Deep Learning for NLP - Focus on Medical Applications

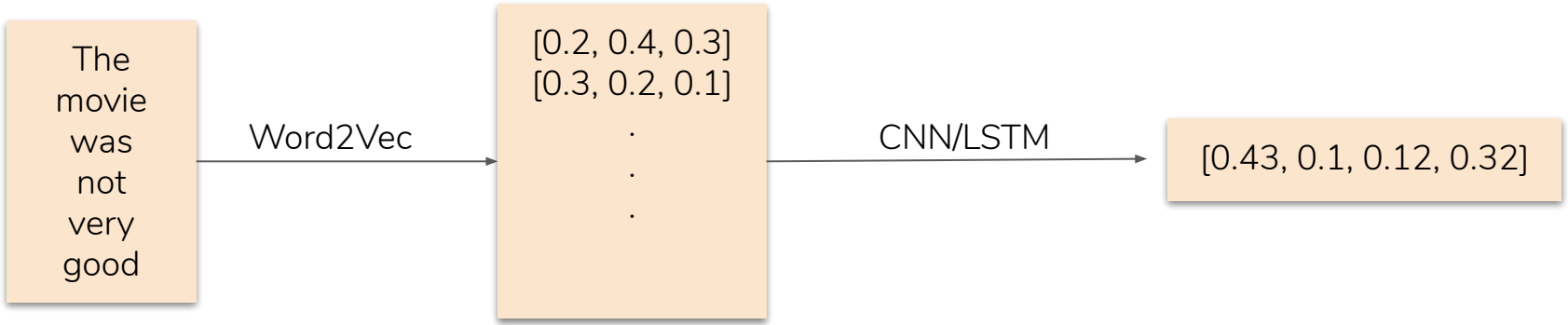
Representation Learning and Convolutional Networks

1. What did we do wrong in the Tweets Classification Task

- Fully Connected Networks
  - Linear layers
- Word2Vec
  - Learns word representations
- Classifying words
- Sentence representations
  - Averaging is not the best solution
    - Significant information loss
  - Works when order of words is not important
    - Disease mention

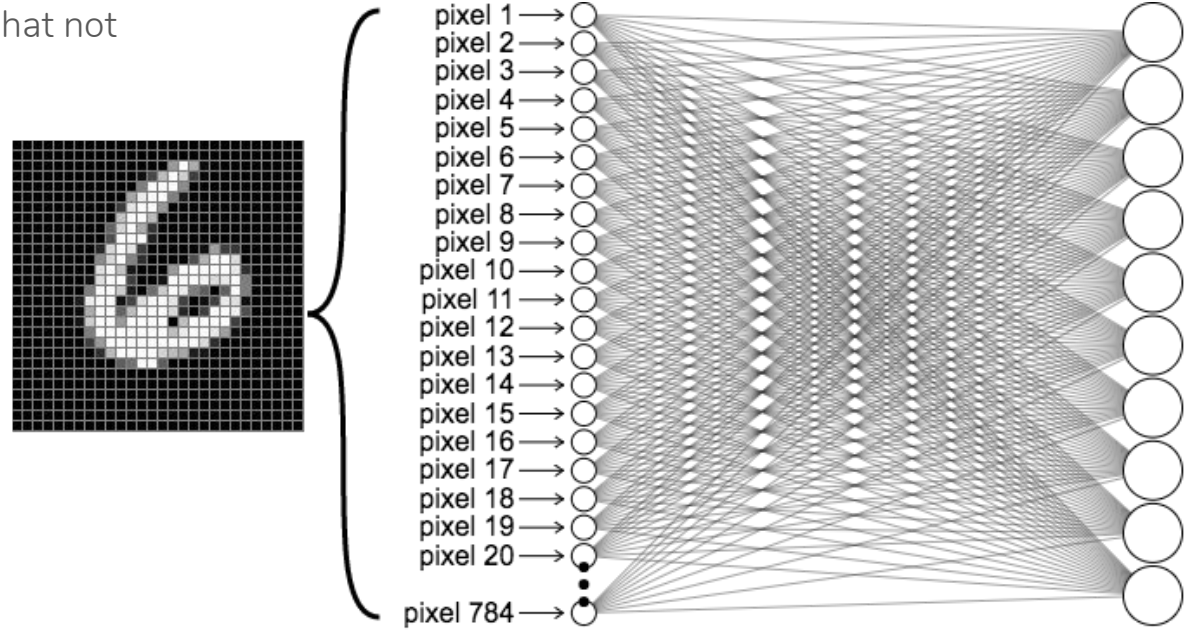
2. Representation learning

- Language Modeling
- Words to sentences
- Images to vectors



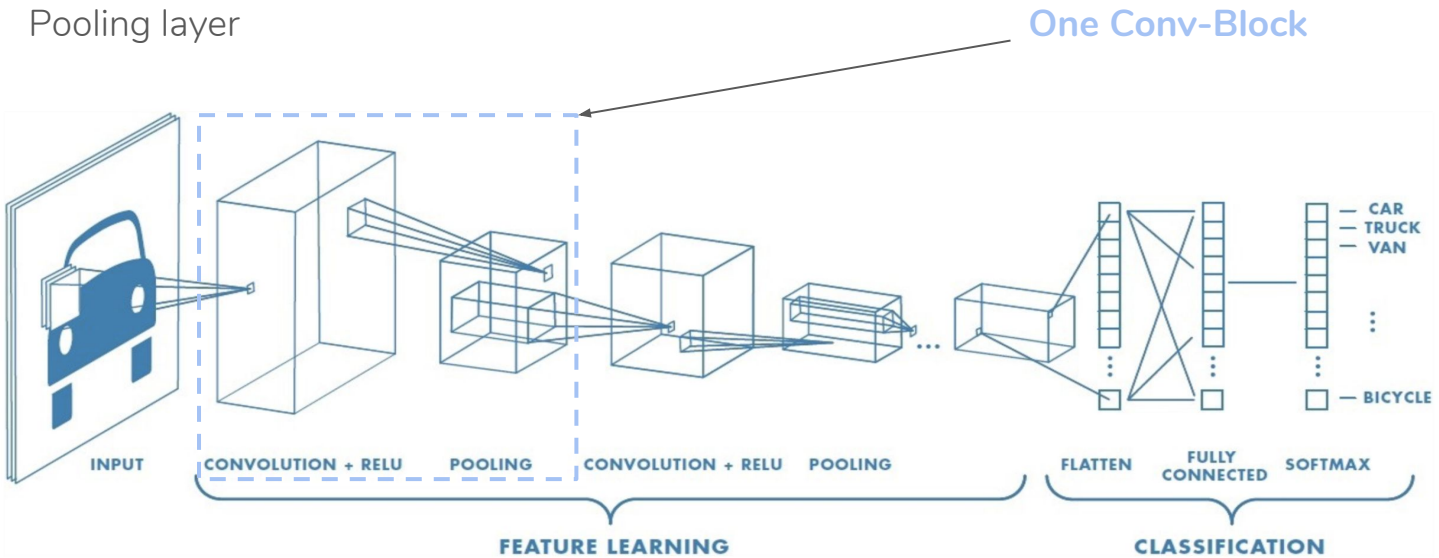
2. Representation Learning - Images

- The network has to learn everything
  - Background
  - Position
  - What is important what not
  - Colors/Shades



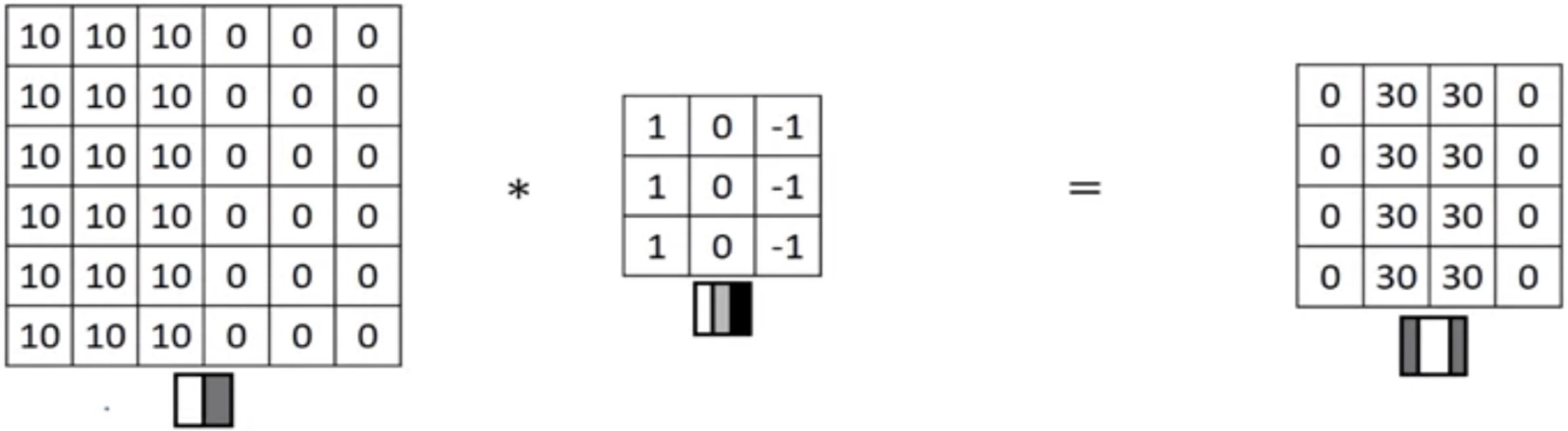
3. Convolutional Networks

- We need something a bit biased towards detecting features in images
  - Eges, Shapes, Parts
- One conv-block consists of:
  - Convolution + Activation = Conv Layer
  - Pooling layer



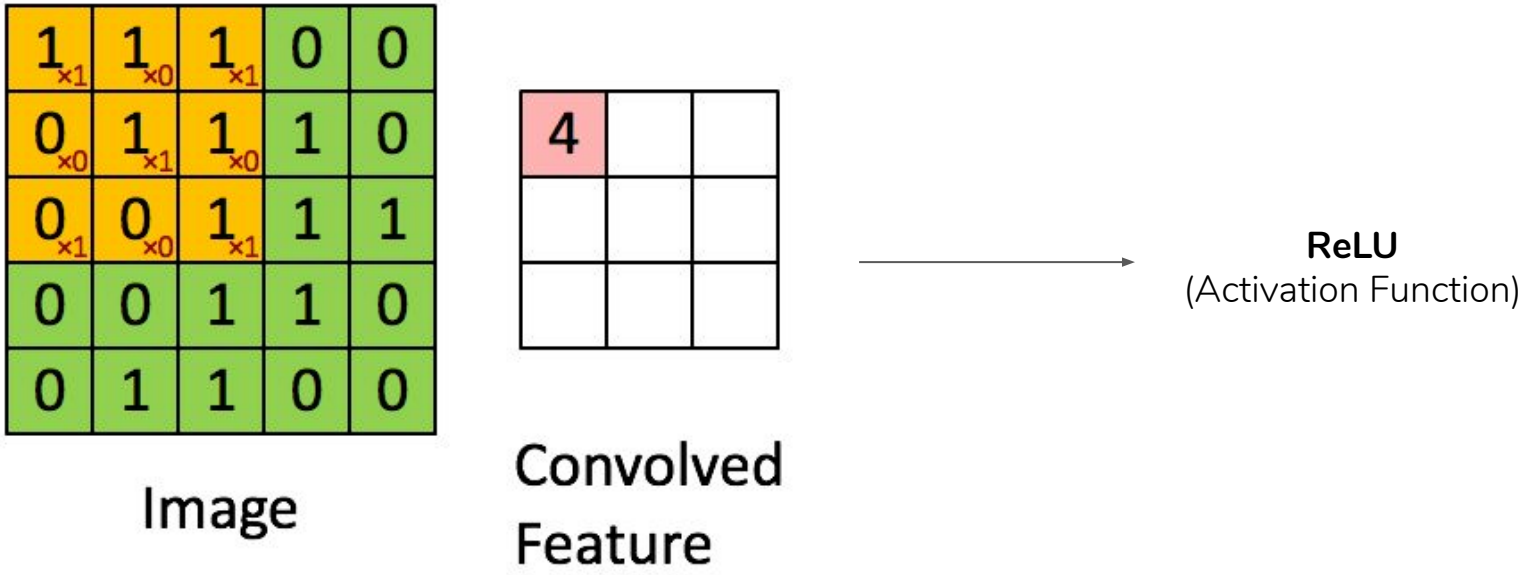
3. Convolutional Networks - filters/kernels

- We are learning the weights of the filters/kernels
- Detecting vertical edges *(see colab on edge detection)*



**Example taken from* the Deep Learning course by Andrew Ng.

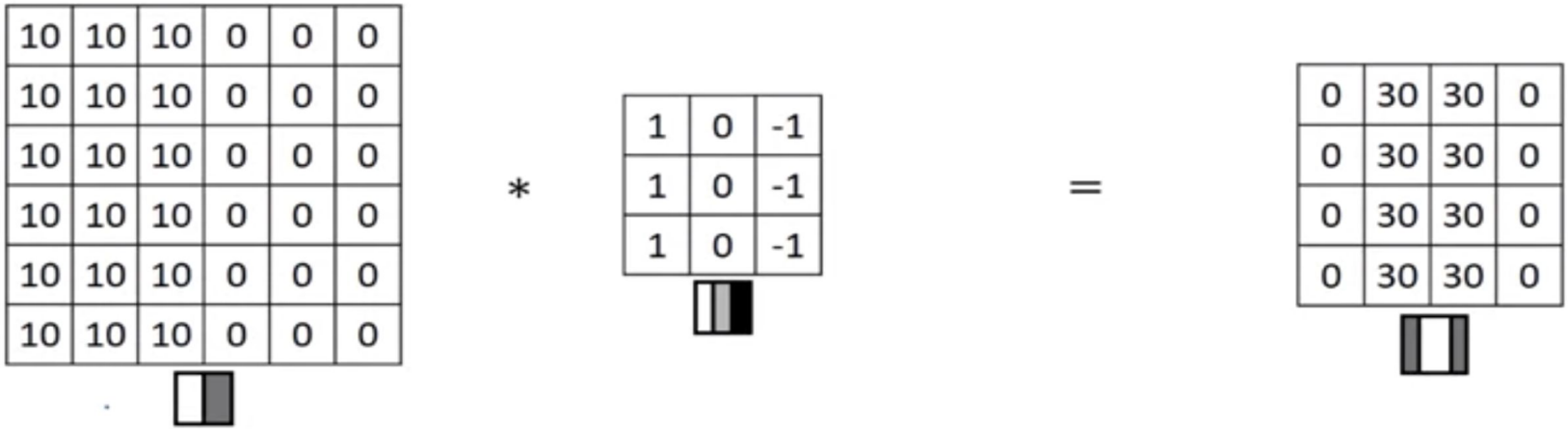
3. Convolutional Networks - filters/kernels



Example taken from the Deep Learning course by Andrew Ng.

3. Convolutional Networks - filters/kernels

- We are learning the weights of the filters/kernels
- Detecting vertical edges

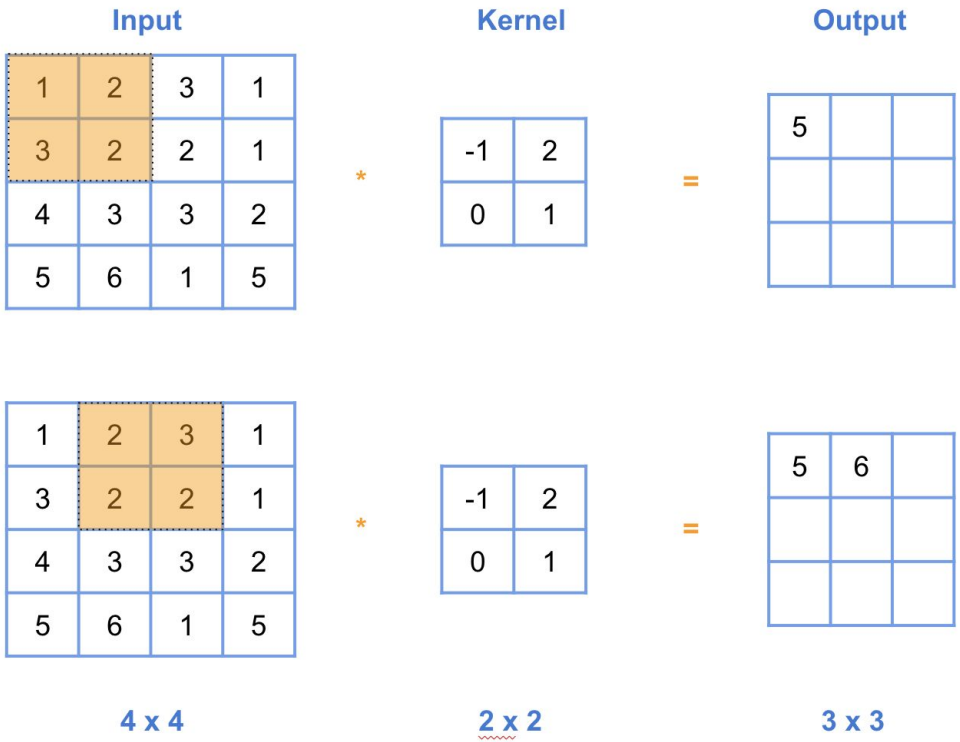


6 x 6 3 x 3 4 x 4

Example taken from the Deep Learning course by Andrew Ng.

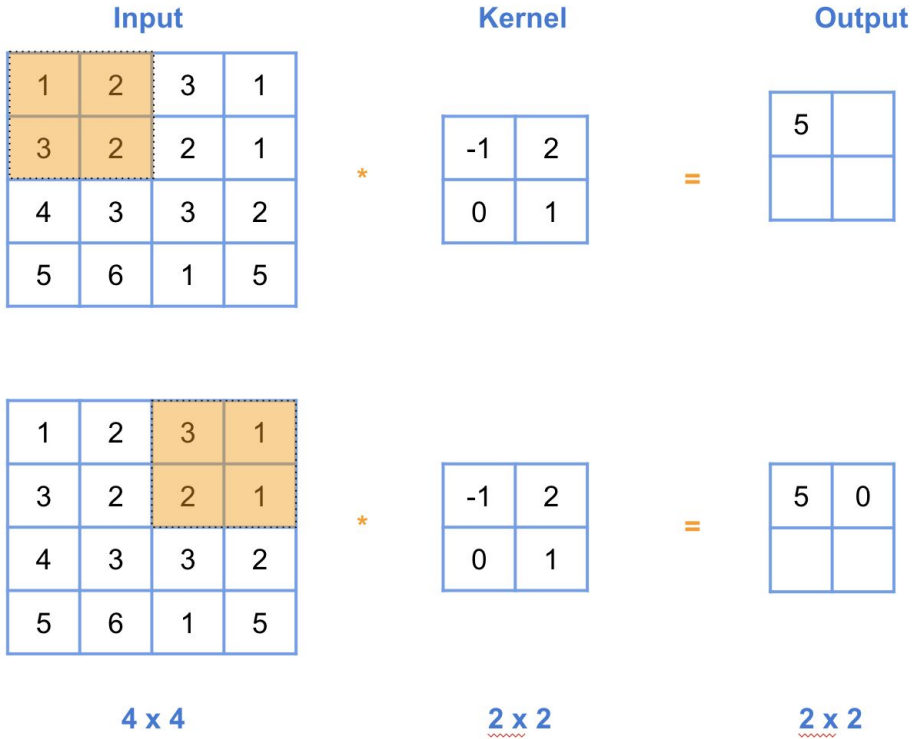
3. Convolutional Networks - filters/kernels

- **stride = 1**
- padding
- dilation



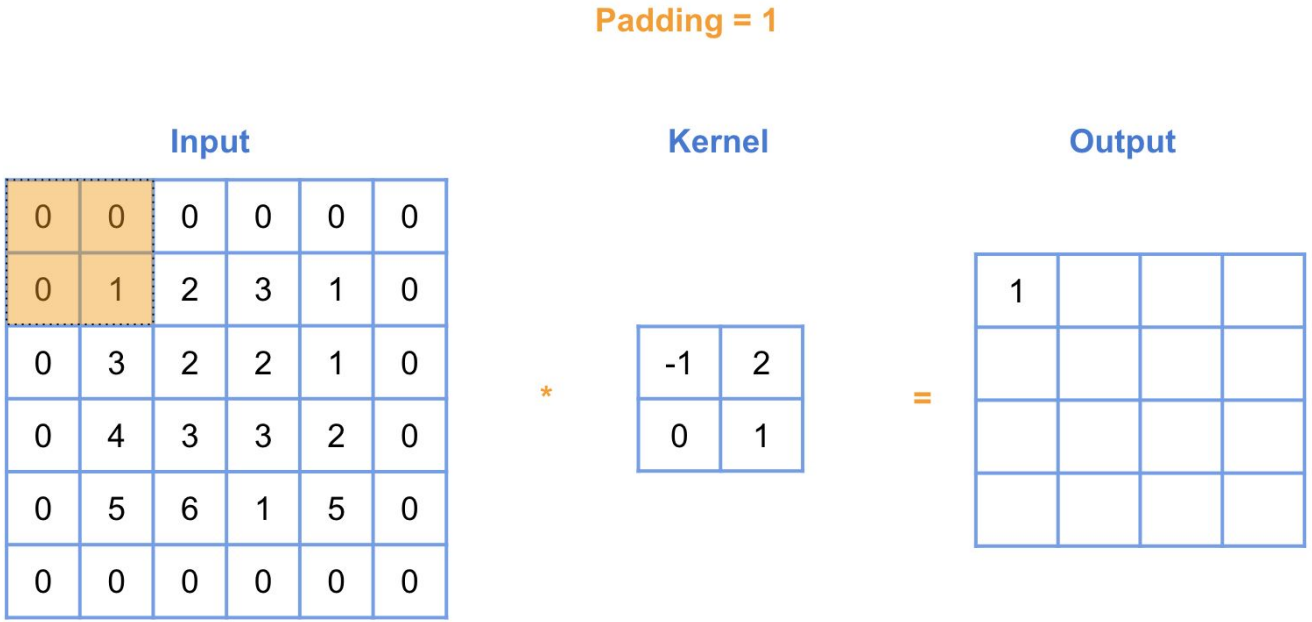
3. Convolutional Networks - filters/kernels

- **stride = 2**
- padding
- dilation



3. Convolutional Networks - filters/kernels

- stride
- **padding**
- dilation

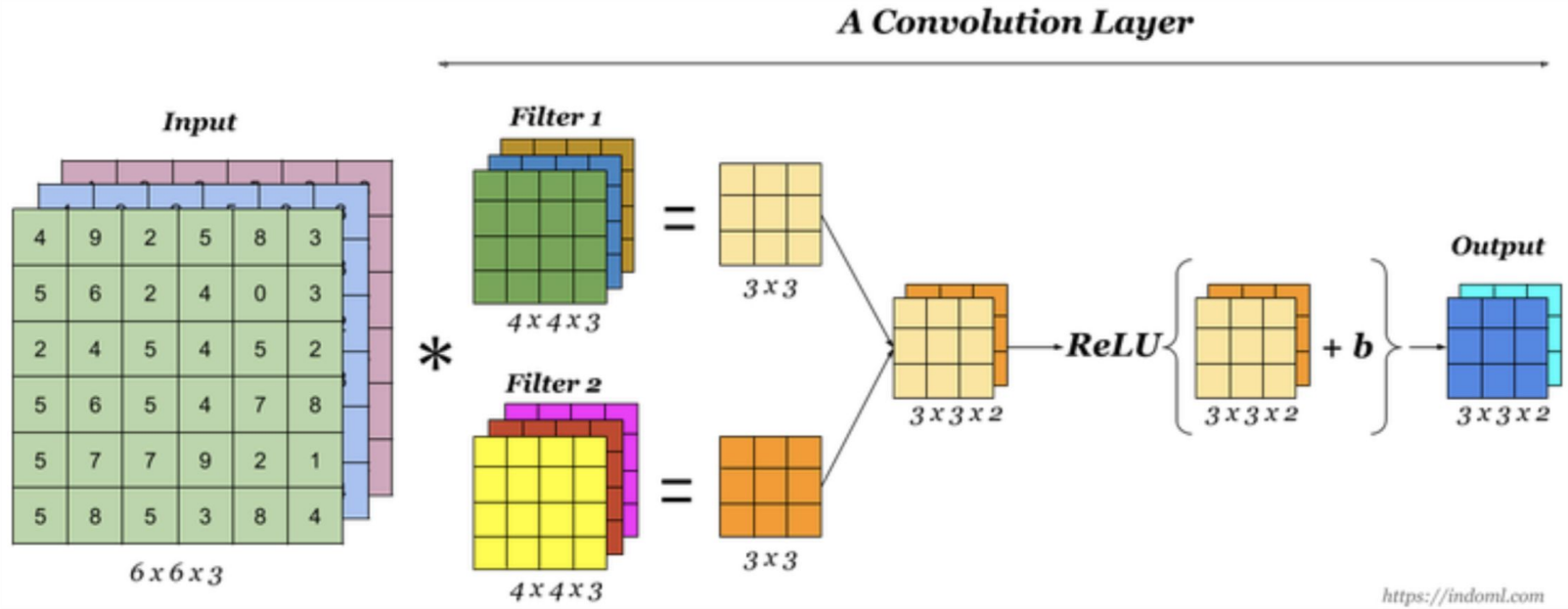


3. Convolutional Networks - filters/kernels

- stride (usually 1)
- padding (usually no padding)
- dilation (not easy to explain, but skipping pixels)

*dilation: https://github.com/vdumoulin/conv_arithmetic/blob/master/README.md

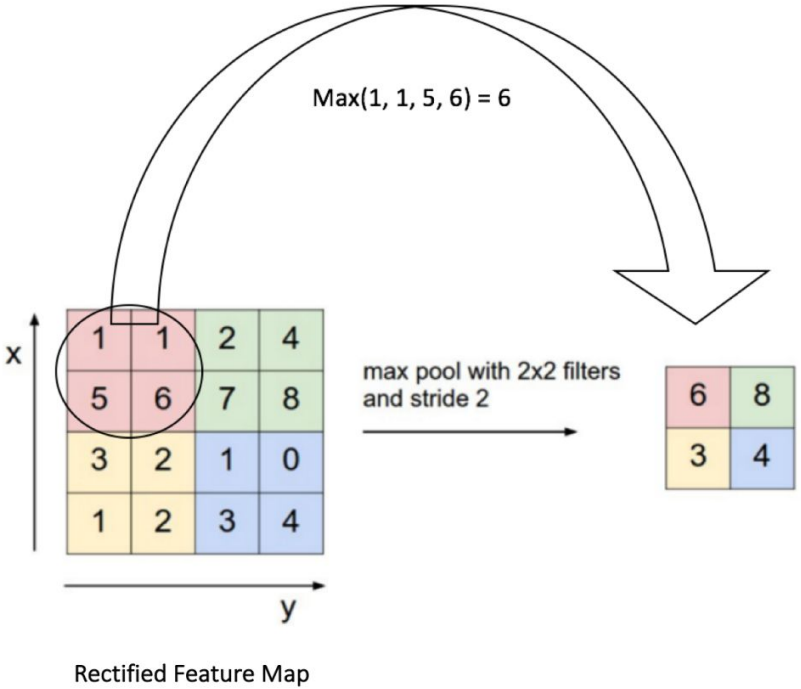
3. Convolutional Networks - filters/kernels



Convolution layer

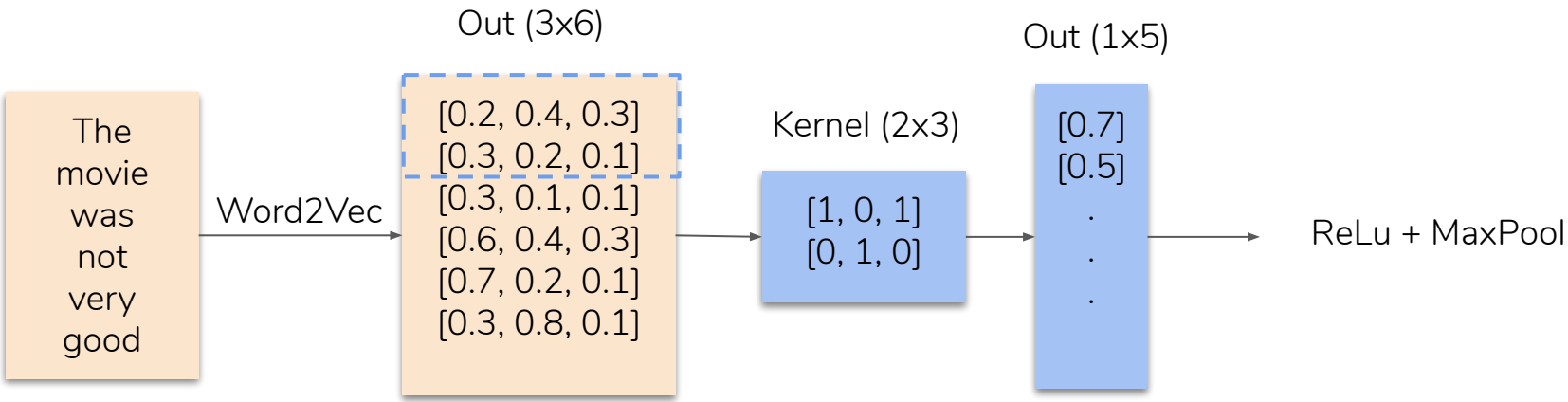
* http://scs.ryerson.ca/~aharley/vis/conv/

3. Convolutional Networks - Max pooling

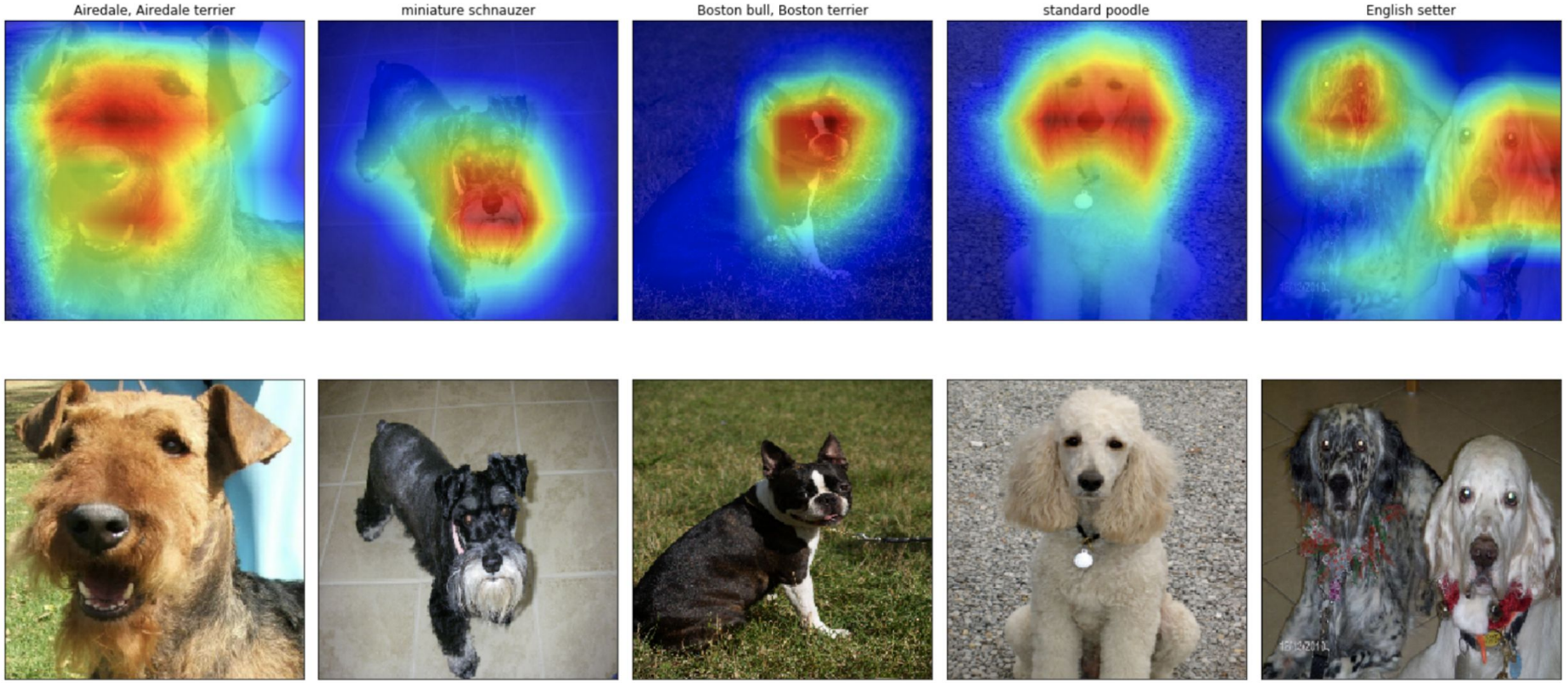


Example taken from the Deep Learning course by Andrew Ng.

4. ConvNets applied to Text



4. ConvNets Interpretability - Class Activation Map



4. ConvNets - Problems

- Not made for text classification
- Very good for finding short patterns
- No long dependencies, or sequential reading of text
- Loss of semantics

Summary

- What is Representation Learning
- How are fully connected networks used
- What did we do wrong
- Convolutional Networks (CNNs)
  - How do they work
  - Training a CNN
  - Advantages
  - Problems
  - Interpretability