What is ImageNet competition?

ImageNet Competition: The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual computer vision competition held to evaluate the performance of computer vision models on a large scale image classification task.

Significance: The ImageNet competition has been instrumental in advancing the state-of-the-art in computer vision, with deep learning models winning the competition since 2012.

What is Top-K Accuracy?
- Top-k accuracy refers to the accuracy of a model's top k predictions compared to the actual target classes.

- For example, if we have a multi-class classification problem with 5 classes and a model makes the following predictions:

- Prediction: [0.1, 0.3, 0.5, 0.8, 0.2] | True class: 3 If we set k=3, the model's top 3 predictions are classes 2, 3, and 4.

- Since class 3 is the true class, the top-k accuracy would be 1.0, indicating a correct prediction.

Alexnet Architecture
- AlexNet is the first CNN to win ImageNet Competition in 2012.
- It was primarily designed by Alex Krizhevsky.
- They achieved a top 5 accuracy of 84.7%, while runner-up was at 74!
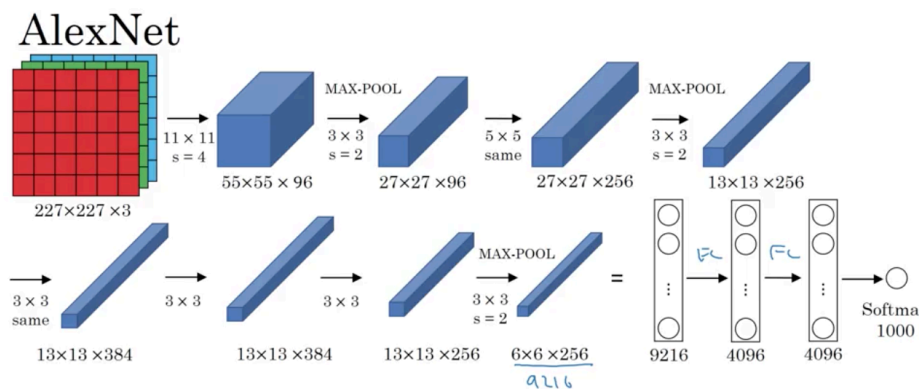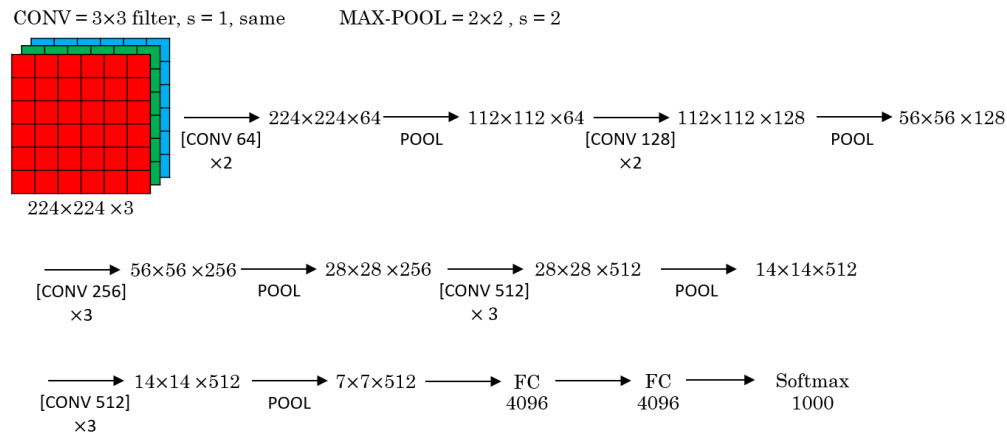

What was so new about AlexNet?
1. Graphics processing units (GPUs)
2. Rectified linear unit (ReLU) activation
3. Regularization   (Dropout and Data Augmentation)
4. Pooling layers
5. Increasing the Depth

VGGNet Architecture

VGG19 won second place in 2014 It was designed by the Visual Geometry Group (VGG) at the University of Oxford, UK.

VGG16 has top 5 acc of 91.9 and top 1 acc of 74.4 on ImageNet dataset.

CONV = 3×3 filter, s = 1, same   MAX-POOL = 2×2 , s = 2

224×224 ×3 → [CONV 64] ×2 → 224×224×64 → POOL → 112×112 ×64 → [CONV 128] ×2 → 112×112 ×128 → POOL → 56×56 ×128

→ [CONV 256] ×3 → 56×56 ×256 → POOL → 28×28 ×256 → [CONV 512] × 3 → 28×28 ×512 → POOL → 14×14×512

→ [CONV 512] ×3 → 14×14 ×512 → POOL → 7×7×512 → FC 4096 → FC 4096 → Softmax 1000

Improvements of VGG-16 over AlexNet
1. Deeper Network
2. Smaller Filters
3. Pooling Layers: VGG-16 employs more pooling layers
4. Uniform Filter Size - VGG-16 uses the same filter size (3x3) throughout the network

What is Transfer Learning?

Definition: Transfer learning is a machine learning technique in which a pre-trained model is used as a starting point to solve a similar but different problem.

Advantages:
1. Transfer learning is often used in computer vision because it allows one to leverage the knowledge learned from large amounts of data in one task to another, similar task.
2. This can result in improved performance and faster training times, as well as reduced data and computational requirements compared to training a CNN from scratch.
3. Additionally, transfer learning can be useful when there is limited annotated data available for a particular problem.

General Transfer Learning Architecture
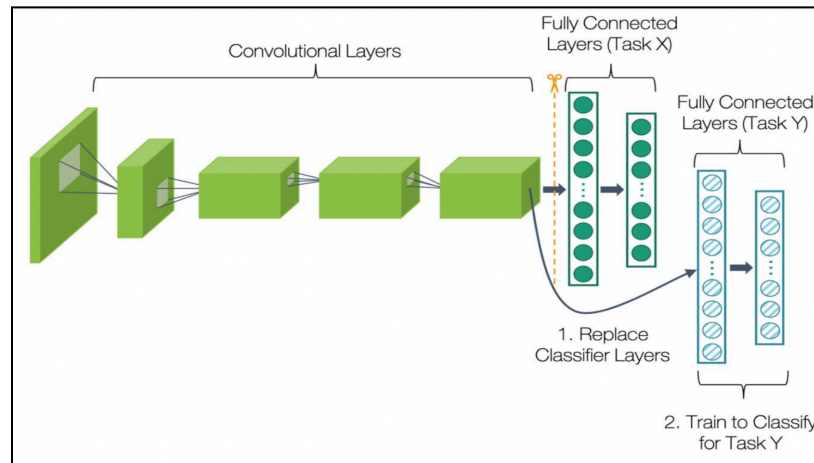
What are the shortcomings of VGG-16

1. **Complexity:** VGG-16 has a large number of parameters, making it computationally expensive to train and difficult to deploy on low-power devices.

2. **Overfitting:** Due to its large number of parameters, VGG-16 is prone to overfitting if not regularized properly, which can result in poor performance on unseen data.

3. **Large Memory Requirements:** VGG-16 requires a large amount of memory to store its parameters, making it challenging to run on smaller devices with limited memory.

4. **Difficulty in Training:** VGG-16's deep architecture and large number of parameters can make training challenging and time-consuming, especially when using limited amounts of annotated data.

5. **Lack of Spatial Information:** VGG-16 uses max pooling to reduce the spatial resolution of the feature maps, which can result in the loss of spatial information and reduce the model's ability to capture fine-grained details in images.