EPFL

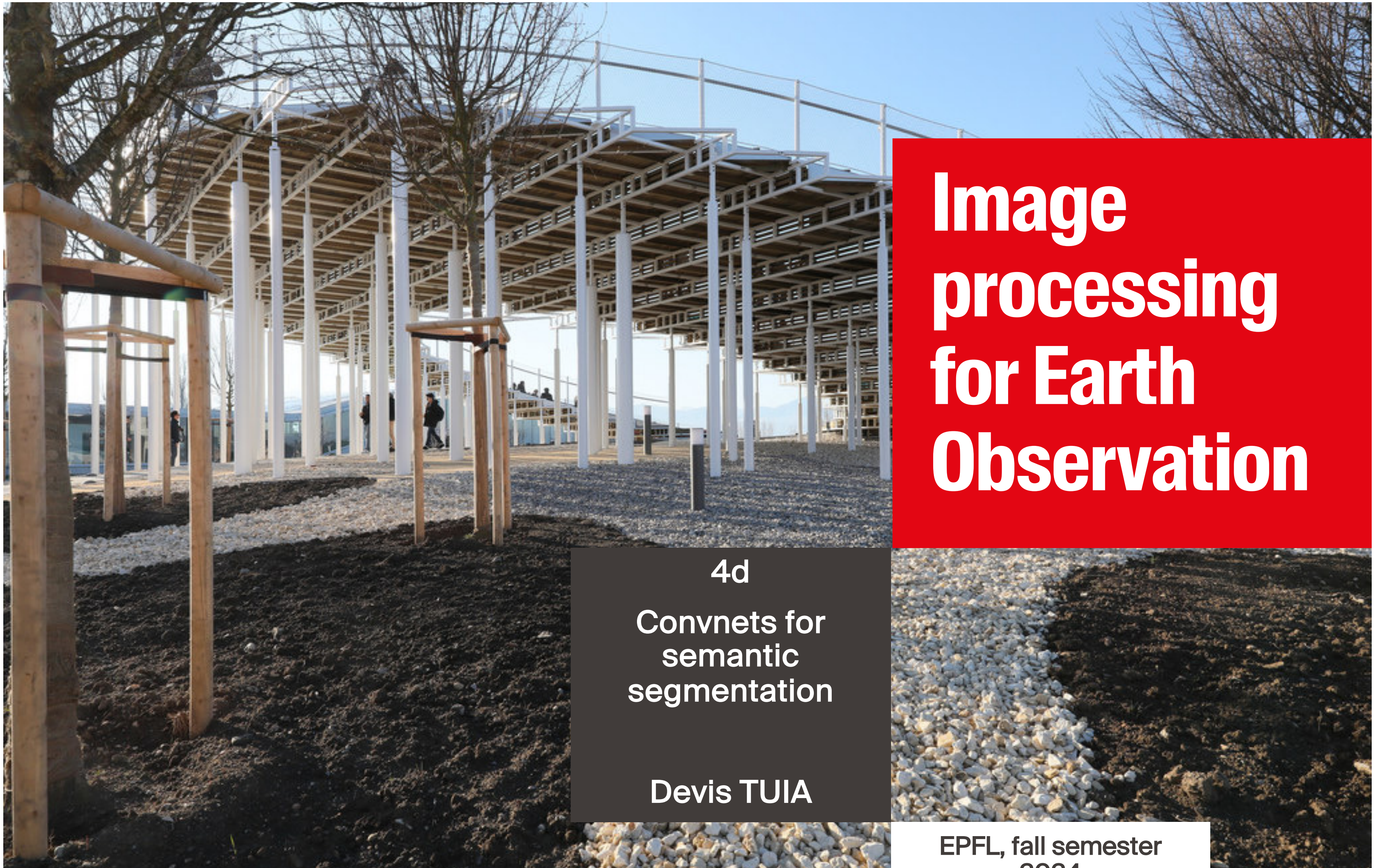


Image processing for Earth Observation

4d

Convnets for semantic segmentation

Devis TUIA

EPFL, fall semester 2024

École polytechnique fédérale de Lausanne

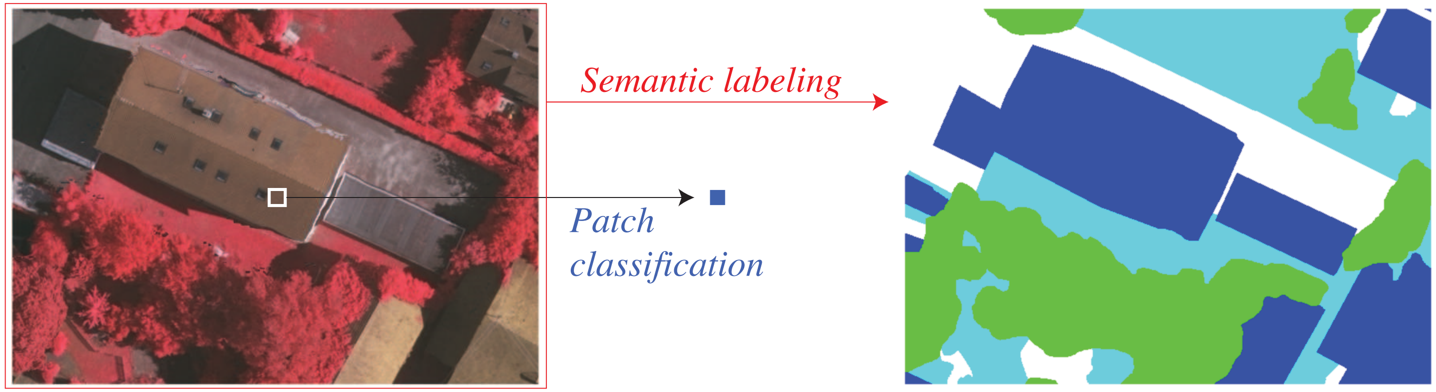
Content (6 weeks)

- W1 General concepts of image classification / segmentation
  Traditional supervised classification methods (RF)
- W2 Traditional supervised classification methods (SVM)
  Best practices

- W3 Elements of neural networks
- W4 Convolutional neural networks
- W5 Convolutional neural networks for semantic segmentation
- W6 Sequence modeling, change detection

What we saw so far is not enough for pixel classification (or semantic segmentation)

- The classifiers so far predict one value per patch
- Usually we want one prediction per pixel



- First DL models were predicting patch by patch and sliding through the image

EPFL

Patch sliding was a bit slow…

[Volpi and Tuia, IEEE TGRS, 2017]

- Prediction time at test was very slow

	CNN-PC	CNN-SPL	CNN-FPL
time (s/image)	2768 (1360.6†)	1.8	6.2

† Inference time with stride 2

- Prediction was also ambiguous. When sliding a few pixels…

Expected prediction: animal

Expected prediction: ground

New models surfaced, made for pixel classification (semantic segmentation)

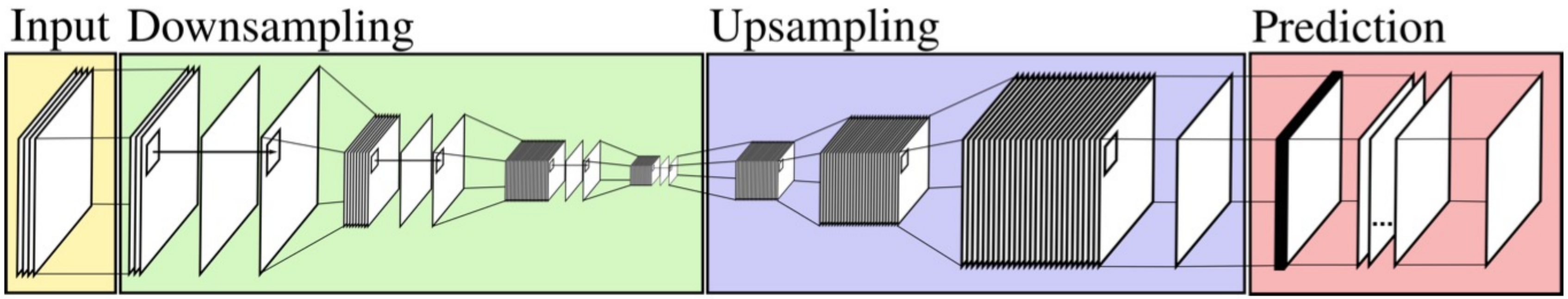
(a) Image classification

(b) Semantic segmentation

D. Tuia, D. Marcos, K. Schindler, and B. Le Saux. Deep learning-based semantic segmentation in remote sensing. In *Deep learning for Earth Sciences - A comprehensive approach to remote sensing, climate science and geosciences*. Wiley, 2021.

Semantic segmentation models usually have 2 steps

- 1. Encoder → Compresses information in features semantically interesting (= good for classification)
    - They summarize the image content, but you lose spatial detail (activations are coarse spatially)
    - Often CNN as those we saw last week
- 2. Decoder
    - Retrieves spatial details often by oversampling (either hardcoded or learned)

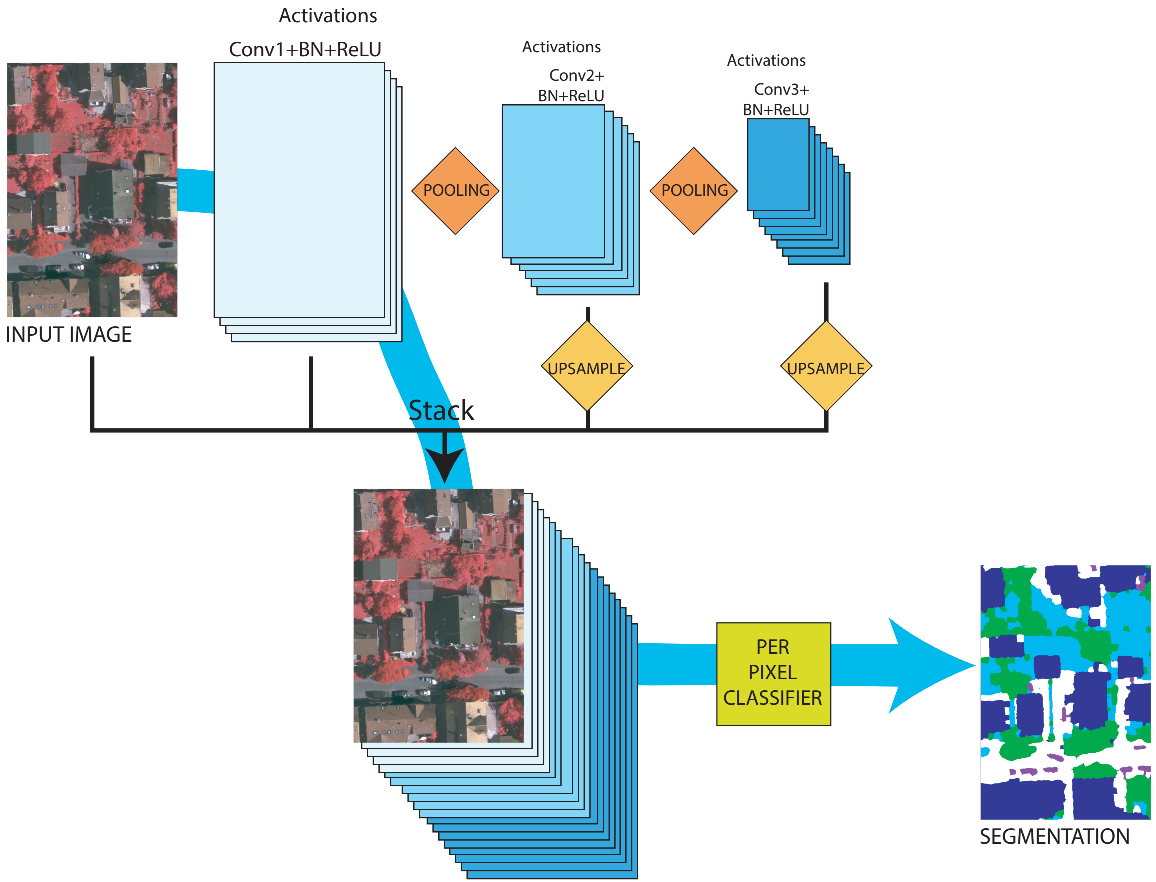


M. Volpi and **D. Tuia**. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.*, 55(2):881–893, 2017.

Hard-coded decoding: hypercolumns

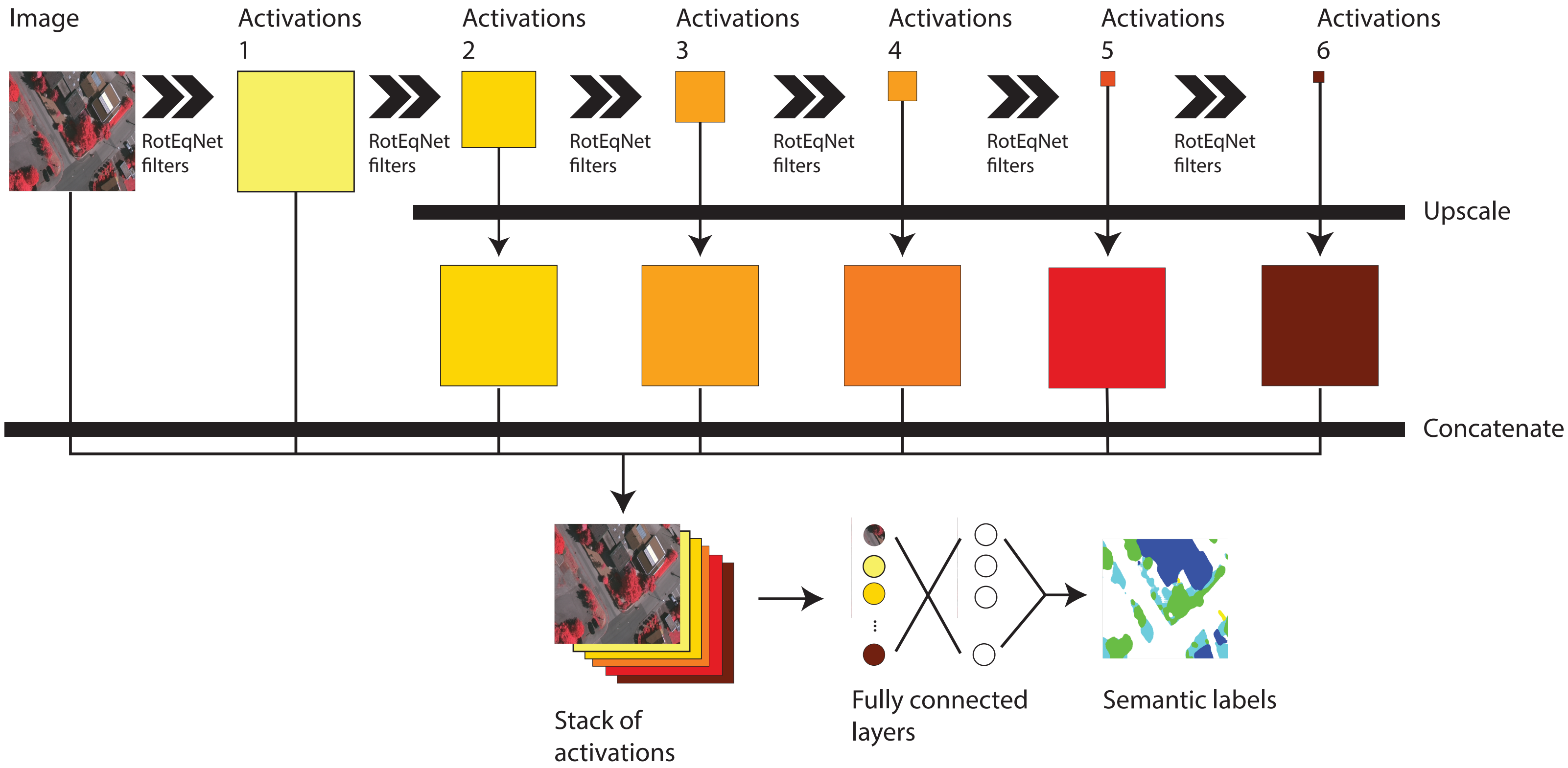
Upsample your features to the original resolution

Learn a per pixel classifier on the resulting datacube (an MLP)



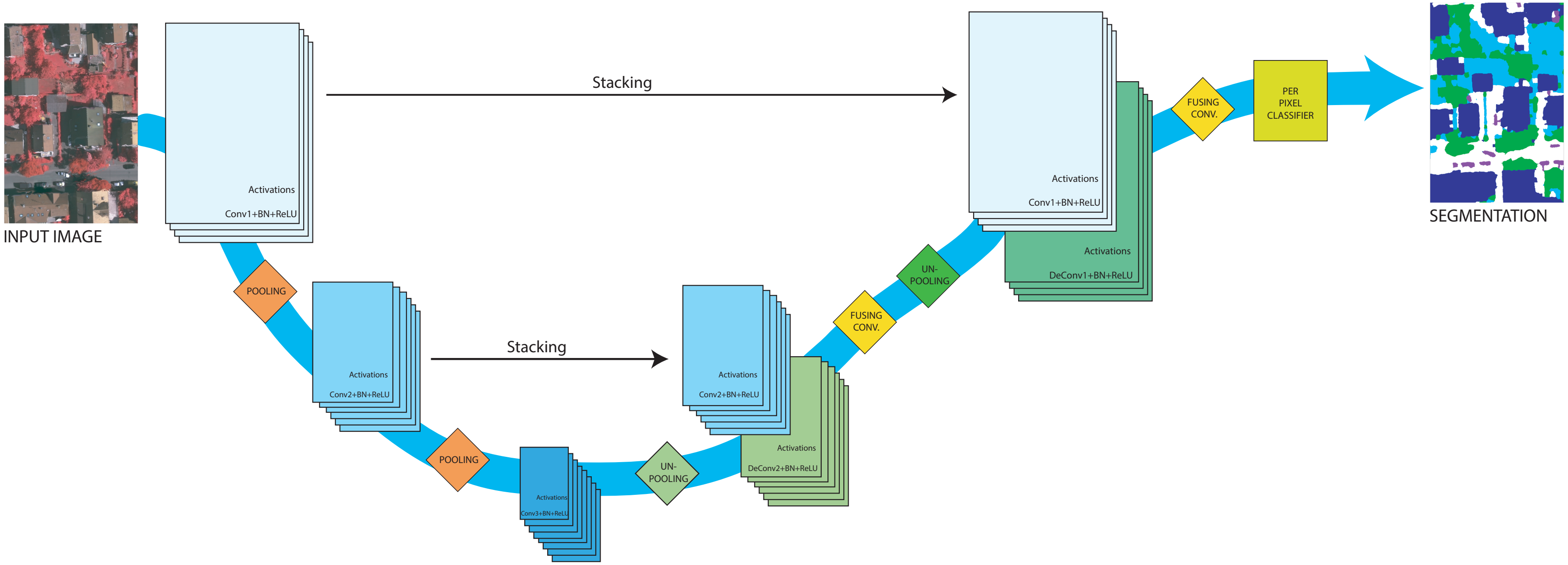
D. Marcos, M. Volpi, B. Kellenberger, and **D. Tuia**. Land cover mapping at very high resolution with rotation equivariant CNNs: towards small yet accurate models. *ISPRS J. Int. Soc. Photo. Remote Sens.*, 145(A):96–107, 2018.

Another view on hypercolumns



EPFL

Learned decoding: U-Net

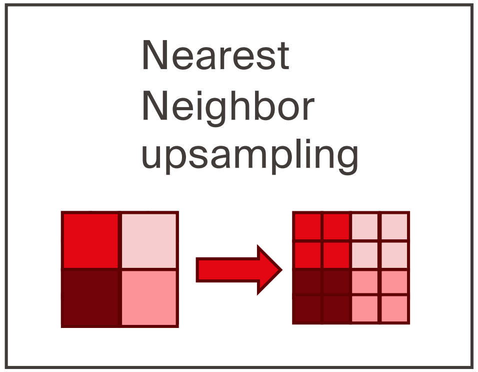


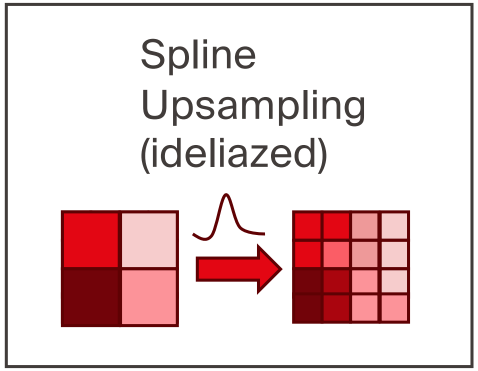
D. Tuia, D. Marcos, K. Schindler, and B. Le Saux. Deep learning-based semantic segmentation in remote sensing. In *Deep learning for Earth Sciences - A comprehensive approach to remote sensing, climate science and geosciences*. Wiley, 2021.

EPFL

How to ensure we have the details?

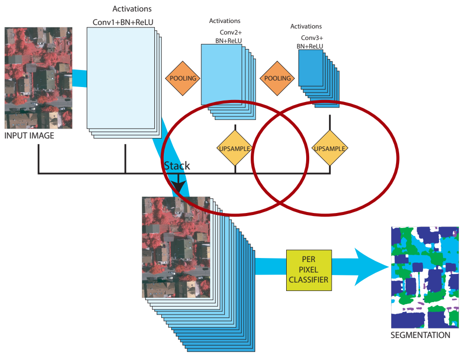
- Both models rely on upsampling
- If you use traditional upsampling models
  - Nearest neighbors (you copy the value to a k x k window)
  - Splines
  - Bicubic interpolations
- Your output will be smooth, and so the final output, which is unsatisfactory.
- 2 approaches are common.

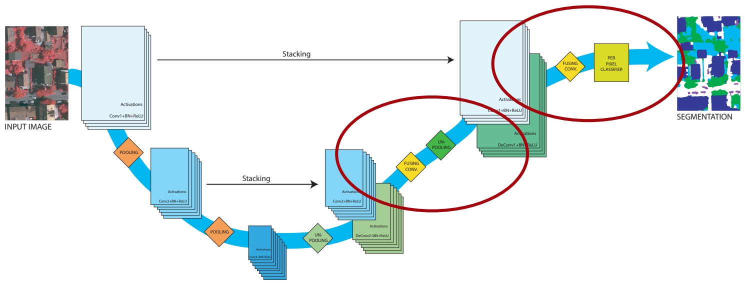




1. Learning the upsampling

- The upsampling is also a place where one can learn parameters leading to sharper maps!

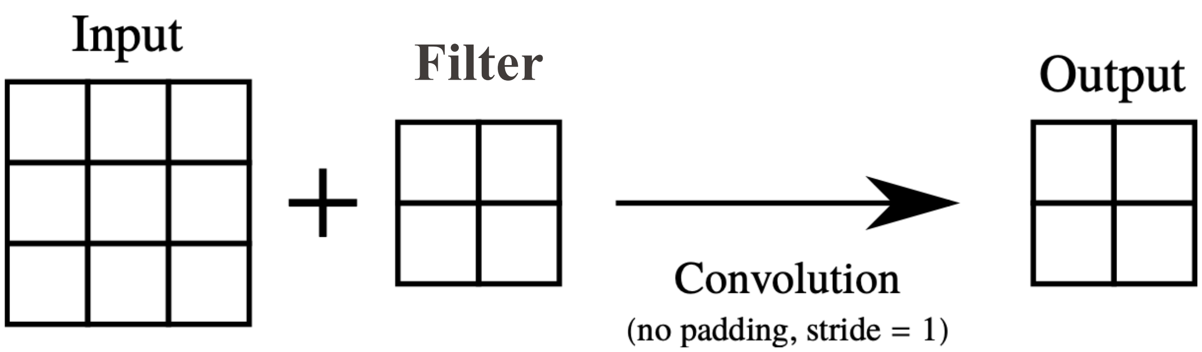




- Instead of using an existing interpolator, you learn its weights, so that it leads to sharp maps. It's backprop again!

1. Transposed convolutions

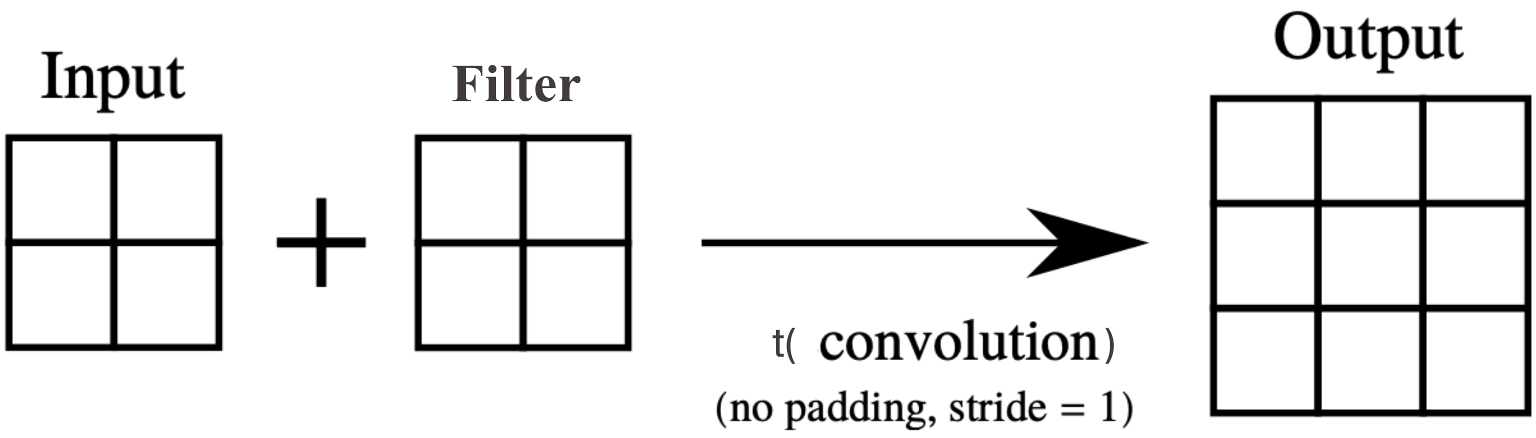
- We learn a filter that upsamples well.
- It’s another convolutional filter
- Instead of summarising, it “de-summarizes”
- As an example, take a convolution (see course of last week):



- Convolutions downsample.

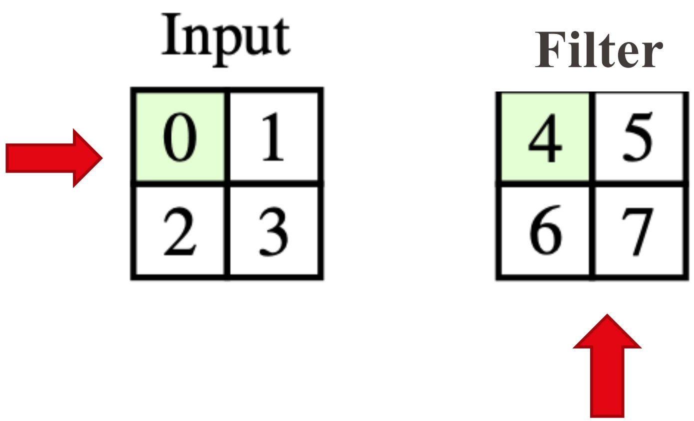
1. Transposed convolutions

- We learn a filter that upsamples well.
- It's another convolutional filter
- Instead of summarising, it "de-summarizes"
- The t(convolution) does instead :



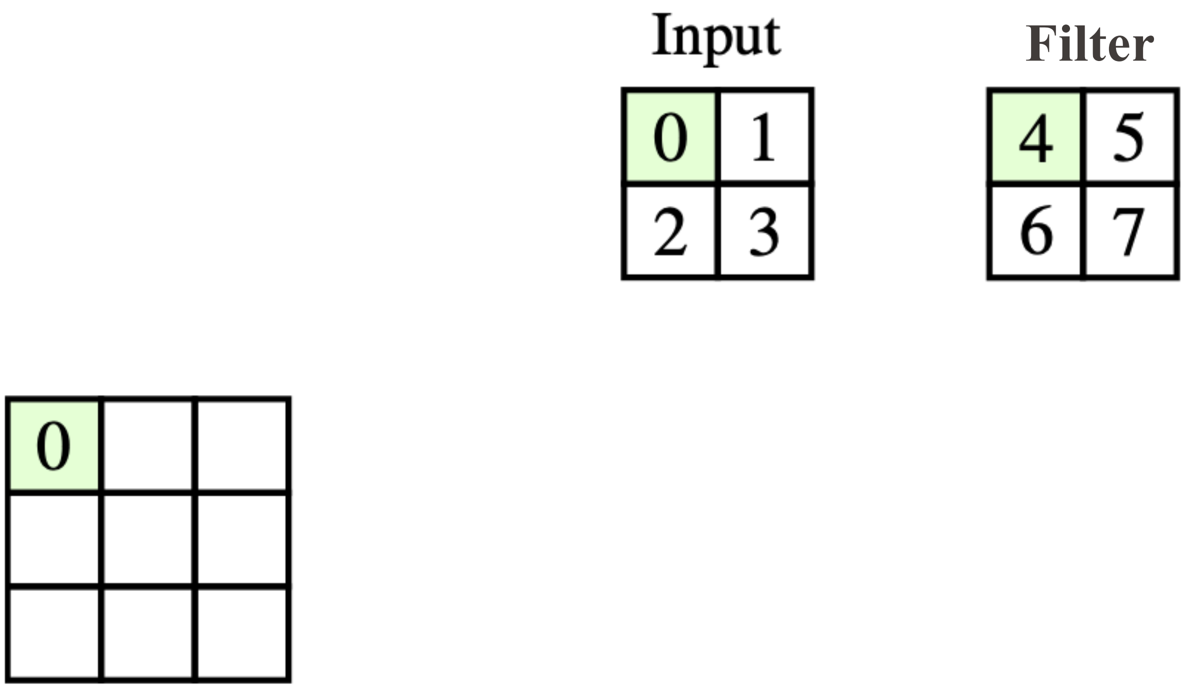
1. Transposed convolution: how?

Former activation map
(to be upsampled)

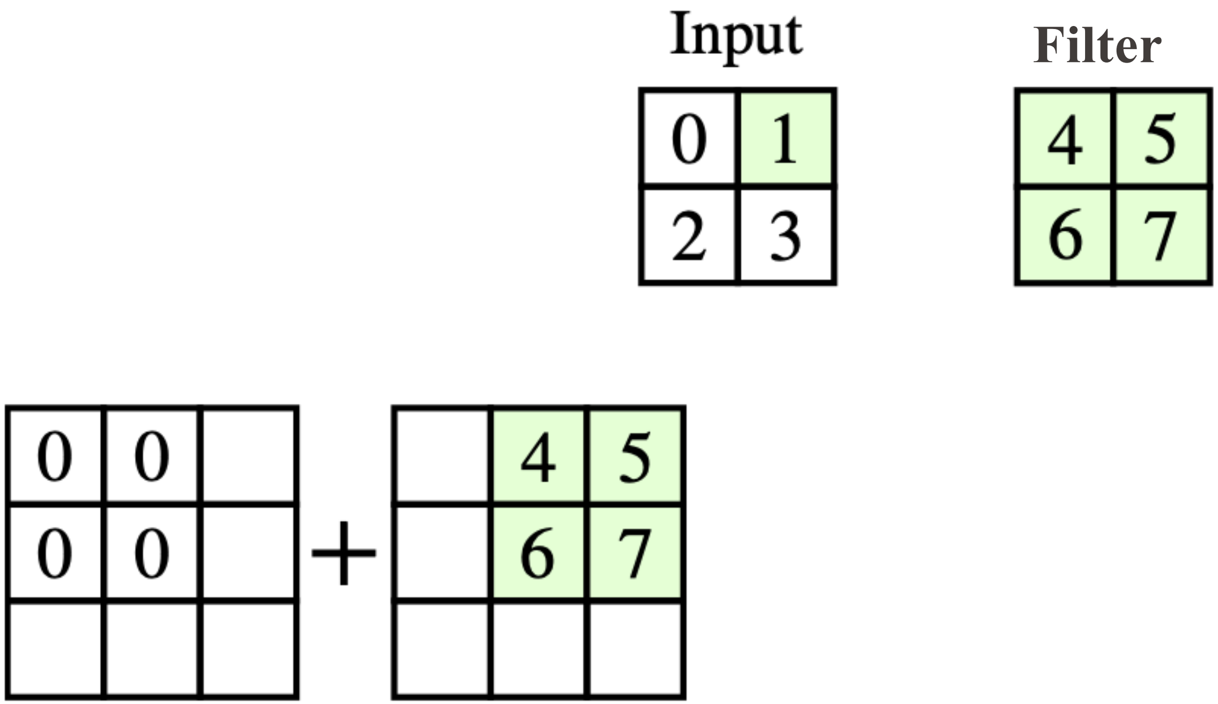


The parameters of the filter
<u>are to be learned</u> as any other filter!

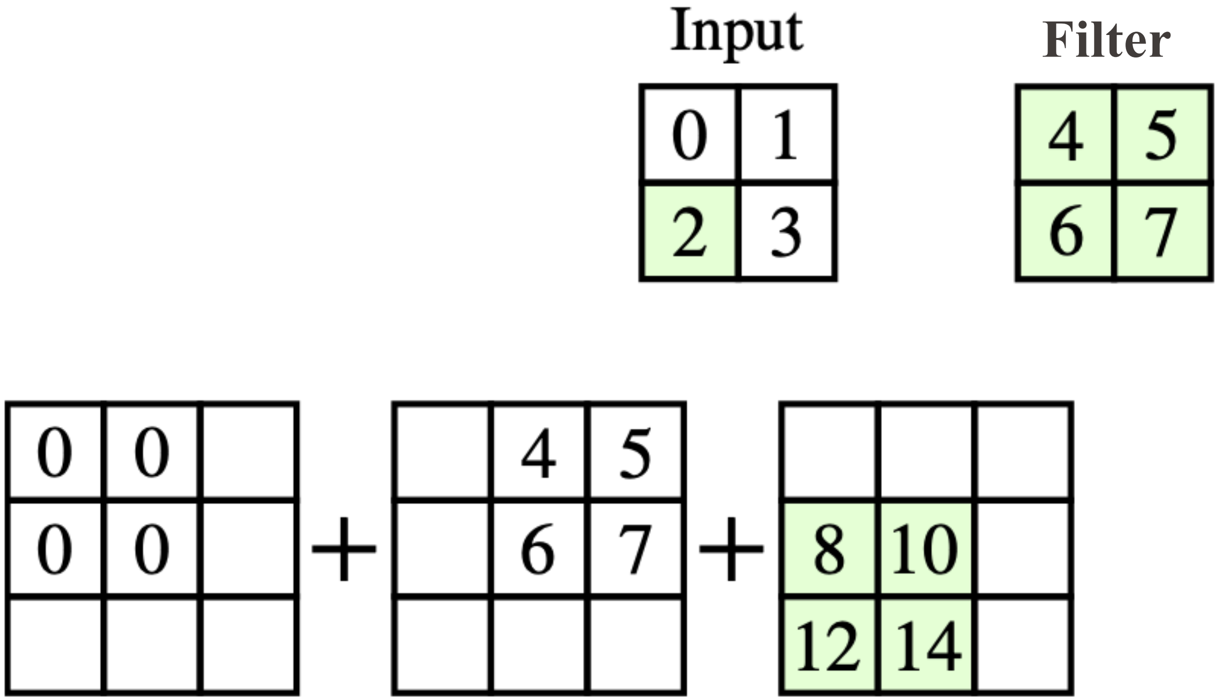
1. Transposed convolution: how?



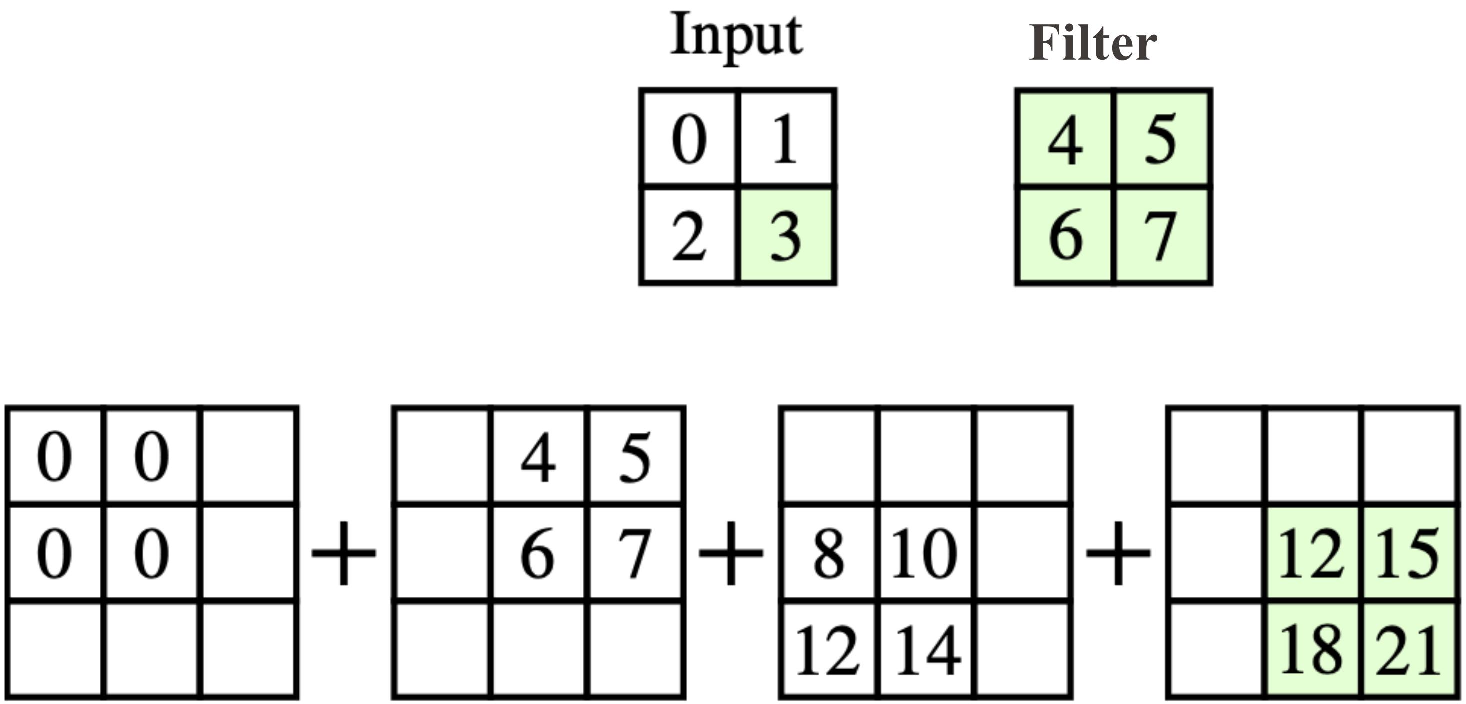
1. Transposed convolution: how?



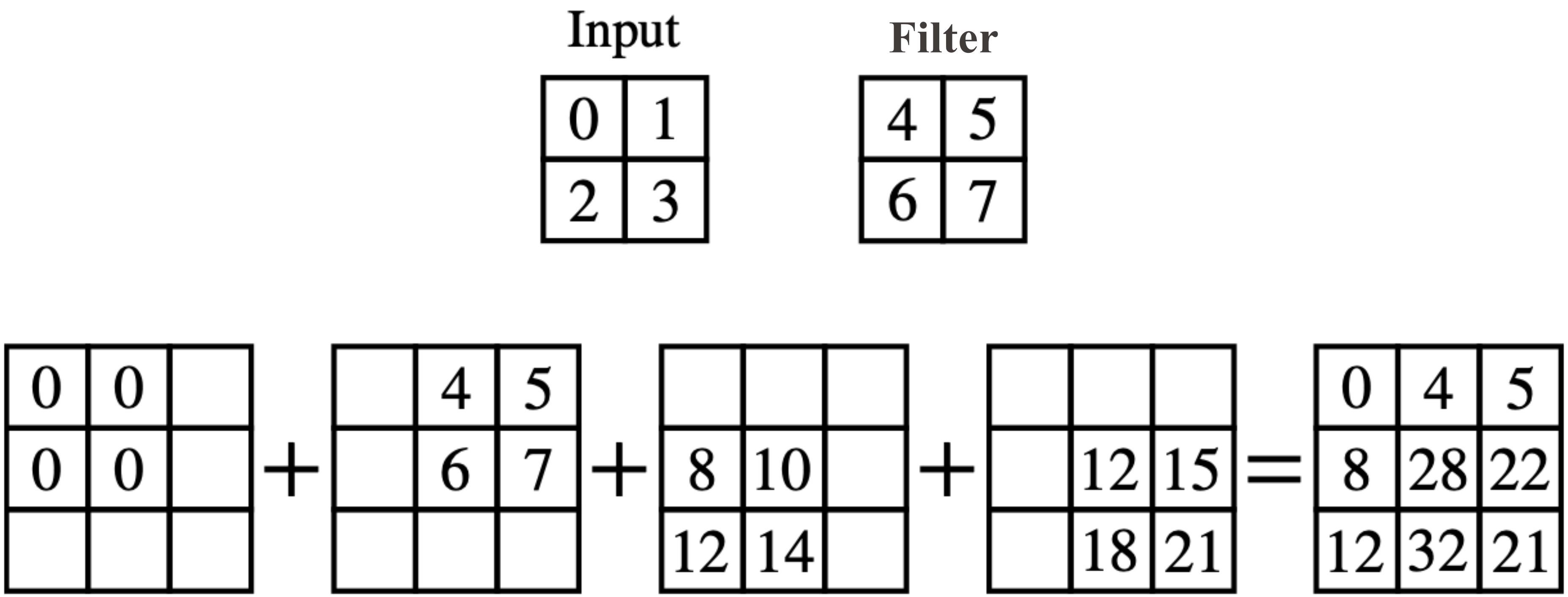
1. Transposed convolution: how?



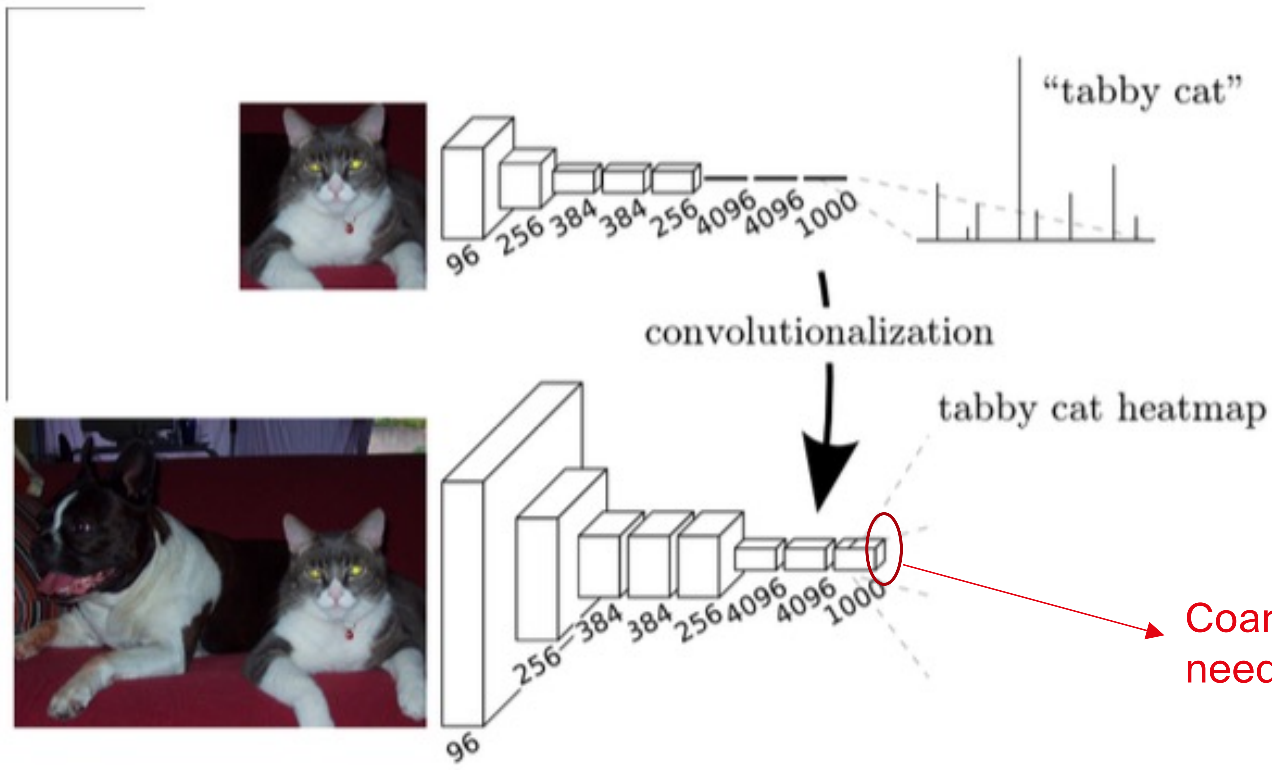
1. Transposed convolution: how?



1. Transposed convolution: how?

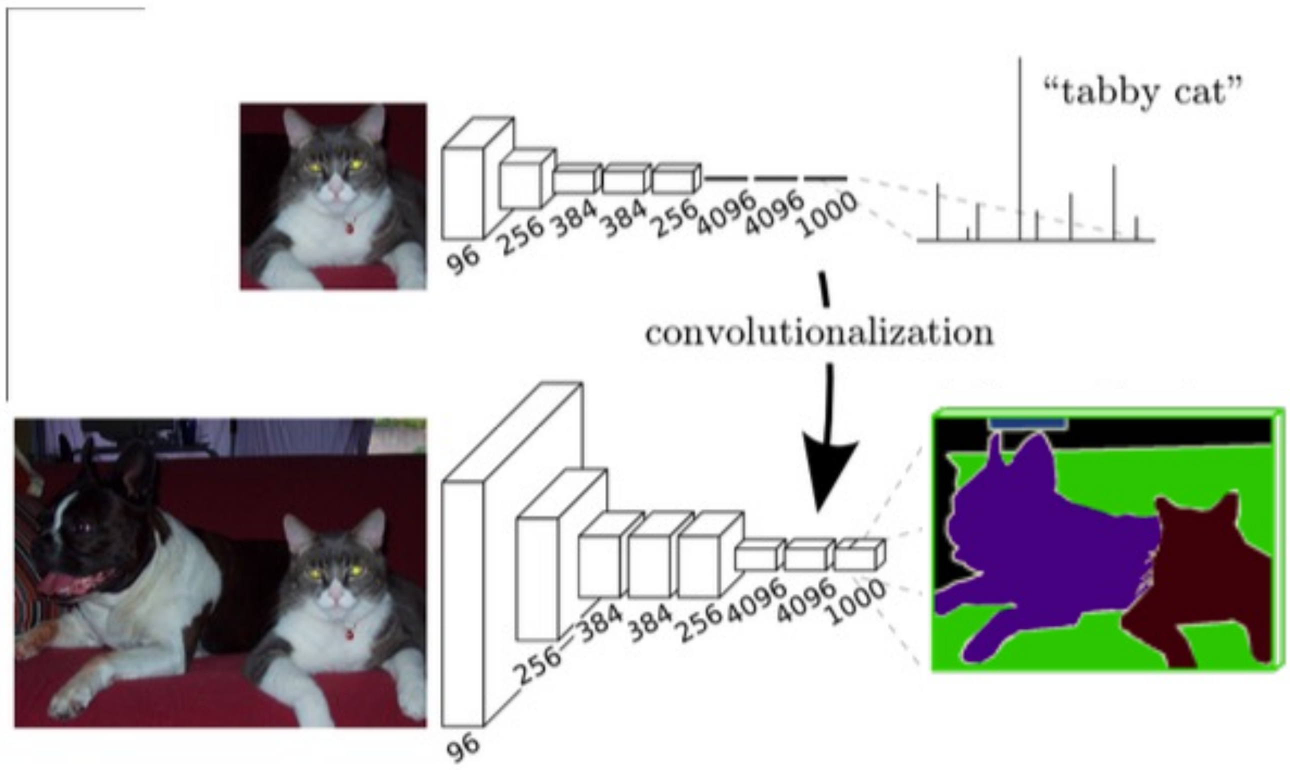


This is what launched the Fully convolutional networks



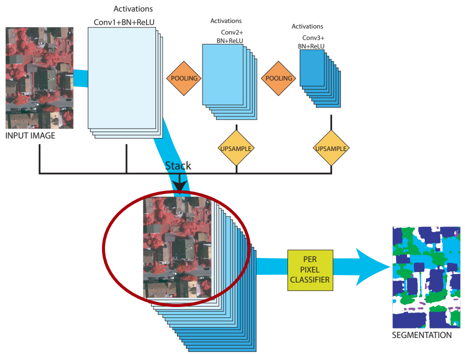
Coarse map, needs upsampling

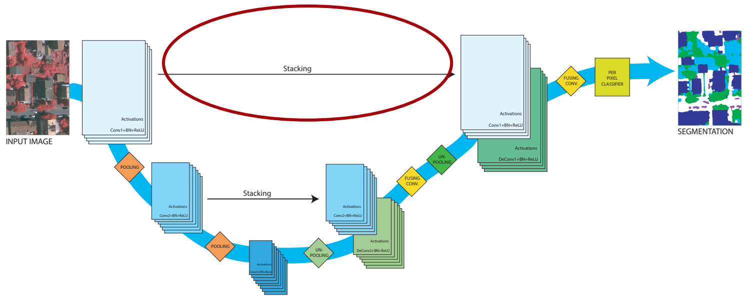
This is what launched the Fully convolutional networks



2. Skip connections

- You might have noticed, both models re-inject the high res information during decoding
- Usually via stacking





- This helps keeping a sharp response, since you re-inject features before spatial pooling
- It is a key to the success of semantic segmentation models

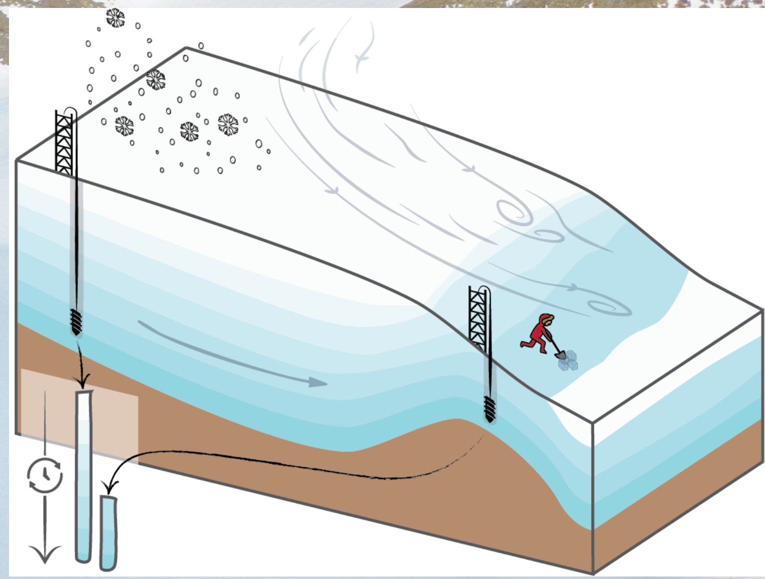
Is this stuff really useful?

2-3 examples from the ECEO lab current research

Cryosphere: treasure hunting for blue ice

- Blue ice areas are resurgences of very old ice
- They also hold most meteorites found of Earth
- They are easy to spot
- But Anctartica is BIG!

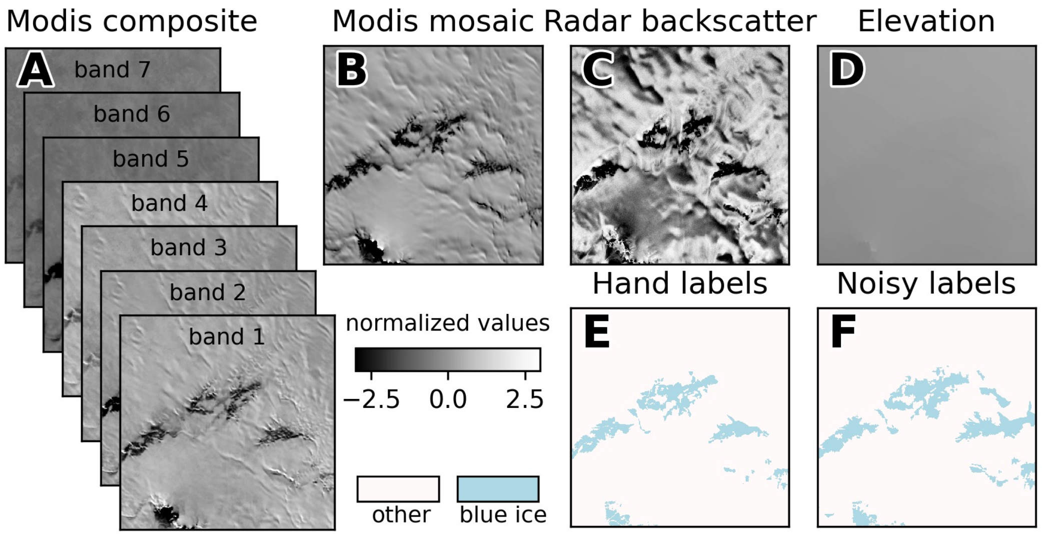
e.g., Yan et al., Nature (2019)



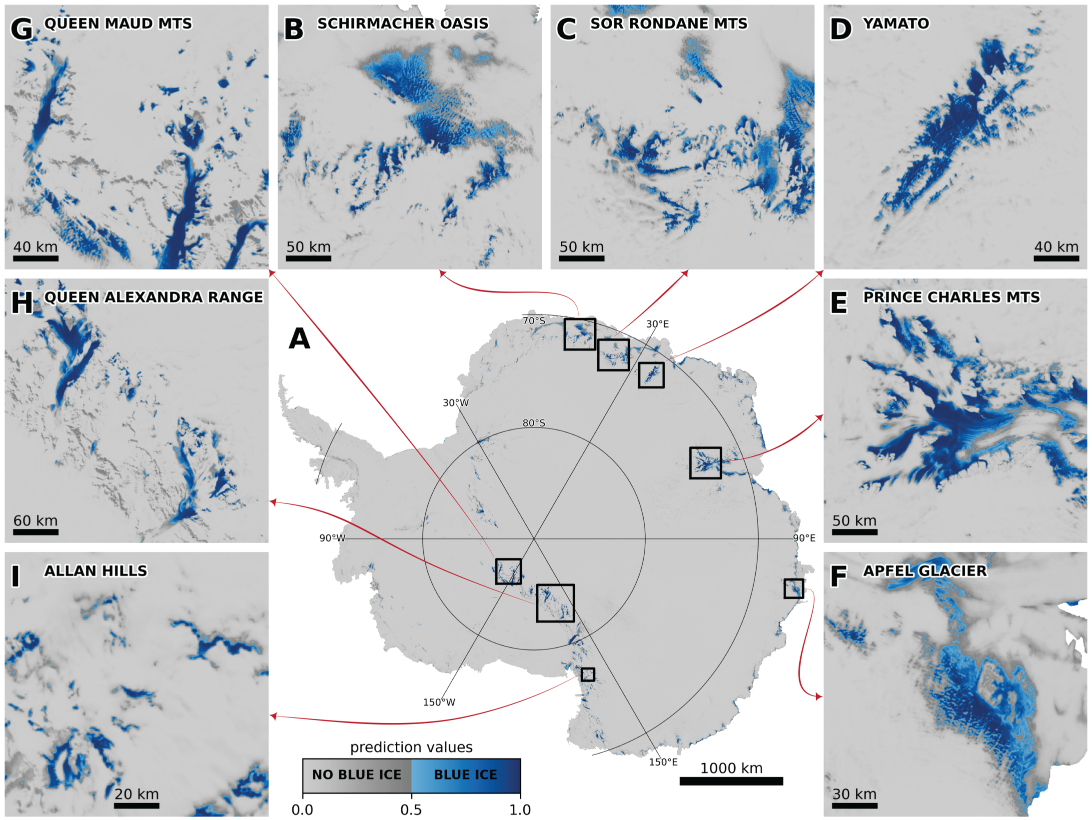
EPFL

Cryosphere: treasure hunting for blue ice

- We collected continental-scale remote sensing data
  - Optical (MODIS)
  - Rarar (RADARSAT)
  - Elevation
- And learned a U-Net network to predict blue ice locations

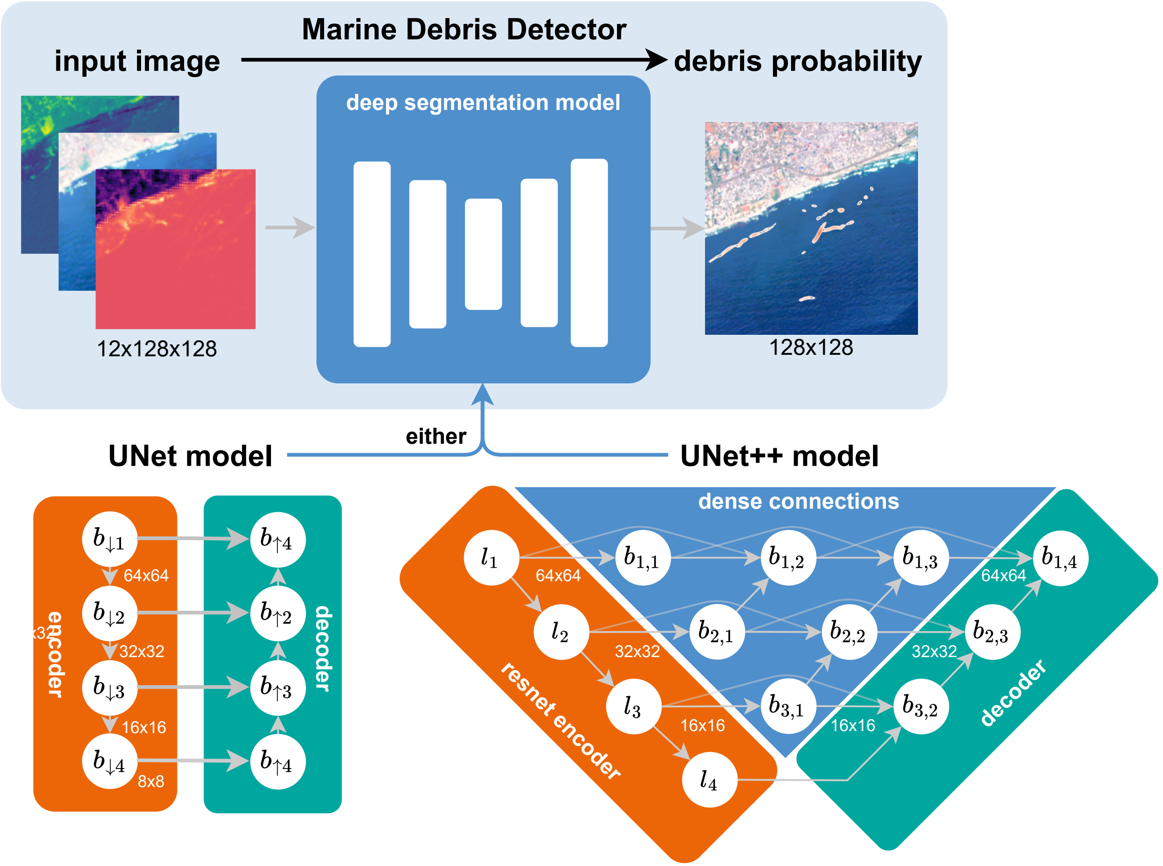


Cryosphere: treasure hunting for blue ice



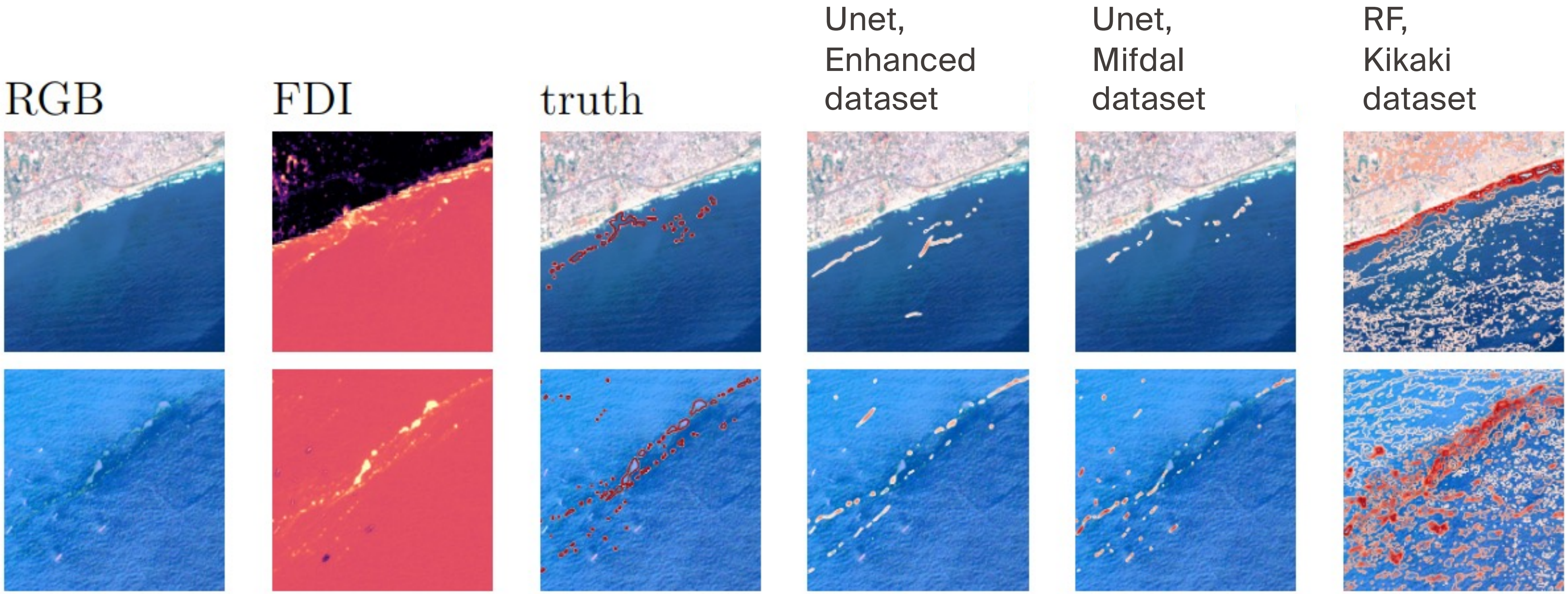
Tollenaar et al., Where the white continent is blue: deep learning locates bare ice in Antarctica.
Geophysical Research Letters, 51(3):e2023GL106285, 2024. https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2023GL106285

Segmenting floating plastics in the Oceans



M. Rußwurm, S. J. Venkatesa, and **D.Tuia**. Large-scale Detection of Marine Debris in Coastal Areas with Sentinel-2. *iScience, 108402,* 2023.
Available: https://www.sciencedirect.com/science/article/pii/S2589004223024793

Prediction examples



Qualitative examples: https://marcrusswurm.users.earthengine.app/view/marinedebrisexplorer

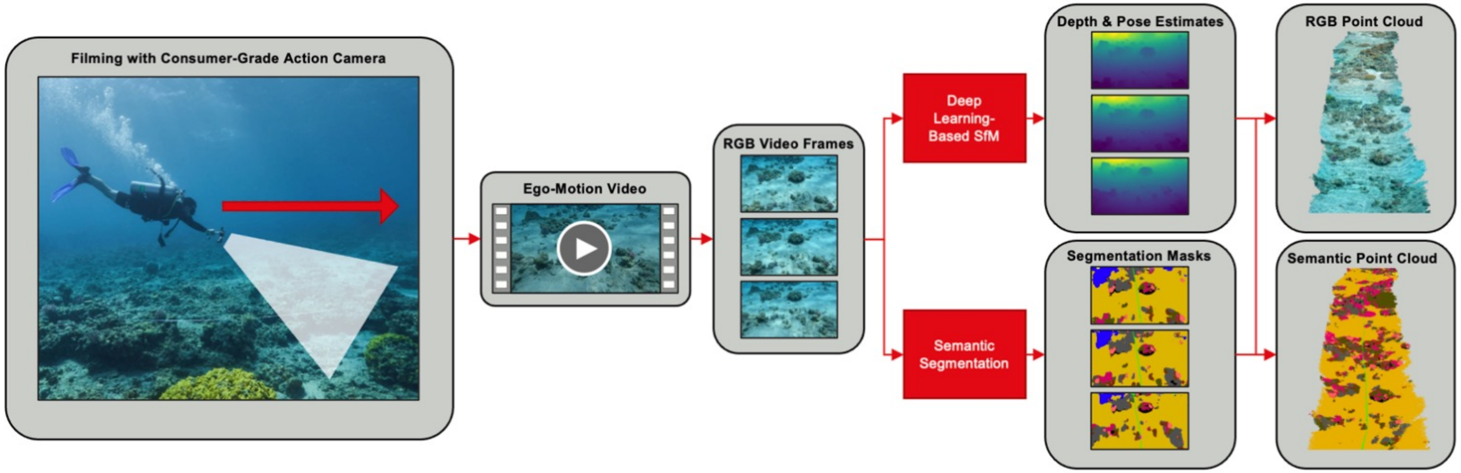
EPFL

Transnational Red Sea Center
Bridging Science & Diplomacy for the Future of Corals

Enabling scalable reef monitoring: Open source, fast, large scale.

- A model that works on videos, leveraging 2 tasks

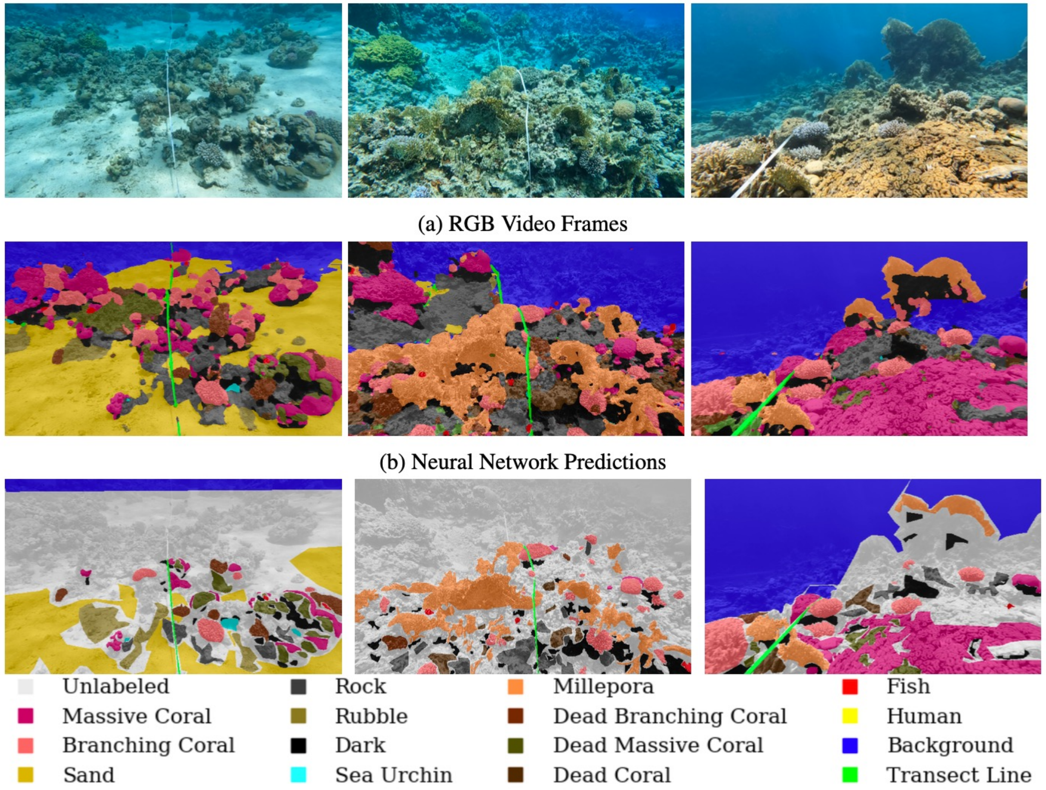
- Once trained, 3D reconstructs a 100m transect in playing video time
- Tested in Israel, Jordan and Djibouti in 2022/2023



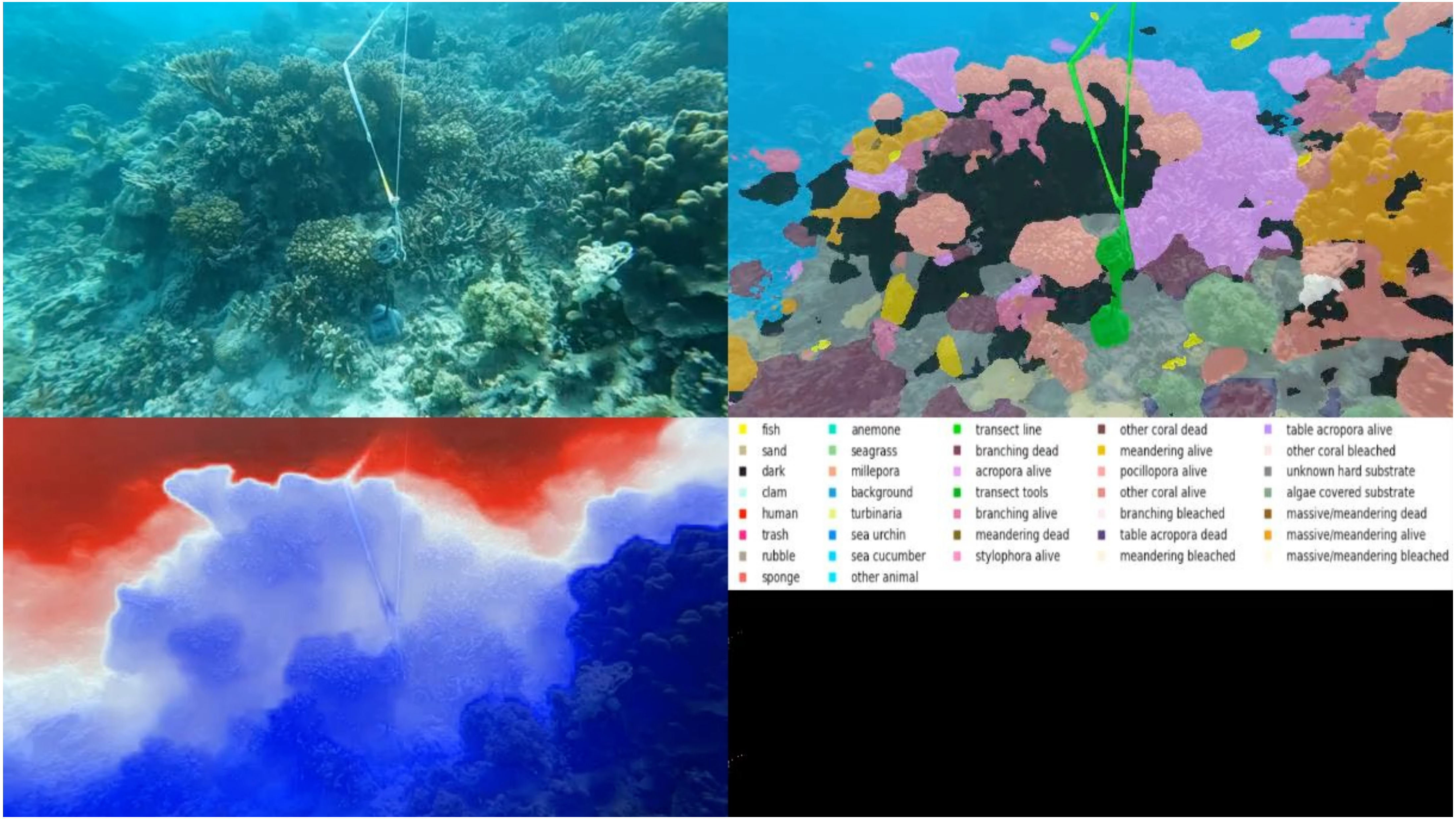
J. Sauder, G. Banc-Prandi, A. Meibom, and **D. Tuia**. Scalable semantic 3d mapping of coral reefs with deep learning. *Methods in Ecology and Evolution*, 2024. https://arxiv.org/abs/2309.12804


Semantic segmentation

- Unet with ResNeXt backbone
- ~85% accurate in Jordanian and Israeli reefs



J. Sauder, G. Banc-Prandi, A. Meibom, and **D. Tuia**. Scalable semantic 3d mapping of coral reefs with deep learning. *Methods in Ecology and Evolution*, 2024. https://arxiv.org/abs/2309.12804


fish
sand
dark
clam
human
trash
rubble
sponge
anemone
seagrass
millepora
background
turbinaria
sea urchin
sea cucumber
other animal
transect line
branching dead
acropora alive
transect tools
branching alive
meandering dead
stylophora alive
other coral dead
meandering alive
pocillopora alive
other coral alive
branching bleached
table acropora dead
meandering bleached
table acropora alive
other coral bleached
unknown hard substrate
algae covered substrate
massive/meandering dead
massive/meandering alive
massive/meandering bleached

Summary – Semantic segmentation

- The techniques are still based on the same concepts, but focusing on responses at the pixel level.
- Convnets are great and are now the state of the art in research and industry.
- In the NLP domain, the models called Transformers are becoming the new standard, but in image semantic segmentation Convnets are still more common.
- You are now familiarized with the concepts, but the best way of mastering them is by doing!