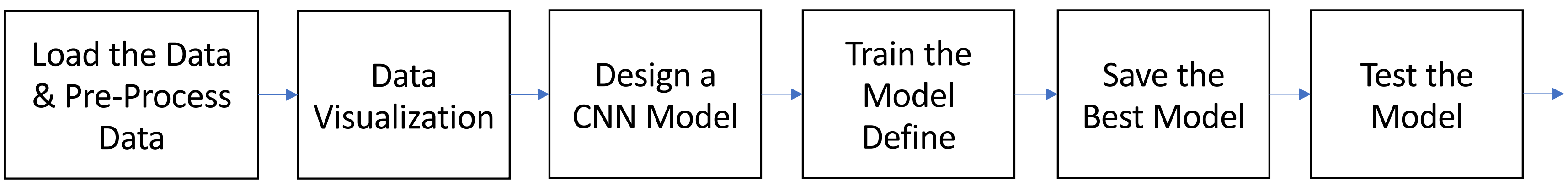
Probabilistic Machine Learning and AI

Outline of the lecture

This lecture introduces you to the fascinating subject of classification and regression with convolutional neural networks.

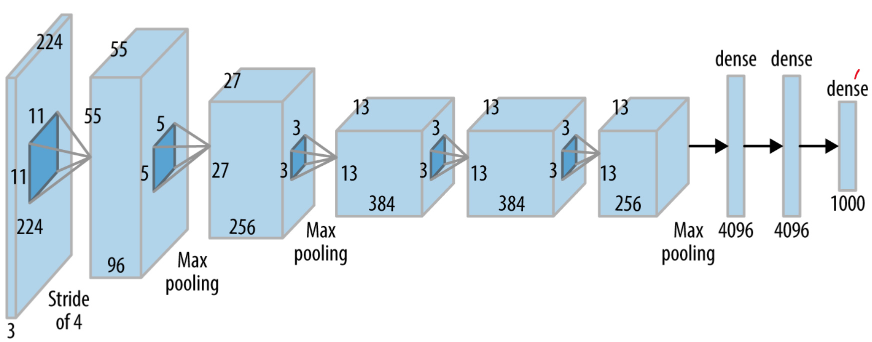
- Different CNN Models and Datasets < http://www.image-net.org/challenges/LSVRC/ >
  - AlexNet
  - VGGNet
  - ResNet
- Transfer Learning
- Visualizing Convolution Neural Network

Process


Load the Data & Pre-Process Data
Data Visualization
Design a CNN Model
Train the Model Define
Save the Best Model
Test the Model

AlexNet

AlexNet is the first deep architecture and was introduced by Geoffrey Hinton and his colleagues. It is a very simple layout but a substantial network architecture that consist of convolution and pooling layers placed one by one which is completely connected at the top. One of the most remarkable features of this architecture is the pace at which it performs various tasks; it has the ability to speed up the training by 10 times through GPU. The network which was given by Geoffrey could be used for classification with 1000 possible categories. It also uses ReLU for nonlinearity functions and data augmentation techniques which comprise of various reflections, patch extractions and image translations. It also applies drop out layers to overcome the overfitting training data problem. Though presently we have more updated architecture but AlexNet is still applied for the deep neural network for different tasks.

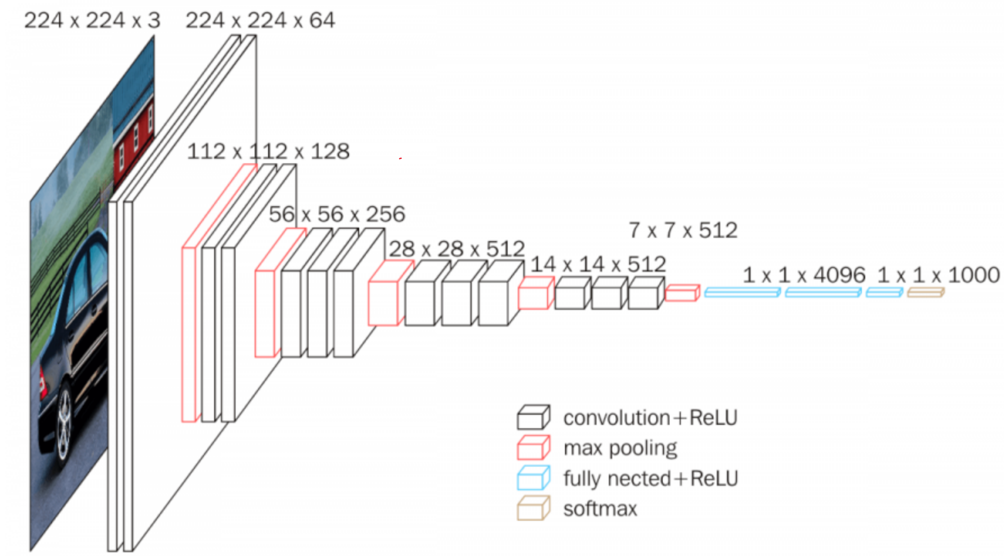


AlexNet Network - Structural Details												
Input			Output			Layer	Stride	Pad	Kernel size		in	out	# of Param
227	227	3	55	55	96	conv1	4	0	11	11	3	96	34944
55	55	96	27	27	96	maxpool1	2	0	3	3	96	96	0
27	27	96	27	27	256	conv2	1	2	5	5	96	256	614656
27	27	256	13	13	256	maxpool2	2	0	3	3	256	256	0
13	13	256	13	13	384	conv3	1	1	3	3	256	384	885120
13	13	384	13	13	384	conv4	1	1	3	3	384	384	1327488
13	13	384	13	13	256	conv5	1	1	3	3	384	256	884992
13	13	256	6	6	256	maxpool5	2	0	3	3	256	256	0
						fc6			1	1	9216	4096	37752832
						fc7			1	1	4096	4096	16781312
						fc8			1	1	4096	1000	4097000
						Total							**62,378,344**

https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

VGG Net

The Researchers at Visual Graphics Group at Oxford gave the VGC Net architecture. The network is basically of pyramidal shape with deep top layers and the bottom layers are placed closer to the image are wide. It is one of the advisable architecture for benchmarking on a particular task. It consists of convolutional layer preceded by pooling layer due to which the layers are narrower. Pre-trained networks of this particular architecture are easily available on the internet due to which it is highly used. Though it is slow to train, if training needs to be followed from a very basic level. It works considerably well on both image classification and localization tasks. The number of filters doubles after each maxpool layer due to which spatial dimension shrinks, but grows in terms of depth.

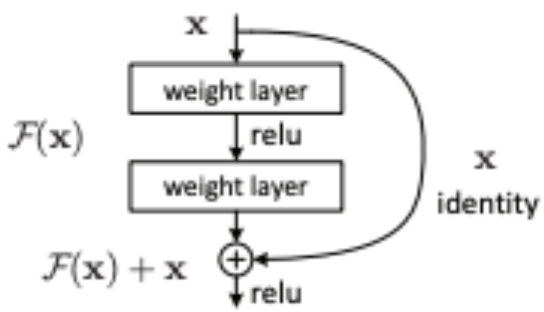


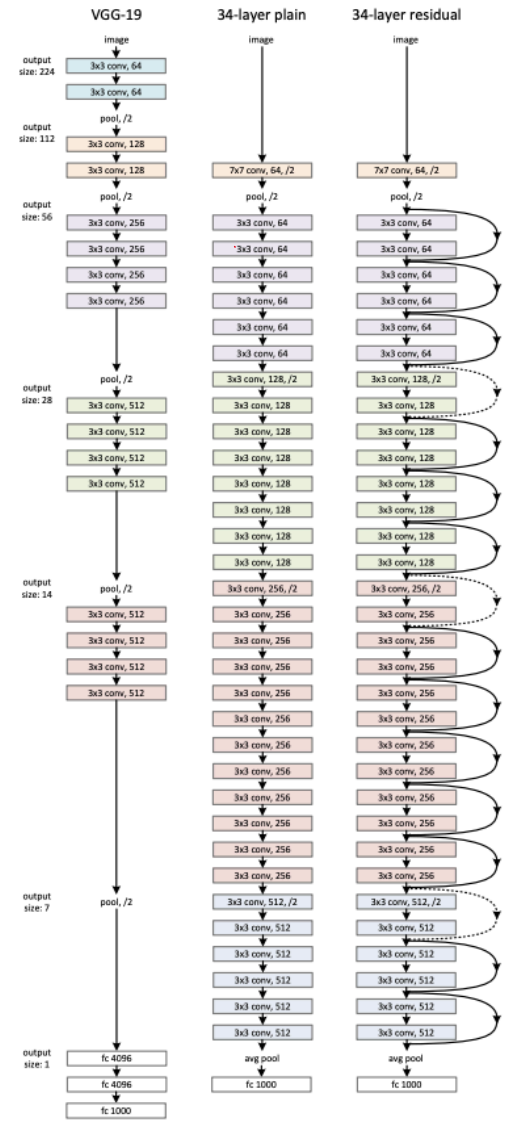
VGG16 - Structural Details

#	Input Image			output			Layer	Stride	Kernel		in	out	Param
1	224	224	3	224	224	64	conv3-64	1	3	3	3	64	1792
2	224	224	64	224	224	64	conv3064	1	3	3	64	64	36928
	224	224	64	112	112	64	maxpool	2	2	2	64	64	0
3	112	112	64	112	112	128	conv3-128	1	3	3	64	128	73856
4	112	112	128	112	112	128	conv3-128	1	3	3	128	128	147584
	112	112	128	56	56	128	maxpool	2	2	2	128	128	65664
5	56	56	128	56	56	256	conv3-256	1	3	3	128	256	295168
6	56	56	256	56	56	256	conv3-256	1	3	3	256	256	590080
7	56	56	256	56	56	256	conv3-256	1	3	3	256	256	590080
	56	56	256	28	28	256	maxpool	2	2	2	256	256	0
8	28	28	256	28	28	512	conv3-512	1	3	3	256	512	1180160
9	28	28	512	28	28	512	conv3-512	1	3	3	512	512	2359808
10	28	28	512	28	28	512	conv3-512	1	3	3	512	512	2359808
	28	28	512	14	14	512	maxpool	2	2	2	512	512	0
11	14	14	512	14	14	512	conv3-512	1	3	3	512	512	2359808
12	14	14	512	14	14	512	conv3-512	1	3	3	512	512	2359808
13	14	14	512	14	14	512	conv3-512	1	3	3	512	512	2359808
	14	14	512	7	7	512	maxpool	2	2	2	512	512	0
14	1	1	25088	1	1	4096	fc		1	1	25088	4096	102764544
15	1	1	4096	1	1	4096	fc		1	1	4096	4096	16781312
16	1	1	4096	1	1	1000	fc		1	1	4096	1000	4097000
							Total						138,423,208

Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

Residual Networks





Practice assignment:
Please read the following link:
http://neuralnetworksanddeeplearning.com/chap5.html

and write 1 page on vanishing gradient

https://arxiv.org/pdf/1512.03385v1.pdf

CNN Benchmarks

https://github.com/jcjohnson/cnn-benchmarks

Transfer Learning

Transfer learning involves taking a **pre-trained neural network** and adapting the neural network to a new, different data set. Depending on both:

- The size of the new data set, and
- The similarity of the new data set to the original data set

Transfer Learning

The approach for using transfer learning will be different. There are four main cases:

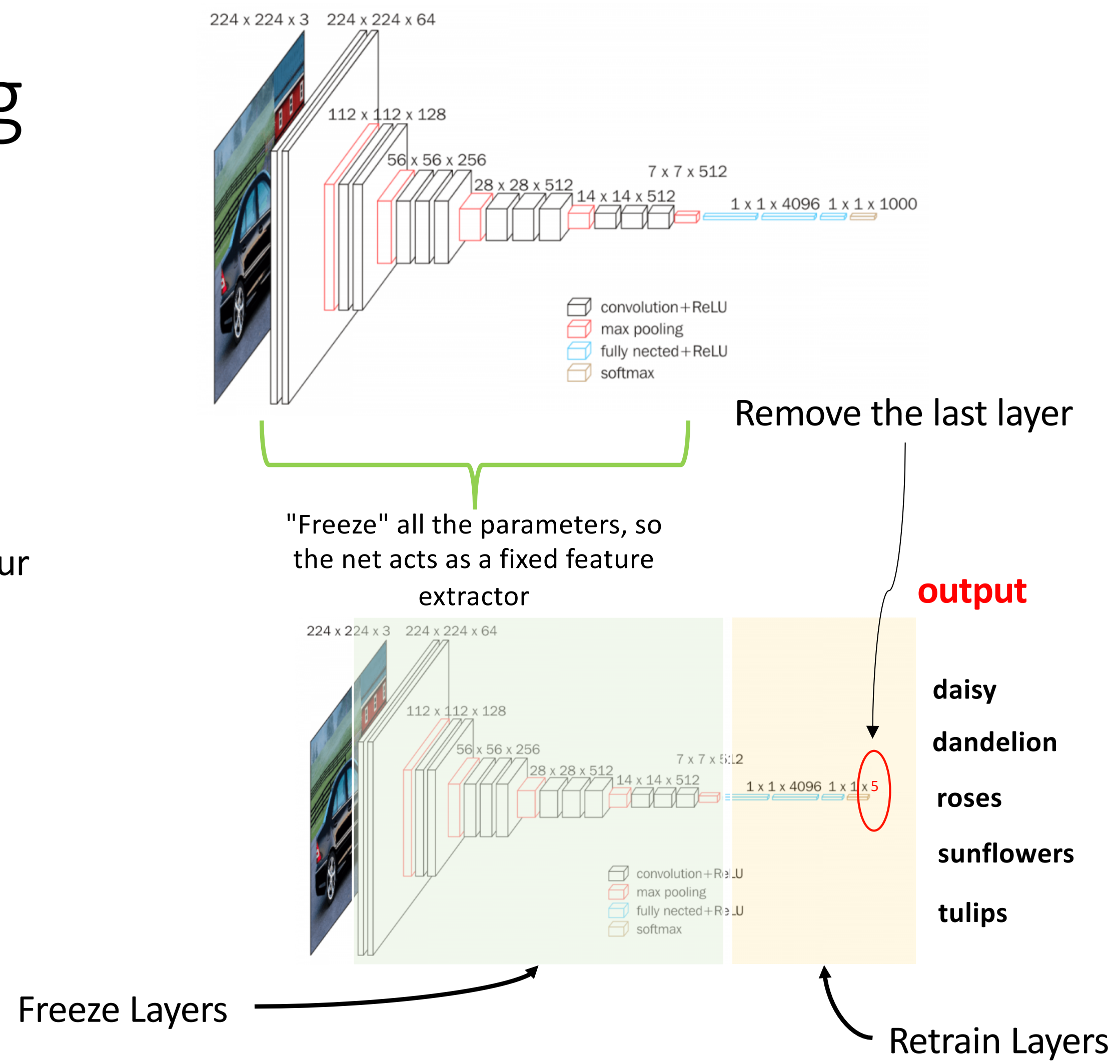
- New data set is small, new data is similar to original training data.
- New data set is small, new data is different from original training data.
- New data set is large, new data is similar to original training data.
- New data set is large, new data is different from original training data.

More to read: https://cs231n.github.io/transfer-learning/#tf

Ex. Transfer Learning

Sample Transfer Learning Steps:

- Load in a pre-trained VGG16 model
- "Freeze" all the parameters, so the net acts as a fixed feature extractor
- Remove the last layer
- Replace the last layer with a linear classifier of our own
- Train FC layers using the data set



Assignment: Visualizing and Understanding Convolutional Networks

Please read the following Paper by Zeiler and Fergus (NYU)

https://cs.nyu.edu/~fergus/papers/zeilerECCV2014.pdf

Here is Matt Zeiler's presentation

https://www.youtube.com/watch?v=ghEmQSxT6tw

Code: https://github.com/FHainzl/Visualizing_Understanding_CNN_Implementation

Assignment 3: Implement this paper in Pytorch