JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Comprehensive Analysis of Various Deep Learning Based Facemask Detection Techniques

Trupal Chaudhary[1], Om Brahmbhatt[2], Qusai Onali[2]
[1]*B. Tech. Student, U & P.U. Department of Computer Engineering, CSPIT, Anand, India*
[2]*B. Eng. Student, Department of Information Technology, GCET, Anand, India*

Abstract: Uniform spread of the coronavirus has caused a sustained increase in the number of victims in different countries of the world since the day it took off. Detecting criminals is extremely necessary to ensure that the spread of the virus is constantly reduced. This article shows the use of deep learning methods which deal with different person who does not wear a facemask. This mask detection dataset includes with and without mask images, we will perform face detection in real time from a live feed through our webcam. We will likely find out if the person is wearing a face mask in still images, as well as in computer video streaming and deep learning procedures. To think of a framework that will be updated broadly, multiple classifiers with multiple optimizers must be evaluated. This research article presents a review of classifiers such as SGD, ADAM, MobileNetV2, each with ADAGRAD, VGG16 and RESNET50 optimizers.

***IndexTerms* - Python, Deep Learning, Computer Vision, OpenCV, Tensor Flow, Keras, Pytorch, COVID-19**

I. INTRODUCTION

According to World Health Organization (WHO) situation reports, Coronavirus infection 2019 (COVID-19) has infected over 150,000,000 people and resulted in over 3,170,000 deaths. Furthermore, there have been a number of comparative large-scale true respiratory illnesses in recent years, such as severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS). It was stated that the Coronavirus conceptive number is higher than that of SARS. As a result, an increasing number of people are concerned about their well-being, and governments regard general well-being to be their top priority.

Furthermore, several public service providers require clients to access the service only if they are wearing face masks. As a result, mask identification has become a key duty for assisting the worldwide society, yet mask detection research is limited. Mask detection points are used to determine whether or not a person is wearing a mask and where the mask is located. The problem is clearly defined as generic article identification, which is used to identify several types of things, and face discovery, which is used to recognize a specific type of item, such as a face. Article and face discovery are used in a variety of settings, including self-driving, education, and reconnaissance.

Deep learning-based article identities have recently demonstrated exceptional performance and deal with the advancement of present item identifiers. Deep learning allows neural networks to learn highlights in a start-to-finish manner without the use of previous information for molding component extractors. There are two types of important learning-based article identifiers: one-stage and two-stage. Single shot identifier (SSD) and You Just Look Once are examples of one-stage indicators that use a single neural organization to distinguish items (YOLO).

Two-stage indicators, on the other hand, use two organizations, such as Area Based Convolutional Neural Organization (R-CNN) and faster R-CNN, to play out a coarse-to-fine location. Face detection, in essence, receives similar engineering to broad article locator, but includes extra face-related highlights, such as facial milestones, to boost face identification accuracy. Nonetheless, there is a rare inspection that focuses on the placement of the face mask.

In the midst of this, a performance evaluation of multiple facemask detection frameworks executing various classifiers is necessary, so the finest framework in terms of exactness and loss can be used to achieve large-scale success [6]. The proposed model can be used in conjunction with spy cameras to block COVID-19 transmission by allowing the identification of people wearing or not wearing face masks. Traditional Machine Learning and Deep Learning are combined in the model. For feature extraction, we used Deep Transfer Learning with three traditional machine learning methods.

II. CONCEPTS

A. Computer Vision

Computer vision is interdisciplinary logical domain that manages how PCs can acquire a significant level of understanding from images or computer recordings [5]. From a design point of view, it tries to understand and robotize the careers that the human visual framework can perform. Delivering mathematical or emblematic data, for example in the types of choice, understanding the context implies the change of visual images into representations of the world that augur many ways of thinking and can evoke an appropriate activity. This understanding of images can be viewed as the unraveling of iconic data from image information using models developed with the guidance of math, physical science, measurement, and learning hypotheses.

The purpose of computer vision is to figure out what advanced imaging is all about. Image data can have a variety of forms, such as video footage, several camera views, or multidimensional information from a clinical scanner. Computer view aims to make the building of computer vision frames mechanically controlled by its suppositions and models.

B. Deep Learning

Deep learning algorithms are aimed at taking into account highlight chains with climaxes at higher levels of the hierarchical order [2], which are produced by the arrangement of lower-level attributes. Deep learning is simulated intelligence activity that mimics the human mind's operations in preparing information for use in distinguishing things, hearing speech, deciphering dialects, and making decisions. Deep learning-based computational intelligence can learn without human supervision, relying on unstructured and untagged data. Deep learning decodes large amounts of unstructured data that would otherwise take a long time to comprehend and measure.

The picture is deep, with many layers, if we were to make a graph depicting how these ideas build on each other. As a result, we refer to this approach to dealing with simulated intelligence as deep learning. Deep learning reigns supreme in challenging environments where data sources (and, shockingly, performance) are simple. That is, it is images of pixel information, text information reports, or sound information recordings, not just a few variables in a simple arrangement. Deep learning allows computer models with multiple levels of preparation to learn information representations with varied degrees of thoughtfulness.

C. OpenCV

OpenCV is a computer vision and artificial intelligence programming library. OpenCV was created as the foundation for computer vision applications and to enable companies to apply machine judgment to trade items more quickly. Because OpenCV is a BSD licensed project, organizations can easily use and modify it. The collection includes more than 2,500 optimized algorithms, covering a wide range of traditional and sophisticated machine learning and computer vision techniques. These algorithms can be used to identify things, to classify human behavior in a video, to extract 3D models from stereo camera items, to track moving objects, to pin together images in order to create full scene resolution images.

D. TensorFlow

TensorFlow is a programming framework for open-source data flow that is capable of meeting different challenges. It is a library used in neural networks and other apps for machine learning. The second-generation Google Brain system, TensorFlow. The latest release was released in February. Unlike the implementation of reference, TensorFlow can run on several CPUs and GPUs.

III. PROPOSED SYSTEMS

The proposed method uses a face mask to identify the individual in real-time image / video sequences using computer vision classifiers and deep learning. [15] The proposed method considers a data collection of 1300 size photographs.

Fig. 1. Facemask Dataset Images

Fig. 2. Facemask-free images data set

Photos are classified using a two-step trained template:

i. Phase 1: To create a ready model, the mask dataset is stacked in the frame and various classifiers such as MobileNetV2, ResNet50 and VGG 16 are used.
ii. Phase 2: Load the mask classifier template, search for the faces in image / video sequence, apply classifier to all the faces, and confidently categorize as "hidden" or "unmasked".

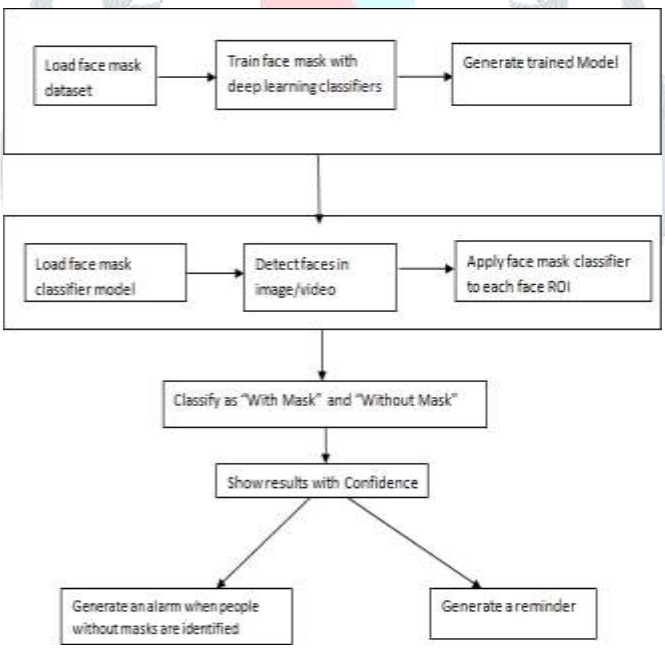

Fig. 3. Architecture of the detection system for the face mask

Best performance should be considered when selecting a system. Consequently, the above performance factors can be taken into account when designing optimal systems for large-scale deployment.

Following *Classifiers* were tested in this system:

1. ***MobileNetV2:***

This is a convolutional neural organization engineering that seems to work well on mobile phones. Depends on an inverted remaining layout where persistent associations are between bottleneck layers.

The mid-spread layer uses light depth smart convolutions to channel glare as a source of non-linearity. In general, MobileNetV2 engineering contains the underlying fully convolutional layer with 32 channels, followed by 19 persistent bottleneck layers. [11]

This classifier uses the distinguishable convolution of Profundity, which can greatly reduce the cost of the complexity and the size of the organization model, and therefore it is reasonable for cell phones.

It's time to introduce the inverted residual structure. In narrow layers, non-linearity is eliminated. MobileNetV2 as the feature extraction backbone yields the best results for object detection and semantic segmentation.

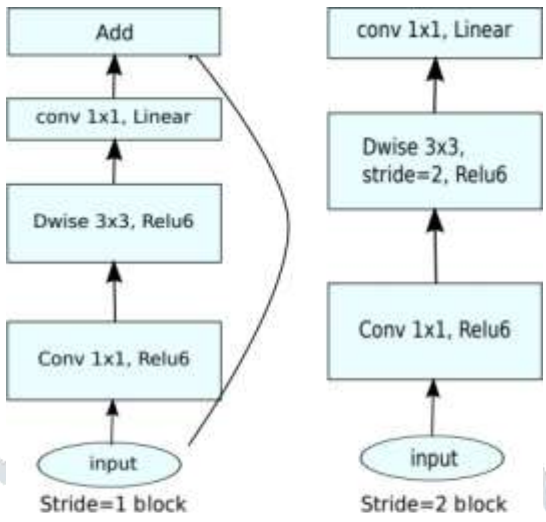

Fig. 4. MobileNet V2 Architecture

2. *ResNet50:*

Resnet is the short name for the Residual Network that supports residual learning. The 50 shows how many layers it has. Therefore, Resnet50 represents persistent organization with 50 layers. The deep convolutional networks caused a series of advances in the order of the images. Typically, the model goes through a deeper number of layers to fine-tune complex mappings and increase layout precision and recognition.

However, as we delve deeper into neural organizations, precision begins to sink and then becomes corrupted. Persistent preparation tries to fix this problem. In a deep convolutional global neural organization, many layers are stacked and prepared for the work to be done. In persistent learning, instead of trying to become familiar with certain strengths, try to master certain leftovers. The rest can be seen mainly as a deduction of the highlight obtained from the contribution of this layer.

ResNet does this by using alternate path associations (direct contribution of the layer interface to a few $(n + x)^{th}$ layers [11]. It has shown that preparing such organizations is simpler than preparing of simple convolutional layers deep neural organizations and furthermore solves the problem of degrading precision.

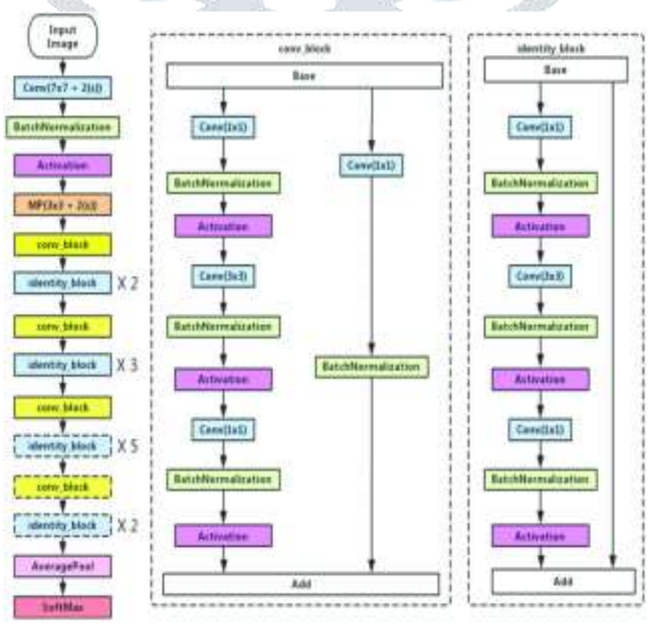

Fig. 5. ResNet 50 Architecture

3. VGG16:

VGG16 is a convolutional neural organization model proposed by K. Simonyan and A. Zisserman. The model achieves a test accuracy of 92.7% in the top 5 on ImageNet, which is a data set of more than 14 million images that have a place with 1000 classes.

The most innovative thing about VGG16 is on behalf of having countless hyper boundaries, they focused on 3x3 channel convolutional layers with step 1. It follows this layered and convolutional plane - maximum pool capacity reliably across engineering. Finally, there are two FCs (fully associated layers) pulled by a SoftMax for performance. The 16 in VGG16 refers to 16 layers that have fillings.

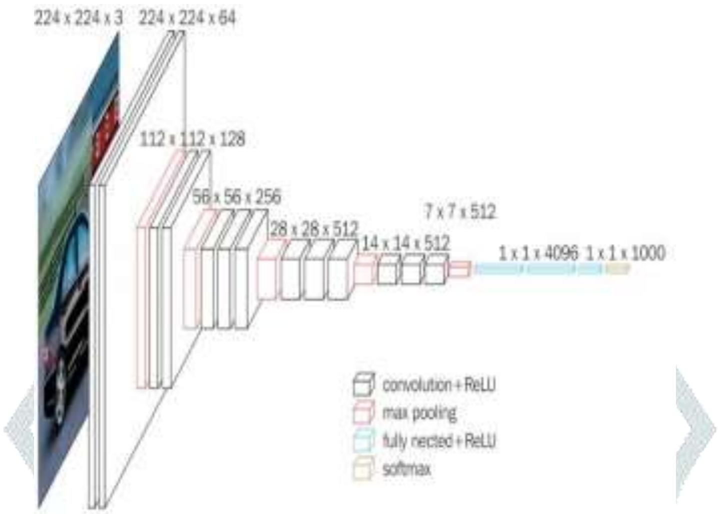

Fig. 6. VGG 16 Architecture

To test the system's performance, three optimizers were applied to each of these classifiers:

a) ADAM:

Adam is an algorithm for on-demand slope-based improvement of stochastic target capabilities, which relies on general-purpose evaluations of lower-demand minutes. This approach is computationally efficient and requires little memory to run.

This is a corner-to-corner resizing invariant, suitable for problems that are huge in terms of information and boundaries. Hyperbounds have knee-jerk translations and generally require little adjustment. The exact results show that Adam works admirably and takes a good look at other stochastic advance techniques.

b) ADAGRAD:

Adagrad is a parser with explicit limit learning rates, which adjust based on how often a limit is revised during setup. The more updates a cap receives, the more modest the updates will be. [14]

c) SGD:

In Stochastic Slope Drop, some examples are chosen at random rather than a complete collection of information for each cycle. In Slope Plunge, "cluster" indicates the total number of tests in a data set that is used to calculate the angle for each accent [18]. Using the full data set is useful for reaching the minima in a quieter and less arbitrary way.

The challenge arises when the datasets get too big. Assume that the dataset contains millions of examples, to use a normal slant drop improvement strategy, use all 1,000,000 examples to cycle while playing angle drop, and this should be accomplished for each accent until they are reached the minimums. Subsequently, it turns out to be extravagant in terms of the calculation to be made. This problem can be solved using the Stochastic Angle Plunge.

In SGD, you only use a lone example, that is, a group size of one, to reproduce each loop. The example is rearranged at random and chosen to reproduce the cycle. In SGD, since only one example of the data set is selected indiscriminately for each cycle, the computation's path to the minima is generally noisier than a standard elevation gain calculation.

IV. RESULTS

The following classifiers and optimizers are used to evaluate the experimental results of the system performance:

Table 1. Results of the MobileV2 classifier system proposed

Classifier	Epochs	Train/test size	Optimizer	Train loss	Train Accuracy	Test Loss	Test accuracy
Mobilenet V2	20	90/10	ADAM	0.0090	0.9981	0.0071	1.0000
			ADAGRAD	0.2454	0.9148	0.1811	0.9819
			SGD	0.1549	0.9502	0.0216	0.9855

When compared to the other two optimizers ADAGRAD and SGD, in both training and testing, the ADAM optimizer performs well., as shown in Table 1.

Fig. 7. MobilenetV2 with ADAM Optimizer Training Loss and Accuracy Plot and Results

Table 2. The proposed system's results with the ResNet-50 Classifier

Classifier	Epochs	Train/test size	Optimizer	Train loss	Train Accuracy	Test Loss	Test accuracy
Resnet50	20	90/10	ADAM	0.0068	0.9975	0.0557	0.9856
			ADAGRAD	0.1087	0.9693	0.0019	1.0000
			SGD	0.1114	0.9693	0.0100	1.0000

In comparison to the other two optimizers, ADAGRAD and SGD, the ADAM optimizer performs well in both training and testing, as shown in Table2, and all test accuracies are good.

Fig. 8. ResNet-50 with ADAM Optimizer Training Loss and Accuracy Plot and Results

Table 3. The proposed system's results with the VGG-16 Classifier

Classifier	Epochs	Train/test size	Optimizer	Train loss	Train Accuracy	Test Loss	Test accuracy
VGG16	20	90/10	ADAM	0.2145	0.9826	0.0006	1.0000
			ADAGRAD	1.7911	0.8425	0.4243	0.9638
			SGD	0.5133	0.9536	0.1055	0.9928

When compared to the other two optimizers ADAGRAD and SGD, in both training and testing, the ADAM optimizer performs well, as shown in Table 3, although SGD test accuracy is roughly similar to ADAM.

Fig. 9. VGG-16 Results with ADAM Optimizer and Training Loss and Accuracy Plot

V. CONCLUSIONS

i. According to the results of the different classifiers, the performance of the ADAM Optimizer is fairly good, and the test accuracy of SGD is nearly equivalent to ADAM for the three classifiers described above.

ii. During testing, it was discovered that the MobileNetV2 classifier produces the greatest results with the highest accuracy.

iii. Face mask detection, which would allow us to detect whether or not someone is wearing a mask and admit them, would be immensely valuable to society.

VI. FUTURE SCOPES

The present system is evaluated using various classifiers. The best system, as well as the interaction with the alarm and alert systems, could be built in the near future.

This system could be integrated with a system of social separation to create a healthy system that has a significant impact on disease transmission [16].

The aforementioned use cases are only a handful of the many possibilities accessible with this solution. Many more scenarios, we believe, might be added to our system to provide people a fuller sense of security. A list of some of the features that are currently in development may be found below.

i. Coughs and sneezes: According to WHO standards, chronic coughs and sneezes are one of main signs of COVID-19 infection and one of main routes of disease transmission to the uninfected public [9]. By augmenting our proposed method with body gesture analysis to determine whether a person is sneeze and coughing in public locations while removing the face mask and distance requirements, the deep learning-based technique can be beneficial in detecting and limiting the transmission of disease. social and reliant Results can be communicated to execution authorities.

ii. Another major indicator of COVID-19 infection is an increase in body temperature; today, thermal monitoring is done utilising non-contact handheld infrared thermometers where the healthcare practitioner should be. Closer examination is required, putting health care professionals at risk of infection. Furthermore, because it is nearly difficult to capture the temperature of every individual in the public places, proposed use case can be equipped with image-based detection thermals to assess the temperature of the body of persons in the public places, which can aid law enforcement organisations. in order to properly combat the epidemic.

VII. REFERENCES

[1] N. H. Leung, D. K. Chu, E. Y. Shiu, K.-H. Chan, J. J. McDevitt, B. J. Hau, H.-L. Yen, Y. Li, D. KM, J. Ipet al., "Respiratory virus shedding in exhaled breath and efficacy of face masks."

[2] S. Feng, C. Shen, N. Xia, W. Song, M. Fan, and B. J. Cowling, "Rational use of face masks in the covid- 19pandemic,"

The Lancet Respiratory Medicine, 2020.

[3] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Peiet al., “Masked face recognition dataset and application,” arXiv preprint arXiv:2003.09093, 2020.

[4] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, “Object detection with deep learning: A review,” IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212–3232, 2019.

[5] A. Kumar, A. Kaur, and M. Kumar, “Face detection techniques: a review,” Artificial Intelligence Review, vol. 52, no. 2, pp. 927–948, 2019.D.-H. Lee, K.-L. Chen, K.-H. Liou, C.-L. Liu, and J.-L. Liu, “Deep learning and control algorithms of direct perception for autonomous driving”, 2019.

[6] S. Wang Chen, Horby Peter W, Hayden Frederick G, Gao George F. A novel coronavirus epidemic of global concern for health.2020;395(10223):470-473. 10.1016 / S0140-6736(20)30185-9.

[7] Chen, S., Zhang, C., Dong, M., Le, J., R., M., 2017b. Using ranking CNN for age estimates, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[8] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2: Inverted residues and linear bottlenecks," IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510-4520. 2018.

[9] Zhu, X. 2006. Semi-supervised study of learning literature survey. Computer Science, University of Wisconsin-Madison 2(3):4.

[10] X. Jia, "Image recognition method based on deep learning," 2017, 29th Chinese Control and Decision Conference (CCDC), Chongqing, 2017, pp. 4730-4735 doi:10.1109/CCDC.2017.7979332.

[11] Adaptive Sub gradient Methods for Online Learning and Stochastic Optimization - John Duchi, Elad Hazan, Yoram Singer.

[12] Image Classification Using Convolutional Neural Networks Deepika Jaswal, Sowmya.V, K.P.Soman.

[13] A. Kumar, A. Kaur, and M. Kumar, “Face detection techniques: a review, “Artificial Intelligence Review, vol. 52, no. 2, pp. 927–948, 2019.D.-H. Lee, K.-L. Chen, K.-H. Liou, C.-L. Liu, and J.-L. Liu, “Deep learning and control algorithms of direct perception for autonomous driving, 2019.

[14] Ejaz, M.S., Islam, M.R., Sifatullah, M., Sarker, A., Implementation of principal component analysis on masked and non-masked face recognition, 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT),2019, pp. 1–5.

[15] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278-2324, November 1998.

[16] Dalal, N., and Triggs, B., Histograms of oriented gradients for human detection, CVPR '05, 2005.

[17] Girshick, R., Fast R-CNN, Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.