CSCE 636 Homework 3

Solutions Keys

November 2019

1

Rubrics: You are expected to achieve the accuracies higher than 90% for both v1 and v2. Considering the limited time and computational resources you have, no deductions are applied if your accuracies are higher than 80%.

2

For standard residual block with $128 \times 16 \times 16 \times 32$ inputs and outputs, there are two 3×3 convolution layers, each contains 32 filters with 32 channels, so the total number of parameters is $2 \times (32 \times 3 \times 3 \times 32) = 18432$.
When considering batch normalization: there are 2 batch normalization layers, each contains 32×2 parameters. The total number of parameters is $2 \times (32 \times 3 \times 3 \times 32 + 32 \times 2) = 18560$.
You can also consider bias as long as your computation is correct.

For bottleneck residual block with $128 \times 16 \times 16 \times 128$, there are one 1×1 convolution layer containing 32 filters with 128 channels, one 3×3 convolution layer containing 32 filters with 32 channels, and one 1×1 convolution layer containing 128 filters with 32 channels. So the total number of parameters is $32 \times 128 + 32 \times 3 \times 3 \times 32 + 128 \times 32 = 17408$.
When considering batch normalization: there are three batch normalization layers whose number of parameters is two times of input channels. So the total number of parameters is $32 \times 128 + 32 \times 2 + 32 \times 3 \times 3 \times 32 + 32 \times 2 + 128 \times 32 + 128 \times 2 = 17792$.
You can also consider bias as long as your computation is correct.

We can find that the bottleneck residual block has roughly the same number of parameters as the standard residual block. The advantage of bottleneck residual block is that it can stack a deeper network with less parameters than the standard residual block, and speed up the computing because 1×1 convolution is faster than 3×3 convolution. But the outputs from the bottleneck block have more channels, resulting the increasing of the number of parameters in the softmax layer, Also, the bottleneck block doesn't enlarge the receptive field, which is still 3×3. However, the receptive field of using two convolutional layers in the standard residual block is 5×5.

3

(a)The shape of the mean and variance are both $1 \times C$.
(b)The shape of the mean and variance are both $1 \times 1 \times 1 \times C$.

4

(a)

$$A = \begin{bmatrix} \omega_1^{11} & \omega_2^{11} & \omega_3^{11} & 0 & \omega_1^{21} & \omega_2^{21} & \omega_3^{21} & 0 \\ 0 & \omega_1^{11} & \omega_2^{11} & \omega_3^{11} & 0 & \omega_1^{21} & \omega_2^{21} & \omega_3^{21} \\ \omega_1^{12} & \omega_2^{12} & \omega_3^{12} & 0 & \omega_1^{22} & \omega_2^{22} & \omega_3^{22} & 0 \\ 0 & \omega_1^{12} & \omega_2^{12} & \omega_3^{12} & 0 & \omega_1^{22} & \omega_2^{22} & \omega_3^{22} \end{bmatrix}$$

(b)

$$\frac{\partial L}{\partial \widetilde{X}} = \begin{bmatrix} \omega_1^{11} & 0 & \omega_1^{12} & 0 \\ \omega_2^{11} & \omega_1^{11} & \omega_2^{12} & \omega_1^{12} \\ \omega_3^{11} & \omega_2^{11} & \omega_3^{12} & \omega_2^{12} \\ 0 & \omega_3^{11} & 0 & \omega_3^{12} \\ \omega_1^{21} & 0 & \omega_1^{22} & 0 \\ \omega_2^{21} & \omega_1^{21} & \omega_2^{22} & \omega_1^{22} \\ \omega_3^{21} & \omega_2^{21} & \omega_3^{22} & \omega_2^{22} \\ 0 & \omega_3^{21} & 0 & \omega_3^{22} \end{bmatrix} \frac{\partial L}{\partial \widetilde{Y}}$$

The relationship is $B = A^T$.
(c)It can been as the convolution on padded $\frac{\partial L}{\partial Y}$, that is

$$\begin{bmatrix} 0 & 0 & \frac{\partial L}{\partial y_{11}} & \frac{\partial L}{\partial y_{12}} & 0 & 0 \\ 0 & 0 & \frac{\partial L}{\partial y_{21}} & \frac{\partial L}{\partial y_{22}} & 0 & 0 \end{bmatrix}$$

The kernel can be write as $U^{ij} = [\omega_3^{ji}, \omega_2^{ji}, \omega_1^{ji}], i = 1, 2, j = 1, 2$, which scans the i-th channel of padded $\frac{\partial L}{\partial Y}$ and contributes to the j-th channel of outputs.