ImageNet Classification with Deep Convolutional Neural Networks

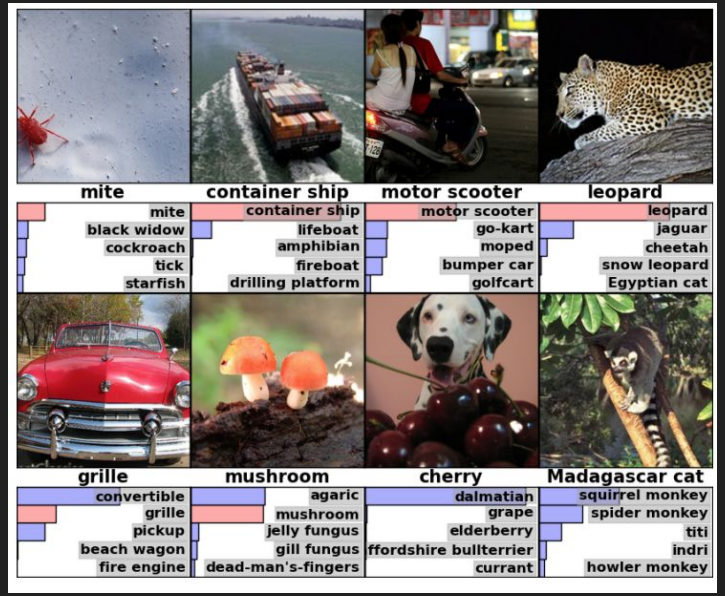
authors: Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton

presented by: John Fang

motivation and overview

- problem: image classification

- ImageNet / ILSVRC
  - 1000 classes
  - 1.2 million labeled training images
  - 150k testing images
  - measured on top-1 and top-5 error rates

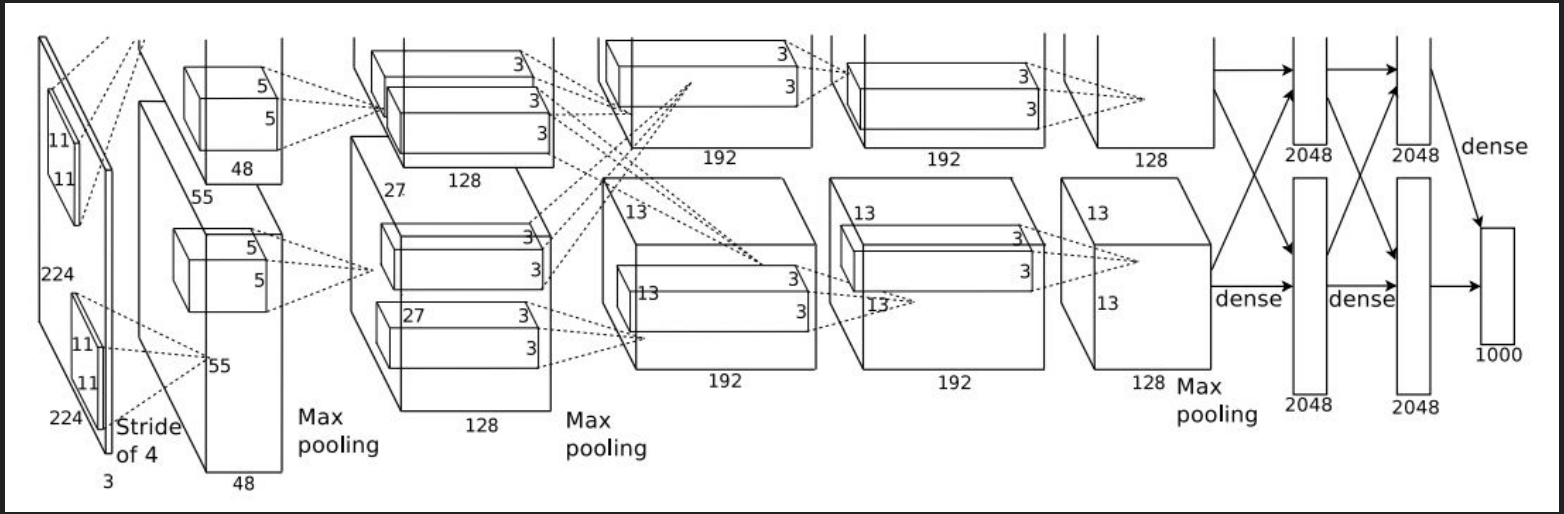


contributions

- created a deep CNN
- evaluated techniques to make this feasible
- publicized their GPU implementation

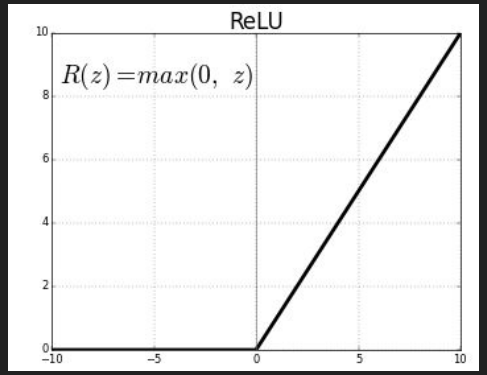
architecture

- 5 convolutional layers, 3 fully connected layers, 1 softmax

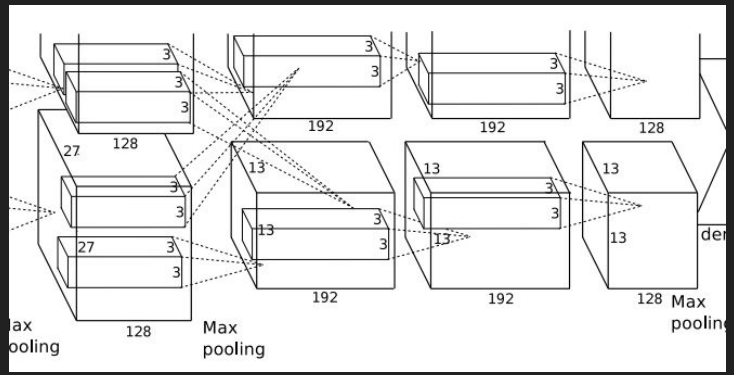


optimizations

- ReLU non-linearity
  - motivation: non-saturating linearity
  - result: trains several times faster



- training across two GPUs
  - motivation: increase size of network
  - result: reduces error by 1.7% (top 1) and 1.2% (top 5)



optimizations continued

- overlapping pooling
  - results: reduces error rates by 0.4% (top 1) and 0.3% (top 5), and reduces overfitting

- local response normalization

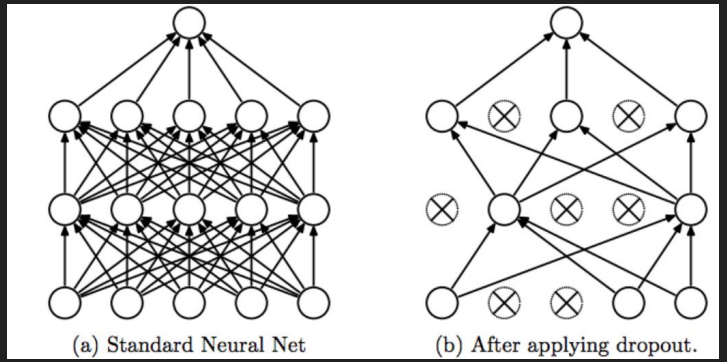
$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^{\beta}$$

  - motivation: "implements a form of lateral inhibition...creates competition for big activities amongst neuron outputs computed using different kernels"
  - results: reduces error by 1.4% (top 1) and 1.2% (top 5)

optimizations to prevent overfitting

- data augmentation
  - increased training set size by 2048x

- dropout



results

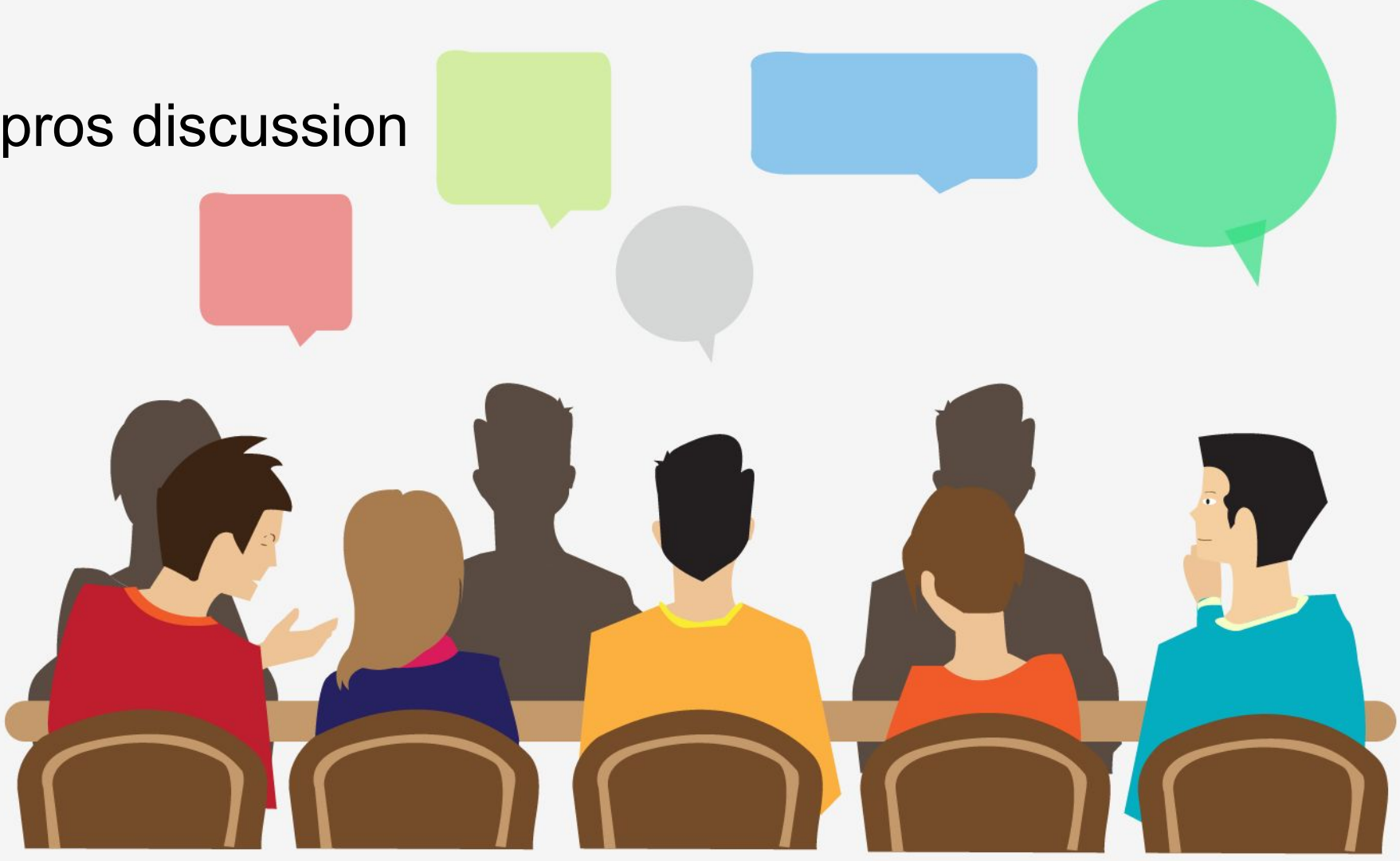
- beat the ILSVRC-2010 winner

- won ILSVRC-2012

Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

Table 1: Comparison of results on ILSVRC-2010 test set. In *italics* are best results achieved by others.

pros discussion



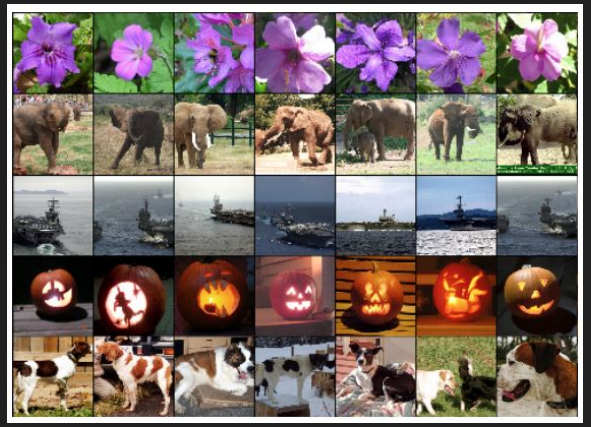
clear writing

- organization is clear

factor of two during training. The recently-introduced technique, called "dropout" [10], consists of setting to zero the output of each hidden neuron with probability 0.5. The neurons which are "dropped out" in this way do not contribute to the forward pass and do not participate in back-propagation. So every time an input is presented, the neural network samples a different architecture, but all these architectures share weights. This technique reduces complex co-adaptations of neurons, since a neuron cannot rely on the presence of particular other neurons. It is, therefore, forced to learn more robust features that are useful in conjunction with many different random subsets of the

- sections are self-contained and readable

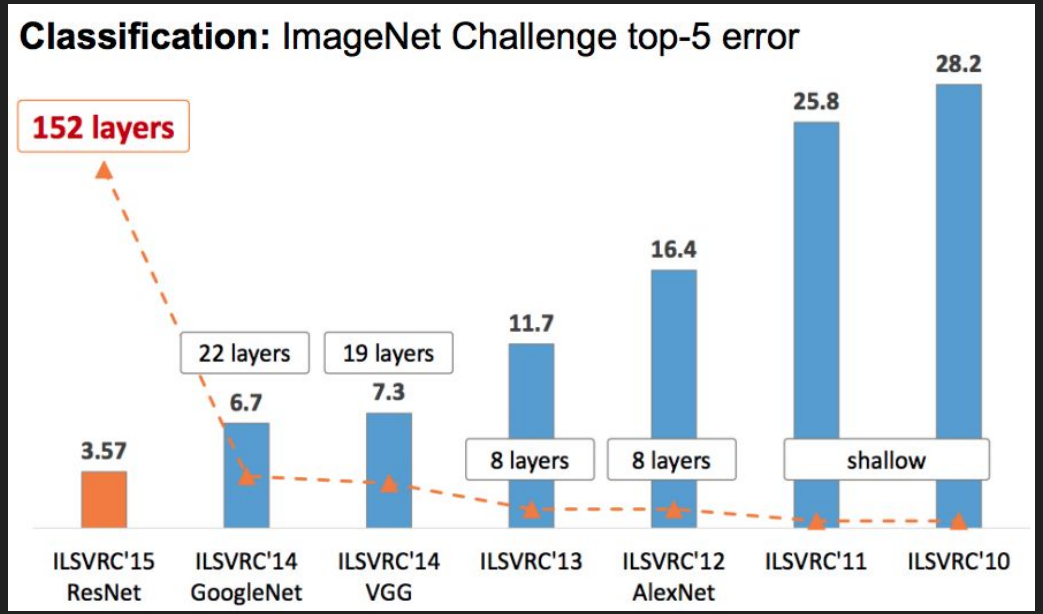
- figures help understand networks



historical significance

- massive improvement compared to previous years

- shift to deep learning



publicization of methods

- makes research reproducible

- could help other researchers

competitions [2] and achieved by far the best results ever reported on these datasets. We wrote a highly-optimized GPU implementation of 2D convolution and all the other operations inherent in training convolutional neural networks, which we make available publicly[1]. Our network contains