

## AlexNet and VGGNet

**Paper 1:** AlexNet: ImageNet Classification with Deep Convolutional Neural Networks

**Paper 2:** VGGNet: Very Deep Convolutional Networks for Large-Scale Visual Recognition

1. What are the assumptions of convolutions presented in paper 1. Describe each of these assumptions briefly.
2. Why CNNs are easier to train as compared with fully connected networks? In paper 1, why does the author say, "their (CNNs) theoretically-best performance is likely to be only slightly worse"? explain.
3. Why is ReLU a non-saturating\* non-linearity? What are its advantages over other activation functions like sigmoid, tanh, and soft sign. How it helps in training deep neural network models?
4. Define briefly (in context of paper 1): (a) Local Response Normalization, (b) Overlapping pooling, (c) Data augmentation, (d) Dropout, (e) top-1 accuracy, and (f) top-5 accuracy
5. Explain the sentence, "initializing biases with 1 instead of 0 accelerates the early stages of learning in deep model". Also, how do we choose, when to initialize biases with 1 or 0?
6. How can you design an effective receptive field of 5x5 and 7x7 using only 3x3 convolutional layers? Why there should not be spatial pooling in between these convolutional layers?
7. Explain, "The incorporation of 1x1 convolutional layers is a way to increase the non-linearity of the decision function without affecting the receptive fields of the convolutional layers?"
8. In paper 1 and paper 2, while training, the learning rate is reduced by some constant factor once validation error stops improving. Why?
9. According to paper 2, how does using a small filter size (3x3) all across the network helped them in training?