

A Physics-informed Diffusion Model for HF Flow Field Reconstruction

ANDRÉS FORERO, Technical University of Munich

This report examines and critically assesses the diffusion-based computational fluid dynamics (CFD) super-resolution framework of [Shu et al. 2023], which introduces a physics-informed Denoising Diffusion Probabilistic Model (DDPM) [Ho et al. 2020] for reconstructing high-fidelity (HF) flow fields from low-resolution or sparse inputs. The core idea of the method is to train the diffusion model exclusively on HF data and to treat reconstruction as a conditional denoising problem: at inference time, low-fidelity (LF) observations are first projected onto the HF grid and mixed with Gaussian noise so that they match the distribution of noisy training samples, after which the reverse diffusion process produces a HF field. This noise-induced alignment decouples the reconstruction model from the specific way the LF data are obtained, and is the main mechanism behind its robustness to different resolutions, filters, and sparsity levels. [Shu et al. 2023] additionally propose two physics-informed guidance strategies based on the Navier–Stokes residual, which act as optional corrections during sampling.

Building on this foundation, the present study provides a detailed analysis of the diffusion formulation and its physics-guided extensions, and implements a minimal DDPM on synthetic two-dimensional CFD-style fields to probe the training dynamics in practice. The experiments confirm that the ε -prediction mechanism of the DDPM behaves as expected under the standard denoising objective, but also show that naive additions of physics-inspired penalty terms can destabilise optimisation and do not automatically yield more realistic samples. This observation supports the view that the main strength of the framework lies in the HF-only diffusion formulation and the associated noise alignment, whereas the current physics-guidance mechanisms are effective only under idealised assumptions (periodic domains, full DNS trajectories, exact PDE knowledge). The report identifies key limitations in terms of computational cost, scalability beyond two-dimensional periodic turbulence, and the lack of explicit uncertainty quantification, and outlines future research directions including solver-consistent self-guided diffusion, latent-space and transformer-based architectures for three-dimensional flows, and more principled integration of uncertainty estimation in diffusion-based CFD surrogates.

Additional Key Words and Phrases: Computational fluid dynamics, diffusion models, physics-informed learning, turbulence reconstruction, uncertainty quantification

1 SUMMARY OF PAPER

1.1 Problem Context

HF CFD simulations provide accurate representations of turbulent flows but remain computationally prohibitive at high Reynolds numbers due to the cost of resolving all spatial and temporal scales. Classical alternatives such as Reynolds Averaged Navier-Stokes (RANS) [Launder and Sharma 1974], Large-eddy simulation (LES) [Pitsch 2006], and hybrid RANS–LES [Shur et al. 2008] reduce this cost by introducing modeling assumptions, but at the expense of losing fine-scale structure. In recent years, machine learning models particularly CNN- and GAN-based super-resolution networks have

been applied to reconstruct high-resolution flow fields from under-resolved numerical or experimental inputs. While these architectures perform well under fixed training conditions, they generalize poorly: each model is tailored to a specific LF dataset and tends to fail when the degradation process differs from the one seen during training.

This motivates the need for a reconstruction method that is not tied to a particular downsampling, filtering, or sparsification operator and that can generalize across LF input distributions without retraining.

1.2 Methodology

Shu et al. [Shu et al. 2023] addresses this problem by reformulating CFD super-resolution as a *conditional denoising* task solved with a DDPM. Unlike conventional approaches that learn a direct mapping from LF to HF data, their framework trains exclusively on HF flow fields. The model learns to invert a Gaussian corruption process: during training, clean HF samples are progressively noised; during inference, the LF observation is embedded into this same noisy HF manifold and the reverse diffusion process denoises it back into a HF field.

Concretely, any LF input $\mathbf{x}^{(g)}$ is first projected to the HF grid and then mixed with Gaussian noise at a chosen diffusion time-step,

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} g(\mathbf{x}^{(g)}) + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon},$$

so that its distribution matches that of noisy HF samples seen during training. This noise-induced alignment decouples the reconstruction model from the specifics of the LF generation process and is the main driver of the strong generalization reported in the paper.

1.2.1 Problem formulation as conditional denoising. Let $\mathbf{x}_0 \in \mathcal{Y}$ denote a HF vorticity field from the Kolmogorov flow dataset [Shu et al. 2023], and let $\mathbf{x}^{(g)} \in \mathcal{X}$ denote a LF observation (e.g. uniformly downsampled, filtered, or sparse measurements on the grid). Classical super-resolution methods explicitly learn a mapping $f_\phi : \mathcal{X} \rightarrow \mathcal{Y}$ such that $f_\phi(\mathbf{x}^{(g)}) \approx \mathbf{x}_0$. This coupling to a specific \mathcal{X} makes them sensitive to distribution shifts in the LF data.

Instead, the diffusion-based framework keeps the *model* purely in the HF space and uses the LF input only as a *guide* for the denoising process. Formally, the DDPM learns a family of conditional distributions $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ on \mathcal{Y} , while the LF input $\mathbf{x}^{(g)}$ is injected via the initialization of the reverse process at a chosen time-step t .

1.2.2 Forward diffusion process on HF data. During training, only HF samples $\mathbf{x}_0 \sim p_{\text{data}}(\mathcal{Y})$ are used. A forward (diffusion) process gradually corrupts \mathbf{x}_0 into Gaussian noise through a Markov chain

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

where $\{\beta_t\}_{t=1}^T$ is a predefined variance schedule and $\alpha_t := 1 - \beta_t$. This chain has the closed-form expression

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}), \quad \bar{\alpha}_t := \prod_{i=1}^t \alpha_i. \quad (2)$$

Intuitively, \mathbf{x}_t is a *controlled mixture* of signal and noise:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (3)$$

For small t , \mathbf{x}_t is close to the original flow field, whereas for large t it approaches pure Gaussian noise.

1.2.3 Reverse process and DDPM training objective. The generative model is defined by a reverse Markov chain

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t), \quad p(\mathbf{x}_T) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (4)$$

with Gaussian transitions

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \Sigma_t \mathbf{I}). \quad (5)$$

Following Ho et al. [Ho et al. 2020], the mean $\boldsymbol{\mu}_\theta$ is parameterized in terms of a learned *noise predictor* $\boldsymbol{\epsilon}_\theta$:

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right). \quad (6)$$

The network $\boldsymbol{\epsilon}_\theta$ is implemented as a U-Net [Ronneberger et al. 2015], which operates on vorticity fields and is conditioned on the discrete time-step t .

The model is trained by minimizing the simplified denoising objective

$$\mathcal{L}_{\text{simple}}(\theta) = \mathbb{E}_{\mathbf{x}_0, t, \boldsymbol{\epsilon}} \left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t) \right\|_2^2 \right], \quad (7)$$

where $\mathbf{x}_0 \sim p_{\text{data}}(\mathcal{Y})$, $t \sim \mathcal{U}\{1, \dots, T\}$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Thus, the DDPM learns to *invert* the forward diffusion process by predicting the noise that was added to a HF flow field at any time-step t .

1.2.4 Conditional sampling with LF guidance. At test time, the goal is to reconstruct a HF field from a LF observation $\mathbf{x}^{(g)} \in \mathcal{X}$, without retraining the model for each new type of LF data. Instead of starting the reverse chain from pure Gaussian noise $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, Shu et al. [Shu et al. 2023] perform *partial backward diffusion* starting from an intermediate time-step t initialized by a noisy version of the guidance input:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} g(\mathbf{x}^{(g)}) + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (8)$$

where $g(\cdot)$ denotes a preprocessing operator that lifts the LF observation to the HF grid (e.g. interpolation or nearest-neighbor padding for sparse data). Starting from \mathbf{x}_t , a DDPM/DDIM-style reverse process (e.g. the implicit sampler of Song et al. [Song et al. 2020]) is run from t down to 0:

$$\mathbf{x}_{t-1} \sim p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t), \quad \dots, \quad \mathbf{x}_0 \sim p_\theta(\mathbf{x}_0 | \mathbf{x}_1). \quad (9)$$

The final state \mathbf{x}_0 is taken as the reconstructed HF flow field.

1.2.5 Noise-induced distribution alignment and generalization. The crucial effect of the initialization in (8) is that it maps diverse LF inputs into the *same noisy HF manifold* that the DDPM was trained on. During training, all inputs to $\boldsymbol{\epsilon}_\theta$ have the form

$$\mathbf{x}_t^{\text{train}} = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, \quad \mathbf{x}_0 \sim p_{\text{data}}(\mathcal{Y}).$$

At inference, after preprocessing and noise injection, the guided initialization $\mathbf{x}_t^{\text{test}}$ in (8) has the *same statistical form*, regardless of how $\mathbf{x}^{(g)}$ was generated (different filters, resolutions, or sparsity patterns). By choosing t appropriately, the model ensures that:

- the signal component $\sqrt{\bar{\alpha}_t} g(\mathbf{x}^{(g)})$ carries the large-scale structure of the LF observation, while
- the noise component $\sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_t$ pushes the sample into the region of the space where the DDPM has learned a reliable denoising mapping.

In other words, the model is **not** trained to map LF \rightarrow HF. It is trained to map *noisy HF-like inputs* to clean HF fields. By first transforming the LF input into such a noisy sample, the reverse diffusion process can be reused for a wide range of LF data distributions without retraining. This noise-induced alignment is the main mechanism behind the improved robustness and generalization reported in [Shu et al. 2023].

1.2.6 Optional physics-informed guidance. On top of the purely data-driven DDPM reconstruction, Shu et al. [Shu et al. 2023] propose to incorporate information from the governing Navier–Stokes equation in vorticity form,

$$G(\omega) := \partial_t \omega + \mathbf{u} \cdot \nabla \omega - \frac{1}{\text{Re}} \nabla^2 \omega - f(\mathbf{x}) = 0, \quad (10)$$

by using the *PDE residual* as a guidance signal during sampling. Given a predicted vorticity field ω (or, in practice, a triplet of frames $[\omega_{t-1}, \omega_t, \omega_{t+1}]$), the residual $r := G(\omega)$ is computed numerically and its gradient with respect to the field,

$$\mathbf{c} := \frac{\partial r}{\partial \omega}, \quad (11)$$

is used as a physics-informed conditioning variable.

The first variant, referred to as *learned residual guidance*, augments the DDPM network input with \mathbf{c} and trains the U-Net to predict the noise $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{c})$. A classifier-free guidance scheme [Ho and Salimans 2022] is then used at inference time to combine the conditional and unconditional predictions,

$$\tilde{\boldsymbol{\epsilon}}_\theta = \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{c}) + w [\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \mathbf{c}) - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t, \emptyset)], \quad (12)$$

where w is a guidance strength. Intuitively, the score estimate is nudged towards states that both look like realistic turbulence and yield a smaller PDE residual.

The second variant, *direct residual descent*, keeps the original DDPM architecture unchanged and instead injects the residual gradient directly into the Denoising Diffusion Implicit Model (DDIM) update developed by [Song et al. 2020]. In this case, each reverse step is modified as

$$\mathbf{x}_{t-1} = \text{DDIMStep}(\mathbf{x}_t, \tilde{\boldsymbol{\epsilon}}_\theta) - \lambda \mathbf{c}, \quad (13)$$

with a step size $\lambda > 0$. Here the role of \mathbf{c} is analogous to taking a small gradient-descent step to reduce the PDE residual at every denoising iteration. Both variants leave the *core* reconstruction mechanism unchanged: the DDPM still denoises noisy, HF-like inputs into HF flow fields, and the physics-based terms act only as additional guidance during sampling.

Through this framework, [Shu et al. 2023] demonstrate that their diffusion-based approach can reconstruct HF flow fields from both low-resolution and sparsely sampled inputs with strong accuracy and generalization. By combining probabilistic denoising with physics-informed conditioning, the model bridges the gap between data-driven learning and physical laws, providing a scalable and robust solution for CFD super-resolution.

2 EXPERIMENTAL SETUP

To complement the theoretical discussion of diffusion-based reconstruction, I implemented a minimal DDPM in PyTorch targeting small two-dimensional CFD-style fields. The goal of this experiment was not to reproduce the full physics-informed framework, but to validate the core denoising mechanism on a lightweight synthetic dataset and to inspect the training and sampling behaviour in practice.

Data and preprocessing. All data are stored as NumPy arrays with shape $[N, C, H, W]$, where $C = 3$ and $H = W = 64$ by default. The channels can be interpreted as three scalar fields on the same grid (e.g. vorticity snapshots at neighbouring times or different flow components), but in this implementation they simply serve as generic “CFD-style” fields rather than true DNS solutions. Synthetic datasets are generated by a generic synthetic script, which writes both the main array and an associated per-sample conditioning array with three scalar parameters, intended to represent, for example, forcing amplitude, dissipation scale and Reynolds number.

For future work, the U-Net is configured to accept concatenated data and conditioning channels, with a default of `in_channels = 6` and `out_channels = 3`. In the experiments reported here, however, the conditioning is not used: only the three data channels are loaded during training, and the conditioning array is ignored.

Model architecture. The score network is a compact U-Net. The design follows a standard encoder–decoder structure with skip connections, but is deliberately small (depth = 3, base number of feature channels = 32) to keep training feasible on a single GPU within a few days.

To mimic the periodic boundary conditions used in Kolmogorov-flow benchmarks, every Conv2d layer is instantiated with circular padding, so that the network is equivariant to toroidal shifts instead of imposing artificial zero-padding at the domain boundaries. The U-Net also includes sinusoidal time-step embeddings that are passed through a small MLP and used as FiLM-style modulation parameters inside intermediate blocks.

Conceptually, this would allow the model to condition on the diffusion time-step t and, in a later extension, on external parameters such as the synthetic flow settings. In the current training script, however, the DDPM is used in its simplest form: the training loop minimises the standard noise-prediction loss and, in the version used for this report, does not yet feed either the time-step or the conditioning vector into the network. Practically, the model learns a mapping $\epsilon_\theta : \mathbf{x}_t \mapsto \hat{\epsilon}$ that predicts the added noise from the noisy input alone.

Diffusion process and training. A linear noise schedule $\{\beta_t\}_{t=1}^T$ is used, constructed as

$$\beta_t \in \text{linspace}(10^{-4}, 2 \cdot 10^{-2}, T),$$

with corresponding $\alpha_t = 1 - \beta_t$ and cumulative products $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ cached for efficiency. The model is trained with the usual DDPM objective of Ho et al. [Ho et al. 2020], i.e. predicting the Gaussian noise ϵ added in the forward process at a randomly sampled time-step t .

All hyperparameters are defined in a configuration file, including the learning rate, batch size, image size, number of channels, base channels, U-Net depth, the number of diffusion steps T (default $T = 50$), total epochs and checkpoint frequency. Checkpoints store both the model and optimiser state, making it possible to resume training or run sampling experiments from intermediate models. The implementation assumes powers-of-two image sizes to match the encoder/decoder strides; minor mismatches are handled by interpolation in the decoder.

An optional utility is provided to downsample external fields to the target resolution and stack them into the expected $[N, C, H, W]$ format, but the experiments in this report rely solely on the synthetic generator.

Sampling procedure. Sampling is performed via the ancestral DDPM reverse process. Given a desired output shape (e.g. `[batch_size, 3, 64, 64]`), the sampler initialises \mathbf{x}_T as standard Gaussian noise and iteratively applies the learned reverse transitions $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ down to $t = 0$.

No LF input, PDE residual or external conditioning is provided to the sampler. Instead, the experiment focuses on the ability of the trained DDPM to transform pure Gaussian noise into coherent CFD-style patterns that qualitatively resemble the synthetic training set.

Evaluation. Given the pedagogical scope and the synthetic nature of the data, evaluation is primarily qualitative. During training, I monitored the denoising loss as a proxy for how well the model learns the forward process. After training, I generated batches of samples from the latest checkpoint and compared them visually against training examples. The comparisons focus on the presence of large-scale coherent structures and the diversity of generated fields, rather than on strict quantitative measures such as RMSE or energy spectra.

Overall, this minimal implementation serves as a controlled environment to test and better understand the core DDPM mechanism discussed in Section 1.2, before introducing additional complexity such as explicit LF conditioning or physics-informed guidance.

3 RESULTS

3.1 Training dynamics

The loss curve for training and validation in Appendix A.1 shows the evolution of the training and validation losses over 20 epochs for the minimal DDPM implementation. The loss plotted on the validation set is the standard denoising objective used in DDPMs, i.e. the mean squared error between the predicted noise and the Gaussian noise that was added in the forward process. On the training set, however, I experimented with augmenting this loss by an additional physics-inspired penalty term (e.g. based on discrete differential operators), applied only during training and not in validation.

The resulting curves exhibit an unusual pattern. The validation loss (orange) decreases smoothly and monotonically from approximately 0.35 to 0.21, indicating that the model does learn to denoise the synthetic fields in the sense of the standard DDPM objective. In contrast, the training loss (blue) initially decreases during the first

few epochs, but then exhibits a sharp spike and remains noticeably higher than the validation loss, with strong fluctuations.

This behaviour is not consistent with classical overfitting (where one would expect the training loss to continue decreasing while the validation loss stagnates or increases). Instead, it strongly suggests an *objective mismatch* between training and validation: once the additional physics term is activated, the reported training loss no longer measures the same quantity as the validation loss. In other words, the DDPM part of the model appears to improve steadily, but the extra penalty term dominates and destabilises the training objective.

Several practical factors likely contribute to this instability:

- The physics term is not carefully scaled relative to the ϵ -MSE, so its magnitude can be orders of magnitude larger and overwhelm the denoising loss.
- The discrete differential operators used in the penalty (e.g. gradients or Laplacians) are sensitive to boundary conditions, aliasing and grid spacing; if their implementation does not match the data generation process, the penalty effectively introduces structured label noise.
- The learning rate and optimisation settings were tuned for the pure DDPM loss. Adding a stiff additional term without re-tuning can easily produce the kind of noisy, non-monotonic training curve observed here.

From this experiment, the main takeaway is that the *core* DDPM denoising mechanism can be trained stably on synthetic CFD-style fields (as indicated by the steadily decreasing validation loss), but naively adding physics-inspired penalties to the objective leads to unstable and difficult-to-interpret training dynamics. This mirrors, in a simplified setting, one of the conceptual issues in the original paper: while physics-informed guidance is attractive in theory, in practice it requires very careful numerical design and scaling to avoid degrading the behaviour of the underlying diffusion model rather than improving it.

4 CONCLUSIONS

4.1 Advantages and Scientific Impact

The main conceptual advance of [Shu et al. 2023] is not the specific physics add-ons, but the reformulation of CFD super-resolution as a *conditional denoising* problem solved by a DDPM. By training the model exclusively on HF data and learning to invert a fixed Gaussian corruption process, the network becomes independent of any particular LF generation pipeline. At inference time, arbitrary LF observations (downsampled, filtered or sparsely sampled) are first embedded into the noisy HF manifold via noise mixing, and the same reverse diffusion process is reused to reconstruct the flow. This decoupling of the reconstruction model from the LF distribution addresses a key weakness of classical CNN or GAN approaches, which typically must be retrained whenever the measurement operator or resolution changes and tend to fail under distribution shift.

A second important advantage is the robustness and stability inherited from the diffusion framework itself. DDPMs are trained with a well-behaved denoising objective, avoiding the mode-collapse and adversarial instabilities that often plague GAN-based surrogates. In

the context of CFD, this translates into generative models that reliably produce physically plausible flow fields across a range of noise levels and degradation types. The experiments in [Shu et al. 2023] show that, even without sophisticated conditioning mechanisms, the diffusion model can recover fine-scale structures and maintain reasonable energy spectra from severely corrupted inputs. In my own minimal implementation, the DDPM’s ϵ -prediction mechanism behaved consistently with the theory: the validation loss decreased smoothly as the model learned to denoise the synthetic fields, even though the generated samples were still far from realistic CFD solutions. While the experiment revealed that naive additions such as physics-style penalty terms can destabilize training, it also reinforced the central insight of [Shu et al. 2023]: the HF-only diffusion formulation provides a stable and well-posed learning problem, independent of the specifics of the LF data or the reconstruction task. In this sense, the strength of the framework stems from the DDPM core, not from task-specific architectural modifications.

The proposed framework also offers a natural interface for incorporating physics in a modular way. When the governing equations and discretisation are known exactly, the Navier-Stokes residual can be evaluated and used as an additional guidance signal during sampling, nudging the reverse diffusion trajectory toward states that better satisfy the PDE. Although the quantitative improvements reported in [Shu et al. 2023] are modest and largely confined to idealised settings, this mechanism illustrates how diffusion models can act as flexible priors that can be combined with physics-based operators or numerical solvers. More broadly, the approach positions diffusion models as a bridge between data-driven learning and traditional CFD: the same generative model can, in principle, be coupled with different measurement operators, embedded into hybrid CFD–ML workflows, or extended with latent-space or transformer-based architectures, providing a structured path toward more general and physically grounded flow reconstruction methods.

4.2 Limitations and Open Challenges

Despite the methodological elegance of the diffusion-based framework, several important limitations must be acknowledged before such models can be deployed in realistic CFD scenarios. The most substantive limitation concerns the physics-informed guidance proposed in [Shu et al. 2023]. Although attractive in principle, the approach relies on assumptions that hold only in highly idealised settings. The PDE residual is computed from three consecutive DNS frames using a central-difference approximation of the time derivative, which implicitly requires access to the “future” frame ω_{t+1} . This violates the causality constraints of any forecasting or online reconstruction task and restricts the method to offline post-processing where the full trajectory is already known. Moreover, the residual itself is evaluated using Fourier-based operators under periodic boundary conditions and a fixed Reynolds number. This makes the physics correction highly dependent on the availability of perfect PDE knowledge and an exact spectral discretisation. As a result, the physics-guided variants improve residuals mainly in the same synthetic Kolmogorov-flow setting in which they are trained, but do not readily generalise to more realistic geometries, imperfect data, or unknown governing parameters.

A second central limitation is computational efficiency. Diffusion models rely on hundreds of reverse denoising iterations, each requiring a full U-Net pass. While this iterative structure contributes to the model’s stability and robustness, it comes at a significant computational cost compared to single-pass neural surrogates, reduced-order models, or learned PDE solvers. For practical CFD applications especially three-dimensional flows, high-resolution grids, or parameter sweeps this sampling overhead becomes prohibitive. Recent benchmark studies confirm that diffusion-based surrogates offer strong physical consistency but still face substantial barriers in terms of runtime and scalability [Liu and Thuerey 2024].

Finally, although diffusion models are probabilistic in nature, [Shu et al. 2023] does not explicitly quantify uncertainty in the reconstructed fields. The absence of calibrated uncertainty estimates limits the interpretability of the predictions and restricts the use of such models in safety-critical or engineering design workflows, where reliability and confidence bounds are essential. Together, these challenges highlight that while diffusion models provide a promising and conceptually powerful framework, substantial work remains before they can be integrated into production-level CFD pipelines.

4.3 Future Work and Research Directions

Recent developments since [Shu et al. 2023] point toward two concrete and promising directions for advancing diffusion-based CFD models beyond the limitations of the original framework. A first direction involves replacing the idealised downsampling–upsampling setting of Shu et al. with physically consistent guidance derived directly from coarse numerical solvers. Self-guided diffusion models such as SG-Diff [Li et al. 2025] introduce a principled mechanism for reconstructing high-resolution flows from solver-generated low-resolution fields using a predictor–corrector–advancer sampling scheme and importance-weighted training that emphasises fine-scale structures. Unlike the physics-informed guidance in [Shu et al. 2023], which requires access to future frames and perfect spectral derivatives, these self-guided methods operate on causally valid inputs and have demonstrated the ability to recover coherent vortical features from realistic coarse simulations without retraining. Extending the diffusion framework in this direction would address one of the central practical limitations identified in Section 4.2: the reliance on idealised DNS triplets and a rigid periodic-domain setting.

A second direction focuses on scalability to three-dimensional and geometrically complex flows. Latent-space diffusion models such as CoNFILD [Wang et al. 2024] compress the flow field into a low-dimensional implicit representation and perform the diffusion process in this latent space, drastically reducing memory and computational cost while preserving physical structure. These approaches have been successfully applied to wall-bounded turbulence and separated flows, offering zero-shot generalisation across geometries and boundary conditions. Complementary work on transformer-based diffusion architectures [Lei 2024] further demonstrates that replacing the U-Net backbone with windowed and plane-wise attention significantly improves the model’s ability to handle high-dimensional 3D domains. Together, these advances suggest a path

toward diffusion models that are both physically aware and computationally feasible for practical CFD applications.

More broadly, integrating these two lines of work–solver-consistent guidance and geometry-aware latent diffusion–would enable hybrid CFD–ML pipelines capable of operating on complex meshes, irregular boundaries and realistic simulation outputs. Such models would extend far beyond the controlled Kolmogorov–flow setting of [Shu et al. 2023] and represent a major step toward making diffusion-based surrogates a standard tool in next-generation CFD workflows.

A ADDITIONAL FIGURES

A.1 Training and Validation Loss Curves

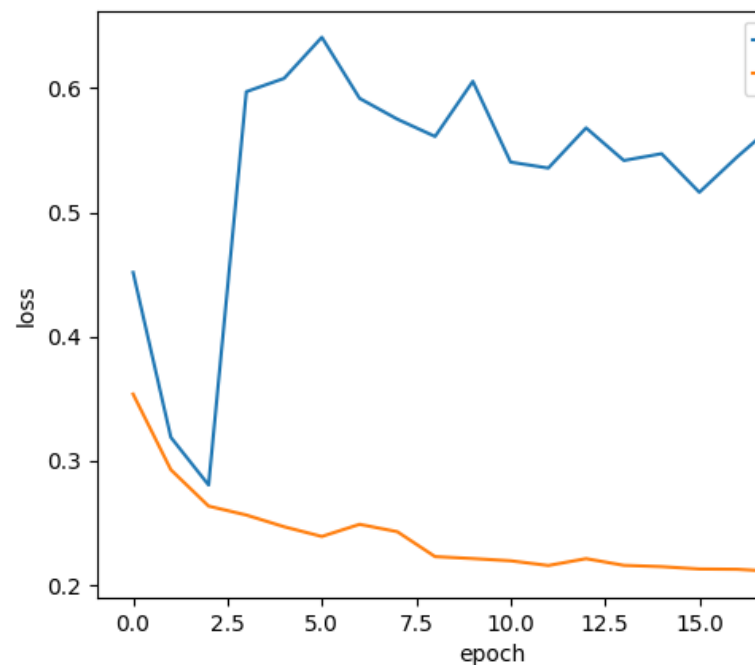


Fig. 1. Training and validation loss over 20 epochs for the minimal DDPM model. The validation loss uses pure ε -MSE, while the training loss includes an additional physics-inspired penalty term.

REFERENCES

- Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. 33 (2020), 6840–6851. <https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html>
- Jonathan Ho and Tim Salimans. 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598* (2022).
- Brian Edward Launder and Bahrat I Sharma. 1974. Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc. *Letters in heat and mass transfer* 1, 2 (1974), 131–137.
- Fan Lei. 2024. Reconstructing 3D Flow from 2D Data with Diffusion Transformer. *arXiv preprint arXiv:2502.02593* (2024).
- Ruoyan Li, Zijie Huang, Haixin Wang, Guancheng Wan, Yizhou Sun, and Wei Wang. 2025. Self-Guided Diffusion Model for Accelerating Computational Fluid Dynamics. (2025). *arXiv:cs.CE/2504.04375* <https://arxiv.org/abs/2504.04375>

- Qiang Liu and Nils Thuerey. 2024. Uncertainty-aware surrogate models for airfoil flow simulations with denoising diffusion probabilistic models. *AIAA Journal* 62, 8 (2024), 2912–2933.
- Heinz Pitsch. 2006. Large-eddy simulation of turbulent combustion. *Annu. Rev. Fluid Mech.* 38, 1 (2006), 453–482.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. (2015). arXiv:cs.CV/1505.04597 <https://arxiv.org/abs/1505.04597>
- Dule Shu, Zijie Li, and Amir Barati Farimani. 2023. A physics-informed diffusion model for high-fidelity flow field reconstruction. *J. Comput. Phys.* 478 (April 2023), 111972. DOI: <https://doi.org/10.1016/j.jcp.2023.111972> Elsevier BV.
- ML Shur, PR Spalart, MK Strelets, and AK Travin. 2008. A hybrid RANSLES approach with delayed-DES and modeled LES capabilities. *International Journal of Heat and Fluid Flow* 29 (2008).
- Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).
- Jian-Xun Wang, Pan Du, Meet Parikh, Xiantao Fan, and Xin-Yang Liu. 2024. CoN-FILD: Conditional Neural Field Latent Diffusion Model Generating Spatiotemporal Turbulence. (2024).