CV / VLM

Unit 3: Transfer Learning for CV

3.2.1

Pre-trained Models & Transfer Learning

Introduction to popular pre-trained models

TIL-AI
TODAY I LEARNED AI

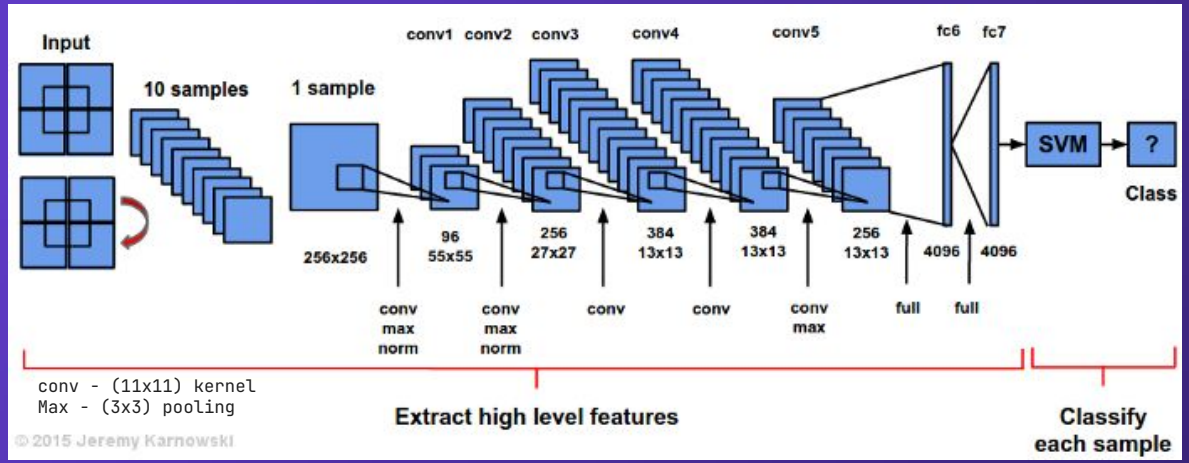
AlexNet

AlexNet competed in the ImageNet Challenge in Sep 2012

- Achieved top-5 error of 15.3%

Its author, Alex Krizhevsky, proposed that the **depth of the model was essential for its high performance.**

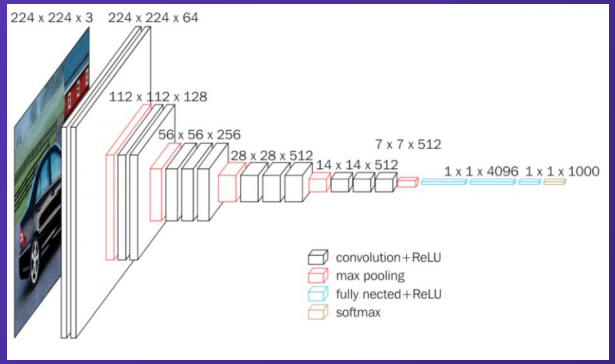
- This was made computationally feasible by being among the first to use GPUs for training



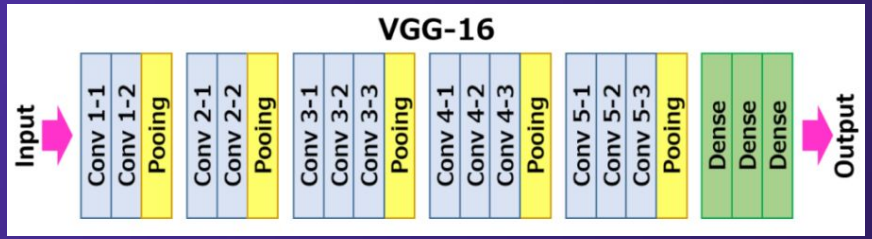
Schematics of Alexnet + SVM, Adopted from AlexNet Visualization | Jeremy Karnowski

Visual Geometric Group (VGG)

- **VGG is characterized by its depth.** A VGG16 has 16 weight layers (hence the name).
- It is proposed in 2014 for ImageNet.
- The architecture consists of a consistent arrangement of convolution layers and max-pooling layers.
- The input tensor size for VGG16 is 224x244 with 3 RGB channels.
- Focused on usage of small 3x3 convolution filters and 2x2 max_pooling layers.



(top) VGG-16 | CNN model for Classification and Detection - All about Machine Learning (techcraft.org)



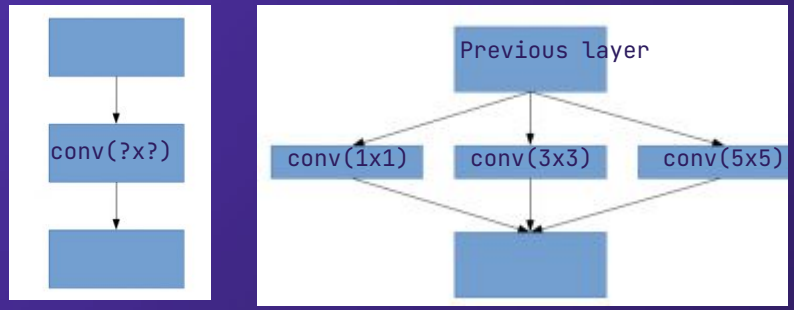
Very Deep Convolutional Networks for Large-Scale Image Recognition - Proposed by K. Simonyan and A. Zisserman

InceptionNet (GoogLeNet)

InceptionNet was developed to improve upon the performance of previous CNNs in the ImageNet Challenge in 2014. It features:

Inception modules

- Definitions: A combination of 1x1, 3x3, and 5x5 convolutions on the previous layer.
- Details: its researchers suggest that, instead of just a deeper network, we would also benefit from sparsely connected network architectures (instead of fully connected networks) within convolutional layers.



(top left) Densely connected architecture - Convolution Modules
(top right) Sparsely connected architecture - Inception Modules

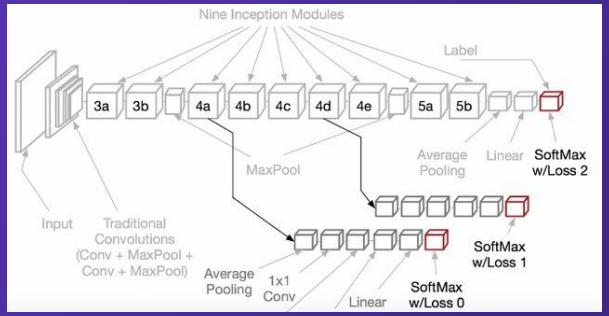
Schematics of Alexnet + SVM, Adopted from AlexNet Visualization | Jeremy Karnowski

InceptionNet (GoogLeNet)

Auxiliary Classifier for Training

- The use of intermediate classifiers (softmax).

  It utilized a total of three loss layers instead of a single layer.



(top) Visual schematics of InceptionNet, with focus on SoftMax layers

These Features are designed to address:

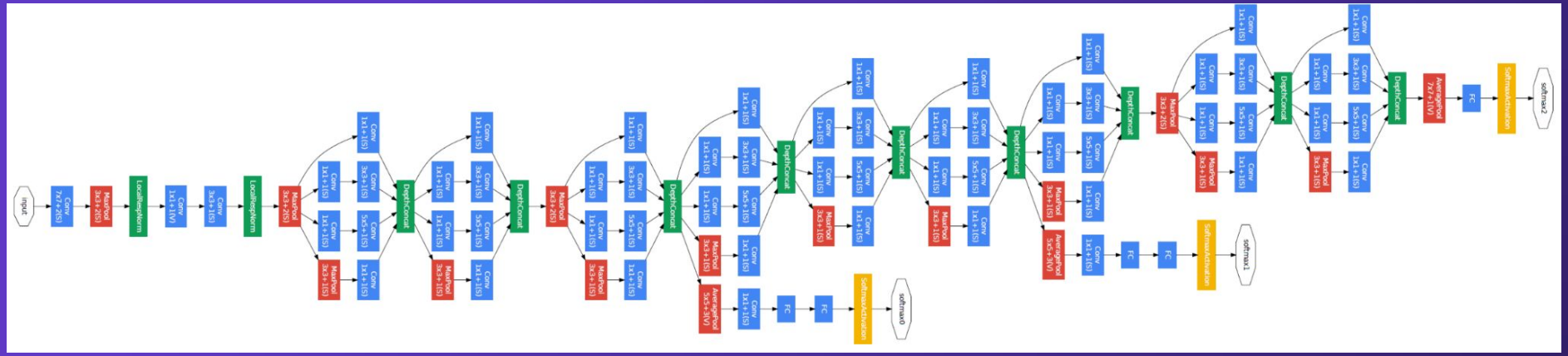
Vanishing gradients problem during training (back-propagation algorithm)

- Intuitively, with deeper layers, the gradients carry less and less information with each subsequent depth, thereby hindering parameters to be learnt.

Inception Network | Implementation Of GoogleNet In Keras (analyticsvidhya.com)
Understanding GoogLeNet Model - CNN Architecture - GeeksforGeeks
deep learning - Google Inception model:why there is multiple softmax? - Cross Validated (stackexchange.com)

InceptionNet (GoogLeNet)

The architecture Overview



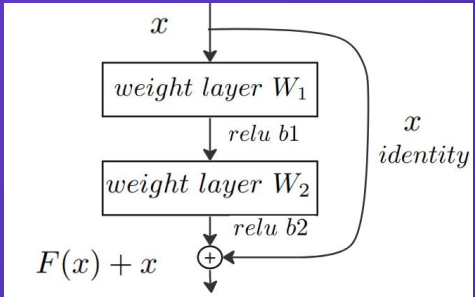
(top) InceptionNet's architecture, with inception modules (Conv), and Auxiliary Classifier

Schematics of Alexnet + SVM, Adopted from AlexNet Visualization | Jeremy Karnowski

Residual Neural Networks (ResNet)

- Proposed in 2015 for ImageNet Challenge.
- Introduced to address ‘vanishing gradient problem’
- It circumvents the vanishing gradient problem by providing an alternative path for the gradient:

To be explored in more depth in Unit 6!

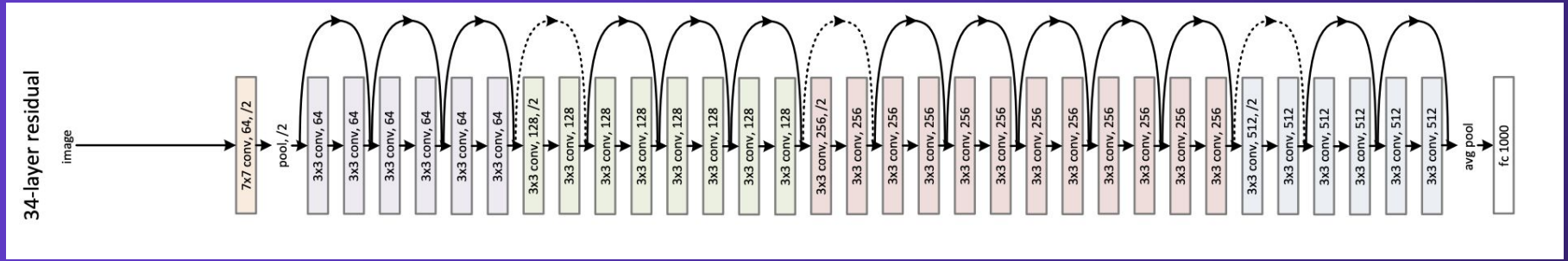


(left) Directy summing the (input - X) previous layers to the results of the next layers (output F(X) + X).

ResNet 34 Classification Model: What is, How to Use (roboflow.com)
Deep Residual Learning for Image Recognition

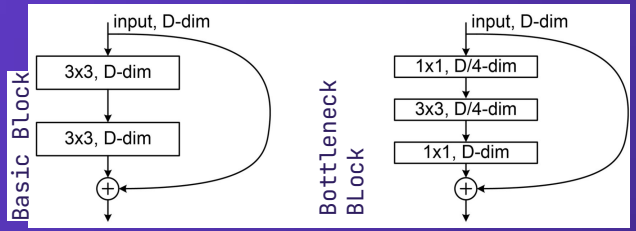
Residual Neural Networks (ResNet)

- Overview



(top) Resnet34 with its residuals layers (the Bolded Curved Line) in between the layers
There are two variants of Convolutional Resnet block

- **Solid Lines - Basic Block**
- **Dotted Lines - Bottleneck Block** (has an extra 1x1 convolutional layers for dimension reduction and restoration)



ResNet 34 Classification Model: What is, How to Use (roboflow.com)
Deep Residual Learning for Image Recognition