Deep Learning

744.

Convolutional Neural Networks

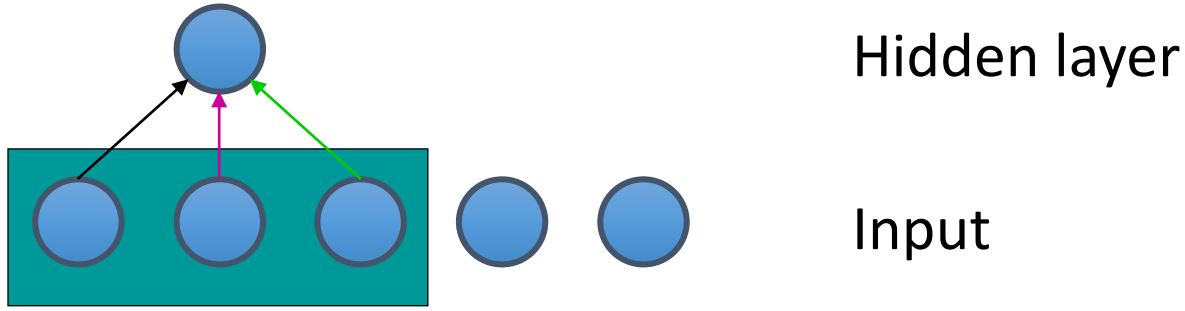
Cats’ Brains and CNNs

(Don’t do this to your cats. It can traumatize them.)

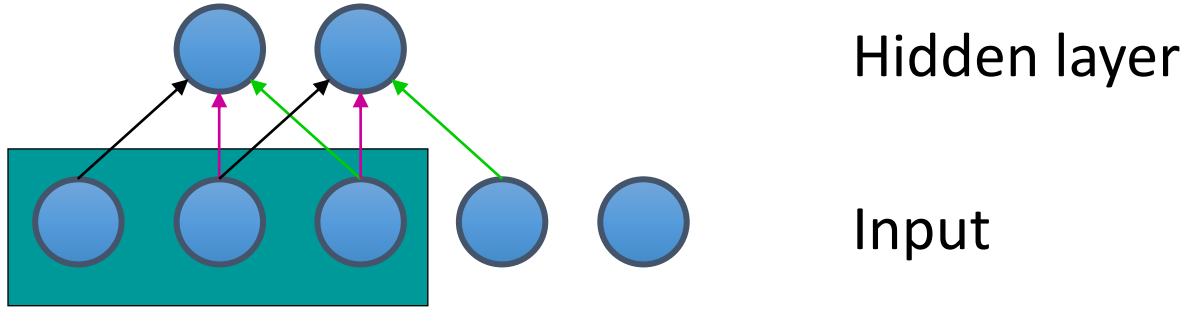
Short history of CNN

- Introduced by Hubel and Wiesel in the 50s and 60s
  - Neurons in the mammalian visual cortex respond to specific small patterns in the visual field
- Fukushima 1980
  - Propagating local features to higher layers
- LeCun 1989-
  - Training for handwriting recognition (MNIST) etc.
- Collobert 2011
  - use in NLP for semantic role labeling
- Krizhevsky et al. 2012
  - Object detection in ImageNet

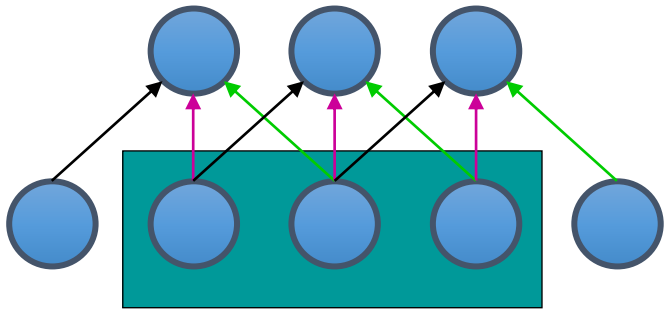
Basic Idea of CNNs



Basic Idea of CNNs



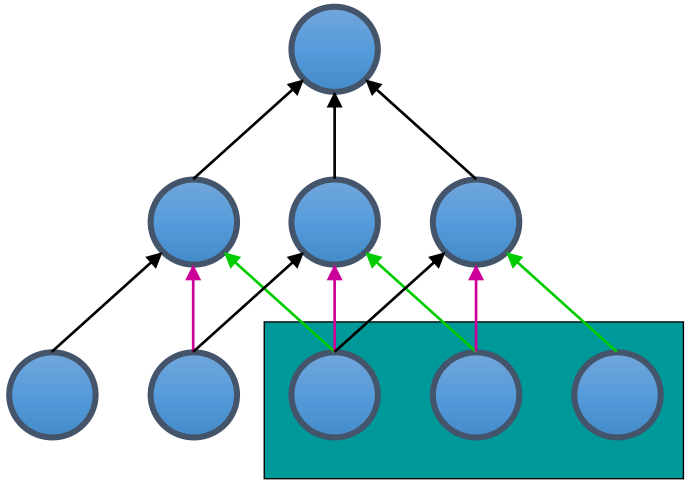
Basic Idea of CNNs



Hidden layer

Input

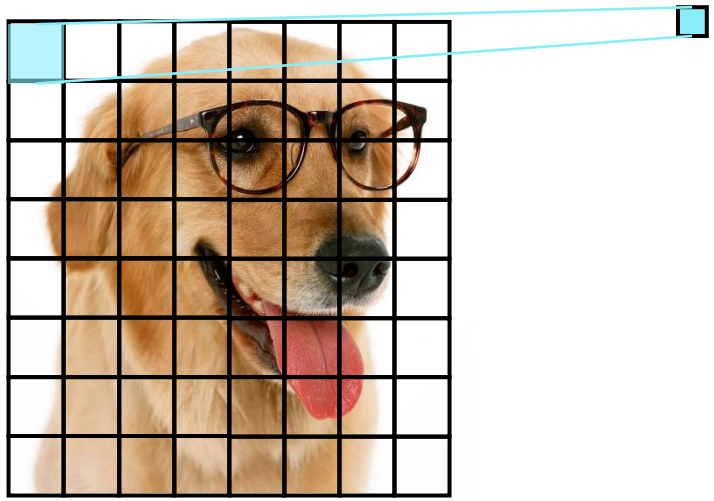
Basic Idea of CNNs



Hidden layer

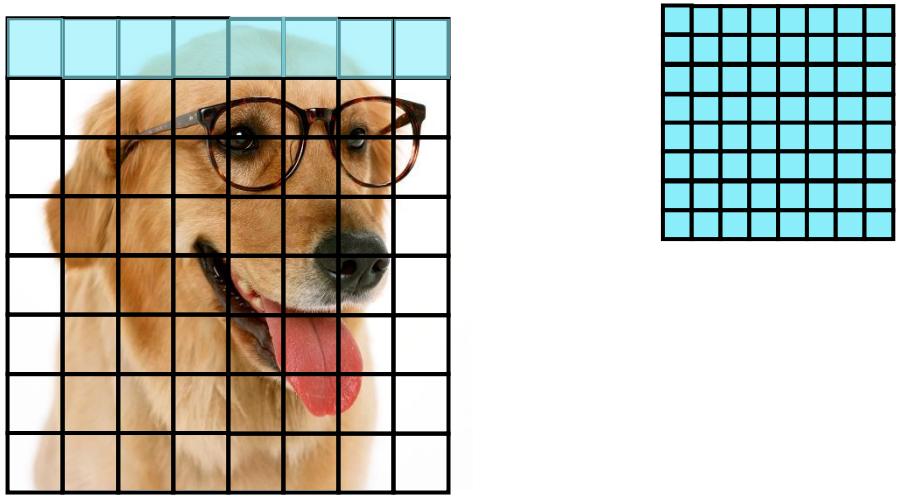
Input

CNN for Image Classification



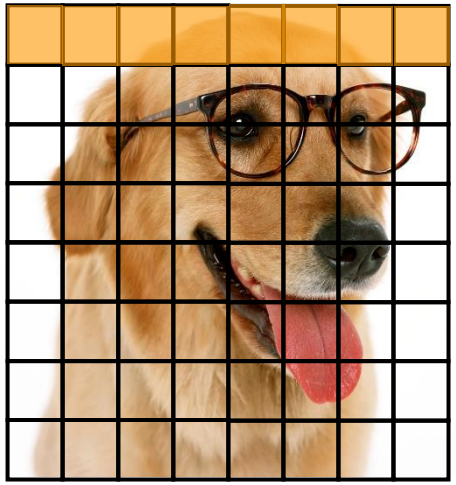
Convolutional Layer

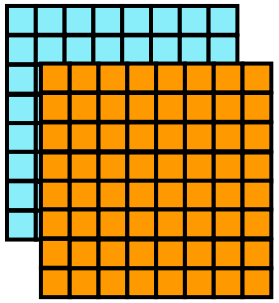
CNN for Image Classification



Convolutional Layer

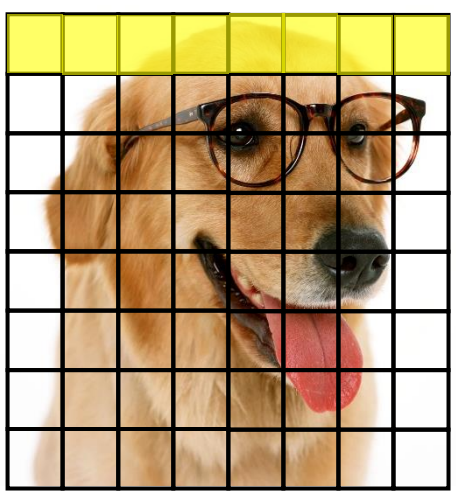
CNN for Image Classification

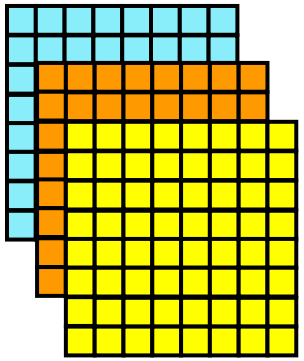




Convolutional Layer

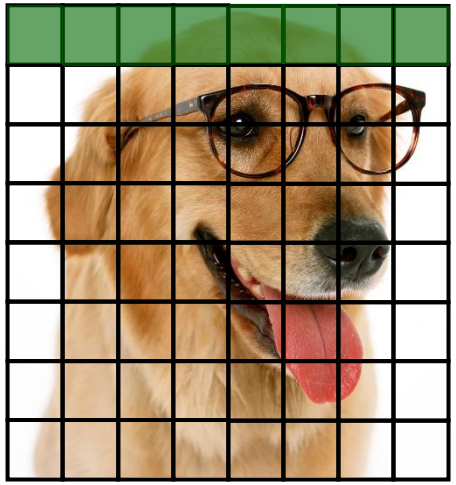
CNN for Image Classification

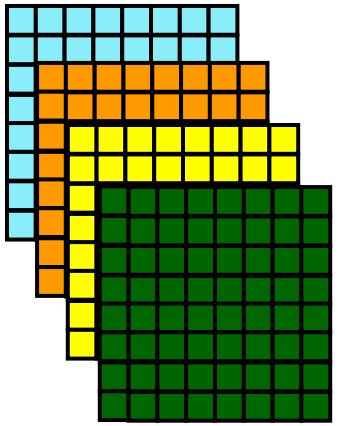




Convolutional Layer

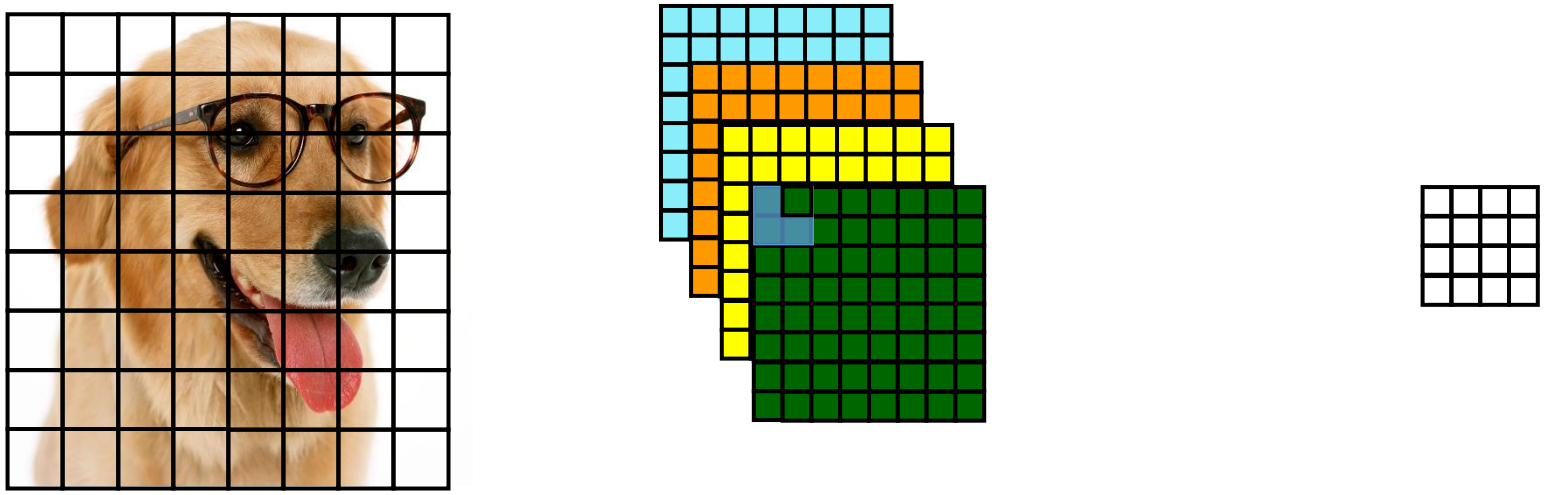
CNN for Image Classification





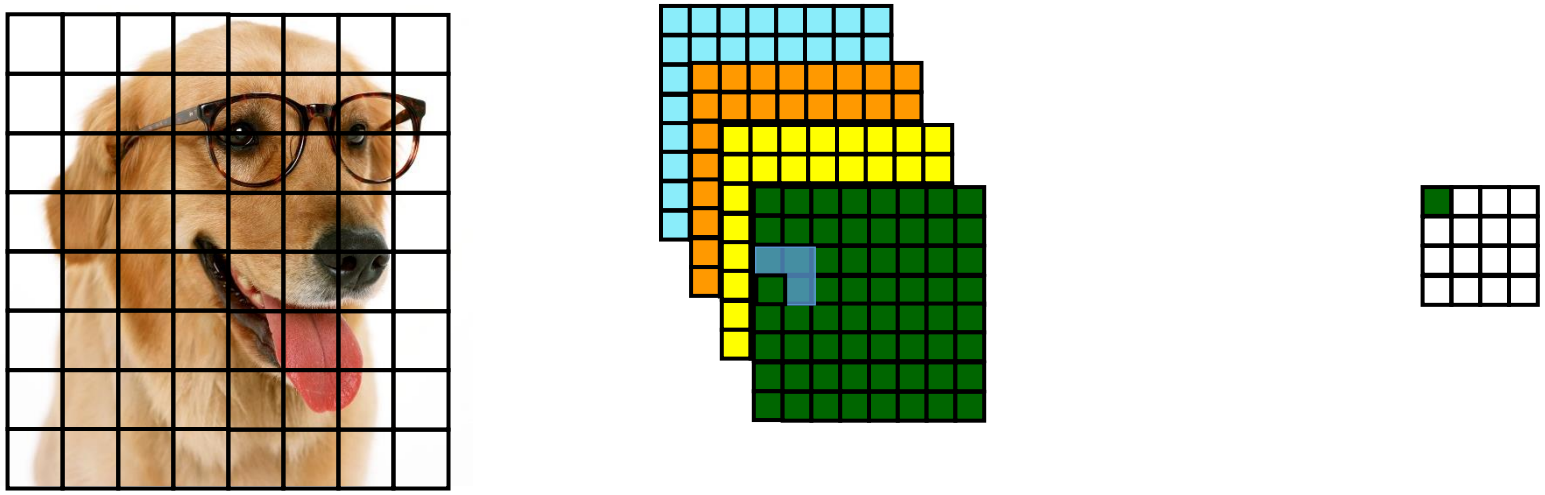
Convolutional Layer

CNN for Image Classification



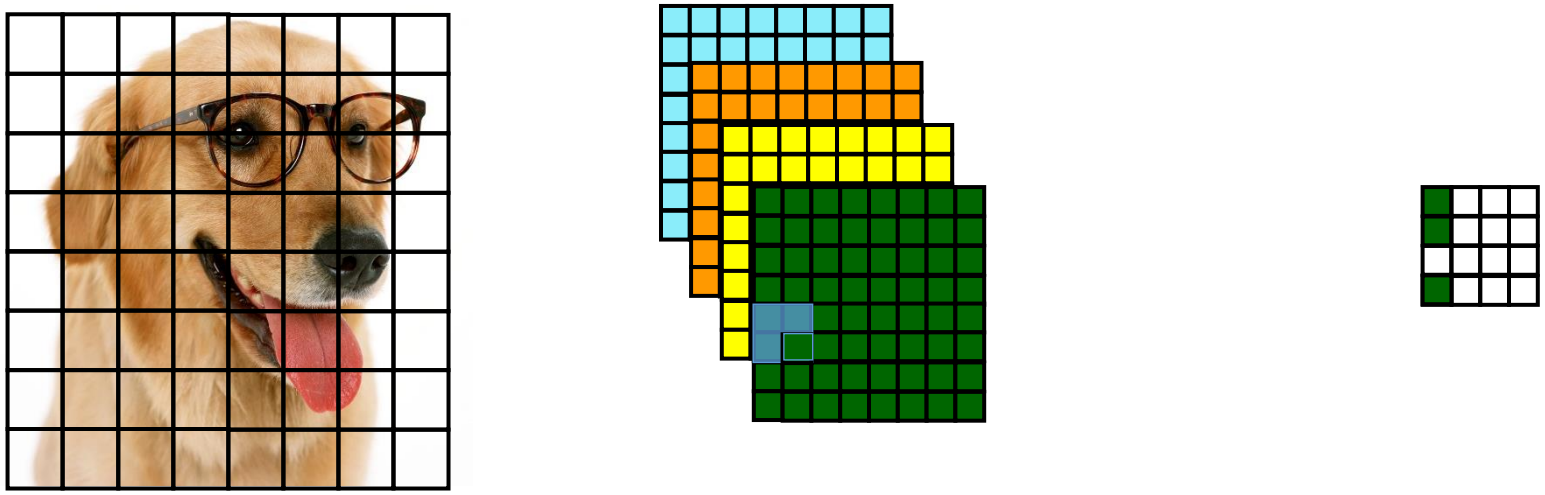
Max Pooling

CNN for Image Classification



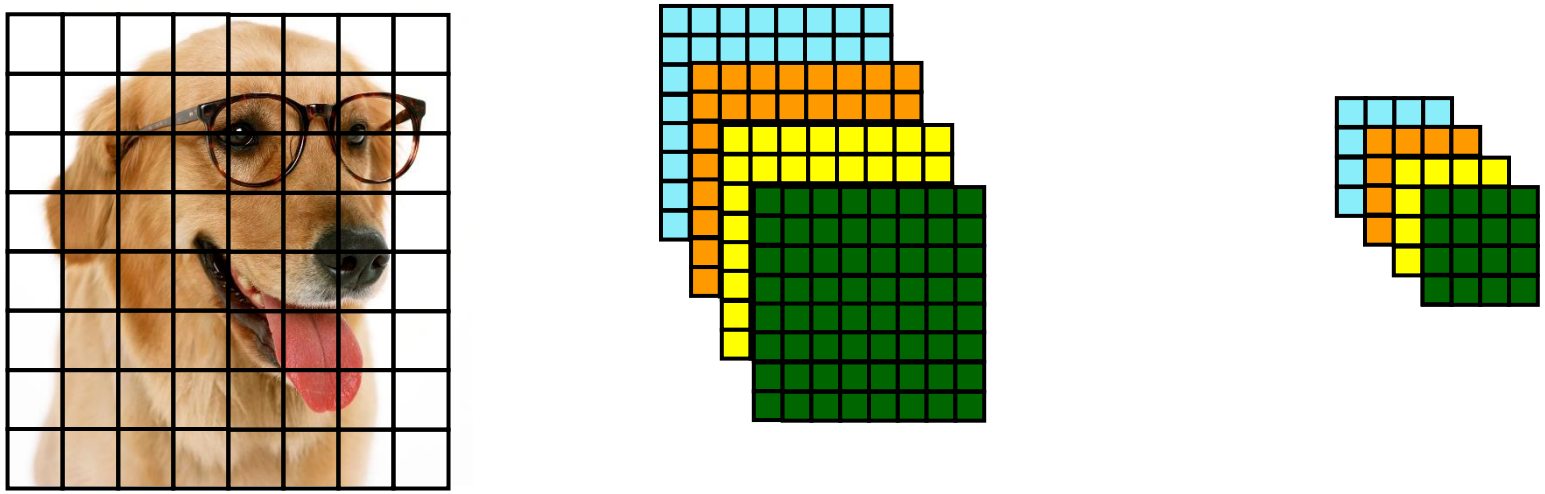
Max Pooling

CNN for Image Classification



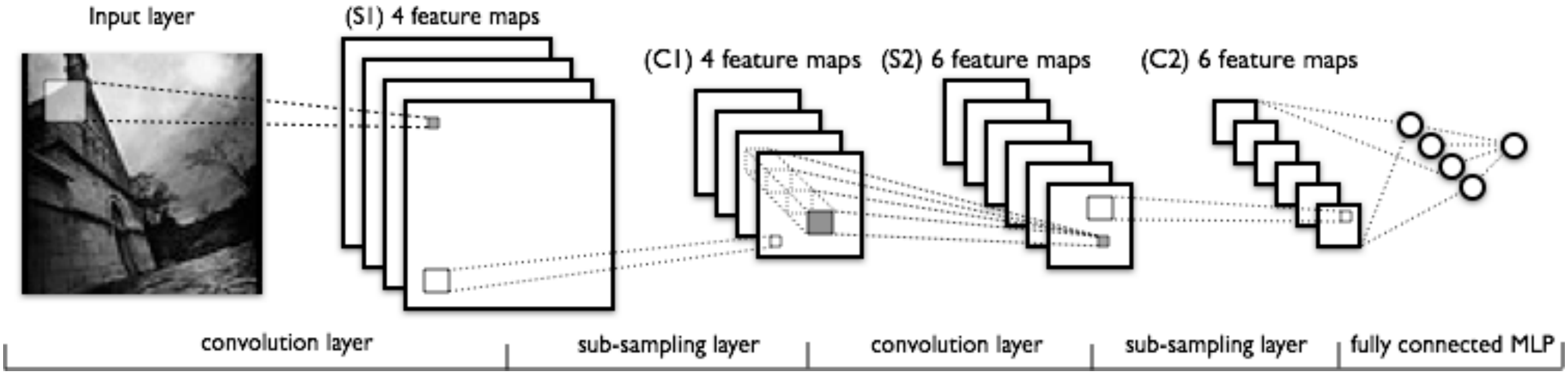
Max Pooling

CNN for Image Classification



Max Pooling

CNN for Image Classification



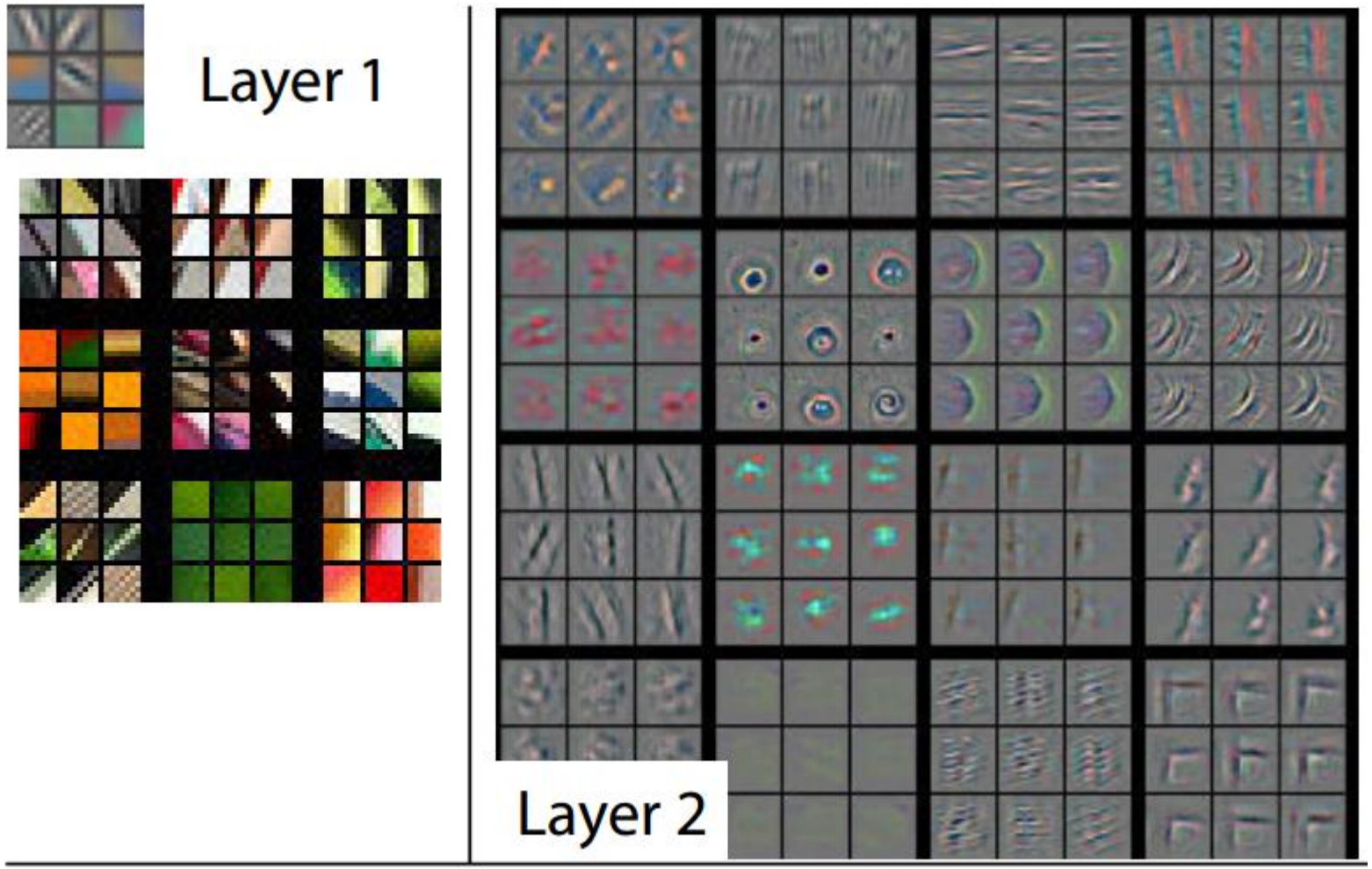
http://deeplearning.net/tutorial/lenet.html

[LeCun]

CNN for Image Classification

- How good are CNNs?
- “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.”
  - Krizhevsky et al., 2012: ImageNet Classification with Deep Convolutional Neural Networks
  - Competition to classify photos from ImageNet, http://www.image-net.org/

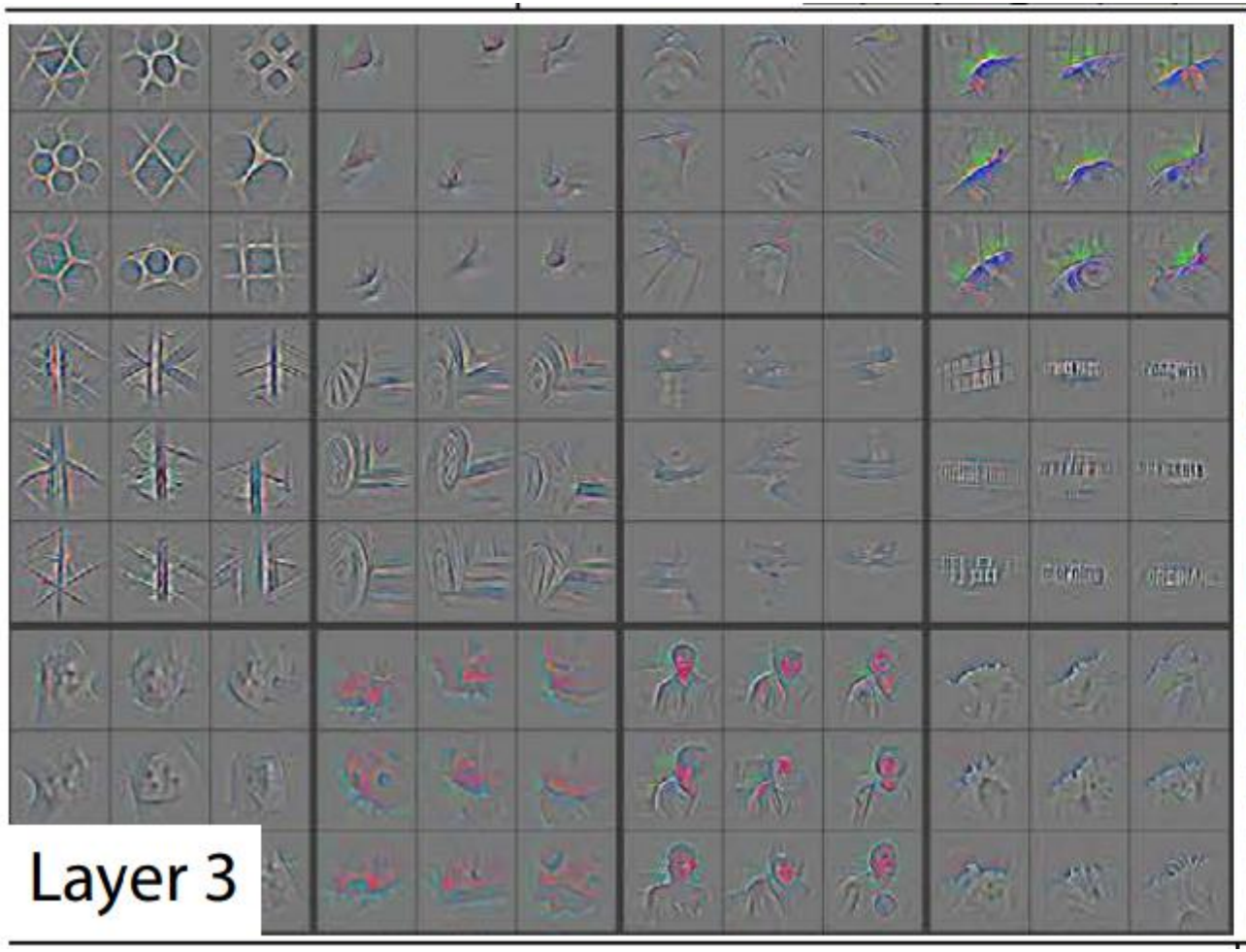
CNN for Image Classification



Learned Filters

Zeiler and Fergus 2014

CNN for Image Classification



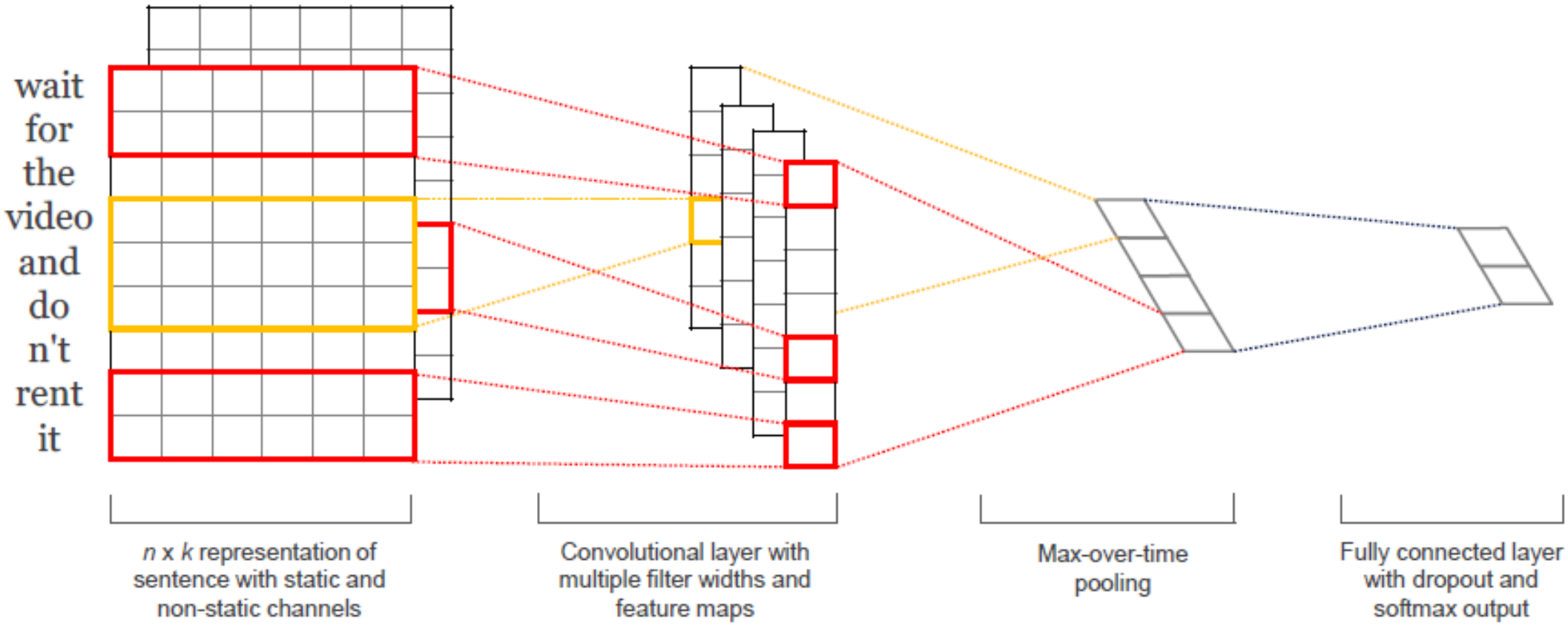
Zeiler and Fergus 2014

CNN for Image Classification

Layer 4

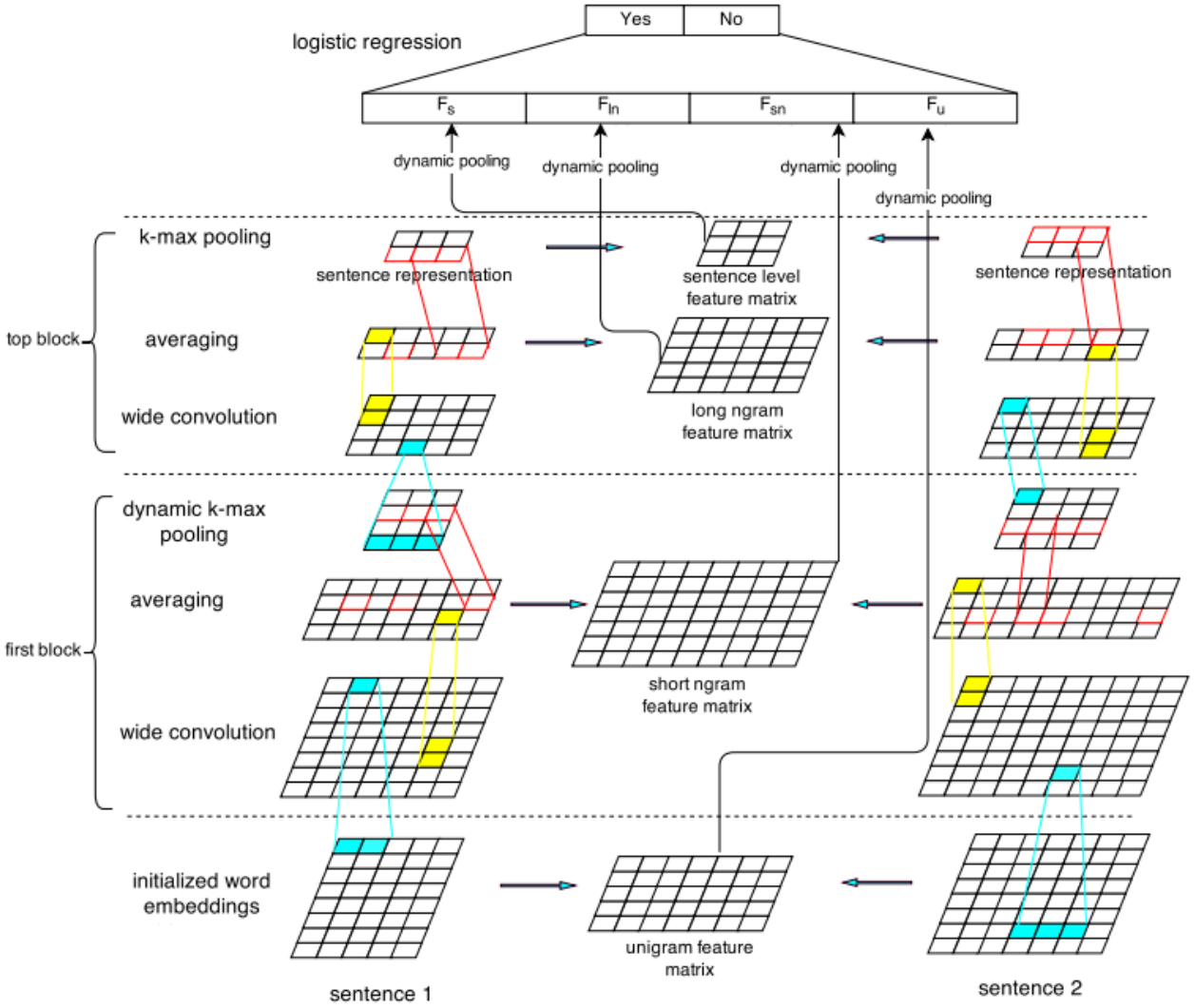
Zeiler and Fergus 2014

Sentence Classification



[Kim 2014]

Paraphrase Detection



[Yin & Schütze 2015]

Adjustments to CNN

- Striding
  - Skip some of the possible input substrings
  - E.g., start at every other word
- Pooling
  - Reduction, e.g., average, max
  - k-max: did this feature appear at least k times
- Stacking
  - Same idea as with RNN and LSTM
- Dilating
  - (see next slide)

Dilated Convolution

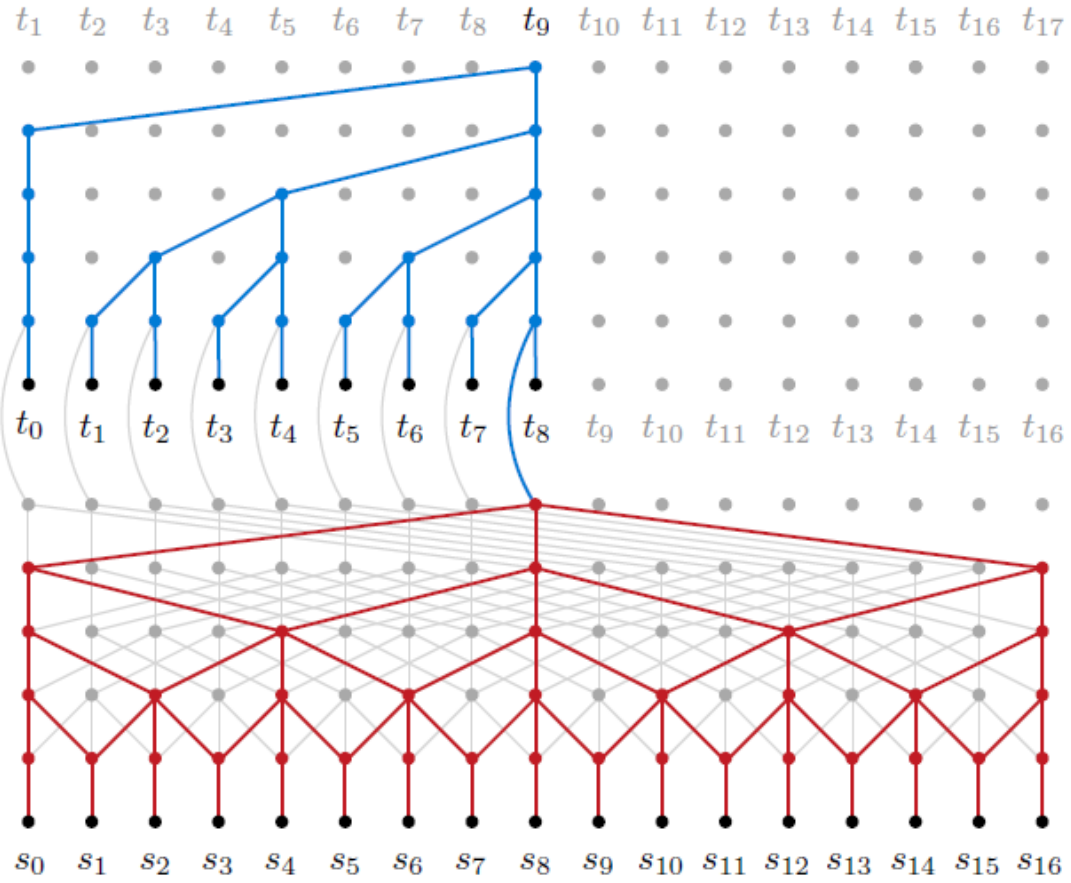


Figure 1. The architecture of the ByteNet. The target decoder (blue) is stacked on top of the source encoder (red). The decoder generates the variable-length target sequence using dynamic unfolding.

[Kalchbrenner et al. 2016]

Using CNNs for NLP

- Convolutional Neural Network for Paraphrase Identification (Yin & Schütze 2015)
- Summarization-based Video Caption via Deep Neural Networks (Li et al. 2015)
- Question Answering over Freebase with Multi-Column Convolutional Neural Networks (Dong et al. 2015)
- Convolutional Neural Network Architectures for Matching Natural Language Sentences (Hu et al. 2015)
- Learning Semantic Representations Using Convolutional Neural Networks for Web Search (Sheng et al. 2015)
- Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts (dos Santos & Gatti 2014)
- Relation Extraction: Perspective from Convolutional Neural Networks (Nguyen & Grishman 2015)
- Modeling Mention, Context and Entity with Neural Networks for Entity Disambiguation (Sun et al. 2015)
- Modeling Interestingness with Deep Neural Networks (Gao 2015)