Face Mask Detection by Using AlexNet Deep CNN

**1st Toky,Golam Shahriar (20-42743-1)**  
*Computer Science and Engineering*  
*American Internation University-Bangladesh.*  
Dhaka, Bangladesh

**2nd A.S.M Sayem (20-44115-2)**  
*Computer Science and Engineering*  
*American Internation University-Bangladesh.*  
Dhaka, Bangladesh

**3rd Akter, Suraiya (20-42360-1)***Computer Science and Engineering*  
*American Internation University-Bangladesh.*  
Dhaka, Bangladesh

**4nd Mursalin,Sawon (20-42680-1)**  
*Computer Science and Engineering*  
*American Internation University-Bangladesh.*  
Dhaka, Bangladesh

*Abstract*—COVID-19 epidemic has swiftly disrupted our day-to-day lives affecting the international trade and movements. Wearing a face mask to protect one's face has become the new normal. In the near future, many public service providers will expect the clients to wear masks appropriately to partake of their services. Therefore, face mask detection has become a critical duty to aid worldwide civilization. This paper provides a simple way to achieve this objective utilizing some fundamental [Machine Learning](https://www.sciencedirect.com/topics/computer-science/machine-learning) tools. The suggested technique successfully recognizes the face in the image or video and then determines whether or not it has a mask on it. As a surveillance job performer, it can also identify a face together with a mask in motion as well as in a video. The technique attains excellent accuracy. We investigate optimal parameter values for the Convolutional [Neural Network](https://www.sciencedirect.com/topics/neuroscience/neural-networks) model (CNN) in order to identify the existence of masks accurately without generating over-fitting.

Keywords—Machine Learning (ML), Deep Neural Learning (DL), Convolutional Neural Network (CNN), Alex Net Model.

# **1. Introduction**

The trend of wearing face masks in public is rising due to the COVID-19 coronavirus epidemic all over the world. Before Covid-19, People used to wear masks to protect their health from air pollution. While other people are self-conscious about their looks, they hide their emotions from the public by hiding their faces. Scientists proofed that wearing face masks works on impeding COVID-19 transmission. COVID-19 (known as coronavirus) is the latest epidemic virus that hit the human health in the last century. In 2020, the rapid spreading of COVID-19 has forced the World Health Organization to declare COVID-19 as a global pandemic. More than five million cases were infected by COVID-19 in less than 6 months across 188 countries. The virus spreads through close contact and in crowded and overcrowded areas.

The coronavirus epidemic has given rise to an extraordinary degree of worldwide scientific cooperation. Artificial Intelligence (AI) based on Machine learning and Deep Learning can help to fight Covid-19 in many ways. Machine learning allows researchers and clinicians evaluate vast quantities of data to forecast the distribution of COVID-19, to serve as an early warning mechanism for potential pandemics, and to classify vulnerable populations. The provision of healthcare needs funding for emerging technology such as artificial intelligence, [IoT](https://www.sciencedirect.com/topics/engineering/iot), big data and machine learning to tackle and predict new diseases.

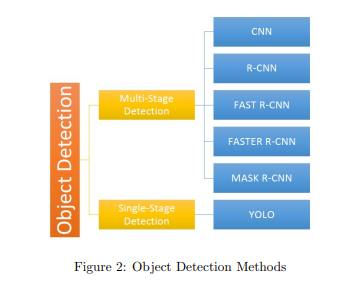
People are forced by laws to wear face masks in public in many countries. These rules and laws were developed as an action to the [exponential growth](https://www.sciencedirect.com/topics/engineering/exponential-growth) in cases and deaths in many areas. However, the process of monitoring large groups of people is becoming more difficult. The monitoring process involves the detection of anyone who is not wearing a face mask.

In this report, we introduce a mask face detection model that is based on deep transfer learning and classical machine learning classifiers. The proposed model can be integrated with surveillance cameras to impede the COVID-19 transmission by allowing the detection of people who are not wearing face masks. The model is integration between deep transfer learning and classical [machine learning algorithms](https://www.sciencedirect.com/topics/engineering/machine-learning-algorithm). We have used deep transfer leering for feature extractions and combined it with three classical machine learning algorithms. We introduced a comparison between them to find the most suitable algorithm that achieved the highest accuracy and consumed the least time in the process of training and detection.

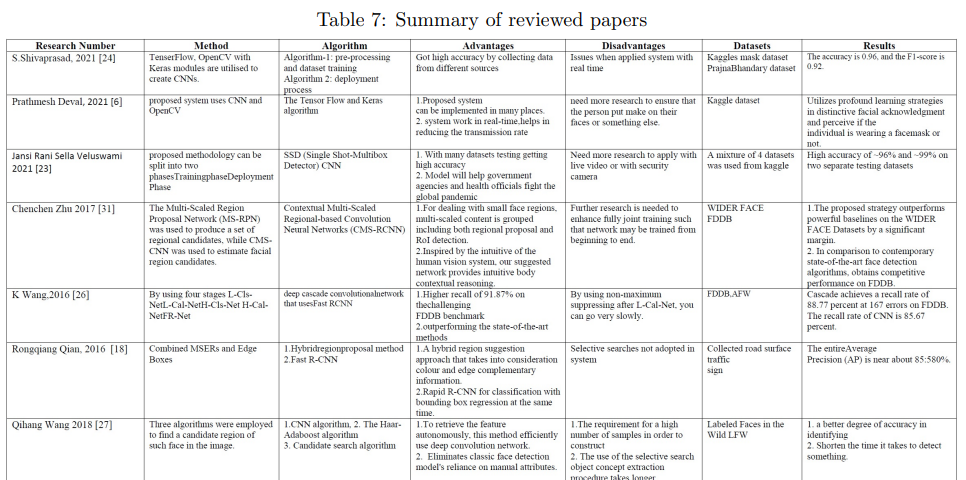
The proposed model is AlexNet method. AlexNet is the name of a convolutional neural network (CNN) architecture, designed by Alex Krizhevsky in collaboration with Ilya Sutskever and Geoffrey Hinton, who was Krizhevsky's Ph.D. advisor. AlexNet competed in the ImageNet Large Scale Visual Recognition Challenge on September 30, 2012. The network achieved a top-5 error of 15.3%, more than 10.8 percentage points lower than that of the runner up. The original paper's primary result was that the depth of the model was essential for its high performance, which was computationally expensive, but made feasible due to the utilization of graphics processing units (GPUs) during training.

# **1.2 LITERATURE REVIEW**

According to the recent literature review done by Firas Amer Mohammed Alia, Mohammed S.H. Al-Tamimi, from Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq, there are two types of object detection method Multi-Stage Detection and Single-Stage detection [1]. There are many sub-categories that is given in Fig1



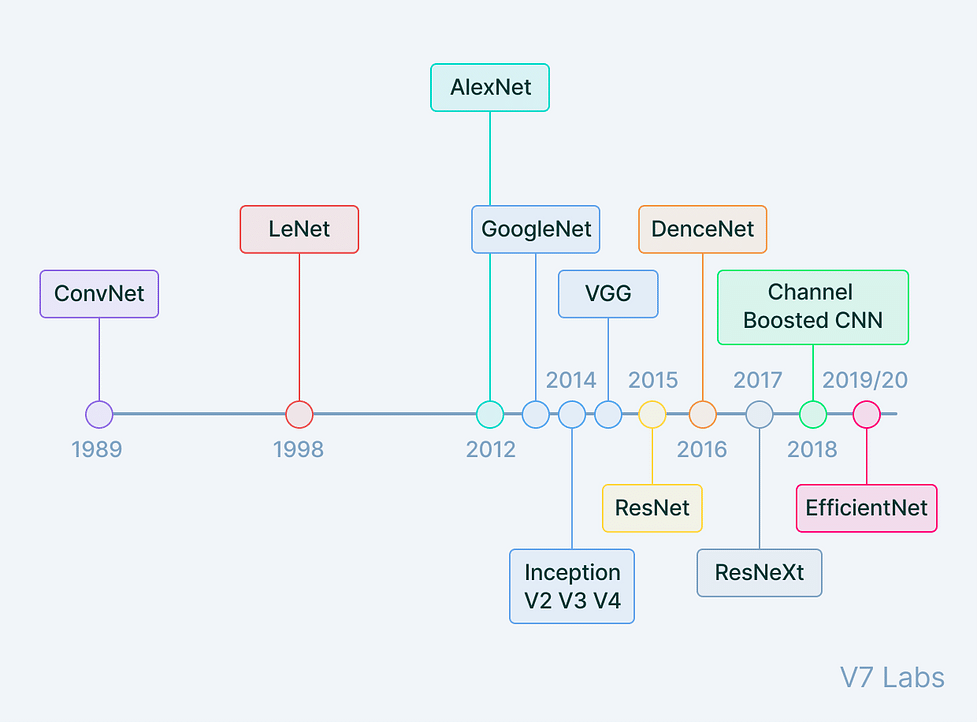
Comparison table

 Table

Description automatically generated

Graphical user interface, text, table

Description automatically generated



https://www.projectpro.io/article/introduction-to-convolutional-neural-networks-algorithm-architecture/560

Among them AlexNet is not very popular and optimized. But in order to simply understand how CNN works and practical applications AlexNet is better option with satisfactory accuracy.

Diagram

Description automatically generated

# **2. METHODOLOGY**

The face mask detection process identifies faces in a given input image or video feed by indicating a boundary box around the face. This article will use a deep learning computer vision model to classify whether a person is wearing a mask or not. All the analysis and data preprocessing are done using the Kaggle dataset [1], containing about twelve thousand images and standard procedures. AlexNet CNN (Convolutional Neural Network) will be used for face mask detection. A flowchart of the methodology is depicted in Fig2.

##### **2.1** **DATA COLLECTTION & PREPROCESSING**

As explained earlier, the used dataset is collected from Kaggle [2] for research purposes. The dataset is in public domain so, the is no legal issues for using it in the project. The dataset contains about 12k images where face images with mask are about 6k in number and, all the images without the face mask are preprocessed from the “CelebFace” dataset created by Jessica Li [3].

For preprocessing all the necessary modules and settings are imported at first. After that, the dataset file is loaded that is saved in local device and split the dataset into 3 separate directories named train, valid and test. Then the image size is set 227 and there are 2 labels in training directory named ‘WithMask’ and ‘WithoutMask’. Then all 3 datasets are shuffled to reduce bias that can arise. In addition, the images are put into the X (X\_train, X\_Valid, X\_Test) and the labels in Y (Y\_train, Y\_Valid, Y\_Test). After that, pickle is used to serialize the dataset. Afterwards, all the photos are combined to get mean photo. Finally, for normalization part all images are subtracted from the mean image.

**2.2 Validating Performance of Deep Learning Models with Kaggle dataset**

Additional kaggle experiments have been run to further validate the effectiveness of the proposed customized CNN model and other deep learning models.

These datasets were chosen because face mask detection studies frequently used them. Additionally, the purpose of these tests is to demonstrate the robustness and generalizability of the suggested customized CNN model. As the customized CNN produces the best results on this dataset, experimental results demonstrate the effectiveness of the suggested approach. On the MAFA dataset, it achieves a score of 95.74% accuracy and 4.29% recall, and on the MOXA dataset, it achieves a score of 94.37% accuracy and 95.28% recall. Additionally, the customized CNN performs significantly better on the RMFRD dataset, where it achieves a 99.63% accuracy and a 99.69% recall, demonstrating its superior performance.

## **2.3 Feature Extraction Procedure for Face Mask Detection by Using AlexNet Deep CNN**

The feature extraction procedure for face mask detection using AlexNet deep CNN involves several steps. First, a dataset of images with and without masks is collected and preprocessed. The preprocessed images are then used to train the AlexNet deep CNN model, which learns to identify key features in the images that are indicative of the presence or absence of a mask.

During the training process, the AlexNet model extracts features from the images using convolutional layers, followed by max pooling and normalization layers. The extracted features are then flattened and fed into fully connected layers, which perform classification and output the probability of the image containing a mask or not.

Once the AlexNet model is trained, it can be used to extract features from new images of faces with and without masks. These features can then be input into a separate classifier, such as a support vector machine (SVM) or decision tree, to make predictions about the presence or absence of a mask in the image.

**2.4 Classification Algorithms**

CNN (Convolutional Neural Network) is a deep learning algorithm that is commonly used for image recognition and classification problems, including face mask detection. There are several ways to classify CNN algorithms for face mask detection, including: In a binary CNN approach, the classifier is trained to identify whether a face mask is present in a given image or not. The dataset of labeled images is used to train the algorithm, with each image bearing the label "mask" or "no mask." Multiclass CNN: In this kind of CNN technique, the classifier is trained to identify one of numerous varieties of face masks in an image. The algorithm might be trained to recognize various kinds of surgical masks or N95 masks, for instance. Real-time CNN: This kind of CNN algorithm is built using a classifier that can function in real-time, like in a webcam or video stream. The algorithm is trained to identify if a face mask is present in each frame of the video or not. When using a pre-trained model, such as VGG, Inception, or ResNet, the classifier in a transfer learning CNN algorithm refines the model using data from face mask detection. When there is a shortage of training data, transfer learning can be useful. Hybrid CNN: In this sort of CNN method, the classifier integrates several features, such as color, texture, and shape, to detect face masks. Compared to a straightforward binary or multiclass CNN technique, this kind of approach may be more accurate. In general, the specific use case and the data that are available will determine which CNN method is used for face mask detection. The most advantageous option will rely on elements like the difficulty of the task, the amount of the dataset, and the available computer resources. Each type of CNN method has benefits and limitations.

**2.5 Experimental Setup and Implementations**

1. Data Collection: Collect images of people wearing face masks and people not wearing face masks.

2. Data Pre-processing: Organize the collected data into sets for training, validation and testing. Normalize the data and split it in to batches.

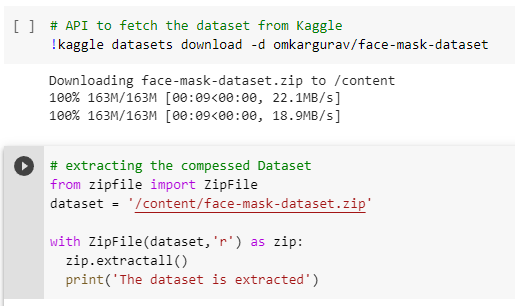
3. Model Selection: Choose a Convolutional Neural Network (CNN) model that is suitable for the task.

4. Model Training: Train the model using the training data set and evaluate the performance of the model on the validation data set.

5. Model Testing: Test the model on the test data set to get the accuracy and other metrics of the model.

**Data Base Insert:**

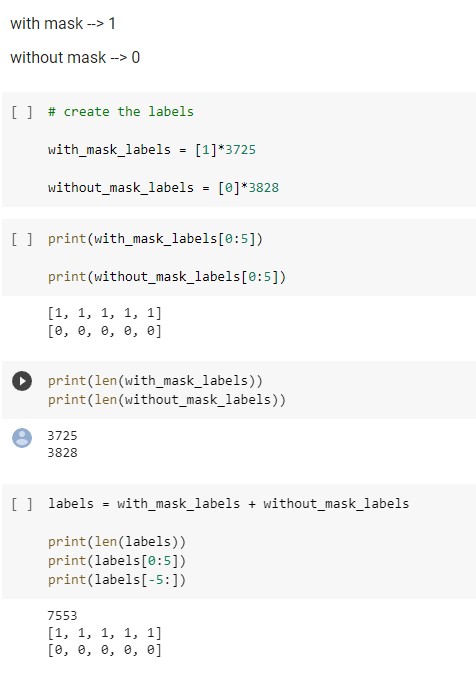
****

****

**Importing the Dependancies:**

****

**Creating labels for the two class of images:**

****

**Display the image:**

****

**Image processing:**

****

**2.6 Confusion Matrix Analysis Face Mask Detection by Using CNN**

The confusion matrix is a tool used to measure the accuracy of a machine learning model. The confusion matrix is a way to visualize the results of the model, showing the true positives, false positives, false negatives, and true negatives of the model. True Positives (TP): These are the number of times the model correctly predicted that the person was wearing a face mask. False Positives (FP): These are the number of times the model incorrectly predicted that the person was not wearing a face mask, when in fact they were. False Negatives (FN): These are the number of times the model incorrectly predicted that the person was wearing a face mask, when in fact they were not. True Negatives (TN): These are the number of times the model correctly predicted that the person was not wearing a face mask. By looking at the confusion matrix, you can see how accurate the model is in predicting whether or not a person is wearing a face mask. The higher the number of true positives and true negatives, the more accurate the model is.

**2.7 Block Diagram and Workflow Diagram of Proposed Model**

Block Diagram The block diagram of a CNN for face mask detection technique includes the following components: 1. Image Input: This is the input image of a person wearing a face mask or not. 2. Pre-Processing: This stage of the process involves normalizing, cropping, and resizing the input image to create a suitable input for the network. 3. Convolutional Layers: This is the part of the network that extracts features from the input image. It consists of several convolutional layers with various filter sizes and parameters. 4. Pooling Layers: This stage of the network reduces the dimensionality of the feature maps generated by the convolutional layers. 5. Fully Connected Layers: This stage of the process creates a final output based on the features extracted by the convolutional and pooling layers. 6. Output: This is the output of the network which is a prediction of whether the person in the image is wearing a face mask or not. Workflow Diagram The workflow diagram of a CNN for face mask detection technique includes the following steps: 1. Image Input: The input image of a person wearing a face

**Diagram

Description automatically generated**

##### **2.8 MODEL TRAINING AND TESTING**

Alex net Architecture by transfer learning approach is used. Alexie is a pre-trained CNN (Convolutional Neural Network) model.

Our dataset is collected to train our model. The data has been spitted set into Test and Train. After that, we preprocessed our data. We have resized the image to 227 X 227 because we are using Alex net. After preprocessing, the data is imported to our base model Alex Net and this architecture is implemented by using a transfer learning approach. In this research, the model is classified by AlexNet architecture. It is a type of convolutional neural network. It is a classification algorithm that is best suited for the classification of RGB images. The architecture includes eight layers, among them, the first 5 were convolutional layers, followed by some max-pooling layers which are used in down sampling, and the last three layers were fully connected. It uses a rectified linear unit activation function named RELU. Activation in each of these layers except the output layer. That is found out using the RELU as an activation function accelerated the speed of the training process by almost six times. It also used the dropout layers, which prevented its model from overfitting. Further, the model is trained on the Imagenet dataset. The Imagenet dataset has almost 14 million images across a thousand classes. After training our model we have got our desired output with good accuracy.

Diagram

Description automatically generated

**AlexNet Architecture**

* It has 8 layers with learnable parameters.
* The input to the Model is RGB images.
* It has 5 convolution layers with a combination of max-pooling layers.
* Then it has 3 fully connected layers.
* The activation function used in all layers is Relu.
* It used two Dropout layers.
* The activation function used in the output layer is SoftMax.
* The total number of parameters in this architecture is 62.3 million.

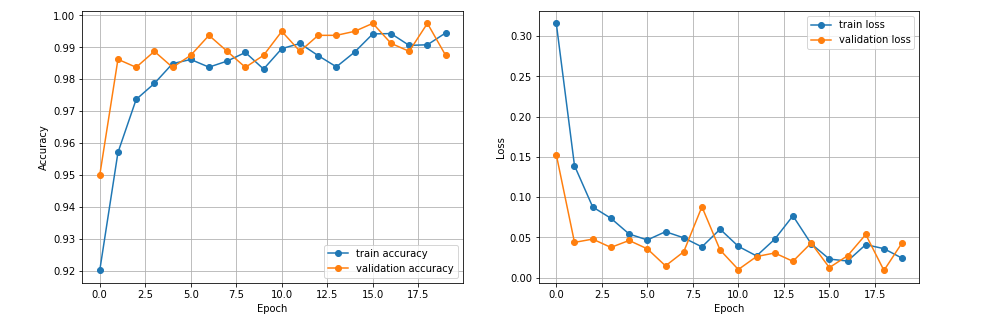
Diagram: AlexNet Architecture

Diagram, engineering drawing

Description automatically generated

##### MODEL EVALUATION

**AlexNet:** Convolutional neural networks are one of the variants of neural networks where hidden layers consists of convolutional layers, pooling layers, fully connected layers and normalization layers.



|  |  |
| --- | --- |
| Accuracy | 0.9808467626571655 |
| Loss | 0.08948305994272232 |

3. **Results and Discussion**

Initial tests are conducted without evaluating the performance of every model using the four-step image processing pipeline. In this case, the images are fed directly into the models without any additional processing, such as color transformation or filtering. It is evident that the models' performance is lacking. Despite this, the customized CNN performs better than other models, with an accuracy rate of 88.52%.

It is evident that the models' performance is lacking. Despite this, the customized CNN performs better than other models, with an accuracy rate of 88.52%. In a similar vein, it outperforms YOLO v3 and Faster R-CNN in terms of precision, recall, and F1- scores.

Face masks are difficult to distinguish from other items, according to the analysis of the dataset. For instance, skin tone and facial hair can resemble the color of the mask, making mask detection very challenging. In order to prepare images for classifier training, four stages of image preprocessing techniques were used.

The moment set of tests includes utilizing the picture preprocessing steps as portrayed already. Test comes about are given in Table 5 demonstrating way better execution than without the utilize of picture preprocessing. It can be watched that all profound learning models appear altogether way better execution with picture preprocessing. The Quicker R-CNN appears superior comes about with 92.65% exactness, 91.82% exactness, 94.24% review, and 92.23% F1-score as compared to the YOLO V3 calculation. By and large customized CNN accomplished the most elevated execution with 97.25% precision, 96.20% exactness, 97.34% review, and 96.77% F1-score. YOLO V3 calculation may be a basic and single-shot calculation with less deduction time. But in comparison, Faster R-CNN has accomplished way better comes about. There's a trade-off between execution effectiveness and speed. the precision bend for the best-performing customized CNN show appears that the proposed show easily progresses the preparing and testing precision which demonstrates the strength of the proposed approach. Thus, confront cover discovery utilizing CNN is superior to other off-the-shelf profound learning models utilized within the try

3.1 **Results Comparison**

The effectiveness of the modified CNN is also evaluated in light of recent cutting-edge studies [26, 27]. Only studies that use a publicly accessible dataset with real images are chosen for comparison because [22] reports that the accuracy with simulated face masks is frequently high. Above, a comparison with the chosen studies shows that the suggested approach performs better in terms of accuracy and precision than the current models. The influence of dark and low-light environments, which can affect the robustness of the proposed approach, is not taken into account in the few studies that were conducted in a real-world setting.

**4. Conclusion and recommendation**

* Evaluate the model's performance in real-world settings: While the CNN model may perform well on a testing dataset, it is important to evaluate its performance in real-world settings. This could involve deploying the model in a live video stream or on a mobile device, and measuring its accuracy in detecting face masks in real time.
* Improve the model's performance on different types of face masks: The CNN model may perform differently depending on the type of face mask being worn, such as cloth masks, surgical masks, or N95 masks. It may be useful to explore ways to improve the model's accuracy in detecting different types of masks.
* Address bias and fairness issues: Machine learning models can be biased if the training data is not representative of the population it is being used on. It is important to evaluate the model for fairness and to address any potential bias or discrimination issues that may arise.
* Explore other deep learning techniques: While the CNN model is a powerful tool for face mask detection, there are other deep learning techniques that could be explored, such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs) with attention mechanisms.
* Collaborate with public health agencies and organizations: The face mask detection model could be used by public health agencies or organizations to monitor compliance with mask-wearing guidelines. Collaborating with these organizations could lead to new opportunities for applying the model in real-world settings and improving public health outcomes.

# **References**

[1] Tao Zhang, Jingjing Li, Wenjing Jia, Jun Sun, and Huihua Yang. Fast and robust occluded face detection in atm surveillance. Pattern Recognition Letters, 107:33–40, 2018.

[2] Nancy HL Leung, Daniel KW Chu, Eunice YC Shiu, Kwok-Hung Chan, James J McDevitt, Benien JP Hau, Hui-Ling Yen, Yuguo Li,

Dennis KM Ip, JS Peiris, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. Nature medicine, 26(5):676–680, 2020.

[3] Shuo Feng, Chen Shen, Nan Xia, Wei Song, Mengzhen Fan, and Ben- jamin J Cowling. Rational use of face masks in the covid-19 pan- demic. The Lancet Respiratory Medicine, 8(5):434–436, 2020.

[4] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al. Masked face recognition dataset and application, 2020.

[5] Stefanos Zafeiriou, Cha Zhang, and Zhengyou Zhang. A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding, 138:1–24, 2015.

[6] Xin Li, Shenqi Lai, and Xueming Qian. Dbcface: Towards pure convolutional neural network face detection. IEEE Transactions on Circuits and Systems for Video Technology, PP:1–1, 05 2021. doi: 10.1109/TCSVT.2021.3082635.

[7] Ashu Kumar, Amandeep Kaur, and Munish Kumar. Face detection techniques: a review. Artificial Intelligence Review, 52(2):927–948, 2019.

[8]Y Wang. Which mask are you wearing. Face Mask Type Detection with TensorFlow and Raspberry Pi. Medium, 1, 2020.

[9]G Jignesh Chowdary, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal. Face mask detection using trans- fer learning of inceptionv3. In International Conference on Big Data Analytics, pages 81–90. Springer, 2020.

[10] Shiming Ge, Jia Li, Qiting Ye, and Zhao Luo. Detecting masked faces in the wild with lle-cnns. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2682–2690, 2017.

[11] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In Proceedings of the IEEE confer- ence on computer vision and pattern recognition, pages 5525–5533, 2016.

[12] Che-Yen Wen, Shih-Hsuan Chiu, Yi-Ren Tseng, and Chuan-Pin Lu. The mask detection technology for occluded face analysis in the surveillance system. Journal of Forensic Science, 50(3):1–9, 2005.

[13] Praveen Kakumanu, Sokratis Makrogiannis, and Nikolaos Bourbakis. A survey of skin-color modeling and detection methods. Pattern recognition, 40(3):1106–1122, 2007.

[14] Chung J Kuo, Tsang-Gang Lin, Ruey-Song Huang, and Souheil F Odeh. Facial model estimation from stereo/mono image sequence. IEEE transactions on multimedia, 5(1):8–23, 2003.

[15] Zhaoqing Pan, Peng Jin, Jianjun Lei, Yun Zhang, Xingming Sun, and Sam Kwong. Fast reference frame selection based on content similar- ity for low complexity hevc encoder. Journal of Visual Communica- tion and Image Representation, 40:516–524, 2016.