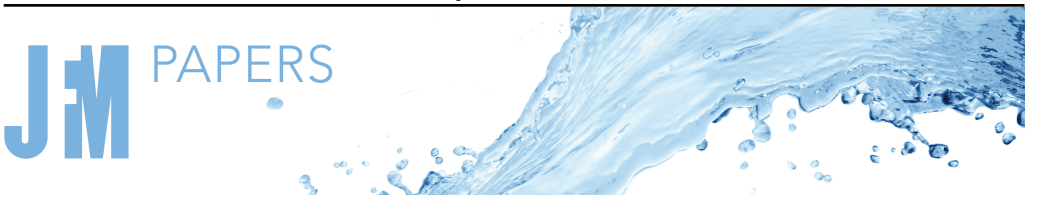
J. Fluid Mech. (2026), *vol.* 1035, A3, doi:10.1017/jfm.2026.11511



High-fidelity turbulent flow field reconstruction using a residual-based diffusion model

Golsa Tabe Jamaat[1], **Takayuki Okatani**[1,2] **and Yuji Hattori**[3]

[1]Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi 980-8579, Japan

[2]RIKEN Center for AIP, Tokyo 103-0027, Japan

[3]Institute of Fluid Science, Tohoku University, Sendai, Miyagi 980-8577, Japan

Corresponding author: Golsa Tabe Jamaat, tabe.jamaat.golsa.d2@tohoku.ac.jp

(Received 27 June 2025; revised 28 February 2026; accepted 30 March 2026)

In this study, we develop a super-resolution (SR) model for homogeneous isotropic turbulence (HIT) inspired by the recently proposed low-inference-cost ResShift diffusion model. The training data are obtained from direct numerical simulation of two three-dimensional HIT cases with varying grid resolutions and Reynolds numbers ($Re_\lambda = 94$ and 173) to increase the model's generalisability. The model is trained on two-dimensional snapshots rather than full three-dimensional fields, as training and inference on three-dimensional data would increase the computational cost significantly. Both the data from the whole domain and the data from a quarter of the domain are considered in the dataset to increase the diversity and quantity of training samples. This strategy also helps the model learn more localised flow structures and reduces dependence on global domain-specific patterns. The model is trained using single snapshots of velocity components for three upsampling factors of 4, 8 and 16. To assess the generalisability of the trained model, it is tested for flows under conditions different from those of the training data. Additionally, the high-resolution reconstruction of flow fields from low-resolution turbulent boundary layer data is performed to evaluate the model's performance in anisotropic turbulence. The results show that the diffusion model presented in this study performs well in predicting the velocity field even for high upsampling factors, and unlike bicubic interpolation, convolutional neural network (CNN)- and U-Net-based models, it does not generate a visually blurry flow field when applied to high upsampling factors. It also outperforms bicubic interpolation, CNN- and U-Net-based models, as well as the traditional conditional denoising diffusion probabilistic model designed for SR, in predicting flow statistics. The model effectively extracts flow features, generates flow structures of varying sizes and shows strong performance in predicting vorticity. It also reproduces the energy

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

spectrum at high wavenumbers with reasonable accuracy, indicating the recovery of small-scale structures often lost in coarse data. This capability is valuable for subgrid-scale stress estimation and helps improve the physical fidelity of large eddy simulation frameworks.

Key words: turbulent flows

1. Introduction

In recent years, machine learning has gained significant importance in fluid dynamics, and numerous studies have explored its application to address various challenges in this field (Duraisamy, Iaccarino & Xiao 2019; Brunton, Noack & Koumoutsakos 2020; Zhang *et al.* 2021; Vinuesa & Brunton 2022; Zhang *et al.* 2023*b*). One area of its application is turbulence modelling, which includes both Reynolds-averaged Navier–Stokes (RANS) and large eddy simulation (LES). In the context of RANS, machine learning has been utilised to improve Reynolds stress closure modelling (Ling, Kurzawski & Templeton 2016; Jiang *et al.* 2021; Sun *et al.* 2022*b*). For LES, it has been applied to enhance subgrid-scale (SGS) modelling (Gamahara & Hattori 2017; Wang *et al.* 2018*b*; Zhou *et al.* 2019; Park & Choi 2021; Tabe Jamaat & Hattori 2022; Zhao, Jin & Zhou 2022; Liu *et al.* 2022) as well as wall modelling (Yang *et al.* 2019; Bae & Koumoutsakos 2022; Tabe Jamaat & Hattori 2023; Lozano-Durán & Bae 2023; Tabe Jamaat *et al.* 2024; Maejima, Tanino & Kawai 2024). Additionally, machine learning techniques have been employed for predicting complex turbulent flows in engineering and geophysical applications (Frank, Drikakis & Charissis 2020; Zhang *et al.* 2022; Chen *et al.* 2023; Zhang *et al.* 2024*a*,*b*), flow field reconstruction (Fan *et al.* 2023; Yousif *et al.* 2023; Matsuo *et al.* 2024; Fan *et al.* 2025*a*), flow control (Guastoni *et al.* 2023; Vignon, Rabault & Vinuesa 2023; Ishize, Omichi & Fukagata 2024) and solving partial differential equations (PDEs) for flow simulation (Li *et al.* 2023; McCabe *et al.* 2024; Rahman *et al.* 2024; Fan *et al.* 2025*b*).

Another important application of machine learning in fluid dynamics is super-resolution (SR) reconstruction of flow fields, which has gained significant attention in recent years and was first introduced by Xie *et al.* (2018) and Fukami, Fukagata & Taira (2019). Single-image SR is a fundamental low-level vision task that focuses on generating a high-resolution (HR) image from a low-resolution (LR) input while preserving structural and perceptual consistency. This task is challenging because it is an inverse problem, where a single LR image can correspond to multiple possible HR outputs.

Super-resolution techniques are recognised as effective tools for enhancing turbulent flow resolution, enabling more accurate predictions at lower computational costs. Direct numerical simulation (DNS), which is an HR simulation, resolves all flow scales but is computationally expensive, limiting its practical use. As a solution to these computational limitations, SR techniques can be employed to reconstruct fine-scale turbulence structures from coarse data while preserving the essential flow dynamics. These approaches can enhance the efficiency of numerical simulations and experimental measurements, bridging the gap between LR and HR data, and offer considerable potential for advancing turbulence research and its applications in engineering and industry.

The simplest SR method is interpolation, which fills in missing details using nearby data. It assumes smooth transitions, making it ineffective for capturing the chaotic and multi-scale nature of turbulence. In image SR, regression-based methods (Dong *et al.* 2015; Liang *et al.* 2021; Sun *et al.* 2022*a*) like the convolutional neural network (CNN)

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

establish a direct mapping from LR to HR images using a loss function, typically the mean squared error (MSE). However, these approaches tend to produce outputs with overly smoothed features, failing to capture fine details effectively (Gao *et al.* 2023), especially as the upsampling factor increases.

Deep generative models, such as autoregressive models (Chen *et al.* 2018; Lee *et al.* 2022), variational autoencoders (VAE) (Kingma & Welling 2013; Van Den Oord *et al.* 2017) and the generative adversarial network (GAN) (Goodfellow *et al.* 2014), have achieved notable success in generating high-quality images by learning the intricate distributions of images. These models have also been applied to the image SR task (Ledig *et al.* 2017; Liu *et al.* 2021*b*; Chira *et al.* 2022; Guo *et al.* 2022) to enhance fine details. However, GAN-based methods for SR often encounter issues like mode collapse, where the model generates a narrow range of similar outputs with limited diversity. Moreover, the training process is generally unstable and can take a long time to converge. To stabilise training, an additional discriminator is used, which is crucial for optimisation but unnecessary during inference, adding complexity to the model without providing direct benefits during testing (Isola *et al.* 2017). On the other hand, VAE models underperform compared with GAN-based models, and autoregressive models are computationally expensive, limiting their application to LR images (Gao *et al.* 2023).

There have been numerous studies on developing data-driven models for SR reconstruction of turbulent flows. Fukami, Fukagata & Taira (2021) developed a CNN-based spatio-temporal SR model and applied it to spatio-temporal data for enhanced turbulent flow field reconstruction. Their results show that the model successfully reconstructs fine-scale turbulence structures, closely matching DNS results, and demonstrate the model's ability to capture the key flow dynamics. However, their method struggles to reconstruct turbulent flow field in time when the temporal two-point correlation coefficient is low. Zhang *et al.* (2023*a*) developed a U-Net-based architecture (Ronneberger, Fischer & Brox 2015) combined with additional convolutional layers for SR reconstruction of two-dimensional cylinder wake flow and hydrofoil velocity fields. They showed that their model outperforms both the standard U-Net and conventional CNN models. In a more recent study, Fukami & Taira (2024) explored SR reconstruction using single snapshot data. They demonstrated that their approach can effectively recover HR turbulence structures, such as vortices and eddies, from limited input. They designed the test dataset to produce snapshots with vortical structures similar in size to those in subdomains of a single snapshot, which may limit the model's ability to generalise to other flow structures. Additionally, since the model is CNN-based and the upsampling factor is high, the reconstructed flow field remains somewhat blurry.

In addition to CNN, the ability of transformer and GAN has been tested for SR reconstruction of turbulent flows. Xu *et al.* (2023) developed a transformer-based model (Chen *et al.* 2021*a*) for SR reconstruction of turbulent channel flow and homogeneous isotropic turbulence (HIT). However, their model did not show significant improvement over CNN-based models, and they did not test it on data with an unseen Reynolds number to assess its generalisability to different flow conditions. Kim *et al.* (2021) applied GAN (Ledig *et al.* 2017) and cycle-consistent GAN (CycleGAN) (Zhu *et al.* 2017; Yuan *et al.* 2018), which are supervised and unsupervised learning approaches, respectively, to the SR reconstruction of HIT and turbulent channel flow. They showed that, for paired data, the GAN model results in lower error than the CycleGAN model. There are also other studies which have explored SR reconstruction of turbulent flows using CNN-based methods (Liu *et al.* 2020; Zhou *et al.* 2022), GAN-based approaches (Xie *et al.* 2018; Yousif, Yu & Lim 2021; Chen *et al.* 2021*b*; Yousif, Yu & Lim 2022; Yu *et al.* 2022; Ward 2023) and transformer-based methods (Tang *et al.* 2024; Zeng *et al.* 2024; Liu *et al.* 2025).

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

However, there is ongoing research to explore new deep learning methods for the SR reconstruction of turbulent flows and to develop a model with enhanced performance, capable of capturing flow structures of different sizes and applicable to various upsampling factors without producing a visually blurry flow field.

Recently, diffusion models have gained significant success in image generation (Ho, Jain & Abbeel 2020; Dhariwal & Nichol 2021), achieving state-of-the-art performance across various computer vision tasks. The iterative process of diffusion models has proven to be highly effective in learning complex distributions. These models have outperformed GANs, particularly excelling in generating high-quality, diverse and realistic images. Unlike GANs, which depend on adversarial training and often suffer from mode collapse, diffusion models employ a systematic process of transforming noise into the structured data through iterative denoising steps, resulting in more stable and higher-quality outputs. There are also several studies that have explored their application to image SR, demonstrating their effectiveness in enhancing image quality (Saharia *et al.* 2022; Li *et al.* 2022; Gao *et al.* 2023; Yue *et al.* 2023).

Following their success in image generation, diffusion models have also been applied to the fluid dynamics field. Du *et al.* (2024) developed a framework based on the latent diffusion model (Rombach *et al.* 2022) for generating spatio-temporal turbulence. They proposed a framework that combines a conditional neural field and a latent diffusion model to generate spatio-temporal turbulence. Their model effectively simulates complex turbulent flows across varying domains by conditioning the diffusion process on relevant spatio-temporal data. However, the model's performance heavily depends on the size and diversity of the training data, requiring a sufficiently large set of subsequences to ensure effective generalisation. Shu, Li & Farimani (2023) used the denoising diffusion probabilistic model (DDPM) (Ho *et al.* 2020; Saharia *et al.* 2022) with modifications for SR reconstruction of two-dimensional Kolmogorov flow and incorporated physics into their model by using the gradient of the PDE residual as conditioning information. The results of their study show that, compared with bicubic interpolation, the model can generate relatively high-quality flow fields without blurriness. However, it struggles to capture the flow details accurately, especially for the upsampling factor of 8. The obtained energy spectrum does not match the DNS results at high wavenumbers, indicating that the model has difficulty capturing the small-scale features of the flow. Additionally, the predicted vorticity distribution is less accurate than that produced by bicubic interpolation and deviates from the DNS results even for the upsampling factor of 4. Qi & Ma (2024) implemented the conditional denoising diffusion model (Saharia *et al.* 2022) to the SR reconstruction of turbulent flows. They found that, although the model learns the distribution, it does not achieve sufficient accuracy. To address this issue, they incorporated a residual network (He *et al.* 2016) to further improve the model's performance. They tested their model on flow data with the same Reynolds number as that used in the training process, and showed that their model requires significantly more time for reconstruction compared with the CNN- and GAN-based models.

Despite their widespread use, diffusion models suffer from slow inference since they require hundreds or even thousands of sampling steps. Although accelerated sampling techniques can speed up computation, they often compromise performance, producing overly blurred SR results. In the present study, we aim to develop a data-driven model inspired by the strategy of the recently proposed ResShift diffusion model (Yue *et al.* 2023) for the SR reconstruction of three-dimensional HIT, which can capture flow details well and generate clear HR flow fields without blurriness, even for high upsampling factors. Unlike the diffusion models used in previous studies on SR reconstruction of turbulent flows, the developed model has low inference time while being capable of generating HR

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

flow fields. The ResShift model was originally developed for HR real-world and synthetic image reconstruction for the upsampling factor of 4, and has been applied, with some modifications, to the SR of microscopy images (Gao *et al.* 2025) as well. Handling high compression rates in latent diffusion models (LDMs) is a significant challenge, as such rates can affect detail consistency and may lead to quality degradation during image reconstruction, especially as the compression rate increases (Lin *et al.* 2023). Additionally, because LDMs typically require an autoencoder tailored to a specific upsampling factor, using a pixel-wise formulation makes the model more flexible and easier to train without such constraints. Therefore, since capturing details is crucial for the SR of turbulent flows and we aim to train the model for different upsampling factors, unlike the original ResShift model, we construct the model in pixel space rather than latent space. Additionally, we include the perceptual loss in the training process to enhance the reconstruction quality and preserve fine flow structures.

In this study, we will use a large dataset obtained from the DNS of two HIT simulations with completely different parameters, including varying Reynolds numbers and grid sizes, to increase the generalisation ability of the model. We will also employ data from the whole domain as well as data from a quarter of the domain to increase the diversity of training samples, reduce the dependence of the model on domain-specific patterns and enable it to learn localised flow structures. The model will be trained for three SR ratios of 4, 8 and 16, and evaluated on data from the same simulations used for training as well as simulations with different parameters. The performance of the model will also be compared with that of bicubic interpolation, a CNN- and a U-Net-based model (Ronneberger *et al.* 2015) and the conditional DDPM designed for SR Saharia *et al.* (2022).

The paper is organised as follows. In § 2, we introduce the overall methodology, including the dataset details, the diffusion model description and the training set-up. The results of applying the diffusion model are presented and discussed in § 3, and the conclusions of the study are given in § 4.

2. Methodology

In the present study, we develop a diffusion model for the SR reconstruction of HIT. The model will be trained to predict the HR velocity components using the corresponding LR inputs. In this section, first, we provide the details of the dataset, which plays a crucial role in developing a reliable model. Next, we explain the diffusion model, which is used in this study for establishing an SR model for reconstructing the HR flow field. Finally, we present the training set-up, detailing the training process and hyperparameter choices.

2.1. *Dataset preparation*

To generate the training dataset, we conduct DNS for three-dimensional HIT. Direct numerical simulation was conducted using the same approach as Ishihara *et al.* (2007) and Miyazaki & Hattori (2020). We consider an incompressible fluid with a uniform density of one, governed by the Navier–Stokes (NS) equation

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j}(u_i u_j) = -\frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_k x_k} + f, \tag{2.1}$$

and continuity equation

$$\frac{\partial u_i}{\partial x_i} = 0, \tag{2.2}$$

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

$Case$	N	Re_λ	k_{max}	$10^3\Delta t$	$10^4\nu$	λ	$10^3\eta$	ε	$\mathcal{T}$
HIT1	256	94	121	1.0	20.0	0.327	17.1	0.0936	1.91
HIT2	512	173	241	1.0	7.0	0.210	8.1	0.0795	2.10
HIT3	256	60	121	1.0	50	0.521	34.75	0.0857	1.60
HIT4	1024	268	483	0.4	2.8	0.130	4.03	0.0829	1.94

Table 1. Direct numerical simulation parameters and turbulence characteristics of HIT simulations.

where u, p, ν and f denote the velocity, pressure, kinematic viscosity and forcing term, respectively. The turbulence field is considered periodic in all directions of the Cartesian coordinate system, with a periodic domain of size 2π, allowing it to be represented as a Fourier series. The NS equations are solved using spectral method, and the dealiasing is performed by the phase-shift method, retaining Fourier modes satisfying $k < k_{max} = \sqrt{2}N/3$, where k and N represent the wavenumber and number of grid points in each direction of the Cartesian coordinate system. The fourth-order Runge–Kutta method is used for time marching.

The forcing term $\hat{f}(k) = -c\hat{u}(k)$ is applied in wavenumber space, where $\hat{f}$ and $\hat{u}$ denote the Fourier transforms of f and u, respectively. The coefficient c is non-zero only for wavenumbers within the range $k < K_c = 2.5$ and is dynamically adjusted at each time step to maintain the total kinematic energy E nearly constant ($\approx$0.5). By restricting the forcing to lower wavenumbers, its direct influence on the small-scale dynamics is minimised. For the training dataset, we conduct two forced HIT simulations, HIT1 and HIT2, with different flow parameters, as shown in table 1, where Re_λ, Δt, λ, η, ε and $\mathcal{T}$ represent the Reynolds number based on the Taylor microscale, time-step size, Taylor microscale, Kolmogorov length scale, mean energy dissipation rate and the eddy turnover time, respectively.

In previous studies, various methods have been employed to obtain degraded LR data for the training process, including max and average pooling (Fukami *et al.* 2019; Kim *et al.* 2021), and selecting discrete points in the flow domain (Yu *et al.* 2022). In the present study, we apply the Gaussian filter to the DNS data, which can produce smooth and physically consistent coarse-grained flow fields. This method has been commonly used in data-driven turbulence modelling for LES as it helps make the data more representative of LES flow fields and has been shown to yield more accurate models than other filter types, such as sharp spectral and top-hat filters (Zhou *et al.* 2019; Miyazaki & Hattori 2020; Tabe Jamaat & Hattori 2022). The Gaussian filter kernel is defined as

$$\hat{G}(k) = \exp\left(-\frac{k^2\Delta^2}{24}\right), \tag{2.3}$$

in the spectral space, where G and Δ denote the filter kernel and filter width, respectively. After applying the Gaussian filter, we coarse grain the results to the desired number of grid points in the LR flow field (Zanna & Bolton 2020; Guan *et al.* 2022). The coarse graining is performed in spectral space as well by applying a cutoff filter to the data

$$\hat{G} = H(k_c - |k|), \tag{2.4}$$

where $k_c = \pi/\Delta$ is the cutoff wavenumber and H denotes the Heaviside step function. The filter size Δ is determined based on the upsampling factor. For instance, when $N = 256$ and the upsampling factor is 4, $\Delta = 2\pi/64$, where 64 corresponds to the number of grid points in the LR flow field.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

There have been different methods used for normalising the training dataset. In the studies by Yousif *et al.* (2021), Tang *et al.* (2024) and Yu *et al.* (2022), the data are non-dimensionalised using min–max normalisation. However, in a study by Fukami & Taira (2024) on the SR reconstruction of two-dimensional decaying isotropic turbulence, the input and output data are normalised using the instantaneous maximum absolute vorticity to consider the magnitude variations in the vorticity field at different Reynolds numbers. In the present study, we normalise the data of each velocity component using the maximum absolute value of the corresponding velocity component. Our training dataset includes 25 600 snapshots, each of size 256^2, which is considerably larger than the datasets used in the previous studies (Liu *et al.* 2020; Xu *et al.* 2023), and the validation dataset includes 800 snapshots. In more detail, for training, 800 time steps with a uniform time interval of 0.2 are selected for HIT1, and 200 time steps with a uniform time interval of 0.1 are chosen for HIT2. For the validation dataset, 10 time steps are uniformly selected for both HIT1 and HIT2, with time intervals of 2 and 1, respectively. In each time step, 16 x–y planes are extracted. We have also evaluated the impact of training data size by comparing the model trained with the original dataset to one trained with twice the dataset size, ensuring sufficient data for the training process.

For data from the simulation on the grid size of 512^3 (HIT2), we divide each two-dimensional snapshot into four subdomains, each of size 256^2. This form of data augmentation increases both the diversity and number of training samples. Furthermore, it reduces the model's reliance on global, domain-specific patterns and enhances its ability to learn localised flow structures, which is beneficial for reconstructing turbulent features at high upsampling factors. This choice of domain size for HIT2 satisfies the condition that the domain length should be larger than the integral length scale.

Since the training data come from different simulations and the dataset size is large, the model is expected to generalise reasonably well to flow fields with varying Reynolds numbers and grid sizes. The data-driven SR model will be trained for three upsampling factors of 4, 8 and 16. In all cases, the output size will be fixed, and the input sizes will be 64, 32 and 16, respectively. In Appendix A, additional details have been provided on the relationship between the LR grid spacing and the characteristic turbulence scales.

For evaluation, the model is tested on data from simulations with Reynolds numbers lower and higher than those of the training data. Furthermore, the model will be tested on input sizes significantly larger than those of the training data. The test dataset includes data from two different simulations, as well as data from the same simulations used for training but at different time steps. The simulation settings for the test cases are similar to those of the training data, with grid sizes and viscosities set to 256^3 and 5×10^{-3} for the lower Reynolds number flow, and 1024^3 and 2.8×10^{-4} for the higher Reynolds number flow. The details of the simulations for the test cases HIT3 and HIT4 are provided in table 1. In Appendix B, we also test the model on the LR data obtained using average pooling to assess the model's sensitivity to the coarse-graining method.

2.2. *Diffusion model for SR reconstruction*

In this study, we develop a diffusion model based on the ResShift model (Yue *et al.* 2023) for SR reconstruction of HIT. The model employs a Markov chain to transfer between HR and LR data by shifting the residuals between them, greatly improving the transition efficiency. A detailed noise schedule also manages the shifting speed and noise strength during the diffusion process. Unlike the original ResShift model, which was developed for image SR in latent space at a fixed upsampling ratio, the present model operates in pixel space on turbulent flow fields, enabling easy applicability to different upsampling factors

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

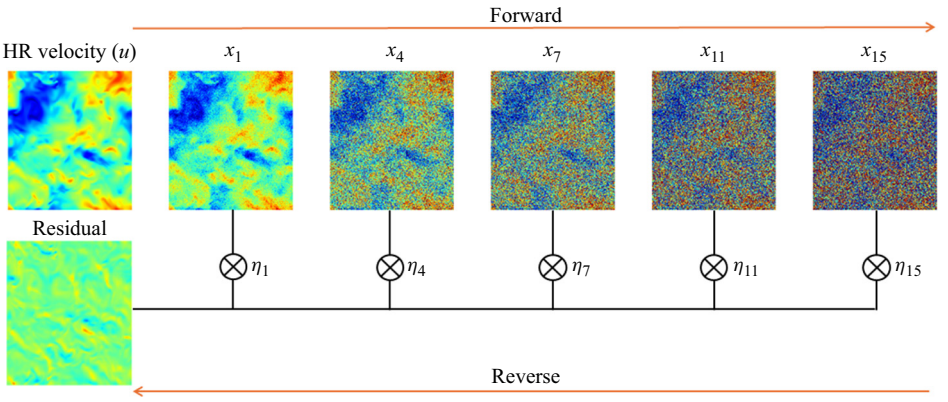


Figure 1. Schematic of ResShift diffusion model.

(4, 8, 16) and improving pixel-wise accuracy. As will be discussed later, perceptual loss is incorporated to better capture fine-scale vortical structures, and a cosine learning rate scheduler enhances training stability and generalisation. These modifications allow more accurate reconstruction of multi-scale turbulence while preserving high-frequency details.

In the SR task using diffusion models, one common approach is to input the LR image into the model and retrain it from scratch using the training data. Alternatively, an unconditional pretrained diffusion model can be used as a prior, adjusting the reverse path to generate the HR image. However, both approaches often suffer from inefficiencies during inference, requiring hundreds or even thousands of sampling steps (Ho *et al.* 2020; Saharia *et al.* 2022). Furthermore, existing acceleration techniques aimed at speeding up inference and the sampling process usually compromise performance, leading to over-smoothed results (Song, Meng & Ermon 2020; Nichol & Dhariwal 2021). The strategy used in the ResShift diffusion model addresses both efficiency and performance. Diffusion models generally consist of two main stages: the forward process and the backward process. In this section, we outline these stages within the context of the ResShift model, which is focused on the SR task. A schematic of the ResShift diffusion model is shown in figure 1.

The ResShift model facilitates the transition from an HR image (x_0) to an LR image (y) by gradually shifting the residual, defined as $d = y - x_0$, through a Markov chain over T steps. T is set to 15, which is significantly lower than that of other diffusion models (Ho *et al.* 2020; Song *et al.* 2020; Rombach *et al.* 2022), leading to a notable reduction in the computational cost of inference. To ensure both images share the same spatial resolution, the LR image y is first upsampled using bicubic interpolation. The Markov chain serves as a bridge between the HR and LR images, while the diffusion process integrates information from the LR image, enabling SR through reverse sampling conditioned on it. The shifting sequence $\{\eta_t\}_{t=1}^{T}$ increases monotonically with time step t, satisfying $\eta_1 \to 0$ and $\eta_T \to 1$. In the forward process, the marginal distribution at any time step $t \in \{1, 2, \ldots, T\}$ can be obtained as follows:

$$q(x_t|x_0, y) = \mathcal{N}\big(x_t; x_0 + \eta_t d, k^2\eta_t I\big), \tag{2.5}$$

where k is a hyperparameter controlling the noise variance, which is set to 2, I is the identity matrix and $\mathcal{N}(x_t; \mu_t, \sigma_t^2)$ denotes the Gaussian (normal) distribution over x_t with mean μ_t and variance σ_t^2, which can be equivalently written as

$$x_t = \mu_t + \sigma_t\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{2.6}$$

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

where $\mu_t = x_0 + \eta_t d$ and $\sigma_t^2 = k^2 \eta_t I$. The schedule for η_t is

$$\sqrt{\eta_t} = \sqrt{\eta_1} \times b_0^{\beta_t}, \tag{2.7}$$

where $t = \{2, \ldots, T-1\}$. b_0 is

$$b_0 = \exp\Big[\frac{1}{2(T-1)} \log \frac{\eta_T}{\eta_1}\Big], \tag{2.8}$$

and β_t is defined as

$$\beta_t = \Big(\frac{t-1}{T-1}\Big)^p \times (T-1), \tag{2.9}$$

where $p = 0.3$.

The aim of the reverse process is to gradually denoise a noisy sample by reversing the forward diffusion process and estimating the posterior distribution $p(x_0|y)$. The inverse process can be performed as follows:

$$p_\theta(x_{t-1}|x_t, y) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, y, t), k^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t I), \tag{2.10}$$

where $\alpha_1 = \eta_1$ and $\alpha_t = \eta_t - \eta_{t-1}$ for $t > 1$. Therefore, x_{t-1} can be calculated as follows:

$$x_{t-1} = \mu_\theta + \sigma_t \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{2.11}$$

where $\sigma_t^2 = k^2(\eta_{t-1}/\eta_t)\alpha_t I$. The mean parameter μ_θ is reparameterised as

$$\mu_\theta(x_t, y, t) = \frac{\eta_{t-1}}{\eta_t} x_t + \frac{\alpha_t}{\eta_t} f_\theta(x_t, y, t), \tag{2.12}$$

where f_θ is a deep network with trainable parameters (weights) θ. The network f_θ is trained to predict x_0 from the noisy intermediate state x_t, conditioned on the LR flow field (y) and the time step t, by minimising the loss function between the predicted and ground truth x_0. The reverse process starts from the final time step T, and the model denoises step-by-step using the learned posterior p_θ, progressively reconstructing less noisy flow fields until the final HR flow field x_0 is obtained. The same network f_θ is used for all steps during both training and inference, and since the forward schedule $\{\eta_t\}$ controls the noise level at each step, errors naturally vary with t.

2.3. *Model training set-up*

We utilise the Adam optimiser (Kingma & Ba 2014) with a cosine learning rate scheduler for smoother and improved convergence, where the minimum learning rate is set to 2×10^{-5}. The batch size is 64 and the training is performed for 500 epochs.

Most previous studies on SR reconstruction of turbulent flows, particularly those employing CNN-based models, have relied solely on MSE as the loss function (Fukami *et al.* 2019, 2021; Xu *et al.* 2023; Zhang *et al.* 2023*a*). However, there have been some recent studies that employed perceptual loss (Johnson, Alahi & Fei-Fei 2016; Ledig *et al.* 2017; Zhang *et al.* 2018) together with MSE (Yousif *et al.* 2021; Yu *et al.* 2022; Tang *et al.* 2024). The perceptual loss measures the difference in the extracted features between the ground truth and predicted data. It improves the perceptual quality of the reconstructed images by focusing on high-level features rather than pixel accuracy. By emphasising feature-level consistency, it helps preserve important structures and patterns, reduces artefacts and blurriness and improves the overall clarity and realism of the generated outputs.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

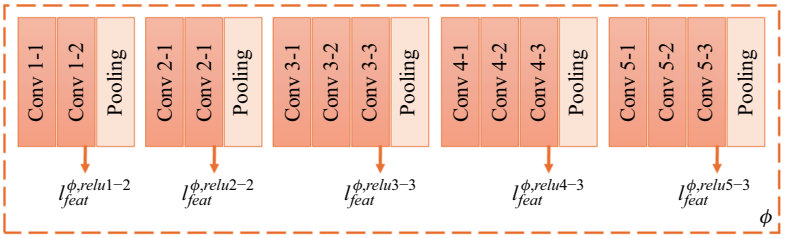


Figure 2. A schematic of the VGG16 network.

To capture perceptual quality more effectively, we adopt the learned perceptual image patch similarity (LPIPS) loss, a perceptual loss based on deep feature distance. The LPIPS captures differences in learned representations, making it suitable for preserving complex spatial patterns in turbulent flows that pixel-wise losses may not capture well. We use the pretrained deep convolutional network VGG16 (Simonyan & Zisserman 2014) to extract features. A schematic of VGG16 and its feature extraction process in different layers is shown in figure 2. The three velocity components are used as the input channels to VGG16. While the study by Wang *et al.* (2018*a*) used only a single layer of VGG19 for feature extraction, subsequent research (Yousif *et al.* 2021) has shown that leveraging more layers can improve training stability. Therefore, we employ five layers from VGG16 to extract multi-scale features. The features in different blocks are extracted at the end of each block, just before the pooling layer.

In this study, the objective function is the combination of reconstruction loss L_{MSE}, which is the MSE, and the perceptual loss L_p

$$L_{tot} = L_{MSE} + \lambda L_p, \tag{2.13}$$

where λ is set to 0.005 and L_{MSE} is defined as

$$L_{MSE} = \frac{1}{N}\sum_{j=1}^{N} \|\hat{u}_j - u_j\|_2^2, \tag{2.14}$$

where $\hat{u}_j$ is the predicted velocity for the jth data point, u_j is the corresponding ground truth velocity and N is the total number of data points in the batch. Also, L_p is given by

$$L_p = \sum_{l=1}^{L} w_l ||\phi_l(\hat{u}_j) - \phi_l(u_j)||_2^2, \tag{2.15}$$

where ϕ_l represents the feature map at the lth layer of the pretrained network, which is VGG16 in this study, w_l is the learned weight of the lth layer and L denotes the layers used for the loss calculation. A schematic of LPIPS loss calculation is shown in figure 3.

To predict the HR flow field (x_0) using f_θ in (2.12), the U-Net model (Ronneberger *et al.* 2015) is used. A schematic of the U-Net model is presented in figure 4. The U-Net architecture used in DDPM (Ho *et al.* 2020), combined with the Swin transformer (Liu *et al.* 2021*a*) block instead of the self-attention layer, is employed as the deep network to improve the model's performance.

The hierarchical structure and shifted window mechanism of the Swin transformer enable effective modelling of local and global dependencies while allowing the network to handle input sizes. Attention is computed within local windows and gradually aggregated across shifted windows, providing flexibility for various input resolutions (Liu *et al.* 2021*a*). Combined with the convolutional blocks, which are inherently resolution agnostic,

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

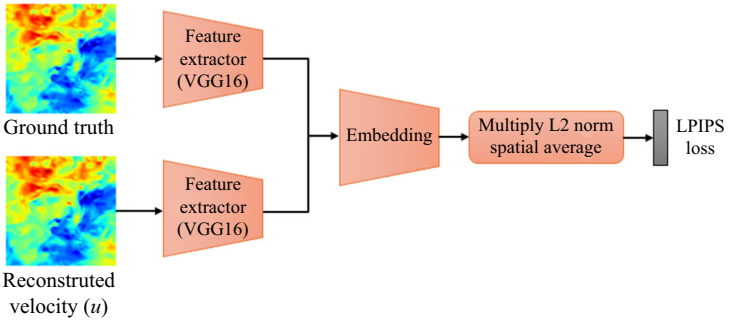


Figure 3. Illustration of LPIPS loss function computation process.

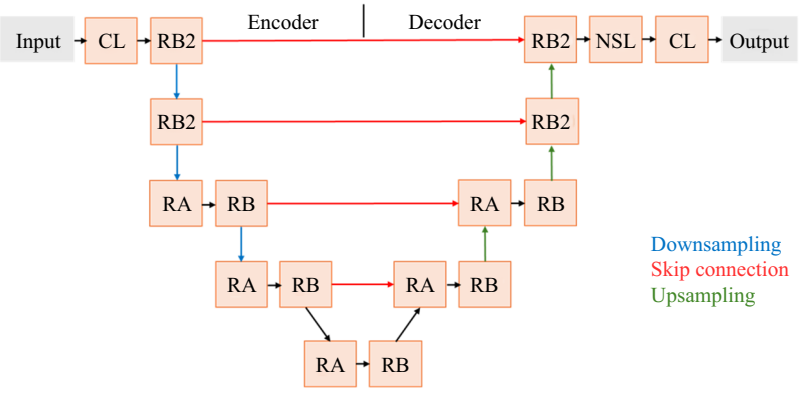


Figure 4. A schematic of U-Net.

the network can accommodate different input sizes without retraining. Since we evaluate the model across different resolutions, this design can enhance the robustness and adaptability of the model to various input sizes. The details of the U-Net architecture are presented in Appendix C. Additionally, in Appendix D, the advantages of using pixel-wise diffusion model, the LPIPS loss and the Swin transformer are further discussed.

3. Results

In this section, the trained diffusion model is assessed on various cases, including data from the same simulations used in the training process at time steps different from those of the training data, and data from two other simulations, HIT3 and HIT4, as presented in table 1. The ability of the diffusion model in predicting the HR flow field is evaluated for three upsampling factors of 4, 8 and 16.

We will also compare the performance of the diffusion model presented in this study with bicubic interpolation, a CNN model, a U-Net-based model and the conditional DDPM designed for SR (Saharia *et al.* 2022) to investigate the improvements achieved by using the developed diffusion model. The diffusion model presented in this study and the conditional DDPM (Saharia *et al.* 2022) are referred to as residual-based diffusion model (DM) and DDPM in this study, respectively. The CNN model we use for comparison includes only convolutional layers with the kernel size of 3×3 and upsampling layers, with leaky rectified linear unit (LReLU) as the activation function. A schematic of the CNN model used in the present study is shown in figure 5. The U-Net architecture used

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

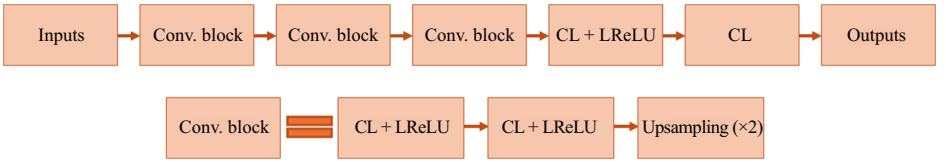


Figure 5. A schematic of the CNN model architecture for the upsampling factor of 8. Here, CL denotes a convolutional layer.

for comparison in this study is adapted from the original U-Net model (Ronneberger *et al.* 2015) for the SR task with some modifications. It takes three-channel inputs and outputs, where the inputs correspond to bicubic-interpolated LR flow fields and the outputs are HR flow fields. Downsampling is performed using average pooling, while upsampling is achieved through bicubic interpolation to ensure smoother reconstruction and improved preservation of flow structures.

3.1. *Qualitative analysis of reconstructed HR flow fields*

The velocity distribution for HIT2 is presented in figure 6. In this case, the input size is 512^2, which differs from the input size of 256^2 used during the training process. As seen in figure 6, all models are able to predict the flow field roughly well for the upsampling factor of 4. As the upsampling factor increases, bicubic interpolation and the CNN model generate blurry HR reconstructions and fail to capture the small-scale flow structures. The DDPM tends to underpredict or overpredict the flow field for all upsampling factors. This issue with DDPM has been observed in the previous study by Qi & Ma (2024) as well. The U-Net model shows better performance than bicubic interpolation and the CNN model although it also predicts almost a blurry flow field for the upsampling factor of 16. However, the residual-based DM can generate a clear HR flow field with no blurriness, capable of capturing small-scale features, even at high upsampling factors.

To further investigate the performance of the model, it is tested on two additional cases, HIT3 and HIT4. As seen in table 1, the viscosities in HIT3 and HIT4 are significantly smaller and larger, respectively, compared with those of the training data, making them challenging for the model and providing a rigorous test of its generalisation capability. The velocity distribution of HIT3 at the upsampling factor of 8 is presented in figure 7. As seen in this figure, although HIT3 differs from the training data, it is not difficult for the models to achieve reasonable accuracy in predicting the velocity field. In fact, due to its lower Reynolds number, there are fewer complex, chaotic structures compared with the higher Reynolds number case, HIT2, and it exhibits smoother, more predictable behaviour. This makes it easier for the models to capture and upscale the flow field without blurriness. Nevertheless, even in this case, the velocity predictions by bicubic interpolation, the CNN and the U-Net models is over-smoothed, and the residual-based DM achieves a slightly better prediction.

The results of testing the models on HIT4 data are shown in figure 8. To make the differences clearer, the results are shown for a quarter of the domain. In this case, the Reynolds number is much higher than those of the training data, and the grid size is also four and two times larger than those of the training data, HIT1 and HIT2, respectively. As seen in figure 8, the models are insensitive to the grid size, as the input data with the size of 1024^2 are directly fed into the CNN and U-Net models, and residual-based DM, while they have been trained on the input size of 256^2, and there is no noticeable change in model performance compared with HIT2 in figure 6. The developed models show satisfactory generalisability to the cases with the conditions completely different from those of the

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

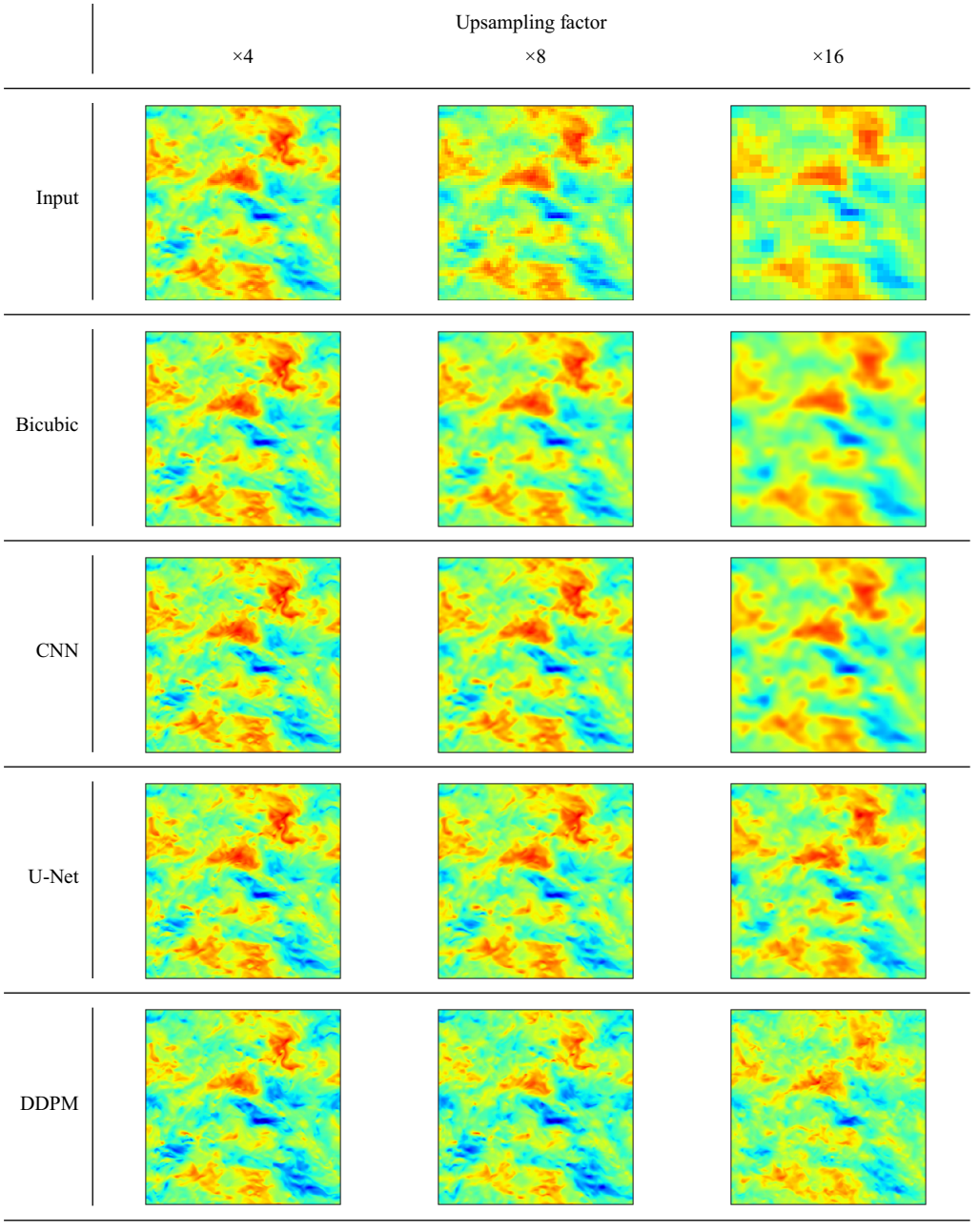


Figure 6. For caption see next page.

training data. One reason for this decent generalisability may be the use of data from different simulations in the training dataset, which can help the model learn the features of HIT under various conditions. By comparing the velocity distribution predicted by the diffusion model with those of bicubic interpolation, CNN and U-Net models, we can infer that, unlike the other models, the residual-based DM is capable of capturing small flow structures without any blurriness. Since HIT4 is a high Reynolds number flow and is more

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

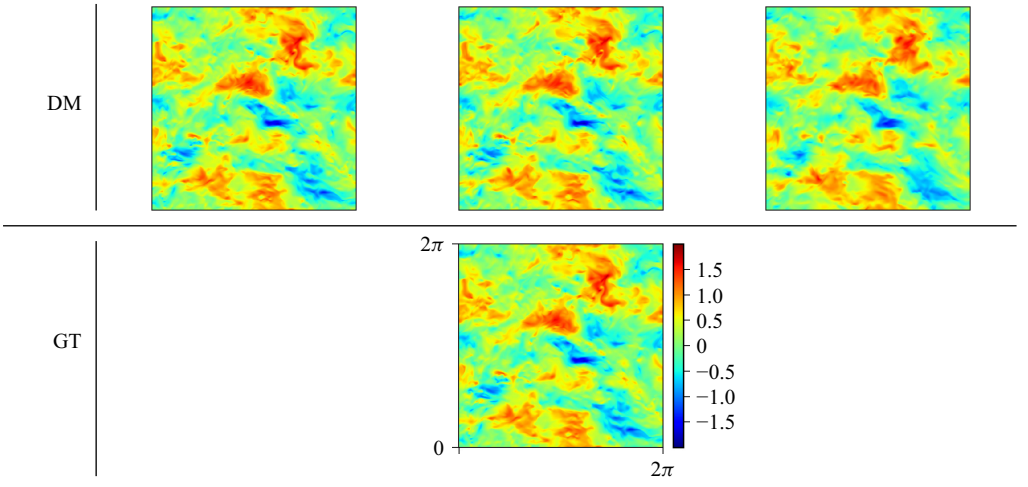


Figure 6. (cntd). Reconstructed instantaneous velocity field (u) from the LR input for HIT2. Here, DM and GT denote the residual-based diffusion model and ground truth (DNS), respectively.

challenging than HIT3, the superior performance of the residual-based DM compared with bicubic interpolation, CNN and U-Net models becomes more evident.

To assess the performance of the residual-based DM in more detail, we compute the vorticity from the predicted velocity field. Figure 9 shows the vorticity field (ω_z) for HIT2 across different upsampling factors. Compared with the velocity field, the difference between the residual-based DM and the other approaches, bicubic interpolation, the CNN and U-Net models and DDPM, becomes clearer when we compare the vorticity. As shown in figure 9, due to the CNN model's tendency to generate smooth outputs, it cannot predict the vorticity field at high upsampling factors. Especially, for upsampling factor of 16, the velocity predicted by bicubic interpolation and the CNN model fails to capture sufficient details, resulting in vorticity fields that resemble the LR input rather than the HR ground truth. The U-Net model outperforms bicubic interpolation and the CNN model, capturing more flow details and predicting a vorticity field with less excessive smoothing. On the other hand, DDPM achieves improved performance as it avoids over-smoothing and produces a sharper vorticity field than the U-Net model. However, the residual-based DM shows robust performance across various upsampling factors and consistently outperforms the other models, confirming its ability to capture small flow structures effectively.

We further evaluate the models' ability to predict the vorticity field by applying them to the test cases HIT3 and HIT4, as presented in figure 10. Similar to figure 8, the results for HIT4 are shown for a quarter of the domain to highlight the differences more clearly. As seen in figure 10, the residual-based DM demonstrates strong capability in predicting the vorticity field for test data with either smaller or larger viscosity than that of the training data. It noticeably outperforms bicubic interpolation, the CNN and U-Net models, especially at the higher upsampling factor of 8. Furthermore, its performance does not degrade even when tested on HIT data with completely different specifications from those of the training data.

To better show the differences between the CNN model and the residual-based DM, the magnified vorticity fields for HIT3 and HIT4 at the upsampling factor of 8 are presented in figures 11 and 12, respectively. As seen in figure 11, the vorticity obtained using the

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

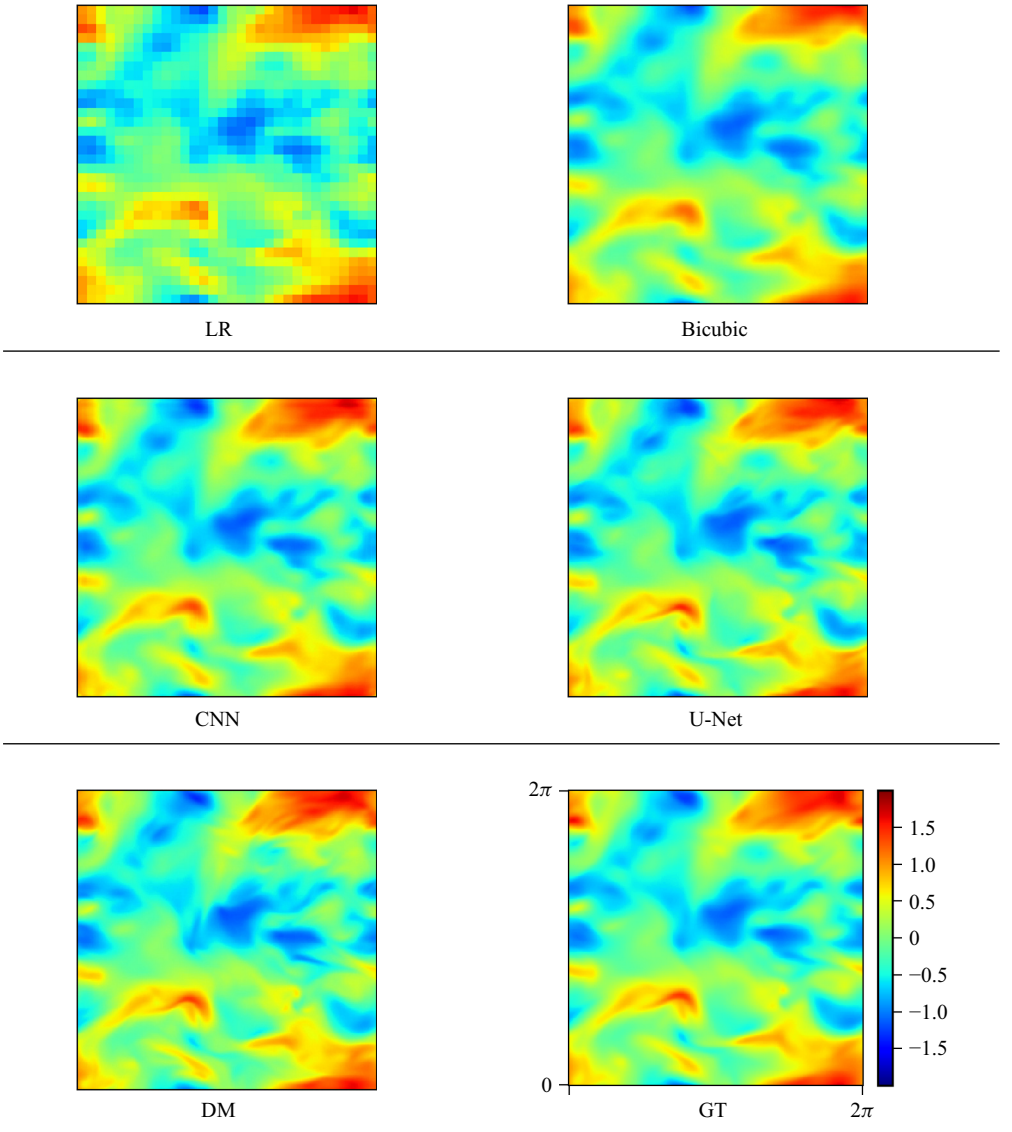


Figure 7. Reconstructed instantaneous velocity field (u) from the LR input for HIT3 (upsampling factor = 8). Here, DM and GT refer to the residual-based diffusion model and ground truth (DNS), respectively.

reconstructed HR velocity from the CNN model is patchy in some regions. The CNN model tends to smooth the predicted velocity, which limits its ability to capture the subtle gradients necessary for computing derivative-based quantities like vorticity. As a result, even small inconsistencies in the velocity field can be amplified when calculating vorticity, leading to a patchy or noisy vorticity field. On the other hand, the vorticity predicted by the residual-based DM does not show patchiness or noise, demonstrating its ability to capture the fine details and intricate features of the velocity field, which enables it to have a more consistent prediction.

In figure 12, we can clearly observe the notable difference between the vorticity fields obtained using the predicted velocity from the CNN model and the residual-based DM for the upsampling factor of 8 in the high Reynolds number flow case,

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

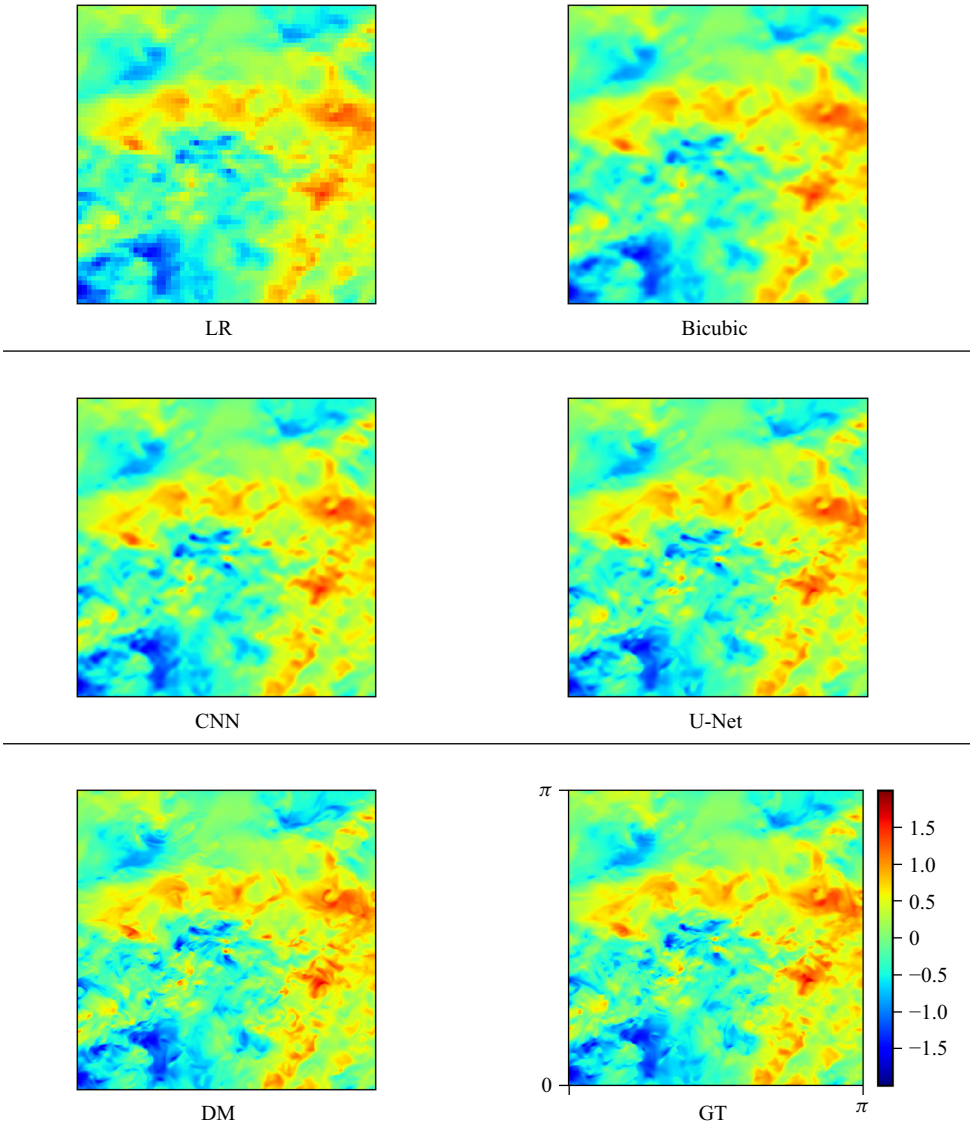


Figure 8. Reconstructed instantaneous velocity field (u) from the LR input for HIT4 (upsampling factor = 8). Here, DM and GT denote the residual-based diffusion model and ground truth (DNS), respectively.

HIT4. This is a challenging case due to the high Reynolds number and high upsampling factor, which make it difficult to accurately predict the vorticity. Since the vorticity is calculated from the predicted velocity field, any discrepancy in the velocity prediction can have a substantial negative impact on the vorticity results. Unlike the CNN model, the residual-based diffusion model's performance remains roughly unaffected by the high upsampling factor. Even in this case, the residual-based DM produces clear results without patchiness or noise. In contrast, the CNN model shows a considerable drop in performance. It struggles to reconstruct small-scale turbulence, and consequently, the vorticity becomes both patchy and blurry, leading to a noticeable decrease in accuracy. Therefore, despite being a challenging case, the residual-based DM still performs well.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

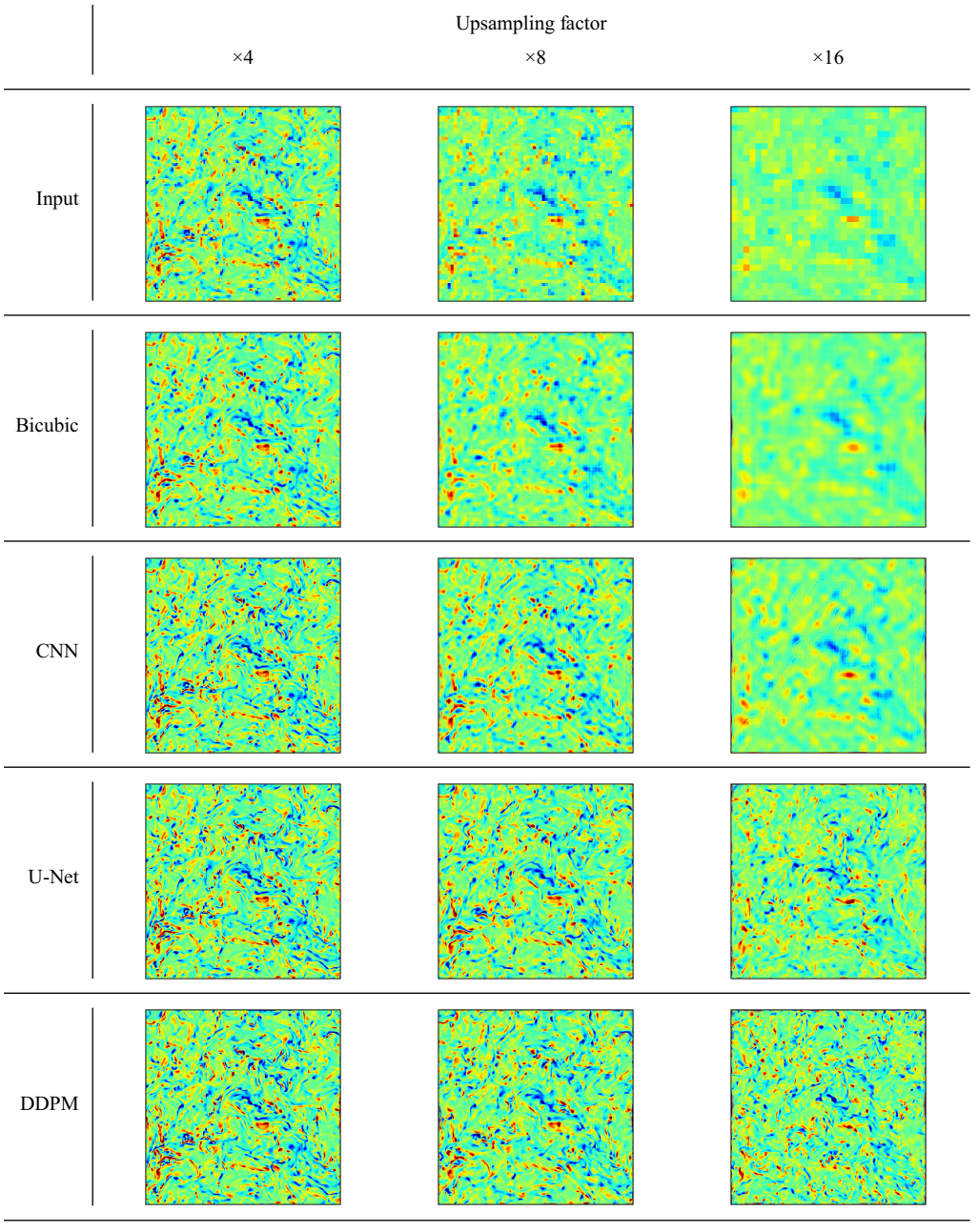


Figure 9. For caption see next page.

3.2. *Performance and computational cost metrics for turbulent flow SR*

We use various metrics to quantitatively evaluate the model's predictions, including MSE and LPIPS, which were explained in § 2.3, as well as the correlation coefficient. The MSE and LPIPS measure the pixel-wise error and perceptual differences between the predicted and ground truth flow fields, respectively. Correlation coefficient measures how closely the model prediction and ground truth are related and is defined as

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

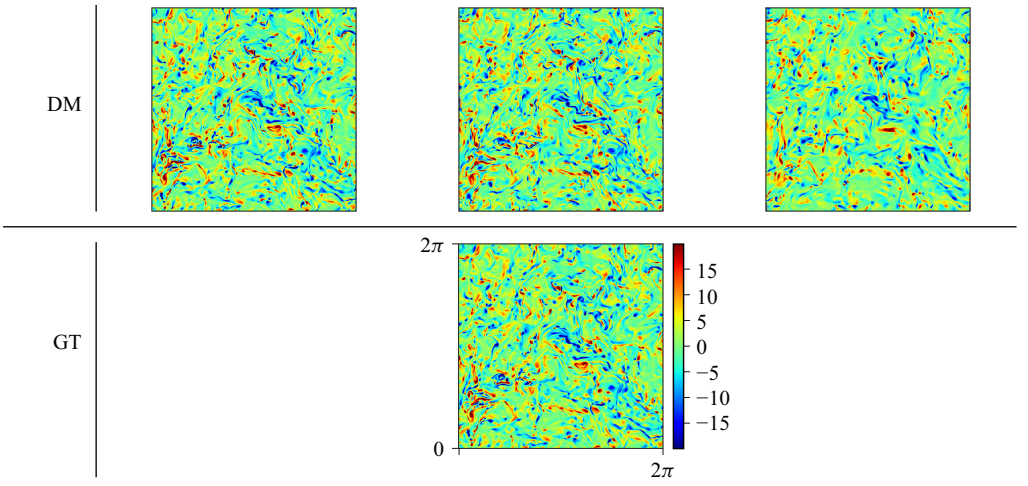


Figure 9. (cntd). Reconstructed instantaneous vorticity field (ω_z) for HIT2. Here, DM and GT refer to the residual-based diffusion model and ground truth (DNS), respectively.

$$C.C. = \frac{\langle(\hat{u}_i - \langle\hat{u}_i\rangle)(u_i - \langle u_i\rangle)\rangle}{\sqrt{\langle(\hat{u}_i - \langle\hat{u}_i\rangle)^2\rangle\langle(u_i - \langle u_i\rangle)^2\rangle}}, \tag{3.1}$$

where $\hat{u}_i$ and u_i denote the velocity predicted by the model and the ground truth velocity obtained from DNS, respectively, and $\langle u_i\rangle$ denotes the mean velocity, which is zero for HIT.

The results of comparing the metrics for velocity (u) and vorticity (ω_z) for HIT2 are shown in tables 2–4. As seen in these tables, the residual-based DM outperforms DDPM in both metrics, the correlation coefficient and MSE, across all upsampling factors. At the upsampling factor of 4, the residual-based DM shows a higher correlation coefficient and smaller MSE than bicubic interpolation and the CNN model for both velocity and vorticity. Overall, its performance is also comparable to that of the U-Net model. However, as shown in tables 3–4, at higher upsampling factors, the residual-based DM generally shows a lower correlation coefficient and a larger MSE for velocity compared with the CNN and U-Net models. This issue has been observed in the study by Kim *et al.* (2021) as well, where they analysed the performance of their GAN-based SR model quantitatively. Their results showed that even for the upsampling factor of 4, the GAN-based model has a larger MSE than bicubic interpolation, although it outperforms bicubic interpolation both qualitatively and statistically.

In fact, in turbulent flows, where the LR input can correspond to multiple HR flow fields, especially for higher upsampling factors, the CNN and U-Net models tend to produce the mean of possible HR flow fields, which makes the output smooth and can lead to smaller MSE. However, the CNN and U-Net models result in the blurry HR reconstruction as the upsampling factor increases and the model will not be able to capture the inherent multi-scale variability in turbulence. In contrast to the CNN and U-Net models, which are regression-based approaches, the residual-based DM utilises a probabilistic generative framework. This means that, instead of regressing to an averaged prediction, it samples from the distribution of possible HR flow fields conditioned on the LR input. This enables the residual-based DM to reconstruct more realistic flow structures and recover the inherent multi-scale structure of turbulent flows, even though the individual samples may have larger local discrepancies, including larger MSE. Additionally, correlation coefficient

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

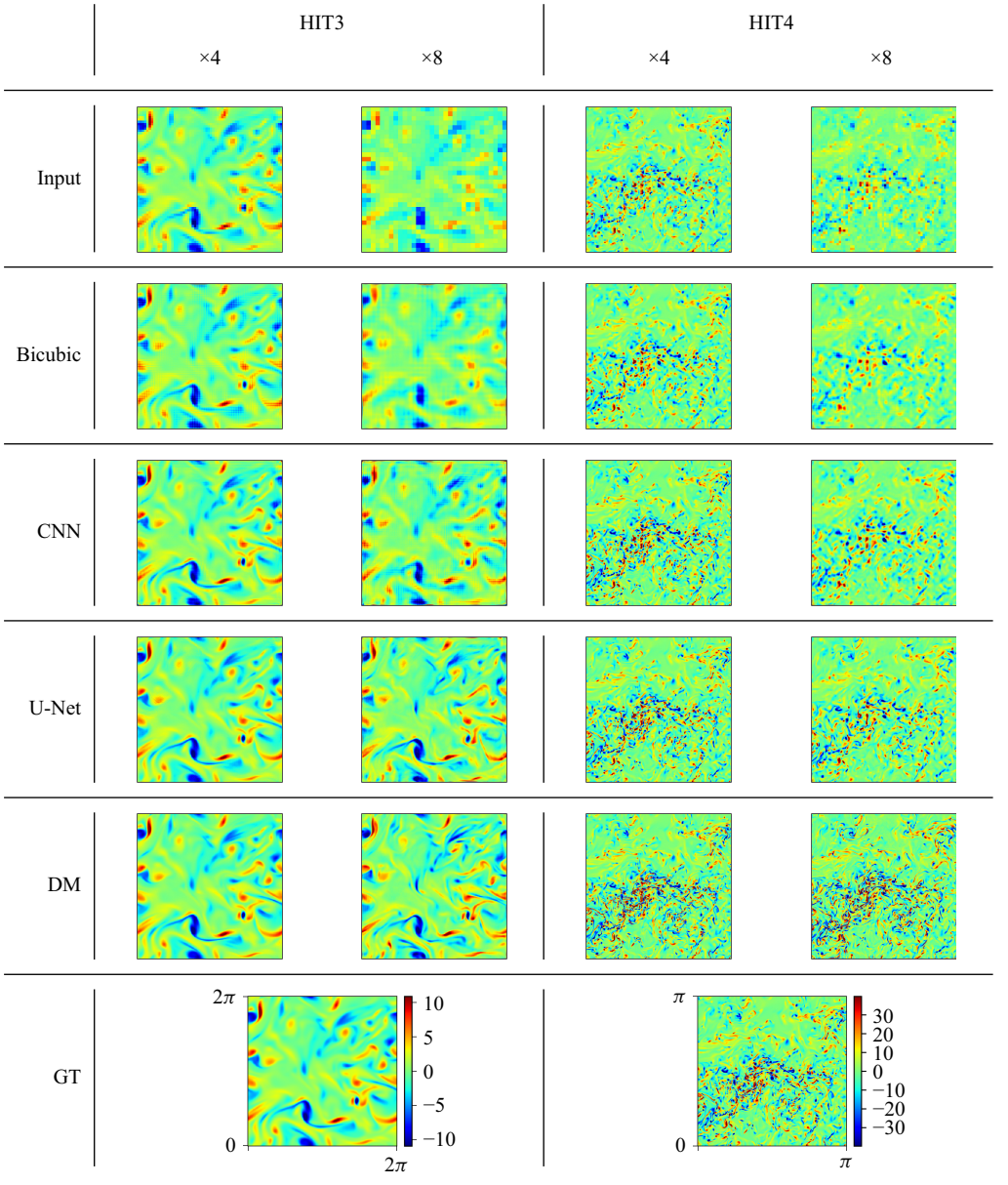
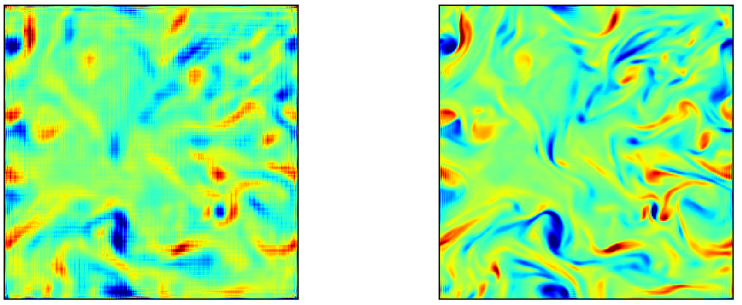


Figure 10. Reconstructed instantaneous vorticity field (ω_z) for HIT3 and HIT4. Here, DM and GT represent the residual-based diffusion model and ground truth (DNS), respectively.

and MSE are pixel-wise metrics that penalise point-wise differences but not structural similarity. Therefore, they are sensitive to small shifts, which may not be physically important in fluid dynamics. For instance, high-quality features like vortices can slightly shift while still being physically correct; however, these metrics interpret such shifts as large errors.

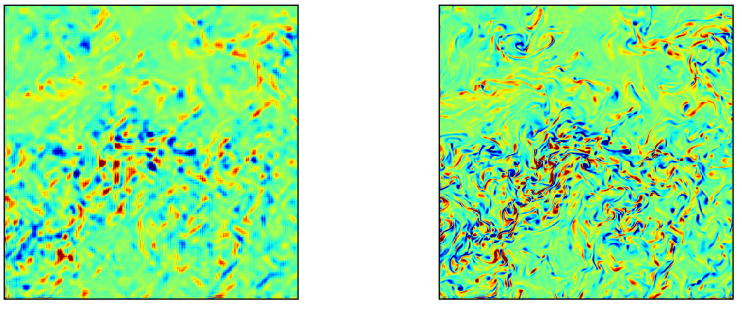
As shown for vorticity in tables 2–4, similar trends to those observed for velocity are seen across the upsampling factors. This inconsistent behaviour at different upsampling factors for both velocity and vorticity makes the correlation coefficient and MSE unreliable

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press



CNN DM

Figure 11. Zoomed-in reconstructed instantaneous vorticity field for HIT3 (upsampling factor = 8). Here, DM denotes the residual-based diffusion model.



CNN DM

Figure 12. Zoomed-in reconstructed instantaneous vorticity field for HIT4 (upsampling factor = 8). Here, DM represents the residual-based diffusion model.

	Velocity (u)					Vorticity (ω_z)				
Metric	Bicubic	CNN	U-Net	DDPM	DM	Bicubic	CNN	U-Net	DDPM	DM
C.C.	0.9829	0.9988	0.9992	0.9978	0.9992	0.7709	0.9364	0.9610	0.9156	0.9584
MSE	0.0025	0.0002	0.0001	0.0071	0.0001	0.0018	0.0005	0.0003	0.0007	0.0004
LPIPS	0.1751	0.0856	0.0477	0.0954	0.0452	0.3559	0.2261	0.1190	0.2204	0.1150

Table 2. Metrics for reconstructed velocity (u) and vorticity (ω_z) fields at upsampling factor = 4 for HIT2. Here, DM denotes the residual-based diffusion model.

and uninterpretable for analysing the performance. However, unlike these contradictory results in metrics, we can clearly see the better performance of the residual-based DM compared with the other models for both velocity and vorticity in figures 6 and 9 at all upsampling factors. On the whole, for quantitative analysis, statistical and spectral measures like the probability distribution and energy spectrum are more effective and informative than pixel-wise metrics, especially for assessing physical accuracy in turbulent flows. A detailed and comprehensive analysis using these measures is presented in § 3.4, where the accuracy and advantages of the residual-based DM are more clearly demonstrated.

Interestingly, when comparing LPIPS, a metric that uses deep features to assess similarity, we see that the residual-based DM consistently shows smaller error than the

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

	Velocity (u)					Vorticity (ω_z)				
Metric	Bicubic	CNN	U-Net	DDPM	DM	Bicubic	CNN	U-Net	DDPM	DM
C.C.	0.9377	0.9880	0.9869	0.9723	0.9843	0.5030	0.7244	0.7107	0.6338	0.6803
MSE	0.0088	0.0017	0.0019	0.0128	0.0023	0.0033	0.0021	0.0023	0.0031	0.0029
LPIPS	0.4252	0.3182	0.2672	0.3095	0.2588	0.5609	0.5352	0.3552	0.4106	0.3261

Table 3. Metrics for reconstructed velocity (u) and vorticity (ω_z) fields at upsampling factor = 8 for HIT2. Here, DM represents the residual-based diffusion model.

	Velocity (u)					Vorticity (ω_z)				
Metric	Bicubic	CNN	U-Net	DDPM	DM	Bicubic	CNN	U-Net	DDPM	DM
C.C.	0.8424	0.9434	0.9214	0.8884	0.9159	0.2752	0.3952	0.2978	0.2398	0.3114
MSE	0.0213	0.0080	0.0110	0.0158	0.0120	0.0041	0.0038	0.0050	0.0061	0.0053
LPIPS	0.5785	0.5218	0.4481	0.4498	0.4269	0.6116	0.6173	0.5112	0.4852	0.4439

Table 4. Metrics for reconstructed velocity (u) and vorticity (ω_z) fields at upsampling factor = 16 for HIT2. Here, DM represents the residual-based diffusion model.

other models for different upsampling factors. This is because LPIPS is sensitive to the perceptual quality and structural details in the flow field rather than just point-wise differences. By sampling from the distribution of HR flow fields, the residual-based DM can manage the reproduction of local features of turbulent flows remarkably better than the CNN model. This means that despite larger MSE at high upsampling factors, the velocity field reconstructed by the DM is perceptually more promising and contains more physically relevant details than those predicted by bicubic interpolation, the CNN and U-Net models, which tend to generate overly smoothed outputs.

To evaluate the prediction of turbulence parameters by the models, we examine the Taylor microscale (λ), energy dissipation rate (ε) and root-mean-square (r.m.s.) streamwise velocity (u_{rms}) for the reconstructed flow fields, where λ is defined as

$$\lambda = (15\nu u'^2/\varepsilon)^{\frac{1}{2}}, \tag{3.2}$$

where

$$u' = \left(\frac{\langle u_i u_i \rangle}{3}\right)^{1/2}, \tag{3.3}$$

and energy dissipation rate is

$$\varepsilon = 2\nu \langle S_{ij} S_{ij} \rangle, \tag{3.4}$$

where $S_{ij} = (1/2)((\partial u_i/\partial x_j) + (\partial u_j/\partial x_i))$ denotes the strain rate. As presented in tables 5–7, across all upsampling factors, the residual-based DM consistently provides predictions that are close to the ground truth (GT), outperforming bicubic interpolation, the CNN and U-Net models, and DDPM. For the Taylor microscale, residual-based DM achieves values very close to GT, at all upsampling factors, while other models, including CNN and U-Net models, overpredict λ. Similarly, the predicted energy dissipation rate ε from residual-based DM aligns well with GT, whereas the CNN and U-Net models significantly underpredict ε at higher upsampling factors. For u_{rms}, the residual-based DM shows excellent agreement with GT, while the other models generally exhibit larger

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

Variable	LR	Bicubic	CNN	U-Net	DDPM	DM	GT
λ	0.32379	0.31907	0.28403	0.28005	0.24878	0.26888	0.27433
ε	0.03220	0.03324	0.04255	0.04338	0.05183	0.04758	0.04571
u_{rms}	0.32152	0.32230	0.32688	0.32405	0.30549	0.32759	0.32764

Table 5. Turbulence characteristics for reconstructed flow fields at upsampling factor = 4. Here, DM refers to the residual-based diffusion model.

Variable	LR	Bicubic	CNN	U-Net	DDPM	DM	GT
λ	0.42027	0.40523	0.32834	0.29925	0.25722	0.26844	0.27433
ε	0.01836	0.01976	0.03135	0.03722	0.04474	0.04465	0.04571
u_{rms}	0.30883	0.30910	0.32186	0.31741	0.28195	0.30645	0.32764

Table 6. Turbulence characteristics for reconstructed flow fields at upsampling factor = 8. Here, DM denotes the residual-based diffusion model.

Variable	LR	Bicubic	CNN	U-Net	DDPM	DM	GT
λ	0.63786	0.52033	0.42810	0.32994	0.25299	0.30426	0.27433
ε	0.00726	0.01088	0.01737	0.02937	0.03227	0.03693	0.04571
u_{rms}	0.28149	0.28055	0.30312	0.30448	0.19669	0.32560	0.32764

Table 7. Turbulence characteristics for reconstructed flow fields at upsampling factor = 16. Here, DM represents the residual-based diffusion model.

deviations, especially DDPM, which underpredicts fluctuations at all upsampling factors. Overall, these results demonstrate that the residual-based DM accurately captures the essential turbulence characteristics and provides superior performance compared with other SR approaches. In Appendix A, the relative errors of turbulence characteristics of the reconstructed flow fields have been provided as well to further facilitate the comparisons.

To assess the complexity and computational cost of the models, we compare the total number of parameters, the number of floating-point operations (FLOPs) and the computational requirements for training and inference. Giga FLOPs (GFLOPs) are reported per forward pass for CNN and U-Net models, and for the entire inference process (including all diffusion steps) for DDPM and residual-based DM. Here, GFLOP represents the total number of computational operations required to process the input, while the training and inference times reflect the actual cost of computation for model development and deployment. All computations for both training and inference were performed on a single NVIDIA RTX A6000 GPU to evaluate the computational cost. The results are shown in table 8. As shown in this table, the residual-based DM exhibits a higher training time and parameter count compared with the CNN and U-Net models, and DDPM due to its deeper architecture. However, its inference time and GFLOPs remain significantly lower than those of DDPM, demonstrating its computational efficiency as a diffusion model. This efficiency stems from employing a reduced-time-step diffusion process, which requires only 15 time steps ($T = 15$) during the reverse process in this study. In contrast, standard DDPM adapted for SR (Saharia *et al.* 2022) commonly requires 1000–2000 time steps. The number of time steps for DDPM used in this study is 1000. The reduced number

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

	Bicubic	CNN	U-Net	DDPM	DM
Training time (s)	—	0.1859	0.9176	0.8559	2.2823
Inference time (s)	0.0001	0.0009	0.0588	78.9656	2.4378
Parameter count (M)	—	0.298	100.903	92.557	114.653
GFLOPs	0.022	3.801	3372.1	1426900	77449.8

Table 8. Computational time and number of parameters for models for the upsampling factor of 16. The training time indicates the computational time required per iteration for a batch size of 16, while the inference time represents the time needed to generate predictions for a single sample (batch size of 1) using the trained model; GFLOPs measures the total number of operations, in billions and are reported per forward pass for CNN and U-Net models, and for the entire inference process (including all diffusion steps) for DDPM and residual-based DM.

of sampling steps drastically lowers the inference cost while maintaining reconstruction quality, making residual-based DM a low-inference-cost diffusion model.

3.3. *Deep feature-based analysis of super-resolved flow fields*

To assess perceptual similarity and feature representation, we compare the features extracted for vorticity from different layers of VGG16, used for the LPIPS calculation, in figure 13.

In the context of computer vision, in CNN-based models like VGG16, earlier layers capture low-level features such as edges, textures and colours, while as the network deepens, the layers progressively capture more complex, high-level features like shapes, objects and semantic structures, enabling the network to understand more abstract aspects of the image. When applied to vorticity fields, these extracted features provide a hierarchical representation of spatial variations in the data at different network depths.

Comparing the feature maps extracted from the SR models with those of DNS can therefore be considered a perceptual, multi-scale comparison of structural similarity. The similarity in patterns between an SR model and DNS feature maps indicates that the model reproduces spatial structures consistently in the feature space of the network. To obtain the feature maps for vorticity, since it has a single channel, the field is repeated across three channels to form a 3-channel input tensor before passing it to the VGG16 network. For the visualisation purpose, a single representative channel is displayed as a two-dimensional contour map to qualitatively illustrate the spatial features learned by each model. This visualisation allows assessment of how the reconstructed flow fields preserve or deviate from the DNS feature representations at different network depths.

In figure 13, the second block refers to the feature maps extracted from the second block of the VGG16 network, corresponding to relatively earlier convolutional layers. The third block corresponds to deeper layers of VGG16, which encode more spatially aggregated and structurally complex activations. The blocks of the VGG16 network are shown in figure 2.

As seen in figure 13, for the upsampling factor of 4, the models can capture the features extracted from the second block, which correspond to the earlier layers, fairly well; however, in the deeper layers, the extracted features from bicubic interpolation and the CNN model are not similar to those of the DNS. This discrepancy becomes more pronounced when we compare the extracted features for the upsampling factor of 8, as we see that the extracted features from both the second and third blocks are not similar to those of the DNS. However, we observe that for all tested cases, the extracted features from the residual-based DM are closely similar to those of the DNS, indicating improved structural consistency with the reference solution in feature space. The U-Net model and DDPM

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

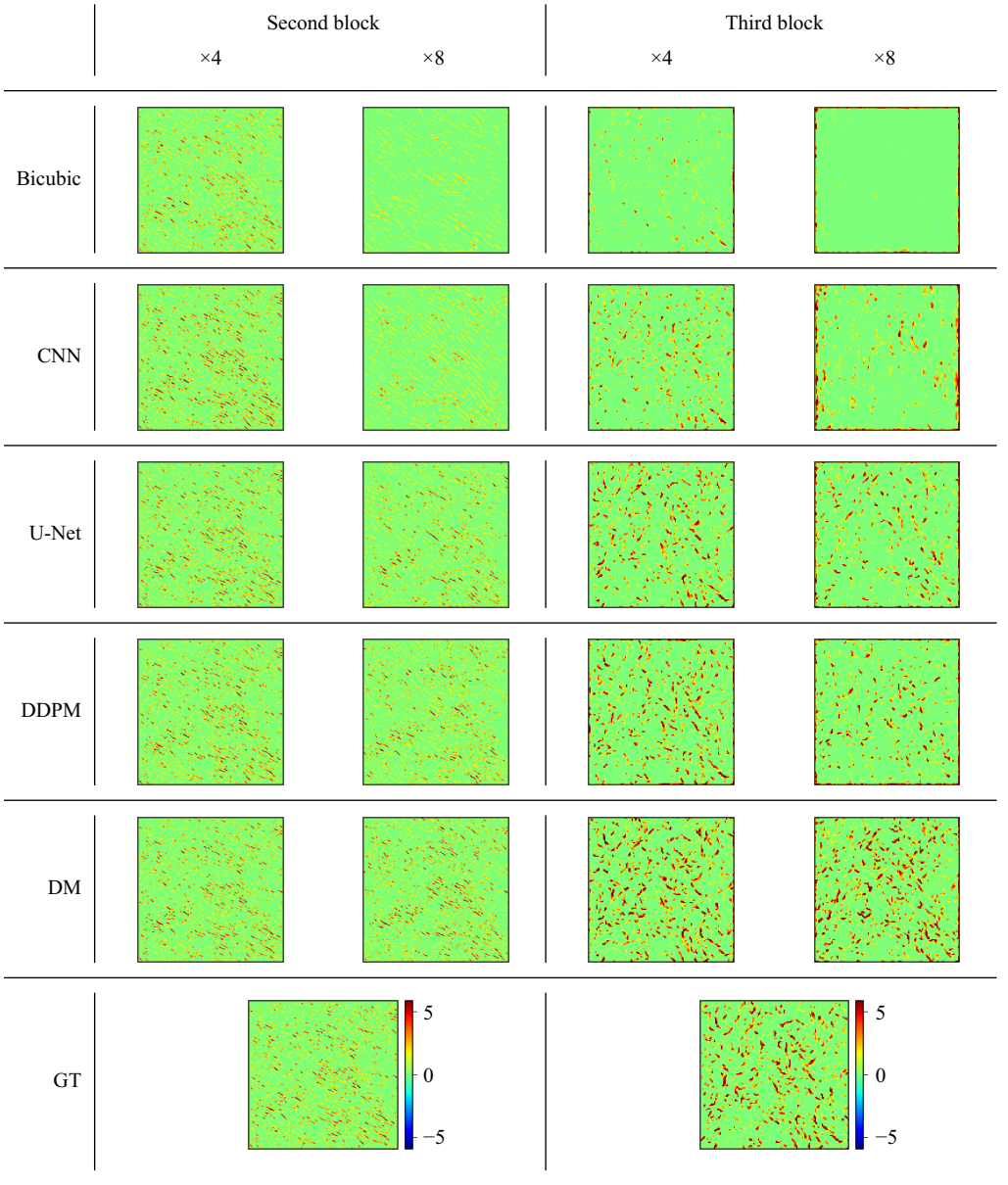


Figure 13. Feature extraction (single-channel visualisation) of vorticity (ω_z) for HIT2 at upsampling factors of 4 and 8. Here, DM and GT represent the residual-based diffusion model and ground truth (DNS).

show better performance than the CNN model although they still underperform compared with the residual-based DM. We quantify the comparison in table 9. For the extracted features from the second and third blocks at the upsampling factor of 4, the residual-based DM generally achieves higher correlation coefficients and lower MSE values than the other models, while remaining competitive with U-Net in terms of MSE. These results quantitatively demonstrate the strong feature extraction capability of the residual-based diffusion model. At the upsampling factor of 8, the correlation coefficient and MSE exhibit trends that differ from those observed at the upsampling factor of 4 across the evaluated feature blocks. Although the visualisations, similar to those for vorticity, indicate superior

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

Ratio	Metric	Bicubic	CNN	U-Net	DDPM	DM
4	$C.C._2$	0.5995	0.8055	0.8816	0.7811	0.8828
	MSE_2	0.3730	0.1961	0.1235	0.2308	0.1281
	$C.C._3$	0.5610	0.7588	0.8623	0.7403	0.8722
	MSE_3	1.0810	0.6547	0.3940	0.7515	0.4078
8	$C.C._2$	0.4220	0.4341	0.5150	0.4371	0.4918
	MSE_2	0.4867	0.5067	0.4726	0.5704	0.5595
	$C.C._3$	0.2028	0.3174	0.4637	0.3896	0.4937
	MSE_3	1.6695	1.6724	1.4245	1.6806	1.6146

Table 9. Metrics for HIT2 at upsampling factors of 4 and 8. The subscripts 2 and 3 refer to the second and third blocks, respectively. Here, DM represents the residual-based diffusion model.

performance of the residual-based DM, this superiority is not consistently reflected in the correlation coefficient and MSE values. This behaviour is consistent with the vorticity results shown in figures 6 and 9 and in tables 2–4, further highlighting the inconsistency of these metrics in assessing the relative performance of the models as the upsampling factor increases.

3.4. *Statistical and spectral analysis of turbulent structures*

The probability density function (PDF) for the velocity and vorticity is shown in figures 14–15. As seen in figure 14, at the upsampling factor of 4, the difference in the PDF of velocity predicted by different models is less apparent. However, as the upsampling factor becomes higher, the differences become more visible. The bicubic interpolation tends to behave similarly to the LR input in predicting the PDF. The PDF obtained by the CNN and U-Net models lies between the GT and LR inputs. The most significant difference is seen for the upsampling factor of 16, where bicubic interpolation, the CNN and U-Net models noticeably overpredict the peak of the PDF, which occurs near zero for all velocity components, and underpredict the PDF in the other regions. This significant difference at high upsampling factors is due to the smooth velocity predicted by bicubic interpolation and the CNN and U-Net models. On the other hand, we see that the PDF of residual-based DM follows that of GT closely for all velocity components and upsampling factors. The results of the PDF, which assess the models' capabilities statistically, are consistent with the results presented for HIT2 for the velocity distribution earlier in this section. For DDPM, we also see that the PDF of velocity is not predicted well for any upsampling factor and it significantly underperforms compared with the residual-based DM. Such inaccuracy in the prediction of PDF is seen in the study by Shu *et al.* (2023) as well which can be caused by the noise estimation error of DDPM (Shu *et al.* 2023).

By comparing the PDF of vorticity in figure 15, we further understand the weakness of bicubic interpolation, the CNN model, the U-Net model and DDPM in capturing various flow scales and gradients. As seen in this figure, even for the upsampling factor of 4, the predicted vorticity obtained from bicubic interpolation is nearly identical to that of the LR input. On the whole, DDPM is not able to predict the peak of PDF correctly, and the residual-based DM outperforms DDPM. For the CNN and U-Net models, the difference becomes more apparent for the upsampling factor of 8, and the PDF becomes significantly different for the upsampling factor of 16. However, the PDF of vorticity resulting from the residual-based DM matches that of the GT, and only a small over- or underprediction in the peak of PDF can be seen for this model, which statistically confirms the ability of

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

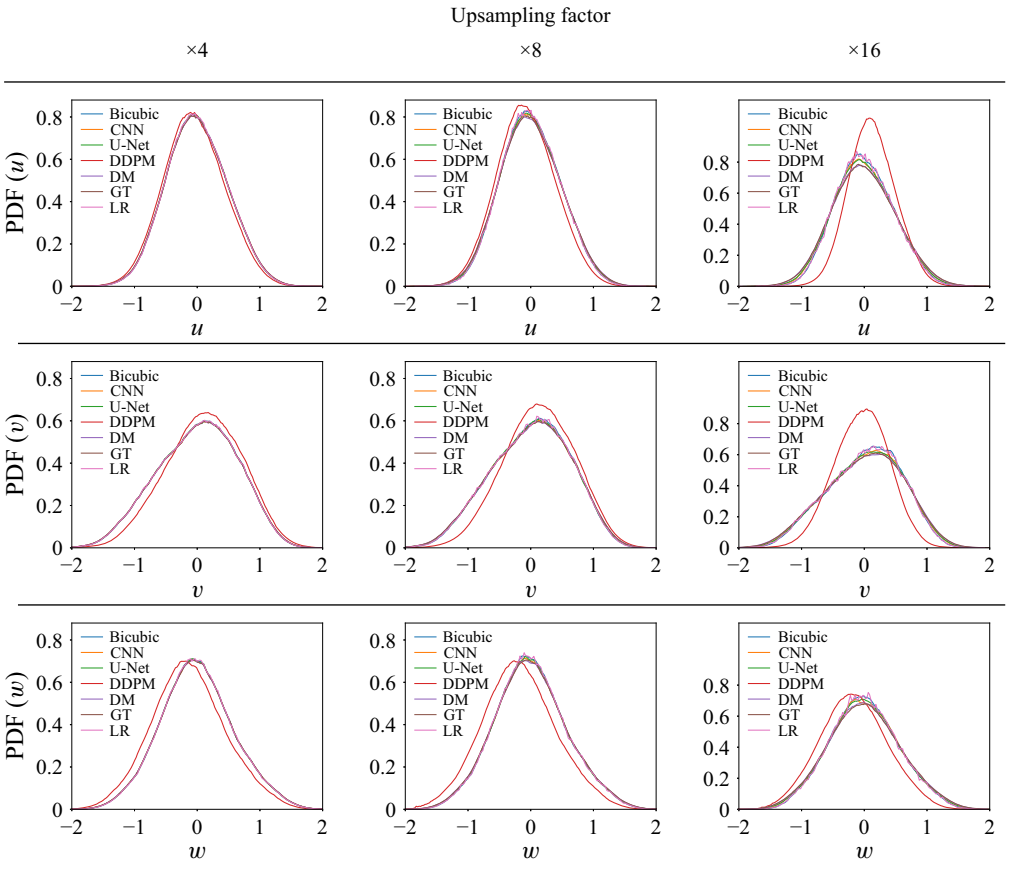


Figure 14. The PDFs of the velocity components, from top, u, v and w, respectively, for HIT2. Here, DM and GT refer to the residual-based diffusion model and ground truth (DNS), respectively.

the residual-based DM in reconstructing the HR flow field using LR inputs at different upsampling factors.

We further evaluate the residual-based diffusion model's performance in reconstructing HR flow fields from the statistical perspective by comparing the energy spectra. As seen in figure 16, DDPM has a reasonable performance in predicting the energy spectrum, although it does not outperform the residual-based DM. At the upsampling factor of 4, bicubic interpolation fails to recover the DNS energy spectrum, while the CNN and U-Net models yield a spectrum lying between that of the LR input and the DNS. However, the U-Net model outperforms the CNN model. In contrast, the residual-based DM accurately reproduces the DNS spectrum. As the upsampling factor increases, both bicubic interpolation and CNN model results increasingly resemble the LR input, indicating their limitations in recovering fine-scale features. The residual-based DM, however, continues to closely match the DNS spectrum even at higher upsampling factors, demonstrating superior statistical accuracy.

This analysis highlights the diffusion model's ability to reconstruct the energy spectrum at high wavenumbers, effectively recovering small-scale structures typically lost in coarse-resolution data. Additionally, accurate reconstruction of the energy spectrum at high wavenumbers suggests reliable enstrophy prediction, reflecting the model's ability to

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

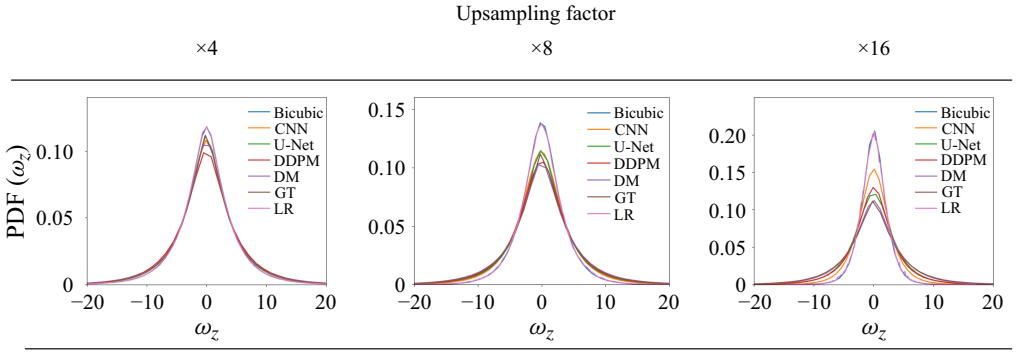


Figure 15. The PDFs of vorticity (ω_z) for HIT2. Here, DM and GT denote the residual-based diffusion model and ground truth (DNS), respectively.

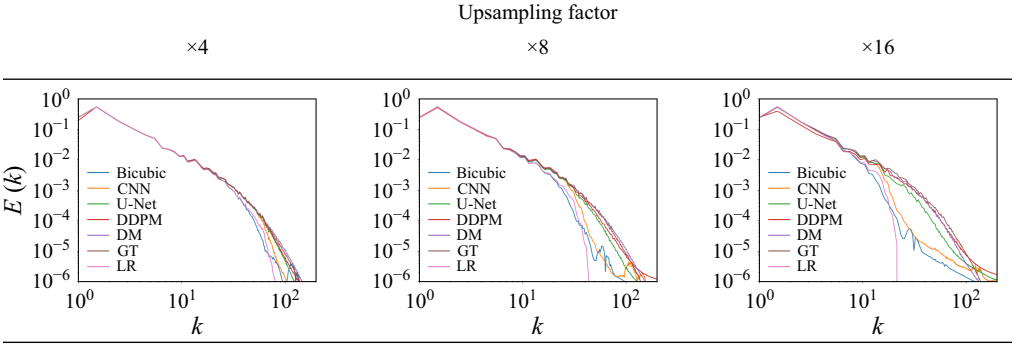


Figure 16. Energy spectrum of HIT2 at upsampling factors of 4, 8 and 16. Here, DM and GT represent the residual-based diffusion model and ground truth, respectively.

capture small-scale vortical structures. Such capability is crucial for turbulence modelling, where accurate representation of fine-scale motions is important for physical processes such as energy loss and the SGS dynamics. Given its capacity to super-resolve turbulent fields in a physically consistent manner, the diffusion model shows strong potential as a data-driven approach for improving SGS stress estimation and enhancing the fidelity of LES.

3.5. *Applicability of model to anisotropic turbulence*

To evaluate the model's performance in anisotropic turbulence, we reconstruct HR flow fields from LR transitional boundary layer data, with the friction Reynolds number (Re_τ) ranging from Re_{τ_i} to Re_{τ_o} (481 to 1024), where the subscripts i and o refer to inflow and outflow, respectively. The HR dataset is obtained from the DNS of boundary layer flow available in the database of Towne *et al.* (2023), where the DNS is carried out by solving the incompressible NS equations numerically. The inputs and outputs of the model, which are the velocity components, are initially normalised using the local friction velocity (u_τ), and for training, each velocity component is further normalised by its maximum absolute value. The LR data are obtained by performing average pooling (Fukami *et al.* 2019; Fukami & Taira 2025). The size of HR and LR flow fields are 128^2 and 16^2, respectively. The two-dimensional snapshots are taken from the $x-y$ plane of the flow

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

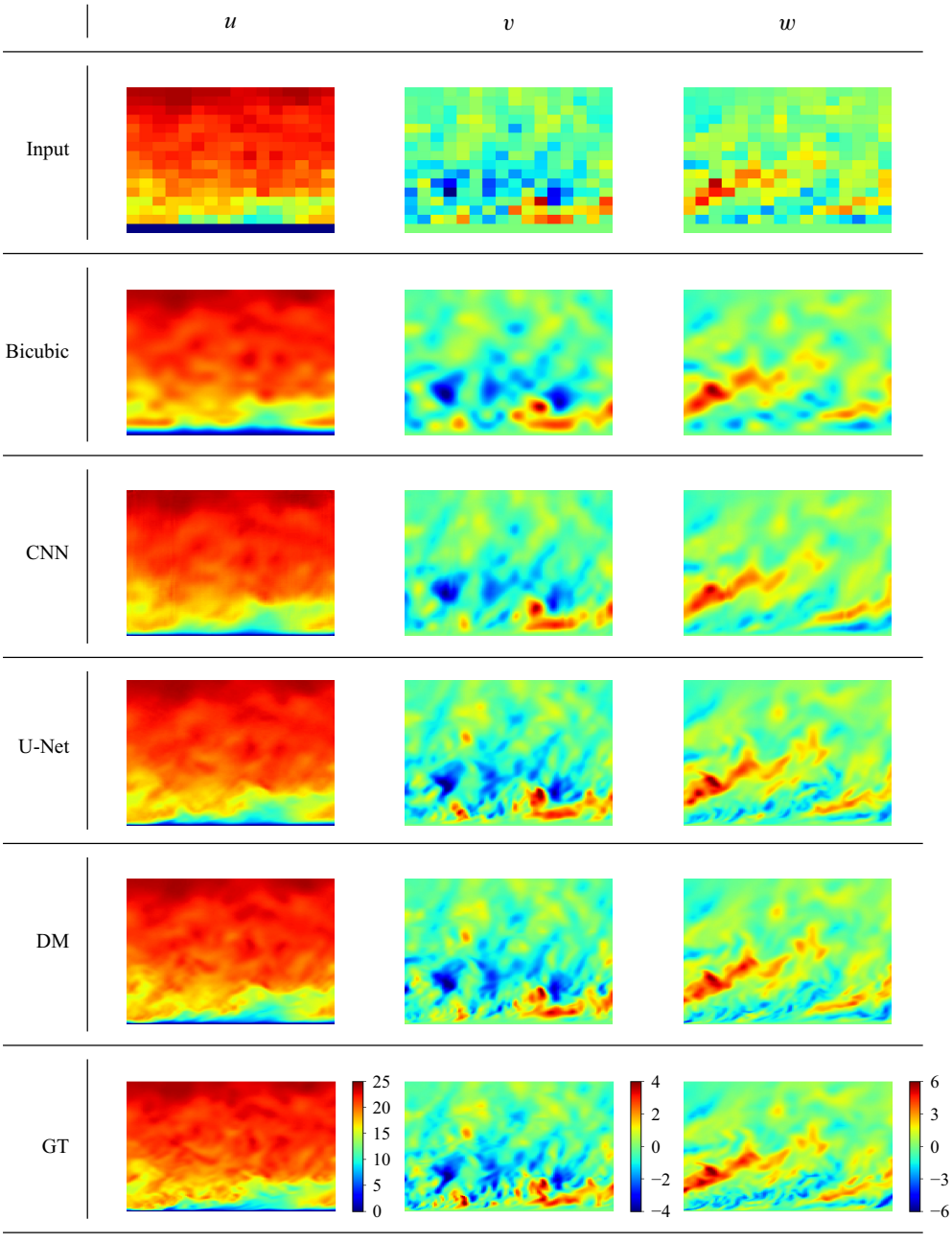


Figure 17. Reconstructed instantaneous velocity field from the LR input for turbulent boundary layer data. Here, DM and GT denote the residual-based diffusion model and ground truth (DNS), respectively.

field, representing the streamwise and wall-normal directions of the boundary layer. The training and validation datasets consist of 26 000 and 1500 samples, respectively, which are randomly sampled over time and streamwise directions.

The velocity distributions for the streamwise, wall-normal and spanwise components are shown in figure 17. As seen in this figure, the residual-based DM accurately predicts the velocity field for all components and, similar to the HIT cases, effectively captures the

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

	u				v				w			
Metric	Bicubic	CNN	U-Net	DM	Bicubic	CNN	U-Net	DM	Bicubic	CNN	U-net	DM
C.C.	0.9613	0.9843	0.9900	0.9904	0.8323	0.8342	0.9147	0.9109	0.8937	0.9021	0.9510	0.9480
MSE	0.0030	0.0008	0.0005	0.0005	0.0159	0.0150	0.0081	0.0085	0.0088	0.0081	0.0041	0.0044
LPIPS	0.2617	0.2244	0.1422	0.1324	0.4148	0.3534	0.1764	0.1878	0.3917	0.3034	0.1579	0.1531

Table 10. Metrics for reconstructed velocity field (u, v, w) for turbulent boundary layer flow. Here, DM represents the residual-based diffusion model.

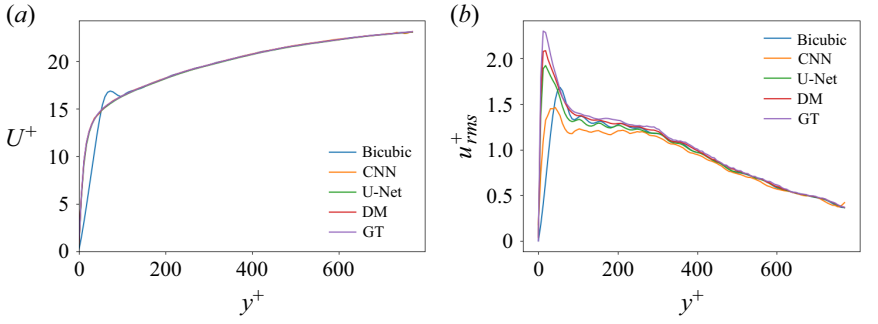


Figure 18. Mean velocity profile (a) and r.m.s. of streamwise velocity fluctuations (b). Here, DM and GT represent the residual-based diffusion model and ground truth (DNS), respectively.

fine-scale flow structures. In contrast, the velocity fields predicted by bicubic interpolation and the CNN model fail to reproduce these flow details and small-scale structures. The U-Net model performs better than the CNN model in capturing flow details but still underperforms compared with the residual-based DM, showing slightly more over- and underpredictions in the flow field.

The mean streamwise velocity and r.m.s. of streamwise velocity fluctuations are presented in figure 18. As seen in the figure, all tested models, except for bicubic interpolation, accurately predict the mean velocity profile. In contrast, the models tend to underpredict the r.m.s. velocity in the near-wall region. Bicubic interpolation fails to reproduce even the location of the r.m.s. peak, while the other models underestimate its magnitude. However, the residual-based DM closely matches the DNS results, providing a more accurate prediction of the r.m.s. velocity compared with the other approaches.

To compare the models' predictions quantitatively, the correlation coefficient, MSE and LPIPS loss are shown in table 10. As seen in this table, across all three components, the residual-based DM consistently achieves higher correlation coefficients and lower MSE than bicubic interpolation and the CNN model, indicating superior accuracy in reproducing the true flow field. The residual-based DM performs similarly to the U-Net model in these metrics, with both achieving high correlation coefficients and low MSE values. However, the results in figure 18 indicate overall superior performance of the residual-based DM. Furthermore, the LPIPS loss demonstrates the residual-based diffusion model's advantage in perceptual quality, with generally lower values compared with the other methods. This suggests that the residual-based DM effectively preserves the fine-scale structures and the overall spatial coherence of the turbulent flow field. A brief test on the applicability of the model to the experimental data is provided in Appendix E.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

4. Conclusions

In the present study, an efficient diffusion model based on ResShift, referred to as residual-based DM, was developed for SR reconstruction of high-fidelity turbulent flow. Unlike the original ResShift model, which operates in latent space, this model was constructed in pixel space to better capture fine details crucial for SR of turbulent flows. The model was trained at upsampling factors of 4, 8 and 16 using data from two simulations with varying flow parameters to enhance generalisability. Additionally, data from both the full domain and a quarter of the domain were included to reduce reliance on domain-specific patterns. To assess the improvements achieved by the residual-based DM, its performance was compared with bicubic interpolation, a CNN model, a U-Net model and DDPM as well.

The qualitative results demonstrated that the residual-based DM successfully generates HR velocity fields without blurriness, even at high upsampling factors, and maintains reasonable accuracy across different upsampling factors. In contrast, bicubic interpolation, the CNN and U-Net models exhibit significant degradation in performance as the upsampling factor increases, particularly for high Reynolds number flows where smaller flow structures are more pronounced. The residual-based DM also outperformed the other methods in producing HR vorticity fields, providing higher perceptual quality and demonstrating better ability in capturing small-scale structures. In the feature-based analysis of the reconstructed flow fields, we further investigated the similarity between the features extracted from the DNS and those extracted from the models. The results showed that the residual-based DM produces features more similar to those of the DNS, indicating better perceptual similarity compared with the other models. Furthermore, the analysis of turbulence parameters demonstrated that the residual-based DM accurately reconstructs the Taylor microscale, energy dissipation rate and r.m.s. streamwise velocity across all upsampling factors, and it outperforms bicubic interpolation, the CNN and U-Net models and DDPM in predicting turbulence characteristics.

We also evaluated the performance of the residual-based DM statistically by computing the PDF of velocity and vorticity. The results showed that, for both velocity and vorticity, the PDF obtained by the residual-based DM closely follows that of the DNS across all upsampling factors. However, bicubic interpolation, the CNN model, the U-Net model and DDPM show strong sensitivity to the upsampling factor and experience a noticeable drop in performance as the upsampling factor increases. Especially, the difference between the residual-based DM and the other approaches becomes more apparent when comparing the PDF of vorticity, as the other methods tend to overpredict the peak and underpredict other regions of the PDF. In contrast, the PDF obtained by the residual-based DM captures the peak reasonably well and matches the DNS in other regions across all upsampling factors, further confirming the strong statistical performance of the residual-based DM. We further assessed the residual-based DM statistically by comparing the energy spectrum. The results showed that, unlike the other models used for comparison, the residual-based DM can represent the energy spectrum at high wavenumbers well and successfully recovers information not present in the LR input data. This suggests its potential usefulness in turbulence modelling tasks that depend on small-scale flow features, such as SGS stress estimation in LES.

The applicability of the residual-based DM to anisotropic turbulence was also investigated by performing HR reconstruction of turbulent flows from LR turbulent boundary layer data. The results showed that residual-based DM captures the mean velocity profile and r.m.s. streamwise velocity with reasonable accuracy. Qualitative comparisons of the reconstructed flow fields, distributions of velocity components

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

and quantitative metrics, including correlation coefficients, MSE and LPIPS loss, demonstrated that, on the whole, the residual-based DM outperforms bicubic interpolation, the CNN and U-Net models in HR flow flow field reconstruction of turbulent boundary layer flow.

Although the developed diffusion model demonstrated strong capabilities in generating HR flow fields from LR data with high perceptual quality, there is still room for improvement in pixel-wise accuracy at high upsampling factors. To address this, integrating the diffusion model with a CNN-based model can help enhance local-level precision and reduce point-wise errors such as MSE. Thus, while the diffusion model effectively captures a wide range of turbulent structures, combining it with an additional network may further reduce pixel-level discrepancies and improve overall accuracy. Additionally, as a future direction, the model can be extended to three-dimensional HR flow field reconstruction, which can be achieved by adopting the three-dimensional version of the U-Net-based architecture, to further capture volumetric correlations and ensure inter-slice coherence. Our code will be made publicly available as open source upon publication. Additional details are provided in Appendix F.

Acknowledgements. This work is partially supported by the supporting program from the Center for Diversity, Equity, and Inclusion (TUMUG) at Tohoku University. Numerical computations have been performed on the supercomputer system 'AFI-NITY' at the Advance Fluid Information Research Center, Institute of Fluid Science, Tohoku University.

Declaration of interests. The authors report no conflicts of interest.

Data availability statement. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix A. The LR data details and turbulence parameters

Table 11 shows the LR grid size relative to the Kolmogorov scale, Taylor microscale and integral length scale for HIT2. For the finest LR case (128×128), the grid spacing satisfies $\Delta x/\lambda \approx 0.15$ but $\Delta x/\eta \approx 4.8$, indicating that, while part of the large-scale and early inertial-range dynamics is resolved, dissipation-range motions are entirely unresolved. As the resolution decreases, $\Delta x/\lambda$ increases to 0.23 (64×64) and 0.31 (32×32), while $\Delta x/\eta$ increases to approximately 8.4 and 13.3, respectively, implying that progressively larger portions of the inertial subrange are also truncated. These scale separations indicate that the LR inputs predominantly retain energy-containing motions and, at best, the upper part of the inertial subrange, while dissipation-range physics is absent in all LR cases. Consequently, the reconstruction task primarily involves recovering the inertial-range dynamics rather than true dissipation-range structures. This interpretation is consistent with the observed deviation of the LR energy spectra from DNS at wavenumbers corresponding to length scales well above the Kolmogorov scale, reinforcing that the effective resolution is set by the LR cutoff rather than by η. Overall, this analysis clarifies that the proposed method reconstructs physically meaningful small-scale features by inferring unresolved inertial-range content from large-scale information, rather than directly resolving dissipation-range physics, which provides important physical insight beyond purely metric-based evaluation.

Tables 12–14 present the absolute relative errors with respect to the GT for both the LR flow field and the HR flow field predicted by the models, enabling a fair comparison among the different approaches. The results show that the residual-based DM yields the lowest

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

LR grid size	Δx	$\Delta x/L$	$\Delta x/\lambda$	$\Delta x/\eta$
128×128	0.04909	0.04353	0.15161	4.83169
64×64	0.09817	0.08430	0.23359	8.39778
32×32	0.19635	0.15996	0.30783	13.32090

Table 11. Grid resolution of LR samples relative to integral length scale L, Taylor microscale λ and Kolmogorov scale η.

Variable	LR	Bicubic	CNN	U-Net	DDPM	DM
λ	0.18029	0.16309	0.03536	0.02085	0.09314	0.01987
ε	0.29556	0.27281	0.06913	0.05097	0.13389	0.04091
u_{rms}	0.01868	0.01630	0.00232	0.01096	0.06760	0.00015

Table 12. Absolute relative errors with respect to GT at upsampling factor = 4.

Variable	LR	Bicubic	CNN	U-Net	DDPM	DM
λ	0.53199	0.47716	0.19688	0.09084	0.06237	0.02147
ε	0.59834	0.56771	0.31415	0.18574	0.02122	0.02319
u_{rms}	0.05741	0.05659	0.01764	0.03122	0.13945	0.06467

Table 13. Absolute relative errors with respect to GT at upsampling factor = 8.

Variable	LR	Bicubic	CNN	U-Net	DDPM	DM
λ	1.32516	0.89673	0.56053	0.20271	0.07779	0.10910
ε	0.84117	0.76198	0.62000	0.35747	0.29403	0.19208
u_{rms}	0.14086	0.14372	0.07484	0.07069	0.39968	0.00623

Table 14. Absolute relative errors with respect to GT at upsampling factor = 16.

errors in the turbulence characteristics, highlighting its strong capability in reconstructing the HR flow field.

Appendix B. Testing model on LR input obtained using average pooling

We tested the model on LR data obtained via average pooling to evaluate its performance when the LR data are generated using a method different from that used for training. As shown in figure 19, although the ResShift model was trained on LR data produced using the Gaussian filter, it performs roughly well on LR data obtained through average pooling. This is because the model primarily learns to reconstruct missing high-frequency details from the preserved low-wavenumber structures. As long as the coarse-graining method retains the main flow features, the model can generalise across different types of downsampling. Only slight over- and under-predictions are observed in the velocity field when the LR data are generated via average pooling, which is expected given that the

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

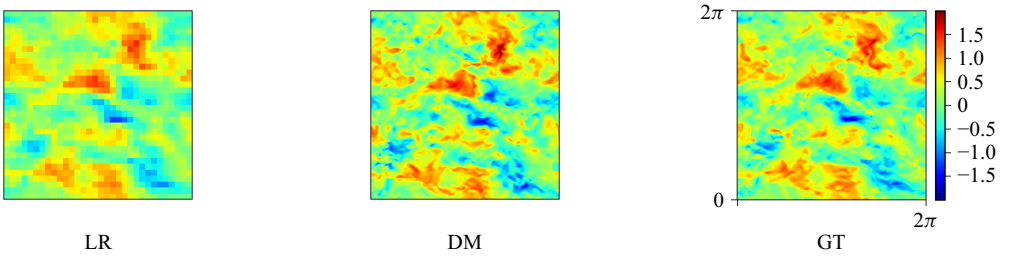


Figure 19. Testing model on LR data obtained using average pooling. Here, DM refers to the residual-based diffusion model.

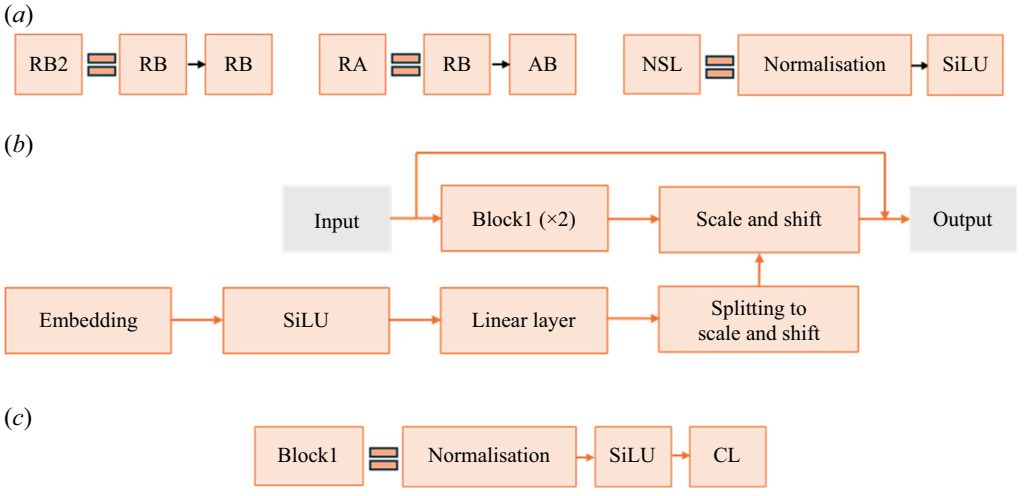


Figure 20. Details of U-Net model. (*b*) Shows the details of residual block (RB) in (*a*), and AB represents the attention block, which is a Swin transformer in the present study.

model was trained on Gaussian-filtered data. Although this difference is minor, including LR data generated using multiple degradation methods during training could further reduce the model's sensitivity to the specific downsampling approach. Another approach could be to fine-tune the model using the corresponding LR data obtained with a different downsampling method.

Appendix C. Details of network architecture

In figure 4, the HR and interpolated LR flow fields are concatenated along the channel dimension and used as input to the U-Net model. Downsampling is performed by convolution with a stride of 2 and 3×3 kernel. During upsampling, spatial resolution is doubled using nearest-neighbour interpolation, followed by a 3×3 convolution to refine the upsampled features. Figure 20 shows the components of the U-Net model presented in figure 4. In order to obtain the embedding in figure 20, the time step is first mapped to a sinusoidal vector, which is then processed by a feedforward network consisting of two linear layers with a sigmoid linear unit (SiLU) activation function between them. In figure 20, AB denotes the Swin transformer block (Liu *et al.* 2021*a*), where each block has a depth of 2, embedding dimension of 192, window size of 8×8 and an MLP ratio

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

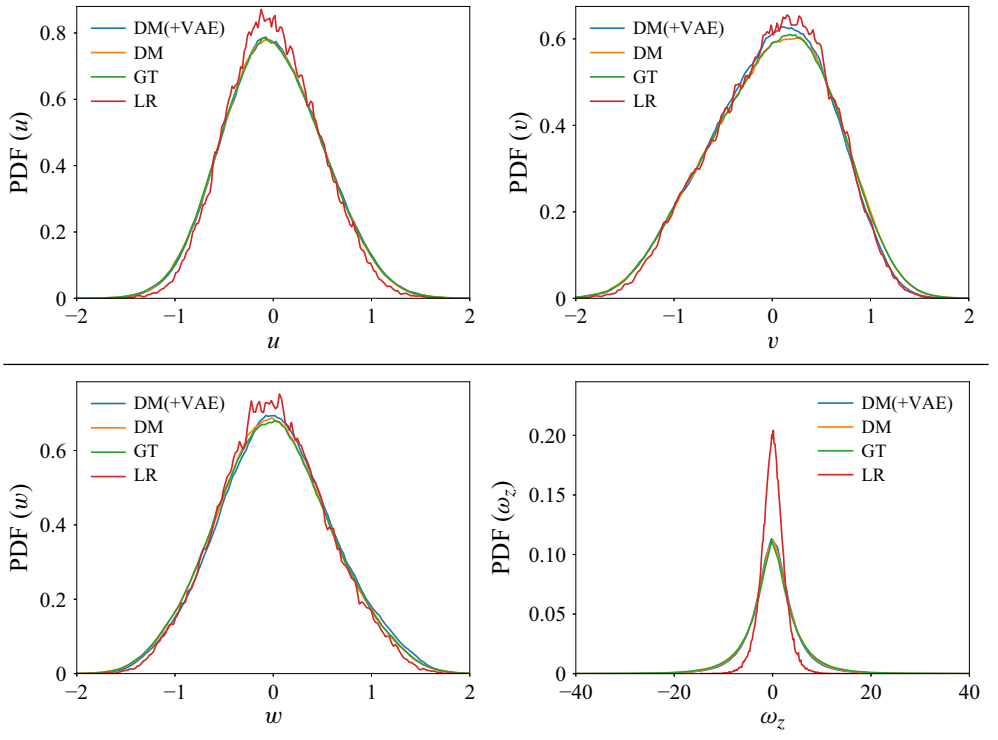


Figure 21. The PDFs of velocity components and vorticity for the upsampling factor of 16. Here, DM (+VAE) denotes the residual-based diffusion model operating in the latent space.

of 4, with 32 channels per attention head. This allows the network to capture long-range dependencies while maintaining the hierarchical structure of the U-Net.

Appendix D. Ablation study

D.1. *Pixel-space diffusion model*

To examine the difference between the presented pixel-wise diffusion model (residual-based DM) and the diffusion model operating in the latent space, we employed the latent representation of VQGAN (Esser, Rombach & Ommer 2021; Yue *et al.* 2023). In this approach, the original flow field is encoded into a compressed latent space using a pretrained VQGAN encoder, with a spatial downsampling factor of four along each dimension. The diffusion process is then performed in this latent domain, and the final output is reconstructed in the physical space using the VQGAN decoder. This experiment is conducted for the upsampling factor of 16. The results for the PDFs of the velocity components and vorticity are shown in figure 21. As seen in this figure, the LDM tends to overpredict the peak values of the velocity and vorticity PDFs. Moreover, the LPIPS loss for the LDM in predicting the vorticity field is 0.5920, which is significantly higher than that of the pixel-wise diffusion model (0.4628). This demonstrates that the LDM produces flow fields with lower perceptual quality and weaker structural similarity to the GT, whereas the pixel-wise diffusion model achieves higher reconstruction fidelity, effectively preserving fine-scale vortical structures and overall spatial coherence of the turbulent flow.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

D.2. *The LPIPS loss*

In the present study, we added perceptual loss to the training objective to enhance the reconstruction of turbulent flow fields. We observed that, in our study, incorporating LPIPS loss improves both perceptual and pixel-wise accuracy: models trained with only MSE tend to produce mildly smooth velocity and vorticity fields. In particular, LPIPS emphasises consistency of multi-scale flow features, including fine-scale vortical structures that are crucial in turbulence. By combining MSE with LPIPS loss, the model not only achieves accurate pixel-level predictions but also preserves the spatial coherence and realism of the reconstructed flow fields, reducing artefacts commonly observed in purely MSE-trained models. This improvement indicates that LPIPS enhances the model's ability to capture perceptually important flow features beyond simple pixel-wise fidelity. These observations suggest that incorporating LPIPS complements the MSE loss, reinforcing both qualitative realism and overall accuracy of the super-resolved flow fields.

D.3. *Swin transformer*

In this study, within the U-Net architecture used in the residual-based DM, the conventional self-attention block is replaced with a Swin transformer block. Unlike standard self-attention, which exhibits quadratic computational complexity with respect to spatial resolution, the Swin transformer maintains linear complexity, thereby reducing both memory usage and inference time. The average inference times are 2.4378 and 4.2427 seconds for the Swin transformer and self-attention versions, respectively, demonstrating that the Swin transformer is significantly faster and can substantially reduce computational time during inference. The computations were performed on a single NVIDIA RTX A6000 GPU.

Appendix E. Applicability of the model to experimental data

To assess the applicability of the residual-based DM trained on DNS data to experimental measurements, we test the model using planar particle image velocimetry (PIV) data of turbulent boundary layer flow provided by Towne *et al.* (2023). The friction Reynolds number for the tested PIV data is $Re_\tau = 987$. As the spanwise velocity component is not available in the experimental dataset, following § 3.5, the model is trained using paired filtered DNS and DNS data for the streamwise velocity component only. The trained residual-based DM is then tested on the PIV data.

The results of testing the model on the PIV data are shown in figure 22, where HR denotes the HR data obtained from the experimental measurements. As shown in the figure, the PIV measurements cannot fully resolve the near-wall region, particularly for $y^+ < 15$ (Fukami & Taira 2025). Nevertheless, when LR data obtained from downsampled PIV measurements are fed into the residual-based DM, it is roughly able to capture the near-wall region, while providing roughly accurate predictions in regions farther from the wall. These results demonstrate the potential of the trained residual-based DM for application to experimental data.

For improved accuracy, the model can be further fine tuned using PIV data. Several strategies can be employed to achieve this. For example, the model can be embedded into a self-supervised learning framework, which does not require paired data (Chen *et al.* 2024). Alternatively, the model can be combined with an unsupervised approach such as cycleGAN (Zhu *et al.* 2017), where the cycleGAN performs unsupervised learning on unpaired PIV and filtered DNS data, and the residual-based DM subsequently reconstructs DNS-quality flow fields from the filtered DNS data. This strategy can also be applied to

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

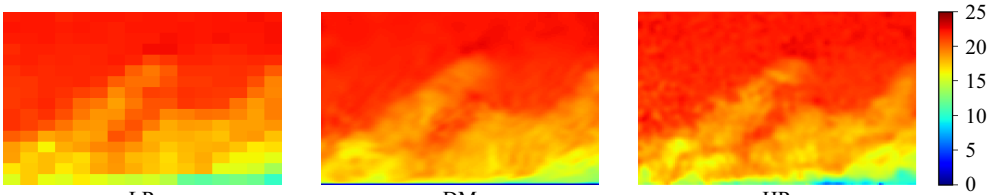


Figure 22. Evaluation of the residual-based DM trained on DNS data using experimental (PIV) data.

LES datasets. Such approaches provide a foundation for developing data-driven turbulence models, including SGS models in LES, in the absence of GT DNS data (Maejima & Kawai 2025).

Appendix F. Code availability

We created a public repository, https://github.com/golsatabe/SR-Turb-DM, where the code will be released once the paper is published.

References

Bae, H.J. & Koumoutsakos, P. 2022 Scientific multi-agent reinforcement learning for wall-models of turbulent flows. *Nat. Commun.* **13** (1), 1443.

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

Chen, H., Li, W., Gu, J., Ren, J., Sun, H., Zou, X., Zhang, Z., Yan, Y. & Zhu, L. 2024 Low-res leads the way: improving generalization for super-resolution by self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25857–25867.

Chen, H. Wang, Y. Guo, T. Xu, C. Deng, Y. Liu, Z. Ma, S. Xu. C. XU, C & Gao, W. 2021*a* Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12299–12310.

Chen, L., Han, B., Wang, X., Zhao, J., Yang, W. & Yang, Z. 2023 Machine learning methods in weather and climate applications: a survey. *Appl. Sci.* **13** (21), 12019.

Chen, S., Sammak, S., Givi, P., Yurko, J.P. & Jia, X. 2021*b* Reconstructing high-resolution turbulent flows using physics-guided neural networks. In *2021 IEEE International Conference on Big Data*, pp. 1369–1379. IEEE.

Chen, X., Mishra, N., Rohaninejad, M. & Abbeel, P. 2018 Pixelsnail: an improved autoregressive generative model. In *International Conference on Machine Learning*, pp. 864–872. PMLR.

Chira, D., Haralampiev, I., Winther, O., Dittadi, A. & Liévin, V. 2022 Image super-resolution with deep variational autoencoders. In *European Conference on Computer Vision*, pp. 395–411. Springer.

Dhariwal, P. & Nichol, A. 2021 Diffusion models beat GANs on image synthesis. *Adv. Neural Inform. Proc. Syst.* **34**, 8780–8794.

Dong, C., Loy, C.C., He, K. & Tang, X. 2015 Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38** (2), 295–307.

Du, P., Parikh, M.H., Fan, X., Liu, X.-Y. & Wang, J.-X. 2024 Conditional neural field latent diffusion model for generating spatiotemporal turbulence. *Nat. Commun.* **15** (1), 10416.

Duraisamy, K., Iaccarino, G. & Xiao, H. 2019 Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51** (1), 357–377.

Esser, P., Rombach, R. & Ommer, B. 2021 Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12873–12883.

Fan, H., Cheng, S., de, N., Audrey, J. & Arcucci, R. 2025*a* Vitae-sl: a vision transformer-based autoencoder and spatial interpolation learner for field reconstruction. *Comput. Phys. Commun.* **308**, 109464.

Fan, J., Duan, W., Huang, L., Zhang, L. & Yang, K. 2023 High-fidelity flow field reconstruction model for incompressible fluid with physical constraints. *Ocean Engng* **280**, 114597.

Fan, X., Akhare, D. & Wang, J.-X. 2025*b* Neural differentiable modeling with diffusion-based super-resolution for two-dimensional spatiotemporal turbulence. *Comput. Meth. Appl. Mech. Engng* **433**, 117478.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

FRANK, M., DRIKAKIS, D. & CHARISSIS, V. 2020 Machine-learning methods for computational science and engineering. *Computation* **8** (1), 15.
FUKAMI, K., FUKAGATA, K. & TAIRA, K. 2019 Super-resolution reconstruction of turbulent flows with machine learning. *J. Fluid Mech.* **870**, 106–120.
FUKAMI, K., FUKAGATA, K. & TAIRA, K. 2021 Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows. *J. Fluid Mech.* **909**, A9.
FUKAMI, K. & TAIRA, K. 2024 Single-snapshot machine learning for super-resolution of turbulence. *J. Fluid Mech.* **1001**, A32.
FUKAMI, K. & TAIRA, K. 2025 Observable-augmented manifold learning for multi-source turbulent flow data. *J. Fluid Mech.* **1010**, R4.
GAMAHARA, M. & HATTORI, Y. 2017 Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids* **2** (5), 054604.
GAO, D., GONG, Y., CAO, J., WANG, B., ZHANG, H., DONG, J. & QIU, J. 2025 ResShift-4E: improved diffusion model for super-resolution with microscopy images. *Electronics* **14** (3), 479.
GAO, S., LIU, X., ZENG, B., XU, S., LI, Y., LUO, X., LIU, J., ZHEN, X. & ZHANG, B. 2023Implicit diffusion models for continuous super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10021–10030.
GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. & BENGIO, Y. 2014 Generative adversarial nets. In *Advances in Neural Information Processing Systems* 27.
GUAN, Y., CHATTOPADHYAY, A., SUBEL, A. & HASSANZADEH, P. 2022 Stable a posteriori LES of 2D turbulence using convolutional neural networks: backscattering analysis and generalization to higher Re via transfer learning. *J. Comput. Phys.* **458**, 111090.
GUASTONI, L., RABAULT, J., SCHLATTER, P., AZIZPOUR, H. & VINUESA, R. 2023 Deep reinforcement learning for turbulent drag reduction in channel flows. *Eur. Phys. J. E* **46** (4), 27.
GUO, B., ZHANG, X., WU, H., WANG, Y., ZHANG, Y. & WANG, Y.-F. 2022 Lar-sr: A local autoregressive model for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1909–1918.
HE, K., ZHANG, X., REN, S. & SUN, J. 2016 Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
HO, J., JAIN, A. & ABBEEL, P. 2020 Denoising diffusion probabilistic models. *Adv. Neural Inform. Proc. syst.* **33**, 6840–6851.
ISHIHARA, T., KANEDA, Y., YOKOKAWA, M., ITAKURA, K. & UNO, A.2007 Small-scale statistics in high-resolution direct numerical simulation of turbulence: Reynolds number dependence of one-point velocity gradient statistics. *J. Fluid Mech.* **592**, 335–366.
ISHIZE, T., OMICHI, H. & FUKAGATA, K. 2024 Flow control by a hybrid use of machine learning and control theory. *Intl J. Numer. Meth. Heat Fluid Flow* **34** (8), 3253–3277.
ISOLA, P., ZHU, J.-Y., ZHOU, T. & EFROS, A.A. 2017 Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134.
JIANG, C., VINUESA, R., CHEN, R., MI, J., LAIMA, S. & LI, H. 2021 An interpretable framework of data-driven turbulence modeling using deep neural networks. *Phys. Fluids* **33** (5), 055133.
JOHNSON, J., ALAHI, A. & FEI-FEI, L. 2016 Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pp. 694–711. Springer.
KIM, H., KIM, J., WON, S. & LEE, C. 2021 Unsupervised deep learning for super-resolution reconstruction of turbulence. *J. Fluid Mech.* **910**, A29.
KINGMA, D.P. & BA, J. 2014 Adam: a method for stochastic optimization. arXiv:1412.6980.
KINGMA, D.P. & WELLING, M. 2013 Auto-encoding variational Bayes. arXiv:1312.6114.
LEDIG, C., *et al.* Wenzhe 2017 Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690.
LEE, D., KIM, C., KIM, S., CHO, M. & HAN, W.-S. 2022 Autoregressive image generation using residual quantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11523–11532.
LI, H., YANG, Y., CHANG, M., CHEN, S., FENG, H., XU, Z., LI, Q. & CHEN, Y. 2022 SRDiff: single image super-resolution with diffusion probabilistic models. *Neurocomputing* **479**, 47–59.
LI, Z., SHU, D. & BARATI FARIMANI, A. 2023 Scalable transformer for PDE surrogate modeling. *Adv. Neural Inform. Proc. Syst.* **36**, 28010–28039.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

LIANG, J., CAO, J., SUN, G., ZHANG, K., VAN, G., Luc, & TIMOFTE, R. 2021 Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1833–1844.

LIN, X., LI, Y., HSIAO, J., HO, C. & KONG, Y. 2023 Catch missing details: Image reconstruction with frequency augmented variational autoencoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1736–1745.

LING, J., KURZAWSKI, A. & TEMPLETON, J. 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.* **807**, 155–166.

LIU, B., TANG, J., HUANG, H. & LU, X.-Y. 2020 Deep learning methods for super-resolution reconstruction of turbulent flows. *Phys. Fluids* **32** (2), 025105.

LIU, B., YU, H., HUANG, H., LIU, N. & LU, X. 2022 Investigation of nonlocal data-driven methods for subgrid-scale stress modeling in large eddy simulation. *AIP Adv.* **12** (6), 065129.

LIU, X., ZHANG, Y., GUO, T., LI, X., SONG, D. & YANG, H. 2025 A multi-scale hybrid attention Swin-transformer-based model for the super-resolution reconstruction of turbulence. *Nonlinear Dyn.* **113**, 15815–15844.

LIU, Z., LIN, Y., CAO, Y., HU, H., WEI, Y., ZHANG, Z., LIN, S. & GUO, B. 2021*a* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022.

LIU, Z.-S., SIU, W.-C. & WANG, L.-W. 2021*b* Variational autoencoder for reference based image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 516–525.

LOZANO-DURÁN, A. & BAE, H.J. 2023 Machine learning building-block-flow wall model for large-eddy simulation. *J. Fluid Mech.* **963**, A35.

MAEJIMA, S. & KAWAI, S. 2025 Unsupervised and supervised machine learning pipeline for super-resolution-based subgrid scale modelling in coarse-grid large-eddy simulations. *J. Fluid Mech.* **1013**, A28.

MAEJIMA, S., TANINO, K. & KAWAI, S. 2024 Physics-informed machine-learning solution to log-layer mismatch in wall-modeled large-eddy simulation. *Phys. Rev. Fluids* **9** (8), 084609.

MATSUO, M., FUKAMI, K., NAKAMURA, T., MORIMOTO, M. & FUKAGATA, K. 2024 Reconstructing three-dimensional bluff body wake from sectional flow fields with convolutional neural networks. *SN Comput. Sci.* **5** (3), 306.

MCCABE, M., *et al.* 2024 Multiple physics pretraining for spatiotemporal surrogate models. *Adv. Neural Inform. Proc. Syst.* **37**, 119301–119335.

MIYAZAKI, S. & HATTORI, Y. 2020 Improving accuracy of turbulence models by neural network. arXiv:2012.01723.

NICHOL, A.Q. & DHARIWAL, P. 2021 Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 8162–8171. PMLR.

PARK, J. & CHOI, H. 2021 Toward neural-network-based large eddy simulation: application to turbulent channel flow. *J. Fluid Mech.* **914**, A16.

QI, J. & MA, H. 2024 A combined model of diffusion model and enhanced residual network for super-resolution reconstruction of turbulent flows. *Mathematics* **12** (7), 1028.

RAHMAN, M.A., *et al.* 2024 Pretraining codomain attention neural operators for solving multiphysics PDEs. *Adv. Neural Inform. Proc. Syst.* **37**, 104035–104064.

ROMBACH, R., BLATTMANN, A., LORENZ, D., ESSER, P. & OMMER, B. 2022 High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695.

RONNEBERGER, O., FISCHER, P. & BROX, T. 2015 U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241. Springer.

SAHARIA, C., HO, J., CHAN, W., SALIMANS, T., FLEET, D.J. & NOROUZI, M. 2022 Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal.* **45** (4), 4713–4726.

SHU, D., LI, Z. & FARIMANI, A.B. 2023 A physics-informed diffusion model for high-fidelity flow field reconstruction. *J. Comput. Phys.* **478**, 111972.

SIMONYAN, K. & ZISSERMAN, A. 2014 Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

SONG, J., MENG, C. & ERMON, S. 2020 Denoising diffusion implicit models. arXiv:2010.02502.

SUN, L., PAN, J. & TANG, J. 2022*a* Shufflemixer: an efficient convnet for image super-resolution. *Adv. Neural Inform. Proc. Syst.* **35**, 17314–17326.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

SUN, X., CAO, W., LIU, Y., ZHU, L. & ZHANG, W. 2022*b* High Reynolds number airfoil turbulence modeling method based on machine learning technique. *Comput. Fluids* **236**, 105298.
TABE JAMAAT, G. & HATTORI, Y. 2022 Development of subgrid-scale model for LES of Burgers turbulence with large filter size. *Phys. Fluids* **34** (4).
TABE JAMAAT, G. & HATTORI, Y. 2023 A priori assessment of nonlocal data-driven wall modeling in large eddy simulation. *Phys. Fluids* **35** (5).
TABE JAMAAT, G., HATTORI, Y. & KAWAI, S. 2024 A posteriori study on wall modeling in large eddy simulation using a nonlocal data-driven approach. *Phys. Fluids* **36** (6), 065164.
TANG, L., LI, C., ZHAO, Z., XIAO, Y. & CHEN, S. 2024 Super-resolution reconstruction of wind fields with a swin-transformer-based deep learning framework. *Phys. Fluids* **36** (12), 125110.
TOWNE, A., *et al.* 2023 A database for reduced-complexity modeling of fluid flows. *AIAA J.* **61** (7), 2867–2892.
VAN DEN OORD, A. *et al.* 2017 Neural discrete representation learning. *Adv. Neural Inform. Proc. Syst.* **30**.
VIGNON, C., RABAULT, J. & VINUESA, R. 2023 Recent advances in applying deep reinforcement learning for flow control: perspectives and future directions. *Phys. Fluids* **35** (3), 031301.
VINUESA, R. & BRUNTON, S.L. 2022 Enhancing computational fluid dynamics with machine learning. *Nature Comput. Sci.* **2** (6), 358–366.
WANG, X., YU, K., WU, S., GU, J., LIU, Y., DONG, C., QIAO, Y. & CHANGE LOY, C. 2018*a* Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.
WANG, Z., LUO, K., LI, D., TAN, J. & FAN, J. 2018*b* Investigations of data-driven closure for subgrid-scale stress in large-eddy simulation. *Phys. Fluids* **30** (12), 125101.
WARD, N.J. 2023 Physics-informed super-resolution of turbulent channel flows via three-dimensional generative adversarial networks. *Fluids* **8** (7), 195.
XIE, Y., FRANZ, E., CHU, M. & THUEREY, N. 2018 TempoGAN: a temporally coherent, volumetric GAN for super-resolution fluid flow. *ACM Trans. Graph. (TOG)* **37** (4), 1–15.
XU, Q., ZHUANG, Z., PAN, Y. & WEN, B. 2023 Super-resolution reconstruction of turbulent flows with a transformer-based deep learning framework. *Phys. Fluids* **35** (5), 055130.
YANG, X., ZAFAR, S., WANG, J.-X. & XIAO, H. 2019 Predictive large-eddy-simulation wall modeling via physics-informed neural networks. *Phys. Rev. Fluids* **4** (3), 034602.
YOUSIF, M.Z., YU, L., HOYAS, S., VINUESA, R. & LIM, H.C. 2023 A deep-learning approach for reconstructing 3D turbulent flows from 2D observation data. *Sci. Rep.* **13** (1), 2529.
YOUSIF, M.Z., YU, L. & LIM, H.-C. 2021 High-fidelity reconstruction of turbulent flow from spatially limited data using enhanced super-resolution generative adversarial network. *Phys. Fluids* **33** (12), 125119.
YOUSIF, M.Z., YU, L. & LIM, H.-C. 2022 Super-resolution reconstruction of turbulent flow fields at various Reynolds numbers based on generative adversarial networks. *Phys. Fluids* **34** (1), 015130.
YU, L., YOUSIF, M.Z., ZHANG, M., HOYAS, S., VINUESA, R. & LIM, H.-C. 2022 Three-dimensional ESRGAN for super-resolution reconstruction of turbulent flows with tricubic interpolation-based transfer learning. *Phys. Fluids* **34** (12), 125126.
YUAN, Y., LIU, S., ZHANG, J., ZHANG, Y., DONG, C. & LIN, L. 2018 Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 701–710.
YUE, Z., WANG, J. & LOY, C.C. 2023 Resshift: efficient diffusion model for image super-resolution by residual shifting. *Adv. Neural Inform. Proc. Syst.* **36**, 13294–13307.
ZANNA, L. & BOLTON, T. 2020 Data-driven equation discovery of ocean mesoscale closures. *Geophys. Res. Lett.* **47** (17), e2020GL088376.
ZENG, K., ZHANG, Y., XU, H. & FENG, X. 2024 Super-resolution reconstruction of turbulent flows with a hybrid framework of attention. *Phys. Fluids* **36** (6), 065107.
ZHANG, J., LIU, J. & HUANG, Z. 2023*a* Improved deep learning method for accurate flow field reconstruction from sparse data. *Ocean. Engng* **280**, 114902.
ZHANG, R., ISOLA, P., EFROS, A.A., SHECHTMAN, E. & WANG, O. 2018 The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595.
ZHANG, Z., FLORA, K., KANG, S., LIMAYE, A.B. & KHOSRONEJAD, A. 2022 Data-driven prediction of turbulent flow statistics past bridge piers in large-scale rivers using convolutional neural networks. *Water Resour. Res.* **58** (1), e2021WR030163.
ZHANG, Z., HAO, X., SANTONI, C., SHEN, L., SOTIROPOULOS, F. & KHOSRONEJAD, A. 2023*b* Toward prediction of turbulent atmospheric flows over propagating oceanic waves via machine-learning augmented large-eddy simulation. *Ocean. Engng* **280**, 114759.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press

ZHANG, Z., SANTONI, C., HERGES, T., SOTIROPOULOS, F. & KHOSRONEJAD, A. 2021 Time-averaged wind turbine wake flow field prediction using autoencoder convolutional neural networks. *Energies* **15** (1), 41.

ZHANG, Z., SOTIROPOULOS, F. & KHOSRONEJAD, A. 2024*a* A deep-learning approach for 3d realization of mean wake flow of marine hydrokinetic turbine arrays. *Energy Rep.* **12**, 2621–2630.

ZHANG, Z., SOTIROPOULOS, F. & KHOSRONEJAD, A. 2024*b* Predicting turbulent wake flow of marine hydrokinetic turbine arrays in large-scale waterways via physics-enhanced convolutional neural networks. *Phys. Fluids* **36** (4), 045156.

ZHAO, Q., JIN, G. & ZHOU, Z. 2022 Deep learning method for the super-resolution reconstruction of small-scale motions in large-eddy simulation. *AIP Adv.* **12** (12), 125304.

ZHOU, Z., HE, G., WANG, S. & JIN, G. 2019 Subgrid-scale model for large-eddy simulation of isotropic turbulent flows using an artificial neural network. *Comput. Fluids* **195**, 104319.

ZHOU, Z., LI, B., YANG, X. & YANG, Z. 2022 A robust super-resolution reconstruction model of turbulent flow data based on deep learning. *Comput. Fluids* **239**, 105382.

ZHU, J.-Y., PARK, T., ISOLA, P. & EFROS, A.A. 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232.

https://doi.org/10.1017/jfm.2026.11511 Published online by Cambridge University Press