CONVOLUTIONAL NEURAL NETS

HLML – 06/02/2019

THE PAPER THAT STARTED IT ALL

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

THE PROBLEM | IMAGENET LARGE-SCALE VISUAL RECOGNITION CHALLENGE

>1.2 million high-res images of different sizes to train with

1000 categories

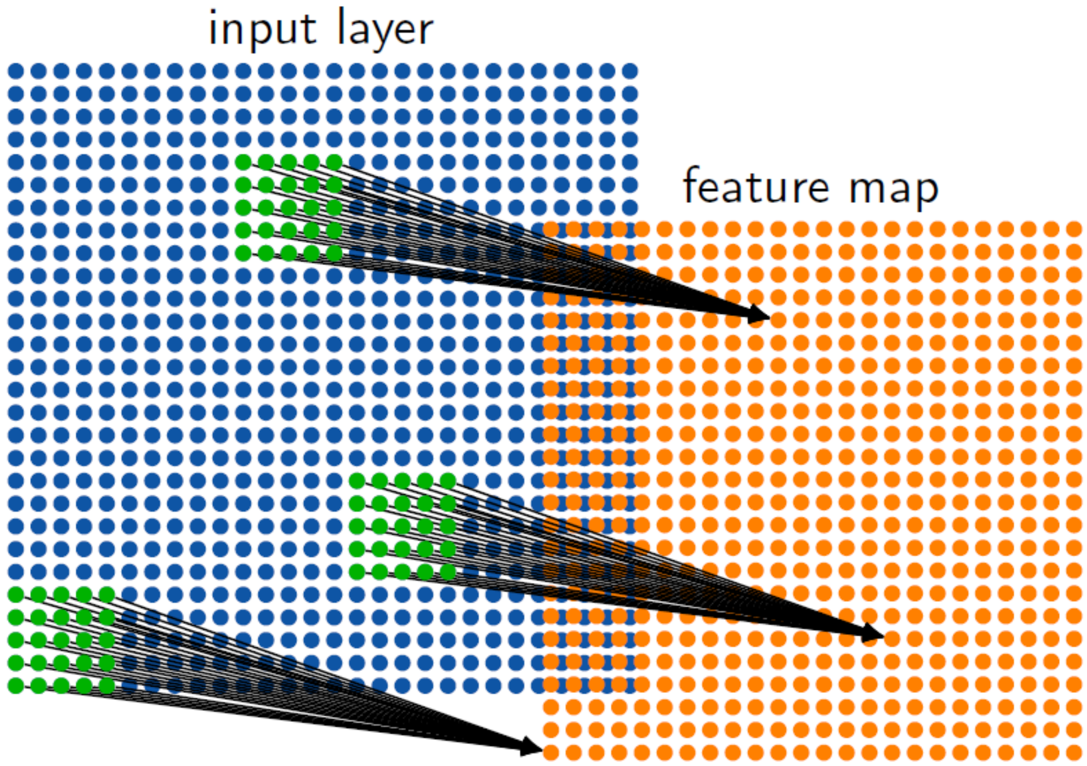
Examples:

CHERRY

CHERRY

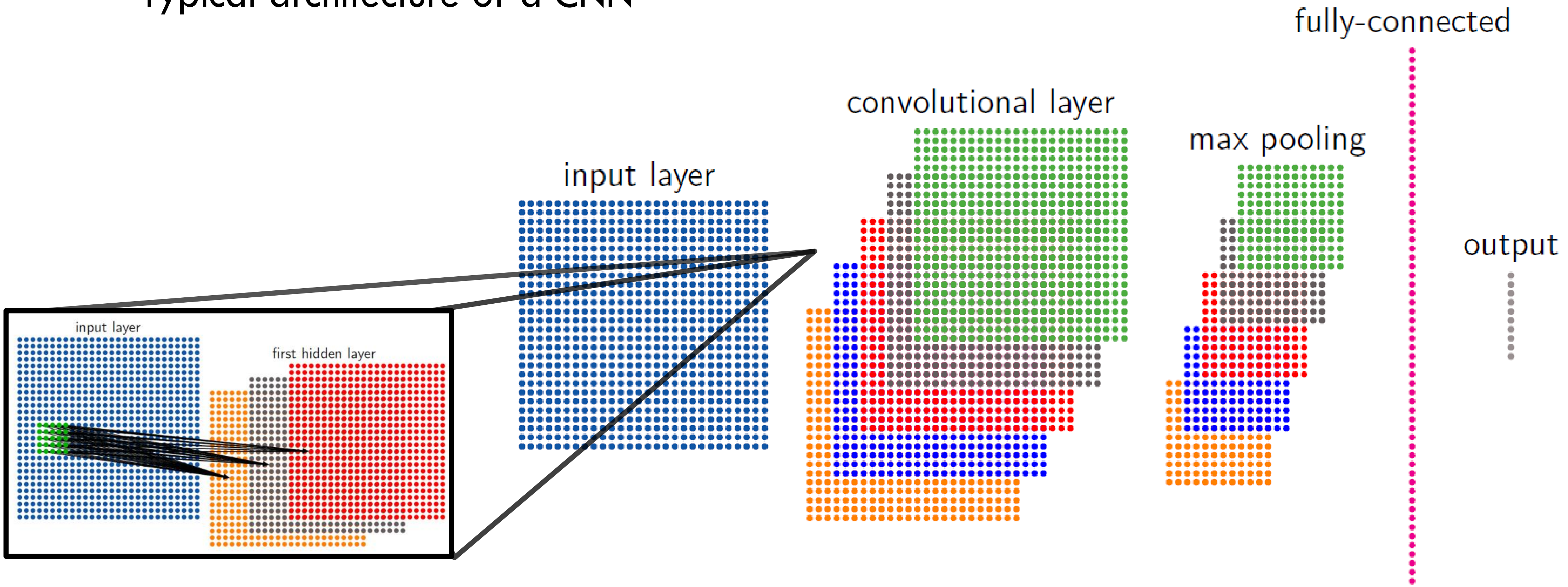
THE APPROACH | CONVOLUTIONAL NEURAL NETWORKS

We want a NN which can extract features from anywhere in an image



THE APPROACH | CONVOLUTIONAL NEURAL NETWORKS

Typical architecture of a CNN

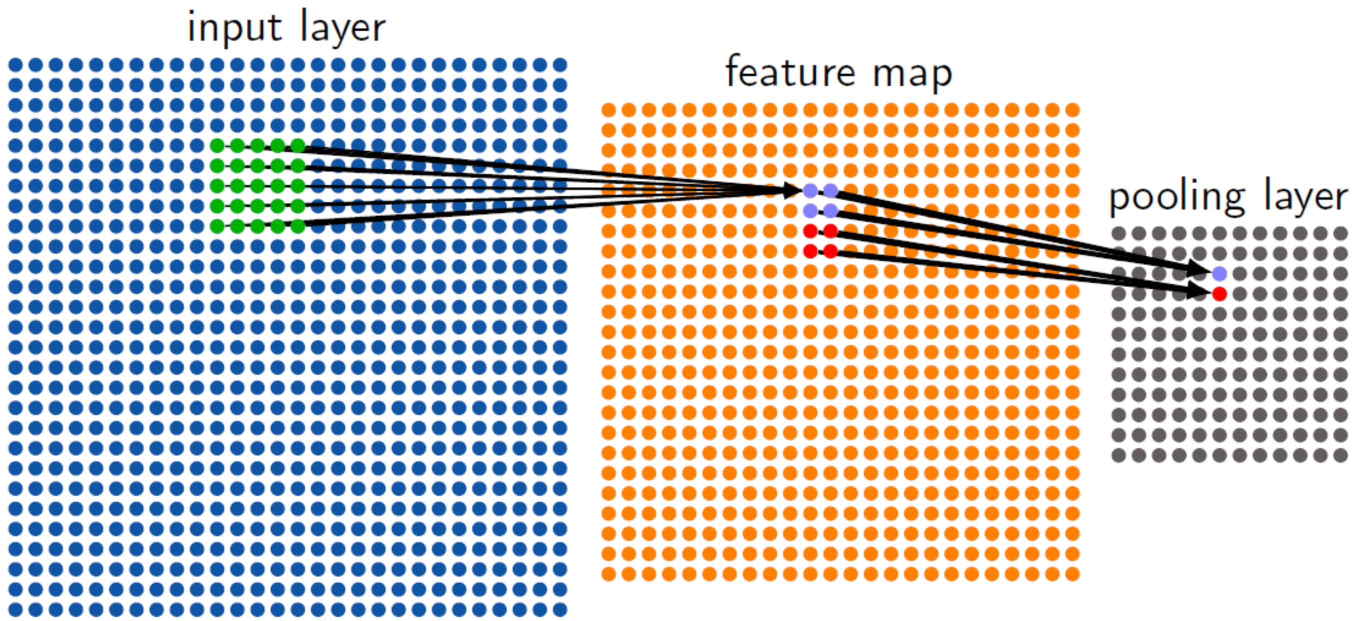


ARCHITECTURE OF A CNN

"Filters" or "kernels" refer to feature maps

There must be multiple filters in a given layer for an effective CNN

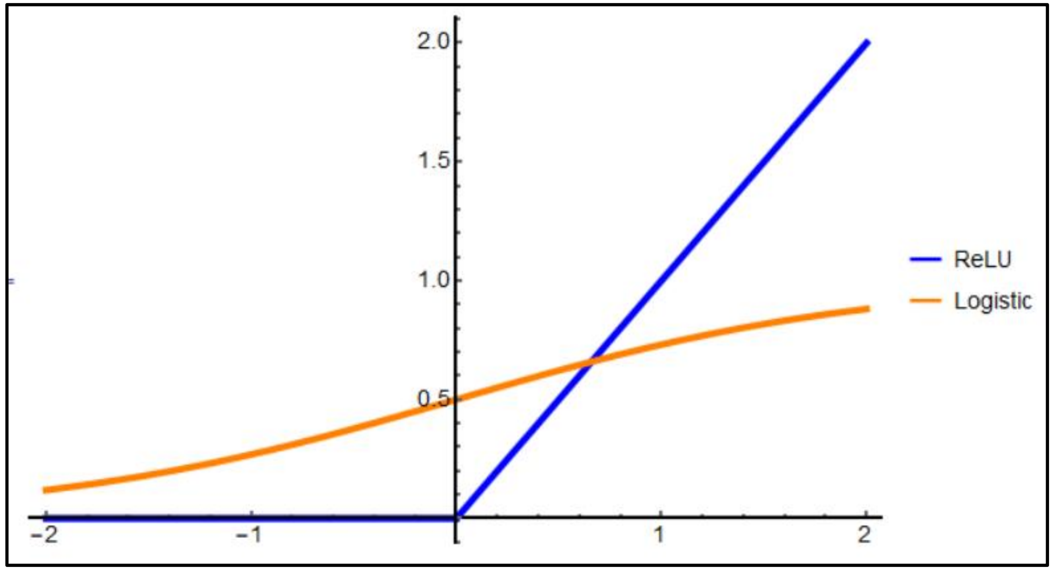
Max pooling is a quick-and-dirty dimensionality reduction technique



THE SUCCESS OF ALEXNET

1. ReLU activation function necessary to train such a large network in reasonable time

$$ReLU(x) = \max(0, x)$$



THE SUCCESS OF ALEXNET

2. Local response normalization

$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^\beta$$

3. Overlapping pooling

   Using stride length less than window helps prevent overfitting

THE SUCCESS OF ALEXNET

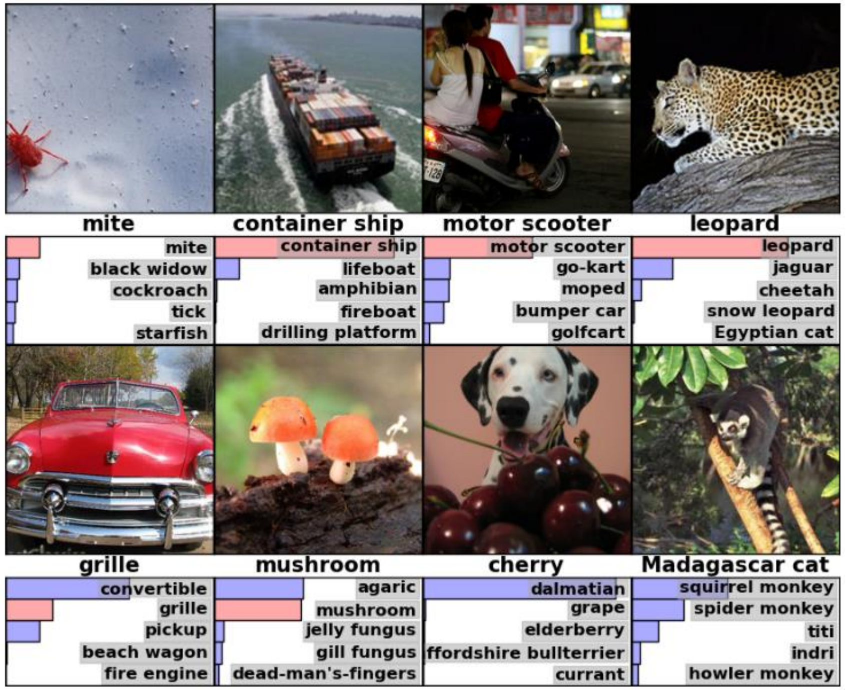
4. Data augmentation
   - Image translations and horizontal reflections
   - Adding colour noise

5. Dropout
   - Used 50% dropout

6. Weight decay

$$v_{i+1} := 0.9 \cdot v_i - 0.0005 \cdot \epsilon \cdot w_i - \epsilon \cdot \left\langle \frac{\partial L}{\partial w}\Big|_{w_i} \right\rangle_{D_i}$$

$$w_{i+1} := w_i + v_{i+1}$$



SOURCES

- Krizhevsky, A., I. Sutskever, and G. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. Proceeding NIPS'12 Proc. 25th Int. Conf. Neural Inf. Process. Syst. 1: 1097–1105.
- DAT112 Lecture Slides – Erik Spence, SciNet HPC Consortium