TU Delft

PHYSHADE-Net:

Leveraging Geometric Priors in Physics-Guided Neural Networks for Building Shadow Segmentation and Height Estimation

P5 Presentation | Lars Huizer

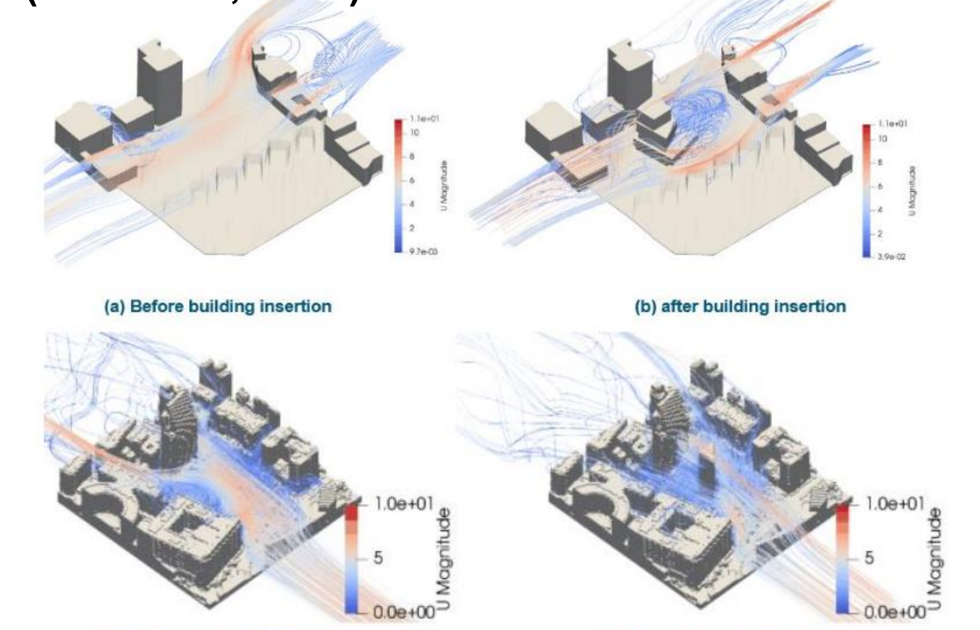
18/06/2025

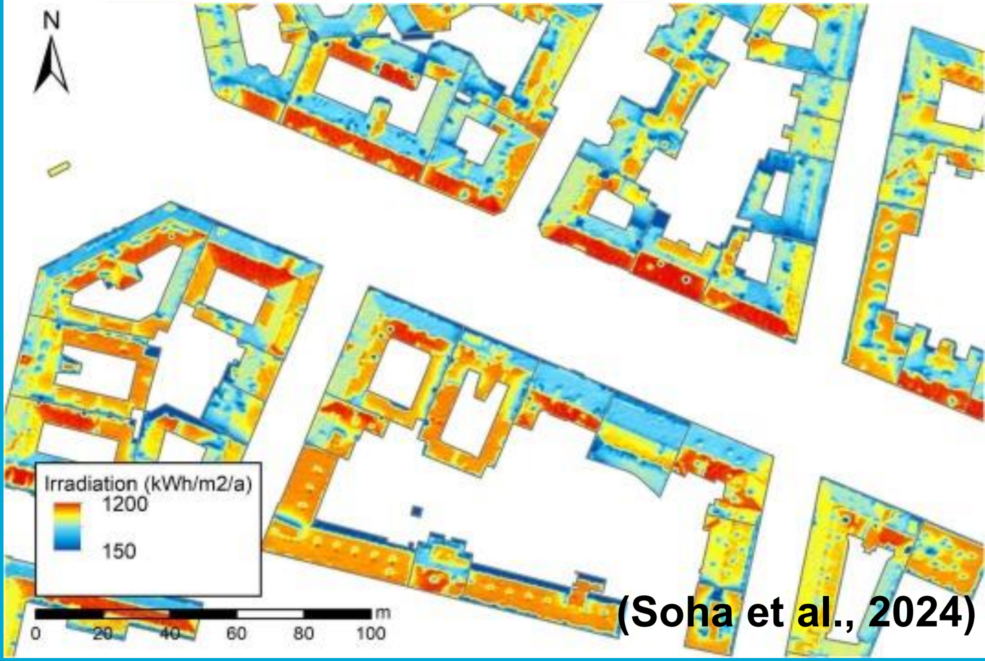
Table of Contents

Problem

- Accurate building heights are critical for a variety of purposes, such as urban planning, energy modelling, and digital twins.
- While LiDAR delivers precise data, it is costly and inaccessible to communities with limited resources
- This leads to a **division in data access**

(Tsai et al., 2024)

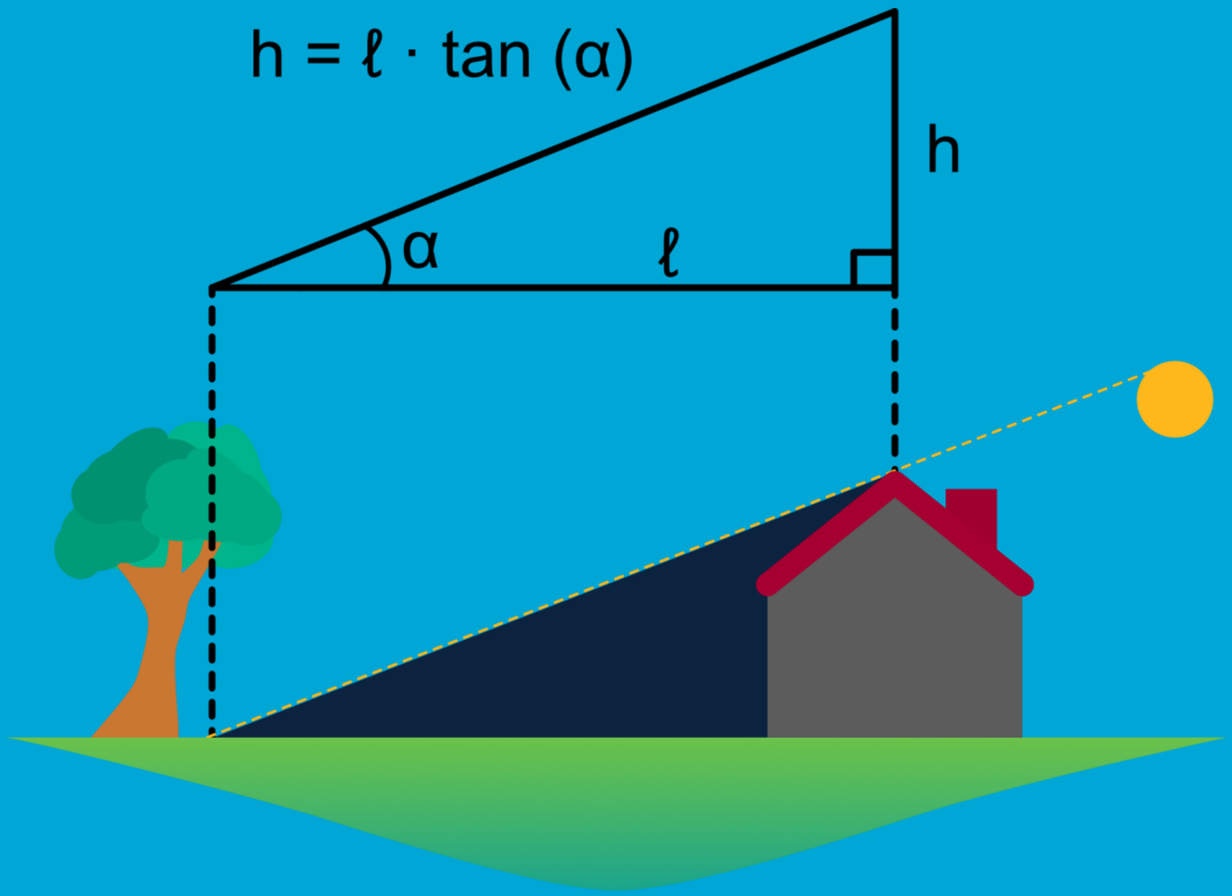




(Soha et al., 2024)

Research Gap

One solution is the derivation of building heights from **Shadow Lengths,** and **Date/Time Metadata** for solar angle.



Research Gap

- This concept of height estimation through shadows has been explored before. However, it requires the accurate detection of shadows in images for accurate results.
- Existing deep-learning approaches are relatively robust, but lack

  Contextual Awareness: What is a building shadow, and what is a regular shadow?

  Generalization: If I apply the model outside of the domain it was trained in, will it perform as well? Different lighting conditions? Places?

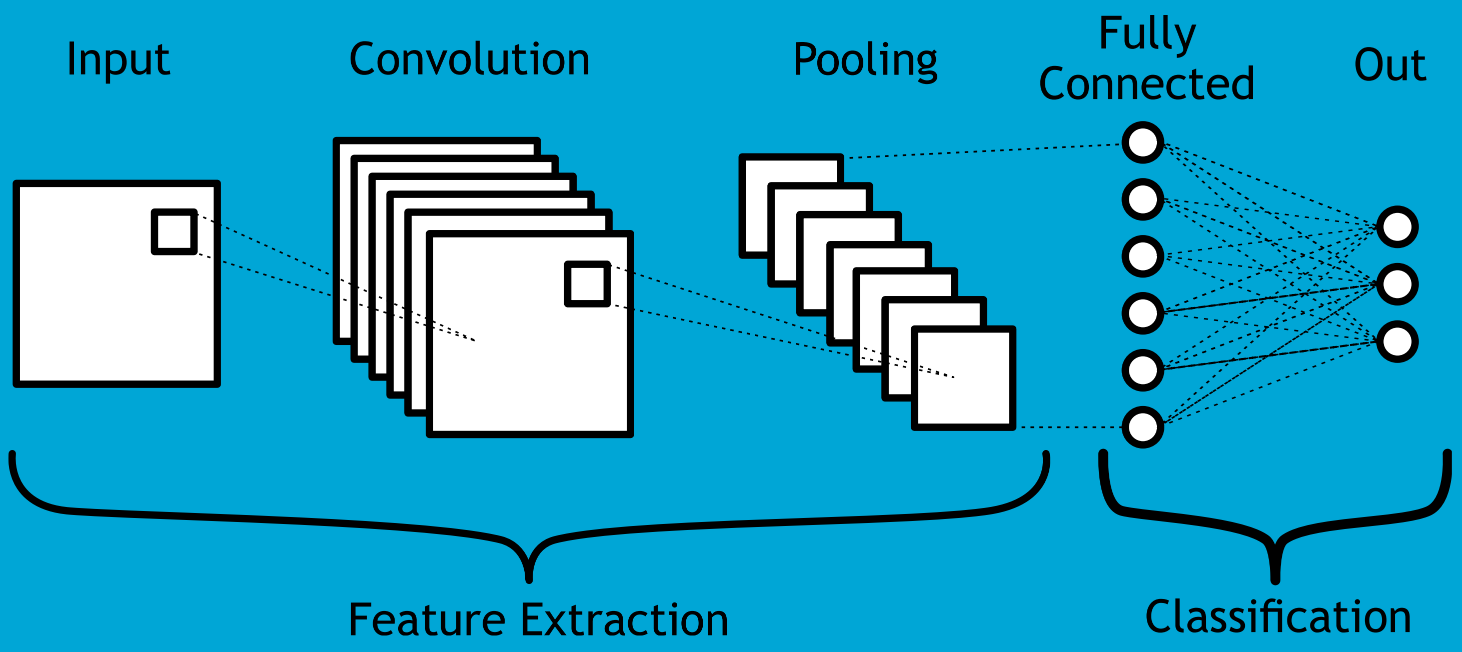
  Usage of physical priors: How can we predict where building shadows may be based upon building geometry and the position of the sun?

Research Questions

- **[Q. Main] How and to what extent does injecting geometric priors in the form of shadow masks into a U-Net architecture improve the accuracy and robustness of building-shadow segmentation in aerial imagery?**
- [Q1.] What is the baseline performance of an RGB U-Net trained on the Luo et al. (2020) dataset when evaluated on Dutch aerial imagery and compared to the original dataset?
- [Q2.] How does adding the pre-calculated shadow mask channel derived from building footprints affect segmentation accuracy across different urban morphologies, seasons, and solar geometries?
- [Q3.] Which loss formulations and weighting schemes most effectively balance appearance-based learning with the geometric prior?
- [Q4.] How effective are the inferred building shadows at estimating building height using the raster-based raycasting algorithm?

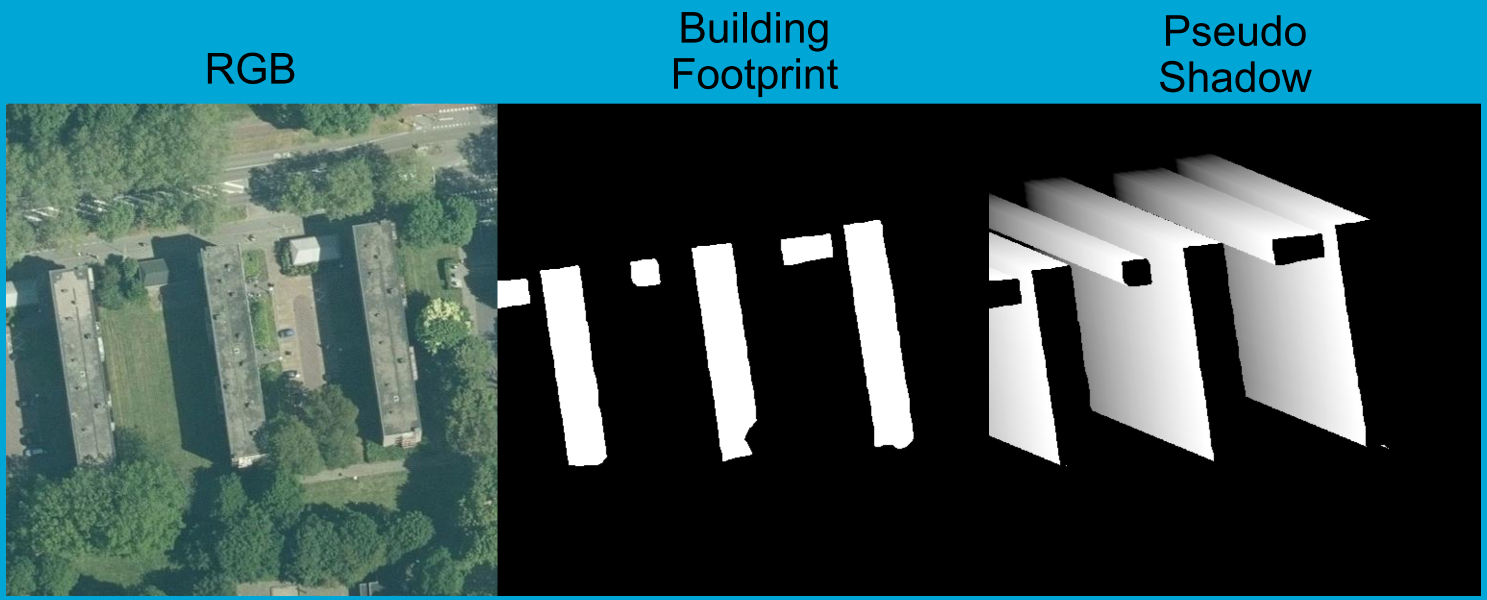
Key Concepts

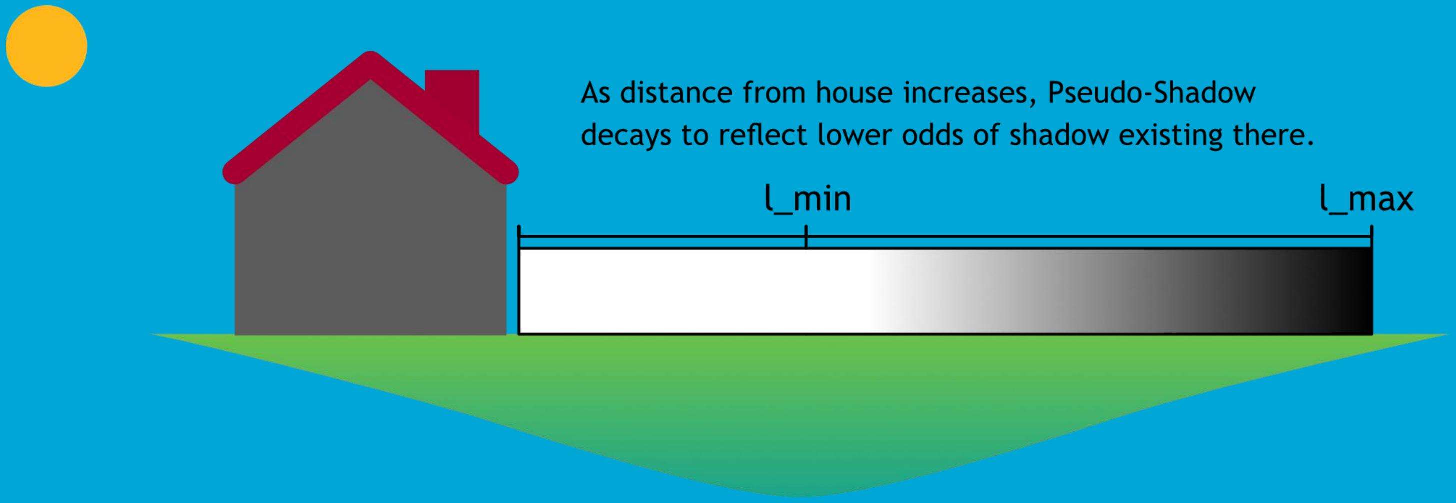
Shadow Segmentation using a CNN



$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_i p_i y_i + \varepsilon}{\sum_i (p_i + y_i) + \varepsilon}$$

- CNNs can detect features in images through **kernels**, learned **in the convolution layer**.
- A CNN learns through a **loss function**, which describes how wrong the model is by comparing its prediction to the correct answer
- This loss is then used in **backpropagation**, where the kernels are adjusted based on their contribution to the error
- Some loss functions model laws of physics to make them more accurate; these Physics-Informed Neural Networks are an inspiration for the proposed method in this thesis.



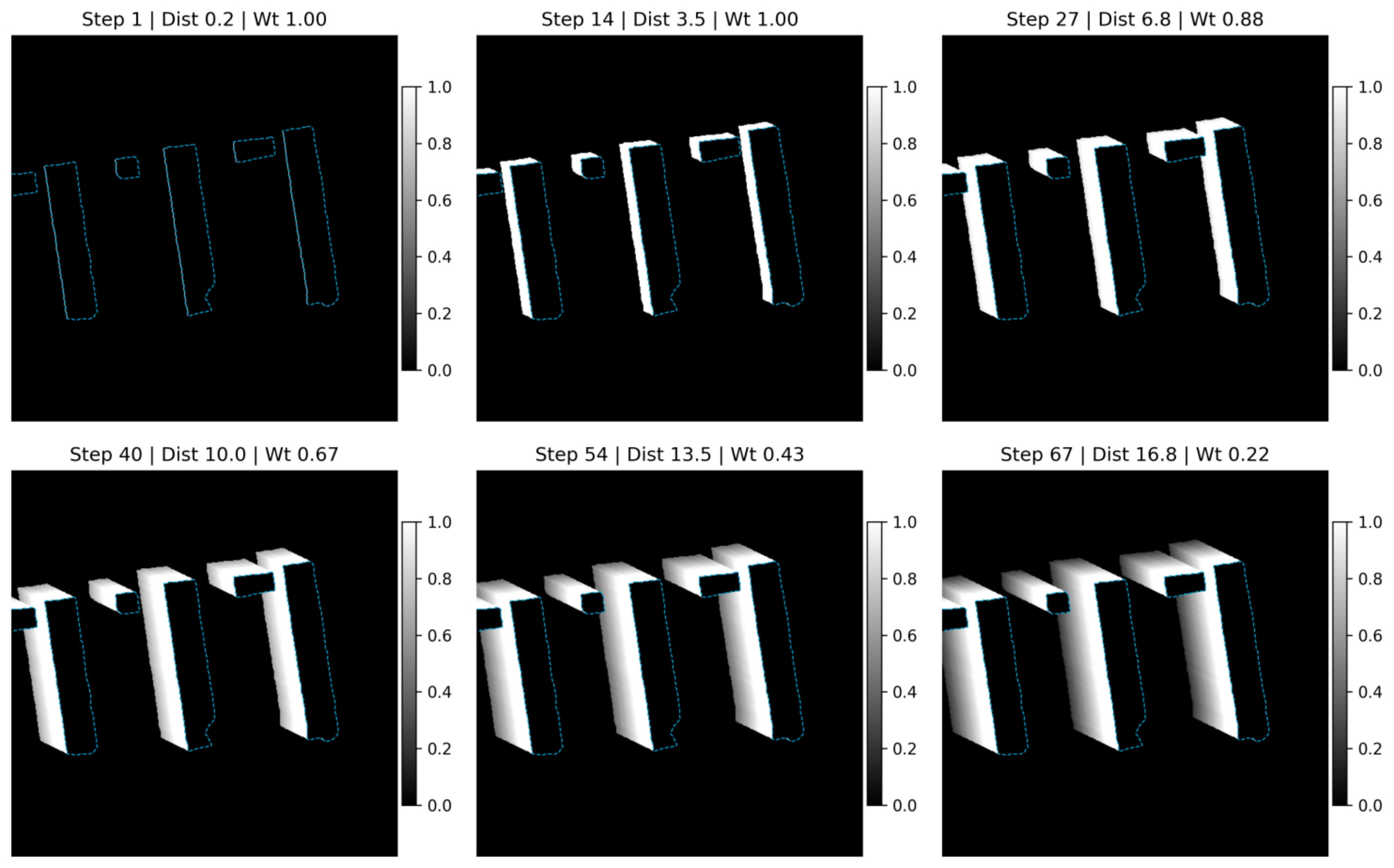


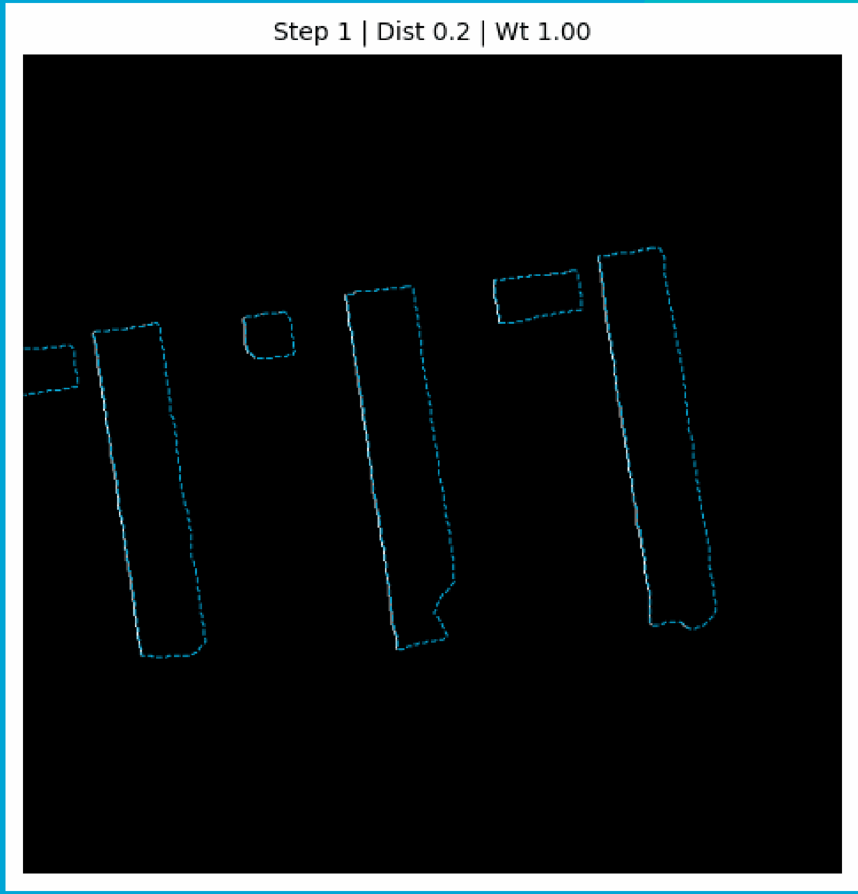
Key Concepts

Integration of Geometric Prior

- The CNN detects features by looking at the RGB channels
- However, it has no concept of how physics may influence the projection of shadows
- The model proposed in this thesis introduces this to the CNN by adding a probability field based on the buildings and the position of the sun: the **Pseudo-Shadow**
- l_min and l_max represent confidence of a shadow existing to that point.
- l_min = 2 meters,
  l_max = 42.90 meters (≈ 95th percentile of building heights in Netherlands)

TU Delft


Step 1 | Dist 0.2 | Wt 1.00
Step 14 | Dist 3.5 | Wt 1.00
Step 27 | Dist 6.8 | Wt 0.88
Step 40 | Dist 10.0 | Wt 0.67
Step 54 | Dist 13.5 | Wt 0.43
Step 67 | Dist 16.8 | Wt 0.22
1.0
0.8
0.6
0.4
0.2
0.0


Step 1 | Dist 0.2 | Wt 1.00

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_i p_i y_i + \varepsilon}{\sum_i (p_i + y_i) + \varepsilon}$$

Key Concepts

Loss Functions

- Dictate how a deep-learning network learns by evaluating difference between ground truth and model output.
- Depending on the formularization of the loss function, loss and thus the training of the model behaves differently
- Two standard loss functions (BCE and Dice) were used, with three new geometric-prior based being introduced (Attentive BCE, Attentive Dice, BCE/Dice Physics Weighted)

Physics-Guided Loss Functions

$$\mathcal{L}_{\mathrm{BCE}}^{\mathrm{att}} = \frac{1}{N}\sum_{i=1}^{N}(1+\alpha s_i)\left[-y_i \log p_i - (1-y_i)\log(1-p_i)\right]$$

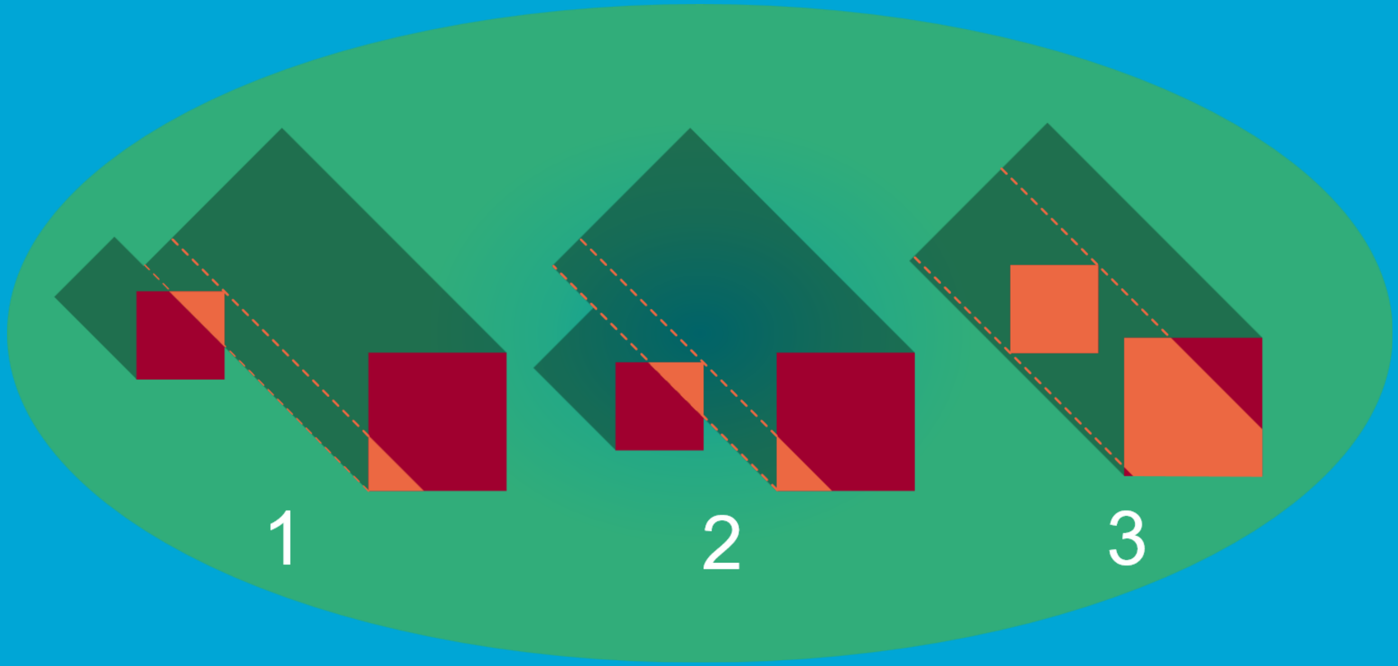
$$\mathcal{L}_{\mathrm{Dice}}^{\mathrm{att}} = 1 - \frac{2\sum_i (1+\alpha s_i)\, p_i y_i + \varepsilon}{\sum_i (1+\alpha s_i)\,(p_i + y_i) + \varepsilon}$$

$$\mathcal{L}_{\mathrm{BCE/Dice}}^{\mathrm{Phys}} = \lambda_{BCE}\cdot\mathcal{L}_{BCE} + \lambda_{Dice}\cdot\mathcal{L}_{Dice} + \lambda_{Phys}\cdot\mathcal{L}_{Phys}$$

$$\mathcal{L}_{Phys} = -\frac{1}{N}\sum_{i=1}^{N}\left[s_i \log(p_i) + (1-s_i) log(1-p_i)\right]$$

$$\mathcal{L}_{Phys} = 1 - \frac{2\sum_i p_i s_i + \varepsilon}{\sum_i (p_i + s_i) + \varepsilon}$$

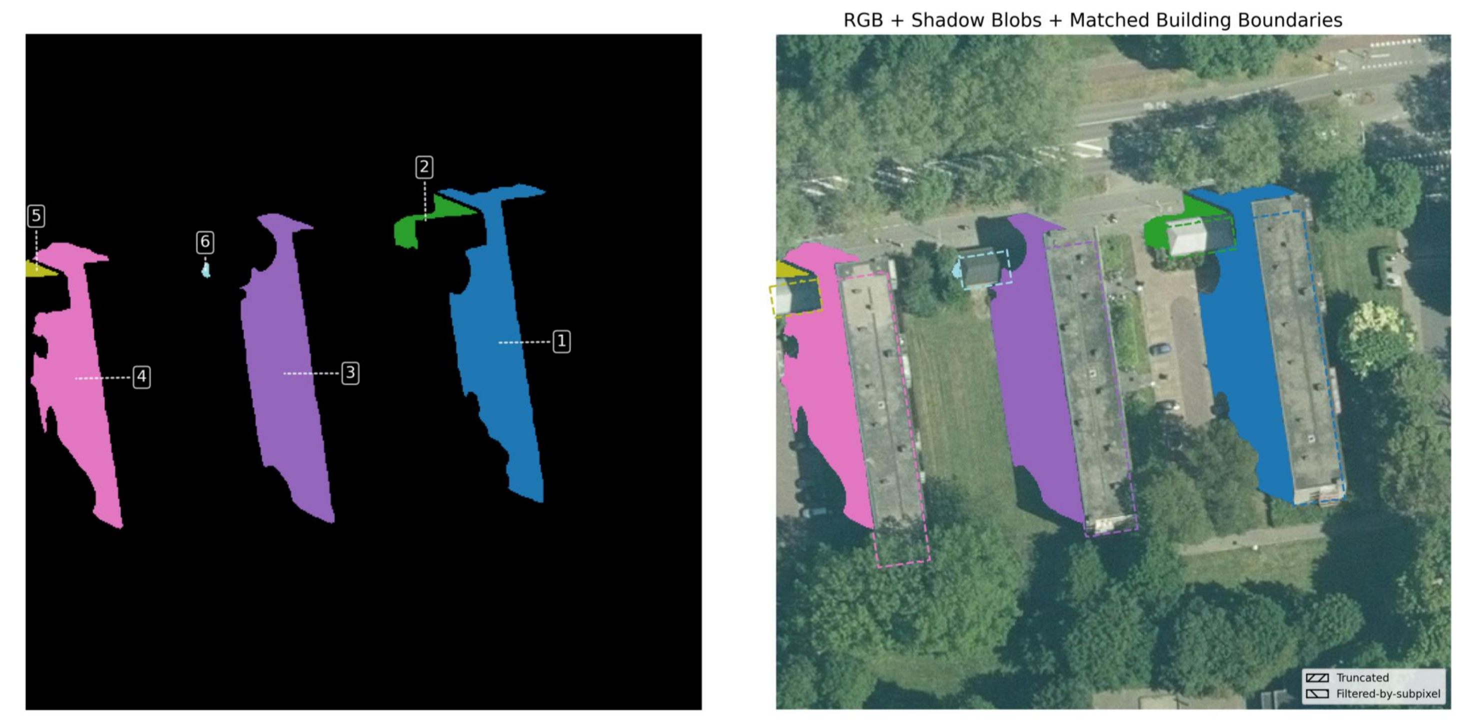
- In addition, three extra loss functions were added that implicitly use the pseudo-shadow
- For the attentive losses, inference is judged more harshly within the regions of the pseudo-shadow
- For the physics loss below, an additional error term is added to the default loss function.
- As such, attentive loss is multiplier based on pseudo-shadow, physics loss is additive

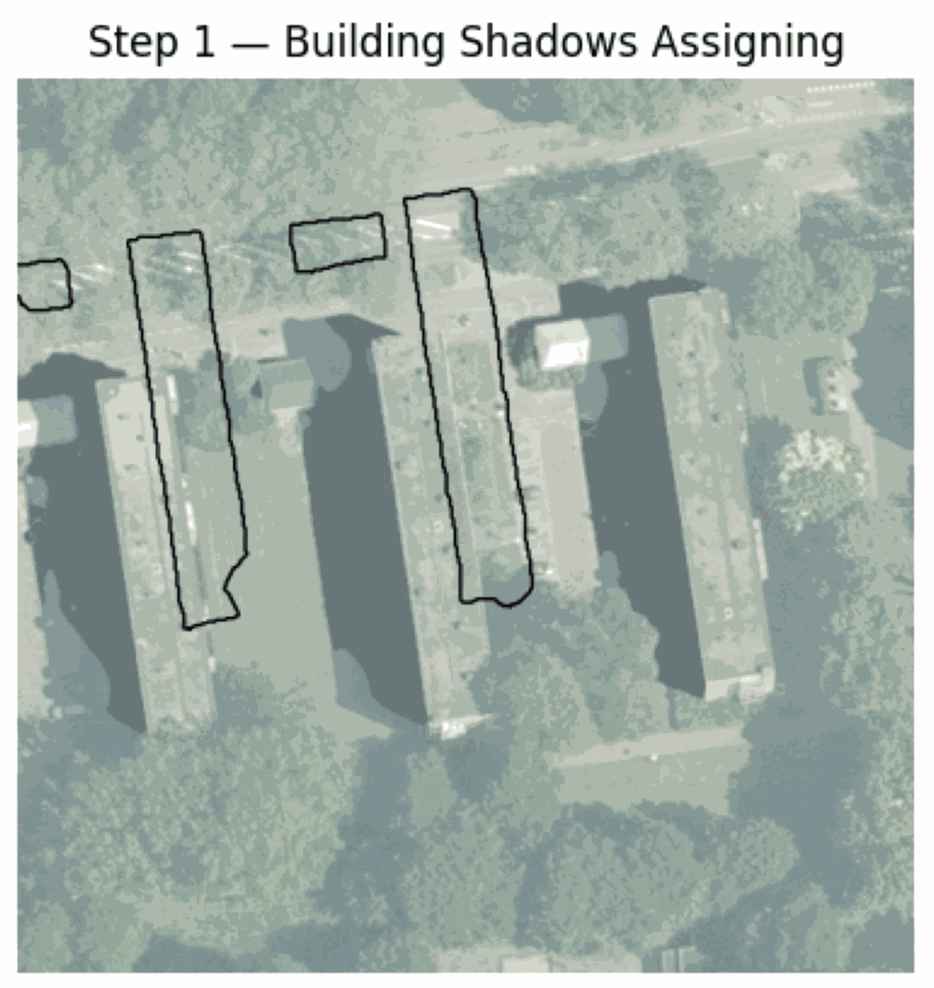


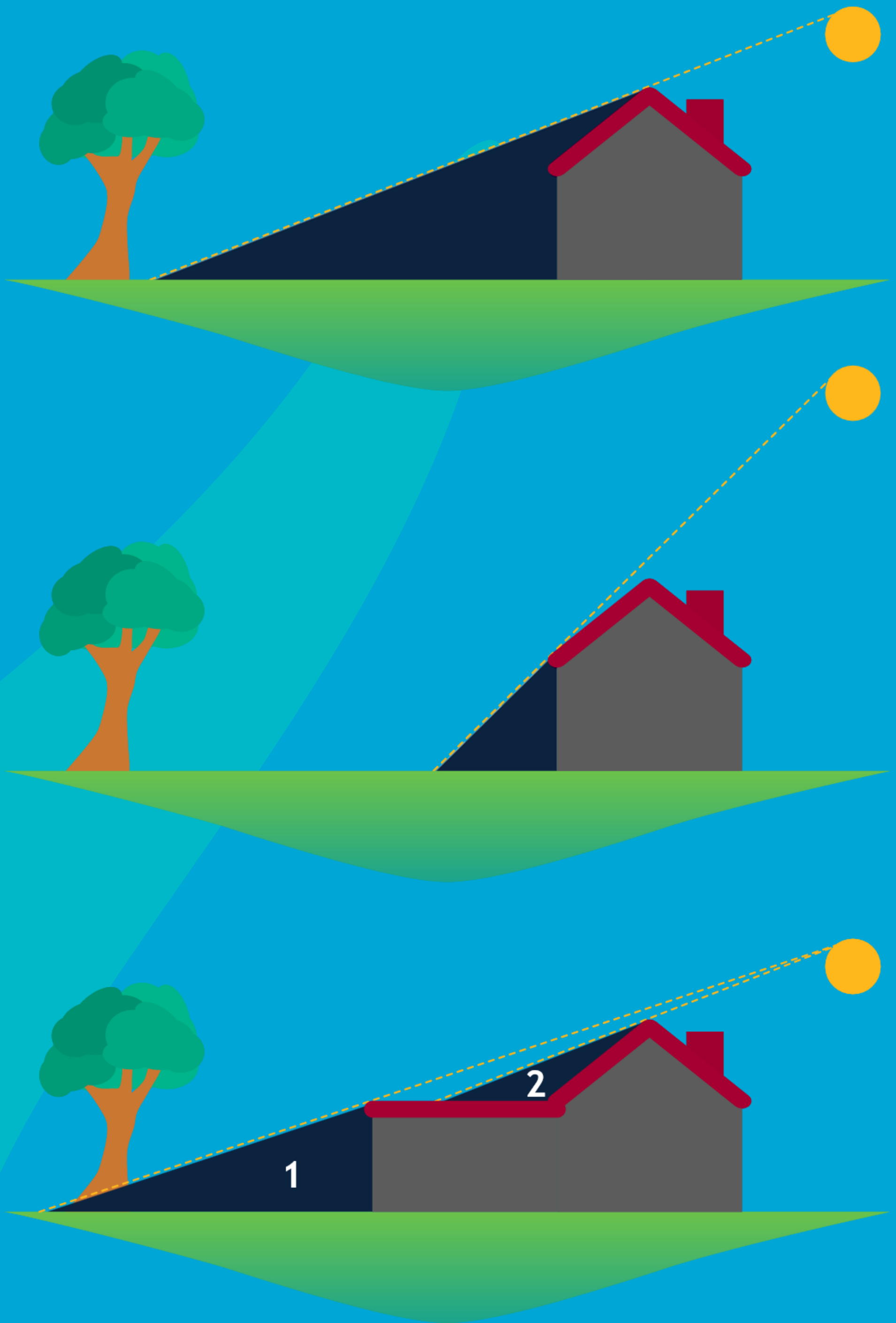
Key Concepts

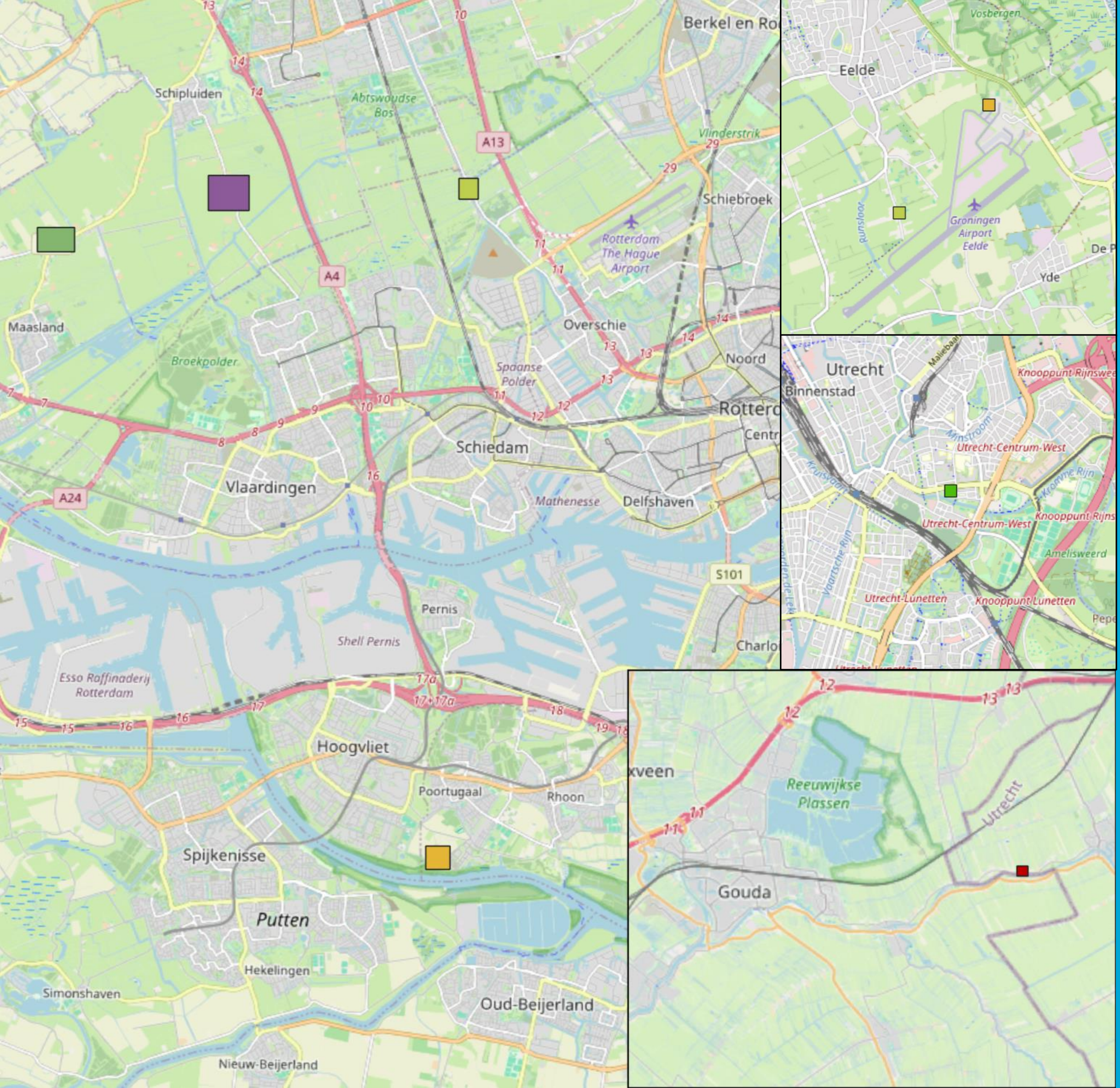
Height Estimation from Shadows

- To estimate the heights of buildings, the lengths of a given shadow need to be found
- However, shadows from different buildings bleed into one another, leading to shadow-to-building mapping issues
- In addition, ambiguity may occur in cases where buildings overlap shadows
- To solve in part, this thesis employs a similar algorithm to the pseudo-shadow smearing algorithm, but in reverse to break up and assign shadows.


RGB + Shadow Blobs + Matched Building Boundaries
1
2
3
4
5
6
Truncated
Filtered-by-subpixel


Step 1 — Building Shadows Assigning

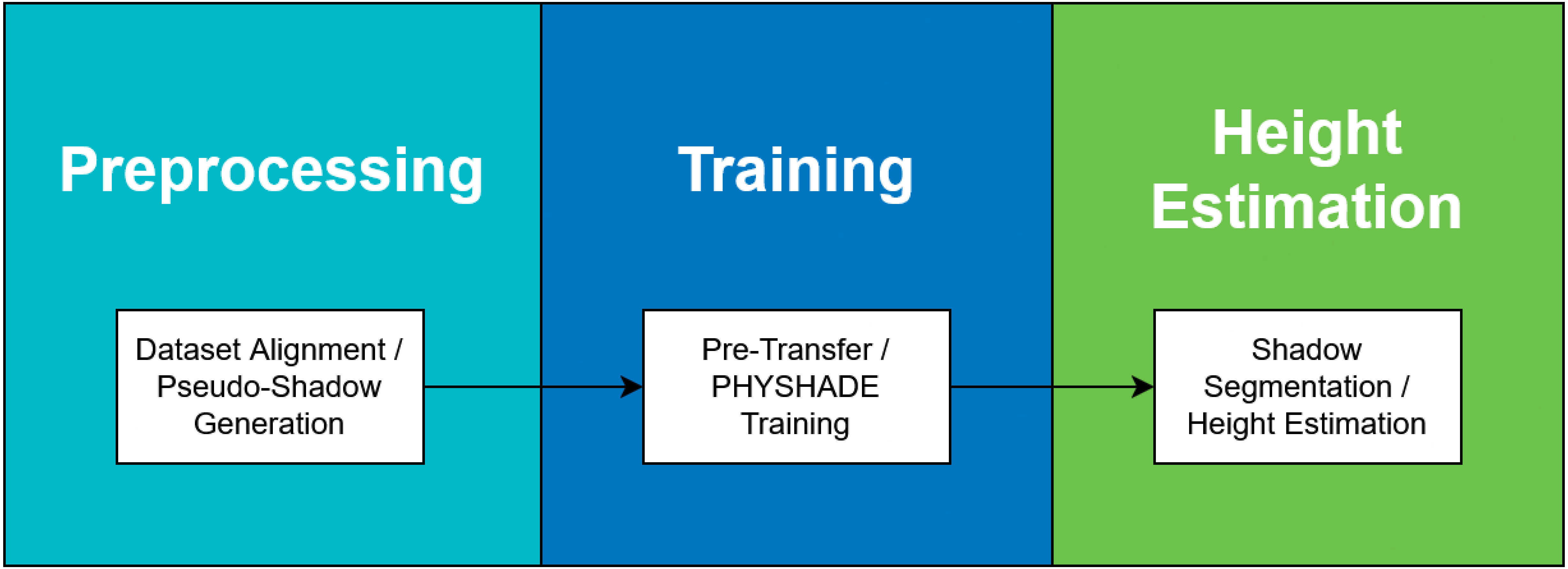

1
2



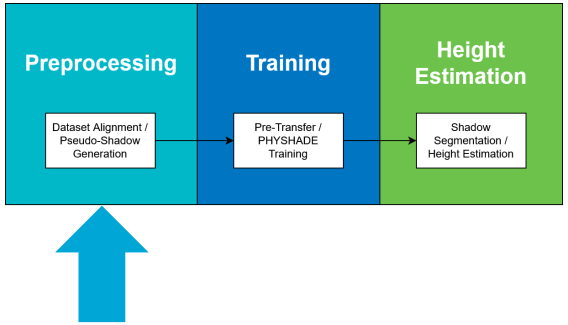
Case Study Area

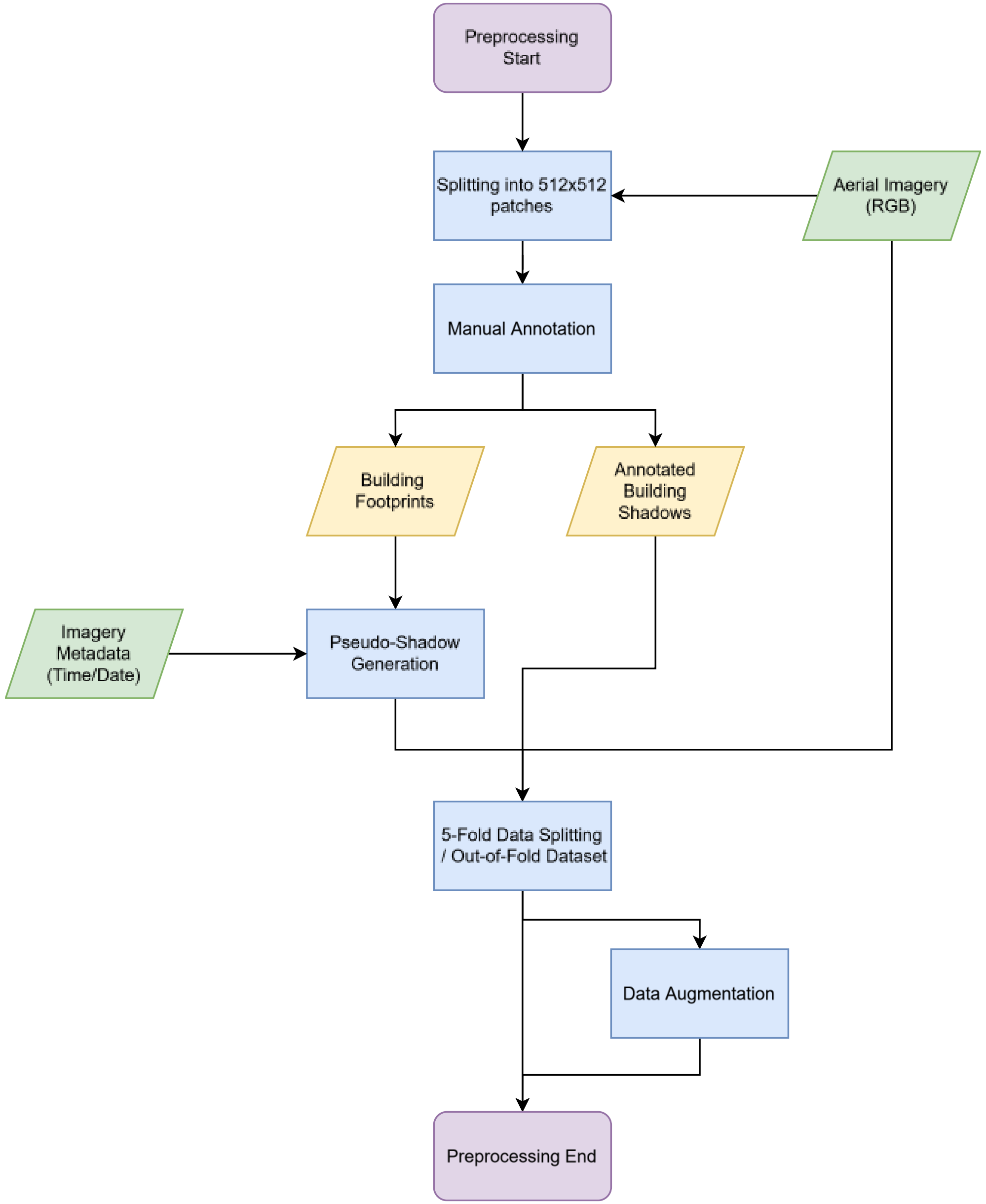
- Imagery was collected from 7 different areas within the Netherlands
- Rough selection of 206 tiles based upon study area viability
- Out of these, 41 were selected as good fits, based on:
  - At least one building with clearly visible shadow
  - Majority of shadows are relatively clearly discernible
  - Image also contains shadows not belonging to buildings
  - Some tiles should contain water
- Summer-Winter paired imagery, i.e. per location two images
- Data split up into cross-validation set, and out-of-fold dataset for full training / height estimation

Methodology at a glance

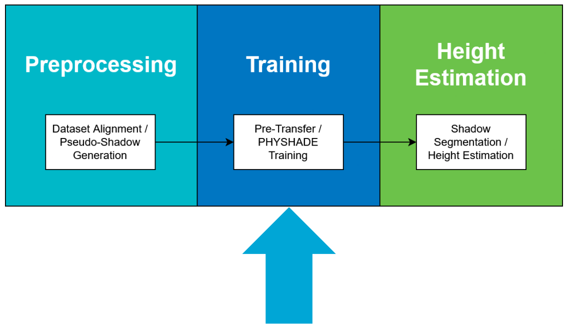


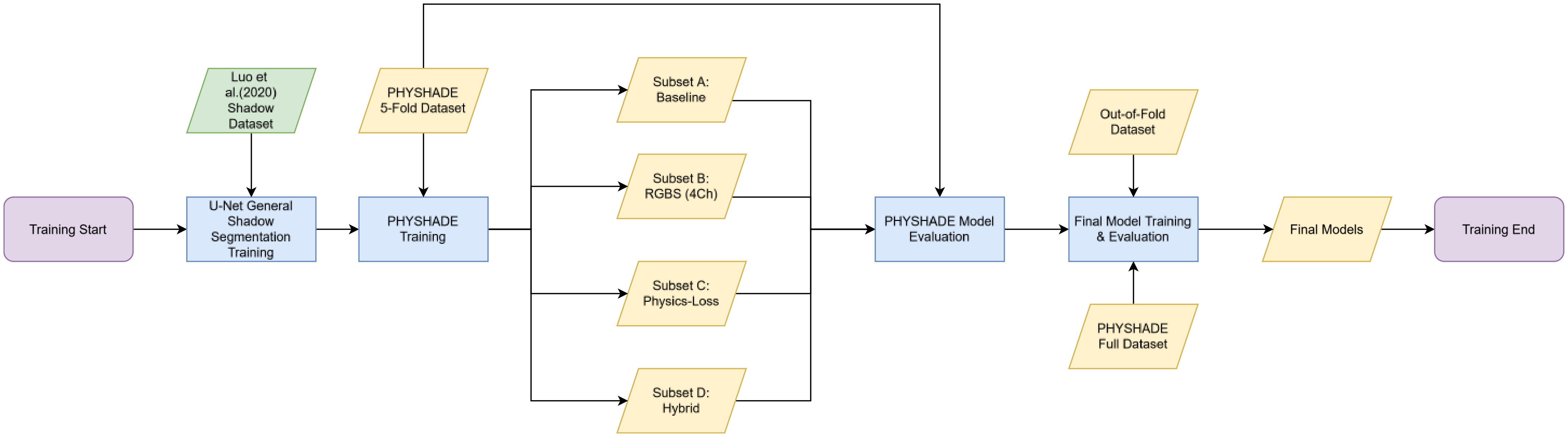
Methodology at a glance



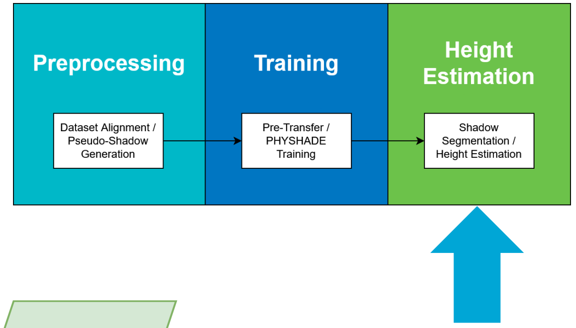


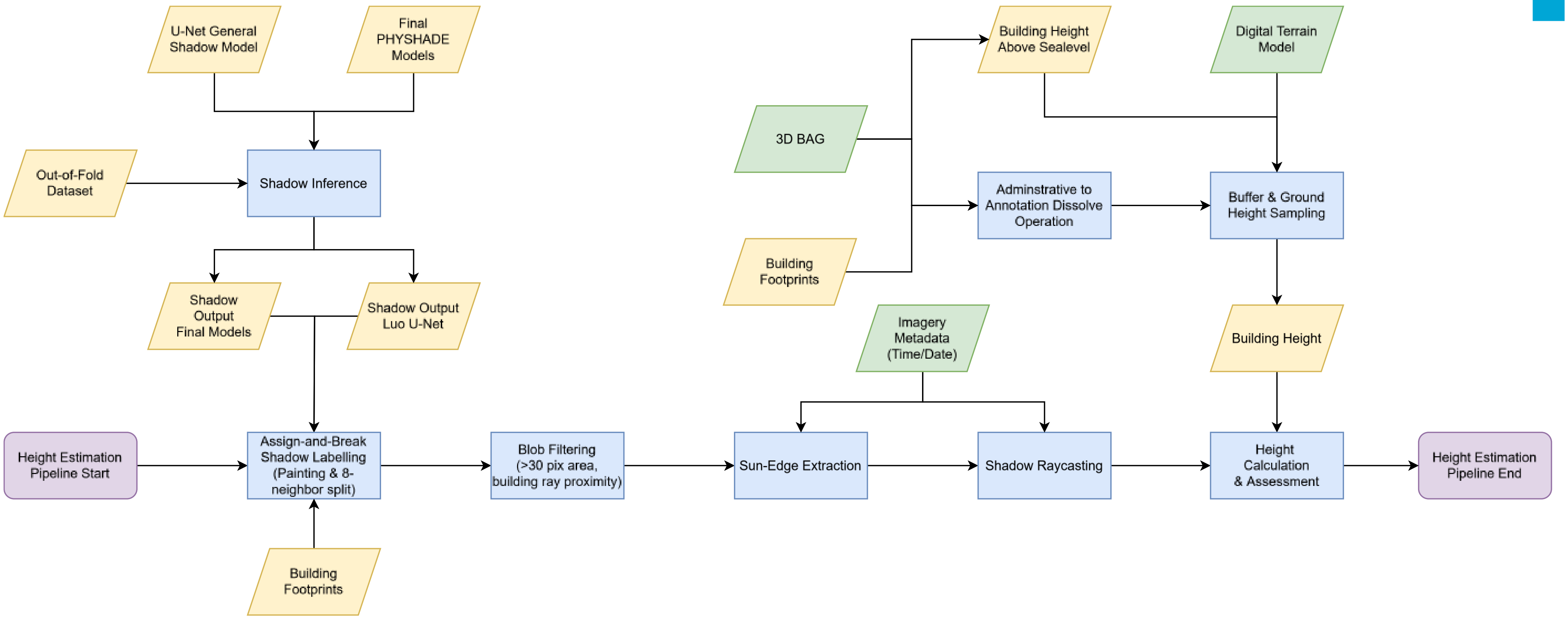
Methodology at a glance



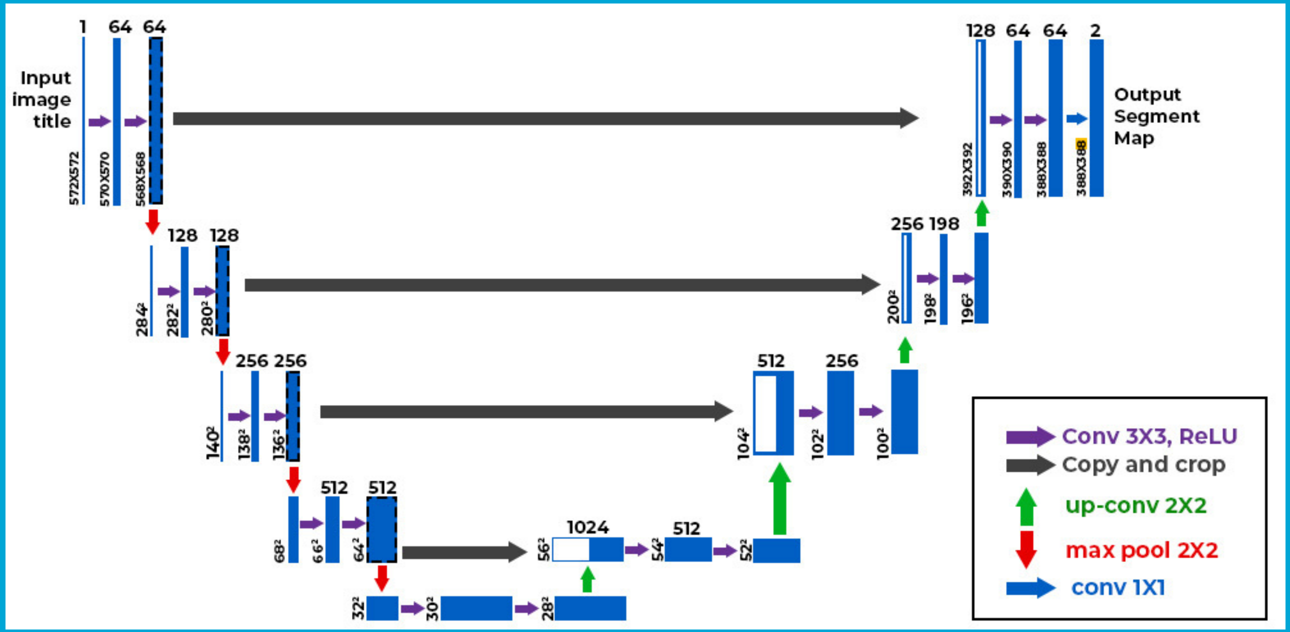


Methodology at a glance





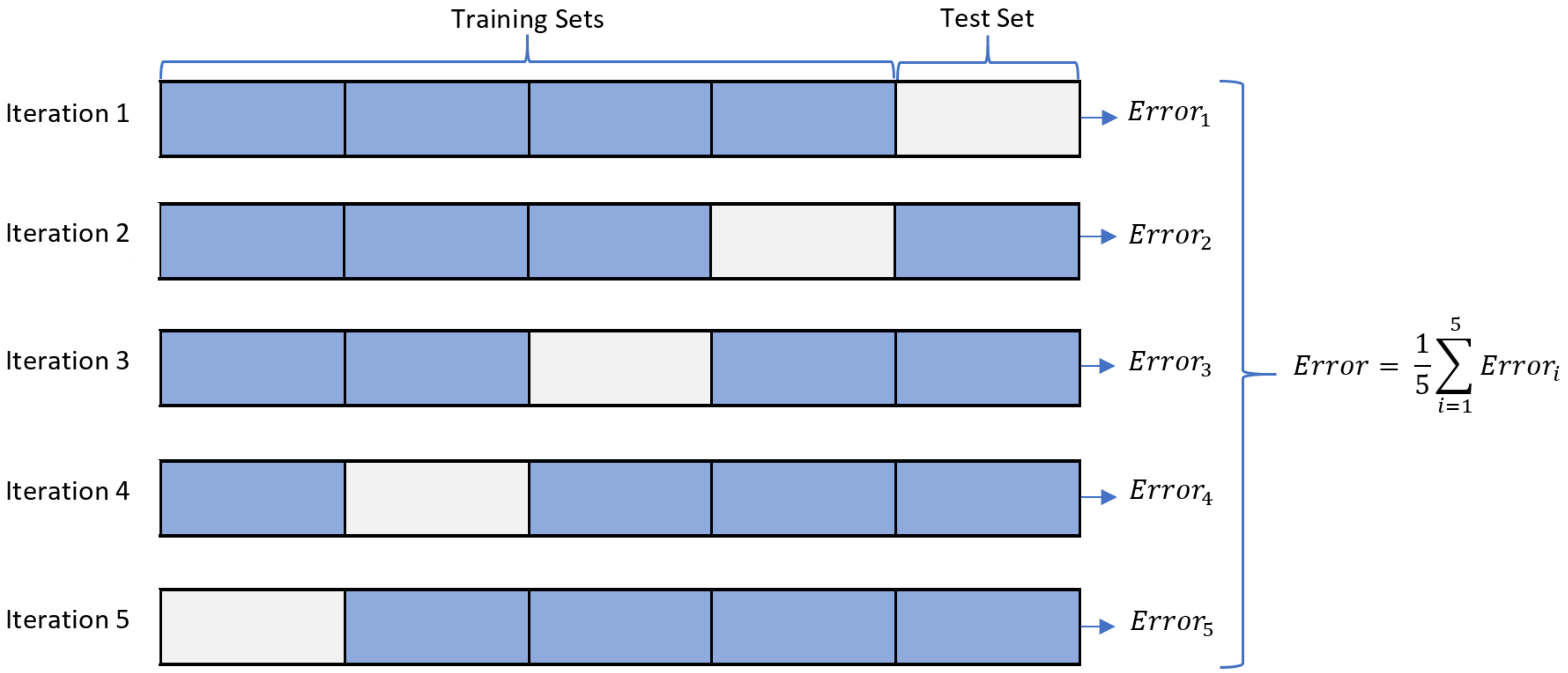
Model Training



- U-Net, a famous CNN used in biomedical image recognition, was used as a baseline model architecture.
- Since availability of data is low (35 images), transfer learning was employed in combination with 5-fold cross validation.
- In addition, each image had data augmentation performed, increasing cross-validation size to 280 images
- The U-Net model before transfer learning was trained on a dataset by Luo et al. (2020), and used Binary Cross Entropy as the loss function

Table 3: An overview of the different augmentation operations, based upon a multiplicative factor.

Multiplication	Operation
1	Original Image
2	Horizontal Flip
3	Vertical Flip
4	Rotated 90 Degrees
5	Rotated 180 Degrees
6	Rotated 270 Degrees
7	Horizontal Flip, Rotated 90 Degrees
8	Vertical Flip, Rotated 90 Degrees



From Rebecca Patro's "Cross Validation: K Fold vs Monte Carlo", 2021

https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b/

Table H.1: An overview of the trained models, with varying hyperparameters for the purpose of ablation and parameter tuning.

Experiment ID	Name	Hyperparameters	4th Channel Pseudo-Shadow Enabled
BASE BCE	BCE	-	False
BASE DICE	Dice	-	False
RGB BCE30 DICE70	BCE / Dice	BCE: 0.3, Dice: 0.7	False
RGB BCE50 DICE50	BCE / Dice	BCE: 0.5, Dice: 0.5	False
RGB BCE70 DICE30	BCE / Dice	BCE: 0.7, Dice: 0.3	False
RGBS BCE30 DICE70	BCE / Dice	BCE: 0.3, Dice: 0.7	True
RGBS BCE50 DICE50	BCE / Dice	BCE: 0.5, Dice: 0.5	True
RGBS BCE70 DICE30	BCE / Dice	BCE: 0.7, Dice: 0.3	True
PHYS ATT 1.0	Attentive BCE / Dice	Attention: 1, BCE: 0.5, Dice: 0.5	False
PHYS ATT 0.5	Attentive BCE / Dice	Attention: 0.5, BCE: 0.5, Dice: 0.5	False
PHYS ATT 0.1	Attentive BCE / Dice	Attention: 0.1, BCE: 0.5, Dice: 0.5	False
PHYS BCE 33	Physics BCE	Physics: 0.3, BCE: 0.3, Dice: 0.3	False
PHYS BCE 10	Physics BCE	Physics: 0.1, BCE: 0.45, Dice: 0.45	False
PHYS BCE 50	Physics BCE	Physics: 0.5, BCE: 0.25, Dice: 0.25	False
PHYS DICE 33	Physics Dice	Physics: 0.3, BCE:0.3, Dice: 0.3	False
PHYS DICE 10	Physics Dice	Physics: 0.1, BCE: 0.45, Dice: 0.45	False
PHYS DICE 50	Physics Dice	Physics: 0.5, BCE: 0.25, Dice: 0.25	False
HYB BCE PHYS30	Physics BCE	Physics: 0.3, BCE: 0.3, Dice: 0.3	True
HYB BCE PHYS50	Physics BCE	Physics: 0.5, BCE: 0.25, Dice: 0.25	True
HYB BCE PHYS10	Physics BCE	Physics: 0.1, BCE: 0.45, Dice: 0.45	True
HYB DICE PHYS30	Physics Dice	Physics, 0.3, BCE: 0.3, Dice: 0.3	True
HYB DICE PHYS50	Physics Dice	Physics: 0.5, BCE: 0.25,Dice: 0.25	True
HYB DICE PHYS10	Physics Dice	Physics: 0.1, BCE: 0.3, Dice: 0.3	True
HYB ATT 1.0	Attentive BCE / Dice	Physics: 1, BCE: 0.5, Dice: 0.5	True
HYB ATT 0.5	Attentive BCE / Dice	Physics: 0.5, BCE: 0.5, Dice: 0.5	True
HYB ATT 0.1	Attentive BCE / Dice	Physics: 0.1, BCE: 0.5, Dice: 0.5	True

Model Configurations

- 26 Model configurations were formulated to test the effectiveness of the addition of geometric priors in the form of pseudo-shadows
- Configuration Subset A:
  Baseline Performance (No geometric priors)
- Configuration Subset B:
  4th Channel Pseudo-Shadows added.
- Configuration Subset C:
  Physics-guided loss, where loss functions are modified based upon pseudo-shadows.
- Configuration Subset D:
  Hybrid of B and C.

Table 4: Summary table indicating the purpose of each individual ablation.

Ablation Subsets	**Purpose**
A & B	To establish the performance difference unique to the models' interpretation of the pseudo-shadows as an extra channel.
A & C	To establish the performance difference unique to the usage of physics-guided loss based on the pseudo-shadows without vision on the pseudo-shadows.
B & D	To establish the performance difference unique to the usage of various physics-guided losses based on the pseudo-shadows with vision on the pseudo-shadows.

PHYSHADE Cross-Validation Results

Table 7: An overview of the averaged fold statistics for experimental subset A.

Experiment	Mean Dice	Std. Dice	Mean Precision	Std. Precision	Mean Recall	Std. Recall	Mean Loss	Std Loss	Mean Epochs Run	Dice CI95 Lower	Dice CI95 Upper
BASE BCE	0.4978	0.0674	0.5447	0.1145	0.3778	0.0335	0.0418	0.0136	39.4000	0.4387	0.5569
BASE DICE	0.5309	0.0331	0.5885	0.0944	0.4301	0.0610	0.4915	0.0566	46.8000	0.5019	0.5598

- Experimental Subset A: Baseline models
- Like expected, models trained by baseline to only recognize building shadows do not perform well.
- No statistically significant difference between Dice-based loss and BCE-based loss in this case

PHYSHADE Cross-Validation Results

Table 8: An overview of the averaged fold statistics for Experimental Subset B, comparing the RGB vs the RGBS models.

Experiment ID	Mean Dice	Std. Dice	Mean Precision	Std. Precision	Mean Recall	Std. Recall	Mean Loss	Std. Loss	Dice CI95 Lower	Dice CI95 Upper
RGBS BCE30 DICE70	0.8475	0.0259	0.8435	0.0552	0.8239	0.0283	0.1245	0.0126	0.8248	0.8702
RGBS BCE50 DICE50	0.8455	0.0242	0.8488	0.0380	0.8172	0.0321	0.0971	0.0095	0.8242	0.8667
RGBS BCE70 DICE30	0.8487	0.0277	0.8470	0.0403	0.8118	0.0391	0.0672	0.0097	0.8244	0.8730
RGB BCE30 DICE70	0.5225	0.0313	0.5662	0.1033	0.4295	0.0754	0.3837	0.0323	0.4951	0.5499
RGB BCE50 DICE50	0.5208	0.0271	0.5578	0.0875	0.4186	0.0744	0.3007	0.0301	0.4971	0.5446
RGB BCE70 DICE30	0.5249	0.0230	0.5550	0.0955	0.4070	0.0349	0.2011	0.0219	0.5047	0.5451

Table 9: An overview of the Experimental Subset B ablation using paired t-testing, ran between RGB versus RGBS channels.

Experiment	RGBS Mean Dice	RGB Mean Dice	Delta Mean Dice	Delta Std. Dice	Delta Mean Precision	Delta Std. Precision	Delta Mean Recall	Delta Std. Recall	Delta Dice CI95 Lower	Delta Dice CI95 Upper	p-value (Dice)
BCE30 DICE70	0.8475	0.5225	0.3250	0.0264	0.2772	0.0556	0.3944	0.0709	0.2991	0.3509	0.0000
BCE50 DICE50	0.8455	0.5208	0.3247	0.0242	0.2910	0.0579	0.3986	0.0693	0.3009	0.3484	0.0000
BCE70 DICE30	0.8487	0.5249	0.3239	0.0280	0.2920	0.0649	0.4048	0.0415	0.2964	0.3513	0.0000

- Experimental Subset B: 4th Channel Pseudo-Shadows Added
- Addition of geometric priors in the form of Pseudo-Shadows as a fourth channel significantly increase segmentation performance by around 0.32.
- The difference between Dice score means of the RGB vs RGBS models is tested as statistically significant

PHYSHADE Cross-Validation Results

Table 10: An overview of the ablation ran using Experimental Subset C.

Physics Config	Mean Dice	Std. Dice	Mean Precision	Std. Precision	Mean Recall	Std. Recall	Base Mean Dice	Delta Mean Dice	Delta Std. Dice	Delta Dice CI95 Lower	Delta Dice CI95 Upper	p-value	Significant?
ATT 0.1	0.5264	0.0177	0.5436	0.0783	0.4153	0.0591	0.5208	0.0055	0.0096	-0.0038	0.0149	0.3105	No
ATT 0.5	0.5394	0.0268	0.5465	0.1029	0.4119	0.0628	0.5208	0.0186	0.0156	0.0033	0.0338	0.0752	No
ATT 1.0	0.5325	0.0214	0.5578	0.0812	0.4307	0.0564	0.5208	0.0116	0.0135	-0.0016	0.0249	0.1597	No
BCE 10	0.5306	0.0341	0.5754	0.0965	0.4346	0.0365	0.4978	0.0328	0.0368	-0.0033	0.0689	0.1497	No
BCE 33	0.5234	0.0577	0.5543	0.1317	0.4161	0.0397	0.4978	0.0256	0.0203	0.0058	0.0455	0.0647	No
BCE 50	0.5108	0.0774	0.5076	0.1378	0.4384	0.0359	0.4978	0.0130	0.0136	-0.0003	0.0264	0.1275	No
DICE 10	0.5327	0.0247	0.5548	0.0914	0.4149	0.0517	0.5309	0.0019	0.0297	-0.0272	0.0310	0.9047	No
DICE 33	0.5267	0.0292	0.5479	0.0875	0.4330	0.0480	0.5309	-0.0041	0.0346	-0.0381	0.0298	0.8229	No
DICE 50	0.5396	0.0413	0.5108	0.0832	0.4413	0.0343	0.5309	0.0087	0.0429	-0.0333	0.0508	0.7052	No

- Experimental Subset C: Physics-Guided Loss
- Like expected, physics-guided loss without being able to “see” the pseudo-shadows does not improve performance
- Physics term causes more erratic behaviour as signified in increases in Std. Dice across the board. However, none tested as significantly different

PHYSHADE Cross-Validation Results

Table 11: An overview of the ablation ran using Experimental Subset D

Physics Config	Mean Dice	Std. Dice	Mean Precision	Std. Precision	Mean Recall	Std. Recall	Base Mean Dice	Delta Mean Dice	Delta Std. Dice	Delta Dice CI95 Lower	Delta Dice CI95 Upper	p-value	Significant?
BCE PHYS10	0.8523	0.0258	0.8498	0.0344	0.8320	0.0222	0.8455	0.0068	0.0031	0.0038	0.0099	0.0119	Yes
BCE PHYS30	0.8349	0.0270	0.8036	0.0402	0.8453	0.0253	0.8455	-0.0105	0.0075	-0.0179	-0.0032	0.0475	Yes
BCE PHYS50	0.8077	0.0327	0.7237	0.0599	0.8669	0.0288	0.8455	-0.0377	0.0097	-0.0472	-0.0282	0.0015	Yes
DICE PHYS10	0.8443	0.0266	0.8329	0.0494	0.8317	0.0277	0.8455	-0.0012	0.0029	-0.0040	0.0016	0.4546	No
DICE PHYS30	0.8447	0.0286	0.8139	0.0544	0.8540	0.0310	0.8455	-0.0008	0.0041	-0.0048	0.0032	0.7065	No
DICE PHYS50	0.8182	0.0267	0.6569	0.1132	0.8944	0.0185	0.8455	-0.0273	0.0087	-0.0358	-0.0187	0.0033	Yes
ATT 0.1	0.8472	0.0235	0.8433	0.0395	0.8205	0.0245	0.8455	0.0017	0.0019	-0.0002	0.0035	0.1476	No
ATT 0.5	0.8467	0.0277	0.8489	0.0378	0.8235	0.0333	0.8455	0.0012	0.0034	-0.0021	0.0045	0.5237	No
ATT 1.0	0.8417	0.0243	0.8403	0.0387	0.8189	0.0273	0.8455	-0.0038	0.0033	-0.0071	-0.0005	0.0847	No

- Experimental Subset D: Hybrid Models
- Combining the RGBS configuration together with physics-guided loss in the majority of configurations led to no significant differences
- Of the four models that had significant changes in score, three led to lower scores, whereas one led to the highest score of all models
- Exploration of sub 0.1 hyperparameters can be interesting following trends

PHYSHADE Segmentation Summary

Table 13: Table of all experimental configurations trained, averaged and sorted by mean Dice score.

Experiment ID	Mean Dice	Std. Dice	Mean Precision	Std. Precision	Mean Recall	Std. Recall	Mean Loss	Std. Loss	Mean Epochs Run
HYB BCE PHYS10	0.8523	0.0258	0.8498	0.0344	0.8320	0.0222	0.1402	0.0214	58.2
RGBS BCE70 DICE30	0.8487	0.0277	0.8470	0.0403	0.8118	0.0391	0.0672	0.0097	53.2
RGBS BCE30 DICE70	0.8475	0.0259	0.8435	0.0552	0.8239	0.0283	0.1245	0.0126	58.6
HYB ATT 0.1	0.8472	0.0235	0.8433	0.0395	0.8205	0.0245	0.0975	0.0097	50.2
HYB ATT 0.5	0.8467	0.0277	0.8489	0.0378	0.8235	0.0333	0.0963	0.0127	55.2
RGBS BCE50 DICE50	0.8455	0.0242	0.8488	0.0380	0.8172	0.0321	0.0971	0.0095	54.2
HYB DICE PHYS30	0.8447	0.0286	0.8139	0.0544	0.8540	0.0310	0.2614	0.0241	58.8
HYB DICE PHYS10	0.8443	0.0266	0.8329	0.0494	0.8317	0.0277	0.1573	0.0147	55.2
HYB ATT 1.0	0.8417	0.0243	0.8403	0.0387	0.8189	0.0273	0.1018	0.0125	52.6
HYB BCE PHYS30	0.8349	0.0270	0.8036	0.0402	0.8453	0.0253	0.1947	0.0301	55.4
HYB DICE PHYS50	0.8182	0.0267	0.6569	0.1132	0.8944	0.0185	0.3782	0.0347	46.8
HYB BCE PHYS50	0.8077	0.0327	0.7237	0.0599	0.8669	0.0288	0.2433	0.0390	57.8
PHYS DICE 50	0.5396	0.0413	0.5108	0.0832	0.4413	0.0343	0.5432	0.0231	39.2
PHYS ATT 0.5	0.5394	0.0268	0.5465	0.1029	0.4119	0.0628	0.2841	0.0321	37.8
PHYS DICE 10	0.5327	0.0247	0.5548	0.0914	0.4149	0.0517	0.3451	0.0263	42.0
PHYS ATT 1.0	0.5325	0.0214	0.5578	0.0812	0.4307	0.0564	0.2810	0.0295	36.2
BASE DICE	0.5309	0.0331	0.5885	0.0944	0.4301	0.0610	0.4915	0.0566	46.8
PHYS BCE 10	0.5306	0.0341	0.5754	0.0965	0.4346	0.0365	0.3244	0.0244	41.4
PHYS DICE 33	0.5267	0.0292	0.5479	0.0875	0.4330	0.0480	0.4183	0.0136	42.0
PHYS ATT 0.1	0.5264	0.0177	0.5436	0.0783	0.4153	0.0591	0.2913	0.0275	38.2
RGB BCE70 DICE30	0.5249	0.0230	0.5550	0.0955	0.4070	0.0349	0.2011	0.0219	35.2
PHYS BCE 33	0.5234	0.0577	0.5543	0.1317	0.4161	0.0397	0.3242	0.0372	43.8
RGB BCE30 DICE70	0.5225	0.0313	0.5662	0.1033	0.4295	0.0754	0.3837	0.0323	38.6
RGB BCE50 DICE50	0.5208	0.0271	0.5578	0.0875	0.4186	0.0744	0.3007	0.0301	37.6
PHYS BCE 50	0.5108	0.0774	0.5076	0.1378	0.4384	0.0359	0.3570	0.0462	43.8
BASE BCE	0.4978	0.0674	0.5447	0.1145	0.3778	0.0335	0.0418	0.0136	39.4

- Subset B: The addition of Pseudo-shadows as a fourth channel increases Dice score by around 0.32, meaning that segmentation quality improves significantly.
- Subset C: The addition of physics-guided loss based on the Pseudo-Shadows by itself does not lend to meaningful increases in segmentation quality
- Subset D: The combination of fourth channel pseudo-shadows and physics-guided loss in most cases leads to decreased performance, but also creates the best performing model.

Final Training of Three Models

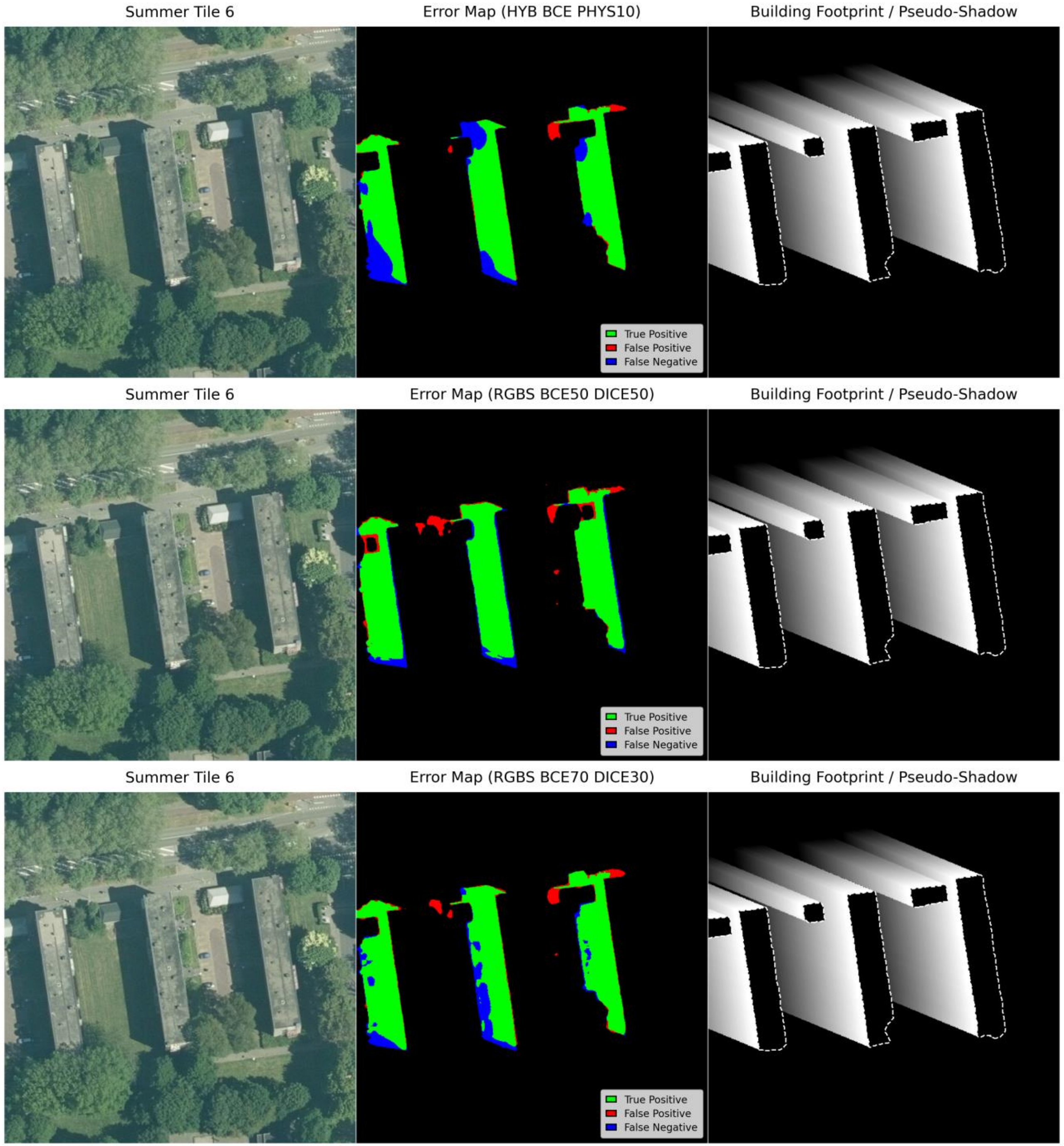
Table 15: An overview of the statistics when applying the final models to the out-of-fold dataset.

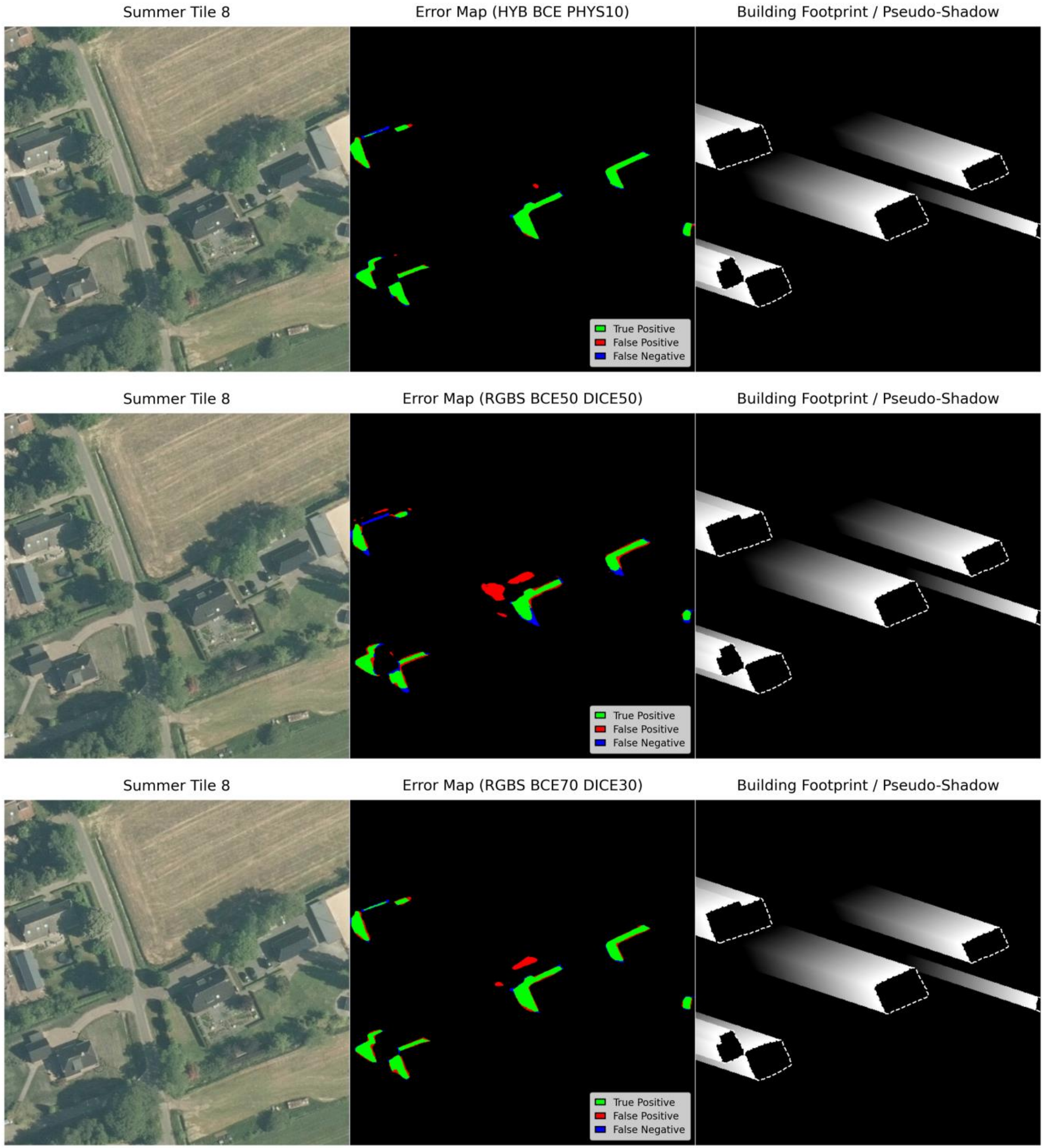
Experiment	Mean Dice	Std. Dice	Mean Precision	Std. Precision	Mean Recall	Std. Recall	Delta Dice	Delta Precision	Delta Recall
HYB BCE PHYS10	0.7636	0.2612	0.9327	0.043	0.7109	0.3196	-0.0887	0.0829	-0.1211
RGBS BCE70 DICE30	0.7213	0.3266	0.8654	0.1267	0.7192	0.3439	-0.1275	0.0184	-0.0926
RGBS BCE50 DICE50	0.7466	0.1897	0.7548	0.1951	0.7939	0.2279	-0.0989	-0.094	-0.0233
LUO UNET	0.4483	0.2049	0.3562	0.1828	0.8593	0.2443	-	-	-

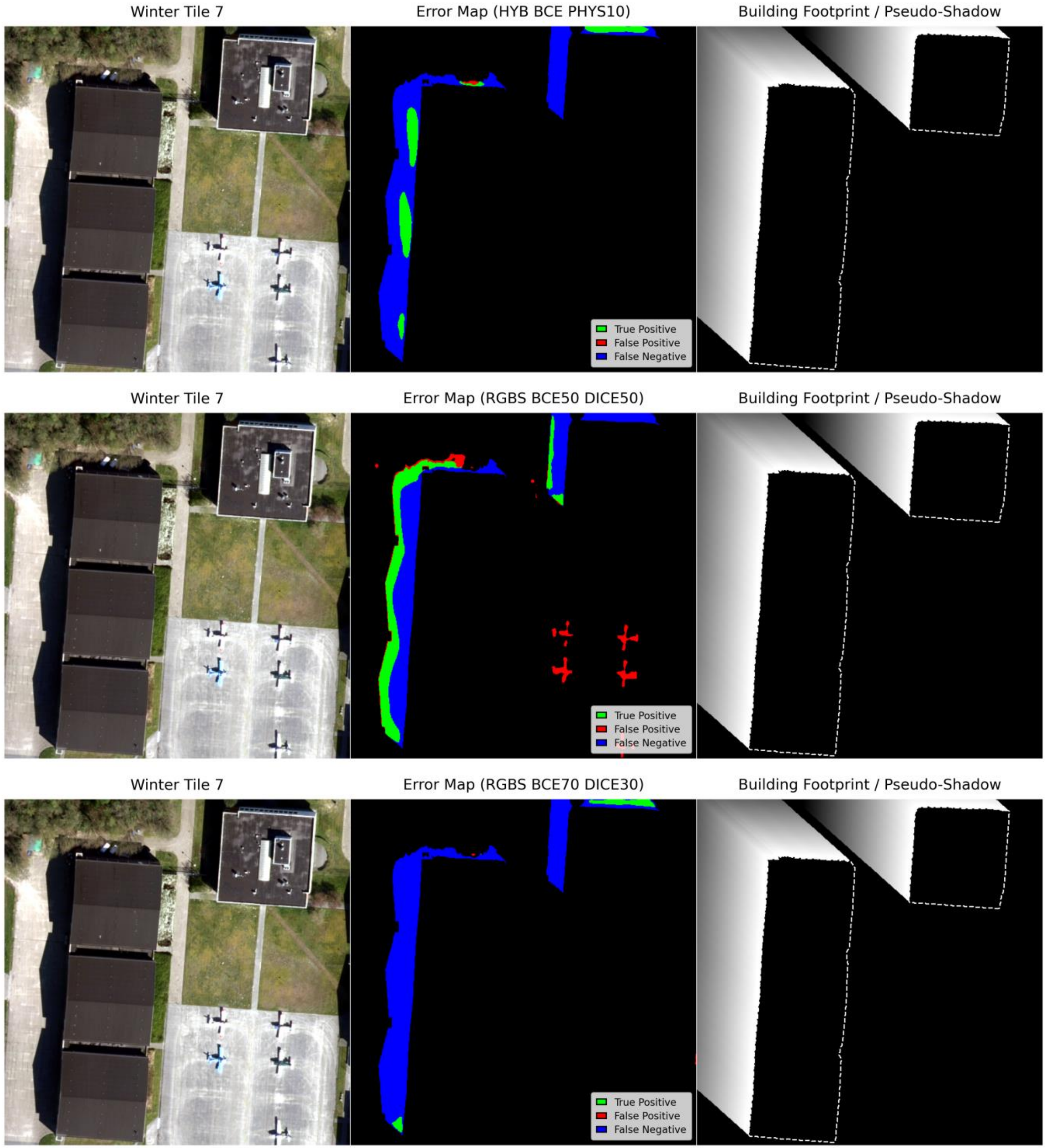
- To establish the performance of height estimation, a few different models were chosen to be trained on the full set instead of the cross-validation set:

  - HYB BCE PHYS10, as this was the model with the highest Dice scores
  - RGBS BCE50 DICE50, as this model is the ablational counterpart of the above
  - RGBS BCE70 DICE30, as this model was the second best performing model.

- These models were then assessed on the out-of-fold dataset, alongside the pre-transfer learning model LUO UNET to establish a baseline


Summer Tile 6
Error Map (HYB BCE PHYS10)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative
Summer Tile 6
Error Map (RGBS BCE50 DICE50)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative
Summer Tile 6
Error Map (RGBS BCE70 DICE30)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative

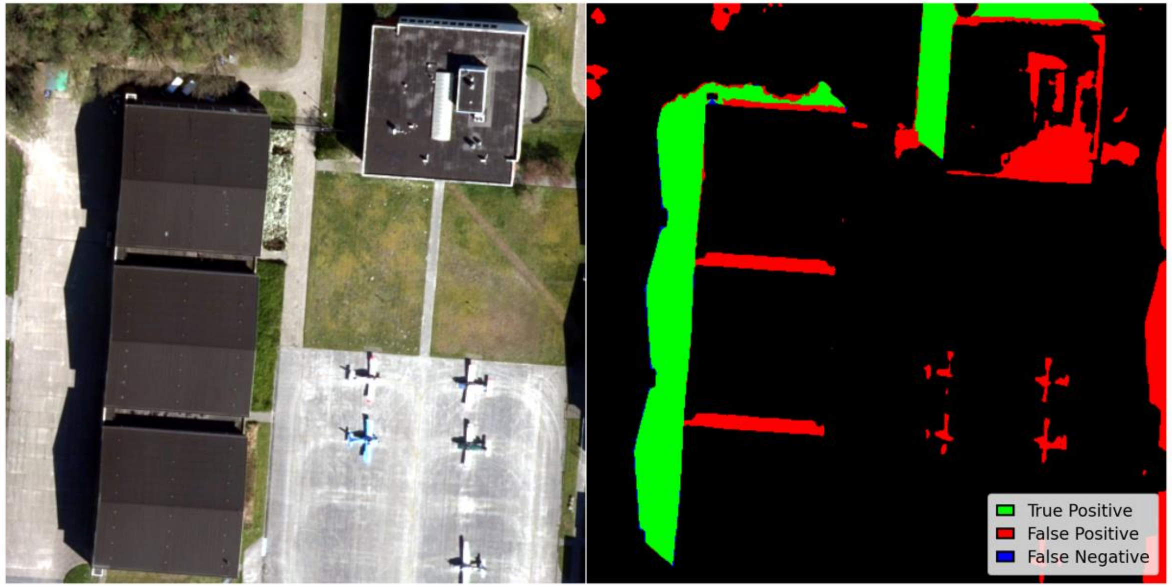

Summer Tile 8
Error Map (HYB BCE PHYS10)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative
Summer Tile 8
Error Map (RGBS BCE50 DICE50)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative
Summer Tile 8
Error Map (RGBS BCE70 DICE30)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative



Winter Tile 7
Error Map (HYB BCE PHYS10)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative
Winter Tile 7
Error Map (RGBS BCE50 DICE50)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative
Winter Tile 7
Error Map (RGBS BCE70 DICE30)
Building Footprint / Pseudo-Shadow
True Positive
False Positive
False Negative

Summer Tile 7 | Error Map (LUO UNET)

Winter Tile 7 | Error Map (LUO UNET)



Summer Tile 8 | Error Map (LUO UNET)

Table 16: An overview of the error statistics when running the raster algorithm on synthetic shadows generated by the smearing algorithm.

Ray Length Percentile	Mean Error	RMSE	Std. Error	Median Error	P90 Error	P95 Error	Max Error
100	0.0169	0.0822	0.0807	0.0	0.0	0.0	0.5
99	0.0056	0.0317	0.0313	0.0	0.0	0.0	0.25
95	0.0236	0.0964	0.0938	0.0	0.0	0.25	0.75
90	0.0446	0.2113	0.2073	0.0	0.0	0.25	2.0
75	0.2183	0.8021	0.7744	0.0	0.325	1.4125	5.625

Table 18: An overview of the average performance of height estimation per model of PHYSHADE applied to the out-of-fold dataset.

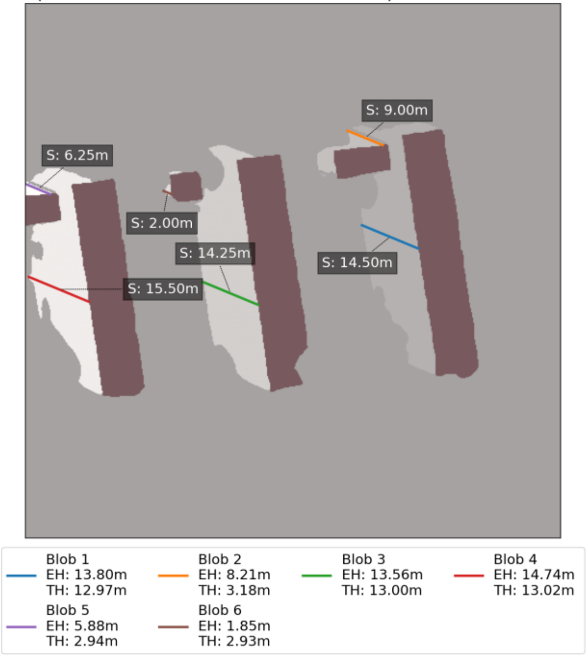
Model	Mean Error	Mean Absolute Error	RMSE	Std. Residuals	Mean Est. Height	Mean True Height	Min Est. Height	Max Est. Height
HYB BCE PHYS10	-0.8976	1.9068	2.7489	2.6477	5.9569	6.8545	1.851	14.7363
RGBS BCE50 DICE50	0.3397	1.1832	1.6948	1.6909	7.2132	6.8735	2.5958	14.5126
RGBS BCE70 DICE30	-0.1453	1.2849	1.774	1.8031	6.6888	6.834	2.0889	14.5126
Blob Weighted Average	-0.2284	1.4571	2.0716	2.0458	6.6261	6.8545	2.1848	14.5872
LUO UNET	0.2616	1.0805	1.5315	1.5377	7.1161	6.8545	1.6559	15.2264

Height Estimation Results

- For the PHYSHADE models, an RMSE was achieved of 2.07 and an MAE of 1.46 over the out-of-fold dataset.
- However, the baseline LUO UNET outperformed the PHYSHADE models that were made for segmenting especially building shadows
- Overlap between post-processing steps and purpose of PHYSHADE means that a higher recall is more important than precision, as relevant false positives are filtered out
- However, due to limited height estimation dataset size, performance related statements remain tentative

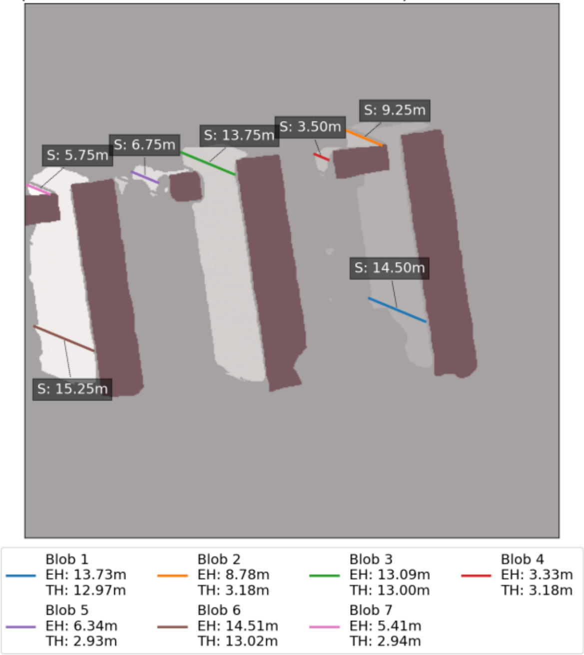
Second

Percentile (99th) Rays per Blob
(Azimuth: 112.54° for HYB BCE PHYS10) in Summer Tile 6



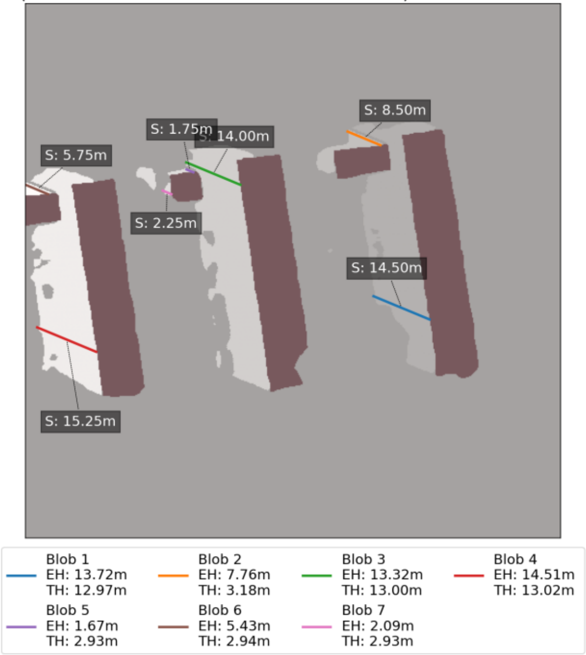
Fourth

Percentile (99th) Rays per Blob
(Azimuth: 112.54° for RGBS BCE50 DICE50) in Summer Tile 6



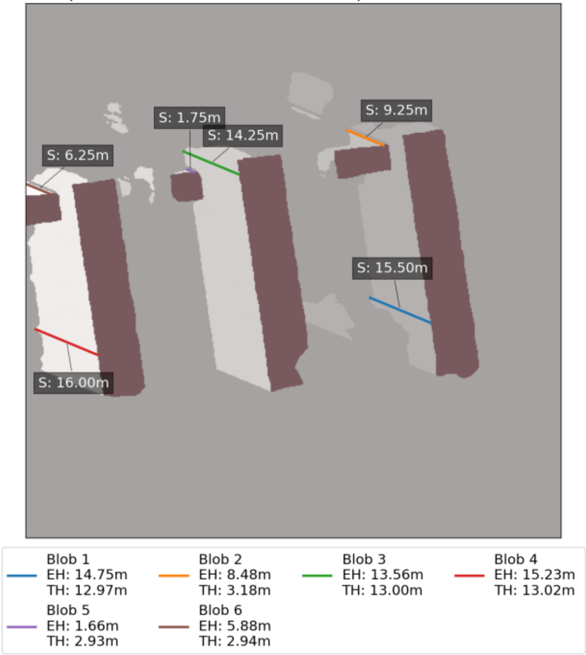
First

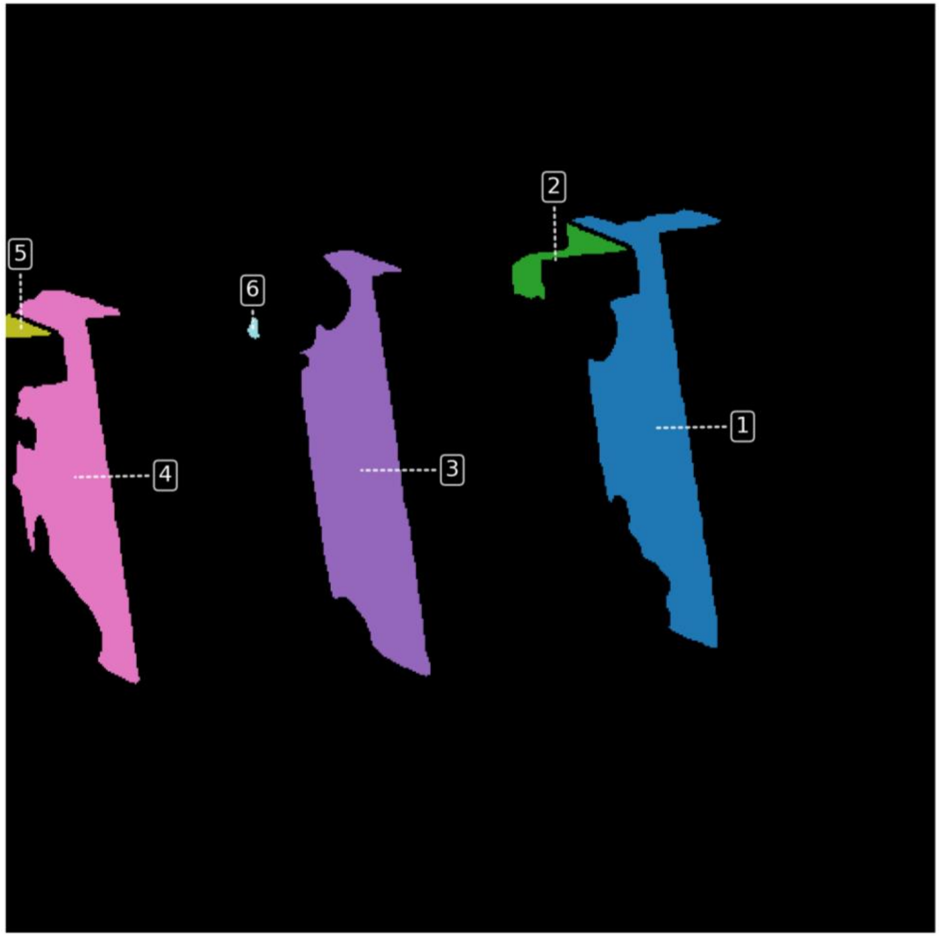
Percentile (99th) Rays per Blob
(Azimuth: 112.54° for RGBS BCE70 DICE30) in Summer Tile 6



Third

Percentile (99th) Rays per Blob
(Azimuth: 112.54° for LUO UNET) in Summer Tile 6

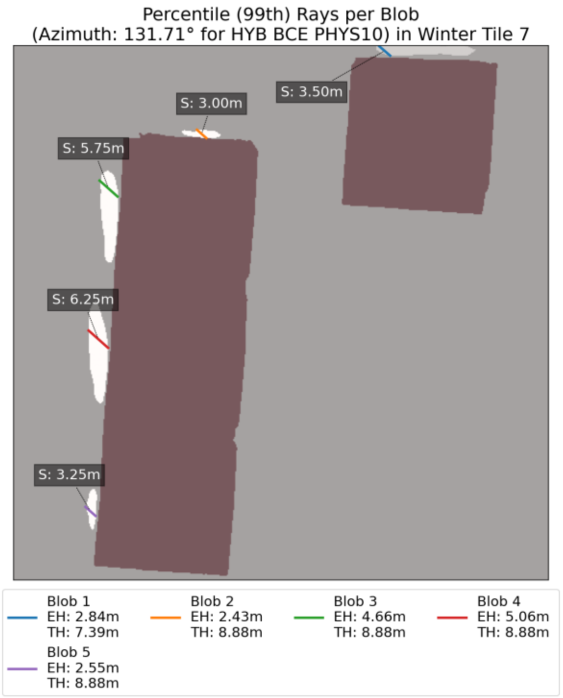




2
5
6
1
4
3

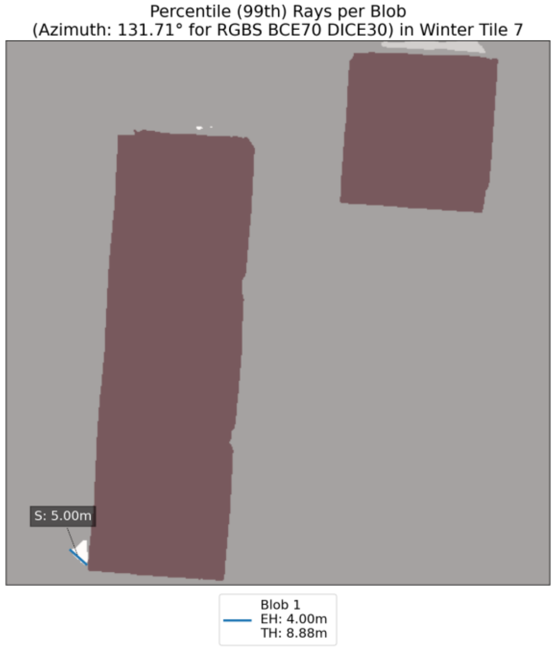
RGB + Shadow Blobs + Matched Building Boundaries
Truncated
Filtered-by-subpixel

Third



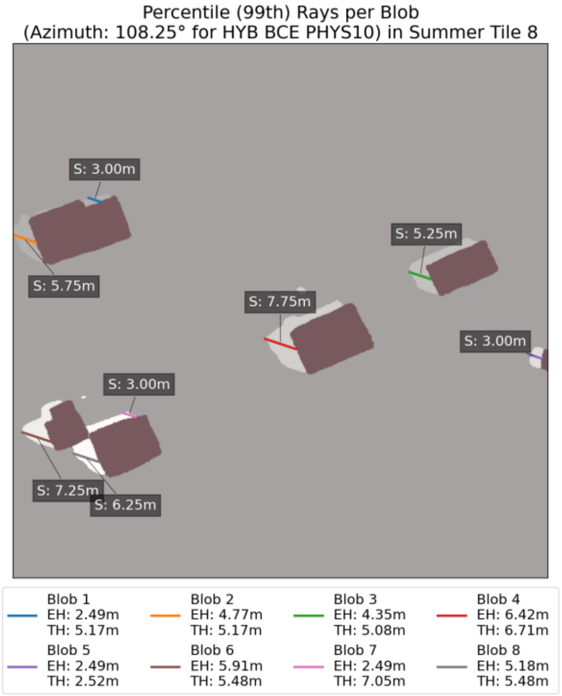
Second

Fourth

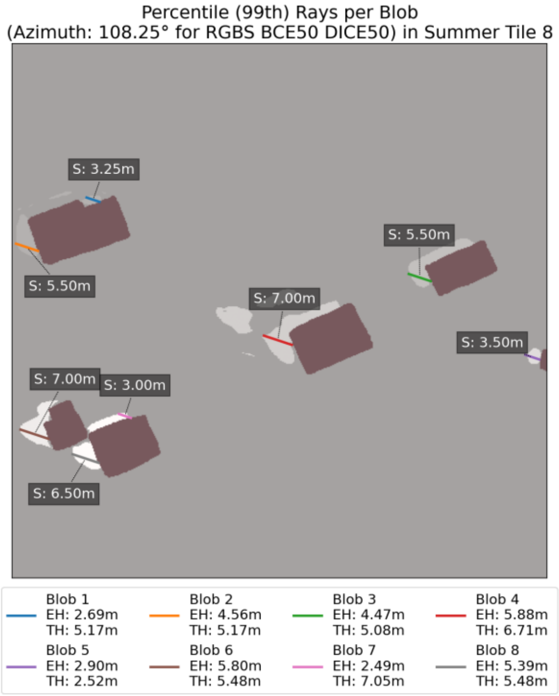


First

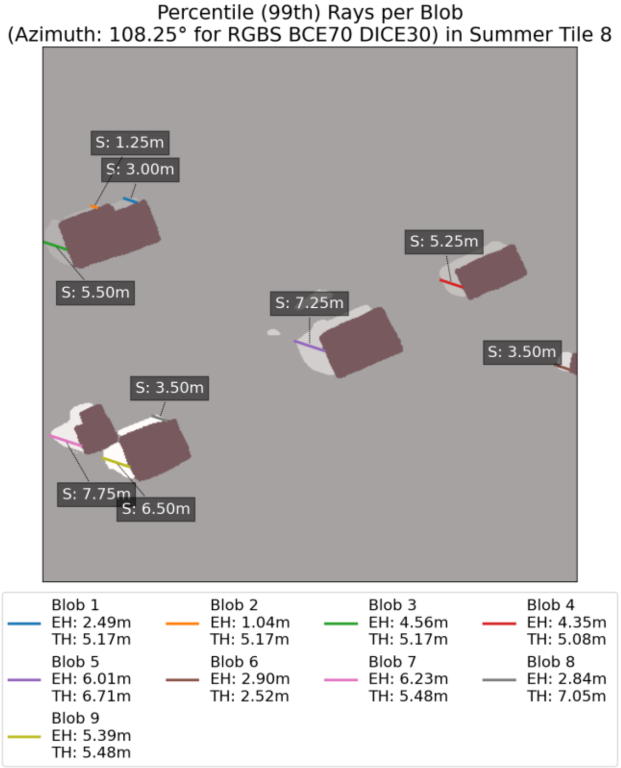
Second



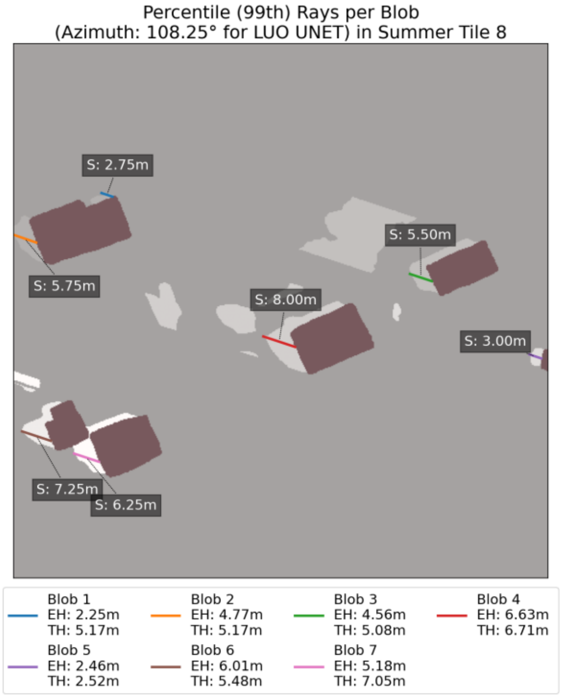
Third



Fourth



First



Discussion & Conclusion

- The addition of Pseudo-Shadows as a fourth channel allows for PHYSHADE to selectively segment building shadows with relatively good performance, getting Dice scores up to 0.847 during cross-validation
- The addition of Pseudo-Shadows in the form of physics guided loss rarely provided benefit, although one model combining the fourth channel pseudo-shadows together with the physics-guided loss led to the best performing model of all different configurations.
- For height estimation, the RMSE and MAE values lie around 2 meters and 1.46 meters respectively, meaning that they can potentially be a viable alternative provided that the accuracy required is not mission-critical.
- For the height estimation method used in this study combined with the pre-processing steps, a higher inference recall is valued more than precision for the purposes of height estimation

Limitations

- Since shadows carry ambiguity if they overlap with one another, additional error may be introduced such as when shadows are misattributed.
- This error was less visible within the out-of-fold dataset employed in this study, but would be exacerbated in more dense urban contexts.
- Dataset size for training, validation and height estimation is small, with the main training set only containing 35 images and the out-of-fold validation set only containing 6 images. As such, this limits the claims about the generalisability of the methods employed.

Future Work

- Future studies could focus on repeating the same study with a larger data sample to verify the performances established in this thesis.
- Additionally, building segmentation networks could be combined with a similar pseudo-shadow based method here to develop fully automated pipelines. For example, a multi-headed CNN that segments both buildings and building shadows at the same time.
- On top of that, raycasting could be foregone by including the height estimation into a CNN structure as well, such as in the work by Li et al. (2020). Combining this with PHYSHADE’s pseudo-shadows may help quality of segmentation and estimation there.
- Finally, additional methods distinguishing shadows from different buildings should be explored as overlapping shadows are currently causing height estimation performance degradation.

Thank you for your attention!

P5 Presentation | Lars Huizer

19-6-2025

TU Delft

- Tsai, B.-S., Huizer, L., Giampaolo, M., Monté, S., Gong, S., Garcia, G., & Agugiaro, G. (2024). Integration of GIS and CAD data to perform interactive preliminary environmental analyses at district scale. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-4/W10-2024*, 169–176.

- Soha, T., Sugár, V., & Hartmann, B. (2024). City-scale analysis of PV potential and visibility in heritage environment using GIS and LiDAR. *Energy and Buildings*, *311*, 114124. https://doi.org/10.1016/j.enbuild.2024.114124
https://doi.org/10.5194/isprs-archives-XLVIII-4-W10-2024-169-2024

- Li, Weijia & He, Conghui & Fang, Jiarui & Zheng, Juepeng & Fu, Haohuan & Yu, Le. (2019). Semantic Segmentation-Based Building Footprint Extraction Using Very High-Resolution Satellite Images and Multi-Source GIS Data. Remote Sensing. 11. 403. 10.3390/rs11040403.

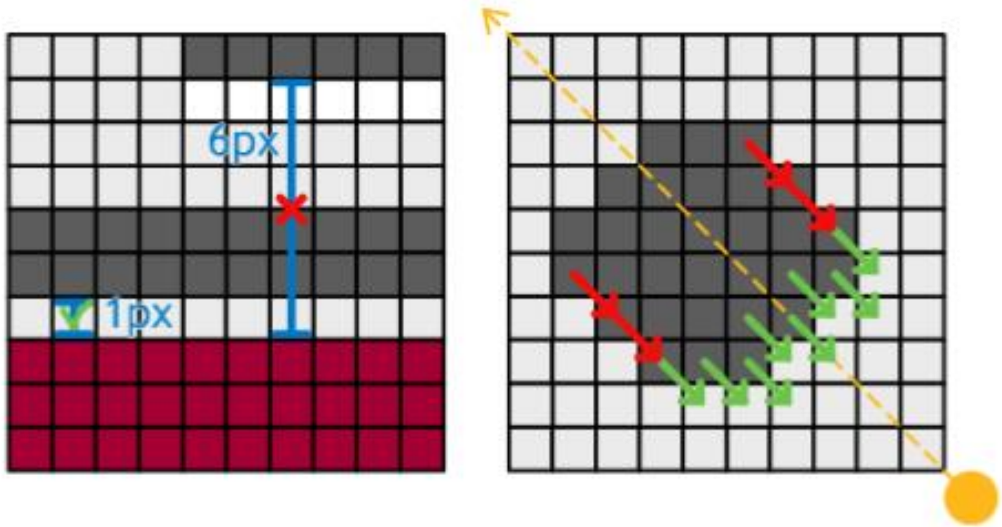


Figure 11: Image showcasing acceptance of close-by blobs and rejection of far-away blobs from buildings on the left, and the edge finding mechanism for finding the starting points for the height estimation ray-casts on the right.

Table I.1: An overview of the used hyperparameters to train the baseline model by Luo et al. (2020). These settings were also used for the the training of the derivative transfer-trained models, albeit with different loss functions.

Name	Value	Description
Batch Size	8 (Base), 16 (PHYSHADE)	Number of samples per training batch
Stride	128	The stride used to extract patches from images
Patch_size	512	The size of the patches cropped from the imagery
Seed	15	Seed for the purposes of reproducibility
Epochs	150	The maximum number of training epochs
Early Stop Patience	25	Epochs to wait without improvement to Dice loss before stopping
Optimizer	AdamW	Optimizer used for training
Learning Rate	1e-4	Initial learning rate
Weight Decay	1e-4	Modulates training regularization
Scheduler	ReduceLROnPlateau	Reduces the learning rate when validation loss plateaus
Factor	0.5	Learning rate is reduced by this factor
Patience	5	The number of epochs without improvement before educing the learning rate
Min_lr	1e-6	Lower bound on learning rate
Mode	'min'	Reduce learning rate
Loss function	BCEWithLogitsLoss	Binary Cross Entropy with logits

Algorithm J.1 Pseudocode describing the smearing algorithm responsible for generating pseudo-shadows from building footprints

function SMEAR($M, \theta, (l_{\min}, l_{\max}), T, \text{fade}$) ▷ M: binary mask, θ: azimuth, T: transform

  Compute unit vector: $(u_x, u_y) \leftarrow (\sin(\theta+\pi), \cos(\theta+\pi))$

  Extract pixel size: $(r_x, r_y) \leftarrow$ resolution from T

  Compute Step Distance: $r_{\text{eff}} \leftarrow \sqrt{(u_x r_x)^2 + (u_y r_y)^2}$

  Compute number of steps: $N \leftarrow \lceil l_{\max} / r_{\text{eff}} \rceil$

  Calculate Step Distance: $\Delta\ell \leftarrow l_{\max}/N$

  Initialize empty matrix S, same size as M

  for $iteration = 1, 2, \ldots, N$ **do**

    $\ell_i \leftarrow i \cdot \Delta\ell$

    Compute decay weight w_i based on ℓ_i:

$$w_i = \begin{cases} 1, & \text{if fade = solid} \\ 1, & \text{if } \ell_i \leq l_{\min} \\ 0, & \text{if } \ell_i \geq l_{\max} \\ 1 - \dfrac{\ell_i - l_{\min}}{l_{\max} - l_{\min}}, & \text{otherwise} \end{cases}$$

    Compute shift:

$$\Delta x = \frac{u_x \cdot \ell_i}{r_x}, \quad \Delta y = -\frac{u_y \cdot \ell_i}{r_y}$$

    Shift mask: $M_i \leftarrow \text{warp}(M, \Delta x, \Delta y)$

    Accumulate: $S \leftarrow \max(S, w_i \cdot M_i)$

  end for

  if fade is gradient **then**

    Normalize $S \leftarrow S/\max(S)$

  end if

  return S

end function

Algorithm J.2 Pseudocode describing the raymarching-based height estimation method using subpixel interpolation and percentile-based aggregation.

function EstimateHeight$(M, B, \theta, e, p, \Delta, N)$ ▷ M: shade mask, B: building mask, θ: azimuth, e: elevation, p: percentile, Δ: pixel size

Compute direction unit vector: $(u_x, u_y) \leftarrow (\sin(\theta+\pi), -\cos(\theta+\pi))$
Create binary masks: $M_b \leftarrow M > 0.5$, $B_b \leftarrow B > 0.5$
Label shadow blobs: $L_M, n_M \leftarrow \text{label}(M_b)$
Label combined connectivity: $L_C \leftarrow \text{label}(M_b \vee B_b)$
Identify connected blob IDs:

$$C \leftarrow \{\, i \mid \exists (x,y) \in M_b \text{ with } L_C[x,y] \in L_C[B_b > 0] \,\}$$

Initialize height map $H \leftarrow 0$ and blob ID map $I \leftarrow 0$
$k \leftarrow 1$ ▷ Blob counter
for all $i \in C$ **do**
  Extract blob mask: $S_i \leftarrow L_M == i$
  if $\text{area}(S_i) < 30$ **then** ▷ Helps ignore sliver artifacts from smearing
    continue
  end if
  Collect edge starting coordinates: $P \leftarrow \{(x,y) \in S_i : (x - \text{sign}(u_x), y - \text{sign}(u_y)) \notin S_i\}$
  Initialize ray length list: $R \leftarrow []$
  for all $(x_0, y_0) \in P$ **do**
    Initialize ray value list: $v \leftarrow []$
    for $j = 1, 2, \ldots, N$ **do**
      $x \leftarrow x_0 + u_x \cdot j$, $y \leftarrow y_0 + u_y \cdot j$
      if (x,y) out of bounds **then**
        break
      end if
      $v_j \leftarrow \text{interpolate}(M, (x,y))$
      if $v_j < 0.25$ **then** ▷ Min. ray length, change depending on resolution
        break
      end if
      Append v_j to v
    end for
    Append $|v|$ to R
  end for
  if $R \neq \emptyset$ **then**
    $\ell \leftarrow \text{percentile}(R, p)$ ▷ Prevents outliers common outside 8-point compass dir.
    $h \leftarrow \ell \cdot \tan(e)$
    Set $H[S_i] \leftarrow h$
    Set $I[S_i] \leftarrow k$
    $k \leftarrow k + 1$
  end if
end for
return H, I
end function

Figure K.1: A mosaic of the out-of-fold dataset used to assess the final models of PHYSHADE and to do the height estimation on. From left to right: Tile 6, Tile 7 and Tile 8.