Face Mask Recognition to Identify People Wearing Masks to Support Covid-19 Prevention Policies

HASAN ÖMER YILMAZ

Piri Reis University, Turkey

VOLKAN USLAN

Piri Reis University, Turkey

HUSEYİN ŞEKER*

Birmingham City University, United Kingdom

Nowadays, Covid-19 pandemic has penetrated all the human beings daily life. One of the most common and strongest way to prevent spread is wearing a face mask in public places. People gained a new term to their vocabulary; “new normal”. The new normal contains wearing a face mask as well. Therefore, a face mask recognition system is a vital need for helping daily life processes. This paper acquaints a face mask recognition to identify masked and unmasked faces to support Covid-19 policies. The face mask recognition in this paper developed by deep learning algorithm using the CNN architecture VGG-16. Our results suggest that deep learning-based method achieved high accuracy (99%) in both the validation and testing datasets.

Additional Keywords and Phrases: Face Mask Recognition, Deep Learning, Covid-19

1 INTRODUCTION

In today’s world, Artificial Intelligence (AI) and Neural Networks are crucial technologies for human life [19]. The main logic at the backstage of artificial neural networks is teaching machines solving problems without human interference. Therefore, computers are trained to built models based on datasets using machine learning methods to predict solution with the highest accuracy.

2020 had been a tough year for all human beings around the world because of the Covid-19 disease caused from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [10]. The virus causing the disease spread so fast all over the world and became pandemic. Covid-19 disease have caused deaths, sufferings, economic problems, and sociological problems for most of the countries. The pandemic changed society’s all aspects and daily routines. Governments and institutions take steps to prevent spreading of the virus for society’s safety. The coronavirus spread in different ways such as airborne transmission, contact with the infected person etc. The airborne transmission way is one of the most common spread method, the virus

* Corresponding author (Huseyin Seker)

(SARS-CoV-2) may remain in the air for a long time and it can affect the person around 1 meter distance [1]. Recent research reported the efficacy of wearing a face mask on prevention of respiratory infections [9], [12], [13]. Governments set wearing a face mask as a policy for Covid-19 pandemic in public places and social environments since the beginning of the pandemic because it spreads through close contact easily. Therefore, one of the new daily routines of human beings have became wearing a face mask to avoid airborne transmitted virus, when they step out their houses. However, some people unwilling to comply this rule and endanger society's health and daily life. A face mask recognition system can be developed for separating that people who wears a face mask from people who does not. Camera systems using computer vision and deep learning algorithms can be used together to prevent Covid-19 disease and the virus causing the disease from spreading. There are huge amounts of images to be analyzed, so it's impossible to analyze and follow those images in realtime by a human. Neural networks are widely used in image processing and numerous applications of neural networks in image processing were already appeared in research studies [18]. For this reason, we considered that computer vision and deep learning together would be efficient to solve such a real life problem.

The purpose of this paper is creating a face mask recognition model for detecting people who are not wearing the face mask to support Covid-19 prevention policies. This model is developed on deep learning [16] using Convolutional Neural Network (CNN) architecture. The developed model is trained and tested using newly created dataset. The rest of the paper starts by presenting materials and methods that includes the convolutional neural networks, VGG-16, and transfer learning in Section II. The proposed deep learning based model is described in Section III. Results are presented in Section IV. Discussion and conclusion are provided in Section V.

2 MATERIALS AND METHODS

Convolutional Neural Networks

Neural networks consist of computational units (neurons) that model the way the human brain works [2]. There are different types of neural networks. Convolutional neural networks [21] is a type of deep learning network for recognition of patterns in images (e.g. face expression recognition [20]). Up to the present, various applications have already made on this new method [22] as well as for image classification [23]. It is based on the deep learning techniques of artificial intelligence applications that generates quick and almost fully optimized

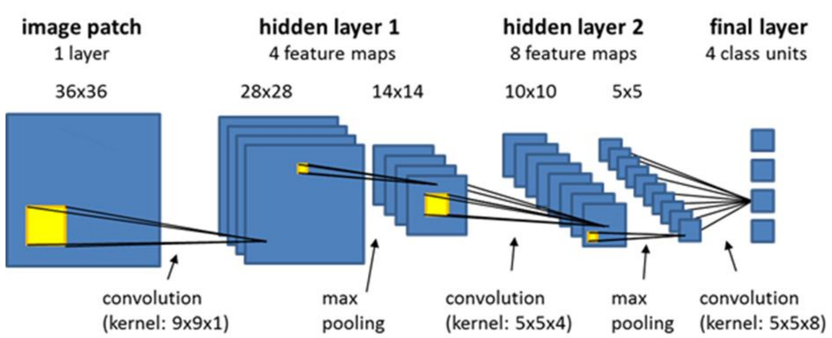

Figure 1: Sample Convolutional Neural Network Scheme (Image from [3]).

solutions. Convolutional neural networks use photograph or video files represented in matrix formats as inputs to extract unique features. As illustrated in Figure 1, CNN architecture consists of three layers; convolutions layers, pooling layers, and fully connected layers.

CNN applies filters to detect unique features from the input images or videos to separate or compare them among each other. Convolution layers are basically filters that learn some feature. The output from a convolutional network called as activation map or feature map. Every convolutional network come out a group of stacked filters that produce the output value. The first step of neural network process that named as convolution function is masking image with a filter. Activation stage comes after convolution stage. ReLU function deletes negative values in images. Last but not least, the pooling layers reduce the parameters count and control the overfitting. There are two types of pooling layers; average pooling and max pooling. Convolutional networks hold filters that is important to know meaning of hypermeters. Filter size, stepping range, adding zeros, fitting parameters at layers, and also understanding complexity of the model are important factors for hypermeters of CNN [5]. The way the network operates based on the activations in one layer, determine the activations of the next layer. The network decreases the complexity of the model by optimizing parameters and depth.

VGG-16

There are many types of convolutional neural networks. Among them VGG-16 [15] is used as a convolutional neural network architecture for this study because it is a very dense network which is valuable when extracting the features to classify and identify masks. VGG is named after Visual Geometry Group -a research group- established at University of Oxford. Every convolutional step aims to create a deeper feature map but smaller in size. This is based on the deepening technique of the conventional neural network to improve its performance. VGG-16 operates with filters that are quite small in size, typically 3 x 3 kernel-sized filters. Filters with different weights are calculated at the output of each convolution layer of the model, and as the number of layers increases, the features formed in the filters symbolize the depths of the image. One of the reasons why VGG-16 is preferred as architecture is that its layers can be used as trio and this is one reason that VGG-16 works very accurate with medium and huge volume datasets.

Transfer Learning

A convolutional neural network is one of the best algorithms for image processing. However, training a neural network from scratch is a very long process and it requires very powerful and upscale computing system. At this point, transfer learning helps us to speed up the process. Transfer learning is basically using the features that are learned from another similar or different deep learning task. Transfer learning is a big time and process saver for the second task. There are different transfer learning approaches for convolutional neural networks, from those pre-trained model approach is used for our model. Transferring the weights from a pre-trained model had proven high accuracy for CNN [4]. In this study, we used ImageNet for transfer learning [11]. The weights learned from 14 million images were transferred to our model.

Figure 2: Sample Unmasked and Masked Face Image from the Generated Dataset

3 DEEP LEARNING BASED MODEL

The first step of developing a deep learning model is collecting image data. Dataset is vital for training the face mask recognition model. Therefore, we collected data from various resources to form a dataset. First one is Flickr-Faces-HQ Dataset (FFHQ) [6] which provides us unmasked face images. Second one is Correctly Masked Faces Dataset (CMFD) [7] which is created based on Flickr-Faces-HQ Dataset. In the generation of second dataset, image augmentation method was used so that the data collecting and pre-processing tasks became much faster. Image augmentation is the operation of generating more images with some modifications from existing dataset and including them into the same dataset. Figure 2 illustrates a sample unmasked and masked face image from the generated dataset. There are 11758 images collected in the dataset; half of them are unmasked faces and the other half is masked face images. These images were splitted into 9406 train (80%), and 2352 test (20%) images. All of the image sizes are 1024x1024. Lastly, It is worth to be mentioned that the minimum age for wearing face masks changing depends on the country. The categories is split into classes after each sample initialized with a label belongs to. Lastly, labels were arranged as binary variables to be used in the machine learning model before the training process starts. Pre-processing is the last step before training and testing of the model begins. In pre-processing part, images resized to 224x224, then images converted to arrays and pixels are scaled to values in input images to the range of [-1,1]. Resizing operation provided an increased efficiency for training the model. The smaller image sizes mean faster to train. Image sizes got smaller but pixel values remained the same. The conversion of an image into an array is necessary to make images callable by loop functions. At the end of the pre-processing step, the dataset is shaped into sizes (11758, 224, 244, 3).

The pre-trained VGG-16 model downloaded from ImageNet to classify people with face mask and without face mask. VGG-16 has 16 layers, but the last layer is deleted and a dense layer added instead because the last layer of VGG-16 has thousands of classes but the required class count is only two (masked, unmasked).

The pre-trained model has 14,714,688 parameters and all of them are trainable parameters. However, we don't need to set any of these parameters as they are pre-trained already using the transfer learning. Transfer learning carried all of the feature extractions from the pre-trained network to the specific model. The output of the pre-trained model applied to our specific model that is generated by using 6 x 6 sized pooling layers to avoid increasing complexity of model. In the flatten layer, the pooled feature map converted to a single column. The flatten layer, followed by connected 128 neuron layers are created by activation ReLU function is used in the dense layer and dropout layer set with dropout 0.5 rate. Dropout layer prevents the model from overfitting. If the model was overfitted then it would generalize poorly and error rates in predicting new samples not included in the training set would increase. Dropout layer provides to learn and avoids to memorize the training samples. Lastly, dense layer with two neurons with sigmoid and softmax function is added in order to classify masked and unmasked faces. The final model has 14,780,610 total parameters; 65,922 of the parameters are trainable and 14,714,688 of the parameters are non-trainable. The transfer learning algorithm reduced the calculation time with the skip of training those 14,714,688 parameters.

There is one more step remained before the training of the model gets completed. Learning rate, batch size, epochs must be defined. Learning rate determines how fast the model reach minimum loss, epochs means how many iterations will be needed to update model weights, and batches size means how many samples will be trained at the same time, RAM capacity is important for batches size. In this model, learning rate was set to 0.001, epochs was set to 10, and batches size was set to 60. Adam optimizer [17] was chosen to optimize model on gradient descent. Finally, augmented images added to model and then training of the model was conducted. When the training of the model was over, prediction can take place on testing data with the same batch size to find output that gives the maximum probability of classification by the argmax function. Accuracy and loss of the training, validation and testing phases are then evaluated to consider how well the model succeeds in classifying the images.

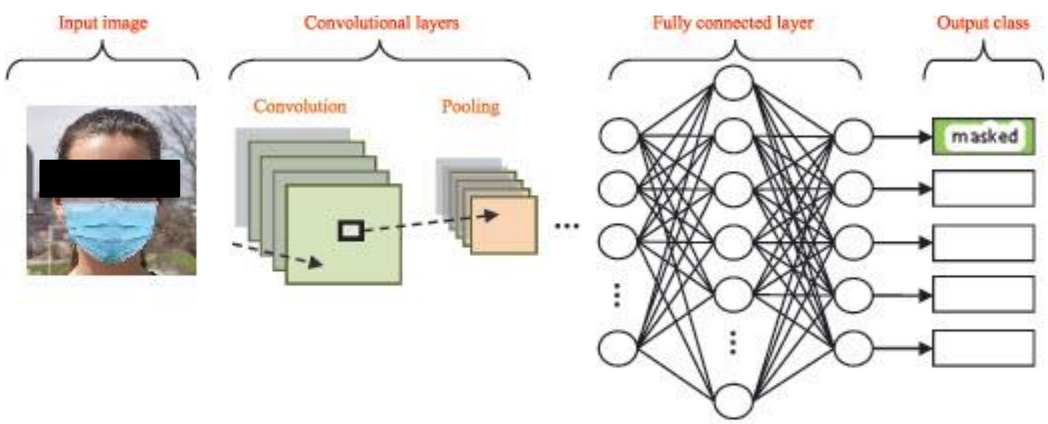

Figure 3: Illustration of the face mask recognition pipeline. Image adapted from [14].

4 RESULTS

Our aim is to propose a model that will distinguish people who wear a mask from those who do not. Therefore a dataset is generated with images collected from various resources. Faces with mask images are taken from CMFD, and faces without mask images are taken from FFHQ datasets. The image database consisted of 11758 images (with face mask or not). The proposed face mask recognition model is trained, validated and then tested. The number of iterations, n=10, has been used and it was seen the accuracy reached a good level. During the training process even though accuracy and loss values have experienced undulation in values but in general accuracy is increased and loss is decreased. Accuracy, precision, recall and f1-score are used as metrics to evaluate our model's performance. The analysis comprises of plotting the training loss, validation loss, training accuracy and validation accuracy which are also useful to visually see the performance of the model (Figure 3).

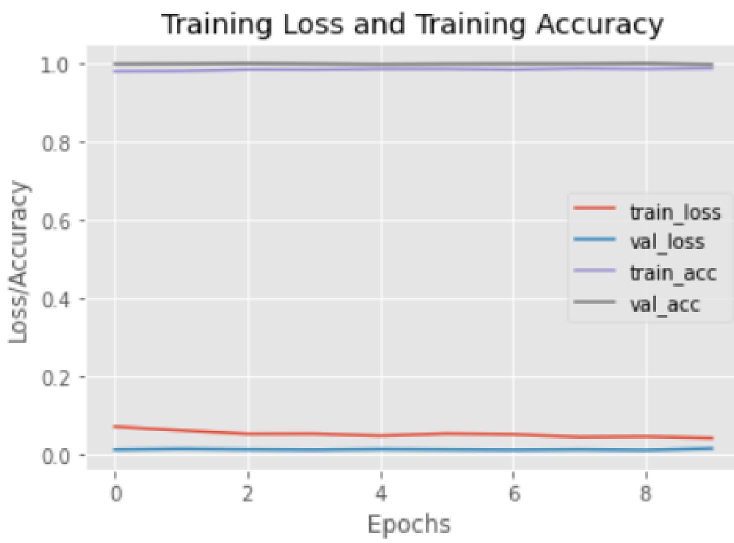

Figure 3: Training and Validation Loss and Accuracy

The precision values are 1.00 and 0.99 for masked and unmasked, respectively. The recall values are 0.99 and 1.00 for masked and unmasked faces, respectively. The f1-score is the harmonic mean of the precision and recall. The f1-score values are 1.00 and 1.00 for masked and unmasked faces, respectively. As a result, we obtained an accuracy percentage slightly more than the related works presented in recent literature [8]. We have also designed the face mask recognition process as a realtime process, we have used video streaming and time in combination to detect and predict masks based on a rate of interest in the parts of a face (as illustrated in Figure 4). The trained mask classifier model employed in realtime to predict a face that is masked or unmasked provided a very good accuracy.

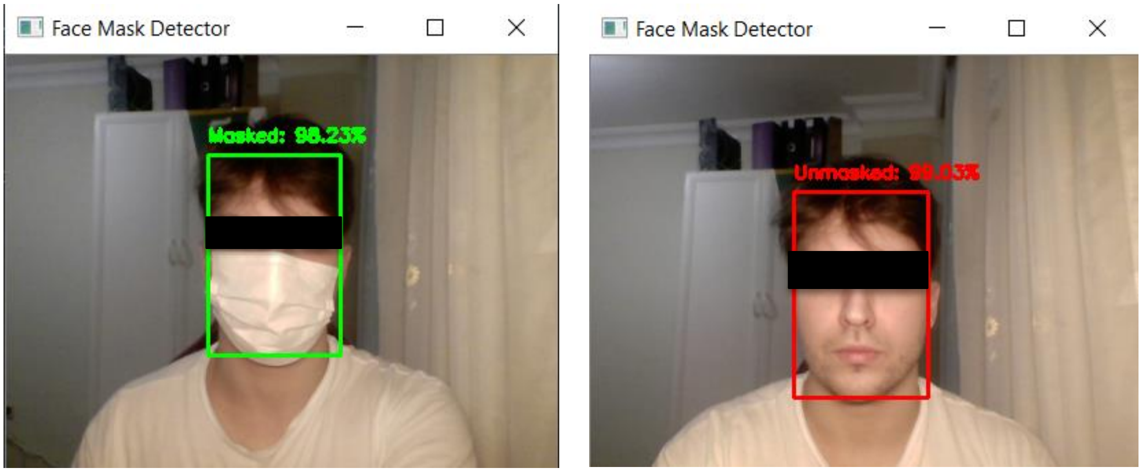

Figure 4: Masked and Unmasked Face Recognition in RealTime Video Streaming

5 DISCUSSION AND CONCLUSION

The Covid-19 pandemic that human beings face all around the world is hard to handle till all the people are vaccinated. The vaccines don't give certain results against the coronavirus also most of the countries don't have enough vaccine dose to vaccinate all the citizen. The transmission of the virus is still hard to prevent because of these reasons. Also, it's a proven fact that the virus spread more easy and faster in crowded places. Researches have proved wearing face masks reduce the transmission rate of the coronavirus. A face mask recognition system with computer vision is useful in order to distinguish easily the people who do not wear face masks in public places. This system helpful to maintain flow of daily life and society health. The motivation of this study is coming from the aforementioned issues. We briefly explained the methods and tools briefly before presenting the deep learning based face mask recognition system. A dataset which composed of 11758 images is created and used in developing deep learning model to recognize of masked faces from unmasked ones. Images of dataset are collected from several sources. Face mask detector developed using VGG-16 architecture and transfer learning algorithm. Image augmentation methods used to improve our training dataset variety without having an extra-large dataset. The model achieved 99% accuracy which is nearly perfect but sometimes performance of recognition depends on some variables such as light position, light power, resolution etc. These variables can affect the performance of face mask detector but the model still work smooth and issueless. The developed model is vital to recognize people who are not wearing face masks in closed areas and public places. In the future, the model can be developed to identify type of masks such as basic medical masks, n95 masks or fabric masks. On the other hand, there are many people who are not wearing face masks correctly in public. As a future work, these issues can be considered in creating a new dataset and improve the model including detecting the mask on a face worn correctly or incorrectly.

REFERENCES

[1] World Health Organization (WHO). 2020. Modes of transmission of virus causing COVID-19: implications for IPC precaution

recommendations: scientific brief.

[2] D.S. Sisodia, R.B. Pachori, L. Garg. 2020. Handbook of research on advancements of artificial intelligence in healthcare engineering. IGI Global.

[3] eCognition. 2019. Reference Book eCognition Developer. Trimble Inc.

[4] K. Fırıldak, M.F. Talu. 2019. Investigation of Transfer Learning Approaches Used Convolution Neural Networks. Anatolian Journal of Computer Science, 4, 2, 88-95.

[5] A. Amidi, S. Amidi. 2019. Convolutional Nural Networks cheatsheet. Stanford CS 230 – Deep Learning.

[6] Tero Karras, Samuli Laine, Timo Aila.2019. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4401-4410.

[7] A. Cabani et al. 2021. MaskedFace-Net -- A Dataset of Correctly/Incorrectly Masked Face Images in the Context of COVID-19. Smart Health 19, 100144.

[8] B. Qin, D. Li. 2020. Identifying Facemask-wearing Condition Using Image Super-Resolution with Classification Network to Prevent. Sensors, 20, 5236.

[9] B.J. Cowling et al. 2009. Facemasks and hand hygiene to prevent influenza transmission in households: A cluster randomized trial. Ann. Intern. Med., 151, 437–446.

[10] World Health Organization (WHO). 2020.Naming the coronavirus disease (COVID-2019) and the virus that causes it.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR).

[12] S.M. Tracht et al. 2010. Mathematical modeling of the effectiveness of facemasks in reducing the spread of novel influenza A (H1N1). PLoS ONE, 5, e9018.

[13] T. Jefferson et al. 2011. Physical interventions to interrupt or reduce the spread of respiratory viruses. Cochrane Database Syst. Rev. 7, CD006207.

[14] Waseem Rawat, Zenghui Wang. 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput, 29, 9, 2352–2449.

[15] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[16] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature. 521.7553, 436-444.

[17] D.P. Kingma, J. Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[18] M. Egmont-Petersen, D. de Ridder, H. Handels. 2002. Image processing with neural networks—a review. Pattern Recognition, 35, 10, 2279-2301.

[19] Y. Hu, Maryam Vaziri-Pashkam. 2021. Limits to visual representational correspondence between convolutional neural networks and the human brain. Nat Commun. 12, 2065.

[20] Hilal Hazel Cumhuriyet, Volkan Uslan, Ersin Yavas, Huseyin Seker. 2021. Face Expression Recognition using Deep Learning in Cloud Computing, 5th International Conference on Cloud and Big Data Computing.

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25, 1097-1105.

[22] Jiuxiang Gu et al. 2018. Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377.

[23] Waseem Rawat, Zenghui Wang. 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput, 29, 9, 2352–2449.