CID : 01752815

Q1.

	Number of output neurons	Number of parameters
Conv1	145200	17472
MaxPool	34992	\
Conv2	86528	221312
MaxPool	18432	\
Conv3	27648	614592
Conv4	23232	590016
Conv5	15488	221312
MaxPool	2048	\
FC1	\	3278400
FC2	\	2561600
Softmax	\	16000

Total Neurons: 357768

Total Parameters : ~~9105704~~ 9105704

Why 224x224 to 227x227?

Answer : To make all the size of output image to be integers

Q2:

The first improvement is applying PReLU as activation function.

The PReLU sets the coefficient that controls negative derivatives as Learnable parameter. Meanwhile, it introduced a very small number of extra parameters, which is negligible when considering total number of weights. But it can improve the performance of networks

Comparison:

On the one hand, he conducted comparisons on a deep but efficient model with 14 weight layers based on a known paper. By comparing the top-1 and top-5 error rate, it can basically ensure the PReLU is better than ReLU (1.1% gain).

On the other hand, based on 1000-class ImageNet 2012 dataset. He set the same total number of epochs, and the learning rates are also switched after running the same number of epochs. After analyze the table of erro rate, it can justify that PReLU improves both small and large models.

Q3.

The improvement is setting the initialized weight by zero-mean Gaussian distribution whose standard deviation is $\sqrt{\frac{2}{n_l}}$ for ReLU method. As for PReLU method, the standard deviation is $\sqrt{\frac{2}{n_l(1+a^2)}}$.

The improvement is aimed to equip the networks with a robust initialization method which can remove the obstacle of traing extremely deep rectifier networks.

As for Glorot and Bengio, they applied the uniform distribution with range of $[-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}]$, which is called "Xavier" method. But its derivation is based on assumption that the activations are linear which is invalid for ReLU and PReLU.

As for Alex paper, they applied zero-mean Gaussian distribution with standard deviation 0.01 in each layer. But it has difficult to converge when neural networks are extremely deep.

Comparison.

In He et al paper, they compared the improved initialization method with that of "Xavier" on extremely deep models with 22 and 30 la There is not much difference for 22 layers. But for 30 layers, the initialization is able to make the model converge. On the contrary, "Xavier" method completely stalls the learning, and the gradients are diminishing as monitored in the experiments. But both method have the similar accuracy.

Q4

As for comparisons between ReLU and PReLU, they compared ReLU and PReLU on the large model A with same setting. And they found that PReLU improves both small and large models.

As for comparisons of single-model results, they compared their five models (model A with ReLU and eta.) with VGG-16 and GoogLeNet. And their baseline model outperforms VGG-19.

As for comparisons of multi-model results, they combined six models and this model shows 1.7% better than ILSVRC 2014 winner.

As for comparisons with human performance, their result (4.94%) exceeds the reported human-level performance from Russakovsky based on ImageNet.