On Recognizing Occluded Faces in the Wild

Mustafa Ekrem Erakın*
Dept. of Computer Engineering
Istanbul Technical University
Istanbul, TURKEY
erakin20@itu.edu.tr

Uğur Demir*
Dept. of Computer Engineering
Istanbul Technical University
Istanbul, TURKEY
demirug16@itu.edu.tr

Hazım Kemal Ekenel
Dept. of Computer Engineering
Istanbul Technical University
Istanbul, TURKEY
ekenel@itu.edu.tr

2021 International Conference of the Biometrics Special Interest Group (BIOSIG) | 978-1-6654-2693-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIOSIG52210.2021.9548293

***Abstract*—Facial appearance variations due to occlusion has been one of the main challenges for face recognition systems. To facilitate further research in this area, it is necessary and important to have occluded face datasets collected from real-world, as synthetically generated occluded faces cannot represent the nature of the problem. In this paper, we present the Real World Occluded Faces (ROF) dataset, that contains faces with both upper face occlusion, due to sunglasses, and lower face occlusion, due to masks. We propose two evaluation protocols for this dataset. Benchmark experiments on the dataset have shown that no matter how powerful the deep face representation models are, their performance degrades significantly when they are tested on real-world occluded faces. It is observed that the performance drop is far less when the models are tested on synthetically generated occluded faces. The ROF dataset and the associated evaluation protocols are publicly available at the following link https://github.com/ekremerakin/RealWorldOccludedFaces.**

***Index Terms*—Face recognition, face occlusion, deep learning, real-world occluded faces**

I. Introduction

With the recent advancements in deep learning and its application to computer vision problems, state-of-the-art face recognition systems have achieved excellent results on various datasets, such as LFW [1], AgeDB-30 [2], and MegaFace [3]. As the performance on these well-known datasets converges, researchers started to divert their attention towards more challenging problems. One of these challenges is recognizing occluded faces in the wild [4]. To catalyze further research on this topic, in this paper, we present the Real World Occluded Faces (ROF) dataset, that contains faces with both upper face and lower face occlusions. To test the authenticity of the dataset, we participated in a masked face recognition challenge [5]. Our model, fine-tuned on real life masked images, outperformed models trained on larger, synthetically generated masked face training sets, leading to the best performance among 16 other academic submissions [5].

There have been several works that studied the effects of many different appearance variations on face recognition performance [6], [7], [8]. In this paper, we will be addressing specifically the occlusion problem using a real-world occluded face dataset. [6] used AR face dataset [9] that contains occluded face images collected in a constrained environment, while [7] and [8] used synthetic occlusions built on top of LFW [1]. Our experiments show that real world occlusions are more challenging than their synthetic counterparts.

Previous studies show that deep CNN based face recognition models trained on VGGFace [10], faces major performance drops when confronted by sunglasses and scarves. [6] reports the performance with sunglasses occlusion in the range of 30-35% on the AR face dataset [9], which is a 110 identity face image dataset that is collected in a controlled environment with cooperating subjects, a rather easy benchmark for a modern face recognition model. Another study [7] uses synthetic occlusions to test the face recognition performance again using a VGGFace pre-trained model [10]. Occluded face images are generated by applying black boxes on samples from the LFW dataset [1]. Different occlusion types are simulated by applying these black boxes in different locations. The study reports 25.94% face recognition accuracy against sunglasses effect.

The main contributions of this paper can be summarized as follows: (i) We introduced an in-the-wild occlusion dataset for face recognition, (ii) we proposed two evaluation protocols and analyzed the impact of upper face and lower face occlusion on face recognition performance, (iii) we show that real-world face occlusion poses a more challenging problem for face recognition systems. We also visualized the results and discuss the outcomes in detail.

II. The Dataset

Real World Occluded Faces (ROF) dataset contains face images with real-life upper-face and lower-face occlusions, due to sunglasses and face masks, respectively. The dataset consists of 6421 neutral face images, 4627 face images with sunglasses, and 678 face images with masks. There are 47 subjects with neutral, masked, and sunglasses images, 114 subjects with neutral and sunglasses images, while 20 subjects have only neutral and masked images. The identities are collected from a list of celebrities and politicians. All of the images are from real-life scenarios and contain large variations in terms of pose and illumination. The images were downloaded from Google Image Search using the pipeline described in VGGFace2 study [11]. On average there are 50 neutral images, 30 sunglasses images and 15 masked images per identity.

*Equal contribution

Dataset Collection

Dataset collection is done using a modified version of the pipeline described in VGGFace2 [11]. A name list consisting of public figures, i.e., politicians, celebrities, sports players, etc., were collected. For every name in the list 100 images were downloaded for each type of face image we are after. A reference image was extracted from the collected neutral images for every name using the image size and face count within the image to try to get the best possible reference. Duplicates were removed using perceptual hashing and faces were detected and cropped from the remaining images using a combination of RetinaFace [12] and MTCNN [13].

Then using the reference images and a ResNet50 [14] trained on VGGFace2 [11], face embeddings were extracted for every remaining image and compared with the subject's respective reference image's embedding, using cosine distance as the similarity metric. For neutral images, candidate images with a similarity above 0.5 were selected while for occluded images, the threshold was set to 0.2. Finally, filtered face images were manually verified and stored. Overall, manual work was limited and the bottleneck was finding the appropriate identities that would have both sunglasses and face mask images, which proved to be a niche category. Figure 1 shows sample images from the ROF dataset.

Fig. 1. Samples of neutral, masked, and sunglasses images for the same subjects from the ROF dataset

III. Deep Face Models

We utilized three different deep learning architectures to examine the performance degradation when encountered with facial occlusions, namely ArcFace [15], VGGFace2 [11], and MobilFaceNet [16].

ArcFace [15] is a state-of-the-art face recognition model that achieved excellent performance on various face recognition datasets, such as LFW [1], AgeDB-30 [2], and MegaFace [3]. ArcFace is trained on MS1MV2 [15], which is a revised version of the MS-Celeb-1M dataset [17]. MS1MV2 contains 85,000 identities and 5.8 million images. In this work, we used three different ArcFace architectures to represent various model complexities. Different Arcface architectures are denoted as Arcface-N which corresponds to a ResNet-N model pre-trained on MS1MV2 dataset.

VGGFace2 is a large-scale dataset containing 9131 identities and 3.3 million samples. Researchers used the dataset to train deep learning models, and it was one of the state-of-the-arts. We used their ResNet-50 pre-trained model throughout our experiments [11].

Since ArcFace and VGGFace2 mainly use ResNet as the backbone architecture, their model complexities are not suitable for mobile devices. Therefore, we tested the model also on MobilFaceNet [16] to analyze the performance degradation of a smaller model. MobilFaceNet used in this work is trained on the MS1MV2 dataset [15].

Throughout our experiments, we employed 512-dimensional feature embeddings. For distance metrics, Euclidean distance is utilized for ArcFace and MobilFaceNet, and cosine similarity is used for VGGFace2. ArcFace and MobilFaceNet pre-trained models are adopted from the Insightface repository[1]. VGGFace2 pre-trained model is adopted from the verified VGGFace2 repository.

TABLE I
TOTAL NUMBER OF IDENTITIES AND IMAGES FOR EACH PROTOCOL

Protocol	Identities	Gallery	Synthetic	Sunglasses	Masked
1	161	483	5322	4627	-
2	67	199	1800	-	464

IV. Experimental Setup

In this section, we present the experimental setups to evaluate the occlusion robustness of the deep face recognition methods. We also analyze and compare the differences between the effects of synthetically crafted and real-world occluded face images. We present two experiment protocols. The first protocol investigates the effects of upper face occlusions, while the second one assesses the performance against lower face occlusions. Both protocols also probe with synthetic occlusions and compare the results with the ones obtained on the real world occluded samples. The image and identity counts across protocols are given in Table I.

For data preprocessing, CosFace [18] and SphereFace [19] papers are followed. First, the bounding box and five facial landmarks, namely, eyes, nose, and mouth corners, are obtained using MTCNN [20]. Afterwards, similarity transform is applied to images for face normalization. Then, images are cropped and resized to 112×112.

For testing against upper-face occlusions, we used the ROF sunglasses dataset in the first protocol. We also generated synthetic fixed upper-face occlusions that cover the eye region. Real sunglasses and synthetic occlusions can be seen in Figure 2b and 2d, respectively.

In the second protocol, for lower-face occlusions we used the ROF mask dataset. Synthetic lower face occlusions are generated by fixing the nose and mouth area and using the mask generator [21] published in a recent study. Samples from real and synthetic lower face occlusions can be seen in Figure 2c and 2f, respectively.

[1]https://github.com/deepinsight/insightface

TABLE II
FACE IDENTIFICATION RESULTS USING PROTOCOL 1 (ARCFACE-N DENOTES THE RESNET ARCHITECTURE WITH N LAYERS, FOR VGGFACE2 RESNET50 WAS USED)

	Arcface-100	Arcface-50	Arcface-34	MobilFaceNet	VGGFace2
Occlusion Type	**Top 1**	**Top 1**	**Top 1**	**Top 1**	**Top 1**
No occlusion	99.57%	99.34%	99.17%	98.89%	98.12%
Wearing sunglasses	86.60%	84.18%	83.51%	77.16%	76.83%
Upper occlusion	98.25%	95.92%	95.43%	83.13%	75.65%
Lower occlusion	98.21%	96.81%	96.64%	86.98%	88.56%
Synthetic masked	98.53%	97.16%	96.56%	89.57%	89.59%

TABLE III
FACE IDENTIFICATION RESULTS USING PROTOCOL 2

	Arcface-100	Arcface-50	Arcface-34	MobilFaceNet	VGGFace2
Occlusion Type	**Top 1**	**Top 1**	**Top 1**	**Top 1**	**Top 1**
No occlusion	99.61%	99.33%	99.39%	99.06%	99.28%
Wearing mask	85.34%	76.08%	73.71%	70.04%	79.31%
Upper occlusion	98.39%	96.89%	96.67%	89.78%	89.28%
Lower occlusion	98.83%	97.67%	97.11%	92.06%	93.94%
Synthetic masked	99.00%	97.78%	97.78%	93.22%	94.83%

(a) Neutral Image (b) Wearing Sunglasses (c) Wearing Mask

(d) Upper-face Occlusion (e) Lower-face Occlusion (f) Synthetic Mask

Fig. 2. a) Non-occluded face image, b) Upper face occlusion due to wearing sunglasses, c) Lower face occlusion due to wearing a mask, c) Synthetic upper face occlusion, d) Synthetic lower face occlusion, e) Synthetic mask generation for lower face occlusion

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we present the experimental results. We performed both face identification and verification experiments and assessed the effect of occlusion in both scenarios.

A. Impact of Occlusions on Face Identification

Protocol 1: The experimental results are presented in Table II for all used deep face models. Each row corresponds to the obtained correct classification rates on a specific probe set. Samples from probe sets are shown in Figure 2. All models are found to be very successful when classifying face images that do not contain occlusion, as can be seen from the first row. Arcface is found to be more robust compared to MobileFaceNet and VGGFace2, when a part of the face is occluded synthetically, either by painting the corresponding region with black or generating an artificial mask. However, even Arcface's performance deteriorates when it is tested on real world occluded faces that contain sunglasses.

Protocol 2: In Table III, we present the experimental results using protocol 2. The outcomes are similar to the ones obtained using protocol 1. The deep face models are found to be very successful when there is no occlusion in the probe images. Arcface is found to be superior to MobileFaceNet and VGGFace2, when a part of the face is occluded synthetically, either by painting the corresponding region with black or generating an artificial mask. However, again, even Arcface's performance deteriorates when it is tested on real world occluded faces that contain masks.

These results show that synthetically generated occlusions do not reflect the nature of the real-world occlusions. One reason could be due to the fact that the synthetic occlusions contain the same texture and covers the same regions across different faces. However, real world occlusions contain different textures and cover different parts of the faces depending on the style of the sunglasses or the type of the mask and the way the person wears it.

To analyze the results further, we also visualized the regions that the deep face model focuses using Grad-CAM method [22]. The obtained results are illustrated in Figure 3. As can be seen the model mainly focuses on the inner face region, where eye and nose are contained. This is expected, since, especially, eye region is known to have a high discrimination power. However, as the models learn from the data the highly discriminative parts and focuses on these, when they are occluded they suffer from a performance loss. Therefore, while developing an occlusion-robust deep face recognition system, this fact has to be taken into account.

B. Impact of Occlusions on Face Verification

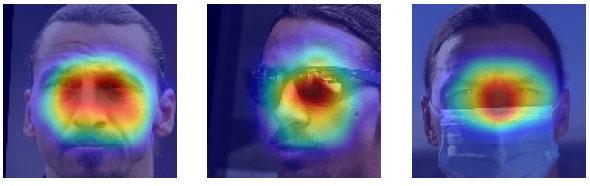
For the sake of completeness, we also run face verification experiments using the proposed ROF dataset. The results of experiments using protocol 1 and 2 are presented in Tables IV and V, respectively. Similar observations can be derived

TABLE IV
FACE VERIFICATION RESULTS USING PROTOCOL 1

	Arcface-100	Arcface-50	Arcface-34	MobilFaceNet	VGGFace2
Occlusion Type	**EER**	**EER**	**EER**	**EER**	**EER**
No occlusion	0.011	0.014	0.014	0.021	0.017
Wearing sunglasses	0.088	0.091	0.095	0.106	0.096
Upper occlusion	0.024	0.036	0.041	0.073	0.076
Lower occlusion	0.021	0.028	0.033	0.054	0.053
Synthetic masked	0.025	0.033	0.036	0.068	0.059

TABLE V
FACE VERIFICATION RESULTS USING PROTOCOL 2

	Arcface-100	Arcface-50	Arcface-34	MobilFaceNet	VGGFace2
Occlusion Type	**EER**	**EER**	**EER**	**EER**	**EER**
No occlusion	0.013	0.019	0.021	0.024	0.017
Wearing mask	0.083	0.119	0.119	0.119	0.082
Upper occlusion	0.031	0.035	0.043	0.067	0.058
Lower occlusion	0.019	0.036	0.038	0.054	0.045
Synthetic masked	0.028	0.036	0.043	0.074	0.049



(a) Neutral Image (b) Wearing Sunglasses (c) Wearing Mask

(d) Upper-face Occlusion (e) Lower-face Occlusion (f) Synthetic Mask

Fig. 3. The regions that VGGFace2 targets the most during embedding extraction [22].

from these experiments: ArcFace is found to be more robust to synthetic occlusions. The EER increases significantly when the models are tested on real world occluded faces.

VI. CONCLUSION

In this study, we present a real-world occluded face dataset and explore the effects of occlusion on the state-of-the-art face recognition methods' performance. We have shown that synthetically generated occlusions do not reflect the nature of the real-world occlusions. We have observed significant performance drops when deep face models are tested on real world occluded faces that contain masks or sunglasses. Visualization of the results indicate that the deep face models mainly focus on the inner face region. Therefore, the models experience a performance loss, when this region is occluded. For our future work, we aim to expand the collected dataset and develop an occlusion-robust deep face recognition system by benefiting from the findings of this work.

ACKNOWLEDGMENT

This study is supported by the Istanbul Technical University Research Fund, ITU BAP, project no. 42547 and by the Scientific and Technological Research Council of Turkey (TUBITAK) project no. 120N011.

REFERENCES

[1] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[2] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "Agedb: the first manually collected, in-the-wild age database," in *CVPR Workshop*, vol. 2, no. 3, 2017, p. 5.

[3] I. Kemelmacher-Shlizerman, S. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *CVPR*, 2016, pp. 4873–4882.

[4] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," in *https://arxiv.org/abs/2006.11366*, 2020.

[5] F. Boutros *et al.*, "Mfr 2021: Masked face recognition competition," in *IJCB 2021*, 2021, pp. 1–10.

[6] M. Ghazi and H. Ekenel, "A comprehensive analysis of deep learning based representation for face recognition," in *CVPR workshops*, 2016, pp. 34–41.

[7] S. Karahan, M. Yildirum, K. Kirtac, F. Rende, G. Butun, and H. Ekenel, "How image degradations affect deep cnn-based face recognition?" in *BIOSIG*. IEEE, 2016, pp. 1–5.

[8] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *IET Biometrics*, vol. 7, no. 1, pp. 81–89, 2018.

[9] A. Martinez and R. Benavente, "The ar face database: Cvc technical report, 24," 1998.

[10] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[11] Q. Cao, L. Shen, W. Xie, O. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *FG 2018*. IEEE, 2018, pp. 67–74.

[12] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-stage dense face localisation in the wild," *arXiv preprint arXiv:1905.00641*, 2019.

[13] J. Xiang and G. Zhu, "Joint face detection and facial expression recognition with mtcnn," in *2017 4th ICISCE*. IEEE, 2017, pp. 424–427.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[15] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019, pp. 4690–4699.

[16] S. Chen, Y. Liu, X. Gao, and Z. Han, "Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices," in *CCBR*. Springer, 2018, pp. 428–438.
[17] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *ECCV*. Springer, 2016, pp. 87–102.
[18] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," in *CVPR*, 2018, pp. 5265–5274.
[19] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *CVPR*, 2017, pp. 212–220.
[20] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, Oct 2016. [Online]. Available: http://dx.doi.org/10.1109/LSP.2016.2603342
[21] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," in *https://arxiv.org/abs/2008.11104*, 2020.
[22] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *IJCB*, vol. 128, no. 2, pp. 336–359, Oct 2019. [Online]. Available: http://dx.doi.org/10.1007/s11263-019-01228-7