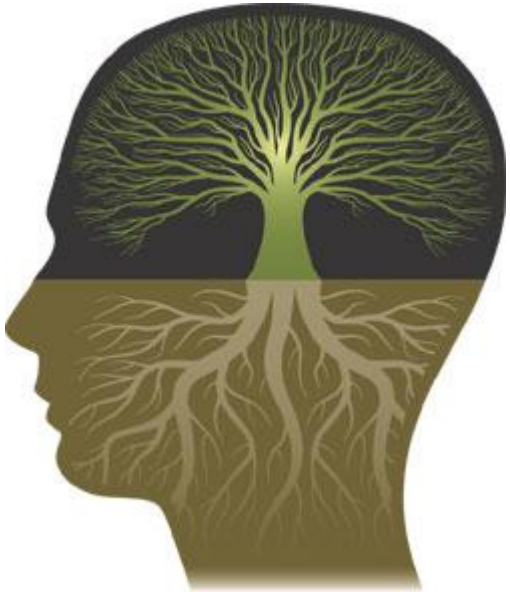
Neural Networks for Vision

Kyle Kastner
University of Texas - San Antonio

What is “Deep Learning”?

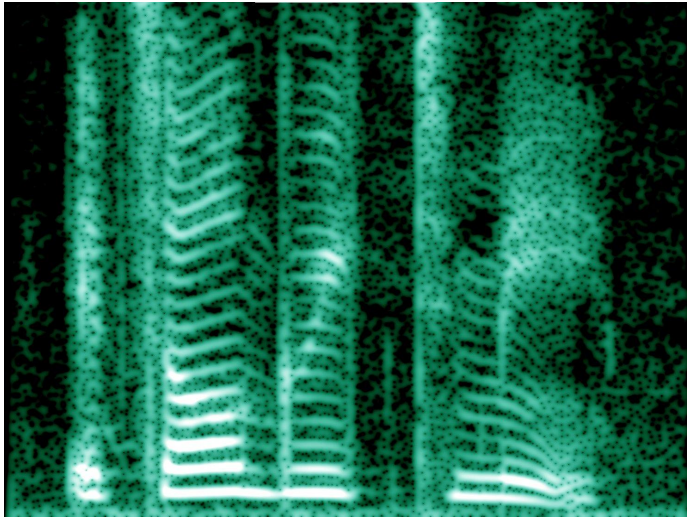
- Improved neural networks
  - Better optimization methods
  - New activation functions
- More layers
- More data, more compute power
- Less feature engineering
- Goal: Hierarchical representations



http://deeplearningblog.com

Modern Applications

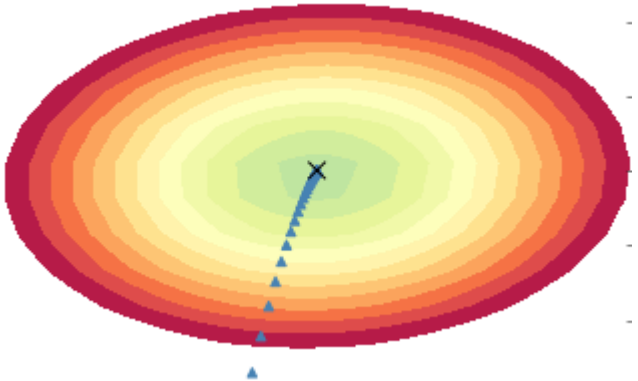
- State of the art in many areas
- Natural Language Processing
- Speech Processing
- Image Processing



http://wikipedia.org

Techniques

- Zero phase components (ZCA) preprocessing
- Trained by stochastic gradient descent
- Convolutional units
- Maxout units
- Dropout
- Softmax classification



http://kastnerkyle.github.io

Preprocessing

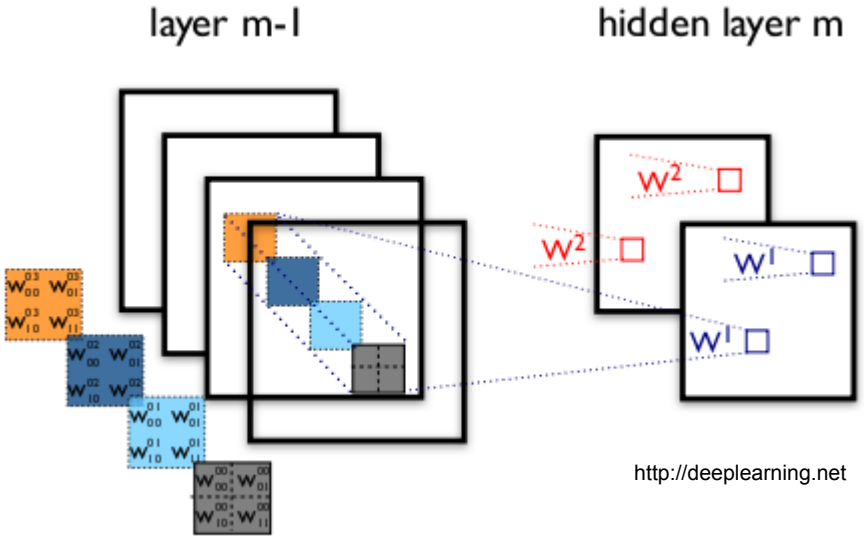
- ZCA mimics the processing of the human visual system
- Similar concept to Principle Component Analysis (PCA)
- Typically used only for images
- **Bell A.J. and Sejnowski T.J. 1996.** *The `Independent Components' of natural scenes are edge filters*

$$W = (XX^T)^{\frac{1}{2}} = ED^{\frac{1}{2}}E^T$$

$$X_w = XW$$

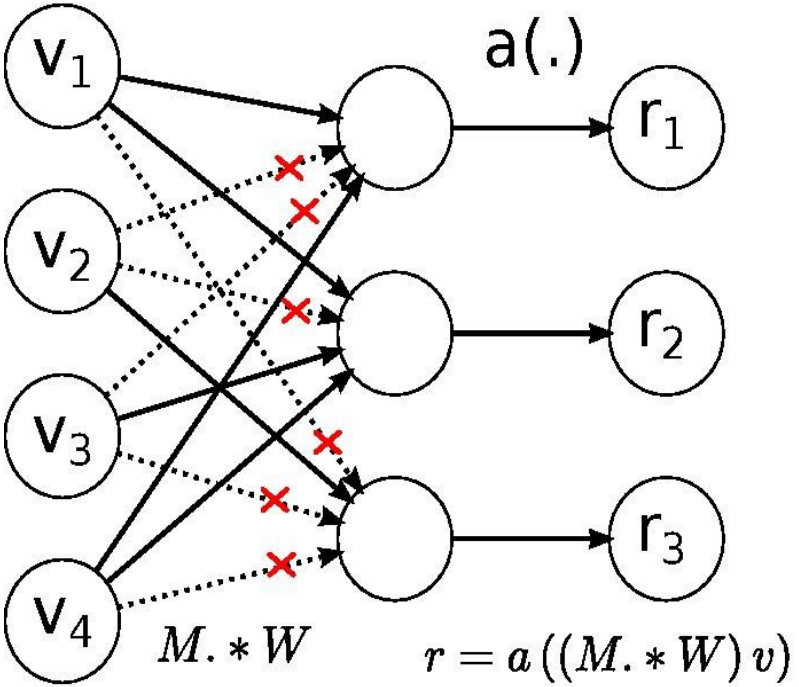
Convolutional Units

- Inspired by research on cat visual cortex
- Used to exploit correlation in input pixels
- Primarily for vision tasks



Dropout

- Randomly drop ~50% of input
- Typically keep 80% in first layer
- Equivalent to training many nets
- Very strong regularizer
- Limit weight adaptation
- Typically ReLU or Maxout units



http://cs.nyu.edu

Maxout

- Uses localized “max” units to
- Shown to be an approximator for many functions
- Designed to be used in conjunction with dropout

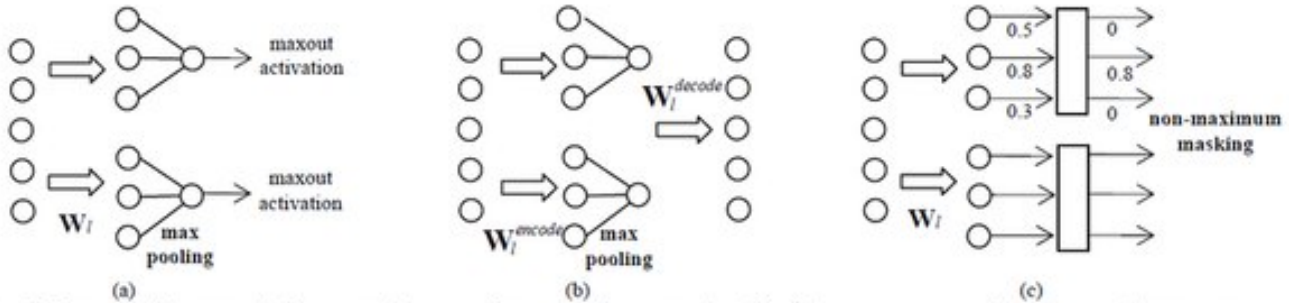
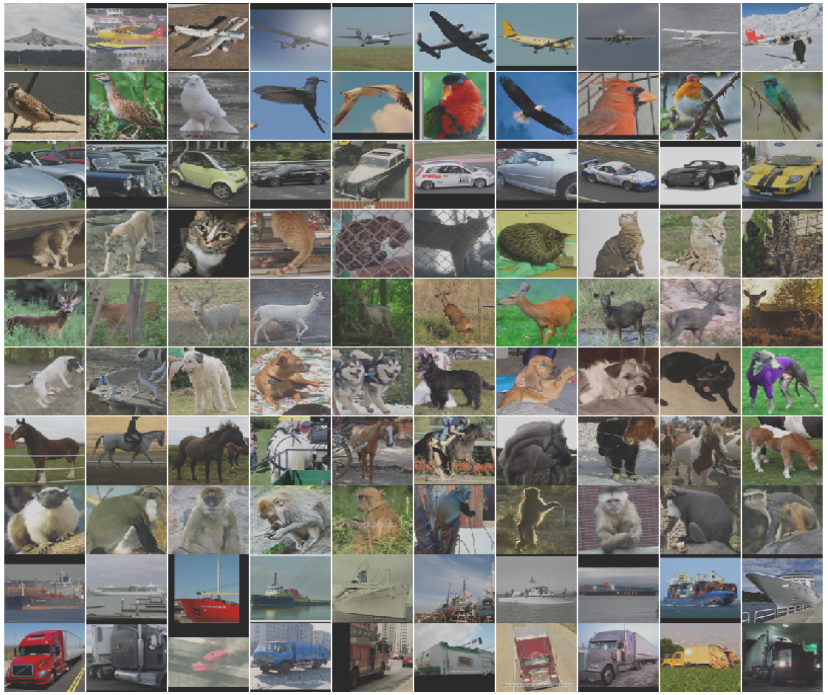


Fig. 1. Maxout architectures in this paper: (a) maxout layer with the group size of 3; (b) maxout autoencoder; (c) sparse feature extractor.

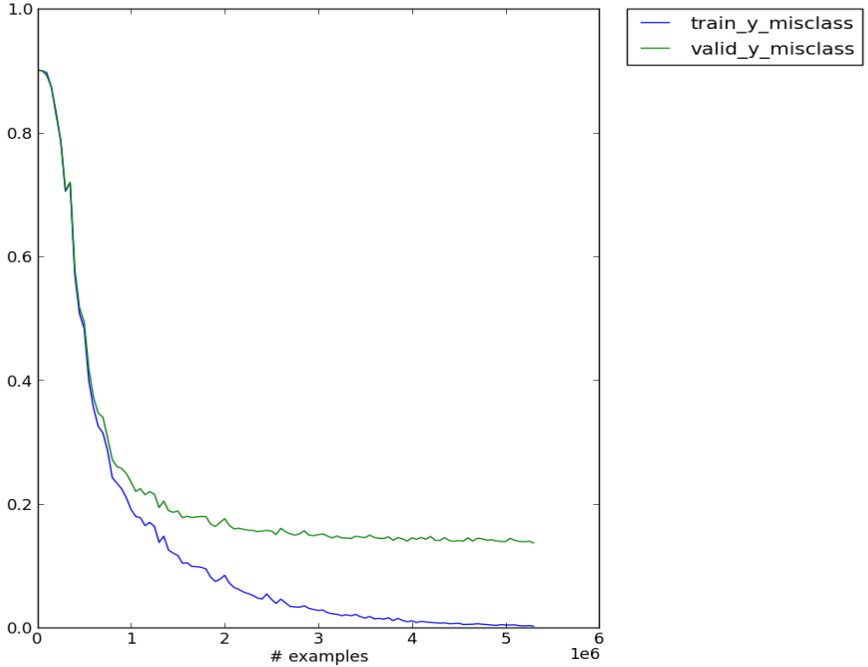
https://portal.xsede.org/publications/-/publications/715

First Experiment

- CIFAR10 dataset
- Standard benchmark for vision
- Replicate performance
- Planned to extend for Asirra
- Results: **86.25**% accuracy
- Similar:
  - CIFAR100
  - ImageNet, STL-10



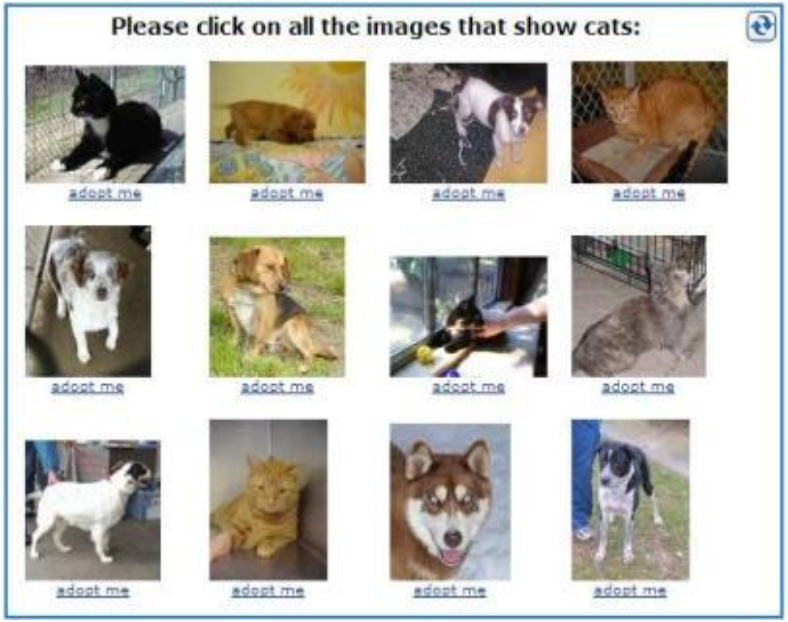
http://www.stanford.edu/~acoates//stl10/



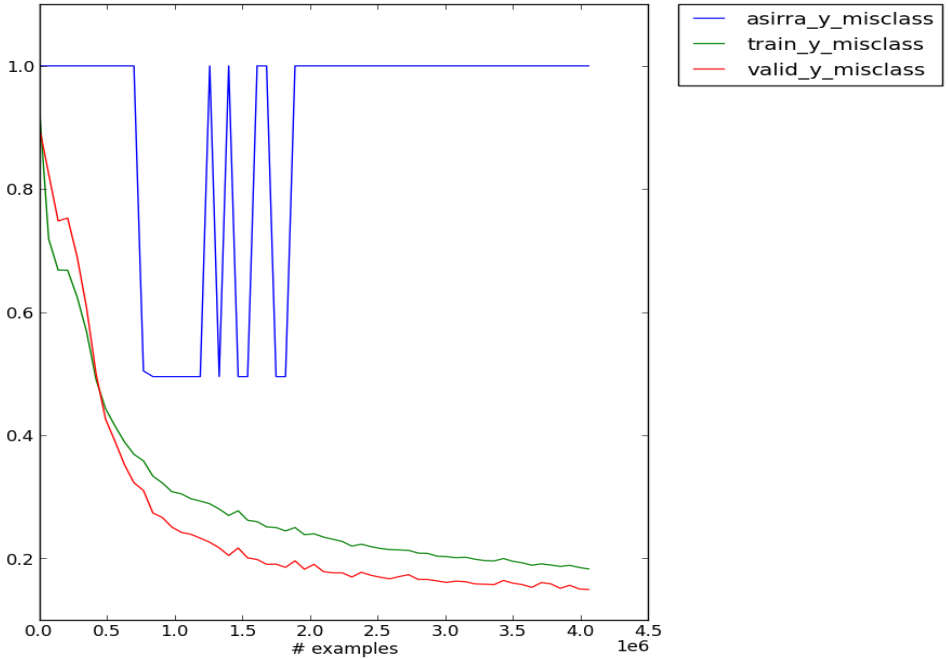
Results from CIFAR10 Training

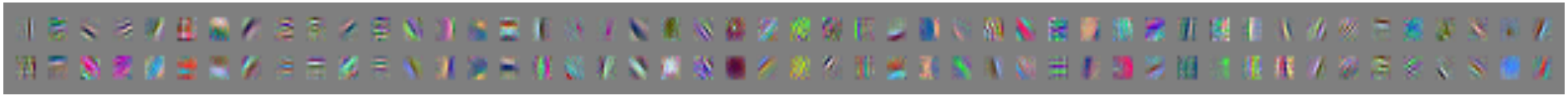
Failed Extension

- Image based CAPTCHA
- Developed by Microsoft Research
- Kaggle Dogs vs. Cats
- SOTA: **96**%
- Best accuracy: **50%**
- This is **BAD!**



http://www.peta.org





Failed results from Asirra + CIFAR10 Training

Future Work

- Validated work done at University of Montreal
- Extension to (very similar) dataset is not working
- Need better class discrimination
- Get Asirra dataset working!
- Extend to problems with time series (accelerometer, speech)

?