BUFFER-X: Towards Zero-Shot Point Cloud Registration in Diverse Scenes

ICCV 2025 🌟Highlight🌟

Minkyun Seo*, Hyungtae Lim*, Kanghee Lee,
Luca Carlone, Jaesik Park†

*These authors contributed equally to this work.

†Corresponding author.

(Unit: Success rate [%])

This video includes audio narration

Three Key Factors

Three factors crucial for achieving domain generalization in registration

1. Proper voxel size and search radius

2. Robust keypoint detection for out-of-domain scenes

3. Input Scale normalization

Three Key Factors

Three factors crucial for achieving domain generalization in registration

1. Proper voxel size and search radius

2. Robust keypoint detection for out-of-domain scenes

3. Input Scale normalization

Issue #1: Most approaches require manual tuning of voxel size and search radius by users.

Three Key Factors

Three factors crucial for achieving domain generalization in registration

1. Proper voxel size and search radius

2. Robust keypoint detection for out-of-domain scenes

3. Input Scale normalization

Issue #2: Learning-based keypoint extractor modules are empirically brittle to out-of-domain data.

Three Key Factors

Three factors crucial for achieving domain generalization in registration

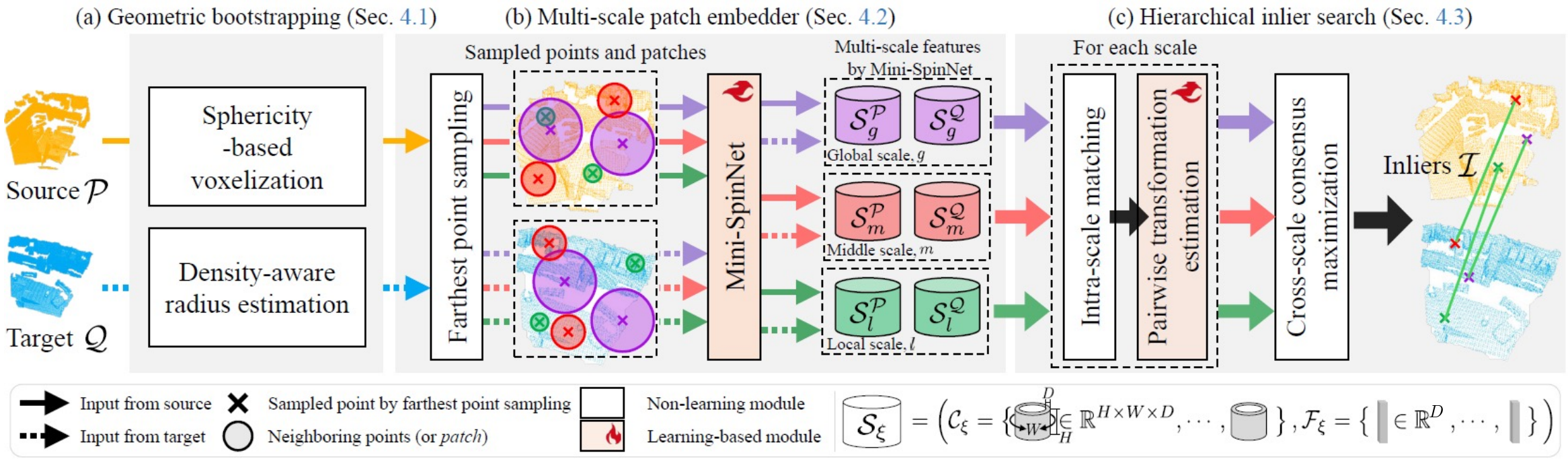
1. Proper voxel size and search radius

2. Robust keypoint detection for out-of-domain scenes

3. Input Scale normalization

Issue #3: Directly feeding raw x, y, and z values into the network leads to strong in-domain dependency

BUFFER-X: Multi-scale patch-based method for *zero-shot registration*

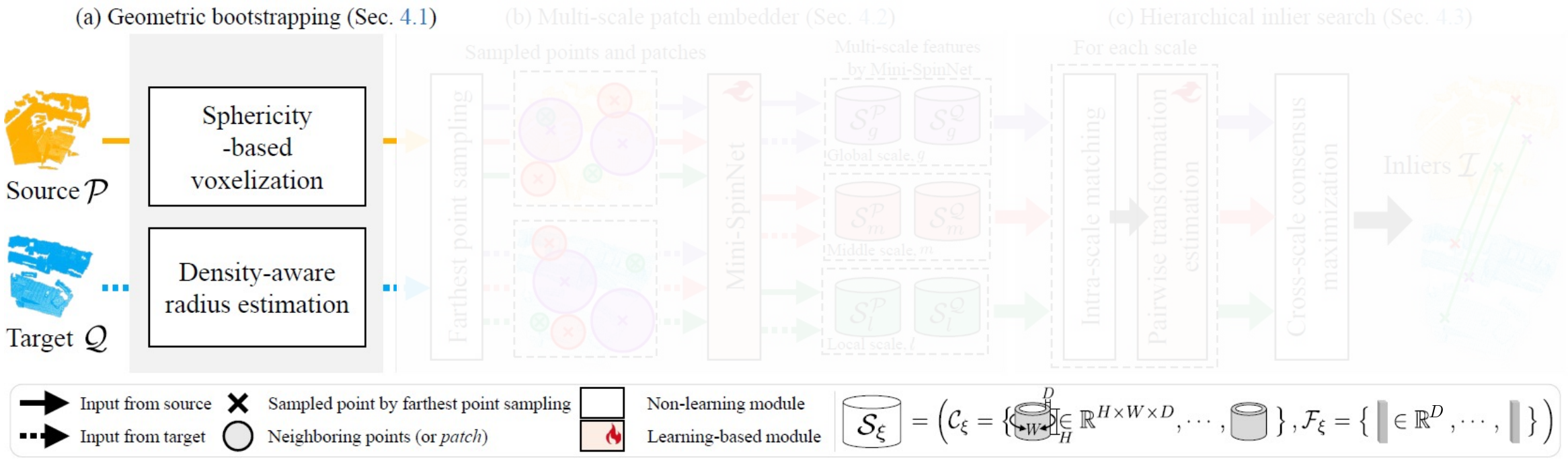


This video includes audio narration

BUFFER-X: Multi-scale patch-based method for *zero-shot registration*

(Cont'd)

※ The highlighted part below addresses the issue above.

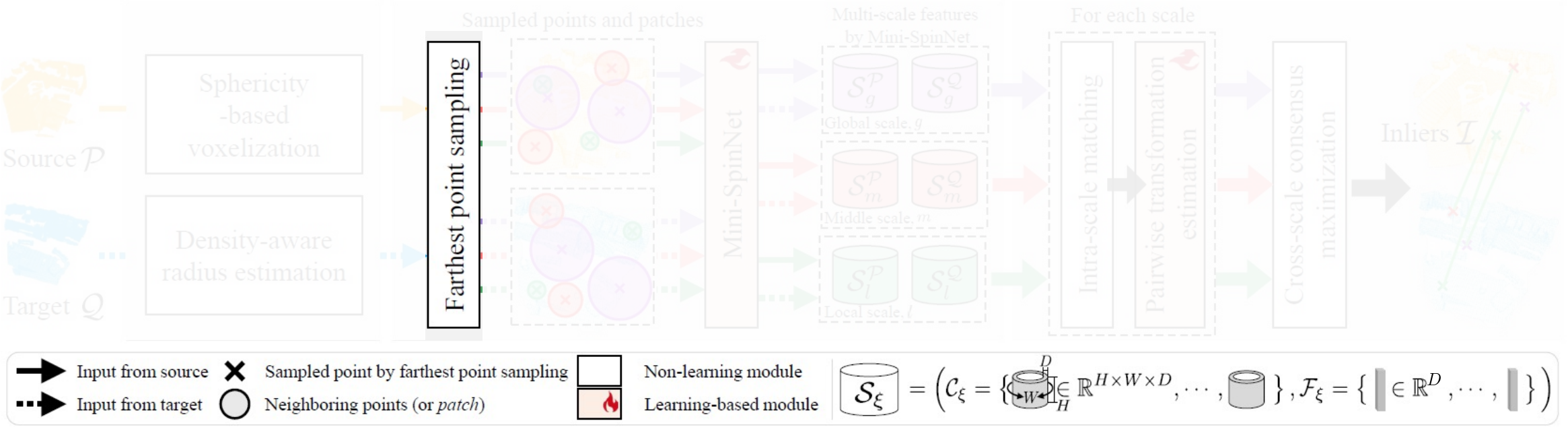


This video includes audio narration

BUFFER-X: Multi-scale patch-based method for *zero-shot registration*

(Cont'd)

※ The highlighted part below addresses the issue above.

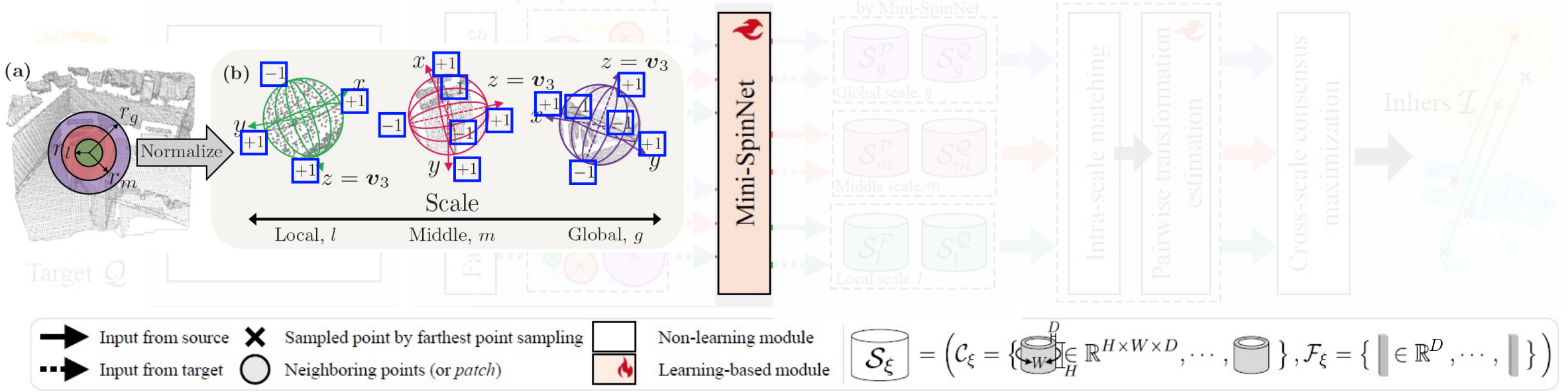


This video includes audio narration

BUFFER-X: Multi-scale patch-based method for *zero-shot registration*

(Cont'd)

※ The highlighted part below addresses the issue above.

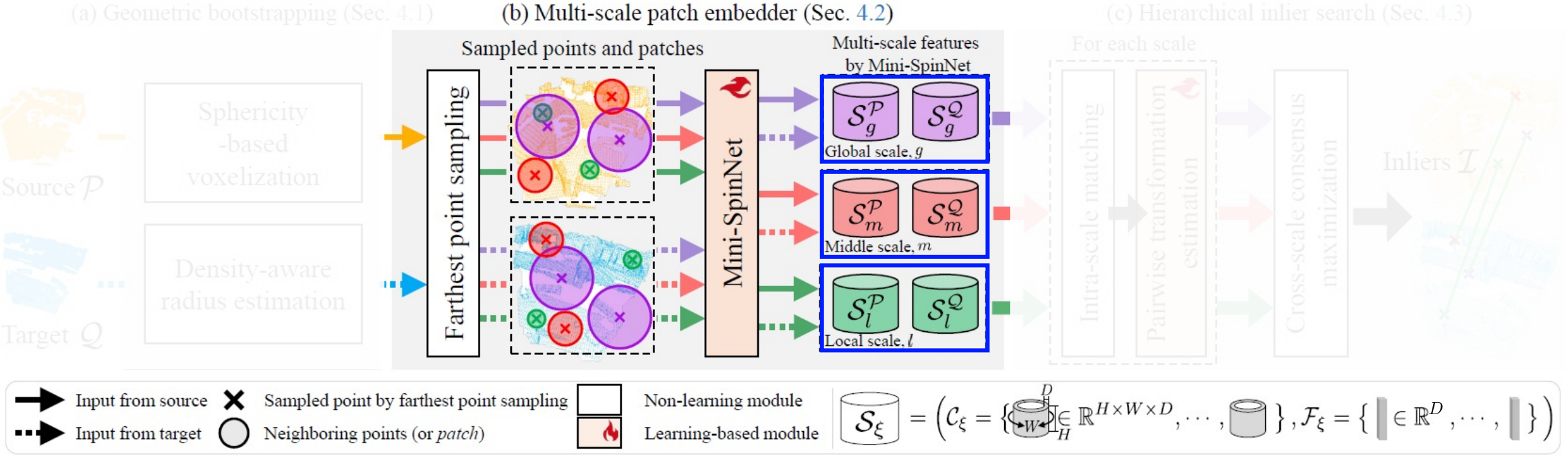


$$\mathcal{S}_\xi = \left(\mathcal{C}_\xi = \{ \in \mathbb{R}^{H\times W\times D}, \cdots, \}, \mathcal{F}_\xi = \{ \in \mathbb{R}^D, \cdots, \}\right)$$

This video includes audio narration

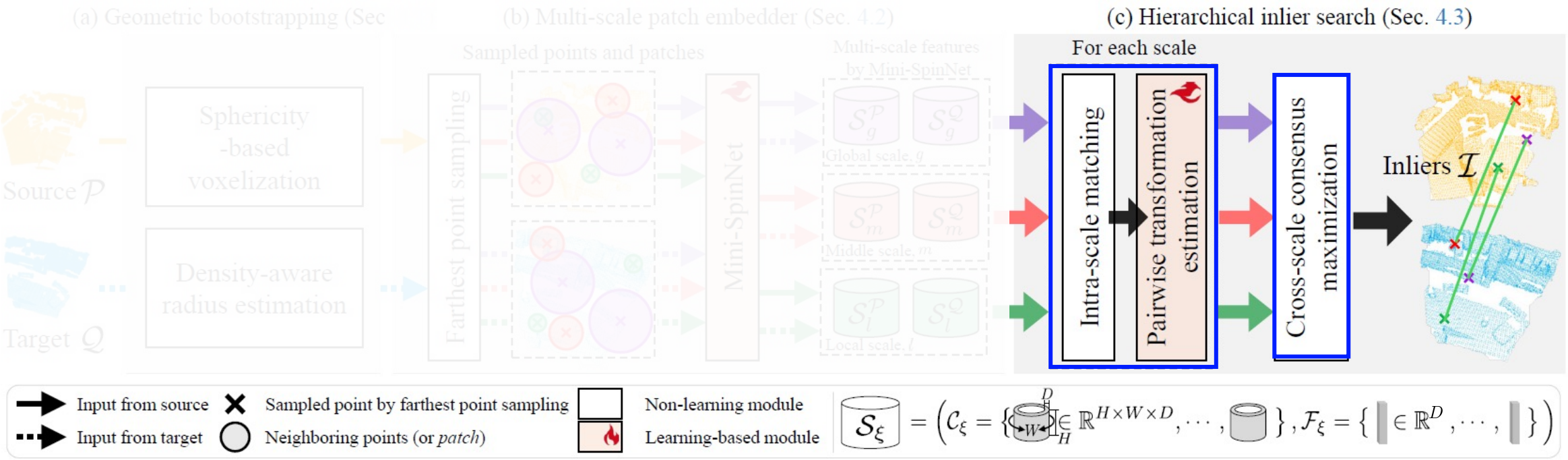
BUFFER-X: Multi-scale patch-based method for *zero-shot registration*

(Cont'd)



BUFFER-X: Multi-scale patch-based method for *zero-shot registration*

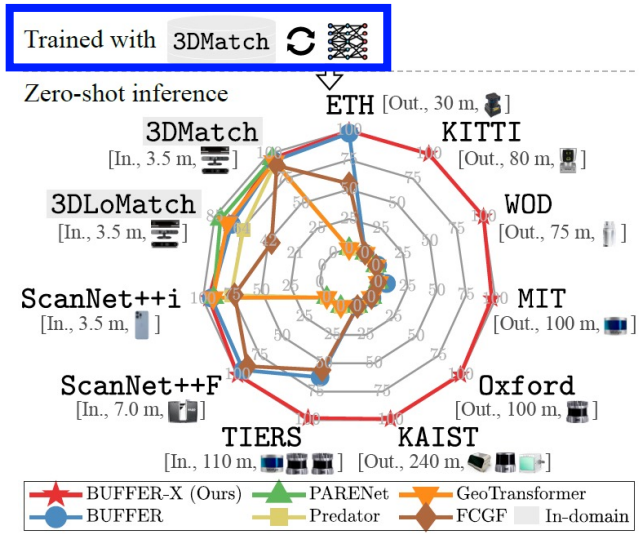
(Cont'd)



This video includes audio narration

Visual & Geometric Intelligence Lab. SPARKlab

Experimental Results



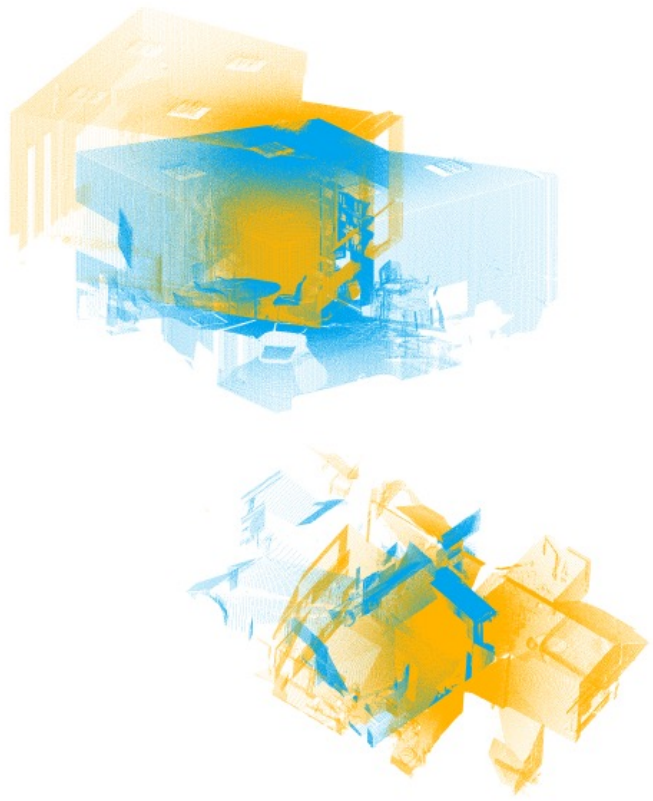
	Env.	Indoor					Outdoor						Average rank
	Dataset	3DMatch	3DLoMatch	ScanNet++i	ScanNet++F	TIERS	KITTI	WOD	KAIST	MIT	ETH	Oxford	
Conventional	FPFH [63] + FGR [94] + 👁	62.53	15.42	77.68	92.31	80.60	98.74	**100.00**	89.80	74.78	91.87	99.00	9.55
	FPFH [63] + Quatro [43] + 👁	8.22	1.74	9.88	97.27	86.57	99.10	**100.00**	91.46	79.57	51.05	91.03	10.73
	FPFH [63] + TEASER++ [77] + 👁	52.00	13.25	66.15	97.22	73.13	98.92	**100.00**	89.20	71.30	93.69	99.34	10.00
Deep learning-based	FCGF [22]	88.18	40.09	72.90	88.69	55.96	0.00	0.00	0.00	0.00	54.98	0.00	15.00
	+ 👁	88.18	40.09	85.87	88.69	78.62	90.27	97.69	92.91	92.61	54.98	93.68	10.18
	+ 👁 +	88.18	40.09	85.87	88.69	80.11	94.41	97.69	93.55	93.04	55.53	95.68	9.55
	Predator [32]	90.60	62.40	75.94	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	15.73
	+ 👁	90.60	62.40	75.94	29.81	56.44	0.00	0.00	0.95	0.00	0.14	0.33	14.55
	+ 👁 +	90.60	62.40	75.94	86.01	75.74	77.29	86.92	87.09	79.56	54.42	93.68	11.82
	GeoTransformer [59]	92.00	75.00	91.18	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	14.00
	+ 👁	92.00	75.00	91.18	7.54	5.06	0.36	0.77	0.25	0.87	0.00	0.33	13.09
	+ 👁 +	92.00	75.00	92.72	97.02	92.99	92.43	89.23	91.86	95.65	71.53	97.01	6.27
	BUFFER [5]	92.90	71.80	92.72	93.75	62.30	0.00	1.54	0.50	6.96	97.62	0.66	10.45
	+ 👁	92.90	71.80	93.01	94.69	88.96	99.46	**100.00**	97.24	95.65	99.30	99.00	3.82
	+ 👁 +	92.90	71.80	93.01	94.69	88.96	99.46	**100.00**	97.24	95.65	99.30	99.00	3.82
	PARENet [80]	95.00	**80.50**	90.84	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	13.27
	+ 👁	95.00	**80.50**	90.84	43.75	6.21	0.18	0.77	0.75	1.30	1.40	1.66	11.55
	+ 👁 +	95.00	**80.50**	90.84	87.95	75.06	84.86	92.31	86.44	84.78	69.42	93.36	8.82
	Ours with only r_m	93.38	71.69	93.10	99.60	90.80	**99.82**	**100.00**	99.05	95.65	99.30	99.34	3.00
	Ours	**95.58**	74.18	**94.99**	**99.90**	**93.45**	**99.82**	**100.00**	**99.15**	**97.30**	**99.72**	**99.67**	**1.55**

(Unit: Success rate [%])

Experimental Results (Cont'd)

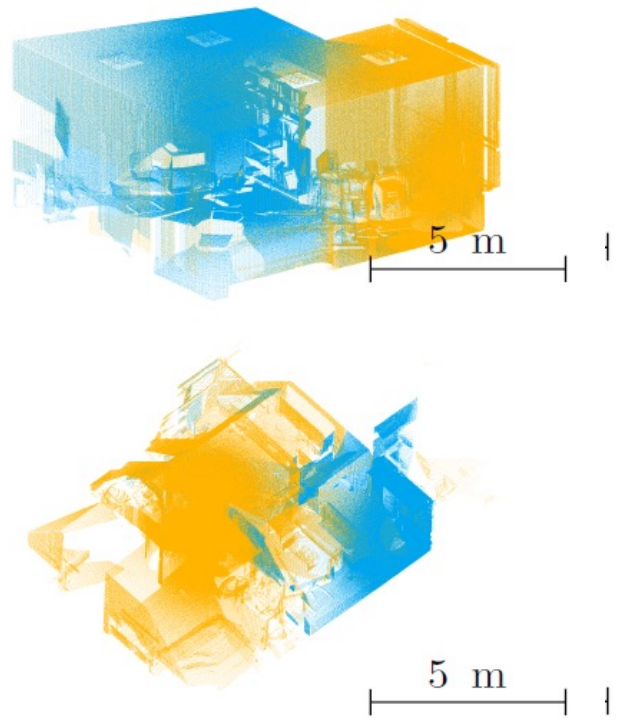
Outdoor scenes acquired by 

*which qualitatively demonstrate applicability across **various sensors and diverse scenes***



Source and target (input)

BUFFER-X (Ours)



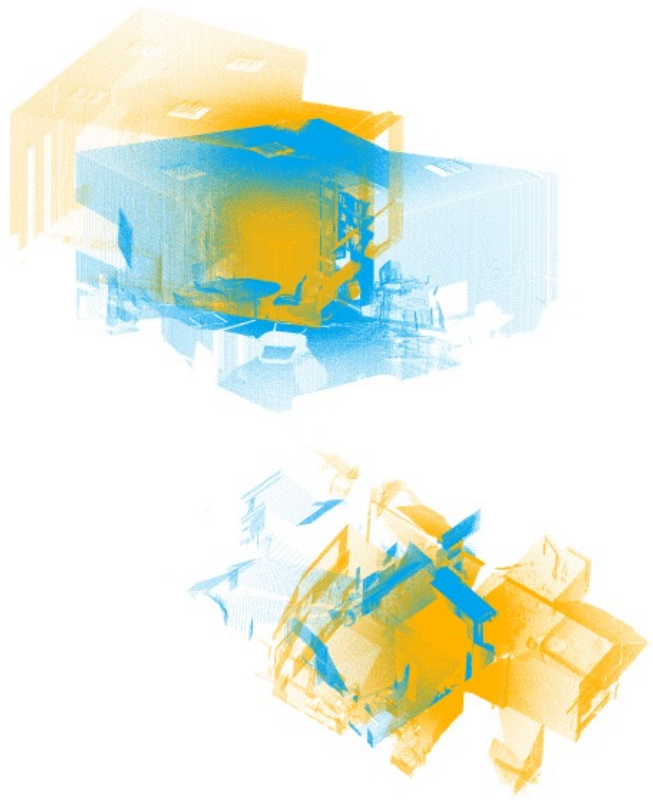
Ground truth

This video includes audio narration

Experimental Results (Cont'd)

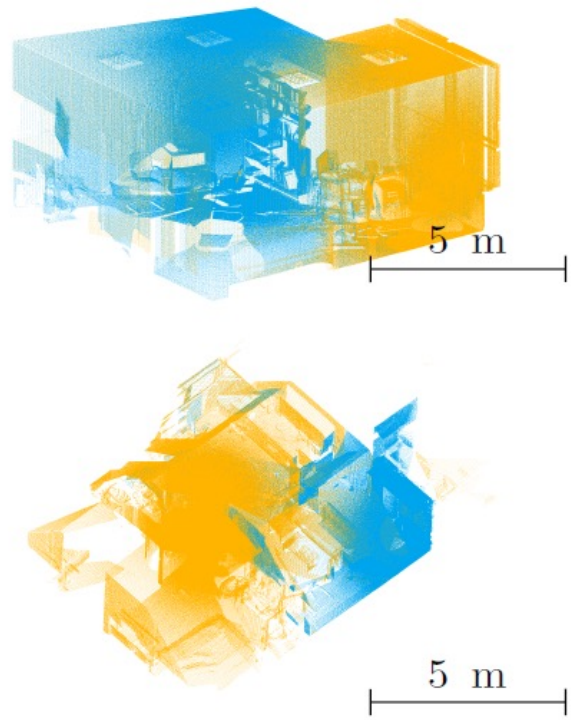
Indoor scenes acquired by 

*which qualitatively demonstrate applicability across **various sensors and diverse scenes***



Source and target (input)

BUFFER-X (Ours)



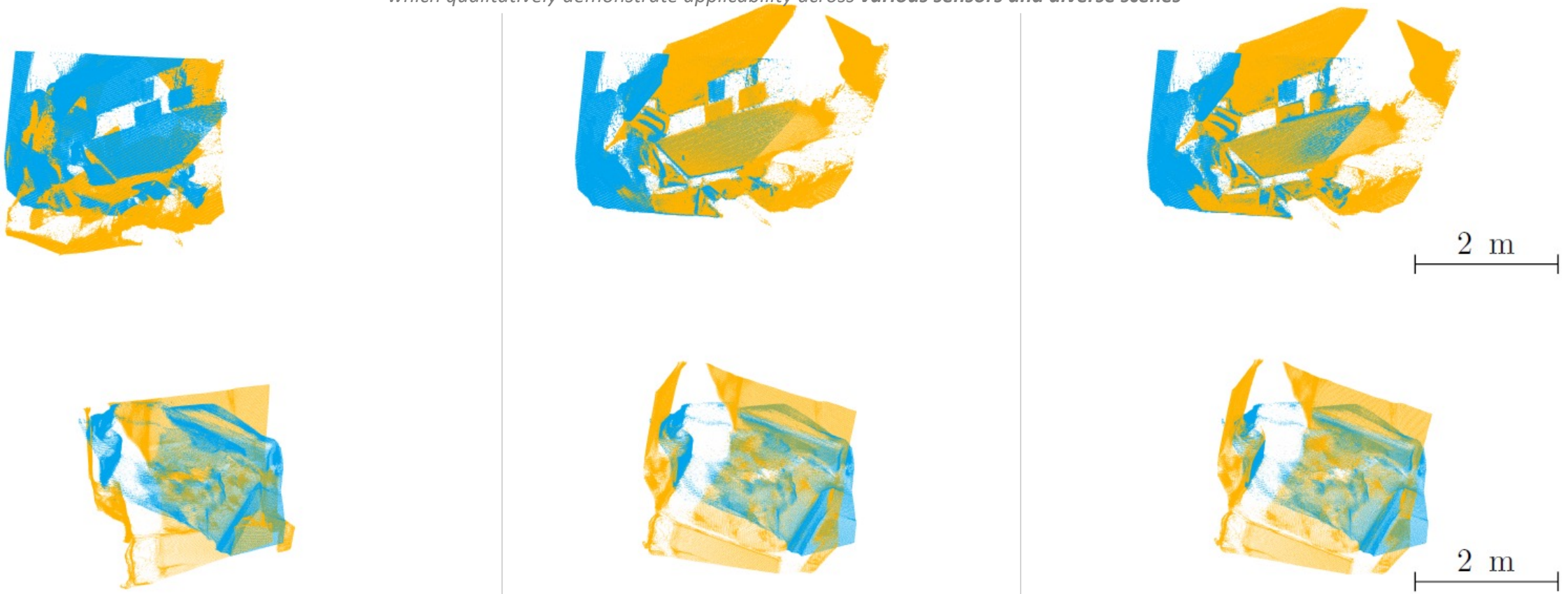
Ground truth

This video includes audio narration

Experimental Results (Cont'd)

Indoor scenes acquired by 

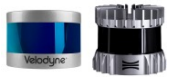
*which qualitatively demonstrate applicability across **various sensors and diverse scenes***



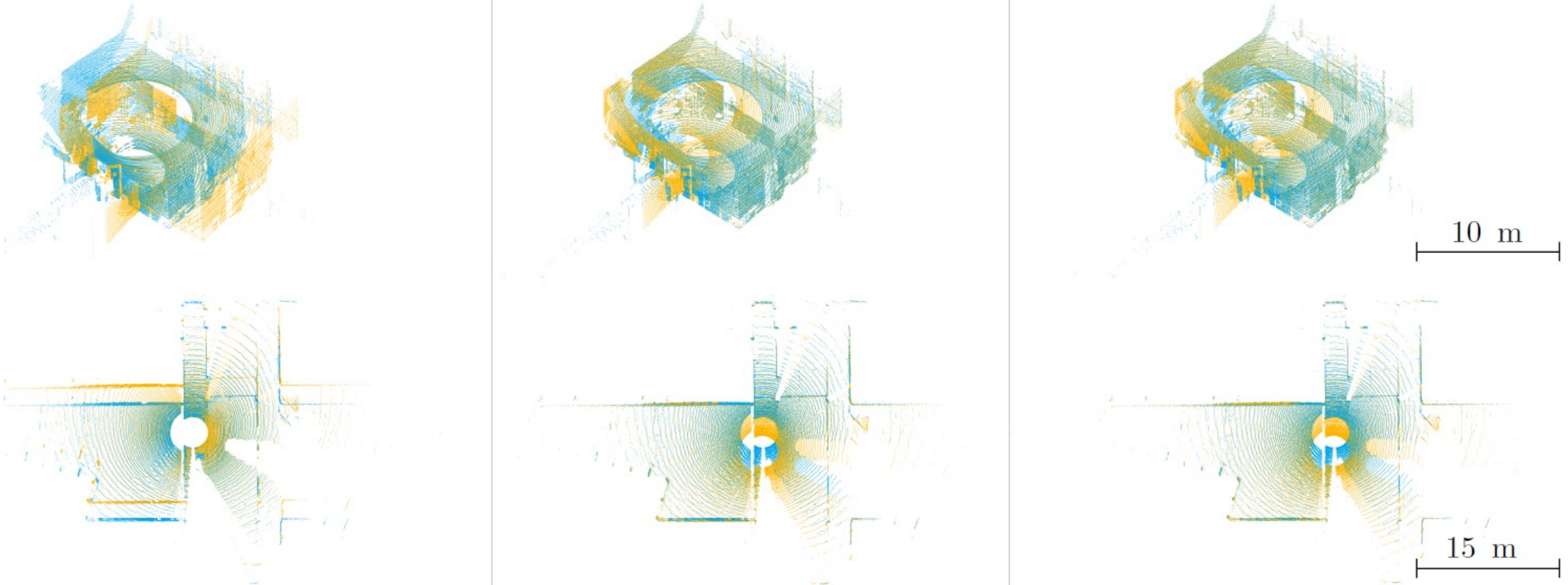
Source and target (input) | BUFFER-X (Ours) | Ground truth

This video includes audio narration

Experimental Results (Cont'd)

Indoor scenes acquired by 

*which qualitatively demonstrate applicability across **various sensors and diverse scenes***



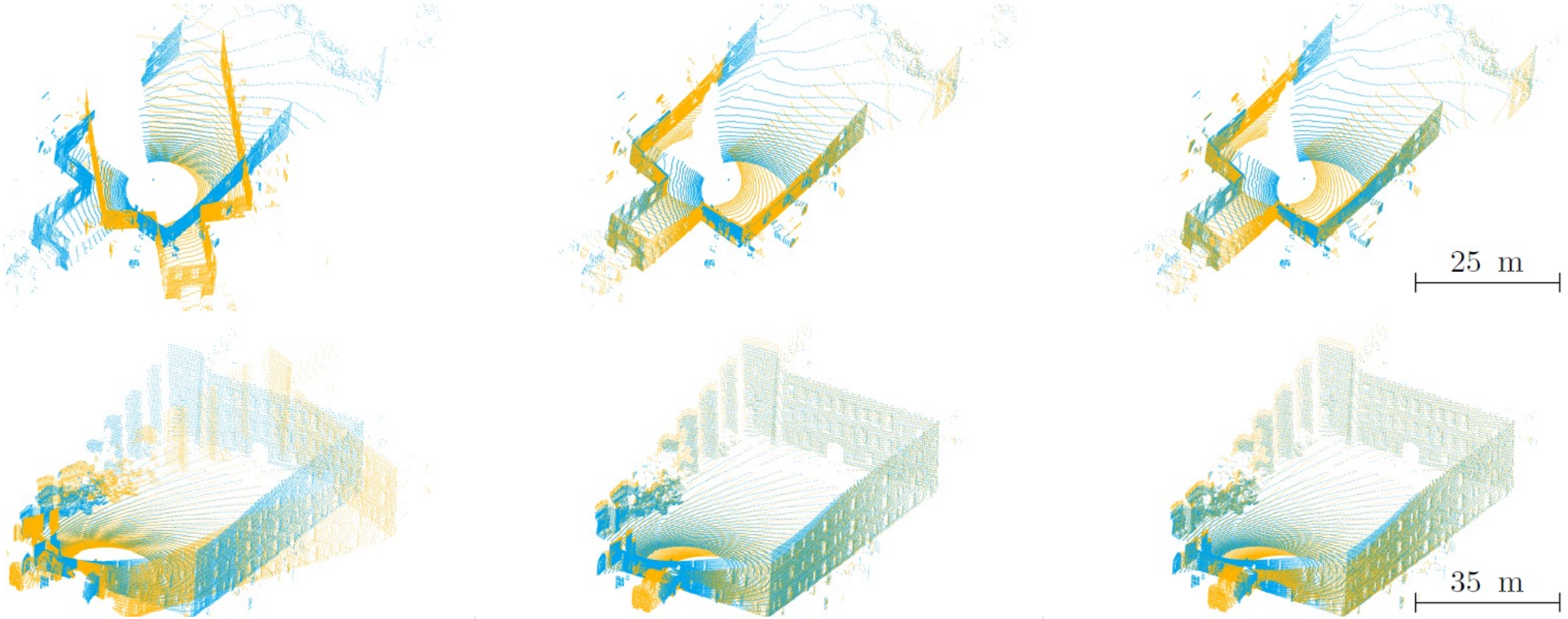
Source and target (input) | BUFFER-X (Ours) | Ground truth

This video includes audio narration

Experimental Results (Cont'd)

Outdoor scenes acquired by 

*which qualitatively demonstrate applicability across **various sensors and diverse scenes***



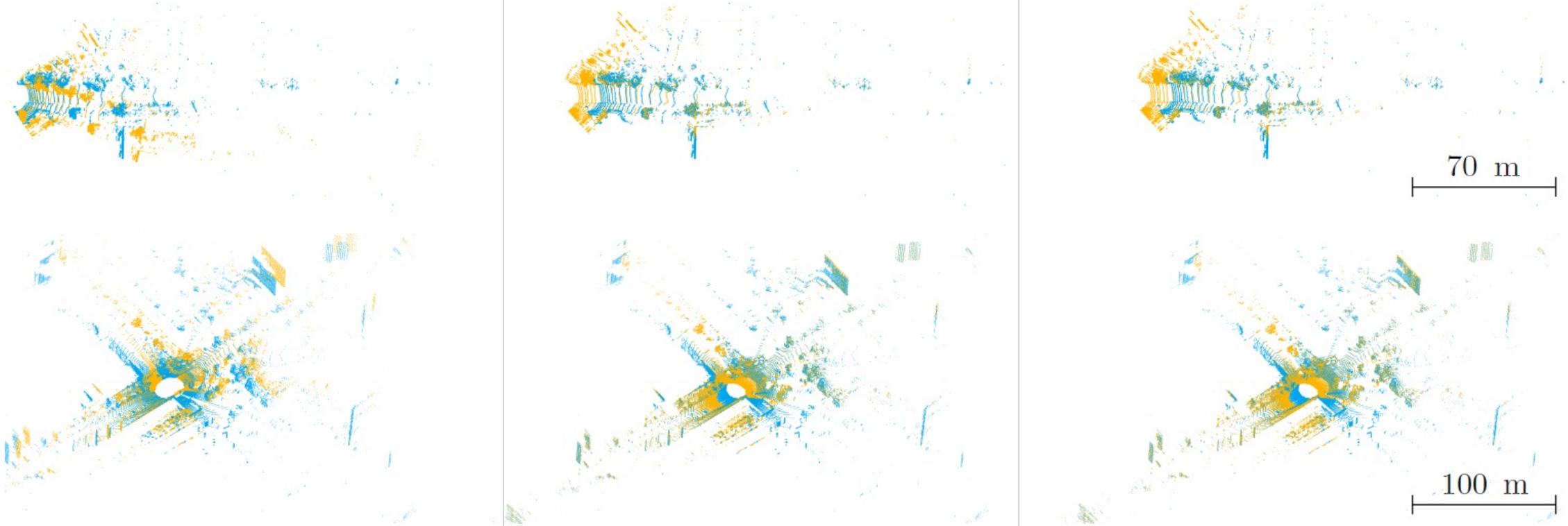
Source and target (input) | BUFFER-X (Ours) | Ground truth

This video includes audio narration

Experimental Results (Cont'd)

Outdoor scenes acquired by 

*which qualitatively demonstrate applicability across **various sensors and diverse scenes***



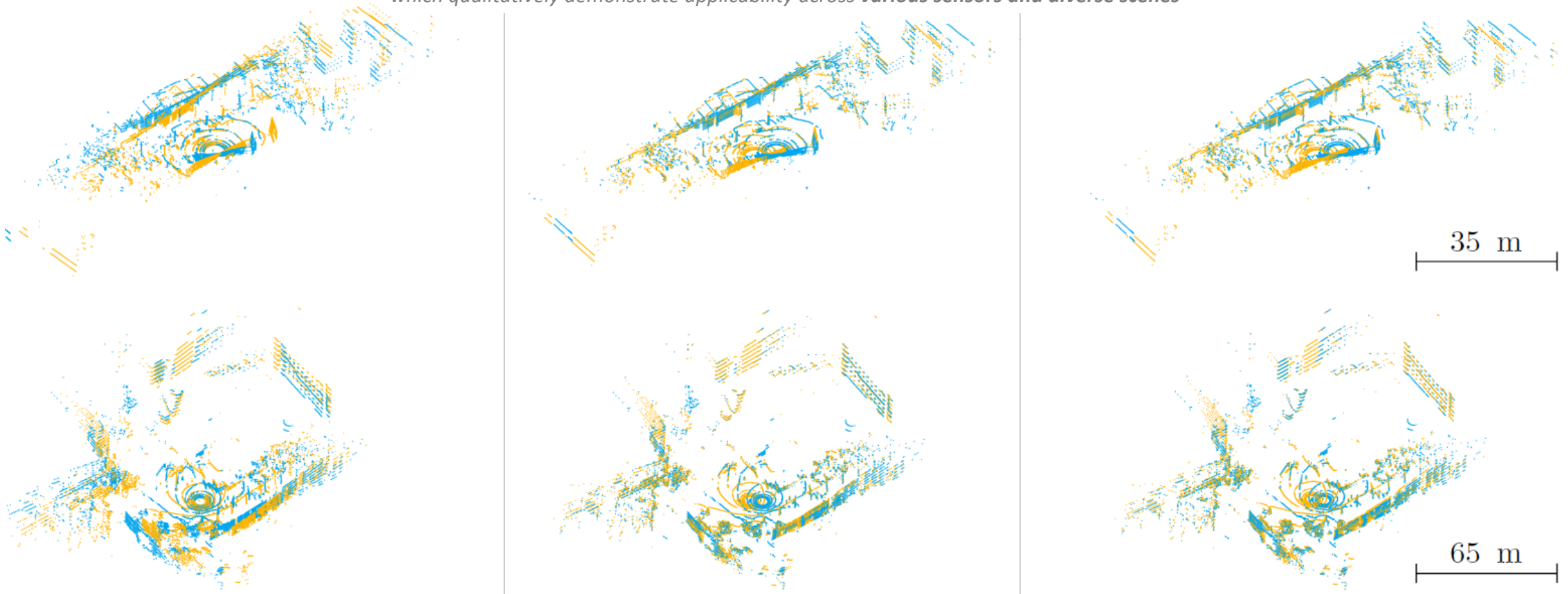
Source and target (input) | BUFFER-X (Ours) | Ground truth

This video includes audio narration

Experimental Results (Cont'd)

Outdoor scenes acquired by 

*which qualitatively demonstrate applicability across **various sensors and diverse scenes***



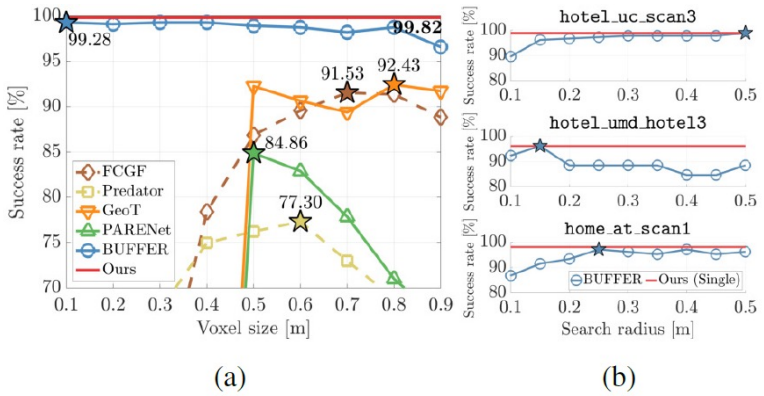
Source and target (input) | BUFFER-X (Ours) | Ground truth

This video includes audio narration

Experimental Results (Cont'd)

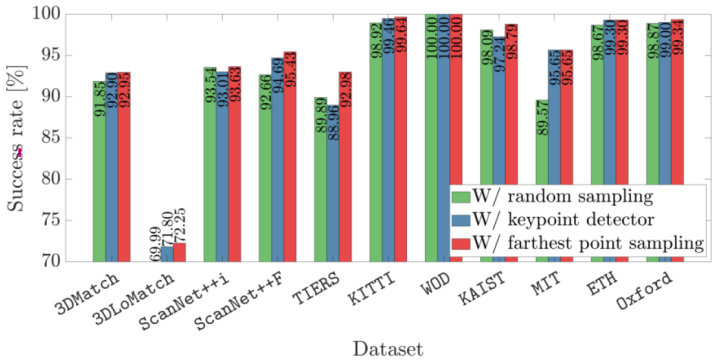
+ Ablation studies are also provided

Impact of geometric bootstrapping



Success rate changes depending on voxel size and search radius

Learning-based keypoint detector vs. Farthest point sampling

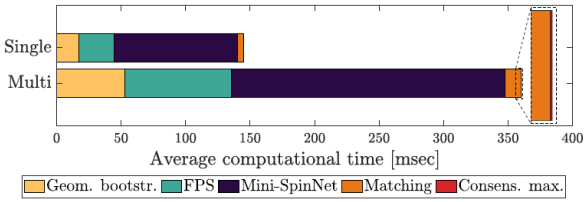


Success rate comparison across keypoint selection strategies

Impact of multi-scale

Local	Middle	Global	RTE [cm] ↓	RRE [°] ↓	Succ. rate [%] ↑	Hz ↑
✓			6.57	2.15	84.06	**5.61**
	✓		5.87	1.85	93.38	5.47
		✓	6.06	1.91	93.57	5.49
✓	✓		**5.73**	1.81	94.31	2.35
✓		✓	5.77	1.81	94.02	2.36
	✓	✓	5.78	1.81	94.62	2.33
✓	✓	✓	5.78	**1.79**	**95.58**	1.81

Ablation study of multi-scale design



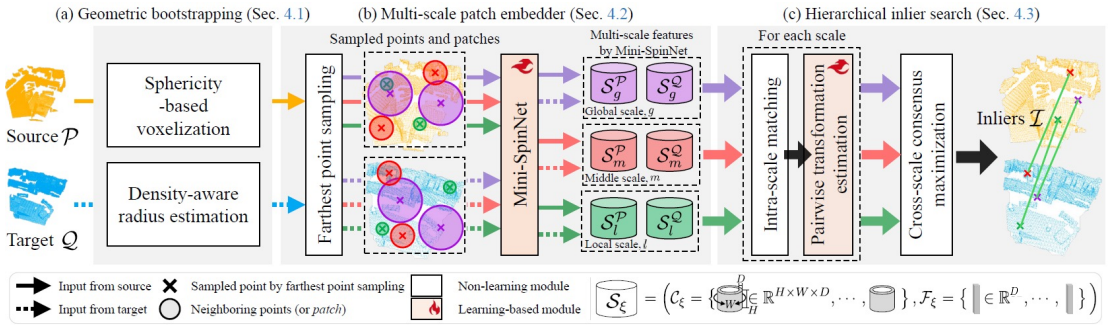
Runtime analysis

This video includes audio narration

Conclusion

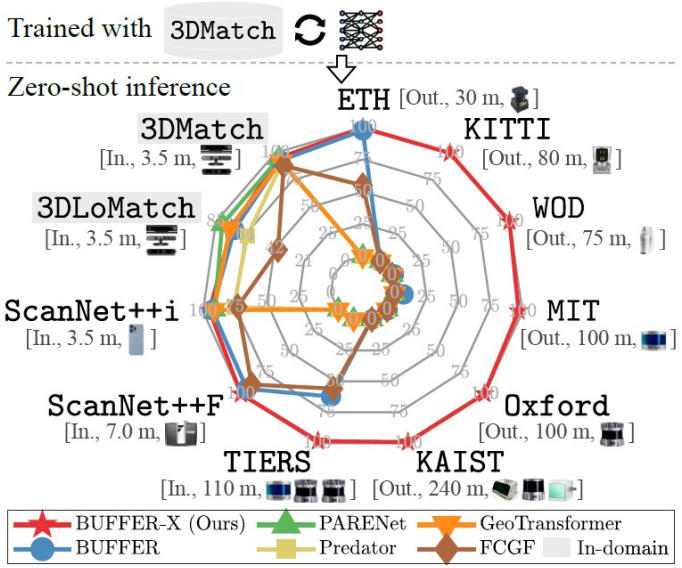
Key contribution #1:

Zero-shot registration pipeline called *BUFFER-X*



Key contribution #2:

Comprehensive benchmark to evaluate generalization capability



This video includes audio narration

Visual & Geometric Intelligence Lab.

SPARKlab

Thank you!
More results in the paper
& code is available

arXiv version paper

GitHub code