SynRS3D: A Synthetic Dataset for Global 3D Semantic Understanding from Monocular Remote Sensing Imagery

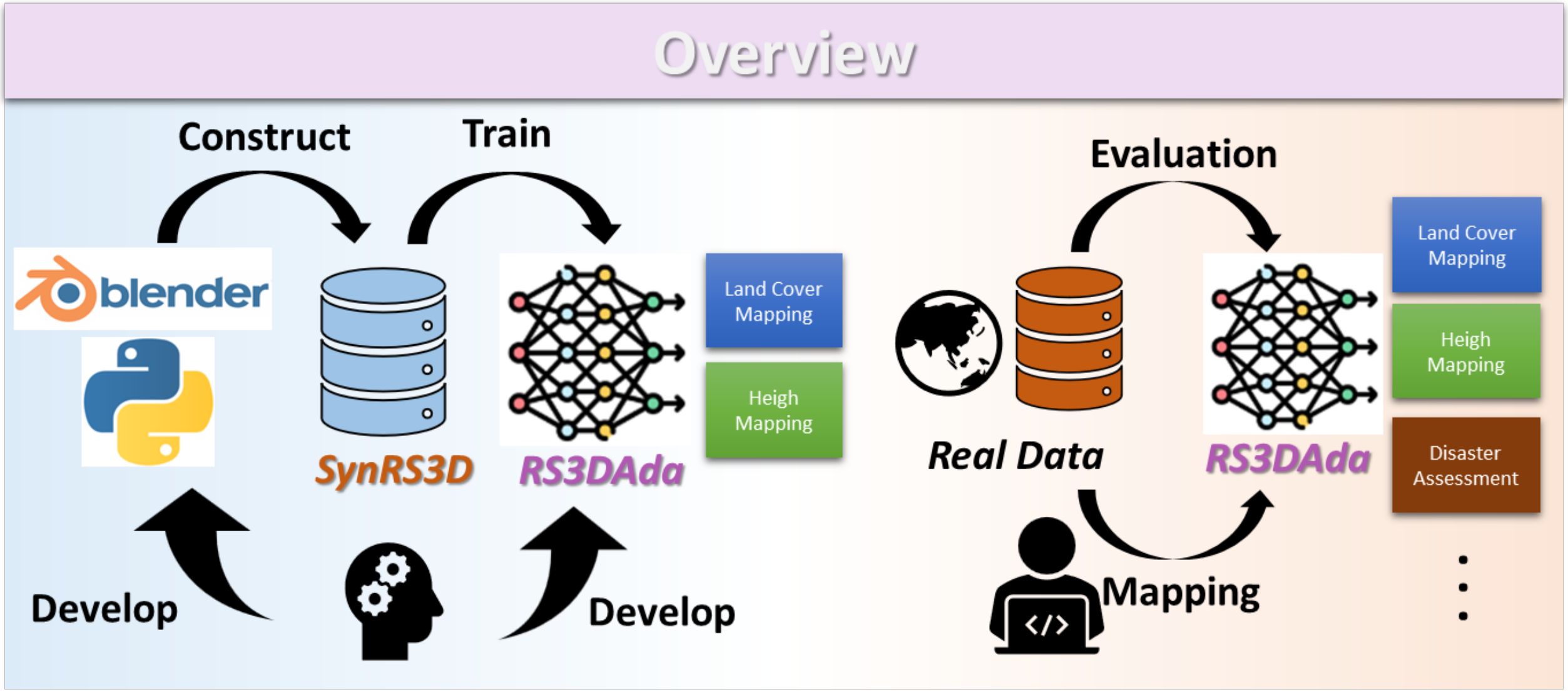
Jian Song[1,2] Hongruixuan Chen[1,2] Weihao Xuan[1,2] Junshi Xia[2] Naoto Yokoya[1,2]

[1]The University of Tokyo, Japan
[2]RIKEN, Japan

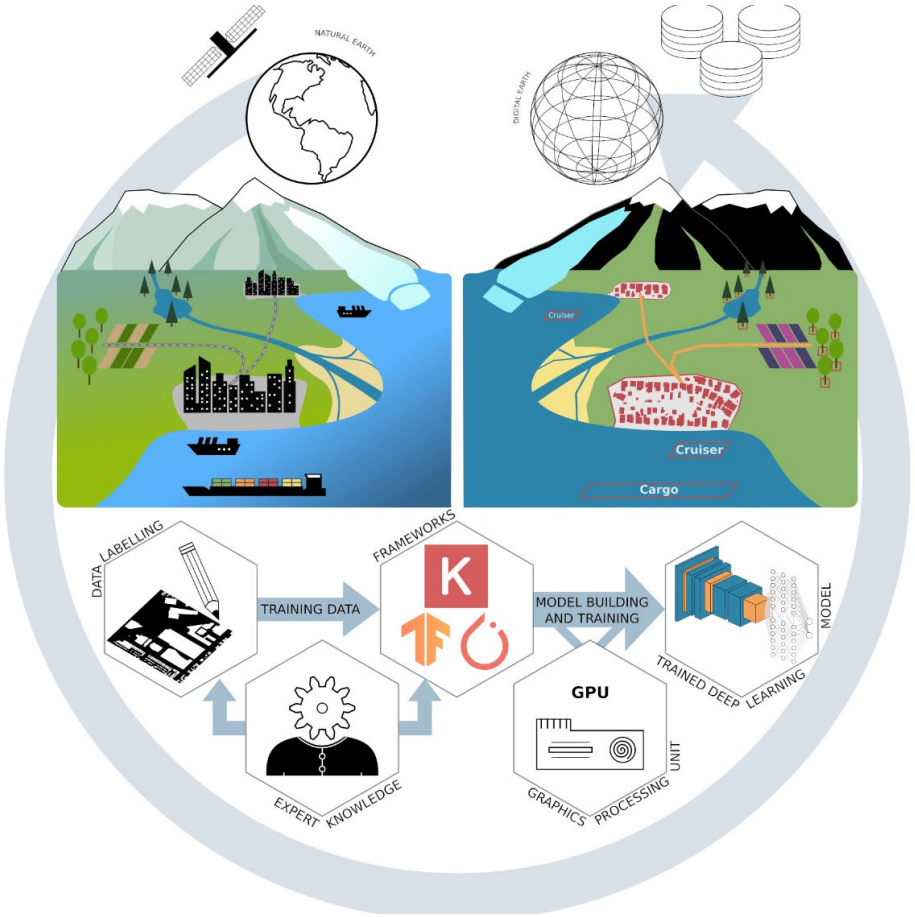
東京大学
THE UNIVERSITY OF TOKYO

RIKEN AIP

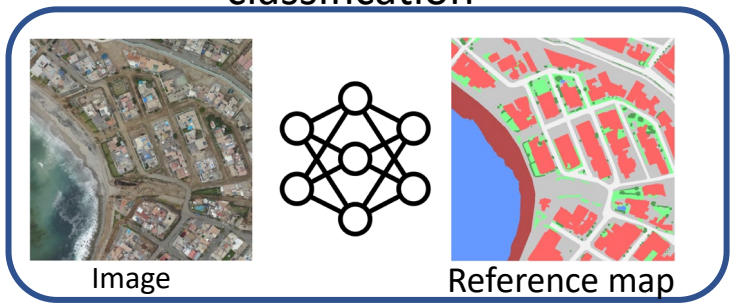
Overview


Overview
Construct
Train
Evaluation
blender
SynRS3D
RS3DAda
Land Cover Mapping
Heigh Mapping
Real Data
RS3DAda
Land Cover Mapping
Heigh Mapping
Disaster Assessment
Develop
Develop
Mapping

Remote Sensing Image Processing

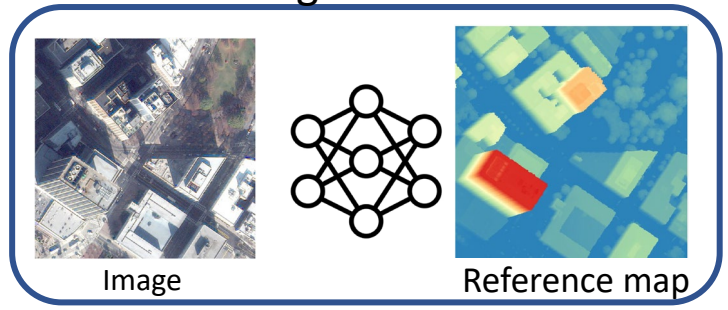


classification



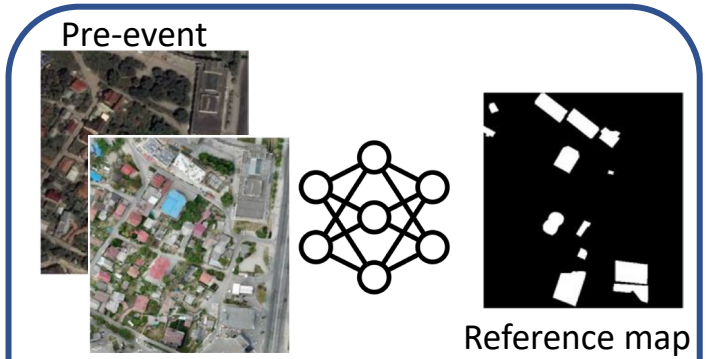
Task 1: Land Cover Mapping

regression



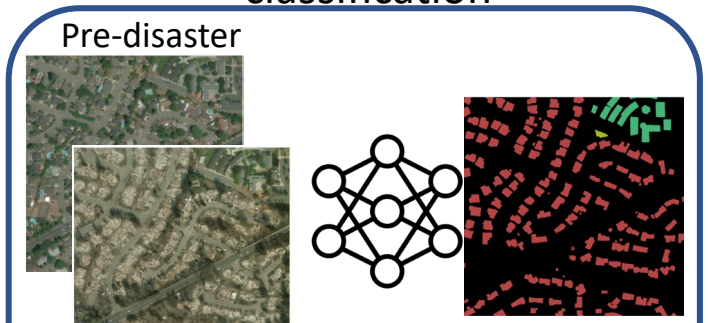
Task 2: Height Estimation

classification



Task 3: Building Change Detection

classification



Task 4: Disaster Mapping

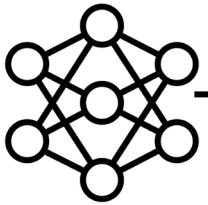
Fig1: Villalobos, Pablo, et al. "Machine Learning Model Sizes and the Parameter Gap." *arXiv:2207.02852,* 2022.

Challenges

- Expensive to build a real-world remote sensing benchmark dataset

- Lack of publicly available open-source data.
- Commercial data is expensive.
- Regional policies restrict data access.

High-Resolution RGB



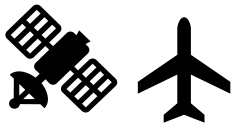
Labels of Land Cover Maps

- Requires expert knowledge.
- Time-consuming.
- Expensive.

Elevation Data

- Requires extremely expensive equipment.
- Noise exists.
- Lack of publicly available open-source data.

- Domain shift & Covariate Shift

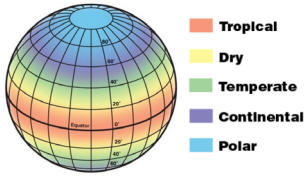


Different sensors

Different continents

Different districts

Different seasons



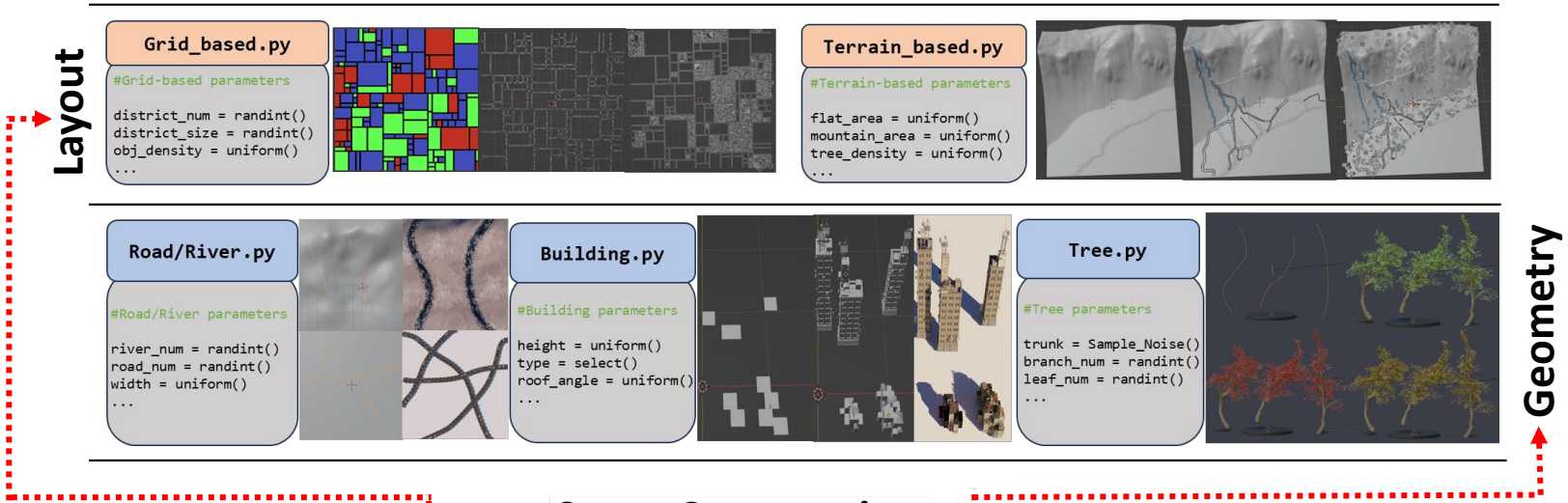
Different climate zones

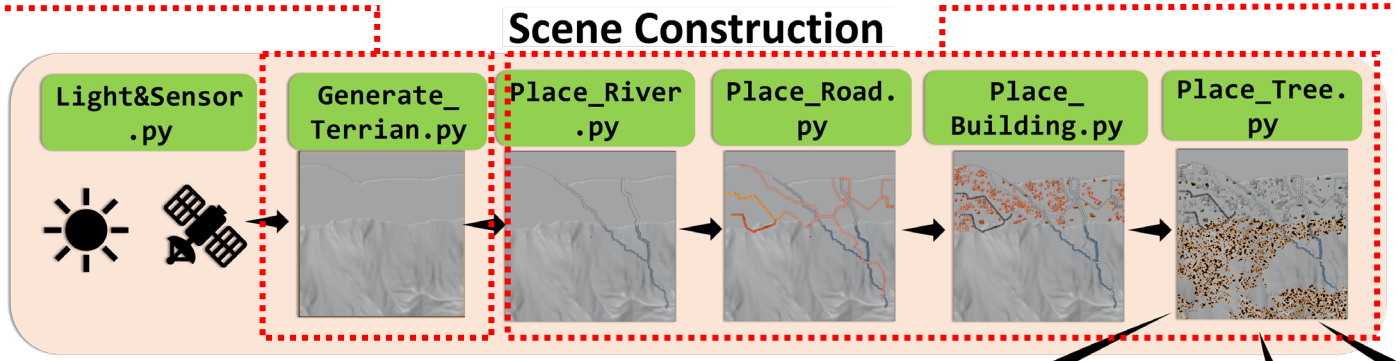
Way to Solve the Challenges:

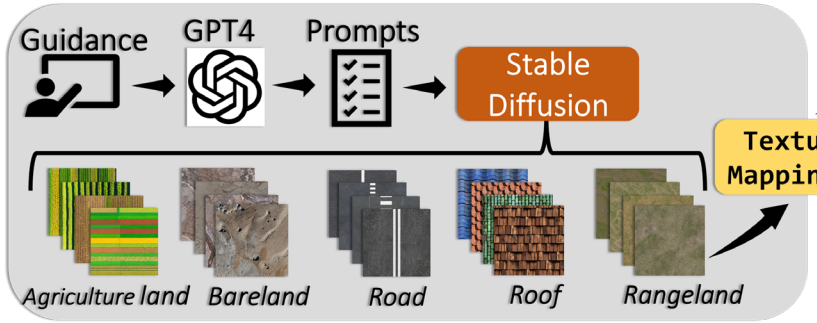
Synthetic Dataset with **UDA**

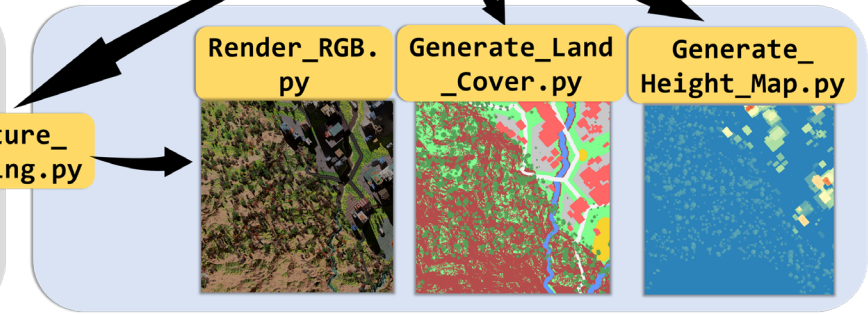
Fig1: http://www.mychimneyprofessional.com/blog
Fig2: https://sciencetrek.org/sciencetrek/topics/climate/facts.cfm

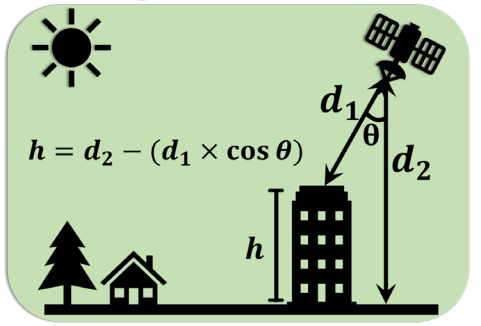
The SynRS3D Acquisition Protocol

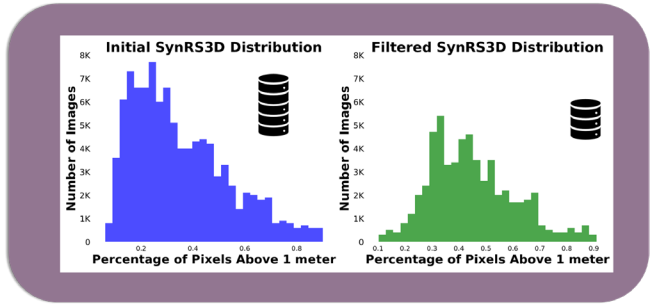






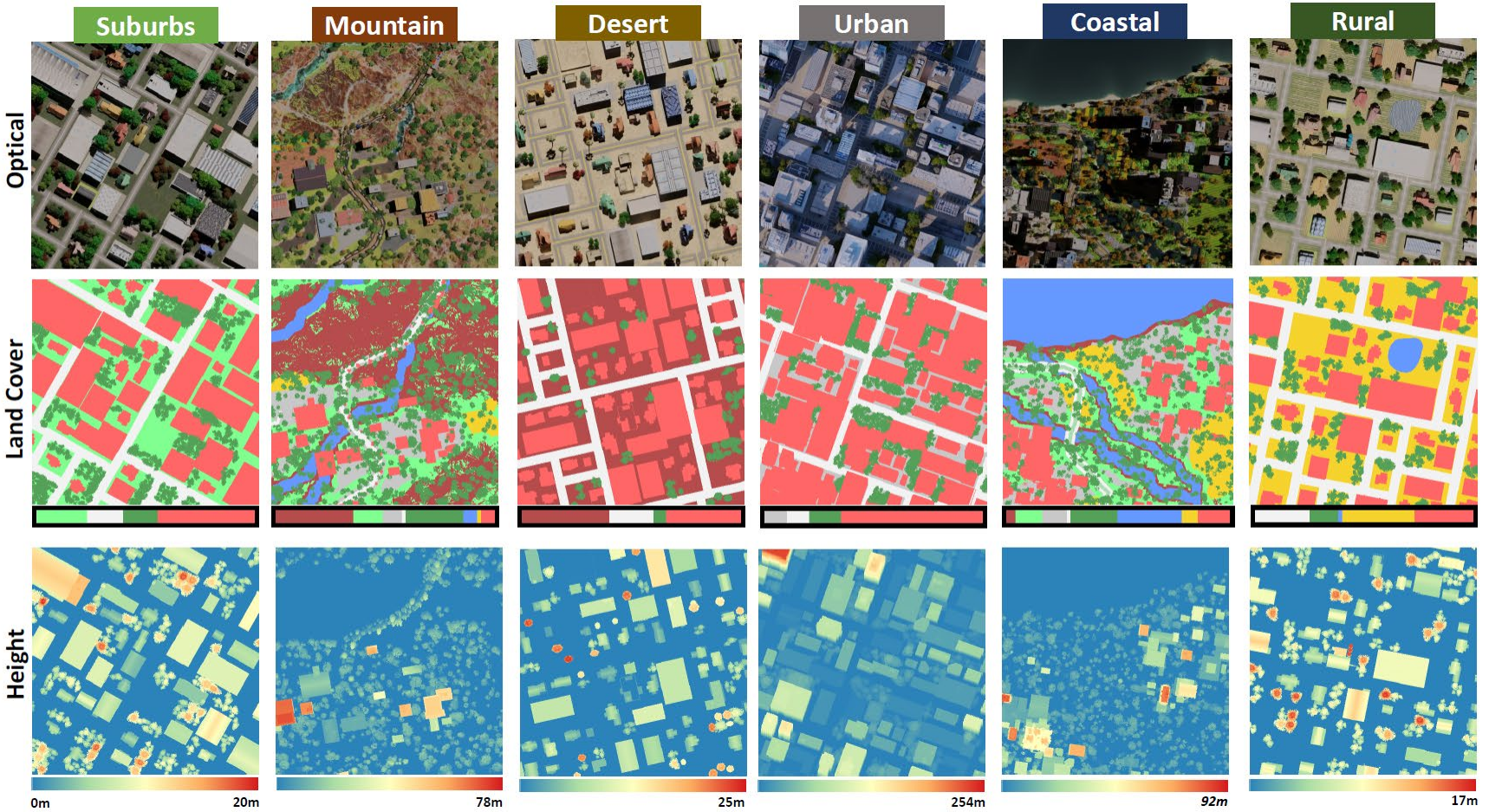


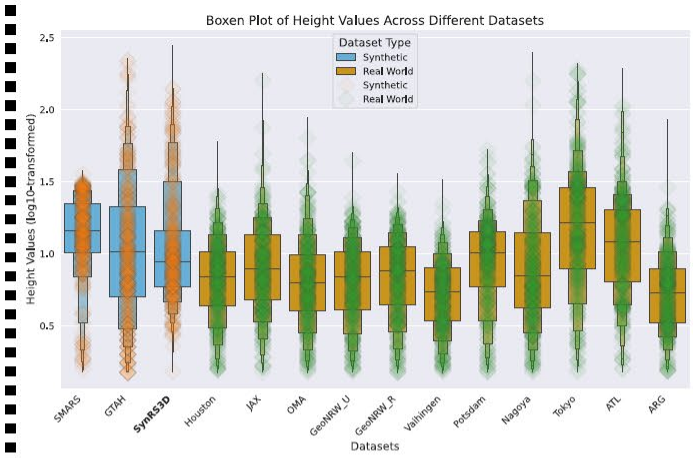


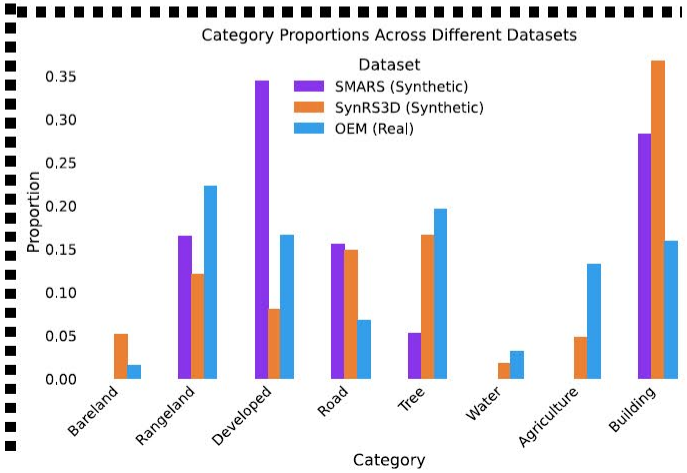


Stable Diffusion project: https://huggingface.co/spaces/stabilityai/stable-diffusion

Examples of SynRS3D Dataset & Statistics







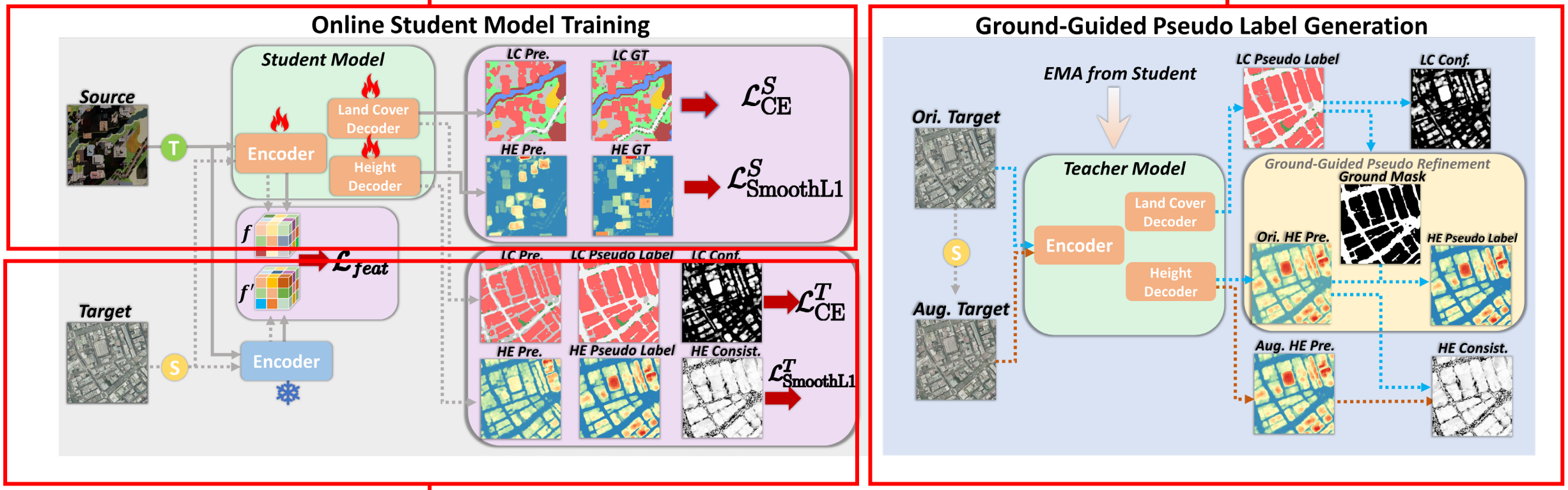
Main Characteristics

- Mimicking six city styles from around the world.
- Height distribution and land cover category distribution more closely resembling Real-World data.

RS3DAda: Unlock the Potential of the SynRS3D:

Step 1: Train on source domain dataset (Synthetic)

Step 2: Generate pseudo labels from target dataset (Real-World)

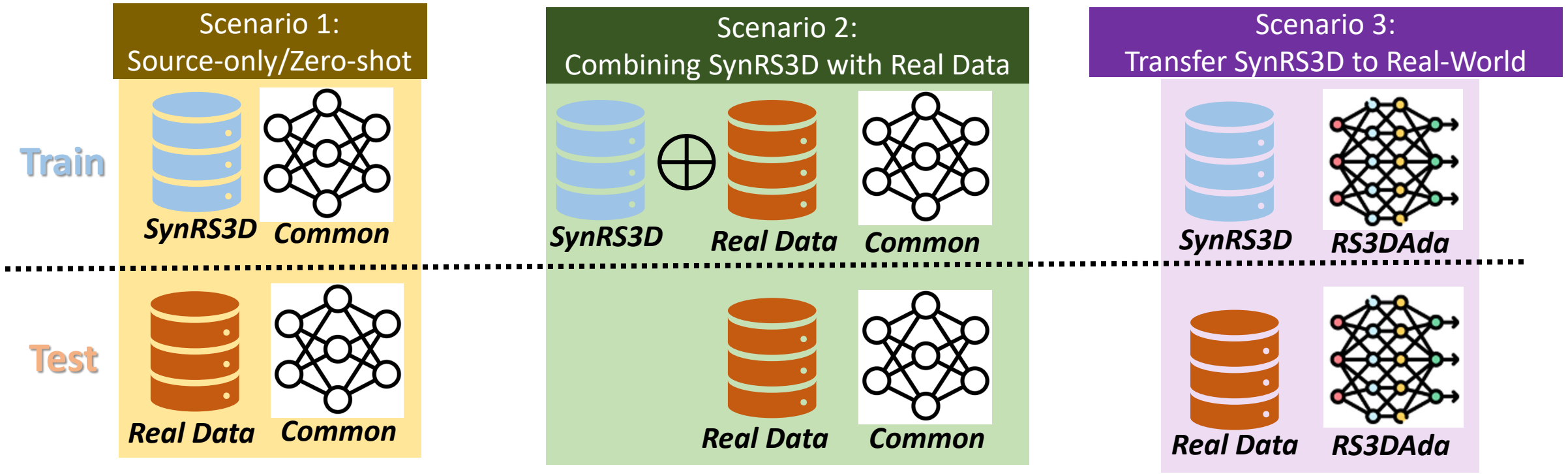


Step 3: Train on target dataset with pseudo labels

Key Ideas

- Unleashing the power of large-scale unlabeled Real-World data.
- Forcing cooperation between land cover mapping and height estimation tasks.
- Utilizing prior knowledge of remote sensing images.

Experimental Setting & Evaluation Datasets



Relatively easy to access and simple real-world datasets.

Difficult to access and more challenging real-world datasets.

Real-World Height Estimation Datasets

Types	Datasets	Region	Height mean&std	Channel
Target Domain 1	Houston [103]	US	[3.07, 5.02]	RGB
	JAX [42]	US	[4.73, 9.02]	RGB
	OMA [42]	US	[2.37, 5.27]	RGB
	GeoNRW_Urban [4]	Germany	[2.46, 4.31]	RGB
	GeoNRW_Rural [4]	Germany	[2.03, 4.21]	RGB
	Potsdam [77]	Germany	[3.02, 5.68]	RGB
Target Domain 2	ATL [12]	US	[8.40, 13.41]	RGB
	ARG [12]	Argentina	[3.90, 4.29]	RGB
	Nagoya [14]	Japan	[7.36, 11.84]	RGB
	Tokyo [14]	Japan	[15.73, 22.77]	RGB
	Vaihingen [77]	Germany	[2.36, 3.57]	NIR, G, B

(a) Real-world height estimation datasets.

Real-World Land Cover Mapping Datasets

Types	Datasets	Region	Categories
Target Domain	OEM [98]	Global	8
	Vaihingen [77]	Germany	6
	Potsdam [77]	Germany	6
	JAX [42]	US	6
	OMA [42]	US	6

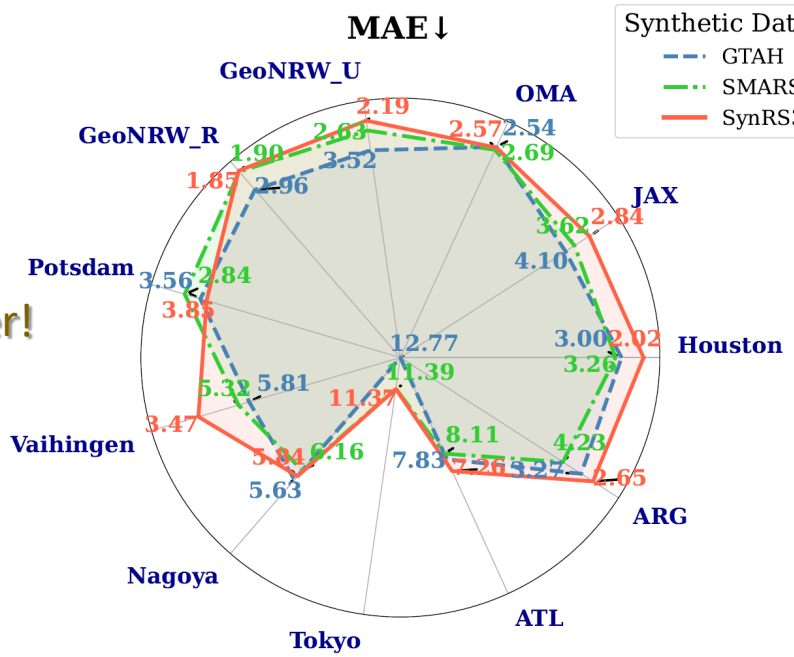
(b) Real-world land cover mapping datasets.

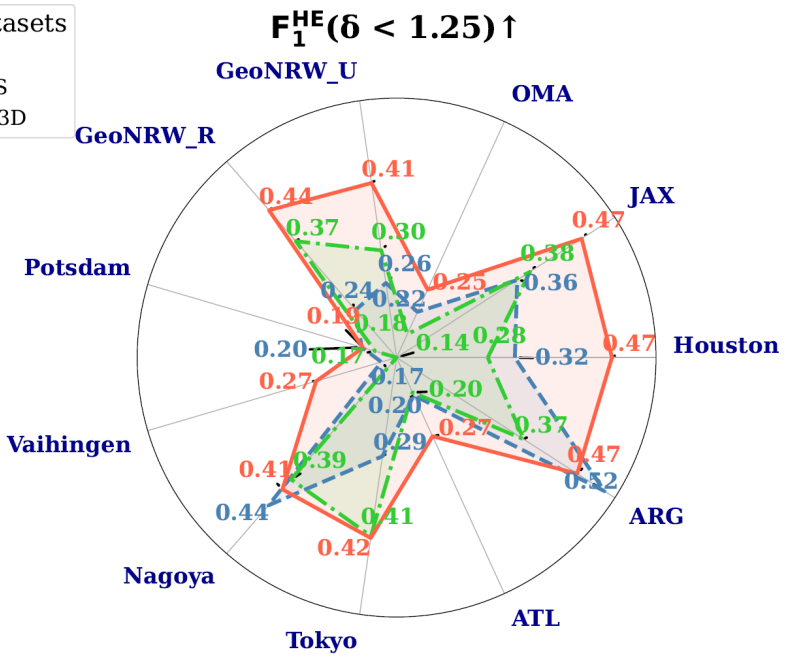
OEM spans 77 regions across 44 countries on six continents!

Scenario 1: Source-only/Zero-shot

Task: Height Estimation

The smaller the better!





The larger the better!

Task: Land Cover Mapping

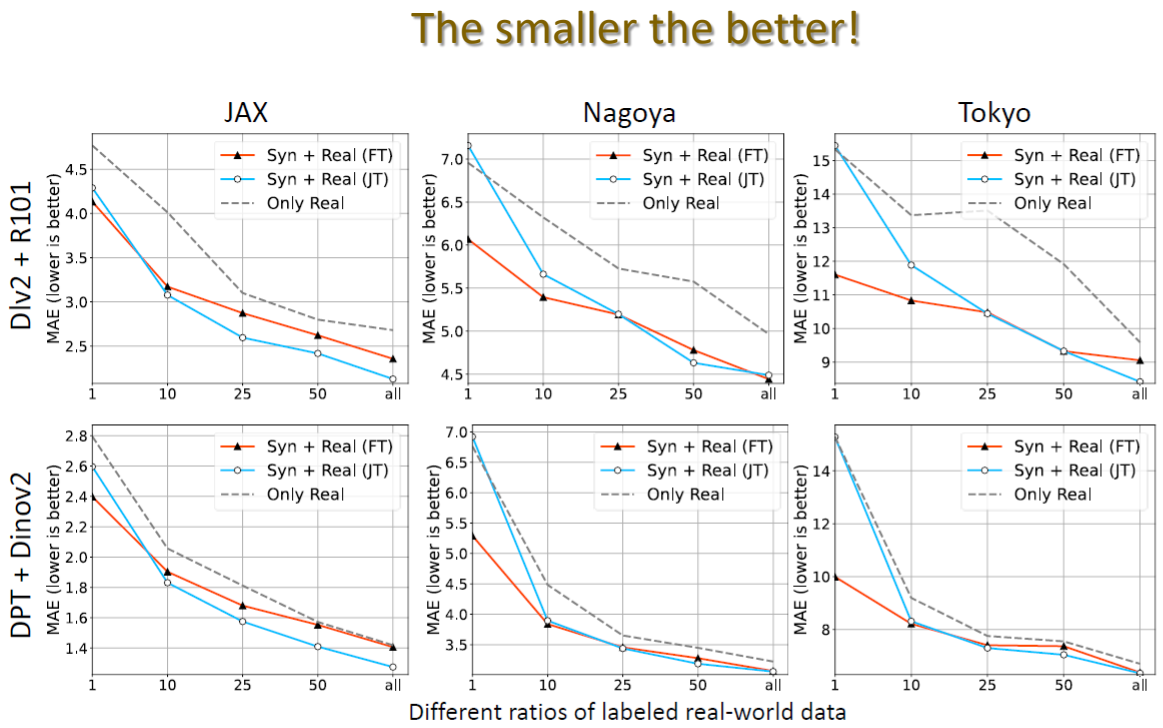
The larger the better!

Datasets	JAX [42]				OMA [42]				Vaihingen [77]				Potsdam [77]			
	Ground	Tree	Building	mIoU	Ground	Tree	Building	mIoU	Ground	Tree	Building	mIoU	Ground	Tree	Building	mIoU
SMARS [73]	76.02	43.13	61.28	60.14	82.17	17.25	59.94	53.12	74.10	58.40	74.35	68.95	68.56	5.35	57.51	43.81
SyntheWorld [83]	74.63	54.74	64.18	64.52	81.29	**45.83**	56.56	61.23	72.69	68.09	75.67	72.15	69.09	32.49	55.88	52.49
SynRS3D	**77.69**	**57.03**	**68.96**	**67.89**	**83.96**	41.08	**62.28**	**62.44**	**75.66**	**68.58**	**79.61**	**74.61**	**74.26**	**35.34**	**69.46**	**59.69**

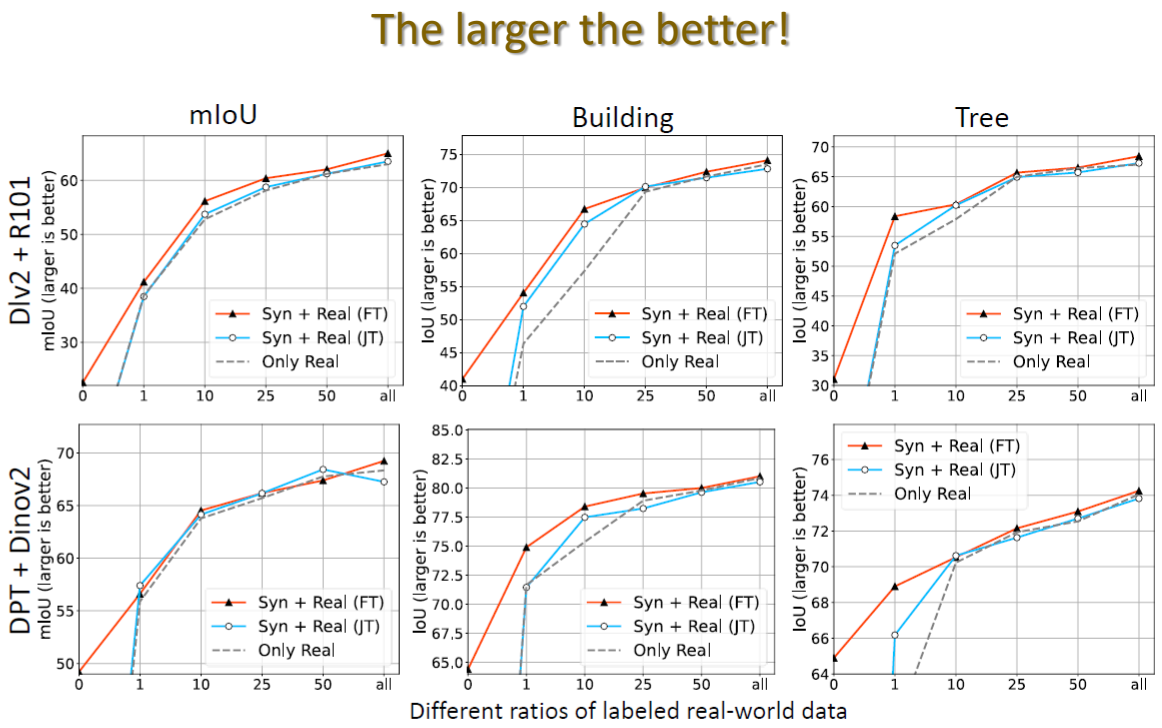
Takeaway

- The SynRS3D dataset outperforms all existing RS synthetic datasets.

Scenario 2: Combining SynRS3D with Real Data



(a) Height Estimation.



(b) Land Cover Mapping.

Takeaway

- The SynRS3D dataset serves as a valuable complement to real-world datasets, especially when available data is limited.

Scenario 3: Transfer SynRS3D to Real-World (RS3DAda)

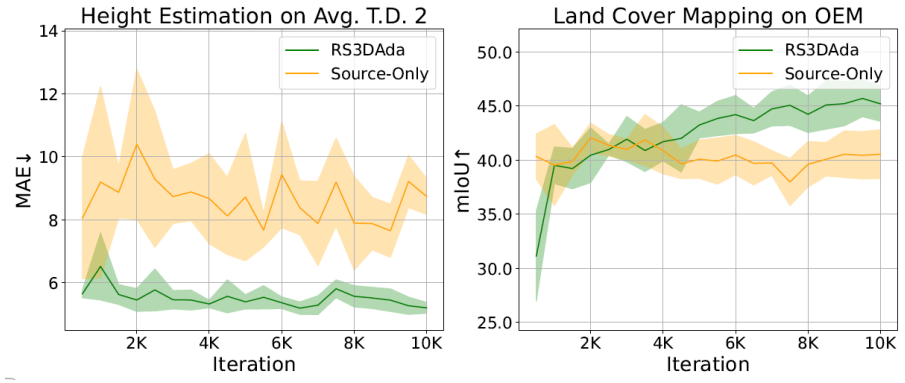
Task: Land Cover Mapping

Model	Bareland	Rangeland	Developed	Road	Tree	Water	Agriculture	Buildings	**mIoU**
Source-only	8.69	37.95	**22.54**	**49.05**	60.16	46.64	35.40	**65.19**	40.70
DAFormer [33]	12.54	41.16	10.88	43.88	**62.56**	**77.55**	62.62	59.10	46.29
RS3DAda	**19.92**	**47.61**	18.41	44.06	61.04	71.66	**63.73**	59.42	**48.23**
Train-on-OEM	50.04	59.10	58.18	65.39	73.07	83.65	76.36	80.88	68.34

Task: Height Estimation

Model	MAE ↓		RMSE ↓		Accuracy Metrics [17] ↑			F_1^{HE} ↑		
Avg. T.D.1	**Whole**	**High**	**Whole**	**High**	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Train-on-T.D.1	**1.272**	**3.363**	**2.381**	**4.329**	**0.379**	**0.463**	**0.510**	**0.617**	**0.710**	**0.742**
Source Only	2.557	5.617	4.128	6.705	0.123	0.192	0.246	0.372	0.491	0.552
RS3DAda	2.148	4.921	3.593	6.024	0.185	0.258	0.318	0.418	0.554	0.623
Avg. T.D.2	**Whole**	**High**	**Whole**	**High**	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Train-on-T.D.1	5.378	8.302	8.301	10.714	0.146	0.244	0.336	0.384	0.535	0.627
Source Only	6.117	8.923	9.221	11.443	0.125	0.223	0.312	0.365	0.514	0.601
RS3DAda	**4.866**	**7.227**	**7.584**	**9.594**	**0.182**	**0.299**	**0.389**	**0.485**	**0.621**	**0.689**

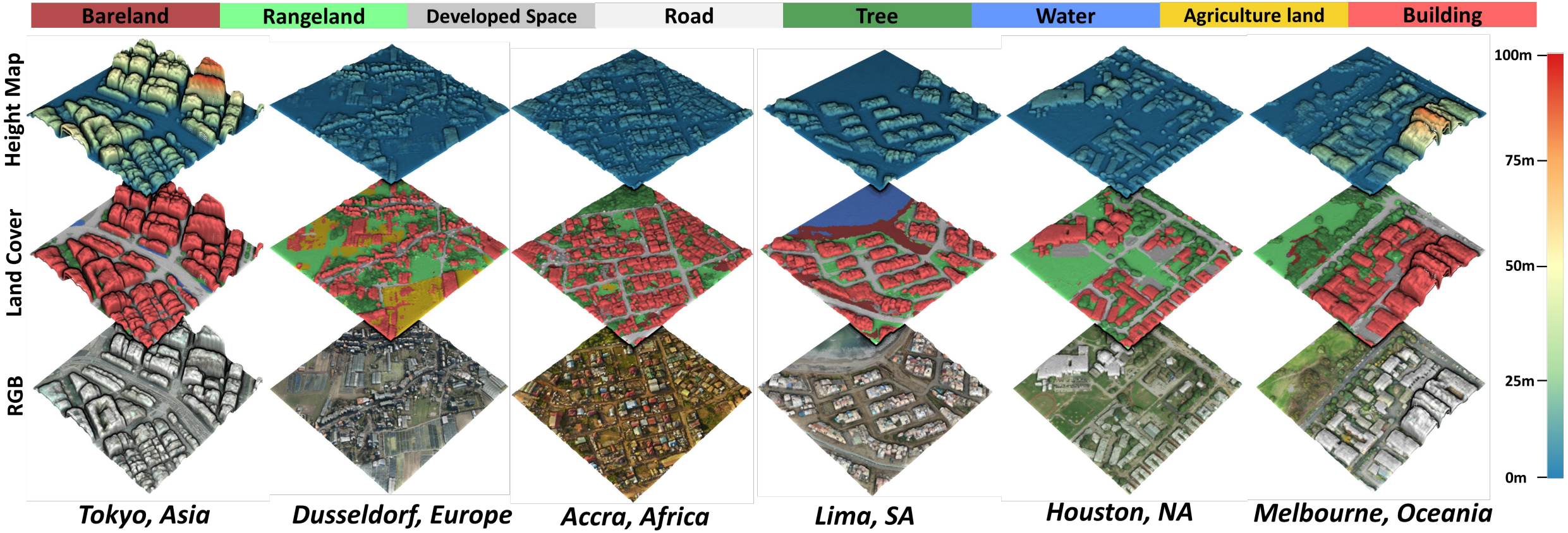
Training curves



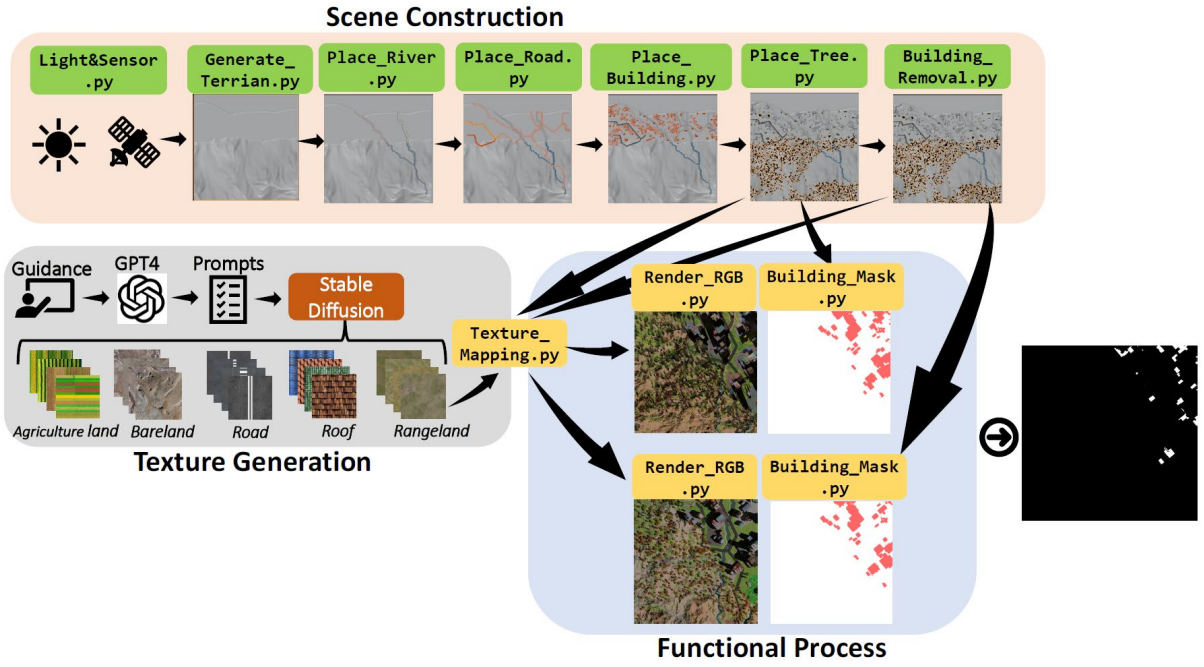
Takeaway

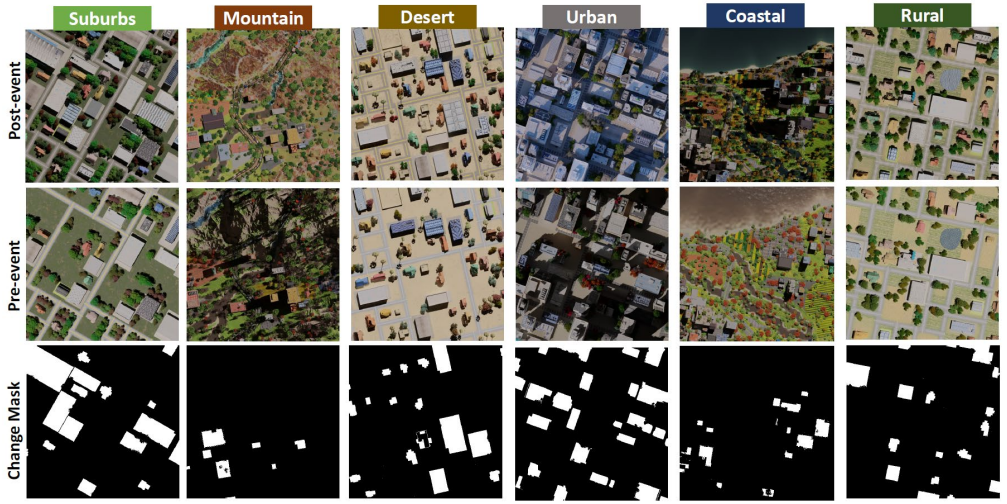
- RS3DAda has greatly unlocked the potential of the SynRS3D dataset.
- RS3DAda has significantly stabilized the training process of the SynRS3D dataset.

3D Semantic Reconstruction Outcomes



By-Product 1: Building Change Detection





Peformance evaluation of building change detection task on WHU-CD [38] dataset.

Train on	DTCDSCN [55]		ChangeFormer [5]		ChangeMamba [9]		DinoMamba	
	IoU	F1	IoU	F1	IoU	F1	IoU	F1
SMARS [73]	26.84	42.55	18.67	31.88	42.50	59.63	48.11	64.87
SyntheWorld [83]	30.17	46.53	**41.73**	**58.87**	47.26	64.10	54.20	70.14
SynRS3D	**33.09**	**49.84**	35.00	51.94	**52.94**	**69.08**	**61.60**	**76.00**
Real	58.31	73.67	79.98	88.88	88.44	93.87	87.57	93.38

Peformance evaluation of building change detection task on LEVIR-CD+ [8] dataset.

Train on	DTCDSCN [55]		ChangeFormer [5]		ChangeMamba [9]		DinoMamba	
	IoU	F1	IoU	F1	IoU	F1	IoU	F1
SMARS [73]	11.70	21.53	15.67	27.58	27.50	42.50	30.85	47.31
SyntheWorld [83]	21.16	35.28	23.31	38.12	28.28	44.30	48.78	65.46
SynRS3D	**25.82**	**41.30**	**23.33**	**38.14**	**30.39**	**46.78**	**49.63**	**66.23**
Real	63.44	77.63	67.48	80.58	77.39	87.25	74.12	85.14

Peformance evaluation of building change detection task on the SECOND [105] dataset.

Train on	DTCDSCN [55]		ChangeFormer [5]		ChangeMamba [9]		DinoMamba	
	IoU	F1	IoU	F1	IoU	F1	IoU	F1
SMARS [73]	17.26	29.88	23.30	38.09	29.85	46.15	35.20	51.07
SyntheWorld [83]	21.00	35.07	26.44	42.06	27.23	43.02	37.61	54.71
SynRS3D	**33.52**	**50.32**	**31.36**	**47.90**	**38.88**	**56.02**	**39.18**	**56.33**
Real	58.78	74.04	60.08	75.06	67.61	80.68	67.65	80.71

By-Product 2: Disaster Mapping

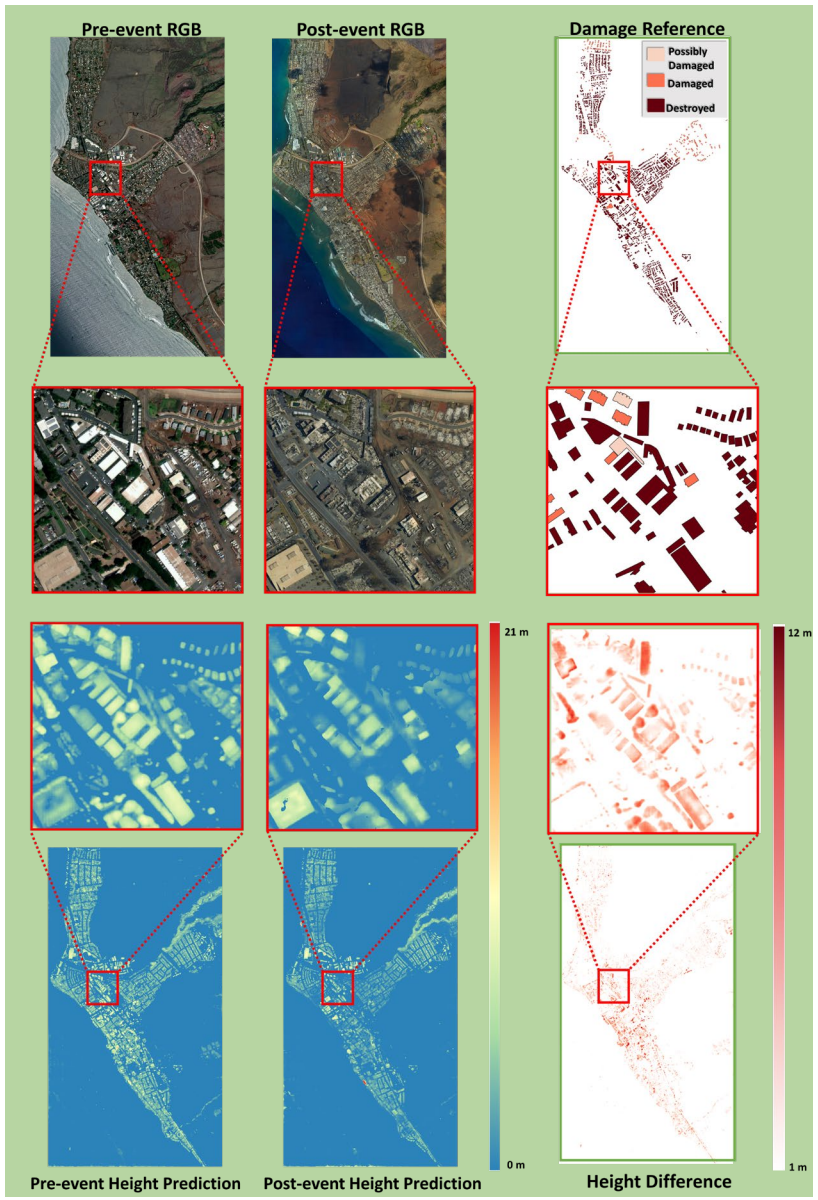
Paradigm

Pre-event RGB → RS3DAda → Pre-event Height Prediction

Post-event RGB → RS3DAda → Post-event Height Prediction

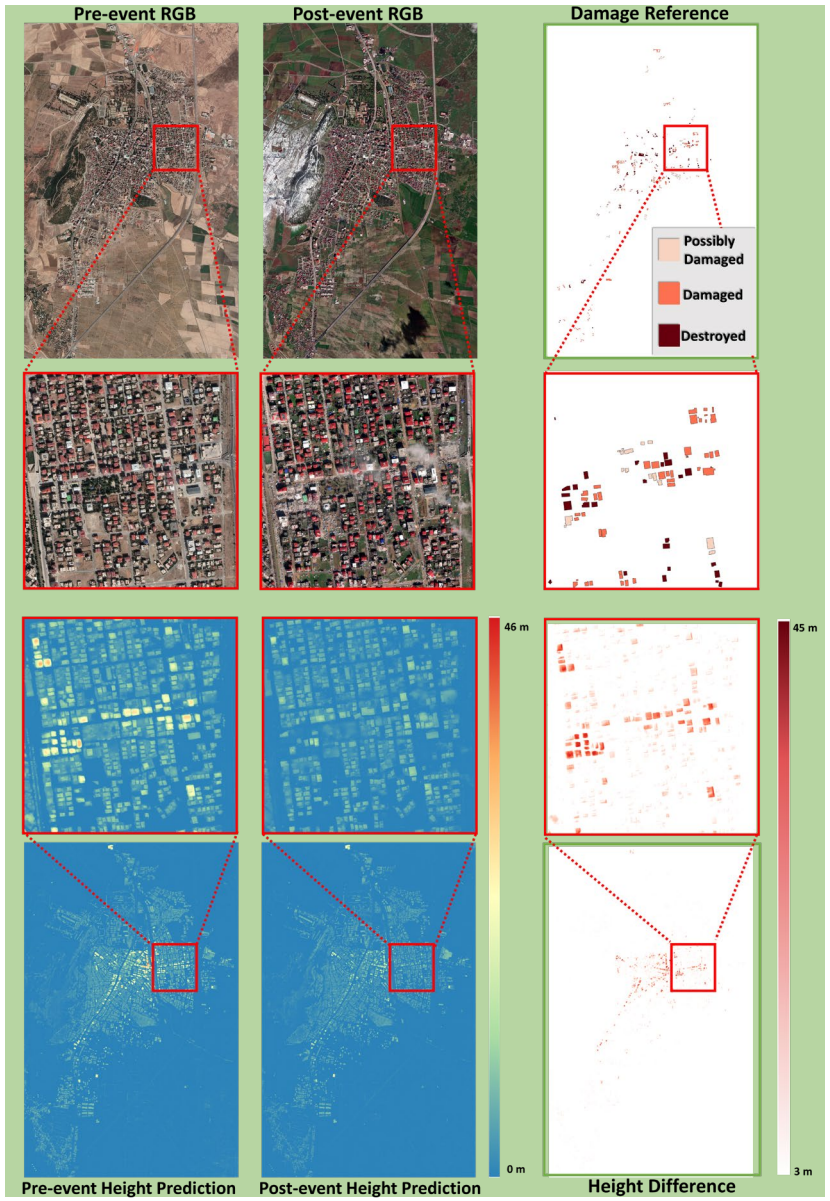
Pre-event Height Prediction − Post-event Height Prediction = Height Difference

Study Case 1: 2023 Hawaii Wildfire



Images@Google Satellite & Bing Satellite.

Study Case 2: 2023 Turkey Earthquake



Images@CNES/Airbus, Maxar Technologies.

Summary

- **Dataset**
  - **SynRS3D** – the largest synthetic dataset for remote sensing, with high-res RGB images, land cover classes, and height references.
- **Method**
  - **RS3DAda** – the first multi-task UDA model for RS syn-to-real adaptation.
- **Real-world Validation**
  - Benchmarked in **3 scenarios** on real RS datasets, showing strong performance gains.
- **Broad Applicability**
  - Building change detection
  - Disaster mapping