An Application of Detecting Faces with Mask and without Mask using Deep Learning Model

Ratnesh Kumar Shukla[1], Arvind Kumar Tiwari[2]
[1]*Department of Computer Science & Engineering, Dr. APJ Abdul Kalam Technical University, Lucknow, Uttar Pradesh, India*
[2]*Department of Computer Science & Engineering, Kamla Nehru Institute of Technology, Sultanpur, Uttar Pradesh, India*

Keywords: Face detection model, MobileNetV2, Facial action unit, Convolution neural network, Random convolution neural network, Deep neural network, Biometric machine etc.

Abstract: The proposed model is stronger as it naturally will identify people with masks and without mask. This approach reduces the deep learning process to a single stage and the mask detector model is added to identify with mask and without mask. What we need to do is to use the learning algorithm to provide us with bounding cases in one forward network pass for both people with masks and without masks. The Keras classifier is based on the MobileNetV2 neural net architecture. This model was tested in real time with pictures and video streams. Although the exactness of the prototype is around 98% and model optimisation is a continuous process by setting the hyper-parameters. We are finding a highly precise solution. Size and computer costs are highly optimized and tailored for object detection tasks on-device such as a cell phone or camera streams.

1 INTRODUCTION

Deep Learning is a process to solve and the image related problem to using different algorithms. In this process, we are observing how machine works and learn the process. This can be improving the previous experience and benchmark for the classification of the image, text and video in algorithm. Before beginning our work first we must know about machine learning, artificial intelligence, deep learning and building there connection. Deep learning is a subset of machine learning and artificial intelligence. Artificial Intelligence plays vital role to develop computers which think intelligently. Machine Learning is using data insight algorithms to improving the result of the output. But Deep Learning is also using that particular algorithm and improves the features of neural network algorithms. It is just a kind of algorithm or model that seems to work very well to predict output. After collection of input values or underlying data, they will be transferred the data through this networks. Hidden layers are converting this value in output layer. We find the output by using fully connected layer and flatten layer. Hidden layers of the neural network filter the data so that they finally find what their features to the target variable and each node has a weight and bias to multiply the input value. Do this over a few layers and the network is basically capable of translating data into a meaningful object. For one reason, Deep Learning is very exciting for us in human life but we have been able to achieve practical and useful precision in the tasks that matter. Machine Learning has been used to classify images and texts for decades, but it has not reached the threshold. There is a basic precision that these algorithms need to work with in applications. Deep learning finally helps us to cross the line and change it. Deep learning senses the named data and begins to identify the true human being. A typical neural network (NN) consists of a number of basic, interconnected processors called neurons, each generating a sequence of real-value activation functions. Input neurons are stimulated by receptors which feel the environment while other neurons are stimulated by a weighted association of previously active neuron and neuron can have a behavioural impact on the environment. Deep Learning is giving credit to about knowing the weights that make NN the desired task, such as driving a car. Depending on the issue and how the neuron is related, such actions can involve a long, causal chain of computational

Shukla, R. and Tiwari, A.
An Application of Detecting Faces with Mask and without Mask using Deep Learning Model.
DOI: 10.5220/0010562500003161
In *Proceedings of the 3rd International Conference on Advanced Computing and Software Engineering (ICACSE 2021)*, pages 54-60
ISBN: 978-989-758-544-9

stages and aggregate network activation transforms every single point.

1.1 Basic Working of Neural Network

Nowadays, there are many types of deep learning neural networks model that are used for a number of purposes. In this article, we will go through the topologies most widely used in neural networks. Briefly introduces how they can work along with some of their applications and task with real-world challenges. Today many kinds of deep learning neural networks are being used for a different number of purposes. The neural network functions in the same way as the human brain and the brain is made up of billions of cells called neurons. They link up like a network, the perceptron links a web point that solves a basic but complex problem. In general, it is a mathematical function that collects and classifies information according to a specific architecture. The network has close parallels with mathematical approaches such as curve fitting and regression analysis. Neural networks were first proposed in 1944 by Warren McCullough and Walter Pitts in Chicago University scholars who moved to MIT in 1952 as founding members of what is often referred to as the first cognitive science department. When a neural network is formed, all its weights and input values are initially set to random values. Training data is feed to the bottom layer in the model and transferred to the next layers, multiplying and adding together dynamically, before the output layer finally arrives dramatically updated. During practise the weights and thresholds are constantly changed until the results of the testing with the same marks consistently produce the same outputs. The dataset we use consists of images of different colours, sizes and directions. We then need to convert all images to grayscale, so we need to make sure the colour is not a critical point for mask detection.

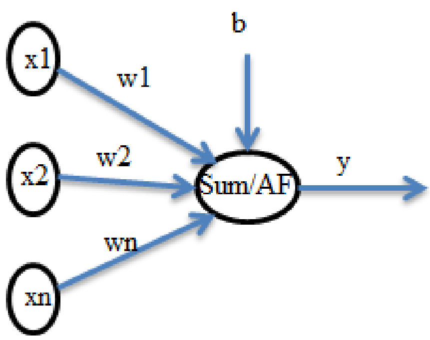

Figure 1: Basic architecture of neural network with weight and bias.

Figure 1 demonstrates the basic neural network architecture. It is often used as a single-layer network model. This figure1 shows the input layer and the output layer with a mixture of weight and bias. Input layer (x1, x2.................xn), weight (w, w2...w) with a mixture of bias (b) and output (y).

$$y= AF\ (\textstyle\sum_{i=1}^{n} xi)\ +b \qquad (1)$$

Equation (1) indicates the logical feature of the neural network. It is also referred to as the mathematical activity of biological neurons. There is no hidden layer in the neural network as it is a clear model of biological neurons. The input is taken and the weighted input and bias of each node are determined. It then uses the activation function (primarily the sigmoid function) for grouping purposes.

2 RELATED WORK OF FACE DETECTION AND RECOGNITION USING DEEP LEARNING MODEL

Face detection and recognition has become an active and common area of research for deep learning researchers in a wide range of disciplines, such as neuroscience, psychology and computer vision. In a wide range of applications including biometrics, face recognition, safety, security, advanced driver tracking, entertainment, and virtual reality. In particular with regard to computer vision, impact and substantial capacity for face detection and recognition increased. Since deep learning is a popular approach in terms of quantitative financing, practical learning calls for beginners. However, train practical deep learning deals to determine where to sell at what price and how much to produce and debug error prone and stressful (Liu, Yang, Chen, Zhang, & Yang Xiao & Wang, 2020). We also provided a way of examining the functions of and network unit in order to better understand how the network operates. In the classifier, the units display how the network breaks down complex class recognition into single view concepts for each class (Bau, Zhu, Strobelt, Lapedriza, Zhou, & Torralba, 2020). Transformer or object detection is a set-based global loss that forces such predictions into a small community of learned object queries through bipartite matching and transforms encoder decoder architecture. The accuracy of ResNet-50 has increased from 76.3% to 82.78%, mCE has increased from 76.0% to 48.9% and mFR has improved from 57.7% to 32.3% compared with the validation package of ILSVRC2012. This contributed to a decrease in the amount of deduction from 536 to 312 (Lee, Won, Lee, Lee, Gu, & Hong, 2020). They use several additional features to deliver cutting edge of results such as 43.5% AP to 65.7% AP50 for Tesla V100's MS COCO dataset with 65 FPS in real time (Bochkovskiy, Wang, & Liao, 2020). We present our custom neural network model and discuss research results. We are finding accurate solution with proposed models with numerical and alpha-numeric tested datasets and our network's cracking accuracy is high at 98.94 percent and 98.31 percent respectively (Nouri & Rezaei, 2020). GreedyNAS is an easy approach and the experimental results of ImageNet data sets indicate, however, that the Top1 accuracy can be improved in the same search field, and the latency ratios, with a supernet training rate of only 60%. Our GreedyNAS can also learn new cutting-edge architectures by searching for a greater space (You, Huang, Yang, Wang, Qian, & Zhang, 2020). This design enables a deep network to be trained from scratch, without using the image grading feature. Two variants of the architecture that are proposed to enable the application in different environments, U2 Net (176.3 MB, 30 FPS GTX 1080Ti GPU) and U2 Net Connection (4.7 MB, 40 FPS). In six SOD datasets, both models achieve reasonable efficiency (Lee C. T., 2019). In ImageNet classification MobileNetV Wide was 3.2% more effectively and latency was reduced by 20% than MobileNetV2. The relatively insignificant MobileNetV3 was 6.6% more accurate than the comparable latency MobileNetV2 model. MobileNetV3 the big detection at the COCO detection is over 25 percent faster than MobileNetV2. MobileNetV3 Wide LRASPP is 34% faster for the segmentation of MobileNetV2 R-ASPP than MobileNetV2 (Howard, et al., 2019). In this article, the architecture that uses k-mean embedding to describe a sequence, a fully convolutional layer and a repeating layer is superior to all other approaches in terms of model accuracy. They offer advice that will assist the professional in finding the right architecture for the mission. At the same time and gives some insight into the discrepancies between the models learned from the convolution and recurrent networks (Trabelsi, Chaabane, & Ben-Hur, 2019). The various NLP attributes can be used to characterise the successful and unsuccessful performance of the CPS. The ML-based framework promotes the development of evidence-based design for collaboration skills mapping which seeks to help teams perform successfully in a dynamic situation (Chopade, Edwards, Khan, Andrade, & Pu, 2019, November). Such findings can be enhanced if data pre-processing methods have been applied in data sets (Agarap, 2019). This paper is limited to supervise machine learning and aims to explain only the basics of this dynamic procedure (Pahwa & Agarwal, 2019,February). The paper suggested Sparse several to one encoder (SF) and a random collaborative face (RF). They focused on presenting the invariant representation of the face and detecting the faces. The author uses Multi PIE pose databases to function on various document you tube datasets (YTF) and data sets in the real world. Output in facial detection and identification increased from 7 to 14% (Shao, Zhang, & Fu, 2017). CNN based model has to address facial detection and recognition problems for categorised web data features categories using a progressive learning algorithm. CNN has increased the efficiency of various types and categorised databases. They also deliver better results in the face detection (Yang, Sun, Lai, Zheng, & Cheng, 2018). An unsupervised framework for learning and a standard framework for optimization, this methodology have improved the co-segmentation mask to increase the characteristics of co-salience. They discuss the idea of objectivity and salience in different forms of multiple images or datasets Cosal2015, iCoseg, Image Pair and MSRC data sets deliver high-quality co-salinity and co-segmentation outcomes (Tsai, Li, Hsu, Qian, & Lin, 2018). This model is improving the importance of structure information in the neural convolution network, offering a better approach for facial identification and face recognition. This algorithm used structural knowledge to take advantage of facial blur and noise. This approach also produces a successful initialization of the ears. Using CNN based approach has more precision to improve distorted faces. This approach also worked on different forms of frameworks (Pan, Ren, Hu, & Yang, 2018). A non-blind deconvolution approach has been proposed to eliminate ringing objects lighted for facial identification and recognition. The non-blind deconvolution process senses light stretches for corrupted images and integrates them into the optimising facial identification and recognition framework. The author used low light conditions to work on png and jpeg images (Hu, Cho, Wang, & Yang, 2014). Blind face detection and identification deblurring algorithms for real faces have been performed and increased image accuracy has been achieved. Usage of a repeated pattern between patches to identify and recognise faces and have a 100% performance rate in the SAuntel, Perrove and Favaro databases (Pan, Sun, Pfister, & Yang, 2017). Deblurring for class-specific issues and blind deconvolution of class genetics has been used to address current process shortcomings and high frequency failure when working with blurred picture. They concentrate only blur id images containing a single object and class-specific training through the use of data sets CMUPIE, Vehicle, FTHZ and INRIA (Anwar, Huynh, & Porikli, 2018). Mixed form vON Mises-Fisher (vMF) is used by CNN for facial identification and recognition. The vMF has a blending model that has been worked on discriminatory characteristics and interprets the relationship between the parameters and the function of the entity. The vMF function is learning and the accomplishment of the discriminatory learning characteristics (Hasnat, Bohné, Milgram, Gentric, & Chen, 2017). Face identification and recognition use Temporary Non-Volume Preservation (TNVP) and Generative Adversarial Network (GAN) in FGNET, MORPH, CACD and AGFW datasets. TNVP measured both the clear synthesising of advanced age faces and the cross-face verification age. The function of attractive density was guaranteed by TNVP. They collected information on features and inferred the importance of consecutive faces in the assessment of embedded datasets (Nhan Duong, Gia Quach, Luu, Le, & Savvides, 2017). In order to detect and classify faces, CNN used the data sets IJB-A, JanusCS2, VGG-NET, CASIA and RLSVRC. Not only is CNN's rendering technique fast, it also increases the accuracy of recognition. It makes the fast generation of CNNs' big faces. There are a lot of required aspect differences (Masi, Hassner, Tran, & Medioni, 2017). Digital Infrared Verses (VSS-NIR) and Invariant Deep Representation (IDR) using the datasets CASIA, NIR-VIS2.0 and Broad Scale VIS Facial Identification and Recognition. The findings are checked by 94% compared with the state-of-the-art VIS results. Just a tiny 64% raises the error rate by 58% (He, Wu, Sun, & Tan, 2017).

3 PROPOSED MODEL FOR FACE DETECTION IN WITH MASK AND WITHOUT MASK

The proposed model (figure2) is representing for identifying the images from dataset in with mask or without mask. Using this model we are identifying with mask and without mask using convolutional neural network (CNN). In CNN we are collect the data from input live or from datasets, after getting input we are pre-processed the input data. Pre-processing is convoluted the input image and trained input image for next session. Filter is generally known as to identify the facial feature using HaarCascade classifier, after these steps we find the image with mask and without mask. If we find without mask, then we go for recognition otherwise we go to previous step.

Figure 2: Proposed model of application to identified with mask and without mask using CNN in deep learning.

Result Analysis: First of all, in contrast to other solutions, we will speak about our motivation and particular challenge. We will then display what is needed to detect people reliably in the feed of the camera and count them. Expect some interesting things like a pose assessment. Then we will immerse ourselves in the way any person wears a with mask or without mask. This is complemented by demonstrating how the use of videos will help us make a better decision on with mask and without mask to over independent frames. Finally, the following steps and potential changes are seen real results in a longer video and brainstorm.

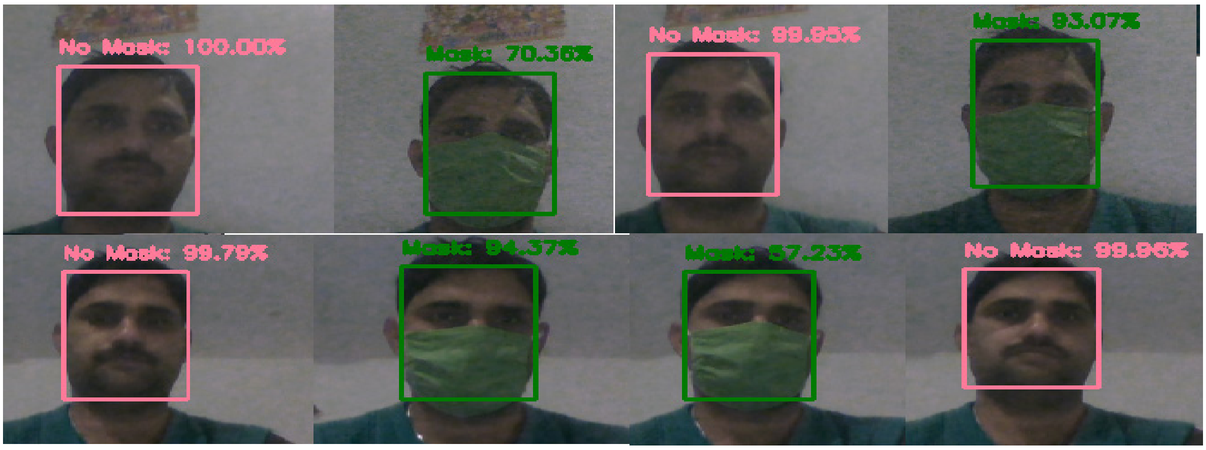

Figure 3: Output of the with mask and without mask in real time processing using proposed model.

In Figure 3 the image is constrained by a neural network like MobileNetV2 and is categorised in two groups around each detected face (with mask/without mask). The pre-trained model tends to be very useful to find out with masks and without mask detection. Our model fits well with details and is 98% mask-free on average, though 90% mask-free on average. There is also no need to retrain the model. We can group the faces of the same person in different frames by tracking. This helps us to run an image size classier and combine the results with a mask and without a mask into a single decision. Our model suit better, the results from Table 1 reveal. We have over 1800 images of masks and without masks to test the accuracy of the model. We are founding a better precision than 98 percent has been found. They are performing better solution compare to other algorithms. We will found the with mask and without mask easily. We are comparing different methodology, than can be used by OpenCV, Pytorch and CNN with different datasets. We find out whether people wear face masks or not. This model has been checked in real time with pictures and videos. Although this model is about 98 percent accurate, model optimization is an on-going task, and by setting hyper parameters we are creating a highly accurate solution. To create the same mobile version, MobileNetV2 was used. This particular model may be used as an implementation case for edge analysis.

Table 1: Output of the accuracy of the with mask and without mask using CNN based deep learning model.

	precision	recall	F1-Score	Support
with mask	0.98	0.98	0.98	385
without mask	0.98	0.98	0.98	394
accuracy			0.98	769
macro average	0.99	0.99	0.99	769
weighted average	0.99	0.99	0.99	769

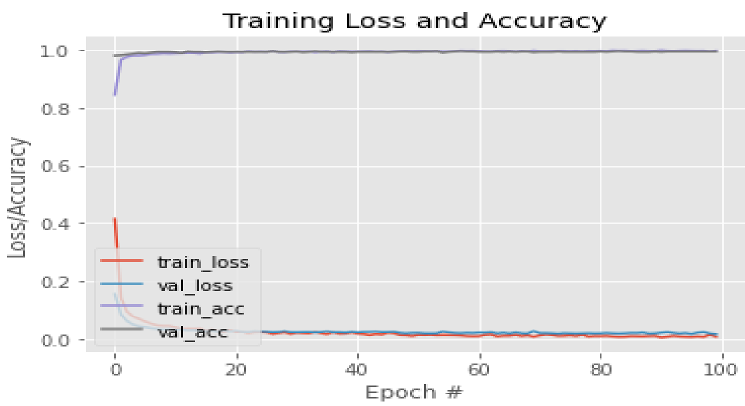

Figure 4 Result analysis of the Loss and Accuracy using given dataset and find out the result.

Figure 4 shows the graph between the accuracy of training and loss. In this figure, the accuracy of the training and the accuracy of the validation continued to increase. Loss and loss of training decreased. The model is still under-fitting. First of all, a common protocol is to configure the network from one task, weights (and bias) with a pre-trained group of data based on a large-scale data set and retrain these parameters for another new target task. With the exception of the first and last layers, these pre-trained weights are sufficient for initialization on all CNN layers, considering the input resolution or the number of classes. The current task dataset can vary from the dataset used for pre-training.

4 CONCLUSION

The proposed model works better than naturally detecting people wearing masks and without mask, that otherwise the face detector would have been unable to detect the faces. Since so much of the face was hidden. This method reduces the vision pipeline to one single step and implements the model of the face mask sensor. All we need to do is apply the object detector in a single network for both with mask and without mask bounding boxes. This Keras classification is based on the MobileNetV2 architecture to classify the features of the faces use of the neural network. The size and computing costs are significantly optimised and are suited for object detection on-device tasks like a cell telephone or webcam in real time.

REFERENCES

Agarap, A. F. (2019). An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv preprint arXiv: 1712.03541.*

Anwar, S., Huynh, C. P., & Porikli, F. (2018). Image deblurring with a class-specific prior. , 41(9), 2112-2130. *IEEE transactions on pattern analysis and machine intelligence, 41*(9), 2112-2130.

Bau, D., Zhu, J. Y., Strobelt, H., Lapedriza, A., Zhou, B., & Torralba, A. (2020). Understanding the role of individual units in a deep neural network. *Proceedings of the National Academy of Sciences, 117*(48), 30071-30078.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934.*

Chopade, P., Edwards, D., Khan, S. M., Andrade, A., & Pu, S. (2019, November). CPSX: Using AI-Machine Learning for Mapping Human-Human Interaction and Measurement of CPS Teamwork Skills. *In 2019 IEEE International Symposium on Technologies for Homela*.

Hasnat, M., Bohné, J., Milgram, J., Gentric, S., & Chen, L. (2017). von mises-fisher mixture model-based deep learning:. *Application to face verification. arXiv preprint arXiv:1706.04264*.

He, R., Wu, X., Sun, Z., & Tan, T. (2017, February). Learning invariant deep representation for nir-vis face recognition. *In Proceedings of the AAAI Conference on Artificial Intelligence, 31*(1).

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., et al. (2019). Searching for mobilenetv3. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1314-1324.

Hu, Z., Cho, S., Wang, J., & Yang, M. H. (2014). Deblurring low-light images with light streaks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3382-3389.

Lee, C. T. (2019). Forging Pathways to Enable Multiscale Modeling of Cellular Scale Phenomena. *University of California, San Diego*.

Lee, J., Won, T., Lee, T. K., Lee, H., Gu, G., & Hong, K. (2020). Compounding the performance improvements of assembled techniques in a convolutional neural network. *arXiv preprint arXiv:2001.06268*.

Liu, X. Y., Yang, H. Q., Chen, R., Zhang, L., & Yang Xiao & Wang, C. D. (2020). FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance. *arXiv preprint arXiv: 2011.09607*.

Masi, I., Hassner, T., Tran, A. T., & Medioni, G. (2017, May). Rapid synthesis of massive face sets for improved face recognition. *In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 604-611.

Nhan Duong, C., Gia Quach, K., Luu, K., Le, N., & Savvides, M. (2017). Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. *In Proceedings of the IEEE International Conference on Computer Vision*, 3735-3743.

Nouri, Z., & Rezaei, M. (2020). Deep-CAPTCHA: a deep learning based CAPTCHA solver for vulnerability assessment . *Available at SSRN 3633354.*

Pahwa, K., & Agarwal, N. (2019,February). Stock market analysis using supervised machine learning. (IEEE, Ed.) *In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* , 197-200.

Pan, J., Ren, W., Hu, Z., & Yang, M. H. (2018). Learning to deblur images with exemplars. *IEEE transactions on pattern analysis and machine intelligence, 46*(6), 1412-1425.

Pan, J., Sun, D., Pfister, H., & Yang, M. H. (2017). Deblurring images via dark channel prior. *IEEE transactions on pattern analysis and machine intelligence, 40*(10), 2315-2328.

Shao, M., Zhang, Y., & Fu, Y. (2017). Collaborative random faces-guided encoders for pose-invariant face representation learning. *IEEE transactions on neural networks and learning systems, 29*(4), 1019-1032.

Trabelsi, A., Chaabane, M., & Ben-Hur, A. (2019). Comprehensive evaluation of deep learning architectures for prediction of DNA/RNA sequence binding specificities. *Bioinformatics, 35*(14), i269-i277.

Tsai, C. C., Li, W., Hsu, K. J., Qian, X., & Lin, Y. Y. (2018). Image co-saliency detection and co-segmentation via progressive joint optimization. , 28(1), . *IEEE Transactions on Image Processing, 28*(1), 56-71.

Yang, J., Sun, X., Lai, Y. K., Zheng, L., & Cheng, M. M. (2018). Recognition from web data: A progressive filtering approach. *IEEE Transactions on Image Processing, 27*(11), 5303-5315.

You, S., Huang, T., Yang, M., Wang, F., Qian, C., & Zhang, C. (2020). Greedynas: Towards fast one-shot nas with greedy supernet. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1999-2008.