LAB 5 CVAE For Video Prediction

謝宇恆 411551022

1. Introduction

在本實驗中，我們的任務是做出一個Conditional VAE，用於視頻預測。這個模型可以根據過去的frame進行預測。具體來說，當我們將上一個frame 輸入encoder時，它會生成一個latent code 。接著，我們會從fixed prior中採樣 。最後，我們將encoder的輸出和 以及動作和位置 (即Condition) 一起作為decoder的輸入，期望輸出的影格為下一個影格 。訓練和生成的過程可以參考圖1。

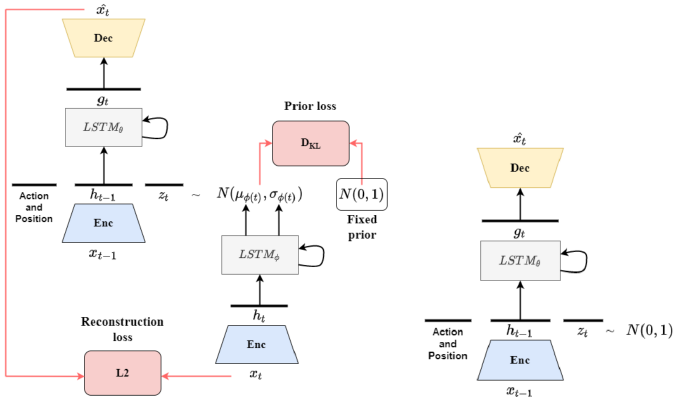


圖 1. Training Procedure (左)，Generation Procedure (右)

2. Derivation of CVAE

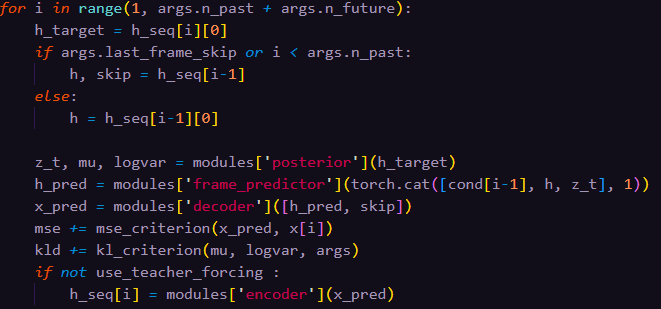
*The goal of CVAE is to minimize*

*since*

3. Implementation details:

* Train\_fixed\_prior.py

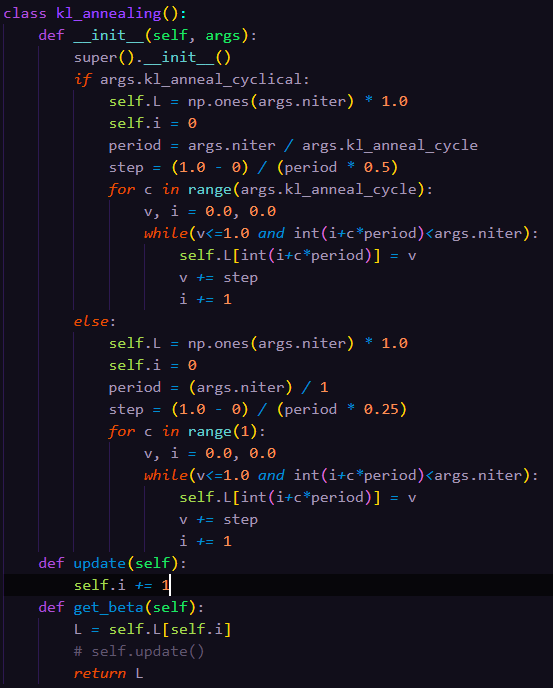
在這邊我們會利用前兩個Frame作為條件，預測接下來10個frame。



* KL\_annealing:

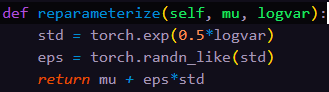
KL\_annealing是為了動態調整KL Divergence的權重以避免模型在學習時只專注於降低KL Divergence。

在本次實驗種我們要實作兩種不同的KL\_annealing分別是Monontic and cyclical。



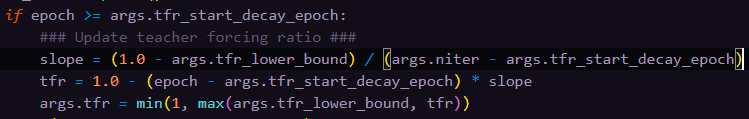
* Reparameterize Trick in LSTM:

為了訓練end-to-end的model，我們使用Reparameterize Trick。在此實驗中，我們使用log variance作為模型的輸出。



* Describe Teacher forcing ratio

Teacher forcing ratio衰減的程式如下：我使用線性方式進行衰減，因此當到達最後一個訓練 epoch 時，Teacher forcing ratio會衰減至其下限。



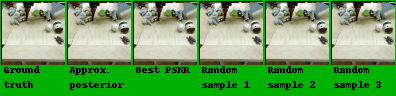
在訓練模型時，使用teacher forcing會將真實的frame輸入到encoder中，以獲取latent code，然後預測下一個時間步的frame。如果不使用teacher forcing，則會將預測的frame輸入到encodder中，然後預測下一個時間的frame，以此類推直到訓練結束。使用teacher forcing的好處在於讓模型學習正確的輸入以預測下一個frame。缺點則是我們的目標是使用預測的frame來預測下一個時間步的frame，而輸入數據可能會存在一些偏差。因此，下一個時間步預測的frame也可能會存在一些偏差。

為了避免這種情況，我們使用teacher forcing比率衰減策略。一開始，讓模型學習真實的輸入以預測下一個時間的frame。在幾個 epoch 後，逐漸降低teacher forcing比率，讓模型學習有偏差的輸入以預測下一個時間步的影格。

4. Results and discussion:

4.1 Show your results of video prediction

下圖為video prediction結果的screenshot。

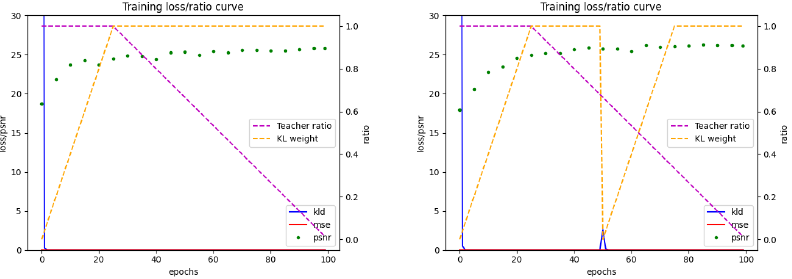


下圖為prediction at each time step。在圖中最前面兩張為groud truth，接下來10張圖是它們predict出來的結果



4.2 Plot the KL loss and PSNR curves during training

* epoch: 100
* beginning tfr: 1.0
* tfr lower bound: 0.0
* tfr start decay epoch
* learning rate: 0.002



4.3 Discuss the results according to your setting of teacher forcing ratio, KL weight,

and learning rate.

4.3.1 KL\_annealing:

在訓練的初期，我們希望模型能夠專注於MSE。另外一個原因是 KL Cost相對於MSE過高（約為150），這可能導致Cost被 KL Cost主導。為了解決這個問題，我們使用週期性的 KL Cost調整方法。如圖5所示，第50個訓練 epoch 的 KL 損失突然增加到約2.6。儘管如此，該 KL 損失值仍然低於第1個 epoch 的 KL 損失值。通過這種方式，我們可以獲得比單調方式更好的PSNR，如下表所示。

|  |  |  |
| --- | --- | --- |
|  | Monotonic | Cyclic |
| PSNR | 25.75 | 26.03 |

4.3.2 Teacher Forcing Ratio:

在訓練模型時，我們不能在一開始就將Teacher forcing ratio設置為0。這是因為預測的輸入數據可能存在偏差或模糊，這將影響我們模型對正確frame的重建。為了避免這種情況，我將Teacher forcing ratio初始設為1，並在前25個訓練 epochs 線性地衰減到0。這樣可以讓模型在初始階段學習使用真實輸入數據進行正確的重建，然後再學習在具有偏差的輸入數據下進行重建。