Computer Vision Discussion Quiz – 2

1. Answer:

   Paper 1's assumptions about convolutions are local pixel interdependence and image statistics. The local pixel dependencies assumption is that nearby pixels are frequently associated, but remote pixels are usually not. This is the local pixel dependency assumption that's been made in paper 1. And, the other assumption is image statistics, this includes to identify the same feature in various parts of the image we could use the same filter. This assumption makes the image regardless of the time series. These are the two assumptions that are been made in the paper 1.

2. Answer:

   Computational Neural Networks are less difficult to train than fully connected networks as in CNNs we move a filter, and CNNs have less connections compared to fully connected networks and every node relates to other nodes. This helps in reducing overfitting. With, CNNs we can detect every corner. When we apply pooling to CNNs it gives us location invariant feature detection. CNNs are computationally less costly and more efficient than fully connected networks as in CNNs there are only few parameters, so parameter sharing is feasible. CNNs are simpler to train as they contain fewer parameters and connections. When the author said about CNNs performance what he was referring to is that even though CNNs have so many advantages, when it comes to big scale CNNs are still extremely costly to apply to high-resolution photographs. But in current scenarios, GPUs, combined with a well-optimized 2D convolution implementation, are strong enough to allow for the training of huge CNNs.

3. Answer:

   ReLu, Rectified Linear Unit activation function takes a feature map and substitutes the negative value with zero. So, in this activation function as the limit of the function is absolutely in the positive range, this makes ReLu non-saturating function. And, as we replace all negative values with zero, the graph would bend near zero and cannot be a straight line all through the graph. So, ReLu is a non-linear activation function. ReLu's benefits over sigmoid, tanh and softsign activation functions are ReLu introduces non-linearity, with ReLu activation function the issue with the Vanishing Gradient has been resolved fully unlike in other activation functions. And as we replace negative values with zero in ReLu, its derivative is faster than other activation functions as the slope is either zero or positive. So, this speeds up the training of neural networks and makes the computation faster.

4. Answer:

   a. Local Response Normalization: When the activation function values of one feature map is very big than the activation values of other feature maps, then that feature map with big activation value is prioritized more and other feature maps become less important. So, to make the other feature maps also prioritized to some extent, we normalize the activation function values per pixel of all the feature maps. This kind of normalization is called as Local Response Normalization.

   b. Overlapping Pooling occurs when windows converge. So, let F be the size of the window and S be the value of stride, then if F value is less than S value then overlapping occurs. Overlapping pooling contributes more in reducing over-fitting, reduces the number of dimensions and computations. And, overlapping pooling makes the model tolerant towards small distortions.

c. When we overtrain the data or sometimes with training data we get the desired outputs with greater accuracy but fails to get desired outputs with test data. This problem is called over-fitting. Data Augmentation is a technique used to handle the over-fitting problem. In Data Augmentation, we generate new samples from the existing samples of data and raise the size of the data. For example, we can generate horizontal flip of the original image, rotation of the original image, contrast of the original image, zoom of the original image and few more.

d. Dropout regularization is also used to reduce over-fitting. In drop out regularization we randomly drop few neurons in one stage and train the remaining neurons and in next stage we drop from these previously trained neurons and train the remaining neurons in a random manner. This helps the neural network to work well not only with training data but also obtain good accuracy with test data. As the neurons might be dropped out randomly, they cannot rely on one input. This way neurons will not learn redundant details of input, and this helps in increasing the accuracy of testing with the test data.

e. Top-1 accuracy is nothing but the output from the model has to be similar to the desired output. This type of accuracy where the model provides best result on predicting the desired output and is hundred percent accurate. This kind of accuracy is called as Top-1 accuracy.

f. Top-5 accuracy is nothing but when any one of the five output predictions from the model is similar to the desired output. So, the desired output must be there in any of the five output predictions that the model makes. This kind of accuracy is referred to as Top-5 accuracy as we are considering top-5 predictions of the model.

5. When we initialize biases with 1, we supply positive inputs to ReLu, however when we execute a weighted sum on unlearned weights, the outcome might be tiny or negative at first. So here, we started biases at 1 during training, and the negative weight output values turn positive as the bias value is applied to them. And the ReLu output is positive for positive values. So, when we applied ReLu here, the neuron will not die, and the gradient value will not vanish. As a result, the training is more efficient. I suppose, initializing biases with 1 when using ReLu activation function makes the training efficient and for other activation functions initializing biases with 0 would suffice.

6. We could create both 5x5 and 7x7 with active receptive field using 3x3 convolutional layers, and the feature extraction that we do with the 5x5 layer could be done with two 3x3 convolutional layers. One 5x5 filter and two 3x3 filters yield the same feature extraction results. We check for 25 surrounding pixels with the 5x5 filter, and we check for 18 neighboring pixels with the two 3x3 filters. As a result, utilizing two 3x3 instead of one 5x5 is more efficient since the two 3x3 have a larger effective receptive field than one 5x5. In the same way, the feature extraction results of one 7x7 filter and three 3x3 filters are identical. And the 7x7 filter looks for 49 surrounding pixels, whilst the three 3x3 filters look for a total of 27 neighboring pixels. As a result, three 3x3 convolution filters have a more effective receptive field than one 7x7 convolution filter. There should be no spatial pooling between the 3x3 filters since the feature extraction output of 5x5 would be the same as two 3x3, but if we use spatial pooling between these 3x3 filters, the result would be different from 5x5 or 7x7. As a result, no spatial pooling should occur between these 3x3 convolutional layers.

7. The 1x1 convolution filter can increase the non-linearity in the data as the 1x1 convolution filter is followed by ReLu. ReLu substitutes the negative value with zero in the feature map, and this helps in increasing non-linearity. 1x1 convolution filter makes sure that each pixel has a linear transformation included in it. In this way 1x1 convolutional layers increases the non-linearity of the decision function without compromising the convolutional layers' receptive fields.

8. When the validation error quits improving during training, the learning rate is lowered by a constant factor, this could be to check if there are any other minimum values below that we are missing, as the global minima helps become more efficient and shouldn't miss that. So we lower the learning rate by constant to ensure that it does not bypass the global minimum.

9. Using 3x3 filter size throughout the network in training from paper 2, helped in efficiency of computation by lowering their costs and it helped in sharing weights which resulted in lowering back-propagation weights.

Citation: https://medium.com/analytics-vidhya/how-to-choose-the-size-of-the-convolution-filter-or-kernel-size-for-cnn-86a55a1e2d15

https://www.aitude.com/comparison-of-sigmoid-tanh-and-relu-activation-functions/

https://ai.stackexchange.com/questions/13683/what-is-the-meaning-of-stationarity-of-statistics-and-locality-of-pixel-depen