Analysis of face detection, face landmarking, and face recognition performance with masked face images

Ožbej Golob

Faculty of Computer and Information Science,
University of Ljubljana, Večna pot 113, 1000 Ljubljana
E-mail: ozbej.golob@gmail.com

Abstract

Face recognition has become an essential task in our lives. However, the current COVID-19 pandemic has led to the widespread use of face masks. The effect of wearing face masks is currently an understudied issue. The aim of this paper is to analyze face detection, face landmarking, and face recognition performance with masked face images. HOG and CNN face detectors are used for face detection in combination with 5-point and 68-point face landmark predictors and VGG16 face recognition model is used for face recognition on masked and unmasked images. We found that the performance of face detection, face landmarking, and face recognition is negatively impacted by face masks.

1 Introduction

Face recognition has become an essential task in our daily life. The wide availability of powerful and low-cost computing systems has popularized face recognition for a variety of applications, including biometric authentication, surveillance, human-computer interaction, and multimedia management [1]. Many highly accurate face recognition models were developed. Some of those models are FaceNet [2], DeepFace [3], and DeepID3 [4]. However, face occlusion can negatively impact the accuracy of face recognition models. This problem has already been addressed in the scope of face detection by Opitz et al. [5]. Some research has also been carried out by Song et al. [6] to develop occlusion invariant face recognition solutions. However, this research focuses on typical in-the-wild occlusion scenarios such as wearing sunglasses.

The current COVID-19 pandemic has led to the widespread use of face masks to prevent the spread of the disease. For this reason, it is essential to research the specific effect of wearing face masks on the performance of face recognition models.

Face landmarking represents one of the first steps in a standard face recognition pipeline. During landmarking, the location of certain facial features (eye corners, the tip of the nose, etc.) is identified in the face images and used to align faces prior to feature extraction. If the landmarks are not detected properly, the alignment procedure will fail and result in poorly aligned or partial facial areas that will ultimately affect face recognition performance. In this paper, we explore how face detection, face landmarking, and face recognition work with masked faces.

2 Related work

Damer et al. [7] study the effect of masked faces on the behavior of three top-performing face recognition systems. Two of these algorithms are academic approaches, namely the ArcFace [8] and SphereFace [9]. The third algorithm is a commercial off-the-shelf (COTS) from the vendor Neurotechnology [10]. Authors evaluate the face verification performance without masks and compare results to the face verification performance with masks. The verification performance of the ArcFace and SphereFace is negatively affected when the faces are masked, while the COTS is not significantly affected by masked faces.

Wang et al. [11] propose three publicly available masked face datasets: Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD), and Simulated Masked Face Recognition Dataset (SMFRD). Authors propose a face-eye-based multi-granularity recognition model where they apply different attention weights to key features visible in masked faces (face contour, ocular and periocular details, forehead, etc.). The model improves the recognition accuracy of masked faces from the initial 50% to 95%.

3 Methods

3.1 Face detection

We implemented two face detectors: (i) Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, an image pyramid, and sliding window detection scheme, and (ii) Max-Margin Convolutional Neural Network (CNN) face detector. The HOG detector is accurate and computationally efficient while the CNN detector is accurate and robust, capable of detecting faces from varying viewing angles, lighting conditions, and occlusion. Both face detectors are implemented in Dlib [12].

3.2 Face landmarking

We implemented two landmark predictors: (i) 68-point landmark predictor (see Figure 1), and (ii) 5-point landmark predictor (see Figure 1, points 34, 37, 40, 43, and

46). Both landmark predictors are Dlib's implementation of [13].

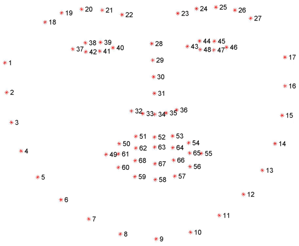

Figure 1: 68-point landmarks.

Face landmarking performance was evaluated with the normalized root mean square error (NRMSE). The normalization is done with respect to the inter-ocular distance (IOD), which is the distance between the two eye centers. Normalizing landmark localization errors with the IOD makes the performance evaluation independent of the face size or the camera zoom factor.

The normalized distance δ is computed as the Euclidean distance $d(.,.)$ between the ground-truth landmark coordinates (x, y) and the predicted landmark coordinates $(\tilde{x}, \tilde{y})$, normalized by the IOD. Equation 1 shows the formula for the normalized distance of each landmark, where subscript k indicates one of the landmarks.

$$\delta_k = \frac{d\{(x_k, y_k), (\tilde{x_k}, \tilde{y_k})\}}{IOD} \tag{1}$$

NRMSE of each image ($NRMSE_{local}$) is calculated by the formula shown in Equation 2, where n is the number of landmarks.

$$NRMSE_{local} = \sqrt{\frac{\sum\limits_{k=1}^{n} \delta_k^2}{n}} \tag{2}$$

3.3 Face recognition

VGG16 architecture which is pre-trained on a huge ImageNet database with more than 1 million images belonging to 1000 different categories is used to train the input face images. The Softmax layer is removed in order to get image feature vectors. The extracted features are fed as input to the Fully Connected Layer and Softmax activation. Figure 2 shows the VGG16 architecture.

4 Experiments

4.1 Datasets

To evaluate face landmarking performance, we identified a couple of datasets with annotated facial landmarks. These datasets are Labeled Face Parts in the Wild (LFPW) [14], Annotated Faces in the Wild (AFW) [15], HELEN [16], and IBUG [17]. All used datasets are collected in the wild. The LFPW dataset consists of 1,432 faces, where 1,132 images are a part of the train set and 300 images are a part of the test set. The HELEN dataset is composed of 2330 face images, where 2000 images are a part of the train set and 330 images are a part of the test set. The AFW dataset contains 205 images with 468 faces. The IBUG dataset includes 135 images with extreme poses and expressions. Evaluation on LFPW and HELEN was performed on the test set only while evaluation on AFW and IBUG was performed on the whole dataset. All used datasets were re-annotated with 68 landmarks as a part of 300 Faces In-the-Wild Challenge (300-W) [18] using the mark-up of Figure 1.

To evaluate face landmarking performance on masked faces, we generated masked versions of LFPW, AFW, HELEN, and IBUG datasets with the help of the MaskTheFace tool [19].

To evaluate face recognition performace, we used a subset of CASIA dataset. 960 train images and 417 test images, belonging to 10 identities, were used for evaluation. To evaluate face recognition performance on masked faces, we generated masked images with the help of the MaskTheFace tool.

4.2 Experimentation details

Face detection results were measured as a percentage of annotated faces that the detector was able to detect. Face landmarking results were measured as NRMSE. Figure 3 shows positions of ground truth and predicted landmarks. The distance between ground truth and predicted landmarks (see Equation 1) is calculated for each landmark and NRMSE (see Equation 2) is calculated for each image. NRMSE is then averaged for the whole dataset.

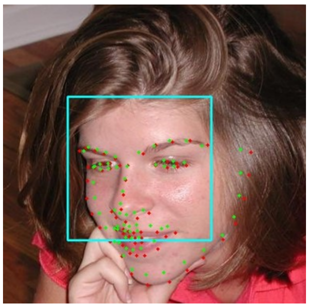

Figure 3: Face landmarks (red dots represent ground truth landmarks, green dots represent predicted landmarks).

We ignored images where zero faces were detected and images where the annotated face was not detected. Images with errors were ignored because they increased the NRMSE significantly.

VGG16 model for face recognition was initialized with weights learned from a pre-trained model. VGG16 model was additionaly trained on unmasked images and evaluated on unmasked and masked images. Model was then re-trained and evaluated on masked images.

5 Results

5.1 Face detection

Table 1 shows the HOG face detection accuracy on original and masked datasets.

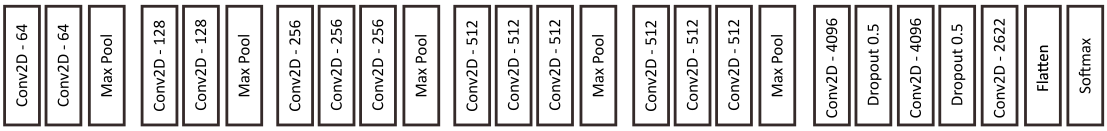

Figure 2: VGG16 architecture.

Table 1: HOG face detection accuracy.

Dataset	Original	Masked
HELEN	0.967	0.735
LFPW	0.987	0.801
AFW	0.896	0.565
IBUG	0.711	0.487
Average	0.890	0.647

Table 2 shows the CNN face detection accuracy on original and masked datasets.

Table 2: CNN face detection accuracy.

Dataset	Original	Masked
HELEN	1.000	0.892
LFPW	1.000	0.878
AFW	1.000	0.708
IBUG	0.941	0.635
Average	0.985	0.778

5.2 Face landmarking

Table 3 shows the Dlib 68-point face landmarking NRMSE with HOG face detector on original and masked datasets.

Table 3: Dlib 68-point face landmarking NRMSE with HOG face detector.

Dataset	Original	Masked
HELEN	0.044	0.153
LFPW	0.054	0.159
AFW	0.063	0.169
IBUG	0.102	0.204
Average	0.066	0.171

Table 4 shows the Dlib 68-point face landmarking NRMSE with CNN face detector on original and masked datasets.

Table 5 shows the Dlib 5-point face landmarking NRMSE with HOG face detector on original and masked datasets.

Table 6 shows the Dlib 5-point face landmarking NRMSE with CNN face detector on original and masked datasets.

Table 4: Dlib 68-point face landmarking NRMSE with CNN face detector.

Dataset	Original	Masked
HELEN	0.065	0.187
LFPW	0.073	0.189
AFW	0.114	0.214
IBUG	0.209	0.266
Average	0.115	0.214

Table 5: Dlib 5-point face landmarking NRMSE with HOG face detector.

Dataset	Original	Masked
HELEN	0.034	0.105
LFPW	0.050	0.113
AFW	0.062	0.123
IBUG	0.092	0.172
Average	0.060	0.128

Table 6: Dlib 5-point face landmarking NRMSE with CNN face detector.

Dataset	Original	Masked
HELEN	0.037	0.114
LFPW	0.052	0.115
AFW	0.070	0.136
IBUG	0.107	0.152
Average	0.067	0.129

5.3 Face recognition

VGG16 model trained on unmasked images achieved test accuracy of 0.966 on unmasked images and 0.867 on masked images. After the model was re-trained on masked images, the model achieved 0.952 accuracy on masked images.

6 Conclusion

In the scope of this paper, we analyzed face detection, face landmarking, and face recognition performance with masked face images. We observed a negative impact of face masks on face detection. Face detection accuracy dropped by 0.243 on average for the HOG face detector and by 0.207 on average for the CNN face detector for masked images. Face landmarking performance was also negatively impacted by face masks. NRMSE of the 68-point face landmarking predictor increased nearly 3-times with the HOG face detector and nearly 2-times with the CNN face detector for masked images. NRMSE of the 5-point face landmarking predictor increased by approximately 2-times for HOG and CNN face detector for masked images. Face recognition performance was also negatively impacted by face masks, lowering the accuracy by 0.099 on masked images. We were able to improve the accuracy to 0.952 by additionally training the model on masked images.

We conclude that face masks have a negative impact on face detection, face landmarking, and face recognition. This implies that facial areas beneath face masks (mouth, nose, etc.) hold significant information for face detection, face landmarking, and face recognition. We found that certain models can be improved by additionally training the model on masked images.

References

[1] Anil K Jain and Stan Z Li. *Handbook of face recognition*, volume 1. Springer, 2011.

[2] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[3] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[4] Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang. Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*, 2015.

[5] Michael Opitz, Georg Waltner, Georg Poier, Horst Possegger, and Horst Bischof. Grid loss: Detecting occluded faces. In *European conference on computer vision*, pages 386–402. Springer, 2016.

[6] Lingxue Song, Dihong Gong, Zhifeng Li, Changsong Liu, and Wei 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 (ICCV)*, October 2019.

[7] Naser Damer, Jonas Henry Grebe, Cong Chen, Fadi Boutros, Florian Kirchbuchner, and Arjan Kuijper. The effect of wearing a mask on face recognition performance: an exploratory study. In *2020 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–6. IEEE, 2020.

[8] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[9] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017.

[10] Neurotechnology MegaMatcher 11.2 SDK. `https://www.neurotechnology.com/`. Accessed: 2022-02-01.

[11] Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al. Masked face recognition dataset and application. *arXiv preprint arXiv:2003.09093*, 2020.

[12] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[13] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[14] Peter N Belhumeur, David W Jacobs, David J Kriegman, and Neeraj Kumar. Localizing parts of faces using a consensus of exemplars. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2930–2940, 2013.

[15] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2879–2886. IEEE, 2012.

[16] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang. Interactive facial feature localization. In *European conference on computer vision*, pages 679–692. Springer, 2012.

[17] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 397–403, 2013.

[18] 300 Faces In-the-Wild Challenge (300-W), ICCV 2013. `https://ibug.doc.ic.ac.uk/resources/300-W/`. Accessed: 2022-02-01.

[19] Aqeel Anwar and Arijit Raychowdhury. Masked face recognition for secure authentication. *arXiv preprint arXiv:2008.11104*, 2020.