shouldn't be very small or very large. Normally for the facial recognition stride of 1 works fair enough. No doubt certain information is lost but the important information is retained. We have many filters maps as we

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 8, August 2018

have classify the input image on basis of various features. The convolutional layers are lined up one after the one.

2.2.2 ReLu

The function of the ReLu layer is to increase the non-linearity in our model. ReLu is a rectifier and it includes units which employ rectification.

The ReLu function is defined in mathematical form as

$$f(x) = \max(0, x)$$

where x is the input to a neuron

To reduce the the linearity all the negative values are replaced by 0. This breaks the linearity of the input image.

2.2.3 Max Pooling

Pooling layer is used to reduce the dimensionality measure of activation mapping. The pooling layer reduces the spatial dimensions. The depth isn't lost due to pooling in CNN. In addition the computation power increases as well. Max-pooling simply outputs the maximum value within a rectangular neighbourhood of the activation map. The position of this layer is generally between the convolutional layers.

2.2.4 Flattening

Flattening layer is the one of the most simple layers in CNN. The output, that is, the pooled feature map which is stored as an 2-D array is converted to a 1-D array. The purpose of flattening is so that it is suitable to feed the further data for classification into the neural network.

2.2.5 Fully connected layer

The final layer, Fully Connected Layer (FCL) term refers to that every filter in the previous layer is connected to every filter in the next layer. The main purpose is to combine the features into more attributes that predict the classes in a better way. The input to this layer is the flattened layer and gives output by specifying the class the image belongs to.

2.3 Softmax

To make sum of the probability of the different classes getting predicted equal to 1, we introduce the softmax layer.

Softmax layer is mathematically defined as :

$$P(y = j | \mathbf{x}) = \frac{e^{\mathbf{x}^T \mathbf{w}_j}}{\sum_{k=1}^K e^{\mathbf{x}^T \mathbf{w}_k}}$$

2.4 Algorithm used :

1. Detect the faces in an image of S*S grid.
2. Put a bounding box around the faces detected.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 8, August 2018

3. Convert all the images to 16 X 16 X 1, 32 X 32 X 1, 64 X 64 X 1.
4. Add the convolutional layers to classify between the predefined classes.
5. Later, to correctly classify the faces into appropriate folder using Softmax function.

III. CONCLUSION

This paper represents face recognition using YOLO and CNN Architecture, for face detection YOLO is used and image classification CNN is used. Now, the mixture of input images can be classified into the different folders according to the name of the person. The image whose final value doesn't reach the specified threshold value are put in a anonymous folder, which contains all the unclassified images. The minimum threshold value for classification could be adjusted according number of training images available, it again depends on the number of folders. So the burden to categorize the mixture of input image is reduced.

REFERENCES

- [1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788
- [2] Erik Hjelm, Boon Kee Low, "Face Detection: A Survey", Computer Vision and Image Understanding 83, 236-274 (2001)
- [3] Yicheng An, Jiafu Wu, Chang Yue, "CNNs for Face Detection and Recognition", Stanford Report
- [4] Ion Marques, Manuel Grana, "Face Recognition Algorithms", Proyecto Fin de Carrera June 16, 2010
- [5] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D, "Face Recognition: A Convolutional Neural-Network Approach", IEEE Transactions On Neural Networks, Vol. 8, No. 1, January 1997
- [6] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, Gang Hua, "A Convolutional Neural Network Cascade for Face Detection", IEEE proceedings
- [7] Masakazu Matsugu, Katsuhiko Mori, Yusuke Mitari, Yuji Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network", Neural Networks 16 (2003) 555-559, 2003 Special issue
- [8] M. Coşkun, A. Uçar, Ö. Yildirim and Y. Demir, "Face recognition based on convolutional neural network," 2017 International Conference on Modern Electrical and Energy Systems (MEES), Kremenchuk, 2017, pp. 376-379.
- [9] P. Ren, W. Fang and S. Djahel, "A novel YOLO-Based real-time people counting approach," 2017 International Smart Cities Conference (ISC2), Wuxi, 2017, pp. 1-2.
- [10] W Zhao, R. Chellappa, P. J. Philips, A. Rosenfeld "Face recognition: A literature survey", Journal ACM Computing Surveys Volume 35 Issue 4, Dec 2003
- [11] Benjamin J. Balas, Pawan Sinha, "Region-based representations for face recognition", October 2006 ACM Transactions on Applied Perception (TAP): Volume 3 Issue 4, October 2006
- [12] Mankar, Vijay & G Bhele, Sujata. (2012). A Review Paper on Face Recognition Techniques. International Journal of Advanced Research in Computer Engineering & Technology. 1. 339-346.