BWCFace: Open-set Face Recognition using Body-worn Camera

Ali Almadan, Anoop Krishnan, Ajita Rattani
Dept. of Electrical Eng. and Computer Science
Wichita State University, Wichita, USA
aaalmadan@shockers.wichita.edu; ajita.rattani@wichita.edu

Abstract—With computer vision reaching an inflection point in the past decade, face recognition technology has become pervasive in policing, intelligence gathering, and consumer applications. Recently, face recognition technology has been deployed on body-worn cameras to keep officers safe, enabling situational awareness and providing evidence for trial. However, limited academic research has been conducted on this topic using traditional techniques on datasets with small sample size. This paper aims to bridge the gap in the state-of-the-art face recognition using body-worn cameras (BWC). To this aim, the contribution of this work is two-fold: (1) collection of a dataset called BWCFace consisting of a total of 178K facial images of 132 subjects captured using the body-worn camera in in-door and daylight conditions, and (2) open-set evaluation of the latest deep-learning-based Convolutional Neural Network (CNN) architectures combined with five different loss functions for face identification, on the collected dataset. Experimental results on our BWCFace dataset suggest a maximum of 33.89% Rank-1 accuracy obtained when facial features are extracted using SENet-50 trained on a large scale VGGFace2 facial image dataset. However, performance improved up to a maximum of 99.00% Rank-1 accuracy when pretrained CNN models are fine-tuned on a subset of identities in our BWCFace dataset. Equivalent performances were obtained across body-worn camera sensor models used in existing face datasets. The collected BWCFace dataset and the pretrained/fine-tuned algorithms are publicly available to promote further research and development in this area. A downloadable link of this dataset and the algorithms is available by contacting the authors.

Index Terms—Face Recognition, Body-worn Camera, Deep Learning, Person Identification.

I. INTRODUCTION

A facial recognition is a technology capable of verifying or identifying a person from a digital image or a video frame from a video source. A typical face recognition pipeline consists of face image acquisition, face detection, facial image representation and matching [1], [21]. Face recognition can be categorized as face verification and identification. The former determines whether a pair of face images belong to the same identity, while the latter classifies a face to a specific identity. This is done by comparing the probe (test image) to the gallery (template) set of all the identities in the dataset. Face recognition is a widely adopted technology in surveillance, border control, healthcare, banking services, and lately, in mobile user authentication with Apple introducing Face ID moniker with iPhone X [1], [3].

Over the course of time, various facial image representation methods have been proposed ranging from holistic, such as Principal Component Analysis (PCA), to local image descriptors, such as Local Binary Patterns (LBP) and Scale Invariant Feature Transform (SIFT) [20], [21], [32]. Recent years have witnessed the great success of convolutional neural networks (CNNs) in face recognition [29]. Owing to advanced network architectures of CNNs and discriminative learning approaches with the introduction of the loss functions, face recognition performance has been boosted to an unprecedented level. Loss functions such as ArcFace [14], CosFace [15], SphereFace [19], and AdaCos [16] have been introduced to learn better feature representation with small intra-class and large inter-class distance for enhancement of the overall accuracy.

Recent interest has been drawn on deploying face recognition technology on body-worn cameras (BWC) by law enforcement practitioners in order to keep officers safe, enabling situational awareness and providing evidence for trial [24]–[26]. A 2012 report by the National Institute of Justice, within the U.S. Department of Justice, defines body-worn cameras as “mobile audio and video capture devices that allow officers to record what they see and hear. Devices can be attached to various body areas, including the head or to the body by pocket, and they have the capability to record officer interactions that previously could only be captured by in-car or interrogation room camera systems” [26]. The widespread
The statistical validation of the reported results could not be performed. In order to be more relevant to the real-world applications at scale, the system needs to perform well in an open-set evaluation where identities in the test set are disjoint from those in the training set. Figure 1 illustrates the schema of the deep-learning-based open-set face recognition where the deep features are extracted from a pretrained CNN model. The extracted features from the gallery and the test image are compared using a distance/similarity metrics for the final identity assignment.

To further advance the research and development, the contribution of this work is as follows:

- A publicly available BWCFace dataset consisting of a total of 178K facial images from 132 subjects captured using body-worn cameras in indoor and daylight conditions, and
- Open-set evaluation of the latest deep-learning-based face recognition algorithms[2] based on ResNet[8] and SNENet[23] architectures along with five different loss functions, i.e., Softmax, ArcFace[11], CosFace[13], SphereFace[19], and AdaCos[16] for face identification on our collected dataset. Rank-1 to Rank-10 identification accuracy values are used for performance evaluation.

The collected dataset and the evaluated algorithms serve as a baseline for further research and development. The dataset and the pretrained CNN models are available by contacting the authors.

This paper is organized as follows: in section II literature review of the existing methods on body-worn camera is performed. Section III discusses the complete data collection process. Section IV discusses the models used and the experimental protocol. Section V discusses the experiments performed and the reported results.

II. Prior Work

One of the earliest studies on face recognition using body-worn cameras can be dated back to 2011 by Al-Obaydy and Sellahewa[28]. Authors collected videos from 20 subjects using an iOPTEC-P300 body-worn digital video camera and have evaluated traditional subspace methods, namely, PCA, LDA, and discrete wavelet transforms, for facial image representation. The cosine similarity metric was used to compute scores between feature vectors from the pair of facial images for the final identity assignment. Rank-1 recognition accuracy values in the range [65.83%, 76.04%] was obtained. This dataset is referred to as University of Buckingham High and Standard Definition (UBHSD) dataset.

Jason et al.[6] and Suss et al.[7] discussed that the design issues such as video quality, camera-mounting positions, camera activation, and data transfer, are likely to enhance adoption, usability, and acceptance of BWC. The authors in[7] also laid down future research directions such as biometric identification, live streaming of videos, and training of the cops.

Brown and Fan[22] evaluated three face detectors: Viola & Jones, Aggregate Channel Feature, and Faster R-CNN over a collected dataset containing 638 face images captured from BWC. R-CNN approach achieved the highest accuracy of 94% over other face detection methods. The challenges included detecting blurry, dark, and occluded faces.

Bryan[4] studied the impact of the location of the body-worn camera on the facial detection performance. The reported results suggest that the most ideal and preferred location for camera mounting is at a center chest.

In our prior work[33], deep CNN models, i.e., VGG and ResNet, were evaluated for face identification on an in-house dataset consisting of 136k frames from 102 subjects. In this study, closed-set evaluation was adopted. Overall, the mean performance Rank-1 accuracy of 99.5% was obtained for all

| Reference | Dataset Size | Method |
|-----------|--------------|--------|
| Al-Obaydy and Sellahewa[28] (UBHSD dataset) | 20 subjects x 96 images | PCA, LDA and Discrete Wavelet |
| Brown and Fan[22] | 638 face images | Viola & Jones, Agg. Feature, Faster R-CNN |
| Al-Obaydy[28] | 3,600 face images | Neurotechnology SDKs |
| Al-Obaydy et al.[33] | 102 subjects x 136K images | CNN-based face identification (Closed-set evaluation) |
| PROPOSED | 132 subjects x 178K (BWCFace dataset) | CNN-based face identification (Open-set evaluation) |

TABLE I Summary of the existing facial analysis studies using body-worn camera.

Footnotes:
1Face recognition and identification are used interchangeably in this paper.
2Algorithms and models are used interchangeably.
Fig. (2) Face detection on the video frame of a randomly chosen subject from BWCFace dataset using Dlib library [5].

the models in the same lighting conditions, but the accuracy dropped to 98.26% in the cross-lighting conditions.

Table I summarizes the existing studies on facial analysis using a body-worn camera. As can be seen, mostly traditional methods and small size datasets are used in existing studies.

III. BWCFace Data Collection

Students from Wichita State University were recruited for the dataset collection using the Body-worn Camera (BWC). Videos from 132 subjects were collected from Boblov 1296P WiFi Body Mounted Camera [3]. The subjects were recorded using a 30s high definition video of resolution 1920 × 1080 pixels at 30fps in in-door and daylight settings (10 to 15 minutes apart). The body-worn camera was mounted on the center chest of the Research Assistant following the recommendation in [4]. During recording, subjects were asked to act naturally in order to represent realistic conditions. The actions captured from the subjects include talking, facial expressions, and head movements. The recording was done in an uncontrolled environment with non-uniform background and varying distance between the subject and the body-worn camera. However, the distance remained reasonable, subjects collaborated with BWC carrier to represent an active surveillance scenario. Ethnic groups of the captured subjects include Asian, White American, Black, and Middle Eastern.

Video clips were processed on a frame-by-frame basis for face detection using Dlib library [5]. The Dlib face detector is based on Histogram of Oriented Gradients (HOG) and Support Vector Machines (SVM). False positives in the form of non-face images were manually removed. This results in the collection of a total of 178K images from 132 subjects (varying size facial crops) equally divided between in-door and daylight conditions. Figure 2 shows the face region cropped from one of the video frames of the captured subject using Dlib library. Figure 3 shows the subset of cropped face images in the dataset from randomly selected subjects. The intra-class variations such as facial expressions, pose, and motion blur are clearly visible in the images.

Fig. (3) Subset of cropped face images from randomly selected subjects in the BWCFace dataset. Facial expression, motion blur, and poses are clearly visible in the samples.

IV. Implementation Details

In this section, we discuss the CNN architectures used, along with several loss functions and the experimental protocol. All the experiments are conducted on a workstation configured with AMD Ryzen 5 processor and GeForce RTX 2060 GPU. PyTorch [4] framework was used for CNN model implementation and evaluation.

A. CNN Architectures

We used the popular ResNet [8] and SE-ResNet [23] architectures as they are widely used for face recognition. These two backbone architectures are described below.

- **ResNet-50**: ResNet is a short form of residual network based on the idea of “identity shortcut connection” where input features may skip certain layers [8]. In this study, we used ResNet-50 which has 23.5M parameters.

- **SE-ResNet-50**: The added block SE, the Squeeze-and-Excitation, utilizes global information in selecting representative features. In other words, the block selectively adds weights to the feature maps that are representative [23].

Briefly, ResNet-50 [8] and SE-ResNet-50 [23] (SENet) are trained from scratch on VGGFace2 [30] and MS1Mv2 [31] along with five different loss functions i.e., CosFace [13], ArcFace [11], SphereFace [12], AdaCos [16], and the default Softmax [11]. These cosine loss functions (ArcFace, CosFace, SphereFace, and AdaCos) aim to maximise face class separability. Readers are referred to the cited papers on more information on these loss functions.

B. Network Training

The CNN models are trained from scratch on MS1Mv2 [11] and VGGFace2 [30] datasets, separately. We have used the

3https://www.boblov.com/

4https://pytorch.org/
curated version of these datasets for training purposes. For the purpose of experiments, all the images were resized to 224 × 224. VGGFace2 [30] dataset contains 3.31 million images from 9,131 celebrities. For training the models, we use a training split of 3.14 million images from 8,631 subjects. The MS1Mv2 dataset [11], is a cleaned version of the MS1M dataset [31], containing around 5.8 million images from 85,742 subjects were used for training the models.

For ResNet-50 model based on cosine loss functions (ArcFace, CosFace, SphereFace, and AdaCos), Batch-normalization, Drop-out, and fully connected layers of 2048 and 512 were added after the last convolutional layer. This was followed by the last fully connected layer of 512 and the final output layer equal to the number of classes for the network training. The angular margin for ArcFace, CosFace, and SphereFace was set to 0.50, 0.40 and 4.0 following the default setting reported in their respective papers. AdaCos automatically adjusts the scale parameter. The networks were trained using an Adam optimizer [10] with a batch size of 128 for 25 epochs using an early stopping mechanism on the validation set. The learning rate was set equal to 1e-4 and a decay of 5e-4.

Table II shows the validation accuracy obtained on training the network on two datasets along with five different loss functions. The validation accuracy in the range of [94.63%, 99.85%] was obtained. Mostly, AdaCos and Softmax obtained equivalent validation accuracy values.

### Table (II) Validation accuracy for CNN models trained on MS1Mv2 [11] and VGGFace2 [30] datasets.

| CNN        | Trained | Loss Function | Accuracy [%] |
|------------|---------|---------------|--------------|
| ResNet-50  | MS1Mv2  | CosFace       | 97.88        |
|            |         | ArcFace       | 97.67        |
|            |         | SphereFace    | 99.04        |
|            |         | AdaCos        | 99.75        |
|            |         | Softmax       | 99.58        |
| ResNet-50  | VGGFace2| CosFace       | 97.83        |
|            |         | ArcFace       | 97.58        |
|            |         | SphereFace    | 99.13        |
|            |         | AdaCos        | 99.58        |
|            |         | Softmax       | **99.63**    |
| SENet-50   | MS1Mv2  | CosFace       | 95.29        |
|            |         | ArcFace       | 94.63        |
|            |         | SphereFace    | 97.54        |
|            |         | AdaCos        | 98.92        |
|            |         | Softmax       | **99.21**    |
| SENet-50   | VGGFace2| CosFace       | 96.96        |
|            |         | ArcFace       | 97.08        |
|            |         | SphereFace    | 98.67        |
|            |         | AdaCos        | 99.50        |
|            |         | Softmax       | **99.85**    |

### C. Experimental Protocol

For the open-set evaluation of the trained deep learning models for face identification on BWCFace dataset, the gallery set consists of 12 face images per subject. The probe set consists of 100 face images randomly selected per subject.

From the gallery and probe set of all the subjects, deep features are extracted using the pretrained ResNet-50 and SENet-50 models (discussed in Section IV). The matching score between the deep features from the pair of gallery and probe image is computed using cosine similarity, as given below:

\[
\text{sim}(u, v) = \frac{u \cdot v}{||u|| ||v||} = \frac{\sum_{i=1}^{N} u_i v_i}{\sqrt{\sum_{i=1}^{N} u_i^2} \sqrt{\sum_{i=1}^{N} v_i^2}}
\]

where \( u \) and \( v \) are the two deep feature vectors: \( u = \{u_1, u_2, \ldots, u_N\} \) and \( v = \{v_1, v_2, \ldots, v_N\} \).

The average of the scores from the multiple gallery images per subject is used for final identity assignment. In cross-lighting conditions, the template (gallery set) was selected from a different lighting condition that of probe set for all the identities.

### V. Evaluation Results

In this section, we evaluate the open-set performance of the face recognition on our BWCFace dataset collected using a body-worn camera. All the models are evaluated in the same and cross-lighting conditions. Next, we discuss the experiments performed and the results obtained.

1. **Pre-trained CNNs:** First, we evaluated the open-set performance of the face recognition system on BWCFace dataset when deep features are extracted from ResNet-50 and SENet-50, trained on MS1Mv2 [11] and VGGFace2 [30] datasets, in the same and cross-lighting conditions.

Table III shows the open-set evaluation of these pretrained face recognition models on BWCFace dataset. Rank-1, Rank-5, and Rank-10 accuracy values are recorded. These results...
are shown only for the Softmax loss function. Equivalent performances were obtained for other loss functions, not shown for the sake of space. Cross-lighting scenario is shown using Day vs. Office entry and vice-versa. Day vs. Office entry means that gallery set is acquired in daylight and probe set in office light conditions.

It can be seen that overall low accuracy values are obtained. SENet-50 has slightly better performance than ResNet-50. Day vs. Day obtained better accuracy values over Office vs. Office. Maximum accuracy of 33.89%, 61.69%, and 75.06% are consistently obtained at Rank-1, Rank-5, and Rank-10 for SENet-50 when trained on VGGFace2 dataset. When compared with the same lightning settings, cross-lighting evaluation degraded the accuracy values by a factor of 5.18, 2.35, and 1.75 at Rank-1, Rank-5, and Rank-10, respectively. We have also reported an open-set evaluation of the pretrained models on the UBHSD dataset [28]. Equivalent accuracy values were obtained when compared to the same lighting condition results on our BWCFace dataset. UBHSD dataset did not annotate the images based on the lighting condition. Therefore, cross-lighting comparisons could not be performed.

Overall, low accuracy values were obtained when pretrained CNN models are used for face recognition on our BWCFace and UBHSD dataset acquired using the body-worn camera. This could be due to existing face image datasets scraped from the web are not representative of the images captured using the body-worn camera.

2. Pretrained CNNs Fine-tuned on BWCFace Dataset: Table IV shows the performance of the ResNet-50 model, pretrained on VGGFace2 [30] and MS1Mv2 [11], when fine-tuned on subset of 30 subjects from our BWCFace dataset. It can be seen that the validation accuracy of 99% was obtained by SphereFace, AdaCos, and Softmax. CosFace and ArcFace obtained around 97% validation accuracy.

Open set evaluation obtained a maximum of 99% Rank-1 for Softmax loss function in Day vs. Day settings. At Rank-5, all the loss functions obtained around 99% accuracy for Day vs. Day settings. At rank-10, all the loss functions hit 99% accuracy in the same lighting conditions.

By comparing with the mean accuracy over all lighting conditions, the average drop in the performance for cross-lighting conditions was 8.97%, 2.76%, and 1.31% at Rank-1, Rank-5, and Rank-10, respectively. Overall, the drop in the performance was higher for Day vs. Office in comparison to Office vs. Day condition.

Table IV shows the performance of the SENet-50 when trained on MS1Mv2 [11] and VGGFace2 [30] and fine-tuned on 30 subjects from our BWCFace dataset. It can be seen that the validation accuracy of 99% was obtained by AdaCos and Softmax. CosFace and ArcFace obtained around 97% validation accuracy. Open set evaluation obtained a maximum of 99% Rank-1 for Softmax loss function in Day vs. Day settings, when pre-trained weights are obtained from the MS1Mv2 dataset. At Rank-5, most of the models obtained around 99% accuracy in day vs. day setting. At Rank-10, most of the models obtained 99% accuracy for the same lighting condition. The average drop in the performance for cross-lighting conditions was 7.23%, 2.02% and 0.80% at Rank-1, Rank-5, and Rank-10, respectively. The drop in the performance was higher for Day vs. Office in comparison to Office vs. Day condition.

The comparison in terms of architectures showed that SENet-50 obtained slightly better identification performance compared to ResNet-50 for both the datasets. In fact, the average Rank-1 accuracy for ResNet-50 is 88.55%, and 89.09% for SENet-50. This can be due to the significant influence of the global information that the SE block adds. Overall, fine-tuning the models on a small subset of subjects captured from body-worn camera has obtained hallmark performance.

3. Fine-tuned Models Evaluated on UBHSD Dataset [28]:
We also evaluated open-set face recognition performance of the models fine-tuned on our BWCFace dataset and tested on the UBHSD dataset [28]. The UBHSD dataset contained 96 face images from each of the 20 distinct subjects. This dataset was collected using an iOPTEC-30 BWC camera sensor which is different from our sensor. The dataset comprised of blurred and occluded face images with eyes half-closed. For this particular dataset, templates (gallery set) consist of 12 face images and probe set consists of 84 images. The deep features from the pair of template and probe images are extracted using the fine-tuned models. As this dataset is not annotated with the lighting condition, cross-lighting evaluation could not be performed. Table IV shows the Rank-1, Rank-5 and Rank-10 accuracy values of the CNN models fine-tuned on our BWCFace dataset and tested on UBHSD dataset. It can be seen that maximum Rank-1 accuracy of around 98% (1% performance drop) for most of the loss functions. At Rank-5 and Rank-10, most of the loss functions obtained 99% accuracy. At Rank-1 and Rank-5, ResNet-50 obtained better performance than SENet-50. ResNet-50 obtained a mean of

| CNN           | Trained      | Metrics   | Rank-1 [%] | Rank-5 [%] | Rank-10 [%] |
|---------------|--------------|-----------|------------|------------|-------------|
| ResNet-50     | MS1Mv2       | CosFace   | 98.21      | 99.52      | 99.94       |
|               |              | ArcFace   | 97.92      | 99.35      | 99.88       |
|               |              | SphereFace| 98.51      | 99.70      | 99.94       |
|               |              | AdaCos    | 97.26      | 99.40      | 99.82       |
|               |              | Softmax   | 98.99      | 99.88      | 100.00      |
| ResNet-50     | VGGFace2     | CosFace   | 98.39      | 99.70      | 99.94       |
|               |              | ArcFace   | 98.15      | 99.64      | 99.82       |
|               |              | SphereFace| 98.27      | 99.64      | 99.76       |
|               |              | AdaCos    | 97.92      | 99.52      | 99.64       |
|               |              | Softmax   | 98.75      | 99.64      | 99.76       |
| SENet-50      | MS1Mv2       | CosFace   | 98.01      | 99.21      | 99.11       |
|               |              | ArcFace   | 95.60      | 98.33      | 98.93       |
|               |              | SphereFace| 96.19      | 98.39      | 98.87       |
|               |              | AdaCos    | 95.83      | 98.51      | 99.40       |
|               |              | Softmax   | 95.71      | 98.69      | 99.82       |
| SENet-50      | VGGFace2     | CosFace   | 97.02      | 98.57      | 99.35       |
|               |              | ArcFace   | 97.08      | 98.63      | 99.52       |
|               |              | SphereFace| 97.44      | 98.57      | 99.40       |
|               |              | AdaCos    | 96.55      | 98.81      | 99.40       |
|               |              | Softmax   | 97.98      | 99.52      | 99.82       |
98.24% and 99.60% for Rank-1 and Rank-5, respectively, whereas SENet-50 obtained a mean of 96.55% for Rank-1 and 98.63% for Rank-5, respectively.

It was observed that the performance for all models and loss functions was comparable with minimum drop, despite that they were fine-tuned on our dataset (see Table IV). This also suggests that equivalent performances of fine-tuned models could be obtained across different body-worn camera models.

VI. CONCLUSION

This is the first study evaluating open-set face recognition on our BWCFace data captured using the body-worn camera. To this aim, the latest deep learning models with different loss functions are evaluated on our dataset. Experimental results suggest a maximum of 33.89% Rank-1 accuracy obtained. The reason could be datasets such as VGGFace2 and MS1MV2 scraped from the web do not contain images captured from Body-worn cameras. Therefore, these datasets may not be representative of the samples acquired using a body-worn camera. However, on fine-tuning the models using a small subset of subjects captured using the body-worn camera, hallmark Rank-1 accuracy of 99% was obtained. Further, higher accuracy rates were also obtained on data samples captured from a different model of the body-worn camera from the UBHSD dataset [28]. No significant differences were observed across the loss functions. As a part of future work, lightweight CNN models such as lightCNN [34] will be evaluated. Cross-sensor evaluation will be performed using data captured by mobile devices [3]. Note that the study on the bias of the face recognition technology across demographic variations is a different research topic that is beyond the scope of this current study.

ACKNOWLEDGMENT

This work is supported in part by Award for Research/Creative projects by Wichita State University.

REFERENCES

[1] Anil K Jain and Stan Z Li. Handbook of face recognition, volume 1. Springer, 2011.
[2] Timo Ahonen, Abdounor Hadid, and Matti Pietikainen. Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12):2037–2041, December 2006.
[3] Ajita Rattani, Reza Derakhshani and Arun Ross. Selfie Biometrics. In Springer, 2019.
[4] Julia Bryan. Effects of Movement on Biometric Facial Recognition in Body-Worn Cameras. PhD thesis, Purdue University Graduate School, 2020.
[5] Davis E King. Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10(Jul):1755–1758, 2009.
[6] Jason J Corso, Alexandre Alahi, Kristen Grauman, Gregory D Hager, Louis-Philippe Morency, Harpreet Sawhney, and Yaser Sheikh. Video analysis for body-worn cameras in law enforcement. arXiv preprint arXiv:1604.03130, 2016.
[7] Joel Suss, Alexis Raushel, Adam Armanjo, and Brian White. Design considerations in the proliferation of police body-worn cameras. ergonomics in design, 26(3):17–22, 2018.
[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. In arXiv, 2015.
[9] S. Liu and W. Deng. Very deep convolutional neural network based image classification using small training sample size. In 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.
[10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[11] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4685–4694, Long Beach, CA, USA, June 2019. IEEE.
[12] Wei yang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. SphereFace: Deep Hypersphere Embedding for Face Recognition. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 6738–6746, Honolulu, HI, July 2017. IEEE.
[13] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Zifeng Li, Dihong Gong, Jingchao Zhou, Wei Liu CosFace: Large Margin Cosine Loss for Deep Face Recognition. In CoRR, 2018.
[14] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 4685–4694, Long Beach, CA, USA, June 2019. IEEE.
[15] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face Description could be obtained across different body-worn camera models. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 6738–6746, Honolulu, HI, July 2017. IEEE.
| Loss Function | Validation Accuracy [%] | Light. Condition | Rank-1 [%] | Rank-5 [%] | Rank-10 [%] |
|---------------|--------------------------|------------------|------------|------------|-------------|
| **ResNet-50 - MS1Mv2** | | | | | |
| CosFace | 97.88 | Office vs. Office | 94.19 | 98.33 | 99.20 |
| | | Day vs. Day | 98.08 | 99.46 | 99.63 |
| | | Day vs. Office | 76.45 | 90.41 | 94.27 |
| | | Office vs. Day | 83.83 | 93.98 | 96.89 |
| ArcFace | 97.66 | Office vs. Office | 93.41 | 98.11 | 99.04 |
| | | Day vs. Day | 97.83 | 99.38 | 99.68 |
| | | Day vs. Office | 75.22 | 90.27 | 94.25 |
| | | Office vs. Day | 82.55 | 93.15 | 96.43 |
| SphereFace | 99.04 | Office vs. Office | 94.14 | 98.28 | 99.17 |
| | | Day vs. Day | 98.34 | 99.44 | 99.65 |
| | | Day vs. Office | 77.04 | 90.78 | 94.67 |
| | | Office vs. Day | 84.93 | 94.58 | 97.26 |
| AdaCos | 99.75 | Office vs. Office | 89.71 | 97.22 | 98.66 |
| | | Day vs. Day | 95.55 | 99.03 | 99.57 |
| | | Day vs. Office | 69.27 | 88.67 | 93.80 |
| | | Office vs. Day | 77.73 | 90.84 | 94.56 |
| Sofmax | 99.58 | Office vs. Office | 96.36 | 98.42 | 99.34 |
| | | Day vs. Day | 99.08 | 99.83 | 99.85 |
| | | Day vs. Office | 85.02 | 96.20 | 98.31 |
| | | Office vs. Day | 95.19 | 98.97 | 99.32 |
| **ResNet-50 - VGGFace2** | | | | | |
| CosFace | 97.83 | Office vs. Office | 94.25 | 98.32 | 99.02 |
| | | Day vs. Day | 98.82 | 99.79 | 99.83 |
| | | Day vs. Office | 76.39 | 93.57 | 96.96 |
| | | Office vs. Day | 83.40 | 94.78 | 98.39 |
| ArcFace | 97.58 | Office vs. Office | 94.16 | 98.37 | 99.07 |
| | | Day vs. Day | 98.73 | 99.74 | 99.87 |
| | | Day vs. Office | 75.49 | 93.39 | 97.10 |
| | | Office vs. Day | 83.61 | 95.63 | 98.78 |
| SphereFace | 99.13 | Office vs. Office | 94.25 | 98.39 | 99.07 |
| | | Day vs. Day | 98.84 | 99.72 | 99.89 |
| | | Day vs. Office | 77.92 | 93.96 | 97.33 |
| | | Office vs. Day | 84.16 | 96.02 | 98.72 |
| AdaCos | 99.58 | Office vs. Office | 92.43 | 97.84 | 98.78 |
| | | Day vs. Day | 98.26 | 99.78 | 99.85 |
| | | Day vs. Office | 72.84 | 92.12 | 96.49 |
| | | Office vs. Day | 80.12 | 94.84 | 97.98 |
| Sofmax | 99.63 | Office vs. Office | 96.33 | 98.47 | 99.00 |
| | | Day vs. Day | 99.36 | 99.81 | 99.85 |
| | | Day vs. Office | 84.25 | 95.61 | 98.06 |
| | | Office vs. Day | 94.45 | 98.99 | 99.63 |
## TABLE (VI)  
Rank-1, Rank-5 and Rank-10 accuracy values of the SENet-50 trained on MS1Mv2 [11] and VGGFace2 [30] datasets, fine-tuned and evaluated on non-overlapping subjects from BWCFace dataset.

| Loss Function | Validation Accuracy [%] | Light. Condition | SENet-50 - MS1Mv2 | SENet-50 - VGGFace2 |
|---------------|--------------------------|------------------|-------------------|-------------------|
| CosFace       | 95.29                    | Office vs. Office| Rank-1 [%] | Rank-5 [%] | Rank-10 [%] | Office vs. Office| Rank-1 [%] | Rank-5 [%] | Rank-10 [%] |
|               |                          | Day. vs. Day     | 93.431          | 97.902           | 98.755          | Day. vs. Day     | 94.019          | 97.294           | 98.196          |
|               |                          | Day vs. Office   | 97.665          | 99.570           | 99.821          | Day vs. Office   | 97.553          | 99.477           | 99.776          |
|               |                          | Office vs. Day   | 78.549          | 93.373           | 96.804          | Office vs. Day   | 85.867          | 96.331           | 98.524          |
| ArcFace       | 94.63                    | Office vs. Office| 92.754          | 97.588           | 98.598          | Office vs. Office| 96.088          | 98.294           | 99.196          |
|               |                          | Day. vs. Day     | 97.422          | 99.514           | 99.813          | Day. vs. Day     | 97.553          | 99.477           | 99.776          |
|               |                          | Day vs. Office   | 77.627          | 93.510           | 97.137          | Day vs. Office   | 85.867          | 96.331           | 98.544          |
|               |                          | Office vs. Day   | 84.586          | 94.448           | 97.554          | Office vs. Day   | 83.479          | 96.059           | 98.466          |
| SphereFace    | 97.54                    | Office vs. Office| 93.362          | 97.990           | 98.774          | Office vs. Office| 95.000          | 98.127           | 98.755          |
|               |                          | Day. vs. Day     | 97.646          | 99.514           | 99.795          | Day. vs. Day     | 99.197          | 99.738           | 99.813          |
|               |                          | Day vs. Office   | 78.529          | 93.490           | 97.000          | Day vs. Office   | 87.392          | 96.059           | 98.039          |
|               |                          | Office vs. Day   | 84.974          | 94.040           | 97.204          | Office vs. Day   | 94.545          | 98.796           | 99.340          |
| AdaCos        | 98.92                    | Office vs. Office| 90.666          | 97.039           | 98.157          | Office vs. Office| 91.107          | 98.294           | 99.196          |
|               |                          | Day. vs. Day     | 96.507          | 99.477           | 99.720          | Day. vs. Day     | 97.553          | 99.477           | 99.776          |
|               |                          | Day vs. Office   | 73.333          | 91.549           | 96.000          | Day vs. Office   | 85.673          | 96.855           | 98.758          |
|               |                          | Office vs. Day   | 81.208          | 92.972           | 96.252          | Office vs. Day   | 83.479          | 96.059           | 98.466          |
| Sofmax        | 99.21                    | Office vs. Office| 93.000          | 98.127           | 98.755          | Office vs. Office| 94.545          | 98.796           | 99.340          |
|               |                          | Day. vs. Day     | 99.197          | 99.738           | 99.813          | Day vs. Office   | 87.392          | 96.059           | 98.039          |
|               |                          | Day vs. Office   | 95.000          | 98.127           | 98.755          | Office vs. Day   | 94.545          | 98.796           | 99.340          |
|               |                          | Office vs. Day   | 94.545          | 98.796           | 99.340          | Office vs. Day   | 91.322          | 98.156           | 99.495          |