Data Mining: Introduction to Deep Learning

马锦华

Edited based on Ian Goodfellow's slides from

www.deeplearningbook.org

数据科学与计算机学院

中山大学

Introduction to Deep Learning

- **Architecture Design**

- Convolutional Neural Networks (CNN)

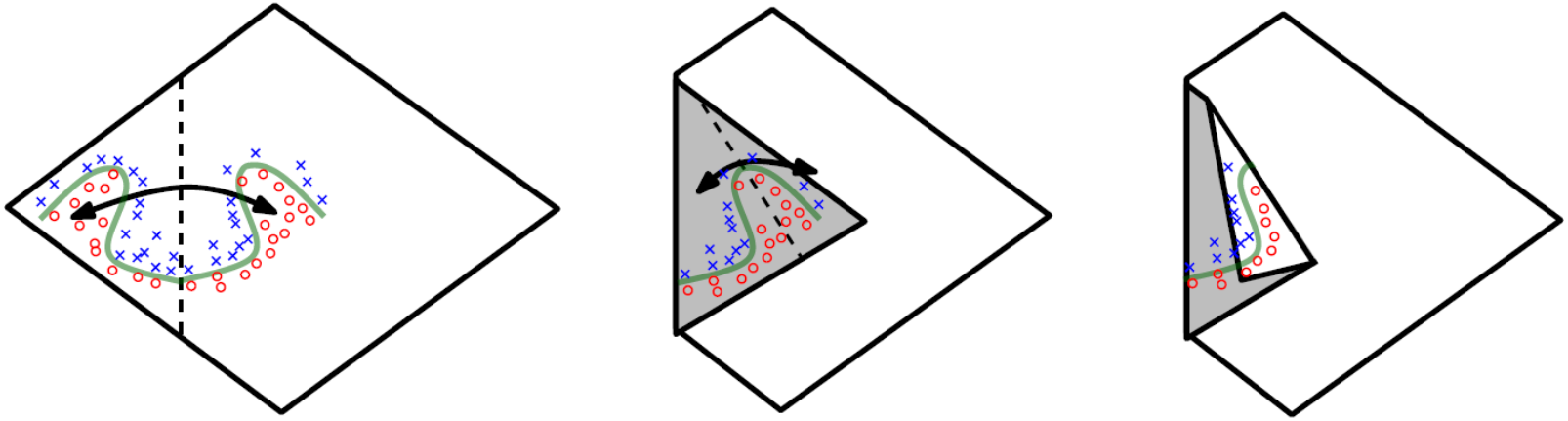
Output Types

Output Type	Output Distribution	Output Layer	Cost Function
Binary	Bernoulli	Sigmoid	Binary cross-entropy
Discrete	Multinoulli	Softmax	Discrete cross-entropy
Continuous	Gaussian	Linear	Gaussian cross-entropy (MSE)
Continuous	Mixture of Gaussian	Mixture Density	Cross-entropy
Continuous	Arbitrary	See part III: GAN, VAE, FVBN	Various

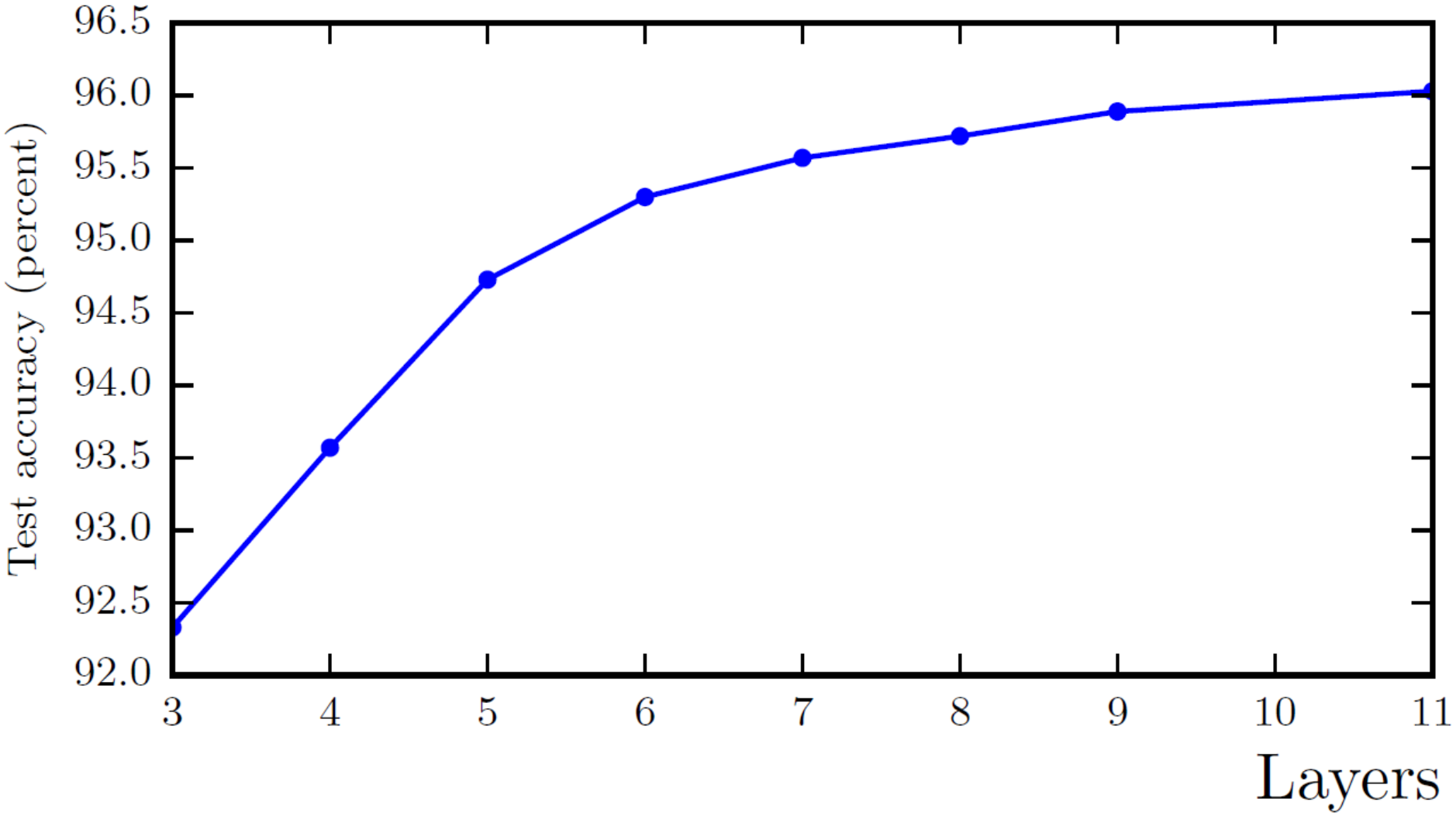
Universal Approximator Theorem

- Why deeper?
  - Shallow net may need (exponentially) more width
  - Shallow net may overfit more

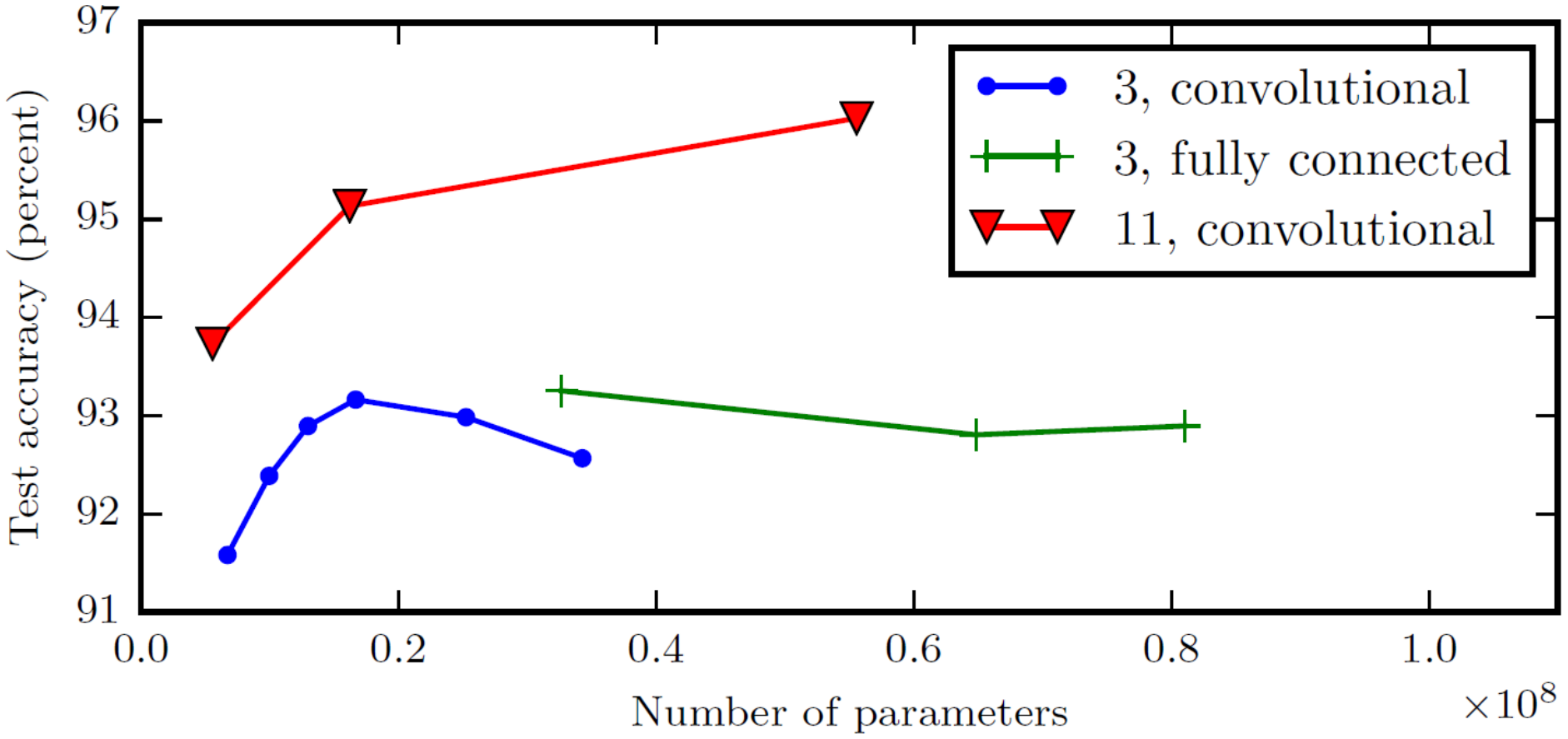
- Illustration of advantage of depth



Better Generalization with Greater Depth



Large, Shallow Models Overfit More



Dataset Augmentation

Affine Distortion

Noise

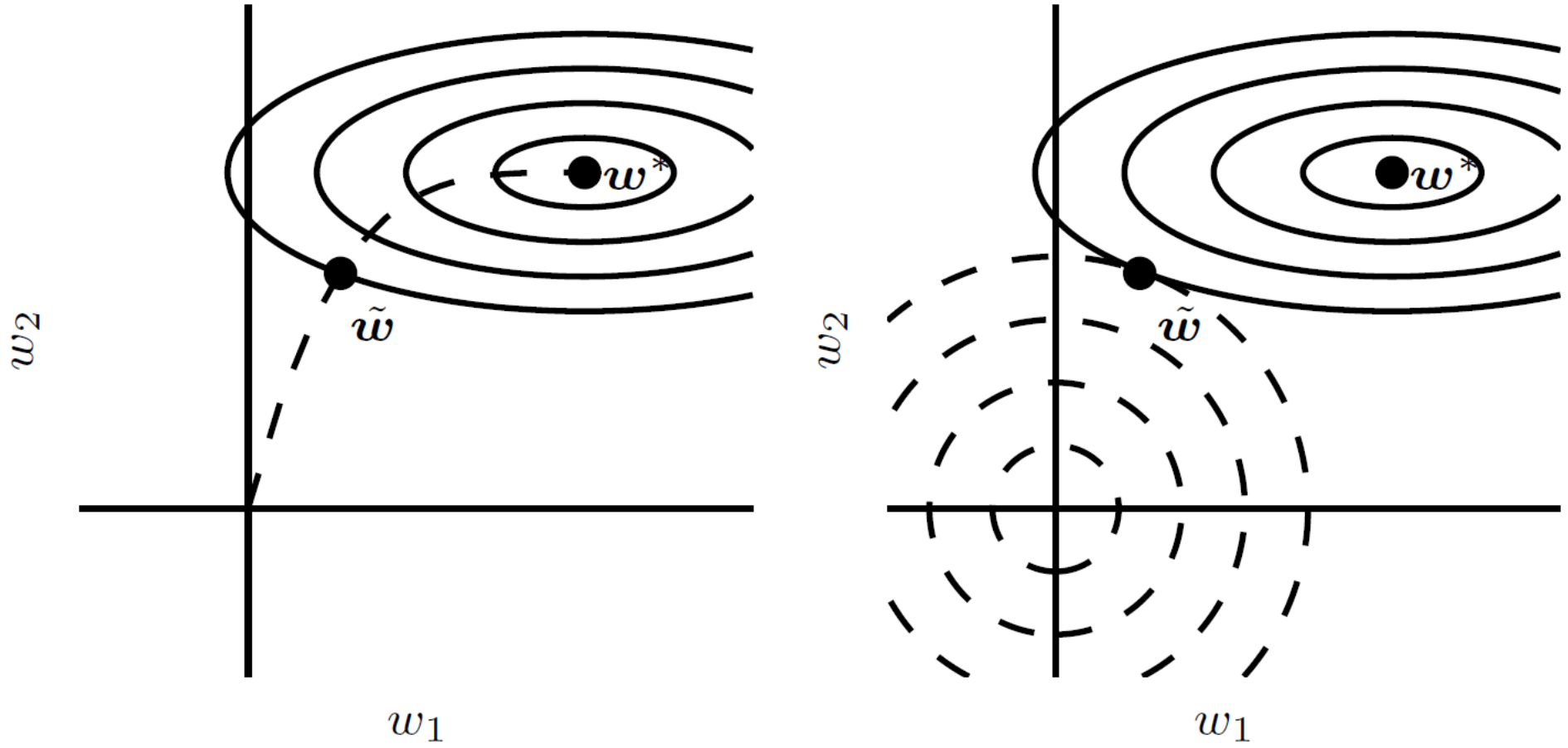
Elastic Deformation

Horizontal flip

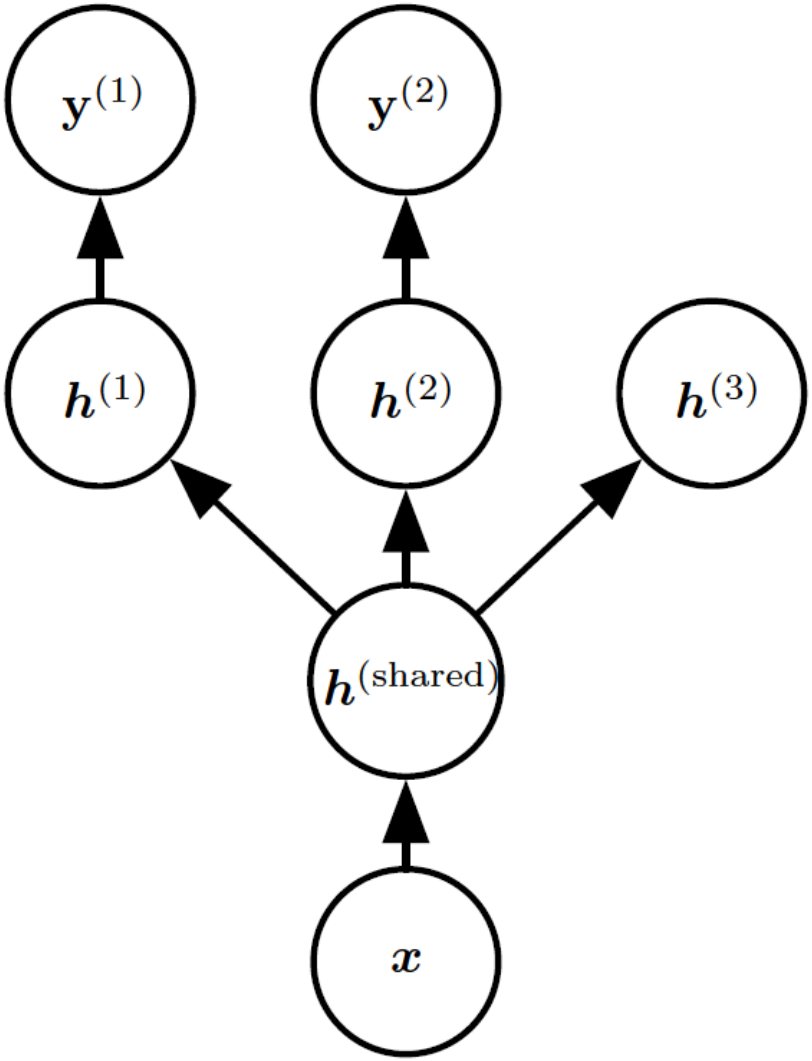
Random Translation

Hue Shift

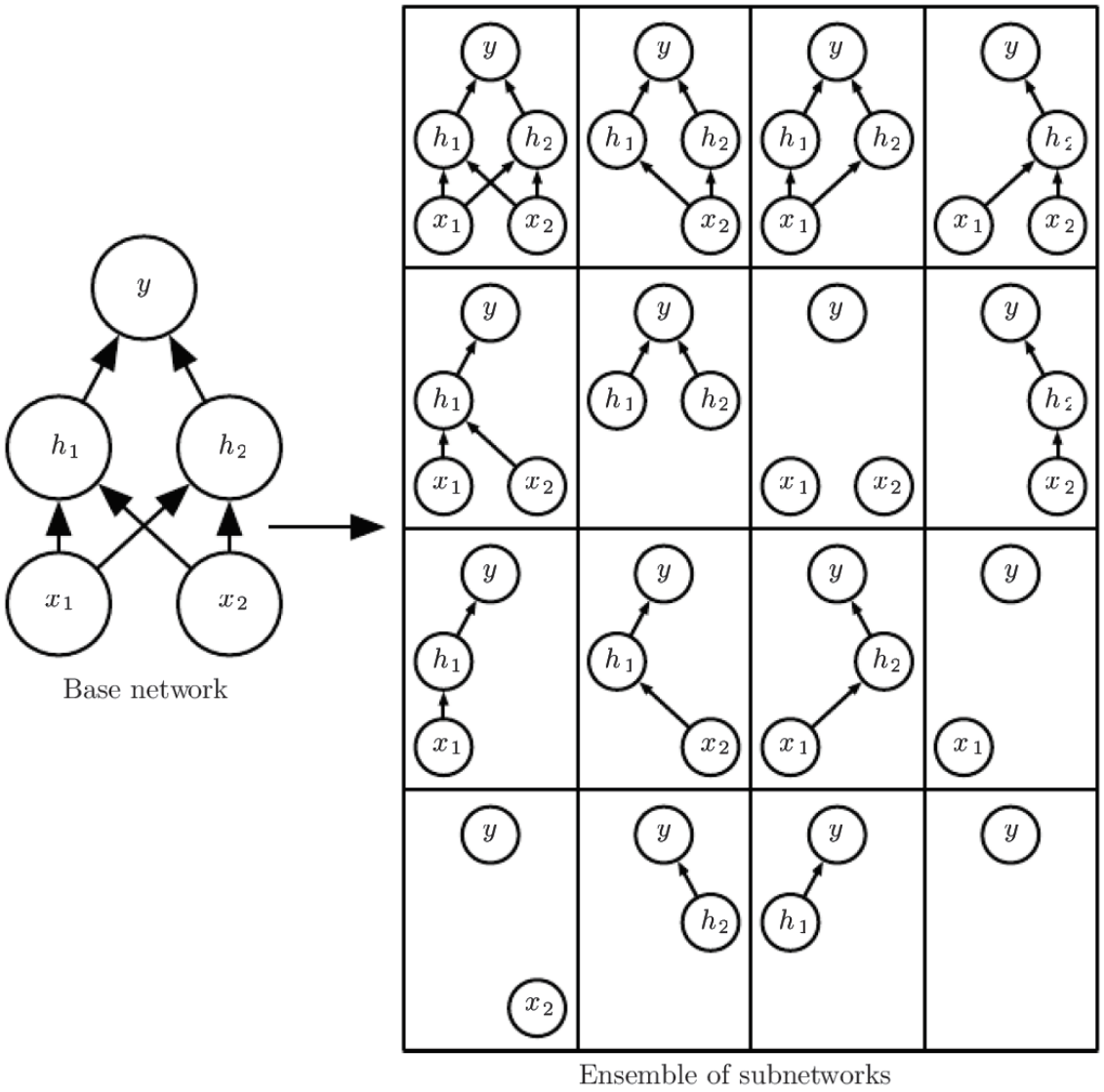
Early Stopping as Regularizer



Parameter Sharing



Dropout



Introduction to Deep Learning

- Architecture Design
- **Convolutional Neural Networks (CNN)**

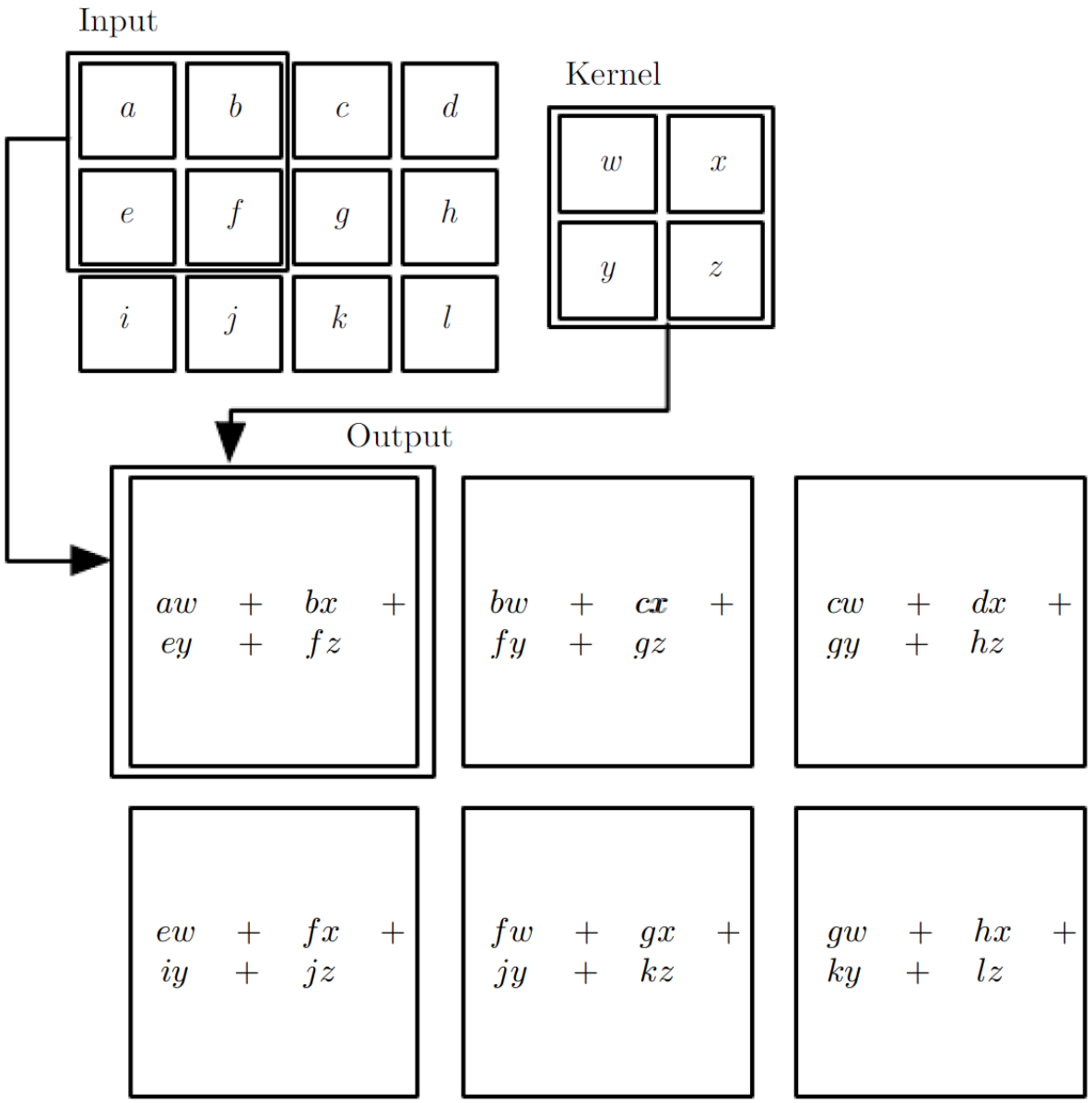
Convolutional Networks

- Scale up neural networks to process very large images / video sequences
  - Sparse connections
  - Parameter sharing
- Automatically generalize across spatial translations of inputs
- Applicable to any input that is laid out on a grid (1-D, 2-D, 3-D, …)

Key Idea

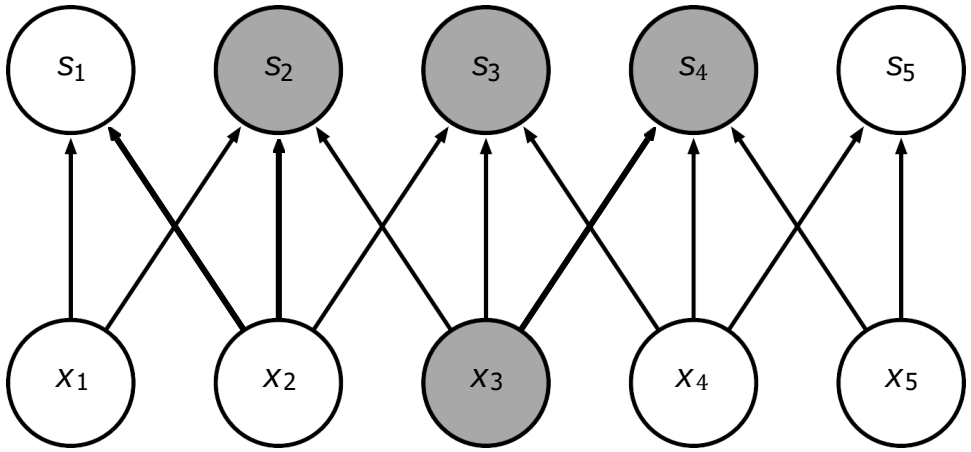
- Replace matrix multiplication in neural nets with convolution
- Everything else stays the same
  - Maximum likelihood
  - Back-propagation
  - etc.

2D Convolution

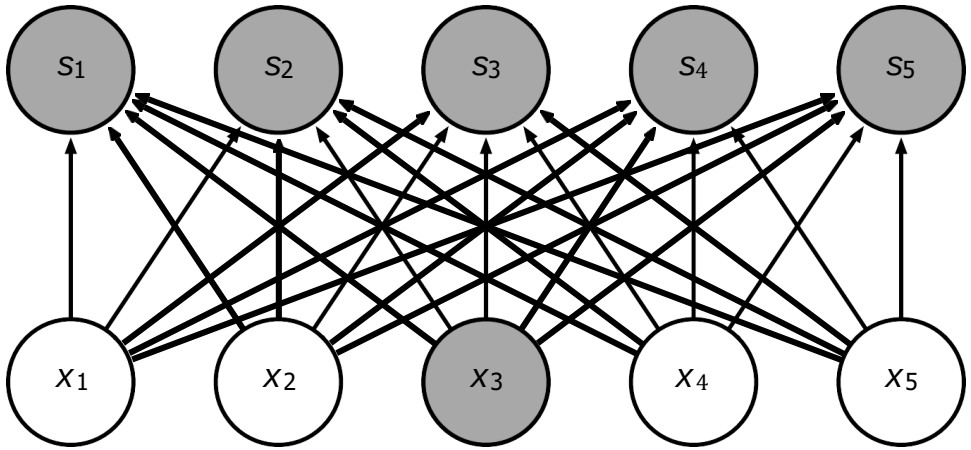


Sparse Connectivity

Sparse connections due to small convolution kernel

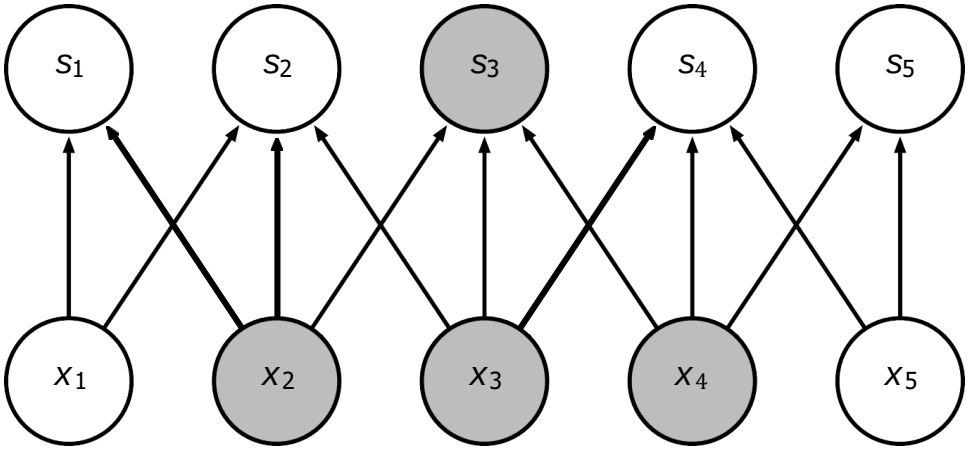


Dense connections

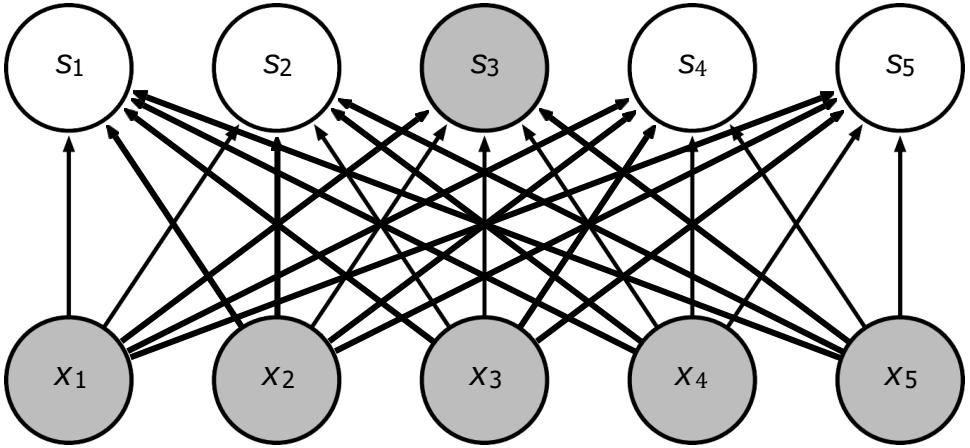


Sparse Connectivity

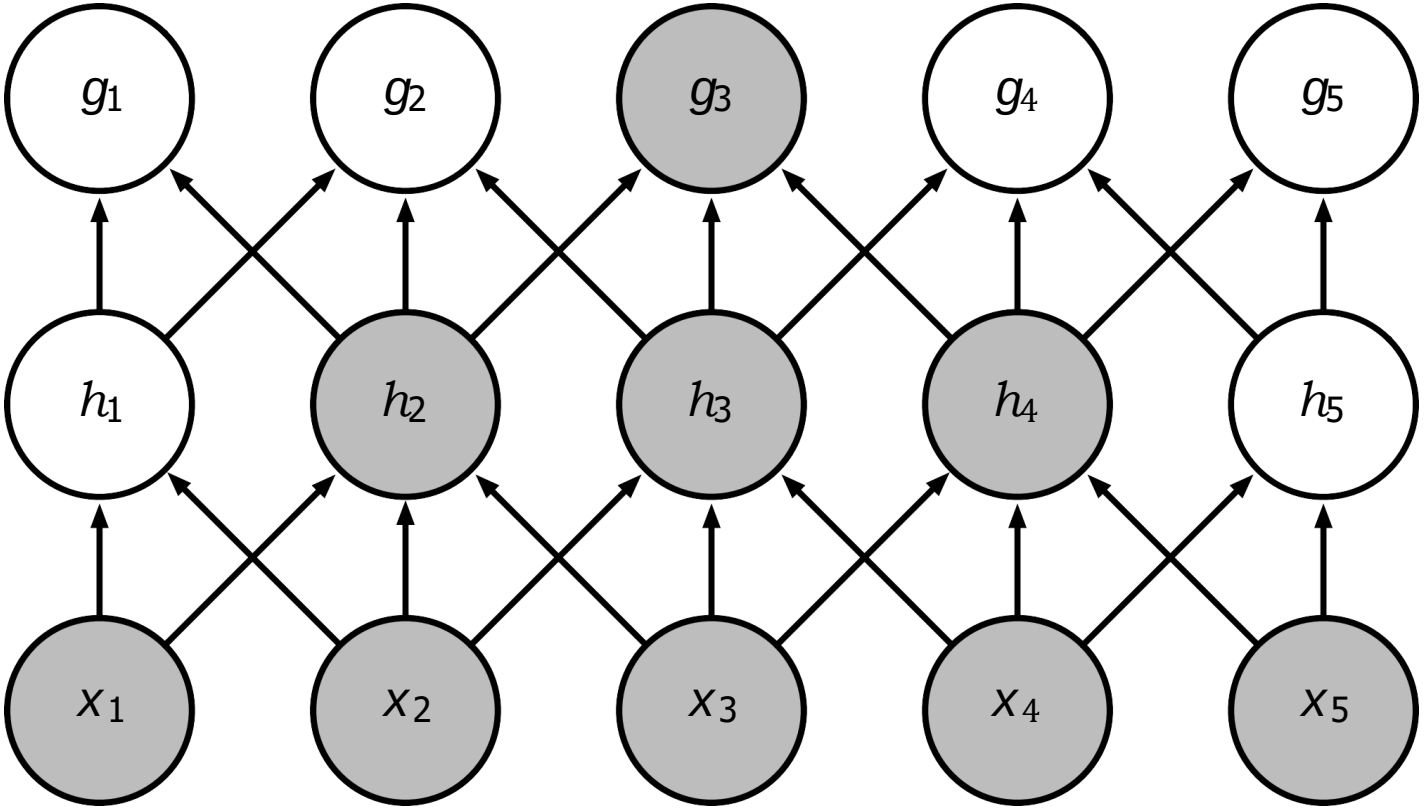
Sparse connections due to small convolution kernel



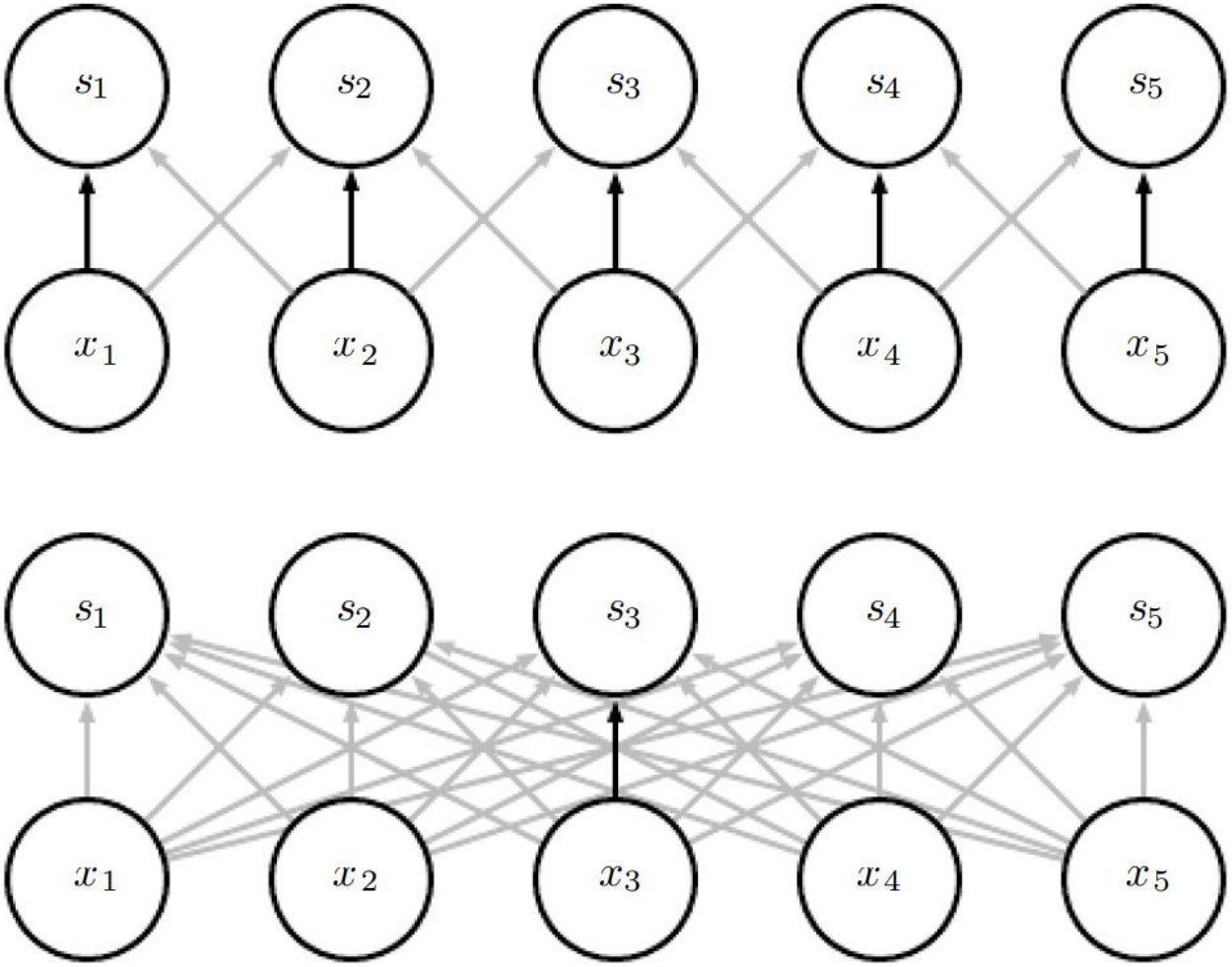
Dense connections



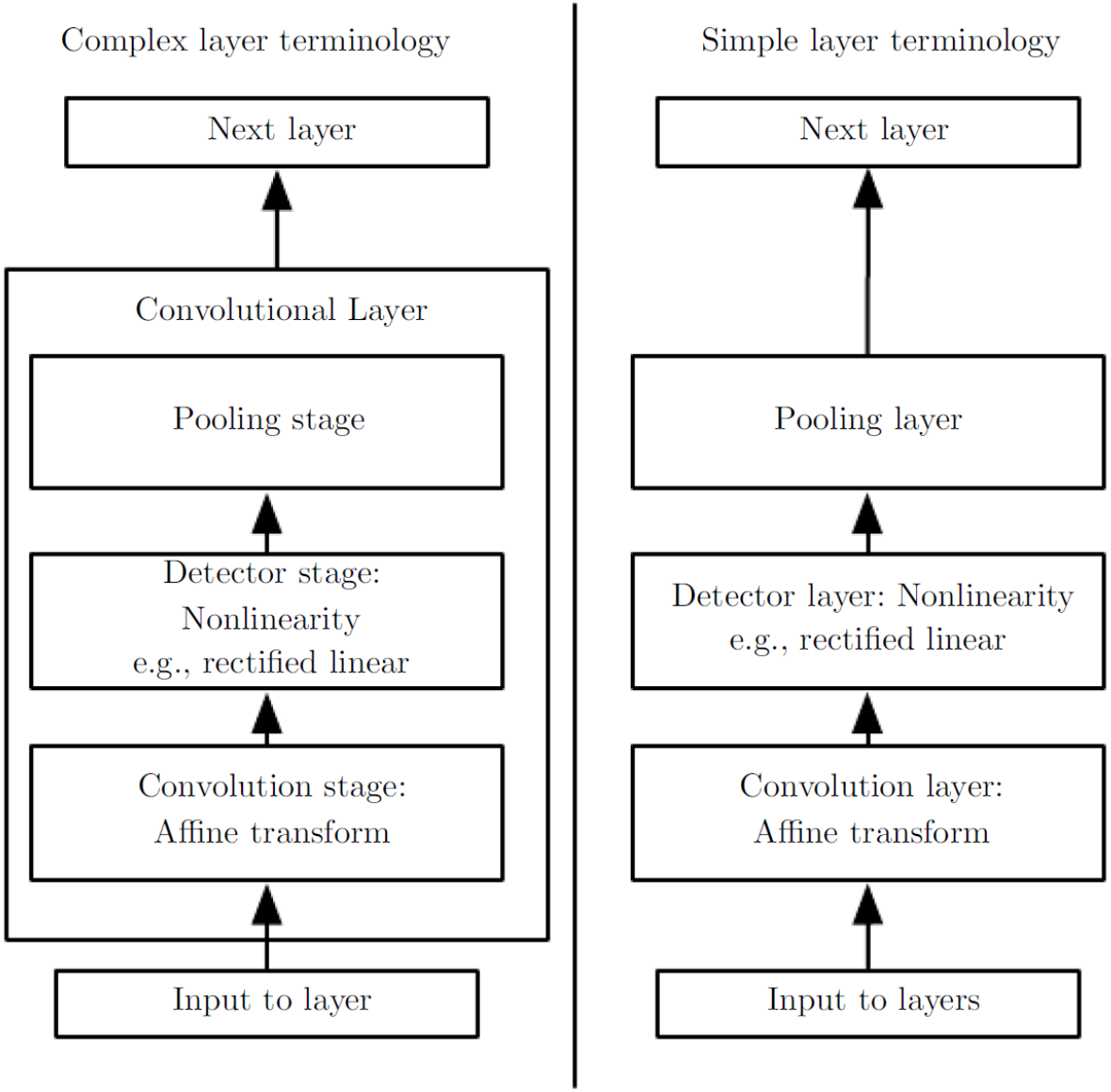
Growing Receptive Fields



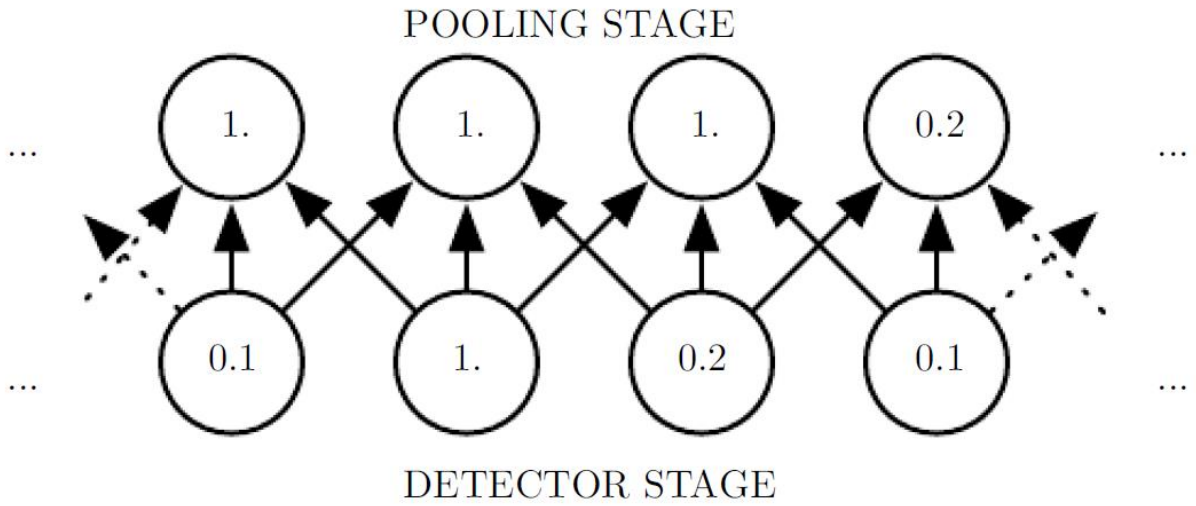
Parameter Sharing

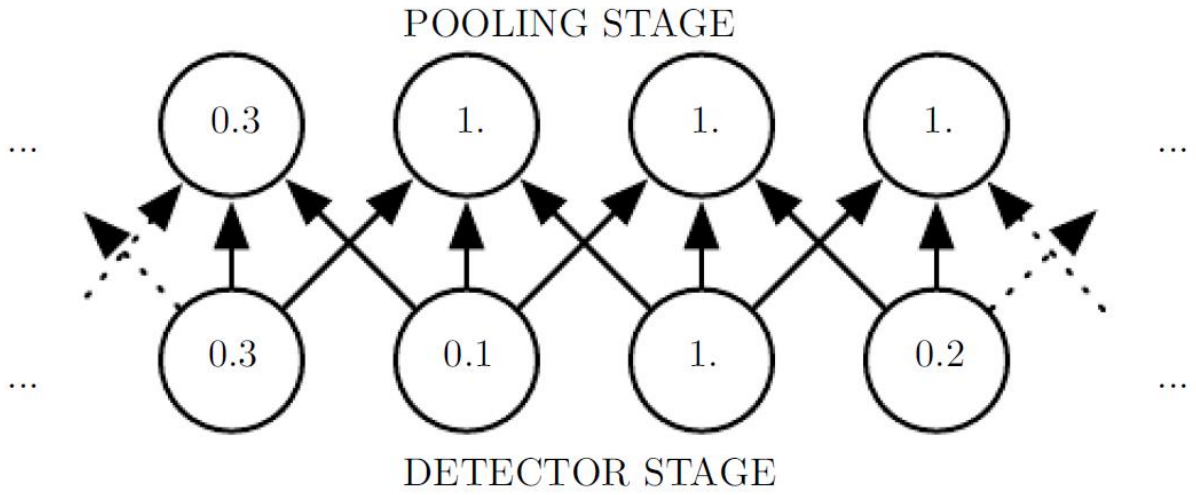


Convolutional Network Components

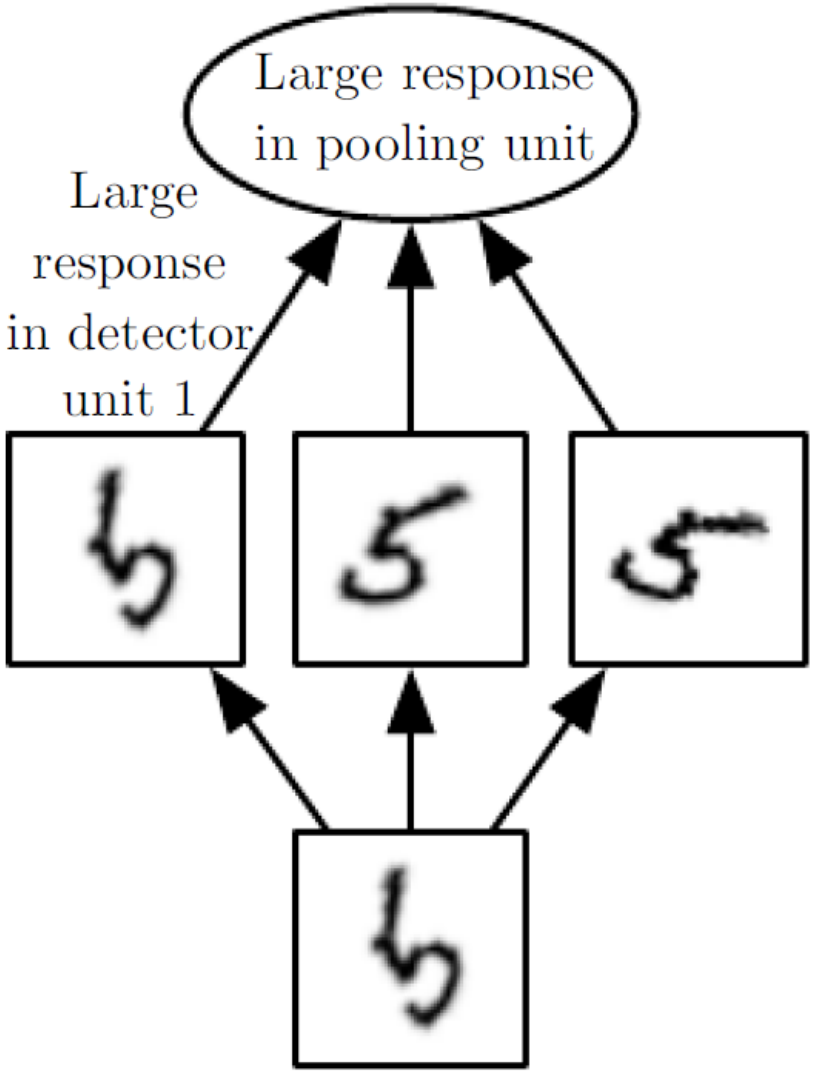


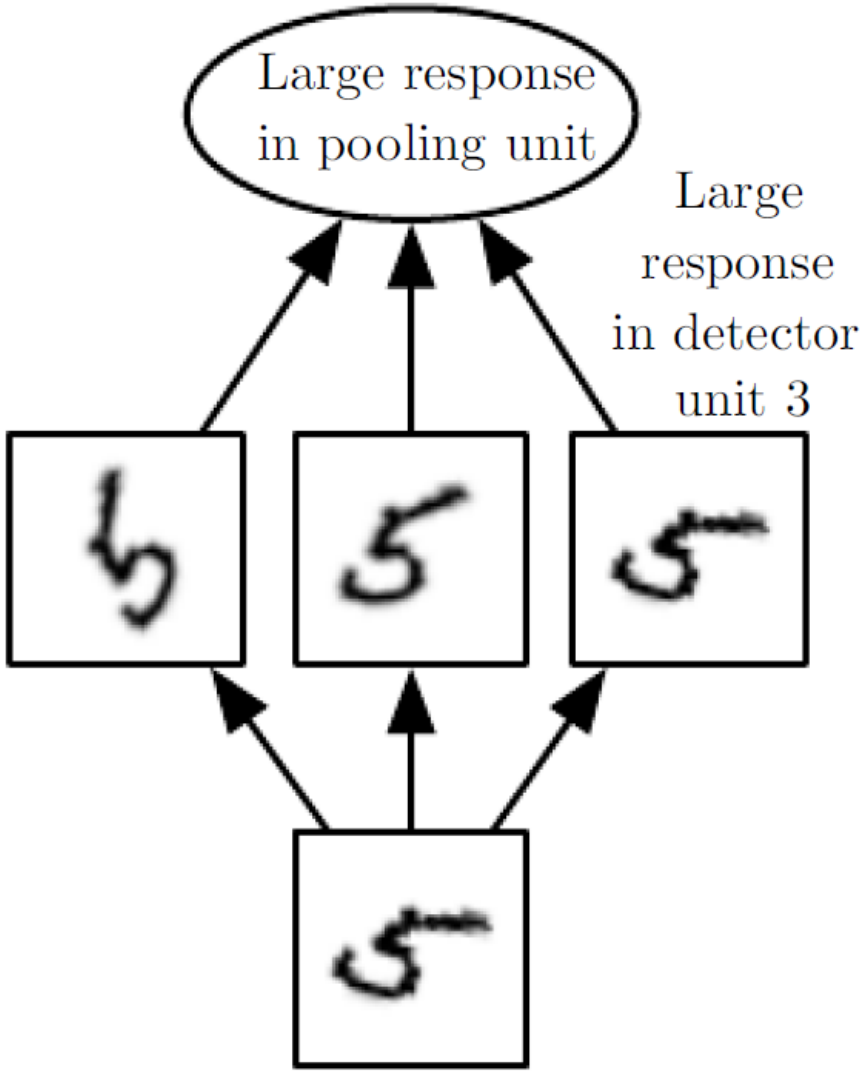
Max Pooling and Invariance to Translation



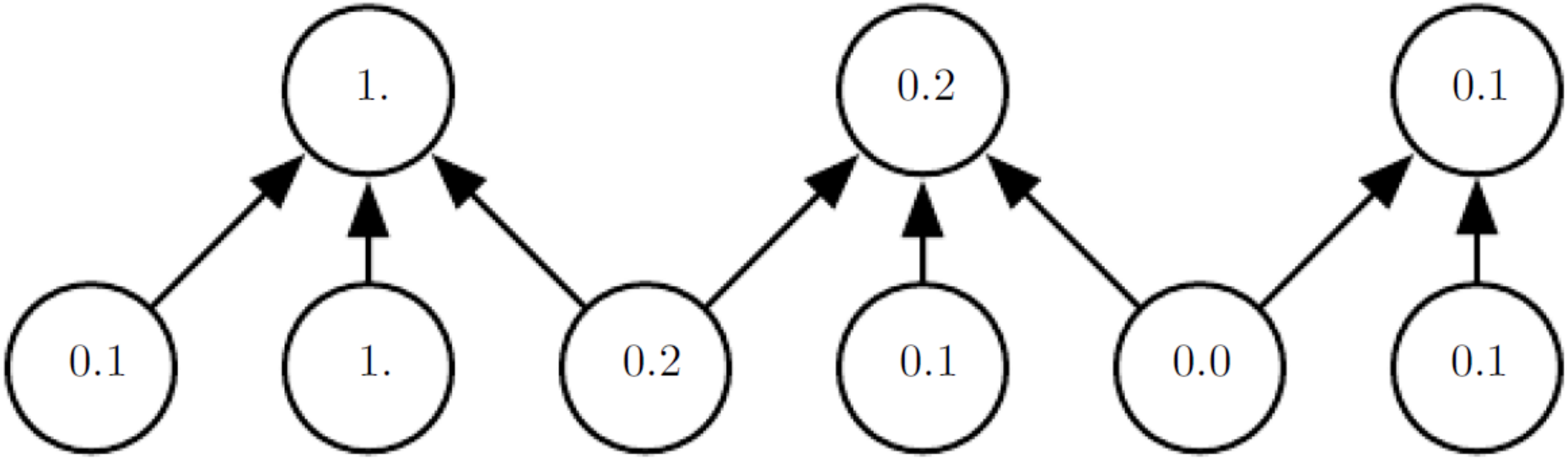


Cross-Channel Pooling and Invariance to Learned Transformations

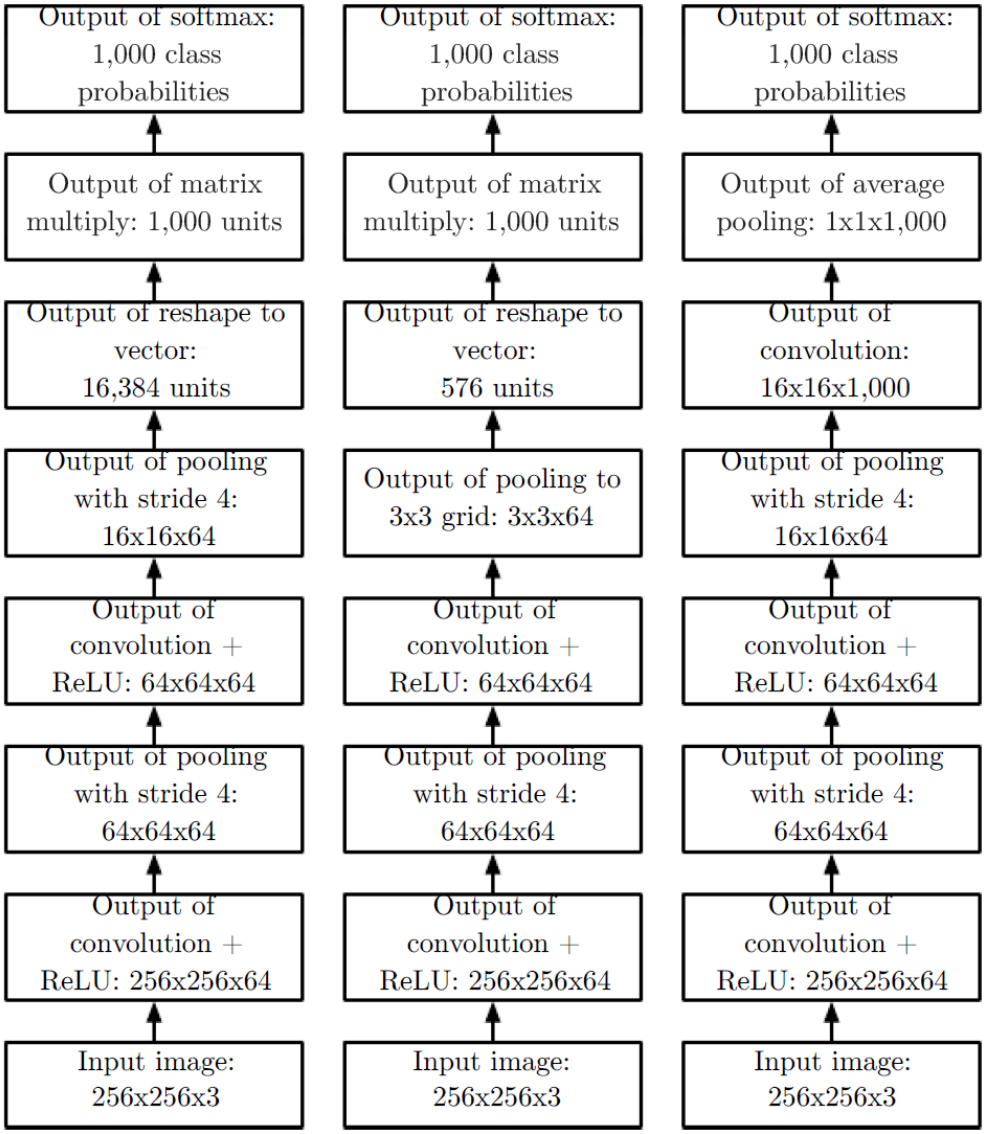




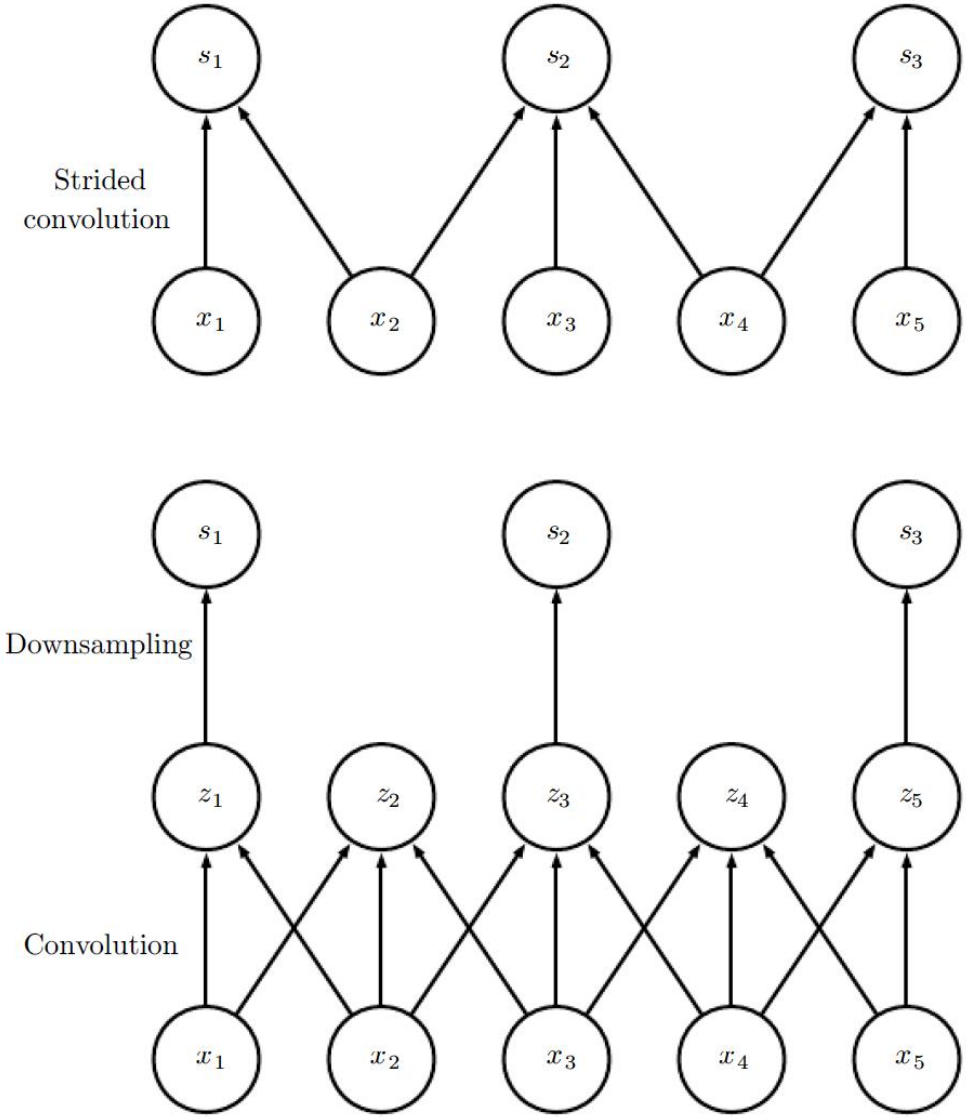
Pooling with Downsampling



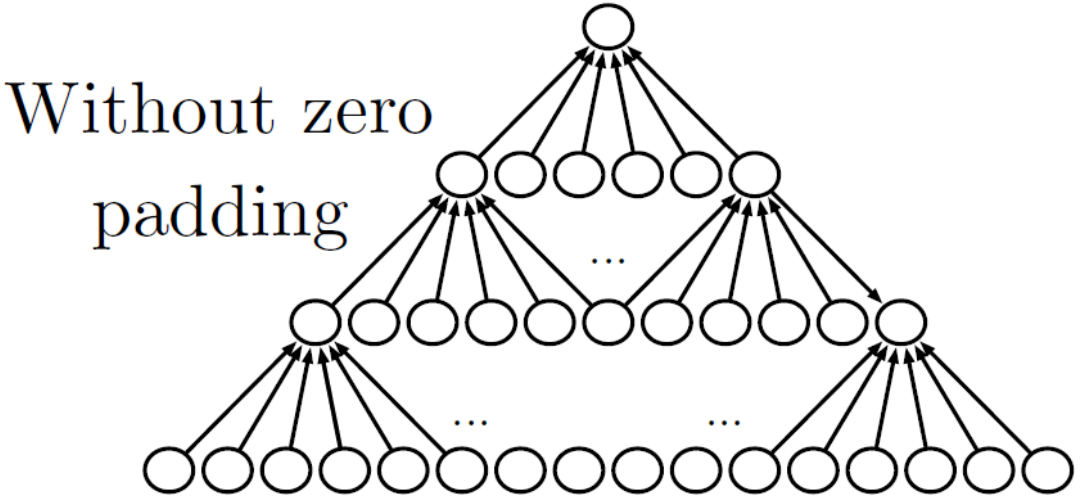
Example Classification Architectures



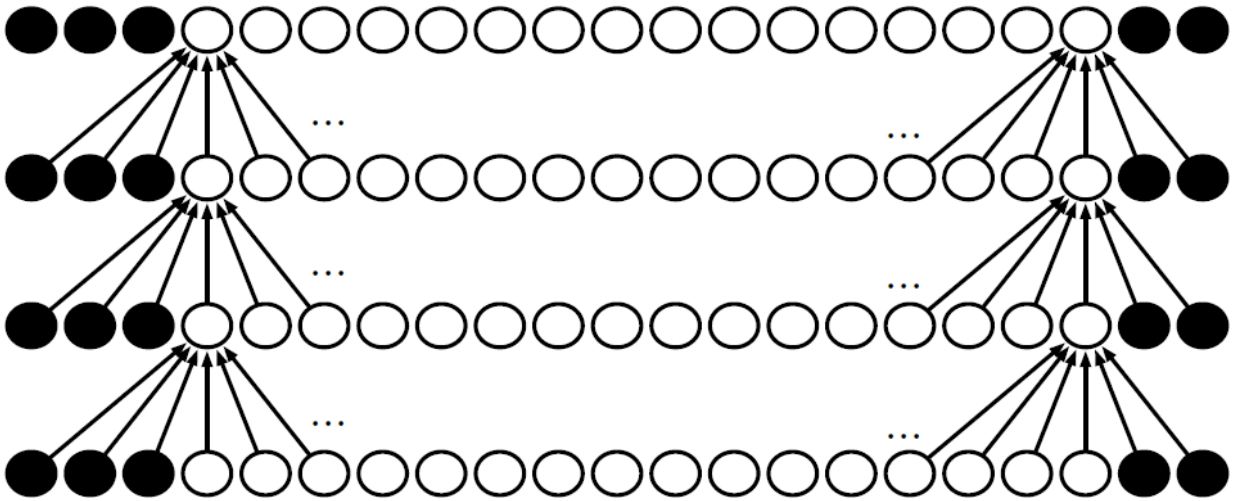
Convolution with Stride



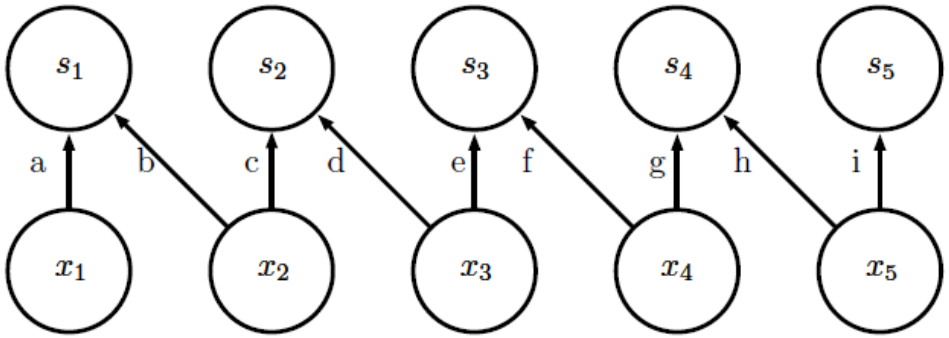
Zero Padding Controls Size



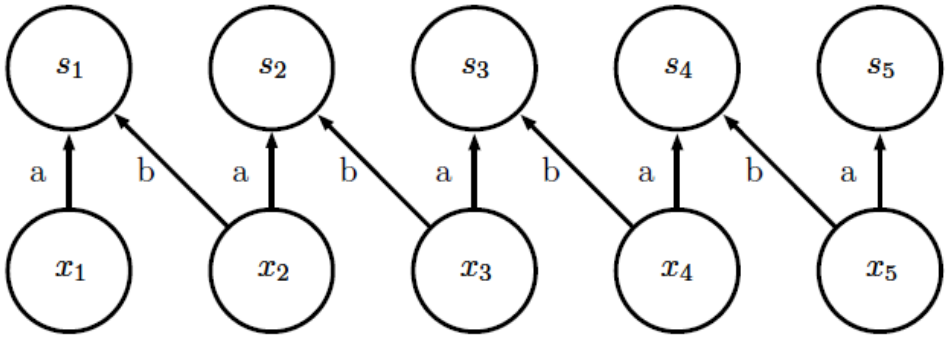
With zero
padding



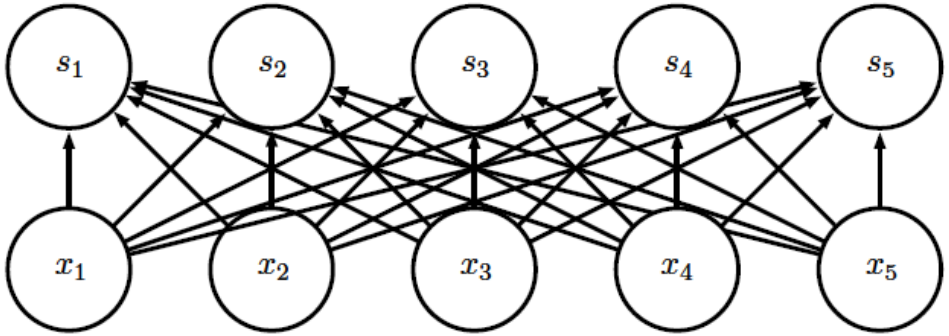
Kinds of Connectivity



Local connection: like convolution, but no sharing

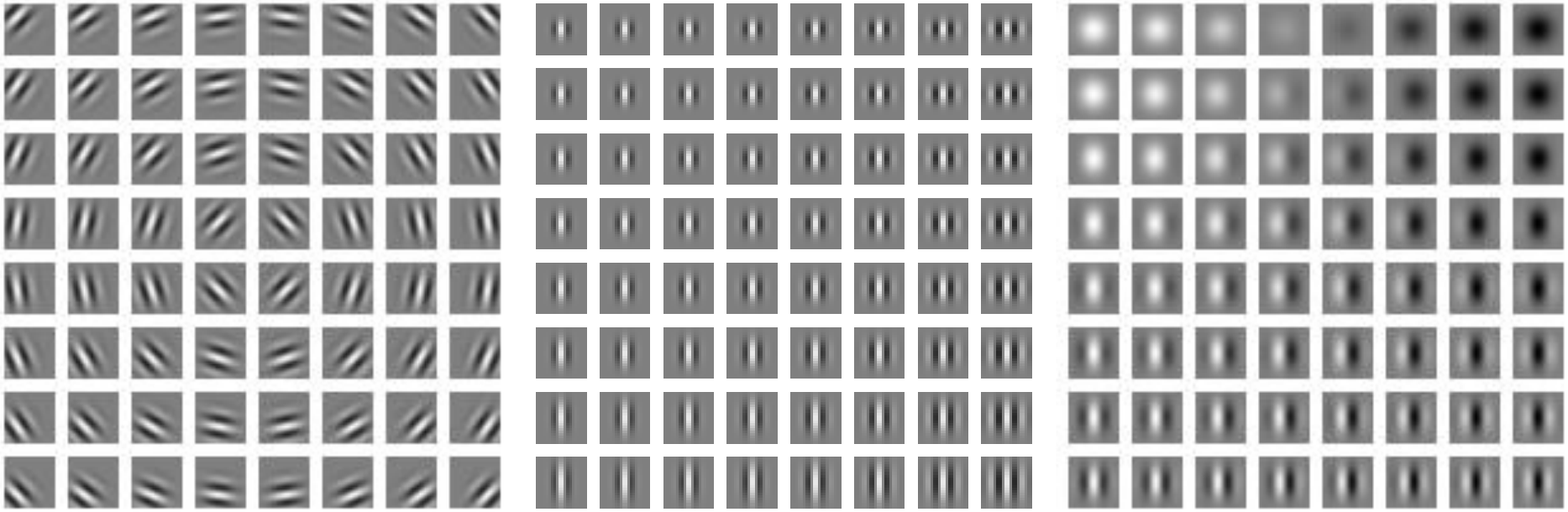


Convolution

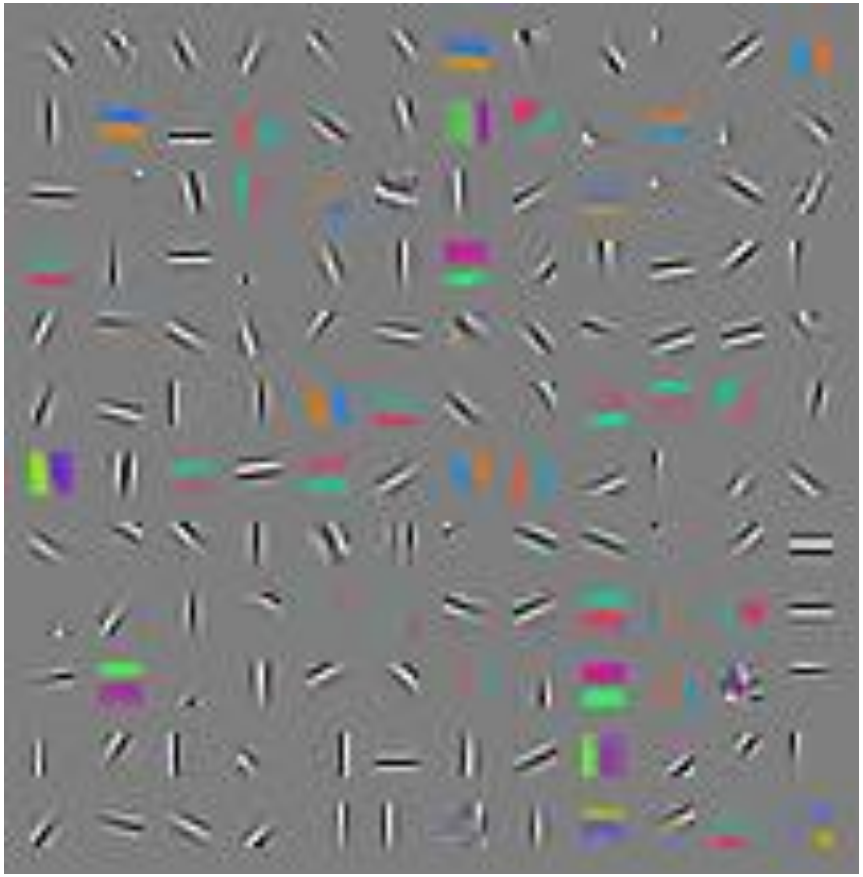


Fully connected

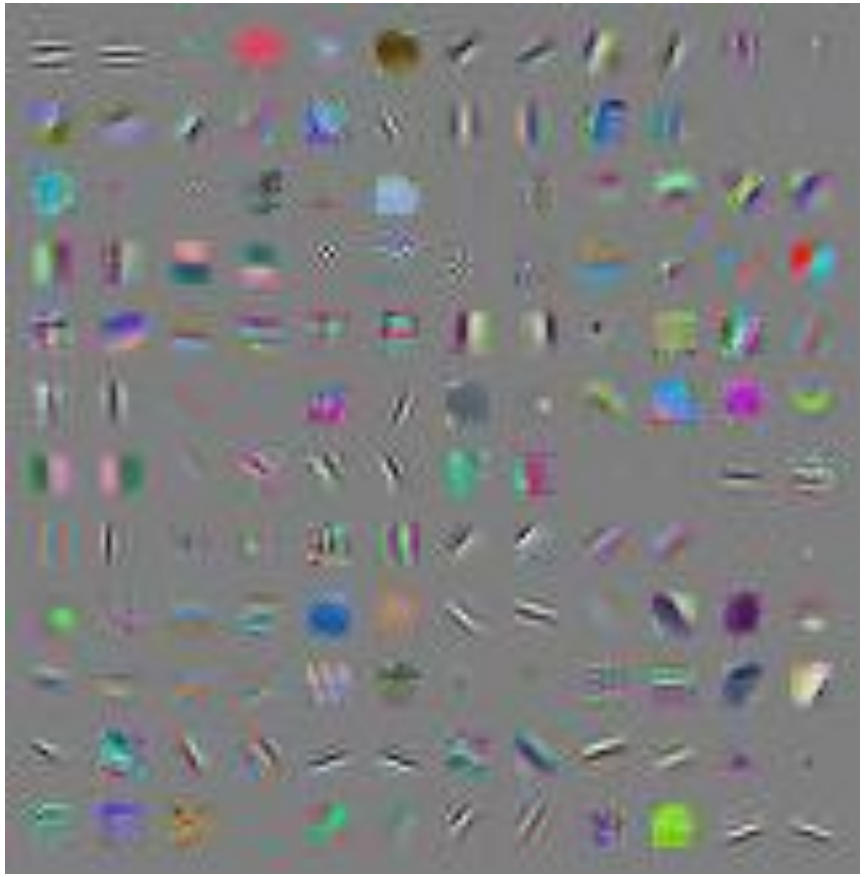
Gabor Functions



Gabor-like Learned Kernels



sparse coding



CNN

Major Architectures

- Spatial Transducer Net: input size scales with output size, all layers are convolutional
- All Convolutional Net: no pooling layers, just use strided convolution to shrink representation size
- Inception: complicated architecture designed to achieve high accuracy with low computational cost
- ResNet: blocks of layers with same spatial size, with each layer's output added to the same buffer that is repeatedly updated. Very many updates = very deep net, but without vanishing gradient.