Roof Structure Extraction from Remote Sensing Images

Hsin-Yu Cheng

First Supervisor: Dr. Liangliang Nan

Second Supervisor: Dr. Weixiao Gao

Co-reader: Dr. Seyran Khademi

Search your house on Google Map

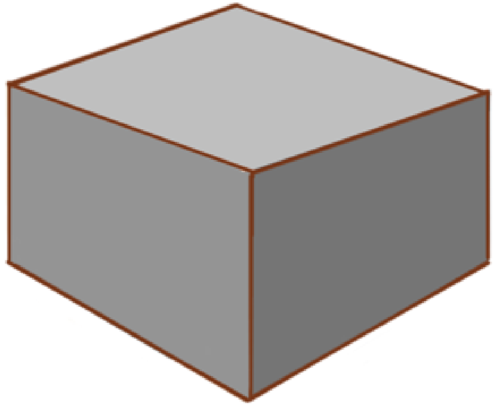
Image from Google Maps, © 2025 Google

01

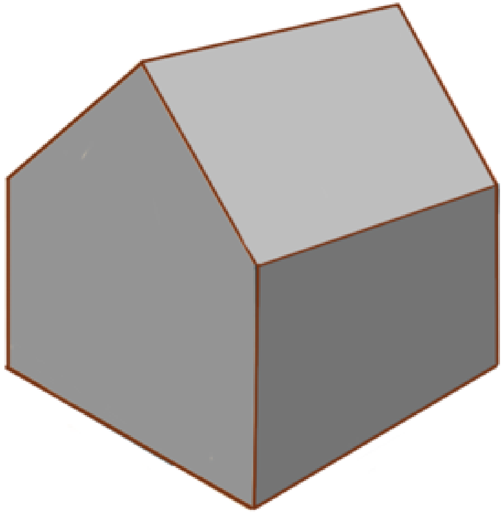
Introduction

Why Structure Matters Beyond Segmentation

Building



Segmentation output



Structure extraction

Applications that Depend on Roof Structure

- Importance of roof part structural information

Solar Potential Estimation

Drainage Simulation

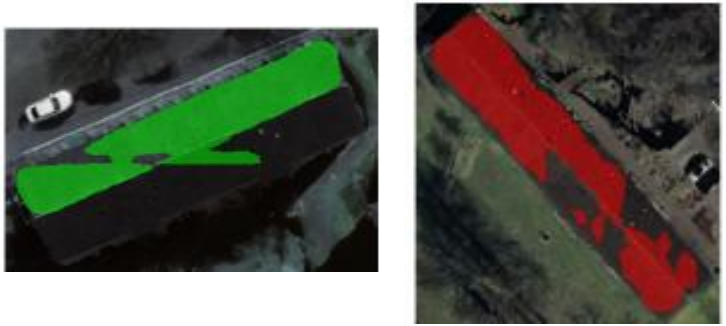
Green Roof Design

Building Safety Monitoring

TU Delft

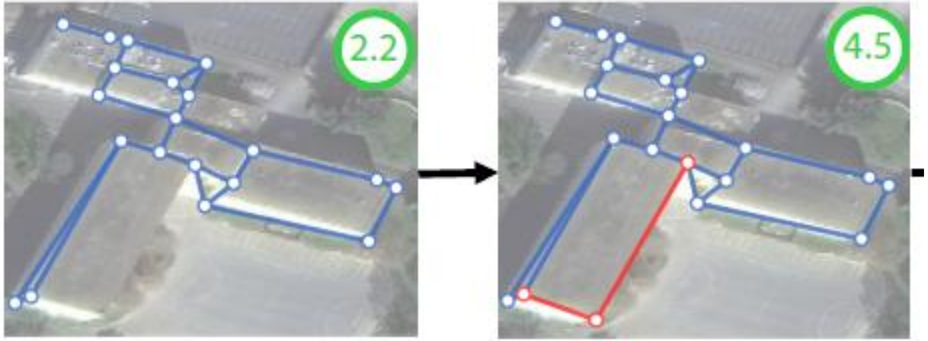
But Roof Structure is not that simple...

- Predict **pixel-based mask** → **Extract polygon outlines**
- Predict **corners and edges** → **Connect or refine** geometry



Pixel-based mask lacks geometric structure information, leading to:

- Blurry boundaries
- Polygon distortion



Zhang et al., CVPR 2021

- Unclosed shapes

02

Related Work

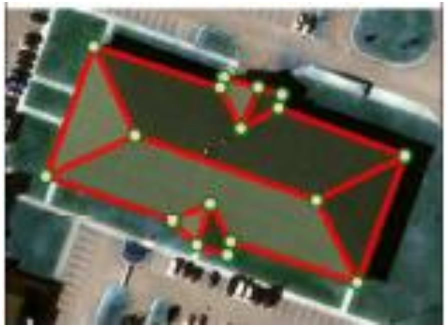
Related Work: Segmentation Methods

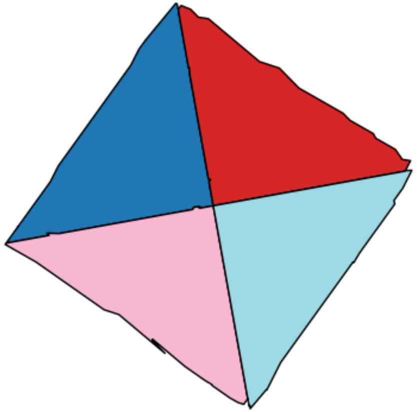
Building Segmentation

Chicchon et al., 2024

Building Instance Segmentation

Related Work: Structure-based Approaches





Roof Line Extraction

Roof Planar Structure Extraction

Zorzi et al., CVPR 2022

Nauata et al., CVPR 2020

Related Work: Roof Planar Extraction Methods

Reference	Contribution	Limitation
Lussange et al., 2023	2-stage: predict corners & height, bypass mask-to-polygon step	No recovery if mask is wrong
Xu et al., 2024	DSM + MLP + CCA to assemble polygon planes from corner lines	Sensitive to DSM quality; fails on blurry edges, computationally heavy

Research Question

- Main Research Question:

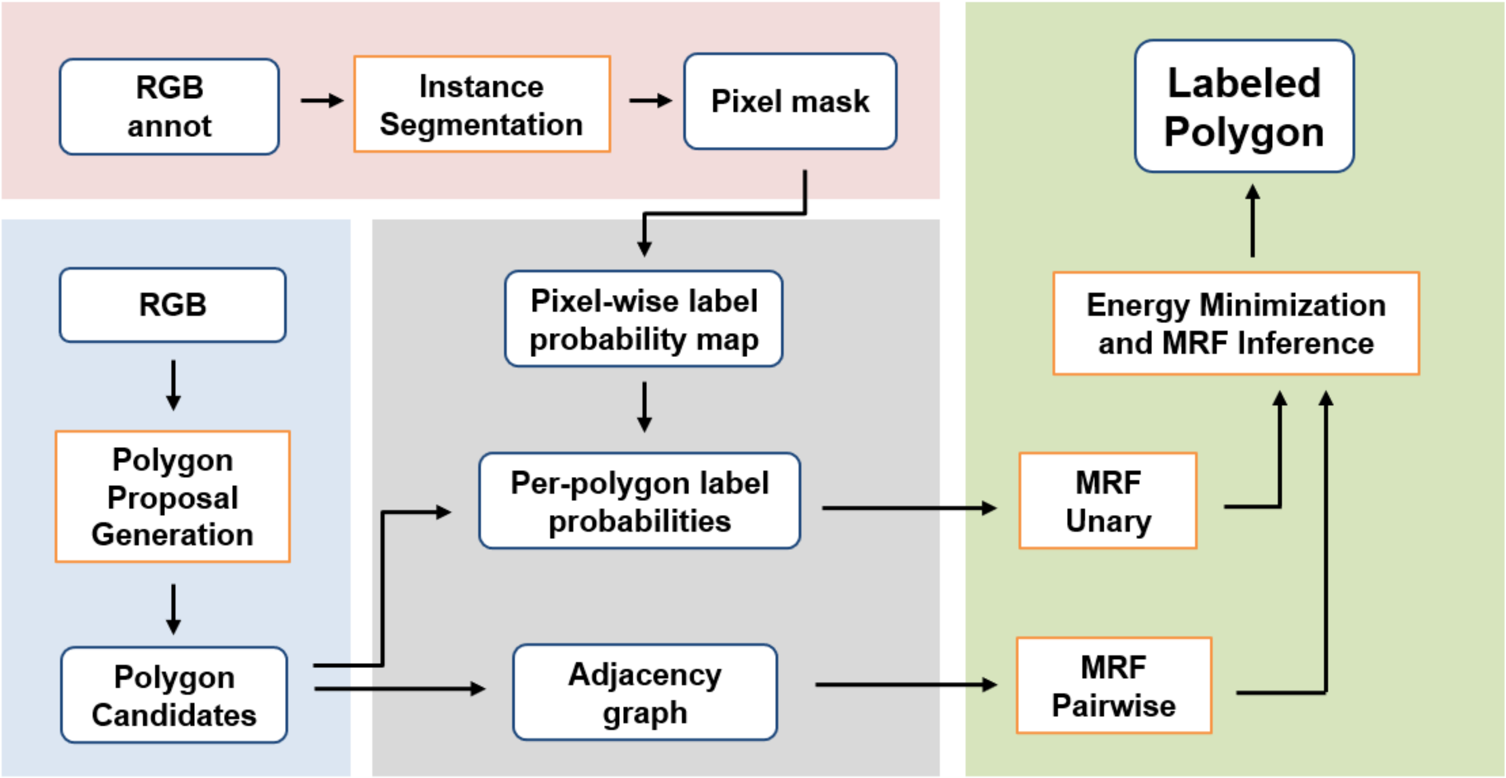
How can we extract roof planes from remote sensing images?

- Sub-Questions:
  - **Polygonal Conversion**: From pixel-based masks to structured polygons
  - **Plane Assignment**: Grouping roof fragments into coherent planes
  - **Efficiency and Compactness**: Doing all of this reliably and fast

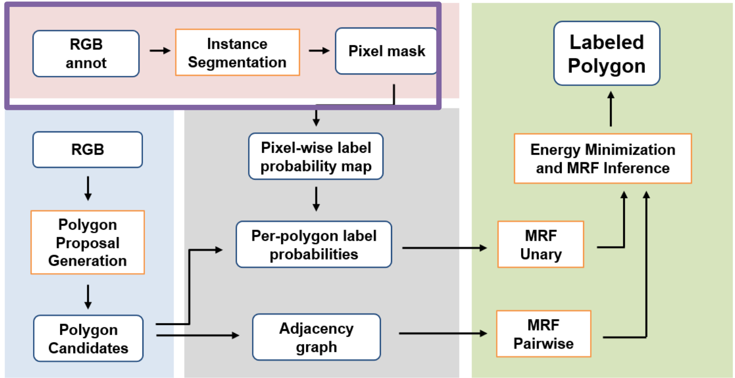
03

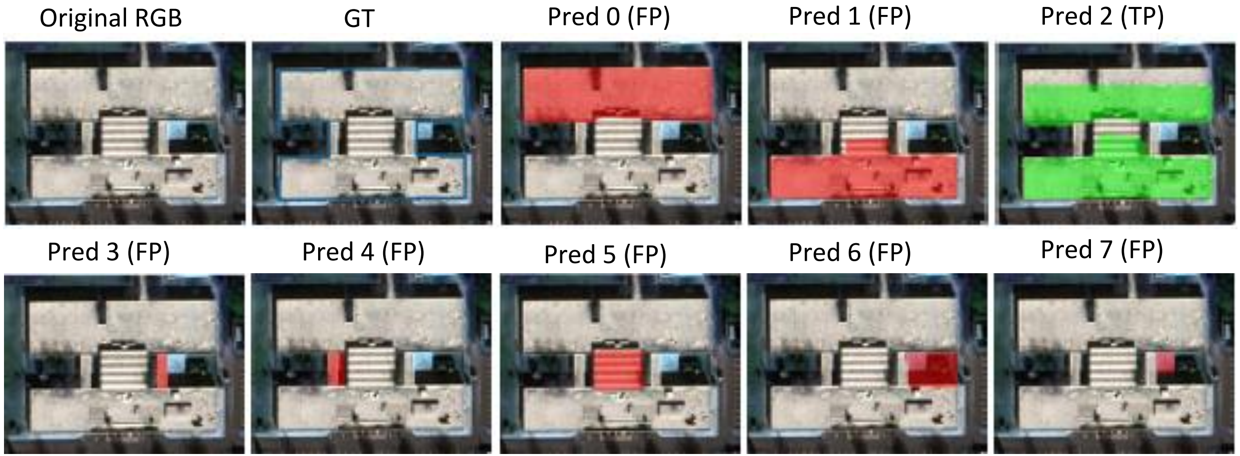
Methodolgy

Overview of the Framework



Instance Segmentation



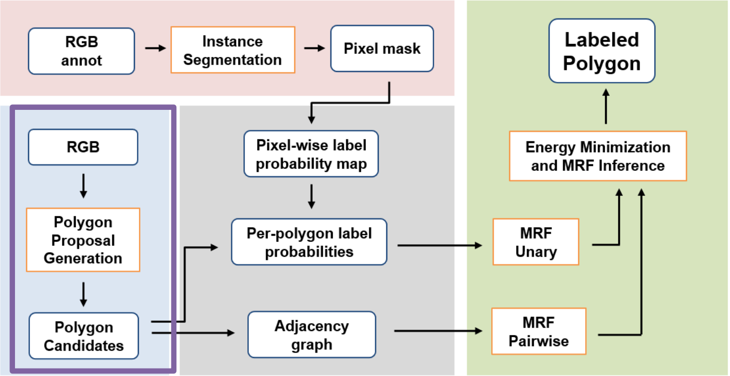


- Input: RGB, annot
- Output: pixel-wise semantic segmentation mask, confidence score

Problem:

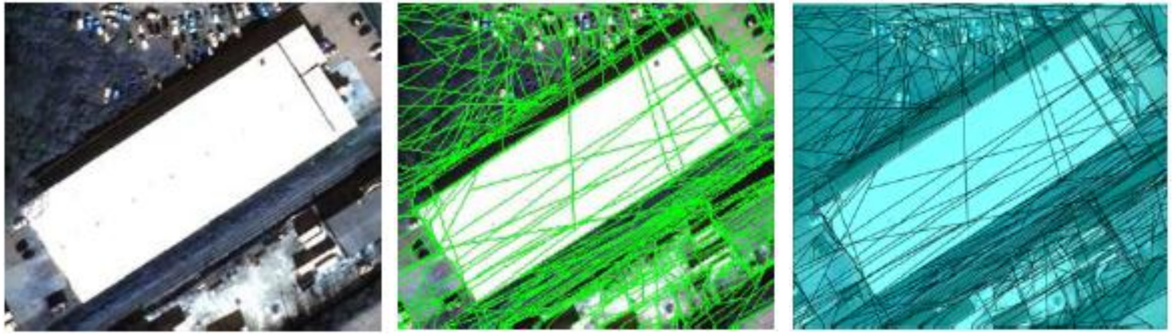
- Fuzzy boundaries
- Fragmented shapes
- Redundant overlapping predictions

Polygon Proposal Generation



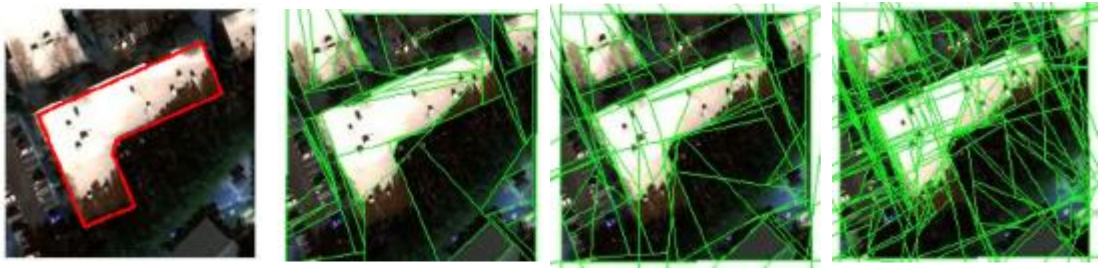
KInetic Polygonal Partitioning of Images

Original image → **Roofline detected using KIPPI** → **Aggregate rooflines into polygons**



- Input: RGB
- Output: polygons

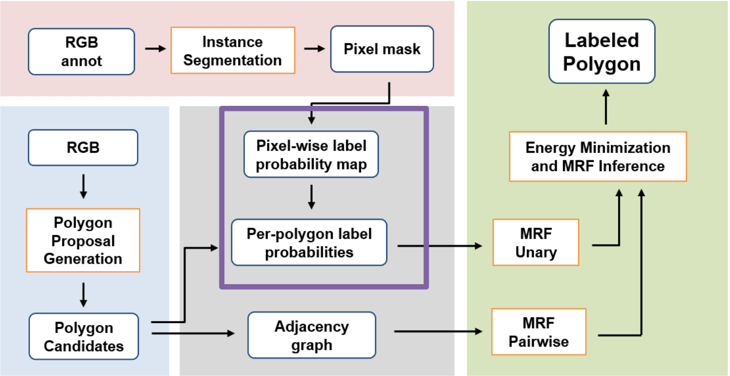
Different Parameter result



- Lsd_scale
- Number_intersection

TU Delft

Pixel-wise Probabilities
Per-polygon Label Probability



Confidence-weighted mask probability :

$$P_i(x,y) = s_i \cdot M_i(x,y) \qquad P_{bg}(x,y) = 1 - \max_i P_i(x,y)$$

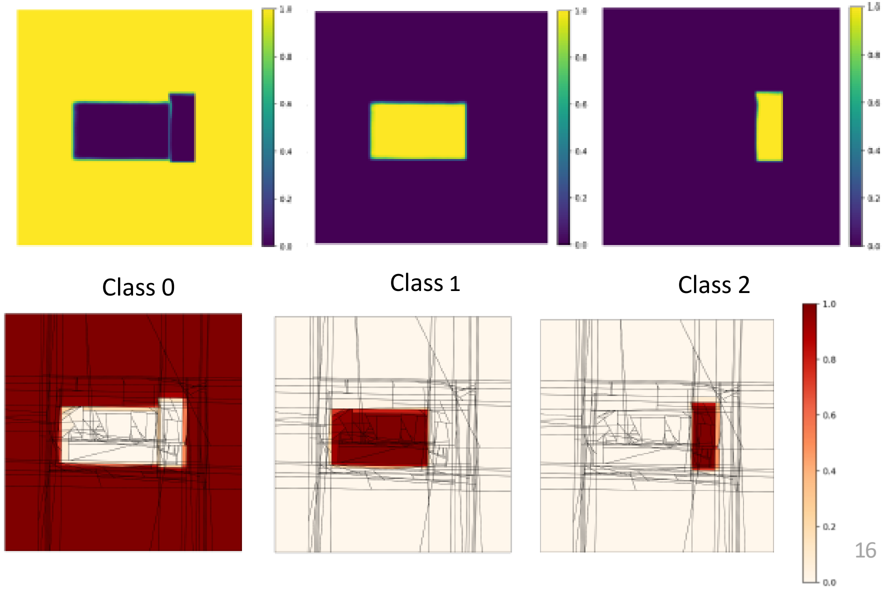
probability $M_i(x,y)$
classification confidence score s_i

Per-polygon label probability :

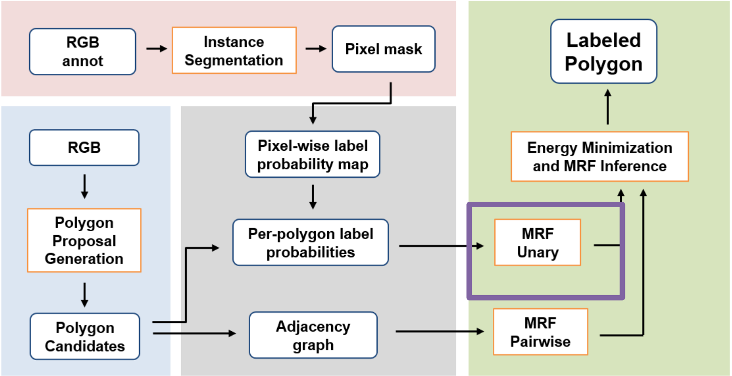
$$p_{k,j} = \frac{1}{|\Omega_k|} \sum_{(x,y)\in\Omega_k} P_j(x,y)$$

binary mask Ω_k

Original RGB



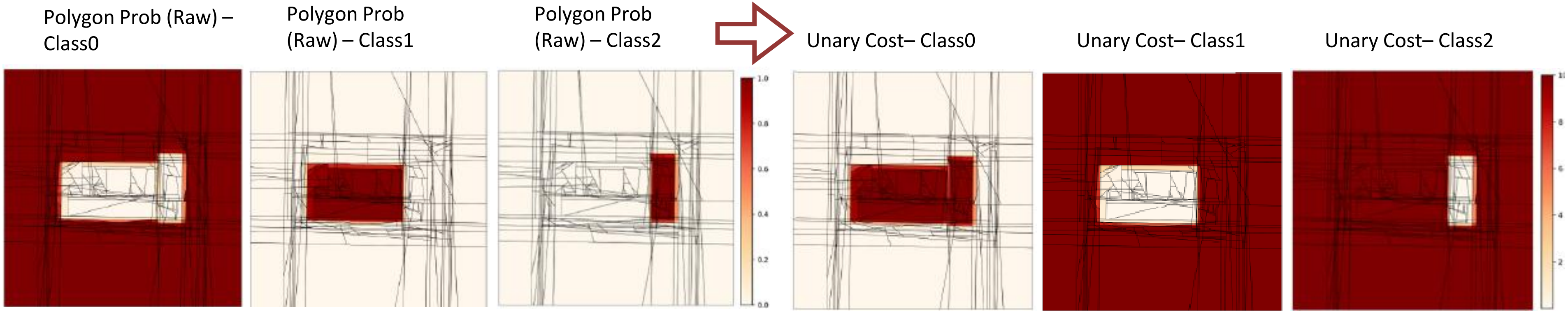
MRF Unary



Unary Cost Transformation :

$$U_{k,j} = (1 - p_{k,j}) \cdot \alpha$$

α scaling parameter

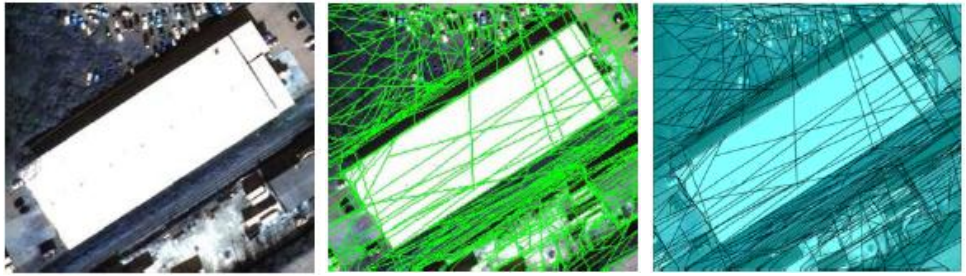


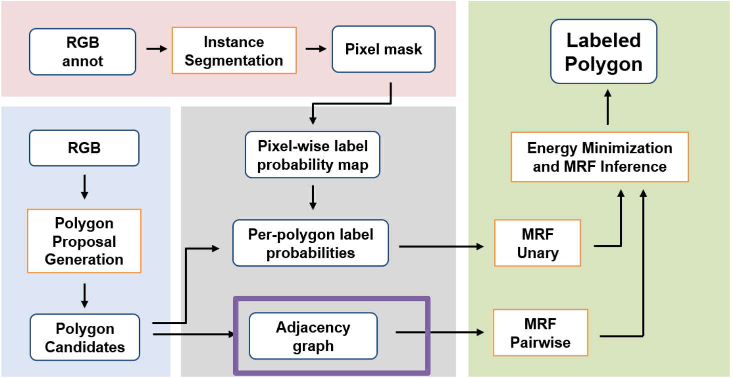
Unary Cost: Classification confidence of each polygon (internal evidence, polygon itself)

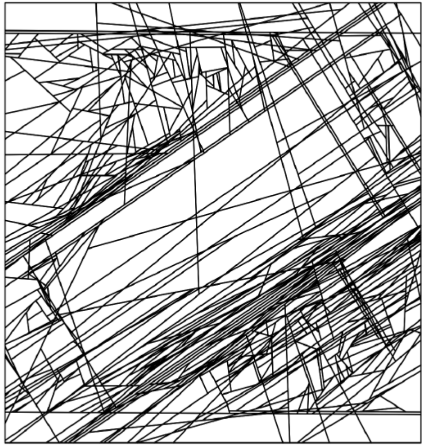
```
# Shape: (num_polygons, num_classes)
[[0.9313, 0.0, 0.0687],
 [0.6558, 0.0, 0.3441],
 [1.0000, 0.0, 0.0000]]
```

```
# After cost transform (alpha=10):
[[0.6867, 10.0, 9.3133],
 [3.4415, 10.0, 6.5585],
 [0.0000, 10.0, 10.0000]]
```

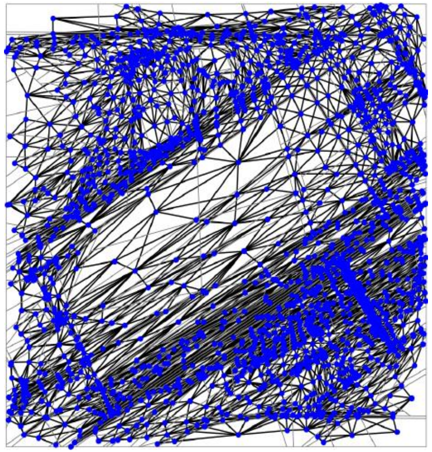
Adjacency Graph



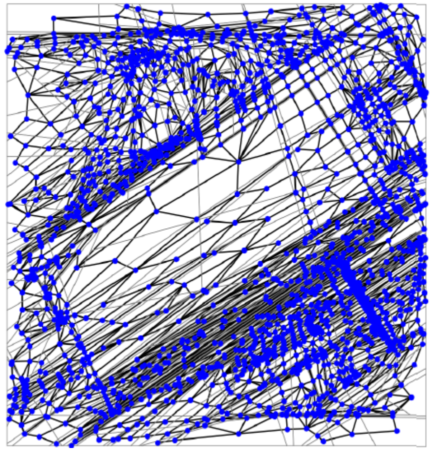




Original Polygons

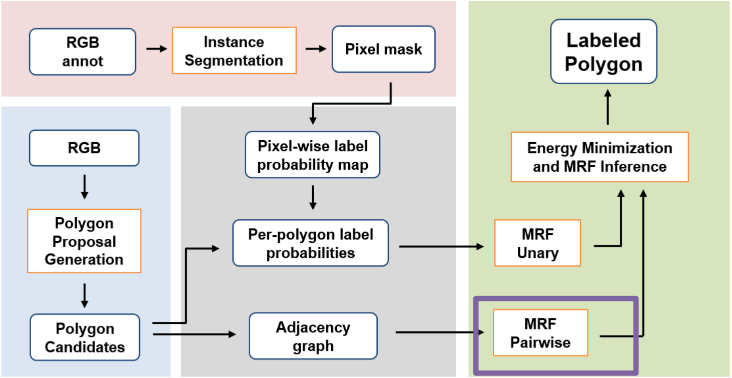


Point Adjacency



Line Adjacency

MRF Pairwise

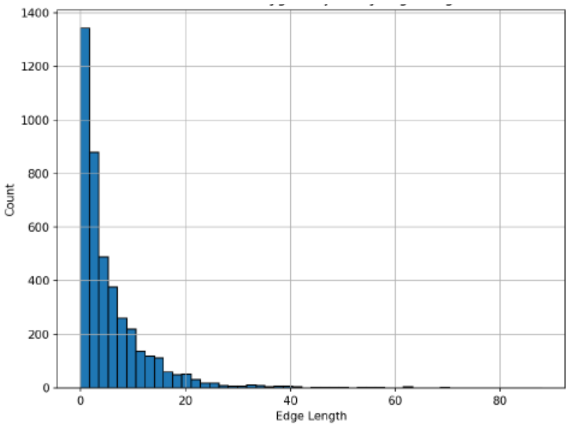


Edge weight normalization:

$$w_{ij} = f\left(l_{ij}\right) = \sqrt{l_{ij}/l_{max}} \cdot scale + offset$$

$l_{max} = 80$ pixels, $scale = 5$, and $offset = 1$

shared boundary length l_{ij}



Label penalty matrix:

$$V(a,b) \begin{cases} 0 & if \;\; a = b \\ 1 & \text{otherwise} \end{cases}$$

Final pairwise cost:

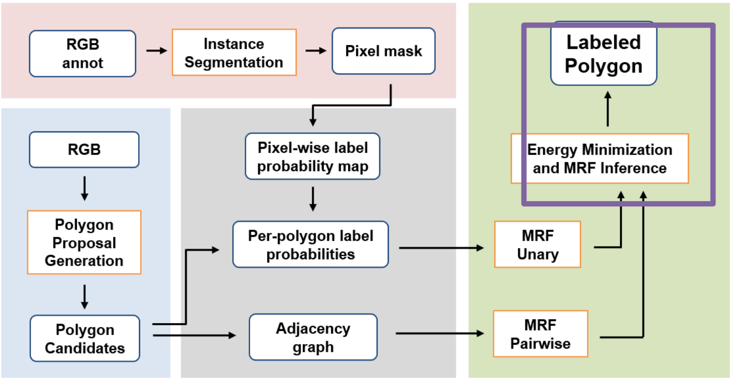
$$\text{PairwiseCost}_{ij}(a,b) = \lambda \cdot w_{ij} \cdot V(a,b)$$

hyperparameter λ $\lambda \in \{0.001, 0.01, 0.1, 1, 10, 100\}$

MRF Inference

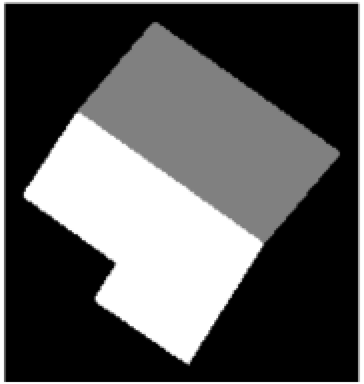
Given the **Unary term** and the **Pairwise term**, the final label configuration L = $\{L_1, L_2, \dots, L_N\}$ is obtained by minimizing the total energy:

$$L^* = \arg\min_{L}\left(\sum_{k} U_k(L_k) + \sum_{(i,j)} \lambda \cdot w_{ij} \cdot V(L_i, L_j)\right)$$



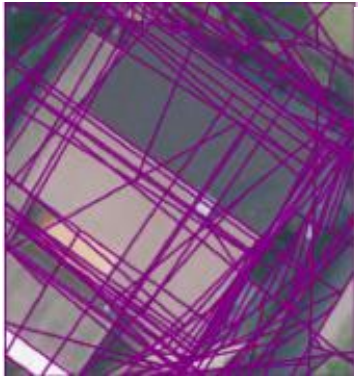
RGB

GT

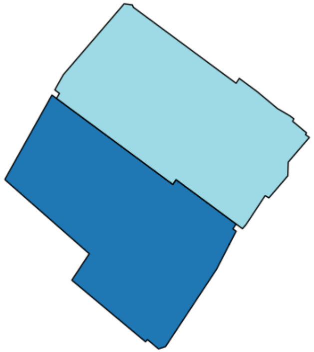


Instance segmentation

Polygon candidate



MRF



04

Results and Evaluation

Results

$$\text{IoU} = \frac{|\text{Prediction} \cap \text{GroundTruth}|}{|\text{Prediction} \cup \text{GroundTruth}|}$$

Smoothness Cost	TP	FP	FN	Mean IoU
0.001	749	319	200	0.8331
0.01	749	316	200	0.8331
0.05	749	313	200	0.8333
0.1	748	306	201	0.8336
0.5	738	277	211	0.8260
1	723	262	226	0.8179
10	373	84	576	0.5064
100	0	1	949	0.0000

MRF sweep results on the Cities dataset (unary scale = 10)

Smoothness Cost	TP	FP	FN	Mean IoU
0.001	3265	111	90	0.9125
0.01	3264	109	91	0.9127
0.05	3263	99	92	0.9130
0.1	3262	94	93	0.9130
0.5	3261	73	94	0.9117
1	3248	70	107	0.9049
10	1863	96	1492	0.5568
100	14	69	3341	0.0063

MRF sweep results on the RoofVec dataset (unary scale = 10)

Datasets	TP	FP	FN	Precision
Cities	870	600	197	0.5918
RoofVec	3338	356	48	0.9036

Datasets: Train/Validation/Test
Cities: 1192/397/397
RoofVec: 4584/1528/1528

Comparison of pixel-base segmentation on the **Cities** and **RoofVec** datasets

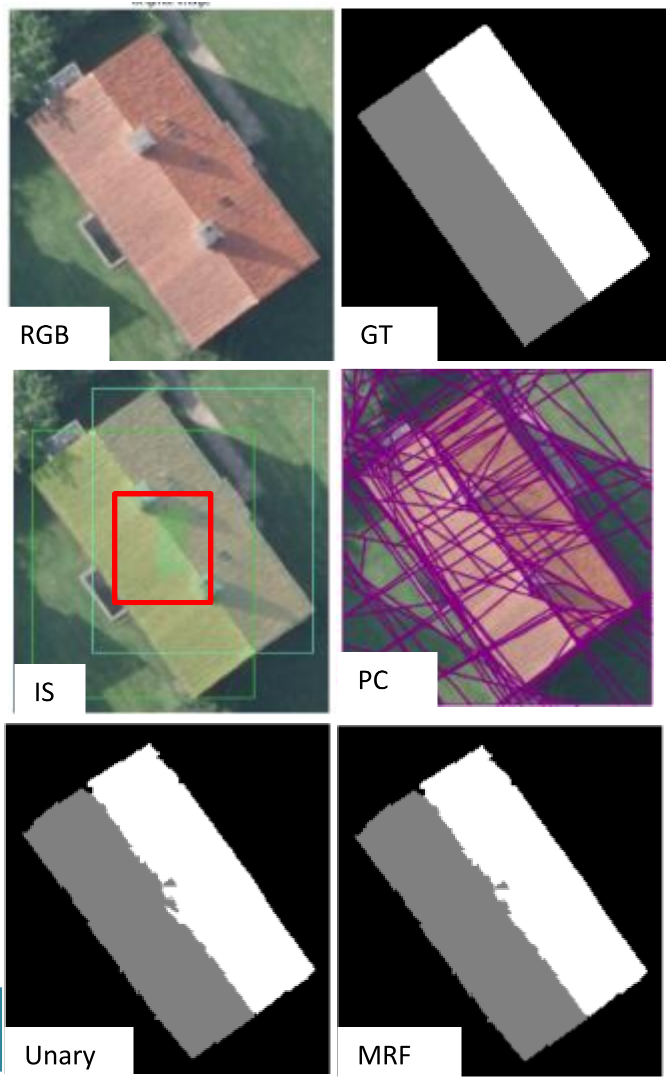
Cities	TP	FP	FN	Precision	Mean IoU
Unary term only	749	319	200	0.7012	0.8331
Whole MRF (λ = 0.1)	748	306	201	0.7097	0.8336

RoofVec	TP	FP	FN	Precision	Mean IoU
Unary term only	3265	111	90	0.9362	0.9125
Whole MRF (λ = 0.1)	3262	94	93	0.9720	0.9130

Comparison of MRF with or without pairwise term segmentation on the **Cities** and **RoofVec** datasets

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

IS: Instance Segmentation
PC: Polygon Candidate



Compare to Zhang et al. (2021)

- Use the same Cities Datasets

Table 1. Comparison against two state-of-the-art RL-based systems, Ellis *et al.* and Lin *et al.* based on the corner/edge/region f1 scores. All systems use Conv-MPN reconstructions as the initial models. orange and cyan indicate best and second best scores.

Algorithm	Corner	Edge	Region
Conv-MPN [22]	78.8	58.1	54.9
REPL [7]	72.6	43.8	15.9
Lin *et al.* [15]	74.7	51.6	35.8
+ Ours	83.2	67.1	63.5

Adapted from Zhang et al., CVPR 2021

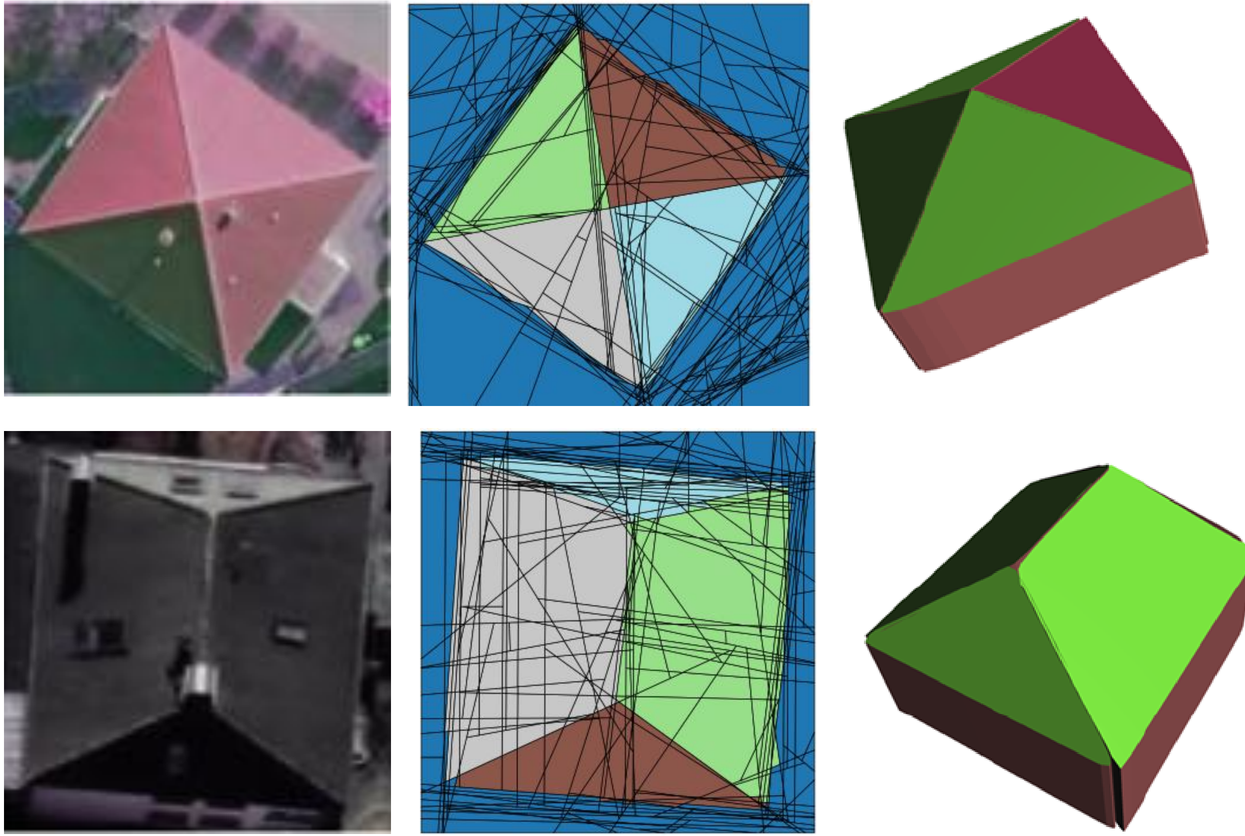
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

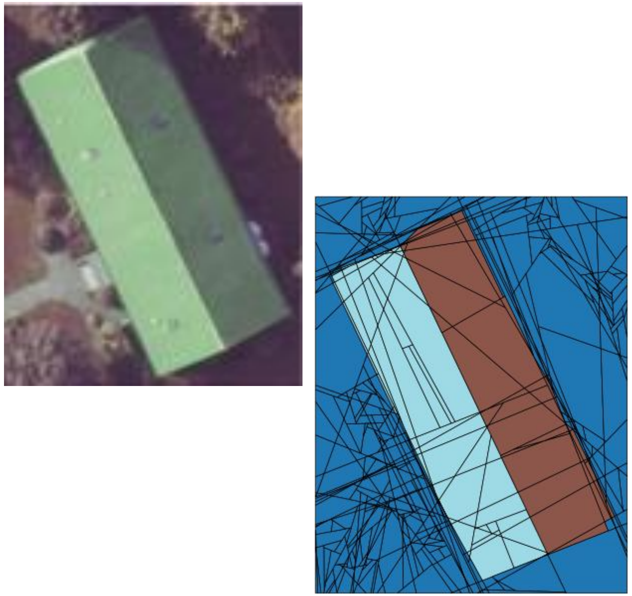
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

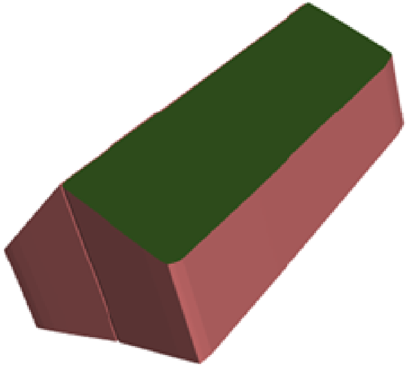
$$\text{F}_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Our Region F1-score = 0.661 × 100 = 66.1

Application Demonstration







05

Discussion and Limitation

TUDelft

Discussion

- Improves segmentation quality
- Stable performance across datasets
- Robust and efficient geometric module
- Model-agnostic and generalizable

Table 6.1.: Average runtime per step

Step	Time (s)
Load image	0.1169
Model inference	0.2164
Generate pixel probability map	0.0005
Compute polygon probabilities	1.9052
Prepare unary	0.0000
Build graph	0.6838
GCO labeling	0.1035
Rasterize prediction	0.1056
Evaluate metrics	0.0027
Smoothness stats	0.0034

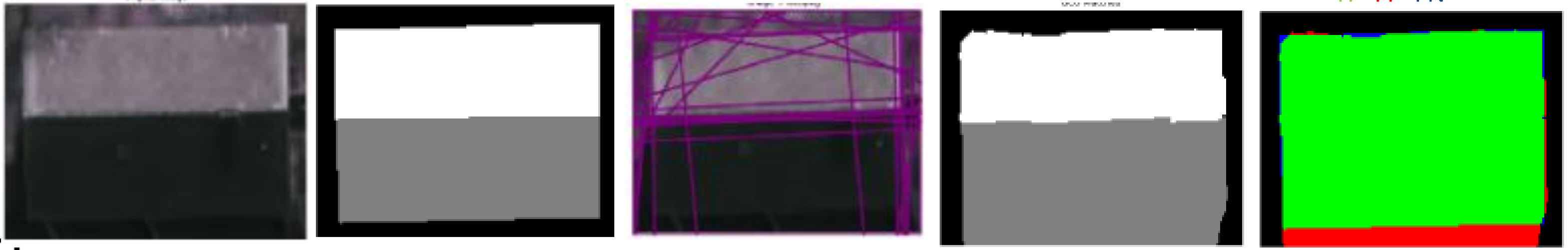
Limitation

RGB GT Instance segmentation MRF

- Limited by poor initial pixel-mask predictions
- MRF tuning requires trial and error
- Depends on polygon Candidate
- Depends on annotation quality

RGB GT Polygon candidate MRF Evaluate by pixel

TP FP FN



06

Conclusion and Future Work

TU Delft

Conclusion

Research Question	Contribution / Answer
Structured polygon from pixel masks?	Polygon proposal module (line detection + spatial partitioning) that transforms raster masks into clean, geometry-aware polygons.
Group roof parts into coherent planes?	MRF-based label refinement groups nearby polygons using classification confidence(Unary term) and geometric connectivity(Pairwise Term).
Efficient extraction with quality?	Fast and model-agnostic; processes each image in under 4 seconds while improving performance.

Future Work

- **Better Instance Segmentation**
  Explore stronger backbones to reduce FP and improve mask coverage.

- **Polygon Proposal Analysis**
  Connect proposal parameters to MRF outcomes; explore filtering/merging strategies.

- **MRF Term Refinement**
  Improve unary weighting and pairwise design for better structure in noisy areas.

Reference

Lussange, J., Yu, M., Tarabalka, Y., and Lafarge, F. (2023). 3d detection of roof sections from a single satellite image and application to lod2-building reconstruction. arXiv preprint, arXiv:2307.05409.

Muneeb, M., and Lin, Y. (2020). Building semantic segmentation using UNet convolutional network on SpaceNet public data sets. In Proceedings of the 2020 International Conference on Image, Video Processing and Artificial Intelligence, pages 89–96. SPIE.

Nauata, N., and Furukawa, Y. (2020). Vectorizing world buildings: Planar graph reconstruction by primitive detection and relationship inference. arXiv preprint, arXiv:1912.05135.

Xu, Y., Jubanski, J., Bittner, K., and Siegert, F. (2024). Roof plane parsing towards lod-2.2 building reconstruction based on joint learning using remote sensing images. International Journal of Applied Earth Observation and Geoinformation, 133:104096.

Zhang, F., Xu, X., Nauata, N., and Furukawa, Y. (2021). Structured outdoor architecture reconstruction by exploration and classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 12427–12435.

Zorzi, S., Bazrafkan, S., Habenschuss, S., and Fraundorfer, F. (2022). PolyWorld: Polygonal building extraction with graph neural networks in satellite images. arXiv preprint, arXiv:2111.15491.

Thank you!

Hsin-Yu Cheng

First Supervisor: Dr. Liangliang Nan

Second Supervisor: Dr. Weixiao Gao

Co-reader: Dr. Seyran Khademi