



Deep Learning for Computer Vision

Group 30 So_Ez_Peasy

薛孟謙 胡捷翔 李唯 謝昌諭 Poh-Yuen Chen

國立臺灣大學

R09631012 R10921120 B07505046 B06602052 R09942151

介紹

數據量分布不均的資料集在任何實驗收集中都可能發生，對於此種大量 (large-scale)、細微 (fine-grained) 及長尾 (long-tailed) 分布的食物分類任務而言，最困難的是分辨出類別與類別之間的細微差異，且同時少量類別要有不錯的準確率，因此，影像數量便是使模型訓練困難的關鍵。本組在報告中，提出三種方法，在訓練的過程中，改變採樣 (sample) 的方式，讓少量影像的類別多次出現，使得模型更加熟悉該類別的特徵。

方法

本組共使用了三種方法來改變採樣：
(1) Mixup training，
(2) 使用 weighted random sampler
(3) resample 搭配 semi-supervised learning。

1

Mixup training

訓練時在每一個 batch 訓練前，先從 Beta 分佈中 sample 出一個值 λ 後，隨機取出兩筆資料 $(x_i, y_i), (x_j, y_j)$ ，再用式 (1) 與 (2) 去合併兩筆資料。這個方法除了也可以直接用在 input data 上也可以用在中間 hidden layer 的 output 上。

$$\hat{x} = \lambda x_i + (1 - \lambda) x_j \quad (1)$$

$$\hat{y} = \lambda y_i + (1 - \lambda) y_j \quad (2)$$

Scenario	Accuracy
Baseline	
Input mixup	
Hidden layer output mixup	

2

Weighted Random Sampler

先取得每一個 class 的圖片數量，再算出 imbalanced data = 該類別標籤 * 圖片數量，進而計算平衡機率 (1/每個類別出現次數) 使每個類別被抽到的機率相等，最後再使用 weighted random sampler，model 使用 densenet121。

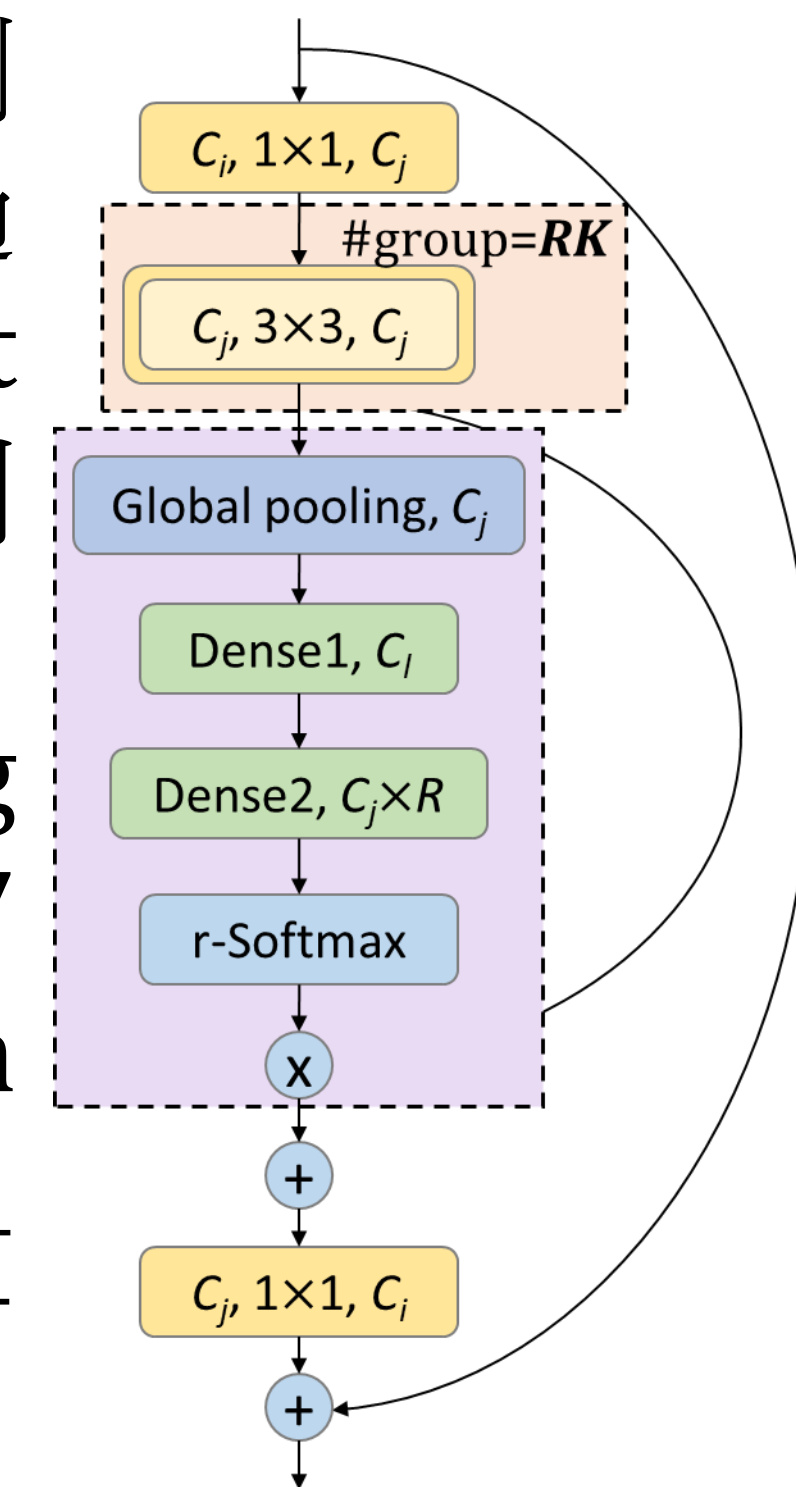
3

Resampling + semi-supervised

Resampling 先計算各類別影像數量占比 (f_c)，再透過式 (3) 求該類別的 repeat factor (r_c)，若為非整數則取無條件進位至整數。Semi-supervised learning 當 validation accuracy ≥ 0.7 時啟動，預測 validation set 的 pseudo label，且 confidence score ≥ 0.99 才會採用此影像。

$$r_c = \max(1, \sqrt{t/f_c}) \quad (3)$$

Block of ResNeSt-101



Scenario	Public accuracy (%)			
	Main	Freq.	Com.	Rare
ResNeSt-101 [1]	66.5	84.4	61.8	23.0
+ resampling [2] ($t=1e-6$)	68.1	85.7	64.4	24.5
+ semi-supervised	68.4	86.9	66.6	24.7

結論

本組共使用了三種方法，其中使用 resampling 及 semi-supervised learning 的方法最佳，可在 Main 準確率達 68.4%，freq. 達 86.9%，com. 達 66.6% 及 rare 達 24.7%。

參考資料

[1] ResNeSt: Split-Attention Networks

[2] LVIS: A Dataset for Large Vocabulary Instance Segmentation