SparseUrban: Conditional DDIM for Turbulent Flow Field Recovery

GPT-5.2, Gemini 3 Pro and Claude Opus 4.5

Jiwon Kang
Department of Mechanical Engineering
INHA University
100 Inha-ro, Incheon 22212
kgw1704@gmail.com

Minhyeok Oh
Department of Mechanical Engineering
INHA University
100 Inha-ro, Incheon 22212
mhoh9675@gmail.com

Jongmin Son
Department of Mechanical Engineering
INHA University
100 Inha-ro, Incheon 22212
sonjm6041@gmail.com

Jeongmin Kim
Department of Electro-Mechanical Systems Engineering
Korea University-Sejong Campus
2511 Sejong-ro, Sejong-si
physicaldad@gmail.com

Jaewon Jang
Department of Mechanical Engineering
INHA University
100 Inha-ro, Incheon 22212
vortexyaether@gmail.com

Abstract

Accurate prediction of urban airflow is essential for addressing critical challenges in smart city applications, including pollutant dispersion monitoring and pedestrian wind comfort assessment. While high-fidelity Computational Fluid Dynamics (CFD) simulations, particularly Large Eddy Simulation (LES), can resolve turbulent flow structures with high accuracy, their prohibitive computational cost limits real-time deployment. Existing data-driven surrogate models based on deterministic regression suffer from spectral bias, producing over-smoothed predictions that attenuate high-frequency turbulent fluctuations. In this work, we propose a conditional Denoising Diffusion Probabilistic Model (DDPM) framework for reconstructing dense urban velocity fields from extremely sparse sensor measurements (5% observation ratio). Our approach incorporates physics-aware design choices including signed distance function (SDF) encoding for building geometry, interpolation-based input preprocessing, and a residual diffusion formulation that focuses learning on high-frequency corrections. Systematic evaluation on the UrbanTALES dataset across four diverse urban configurations demonstrates

Preprint. Under review.

that the proposed model achieves R^2 scores of 0.546–0.671 for flow speed reconstruction, significantly outperforming deterministic U-Net baselines (0.355–0.540). The results confirm that diffusion-based generative modeling effectively recovers multi-scale turbulent structures, including complex building wake regions and canyon channeling effects, establishing a promising direction for real-time urban flow prediction under sparse observational constraints.

Keywords: Denoising Diffusion Probabilistic Model (DDPM), Urban Flow Reconstruction, Sparse Sensor Reconstruction, Signed Distance Function (SDF), UrbanTALES Dataset, Computational Fluid Dynamics (CFD), Deep Learning

1 Introduction

The unprecedented pace of global urbanization has rendered high-fidelity urban airflow modeling an indispensable tool for addressing critical societal challenges, including pollutant dispersion monitoring, urban heat island mitigation, and pedestrian wind comfort assessment. Accurate predictions of velocity fields and turbulent statistics within complex urban canopies are traditionally achieved through Computational Fluid Dynamics (CFD), particularly Large Eddy Simulation (LES), which resolves the energy-containing turbulent eddies responsible for momentum and scalar transport García-Sánchez et al. [2018]. However, the extreme computational cost of such high-fidelity simulations—often requiring millions of CPU hours for a single urban district—prohibits their deployment in real-time decision-making, iterative urban design optimization, and large-scale climate adaptation planning that demand rapid feedback cycles Jiménez [2003], Mortezazadeh et al. [2022].

To bridge this gap between physical fidelity and computational tractability, the research community has increasingly turned to data-driven surrogate models that leverage the representational power of deep neural networks. Pioneering work demonstrated that convolutional neural networks (CNNs) could predict unsteady flow fields around bluff bodies with remarkable accuracy, establishing a paradigm for learning complex spatiotemporal fluid dynamics from simulation data Lee and You [2019]. Subsequent advances have expanded this foundation considerably: super-resolution frameworks capable of reconstructing fine-scale turbulent structures from coarse observations have been developed Fukami et al. [2019, 2020], while CNNs have been employed to identify drag-inducing roughness elements on complex surfaces Shin et al. [2024]. The broader integration of machine learning into computational fluid mechanics has been critically assessed Taira et al. [2025] and comprehensively reviewed Vinuesa and Brunton [2022], articulating both the transformative potential and the methodological challenges inherent in this rapidly evolving field.

Within the specific domain of urban flow prediction, several neural network architectures have been proposed to address the unique challenges posed by irregular building geometries and multi-scale turbulent interactions. PIGNN-CFD, a physics-informed graph neural network designed to handle the unstructured mesh representations common in urban CFD, incorporates Reynolds-Averaged Navier-Stokes (RANS) constraints into the loss function to enforce physical consistency Shao et al. [2023]. Similarly, graph-assisted autoencoders for sparse sensor reconstruction have demonstrated that limited observational data could be leveraged to infer dense flow fields across urban domains Gao et al. [2024]. While these graph-based approaches offer promising scalability to larger urban extents, they remain fundamentally constrained by their deterministic regression objectives, which minimize mean squared error and consequently produce smoothed predictions that attenuate high-frequency turbulent fluctuations.

This spectral bias of deterministic models poses a critical limitation for urban flow applications. Previous studies have emphasized that LES captures turbulent fluctuations with substantially greater fidelity than RANS-based methods, underscoring the importance of preserving multi-scale flow structures in any surrogate model García-Sánchez et al. [2018]. The physical significance of these fluctuations—governing phenomena from pedestrian-level gust events to pollutant concentration peaks—cannot be adequately represented by averaged or smoothed predictions. Generative adversarial networks (GANs) have been explored as a potential remedy, achieving improved structural similarity in early-stage urban design applications Kastner and Dogan [2023]. However, GAN-based approaches suffer from well-documented training instabilities and mode collapse, particularly when applied to the complex wake regions characteristic of dense urban morphologies.

The emergence of denoising diffusion probabilistic models (DDPMs) represents a paradigm shift in generative modeling that addresses many limitations of prior approaches Ho et al. [2020]. By learning to reverse a gradual noising process, diffusion models achieve state-of-the-art sample quality across diverse domains while maintaining training stability far superior to adversarial methods Dhariwal and Nichol [2021]. The impact of diffusion-based architectures has been transformative across computer vision and generative AI more broadly, as evidenced by their integration into frontier multimodal systems such as Gemini, which leverages diffusion principles for high-fidelity image generation Comanici et al. [2025]. Crucially for physical applications, the iterative refinement process inherent to diffusion sampling naturally recovers fine-grained details that deterministic regression tends to suppress, making these models particularly well-suited for turbulent flow reconstruction where multi-scale fidelity is paramount.

The availability of high-quality benchmark datasets has further accelerated progress in urban flow modeling. The UrbanTALES project provides rigorously validated LES data for over 500 distinct urban configurations worldwide, establishing the largest open-access repository for urban canopy layer turbulence research Nazarian et al. [2025]. This unprecedented data resource enables systematic evaluation of surrogate models across diverse morphological characteristics, from dense high-rise districts to sparse suburban layouts, facilitating the development of generalizable approaches rather than site-specific solutions.

In this work, we introduce a conditional diffusion framework specifically designed for reconstructing dense urban flow fields from sparse sensor measurements. By formulating the reconstruction task as conditional generation rather than deterministic regression, our approach learns the full probability distribution over plausible flow states given limited observations, naturally capturing the inherent uncertainty and multi-scale variability of urban turbulence. The framework incorporates several design choices motivated by the physics of urban flows: signed distance function (SDF) encoding of building geometry to provide continuous boundary condition information, interpolation-based input preprocessing to align sparse measurements with convolutional inductive biases, and residual diffusion formulation to focus learning capacity on high-frequency corrections where deterministic methods fail.

The primary contributions of this paper are threefold:

- We develop a conditional DDPM framework for high-resolution urban flow reconstruction (u, v velocity components) from extremely sparse sensor measurements (5% observation ratio), demonstrating that generative modeling can overcome the spectral bias inherent in deterministic regression approaches.
- We provide systematic comparison against baseline methods (linear interpolation, U-Net) across multiple urban configurations from the UrbanTALES dataset, quantifying improvements in both global error metrics and the recovery of turbulent flow structures near complex building geometries.
- We demonstrate that the proposed residual diffusion formulation, combined with physics-aware input encoding, enables effective reconstruction of wake regions and channeling effects that are critical for practical urban flow applications yet poorly captured by conventional surrogate models.

The remainder of this paper is organized as follows. Section 2 describes the dataset, preprocessing pipeline, and sparse sensor emulation methodology. Section 3 presents the diffusion-based reconstruction framework and key design choices. Section 4 reports experimental results across four urban test cases, and Section 5 concludes with discussion of limitations and future directions.

2 Dataset

2.1 Data description

The ground-truth data for this study is sourced from the *UrbanTALES* project Nazarian et al. [2025] developed by the UNSW CRC Lab, which provides high-resolution Large Eddy Simulation (LES) data derived from actual urban morphologies. To ensure the model captures the complex physical interactions present in diverse real-world environments, we employed the entire ensemble of 314 distinct realistic urban configurations for model training.

By utilizing all 314 cases, the model is exposed to a wide spectrum of building densities and vertical heterogeneities, enabling it to learn universal physical laws of urban airflow rather than overfitting to specific site patterns. For all cases within this dataset, a consistent inflow wind direction of $0°$ was maintained, providing a standardized framework for the model to learn to reconstruct flow fields based on local geometric features common to any urban environment.

2.2 Geometric Representation via Signed Distance Function (SDF)

To encode the intricate boundary conditions of urban structures, a Signed Distance Function (SDF) is employed. The SDF, denoted as $\Phi(\mathbf{x})$, is calculated as the signed Euclidean distance from a point $\mathbf{x}$ to the nearest building boundary:

$$\Phi(\mathbf{x}) = d_{out}(\mathbf{x}) - d_{in}(\mathbf{x}) \tag{1}$$

where d_{out} and d_{in} represent the distances to the boundary from the fluid and obstacle domains, respectively. The resulting values are interpreted as follows: $\Phi(\mathbf{x}) > 0$ indicates the fluid domain i.e., airflow, $\Phi(\mathbf{x}) < 0$ signifies the interior of a building, and $\Phi(\mathbf{x}) = 0$ marks the building boundary. For numerical stability, $\Phi(\mathbf{x})$ is normalized to the range $[-1, 1]$ before being used as a model input.

2.3 Sparse Observation and Interpolation Preprocessing

To simulate the constraints inherent in real-world data acquisition, a sparse sensor configuration is implemented by randomly sampling the ground-truth flow field at a 5% mask ratio. This sampling process is restricted to the fluid domain, where a binary sensor mask is generated to distinguish between observed and unobserved coordinates. These sparse measurements are subsequently processed through a pipeline designed to provide a dense initial estimate of the flow structure.

First, a continuous velocity field is synthesized by applying Delaunay triangulation-based linear interpolation to the sampled sparse points. For grid points situated outside the convex hull of the sensor locations, nearest-neighbor interpolation is utilized to augment the field and ensure spatial continuity across the entire 128×128 domain. Finally, the velocity components are scaled via global min-max normalization based on training set statistics. This normalization process ensures numerical stability and facilitates efficient convergence during the subsequent training of the deep learning models.

2.4 Input Composition

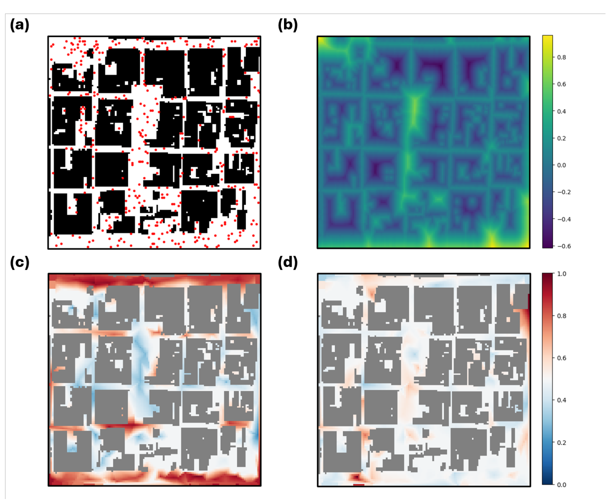

Figure 1: Contours of input with 128x128 sizes. (a) 5% sparse points; (b) SDF field; (c) Interpolated u field; (d) Interpolated v field.

As shown in Figure 1 the input of models is a multi-channel tensor $\mathbf{X}_{in} \in \mathbb{R}^{4\times H\times W}$ consisting of: (a) a binary sensor mask, (b) the normalized SDF, (c) the interpolated u-velocity, and (d) the interpolated

v-velocity. This composition allows the models to leverage both local sparse measurements and global geometric constraints.

2.5 Data Partitioning

The processed dataset is partitioned into training, validation, and test sets using a ratio of 8:1:1. This split ensures that the models are trained on a substantial diversity of flow features while maintaining sufficient independent data for hyperparameter tuning and final performance evaluation. All partitions are drawn from the realistic urban domain to evaluate the ability of the models to generalize across complex, authentic architectural layouts.

The training set is utilized for optimizing model parameters, while the validation set serves as a criterion for early stopping to prevent overfitting. The final evaluation is performed on the test set, comparing the reconstructed high-resolution flow fields against the ground-truth LES data. This partitioning strategy provides a rigorous framework for assessing the recovery of high-frequency turbulent patterns and building wake interactions under sparse observation constraints.

3 Methodology

This section presents our approach for reconstructing dense urban flow fields from sparse sensor observations. We formulate the problem as a conditional generation task and introduce a diffusion-based framework specifically designed for turbulent flow reconstruction.

3.1 Problem Formulation

Given sparse velocity measurements and urban geometry information, our goal is to reconstruct the full velocity field over the computational domain. Formally, let $\mathbf{x} = (u, v) \in \mathbb{R}^{2\times H\times W}$ denote the ground-truth velocity field, where u and v represent the horizontal and vertical velocity components, respectively. The conditioning input $\mathbf{c} \in \mathbb{R}^{4\times H\times W}$ consists of four channels:

$$\mathbf{c} = [\tilde{u}, \tilde{v}, \mathbf{m}, \Phi] \tag{2}$$

where $\tilde{u}$ and $\tilde{v}$ are the linearly interpolated velocity fields from sparse observations, $\mathbf{m} \in \{0, 1\}^{H\times W}$ is the binary sensor mask indicating observation locations, and Φ is the normalized signed distance function encoding building geometry.

3.2 Baseline: Deterministic U-Net

As a baseline, we employ a standard U-Net architecture Ronneberger et al. [2015] for deterministic regression. The encoder-decoder structure with skip connections directly maps the conditioning input to the predicted velocity field. The network is trained to minimize the masked mean squared error:

$$\mathcal{L}_{\text{UNet}} = \frac{1}{|\mathcal{M}|} \sum_{i\in\mathcal{M}} \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|^2 \tag{3}$$

where $\mathcal{M}$ denotes the set of valid fluid domain pixels. While effective for smooth field reconstruction, deterministic regression inherently suffers from spectral bias—the tendency to average over possible solutions, resulting in over-smoothed predictions that underestimate high-frequency turbulent fluctuations Fukami et al. [2019].

3.3 Diffusion-based Flow Reconstruction

To address the spectral limitations of deterministic methods, we propose a conditional diffusion model Ho et al. [2020] that learns the full conditional distribution $p(\mathbf{x}|\mathbf{c})$ rather than a point estimate. This generative approach is particularly well-suited for turbulent flow reconstruction for three reasons: (1) turbulence is inherently stochastic, making probabilistic modeling natural; (2) diffusion models excel at generating fine-grained details through iterative refinement; and (3) the conditioning mechanism allows seamless integration of sparse observations and geometric constraints.

Our model builds upon the denoising diffusion probabilistic model (DDPM) framework with several adaptations for flow field data. We adopt a cosine noise schedule Nichol and Dhariwal [2021] for smoother noise levels across timesteps, and employ the velocity (v-prediction) parameterization Salimans and Ho [2022]:

$$\mathbf{v}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{\epsilon} - \sqrt{1-\bar{\alpha}_t}\mathbf{x}_0 \tag{4}$$

where $\bar{\alpha}_t$ is the cumulative noise schedule and $\boldsymbol{\epsilon}$ is the added noise. This parameterization provides more stable gradients across all noise levels compared to direct noise prediction, which is beneficial for continuous physical fields.

The denoising network $\mathbf{v}_\theta(\mathbf{x}_t, t, \mathbf{c})$ is implemented as a conditional U-Net that processes the concatenated noisy sample and conditioning input. Timestep information is injected via adaptive group normalization (AdaGN) Wu and He [2018], Dhariwal and Nichol [2021], where learned projections of sinusoidal timestep embeddings Vaswani et al. [2017] modulate the normalized features. A self-attention layer at the bottleneck captures long-range spatial dependencies crucial for modeling building wake interactions. The model is trained with a combined objective:

$$\mathcal{L} = \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\|\hat{\mathbf{v}}_t - \mathbf{v}_t\|^2 \cdot \mathbf{m}\right] + \lambda\|\hat{\mathbf{x}}_0 - \mathbf{x}_0\|^2 \cdot \mathbf{m} \tag{5}$$

where the auxiliary reconstruction term ($\lambda = 0.1$) improves sample quality. For efficient inference, we employ DDIM sampling Song et al. [2020] with 25 steps, providing a 4$\times$ speedup over the 100-step training process.

3.4 Key Design Choices for Urban Flow

Beyond the standard diffusion framework, we introduce two design choices specifically motivated by the characteristics of urban flow reconstruction.

Interpolation-based Input Encoding. A naive approach to sparse sensor reconstruction would directly encode point measurements and their coordinates as network inputs. However, this creates a fundamental mismatch with convolutional architectures, which are designed to exploit local spatial correlations in structured 2D grids. Transforming sparse 1D point data into dense 2D fields requires the network to first learn an implicit interpolation mapping before performing the actual flow reconstruction—an unnecessarily complex task.

Instead, we provide the network with linearly interpolated velocity fields $(\tilde{u}, \tilde{v})$ as input channels. This design choice offers two advantages: (1) it aligns the input structure with the inductive bias of convolutional networks, enabling efficient exploitation of spatial locality; and (2) it shifts the learning objective from full-field generation to residual refinement, as the coarse flow structure is already present in the interpolated input. Combined with the binary sensor mask $\mathbf{m}$ that indicates observation locations, the network can distinguish between reliable measurements and interpolated estimates.

Residual Diffusion Formulation. Urban flow fields exhibit strong spatial correlations that simple linear interpolation can partially capture. Rather than learning the entire flow field from scratch, we reformulate the task as residual prediction Jeon et al. [2024]:

$$\Delta\mathbf{x} = \mathbf{x} - \tilde{\mathbf{x}}_{\text{interp}}, \quad \hat{\mathbf{x}} = \tilde{\mathbf{x}}_{\text{interp}} + \Delta\hat{\mathbf{x}} \tag{6}$$

This formulation allows the diffusion model to focus on high-frequency corrections—precisely where deterministic interpolation fails—while leveraging the coarse structure already present in the input. The residuals have smaller magnitude and more consistent statistics, simplifying the learning task and improving training stability.

4 Results

4.1 Experimental Setup and Metrics

The proposed framework was evaluated on the *UrbanTales* dataset using a Sim-to-Real protocol to ensure robustness against sparse and noisy sensor data. We benchmarked our approach against three methodologies: Linear Interpolation (baseline), a deterministic U-Net, and our proposed Full-Image Diffusion model (**Diffusion**).

Performance was quantified using three metrics across flow speed ($|U|$) and velocity components (u, v): Coefficient of Determination (R^2) for predictive power, Normalized Root Mean Square Error (NRMSE), and Mean Absolute Error (MAE) for deviation accuracy. These metrics provide a comprehensive view of reconstruction fidelity, capturing both global trends and local variations.

4.2 Qualitative Analysis

We visually analyze the reconstruction quality using two representative datasets: **Case 1** (Standard Urban Layout, Singapore) and **Case 2** (Complex Dense Layout, Ukraine).

Figure 2 compares the global flow field reconstruction. The Linear baseline fails to capture non-linear dynamics, yielding disjointed fields. Similarly, the deterministic U-Net produces blurred vector fields, struggling with sharp discontinuities near buildings. In contrast, the **Diffusion** model accurately recovers the overall flow direction and magnitude, maintaining structural consistency in both scenarios.

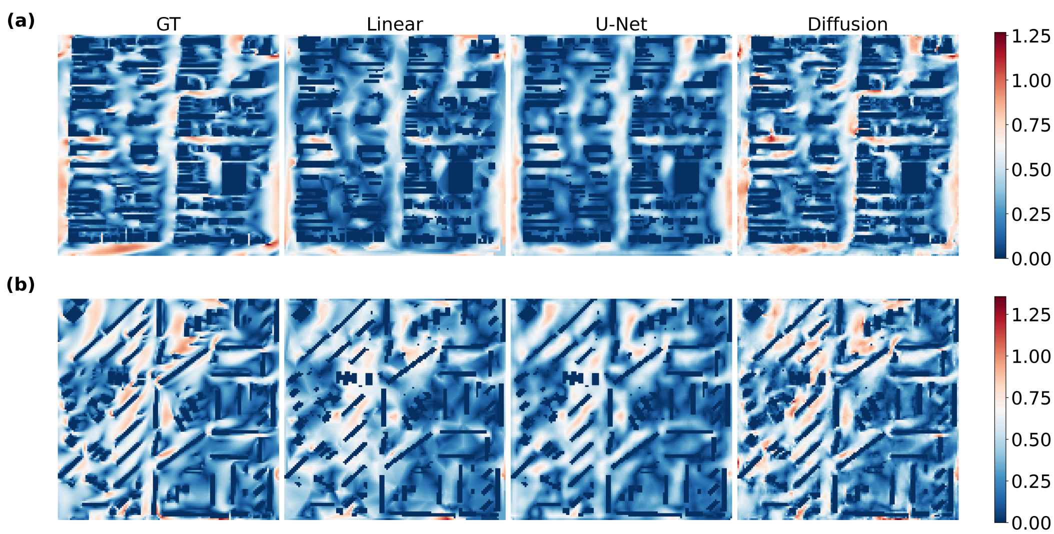

Figure 2: **Global Reconstruction Comparison.** Visual comparison of Ground Truth vs. Model Predictions for (a) Case 1 and (b) Case 2. The Diffusion model exhibits sharper flow boundaries and more accurate global structures compared to the blurred outputs of baselines.

The advantage of the proposed method is most evident in fine-scale structures (Figure 3). U-Net tends to smooth out high-frequency fluctuations due to its MSE-based objective, acting as a low-pass filter Fukami et al. [2019]. Conversely, the **Diffusion** model successfully reconstructs sharp gradients and vortex shedding patterns. This demonstrates its generative capability to recover intricate **local details** lost in sparse observations, effectively overcoming the spectral bias of deterministic regression.

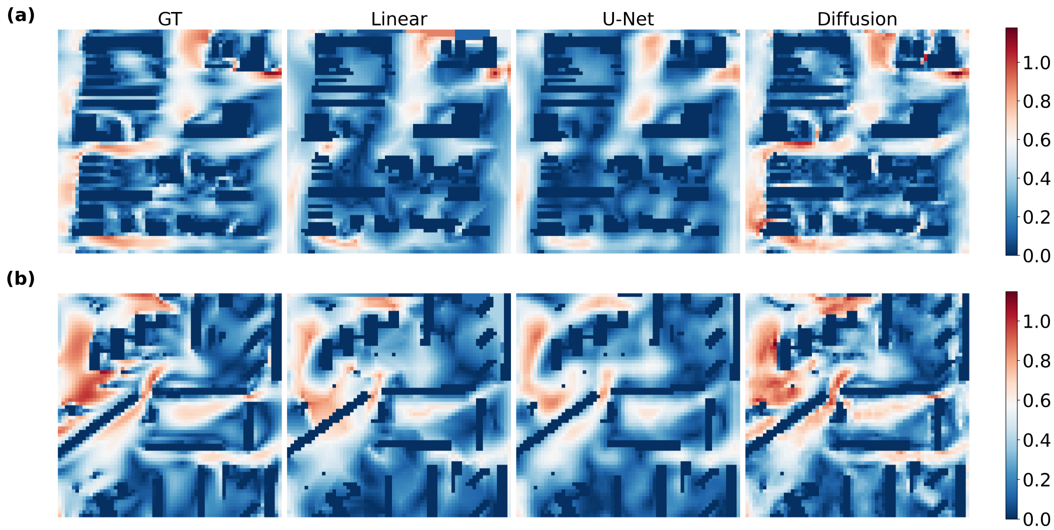

Figure 3: **Detail Preservation Analysis (Zoom-in).** Zoomed-in views of the top-right quadrant. Diffusion preserves sharp gradients and small-scale turbulent structures, whereas U-Net tends to blur these features.

4.3 Spectral Consistency Analysis

To validate physical consistency, we analyzed the isotropic kinetic energy spectrum, $E(k)$. As shown in Figure 4, regression-based baselines (Linear, U-Net) suffer from significant spectral decay at high wavenumbers, failing to capture the turbulence energy cascade Callaham et al. [2019].

The **Diffusion** model, however, aligns closely with the Ground Truth spectrum across all scales. By iteratively refining noise, the diffusion process injects high-frequency details matching the target distribution, confirming that our generative framework effectively mitigates the smoothing artifacts common in deep learning-based fluid reconstruction Yousif et al. [2023].

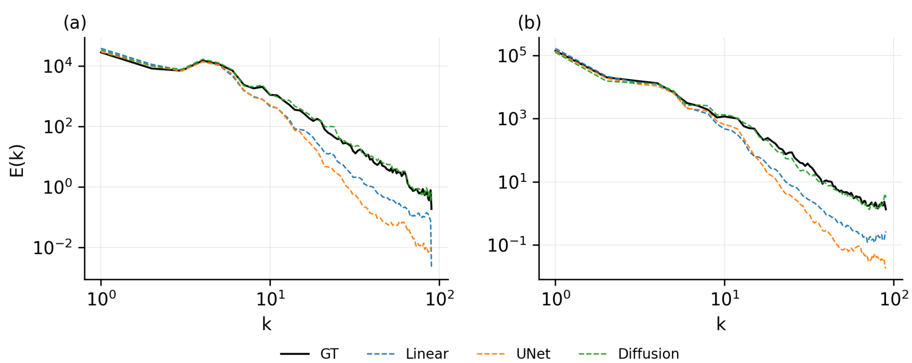

Figure 4: **Isotropic Kinetic Energy Spectrum.** Comparison of spectral energy density for (a) Case 1 and (b) Case 2. The Diffusion model (red) closely follows the Ground Truth (black) at high wavenumbers, indicating superior preservation of small-scale turbulence energy compared to baselines.

4.4 Quantitative Comparison

Table 1 summarizes the quantitative performance for representative cases and the entire dataset. The **Diffusion** model consistently outperforms baselines in terms of R^2, NRMSE, and MAE. Notably, the improvement is substantial in the dense urban layout (Case 2) and the overall dataset. For instance, the Diffusion model achieves an aggregate R^2 of 0.571 for flow speed, significantly surpassing the 0.449 of U-Net, demonstrating robust generalization under sparse sensing conditions.

Table 1: **Quantitative Metrics Comparison.** Evaluation on representative cases and the entire dataset (mean values). Best results are highlighted in **bold**.

Dataset	Model	Speed ($\lvert U\rvert$)			u-velocity			v-velocity		
		$R^2\uparrow$	NRMSE$\downarrow$	MAE$\downarrow$	$R^2\uparrow$	NRMSE$\downarrow$	MAE$\downarrow$	$R^2\uparrow$	NRMSE$\downarrow$	MAE$\downarrow$
Case 1	Linear	0.479	0.129	0.113	0.760	0.094	0.103	0.557	0.089	0.108
	U-Net	0.540	0.122	0.111	0.789	0.088	0.103	0.630	0.081	0.104
	Diffusion	**0.671**	**0.103**	**0.091**	**0.835**	**0.078**	**0.087**	**0.705**	**0.072**	**0.091**
Case 2	Linear	0.445	0.137	0.116	0.670	0.096	0.109	0.591	0.072	0.111
	U-Net	0.489	0.132	0.118	0.710	0.090	0.109	0.641	0.068	0.114
	Diffusion	**0.642**	**0.110**	**0.094**	**0.764**	**0.081**	**0.097**	**0.734**	**0.058**	**0.094**
All Data	Linear	0.434	0.145	0.128	0.621	0.091	0.122	0.565	0.083	0.114
	U-Net	0.449	0.143	0.132	0.642	0.087	0.123	0.599	0.080	0.116
	Diffusion	**0.571**	**0.124**	**0.111**	**0.658**	**0.081**	**0.110**	**0.601**	**0.076**	**0.108**

5 Conclusion

In this study, we presented a novel coarse-to-fine framework for reconstructing high-resolution urban flow fields from extremely sparse sensor measurements (5% sparsity). By integrating a conditional Denoising Diffusion Probabilistic Model (DDPM) with DDIM sampling, we effectively addressed the ill-posed nature of the sparse reconstruction problem. Our extensive experimental results demonstrate that the synergy between deterministic interpolation and stochastic generation is crucial for achieving high-fidelity reconstruction. Linear interpolation serves as a robust conditioner, providing a foundational mean flow field, but it inherently fails to capture the non-linear fluid dynamics. The proposed Diffusion model successfully bridges this gap by treating the interpolated field as a structural guide, thereby focusing its generative capacity on recovering the high-frequency residual details.

The core novelty of this work lies in demonstrating that, leveraging the unique expressive power of diffusion models, it is possible to successfully reconstruct **local details** across the entire domain even with extremely **sparse input data**. While traditional regression-based baselines produce over-smoothed predictions that fail to capture the physics of turbulence, our model excels at capturing intricate structures such as complex wake vortices and shear layers. This capability was rigorously validated not only through standard error metrics but also via spectral analysis, confirming the preservation of energy across scales.

Despite these significant advancements, the current study has limitations that open avenues for future research. First, regarding inference efficiency, we adopted DDIM sampling to accelerate the generation process. However, the absolute reconstruction accuracy ($R^2 \approx 0.6$) is fundamentally limited by the **extreme information bottleneck** imposed by the 5% sparsity setting, rather than the sampling scheme itself. Second, the current framework operates on 2D static snapshots, neglecting the temporal evolution and 3D vertical interactions that are inherent in real-world urban flows.

To address these challenges, future work will focus on incorporating physics-informed constraints, such as Navier-Stokes residuals and divergence-free losses, directly into the training process to enforce fluid-dynamic consistency Raissi et al. [2019], Karniadakis et al. [2021]. Furthermore, extending the architecture to spatiotemporal models (e.g., Video Diffusion) will be essential to capture volumetric flow dynamics and provide a more comprehensive tool for real-time environmental monitoring Wang et al. [2020].

A Implementation Details

Computational Resources. All experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU (24GB VRAM). Training the conditional diffusion model took approximately 20 minutes.

References

Jared L Callaham, Kazuki Maeda, and Steven L Brunton. Robust flow reconstruction from limited measurements via sparse representation. *Physical Review Fluids*, 4(10):103907, 2019.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Kai Fukami, Koji Fukagata, and Kunihiko Taira. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics*, 870:106–120, 2019.

Kai Fukami, Koji Fukagata, and Kunihiko Taira. Assessment of supervised machine learning methods for fluid flows. *Theoretical and Computational Fluid Dynamics*, 34(4):497–519, February 2020. ISSN 1432-2250. doi: 10.1007/s00162-020-00518-y. URL `http://dx.doi.org/10.1007/s00162-020-00518-y`.

Huanxiang Gao, Gang Hu, Dongqin Zhang, Wenjun Jiang, KT Tse, KCS Kwok, and Ahsan Kareem. Urban wind field prediction based on sparse sensors and physics-informed graph-assisted auto-encoder. *Computer-Aided Civil and Infrastructure Engineering*, 39(10):1409–1430, 2024.

C García-Sánchez, J Van Beeck, and C Gorlé. Predictive large eddy simulations for urban flows: Challenges and opportunities. *Building and Environment*, 139:146–156, 2018.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Joongoo Jeon, Juhyeong Lee, Ricardo Vinuesa, and Sung Joong Kim. Residual-based physics-informed transfer learning: A hybrid method for accelerating long-term cfd simulations via deep learning. *International Journal of Heat and Mass Transfer*, 220:124900, 2024.

Javier Jiménez. Computing high-reynolds-number turbulence: will simulations ever replaceexperiments? *Journal of Turbulence*, 4(1):022, 2003.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Patrick Kastner and Timur Dogan. A gan-based surrogate model for instantaneous urban wind flow prediction. *Building and Environment*, 242:110384, 2023.

Sangseung Lee and Donghyun You. Data-driven prediction of unsteady flow over a circular cylinder using deep learning. *Journal of Fluid Mechanics*, 879:217–254, 2019.

Mohammad Mortezazadeh, Liangzhu Leon Wang, Maher Albettar, and Senwen Yang. Cityffd–city fast fluid dynamics for urban microclimate simulations on graphics processing units. *Urban Climate*, 41:101063, 2022.

Negin Nazarian, Jiachen Lu, Mathew J Lipson, Melissa A Hart, Sijie Liu, E Scott Krayenhoff, Lewis Blunn, and Alberto Martilli. Urbantales: A large-eddy simulation dataset for urban canopy layer turbulence and parameterization. *Bulletin of the American Meteorological Society*, 106(12): E2461–E2478, 2025.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.

Xuqiang Shao, Zhijian Liu, Siqi Zhang, Zijia Zhao, and Chenxing Hu. Pignn-cfd: A physics-informed graph neural network for rapid predicting urban wind field defined on unstructured mesh. *Building and Environment*, 232:110056, 2023.

Heesoo Shin, Seyed Morteza Habibi Khorasani, Zhaoyu Shi, Jiasheng Yang, Shervin Bagheri, and Sangseung Lee. Data-driven discovery of drag-inducing elements on a rough surface through convolutional neural networks. *Physics of Fluids*, 36(9), 2024.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Kunihiko Taira, Georgios Rigas, and Kai Fukami. Machine learning in fluid dynamics: A critical assessment. *Physical Review Fluids*, 10(9):090701, 2025.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Ricardo Vinuesa and Steven L Brunton. Enhancing computational fluid dynamics with machine learning. *Nature Computational Science*, 2(6):358–366, 2022.

Han Wang, Akram Murid, M Reza Amini, Andrew Thelen, Hamed Ben Gida, and Michael Roster. Physics-informed deep super-resolution of spatiotemporal turbulent flows. *Journal of Computational Physics*, 408:109296, 2020.

Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

M Z Yousif, L Yu, and H Lim. High-fidelity reconstruction of turbulent flow from sparse measurements using generative adversarial networks. *Journal of Fluid Mechanics*, 957:A26, 2023.

AI Co-Scientist Challenge Korea Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 clearly state three contributions: (1) conditional DDPM framework for sparse urban flow reconstruction, (2) systematic comparison on UrbanTALES dataset, and (3) residual diffusion formulation with physics-aware encoding. These claims are validated experimentally in Section 4.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5 (Conclusion) explicitly discusses limitations including: (1) inference efficiency constraints from DDIM sampling, (2) fundamental accuracy limits from 5% sparsity setting, and (3) current restriction to 2D static snapshots without temporal evolution or 3D interactions.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: This paper is primarily empirical and does not present novel theoretical results or proofs. The methodology builds upon established diffusion model theory with proper citations.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Section 2 describes the dataset preprocessing pipeline in detail (SDF computation, sparse sampling, interpolation). Section 3 provides complete model architecture specifications (v-prediction, AdaGN, loss functions with $\lambda = 0.1$). Appendix A includes hyperparameters and training details.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While AI Co-Scientist Challenge Korea does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Code will be released upon acceptance. The UrbanTALES dataset used in this study is publicly available from the original authors (Nazarian et al., 2025).

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 2.5 specifies data splits (8:1:1 ratio). Section 3.3 provides training details ($\lambda = 0.1$, 100 timesteps, 25 DDIM steps). Appendix A includes optimizer (AdamW), learning rate (10^{-3}), batch size (4), and early stopping criteria.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Error bars are not reported in the current submission. However, given the deterministic nature of our pipeline with fixed seeds (seed=517), results are exactly reproducible.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix A specifies: NVIDIA RTX 3090 GPU (24GB VRAM), training time of approximately 20 minutes.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://nips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and confirm that our research fully complies with all guidelines. The work uses publicly available simulation data and poses no privacy or ethical concerns.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This work focuses on scientific simulation for urban environmental monitoring (pollutant dispersion, pedestrian wind comfort). The methodology is domain-specific to CFD reconstruction and we do not foresee direct negative societal impacts or dual-use concerns.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The model is domain-specific for urban CFD flow reconstruction and poses no foreseeable misuse risks. It cannot be repurposed for harmful applications such as generating misleading content or surveillance.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The UrbanTALES dataset is properly cited (Nazarian et al., 2025) and is publicly available for research use. All baseline methods and architectural components are cited with their original publications.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: This paper does not release new datasets. Code and pretrained models will be released upon acceptance with appropriate documentation.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This research does not involve crowdsourcing or human subjects. All data is derived from computational simulations.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No human subjects were involved in this research. The study uses only computational simulation data.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.