Analysis of Masked Facial Recognition Algorithms

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Edward Shen
Spring, 2021

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature Edward Shen Date 5/6/2021
Edward Shen

Approved ____ Date __________
David Evans, Department of Computer Science

Analysis of Masked Facial Recognition Algorithms

Edward Shen
Department of Computer Science
University of Virginia
es6xs@virginia.edu

ABSTRACT

The use of Facial Recognition technology in society has seen considerable increases in its applications and uses. One notable example is Face ID, a common feature among smartphones that allows a user to authenticate using a live image of their face. However, with the onset of widespread mask usage largely due to the ongoing pandemic dubbed COVID-19, Face ID systems are no longer able to identify a person as their lower face is covered. This research investigates novel approaches to facial recognition for users wearing masks and analyzes the current research done regarding this problem. One notable peer reviewed paper recently published on this topic was analyzed to determine the methodology needed for accurate facial recognition systems to perform well when identifying masked individuals. Alongside this paper, two others were analyzed in order to support the findings in the original. From these papers, we see that facial recognition can reach levels of accuracy comparable to contemporary facial recognition algorithms even with face masks obscuring features.

1 INTRODUCTION

Today's world largely hinges on the ongoing pandemic. Right now, both the positive and negative effects of COVID-19 determine how we go about our daily lives. One of the biggest changes that we as a society have adopted is the widespread use of facial masks in order to reduce the spread of the virus. While this change seems rather small, considering many countries in parts of Asia already entertained the idea of using masks to decrease the transmission of germs, this idea is very foreign for many western nations. One technically significant problem that has been introduced as a byproduct of wearing masks is accurate facial recognition. Masks, while medically beneficial, create algorithmic problems when it comes down to programs for facial recognition tasks. Naturally, wearing a mask will hide many of one's discerning facial features. Facial features that the most common modern day facial recognition technologies specifically attempt to identify and use in the mapping of one's face.

There are several means by which this problem can be addressed. Specifically, the means that are outlined in "Efficient Masked Face Recognition Method During the COVID-19 Pandemic" by Walid Hariri alongside "Towards Facial Recognition Problem in COVID-19 Pandemic" by Mundai et. al. and "Masked Face Recognition for Secure Authentication" by Anwar and Raychowdhury. As outlined in Hariri's paper, facial recognition with masks encounter the following problems: 1) fraudsters and thieves take advantage of the mask, stealing and committing crimes without being identified, 2) community access control and face authentication have become very difficult tasks when a grand part of the face is hidden by a mask, 3) existing face recognition methods are not efficient when wearing a mask which cannot provide the whole face image for description, 4) exposing the nose region is very important in the task of face recognition since it is used for face normalization, pose correction, and face matching. Due to these problems face masks pose a significant challenge to existing facial recognition methods. These challenges can be separated into two different problems, face mask recognition and masked facial recognition. Face mask recognition is the problem of using computer vision in order to analyze an image of a face to determine whether or not the subject is wearing a face mask. Masked facial recognition is the problem of identifying or labeling someone as non-masked facial recognition algorithms do with people without masks on, with the extra challenge of having occluded features on a face.

With any machine learning project, having a large dataset is paramount to achieving high accuracy on a model. Because the COVID-19 pandemic is a very recent development, open-source images of people wearing face masks are not very common relative to contemporary facial recognition databases such as ImageNet. Thus, it is necessary to gather enough data in the form of masked faces in order to train a model to recognize and identify masked faces. Towards this, Anwar and Raychowdhury proposed a new open-source tool dubbed MaskTheFace, which can extend a face dataset to include masked faces. This tool uses a dlib (deep learning library in C++) based landmarks detector to identify six key features of the face necessary for applying a mask. From this, MaskTheFace places a mask from a predetermined collection of masks onto the face. This tool is capable of converting any normal face dataset into a masked face dataset. MaskTheFace proposed a series of features beneficial to generating a masked face dataset. Support for multiple masks types allows blue surgical masks, N95 masks, and three others, and allows the addition of other mask types. It also includes mask variations, adding colors and patterns to masks to represent real world variations in masks color and pattern. Its wide face angle coverage feature allows for faces of varying degrees of tilt to also be used, therefore keeping the dataset as variable as possible in order to ensure a higher model accuracy.

The primary work analyzed in this paper, Efficient Masked Face Recognition Method during the COVID-19 Pandemic, proposes new deep learning-based methodology to tackle masked facial recognition. The methodology used in Hariri's work will be discussed in the following section. The results of Hariri's work will be discussed in the Results section, and will be compared to the results of the similar work done by Mundial et. al.. The analysis section of this paper will discuss Hariri's work, identifying potential flaws, questions, potential insights, and further work with supporting information from Mundial et. al. and Anwar and Raychowdhury. The conclusion section of this paper will summarize key findings and discuss the significance of this work.

2 RELATED WORK

Besides the Efficient Masked Fact Recognition Method to tackle masked facial recognition, there are other approaches to handle the facial occlusion challenge. To be more specific, they can be classified into three categories. The first category is the local matching approach. This approach given the name, compares images using a matching algorithm. Usually, the target face image is compared with a sample of other patches, all of which are roughly the same size. Consequently, the features are extracted from each patch and finally, a matching algorithm is used between probe and gallery faces. Furthermore, instead of grouping them into patches of the same size, other approaches strive to detect for keypoints in the facial image. To be specific, this is where they pinpoint key points of the face and textural with geometrical extraction is applied onto these keypoints. Then, set point matching is applied to fully match the distinct and similar features, thus able to recognize persons of interests given partial faces. Other similar methods like the former include Keypoint based matching where this approach uses a SIFT descriptor to find the keypoints as well as applying Gabor ternary pattern for the matching algorithm.

Moving on, the second category is the restoration approach. Here, the occluded regions in the facial image are restored by comparing them with gallery faces. This can be carried out in many ways. One way is to calculate the depth map values of the 3D images by applying thresholding methods and finally applying Principal Component Analysis. Another way is to estimate the covered regions of the facial image. Essentially, apply statistical models or algorithms, such as the iterative closest point algorithm, to predict the partial facial curves in the occluded regions.

Last but not least, the third category is the discard occlusion-based approach. Here, this approach aims to find regions determined to be occluded and remove them from the facial image to ease the feature extraction and classification process. To do this, there is a Segmentation based approach in which it divides the facial image into smaller segments and uses SVM classifiers to detect which patches to remove. At the end, the Mean based weight matrix is applied onto the revised facial image. Other ways besides the former is to use global masked projection or partial matching mechanisms to remove the occluded segments in facial images and for the restoration, approaches such as the partial Gappy PCA, which utilizes eigenvectors, are quite effective.

3 SYSTEM DESIGN

The overarching idea of Hariri's method is to separate the masked part of an image from the rest of the face and run a deep learning model on the resulting images. A cropping filter is applied to the base image that identifies the face and the mask and removes the mask, leaving only the eyes, eyebrows, and forehead. Another idea presented in this paper is an alternative to the standard computationally expensive deep learning architectures. Hariri proposes the use of an efficient quantization based pooling method for face recognition using the VGG-16 pre-trained model. To do this, only the feature maps on the last layer of the convolutional neural network are considered using the Bag of Words paradigm. The key concept of Bag of Words is to represent features as an orderless set. Local features are first extracted from training images, with each feature representing a part of the image. These features are then quantized to form a codebook. When testing, features from a test image are assigned to the nearest feature in the codebook and used to determine an output. In this circumstance, Bag of Words is used as a pooling layer in the convolutional neural network in order to reduce the number of parameters in the network.

The deep quantization technique specified above allows a lightweight representation of real world masked facial recognition. Additionally, this deep quantization method has the ability to classify images of different sizes, in the case where unmasked regions of a processed image are of different sizes. Furthermore, the quantization method allows it to be trained alongside the rest of the model, allowing a better quantization scheme compared to a fixed quantization model. Note, this method doesn't need to be retrained on maskless images, it can be used instead to improve generalizations of facial recognition with masks during the COVID-19 pandemic.

3.1 Preprocessing and Cropping Filter

The dataset used by Hariri was pre cropped to the subject's face, eliminating the need for a face mask recognition algorithm to identify faces and center them in an image. In order to be able to remove the mask from the image as stated previously, it is necessary to rotate the image. To do so, facial landmarks are detected using the dlib open-source computer vision library and rotated using a 2D rotation based on the positions of landmarks identified. To extract the non-masked portion, the rotated image is scaled to a 240x240 pixel image. Next, the image is partitioned into a total of one hundred 24x24 blocks. Then, blocks containing parts of the mask are dropped, leaving only the upper portion of the face in the image.

3.2 Feature Extraction Layer

Deep features are extracted using the VGG-16 face CNN descriptor from the images. The VGG-16 model is a pretrained model trained on the ImageNet dataset. This model contains a variety of layers consisting of convolution layers, max pooling layers, activation layers, and fully connected layers. Specifically, VGG-16 consists of 13 convolutional layers, 5 max pooling layers, 3 dense layers, for a total of 21 layers and 16 weighted layers. For this research, only the feature maps outputted from the last convolutional layer, also called channels, are used in the quantization stage.

3.3 Deep Bag of Features Layer

From the last layer specified above, for any image feature maps are extracted in this way. To measure the similarity of the feature vectors from the mapping and the feature vectors, also called codewords in Hariri's paper, stored in the codebook, and RBF kernel is used as a similarity metric. Thus, the RBF kernel is the next sublayer added to the VGG-16 CNN, with each neuron in the layer representing a codeword. The RBF neurons are initialized either manually or automatically on all the extracted feature vectors across the entire dataset. The main automatic algorithm used to determine the codewords is k-means, which is an obvious choice as it easily groups the feature vectors into n codewords. This process can be done afterwards to get the best RBF centers.

The quantization step uses the RBF neurons to extract a histogram with the number of bins equal to the number of codewords. Another layer is added after the RBF layer called the quantization layer which collects the output of the RBF layer and consolidates it into a histogram. After determining the final histogram, it is passed to the classification stage where test images are assigned an identity. The classifier used is the Multilayer perceptron classifier, which represents each face as a term vector. The Deep Bag of Features network is capable of being trained with back-propagation and gradient descent, providing increased accuracy of this layer. The Multilayer perceptron classifier uses the term vector as input values and uses a sigmoid function to sum the weights and map to the output, an identity.

RESULTS

The efficacy of Hariri's proposed algorithm was assessed on publicly available masked face datasets. The main dataset used was the Real-World Masked Face Dataset, curated in order to improve facial recognition performance on masked faces during the COVID-19 pandemic. Specifically, there are three subsets inside this dataset: Masked Face Detection Dataset, Real World Masked Face Recognition Dataset, and Simulated Masked Face Recognition Dataset. The first subset contains 24771 masked face images, the second subset contains 5000 pictures of 525 people wearing masks and 90000 pictures of the same 525 people without masks, and the third contains 500000 simulated masked images of 10000 different people. In Hariri's paper, he evaluated his proposed model on the second set.

After processing an image and passing it through the model as specified in the methodology section, Hariri tested three different bin sizes, 50, 60, and 70. In the experiment, a 10 fold cross-validation strategy was used to evaluate the accuracy of the model. These experiments were each repeated 10 times on the second dataset, where 9 samples are used as the training set and the remaining sample used as a testing set. The following table details the results of the experiment. The best recognition rate was 91.3% with the third feature mapping with 60 bins.

Method	**Size 1**	**Size 2**	**Size 3**
term vectors	50	60	70
Conv5 FM1	88.5%	89.2%	87.1%
Conv5 FM2	90.8%	87.4%	87.2%
Conv5 FM3	91.0%	**91.3%**	90.1%

Table 1: Results

The high accuracy is attributed to having the best features being extracted from the last convolutional layer of the VGG-16 CNN and the high efficiency of the Bag of Features paradigm that gives a lightweight alternative to a traditional softmax function. Only applying the algorithm on informative regions while eliminating non-informative regions and the high level of generalization of the algorithm makes it much faster and applicable to real time applications. Notice the middle-ranged quantity of term vectors had the best performing recognition rate even compared to one with more term vectors. This behavior could be attributed to the proposed Deep Bag of Features method depending on the number of extracted deep features in an image. Changing the size of the pooling layer can increase scale-invariance and bring more spatial information to the fully connected layer.

ANALYSIS

Hariri's work proposing a new process of tackling the problem of masked facial recognition is a fascinating approach to problems usually solved by brute force in machine learning. Instead of simply passing a model massive quantity of data for it to eventually learn how to classify masked images, he redesigns the architecture itself to better suit this problem. The preprocessing step of scaling images to 240x240 pixels is a standard preprocessing step but dividing it into boxes and truncating it so masked regions are completely removed is not. Furthermore, the bag of features method proposed seems extremely useful in other machine learning problems, as it can take in images of different sizes and still be able to map it into a machine learning model. However, one mention about the preprocessing done in Hariri's work, the dataset used was already pre cropped to contain only the face. In a real-world scenario, the face is not guaranteed to at the center of the image. If this process was applied on data that was not already cropped in this way, the algorithm would fail to perform the necessary preprocessing steps and would not function as proposed. This could be easily fixed by addressing the other problem with facial recognition brought up by masks, face mask recognition. This exact process was defined in Mundial et. al., where a face detector was applied to raw images in the dataset and the detected face was isolated and resized and normalized after. Here, if an algorithm can identify a face mask on a person's face, it can isolate the face and perform the cropping as a postprocessing step, and feed that into the preprocessing step of the masked facial recognition algorithm proposed by Hariri.

Noting the dataset choice by Hariri, he chose to use the Real-World Masked Face Dataset. This choice was justifiable as it is the only large dataset of face mask images. However, his choice to use the smaller subset, the Real-World Masked Face Recognition Dataset, as opposed to the significantly larger Simulated Masked Face Recognition Dataset, raised some concerns. The Simulated Masked Face Recognition Dataset consists of images of people from normal face datasets, but have masks digitally added to the face. This process is detailed in Anwar and Rawchowdhury's paper, Masked Face Recognition for secure authentication. Their tool MaskTheFace takes normal face images and superimposes a variety of masks onto the image. Hariri's choice of dataset consisted only of 5000 masked images and 525 unique people. In a machine learning context, this amount of data may not be significant enough to come to a significant conclusion. Here, further work can be done continuing the proposed method and testing it on a larger dataset or the Simulated Masked Face Recognition Dataset.

Looking at the architecture of Hariri's proposed system, it is a different approach to facial recognition compared to the well known deep learning models such as ResNet, AlexNet, etc. The output of the convolutional neural network here isn't a classification, but rather a feature mapping that is used as an input. This is similar to a more contemporary CNN where all but the last few layers are used to generate inputs similar to a feature mapping, and the last few layers are used to classify those inputs. Hariri's histogram based deep bag of features idea, however, is not present in other models. The trainability of the histogram was also an interesting point made, as it would improve its distribution over epochs to better organize the bins. The time advantages of this method make it lucrative for applying facial recognition with masks in real time, however, as computing technologies have advanced and can operate in real time with heavier models, it is not necessarily a major benefit when trading off accuracy.

The highest accuracy of 91.3% of Hariri's proposed model seems to perform well based on the lightweight nature of his model and the inherent difficulty of masked facial recognition operating with significantly reduced facial features to process. Perhaps with more training data the model could reach higher accuracy, as with only 5000 masked images it may not have enough training data. On the other hand, because there are only 5000 masked images used, it could be possible that the model suffers from a degree of overfitting, which can explain lower testing accuracy.

CONCLUSION

The modern world is constantly evolving to present new issues within the field of biometric security. Changes in the lifestyle and functionality of human life often provide the spark needed for computer scientists to develop much more complex and intricate solutions to the problems which arise. A concept which is made only even more apparent when considering the circumstance of the present day and the ongoing COVID-19 pandemic. Facial recognition as an approach was forced to adapt as well. The mask mandate moved to only generate complexity in the nature of the algorithms which were dedicated to facial analysis. In the process of covering up half of an individual's face, data that was already limited was cut down even further. A situation arose which demanded a solution from the field it came from. A solution which would have to put in almost double the work that is done by modern day facial recognition algorithms with half the amount of data that is offered. A new approach to a problem which had already exemplified itself in the past for being notoriously complex in the very first place.

There is not necessarily one solution which can be considered ideal in this circumstance. Evidence of which can be seen from what we previously discussed. This extension of facial recognition with masks can be chalked up to being relatively unexplored and extremely novel. However, after the events that have transpired between the years of 2020 and 2021, society as we know it has changed. The idea of utilizing masks will most likely stick around long after the conclusion of this pandemic and far into the days of the future.

REFERENCES

[1] Walid Hariri. 2020. Efficient Masked Face Recognition Method During the COVID-19 Pandemic. In *Research Gate. DOI: 10.21203/rs.3.rs-39289/v1*

[2] Aqeel Anwar and Arijit Raychowdhury. 2020. Masked Face Recognition for Secure Authentication. In *arXIv.org.* Retrieved from: https://arxiv.org/abs/2008.11104v1

[3] Imran Mundial, Sohaib Ul Hassan, Islam Tiwana, and Waqar Qureshi. 2020. Towards Facial Recognition Problem in COVID-19 Pandemic. In *2020 4th International Conference on Electrical, Telecommunication and Computer Engineering. DOIi: 10.1109/ELTICOM50775.2020.9230504.*