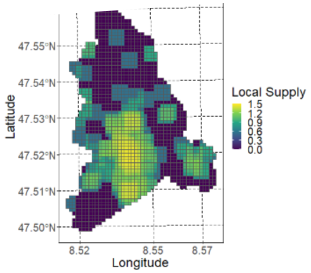
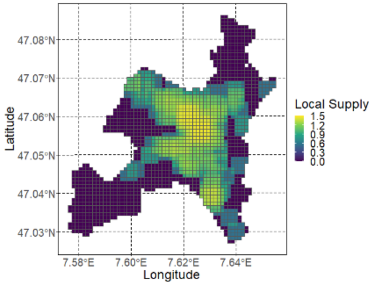
FH
GR
Fachhochschule Graubünden
University of Applied Sciences

Convolutional Neural Networks

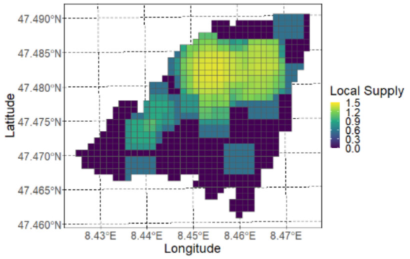
Dr. Yves Staudt



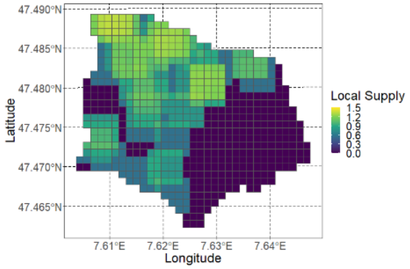
(a)
Grenchen



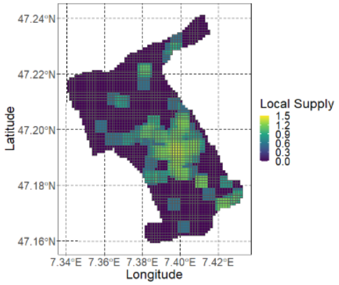
(b)
Oberglatt



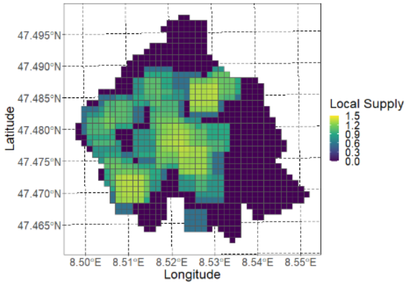
(c)
Reinach



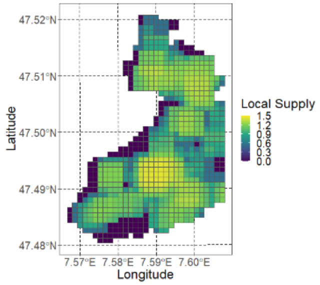
(d)
Rubigen



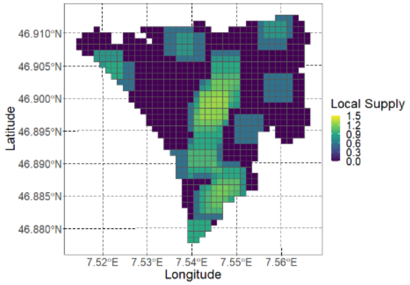
(e)
Buelach



(f)
Burgdorf



(g)
Dielsdorf



(h)
Dornach

Lernziel

Die Studierende sind in der Lage

- Convolutional Neural Networks anzuwenden und zu beschreiben.

Why Convolutional Networks

- Dense layers learn global patterns in their input feature space

- Convolutional layers learn local patterns

(Chollet, 2018)

Key Characteristics Convolutional Neural Networks (CNN)

```
CNN:
(128, 28, 28, 1) input shape
-2 Dimensionen durch conv 2D (64)
(128, 26, 26, 64) (64 ist Batch)
werden 2 von 28 pixel reduziert
geht so laufend in quadraten durch und erkennt die Muster "oben links (lokal)

FFN: einzelne Nodes in 2D, sind alle verbunden. Erkennt Muster nur als
Ganzes(global)
```

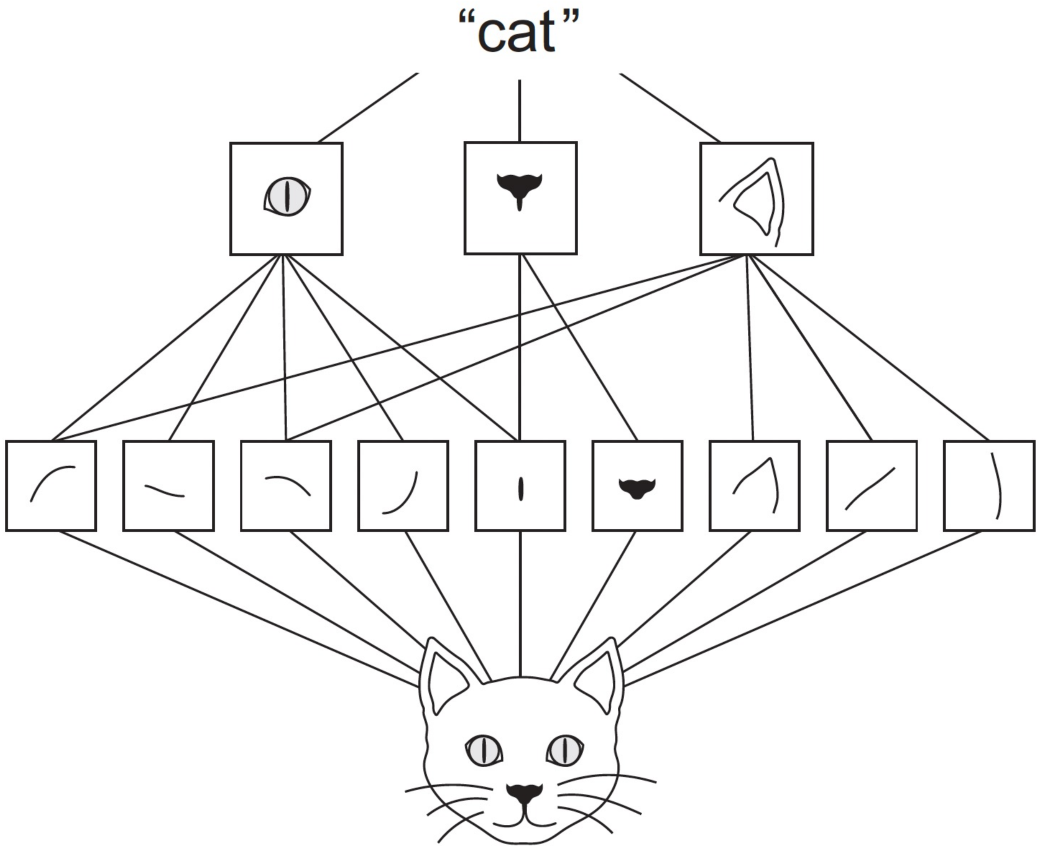
2 Key Properties of Convolutional Networks:

1. Learned patterns are translation invariant
   - After learning a certain pattern in a picture, a convnet can recognize it anywhere
   - Visual world is fundamentally translation invariant
   - Need of fewer training samples to learn representations that have generalization power

2. convnet learn spatial hierarchies of patterns
   - First convolution layer will learn small local pattern such as edges
   - Second convolution layer will learn larger patterns made of the features of the first layers
   - Convnets are allowed to learn efficiently increasingly complex and abstract visual concepts

(Chollet, 2018)

Visual Representation Spatial Hierarchies of Patterns (Chollet, 2018)



Tensors in CNN

- Convolutions operate over **3D tensors**
- Feature maps: two **spatial axes** (height and width) and a **depth axis** (channel axis)
- In the case of an RGB image: depth axis has a dimension of 3, 3 colors
- Black and white image: depth axis of dimension 1

(Chollet, 2018)

Feature Map

- Convolution operation extract **patches** from its input feature map
- Convolution applies the **same transformation** to all of these patches, producing an output called **feature map**
- Output feature map is a **3D tensor** with width and height and an arbitrary depth
- **Depth** is a parameter of the layer
- Depth axis no longer stands for specific colors
- Depth axis represent filters
- **Filters** encode specific aspects of the input data
- For example: a filter could encode the concept of presence of a face in the input
- Every dimension in the depth axis is a feature (or filter)
- The 2D tensor output is the 2 spatial map of the response of this filter over the input

(Collet, 2018)

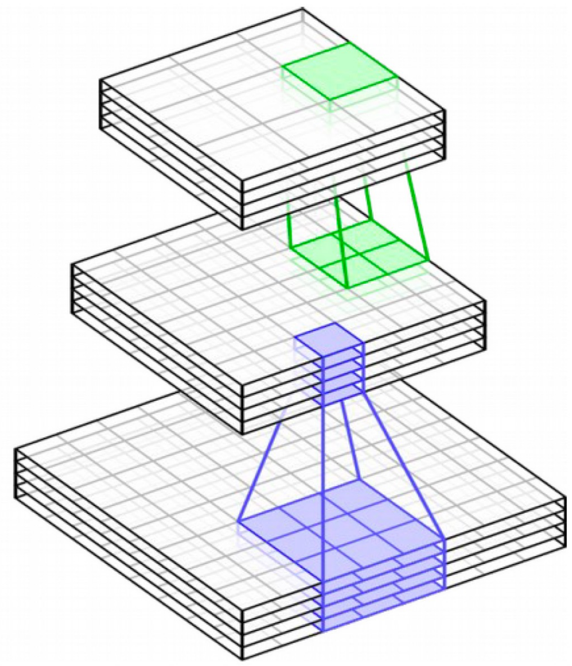
Definiton Two Key Parameters of Convolutions

1. Size of patches extracted from the inputs

   - ➢ Typically 3x3 or 5x5

2. Depth of the output feature map

(Chollet, 2018)

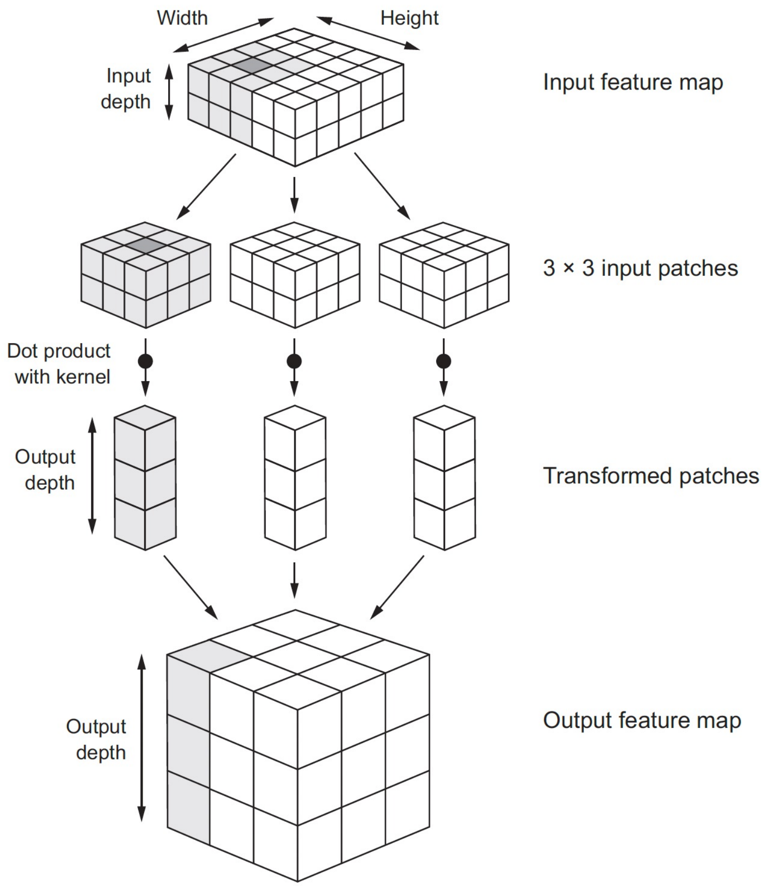


Visualisation of the Patch Description (Paulin et al., 2016)

FH
GR

Process of Convolutional Neural Networks

- Convolution works by sliding the windows of size 3x3 or 5x5 over the 3D input
- Windows stop at every possible location
- Each window extracts a 3D Patch of surrounding features (shape (window height, window width, input depth)
- Each 3D Patch is transformet into a 1D vector of shape (output depth)
- All vectors are spatially reassembled into a 3D output map of shape (height, width, output depth)
- Every spatial lcoation in the output feature corresponds to the same location in the input feature



Process of the CNN (Chollet, 2018)

Border Effects

Output width and height may differ from the input width and height

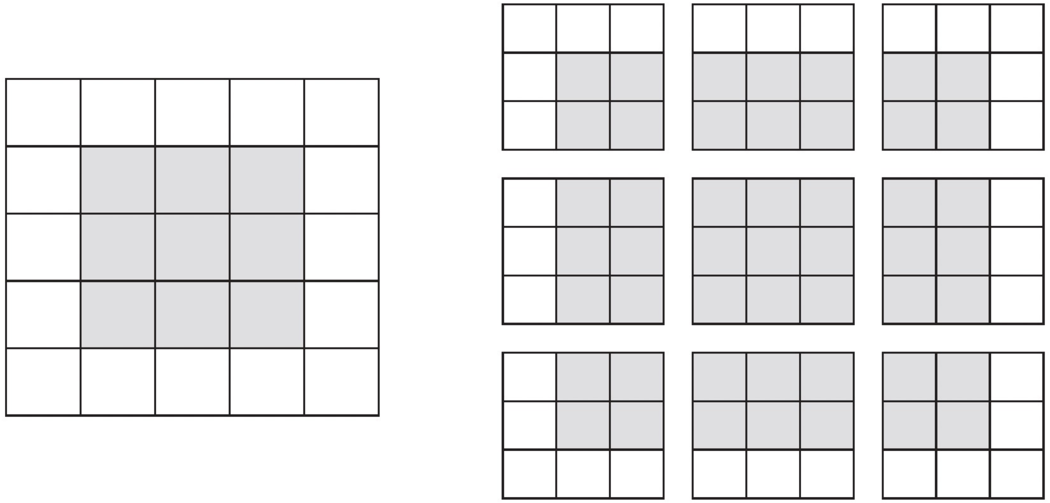
They may differ for two reasons:

1. Border effects can be encountered by padding the input feature map
2. Use of strides

Understanding Border Effects:

- 5x5 feature map (25 tiles)
- Windows of 3x3
- Output of 3x3
- Input shrinks by exactly two tiles alongside each dimension

(Chollet, 2018)



Application of 3x3 windows on a 5x5 feature map (Chollet, 2018)

Padding

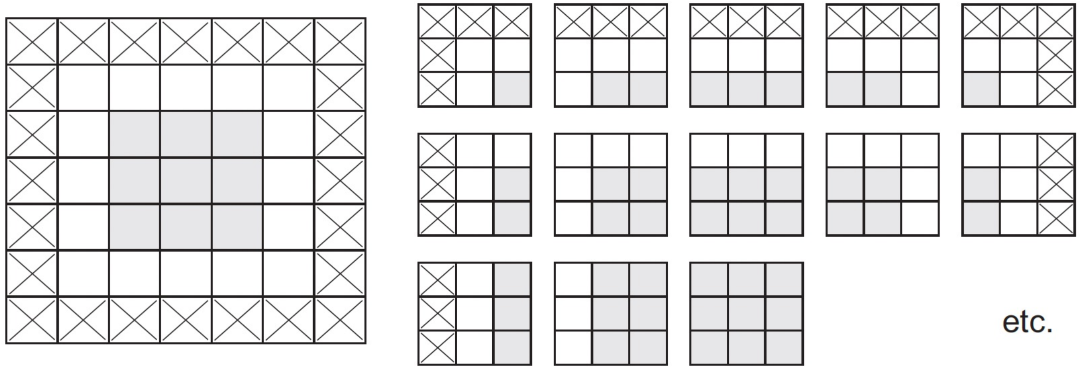
- **Goal**: Output feature map with the same spatial dimension as input feature map

- **Solution**: Padding

- Padding consists of adding an appropriate number of rows and columns on each side of the input feature map

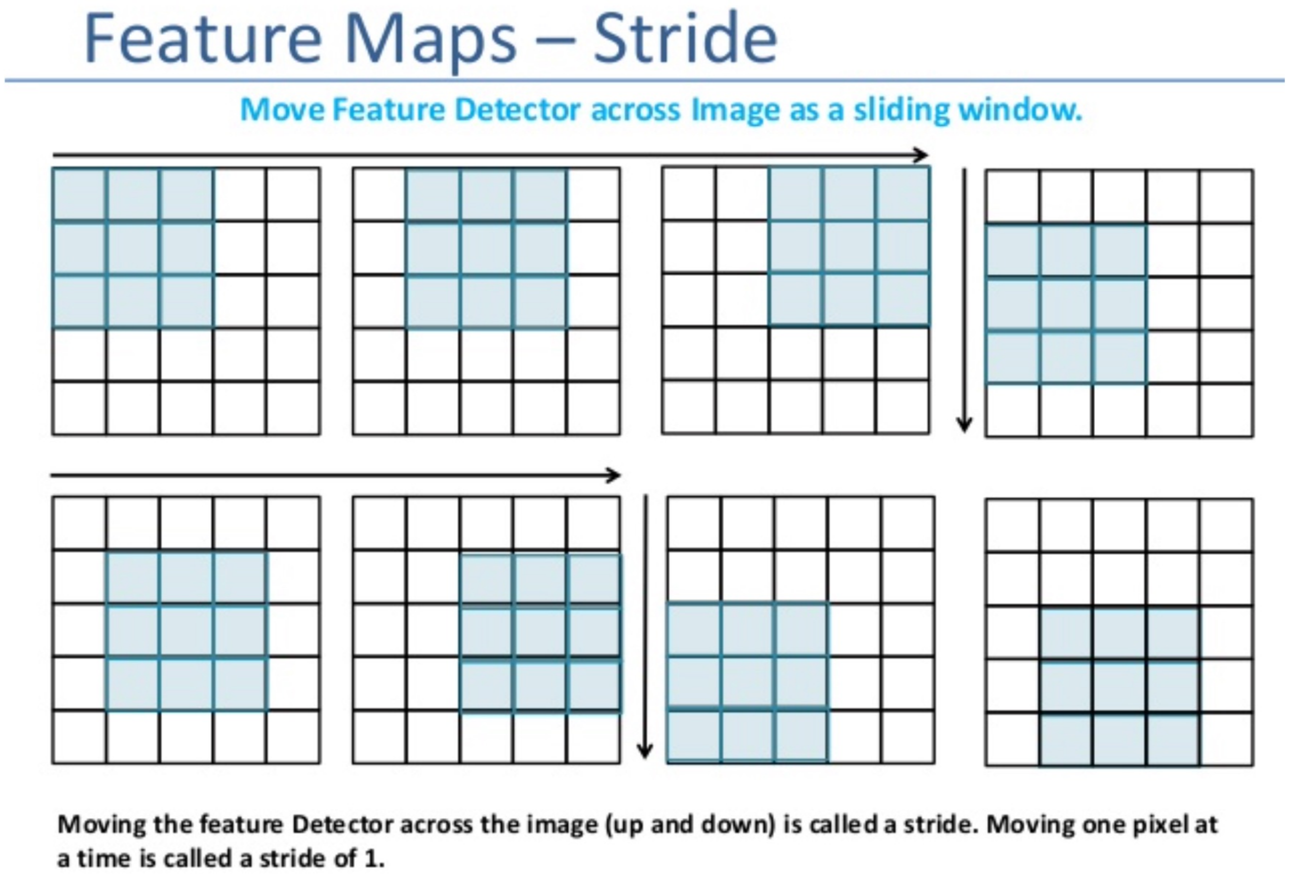
- Make it possible to fit center convolution windows around every input tile

(Chollet, 2018)



Visualisation of padding for 5x5 feature map (Chollet, 2018)

Strides (Mishra, 2023)



Feature Maps – Stride

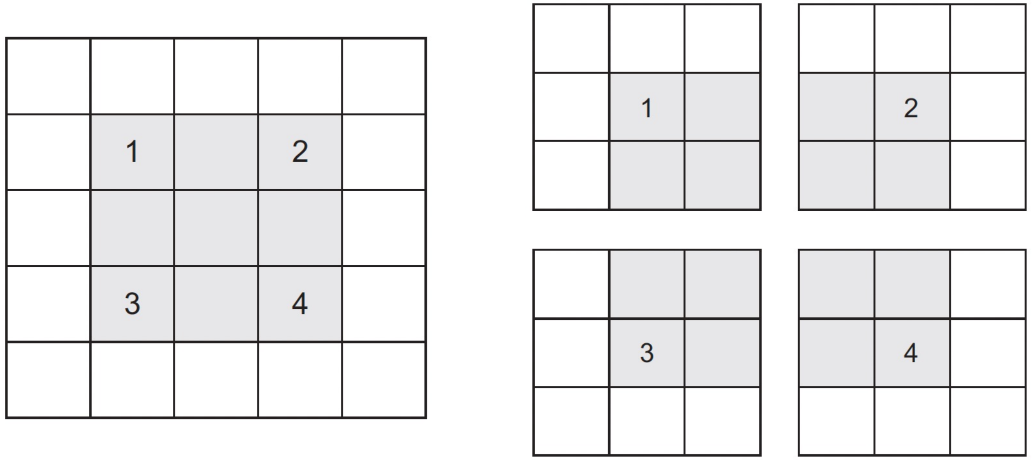
Move Feature Detector across Image as a sliding window.

Moving the feature Detector across the image (up and down) is called a stride. Moving one pixel at a time is called a stride of 1.

Stride

- Output size is influenced by strides
- Mostly convolution windows are all contiguous
- Using stride 2 means the width and height of the feature map are downspampled by factor of 2
- Strided convolutions are rarely applied in practice
- Max-Pooling is mostly used for downsampling the feature space.

(Chollet, 2018)

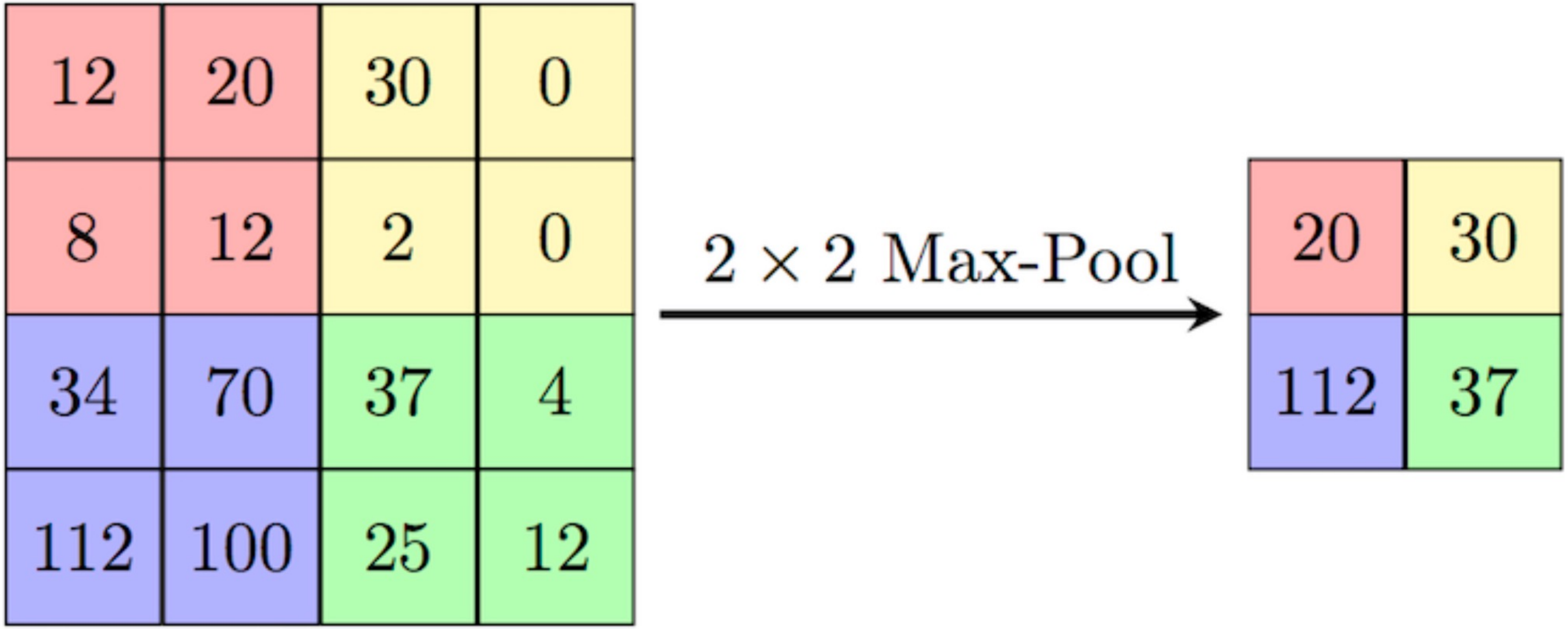


3x3 convolution patches with 2x2 strides (Chollet, 2018)

Max-Pooling

- **Goal of Max-Pooling:** To agressively downsample feature maps
- Max-pooling consist of extracting windows from the input feature maps and outputting the max value of each channel
- **Convolution Kernel:** local patches transformation via learned linear transformations
- **Max-pooling Kernel:** transformation via hardcoded max tensor operation
- Max-pooling is usually done with 2x2 windows and stride 2, in order to downsample the feature maps by a factor of 2

Visualisation Pooling (Mishra, 2023)



12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

20	30
112	37

Why Max-Pooling?

Max-Pooling is important for:

- To learn spatial hierarchy features – when max-pooling is not applied, initial input may not be enough to learn features for classifying the pictures

- To reduce the space – when max-pooling is not applied, the number of parameters can be far too large to be flattened

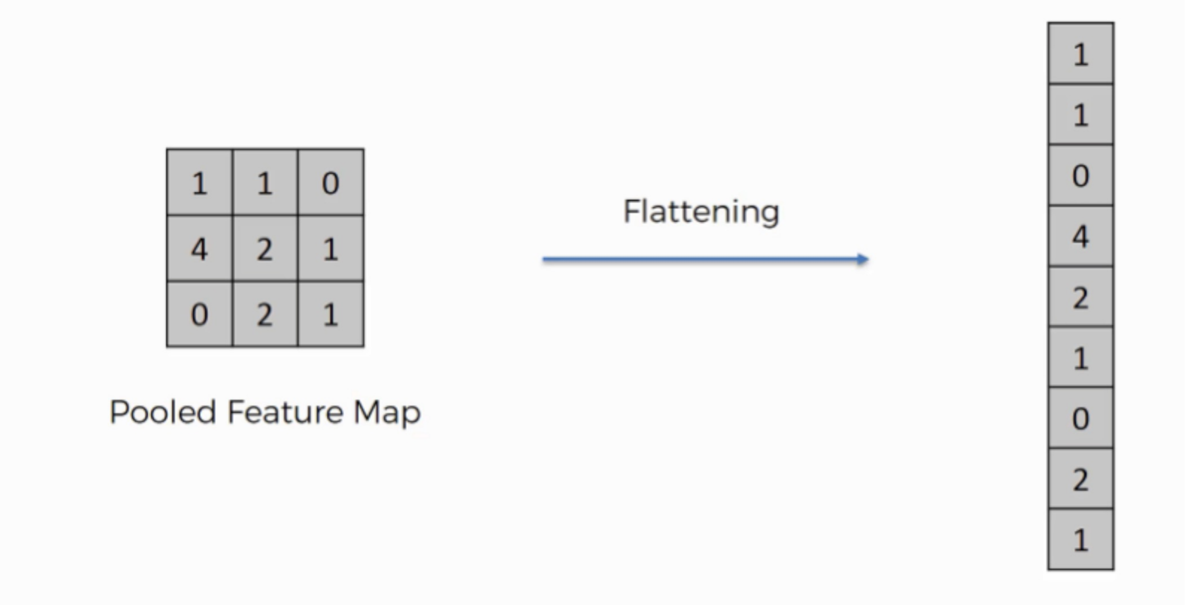
(Chollet, 2018)

Average Pooling

- Average pooling instead of max-pooling
- **Average pooling:** each local input patch is transformed by taking the average value of each channel over the patch
- Max-pooling tends to work better
- Reason of max-pooling: features tend to encode the spatial presence of some pattern or concept over the different tiles of the feature map
- Looking at the maximal presence of different features than at the average presence is more informative
- **Reasonable subsampling strategy:** to first produce dense maps of features (via unstrided convolutions) and then look at the maximal activation of the features over small patches, rather than looking at sparser windows of the inputs

(Chollet, 2018)

Visualisation of Flattening (Mishra, 2023)

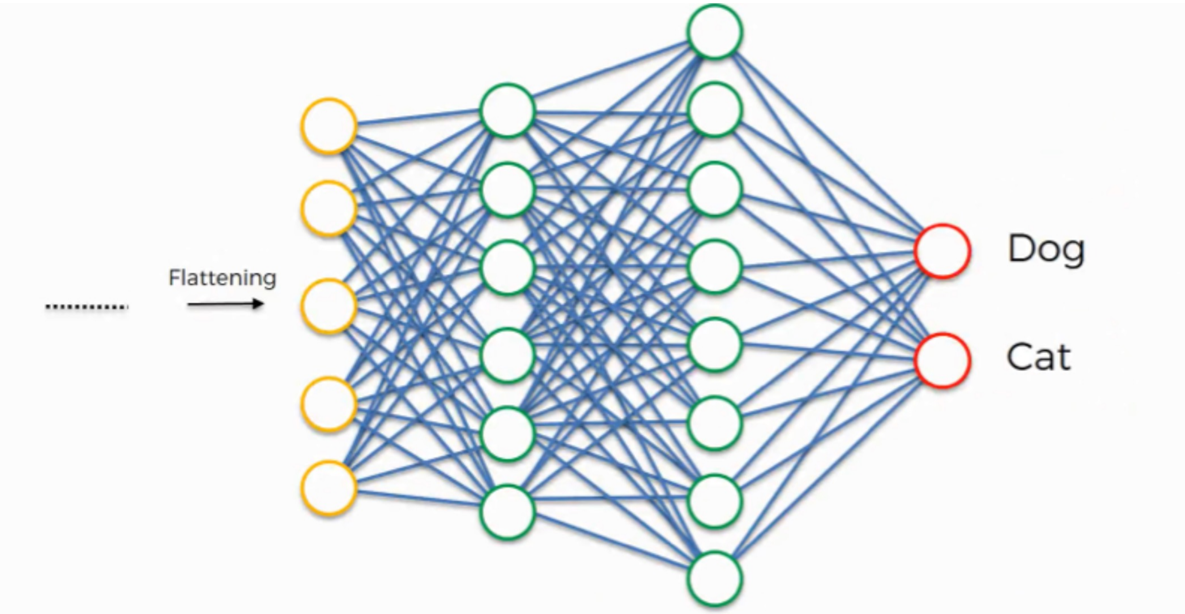


1	1	0
4	2	1
0	2	1

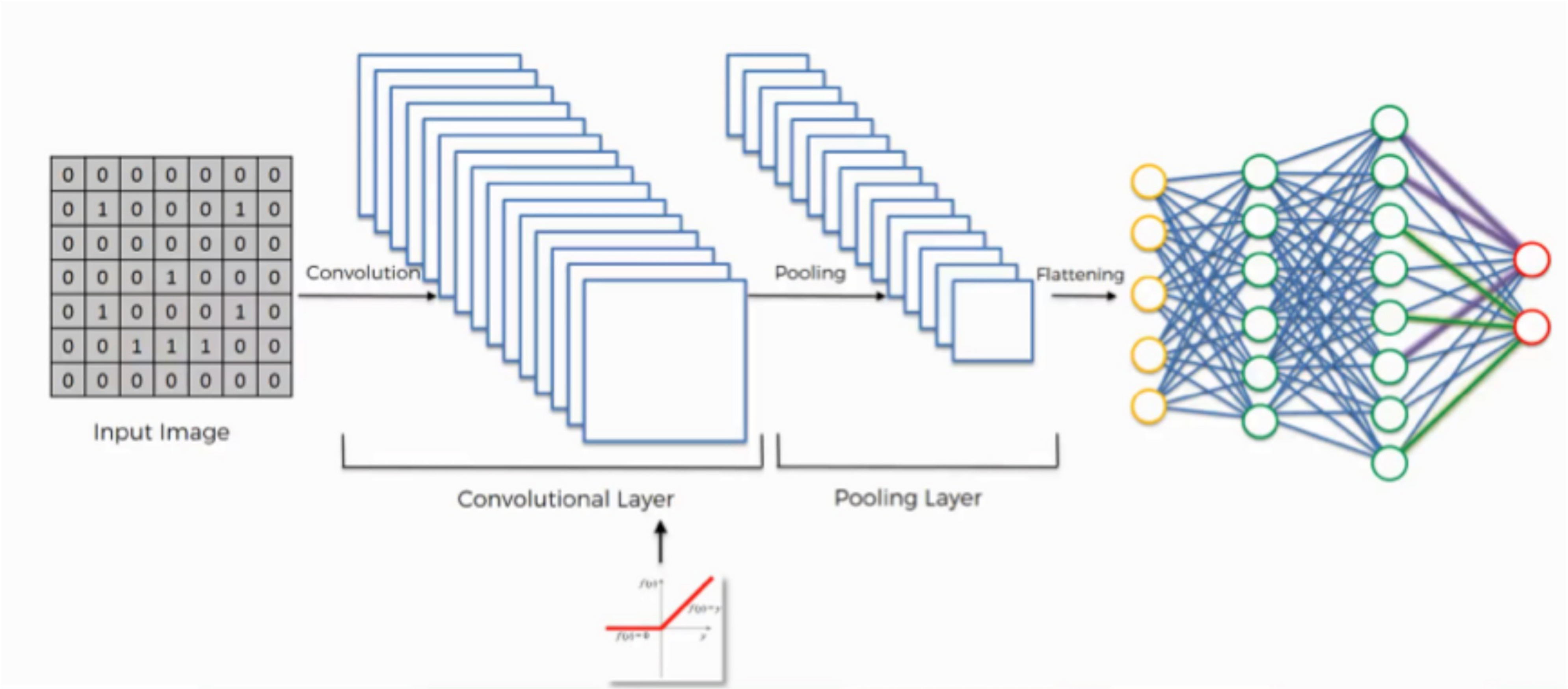
Pooled Feature Map

Flattening

1
1
0
4
2
1
0
2
1



Summary of Steps of Convolutional Neural Network (Mishra, 2023)



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

Aufgabe

Als Aufgabe sollen Sie nun einen Convolutional Neural Network von der Grundbasis für folgenden Katzen und Hunden Datensatz aufbauen.

Den Datensatz können Sie hier runterladen: https://www.kaggle.com/competitions/dogs-vs-cats/data

Für die Aufbereitung der Daten finden Sie einen Code auf Moodle.

Als zweiter Schirtt versuchen Sie den Code auf den Working Stations zum laufen zu bekommen.

Referenzen

- Francois Chollet, Deep Learning with Python, Manning, 2018.
- Mattis Paulin, Julien Mairal, Matthijs Douze, Zaid Harchaoui, Florent Perronnin, Cordelia Schmid, Convolutional Patch Representations for Image Retrieval an Unsupervised Approach, arXiv, 2016.
- Shriyashish Mishra, Object Detection, https://github.com/shriyashish/objectdetection, accessed on 23.03.2023.

Fragen

Darstellung eines Fragesymbol aufgerufen von der Webseite https://www.qnigge.de/news/detail/modul-v/#images am 12.07.2021.

FH
GR

Fachhochschule Graubünden
Pulvermühlestrasse 57
7000 Chur
T +41 81 286 24 24
info@fhgr.ch

Vielen Dank für Ihre Aufmerksamkeit.
Grazia fitg per l'attenziun.
Grazie per l'attenzione.

swissuniversities

SCHWEIZERISCHER AKKREDITIERUNGSRAT
CONSEIL SUISSE D'ACCRÉDITATION
CONSIGLIO SVIZZERO DI ACCREDITAMENTO
SWISS ACCREDITATION COUNCIL

Institutionell akkreditiert nach
HFKG 2018-2025