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

MASKED FACE RECOGNITION WITH LATENT PART DETECTION

Arshi Afreen, As. Prof. Dr. R.K. Banyal

Rajasthan Technical University, Kota

ABSTRACT

The present study introduced an innovative method for recognising masked faces through the utilisation of Latent Part Detection (LPD). The method under consideration employs a convolutional neural network (CNN) for the purpose of identifying masked faces. The identified faces are subsequently processed by a secondary network to produce an embedding for the masked face. The process of utilising this embedding involves comparing the facial features of a masked individual with a pre-existing database of recognised individuals. Comparing the proposed approach to other facial recognition techniques, the findings demonstrate that it achieves high accuracy. Furthermore, the VGG-16 model under consideration has undergone training not only with the ALexNet architecture but also with the ResNet architecture, in conjunction with data obtained from train sets. The present study employed Alex and Res Net forthe purpose of evaluating and contrasting the efficacy of the proposed model with respect to theperformance metrics. The results indicated a 97% accuracy rate, 97.62% specificity rate, and 100% sensitivity rate, which surpasses the performance of previous models. The article additionally offers valuable perspectives on the characteristics of the data and elucidates the potential applications of LPD in enhancing the precision of facial recognition. Our aim in the forthcoming period is to broaden the scope of our masked dataset to encompass supplementary applications that involve masked faces. Simultaneously, we endeavour to consistently augment the efficacy of masked face recognition.

Keywords: Masked face, LPD, biometric technology, CNN.

INTRODCUTION

Deep learning, especially deep CNN, has enabled significant advancements in facial recognition as a technique for identifying or verifying a person's identification using a face picture[1]. The poor picture quality is one of the ongoing difficulties, however. Pose, blur, occlusion, lighting, and other variables might result in photographs of poor quality. Traditional face recognition faces a new challenge in the face mask, which is an efficient technique to stop the virus from spreading. Traditional face recognition algorithms may not be successful since the mask covers a significantportion of the face that has many characteristics[2]. As a result, it is important to comprehend how face recognition algorithms handle masked faces as a unique instance of occlusion.

These parts include facial expressions, skin texture, and facial shape, which are all unique to each individual and can be used to accurately distinguish one person from another. The technology works by first extracting features from the face, such as skin texture, facial shape, and facial expressions [3]. Next, it uses a deep learning algorithm to train a model on these features, allowing it to recognize individuals even when their face is obscured. Finally, the model Masked face recognition with latent part detection is a novel approach to facial recognition that takes into account the presence of facial masks and other occluded facial features.

It is based on the idea that a face can be identified based on its latent features, which are features that are not visible due to occlusion. This approach uses a deep learning model to detect and recognize a face even when it is partially or fully occluded[4]. The model is trained on a dataset of faces with various occlusions, such as masks or facial hair. The model then learns to detect and recognize the latent features of these faces, such as the eyes and mouth, even when they are covered up. This approach has many potential applications in the field of facial recognition [5]. For example, it could be used for security purposes, such as in airports, where facial recognition is used to identify travellers. It could also be used for biometrics, such as in mobile devices, where it could be used to unlock the device. The approach works by first training a deep learning model on a dataset of faces with various occlusions[6]. The model learns to detect and recognize the latent features of the faces, such as the eyes and mouth, even when they are covered up. The model then learns to recognize the features even when the face is partially or fully occluded. Themodel is then applied to a new image, such as a face with a mask or facial hair. The model then detects and recognizes the latent features of the face, such as the eyes and mouth, even when theyare covered up. This allows the model to recognize the face even when it is partially or fully occluded [7].

RELATED WORK

Based on feature set matching, other studies attempted to identify incomplete faces[8]. Itsuggested to find the greatest matching area (LMA) in testing pictures that can be represented by training photos in line with the concept of sparse representation (Wright et al. 2008). In order to optimise, changed the rank minimization issue into the nuclear norm minimization problem.

A Laplacian-uniform mixture function was used in Laplacian-uniform mixture-driven iterative robust coding (LUMIRC) [9] to simulate the distribution of the reconstruction residuals. Although their theoretical contributions are sound, the complexity of real-world scenarios restricts their usefulness. Deep learning has recently become quite popular in occlusion-resistantface recognition.

Wan et al. proposed the use of MaskNet to acquire diverse weights for the spatial positions of feature maps in the medial layer of a deep face network [10]. Through the utilisation of the pairwise differential siamese network (PDSN), feature maps extracted from a partially obscured facial image can be compared to those of its unobstructed counterpart, Song et al. suggested learning masks[11]. In order to build a reliable face representation on noisy labelled datasets, a Light CNN framework was created in by using Max-Feature-Map (MFM), which can distinguish between noisy and informative signals[5]. The researchers [12] suggested an LSTM-autoencoder model to identify occlusions as well as restore natural faces for the face de-occlusion problem. Alexious et.al., said that in order to evaluate masked face recognition algorithms, [13]gathered two datasets called MFV and MFI In order to locate the latent face component, they created a latent part detection (LPD) model, according to Ding and others [14].

Organization of The Study

The work may be broken down into the following sections: Background of the proposed methodsare provided in Section II, an explanation of the methodology that underpins the recommended algorithm is provided in Section III, Section IV delivers the results of the study, and Section V provides a conclusion.

II. BACKGROUND

In this section of Background of masked face recognition classification, the architecture of the previously developed models Alex Net and Res Net 50, in addition to the design of our suggestedmodel VGG16, LPD is discussed.

VGG16 Architecture

The VGG16 challenge marked a significant milestone in the pursuit of endowing computers with visual perception capabilities, representing a pivotal moment in human history. Over several decades, considerable efforts have been directed towards enhancing this ability in the domain of computer vision (CV)[15]. The significant advancement identified as VGG16 set the path for a number of further advancements in this area of study. The Convolutional Neural Network model (often known as CNN) was developed by Andrew Zisserman and Karen Simonyan of the University of Oxford. The first presentation of the vehicle's idea took place in 2013, but the submission of the vehicle's complete model didn't take place until the ILSVRC ImageNet Challenge in 2014[16]. A competition held annually and known as the ImageNet Wide Area Visual Classification Challenge (ILSVRC) assessed the effectiveness of several methods for theclassification of images (as well as the identification of objects).

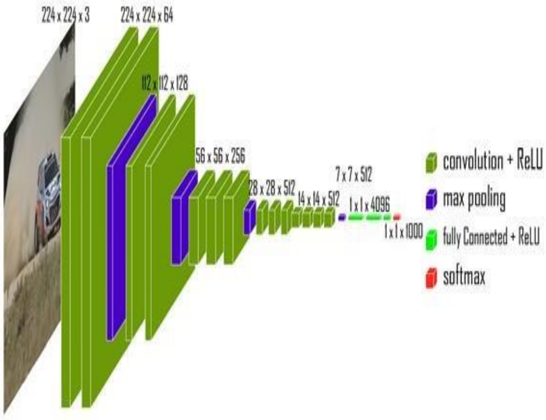


Figure 1 VGG 16 Architecture

Alex Net Architecture

The architectural design comprises a grand total of eight layers, with the first five being convolutional layers and the last three being the concluding layers of which are completely connected layers [17]. In order to get as many features as is practically possible, the first two convolutional layers of the network are linked to overlaid portions of max-pooling. The fully- connected layers are immediately coupled to the fourth, third, and five convolutional layers through direct links. All of the layers' outputs are linked to the ReLu non-linear function of activation[18].

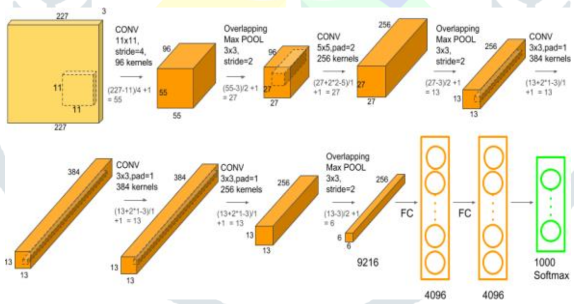


Figure 2 Alex Net Architecture

Res Net 50 Architecture

Microsoft Research created a convolutional neural network (CNN) architecture called ResNet-50. It is a deep neural network with 50 layers comprised of numerous residual blocks with shortcut connections ResNet-50 is designed to enhance the precision of image classification tasks.[19]A convolutional layer is followed by many residual blocks in the ResNet-50 architecture. Each residual block consists of three convolutional layers with a direct input link. The fast connection is used to deepen the network, enabling the learning of more complicated elements [20]. The convolutional layers use batch normalization and ReLU activation algorithms to enhance the network's precision and speed. Many applications, such as image classification, object identification, and semantic segmentation, have used the ResNet-50 architecture. In numerous benchmark datasets, including ImageNet, CIFAR-10, and Pascal VOC, the system hasattained a level of performance that is currently considered to be the most advanced in its field. Moreover, the network is often used for transfer learning[21]. Because to its precision and scalability, ResNet-50 has emerged as one of the most prominent deep learning architectures. It has become a vital component of many computer vision applications and is an excellent option for anyone seeking to develop a deep learning model [22]

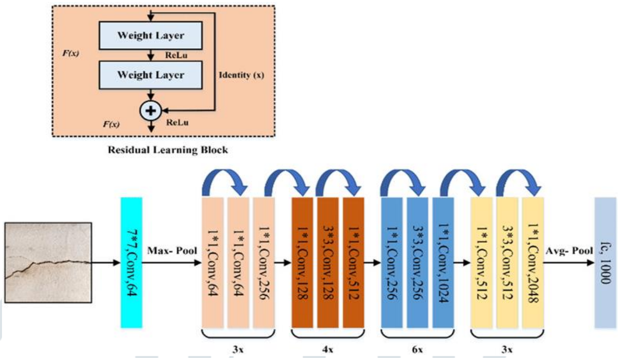


Figure 3 Res Net 50 Architecture

Latent Part Detection (LPD)

Latent Part Detection is a technique for finding and deleting portions of an image that are maskedor concealed by other picture elements. Latent Part Detection may be used to find items in imagesor movies that are difficult for the human eye to see. It may also be used to find details in a picture that aren't immediately evident, including wall fissures, things hidden behind other things, or even tiny creatures or insects.

Edge detection, texture analysis, and pattern recognition are often used in conjunction to performlatent part detection. Techniques for edge detection are used to find items that seem to be concealed or veiled and to recognise sudden changes in an image's intensity. When identifying objects with similar texture characteristics, such as cracks or rough surfaces, texture analysis techniques are used. Finally, to recognise items that have a certain form, size, or colour, pattern recognition algorithms are used.

A scene may be more fully and accurately shown after the concealed or obscured portions of a picture have been found and deleted or restored. This may be accomplished by cropping the picture or by removing the sections using image-editing software. In other instances, the pieces would need to be changed out with fresh components, such adding flowers or trees to a landscapeor a person to a group shot.

Forensic investigators, law enforcement personnel, and other professionals who may need to identify or examine things that are not readily apparent to the naked eye might consider using latent part detection as a valuable tool. For scientific or aesthetic reasons, it may also be utilisedto improve the quality and accuracy of photographs.

III. METHODOLOGY

Masked Face Recognition with Latent Part Detection is a technique used to identify people whoare wearing masks in a given image. This technique utilizes a combination of facial recognition algorithms, deep learning models, and latent part detection to accurately identify individuals whoare wearing masks. This technique is commonly used in security and surveillance applications, where it can be used for the purpose of detecting and identifying individuals who are wearing masks. public areas.

The method for Masked Face Recognition with Latent Part Detection begins with the acquisitionof an image or video of the person whose identity is to be determined. This image can be taken from a variety of sources, such as a CCTV camera or a smartphone. Masked Face Recognition with Latent Part Detection is a technique used to identify people who are wearing masks in a given image. In this study we will be using our two-branch CNN, which consists of a global branch for latent part detection and prejudiced global characteristic and focuses on the acquisition of knowledge and the development of a specialised subfield dedicated to identifying latent components and acquiring partial features. The sharing of parameters between the two branches of a Convolutional Neural Network (CNN) results in mutual benefits and reduced size.Data description

For this study, we will be collecting data from this web site.

https://www.kaggle.com/datasets/omkargurav/face-mask-dataset

Pre-processing steps for image classification:

This experiment's goal is to demonstrate how the precision of a fundamental convolutional network is subject to variation upon the application of several established pre-processing techniques topologies. Specifically, experiment will focus on how these changes affect image classification. The procedures listed below are a few instances of pre-processing methods.

- Read the picture
- Resample the image
- Eliminate the noise

Read the picture

After saving the link to our image database in a variable, we built a way to load photo-containing foldersinto arrays so that we could read the image. This allowed us to read the image.

Resample the image

During this step of the process of resizing pictures, we will develop two different methods to display the photographs: one technique will display a single image, while the other approach willdisplay a pair of photos. This will allow us to observe how the size of the images has changed. After that, we build a technique that we call processing, which takes photographs alone as its soleacceptable kind of input.

Eliminate the noise

In order to quiet down the commotion A gaussian blur is the outcome of applying a Gaussian function on an image in order to blur its edges. This procedure is referred to as a gaussian blur. One of the ways that this effect is put to use in graphic design software is to reduce the quantity of visual noise that is present in an image. The practice of the gaussian flattening is frequently utilized as a step in the pre-processing phase of computer vision algorithms. This process is carried out in order to improve the picture structures at varying levels of magnification.

Feature selection using LPD

A novel latent component detection model was devised for the purpose of masked face recognition. The model was inspired by the finding that the human visual system assigns a higher degree of importance to the non-occluded areas when identifying faces. The objective of this model is to ascertain the latent region that exhibits the highest degree of discrimination in masked facial images. The region under consideration is identified as the latent portion that the model endeavours to detect. This is because, according to previous research, the non-masked region, particularly the forehead-eye area, is the primary focus of the visual system when mask-occlusiontakes place.

Finding the latent component is the next step. To find the forehead-eye region, one option is to employ landmark detection. This approach often fails because masks obscure several markers that are important for locating the forehead-eye area, such as the nose tip. We offer the latent partdetection model to learn the latent component expressed as a bounding box in order to overcomethe aforementioned flaw.

$$B^{s.} = [(X_i, Y_i), (X_{r.}, Y_{r.})] \quad (1)$$

Where (X, Y_i) stands for the upper-left corner's coordinate and $(X_{r.}, Y_{r.})$ for the bottom-rightcorner's coordinate.

We hope that our network can learn the bounding box $B^{s.}$ coordinates and that they are differentiable with regard to the loss function.

Train test split:

When all of the pre-processing procedures have been completed, the data set is divided into testing and training sets according to the split ratio that the user has specified. In the future, the split training information will be used for training the algorithms, while the test results will be utilized for testing the models. Both sets of data will be separated into two separate sets.

Train the network:

Train data are used in the training of the proposed VGG-16 model as well as Alex, Res Net 50. The performance of the recommended model is assessed and contrasted using these two more instances of Alex and VGG-16.

The following metrics may be used to assess the model's performance.

Performance Matrices:

When evaluating the usefulness of a method, the accuracy, sensitivity, precision, and F1-scoreof the confusion matrix are all taken into consideration.

Accuracy: It refers to the proportion of the total number of topics that were understoodadequately.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

Sensitivity: The percentage of appropriately favourable labels that the machine we use recognizes as being labels is referred to as the recall, which is also known by its other name, sensitivity. This percentage might be either positive or negative.

$$Sensitivity = \frac{TP}{TP + FN} \qquad (2)$$

Precision: By taking into account the total quantity of accurate forecasts, it is possible to assessthe precision of a projection. This concept is alternatively referred to as predictive value.

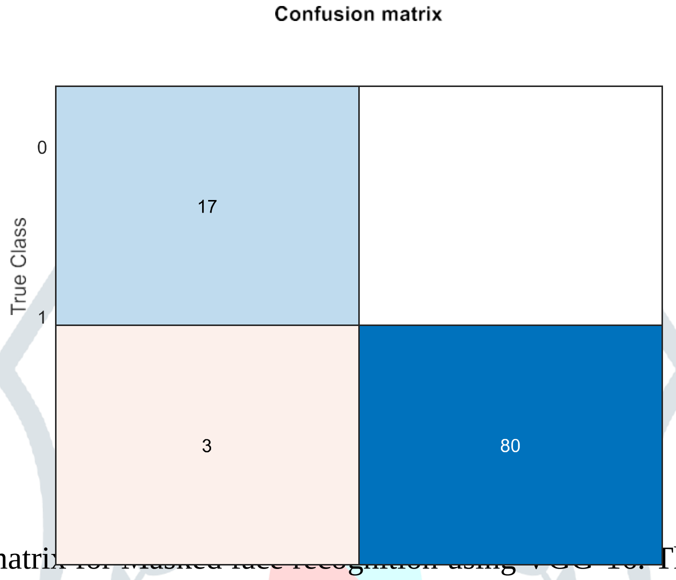
$$\Pr ecision = \frac{TP}{TP + FP} \qquad (3)$$

Specificity: The algorithm has properly classified the negative as specificity, which is theintended meaning of the phrase.

$$specificity = \frac{TN}{TN + FP} \qquad (4)$$

IV. RESULTS

Performance measures are utilized to compare the proposed algorithm against existing classification methods, such as Res Net 50, Alex net, and VGG-16, in order to illustrate the effectiveness of the suggested approach. This comparison is carried out in order to provide evidence of the efficiency of the algorithm.



From the above confusion matrix for Masked face recognition using VGG-16. The confusion matrix displays two classes in which class 0 shows normal face and class 1 shows masked face. Our proposed model accurately predicted 80 times as masked face and 15 times as normal face. In this 3 misclassifications have also occurred.

To assess the efficacy of the architectures, the performance metrics, such as Accuracy, Sensitivity, and Specificity, are computed and represented as follows.

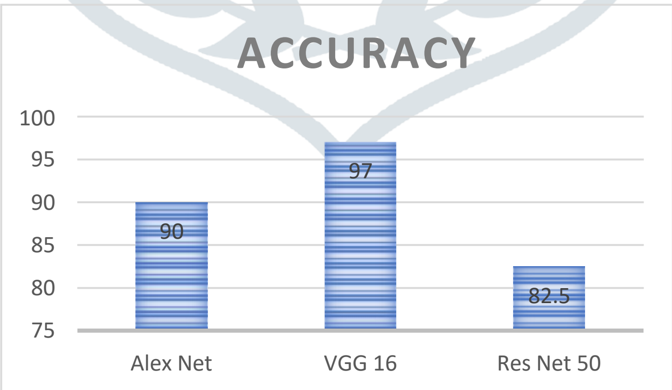


Figure 1 Accuracy

The average accuracy of Res Net 50 architectures is 82.5 percent, ALEXNET architectures are 90 percent, and VGG-16 architectures are 97 percent. This demonstrates that the VGG-16 architecture are more accurate in recognising Masked face.

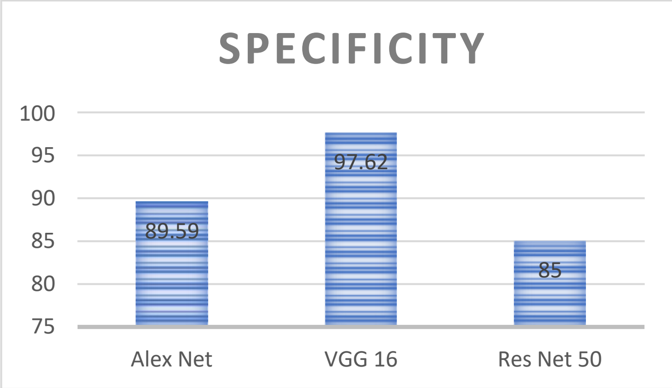


Figure 2 Specificity

The average specificity of Res Net 50 architectures is 85 percent, ALEXNET architectures are 89.59 percent, and VGG-16 architectures are 97.62 percent. This demonstrates that the VGG 16architecture is more accurate in recognizing masked face.

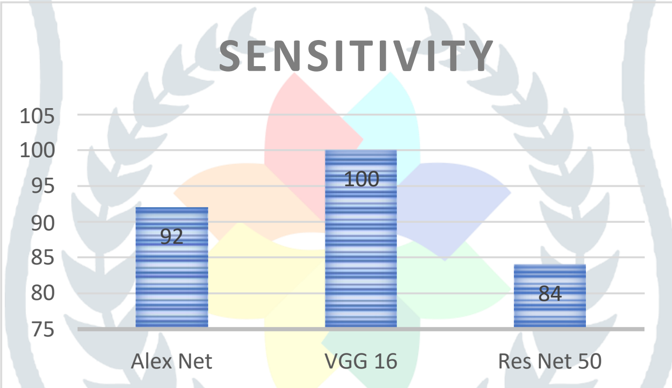


Figure 3 Sensitivity

The average sensitivity of Res Net 50 architectures is 84 percent, ALEXNET architectures are 92 percent, and VGG-16 architectures is 100 percent. This demonstrates that the Res Net 50 architecture are more accurate in recognising Masked face.

Architectures	Accuracy (%)	Specificity (%)	Sensitivity (%)
Alex Net	90	89.59	92
VGG-16	97	97.62	100
Res Net 50	82.5	85	84

V. CONCLUSION

For the purpose of this research, we devised a CNN architecture for the recognition andclassification of masked face that was based on the VGG-16 model. Furthermore, a novel model for detecting latent facial parts (referred to as the Latent Part Detection or LPD model) has been introduced, which is capable of accurately identifying such parts even in the presence of masks. Following the completion of any necessary pre-processing, the divided train data is put to use inorder to instruct the algorithms, while the test data is put to use in order to assess the performance of the model. The proposed VGG-16 model is trained using not just Alex Net but also Res Net, as well as data from trains. These two, Alex and Res Net, are used in order to analyze and compare the performance of the suggested model in accordance with the performance measures, and we obtained 97 percent accuracy, 97.62 percent specificity, and 100 percent sensitivity, which are higher than past models. Our objective for the future is to expand our masked dataset to encompass additional applications involving masked faces, while also striving to consistently enhance the performance of masked face recognition.

References

[1] G. Jeevan, G. C. Zacharias, M. S. Nair, and J. Rajan, "An empirical study of the impact of masks on face recognition," *Pattern Recognit.*, vol. 122, p. 108308, 2022.

[2] E. Rezende, G. Ruppert, T. Carvalho, F. Ramos, and P. De Geus, "Malicious software classification using transfer learning of resnet-50 deep neural network," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2017, pp. 1011–1014.

[3] R. Natsume, T. Yatagawa, and S. Morishima, "Rsgan: face swapping and editing using face and hair representation in latent spaces," *arXiv Prepr. arXiv1804.03447*, 2018.

[4] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," *Electron.*, vol. 9, no. 8, pp. 1–53, 2020, doi: 10.3390/electronics9081188.

[5] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar, "Detecting silicone mask- based presentation attack via deep dictionary learning," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 7, pp. 1713–1723, 2017.

[6] P. Wlodarczak, J. Soar, and M. Ally, "Multimedia data mining using deep learning," in *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*, 2015, pp. 190–196.

[7] L. Qi, J. Huo, L. Wang, Y. Shi, and Y. Gao, "Maskreid: A mask based deep ranking neural network for person re-identification," *arXiv Prepr. arXiv1804.03864*, 2018.

[8] and T. 2016 Weng, Lu, "MASKED FACE RECOGNITION," 2016.

[9] Z. Zuo *et al.*, "Convolutional recurrent neural networks: Learning spatial dependencies for image representation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 18–26.

[10] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, 2018.

[11] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential siamese network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 773–782.

[12] S. Singh, A. K. Aggarwal, P. Ramesh, L. Nelson, P. Damodharan, and M. T. Pandian, “COVID 19: Identification of Masked Face using CNN Architecture,” in *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2022, pp. 1045–1051.

[13] M. Alexiou, I. P. Ktistakis, and G. Goodman, “Towards a Masked Face Recognition Algorithm: A Novel Rule Based Hybrid Algorithm,” in *2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, 2021, pp. 1–6.

[14] F. Ding, P. Peng, Y. Huang, M. Geng, and Y. Tian, “Masked face recognition with latent part detection,” in *Proceedings of the 28th ACM international Conference on multimedia*,2020, pp. 2281–2289.

[15] T. Besiroglu, N. Emery-Xu, and N. Thompson, “Economic impacts of AI-augmented R&D,” *arXiv Prepr. arXiv2212.08198*, 2022.

[16] J. del Río Fernandez, “MARTECH 2021: 9th International Workshop on Marine Technology,” in *International Workshop on Marine Technology (9th: 2021: Virtual)*, 2021, no. 21, pp. 1–96.

[17] T. Shanthi and R. S. Sabeenian, “Modified Alexnet architecture for classification of diabetic retinopathy images,” *Comput. Electr. Eng.*, vol. 76, pp. 56–64, 2019.

[18] X. Han, Y. Zhong, L. Cao, and L. Zhang, “Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification,” *Remote Sens.*, vol. 9, no. 8, p. 848, 2017.

[19] S. A. ElGhany, M. R. Ibraheem, M. Alruwaili, and M. Elmogy, “Diagnosis of Various Skin Cancer Lesions Based on Fine-Tuned ResNet50 Deep Network,” *Comput. Mater. Contin.*, vol. 68, no. 1, pp. 117–135, 2021, doi: 10.32604/cmc.2021.016102.

[20] B. Mandal, A. Okeukwu, and Y. Theis, “Masked face recognition using resnet-50,” *arXiv Prepr. arXiv2104.08997*, 2021.

[21] M. Sankupellay and D. Konovalov, “Bird call recognition using deep convolutional neural network, ResNet-50,” in *Proc. Acoustics*, 2018, vol. 7, no. 2018, pp. 1–8.

[22] A. Naeem, T. Anees, M. Fiza, R. A. Naqvi, and S. W. Lee, “SCDNet: A Deep Learning- Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images,” *Sensors*, vol. 22, no. 15, 2022, doi: 10.3390/s22155652.