An Automatic CNN-based Face Mask Detection Algorithm Tested During the COVID-19 Pandemics

Giorgio De Magistris[1], Emanuele Iacobelli[1], Rafał Brociek[2] and Christian Napoli[1,3]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Via Ariosto 25, Roma, 00185, Italy
[2]Department of Mathematics Applications and Methods for Artificial Intelligence, Faculty of Applied Mathematics, Silesian University of Technology, Gliwice, 44-100, Poland
[3]Institute for Systems Analysis and Computer Science, Italian National Research Council, Via dei Taurini 19, Roma, 00185, Italy

Abstract

The ongoing COVID-19 pandemic has highlighted the importance of wearing face masks as a preventive measure to reduce the spread of the virus. In medical settings, such as hospitals and clinics, healthcare professionals and patients are required to wear surgical masks for infection control. However, the use of masks can hinder facial recognition technology, which is commonly used for identity verification and security purposes. In this paper, we propose a convolutional neural network (CNN) based approach to detect faces covered by surgical masks in medical settings. We evaluated the proposed CNN model on a test set comprising of masked and unmasked faces. The results showed that our model achieved an accuracy of over 96% in detecting masked faces. Furthermore, our model demonstrated robustness to different mask types and fit variations commonly encountered in medical settings. Our approaches reaches state of the art results in terms of accuracy and generalization.

Keywords

COVID-19, Face Mask, CNN, ResNet50

1. Introduction

The use of face masks has become a critical preventive measure in controlling the spread of infectious diseases, particularly in the context of the ongoing COVID-19 pandemic. In medical settings, such as hospitals and clinics, healthcare professionals and patients are required to wear surgical masks to minimize the risk of transmission. However, the use of masks can hinder facial recognition technology, which is commonly used for identity verification and security purposes. Accurate and efficient detection of faces covered by surgical masks is thus crucial for maintaining security measures while adhering to infection control protocols.

Traditional approaches for face detection and recognition may face challenges in the presence of masks, as masks can alter facial features, obstructing key facial landmarks and reducing facial visibility. To address this challenge, convolutional neural networks (CNNs), a type of deep learning model known for their ability to automatically learn hierarchical features from images, have been proposed as a promising solution. CNNs have shown great success in various computer vision tasks, including object detection, image classification, and image segmentation. They have the potential to learn complex patterns and representations from large datasets, which can aid in accurately detecting faces covered by surgical masks.

In this paper, we propose a CNN-based approach for detecting faces covered by surgical masks in medical settings. We aim to develop a model that can accurately and robustly identify masked faces, considering the unique challenges posed by different mask types, colors, and fit variations commonly encountered in medical environments. We collect a dataset of facial images with individuals wearing surgical masks, and fine-tuned a ResNet50 model on the specific task. The contributions of our work include the development of a CNN-based approach tailored for face mask detection in medical settings, and the investigation of model performance and generalization in the presence of different mask types and fit variations. The proposed approach has the potential to enhance security measures while maintaining infection control protocols, and can have wide-ranging applications in various real-world scenarios. The rest of the paper is organized as follows: in Section 2, we review related works in the field; in Section 3, we describe the dataset used in our study; in Section 4, we introduce the proposed method; in Section 5 we present the experimental results and discuss the findings; and finally, in Section 6, we conclude the paper and outline future directions of research in this area.

ICYRIME 2022: International Conference of Yearly Reports on Informatics, Mathematics, and Engineering. Catania, August 26-29, 2022
demagistris@diag.uniroma1.it (G. D. Magistris); iacobelli@diag.uniroma1.it (E. Iacobelli); Rafal.Brociek@polsl.pl (R. Brociek); cnapoli@diag.uniroma1.it (C. Napoli)
0000-0002-3076-4509 (G. D. Magistris); 0000-0002-3336-5853 (C. Napoli)

2. Related Works

Machine learning and convolutional neural networks have been widely applied to the general field of face recognition. Face recognition is a rapidly developing field that has seen significant advancements in recent years, driven by the increasing availability of large-scale datasets and powerful machine learning algorithms. Deep learning has become the dominant approach for face recognition due to its superior performance on large-scale datasets. Some of the most widely used deep learning architectures for face recognition include Convolutional Neural Networks (CNNs), Siamese Networks, Triplet Networks, and Deep Belief Networks (DBNs). These models are trained using large-scale datasets such as VGGFace [1], FaceNet [2], and IE-CNN models [3], which contain millions of face images. Deep learning models have achieved state-of-the-art results on various benchmarks such as the Labeled Faces in the Wild (LFW) [4] and MegaFace [5] datasets. Many other applications have been developed using machine learning and face recognition algorithms [6, 7, 8, 9, 8] 3D face recognition is an emerging field that uses 3D information to improve the accuracy of face recognition systems. 3D face models can capture additional facial details such as the depth of the facial features, which are not present in 2D images. Some of the popular approaches for 3D face recognition include the use of 3D morphable models (3DMM), depth-based methods, and multi-view based methods. Face recognition can be divided into two categories: verification and identification. Verification aims to determine if two face images belong to the same person, while identification aims to identify a person from a set of images. Face verification systems are commonly used for security applications, while face identification systems are used in large-scale surveillance applications. Recent advancements in face recognition have focused on improving the accuracy of both verification and identification systems. Hybrid approaches combine multiple techniques to improve the performance of face recognition systems. For example, a hybrid approach may combine deep learning models with 3D face recognition techniques to improve accuracy. Another approach is to use facial landmark detection algorithms to improve the alignment of face images before recognition. As face recognition technology becomes more prevalent, there are growing concerns about privacy and security. Several approaches have been proposed to address these concerns, such as anonymization techniques, which modify the facial features to protect the privacy of the individuals in the images. Other approaches include adversarial attacks, which aim to fool the face recognition system into misidentifying a person. Face recognition is a rapidly developing field that has seen significant advancements in recent years, driven by the increasing availability of large-scale datasets and powerful machine learning algorithms. The latest algorithms for face recognition are based on deep learning architectures, 3D face recognition, and hybrid approaches. As face recognition technology becomes more prevalent, there is a growing need to address privacy and security concerns. ArcFace [10] is a face recognition algorithm that uses a margin-based softmax loss function to optimize feature representations for face recognition, similarly CosFace [11], another deep learning-based face recognition algorithm, uses a cosine-based loss function to improve the discriminability of the learned features. Like ArcFace, it has achieved state-of-the-art performance on several benchmark datasets. Another algorithm, named SphereFace [12], uses a angular-based softmax loss function to improve the discriminability of the learned features. While the said algorithm have many similarities, another system named DeepID [13, 14], has beeen developed to offer a multi-task learning approach to learn multiple levels of features for face recognition. However it has been surpassed by more recent approaches such as VGGFace2 [1]. VGGFace2 is a large-scale face recognition dataset that has been used to train several deep learning-based face recognition algorithms, including some of the approaches mentioned above. It contains over 3 million face images of over 9,000 subjects, making it one of the largest face recognition datasets available. Another convolutional model is FaceNet [2] that, differently from the previous approaches, uses a triplet loss function to learn discriminative features for face recognition, being widely adopted in industry. With the outbreak of the COVID-19 pandemic, the use of face recognition algorithms has been applied to the recognition of people wearing (or not wearing) face masks. One common approach is to use deep learning-based object detection methods to detect the presence of a face and then classify whether the face is wearing a mask or not. This approach typically involves training a CNN on a dataset of masked and unmasked faces. The CNN learns to extract features from the face images and use them to classify whether a mask is present or not. Several studies have reported high accuracy rates for face mask detection using deep learning algorithms. In fatct during the COVID-19 pandemic the use of convolutional neural networks (CNNs) for face mask detection has gained significant attention in literature. Several studies have proposed CNN-based approaches for detecting masked faces. In [15] the authors appliy a Long Short-Term Memory (LSTM) network to model the time-dependencies in order detect whether a person wears a face mask while speaking. In [16] the authors are able to detect if face masks are worn by people in a closed envirоinment. Overall, deep learning algorithms have shown promise for detecting whether a person is wearing a face mask [17], and their use could help to improve public health measures during the COVID-19 pandemic. For example, the authors of [18] propose a

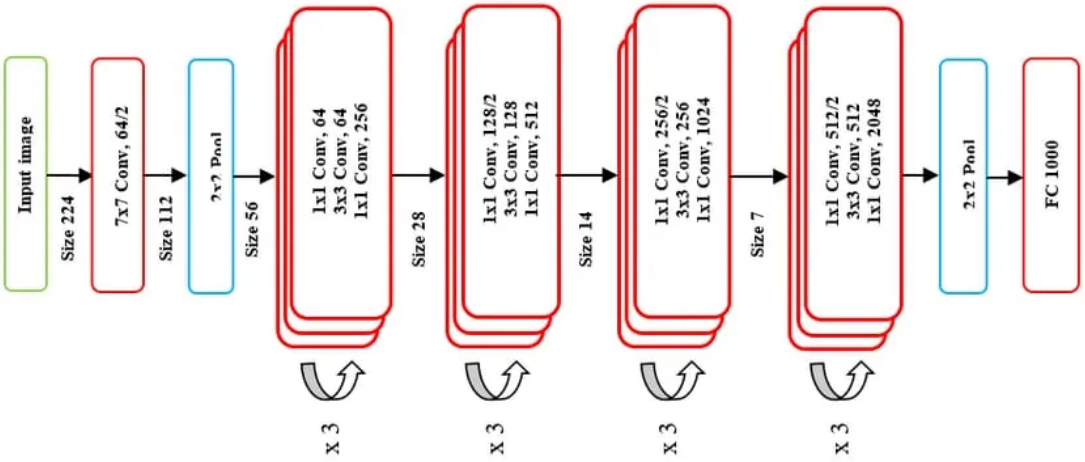

Figure 1: ResNet50 architecture

slightly modified version of LeNet [19] to detect masked faces in the wild. Similarly the authors of [20] propose a novel CNN architecture tailored for the specific task and evaluate their model on a custom dataset. For more information about the existing methods of covered face detection we refer the reader to the recent survey [21]. Our method differs from the other works in literature both in the dataset used and the training strategy. In particular we will address the task as a multi-label classification problem. Previous works [22, 23] have shown that this approach allows a better understanding of the context. We will see that this approach allows to reduce overfitting and consequently to generalize better on unseen data.

3. Dataset

To train our model we created a custom dataset containing 151 images of uncovered faces extracted from the Flickr-Faces-HQ dataset[24]. The Flickr-Faces-HQ (FFHQ) dataset is a large-scale dataset of high-quality facial images collected from the photo-sharing website Flickr. The dataset was created by NVIDIA Research in 2019 and contains 70,000 images of 1,024x1,024 resolution, with a diverse range of ages, ethnicities, and genders. The images in the FFHQ dataset are highly curated and filtered for quality, ensuring that they are of high fidelity, high resolution, and well-lit. The dataset is designed to be used for training and evaluating machine learning models for various computer vision tasks, including face recognition, facial expression analysis, and face synthesis. One of the key features of the FFHQ dataset is that it includes a wide range of facial expressions, poses, and lighting conditions, which makes it more challenging than other facial image datasets such as the popular Labeled Faces in the Wild (LFW) dataset. The FFHQ dataset also includes annotations for facial landmarks, which can be used for tasks such as face alignment and face tracking. The FFHQ dataset has been used in a range of computer vision research projects, including the development of generative models for face synthesis and style transfer, as well as the development of deep learning-based models for facial expression recognition and emotion detection. The availability of high-quality facial images in the FFHQ dataset has helped to advance the state-of-the-art in these and other computer vision tasks, and it is likely to continue to be an important resource for researchers working in the field of computer vision. In this study 1000 images of faces uncovered and covered with masks have been selected from the FFHQ dataset and merged, to preserve generality, with a large portion of Kaggle face mask dataset[25] while others were added manually in order to increase the variability in the masks types and colors; 160 images of masks with no faces scraped from the web and 150 images containing different classes of objects all unrelated to faces and to masks also scraped from the web.

4. Method

For the classification task we used ResNet50 [26] a dResNet50 is a 50-layer deep neural network that is based on residual connections, which allow for the reuse of features from previous layers in the network. The architecture includes several blocks of convolutional layers, batch normalization, and pooling layers, as well as shortcut connections that bypass one or more layers. These shortcut connections enable the network to learn resid-

ual functions, which can be more easily optimized during training and help to mitigate the vanishing gradient problem that occurs in very deep networks. ResNet50 has been trained by means of ImageNet [27]. ImageNet is a large-scale dataset of labeled images that has been used as a benchmark for evaluating and comparing the performance of computer vision algorithms. The dataset was created in 2009 by researchers at Stanford University and contains over 14 million images, each labeled with one of 21,841 categories. When trained with ImageNet, ResNet50 and has achieved state-of-the-art performance on a range of computer vision tasks, including image classification, object detection, and semantic segmentation. In particular, ResNet50 has been used as a pre-trained model for transfer learning, where the network is used as a feature extractor for other tasks, such as fine-grained classification, image retrieval, and face recognition. The ResNet50 architecture has inspired many other variants, such as ResNet101 and ResNet152, which are even deeper neural networks that have achieved even better performance on some tasks. Overall, ResNet50 has become a widely used and popular neural network architecture in computer vision, and its success has helped to advance the field of deep learning. The architecture ResNet50 is represented in figure 1. We first trained the network for the binary classification task of detecting covered versus uncovered faces, but this first attempt resulted in severe overfitting. To reduce the gap between training and generalization error we adopted a different training strategy. In particular we cast the problem into a multi-label classification problem where the labels are: person, mask and covered. The three labels are not mutually exclusive, such that the network is able to distinguish the following cases: there are people and no masks, there are masks and no people, there are people wearing masks and there are people and masks but people are not wearing masks. This training strategy was motivated by different reasons: it increased the size of the dataset, that is a good remedy for overfitting, it allowed the network to distinguish between different scenarios without relying on segmentation, that requires an expensive process of annotation, and, finally, the the introduction of negative examples (objects unrelated with faces and masks) helped the network to focus on the right set of features. In particular we observed that before the introduction of the third class, the network had difficulty in correctly classifying images with many objects in the background, while after the introduction of the negative examples the network is able to ignore irrelevant objects in the image [28, 29, 30].

5. Results

We fine-tuned the network for 200 epochs on images with size 224x224 pixels. We used the feature extractor of the ResNet50 pretrained on ImageNet and we added a single classification layer with sigmoid activation (remember that the labels are not mutually exclusive). We used the standard binary cross entropy loss: considering the batch size B, the predicted value $\hat{y}$ and the true value y we get:

$$L(y, \hat{y}) = -\frac{1}{B} \sum_{i=1}^{B} \sum_{j=1}^{3} Y_{ij} \quad (1)$$

where

$$Y_{ij} = y_i[j] \log(\hat{y}_i[j]) + (1 - y_i[j]) \log(1 - \hat{y}_i[j]) \quad (2)$$

With this training strategy we obtained an accuracy of 96% on a balanced testset, against the 90% of accuracy obtained by the network trained to classify only covered versus uncovered faces. To make a fair comparison between the two approaches, the accuracy in the multilabel classification problem is computed considering only the label that indicates if the face is covered or not, which is in fact a binary classification problem. Figure 2 shows some samples along with the network predictions.

6. Conclusion

Our findings suggest that CNNs can effectively detect faces covered by surgical masks in medical settings, which can be beneficial for enhancing security measures while maintaining infection control protocols. The proposed model has potential applications in healthcare facilities, airports, and other settings where face mask detection is critical for security and safety purposes. Future work can explore additional data sources and further optimization techniques to improve the model's performance and real-world applicability.

References

[1] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserman, Vggface2: A dataset for recognising faces across pose and age, in: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), IEEE, 2018, pp. 67–74.

[2] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 815–823.

[3] A.-P. Song, Q. Hu, X.-H. Ding, X.-Y. Di, Z.-H. Song, Similar face recognition using the ie-cnn model, IEEE Access 8 (2020) 45244–45253.

[4] G. B. Huang, E. Learned-Miller, Labeled faces in the wild: Updates and new reporting procedures, Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep 14 (2014).

Figure 2: Some samples from the dataset with the true labels and the labels predicted by the model. The three labels are: person, mask and face covered

[5] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, E. Brossard, The megaface benchmark: 1 million faces for recognition at scale, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 4873–4882.

[6] V. Marcotrigiano, G. Stingi, S. Fregnan, P. Magarelli, P. Pasquale, S. Russo, G. Orsi, M. Montagna, C. Napoli, C. Napoli, An integrated control plan in primary schools: Results of a field investigation on nutritional and hygienic features in the apulia region (southern italy), Nutrients 13 (2021). doi:10.3390/nu13093006.

[7] G. Capizzi, C. Napoli, S. Russo, M. Woźniak, Lessening stress and anxiety-related behaviors by means of ai-driven drones for aromatherapy, in: CEUR Workshop Proceedings, volume 2594, 2020, pp. 7–12.

[8] V. Ponzi, S. Russo, V. Bianco, C. Napoli, A. Wajda, Psychoeducative social robots for an healthier lifestyle using artificial intelligence: a case-study, in: CEUR Workshop Proceedings, volume 3118, 2021, pp. 26–33.

[9] G. De Magistris, R. Caprari, G. Castro, S. Russo, L. Iocchi, D. Nardi, C. Napoli, Vision-based holistic scene understanding for context-aware human-robot interaction, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 13196 LNAI (2022) 310–325. doi:10.1007/978-3-031-08421-8_21.

[10] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4690–4699.

[11] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, Cosface: Large margin cosine loss for deep face recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5265–5274.

[12] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, Sphereface: Deep hypersphere embedding for face recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 212–220.

[13] Y. Sun, D. Liang, X. Wang, X. Tang, Deepid3: Face recognition with very deep neural networks, arXiv preprint arXiv:1502.00873 (2015).

[14] W. Ouyang, X. Zeng, X. Wang, S. Qiu, P. Luo, Y. Tian, H. Li, S. Yang, Z. Wang, H. Li, et al., Deepid-net: Object detection with deformable part based convolutional neural networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2016) 1320–1334.

[15] S. Liu, A. Mallol-Ragolta, T. Yan, K. Qian, E. Parada-Cabaleiro, B. Hu, B. W. Schuller, Capturing time dynamics from speech using neural networks for surgical mask detection, IEEE Journal of Biomedical and Health Informatics 26 (2022) 4291–4302.

[16] Q. Chen, L. Sang, Face-mask recognition for fraud prevention using gaussian mixture model, Journal of Visual Communication and Image Representation 55 (2018) 795–801.

[17] R. Brociek, G. Magistris, F. Cardia, F. Coppa, S. Russo, Contagion prevention of covid-19 by means of touch detection for retail stores, in: CEUR Workshop Proceedings, volume 3092, 2021, pp. 89–94.

[18] S. Lin, L. Cai, X. Lin, R. Ji, Masked face detection via a modified lenet, Neurocomputing 218 (2016) 197–202.

[19] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (1998) 2278–2324.

[20] W. Bu, J. Xiao, C. Zhou, M. Yang, C. Peng, A cascade framework for masked face detection, in: 2017 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM), IEEE, 2017, pp. 458–462.

[21] B. Wang, J. Zheng, C. P. Chen, A survey on masked facial detection methods and datasets for fighting against covid-19, IEEE Transactions on Artificial Intelligence 3 (2021) 323–343.

[22] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, W. Xu, Cnn-rnn: A unified framework for multi-label image classification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2285–2294.

[23] G. De Magistris, R. Caprari, G. Castro, S. Russo, L. Iocchi, D. Nardi, C. Napoli, Vision-based holistic scene understanding for context-aware human-robot interaction, in: AIxIA 2021–Advances in Artificial Intelligence: 20th International Conference of the Italian Association for Artificial Intelligence, Virtual Event, December 1–3, 2021, Revised Selected Papers, Springer, 2022, pp. 310–325.

[24] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4401–4410.

[25] Kaggle, Face mask detection, https://www.kaggle.com/datasets/andrewmvd/face-mask-detection, 2020.

[26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.

[28] N. Brandizzi, S. Russo, R. Brociek, A. Wajda, First studies to apply the theory of mind theory to green and smart mobility by using gaussian area clustering, in: CEUR Workshop Proceedings, volume 3118, 2021, pp. 71–76.

[29] V. Ponzi, S. Russo, A. Wajda, R. Brociek, C. Napoli, Analysis pre and post covid-19 pandemic rorschach test data of using em algorithms and gmm models, in: CEUR Workshop Proceedings, volume 3360, 2022, pp. 55–63.

[30] G. Magistris, C. Rametta, G. Capizzi, C. Napoli, Fpga implementation of a parallel dds for wide-band applications, in: CEUR Workshop Proceedings, volume 3092, 2021, pp. 12–16.