Novel Face Mask Detection Technique using Machine Learning to control COVID'19 pandemic

Sandeep Gupta [a], S.V.N. Sreenivasu [b], Kuldeep Chouhan [c], Anurag Shrivastava [d,*], Bharti Sahu [e], Ravindra Manohar Potdar [f]

[a] Department of Electrical Engineering, JECRC University, Jaipur, India
[b] Department of Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India
[c] Department of Computer Science and Engineering, I.T.S Engineering College, Greater Noida, India
[d] Lakshmi Narain College of Technology and Science Indore, India
[e] Department of Computer Science and Engineering, Dr. D.Y. Patil Institute of Technology, Pimpri, Pune, India
[f] Department of Electronics and Telecommunication Engineering, Bhilai Institute of Technology, Durg, Chhattisgarh, India

ARTICLE INFO

ABSTRACT

The COVID-19 pandemic has been scattering speedily around the world since 2019. Due to this pandemic, human life is becoming increasingly involutes and complex. Many people have died because of this virus. The lack of antiviral drugs is one of the reasons for the spreading of COVID-19 virus. This disease is spreading continuously and easily due to some common mistakes by people, like breathing, coughing and sneezing by infected persons. The main symptom is the normal flu. Therefore, in the present condition, the best precaution for this disease is the face mask, which covers both areas of mouth & nose. According to the government and the World Health Organization, everyone should wear a face mask in busy places like hospitals and marketplaces. In today's environment, it's difficult to tell if someone is wearing a mask or not, and physical inspection is impractical since it adds to labour costs. In this research, we present a mask detector that uses a machine learning facial categorization system to determine whether a person is wearing a mask or not, so that it may be connected to a CCTV system to verify that only persons wearing masks are allowed in.

1. Introduction

As of 2020, respiratory infectious disease is spreading over the world, affecting the entire population. This virus causes respiratory tract infections in humans, which can range from mild to fatal. This sickness has spread throughout China and is now afflicting the entire population. Every country is working hard to develop a vaccine that will either prevent or cure this disease. This disease is dangerous because it is easily transmitted through simple means, such as visiting public areas, touching infected objects, coming into close contact with an infected person, or their nasal and cough droplets. This is how anyone can become sick and subsequently infect others. As a result, the authorities advised wearing masks whenever going out in public places and maintaining social distance, both now and in the future, as a precautionary measure to protect yourself from this sickness. It is now required, but some are breaching the law by not wearing a mask as they walk out, necessitating the presence of a person at the entrance to verify whether or not a person is wearing a mask. Security personnel, on the other hand, will be at risk, as it is impossible to track every individual, and any inaccuracy may accelerate virus propagation. In our investigation, we used a detector for the face with a mask; typical detectors for auto-focusing [1,2] and human–computer interaction [3,4] are available. For example, in [5], C2D-CNN is used to distinguish a person with a mask or a covered face using a mask detection technique. In essence, we're using a face detection algorithm to distinguish human faces in digital data, such as a picture or a video clip. It can also be considered an example of object-class detection. Face recognition has long been a difficult challenge for computer vision,

* Corresponding author.
E-mail address: anuragshri76@gmail.com (A. Shrivastava).

S. Gupta, S.V.N. Sreenivasu, K. Chouhan et al.

but as technology progresses, it is becoming easier [6,7]. There was semi-automated equipment for face identification, just like in the 1960s. The first semi-automatic facial recognition system was deployed in 1988, and the automated system was displayed at the Super Bowl event in 2001, capturing surveillance photographs and attracting the attention of a large number of people. Nowadays, everyone uses this technology for various reasons and current trends, and it is extremely important in the IT sector and the field of computer vision. It's employed in practically every task, from simple camera detection to a wide range of scientific applications. Face detection techniques include feature-based, appearance-based, knowledge-based, and template-matching methods, among others. Deep learning approaches such as face detection of masked and non-masked people, as well as image categorization of both datasets, are used in our research. This may be a notification system that works with any CCTV or sensor, allowing only those wearing masks on their faces to enter. We employed two datasets in this study, one with people wearing masks and the other without masks, and we used numerous stages, such as data pre-processing, face detection, and image classification utilizing convolution neural networks and the Alex Net architecture. This study can be used in a variety of settings, including supermarkets, schools, hospitals, and other public areas, to detect those who are breaking the rules and to make it easier for security officials to follow each and every individual passing through.Tables 1 and 2

2. Database used

The Real-World Masked Face Dataset (RMFD) [8], which is accessible on a GitHub repository and is the world's biggest masked face dataset, is the first dataset we use for masked persons. Celeb Faces Attributes (CelebA) [9] is the other dataset we used. As

Table 2
Metrics on testing data.

Accuracy: 0.9839666154184055
Precision: 0.9904761904761905
Recall: 0.839677047289504

Fig. 1. Images of people without Mask (Celebrity face dataset).

Table 1
Programming input & output data.

```
_____
Layer (type) Output Shape Param #
========================================================
input_1 (InputLayer) (None, 250, 250, 3) 0_____
conv2d_1 (Conv2D) (None, 63, 63, 96) 34944
_____
max_pooling2d_1 (MaxPooling2 (None, 32, 32, 96) 0
_____
conv2d_2 (Conv2D) (None, 32, 32, 384) 332160
_____
conv2d_3 (Conv2D) (None, 32, 32, 384) 1327488
_____
conv2d_4 (Conv2D) (None, 32, 32, 256) 884992
_____
max_pooling2d_2 (MaxPooling2 (None, 16, 16, 256) 0
_____
flatten_1 (Flatten) (None, 65536) 0
_____
dense_1 (Dense) (None, 4096) 268439552
_____
dropout_1 (Dropout) (None, 4096) 0
_____
dense_2 (Dense) (None, 4096) 16781312
_____
dropout_2 (Dropout) (None, 4096) 0
_____
dense_3 (Dense) (None, 1000) 4097000_____dropout_3 (Dropout) (None, 1000) 0
_____
dense_4 (Dense) (None, 1) 1001
========================================================
Total params: 291898449
Trainable params: 291898449
Non-trainable params: 0
_____
```

S. Gupta, S.V.N. Sreenivasu, K. Chouhan et al.

Fig. 2. Images of people with mask (RMFD Dataset).

illustrated in Figs. 1 and 2, the Kaggle dataset comprises over 200 k photographs of celebrities with 40 binary feature annotations.

2.1. Face-detection

For face identification in this article, we employed C2D CNN [9,10,16,17] or 2D colour PCA-Principal component Analysis [11] – convolutional neural networks [5,13–15]. In this procedure, we estimate a mixture of features learnt from original pixels and the representation learnt by CNN, and then estimate the outcome or make a decision. A colour 2-dimensional principal component analysis-convolutional neural network is known as a C2D CNN.

2.1.1. Convolution neural network

Convolution is a process to mix two signals and create a new one. It is one of the most important techniques in signal processing, as image is a 2-dimensional signal so convolution is applied to image to create a new image with modified properties according

to need. In order to understand convolution, consider an image of shape (10x10) and a kernel/filter of shape (3x3) having pixel intensities from I_1 to I_{64} and F_1 to F_9 respectively as per followings.

I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}
I_{17}	I_{18}	I_{19}	I_{20}	I_{21}	I_{22}	I_{23}	I_{24}
I_{25}	I_{26}	I_{27}	I_{28}	I_{29}	I_{30}	I_{31}	I_{32}
I_{33}	I_{34}	I_{35}	I_{36}	I_{37}	I_{38}	I_{39}	I_{40}
I_{41}	I_{42}	I_{43}	I_{44}	I_{45}	I_{46}	I_{47}	I_{48}
I_{49}	I_{50}	I_{51}	I_{52}	I_{53}	I_{54}	I_{55}	I_{56}
I_{57}	I_{58}	I_{59}	I_{60}	I_{61}	I_{62}	I_{63}	I_{64}

F_1	F_2	F_3
F_4	F_5	F_6
F_7	F_8	F_9

Convolution will be performed by moving the filter over the picture from top left to bottom right. Equation explains it mathematically (1),

$$\sum_{k=1}^{m} \sum_{l=1}^{n} I(i+k-1, j+1-1)K(k,l) \tag{1}$$

where, i = 1 to (M – m + 1) and j = 1 to (N – n + 1)

As seen in Fig. 3, the convolution process is now repeated across layers to produce a network-like structure known as a convolutional neural network. This diagram describes the colour 2-dimensional principal component analysis-convolutional neural network topology of this network (C2D-CNN).

This network takes colored face training images into input and passes through 11 layers to generate SoftMax output. To train this network backpropagation algorithm is used and gradients for that are calculated by using the following equations,

$$\frac{\partial L}{\partial \sigma^2} = -\frac{1}{2} \sum_{i=1}^{s} \frac{\partial L}{\partial y_i}(x_i - \mu)(\sigma^2)^{\frac{3}{2}} \tag{2}$$

$$\frac{\partial L}{\partial \mu} = \left(\sum_{i=1}^{s} \frac{\partial L}{\partial y_i} * \frac{-1}{\sigma} \right) + \frac{\partial L}{\partial \sigma^2} \frac{-2\sum_{i=1}^{s}(x_i - \mu)}{s} \tag{3}$$
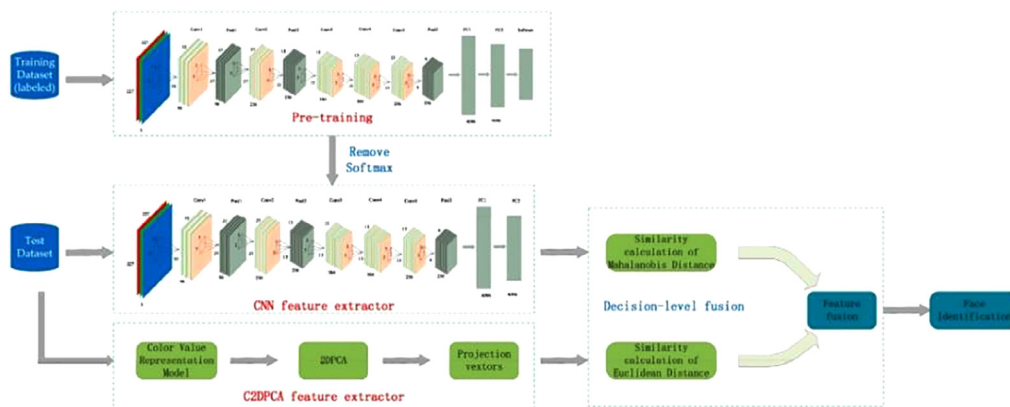

Fig. 3. The construction of convolutional neural network algorithm.

S. Gupta, S.V.N. Sreenivasu, K. Chouhan et al.

$$\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial y_i} * \frac{1}{\sigma} + \frac{\partial L}{\partial \sigma^2} * \frac{2(x_i - \mu)}{s} + \frac{1}{s}\frac{\partial L}{\partial \mu} \qquad (4)$$

$$\frac{\partial L}{\partial \varphi} = \frac{\partial L}{\partial \mu_B}\frac{\partial \mu_B}{\partial \varphi} = \frac{\partial L}{\partial \mu_B}(\mu - \mu_B) \qquad (5)$$

$$\frac{\partial L}{\partial \psi} = \frac{\partial L}{\partial (\sigma_B^2)}\frac{\partial (\sigma_B^2)}{\partial \psi} = \frac{\partial L}{\partial (\sigma_B^2)}(\sigma^2 - \sigma_B^2) \qquad (6)$$

In these equations, L is the loss function and φ and ψ are the parameters that the model will learn. Now, the same image that was given as input to above CNN is passed through PCA to obtain projection vectors.

2.1.2. Principal component examination

The Principal Component Analysis (PCA) is used to remove linearly dependent vector from a data/matrix. It removes the features by projecting the whole data to a new rotated vector space where the data acts as ideal data. The whole process revolved around making the covariance matrix of data to a sparse matrix. This makes the inter feature covariance negligible or very less which is used to make features less linearly dependent.

Fig. 4 explains the rotated vector space. The vectors with b_1 & b_2 as basis vectors are original vectors and vectors with μ_1 & μ_2 are rotated vectors. This is how whole data is projected to a new vector space and then vectors with maximum individual variance and rest of the features are removed to reduce the dimensions.

2.2. Mask detection

In this paper we did the mask recognition by a special convolutional architecture. It is basically a classification algorithm that is best suited for the classification of RGB images. The architecture consists of eight layers, the first five of which are convolutional layers, followed by some max-pooling layers, and the last three of which are fully linked layers. It primarily employs the ReLU activation function, which outperforms tanh and sigmoid.

3. Topology with results

3.1. Data preprocessing

Data preprocessing is a mining technique that includes transforming any primitive data into an easily understandable format. Real data is sometimes inconsistent or lacking in some trends, and can contain many errors. We are taking an image data set, so it contains lots of unnecessary images, or shape issues, dimension issues so we are pre-processing our data to make it suitable for our problem. We have taken two data sets in our research i.e., first is RMFD which is a masked image dataset and second is a celebrity
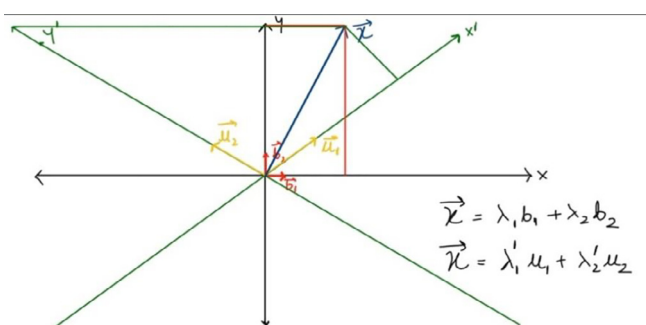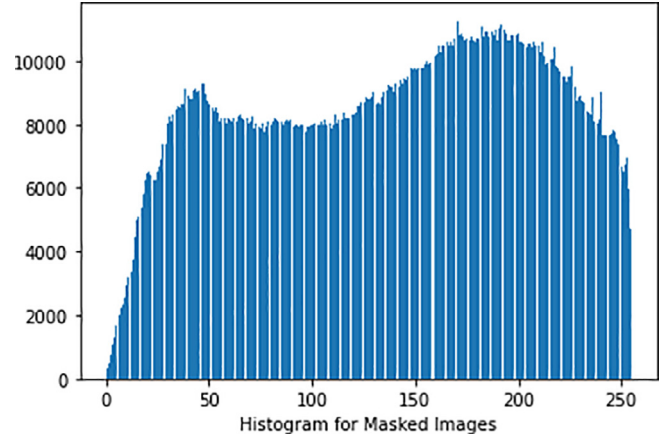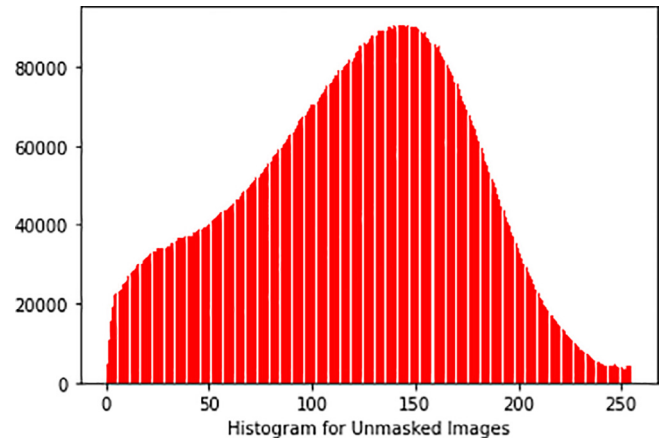

Fig. 5. Mean data of all the masked images.

Fig. 6A. Mean data of all the unmasked images.

dataset, so we are plotting or reshaping them to process them as shown in Fig. 5 and Fig. 6.Fig. 6AFig. 6B.

3.2. Detected faces out of the image

For face recognition, we employed a deep C2D CNN (colour 2-dimensional principal component analysis – convolutional neural network). We mix the attributes of original pixels with the picture representation learnt by CNN in this procedure, and then perform decision fusion to enhance our face recognition model. There are several steps in this network, first, the introduction of the normalization layer in the network to shorten the time for training. Second, the introduction of a layered activation function which will prevent gradient diffusion. Lastly, the application of probabilistic max-pooling for the preservation of feature information to the maximum level. We are using this algorithm because in HOG, the algorithm is based on shapes and people with a mask cannot be detected using that, so we are using this C2D-CNN as it is a pixel-oriented algorithm and favors our process.

3.3. AlexNet

In this research we are doing classification by AlexNet architecture. It is a type of convolutional neural network. It is basically a classification algorithm that is best suited for the classification of RGB images. Architecture includes eight layers, among them, the first 5 were convolutional layers, followed with some max-
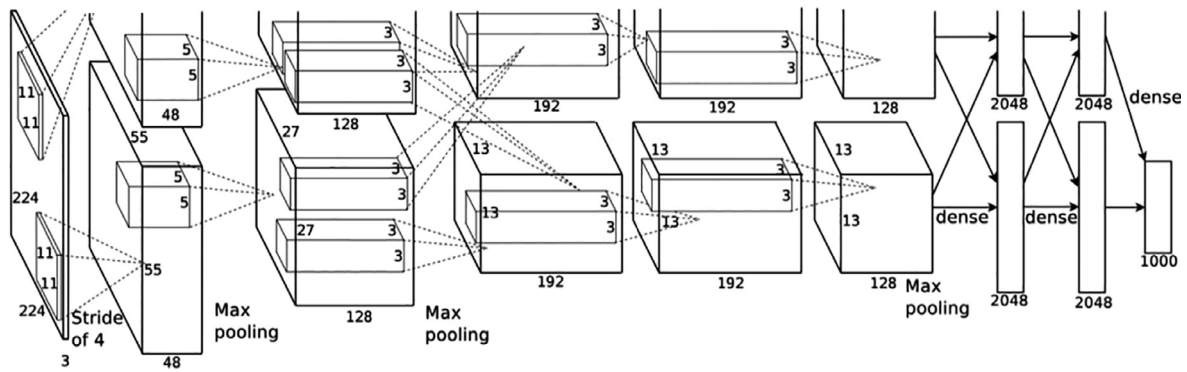

Fig. 4. Different vector space rotation.
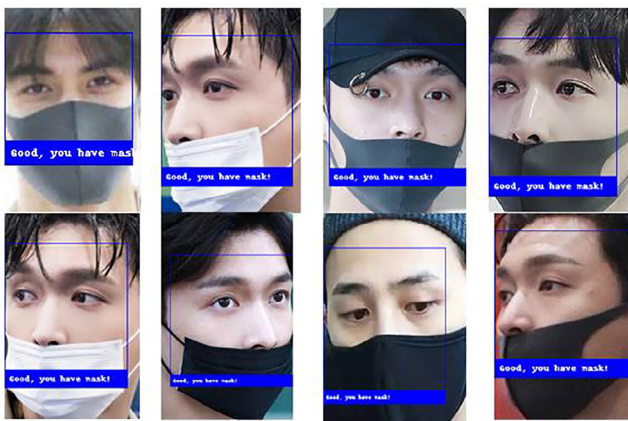
Fig. 6B. Detecting masks on faces.

Fig. 7. Alex Net Architecture [12]

pooling layers which are used in down sampling and the last three layers were fully connected layers. It uses a rectified linear unit activation function. The basic description of Alex Net architecture is shown in Fig. 7.

3.4 Results

.

4. Conclusion

COVID-19, as we all know, is a worldwide epidemic that has harmed all of the world's main powers. As a result, it becomes everyone's national responsibility to stop it from spreading and to teach people to follow the norms. Mask recognition may be used in public surveillance cameras to look for persons who aren't wearing a mask, and then analyzing the data may reveal a lot about the disease's progress. This corona-virus will become ingrained in our society; we will have to be more careful when interacting with crowds, and wearing masks will become the norm. Based on the findings, we can conclude that the ML based topology provides better results with higher accuracy and is more effective in controlling the COVID-19 pandemic.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Huang, H. Ai, Y. Li, S. Lao, High-performance rotation invariant multi-view face detection, IEEE TPAMI 29 (4) (2007) 671–686.

[2] E.N. Perencevich, D.J. Diekema, M.B. Edmond, Moving personal protective equipment into the community: face shields and containment of COVID-19, Jama 323 (22) (2020) 2252–2253.

[3] B. Jun, I. Choi, D. Kim, Local transform features and hybridization for accurate face and human detection, IEEE TPAMI 35 (6) (2013) 1423–1436.

[4] Kushal Sharma and Sandeep Gupta. "Positive and Negative Vibe Classifier by Converting Two-Dimensional İmage Space into One-Dimensional Audio Space Using Statistical Techniques", Advances in Renewable Energy and Sustainable Environment. Springer, Singapore, pp. 347-355, 2021.

[5] Li, Jing Qiu, Tao, Wen, Chang, Xie, Kai, Wen, Fangqing, "Robust Face Recognition Using the Deep C2D-CNN Model Based on Decision-Level Fusion", Sensors, vol. 18, pp. 2080, 2018.

[6] Bhattacharya, Sweta, et al. "Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey", Sustainable cities and society, vol. 65, pp. 102589, February 2021.

[7] Wan, Lihong, Na Liu, Hong Huo and Tao Fang. "Face Recognition with Convolutional Neural Networks and subspace learning." 2017 2ndInternational Conference on Image, Vision and Computing, pp. 228-233, 2017.

[8] T. K. Das, A. Banik, S. Chattopadhyay and A. Das, "FFT based Classification of Solar Photo Voltaic Microgrid System," 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 2019, pp. 1-5,

[9] A. Banik and A. Sengupta, "Scope, Challenges, Opportunities and Future Goal Assessment of Floating Solar Park," 2021 Innovations in Energy Management and Renewable Resources (52042), Kolkata, India, 2021, pp. 1-5.

[10] A. Banik, A. Shrivastava, R. Manohar Potdar et al., Design, Modelling, and Analysis of Novel Solar PV System using MATLAB, Materials Today: Proceedings, 10.1016/j.matpr.2021.06.226

[11] Das T.K., Banik A., Chattopadhyay S., Das A. (2021) Energy-Efficient Cooling Scheme of Power Transformer: An Innovative Approach Using Solar and Waste Heat Energy Technology. In: Ghosh S.K., Ghosh K., Das S., Dan P.K., Kundu A. (eds) Advances in Thermal Engineering, Manufacturing, and Production Management. ICTEMA 2020. Lecture Notes in Mechanical Engineering. Springer, Singapore. 10.1007/978-981-16-2347-9_17.

[12] Sharma, Kushal, Chirag Gupta, and Sandeep Gupta. "Infant Weeping Calls Decoder using Statistical Feature Extraction and Gaussian Mixture Models." In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6. IEEE, 2019.

[13] S. Gomathi, R. Kohli, M. Soni, G. Dhiman, R. Nair, Pattern analysis: predicting COVID-19 pandemic in India using AutoML, in: World Journal of Engineering, Vol.ahead-of-print., 2020, https://doi.org/10.1108/WJE-09-2020-0450.

[14] R. Nair, S. Vishwakarma, M. Soni, T. Patel, S. Joshi, Detection of COVID-19 cases through X-ray images using hybrid deep neural network, in: World Journal of Engineering, Vol.ahead-of-print., 2021, https://doi.org/10.1108/WJE-10-2020-0529.

[15] M. Soni, S. Chauhan, B. Bajpai and T. Puri, "An Approach To Enhance Fall Detection Using Machine Learning Classifier," 2020 12th International Conference on Computational Intelligence and Communication Networks