

Lab02 Report – DeepLearning

學生：梁皓鈞，學號：314580042，系所：前瞻半導體研究所晶片組

Your report should include/answer

I. Required

- Screenshot of Task 1 (MAE on public testing data ≤ 2.4)

Answer:

```
[INFO] Loaded model state dict from checkpoint: model.pth
Evaluating: 0%|          | 0/13 [00:00<?, ?it/s]
[TEST] MAE[r,l,t] = [2.5879, 1.8609, 1.5457] | Avg MAE = 1.9982
[TEST] RMSE[r,l,t] = [2.8653, 2.2349, 1.9857] | Avg RMSE = 2.3620
```

- Screenshot of Task 2 (MAE on public testing data ≤ 2.0)

Answer:

```
[INFO] Loaded model state dict from checkpoint: model.pth
Evaluating: 0%|          | 0/13 [00:00<?, ?it/s]
[TEST] MAE[r,l,t] = [2.5047, 1.4540, 0.9329] | Avg MAE = 1.6305
[TEST] RMSE[r,l,t] = [2.7168, 1.7530, 1.1951] | Avg RMSE = 1.8883
```

- In Task 2
 - What model did you choose?

Answer:

我設計的架構如下：首先透過 1×1 卷積將輸入張量由 $[B, 1, H, W]$ 映射為 $[B, 3, H, W]$ ，接著依據 ImageNet-1K 的統計量進行正規化：均值為 $\text{mean} = [0.485, 0.456, 0.406]$ ，標準差為 $\text{std} = [0.229, 0.224, 0.225]$ 。正規化後，我採用移除最終全連接層的預訓練 ResNet-18 作為骨幹來提取特徵，輸出維度為展開後的 $[B, 512]$ 。最後，我在此基礎上實作一個分類頭，其結構為 $\text{Linear}(512 \rightarrow 512) \rightarrow \text{ReLU} \rightarrow \text{Dropout}(p = 0.2) \rightarrow \text{Linear}(512 \rightarrow 3) \rightarrow \text{Softplus}$ 。

- Why did you choose this model? What advantages does it offer?

Answer:

由於輸入為灰階影像張量 $[B, 1, H, W]$ ，而預訓練骨幹網路需要三通道輸入，因此在最前端我透過一個 1×1 卷積，將輸入張量由 $[B, 1, H, W]$ 映射為 $[B, 3, H, W]$ ；這樣一來，不僅保留網路在訓練過程中自動學習更佳通道混合的彈性，還能使後續特徵圖契合預訓練模型的輸入

形狀。特徵抽取部分選用 ResNet-18 作為骨幹，並移除其末端給 ImageNet-1K 資料集的全連接層分類頭，原因是 ResNet-18 參數量少、推論速度快，對於 UCSD 小型、低解析度的行人資料集而言，不僅能降低過擬合風險，也更適合進行遷移學習。最後，我於自行實作的分類頭中加入 Dropout，以進一步降低過擬合風險；輸出層則使用 Softplus 激活函數，確保結果為非負值，避免出現物理上不合理的負計數。

- **Compare the characteristics of MAE (Mean Absolute Error) and RMSE (Root Mean Square Error). In what types of scenarios might one be preferred over the other?**

Answer:

MAE（平均絕對誤差，Mean Absolute Error）的公式為：

$$\frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

，而 RMSE（均方根誤差，Root Mean Squared Error）的公式為：

$$\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

，其中，N 表示批量大小， y_i 是第 i 個樣本的實際值， \hat{y}_i 是第 i 個樣本的預測值。對於 RMSE 而言，誤差會因平方項而放大，大誤差項因而受到更嚴重的懲罰，因此在某些對誤差極度敏感的應用中，例如：醫療診斷、自動駕駛，RMSE 往往更受青睞。相較之下，MAE 對所有誤差一視同仁，不會特別放大大誤差，適合用於觀察整體預測偏差的情境。

- **Another popular method in crowd counting is “density map estimation.” Briefly explain what density map estimation means in the context of crowd counting.**

Answer:

在進行人群總數估計時，密度圖估計法會將輸入的人群圖像轉換為與原圖相同尺寸的像素密度圖；其中，每個像素的數值表示該位置附近出現人頭的機率。換言之，可以將每一個人頭位置視為存在一以其為中心的高斯分布，所有人頭的高斯分布疊加後，便構成了完整的密度估計圖；若將密度圖中所有像素值進行累加，所得結果便是對應於圖像中的總人數。

- **How does it differ from the regression-based approach used in our implementation?**

Answer:

我們的回歸方法直接將整張影像映射為總人數，訓練目標直觀明確，但此作法缺乏空間分布

資訊，因此在面對場景變化或極端情況時，較難進行局部校正。相較之下，密度圖估計方法則將影像映射為以像素為單位的密度場，其總和對應人數，不僅保留了空間分布資訊，也更容易進行局部校正；然而，此方法的資料標註成本相對較高。

- **Give at least one metrics used to evaluate it.**

Answer:

密度圖估計法將像素加總後即為總人數預測值，可透過 MAE、RMSE、來計算誤差。

II. Can include but not limited to

- **Anything you do to improve the accuracy.**

Answer:

在權重初始化部分，卷積層採用了 Kaiming 初始化，全連接層則使用 Xavier 初始化，而批次正規化則採取恆等映射，以提升初期梯度傳遞的穩定性並加速模型收斂。在損失函數的設計上，選用了 Smooth L1 Loss，其對離群值更具魯棒性，能有效減少訓練過程中的不穩定影響。優化器方面使用 Adam，並搭配 Weight Decay 以降低過擬合風險。最後，學習率調度器採用了 Reduce Learning Rate On Plateau，當模型在指定的 Epoch 數內未獲得明顯改善時，會自動降低學習率，以進一步促進模型的快速且穩定收斂。

- **Discuss any challenges you faced.**

Answer:

在訓練過程中，最大的挑戰在於反覆調整 Learning Rate、Weight Decay 等超參數所帶來的不確定性。每一次嘗試都伴隨著對最終模型表現的焦慮與未知感，在找到合適的設定之前，幾乎無法確定調整方向是否正確。直到成功確立適切的超參數組合後，才真正確信這樣的訓練策略是可行的，並感受到如釋重負的安心感。