IN3060/INM460 Computer Vision Coursework report

- **Student name, ID and cohort:** Shreenika Aldur Krishnegowda (230016400) - PG
- **Google Drive folder:** https://drive.google.com/drive/folders/1ri-uleIVRbZPyXYcEnPRDWMPkCED3jLa?usp=drive_link

This study delves into applying and comparing Multilayer Perceptron, Support Vector Machine and Convolutional Neural Network on face mask detection dataset to observe and analyse which model architecture gives more accurate results. The Best trained model is then run on a test video to detect face covering in the video "in the wild". After careful analysis, it was observed that Convolutional Neural Network performs slightly better than Multilayer Perceptron and Support Vector Machine architecture.

Data

The image dataset provided for the coursework is a collection of images of people wearing face mask, not wearing them, and wearing them improperly. It consists of 2394 images for training and 458 images for testing of the human face. Each label is accompanied by its label in a txt file. The file contains labels as integers ranging from 0 – 2.
O: No mask is worn

1: Mask is worn

2: Mask is worn improperly

This distribution of data offers clarity for classification task paving for accurate identification of whether the image depicts wearing masks, not wearing masks, or wearing them improperly. For the Video testing operation a creative commons video was downloaded from YouTube which can be found in the below link: *https://www.youtube.com/watch?v=W_9jLju5FuQ&t=15s*

Implemented methods

Both Multilayer Perceptron and Support Vector Machine are traditional machine learning techniques and in this study, they both leverage Histogram of Oriented Gradients (HOG) for feature extraction. By computing histograms of gradient descent HOG provides a concise representation of local object appearance and shape, making it suitable for face mask detection using Multilayer Perceptron and Support Vector Machine [1]. For Convolutional Neural Networks, this study leverages Data Augmentation to apply transformations like rotation, scaling, and flipping to generate new training samples. It diversifies the training data to increase the models' performance on unseen variables [2].

Both Multilayer Perceptron and Support Vector Machine architectures use Histogram of Oriented Gradients feature extraction for pre-processing by checking if the input images are in grayscale or color. Multilayer Perceptron architecture is defined with multiple parameters such as hidden layers, activation functions, regularisation strength and learning rate schedule. Grid search iterates over all the combinations of hyperparameters and using cross-validation identifies the best hyperparameters. Then a new instance of Multilayer Perceptron classifier is initialised with the best parameters and is trained on the extracted HOG features. The final Multilayer Perceptron model is evaluated on both the training and validation sets for its accuracy.

Support Vector Machine architecture is trained on a set of hyperparameters such as regularisation, kernel type and kernel coefficient. Grid search is implemented to iterate over various combinations of hyperparameters and cross-validation is used to identify the best hyperparameters, and a new SVM classifier instance is created. The implementation of both Multilayer perceptron and Support Vector Machine primarily remains the same.

Convolutional Neural Network architecture in this study is defined with three convolutional layers which are used to extract features from input images. The first layer has 32 filters with (3,3) kernel size, to second layer has 64 filters and finally the third has 128 filters. The first layer input shape is (128, 128, 3). The architecture has three max pooling layers with each pooling layer having a pool size of (2,2). Further, the 3D feature maps are flattened into a 1D vector. Finally, the two fully connected layers of which the first dense layer consists of 128 neurons with ReLU activation function and dropout rate of 0.5. The second dense layer consists of 3 neurons which output the probability of no mask, mask worn correctly, mask work improperly.

Results

The final evaluation of all three models clearly indicates that the Convolutional Neural Network performs slightly better then Multilayer Perceptron and Support Vector Machine.

Multilayer Perceptron: *The model achieved an overall training accuracy of 84% and a validation accuracy of 85%. Class-wise results vary significantly, however, with No mask and Mask worn showing higher and more reasonable performance while Mask worn improperly performs poorly. Overall, the Multilayer perceptron model performs decently but struggles to detect when masks are worn improperly.*

Support Vector Machine: *The model achieves an accuracy of 100% in training accuracy and gains 90% in Validation accuracy. This is a clear indication that Support Vector Machine undergoes overfitting, basically, the disparity between the training and validation accuracies clearly suggests that the trained SVM model has memorized the training set rather than learning the underlying patterns. The support Vector Machine model fails to generalize well to the new data.*

Convolutional Neural Network: *Convolutional Neural Network model achieves superior functions than all the other models in this study with an impressive 93% accuracy and a validation accuracy of 94%. Convolutional Neural Network performs exceptionally well for detecting No mask and Mask is worn however, it struggles with indicating Mask is worn improperly class.*

Table 1 - Class-Wise Performance:

Model	***Class***	***Training Accuracy***	***Validation Accuracy***
MLP	*0 (No mask)*	*37%*	*31%*
	1 (Mask worn)	*>97%*	*>97%*
	2 (Mask worn improperly)	*0%*	*0%*
SVM	*0 (No mask)*	*100%*	*60%*
	1 (Mask worn)	*>98%*	*>98%*
	2 (Mask worn improperly)	*0%*	*12%*
CNN	*0 (No mask)*	*96%*	*98%*
	1 (Mask worn)	*>97%*	*>97%*
	2 (Mask worn improperly)	*0%*	*0%*

MaskDetection(Path_to_testset, Model_type):

All three trained models run as expected with high precision obtained by CNN model and SVM providing the lowest precision. However, all the models yet fail to identify Mask is worn improperly class.

MaskVideoDetection(Video_path, Model_Type):

The best model which is the CNN is implemented to perform predictions and the code divides the video into 50 random frames and then performs face mask detection prediction which gives accurate predictions on the trained model.

Conclusion: *The Multilayer perceptron model demonstrates decent performance generalisation of new data but struggled to determine the Mask worn improperly class. Support Vector Machine on the other hand displayed overfitting with higher discrepancies between the training and validation accuracies. Finally, the Convolutional Neural Network which displayed superior generalisation ability achieved higher accuracy scores in both training and validation sets but faced challenges with Masks worn improperly in class. Multilayer perceptron and Support Vector Machine should be enhanced with more hyperparameters to avoid overfitting and the Convolutional Neural Network should be incorporated with multiple hyperparameters and different optimization techniques to improve its weak points at Mask worn improperly class. Further, increasing the number of samples in the Mask worn improperly could result in higher accuracy.*

References

1. *Histograms of Oriented Gradients for Human Detection by Navneet Dalal and Bill Triggs (2005)- https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf*

2. *Convolutional Neural Networks for Image Recognition by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton (2012)- https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks*

3. *Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford University Press- https://people.sabanciuniv.edu/berrin/cs512/lectures/Book-Bishop-Neural%20Networks%20for%20Pattern%20Recognition.pdf*

4. *Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297- https://www.scirp.org/reference/referencespapers?referenceid=1150668*