DEEP CONVOLUTION NEURAL NETWORKS FOR BINARY CLASSIFICATION OF MASKED FACE IMAGES USING SVM AND TRANSFER LEARNING

Project report in partial fulfillment of the requirement for the award of the degree of
Bachelor of Technology
In
COMPUTER SCIENCE

Submitted By

Aniket Nayek	University Roll No. 12020009023004
Naiwrit Mullick	University Roll No. 12020009001041
Vishal Kumar	University Roll No. 12020009001047
Shreyon Sinha	University Roll No. 12020009001011
Aditya Kumar	University Roll No. 12020009023126
Abhishek Das	University Roll No. 12020009023039
Subham Das	University Roll No. 12020009028001

Under the guidance of

Prof. Dr. Maumita Chakraborty

&

Prof. Prasenjit Kumar Das

Department of Computer Science

UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

University Area, Plot No. III – B/5, New Town, Action Area – III, Kolkata – 700160.

CERTIFICATE

This is to certify that the project titled **Deep Convolution Neural Networks for Binary Classification of Masked Face images using SVM and Transfer Learning** submitted by **Aniket Nayek** (University Roll No. 12020009023004**),** **Naiwrit Mullick** (University Roll No. 12020009001041**)**, **Vishal Kumar** (University Roll No. 12020009001047**),** **Shreyon Sinha** (University Roll No. 12020009001011), **Aditya Kumar** (University Roll No. 12020009023126), **Abhishek Das** (University Roll No. 12020009023039) and **Subham Das** (University Roll No. 12020009028001), students of UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA, in partial fulfilment of requirement for the degree of Bachelor of Computer Science, is a bona fide work carried out by them under the supervision and guidance of Prof. Dr. Maumita Chakraborty and Prof. Prasenjit Kumar Das during 3rd Semester of academic session of 2020 - 2021. The content of this report has not been submitted to any other university or institute. I am glad to inform that the work is entirely original, and its performance is found to be quite satisfactory.

Prof. Dr. Maumita Chakraborty
Project Guide
HOD, Department of Computer Science and Technology
UEM, Kolkata

Prof. Prasenjit Kumar Das
Project Guide and Assistant Professor
Department of Computer Science and Technology
UEM, Kolkata

Prof. Dr. Sukalyan Goswami
Head of the Department
Department of Computer Science and Engineering
UEM, Kolkata

ACKNOWLEDGEMENT

We would like to take this opportunity to thank everyone whose cooperation and encouragement throughout the ongoing course of this project remains invaluable to us.

We are sincerely grateful to our guide Prof. Dr. Maumita Chakraborty and Prof. Prasenjit Kumar Das of the Department of Computer Science, UEM, Kolkata, for his wisdom, guidance and inspiration that helped us to go through with this project and take it to where it stands now.

We would also like to express our sincere gratitude to Prof. Dr. Sukalyan Goswami, HOD, Computer Science, UEM, Kolkata and all other departmental faculties for their ever-present assistant and encouragement.

Last but not the least, we would like to extend our warm regards to our families and peers who have kept supporting us and always had faith in our work.

Aniket Nayek
Naiwrit Mullick
Vishal Kumar
Shreyon Sinha
Aditya Kumar
Abhishek Das
Subham Das

TABLE OF CONTENTS

DEEP CONVOLUTION NEURAL NETWORKS FOR BINARY CLASSIFICATION OF MASKED FACE IMAGES USING SVM AND TRANSFER LEARNING

ABSTRACT

The project presents a hybrid model of integrating the synergy of two superior classifiers: pre-trained Deep Convolution Neural Network (DCNN) i.e., "Alexnet" and Support Vector Machine (SVM), which has achieved better results in recognizing different types of patterns from masked face and non-masked face. In this project, CNN model "Alexnet" works as a terrible feature extractor and SVM performs as a recognizer. Our hybrid model will automatically extract 1000 features from the raw images and generates the predictions. Experiment have been conducted on dataset from RMFD GitHub. We have tried to achieve maximum recognition and classification accuracy using our fusion model. These performances have been analyzed with reference to those by human subjects.

1. INTRODUCTION

Due to the recent pandemic of COVID-19 virus around the world, people across the country wear mask and there appear a large number of masked face samples. Based on the samples of face masked, the corresponding masked face detection, recognition algorithms, and possible intelligent management and control system are designed for public safety events in the future. In addition, the upgrade of face recognition, facial attendance machines, and facial security checks at public places can be adapted.

The surveillance cameras in different public places routinely collect image data in effort to maintain security. Since in these places humans are hired for classification of masked and non-masked citizens as part of precautions against COVID-19 virus. We can use our CNN-SVM model to detect and recognize the masked as well non-masked citizens in cost-effective way with a high accuracy level. In this way, by developing appropriate algorithm for recognition will significantly leads in the reduction of cost. We have come a long way from manual visual of security checkers to acquiring face data using high-speed digital cameras with automated face capture features.

Once the high-resolution digital images of the faces are obtained, they are processed through a compression sub-system to achieve a size and pixel reduction without loss of quality before they are stored. The images are then processed using various algorithms to extract different image features (i.e., orientation, density, location) in a form a 2-D matrix are then stored in the local database and can be linked with our hybrid CNN-SVM model.

Feature extraction is one key factor in the success of a recognition system. It requires that features should have the most distinguishable characteristics among different classes while retaining invariant characteristic within the same class. Traditional hand-designed feature extraction is tedious and time-consuming task and cannot process raw images, while the automatic extraction methods can retrieve features directly from raw images. Szarvas et al. [1] evaluated the automatically optimized features learned by the Convolution Neural Network on pedestrian detection and showed the CNN-features plus SVM combination generated the highest accuracy. Mori et al. [2] trained the convolutional spiking neural network with time domain encoding schemes module by module using different fragment images. The outputs of each layer in the model were fed as features to the SVM. 100% face recognition rate was obtained on the 600 images of 20 people. Inspired by this particular work, we have tried the hybrid CNN-SVM model. Since both CNN and SVM have already achieved superb performances for recognition, we focus our project on the fusion of CNN and SVM to bring out their best qualities.

In this project, we have tried a hybrid CNN-SVM model for masked face and non-masked face detection and recognition. This model automatically retrieves features based on the CNN architecture of "Alexnet" and recognizes the unknown patterns using the SVM recognizer. By transfer learning from a popular pre-trained seven-layered "Alexnet-CNN" model have returned a set of 1000 valuable parameters as features extracted from the image dataset, which has obtained the low error rates of 1.4511%. To verify the feasibility of the methodology, the RMFD GitHub dataset is tested. The extracted features are then fed into linear SVM classifier as a training data for binary classification of two different classes i.e., masked face and non-masked face.

2. LITERATURE SURVEY

Previous research also has been done by F. Lauer that automatically extracts feature using a trainable feature extraction technique based on the Convolution Neural describe in [3], which showed a high-performance of handwritten digit recognition. In the same field, Szarvas et al. [1] has also tried to evaluate the automatically optimized features learned by the CNN on pedestrian detection and showed CNN plus SVM combinations. In another paper [2], the convolution spiking neural network with time domain encoding schemes module by module using different fragmented images has achieved a 100% face detection rate when the features were fed into SVM classifier. Xiao-Xiao Niu [4] trained a novel hybrid CNN-SVM classifier for handwritten digit recognition has achieved an accuracy of 99.65%.

3. PROBLEM STATEMENT

Remarkable progress has been made in image classification, primarily due to the availability of large-scale annotated datasets and a decent accuracy score has been achieved. In this project, we will try to improve the accuracy score of Real-World Masked Face (RMFD) by implementing pre-trained Deep Convolution Neural Network (DCNN) and Support Vector Machine (SVM) algorithm for classification. The primary objective of this project is to redesign a supervised deep convolution neural network for the binary classification of real-world masked face images and to improve the accuracy of the prediction using some transfer learning from “Alexnet”. The SVM algorithm will complement the network model for getting more accurate and better results.

4. EXPERIMENTAL SETUP

The goal of this project is to develop a simplified and effective vision-based detection system by using a pre-trained deep learning model to detect masked face from images through transfer learning and using linear SVM classification. There are currently some major techniques that successfully employ CNNs to image classification: training the CNN from scratch, using the off-the-shelf pre-trained CNN networks, etc. But here we will be using a hybrid CNN-SVM classifier for recognizing real-world masked and non-masked faces. Generally, CNNs enable training data-driven, highly representative, layered hierarchical image features from sufficient training data. Here at first, we discuss the images dataset which come from an open-source, RMFD GitHub. The overall methodology is then described which involves the use of pre-trained Alexnet DCNN for feature extraction from the Fully Connected eighth layer (fc8), and then train a machine learning classifier i.e., Support Vector Machine (SVM) to predict the labels.

4.1 Dataset

We used a subset of the Real-World Masked Face dataset from the RMFD GitHub which contains research quality image dataset for over 5,000 masked faces of 525 people and 9,000 normal faces throughout the People's Republic of China. We used a subset of the original dataset, which contain 2,165 masked face and 1,930 normal faces. Further we divide the dataset into training and testing. The training set contains a total 3,239 randomly chosen images, and the rest 810 randomly chosen images were considered for testing. Once again, the training set was used for cross-validation.

4.2 Deep Convolution Neural Network and Transfer Learning

Deep Neural Networks (DNNs), more commonly referred to as Deep Learning, employ deep NN architectures to automatically learn hierarchy of features from raw input data without the need for feature engineering [5]. Loosely inspired by how the mammalian brain uses different areas of the cortex to abstract different levels of features when given an input percept, deep learning methods are characterized by deep architectures with several hidden layers that allow them to learn many levels of abstraction, as opposed to shallow architectures with 1 or 2 hidden layers.

DCCNs have shown to be highly effective in processing visual data, such as images and videos. DCNNs take raw input data at the lowest level and transforms them by processing them through a sequence of basic computational units to obtain representations that have intrinsic values for classification in the higher layers [6,7]. A DCNN typically consists of three-layer types (Fig. 1): convolution layers, pooling layers, and fully connected layers. A convolutional layer is parametrized by the number of channels, kernel size, stride factor, border mode, and the connection table. The convolution layer takes the input image and applies convolution filter on it to produce the output image. Multiple convolutional layers are used to take into consideration the spatial dependencies among image pixels. The max-pooling layer is used to make the neural network more invariant and robust which lead to lead to faster convergence and better generalization. It is common to use multiple fully connected layers after several rounds of convolution and the resulting structure of the last convolutional layer is flattened before connecting to the following fully connected layer.

DCNNs typically require large, annotated image datasets to achieve high predictive accuracy. However, in many domains, acquisition of such data is difficult and labelling them is costly. In light of these challenges, the use of well-established pre-trained DCNNs such as VGG-16, AlexNet, and GoogLe-Net has shown to be very useful for solving cross domain image classification problems through the concept of transfer learning and fine-tuning [8]. The idea behind transfer learning is that it is cheaper and efficient to use deep learning models trained on ''big data" image datasets (like ImageNet) and ''transfer" their learning ability to new classification scenario rather than train a DCNN classifier from scratch [9]. With adequate fine-tuning, pretrained DCNN has been shown to outperform even DCNN trained from scratch model [8,10].

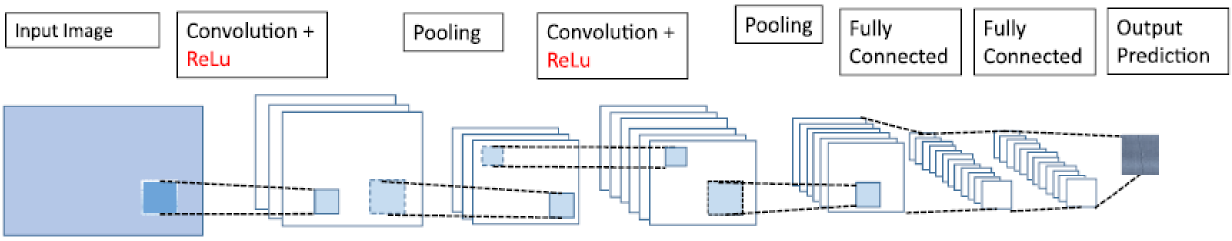

Fig. 1: Schematic of Convolution neural network architecture.

Image Reference: K. Gopalakrishnan, Siddhartha K. Khaitan, Alok Choudhury, Ankit Agrawal, Deep Convolution Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection, p 324. [16]

4.3 The hybrid CNN-SVM Algorithm

Our proposed algorithm was designed to integrate the CNN and SVM classifiers. We will briefly introduce the Alexnet-CNN structure in Section 4.3.1, and the SVM theory in Section 4.3.2. Then, the hybrid CNN-SVM trainable feature extractor model will be presented in Section 4.3.3.

4.3.1 Alexnet-CNN

A Convolutional Neural Network [11] is a multi-layer neural network with a deep supervised learning architecture that can be viewed as the composition of two parts: an automatic feature extractor and a trainable classifier. The feature extractor contains feature map layers and retrieves discriminating features from the raw images via two operations: convolutional filtering and down sampling. And in this project instead of Alexnet classifier, we will be using SVM classifier.

The Alexnet architecture was primarily published by L. Sutskever and Krizhevsky in [12], achieved significantly improved performance over the other non-deep learning methods for ImageNet Large Scale Vision Recognition Challenge (ILSVRC) 2012. AlexNet architecture consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer, as shown in Fig. 2. Each convolutional layer consists of convolutional filters and a nonlinear activation function ReLU. The pooling layers are used to perform maximum pooling. Input size is fixed i.e., [227 227 3] due to the presence of fully connected layers and has a batch size of 128. It overall has 60 million parameters. In this project, we will extract all the features before the softmax and classification layer, and then we fed them in SVM classier.

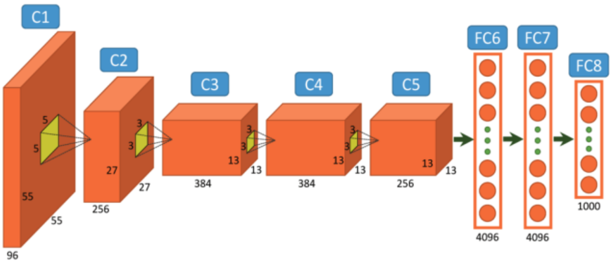

Fig. 2: A schematic of the Alexnet Deep Convolutional Neural Network (DCNN) architecture trained on ImageNet

Image Found on the Web
Sébastien Collet, G. (2017), What is Object Detection? Retrieved from https://www.saagie.com/blog/object-

4.3.2 Linear Support Vector Machine (SVM) classifier

Support Vector Machines [13] with different kernel functions can transform a non-linear separable problem into a linear separable problem by projecting data into the feature space and then finding the optimal separate hyperplane. This method was initially proposed to solve binary class classification problems. Later, a few strategies were suggested to extend this technique to multi-class classification problems. If we have a given a training dataset of points of the form $(x_1, y_1), \ldots\ldots, (x_n, y_n)$, where the y_i either 1 or −1, each indicating the class to which the point belongs. Each x_i is a p-dimensional real vector. The maximum-margin hyperplane that divides the group of points x_i for which $y_i = 1$ from the group of points for which $y_i = -1$, which is defined so that the distance between the hyperplane and the nearest point x_i from either group is maximized. The hyperplane can be written as the set of points x satisfying $w^T x - b = 0$, where w is the normalized vector to the hyperplane.

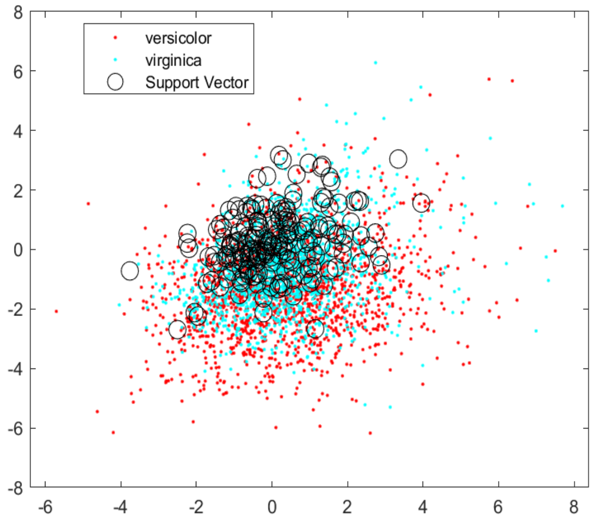

Fig. 3: Linear SVM Scatter Plot of all Support Vectors

4.3.3 Hybrid CNN-SVM

The architecture of the hybrid CNN–SVM model was designed by replacing the last output layer of the CNN model with an SVM classifier. For output units of the last layer in the CNN network, they are the estimated probabilities for the input sample. Each output probability is calculated by an activation function. The input of the activation function is the linear combination of the outputs from the previous hidden layer with trainable weights, plus a bias term. Looking at the output values of the hidden layer (see Fig. 4) is meaningless, but only makes sense to the CNN network itself; however, these values can be treated as features for any other classifiers. Fig. 5 shows the structure of the hybrid CNN–SVM model. Firstly, the normalized and centered input images are sent to the input layer, and the original CNN with the output layer is trained with several epochs until the training process converges. Then, the SVM with a linear kernel function replaces the output layer. The SVM takes the outputs from the hidden layer as a new feature vector for training. Once the SVM classifier has been well trained, it performs the recognition task and makes new decisions on testing images with such automatically extracted features.

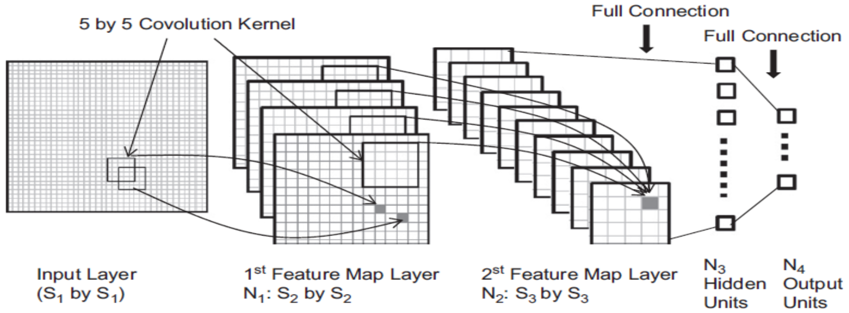

Fig. 4: Structure of the adopted CNN.

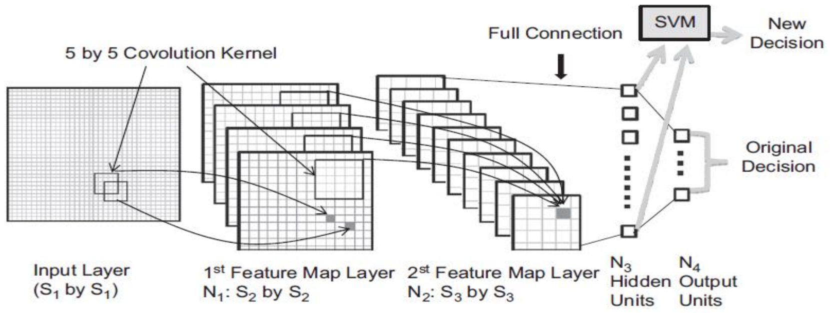

Fig. 5: Structure of the hybrid CNN-SVM model.

Image Reference: Xiao-Xiao Niu, Ching Y. Suen. A novel hybrid CNN–SVM classifier for recognizing handwritten digits. p 1320. [4]

5. RESULT ANALYSIS

The hybrid CNN–SVM model was built and trained. The last fully connected layer of CNN was replaced by an SVM classifier to predict labels of the input patterns. One thousand values from the fully connected layer (i.e., fc8) of the trained CNN network were used as a new feature vector to represent each input pattern and were fed to the SVM (using fitcsvm function in MATLAB) for learning and testing. To build the SVM in the hybrid model, we first applied the Alexnet-CNN, and the extracted feature vector were stored in a variable, and then it was fed into SVM for final classification, and the end we apply the 5-fold cross validation method on the training dataset.

Some papers [3,14,15] have proven that better generalization can be achieved with an expanded training dataset by using distortion techniques. In our experiments, the hybrid CNN-SVM model has been used. In the Fig. 6, the number of function evaluations is the number of iterations of the objective function. The minimum Objective is the minimum value of the objective function reached in that iteration. Here the training procedure was stopped after a 30 iteration, and both converged at a minimum objective function value. With this setup, the hybrid CNN-SVM learning classifier produced an error of 1.45% on the training dataset, and 1.1% on the testing dataset.

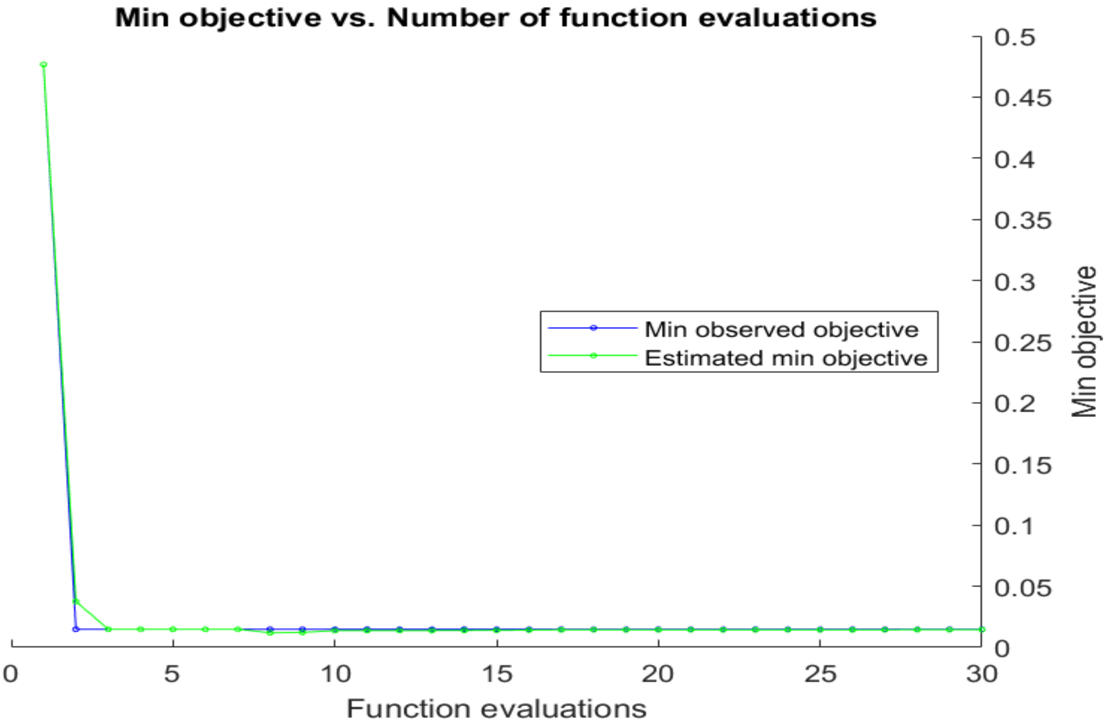

Fig. 6: Number of objective evaluations with iteration.

The above discussed algorithm based on SVM generates a line or hyperplane which separates the two classes as shown in the Fig. 8. The function used for the hyperplane is:

$$f(x) = w^T x - b = 0$$

where *w* are the weight vector and *b* are the bias vector of the training dataset. We can see that the hybrid CNN-SVM model have successfully separates the two classes.

However, a significant achievement was made by the hybrid method with the lowest error of 2.4%, which boosted the performance accuracy compared with the other recognition result up to date. After successfully running the experiment for several times, we have achieved an accuracy score of 98.54% (as shown in Fig. 7). Fig. 7 also shows there are 47 misclassified samples, and it presents the confusion matrix. After analyzing these errors, we found that they can be categorized into two types:

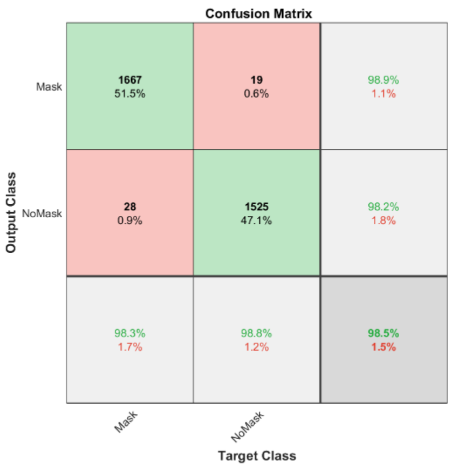

Fig. 7: Confusion matrix of the hybrid model on the RMFD dataset.

(1) Some images in the dataset are not properly captured, it doesn't have all the features of a face and masked face. The images were cropped badly, and in somewhere there are few texts on the image, that's the reason why the hybrid model failed in properly identifying some of the image class.
(2) One more observation about the unclassified images is about the degraded quality of masked images, they are blurred to extract the proper image features. The image was not as clear as it was expected. That why the model misclassified the image.

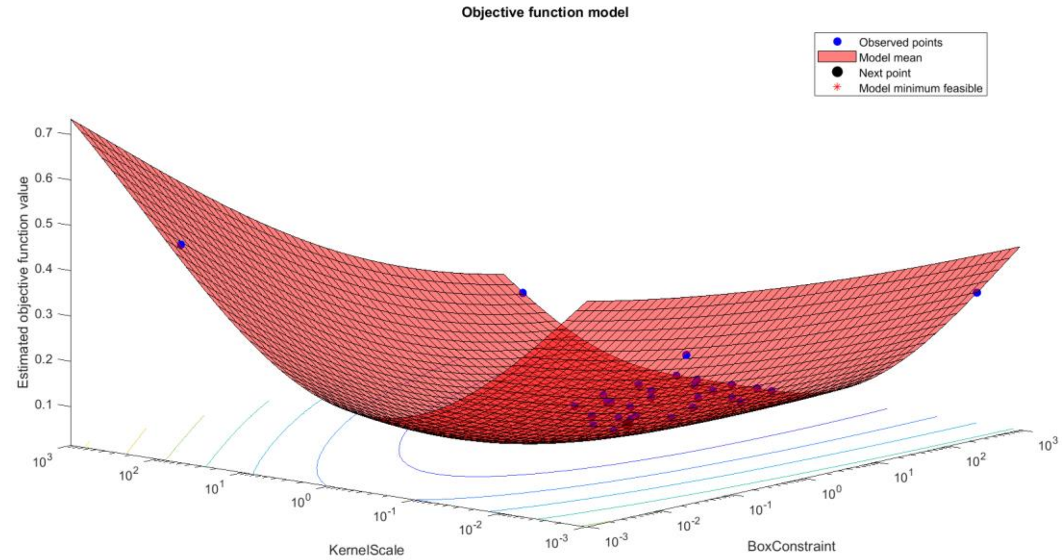

Fig. 8: SVM generated hyperplane for Binary Classification.

5.1 Merits of hybrid model

Our expectation that the hybrid CNN–SVM model will outperform many classifiers is based on the fact that the hybrid system compensates the limits of the CNN and SVM classifiers by incorporating the merits of both classifiers. The SVM classifier aims to minimize the generalization errors on the unseen data with a fixed distribution for the training set, by using the Structural Risk Minimization principle. The separating hyperplane is a global optimum solution. It is calculated by solving the quadratic programming problem, and the margin area between two classes of training samples reaches its maximum. As a result, the generalization ability of SVM is maximized to enhance the classification accuracy of the hybrid model after replacing the output layer in the Alexnet-CNN.

The advantage of the Alexnet-CNN classifier is that it automatically extracts the salient features of the input image. The features are invariant at a certain degree to the shift and shape distortions of the input characters. This invariance occurs because CNN adopts the weight sharing technique on one feature map. On the contrary, the hand-designed feature extractor needs elaborately designed features or even applies different types of features to achieve the distortion invariance of the individual characters. CNN uses the receptive field concept successfully to obtain such local visual features. However, the hand-designed feature extraction methods ignore and lose such topology of the input in most cases. Therefore, the trainable features of CNN can be used instead of the hand-designed features to collect more representative and relevant information.

Table 1: Results of different classifiers on RMFD dataset.

Performance (%)	Alexnet-CNN	Hybrid CNN-SVM
Training	97.20%	**98.54%**
Testing	97.72%	**98.89 %**

In Fig. 8, the Parallel Coordinates Plot shows comparison of different features of the dataset on the Standard Deviation axis. Out of the 1000 features extracted from the dataset, the graph shows only the first 10 features which belong the two classes i.e., masked face and non-masked face. The horizontal lines intersect at different values of the features, and on the left it shows the amount of deviation by each image. This shows the upper and lower bounds of deviation stands between a range from -2 to 4 for the first 10 features extracted from image dataset. It also shows the maximum density of predicted labels is between -1 to 1. The few dotted blue and red lines show the data points which have failed to classify the images compared to their respective labels. Further if we observe carefully, we can see the deviation is more for the masked images, it means our hybrid model has identified the non-masked face easily, but it faced difficulty for identifying the masked faces, therefore the deviation has increased for the masked dataset.

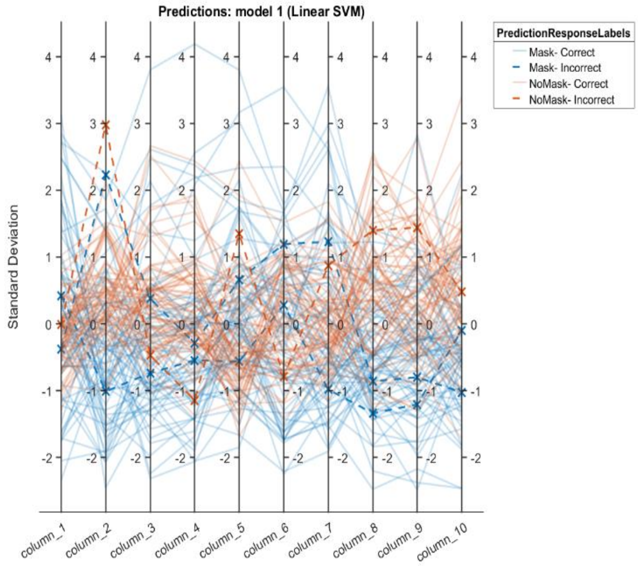

Fig. 9: Parallel Coordinated Plot for 10 features with Standard Deviation.

6. CONCLUSION AND FUTURE SCOPE

In this project, a hybrid CNN-SVM model has been designed to solve the masked and non-masked face recognition problem. The model took the Alexnet-CNN as an automatic feature extractor and it later the output was passed into SVM for the final output predictions. The efficiency and feasibility of the model were evaluated by the recognition accuracy. Experimental results on the RMFD dataset showed great benefit of the hybrid model. It achieved a loss of 1.45%.

Our results indicate that the hybrid model is quite a promising classification method for the masked faced recognition domain due to the three properties: (1) one is that the salient features can be automatically extracted by the hybrid model, while the success of most of other traditional classifiers relies largely on the retrieval of good masked and non-masked face features which is laborious and time-consuming task. (2) The second lies in that the hybrid model combines the advantages of SVM and Alexnet-CNN, as both are the most popular and successful classifiers in the image recognition field. (3) The third is that the complexity of the hybrid model in the decision process just increases a little bit when compared with other CNN classification models.

There are several research papers published on the CNN-SVM learning model. The performance of the hybrid model can be further improved through the fine tuning of its structure and its parameters. For example, improvements might be made based on the size of the input layer, the kernel functions used in the model, etc.

Extending the hybrid model to other applications is a task worth investigating. It can further study that how to improve the model for face detections over a mask, which means the model further can try to identify the faces with a mask, which will improve the security aspects in different public gatherings. Furthermore, we can try this model on peoples of different countries and cultures for a generalization.

REFERENCES

[1] M. Szarvas, A. Yoshizawa, M. Yamamoto, J. Ogata, Pedestrian detection with convolutional neural networks, in: Proceedings of the IEEE on Intelligent Vehicles Symposium, June 2005, pp. 224–229.

[2] K. Mori, M. Matsugu, T. Suzuki, Face recognition using SVM fed with intermediate output of CNN for face detection, in: Proceedings of the IAPR Conference on Machine Vision Applications, Tsukuba Science City, Japan, May 2005, pp. 410–413.

[3] F. Lauer, C.Y. Suen, G. Bloch, A trainable feature extractor for handwritten digit recognition, Pattern Recognition 40 (6) (June 2007) 1816–1824.

[4] Xiao-Xiao Niu, Ching Y. Suen. A novel hybrid CNN–SVM classifier for recognizing handwritten digits, Pattern Recognition 45 (2012) 1318–1325.

[5] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (2006) 1527–1554, https://doi.org/10.1162/neco.2006.18.7.1527.

[6] S. Bai, Growing random forest on deep convolutional neural networks for scene categorization, Expert Syst. Appl. 71 (2017) 279–287, https://doi.org/10.1016/j.eswa.2016.10.038.

[7] H. Goh, N. Thome, M. Cord, J.H. Lim, Learning Deep Hierarchical Visual Feature Coding, IEEE Trans. Neural Networks Learn. Syst. 25 (2014) 2212–2225, https://doi.org/10.1109/TNNLS.2014.2307532.

[8] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (2016) 1285–1298.

[9] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, H. Greenspan, Chest pathology detection using deep learning with non-medical training, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). Presented at the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015, pp. 294–297. 10.1109/ISBI.2015.7163871.

[10] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med. Imaging 35 (2016) 1299–1312, https://doi.org/10.1109/TMI.2016.2535302.

[11] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in NIPS, 2012.

[13] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[14] M. Ranzato, C. Poultney, S. Chopra, Y. LeCun, Efficient learning of sparse representations with an energy-based model, in: John Platt Bernhard Sch¨ olkopf, Thomas Hofmann (Eds.), Advances in Neural Information Processing Systems, MIT Press, 2006, pp. 1134–1144.

[15] P.Y. Simard, D. Steinkraus, J.C. Platt, Best practice for convolutional neural networks applied to visual document analysis, in: Proceedings of the International Conference on Document Analysis and Recognition, Edinburgh, Scotland, 2, 2003, pp. 958–962.

[16] K. Gopalakrishnan, Siddhartha K. Khaitan, Alok Choudhury, Ankit Agrawal, Deep Convolution Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection, September 18, 2017, pp. 322–930.