Sharfaraz Hassan

**Biological Analogies**

* Wiesel and Hubel experiment

1. Through their experiment, they discovered that only certain neurons fired only when was perceived in at a particular angle

* Image reconstruction and formation takes up 50-60% of the brain’s computing resources

**Convolution Functions**

* Many layers of convolution starting from individual pixel to more high level detail
* Multi-dimensional convolutions follow a sequence of chaining equations

**CNN Layers – Part 1**

* A convolutional layer specifies a filter to another layer
* The convolution layer has features: special extent, depth width and height, and zero padding

**CNN Layers – Part 2**

* Stride size must have constraints otherwise the convolution layer would not know how to handle it
* If given a lot of parameters, it is possible to share parameters across convolutional layers via filters.
* Requirements of layers:

1. Width, height and depth
2. Depth determines the number of filters
3. Spatial extent of filters
4. The stride
5. The amount of zero padding

* Using interleaving pooling layers reduce the size of the overall layer

**CNN architectures – Part 1**

1. Lenet-5

* 2 convolution layers with 2 fully connected layers

1. Alexnet

* Similar to Lenet-5 but with more layers
* Includes Relu activation – allows infinite error feedback to a network which makes training the network much faster
* Dropout – randomly drops neurons from the network which reduces overfitting issue

**CNN Architectures – Part 2**

1. Googlenet

* Different paths dealt with different types of images based on the complexity of the image
* Introduced many more layers but with far less parameters in each layer

1. Resnet

* Reduced the error rate of calculations by introducing arcs in which layers can skip around to other layers through back propagation

**CNN and computer vision tasks**

* Identifying an image can be solved either via regression or classification method
* Through regression, it is easy to implement if there exists a pre-trained model
* Sliding window creates smaller windows and takes the average score

**CNN Image classification**

* If there exist multiple objects of the same object in an image, it is possible to identify the number of objects using continuous classification with sliding-window
* Because of massive performance issue, instead of using sliding-window, region proposal is used to create an approximate algorithm to generate blob-like structures
* Classifier training includes intruding both positive (the images of the correct object) and negative (the images of the incorrect object that might look like the correct object) samples.

**CNN Segmentation**

* Segmentation – the grouping of an image into several parts
* Two types of segmentation: instanced and semantic
* Sematic: label and understand the image at the pixel level
* Instance: given the category, come up with all instances of category and the mask
* Proposal generation -> feature extraction -> regional classification

**CNN Visualization**

* There are several ways to visualize a convNet: measuring performance, watching the raw weight, t-SNE style visualization, looking at the raw activation, deconvolution based approach, criterion approach
* T-SNE is used when trying to visualize data with many dimensions
* Class score optimization

**Deep dream**

* Feed-forward -> back propagation -> update the image
* Images might consistently include animal faces because the machine rendering the images was fed a training set consisting of many animal faces
* Style transfer – transfer the aesthetic style of one image to another

1. Feed forward the content image into the network and collect the activation at certain layers and do this for both images
2. Calculate the gram matrix
3. Minimize a loss function on content and style which are weight across multiple layers

**Super Resolution**

* Enables lost information of the image to be restored so that the image can be dynamically upscaled (used for enhancing images to HD)
* Typical super-resolution method:

1. Upsample the image
2. Apply interpolation: nearest neighbor, bilinear, bi-cubic

**General Adversarial Networks**

* GAN is used for unsupervised learning where the machine is fed noise rather than clearly defined dataset. It is the machine’s job to make sense of the noise and formulate data based on it.
* Two types of GAN training:

1. Variational

* Pose the network as a graphical model, then optimize the variation lower bound

1. Conditional

* Given the first row/column of the pixel, try to determine what should be next
* Discriminator is used to differentiate between good and bad images