

111 學年度大學部專題競賽



國立清華大學資訊工程學系

Department of Computer Science, National Tsing Hua University

SensatUrban : 3D semantic segmentation challenge with limited resources

組員姓名：108062118 賴翰霖 108062203 江浩辰 108062315 莊子郁

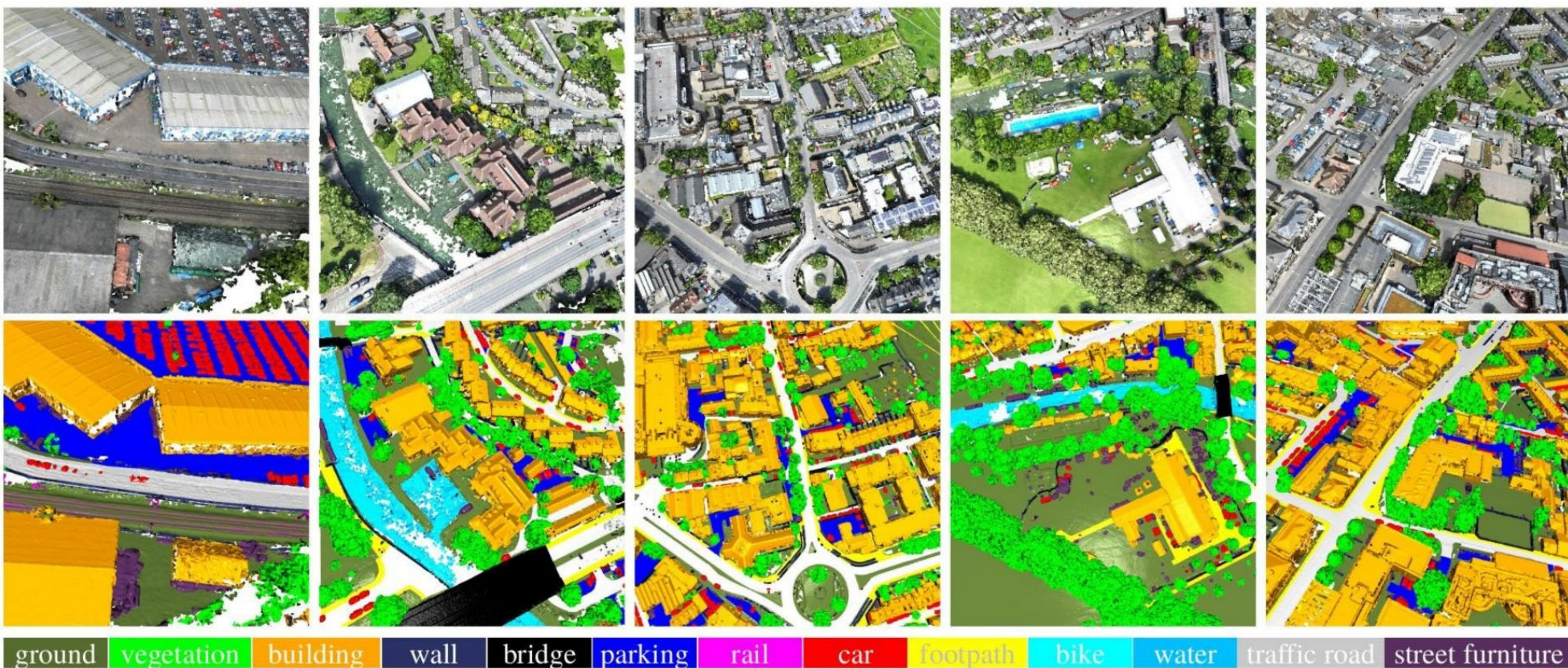
摘要

為了針對大尺度空間的語義分割提供一個辨別能力強、訓練速度和穩定性兼具的模型，我們參考現有的模型和演算法進行研究。在研究過程中發現對資料的預處理及採樣方法在模型訓練上有顯著的影響，並朝此方向進行改進，試圖突破語義分割在超大型資料集訓練上的極限，希望藉由這塊領域的進步，廣泛應用在超大型計畫中，比如城市規劃、土地勘查等城市規模的企劃。

問題描述

Urban 3D Challenge為歐洲計算機視覺會議（ECCV 2022）旨在建立的城市尺度（Urban-Scale）點雲分割基準所推出的挑戰，在3D語義分割中此挑戰所使用的資料集為SensatUrban，此資料集為目前世界最大的點雲資料集，內含7.6平方公里中近 30 億個具有詳細分類並標註的點，而此資料集表明了真實場景中的類別分布不均，故若需對此做語義分割，最需解決的即是極大的資料量和分布嚴重不均的類別。

於是我們決定以 Urban 3D Challenge 其中的3D語義分割研究作為我們的專題主題，試著在實驗室較為有限的計算資源下進行訓練，以盡量達到最好的表現



設計原理

我們主要以目前 Paper with Code 上面針對各模型在 SensatUrban 資料集上的分析排名作參考，並以工作坊所提供基於點的訓練模型—RandLA-Net 與基於投影（Projection）的訓練模型—BEV-Seg-Net 作為修改基礎及比較對象：

1. RandLA-Net

RandLA-Net具計算效率高和記憶體占用少兩大特性，且在大規模點雲具有很高的應用性。其重要結構有二：隨機採樣和局部特徵聚合模塊。

我們將 RandLA-Net 的研究與輸入切割以下三個部分：

(1)資料集預處理：

由於資料集中種類分布不均，我們先對資料集進行資料增強，來增加資料的多樣性、隨機性和總數。

(2)採樣方式：

a. 自適應採樣 (Adaptive Sampling)

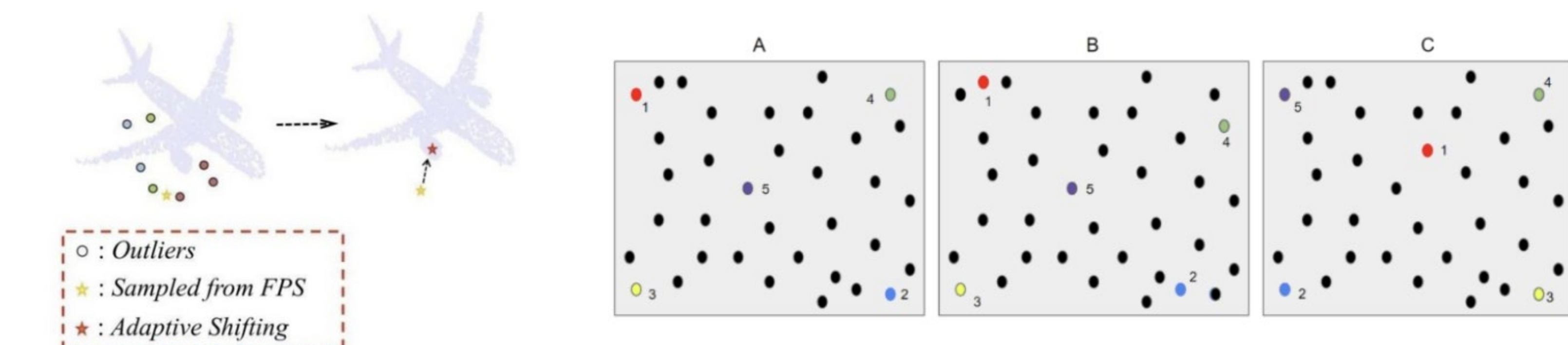
採樣的點可能為噪點，故使用自注意力機制學習將噪點移至正常的位置。

b. 隨機採樣 (Random Sampling)

為 RandLA-Net 預設的採樣方式。

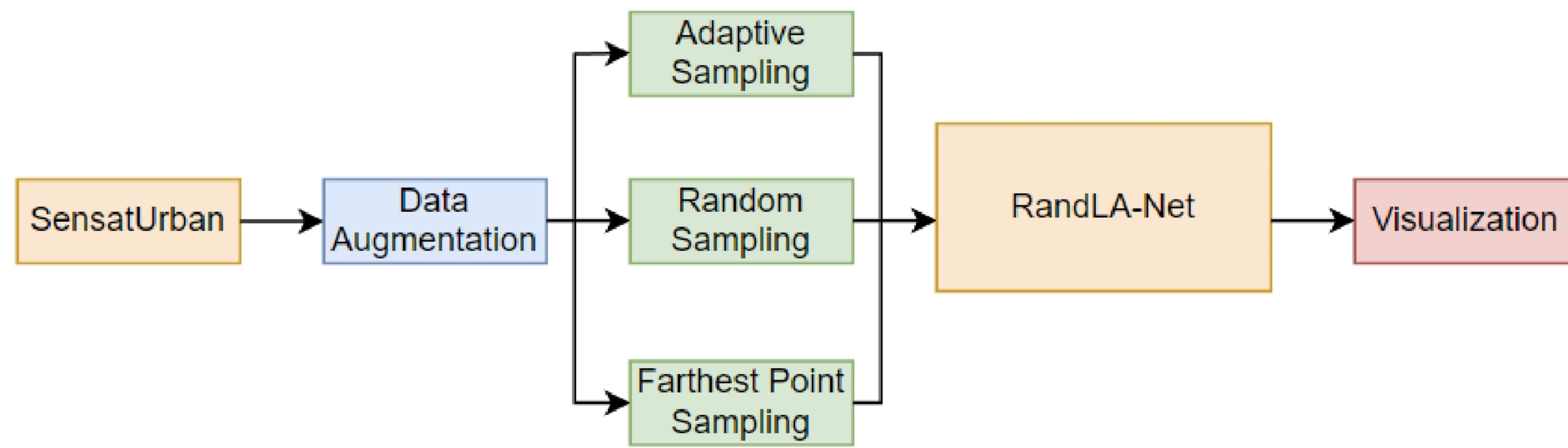
C. 最遠點採樣 (Farthest Point Sampling)

每次挑選最遠點作為採樣點，故而達到採樣均勻的結果。



(3)模型與點雲可視化

點雲可視化可以讓我們簡易的察看驗證集與測試集在我們預測的表現下正解的差別，從而觀察出模型在哪些部分的預測需要加強，有助於我們了解對模型進行改良的方向。



2. BEV-Seg-Net

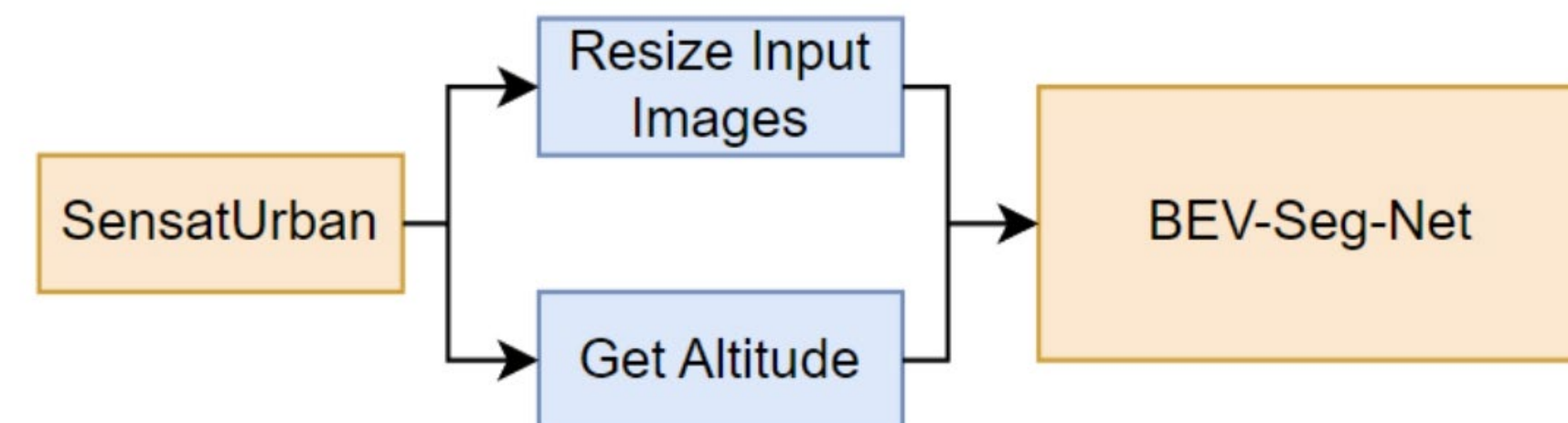
BEV-Seg-Net先將3D 點雲投影至2D 的鳥瞰圖，以 2D 形式做語義分割，再重新將分類映射回 3D 的點雲。BEV-Seg-Net 的主要結構有鳥瞰圖投影和2D多模態分割。

(1)輸入圖像大小

由於 GPU 的硬體限制，我們對論文中的原始尺寸（500x500）進行縮放到三種大小，分別是100x100、200x200和300x300。

(2)輸入圖像僅使用 R、G、B 三個通道

因為鳥瞰圖導致原本3D空間的物體缺少了高低差，於是我們增加了第四個通道，即為高度（Altitude）來做訓練。



實驗結果

我們使用的評估指標為 Mean Intersection over Union（mIOU）。

主硬設備為GPU：GeForce GTX 1080（VRAM 8GB）和GeForce GTX TITAN X（VRAM 12GB）

1.RandLA-Net

我們試更改的三個變因其結果：

- 資料增強：隨機平移、隨機縮放、增加高斯雜訊
- 批次大小與資料大小的比例
- 採樣方式：最遠點採樣法、自適應採樣

2. BEV-Seg-Net

我們嘗試更改的兩個變因及其結果

- 輸入圖像大小100x100、200x200、300x300
- 第四通道：高度

Method	mIoU	Ground	Veg.	Buildings	Walls	Bridge	Parking	Rail	Traffic oads	Street	Cars	Footpath	Bikes	Water
RandLA-Net Default	48.20	72.23	95.23	93.41	34.08	8.48	42.91	11.16	53.26	28.33	74.43	16.72	0.00	47.50
RandLA-Net Epoch200	49.24	74.29	95.23	94.09	35.84	14.21	47.86	12.29	53.87	28.65	75.33	16.72	0.00	58.52
RandLA-net w/ Titan X	50.02	72.12	96.41	94.00	47.06	16.14	49.55	7.62	59.06	33.29	79.31	20.41	3.21	63.69
RandLA-net 6*40000	54.09	74.12	96.37	94.44	51.92	48.08	47.73	14.44	60.60	39.16	82.63	19.49	9.39	64.77
RandLA-net 24*10000	55.6	73.74	96.66	94.41	52.69	59.04	55.32	20.89	61.37	33.34	83.01	17.99	0.00	74.30
Hand Augmentation	40.88	67.20	94.87	90.25	34.12	4.15	32.93	0.72	54.95	23.57	72.67	17.08	0.00	38.96
tf. augmentation	41.25	66.60	95.22	90.83	32.34	4.34	38.35	0.03	54.80	23.79	70.48	15.83	0.00	43.70
FPS – method1	30.16	69.19	94.92	84.56	26.83	0.04	1.06	0.00	45.02	13.05	57.41	0.00	0.00	0.00
FPS – method2	38.70	67.73	93.53	87.77	29.77	2.34	23.97	0.00	38.72	26.22	73.99	12.11	0.00	46.95
FPS – method3	40.74	69.79	93.51	87.72	31.83	7.42	33.24	0.00	46.12	25.15	72.26	15.93	0.00	46.70
BEV-RGB 100x100 Epoch300	16.32	47.26	61.39	37.72	0.00	0.00	0.00	0.00	23.22	0.00	0.00	0.00	0.00	0.00
BEV-RGBD 100x100 Epoch300	64.93	88.78	90.44	93.96	38.25	82.43	77.78	0.00	85.42	40.48	70.88	69.01	0.00	76.12
BEV-RGBD 100x100 Epoch200	57.61	79.47	79.79	85.69	24.47	74.52	69.02	0.00	78.50	36.83	69.47	58.84	0.00	66.55
BEV-RGBD 200x200 Epoch200	46.93	68.22	88.48	90.72	33.86	7.41	40.19	0.00	58.69	28.93	68.12	20.45	0.00	65.60
BEV-RGBD 300x300 Epoch80	42.07	48.50	66.43	68.38	24.42	40.04	27.45	22.29	51.19	40.00	58.31	28.01	0.00	31.80