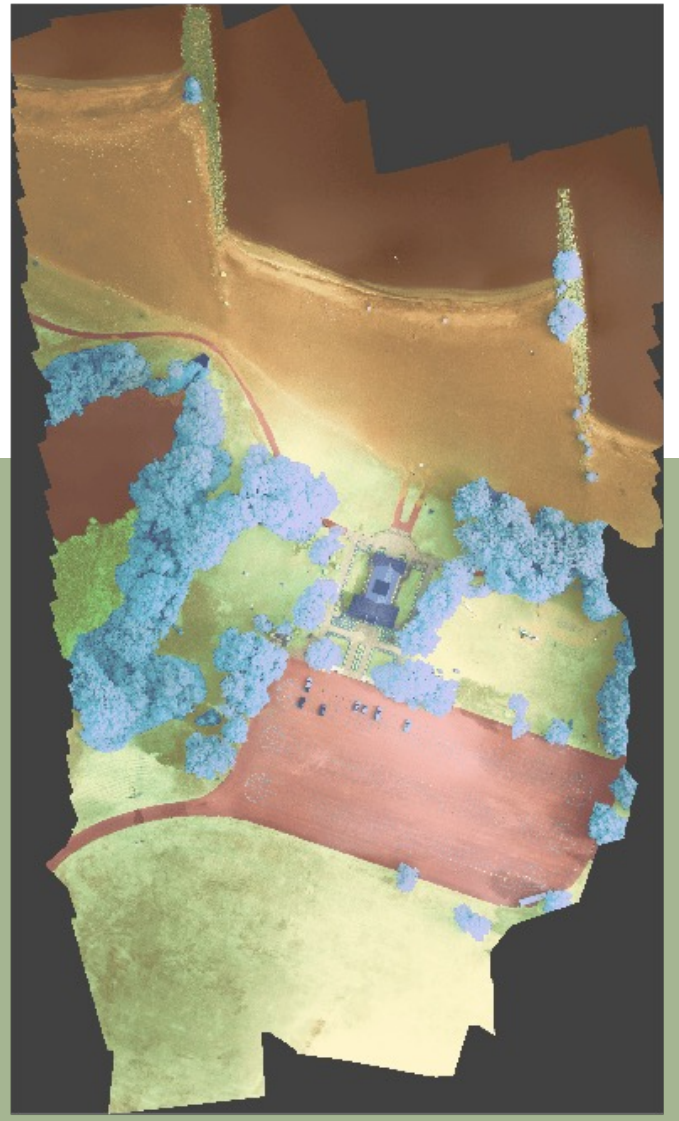
Università degli Studi di Milano – Bicocca
Master's Degree in Data Science
Academic Year 2021/2022

Foundations of Deep Learning

Semantic Segmentation for Multispectral Images



Silvia GROSSO, 881993
Niccolò ROCCHI, 881404
Julia BUI XUAN, 882385

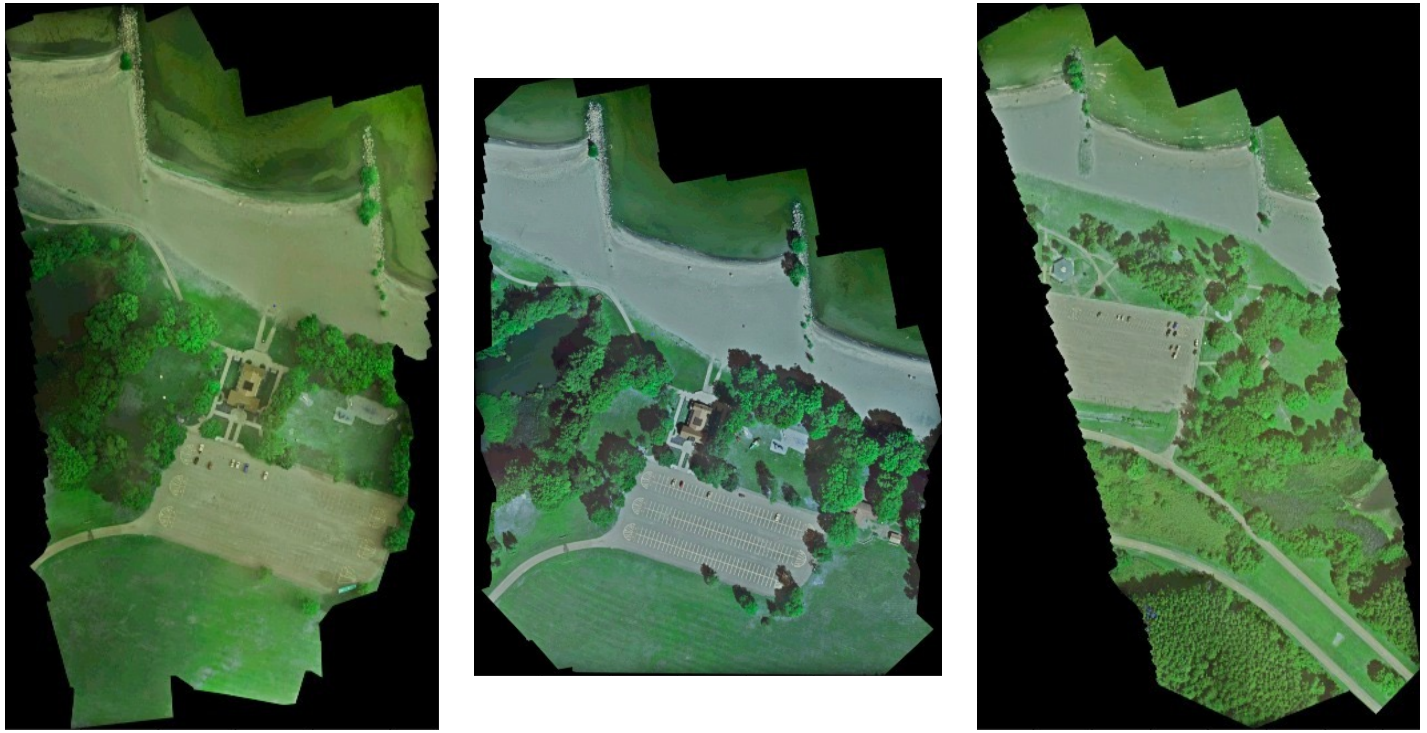
RIT-18 DATASET

Train data
Image: 9393x5642x7
Mask: 9393x5642x1

Validation data
Image: 8833x6918x7
Mask: 8833x6918x1

Test data
Image: 12446x7654x7

RGB component of Training, Validation and Test Image (Left to Right)



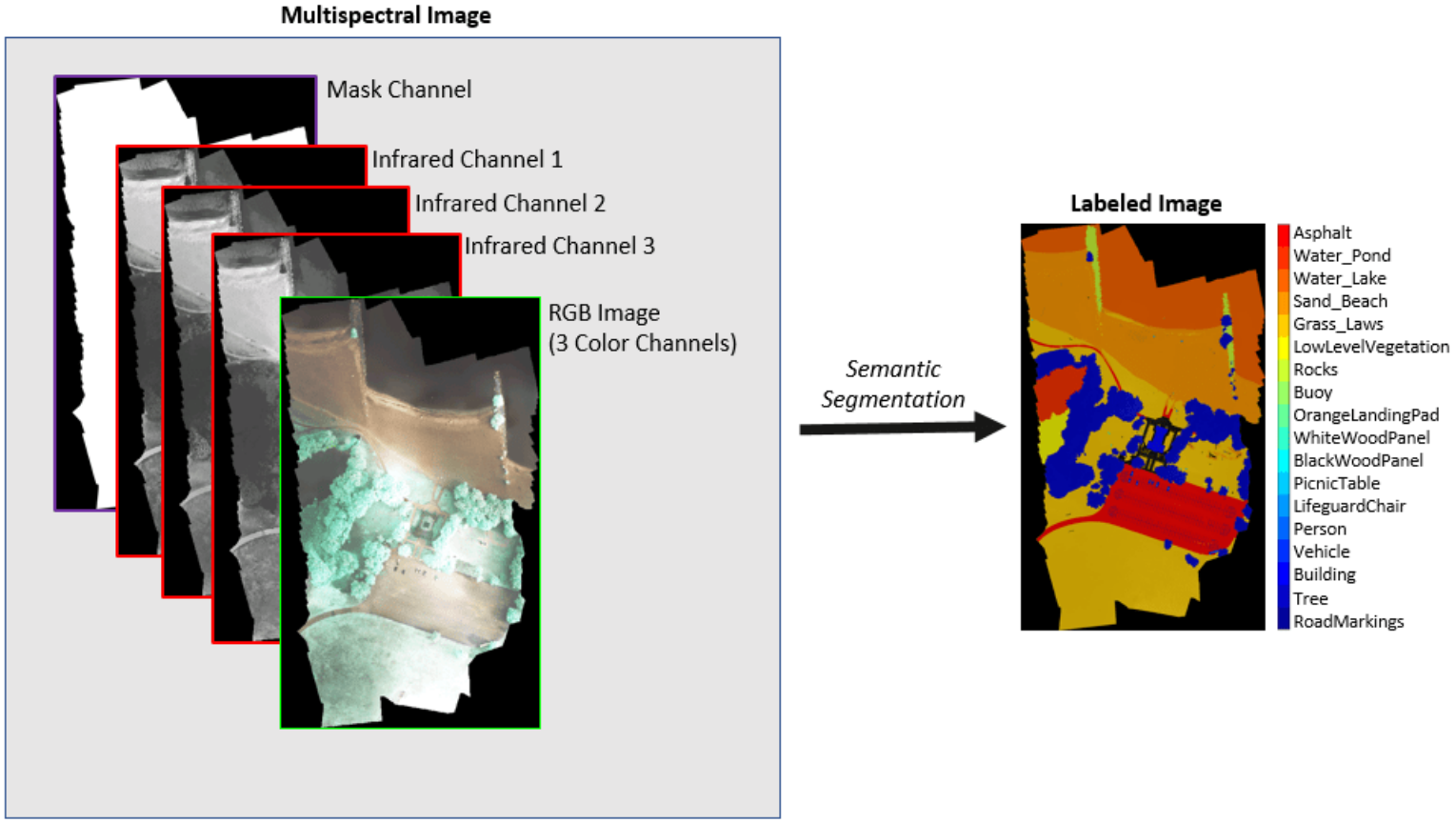
OBJECTIVES

RIT-18 image dataset

- Captured by a drone over a park
- 18 classes
- 3 near-infrared channels that provide a clearer separation of the classes

Main purposes

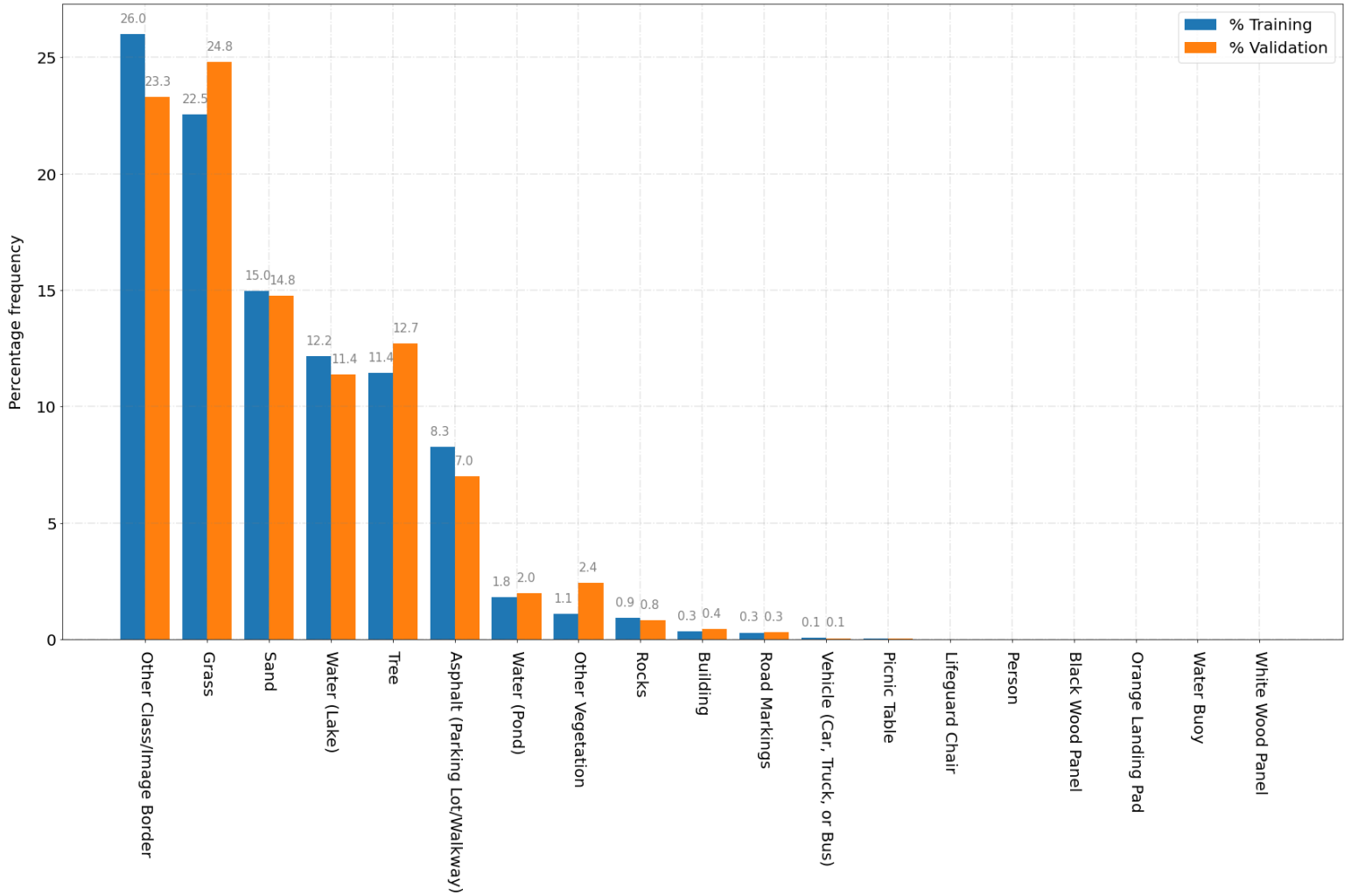
- ***Semantic segmentation***, that involves labelling each pixel in an image with a class
- Track vegetation cover for environmental purposes



DATA ANALYSIS

Class imbalance problem

- *Other Class/Image Border* is the most represented class (26% for training, 23.3% for validation)
- 11 classes are present with a frequency under 1%



U-NET ARCHITECTURE

The network is based on a **fully convolutional network**, whose architecture was modified to yield more precise segmentation.

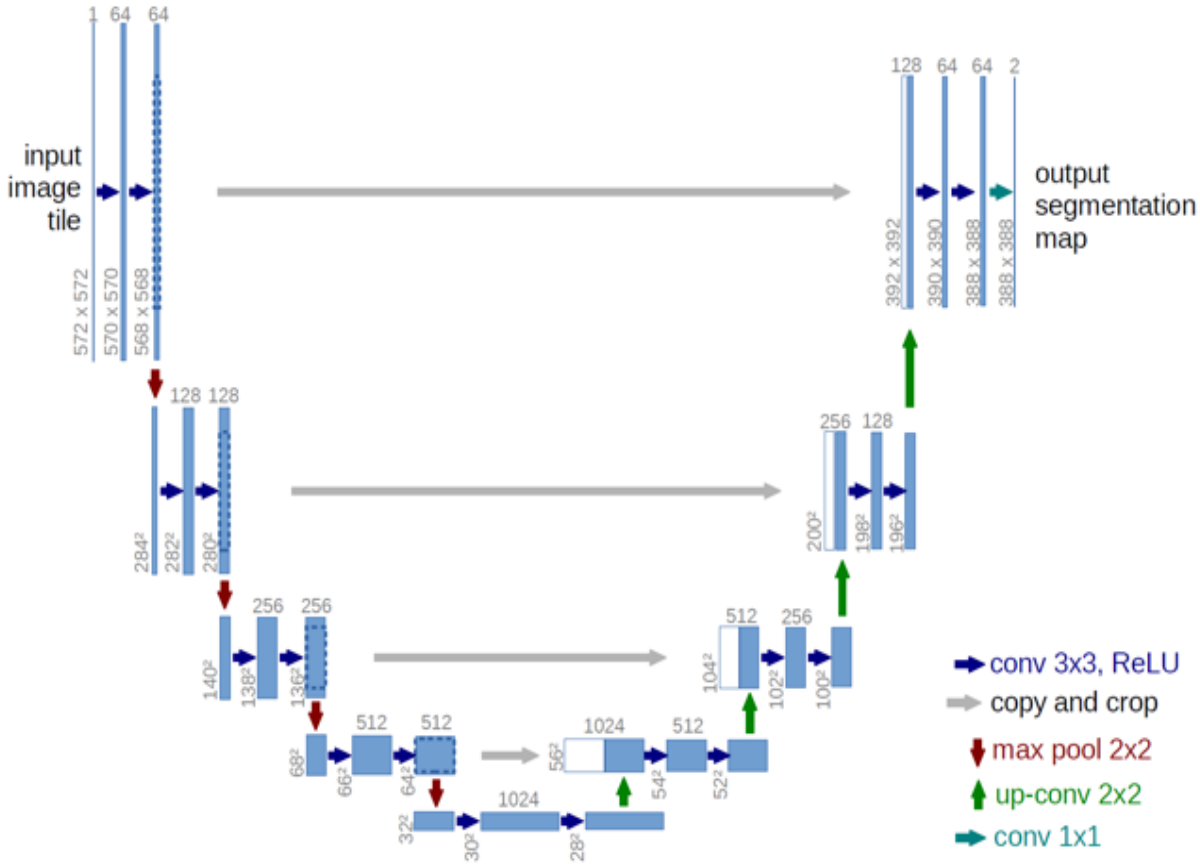
Symmetric architecture

1. Contracting path
- Convolution blocks
- Downsampling

Low-resolution, highly efficient feature maps

2. Expansive path
- Upsampling & skip connections
- Convolution blocks

Full-resolution segmentation maps



DATA AUGMENTATION - IMAGE MANIPULATIONS

Pair of train – mask image must be resized to **256x256xC .**

no overlapped crops ⇨ 800 images

overlapped crops ⇨ 1600 images

Overlapped crop is always random.
In 90% of the cases:

- Horizontal flip (prob = 0.4)
- Vertical flip (prob = 0.4)
- Colour transformation (prob = 0.5)
- Greed distortions (prob = 1)

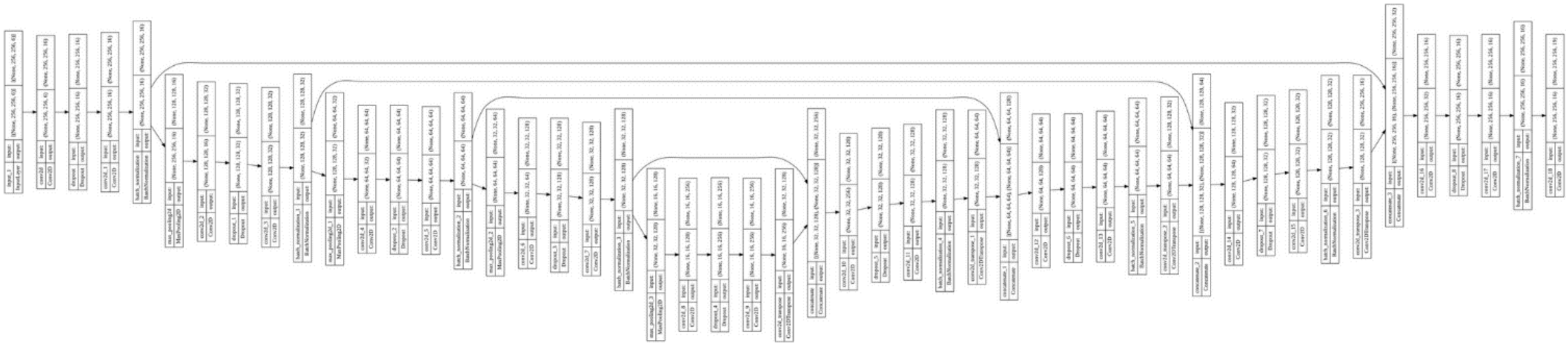
FIRST APPROACH
U-NET 1

Layers of Downsampler block x4
- 2 Convolutions
- Dropout
- MaxPooling2D

Layers of Base block
- 2 convolutions
- Dropout

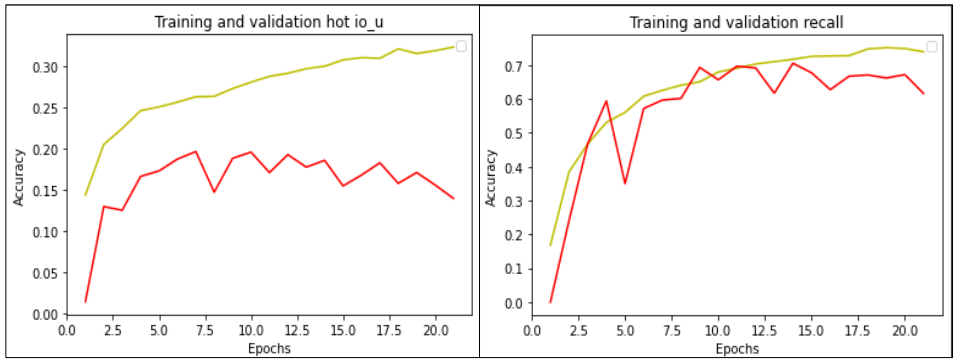
Layers of Upsampler block x4
- Transpose Convolution
- Skip Connections
- 2 Convolutions
- Dropout

Output layer
- Convolution



SECOND APPROACH
U-NET 1

Small improvements obtained by adding ***Batch Normalization*** layers at Downsampler and Upsampler blocks of the same net.

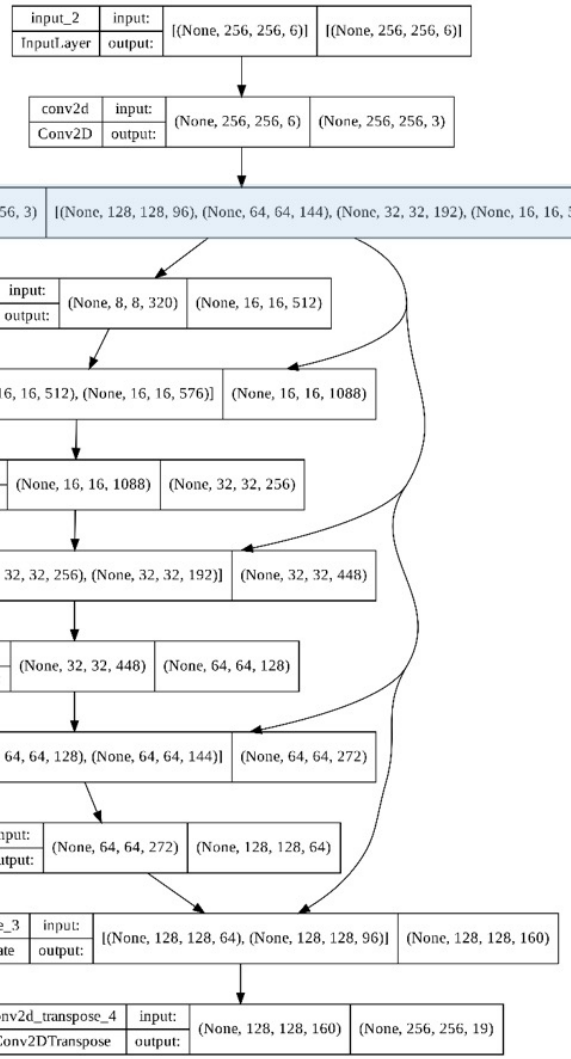


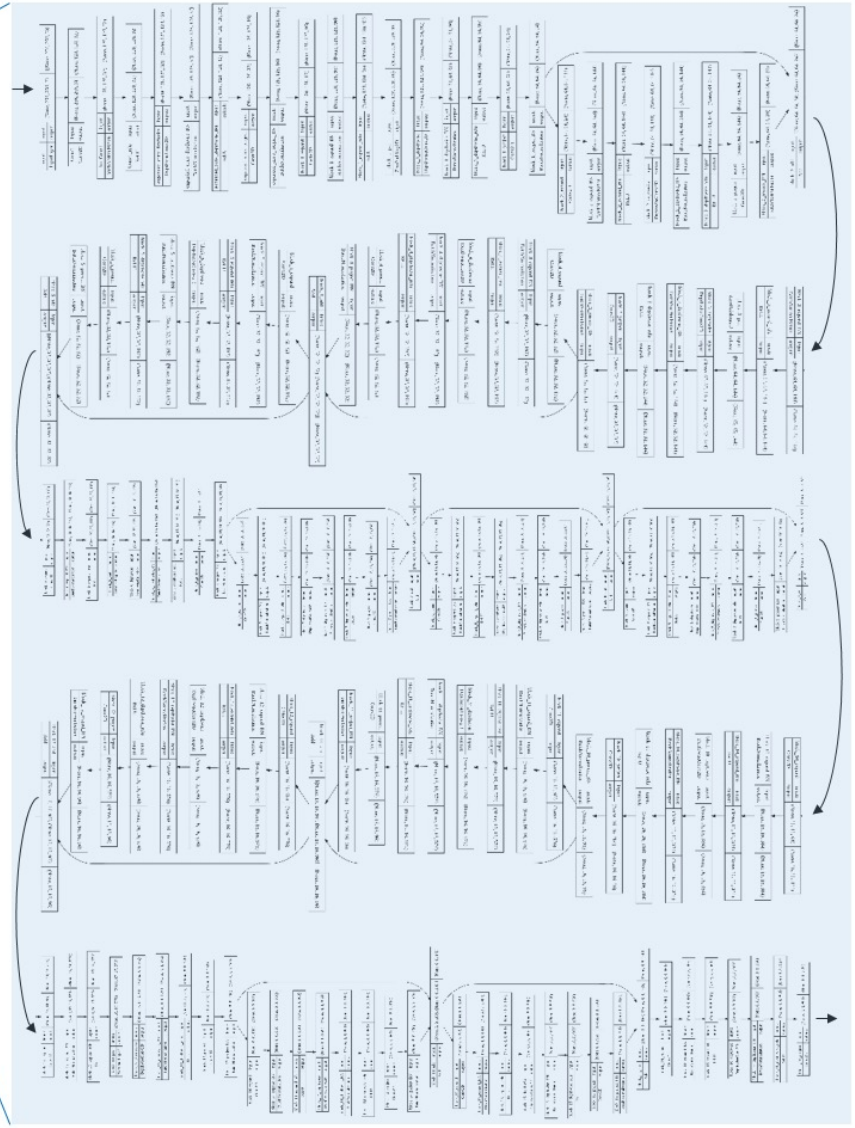
Results		
	Train	Validation
Accuracy	87.4%	65.6%
Precision	95.0%	72.6%
Recall	76.4%	61.6%
OneHotMean IoU	30.8%	14.0%
Focal Loss	0.004	0.015

- Early stopping: patience = 10, monitor = validation loss
- Epochs = 50
- Optimizer = Adam
- Learning rate = 0.001
- Bach size = 8

U-NET 2

MobileNetV2

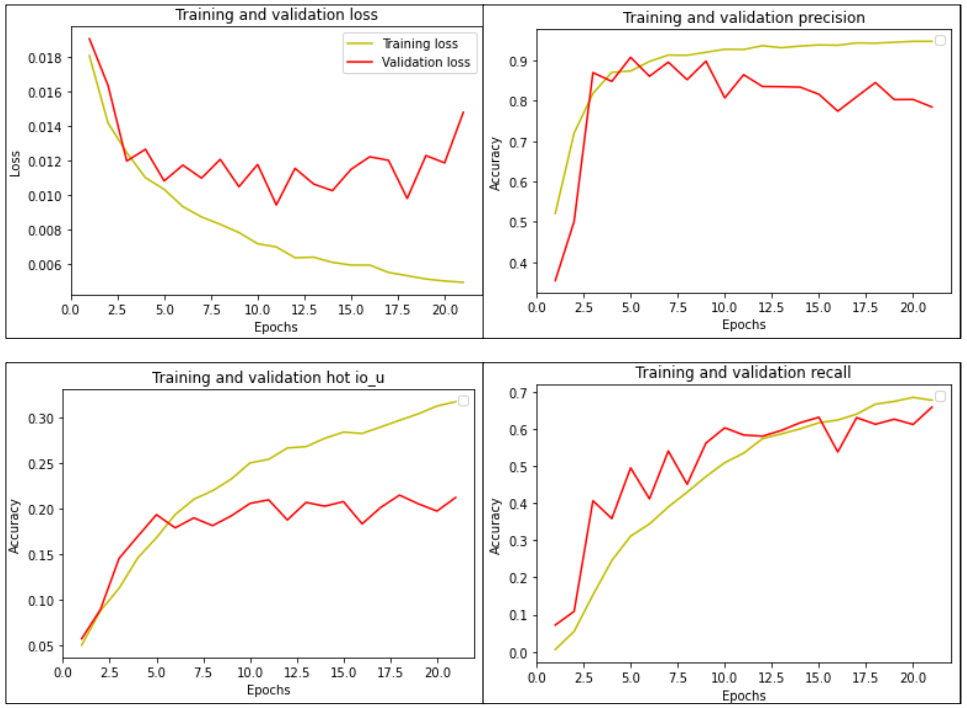




THIRD APPROACH
U-NET 2

Two methods were implemented based on different **learning rates:** 0.0001 and 0.001.

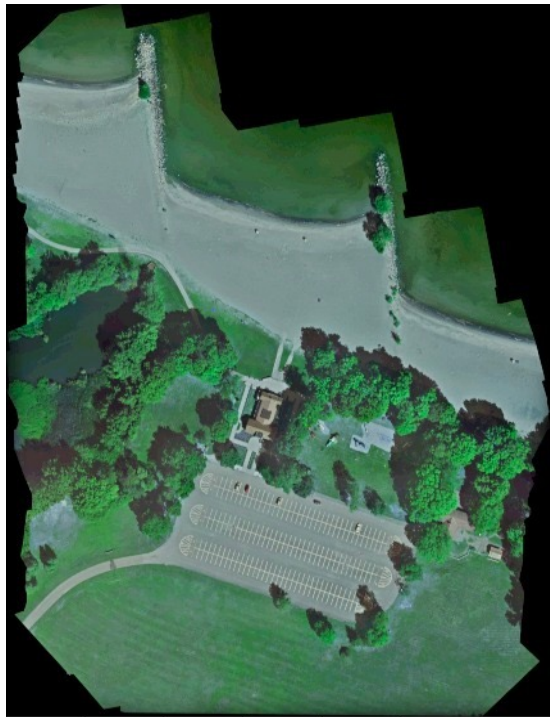
The latter one was chosen, providing better results.



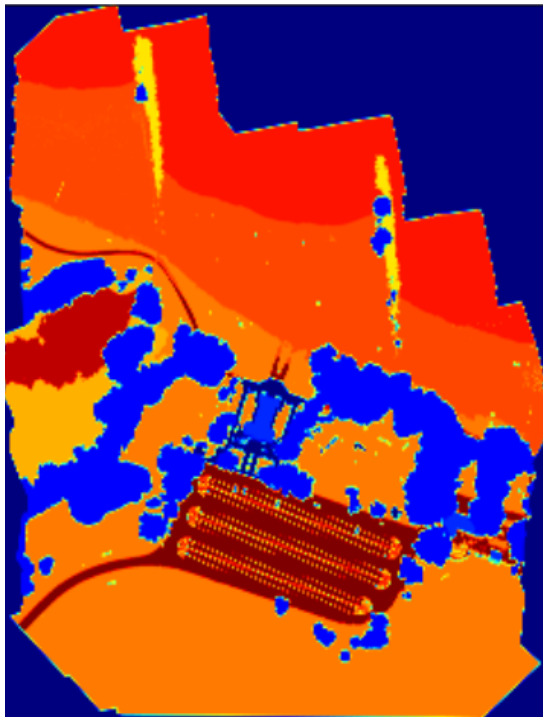
Results		
	Train	Validation
Accuracy	84.2%	72.8%
Precision	92.1%	78.4%
Recall	74.8%	65.8%
OneHotMean IoU	29.8%	21.2%
Focal Loss	0.005	0.015

- Early stopping: patience = 10, monitor = validation loss
- Epochs = 50
- Optimizer = Adam
- Learning rate = 0.001
- Bach size = 8

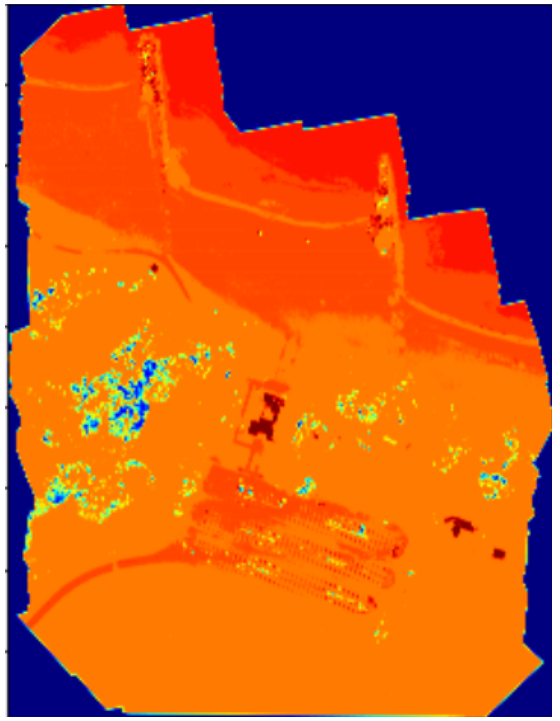
RESULTS – VALIDATION IMAGE



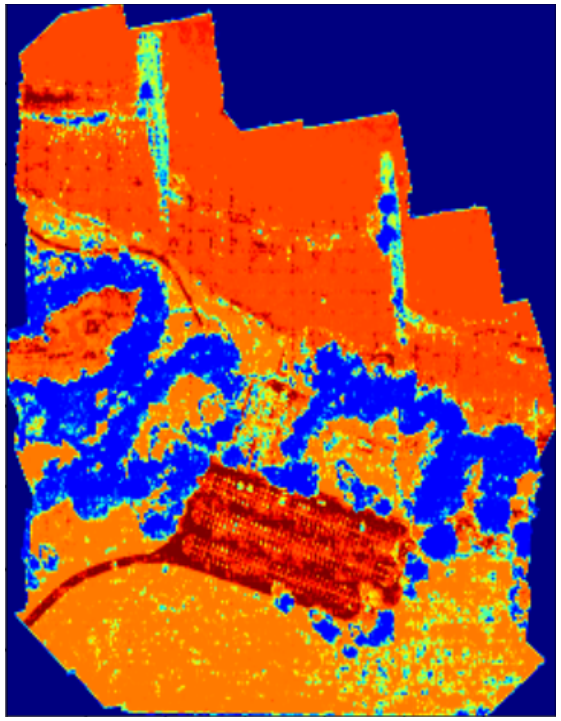
Validation Image



Mask



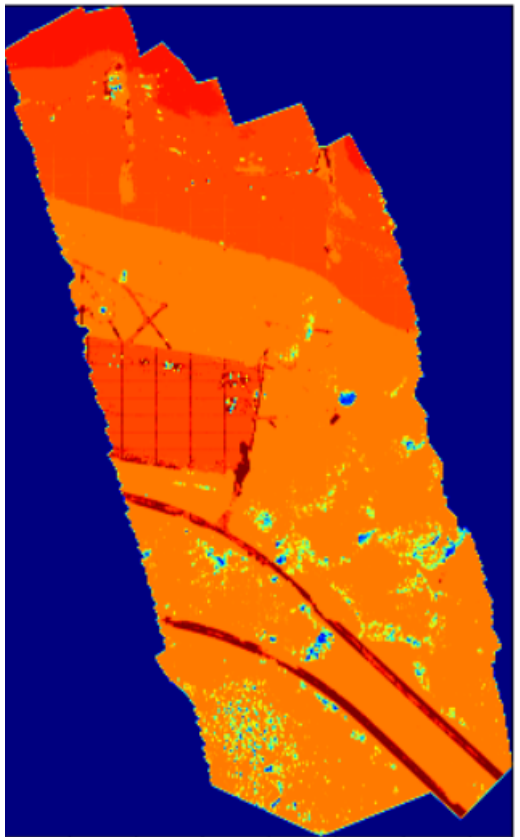
Prediction U-Net 1



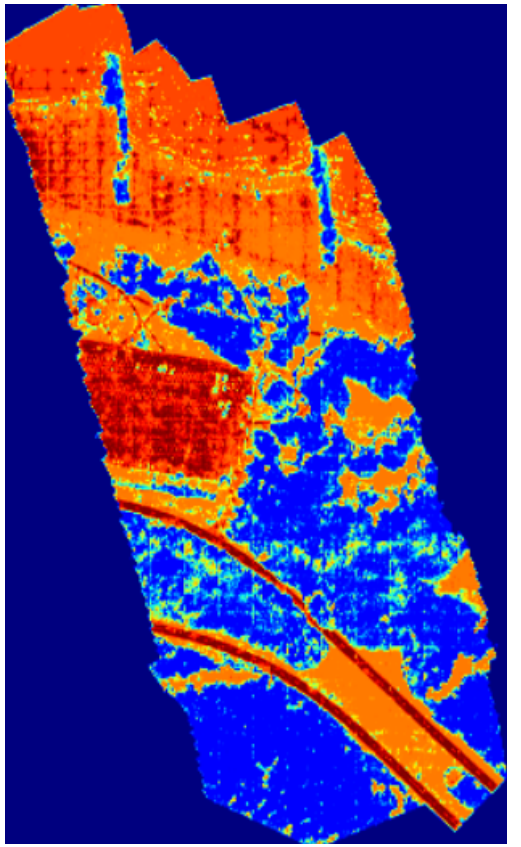
Prediction U-Net 2

RESULTS – TEST IMAGE

Test Image



Prediction U-Net 1



Prediction U-Net 2

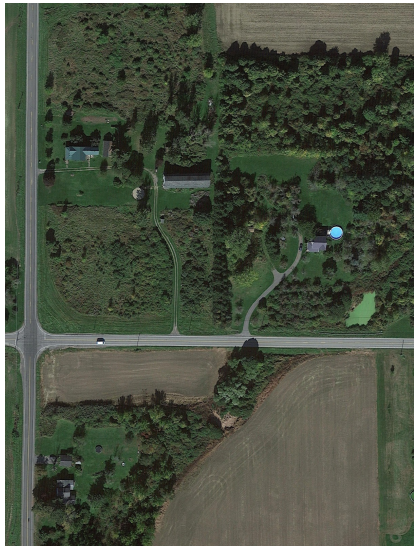
DATA AUGMENTATION - DEEP LEARNING APPROACH

A Generative Adversarial Network (**GAN**), defined as *pix2pix*, was trained with the previously obtained 1600 images.

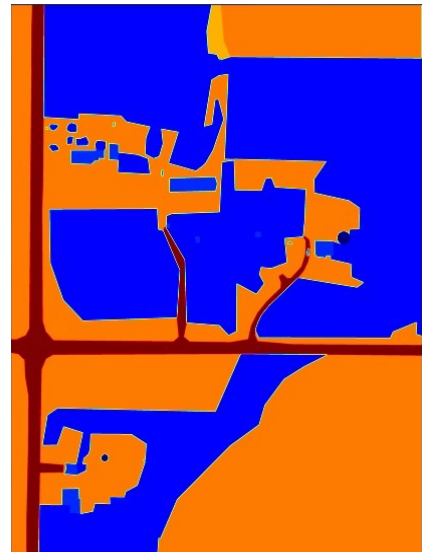
It is composed by:

1. A generator ("*the artist*")
2. A discriminator ("*the critic*")

This net was fed with a new mask cropped into 800 images of size 256x256x1, generating new data of size 256x256x6.



New image of the park downloaded from the web

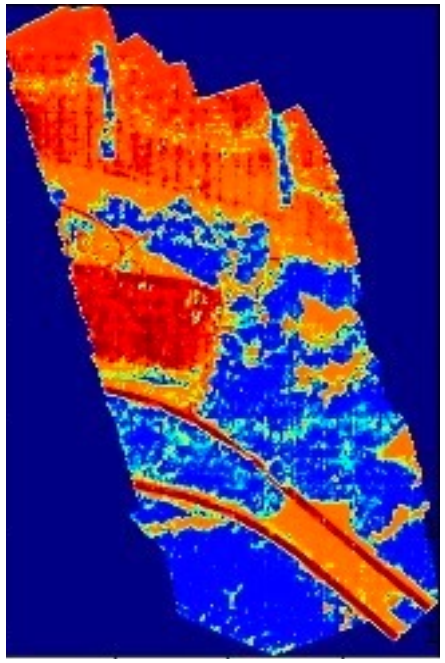


Corresponding mask obtained with the GAN

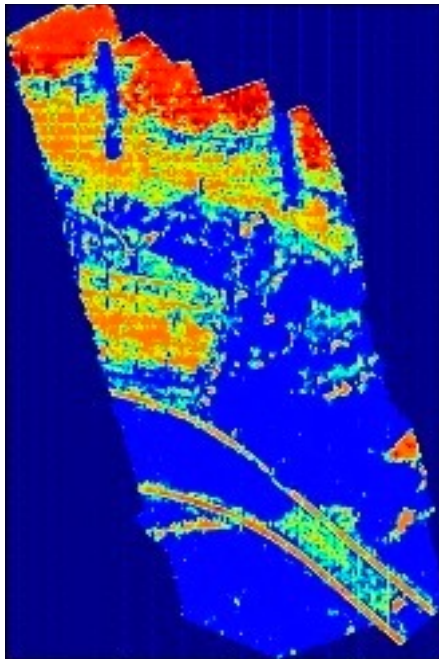
- Learning rate = 0.001
- Loss function = Cross-entropy based
- Epochs = 20
- Batch size = 1

FOURTH APPROACH
U-NET 2

U-NET 2 trained with the approach of **data generation.**



Prediction without data generation



Prediction with data generation

Results		
	Train	Validation
Accuracy	66.7%	52.1%
Precision	85.6%	61.0%
Recall	48.7%	43.6%
OneHotMean IoU	21.6%	12.9%
Focal Loss	0.010	0.025

- Early stopping: patience = 10, monitor = validation loss
- Epochs = 50
- Optimizer = Adam
- Learning rate = 0.001
- Bach size = 8

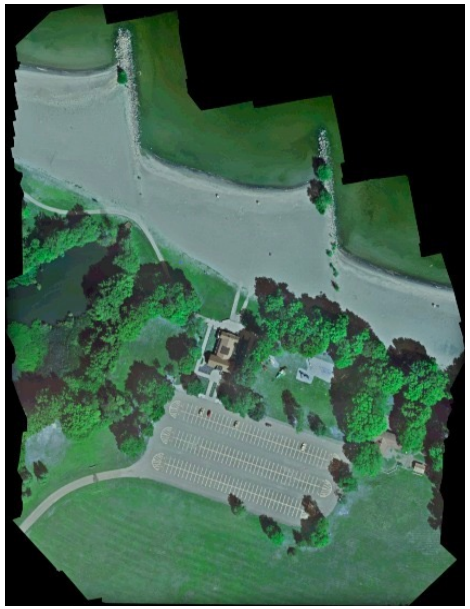
CONCLUSIONS

Best model: **U-NET 2 without data generation and a learning rate of 0.001**

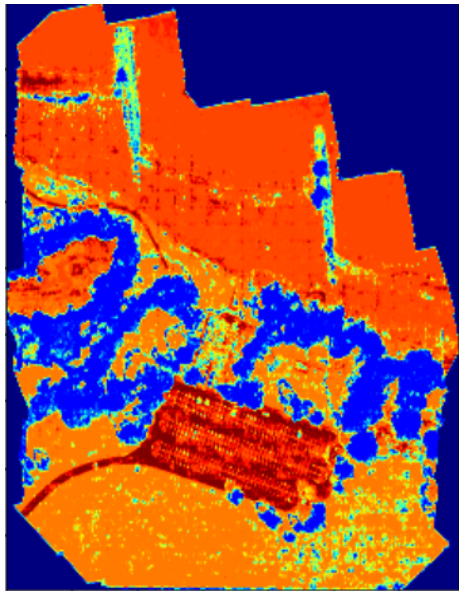
- Accuracy = 72.8%
- OneHotMean IoU = 21.2%
- Focal Loss =0.015

Finally, the percentage of **vegetation cover** was evaluated:

- Validation image: 52.06%
- Predicted image: 56.31%



Validation Image



Segmented Validation Image

Thank you for your attention!

s.grosso9@campus.unimib.it
n.rocchi@campus.unimib.it
j.buixuan@campus.unimib.it