



DigiFace-1M: 1 Million Digital Face Images for Face Recognition

Winter Conference on Applications of Computer Vision 2023

Gwangbin Bae Martin de La Gorce Tadas Baltrušaitis Charlie Hewitt Dong Chen
Julien Valentin Roberto Cipolla Jingjing Shen

Paper

arXiv

Dataset



Abstract

State-of-the-art face recognition models show impressive accuracy, achieving over 99.8% on Labeled Faces in the Wild (LFW) dataset. However, these models are trained on large-scale datasets that contain millions of real human face images collected from the internet. Web-crawled face images are severely biased (in terms of race, lighting, make-up, etc) and often contain labeling noise. Most importantly, these face images are collected without explicit consent, raising more pressing privacy and ethical concerns. To avoid the problems associated with real face datasets, we introduce a large-scale synthetic dataset for face recognition, obtained by photo-realistic rendering of diverse and high-quality digital faces using a computer graphics pipeline. We compare our method to SynFace, a recent method trained on GAN-generated synthetic faces, and reduce the error rate on LFW by 52.5% (accuracy from 91.93% to 96.17%). We first demonstrate that aggressive data augmentation can significantly help

reduce the domain-gap between our synthetic faces and real face images. Taking advantage of having full control over the rendering pipeline, we also study how each attribute (e.g., variation in facial pose, accessories, and textures) affects the accuracy. Finally, by fine-tuning the network on a smaller number of real face images that could reasonably be obtained with consent, we achieve accuracy that is comparable to the methods trained on millions of real face images, while alleviating the problems associated with large datasets.

Motivation

State-of-the-art face recognition models are trained on millions of real human face images collected from the internet. **DigiFace-1M** aims to tackle three major problems associated with such large-scale face recognition datasets.

- **Ethical issues** - Many existing datasets are obtained by collecting web images without explicit consent. Our digital faces are created using a generative model built from high quality head scans of a small number of individuals *obtained with consent*.
- **Labeling noise** - Web images collected by searching the names of celebrities often contain errors. Our synthetic data has *guaranteed correctness of labels*.
- **Data bias** - Face recognition models are generally trained and tested on celebrity faces, many of which are taken with strong lighting and make-up. They also have imbalanced racial distribution. Our synthetic data generation pipeline allows us to control the distribution of the data and ensure a *fair* dataset.

About the Dataset

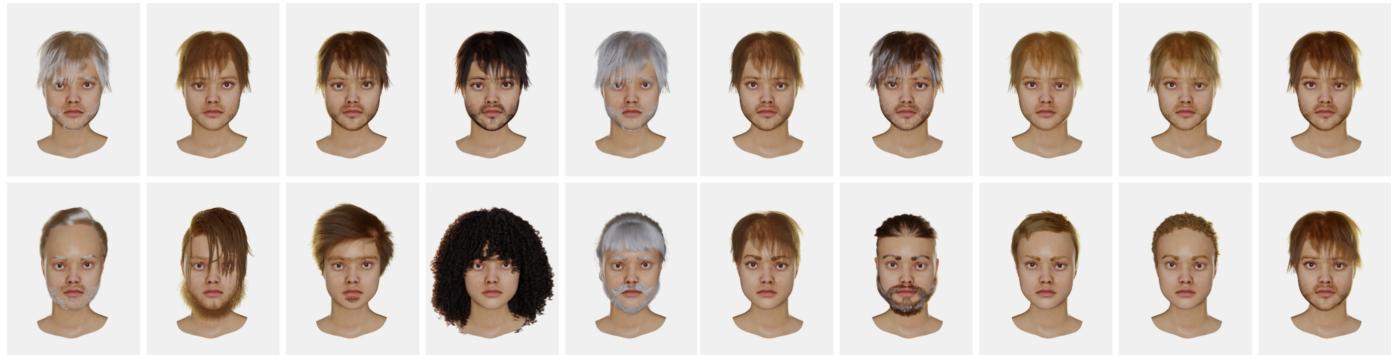
We build on the synthetic face generation framework of [Wood et al.](#) to create a dataset of **over one million** synthetic face images. This synthetics dataset helps us to overcome three primary shortcomings of existing large-scale face recognition datasets:

We define identity as a unique combination of facial geometry, texture, eye color and hair style. For each identity, we sample a set of accessories including clothing, make-up, glasses, face-wear and head-wear.

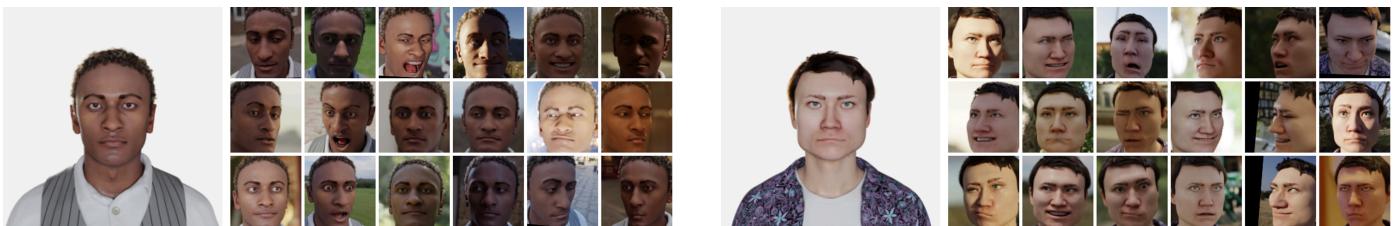


While hair style can change for an individual, most people maintain similar hair style (for both facial and head hair) which makes hair style an important cue for the person's identity.

Therefore, for the same identity, we randomize only the color, density and thickness of the hair (top row) and avoiding the impression of changing identity (bottom row). This simulates aging to some extent as hair typically becomes grayer, sparser and thinner during aging. The hair style is only changed when the added head-wear is not compatible with the original hair style.



After sampling the identity and the accessories, we can render multiple images by varying the pose, expression, environment (lighting and background) and camera.



DigiFace-1M is split into two parts:

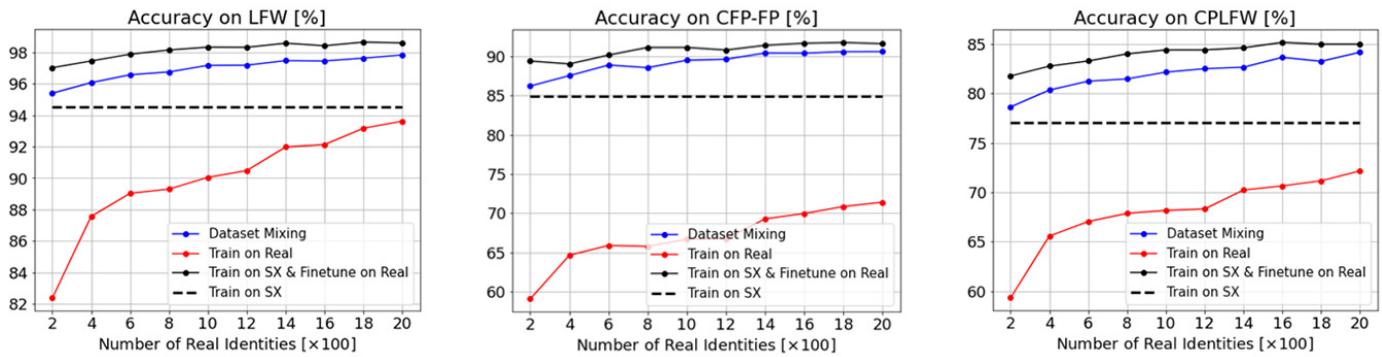
- 720K images with 10K identities (72 images per identity). For each identity, 4 different sets of accessories are sampled and 18 images are rendered for each set.
- 500K images with 100K identities (5 images per identity). For each identity, only one set of accessories is sampled.

Results

SynFace is the current state-of-the-art for face recognition model trained on synthetic faces. They used DiscoFaceGAN to generate 500K synthetic faces of 10K unique identities. We significantly outperform SynFace across all datasets, suggesting that our rendered synthetic faces are better than GAN-generated faces for learning face recognition. This is likely because GAN-generated images do not enforce identity or geometric consistency, and are not effective at changing accessories. The GAN models also have unresolved ethical and bias concerns as they are typically trained on large-scale real face datasets.

| Method | #images | lfw | CFP-FP | CPLFW | AgeDB | CALFW | Avg |
|---------|--------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| SynFace | 500K (10K×50) | 91.93 | 75.03 | 70.43 | 61.63 | 74.73 | 74.75 |
| Ours | 500K (10K×50) | 95.40 | 87.40 | 78.87 | 76.97 | 78.62 | 83.45 |
| Ours | 1.22M (10K×72+100K×5) | 95.82 | 88.77 | 81.62 | 79.72 | 80.70 | 85.32 |

When a small number of real face images are available, we can use them to fine-tune the network that is pre-trained on our synthetic data. Such fine-tuning significantly improves the accuracy across all datasets.



However, there remains a substantial accuracy gap from the state-of-the-art methods that are trained on large-scale real face datasets. This gap can be reduced by adopting better data augmentation or by improving the realism of the face generation pipeline. We leave this as future work.

| Method | #synth images | #real images | LFW | CFP-FP | CPLFW | AgeDB | CALFW | Avg |
|-------------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Ours | 1.22M | 0 | 96.17 | 89.81 | 82.23 | 81.10 | 82.55 | 86.37 |
| Ours + Real | 1.22M | 120K | 99.33 | 95.93 | 89.47 | 91.55 | 91.78 | 93.61 |
| CosFace | 0 | 5.8M | 99.78 | 98.26 | 92.18 | 98.17 | 96.18 | 96.91 |
| MagFace | 0 | 5.8M | 99.83 | 98.46 | 92.87 | 98.17 | 96.15 | 97.10 |
| AdaFace | 0 | 5.8M | 99.82 | 98.49 | 93.53 | 98.05 | 96.08 | 97.19 |

BibTeX

```
@inproceedings{bae2023digiface1m,
    title={DigiFace-1M: 1 Million Digital Face Images for Face Recognition},
    author={Bae, Gwangbin and de La Gorce, Martin and Baltru{\v{s}}aitis, Tadas},
    booktitle={2023 IEEE Winter Conference on Applications of Computer Vision (\year=2023)},
    organization={IEEE}
}
```

Work conducted at [Mesh Labs – Cambridge](#).

