Masked Face Recognition

A DISSERTATION

submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

by

Raushan Raj	(IIT2018031)
Bindu	(IIT2018105)
Ayushi Gupta	(IIT2018118)
Sanjana	(IIT2018120)

Under the supervision of

Dr. SHIV RAM DUBEY & Dr. SATISH KUMAR SINGH

Assistant Professor

DEPARTMENT OF INFORMATION TECHNOLOGY

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

ALLAHABAD - 211015, (UP) INDIA

DECLARATION

We hereby certify that this project report entitled "**Masked Face Recognition**" is submitted in partial fulfilment of the requirement of the Degree of Bachelor of Technology in Department of Information Technology, Indian Institute of Information Technology, Allahabad.

We have given due credit to the original authors/ sources through proper citation for all the words, ideas, diagrams, graphics, computer programs, experiments, results, websites, that are not my original contribution. We have used quotation marks to identify verbatim sentences and given credit to the original authors/sources.

We affirm that no portion of our work is plagiarized. In the event of a complaint of plagiarism, we shall be fully responsible. We understand that our Supervisor may not be in a position to verify that this work is not plagiarized.

Place : Allahabad, Prayagraj		Raushan Raj	(IIT2018031)
Date : 11-05-2021		Bindu	(IIT2018105)
		Ayushi Gupta	(IIT2018118)
		Sanjana	(IIT2018120)

CERTIFICATE

It is certified that the work contained in the project titled "**Masked Face Recognition**" has been carried out under my supervision and that this work has not been submitted elsewhere.

	Signature of Supervisor	Signature of Supervisor
Date: 11-05-2021	Dr.Satish Kumar Singh Assistant Professor	Dr Shiv Ram Dubey Assistant Professor

ACKNOWLEDGEMENT

We would like to express our gratitude to the mercy of Almighty, that gave us enough strength to complete this project work. We are very thankful to our supervisor **Dr. Shiv Ram Dubey,** Assistant Professor, Department of Information technology, IIIT-Allahabad to guide us throughout the project work and always motivated us to work hard. The regular discussions and extensive comments helped us to improve the project and a direct impact can be seen on the project.

We are thankful to **Dr. Satish Kumar Singh**, Assistant Professor, IIIT-Allahabad, for providing valuable guidance at the beginning of the project. We are also grateful to all the faculty members and research scholars who have been very supportive in providing resources to complete this project.

At the end, we wish to thank all our friends who have always been a great support to us and a source of motivation to work forward.

Raushan Raj (IIT2018031)
Bindu (IIT2018105)
Ayushi Gupta (IIT2018118)
Sanjana (IIT2018120)

ABSTRACT

In this technological era, artificial intelligence has become the new powerhouse of data analysis. With the advent of various machine learning and computer vision algorithms, their application in data analysis has become a general trend. However, the application of deep neural networks to various tasks of analyzing masked face data and studying the performance of these models has yet to be explored to a great extent. So in this article we have proposed a model, which has been trained as such when you feed an image as input , our model will be in a position to recognize that and will print the name of the person. Our proposed models achieved fairly high precision with a low cross entropy rate.

Contents

Motivation

Every domain is becoming more automated as the number of Artificial Intelligence applications grows. Machine learning has allowed systems to learn the process on their own, reducing the need for human intervention. While AI has not been widely used in masked face recognition, there are some places where it can be extremely beneficial. We are trying to develop a deep learning model for recognition of face under mask. The main purpose of performing such activities is to teach the model to train in order to recognize the face under a mask , which can help cops to catch the thief or any central agencies like **CBI, IB** to catch the infiltrators / spy.

1 Problem Statements

(i) Develop a SoftMax based convolutional neural network for classification of cricket shots.
(ii) Develop a SVM based convolutional neural network for classification of cricket shots.
(iii) Compare the performance of both models using different metrics.

2 Introduction

Cricket is a popular sport in many countries across the world. Batting is a part of the game, which is the art of hitting the ball bowled by bowler in order to score runs. In recent years, There is a revolution of Technology in the field of Cricket also. With the growth in the application of Artificial Intelligence, Machine Learning has enabled the system to learn the process on their own in order to reduce human labour. Batting is one of the most complicated activities in Cricket, where each batsman plays a different kind of shots according to the cricket pitch. So, we can apply AI techniques in Cricket, which can be of great help to broadcasters in doing automated commentary in near future. It will make their job easier. Therefore, we have developed a Deep Learning model to detect which kind of shot batsman has played . Our model will classify a cricket shot through an image. We use feature extraction for dimensionality reduction. So, we have proposed a SoftMax based CNN and a SVM based CNN model. Later we have compared these two models based on their performance and accuracy. In this paper, we have 4200 images , out of which we have 700 images per shot and out of these 700 images , 600 images are augmented and 100 images are original. Currently, the goal of our developed model is to classify 6 types of cricket shots. On our dataset , the developed model will detect the type of cricket shots and apart from that it will also recognize the similarities and dis-similarities between different cricket shots. The

2nd section of this paper elaborates the literature review and about the related work that has been done in this field. Further, 3rd section deals with the proposed methodology, the dataset and the background of CNN.

3 Literature Review

Paper Title : Implementation of principal component analysis on masked and non - masked face recognition

Objective : The main objective of this paper is to represent an implementation of Principal component analysis (PCA) on masked and non masked.

Methodology : During the training phase , Viola - Jones algorithm was used and then PCA was used to effectively represent the gallery of images. PCA feature space is obtained which is called which eigen faces. In the test phase the targeted face i.e masked and non - masked is recognized.The nearest neighbour(NN) classifier distance is taken for recognition.

Conclusion : The masked face recognition achieved an accuracy of 72% while non-masked face recognition achieved an accuracy of of 95%.

*Masked face recognition had a 72 percent accuracy rate, whereas non-masked face identification had a 95 percent accuracy rate.

Paper Title : Efficient masked face recognition method during the COVID - 19 pandemic

Objective : The main objective of this paper is to do two different tasks - face mask recognition and masked face recognition.The first one checks whether the person is wearing a masks or no.Face recognition with a mask, on the other hand, focuses on the eyes and forehead areas to recognise a face with a mask. In this research, they used a deep learning-based strategy to solve the second problem. To extract features from the unmasked facial areas, they employ a pre-trained deep learning-based model (out of the mask region).

Methodology : Firstly, the preprocessing is done , cropping filter is applied to extract only the non - masked region. Secondly , there is a feature extraction layer , in which deep features are extracted using VGG - 16 face CNN descriptor from the 2 D images.Thirdly , there is a deep bag feature layer.

Conclusion : The proposed method achieved the high recognition performance on masked face recognition .

*On masked face recognition, the proposed technique achieved a high recognition performance.

Paper Title : Face Mask Recognition System with YOLOV5 Based on Image Recognition

Objective : This study recommends that manual inspection be replaced with a deep learning method, and that YOLOV5, the most powerful objection detection algorithm currently available, be used to better apply it in the real world, particularly in the supervision of people wearing masks in public spaces.

Methodology : The method was divided into 4 parts :facial mask image enhancement, facial mask image segmentation, facial mask image recognition, and interface interaction.

Conclusion : In this paper YOLOV5 is applied to identify whether the face mask is worn or not. Through testing , this experiment had a success rate of 97.9%.

Paper Title : Face recognition in the scene of wearing mask

Objective : The main objective of the paper is to use model training to improve recognition of critical areas like the eyes and brows. The LBP (Local Binary Pattern) feature efficiently depicts the texture of face photos and can improve the robustness of face recognition algorithms.

Method : To detect face and mask occlusion zones, they have used the upgraded MTCNN. They extracted the LBP features of the non occlusion area after detecting the mask occlusion area, then fed the extracted features into the support vector machine to recognise the face. The non-occluded area of the facial image is used to recognise the face.

Conclusion : The classifier used to recognise the face determines the performance of any facial recognition system.It will increase the performance of the face recognition system by removing the occlusion area's interference.

Paper Title : Masked face recognition with latent part detection.

Objective : The main objective of this paper is to use latent part detection (LPD) model on Masked Face dataset for verification (MFV) and Masked Face dataset for Identification Dataset(MFI) is suggested to locate the latent facial part which is robust to mask wearing.

Methodology : This paper collected two datasets for masked face recognition MFV and MFI. It proposed a latent part detection method to extract discriminative facial features which in turn is fed to a CNN model for facial recognition.

*It suggested a method for extracting discriminative face characteristics via latent component detection, which is then input into a CNN model for facial recognition.

Conclusion : The proposed model in this paper achieved an accuracy of about 94%.

Paper Title : MFCosface- A Masked-Face Recognition Algorithm Based on Large Margin Cosine Loss

Objective : The main objective is to have a system of contactless recognition which is not possible in fingerprint and ID cards.An algorithm is proposed in the paper based on large margin cosine loss to have high recognition accuracy.

Methodology : In this research, A high-recognition-accuracy masked-face recognition algorithm was proposed by the author. They employed the recognition of essential facial traits to cover face photos with common mask templates to build related datasets to address the problem of insufficient masked-face images. They introduced an attention method to make the model focus on effective regions to extract more relevant feature information because the mask eliminates some of the face feature information.

Conclusion : After performing the experiment we concluded that the algorithm is giving high accuracy and can extend on a dataset having occluded objects like glasses or scarves.

Paper Title : Cropping and attention based approach for masked face recognition

Objective : In this paper attention-based approach and cropping-based approach is proposed for face recognition.

Methodology : Two ways were suggested in this study to address the mentioned issues: An attention-based approach and a cropping-based approach. In an attention-based method, the CBAM attention module is used to focus on the area around the eyes, which outperforms the other attention modules on masked face recognition.

Conclusion : The integration of the optimal cropping and CBAM module achieves the best recognition accuracy for Masked Face Recognition.

4 Proposed Methodology

A. DATASET

We will use the ShotNet [11] dataset to train and evaluate our model. It contains 3600 images with 6 classes of cricket shots, namely Cut shot, Cover drive, Straight drive, Pull shot, Leg glance and Scoop shot, with 600 images in each class. The authors collected 100 images per class and later augmented it using five different techniques to expand the dataset to 600 images per class.

Figure 1: Some example images

The augmentation techniques used were Rotation by $\pm 30°$, Shearing, Addition of salt and pepper noise and shading.

B. BACKGROUND OF CNN

Deep learning makes use of artificial neural networks. Neural networks work just like our brains. Convolutional Neural Networks (CNNs) are one of the most efficient deep learning networks. This is an artificial neural network also known as ANN feedforward. Knowledge circulates across networks in a "feed-forward" network. CNN functions similarly to a biological visual cortex. CNN is one of the most common image classification models. CNN outperforms all other image classification algorithm in terms of classification accuracy. We don't have to pick features in CNN, but we do in other image classification algorithms. Different types of layers are used in CNN. A filter or moving centre runs through the picture in the convolutional layer. It usually occupies a specific part of a 2D matrix (image representation), applies point multiplication, and stores the result in another matrix.

Convolution is represented by following mathematical formula,
(the size of filter is $(2a + 1)\times(2b + 1)$)

$$h(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} F(s, t) I(x + s, y + t) \tag{1}$$

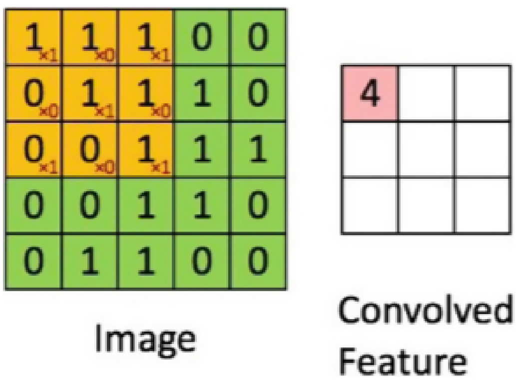

Figure 2: Convolution of an image with a filter

The dimension of the output matrix can be calculated by an equation. We can see an equation below where:

O – Output dimension
N – Input dimension
F – Window size
S – Stride
P – Padding

$$O = \lfloor \frac{N-F+2P}{S} \rfloor + 1 \quad (2)$$

The pooling layer is generally adjacent to the convolution layer. It was mainly used for memory reduction and quick calculation. Decreases the volume. Max pooling is one of the most used levels by CNN. Set up a kernel and find the maximum number from the array.

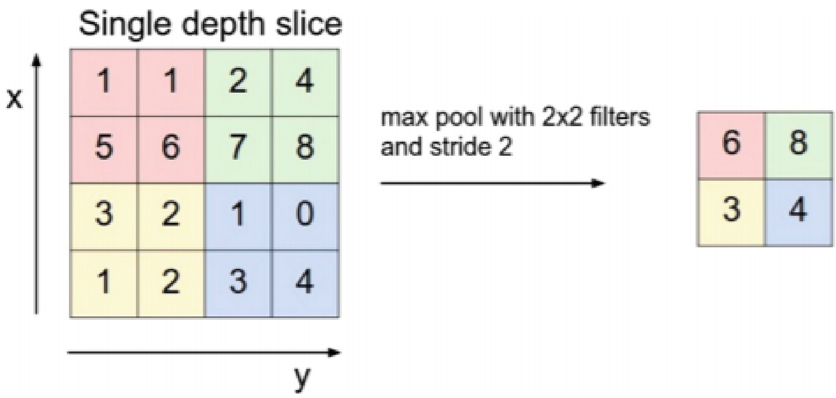

Figure 3: Max pooling with 2×2 filter with stride 2

A fully connected layer receives a 2D or 3D array from the previous layer as input and converts it to a 1D array. The probability of the classes is shown in the output layer of a convolutional neural network. The "SoftMax" function is used to measure it. The probability calculation equation is given below.

$$\sigma(X_i) = \frac{e^{X_i}}{\sum_j e^{X_j}} \quad (3)$$

C. BACKGROUND OF SVM

i. Support Vector Machines:

Given training data and its corresponding labels (x_n, y_n), $n = 1, ..., N$, $x_n \in R^D$, $t_n \in \{-1, +1\}$, SVMs learning consists of the following constrained optimization:

$$\frac{1}{2}w^T w + C\sum_{n=1}^{N} \xi_n \quad (4)$$

$$s.t.\ w^T x_n t_n \geq 1 - \xi_n \ \forall n$$

$$\xi_n \geq 0 \ \forall n$$

ξ_n are slack variables which penalize data points which violate the margin requirements. The corresponding unconstrained optimization problem is:

$$\frac{1}{2}w^T w + C\sum_{n=1}^{N} max(1 - w^T x_n t_n, 0) \quad (5)$$

L1-SVM is not differentiable. So, we have the need of a differentiable SVM, which is L2-SVM. L2-SVM minimizes the squared hinge loss:

$$\frac{1}{2}w^T w + C\sum_{n=1}^{N} \left(1 - w^T x_n t_n, 0\right)^2 \quad (6)$$

To predict the class label of a test data x:

$$\left(w^T x\right)t \quad (7)$$

ii. Multiclass SVM

K linear SVMs will be trained separately for k class problems, with data from the other classes forming the negative cases. The output of the k^{th} SVM is

$$a_k(x) = w^T x \quad (8)$$

The predicted class is:

$$a_k(x) \quad (9)$$

SVMs essentially seek out the maximum margin between data points from various classes., while SoftMax layer maximizes the likelihood or minimizes cross-entropy.

The feature extraction can be done using the convolutional layers which will then be used as the input to SVM. Transfer learning can also be used where some pretrained models can be used and SVM added after removing the output layer.

D. Loss Functions

1. Categorical cross entropy for CNN-softmax model

$$L_{CE} = -\sum_{i=1}^{n} t_i log(p_i) \text{for } n \text{ classes}$$

where t_i is the truth label and p_i is the softmax probability for the i^{th} class.

2. Square hinge loss for CNN-SVM model

$$L(y, y') = \sum_{i=0}^{n} (max(0, 1 - y_i \cdot y_i')^2)$$

where ŷ the predicted value and y is either 1 or -1.

D. Proposed Model

We came up with our own CNN model which is a slight modification of the model proposed by Md Ferdouse Ahmed Foysal et al. [11]. Our model has 19 layers. We named our model as Shotnet-Pro. There are five convolutional layers:

- First layer has 16 - 3 × 3 filters and 'relu' as activation function.
- Second layer has 32 - 3 × 3 filters and 'relu' as activation function.
- Third layer has 64 - 3 × 3 filters and 'relu' as activation function.
- Fourth layer has 128 - 3 × 3 filters and 'relu' as activation function.
- Fifth layer has 256 - 3 × 3 filters and 'relu' as activation function.

In addition, there are five max-pooling layers, each of size 2 × 2, each after every convolution layer. There are six dropout layers with parameter 0.20. We have a flatten layer, in the model. Lastly, there are two dense layers, where in one we used 'relu' as activation function and in the other we used 'softmax' as activation function. We used softmax activation function to get the probability of each class.

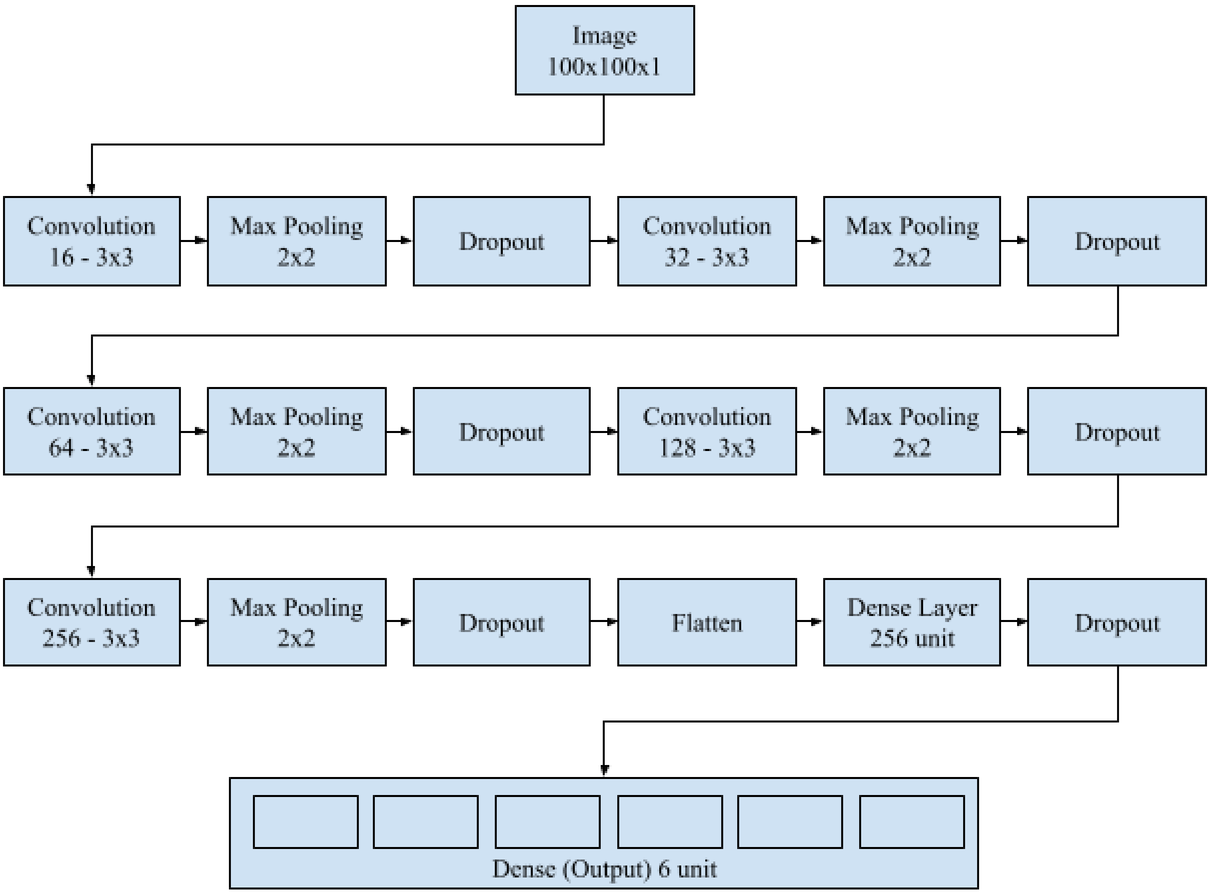

Figure 4: Architecture of our model

For the CNN-SVM model, the last layer "linear" activation function is used is replaced by a L2-SVM (Squared-Hinge Loss).

D. Training the Model

For both the models, we have used Adam optimizer to compile our model. We have used 80% of the dataset for training and 20% for testing. The training dataset has 3360 images with 540 images per shot. The test dataset has 840 images with 140 images per shot. We have used 5-Fold cross validation for evaluating the performance. In 5-Fold cross validation the dataset is split into five parts, the model is trained on four parts, and validated on the remaining one. This process is repeated five times with every part used as validation set once. We have trained the network for 40 epochs in every fold.

E. Evaluation of the Model

The results of one fold training are shown below

Training Phase

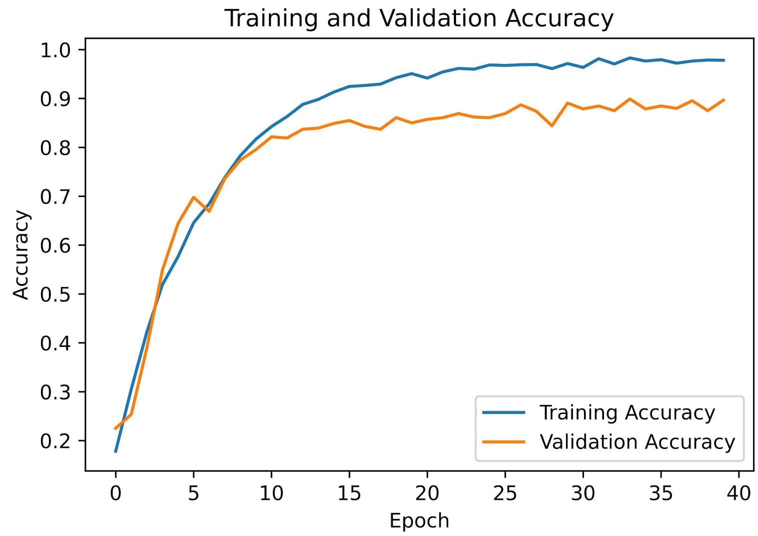

Figure 5: Training and validation accuracy of CNN-softmax model

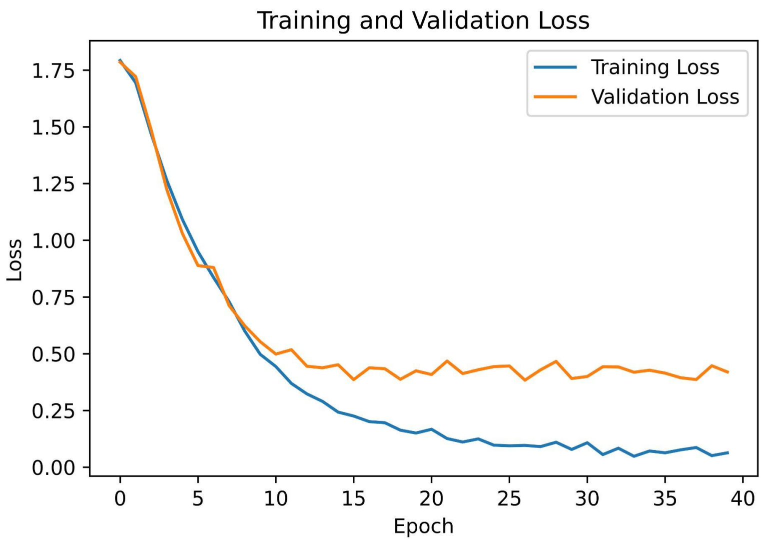

Figure 6: Training and validation loss plot of CNN-softmax model

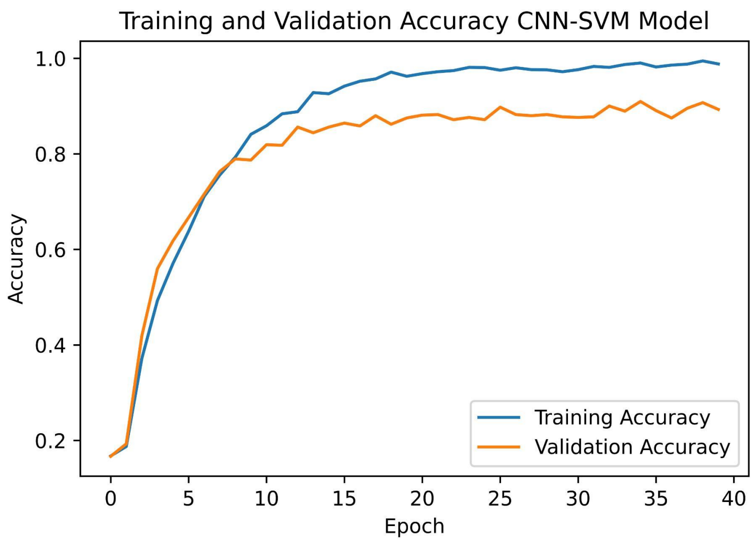

Figure 7: Training and validation accuracy of CNN-SVM model

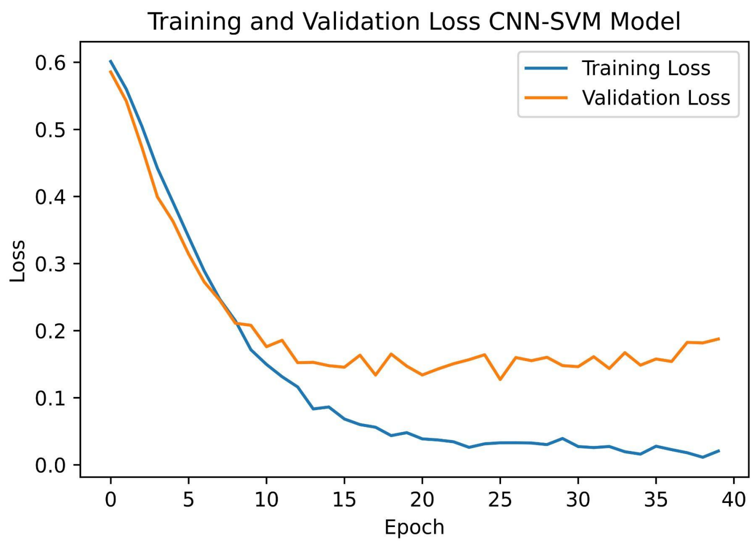

Figure 8: Training and validation loss of CNN-SVM model

Test Phase

We evaluated the trained model on the following metrics: precession, F1-score and recall. The following classification reports were obtained for both the models along with the classification report of the shotnet model.

Table 1: Classification report of CNN-Softmax Model

5 Result Discussion

6 Future Work

We may distinguish various forms of cricket strokes using our proposed methods. Convolutional neural networks were used to model our results. Our long-term aim is to develop a more accurate neural network. We can upgrade this network to perform image classification on RGB images. The model can also be modified to perform classification on video dataset. We intend to perform classification based on 3D depth images using deep learning. We can use different types of algorithms choosing which one is efficient.

7 Conclusion

In this paper, we present a method for recognizing masked faces . We used CNN to get this work done. We used the RESNET model (RESNET34) . The final achieved result we've found from the mode; is so promising. We hope that in the future, these approaches can be adapted into real-world applications, which will help cops and other central agencies in investigation. It will play a crucial role.

● **Timeline**

25 Jan – 7 Feb: Read research papers on basics of CNN and use of CNN in the sports field.
8 Feb – 21 Feb: Read research papers on CNN involving SVM.
3 Mar – 21 Mar: Building SoftMax based model.
22 Mar – 11 Apr: Building SVM based model.
12 Apr – 20 Apr: Compare the models and write the final report.

8 References

[1] Md. Sabbir Ejaz, Md Rabiul Islam, Ananya Sarker, Md Sifatullah "Implementation of principal component analysis on masked and non-masked face recognition."(ICASERT2019)

[2] Walid Hariri "Efficient Masked Face Recognition Method during the COVID-19 Pandemic"

[3] GuiLing Wu "Masked Face Recognition Algorithm for a Contactless delivery."

[4] Hongxia Deng, Zijian Feng, "A Masked-Face Recognition Algorithm Based on Large Margin Cosine Loss"

[5] Vibhaakar Sharma, Swathi Gangaraju, Vishal K. Sharma "Masked Face Recognition"

[6] Anwar, Arijit "Masked Face Recognition for Secure Authentication"