



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# ROBUST VISUAL REPRESENTATION ACROSS MODALITIES IN SEMANTIC SCENE UNDERSTANDING

24 March 2025



PHD CANDIDATE: ELENA CAMUFFO

ADVISOR: PROF. SIMONE MILANI

## CLASSIFICATION



animals

# SEMANTIC SCENE UNDERSTANDING

- Understanding **spatial and semantic relationships** between objects in a scene.
- Includes various tasks: **classification**, **object detection**, and **segmentation**.

## OBJECT DETECTION



dog

cat

# SEMANTIC SCENE UNDERSTANDING

- Understanding **spatial and semantic relationships** between objects in a scene.
- Includes various tasks: **classification, object detection, and segmentation.**

## SEMANTIC SEGMENTATION



dog

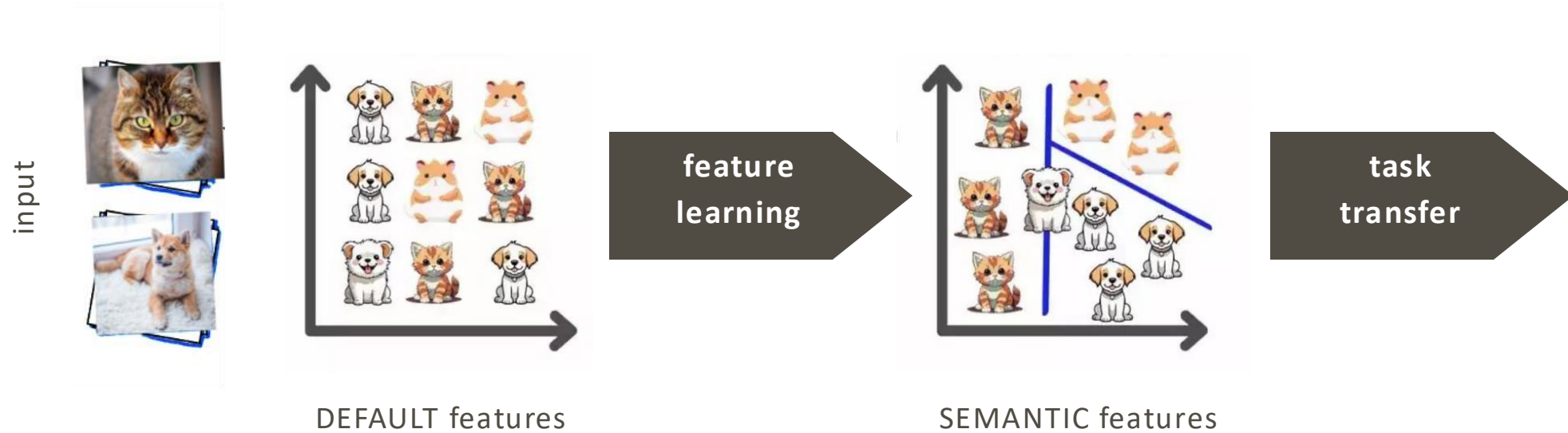
cat

# SEMANTIC SCENE UNDERSTANDING

- Understanding **spatial and semantic relationships** between objects in a scene.
- Includes various tasks: **classification, object detection, and segmentation.**



# ROBUST VISUAL REPRESENTATION



# PROBLEMS



01 **Adapt** to new tasks and integrate data from multiple sensors.



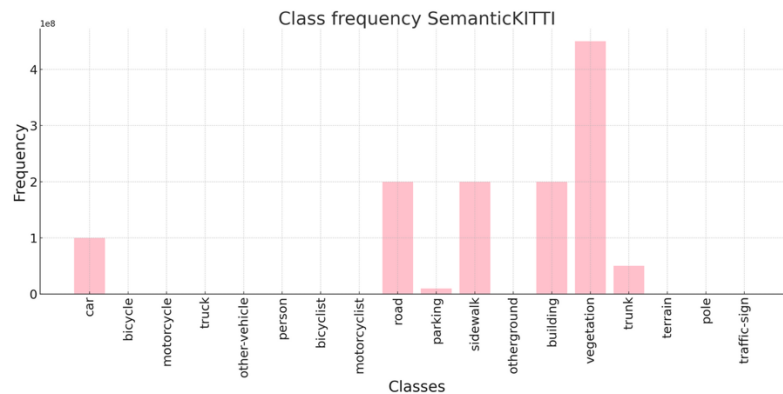
03 **Understand** the environment accurately.



02 **Handle** imbalance datasets, limited training data and corrupted inputs.

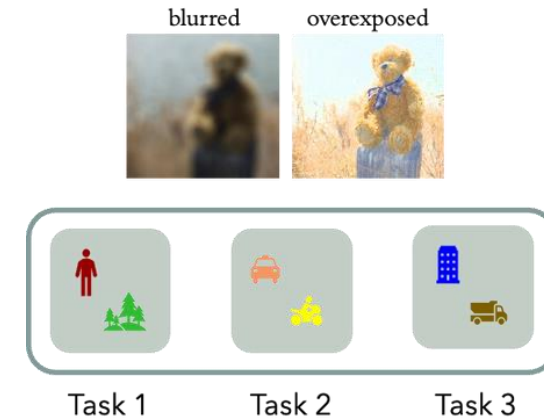
# KEY CHALLENGES

01 **Adapt** to new tasks and integrate data from multiple sensors.



DATA IMBALANCE AND SCARCITY

02 **Handle** imbalance datasets, limited training data and corrupted inputs.



DATA AND LABEL DISTRIBUTION SHIFTS

# INNOVATIONS TO TACKLE CHALLENGES





# INNOVATIONS TO TACKLE CHALLENGES

## DATA IMBALANCE AND SCARCITY

### 1. Class Imbalance in 3D Data

Hierarchical Learning for improved class balance.  
Semantic-guided transmission.

### 4. Few-Shot and 3D Reconstruction

Heritage Point Cloud Instance Collection.  
Improved Scan-to-BIM via Few-Shot learning.

# INNOVATIONS TO TACKLE CHALLENGES

## DATA AND LABEL DISTRIBUTION SHIFTS

### 1. Class Imbalance in 3D Data

### 2. Continual and Multimodal Learning

Enhanced Continual Learning on LiDAR data.

Multimodality for resilient architectures to losses.

### 3. Robustness to Corrupted Data

Detect corrupted images using FFT and a deep network,  
adapt BN layers of scene understanding models.

### 4. Few-Shot and 3D Reconstruction

# INNOVATIONS TO TACKLE CHALLENGES

**1. Class Imbalance  
in 3D Data**

**2. Continual and  
Multimodal Learning**

**3. Robustness to  
Corrupted Data**

**4. Few-Shot and 3D  
Reconstruction**

**5. Conclusions**

Limitations and Future Work.  
Bibliography and Acknowledgments.

# INNOVATIONS TO TACKLE CHALLENGES

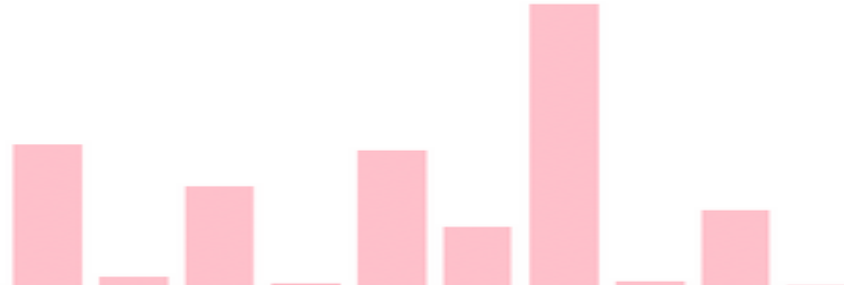
**1. Class Imbalance  
in 3D Data**

**2. Continual and  
Multimodal Learning**

**3. Robustness to  
Corrupted Data**

**4. Few-Shot and 3D  
Reconstruction**

**5. Conclusions**



# 1. Class Imbalance in 3D Data

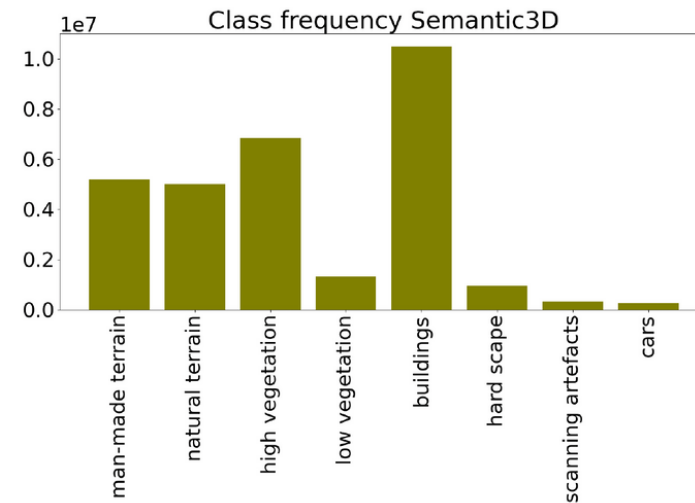
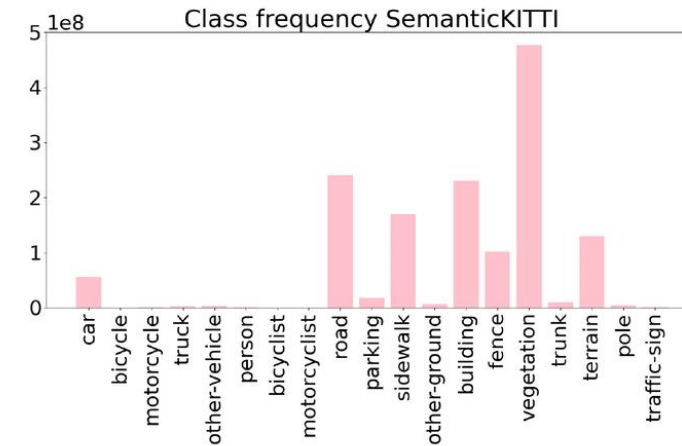
- **LEAK:** Hierarchical learning for better class balance in semantic segmentation [1].
- Improved performance on minority classes and efficient **semantic-guided transmission** [2].

[1] **Camuffo E.**, Michieli U., Milani S. "Learning from Mistakes: Self-Regularizing Hierarchical Semantic Representations in Point Cloud Segmentation", IEEE Transactions on Multimedia, 2023.

[2] Mari D., **Camuffo E.**, Milani S. "CACTUS: Content-Aware Compression and Transmission Using Semantics for Automotive LiDAR data", Sensors, 2023.

# CLASS IMBALANCE IN 3D DATA

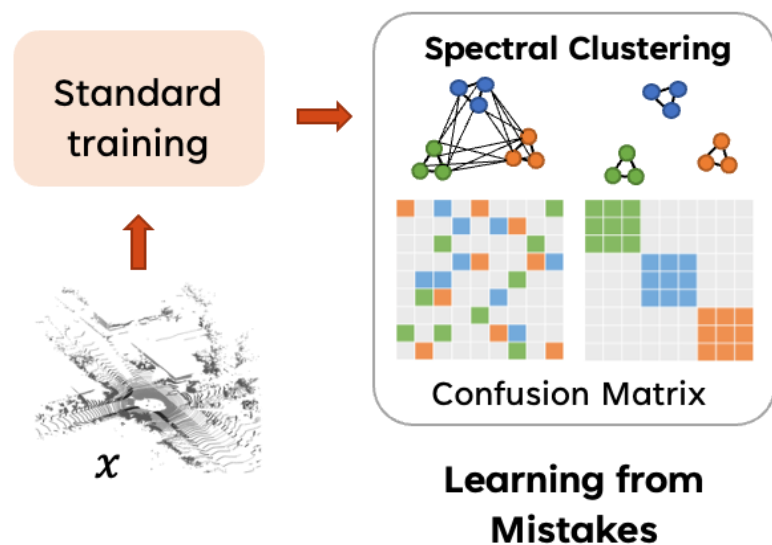
- Point Cloud datasets are usually **not properly balanced**, and classes with less points are often misclassified by semantic understanding models.
- Most of the errors occur within classes that are **sematically similar**.



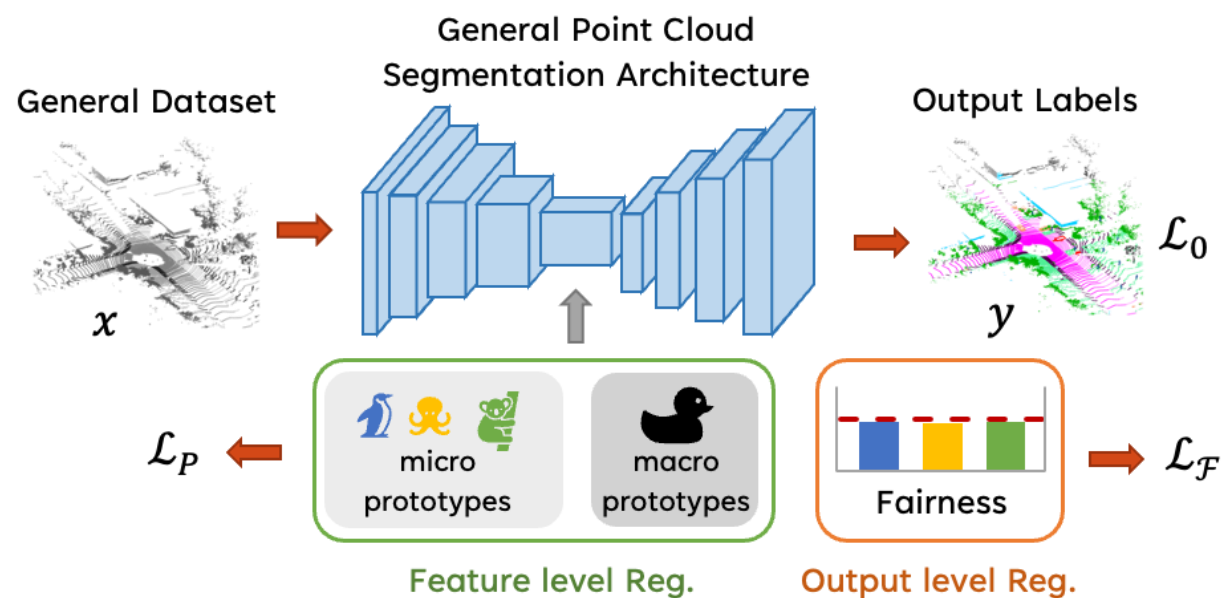


# LEAK ARCHITECTURE

BUILD CLASS HIERARCHIES

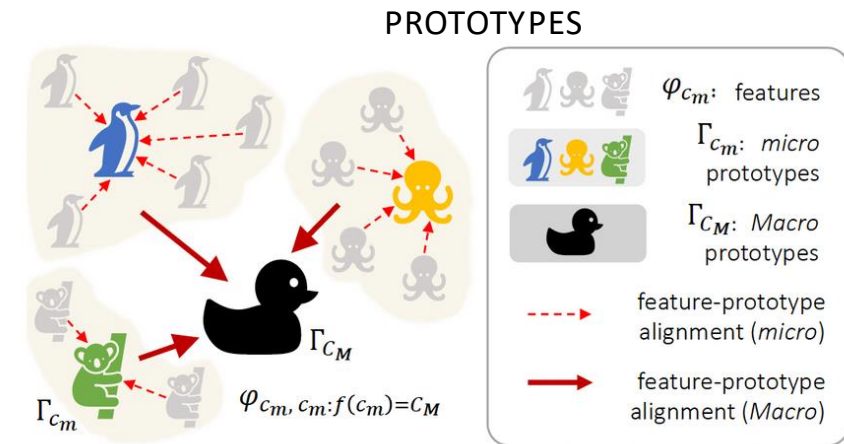
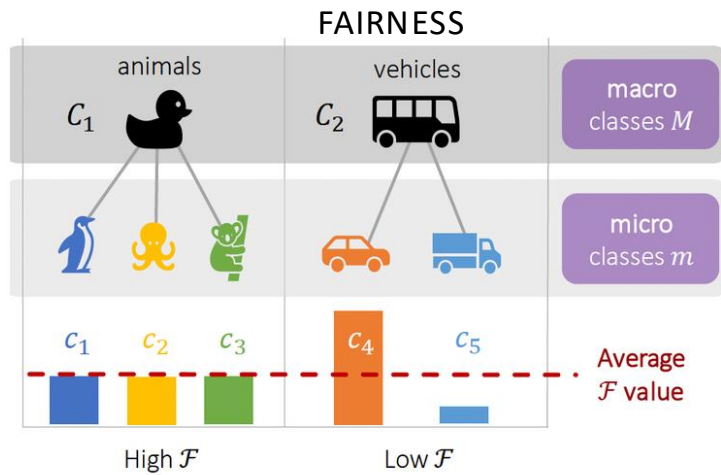


HIERARCHICAL SELF-REGULARIZATION



[1] Camuffo E., Michieli U., Milani S. "Learning from Mistakes: Self-Regularizing Hierarchical Semantic Representations in Point Cloud Segmentation", IEEE Transactions on Multimedia, 2023.

# LEAK COMPONENTS

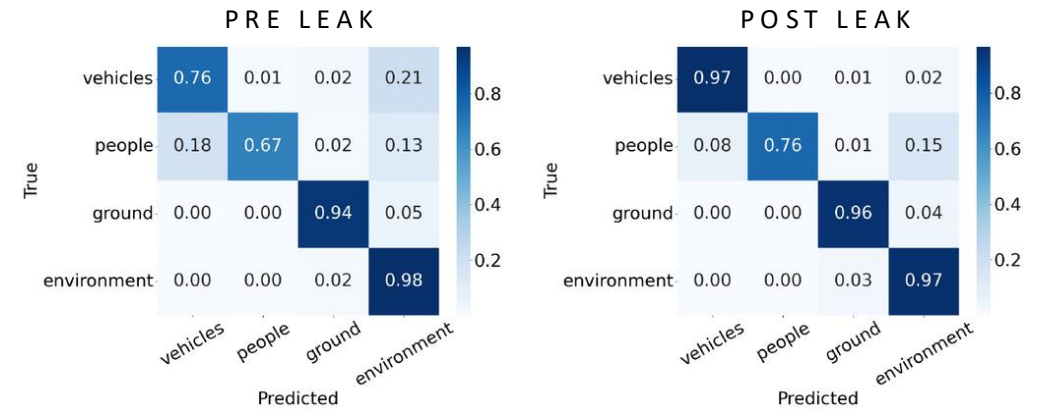
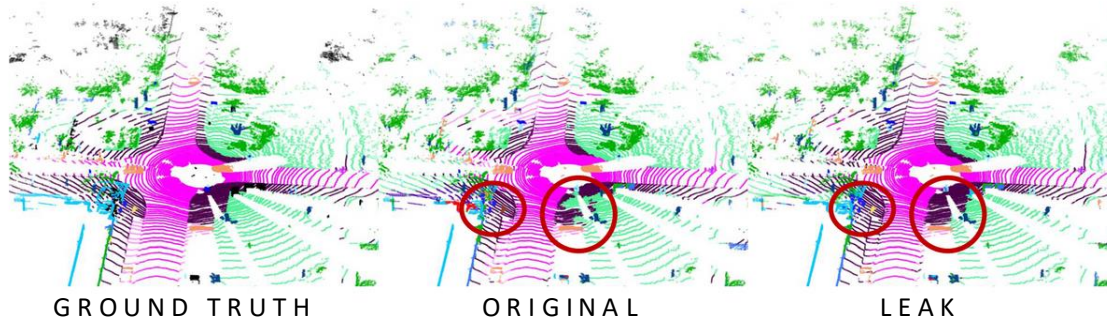


- **Macro-aware Fairness** score on the micro classes promotes homogeneous scores in macro clusters.
- Class-conditional latent **features-to-prototype alignment** at 2 levels (micro and macro) improves class-wise feature discrimination.

$$\mathcal{L}_{LEAK} = \mathcal{L}_0 + \lambda_{P_m} \cdot \mathcal{L}_{P_m} + \lambda_{P_M} \cdot \mathcal{L}_{P_M} + \lambda_{\mathcal{F}} \cdot \mathcal{L}_{\mathcal{F}}$$

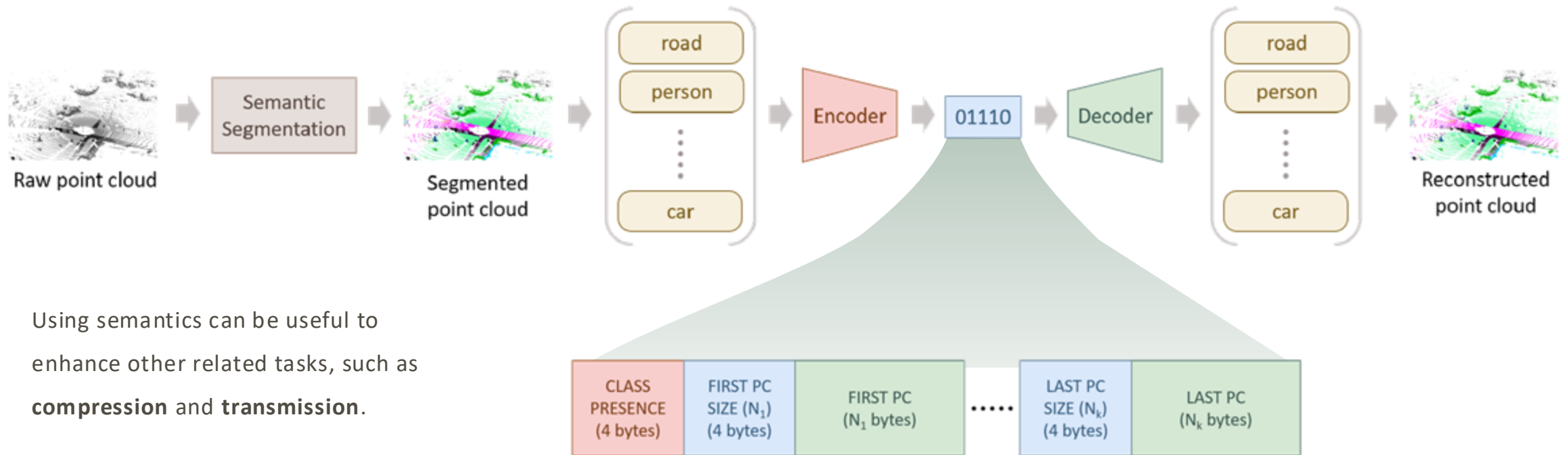
[1] **Camuffo E.**, Michieli U., Milani S. "Learning from Mistakes: Self-Regularizing Hierarchical Semantic Representations in Point Cloud Segmentation", IEEE Transactions on Multimedia, 2023.

# LEAK RESULTS



[1] **Camuffo E.**, Michieli U., Milani S. "Learning from Mistakes: Self-Regularizing Hierarchical Semantic Representations in Point Cloud Segmentation", IEEE Transactions on Multimedia, 2023.

# SEMANTIC GUIDED TRANSMISSION



Using semantics can be useful to enhance other related tasks, such as **compression** and **transmission**.

[2] Mari D., Camuffo E., Milani S. "CACTUS: Content-Aware Compression and Transmission Using Semantics for Automotive LiDAR data", Sensors, 2023.

# INNOVATIONS TO TACKLE CHALLENGES

**1. Class Imbalance  
in 3D Data**

**2. Continual and  
Multimodal Learning**

**3. Robustness to  
Corrupted Data**

**4. Few-Shot and 3D  
Reconstruction**

**5. Conclusions**



## 2. Continual and Multimodal Learning

- Enhanced **Continual Learning** on LiDAR data via knowledge distillation and self-inpainting [3].
- Robust **Multimodal** (LiDAR + RGB) architecture, resilient to sensory loss and malfunctioning [4].

[3] Camuffo E., Milani S., “Continual Learning for LiDAR Semantic Segmentation: Class-Incremental and Coarse-to-Fine strategies on Sparse Data”, CVPRW, 2023.

[4] Barbato F., Camuffo E., Milani S., Zanuttigh P., “Multi-Modal Continual Learning for Semantic Segmentation”, ICIP, 2024.



# CONTINUAL AND MULTIMODAL LEARNING



## CONTINUAL LEARNING

Learning **continuously and adaptively**, enabling the autonomous incremental development of ever more complex skills and knowledge.



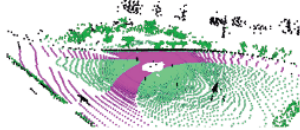
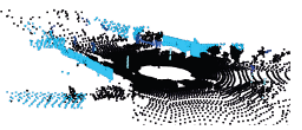

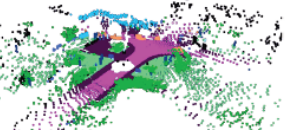
## MULTIMODAL LEARNING

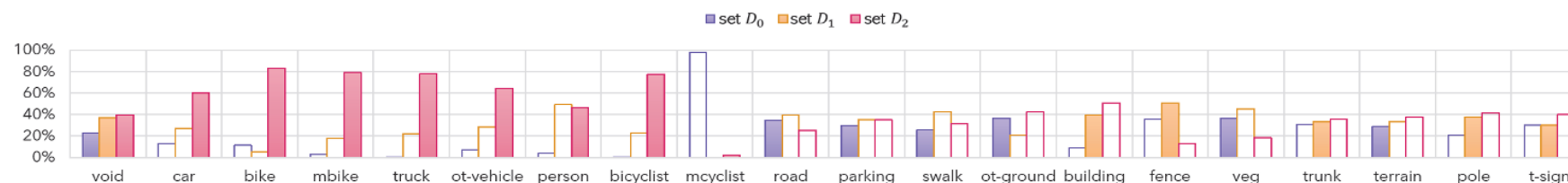
Inclusion of **heterogeneous data** to learn more information about the scene 2D and 3D information together.

[3] **Camuffo E.**, Milani S., “Continual Learning for LiDAR Semantic Segmentation: Class-Incremental and Coarse-to-Fine strategies on Sparse Data”, CVPRW, 2023.

[4] Barbato F., **Camuffo E.**, Milani S., Zanuttigh P., “Multi-Modal Continual Learning for Semantic Segmentation”, ICIP, 2024.

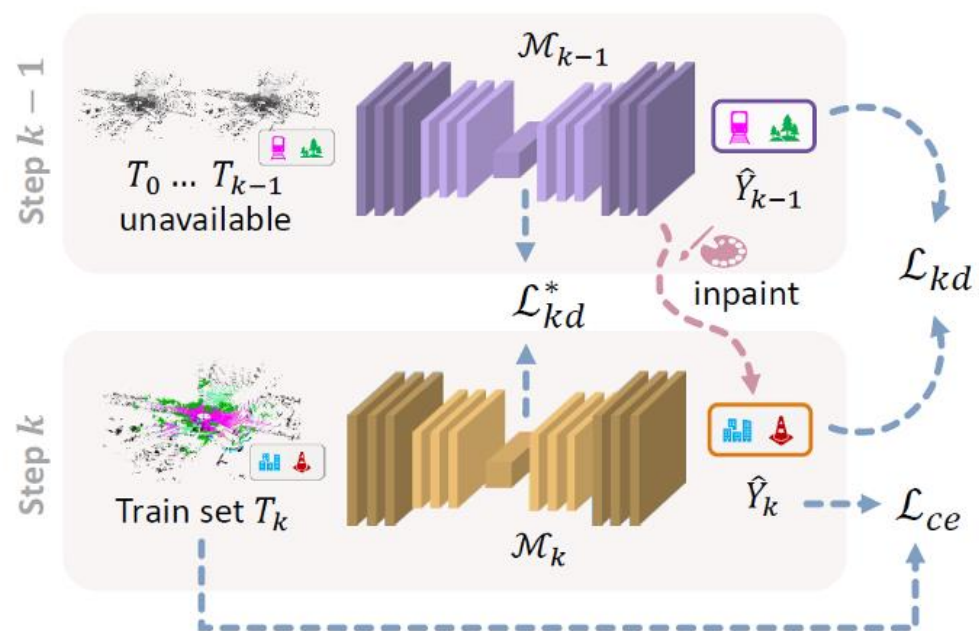
# DISCOVERY OF NOVEL SEMANTIC CATEGORIES

	Training Subset $T_0$	Training Subset $T_1$	Training Subset $T_2$	Validation Set
	SemanticKITTI train			SemanticKITTI val
Sequences	$D_0 = \{01, 02, 03\}$	$D_1 = \{04, 05, 09, 10\}$	$D_2 = \{00, 06, 07\}$	$D = \{08\}$
Labeled classes	$C_0 = \{\text{road, parking, sidewalk, other-ground, vegetation, terrain}\}$	$C_1 = \{\text{building, fence, trunk, pole, traffic-sign}\}$	$C_2 = \{\text{bicycle, motorcycle, truck, other-vehicle, person, bicyclist, motorcyclist, car}\}$	$C = \{C_0 \cup C_1 \cup C_2\}$
				
# Clouds	6563	4623	4541	4071
# Points	355280	375140	121494	19130
% Labeled pt.	81.73%	19.21%	8.82%	95.53%



[3] **Camuffo E.**, Milani S., "Continual Learning for LiDAR Semantic Segmentation: Class-Incremental and Coarse-to-Fine strategies on Sparse Data", CVPRW, 2023.

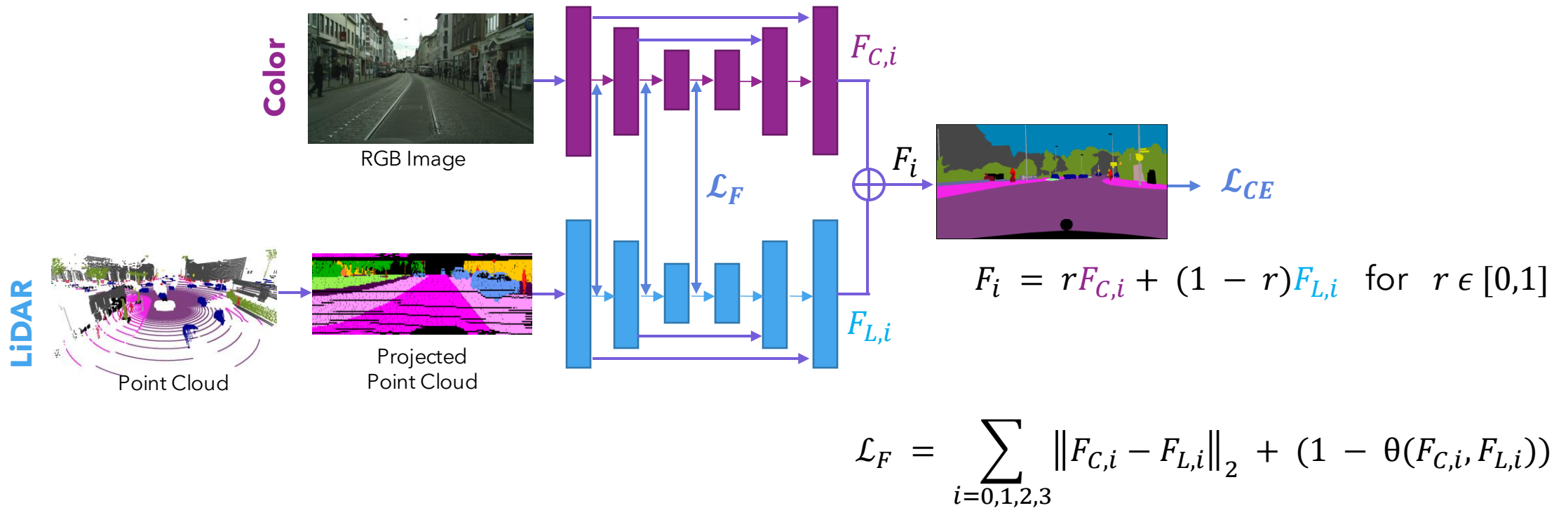
# DISCOVERY OF NOVEL SEMANTIC CATEGORIES



- **Feature Level Distillation ( $\mathcal{L}_{kd}^*$ ):**  $L_p$  ( $L_1$  or  $L_2$ ) norm between the features of current and previous models.
- **Output Level Distillation ( $\mathcal{L}_{kd}$ ):** additional Cross-Entropy, where labels for past classes are obtained from previous model predictions.
- **Background Self-Inpainting:** labels for old classes  $\mathcal{C}_{k-1}$  are predicted by previous model  $\mathcal{M}_{k-1}$  and inpainted onto background labels of current training set  $T_k$ .

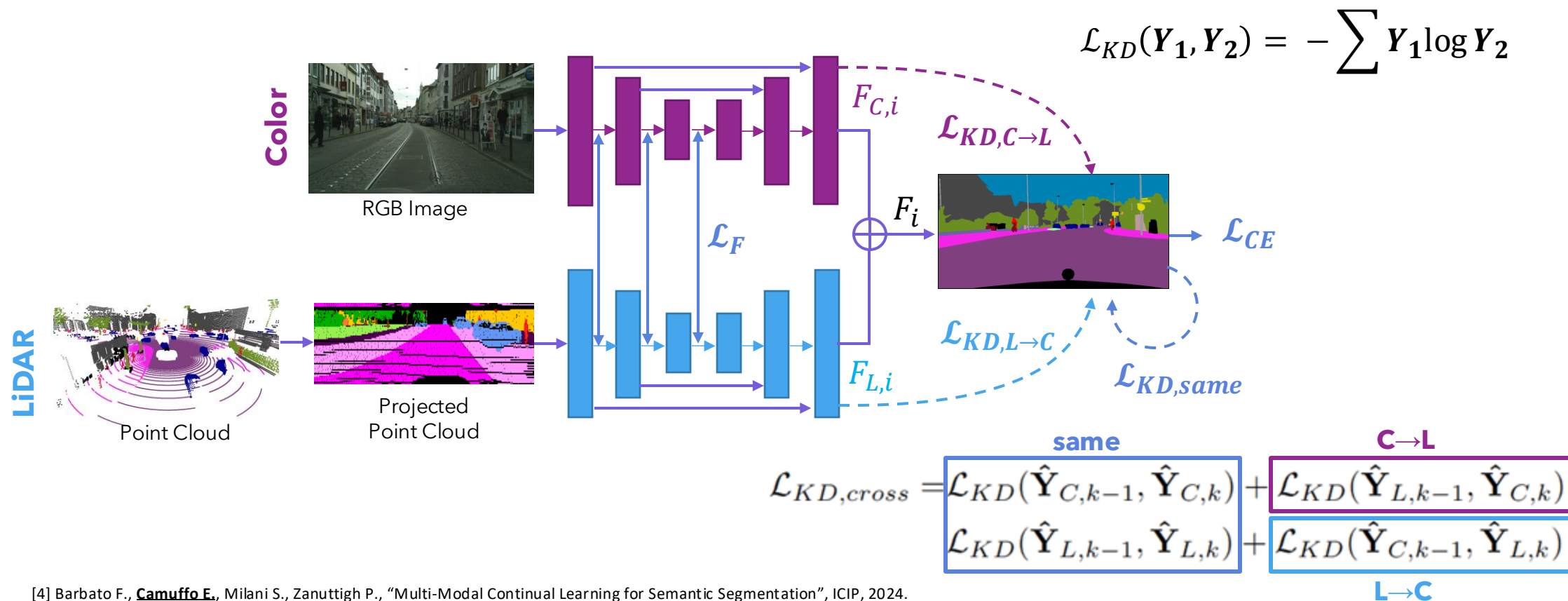
[3] Camuffo E., Milani S., "Continual Learning for LiDAR Semantic Segmentation: Class-Incremental and Coarse-to-Fine strategies on Sparse Data", CVPRW, 2023.

# MULTIMODAL DATA INTEGRATION



[4] Barbato F., **Camuffo E.**, Milani S., Zanuttigh P., “Multi-Modal Continual Learning for Semantic Segmentation”, ICIP, 2024.

# MULTIMODAL DATA INTEGRATION



[4] Barbato F., **Camuffo E.**, Milani S., Zanuttigh P., "Multi-Modal Continual Learning for Semantic Segmentation", ICIIP, 2024.

# INNOVATIONS TO TACKLE CHALLENGES

1. Class Imbalance  
in 3D Data

2. Continual and  
Multimodal Learning

3. Robustness to  
Corrupted Data

4. Few-Shot and 3D  
Reconstruction

5. Conclusions





### 3. Robustness to Corrupted Data

- FROST and PAN frameworks to detect corruptions using FFT and deep models and mitigate corruptions via adaptation of BN layers [5,6,7].

[5] **Camuffo E.**, Michieli U., Moon J., Kim D., Ozay M., “FFT-based Selection and Optimization of Statistics for Robust Recognition of Severely Corrupted Images”, ICASSP, 2024.

[6] **Camuffo E.**, Michieli U., Milani S., Moon J., Ozay M., “Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics”, IROS, 2024.

[7] Michieli U., Ozay M., Moon J., Kim D., **Camuffo E.**, “Performing a Computer Vision Task”, US Patent App. 18/933,406, 2025.

# ROBUSTNESS TO CORRUPTED DATA



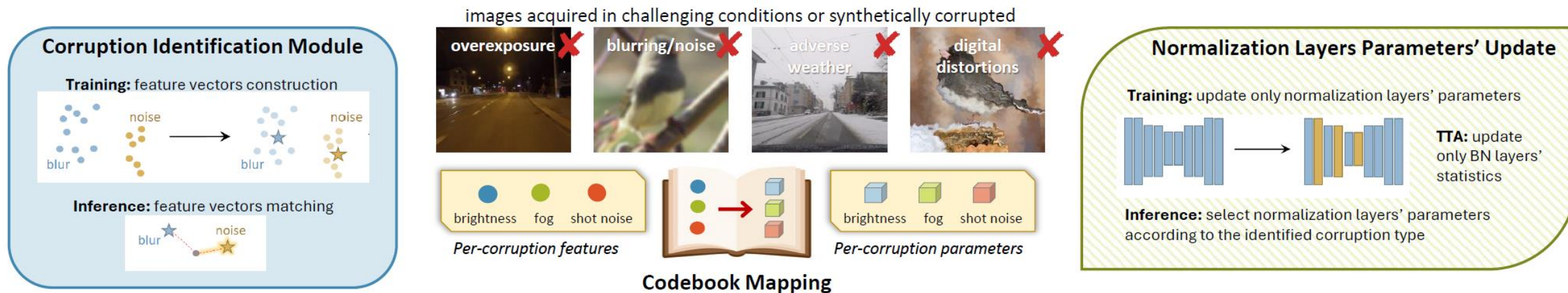
- Scene understanding models suffer corrupted images
- **Real** vs **Synthetic** Corruptions (approximations)

$$\tilde{x} = \underset{\text{clean image}}{x} + \underset{\text{corruption}}{\psi(x)} \quad \text{Corrupted Image}$$

- $\psi(\cdot)$ : corruption operator – introduces **noise**, shift

[5] **Camuffo E.**, Michieli U., Moon J., Kim D., Ozay M., “FFT-based Selection and Optimization of Statistics for Robust Recognition of Severely Corrupted Images”, ICASSP, 2024.  
[6] **Camuffo E.**, Michieli U., Milani S., Moon J., Ozay M., “Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics”, IROS, 2024.  
[7] Michieli U., Ozay M., Moon J., Kim D., **Camuffo E.**, “Performing a Computer Vision Task”, US Patent App. 18/933,406, 2025.

# HANDLE CORRUPTED IMAGES

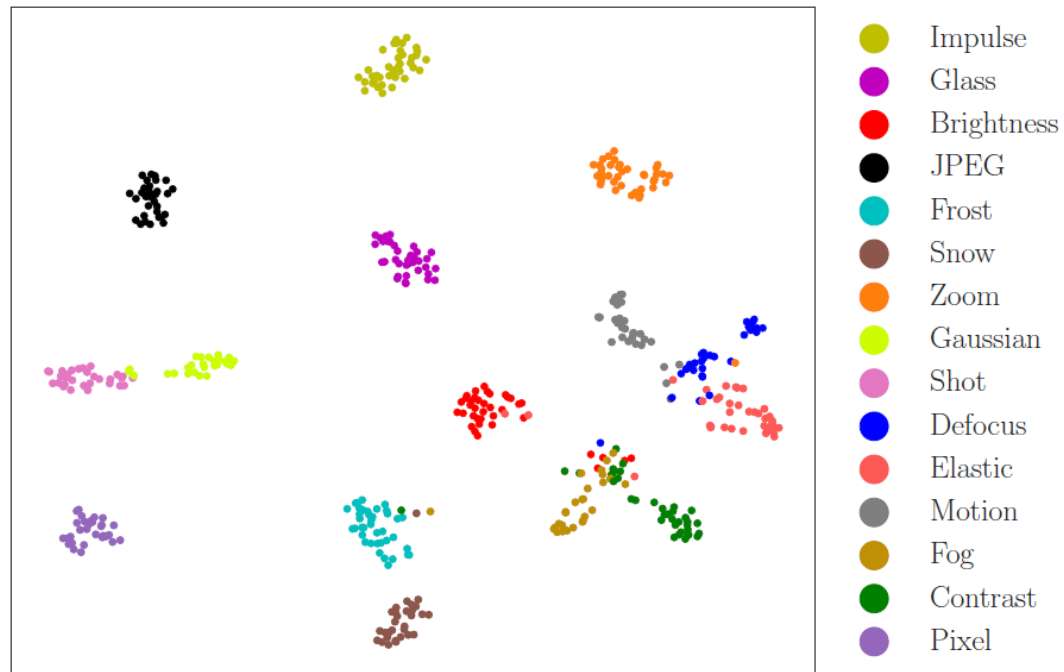


1. **FROST:** FFT features
2. **PAN:** CIM network

1. **FROST:** trained BN layers
2. **PAN:** adapted BN layers

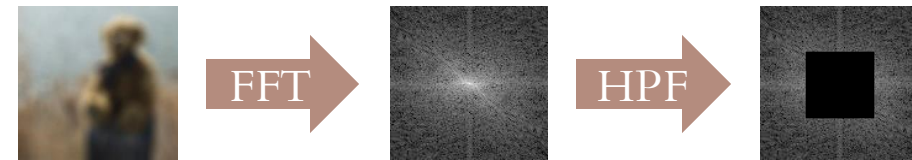
[5] **Camuffo E.**, Michieli U., Moon J., Kim D., Ozay M., "FFT-based Selection and Optimization of Statistics for Robust Recognition of Severely Corrupted Images", ICASSP, 2024.  
 [6] **Camuffo E.**, Michieli U., Milani S., Moon J., Ozay M., "Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics", IROS, 2024.  
 [7] Michieli U., Ozay M., Moon J., Kim D., **Camuffo E.**, "Performing a Computer Vision Task", US Patent App. 18/933,406, 2025.

# CORRUPTION IDENTIFICATION

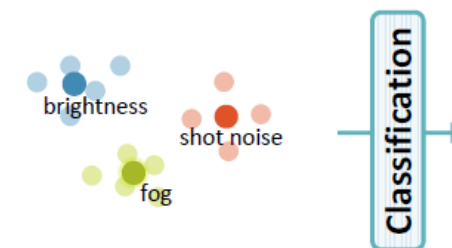


TSNE OF CORRUPTIONS FEATURES (PAN)

**FROST:** Uses High-Frequency FFT amplitudes.



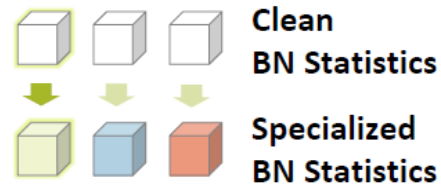
**PAN:** Uses a Deep Convolutional Network.



- [5] **Camuffo E.**, Michieli U., Moon J., Kim D., Ozay M., “FFT-based Selection and Optimization of Statistics for Robust Recognition of Severely Corrupted Images”, ICASSP, 2024.  
[6] **Camuffo E.**, Michieli U., Milani S., Moon J., Ozay M., “Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics”, IROS, 2024.  
[7] Michieli U., Ozay M., Moon J., Kim D., **Camuffo E.**, “Performing a Computer Vision Task”, US Patent App. 18/933,406, 2025.

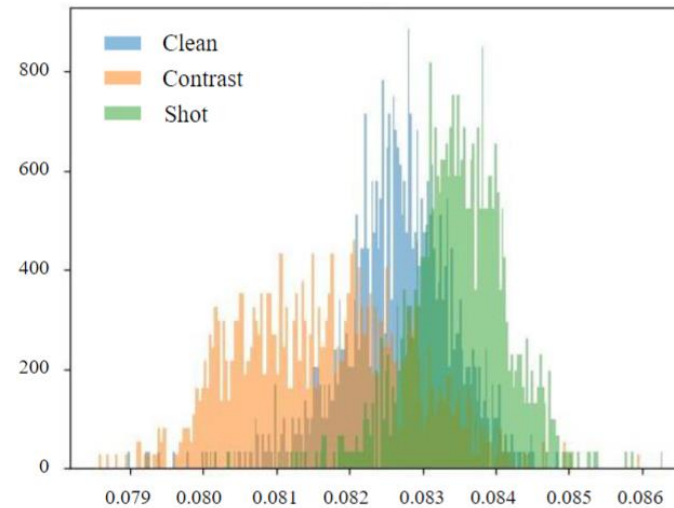
# NORMALIZATION LAYERS' ADAPTATION

**FROST:** Train normalization layers' params for each corruption type separately and use them at inference time.

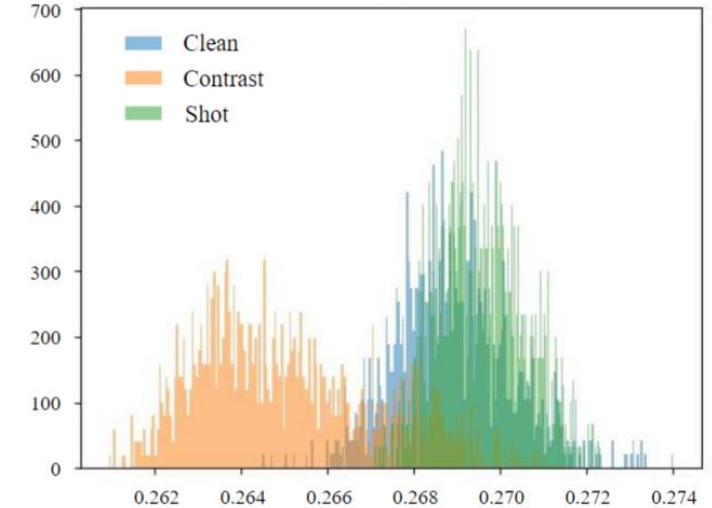


**PAN:** Adjust normalization layers for each corruption type at inference time.

AVERAGE



VARIANCE



They respond differently depending on the corruption type.

- [5] **Camuffo E.**, Michieli U., Moon J., Kim D., Ozay M., "FFT-based Selection and Optimization of Statistics for Robust Recognition of Severely Corrupted Images", ICASSP, 2024.
- [6] **Camuffo E.**, Michieli U., Milani S., Moon J., Ozay M., "Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics", IROS, 2024.
- [7] Michieli U., Ozay M., Moon J., Kim D., **Camuffo E.**, "Performing a Computer Vision Task", US Patent App. 18/933,406, 2025.

# RESULTS

PAN RESULTS	Model	Backbone	Dataset	Task	Real	Src	PAN	Gain (%)	Metric	Model (MB)	BN (MB)
	ResNet18 [110]	-	ImageNet-C [113]	OR		31.7	39.0	23.0	CA ↑	44.6	0.04
	ResNet50 [110]	-	ImageNet-C [113]	OR		46.1	47.7	3.5	CA ↑	97.7	0.20
	ResNet101 [110]	-	ImageNet-C [113]	OR		53.0	55.5	4.7	CA ↑	170.3	0.40
	MobileNetV3 [121]	-	ImageNet-C [113]	OR		32.9	34.4	4.6	CA ↑	9.7	0.05
	ResNeXt50 [337]	-	ImageNet-C [113]	OR		49.6	51.3	3.4	CA ↑	95.7	0.26
	Wide-ResNet50 [346]	-	ImageNet-C [113]	OR		49.0	50.2	2.4	CA ↑	263.0	0.26
	ResNet50 [110]	-	VizWiz [13]	OR	✓	39.1	43.8	12.0	CA ↑	97.7	0.20
	ResNet50 [110]	-	OpenLORIS [270]	OR	✓	42.5	43.8	3.1	CA ↑	97.7	0.20
	YOLOv8n [142]	CSPNet [315]	VOC-C [197]	OD		34.6	36.3	4.9	mAP <sup>50-95</sup> ↑	12.1	0.04
	YOLOv8n [142]	CSPNet [315]	ExDARK [176]	OD	✓	39.4	40.3	2.3	mAP <sup>50-95</sup> ↑	12.1	0.04
	DeepLabV2 [328]	MobileNetV2 [266]	Cityscapes-C [197]	SS		34.5	41.5	20.3	mIoU ↑	42.2	0.11
	DeepLabV2 [328]	MobileNetV2 [266]	ACDC [262]	SS	✓	37.8	40.1	6.1	mIoU ↑	42.2	0.11

- Applicable on both **synthetic** and **real** corruptions.
- Applicable to different models, datasets and tasks.

[5] **Camuffo E.**, Michieli U., Moon J., Kim D., Ozay M., “FFT-based Selection and Optimization of Statistics for Robust Recognition of Severely Corrupted Images”, ICASSP, 2024.

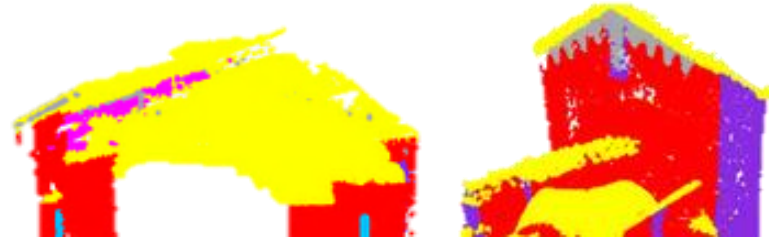
[6] **Camuffo E.**, Michieli U., Milani S., Moon J., Ozay M., “Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics”, IROS, 2024.

[7] Michieli U., Ozay M., Moon J., Kim D., **Camuffo E.**, “Performing a Computer Vision Task”, US Patent App. 18/933,406, 2025.



# INNOVATIONS TO TACKLE CHALLENGES



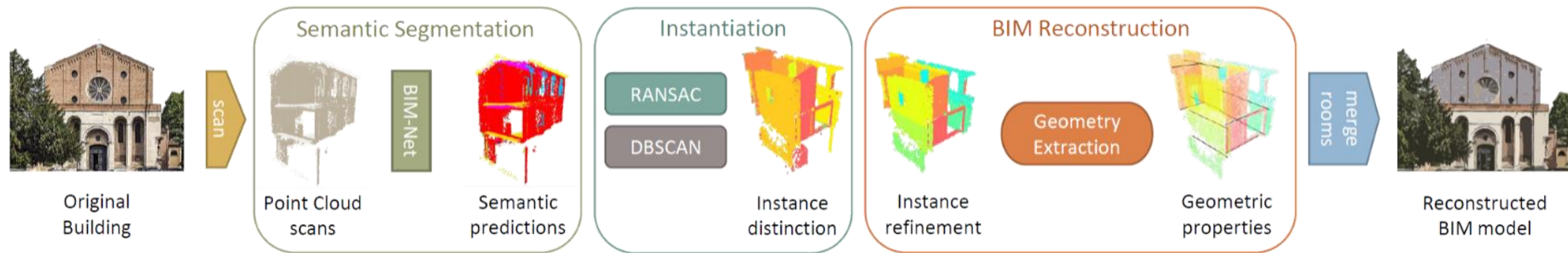


## 4. Few-Shot and 3D Reconstruction

- **HePIC** Dataset: heritage point cloud instance collection of large-scale buildings.
- **Few-shot** learning for Scan-to-BIM instances reconstruction [8].

[8] Campagnolo D., **Camuffo E.**, Michieli U., Borin P., Milani S., Giordano A., “Fully-Automated Scan-to-BIM via Point Cloud Instance Segmentation”, ICIP, 2023.

# SCAN-TO-BIM RECONSTRUCTION



## FULLY AUTOMATED SCAN-TO-BIM PIPELINE

- **Heritage Point cloud Instance Collection (HePIC).**
- **Novel ad hoc deep network (BIM-Net++)**
- **Novel model pre-training and class re-weighting.**

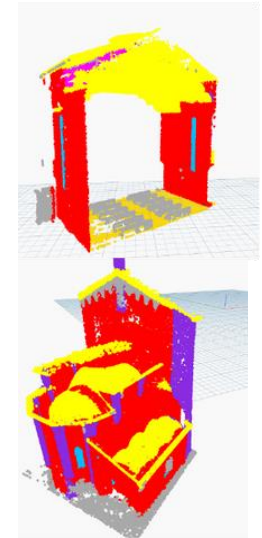
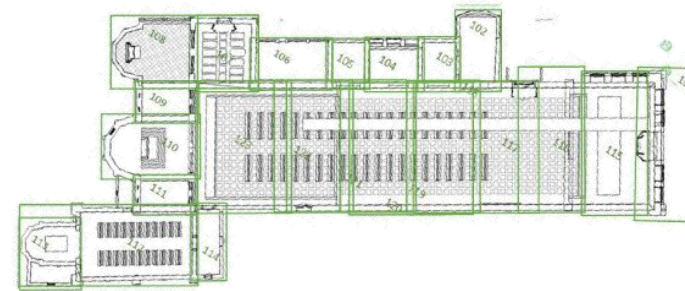
[8] Campagnolo D., **Camuffo E.**, Michieli U., Borin P., Milani S., Giordano A., "Fully-Automated Scan-to-BIM via Point Cloud Instance Segmentation", ICIP, 2023.

# HEPIC DATASET

Acquired via terrestrial laser scanning and labeled using BIM models manually generated.

91 rooms, each composed of 100k points with both semantic and instance labels.

Class	# items			total # points		
	Church	Castle	Tot	Church	Castle	Tot
<i>unassigned</i>	23	66	89	216132	296201	512333
<i>beams</i>	147	1809	1956	6393	415069	421462
<i>columns</i>	18	238	256	18951	74588	93539
<i>doors</i>	27	265	292	13962	133257	147219
<i>floors</i>	83	314	397	146516	745314	891830
<i>roofs</i>	218	535	753	564371	632383	1196754
<i>stairs</i>	16	104	120	2394	77089	79483
<i>walls</i>	417	1141	1558	1229115	4229447	5458562
<i>windows</i>	189	81	270	62056	46512	108568
<b>Tot</b>	<b>1250</b>	<b>4441</b>	<b>5691</b>	<b>2299983</b>	<b>6609767</b>	<b>8909750</b>

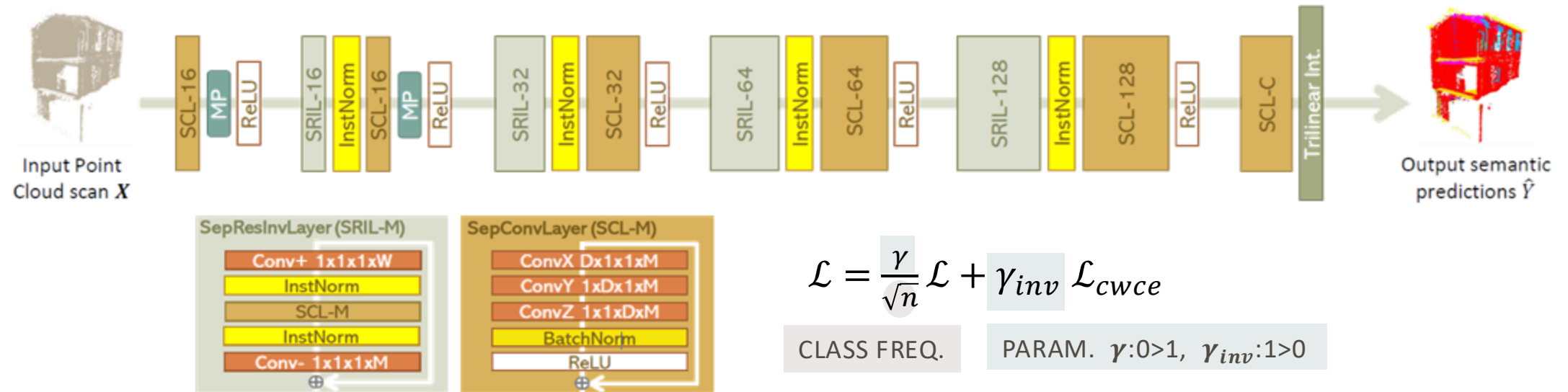


FLOORMAP AND SAMPLES  
FROM EREMITANI CHURCH

[8] Campagnolo D., **Camuffo E.**, Michieli U., Borin P., Milani S., Giordano A., "Fully-Automated Scan-to-BIM via Point Cloud Instance Segmentation", ICIP, 2023.

# FEW-SHOT AND REWEIGHTING SCHEMES

Ad hoc lightweight Network (BIM-Net) using **few-shot learning** (BIM-Net++) and loss weighting schemes.

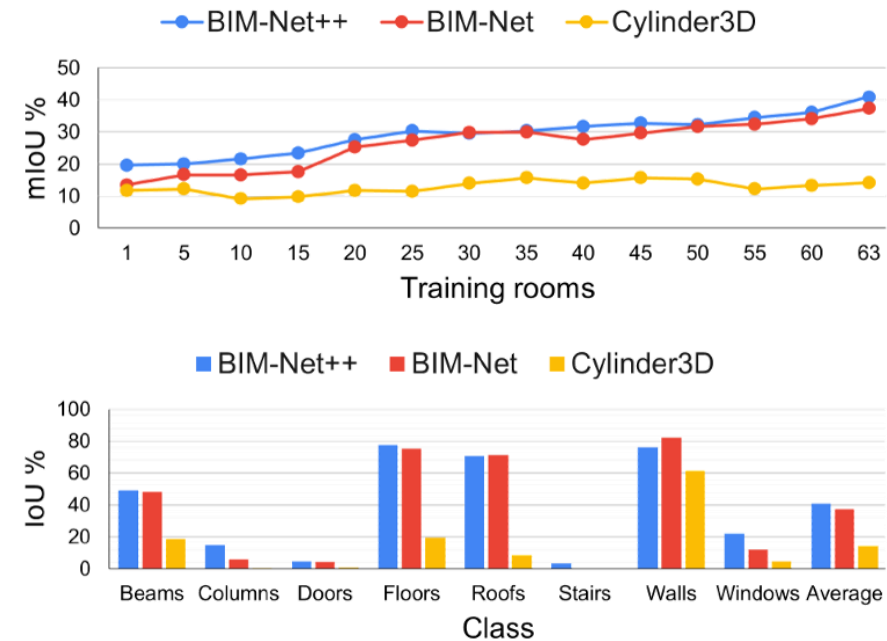


[8] Campagnolo D., **Camuffo E.**, Michieli U., Borin P., Milani S., Giordano A., "Fully-Automated Scan-to-BIM via Point Cloud Instance Segmentation", ICIP, 2023.

# SCAN-TO-BIM RESULTS

model	5k steps			25k steps		
	PA %	PP%	IoU%	PA %	PP%	IoU%
SegCloud [8]	17.6	24.7	13.2	17.6	24.7	13.2
Cylinder3D [7]	21.0	23.2	14.2	21.0	23.2	14.2
RandLA-Net [9]	35.4	49.3	27.8	35.6	56.2	28.8
PVCNN [11]	40.7	45.3	32.9	43.3	48.1	34.9
BIM-Net	44.0	<b>54.4</b>	37.3	47.1	<b>58.9</b>	40.6
<b>BIM-Net++</b>	<b>56.2</b>	49.4	<b>40.9</b>	<b>59.1</b>	53.0	<b>43.7</b>

## COMPARISON WITH OTHER STATE-OF-THE-ART ARCHITECTURES



[8] Campagnolo D., **Camuffo E.**, Michieli U., Borin P., Milani S., Giordano A., "Fully-Automated Scan-to-BIM via Point Cloud Instance Segmentation", ICIP, 2023.

# INNOVATIONS TO TACKLE CHALLENGES



## 5. Conclusions

[8] Campagnolo D., **Camuffo E.**, Michieli U., Borin P., Milani S., Giordano A., “Fully-Automated Scan-to-BIM via Point Cloud Instance Segmentation”, ICIP, 2023.



# LIMITATIONS AND FUTURE WORK: SEMANTICS FOR NOVEL REPRESENTATIONS

# CONCLUSION

# BIBLIOGRAPHY



- [1] Camuffo E., Michieli U., Milani S. “Learning from Mistakes: Self-Regularizing Hierarchical Semantic Representations in Point Cloud Segmentation”, IEEE Transactions on Multimedia, 2023.
- [2] Mari D., Camuffo E., Milani S. “CACTUS: Content-Aware Compression and Transmission Using Semantics for Automotive LiDAR data”, Sensors, 2023.
- [3] Camuffo E., Milani S., “Continual Learning for LiDAR Semantic Segmentation: Class-Incremental and Coarse-to-Fine strategies on Sparse Data”, CVPRW, 2023.
- [4] Barbato F., Camuffo E., Milani S., Zanuttigh P., “Multi-Modal Continual Learning for Semantic Segmentation”, ICIP, 2024.
- [5] Camuffo E., Michieli U., Moon J., Kim D., Ozay M., “FFT-based Selection and Optimization of Statistics for Robust Recognition of Severely Corrupted Images”, ICASSP, 2024.
- [6] Camuffo E., Michieli U., Milani S., Moon J., Ozay M., “Enhanced Model Robustness to Input Corruptions by Per-corruption Adaptation of Normalization Statistics”, IROS, 2024.
- [7] Michieli U., Ozay M., Moon J., Kim D., Camuffo E., “Performing a Computer Vision Task”, US Patent App. 18/933,406, 2025.
- [8] Campagnolo D., Camuffo E., Michieli U., Borin P., Milani S., Giordano A., “Fully-Automated Scan-to-BIM via Point Cloud Instance Segmentation”, ICIP, 2023.
- [9] Camuffo E., Battisti F., Pham F., Milani S. “3D Model Optimization for Immersive and Interactive Applications”, EUVIP, 2022.
- [10] Camuffo E., Mari D., Milani S., “Recent advancements in learning algorithms for point clouds: An updated overview”, Sensors, 2022.



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# THANK YOU

*"It is kind of fun to do the impossible"*

Walt Disney