18AIC301J: DEEP LEARNING TECHNIQUES

B. Tech in ARTIFICIAL INTELLIGENCE, 5th semester

Faculty: **Dr. Athira Nambiar**

Section: A, slot:D

Venue: TP 804

Academic Year: 2022-22

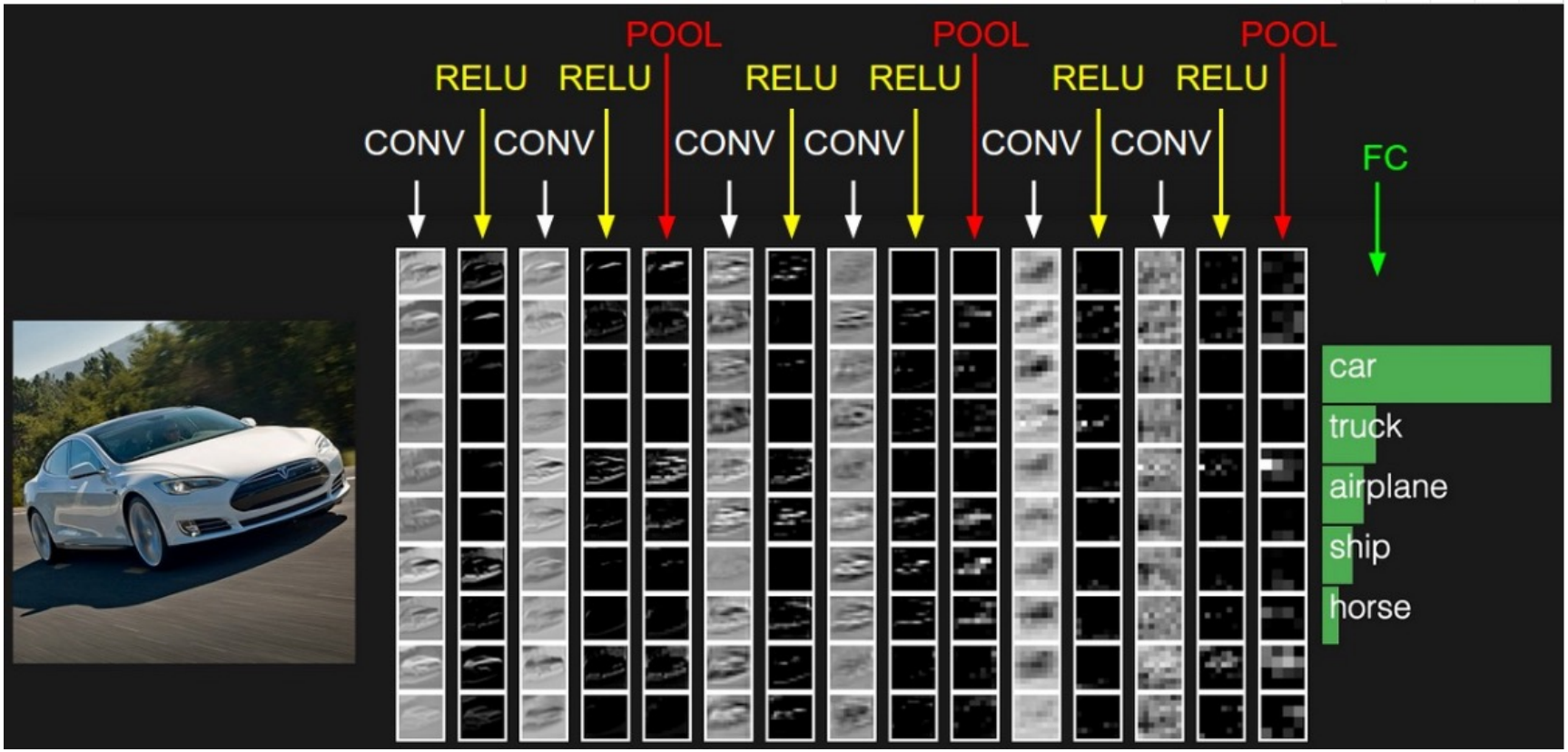
UNIT-3

One hot representation of words, Distributed representation of words
SVD for learning word Representations, Continuous bag of words model, Skip-gram model, Hierarchical Softmax
Implement skip gram model to predict words within a certain range before and after the current word
Introduction to Convolution Neural Networks, Kernel filters
The convolution operation with Filters, padding and stride, Multiple Filters, Max pooling and non-linearities
Implement LeNet for image classification
Classic CNNs architecture- The ImageNet challenge, Understanding Alex Net architecture
ZFNet, The intuition behind GoogleNet, Average pooling, Residual CNN-ResNet architecture
Implement ResNet for detecting Objects.

UNIT-3

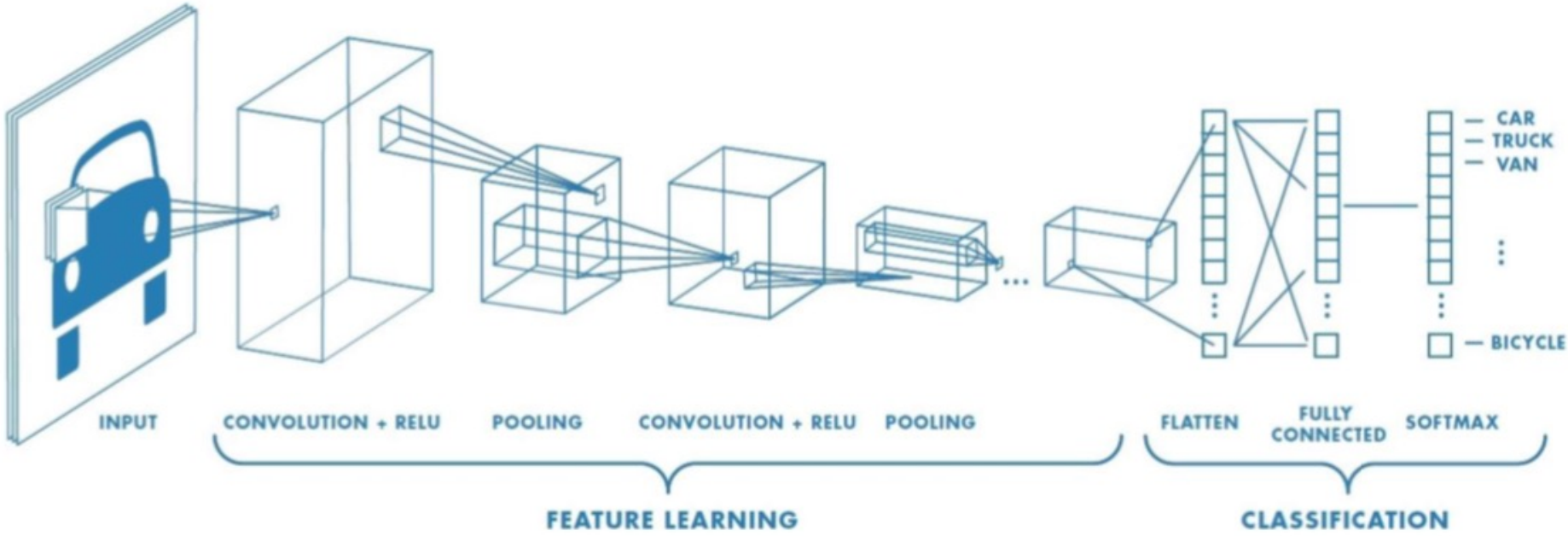
One hot representation of words, Distributed representation of words
SVD for learning word Representations, Continuous bag of words model, Skip-gram model, Hierarchical Softmax
Implement skip gram model to predict words within a certain range before and after the current word
Introduction to Convolution Neural Networks, Kernel filters
The convolution operation with Filters, padding and stride, Multiple Filters, Max pooling and non-linearities
Implement LeNet for image classification
Classic CNNs architecture- The ImageNet challenge, Understanding Alex Net architecture
ZFNet, The intuition behind GoogleNet, Average pooling, Residual CNN-ResNet architecture
Implement ResNet for detecting Objects.

A simple CNN structure



CONV: Convolutional kernel layer
RELU: Activation function
POOL: Dimension reduction layer
FC: Fully connection layer

CNN for classification

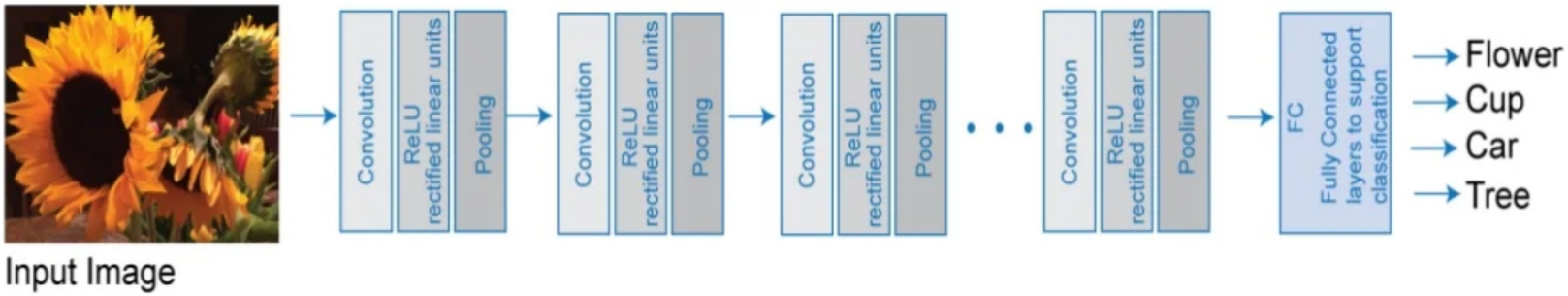


1. **Feature Learning:** Conv + ReLU + pooling
2. **Classification:**
   - Fully connected layer uses these features for classifying input image
   - Express output as **probability** of image belonging to a particular class

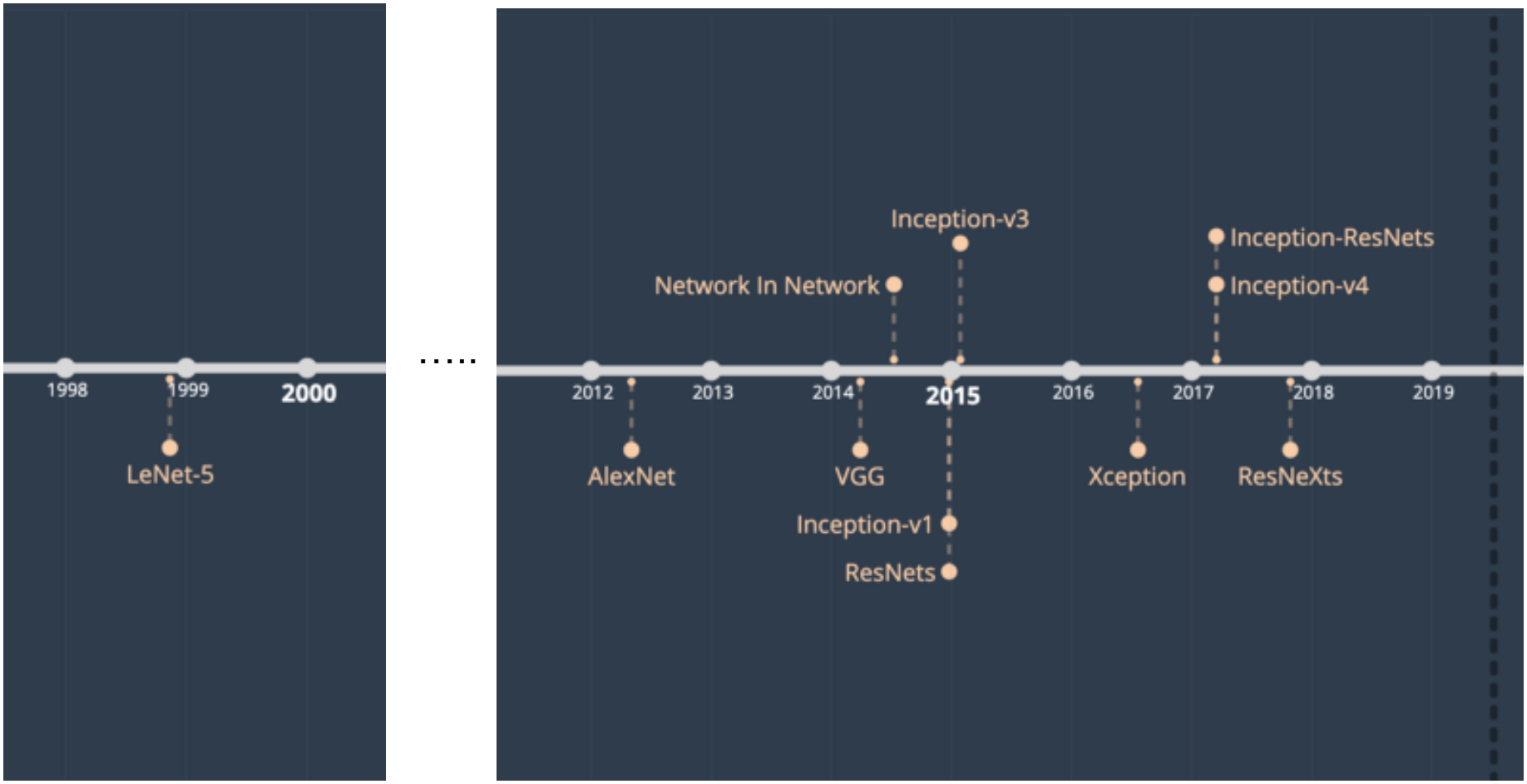
$$\text{softmax}(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

Classic CNNs architecture

- **Convolutional Neural Network (CNN)**is a multi-layer neural network
- **Convolutional Neural Network** is comprised of one or more **convolutional layers** (often with a **pooling layers**) and then followed by one or more **fully connected layers**.



Classic CNNs architecture

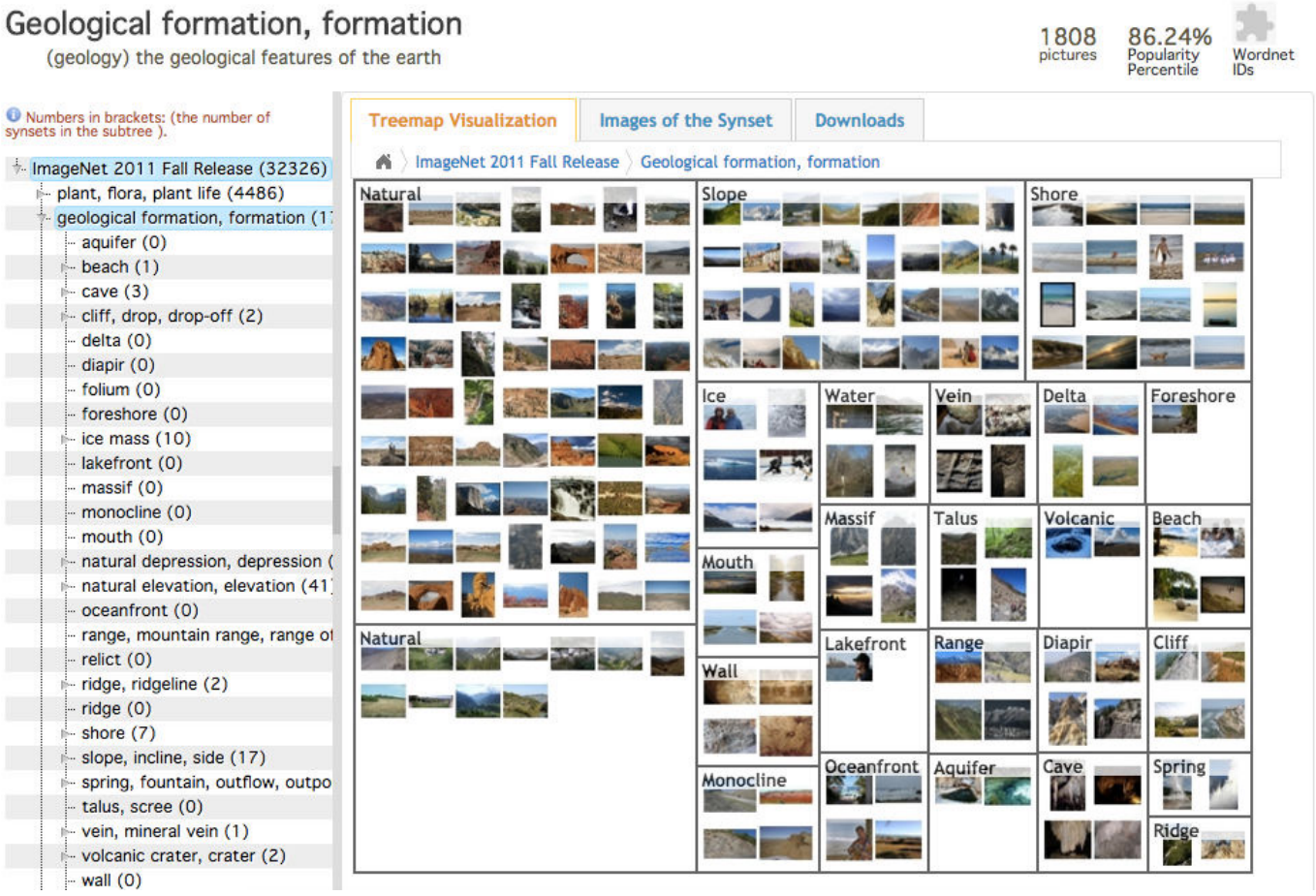


The ImageNet challenge

- The ImageNet dataset is long-standing landmark in computer vision.
- The impact ImageNet has had on computer vision research is driven by the dataset's size and semantic diversity.

The ImageNet challenge

ImageNet consists of 14,197,122 images organized into 21,841 subcategories. These subcategories can be considered as sub-trees of 27 high-level categories. Thus, ImageNet is a well-organized hierarchy that makes it useful for supervised machine learning tasks.



The ImageNet challenge

ILSVRC

- **ImageNet Large Scale Visual Recognition Challenge** is image classification challenge to create model that can correctly classify an input image into 1,000 separate object categories.
- Models are trained on 1.2 million training images with another 50,000 images for validation and 150,000 images for testing

The **ImageNet Large Scale Visual Recognition Challenge (ILSVRC)** is the most commonly used subset of the ImageNet dataset. Within this subset:

- ImageNet contains 1,281,167 training images
- ImageNet contains 50,000 validation images
- ImageNet contains 100,000 test images
- ImageNet contains 1000 object classes

The ImageNet challenge

https://www.image-net.org

IMAGENET

14,197,122 images, 21841 synsets indexed

Home Download Challenges About

Not logged in. Login | Signup

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Competition

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale. One high level motivation is to allow researchers to compare progress in detection across a wider variety of objects -- taking advantage of the quite expensive labeling effort. Another motivation is to measure the progress of computer vision for large scale image indexing for retrieval and annotation.

For details about each challenge please refer to the corresponding page.

- ILSVRC 2017
- ILSVRC 2016
- ILSVRC 2015
- ILSVRC 2014
- ILSVRC 2013
- ILSVRC 2012
- ILSVRC 2011
- ILSVRC 2010

Workshop

Every year of the challenge there is a corresponding workshop at one of the premier computer vision conferences. The purpose of the workshop is to present the methods and results of the challenge. Challenge participants with the most successful and innovative entries are invited to present. Please visit the corresponding challenge page for workshop schedule and information.

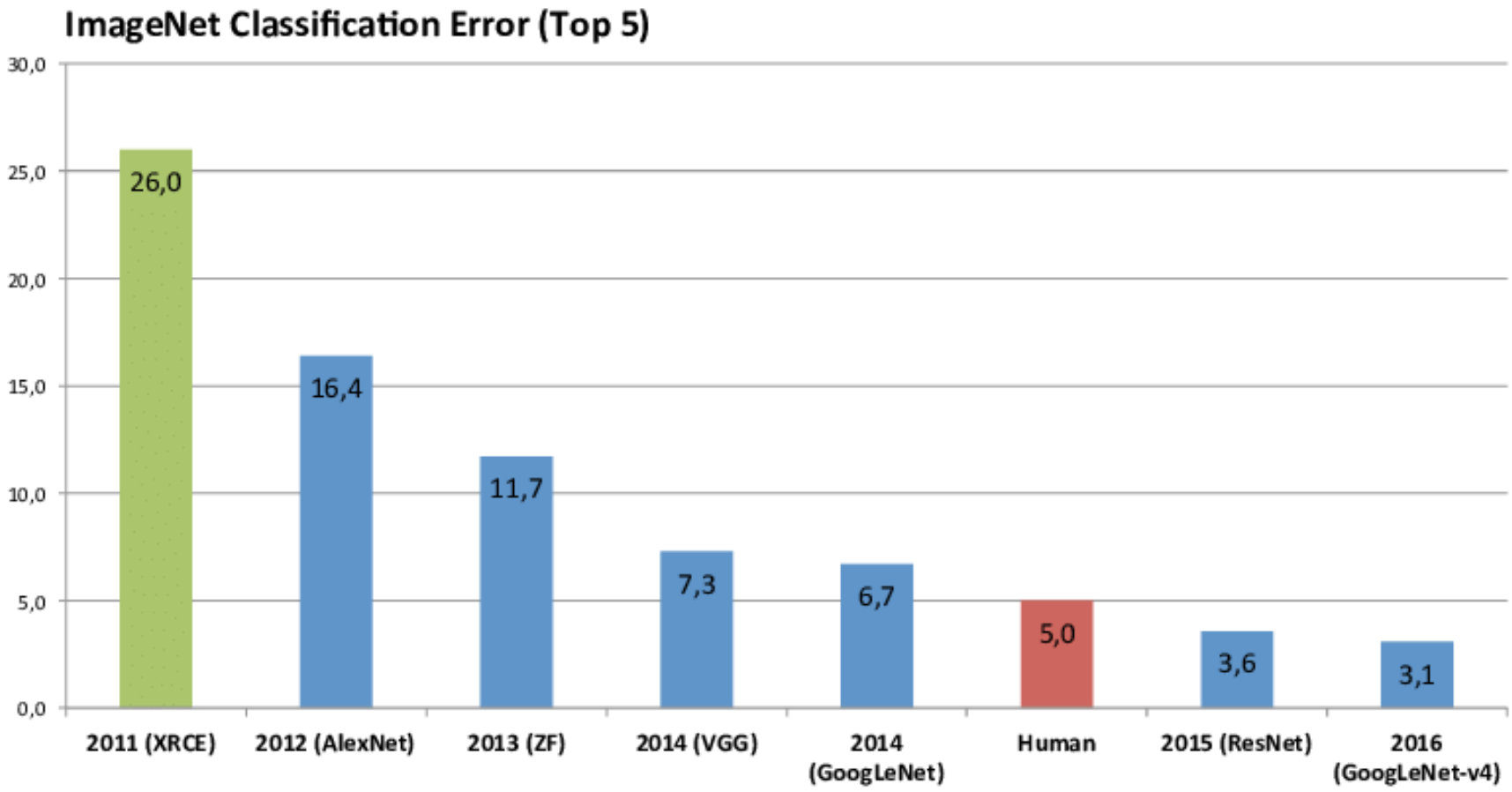
Download

The most popular challenge is the ILSVRC 2012-2017 image classification and localization task. It is available on Kaggle. For all other data please log in or request access.

Evaluation Server

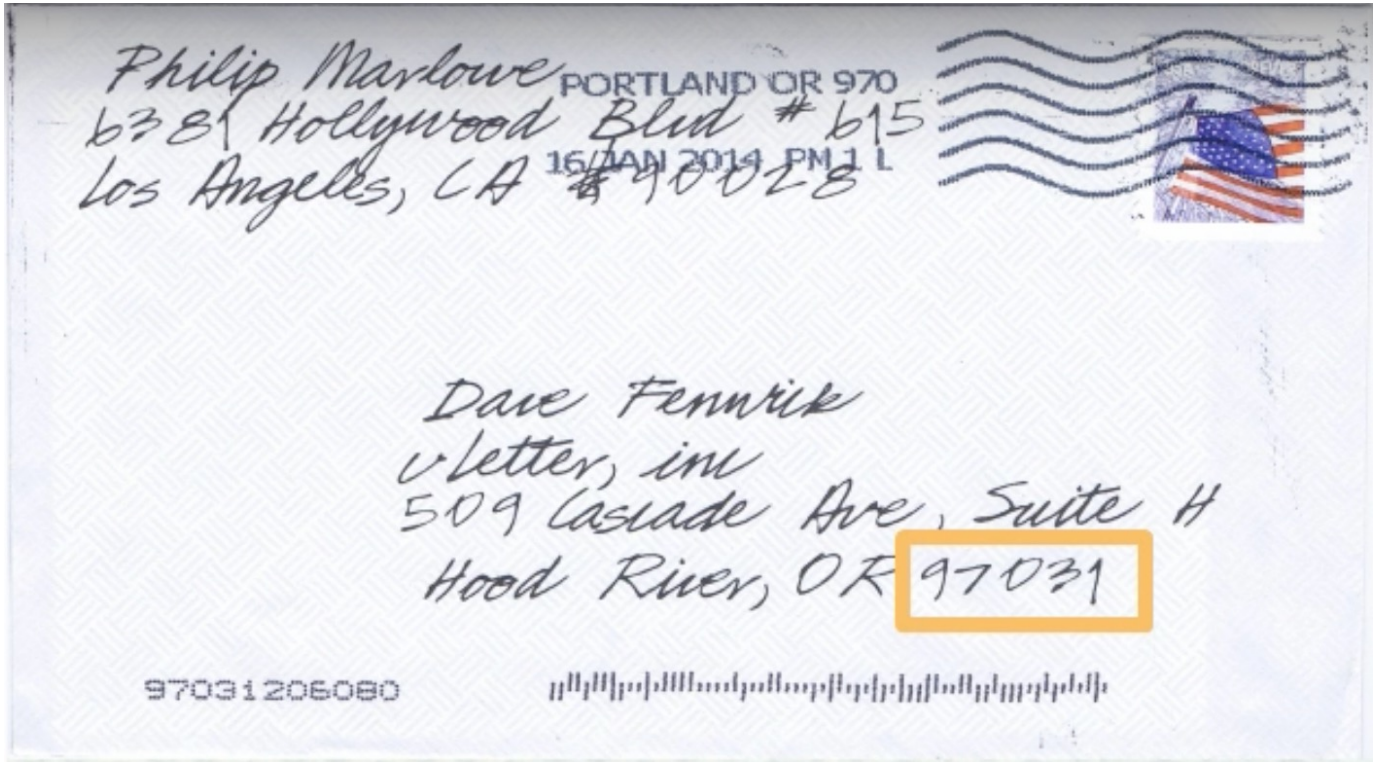
The evaluation server can be used to evaluate image classification results on the test set of ILSVRC 2012-2017. Please see here for our submission policy. Importantly, you should not make more than 2 submissions per week.

The ImageNet challenge



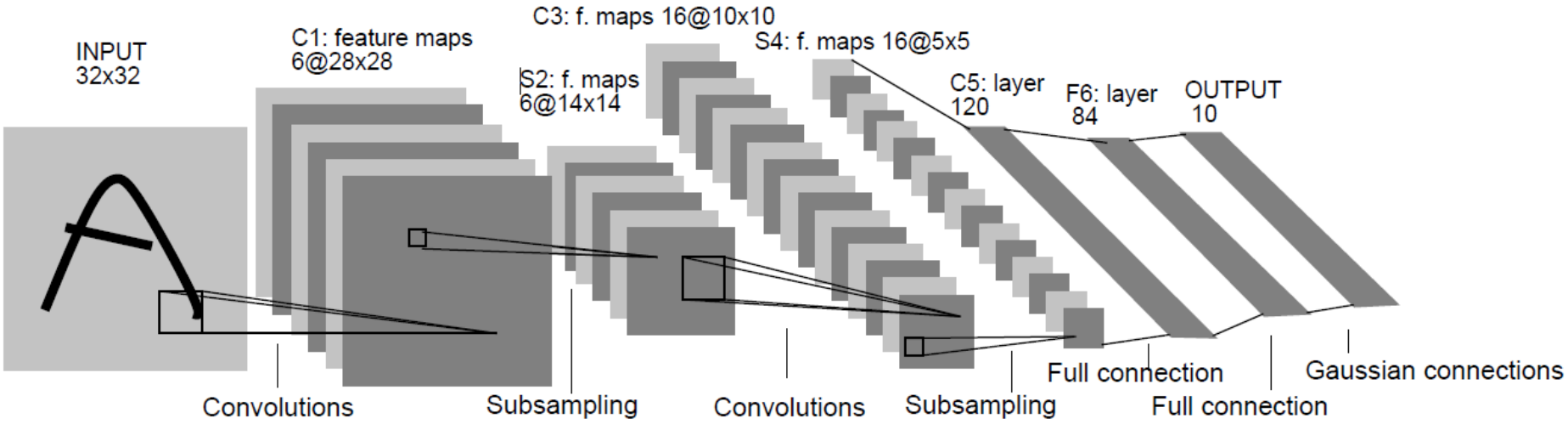
LeNet

Yann LeCun created **LeNet – 5** in 1998.
Handwritten and machine printed character recognition are the main applications of this network architecture.



Handwritten digit recognition

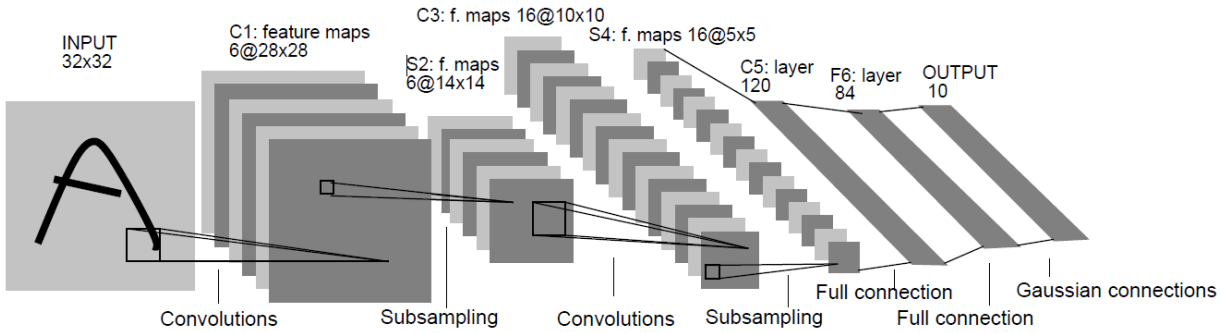
LeNet



MNIST

- Centered and scaled
- 50,000 training data
- 10,000 test data
- 28 x 28 images
- 10 classes

LeNet



Layer	Size
Input	32×32
Convolution	28×28
Pooling	14×14
Convolution	10×10
Pooling	5×5
Convolution	1×1
Fully Connected	84
Fully Connected (Output)	10

LeNet

The Architecture of the Model

Layer	# filters / neurons	Filter size	Stride	Size of feature map	Activation function
Input					
Conv 1					
Avg. pooling 1					
Conv 2					
Avg. pooling 2					
Conv 3					
Fully Connected 1					
Fully Connected 2					

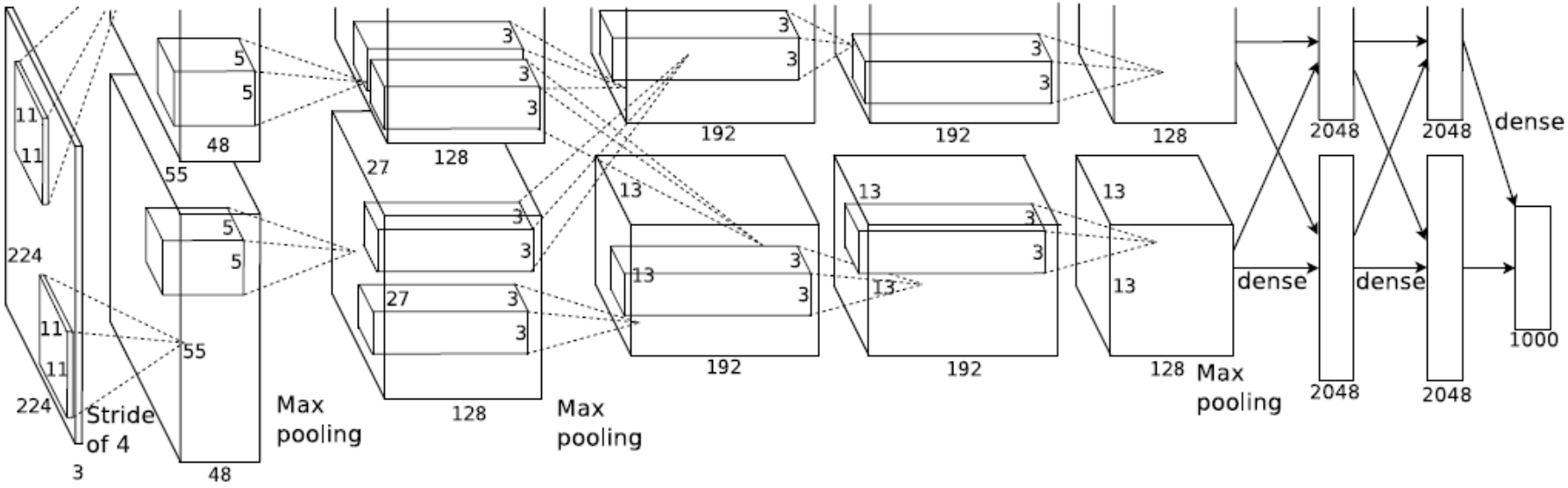
?

AlexNet architecture

- AlexNet was developed by Alex Krizhevsky et al. in 2012 to compete in the **ImageNet competition**.
- The general architecture is quite similar to LeNet-5, although this model is considerably larger.
- The success of this model (which took **first place in the 2012 ImageNet competition)** convinced a lot of the computer vision community to take a serious look at deep learning for computer vision tasks.
- In 2012, AelxNet won the ILSVRC challenge with a top 5 error rate of 16%, which was almost 10% less than the runner-up model (26%).

AlexNet architecture

- It consists of five convolutional layers and three fully connected dense layers, a total of eight layers. The activation function is ReLU for all the layers except the last one which is softmax activation.

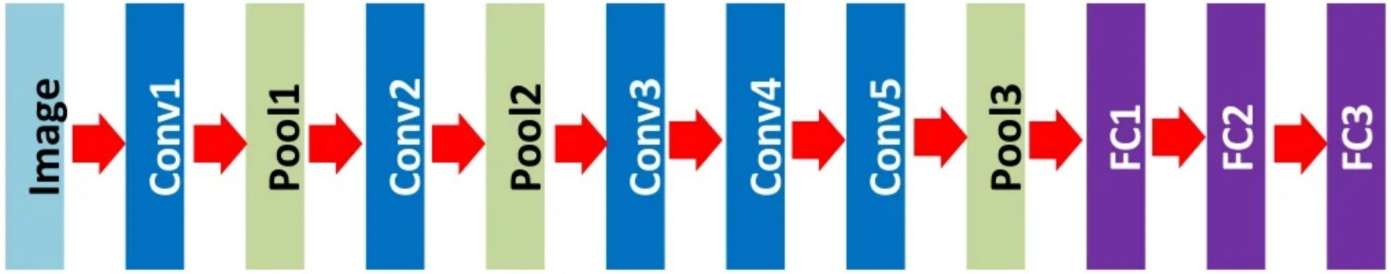


Parameters: 60 million

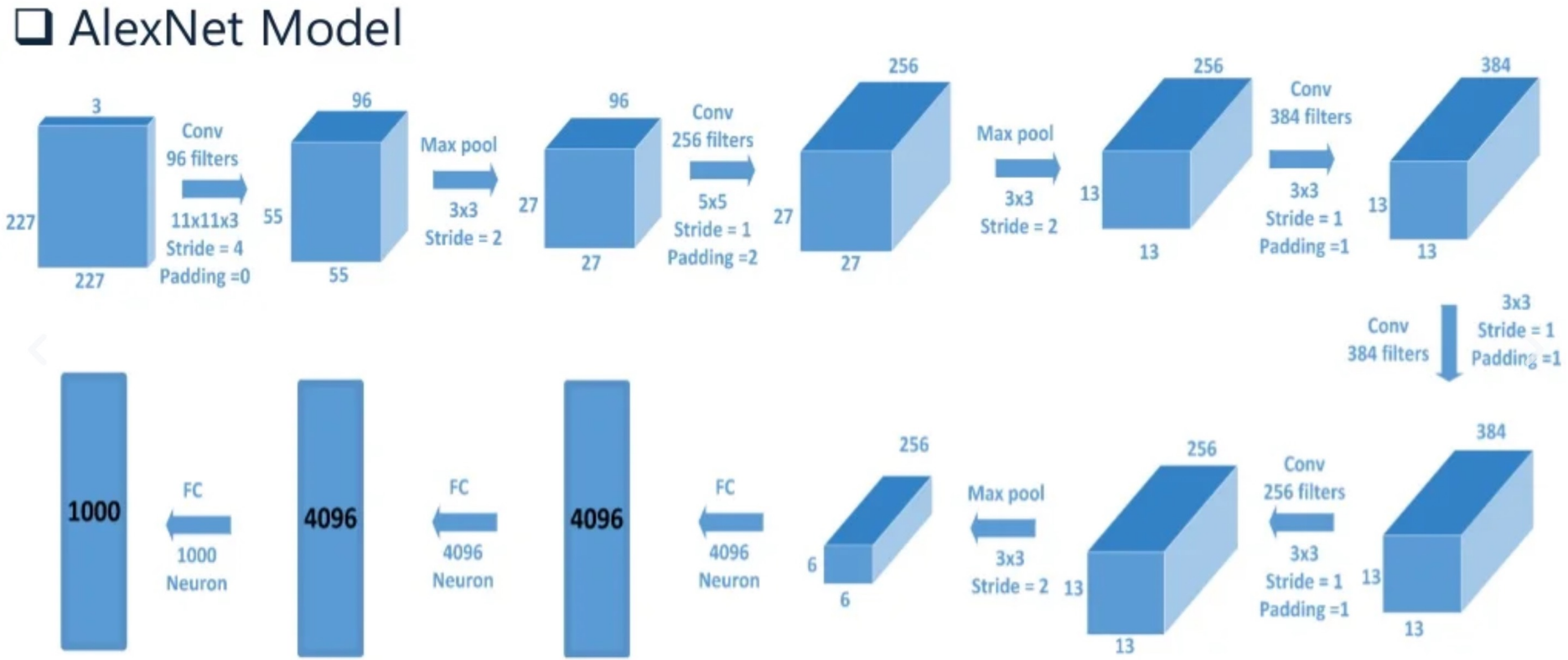
Paper: ImageNet Classification with Deep Convolutional Neural Networks

AlexNet architecture

- **AlexNet** achieve on ILSVRC 2012 competition **15.3%** Top-5 error rate compare to 26.2% achieved by the second best entry.
- **AlexNet** using batch stochastic gradient descent on training, with specific values for momentum and weight decay.
- **AlexNet** implement dropout layers in order to combat the problem of overfitting to the training data.
- **AlexNet** has 8 layers without count pooling layers.
- **AlexNet** use ReLU for the nonlinearity functions
- **AlexNet** trained on two GTX 580 GPUs for **five to six days**

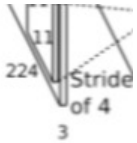


AlexNet architecture



AlexNet architecture

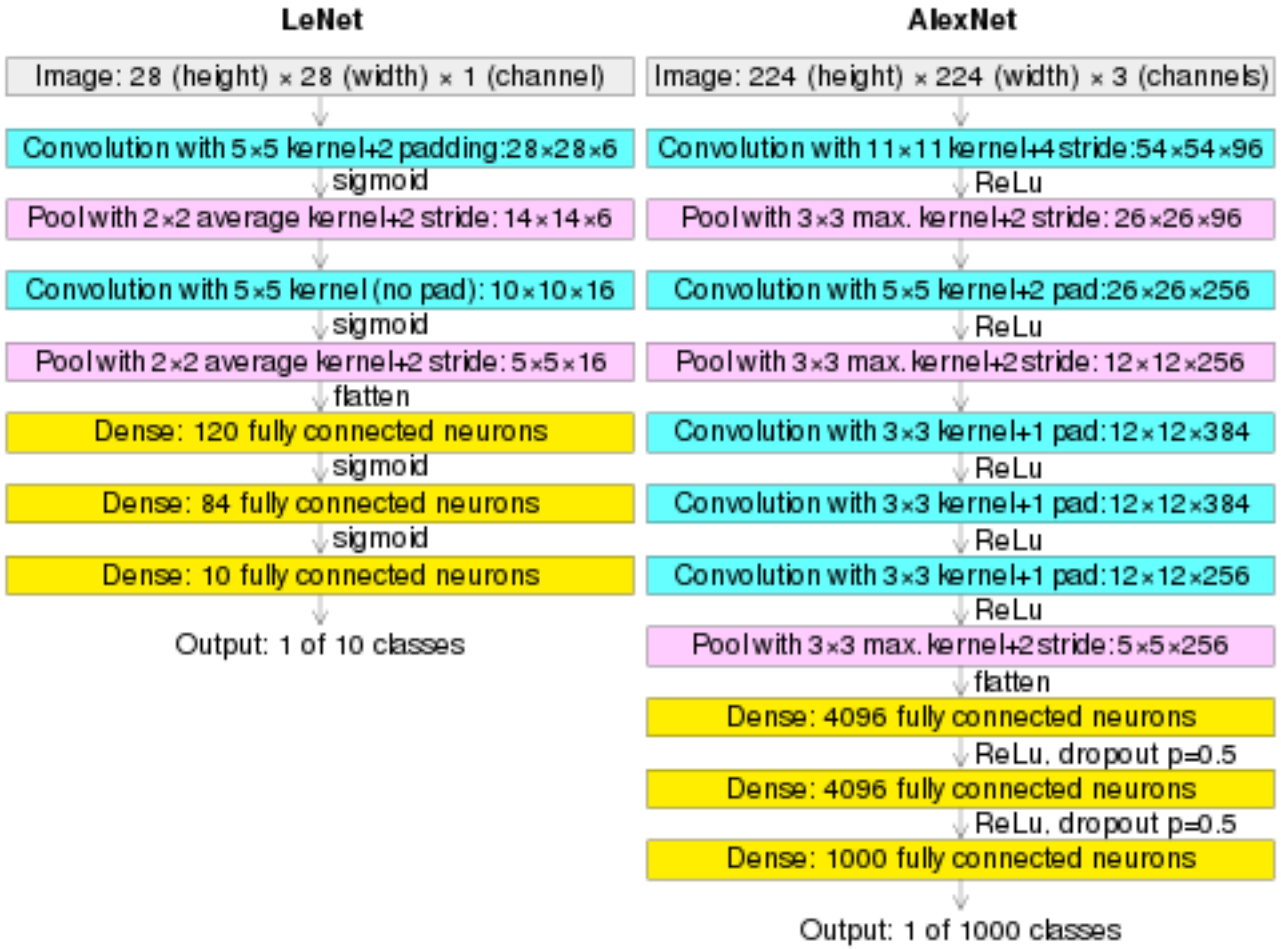
Full (simplified) AlexNet architecture:



[227x227x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)

AlexNet architecture

Comparison of the LeNet and AlexNet convolution, pooling, and dense layers (AlexNet image size should be 227×227×3, instead of 224×224×3, so the math will come out right. The original paper said different numbers, but Andrej Karpathy, the head of computer vision at Tesla, said it should be 227×227×3 (he said Alex didn't describe why he put 224×224×3). The next convolution should be 11×11 with stride 4: 55×55×96 (instead of 54×54×96). It would be calculated, for example, as: [(input width 227 - kernel width 11) / stride 4] + 1 = [(227 - 11) / 4] + 1 = 55. Since the kernel output is the same length as width, its area is 55×55.)



Learning Resources

- Charu C. Aggarwal, Neural Networks and Deep Learning, Springer, 2018.
- Eugene Charniak, Introduction to Deep Learning, MIT Press, 2018.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
- Michael Nielsen, Neural Networks and Deep Learning, Determination Press, 2015.
- Deng & Yu, Deep Learning: Methods and Applications, Now Publishers, 2013.

http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture5.pdf

- https://www.youtube.com/watch?v=uapdILWYTzE&t=2172s
- https://www.youtube.com/watch?v=bNb2fEVKeEo&t=2949s

Thank you