1. What are the advantages of a CNN for image classification over a completely linked DNN ?

Ans : CNN can be used to reduce the number of parameters we need to train without sacrificing performance — the power of combining signal processing and deep learning! But training is a wee bit slower than it is for DNN. LSTM required more parameters than CNN, but only about half of DNN.The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. For example, given many pictures of cats and dogs it learns distinctive features for each class by itself. CNN is also computationally efficient.VGG16 is a pre-trained CNN model which is used for image classification. It is trained on a large and varied dataset and fine-tuned to fit image classification datasets with ease.

2. Consider a CNN with three convolutional layers, each of which has three kernels, a stride of two, and SAME padding. The bottom layer generates 100 function maps, the middle layer 200, and the top layer 400. RGB images with a size of 200 x 300 pixels are used as input. How many criteria does the CNN have in total? How much RAM would this network need when making a single instance prediction if we're using 32-bit floats? What if you were to practice on a batch of 50 images ?

Ans :parameters

first convolutional layer kernel-size and RGB channels, plus bias: 3 \* 3 \* 3 + 1 = 28 output feature maps is 100: 28 \* 100 = 2800

second convolutional layer kernel-size and last feature maps, plus bias: 3 \* 3 \* 100 + 1 = 901 output feature maps is 200: 901 \* 200 = 180200

third convolutional layers kernel-size and last feature maps, plus bias: 3 \* 3 \* 200 + 1 =1801 output feautre maps is 400: 1801 \* 400 = 720400

Total parameters is 2800 + 180200 + 720400 = 903400

memories since 32-bit is 4 bytes

first convolutional layer one feature map size: 100 \* 150 = 15000 total output: 15000 \* 100 = 1,500,000

second convolutional layer one feature map size: 50 \* 75 = 3,750 total output: 3750 \* 200 = 750,000

third convolutional layer one feature map size: 25 \* 38 = 950 total ouput: 950 \* 400 = 380, 000

(1,500,000 + 750,000 + 380,000) \* 4 / 1024 /1024 = 10.032 (MB) 903400 \* 4 / 1024 / 1024 = 3.44 (MB) 10.032+ 3.44=13.47(MB)

3. What are five things you might do to fix the problem if your GPU runs out of memory while training a CNN ?

Ans : reduce mini-batch size

reduce dimensionality using a larger stride in one or more layers

remove one or more layers

using 16-bits instead of 32-bit floats

distributed the cnn across multiple devices.

4. Why would you use a max pooling layer instead with a convolutional layer of the same stride ?

Ans : Pooling layers are used to reduce the dimensions of the feature maps. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network. The pooling layer summarizes the features present in a region of the feature map generated by a convolution layer.conv-layer has parameters to learn (that is your weights which you update each step), whereas the pooling layer does not - it is just applying some given function e.g max-function.

5. When would a local response normalization layer be useful ?

Ans : Layer normalization normalizes each of the inputs in the batch independently across all features. As batch normalization is dependent on batch size, it's not effective for small batch sizes. Layer normalization is independent of the batch size, so it can be applied to batches with smaller sizes as well.Filter Response Normalization (FRN) is a type of normalization that combines normalization and an activation function, which can be used as a replacement for other normalizations and activations.The local normalization tends to uniformize the mean and variance of an image around a local neighborhood. This is especially useful for correct uneven illumination or shading artifacts. Thanks to our fact implementation of the Gaussian filtering, the Local Normalization is running very fast.

6. In comparison to LeNet-5, what are the main innovations in AlexNet? What about GoogLeNet and ResNet's core innovations ?

Ans : AlexNet

it is much larger and deeper

stacks convolutional layer directly on top of each convolutional layer

GooLeNet

introduce a inception modules, which make it possible to have much deeper net than previous network

ResNet

introduce a skip connection.

7. On MNIST, build your own CNN and strive to achieve the best possible accuracy.

Ans : 99.8% accuracy

The best known MNIST classifier found on the internet achieves 99.8% accuracy!! That's amazing. The best Kaggle kernel MNIST classifier achieves 99.75% posted here.

8. Using Inception v3 to classify broad images.

Images of different animals can be downloaded. Load them in Python using the matplotlib.image.mpimg.imread() or scipy.misc.imread() functions, for example. Resize and/or crop them to 299 x 299 pixels, and make sure they only have three channels (RGB) and no transparency. The photos used to train the Inception model were preprocessed to have values ranging from -1.0 to 1.0, so make sure yours do as well.

9. Large-scale image recognition using transfer learning.

Make a training set of at least 100 images for each class. You might, for example, identify your own photos based on their position (beach, mountain, area, etc.) or use an existing dataset, such as the flowers dataset or MIT's places dataset (requires registration, and it is huge).

Create a preprocessing phase that resizes and crops the image to 299 x 299 pixels while also adding some randomness for data augmentation.

Using the previously trained Inception v3 model, freeze all layers up to the bottleneck layer (the last layer before output layer) and replace output layer with appropriate number of outputs for your new classification task (e.g., the flowers dataset has five mutually exclusive classes so the output layer must have five neurons and use softmax activation function).

Separate the data into two sets: a training and a test set. The training set is used to train the model, and the test set is used to evaluate it.