Supplementary materials for:
CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training

1. Analysis of the latent vector z

Since the network G is able to reconstruct the input image only with the latent vector z and category label c. Therefore, it is expected that it can conveniently encode all the attribute information, such as pose, color, illumination, and even more complex high level styles, in the latent vector. In this section, we will introduce some interesting findings about the latent vector.

The same latent variable represents the same attribute. One important finding is that, although we do not use any supervision on the attributes, the same latent vector for different labels will generate images with different category labels but with similar attributes. The reason for this phenomenon may be that images with the same attribute present certain resemblance at the pixel level. So the network automatically put them together through unsupervised clustering.

To confirm this, we train a model on a face dataset, FaceScrub, and then extract the latent vector of all the face images. In order to clearly present the distribution, we project all the latent vectors into a two dimensional space by PCA. As shown in Figure S1, the distribution of latent vector is a Gaussian as expected, and the attribute of images, include face pose, illumination, and the background is the same for the same latent vector.

With this property, our algorithm can be used in many other applications, such as attributes transformation which generates images with different category labels but with similar attributes, and attributes retrieval which searches for other images with similar attributes.

1.1. Attributes transformation

In this part, we will experimentally validate the attributes transformation. We test our method on FaceScrub dataset. Given a source image, we first use the encoder network E to extract the latent vector z. Then using this latent vector, we can generate images in any specific category. Figure S2 shows the results of attribute transformation. We generate an image in the target category. And its attribute is similar to the source image.

1.2. Attributes retrieval

Through the above analysis of the latent vector, we come to such an inference: the similar latent vector represents similar attribute. Therefore, we can use our encoder model to extract the attributes features, and use them to search for the image with the most similar attribute in a dataset. We use FaceScrub dataset for this experiment. We first extracted all attribute features by the encoder network E. Then we simply conduct a image retrieval task by using ℓ_2 distance. As shown in Figure S3, we show the top 5 results that most similar to the query image but with different category. We found the faces with similar skin color, viewpoint or emotion.

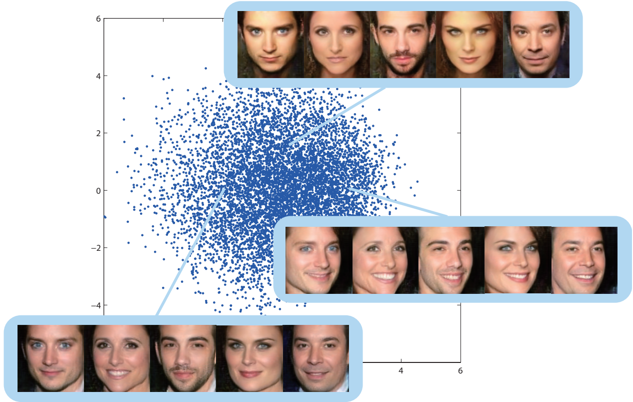
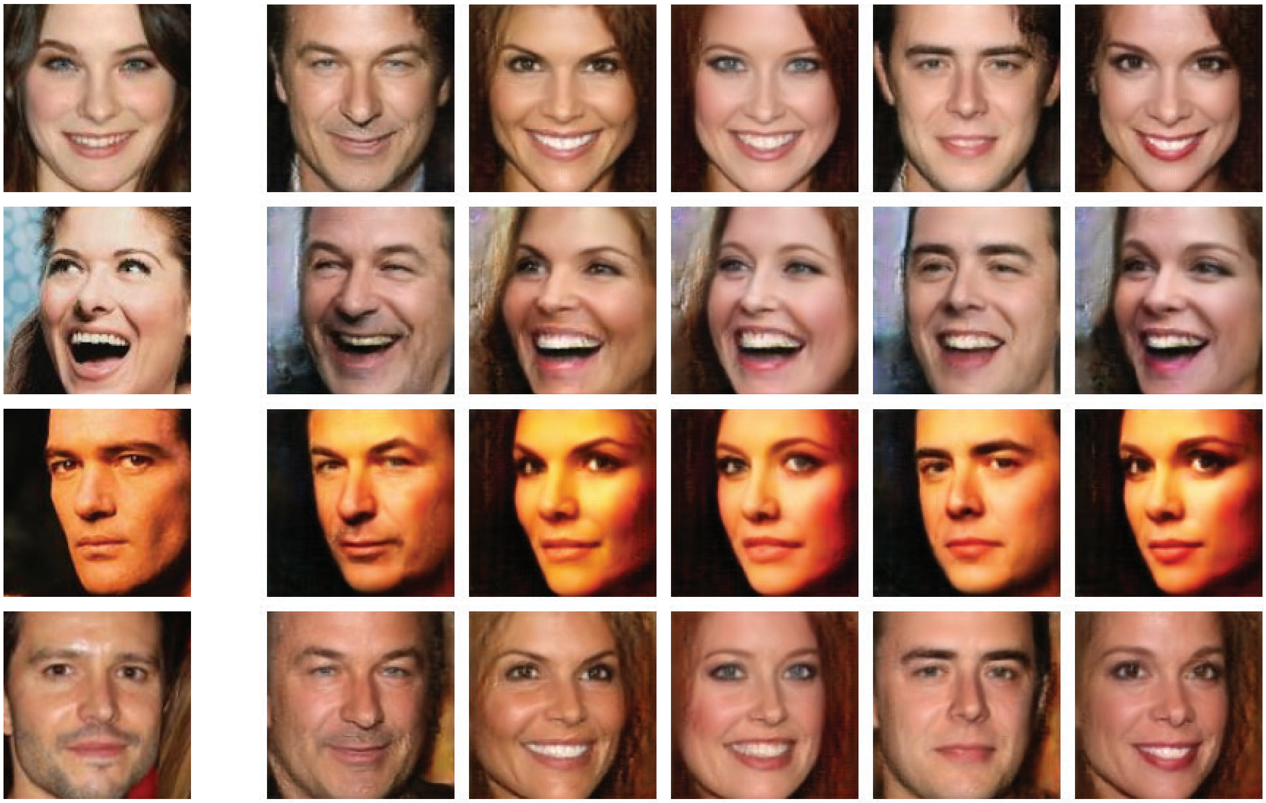


Figure S1. Illustration of the distribution of latent vector and the generated image with the same latent vector. As the readers have observed, images generated with the same latent vector z but different categories will have the same attributes, such as pose, illumination, and background.

2. Nearest Neighbors Test

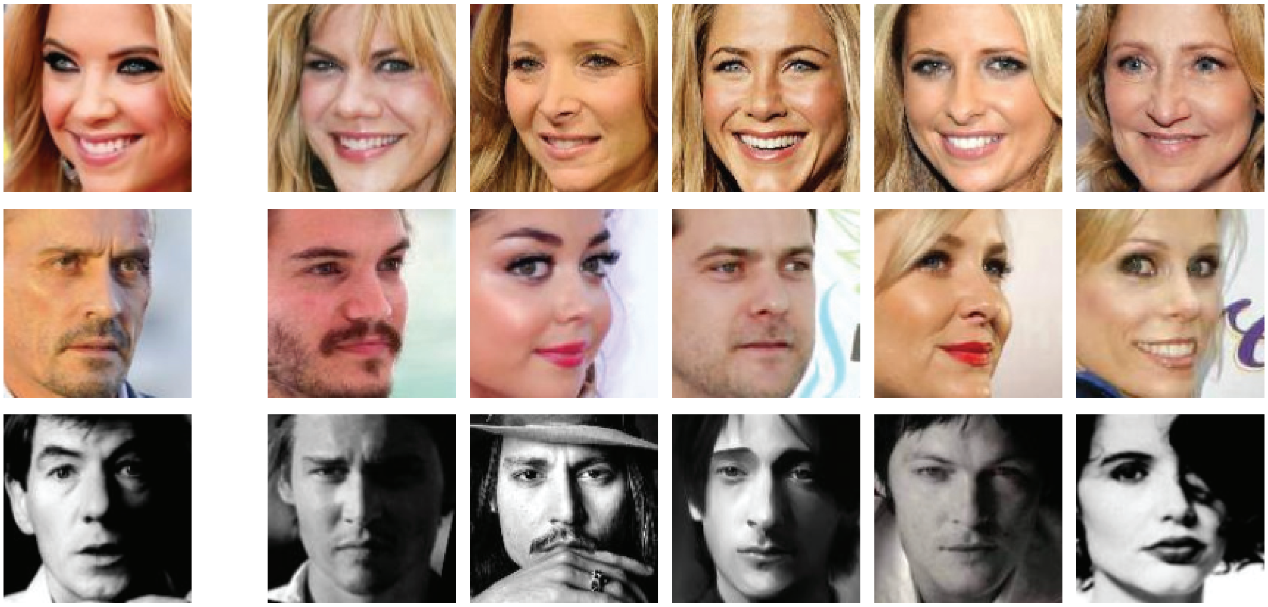
In this section, we want to demonstrate that our model is not just memorizing all training examples in the training process. Our model can generate samples which are not copies of the training samples. We choose FaceScrub dataset for this experiment. Firstly, we randomly generate 8 samples from 8 different categories. Then we simply conduct an image retrieval task using ℓ_2 distance at the pixel level. As shown in Fig. S4, we show the top 5 results that most similar to the generated images, we find that the generated samples have the same identity but different attributes like pose, illumination, and emotion as compared to nearest neighbors.



a) Original images

b) Generated images

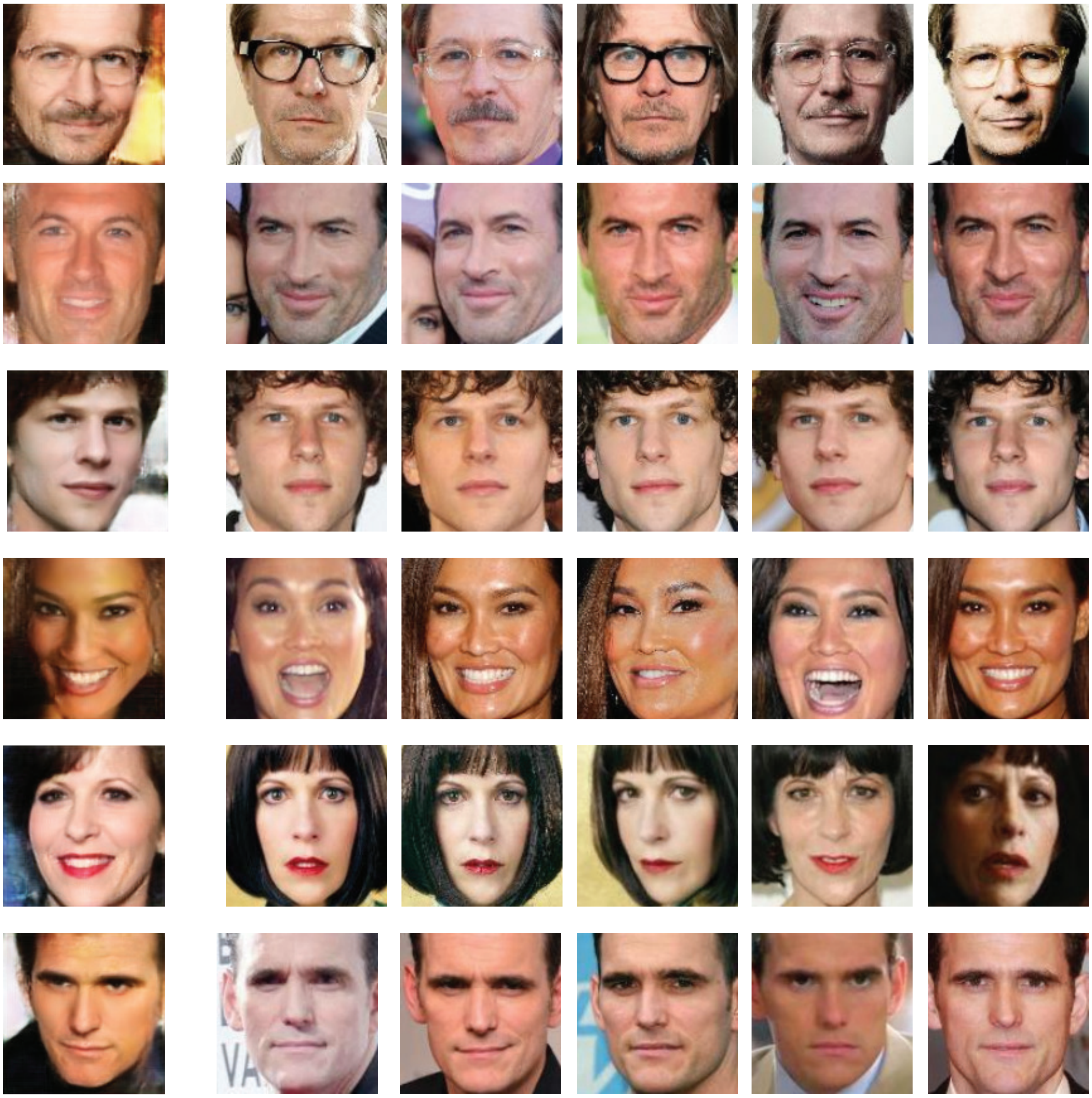
Figure S2. Results of attributes transformation. a) Original images from the Facescrub dataset, which offer the attributes for the generated images. b) Generated images using latent vector of the original images and the target category c. From the results, we can observe that the generated images have the same attributes as the original images.



a) Query images

a) Top 5 answers

Figure S3. Results of same attributes retrieval. a) The query images. b) The top 5 answers in the original datasets.



a) Generated images b) Top-5 nearest neighbor in the origin datasets

Figure S4. Results of nearest neighbors search on the generated samples. a) The generated samples from the CVAE-GAN model. b) The top-5 nearest neighbors real samples, which shows the novelness of generated images.