Coursework May 5th – To be returned on May 11th before 9.00 am.

The following paper has more than 61,000 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 paper showed that a convolutional neural network with 60 million parameters and 650,000 neurons 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 categories. Question 1 below uses a Convolutional Neural Network inspired from that of the AlexNet paper, *but with slightly different parameters to make the question a bit more challenging, meaning that you will not get the same number of parameters and neurons as the AlexNet paper.*

Answer the following questions about the paper:

1. To simplify, assume that there is only one GPU processing the sequence represented at the bottom layer of Figure 2 of the AlexNet paper and as summarized in the table below. For each layer of the sequence, the size f of the convolution window, the padding p and the stride s are also given in the table. First, explain why the size of the input image should be changed from 224x224 (as it is given in Figure 2 of the AlexNet paper) to 227x227. Use this number in the following calculations, which consist of calculating the nine values in the orange cells and the nine values in the yellow cells. You can use the attached unfilled spreadsheet but do not include it in your final answer, just clearly list the values of the orange and yellow cells. Assume that all the convolutional and fully-connected layers have a bias term. (40 points)

	Size of input image n	Number of input channels	f	p	s	Size of output image (n+2p-f)/s+1	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			6	2	1		128			
MaxPool			4	0	2		128			
Conv3			5	2	1		192			
Conv4			4	1	1		192			
Conv5			3	1	1		128			
MaxPool			4	0	3		128			
	Size of input							Number of output neurons		
FC1								1600		
FC2								1600		
Softmax								1000		
							Total Neurons		Total Parameters	

Formula used for last column has been given in the morning course:

Number of Parameters= Number of output channels × (Size of filter+1) (as we assume there is a bias term)

Three years after the AlexNet paper, the paper *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, by He et al, was published (more than

7,700 citations). It proposed two significant improvements to the approach followed by AlexNet.

2. Explain in less than 15 lines what was the first proposed improvement, which was about the activation function used. How did it compare to the AlexNet approach and what was its justification? (20 points)

The second improvement proposed by the *He et al* paper was about the initialization of the weights (you can ignore the bias parameters from now) of a convolutional neural network such as Alexnet. The paper compared their new initialization approach with that of the AlexNet paper and that of another classic paper from 2010 by by Glorot and Bengio: *Understanding the Difficulty of Training Deep Feedforward Neural Networks* (more than 8,400 citations), which focussed only on feed-forward networks with a linear activation function.

3. Explain in less than 25 lines what is this second improvement, how different it is from the approaches of AlexNet and of Glorot and Bengio, and what is its justification. (20 points)

4. Explain in less than 10 lines how He and al quantify the performance of their two above improvements on ImageNet as compared to the Alexnet approach . (20 points)