Lab4 : Conditional VAE For Video Prediction

311511043 李承翰

1. Introduction

這次lab需要實作出 conditional VAE，利用影片前面的 frame 來預測 並生成後面的frame。首先將frame 輸入進encoder，接著使用encode 出來的latent variale 及利用condition(action 還有 position) 得到z當作decoder的輸入，最後就能夠輸出生成我們預測的frame。

AutoEncoder的架構是由encoder 還有 decoder組成，encoder將輸入進行降維處理之後得到latent variale，也就是輸入的主要特徵，接著透過decoder將這個主要特徵還原成輸入本身。而VAE就是透過限制encoder生成的latent variable，使其能夠遵守Normal distribution。這樣就能產生與原始資料類似的輸出資料。而CVAE則是將原始資料和其condition共同作為encoder的輸入，這樣就可以生成特定類別的輸出資料。

1. Derivation of CVAE

推導CVAE的流程和CVAE基本相同，主要是需要將x限制在條件C，也就是p(x) = p(x|c)。而為了要求得模型參數𝜃，要想辦法最大化𝑝(𝑥|𝑐; 𝜃)

由於z有無限多種，我們沒有辦法利用maximum likelihood的原理設計，因此我們需要透過以下方法推導

對兩邊同取log

將兩邊同乘𝑞(Z) 後對 Z積分

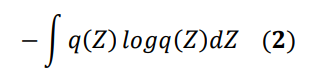
=

= KL(q(Z)||p(Z|X,c;

而因為KL 大於等於0， 要最大化 就是要最大化X,q,

一張含有 文字 的圖片

自動產生的描述



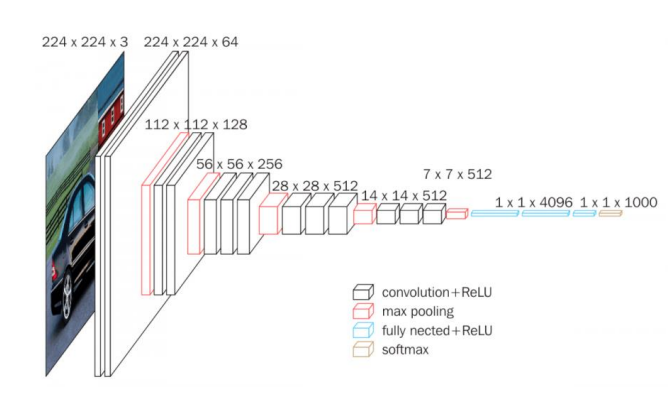


一張含有 文字, 室內, 文件, 螢幕擷取畫面 的圖片

自動產生的描述

1. Implementation details
2. Describe how you implement your model (encoder, decoder, reparameterization trick, dataloader)
3. Encoder、decoder

這次模型當中的架構我採用助教給予的sample code而並沒有去改變他。



這次模型採用的是VGG64的架構，而常見的VGG16架構就如上圖。這兩種網路的差別只在於深度不一樣。透過多個捲積層來堆疊網路，以加深深度。

而在實作上則是透過推疊多個vgg\_layer來達成。

下圖是程式當中的定義，也就是說當我們疊加了多個vgg\_layer，就是堆疊了多個捲積層。

一張含有 文字 的圖片

自動產生的描述

Encoder 的部分是將一個 64\*64 的 3 通道彩色圖片逐層慢慢轉換 成 1\*1 的輸出，提取重要特徵後，透過 decoder上採樣逆向復原成 64\*64\*的圖片。

1. Reparameterization trick

由於從 VAE 參數化分佈的採樣過程是不可微的，這樣會造成梯度 更新的問題，因此需要使用 reparameterization 的技巧。利用高斯分布 的特性，我們可以對一個Normal distribution進行採樣，加上原本均值再乘上 標準差，這樣就可以將隨機元素與學習參數分開

1. Dataloader
2. Describe the teacher forcing (including main idea, benefits and drawbacks
3. Experimental results
4. The highest testing accuracy:

表格中與截圖數據有些微不同，因為是不同次實驗當中擷取的

|  |  |  |
| --- | --- | --- |
|  | With pre-train | Without pre-train |
| KL\_Anneal\_cyclical | 81.52% | 73.38% |
| KL\_Anneal\_monotonic | 82.23% | 73.35% |

1. Comparison figures

|  |  |  |
| --- | --- | --- |
|  | With pre-train | Without pre-train |
| ResNet18 |  |  |
| ResNet50 |  |  |

1. Discussion