Experiments of VAE-based image reconstruction

Benshun Yin
No. 019034910033

***Abstract*—Variational autoencoder (VAE) is a type of generative model that can learn a model to approximate the distribution of input data. This project implement image reconstruction using VAE on MNIST and CIFAR10 dataset and conduct some experiments about the loss function, the learning rate and the dimension of latent vector.**

I. Variational Autoencoder

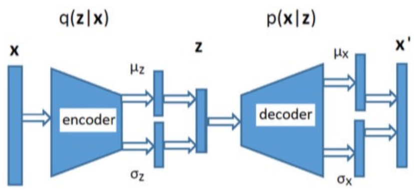


Fig. 1: Framework of variational autoencoder

As shown in Fig. 1, variational autoencoder (VAE) [1] includes a encoder and a decoder. The encoder transforms the input into a Gaussian distribution with mean value μ_z and variance σ_z. Using μ_z and σ_z, the latent vector z is obtained by sampling from the Gaussian distribution. Finally, the decoder estimate the μ_z and σ_z, and thus reconstruct the input data.

II. Experiments

A. Loss Function

This experiment uses three loss functions as follows. Denote x and $\hat{x}$ as the original image and the reconstructed image respectively.

- Case 1: Loss=cross entropy$(x, \hat{x})$
- Case 2: Loss=$\| x - \hat{x} \|^2$
- Case 3: Loss=cross entropy$(x, \hat{x})$ + $\| x - \hat{x} \|^2$

Fig. 3 shows the training results of VAE with different loss functions on MNIST. The cross entropy (CE) loss of case 3 is lower than the others while the mean squared error (MSE) loss is closed to the smallest.

B. Latent Dimension

Fig. 5 shows the training loss with different dimension of latent vector. Obviously, the larger dimension of latent vector is, the lower loss can be achieved. As shown in Fig. 4, the resolution of reconstructed images become higher with a larger dimension of latent vector.

C. Learning Rate

Fig. 6 shows the training process with different learning rate. It obviously that the training speed of VAE with the smallest learning rate is slower than others and the training process with learning rate set as 0.0005 is more efficient.

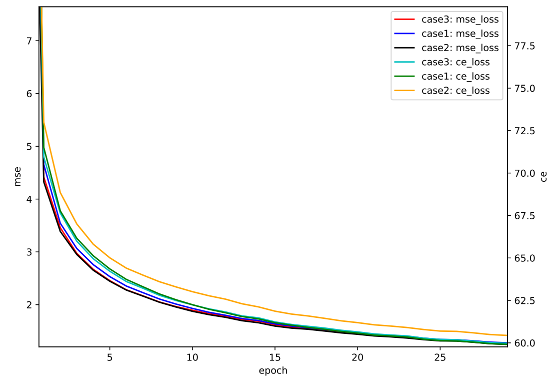
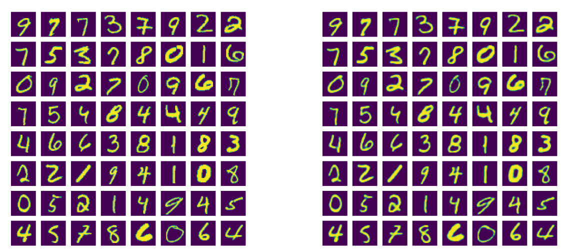


Fig. 2: Training process with different loss functions



(a) (b)

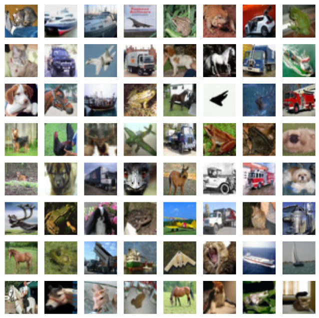
Fig. 3: Test result of VAE trained on MNIST dataset.

D. Feature Visualization

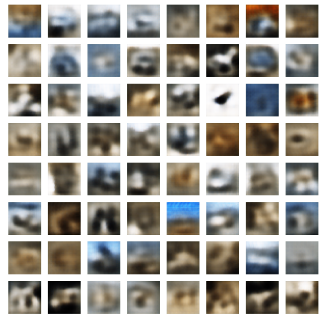
As shown in Fig. 7 and Fig. 8, the pca components of the encoder output and the input data are very similar. Using tSNE, we can see that the distance between different categories is almost unchanged, which is different from classifier.

References

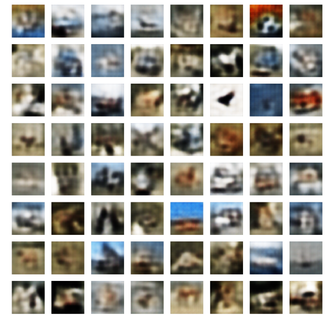
[1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.



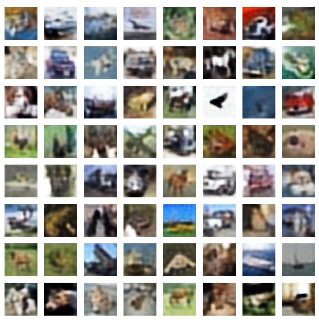
(a)



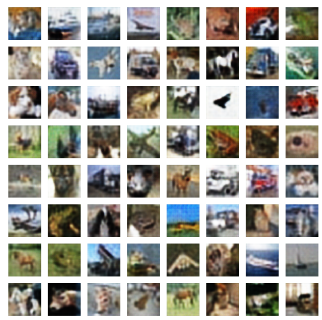
(b)



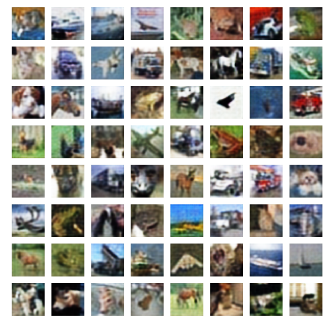
(c)



(d)



(e)



(f)

Fig. 4: Test result of VAE with different dimension of latent vector. (a): original images (b)-(f): reconstructed images with latent dimension set as 32, 64, 128, 256, 512, respectively.

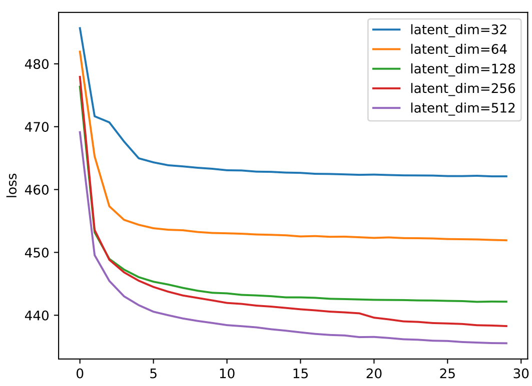


Fig. 5: Training process with different dimension of latent vector

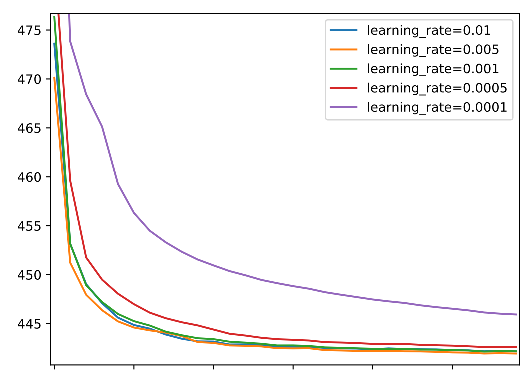


Fig. 6: Training process with different learning rate

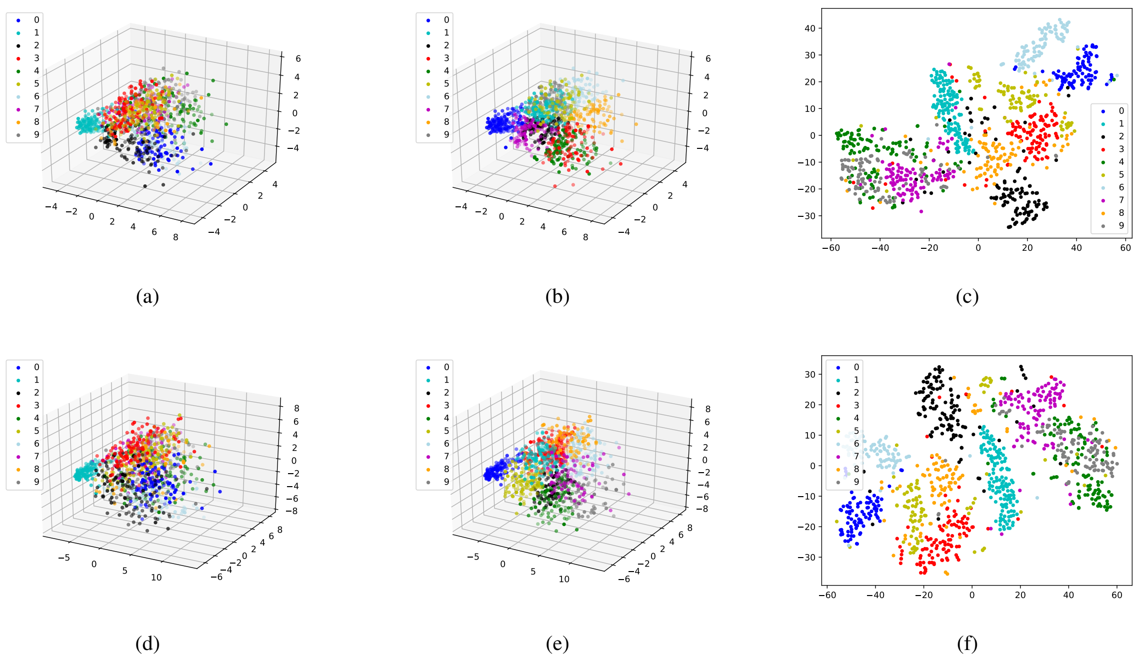


Fig. 7: Feature visualization of MNIST dataset using PCA and tSNE. (a): first three PCA components of raw data, (b): k-means clustering result of the first three PCA components of raw data, (c): tSNE of raw data, (d): first three PCA components of encoder output, (e): k-means clustering result of the first three PCA components of encoder output, (f): tSNE of encoder output.

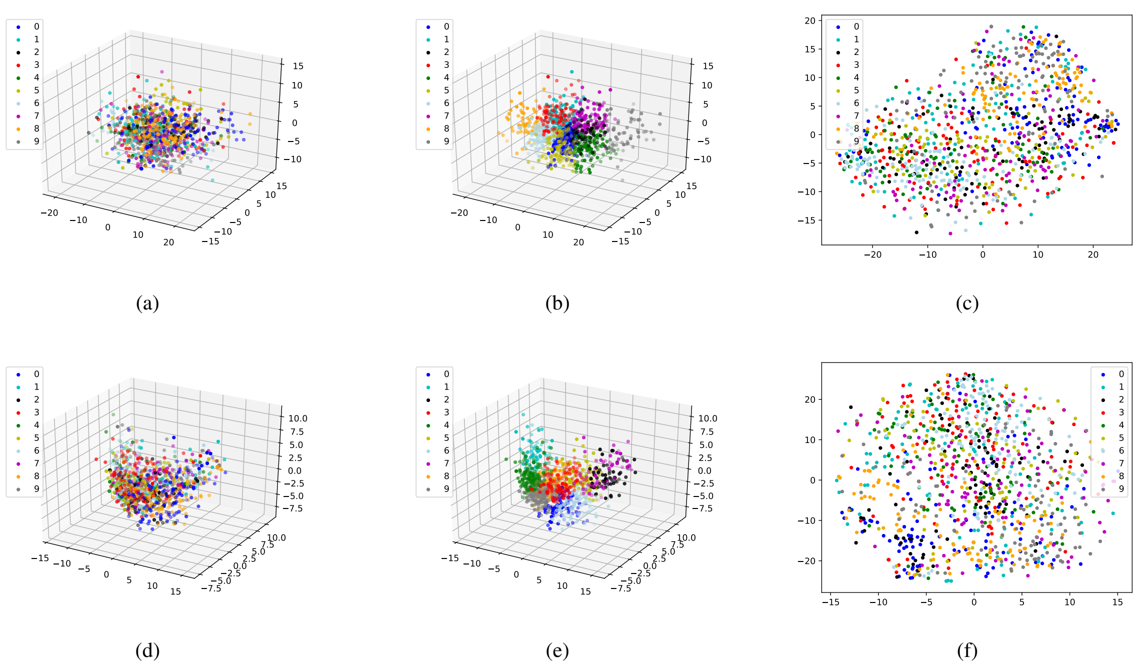


Fig. 8: Feature visualization of CIFAR10 dataset using PCA and tSNE. (a): first three PCA components of raw data, (b): k-means clustering result of the first three PCA components of raw data, (c): tSNE of raw data, (d): first three PCA components of encoder output, (e): k-means clustering result of the first three PCA components of encoder output, (f): tSNE of encoder output.