Multi-modal Foundation Models for Wireless Communication and Sensing

Jun Zhang

iComAI Lab
INTEGRATED COMMUNICATIONS AND AI LAB

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

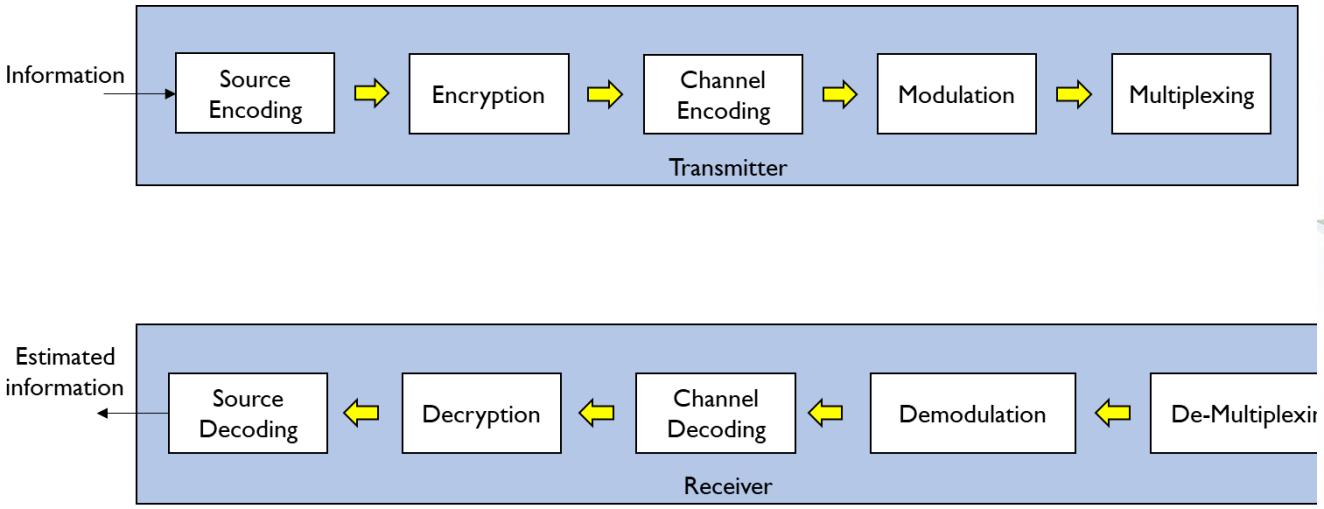
Outline

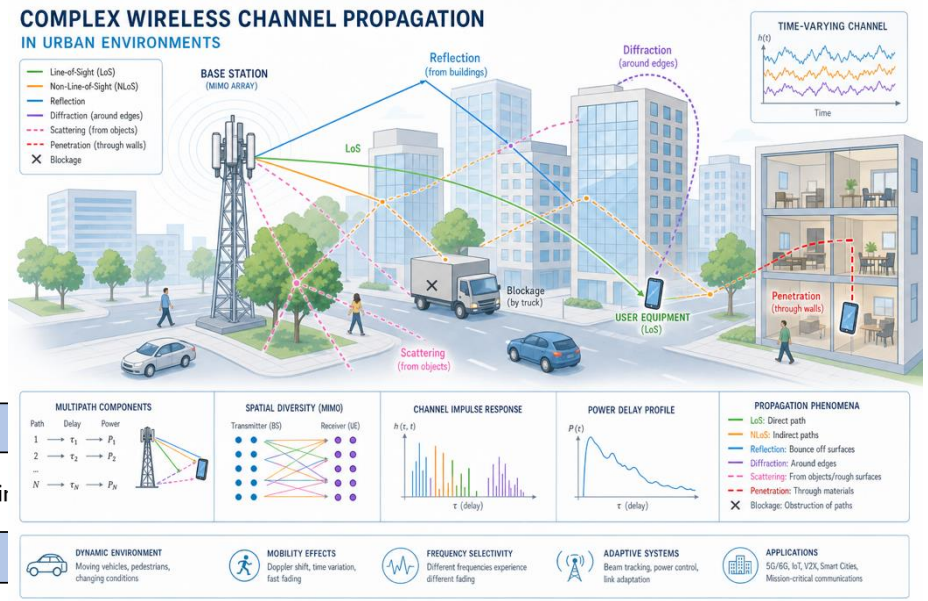
- Towards Multi-modal Wireless World Models
- **WRF-GS**: Wireless Radiation Field Reconstruction via 3D-GS
- **URF-GS**: A Unified Radio-Optical Radiation Field for Wireless
- **Point2Radio**: A Foundation Model for Cross-Scene Radio Fields

Towards Multi-modal Wireless Foundation Models

3D Channel Modeling and Wireless Sensing

The "final" challenge of wireless communications





Perfect CSI → **Pilot-assisted transmission** Imperfect CSI → **Massive MIMO** High-dim. CSI → **Dense networks, mixed-band Ultra-reliable and low latency**

Nice theory

Statistical estimation
It works well for data trans. (4G)

Still open for theory

Pilot overhead ~ 10%-20%

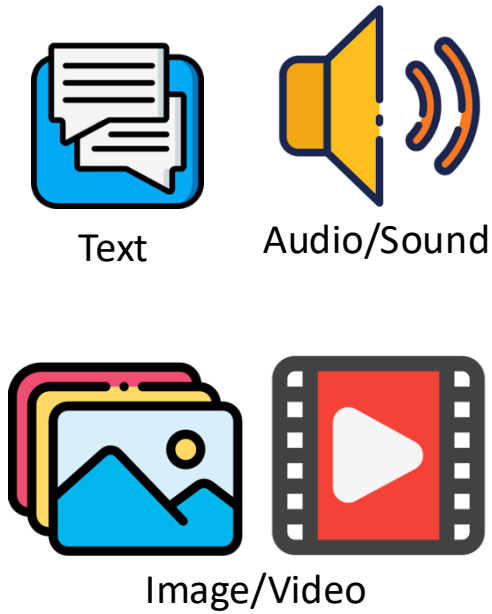
Exploit structural information (sparsity)

Environment-aware commun.
Low-pilot/pilot-free

Q: *How to effectively model and exploit the environment priors for wireless communication and sensing?*

Spatial Intelligence

- Spatial Perception: From text, audio, plane image/video, to 3D multi-view perception







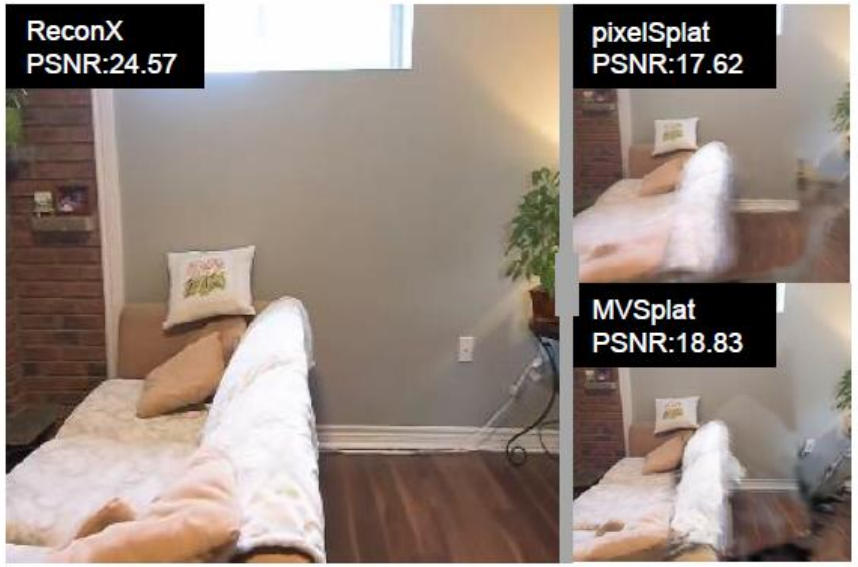
3D spatial perception

Our increasing ability in understanding the spatial world

TIP 26

ReconX:
Reconstruct any scene from sparse views with video diffusion model



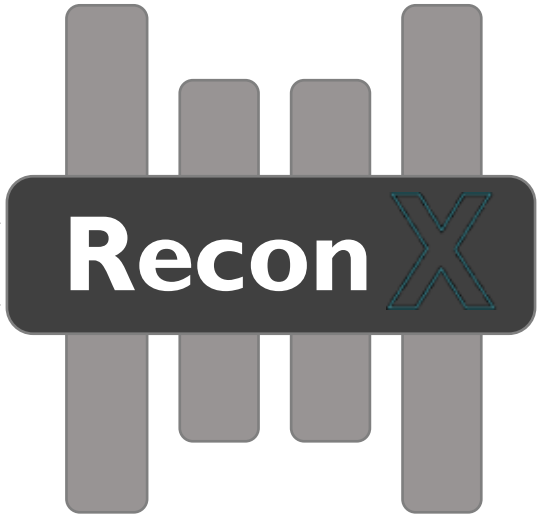


F. Liu*, W. Sun*, H. Wang*, Y. Wang, H. Chen, J. Ye, **J. Zhang**, and Y. Duan, "ReconX: Reconstruct any scene from sparse views with video diffusion model," *IEEE Trans. Image Processing (TIP)*, vol. 35, pp. 2305-2319, February 2026.

Our increasing ability in understanding the spatial world









Our recent endeavors in building world models



Wenqiang SUN
(Tencent Hunyuan, PhD HKUST)

- Learning Consistent and Interactive World Models with Generative Priors
  - **ReconX: 3D Worlds Reconstruction from sparse views [TIP 26]**
  - **DimensionX: 3D/4D worlds generation from a single image [ICCV 25]**
  - **WorldPlay: long-term consistent and real-time interactive worlds [ICML 26]**



Shenyuan GAO
(NVIDIA GEAR, PhD HKUST)

- Building Generalist World Models for Embodied Intelligence
  - **Vista: generalizable driving world model [NeurIPS 24]**
  - **AdaWorld: to enable human-alike adaptability [ICML 25]**
  - **DreamDojo: generalist robot world model [ICML 26, spotlight]**

My interest

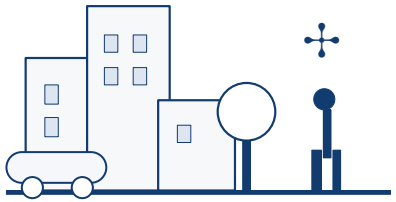
- **To understand how the world works**
  - ***How to reconstruct or create a physically realistic world?***

- **How to apply in wireless communications?**
  - ***How to reconstruct wireless radiation field, with minimal measurements?***
  - ***How to leverage multi-modal environment data?***
    - *Go beyond LLM-based approaches*

Spatial intelligence for wireless (**SI4Wireless**)

1 Physical Environment

What exists in the world

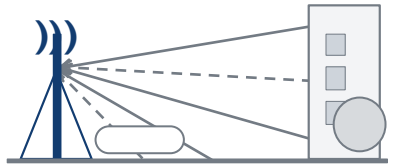


- geometry
- materials
- objects
- motion

causes

2 Wave Propagation Mechanisms

How waves interact with the environment

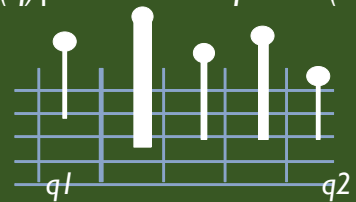


- LoS
- reflection
- diffraction
- scattering
- blockage

encodes

3 Wireless Radiance Field (WRF) H(Tx/Rx) Output

|H(q)| Channel Response (CSI)

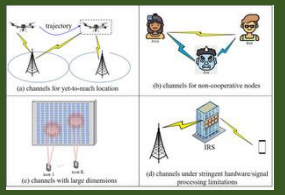


≈ gain α
◷ delay τ
∠ angle θ / φ
∼ Doppler ν

Radio Map

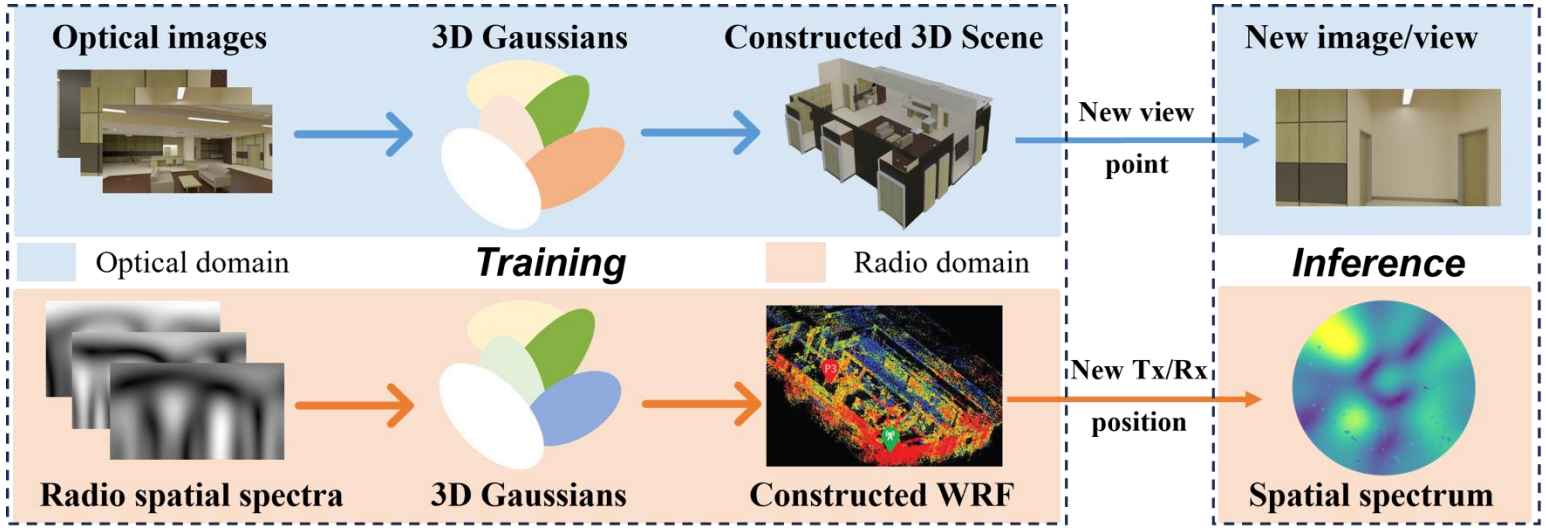


Channel Knowledge Map (CKM)



Spatial + **Intelligence**

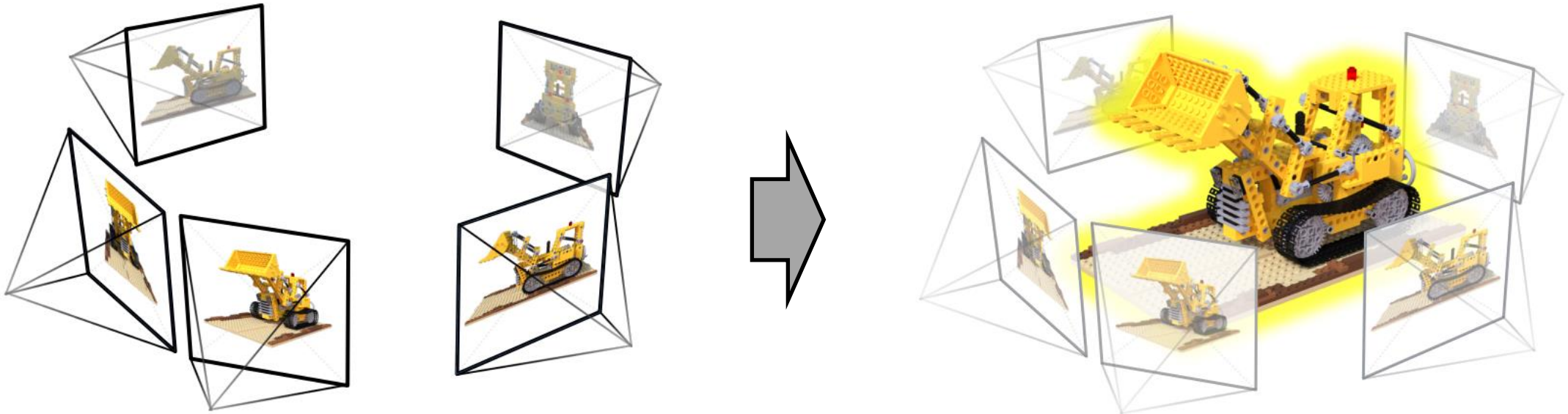
WRF-GS:
Wireless radiation field reconstruction via 3D-GS



C. Wen, J. Tong, Z. Lin, and **J. Zhang**, "WRF-GS: Wireless Radiation Field Reconstruction with 3D Gaussian Splatting," INFOCOM 2025.

3D reconstruction and rendering

- What is 3D reconstruction:
  - Create a 3D scene representation from a set of 2D images and their corresponding camera poses.

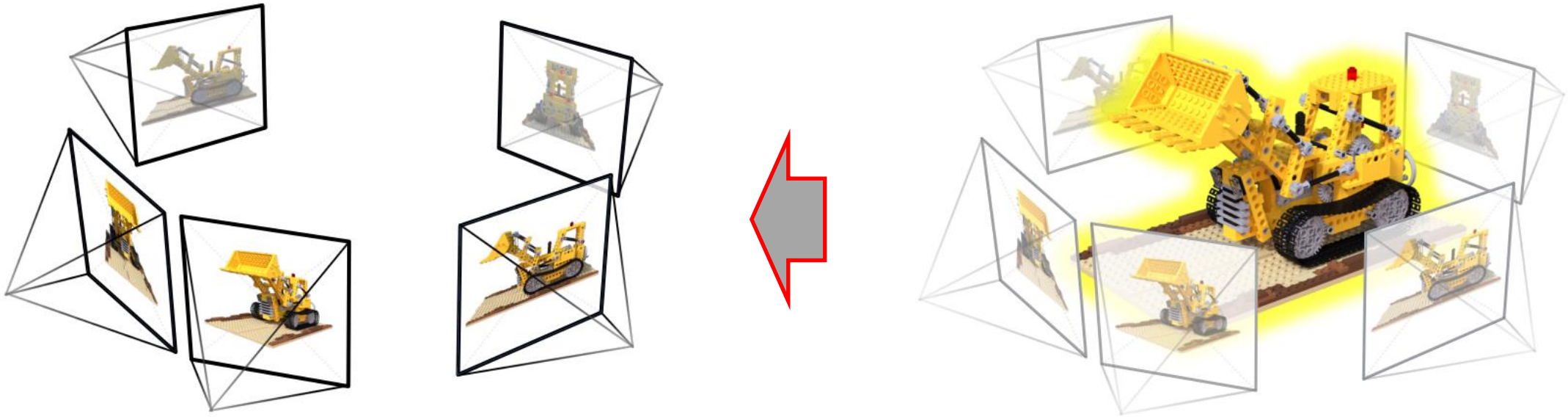


Images & Accurate Camera Poses

3D Scene Representation

3D reconstruction and rendering

- What is rendering:
  - Given a 3D scene representation and a camera pose, create a 2D image.



Images & Accurate Camera Poses

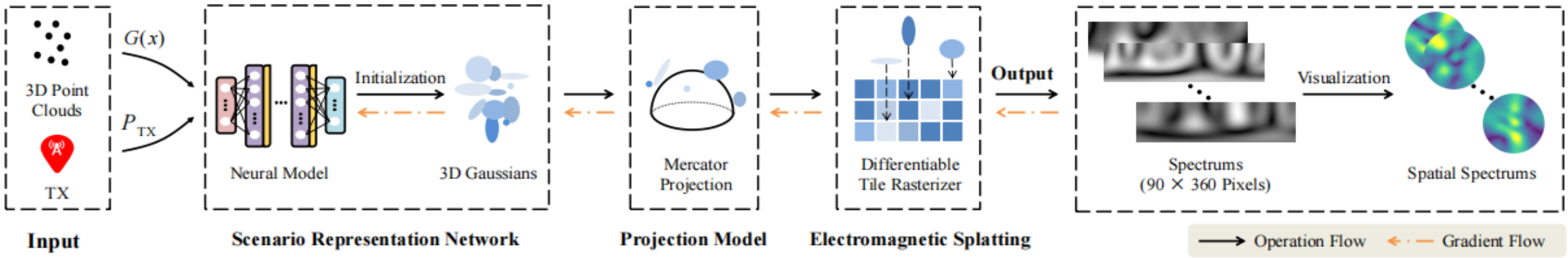
3D Scene Representation

Reconstruction of wireless radiation field



Since light is a form of electromagnetic (EM) wave, we can use optical 3D reconstruction techniques for wireless channel modeling.

Our framework: WRF-GS



Input: The position of the 3D point cloud and TX.

Scenario Representation Network: The signal source is replaced by a series of virtual TXs in space and represented using 3D-GS, where each 3D Gaussian obtains a complex-valued signal and attenuation at a specific location.

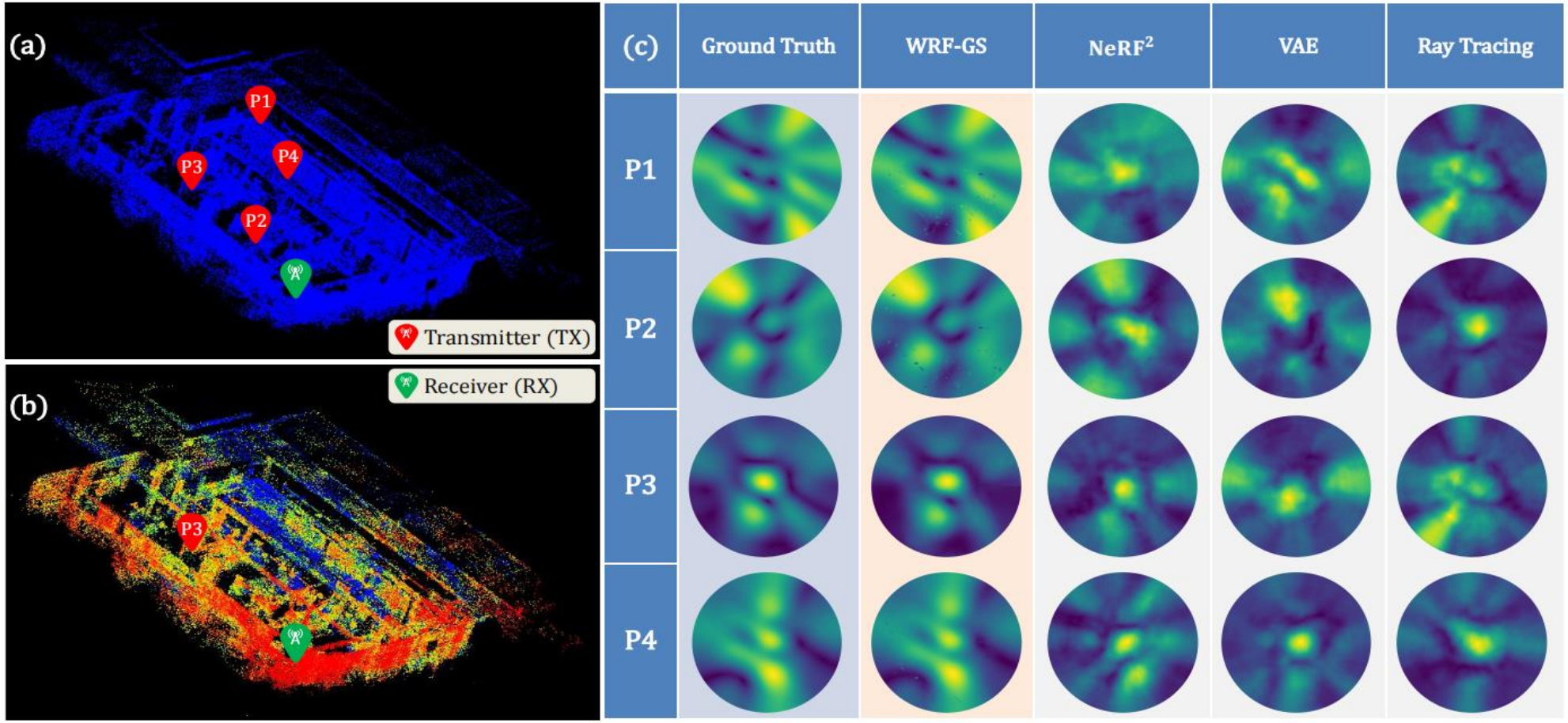
Projection Model: The 3D Gaussian is transformed from world coordinates to pixel coordinates using the Mercator projection.

Electromagnetic Splatting: The signals in each direction at TX are calculated in parallel.

Output: The spatial spectrum at any TX position is the channel feature. For convenience, the RX position expands the spatial spectrum into a 90×360 pixel image, where each pixel corresponds to a signal arrival angle with an angular resolution of 1°.

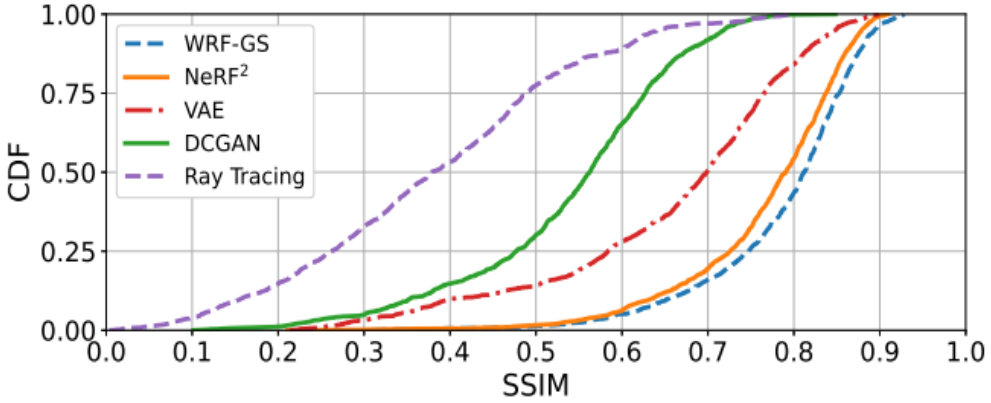
Experiment: WRF reconstruction

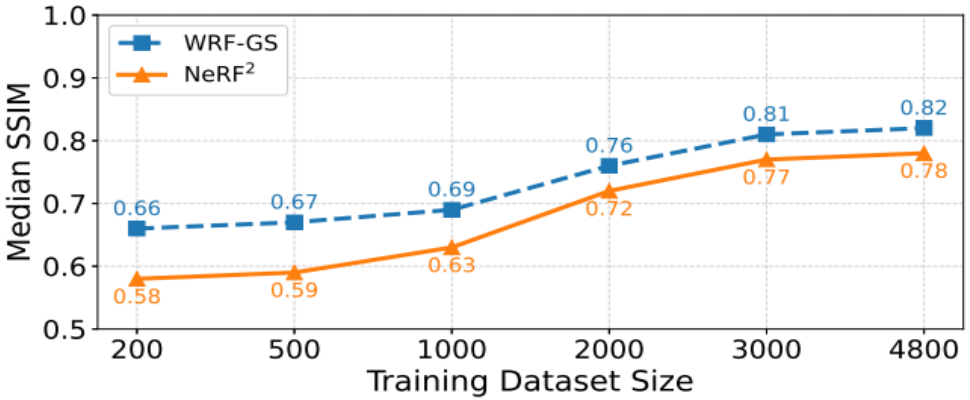
- The synthesized spatial spectrums of four algorithms when the TX is located at four different positions.



Experiment: WRF reconstruction

- We adopt the structural similarity index measure (SSIM) as the performance metric to compare different methods.

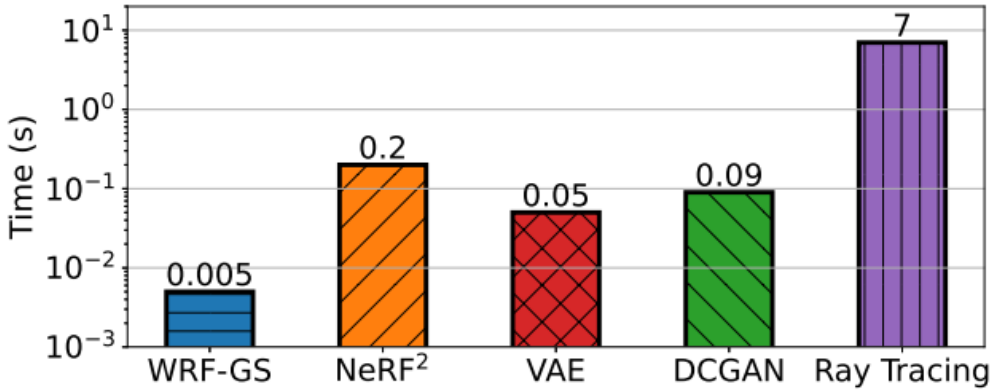






(GitHub: WRF-GS)

- The rendering time required to synthesize a spatial spectrum using different methods under the same settings.

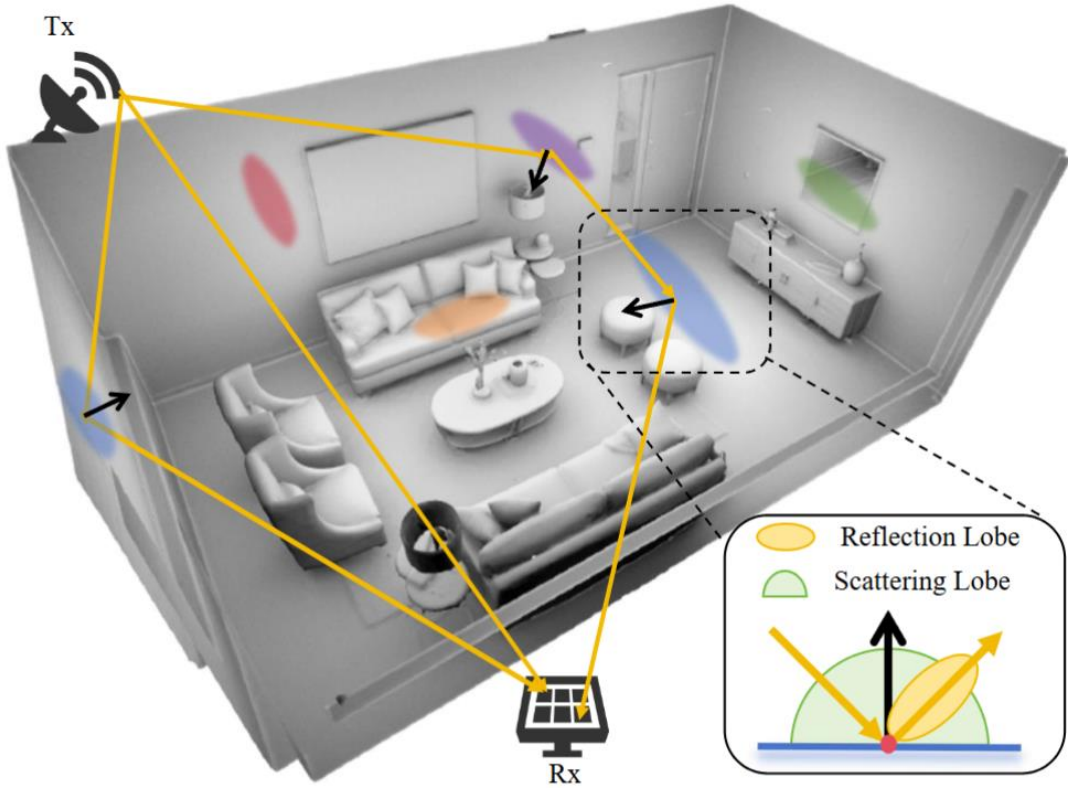




(GitHub:WRF-GS+)

URF-GS:

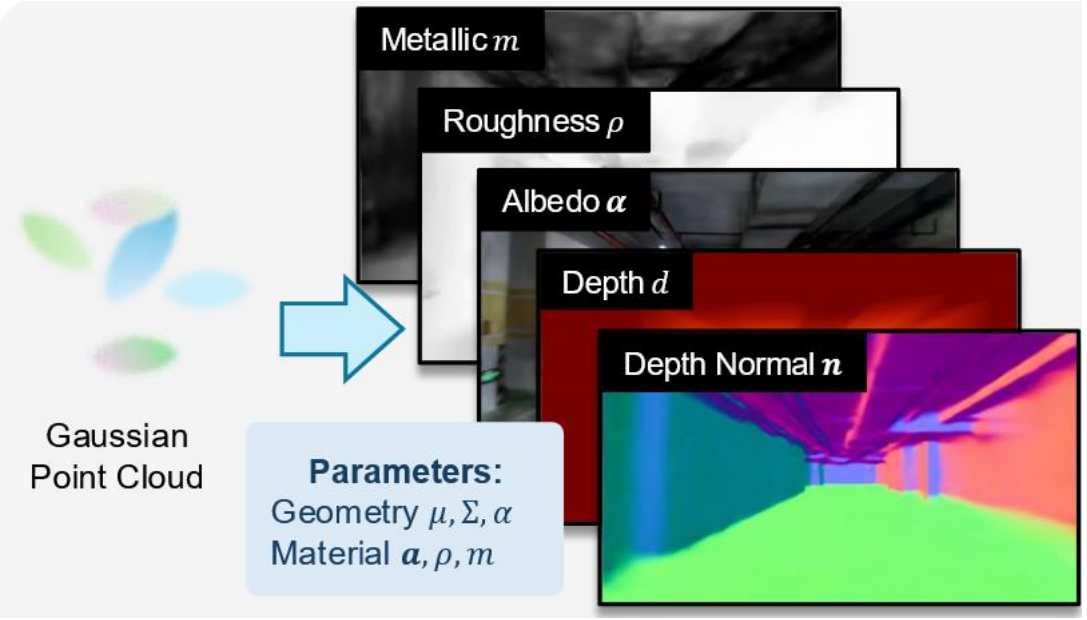
A Unified Radio-Optical Radiation Field for 3D Radio Map Construction



Inverse rendering to exploit rich vision information

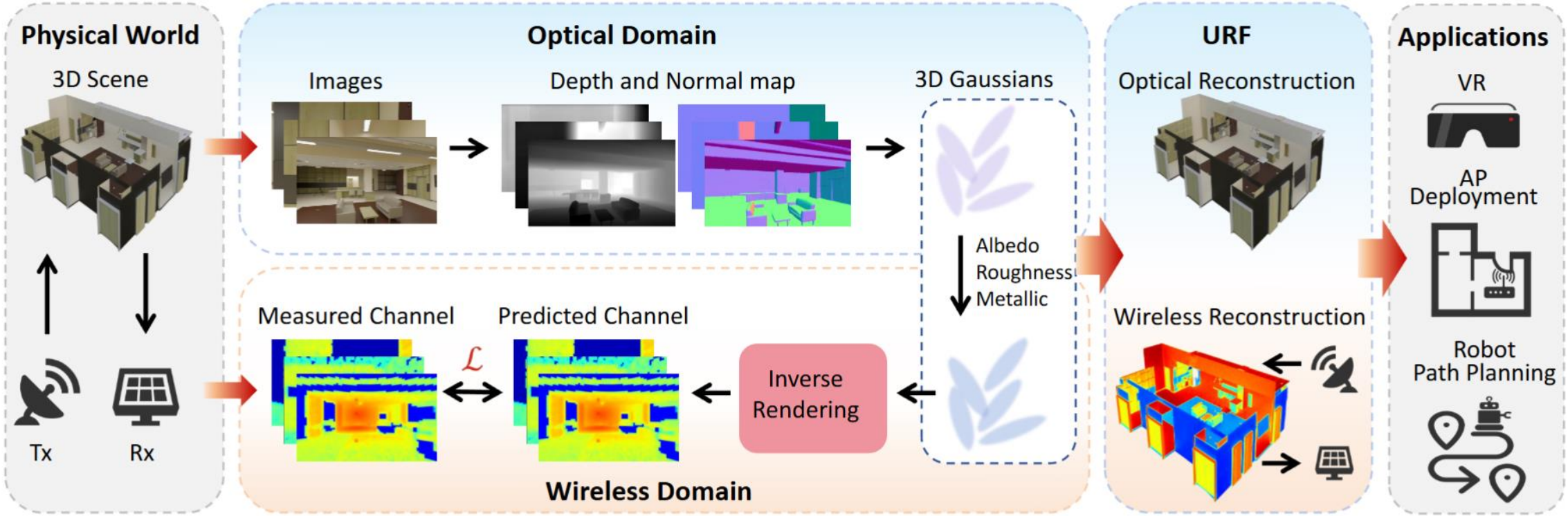
Why inverse rendering?

- Inverse rendering aims to estimate the physical attributes of a 3D scene, such as material properties, geometry, and lighting, from captured images.
- It supports multiple downstream tasks such as scene editing and changing lighting conditions.

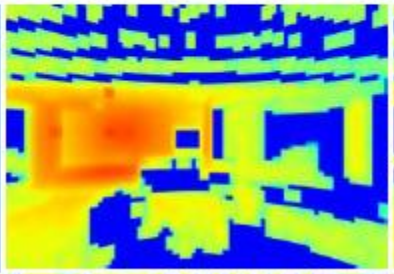
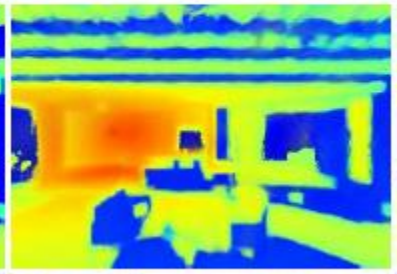
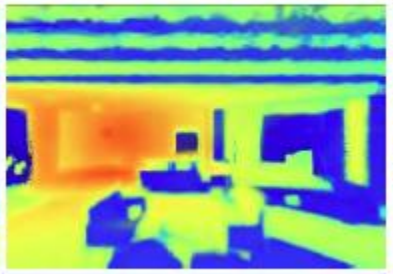
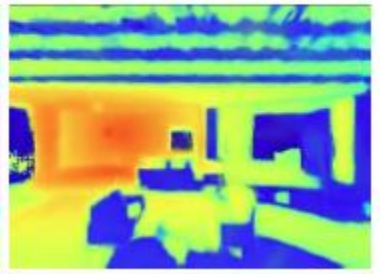
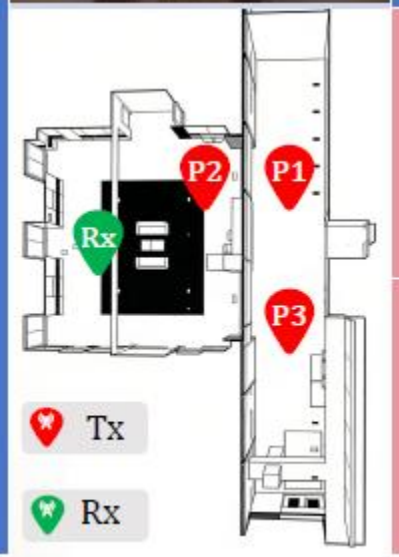
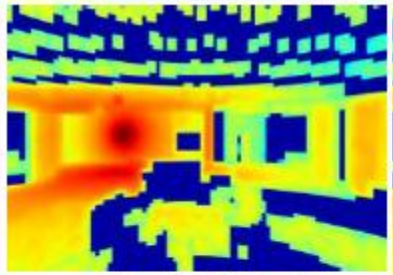
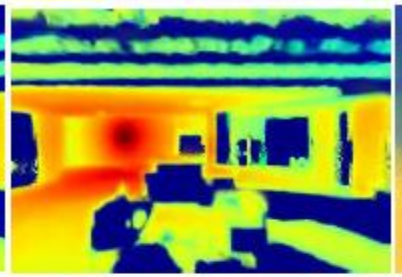
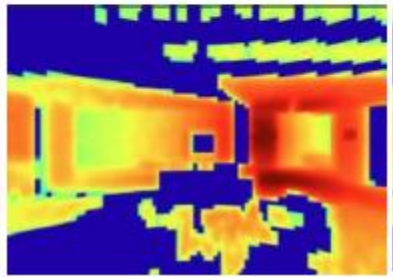
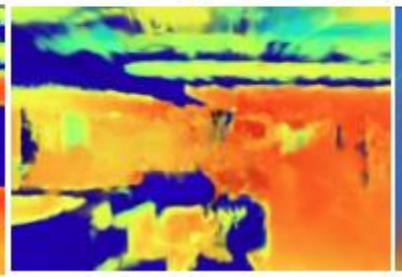
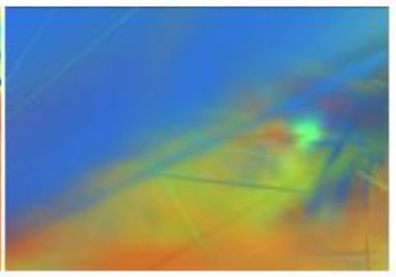


We propose URF-GS, a wireless channel inverse rendering framework based on 3D-GS, to support scene generalization ability!

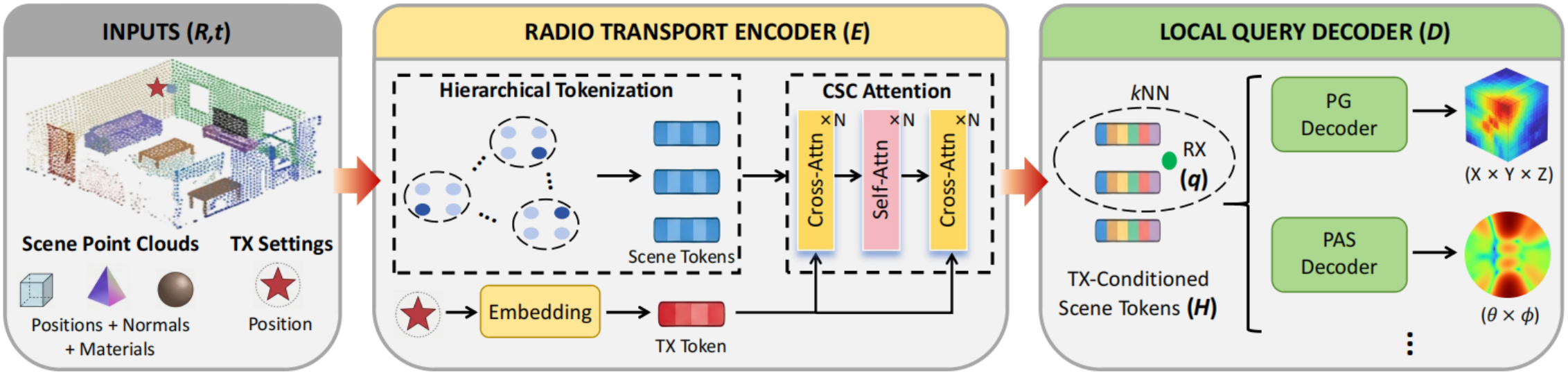
URF-GS: Framework



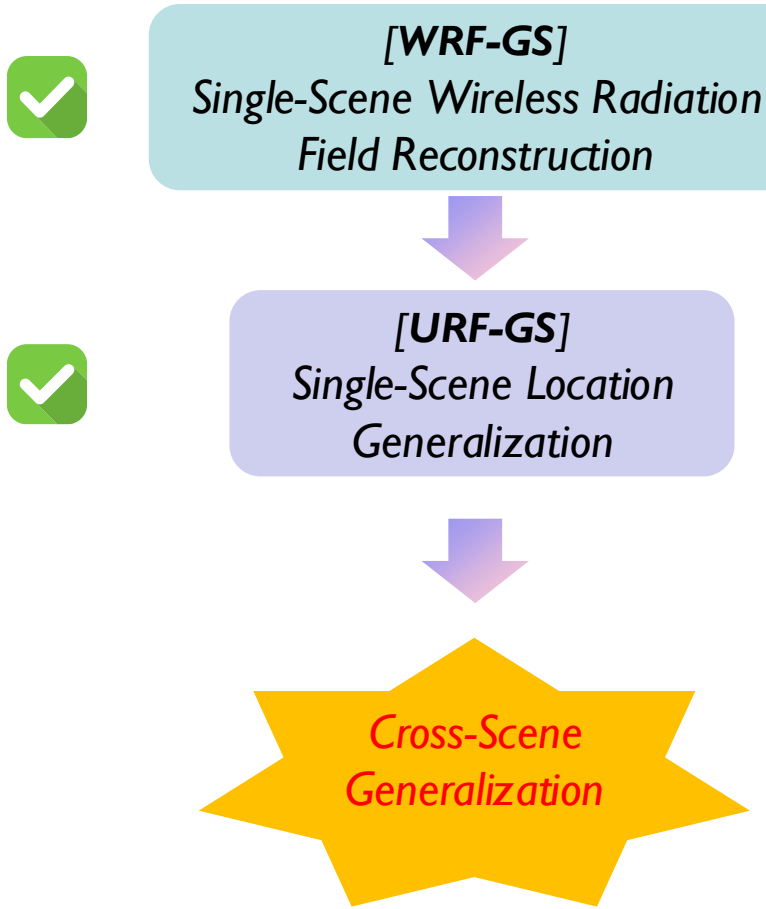
Experiment: 3D radio map construction

(a)	Setting	(b)	Ground Truth	URF-GS (Ours)	WRF-GS+	RF-3DGS
		P1 Full-Training				
		P2 Few-Shot				
		P3 Zero-Shot				

Point2Radio:
A Foundation Model for Cross-Scene Radio Fields from Material-Aware Point Clouds



What's next?

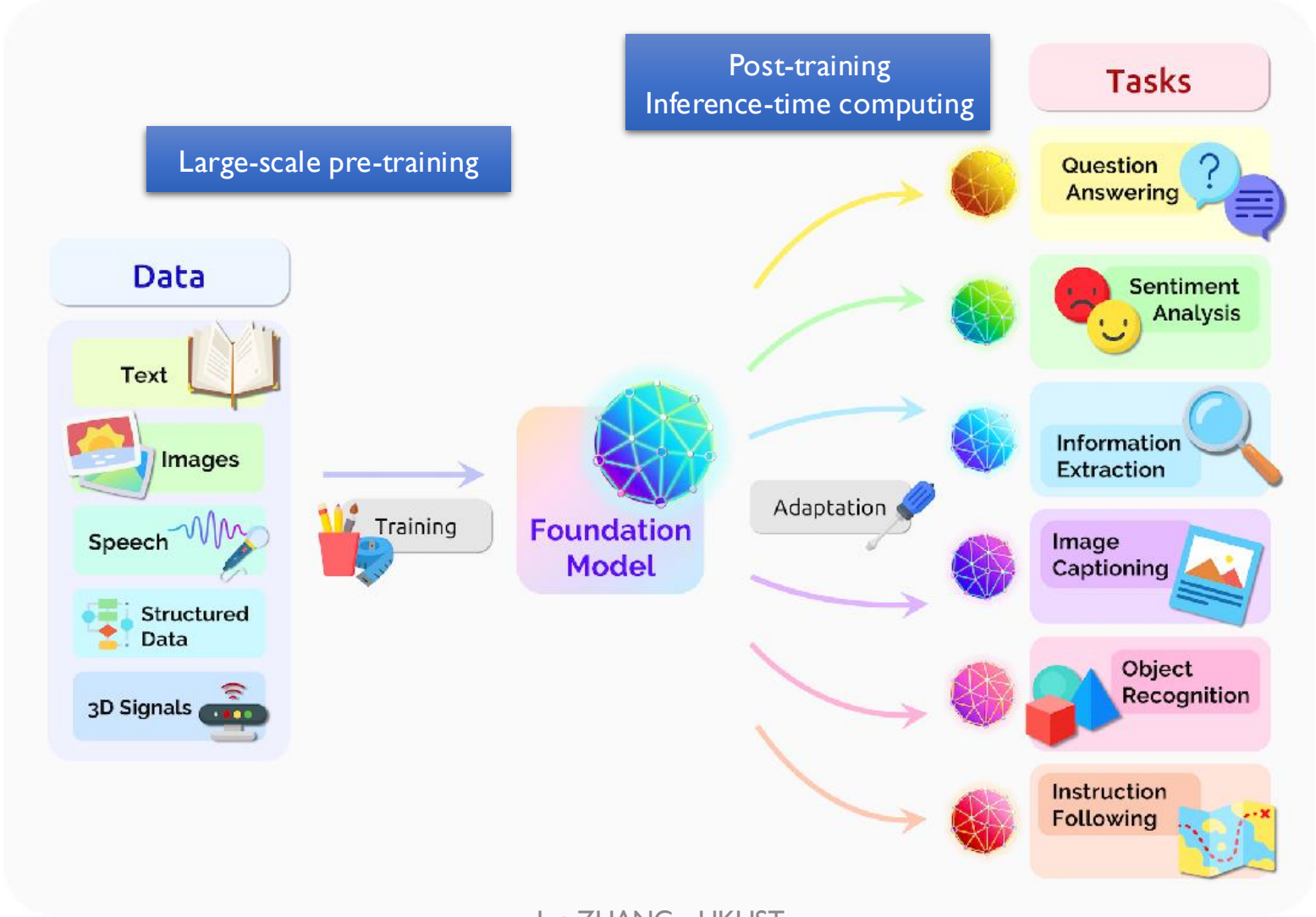


Fails to generalize to arbitrary TX-RX setups from a given configuration.

Overfits to a specific scene and requires retraining for entirely new environments.

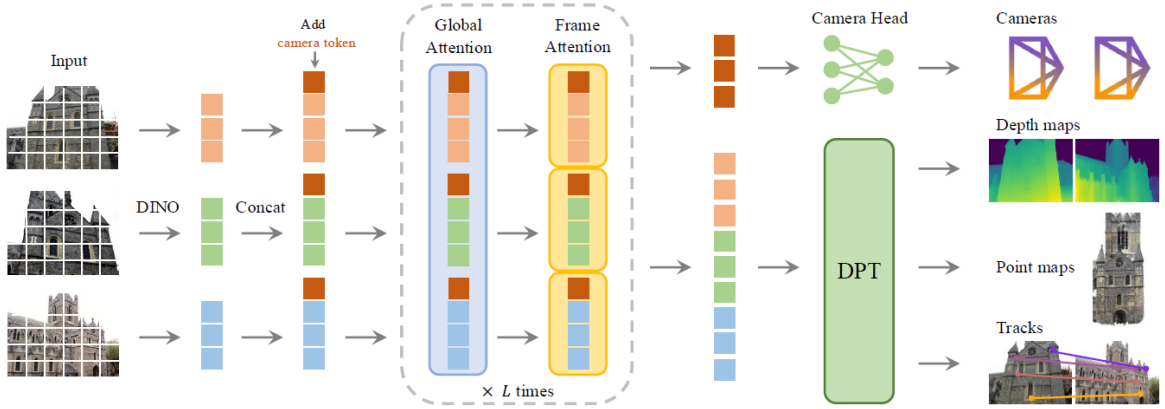
Can we construct a foundational model trained on large-scale data to learn the inherent propagation patterns between TX and RX, thereby achieving true generalization across diverse scenarios?

The era of foundation models

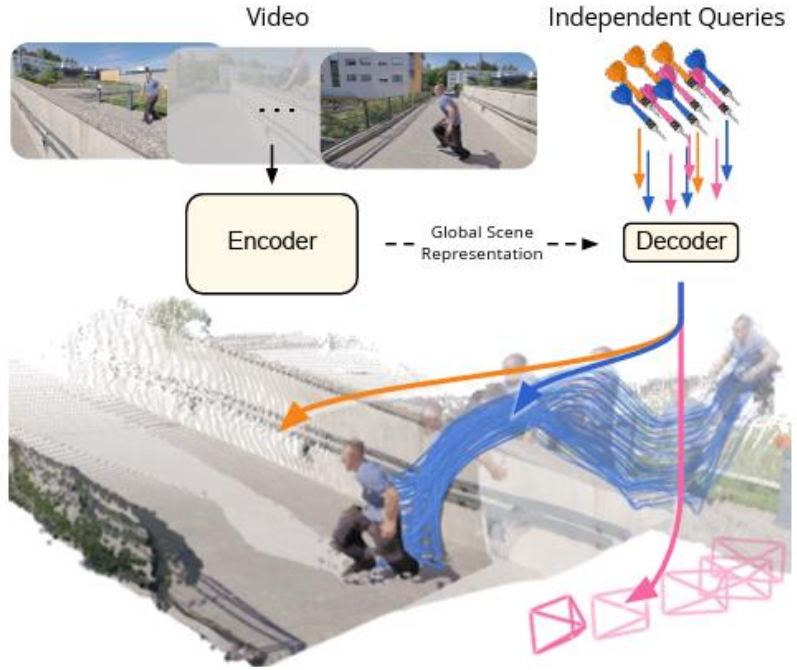


Inspiration: Foundation models for 3D

- **VGGT**: a feed-forward neural network that can directly estimate all key 3D scene properties for hundreds of input views.
  - **CVPR 2025 Best Paper Award**



- **D4RT**: a network that can reconstruct the geometry and motion of dynamic 4D scenes from video.
  - **CVPR 2026 Best Paper Award**



Foundation model for radio fields

Shared backbone + task-specific heads — one transferable propagation prior.

$$H = E(R, t) \qquad \hat{y} = D(H, q)$$

R

Material-aware point cloud
(geometry + EM attributes)

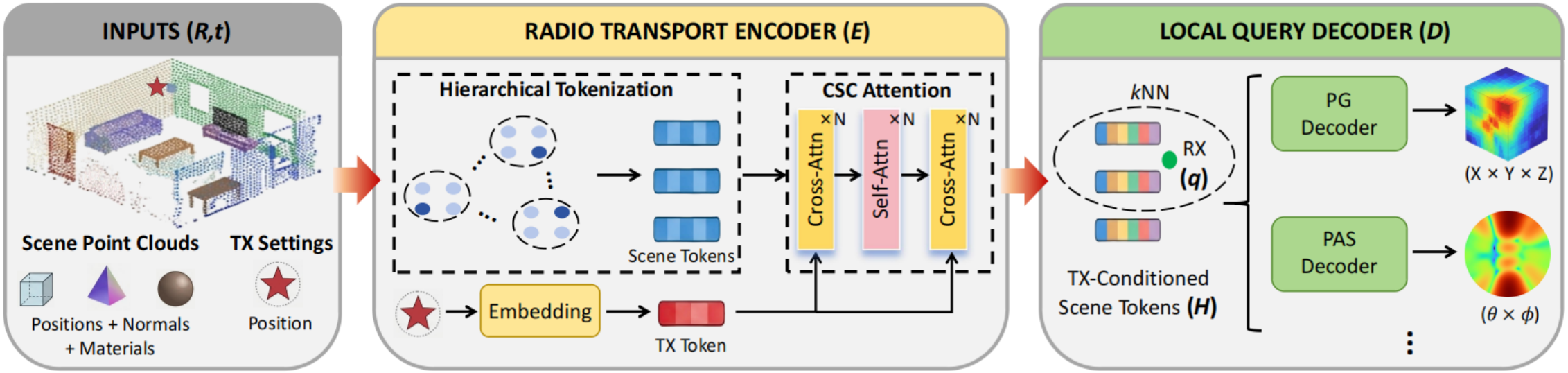
E

Radio transport encoder
→ TX-conditioned codes H

D

Task query decoder
→ PG fields or PAS spectra

Method overview



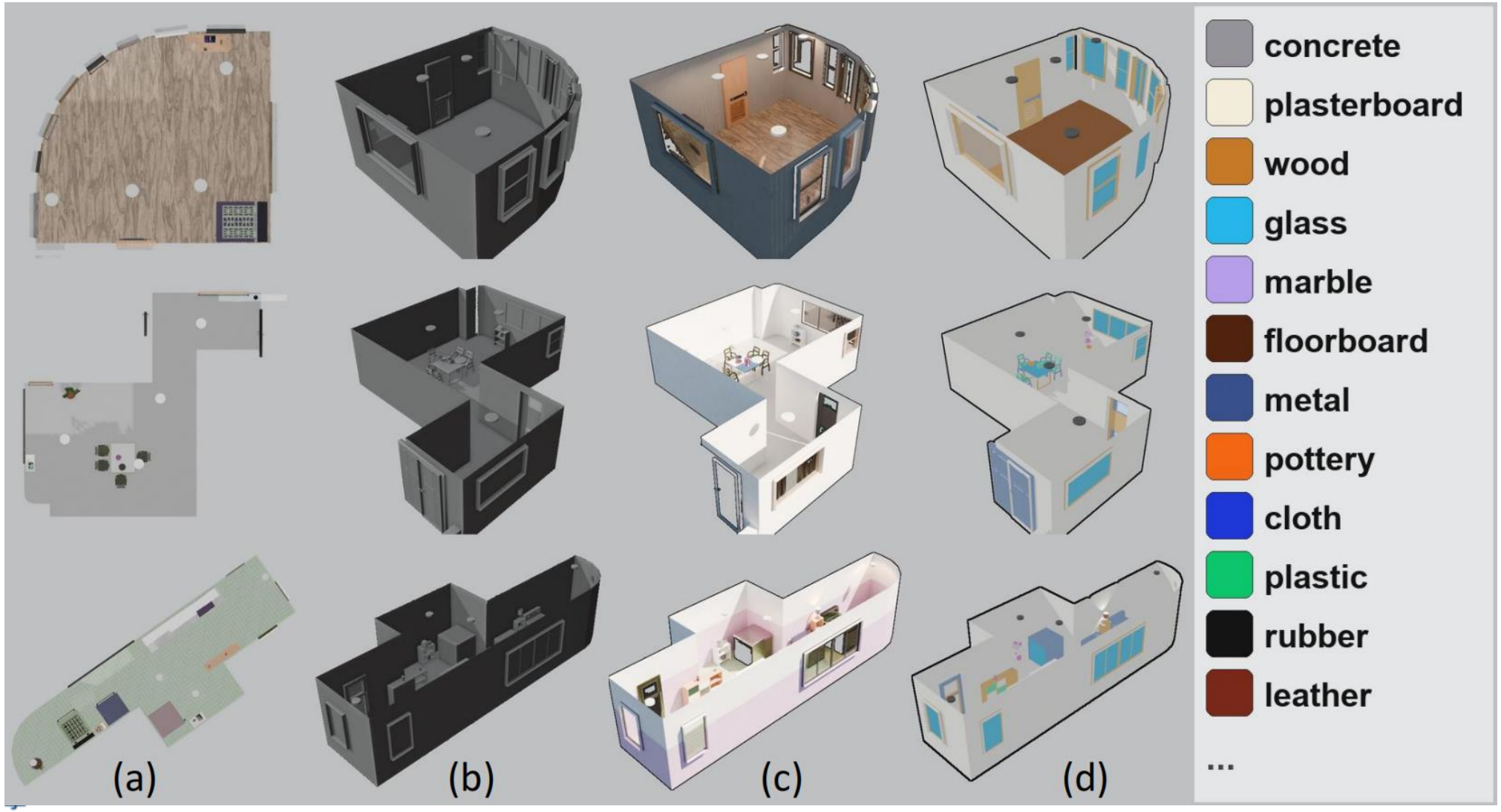
Point features: position, normal, ε, $\log_{10}\sigma$, scattering, XPD, emission flag

Attention as a feedforward surrogate for multipath transport (no online ray tracing)

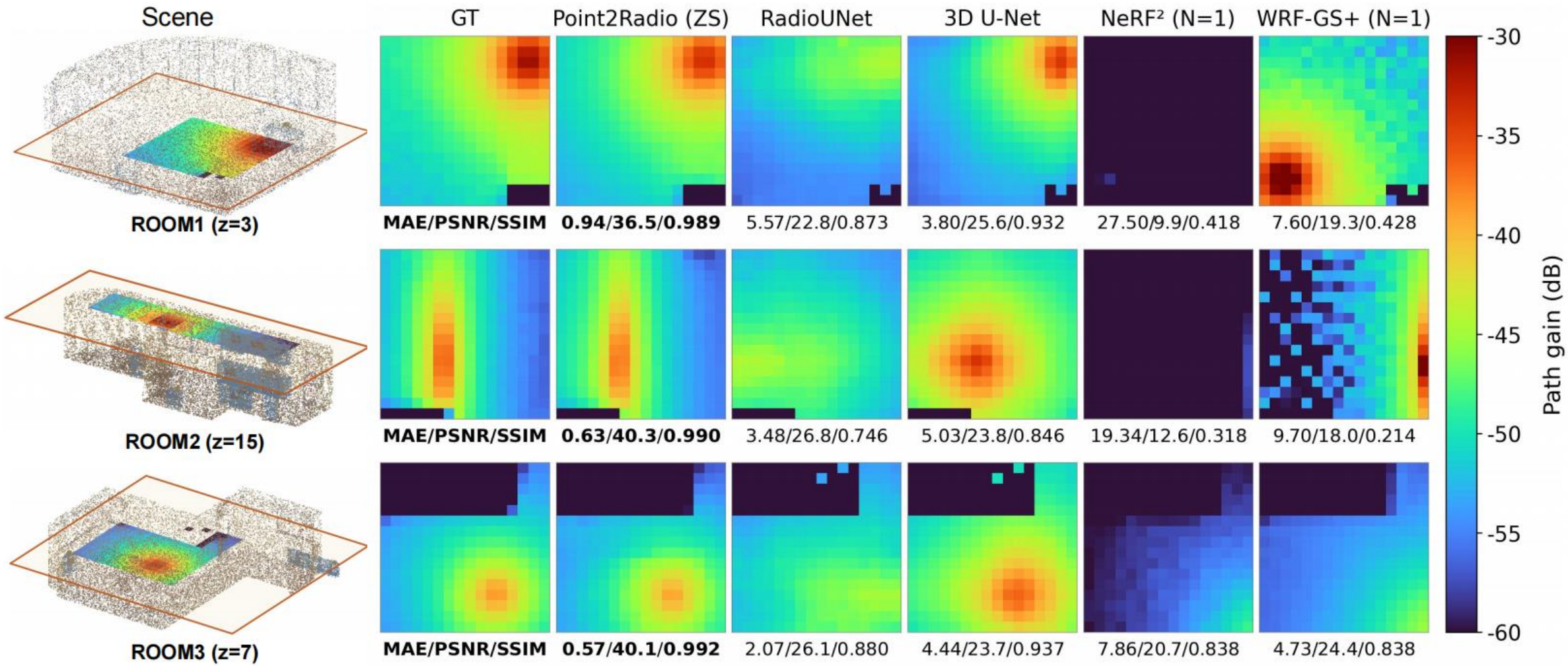
After PG pretraining: freeze encoder → train PAS head; optional residual FT per scene

PG: path-gain; **PAS**: power angular spectra; **ZS**: zero-short; **FT**: fine-tuning.

PRISM dataset



Experiment: Cross-scene path gain



Experiment : Path gain

Zero-shot full-volume PG on 35 held-out test scenes.

Method	MAE ↓	NMSE	PSNR ↑	SSIM ↑
RadioUNet (2D)	3.745	0.0159	25.10	0.836
3D U-Net	3.038	0.0152	26.71	0.884
Point2Radio (ZS)	**0.871**	**0.0025**	**34.83**	**0.954**

76.7%

MAE reduction vs. RadioUNet

0.871 dB

Zero-shot MAE on held-out rooms

0.954

SSIM — strong spatial structure



Takeaways

- **A new frontier:** Spatial Intelligence for Wireless (SI4Wireless)

- **Applications in wireless communications**
  - Spatial/3D channel modeling
  - Environment-aware communication
  - Wireless sensing
  - …

- **Key**: how to efficiently exploit different modalities?
  - To resolve the **data scarcity problem** in wireless

References

- C. Wen, J. Tong, Y. Hu, Z. Lin, and **J. Zhang**, "WRF-GS: Wireless radiation field reconstruction with 3D Gaussian splatting," *IEEE INFOCOM 2025*.
- C. Wen, J. Tong, Y. Hu, Z. Lin, and **J. Zhang**, "Neural representation for wireless radiation field reconstruction: A 3D Gaussian splatting approach," *IEEE Trans. Wireless Commun.*, vol. 25, 7490-7504, 2026.
- C. Wen, J. Tong, Z. Lin, and **J. Zhang**, "Bridging Visual and Wireless Sensing: A Unified Radiation Field for 3D Radio Map Construction," *npj Wireless Technology*, to appear.
- C. Wen, C. Bian, H. Chen, **J. Zhang**, "Point2Radio: A Foundation Model for Cross-Scene Radio Fields from Material-Aware Point Clouds," *preprint*. https://arxiv.org/abs/2607.28994v1



Chaozheng Wen
(PhD HKUST)

A warm welcome

iComAI Lab
INTEGRATED COMMUNICATIONS AND AI LAB

IEEE Hong Kong 6G Wireless Summit

HONG KONG 6G
6G@HK Research Center

3 – 4 December 2026 // Hong Kong, China



IEEE

IEEE ComSoc
IEEE Communications Society