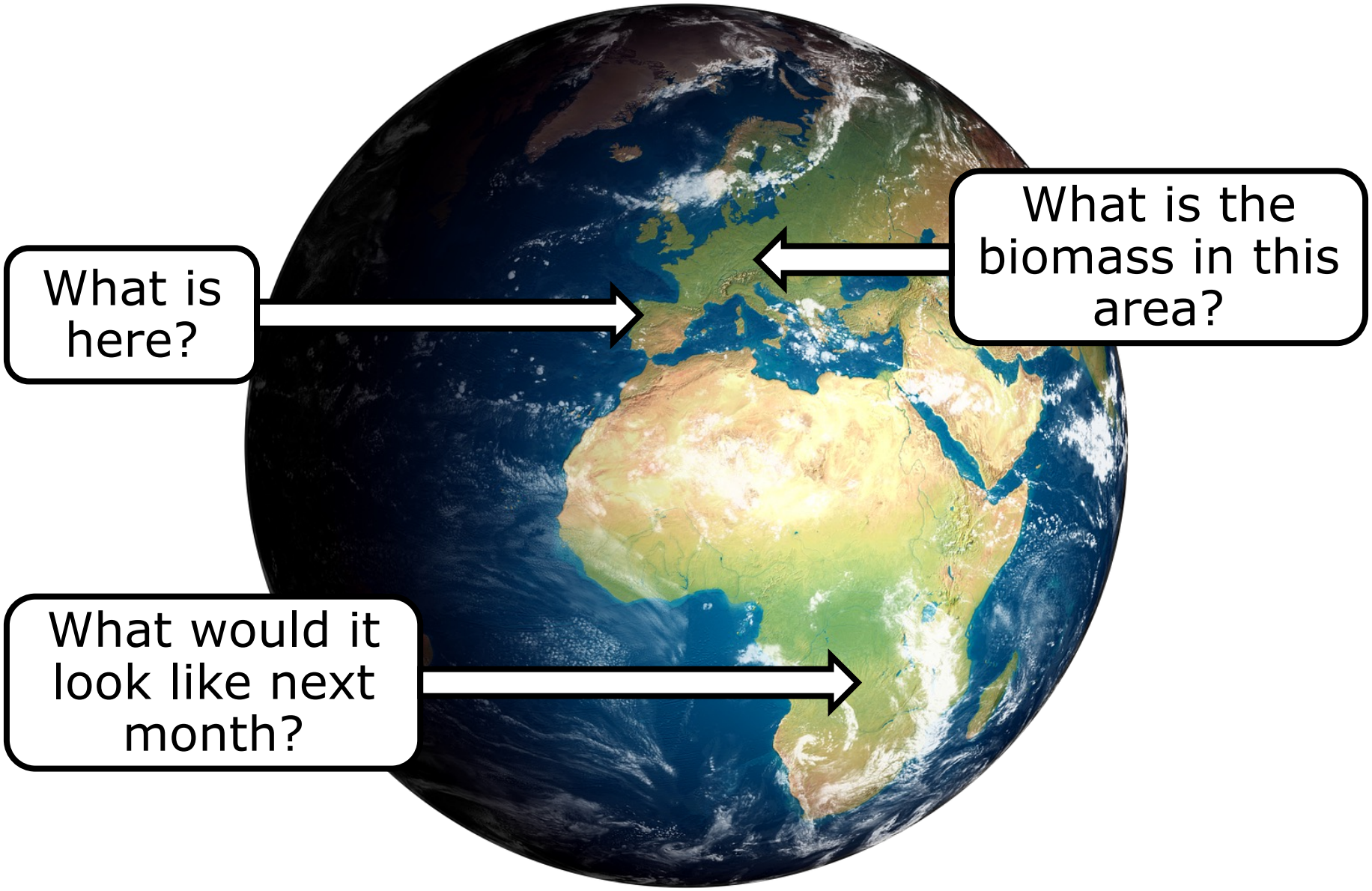
Explainable machine learning with neural networks

Ribana Roscher

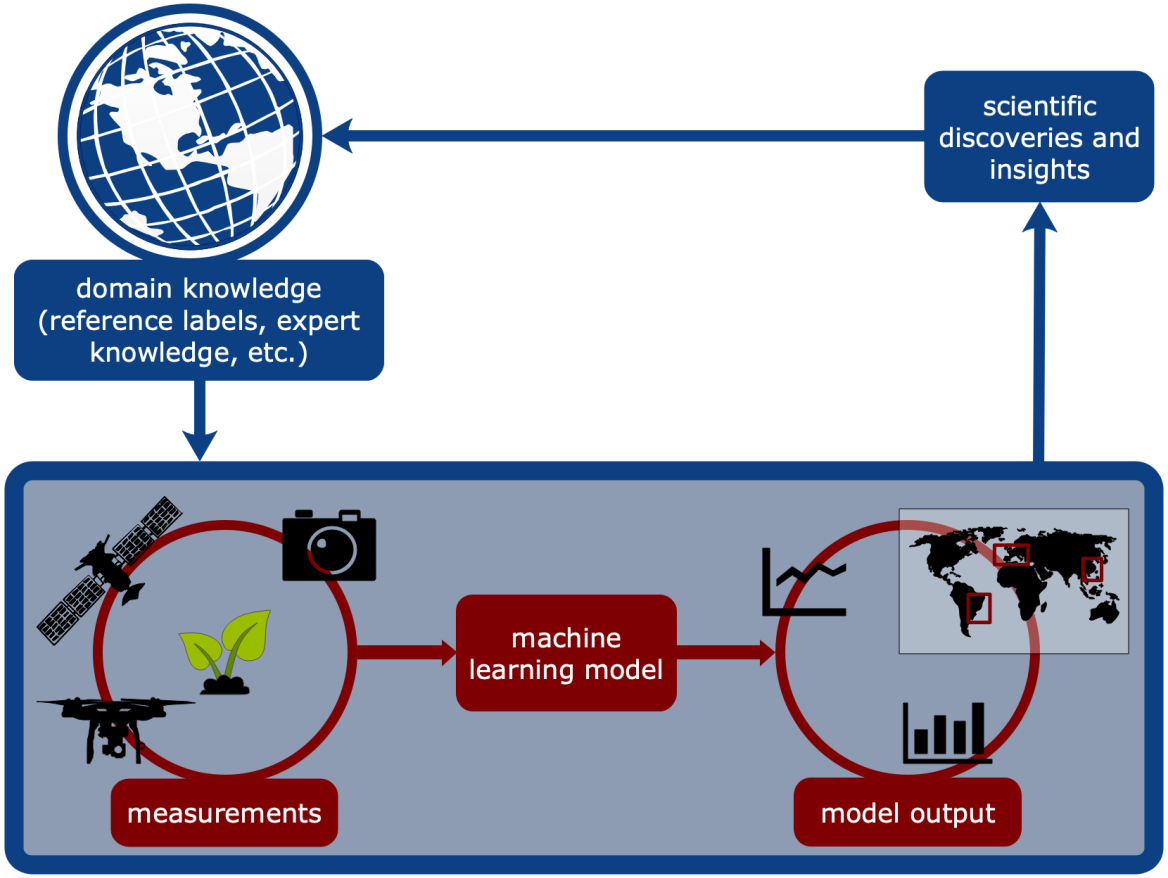
Remote Sensing Group, Universität Bonn

UNIVERSITÄT BONN


What is here?
What is the biomass in this area?
What would it look like next month?

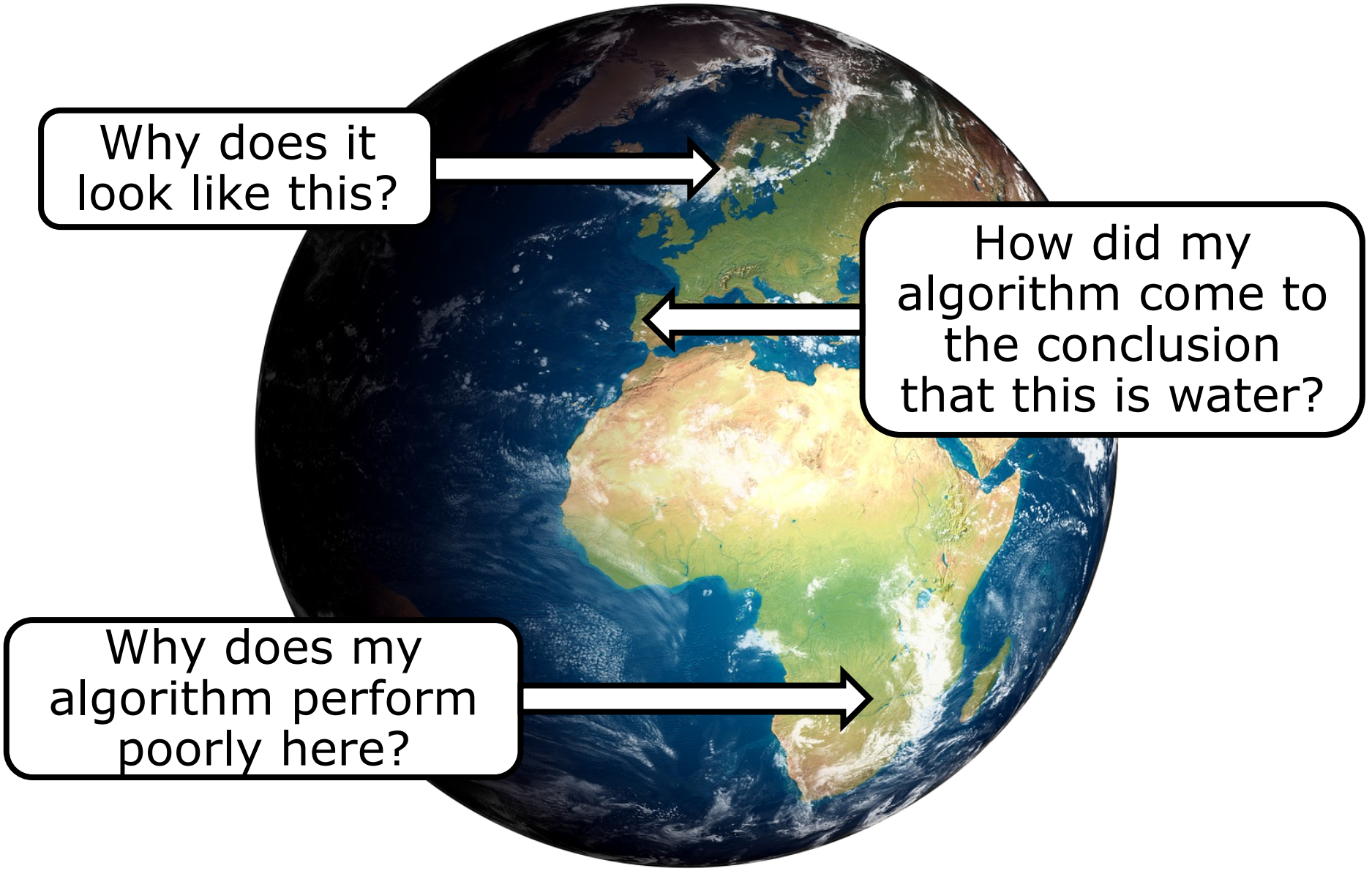
Data science

Field of study that uses scientific approaches to extract knowledge and insights from data



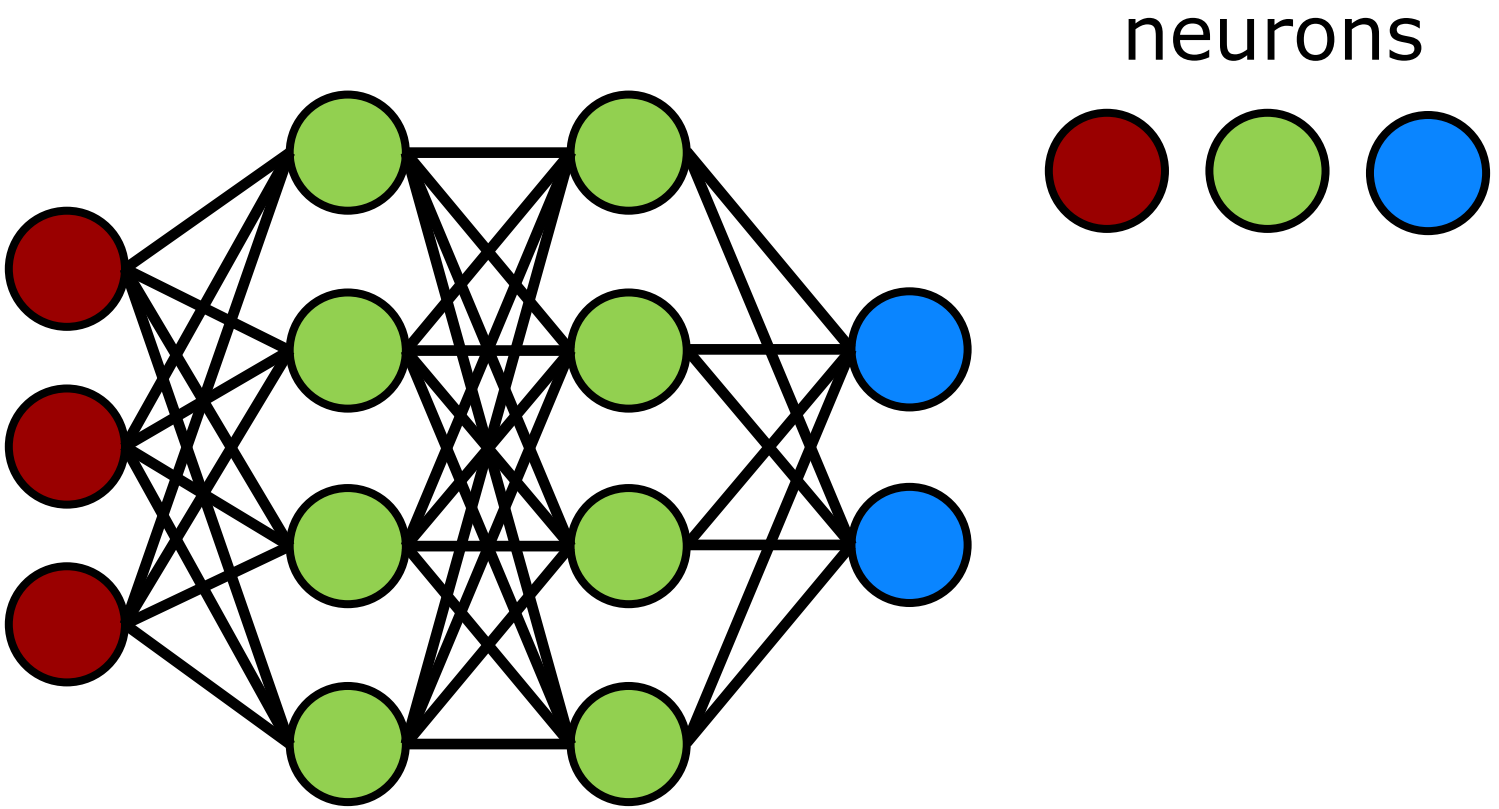
Machine learning

Set of techniques that allow computers to learn from data (e.g., deep learning)


Why does it look like this?
How did my algorithm come to the conclusion that this is water?
Why does my algorithm perform poorly here?

Deep neural networks for image interpretation

Neural networks



Neural networks

Input layer

Observational data

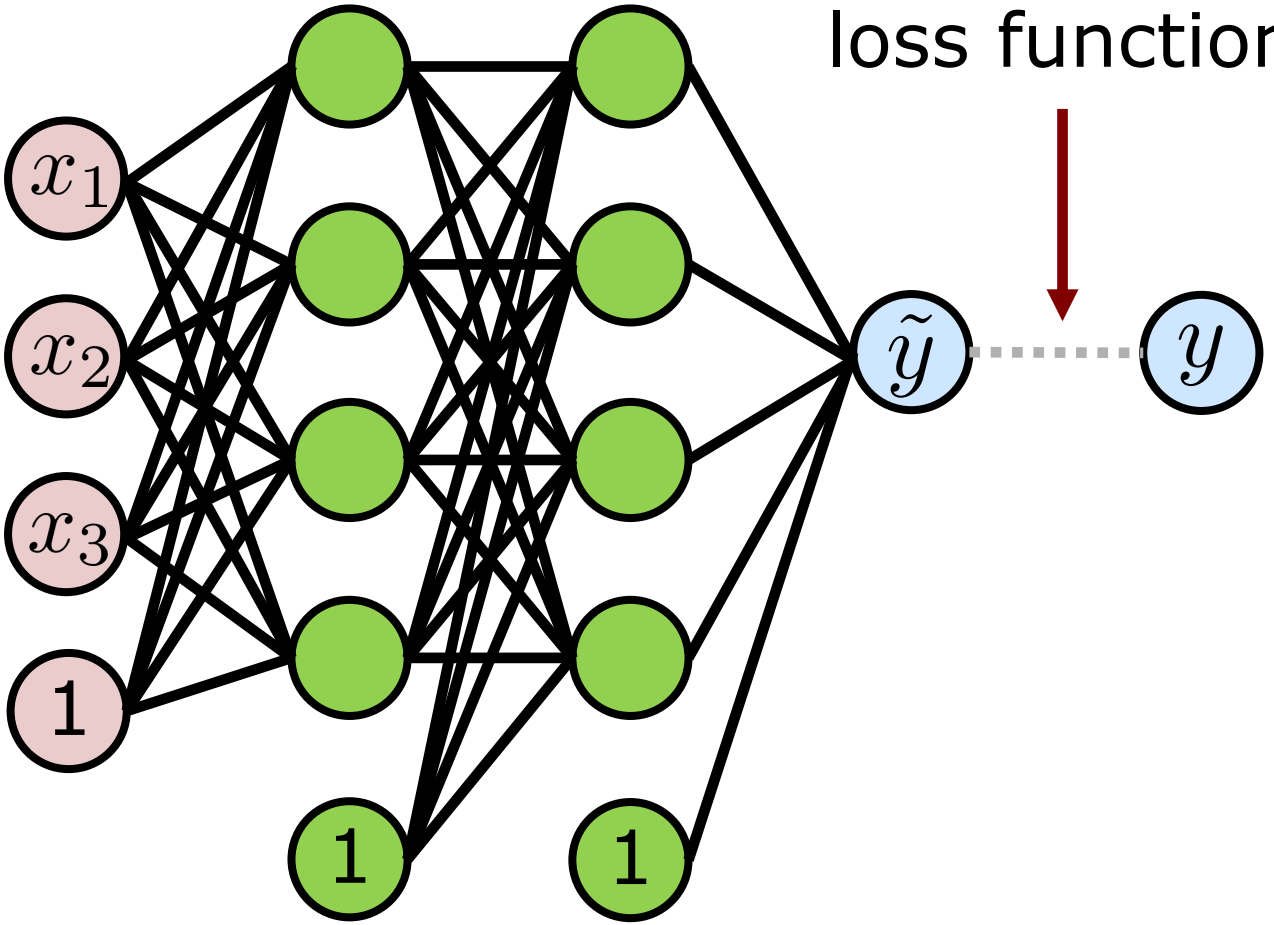
Hidden layer(s)

Layers, which learn a new representation of the data; neurons are functions yielding the representation → **representation learning**

Output Layer

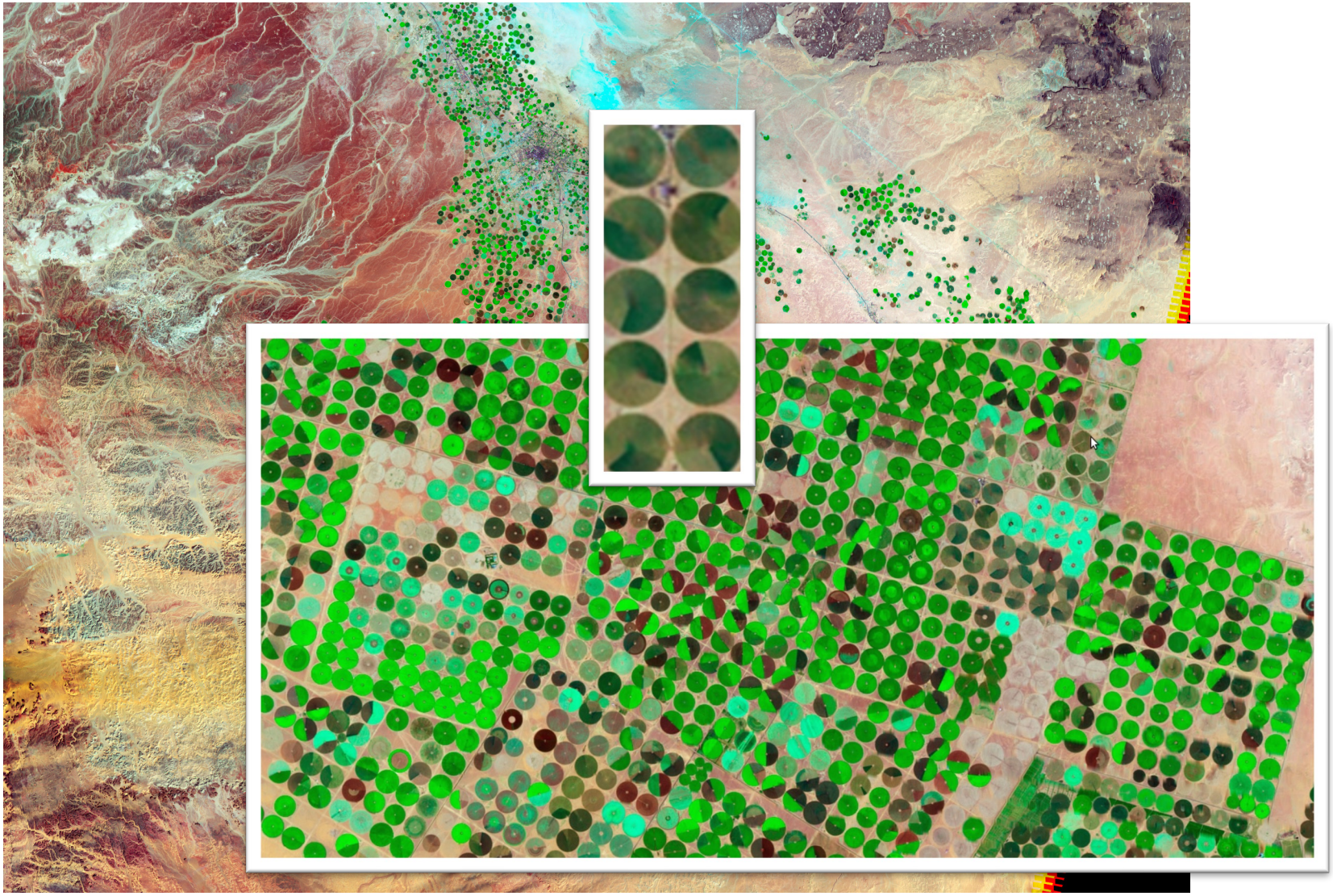
Class labels, estimations, ...

Neural network architecture



$$\tilde{y} = f(\boldsymbol{x}, \boldsymbol{w}) = \mathcal{F}_1 \left(W_1^{\mathsf{T}} \cdot \mathcal{F}_0 \left(W_0^{\mathsf{T}} \cdot \boldsymbol{x} \right) \right)$$

Convolutional neural networks



Basic principle

CNN

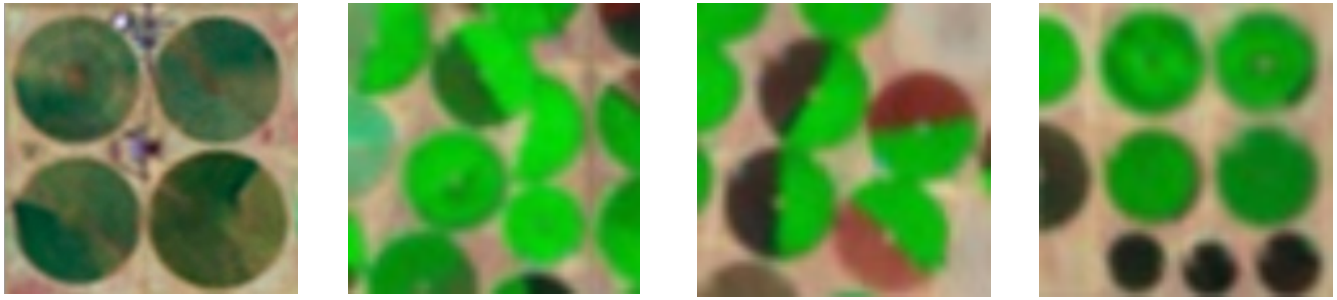
class 1 (agricultural area)

or

class 2 (desert)

Basic principle

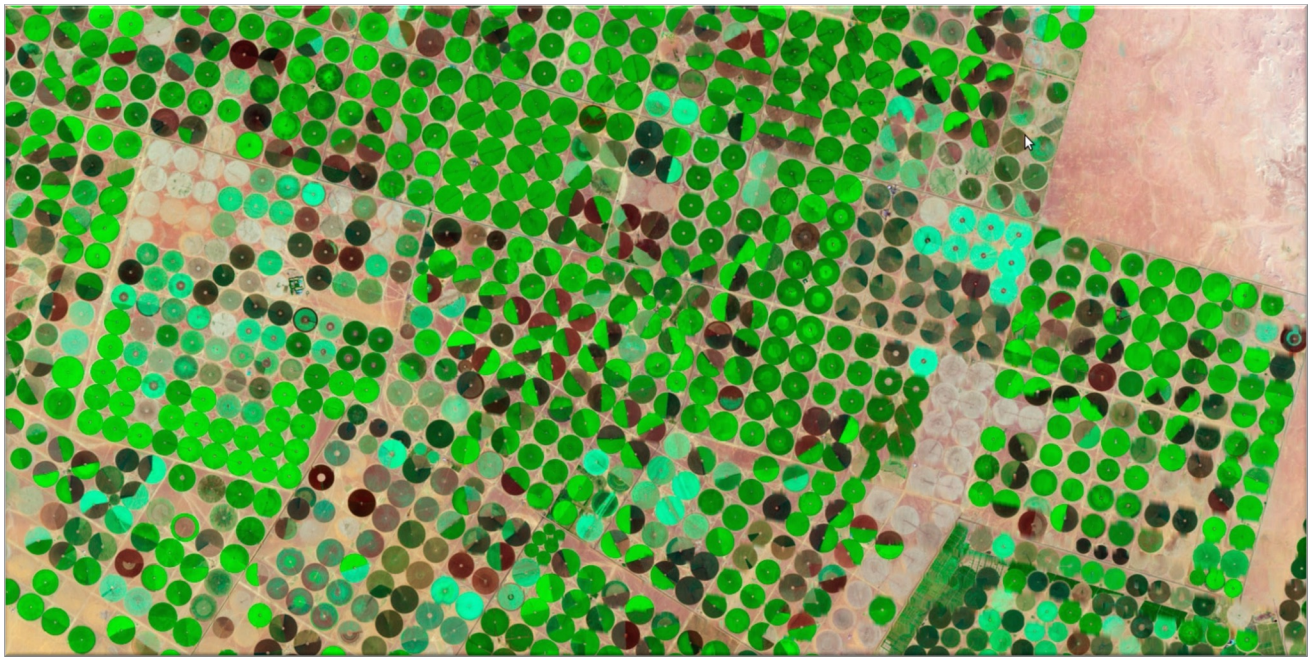
Objects can be translated, scaled, rotated, weighted, ...



Contain similar **structures**

Basic principle

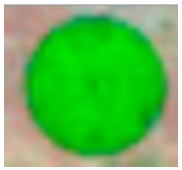
- Find similar parts, which characterize the image



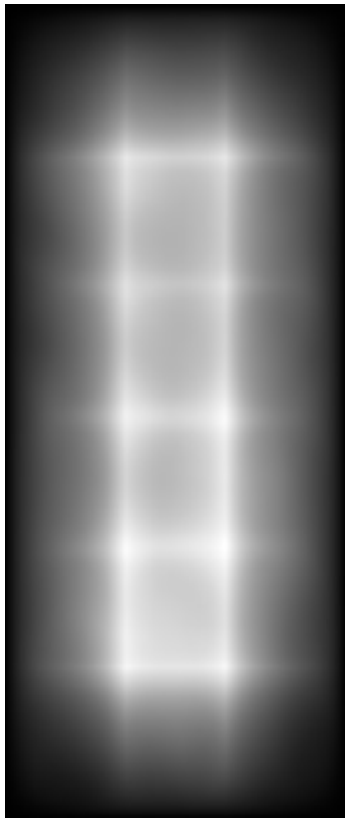
- Math behind: filtering/**convolution**

Convolutional layer

filter



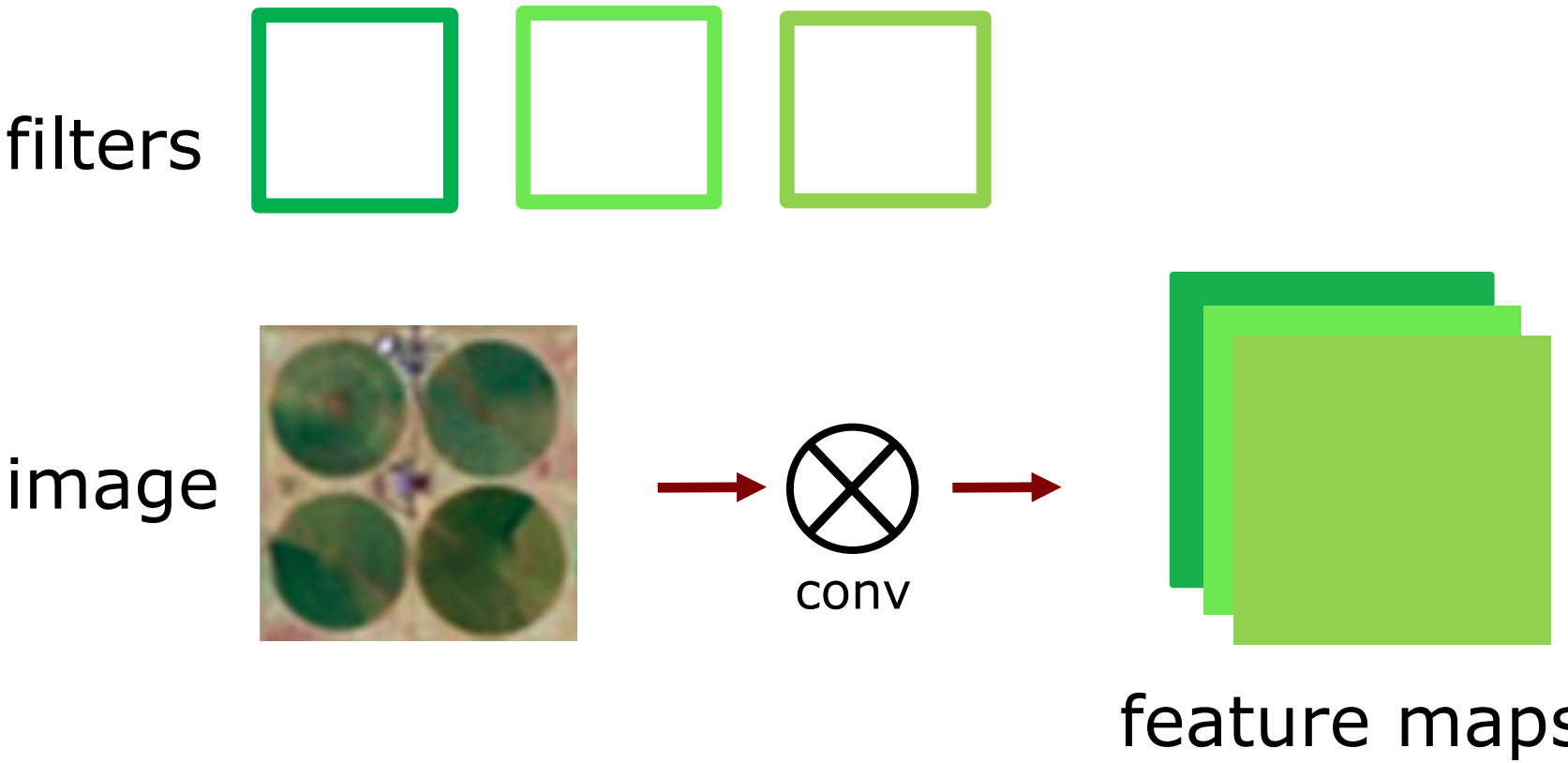
image



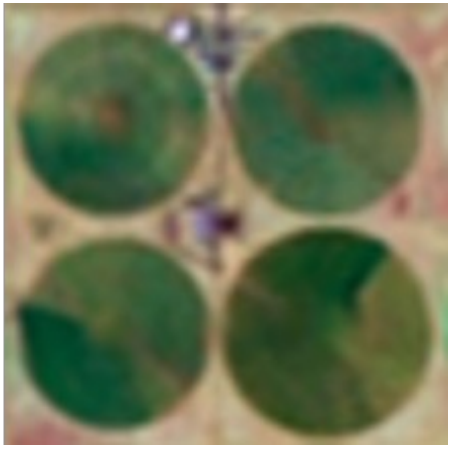
feature map (activation map)

Convolutional layer

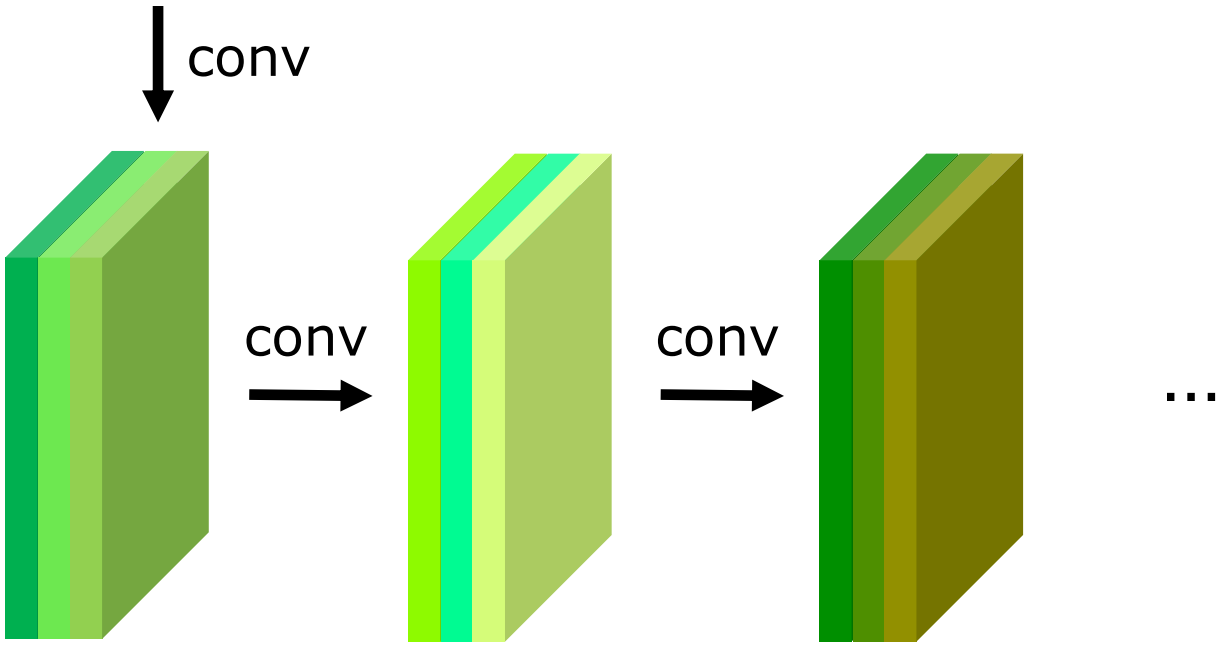
- One image becomes a stack of filtered images



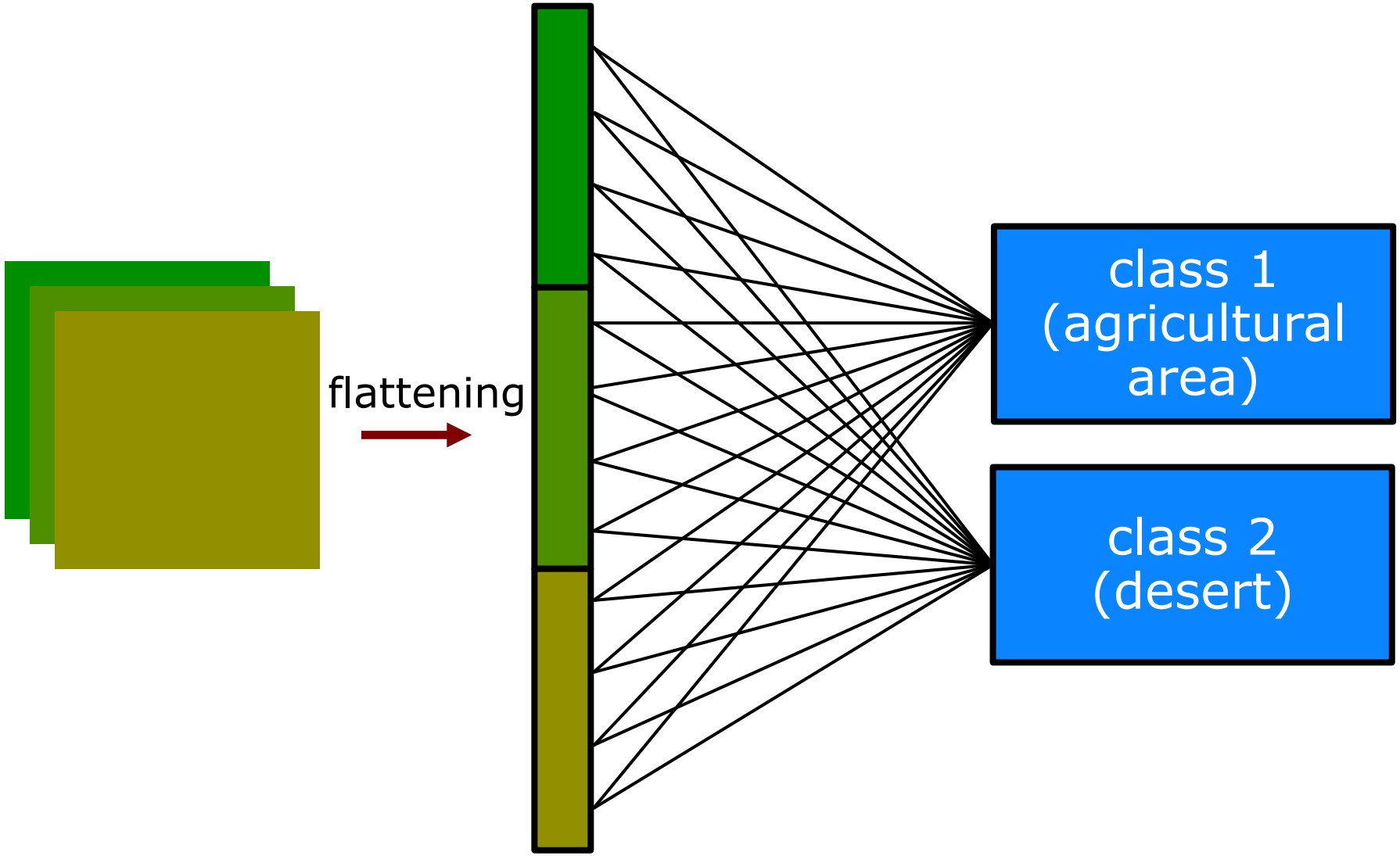
Convolutional neural network



CNNs contain **multiple convolutional layers**



Fully connected layer

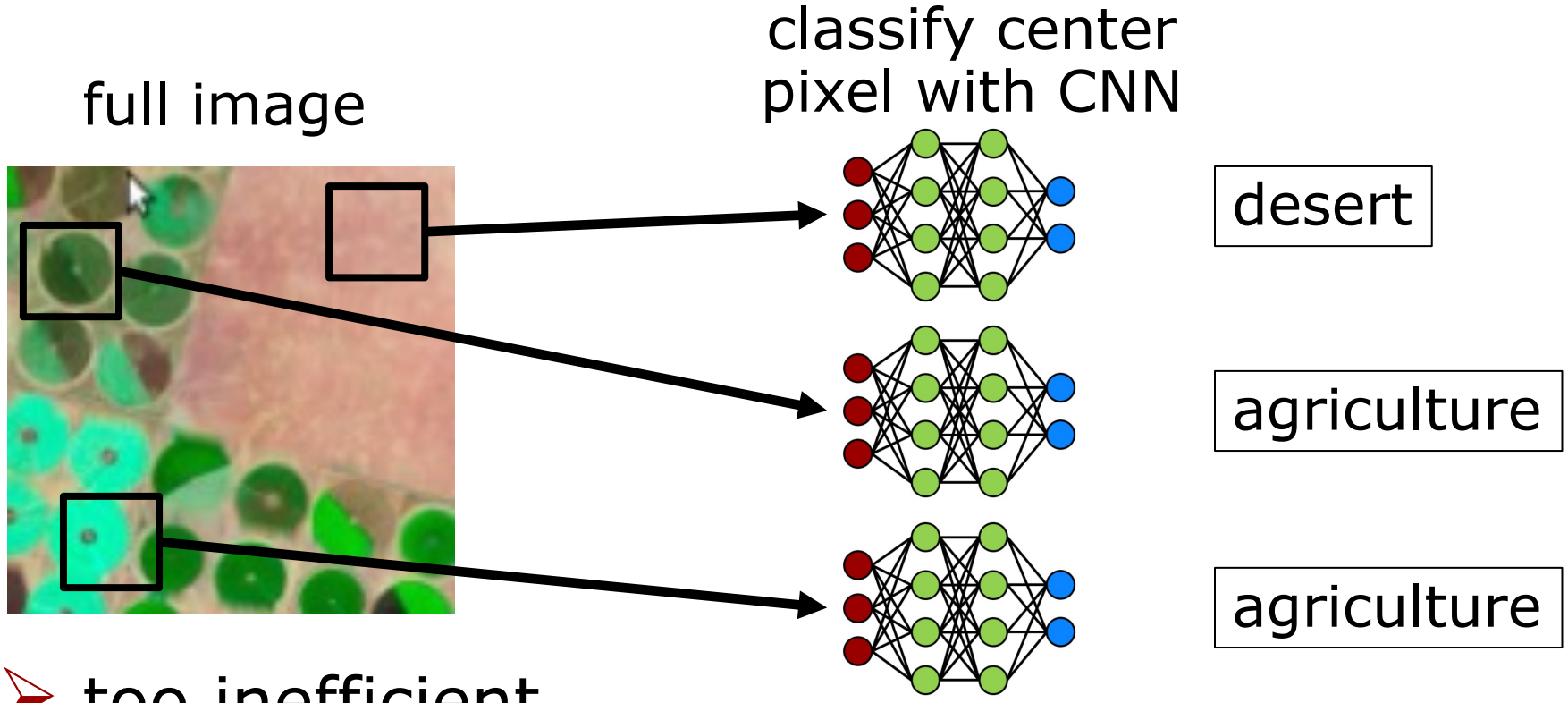


Each feature value becomes a list of votes

Image segmentation

- How can you pixelwise segment the image (instead of categorizing it)?

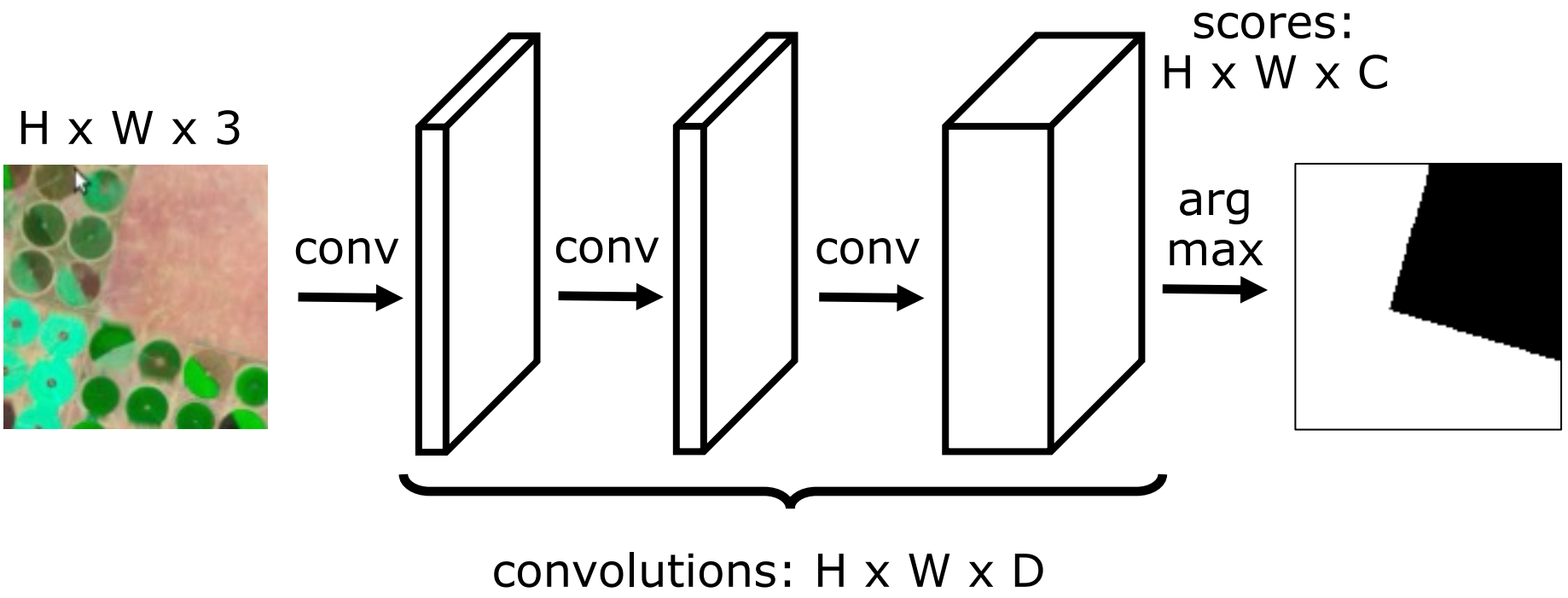
1. possibility: sliding window categorization



➢ too inefficient

Image segmentation

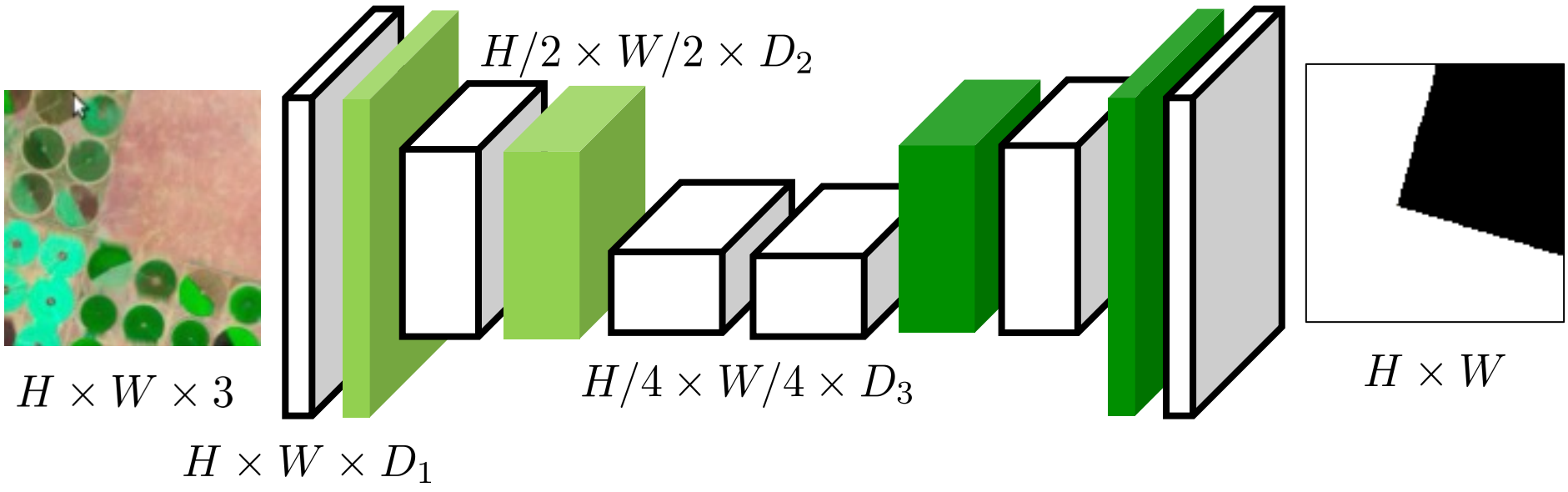
2. possibility: bunch of convolutional layers to make predictions for pixels all at once



- Each convolutional layer is as big as the input
- Computationally expensive

Image segmentation

3. possibility: downsampling and upsampling



Why downsampling?

- Reduces the complexity in the network
- If the convolutional filter size stays the same, the downsampling layers cause an **increase in the receptive field** in deeper layers

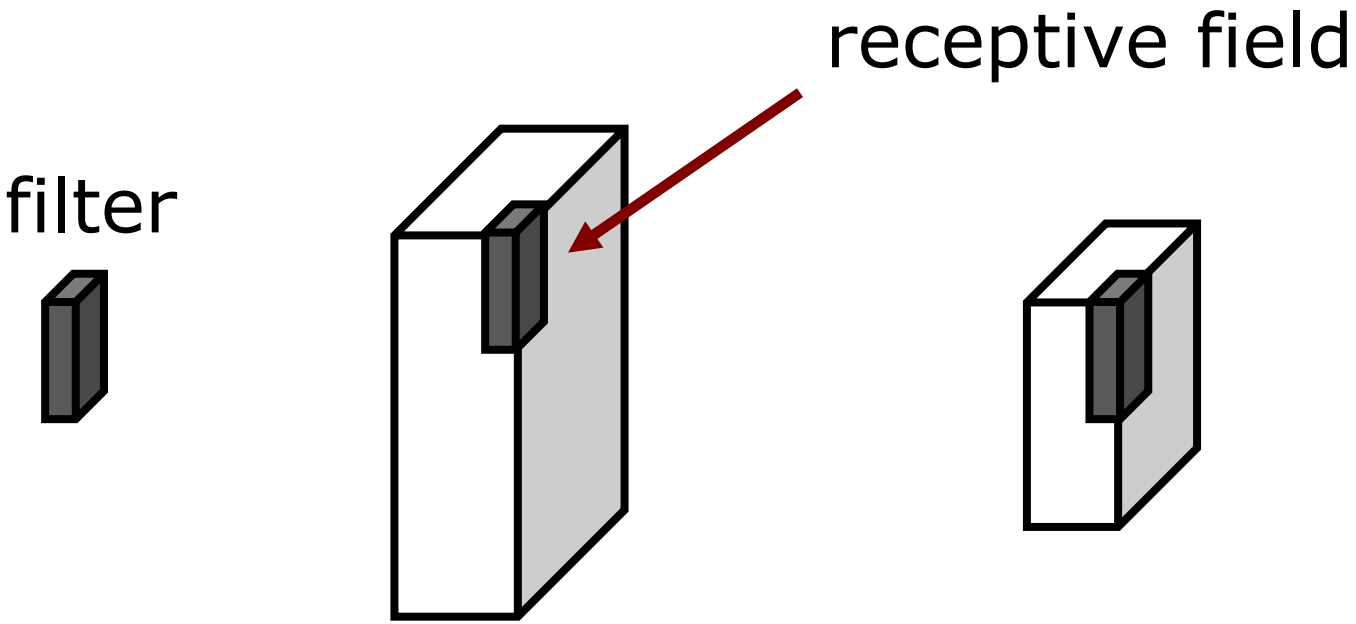
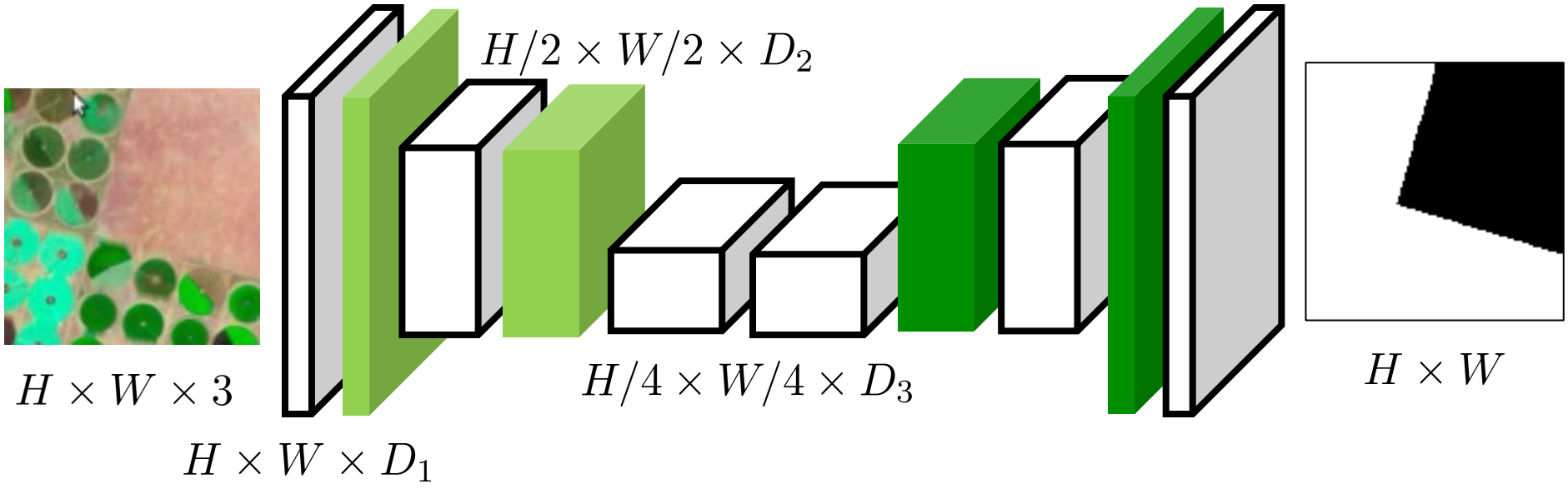


Image segmentation

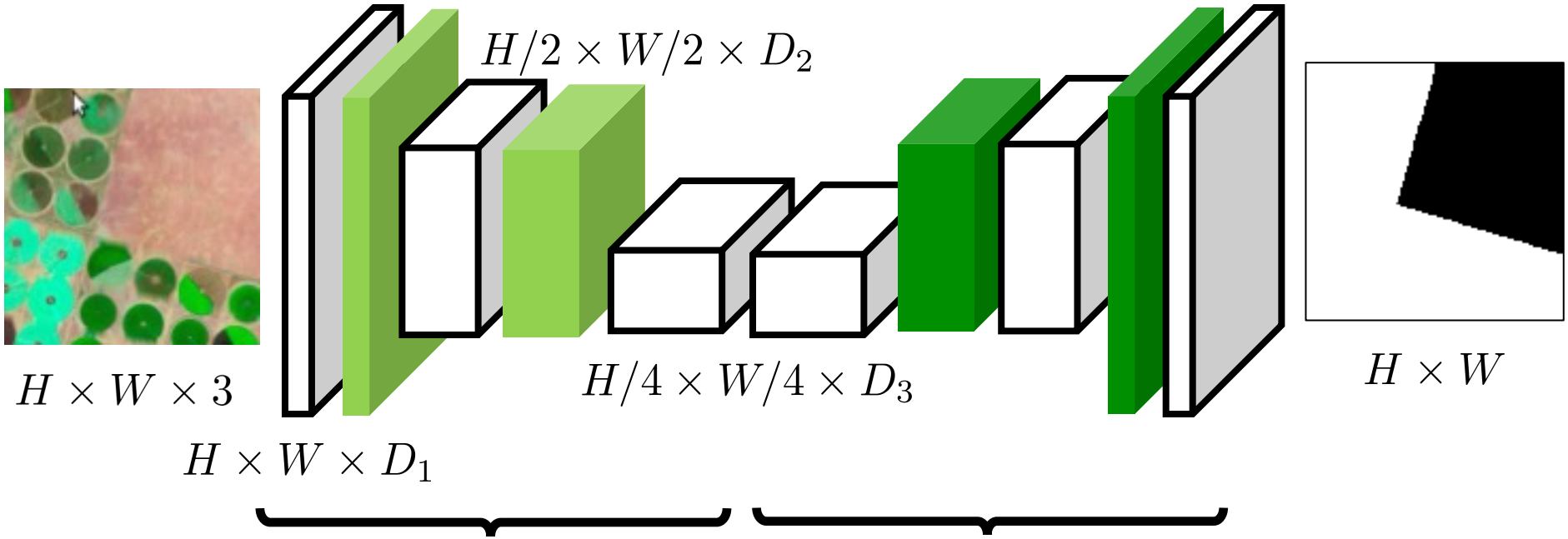


Generally, the number of filters increase with depth (in the encoder), because:

- lower layers with too many filters are too computationally intense
- the correlation between filters decreases with depth

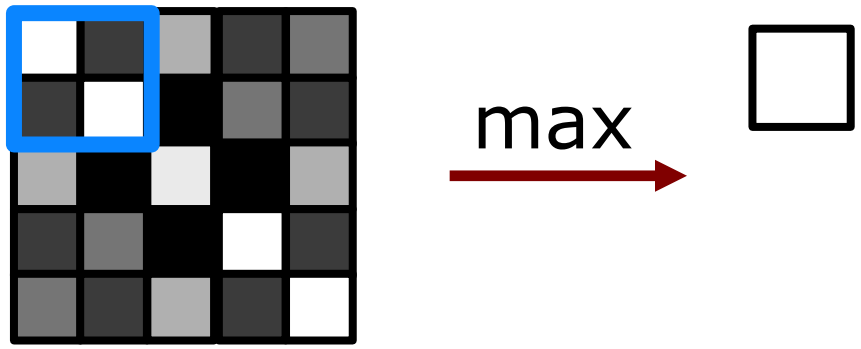
Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1520-1528).

Image segmentation network



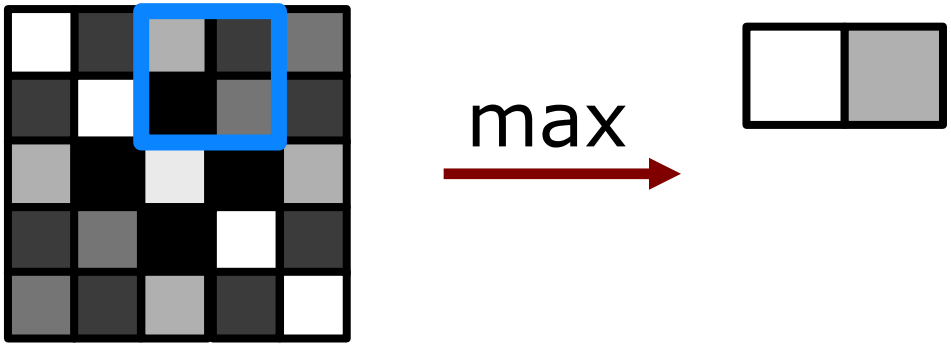
Downsampling: pooling

- Shrinking the image stack
- Commonly used: **Max-pooling** and mean-pooling
- Pooling layer defined by stride (2) and window size (2x2)



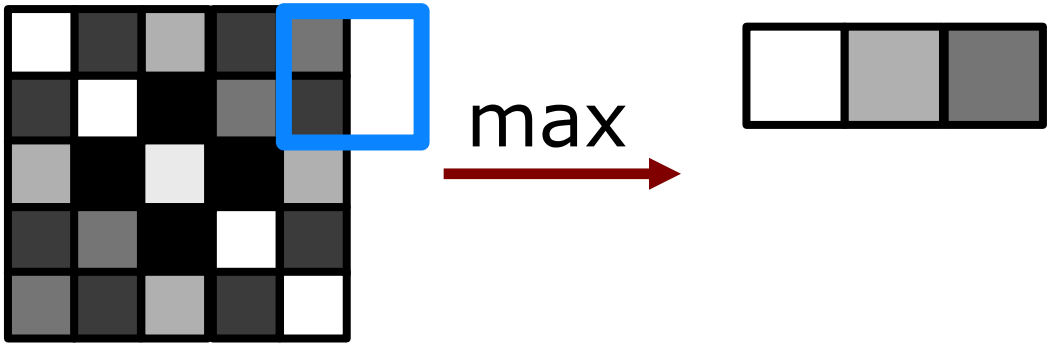
Downsampling: pooling

- Shrinking the image stack
- Commonly used: **Max-pooling** and mean-pooling
- Pooling layer defined by stride (2) and window size (2x2)



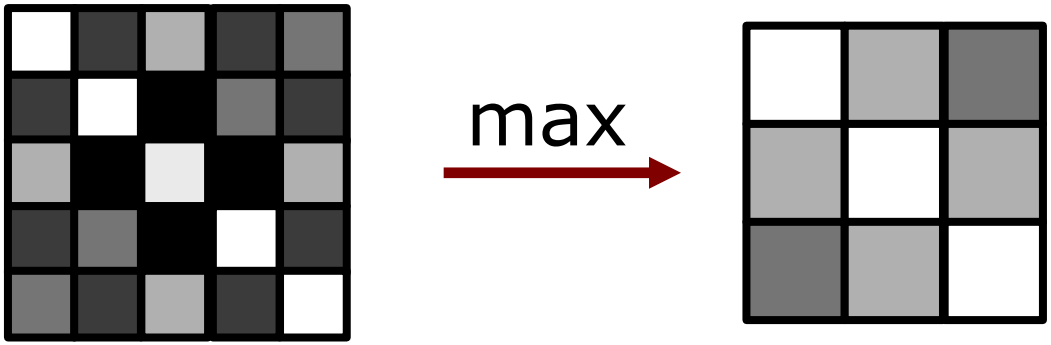
Downsampling: pooling

- Shrinking the image stack
- Commonly used: **Max-pooling** and mean-pooling
- Pooling layer defined by stride (2) and window size (2x2)



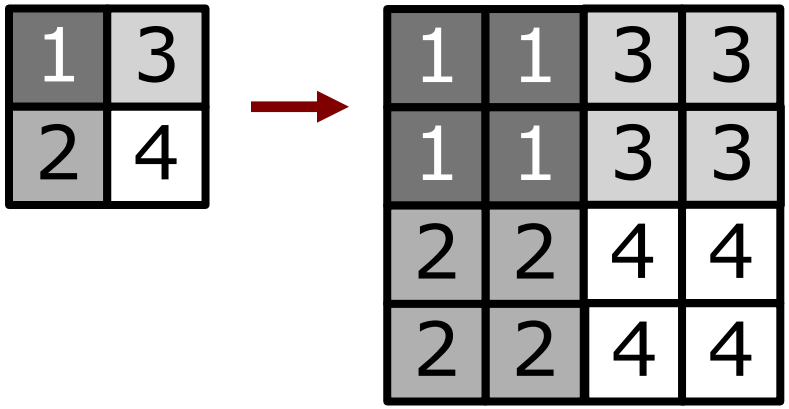
Downsampling: pooling

- Shrinking the image stack
- Commonly used: **Max-pooling** and mean-pooling
- Pooling layer defined by stride (2) and window size (2x2)

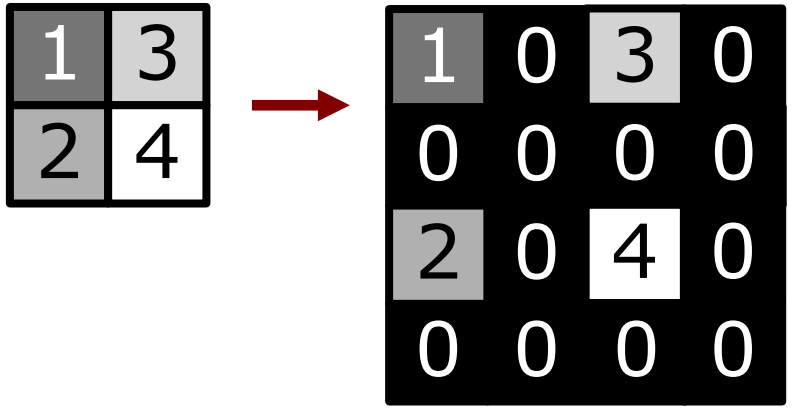


Upsampling: unpooling

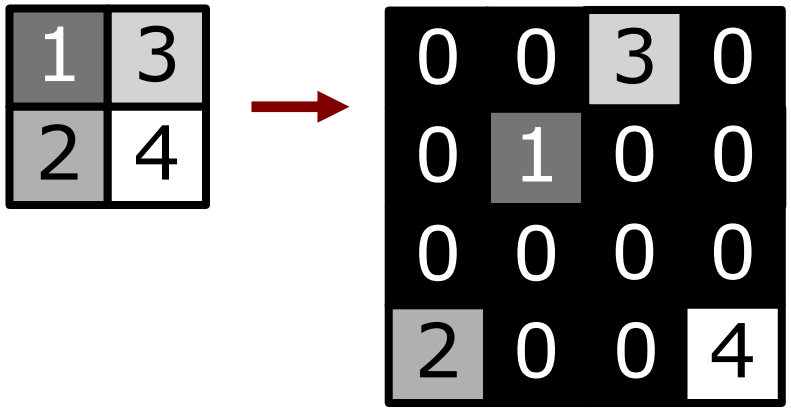
Nearest neighbor



Bed of nails



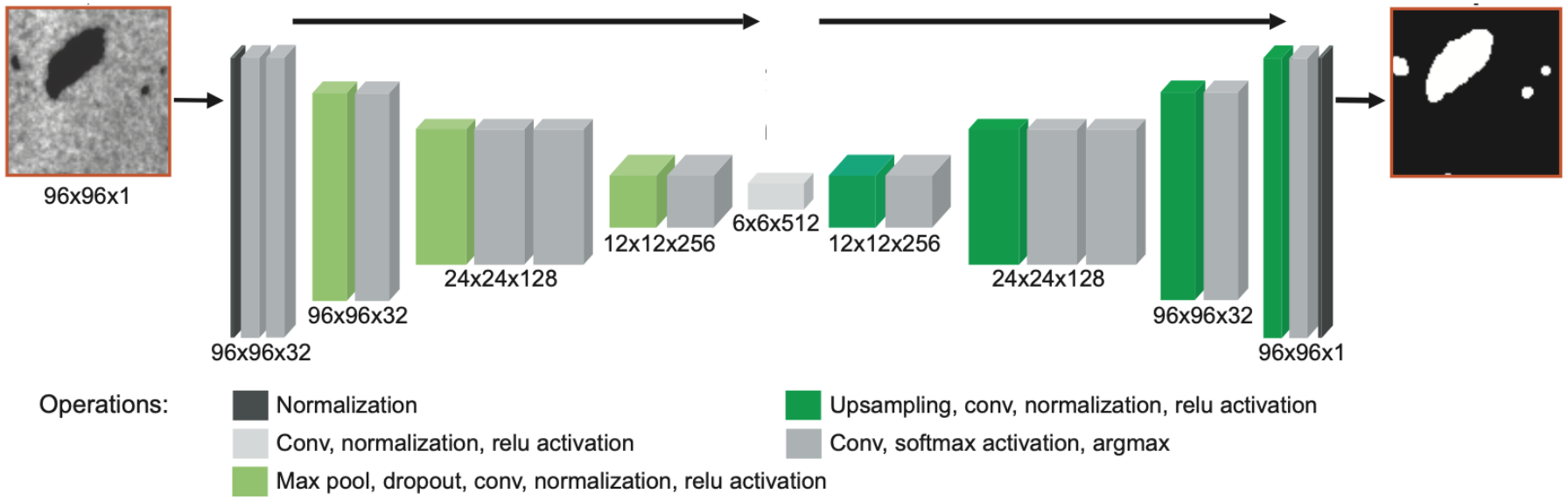
Max unpooling



Positions of max values in downsampling layers are stored

Stacked layers

CNNs oftentimes consist of a sequence of alternating convolutional layers and pooling layers



Learning the parameters

What needs to be learned:

- Filters
- Weights in fully connected/dense layers
- But: Pooling layers have no parameters

How?

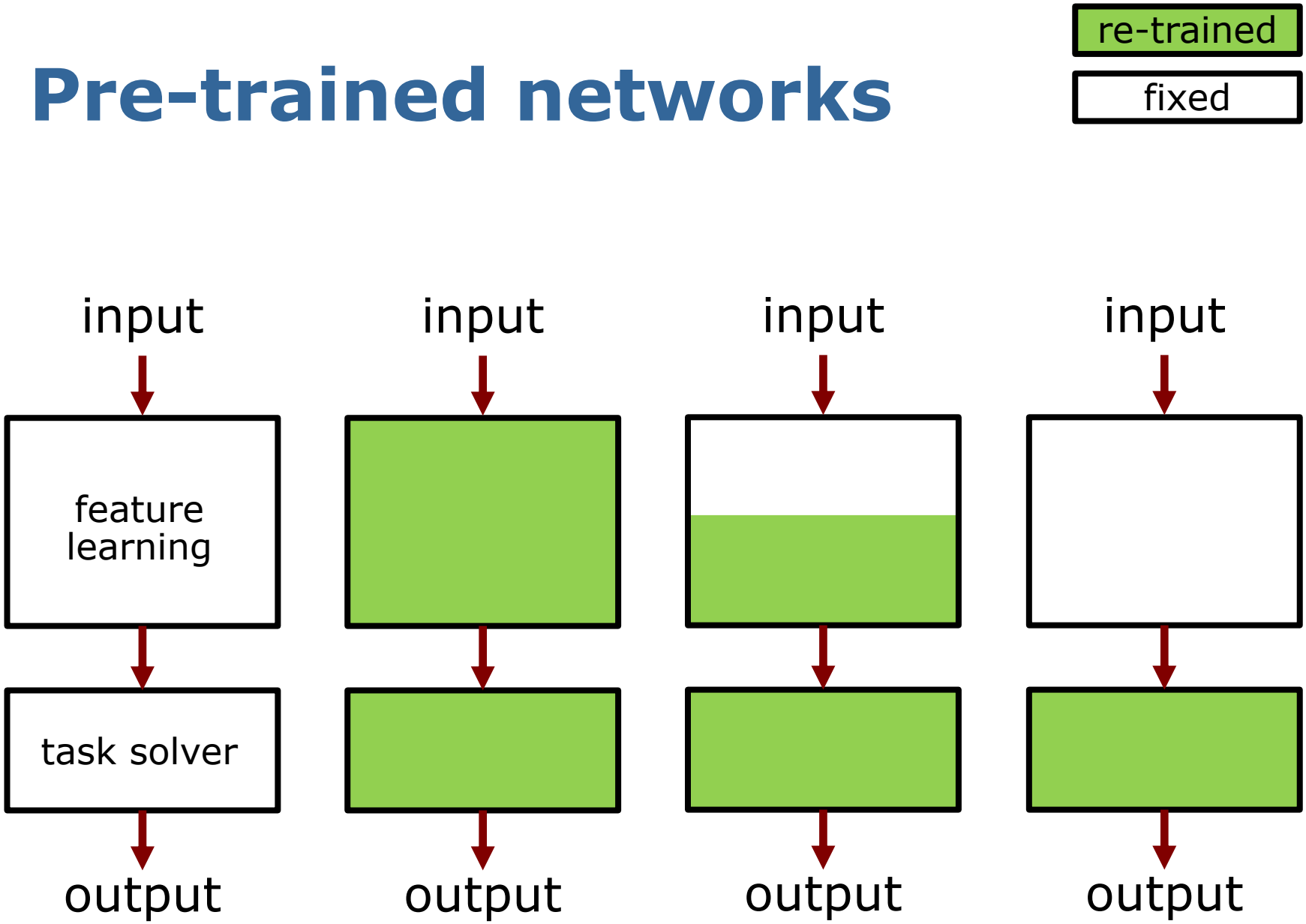
backpropagation

Learning the parameters

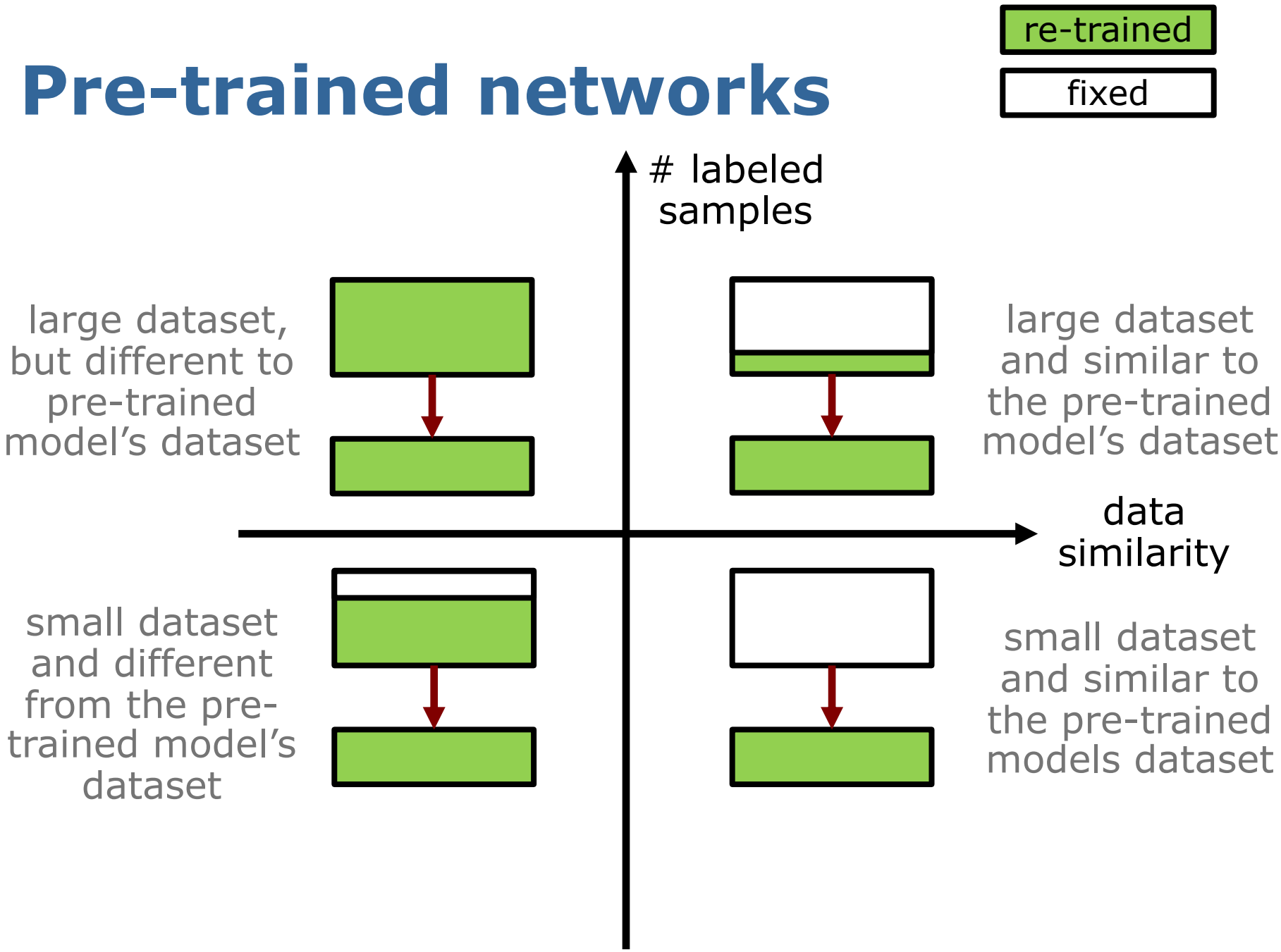
Hyperparameters which need to be set:

- Convolutions
  - Number of filters
  - Size of filters
  - (Window stride)
- Pooling
  - Window size
  - Window stride

Pre-trained networks



Pre-trained networks



Summary

- Convolutional neural networks are suitable to extract structural features (in time and space)
- Commonly used for gridded data, but can also be used for graphs (graph-CNNs)
- Convolutions can be combined with other networks, e.g. convolutional autoencoder