# **Accelerating High-fidelity Airfoil Design via Physics-informed Video Diffusion Model**

Yifeng Ai1#, Yijia Guo2#, Hongfu Zhang3\*, Lei Zhou4\*, Tiejun Huang2, Xuhui He4, Peng Guo4, C.W. Lim1,5\*, Lei Ma2,6\*

1. City University of Hong Kong, Hong Kong SAR, China;

2. National Key Laboratory for Multimedia Information Processing, Peking University, Beijing, China;

3. The Hong Kong Polytechnic University, Hong Kong SAR, China；

4. Central South University, Changsha, China;

5. Zhejiang University, Hangzhou, China;

6. National Biomedical Imaging Center, Peking University

Corresponding email address:

Dr. Hongfu Zhang:

Dr. C.W. Lim: bccwlim@city.edu.hk; Dr. Lei Ma: lei.ma@pku.edu.cn

\*These authors contributed equally: Yifeng Ai#, Yijiang Guo#.

# **Abstract**

Airfoil design remains a challenging multi-objective optimization problem requiring precise geometric control for optimal aerodynamic performance. Data-driven inverse design methods mitigate this complexity but often yield non-physical results when physical constraints are not explicitly incorporated. Here, we construct a high-fidelity NACA–Nek1000 dataset with 53,400 samples from direct numerical simulations to train a dual physics-informed video diffusion model (PVDM). The governing equations of incompressible flow are intrinsically coupled with the diffusion learning process, enabling the reconstruction of full-field velocity and pressure distributions from lift-to-drag response sequences. The PVDM achieves accurate low–Reynolds number airfoil reconstruction and real-time flow generation within 30 seconds per case, approximately 600 times faster than conventional simulations, while maintaining over 90% fidelity. The reconstructed airfoils are further modeled as propeller geometries and experimentally validated which well fit the simulated lift-to-drag responses, establishing a rapid and generalizable framework for high-fidelity, physics-informed inverse aerodynamic design.

# **Introduction**

Airfoil design constitutes a fundamental challenge in aircraft and underwater engineering, representing a complex, multi-objective optimization task. This process centers on maximizing the lift-to-drag ratio (*CL*/*CD*) to ensure optimal aerodynamic efficiency. However, achieving optimal *CL*/*CD* performance across a wide range of angles of attack (AoA) requires intricate geometric balancing. Precise tuning of parameters, specifically thickness distribution, camber magnitude, and maximum camber position are critical, as each exerts a pronounced non-linear influence on the flow field. despite these strict requirements, current workflows remain largely heuristic and dependent on legacy engineering practices. Aerodynamic characteristics are typically evaluated through computationally intensive fluid dynamics simulations, ranging from high-fidelity direct numerical simulation (DNS)1,2,3, large-eddy simulation (LES)4,5,6 to efficient Reynolds-averaged Navier–Stokes (RANS) models7,8,9, or through experimental wind-tunnel validation using particle image velocimetry10,11,12,13.

To overcome these limitations, inverse airfoil design strategies have emerged, deriving profiles directly from aerodynamic targets in turbulence-dominated aircraft or laminar bionic systems. Recent machine learning approaches such as convolutional neural networks14 and reinforcement learning15 capture nonlinear geometry performance mappings, improving design efficiency.

~~but lacking physical interpretability~~. ~~Generative diffusion frameworks, including denoising diffusion probabilistic models (DDPMs)~~~~16~~~~, address this issue by learning multimodal geometry performance distributions through iterative denoising processes.~~

~~DDPMs can predict high-dimensional flow fields from simple conditioning inputs such as~~ *~~C~~~~L~~*~~/~~*~~C~~~~D~~* ~~responses~~~~17~~~~. Their spatiotemporal extensions, video diffusion models (VDMs)~~~~18,19~~~~, generate coherent frame sequences but may introduce non-physical artifacts when physical supervision is absent~~~~20~~~~. Incorporating physics constraints into VDMs to enhance flow fidelity remains~~ ~~underexplored in inverse aerodynamic design.~~

This study introduces a physics-informed video diffusion model (PVDM) for accurate reconstruction of low Reynolds number airfoil geometries. Airfoils are parameterized by the three defining NACA variables: maximum camber, camber position, and chord thickness. The model infers these geometric features from *CL*/*CD* sequences across varying AoA values. Training and testing use the NACA–Nek1000 dataset, comprising 53,400 DNS samples of two-dimensional incompressible flow fields generated via high-order spectral element simulations at three spatial resolutions. The *CL*/*CD* response tokens are extracted from DNS and used as model inputs. Physics guidance is incorporated by embedding normal gradients of flow field into the diffusion framework, coupling the learning process with the incompressible Navier–Stokes equations. The model adopts a dual VDM architecture based on a 3D U-Net with cross-attention layers that inject *CL*/*CD* –AoA sequences into spatial and temporal dimensions. One branch reconstructs global flow contours, and the other refines near-airfoil regions, compensating for the limitations of single 3D U-Net models in resolving high-resolution features. 第二段讲传统机器学习的问题（缺乏物理引导，低纬到高纬映射）；第三段讲我们提出的PVDM（讲VDM的好处，讲物理引导的好处）；第四段为了训练这个模型，我们设计了这个数据集。

The proposed PVDM generates flow-field videos within 30 seconds per case approximately 600 times faster than DNS (~300 minutes for 10 cases) while maintaining > 90% structural fidelity (PSNR: 34–49 *dB*; SSIM: 0.83–0.99). It requires only *CL*/*CD* sequences as inputs, eliminating boundary condition setup and pre-testing overhead. The results demonstrate high temporal and spatial coherence, establishing a scalable paradigm for physics-informed generative modeling in aerodynamic inverse design.

# **Results:**

## **Training and Testing Datasets Generation**

The construction of the NACA-Nek1000 dataset relies on a rigorous pipeline linking geometric parameterization to high-fidelity flow physics, as summarized in Fig. 1. To ensure comprehensive coverage of the design space, we generated airfoil geometries based on the NACA 4-digit parameterization, sweeping through a wide range of camber, thickness, and angle of attack configurations (Fig. 1a).

The high-fidelity flow field of such airfoil profiles is calculated by DNS at Reynolds number (, where the inlet flow velocity , is the corth length , is the kinematic viscosity). We employed a higher order spectral-element method with a self-adaptive meshing strategy to guarantee results accuracy, particularly in the critical near-wall and wake regions (Fig. 1b, c). To standardize the data for learning, the resolved two-velocity-components (, ) and pressure () were cropped and interpolated onto uniform 9696 grids, stored in floating-point format to preserve numerical precision (Fig. 1d, e). Note that all the variables in this paper are dimensionless, which are normalized by the reference velocity and chord length . By labelling these full-field flow snapshots with their corresponding *CL*/*CD* (Fig. 1f), we established a high-fidelity training protocol containing 53,400 paired samples (Fig. 1g). This dataset provides the physical ground truth for training the PVDM.

|  |
| --- |
|  |
| **Fig. 1: Process employed for generating the training and testing datasets:** **a. Definitions of Airfoil:** airfoil envelopes follow the NACA-4 digit, maximum camber A varies from 1 to 9 at a step of 1, location of max camber B%\*10 varies from 1 to 9 at a step of 1, thickness C% varies from 2 to 98 at a step of 2, chord length *Lc*; **b. Calculation Domain Settings:** the mesh is densified significantly around the airfoil and the wake region for promoting the accuracy, the domain area is set as *X/Lc* varies from -12 to 24, *Y/Lc* varies from -8 to 8. The boundary conditions: inflow side as well as two sides is set as *u=U*, *v=*0 and *=*0 (n is the unit normal vector of the boundary), outflow is set as *=*0, and *=*0, the airfoil wall boundary is set as *u=*0, *v=*0, and *=*0; **c. Flow Field Contours:** The pressure contours are superimposed with velocity streamlines of airfoil region and wake region; **d. Data Cutting:** flow fields of *u*, *v* and *p* are cut out which focus on the target region of flow past the airfoil; **e. Data Post-processing:** the cut-out flow fields are linearly interpolated into 9696 structured data and transferred into files as the “npy” format; **f.** ***CL*/*CD* Response Extraction:** the mean *CL*/*CD* responses are extracted from the wall surface of airfoil, which labels the corresponding “npy” file of flow field and build accurate reflection; **g. Datasets Visualization:** the training protocol comprises 1800 airfoil profiles which can be visualized by the span-wise fixed airfoil. Note that all the variables in this paper are dimensionless, which are normalized by the reference velocity *U* and chord length *Lc.* |

## **Physics-informed Video Diffusion Model**

Fig. 2.a shows the training phase of the proposed PVDM. We first perform the tokenization process, in which the *CL*/*CD* values varying with the AoA are encoded as a temporal vector to label each frame. Before temporal embeddings are applied, the shaped blank frames are spatially convolved and appended with spatial attention. These spatio-temporal features are then passed into the 3D U-Net architecture to predict noise at time *t* for DDPM process. The forward process of DDPM can be mathematically represented for a process transitioning over a period *T*. For this model, the Gaussian noise is gradually added to the data which is later denoised in a learned reverse process seeking to reconstruct the original distribution*.* The probability of distribution *q* from point *xt*−1 at time *t−*1 to *xt* at time step *t*, which remains to be learned, is a fixed Markov chain with a given variance schedule:

|  |  |
| --- | --- |
|  | (1) |

where *x1,...,xT* are the latents of the same dimensionality as the data *x0* ~ *q*(*x0*), *β* ∈ (0,1) is the scheduled variance associated with the Gaussian distribution *N*, and *I* denotes the identity matrix. The closed-form sampling process of Eq. 1 is:

|  |  |
| --- | --- |
|  | (2) |

where *z* ~ (0,1) is a noise-predictor and . To train this model, we can optimize the variational lower-bound for . In actual training, we simply set the optimization objective as:

|  |  |
| --- | --- |
|  | (3) |

The variable *zθ* is predicted by the U-Net network, and *z* is obtained through a denoising process. Drawing inspiration from recent advances in video generative models and inverse design models, our architecture leverages variants of the standard 2D (space-only) U-Net, where the applied *CL/CD* dimension is effectively treated as a batch axis, thereby preserving the integrity of the foundational base network. To ensure physical consistency across varying strain steps, by packaging the flow field's velocity and pressure components as a three-channel feature set for network input, we promote a more holistic representation of the flow state. Crucially, this input paradigm provides the necessary basis for introducing physics-informed regularization during subsequent inference. Two specialized models are trained: one focusing on the airfoil region (AR) and the other on the general flow region (GR).

Fig. 2.b shows the inference phase of the proposed Physics-informed Video Diffusion Model. According to Chung et al.23, an estimate of an be derived from the following specialized representation of the posterior mean:

|  |  |
| --- | --- |
|  | (4) |

Further, the inference result at timestep *t1* without correction can be expressed as:

|  |  |
| --- | --- |
|  | (5) |

where *σt* =. It is noted that (and its variants) is always learned by neural networks24. In the case of conditional generative modeling, the data *x* is drawn jointly with conditioning information *c*, such that Eq. (31) can be further represented as:

|  |  |
| --- | --- |
|  | (6) |

here we have used the fact that:

|  |  |  |
| --- | --- | --- |
|  | | (7) |

In our case, we have two conditions: lift-to-drag ratio guidance *cl* and N-S equation guidance . Here can be obtained by classifier-free25 training and sampling. However, cannot be learned through the methods above because it is related to . The diffusion posterior sampling (DPS)26 showed that under Gaussian noise assumption of the sparse measurement operator , the log-likelihood function can be approximated with:

|  |  |
| --- | --- |
| , | (8) |

where 0 is an estimated value of *x0* through Eq. 31. Given the governing equation for flow:

|  |  |
| --- | --- |
| *∇u =* 0 | (9) |
|  | (10) |

we have the N-S physical correction (hyperparameter) as:

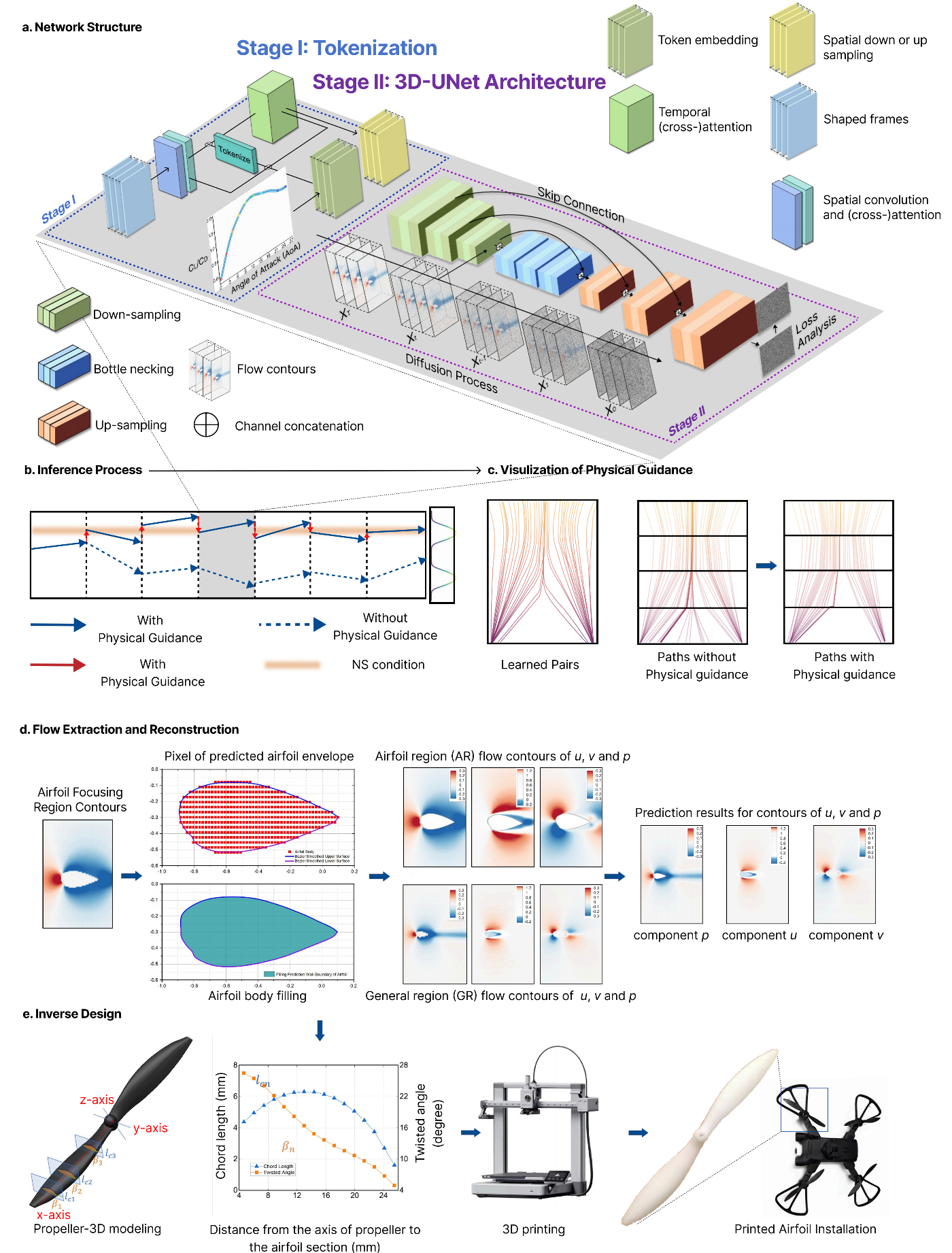
|  |  |
| --- | --- |
|  | (11) |

and:

|  |  |
| --- | --- |
|  | (12) |

where is the result of *CL*/*CD* guidance according to Eq. 32, represents the N-S equation and is the N-S constraint coefficient.

We provide 3 pages for further details of derivation of the PVDM in the SM (Section 2), where the algorithm is given. By integrating N-S constraints, we substantially mitigate the epistemic uncertainty inherent in diffusion models. This yields reconstructed wing geometries and synthesized flow fields that demonstrate enhanced adherence to fundamental physical laws.



|  |
| --- |
| **Fig. 2. PVDM architecture and merger of full-field flow contours: a. Network Structures.** Stage I: Tokenization: the *CL/CD* ratio curves are stored as a vector and then embedded into frames. Blank frames are shaped as: 96\*96\*10, which is attached with the spatial convolution and then transferred into the spatial attention. With the tokenized *CL/CD* values, all of frames based on the increasing rotational angles (AoA) are labeled temporally. Finally, the labeled frames are inserted into the spatial up or down sampling process within the 3D U-Net of Stage II. Stage II: 3D U-Net Architecture: the first DDPM illustrated by the purple region reconstructs global flow contour frames via the denoising process, flow fields encompassing both global and local features are generated by averages of frames stochastically sampled from both models; *CL*/*CD* ratio token sequences are embedded into spatiotemporal convolutional modules via cross-attention mechanisms; physical guidance is added to restrain the down-sampling process. **b. Inference Process.** The diffusion and denoising processes are displayed along a Markov’s chain based on the figure distribution *x0*, ..., *xt-1*, *xt*, *xT*, and the physics guidance can be shown as the PDE loss channeled which to minimize the difference between prediction and real interpolated data. **c. Visualization of Physical Guidance.** Physical guidance steers diffusion trajectories toward physically consistent manifolds, yielding smoother and divergence-free streamline evolution. **d. Flow Field Extraction.** Airfoil profiles are extracted from flow fields and reconstructed using Bezier-curve fitting along the edges, finally fused to generate the flow fields of *u*, *v* and *p* and enhance the inverse design. **e. Inverse Design.** The output airfoil profiles are modelled in 3D as a propeller and then to be produced by 3D printer, the printed propeller is then installed for further experimental tests. |

## **Full-field Flow Predictions**

The generalization performance of the trained PVDM model in predicting the *CL*/*CD* response beyond the training protocol was evaluated using a testing dataset partitioned into three distinct test cohorts: (i) high *CL*/*CD* magnitude variants (1/3 of all samples), (ii) profiles with curvature slopes exceeding all training samples (1/3 of all samples), and (iii) geometrically analogous samples (the remaining 1/3). Three representative airfoils are demonstrated in Fig. 3: NACA5910 in test cohort (i), NACA5431 in test cohort (ii), and NACA7122 in test cohort (iii).

The PVDM demonstrates robust inverse design capabilities. Utilizing the target *CL*/*CD* trajectories obtained from DNS as input (Fig. 3a–c, bold blue lines), the model accurately reconstructs the corresponding airfoil profiles. The accuracy performance of this geometric reconstruction is evidenced in Fig. 3g–i, where the Bezier-fitted predictions exhibit exceptional agreement with the ground-truth theoretical envelopes. To validate the aerodynamic performance of these inversely generated geometries, we performed DNS re-simulations on the predicted profiles. The resulting *CL*/*CD* responses (Fig. 3a–c, orange scatter points) align precisely with the target inputs, confirming the functional accuracy of the design.

Beyond scalar airfoil profile, the model's capacity to predict full-field flow variables was assessed. A comparison of the velocity (*u*, *v*) and pressure (*p*) distributions (Fig. 3j–o) reveals that the PVDM-predicted fields are qualitatively indistinguishable from the DNS baselines across varying angles of attack (AoA). This visual consistency is substantiated quantitatively by high Peak Signal-to-Noise Ratios (PSNR, 34.20–54.35 *dB*) and Structural Similarity Index Measures (SSIM > 0.8) as shown in Fig. 3d–f. These metrics confirm that the proposed approach achieves high-fidelity prediction of 2D flow fields solely through PVDM inference, bypassing the computational cost of traditional solvers. Comprehensive results across all 10 AoA values are detailed in Supplementary Material (SM, Section 5).

|  |
| --- |
|  |
| **Fig. 3. Inverse airfoil prediction and full-field flow reconstruction results for representative NACA airfoils: NACA5190, NACA5431, NACA7122. a-c,** *CL*/*CD* responses at all angles of attack (AoA) for NACA 5910, 5431 and 7122, respectively. **d-f,** PSNR, SSIM values of general flow contours between prediction and DNS simulation. **g-i,** Actual and inversely generated airfoil profiles from Bezier curve fitting. **j-l,** Comparison of predicted and simulated velocity and pressure distributions obtained for selected AoA values within 0◦ to 27◦ at increment of 9◦ at 96 × 96 pixels global regions. **m-o,** Comparison of predicted and simulated velocity and pressure distributions obtained for selected AoA values within 0◦ to 27◦ at increment of 9◦ at 256 × 256 pixels global regions. |

Beyond statistical similarity metrics, we scrutinized the physical validity of the generated flow fields by analyzing the residuals of the governing equations. Figure 4 summarizes the PDE loss characteristics, highlighting the impact of the physics-informed constraints. The cumulative loss maps presented in Fig. 4a, b visualize the normalized per-pixel deviations aggregated over the *u*, *v*, and *p* components for the NACA5431 and NACA7122 airfoils.

A comparative assessment reveals that the inclusion of physical guidance markedly suppresses spatial errors and high-frequency artifacts relative to the unguided baseline. This mitigation confirms that explicitly enforcing incompressible Navier–Stokes constraints (Eqs. 9–10) during inference effectively regularizes the flow field prediction. Furthermore, the sensitivity of the model to the loss-weighting hyperparameter is quantified in Fig. 4c. The optimization landscape demonstrates a convex profile, attaining a global minimum PDE loss of 0.212 (0.037 s.d.) at the optimal hyperparameter setting. This indicates a stable operating regime where the balance between data fidelity and physical consistency is maximized.

|  |
| --- |
|  |
| **Fig. 4: Evaluation of physical consistency and hyperparameter sensitivity. a,** Average PDE loss (means.d.) across all sampling cases as a function of the loss function hyperparameter, identifying an optimal weighting that minimizes physical residuals. **b, c,** Cumulative PDE loss maps for the NACA5431 (a) and NACA7122 (b) airfoils. Both panels contrast the prediction with physical guidance (left) against the unguided baseline (right), visualizing normalized per-pixel residuals (scaled10). |

# **Discussion**

The proposed physics-informed video diffusion model (PVDM) achieves inverse airfoil design with remarkable efficiency by relying solely on compact *CL*/*CD* ratio tokens. Each prediction requires approximately 30 seconds of computation, compared with the several hours typically needed for iterative optimization based on conventional CFD simulations. This accelerated framework demonstrates substantial potential for mission-specific airfoil tailoring in unmanned aerial vehicles (UAVs), enabling rapid adaptation to low-altitude flight regimes relevant to delivery, surveillance, and inspection tasks. The model accurately reconstructs low Reynolds number flow fields, providing dynamic predictions of velocity and pressure distributions across varying AoAs with a mean contour error below 5%. The dual-model architecture reduces training cost by nearly 600-fold relative to a single DDPM configuration while mitigating resolution-dependent artifacts. In addition, the generated NACA–Nek1000 dataset offers publicly accessible, high-precision flow data (*Re* = 1000) for micro–aerial robot research. Collectively, these contributions establish a foundation for real-time, high-fidelity aerodynamic design and visualization.

Despite its promising performance, the PVDM exhibits limited generalization when exposed to distributionally shifted *CL*/*CD* tokens, constraining its extrapolative capability. High-resolution training (≥ 256×256) can introduce duplication or ghosting artifacts in predicted contours, partly due to the local attention of the U-Net backbone, which restricts the modeling of long-range spatial dependencies. These artifacts arise primarily from noise amplification in sparse-mesh regions and discontinuities at airfoil wall interfaces during block-wise processing. Moreover, the current model is trained exclusively on low Reynolds-number laminar flow conditions (), which limits its direct applicability to turbulent or transitional flow regimes. Although diffusion transformers (DiTs) offer global feature extraction and improved coherence, their substantial computational demands and data requirements currently hinder their deployment for high-resolution aerodynamic modeling.

This study outlines a new direction for integrating generative AI with physics-informed modeling in dynamic flow prediction and inverse aerodynamic design. The proposed framework provides an ultra-fast optimization feedback mechanism that could revolutionize the speed of aerodynamic reasoning. Future developments may incorporate datasets spanning a wider range of Reynolds numbers and geometries, enabling physically consistent predictions across laminar, transitional, and turbulent flows. By tokenizing characteristic flow features and embedding them into the diffusion process, the model could achieve tighter conformity with the Navier–Stokes equations. Looking forward, the framework may evolve into a diffusion-informed transformer (DiT) capable of reasoning directly from textual or symbolic design prompts, automatically generating tokenized performance curves under prescribed flow or boundary conditions. Such systems could support applications in frame interpolation, super-resolution reconstruction, and unsteady flow prediction, ultimately leading to customizable, lightweight generative models for complex aerodynamic and microfluidic environments.

# **Methods**

* 1. High-fidelity CFD simulation via the high spectral element method and *CL*/*CD* response extraction

Here we use a high-order spectral/hp element discretization27, which expands the solution with degree-*p* polynomials on each element of an unstructured mesh, combining the geometric flexibility of finite elements with the resolution properties of spectral methods. The formulation supports *h*- (mesh) and *p*- (polynomial) refinement, as well as their *hp* combination, yielding algebraic error decay () or sufficiently smooth fields. We adopt a standard Galerkin weak form with -continuous expansions and element-level operators, which enables efficient solvers (e.g., static condensation) while preserving high-order accuracy. In present study, the polynomial order is set as 4 in the global domain, which is much higher than the conventional finite volume method (second order). Along with DNS simulation, we can get the high-fidelity flow field. Here, we use semi-implicit scheme to directly solve the N-S equations shown in equations (9) and (10). The scheme discretises the momentum equation Eq. (9-10) with a backwards approximation of the time derivative to obtain:

|  |  |
| --- | --- |
|  | (13) |

where is an intermediate velocity and is the summation of previous solutions. With the discrete time derivative, we initially solve a pressure Poisson equation of the form:

|  |  |
| --- | --- |
| () | (14) |

and use consistent Neumann boundary conditions prescribed as:

|  |  |
| --- | --- |
| -[ | (15) |

The advection term is , where the superscript indicates extrapolation from previous solutions.

Subsequently, we solve a Helmholtz equation for new velocity [*, ,*], thus:

|  |  |
| --- | --- |
|  | (16) |

Generally, this scheme solves for the new pressure and velocity based on initial conditions at and boundary conditions. The algorithm proceeds in three steps: first, the advection terms are evaluated; second, a Poisson problem for the updated pressure is solved; finally, a Helmholtz problem is solved for each velocity component.

The validation details of lift and drag coefficients are shown in SM (Subsection 3.3) at . A *p*-th order of is selected for all general flow domain calculation cases to balance computational accuracy with time and storage costs. It should be mentioned that all of the physical variables are dimensionless.

The lift and drag force coefficients (*CL*/*CD*) are extracted from the temporal snapshots involving full-field , , and distribution data over time, as follows.

|  |  |
| --- | --- |
|  | (17) |
|  | (18) |

Here, the working fluid is air, where the air density is assumed to be 1 kg/m3, is applied vertical lift aerodynamic force on the airfoil, and is the horizontal drag aerodynamic force applied on the whole airfoil. The uniform inlet velocity is dimensionless is set as default for all of simulation cases. Therefore, the Reynolds number at the control inlet is set as according to the chord length of the airfoil surface (where ).

## **Normalization of DNS data**

All discretized 2D Reynolds-average full-field flow data (, , and ) and *CL*/*CD* ratio tokens obtained from the DNS process were normalized according to the following process based on their respective global maximum and minimum values obtained over all matrix coordinates, denoted as and , respectively.

|  |  |
| --- | --- |
|  | (19) |

Here, the position of each pixel value is denoted as (*idx*, *idy*). The global maximum and minimum values of the flow data are obtained as and . Therefore, all normalized data has been restrained within the range [−1, 1], while a constant “” is included in the denominator to avoid division by zero under the unlikely case where .

## **2D U-Net framework**

Herein, we summarize the spatial 2D U-Net framework28 at a conceptual level. More comprehensive technical implementation details can be referred to ‘Code availability’. All implementations apply the PyTorch framework (*v*2.1+)29. Because diffusion models operate through iterative noise removal from typically image data, the U-Net architecture must meet the requirement for identical input-output dimensions. Our model extends the foundational 2D U-Net architecture to accommodate the video diffusion framework and its reference implementation30. The applied encoder-decoder structure progressively compresses spatial information while expanding latent features (down-sampling), and then inversely reconstructs spatial resolution from latent representations (up-sampling). Each downsampling/ upsampling stage integrates the following features:

(i) dual residual neural network (ResNet) blocks with convolutional layers and sigmoid linear unit (SiLU) activation functions 31;

(ii) spatially factorized self-attention for enhanced computational efficiency32;

(iii) strided convolutions for resolution scaling.

The bridge module acting between and encoder and decoder similarly employs dual ResNet blocks with full self-attention interleaved33. We deployed four hierarchical resolutions (96×96→12×12) with exponentially increasing latent dimensions (64→512), where each multi-head attention block contains 8 heads (dimension = 32 per head). The critical hyperparameters employed during network training are cataloged in SM (Subsection 3.2) along with details regarding the experimental platforms employed for dataset generation, model training, and model testing.

## **4.3. Extension to temporal 3D U-Net architecture**

The foundational 2D U-Net architecture was extended into a temporal 3D architecture according to a previously reported process34 that integrated a quasi-temporal dimension representing an applied mechanical process of airfoil swinging associated with AoAs, and treated this dimension as a batched axis to maintain structural consistency across all building blocks. This architecture is clearly illustrated schematically in Fig. 2, where, crucially, a temporal self-attention layer is embedded at a point prior to the encoder-decoder framework and iteratively after each spatial attention module, with spatial axes batched to enable attention across 10 discrete *CL*/*CD* intervals. In addition, a previously reported relative positional encoding *Prel*(*ti*, *tj*) is implemented at times *ti* and *tj* to encode AoA-sequence dependencies35. This encoding thereby preserves mechanical chronology while minimizing spectral leakage artifacts, as demonstrated in Figs. 3a, 4a, and 5a.

## **4.4. Conditioning the PVDM on *CL*/*CD* responses to the AoA**

The PVDM was conditioned on the *CL*/*CD* response to AoA by projecting all of 10 AoA values into a latent embedding via a parameterized linear projection . These embeddings are concatenated to spatial attention tokens for cross-attention, where queries derive from pixel embeddings and keys/values derive from conditioning embeddings, while temporal attention tokens integrate all 10 embeddings with the relative positional encoding *Prel* (*ti, tj*) discussed in the preceding subsection to preserve mechanical chronology. Crucially, a conditioning amplifier averages AoA-step embeddings and transforms them via a two-layer multilayer perceptron (MLP) with SiLU activations (*β* = 1.702)31, which projects outputs to match the dimensionality of the latent space of diffusion timestep *t*. This fused representation is additively incorporated into the ResNet blocks of the PVDM, which synchronizes denoising dynamics with aerodynamic conditioning, while suppressing spectral leakage artifacts, as demonstrated in Figs. 3c, 4c, and 5c.

# **Data availability**

The training and validation protocol, including the token response and corresponding full-field flow data distributions (variables *u*, *v*, and *p*) for the DNS-solved global region and airfoil-focused region calculation subdomains is posted at https://huggingface.co/datasets/diff-flow/diff-flow/tree/main, where the pretrained PVDM is also given.

# **Code availability**

The post data processing code and the main code used for training the network applied for mapping connections between *CL*/*CD* ratio response and corresponding frames of full-field *u*, *v*, and *p* distributions is available at <https://github.com/yifengai2/diffusion-model-of-inversely-design-airfoil-flowcontours>

# **Acknowledgement**

The work described in this paper was supported by National Natural Science Foundation of China (Project No. NSFC 12572031).

# **Reference**

1. Ai, Y., Zhou, L., Tse, K. T., & Zhang, H. (2023). Interference and ground effects on flow past two inclined flat plates in tandem arrangement. *Ocean Engineering*, *270*, 113653. <https://doi.org/10.1016/j.oceaneng.2023.113653>
2. Adak, R., Mandal, A., & Saha, S. (2024). Reynolds-averaged Navier–Stokes assisted direct numerical simulation of low Reynolds number flow over airfoils. *Physics of Fluids*, *36*(12). <https://doi.org/10.1063/5.0237871>
3. He, Z., Zhao, Y., Zhang, H., Tang, H., Zhu, Q., Ai, Y., He, X., & Zhou, L. (2025a). Vortex-induced vibrations and galloping of a square cylinder: The impact of damping and mass ratio. *Ocean Engineering*, *320*, 120371. <https://doi.org/10.1016/j.oceaneng.2025.120371>
4. Deardorff, J. W. (1970). A numerical study of three-dimensional turbulent channel flow at large Reynolds numbers. *Journal of Fluid Mechanics*, *41*(2), 453–480. https://doi.org/10.1017/s0022112070000691
5. Wang, R., Wu, F., Xu, H., & Sherwin, S. J. (2021). Implicit large-eddy simulations of turbulent flow in a channel via spectral/hp element methods. *Physics of Fluids*, *33*(3). https://doi.org/10.1063/5.0040845
6. Sun, X., Yan, H., & Chen, F. (2025). Large eddy simulation of droplet breakup in turbulent flow with adaptive mesh refinement. *Physical Review Fluids*, *10*(2). https://doi.org/10.1103/physrevfluids.10.024004
7. Tang, L. (2008). Reynolds-Averaged Navier-Stokes simulation of Low-Reynolds-Number Airfoil aerodynamics. *Journal of Aircraft*, *45*(3), 848–856. <https://doi.org/10.2514/1.21995>
8. Winslow, J., Otsuka, H., Govindarajan, B., & Chopra, I. (2017). Basic understanding of Airfoil characteristics at Low Reynolds numbers (104–105). *Journal of Aircraft*, *55*(3), 1050–1061. https://doi.org/10.2514/1.c034415
9. Catalano, P., & Tognaccini, R. (2010). Turbulence Modeling for Low-Reynolds-Number Flows. AIAA Journal, 48(8), 1673–1685. <https://doi.org/10.2514/1.j050067>
10. Schimpf, A., & Kallweit, S. (2004). Photogrammetric particle image velocimetry. In *Springer eBooks* (pp. 295–300). <https://doi.org/10.1007/978-3-642-18795-7_21>
11. Westerweel, J., Elsinga, G. E., & Adrian, R. J. (2011). Particle image velocimetry for complex and turbulent flows. *Annual Review of Fluid Mechanics*, *45*(1), 409–436. <https://doi.org/10.1146/annurev-fluid-120710-101204>
12. Jiménez-Portaz, M., Chiapponi, L., Clavero, M., & Losada, M. A. (2020). Air flow quality analysis of an open-circuit boundary layer wind tunnel and comparison with a closed-circuit wind tunnel. *Physics of Fluids*, *32*(12). https://doi.org/10.1063/5.0031613
13. Rennie, R. M. (2022). Method of characteristics Analysis of Open-Return, Unsteady-Flow Wind-Tunnel performance. *AIAA Journal*, *60*(10), 6049–6053. <https://doi.org/10.2514/1.j061751>
14. Hu, J., & Zhang, W. (2023). Flow field modeling of airfoil based on convolutional neural networks from transform domain perspective. *Aerospace Science and Technology*, *136*, 108198. <https://doi.org/10.1016/j.ast.2023.108198>
15. Scavella, P., Paolillo, G., & Greco, C. S. (2025). Deep reinforcement learning-based airfoil design and optimization: An aerodynamic analysis. Aerospace Science and Technology, 167, 110638. <https://doi.org/10.1016/j.ast.2025.110638>
16. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Neural Information Processing Systems*, *33*, 6840–6851. <https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf>
17. Du, P., Parikh, M. H., Fan, X., Liu, X., & Wang, J. (2024). Conditional neural field latent diffusion model for generating spatiotemporal turbulence. *Nature Communications*, *15*(1). <https://doi.org/10.1038/s41467-024-54712-1>
18. Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., Fleet D. J., 2022, Video Diffusion Models, *Neural Information Processing Systems 36*, <https://papers.neurips.cc/paper_files/paper/2022/file/39235c56aef13fb05a6adc95eb9d8d66-Paper-Conference.pdf>
19. Yang, R., Srivastava, P., & Mandt, S. (2023). Diffusion probabilistic modeling for video generation. *Entropy*, *25*(10), 1469. <https://doi.org/10.3390/e25101469>
20. Bastek, J., & Kochmann, D. M. (2023). Inverse design of nonlinear mechanical metamaterials via video denoising diffusion models. *Nature Machine Intelligence*, *5*(12), 1466–1475. <https://doi.org/10.1038/s42256-023-00762-x>
21. Du, P., Parikh, M. H., Fan, X., Liu, X., & Wang, J. (2024). Conditional neural field latent diffusion model for generating spatiotemporal turbulence. *Nature Communications*, *15*(1). <https://doi.org/10.1038/s41467-024-54712-1>
22. Chung, H., Ye, J. C., Milanfar, P., & Delbracio, M. (2023). Prompt-tuning latent diffusion models for inverse problems. *arXiv preprint arXiv:2310.01110*.
23. Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2021). Score-Based Generative Modeling through Stochastic Differential Equations. *International Conference on Learning Representations*. <https://openreview.net/pdf?id=PxTIG12RRHS>
24. Ho, J., & Salimans, T. (2022). Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*
25. Patera, A. T. (1984). A spectral element method for fluid dynamics: Laminar flow in a channel expansion. *Journal of Computational Physics*, *54*(3), 468–488. <https://doi.org/10.1016/0021-9991(84)90128-1>
26. Zhao, Y., Chen, J., Zhang, Z., & Zhang, R. (2022). BA-NeT: Bridge attention for deep convolutional neural networks. In Lecture notes in computer science (pp. 297–312). <https://doi.org/10.1007/978-3-031-19803-8_18>
27. Cantwell, C., Moxey, D., Comerford, A., Bolis, A., Rocco, G., Mengaldo, G., De Grazia, D., Yakovlev, S., Lombard, J., Ekelschot, D., Jordi, B., Xu, H., Mohamied, Y., Eskilsson, C., Nelson, B., Vos, P., Biotto, C., Kirby, R., & Sherwin, S. (2015). Nektar++: An open-source spectral/hp element framework. Computer Physics Communications, 192, 205–219. <https://doi.org/10.1016/j.cpc.2015.02.008>
28. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-NET: Convolutional Networks for Biomedical Image Segmentation. In *Lecture notes in computer science* (pp. 234–241).https://doi.org/10.1007/978-3-319-24574-4\_28
29. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E. Z., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., . . . Chintala, S. (2022). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv (Cornell University)*, *32*, 8026–8037. https://doi.org/10.48550/arxiv.1912.01703
30. Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D. P., Poole, B., Norouzi, M., Fleet, D. J., & Salimans, T. (2022, October 5). Imagen Video: High definition video generation with diffusion models. arXiv.org. https://arxiv.org/abs/2210.02303
31. Elfwing, S., Uchibe, E., & Doya, K. (2018). Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, *107*, 3–11. <https://doi.org/10.1016/j.neunet.2017.12.012>
32. Katharopoulos, A., Vyas, A., Pappas, N., & Fleuret, F. (2022). Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention. *arXiv (Cornell University)*, *1*, 5156–5165. https://doi.org/10.48550/arxiv.2006.16236
33. Sergey Z., Kikos K. (2016). Wide Residual Networks, *arXiv (Cornell University),* <https://doi.org/10.48550/arXiv.1605.07146>
34. Chatterjee, S. et al. (2024). DDoS-unet: Incorporating temporal information using dynamic dual-channel unet for enhancing super-resolution of dynamic MRI, IEEE Access, 12, pp. 99122–99136. doi:10.1109/access.2024.3427674.
35. Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*.