Mask-net: Detection of Correct Use of Masks Through Computer Vision

Alexander Kalen Targa [*1] **Alberto Landi Cortiñas** [*1] **Nicolas Araque Volk** [1] **Alejandro Marcano Van Grieken** [1]

Abstract

This paper focuses on creating a system for recognizing the correct use of a mask through computer vision techniques. Research was carried out with aims of establishing the criteria for the creation of custom datasets, which were used to train, validate and test a pair of deep learning models, Mask-net and I-Mask-net. Both were designed with similar architectures, making use of Transfer Learning Techniques. The results given by training showed that the fine tuning carried out was adequate, while the tests carried out showed that the models have an acceptable level of accuracy, reaching 85.47% for Mask-net and 85.96% for I-Mask-net, additionally supported by the obtained precision, recall and F1-Score calculations.

1. Introduction

Since the start of the COVID-19 pandemic, people's lives have been affected all over the world. According to the World Health Organization's (WHO) dashboard (2020b), this disease has infected over 175 million people and has caused over 3.8 million deaths as of June 10th, 2021.

Mandatory emergency measures to control and prevent further spread of the virus have caused major changes in the lifestyle of humanity. One issue regarding compliance with bio-safety measures is the incorrect use of masks; WHO preliminary reports (2020a) indicate that the disease spreads mainly from person to person through droplets expelled from the nose or mouth of an infected person when coughing, sneezing or talking.

Currently, the use of masks is one of most effective preventive measures used worldwide, and the first studies conducted regarding their use show promising results, such as the one carried out by Mitze et al. (2020) where it's stated in a study conducted in Jena, Germany, that the use of masks helped reduce the rate of daily growth of infections by 60%.

Based on the previous statement, a research was carried out to determine the criteria of the correct use of a mask, where according to Maragakis & Johns Hopkins Medicine (2020), it's considered all instances where the nose, mouth and chin of a face are fully covered, thus achieving the greatest effectiveness of the mask. On the other hand, incorrect use of the mask is considered any instance where the nose, mouth or chin of a face are partially or fully exposed.

Computer vision models can be used to solve complex issues relevant to images or videos and could be designed and trained to predict the correct use of a mask. Using a deep learning model to solve this problem without human intervention could prove useful in high-traffic spaces, avoiding the manual inspection of such, thus promoting social distancing measures with a reliable check.

Convolutional Neural Networks (CNNs) are the most common networks in the field of computer vision. According to Goodfellow et al. (2016), they are considered a type of neural network for data processing with a known matrix topology in the form of a grid, being tremendously successful for processing images.

This research's main objective is to create a model that applies a CNN architecture and makes use of computer vision techniques to determine if a subject is wearing a mask correctly. The model has three possible outcomes: "Mask", "Mask Incorrect" and "No Mask". Additionally, it was proposed to create another deep learning model that activates whenever the output of the previous model is "Mask Incorrect", this second model indicates if the mask is located over the chin, over the mouth and chin or over the nose and mouth, to be able to provide a suggestion to the user on how to place it correctly.

2. Methodology

Two datasets named KLD and IMKLD were created by the team. KLD is a dataset consisting of a total of 11,003 images in order to train and validate the main model, named Mask-net. The selected images from this dataset were collected from multiple sets of public domain data:

- Medical Mask Dataset by Humans in the Loop (2020),

[*]Equal contribution [1]Faculty of Engineering, Universidad Metropolitana, Caracas, Venezuela. Correspondence to: Alexander Kalen Targa <alexanderkalen@correo.unimet.edu.ve>.

Proceedings of the LatinX in AI (LXAI) Research workshop at ICML 2021.

which provided images of real people wearing masks correctly.

- Flickr-Faces-HQ by Karras et al. (2019), which provided images of real people, without face mask.

- MaskedFace-net by Cabani et al. (2021), which provided images of people using medical mask correctly and incorrectly.

- Own authorship, a set of images that captured situations of the real world, taking into account shooting angles, backgrounds and use of mask.

The second dataset, IMKLD, is made up of 1,841 images built to train and validate the second model, named I-Mask-net. It's composed of images provided by the MaskedFace-net dataset, which supplied all three classes of the incorrect use of a mask. In addition, some images of own authorship were incorporated to increase the number of real scenarios.

For data pre-processing, it was considered essential to reduce unnecessary information that could generate bias in the data. For this task, the MTCNN algorithm by Zhang et al. (2016) was used. Given an image it allows to detect faces using CNNs, which allows to crop the image with the face of subject of interest.

Another bias issue detected corresponds to the lack of color of masks on both datasets. As these are made up of images of blue surgical masks for the most part. To mitigate this, both datasets were altered by a python algorithm, which detects areas with the presence of a desired color range (utilizing the HSV color space), in this case, light blue color shades present in surgical masks. After detecting said matching areas, other colors and tonalities provided by the BGR color space were used to replace them in such a manner that the variety of masks found were closer to those in reality. The algorithm randomly returned four different mask types: the original unchanged masks, dark colored masks, light colored masks and random colored masks with and without patterns, where the patterns are mixtures of different color tones.

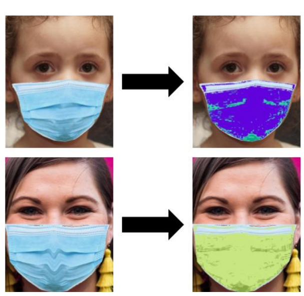

Figure 1. Color changes made to examples.

Both Mask-net and I-Mask-net were designed with similar architectures making use of Transfer Learning, choosing MobileNet as the preferred base model which is pretrained utilizing the ImageNet dataset.

The Mask-net model consist of 85 deep layers, 82 of which are MobileNet attribute extraction layers and the rest consist of fine tuning adjustment layers, which include a hidden layer of 512 nodes followed by a Dropout layer with an index of 0.2, culminating in a Softmax output layer for the three corresponding classes.

On the other hand, I-Mask-net differs from Mask-net, in that it has 84 deep layers, connected to a single Dropout layer with an index of 0.6, culminating in a Softmax output layer for the three corresponding classes.

For Mask-net, the last 23 layers were trainable, and regarding I-Mask-net, the last 34 layers were trainable. By freezing these layers, Mask-net was allowed to retain knowledge of general attributes extracted from the ImageNet dataset, presenting similar classes to KLD such as mask, respirator, gas mask, ski mask, and oxygen mask. On the other hand, I-Mask-net had less frozen layers due to the fact that there were no classes that solved a similar issue, which is the reason why it was necessary to train more layers specific to the task.

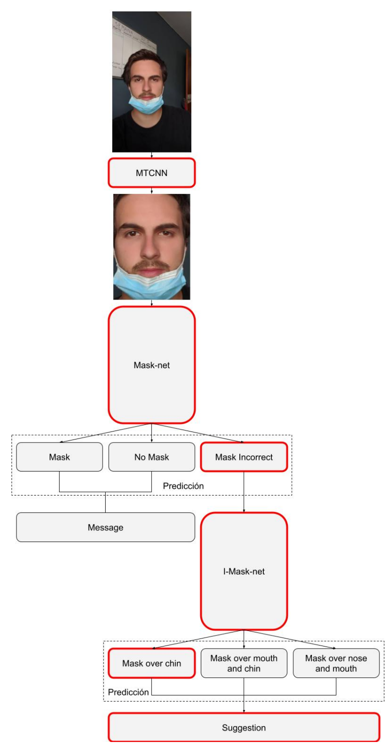

Figure 2. Image of an individual with a mask over their chin. The Mask-net model classifies it as "Mask Incorrect", which promptly feeds the image into the I-Mask-net model, which classifies it as "Mask Over Chin".

3. Results

Table 1. Mask-net's precision, recall and F1-score per class, on KLD Test Set (170 samples).

Class	Precision	Recall	F1-Score
Mask	0.7638	0.9821	0.8593
No Mask	0.9130	0.7368	0.8154
Mask Incorrect	0.9615	0.8771	0.9173

The results found on this research are promising. Using a group of 170 images corresponding to the Test Set of KLD, predictions were run on the Mask-net model where it presented an overall accuracy of 86.47%. As seen on Table 1, precision metrics show that ”No Mask” and ”Mask Incorrect” classes have a low amount of false positives among the whole Test Set, with ”Mask” class having the lowest precision, a score of 0.7638, which remains an acceptable metric. Recall shows correct predictions throughout each class, with ”Mask Incorrect” and ”Mask” classes with a recall score of 0.8771 and 0.9821 respectively, while the ”No mask” class has a recall score of 0.7368, which is still considered a positive performance.

Mask-net presents some prediction faults, particularly in the classes ”Mask Incorrect” and ”No Mask”. The ”Mask Incorrect” false negatives occurred especially in images where the subject's chin was uncovered, this is likely due to the similar visual look of this particular ”Mask Incorrect” case and the ”Mask” class, being mistakenly determined as the later, hence its lower precision score. In contrast, “No Mask” prediction faults are mostly due to subjects attempting to trick the model by using objects, or even their hands, to cover their faces.

Table 2. I-Mask-net's precision, recall and F1-score per class, on IMKLD Test Set (57 samples).

Class	Precision	Recall	F1-Score
Mask over Chin	0.8888	0.8421	0.8648
Mask over Mouth and Chin	0.8750	0.7368	0.7999
Mask over Nose and Mouth	0.8260	1.0000	0.9047

On the other hand, Using a group of 57 images corresponding to the Test Set of IMKLD, predictions were run on the I-Mask-net model obtaining an overall accuracy of 85.96%. As shown in Table 2, precision gets its best results when the predicted class is ”Mask over Chin”, with a precision score of 0.8888, followed by the ”Mask over Mouth and Chin” class with a precision score of 0.8750, and lastly ”Mask over Nose and Mouth” class, with a precision score of 0.8260. In regards to recall, the ”Mask over Nose and Mouth” class obtained a recall score of 1.0 (all labeled images predicted correctly), followed by the classes ”Mask over Mouth and Chin” and ”Mask over Nose and Mouth”, obtaining a recall score of 0.8421 and 0.7368 respectively, an acceptable result. Taking a look at the obtained F1-Scores in Table 2, which is a weighted average between precision and recall, results across all classes show an acceptable performance since they are close to 1.0.

Generally, prediction errors made by I-Mask-net on the IMKLD Test Set occurred most likely due to a lack of generalization in some of the test samples, attributable to using artificially made images for training. This likely happens as a consequence to both the IMKLD Training and Validation Sets not being large enough to properly generalize. The sizing issue also continues on to testing, owed to the small size of the IMKLD Test Set, making it difficult to create a proper statistical representation.

4. Conclusions

This paper proposes a deep learning system for detecting the incorrect use of a mask. From the existing results, it can be concluded that classification CNNs are capable of solving this problem with acceptable levels of accuracy, and not only identify medical masks, but also identify a diversity of cloth masks in a variety of colors and patterns. Also, in addition to detecting the correct and incorrect use of the mask, the I-Mask-net model makes users aware of how they're using it wrongly. Furthermore, such a system could represent a novel way to solve issues created by the COVID-19 Pandemic, such as overseeing mandatory compliance of face mask mandates in applicable areas, like department stores, public transportation, governmental buildings and others.

5. Recommendations

This research is currently a work-in-progress, with possibilities to scale and improve. To do so, the following suggestions are recommended:

- Expand KLD and IMKLD datasets in order to cover more real life scenarios such as images with people in different angles, with different facial hair styles and accessories.
- Alternatively, both datasets could be merged into a single one, regarding five classes: “Mask”, “No Mask”, ”Mask over Chin”, ”Mask over Mouth and Chin” and “Mask over Nose and Mouth”, with the aim of training a multi-class model, such as Mask-net, and thus comparing whether this option is more efficient than using

two independently trained models, such as those found in this paper.

- Finally, it's proposed to test different models of transfer learning in order to judge their feature extraction ability and therefore evaluate if the performance of the models can be improved.

References

Cabani, A., Hammoudi, K., Benhabiles, H., and Melkemi, M. MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart Health*, 19:100144, Mar 2021. ISSN 2352-6483. doi: 10.1016/j.smhl.2020.100144. URL http://dx.doi.org/10.1016/j.smhl.2020.100144.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Humans in the Loop. Medical mask dataset - free ai/ml dataset. https://humansintheloop.org/resources/datasets/medical-mask-dataset/#:~:text=Humans%20in%20the%20Loop%20is,ethnicities%2C%20ages%2C%20and%20regions., 2020.

Karras, T., Laine, S., and Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks, 2019.

Magarakis, L. L. and Johns Hopkins Medicine. How to Properly Wear a Face Mask: Infographic. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/proper-mask-wearing-coronavirus-prevention-infographic, Sep 2020.

Mitze, T., Kosfeld, R., Rode, J., and Wälde, K. Face Masks Considerably Reduce Covid-19 Cases in Germany - A Synthetic Control Method Approach. CESifo Working Paper Series 8479, CESifo, 2020. URL https://ideas.repec.org/p/ces/ceswps/_8479.html.

World Health Organization. Coronavirus disease (COVID-19). https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19, 2020a.

World Health Organization. WHO Coronavirus (COVID-19) Dashboard — WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. https://covid19.who.int/?gclid, 2020b.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, Oct 2016. ISSN 1558-2361. doi: 10.1109/lsp.2016.2603342. URL http://dx.doi.org/10.1109/LSP.2016.2603342.