TU Delft

A Confidence-aware Deep Learning Framework for Refining Laser-scanned Point Cloud Classification

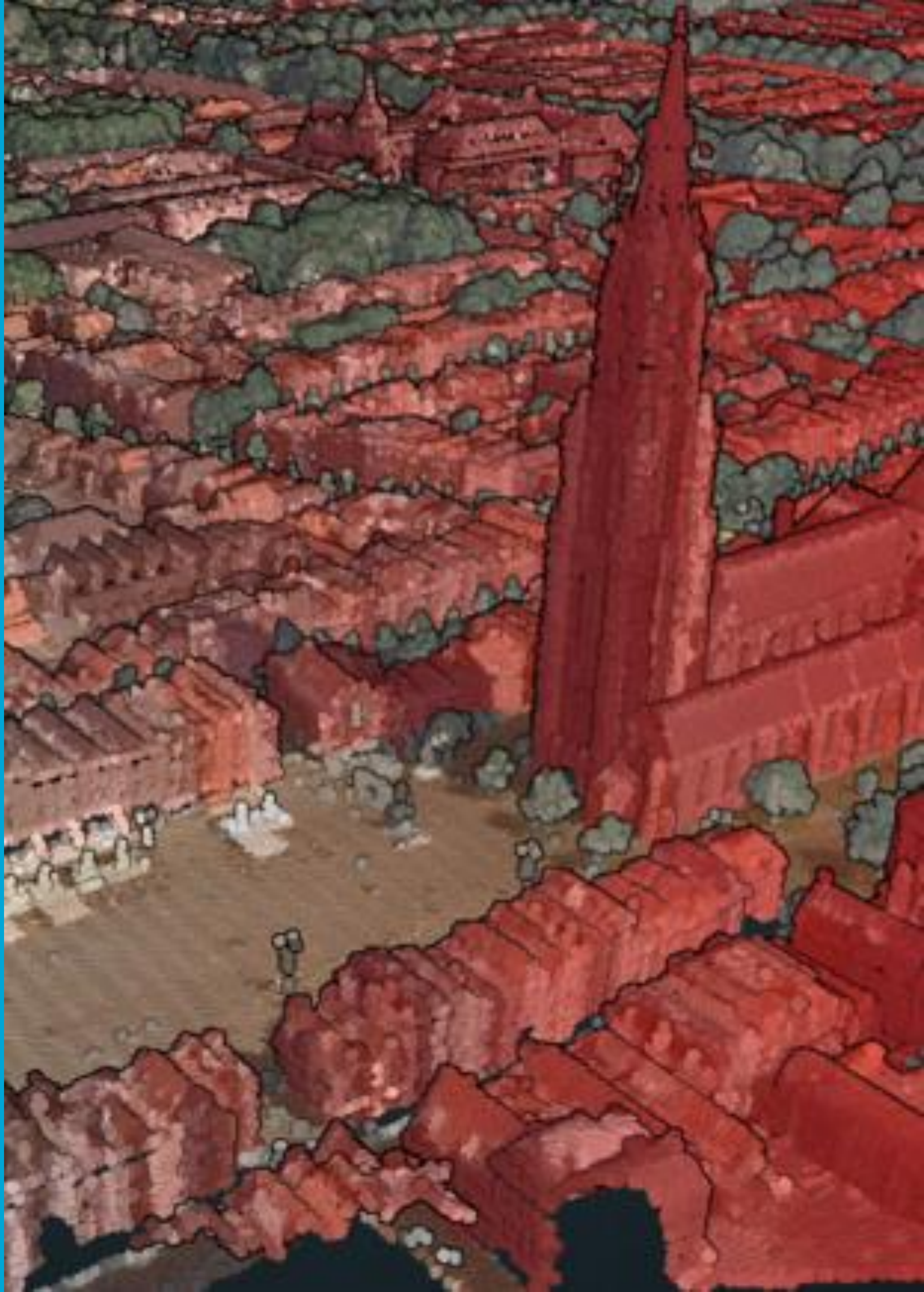
Sharath Chandra

Supervisors

Shenglan

Daan

Jantien



Content

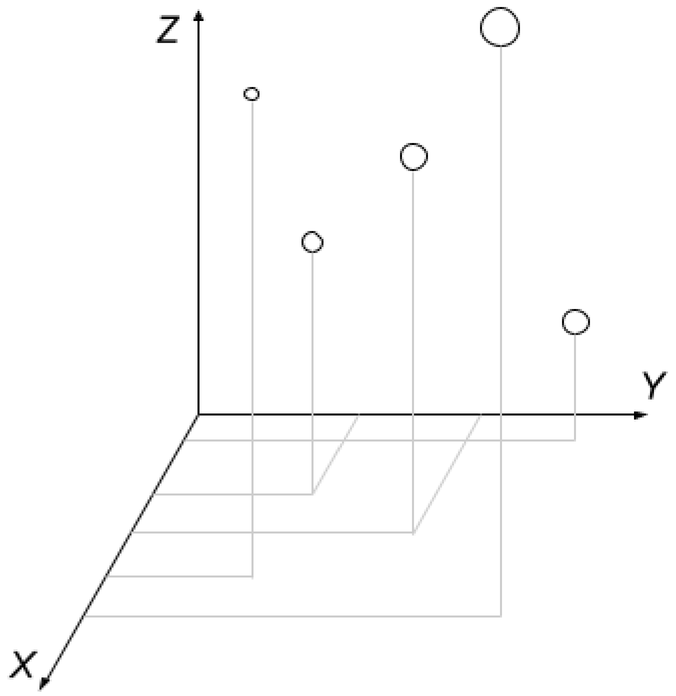
Introduction	Background	Method	Results	Conclusions
point cloud segmentation what it means	*automatically* deep learning	*online learning*	*Segmentation & online updates* performance	*Interpretations* Reasoning

01

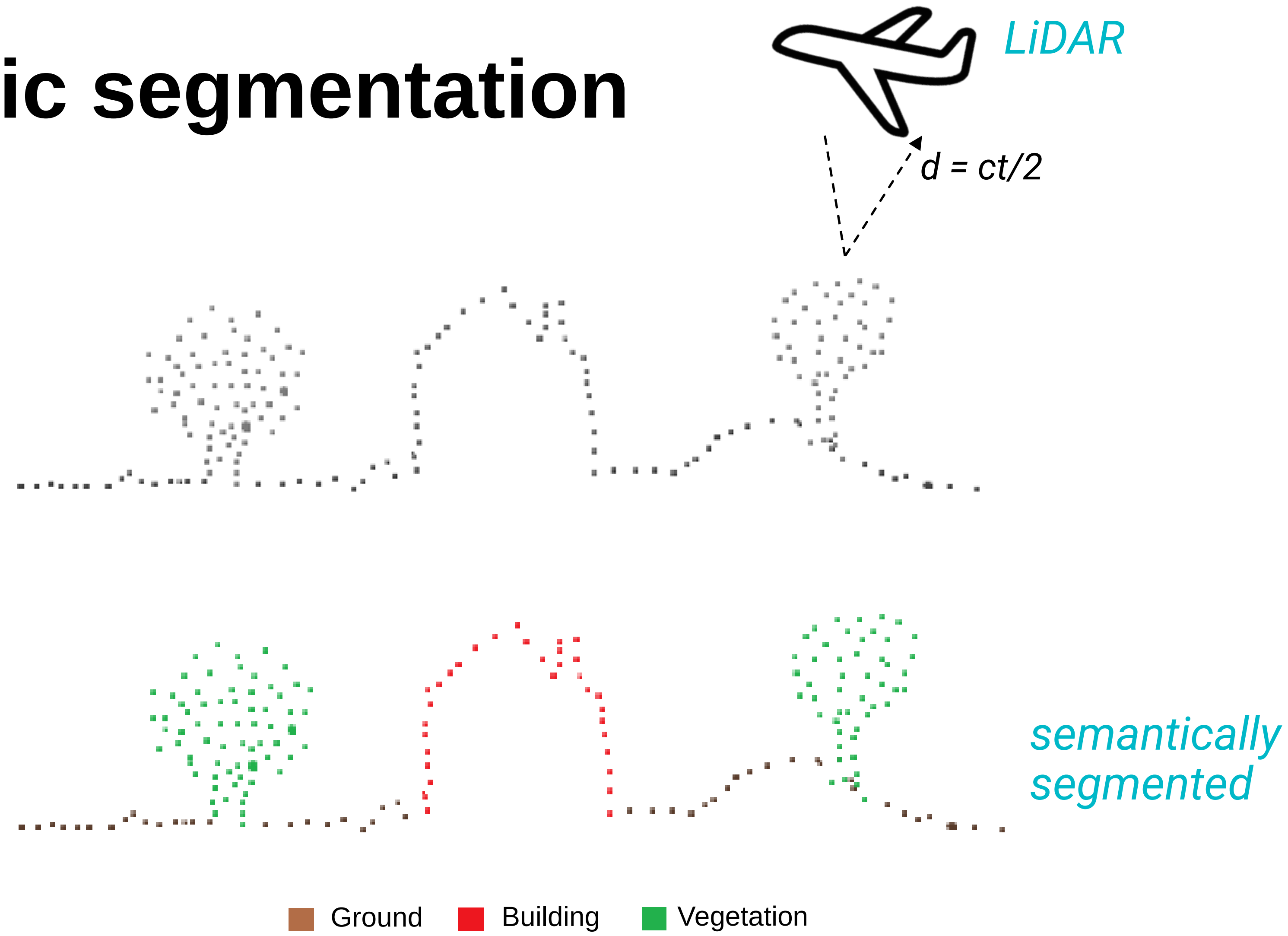
Introduction

Point cloud segmentation

Point cloud semantic segmentation



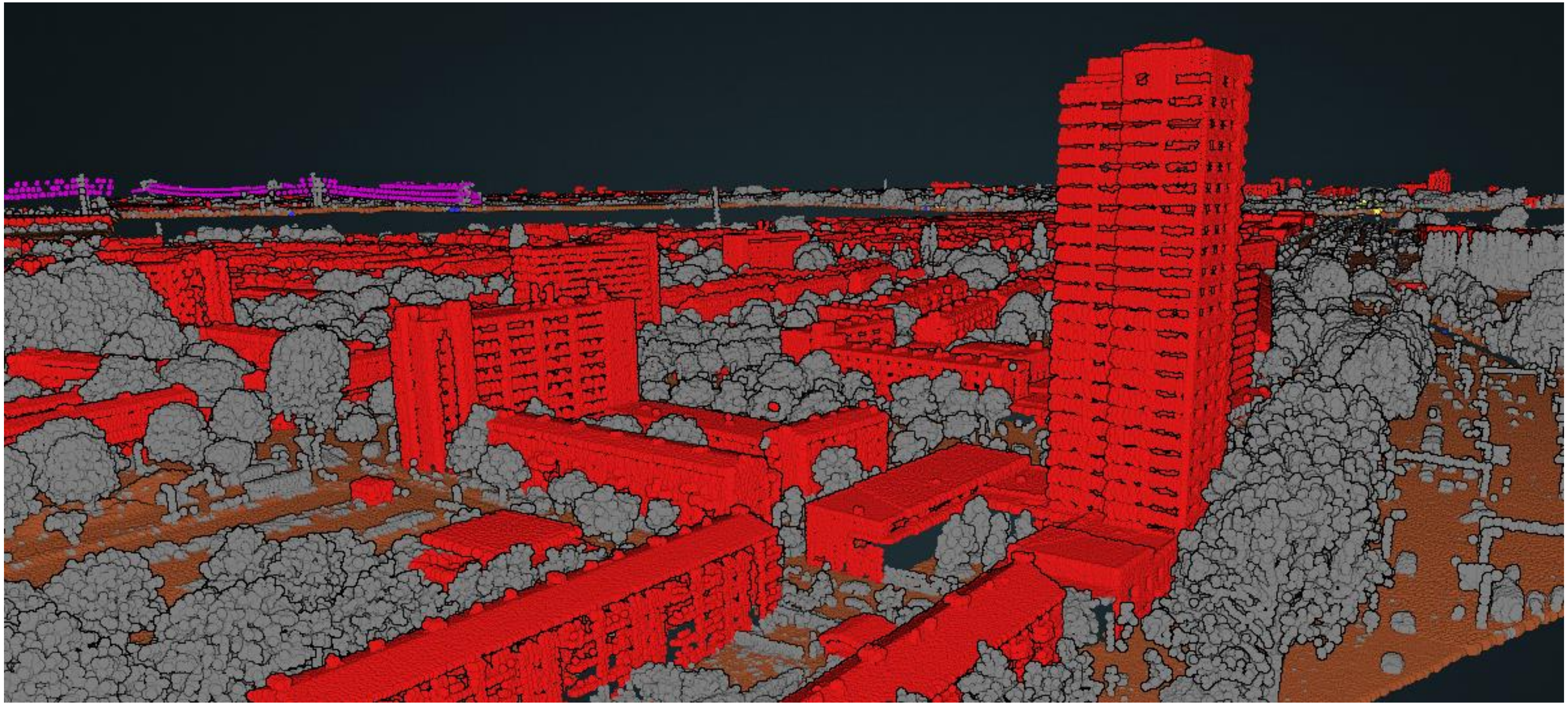
Points in 3D space



Point cloud semantic segmentation

Every point is given a class label

- Ground
- Building
- Civil
- Water
- Others
- High tension



AHN4

Point cloud classification

2D, 3D modelling

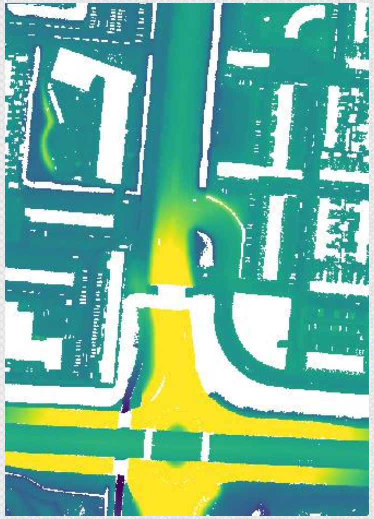
- DTM (from ground)
- DSM (-water)

Digital Twins

- 3D BAG

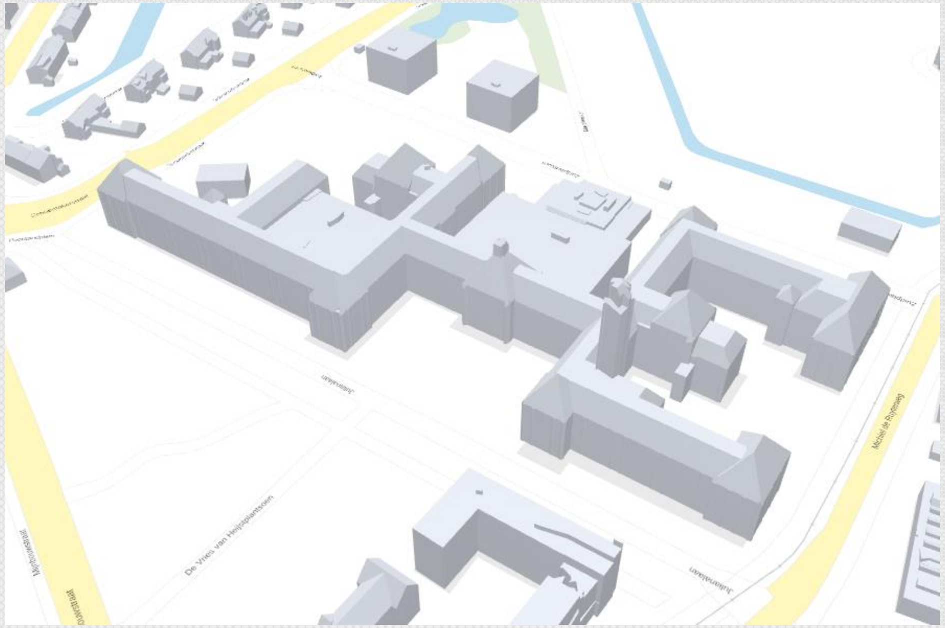
Environment mapping

- Forest
- Coastline



DTM

DSM



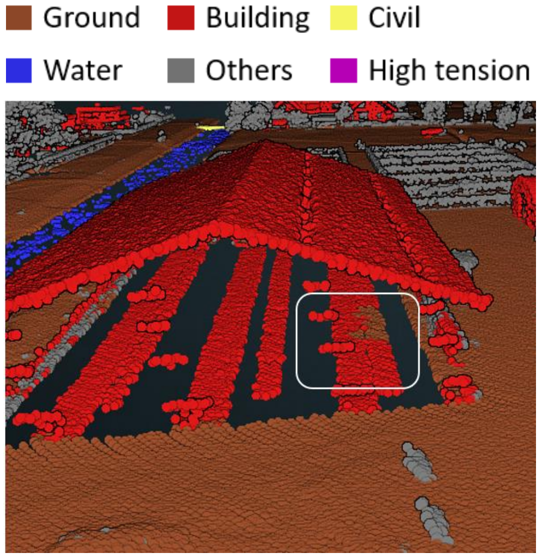
3D BAG

TU Delft

Point cloud classification

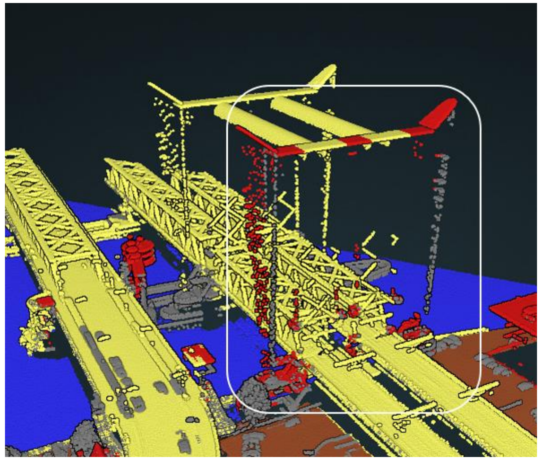
If its wrong!

- Water through buildings
- Buildings on bridges!



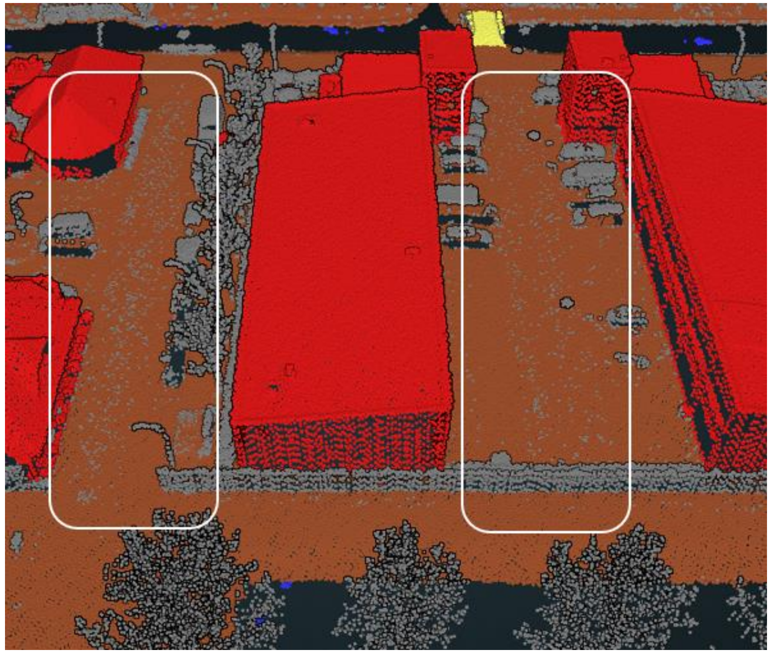
DTM

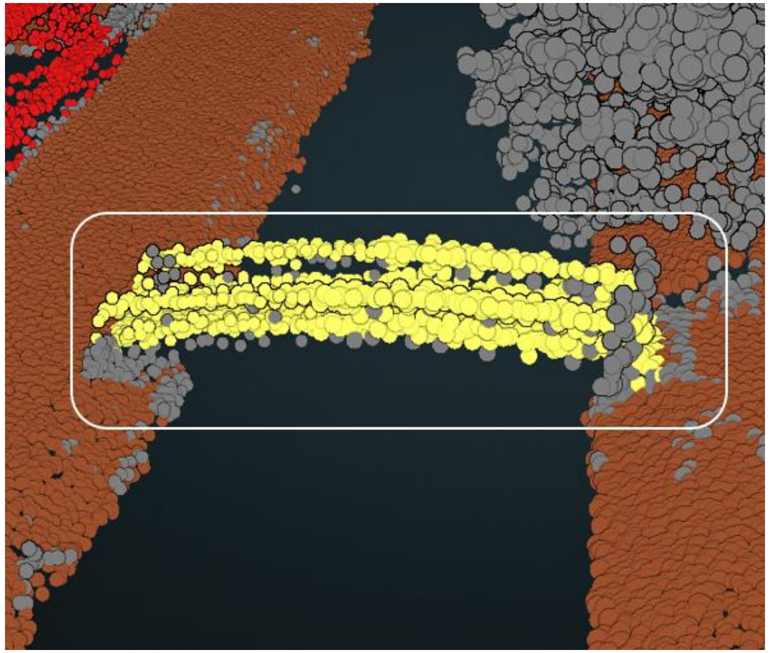
RGB

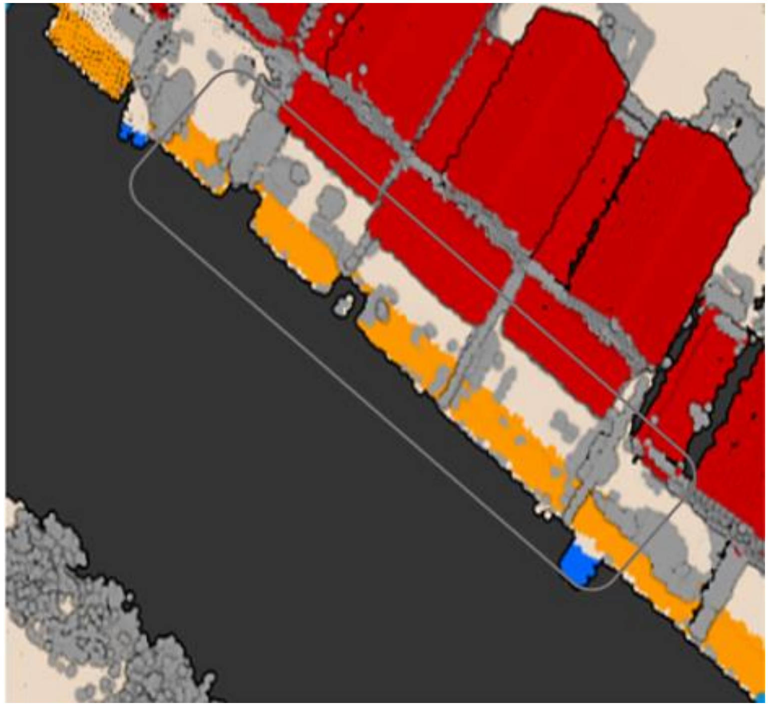


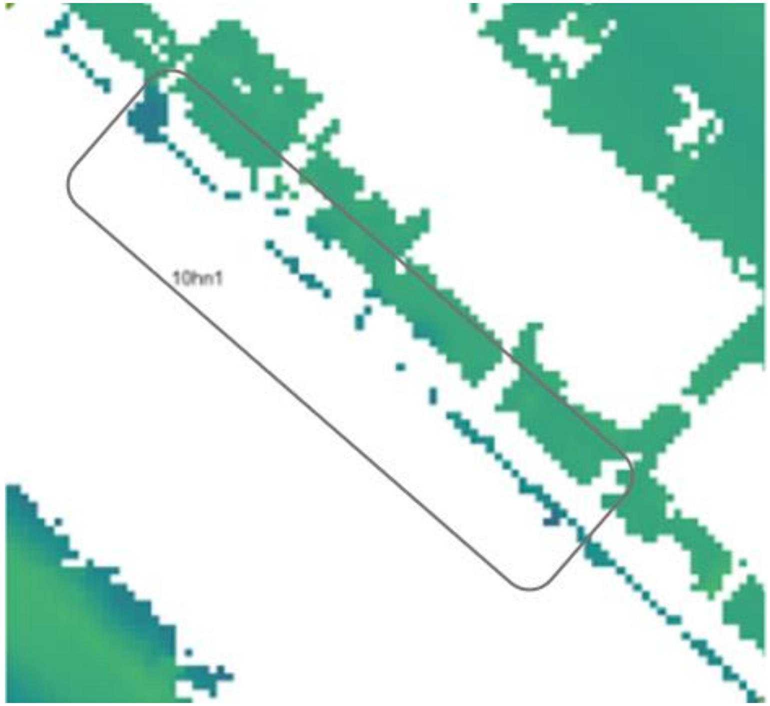
Segmented PC

also noisy labeling









Research goal

How to develop a DL framework to automatically improve the existing classifications of laser-scanned point cloud data by correcting misclassifications?

1. How to incorporate *geospatial knowledge into a DL framework*?

2. Can *Online Learning Strategy enhance the model's ability* to correct misclassifications and improve overall segmentation accuracy compared to traditional training approaches?

3. What is the impact of incorporating *additional spectral features* (such as *NIR* and *RGB*) on the performance?

02

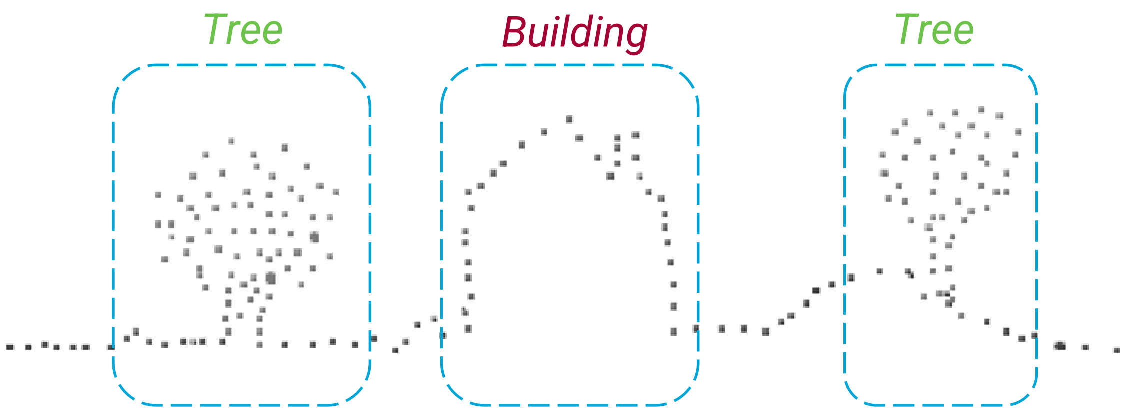
Background

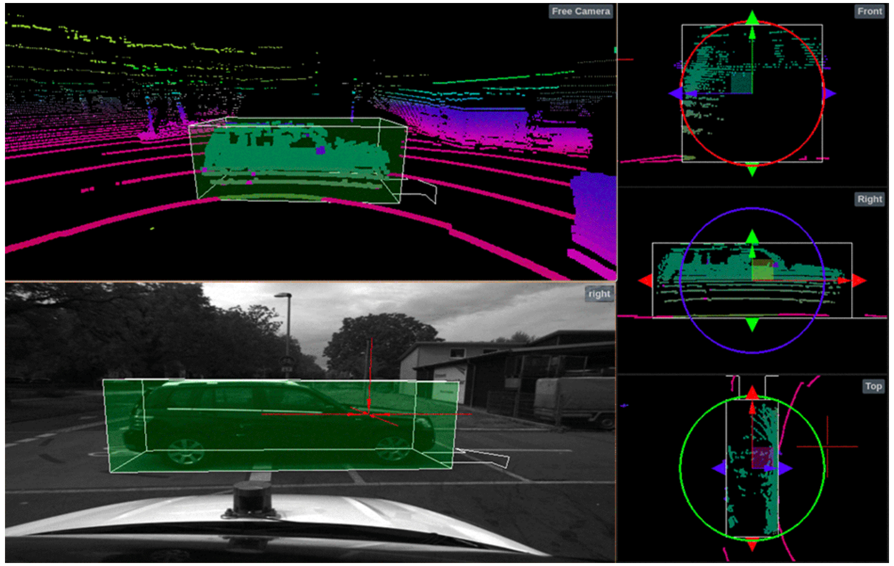
Automatic segmentation

Traditional

Bounding boxes

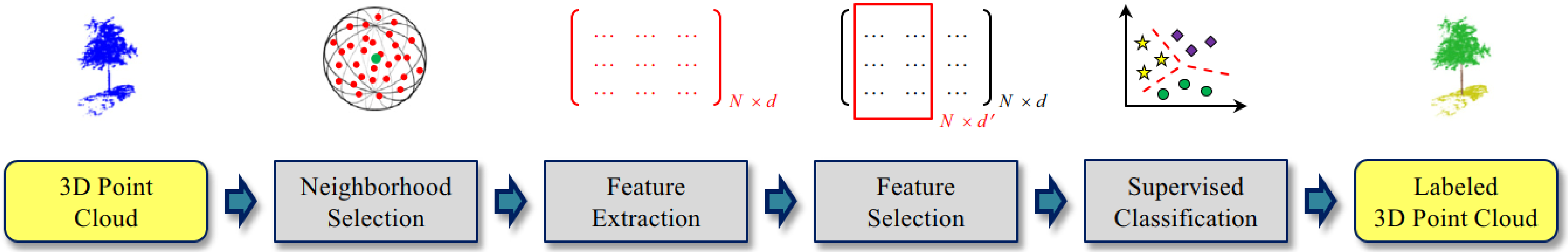
Takes a lot time





Img source: understand.ai

Machine learning

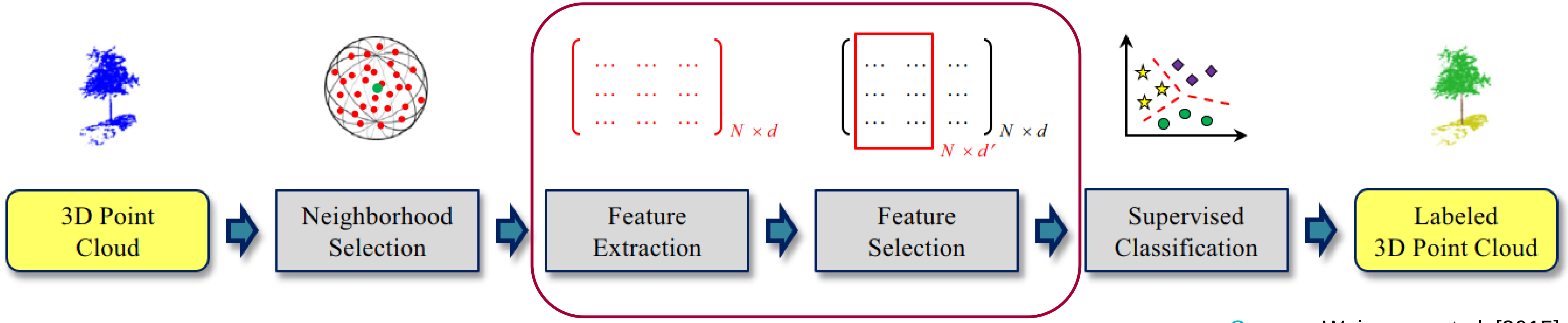


Source: Weinmann et al. [2015]

SEMANTIC INFORMATION!!!

Machine learning

SEMANTIC INFORMATION!!!



Source: Weinmann et al. [2015]

Raw data

+

Information to understand it

Human intervention

- *curvature*
- *normals*
- *shape descriptors*

Deep learning

Less human interaction – automatic features extraction

- *Data is fuel but Scarce*

Normal DL

- *Good performance*
- *Lots of good training data*

Data Efficient DL

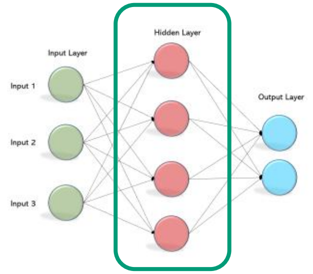
- *Make use of limited training data*

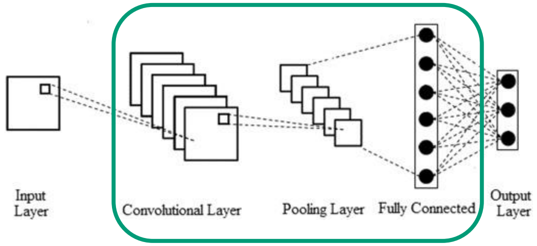
Deep learning

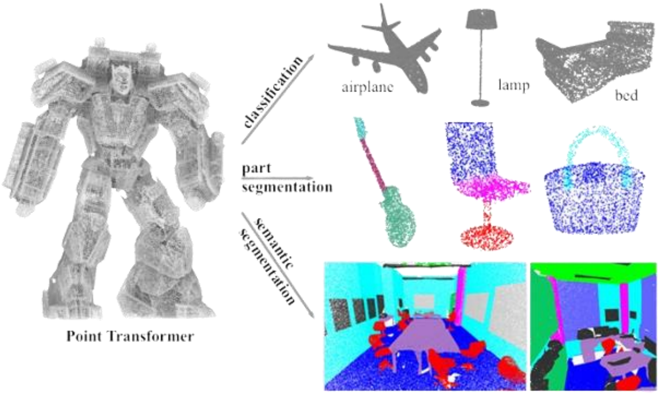
Less human interaction – automatic features extraction

Normal DL

1. *Multi Layered Perceptrons – Basic NN*
   - *Ex: PointNet, PointNet++*
2. *Convolution - Images*
   - *Ex: PointCNN, KPConv*
3. *Transformer - NLP*
   - *Ex: Point Transformer, Superpoint Transformer*







Img source: Zhao et al., 2020

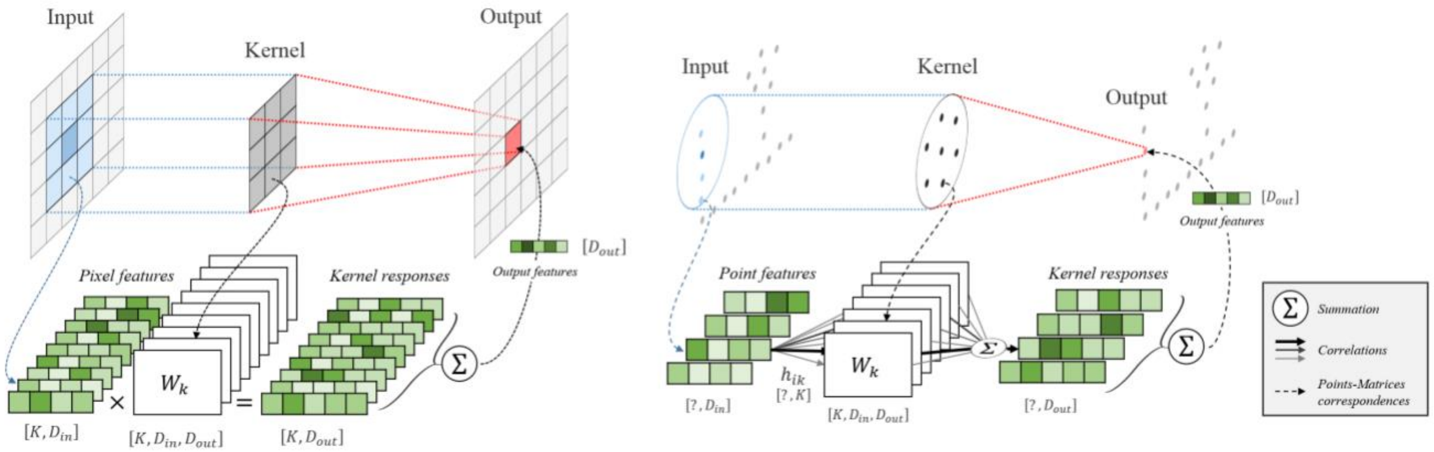
KPConv

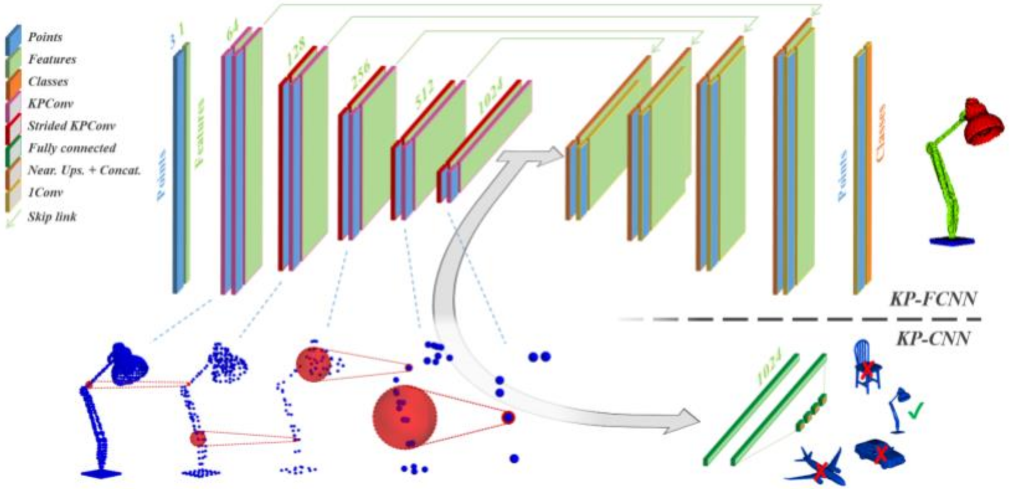
Kernel Point Convolution

Inspired from image CNNs

Backbone!

- *Tradeoff – performance & resources*





Source: KPConv, Thomas et al. [2019]

Data efficient methods

Make the MOST out of limited Training Data

Approaches

1. *Transfer learning*
   - *Finetuned to smaller datasets*
2. *Semi-supervised*
   - *Little labeled - lot of unlabeled data*
3. *Self-supervised*
   - *No labeled data – gives its own labels*

Data efficient methods

Make the MOST out of limited Training Data

Self-training

Progressively expands the limited training data

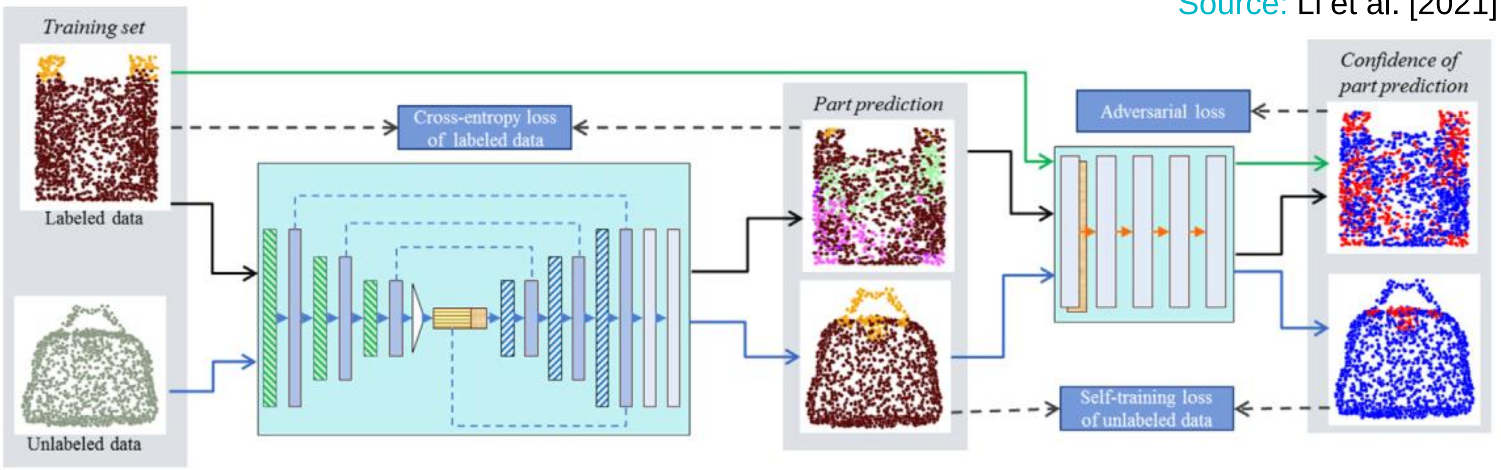
GAN

2 Networks in parallel

Heavy!

GAN

Source: Li et al. [2021]

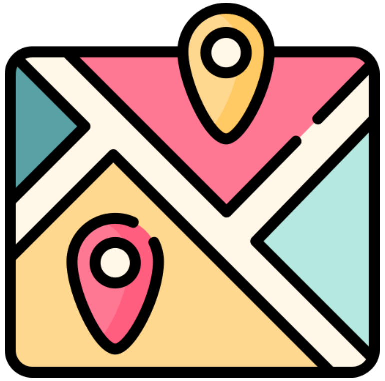


Segmentation network makes predictions

Discriminator network judges predictions

Idea of our approach

- *To keep the network simple, but with the benefits of Data efficient models*



- *Incorporate geospatial knowledge*
- *To have one network*

Light weight!

03

Method

Network training - *Online strategy*

Method

GOAL: DL framework to learn from accurate labels by correcting misclassifications?

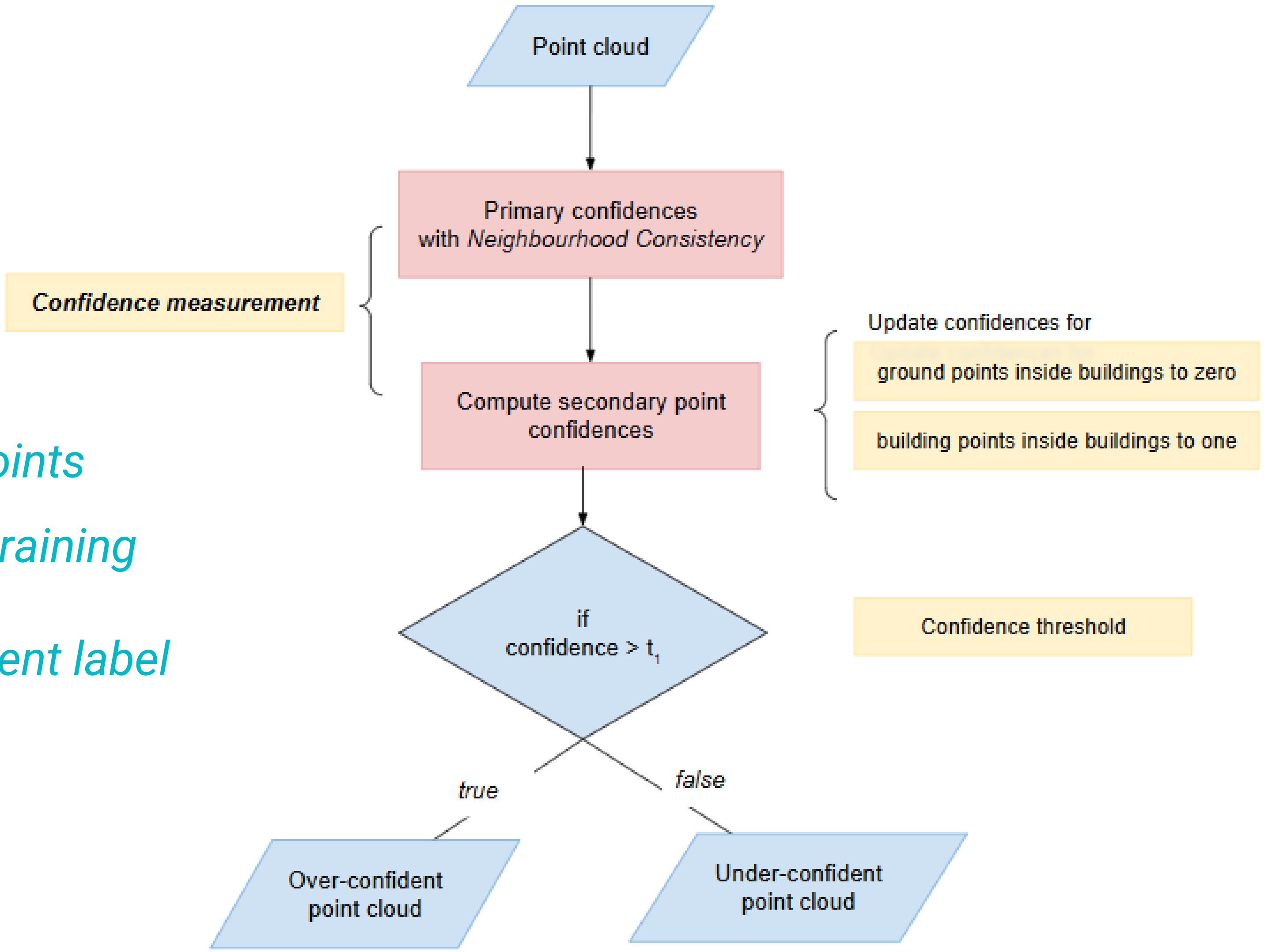
1. *Preprocessing*
    - Separate good from bad samples

2. *Online deep learning*
    - Learns from good labels
    - Correct the bad ones

1 Preprocessing

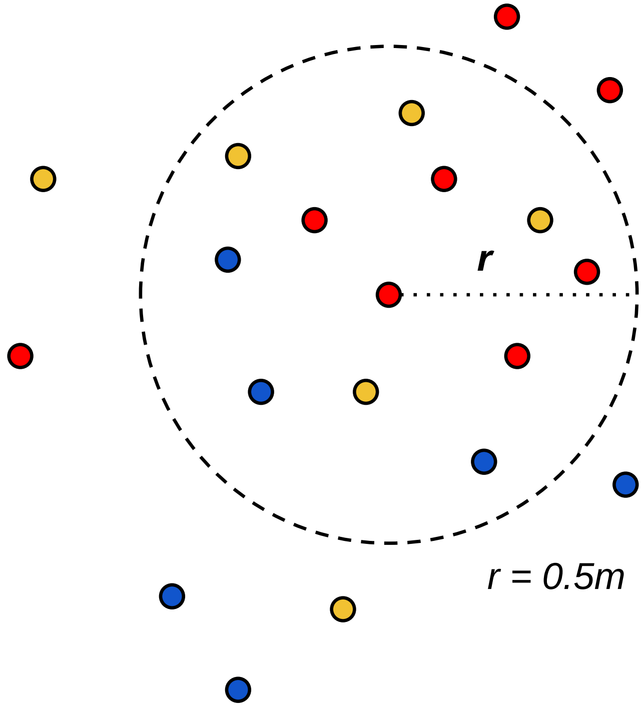
Separate good from bad samples

- *Confidence scores for all the points which decides Participation in training*
- *How confident we are with current label*



TU Delft

1 Preprocessing



1. Primary Confidence

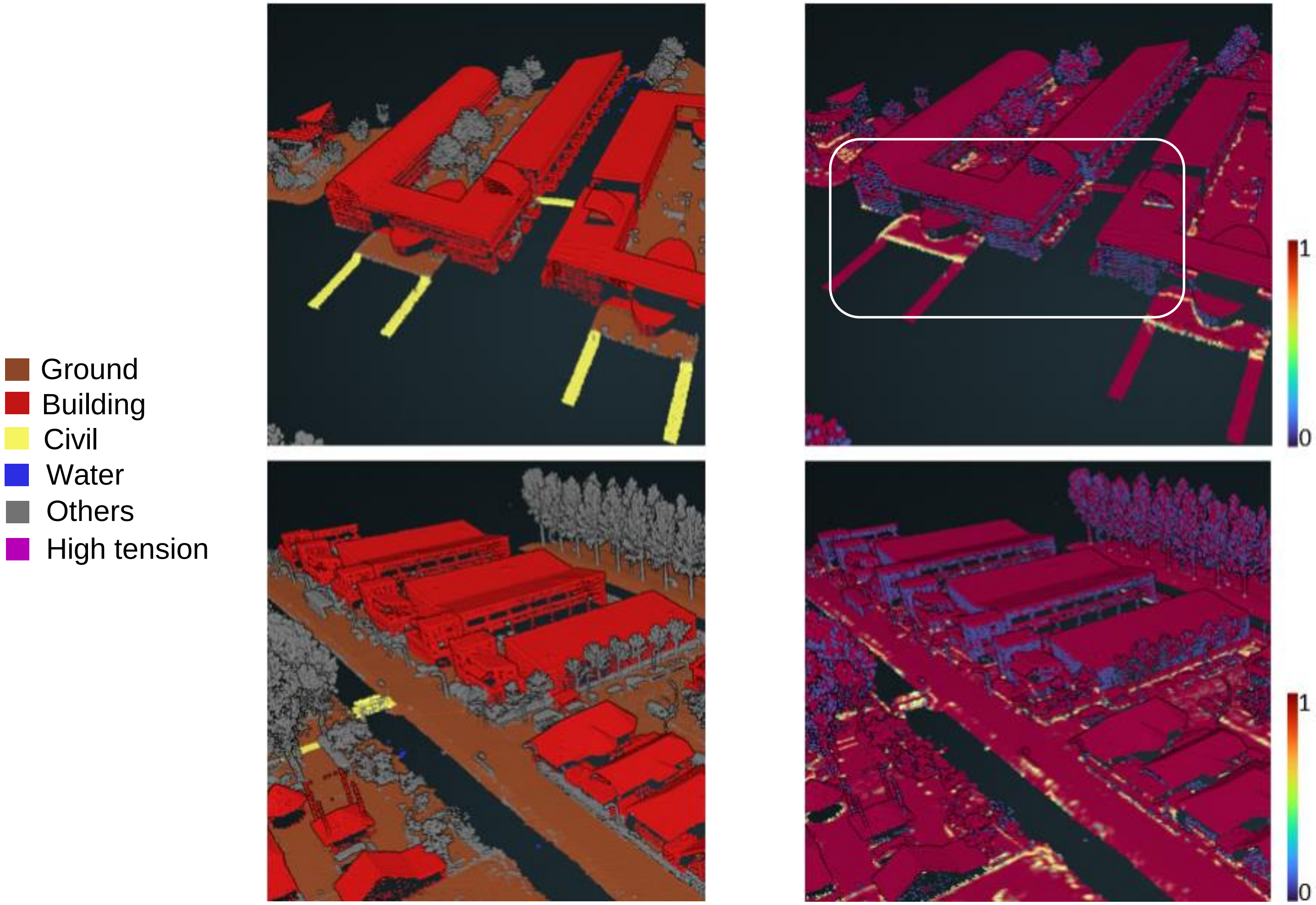
Neighborhood consistency

how well a point is surrounded by points of same classification

$$C = \begin{cases} \frac{N_{sameclass}}{N_{total}} & \text{if } N_{total} >= 5, \\ 0 & \text{if } N_{total} < 5 \end{cases}$$

1 Preprocessing

Confidence scores



Labeled *Primary confidence*

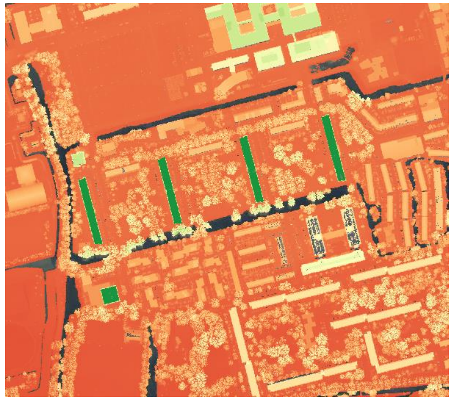
Problem

Building walls

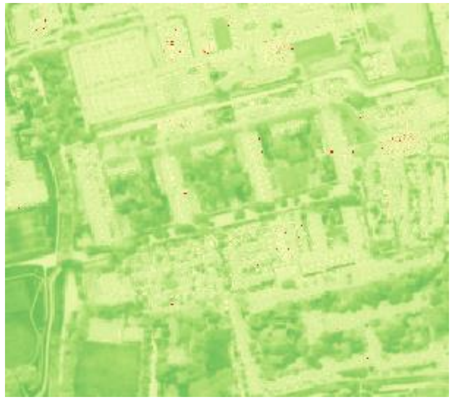
1 Preprocessing

2. Refining Confidence

RGB



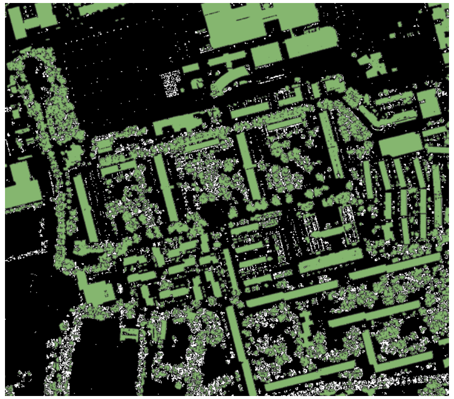
DSM -7.5 41.7



NDVI -1.0 0.95

Inputs

Binary map - polygonised DSM > 2m



After erosion & dilation

Mean NDVI -1.0 0.95

Buildings in ***Red*** *Mean NDVI < 0.3*

Output

Building footprints!

for refinement of

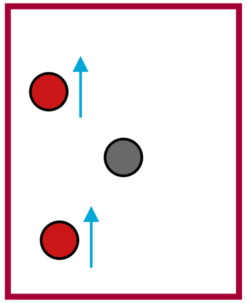
buildings with

additional sources of

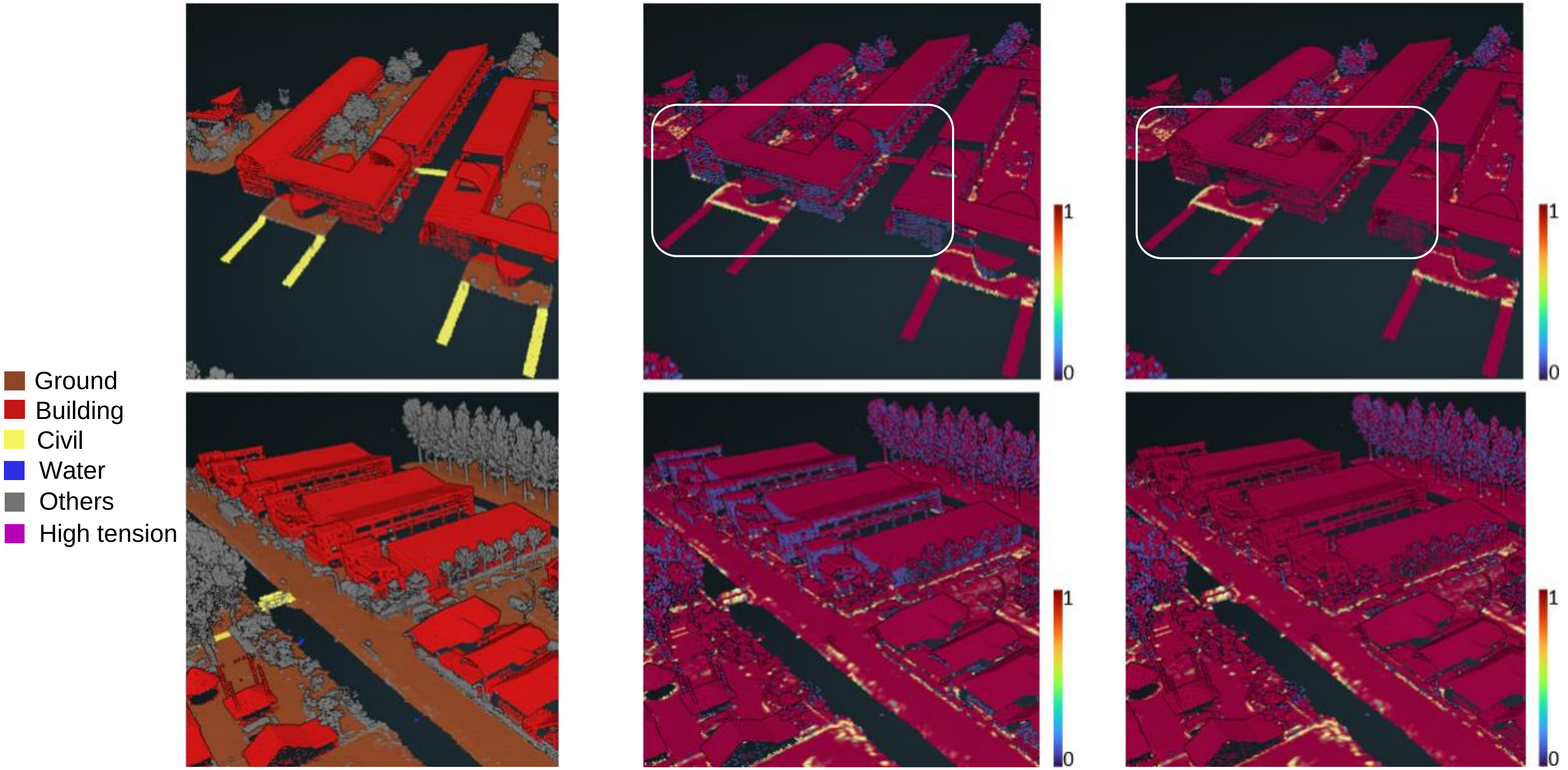
DSM & NDVI

1 Preprocessing

Confidence scores



Building footprint



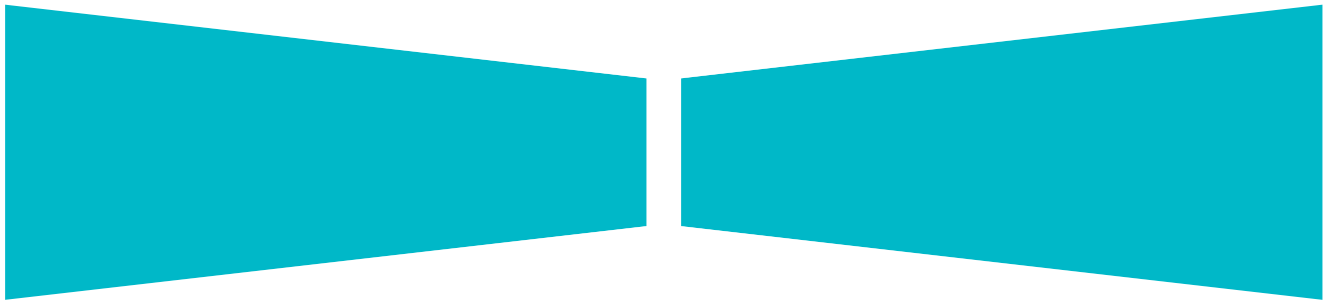
Labeled *Primary confidence* *Refined confidence*

TU Delft

2 Online deep learning

Deep learning

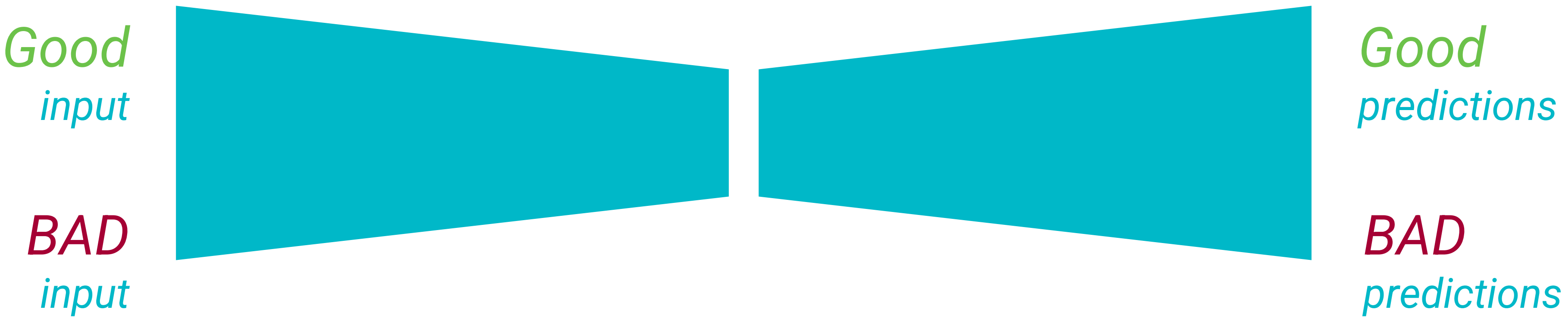
Good
input



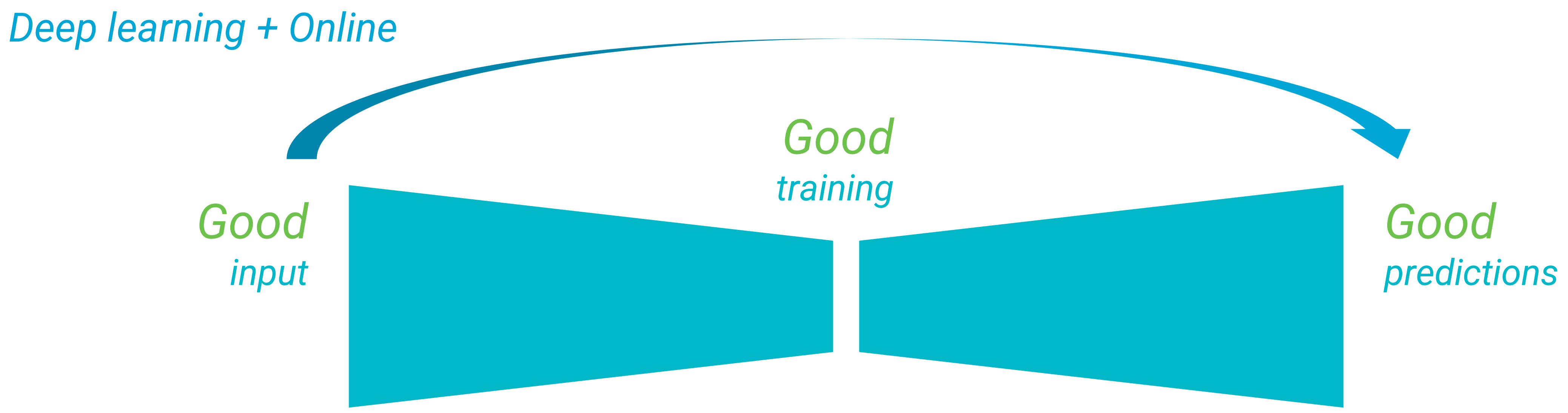
Good
predictions

2 Online deep learning

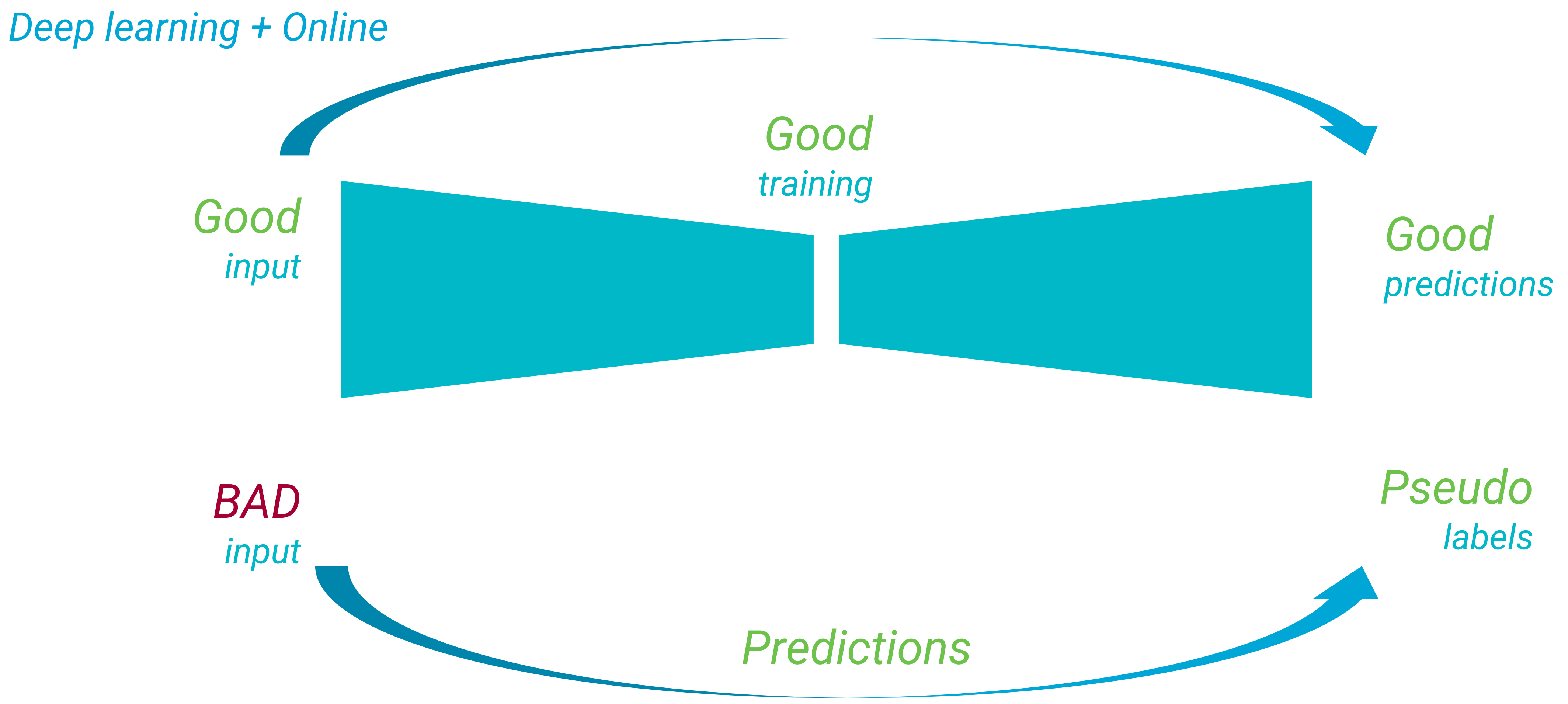
Deep learning



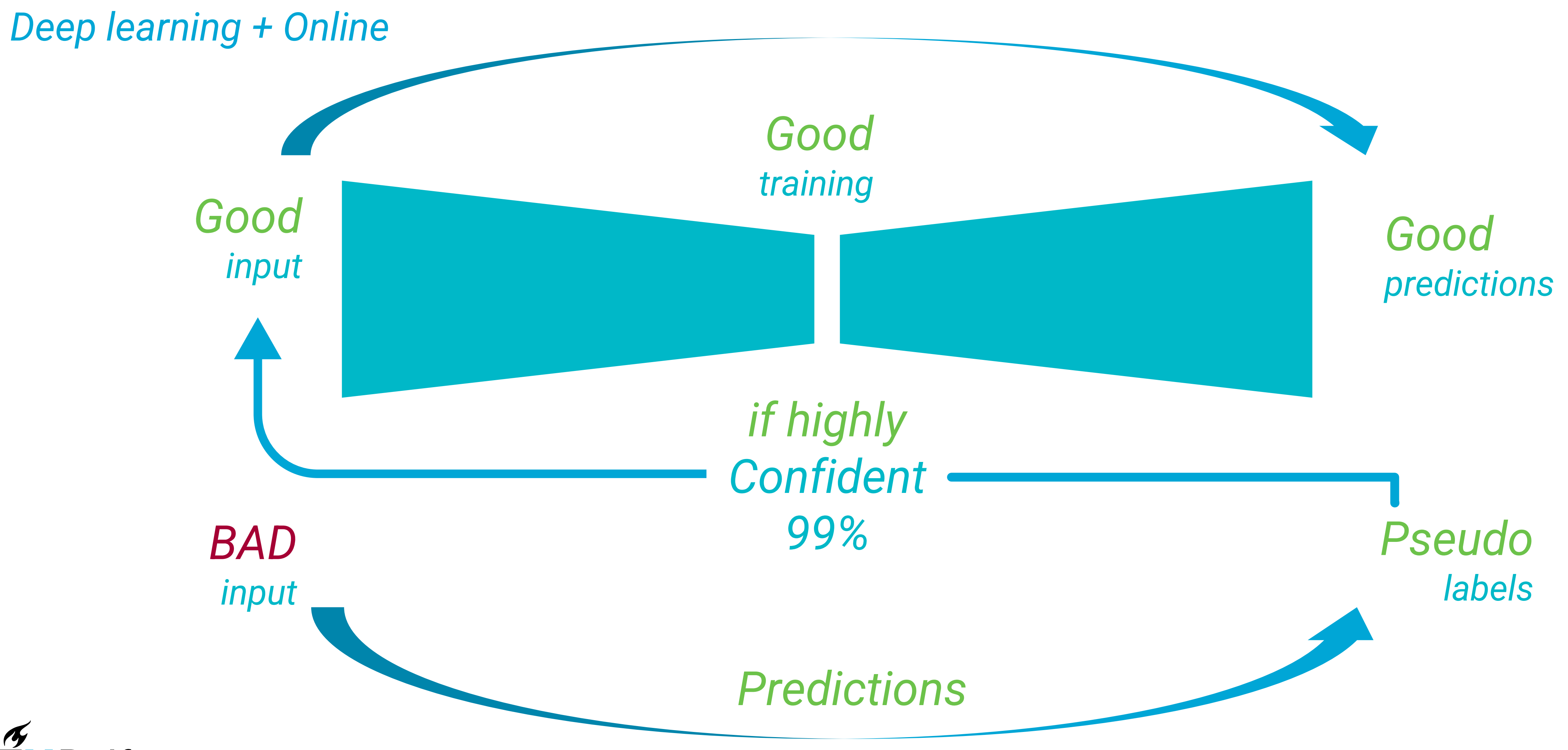
2 Online deep learning



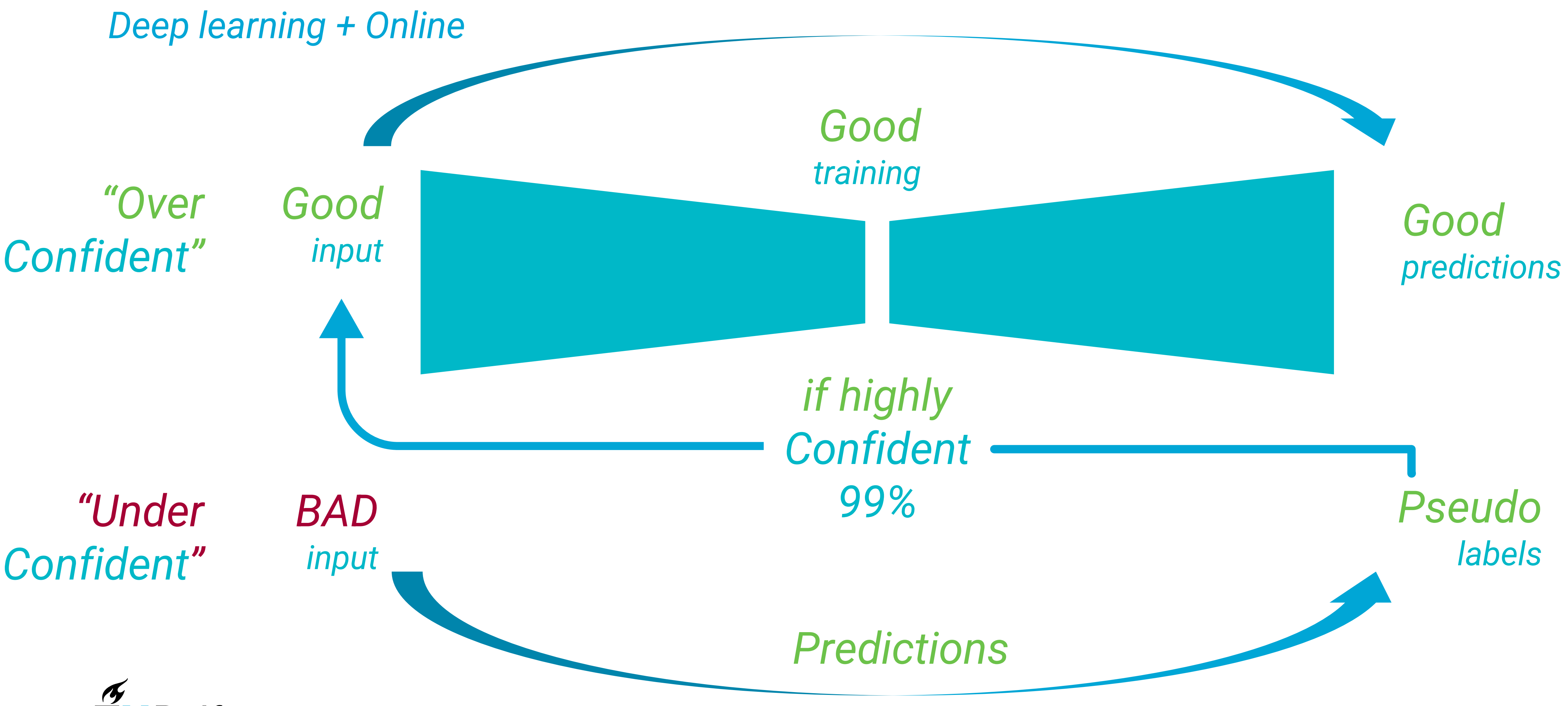
2 Online deep learning



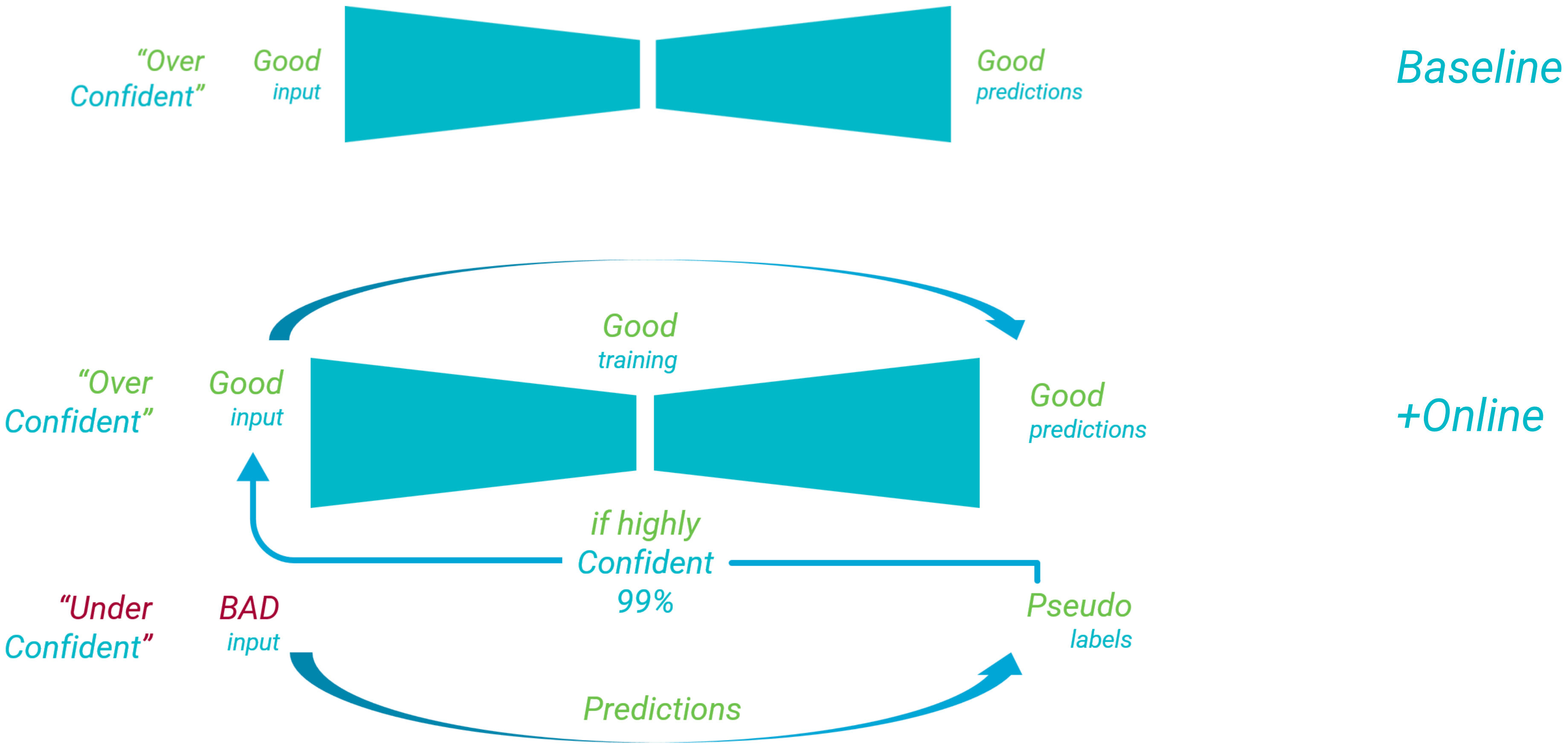
2 Online deep learning



2 Online deep learning



2 Online deep learning



04

Implementation

Point cloud, DSM, MSI

Data

Point cloud

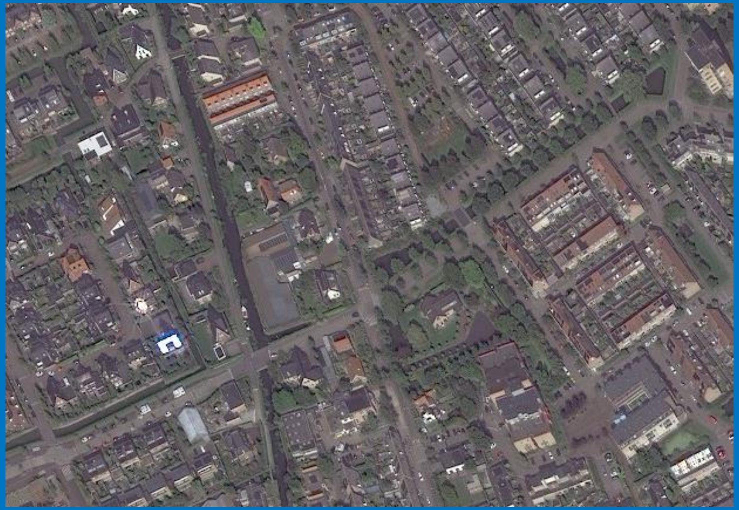
from GeoTiles

AHN4

DSM

from AHN,

0.5m



Aerial MSI

Pléiades Neo from NSO,

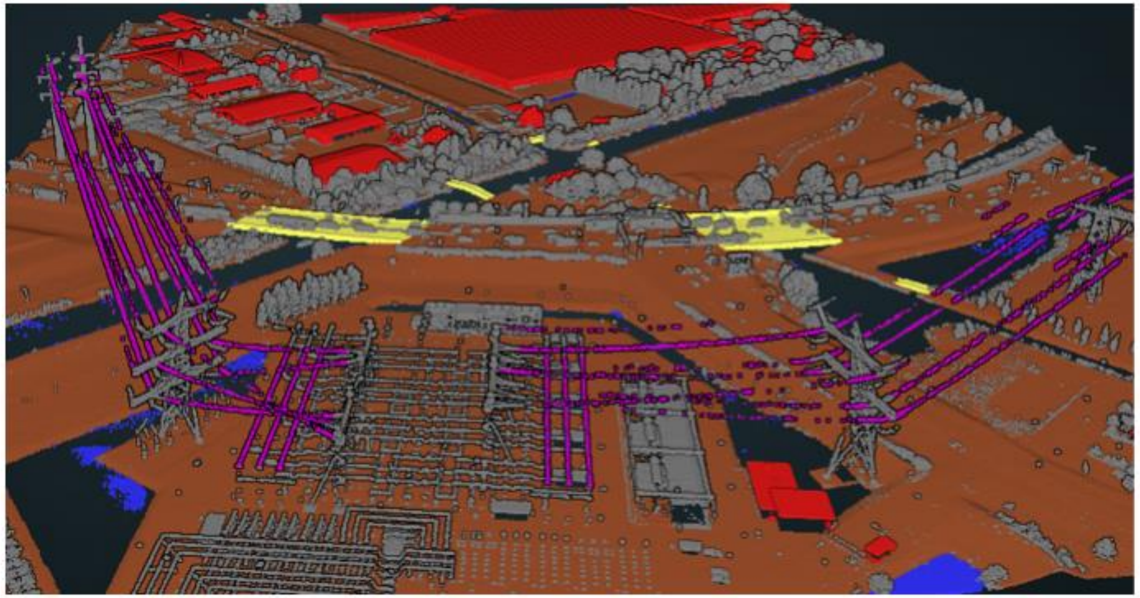
0.5m,

RGB+NIR

Data

Point cloud

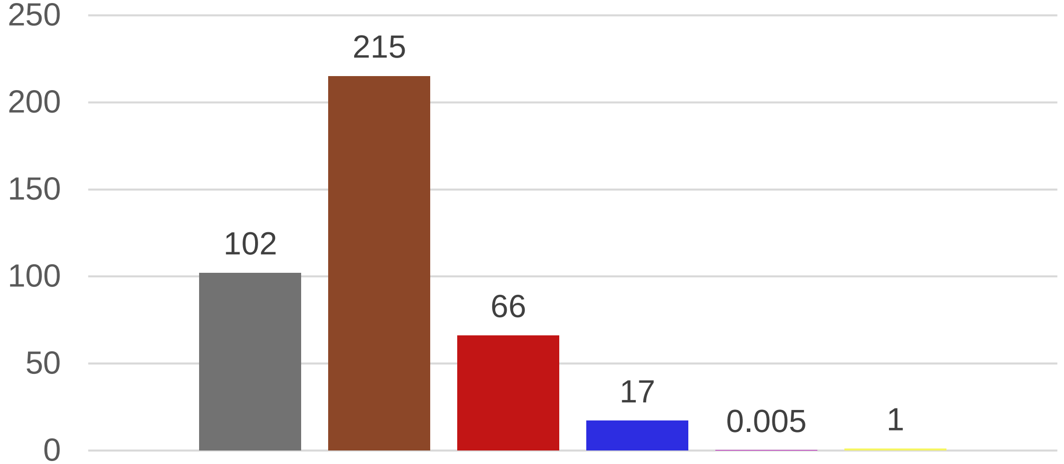
	Training	*Test*	
	52	*8 mini tiles*	*0.25 x 0.3125 km*
	~85%	*15%*	



■ Others ■ Ground ■ Building ■ Water ■ High tension ■ Civils

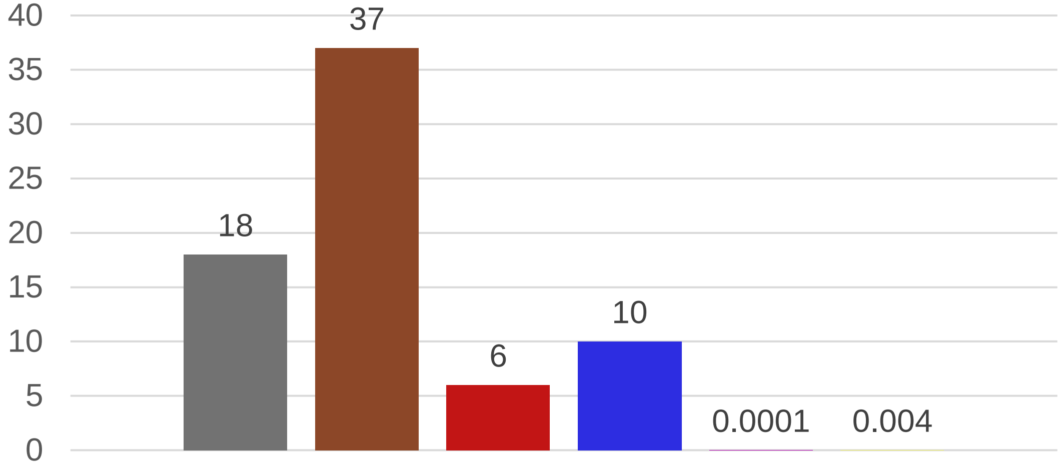
Training

** in millions*



Test

** in millions*



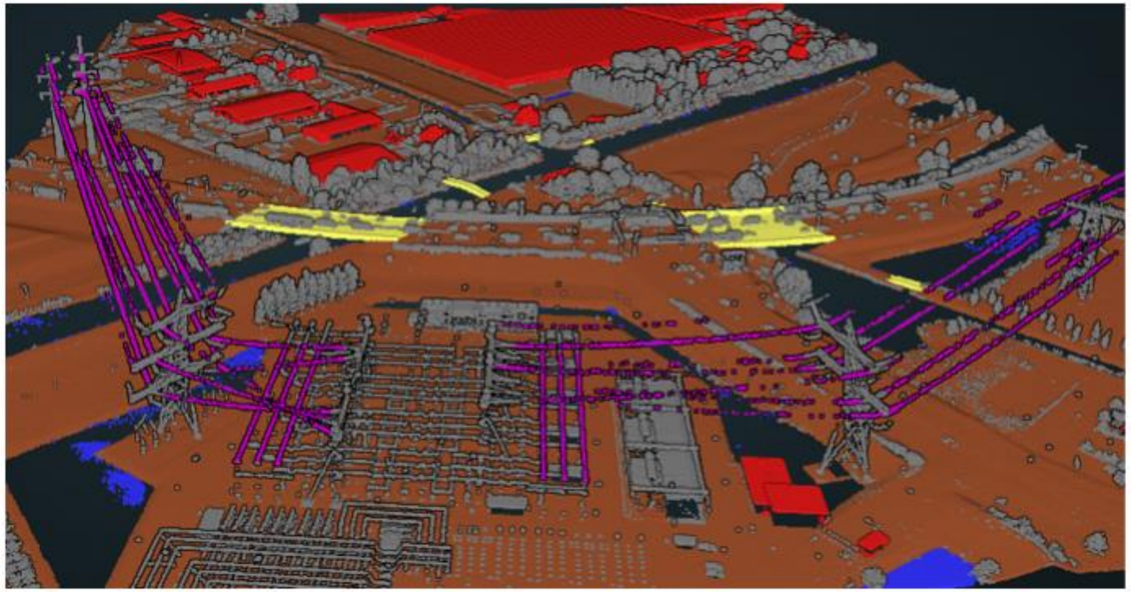
TU Delft

Network supervision

Loss

- *How far model's predictions from true values*

  Penalizing for incorrect predictions



■ Others ■ Ground ■ Building ■ Water ■ High tension ■ Civils

Weighted cross-entropy loss

$$p_c = \frac{n_c}{N}$$

$$w_c = \sqrt[3]{\frac{p_{max}}{p}}$$

$$L_{\text{cross-entropy}}(\hat{y}, y) = -\frac{1}{N}\sum_{j=1}^{N}\sum_{c=1}^{M} w_c y_{c,j} \ln(\hat{y}_{c,j})$$

More weight to minority classes

Hyperparameters

Preprocessing

Hyperparameter	Value
r	$0.5m$
t_1	0.9

Backbone

Hyperparameter	Value
N	300
Epochsteps	300
lr	0.01
in_radius	10.2
kernel points	15

+Online

Hyperparameter	Value
e	150
t_2	0.99

05

Results

Segmentation, online updates

Results with base features

Elevation + intensity - raw features (from LiDAR sensor)

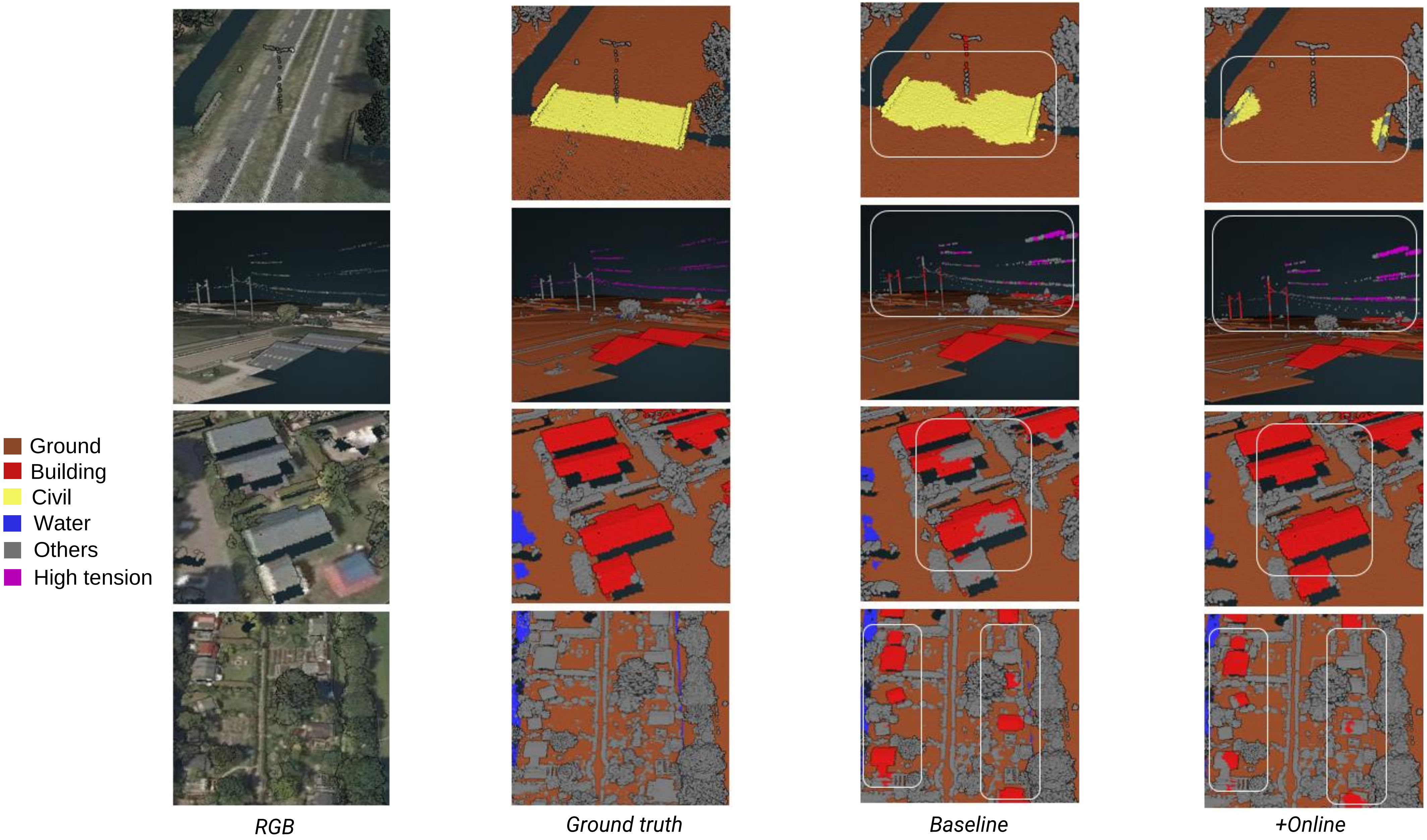
Model features: *elevation, intensity*

	Per class accuracies						mAcc	OA
	Others	Ground	Building	Water	High tension	Civil		
Baseline	89.8	**98.6**	78.1	**99.2**	32.2	**80.0**	**79.6**	94.8
+Online	**90.8**	**98.6**	**79.4**	**99.2**	**33.2**	75.3	79.4	**95.1**
	Per class IoUs						**mIoU**	
	Others	Ground	Building	Water	High tension	Civil		
Baseline	**86.6**	**94.8**	73.5	98.1	27.4	2.6	63.8	
+Online	85.4	**94.8**	**75.4**	**98.4**	**30.4**	**5.7**	**65.0**	

* All values represent the average of three experiments, ensuring fair comparison

Baseline

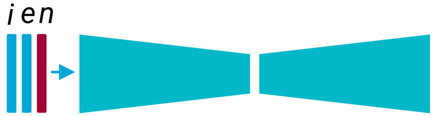
+Online


Ground
Building
Civil
Water
Others
High tension
RGB
Ground truth
Baseline
+Online

TU Delft

Results with additional NIR feature

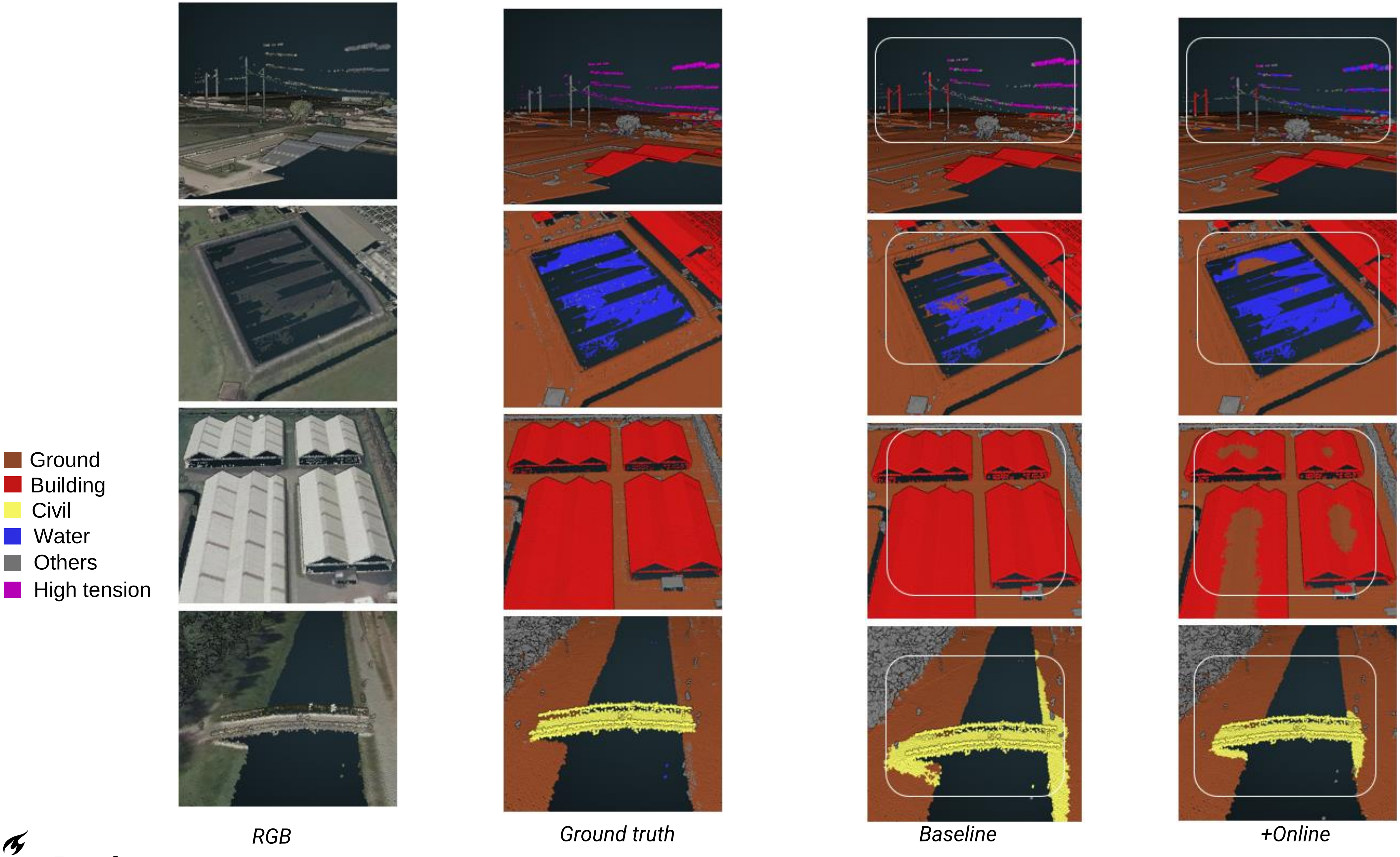
Elevation + intensity *+ NIR* *Additional information from aerial images*



Model features: *elevation, intensity, NIR*

	Per class accuracies						mAcc	OA
	Others	Ground	Building	Water	High tension	Civil		
Baseline	**91** ↑ (1.1)	98.4 ↓ (-0.3)	**73.5** ↓ (-4.6)	99.2 (0)	**44.8** ↑ (12.6)	**79.0** ↓ (-1.0)	**81** ↑ (1.3)	94.6 ↓ (-0.2)
+Online	90.8 (0)	**98.6** (0)	72.9 ↓ (-6.5)	**99.3** ↑ (0.1)	30.9 ↓ (-2.3)	77.6 ↑ (2.3)	78.3 ↓ (-1.1)	**94.7** ↓ (-0.5)
	Per class IoUs						**mIoU**	
	Others	Ground	Building	Water	High tension	Civil		
Baseline	86.5 ↓ (-0.1)	**95.1** ↑ (0.3)	69.1 ↓ (-4.3)	**98.4** ↑ (0.3)	**38.3** ↑ (10.9)	**2.5** ↓ (-0.1)	**65** ↑ (1.2)	
+Online	**87.0** ↑ (1.6)	**95.1** ↑ (0.2)	**69.6** ↓ (-5.9)	97.9 ↓ (-0.5)	28.8 ↓ (-1.7)	2.4 ↓ (-3.3)	63.4 ↓ (-1.6)	

* All values represent the average of three experiments, ensuring fair comparison


Ground
Building
Civil
Water
Others
High tension
RGB
Ground truth
Baseline
+Online

Results with additional RGB features

Elevation + intensity + RGB *Additional information from aerial images*

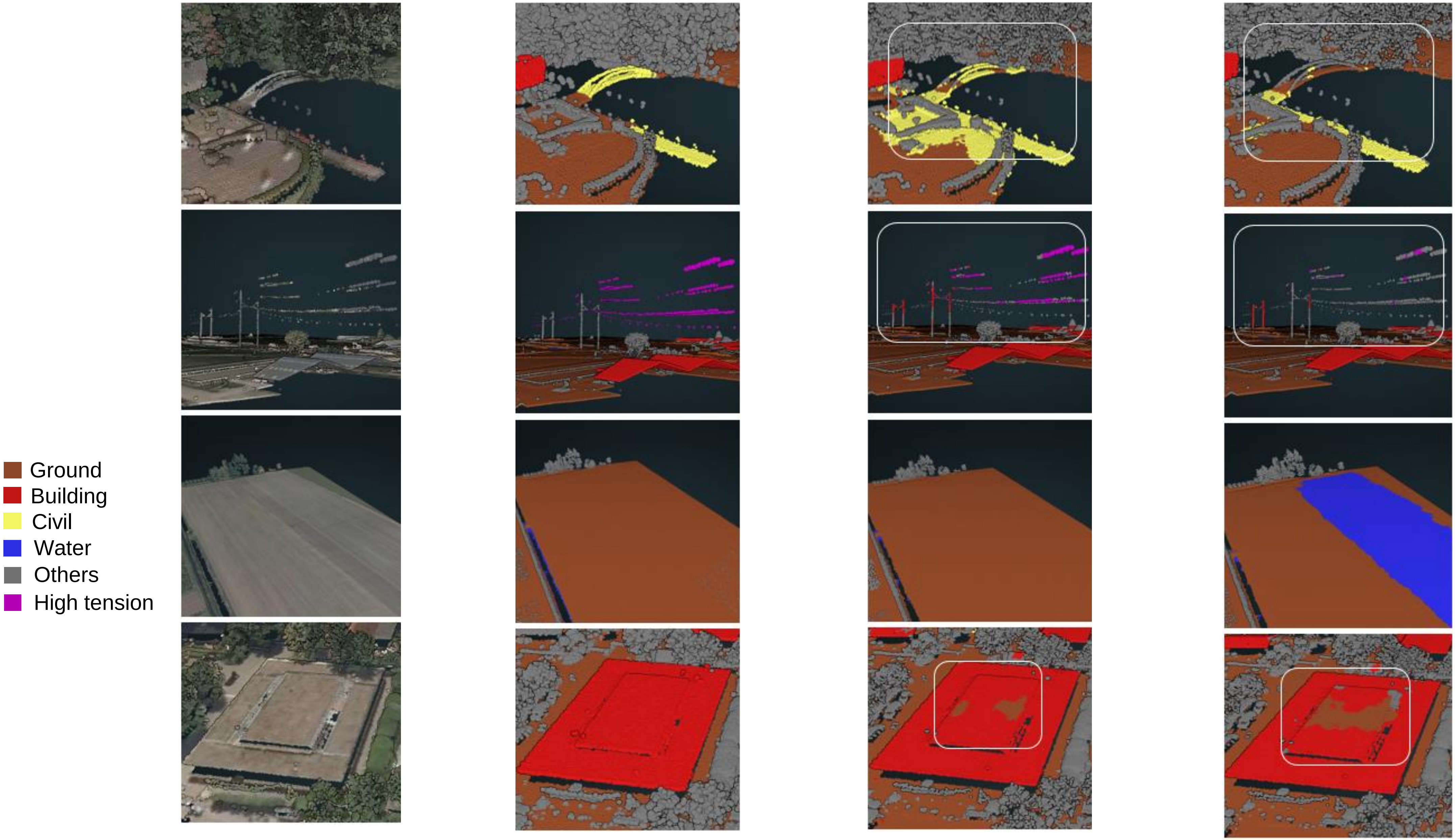
Model features: *elevation, intensity, red, green, blue*

Per class accuracies	Others	Ground	Building	Water	High tension	Civil	mAcc	OA
Baseline	90.5 ↑ (0.6)	97.9 ↓ (-0.8)	**80.3** ↑ (2.2)	**99.3** ↑ (0.2)	**46** ↑ (13.8)	**84.2** ↑ (4.3)	**83** ↑ (3.4)	**94.8** (0)
+Online	**91.7** ↑ (1)	**98** ↓ (-0.6)	68.9 ↓ (-10.5)	97.2 ↓ (-2)	31.3 ↓ (-1.9)	70.4 ↓ (-4.8)	76.3 ↓ (-3.1)	93.9 ↓ (-1.2)

Per class IoUs	Others	Ground	Building	Water	High tension	Civil	mIoU
Baseline	**87.2** ↑ (0.6)	**94.5** ↓ (-0.3)	**75.8** ↑ (2.3)	**96.5** ↓ (-1.6)	**44.5** ↑ (17.1)	**2.7** ↑ (0.2)	**66.9** ↑ (3)
+Online	85.3 ↓ (-0.1)	94.2 ↓ (-0.7)	66 ↓ (-9.4)	95.4 ↓ (-2.9)	25.6 ↓ (-4.9)	2.4 ↓ (-3.3)	61.5 ↓ (-3.5)

* All values represent the average of three experiments, ensuring fair comparison

👍 *Baseline* *+Online*

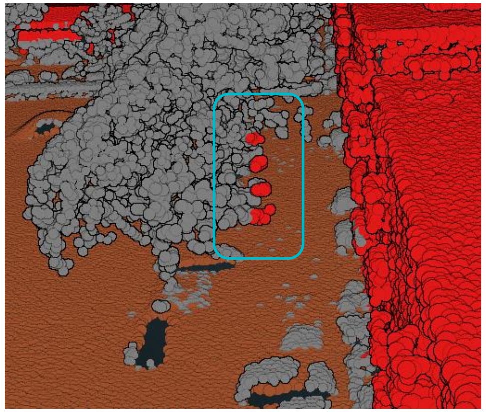


RGB *Ground truth* *Baseline* *+Online*

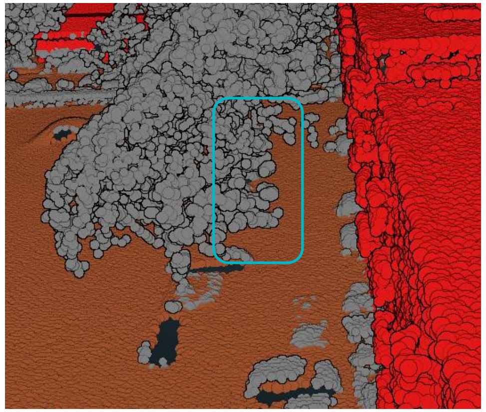
TU Delft

Online updates on Training data

RGB

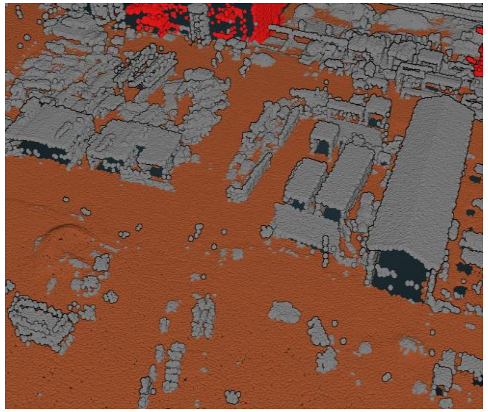


Ground truth

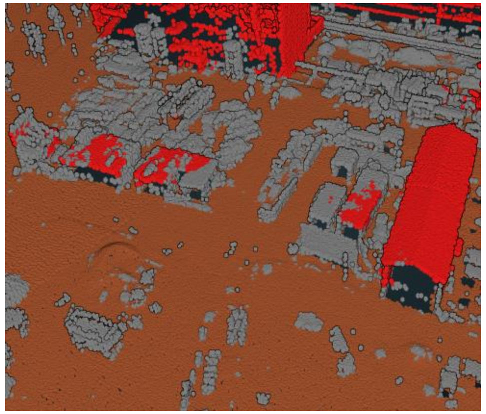


+Online update

RGB



Ground truth

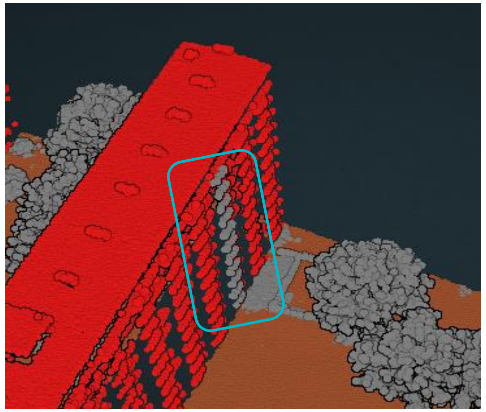


+Online update

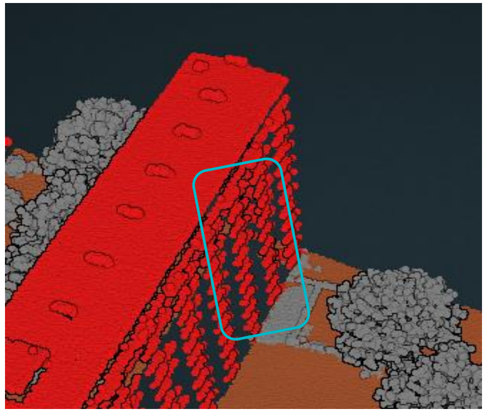
- Ground
- Building
- Civil
- Water
- Others
- High tension

Online updates

RGB

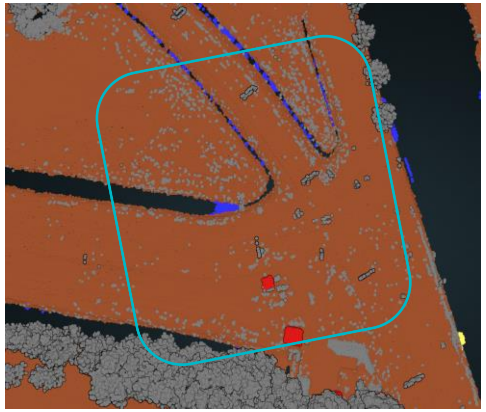


Ground truth

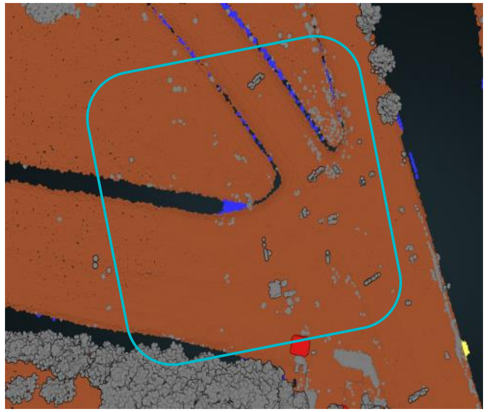


+Online update

RGB



Ground truth



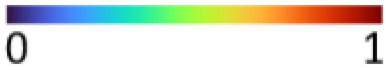
+Online update

- Ground
- Building
- Civil
- Water
- Others
- High tension

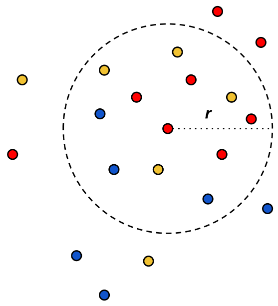
05

Discussions & Conclusions

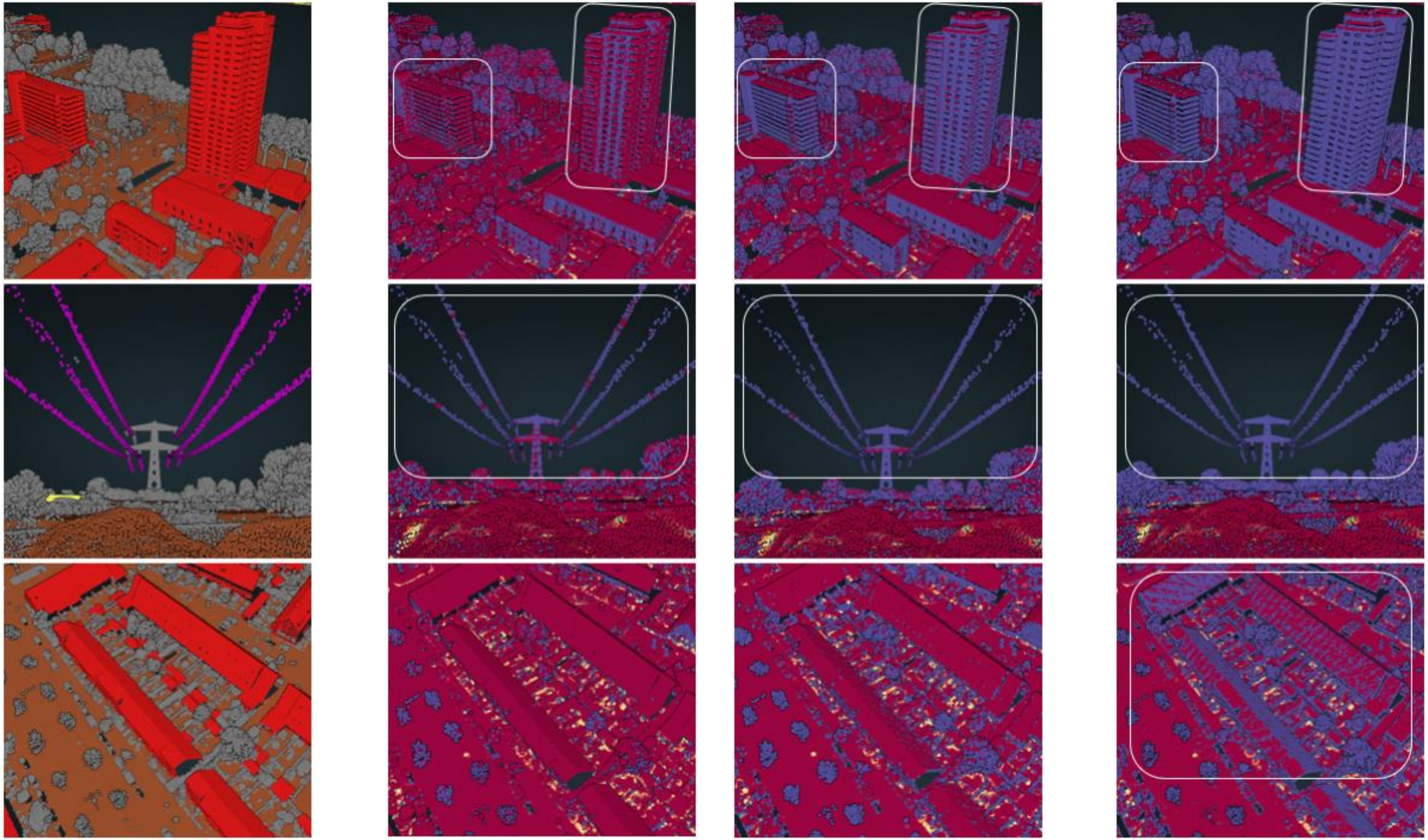
Hyperparameters comparison



r = 0.5m



- Ground
- Building
- Civil
- Water
- Others
- High tension



Ground truth *Threshold 5* *Threshold 10* *Threshold 15*

TU Delft

Our building footprints vs 2D BAG

2D BAG

Extracted Footprints

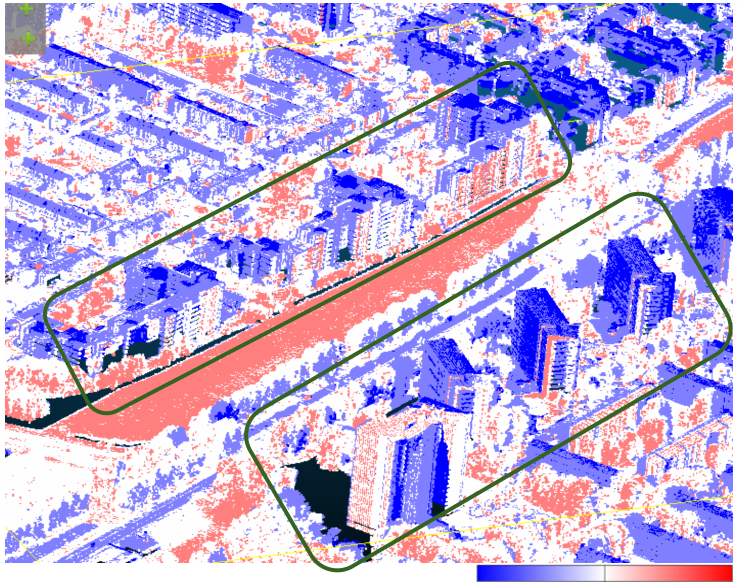
Vegetation (ndvi>0.3)

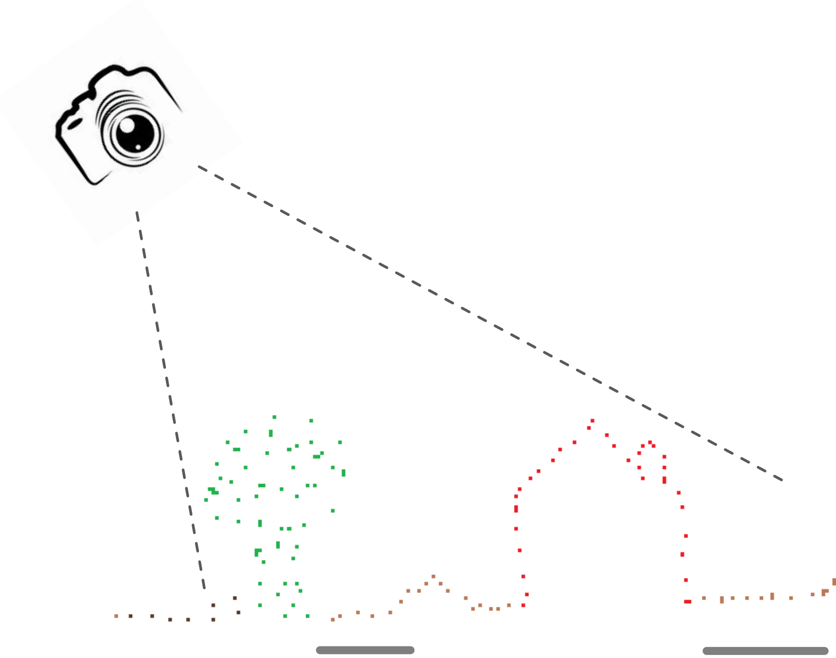
Budildings (ndvi<0.3)

TU Delft

Limitation 1 – Quality of additional features NIR

Buildings – decreased performance!





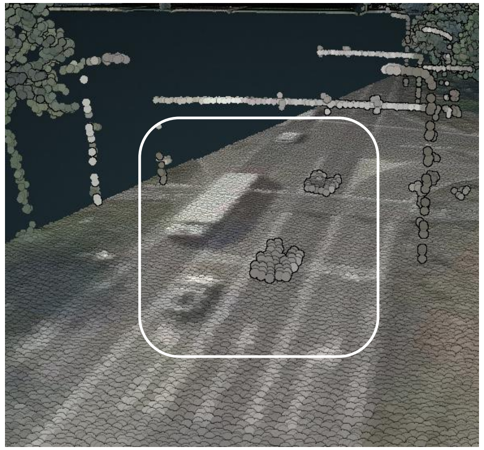
Alignment artifacts – point cloud & aerial images fusion

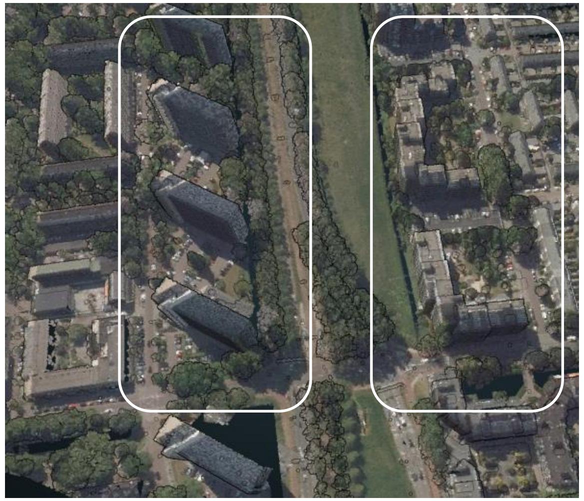
TU Delft

Limitations 1 – Quality of additional features RGB

Ground

decreased performance!

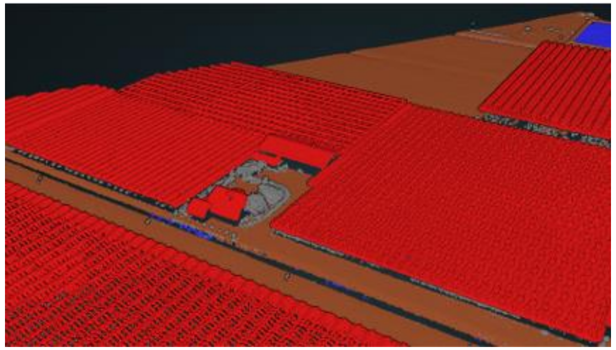


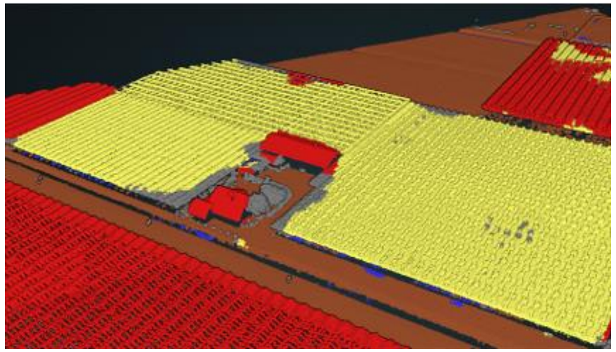
Alignment & Temporal artifacts

Limitation 2 – Missing context

RGB

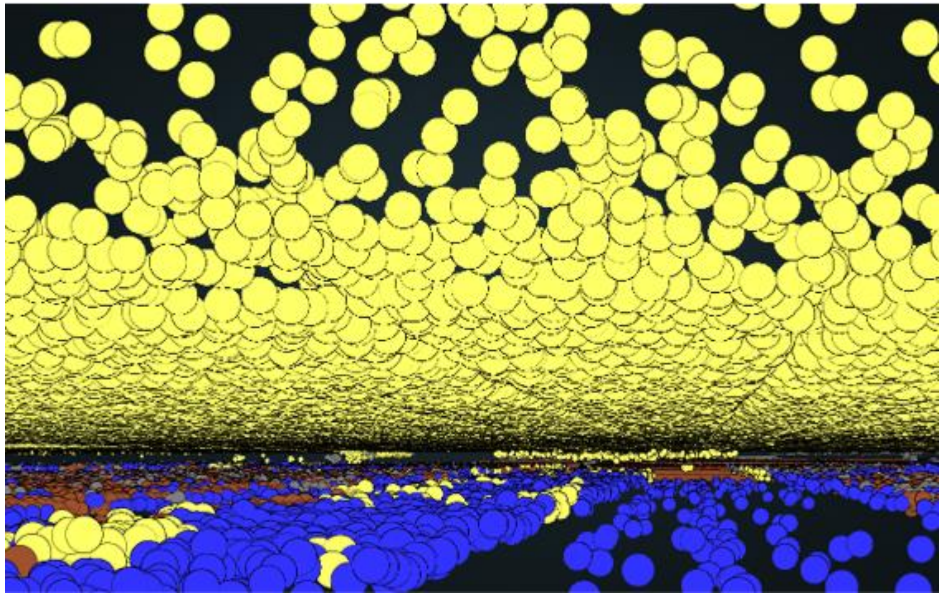


Ground truth



+Online

- Ground
- Building
- Civil
- Water
- Others
- High tension



+Online

High accuracy but very low IoU

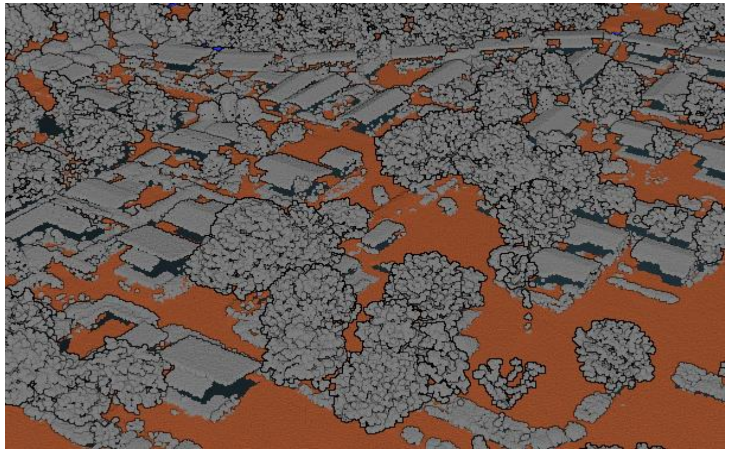
80% accuracy *5% IoU*

- *No training data!*
- *Context is everything*

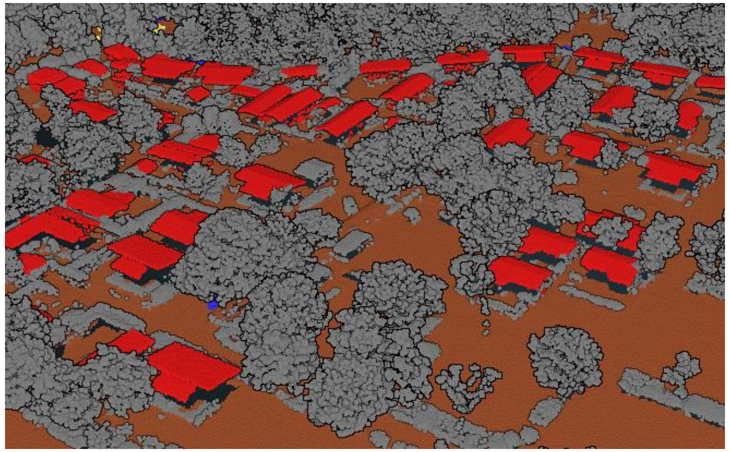
Limitations 3 – Faulty ground truth

Ground Building Civil Water Others High tension

RGB



Ground truth



+Online

Despite the model’s strong predictive ability, these metrics suggest otherwise

Conclusions

Overall

- *If only raw LiDAR*

  +ONLINE – best!

OA	***95.1%***
mAcc	***79.4%***
mIoU	***65***

- *Training data gets better*

  quantify!?

However!

- *for class specific tasks, and*
- *if additional information (NIR, RGB)*

  Baseline with NO Online is better

High tension	***Baseline with RGB***
Others (veg)	***Baseline with NIR***

- +Online is very sensitive to data artefacts

Future scope

1. *Online hypothesis with Transformers*

2. *Generalizability to TLS & MLS*

3. *Incorporation of synthetic data*

Thanks!

Questions?