Deep Learning for Computer Vision HW#2

B05901182 電機四 潘彥銘

Problem 1: (100%)

Collaboration:

B05901027 詹書愷、B05602042 林奕廷、B05901074 陳泓均、B05901060 賴繹文

Reference:

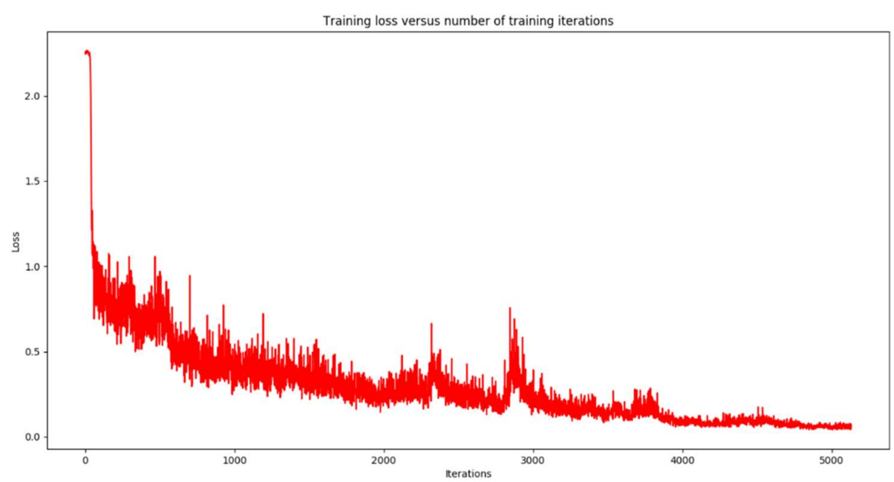
https://arxiv.org/pdf/1802.02611.pdf

Ans:

1. Baseline model:

 1. (5%) I only normalize the data. (by using transforms.Normalize(mean, std))

 2-1. (5%)



 2-2. (5%)

3. (5%)

Class	0
Semantic segmentation result	
Class	1
Semantic segmentation result	
Class	2
Semantic segmentation result	

Class	3
Semantic segmentation result	
Class	4
Semantic segmentation result	
Class	5
Semantic segmentation result	

Class	6
Semantic segmentation result	
Class	7
Semantic segmentation result	
Class	8
Semantic segmentation result	

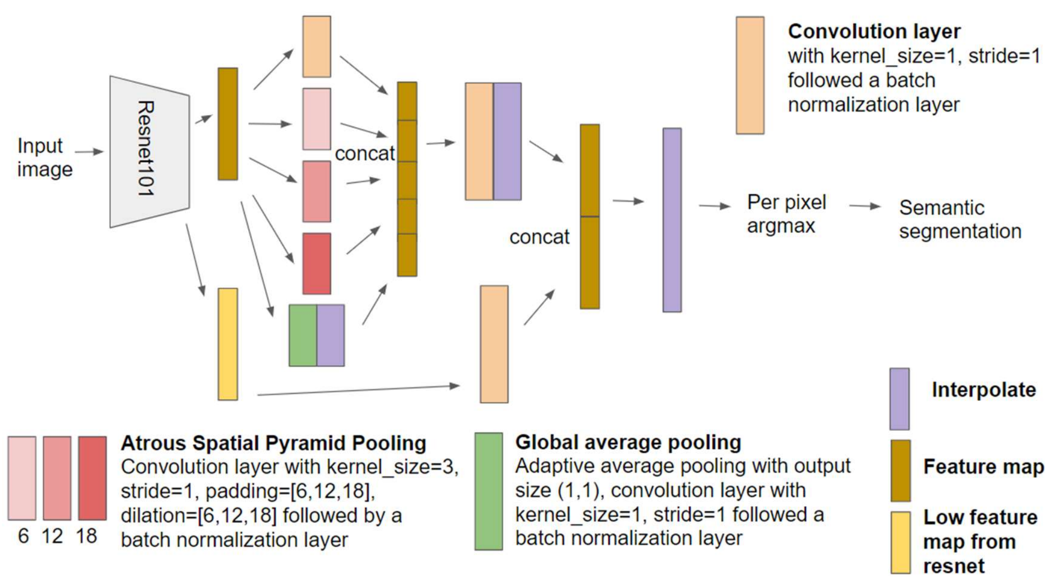
4. (10%)

Class	IoU
0	0.89884
1	0.74058
2	0.63659
3	0.70647
4	0.37073
5	0.47503
6	0.60708
7	0.73580
8	0.63751

- mIoU = 0.645404
- Class 0 has the highest IoU score and class 4 lowest IoU score.
- Maybe it's because the class 4(tv/monitor) has the fewest data and the class 0(background) has the greatest number of data (almost every image has the foreground and the background, or pure background).

2. Improved model:

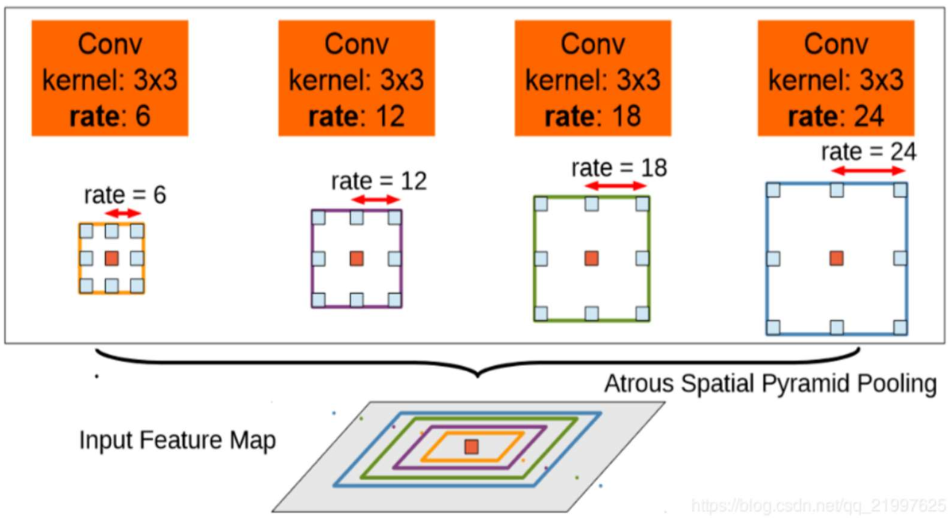
1. (5%) I use the model structure of DeepLabV3+



2. (15%)

- Compared to my baseline model, whose feature extractor is resnet18, my improved model uses resnet101 as its feature extractor. As several experiments have already done, resnet101 has better performance than resnet18, though the computing time is more longer.

- In my improved model, I use a series of special convolution called Atrous Spatial Pyramid Pooling(ASPP, illustrated below) to extract features. Compared to the traditional convolution, the stride of the atrous convolution is larger which can extract more “wider” features. And by deploying the different stride of convolutions, the model can extract multi-scale features

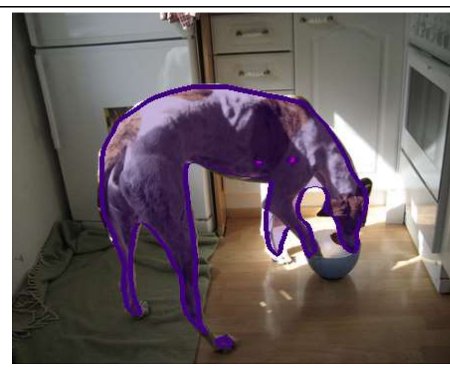


Ref: https://blog.csdn.net/qq_21997625/article/details/87080576

- Compared to the baseline model, which uses the constant learning rate, I use the learning rate scheduler to change my learning rate(original learning rate*0.6)each 4 epochs because at first I noticed that the loss in the training process is hard to decrease after a few of epoches during training process.

3. (10%)

- mIoU = 0.752829

0001.png	
Baseline model	Improved model
	
0014.png	
Baseline model	Improved model
	
0152.png	
Baseline model	Improved model
	

0165.png	
Baseline model	Improved model
	

Problem 2: (10%)

Collaboration: B05901027 詹書愷 B05901060 賴繹文

Ans:

1. (1%)

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} = G(x) * G(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} * \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}}$$

2. (3%)

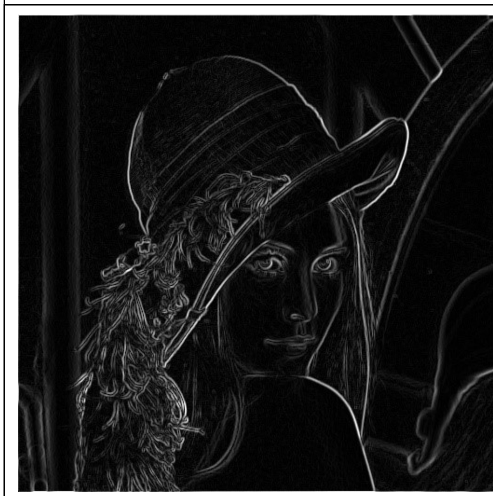
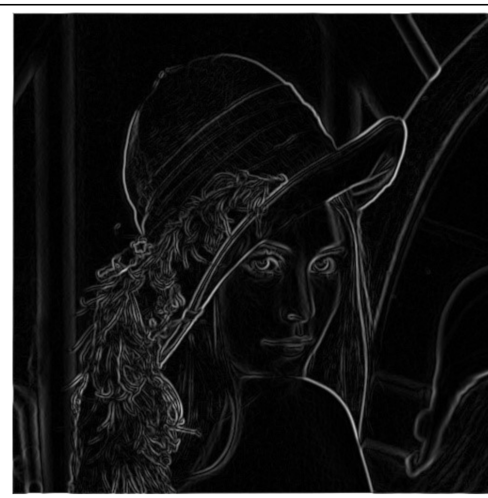
The effect of 2D Gaussian filter can blur the image.

3. (4%)

$$k_x = [-0.5, 0, 0.5] \qquad k_y = [-0.5, 0, 0.5]^{\mathrm{T}}$$

I_x	I_y
	

4. (2%)

lena.png	Gaussian filter image
	

The left image’s egdes are more obvious than that of the right image, and the right image is more blurrer than that of the left image.