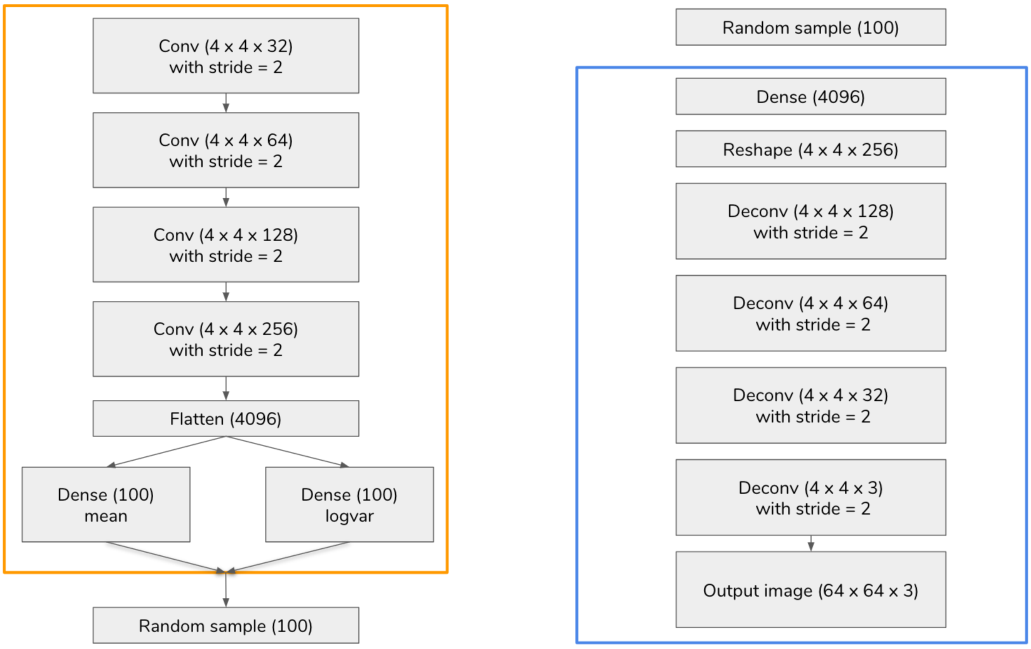
DLCV HW4 - d05921027 Chun-Min Chang

Problem 1. VAE

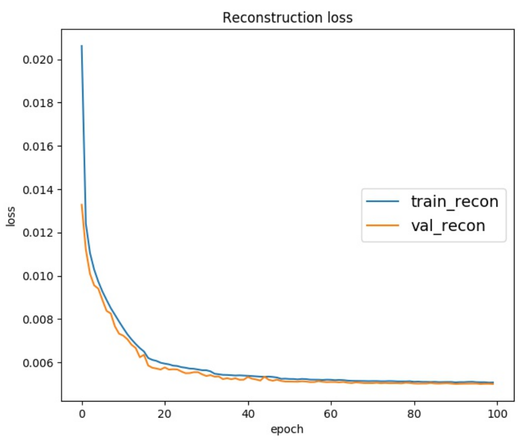
1. Describe the architecture and implementation details of your model

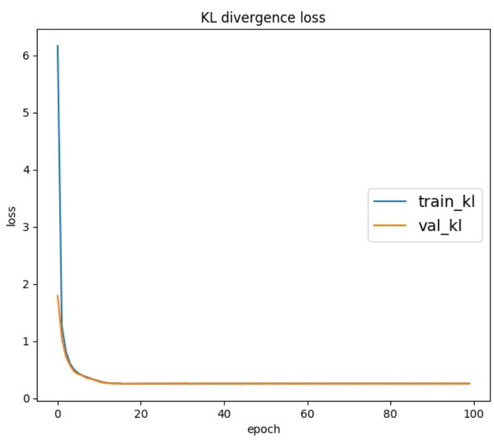
Encoder Decoder



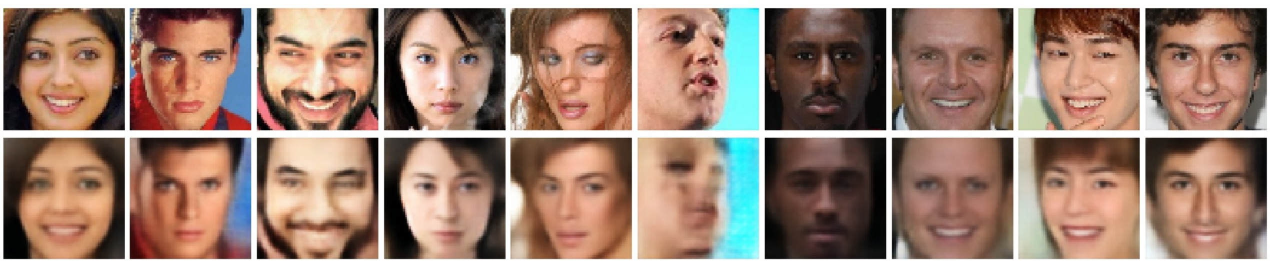
Encoder is composed of 4 conv layers, and each filter size is 4 x 4. We set stride to 2 instead of using pooling layers. The activation function of the output layer uses tanh, and then we rescale to [0, 255] as the corresponding RGB values.

2. Plot the learning curve of your model



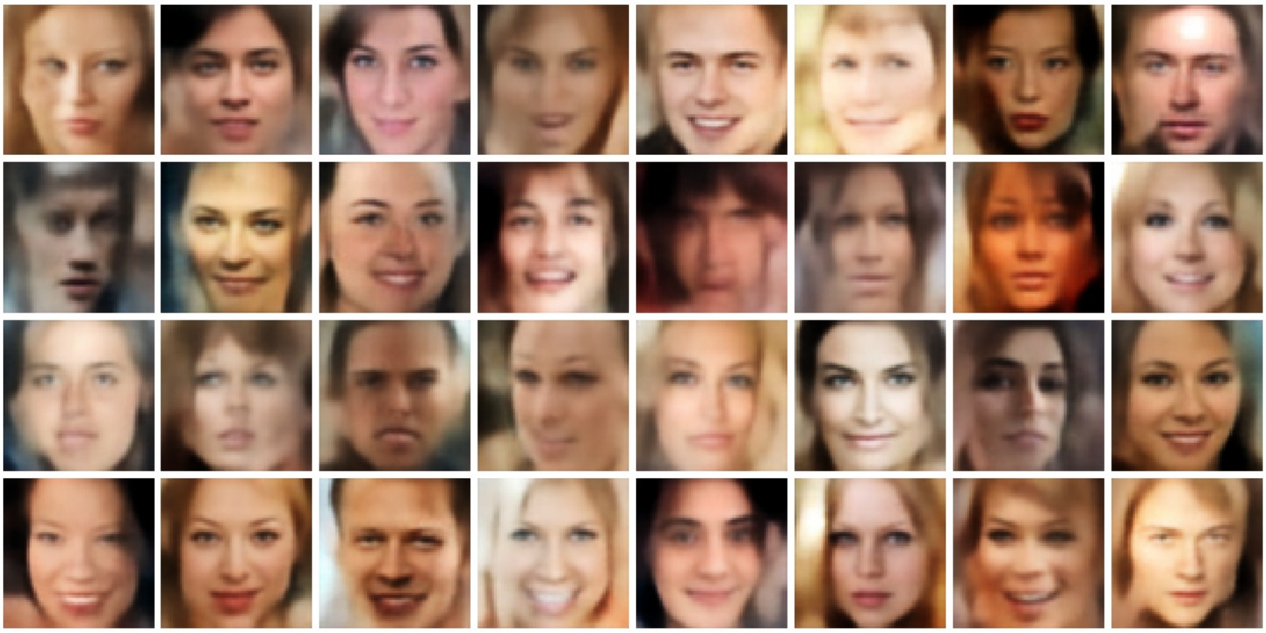


3. plot the first 10 testing images and their reconstruction result of your model and report your testing MSE of the entire test set

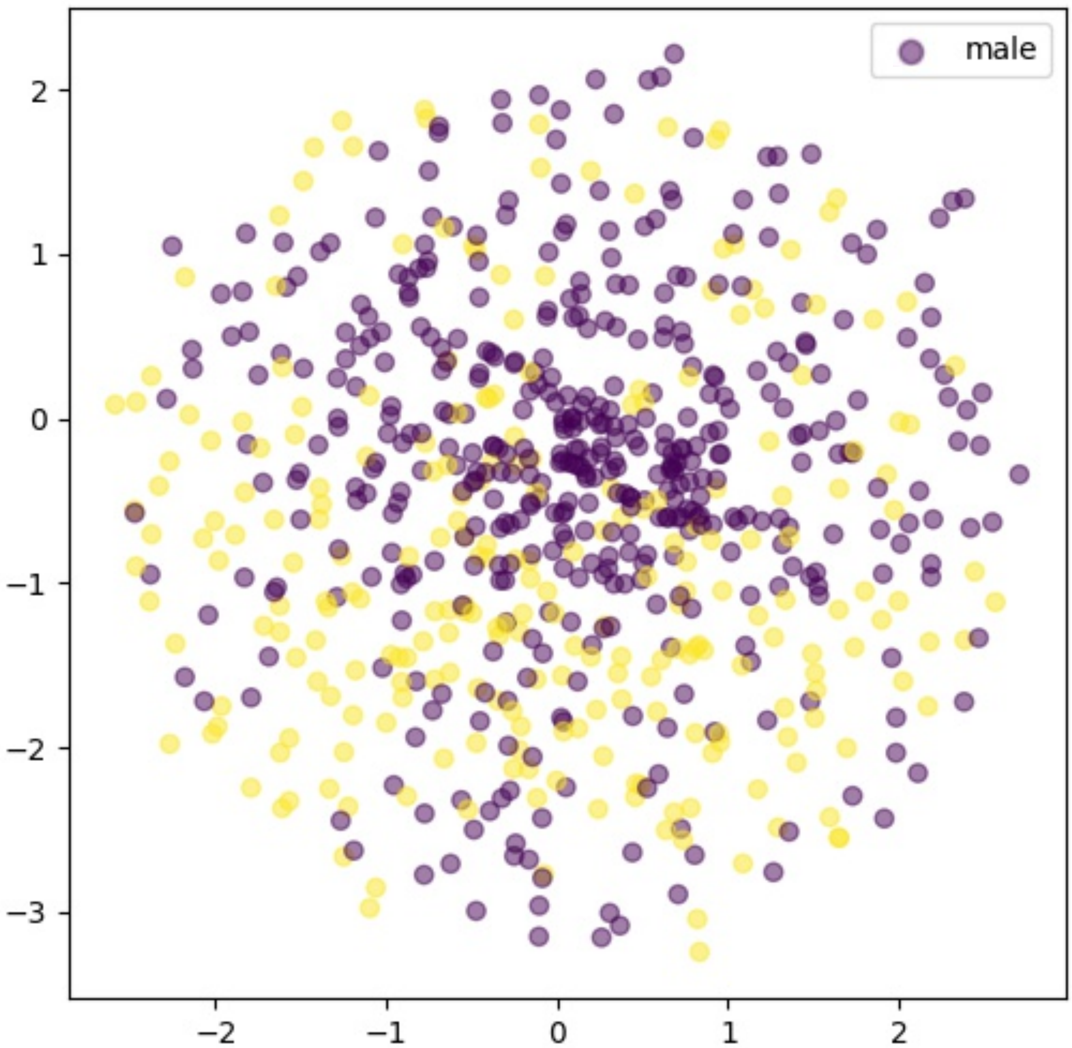


testing MSE = 0.0050

4. Plot 32 randomly generated images of your model



5. Visualize the latent space by mapping test images to 2D space (with t-SNE) and color them with respect to an attribute of your choice



6. Discuss what you've observed and learned from implementing VAE.

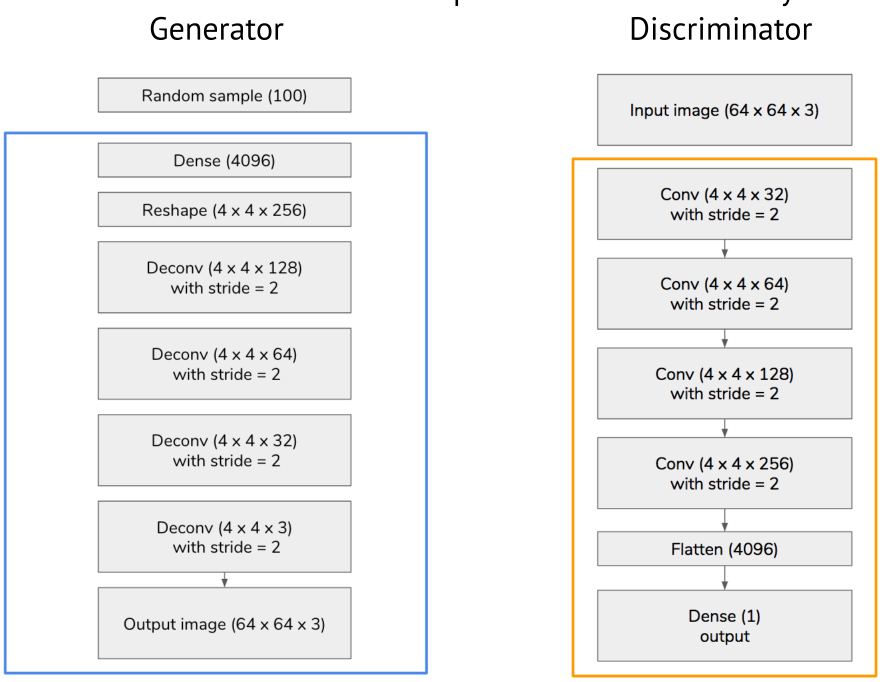
a. The reconstruction loss converges so quickly in comparison with the KL divergence loss.

b. The lambda used to balance the reconstruction loss and KL divergence loss has a decisive effect on image quality. If larger lambda, the quality of generated image will be low-fidelity.

c. VAE is an easy yet effective model since its encoder and decoder has symmetric structure. Besides, the value of loss is highly related to the quality of image, letting us easily understand model's performance.

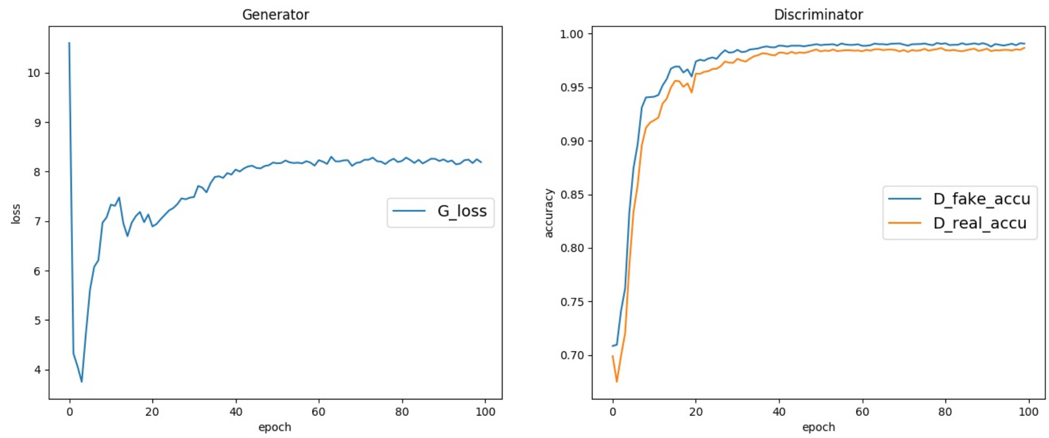
Problem 2. GAN

1. Describe the architecture and implementation details of your model



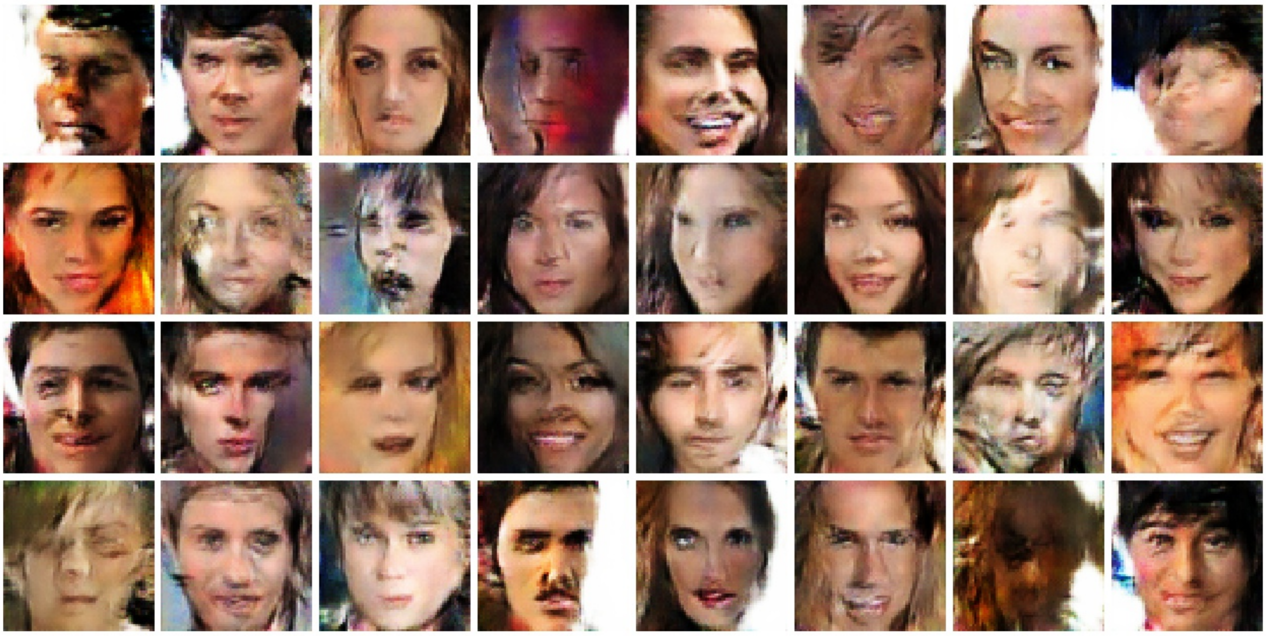
The architecture is similar to VAE. Generator is equivalent to the Decoder in VAE. For Discriminator, we replace the last layer of VAE by a Dense(1) layer, which acts as a binary classification (Real/Fake).

2. Plot the learning curve of your model and briefly explain what you think it represents



- Discriminator: the accuracy stands for the ability to distinguish real/fake.
- Generator: it is hard to interpret the loss.

3. Plot 32 randomly generated images of your model



4. Discuss what you've observed and learned from implementing GAN.

a. The quality of generated images is hard to be quantified. The generated images by the generator of minimal loss are far from good images by human inspection. The loss of the generator cannot represent anything, and this is why GAN is difficult to be well-trained.

b. The alternative training between the generator and discriminator is batch-based instead of epoch-based. If the discriminator is trained for one epoch before training the generator, the discriminator will be too strong and thus the generator will never fool the discriminator no matter what it does.

c. For one batch, I train the discriminator once but train the generator twice in order to make the generator's performance able to come up with the discriminator's performance. They can improve in a fair rate.

5. Compare the difference between image generated by VAE and GAN, discuss what you've observed.

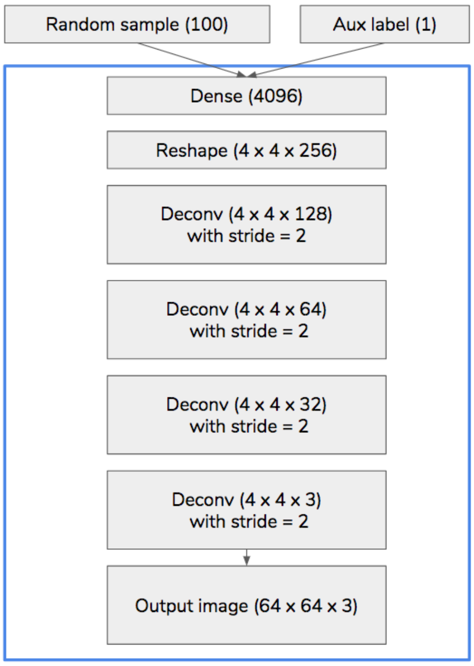
GAN: the generated images by GAN have greater of diversity. They contain blonde-hair, black-hair, smiling man/woman, and have various angles of face orientation. But some images have a large blurred region, and hard to be recognized as human faces.
VAE: the generated images have better fidelity but some of them look so similar. Even if most of them do not have any blurred region, the hair color is close to the skin color and it is sometimes hard to distinguish the boundary between hair and face.

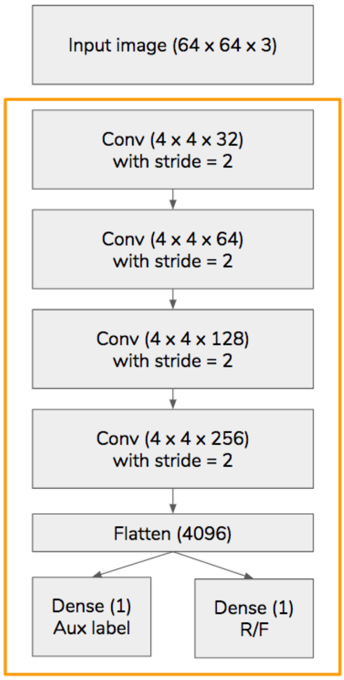
Problem 3. ACGAN

1. Describe the architecture and implementation details of your model

Generator　　　　　　　　　　　　Discriminator

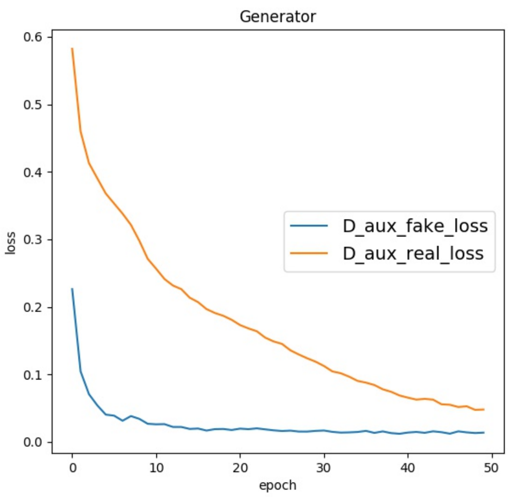
Generator takes a random vector and a given auxiliary label as the input to generate images.

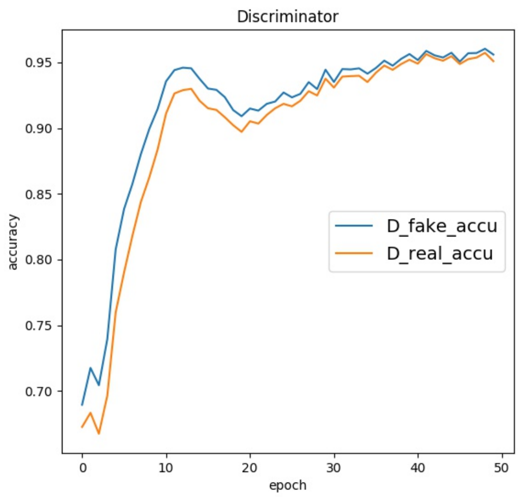




Discriminator has to not only classify real or fake images but also the aux label of images. The structure is similar to GAN, and the only difference is adding aux label into input of generator, and predicting aux label at the output of discriminator.

2. Plot the learning curve of your model and briefly explain what you think it represents





The descent of the aux-label losses implies the generator gradually learned how to generate the pattern of the given label on images. For discriminator, the accuracy stands for the ability to distinguish real/fake.

3. Plot 10 pairs of randomly generated images of your model, each pair generated from the same random vector but with different attribute. This is to demonstrate your model's ability to disentangle feature of interest.

The attribute of the above row is "not smiling"; otherwise, that of the below one is "smiling".