

Face Mask Detection

https://drive.google.com/drive/folders/13CX--tdeNprhpO56vpWZIRoP7jtkkfWL?usp=share_link

1- Data

The goal of this project is to develop a Face covering Detection (FCD) computer vision system. A ready to use dataset of cropped photographs of faces has been given to us to help achieve our task, each image has been categorised into one of three classes: appropriately wearing a face mask, not wearing a face mask, or wearing a face mask incorrectly. The dataset includes testing, validation, and training sets.

The collection's photos show people in a broad range of poses, with various lighting situations and having different face expressions. The images have been marked in order to ensure that the training and testing methods are accurate. The data quality is high and it provides a realistic representation of situations where FCD may be applied in the actual world.

Overall, the dataset and the data preparation provide the foundation for developing accurate and reliable models. The large diversity of images in the dataset will help the models generalise well to a variety of scenarios, and the thorough labelling process ensures that the models are trained on valid data.

2- Implemented Methods

Machine learning models are required to detect human faces in various positions and lighting conditions as well as detect the presence and location of face coverings in order to do this task. In this study, CNN+HOG, SVM+HOG, MLP+HOG, and MLP+LBP were four machine learning combination that were compared based on their accuracy, precision, recall, inference time, and model size. These models were chosen due to their different strengths and weaknesses such as the size difference.

HOG and LBP features were extracted from the original images to create a more condensed representation of the images. While the HOG features record information on gradient orientation, the LBP features encode regional texture patterns. The Standard Scaler was employed to standardise the features in order to get a comparable distribution. In this phase, the mean and standard deviation of the features were subtracted, giving features with a mean of 0 and a variance of 1.

3- Training Process

For each model, the following steps were taken:

- Data pre-processing: We normalized the images, and HOG or LBP features were appropriately extracted. We applied some additional pre-processing, such as sharpening the images using a convolution filter to improve details and equalizing histograms to add image contrast.
- Data Splitting: The dataset was divided into training and validation sets using an 80-20 split ratio. This allowed for separate sets of data to be used for model training and validation, ensuring an efficient evaluation of the model's performance.
- Label Encoding: Labels were converted to one-hot encoding to ease the use of categorical cross-entropy as a loss function during model training.
- Model training: Each model was trained on the corresponding training dataset, with hyperparameters optimized using like grid search, Bayesian optimization. During the training process, real-time data augmentation was employed using the Image Data Generator, which generated transformed images on-the-fly to improve the diversity of the training data.

- Model validation: The models were validated using a separate validation dataset, and performance metrics such as accuracy, precision, and recall were used to assess the models.

4- Results

4.1 Qualitative Examples

We tested our model on images and videos to evaluate its real-world performance. The models should be able to detect face masks in different settings, such as indoor and outdoor environments, and varying lighting conditions.

4.1.1 Images:

We tested the models on a set of 4 random images, which included individuals wearing masks correctly, incorrectly, and not wearing masks at all. In most cases, the models were able to successfully identify the presence and position of face masks, even under challenging lighting conditions. These results indicate the models' effectiveness and potential applicability in real-world situations. Nevertheless, some refinements and improvements are still necessary to achieve optimal performance across all scenarios.



Figure 1 SVM+HOG performance on 4 random images.

4.1.2 Videos:

We further evaluated the performances of all four models on video files, each model was required to process a continuous stream of frames in real-time. Overall, the models were able to maintain a consistent level of accuracy throughout the video, accurately detecting face masks despite changes in subject positions, lighting, and background. However, it is worth noting that some false positive cases were still observed in certain frames. While the general performance of the models is promising, further fine-tuning and optimization may be crucial for deployment in real-world applications where minimizing false positives is essential.



Figure 2 Example of an output frame using the CNN+HOG combination.

4.2 Quantitative Results

The table below shows the performances of the 4 models on the test set:

Model	Accuracy	Precision	Recall	Inference Time	Model Size
CNN+HOG	85.15%	61.70%	34.64%	1.2629s	130.18 MB
SVM+HOG	82.10%	57.43%	56.01%	0.2232s	9.62 MB

<i>MLP HOG</i>	83.62%	79.76%	52.86%	0.0086s	2.50 MB
<i>MLP LBP</i>	83.62%	42.15%	40.28%	0.0020s	1.15 MB

Figure 3 Table of evaluation metrics results on the test set

5- Discussion

The CNN+HOG model achieved the highest accuracy of 85.15%. This model combined the strengths of a deep Convolutional Neural Network (CNN) and the Histogram of Oriented Gradients (HOG) features. However, it had a much larger model size and longer inference time compared to the other models, which may could be a red signal, giving our use case, inference speed is very important and needs to be considered.

The MLP+LBP model provided a good balance between accuracy, speed, and model size, making it a suitable choice for real-time applications where speed and model size are primary concerns. This model used a simple Multi-Layer Perceptron (MLP) architecture with Local Binary Patterns (LBP) features, which are computationally efficient and invariant to lighting changes.

The SVM+HOG and MLP+HOG models achieved comparable accuracy scores, but the MLP+HOG model had the advantage of higher precision and lower inference time. The SVM+HOG model used a Support Vector Machine (SVM) with HOG features, which is a popular combination for object detection tasks. However, it had a longer inference time compared to the MLP-based models.

Overall, the CNN+HOG model delivered the best performance in terms of accuracy, making it the best natural choice for the face mask detection system. However, the MLP+LBP model can be more recommended as speed and model size are important.

This project contributes to the growing body of research on face mask detection using deep learning and computer vision techniques in the context of the COVID-19 pandemic. Our findings are consistent with previous studies that have demonstrated the effectiveness of CNN-based models for this task [1, 2]. In particular, our results show that combining CNNs with HOG features can achieve high accuracy, while MLPs with LBP features can strike a good balance between accuracy, speed, and model size.

6- Conclusion

Based on the results of this project, it is clear that face mask detection using machine learning is a complex task that requires important consideration of model selection. The four models evaluated in this project, including CNN+HOG, SVM+HOG, MLP HOG, and MLP LBP, each had their strengths and weaknesses in terms of accuracy, speed, and model size.

The CNN+HOG model achieved the highest accuracy and is therefore the most suitable model for applications that prioritize accuracy. However, it is also important to consider the inference time and model size, which may be limiting factors for some real-world scenarios. In contrast, the MLP+LBP model provided a good balance between accuracy, speed, and model size, making it a more practical choice for real-time applications. These insights can inform the development of more efficient and accurate face mask detection systems, as well as help to promote compliance with health and safety guidelines in public spaces.

References:

- [1] Wang, L., Lin, Z., Wong, A., "Covid-19 Face Mask Detection Using Deep Learning and Computer Vision Techniques," Image Processing On Line, 10 (2020), pp. 252-262. <https://doi.org/10.5201/ipol.2020.288>
- [2] Elaziz, M. A., Hosny, K. M., Salah, A., Darwish, M. M., Lu, S., & Sahlol, A. T. (2021). Deep Learning for Face Mask Detection in the Era of COVID-19 Pandemic: A Survey. Computers, Materials & Continua, 67(2), 2449-2468. <https://doi.org/10.32604/cmc.2021.014532>