Convolutional Neural Networks

Jeffrey Kelling

23rd September 2018

Convolutional Neural Networks

- overcome problems of MLPs:
  - too many connections required for complex input data
  - overfitting

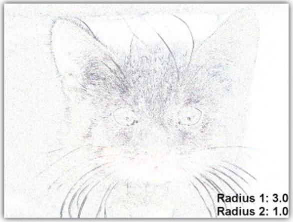
  $\Rightarrow$ very hard or impossible to train
- introduce specialized layers,
  forcing the network to form a specific hierarchy
  - in principle, MLPs could learn the same
  - ... just narrowing the search space

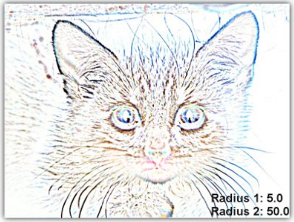
Convolutional Neural Networks II

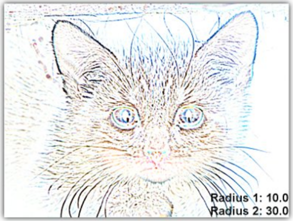
- layers inspired by steps classical approaches use, for images:
  - convolution
    Keras Layers: `Conv2D`; `Conv1D`; `Conv2DTranspose`; ...
  - pooling (max, mean, ...)
    Keras Layers: `MaxPooling2D`; `AveragePooling2D`; ...
- convolutions can also downample when strided

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

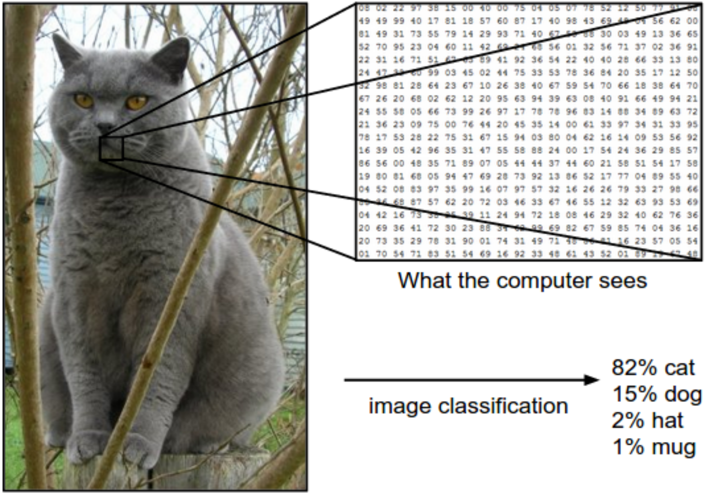
Convolution Filter





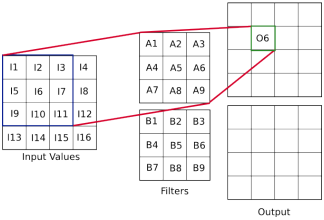


Cat vs Pixels



Convolution I

- Multiply sliding window of input with small filter to produce output activations

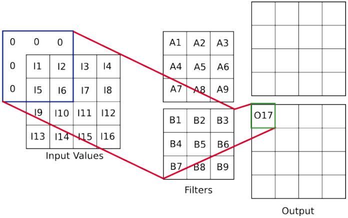


$$
\begin{aligned}
O_6 = {} & A_1 I_1 + A_2 I_2 + A_3 I_3 \\
& + A_4 I_5 + A_5 I_6 + A_6 I_7 \\
& + A_7 I_9 + A_8 I_{10} + A_9 I_{11}
\end{aligned}
$$

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Convolution I

- Multiply sliding window of input with small filter to produce output activations
- Multiple convolution filters lead to multiple activation maps
- Borders may be padded or skipped

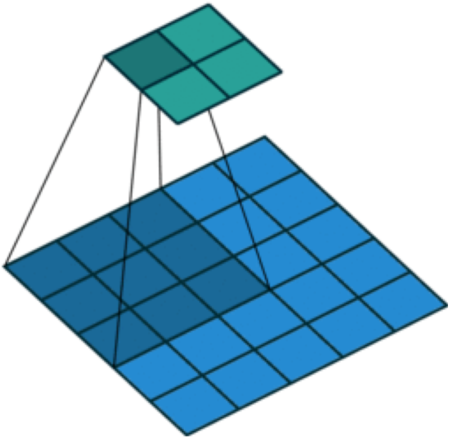


$$O_{17} = B_5 I_1 + B_6 I_2 \\ + B_8 I_5 + B_9 I_6$$

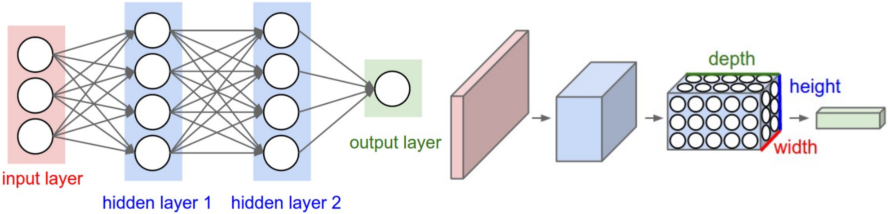
Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Convolution II

- Filters are usually small: 3×3 or 5×5
- Stride > 1 leads to downsampling



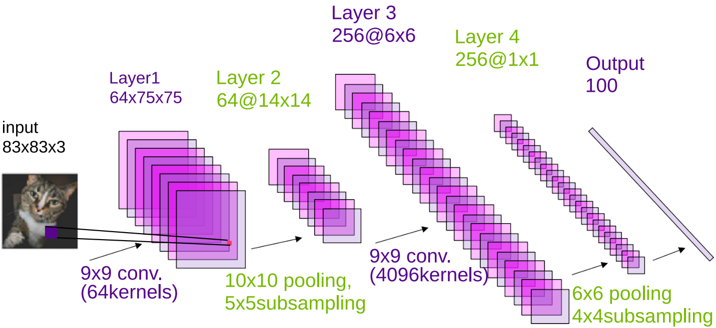
Tensor Representation of Layers



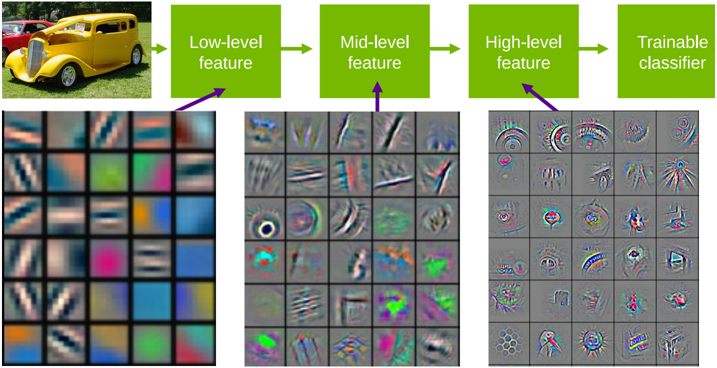
- convolution and pooling layers assign higher-dimensional topology to layers:
  - (x, y) image dimensions $+$ color channels $=$ 3 dimensions
- convolutions sum over all input channels (or excess dimensions)

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Convolutional Neural Networks III



Hierarchical Representations



Feature visualization of convolutional net trained on ImageNet[1]

[1] Zeiler & Fergus 2013

Feature Hierarchies

- DNNs learn a hierarchy of representations
- stages of trainable feature transforms ...
- Image recognition:
  pixel $\rightarrow$ edge $\rightarrow$ texton $\rightarrow$ motif $\rightarrow$ part $\rightarrow$ object
- Text:
  character $\rightarrow$ word $\rightarrow$ word group $\rightarrow$ clause $\rightarrow$ sentence $\rightarrow$ story
- Speech:
  sample $\rightarrow$ spectral band $\rightarrow$ sound $\rightarrow$ ... $\rightarrow$ phoneme $\rightarrow$ word

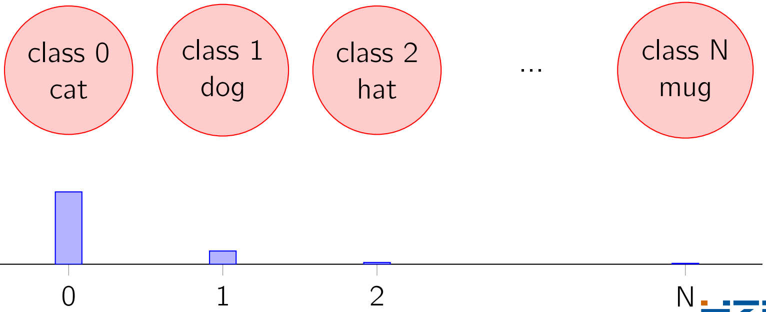
Exercise

`day2/notebooks/convnets_cifar10`

One-Hot Encoding

- Neural network classification results usually use **one-hot encoding**
- One output neuron per class

↓ inference



Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de