Digi2Real Dataset Documentation
https://www.idiap.ch/paper/digi2real/

Anjith George, Sébastien Marcel

Idiap Research Institute, Martigny, Switzerland.

Digi2Real is a synthetic face dataset containing images of 20,000 unique synthetic identities. We introduce a novel framework for realism transfer, designed to enhance the realism of synthetically generated face images. This framework is applied to a subset of the DigiFace dataset to produce photorealistic images, which are more effective for training face recognition models than the original DigiFace dataset. By combining the controllable features of the graphics pipeline with our realism enhancement technique, we present a new approach for creating attribute-controllable face recognition datasets.

1 Description

The Digi2Real dataset includes images of 20,000 unique individuals, with nearly 20 images per identity, totaling 399,355 images. All images are aligned and cropped to a standard resolution of 112×112, commonly used in face recognition research. Along with the images, we provide *MXNet rec files*, a widely-used format in the face recognition community. For further details on the dataset and the creation process, please refer to the reference paper.

An example set of images from the original DigiFace dataset and their transformed versions in Digi2Real is shown in Fig. 1.
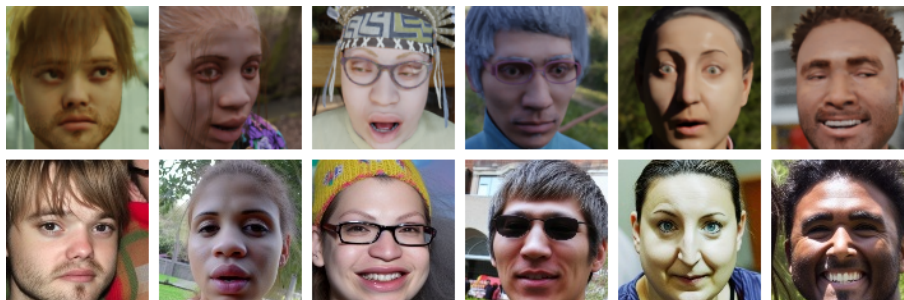

Figure 1: Examples of original images from the DigiFace dataset (first row) and the corresponding transformed images using our approach (second row)

2 Access to the Digi2Real database

The database is available for download free of charge.

3 References

The Digi2Real dataset was originally introduced in the following paper, please cite this paper if you use the dataset:

- Anjith George and Sébastien Marcel, "Digi2Real: Bridging the Realism Gap in Synthetic Data Face Recognition via Foundation Models", *arXiv, 2411.02188*, `https://arxiv.org/abs/2411.02188`.