Data Centric Face Recognition for African Face Authentication

Davy Uwizera, William Bares, Catalin Voss

{davy,bbares,catalin}@smileidentity.com

Abstract

We present the SmileID face recognition system, a commercial system for frontal-face identity verification on mobile handsets in Africa. Our work is a case study in building and deploying a real-world face recognition system that must work primarily on non-caucasian faces. Unlike commercial systems that aim to reduce bias by minimizing accuracy disparity between light-skinned and dark-skinned faces in many lighting conditions and poses, our system focuses specifically on frontal face smartphone-based authentication of dark-skinned, African faces. While much research work has focused on improving model structures and loss functions to reduce modal bias, we show that a data-centric approach – training a state of the art network on African faces – yields strong results. We observe gaps between the accuracy numbers on dark-skinned faces reported by commercial "multi-purpose" systems like AWS Rekognition and their real-world performance once we add the constraints that the images come from low-power handsets as selfies in frontal-only poses. Our model outperforms Rekognition on a benchmark dataset for frontal authentication and achieves an 11% *gain over a baseline ArcFace implementation in this setting by training on an African dataset. On the other hand, it also improves homogeneity by 16% and completeness by 21%.*

1. Introduction

Over the past few years, face recognition models based on Deep Convolutional Neural Networks (CNNs) have closed the gap to human-level performance on academic datasets and enabled a wealth of applications. However, these models have largely been designed and tested with data collected from European and American and predominantly caucasian faces [5]. As a result, many models exhibit significant accuracy disparities across intersections of groups split out by ethnicity, gender, and skin tone [2] [4] [10] [17] [9]. What's more, much research effort has gone into optimizing these models to perform well across a variety of lighting conditions and poses that match the applications of interest of a majority-western customer base. Recently, commercial face recognition providers such as AWS Rekognition, Microsoft, and Google have come under significant criticism for releasing biased models and for enabling surveillance applications. Their response has been to acquire more balanced training data and assess their models for bias, often defined as the accuracy disparity between subgroups of the data, such as dark-skinned women and light-skinned men [13]. However, their goal has remained building robust general-purpose face recognition.

We present a face recognition system targeted specifically at non-surveilance, frontal face authentication to empower smartphone-based authentication in Africa for use cases like banking, lending, or ride sharing. Existing datasets do a poor job of assessing face recognition systems for this use case and that most of existing facial recognition commercial systems, including AWS Rekognition, perform below recently published benchmarks on dark-skinned faces [23] [1].

Our model is a facial embedding model based on the ArcFace loss function [3] that is transfer-learned using a proprietary dataset of African faces. Like most CNN-based face recognition systems, our model follows a metric learning approach, in which the goal is to learn a geometric representation of the input faces that can be assessed with a distance metric, such as the ℓ_2-distance or cosine similarity at inference-time. The goal is to simultaneously achieve intra-class compactness [12] and interclass separability [14] of the pattern learnt by their underlying convolution neural network architecture. Intra-class compactness promotes a low standard deviation of the extracted features for one specific class and separability encourages the distributions of the classes being learnt to be distant from each other [11] [7] [25]. While separability has been the main research focus for several years, compactness has recently gained a lot of interest recently, as models which also focused on improving compactness demonstrated high robustness [22]. We show that by focusing on the right data and using transfer-learning onto African face data greatly improves compactness and outperforms current state-of-art models. In section 5, we provide an analysis of how transfer learning modifies the representations learned by our model with increasing African training data.

2. Related work

Our work follows other efforts using transfer-learning to improve recognition results for specific subgroups and de-bias face recognition models. Luttrell et al [15] show that transfer-learning can be a viable strategy for face recognition even on small datasets. Smith et al [20] demonstrate that transfer learning from face recognition models can be used to perform alternate tasks like gender classification and age estimation. Yin et al [26] use the MS-Celeb-1M dataset and combined transfer learning and data augmentation techniques to build a center-based feature transfer facial recognition framework that could capture under-represented data. Other work [18] [21] shows promise for de-biasing pre-trained models using transfer learning with balanced data. However, this approach faces significant challenges: identifying a representative datasets for each subgroup is challenging, especially if the learning method requires that the data for different subgroups come from the same distribution. Data augmentation techniques have shown promise, but haven't been accurate enough to successfully boost the numbers for under-represented classes [19]. Moreover, Wang et al showed that balanced data may not be enough, due to the fact that in some cases, the correlation between target labels and features learnt may be amplified by the models [24].

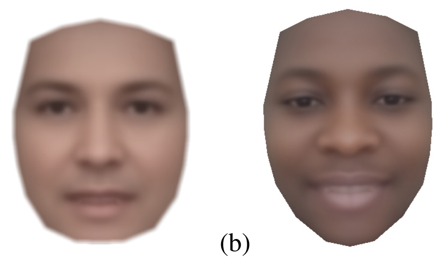
3. Approach

3.1. Data

We collected a proprietary dataset for frontal face authentication in Africa. Images were collected from individuals authenticating with a service through the SmileID mobile SDK, built into third-party apps built by SmileID's customers in Africa. For each individual, selfies were captured using the front-facing camera over a few-second interval; in addition, a single picture was captured for each re-authentication. A true value comparison was performed between the enrollment picture and the authentication picture by a trained human labeler. In this research we used a total number of 22,330 images. We reserve 30% of the dataset for evaluation purposes, splitting at the subject-level. In Figure 1 we illustrate the composition of our dataset compared to the popular Celeb-1M dataset [5] by sampling 500 aligned faces from each dataset. For further testing we also use LFW [8]. We also extract only dark faces from LFW and do random pair matching to create dark face pair dataset, LFWB, of 10817 face pairs.

Table 1. SmileID FFD Summary Statistics

Distinct Individuals	Average Faces	Median Faces
638	35	15



(a) (b)

Figure 1. Blended faces for (a) Celeb1M, our pre-training dataset, (b) the SmileID FFD dataset, our transfer-learning dataset.

3.2. Model and Loss

Our model is based on the ResNet50 architecture [6] made up of convolutional and fully-connected layers. For performance reasons and because many of the images captured on handsets in our application are of low resolution, the input size of 112x112 is used. We also consider the embedding size of 512.

We use the Additive Angular Margin ("ArcFace") [3] loss, which is effectively a softmax signal over an angular margin metric, to learn a space of face representations in which the cosine similarity between embeddings provides is a robust distance metric:

$$L = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}},$$

where $\theta_j := \theta_j(x_i)$ is the angle between a weight W_j and the last deep layer feature x_i and y_i is the 'correct' class for sample i, so θ_{y_i} is the target ground truth angle, m is an angular penalty margin, and s is a feature scale.

3.3. Pre-training

Our model is pre-trained on the MS-Celeb1M dataset [5], the most popular large open dataset used for face recognition consisting of predominantly white faces. ArcFace feature scale s and margin m are set to 64 and 0.5 respectively, momentum to 0.9 and weight decay to 5e-4. A batch size of 128 is used and trained for 180K iterations with a stochastic gradient descent optimizer.

3.4. Transfer Learning

Our model is then fine-tuned on the train split of the SmileID Dataset. We freeze all pre-trained layers except for the last 4 ResNet layers, consistent with the layer freezing approach [36]. We do this to help our model focus on the details of our task and keep the low-level features learnt during the pre-training process. The ArcFace feature scale s and margin m are set to 64 and 0.5 respectively, momentum to 0.9 and weight decay to 5e-4 just like in pre-training. We used a batch size of 16 and trained for 16 epochs.

3.5. Preprocessing

During training and inference, image values are normalized between zero and one, and for each image in the pipeline, a face is located, cropped and aligned to a 112x112 bounding box. Images with no face found during training are automatically removed from the pipeline.

3.6. Data Augmentation

Data augmentation is applied during both the pre-training and transfer-learning process. Using random choice probabilistic approach, random saturation and brightness is applied on train images with 50% probability, vertical flipping introduced with 40% probability, and with 10% probability minor gaussian and salt pepper noise is also applied. We use 0.6 for the lower bound for the random saturation factor, and 1.4 on the upper bound. On brightness, we use 0.4 for our max delta.

3.7. Evaluation

For testing we use LFW data, 30% out of sample of FDD data and LFWB data. LFW, FDD, LFWB are described in section3.1. For each model considered, we look at the overall weighted accuracy on test data. We also compute homogeneity and completeness on dark faces between the ArcFace pre-trained model and African-face transfer-learned ArcFace model, to assess compactness and separability of the generated clusters between the two models. Using TSN-e, we also visualize the cluster output of both models for 10 randomly selected individual facial clusters.

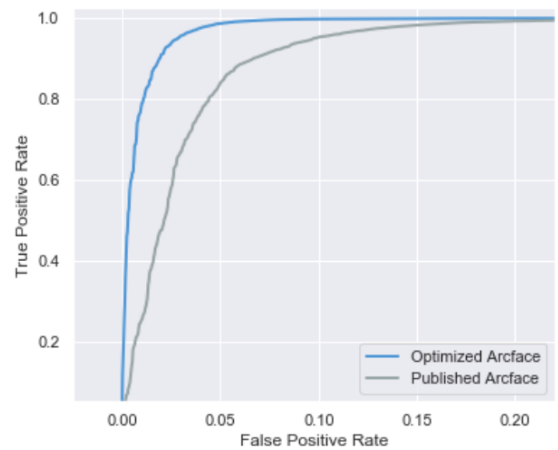


Figure 2. ROC curve on FDD

4. Results

We compare our model to AWS Rekognition and the published ArcFace implementation trained on Celeb1M only. We perform experiments using LFW, LWFB and the SmileID FDD dataset. Accuracy evaluations are done at the equal error rate threshold. LWFB is composed of random pairing of dark faces in LFW.

As shown in Table 2, we achieve an 11% gain over the baseline ArcFace implementation on African faces. An ROC curve is given in Figure 2.

Table 2. Overall model results comparison

Model	LFW	FDD	LFWB
AWS Rekognition	99.2	97.4	96
Arcface Baseline	**99.3**	88	98.1
Optmized Arcface	98.8	**98.9**	**99.8**

Table 3. Model homogeneity and completeness comparison on African faces, Arcface Baseline vs Optimized Arcface

Model	Homogeneity	Completeness
Arcface Baseline	0.798	0.740
Optmized Arcface	**0.959**	**0.963**

5. Analysis

Using TSN-e [16] visualization on classification results per cluster, we visualize the separability and compactness of the two models on both Caucasian and melanin faces. We use a sample of ten individuals with their multiple face snapshots.

Looking at Figure 3, we see that the optimized model improves over the published ArcFace baseline on African faces both on separability and compact-

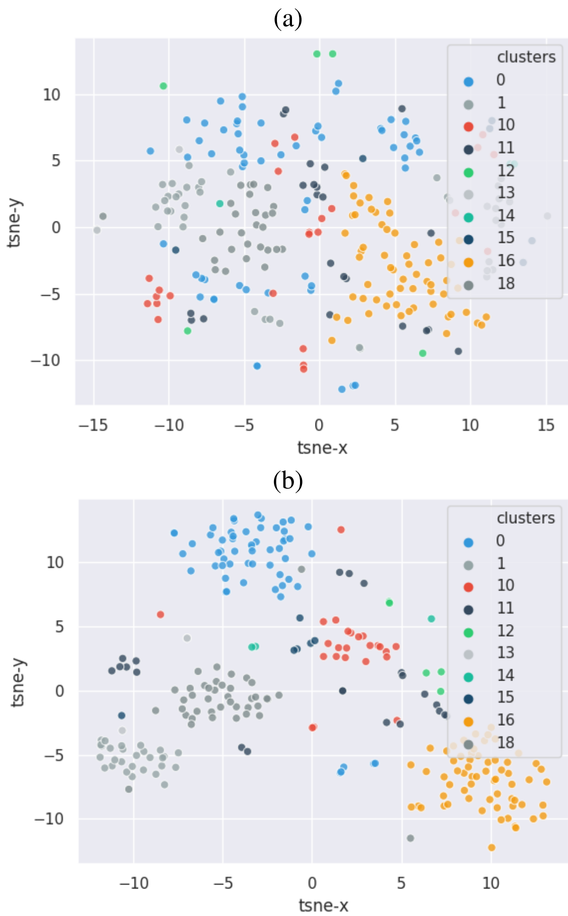


Figure 3. (a) Published ArcFace Embedding Clusters on African Faces, (b) Optimized ArcFace Embedding Clusters on African Faces

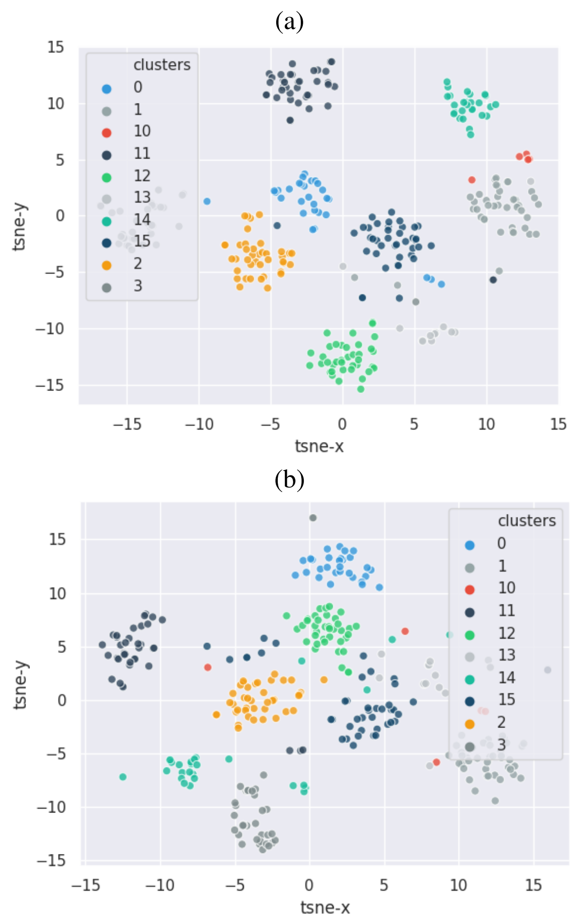


Figure 4. (a) Published ArcFace Embedding Clusters on Caucasian Faces, (b) Optimized ArcFace Embedding Clusters on Caucasian Faces

ness. The pattern in Figure 3 (a), shows that published ArcFace performs poorly on separability on African faces. However, it does fairly well on compactness with limited performance. This may imply that, trained on majority Caucasian faces, the model understands the low-level features well but fails to grasp task specific features due to the differences between African and Caucasian faces.

Using MS-Celeb-1M we also visually analyze the compactness and separability of the improved model on source data compared to the pubished ArcFace. The analysis is done on 10 randomly selected individuals' clusters.

Figure 4, shows that our approach keeps good separability and compactness on the source target after the transfer-learning process with very minimal accuracy trade off. The overall accuracy trade-off on the LFW dataset is less than one percent as shown in Table-2. We also compute homogeneity for each model on

African faces, Table-3 shows that the method used yield better results.

6. Conclusion

We present results from transfer-learning a state-of-the-art face recognition model for African face authentication. Despite the large amount of effort that has gone into debiasing commercial systems, we are able to outperform AWS Rekognition on this face recognition subtask by 1.5%. We improve by 11% over the baseline ArcFace implementation and show that transfer-learning greatly improves the learned representations for African faces in compactness and separability. We believe that frontal-face, smartphone-image-based authentication is an important benchmark task for face recognition as it enables consented authentication for several important use cases in the developing world, where other means of identity verification are often prohibitively expensive.

In future work, we plan to explore how we can improve recognition robustness for African faces using different transfer learning techniques, loss functions, and preprocessing. We also plan to analyze how error and bias in face recognition accumulate in the pipeline from cellphone image capture to model inference and describe techniques for improving accuracy on African face recognition in the capture, preprocessing, and alignment steps.

References

[1] Joy Buolamwini and Timnit Gebru. “Gender shades: Intersectional accuracy disparities in commercial gender classification”. In: *Conference on fairness, accountability and transparency*. 2018, pp. 77–91.

[2] Jacqueline G Cavazos et al. “Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?” In: *IEEE Transactions on Biometrics, Behavior, and Identity Science* (2020).

[3] Jiankang Deng et al. “Arcface: Additive angular margin loss for deep face recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 4690–4699.

[4] Clare Garvie and Jonathan Frankle. “Facial-recognition software might have a racial bias problem”. In: *The Atlantic* 7 (2016).

[5] Yandong Guo et al. “Ms-celeb-1m: A dataset and benchmark for large-scale face recognition”. In: *European conference on computer vision*. Springer. 2016, pp. 87–102.

[6] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[7] Lanqing He et al. “Softmax dissection: Towards understanding intra-and inter-clas objective for embedding learning”. In: *arXiv preprint arXiv:1908.01281* (2019).

[8] Gary B Huang et al. “Labeled faces in the wild: A database forstudying face recognition in unconstrained environments”. In: *Workshop on faces in’Real-Life’Images: detection, alignment, and recognition*. 2008.

[9] Adam Kortylewski et al. “Analyzing and reducing the damage of dataset bias to face recognition with synthetic data”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2019, pp. 0–0.

[10] Susan Leavy. “Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning”. In: *Proceedings of the 1st international workshop on gender equality in software engineering*. 2018, pp. 14–16.

[11] Xiaobin Li and Weiqiang Wang. “Learning discriminative features via weights-biased softmax loss”. In: *Pattern Recognition* 107 (2020), p. 107405.

[12] Weiyang Liu et al. “Large-margin softmax loss for convolutional neural networks.” In: *ICML*. Vol. 2. 3. 2016, p. 7.

[13] Steve Lohr. “Facial recognition is accurate, if you’re a white guy”. In: *New York Times* 9 (2018).

[14] Yan Luo et al. “G-Softmax: Improving Intraclass Compactness and Interclass Separability of Features”. In: *IEEE transactions on neural networks and learning systems* 31.2 (2019), pp. 685–699.

[15] Joseph Luttrell et al. “A deep transfer learning approach to fine-tuning facial recognition models”. In: *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE. 2018, pp. 2671–2676.

[16] Laurens van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.Nov (2008), pp. 2579–2605.

[17] Shruti Nagpal et al. "Deep learning for face recognition: Pride or prejudiced?" In: *arXiv preprint arXiv:1904.01219* (2019).

[18] Sinno Jialin Pan and Qiang Yang. "A survey on transfer learning". In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359.

[19] Connor Shorten and Taghi M Khoshgoftaar. "A survey on image data augmentation for deep learning". In: *Journal of Big Data* 6.1 (2019), p. 60.

[20] Philip Smith and Cuixian Chen. "Transfer learning with deep CNNs for gender recognition and age estimation". In: *2018 IEEE International Conference on Big Data (Big Data)*. IEEE. 2018, pp. 2564–2571.

[21] Chuanqi Tan et al. "A survey on deep transfer learning". In: *International conference on artificial neural networks*. Springer. 2018, pp. 270–279.

[22] Athanasios Voulodimos et al. "Deep learning for computer vision: A brief review". In: *Computational intelligence and neuroscience* 2018 (2018).

[23] Mei Wang et al. "Racial faces in the wild: Reducing racial bias by information maximization adaptation network". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 692–702.

[24] Tianlu Wang et al. "Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 5310–5319.

[25] Yudong Wu et al. "Rotation consistent margin loss for efficient low-bit face recognition". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 6866–6876.

[26] Xi Yin et al. "Feature transfer learning for deep face recognition with under-represented data". In: *arXiv preprint arXiv:1803.09014* (2018).