Semantic Segmentation of Roof Superstructures

MSc thesis presentation | June 28th, 2022 | Irène Apra

TU Delft

TUM

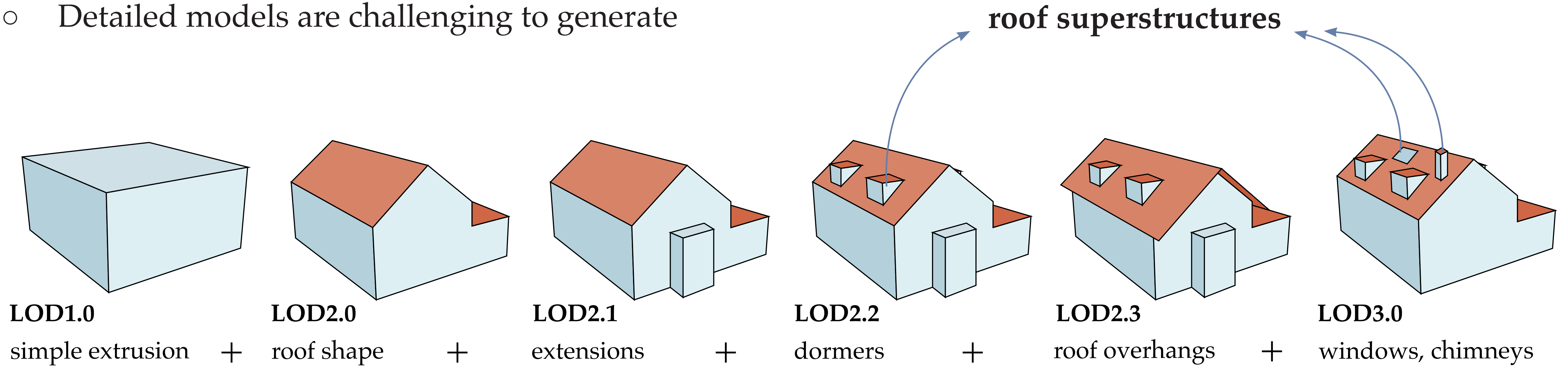
Content

Introduction

Motivation

Semantic city models:

- Different levels of details (LODs)
- Different applications
- Detailed models are challenging to generate



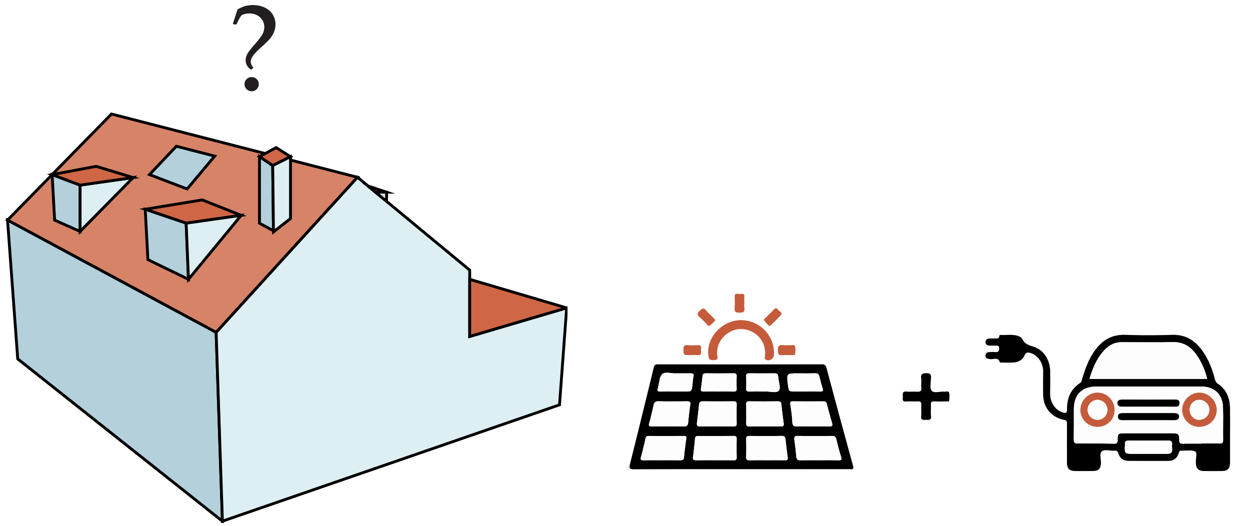
source: Biljecki et al. 2014

Introduction

Motivation

PV potential of buildings:

- ○ Project at the TU Munich
- ○ Need to assess the building's geometrical potential
- ⇨ Estimation of roof surface available for panels

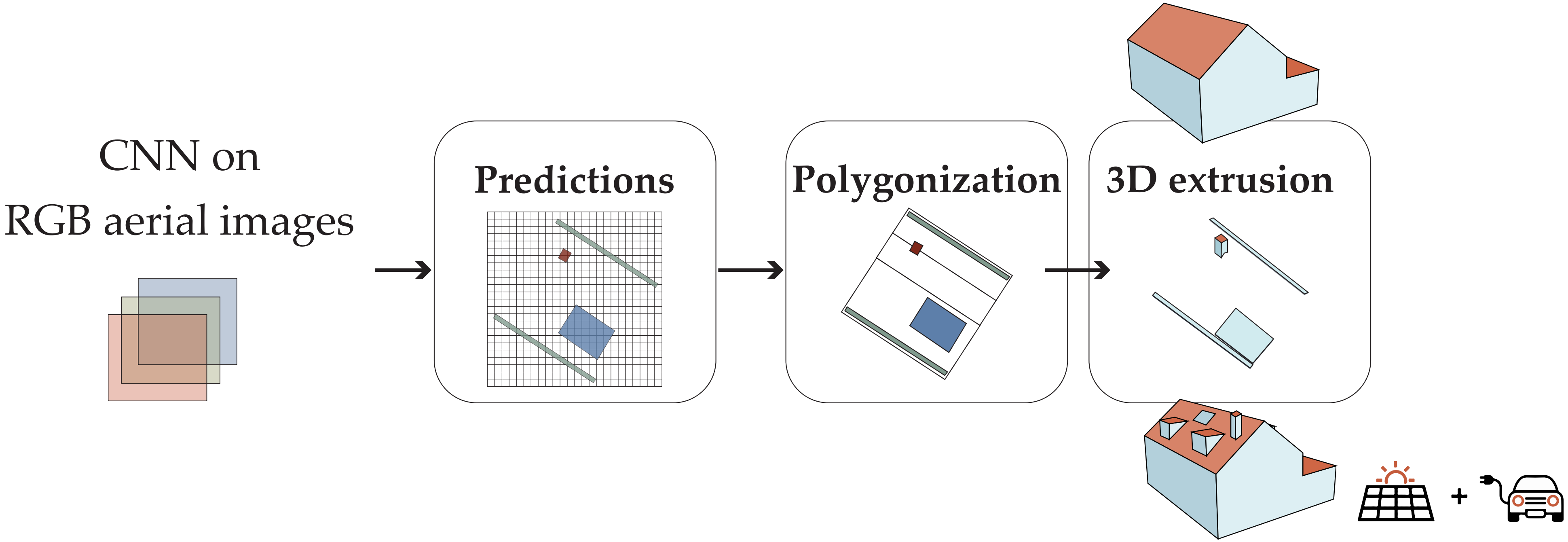


Introduction

Motivation

Approach of the existing project:

1. Detect superstructures through Convolutional Neural Network (CNN)
2. Vectorize and model them in 3D
3. Add them to a simple 3D model available

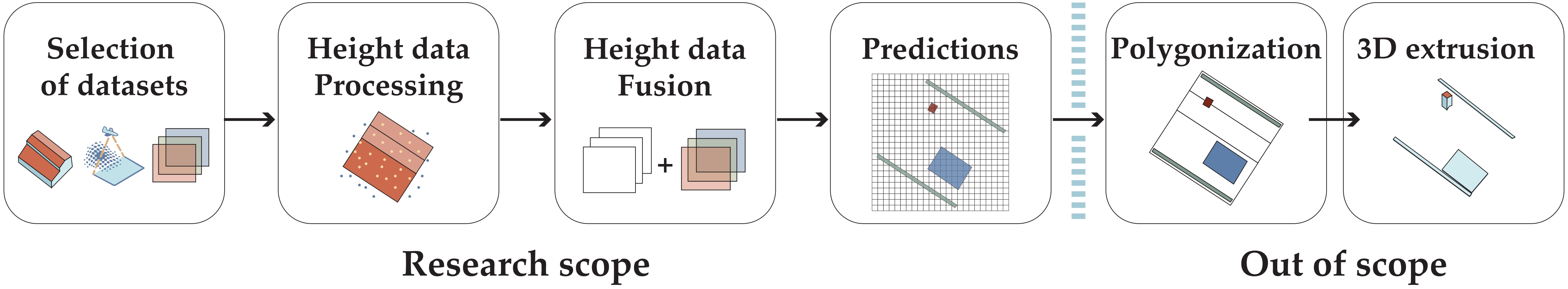


Introduction

Research scope

Research goal and steps:

- Improve superstructure detection by incorporating 3D data sources
- Selection of appropriate 3D data (LiDAR)
- LiDAR preprocessing to obtain height maps (absolute and normalized)
- Implementation of fusion network



Introduction

Problem & Hypotheses

Main question:

How can building height data fused to a Convolutional Neural Network (CNN) on RGB aerial images improve the semantic segmentation of roof superstructures?

Hypotheses:

1. Improvement for some classes only: the ones with relief
2. Relative height yield better results than the absolute one since it would highlight only reliefs
3. Interpolated height data yield better results than no interpolation

1. Related work

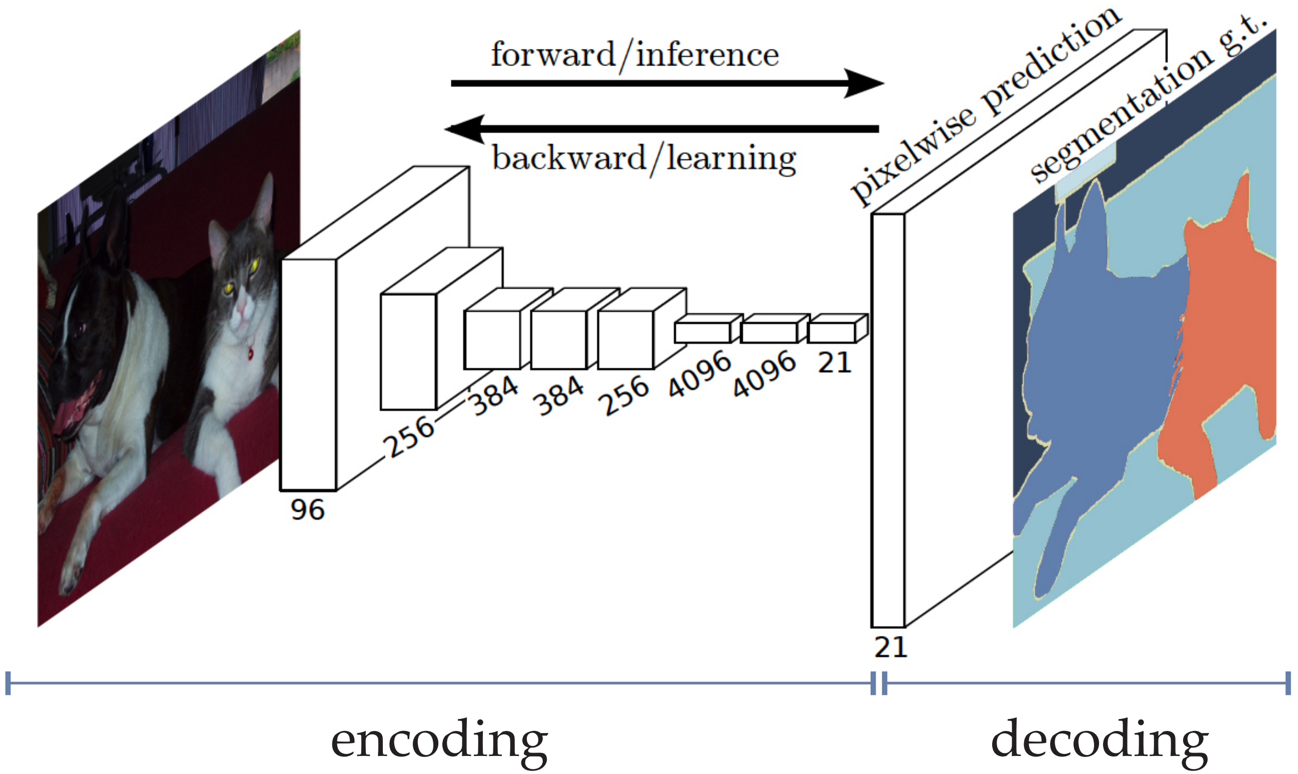
Semantic segmentation | Definition

Semantic segmentation:

- Assigns a class per pixel
  (classes = superstructure categories)

Network:

- CNN on image input
- Encoding part to obtain feature maps
- Decoding back to obtain back the image size



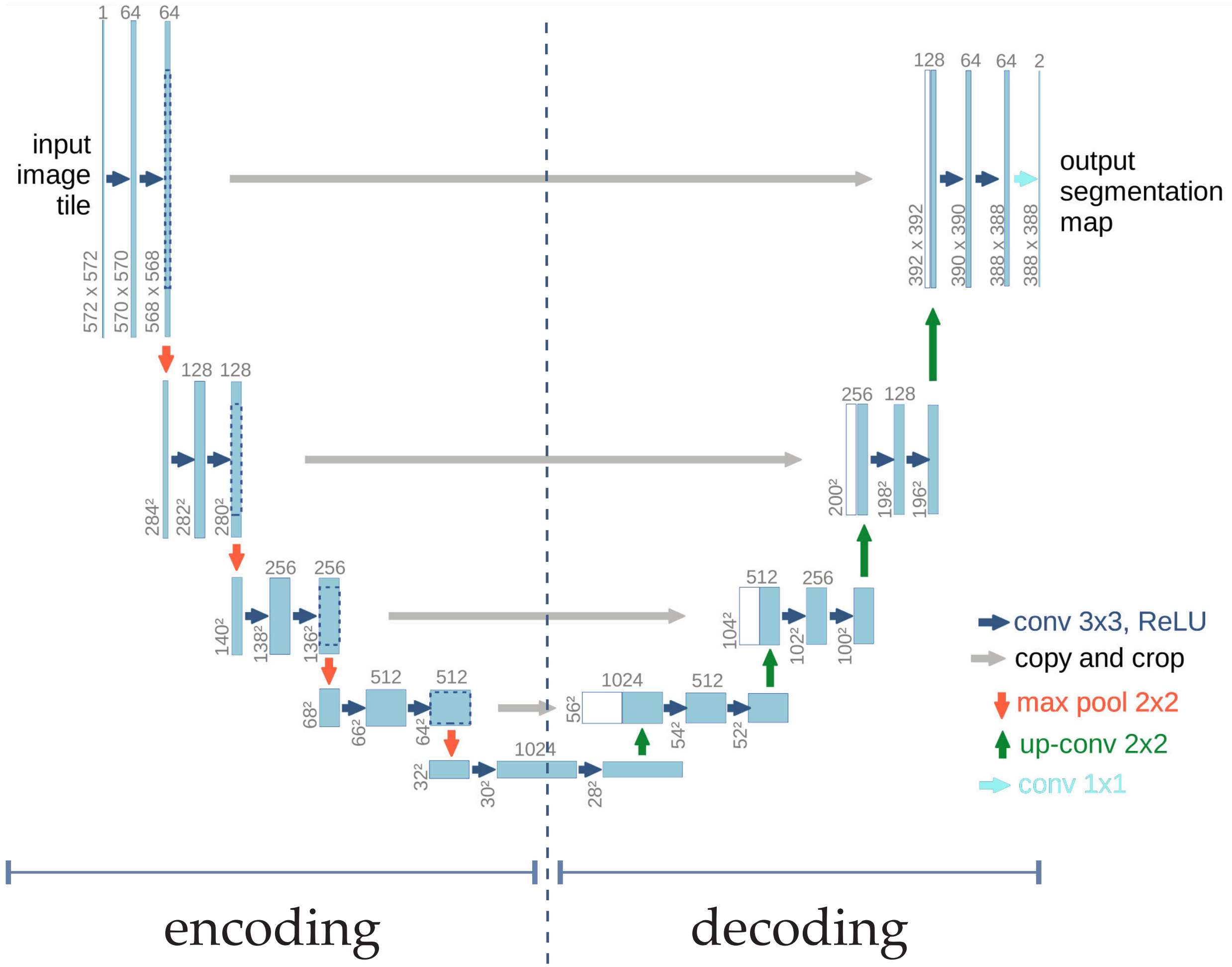
FCN; source: Long et al. 2015

1. Related work

Semantic segmentation | U-Net

U-Net:

- One of the most used CNN architectures
- Symmetric architecture
- Used in the PV-assessment pipeline



source: Ronneberger et al. 2015

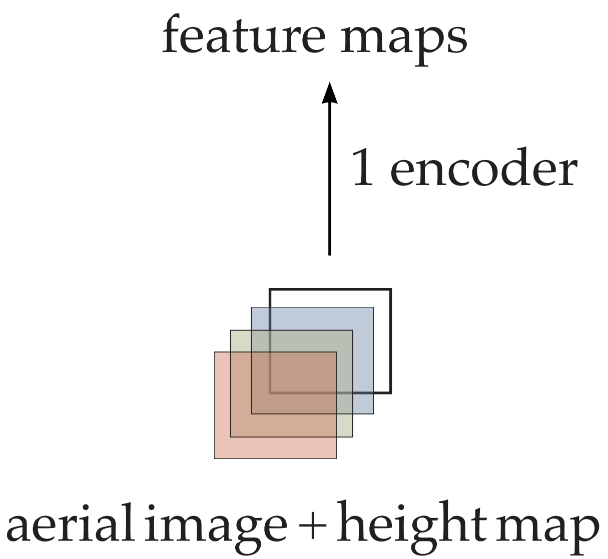
1. Related work

Multimodal networks | Concepts

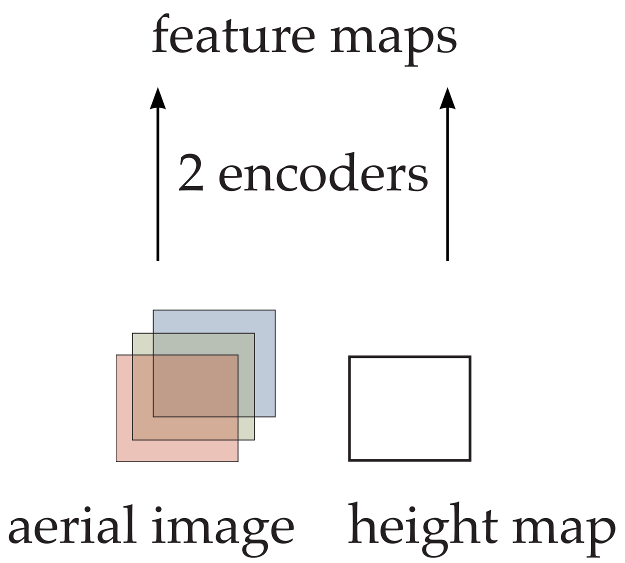
Height data fusion:

1. Stacking on an additional channel
2. Fusing by encoding on another network branch

- According to the literature, fusion yields better results (e.g. land-use segmentation)



STACKING



FUSION

1. Related work

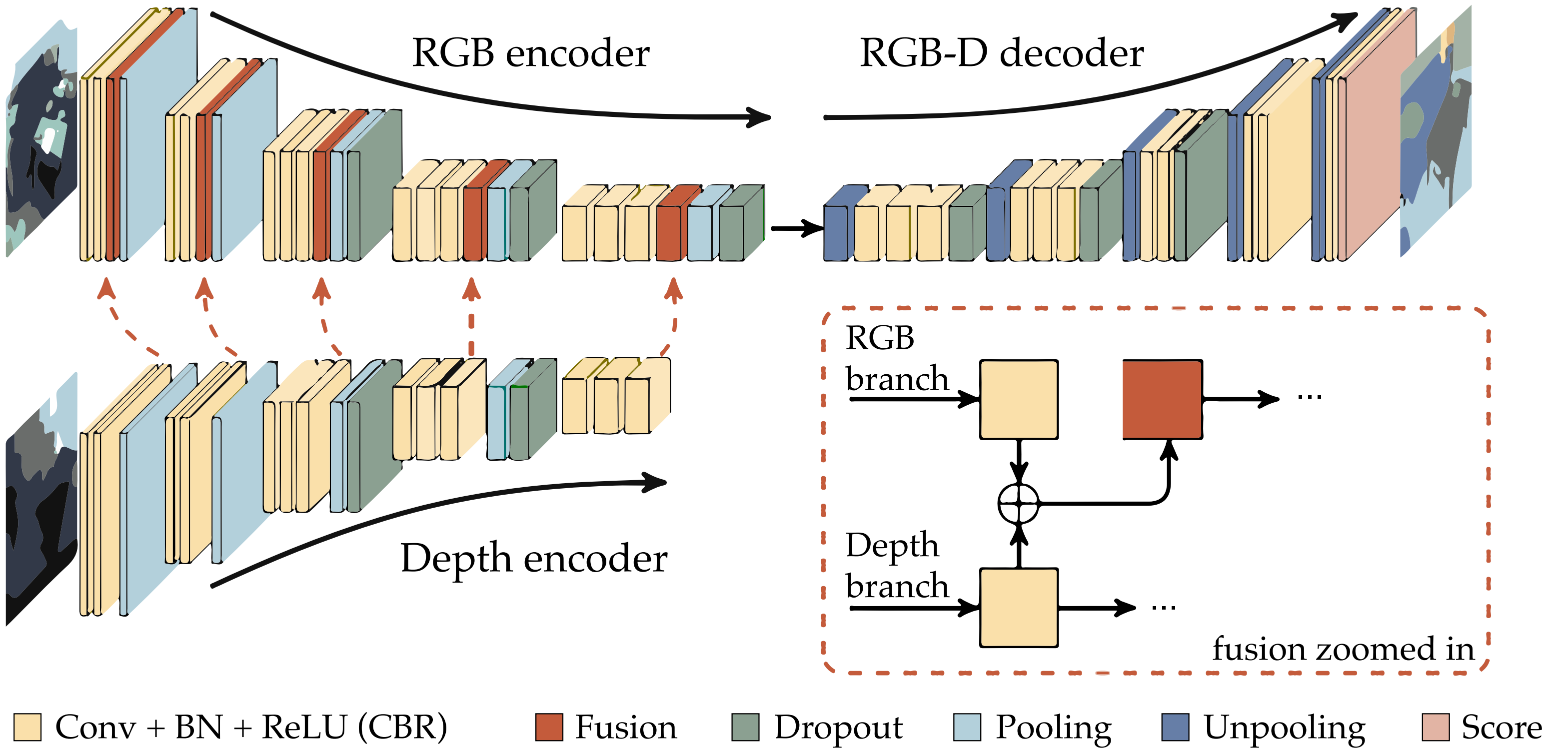
Multimodal networks | FuseNet

FuseNet functioning:

- Fusion network
- Fusion module: activations of the auxiliary branch are fused to the main branch during encoding

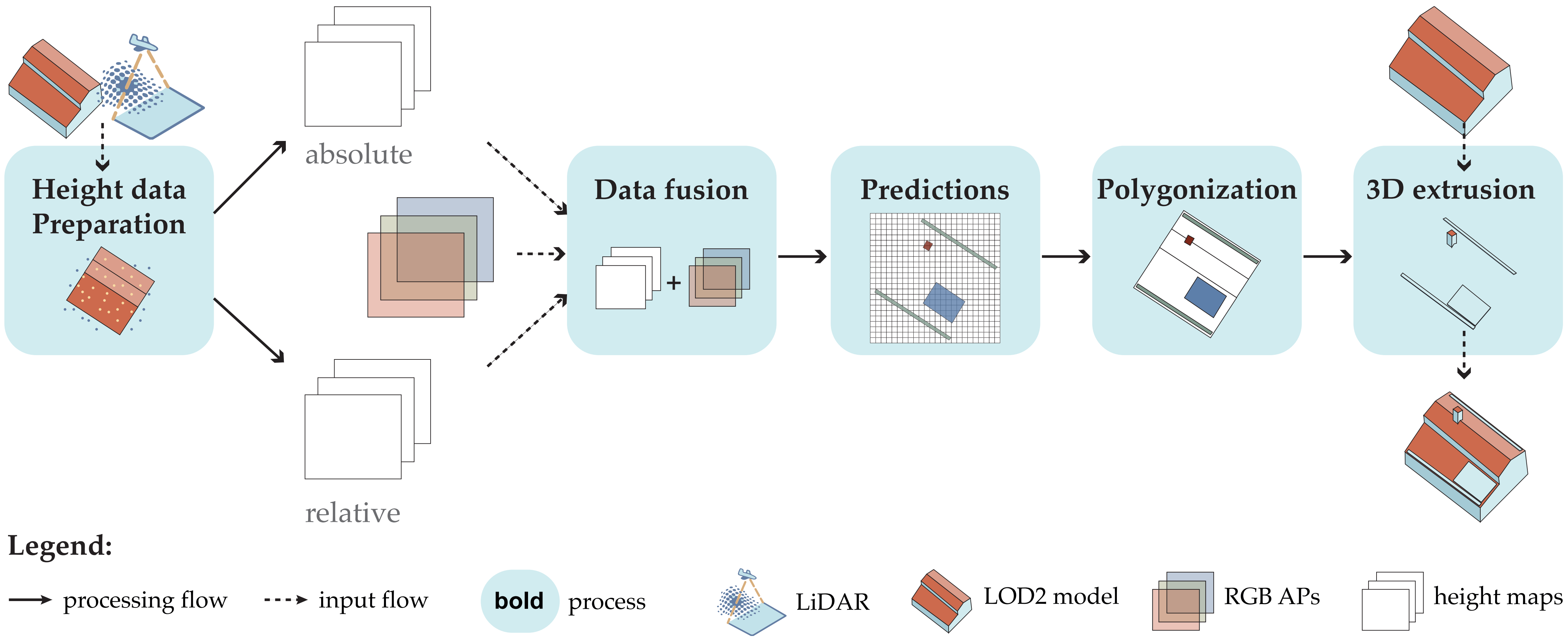
Experiments:

- Comparison of U-Net and FuseNet results



source: adapted from Hazirbas and Aygun 2018

2. Methodology
Overview

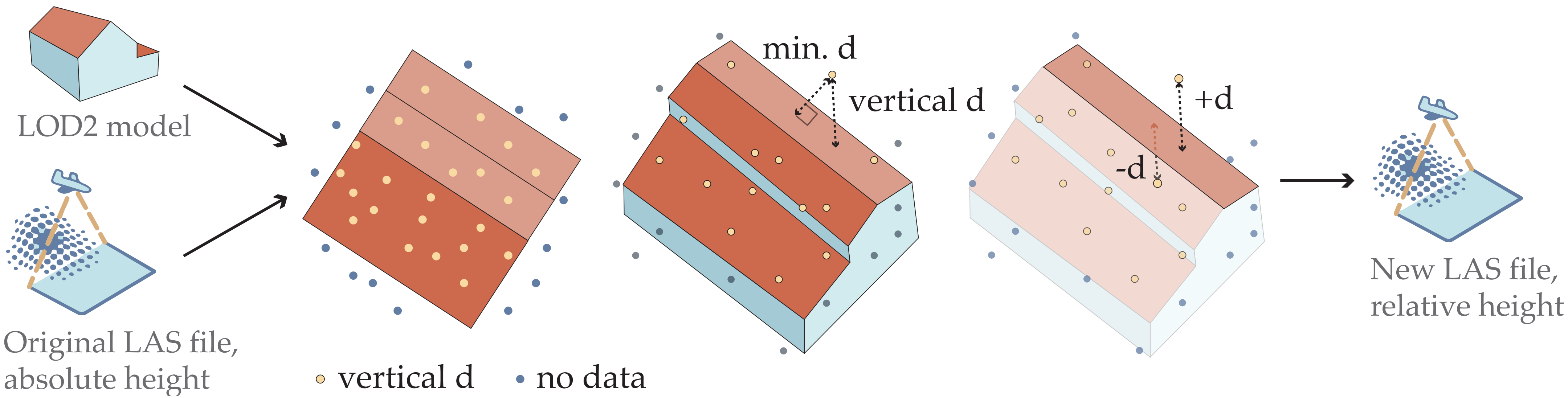


2. Height data preparation

Relative height calculation | Concept

Relative height calculation:

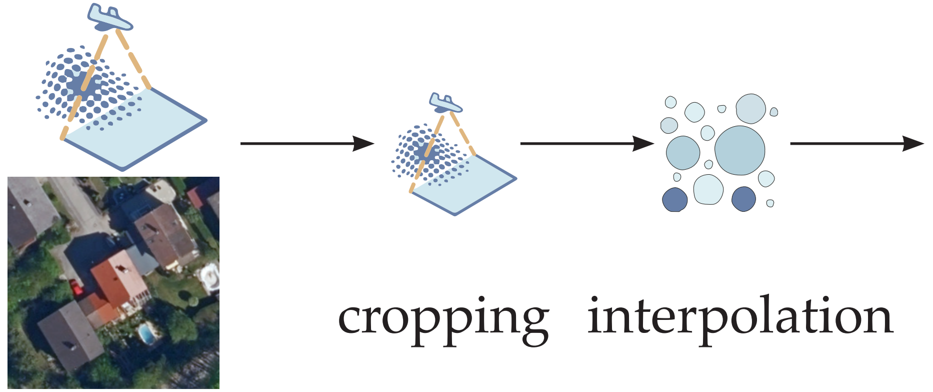
- Extraction of roof polygons
- Retrieval of underlying polygon per point
- Vertical distance calculation

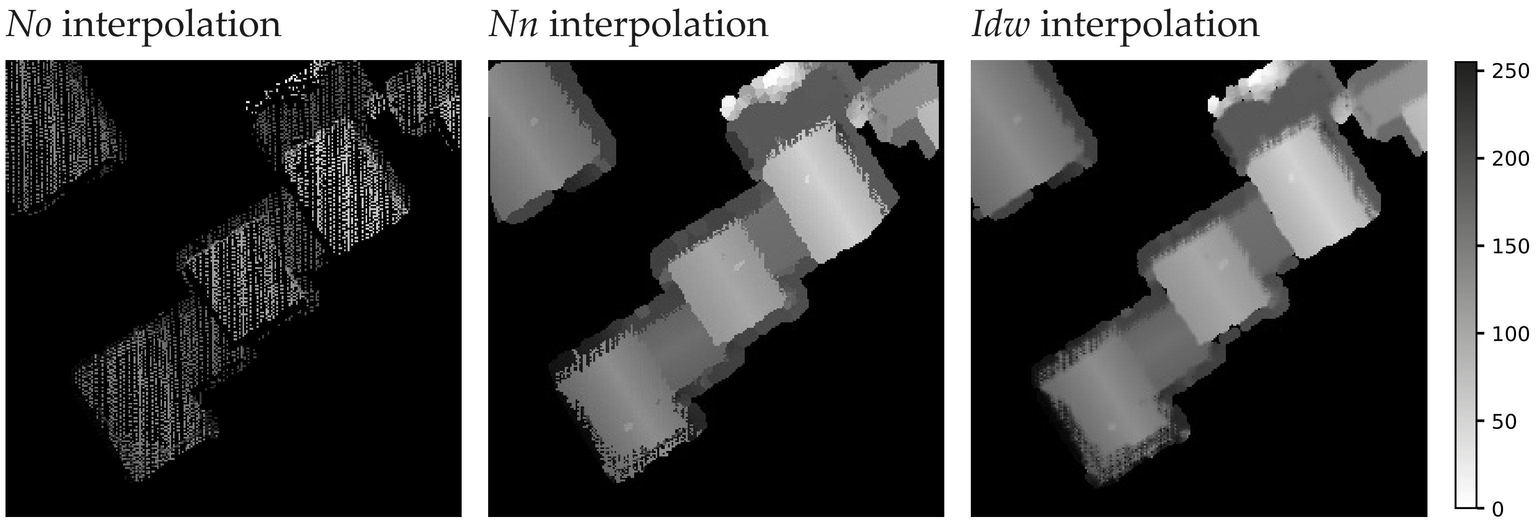


2. Height data preparation

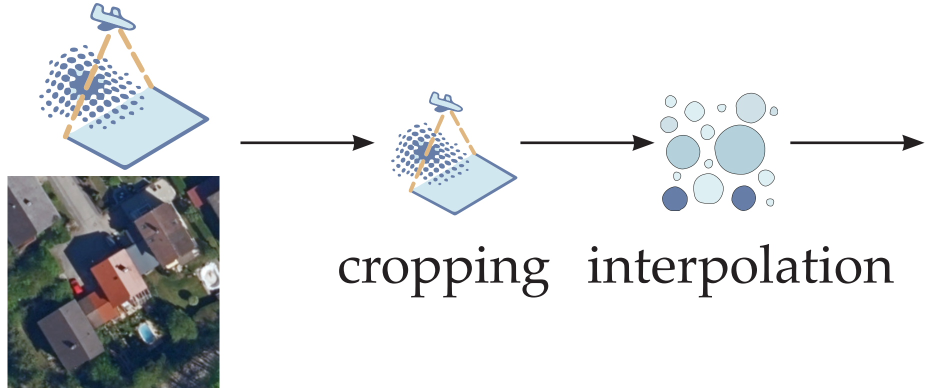
Interpolation | Implementation & Results

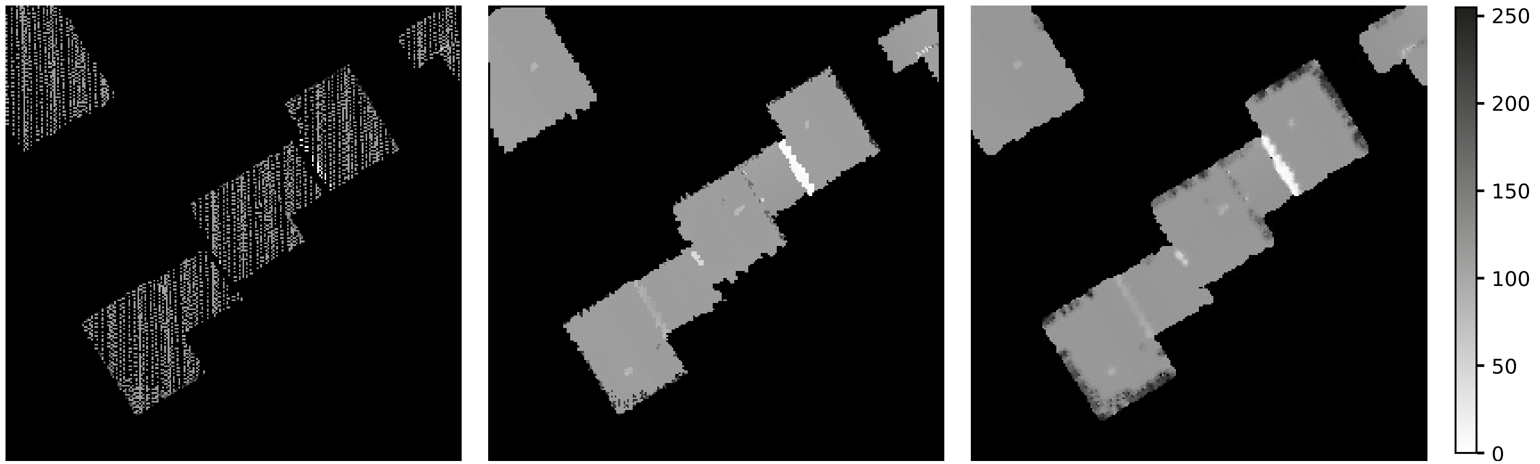
Absolute height





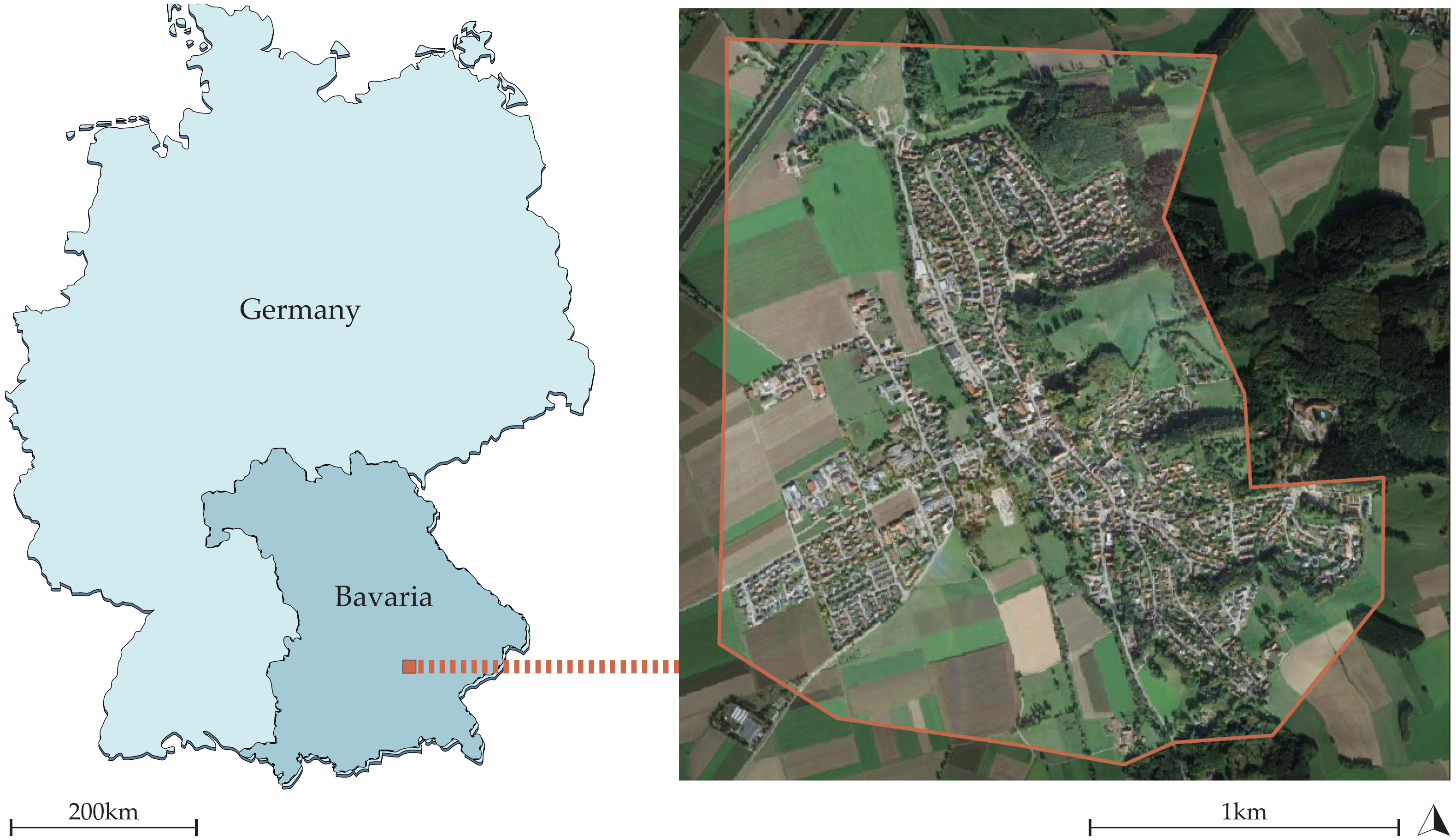
Relative height





3. Experiments
Image dataset | Location, Germany

Image dataset:
- Bavarian village, Wartenberg
- Exisiting implementation uses non-ortho-rectified images
- Roof centered images

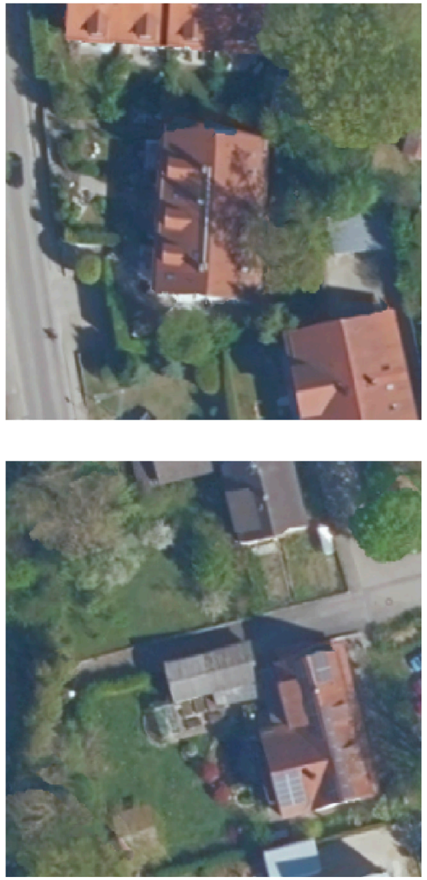


3. Experiments

New training data | Ortho-labels

1 880 roofs (true ortho-photos)

256 × 256 pixels



6 classes

- PV module
- dormer
- window
- ladder
- chimney
- unknown

3. Experiments

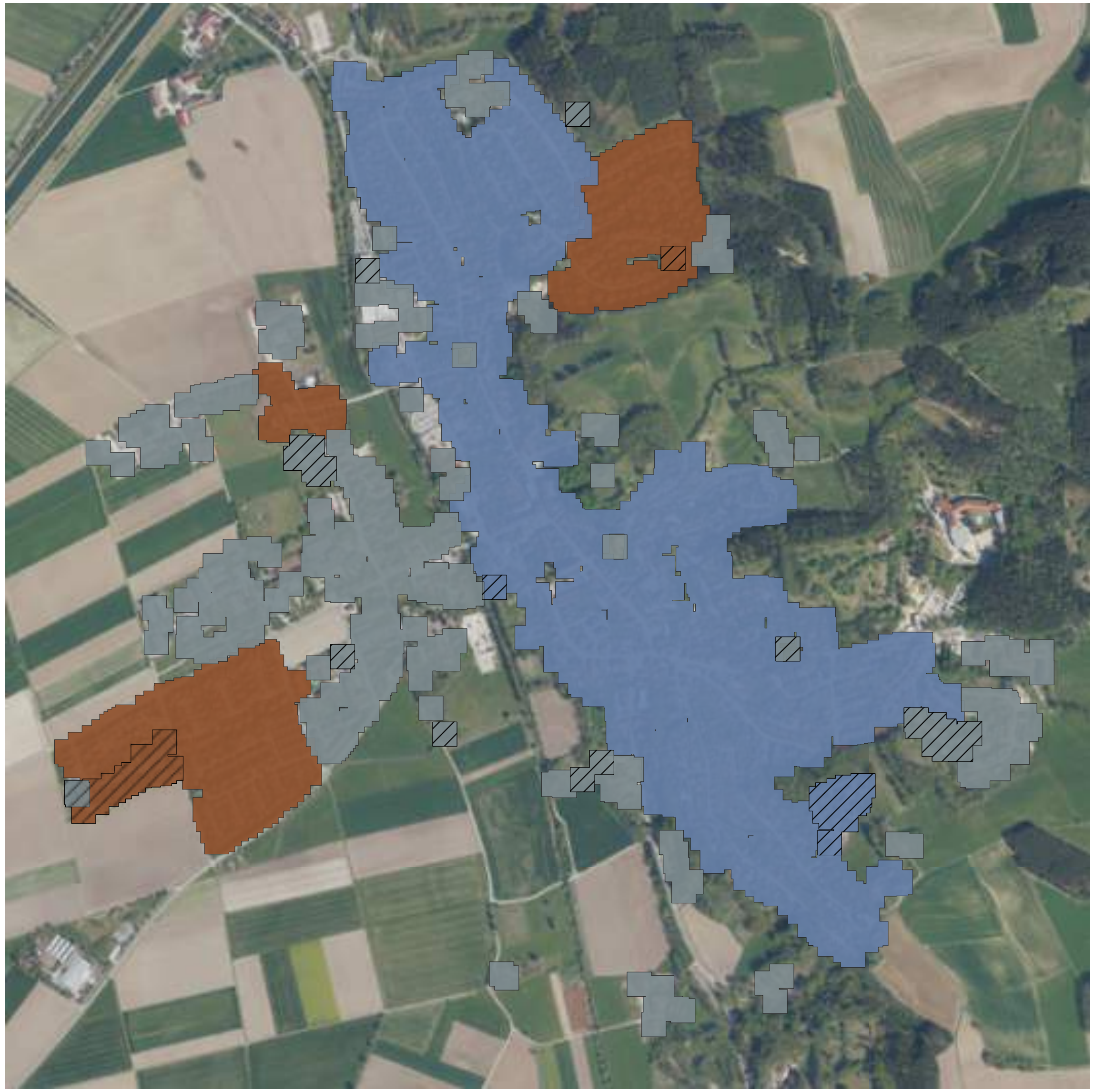
New training data | Data split

Split:

- Train, validation and test datasets
- Independent datasets
- Test dataset is labeled to assess the network performance quantitatively

training dataset validation dataset test dataset

missing height data DOP 20cm/pixel 100m



3. Experiments

Semantic segmentation results | Assessment concept

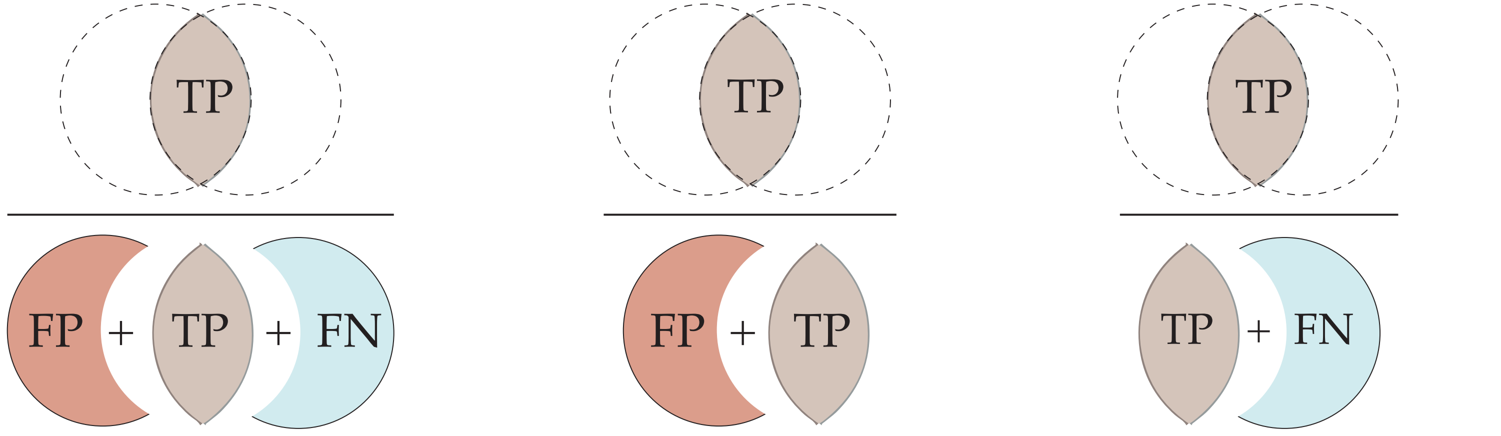
Metrics on test dataset:

1. Intersection over Union (IoU),
   a global score including background
2. Accuracies per class

○ Calculated through confusion matrices

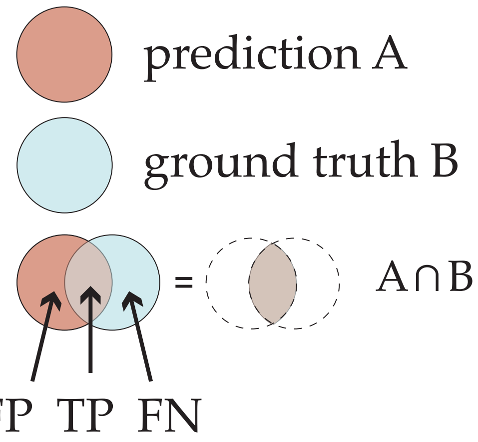
Ground truth (classes) × Prediction (classes)

3E+05	471	8286	7806	75	178	70311	0.79
1863	42791	1175	85	215	134	51566	0.44
4127	557	36633	373	494	604	37315	0.46
9	61	978	4240	221	1064	7908	0.29
32	63	614	267	15531	707	4900	0.7
37	308	2427	1660	1160	5601	22926	0.16
15017	17677	21047	13712	6394	16990	2E+07	1
0.94	0.69	0.51	0.15	0.64	0.22	0.99	0.99



IoU **Prediction accuracy** **Accuracy on GT**

Legend:

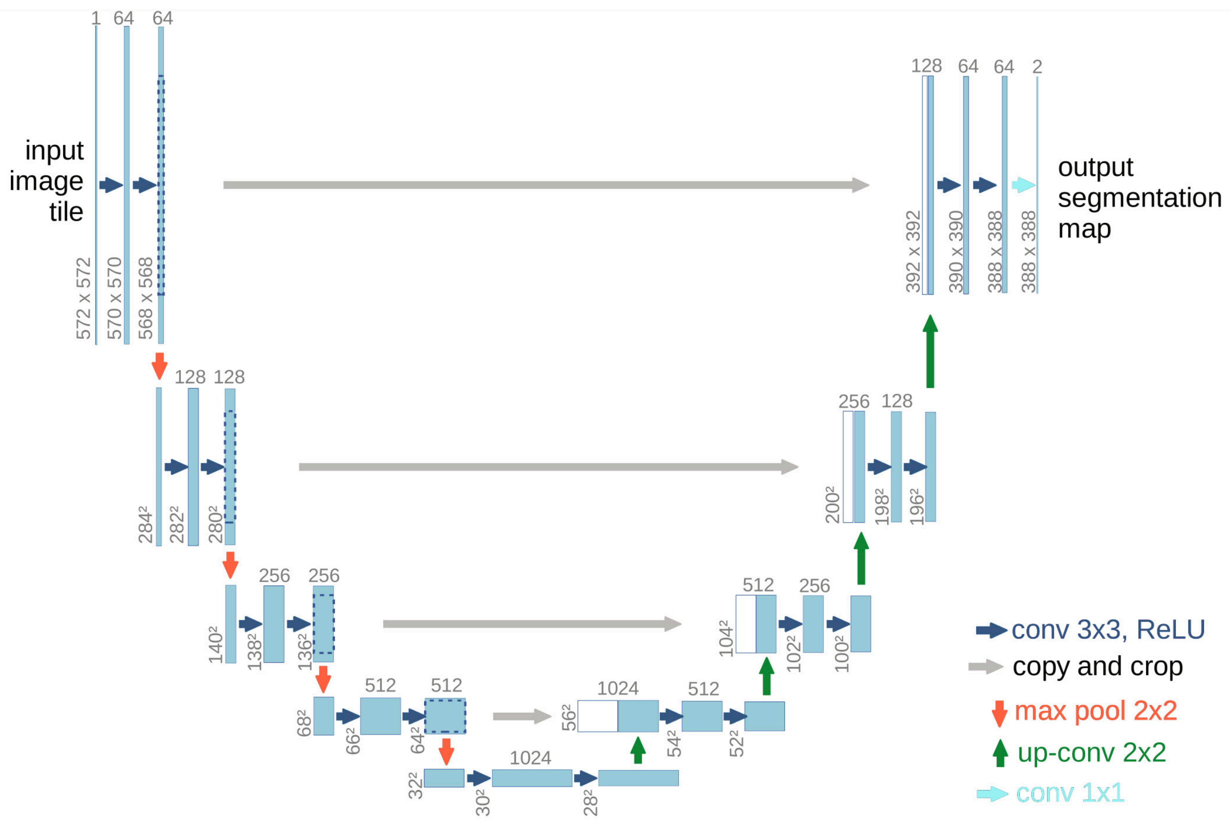


3. Experiments

U-Net | Scope

Scope of U-Net experiments:

- ○ U-Net is only based on aerial images
- ○ Provide a comparison mean for Fusion results
- ○ Have to be run on new labels



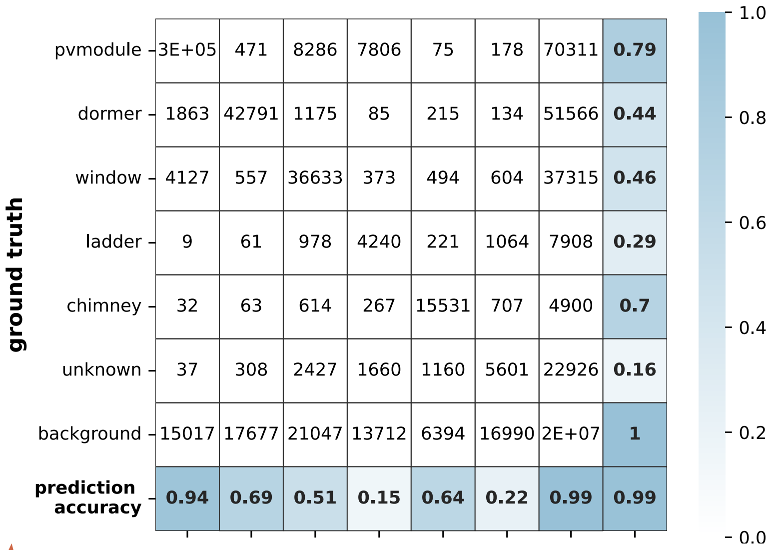
3. Experiments

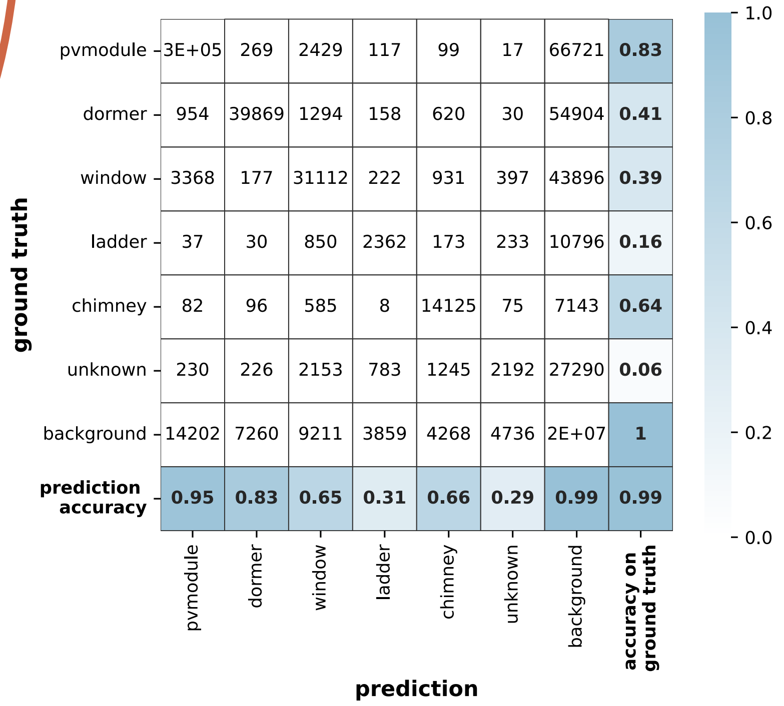
U-Net | Results

Results of U-Net experiments:

- Several sets of parameters are tested
- Two best confusion matrices are kept as reference

id	split	epoch	loss	base	augmentation	mean Acc. GT	mean Pr.Acc.	IoU
1	01	40	CFL+Dice	vgg19	None	**0.47**	**0.53**	**0.45**
2	01	40	CFL+Jaccard	vgg19	None	**0.43**	**0.58**	**0.45**
3	01	40	0.5(CFL+Dice) +0.5(CFL+Jacc.)	vgg19	None	**0.43**	**0.61**	**0.44**
4	01	40	CFL+Jaccard	resnet152	None	**0.40**	**0.53**	**0.43**
5	02	40	CFL+Jaccard	vgg19	None	**0.44**	**0.63**	**0.46**
6	02	40	CFL+Jaccard	resnet152	None	**0.42**	**0.62**	**0.45**
7	01	80	CFL+Jaccard	vgg19	None	**0.42**	**0.62**	**0.45**
8	02	80	CFL+Jaccard	vgg19	None	**0.46**	**0.64**	**0.47**
9	01	40	0.5(CFL+Dice) +0.5(CFL+Jacc.)	resnet152	Train Val.	**0.48**	**0.58**	**0.47**
10	01	40	0.5(CFL+Dice) +0.5(CFL+Jacc.)	resnet152	Train Val. Test	**0.46**	**0.58**	**0.46**



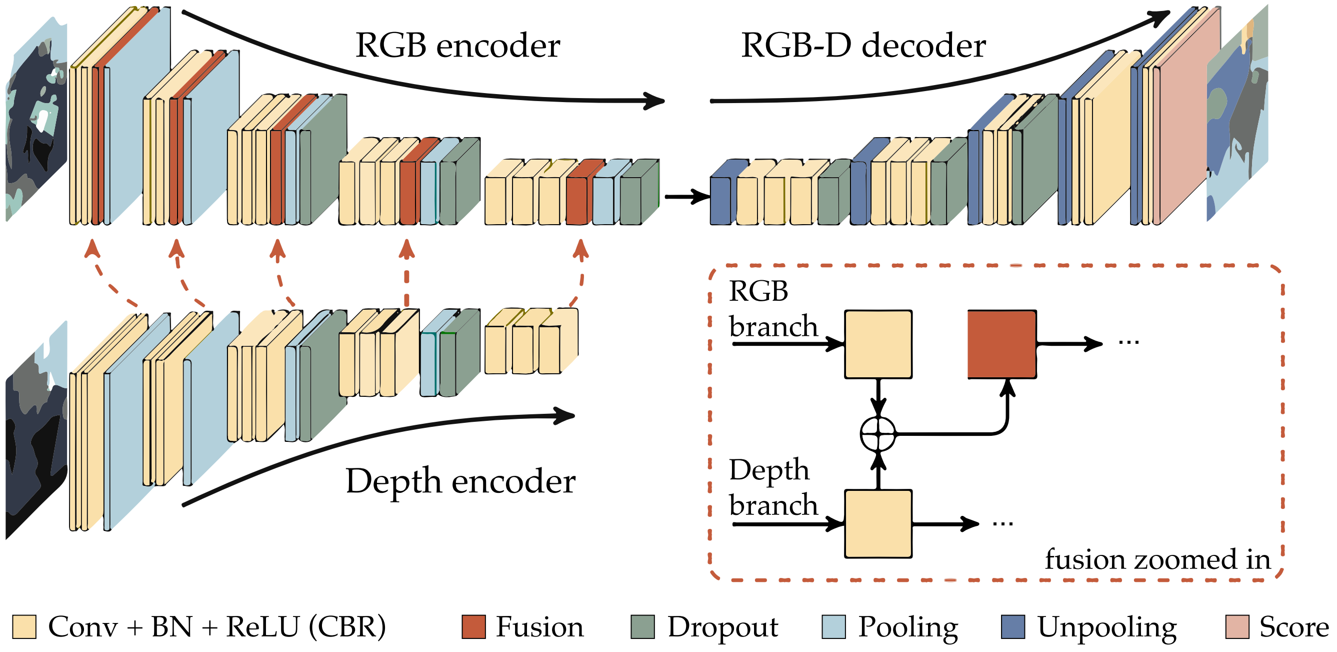


3. Experiments

FuseNet | Scope

Scope of FuseNet experiments:

- Determine best parameters
- Fuse all different height datasets (x6)
- Assess which dataset yields best results



4. Results

Results on all datasets

U-Net best results

id	mean GT	Acc.	mean Pr.Acc.	IoU
1	**0.47**		**0.53**	**0.45**
7	**0.42**		**0.62**	**0.45**

FuseNet results, absolute height

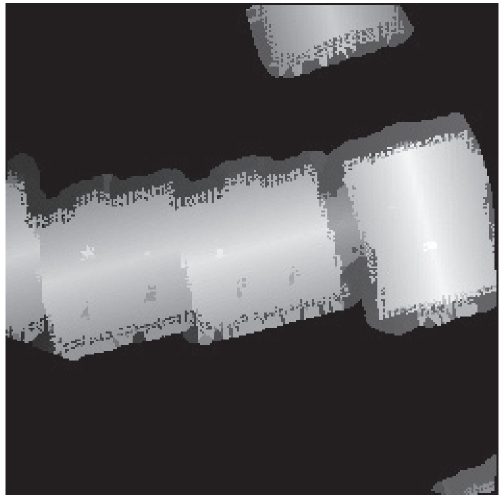
id	height data	mean GT	Acc.	mean Pr.Acc.	IoU
i	*no*	**0.40**		**0.51**	**0.40**
ii	*nn*	**0.51**		**0.56**	**0.46**
iii	*idw*	**0.49**		**0.56**	**0.46**

FuseNet results, relative height

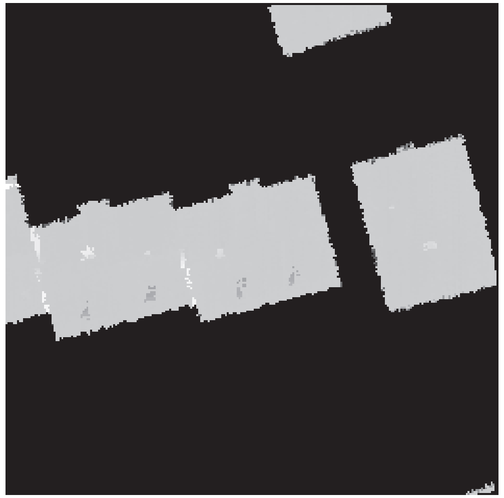
id	height data	mean GT	Acc.	mean Pr.Acc.	IoU
iv	*no*	**0.27**		**0.54**	**0.34**
v	*nn*	**0.51**		**0.51**	**0.45**
vi	*idw*	**0.46**		**0.57**	**0.45**

Enhanced structural information:

Absolute (NN)



Relative (NN)

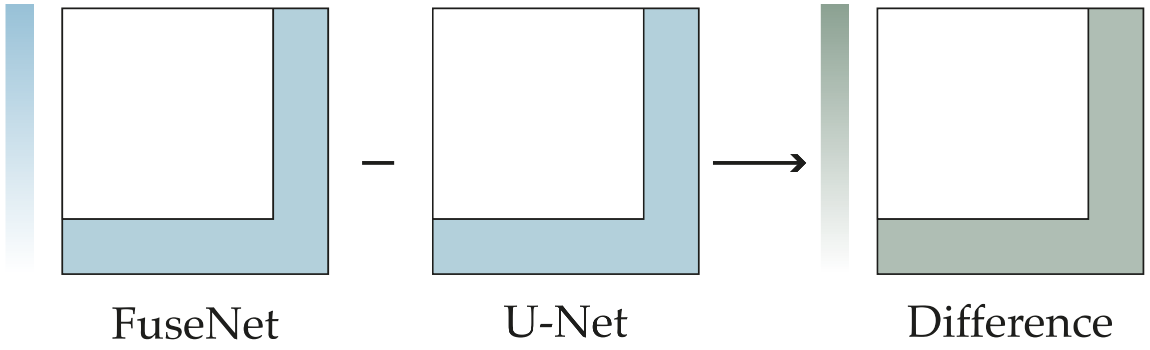


5. Analysis

Quantitative analysis | Comparison U-Net and FuseNet

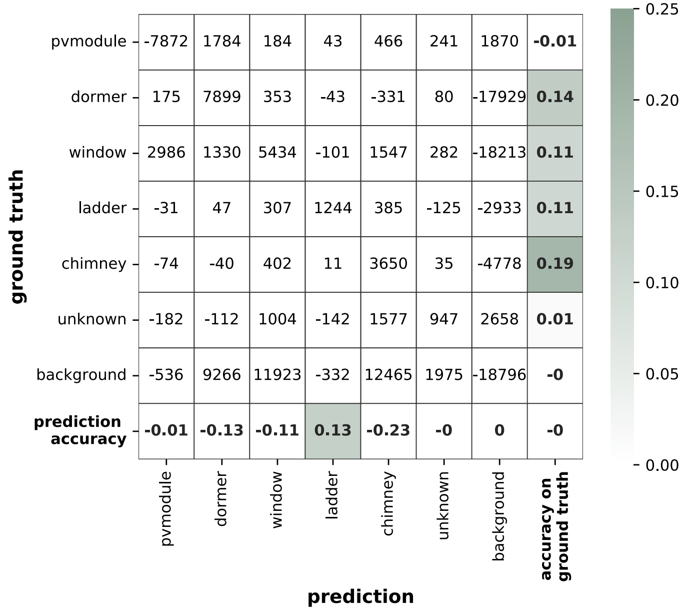
Concept:

- Subtraction of U-Net to FuseNet results
- Comparison per class



Calculation:

- Example result with green scale highlighting improved classes
- 6 FuseNet best results are compared to 2 U-Net best results
- A general trend can be inferred through a scatter plot



ground truth \ prediction	pvmodule	dormer	window	ladder	chimney	unknown	background	accuracy on ground truth
pvmodule	-7872	1784	184	43	466	241	1870	**-0.01**
dormer	175	7899	353	-43	-331	80	-17929	**0.14**
window	2986	1330	5434	-101	1547	282	-18213	**0.11**
ladder	-31	47	307	1244	385	-125	-2933	**0.11**
chimney	-74	-40	402	11	3650	35	-4778	**0.19**
unknown	-182	-112	1004	-142	1577	947	2658	**0.01**
background	-536	9266	11923	-332	12465	1975	-18796	**-0**
prediction accuracy	**-0.01**	**-0.13**	**-0.11**	**0.13**	**-0.23**	**-0**	**0**	**-0**

5. Analysis

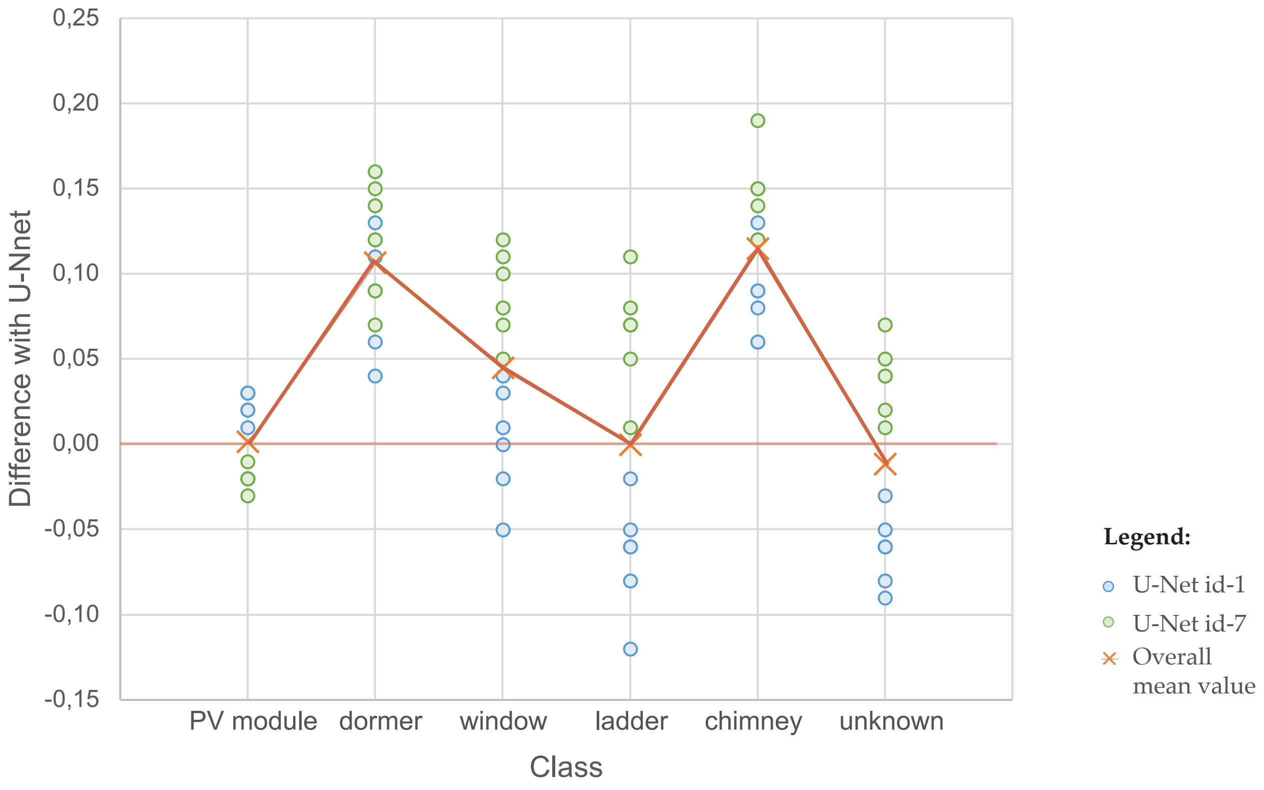
Quantitative analysis | Comparison U-Net and FuseNet

Accuracy on ground truth:

✓ Improved for dormers, chimneys and windows

ground truth \ prediction	pvmodule	dormer	window	ladder	chimney	unknown	background	accuracy on ground truth
pvmodule	-7872	1784	184	43	466	241	1870	**-0.01**
dormer	175	7899	353	-43	-331	80	-17929	**0.14**
window	2986	1330	5434	-101	1547	282	-18213	**0.11**
ladder	-31	47	307	1244	385	-125	-2933	**0.11**
chimney	-74	-40	402	11	3650	35	-4778	**0.19**
unknown	-182	-112	1004	-142	1577	947	2658	**0.01**
background	-536	9266	11923	-332	12465	1975	-18796	**-0**
prediction accuracy	**-0.01**	**-0.13**	**-0.11**	**0.13**	**-0.23**	**-0**	**0**	**-0**

Colour scale: 0.00 – 0.25

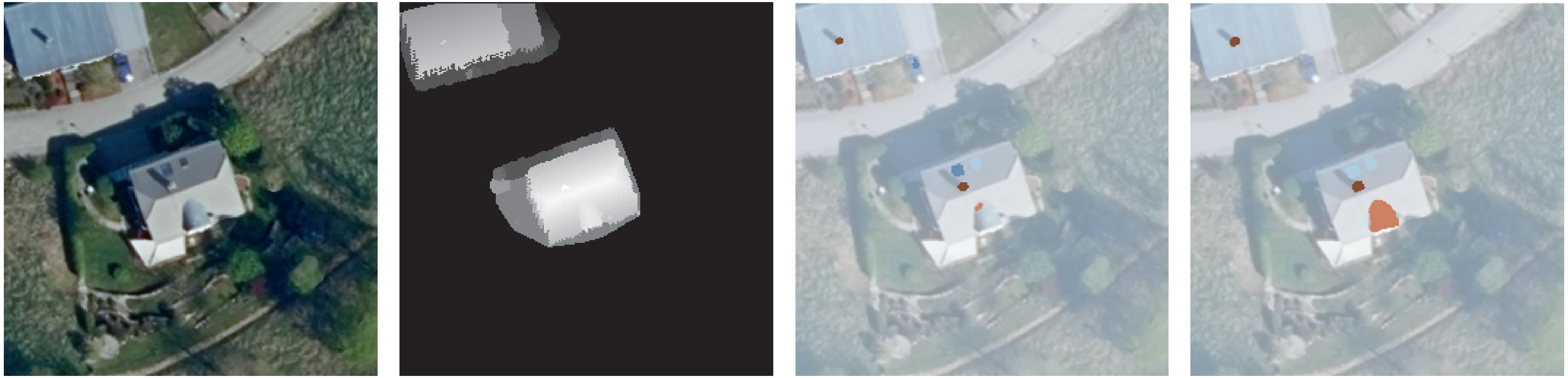


5. Analysis

Qualitative analysis | Comparison U-Net and FuseNet

Volumetric classes are better recognized: example, dormers

AP | Height map, abs *nn* | U-Net prediction | FuseNet prediction, abs *nn*

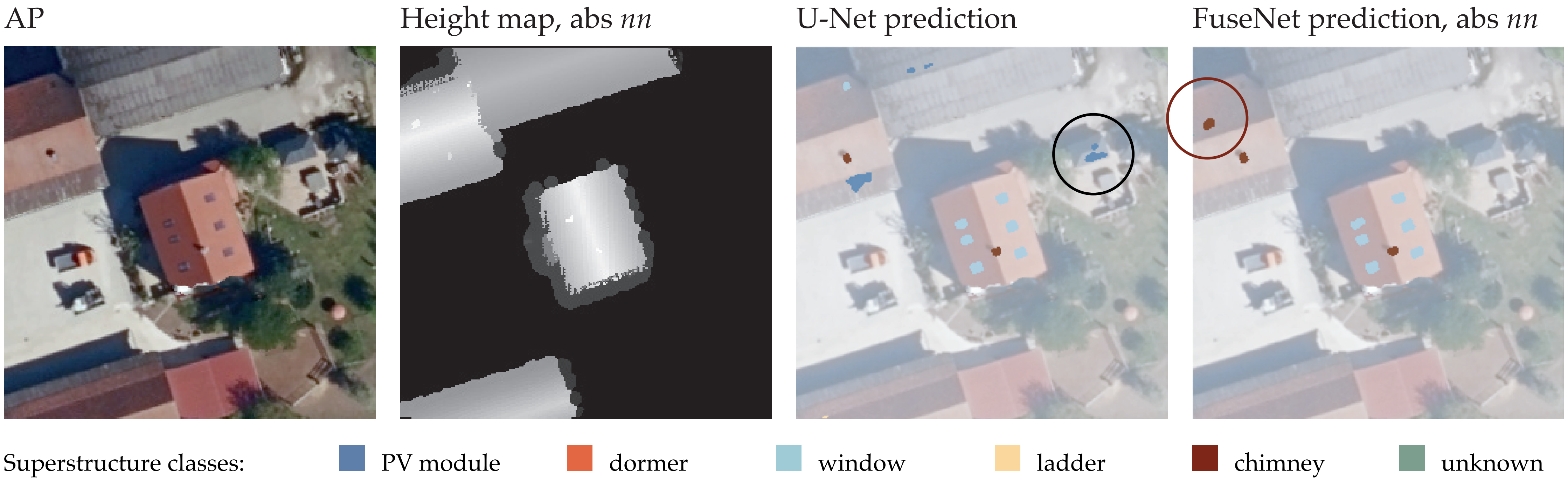


Superstructure classes: PV module, dormer, window, ladder, chimney, unknown

5. Analysis

Qualitative analysis | Comparison U-Net and FuseNet

- Superstructures always located in buildings' footprints
- Point cloud outliers or temporal mismatches ⇨ wrong detections



5. Analysis

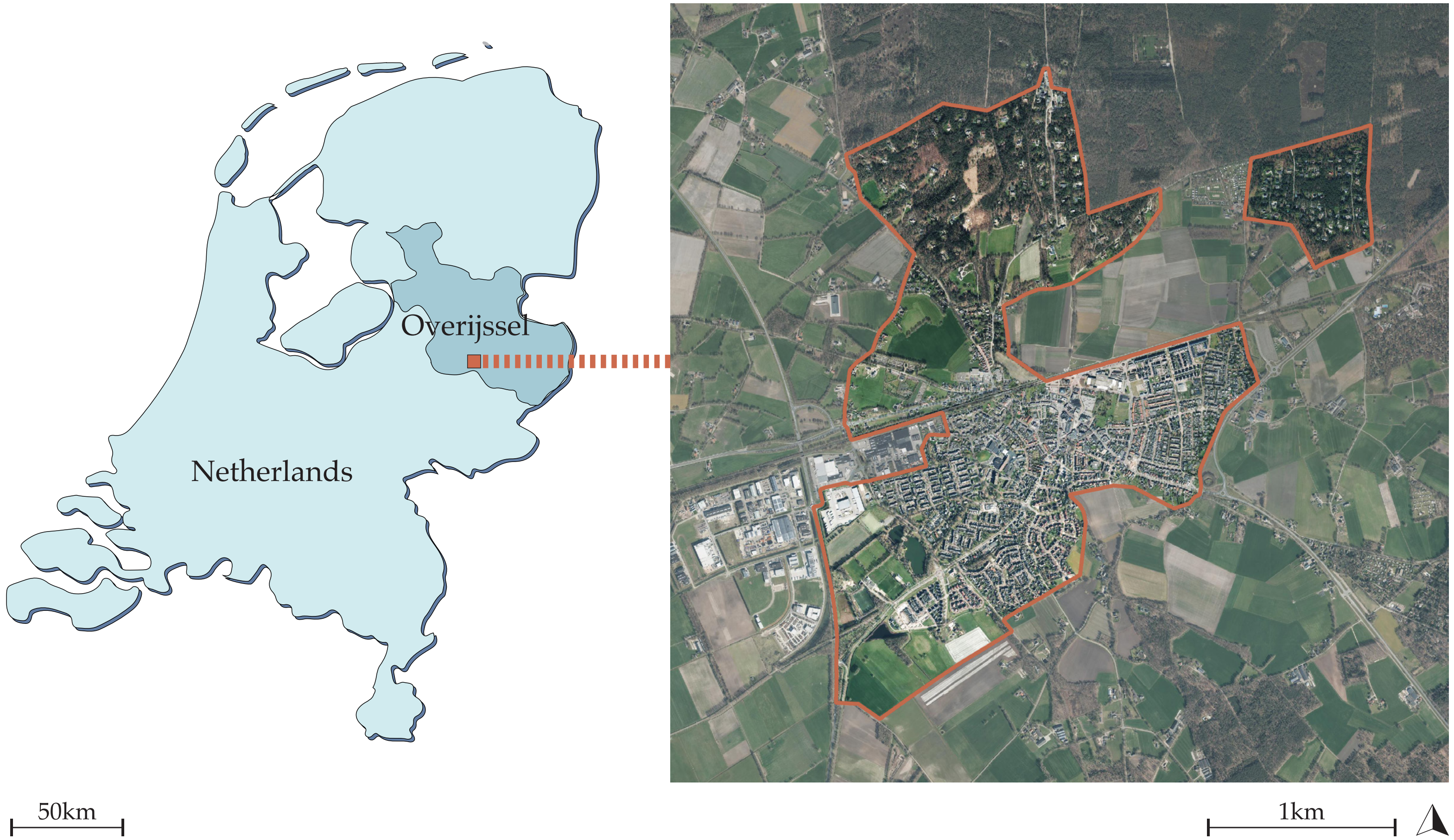
Dutch test dataset | Data preparation

New test set, aims:

- Evaluate performance on a different geographical area
- Confirm scalability of the method

New test set, implementation:

- Holten, the Netherlands
- Usage of datasets openly available (PDOK, 3D-BAG and AHN3)
- Qualitative analysis only



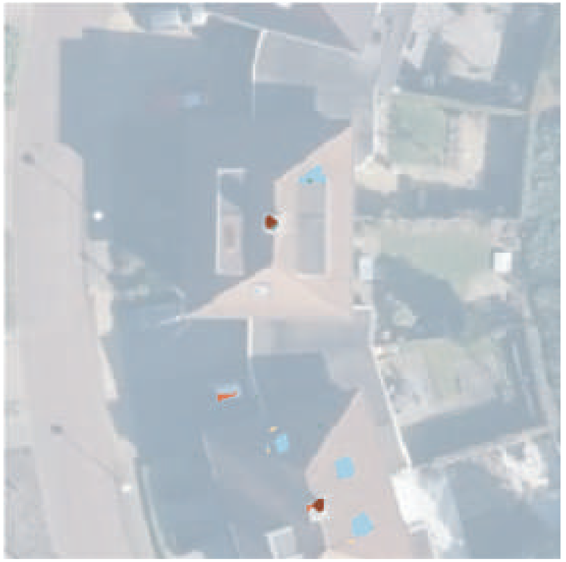
5. Analysis

Qualitative analysis, Holten | Comparison U-Net and FuseNet

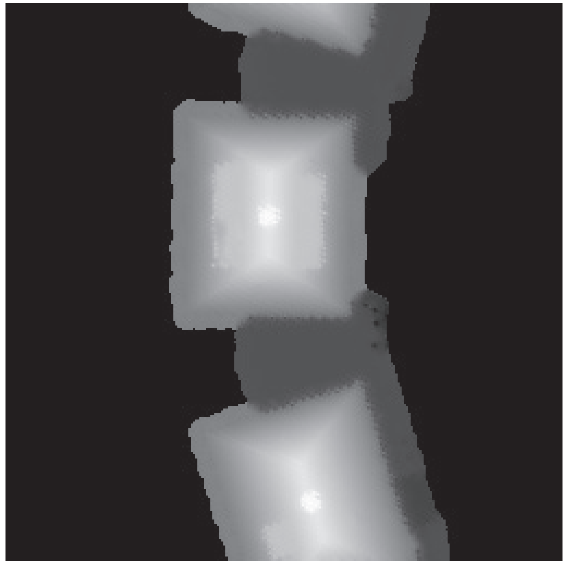
✓ FuseNet detects better chimneys (and dormers)

AP

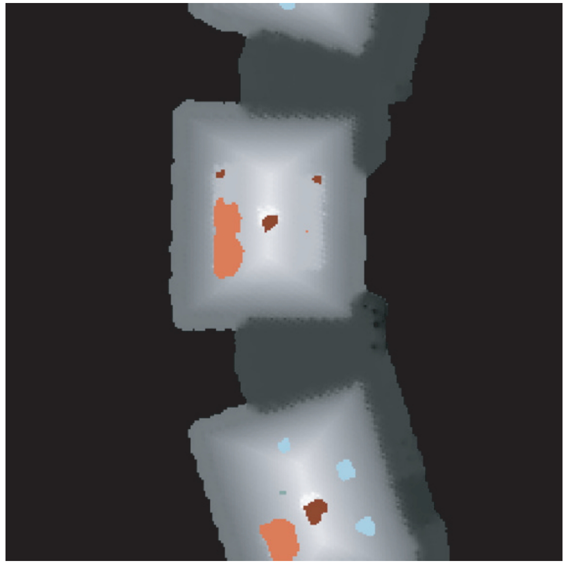
U-Net, with augm.



Height map



FuseNet, with augm.



Superstructure classes: PV module, dormer, window, ladder, chimney, unknown

5. Analysis

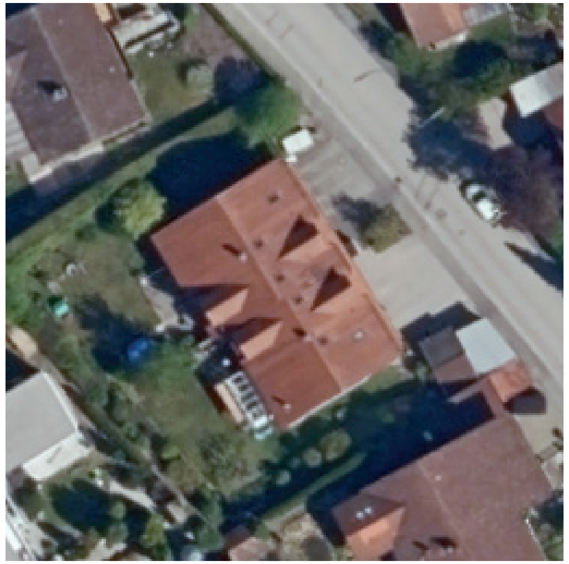
Qualitative analysis, Holten | Comparison train and test datasets

Impact of architectural typologies:

○ Dormers are different between training and test set (gabled versus flat boxes, and different materials)

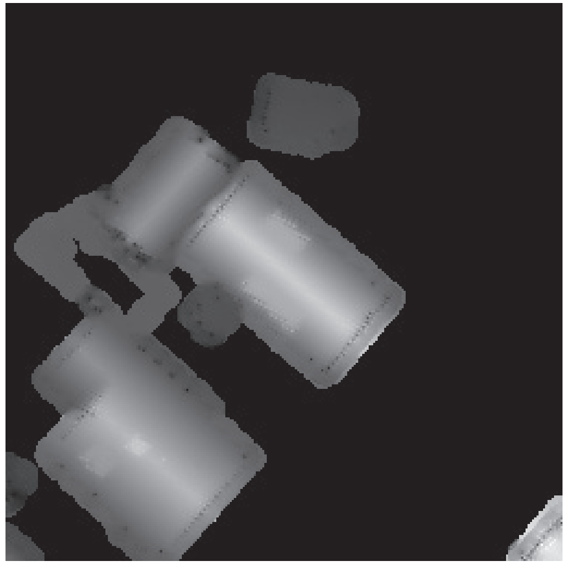
➪ Narrow dormers are detected. Flat dormers are partly detected or confused with PV modules

Dormers Wartenberg



Dormers Holten

Height map



FuseNet, with augm.

Superstructure classes: PV module, dormer, window, ladder, chimney, unknown

5. Analysis

Qualitative analysis, Holten | 3D model

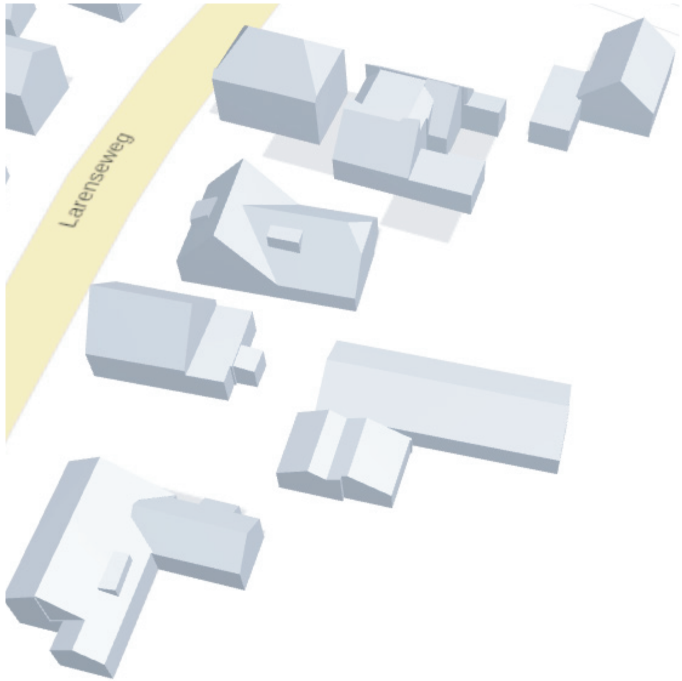
3D modelization of predictions (Bruhse 2022):

- Volumetric and planar LOD3 geometries, including usage description
- Comparison to ground truth and LOD2.2 model available for the Netherlands (3D BAG)

GT, Holten

Modelization of predictions, Holten

3D BAG, LOD2.2 model



5. Analysis

Wartenberg | 3D model application

Modelization of ortho-labels:

- Unprecise geometries
- But one can deduce the available surface

PV-application:

- Available surface ➪ PV disposition
- Roof azimuth ➪ Power generation

source: Willenborg 2022

Conclusion

Answer to problem & hypotheses

Main question:

How can building height data fused to a Convolutional Neural Network (CNN) on RGB aerial images improve the semantic segmentation of roof superstructures?

Hypotheses:

- ✓ Improvement for specific classes → Chimneys & dormers
- ✗ Relative height provides better results than the absolute one → Absolute one is preferred
- ✓ Interpolation methods provide better results than « no » interpolation → Provides more context

Conclusion

Contributions & future work

Contributions:

1. Semantic segmentation at the building scale is improved through the fusion of height data (LiDAR)
2. Knowledge about the labeling process:
   - Usage of ortho-rectified datasets,
   - Impact of geographical location on label classes
3. Outcomes on height data type to use:
   - Absolute height data extracted from LiDAR data → provides more structural information
   - IDW or NN interpolation → provides information about the superstructure scale

Conclusion

Contributions & future work

Future work:

- Improve the evaluation metrics by implementing an uncertainty factor per pixel;
  IoU is a limited criterion, not adapted to highly imbalanced classes
- Try other fusion network architectures
  (Symmetric, more recent networks)
- Improve pipeline (speed, process) to generate height data grids and 3D city model refinement

Thank you for your attention !