2020 IEEE REGION 10 CONFERENCE (TENCON)
Osaka, Japan, November 16-19, 2020

Convolutional Neural Network Based Criminal Detection

Harsh Verma[1], Siddharth Lotia[1], and Anurag Singh[1]

[1]Department of Electronics and Communication Engineering , IIIT Naya Raipur , India
{*harsh18101,siddharth18101,anurag*}@*iiitnr.edu.in*

***Abstract*—Various recent advancements in deep learning models have greatly boosted the performance of semantic pattern recognition using images. Various state estimation of an individual like emotional state and other certain character features or traits can be estimated from the facial images. With this motivation, in this work we are attempting to infer criminal tendency or (crime prediction/detection) from facial images by using the learning capabilities of various deep learning architectures. More precisely two type of deep learning models we have used in this study: standard convolutional neural network(CNN) architecture and pre-trained CNN architectures, namely VGG-16, VGG-19, and InceptionV3. We have done a performance comparative analysis among these models for efficiently capturing criminal traits from a human face. The efficacy of the above deep learning models was evaluated on a public database, National Institute of Standards and Technology (NIST). To avoid any discrepancies, we have only used male images in this work. It was found that VGG CNN models are best performing models, especially in a limited data scenario producing the classification accuracy of 99.5% in identifying criminal faces.**

***Index Terms*—Image Classification, Facial images, Semantic pattern Recognition , Personality Traits , Deep Learning, Image Processing**

I. Introduction

Our faces might disclose more than what we expect. such as race, gender, age, health, emotion, psychology, and profession. We can know about a person's trait and infer feelings of people by looking at their face. From various studies like Lombroso's study [1], which presented a physiological perspective of facial expressions and reveals the fact that criminals can be identified by their facial images using their facial structure and emotions at a particular time. Unlike the physiology and psychiatry perspective, if we could make a machine to learn and identify a person using his facial images, it would have a great impact on controlling crime activities and applications. Motivated by this idea, people have started research in this domain to find useful facial traits or features using which machine can be trained to identify criminal tendencies of human being, however sufficient reliability and proficiency have not been achieved so far. With the significant advancements in the computation capability, deep learning based neural network architectures have gained much attention these days in all walks of engineering applications. This paper examines the proficiency of different convolutional neural network (CNN)-based deep learning algorithms in getting us to find the difference between the facial characteristic of criminal and non-criminal faces. We have taken two prominently known deep learning models standard CNN and pre-trained CNN models VGG-16, VGG-19, and InceptionV3. They are trained with over 4000 facial images that are mixed color, race, gender and then applied to differentiate criminal and non-criminal. The idea behind exploring pre-trained models is that we are having limited data-sets and deep learning models are always data hungry for their reliable performance. However, best thing about pre-trained models is that they don't require training as they are already fully trained on a wide general data-sets and ready-to-use directly for a detection/classification task. So, in this way scarcity of the data will no more be a challenge and we can go for directly testing without compromising with the model training. This study may be considered as a new kind of department that can be called as cyber-forensic for dealing the crime by predicting the behavior of criminals and detecting the nature of crime to be done by criminals.The disclaimer of this work is that it is limited to technical and analytical aspects and not questioning social aspects as it requires a high level of caution and supervision. This work can be further improved with the availability of a large and variety of available data-set. Large corpus will also help in applying multiple deep learning algorithms for getting results that are more accurate. The rest of the paper is organized as follows: Section 2 discusses the related work, methodology has been discussed in Section 4. Results and discussion has been presented in Section 5 and finally conclusion are drawn in Section 5.

II. Related Works

In the modern era, advanced machine learning tools have been key to crime prevention, identification and surveillance applications. Most of the crime-based analysis [2-3] are being done today using some or other machine learning approaches. Crime rate diagnosis against women using machine learning approach has been reported by Tamilarasi et. al [2]. The authors have used previous data to predict the crime. Authors in [4] proposed a methodology to identify criminal activities through a video stream by capturing the abnormal activities by the person in successive video frames. Different CNN architectures like DCNN[5], RNN (Recurrent Neural Network)[6], etc. have been employed to capture the abnormal behavior in the video frames. The DCNN can help in identifying important features from the frames with the help of the HDL algorithm. Navalgund et. al. [7] show that it is possible to detect the intention of the criminal in real-time using videos, images, and alert can be sent to a nearby police station. The pre-trained deep learning models like VGG-19 [8] and GoogleNet [9] have been used in the related literature to detect knife, gun in the hand of a person and pointing it to another person. Real-time crime detection using Machine Learning and Deep Learning for the prevention of crime have been proposed by authors in [10]. Authors in above work proposed an application that helps police officers to know about the possible incident which may happen around in real-time and also gives prediction about possible crimes that might take place in the near future. Among the pre-trained CNN models, Resnet50 [11] gives the best result in detecting the abnormal behaviors followed by InceptionV3. This motivated us to explore pre-trained models for detecting criminal tendencies using face images in this work.

III. Motivation

Identifying criminal out of a large crowd or predicting the ill intentions of someone who is going to commit some crime is really a challenging task. However, it is well said that your face is the reflection of your mind/thought process, which was physiologically proved by Lombroso's Study [1]. The study reveals that the criminals have a set of facial traits/features, which if captured precisely can be employed to identify any criminal tendency by an unknown person. Lombroso's method was clinical and descriptive, with accurate details of skull dimension and other things. He considered that and was convinced that specific criminals can be differentiate by certain characteristics/features like sloping forehead, ears of unusual size, asymmetry of the face, prognathism, excessive length of arms, asymmetry of the cranium, and other physical stigmata. However, these features are hard to capture manually. Due to rapid advancements in the computational capability of machines, deep learning based neural network architecture have gained wide popularity in solving complex and practical classification and identification problems. This motivated us to explore different deep learning based architectures in this work and present a comprehensive performance analysis for criminal tendency detection using facial images. For visual perception, a snapshot of multiple images having criminal/non-criminal tendencies/record is shown in Fig. 2.

IV. Methods

A general work-plan of this work is shown in Fig. 1 in the form of a block diagram. After getting a facial image, the first step is image pre-processing. In the pre-processing stage, firstly we have used haarcascade to detect frontal images from the data-set,then the images are converted into grayscale followed by their resizing to 128*128. After pre-processing, train and test set were created. Training sets have been used to train the self-created CNN model and test-sets were used with pre-trained models, namely VGG-16, VGG-19 and Inception V3.

A. Data Preparation

The proposed methodology was evaluated on a public database, called National Institute of Standards and Technology (NIST) [12]. A total of 8401 images are present in the NIST database. This database mostly contains images of the criminals. We only need a frontal face for our work so we used a haarcascade [13] to obtain a frontal image of the criminals and remove the other images. For non criminal we use around 2300 images taken from various databases. We have removed all the duplicates from our database. For non criminal most of the images are coloured so after getting frontal faces through haarcascading, all the colored images are converted to gray scale. After gray scale conversion, the image size is reduced to 128*128 for further processing and analysis [14-16].

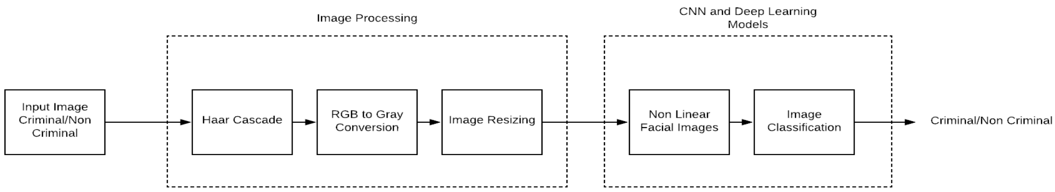

Fig. 1: Block diagram of the Proposed Method

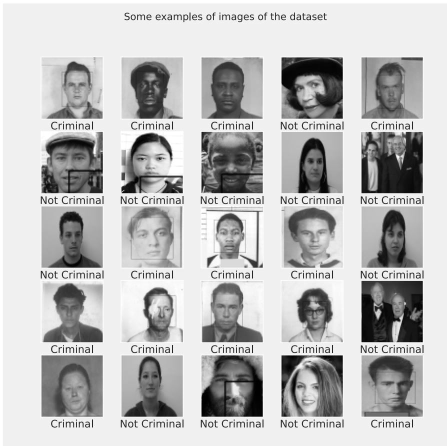

Fig. 2: A snapshot of the Database after using haar cascade and reduce their size to 128*128

B. HaarCascade

Haarcascade is a computer vision algorithm used for detecting objects in an image, video. It has various applications such as vehicle detection in a streaming video, face and eye detection etc.

C. CNN Architectures

CNN's basically consists of an input layer and in between, there are several hidden layers i.e multilayer neural networks. At the end, there is the output layer. These layers are in the form of optimization layers and pooling layers which include max pooling,average pooling, activation function. Activation function is basically a non linear transformation is performed before transferring neuron's output to the next layer. Dropout layer helps in regularization and save the model from over-fitting. At the there is a fully connected layers where all the input layer neurons are connected to the output layer's neuron Fig 3 shows the general architecture of the CNN model we used in this work. The layers details have been discussed in the subsequent sections. There are various pre-trained CNN architectures like Alexnet, VGG-19, VGG-16, Resnet50, Resnet18, Resnet22, GoogleNet, Inception Net. These pre-trained architectures were developed in the imagenet challenge.
Convolutional Layer: Convolution filters are there which calculate the correlation between the different pixels and extract the non-linear features from the input image. Each neuron remembers different features in the neural network.

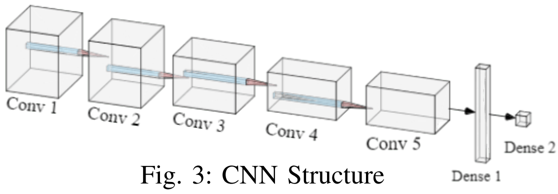

Fig. 3: CNN Structure

Relu: The Relu is a linear activation function called Rectified Linier Unit (Relu). It gives output directly if it is positive, otherwise, it will output zero. It is easier to train and often achieves better performance in neural networks.
It can be defined as follows:
$R(z) = \max(0, z)$

Pooling Layers: Pooling layers can help us in reducing the proportions of the feature maps.It helps in reducing the number of variables to learn thus reducing the no of computations required to train the network. There are two popular functions which commonly use for polling :
Max Pooling[17]: It gives the maximum value of each feature map.
Average Pooling[18]: It gives the average value of each feature map.

Sigmoid neuron: One of the limitations of the Perceptron model that it can work only on linearly separable data. The mathematical equation of the sigmoid is:

$$y = \frac{1}{1 + e^{-(W^T X + b)}} \quad (1)$$

Sigmoid is generally harsh on the boundary. Whereas sigmoid function gives smoother results in the form of probability and less harsh, unlike perceptron. We have used sigmoid activation function for classification of 2 classes.

$y = 1$ if $\sum_{i=1}^{n} w_i x_i \geq b$ (2)

y=0 $otherwise$

D. Optimization Algorithm

1) SGD (Stochastic Gradient Descent): Gradient descent is one of the most famous algorithms in deep learning and machine learning. Meaning of the stochastic is linked with the random probability. In this optimization, batches are taken from the whole data-sets which is used for training the networks to find the local minima and minimize the loss. The mathematical equation of Gradient Descent is given as follows:

Gradient Descent Algorithm:

$$\theta_{n+1} = \theta_n - (\alpha\delta/\delta\theta_n)\, J\,(\theta_n) \tag{3}$$

where

$\theta_n \rightarrow$ Parameter Vector

$J \rightarrow$ cost Function

$\alpha \rightarrow$ Slope Parameter

When data-sets are very large it becomes a computationally intensive algorithm. Other optimizers are preferred when data-sets are too large.

2) ADAM: Adam optimizer can be looked as a combination of the two ideas of RMSprop and stochastic gradient descent with momentum. The mathematical equation of the Adam optimizer is given below:

For every parameter w^j

$v_t = \alpha_1 \times v_{t-1} - (1 - \alpha_1) \times g_t$ (4)

m$_t = \alpha_2 \times m_{t-1} - (1 - \alpha_2) \times g_t^2$ (5)

$\triangle w_t = \Delta w_t - \eta \times (v_t \div \sqrt{m_t + \epsilon}) \times g_t$ (6)

w$_{t+1} = w_t + \Delta w_t$ (7)

$where$

η : Initial learning rate

v_t : Exponential Average of Gradients along w_j

s_l : Exponential Average of squares of Gradients along w_j

g_t : Gradients at time t along w_j

α_1, α_2 : Hyperparameters

E. Transfer Learning

Transfer learning is popular methodology in machine learning which basically deals with keeping the knowledge acquired from solving one problem and use it to solve a related one. It is one of the popular approaches where pre-trained models are used for solving the classification problems.The transfer learning model generally performs better than building a new CNN model with random initialization, convergence, etc. Some of the popular pre-trained models which have been used in work for criminal tendency prediction are discussed below:

1) VGG-16 and VGG-19: VGG-16 [19] and VGG-19 are pre-trained CNN architectures[20]. In VGG the size of all the filters is 3*3 with a fixed stride of 1 and max pool layer have filter of size 2*2 with stride 2 and have the same padding. It follows the same architecture of Convolution and Max pool throughout the architecture. In the end it has 2 feed forward neural networks followed by softmax layer.

2) Inception V1, V2, and V3: Inception Network [21] is one of the best architectures to solve most of the deep learning problems. Inception performs convolution using three different types of filters 1*1, 3*3, and 5*5. Inception due to a large number of parameters, it can suffer from a vanishing gradient problem to solve this problem there are generally 2 auxiliary classifiers.

V. Result

The evaluation of the proposed approach has been carried out using NIST public database. First we passed all the facial images taken from NIST database from haarcascade to get the frontal images only. After getting frontal images, we converted them to gray scale followed by resizing to 128*128. The size is chosen as per our GPU capacity. Table 1 depicts the number of Criminal and Non Criminal Faces in our data-set. We have taken a total of 2492 train images and 1224 validation images.

TABLE I: Data-set information

	Criminal	Not Criminal
Train data-set	1100	1392
Test data-set	422	800

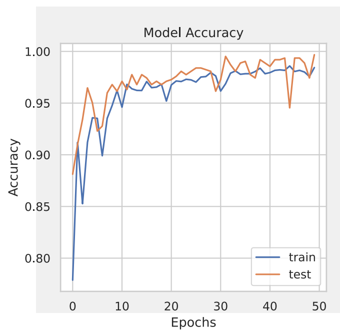

Fig. 4: Our Model Accuracy

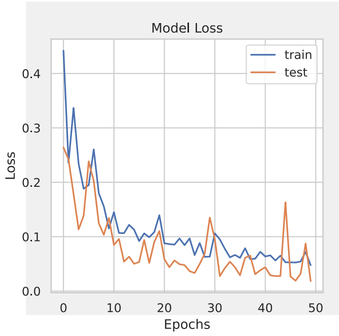

Fig. 5: Log Loss

Using the pre-processed facial images, we have trained the two types of CNN models, first our own made CNN model and second the pre-trained CNN models with fine tuning. Our own CNN model was trained with a learning rate of 0.001 and 5 hidden layers with a max pooling stage. And finally we add a feed forward neural network of 2 stages. We did use the sigmoid function for classification purposes. Model's train and test accuracies and corresponding losses with respect to no. of epochs have been shown in Fig. 4 and Fig. 5. From these two plots, we can say that the model is trained properly as the the test accuracy reaches to maximum close to training accuracy in around 15-20 epochs and corresponding loss also saturates to the minimum value in similar no. of epochs. We have used pre-trained models VGG-16, VGG-19 and InceptionV3 with fine tuning. This implies, we did not perform conventional training as we have done in previous CNN model we have used the already trained weights of the hidden network with fine tuning by adding a feed forward layer followed by sigmoid function for classification. The hyperparameters of the pre-trained models are tuned properly for validation and testing on out data-set. The model accuracies against the no. of epochs for VGG-16, VGG-19 and InceptionV3 models are plotted in Fig 6, Fig 7, Fig 8 respectively. From these plots, we can conclude that the VGG-16 and VGG-19 models are more fine tuned compared to InceptionV3 and this reflects in their performance too. A performance comparison among all the models used in this work is done done and the obtained performance measures in terms of precision, recall, AUC, and accuracy are shown in the Table III. From these qulity metrics, we can conclude that the VGG-16 and VGG-19 models are more suitable for our task of criminal tendency identification as they produce best results in terms of accuracy. The VGG-16 achieves a maximum of 99.5% accuracy whereas VGG-19 gets 99.4% accuracy after 5-fold cross validation. To see how many images from both criminal and non-criminal class got correctly classified, we have given a confusion matrix in Table III. We can observe from here that only 9 images from criminal and 8 from non-criminal classes are getting mis-classified. We also try to find the percentage of criminal and Not Criminal in our data-set and have found around 56% percent is not criminal and 44% are criminal. It is also important to find the number of criminals and non criminals images in our data-set because there can be problem of imbalance class. We ensured that our model is not suffering from imbalance class problem.

We have also compared our work with some of the existing similar works and presented it in Table IV. After comparing our work with recent similar kind of work in this domain we found out our work has attained higher accuracy than others.

TABLE II: Classification Report of Our CNN model

	Precision	Recall	AUC	Accuracy
CNN Model	0.99	0.99	0.99	99.0%
VGG-16	0.980	0.978	0.997	99.5%
VGG-19	0.984	0.985	0.996	99.4%
Inception-V3	0.914	0.904	0.928	91%

TABLE III: Confusion Matrix

Criminal	413	9
Non-Criminal	8	792

TABLE IV: Comparison of Our Work with Recent Work in This Domain

Work	Learning Rate	Accuracy	Initializer	HiddenLayer
Our Work	0.001	99.5%	ImageNet	13
Mikolov et al [16]	0.0001	97%	Random Normal	2
Navalgund et al. [7]	0.001	92%	ImageNet	16

VI. CONCLUSION

Classification of any person requires effort, but more care and seriousness is needed to classify a criminal or a suspect. The shortcoming of this work can be in its some imperfection, because any wrong classification can have serious effects. It will be very biased and too optimistic for us to say that 99 percent accuracy that has been achieved by CNN is cent percent acceptable. This is because of various reasons like, small size of data-set, all images taken are may not be taken in the same

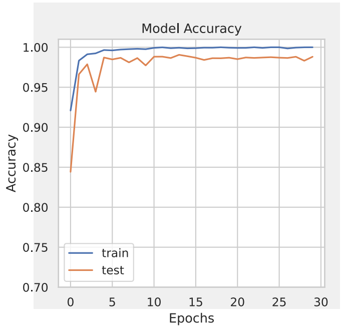

Fig. 6: Accuracy Using VGG-16

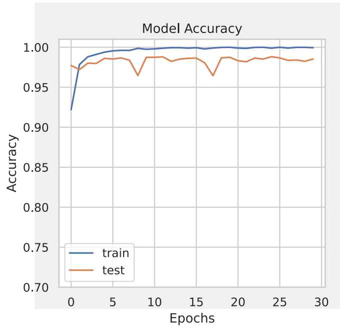

Fig. 7: Accuracy Using VGG-19

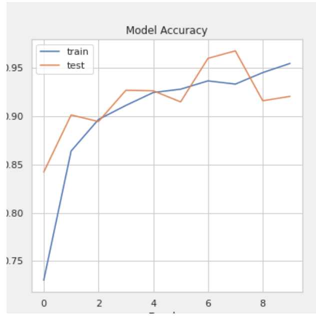

Fig. 8: Accuracy Using InceptionV3

conditions, which can raise questions in this classification. Majorly facial Images are classified using facial emotions and age, so first neutral images and elderly, children images were eliminated .We tried to remove this bias by using haarcascade by cropping the facial part out of the images, but also shown they have less impact on results . So if we create a greater data-set ,by taking in account the various factors mentioned above and detecting other personality traits/features can be our future scope of study.

References

[1] Zebrowitz LA, Montepare JM. Social psychological face perception: why appearance matters. Soc Personal Psychol Compass. 2008;2(3):1497–517.

[2] Tamilarasi, P., and R. Uma Rani. ”Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning.” 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.

[3] Kim, Suhong, et al. ”Crime analysis through machine learning.” 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2018.

[4] Chackravarthy, Sharmila, Steven Schmitt, and Li Yang. ”Intelligent crime anomaly detection in smart cities using deep learning.” 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC). IEEE, 2018.

[5] Ren, Ao, et al. ”Sc-dcnn: Highly-scalable deep convolutional neural network using stochastic computing.” ACM SIGPLAN Notices 52.4 (2017): 405-418.

[6] Mikolov, Tomáš, et al. ”Extensions of recurrent neural network language model.” 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2011.

[7] Navalgund, Umadevi V., and K. Priyadharshini. ”Crime Intention Detection System Using Deep Learning.” 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET). IEEE, 2018.

[8] Mateen, Muhammad, et al. ”Fundus image classification using VGG-19 architecture with PCA and SVD.” Symmetry 11.1 (2019): 1.

[9] .Zhong, Zhuoyao, Lianwen Jin, and Zecheng Xie. ”High performance offline handwritten chinese character recognition using googlenet and directional feature maps.” 2015 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2015.

[10] Rajapakshe, Charuni, et al. ”Using CNNs RNNs and Machine Learning Algorithms for Real-time Crime Prediction.” 2019 International Conference on Advancements in Computing (ICAC). IEEE, 2019.

[11] Akiba, Takuya, Shuji Suzuki, and Keisuke Fukuda. ”Extremely large minibatch sgd: Training resnet-50 on imagenet in 15 minutes.” arXiv preprint arXiv:1711.04325 (2017).

[12] NIST Special Database 18. 2010. https://www.nist.gov/srd/nist-special-database-18. Accessed 12 Apr 2019

[13] Cuimei, Li,et al. ”Human face detection algorithm via Haarcascade classifier combined with three additional classifiers.” 2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI). IEEE, 2017.

[14] Face Recognition Database. http://cbcl.mit.edu/software-data-sets/heisele/facerecognition-database.html Accessed 02 June 2018.

[15] FEI Face Database. http://fei.edu.br/ cet/facedatabase.html. Accessed 02 June 2018

[16] Hashemi, Mahdi, and Margeret Hall. ”Criminal tendency detection from facial images and the gender bias effect.” Journal of Big Data 7.1 (2020): 1-16.

[17] Nagi, Jawad, et al. ”Max-pooling convolutional neural networks for vision-based hand gesture recognition.” 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA). IEEE, 2011.

[18] Lee, Chen-Yu, Patrick W. Gallagher, and Zhuowen Tu. ”Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree.” Artificial intelligence and statistics. 2016.

[19] Qassim, Hussam, Abhishek Verma, and David Feinzimer. ”Compressed residual-VGG-16 CNN model for big data places image recognition.” 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2018.

[20] Shin, Hoo-Chang, et al. ”Deep convolutional neural networks for computer-aided detection: CNN architectures, data-set characteristics and transfer learning.” IEEE transactions on medical imaging 35.5 (2016): 1285-1298.

[21] Szegedy, Christian, et al. ”Rethinking the inception architecture for computer vision.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.