Convolutional Neural Networks (CNNs)

FH University of Applied Sciences TECHNIKUM WIEN

Lecturer: Sharwin Rezagholi
Authors: Lars Mehnen, Matthias Blaickner, Bernhard Knapp, Sharwin Rezagholi

Classification
KNN classification
Classification trees
Ensembles & Boosting
Random Forest
Logistic regression
Naive Bayes
Support vector machines
Neural networks

Regression
KNN regression
Regression trees
Linear regression
Multiple regression
Ridge and Lasso regression
Neural networks

Supervised learning

AI

Machine learning process
Data handling
EDA, data cleaning
Training and testing
Feature selection
Class balancing
etc

Non-supervised learning

Clustering
k-means
Hierachical clustering
DB-scan

Dimensionality reduction
PCA / SVD
tSNE
Multi dimensional scaling
Linear discriminant analysis

Reinforcement learning
Other course.

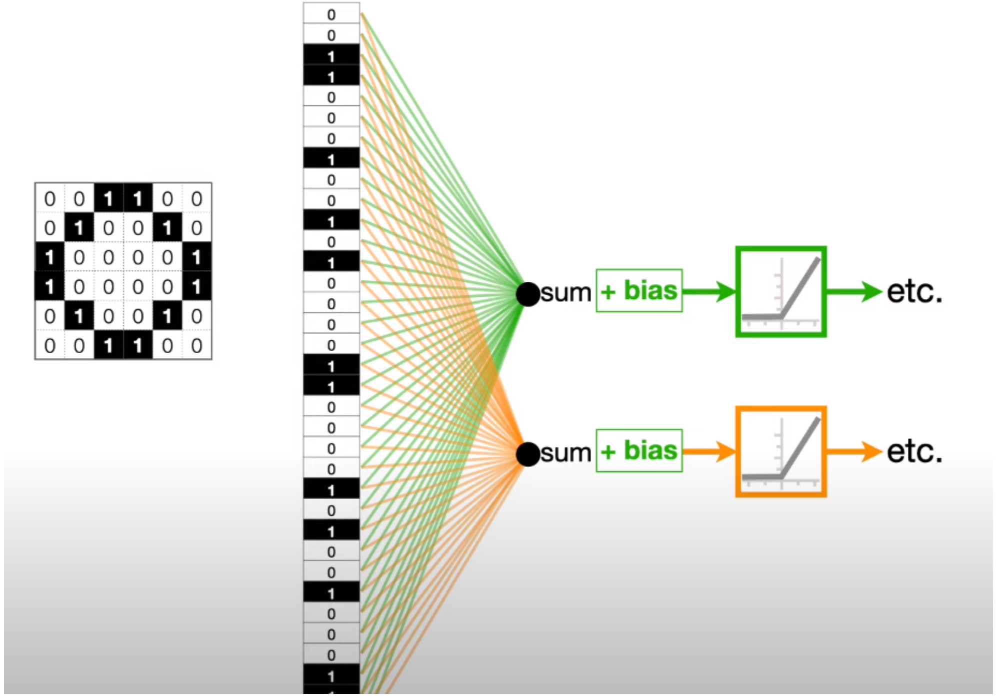
Generative AI
GANs

CNNs: Introduction

- CNNs emerged in the field of computer vision (e.g. recognition of handwritten letters).
- In computer vision our human visual intuition suggests that it may be important to recognize certain forms and patterns in images to classify correctly.
- CNNs are neural networks who are able to find such latent patterns and use them efficiently.
- To this day CNNs are mostly used in computer vision, but they have also shown good performance in other domains, where their initial visual intuition no longer applies.

CNNs: Motivation

- An image is a 2-dimensional array of color values.
- Therefore computer vision data is usually very high-dimensional.
- Example (RGB image):
- dimensionalityOfInput = numberOfPixels * 3
- = numberHorizontal * numberVertical * 3
- The number of weights and biases in a feedforward neural network that need to be estimated is huge.
- Also: The feedforward structure does not seem to be well-adjusted to the task of computer vision. Why?



From [3]

CNNs: Motivation

There are certain changes in input that can trick a feedforward neural network, but would never trick a human being.

FH University of Applied Sciences TECHNIKUM WIEN

CNNs: Motivation

Often it is not necessary to “understand” all pixel values equally well.

Blue pixels tend to be close to other blue pixels.

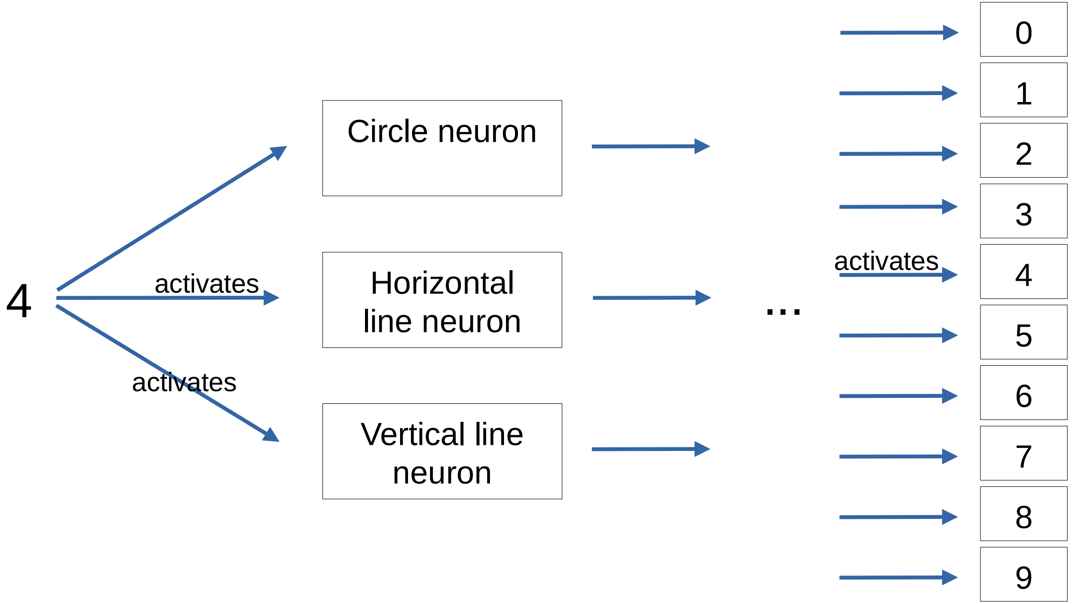
Green pixels tend to be close to other green pixels.

FH University of Applied Sciences TECHNIKUM WIEN

CNNs: Motivation

Suppose a neural network could have neurons that are sensitive to a certain pattern in the image.

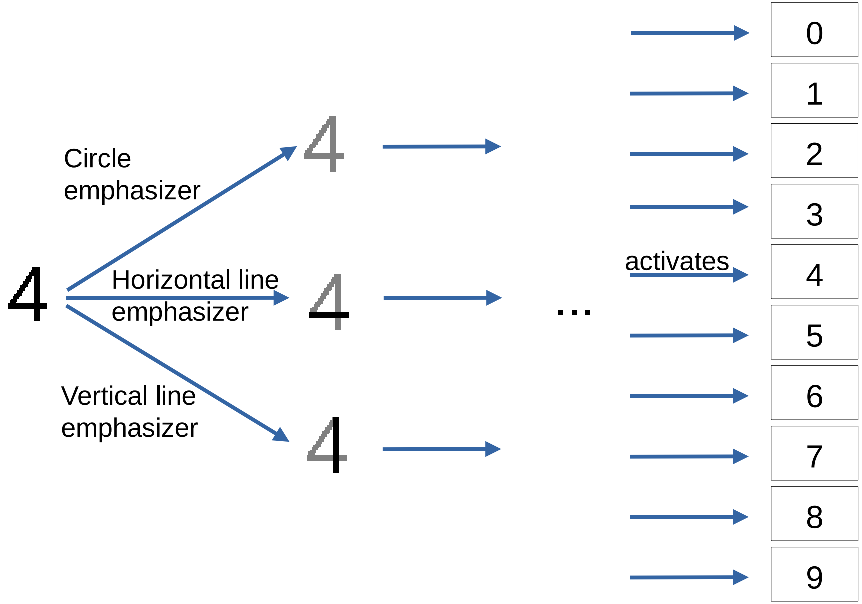
Example: Digit recognition.



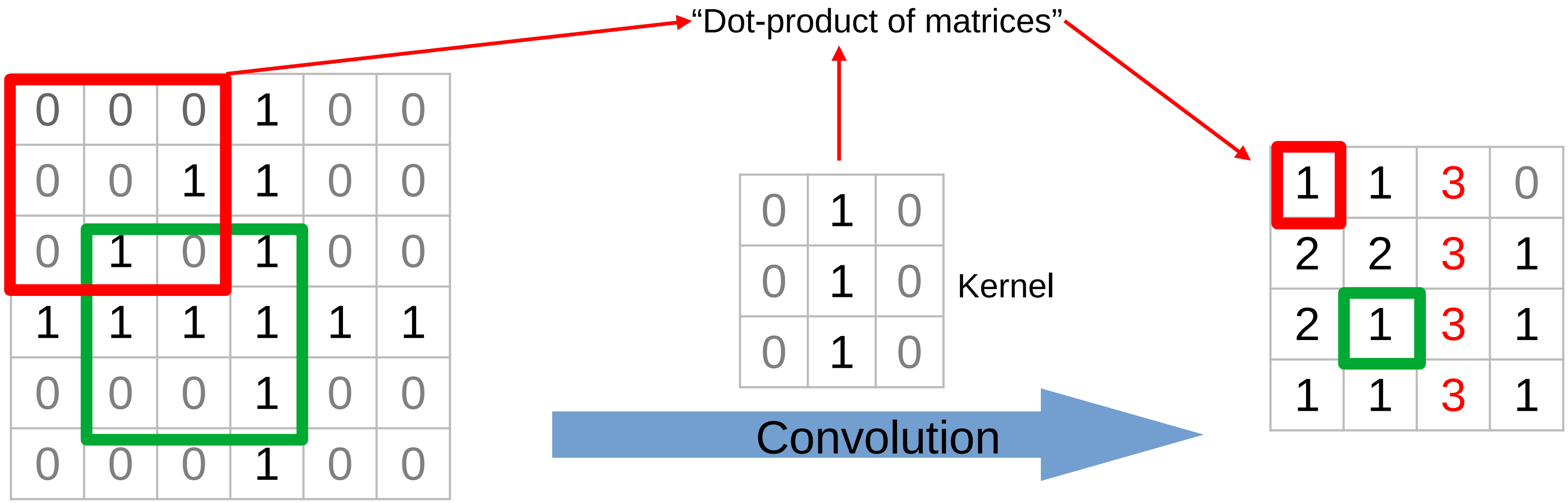
CNNs: Motivation

Suppose a neural network could have neurons that are sensitive to a certain pattern in the image. One way to implicitly achieve this is to obtain copies of the original input where certain properties have been emphasized.

Example: Digit recognition.

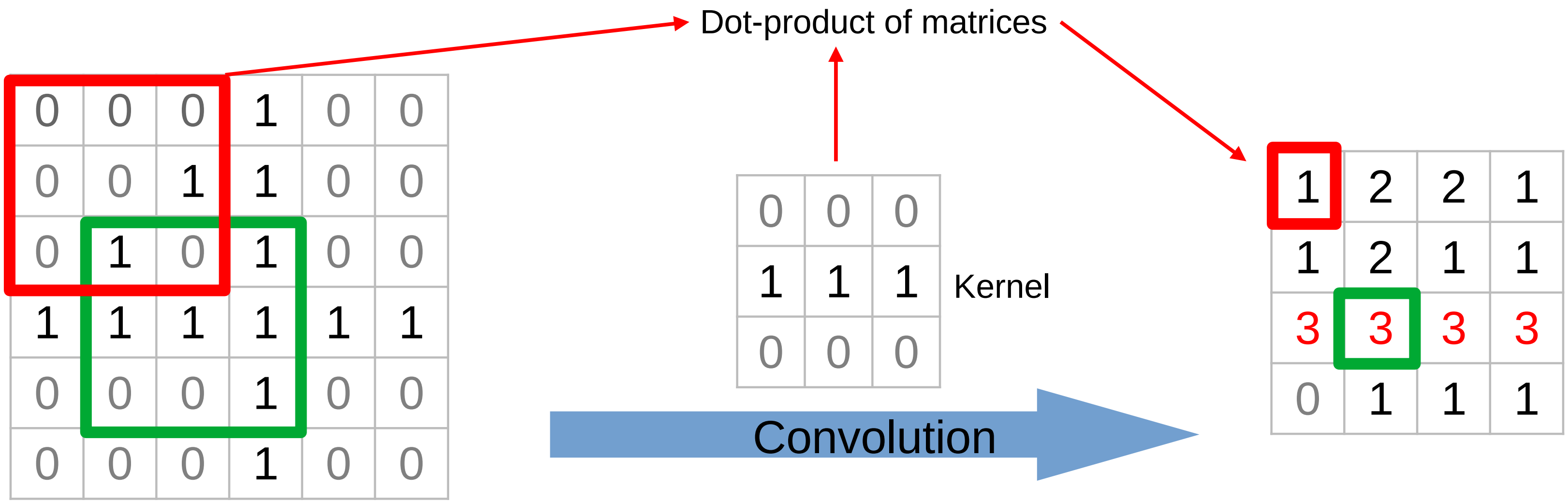


CNNs: Discrete Array Convolution



This emphasized the vertical lines in the image!

CNNs: Discrete Array Convolution



This emphasized the horizontal lines in the image!

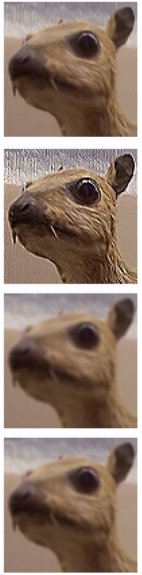
FH University of Applied Sciences TECHNIKUM WIEN

CNNs: Convolution Kernels

Even seemingly simple kernels can be very effective.

But most kernels are not human-interpretable. Still, they correspond to latent features that carry information for the task.

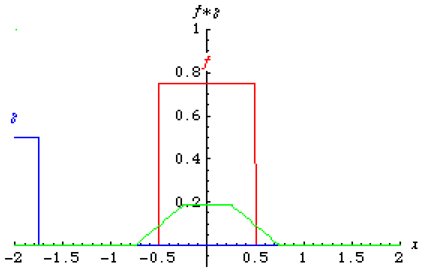
Identity	$\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$
Sharpen	$\begin{matrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{matrix}$
Box blur	$\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$
Gaussian blur	$\begin{matrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{matrix}$

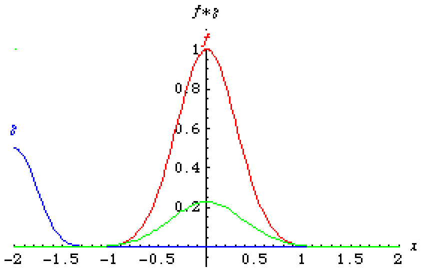


https://wikipedia.org/wiki/Kernel_(image_processing)

Aside: Convolution in Calculus

You may have encountered the convolution of functions with (smoothing) kernel functions in calculus.
This is a common approach to smooth an irregular function, for example to obtain a smooth empirical density in statistics.

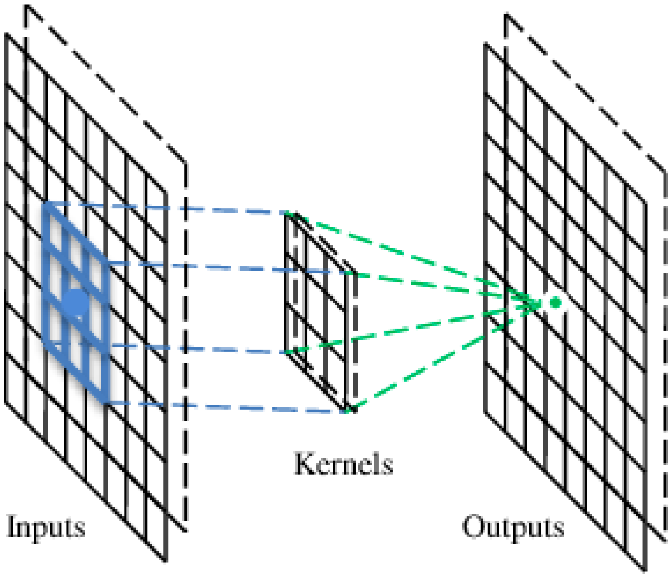




[Image source: http://mathworld.wolfram.com/Convolution.html]

CNNs: Convolutional Layer

A CNN is a neural network that contains convolutional layers.
A convolutional layer contains kernels that correspond to the respective convolution.



CNNs: Padding

0	0	0	1	0	0
0	0	1	1	0	0
0	1	0	1	0	0
1	1	1	1	1	1
0	0	0	1	0	0
0	0	0	1	0	0

Zero-padding

0	0	0
0	0	1
0	1	1

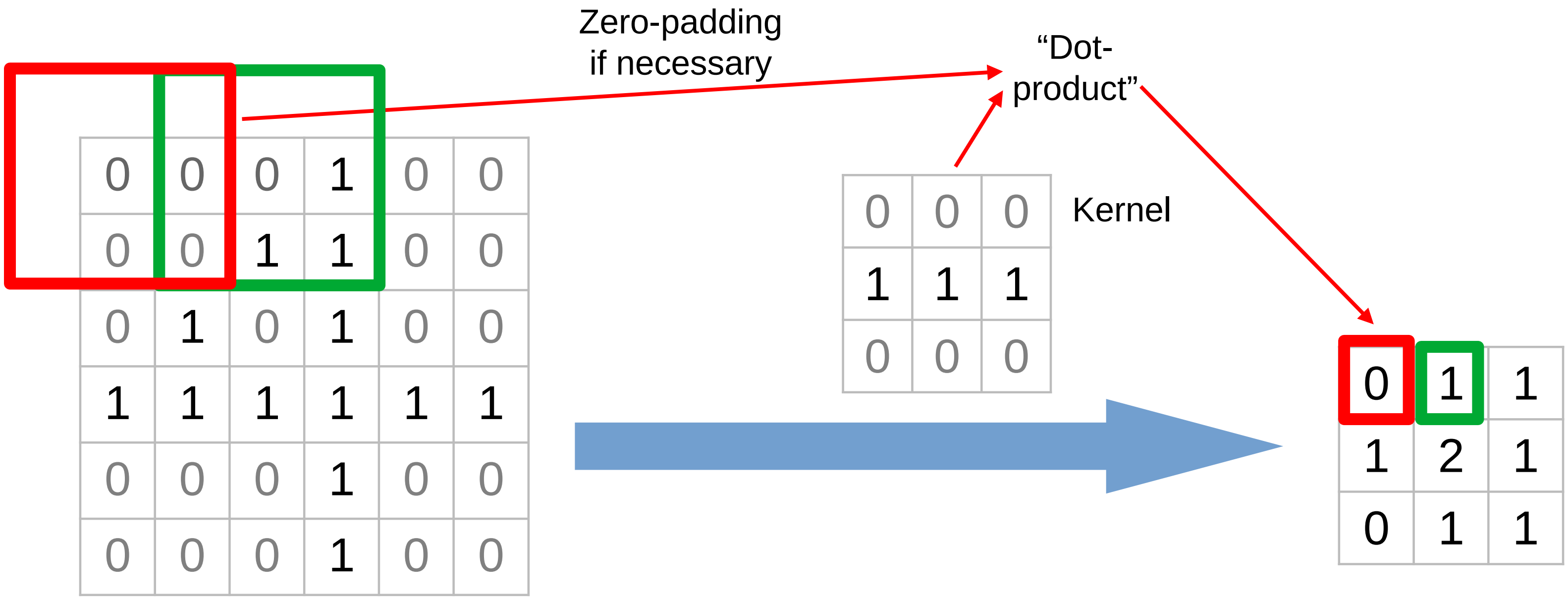
"Dot product"

0	0	0
1	1	1
0	0	0

Kernel

0	0	$\boxed{1}$	1	1	0
0	1	2	2	1	0
1	1	2	1	1	0
2	3	3	3	3	2
0	0	1	1	1	0
0	0	1	1	1	0

CNNs: Stride Parameter



The stride parameter controls the number of indices by which we shift during convolution. Here: Stride parameter = 2.

CNNs: Dimension of Kernel

0	0	0	1	0	0
0	0	1	1	0	0
0	1	0	1	0	0
1	1	1	1	1	1
0	0	0	1	0	0
0	0	0	1	0	0

“Dot-product”

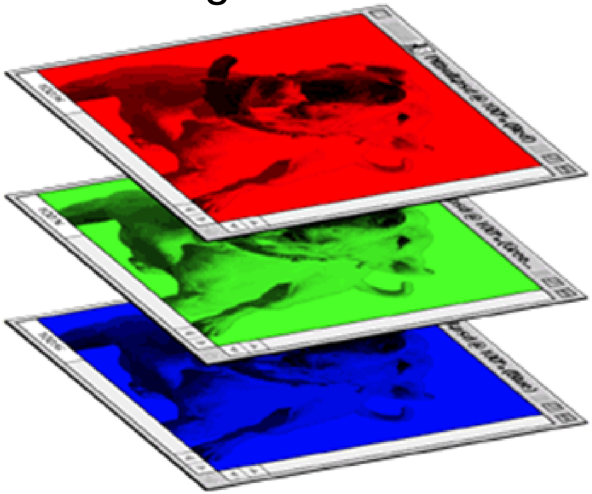
0	0	0	0	0
0	0	0	0	0
1	1	1	1	1
0	0	0	0	0
0	0	0	0	0

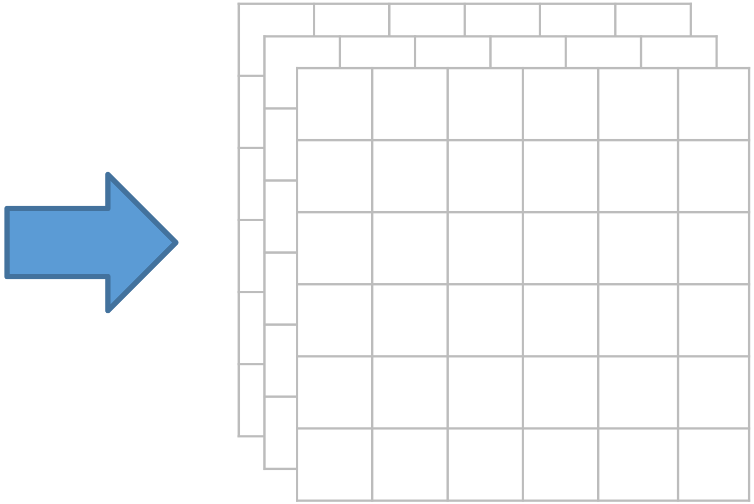
2	2
5	5

The (5*5)-kernel reduces the size of the output even more.

CNNs: RGB image (3 channels)

Color image





Convolution Layers vs. Fully Connected Layers

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \underset{\text{Convolution}}{*} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} aA+bB+dC+eD & bA+cB+eC+fD \\ dA+eB+gC+hD & eA+fB+hC+iD \end{bmatrix}$$

Flattening

Flattening

$$\begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \end{bmatrix} \begin{bmatrix} A & 0 & 0 & 0 \\ B & A & 0 & 0 \\ 0 & B & 0 & 0 \\ C & 0 & A & 0 \\ D & C & B & A \\ 0 & D & 0 & B \\ 0 & 0 & C & 0 \\ 0 & 0 & D & C \\ 0 & 0 & 0 & D \end{bmatrix} = \begin{bmatrix} aA+bB+dC+eD \\ bA+cB+eC+fD \\ dA+eB+gC+hD \\ eA+fB+hC+iD \end{bmatrix}$$

Convolution Layers vs. Fully Connected Layers

- Suppose the input array has size n*n.
- A k*k convolutional layer is equivalent to a certain n^2 to (n-k+1)^2 linear layer.
- The set of such convolutions “is” a (k*k)-dimensional subset of the n^2 (n-k+1)^2-dimensional set of matrices.
- Due to lower-dimensional subspace:
  - More bias
  - Less variance
  - Easier to learn
- Why do you think that convolutional layers work (in computer vision)?

FH University of Applied Sciences TECHNIKUM WIEN

CNNs: Max-Pooling

Max pooling is used to reduce the dimensionality of input.
This reduces computational burden, can prevent overfitting, and provides some translation invariance to the representation.
Max-pooling applies a max-filter to (usually) non-overlapping subregions of the initial representation.

0	0	0	1	0	0
0	0	1	1	0	0
0	1	0	1	0	0
1	1	1	1	1	1
0	0	0	1	0	0
0	0	0	1	0	0

(2*2)-max-pooling

0	1	0
1	1	1
0	1	0

CNNs: Max-Pooling

Max pooling is often used after a convolution.

0	0	0	1	0	0
0	0	1	1	0	0
0	1	0	1	0	0
1	1	1	1	1	1
0	0	0	1	0	0
0	0	0	1	0	0

Kernel

0	0	0
1	1	1
0	0	0

With Zero-padding

0	0	1	1	1	0
0	1	2	2	1	0
1	1	2	1	1	0
2	**3**	**3**	**3**	**3**	**2**
0	0	1	1	1	0
0	0	1	1	1	0

(2*2)-max-pooling

1	2	1
3	**3**	**3**
0	1	1

The emphasized horizontal line is still “visible”.

FH University of Applied Sciences TECHNIKUM WIEN

CNNs: Average-Pooling

Pooling can also be done by averaging.

0	0	0	1	0	0
0	0	1	1	0	0
0	1	0	1	0	0
1	1	1	1	1	1
0	0	0	1	0	0
0	0	0	1	0	0

Kernel

0	0	0
1	1	1
0	0	0

With Zero-padding

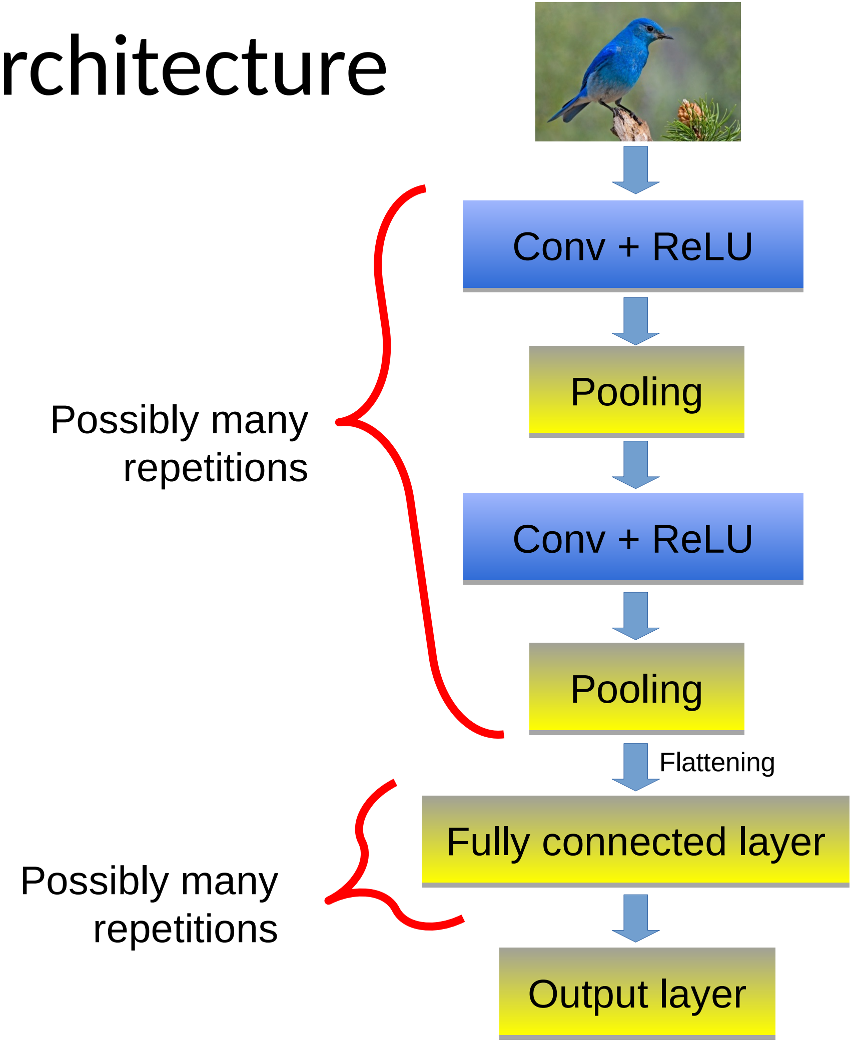
0	0	1	1	1	0
0	1	2	2	1	0
1	1	2	1	1	0
2	**3**	**3**	**3**	**3**	**2**
0	0	1	1	1	0
0	0	1	1	1	0

(2*2)-avg-pooling

0.25	1.5	0.5
1.75	**2.5**	**1.5**
0	1	0.5

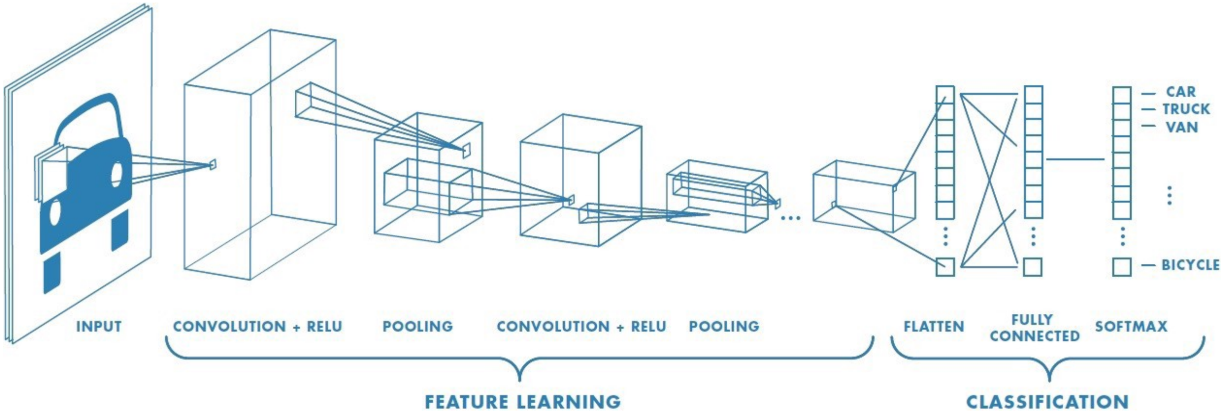
Average pooling leads to more homogeneous output arrays.

CNNs: The whole Architecture



FH University of Applied Sciences TECHNIKUM WIEN

CNNs: Parameters



Feature learning part:

- Values in the kernel matrices

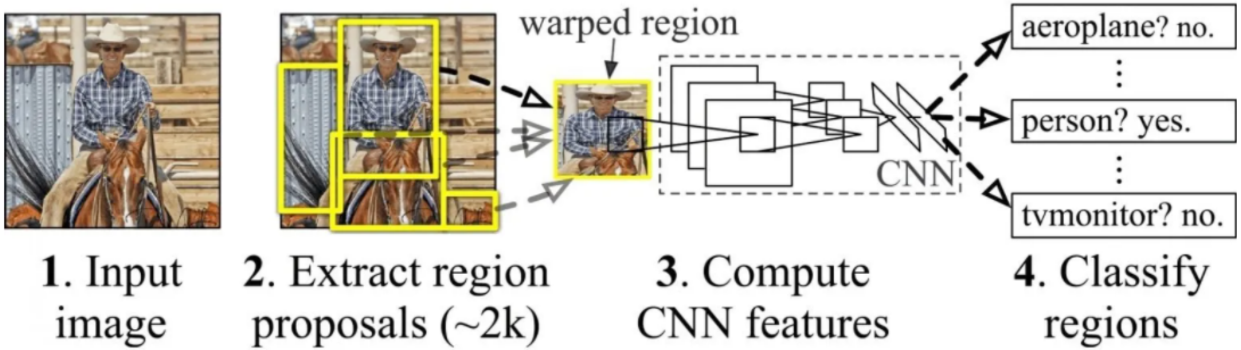
Classification part:

- Weights of artificial neurons
- Biases of artificial neurons

FH University of Applied Sciences TECHNIKUM WIEN

Region-based CNNs (RCNNs)

Utilizes bounding boxes across regions, then evaluates convolutional networks independently on regions of interest to identify different objects in one image.



Concept of R-CNN – Region-based Convolutional Networks

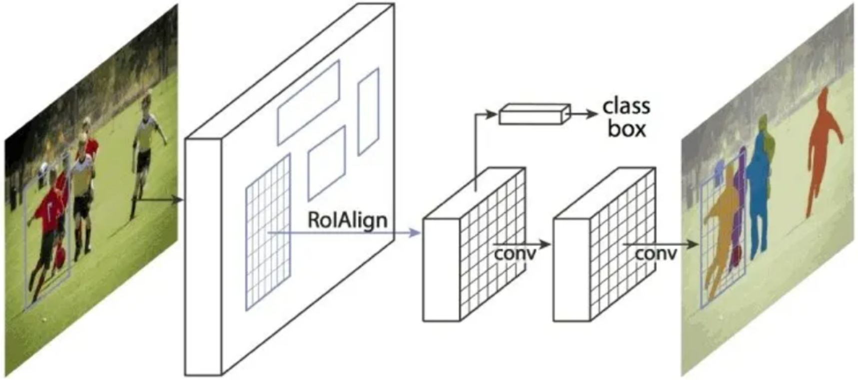
https://viso.ai/deep-learning/mask-r-cnn/

FH University of Applied Sciences TECHNIKUM WIEN

Mask RCNNs

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels). Mask RCNNs additionally output the object mask.

Example: Organ/tumor segmentation in medical imaging.



Mask R-CNN – The Mask R-CNN Framework for Instance Segmentation

https://viso.ai/deep-learning/mask-r-cnn/

CNNs for Face Detection and Facial Expression Recognition

Detection of smiles with recognition rate of 97.6%.

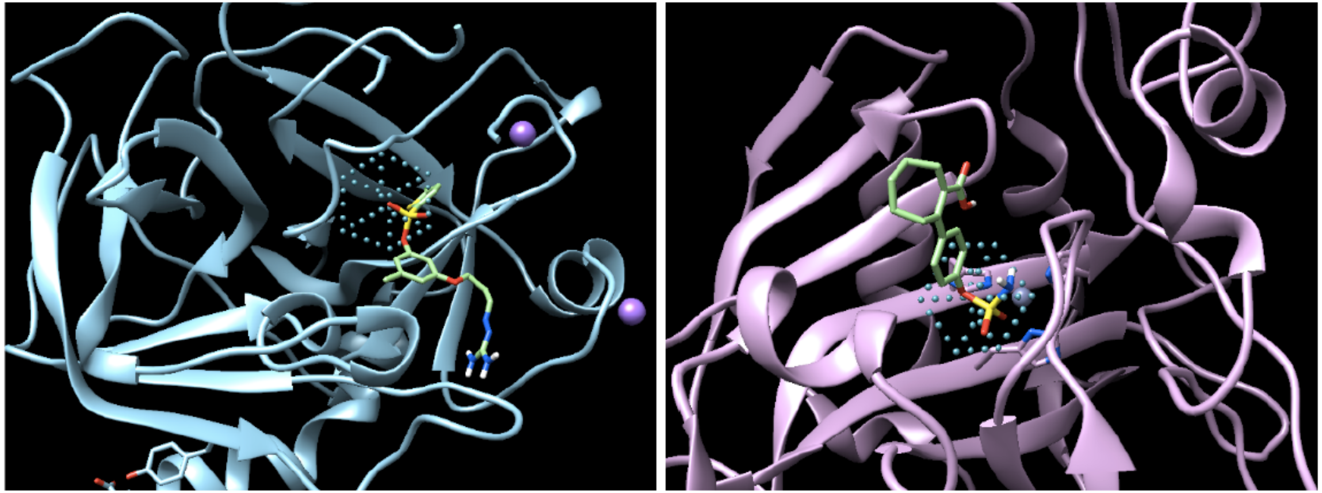
See reference [1].

CNNs for Drug Discovery

Sulfonyl/sulfonamide detector. A CNN was able to infer a meaningful spatial arrangement of input atom types without any chemical prior knowledge.

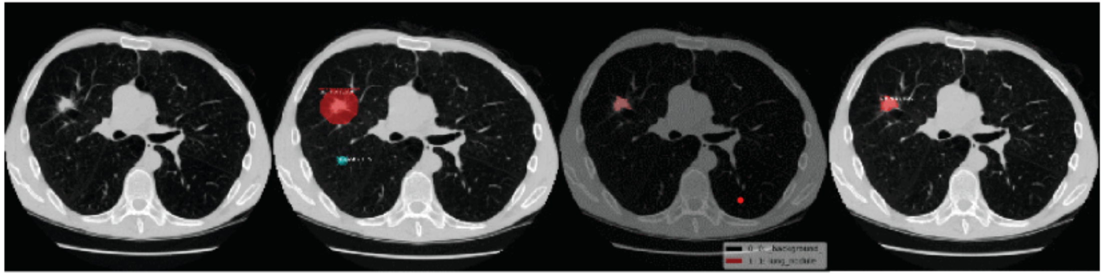
Note: This is an application outside of computer vision.

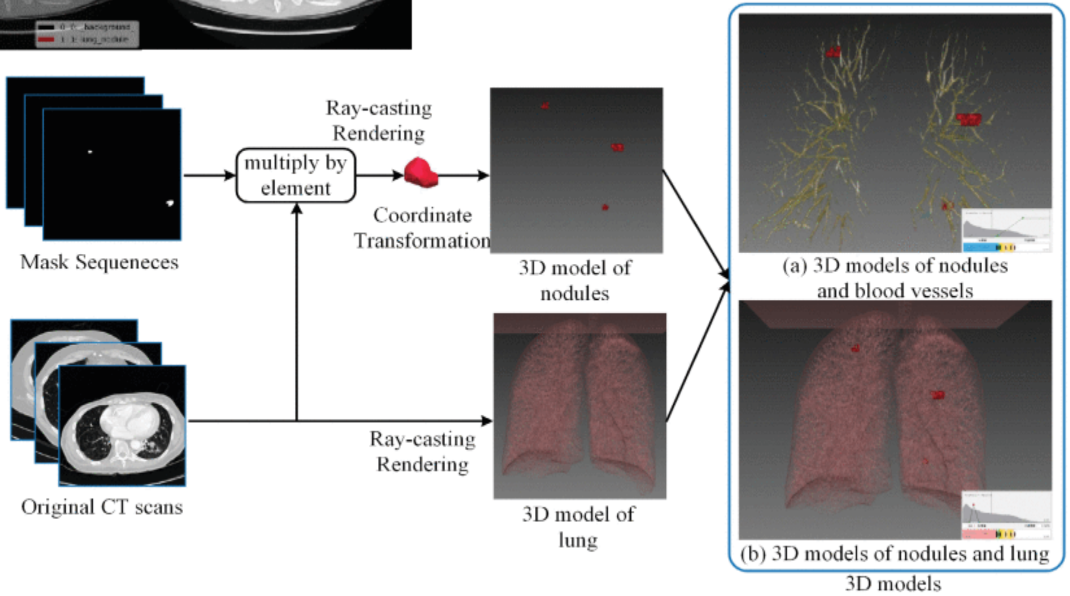
See reference [2].



FH University of Applied Sciences TECHNIKUM WIEN

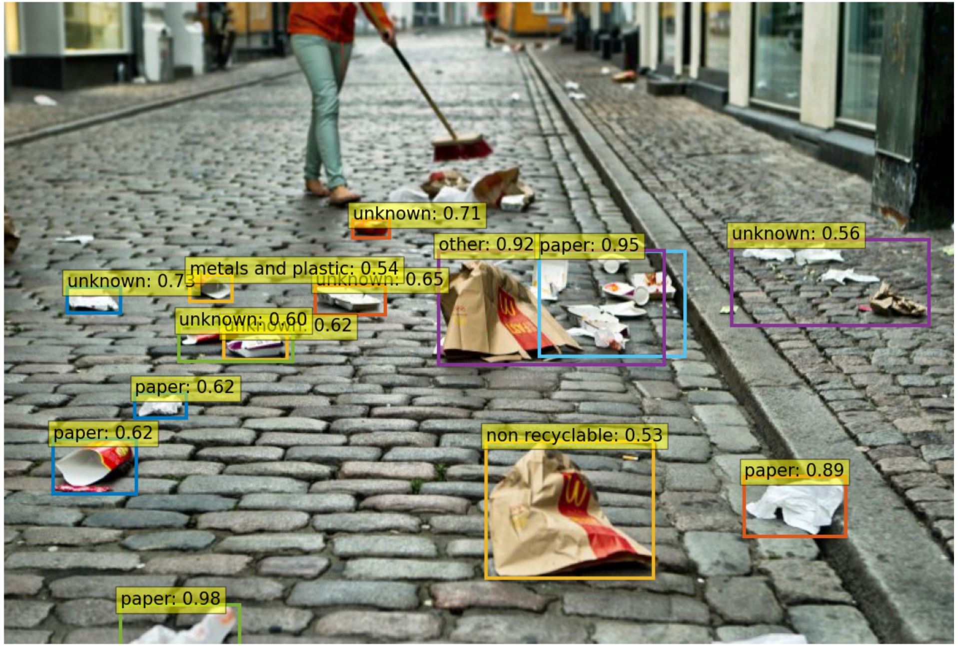
CNNs for Medical Image Segmentation





See reference [4].

CNNs for Waste Detection



https://awesomeopensource.com/project/wimlds-trojmiasto/detect-waste

References

[1] Matsugu M et al. Subject independent facial expression recognition with robust face detection using a convolutional neural network, Neural Networks 16 (2003) 555–559.

[2] Wallach I et al. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. arXiv:1510.02855.

[3] StatQuest with Josh Starmer: Neural Networks Part 8: Image Classification with Convolutional Neural Networks. https://www.youtube.com/watch?v=HGwBXDKFk9I

[4] Cai L et al. AtomNet: Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis. in IEEE Access, vol. 8, pp. 44400-44409, 2020, doi: 10.1109/ACCESS.2020.2976432.

FH University of Applied Sciences TECHNIKUM WIEN

Assignment: CNNs

a) Explain CNNs using 3 slides/pages.

Use self-made images or even hand drawings (of which you take a photo).

Use self-written explanations.

Do not copy from the lecture slides or the internet (neither text nor images).

b) Coding exercise

- Use a library (e.g. SciKit-Learn, or Keras) to compare the performance of CNNs and Multilayer perceptrons in computer vision.
- Let them compete on an image classification dataset of your choice (e.g. MNIST). If both always perform well, shift images or add noise.
- Compare the development of training and test error during gradient descent between CNN and ANN.
- Compare the number of parameters used for the CNN and the ANN.

FH University of Applied Sciences TECHNIKUM WIEN