

DLCV HW2 Report

B05901027 電機四 詹書愷

Problem 1

Baseline Model

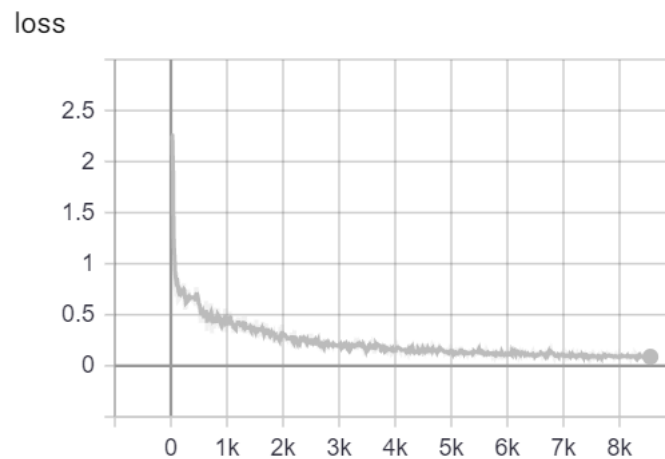
1. Data Processing

除了將 data 轉成 0~1 的 tensor 以外，為了增加資料的多樣性，我還作了以下三個處理：

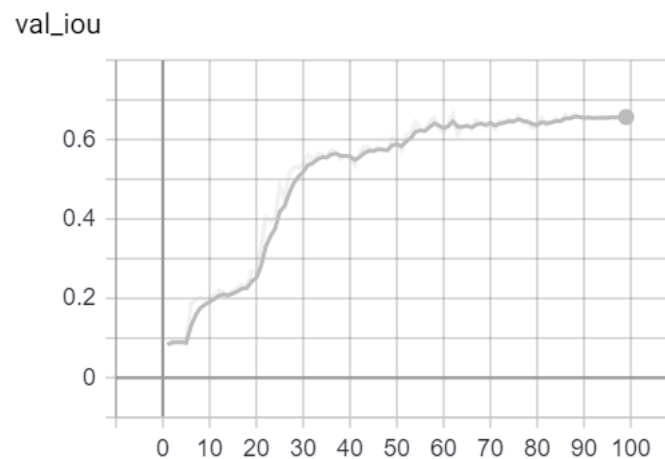
- 1) Normalization: 將資料標準化。為求方便，mean 與 std 皆設為 0.5。
- 2) Random Flip: 每張圖片有 0.5 的機率會水平翻轉。
- 3) Random Scale: 每張圖片皆會被縮放，而縮放的倍率會從[0.5, 1.5]中隨機取樣。

2. Figures

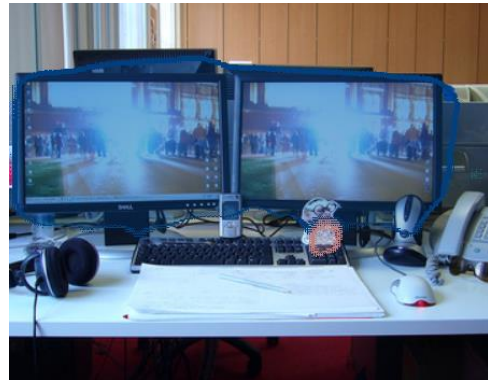
- 1) Training loss v.s. iteration



- 2) mIOU score of validation v.s. epoch



3. Visualization



4. Results

- 1) mIOU score of validation set: 66.46%
- 2) mIOU score per class:

```
class #0 : 0.90832
class #1 : 0.76403
class #2 : 0.64043
class #3 : 0.75714
class #4 : 0.39772
class #5 : 0.50266
class #6 : 0.57822
class #7 : 0.73290
class #8 : 0.70064
```

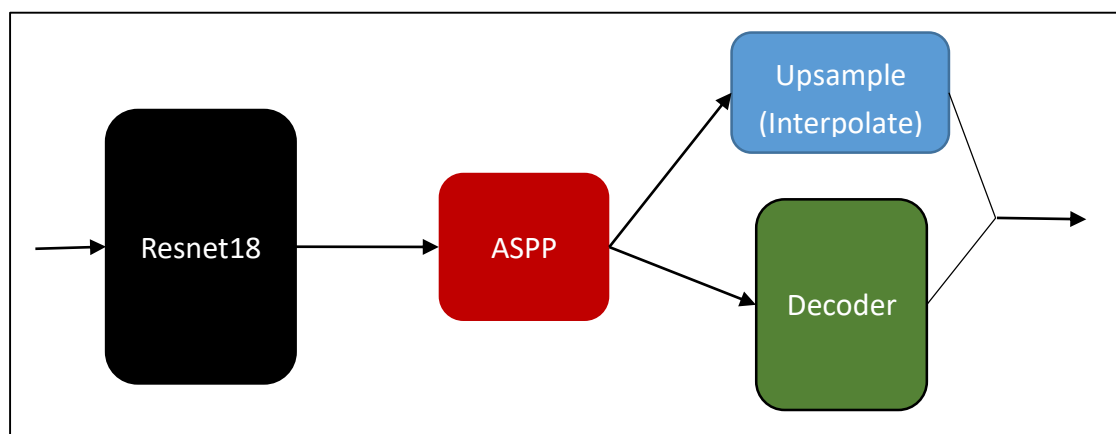
3) Reason

mIOU 成績最高的是 background，最低則是 TV/Monitor。個人認為這樣的結果是因為 data 量不均所造成的：幾乎每張圖片都有 background，故這一類別有非常充足的資料，可以學到很多細節，像是甚麼樣的顏色變化、模糊度可以代表 background 等等。而相反的，TV/Monitor 類資料量便少了許多。另外，validation set 中有一些圖片顯示了螢幕的背面(如下圖)，而螢幕的不同方向投影出的 feature 相當不同，這些情況在 training set 中又少之又少；再加上電視螢幕中可以顯示各式各樣的畫面，較難以顏色作為判斷的標準。network 可能必須學到螢幕的材質、形狀等等 feature 才能判斷出來，顯著的增加了成功預測的難度。



Improved Model

1. Model Architecture



上圖為我的 model 架構，其中 Resnet18 與 baseline 相同，使用了抓出 feature 的部分(convolution)。ASPP 為 Atrous Spatial Pyramid Pooling，為 DeepLab V2¹ 所提出的架構，透過多個 dilation rate 不同的 convolution，去抓出 feature map 中不同規模的資訊，並將其作為分類的依據。

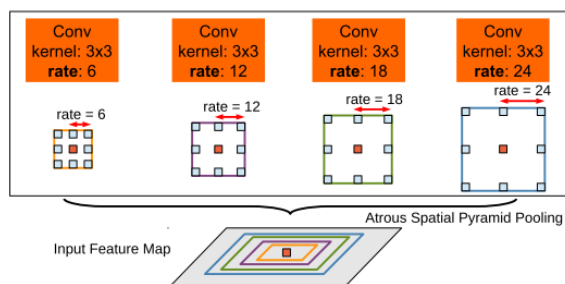


Figure 1. ASPP 示意圖

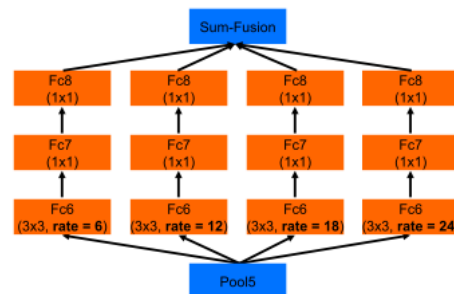


Figure 2. DeepLab V2 所提出的 ASPP 架構

Decoder 與 baseline 類似，為多層的 deconvolution (有更改 filter 的參數)；而 Upsample 使用 bilinear Interpolation，將經過 ASPP 的 feature map 直接放大。經過 Upsample 的 feature 與經過 Decoder 的 feature 最後會相加，成為最終 predict 的結果。

2. Reasons for Improvement

我的 model 能有比較好的表現主要有兩個原因：

1) ASPP

ASPP 在 DeepLab V2 被提出來時，是為了將一類別各種規模的 feature 都列入考慮，同時去除 scale 對於辨別物體的影響(Random scale 也在改善這個問題)。而在此次作業也是一樣，ASPP 能轉換 baseline model 得到的 feature map，考量不同層次的資訊。

下面為 baseline model 與改用 ASPP + bilinear interpolation 的 model(不加 deconvolution)之比較。可看到，ASPP + bilinear interpolation 的 val_iou 高了許多，loss 也收斂的快了許多。

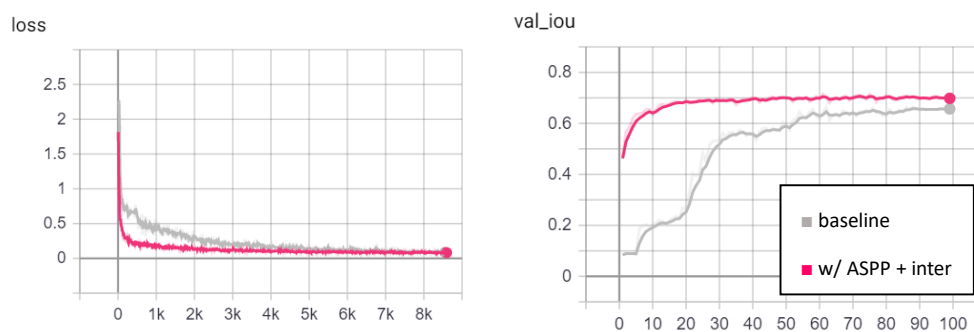


Figure. loss 與 val_iou 曲線

¹ Ref: <https://arxiv.org/abs/1606.00915>

2) Decoder 與 interpolation 並行的架構



Figure. Baseline(左) / ASPP + inter(右) 的 inference 結果

縱然 ASPP + bilinear interpolation 的準確率高了許多，但由上圖可見，baseline model 所得到的 inference 結果平滑許多，也比較能抓到細節。這是因為，我為了維持與 baseline 相近的架構，並沒有依照 DeepLab 的 paper 調整 Resnet 的架構，而這也導致了我最後必須使用 interpolation 將 ASPP 的結果放大 32 倍，也使得結果相當粗糙。

為了改善此結果，我決定合併使用 decoder 和 bilinear interpolation，讓 infer 出來的圖片能變得較不粗糙，又能保有 interpolation 快速收斂的特性，使兩者的優點互相加成。有趣的是，這樣的設計也有一些 Residual 的特性(DC + AC)。

下圖為 ASPP + bilinear interpolation + decoder (improved model)與前述兩個 model 的比較。可見 improved model 比起 ASPP + inter，val_iou 又能提升一些。

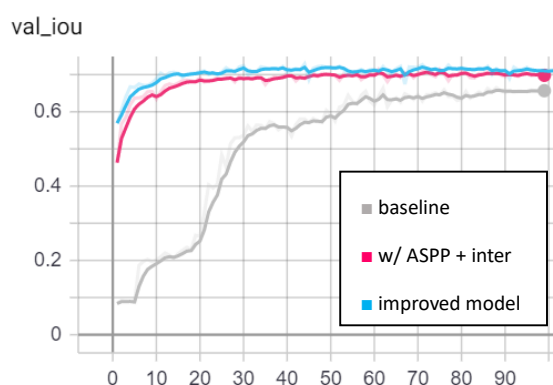


Figure. val_iou 曲線

3. Visualization



4. Results

- 1) mIOU score of validation set: 72.87%
- 2) mIOU score per class:

```
class #0 : 0.90760
class #1 : 0.75393
class #2 : 0.71014
class #3 : 0.79292
class #4 : 0.37528
class #5 : 0.73506
class #6 : 0.72343
class #7 : 0.80046
class #8 : 0.75958
```

Problem 2

1. Proof:

$$\begin{aligned} & I(x, y) * G(x, y) \\ &= \iint I(x, y) \times \frac{1}{2\pi\sigma^2} e^{-\frac{(x-t)^2 + (y-\tau)^2}{2\sigma^2}} dt d\tau \\ &= I(x, y) \times \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-t)^2}{2\sigma^2}} dt \times \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\tau)^2}{2\sigma^2}} d\tau \\ &= I(x, y) * G(x) * G(y) \end{aligned}$$

2.

- 1) Effect of the filter: to blur the image
- 2)



Original



Filtered

3.

$$1) \begin{aligned} k_x &= \begin{bmatrix} -0.5 & 0 & 0.5 \end{bmatrix} \\ k_y &= \begin{bmatrix} -0.5 & 0 & 0.5 \end{bmatrix} \end{aligned}$$

2)



I_x



I_y

4.



I_m of Original



I_m after filter