Vehicular Vision System

HW1 - Autonomous Driving Dataset - Report

Group 16 (309605008黃筱晴 , 0851918嚴麗芯)

Dataset : ApolloScape open dataset for autonomous driving
Topic : Self Localization

1. Introduction of the topic/benchmark

ApolloScape是一個關於Autonomous driving開源的Dataset，他們使用一台裝有Laser scanner, video camera、GPS 和IMU/GNSS system感測器的車子收集資料。Dataset提供 camera images, poses, dense LIDAR point clouds, 3d semantic maps, 3d lane markings, 2d segmentation labels.其獲得的資料可應用於 visual 3D scene reconstruction, self-localization, emantic parsing, semantic instance understanding, object 3D instance understanding等任務上。[1]

Fig. 2: Acquisition system consists of two laser scanners, up to six video cameras, and a combined IMU/GNSS system.

ApolloScape	✓	cm	various weather day time 4 regions in China	3D semantic point 70K 3D fitted cars	pixel: 140k	✓	3D / 2D Video 27 classes

[1] database is not open to public yet.

針對Self Localization任務，他們提供的dataset主要可以應用於Semantic Visual Localization研究上。官網提到有研究使用這個dataset去訓練DeLS-3D演算法[2]。另外，找到資料測試PoseNet來使用此Dataset，以下再作說明。

2. Introduction of the data-self (including how it be collected or organized)

how data collected :
他們在中國北京名為Zpark的地方，共6天的不同的時段(早、中、晚)設定起點與終點，從街道的起點行駛到終點，再從終點駛回起點來收集資料，以時速30行駛，每行駛1米就會觸發相機，一共行駛約3公里。

how data organized :

```
{split}/{scene_names}/{data_type}/{record time}/{record id}/{camera id}/{image_name}
```

split : 會分成Train與Test
　　Train : 包括有image,pose,split
　　Test : 只有image用來測試結果
scene_name : 不同的道路
data_type :
　　image : RGB
　　pose : 在地圖上的絕對位置(roll,pitch,yall,x,y,z)
　　split : 分train與val
record time : 拍攝的時間
record id : 相對應的影像檔
canera id : 鏡頭的ID
image_name : {timestamp}_{camera id}

3. Introduction of the label/ground truth (how It be defined and generated)

how label defined

```
1 170427_222949577_Camera_1.jpg -1.7889,0.0250,-1.4811,492.4870,-1880.7770,40.0860
```

一張影像(RGB)對應該相機在地圖上的絕對位置(roll,pitch,yall,x,y,z)

how ground truth generated

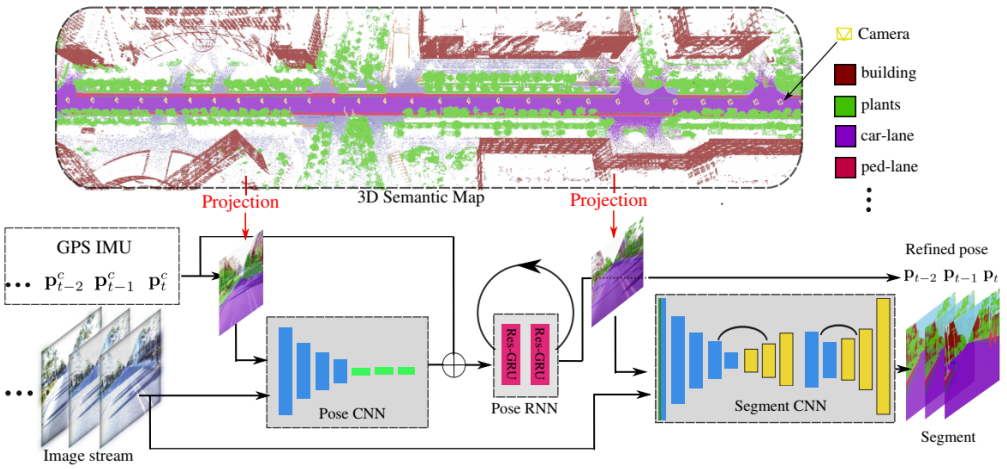
使用Riegl 收集3D pointcloud加上兩個高分辨率鏡頭進行同步和校正，以30 fps的幀率進行記錄。每個相機都有高精度的 GPS 和 IMU，實時記錄準確的相機姿勢。

4. Evaluation metrics

$$e_{translation} = median(\{\|t_i - t_i^*\|_2\}_{i\in\{1,N\}})$$

$$e_{rotation} = median(\{\arccos(|q(r_i) \cdot q(r_i^*)|)\}_{i\in\{1,N\}})$$

5. Bonus Supplement: (e.g., other dataset contains the same topic, the existing projects utilize it…)
   a. DeLS-3D: Deep Localization and Segmentation with a 3D Semantic Map use the dataset [2]
   有使用這個dataset。DeLS-3D演算法的訓練方法是從LIDAR得到point cloud，初始會先GPS/IMU得到粗略的相機姿態，之後透過semantic label得出渲染地圖區域，其集合RGB圖像聯合輸入CNN網絡計算出相對旋轉和平移，並產生修正後的相機姿勢。然後再傳到RNN進一步提高streaming時的估計精度。

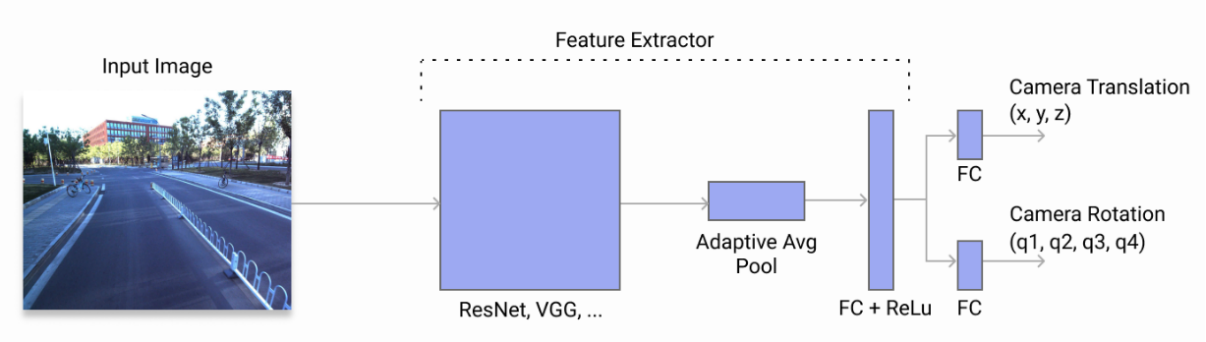


沒有提供Pretrain model 與 Training pipeline

b. PoseNet[3]
這個是一個end-to-end的方法，輸入一張RGB的影像經過CNN的網絡，輸出相機的姿態。論文提出，在室外場景可達2m + 6deg的誤差。論文所使用的Dataset是由作者們在劍橋大學Landmarks 的King's College 場景周圍拍攝的。包含原始視頻，提取的影像有 6 自由度相機姿勢和場景的視覺重建[3c]。而他們所使用的ground truth measurements是Towards Linear-time Incremental Structure from Motion論文所得出的方法。

6. Implement - PoseNet using ApolloScape Dataset

1. 模型架構：



PyTorch實做參考自https://github.com/bexcite/apolloscape-loc。
其中，Input Image在參考repo中原為250*250，但是因為電腦GPU VRAM不夠，改為125*125。
Feature Extractor第一層使用ResNet18，並load pretrained weight。

2. 訓練資料：
使用從http://apolloscape.auto/self_localization.html下載的Sample data。檔案結構如下。在trainval_split中已經幫我們分為train set 和validation set，並搭配上述參考repo中的dataloader來處理

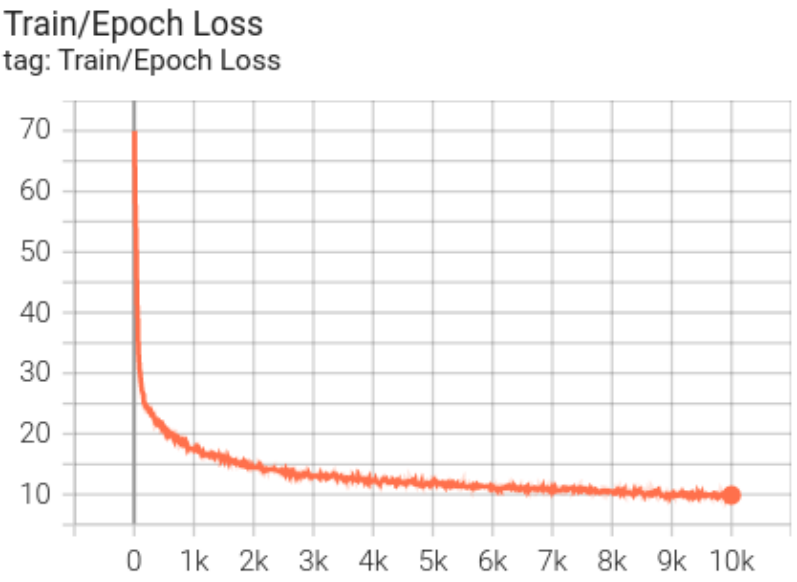
```
yellow@yellow-arg7:~/VSS_HW1/self-localization-sample/self-localization-sample$ tree -d -L 3
.
└── zpark
    ├── camera_params
    ├── image
    │   ├── Record001
    │   ├── Record002
    │   ├── Record003
    │   ├── Record004
    │   ├── Record006
    │   ├── Record007
    │   ├── Record008
    │   ├── Record009
    │   ├── Record010
    │   ├── Record011
    │   ├── Record012
    │   ├── Record013
    │   └── Record014
    ├── pose
    │   ├── Record001
    │   ├── Record002
    │   ├── Record003
    │   ├── Record004
    │   ├── Record006
    │   ├── Record007
    │   ├── Record008
    │   ├── Record009
    │   ├── Record010
    │   ├── Record011
    │   ├── Record012
    │   ├── Record013
    │   └── Record014
    └── trainval_split
```

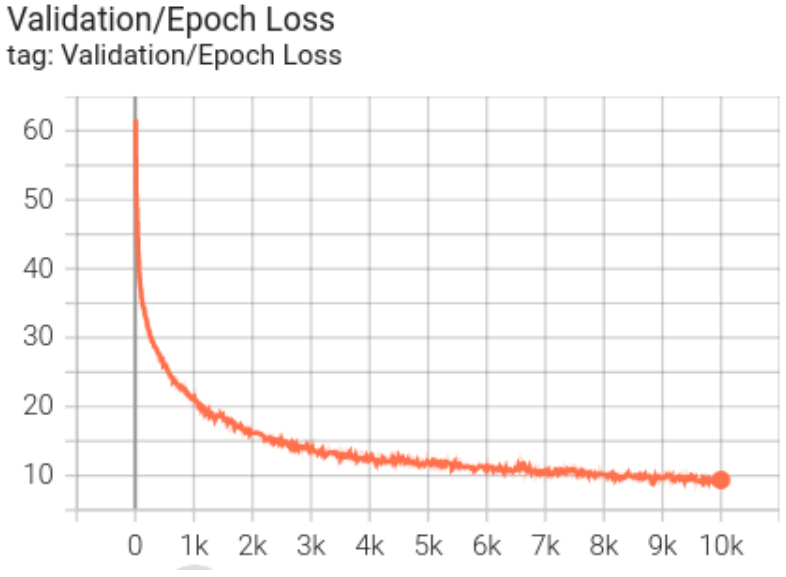
```
train_record = 'Record001'
train_dataset = Apolloscape(root=os.path.join(APOLLO_PATH), road="zpark",
                            transform=transform, record=train_record, normalize_poses=True,
                            pose_format='quat', train=True, cache_transform=True, stereo=stereo)
val_record = 'Record013'
val_dataset = Apolloscape(root=os.path.join(APOLLO_PATH), road="zpark",
                            transform=transform, record=val_record, normalize_poses=True,
                            pose_format='quat', train=False, cache_transform=True, stereo=stereo)
```

共有242張traning data和240張validation data。其中，validation set是和train set同路段但是不同timestamp紀錄的資料。

3. 訓練設定
optimizer : Adam
learning rate : 1e-4
weight decay: 0.0005
batch size: 64
epoch: 10000

4. Learning Curve

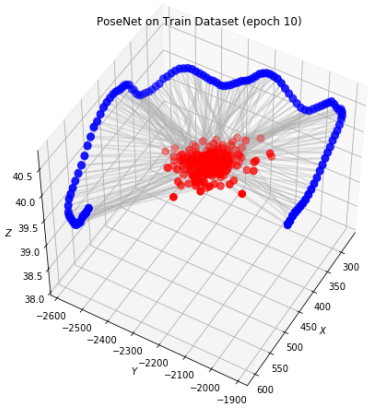


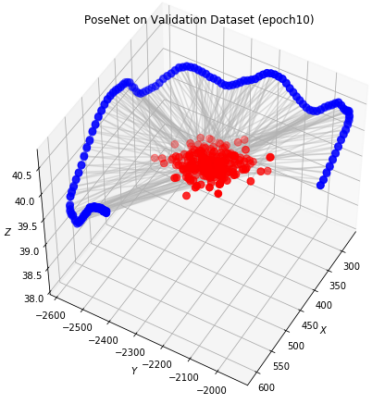


5. 視覺化訓練過程與結果表現

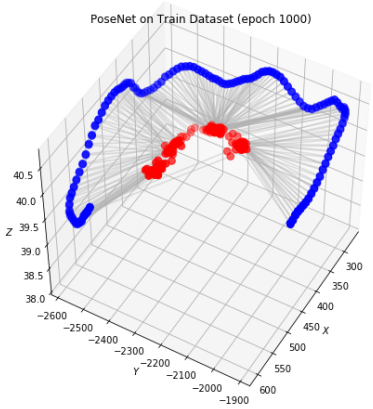
藍色為ground truth，紅色為PoseNet Model預測結果。

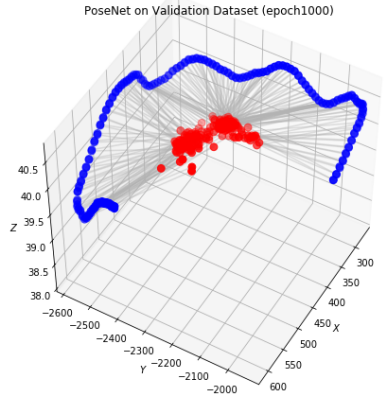
epoch = 10



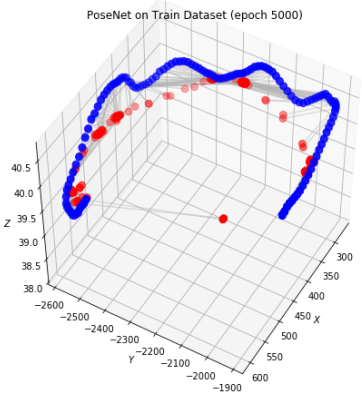


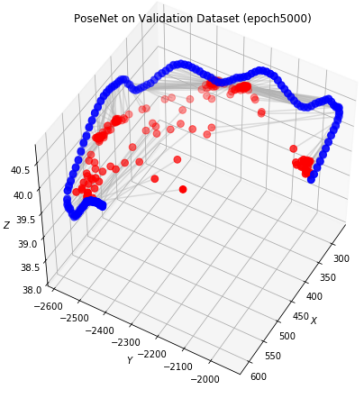
epoch = 1000



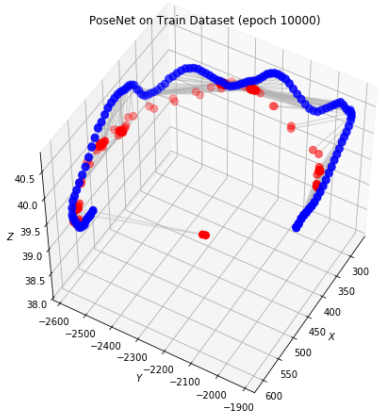


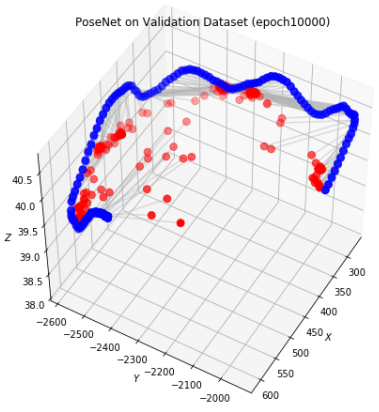
epoch = 5000





epoch = 10000





6. Evaluation with Evaluation metrics

```
=== Test Training Dataset ======
gt_poses = (242, 7)
pred_poses = (242, 7)
poses_std = 280.970
Translation(T) error in meters and Rotation(R) error in degrees:
T: median = 87.985, mean = 120.299
R: median = 1.324, mean = 1.757

=== Test Validation Dataset ======
gt_poses = (240, 7)
pred_poses = (240, 7)
poses_std = 280.970
Translation(T) error in meters and Rotation(R) error in degrees:
T: median = 101.080, mean = 130.990
R: median = 1.876, mean = 8.451
```

7. Disscusion
   a. 用PoseNet方法利用ApolloScape提供的dataset沒有很好的結果
      原本PoseNet所使用的dataset多以在行人視角與相似街景，而ApolloScape dataset是以車子視角與不同街景，因此在套用PoseNet在ApolloScape Dataset沒有得到PoseNet提出的2m + 6deg的誤差。

 PoseNet Dataset	 ApolloScape Dataset

8. Reference

[The ApolloScape Open Dataset for Autonomous Driving and its Application](#) [1]

a. website : http://apolloscape.auto/self_localization.html
b. github : https://github.com/ApolloScapeAuto/dataset-api/tree/master/self_localization

[DeLS-3D: Deep Localization and Segmentation with a 3D Semantic Map use the dataset](#) [2]

a. github : https://github.com/pengwangucla/DeLS-3D

[PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization](#)[3]

a. tutorial : https://capsulesbot.com/blog/2018/08/24/apolloscape-posenet-pytorch.html
b. github : https://github.com/bexcite/apolloscape-loc
c. Dataset : https://www.repository.cam.ac.uk/handle/1810/251342