Computer Vision HW4 Report

Q1

Before Correcting the Augmentation Strategy

['heavy_makeup', 'no_heavy_makeup', 'no_heavy_makeup', 'heavy_makeup']
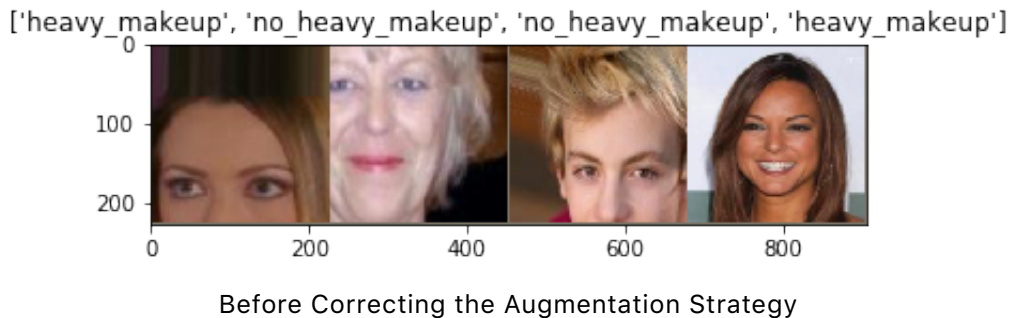

Before Correcting the Augmentation Strategy

1. Please report the validation accuracy of a pretrained Alexnet used as a feature extractor in the two-class classification problem. (5 pts)
   Ans.
   ```
   Best val Acc: 0.810000
   ```

2. Please report the validation accuracy of a pretrained Alexnet after it is fine-tuned in the two-class classification problem. (5 pts)
   Ans.
   ```
   Best val Acc: 0.875000
   ```

3. Please report the validation accuracy of a non-pretrained Alexnet after it is trained in the two-class classification problem. (5 pts)
   Ans.
   ```
   Best val Acc: 0.672500
   ```

4. Please discuss the results of Q1-1, Q1-2, & Q1-3. (5 pts)
   Ans.
   從 Q1-1、Q1-2、Q1-3 中可以看出,有預訓練且做 fine-tune 的準確率較高、有預訓練只訓練最後一層的準確率次之、而從頭訓練起來的準確率最低,這是因為經過預訓練的模型已經能夠對圖片做分類,也因此可以更正確的判斷出圖像中的人物部分,使得其判斷人物有沒有 heavy makeup 可以更準確。相較之下,從頭訓練起的模型在抽取人像的效果可能較差,這就是造成有預訓練和沒預訓練準確率差距的原因。
   而只訓練最後一層的準確率比做 fine-tune 的準確率差,則是因為預訓練得到的前面幾層對於特徵的抽取可能較不適用於現在的 dataset,但經過 fine-tune 調整後,既可以保留預訓練對於基本特徵的抽取效果,也能更 fit 現在的 dataset 所以準確率會比較高。

After Correcting the Augmentation Strategy

['heavy_makeup', 'no_heavy_makeup', 'no_heavy_makeup', 'heavy_makeup']

After Correcting the Augmentation Strategy

5. Please try to correct the data augmentation strategy in order to let the entire face of each image be seen and report the validation accuracy of a pre-trained Alexnet as a feature extractor in the two-class classification problem. (5 pts)
   Ans.

   ```
   Best val Acc: 0.805000
   ```

6. Please try to correct the data augmentation strategy in order to let the entire face of each image be seen and report the validation accuracy of a pretrained Alexnet after it is fine-tuned in the two-class classification problem. (5 pts)
   Ans.

   ```
   Best val Acc: 0.897500
   ```

7. Please discuss the results of Q1-5 & Q1-6. (5pts)
   Ans.

   原本助教提供的 augmentation strategy 是對圖片做隨機的水平翻轉和 random crop，在 random crop 這部分會導致切出來的圖有可能沒有包含完整的臉部資訊，使得神經網路在學習人像是否有 heavy makeup 時的效果較差。

   在將 augmentation strategy 換成對圖片做45度角內隨機旋轉再做 center crop 後，並保留原本隨機水平翻轉的部分，即可讓臉部完整出現在 dataset 的圖片中，使得神經網路學習的圖較符合我們要的結果。

   從 Q1-5 和 Q1-6 的結果可以看出使用修改過的 augmentation strategy，使得模型對 validation dataset 的準確度有稍微改善。

8. Please try to achieve validation accuracy higher than 89.5% using a CNN other than Alexnet & ResNet-18 in the fine-tuning case. (20pts)
   Ans.

   用 VGG16 達到的 validation set 準確率為：

   ```
   Best val Acc: 0.915000
   ```

9. Please discuss the results of Q1-9 (5pts if your meet the requirement of Q1-8)
   Ans.
   使用經過預訓練的 VGG16 模型，由於相較於一開始的 alexnet，VGG16 模型架構的深度更深，可能可以讓特徵抽取的效果比 alexnet 來的好，近一步使得訓練出來的準確率更高。
   但也因為深度更深，fine-tuning 時所要調整的參數量也較多，代價就是訓練時間較長，使用 alexnet 的架構可以在 6 分鐘內跑完 24 次 epoch；相較之下使用 VGG16 的架構則需耗時 12 多分鐘才能跑完 24 次 epoch。

Q2

1. Please try to "eliminate" the skip-connection so the output of convolution layers of FCN8s will be directly upsampled for 32x. Please report pixel accuracy and mIOU before and after. (10 pts)

```
# with skip-connection
The highest mIOU is 0.407460345745489 and is achieved at epoch-20
The highest pixel accuracy is 0.8508900451660156 and is achieved at epoch-1

# without skip-connection
The highest mIOU is 0.37736843620731103 and is achieved at epoch-20
The highest pixel accuracy is 0.8194914245605469 and is achieved at epoch-2
```

2. Please discuss the results of Q2-1. (10 pts)
   Ans.
   從 Q2-1 的結果中可以看出，有 skip-connection 的 mIOU 和 pixel accuracy 有比較高一些。因為在做 upsampling 的時候，是拿較少的資訊下去做 deconvolution 來放大，這時候相較於沒用 skip-connection 的少量資訊，用 skip-connection 拿放大後那層的資訊下去補，就可以獲得較好的效果。

3. Please try to further reduce the number of classes from 11 to 3 and report the pixel accuracy & mIOU of FCN8s. (10 pts)

```
# with skip-connection
The highest mIOU is 0.8745747501613662 and is achieved at epoch-18
The highest pixel accuracy is 0.9748773193359375 and is achieved at epoch-1

# without skip-connection
The highest mIOU is 0.8268486167447682 and is achieved at epoch-20
The highest pixel accuracy is 0.9640892028808594 and is achieved at epoch-2
```

4. Please discuss the results of Q2-3. Was mIOU increased when the number of classes reduce? Please explain why! (10 pts)

   Ans.

   是的,當只分成三類時,mIOU的確較分成11類時來的高。mIOU 的公式如下:

   $$mIoU = \frac{1}{N} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}}$$

   其中 n_{ij} 為class i 被當成 class j 的 pixel 數;$t_i = \sum_j n_{ij}$;N 為總 class 數。

   公式算的是 true positive 佔 false positive 連集 false negative 連集 true positive 的比例。

   當總分類數較少時,N 變小,所以 mIOU 會提升。另一方面,總分類數較少時,也會使得神經網路的學習成本下降,學習效果較好。