Supplementary Material

1. Image morphing

The embedding loss of the X-GAN ensures a smooth transition in the latent space. To visually underpin this, in figure 1 and figure 2 we show the results of an image morphing experiment. The corners of the shown image arrays are real images while the space in between is filled with fake samples. The transition between fake and real image is hardly visible. Even fine textural details are maintained. This shows, that the X-GAN is capable of synthesizing images from the continuous manifold which is defined by discrete samples from the underlying real dataset. The fake images contain features from all adjacent neighbor. While pure autoencoder based approaches yield unnatural looking results superimposing several features, the GAN component of our model ensures realistic looking synthesized images.

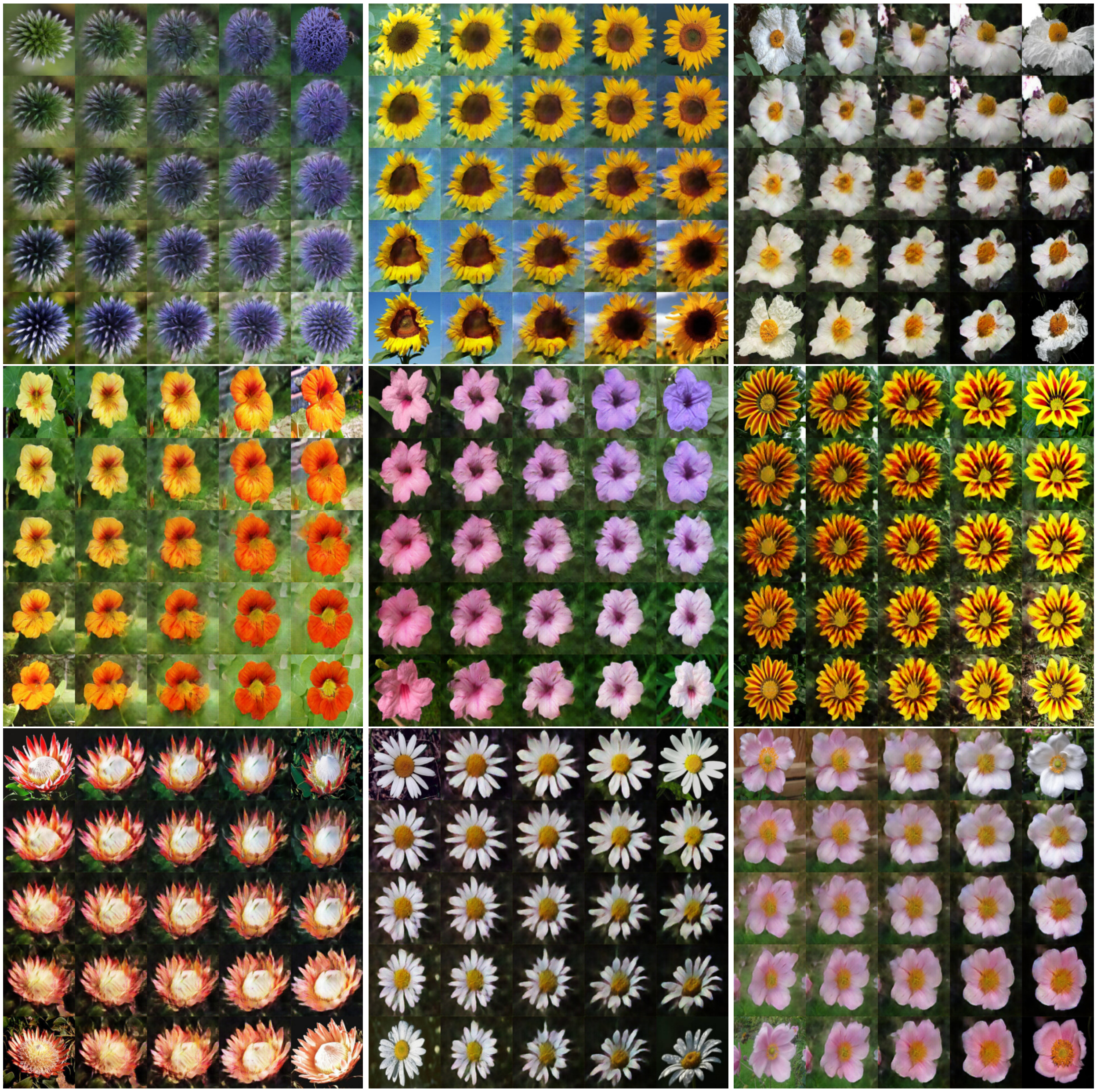
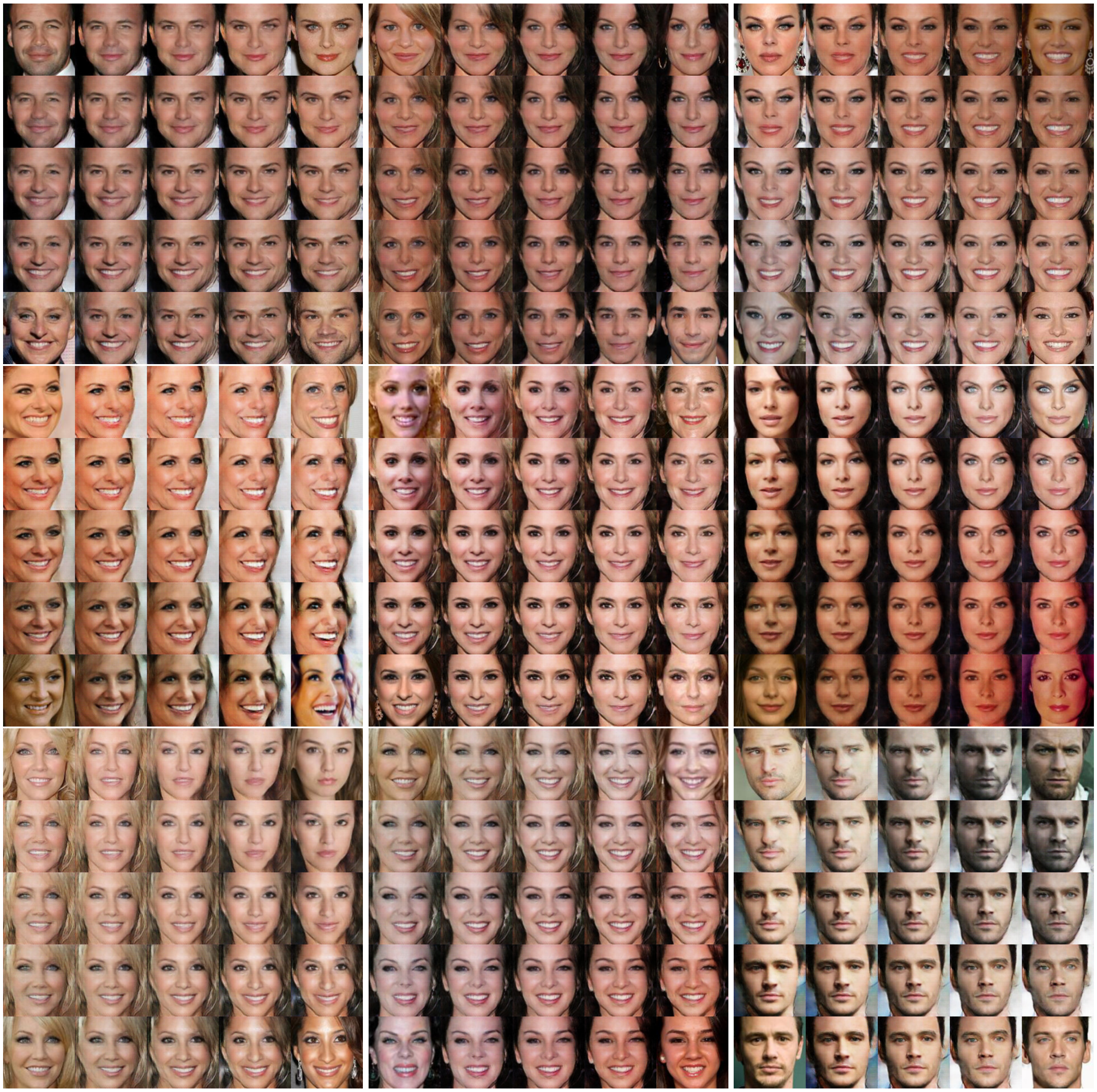


Figure 1. The figure shows several image morphing results for a X-GAN trained on the 102 Category Flowers dataset. For each of the shown image-arrays, the four corner images are real and the space in between is filled with fake images gained from weighted convex combinations.



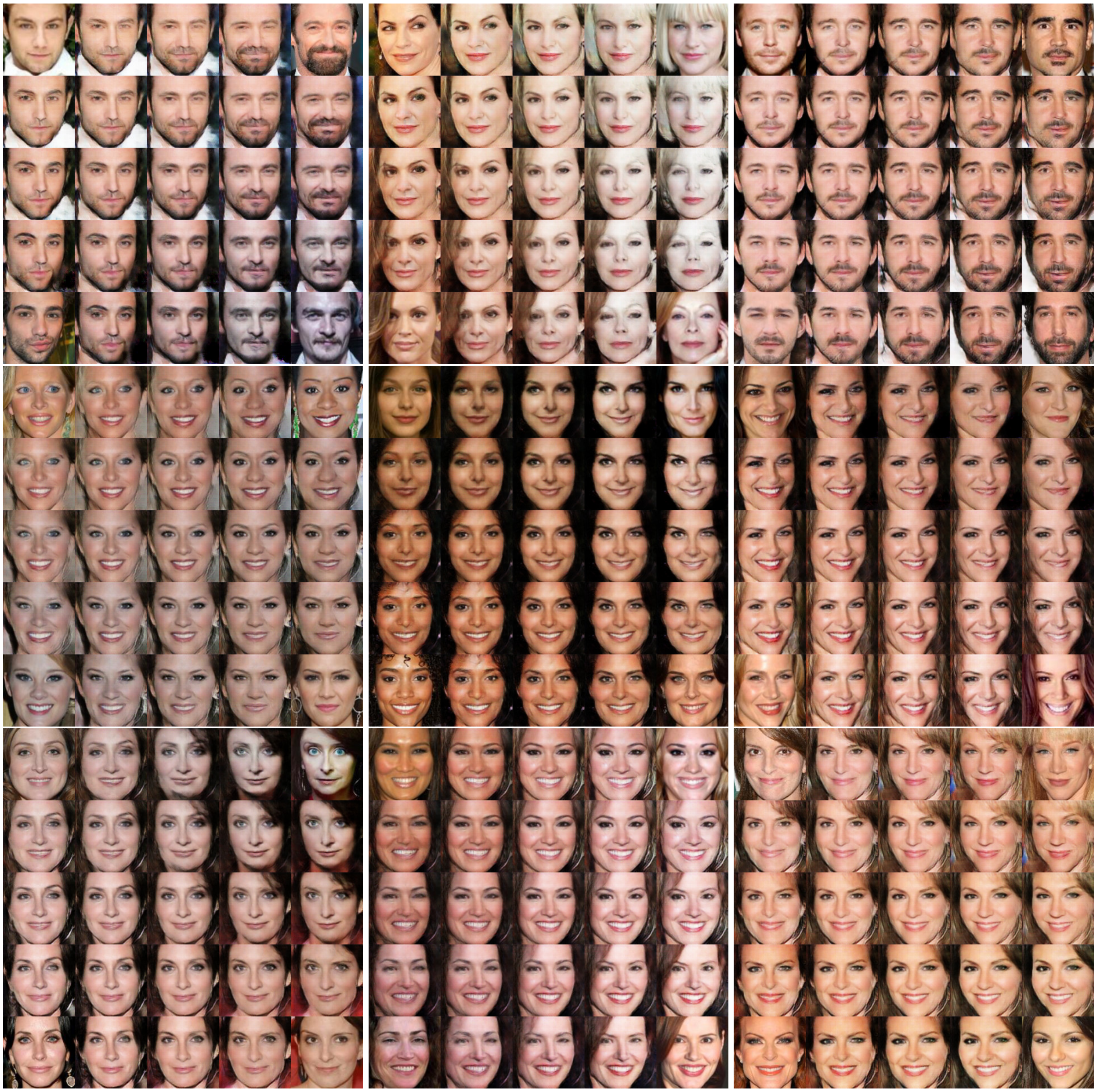


Figure 2. The figure shows several image morphing results for a X-GAN trained on the FaceScrub dataset. For each of the shown image-arrays, the four corner images are real and the space in between is filled with fake images gained from weighted convex combinations.

2. Reconstruction of Unseen Test Samples

In figure 3 we show that our model is capable of reproducing unseen test images. The X-GAN combines the features of several training images to produce a new individual synthesized image. This way the X-GAN is able to reconstruct the full continuous manifold. This is shown by the fact that in many cases the X-GAN is able to generate images that are very close to unseen test samples that are also expected to lay on this manifold. In figure 3 for several test images (a) we show the nearest neighbor in 1000 X-GAN synthesized images (b), 1000 CVAE-GAN synthesized images (c) and ~ 100 real training images (d). Hereby the test images, the X-GAN synthesized images, the CVAE-GAN synthesized images and the training images have the same class label. For this experiment we utilize the cosine distance of the FaceNet embedding as metric for the nearest neighbor search.

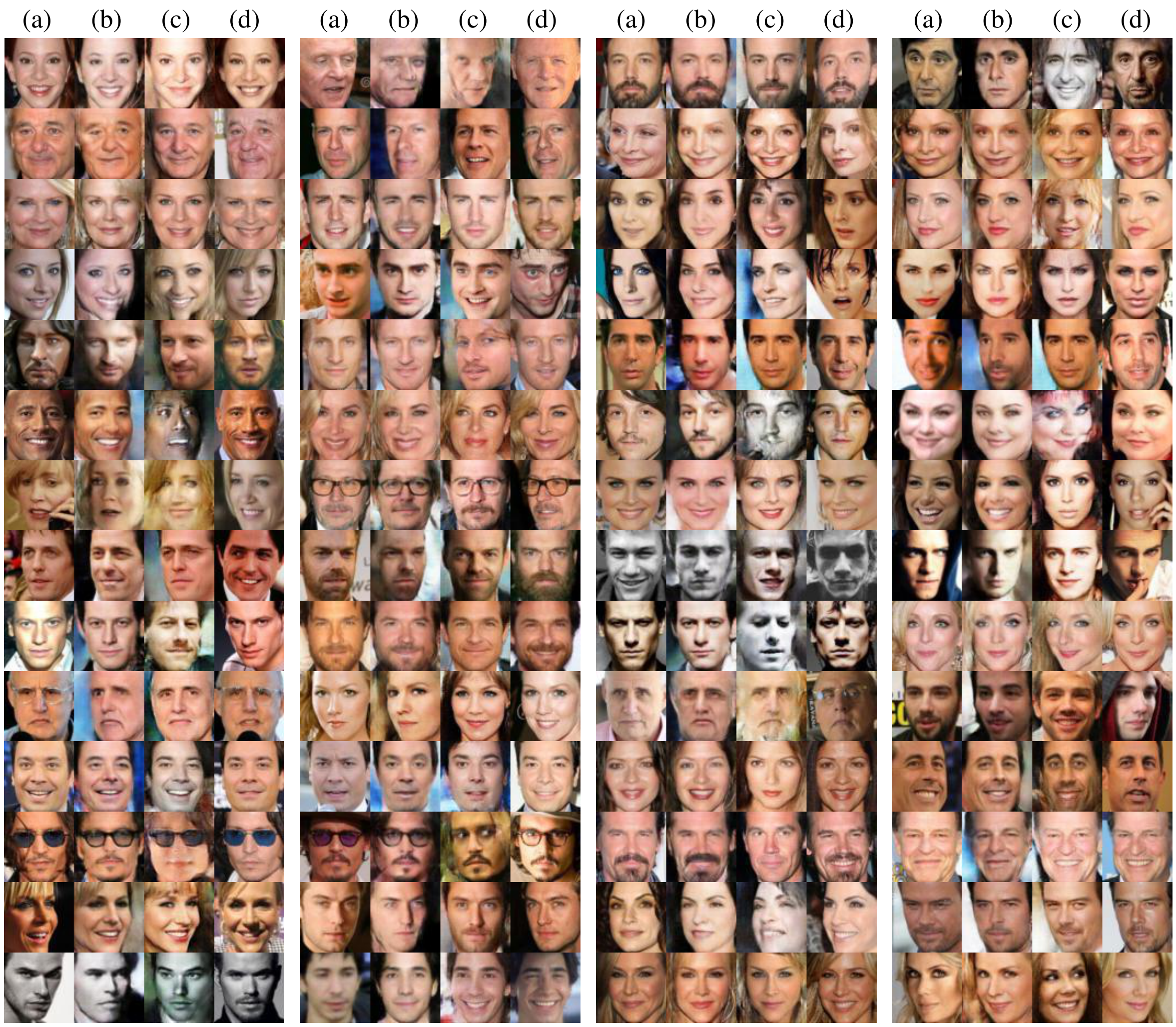


Figure 3. This figure shows, that the X-GAN is capable of reproducing unseen test data. (a) test images (b) nearest neighbor in X-GAN synthesized images (c) nearest neighbor in CVAE-GAN synthesized images (d) nearest neighbor in training set. The nearest neighbor is computed with the distance defined by the FaceNet embedding.