Background:

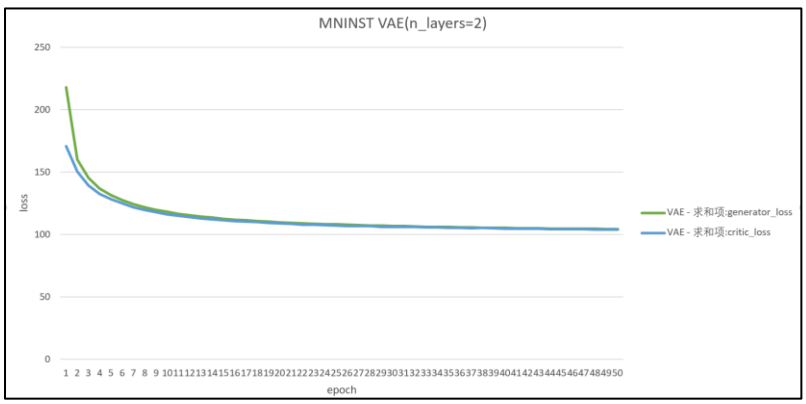
Three models - VAE, GAN, and WGAN – are applied on two datasets, and two hyper parameters - the dimension of the latent space(z_dim) and number of network layers(n_layers) – are tunned. Each training was run 50 epochs(n_epochs=50). Finally, I plot the loss function for VAE, estimated JSD for GAN, and estimated EMD for WGAN and make comparison with them.

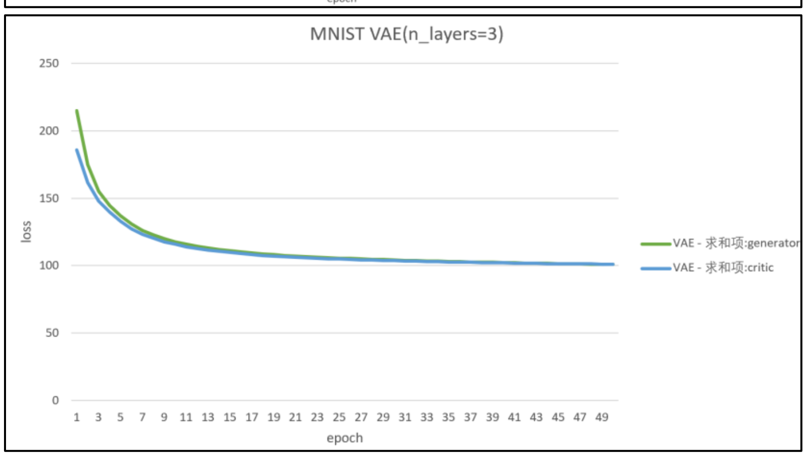
Analysis:

IMAGE COMPARISION:

VAE

- the loss function for VAE for MNIST(z_dim=50)
  Both training loss and validation loss are decreasing, along with the increasing of epoch. When I change the network layer from 2 to 3 for MNIST dataset, their trend and loss seem similar, the minimum loss of higher layer is just a bit smaller than that of lower layer.

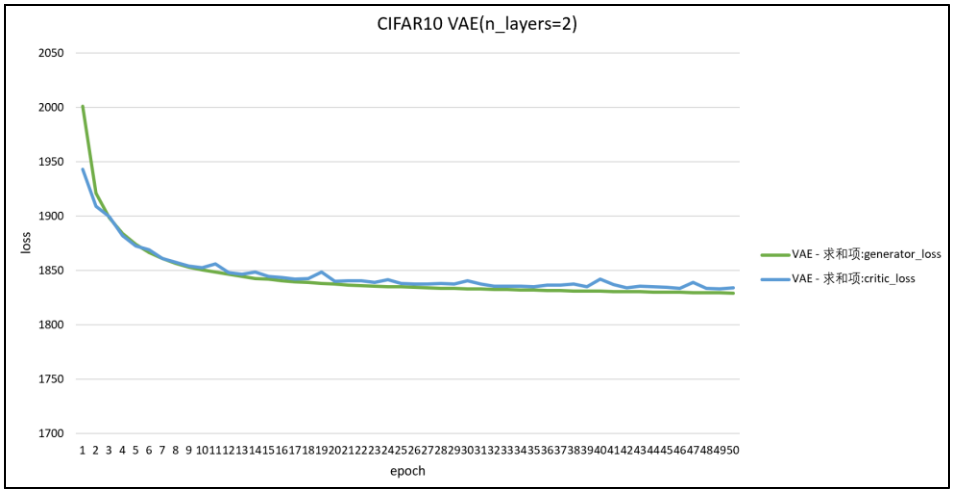


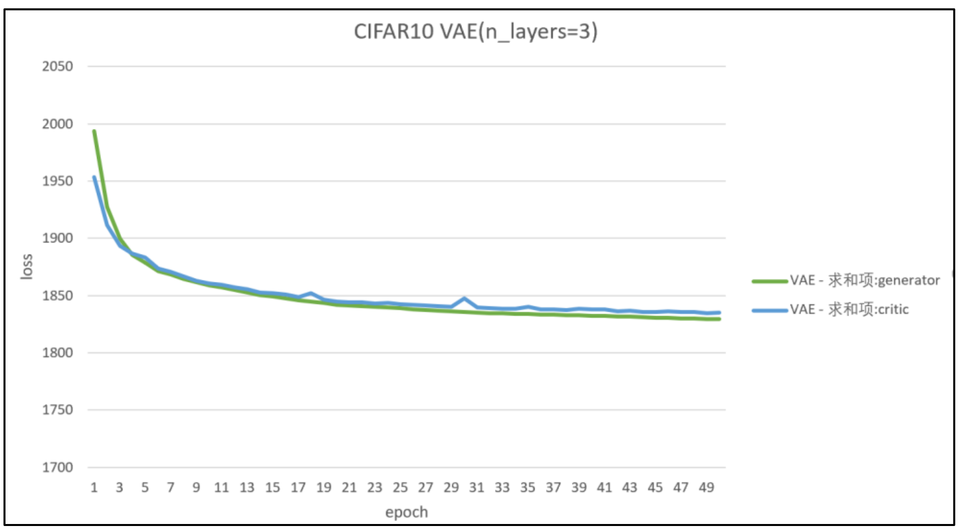


(note: the so called 'generator loss' in VAE represents training loss, and critic loss in VAE represents validation loss,)

- the loss function for VAE for CIFAR10(z_dim=50)

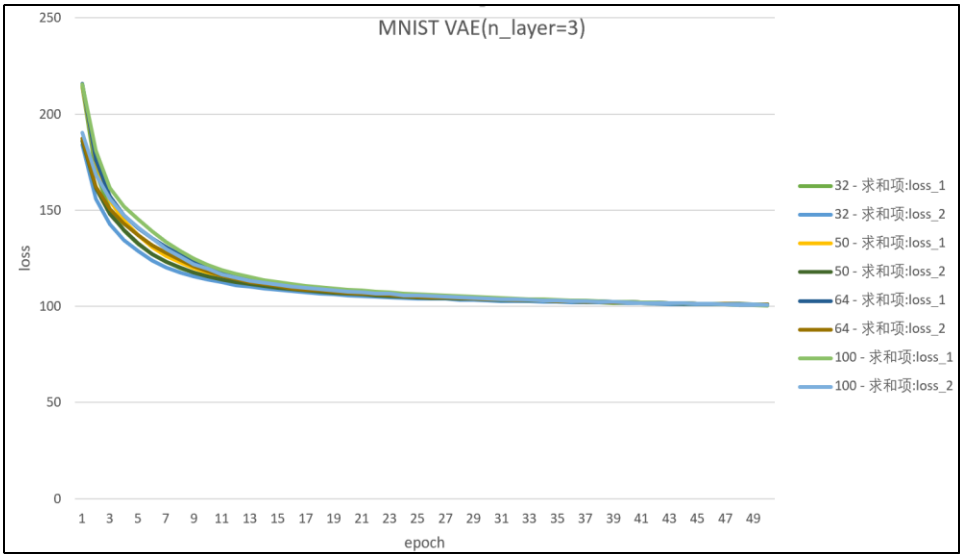
The trend of training loss and validation loss are decreasing, along with the increasing of epoch, however the validation loss fluctuate sometimes.
When I change the network layer from 2 to 3 for CIFAR10 dataset, their trend and loss seem similar. But the validation losses from the network with higher layers fluctuate less.
The loss of CIFAR10 is much bigger than MNIST's: 1825>>100.

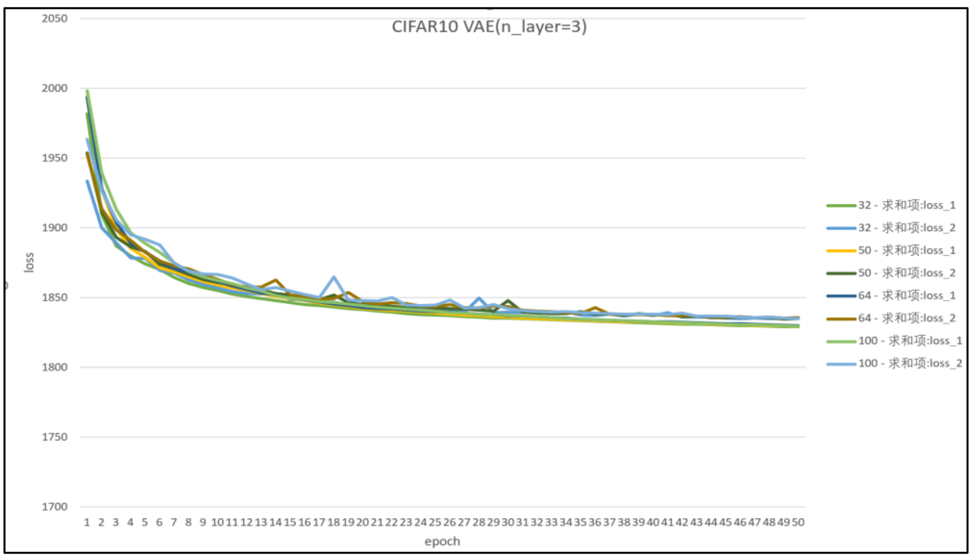




They have less identification, comparing to MNIST dataset.

- The loss function under different z_dim while n_layers=3:
  Loss_1 means the training loss, loss_2 means the validation loss.
  The trend of different z_dim seems similar. However, model for MNIST with higher dimension of the latent space converge a bit quickly.

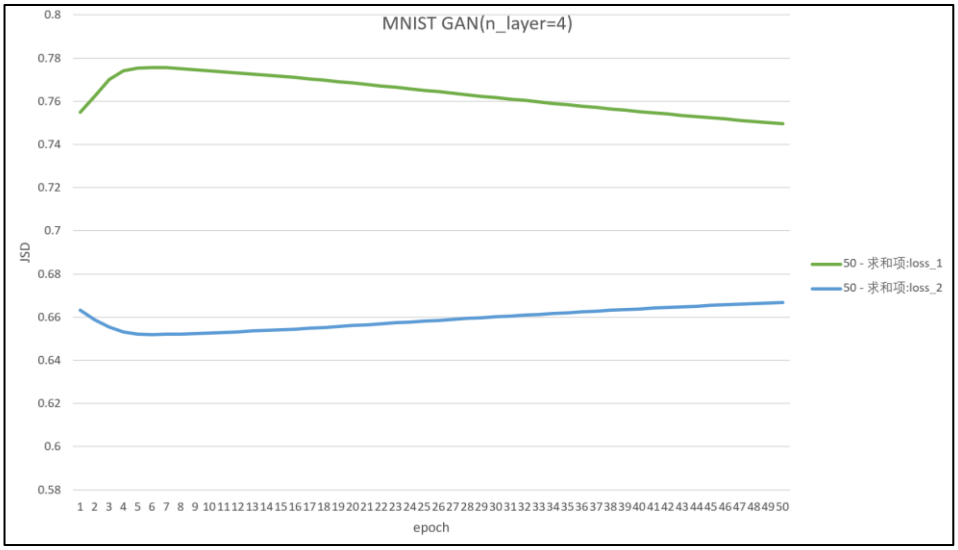


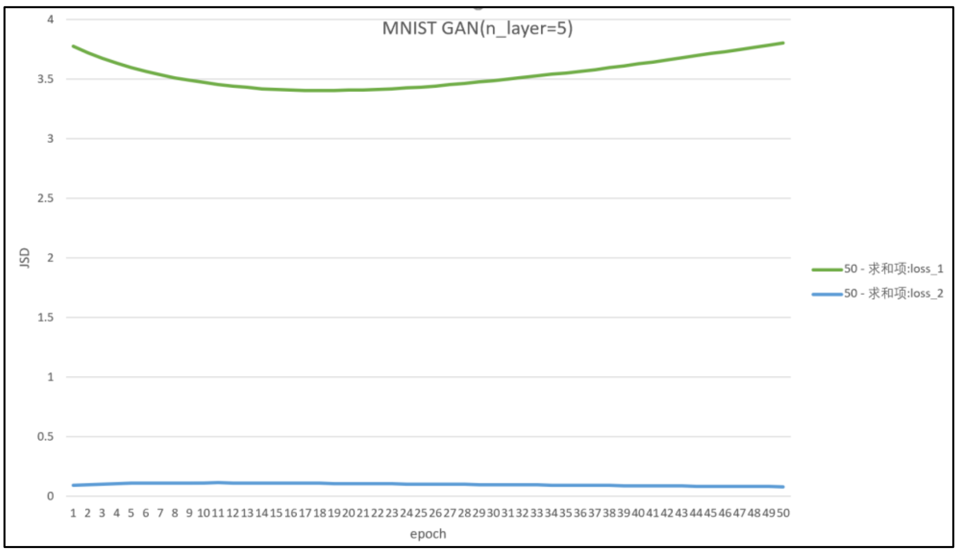


In conclusion, when I change the layers of network from 2 to 3, and the dimension of laten space from 32 to 100, the trend and loss of two datasets themselves didn't vary much.

GAN

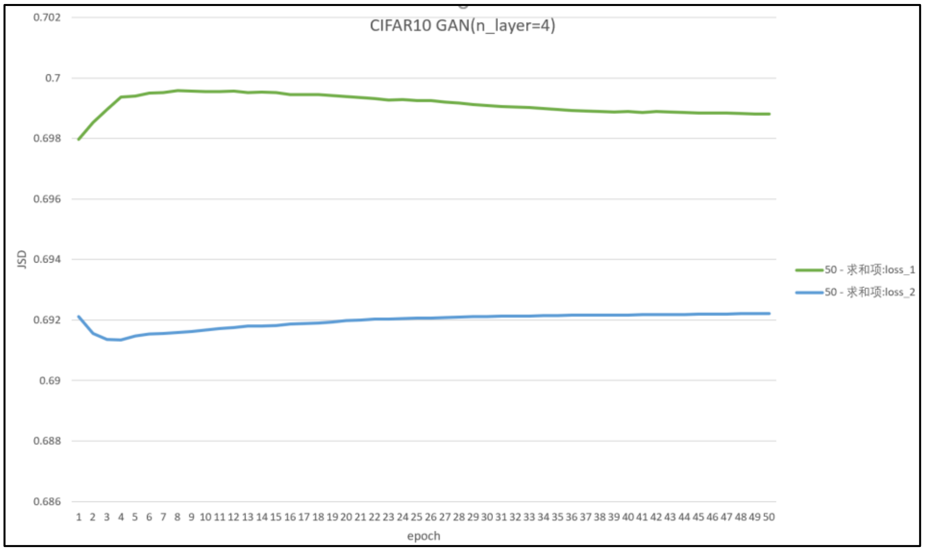
- Estimated JSD of GAN for MNIST(z_dim=50)
  It performs better with lower network(n_layers=4) and the JSD seems to closer to each other and become smaller.

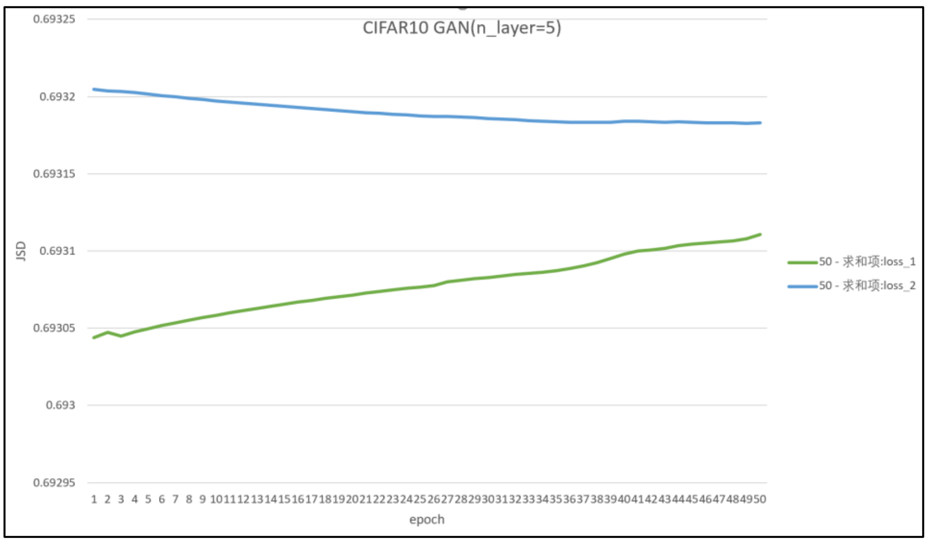




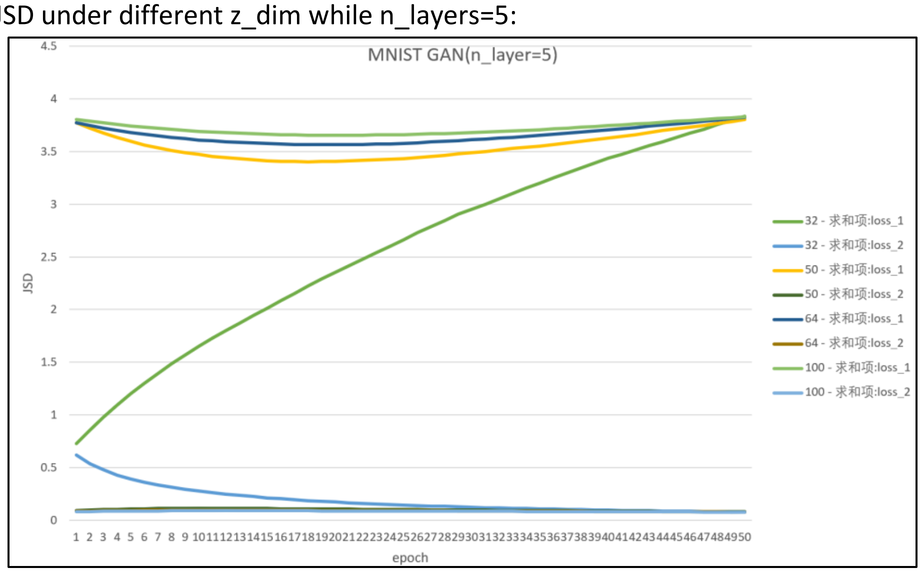
Note: loss_1: generation loss; loss_2: discrimination loss

- Estimated JSD of GAN for CIFAR10(z_dim=50)
  It performs better with lower network(n_layers=4) and the JSD seems to closer to each other and become smaller.



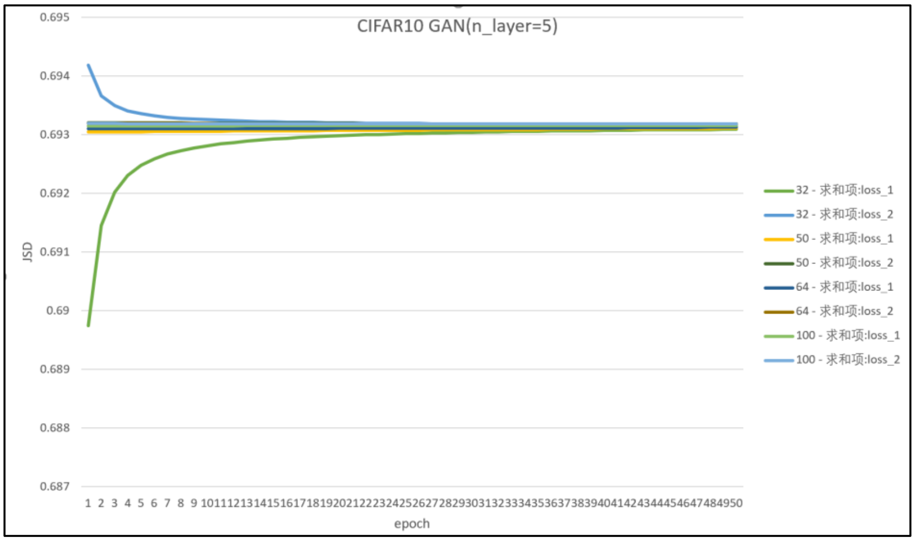


- The JSD under different z_dim while n_layers=5:



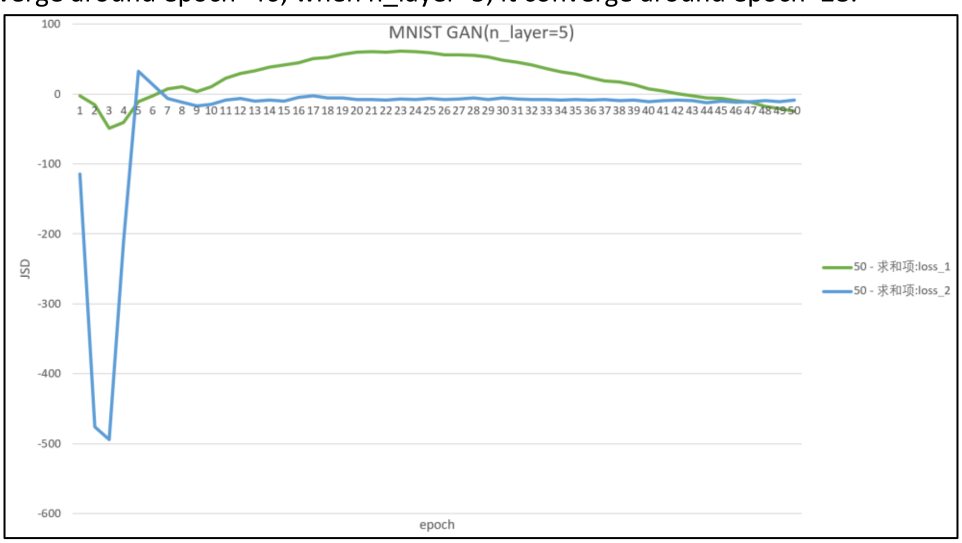
It is hard to judge which z_dim is the best according to above picture, we can do a comparison of a series of images under different z_dim later.

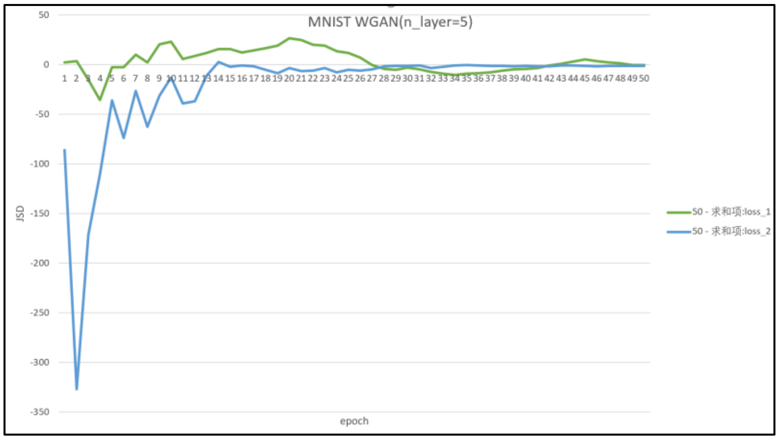
For CIFAR10, except low laten dimension, the others' JSD starts from lower value and converge quickly to lower value, too.



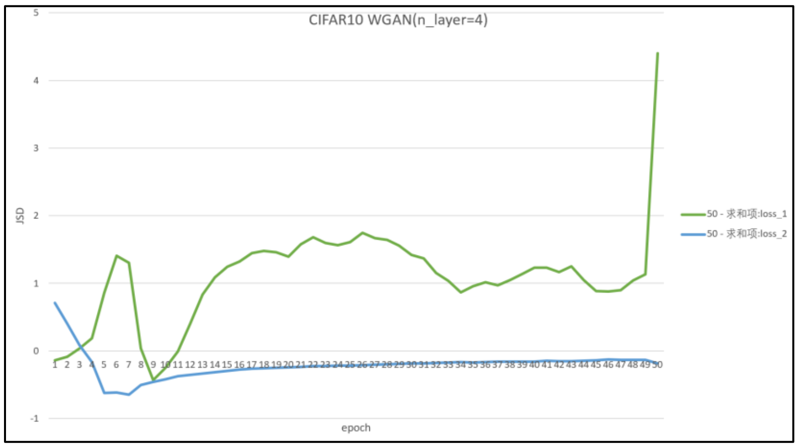
<u>WGAN:</u>

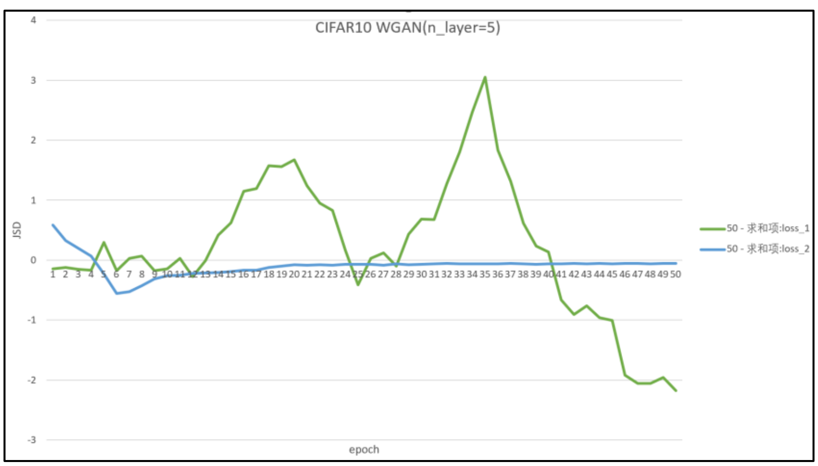
- EMD of WGAN for MNIST(z_dim=50)
  Note: the first is MNST WGAN(n_layer=4)
  WGAN EMD can converge more quickly with more network layer. When n_layer=4, it converge around epoch=46; when n_layer=5, it converge around epoch=28.



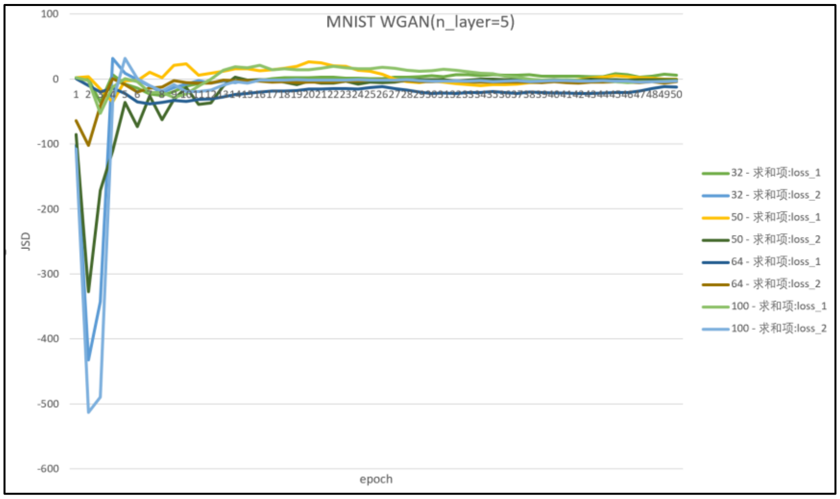


- EMD of WGAN for CIFAR10(z_dim=50)
  n_epochs=50 seems not enough for GAN, it need more training to perform better.





- The EMD under different z_dim while n_layers=5:
  n_epochs=50 seems not enough for GAN, it need more training to perform better.
  Changing n_layers or z_dim just impact little.



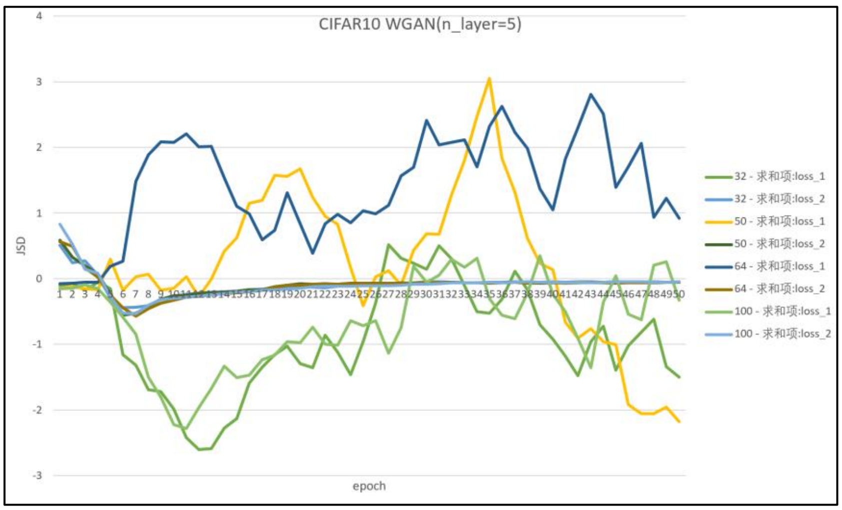
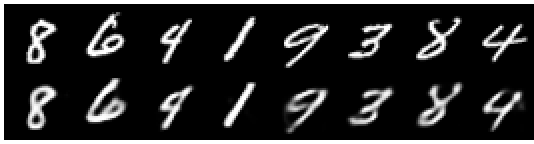


IMAGE COMPARISION:

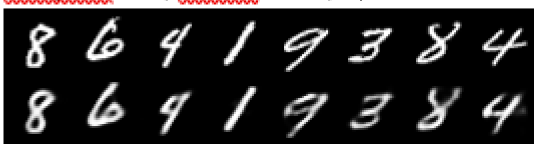
- VAE

There are all generated images by VAE. We can see, after 50 epoch, MNIST are generated much better than CIFAR10.

n_layers = 2, z_dim = 32, epoch 50

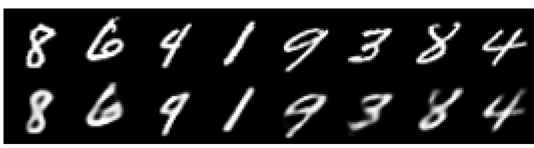


n_layers = 2, z_dim = 50, epoch 50



n_layers = 2, z_dim = 64, epoch 49

n_layers = 2, z_dim = 100, epoch 50



n_layers = 3, z_dim = 32, epoch 50

n_layers = 3, z_dim = 50, epoch 50

n_layers = 3, z_dim = 64, epoch 50

n_layers = 3, z_dim = 100, epoch 50

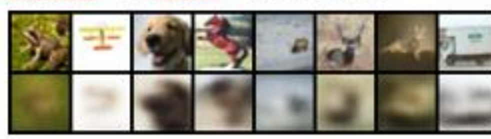
n_layers = 2, z_dim = 32, epoch 50

n_layers = 2, z_dim = 50, epoch 50

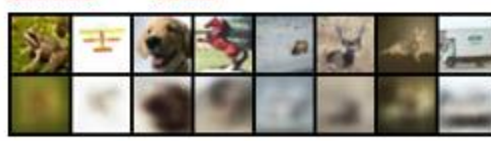
n_layers = 2, z_dim = 64, epoch 50

n_layers = 2, z_dim = 100, epoch 50

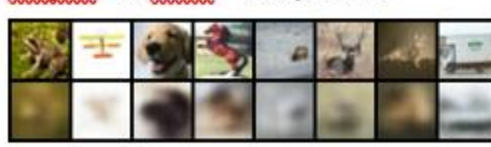
n_layers = 3, z_dim = 32, epoch 50



n_layers = 3, z_dim = 50, epoch 50

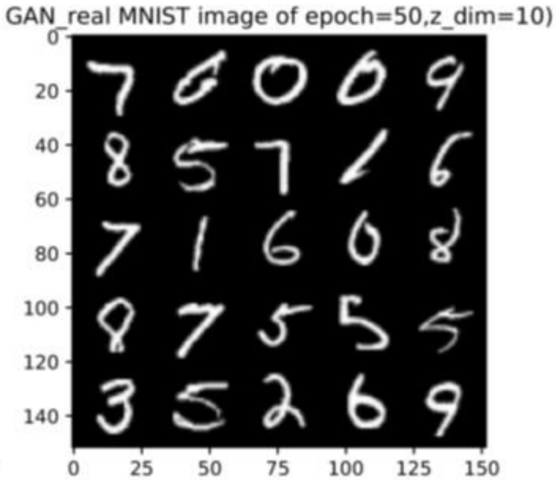


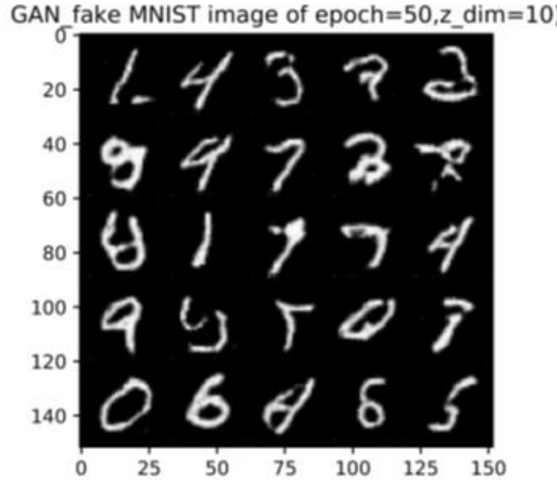
n_layers = 3, z_dim = 64, epoch 50

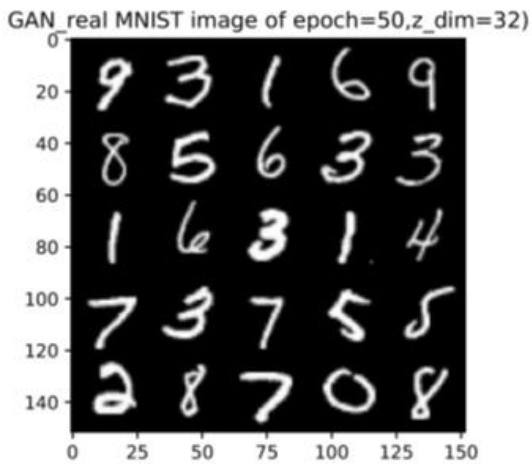


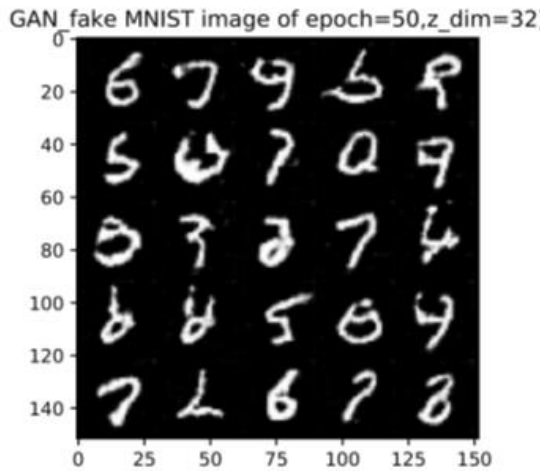
n_layers = 3, z_dim = 100, epoch 50

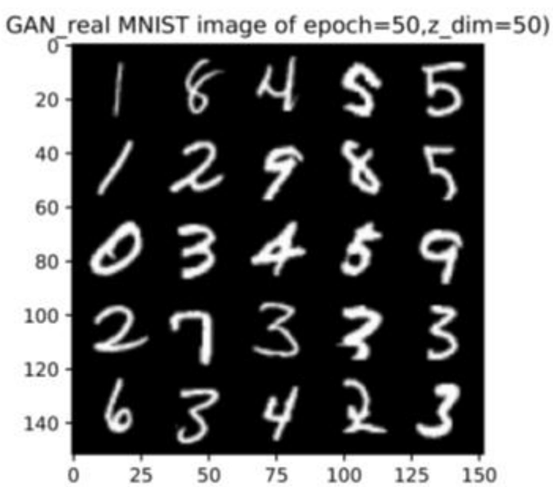
GAN:

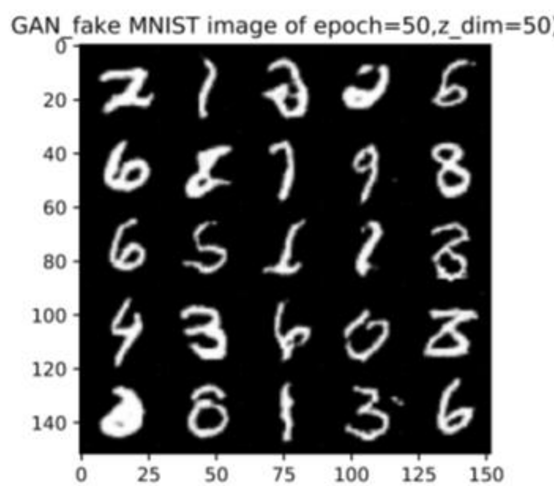

GAN_real MNIST image of epoch=50,z_dim=10)

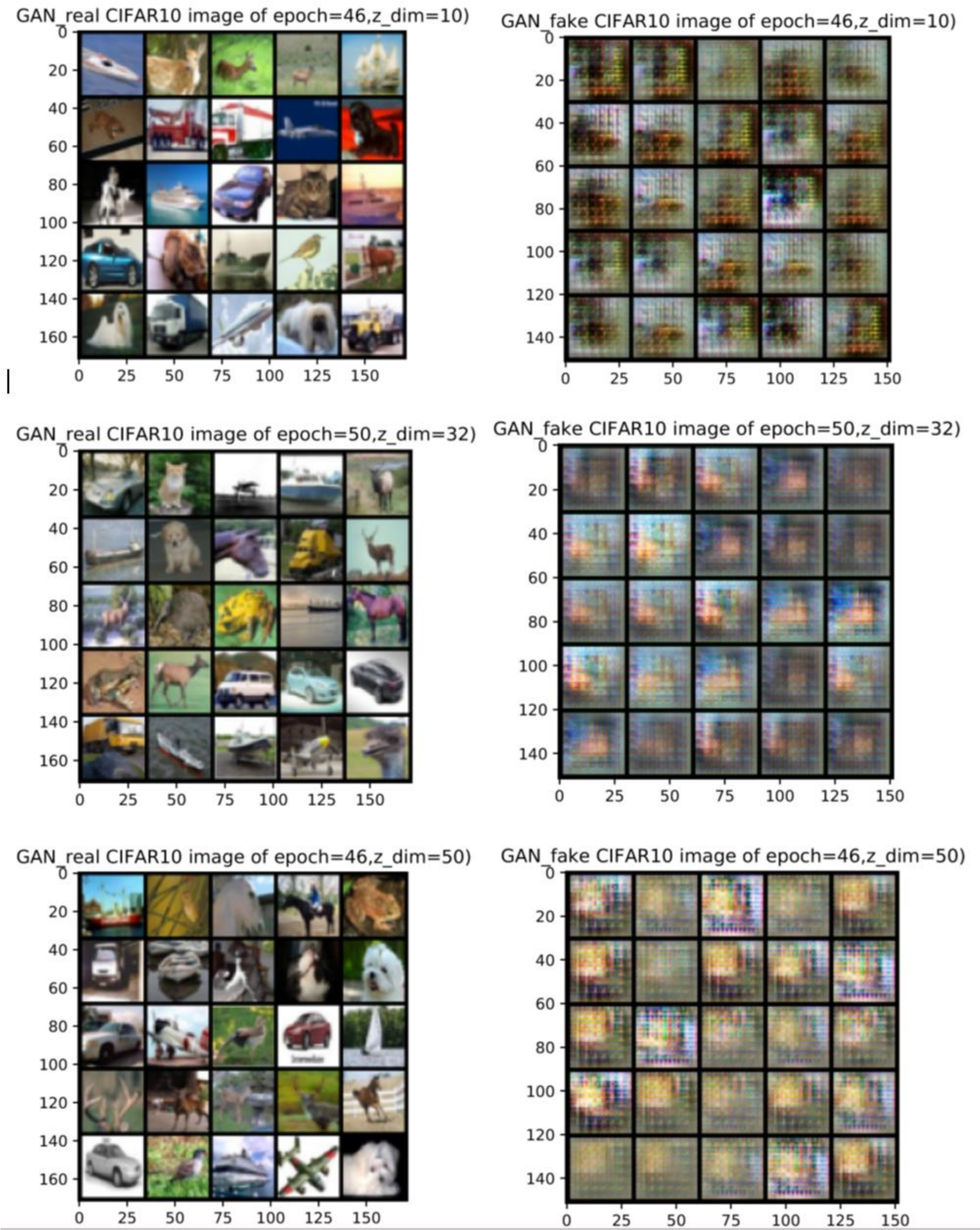

GAN_fake MNIST image of epoch=50,z_dim=10)


GAN_real MNIST image of epoch=50,z_dim=32)


GAN_fake MNIST image of epoch=50,z_dim=32)


GAN_real MNIST image of epoch=50,z_dim=50)


GAN_fake MNIST image of epoch=50,z_dim=50)


GAN_real CIFAR10 image of epoch=46,z_dim=10)
GAN_fake CIFAR10 image of epoch=46,z_dim=10)
GAN_real CIFAR10 image of epoch=50,z_dim=32)
GAN_fake CIFAR10 image of epoch=50,z_dim=32)
GAN_real CIFAR10 image of epoch=46,z_dim=50)
GAN_fake CIFAR10 image of epoch=46,z_dim=50)

WGAN:

In conclusion, when n_epochs=50, WGAN can generate the best images than VAE and GAN.

MNIST are easier to generate than CIFAR10