

ACSE - Module 8

Coursework 1, May 7th

The following paper currently has about 80,000 citations:

Krizhevsky A., Sutskever I., Hinton G., 2012. ImageNet Classification with Deep Convolutional Neural Networks. In NIPS Proceedings, 2012.

The network presented in this paper is the famous AlexNet (from the first author's first name), which changed the history of Convolutional Neural Networks, and of Artificial Intelligence, in 2012. The convolutional neural network produced by AlexNet could beat by a large margin all other image classification programs in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). The network classifies about 1.2 million images into 1,000 classes.

- Using the attached spreadsheet below, you are asked to calculate the total number of parameters and of neurons. We use a simplified version of AlexNet ignoring the split between two GPUs, meaning that you will not get the same number of parameters and neurons as the paper. You also need to explain first why the number of inputs in the input layer should be 227x227 instead of the 224x224 figure given by AlexNet (25 marks).

	Size of input image	Number of input channels	f	p	s	Size of output image	Number of output channels or filters	Number of output neurons	Size of Filter + 1	Number of Parameters
Conv1	227	3	11	0	4		48			
MaxPool			3	0	2		48			
Conv2			5	2	1		128			
MaxPool			3	0	2		128			
Conv3			5	2	1		192			
Conv4			3	1	1		192			
Conv5			3	1	1		128			
MaxPool			3	0	2		128			
	Size of input							Number of output neurons		
FC1								2048		
FC2								2048		
Softmax								1000		
							Total Neurons		Total Parameters	

- What is for the authors the main advantage of CNNs over feed-forward neural networks, and what is the main remaining obstacle to using even larger CNNs? (5 marks)
- What is the reason for choosing the ReLU function rather than a sigmoid or other activation functions ? (5 marks)
- Explain what is the last layer of the network and what it does exactly (10 marks).
- What is the loss function used to optimize the weights? Express it as a formula (10 marks).
- At the beginning of Paragraph 4, the authors write that "Although the 1000 classes of ILSVRC make each training example impose 10 bits of constraint on

the mapping from image to label, this turns out to be insufficient to learn so many parameters without considerable overfitting". Explain where these 10 bits come from (5 marks).

7. How are the different weights initialized (10 marks)?
8. Do you agree with the fact that the "first form" of data augmentation increases the training set by a factor of 2048 (5 marks)?
9. What was the batch size and the rough number of epochs used for the training (10 marks)?
10. Some Model Averaging was used to obtain the final results. Then, Transfer Learning (although the terminology was different) was performed to try improving these results. Explain how it was done and with which results for the Top-5 error rate (15 marks).

Total: 100 marks