BRIDGING THE SIM-TO-REAL GAP IN RF LOCALIZATION WITH LARGE-SCALE SYNTHETIC PRETRAINING*

Armen Manukyan[1,2], **Rafayel Mkrtchyan**[1,2], **Ararat Saribekyan**[1,2],
Theofanis P. Raptis[3], **Hrant Khachatrian**[1,2]
[1]Yerevan State University, Armenia
[2]YerevaNN, Armenia
[3]Institute of Informatics and Telematics, National Research Council, Italy
{armen.manukyan, rafayel.mkrtchyan, ararat.saribekyan}@ysu.am,
theofanis.raptis@iit.cnr.it, hrant@yerevann.com

ABSTRACT

Radio frequency (RF) fingerprinting is a promising localization technique for GPS-denied environments, yet it suffers from poor generalization to unmapped areas. Traditional k-nearest neighbor methods perform well where data exists but fail on unseen streets. Deep learning can learn generalizable spatial-RF patterns, but requires far more training data than typical measurement campaigns provide. We investigate whether synthetic data can bridge this gap. Using a real-world dataset from Rome and NVIDIA's Sionna ray-tracing simulator, we generate synthetic datasets under varying fidelity and scale: Dataset B′ uses real base station (BS) locations with Gaussian Process-calibrated signals (53K samples), while Dataset C uses fully simulated BSs and signals (274K samples). Our evaluation reveals a pronounced sim-to-real gap—models achieving 25m error on synthetic data degrade to 184m on real data—yet pretraining on synthetic data reduces real-world error from 323m to 162m, a 50% improvement. Notably, simulation fidelity proves more important than scale: the smaller calibrated dataset outperforms the larger uncalibrated one. We further evaluate cross-city generalization on an unseen Oslo dataset, achieving 132m zero-shot RMSE and 62m after fine-tuning. This work provides a systematic study of synthetic-to-real transfer for RF localization, highlighting the value of simulation-aware pretraining.

1 INTRODUCTION

Radio frequency (RF) fingerprinting has gained significant traction for localization in GPS-denied environments, including urban canyons and indoor spaces Zafari et al. (2019). The approach builds a database of signal measurements at known locations, then matches new measurements to estimate position Yaro et al. (2023). While modern cellular standards provide localization facilities (Cell-ID, timing advance, OTDOA, 5G NR positioning), these face practical limitations: operator APIs are often restricted Cha et al. (2025); Wymeersch et al. (2017), coarse methods yield hundreds-of-meters accuracy, and advanced methods struggle under multipath and NLOS conditions Xie et al. (2024); Koivisto et al. (2017). In emergency, adversarial, or private contexts, RF fingerprinting offers a UE-side, infrastructure-agnostic alternative.

Traditional k-nearest neighbor (k-NN) methods work well for mapped locations but fail on unseen streets, with errors exceeding thousands of meters Miramá et al. (2021). Deep learning (DL) can potentially learn generalizable spatial-RF patterns Khachatrian et al. (2024), but requires far more data than real-world campaigns provide. The Rome dataset Ali et al. (2022) (Dataset A) exemplifies this, containing only $\sim$2,000 measurements. While synthetic datasets like WAIR-D Huangfu et al. (2022) address data scarcity, the sim-to-real transfer problem remains largely unexplored: do models trained on synthetic RF data work on real measurements?

*This work was previously published in the journal *Information Fusion* (Manukyan et al., 2025b) and is included here under the workshop's non-archival policy.

We investigate whether large-scale synthetic data can bridge this generalization gap using the Sionna ray-tracing simulator Hoydis et al. (2023). We create Dataset B (real BS locations, simulated signals), Dataset B′ (calibrated version of B via Gaussian Process optimization), and Dataset C (fully simulated BSs and signals at city scale). We address three questions: (1) Can synthetic pretraining improve generalization to unseen real locations? (2) Does data quantity or fidelity matter more? (3) How large is the sim-to-real gap, and can transfer learning bridge it?

Our contributions include:

- **Sim-to-real evaluation**: The first systematic study in wireless communication of synthetic-to-real transfer in RF localization, revealing that models achieving 25m synthetic error degrade to 184m on real data.
- **Gaussian Process calibration**: A methodology for calibrating synthetic BS parameters against real measurements, improving per-BS correlation by 0.15–0.25.
- **Large-scale synthetic pretraining**: Demonstration that pretraining on 274K synthetic samples enables 50% error reduction (323m→162m) over real-only training.
- **Scale vs. quality trade-off**: Evidence that 53K calibrated samples often outperform 274K uncalibrated ones, highlighting simulation fidelity over quantity.

2 Related Work

Early outdoor fingerprinting relied on proprietary or small-scale drive tests. Kousias *et al.* released the first openly documented city-scale corpus with 4G, NB-IoT, and 5G measurements in Rome and Oslo Kousias et al. (2024), and De Nardis *et al.* built a k-NN benchmark on these data De Nardis et al. (2023). Such campaigns remain costly and geographically narrow.

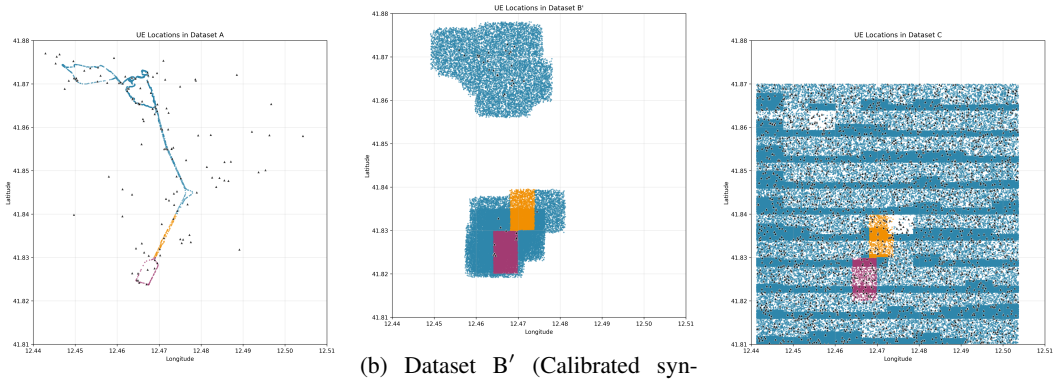
Synthetic data generation has expanded training possibilities. The WAIR-D corpus Huangfu et al. (2022) spans thousands of urban scenes with ray-traced channels. Proprietary simulators—WinProp Altair Engineering Inc. (2021), Wireless InSite Remcom Inc. (2020); Alkhateeb (2019), and NewFasant Gonzalez et al. (2008)—have enabled multipath-aware fingerprinting and mmWave studies. Progress has accelerated with open-source tools like Sionna Hoydis et al. (2023).

Several works demonstrate strong synthetic-only results: Nokia Bell Labs achieved $\sim$1.4m error using WinProp-generated beam-level RSRP fingerprints Butt et al. (2020); Bhattacherjee *et al.* showed sub-meter synthetic accuracy with AOA+RSS+TOA features Bhattacherjee et al. (2020); Del Peral-Rosado *et al.* reported sub-meter errors with delay-based fingerprints from NewFasant Del Corte-Valiente et al. (2019); and Khachatrian *et al.* reached 11.3m NLOS RMSE on WAIR-D Khachatrian et al. (2025). However, none include real-world validation. De Sousa *et al.* de Sousa et al. (2021) represent the closest work, pretraining Random Forest models on WinProp-generated features and reducing mean error from 238m to 149m with 1,000 real calibration samples, though without detailed environmental maps. Network-assisted positioning techniques leverage existing cellular infrastructure for UE localization Shah et al. (2025). Methods like OTDOA utilize time difference measurements from multiple BSs for triangulation, offering improved accuracy with good coverage but limited by cell density and NLOS conditions. Model-based multilateration (TOA, TDOA) achieves high accuracy but requires synchronized clocks and is sensitive to multipath Widdison & Long (2024). Hybrid systems combining GNSS with cellular data enhance reliability in challenging environments Camajori Tedeschini et al. (2023).

Current evidence suggests volume drives representation quality while realism drives sim-to-real transfer Ruah et al. (2023), but the fundamental question of whether synthetic pretraining improves generalization to unseen real locations remains unexplored Akrout et al. (2023). We address this gap with the first systematic study of synthetic-to-real transfer in RF localization.

3 Problem Formulation

We formulate UE localization as predicting a spatial probability distribution. The model receives: (i) a two-channel map $M \in \{0,1\}^{H\times W\times 2}$ encoding building and road footprints; (ii) pixel coordinates of N base stations, $\mathcal{L} = \{\mathbf{l}_i \in \mathbb{R}^2\}_{i=1}^{N}$; and (iii) RF measurements $\mathcal{R} = \{\mathbf{m}_i \in \mathbb{R}^4\}_{i=1}^{N}$, where $\mathbf{m}_i$ comprises RSSI, NSINR, NRSRP, and NRSRQ.



(a) Dataset A (Real-world)

(b) Dataset B′ (Calibrated synthetic)

(c) Dataset C (Fully synthetic)

Figure 1: UE and BS (black triangles) distributions across datasets. Blue: training, orange: validation (unknown streets), purple: test (unknown streets). Validation and test splits are shared across datasets.

The model learns $f_\theta : (M, \mathcal{L}, \mathcal{R}) \mapsto \hat{Y} \in [0,1]^{H \times W}$, a probability heatmap over UE location. Training minimizes pixel-wise BCE loss between prediction $\hat{Y}$ and ground-truth $Y \in \{0,1\}^{H \times W}$. At inference, $\hat{\mathbf{p}} = \arg\max_{(i,j)} \hat{Y}_{ij}$.

4 Data Generation

We use three datasets: **Dataset A** with real-world measurements in Rome; **Dataset B/B′** with real BS locations and simulated (B) or GP-calibrated (B′) signals; and **Dataset C** with fully simulated BSs and signals at city scale. All synthetic datasets produce RSSI, NSINR, NRSRP, and NRSRQ, matching the real data format.

4.1 Dataset A: Real-world measurements

Dataset A comprises a drive-test campaign in central Rome with ≈2,000 unique UE locations and measurements to multiple BSs Kousias et al. (2024); Ali et al. (2022). We rasterize OSM building footprints into binary maps at 1m/px and crop 601×601 tiles centered near each UE, retaining only crops with $\geq$3 visible BSs. UEs are partitioned spatially into training, validation, and test splits based on street coverage.

4.2 Simulation framework

Synthetic data is generated via Sionna RT (v1.1.0) Hoydis et al. (2023) in three stages: (i) *Scene construction*: OSM buildings extruded into 3D solids with height heuristics; (ii) *Channel synthesis*: deterministic ray tracing with diffuse scattering, 10^6 samples/source, 10^4 max paths, 3 reflections, sub-3GHz band; (iii) *Metric extraction*: RSSI, NSINR, NRSRP, NRSRQ computed over 1,008 resource elements and 84 resource blocks.

Dataset B uses real BS locations from Dataset A in two Rome scenes (11 and 27 BSs). Per scene, we sample 64,000 UEs within the central 80% of the BS-induced bounding box and trace all BS–UE channels. Dataset B′ is derived via GP calibration (Section 5).

4.3 Dataset C: Fully simulated

A $25 \times 33\text{km}^2$ region of Rome is divided into a 10×10 grid. Per patch: 10 BSs at 40m height, 5,000 UEs, full ray tracing. Two seeds yield $\sim 10^7$ links. A denser variant concentrates BSs and UEs within tighter sub-regions (Figure 2) to ensure $\geq$3 BS visibility per 600m crop. Datasets B, B′, and C follow the same spatial splits as Dataset A to prevent information leakage.

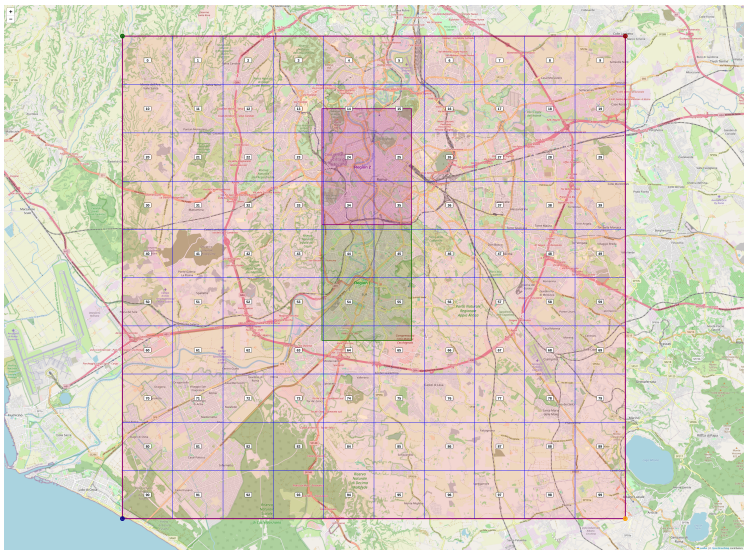


Figure 2: Grid configuration for Dataset C. Purple/green sub-grids denote dense regions maximizing multi-BS visibility within 600m crops.

4.4 OSLO DATASET

To evaluate cross-city generalization, we use the Oslo dataset Kousias et al. (2024) with 5,266 UE locations and 982 BSs (389,124 UE–BS pairs). It serves as both a zero-shot benchmark and a fine-tuning target. Details and splits are in Appendix A.

5 DATASET B′: REDUCING THE SIM-TO-REAL GAP

5.1 IDENTIFYING IMPACTFUL SIMULATION PARAMETERS

To close the sim-to-real gap, we first determined which parameters most affect simulation fidelity. Using 6 BSs with manually verified locations, we performed hundreds of simulations varying configurations and evaluated via (i) Spearman correlation between simulated and real RSSI, and (ii) k-NN localization using four train/test combinations of real [R] and simulated [S] fingerprints: [R→R], [S→S], [R→S], and [S→R]. Simulation-wide parameters (path depth, scattering flags, frequency) showed negligible sensitivity, while BS characteristics (location, altitude, radiation pattern, orientation) were decisive Manukyan et al. (2025a).

5.2 OPTIMIZING BASE STATIONS VIA GAUSSIAN PROCESS

For each BS we optimize four variables: lateral offsets $(x, y) \in [-50, 50]^2$m, height $h \in [20, 100]$m, and azimuth $\phi \in [0, 360]^\circ$. The objective maximizes Spearman correlation between simulated and measured RSSI using expected-improvement acquisition over 90 evaluations per BS. This improves per-BS correlation by 0.15–0.25, yielding Dataset B′.

6 MODEL ARCHITECTURE

6.1 INPUT ENCODING

Each BS is represented by a 6D vector: normalized (x, y) coordinates plus four RF measurements (RSSI, NSINR, NRSRP, NRSRQ). The map crop (601×601) is resized to 224×224; the building map is duplicated across two channels, with the third encoding relative spatial scale. BS locations are overlaid as 7×7 blobs encoding coordinates and RSSI.

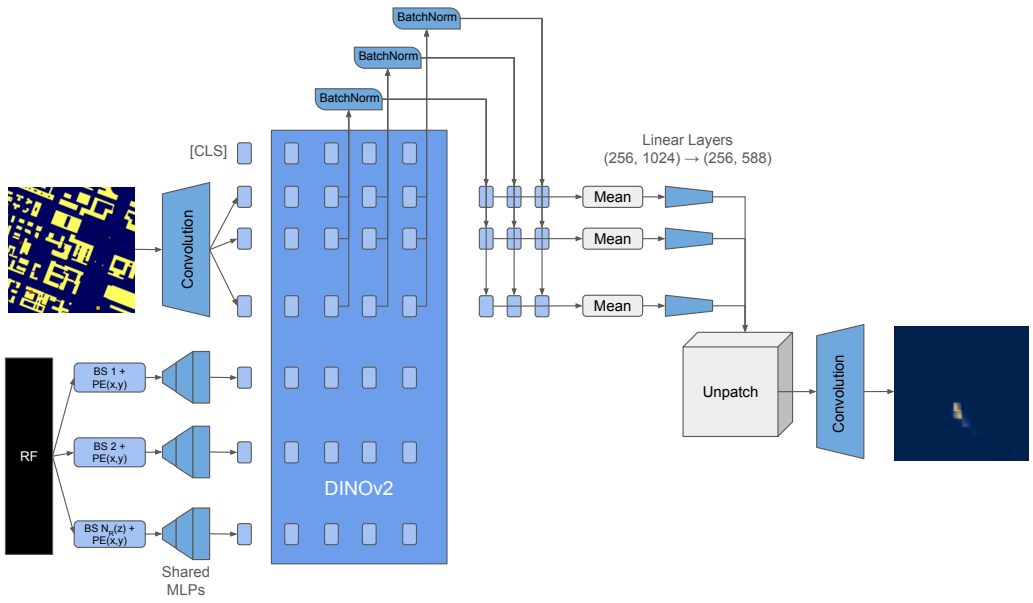


Figure 3: MapRadioFormer+ architecture, adapted from Mkrtchyan et al. (2025).

6.2 MapRadioFormer+ architecture

RF vectors are processed by a shared MLP (256→1,024 neurons, ReLU) to produce *radio tokens* matching the DINOv2 Oquab et al. (2024) hidden dimension. Positional encodings based on BS (x, y) coordinates are added. The map is decomposed into $16 \times 16 = 256$ patch tokens, concatenated with radio tokens, and processed by DINOv2-L/14.

Output image-patch tokens are batch-normalized per layer, averaged across layers, projected to $14 \times 14 \times 3 = 588$ dimensions, and *unpatched* to reconstruct a $224 \times 224 \times 3$ image. A 3×3 convolution followed by sigmoid produces the probability heatmap. Key modifications from Mkrtchyan et al. (2025): (i) removal of cross-token linear layer, using only image-patch output tokens; (ii) addition of spatial positional encodings to radio tokens. See Appendix G for the unpatching operation details.

6.3 Training details

Models are trained on 2×A100 GPUs with DDP Li et al. (2020). Pretraining uses a warmup-stable-decay (WSD) schedule: LR ramps to 2×10^{-4} over 10% of epochs, holds constant, then decays to 0. Pretraining runs 30 epochs (B′) or 14 epochs (C). Fine-tuning on real data runs 100 epochs with linearly decaying LR. Batch size is 31; all models initialize from DINOv2-L/14 pretrained weights.

7 Experiments

7.1 Baselines

k-NN: We reimplement the weighted k-NN baseline from Kousias et al. (2024); Savelli et al. (2024), constructing RSSI fingerprints with missing values infilled at -25dB and distances normalized by non-zero BS count.

k-NN in a square: To fairly compare with DL models (which assume the UE lies within a 600m×600m crop), we restrict k-NN search to the same region, falling back to center prediction when no fingerprints exist.

MLP+UNet: A CNN-based baseline Khachatrian et al. (2025) adapted for RF fingerprinting inputs. See Appendix C for details.

7.2 PRETRAINING PROTOCOL

We pretrain MapRadioFormer+ on Dataset C and Dataset B′ separately, evaluate zero-shot on real data, then fine-tune on Dataset A. We also explore a three-stage pipeline: C→B′→A.

8 RESULTS

Table 1: Localization performance across different experiments.

Experiment	Model	# of UEs (training)	# of training samples	RMSE known streets	RMSE unknown streets
k-NN				35.0	2904.6
k-NN in the square				38.8	200.2
A	MLP+UNet	2,116	320,096	67.0	226.3
B′ zero-shot		53,055	189,293	240.4	249.8
C zero-shot		273,996	4,958,139	242.0	192.0
C → B′ zero-shot		–	–	235.1	209.4
B′ → A		–	–	71.8	190.5
C → A		–	–	67.5	211.3
C → B′ → A		–	–	73.3	213.5
A	MapRadioFormer+	2,116	320,096	38.5	322.7
B′ zero-shot		53,055	189,293	292.0	158.9
C zero-shot		273,996	4,958,139	286.0	184.0
C → B′ zero-shot		–	–	265.2	186.4
B′ → A		–	–	59.2	178.5
C → A		–	–	57.4	162.5
C → B′ → A		–	–	46.9	165.4

8.1 ROME DATASET RESULTS

Table 1 and Figure 4 summarize results. The k-NN baseline achieves 35m on known streets but catastrophically fails on unknown regions (2,905m). The constrained k-NN in a square yields 200m on unknown streets.

Real-only training. MapRadioFormer+ trained on Dataset A achieves 38.5m on known streets (comparable to k-NN) but 322.7m on unknown streets, indicating poor generalization from limited real data. MLP+UNet performs better on unknown streets (226.3m) but worse on known streets (67.0m).

Zero-shot synthetic. MapRadioFormer+ pretrained on B′ achieves 158.9m on unknown streets, outperforming the C-pretrained model (184.0m) by 25m. This gap—despite B′ having 5× fewer samples—demonstrates that calibration quality trumps data quantity for zero-shot transfer. When evaluated on synthetic test data, the C-pretrained model achieves near-perfect generalization (28.1m on known and 26.9m on unknown synthetic streets).

Fine-tuned models. After fine-tuning on real data, C→A achieves the best unknown-street performance at 162.5m—a 50% improvement over real-only training. The three-stage C→B′→A pipeline yields 165.4m, offering no additional benefit. MLP+UNet consistently underperforms MapRadioFormer+ after fine-tuning, confirming the transformer's superior generalization. Domain adversarial

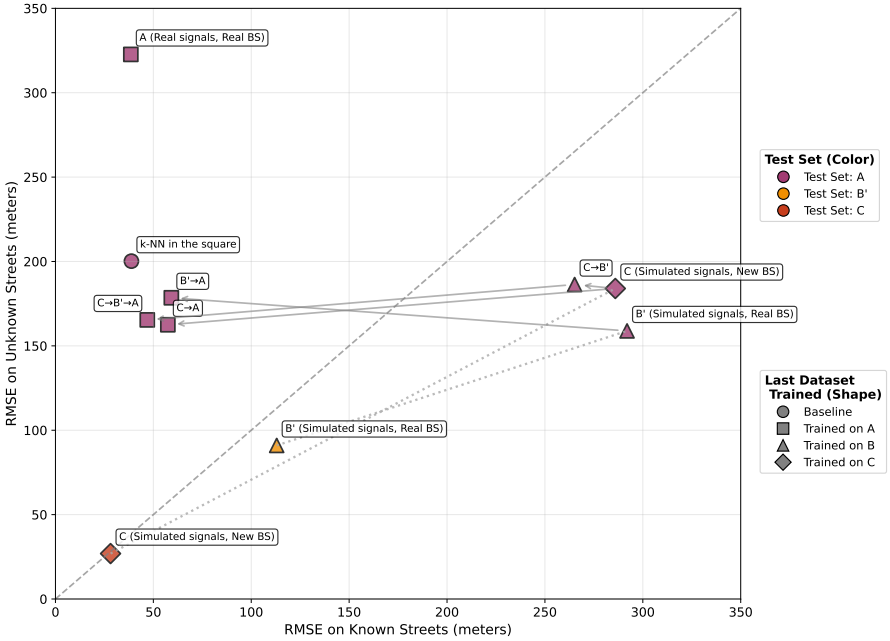


Figure 4: Localization errors of all our models on known (x-axis) and unknown (y-axis) streets.

training further reduces B′→A error to 162.7m, though benefits are inconsistent across settings (Appendix D).

8.2 Cross-city generalization (Oslo)

Table 2: Zero-shot localization performance of the MapRadioFormer+ model on the Oslo out-of-distribution dataset.

Experiment	# of UEs (training)	# of training samples	RMSE
A	2,116	320,096	157.2
B′	53,055	189,293	246.9
C	273,996	4,958,139	299.9
C → B′	–	–	248.1
B′ → A	–	–	149.6
C → A	–	–	132.2
C → B′ → A	–	–	132.6

Table 2 shows zero-shot performance on Oslo. Models trained on synthetic data alone transfer poorly (>240m). The Rome-only model achieves 157.2m, while C→A reduces this to 132.2m—a 16% improvement demonstrating that synthetic pretraining improves cross-city generalization. After fine-tuning on Oslo data, models achieve 61.5m on unknown streets (see Appendix B for the full Oslo fine-tuning matrix). Notably, all synthetically pre-trained models surpass the k-NN in the square baseline (182.4m) on unknown streets, while k-NN retains superiority on known streets (19.9m), confirming that DL models trade known-street precision for generalization. Interestingly,

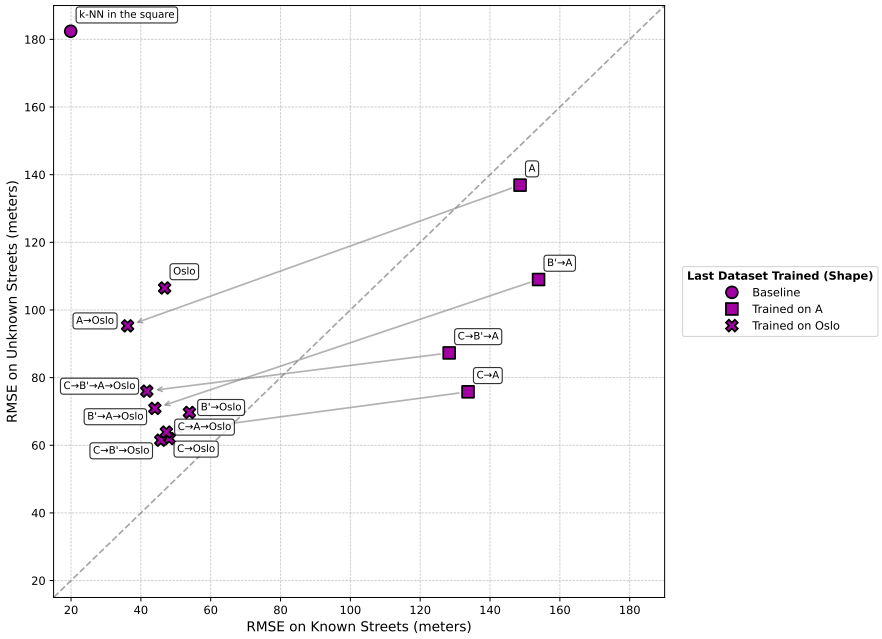


Figure 5: Localization errors of all our models on known (x-axis) and unknown (y-axis) streets on Oslo dataset.

this reverses the Rome finding: for cross-city transfer, synthetic data scale dominates over calibration quality—suggesting that breadth matters more than fidelity when the target domain diverges substantially from the calibration source.

9 Limitations and Future Work

Three main limitations remain. First, the largest contributor to the sim-to-real gap is the quality of 3D building models; simplified geometric approximations (e.g., prism-shaped buildings) cannot capture real-world propagation complexity. Second, MapRadioFormer+ requires a 600m$\times$600m crop centered near the UE, implicitly assuming coarse prior location knowledge. Variable-scale experiments (Appendix E) confirm that performance degrades with increasing map size on unknown streets, underscoring the sensitivity to the 600m crop assumption. future work should remove this assumption. Third, we lack a fine-tuning strategy that simultaneously closes the sim-to-real gap and preserves generalization learned during pretraining. Despite achieving 162m in Rome and 62m in Oslo, these errors remain high for practical deployment. Progress requires more realistic 3D environments, diverse real-world datasets, and greater willingness from network carriers to share BS parameters. Our fully open-source simulation pipeline and forthcoming public datasets aim to support continued research on these challenges.

Additional experimental details—including Oslo data splits, fine-tuning ablations, domain adaptation, variable-scale evaluation, computational cost analysis, and architecture specifics—are provided in the appendices.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Cursor (with Claude-4-Sonnet backend) and GPT5 in order to identify and fix inconsistent terms and definitions across sections, brainstorm on better formulations of the ideas in English, check the grammar, generate visualizations and design certain elements of the page layout. After using those tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

REFERENCES

Mohamed Akrout, Amal Feriani, Faouzi Bellili, Amine Mezghani, and Ekram Hossain. Domain generalization in machine learning models for wireless communications: Concepts, state-of-the-art, and open issues. *IEEE Communications Surveys & Tutorials*, 25(4):3014–3037, 2023. doi: 10.1109/COMST.2023.3326399.

Usman Ali, Giuseppe Caso, Luca De Nardis, Konstantinos Kousias, Mohammad Rajiullah, Ozgu Alay, Marco Neri, Anna Brunstrom, and Maria-Gabriella Di Benedetto. Large-scale dataset for the analysis of outdoor-to-indoor propagation for 5g mid-band operational networks. *Data*, 7(3): 34, 2022. ISSN 2306-5729. doi: 10.3390/data7030034. URL https://www.mdpi.com/2306-5729/7/3/34.

Ahmed Alkhateeb. Deepmimo: A generic deep learning dataset for millimeter wave and massive mimo applications, 2019. URL https://arxiv.org/abs/1902.06435.

Altair Engineering Inc. *WinProp Radio Propagation Software*, 2021. Available: https://web.altair.com/winprop-telecom.

Udita Bhattacherjee, Chethan Kumar Anjinappa, LoyCurtis Smith, Ender Ozturk, and Ismail Guvenc. Localization with deep neural networks using mmwave ray tracing simulations. In *2020 SoutheastCon*, pp. 1–8, 2020. doi: 10.1109/SoutheastCon44009.2020.9249699.

M. Majid Butt, Anil Rao, and Daejung Yoon. Rf fingerprinting and deep learning assisted ue positioning in 5g. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–7, 2020. doi: 10.1109/VTC2020-Spring48590.2020.9128640.

Bernardo Camajori Tedeschini, Mattia Brambilla, Lorenzo Italiano, Simone Reggiani, Davide Vaccarono, Marianna Alghisi, Lorenzo Benvenuto, Alessandro Goia, Eugenio Realini, Florin Grec, and Monica Nicoli. A feasibility study of 5G positioning with current cellular network deployment. *Scientific Reports*, 13(1):15281, September 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-42426-1. URL https://doi.org/10.1038/s41598-023-42426-1.

Hyun-Su Cha, Gilsoo Lee, Amitava Ghosh, Matthew Baker, Sean Kelley, and Juergen Hofmann. 5g nr positioning enhancements in 3gpp release-i8. *IEEE Communications Standards Magazine*, 9 (1):22–27, 2025. doi: 10.1109/MCOMSTD.0001.2400006.

Luca De Nardis, Giuseppe Caso, Ozgu Alay, Marco Neri, Anna Brunstrom, and Maria-Gabriella Di Benedetto. Positioning by multicell fingerprinting in urban nb-iot networks. *Sensors*, 23 (9), 2023. ISSN 1424-8220. doi: 10.3390/s23094266. URL https://www.mdpi.com/1424-8220/23/9/4266.

Marcelo N. de Sousa, Ricardo Sant'Ana, Rigel P. Fernandes, Julio Cesar Duarte, Jose A. Apolinario, and Reiner S. Thoma. Improving the performance of a radio-frequency localization system in adverse outdoor applications. *EURASIP Journal on Wireless Communications and Networking*, 2021(1):123, May 2021. ISSN 1687-1499. doi: 10.1186/s13638-021-02001-6. URL https://doi.org/10.1186/s13638-021-02001-6.

Antonio Del Corte-Valiente, José Manuel Gómez-Pulido, Oscar Gutiérrez-Blanco, and José Luis Castillo-Sequera. Localization approach based on ray-tracing simulations and fingerprinting techniques for indoor–outdoor scenarios. *Energies*, 12(15), 2019. ISSN 1996-1073. doi: 10.3390/en12152943. URL https://www.mdpi.com/1996-1073/12/15/2943.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks, 2016. URL https://arxiv.org/abs/1505.07818.

Ivan Gonzalez, Lorena Lozano, Santiago Cejudo, Francisco Saez de Adana, and Felipe Cetedra. New version of fasant code. In *2008 IEEE Antennas and Propagation Society International Symposium*, pp. 1–4, 2008. doi: 10.1109/APS.2008.4619474.

Jakob Hoydis, Sebastian Cammerer, Fayçal Ait Aoudia, Avinash Vem, Nikolaus Binder, Guillermo Marcus, and Alexander Keller. Sionna: An open-source library for next-generation physical layer research, 2023. URL https://arxiv.org/abs/2203.11854.

Y. Huangfu, J. Wang, S. Dai, R. Li, J. Wang, C. Huang, and Z. Zhang. WAIR-D: Wireless AI research dataset. 2022.

Hrant Khachatrian, Rafayel Mkrtchyan, and Theofanis P. Raptis. Outdoor environment reconstruction with deep learning on radio propagation paths. In *International Wireless Communications and Mobile Computing, IWCMC 2024, Ayia Napa, Cyprus, May 27-31, 2024*, pp. 1498–1503. IEEE, 2024. doi: 10.1109/IWCMC61514.2024.10592367.

Hrant Khachatrian, Rafayel Mkrtchyan, and Theofanis P. Raptis. Deep learning with synthetic data for wireless nlos positioning with a single base station. *Ad Hoc Networks*, 167:103696, 2025. ISSN 1570-8705. doi: https://doi.org/10.1016/j.adhoc.2024.103696. URL https://www.sciencedirect.com/science/article/pii/S157087052400307X.

Mike Koivisto, Mário Costa, Janis Werner, Kari Heiska, Jukka Talvitie, Kari Leppänen, Visa Koivunen, and Mikko Valkama. Joint device positioning and clock synchronization in 5g ultra-dense networks. *IEEE Transactions on Wireless Communications*, 16(5):2866–2881, 2017. doi: 10.1109/TWC.2017.2669963.

Konstantinos Kousias, Mohammad Rajiullah, Giuseppe Caso, Usman Ali, Ozgu Alay, Anna Brunstrom, Luca De Nardis, Marco Neri, and Maria-Gabriella Di Benedetto. A large-scale dataset of 4g, nb-iot, and 5g non-standalone network measurements. *IEEE Communications Magazine*, 62 (5):44–49, 2024. doi: 10.1109/MCOM.011.2200707.

Shen Li, Ge Zhao, Jian Yang, Jian Yu, Hao Chen, Yibo Li, Yang Yu, Haonan Shi, Yufei Li, Min Yang, et al. Pytorch distributed: Experiences on accelerating data parallel training. In *Proceedings of the VLDB Endowment*, volume 13, pp. 3005–3018, 2020.

Armen Manukyan, Hrant Khachatrian, Edvard Ghukasyan, and Theofanis P. Raptis. On the limitations of ray-tracing for learning-based rf tasks in urban environments. *arXiv preprint arXiv:2507.19653v1*, 2025a. URL https://arxiv.org/abs/2507.19653v1.

Armen Manukyan, Rafayel Mkrtchyan, Ararat Saribekyan, Theofanis P. Raptis, and Hrant Khachatrian. Bridging the sim-to-real gap in rf localization with large-scale synthetic pretraining. *Information Fusion*, 2025b. doi: 10.1016/j.inffus.2025.104104.

Víctor F. Miramá, Luis Enrique Díez, Alfonso Bahillo, and Víctor Quintero. A survey of machine learning in pedestrian localization systems: Applications, open issues and challenges. *IEEE Access*, 9:120138–120157, 2021. doi: 10.1109/ACCESS.2021.3108073.

Rafayel Mkrtchyan, Armen Manukyan, Hrant Khachatrian, and Theofanis P. Raptis. Fusion of pervasive rf data with spatial images via vision transformers for enhanced mapping in smart cities, 2025. URL https://arxiv.org/abs/2508.03736.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2024. URL https://arxiv.org/abs/2304.07193.

Remcom Inc. *Wireless InSite: 3D Wireless Prediction Software*, 2020. Available: https://www.remcom.com/wireless-insite-em-propagation-software.

Clement Ruah, Osvaldo Simeone, and Bashir M. Al-Hashimi. A bayesian framework for digital twin-based control, monitoring, and data collection in wireless systems. *IEEE Journal on Selected Areas in Communications*, 41(10):3146–3160, 2023. doi: 10.1109/JSAC.2023.3310093.

Marco Savelli, Luca De Nardis, Giuseppe Caso, Federico Ferretti, Anna Brunstrom, Ozgu Alay, Marco Neri, and Maria-Gabriella Di Benedetto. Range-free positioning in nb-iot networks by machine learning. In *2024 International Conference on Localization and GNSS (ICL-GNSS)*, pp. 1–7, 2024. doi: 10.1109/ICL-GNSS60721.2024.10578446.

Syed Shahid Shah, Chao Sun, Dongkai Yang, Muhammad Wisal, Yingzhe He, Bai Lu, and Ying Xu. Evaluation of 5g positioning based on uplink srs and downlink prs under los and nlos environments. *Applied Sciences*, 15(14), 2025. ISSN 2076-3417. doi: 10.3390/app15147909. URL https://www.mdpi.com/2076-3417/15/14/7909.

Eric Widdison and David G. Long. A review of linear multilateration techniques and applications. *IEEE Access*, 12:26251–26266, 2024. doi: 10.1109/ACCESS.2024.3361835.

Henk Wymeersch, Gonzalo Seco-Granados, Giuseppe Destino, Davide Dardari, and Fredrik Tufvesson. 5g mmwave positioning for vehicular networks. *Wireless Commun.*, 24(6):80–86, December 2017. ISSN 1536-1284. doi: 10.1109/MWC.2017.1600374. URL https://doi.org/10.1109/MWC.2017.1600374.

Liangbo Xie, Yukun Zhang, Changming Zhao, Chenlin Zhang, Mu Zhou, and Xiaolong Yang. Positioning under multipath environments in wireless network: Survey, design, and opportunities. *IEEE Network*, 38(6):476–484, 2024. doi: 10.1109/MNET.2024.3435087.

Abdulmalik Shehu Yaro, Filip Maly, and Pavel Prazak. A survey of the performance-limiting factors of a 2-dimensional rss fingerprinting-based indoor wireless localization system. *Sensors*, 23(5), 2023. ISSN 1424-8220. doi: 10.3390/s23052545. URL https://www.mdpi.com/1424-8220/23/5/2545.

Faheem Zafari, Athanasios Gkelias, and Kin K. Leung. A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, 21(3):2568–2599, 2019. doi: 10.1109/COMST.2019.2911558.

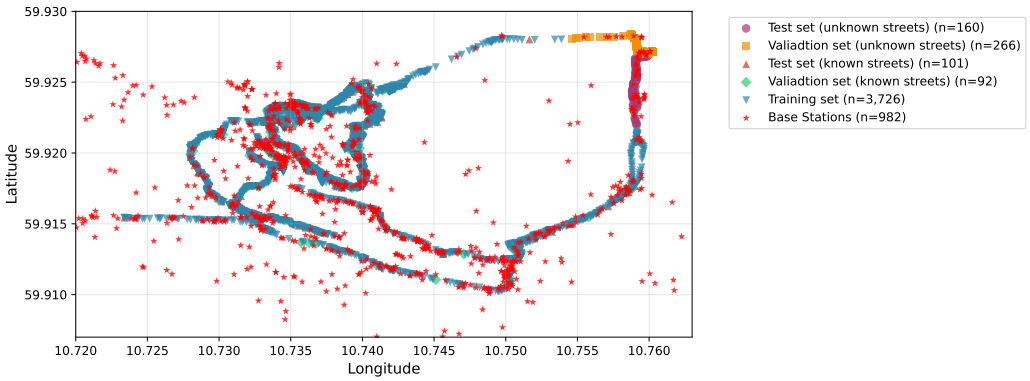


Figure 6: Oslo training, validation, and test split of UEs alongside BS locations (stars).

Table 3: RMSE comparison on known and unknown Oslo streets across pre-training and fine-tuning configurations.

(a) Unknown streets.

Synthetic pre-training	No real data	Rome	Rome→Oslo	Oslo
None		136.9	95.3	106.5
B'	168.3	109.0	70.9	69.7
C	184.7	75.8	63.9	62.0
$C \rightarrow B'$	176.3	87.3	76.0	61.5
k-NN in the square				182.4

(b) Known streets.

Synthetic pre-training	No real data	Rome	Rome→Oslo	Oslo
None		148.6	36.2	46.8
B'	272.9	153.9	44.0	53.9
C	293.2	133.7	47.3	48.1
$C \rightarrow B'$	267.6	128.3	41.7	45.7
k-NN in the square				19.9

A Oslo Dataset Details

The Oslo dataset Kousias et al. (2024) comprises 5,266 distinct UE locations and 982 BS locations, yielding 389,124 measured UE–BS pairs. Figure 6 shows the spatial distribution and data splits. The dataset serves two purposes: (i) zero-shot evaluation for cross-city generalization, and (ii) fine-tuning followed by evaluation on held-out Oslo streets.

B Oslo Fine-tuning Results

Table 3 presents the full Oslo fine-tuning matrix. Key findings: (i) synthetic pre-training consistently improves performance over real-only training; (ii) large-scale Dataset C dominates once real fine-tuning data is available; (iii) direct Oslo fine-tuning yields the best results (61.5m), though Rome→Oslo transfer remains beneficial without synthetic data (106.5m→95.3m). Figure 5 visualizes these results. All MapRadioFormer+ variants surpass the k-NN baseline on unknown streets except the model trained solely on Dataset C.

C MLP+UNet Baseline

The MLP+UNet baseline Khachatrian et al. (2025) combines a convolutional U-Net encoder–decoder with an MLP branch for RF features. The encoder reduces spatial resolution through convolutional blocks (64 to 1,024 channels), while the MLP processes RF inputs through three layers (256, 1,024, 256). At the bottleneck, RF embeddings are spatially expanded and concatenated with encoder features. The decoder reconstructs the heatmap via transposed convolutions with skip connections. Due to architectural constraints, BS count is limited to three. Full MLP+UNet results are included in Table 1 (main text).

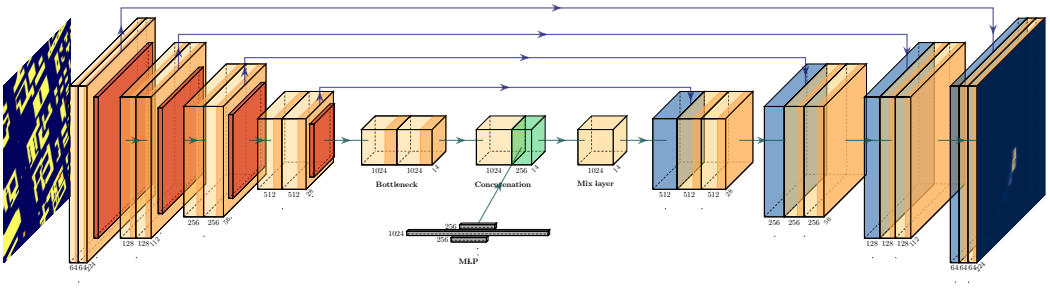


Figure 7: MLP+UNet architecture, adapted from Khachatrian et al. (2025).

D Domain Adaptation

We applied domain adversarial training Ganin et al. (2016) to learn domain-invariant representations. The feature extractor G_f (ViT encoder) feeds both a task predictor G_y (decoder) and a domain classifier G_d (MLP on CLS token), optimizing:

$$\min_{\theta_f,\theta_y} \max_{\theta_d} \mathcal{L}_y[G_y(G_f(x)), y] - \lambda \mathcal{L}_d[G_d(G_f(x)), d]$$

Table 4: Fine-tuning vs. domain adaptation (RMSE on unknown streets).

Source→Target	**Fine-tuning**	**Domain adapt.**
$B' \to A$	178.5	162.7
$B' \to$ Oslo	69.7	75.6
$A \to$ Oslo	95.3	101.1

Results are mixed (Table 4): domain adaptation improves B′→A (178.5→162.7m) but hurts Oslo transfers, suggesting adversarial alignment benefits sim-to-real but not cross-city scenarios with the current setup.

E Variable Scale Evaluation

After generating the variable-scale maps, the data were grouped into uniform bins of 200m width, ranging from 200m to 1,200m. For each bin, model performance was evaluated on the corresponding datapoints, and the RMSE values were averaged within the bin.Figure 8 summarizes the results. For the known streets, all MapRadioFormer+ variants underperform relative to the k-NN in the square baseline. Moreover, no consistent trend is observed with respect to the map scale, as performance remains relatively stable across different size intervals.

F Computational Cost Analysis

Figure 9 summarizes the cost–performance trade-off. The real-only baseline requires ∼15 GPU-hours for 322.7m RMSE. Calibrated B′ achieves 158.9m (zero-shot) at ∼40 GPU-hours (2.7× compute, 51% error reduction). Large-scale C→A reaches 162.5m at ∼88 GPU-hours (5.8× compute, 50% reduction). The three-stage pipeline adds cost (∼125 GPU-hours) without accuracy gains. The cost-performance analysis reveals two viable strategies: calibrated B′ (zero-shot) achieves 51% error reduction at 40 GPU-hours, while uncalibrated C→A achieves 50% reduction at 88 GPU-hours. Calibration provides better efficiency when area-specific tuning is feasible; uncalibrated pretraining offers flexibility when calibration data is unavailable. Both approaches outperform real-only training, demonstrating that synthetic data generates meaningful accuracy gains despite the computational overhead.

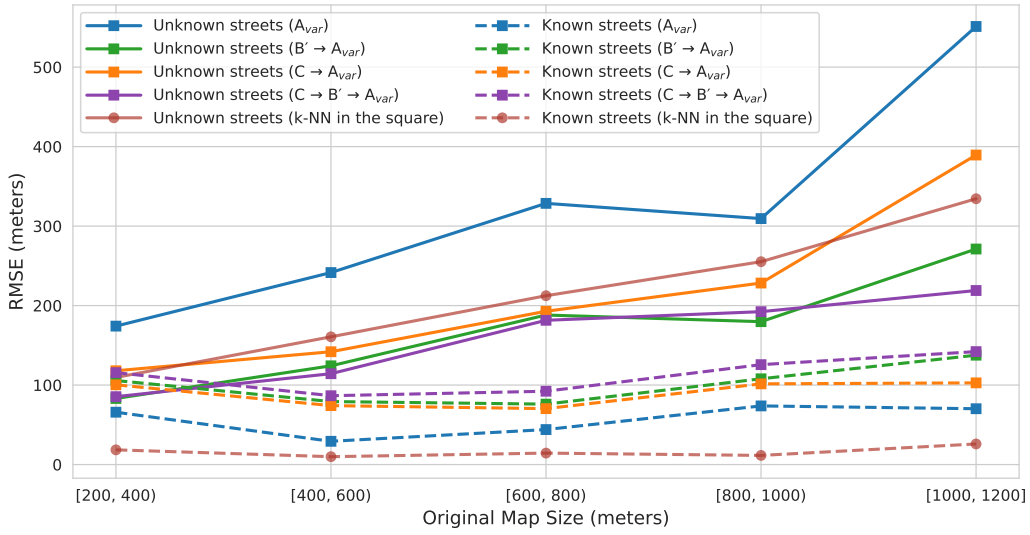


Figure 8: Performance vs. input map scale (200m–1,200m bins).

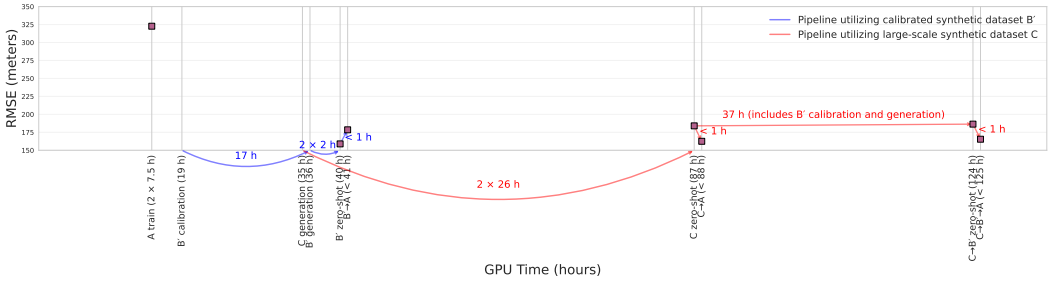


Figure 9: RMSE vs. A100 GPU-hours for different training strategies.

G Architecture Details: Unpatching Operation

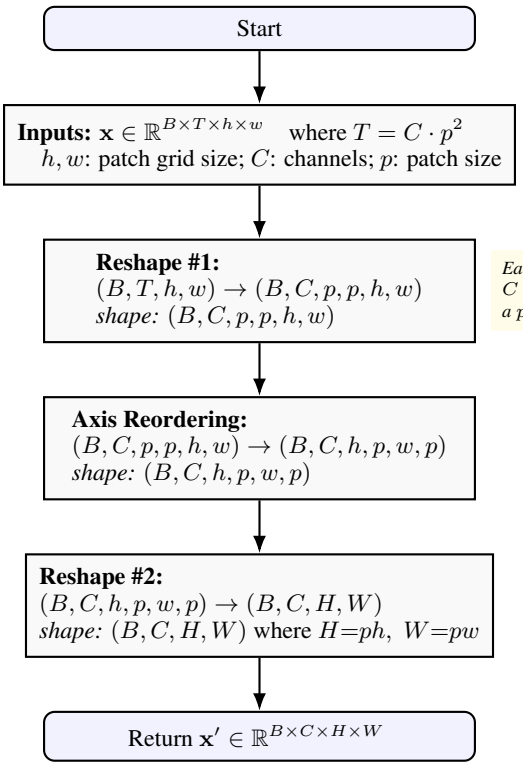
The unpatching operation (Figure 10) reconstructs the output image from transformer patch tokens. Parameters: $h = w = 16$, channels$= 3$, patch_size$= 14$, yielding a $224 \times 224 \times 3$ output.

H Dataset A_{VAR}: Variable Scale Details

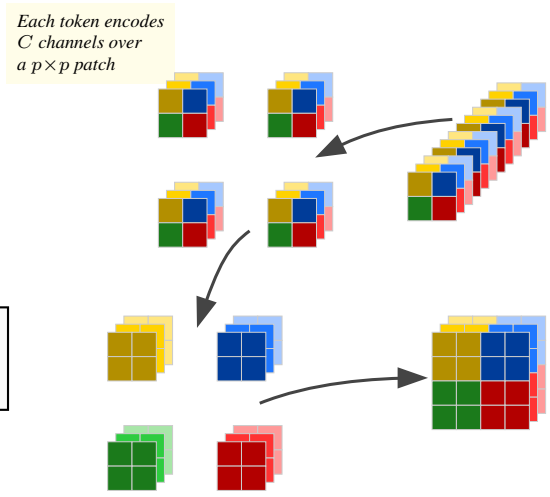
For variable-scale evaluation, we sample random crop half-side lengths from $\mathcal{N}(300, 100)$ constrained to $[100, 600]$m. Data are grouped into uniform 200m bins from 200m to 1,200m for evaluation.

I Simulation Compute Details

All ray-tracing was dispatched as independent single-GPU jobs on NVIDIA A100 (40GB) accelerators, each capped at $\sim$10,000 BS–UE links. For Datasets B/B′ the BS constellation was fixed with resampled UEs; for Dataset C both BS and UE positions were freshly drawn per shard. Up to six A100s processed distinct shards in parallel, scaling linearly with no inter-GPU communication.



(a) Unpatching operation flowchart: reconstructing an image from patch tokens via axis reordering and spatial unfolding.



(b) Visual explanation of the unpatching operation.

Figure 10: The unpatching operation. Input tensor has $T = C \cdot p^2$ tokens per spatial location (h, w), reshaped to output image of size $(H, W) = (ph, pw)$.