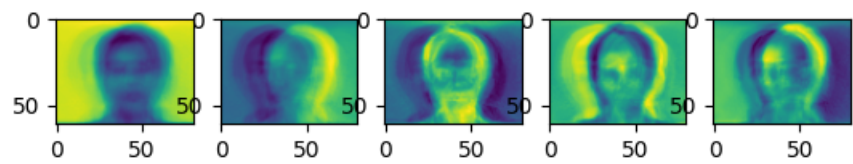
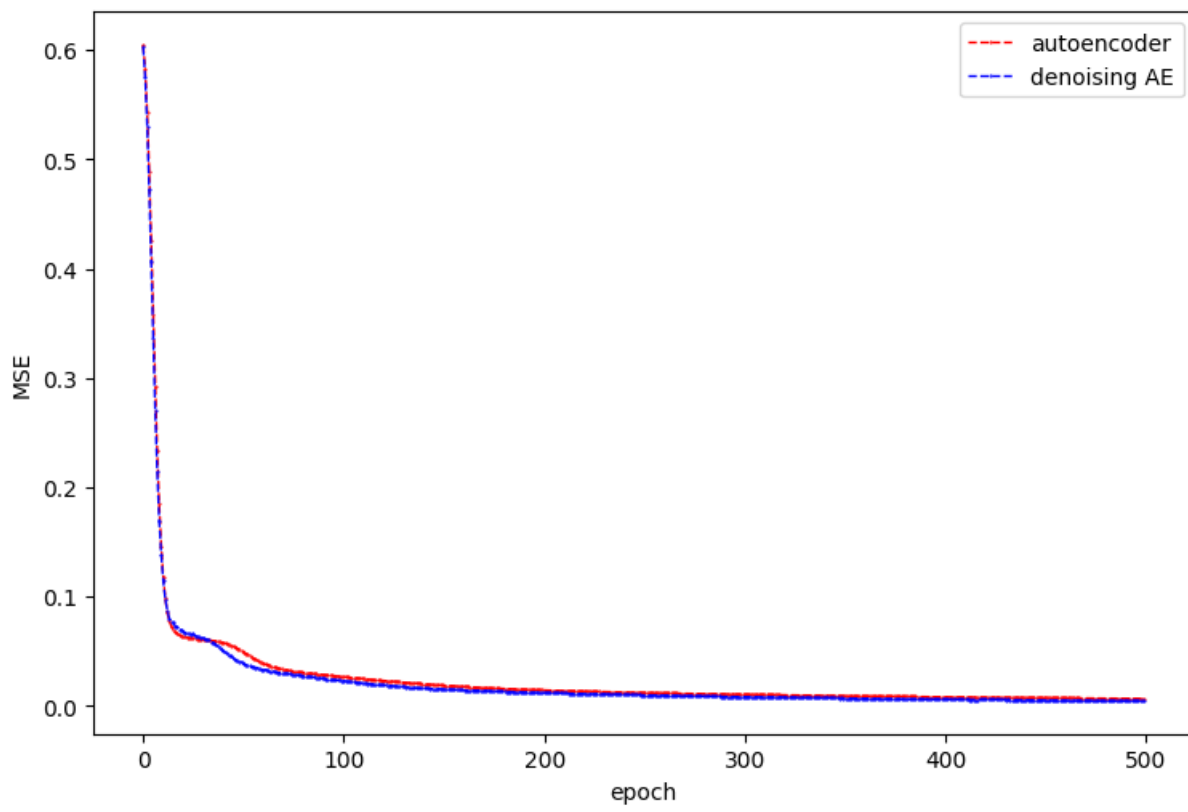


Programming Part Report

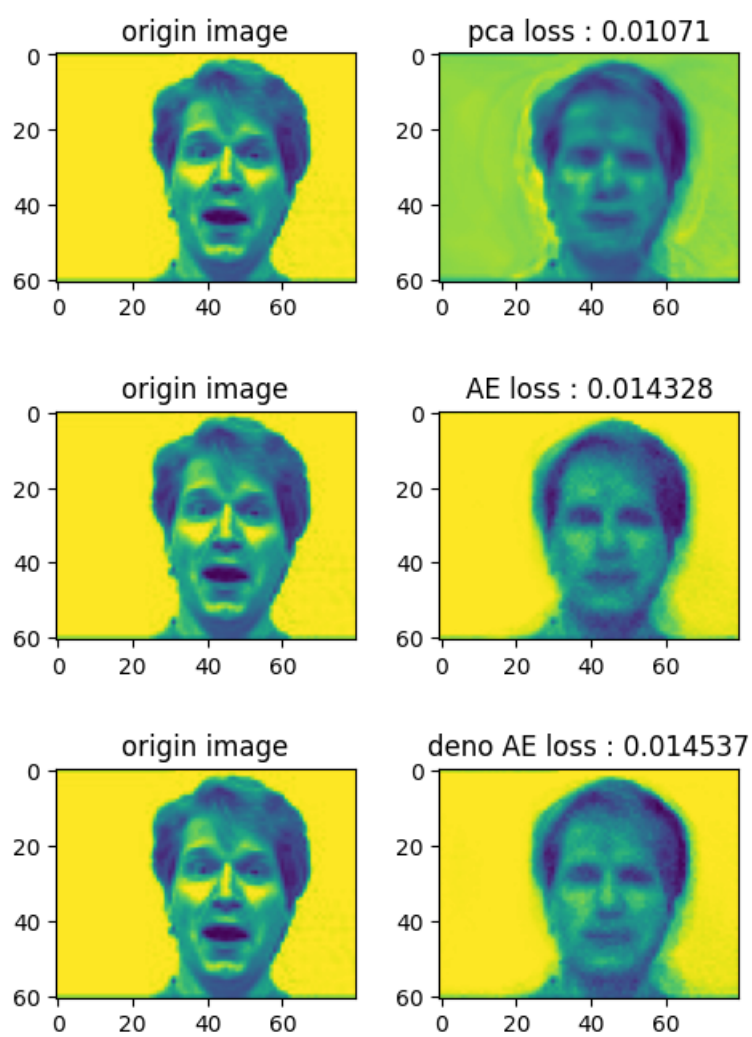
(a)



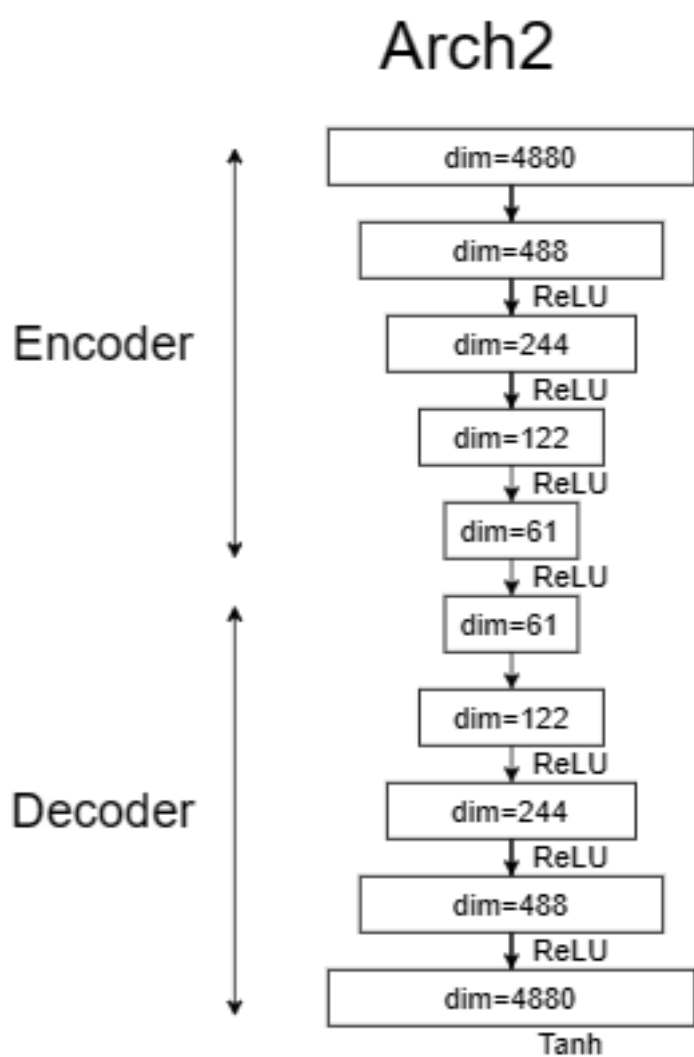
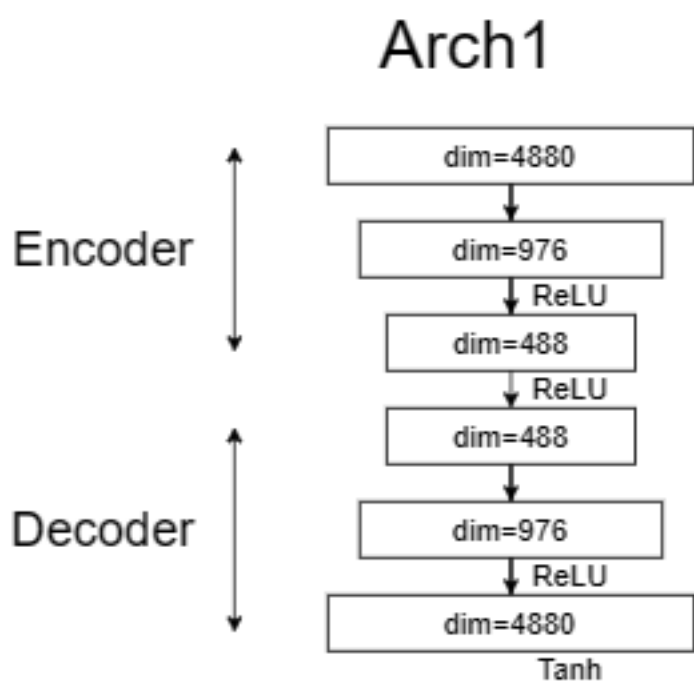
(b)

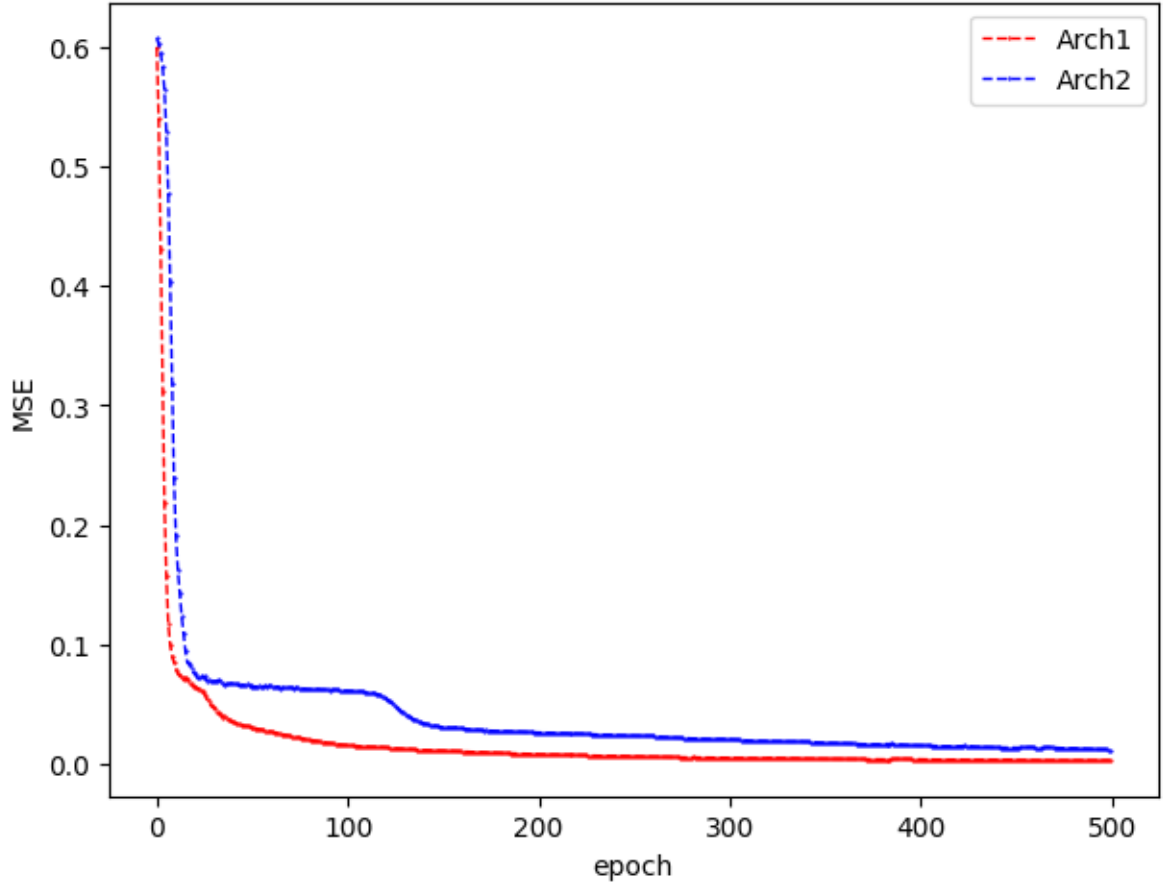


(c)



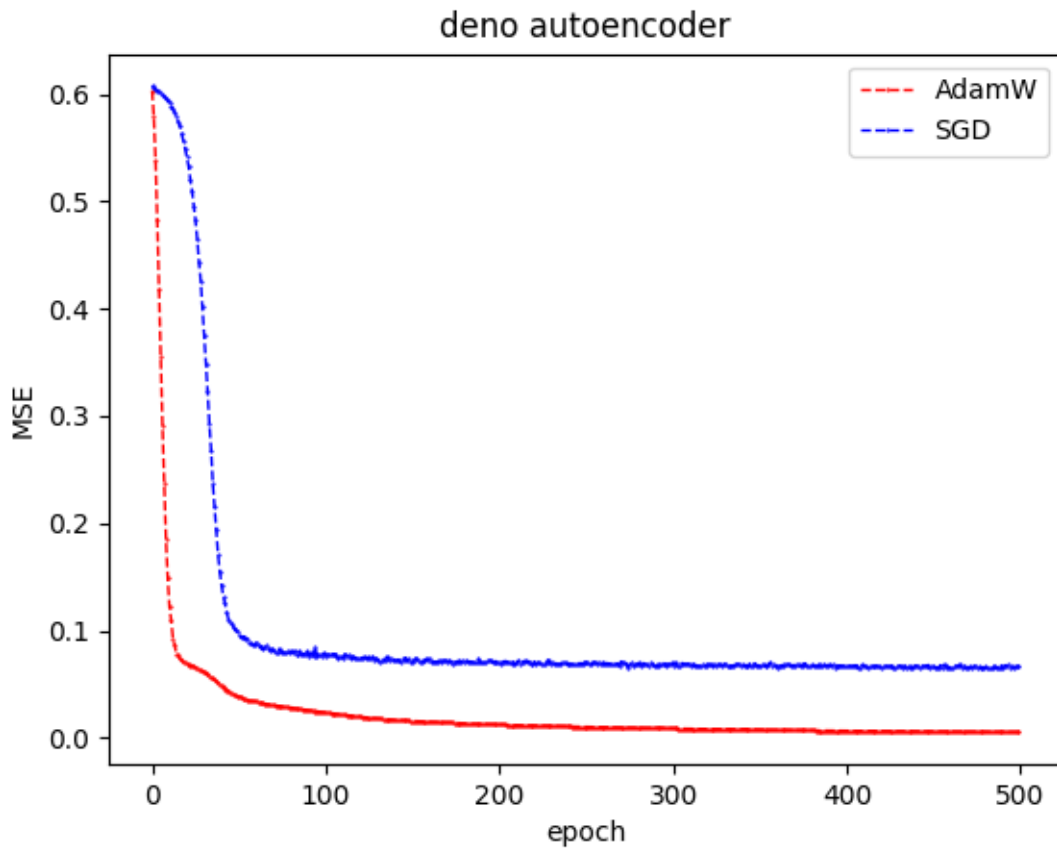
(d)





The two architectures are trained under epoch:500, batchSize:45, learning rates:1e-4 and optimizer:AdamW. Observing the training curve, I find the curve of Arch1 is generally smaller than the curve of Arch2, that is, the MSE loss of Arch1 is smaller than the one of Arch2. Also, when it comes to reconstruction loss, the loss of Arch1 is about 0.01413 while the loss of Arch2 is about 0.01783. Based on these results, I think a shallow but fatter denoising autoencoder may outperform a deeper but thinner one. This might because a fatter network can preserve more information when decoding and a deeper network may be more difficult to optimize.

(e)



I tried two optimizers, one is vanilla SGD and another one is AdamW. Based on the training curve, I find AdamW can converge slightly faster than SGD. However, the overall performance of AdamW defeat the one of SGD totally. The result may because AdamW can dynamically adjust learning rates and use momentum to improve training while vanilla SGD simply train the model.