Research Article

Performance Analysis of Different Architectures on Face Mask Detection

Dr.D.Lakshmi Padmaja[a], G K Sriharsha[b], Dr G N V Ramana Rao[c], Dr K Sudheer Reddy[d]

[a]lakshmipadmajait@cvsr.ac.in, [b]gksriharsha@gmail.com,[c] gnvramanarao@gmail.com, [d]sudheercse@gmail.com

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract: The paper focus on the Face Mask Detection by which we can differentiate people who are wearing mask and not wearing mask. Now a days due to covid19 everyone is instructed to wear mask. This Face Mask Detection Model we can detect people who are not wearing masks. The people who have Diabetic, Hyper tension (BP) and lung diseases are easily affected by the covid19 virus. This study, mainly focuses on how the Convolution Neural Network(CNN) model created,and use the model to identify whether the person is with or without mask.

Keywords: Face Mask Detection; Health, Image Processing; Human Face Detection; CNN Architectures

1. Introduction

Due to the covid19 everyone is restricted to where mask when they are in public places and when they are leaving house for any purpose. But some of the people are refused to wear mask even when are outside or in public places. Some of the people doesn't wear mask properly. We can use CNN [1] and Image processing techniques to identify the who are not wearing mask in the public places and we can warn them to wear the mask. Now the mall, parks and other public places are opened we can use Face Mask Detection Models which are developed using CNN model [1] and image processing to identify the people who are not wearing masks and we can warn to wear mask and we can also where the person is wearing properly or not. We can use the models in entrance of shopping malls and in other places also. The rest of the paper is arranged as follows: Section 2- Literature Survey. Section 3- Challenges/Motivation, Section 4-Existing Methods, Section 5-Proposed Method, Section 6-Experimentation/ Results, Section 7-Discussion and Section 8- Conclusion

2. Literature Survey

The CNN are widely used in creating a model architectures and CNN are used when dataset consists of image. Yamashita, R.K.G et al published Convolutional neural networks: an overview and application in radiology paper [1] which explains structure how the CNN architectures are created and the functioning of the CNN.

The haar cascade algorithms are used for detecting the human face. Padilla et al published Evaluation of Haar Cascade Classifiers for Face Detection paper [2]. This paper says how the haar cascade is came into existence and how it works.

3. Challenges / Motivation

Due to covid19 people are wearing masks to the stop spreading of virus but many people are refused to wear mask as it not comfortable or any other reason. But due to people who are not wearing mask the virus is spreading. We can use Machine Learning and Deep Learning techniques to develop models and light weight applications which can identify the person not wearing mask and the people who are not wearing mask properly and warn them to wear the mask.

4. Existing Methods

The Conv2D, MaxPooling2D, Flatten, Dense algorithms are already available in the keras layers module. Those algorithms are used to create the CNN model [1]. By using ImageDataGenerator algorithms preprocessing image module. The activation function relu [9] and softmax [9], loss function binary_crossentropy [12] and mse [11], optimizer adam [10] and sgd [10] algorithms are already written and used by importing the keras. To detect live image the open cv library is used which is already built. The haar cascade algorithm [1] is used for human face detection which is already built.

The various CNN architectures VGG16 [3], Inception_v3 [5], ResNet50 [6,7], Mobilenet [4] are already available in the keras applications module.

Table 1: Different Architectures

S.NO	ARCHITECTURE	PARAMETERS	DEPTH	SIZE (mb)
1	VGG16	138,357,544	23	528

2	InceptionV3	23,851,784	159	92
3	ResNet50	25,636,712	-	98
4	Mobilenet	4,253,864	88	16

In the Table 1, the architecture refers to the structure of the model and how the layers are arranged and connected. Depth refers to the topological depth of the network. This includes activation layers, batch normalization layers etc. The datasets are taken from the Kaggle.

Figure 1: Image data sets with mask and without mask

FACES WITHOUT MASK

FACES WITH MASK

5. Proposed Method

Using Conv2D, MaxPooling2D, Flatten, Dense algorithms the CNN architecture and model is created and the model is trained on data sets by categorical class mode. To detect the image is with or without mask we convert the image into gray color model and gray colored image is given to the model. Among VGG16 [3], Inception_v3 [5], ResNet50 [6], Mobilenet [4] which one fits better on taken dataset and produces good results.

DATASET DESCRIPTION The dataset collected has 4591 images and among them 2538 images are with mask and 2053 images are without mask (Figure 1). 80% of images are used for the training set and 20% of images are used for the validation set.

TRANSFER LEARNING Transfer learning is the feature extraction from the pretrained models and network. The CNN architectures VGG16 [3], Inception_v3 [5], Mobilenet [4], Resnet50 [6] are the pretrained networks. Their network is not modified but they are trained on the face mask dataset collected.

5. Experimentation / Results

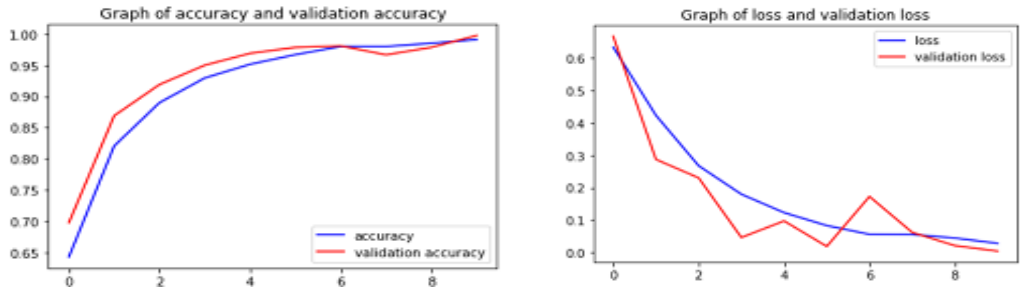
The training set is used for training of the model and validation set is used for testing of the model. Training the model will undergo many iterations and those iterations are called epochs. Epoch number is the nth iteration of training the model. Accuracy and loss are obtained when train dataset is given to the model. Validation accuracy and Validation loss are obtained when test dataset is given to the model. Time is the number of seconds taken to complete the model fit in that epoch. In starting epoch, the accuracy will be less and loss will be more but in final epoch the accuracy grows and loss reduces from Table 2. From Figure 2: Model recognizing with and without mask of an image. The graphs drawn for Table 2 accuracy and loss. After validation the accuracy increased based on the number of epochs increased. The results of various architectures on the taken dataset. In table 3 the architecture is VGG16 [3], Inception_v3 [5], ResNet50 [6], Mobilenet [4]. The accuracy in below table refers to the highest accuracy obtained in 10 epochs and time refers to the average time taken to complete one epoch. The optimizer is used converge the weights to obtain correct output sgd means stochastic gradient descent and loss is used know the difference between obtained results and exact results. mse means mean squared error(mse).

Table 2: Results calculated for accuracy and time

S.NO	EPOCH NUMBER	ACCURACY	LOSS	VALIDATION ACCURACY	VALIDATION LOSS	TIME (sec)
1	2	82%	42%	86%	28%	116
2	3	89%	26%	91%	23%	137
3	4	92%	18%	95%	04%	133
4	5	95%	12%	96%	09%	102
5	6	96%	08%	97%	01%	101
6	7	98%	05%	98%	17%	133
7	8	98%	05%	96%	06%	290

In starting epoch, the accuracy will be less and loss will be more but in final epoch the accuracy grows and loss reduces from Table 2.

Figure 2: Graphical representation of accuracy and loss(data from Table 2 taken).



Note: x-axis is epoch number and y-axis are the accuracy and loss

From Figure 2: Model recognizing with and without mask of an image. The graphs drawn from accuracy and loss. After validation the accuracy increased based on the number of epochs increased. And loss is decreased when for the number of epochs increased, except for the number of epochs is 6.

The results of various architectures on the taken dataset. In below table the architecture is VGG16 [3], Inception_v3 [5], ResNet50 [6,7], Mobilenet [4]. These all architectures are fit on dataset collected and number of epochs is 10. The accuracy in below table refers to the highest accuracy obtained in 10 epochs and time refers to the average time taken to complete one epoch. The optimizer is used converge the weights to obtain correct output sgd means stochastic gradient descent and loss is used know the difference between obtained results and exact results. mse means mean squared error(mse).

$$\text{Precision} = \frac{tp}{tp + fp} \qquad \text{Recall} = \frac{tp}{tp + fn}$$

Where,

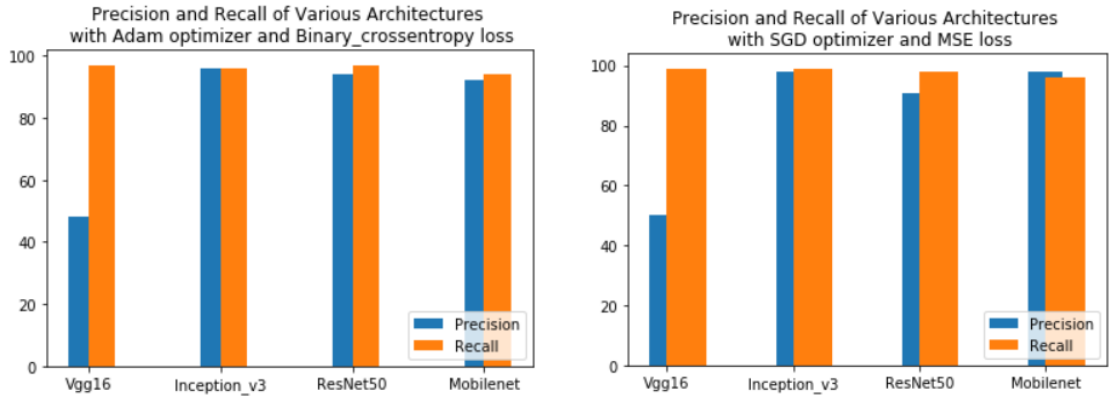
- t_p refers to true positive
- f_p refers to false positive

fn refers to false negative

Table 3: Performance of different Architectures with different optimizers

S. No	Architecture	Optimizer	Loss	Accuracy	Precision	Recall	Time (Sec)	Maximum Epochs
1	VGG16	Adam	Binary_crossentropy	56%	48%	97%	1092	10
		sgd	mse	52%	50%	99%		
2	Inception_v3	Adam	Binary_crossentropy	97%	96%	96%	585	10
		sgd	mse	99%	98%	99%		
3	ResNet50	Adam	Binary_crossentropy	99%	94%	97%	1456	10
		sgd	mse	98%	91%	98%		
4	Mobilenet	Adam	Binary_crossentropy	98%	92%	94%	456	10
		sgd	mse	97%	98%	96%		

Figure 3: Precision and Recall of various architectures for different optimizers and loss functions



The Figure 3 gives information about how various CNN architectures performed on with mask and without mask dataset.

The Table 3 gives information about how various CNN architectures performed on dog/cat dataset to identify given image is cat or dog. This dataset has 3000 images 1500 images are dog and other are cat.

Table 4: Performance analysis of various architectures on dog/cat dataset

S. No	ARCHITECTURE	PRECISION	RECALL
1	VGG16	60%	85%
2	Inception_v3	63%	88%
3	ResNet50	75%	77%
4	Mobilenet	90%	82%

From Table 4 on dog and cat data sets experimented with different architectures to evaluate the performance analysis for precision and recall. From the results Mobilenet precision is more appropriate architecture and Inception_v3 is good for Recall.

6. Discussion

From the Table 1 various parameters name of the architectures, number of parameters, depth, and sizes are placed. If we observe a smaller number of layers used VGG16 and less size used in Mobilenet. Figure 1 is collected images with mask and without mask for calculating precision and recall. The idea about this paper is face mask detection with and without mask and compared with various CNN architectures. From the results placed in Table 2, if the number of epochs is increased accuracy is increased. From the Figure 2 we can observe the accuracy and loss of the Table 2. Precision and Recall are experimented and shown in the Table 3 with different architectures and different optimizers. From the table ResNet50 architecture, Adam optimizer and Inception_v3 architecture and sgd optimizer obtained 99%. As time wise less time taken for Mobilenet architecture next Inception_v3. Figure 3 represents pictorial representation of Table 3 results.

7. Conclusion

Using Deep Learning and Image Processing many models can be created which can makes things easy and efficient. In this covid19 time using the models created by above technologies are very useful like face mask detection, social distancing. From this paper, we conclude that, among VGG16 [3], Inception_v3 [5], ResNet50 [6], Mobilenet [4] architectures ResNet50 gives top accuracy but time taken to train the model is high and space occupied is also more. The Mobilenet also gives high accuracy but little bit less than the ResNet50 [6] but time taken to train model and space occupied is very less compared to ResNet50 [6].

References

1. Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9
2. Padilla, Rafael & Filho, Cicero & Costa, Marly. (2012). Evaluation of Haar Cascade Classifiers for Face Detection.
3. Tammina, Srikanth. (2019). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. International Journal of Scientific and Research Publications (IJSRP). 9. p9420. 10.29322/IJSRP.9.10.2019.p9420.
4. Khasoggi, Barlian & Ermatita, Ermatita & Sahmin, Samsuryadi. (2019). Efficient mobilenet architecture as image recognition on mobile and embedded devices. Indonesian Journal of Electrical Engineering and Computer Science. 16. 389. 10.11591/ijeecs. v16.i1.pp389-394.
5. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826
6. Y. Ma, P. Zhang and Y. Tang, "Research on Fish Image Classification Based on Transfer Learning and Convolutional Neural Network Model," 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 2018, pp. 850-855, doi: 10.1109/FSKD.2018.8686892.
7. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778
8. M. Jiang, X. Fan and H. Yan, Retinamask: A face mask detector, 2020.https://arxiv.org/abs/2005.03950
9. Chigozie Enyinna Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. (2018). Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. https://arxiv.org/abs/1811.03378
10. S. Sun, Z. Cao, H. Zhu and J. Zhao, "A Survey of Optimization Methods From a Machine Learning Perspective," in IEEE Transactions on Cybernetics, vol. 50, no. 8, pp. 3668-3681, Aug. 2020.
11. doi: 10.1109/TCYB.2019.2950779
12. (2011) Mean Squared Error. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_528
13. Y. Zhou, X. Wang, M. Zhang, J. Zhu, R. Zheng and Q. Wu, "MPCE: A Maximum Probability Based Cross Entropy Loss Function for Neural Network Classification," in IEEE Access, vol. 7, pp. 146331-146341, 2019.
14. doi: 10.1109/ACCESS.2019.2946264