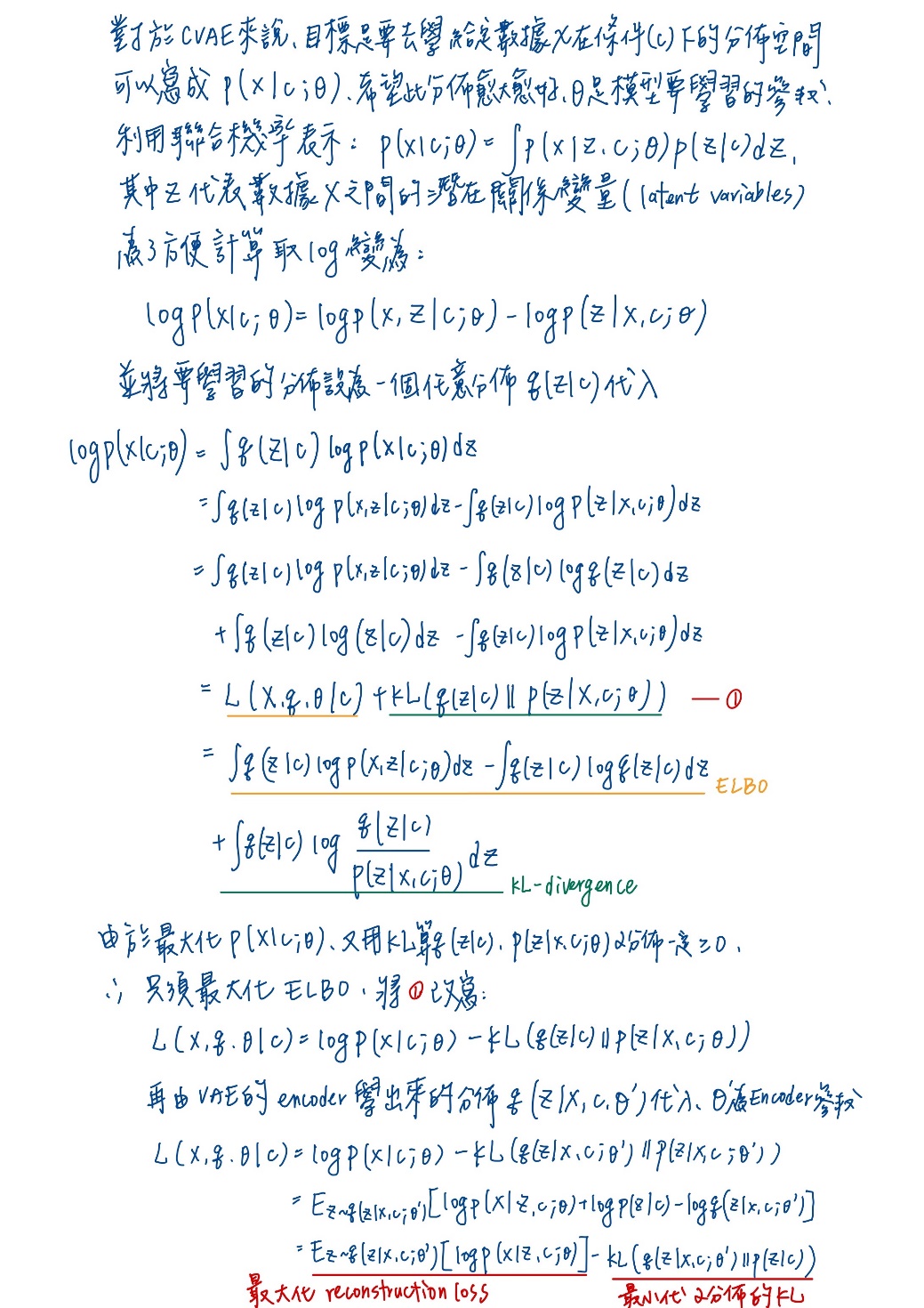
**DLP HW5**

**數據所311554019 宋沛潔**

1. **Introduction:**

這次的作業是使用conditional VAE來實作影片生成的任務。根據已知過去的影片及動作和位置(條件)，利用encoder學出來資料之間的關係分布，再利用encoder輸出、latent vector以及動作和位置（條件）做為input ，去生成未來下一個時間點的影片。

1. **Derivation of CVAE:**
2. **Implementation details**
3. Encoder

使用助教提供的簡化版的vgg64 encoder，包含了五個convolution layer(c1~c5)，由不同數量的vgg\_layer組成。主要是把輸入的照片壓縮成較小的向量，也就是逐漸降維，透過不同convolution layer保留不同的特徵向量。在每個convolution layer都有分別把該層向量存起來，使其可以使用skip，代表在訓練過程中，跳過一些中間層，以解決梯度消失問題。

一張含有 文字 的圖片

自動產生的描述

1. Decoder

使用助教提供的簡化版的vgg64 decoder，包含了五個convolution layer(upc1~ upc5)，由不同數量的vgg\_layer組成。與encoder不同的是要把向量逐漸升維，透過不同convolution layer放大維度到還原圖片。再還原過程中，同時將對應的skip與特徵向量連接起來，利用skip connections技巧還原圖片，以保留更多圖片細節。

一張含有 文字 的圖片

自動產生的描述

1. LSTM

這邊LSTM是用來將encoder的輸入學習特徵，並將其作為decoder的input。而gaussian LSTM 是把latent variable (z)與後驗分布一起使用，以生成更符合真實的數據分布。

一張含有 文字 的圖片

自動產生的描述一張含有 文字 的圖片

自動產生的描述

1. Reparameterization Trick

VAE在decoder抽取latent variable z作為輸入時，在模型計算梯度的部分，利用reparameterization trick將分布視為連續的高斯分布，可以更好計算梯度。先將log-variance 轉成sigma，並從高斯分布抽樣出一個變數與encoder 生成的分布(sigma、mu) sigma相乘並加mu。

一張含有 文字 的圖片

自動產生的描述

1. Data loader

一張含有 文字 的圖片

自動產生的描述一張含有 文字 的圖片

自動產生的描述其中\_\_init\_\_是判斷對哪一種資料集(train、val、test)，以及圖像轉換transform。\_\_len\_\_要計算資料總長度。get\_seq選擇一個路徑從中讀取圖片的順序並將其轉成向量。get\_csv: 從csv對應相對照片來記錄機器人動作和位置(條件)。\_\_getitem\_\_將資料對應打包，把圖片的序列和條件作為向量回傳。

1. Teacher forcing

這部分是用來更新teacher forcing 的比率，原因是teacher forcing用太多會使得訓練模型硬記，造成結果不好。所以用遞減率調整Teacher forcing的程度。使模型在訓練隨著訓練epoch增加，逐漸減少對 teacher forcing的依賴，以提產生更好結果。

一張含有 文字, 螢幕擷取畫面, 屏幕、螢幕, 關閉 的圖片

自動產生的描述

1. Main Idea

Teacher forcing 是一個類似於RNN的技巧，在訓練時，模型會用真實的數據作為t-1的輸入，而不是使用模型預測的輸出。這個方法是為了加速模型的收斂，也可以學到更好的結果。

1. Benefits

它可以使收斂速度變快，因為模型會用真實的數據作為t-1的輸入，而不是根據模型生成的輸出結果。並且減少錯誤，導致預測十一個時間點的誤差使後面序列整個偏掉，提高模型更精確結果。

1. Drawbacks

但模型可能會太依賴這些真實輸入，而不是真的學會這些數據之間的關係，導致模型的泛化能力變差。

1. **Results and discussion**
2. Make videos or gif images for test result

使用的模型參數為epoch: 300、batch size:20、learning rate: 0.002、tfr start decay epoch 100、kl anneal cyclical: True。

GIF是在test時，以前2個做為模型已知結果，去預測剩餘的10張frames。綠色代表已知，紅色為預測。包含approximate posterior，最好的PSNR結果，及隨機選任3個的結果。

一張含有 文字 的圖片

自動產生的描述

1. Output the prediction at each time step

以下Ground Truth 是真實結果，Prediction是在test時以前2個做為模型已知結果，去預測剩餘的10張frames。

Ground Truth:



Prediction:



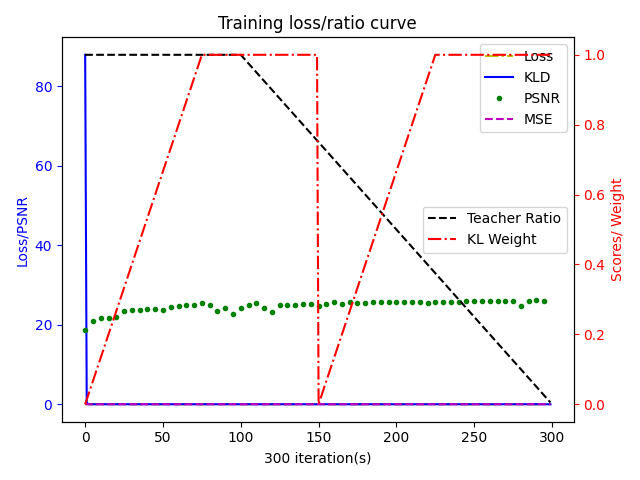
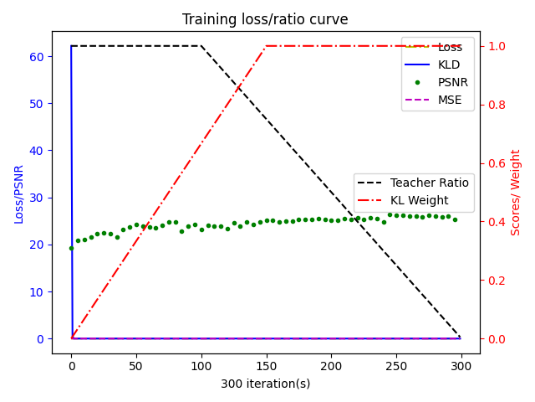
1. Plot the KL loss and PSNR curves
2. kl anneal cyclical:

使用的模型參數為epoch: 300、batch size:20、learning rate: 0.002、tfr start decay epoch 100。

當kl anneal cyclical 設為True的訓練結果比較好，也較穩定，那是因為在kl anneal cyclical 設為True代表會根據周期去更新kl annealing 調整模型。這可以讓整個模型在訓練過程中去 調整KL loss對整體的影響，使模型會以 MSE loss的更新為主，但也可以考慮 KL loss，以提高模型的泛化能力。

1. kl anneal cyclical: False B. kl anneal cyclical: True

Avg. PSNR: 24.618 Avg. PSNR:24.855



1. Teacher forcing:

在訓練模型時，直接在一開始利用teacher forcing時的tfr start decay epoch不設為0，原因是模型學出來與真實輸入差太多，接下來的預測偏差會越來越大，因此我從100 epochs才開始遞減，遞減原因是防止模型太依賴這些真實輸入，而不是真的學會這些數據之間的關係，導致模型的泛化能力變差。則預測結果test有達24.855，但是如果將tfr start decay epoch設較低則test 結果為23.~ 左右。

1. Learning rate:

調整learning rate並沒有差距太多，因此設為跟助教預設一樣的0.002