

Spatial Intelligence: A New Paradigm for Wireless Communications and Sensing

空间智能：无线通信与感知的新范式

Jun Zhang

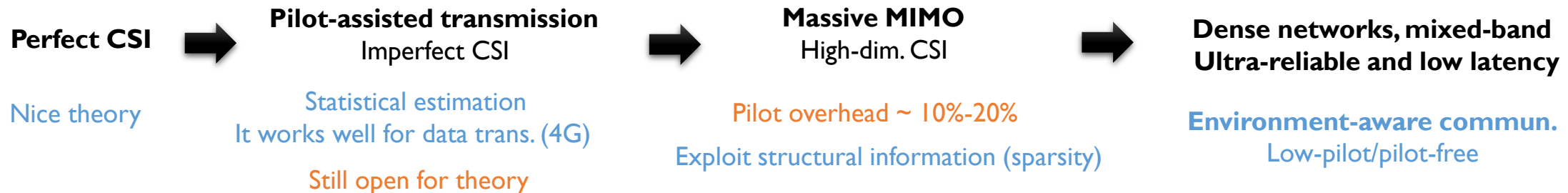
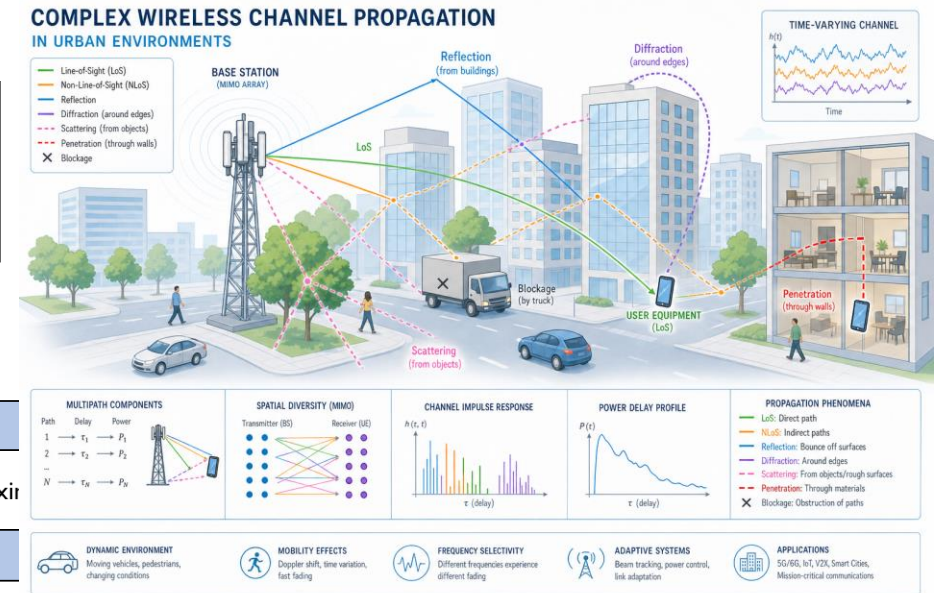
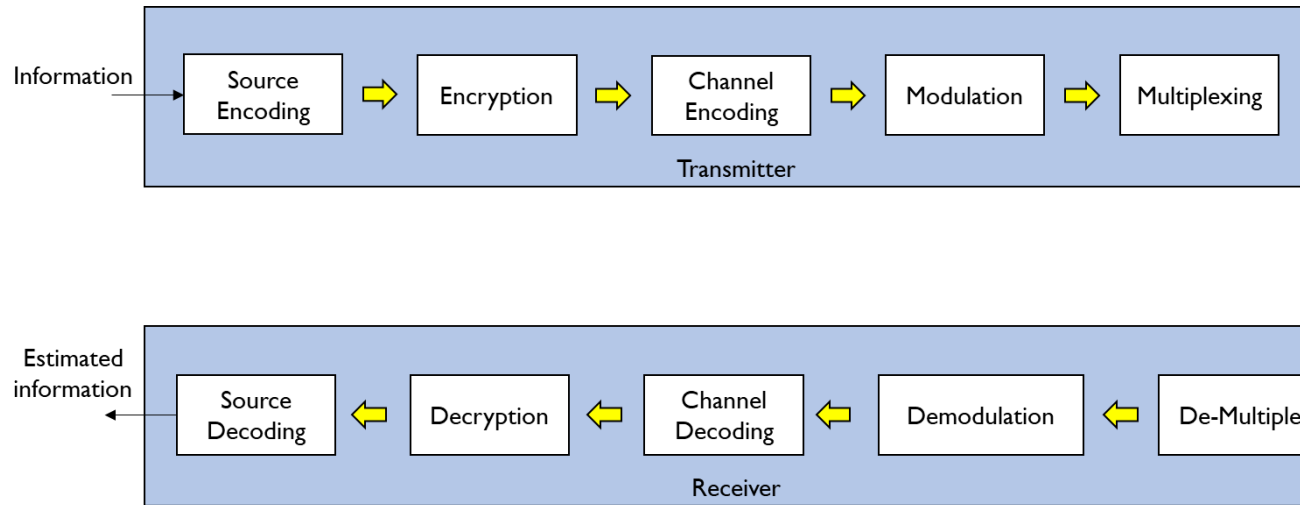


Outline

- 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

The “final” challenge of wireless communications



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

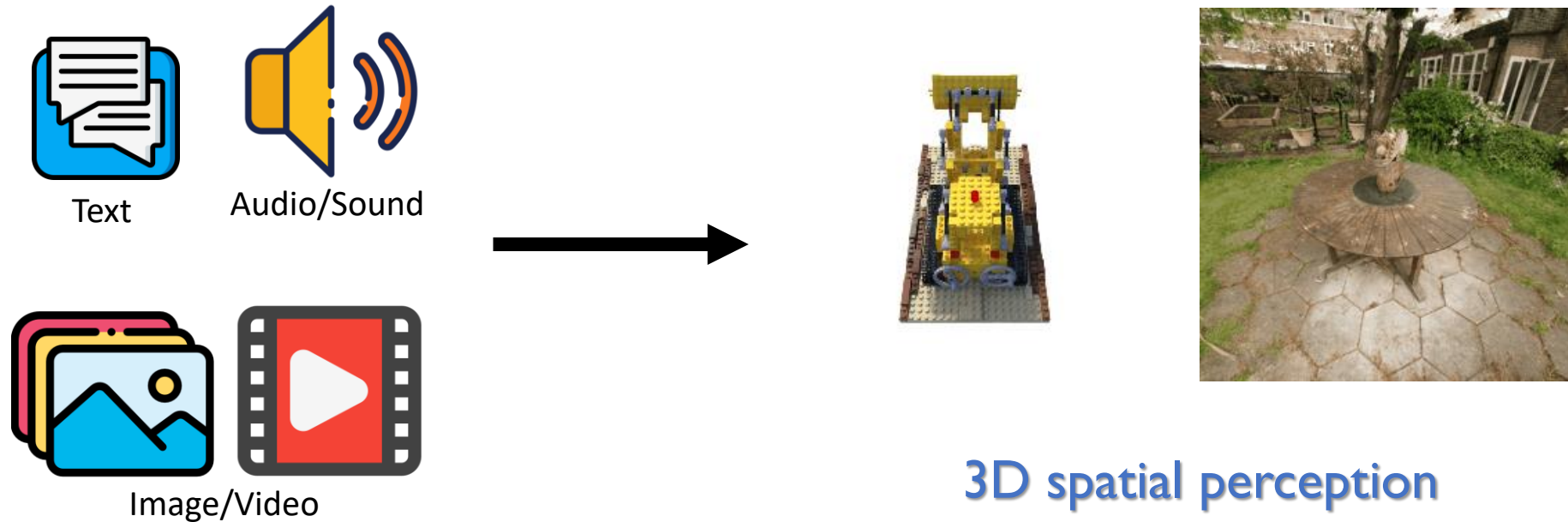
The bottlenecks of “AI for Wireless”

- Radio data scarcity
 - Data, Data, Data
 - Undiversified, low-quality real measurement data
- Generalization/robustness
 - Distribution shift issue
 - Difficult to adapt to new scenes

Q: How to reconstruct/generate diverse radio data, and how to incorporate environmental information for channel modeling?

Spatial Intelligence

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



The rise of world models



AI pioneer Fei-Fei Li's World Labs raises \$1 billion in funding

By Reuters

February 19, 2026 3:22 AM GMT+8 · Updated February 19, 2026



Ex-Meta AI chief Yann LeCun's AMLI raises \$1.03 billion for alternative AI approach

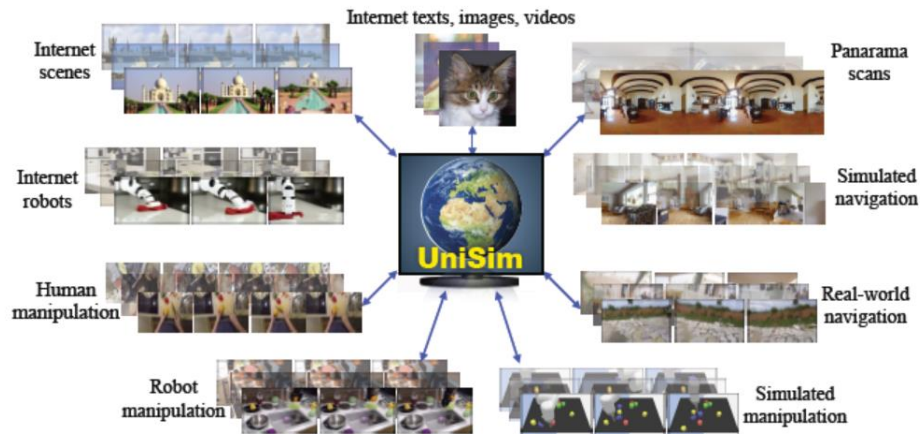
By Katie Paul and Anhata Rooprai

March 10, 2026 1:01 PM GMT+8 · Updated March 10, 2026

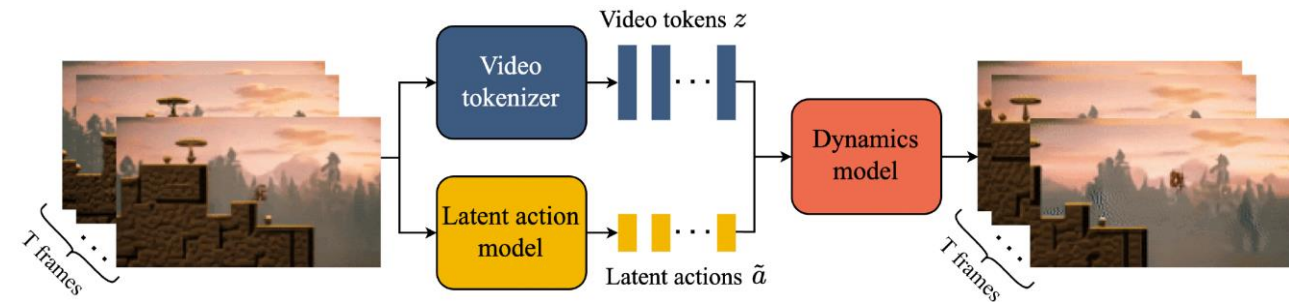


The rise of world models

UniSim [ICLR 24 Outstanding Paper]



Genie [ICML 24 Best Paper]



GAIA-1: A Generative World Model for Autonomous Driving

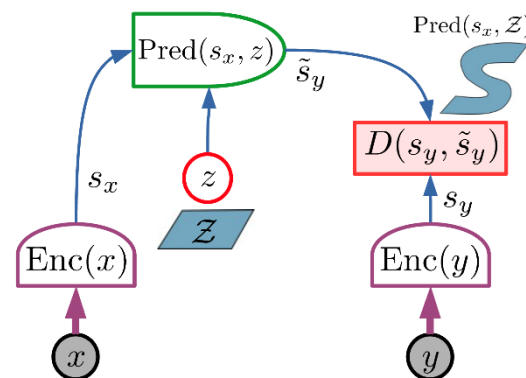
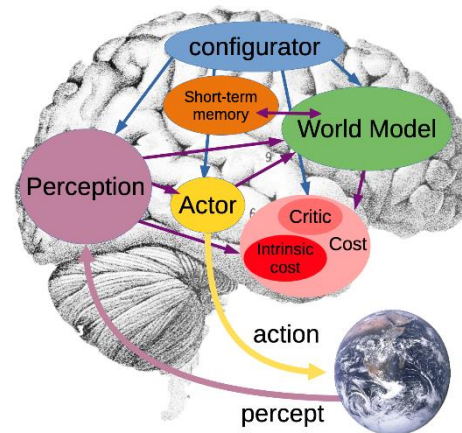
Anthony Hu* Lloyd Russell* Hudson Yeo* Zak Murez George Fedoseev
Alex Kendall Jamie Shotton Gianluca Corrado
Wayve
research@wayve.ai

ADriver-I: A General World Model for Autonomous Driving

Fan Jia^{1*} Weixin Mao^{2*†} Yingfei Liu¹ Yucheng Zhao¹ Yuqing Wen^{3†}
Chi Zhang⁴ Xiangyu Zhang¹ Tiancai Wang^{1†}
¹MEGVII Technology ²Waseda University
³University of Science and Technology of China ⁴Mach Drive

Q1: What is a world model?

- Different World Models

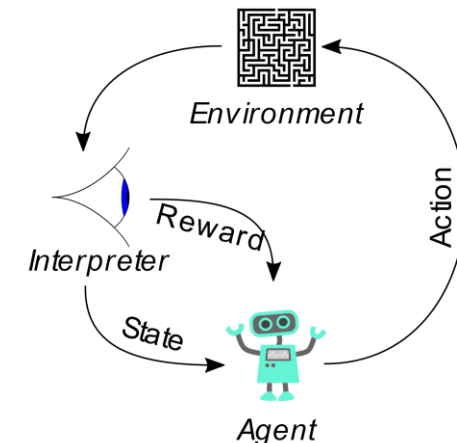


JEPA (representation)

Jun ZHANG - HKUST

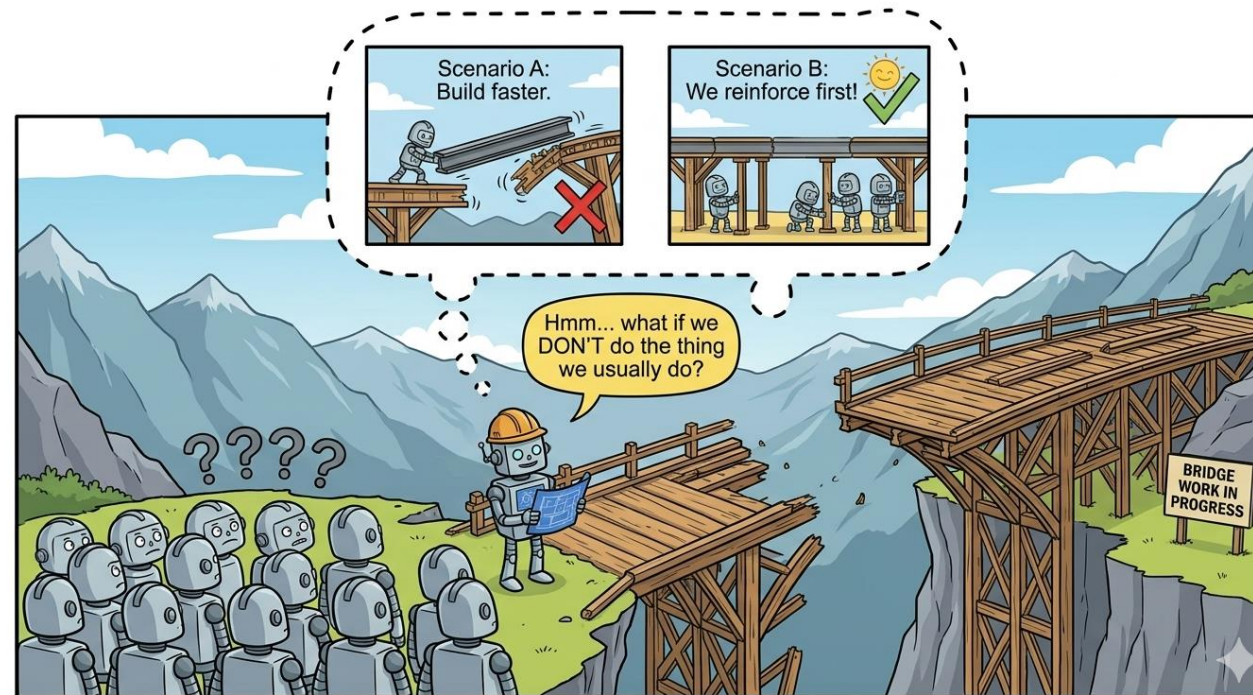
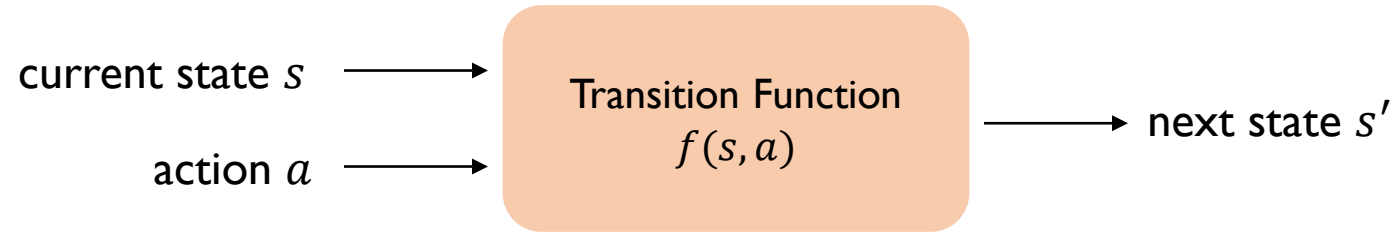


World Labs (3D)



Reinforcement Learning

Q1: What is a world model?



Q2: Why do we want world models?



What I cannot create, I do not
understand.

— *Richard P. Feynman* —

AZ QUOTES

Q2: Why do we want world models?



- **Dr. Strange:** I went forward in time, to **view alternate futures**, to **see all the possible outcomes** of the coming conflict.
- **Peter Quill:** How many did you see?
- **Dr. Strange:** **Fourteen million six hundred and five.**
- **Tony Stark:** How many did we win?
- **Dr. Strange:** **One.**

Q3: Why now?

February 15, 2024 Publication

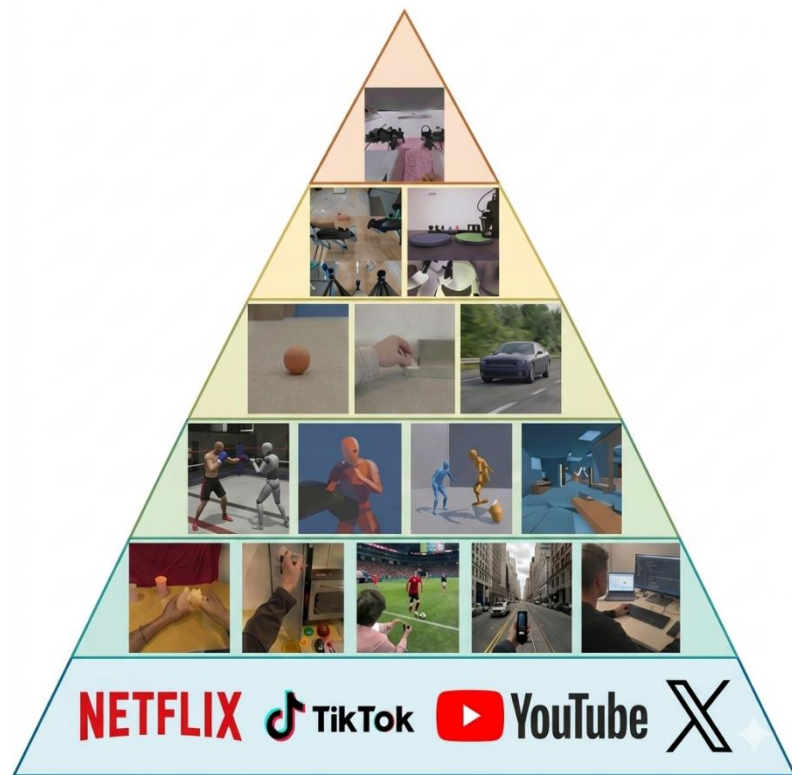
Video generation models as world simulators



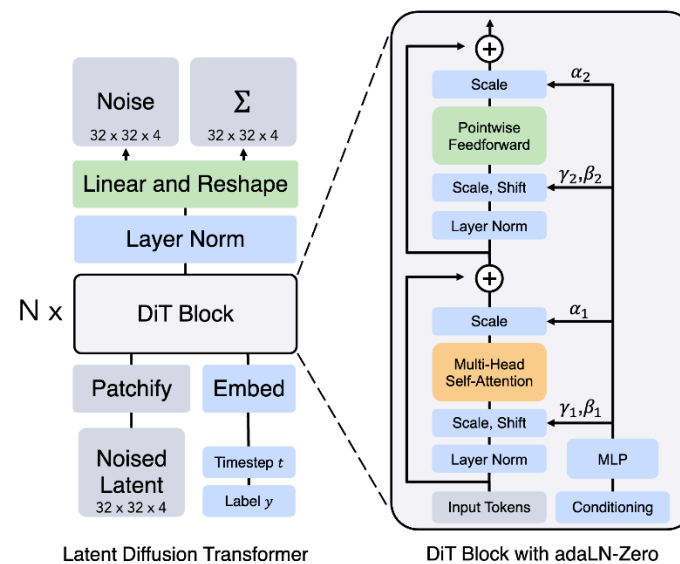
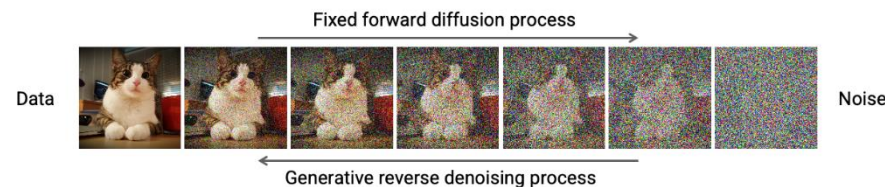
<https://openai.com/index/video-generation-models-as-world-simulators/>

Impressive spatial and temporal consistency!

Video World Models



Scalable Video Data



Scalable Algorithm and Architecture



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]



Shen yuan 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]

Our increasing ability in understanding the spatial world

TIP 26

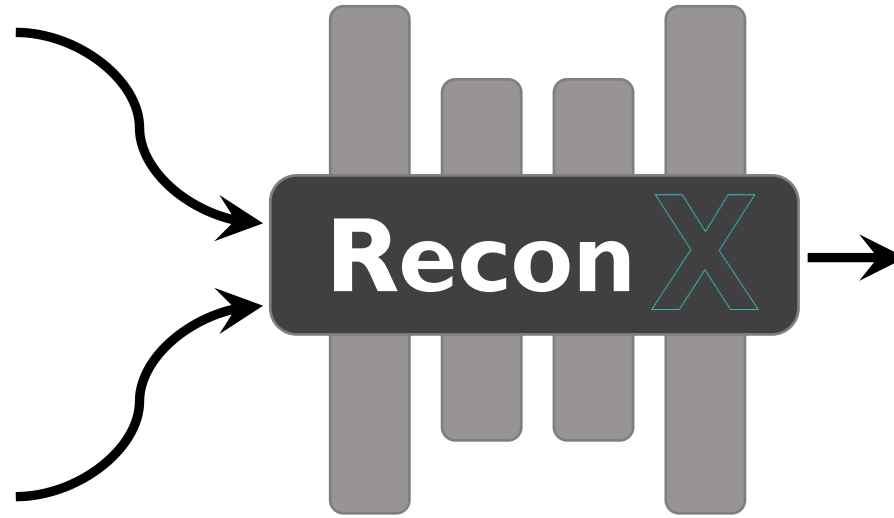


ReconX: Reconstruct any scene from sparse views with video diffusion model



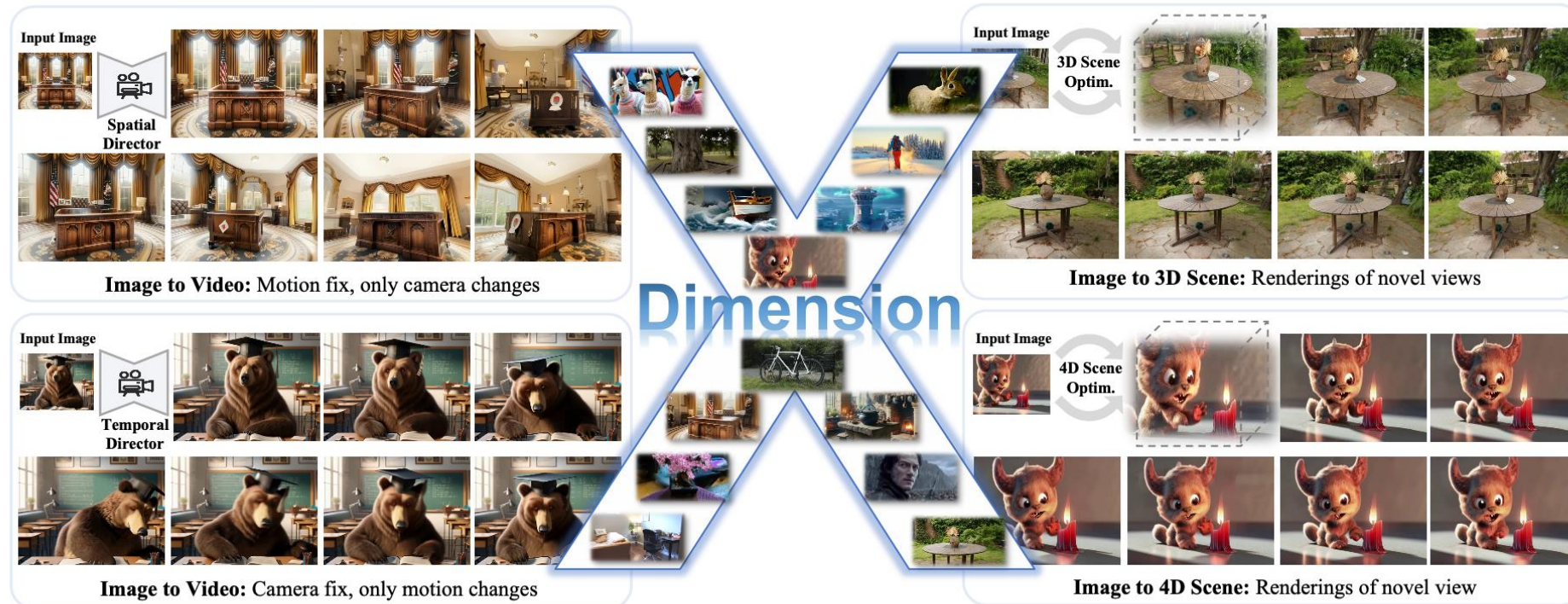
F. Liu*, W. Sun*, H. Wang*, Y. Wang, H. Chen, J. Ye, **J. Zhang**, and Y. Duan, "ReconX: Reconstruct any scene from sparse views with video diffusion model," *IEEE Trans. Image Processing (T-IP)*, vol. 35, pp. 2305-2319, February 2026.

Our increasing ability in understanding the spatial world



Our increasing ability in generating the spatial world

DimensionX: Create Any 3D and 4D Scenes with Controllable Video Diffusion



W. Sun, S. Chen, F. Liu, Z. Chen, Y. Duan, J. Zhang, and Y. Wang, "DimensionX: Create any 3D and 4D scenes from a single image with controllable video diffusion," ICCV 2025. <https://arxiv.org/abs/2411.04928>

Our increasing ability in generating the spatial world



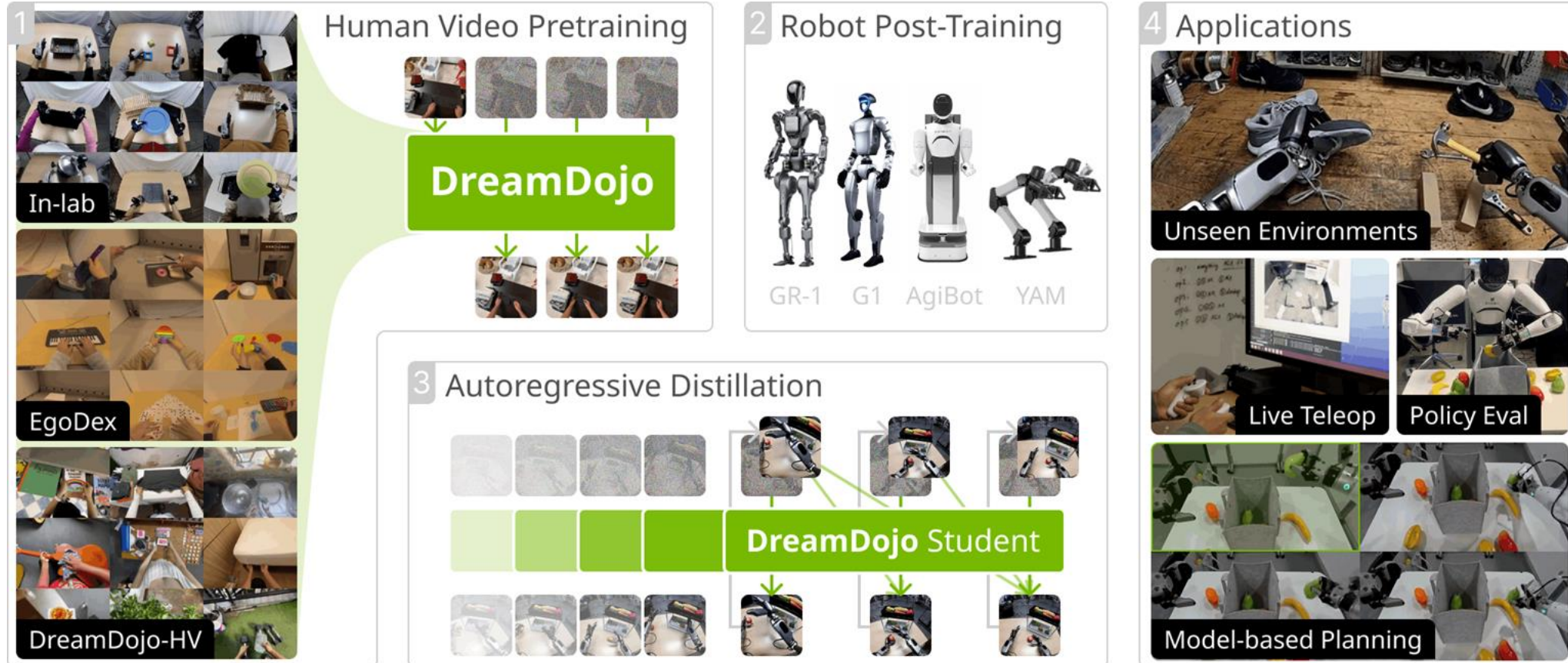
Spatial Control

Temporal Dynamics

Hybrid-dimension Evolution

Our increasing ability in generating control actions

Generalist Robot World Model - DreamDojo



S. Gao, et. al, "DreamDojo: A Generalist Robot World Model from Large-Scale Human Videos," ICML 2026. (Spotlight)

Our increasing ability in generating control actions

- Generalize to Novel Scenarios

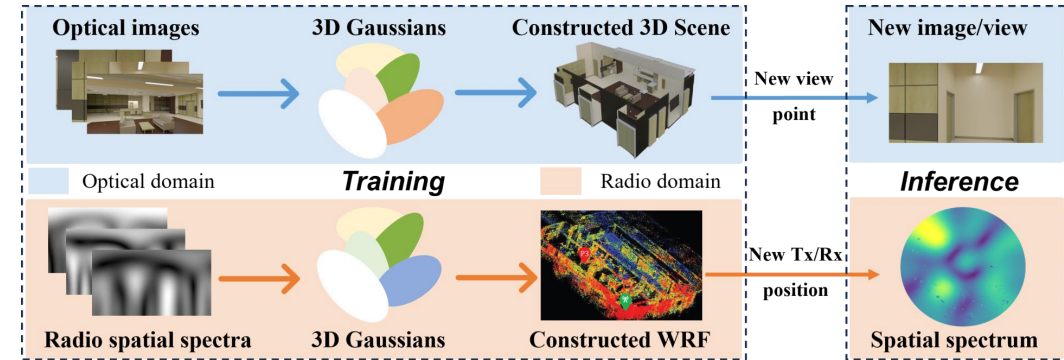


My interest

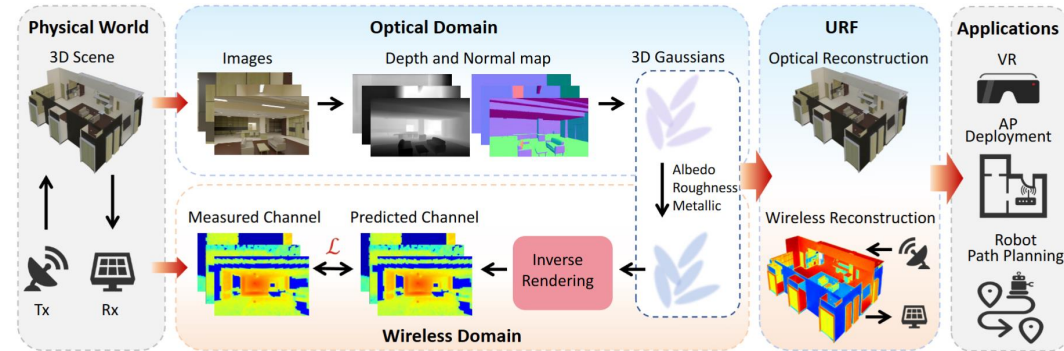
- **To understand how the world works**
 - *How to reconstruct or create a physically realistic world?*
- **How to apply in wireless communications?**
 - *How to reconstruct wireless radiation field, with minimal measurements?*
 - *How to leverage multi-modal environment data?*
 - *Go beyond LLM-based approaches*

Spatial Intelligence for Wireless (SI4Wireless)

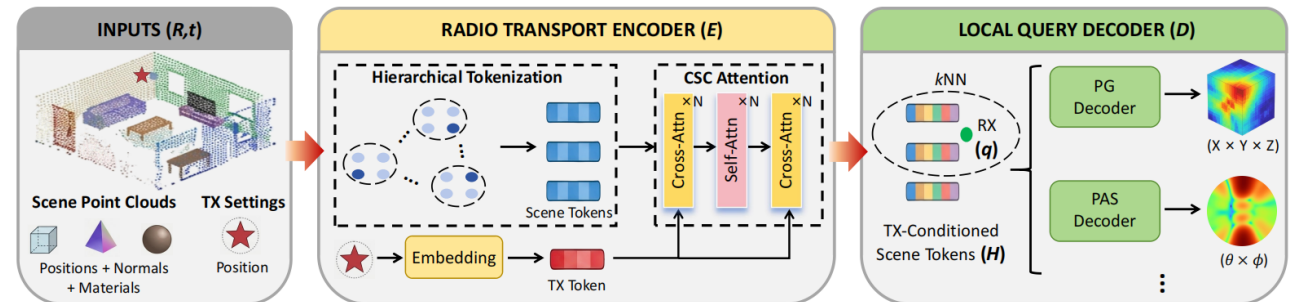
[WRF-GS]
Single-Scene Wireless Radiation
Field Reconstruction



[URF-GS]
A Unified Radio-Optical Radiation Field
for Single-Scene Location Generalization

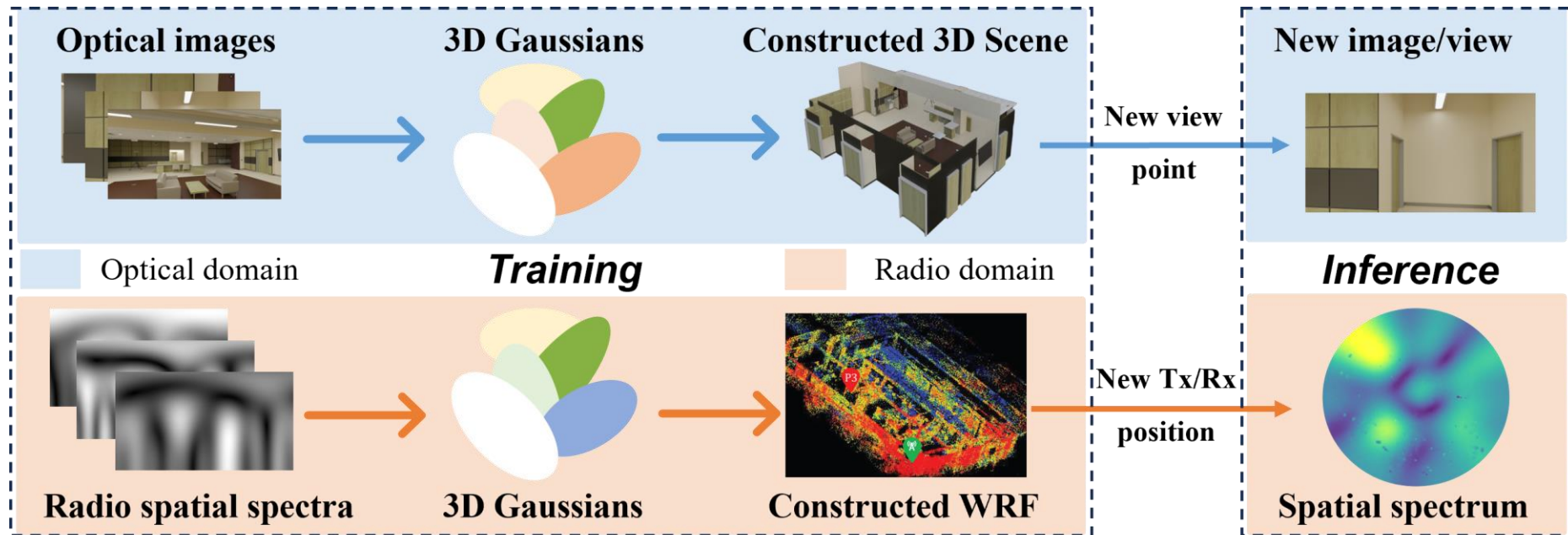


[Point2Radio]
A Wireless Foundation Model for Cross-
Scene Generalization



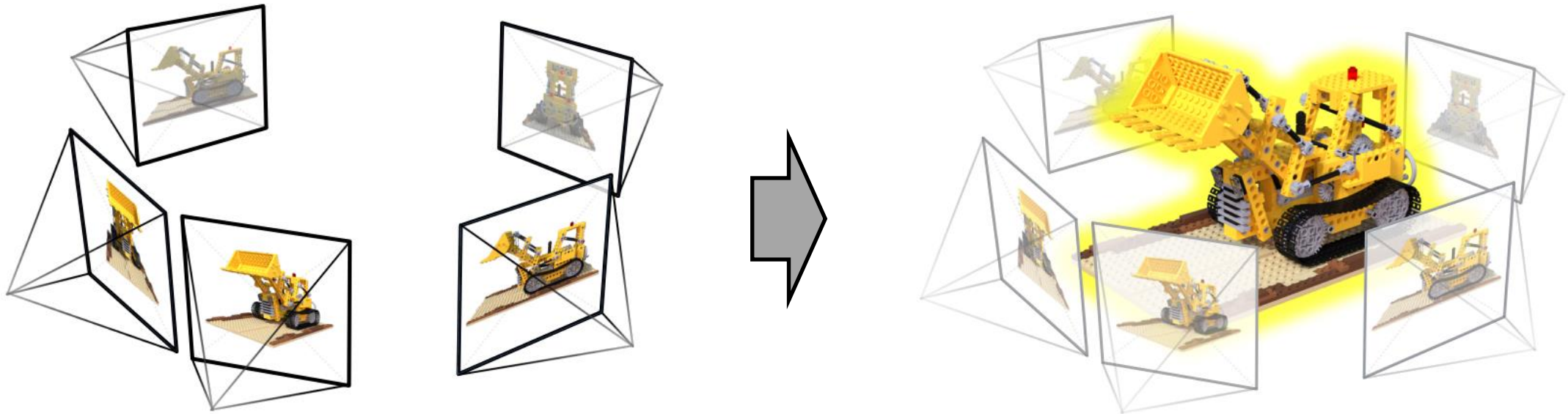
WRF-GS:

Wireless radiation field reconstruction via 3D-GS



3D reconstruction and rendering

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

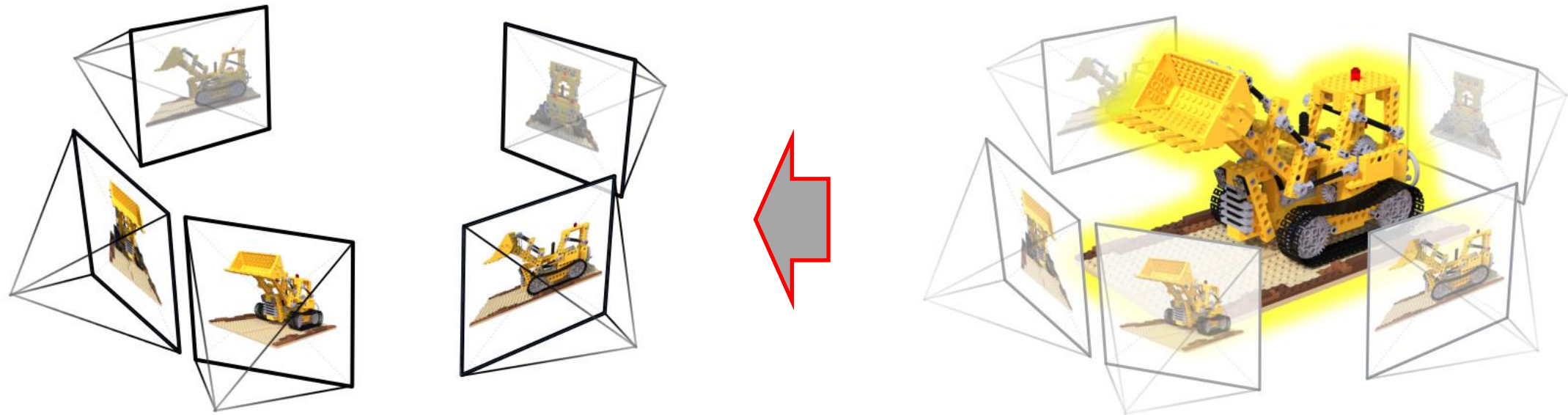


Images & Accurate Camera Poses

3D Scene Representation

3D reconstruction and rendering

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



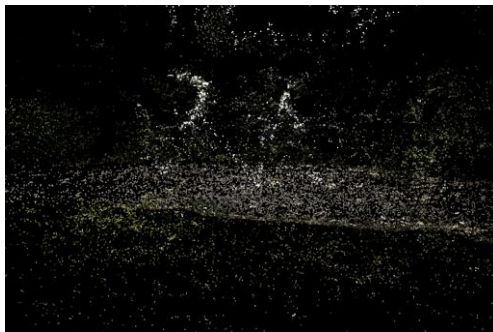
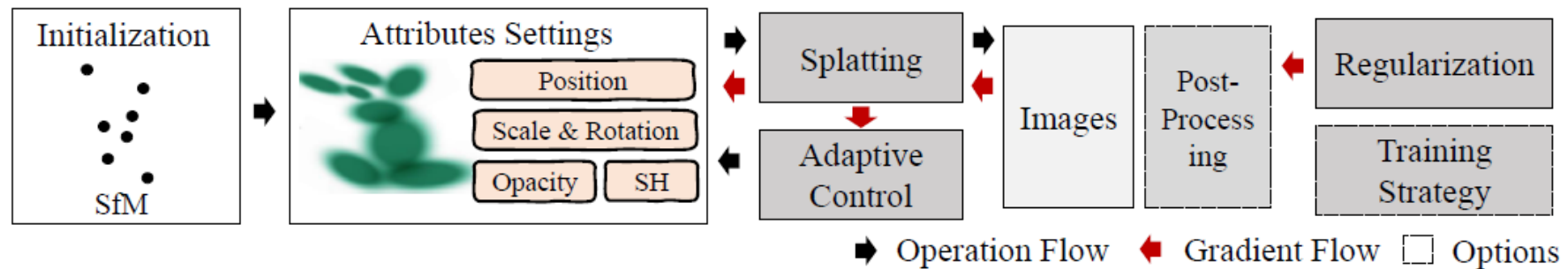
Images & Accurate Camera Poses

3D Scene Representation

3D Gaussian Splatting

- 3D Gaussian Splatting

- The properties of Gaussians are optimized based on stochastic gradient descend, supervised by the rendering quality of available views.
- Tile-based rasterization and splatting, GPU-friendly.



Initialized points



7 million Gaussians



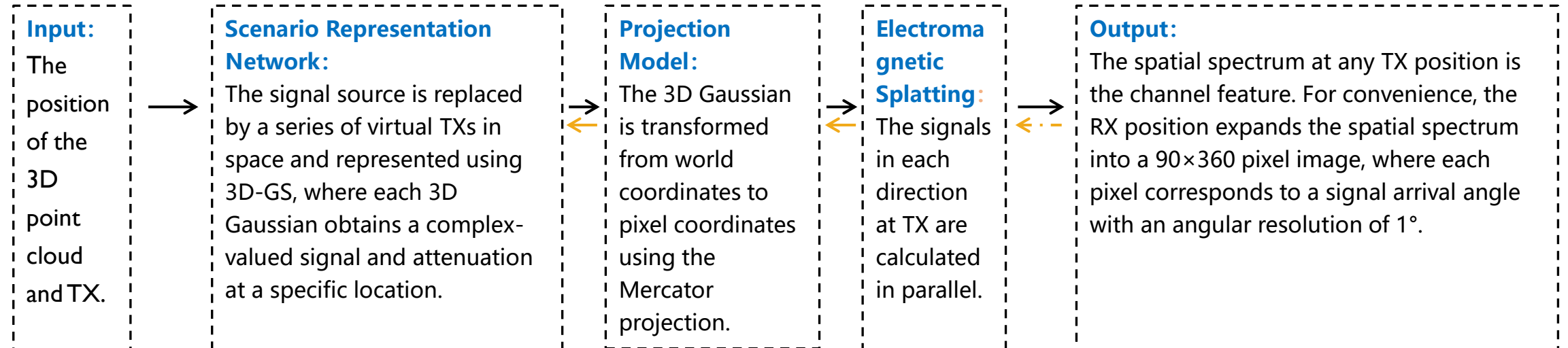
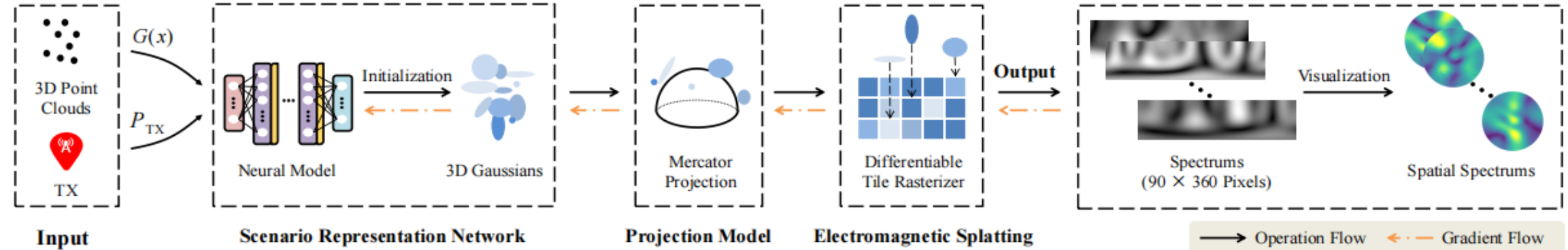
Final optimized result

Reconstruction of wireless radiation field



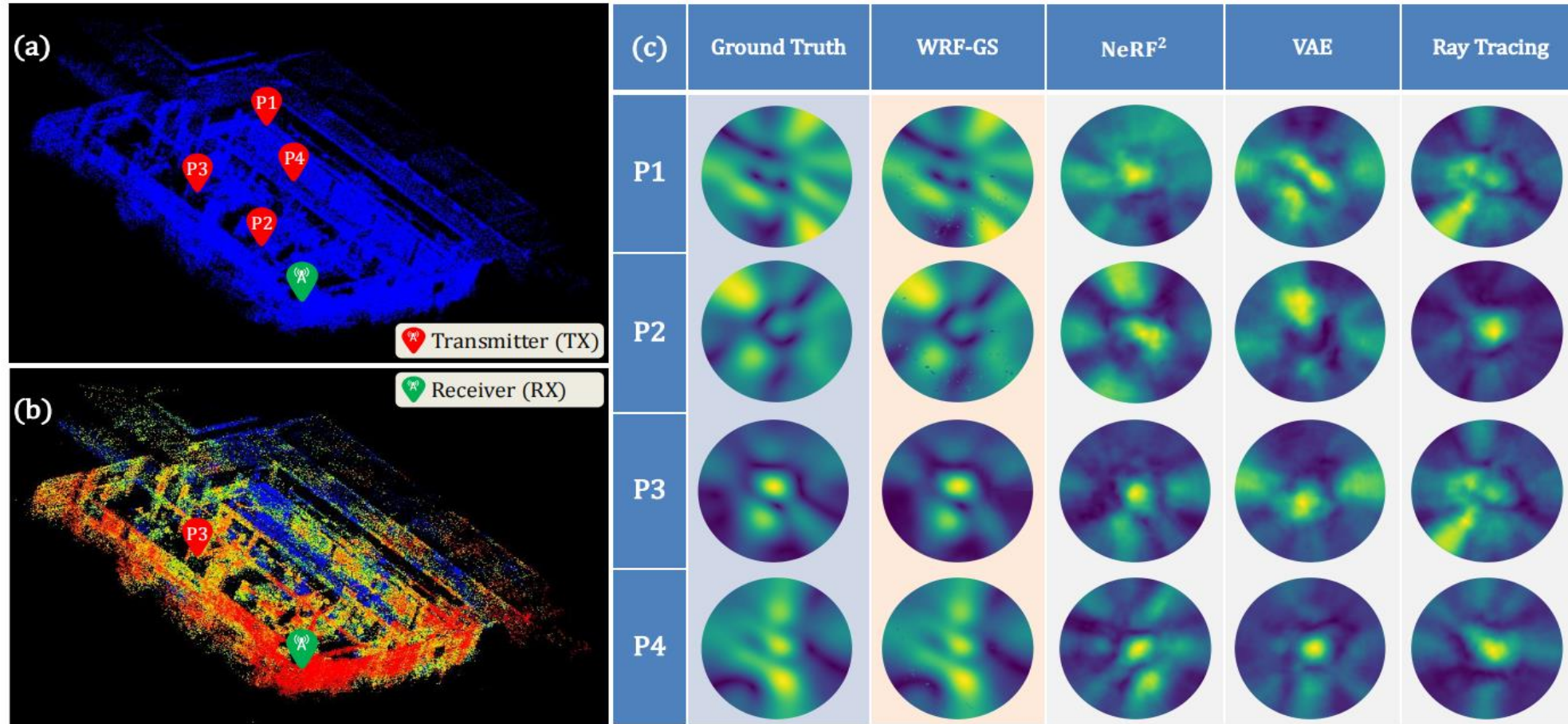
Since light is a form of electromagnetic (EM) wave, we can use optical 3D reconstruction techniques for wireless channel modeling.

Our framework: WRF-GS



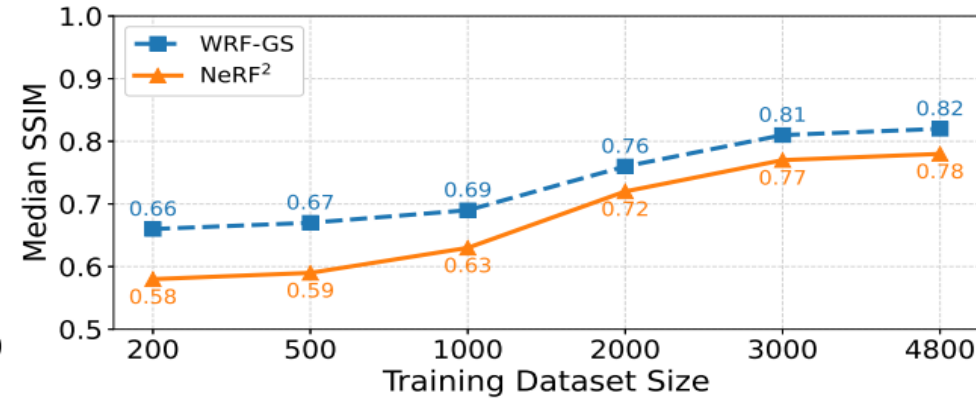
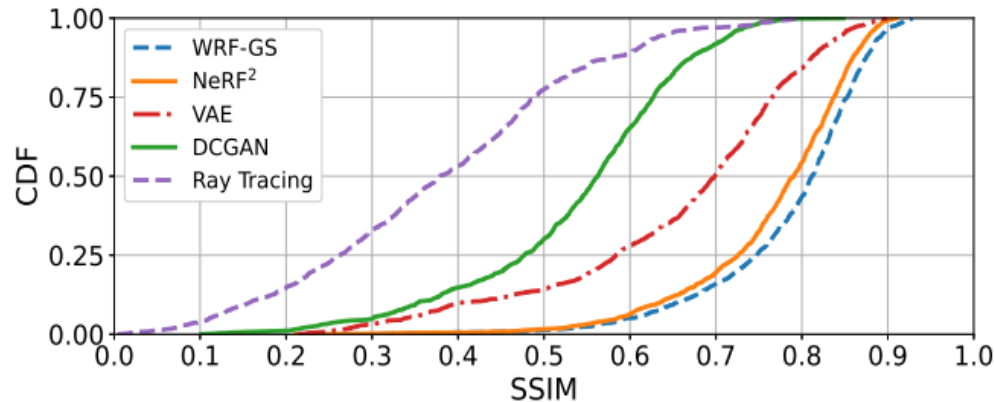
Experiment: WRF reconstruction

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



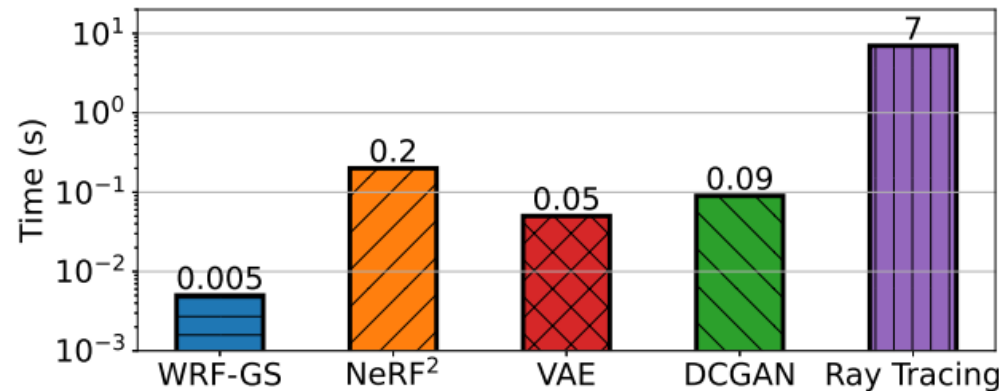
Experiment: WRF reconstruction

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



(GitHub: WRF-GS)

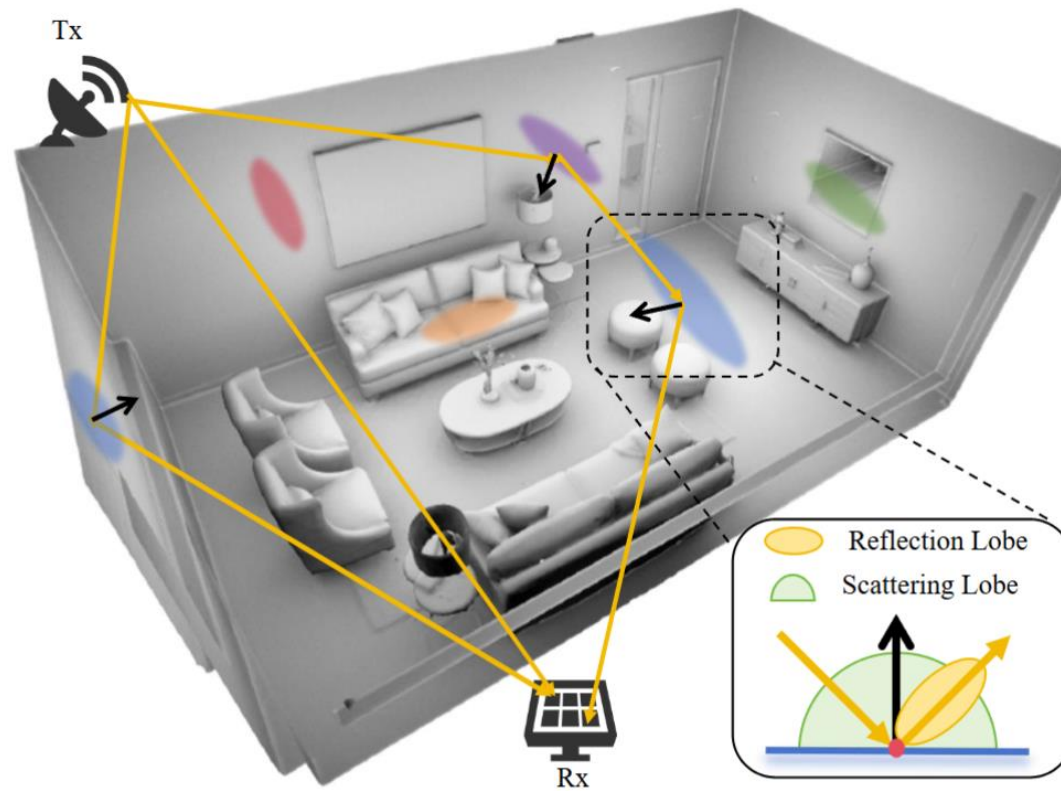
- The rendering time required to synthesize a spatial spectrum using different methods under the same settings.



(GitHub: WRF-GS+)

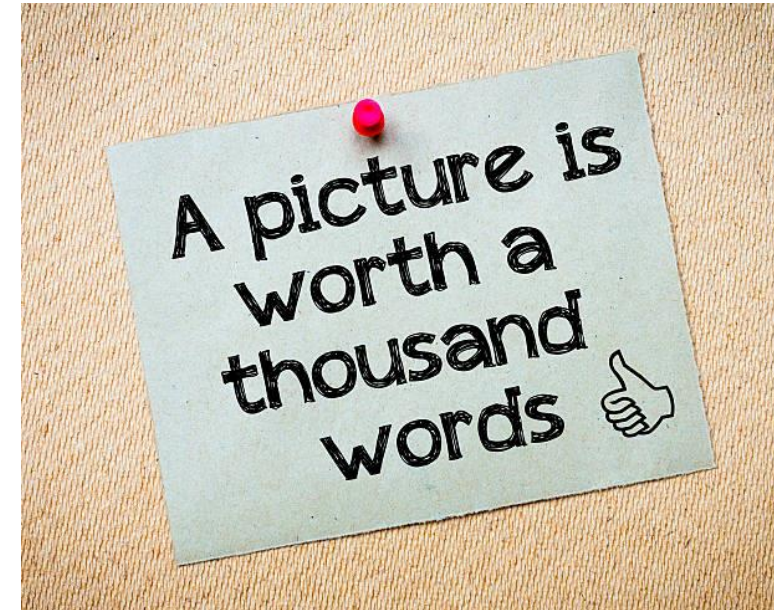
URF-GS:

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



Is radio measurement sufficient?

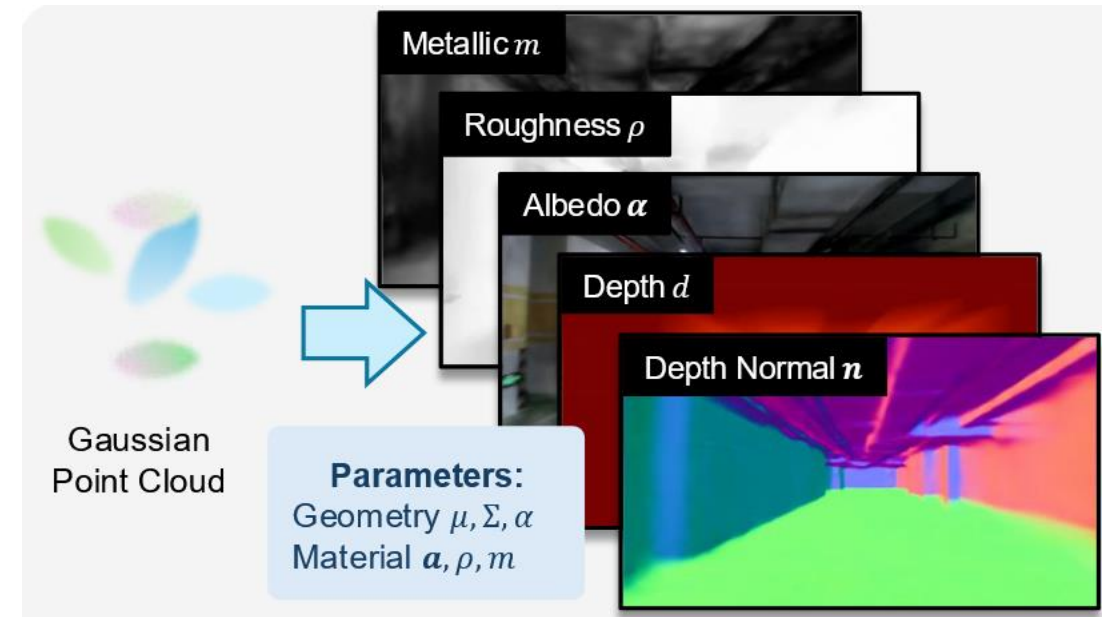
- Radio measurement is limited
 - Costly
 - Scarce
 - Noisy
 - Limited information
- Other modalities for help
 - Images capture much richer information!
 - Much easier to obtain!



Inverse rendering to exploit rich vision information

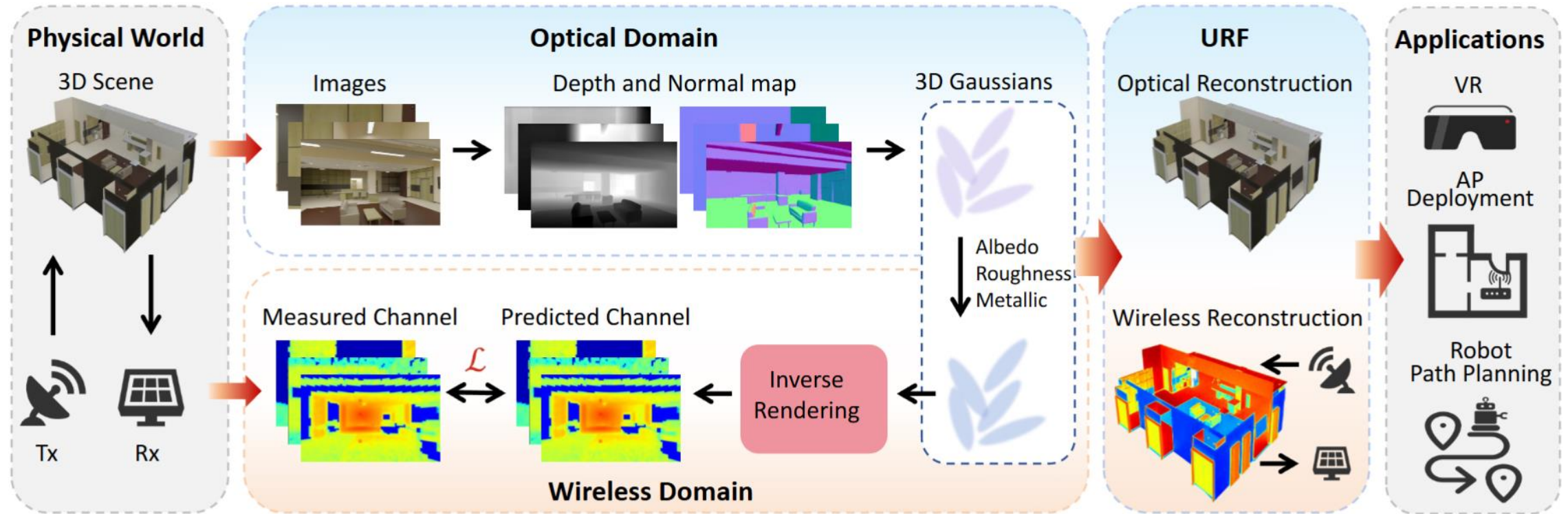
□ Why inverse rendering?

- 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.

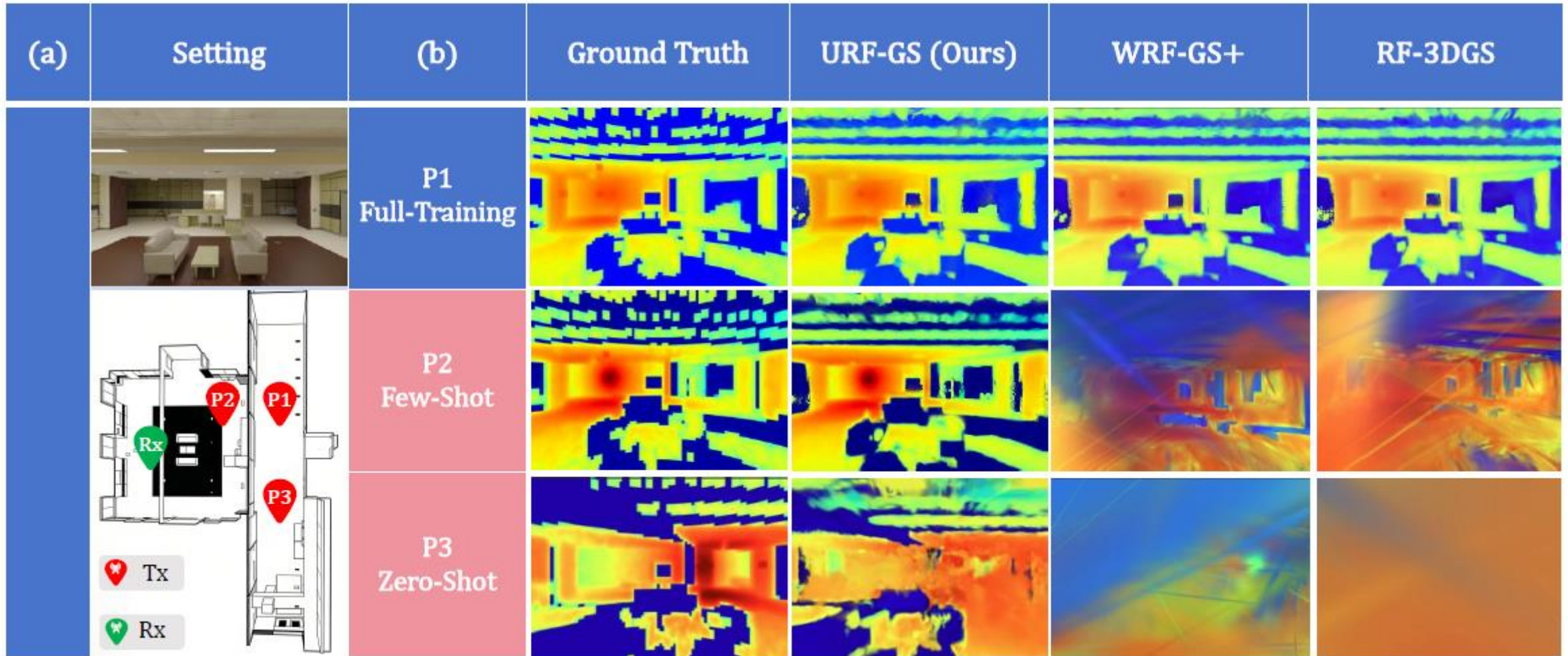


We propose **URF-GS**, a wireless channel **inverse rendering** framework based on **3D-GS**, to support scene **generalization ability**!

URF-GS: Framework



Experiment: 3D radio map construction

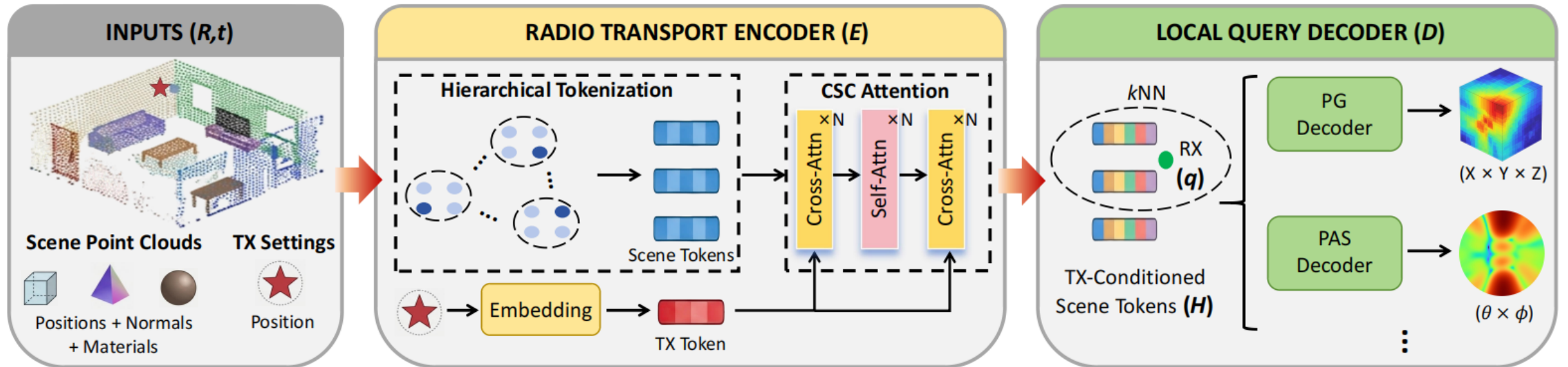


Experiment: 3D radio map construction

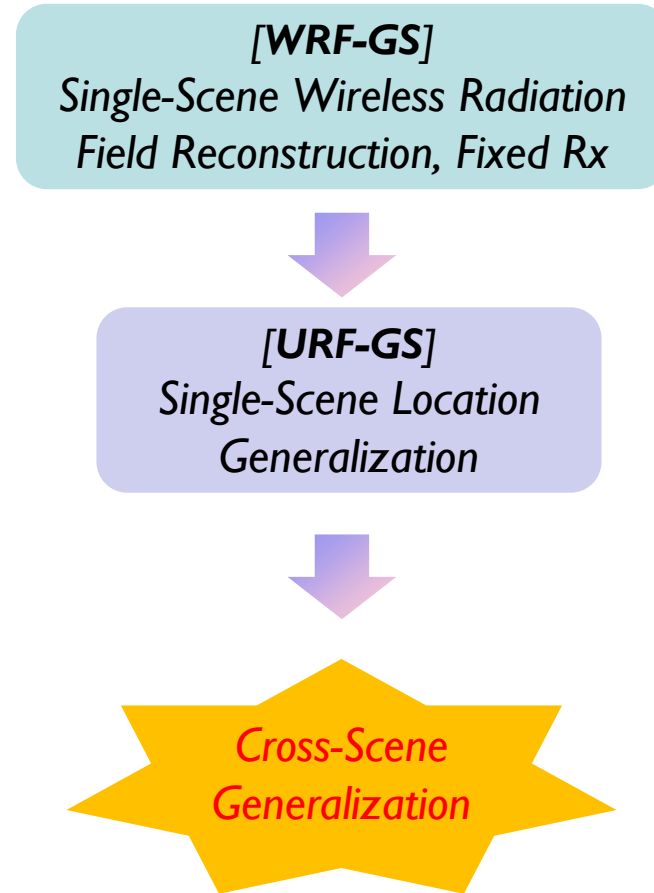
Training Data	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Full-Training	URF-GS (Ours)	17.3818	0.7012	0.3340
	RF-3DGS [12]	16.9949	0.6823	0.3345
	NeRF ² [10]	17.0581	0.5623	0.3979
	WRF-GS+ [5]	17.1085	0.6917	0.3270
Few-Shot	URF-GS (Ours)	12.5612	0.5035	0.5036
	RF-3DGS	8.4301	0.3358	0.7266
	NeRF ²	7.3301	0.3876	0.7777
	WRF-GS+	10.9804	0.4859	0.5639
Zero-Shot	URF-GS (Ours)	9.3316	0.3156	0.7056
	RF-3DGS	7.0124	0.2794	0.8303
	NeRF ²	6.5123	0.2846	0.8202
	WRF-GS+	7.3512	0.2946	0.7564

Point2Radio:

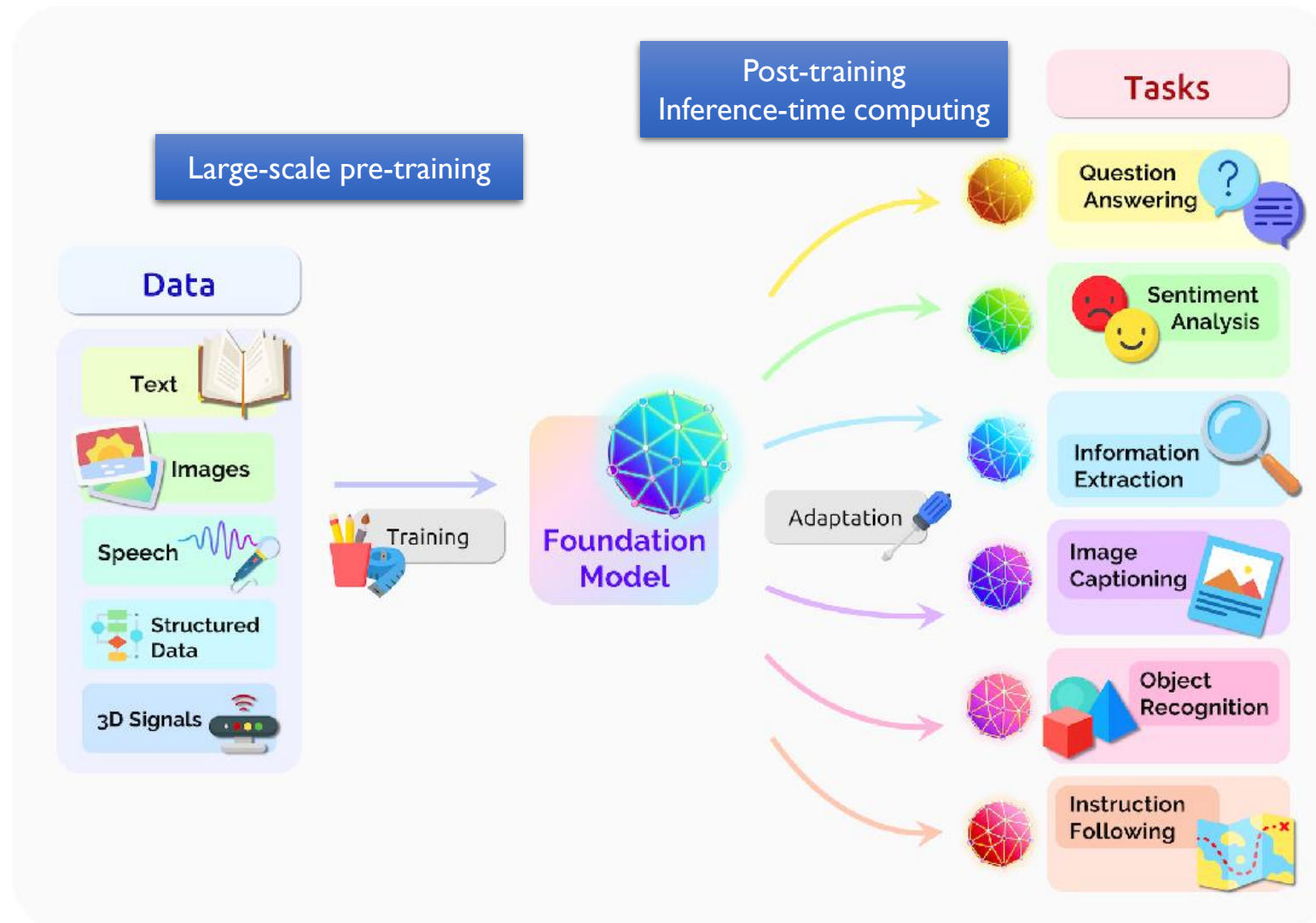
A Foundation Model for **Cross-Scene** Radio Fields from Material-Aware Point Clouds



From **per-scene training** to **cross-scene generalization**



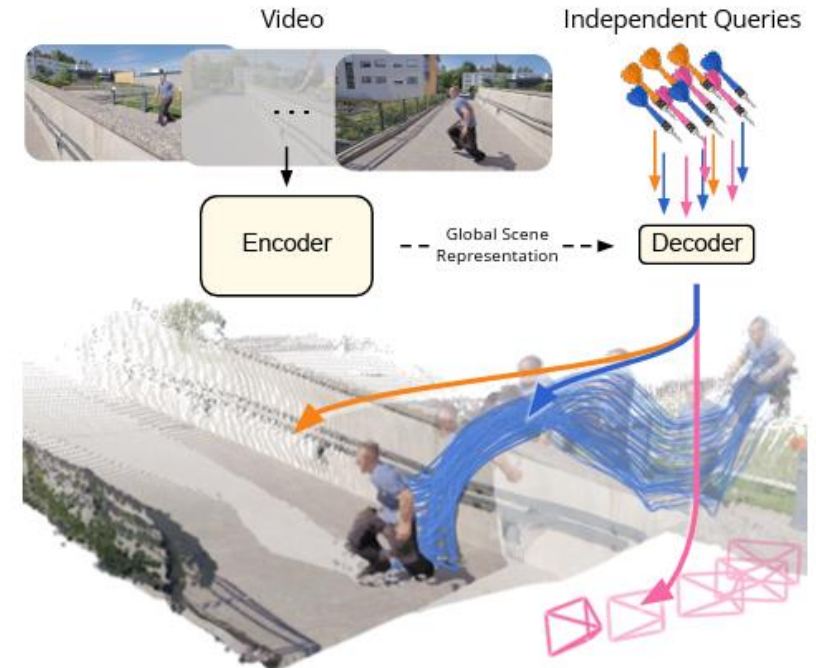
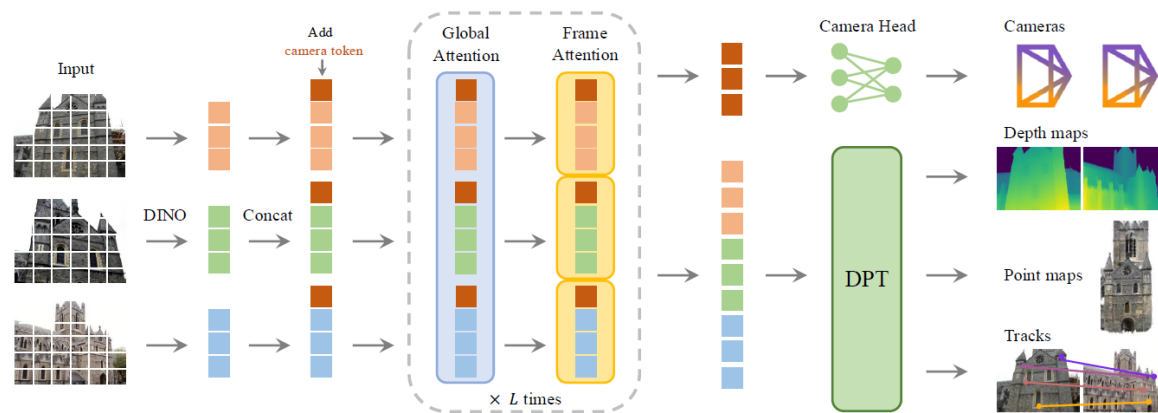
The era of foundation models



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.

$$\mathbf{H} = \mathbf{E}(\mathbf{R}, \mathbf{t})$$

$$\hat{\mathbf{y}} = \mathbf{D}(\mathbf{H}, \mathbf{q})$$

R

Material-aware point cloud
(geometry + EM attributes)

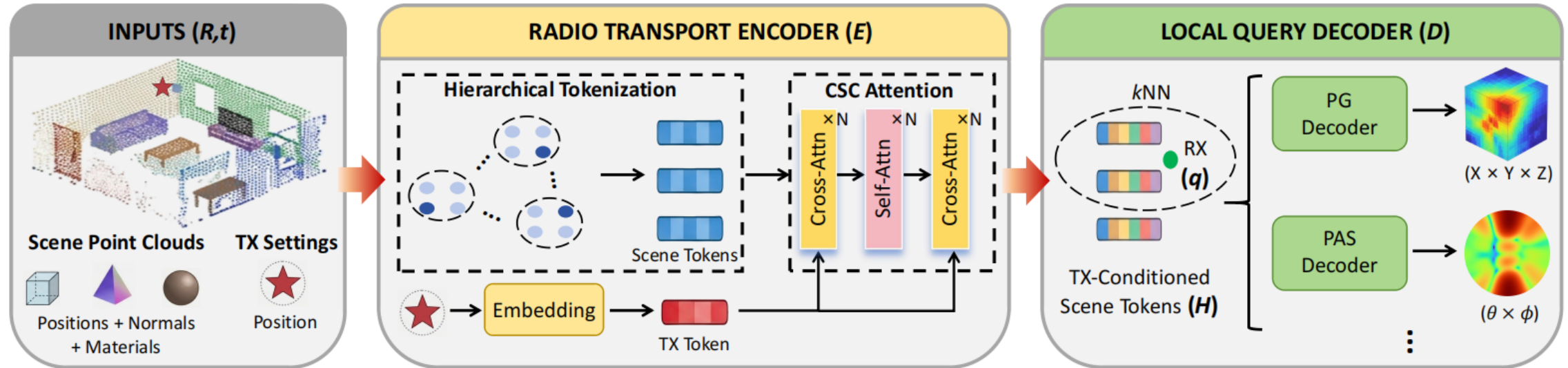
E

Radio transport encoder
→ TX-conditioned codes $\mathbf{H}$

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 $\rightarrow$ train PAS head; optional residual FT per scene

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

PRISM dataset

Procedural Radio–Image Scene Multimodality — Infinigen geometry, Sionna RT labels.

337

Indoor scenes

scene-disjoint split

86,272

TX-conditioned PG fields

3.5 GHz mid-band

15

ITU material classes

ϵ , σ from ITU-R P.2040

269/33/35

Train / Val / Test

held-out rooms

PRISM-PG

Dense 16^3 PG volumes per TX; surface point clouds as model inputs; meshes only for offline labeling.

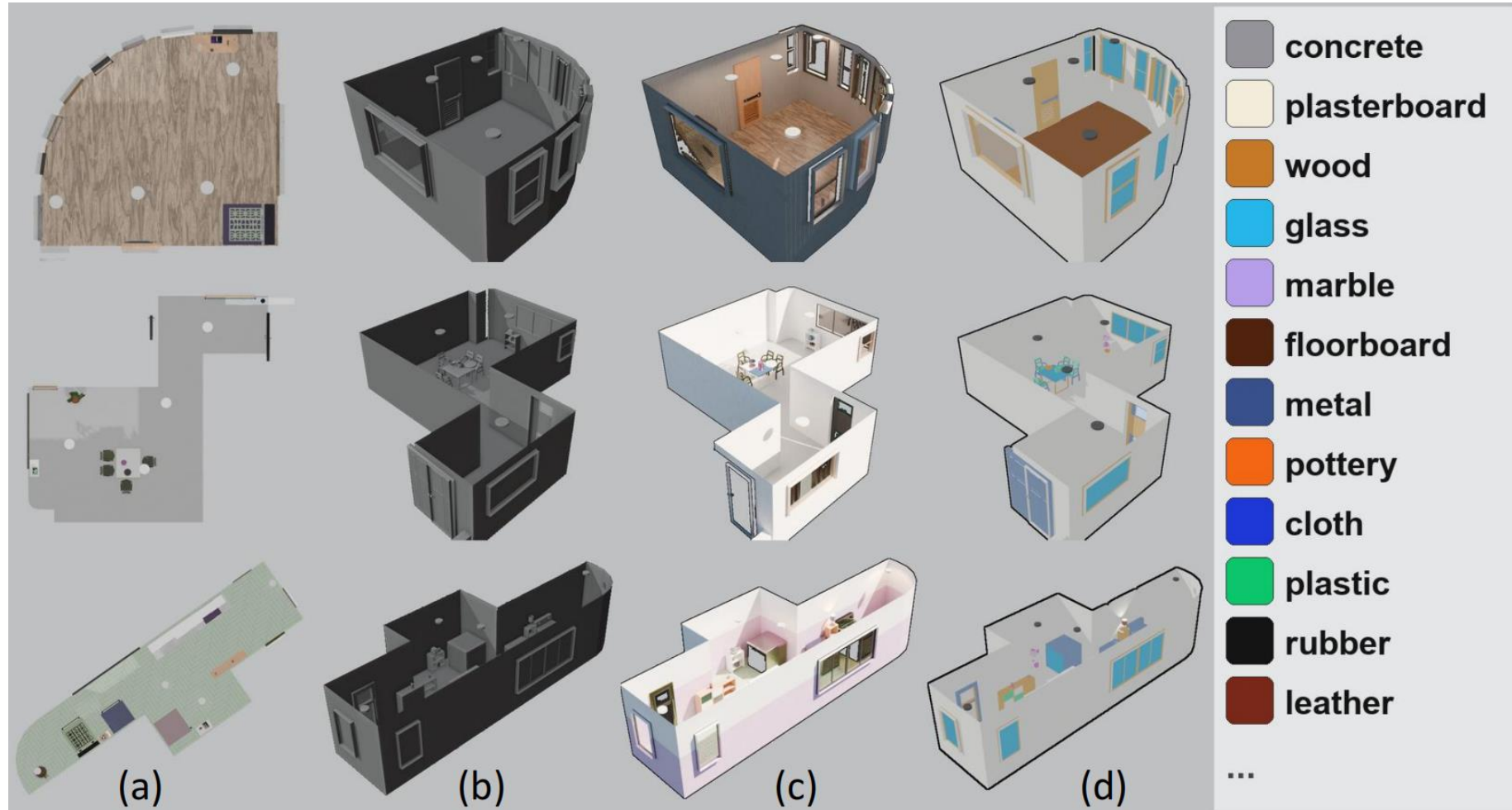
PRISM-PAS

20×20 TX–RX pairs; 4×4 array beamforming → 360×90 az–el spectra; 199 overlapping rooms.

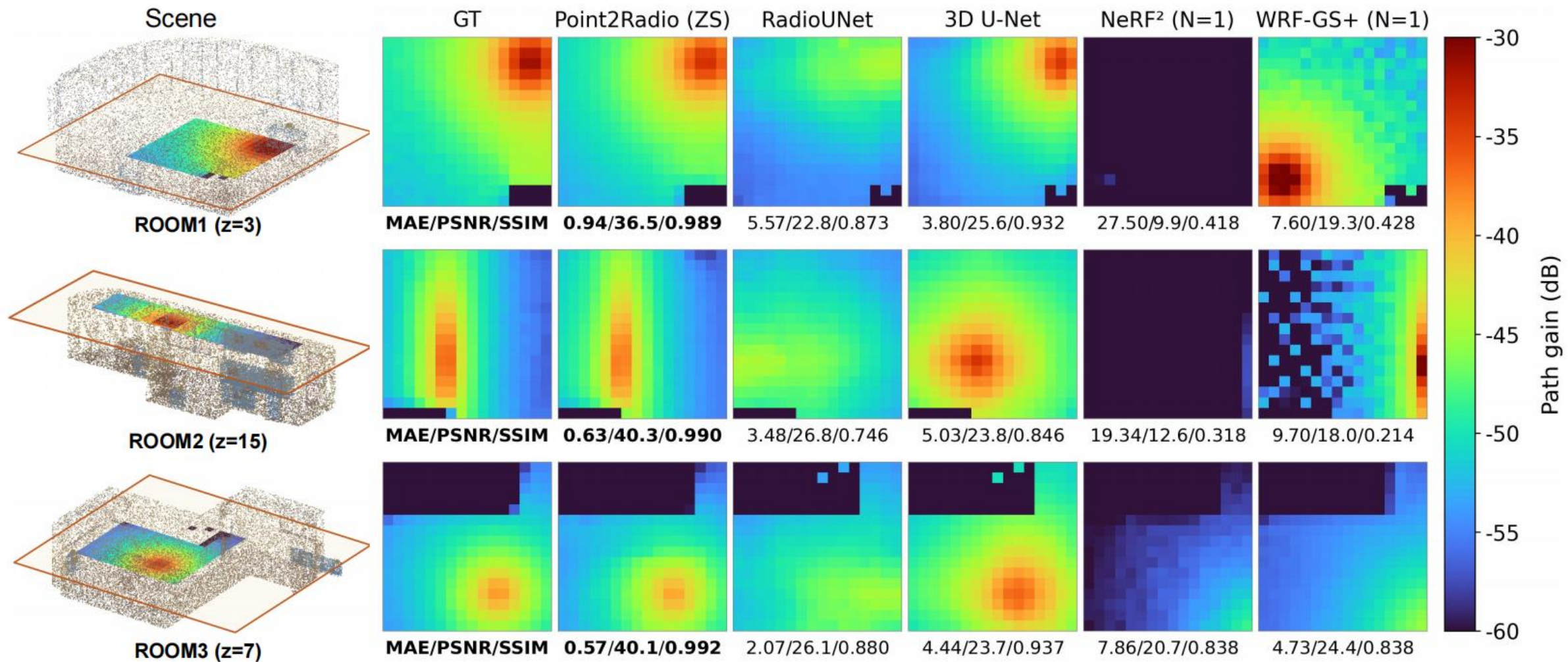
Inference inputs

Material-aware point cloud + transceiver queries only — no mesh, no online path tracing.

PRISM dataset



Experiment: Cross-scene path gain



Experiment: Cross-scene 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

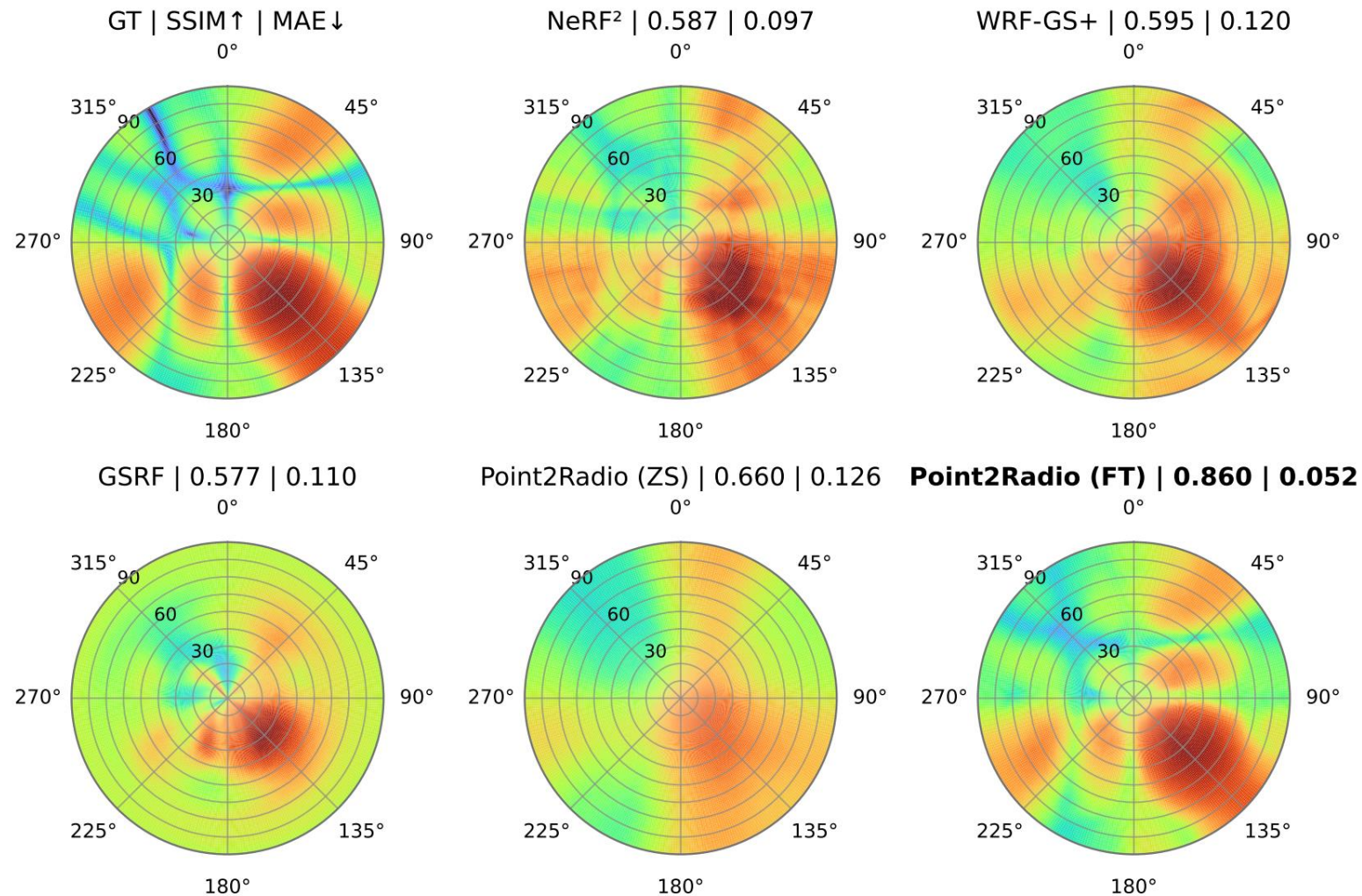
0.871 dB

Zero-shot MAE on held-out rooms

0.954

SSIM — strong spatial structure

Experiment: Power angular spectra



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
 - To enable intelligent wireless networks





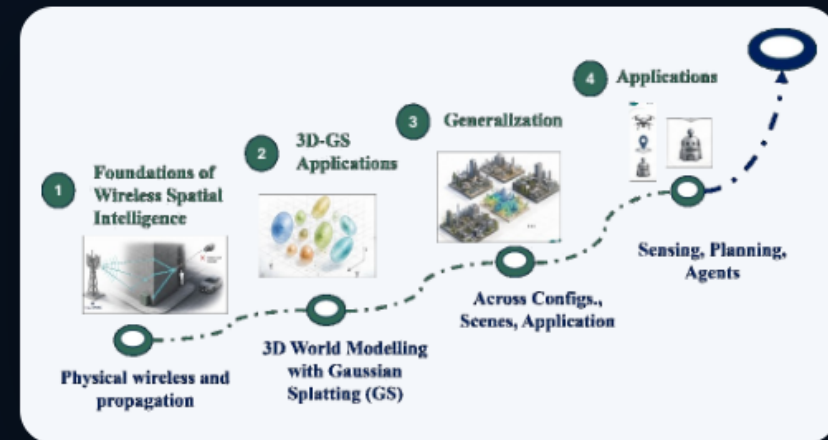
Spatial Intelligence for Wireless Communications and Sensing

Methods to Applications

From wireless radiation fields to world models: a half-day tutorial by HKUST iComAI Lab on reusable spatial knowledge for 6G communications, sensing, and embodied agents.

[View Tutorial Outline](#)

[Download Slides \(PDF\)](#)



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*, 2026.
- 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)

References

• Generative World Models

- F. Liu*, W. Sun*, H. Wang*, Y. Wang, H. Sun, J. Ye, **J. Zhang**, and Y. Duan, “ReconX: Reconstruct any scene from sparse views with video diffusion model,” *T-IP 2026*.
- W. Sun, S. Chen, F. Liu, Z. Chen, Y. Duan, **J. Zhang**, and Y. Wang, “DimensionX: Create any 3D and 4D scenes from a single image with decoupled video diffusion,” *ICCV 2025*.
- W. Sun, H. Zhang, H. Wang, J. Wu, Z. Wang, Z. Wang, Y. Wang, **J. Zhang**, T. Wang, and C. Guo, “WorldPlay: Towards Long-Term Geometric Consistency for Real-Time Interactive World Modeling,” *ICML 2026*.

• Embodied World Models

- J. Yang*, S. Gao*, Y. Qiu*, L. Chen, T. Li, B. Dai, K. Chitta, P. Wu, J. Zeng, **J. Zhang**, A. Geiger, Y. Qiao, and H. Li, “Generalized predictive model for autonomous driving,” *CVPR 2024*. (**Highlight, Top 2.8%**)
- S. Gao, J. Yang, L. Chen, K. Chitta, Y. Qiu, A. Geiger, **J. Zhang**, and H. Li, “Vista: A generalizable driving world model with high fidelity and versatile controllability,” *NeurIPS 2024*.
- S. Gao, S. Zhou, Y. Du, **J. Zhang**, and C. Gan, “AdaWorld: Learning adaptable world models with latent actions,” *ICML 2025*.
- S. Gao, et. al, “DreamDojo: A Generalist Robot World Model from Large-Scale Human Videos,” *ICML 2026*. (**Spotlight, Top 2.2%**)



Wenqiang SUN
(Tencent Hunyuan, PhD HKUST)



Shenyuan GAO
(NVIDIA GEAR, PhD HKUST)

Thank you!



- For more details

<https://eejzhang.people.ust.hk/>



- Our WeChat group account