**Lab4: Conditional VAE for Video Prediction**

**系級:智能系統 學號:312581006 姓名:張宸瑋**

1. Introduction

這次lab要實作基於VAE的模型的條件式視頻預測，在這次的實作主要用到兩個在Paper[1][2]提出的概念，第一個是透過GAN模型，並且使用遇訓練的姿勢估計網路生成的姿勢圖像，以及前一幀，來使模型輸出具有一定質量的下一幀，而第二個是透過基於VAE的模型，並且結合LSTM以及RNN的方式預測未來幀，通過使用兩個參考幀，使模型有能力預測接下來的未來幀。

以下是本次實驗模型的主要架構，左邊的圖為訓練模型的過程，大致如下，Posterior Predictor透過當前幀輸入到frame encoder以及骨架圖像輸入到pose encoder的特徵圖，作為輸入的生成分布。而Generator以當前標籤，跟生成的上一幀畫面，以及由Posterior Predictor預測的分佈所抽樣的噪音作為輸入，生成當前的畫面。而在test的部分噪音會直接從先驗分佈作抽樣。

|  |
| --- |
| 模型架構 |

而這次我們也需要在KL損失函數中添加一個可變的權重beta，來調整KL loss影響的比例，透過不同的策略Cyclical、Monotonic、Without KL annealing strategy來調整beta，並且訓練模型以及分析結果。

|  |
| --- |
|  |
| 週期性策略以及單調策略示意圖 |

1. Implementation details
2. How do you write your training protocol

|  |
| --- |
|  |
| Training Stage |
|  |
| Training One Step  解說:  這段主要是訓練模型的程式碼，在訓練時每個batch的資料會被當作參數，輸入到train\_one\_step函數裡面，並且透過adapt\_TeacherForcing判斷Frame Encoding的輸入是要用Generator生成的還是原本的前一個畫面，並且根據self.train\_vi\_len來決定要預測的影片序列的長度，並且依序將每個畫面輸出。 |

1. How do you implement reparameterization tricks

|  |
| --- |
|  |
| reparameterization tricks |

解說:

這段程式碼主要是透過重新參數化的技巧，來實現反向傳播的計算。

1. How do you set your teacher forcing strategy

|  |
| --- |
|  |
| teacher forcing strategy |

解說:

這段程式碼主要是在描述我的teacher forcing strategy，一開始會先判斷現在的epoch是否小於設定的self.args.tfr\_sde，如果還沒ratio維持1，當epoch大於self.args.tfr\_sde後，每五個epoch透過self.args.tfr\_d\_step依序更新ratio。

|  |
| --- |
| 1. How do you set your kl annealing ratio |
| |  | | --- | |  | | kl annealing ratio | |

解說:

這段程式碼主要是實現KL annealing，透過逐漸增加權重，來幫助訓練模型，程式碼會根據self.kl\_anneal\_type來決定權重的更新是單一性還是週期性。

1. Analysis & Discussion
   * 1. Plot Teacher forcing ratio

|  |
| --- |
|  |
| Teacher forcing ratio |

討論:

我認為透過老師在上課所講的策略，明顯的提升了我們訓練的結果，因為在訓練初期，如果我們直接使用預測的畫面來當作輸入，可能會導致後面模型沒辦法生成正確的序列，導致生成序列的不穩定，因此透過調整ratio，並讓它線性遞減，可以讓訓練前期，透過原本的圖像來當作輸入，使模型學習到正確的重建損失。

* + 1. Plot the loss curve while training with different settings

|  |
| --- |
|  |
| With KL annealing (Cyclical) |
|  |
| With KL annealing (Monotonic) |
|  |
| Without KL annealing |

討論:

在這裡我們可以發現，當沒有採用KL退火策略的時候，代表模型相較於前兩種策略來說，會更關注KL散度的損失，因此我們可以看到圖中沒有採用KL退火策略的KL loss一開始是相對較小的，但這可能會導致模型生成的樣本的品質受到限制，而週期性KL退火策略，想對於單一性來說，因為它交替調整KL散度的權重，因此可以在訓練的時候平衡重建損失以及KL散度損失，而單一性策略，因為在後期更關注KL散度的損失，可能導致訓練的早期就限制了模型的生成能力。

* + 1. Plot the PSNR-per frame diagram in validation dataset

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | | PSNR-per frame diagram in validation dataset | |

|  |  |  |
| --- | --- | --- |
| * + 1. Derivate conditional VAE formula  |  | | --- | |  | | conditional VAE formula | |

1. Discussion

這次的實驗我認為算是目前遇到最困難的，除了要讀熟兩篇paper以外，在訓練的部分也有許多細節需要注意，而當中讓我較有印象的部分是，當時把train\_loader的shuffle設為True，是希望提升模型的泛化能力，但效果好像不是如我預期的那樣，好幾次前幾個epoch在計算loss時，很容易就變成nan，這裡我猜想可能是因為dataset的資料是有時間先後關係的，因此打亂造成了訓練的不穩定性。而在訓練的過程中，也讓我了解到要訓練像VAE這種網路是相對複雜的，因為它要優化的，不是單一個損失函數，因此在超參數的調整方面，是需要非常多的時間去做調整，才能達到理想的結果。