HSLU Lucerne University of Applied Sciences and Arts

Convolutional Neural Networks for Semantic Segmentation

10. November 2023

FH Zentralschweiz

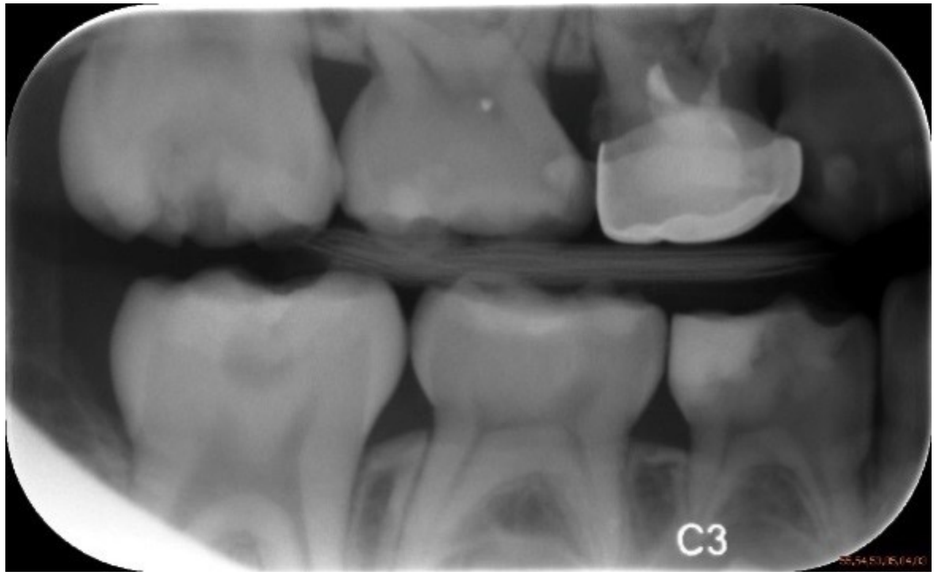
Semantic Segmentation

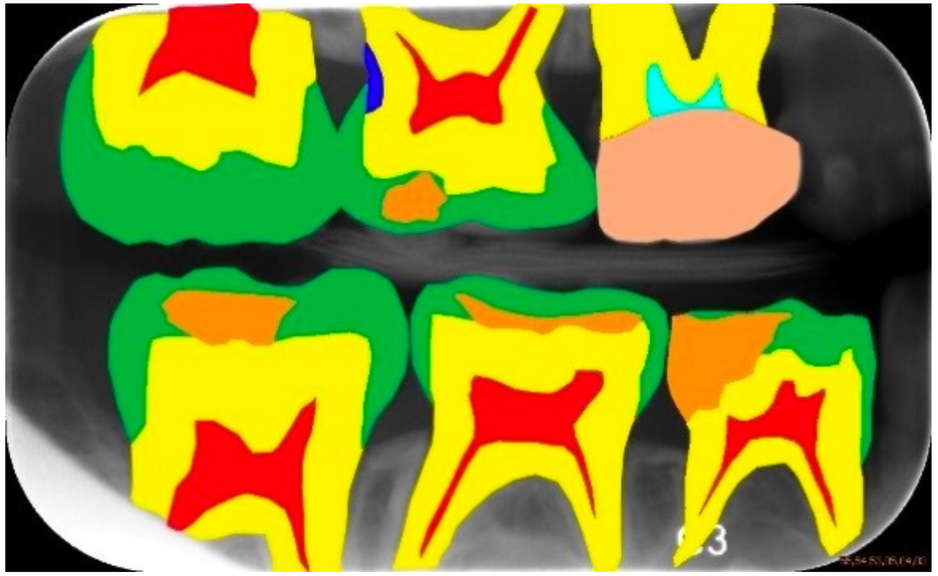
Image classification: classify entire images

Semantic segmentation: **classify each pixel**

Example: Medical Imaging

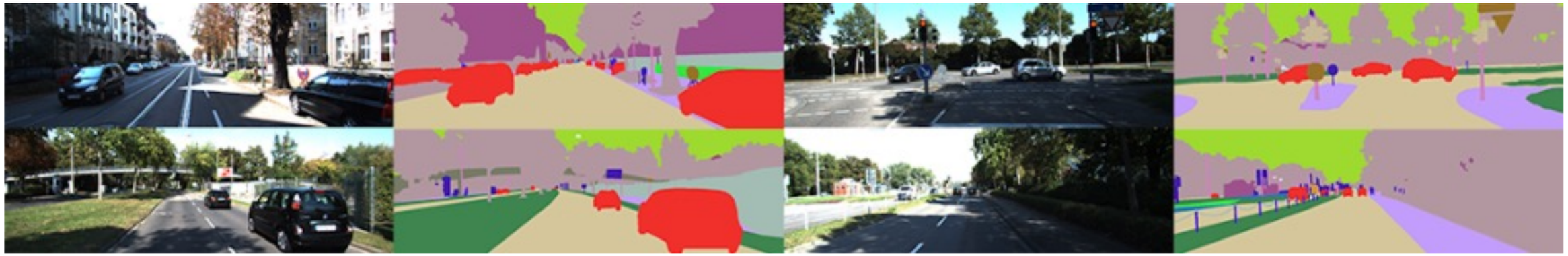
ISBI 2015 Challenge on Dental X-Ray Analysis





Example: Eczema Detection

Example: Autonomous vehicles



http://www.cvlibs.net/datasets/kitti/eval_semseg.php?benchmark=semantics2015

KITTI Dataset

200 train and 200 test images

36 classes

Cityscale Dataset

Fine annotations: 5000 images

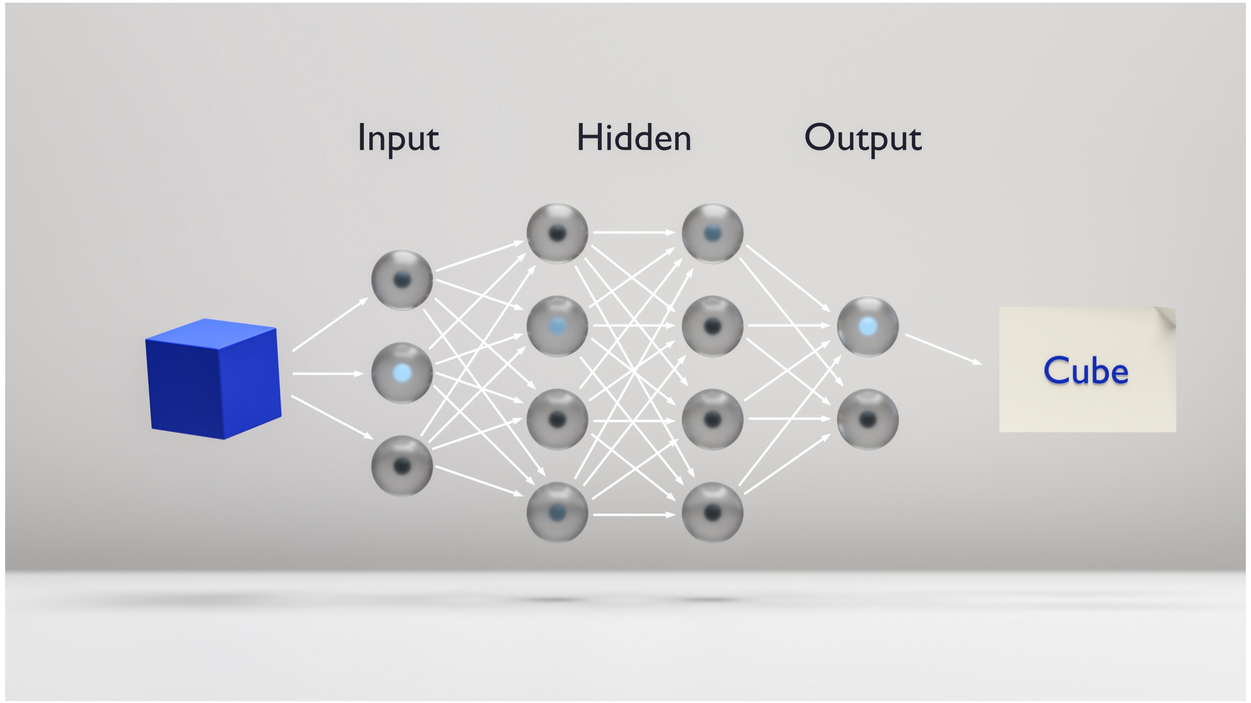
Coarse annotations

CNNs for semantic segmentation

What are the building blocks of a convolutional neural network (CNN) ?

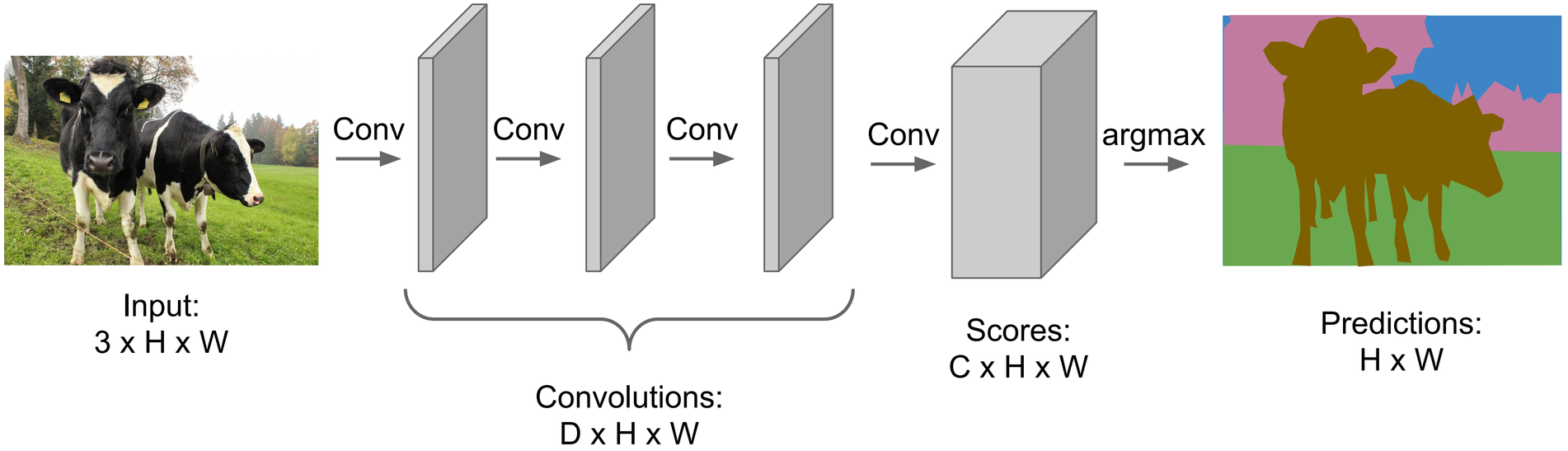
How does a CNN for an image classification task look like?

What happens to the image while it traverses the network?



Fully Convolutional Neural Networks: Idea

Design a network as a bunch of convolutional layers
to make predictions for pixels all at once!

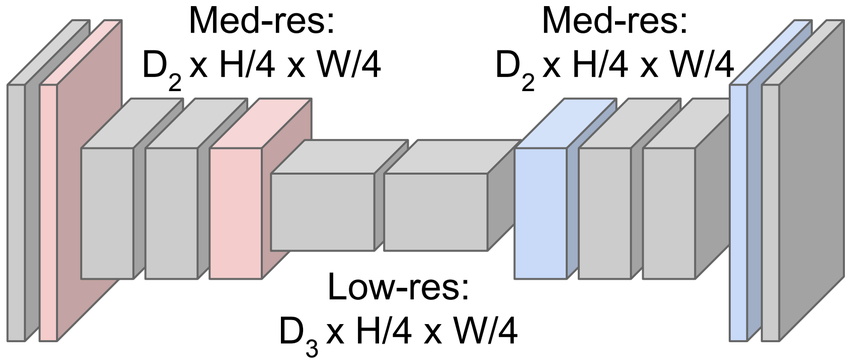


Problem: Might need to be very deep or with large convolutional filters to have a sufficiently large receptive field

FCNs with downsampling and upsampling

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Input:
3 x H x W



High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

Predictions:
H x W

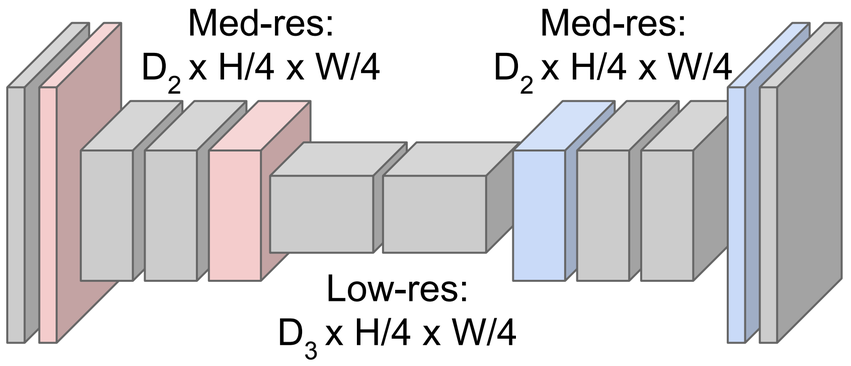
FCNs with downsampling and upsampling

Downsampling:
Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling:
???

Input:
3 x H x W

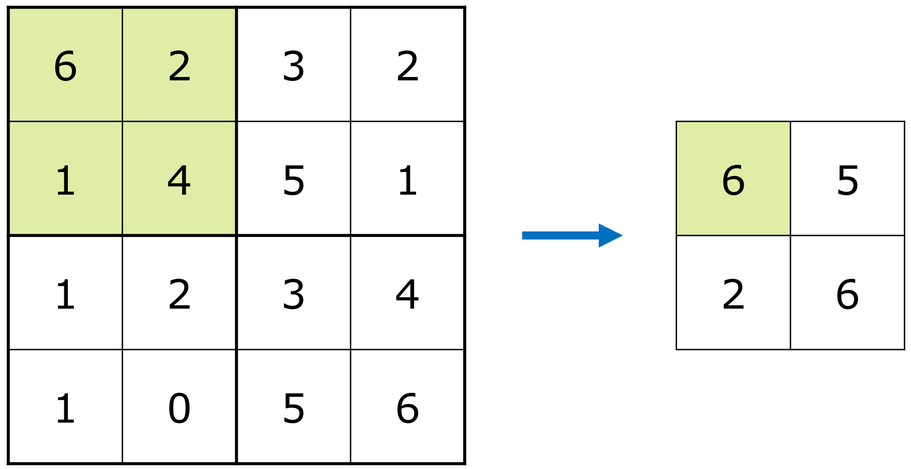


Predictions:
H x W

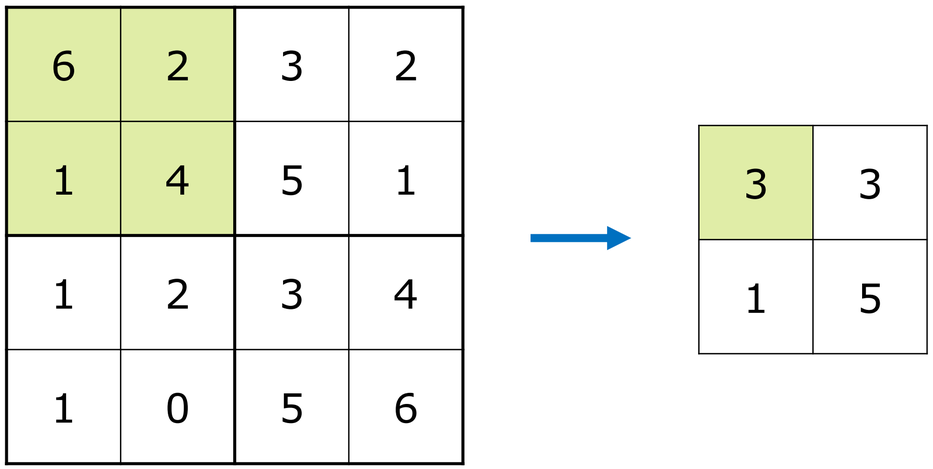
Problem: How to do the upsampling

Remember: Pooling (Max/Avg)

Max Pooling

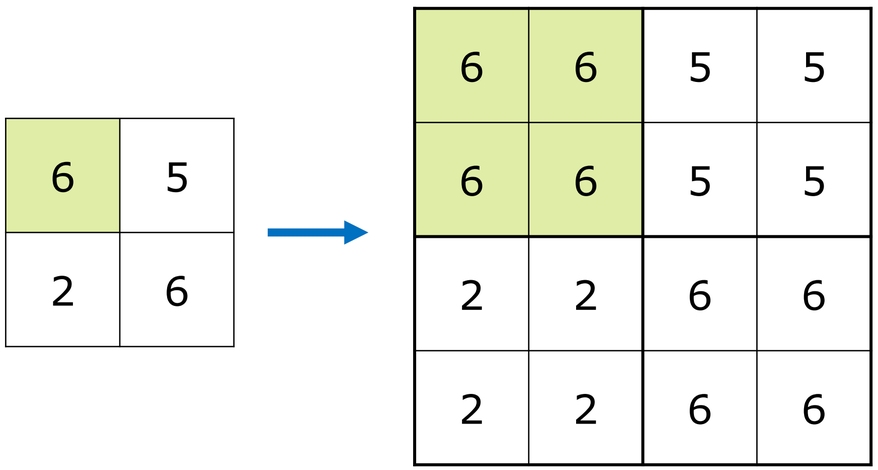


Average Pooling

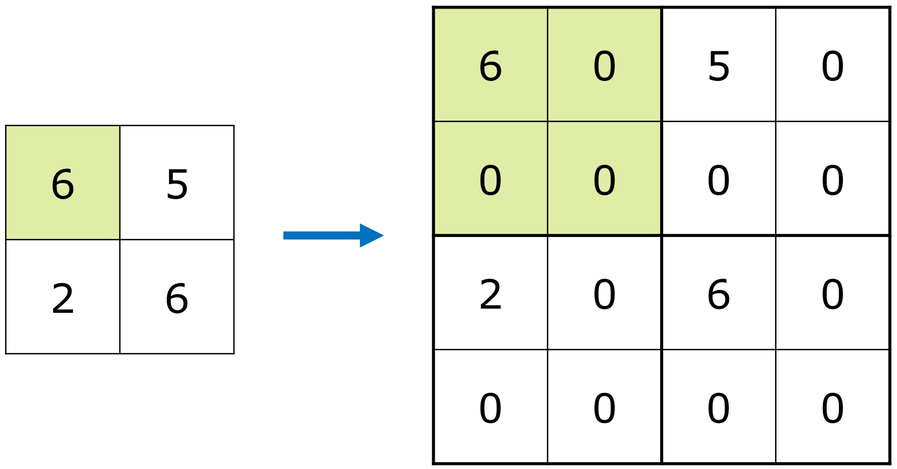


Upsampling: Unpooling

Nearest Neighbor



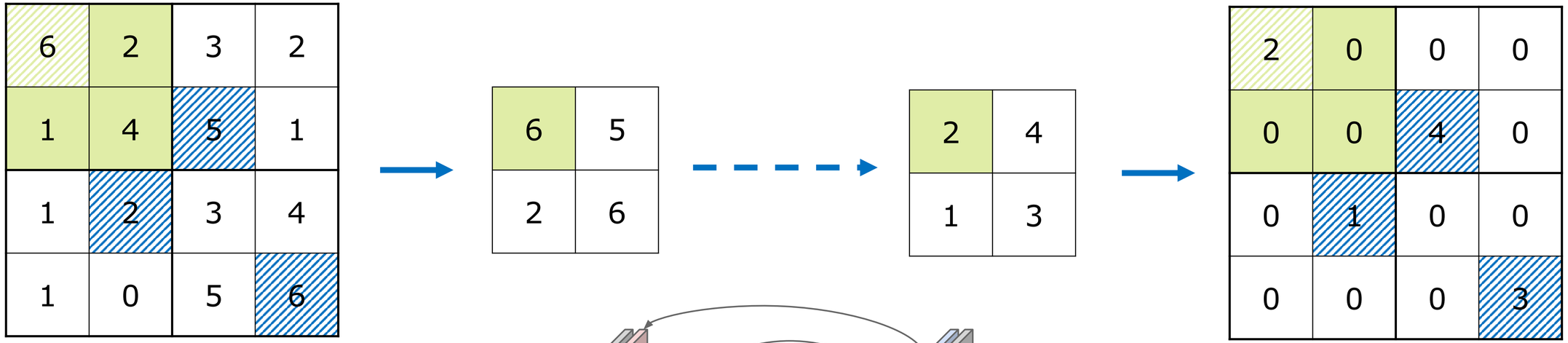
Bed of Nails

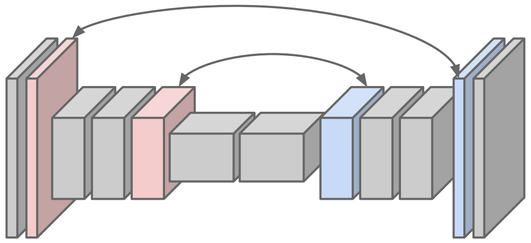


Upsampling: Max Unpooling

Max Pooling: Remember the position of the maximum

Max Unpooling: Put value at the correct position

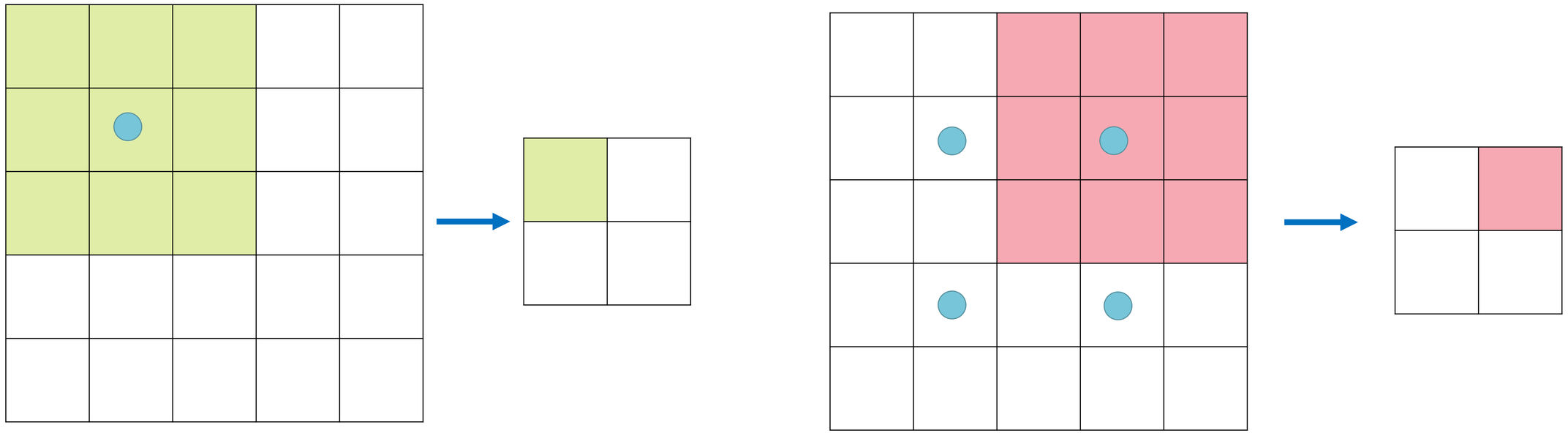




Corresponding layers

Recall: Strided Convolution

Strided 3x3 convolution with stride = 2

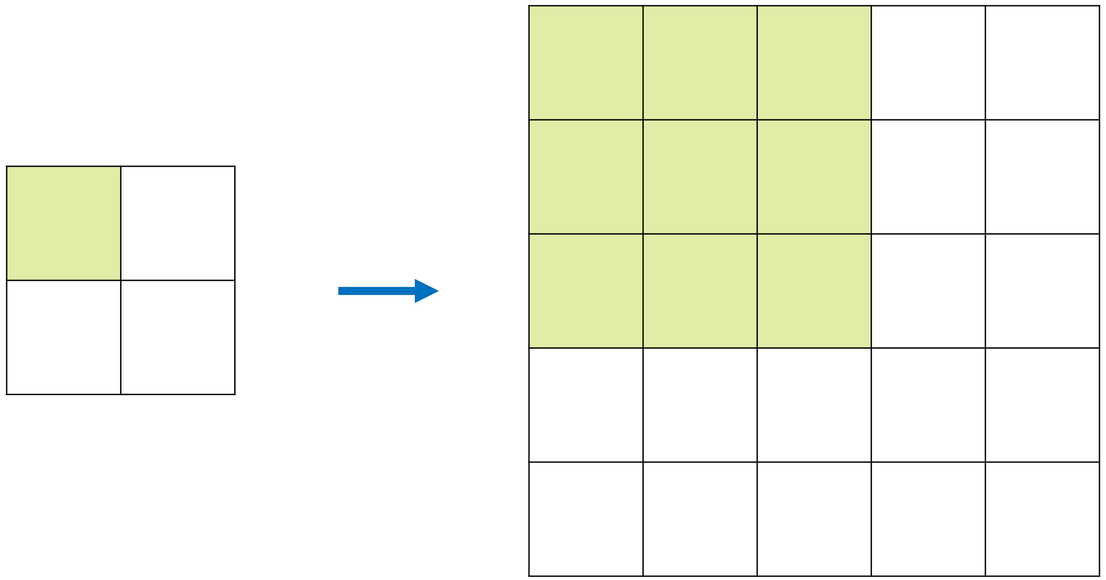


Filter moves 2 positions in **input** for every position in **output**

Transposed convolution

Strided 3x3 transposed convolution with stride = 2

- Multiply input values with filter values
- Add result to output
- Learn filter values through training

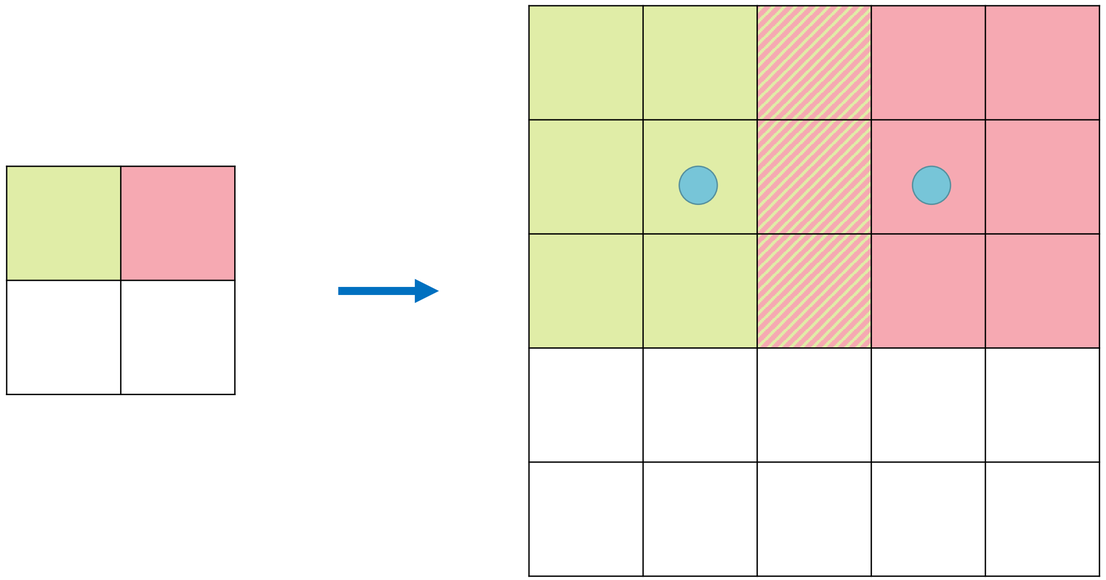


Filter moves 2 positions in **output** for every poition in **input**

Transposed convolution

Strided 3x3 transposed convolution with stride = 2

- Multiply input value with filter values
- Add result to output
- Learn filter values



Summation of values where filter outputs overlap

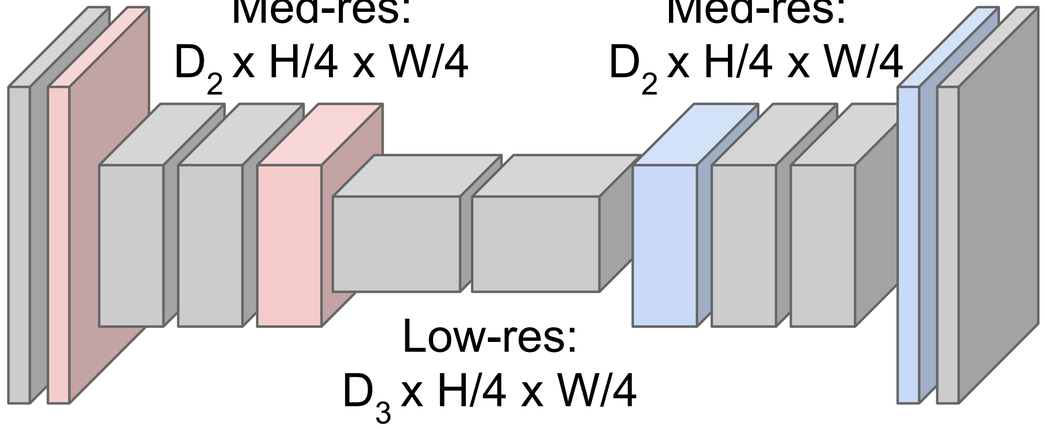
FCS with downsampling and upsampling

Downsampling:
Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling:
Unpooling or strided transpose convolution

Input:
3 x H x W



High-res:
D_1 x H/2 x W/2

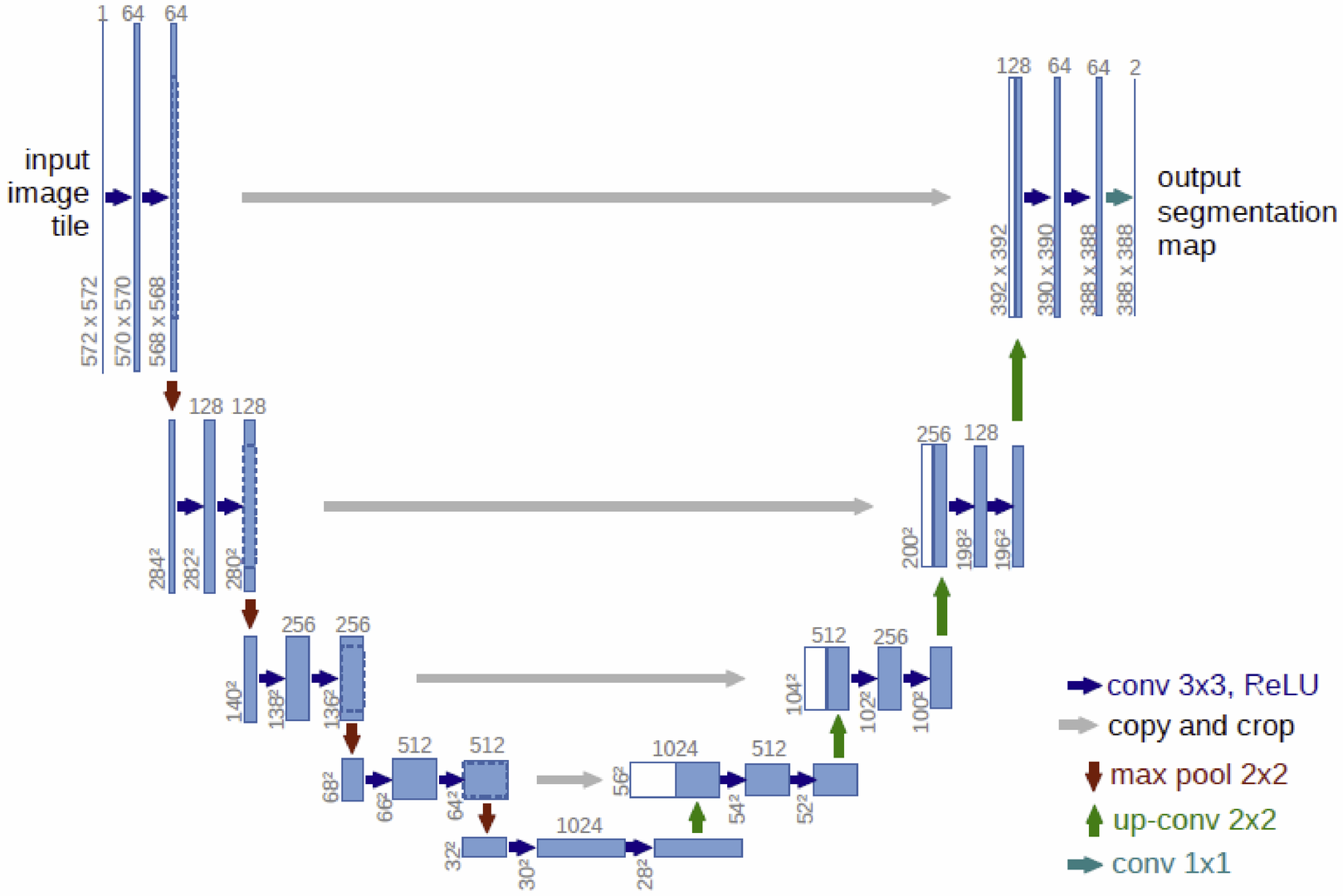
High-res:
D_1 x H/2 x W/2

Predictions:
H x W

Similar to encoder / decoder architecture

U-Net Architecture (Ronneberger et al. 2015)

- Down- and Up-sampling with max pooling
- Additional skip connections between early and late layers in the network



Summary

- Semantic segmentation:
  - pixel-wise classification problem
- Input and output shapes are (almost) identical
- Fully convolutional networks (almost) preserve shape, but may require many layers to reach sufficiently large receptive fields
- Downsampling (e.g., max-pooling) rapidly increases receptive field, but must be undone by an upsampling step (e.g., transposed convolution)
- Problems and techniques from classification apply (e.g., class-specific weighting of loss in case of imbalanced classes).