PMA-Diffusion: A Physics-guided Mask Aware Diffusion Framework for Traffic State Estimation from Sparse Observations

Anonymous Author(s)
Affiliation
Address
email

Abstract

High-resolution Traffic State Estimation (TSE) is a foundational key research topic for building efficient, safe, reliable, and resilient transportation and mobility systems in smart cities. It can be used for various Intelligent Transportation Systems (ITS) applications such as Advanced Transportation Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS). Yet, in practice, urban sensing infrastructures for the transportation and mobility systems, such as loop detectors and probe vehicles, provide data that is sparse, noisy, and unevenly distributed across city networks, limiting their utility for real-time decision-making and long-term planning. We present PMA-Diffusion, a Physics-guided Mask-Aware Diffusion framework designed to reconstruct high-resolution traffic state from incomplete, sparse, noisy observations. PMA-Diffusion learns a mask-aware diffusion prior directly from sparse urban sensing data and employs an iterative posterior sampling that alternates denoising, observation replacement, and physics-guided projection step. On the I-24 MOTION dataset with only 5% sensor coverage, PMA-Diffusion outperformed physics-agnostic baselines and achieved nearly the performance of fully supervised models. By enabling accurate, high-resolution traffic state estimation in data-sparse environments, this work demonstrates how state-of-the-art AI methodologies can be applied to enhance the scalability and robustness of urban transportation and mobility systems. More broadly, our approach highlights a path toward applying physics-guided generative AI to other smart-city applications, such as energy grids, water distribution, and environmental monitoring, where sparse data remains a critical challenge but high-resolution information is essential for reliable, resilient, and sustainable smart city applications.

1 Introduction

Accurate, high-resolution traffic states underpin smart-city mobility analytics and operations. It can support both ATMS (e.g., ramp metering [1], variable speed limit [2], coordinated traffic signal control [3], and incident detection [4]) and ATIS (e.g., travel time reliability measures [5], dynamic route pre-trip planning [6], and en-route traveler advisories [7, 8]). Yet urban mobility sensing (loop detectors, probes) is sparse, noisy, and unevenly distributed [9, 10]. Camera networks can yield dense ground truth for research, but are costly to deploy broadly [11]. This sensing gap motivates Traffic State Estimation (TSE).

TSE aims to infer missing states from sparse, noisy measurements. Classical model-driven approaches leverage macroscopic flow theory (e.g., LWR and variants of this idea [12, 13, 14, 9]), but suffer when the chosen model or calibration mismatches reality. Purely data-driven estimators (ARIMA / PGM / k-NN / GAN / Flow / VAE) [15, 16, 17, 18, 19, 20, 21] can learn rich correlations but often lack

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

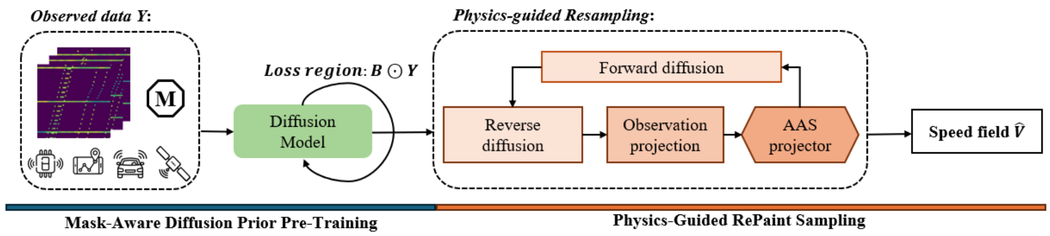


Figure 1: Overview of PMA-Diffusion.

interpretability and require a large amount of representative historical data. Hybrid physics-guided methods embed domain knowledge into learning [22, 23, 24, 25], improving data efficiency, yet they still face how to encode physics robustly while remaining flexible to changing sensing patterns.

Recently, diffusion models have emerged as powerful priors for conditional generation [26]. In TSE, one line treats diffusion as probabilistic inference tools [27, 28, 29]; this is effective but typically assumes fixed or predefined missingness. A second, more flexible line learns a prior first and conditions at inference via posterior sampling [30, 31, 32]. However, (i) training such priors usually requires fully observed data, which are scarce in transportation, and (ii) conditioning without physics can yield reconstructions that conflict with established traffic flow theory.

We address both issues with **PMA-Diffusion**, a physics-guided, mask-aware framework. It learns a mask-aware prior directly from incomplete data with a double-mask strategy inspired by Ambient Diffusion [33], and performs posterior sampling that alternates reverse diffusion, observation projection, and a physics projector that enforces kinematic-wave coherence without altering measurements. These two components jointly enable PMA-Diffusion to reconstruct high-resolution traffic states from sparse and heterogeneous measurements while preserving fundamental traffic flow principles.

2 Methodology

We develop **PMA-Diffusion**, a two-stage framework designed to reconstruct high-resolution spatiotemporal traffic states from sparse, heterogeneous, and noisy measurements, overview see Figure 1.

2.1 Problem Setting

We discretize a freeway segment into S spatial cells and T time steps, yielding a normalized speed field $\mathbf{V} \in [0,1]^{S\times T}$. A binary mask $\mathbf{M} \in \{0,1\}^{S\times T}$ indicates visible entries from fixed loop detectors (fixed spatial pattern) and probe vehicles (stochastic spatiotemporal pattern). We assume negligible probe latency and absorb low-frequency bias into the observation noise; the task is to infer the unobserved entries of $\widehat{\mathbf{V}}$ from $(\mathbf{Y}, \mathbf{M})$, where $\mathbf{Y}$ is observed noisy speed field.

2.2 Mask-Aware Diffusion Prior

Standard DDPMs [26] presume fully observed training data. We instead adopt a *mask-aware Ambient Diffusion* strategy [33]: each training sample is paired with its visibility mask and corrupted with Gaussian noise; the denoiser receives $(\mathbf{Y}_t, t, \mathbf{M})$ and predicts the injected noise only where data are visible. Concretely, with the usual variance schedule and noise-prediction parameterization,

$$\mathcal{L}(\theta) = \mathbb{E}\big[\lambda(t)\,\|\mathbf{M} \odot (h_\theta(\mathbf{Y}_t, t, \mathbf{M}) - \boldsymbol{\eta})\|_F^2\big], \tag{1}$$

which trains a prior $p_\theta(\mathbf{V})$ that captures spatiotemporal structure despite incomplete supervision. Here the denoiser $h_\theta(\cdot)$ is trained to predict the injected noise $\boldsymbol{\eta}$, $\lambda(t)$ re-weights time steps according to the signal-to-noise ratio, $\odot$ is element-wise (Hadamard) product, and the Frobenius norm $\|\cdot\|_F$ sums squared errors over $\mathbf{M}_{s,t} = 1$.

Double-mask for spatial generalization. When some cells are *always* observed (e.g., near detectors), the network risks overfitting their neighborhood and neglecting persistently invisible regions. We therefore introduce a *double-mask* scheme: for each sample we draw an auxiliary mask $\mathbf{B}$

(Bernoulli with small hide probability p_{hide} or from the empirical mask distribution) and define the loss region by $\tilde{\mathbf{M}} = \mathbf{B} \odot \mathbf{M}$. This ensures that *every* cell occasionally contributes to the loss, improving spatial generalization without requiring fully observed ground truth.

2.3 Physics-Guided RePaint Sampling

At inference, we approximate the posterior with a RePaint-style loop [30] that alternates three steps:

1. **Reverse diffusion** (toward the learned prior): $\tilde{\mathbf{Z}}_t = g_\theta(\mathbf{Z}_{t+1}, t{+}1)$.
2. **Observation projection** (exact fidelity on $\mathbf{M}{=}1$): $\hat{\mathbf{Z}}_t = \mathbf{Y} \odot \mathbf{M} + \tilde{\mathbf{Z}}_t \odot (1 - \mathbf{M})$.
3. **Physics projection** (coherence on $\mathbf{M}{=}0$): $\mathbf{Z}_t = P_{\text{phys}}(\hat{\mathbf{Z}}_t, \mathbf{M})$.

Optionally, a forward jump is inserted as in RePaint to improve mixing. The observation projection guarantees measurement consistency; the physics projector modifies *only* unobserved entries:

$$P_{\text{phys}}(\mathbf{V}, \mathbf{M}) \odot \mathbf{M} = \mathbf{V} \odot \mathbf{M} \quad \text{(mask-invariance).} \tag{2}$$

Mask-invariant physics projector. We use a lightweight *Adaptive Anisotropic Smoothing* (AAS) operator guided by traffic characteristics. Let V_s be an anisotropically smoothed surrogate that blends kernels aligned with free-flow and congested wave slopes; let $R \approx \partial_t V + c\, \partial_x V$ be a small first-order transport residual. The update on the missing subspace is

$$\mathbf{V} \leftarrow \operatorname{clip}\Big(\mathbf{V} - \alpha_{\text{smooth}}(\mathbf{V} - V_s) - \alpha_{\text{char}} R,\ 0, 1\Big) \ \text{ on } (1 - \mathbf{M}), \tag{3}$$

which encourages kinematic-wave coherence without requiring density/flow labels or altering observed values. In practice, AAS is no learned parameters at test time once its few hyperparameters are set, and it composes modularity with other projectors if richer physics become available.

3 Experimental Evaluation

We evaluate **PMA-Diffusion** on the I-24 MOTION dataset, focusing on two questions: **RQ1**: How much prior quality is lost when training on incomplete speed fields? **RQ2**: Can RePaint resampling and physics guidance recover this loss under sparse sensing?

3.1 Dataset and Observation Models

I-24 MOTION provides full-resolution traffic states over 4.2 miles of freeway with near-complete coverage from 276 cameras [11]. Trajectories are rasterized into $S{=}64$ spatial cells ($\Delta x = 200$ft) and $T{=}64$ time slices ($\Delta t = 5$s). We synthetically mask the ground truth to simulate loop detectors (5–50% of rows visible) and probe vehicles (Poisson rate $\lambda = 0, 5, 15, 25$).

3.2 Evaluation Metrics and Models

We report performance on unobserved pixels using: (1) **Masked-MSE (2×2)** for low-frequency fidelity; (2) **Sobel-MSE** for edge sharpness; (3) **LPIPS** for perceptual similarity [34].

We compare three training schemes: *Full-obs*, *Single-mask*, and *Double-mask*. For sampling, we test *Ambient one-shot*, *RePaint* [30], and *RePaint + AAS projector*.

3.3 Results

Upper bound (Full-obs). With 5% rows and $\lambda{=}0$, Ambient attains Masked-MSE 0.0945; a single RePaint pass reduces it to 0.0612, and adding physics guidance lowers it further to ~ 0.018 (Table 2). This shows posterior resampling and lightweight physics are beneficial even with a strong prior.

Single-mask. Training strictly on visible pixels can collapse in persistently unseen regions: at 5% rows, $\lambda{=}0$, Ambient has Masked-MSE > 27 and LPIPS ≈ 1.004. RePaint alone may over-extrapolate, but *RePaint+Physics* stabilizes sampling and narrows the gap to *Full-obs* (within ~ 0.035), see Fig. 2.

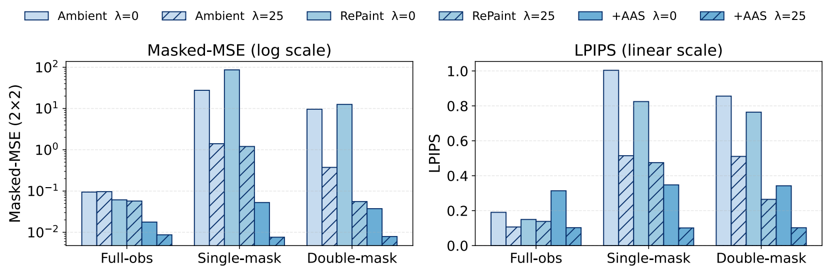
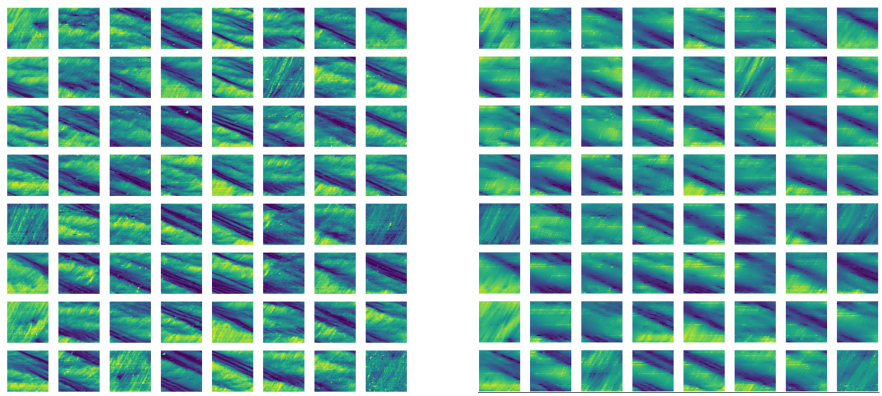


Figure 2: Comparison across training and sampling schemes under extreme sparsity (row= 5%, $\lambda \in \{0, 25\}$).



(a) Ground truth (b) RePaint + Physics (AAS)

Figure 3: Qualitative comparison under *Double-mask* (row= 25%, λ=15).

Double-mask. Randomly hiding some always-visible cells during training markedly improves robustness. In the worst case (5% rows, λ=0) it is an order of magnitude better than *Single-mask* and only 0.0197 above *Full-obs*; gains also hold for LPIPS/Sobel-MSE. At 5% rows, λ=25, *Single-mask+RePaint+Physics* can slightly edge out *Double-mask*, suggesting aggressive random hiding is less critical if every pixel is occasionally observed historically.

Sampler and physics trends. Across all settings (48 combinations in Table 2): (i) *RePaint* outperforms one-shot Ambient, especially with stronger priors; (ii) the AAS projector never degrades metrics and typically reduces both Masked-MSE and Sobel-MSE; (iii) benefits are largest under extreme sparsity (5% rows), where physics guidance sharpens fronts and removes structural artifacts. With a near-optimal *Full-obs* prior, physics may slightly smooth very high-frequency details (Sobel-MSE marginally higher) while preserving perceptual quality.

Qualitative. Fig. 3 compares *Double-mask* + *RePaint+Physics* (row= 25%, λ=15) with ground truth. Reconstructions preserve macroscopic wave propagation and regime transitions in unobserved regions; very fine-scale details are naturally attenuated by posterior regularization.

4 Conclusion

We presented **PMA-Diffusion**, a physics-guided, mask-aware diffusion framework for high-resolution Traffic State Estimation under sparse, noisy, and uneven urban sensing. On the I-24 MOTION dataset with only 5% sensor coverage, PMA-Diffusion outperforms physics-agnostic baselines and approaches the accuracy of fully supervised models. By enabling accurate high-resolution estimates in data-sparse settings, the approach supports scalable and robust ITS operations in smart cities and points toward applying physics-guided generative AI to other urban infrastructures. Limitations include assumptions on time alignment and low-frequency bias; future work will relax these assumptions and profile runtime, energy, and real-time latency across sparsity regimes to facilitate city-scale deployment.

References

[1] Francois Belletti, Daniel Haziza, Gabriel Gomes, and Alexandre M Bayen. Expert level control of ramp metering based on multi-task deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 19(4):1198–1207, 2017.

[2] Yuhang Zhang, Zhiyao Zhang, Junyi Ji, Marcos QuiÃ±ones-Grueiro, William Barbour, Derek Gloudemans, Clay Weston, Gautam Biswas, Daniel B Work, et al. Real-world deployment and assessment of a multi-agent reinforcement learning-based variable speed limit control system. *arXiv preprint arXiv:2503.01017*, 2025.

[3] Afzal Ahmed, Syed Ahsan Ali Naqvi, David Watling, and Dong Ngoduy. Real-time dynamic traffic control based on traffic-state estimation. *Transportation research record*, 2673(5):584–595, 2019.

[4] Austin Coursey, Junyi Ji, Zhiyao Zhang, William Barbour, Marcos Quinones-Grueiro, Tyler Derr, Gautam Biswas, and Daniel B Work. Real-time freeway traffic anomalous event detection system via radar detector sensors. In *Proceedings of the ACM/IEEE 16th International Conference on Cyber-Physical Systems (with CPS-IoT Week 2025)*, pages 1–2, 2025.

[5] Henry X Liu, Will Recker, and Anthony Chen. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A: Policy and Practice*, 38(6):435–453, 2004.

[6] Afzal Ahmed, Dong Ngoduy, and David Watling. Prediction of traveller information and route choice based on real-time estimated traffic state. *Transportmetrica B: Transport Dynamics*, 4(1):23–47, 2016.

[7] Thomas Liebig, Nico Piatkowski, Christian Bockermann, and Katharina Morik. Dynamic route planning with real-time traffic predictions. *Information Systems*, 64:258–265, 2017.

[8] Usue Mori, Alexander Mendiburu, Maite Álvarez, and Jose A Lozano. A review of travel time estimation and forecasting for advanced traveller information systems. *Transportmetrica A: Transport Science*, 11(2):119–157, 2015.

[9] Martin Treiber, Arne Kesting, and R Eddie Wilson. Reconstructing the traffic state by fusion of heterogeneous data. *Computer-Aided Civil and Infrastructure Engineering*, 26(6):408–419, 2011.

[10] Xudong Wang, Yuankai Wu, Dingyi Zhuang, and Lijun Sun. Low-rank hankel tensor completion for traffic speed estimation. *IEEE Transactions on Intelligent Transportation Systems*, 24(5):4862–4871, 2023.

[11] Derek Gloudemans, Yanbing Wang, Junyi Ji, Gergely Zachar, William Barbour, Eric Hall, Meredith Cebelak, Lee Smith, and Daniel B Work. I-24 motion: An instrument for freeway traffic science. *Transportation Research Part C: Emerging Technologies*, 155:104311, 2023.

[12] Michael James Lighthill and Gerald Beresford Whitham. On kinematic waves ii. a theory of traffic flow on long crowded roads. *Proceedings of the royal society of london. series a. mathematical and physical sciences*, 229(1178):317–345, 1955.

[13] Benjamin Coifman. Estimating travel times and vehicle trajectories on freeways using dual loop detectors. *Transportation Research Part A: Policy and Practice*, 36(4):351–364, 2002.

[14] Martin Treiber and Dirk Helbing. Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooper@ tive Tr@ nsport@ tion Dyn@ mics*, 1(3):3–1, 2002.

[15] Zhao Zhang, Xianfeng Terry Yang, and Hao Yang. A review of hybrid physics-based machine learning approaches in traffic state estimation. *Intelligent Transportation Infrastructure*, 2:liad002, 2023.

[16] Ming Zhong, Pawan Lingras, and Satish Sharma. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. *Transportation Research Part C: Emerging Technologies*, 12(2):139–166, 2004.

[17] Daiheng Ni and John D Leonard. Markov chain monte carlo multiple imputation using bayesian networks for incomplete intelligent transportation systems data. *Transportation research record*, 1935(1):57–67, 2005.

[18] Sehyun Tak, Soomin Woo, and Hwasoo Yeo. Data-driven imputation method for traffic data in sectional units of road links. *IEEE Transactions on Intelligent Transportation Systems*, 17(6):1762–1771, 2016.

[19] Zhaobin Mo, Yongjie Fu, and Xuan Di. Quantifying uncertainty in traffic state estimation using generative adversarial networks. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2769–2774. IEEE, 2022.

[20] Han Huang, Jiajia Yu, Jie Chen, and Rongjie Lai. Bridging mean-field games and normalizing flows with trajectory regularization. *Journal of Computational Physics*, 487:112155, 2023.

[21] Guillem Boquet, Antoni Morell, Javier Serrano, and Jose Lopez Vicario. A variational autoencoder solution for road traffic forecasting systems: Missing data imputation, dimension reduction, model selection and anomaly detection. *Transportation Research Part C: Emerging Technologies*, 115:102622, 2020.

[22] Rongye Shi, Zhaobin Mo, and Xuan Di. Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 540–547, 2021.

[23] Chenguang Zhao and Huan Yu. Observer-informed deep learning for traffic state estimation with boundary sensing. *IEEE Transactions on Intelligent Transportation Systems*, 25(2):1602–1611, 2023.

[24] Zeyu Shi, Yangzhou Chen, Jichao Liu, Dechao Fan, and Chaoqiang Liang. Physics-informed spatiotemporal learning framework for urban traffic state estimation. *Journal of Transportation Engineering, Part A: Systems*, 149(7):04023056, 2023.

[25] Jinlei Zhang, Shuai Mao, Lixing Yang, Wei Ma, Shukai Li, and Ziyou Gao. Physics-informed deep learning for traffic state estimation based on the traffic flow model and computational graph method. *Information Fusion*, 101:101971, 2024.

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

[27] Da Lei, Min Xu, and Shuaian Wang. A conditional diffusion model for probabilistic estimation of traffic states at sensor-free locations. *Transportation Research Part C: Emerging Technologies*, 166:104798, 2024.

[28] Huipeng Zhang, Honghui Dong, and Zhiqiang Yang. Tsgdiff: Traffic state generative diffusion model using multi-source information fusion. *Transportation Research Part C: Emerging Technologies*, 174:105081, 2025.

[29] Bo Lu, Qinghai Miao, Yahui Liu, Tariku Sinshaw Tamir, Hongxia Zhao, Xiqiao Zhang, Yisheng Lv, and Fei-Yue Wang. A diffusion model for traffic data imputation. *IEEE/CAA Journal of Automatica Sinica*, 12(3):606–617, 2025.

[30] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022.

[31] Hyungjin Chung and Jong Chul Ye. Deep diffusion image prior for efficient ood adaptation in 3d inverse problems. In *European Conference on Computer Vision*, pages 432–455. Springer, 2024.

[32] Riccardo Barbano, Alexander Denker, Hyungjin Chung, Tae Hoon Roh, Simon Arridge, Peter Maass, Bangti Jin, and Jong Chul Ye. Steerable conditional diffusion for out-of-distribution adaptation in medical image reconstruction. *IEEE Transactions on Medical Imaging*, 2025.

[33] Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, and Adam Klivans. Ambient diffusion: Learning clean distributions from corrupted data. *Advances in Neural Information Processing Systems*, 36:288–313, 2023.

[34] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[35] Fan Wu, Zhanhong Cheng, Huiyu Chen, Zhijun Qiu, and Lijun Sun. Traffic state estimation from vehicle trajectories with anisotropic gaussian processes. *Transportation Research Part C: Emerging Technologies*, 163:104646, 2024.

[36] Seongjin Choi, Zhixiong Jin, Seung Woo Ham, Jiwon Kim, and Lijun Sun. A gentle introduction and tutorial on deep generative models in transportation research. *Transportation Research Part C: Emerging Technologies*, 176:105145, 2025.

[37] 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.

A Methodology Details

A.1 Notation and Observation Model

We discretize the corridor into S spatial cells and T time steps. The latent, normalized speed field is $\mathbf{V} \in [0,1]^{S\times T}$. Our observed data comes from two asynchronous sources: loop detectors and probe vehicles. Their spatial-temporal coverage is illustrated in Figure 4.

- **Loop detectors.** Loop detectors are fixed-point sensors installed at selected highway segments. Their observations follow a static spatial pattern and are represented by a binary mask: $\mathbf{M}^{(\ell)} \in \{0,1\}^{S\times T}$, where $\mathbf{M}^{(\ell)} = 1$ indicates an observed entry. Measurement noise is modeled as $\boldsymbol{\varepsilon}^{(\ell)} \sim \mathcal{N}(0, \sigma_\ell^2)$.
- **Probe vehicles.** Probe vehicle data are generated from moving vehicles equipped with GPS or other communication devices. Unlike loop detectors, probe vehicle data exhibit a stochastic sampling pattern across space and time, represented by $\mathbf{M}^{(p)} \in \{0,1\}^{S\times T}$. Measurement noise is modeled as $\boldsymbol{\varepsilon}^{(p)} \sim \mathcal{N}(0, \sigma_p^2)$.

Here, the combined observation mask is defined as $\mathbf{M} = \mathbf{M}^{(\ell)} \vee \mathbf{M}^{(p)}$, where $\vee$ denotes the element-wise logical OR. In typical highway monitoring scenarios, the coverage ratio $\rho = \|\mathbf{M}\|_0 \big/ (ST)$ is very small, satisfying $\rho \ll 1$. Incorporating a possible GPS latency τ, the observation operator is written as:

$$\mathcal{O}_\rho(\mathbf{V}) = \mathbf{M}^{(\ell)} \odot (\mathbf{V} + b^{(\ell)} + \varepsilon^{(\ell)}) + \mathbf{M}^{(p)} \odot (\mathbf{V} \circ \mathsf{S}_{-\tau} + b^{(p)} + \varepsilon^{(p)}), \tag{4}$$

where $\odot$ denotes the element-wise product, $b^{(a)}$ denotes a *slowly varying bias term specific to sensor* a that is smooth in spatiotemporal and small in magnitude relative to the dynamic range, $\mathsf{S}_{-\tau}$ shifts time backward by the random GPS latency τ.

A.1.1 Empirical Evidence and Assumptions

Recent physics-informed and data-driven TSE studies typically assume that probe latency is negligible relative to the aggregation interval and that low-frequency sensor biases can be absorbed into the observation noise without explicit modeling[22, 35, 10]. Hence we adopt:

Assumption 1 (Negligible latency)**.** *We assume $\tau = 0$ with probability one; probe vehicle data are effectively time-aligned with loop detector data.*

Assumption 2 (Bounded low-frequency drift)**.** *Any smooth bias from sensor a, $b^{(a)}$, is L_b–Lipschitz with $\|b^{(a)}\|_\infty \le \delta$. We absorb $b^{(a)}$ into $\boldsymbol{\varepsilon}^{(a)}$; thus $\mathrm{E}[\boldsymbol{\varepsilon}^{(a)}|\mathbf{V}] = \mathbf{0}$ and the induced score bias is $O(\delta)$.*

Remark 1. *Write the (single-source) conditional log-likelihood as*

$$\log p(Y \mid V) = \log p_\eta\big(Y - V - b^{(a)}\big). \tag{5}$$

Since $b^{(a)}$ is Lipschitz-continuous and small in amplitude ($\|b^{(a)}\|_\infty \le \delta$), a first-order Taylor expansion around $b^{(a)} = 0$ gives

$$\log p_\eta(Y - V - b^{(a)}) = \log p_\eta(Y - V) - \nabla_y \log p_\eta(Y - V)^{\mathrm{T}}\, b^{(a)} + O(\delta^2). \tag{6}$$

Differentiating with respect to $\mathbf{V}$ (to obtain the score) shows that the true score is shifted by at most $O(\delta)$, because $\|\nabla_y \log p_\eta\|$ is bounded under the sub-Gaussian noise model. Hence replacing $(Y - V - b^{(a)})$ by $(Y - V)$ introduces only a vanishing $O(\delta)$ bias in the masked score estimator. The constant L_b controls spatial/temporal smoothness (frequency content), whereas δ controls its amplitude; we do not *assume they are equal ($L_b = \delta$), only that δ is sufficiently small for the expansion to hold.*

Under Assumptions 1-2, we set $\tau = 0$ and drop $\mathsf{S}_{-\tau}$ (both sources share the same Eulerian frame). We thus absorb $b^{(a)}$ into the zero-mean noise term, which perturbs $\mathrm{E}[\boldsymbol{\varepsilon}^{(a)}|\mathbf{V}]$ only by $O(\|b^{(a)}\|_\infty)$,

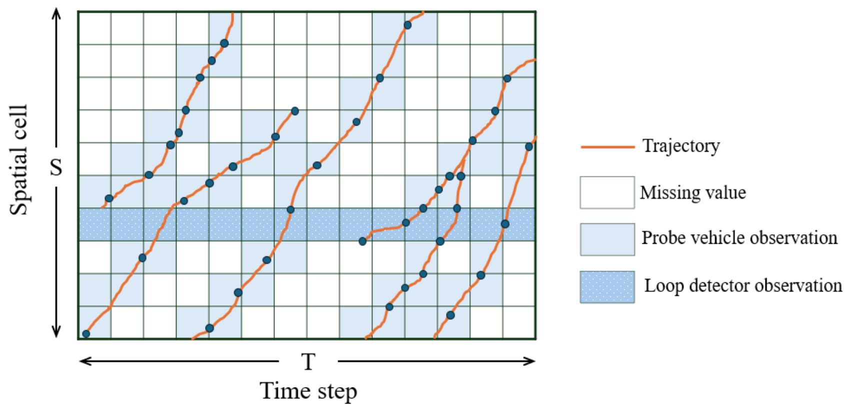


Figure 4: Matrix representation of traffic state variables collected from probe vehicles and loop detectors.

leaving the masked score matching objective unbiased up to a vanishing $O(\delta)$ term. Under these assumptions, the observation model simplifies to

$$\mathbf{Y} \;=\; \mathbf{M}^{(\ell)} \odot \big(\mathbf{V} + \boldsymbol{\varepsilon}^{(\ell)}\big) \;+\; \mathbf{M}^{(p)} \odot \big(\mathbf{V} + \boldsymbol{\varepsilon}^{(p)}\big), \tag{7}$$

which remains statistically equivalent for training and inference in our diffusion framework. To ensure unbiased estimation under masking conditions, we introduce the following assumptions:

Assumption 3 (Noise conditional independence)**.** *Given the latent speed field* $\mathbf{V}$*, loop detector and probe vehicle noise fields are independent:* $\boldsymbol{\varepsilon}^{(\ell)} \perp\!\!\!\perp \boldsymbol{\varepsilon}^{(p)} \mid \mathbf{V}$*. Consequently the joint likelihood factorizes over sources.*

Assumption 4 (Masking missing-at-random (MAR))**.** *Given* $\mathbf{V}$ *the mask process is independent of the noise,*$\mathbf{M}^{(a)} \perp\!\!\!\perp \boldsymbol{\varepsilon}^{(a)} \mid \mathbf{V}$*. Hence, the masked denoising loss used in diffusion training remains an unbiased estimator of the true score field.*

These assumptions guarantee that the masked denoising loss used in our training procedure is an unbiased proxy for the true score function.

A.2 Mask-Aware Ambient Diffusion Prior

Background: Diffusion Models

Diffusion models [26] are a class of likelihood-based generative models that have demonstrated state-of-the-art performance in image and spatio-temporal data generation. They define a two-phase stochastic process: a *forward diffusion process*, which gradually perturbs a clean sample by adding Gaussian noise, and a *reverse denoising process*, which learns to invert this corruption and recover the original data distribution.

Forward process. Given $\mathbf{x}_0 \sim q(\mathbf{x}_0)$, the forward process constructs a Markov chain

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\big(\mathbf{x}_t; \sqrt{1-\beta_t}\,\mathbf{x}_{t-1},\, \beta_t\mathbf{I}\big), \quad t = 1,\dots,T, \tag{8}$$

where $\{\beta_t\}$ is a variance schedule. This admits a closed-form for any t:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\varepsilon}, \quad \bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s),\ \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I}). \tag{9}$$

Reverse process and training. To generate new samples, the model learns a reverse process

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\big(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t,t), \sigma_t^2\mathbf{I}\big), \tag{10}$$

where $\boldsymbol{\mu}_\theta$ is computed from the predicted noise $\boldsymbol{\varepsilon}_\theta$. The training objective minimizes

$$\mathcal{L}(\theta) = \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\varepsilon}}\big[\|\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\varepsilon}, t)\|^2\big]. \tag{11}$$

This generative framework is attractive for highway monitoring because it captures high-dimensional spatiotemporal correlations and provides a principled way to sample from a Bayesian posterior.

A.2.1 Sampling Objective: Bayesian Target Posterior

Under the working assumptions stated in the previous section, the observation model reduces to

$$\mathbf{Y} \;=\; \mathbf{M} \odot \big(\mathbf{V} + \boldsymbol{\varepsilon}\big),\; \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I}). \tag{12}$$

Given $\mathbf{Y}$ and $\mathbf{M}$, our inferential goal is to draw samples from the posterior distribution:

$$p(\mathbf{V} \mid \mathbf{Y}) \;\propto\; \underbrace{\mathbf{1}\big[\mathbf{V} \in \mathcal{P}_{\text{phys}}\big]}_{\text{physics surrogate}} \;\times\; \underbrace{p_\theta(\mathbf{V})}_{\text{mask-aware diffusion prior}} \;\times\; \underbrace{\exp\Big[-\frac{\|(\mathbf{V}\odot\mathbf{M}) - \mathbf{Y}\|_F^2}{2\sigma^2}\Big]}_{\text{masked likelihood}}, \tag{13}$$

where $p_\theta(\mathbf{V})$ is the pre-trained mask-aware prior learned from historical data and $\mathcal{P}_{\text{phys}}$ is the image of the projection operator P_{phys}.

Remark 2. *Bayes' rule factorizes the posterior as*

$$p(\mathbf{V} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \mathbf{V})\, p(\mathbf{V}). \tag{14}$$

Because the noise acts only on visible pixels, the likelihood becomes:

$$p(\mathbf{Y} \mid \mathbf{V}) = \prod_{(s,t):\, M_{s,t}=1} \frac{1}{\sqrt{2\pi}\,\sigma}\, \exp\Big[-\frac{(Y_{s,t} - V_{s,t})^2}{2\sigma^2}\Big], \tag{15}$$

which is exactly the exponential term in (13)*. This* masked likelihood *enforces perfect agreement with the measurements on the support of* $\mathbf{M}$ *while leaving unobserved entries unconstrained. To inject prior knowledge, we multiply by the data-driven prior* $p_\theta(\mathbf{V})$ *and by the indicator* $\mathbf{1}[\mathbf{V} \in \mathcal{P}_{phys}]$*, which restricts the support to speed fields that satisfy the surrogate-physics constraints. Combining these three factors yields Eq.* (13)*.*

To sample from Eq. (13), we need to introduce a *Mask-Aware Ambient Diffusion Prior* and a *Physics-Guided RePaint Sampler* that trains a diffusion prior from sparsely observed speed fields and approximates this posterior by alternating reverse diffusion, observation projection, and physics projection.

A.2.2 Forward corruption

Let $\mathbf{Y}_0 = \mathbf{M} \odot (\mathbf{V} + \boldsymbol{\varepsilon})$ be the partially observed speed field. We apply the VP SDE

$$\mathrm{d}\mathbf{Y}_t = -\tfrac{1}{2}\beta(t)\,\mathbf{Y}_t\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{W}_t, \tag{16}$$

with linear $\beta(t)$ so that

$$\mathbf{Y}_t = \sqrt{\alpha(t)}\,\mathbf{Y}_0 + \sqrt{1-\alpha(t)}\,\boldsymbol{\eta}, \qquad \alpha(t) = \exp\Big(-\int_0^t \beta(s)\,\mathrm{d}s\Big),\;\; \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{17}$$

A.2.3 Masked score matching

A mask-aware denoiser $h_\theta(\cdot)$ predicts the injected noise *only on visible entries*:

$$\mathcal{L}(\theta) = \mathbb{E}\big[\lambda(t)\,\|\mathbf{M} \odot \big(h_\theta(\mathbf{Y}_t, t, \mathbf{M}) - \boldsymbol{\eta}\big)\|_F^2\big], \qquad \lambda(t) = \frac{\alpha(t)}{1-\alpha(t)}. \tag{18}$$

With noise-prediction parameterization, the (conditional) score is approximated by

$$\nabla_{\mathbf{Y}_t} \log p_t(\mathbf{Y}_t) \;\approx\; -\frac{1}{\sigma_t}\, h_\theta(\mathbf{Y}_t, t, \mathbf{M}). \tag{19}$$

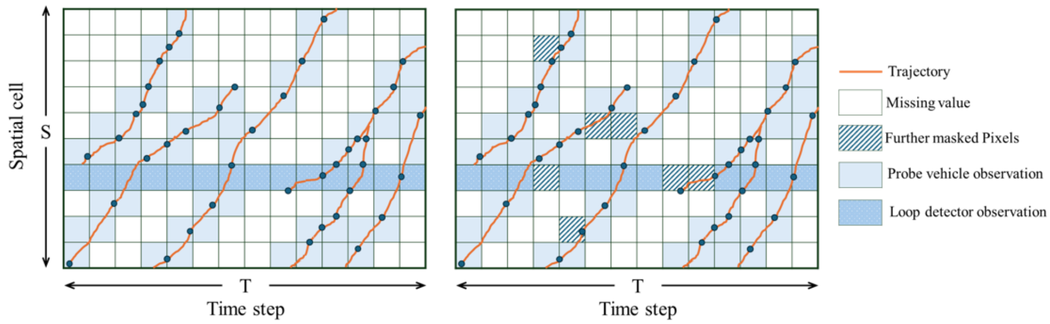


Figure 5: Illustration of the double-mask strategy. Left: original observed data under the single-mask scheme, reflecting actual sparse sensor coverage. Right: the same frame with further masked pixels, simulating the observation pattern under the double-mask scheme.

A.2.4 Mask-selection strategies

Observation masks are heterogeneous: some pixels (e.g., near fixed detectors) are almost always visible, while others remain persistently hidden. To account for this, we consider two complementary training schemes summarized in Table 1.

In the *single-mask* strategy, the true sensor mask $\mathbf{M}$ is left untouched; the network is therefore penalized only on *truly known* pixels. This works well when probe-vehicle coverage is dense and roughly random. Under such conditions every spatial-temporal position is observed at least occasionally. In the *double-mask* strategy, when the overall visibility ratio $\rho = \|\mathbf{M}\|_0/(ST)$ is high, we draw the auxiliary mask $\mathbf{B}$ from the empirical mask distribution $\pi(\mathbf{M})$ captured in the raw data. When ρ is low, we instead sample $\mathbf{B}$ from a Bernoulli distribution with parameter p_{extra}, so that *every* coordinate is occasionally masked and therefore contributes to the loss. In practice we use $p_{\text{extra}} = 0.05$, which improves interpolation near fixed loop locations while incurring negligible additional variance, see Figure 5.

A.2.5 Unbiasedness under MAR

Proposition A.1 (Masked score unbiasedness)**.** *Under Assumptions (Noise conditional independence) and (MAR), and with bounded/small drift absorbed into noise, the masked objective above is an* unbiased estimator *of the conditional score on unobserved entries given the observed ones, up to $O(\delta)$ from the small drift.*

Sketch. Because $\mathbf{M} \perp\!\!\!\perp \varepsilon \mid \mathbf{V}$ and the noise acts only where $\mathbf{M}{=}1$, the masked log-likelihood factorizes over visible entries. The gradient w.r.t. $\mathbf{Y}_t$ equals the conditional score on the visible set; the MAR assumption makes the empirical masked average an unbiased estimator of the population objective. Smooth, bounded bias $b^{(a)}$ perturbs the score by $O(\|b^{(a)}\|_\infty) = O(\delta)$ via a first-order Taylor expansion around zero drift (details in the main text). □

Table 1: Mask-selection strategies for training the ambient diffusion prior.

Strategy	Cells kept	Regions contributing to loss	When to prefer
Single mask	$\mathbf{M}_{s,t} = 1$	$\mathbf{M}$	All pixels have non–zero probability of being observed.
Double mask	$\mathbf{M}_{s,t} = 1$	$\tilde{\mathbf{M}} = \mathbf{B} \odot \mathbf{M}$	Some cells are *always* invisible (e.g. fixed loop detectors)

A.3 Surrogate Physics Posterior and RePaint Sampling

Our target distribution is the *surrogate physics posterior*

$$p(\mathbf{V} \mid \mathbf{Y}) \;\propto\; \underbrace{\mathbf{1}\big[\mathbf{V} \in \mathcal{P}_{\text{phys}}\big]}_{\text{surrogate physics}} \times \underbrace{p_\theta(\mathbf{V})}_{\text{mask-aware prior}} \times \underbrace{\exp\Big[-\tfrac{\|\mathbf{M}\odot(\mathbf{V}-\mathbf{Y})\|_F^2}{2\sigma^2}\Big]}_{\text{masked likelihood}}. \tag{20}$$

Algorithm A.1 (Physics-guided RePaint). Initialize $\mathbf{Z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$; for $t = T-1 \to 0$:

1. **Reverse (denoise):** $\tilde{\mathbf{Z}}_t \leftarrow g_\theta(\mathbf{Z}_{t+1}, t+1)$.
2. **Observation projection:** $\hat{\mathbf{Z}}_t \leftarrow \mathbf{Y} \odot \mathbf{M} + \tilde{\mathbf{Z}}_t \odot (1 - \mathbf{M})$.
3. **Physics projection:** $\bar{\mathbf{Z}}_t \leftarrow P_{\text{phys}}(\hat{\mathbf{Z}}_t, \mathbf{M})$.
4. *(Optional) Forward jump:* $\mathbf{Z}_t \leftarrow f_j(\bar{\mathbf{Z}}_t, t)$ (as in RePaint) to improve mixing.

Repeat to obtain i.i.d. posterior samples $\{\mathbf{Z}_0^{(n)}\}$.

Proposition A.2 (Exact data fidelity at all steps)**.** *If P_{phys} is* mask-invariant*, i.e.*

$$P_{phys}(\mathbf{V}, \mathbf{M}) \odot \mathbf{M} \;=\; \mathbf{V} \odot \mathbf{M}, \tag{21}$$

then every iterate after Step 2 (and hence after Step 3) satisfies $\mathbf{Z}_t \odot \mathbf{M} = \mathbf{Y} \odot \mathbf{M}$.

If P_{phys} coincides with a proximal map of a convex physics penalty $\varphi(\cdot)$ on the missing set, Algorithm A.1 resembles an alternating stochastic proximal sampler for $-\log p_\theta$ plus data and physics terms; our AAS projector is a stable, local surrogate of such a proximal step.

A.4 Physics Projector: Adaptive Anisotropic Smoothing (AAS)

A.4.1 Construction

The current instantiation of P_{phys} employs an *Adaptive Anisotropic Smoothing* (AAS) operator, which extends the Generalized Adaptive Smoothing Method (GASM) [9] into a mask-aware and diffusion-compatible form. Unlike full conservation enforcement, which is infeasible under speed-only observations, AAS acts as a local regularizer that promotes physically plausible structures in unobserved regions while preserving exact fidelity on measured cells, see Algorithm 1.

A.4.2 Stability and non-expansiveness

Lemma A.1 (1-Lipschitz on the missing subspace)**.** *If the convolution kernels are ℓ_1-normalized and $0 \le \alpha_{smooth} \le 1$, $0 \le \alpha_{char} \le \bar{\alpha}$ with $\bar{\alpha}$ sufficiently small, then $P_{phys}(\cdot, \mathbf{M})$ is non-expansive on* $\{\mathbf{Z} : \mathbf{Z} \odot \mathbf{M} = 0\}$.

Sketch. Anisotropic convolution with normalized kernels has operator norm ≤ 1 (Young's inequality). The first term is a convex combination of V and V_s on $(1 - \mathbf{M})$, hence contractive when $\alpha_{\text{smooth}} \in [0, 1]$. The transport residual is a local first-order finite-difference operator; choosing α_{char} small makes the composite map non-expansive. Mask-invariance is immediate because updates are multiplied by $(1 - \mathbf{M})$. □

Practical tip. Use α_{char} at least an order of magnitude smaller than α_{smooth} to ensure numerical damping of high-frequency artifacts while preserving wave alignment.

B Appendix: Experimental Details

We conduct a comprehensive empirical evaluation to assess the effectiveness of **PMA-Diffusion**. The experiments are structured around two central research questions, each corresponding to two stages in the methodology:

RQ1 – Prior quality: *How much accuracy is lost when the diffusion prior is pretrained on incomplete speed fields rather than on fully observed data?*

Algorithm 1 AAS Projector: Physics-guided Mask-invariant Update

Require: Speed field $V \in [0,1]^{S\times T}$; mask $\mathbf{M} \in \{0,1\}^{S\times T}$; physical parameters $(v_{\text{thr}}, \beta, c^{\text{phys}}_{\text{free}}, c^{\text{phys}}_{\text{cong}})$; smoothing weights $(\alpha_{\text{smooth}}, \alpha_{\text{char}})$; kernel stds (σ_{par}, σ_t).

Ensure: Updated field $P_{\text{phys}}(V, \mathbf{M})$.

Compute soft regime gate:

$$p_{\text{free}} \leftarrow 0.5\big(1 + \tanh((V - v_{\text{thr}})/\beta)\big)$$

Convert characteristic speeds to grid units: $c_{\text{free}} \leftarrow c^{\text{phys}}_{\text{free}} \Delta t/\Delta x;\ c_{\text{cong}} \leftarrow c^{\text{phys}}_{\text{cong}} \Delta t/\Delta x$

Build two anisotropic Gaussian kernels $K_{c_{\text{free}}}, K_{c_{\text{cong}}}$ aligned with slopes $c_{\text{free}}, c_{\text{cong}}$ (normalized).

Apply separable convolution: $V_f \leftarrow K_{c_{\text{free}}} * V,\ V_c \leftarrow K_{c_{\text{cong}}} * V$

Compute blended surrogate: $V_s \leftarrow p_{\text{free}} \odot V_f + (1 - p_{\text{free}}) \odot V_c$

Step 1 (smooth relax): $V^{(1)} \leftarrow V - \alpha_{\text{smooth}}(V - V_s) \odot (1 - \mathbf{M})$

if transport correction enabled **then**

  Compute local speed $c_{\text{loc}} \leftarrow p_{\text{free}} c_{\text{free}} + (1 - p_{\text{free}}) c_{\text{cong}}$

  Compute residual: $R \leftarrow (V^{(1)}_{s,t} - V^{(1)}_{s,t-1}) + c_{\text{loc}}(V^{(1)}_{s,t} - V^{(1)}_{s-1,t})$

  Step 2 (residual update): $V^{(2)} \leftarrow V^{(1)} - \alpha_{\text{char}} R \odot (1 - \mathbf{M})$

else

  $V^{(2)} \leftarrow V^{(1)}$

end if

Clip to valid range: $V^{(2)} \leftarrow \text{clip}(V^{(2)}, 0, 1)$

return $P_{\text{phys}}(V, \mathbf{M}) = V^{(2)}$

RQ2 – Sampling guidance: *Given a fixed prior, how much of that loss can be recovered by (i) the RePaint resampling loop and (ii) the physics-guided projector under varying sensing sparsity?*

The proposed method in this study consists of three components: training a diffusion prior directly from sparse observations, applying an iterative resampling strategy during inference, and enforcing physical consistency through a physics-guided projection step. **RQ1** evaluates whether the prior trained under sparse supervision can achieve performance comparable to a model trained on fully observed data. A positive result would indicating that high-quality priors can be learned without full supervision. **RQ2** then investigates whether the proposed sampling strategy, including RePaint-style resampling and physics-guided projection, can effectively compensate for prior limitations by generating physically consistent, high-fidelity traffic state reconstructions under sparse and heterogeneous observations.

B.1 Dataset and Pre-processing

The study corridor corresponds to the Tennessee Department of Transportation's I-24 Mobility Technology Interstate Observation Network (I-24 MOTION) data [11]. I-24 MOTION data captured by 276 pole-mounted high-resolution traffic cameras that provide seamless coverage of approximately 4.2 miles alone the I-24 highway corridor near Nashville, TN. In our data, trajectories are rasterized into speed fields on a fixed Eulerian grid of $S = 64$ cells ($\Delta x = 200$ft), four lanes, and $T = 64$ time slices ($\Delta t = 5$ s). Figure 6 shows random samples from the dataset.

Although I-24 MOTION is camera-based rather than loop-detector-based, it offers two decisive advantages for this study. First, its near-complete spatial and temporal coverage yields a "ground-truth" speed field, which allows us to mask the data synthetically and evaluate reconstruction error at every hidden cell–something impossible with legacy loop-detector archives, where the true state is itself unobserved. Second, the trajectory data include natural traffic variability, lane changes, and congestion waves that would be under sampled by typical loop-detector and probe-vehicle networks. By repeatedly sub-sampling I-24 MOTION to match realistic detector layouts and probe trajectories, we can test PMA-Diffusion under precisely controlled sparsity levels while still benchmarking against a fully observed baseline.

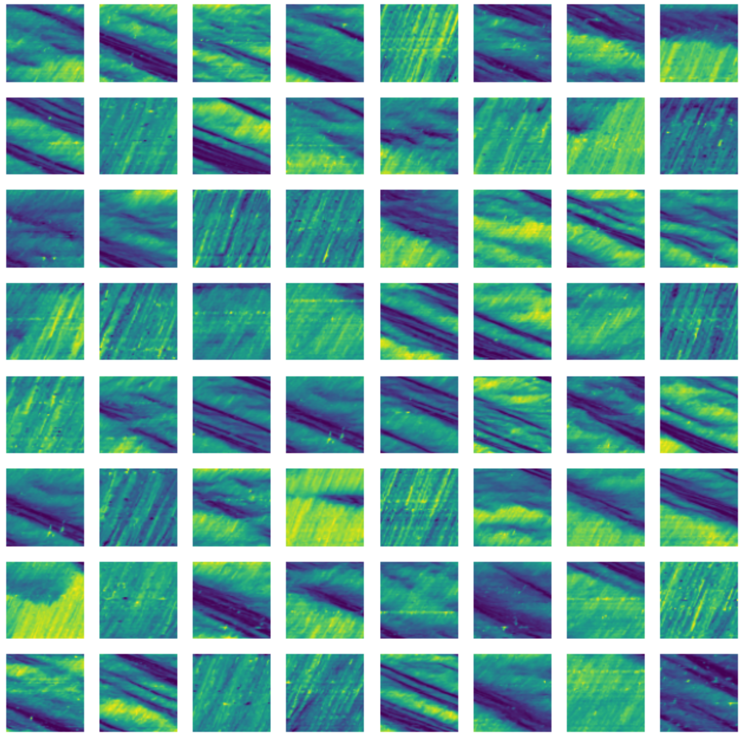


Figure 6: Random speed field samples from I-24 MOTION dataset.

B.1.1 Observation Models

We collect synthetic observation ($\mathbf{Y}$) and mask ($\mathbf{M}$ pairs based on the ground-truth speed field ($\mathbf{V}$) to simulate the real-world scenarios. Our masks have two components:

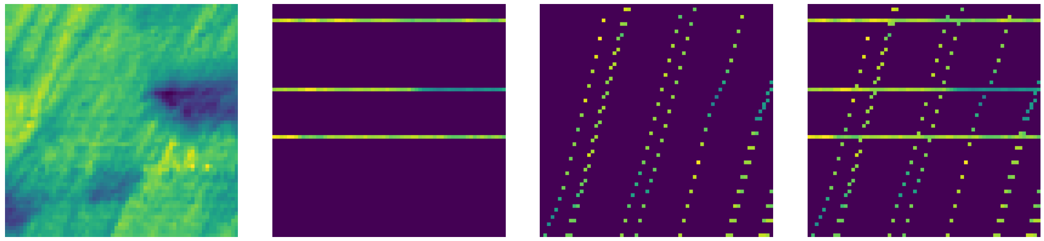
- **Loop detectors mask:** In practice, the Loop detectors are fixed in a specific location. We generate a fixed Loop detector mask $\mathbf{M}^{(l)}$ and mask $5\%, 15\%, 25\%, 50\%$ of the row data as observed data $\mathbf{M}^{(l)} \odot \mathbf{V}$.
- **Probe vehicles mask:** To simulate probe vehicles data, we randomly draw the number of probe vehicles for the current speed slice from a Poisson distribution $N_{\text{seed}} \sim \text{Poisson}(\lambda_{\text{seed}})$. Each vehicle receives a random entry time $t_0^{(j)} \in \{0, \dots, 63\}$, so the set of starting points is $(t_0^{(j)},\, s_0^{(j)})$, where $j = 1, \dots, N_{\text{seed}}$. Based on the starting point and the speed information, we can generate the probe vehicles' trajectory. In our experiments, we tried $\lambda = 0, 5, 15, 25$ respectively.

Building on Assumption 1 and Assumption 2 in methodology, we resample the probe vehicle trajectory directly on the Eulerian grid. The resulting synthetic detectors, therefore, deliver observations that are perfectly aligned with the computational lattice, which removes the interpolation artifacts and timing offsets that would otherwise arise when irregular GPS pings are mapped onto a coarser grid. Figure 7 shows the original speed field and three derived observation masks: (b) induced by sparse loop detectors, (c) from resampled probe vehicle trajectories, and (d) from their union.

B.2 Evaluation Metrics

The reconstruction performance is measured only on the unobserved pixels using three complementary criteria following [36]:

1. **Masked-MSE (2×2):** the mean squared error computed after 2×2 average pooling step, capturing low-frequency fidelity.



(a) Ground-truth data (b) Loop detector (c) Probe vehicle (d) Total observed

Figure 7: Observation Patterns from Loop detectors and Probe vehicles.

2. **Sobel-MSE:** the MSE between Sobel edge maps of the estimate and ground truth, emphasizing high-frequency structure.
3. **LPIPS:** the Learned Perceptual Image Patch Similarity score (AlexNet backbone), accounting for perceptual closeness [34].

Together, these metrics provide a balanced view of numerical accuracy, edge sharpness, and visual realism on the unobserved regions.

B.3 Models

The three training and sampling schemes introduced below are structured specifically to isolate the contributions of the prior and the sampling mechanism with respect to RQ1 and RQ2, respectively.

B.3.1 Pretraining Scheme

We evaluate three training schemes designed to isolate the effects of supervision.

- **Full-obs** The network trained by fully latent speed field data $\mathbf{V}$; the value of **Full-obs** represents the upper bound, showing how well the DDPM architecture alone can fit fully observed data.
- **Single-mask** Each training sample is corrupted by its own visibility mask, exactly mirroring realistic loop detectors + probe vehicles coverage. The model therefore learns only from genuinely observed pixels.
- **Double-mask** On top of the true visibility map we apply an additional mask (see Table 1). Hiding even "always visible" rows encourages the network to reason about persistently unseen locations and improves robustness when some cells are never measured in practice.

All three variants share the same UNet backbone, which is the most widely used backbone neural network structure for diffusion models [37, 26]. Specifically, our UNet backbone is configured with one input channel, 64 base feature channels, three resolution levels with channel multipliers of $(1, 2, 4)$ and 4-head linear attention in the bottleneck. We use a linear β-schedule with T=500 steps ($\beta_0 = 10^{-4}$, $\beta_T = 2{\times}10^{-2}$) and the Huber loss ($\delta = 1$), which is numerically more stable than ℓ_2. We train for ten epochs (roughly 1.8×10^3 iterations) using Adam (lr $= 5{\times}10^{-4}$, batch $= 250$). The model training time is around 15 minutes of GPU time.

B.3.2 Sampling Scheme

Using a prior from Pretraining stage, we compare three inference pipelines:

1. **Ambient one-shot.** a single unconditional reverse trajectory with real time data consistency.
2. **RePaint [30].** It iteratively alternates $j = 1$ forward and one reverse step, repeating $r = 250$ times every 100 iterations ($\eta = 0.9$, keep some randomness) to better mix the conditional and unconditional paths.
3. **RePaint + physics projector (AAS).** We embed the AAS projector into every reverse step, which nudges missing pixels toward first-order traffic consistency while leaving observations unchanged.

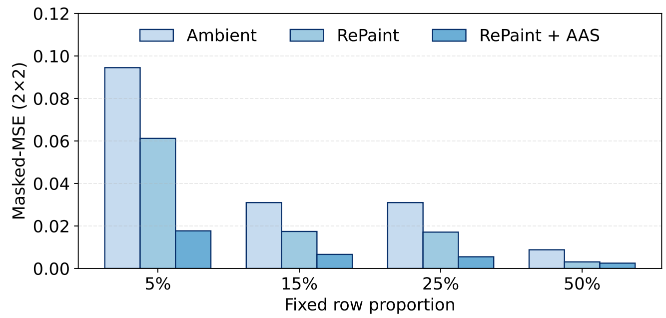


Figure 8: Masked-MSE across samplers based on Full-obs model ($\lambda = 0$).

The β-schedule and timestep (T=500) are shared with Pretraining stage. RePaint uses the same U-Net to predict ϵ_θ; the physics projector is *parameter-free* at test time once the hyper-parameters are fixed.

B.4 Experimental Results

Tables 2 reports LPIPS, Masked-MSE, and Sobel-MSE for every combination of training mode, observation density (row-mask: $5\% \rightarrow 50\%$), probe intensity ($\lambda: \ 0 \rightarrow 25$), and sampler (Ambient, RePaint, RePaint + AAS). To interpret these results correctly, we clarify how visibility is defined across different models. In the **Full-obs model**, the network is trained on fully observed (ground-truth) speed fields, and the reported row visibility and λ values apply only during the *sampling stage*. In contrast, for the **Single-mask** and **Double-mask models**, both training and sampling are conducted under the same partial observation regime; thus, the specified row and λ values reflect the *entire modeling pipeline*. This distinction is essential when comparing performance across models, as it represents the impact of training-time supervision quality versus inference-time robustness to observation sparsity.

B.4.1 Full-obs model

The Full-obs model sets the upper bound. With only 5% fixed rows and no probe vehicles, it already reconstructs the unobserved 95% of the grid with a masked MSE of 0.0945 (Ambient) and pushes the error below 0.0612 after a single RePaint pass (Figure 8). Adding the physics projector reduces the residual again by roughly a factor of three, indicating that, even when a perfect prior is available, posterior resampling and physical nudging remain beneficial.

B.4.2 Single-mask model

In contrast, the Single-mask prior suffers whenever parts of the grid are not visible during training. Under the same 5% row available and $\lambda = 0$ setting, the raw Ambient reconstruction yields a MSE above 27 and an LPIPS of 1.004; RePaint without physics can even amplify this instability because large swaths of the field are extrapolated from virtually no context. Once AAS is inserted, however, the error decreases to only 0.035 worse than the Full-obs and significantly better than the unprojected RePaint result (Figure 2). The takeaway is that an expressive but incompletely trained prior is salvageable if the sampler is endowed with a lightweight physics corrector.

B.4.3 Double-mask model

In Figure 2, the Double-mask model closes most of the gap while keeping the training set identical to Single-mask model. Even the worst case (5% rows, λ = 0) is still an order of magnitude better than its Single-mask counterpart and only 0.0197 worse than the fully supervised Full-obs model. The improvement is consistent in perceptual space and high-frequency structure as well. Interestingly, in the case of Row=5%, $\lambda = 25$, the Masked-MSE of Single-mask model using repaint+AAS is even outperforms Double-mask model, suggesting that if all pixels have non-zero probability of being observed in historic data, Single-mask model can be a good training scheme.

B.4.4 Sampler Effect and Physics Guidance

Across all models and visibility settings, we observe three consistent trends regarding the impact of posterior sampling. First, iterative refinement via RePaint consistently improves over the one-shot Ambient pass, especially when the underlying prior is expressive, namely, in the Full-obs and Double-mask settings. Second, physics guidance through the AAS projector never degrades performance; in all 48 test cases (Table 2), it either improves or preserves LPIPS, Masked-MSE, and Sobel-MSE scores. Third, the benefit of physics-guided sampling is amplified under sparse observations: while RePaint and AAS offer marginal improvements under 50% row visibility, their impact is substantial when only 5% of rows are visible, particularly in sharpening traffic fronts and correcting structural artifacts.

The edge-based Sobel metric further highlights the effect of physics guidance on fine-scale reconstruction. Even when Masked-MSE is already low, AAS consistently reduces Sobel-MSE in the Single- and Double-mask models, indicating that it enhances stop-and-go boundaries rather than applying global smoothing. However, an exception arises when the prior is trained on fully observed data: in this regime, RePaint and AAS may slightly blur high-frequency features, yielding higher Sobel-MSE than the Ambient baseline. This suggests that when the prior is already close to optimal, further physics corrections may inadvertently attenuate sharp details.

B.4.5 Research Questions Revisited

RQ1: How much does prior quality matter? Our results indicate that the learned prior's quality is strongly influenced by the training supervision regime. In the Single-mask model, predictions from Ambient frequently collapse in unobserved regions, and RePaint may over-extrapolate. These pathologies are largely mitigated under the Double-mask model, which introduces randomized dropout to increase spatial generalization. Hence, *prior quality is highly sensitive to the visibility pattern at training time*. A partially supervised prior can match a fully supervised one only if (i) every region is observed at least occasionally, or (ii) aggressive random masking is introduced during training to improve robustness.

RQ2: How much does posterior sampling help? Posterior sampling mechanisms are critical across all settings. RePaint yields consistent improvements over Ambient one-shot sampling, validating the value of iterative resampling. The AAS projector further enhances performance or leaves it unchanged across all cases. Notably, the benefit is most pronounced under sparse observations (e.g., 5% row visibility), where physics-guided sampling not only improves reconstruction metrics but also sharpens spatiotemporal structures. However, in the fully supervised setting, these enhancements may introduce over-smoothing, slightly degrading edge fidelity. Thus, while posterior sampling is generally beneficial, its effects are modulated by the supervision regime and observation density.

B.4.6 Qualitative Illustration

Figure 3 compares the RePaint + AAS reconstruction against the ground truth under the Double-mask model with 25% row visibility and probe intensity $\lambda = 15$. While the generated speed field appears smoother and lacks certain fine-grained variations present in the ground truth, its overall spatiotemporal structure remains consistent and physically coherent. This suggests that the model effectively captures macroscopic traffic patterns, even if some high-frequency details are attenuated during sampling.

B.4.7 Quantitative results

Table 2: Quantitative results for three models (lower is better).

Mode	Row	λ	LPIPS			Masked-MSE (2×2)			Sobel-MSE		
			Ambient	RePaint	+AAS	Ambient	RePaint	+AAS	Ambient	RePaint	+AAS
Full-obs	5%	0	0.1905	0.1497	0.3139	0.0945	0.0612	0.0177	0.0017	0.0017	0.0022
		5	0.1346	0.0970	0.1811	0.0402	0.0198	0.0105	0.0014	0.0014	0.0019
		15	0.1185	0.1212	0.1295	0.0471	0.0358	0.0085	0.0014	0.0017	0.0018
		25	0.1067	0.1386	0.1026	0.0968	0.0573	0.0087	0.0020	0.0022	0.0017
	15%	0	0.0622	0.0508	0.0861	0.0310	0.0174	0.0066	0.0014	0.0013	0.0016
		5	0.3540	0.2115	0.0659	0.2580	0.0913	0.0058	0.0689	0.0136	0.0016
		15	0.0843	0.0480	0.0514	0.0492	0.0148	0.0053	0.0015	0.0013	0.0014
		25	0.1053	0.0554	0.0451	0.0686	0.0200	0.0053	0.0016	0.0013	0.0014
	25%	0	0.0804	0.0628	0.1089	0.0310	0.0171	0.0055	0.0014	0.0014	0.0017
		5	0.1115	0.0833	0.0752	0.0501	0.0217	0.0043	0.0014	0.0016	0.0015
		15	0.1294	0.0625	0.0563	0.0666	0.0128	0.0035	0.0017	0.0013	0.0014
		25	0.1732	0.1481	0.0495	0.2213	0.0668	0.0035	0.0185	0.0042	0.0013
	50%	0	0.0112	0.0053	0.0173	0.0088	0.0031	0.0025	0.0006	0.0005	0.0007
		5	0.0150	0.0064	0.0128	0.0126	0.0040	0.0024	0.0006	0.0005	0.0006
		15	0.0103	0.0054	0.0101	0.0071	0.0026	0.0024	0.0005	0.0004	0.0006
		25	0.0131	0.0047	0.0090	0.0136	0.0029	0.0024	0.0007	0.0004	0.0005
Single-mask	5%	0	1.0040	0.8248	0.3477	27.6183	86.3606	0.0527	6.3083	4.8671	0.00190
		5	0.5620	0.3803	0.1798	0.2823	0.0691	0.0114	1.0646	0.1182	0.00190
		15	0.5705	0.3487	0.1225	0.3102	0.1562	0.00839	1.0823	0.1709	0.00180
		25	0.5146	0.4751	0.1012	1.4005	1.2025	0.00760	2.3589	1.4004	0.00170
	15%	0	0.5919	0.4802	0.0864	0.8678	0.3315	0.00673	3.4730	1.1081	0.00162
		5	1.1527	1.1390	0.0638	64.6652	45.1352	0.00599	179.877	126.190	0.00154
		15	0.4879	0.3356	0.0512	0.7128	0.2561	0.00532	1.6540	0.3211	0.00143
		25	0.4955	0.2698	0.0442	0.6365	0.1263	0.00520	1.6740	0.1745	0.00142
	25%	0	0.5709	0.4666	0.1096	0.8315	0.1997	0.00549	1.9544	0.4157	0.00170
		5	0.5207	0.3961	0.0762	1.4663	0.8395	0.00450	1.3942	0.4572	0.00150
		15	0.9972	0.9915	0.0569	29.1341	12.2471	0.00398	83.1233	35.2213	0.00142
		25	0.9276	0.9189	0.0506	68.9262	49.7180	0.00360	165.865	115.351	0.00130
	50%	0	0.4755	0.2064	0.0173	0.8594	0.2399	0.00250	1.7465	0.2187	0.00070
		5	0.4544	0.1368	0.0125	0.8584	0.2105	0.00250	1.5391	0.1293	0.00060
		15	0.4031	0.0696	0.0101	0.8517	0.1146	0.00244	1.1634	0.0564	0.00056
		25	0.2427	0.0048	0.0094	0.1752	0.00670	0.00246	0.3010	0.00350	0.00050
Double-mask	5%	0	0.8560	0.7641	0.3424	9.6076	12.5773	0.0374	1.5506	0.8953	0.00190
		5	0.4168	0.1902	0.1900	0.1404	0.0820	0.0116	0.1927	0.00610	0.00190
		15	0.5850	0.4850	0.1229	0.5325	0.1693	0.00748	1.6893	0.4743	0.00180
		25	0.5107	0.2648	0.1022	0.3741	0.0556	0.00790	1.2477	0.09630	0.00170
	15%	0	0.5356	0.4172	0.0856	1.2058	0.7597	0.00680	1.9782	0.6346	0.00160
		5	0.5329	0.4692	0.0654	0.9536	0.3093	0.00590	3.0547	0.9078	0.00160
		15	0.5355	0.4412	0.0508	1.0276	0.3548	0.00540	2.6717	0.6638	0.00141
		25	0.4691	0.1911	0.0441	0.3624	0.0424	0.00540	1.0735	0.0758	0.00140
	25%	0	0.5396	0.4025	0.1092	0.5771	0.2698	0.00560	0.9457	0.1379	0.00170
		5	0.5672	0.4994	0.0783	0.8120	0.2590	0.00430	2.2709	0.6650	0.00150
		15	0.5850	0.4850	0.0577	1.0858	0.2954	0.00361	2.0320	0.4874	0.00138
		25	0.5477	0.4498	0.0473	0.9231	0.1944	0.00340	2.2537	0.4694	0.00130
	50%	0	0.2285	0.00970	0.0173	0.0617	0.00420	0.00250	0.1453	0.00370	0.00070
		5	0.1817	0.00810	0.0126	0.0415	0.00900	0.00250	0.0999	0.00300	0.00060
		15	0.4013	0.0230	0.00990	0.4430	0.01550	0.00243	0.8887	0.01490	0.00056
		25	0.4097	0.0395	0.00920	0.6907	0.1017	0.00250	1.0049	0.04470	0.00050

NeurIPS 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 introduction state the three core contributions—(i) a mask-aware diffusion prior trained on incomplete observations, (ii) a RePaint-style posterior sampling loop with exact observation replacement, and (iii) a physics-guided projector—together with assumptions (MAR, bounded drift) and summarized results on I-24 MOTION (e.g., two orders of magnitude reductions at 5% coverage). These match the methods and findings reported in Sections 1, 2.1, 2.2, 2.3, and 3.

   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: The Conclusion explicitly lists limitations (time alignment between probes and loops, small/smooth bias assumption, speed-only surrogate physics) and proposes remedies (joint latency estimation, non-Gaussian/rough bias modeling, density-aware projectors, streaming deployment and transfer) in Section 4.

   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: [Yes]

   Justification: The paper provides a detailed assumptions and proofs in the appendix.

   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 3 specifies dataset construction, observation models, metrics, UNet architecture, loss, optimizer, learning rate, batch size, β-schedule and number of steps (T=500), RePaint hyperparameters (j=1, r=250, η=0.9), and projector details, sufficient to reproduce the reported numbers.

   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 NeurIPS 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: The paper describes all settings needed for reproduction but does not presently include links to code.

   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: Model, losses, optimizer and hyperparameters, training budget, noise schedule, sampling procedures, and evaluation metrics are explicitly given in Sections 3 and appendix; the evaluation protocol is defined via synthetic masking of fully observed frames from I-24 MOTION.

   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: We report point estimates for each configuration without error bars or confidence intervals; multiple seeds will be added in an updated appendix due to compute constraints.

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: We report approximate training time (e.g., $\sim$15 minutes of GPU time, using one RTX 6000 ADA generation).

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://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: No human subjects or personally identifiable information are used; evaluations rely on aggregated grid-based speed fields derived from a publicly documented DOT-operated dataset; anonymity is preserved for submission.

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: [No]

    Justification: The main text focuses on methods and evaluation; we will add a short discussion noting potential benefits (safer, more efficient traffic operations) and risks (privacy in sensing, misuse for surveillance) with mitigations.

    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: [NA]

    Justification: The work does not release high-risk generative models or scraped datasets; outputs are reconstructed speed fields for benchmarks.

    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 dataset (I-24 MOTION) is cited.

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: [NA]

    Justification: We do not introduce a new dataset or benchmark in this submission.

    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: [NA]

    Justification: The work does not involve crowdsourcing or human-subject studies.

    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: [NA]

Justification: Not applicable; no human-subject research was performed.

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.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs are not part of the core methodology or experimental pipeline; any writing assistance does not affect scientific content.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.