International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 07 | July 2018 www.irjet.net p-ISSN: 2395-0072

A Deep Learning Method for Identifying Disguised Faces Using AlexNet and Multiclass SVM

Abhila A.G[1], Sreeletha S.H [2]

[1]Research Scholar, Dept of Computer Science and Engineering, LBSITW, Kerala, India
[2]Associate Professor, Dept of Computer Science and Engineering, LBSITW, Kerala, India

---

Abstract - *Face recognition is an important biometric that is used to uniquely identify an individual. However, the presence of various disguises in the face will diminish the performance of any facial recognition system. Disguised face recognition problem becomes extremely challenging when important facial features get covered up. Because of the increase in popularity and success of deep learning in computer vision problems, features learned by Convolutional Neural Networks (CNN) can be used for face recognition. In this paper, a novel method for identifying and recognizing people who hide their identity by covering face using scarves has been proposed. Recognition is done by using a pre-trained Convolutional Neural Network (Alex-Net Model) for facial feature extraction which is followed by a multi-class Support Vector Machine (SVM) algorithm for performing the classification task. A new dataset of covered faces has also been introduced for training the deep network. This indeed is very useful for law enforcement and other organizations as it would help to identify criminals, Protestants or anyone who hides their identity.*

***KeyWords*: Face Recognition, Deep Learning, Convolutional Neural Network, Support Vector Machine, Transfer Learning**

1. INTRODUCTION

A face is always considered as an important biometric that uniquely identifies an individual. Face recognition technique identifies an individual by comparing the input image to the stored record of images. Such systems usually analyze the characteristics of individuals face and based on the analysis it will recognize the person. It is based on the fact that different individuals have different facial features which makes them unique. In comparison to other existing biometric systems like fingerprint recognition, handwriting recognition etc., face biometric can attain a higher performance in security systems [1].

Numerous algorithms and techniques have been developed for improving the performance of face recognition. These systems are usually susceptible to a number of challenges including illumination, image quality, expression, pose, aging, disguise etc. Among these challenges, recognition of faces with disguise is a major challenge and has only been recently addressed by few researchers.

Disguise is an important and crucial face recognition challenge. There can be both intentional or unintentional changes through which a person become very difficult to be identified. Intentional disguises are those in which a person knowingly changes his identity by wearing a wig, changing hairstyle, wearing glasses, etc [2]. Nowadays it has become very difficult for the cops to identify protestants and criminals as they often find it easier to hide from the law by disguising themselves. The existing face recognition systems are often found to be a failure in case of identifying people who cover their face using scarves with only the eye portion visible. This is because the recognition system could not perform well as the lower facial portion get covered up.

Recently Deep learning has been highly explored in the field of computer vision. Many face recognition systems based on deep learning has been developed and it has been seen that they outperformed all the existing systems. Deep learning methods make use of deep neural networks such as Convolutional Neural Networks (CNN)[3].

Since training a deep model require large dataset and huge computing power, designers often perform transfer learning. Transfer learning[4] is the method by which a model developed for one task can be reused for a different related task. Since the pre-trained network has already learned a number of features, it can be used for a new classification task by just fine-tuning the network. Another advantage is that the number of images required for training and the training time is reduced.

In this paper, we propose a face recognition system for recognizing people whose faces are covered using scarves. Here, a pre-trained CNN AlexNet[3] model is used for extracting features from facial images. The extracted features are then classified by training a multiclass SVM [21].

The rest of this paper is organized as follows: Section 2 gives a brief description on the related works. Section 3 introduces the proposed disguised face recognition system. The experiment, results and analysis are presented in section 4 and 5 respectively. Finally, the Conclusion is given in section 6.

2. RELATED WORKS

Disguised face recognition is a challenging and difficult task. The changes in the face caused by wearing a wig, eyeglasses or even growing a beard can seriously affect the performance of any facial recognition system [2].

Patterson and Baddeley [5] discovered that the ability of facial recognition system decreased when the face get disguised by changing either hairstyle or beard. The performance is much less compared to normal faces' recognition. They also suggested that changes in both beard and hairstyle further reduced the performance.

Ramanathan et al. [6] studied similarity in faces by constructing two eigenspaces corresponding to left and right half of the face respectively. It is an appearance-based algorithm. The algorithm has been successfully tested on the AR face database [7]. It includes faces with slight variations and obtained an accuracy of around 39%.

PCA based algorithm with Mahalanobis angle as the distance metric has been used by Alexander and Smith [8]. The algorithm showed an accuracy of around 45.8% on the AR database [7]. But the performance of this method degrades when the mouth and eye region got covered.

Geometrical feature-based recognition algorithm [9] works on the basis of the distance between geometric features. The algorithm extract features from faces such as eyes, mouth, nose, and ears and computes information of their shapes. Further, the algorithm matches this extracted shape information using Euclidean distance measure to compare two images. This method will work only when all the facial features are clearly visible. When these features are occluded by wearing eyeglasses, beard, and scarf, performance decreased dramatically.

The study made by Righi et al. [10] analyzed the recognition performance by adding or removing things on faces such as wigs and eyeglasses. He concluded that increase in facial disguises decreased the performance of face recognition system. He also suggested that changes in stable facial features greatly affected the recognition system especially when the eye region get altered. Toseeb et al. [11] observed the effect of hair and concluded that no remarkable performance change when participants are shown faces with and without hair.

In another work that compared different face recognition algorithms, including state-of-the-art algorithms and algorithms that are designed to handle disguises. They concluded that performance of appearance and feature-based algorithms degraded when important facial parts are disguised. Also, texture based algorithms will not work well for multiple disguises [12]. These studies suggest that there is a need for a much better facial recognition system to handle disguises particularly when the face is covered using scarves.

Meanwhile, researchers in the machine learning community have designed deep learning models that are capable of extracting important features from the images to perform various computer vision tasks. Early models such as Boltzmann Machines [13], Deep Belief Networks [14] and Stacked Autoencoders [15] produced promising results on datasets of small size. During the ILSVRC competition [16] that involved the task of classifying an image into one of the thousand categories, a Convolutional Neural Network (CNN) named AlexNet[3] produced remarkable classification result. Since then so many deep learning models have come across that has outperformed most of the traditional classification methods. These pre-trained models can be used for related tasks by making use of the transfer learning concept. Deep learning models consist of a number of convolutional layers which extract prominent features from an image. Due to the remarkable performance of deep learning models in the classification task, face recognition using convolutional neural networks started attracting the research group. Several algorithms such as DeepFace [17], DeepID [18], FaceNet [19] and are successful examples of face recognition system using deep learning. The increased face recognition rate using CNN have made researchers to think about adapting CNN to solve the disguised face recognition challenge.

3. PROPOSED SYSTEM

The proposed system uses the Convolutional Neural Network model for extracting important features from the facial images and use those features to train a linear classifier. The development of the proposed disguised face recognition system involves four modules: Preparing the dataset, disguised face detection, feature extraction and classification. Figure 1 shows the structure of the entire system.

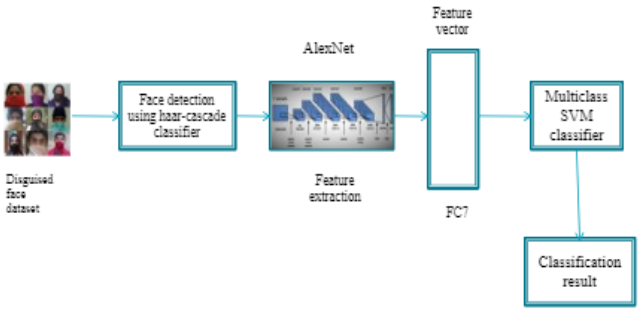

Fig-1: Proposed Architecture

3.1 Preparing the Dataset

Due to the lack of availability of face images covered with scarves, a new dataset has been prepared. It contains 180 images of 3 individuals. For each individual, there are 60 images wearing different types of scarves in different

ways. All images are frontal face with eye region clearly visible. A sample of the dataset is shown in the figure 2:

Fig-2: Sample images in the dataset

3.2 Disguised face detection

Face detection is a essential for face recognition performing, since, at that stage, the faces will be delimited and the essential face features will be extracted. Here, Viola Jones method is used for face detection. It works by sliding a window of size 24*24 pixels. The algorithm works by finding certain haar-like features. These features are used to determine whether there is a face in it or not [20].

.3 Feature extraction

Feature extraction is the process by which important features will be extracted from face for performing the classification task. Here, we are extracting learned features from AlexNet[3] which is a pretrained Convolutional Neural Network. These features will be later used to train a linear classifier. AlexNet is one of the most popular deep networks that is used for various computer vision applications. It has been trained on the ImageNet dataset to classify 1.2 million training images into 1000 different object categories. Figure 3 shows the architecture of the AlexNet model.

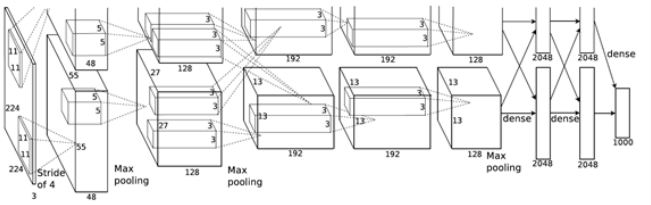

Fig-3: AlexNet Architecture [3]

Alexnet consists of 5 convolutional layers followed by 3 fully connected layers. These convolutional layers extract important features from the image. Each convolutional layer is composed of linear convolution filters which are followed by ReLu activation, normalization and max pooling. The first layer is the input layer which takes images of size 227-by-227-by-3. The first convolution layer has 96 filters each of size 11x11x3 with stride 4 and no padding. The output from the first convolutional layer passes to the ReLu layer which is followed by the max pooling layer. The reason behind the use of ReLu activation function is to prevent the propagation of any negative value in the network. The function of pooling layer is to reduce computation and to control overfitting. The second convolutional layer has 256 filters of size 5x5 with stride 1 and padding 2. The third, fourth and fifth convolution layer performs 3x3 convolution with stride 1 and padding 1. Only convolutional layers 1,2 and 5 contain max-pooling. The convolutional and down-sampling layers are followed by 3 fully connected layers. The last fully connected layer combines features learned from the previous layer to perform the classification task. This layer is followed by a SoftMax layer which will normalize the output.

3.4 Classification Stage

Classification is the process of determining in which class/label a particular data belongs to. So, after extraction of image features using Alexnet, a classifier is required to decide the corresponding label for every test images. For this, a multiclass Support Vector Machine has been used. The idea behind an SVM is to find a hyperplane which is optimal and maximizes training data margin. Binary classification is the simplest one in which an image is classified into one of the two classes. Multiclass classification is comparatively harder than binary classification because the classifier has to learn to find a number of hyperplanes. There are different approaches to multiclass classification problem. Here, we are using the one-against-the-rest method. In order to solve a k-class problem, k-binary models are constructed. An optimal hyperplane will be built that separates one class from the remaining k-1 classes. The performance of the classifier is measured in terms of total classification error or classification accuracy over a set of testing data. Accuracy is 1 if the class label is correctly predicted otherwise it is 0 [21].

The following algorithm lists all the required steps for the proposed system:

Algorithm:

//Input: The input images

//Output: The recognition accuracy

- Detect the face covered using scarf from the prepared dataset using Viola-Jones detection technique.
- Crop the detected face region and store it along with its label.

- Load the input images and its labels.
- Split each category into the similar number of images.
- Load pretrained AlexNet model.
- Pre-process images according to the dimension required for the input layer of AlexNet.
- Split the sets of the images into training and testing data.
- Extract Features from the deeper layers of Alex-Net model.
- Get training labels from the training set
- Train a multiclass SVM classifier using the extracted training features.
- Similarly, extract features from the test set also using AlexNet
- Use the newly trained classifier to predict the labels for the test set
- Get the known labels for the test set
- Display the mean classification accuracy

4. EXPERIMENT

The covered faces after detection using Viola-Jones detection method is loaded along with its labels. Split the dataset into training and testing data. For each person, 70% of the image is used for training and remaining 30% is used for testing. Load the pretrained AlexNet model. Examine the input dimension of the first layer of AlexNet and preprocess the input images according to it. Here the required input dimension is 227-by-227-by-3. If there is any grayscale image it must be converted to RGB since AlexNet can process only RGB images. As already discussed, AlexNet has 5 convolutional layers followed by 3 fully connected layers. The first convolutional layer of the AlexNet learns primitive features from the images such as edges and blobs. The deeper layers of the network combine these low-level features to form higher level image features. The features from the fully connected layer is usually used for various computer vision tasks.

Training features can be easily extracted from AlexNet model. The layer selected for feature extraction depends on our requirement. Here we are extracting features from the second fully connected layer because the disguised face recognition requires deep features. Train the AlexNet on the prepared dataset and extract the trained feature from the selected fully connected layer. The features thus extracted from the CNN is used to train a multiclass SVM. The Stochastic Gradient Descent [22] is used as the optimization algorithm for training the multiclass SVM in which the weight, w is updated on T epochs with a learning rate η for each training data sample chosen at random such that it minimizes the hinge loss and maximizes the margin. A margin is double the distance between a hyperplane and the closest data point.

After training the multiclass SVM, extract test image features in the same way as it is done for the training dataset. Now pass the test features to the newly trained SVM model for performing the classification task. The proposed system performance is evaluated according to the accuracy of the recognition rate.

5. RESULTS AND DISCUSSION

The proposed system was able to detect the disguised face using Viola-Jones Algorithm under different disguises and background conditions. Figure 4 shows the disguised face detection result.

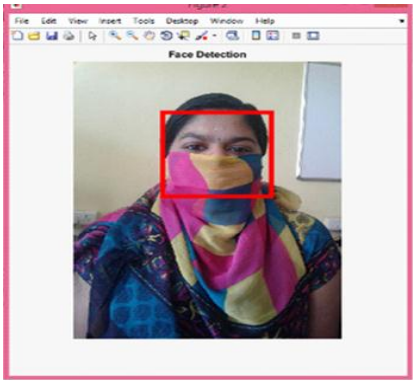

Fig-4: Disguised Face Detection Result

The features extracted from the fully connected layer have been used to successfully train a multiclass SVM classifier and the accuracy is calculated. Accuracy is defined as the number of labels the network predicts correctly. Here the accuracy obtained is around 0.881.

The accuracy is quite convincing and it shows that the system performs well in recognizing faces covered using scarves.

The successful classification result is shown in figure 5. The system correctly classified test images under correct labels.

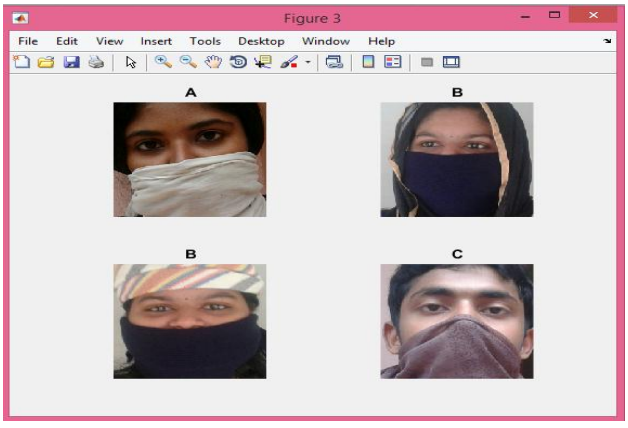

Fig-5: Successful Classification Result

Figure 6 shows a misclassification result. Here the second face in the second row which belongs to label C has been misclassified under label B.

Fig-6: Misclassification Result

The system can be further improved by increasing the images in the database and by using a much deeper convolutional network. However, detecting and recognizing these disguised images when the eye region also gets covered up is still a challenging task.

6. CONCLUSION

Disguised face recognition is an interesting and challenging task. Not much work has been done in recognizing people who hide their identity by covering their face with scarves using deep learning methods. This paper evaluated the extracted learned features from a pre-trained Convolutional Neural Network (Alex-Net) followed by multi-class SVM algorithm to perform recognition of covered faces. The face detection is performed using Viola-Jones detection algorithm. The detected and pre-processed face image is fed as an input to the CNN (Alex-Net). The proposed system is tested on a newly introduced disguise dataset and a high accuracy rate is achieved. Accuracy can be further improved by increasing the number of images in the dataset or by using a much deeper convolutional neural network. This can be very useful for law enforcement and other organizations as it would help to identify criminals, Protestants or anyone who hides their identity.

REFERENCES

[1] Parama Bagchi, Debotosh Bhattacharjee and Mita Nasipuri "Robust 3D face recognition in presence of pose and partial occlusions or missing parts" International Journal in Foundations of Computer Science & Technology (IJFCST), Vol.4, No.4, July 2014

[2] T. I. Dhamecha, R. Singh, M. Vatsa, andA. Kumar. "Recognizing disguised faces: Human and machine evaluation"(2014),Available: https://doi.org/10.1371/journal.pone.0099212

[3] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks", In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, Advances in Neural Information Processing Systems 25, pages 1097–1105. Curran Associates, Inc.

[4] Karl Weiss, Taghi M. Khoshgoftaar and DingDing Wang, "A survey of transfer learning", Journal of Big data (2016)3:9

[5] Patterson, K. E., & Baddeley, A. D. (1977). "When face recognition fails". Journal of Experimental Psychology: Human Learning and Memory, 3, 406 417.

[6] Ramanathan, N.; Chowdhury, A. & Chellappa, R. (2004). "Facial similarity across age, disguise, illumination and pose", Proceedings of International Conference on Image Processing, Vol. 3, pp. 1999–2002

[7] Martinez, A. & Benavente, R. (1998). The AR face database. Computer Vision Center, Technical Report.

[8] Alexander, J. & Smith, J. (2003). "Engineering privacy in public: Confounding face recognition, privacy enhancing technologies", Proceedings of International Workshop on Privacy Enhancing Technologies, pp. 88–106

[9] Cox, I.J.; Ghosn, J. & Yianilos, P.N. (1996). "Feature-based face recognition using mixture distance", Proceedings of International Conference on Computer Vision and Pattern Recognition, pp. 209-216

[10] G. Righi, J. J. Peissig, and M. J. Tarr. "Recognizing disguised faces", Visual Cognition, 20(2):143–169, 2012.

[11] Toseeb U, Keeble DR, Bryant EJ (2012) "The significance of hair for face recognition". PloS ONE 7(3): e34144.

[12] R. Singh, M. Vatsa, and A. Noore. "Recognizing face images with disguise variations". InTech, 2008.

[13] Hinton, G. E. (2002). "Training products of experts by minimizing contrastive divergence", Neural computation, 14(8):1771–1800.

[14] Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006) "A fast learning algorithm for deep belief nets". Neural Computation, 18(7):1527-1554

[15] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. (2010). "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion", The Journal of Machine Learning Research, 11:3371–3408.

[16] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A., and Fei-Fei, L. (2015). "Imagenet large scale visual recognition challenge". International Journal of Computer Vision, 115(3):211–252.

[17] Taigman, Y. and Yang, M. and Ranzato, M. and Wolf, L. 2014. "DeepFace: closing the gap to human-level performance in face verification", In IEEE Conference on Computer Vision and Pattern Recognition, 1701 – 1708

[18] Sun, Y.; Wang, X.; and Tang, X. 2015. "Deeply learned face representations are sparse, selective, and robust", In IEEE Conference on Computer Vision and Pattern Recognition, 2892 – 2900.

[19] Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. "Facenet: A unified embedding for face recognition and clustering", In IEEE Conference on Computer Vision and Pattern Recognition, 815–823

[20] Viola, P., and Jones, M. J. 2004. "Robust real-time face detection"., International Journal of Computer Vision 57(2):137–154

[21] J. Weston, C. Watkins "Multiclass Support Vector Machines" Technical Report CSD-TR-98-04 May 20, 1998

[22] L´eon Bottou NEC Labs America, Princeton NJ 08542, USA leon@bottou.org, "Large-Scale Machine Learning with Stochastic Gradient Descent"