

# Convolutional Neural Networks

Jeffrey Kelling

23rd September 2018

**hzdr**

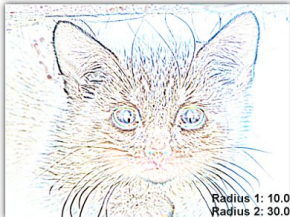
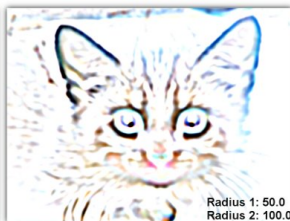
 HELMHOLTZ  
ZENTRUM DRESDEN  
ROSSENDORF

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

⇒ 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

- 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 downsample when strided

# Convolution Filter



# Cat vs Pixels



08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	81	78
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	17	04	56	62	00
81	49	31	73	55	79	14	29	93	71	40	67	52	85	30	03	49	13	36	65
52	70	95	23	04	60	11	42	83	51	88	56	01	32	36	71	37	02	36	91
22	31	16	71	51	63	43	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	33	09	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
31	98	81	28	64	23	67	10	26	38	40	67	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	63	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	63	72
21	36	23	09	75	00	76	44	20	45	35	14	00	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	36	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	68	05	94	47	69	28	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	98	66
05	16	48	87	57	62	20	72	03	46	35	67	46	55	12	32	63	93	53	69
04	42	16	73	35	85	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	69	36	41	72	30	23	88	04	56	89	69	82	67	59	85	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	14	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	61	43	52	01	89	37	67	48

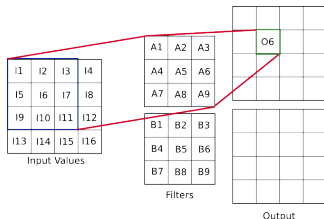
What the computer sees

image classification →

- 82% cat
- 15% dog
- 2% hat
- 1% mug

# Convolution I

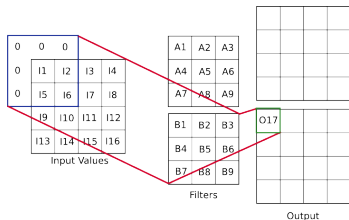
- Multiply sliding window of input with small filter to produce output activations



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

# Convolution I

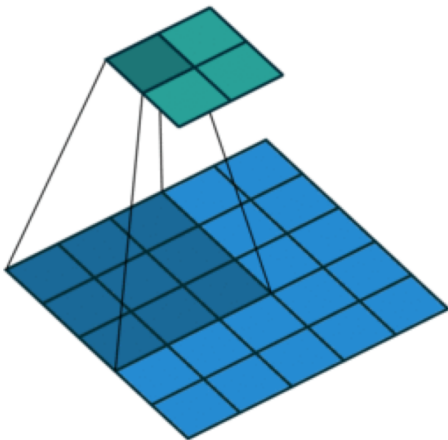
- Multiply sliding window of input with small filter to produce output activations
- Multiple convolution filters lead to multiple activation maps
- Borders may be padded or skipped



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

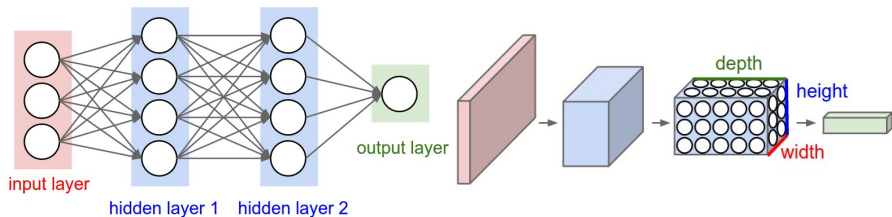
# Convolution II

- Filters are usually small:  $3 \times 3$  or  $5 \times 5$
- Stride  $> 1$  leads to downsampling



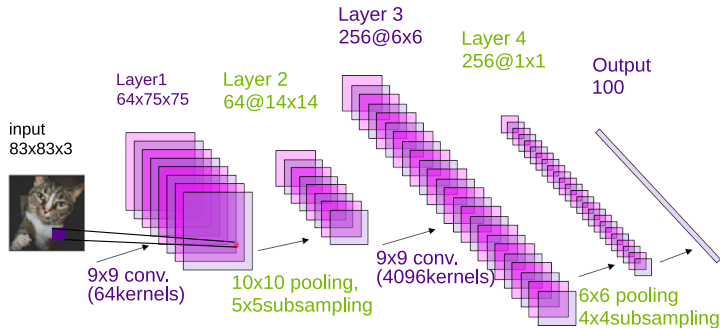


# Tensor Representation of Layers

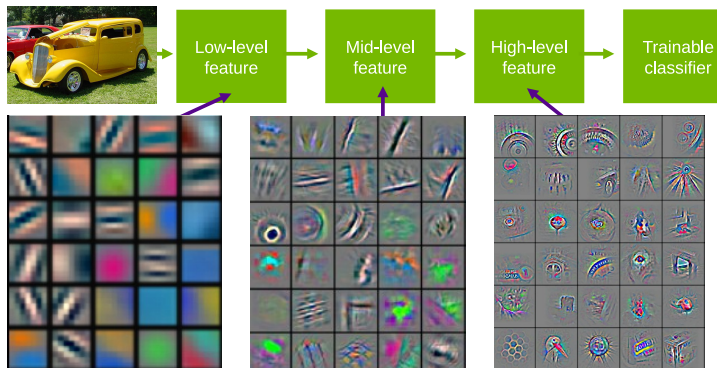


- convolution and pooling layers assign higher-dimensional topology to layers:
  - $(x, y)$  image dimensions + color channels = 3 dimensions
- convolutions sum over all input channels (or excess dimensions)

# Convolutional Neural Networks III



# Hierarchical Representations



Feature visualization of convolutional net trained on ImageNet<sup>1</sup>

<sup>1</sup>Zeiler & Fergus 2013

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

day2/notebooks/convnets\_cifar10

# One-Hot Encoding

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



↓ inference

class 0  
cat

class 1  
dog

class 2  
hat

...

class N  
mug

