

Latent attention on masked patches for flow reconstruction

Ben Eze¹[0009-0009-8594-082X], Luca Magri^{1,2}[0000-0002-0657-2611], and Andrea N3voa^{1,3}[0000-0003-0597-8326]

¹ Aeronautics Dept., Imperial College London, Exhibition Rd, London SW7 2BX, UK

² DIMEAS, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Italy

³ I-X, Imperial College London, 84 Wood Lane, London W12 0BZ, UK
{ben.eze21,l.magri,a.novoa}@imperial.ac.uk

Abstract. Vision transformers have shown outstanding performance in image generation, yet their adoption in fluid dynamics remains limited. We introduce the *Latent Attention on Masked Patches* (LAMP) model, an interpretable regression-based modified vision transformer designed for masked flow reconstruction. LAMP follows a three-fold strategy: (i) partition of each flow snapshot into patches, (ii) patch-wise dimensionality reduction via proper orthogonal decomposition, and (iii) reconstruction of the full field from a masked input using a single-layer transformer trained via closed-form linear regression. We test the method on two canonical 2D unsteady wakes: a laminar wake past a bluff body, and a chaotic wake past two cylinders. On the laminar case, LAMP accurately reconstructs the full flow field from a 90%-masked and noisy input, across signal-to-noise ratios between 10 and 30 dB. Further, the learned attention matrix yields interpretable multi-fidelity optimal sensor-placement maps. LAMP’s performance on the chaotic wake is limited, but outperforms other regression methods such as gappy POD. The modularity of the framework, however, naturally accommodates nonlinear compression and deep attention blocks, thereby providing an efficient baseline for nonlinear, high-dimensional masked flow reconstruction

Keywords: Flow reconstruction · Vision Transformers · POD

1 Introduction

Reconstructing high-dimensional fields from sparse, noisy, and masked measurements is a central challenge in scientific machine learning and fluid mechanics [8, 19, 10]. In experimental settings, particle image velocimetry (PIV) measurements cover only a subset of the domain, and, since each run captures an independent realisation of the flow, PIV is often limited to mean-flow analyses rather than time-resolved full-field measurements in high-dimensional settings, e.g., [1]. To reconstruct full flow fields, data-driven approaches have been employed, such as convolutional neural networks [9, 11], generative adversarial networks [12], or graph neural networks [15], among many others. However, these methods

typically assume that a large fraction of the field is observed and are primarily designed for super-resolution, in-painting, or mean-flow reconstruction tasks. Real-time full-field reconstruction from sparse measurements can alternatively be achieved by coupling reduced-order models with data assimilation [3, 13]. Vision Transformers (ViTs) [4], in particular, have become the state-of-the-art backbone for image generation and large language models [14], but have not been fully adopted yet in the fluid dynamics community. Here, we introduce the *Latent Attention on Masked Patches* (LAMP) model, inspired by ViTs and masked autoencoders [7]. The key idea is to (i) partition each flow snapshot into non-overlapping patches; (ii) compress each patch into a low-dimensional latent representation; here, patch-wise Proper Orthogonal Decomposition (POD) [2]; and (iii) reconstruct the full field from a subset of the patches, i.e., a masked input, using an attention-based patch-to-patch prediction mechanism; here, implemented as a single-layer transformer trained via closed-form linear regression. This design guarantees convergence to a global minimum, eliminates dependence on learning-rate schedules and random weight initialisation, and yields fast, reproducible and physically interpretable results. Further, the modular architecture allows expressive components to be incorporated where performance demands it—such as nonlinear autoencoders [6, 16] or multi-layer transformers. The paper is structured as follows. Section 2 details the proposed LAMP. Section 3 demonstrates LAMP on two canonical unsteady wakes: a laminar bluff-body wake and a chaotic wake over two adjacent cylinders. Section 4 closes the paper with conclusions and future work directions.

2 Latent Attention on Masked Patches (LAMP)

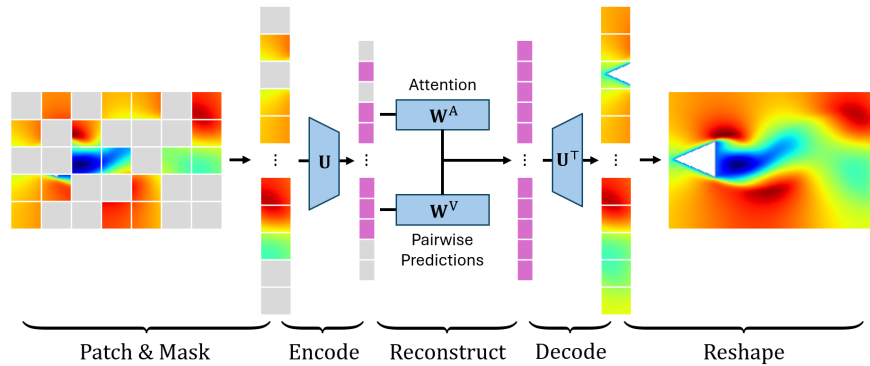


Fig. 1: Pictorial illustration of the patch-wise latent attention model (LAMP). A linear autoencoder embeds the input in a lower-dimensional latent space, where predictions can be made efficiently. The predictions are pair-wise (i.e. each patch predicts every other patch) and weighted with a confidence score.

As illustrated in Fig. 1, LAMP consists of three stages: (i) data patching and reshaping; (ii) patch-wise dimensionality reduction via linear autoencoding with proper orthogonal decomposition (POD); and (iii) attention-based patch-to-patch prediction in the latent space.

Patching. First, we divide the normalized input 2D data (of height H , width W and C components) $\mathbf{X}_{\text{in}} \in \mathbb{R}^{H \times W \times C}$ into N size- P square patches of dimension $D := CP^2$, which are flattened to obtain the patched data $\mathbf{X} \in \mathbb{R}^{N \times D}$ (see [4]).

Patch-wise POD. We apply dimensionality reduction to compress \mathbf{X} into $\mathbf{Z} \in \mathbb{R}^{N \times N_e}$, with embedding (latent) dimension $N_e \ll D$. Each patch vector $\mathbf{x}_n \in \mathbb{R}^D$ is compressed into $\mathbf{z}_n \in \mathbb{R}^{N_e}$ by minimizing

$$\mathcal{L}^{\text{AE}} = \sum_{n=1}^N \|\tilde{\mathbf{x}}_n - \mathbf{x}_n\|_2^2, \quad \text{subject to } \tilde{\mathbf{x}}_n = \mathbf{U}_n \mathbf{U}_n^\top \mathbf{x}_n. \quad (1)$$

The patch-wise optimal orthonormal bases $\mathbf{U}_n \in \mathbb{R}^{D \times N_e}$ are obtained from the training time series as $\mathbf{X}_n^{\text{train}} \approx \mathbf{U}_n \Sigma(\mathbf{V}_n^{\text{train}})^\top$, where the approximation arises as we retain only the first N_e leading singular vectors. Finally, we construct the encoding and decoding operators, \mathbf{E} and \mathbf{D} , as block-diagonal matrices composed of the patch-wise bases $\{\mathbf{U}_n\}_{n=1}^N$ such that

$$\mathbf{Z} = \mathbf{E}\mathbf{X}, \quad \text{and} \quad \tilde{\mathbf{X}} = \mathbf{D}\mathbf{Z}, \quad (2)$$

where the n -th block in \mathbf{E} and \mathbf{D} contain \mathbf{U}_n^\top and \mathbf{U}_n , respectively. The patched latent data \mathbf{Z} is then passed to the transformer. A low reconstruction loss \mathcal{L}^{AE} is a prerequisite for accurate masked predictions.

Masked prediction. We modify the attention equation presented in [18] for the task of single-layer reconstruction. The aim is to predict the full latent snapshot \mathbf{Z}^* from a masked input $\mathbf{Z}^{\text{masked}}$, where $Z_n^{\text{masked}} = 0$ for all masked indices $n \in \mathcal{N}_{\text{masked}}$. Each component in \mathbf{Z}^* is

$$Z_{me}^* = \sum_{n=1}^N \text{softmax}(\mathbf{a}_m)_n V_{mne}, \quad \text{for } \begin{cases} m = 1, \dots, N, \\ e = 1, \dots, N_e \end{cases} \quad (3)$$

where \mathbf{a}_m is the m -th row of the log-attention matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, and $\mathbf{V} \in \mathbb{R}^{N \times N \times N_e}$ is the value tensor; these are defined as

$$A_{mn} = \sum_{e=1}^{N_e} W_{mne}^{\text{A}} Z_{ne}^{\text{masked}}, \quad \text{and} \quad V_{mne} = \sum_{f=1}^{N_e} W_{mnef}^{\text{V}} Z_{nf}^{\text{masked}}, \quad (4)$$

where $\mathbf{W}^{\text{V}} \in \mathbb{R}^{N \times N \times N_e \times N_e}$ and $\mathbf{W}^{\text{A}} \in \mathbb{R}^{N \times N \times N_e}$ are the value and attention weight-tensors, which are trained on *unmasked* data with a two-step approach:

1. The value tensor \mathbf{W}^{V} can be viewed as an $N \times N$ grid of matrices: each block $\mathbf{W}_{mn}^{\text{V}} \in \mathbb{R}^{N_e \times N_e}$ defines a linear map that predicts patch \mathbf{z}_m from patch \mathbf{z}_n . These blocks are obtained by least-squares regression,

$$(\mathbf{Z}^{\text{train}})_m \approx \mathbf{W}_{mn}^{\text{V}} (\mathbf{Z}^{\text{train}})_n, \quad (5a)$$

where the corresponding linear prediction error is

$$\mathcal{L}_{mn}^{\text{lin}} = \|(\mathbf{Z}^{\text{train}})_m - \mathbf{W}_{mn}^{\text{V}}(\mathbf{Z}^{\text{train}})_n\|_2^2. \quad (5b)$$

2. The attention tensor \mathbf{W}^{A} is learned in a second stage and can be decomposed analogously, as a $N \times N$ grid of vectors $w_{mn}^{\text{A}} \in \mathbb{R}^{N_e}$. The error $\mathcal{L}_{mn}^{\text{lin}}$ provides a prediction uncertainty measure, so that predictions with lower $\mathcal{L}_{mn}^{\text{lin}}$ are assigned higher attention weights. In practice, we regress the *log-error*,

$$-\log(\mathcal{L}_{mn}^{\text{lin}}) \approx \mathbf{W}_{mn}^{\text{A}}(\mathbf{Z}^{\text{train}})_n. \quad (6)$$

At inference, the log-error is set to $-\infty$ for masked patches, so that softmax assigns them zero attention weight. The model is then evaluated on unseen test data \mathbf{X}^{test} . The overall prediction error is defined as

$$\mathcal{L}^{\text{pred}} = \|\tilde{\mathbf{X}}^* - \mathbf{X}^{\text{test}}\|_2^2 = \|\mathbf{D}\mathbf{Z}^* - \mathbf{X}^{\text{test}}\|_2^2. \quad (7)$$

3 Results

We deploy LAMP to reconstruct the streamwise u and spanwise v velocity fields of two 2D unsteady wakes. First, we test LAMP on the laminar wake of a triangular bluff body at Reynolds number $Re = 100$ [11] in §3.1 and analyse the effect of varying latent-space dimension, noise intensity, and patch placement. We use 100 snapshots for training and 60 for testing (one shedding period is 32 snapshots). Second, §3.2 showcases LAMP on a chaotic wake past two cylinders of diameter d with centres spaced $1.5d$ apart vertically. The flow has $Re = 200$, Mach number $M = 0.1$, uniform inlet velocity, convective outflow, and periodic side boundary conditions. After discarding transients, 7000 snapshots ($\Delta t = 1.39$) are split 75%/20% for training and testing, discarding the intermediate 5% to reduce temporal leakage. Both datasets are standardized to zero mean and unit variance for a fair comparison between them.

3.1 Laminar wake results

First, we validate the patch-wise POD and examine the loss \mathcal{L}^{AE} (1), which sets the lower bound for the masked reconstruction error. Figure 2 shows that (i) \mathcal{L}^{AE} saturates rapidly with N_e , which is consistent with the well-known energy concentration of laminar bluff-body wakes into a few dominant POD modes; and (ii) larger patches reach the same \mathcal{L}^{AE} at higher compression factors, which indicates that using very small patches may fragment the coherent vortex structures into features that carry little dynamical information.

Figure 3 shows the LAMP masked flow reconstruction results from randomly unmasked patches covering 10% of the domain for noise-free (Fig. 3a) and noisy input (Fig. 3b-c). Noise is added to the input as $\tilde{\mathbf{X}}^{\text{noisy}} = \tilde{\mathbf{X}}^{\text{masked}} + \boldsymbol{\varepsilon}$, where (\cdot) denotes unnormalized data, $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \mathbb{E}[\tilde{\mathbf{X}}^{\text{masked}}]10^{-\text{SNR}/10}$.

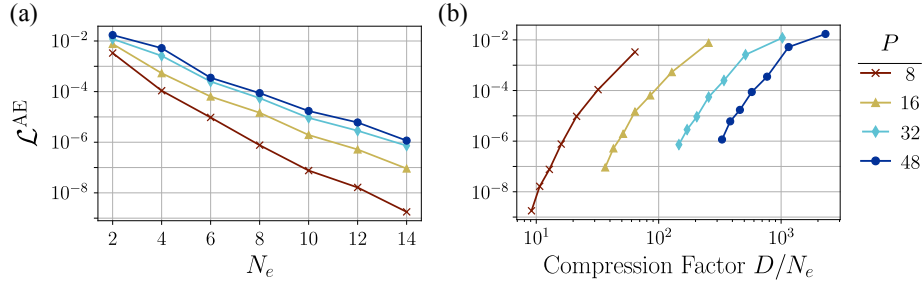


Fig. 2: Patch-wise POD reconstruction performance for varying patch size P . Loss \mathcal{L}^{AE} versus (a) latent dimension N_e and (b) compression factor D/N_e .

In the noisy cases, the LAMP provides flow-field reconstructions with prediction errors $\mathcal{L}^{\text{pred}}$ consistently lower than the variance of the added noise. Increasing the latent dimension N_e reduces $\mathcal{L}^{\text{pred}}$ (Fig. 3a), with larger patches outperforming smaller ones once sufficient modes are retained to resolve the coherent shedding. However, adding Gaussian noise shifts the $\mathcal{L}^{\text{pred}}$ minima toward lower N_e (Fig. 3b-d). Small patches are less robust because their local POD bases amplify sub-patch-scale noise. In contrast, larger patches act as spatial filters that preserve the dominant flow structures. Notably, for moderate SNR a broad range of (P, N_e) yields $\mathcal{L}^{\text{pred}}$ below the noise variance σ_{noise}^2 (horizontal dashed lines). Therefore, the proposed LAMP successfully achieves simultaneous noise removal and flow reconstruction from masked data.

Next, we focus on the attention matrix, which yields as a by-product an interpretable method for informing sensor placement: the predictive-power maps shown in Fig. 4, which are computed by averaging the columns of $-\log \mathcal{L}_{mn}^{\text{lin}}$ (6) and reshaping into the physical space. Patches with highest predictive power are those whose local dynamics most accurately predict the rest of the field, i.e., the optimal sensor placement for reconstruction. Beyond reducing the resolution, increasing the patch size predominantly reduces the range of values in the attention map. By contrast, as N_e increases, high-power regions multiply from 1-2 near-body patches to distributed coverage across shear layers and far wake, mirroring the POD hierarchy. Placing sensors at persistent predictive-power peaks across N_e thus provides optimal multi-fidelity sensor layouts for both coarse and fine robust reconstructions.

3.2 Chaotic wake results

Because of the chaotic nature of the flow, the two-cylinder wake requires larger patches and more POD modes compared to the laminar case, as reflected in the prediction loss in Fig. 5a. The requirement for larger patches to minimise $\mathcal{L}^{\text{pred}}$ indicates that each patch must span a sufficient spatial extent to capture the dominant coherent structures because the relevant dynamical correlations operate over larger spatial scales than in the laminar case.

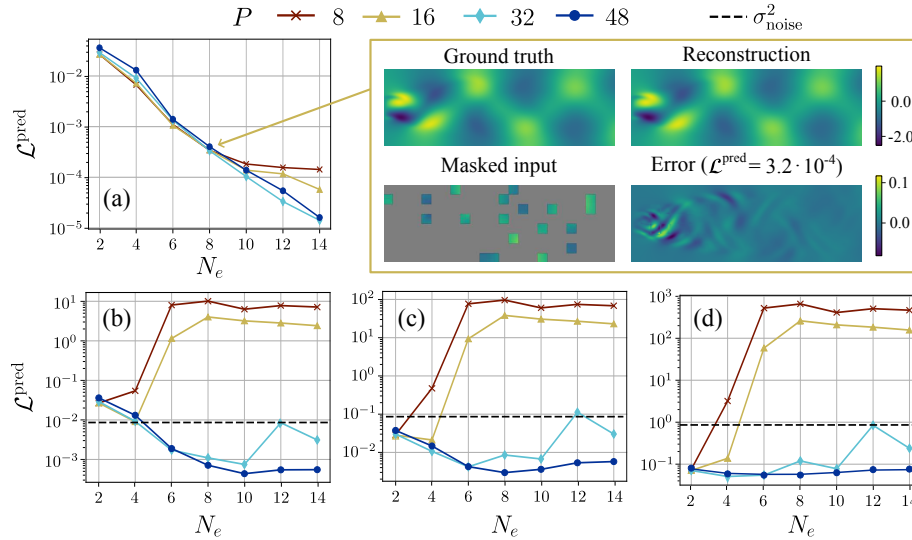


Fig. 3: Reconstruction error $\mathcal{L}^{\text{pred}}$ for varying noise levels, patch size P and latent dimension N_e , all with 10% patches unmasked. The input is noisy masked data with SNR = (a) ∞ (noise-free), (b) 30 dB, (c) 20 dB, (d) 10 dB. The loss is computed against the noise-free target \mathbf{X}^{test} and the model is trained on noise-free data. Horizontal dashed lines show the noise variance. $\mathcal{L}^{\text{pred}}$ is calculated for 25 random patch arrangements and the *median* is plotted. An example reconstruction for the noise-free case with $P = 16$, $N_e = 8$ is shown.

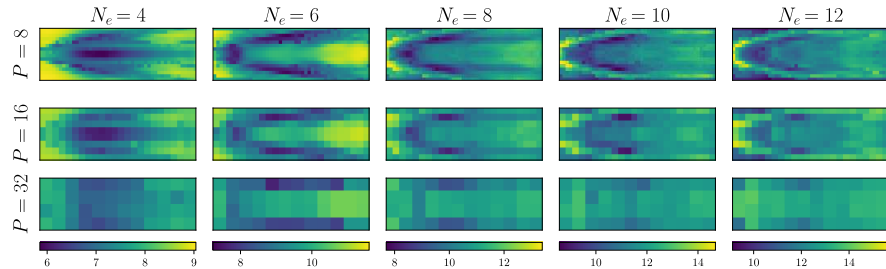


Fig. 4: Predictive-power maps for varying P and N_e . Patches with higher predictive power predict the rest of the patches with lower average loss.

Despite the increased reconstruction loss compared to dataset 1, LAMP achieves a 26% lower reconstruction error $\mathcal{L}^{\text{pred}}$ than gappy POD [5] on identical inputs (results not shown for brevity). Qualitatively, gappy POD tends to overestimate velocity fluctuations in regions of greater uncertainty further downstream, whereas LAMP produces smoother reconstructions. LAMP’s prediction accuracy is highest immediately downstream of unmasked patches and degrades further away. This is due to the spatio-temporally chaotic nature of the flow;

two patches become less correlated with each other the further apart they are. Extending to LAMP with *multi-layer* transformers is required to predict nonlinear relationships between multiple patches. Additionally, time-delay embedding [17] could improve accuracy by exploiting temporal correlations.

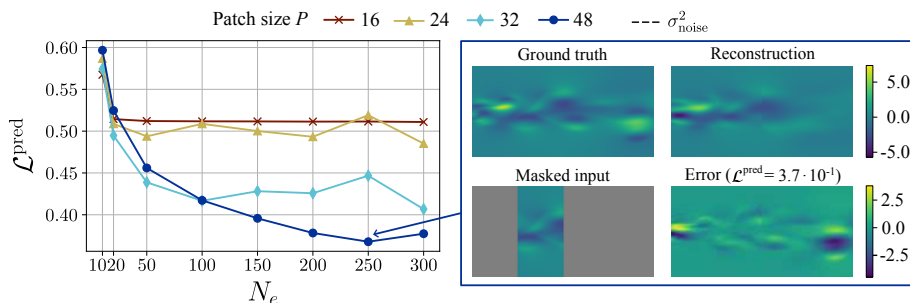


Fig. 5: Masked reconstruction of the chaotic wake from 25% unmasked patches, patches placed at the locations of highest predictive power. The reconstruction of one snapshot with the best model ($P = 48, N_e = 250$) is shown on the right.

4 Conclusion and future work

We introduce the *Latent Attention on Masked Patches* (LAMP) model, which consists of a regression-based architecture inspired by Vision Transformers. LAMP is designed to reconstruct high-dimensional flow fields from noisy and masked measurements. By integrating patch-wise POD with a linearly decomposable attention mechanism, we provide a framework that avoids extensive hyperparameter tuning and back-propagation. We showcase the LAMP on two test cases: a laminar wake past a triangular bluff body and a chaotic wake past two cylinders. On the laminar wake, LAMP successfully reconstructs flow fields from only 10% unmasked patches even when the input patches are corrupted by noise levels up to $\text{SNR} = 10$ dB. The reconstruction error remains an order of magnitude lower than the input noise variance, which shows that LAMP is both a reconstruction and de-noising tool. Beyond reconstruction, the attention maps from the LAMP provide an interpretable multi-fidelity sensor placement approach, which identifies the patches of highest importance for reconstruction. On the two-cylinder wake, LAMP outperforms gappy POD with 26% lower $\mathcal{L}^{\text{pred}}$. The overall error is, however, larger than the laminar case, suggesting that multi-layer attention blocks and nonlinear autoencoding might be required to reconstruct higher-dimensional, spatio-temporally chaotic flows. The modular architecture of the framework provides a natural pathway to more expressive extensions.

Acknowledgments. LM acknowledges funding from ERC Starting Grant PhyCo 949388. AN is supported by the Eric and Wendy Schmidt AI in Science Fellowship.

References

1. Bekoglu, E., Bempedelis, N., Steiros, K.: Formation of turbulent secondary vortex street in absence of vortex shedding instability. *Journal of Fluid Mechanics* **1024**, A7 (2025)
2. Berkooz, G., et al.: The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics* **25**, 539–575 (1993)
3. Cheng, S., Liu, C., Guo, Y., Arcucci, R.: Efficient deep data assimilation with sparse observations and time-varying sensors. *Journal of Computational Physics* **496**, 112581 (2024)
4. Dosovitskiy, A.: An image is worth 16×16 words: Transformers for image recognition at scale. In: *Proc. of IEEE/CVF CVPR*. pp. 45–67 (2021)
5. Everson, R., Sirovich, L.: Karhunen–loève procedure for gappy data. *Journal of the Optical Society of America A* **12**(8), 1657–1664 (1995)
6. Fukami, K., Fukagata, K., Taira, K.: Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics* **870**, 106–120 (2019)
7. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: *Proc. of IEEE/CVF CVPR*. pp. 16000–16009 (2022)
8. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: Learning skillful medium-range global weather forecasting. *Science* **382**(6677), 1416–1421 (2023)
9. Lee, Y., Yang, H., Yin, Z.: Piv-dcnn: cascaded deep convolutional neural networks for particle image velocimetry. *Experiments in Fluids* **58**(12), 171 (2017)
10. Mo, Y., Magri, L.: Reconstruction of three-dimensional turbulent flows from sparse and noisy planar measurements: a weight-sharing neural network approach. *Data-Centric Engineering* **7**, e5 (2026)
11. Mo, Y., Traverso, T., Magri, L.: Decoder decomposition for the analysis of the latent space of nonlinear autoencoders with wind-tunnel experimental data. *Data-Centric Engineering* **5**, e38 (2024)
12. Nista, L., Pitsch, H., Schumann, C.D., Bode, M., Grenga, T., MacArt, J.F., Attili, A.: Influence of adversarial training on super-resolution turbulence reconstruction. *Physical Review Fluids* **9**(6), 064601 (2024)
13. Ózalp, E., Nóvoa, A., Magri, L.: Real-time forecasting of chaotic dynamics from sparse data and autoencoders. *Computer Methods in Applied Mechanics and Engineering* **450**, 118600 (2026)
14. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: *Proc. of IEEE/CVF ICCV*. pp. 4172–4182 (2023)
15. Quattromini, M., Bucci, M.A., Cherubini, S., Semeraro, O.: Mean flow data assimilation using physics-constrained graph neural networks. *Data-Centric Engineering* **6**, e48 (2025)
16. Racca, A., Doan, N.A.K., Magri, L.: Predicting turbulent dynamics with the convolutional autoencoder echo state network. *Journal of Fluid Mechanics* **975**, A2 (2023)
17. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. *Journal of Statistical Physics* **65**(3–4), 579–616 (1991)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
19. Xia, C., Zhang, J., Kerrigan, E.C., Rigas, G.: Active flow control for bluff body drag reduction using reinforcement learning with partial measurements. *Journal of Fluid Mechanics* **981**, A17 (2024)