Final Project: DeepVo

R07921001 李尚倫 R07522625 李佳蓮 R07522653 陳健倫

# Project description

在這次的Final，我們參考這篇Paper: DeepVO: Towards End-to-End Visual Odometry with Deep Recurrent Convolutional Neural Networks，同時參考的範例code為：<https://github.com/ChiWeiHsiao/DeepVO-pytorch>，主要是想藉由實作這篇Paper來了解SfM等相關知識。

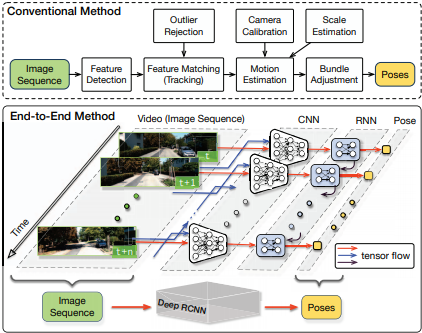


Figure Visual Odometry

# Implementation

* 實作DeepVo的 Model，用自己的DataLoader來做資料處理並訓練。
* 將預測結果用三維圖像的方式匯出。
* 自行實際錄影幾支影片用Blender切割作為Dataset，並丟入Model做Predict。

# How to Run the Program

1. 詳情請看git hub上的 README.md，有較完整的解說，下面為大致流程簡述。

2. 執行getdataset.sh獲得KITTI的資料。

3. 執行Preprocess來對資料作前處理。

4. 執行myMain.py 來做訓練。

5. 執行myTest.py來做預測。

6. 執行myVisualize.py來做2D預測的繪圖。

# DeepVo Model

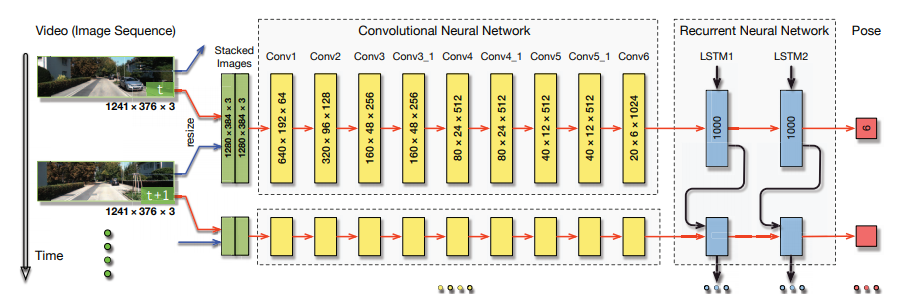
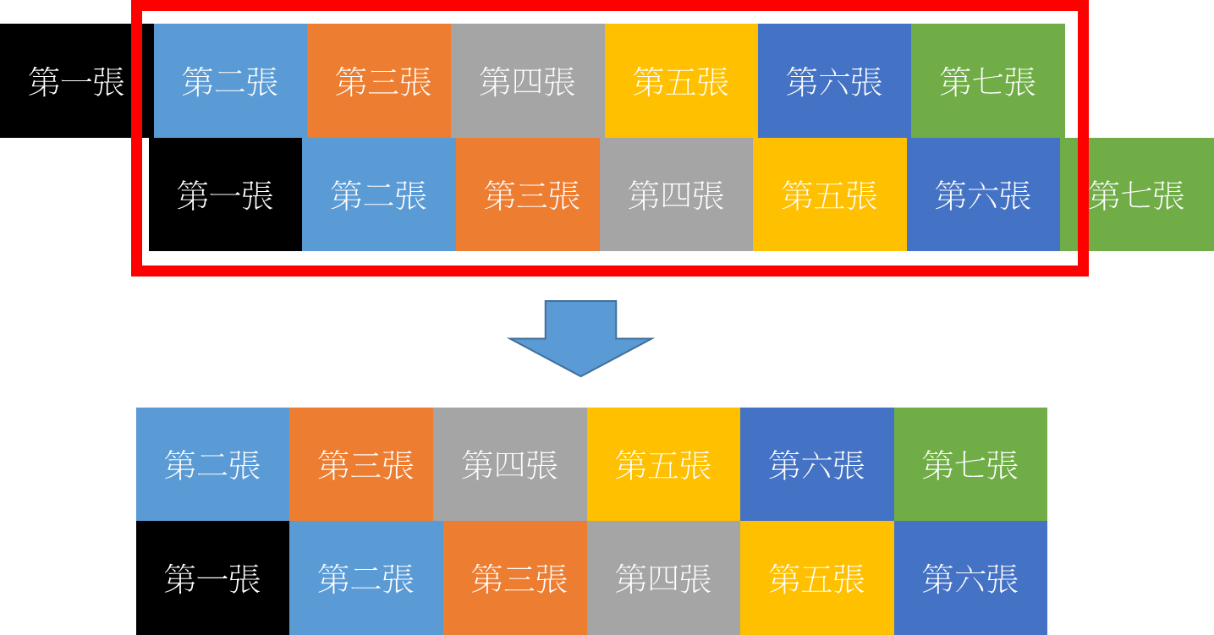
下圖是DeepVo這個架構，是用CNN+RNN作為RCNN的一個Model來做Training，CNN的部分使用網路現有的Flownet來作為Pretrain Model，RNN的部分則是使用pytorch內建的LSTM，Optimizer使用Adam，learning rate為0.0005，batch size為8，Epoch為250，比較需要注意的是這邊影像處理時，0-255的顏色區間必須改為-0.5到0.5之間，這是因為Flownet在訓練時就是對照片做這樣的轉換，因此也要跟著做，Pretrain model才有更大的用意。

Figure DeepVo model

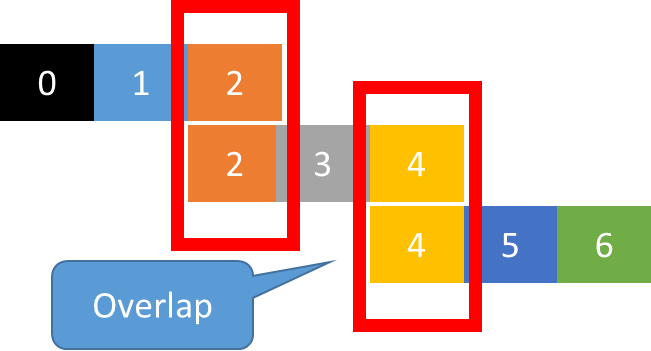
# DataLoader

本Model在預測時的Pose，會有坐標系的問題，所以在Data load進來的時候，必須在ground truth的部分做微處理。假設我們現在一個batch有7張frame，就是1(batch)\*7(frame)\*3(RGB)\*img\_w\*img\_h，在這樣的狀況下，我們把第一個pose視為原點，就是沒有旋轉，同時translate為0，而後六張frame必須減去第一張frame的位置，變成相對坐標系的數值，再來就是要把第frame張的照片剪掉frame-1張的資訊，變成以上一張為基底座標的方式，才能拿來做訓練，也就是每一個predict都是把上一張照片作為原點坐標系去預測下一張所相對移動多少，而這就是此Model比較需要注意的小細節。

解決Pose的問題後，再來就是照片的部分，實際做法我們是把[1:]跟[:-1]這樣的照片concatenate在一起，也就是將2~7張與1~6張疊在一起(如下面示意圖)，作為這個model的input ，dimension 為 1(batch)\*6(frame)\*6(RGB+RGB)\*img\_w\*img\_h，可以看到改變的地方是frame從7變成6，因為去了頭尾的張數，而RGB的通道變成6，原因就是多了一組照片，所以疊的地方是RGB的通道。如此一來 input 為 batch\*6(frame)\*6\*img\_w\*img\_h 、output與label為batch \* 6(frame) \* 6(pose)，這樣就能拿來做訓練。

**Figure 3 Dataloader 擷取示意圖**

而在照片擷取的部分還有的細節就是，我們可以用overlap的技巧來增加data量，假如說資料有1-15張，overlap=1，每個batch所要拿的frame數為3，那麼就是 1-3、3-5、4-6，這樣子拿下去，就能讓Model有更多的Data來訓練。



**Figure 4 overlap示意圖**

# Our dataset

我們共收集五組影片，其中room使用手機拍攝徒步移動軌跡；ntu、campus1、campus2使用手機固定於腳踏車上，並繞行台大一周；NTU1、NTU2使用GoPro安全帽並騎機車繞行台大一周，如下圖所示。



**Figure 6 單車繞行台大校園 (ntu)**



**Figure 7** **教室內步行一圈 (room)**



**Figure 8 椰林大道直線1 (campus1)**



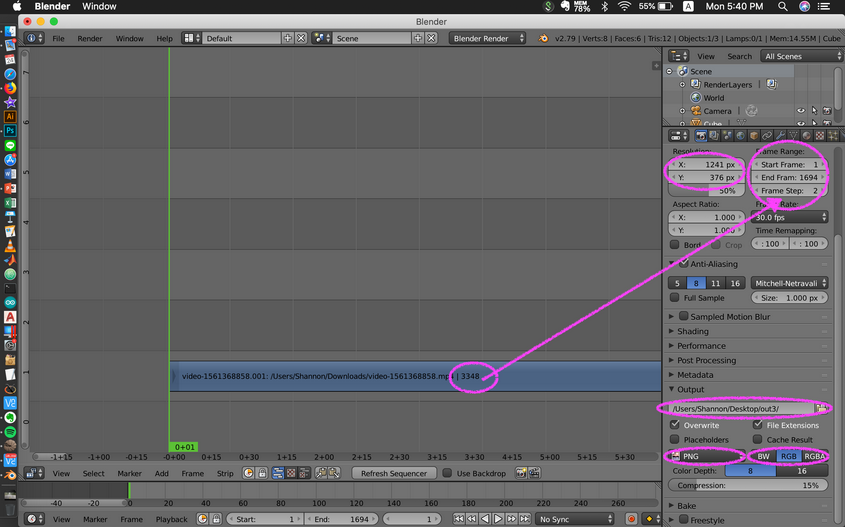
**Figure 9 椰林大道直線2 (campus2)**



**Figure 10 機車繞行台大一周 (NTU1,NTU2)**

# Blender

我們使用Blender對切割影像作切割，並調整影像的尺寸以及取樣頻率等，詳細參數設定如下圖所示。



**Figure 5 Blender參數設定**

# Result

## Test on KITTI dataset

|  |  |  |
| --- | --- | --- |
| paper result | pre-trained model from [alexart13](https://github.com/alexart13) | our model |
|  |  | https://raw.githubusercontent.com/Shining-Zone/VFX_Final/master/result/route_04_gradient.png?token=AIGMJJ5YWLMOD7MT2V5GK425CJBBO |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## test on Self-made dataset video

|  |  |  |  |
| --- | --- | --- | --- |
| ntu | ntu GT | room | room GT |
|  |  |  |  |
| campus1 | campus1 GT | campus2 | campus2 GT |
|  |  |  |  |
| NTU1 | NTU1 GT | NTU2 | NTU2 GT |
|  |  |  |  |

# Conclusion

* Deep VO 對於影像品質非常敏感。

# What do we implement by ourselves

* 訓練時的Script都是自行撰寫，過程有一些部分參考範例Code。
* 意即 training時的Dataloader、Model的建構皆是自行撰寫，並且也自己從頭train。
* 三維度的Visualization是自行撰寫，需要讀取Predict的txt file。
* 自行錄製影片並且用Blender切割成dataset。
* 用自製的dataset去做預測，並且與參考code做比較。

# *What do we not implement by ourselves*

* downloader.sh、myVisualize.py、helper.py、test.py是直接複製網路上的並做一點微調而已。
* 獲得dataset部分，就是直接使用downloader.sh。
* Helper.py內部有一些轉換的函式，像是Euler angle轉換我們是直接使用。
* myVisualize.py是他的二維繪圖，我們直接使用方便與三維的做比較。
* test.py使用的原因在於predict的部分，由於會有坐標系和Dataloader的問題，所以我們是使用自己訓練好的Model搭配他的dataloader以及對Predict的後處理函式來撰寫真正Predict的結果。

# 作業心得

本Project實作了DeepVo這個Model，而我認為最困難的點就是在於座標轉換的地方，然後還有rnn的sequence to sequence training。至於model的撰寫以及dataloader的部分implement不難，但是也因為在predict的時候，貌似需要batch>1才能做predict，所以我的loader沒辦法做predict，不得已只好在最後直接使用網路的loader來做讀檔，但是model至少還是自己train的，所以也可以說對DeepVo要怎麼train有一點基礎上的理解。

## Artifact 1

攝影地點：

## Artifact 2

攝影地點：

## Artifact3

攝影地點：

# Reference