**Lab4: Conditional VAE for Video Prediction**

Student id / name: A113599 / 楊淨富

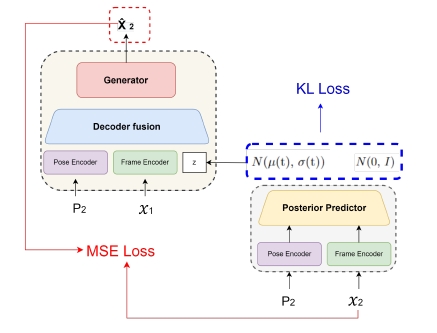
**I. Introduction:**

實作一個conditional VAE model來預測影像，即根據上一幀的資訊預測下一幀。並在最後的testing data上，給予第一幀，去預測接下來的629幀。並和label frame(ground truth)計算出PSNR。

在VAE中，主要是希望學習能夠生成input data的潛在變數表示(latent variable representation)，並利用reparameterization的技巧使得能從中隨機sample來進行back propagation。

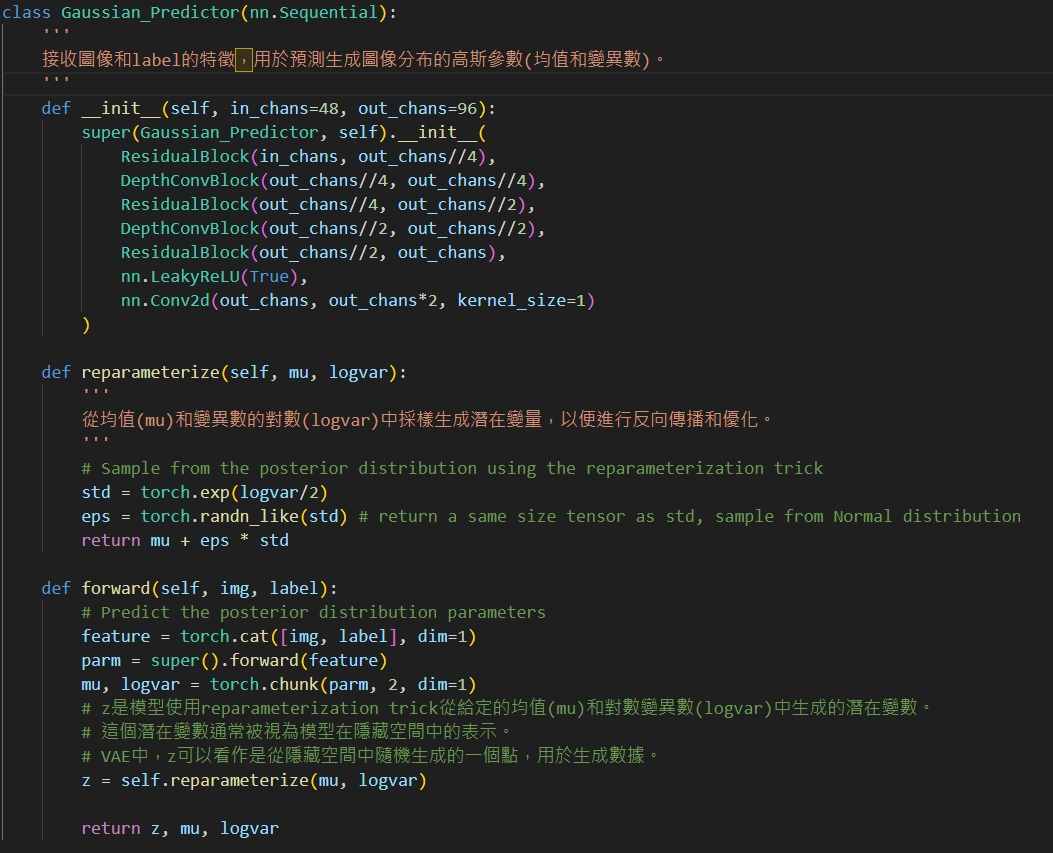
**II. Implementation Details:**

1. How do you write your training protocol



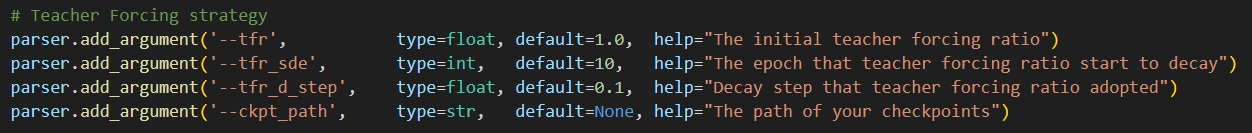
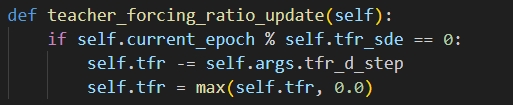
* 總共會有16幀作為training video sequence，即x1, x2, …, x16。以及16幀作為conditional signals，即P1, P2, …, P16。
* 其中第一幀是用來預測未來15幀的。
* 假設目前產生的下一幀為 x̂2，則會需要
  + Pose image (P2): 即label feature。
  + Last generated frame: x1
  + Z sample from prior distribution
    - Sample一個Gaussian distribution的tensor作為noise
* 產出下一幀後，就可以計算mse，會根據adapt\_TeacherForcing來決定是否用ground truth(即x2)或者用自己先前所預測的前一幀來進行比較，即:
  + mse( x̂2, x2) or  mse( x̂2, x1)
* 但loss = mse + kld \* beta，其中:
  + kld為KL divergence，是一種用來衡量兩個機率分佈間差異的指標。它衡量了兩個分布間的相對熵，即一個分布相對於另一個分布的資訊損失。
* 算完loss後就進行back propagrion和更新參數:
  + loss.backward()
  + self.optimizer\_step()
    - 會用到nn.utils.clip\_grad\_norm\_(self.parameters(), 1.0)，主要是用於對模型的梯度進行裁減，防止梯度爆炸的情況。透過將梯度的大小限制在一個合理的範圍內，來緩解這個問題。

1. How do you implement reparameterization tricks

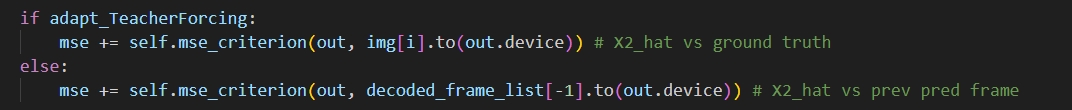


* 在一些生成模型中，需要對潛在變數（latent variables）進行sample，然後使用這些樣本生成數據。但是，直接從潛在分布中sample可能會使反向傳播變得困難，因為梯度無法有效地通過隨機過程進行傳播。
* Reparameterization（重參數化）:從潛在分布中生成樣本，同時允許對樣本進行反向傳播（backpropagation）以更新模型參數。作法是從一個固定的分布（通常是標準正態分布）中sample，然後進行線性變換和平移，得到我們想要的分布。
* reparameterize()這個function從平均(mu)和對數變異數(logvar)中sample，生成潛在變量(z)。這個過程可以通過將標準常態分布中的隨機樣本（eps）進行線性變換和平移得到。因為這個過程可微，所以梯度可以正確傳播，使訓練過程中可以對模型更新參數。
* 重參數化技巧的核心思想是將隨機取樣的過程分解為兩個部分，其中一部分透過模型參數來表達隨機性，而另一部分是從實際分佈中獲取數值，使得整個生成過程可以與反向傳播結合，實現模型的有效訓練和優化。

1. How do you set your teacher forcing strategy

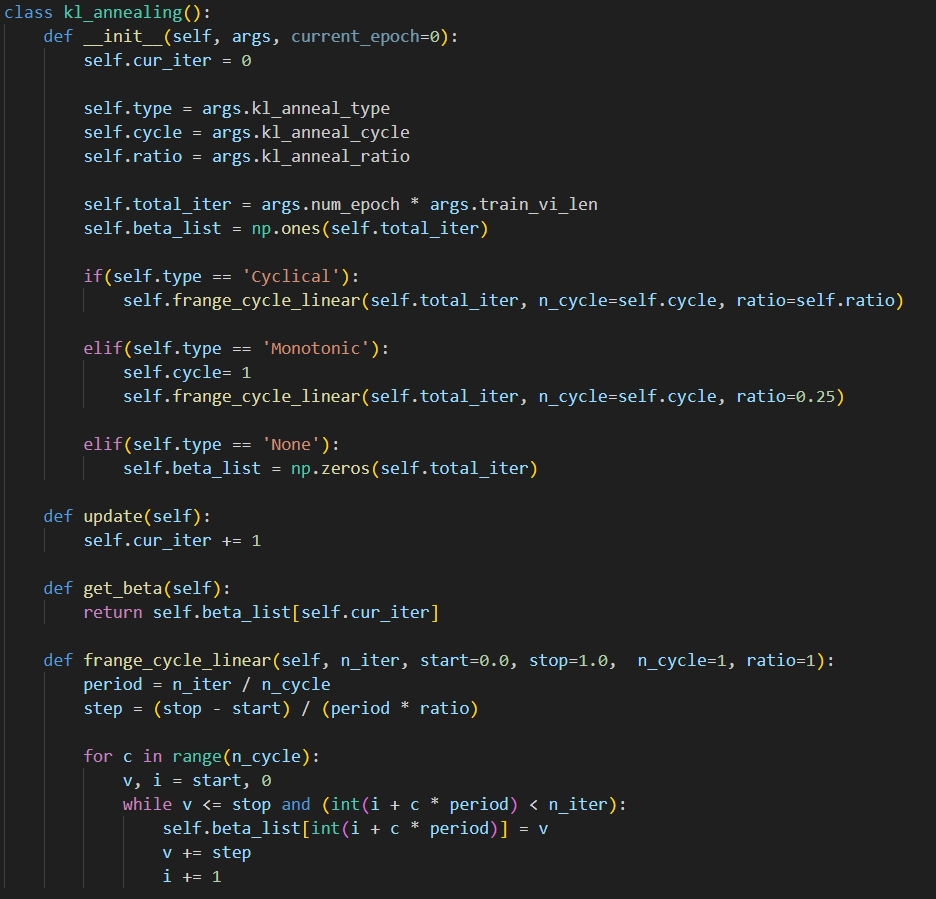


* 一開始的tfr ratio是1.0，即100%使用ground truth。並且每經過tfr\_sde(default = 10)個epoch後，tfr就會減少tfr\_d\_step(default = 0.1)。

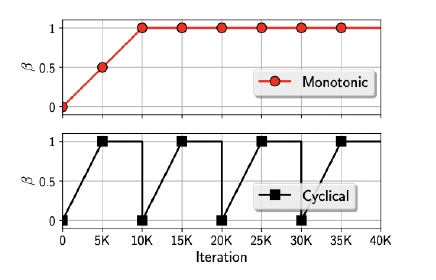


* 如果 adapt\_TeacherForcing 是 True，則使用 ground truth 圖片與生成的圖片計算MSE，即使用真實的輸入來指導模型的訓練。
* 如果 adapt\_TeacherForcing 是 False，則使用先前預測的圖片與生成的圖片計算MSE。不使用真實的輸入，而使用模型自身生成的先前圖片作為輸入，這種策略可以使模型更有能力處理自身生成的結果。

1. How do you set your kl annealing strategy



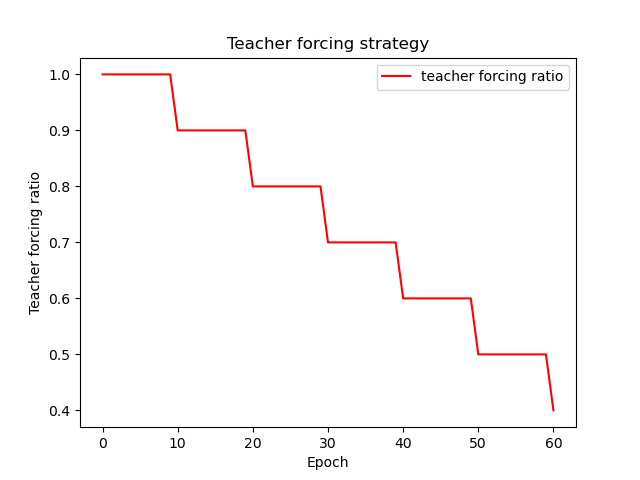
* KL散度(KL divergence)用於評估兩個分布之間的差異。在 VAE 中，希望學習的潛在變量分布（通常是高斯分布）與已知的標準分布（例如，標準高斯分布）之間的 KLD盡量接近零。這意味著學習的潛在變量分布應該盡量接近標準高斯分布，這有助於更好地探索潛在空間。
* 如果在訓練的早期就強制 KLD為零，模型可能會過早地將潛在變量分布縮小到接近標準高斯分布，這可能會導致模型丟失有用的特徵。KL annealing最主要就是避免這種情況。
* KL annealing的想法是在訓練過程中逐步增加 KLD的權重，從而逐漸將潛在變量分布限制在較接近標準高斯分布的範圍內。這可以使模型在早期學習到更多的特徵，然後在訓練後期越來越多地遵循標準高斯分布。這樣的過程有助於更好地平衡生成數據的多樣性和學習有用特徵之間的關係。



* Cyclical: 根據cycle數決定每個「波」，根據ratio決定每個cycle內的每一步(每次period)應該走多少量。
* Monotonic: 全部就是一個cycle，走到1時，後面的nparray全部都是1。

**III. Analysis & Discussion:**

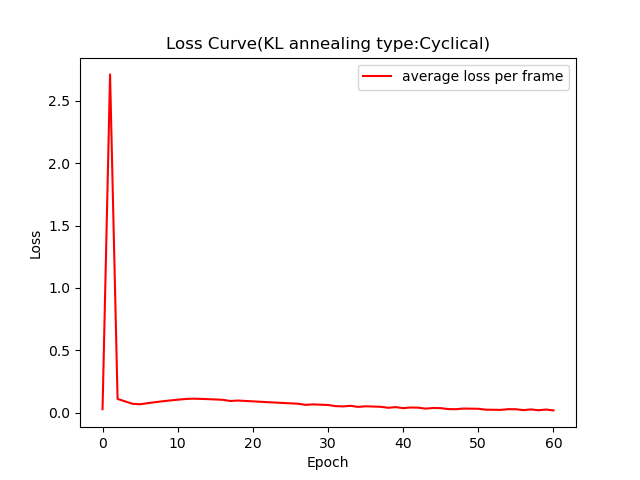
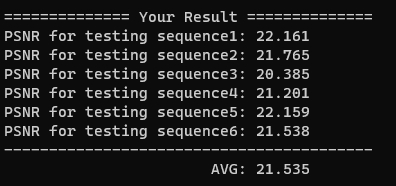
1. Plot Teacher forcing ratio



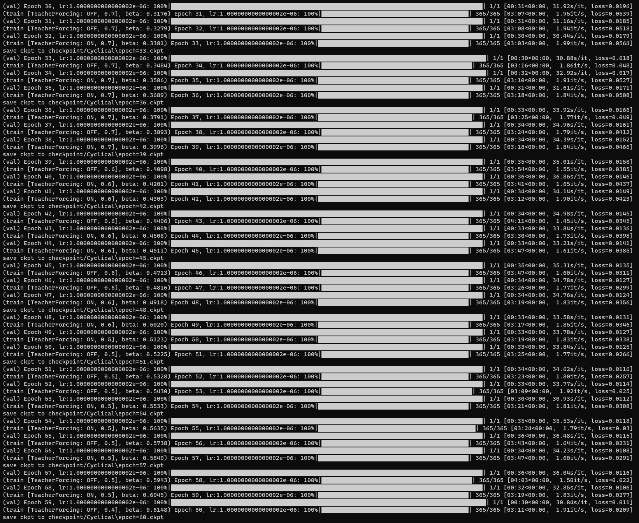
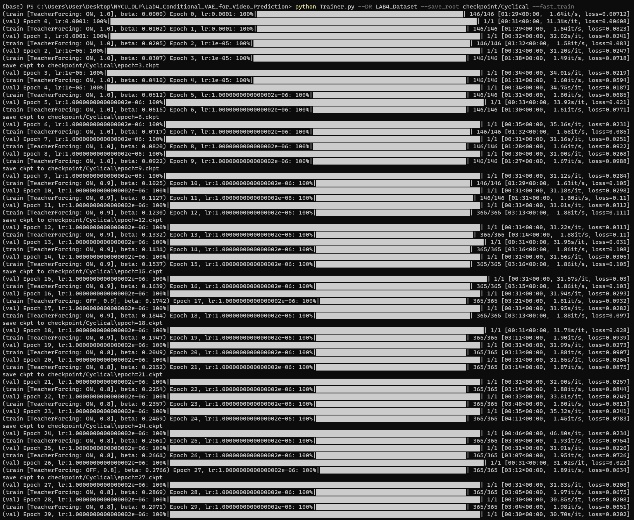
* 我的teacher forcing ratio初始值是1.0，代表100%使用ground truth，然後每個10epochs，則teacher forcing ratio減少0.1。
* Teacher forcing ratio剛開始需要設置比較大的初始值是因為剛開始產出的frame品質比較差，所以拿來計算mse時先以ground truth為主會比較容易讓loss穩定。

1. Plot the loss curve while training with different settings. Analyze the difference between them.

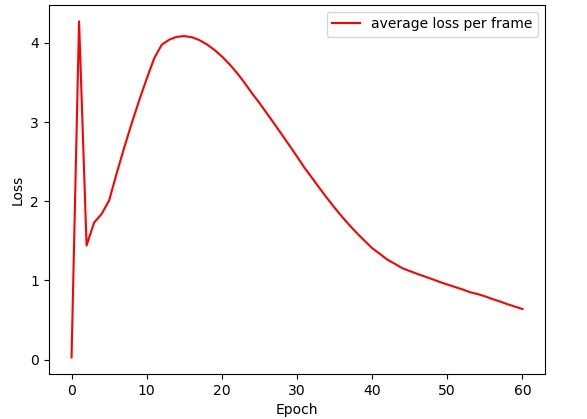
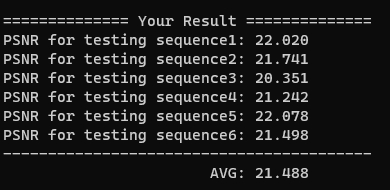
* With KL annealing (Cyclical)

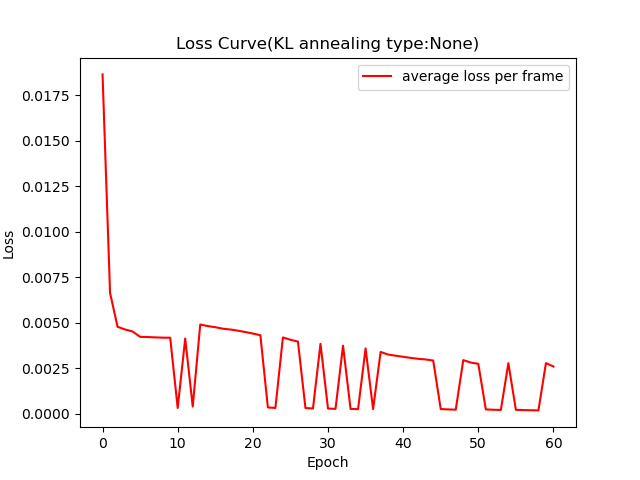
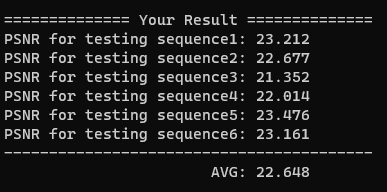
* + 我的loss計算方式是average loss per frame，只有epoch=1時數值很高，理由可能是第一次產出的幀的品質還相當爛，但之後平均每幀的loss便在0.1-0.01左右。



* With KL annealing (Monotonic)

* Without KL annealing

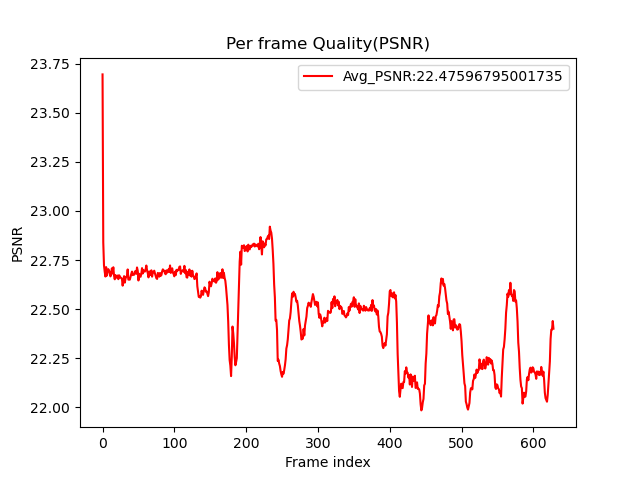
 

* 最好的一次PSNR(Cyclical)



* Anaylze the difference between them
  + 整體而言，Cyclical/ Monotonic/ None三者皆在epoch=1時到達最高點，之後Cyclical是不斷在低點徘徊，Monotonic則又再爬升至一個高點，後面才逐漸穩定減少loss，而None則是有在震盪。就各自的PSNR而言，我的實驗結果反而是Without KL annealing表現最好，而Cyclical和Monotonic兩者差不多。KL annealing並未對我的model產生正面的影響，有可能是我的退火參數(退火速率、最大KLD等等)並未調整好，也有可能是過度限制了model的生成，或者epoch太少。
  + Cyclical KL Annealing在訓練過程中週期性地增加和降低KLD的權重。這可以有助於在不同階段適當地平衡失真和正規化，有助於model更好地適應不同的數據分佈和特徵，讓model更靈活。
  + Monotonic KL Annealing從訓練開始時就增加KLD的權重，但不進行週期性變化。這種策略可能更穩定，但在某些情況下無法適應數據的變化。由於KLD的增加可能會導致模型陷入local minima，可能需要調整相關參數來確保training model的穩定性。

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



* 這張圖的parameters如下:
  + KL annealing type: Cyclical
  + Epoch: 61

1. Derivate conditional VAE formula

