

HW3 DeepQ

電機三 b05901009 高瑋聰

a. 做了哪些比較的實驗

1. optimizer 的選取：實驗過 Adam、SGD、Adamax，最後選用了 Adamax
2. optimizer 的 lr：根據不同的 optimizer 嘗試過各種 lr 的數值，但在 Adam 或 Adamax 這類的 optimizer 上影響不是很大，只要一開始 lr 不要太大即可。
3. batch size：試過 16、32、64，但影響不大。
4. preprocessing：試過對 input 做 RandomCrop、ColorJitter、RandomAffine、RandomHorizontalFlip、RandomVerticalFlip，結果感覺不加入 VerticalFlip 以及 rotation 的 degree 不能太大會有比較好的結果。
5. pretrain：直接使用。
6. mobilenet 的 dropout：試過 0.3~0.7，但結果影響不大。
7. Batch Normalization 的 momentum：試過 0.3 但結果沒有比較好。
8. 存取時的 epoch 數：resnet 大約在 20，mobilenet 大約在 40，總之不需要跑完 50 個 epoch 即可有差不多的 performance。
9. Random Seed：為了比較好比較差異因此設了一個固定的 seed，感覺上對結果比起不設 seed 時沒有太大影響。

b. 最後選擇此值的原因

- optimizer：用 Adamax 比用 Adam、SGD 來得穩定。
- lr：使用了比預設值小的 lr (0.0005, 0.001)，因為考慮到是使用 pretrain 的 model，lr 應該不須太大。
- batch size：感覺影響不大，但考慮到參數更新次數，於是選擇了中間值的 32。
- preprocessing：因為用 pretrain model 後只要 train 一下很容易就 overfit，因此一開始我加了很強的 data augmentation，但發覺太強會讓 model train 不起來，此外加入 vertical flip 與太多 rotation 對 validation acc 沒有太大的幫助。
- pretrain：因為這次的 dataset 很小，epoch 數也少，因此直覺上認為採用被大的 dataset pretrain 過的 model 無論是對收斂速度與最後的表現應該都有正面影響。
- dropout：如上所述，感覺影響不大，可能是該 model 已經 pretrain 得很好了，即使調整 dropout rate 對 validation acc 沒有太大影響。
- Batch Normalization momentum：因為改了 0.3 後沒有變好，應該是預設值就已經很好了。
- epoch：因為使用 pretrain model，收斂速度其實蠻快的，經過觀察大約在 20~40 就存取即可，train 更久表現似乎不會更好。
- Random seed：為了方便做其他參數的比較，於是固定了某個 seed，但其表現與 None 的時候差異不大。