

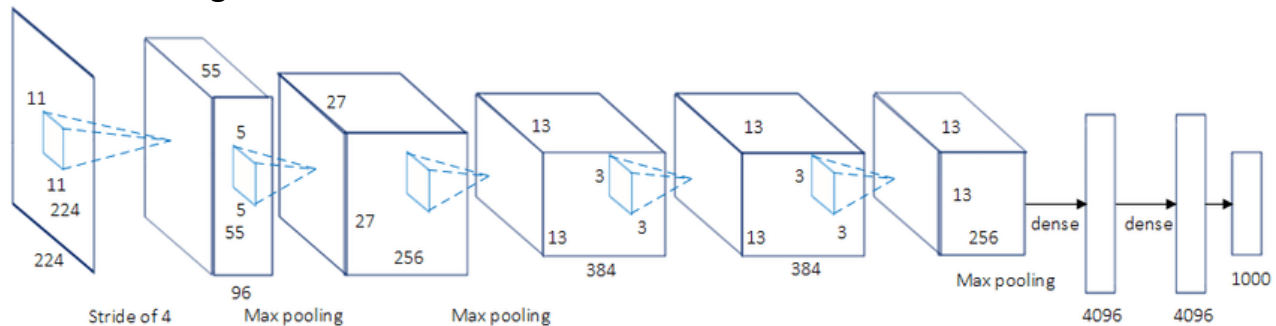


- 1- Assume a neural network with a single hidden layer and a soft-max at the output layer. Derive the gradient descent optimization update equation for the network parameters.
- 2- Write a pseudo code algorithm for the following techniques:-
 - a. Stochastic gradient descent optimization (SGD).
 - b. Min-batch stochastic gradient descent optimization.
 - c. Momentum based gradient descent optimization.
 - d. Nesterov momentum based gradient descent optimization.
 - e. AdaGrad gradient descent optimization.
 - f. RMSProp gradient descent optimization.
 - g. RMSProp with Nesterov gradient descent optimization.
 - h. AdaDelta gradient descent optimization.
 - i. Adam gradient descent optimization.
- 3- Describe the behavior of SGD when dealing with a saddle point. Compare with the other modern techniques.
- 4- Compare the performance of the modern gradient descent optimization techniques against the SGD.
- 5- Derive the parameters update rule using the NR-GD technique. Is this approach practical in training DNNs? Justify your answer. What are the algorithms that overcome the limitation of such a technique?
- 6- Describe how to overcome the vanishing gradients problem? How does this problem affect the training of neural networks?
- 7- What is the overfitting problem? How to avoid such a problem when training neural networks?
- 8- What is the relation between ReLUs and drop out?
- 9- How to obtain the ideal structure of your neural network?
- 10- List some optimization techniques that avoids the fixed learning rate problem.
- 11- How to initialize training parameters?

- 12- What are the main three properties of CNN? Which part of the network is related to which property?
- 13- How can an object be recognized under different scales using a CNN network?
- 14- For the VGGNET structure given below, assume that the filter size is 3X3 in the convolutional layers with a stride and a padding amount of 1. Discuss and calculate the size of each stage.



- 15- The following CNN architecture is called AlexNet:-



- 16- Discuss the network stages different structures listing i) filter sizes, ii) stride and padding amounts, and iii) max-pooling sizes.