Deep Learning Based-Face Mask Detection

Chilukuri Naga Vardhan   
*School of Computer Science and Engineering*  
*VIT-AP University*Amaravati, Andhrapradesh

*Abstract*—— Facial recognition serves numerous purposes, encompassing surveillance, identity verification for login systems, and personalized user experiences. However, its effectiveness is challenged when faced with non-frontal face orientations or when individuals wear accessories that obscure facial features. Even traditional detection systems reliant on facial characteristics encounter difficulties in maintaining high accuracy under these circumstances A pivotal aspect of the system involves utilizing the final feature map to accurately detect faces of varying sizes. This multi-layer detection approach enhances the system's capacity to discern facial attributes amid diverse conditions and orientations.

Keywords— Face detection, CPU, Multiple layer, Deep learning

# **Introduction**

Facial recognition stands as a computer-based AI technology utilized to detect and identify human faces within images, commonly referred to as face recognition. This technology holds applicability across diverse sectors, encompassing biometrics, security, law enforcement, entertainment, and personal safety. Its capabilities encompass real-time monitoring and tracking of individuals. The most recent pandemic, COVID-19, stems from the newly identified coronavirus and has had a profound impact globally. This infectious ailment, caused by the SARS-CoV-2 virus, primarily targets the respiratory system and spreads via airborne transmission, particularly through close proximity. As a response to this outbreak, the World Health Organization (WHO) has advocated various preventive measures, including social distancing, disinfection, and the use of face masks. Wearing masks has proven effective in curbing the virus's dissemination among the populace, leading many countries to mandate their usage in public spaces.

However, manual mask compliance checks are time-intensive, especially in densely populated locales like hospitals, airports, train stations, and shopping malls. Consequently, the need for an automated mask detection system has spurred research efforts. Jignesh et al. employed a learning-to-switch network for automated mask recognition, while Loey et al. introduced an associative deep transfer learning model for the same purpose. In alignment with these endeavors, we have presented the VGG-16 and VGG-19 architectural models for accurate mask recognition .

1. VGG16 Architecture:

VGG16, short for Visual Geometry Group 16, was introduced by Visual Geometry Group at Oxford University in 2014. The architecture consists of 16 layers and is famous for its simplicity and depth. VGG16 consists of a stack of convolutional layers followed by fully connected layers. The convolutional layers are configured with small 3x3 filters, while the maximum compositing operation is used to reduce the spatial size and gradually increase the receptive field. The final layers usually consist of fully connected layers with soft-max activation for multi-class classification tasks. Despite its simplicity, VGG16 has demonstrated outstanding performance in image recognition tasks and serves as a basic model for later architectures.

|  |
| --- |
| Image result for vgg16 architecture |
| Fig. 1. VGG16 Architecture |

1. VGG19 Architecture:

VGG19 is an extension of the VGG16 architecture, also developed by Visual Geometry Group in 2014. As the name suggests, VGG19 consists of 19 layers, making it deeper than its predecessor VGG16. Like VGG16, VGG19 follows a similar pattern with stacked convolutional layers interspersed with maximal clustering layers for spatial shrinking. The deeper architecture allows VGG19 to capture more complex functionality, which can help improve performance in some complex imaging tasks. However, this increased depth also leads to higher computational costs during training and inference.

|  |
| --- |
| Image result for vgg19 architecture |
| Fig. 2. VGG19 Architecture |

Application of VGG16 and VGG19 in Face Mask Detection:

To employ VGG16 and VGG19 for face mask detection, the models are typically fine-tuned on a dataset containing images of individuals with and without face masks. The fully connected layers of the pre-trained VGG16 and VGG19 models are replaced with custom layers suitable for the binary classification task (mask or no mask). Fine-tuning allows the models to adapt to the specific features relevant to face mask detection.

Pre-processing the images involves resizing them to a consistent input size, applying data augmentation techniques to increase the size of the training dataset, and normalizing pixel values for better convergence during training. The models are then trained using appropriate optimization algorithms and loss functions, such as binary cross-entropy, to minimize the classification error.

|  |
| --- |
| Fig. 3. Results of the proposed model |

1. **Motivation**

The driving force behind the release of the mask detection statement using VGG16 and VGG19 lies in the growing need for automated and effective solutions to monitor compliance with mask policy. page. Masks play an important role in reducing airborne disease transmission and protecting public health. Governments, businesses and various organizations recognize the importance of masks in protecting public safety and preventing the spread of infectious diseases.

Traditional manual monitoring of mask compliance in public spaces, workplaces, and other high-traffic areas is resource-intensive and often impractical, especially in regulated environments. There is an urgent need for automated systems that can accurately detect whether individuals are wearing masks. Automated mask detection systems offer a number of benefits, including real-time monitoring, scalability, and reduced burden on human resources.

Deep learning algorithms have become powerful tools in computer vision tasks, exhibiting outstanding performance in image classification and object detection. The VGG16 and VGG19 architectures, renowned for their efficiency and simplicity, have been instrumental in various image recognition applications. By taking advantage of the capabilities of these pre-trained models and adapting them to detect masks, we are able to create a powerful and effective solution to the mask compliance monitoring challenge.

The potential applications for an accurate and reliable mask detection system are vast. In public spaces, such a system could help enforce mask regulations, providing timely warnings to authorities or security personnel if non-compliance is detected. In workplaces and healthcare settings, this can help maintain a safe environment for staff, patients, and visitors.

In addition, mask detection by VGG16 and VGG19 is not limited to pandemic-related situations. The knowledge gained from this research can be adapted to other contexts where automated compliance monitoring is needed, such as industrial safety, security, and access control.

**(B). Contribution**

* Curating and preprocessing a comprehensive dataset for training and evaluation.
* Conducting fine-tuning and transfer learning to adapt the models to the specific task.
* Implementing and optimizing the face mask detection system.
* Comparing the performance of VGG16 and VGG19 in the mask detection task.

# Literature Review

The literature review aims to provide an overview of existing research on mask detection using deep learning techniques, with a particular focus on using the VGG16 and VGG19 architectures. In recent times, the need for automatic mask detection systems has become increasingly important due to public health concerns and safety measures. Deep learning algorithms, especially convolutional neural networks (CNNs), have shown great promise in various computer vision tasks, making them the natural choice for mask detection.

Several studies have demonstrated the effectiveness of deep learning algorithms in mask detection. CNN, which has the ability to automatically learn hierarchical features from images, has been widely used to design accurate and efficient mask detection models. CNN's ability to detect patterns and characteristics associated with masks has proven to be crucial in distinguishing people who wear masks from those who don't.

Transfer learning has become a popular approach in mask detection, where pre-trained CNN models, such as VGG16 and VGG19, are used. The researchers found that refining these models for the specific task of detecting masks would significantly improve their performance. By fine-tuning, models can leverage their prior knowledge from large-scale image datasets to detect mask-related characteristics.

Researchers have explored ensemble methods, which involve combining predictions from multiple models, to improve the performance and robustness of face mask detection systems. Additionally, attention mechanisms have been introduced to focus the model's attention on crucial regions of the face, further enhancing the accuracy of mask detection.

While existing research has made significant progress in face mask detection, challenges remain to be addressed. These include handling occlusions, variations in mask types, and addressing the imbalanced nature of mask vs. no mask class distribution. Future research could explore techniques to handle these challenges and extend the application of face mask detection to more complex and diverse scenarios.

# Face Mask Detection Using VGG16 and VGG19

* **Preprocessing**:

In the mask detection system using VGG16 and VGG19, the preprocessing step is an important step to ensure the effectiveness of the model. The preprocessing steps are as follows:

A diverse dataset of images of individuals wearing and not wearing masks was collected from a variety of sources. The dataset should include different backgrounds, lighting conditions, and variations of mask types.

The collected images are carefully sorted to eliminate any duplicate or unrelated samples that might interfere with the training process.

The image is scaled to a consistent input size in accordance with the requirements of the VGG16 and VGG19 architectures. Image size normalization facilitates smooth training and inference.

To increase the diversity of the training data set and improve the generality of the model, data enhancement techniques such as random rotation, flipping, and luminosity are applied.

The pixel values ​​of the image are normalized to bring them into a specific range, usually [0, 1] or [-1, 1]. Normalization helps to converge faster during training.

* **Proposed Methods Architecture:**

The proposed architecture for mask detection using VGG16 and VGG19 is built on top of the original architecture and customized to specifically solve the masked or unmasked binary classification task. The main components of the proposed architecture are:

The input layer accepts an image of a predefined input size, typically 224x224 or 128x128 pixels, depending on the requirements of VGG16 and VGG19.

A series of convolutional layers with small 3x3 filters are used to extract and learn hierarchical features from the input images. The number of filters and the depth of the convolutional layers can be adjusted depending on the complexity of the data set. ReLU (Rectifier Linear Unit) activation is applied after each convolution layer to introduce non-linearity and improve learning of complex representations of the model. Maximum clustering layers with 2x2 clustering windows are inserted to gradually reduce the spatial size and increase the receptive field, allowing the model to capture more global features. The last layers of the architecture are fully connected layers with soft-max activation for binary classification (masked or unmasked). These classes are responsible for generating output probabilities.

The proposed architecture, suitable for mask detection, leverages the strengths of VGG16 and VGG19 in image feature extraction. By customizing fully connected layers for binary classification, the models are trained to recognize and distinguish people wearing masks from those not wearing masks. VGG16 and VGG19 are well known for their deep and hierarchical structure, which allows them to learn rich and discriminant features. This is important in the context of mask detection, as masks can come in many different forms, including many different shapes, colors, and designs. The convolutional layers of the proposed architecture serve as a powerful feature extractor, capturing the necessary facial features and types related to the mask.

Fully connected final layers with softmax activation facilitate probabilistic interpretation, allowing the model to generate probabilities for each class (masked or unmasked). This provides a quantitative measure of the model's confidence in its predictions, which is important in real-world applications where accuracy and reliability are critical.

During the mask detection system training phase, several important steps are taken to optimize the VGG16 and VGG19 models for accurate mask detection. The selected data set is initially divided into separate subsets, including training, validation and test sets. This division allows the training set to be used to refine model parameters, while the validation set plays a central role in monitoring model performance and protecting against overfitting.

After splitting the dataset, the VGG16 and VGG19 models, pre-initialized with their respective pre-trained weights, are loaded. To adapt the models to mask detection, fully connected custom layers are introduced on top of the existing convolutional layers. This architectural adaptation allows models to learn relevant characteristics to distinguish mask-wearing and non-masking individuals.

Choosing a suitable loss function is important in quantifying the difference between the predicted and actual truth labels during training. For the binary classifier nature of mask detection, the loss function is usually chosen as binary cross entropy. At the same time, an optimizer such as Adam or SGD is used to iteratively adjust the model parameters based on the calculated slope, thereby guiding the model in the direction of minimizing the chosen loss.

The focus of the training phase consists of refining the VGG16 and VGG19 models using back-propagation. This iterative process refines the weights of the models, allowing them to adjust and capture complex features specific to mask detection. By learning from training data, the models gradually improve their ability to distinguish facial attributes that indicate whether individuals are wearing masks or not.

Hyperparameter tuning is performed to configure important parameters that significantly affect the training process. Parameters such as learning rate, batch size, and number of epochs are meticulously tuned for optimal performance. This iterative test aims to strike a balance between speed of convergence, generalization, and avoidance of overfitting, ensuring that models effectively capture and differentiate relevant features. mask. Basically, the training phase consists of a series of strategic actions, from data preparation and model initialization to architecture adaptation, loss function selection, optimization, fine-tuning, and tuning. hyperparameter tuning. Combining these steps leads to models capable of accurately detecting whether or not a mask is worn, contributing to the overall goal of enhancing public health and safety through automated mask detection.

# Dataset Description

The proposed model is trained with Face Mask Detection ~12K Images Dataset dataset that contains two classes those are with mask and without mask. Table I shows the with Face Mask Detection ~12K Images Dataset dataset description that contains ~12k labeled images that are divided into 992 test images, 10000 train images and 800 validation images []. The labels are encoded into two classes in the range from 0 to 1. we used python openCV for face detection where if a face is not covered with the mask it can create a red box, but if the face is covered with the mask it creates a green box with mask label.

**Table1.Dataset Description**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mask Detection Class** | **Train Images** | **Test Images** | **Validation Images** | **Total Images** |
| With Mask | 5000 | 483 | 400 | 5883 |
| Without Mask | 5000 | 509 | 400 | 5909 |
| **Total Images** | 10000 | 992 | 800 | 11792 |

|  |
| --- |
| Fig 4.Dataset Images with Labels |
|  |

# Results and Discussion

Both the model are trained on Kaggle using python script and Tensorflow for 25 epochs on VGG16 and just 20 epochs for VGG19. Adam optimizer and Batch size 32 are used for it. For VGG16, there are total 21,137,729 trainable parameter and there is no non-trainable parameter. For VGG19, There are total 20,040,770 parameter out of which 16,386 are trainable and remaining 20,024,384 non-trainable parameter.

## A) VGG16 results (Batch Size = 32 and epochs = 25)

Accuracy achieved by the model is 97.375% that is model predicted perfectly and classified the two categories of mask and without mask images with an error of just 2.625%. From the figure we can observe that in Y-axis the loss/accuracy graph is plotted when the model is training based on the epochs which is X-axis. At the initial stage when the model started to train at that stage training loss value will be high because back propagation will happen and updating of weights will be done by slowly increasing the number of epochs model gets better and better by gaining good accuracy and less training loss meanwhile value loss optimizes after each iteration and gives the result how poorly model is performed.

|  |
| --- |
|  |
| Fig5. Training Loss for VGG16 |
|  |
|  |
| Fig6.Training Accuracy for VGG16 |
|  |
| Fig7.Confusion Matrix |

*B) VGG19 results(Bacth Size = 32 and epochs =20)*

Accuracy achieved by this model is 99.5% and compared to vgg16 batch size=32 model here we obtained slightly more accuracy due to slightly deeper architecture of VGG19 when compared to VGG16. In VGG19 we have used the Haar Cascade Model detect faces in order to obtain the bounding box coordinates of faces in an image. Object Detection using Haar feature-based cascade classifiers is an effective object detection method proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. We have implemented the social distancing norms for this model and this can be done by iterating over the coordinates of faces and calculating the distance for each possible pair, if the distance for a particular pair is less than MIN\_DISTANCE then the bounding boxes for those faces are colored red. MIN\_DISTANCE must be manually initialized in such a way that it corresponds to the minimum allowable distance in real life (ex. 6ft in India).For this model we have given the minimum distance as 130.

|  |
| --- |
|  |
| Fig8. Bounding Box and Social Distancing Norms Checker |

|  |
| --- |
|  |
| Fig9.Training Loss of VGG19 |
|  |
| Fig10.Training Accuracy of VGG19 |
|  |
| Fig11.Results of the proposed model |

|  |  |  |
| --- | --- | --- |
| Model | Accuracy Rate | Parameters  × 101 million |
| VGG16 | 0.973 | 0.211 |
| VGG19 | 0.995 | 0.200 |

|  |
| --- |
| **Table 2. Tabular comparison of accuracies for both algorithms** |

# Conclusion and Future work

In conclusion, the application of VGG16 and VGG19 architectures in face mask detection has demonstrated their prowess in addressing the critical need for automated mask compliance monitoring. Through meticulous customization and fine-tuning, these deep convolutional neural networks have proven their efficacy in distinguishing between individuals wearing face masks and those without masks. The findings of our research underline the significance of leveraging pre-trained models and adapting them to specific tasks, highlighting the adaptability of VGG16 and VGG19 to the domain of mask detection.

The comparison between VGG16 and VGG19 has revealed intriguing insights into their respective capabilities. While both models exhibit commendable performance, VGG19's slightly deeper architecture appears to provide a subtle edge in accuracy, potentially due to its capacity to learn intricate facial features associated with mask presence. However, the choice between these architectures should be carefully considered based on factors such as computational efficiency and deployment constraints, as both models offer valuable contributions to the face mask detection landscape.

The success of VGG16 and VGG19 in mask detection high lights the potential of deep learning and transfer learning techniques to solve real-world challenges. By driving the automation of mask compliance monitoring, these models contribute to public health and safety efforts, especially in the context of infectious disease prevention and control. . As technology continues to advance, our research illustrates how machine learning can be harnessed to improve public welfare and highlights the importance of continuing innovation in the field of computer vision.

# References

[1] Jignesh, “Face Mask Detection Using Transfer Learning of InceptionV3”

[2] Hensel D. Jahja, and Novanto Yudistira “Mask usage recognition using vision transformer with transfer learning and data augmentation”

[3] Hensel Donato Jahja, Novanto Yudistira, Sutrisno “Mask Usage Recognition using Vision Transformer with Transfer Learning and Data Augmentation”

[4]https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset