Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learning Nonlinear Reduced-Order Models of Fluid Flows with InfoVAE

Manal ZETTAM

m.zettam@uca.ac.ma

Cadi Ayyad University, UCA, National School of Applied Sciences of Safi

Mouhcine MOUDA

Cadi Ayyad University, UCA, National School of Applied Sciences of Safi

Mohammed MADIAFI

Cadi Ayyad University, UCA, National School of Applied Sciences of Safi

Omar ASKOUR

Cadi Ayyad University, UCA, National School of Applied Sciences of Safi

Mustapha OUJAOURA

Cadi Ayyad University, UCA, National School of Applied Sciences of Safi

Jamal EZZAHAR

Cadi Ayyad University, UCA, National School of Applied Sciences of Safi

Research Article

Keywords: Autoencoder, Information Maximizing Variational Autoencoder, Time Series Predictions, Proper Orthogonal Decomposition, Reduced-Order Model

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10594643/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Learning Nonlinear Reduced-Order Models of Fluid Flows with InfoVAE

Manal ZETTAM[1*], Mouhcine MOUDA[2], Mohammed MADIAFI[1], Omar ASKOUR[1], Mustapha OUJAOURA[1], Jamal EZZAHAR[1]

[1*]Cadi Ayyad University, UCA, National School of Applied Sciences of Safi,, Mathematics Computer Science and Communication Systems Laboratory, Route Sidi Bouzid, Safi, BP 63, 46000, Marrakech-Safi, Morocco.

[2]Cadi Ayyad University, UCA, National School of Applied Sciences of Safi,, Materials Processes Environment and Quality Laboratory, Route Sidi Bouzid, Safi, BP 63, 46000, Marrakech-Safi, Morocco.

*Corresponding author(s). E-mail(s): m.zettam@uca.ac.ma;

Abstract

Reduced-order modeling of chaotic fluid flows using autoencoders is a promising approach. However, conventional variational autoencoders often suffer from posterior collapse and a trade-off between reconstruction accuracy and latent regularization. To overcome these limitations, we employ an Information Maximizing Variational Autoencoder (InfoVAE), whose more flexible objective function promotes informative latent representations while preserving reconstruction quality. In this paper, we combine the Information Maximizing Variational Autoencoder (InfoVAE) with an Easy-Attention-based Transformer to develop an efficient reduced-order modeling framework for fluid flow prediction. The proposed approach was evaluated on numerical datasets of two-dimensional viscous flows under both periodic and chaotic conditions. The learned disentangled latent representations provide an interpretable reduced-order model, with features that are consistent with the dominant modes identified by Proper Orthogonal Decomposition (POD). By applying InfoVAE to extract disentangled representations, an interpretable flow model was achieved. The features emerging from this model are similar to those observed in proper orthogonal decomposition.

Keywords: Autoencoder , Information Maximizing Variational Autoencoder , Time Series Predictions, Proper Orthogonal Decomposition, Reduced-Order Model

1 Introduction

Turbulent flows are commonly observed in everyday situations, such as water flowing in rivers or waterfalls, smoke rising from a chimney, or the action of strong winds [1]. These flows are characterized by irregular, unsteady, and seemingly random motion, making the behavior of individual fluid structures difficult to predict [1]. A defining feature of turbulence is its multiscale nature, where large flow structures coexist and continuously interact with much smaller eddies through nonlinear mechanisms. This broad range of spatial and temporal scales, together with the large number of degrees of freedom, makes turbulent flows particularly challenging to analyze and model [2],[3]. Despite this complexity, turbulent flows play a fundamental role in both natural phenomena and engineering applications, including aircraft design and weather forecasting [2]. Consequently, significant research efforts have been devoted to developing reduced-order models (ROMs) capable of capturing the essential flow dynamics while substantially reducing the dimensionality and computational cost of the problem [2],[4],[5].

More detailed observations of turbulent flows can be obtained through laboratory experiments, which reveal that turbulence is characterized by highly irregular concentration and velocity fields spanning a broad range of spatial scales. These observations confirm that turbulent motions exhibit strong variations in both space and time, with large coherent structures interacting continuously with much smaller eddies through nonlinear energy transfer mechanisms [1]. The instantaneous velocity field is commonly denoted by $U(x,t)$, where x represents the spatial position and t the time.

Fluid flows, whether laminar or turbulent, are governed by the Navier–Stokes equations, which accurately describe fluid motion over a wide range of flow regimes [1]. However, for turbulent flows, these equations resolve all spatial and temporal scales, from the largest coherent structures to the smallest dissipative eddies, resulting in an extremely high-dimensional dynamical system [1]. The numerical solution of these equations constitutes the foundation of Computational Fluid Dynamics (CFD), whose governing equations for incompressible flows are

$$\begin{cases} \dfrac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\dfrac{1}{\rho}\nabla p + \nu\nabla^2\mathbf{u}, \\ \nabla\cdot\mathbf{u} = 0. \end{cases} \tag{1}$$

where $\mathbf{u}$ denotes the velocity field, p the pressure, ρ the fluid density, and ν the kinematic viscosity. The nonlinear convective term $(\mathbf{u}\cdot\nabla)\mathbf{u}$ is responsible for the multiscale interactions and chaotic dynamics that characterize turbulent flows [1], [6].

Although Direct Numerical Simulation (DNS) provides an exact numerical solution of the Navier–Stokes equations without turbulence modeling, its computational cost increases dramatically with the Reynolds number. In three-dimensional turbulent flows, the required number of grid points scales approximately as $N \sim \mathrm{Re}^{9/4}$, making DNS impractical for most engineering applications even on modern supercomputers [7], [8]. Consequently, alternative modeling strategies have been developed to reduce computational complexity. Reynolds-Averaged Navier–Stokes (RANS) models employ statistical averaging together with turbulence closure models, whereas Large Eddy

Simulation (LES) explicitly resolves the energy-containing structures while modeling the unresolved subgrid scales. Hybrid approaches such as Detached Eddy Simulation (DES) combine the computational efficiency of RANS with the improved accuracy of LES [8], [9], [10], [11]. Although these approaches make turbulent flow simulations computationally feasible, they inevitably introduce modeling errors and uncertainties, particularly in flows involving separation, recirculation, or strong anisotropy [9], [10].

To further reduce computational cost while preserving the dominant flow dynamics, Reduced-Order Models (ROMs) have become an important research direction. Their objective is to construct low-dimensional representations that enable efficient flow prediction, control, and optimization while significantly reducing the computational expense of high-fidelity CFD simulations [2], [5], [12], [13]. Among the most widely used projection-based techniques is Proper Orthogonal Decomposition (POD), which constructs an optimal orthogonal basis from flow snapshots by extracting the most energetic coherent structures of the system [2]. Another popular linear approach is Dynamic Mode Decomposition (DMD), which identifies coherent modes associated with the temporal evolution of the flow [14]. Although POD and DMD have successfully reduced the complexity of many flow problems [15], their reliance on linear modal representations limits their ability to capture the strongly nonlinear and advection-dominated dynamics typical of turbulent flows [16], [17].

To overcome these limitations, machine learning (ML) and data-driven techniques have emerged as powerful alternatives for turbulence modeling and reduced-order modeling [18], [13], [19], as well as for spatio-temporal flow estimation [20], [21], [22]. These developments have led to Machine Learning Reduced-Order Models (ML-ROMs), in which data-driven models complement or replace traditional projection-based methods. For example, San and Maulik [23] proposed an extreme learning machine framework that outperformed conventional POD-ROMs in quasi-stationary geophysical turbulence. Maulik et al. [24] introduced a probabilistic neural network to predict the temporal evolution of POD coefficients from initial conditions, while Srinivasan et al. [25] demonstrated that Long Short-Term Memory (LSTM) networks accurately capture the temporal dynamics of nonlinear shear-flow systems. These studies illustrate the growing integration of machine learning with classical reduced-order methodologies.

More recently, autoencoder-based architectures have emerged as powerful nonlinear alternatives for dimensionality reduction in fluid mechanics. Milano and Koumoutsakos [26] were among the first to demonstrate that nonlinear autoencoders outperform POD for Burgers' equation and turbulent channel flows, while linear autoencoders reproduce POD behavior [27]. Subsequently, convolutional neural network (CNN)-based autoencoders have become increasingly popular because of their ability to efficiently extract hierarchical spatial features from high-dimensional flow fields [28]. Numerous studies have demonstrated their effectiveness in reduced-order modeling, including the hybrid β-VAE and Transformer framework proposed by Solera-Rico et al. [2], CNN autoencoders for wake-flow analysis [29], CNN-LSTM architectures for predicting unsteady bluff-body flows [30], [31], [32], mode-decomposing CNN autoencoders [33], and nonlinear autoencoder-based ROMs for laminar and turbulent Navier–Stokes flows [34]. Comprehensive reviews of these approaches can be found

in [35], [36], [37], [38], and [39]. Despite these significant advances, developing robust nonlinear reduced-order models capable of accurately representing highly complex turbulent flow topologies while maintaining computational efficiency remains an open research challenge.

The main contributions of this paper are:

- proposing an Information Maximizing Variational Autoencoder (InfoVAE) framework for nonlinear reduced-order modeling of fluid flows;
- comparing the proposed InfoVAE approach with the classical Proper Orthogonal Decomposition (POD) in terms of reconstruction accuracy, latent dimension sensitivity, and captured energy content;
- investigating the ability of InfoVAE to learn nonlinear latent representations capable of capturing complex flow structures beyond the linear subspaces of POD;
- evaluating the effectiveness of the proposed framework on representative fluid flow datasets with different dynamical behaviors.

The remainder of this paper is organized as follows. Section 2 introduces the Information Maximizing Variational Autoencoder (InfoVAE) framework and its adaptation to fluid dynamics. Section 3 describes the dataset used in this study. Section 4 presents the hyperparameters adopted to implement the proposed InfoVAE–easy-attention Transformer framework. Section 5 reports and discusses the experimental results. Finally, Section 6 concludes the paper, summarizes the main findings, and outlines directions for future work.

2 The Information Maximizing Variational Autoencoder

The Information Maximizing Variational Autoencoder framework was first introduced by Shengjia Zhao, Jiaming Song, and Stefano Ermon in 2017 [40]. The Information Maximizing Variational Autoencoder framework has gained significant attention as an effective extension of the standard Variational Autoencoder (VAE). The Information Maximizing Variational Autoencoder (InfoVAE) was designed to address key limitations of VAEs. In particular, posterior collapse and the trade-off between reconstruction quality and latent regularization. Compared to traditional VAEs, InfoVAE has a more flexible objective function. This objective function incorporates alternative divergence measures such as the Maximum Mean Discrepancy (MMD). Thus, InfoVAE enables better control over the information retained in the latent variables [40]. InfoVAE preserves meaningful latent representations while maintaining generative performance, which has contributed to its adoption in representation learning and generative modeling tasks [41], [42]. Furthermore, InfoVAE has demonstrated strong empirical success in a variety of applications including image generation and scientific data modeling [41], [42]. Its ability to learn informative and disentangled latent spaces has made it a competitive alternative to classical VAE variants and other generative models. The remainder of this section will present the Variational Autoencoder framework and its limitations, followed by the formulation of the Information Maximizing Variational Autoencoder (InfoVAE). We first review the conventional ELBO

objective used in variational autoencoders and discuss the challenges associated with posterior collapse, which can limit the informativeness of the learned latent representations. Next, we introduce the aggregated posterior framework and describe how InfoVAE extends the standard objective by explicitly controlling the trade-off between reconstruction fidelity, regularization, and information preservation. We conclude by examining the benefits of this formulation for representation learning and generative modeling.

The Variational Autoencoder (VAE) is a probabilistic generative model that combines deep neural networks with variational inference to learn latent representations of data. The model assumes that observed data x is generated from latent variables z through a generative process defined as:

$$p_\theta(x, z) = p(z)\, p_\theta(x \mid z) \tag{2}$$

where $p(z)$ is typically a standard multivariate Gaussian distribution. Since the true posterior distribution $p_\theta(x \mid z)$ is intractable, it is approximated using a variational distribution $q_\phi(z \mid x)$, parameterized by an encoder network.

The model is trained by maximizing the **E**vidence **L**ower **Bo**und (ELBO),which is defined as follows:

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x \mid z)\right] - D_{\text{KL}} \left(q_\phi(z \mid x) \,\|\, p(z)\right). \tag{3}$$

The first term of $\mathcal{L}_{\text{ELBO}}$ denoted $\mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x \mid z)\right]$ corresponds to the reconstruction log-likelihood. $\mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x \mid z)\right]$ encourages the decoder to reconstruct the input data from the latent representation in accurate way. The second term of $\mathcal{L}_{\text{ELBO}}$ denoted $D_{\text{KL}} \left(q_\phi(z \mid x) \,\|\, p(z)\right)$ is named Kullback–Leibler divergence(kL divergence). The Kullback–Leibler divergence regularizes the approximate posterior to remain close to the prior distribution.

The ELBO objective presents several limitations in practice. One of the issues is posterior collapse, which occurs when the decoder becomes overly expressive and learns to reconstruct the input independently of the latent variable. In this case, the variational posterior collapses to the prior:

$$q_\phi(z \mid x) \approx p(z). \tag{4}$$

As a result, the latent variable z carries little or no information about the input x. This leads to severely limiting the representation learning capability of the model. From an information-theoretic perspective, the mutual information between the observed variable x and the latent variable z under the variational distribution is defined as:

$$I_q(x; z) = \mathbb{E}_{p_{\text{data}}(x)} \left[D_{\text{KL}} \left(q_\phi(z|x) \,\|\, q_\phi(z)\right)\right], \tag{5}$$

The expectation over the data distribution $p_{\text{data}}(x)$ is essential because mutual information measures the *average* amount of information that the latent variable z retains about the input x across the entire dataset, rather than for an individual sample. Consequently, $I_q(x; z)$ quantifies how informative the learned latent representation is with respect to the input data.

It is important to state that maximizing the ELBO does not necessarily produce an accurate approximation of the true posterior $p_\theta(x|z)$ by the amortized inference model $q_\phi(z|x)$. This is another limitation of ELBO objective commonly named Poor amortized inference.

To resolve the issues of posterior collapse and of Poor amortized inference, the Information Maximizing Variational Autoencoder (InfoVAE) explicitly controls the information flow between input data and latent variables. A key concept in InfoVAE model is the aggregated posterior, which defined as:

$$q_\phi(z) = \int q_\phi(z \mid x)\, p_{\text{data}}(x)\, dx. \tag{6}$$

In contrast with Standard VAEs that only impose constraints on the conditional posterior $q_\phi(z|x)$, InfoVAE introduces an additional constraint on the aggregated posterior$q_\phi(z)$. This additional constraint enables better global matching with the prior distribution.

The InfoVAE objective function is defined as:

$$\mathcal{L}_{\text{InfoVAE}} = \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x \mid z)\right] - \alpha D_{\text{KL}}\left(q_\phi(z \mid x) \,\|\, p(z)\right) - \lambda D\left(q_\phi(z) \,\|\, p(z)\right), \tag{7}$$

where α and λ are positive hyperparameters that control the trade-off between reconstruction quality, local posterior regularization, and global distribution matching.

Equation 7 can also be interpreted from an information-theoretic perspective as:

$$\mathcal{L}_{\text{InfoVAE}} = \mathcal{L}_{\text{ELBO}} + (1-\alpha)\, I_q(x; z) - \lambda D\left(q_\phi(z) \,\|\, p(z)\right), \tag{8}$$

where $I_q(x; z)$ denotes the mutual information between the input data x and the latent variable z. When $\alpha < 1$, the objective encourages the latent representation to preserve more information about the input data, thereby mitigating posterior collapse and improving the quality of the learned representations.

$\mathcal{L}_{\text{InfoVAE}}$ provides explicit control over the mutual information between the observed data and latent variables. Another important feature of InfoVAE is its flexibility in choosing the divergence function $\lambda D\left(q_\phi(z) \,\|\, p(z)\right)$. In contrast to standard VAEs, which rely exclusively on the Kullback–Leibler divergence, InfoVAE can employ alternative measures such as Maximum Mean Discrepancy (MMD). This kernel-based metric compares distributions in a reproducing kernel Hilbert space and has been shown to provide more stable optimization and improved empirical performance, particularly in high-dimensional settings.

Overall, InfoVAE generalizes the classical VAE framework by decoupling reconstruction, regularization, and information preservation. By explicitly controlling mutual information and allowing flexible divergence measures, it overcomes key limitations of standard VAEs, particularly posterior collapse, and enables the learning of more informative and meaningful latent representations. This makes InfoVAE particularly suitable for applications in representation learning, semi-supervised learning, and downstream predictive modeling tasks. The overall architecture is illustrated in Fig. 1.

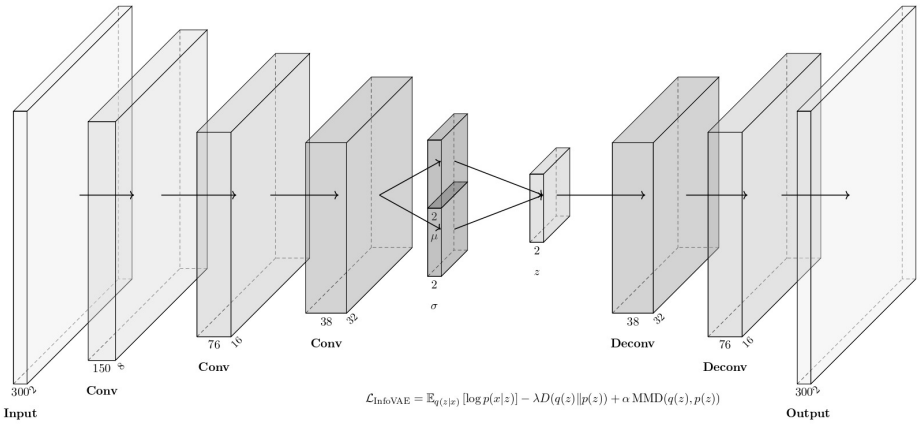


Fig. 1 Architecture of the InfoVAE framework.

3 Data description

All datasets used in this study are publicly available on Zenodo at https://doi.org/10.5281/zenodo.10501215. Two benchmark cases are considered to evaluate the proposed methodology: a periodic flow at (Re=40) and a chaotic flow at (Re=100). Their main characteristics are summarized below.

- **Experiment 1**: A periodic flow with Reynolds number $Re = 40$, defined using the freestream velocity U_∞ and the single-plate chord length c. The two collinear plates are arranged at an angle of 90° with respect to the incoming flow, as illustrated in Fig. 2 [2]. This dataset contains 1,001 temporal snapshots of the velocity field.
- **Experiment 2**: A chaotic flow with Reynolds number $Re = 100$, where the two collinear plates are arranged at an angle of 80° with respect to the incoming flow, as shown in Fig. 3 [2]. This dataset contains 150,000 temporal snapshots of the velocity field.

The two benchmark cases were deliberately selected to evaluate the proposed framework under flow regimes with markedly different dynamical characteristics. Experiment 1 ($Re = 40$) corresponds to a periodic flow exhibiting regular vortex shedding and relatively simple, low-dimensional dynamics. This case provides an appropriate benchmark for assessing the capability of the proposed reduced-order model to learn compact latent representations and accurately reconstruct periodic flow structures. In contrast, Experiment 2 ($Re = 100$) represents a chaotic flow characterized by complex nonlinear interactions and irregular spatio-temporal behavior. The increased flow complexity, together with the substantially larger number of available snapshots, makes this case considerably more challenging for both reduced-order modeling and temporal prediction. By evaluating the proposed approach on both periodic and chaotic flow regimes, the study demonstrates its robustness, scalability, and ability to generalize across different levels of flow complexity.

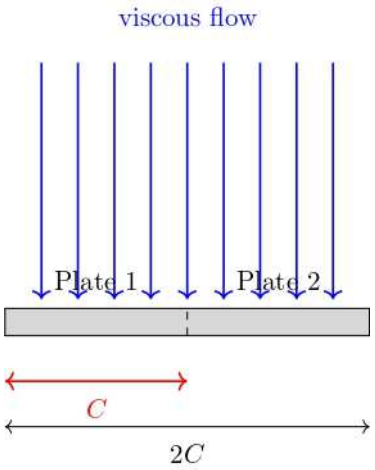


Fig. 2 Experimental configuration of Experiment 1 ($Re = 40$).

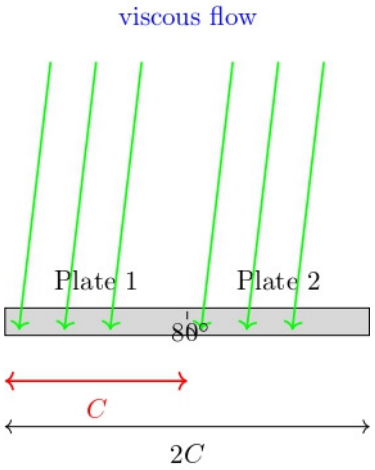


Fig. 3 Experimental configuration of Experiment 2 ($Re = 100$).

For both experiments, the computational domain spans $(96c \times 28c)$ and is discretized on a uniform grid of 300×98 points [2]. The temporal resolution corresponds to one convective time unit, $\Delta t = c/U_\infty = t_c$, for the periodic case and $\Delta t = c/(5U_\infty) = t_c/5$ for the chaotic case [2]. The released datasets are obtained by spatially and temporally downsampling the original high-fidelity numerical simulations [2].

The data are stored in HDF5 format and comprise three arrays: `UV`, `mean`, and `std`. The `UV` array contains the instantaneous two-dimensional velocity fields. Its dimensions are $(1001, 88, 300, 2)$ for the $Re = 40$ dataset and $(150000, 88, 300, 2)$ for the $Re = 100$ dataset, where the first dimension denotes the number of temporal snapshots, the second and third dimensions correspond to the spatial grid, and the last dimension stores the two velocity components, namely the horizontal (U) and vertical (V) velocities.

The `mean` array, of dimensions $(1, 88, 300, 2)$, stores the temporal mean velocity field evaluated at every grid point over the complete dataset, whereas the `std` array, with dimensions $(1, 1, 1, 2)$, contains the global standard deviations of the (U) and (V) velocity components.

Following the definition introduced by Sirovich in the snapshot Proper Orthogonal Decomposition (POD) method [43], a *snapshot* represents the complete state of the flow at a given instant. Accordingly, each snapshot contains the values of both velocity components at every point of the computational grid, providing a complete spatial description of the flow at a particular time. For example, `UV[0]` denotes the first snapshot at time t_0, $UV[1]$ the second snapshot at time t_1, and so forth until the final snapshot.

For visualization and post-processing purposes, each grid point can be represented as one row in a CSV or TXT file. Consequently, a single snapshot defined on an 88×300 grid contains $88 \times 300 = 26\,400$ rows. Each row stores the instantaneous velocity components (U and V), the corresponding temporal mean values (meanU and meanV), the global standard deviations (stdU and stdV), together with the associated snapshot and time indices. Table 1 presents a representative excerpt extracted from snapshot 19 (time index $t = 19$). Each row represents one spatial location of the computational grid at the selected time instant. The associated columns contain the instantaneous velocity components (U and V), the temporal mean velocity at that grid point (meanU and meanV), the global standard deviations (stdU and stdV), as well as the snapshot and time indices.

Table 1 Representative rows extracted from snapshot 19 of the dataset.

	Velocity Components		Statistical Parameters				Snapshot Information	
Row	**U**	**V**	**meanU**	**meanV**	**stdU**	**stdV**	**Snapshot**	**Time**
1	8.860174E-05	-5.769237E-04	1.0006012	-5.9348415E-04	0.21258125	0.04647906	19	19
2	6.6619535E-04	-5.779619E-04	1.0029584	-5.9441590E-04	0.21258125	0.04647906	19	19
3	1.1591126E-03	-5.803539E-04	1.0053436	-5.9657970E-04	0.21258125	0.04647906	19	19
4	1.6626845E-03	-5.840696E-04	1.0077366	-5.9995200E-04	0.21258125	0.04647906	19	19
5	2.1757896E-03	-5.890651E-04	1.0101475	-6.0449380E-04	0.21258125	0.04647906	19	19
6	2.6827260E-03	-5.952780E-04	1.0125797	-6.1015100E-04	0.21258125	0.04647906	19	19
7	3.1806900E-03	-6.026493E-04	1.0150338	-6.1685840E-04	0.21258125	0.04647906	19	19
8	3.7442646E-03	-6.110062E-04	1.0175140	-6.2453600E-04	0.21258125	0.04647906	19	19
9	4.2781183E-03	-6.202735E-04	1.0200305	-6.3308830E-04	0.21258125	0.04647906	19	19
10	4.7979522E-03	-6.302847E-04	1.0225800	-6.4240900E-04	0.21258125	0.04647906	19	19

4 Implementation Details

In this section, the InfoVAE architecture is applied to two flow configurations obtained from numerical simulations of a two-dimensional viscous flow over two collinear flat plates. The experiments were performed on the Marwan HPC cluster, with 64 GB of RAM and 200 GB of memory storage. The experiments were conducted using a modified version of the implementation publicly available at https://github.com/KTHFlowAI/beta-Variational-autoencoders-and-transformers-for-reducedorder-modelling-of-fluid-flows. Several modifications and extensions were introduced to adapt the code to the proposed methodology and to perform the experiments presented in this work. The modified source code developed for this study will be made publicly available upon acceptance of the manuscript.

The proposed InfoVAE is based on a convolutional encoder–decoder architecture. The encoder comprises six convolutional layers with 8, 16, 32, 64, 128, and 256 feature maps, respectively. Each convolutional layer employs a 3×3 kernel with a stride of 2 and a padding of 1, followed by an Exponential Linear Unit (ELU) activation function. The extracted feature maps are subsequently flattened and processed through a fully connected layer with 256 neurons before being projected onto the latent space to estimate the mean and log-variance vectors. The decoder mirrors the encoder through two fully connected layers followed by six transposed convolutional layers, reconstructing the original two-channel flow field. No activation function is applied to the output layer.

The latent variables are generated using the standard reparameterization trick by sampling Gaussian noise from a standard normal distribution. The model is optimized by minimizing the mean squared reconstruction error (MSE) together with Kullback–Leibler (KL) divergence and Maximum Mean Discrepancy (MMD) regularization terms, yielding the following InfoVAE objective:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \alpha \mathcal{L}_{\text{KL}} + \lambda \mathcal{L}_{\text{MMD}}, \tag{9}$$

The latent dimension was selected according to the complexity of the flow. For the periodic flow at $Re = 40$, the POD analysis indicates that the first two modes capture the dominant flow dynamics. Consequently, a two-dimensional latent space ($d = 2$) was adopted. In contrast, the chaotic flow at $Re = 100$ exhibits significantly more complex dynamics, requiring a richer latent representation. Therefore, a latent dimension of $d = 6$ was employed, consistent with previous reduced-order modeling studies reported in the literature.

The model was trained for 1000 epochs using the Adam optimizer with a batch size of 256. The initial learning rate was set to 2×10^{-4} and progressively reduced to 1×10^{-5} using a OneCycle learning-rate scheduler. The InfoVAE hyperparameters were fixed to $\alpha = 0.01$ and $\lambda = 10$ after being selected empirically through a series of preliminary experiments.. No weight decay was applied to either the encoder or the decoder, and the default PyTorch parameter initialization was retained. All experiments were implemented in PyTorch and executed on the CPU computing nodes of the Marwan High Performance Computing (HPC) center. The network architecture, training configuration, and implementation hyperparameters are summarized in Table 2, providing all the information required to reproduce the proposed InfoVAE framework.

Table 2 Implementation details and hyperparameters of the proposed InfoVAE model.

Parameter	**Value**
Encoder architecture	6 convolutional layers
Decoder architecture	6 transposed convolutional layers
Convolution kernel size	3×3
Stride	2
Padding	1
Activation function	ELU
Fully connected layer	256 neurons
Latent dimension	$d = 2$ (Re = 40) or $d = 6$ (Re = 100)[1]
Reconstruction loss	Mean Squared Error (MSE)
Latent regularization	KL divergence + Maximum Mean Discrepancy (MMD)
Optimizer	Adam
Batch size	256
Initial learning rate	2×10^{-4}
Final learning rate	1×10^{-5}
Learning-rate scheduler	OneCycleLR
Training epochs	1000
Weight decay	0
α	0.01
λ	10
Weight initialization	Default PyTorch initialization
Framework	PyTorch
Computing platform	Marwan High Performance Computing (HPC)[2]

[1]The latent dimension was selected according to the flow complexity: $d = 2$ for the periodic flow ($Re = 40$) and $d = 6$ for the chaotic flow ($Re = 100$).

[2]Training was performed on CPU nodes of the Marwan High Performance Computing (HPC) center.

5 Results & Discussion

This section presents the evaluation of the proposed reduced-order modeling framework. The analysis begins with a proof-of-concept study on the periodic flow at Re=40, where the latent representation learned by the InfoVAE is compared with Proper Orthogonal Decomposition (POD) in terms of reconstruction energy, latent-variable correlation, spatial structures, and temporal dynamics. The framework is then coupled with an easy-attention Transformer for latent-space prediction and subsequently assessed on a more challenging nonlinear flow at Re=100.

Unlike classification tasks, where the objective is to assign samples to discrete classes and model performance is commonly evaluated using metrics such as precision, recall, and F1-score, the proposed framework addresses a continuous reconstruction problem in which the autoencoder learns a low-dimensional representation of high-dimensional velocity fields. Consequently, reconstruction-based metrics are more appropriate for assessing model performance [44]. In particular, the evaluation should reflect both the reconstruction accuracy and the physical consistency of the learned latent representation. To this end, three complementary criteria are considered. First, the reconstruction energy (E) quantifies the percentage of the original flow kinetic

energy recovered by the reduced-order representation [2, 45]. Second, the correlation matrix of the latent variables assesses the degree of independence of the learned latent representation, which is desirable for obtaining physically interpretable latent features [2]. Finally, the physical relevance of the latent space is evaluated by comparing the reconstructed spatial structures and the frequency spectra of the associated temporal coefficients with those obtained using Proper Orthogonal Decomposition (POD). Together, these criteria provide a comprehensive assessment of the reconstruction accuracy, latent-space quality, and the ability of the proposed InfoVAE to preserve the dominant flow physics.

To evaluate the performance of both the InfoVAE and Proper Orthogonal Decomposition (POD) methods, the reconstruction energy percentage (E) is employed as a quantitative metric [45]. It is defined as:

$$E = \left(1 - \left\langle \frac{\sum_{p=1}^{N_p} \sum_{i=1}^{2} (u_i - \tilde{u}_i)^2}{\sum_{p=1}^{N_p} \sum_{i=1}^{2} u_i^2} \right\rangle \right) \times 100\%. \tag{10}$$

where $\langle \cdot \rangle$ denotes temporal ensemble averaging, N_p is the number grid points along the spatial domain, u_i represents the i^{th} reference value of fluctuating component velocity, and $\tilde{u}_i$ denotes its low-order reconstruction [2].

To measure the orthogonality of the latent variables, we calculate the correlation matrix $R = (R)_{d \times d}$ is computed as [2]:

$$R_{ii} = 1, \qquad R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}}, \quad 1 \leq i \neq j \leq d, \tag{11}$$

where C_{ij} denotes the components i, j of the covariance matrix C and d is the dimension of the latent-space. A value of 0 is reached when all the variables are completely uncorrelated ($R_{ij} = 0$) and one when they are completely correlated ($R_{ij} = 1$) [2]. This metric reports the degree of correlation between the latent variables[2]. The modes have been ordered according to their cumulative contribution to the reconstructed energy [2], E, following [45].The first mode is chosen as the one with the largest individual contribution to E, and the next are those that have the maximum contribution when added to the previous modes[2]. The periodic case is used as a benchmark to validate the physical soundness of the InfoVAE model representations.

This flow case can be adequately represented only using two POD modes that can represent $E = 98.4\%$ of the kinetic energy of the fluctuations with respect to the base flow, as observed in Fig. 5 similarly to the results obtained by studies conducted by [2]. The Fig. 8 represents POD modes. The first two POD modes of Fig. 8, which are identified as a shedding wake and most energetic modes.

For the InfoVAE case with $\alpha = 0.01$ and $\lambda = 10$, the results in show Figure 8. The InfoVAE spatial modes are defined as the result of using the InfoVAE decoder network. The reconstructed energy in this case is equal to $E = 94\%$ with a good correlation matrix as demonstrated respectively with in Fig. 6 and Fig. 4.

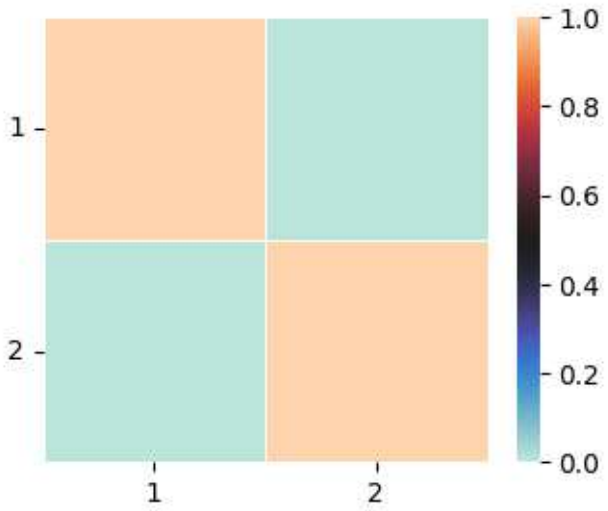


Fig. 4 Correlation matrix of the two-dimensional latent variables learned by the InfoVAE model for the periodic flow at $Re = 40$.

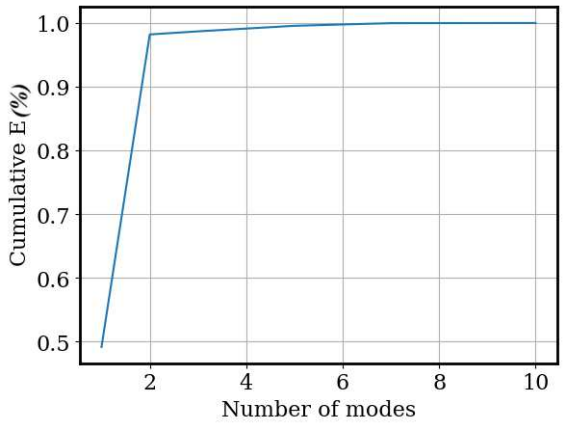


Fig. 5 Cumulative energy percentage, E, captured by the POD modes.

It is observed from Fig. 8 and Fig. 7 that the spatial two first modes obtained from both methods exhibit the same pattern of shedding flow. The spectrum of the temporal coefficients for both methods is shown in the last column of Fig. 8 and Fig. 7. The spectra analysis is from the resulting $r_i(t)$ and $a_i(t)$ time coefficients for the InfoVAE and POD, respectively.

This further confirms that the dynamics associated with the spatial modes are also in good agreement: both methods can capture the same characteristic frequency. These results indicate that the learned latent space captures physically meaningful flow dynamics. Since the InfoVAE architecture requires the user to set a latent-space

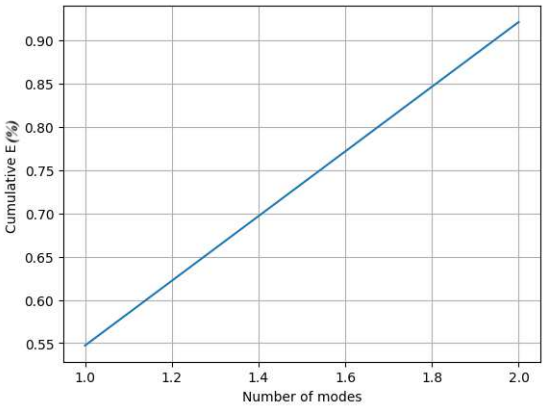


Fig. 6 Cumulative energy percentage, E, recovered by the InfoVAE with a latent-space dimension of d=2.

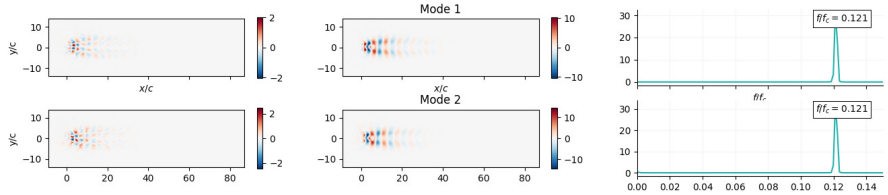


Fig. 7 Proof-of-concept analysis of the periodic flow case using InfoVAE. The first column shows the reconstructed streamwise velocity component associated with each latent feature, obtained by varying the corresponding latent variable with unit amplitude while keeping the remaining latent variables fixed. The second column presents the corresponding crosswise velocity component. The third column displays the frequency spectrum of the temporal coefficient associated with each latent feature, where the characteristic frequency is defined as $f_c = U_\infty/c$.

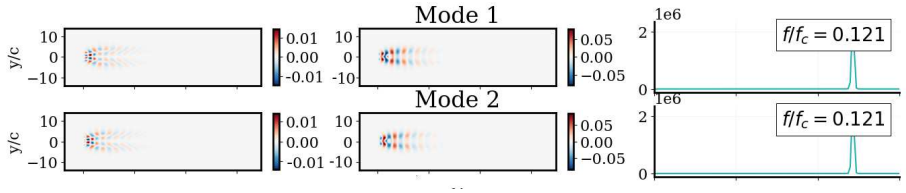


Fig. 8 Proof-of-concept analysis of the periodic flow case using POD. The first column shows the streamwise velocity component of each POD mode normalized to unit amplitude, while the second column presents the corresponding crosswise velocity component. The third column displays the frequency spectrum of the temporal coefficient associated with each POD mode, where the characteristic frequency is defined as $f_c = U_\infty/c$.

dimension d, it can be argued that d can be set to a value larger than 2. However, we observed that models with larger latent-spaces produce only two meaningful modes, as the remaining modes exhibit negligible amplitudes. This behavior shows that the InfoVAE regularization effectively avoids the artificial creation of more modes than necessary to represent the solution.

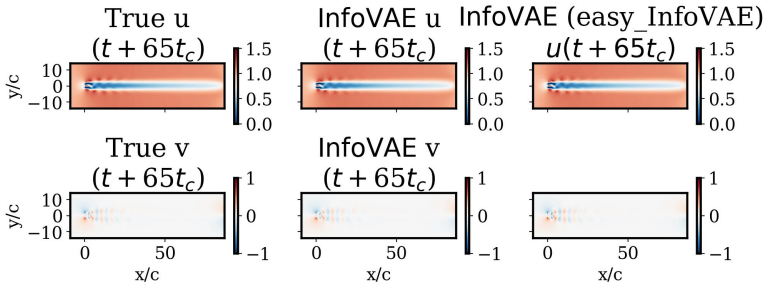


Fig. 9 ROM predicted fields. True, reconstructed and predicted fields for u and v velocity components at t + 65tc: transformer with easy attention.

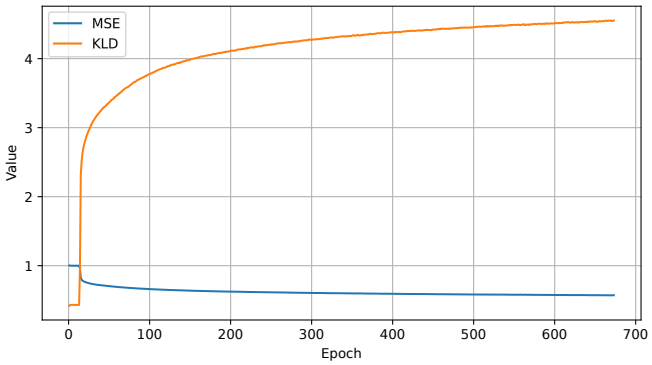


Fig. 10 MSE and KLD evolution over epochs.

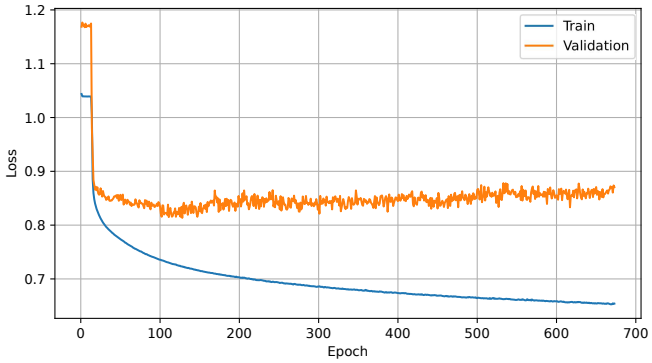


Fig. 11 Training and Validation Loss over Epochs.

The proposed framework first employs an InfoVAE to learn a low-dimensional latent representation of the flow field, capturing its dominant spatial structures. InfoVAE encoder network generates the latent-space time series of the latent-space time series dataset, denoted $r(t)$. This sequence is divided into two-time intervals: $[0,\ t_{\text{train}}]$

and $[t_{\text{train}} + 1,\ t_{\text{end}}]$, where t_{train} correspond to 90% of the available snapshots. The first portion 90% is used for training, while the remaining 10% is reserved for testing. The test data are not used during training and are only encoded using the pre-trained InfoVAE model.

Temporal dynamics are then modeled using an easy-attention-based Transformer. A transformer model is then used to predict the next latent-space vector $r(t + 1)$. The transformer is trained to minimize the difference between the prediction of the next time step and the true data by means of a mean-squared-error loss function. The temporal prediction module is based on an Easy-Attention Transformer operating on the latent representations generated by the InfoVAE. The input and output sequence lengths were both set to 64, with an embedding dimension of ($d_{\text{model}} = 64$). The Transformer consists of four encoder blocks, each employing four attention heads. A time-based embedding strategy was adopted to preserve the temporal ordering of the latent sequences. Residual attention and residual projection connections were enabled to improve information propagation, and the projection dimension was set to 128 with ReLU activation. The Transformer was trained for 100 epochs using the Adam optimizer with a batch size of 256 and a learning rate of (10^{-3}). Early stopping with a patience of 50 epochs was employed to reduce overfitting. Comparative results as shown in Fig. 9 indicates that our approach achieves satisfying performance.

Following the initial assessment of the latent spaces learned by the InfoVAE architectures, a more challenging nonlinear and chaotic flow configuration is investigated at a Reynolds number of $Re = 100$. The case consists of two collinear plates oriented at an angle of 80° with respect to the incoming flow. In addition to the increased Reynolds number, the inclined arrangement introduces further complexity, as the geometry is no longer symmetric relative to the freestream direction.

For this configuration, the latent-space dimension is chosen to be as small as possible while remaining sufficiently large to capture and separate the dominant physical mechanisms governing the flow dynamics.

The training process was monitored through the evolution of the reconstruction loss and the latent regularization term (KL divergence), as illustrated in Fig. 10 and Fig. 11. The reconstruction loss decreases steadily during training before reaching a stable plateau, indicating that the network successfully learns a compact representation of the flow snapshots. Simultaneously, the KL divergence converges to a finite value, showing that the latent variables remain well regularized while preserving meaningful physical information. This balance between reconstruction accuracy and latent regularization prevents posterior collapse and promotes a smooth latent manifold.

The obtained latent representation demonstrates good reconstruction capability for both the training and validation datasets, indicating satisfactory generalization. Unlike linear decomposition techniques such as Proper Orthogonal Decomposition (POD), whose linear basis functions become less effective for strongly nonlinear flow regimes, the nonlinear latent variables learned by the InfoVAE capture more complex spatial structures and interactions between coherent flow features.

Overall, these results confirm that the proposed InfoVAE provides a compact and physically meaningful latent representation suitable for reduced-order modeling of nonlinear CFD flows. The learned latent variables constitute an appropriate

low-dimensional embedding that can subsequently be coupled with temporal prediction models, such as easy-attention-based Transformer, to accurately reproduce the long-term evolution of the flow dynamics.

It should be noted that, for the $Re = 100$ case, the computation of additional quantitative metrics was limited by the available computational resources. Although longer execution times were requested, the training jobs were repeatedly interrupted by HPC node failures occurring before five days of execution. Nevertheless, the stable convergence of both the MSE and KLD, together with the qualitative reconstruction results, indicates that the proposed InfoVAE achieves satisfactory performance. Additional quantitative metrics, including reconstruction energy and latent-variable correlation, will be reported in a revised version of this manuscript or in future work as computational resources permit.

6 Conclusion

This study provides a proof of concept for the effectiveness of modern machine learning techniques in developing compact and robust nonlinear reduced-order models (ROMs) for complex fluid flows. The proposed framework demonstrates the ability to generate accurate predictions while preserving the most relevant flow dynamics and coherent structures, even in highly nonlinear and chaotic regimes. The results show that the InfoVAE-based approach performs satisfactory results.

This paper presented an Information Maximizing Variational Autoencoder (InfoVAE) framework for the reduced-order modeling of two-dimensional fluid flows. The proposed approach was evaluated on both periodic ($Re = 40$) and chaotic ($Re = 100$) flow datasets and compared with the classical Proper Orthogonal Decomposition (POD) method.Indeed, the extracted latent features exhibit behavior similar to the dominant POD modes, indicating that the proposed nonlinear reduced-order model effectively captures the essential flow dynamics. Coupling the learned latent space with an Easy-Attention Transformer further enabled accurate temporal prediction of the reduced dynamics.

The present study is limited to two-dimensional numerical datasets and fixed Reynolds numbers. Furthermore, the hyperparameters of the InfoVAE were selected empirically, and only a limited set of latent dimensions was investigated according to the complexity of the considered flows.

Future work will focus on extending the proposed framework to more challenging fluid dynamics problems, including three-dimensional turbulent flows and higher Reynolds number regimes.

Conflict of interest

The authors declare that they have no conflict of interest.

Availability of data and materials

All datasets used in this study are openly available on Zenodo at https://doi.org/10.5281/zenodo.10501215.

References

[1] Pope, S.B.: Turbulent Flows. Cambridge University Press, Cambridge (2000)

[2] Solera-Rico, A., *et al.*: β-Variational Autoencoders and Transformers for Reduced-Order Modelling of Fluid Flows. Nat. Commun. **15**(1), 1361 (2024) https://doi.org/10.1038/s41467-024-45578-4

[3] Jiménez, J.: Coherent structures in wall-bounded turbulence. J. Fluid Mech. **842**, 1 (2018) https://doi.org/10.1017/jfm.2018.144

[4] Taira, K., *et al.*: Modal analysis of fluid flows: Applications and outlook. AIAA J. **58**(3), 998–1022 (2020)

[5] Taira, K., *et al.*: Modal analysis of fluid flows: An overview. AIAA J. **55**(12), 4013–4041 (2017)

[6] Tsinober, A.: Turbulence: The legacy of A. N. Kolmogorov. J. Fluid Mech. **317**, 407–410 (1996) https://doi.org/10.1017/S0022112096210791

[7] Davidson, P.: Turbulence: An Introduction for Scientists and Engineers, 2nd edn. Oxford University Press

[8] Sagaut, P.: Large Eddy Simulation for Incompressible Flows. Scientific Computation. Springer, Berlin (2006). https://doi.org/10.1007/b137536

[9] Comparative Study on CFD Turbulence Models for the Flow Field in Air Cooled Radiator. Available at https://www.mdpi.com/2227-9717/8/12/1687

[10] Tominaga, Y.: CFD simulations of turbulent flow and dispersion in built environment: A perspective review. J. Wind Eng. Ind. Aerodyn. **249**, 105741 (2024) https://doi.org/10.1016/j.jweia.2024.105741

[11] Comparative CFD Analysis Using RANS and LES Models for NOx Dispersion in Urban Streets with Active Public Interventions in Medellín, Colombia. Available at https://www.mdpi.com/2071-1050/17/15/6872

[12] Lucia, D.J., Beran, P.S., Silva, W.A.: Reduced-order modeling: New approaches for computational physics. Prog. Aerosp. Sci. **40**(1), 51–117 (2004) https://doi.org/10.1016/j.paerosci.2003.12.001

[13] Brunton, S.L., Noack, B.R., Koumoutsakos, P.: Machine learning for fluid mechanics. Annu. Rev. Fluid Mech. **52**(1), 477–508 (2020)

[14] Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. **656**, 5–28 (2010) https://doi.org/10.1017/S0022112010001217

[15] Abreu, L.I., Cavalieri, A.V.G., Schlatter, P., Vinuesa, R., Henningson, D.S.:

Spectral proper orthogonal decomposition and resolvent analysis of near-wall coherent structures in turbulent pipe flows. J. Fluid Mech. **900**, 11 (2020) https://doi.org/10.1017/jfm.2020.445

[16] Berkooz, G., Holmes, P., Lumley, J.L.: The proper orthogonal decomposition in the analysis of turbulent flows. Annu. Rev. Fluid Mech. **25**, 539–575 (1993) https://doi.org/10.1146/annurev.fl.25.010193.002543

[17] Fresca, S., Dede', L., Manzoni, A.: A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized pdes. J. Sci. Comput. **87**(2), 61 (2021) https://doi.org/10.1007/s10915-021-01462-7

[18] Zhang, Y., Zhang, D., Jiang, H.: Review of challenges and opportunities in turbulence modeling: A comparative analysis of data-driven machine learning approaches. J. Mar. Sci. Eng. **11**(7) (2023) https://doi.org/10.3390/jmse11071440

[19] Duraisamy, K.: Perspectives on machine learning-augmented reynolds-averaged and large eddy simulation models of turbulence. Phys. Rev. Fluids **6**(5), 050504 (2021)

[20] Fukami, K., Nabae, Y., Kawai, K., Fukagata, K.: Synthetic turbulent inflow generator using machine learning. Phys. Rev. Fluids **4**(6), 064603 (2019)

[21] Salehipour, H., Peltier, W.R.: Deep learning of mixing by two 'atoms' of stratified turbulence. J. Fluid Mech. **861**, 4 (2019)

[22] Cai, S., Zhou, S., Xu, C., Gao, Q.: Dense motion estimation of particle images via a convolutional neural network. Exp. Fluids **60**(4), 73 (2019)

[23] San, O., Maulik, R.: Extreme learning machine for reduced-order modeling of turbulent geophysical flows. Phys. Rev. E **97**(4), 042322 (2018)

[24] Maulik, R., Fukami, K., Ramachandra, N., Fukagata, K., Taira, K.: Probabilistic neural networks for fluid flow surrogate modeling and data recovery. Phys. Rev. Fluids **5**(10), 104401 (2020)

[25] Srinivasan, P.A., Guastoni, L., Azizpour, H., Schlatter, P., Vinuesa, R.: Predictions of turbulent shear flows using deep neural networks. Phys. Rev. Fluids **4**(5), 054603 (2019)

[26] Milano, M., Koumoutsakos, P.: Neural network modeling for near-wall turbulent flow. J. Comput. Phys. **182**(1), 1–26 (2002)

[27] Baldi, P., Hornik, K.: Neural networks and principal component analysis: Learning from examples without local minima. Neural Netw. **2**(1), 53–58 (1989)

[28] Fukami, K., Fukagata, K., Taira, K.: Assessment of supervised machine learning

methods for fluid flows. Theor. Comput. Fluid Dyn.

[29] Omata, N., Shirayama, S.: A novel method of low-dimensional representation for temporal behavior of flow fields using deep autoencoder. AIP Adv. **9**(1) (2019)

[30] Hasegawa, K., Fukami, K., Murata, T., Fukagata, K.: Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes. Theor. Comput. Fluid Dyn. **34**(4), 367–383 (2020)

[31] Hasegawa, K., Fukami, K., Murata, T., Fukagata, K.: CNN-LSTM based reduced-order modeling of two-dimensional unsteady flows around a circular cylinder at different reynolds numbers. Fluid Dyn. Res. **52**(6), 065501 (2020) https://doi.org/10.1088/1873-7005/abb91d

[32] Nakamura, T., Fukami, K., Hasegawa, K., Nabae, Y., Fukagata, K.: Convolutional neural network and long short-term memory based reduced-order surrogate for minimal turbulent channel flow. Phys. Fluids **33**(2), 025116 (2021) https://doi.org/10.1063/5.0039845

[33] Murata, T., Fukami, K., Fukagata, K.: Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. J. Fluid Mech. **882**, 13 (2020) https://doi.org/10.1017/jfm.2019.822

[34] Fukami, K., Hasegawa, K., Nakamura, T., Morimoto, M., Fukagata, K.: Model order reduction with neural networks: Application to laminar and turbulent flows. SN Comput. Sci. **2**(6), 467 (2021) https://doi.org/10.1007/s42979-021-00867-3

[35] Maulik, R., Lusch, B., Balaprakash, P.: Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders. Phys. Fluids **33**(3), 037106 (2021) https://doi.org/10.1063/5.0039986

[36] Erichson, N.B., Muehlebach, M., Mahoney, M.W.: Physics-Informed Autoencoders for Lyapunov-Stable Fluid Flow Prediction. Preprint at https://arxiv.org/abs/1905.10866 (2019)

[37] Carlberg, K.T., Jameson, A., Kochenderfer, M.J., Morton, J., Peng, L., Witherden, F.D.: Recovering missing CFD data for high-order discretizations using deep neural networks and dynamics learning. J. Comput. Phys. **395**, 105–124 (2019)

[38] Liu, Y., Ponce, C., Brunton, S.L., Kutz, J.N.: Multiresolution convolutional autoencoders. J. Comput. Phys. **474**, 111801 (2023)

[39] Xu, J., Duraisamy, K.: Multi-level convolutional autoencoder networks for parametric prediction of spatio-temporal dynamics. Comput. Methods Appl. Mech. Eng. **372**, 113379 (2020) https://doi.org/10.1016/j.cma.2020.113379

[40] Zhao, S., Song, J., Ermon, S.: InfoVAE: Information Maximizing Variational Autoencoders. Preprint at https://arxiv.org/abs/1706.02262 (2017)

[41] Zhao, S., Song, J., Ermon, S.: InfoVAE: Balancing learning and inference in variational autoencoders. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence. AAAI Press, ??? (2019). https://doi.org/10.1609/aaai.v33i01.33015885

[42] Ye, F., Bors, A.G.: InfoVAEGAN: Learning joint interpretable representations by information maximization and maximum likelihood. In: 2021 IEEE International Conference on Image Processing (ICIP), pp. 749–753 (2021). https://doi.org/10.1109/ICIP42928.2021.9506169

[43] Sirovich, L.: Turbulence and the dynamics of coherent structures. part I: Coherent structures. Quart. Appl. Math. **45**(3), 561–571 (1987) https://doi.org/10.1090/qam/910462

[44] Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge, MA (2016)

[45] Eivazi, H., Le Clainche, S., Hoyas, S., Vinuesa, R.: Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows. Expert Syst. Appl. **202**, 117038 (2022) https://doi.org/10.1016/j.eswa.2022.117038