

Deep Learning Lab 4 - Conditional VAE for Video Prediction

110550088 李杰穎

April 14, 2025

1 Introduction

In this homework, we aim to write and train a variational autoencoder (VAE) to predict the next video frame with given previous frame and next pose image. I train the network with different KL annealing scheduling, including monotonic, cyclical and no annealing. I found that cyclical annealing can lead to better validation PSNR. This is coincide with findings in [1]. The teacher forcing strategy also stabilize the early training process. I achieve ... of PSNR on validation set and ... of PNSR on public test set.

2 Implementation Details

2.1 Training protocol

2.2 Testing protocol

2.3 Reparameterization tricks

When training VAE, in order to make the gradient can be back-propagate to the encoder properly, we need to employ reparameterization trick. Specifically, instead of directly sample latent vector z from $\mathcal{N}(\mu, \sigma^2)$, we first sample a noise n from standard normal distribution $\mathcal{N}(0, 1)$ then we scale and shift the noise n with $z = \mu + n\sigma$. In this way, the gradient can be flowed back to the encoder, whose output is μ and σ^2 . Furthermore, instead of directly output σ^2 , we often make the encoder output the log variance, $\log \sigma^2$, to stabilize the training process. After this modification, the reparameterization becomes $z = \mu + n \exp(\frac{\log \sigma^2}{2})$. The code implementation is as follow:

Code 1: **Implementation of VAE reparameterization trick.**

```
1 def reparameterize(self, mu, logvar):
2     return mu + torch.exp(logvar / 2) *
    ↪ torch.randn_like(mu)
```

2.4 Teacher forcing strategy

The teacher forcing strategy first appears in the fields of reinforcement learning (RL). When the model are asked to generate a sequential data, we often take the previous output as the input to generate the next output. However, when the model's performance are not good, the error will

accumulate and thus harming the training process. Therefore, instead of using the previous output as input, teacher forcing injecting the ground truth of previous step as the next input. In this way, we can stabilize the model training, avoid the degenerate causing by error accumulation. However, if we always employ teacher forcing, the model wouldn't know how to fix the accumulated error, thus, after the model is capable to output reasonable output, we can turn off teacher forcing to make the model more robust. Thus increasing the testing performance where ground truth data is not available.

2.5 KL annealing ratio

3 Analysis & Discussion

References

- [1] Hao Fu et al. "Cyclical Annealing Schedule: A Simple Approach to Mitigating KL Vanishing". In: ed. by Jill Burstein, Christy Doran, and Thamar Solorio. 2019, pp. 240-250. DOI: 10.18653/v1/n19-1021. URL: <https://doi.org/10.18653/v1/n19-1021>.