

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:

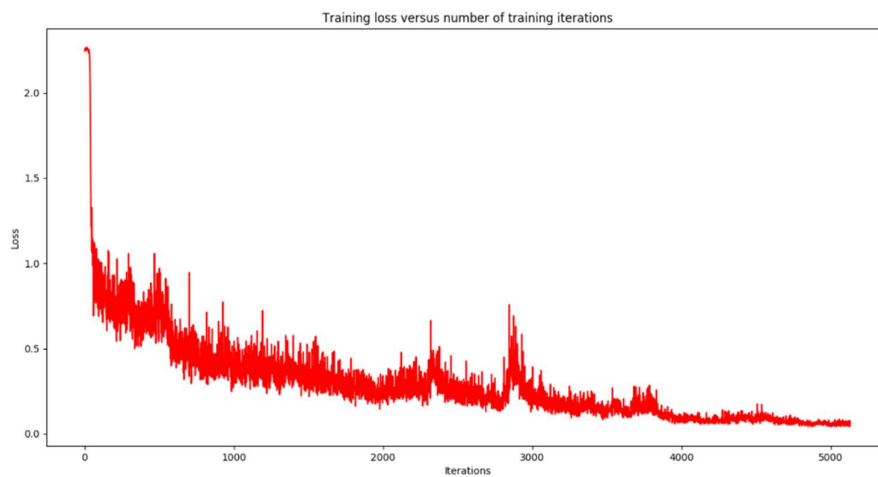
<https://arxiv.org/pdf/1802.02611.pdf>

Ans:

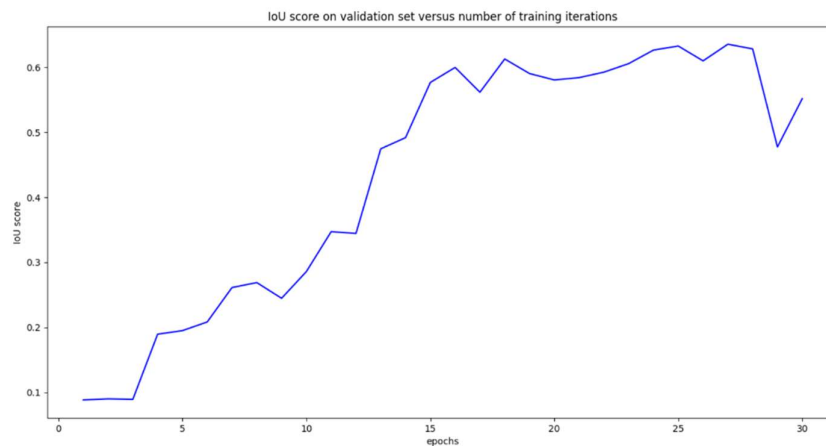
1. Baseline model:

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	 A street scene in London featuring two red double-decker buses. The bus in the foreground is a Routemaster, and the one behind it is a standard double-decker. A building with the sign 'THE VICTORIA' is visible in the background. The image shows the result of semantic segmentation with green outlines highlighting the buses and the building.
Class	4
Semantic segmentation result	 A computer monitor with a blue frame and a small pink bunny on top. The monitor is on a desk, and there are various items on the desk, including a pair of scissors and a small container. The image shows the result of semantic segmentation with a blue outline highlighting the monitor.
Class	5
Semantic segmentation result	 Two horses standing in a field with trees in the background. The horse on the left is white, and the horse on the right is brown. The image shows the result of semantic segmentation with pink outlines highlighting the horses.

Class	6
Semantic segmentation result	 A photograph of a dog standing on a wooden floor, leaning over a blue bowl to eat. The dog's body is outlined with a thick, semi-transparent purple mask, indicating the result of a semantic segmentation process. The background shows a kitchen with white cabinets and a window.
Class	7
Semantic segmentation result	 A photograph of three cats resting together. The cats are outlined with a thick, semi-transparent red mask, representing the semantic segmentation result. The background is a solid red color.
Class	8
Semantic segmentation result	 A photograph of a white sports car, likely a Toyota Camry, parked on a road. The car is outlined with a thick, semi-transparent yellow mask, showing the semantic segmentation result. The car has various sponsor stickers, including 'ADVAN' and 'Powershift'.

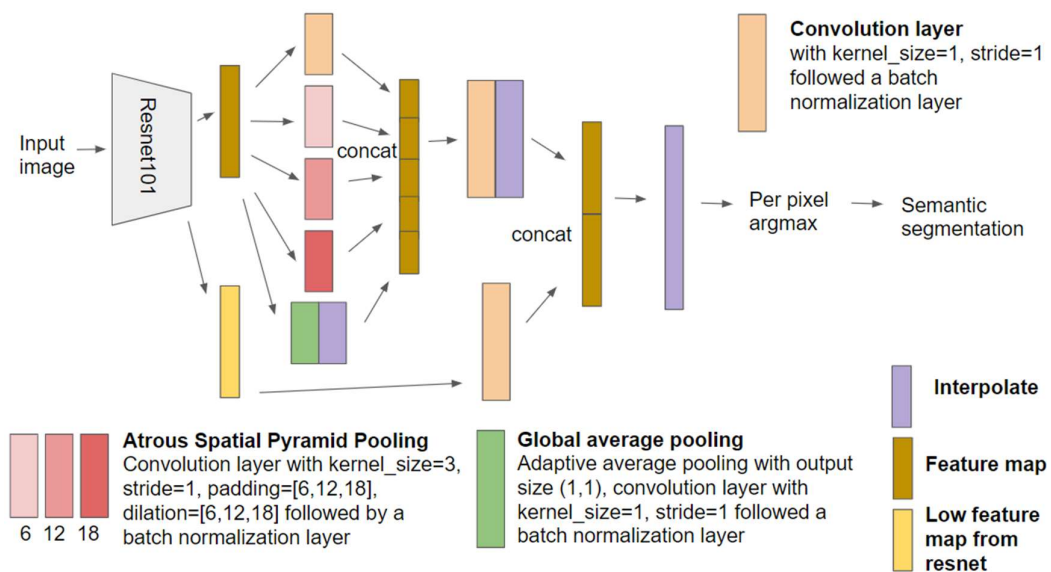
4. (10%)

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).

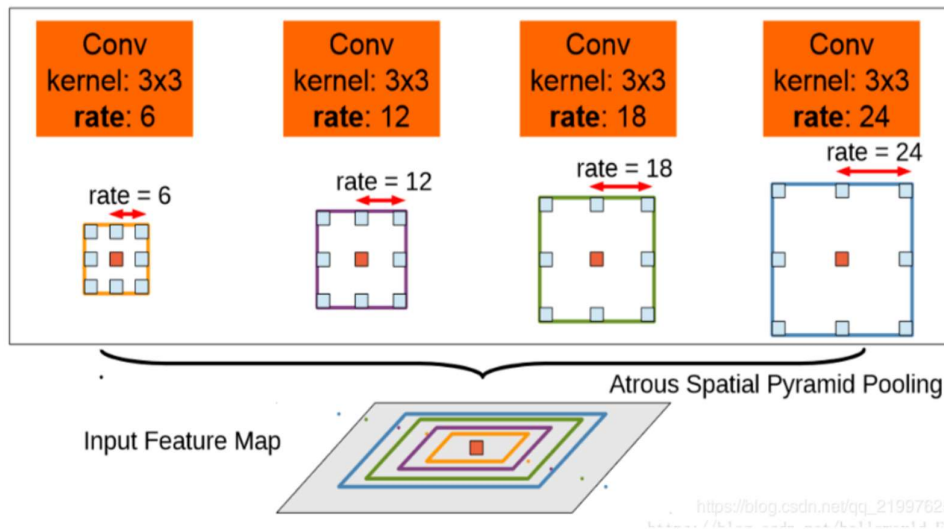
2. Improved model:

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%)

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

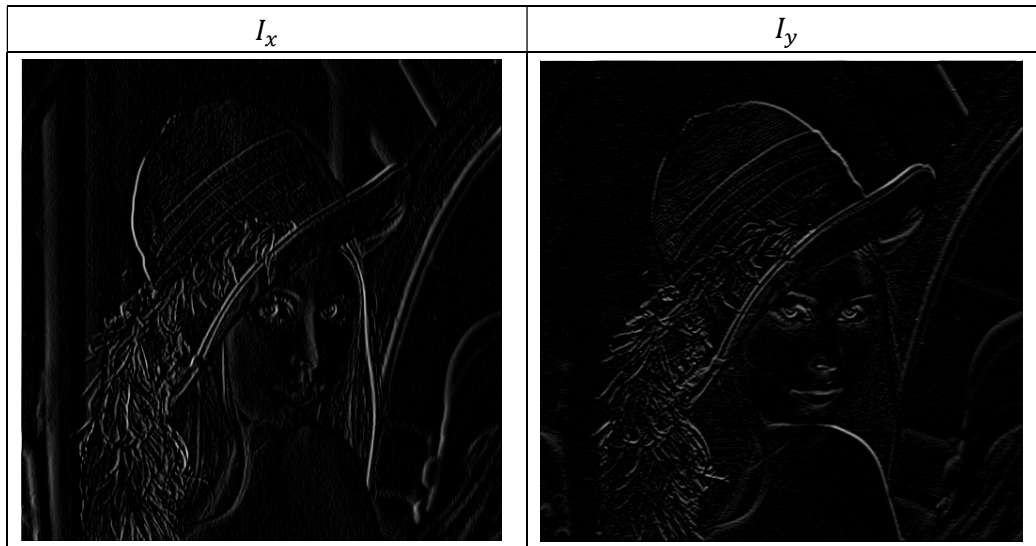
2. (3%)



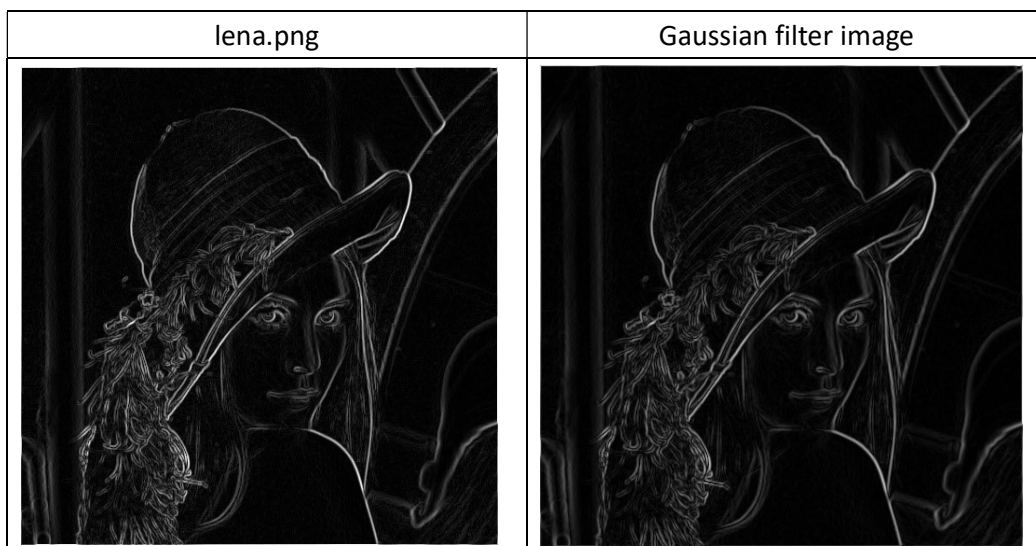
The effect of 2D Gaussian filter can blur the image.

3. (4%)

$$k_x = [-0.5, 0, 0.5] \quad k_y = [-0.5, 0, 0.5]^T$$



4. (2%)



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