
A 3D super-resolution of wind fields via physics-informed pixel-wise self-attention generative adversarial network

Takuya Kurihana

Department of Computer Science
University of Chicago
tkurihana@uchicago.edu

Kyongmin Yeo

IBM Research
kyeo@us.ibm.com

Daniela Szwarcman

IBM Research
daniela.szw@ibm.com

Bruce Elmegreen

IBM Research
bge@us.ibm.com

Karthik Mukkavilli

IBM Research
karthik.mukkavilli@ibm.com

Johannes Schmude

IBM Research
johannes.schmude@ibm.com

Levente Klein

IBM Research
kleinl@us.ibm.com

Abstract

To mitigate global warming, greenhouse gas sources need to be resolved at a high spatial resolution and monitored in time to ensure the reduction and ultimately elimination of the pollution source. However, the complexity of computation in resolving high-resolution wind fields left the simulations impractical to test different time lengths and model configurations. This study presents a preliminary development of a physics-informed super-resolution (SR) generative adversarial network (GAN) that super-resolves the three-dimensional (3D) low-resolution wind fields by upscaling $\times 9$ times. We develop a pixel-wise self-attention (PWA) module that learns 3D weather dynamics via a self-attention computation followed by a 2D convolution. We also employ a loss term that regularizes the self-attention map during pretraining, capturing the vertical convection process from input wind data. The new PWA SR-GAN shows the high-fidelity super-resolved 3D wind data, learns a wind structure at the high-frequency domain, and reduces the computational cost of a high-resolution wind simulation by $\times 89.7$ times.

1 Introduction

Accurate tracing and monitoring of greenhouse gas (GHG) sources is a key measurement to take action for tackling the mitigation of global warming issues [3]. The higher spatial resolution of wind simulation enables precise tracking of GHG emissions from potential sources, and helps decision-making in various fields such as policymakers, agriculture, and renewable energy. However, high-resolution simulation is not always available due to the demands of computational resources.

Artificial Intelligence (AI), including deep neural networks (DNNs), can reduce the computational cost by upscaling the wind fields from low-resolution (LR) to high-resolution (HR) data. Super-resolution (SR) is one of the solutions to achieve the goal: conventional SR techniques rely on convolutional neural networks (CNNs) to reproduce realistic HR wind fields [12, 14, 7, 9, 15, 1]. However, these CNN-based approaches either perform on 2D data or fall short of capturing the 3D dynamics in weather systems because a convolutional kernel truncates the association between

vertical layers, limiting to learning of convection and diurnal cycles induced by incoming solar radiations. Another issue is that widely used weather simulations employ a non-uniform grid in a vertical axis for computational stability. The difference in height between a pair of two adjacent vertical layers becomes larger at higher altitudes. Thus, CNN is not sufficient to capture physics at the same vertical scaling.

To address these issues, we develop a prototype of a novel architecture of physics-informed neural networks that learn multi-scale spatial dynamics. Our goal is to develop a physics-informed neural network that super-resolves 900 m resolution of three-dimensional wind data into 100 m scale high-resolution wind data and captures three-dimensional dynamics in weather systems.

2 3D super-resolution of wind fields

We develop a *pixel-wise self-attention network* (PWA) to learn the three-dimensional dynamics of weather simulations. Using the neural network as a generator, we train a generative network for super-resolving wind velocity fields.

2.1 WRF model dataset

The Weather Research & Forecasting Model (WRF) [11] has been widely used for simulating weather systems from mesoscale to turbulence scales. Large-eddy simulation (LES) is often nested into the WRF framework to allow simulating a convective process to capture a more realistic turbulence structure [2]. Outputs from WRF-LES used for this study are nested in 900 m, 300 m, and 100 m horizontal resolution. The outer 900 m and inner 100 m simulations differ in their model physics and our ultimate goal is to super-resolve 900 m LR data to 100 m HR data. This study, however, focuses on synthesizing the 900 m LR data by spatially averaging 100 m HR data as a preliminary step. Our training and testing data are produced by a WRF nested LES at 100 m horizontal scale and 59 vertical layers, generating outputs every five minutes from 00 UTC to 23:59 UTC. The initialization time window is set for two hours starting at 22:00 UTC a day before. We sample training and testing data randomly from one-month simulation outputs from September 2019. We select three wind velocity fields (U, V, W), all at 100 m spatial resolution over $40 \text{ km} \times 40 \text{ km}$ simulation domain, and use only 8 layers from the surface level. We then extract 22 500 training and 2500 testing data, as the smaller geographical images, each of which is $126 \text{ pixels} \times 126 \text{ pixels}$ ($\sim 12.6 \text{ km} \times 12.6 \text{ km}$) for high-resolution (HR) data. To synthesize low-resolution (LR) data, we average every 9×9 pixels, giving 900 m LR data.

2.2 Physics-informed pixel-wise self-attention

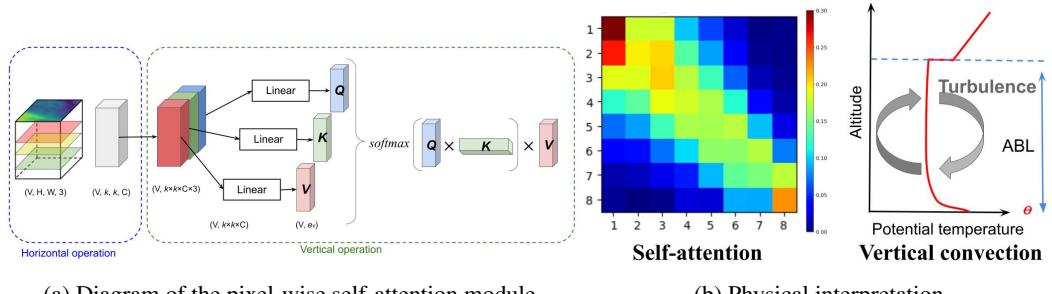


Figure 1: Illustration of the pixel-wise self-attention module: (a) workflow of PWA module. (b) an example 8×8 self-attention map (**left**) shows strong (bright colors) and weak (dark colors) signals between adjacent layers, associating convection in weather systems (**right**).

The fluctuation in the atmospheric boundary layer (ABL) height during both day and night significantly impacts the development of vertical convection intensity, which varies across different altitudes. A standard convolutional filter may truncate these signals and be suboptimal to the non-uniform vertical grid system in numerical models for capturing information at the same scale. We introduce a self-attention computation [13] at each grid column to better embed nonlinear association between

vertical layers into networks: self-attention inherently computes attention scores for each element in sequences, enabling the representation of signal associations between a given vertical layer and others. This versatility extends beyond the constraints of a convolutional kernel that limits associations solely to neighboring layers above and below.

Figure 1 shows the concept of the newly developing pixel-wise self-attention (PWA) module at a high level. Suppose we have an input image X , a 5D Tensor (N, C, V, H, W) where N , C , H , and W depict the size of the mini-batch, the number of variables (i.e., U , V , and W winds), the number of vertical layers, height, and width. PWA module is composed of two parts: *horizontal* and *vertical* operation. For the horizontal operation, we apply the 2D convolutional operation to each image of (V, H, W) by $1 \times 3 \times 3$ convolution kernel. For the vertical operation, we first convert the image shape from (V, H, W, C) to $(V, H \times W \times 3 \cdot C)$ with a linear transformation to create three matrices such that query $Q \in \mathbb{R}^{V \times e_v}$, key $K \in \mathbb{R}^{V \times e_v}$, and value $V \in \mathbb{R}^{V \times e_v}$. We use these three features mapped from physical space to model space to compute a self-attention map M by a simple scaled dot product in Equation 1 as follows;

$$\text{Attn} = \underbrace{\text{softmax} \left(\frac{Q \cdot K^T}{\sqrt{e_v}} \right) \cdot V}_{M}, \quad (1)$$

where e_v is a dimension of intermediate representation and we depict 64 for the dimension. The learned self-attention map M shown in Figure 1b shows strong signals in diagonal and near-diagonal elements, indicating that adjacent vertical layers contribute to letting networks learn vertical dynamics. In particular, near-surface layers (i.e., left top corner of M) exhibit strong attention in wider elements, and this is associated with daytime convection within the ABL.

Effective learning is achieved through the sparsity of weak signals as well as emphasizing key elements in the self-attention map M depicted in Figure 1b. To do so, we add the regularization of M as one of the loss terms as follows:

$$R_i = \sum_{j=1}^V M_{ij}^2, \quad (2)$$

$$R(M) = \sum_{i=1}^V \frac{1}{R_i}, \quad (3)$$

where $M_{ij} = \text{softmax}(QK^T / \sqrt{e_v})$ and R depicts a regularization term. The minimization of $R(M)$ in Equation 2 indicates the elements of strong attention get larger scores as well as weaker attention reduces the scores. One limitation of this approach is that the range is bounded in $[0, 1]$ at each row. To have more attention to strong signals, and in opposite, suppress the weak ones, we introduce a re-scaling scheme to the self-attention map. Simply, we have a trainable matrix Δ , adjusting the value of self-attention map M by $M^{\text{rescale}} = \Delta \cdot M$. Note that we apply the self-attention map regularization scheme to pre-rescaled M .

2.3 PWA SR-GAN: Pixel-wise self-attention super-resolution generative adversarial network

Figure 2 illustrates the architecture of our super-resolution network inspired by SR-GAN [8]. Due to the nature of convolutional filters that pass low-frequency modes, the results of neural networks have smoothed structures. To address this, we additionally include Sobel filter [5] by computing image gradients in horizontal and vertical directions respectively due to the different spatial scales. Overall, we combine multiple loss terms as follows;

$$\mathcal{L} = \mathcal{L}_{\text{content}} + \lambda_R |R(M)| + \lambda_G \underbrace{\sum_{r \in \{u, v\}} |dx/dr - d\hat{x}/dr| + \lambda_{G_v} |dx/dz - d\hat{x}/dz|}_{\text{Image gradient}}, \quad (4)$$

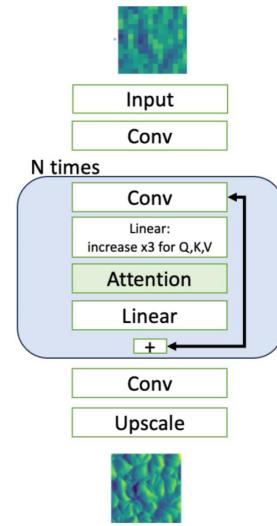


Figure 2: Diagram of the architecture of PWA-SR network. We nest the PWA module (highlighted by light blue color) by three times based on our hyperparameter search.

where $\mathcal{L}_{\text{content}}$ represents mean squared error $\sum_{x_{LR} \in X_{LR}, x_{HR} \in X_{HR}} (G(x_{LR}) - x_{HR})^2$; $G(\cdot)$ depicts the outputs of our neural network. We have three different weight coefficients: λ_R , λ_G , and λ_{G_v} balances self-attention regularization, horizontal gradient, and vertical gradient terms. We use $(\lambda_R, \lambda_G, \lambda_{G_v}) = (0.01, 50, 100)$ in this work. The training uses Adam optimizer [6] with the learning rate at $1e - 3$ on an A-100 GPU for 200 epochs.

CNN-based neural networks are known for filtering low-frequency data in input images and removing high-frequency information [10]. Generative Adversarial Networks (GANs) [4] are our solution to incorporate them into super-resolved images because our training input data x_{LR} has already smoothed out high-frequency information in high-resolution image x_{HR} through averaging the data over 9×9 pixels. Thus, it is challenging to learn the high-frequency data only from low-resolution training inputs. The generator loss in a GAN is typically defined as the negative log-likelihood of the discriminator's output when the generator tries to generate realistic data. The loss term $\mathcal{L}_{\text{adversarial}}$ is often represented as:

$$\mathcal{L}_{\text{adversarial}} = -\mathbb{E}_{\mathbf{x}' \sim p(\mathbf{x}_{LR})} [\log(D(G(\mathbf{x}')))], \quad (5)$$

where \mathbf{x}' is a set of low-resolution wind data sampled from a distribution of low-resolution images; $G(\mathbf{x}')$ is a super-resolved data by the generator; $D(G(\mathbf{x}))$ is a discriminator's output for high-resolution data \mathbf{x} ; and $p(\mathbf{x}')$ is a prior distribution of \mathbf{x}' .

After pre-training PWA-SR network with Equation 4, we modify the loss function for training a generator by adding an adversarial term into the combined loss function;

$$\mathcal{L} = \mathcal{L}_{\text{content}} + \lambda_G \mathcal{L}_{\text{horizontal}} + \lambda_{G_v} \mathcal{L}_{\text{vertical}} + \lambda_a \mathcal{L}_{\text{adversarial}}. \quad (6)$$

For discriminator, we follow the same architecture based on Ledig et al. [8]. See Ledig et al. [8] for further details on the training of GAN.

3 Super-resolution of 3D wind data

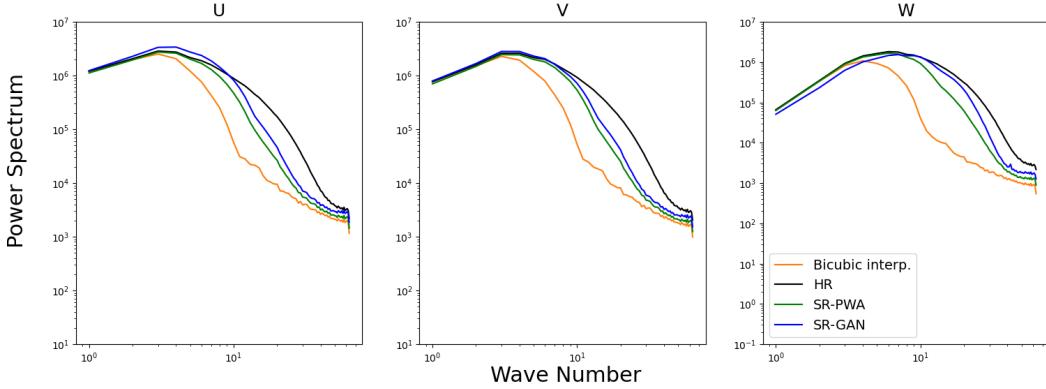


Figure 3: Plots of power spectrum analysis for bicubic interpolation, pixel-wise self-attention SR network (SR-PWA), and pixel-wise self-attention SR-GAN (SR-GAN) as a comparison to HR images. SR-GAN outperforms the SR-PWA (green line) for learning high-frequency domains, especially in vertical wind W

We first investigate whether PWA SR-GAN (hereafter SR-GAN) can generate wind field data that closely approximates the ground-truth HR image in frequency space. We apply the fast Fourier transform to our 2500 test samples for calculating the power spectrum of images instead of mean squared errors (MSEs) because MSEs do not always evaluate if SR models restore high-frequency mode. That is, the test measures how well the model learns the high-frequency mode of wind data through GAN training. Figure 3 shows the results of the power spectrum analysis as a function of wave numbers for U, V, and W wind data respectively. The results compare the power spectrum among different approaches: HR data, bicubic interpolation, SR-PWA, and SR-GAN. The fidelity of capturing realistic wind structures improves as their results of the power spectrum align more closely with HR data (black line). Note that a conventional residual CNN yields a result similar to that of SR-PWA (not reported in Figure 3). We see that the SR-GAN approach (blue line) significantly

outperforms the SR-PWA (green line) for learning high-frequency domains, especially in vertical wind W . However, there is a performance trade-off in low to medium-wave numbers when the learning high-frequency domain achieves well.

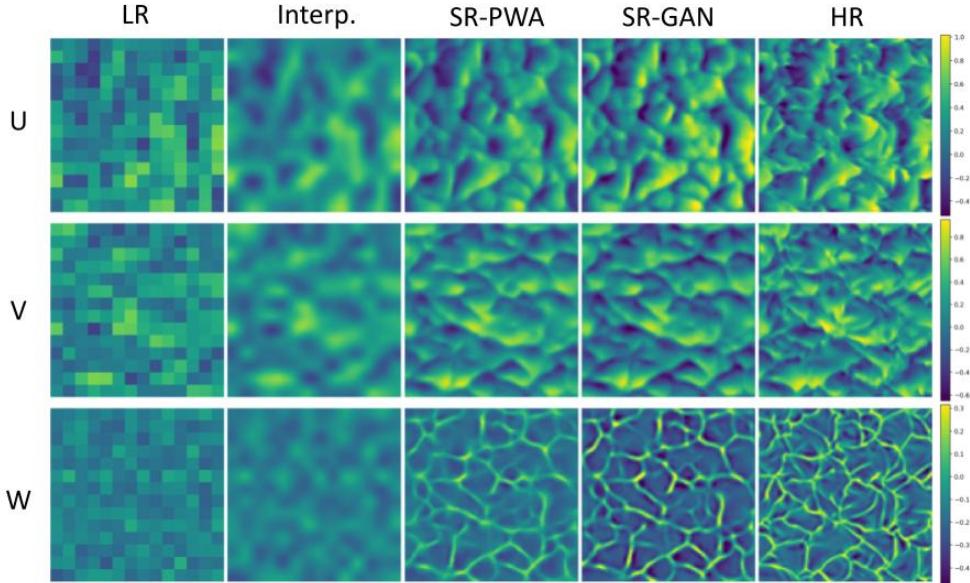


Figure 4: Comparison of qualitative test results based on an example snapshot at the first layer. Each row shows super-resolved or raw images from U , V , or W wind components.

As part of a qualitative evaluation, we present a snapshot example of wind data from the first layer in Figure 4, illustrating the fidelity of super-resolved wind images as a comparison between LR data and HR data. SR-PWA restores major wind structures and the velocity intensities. SR-GAN excels in generating finer wind structures, as evidenced by the results depicted in Figure 3. Additionally, we observe that our SR-GAN decreases computational costs (\sim CPU time \times number of cores) by $\times 89.7$ times compared with the original WRF-LES. Overall, the results indicate that PWA SR-GAN generates realistic 3D high-dimensional wind fields with by lower computing power and resources, enabling simulating advection and diffusion of GHG gas solely using low-resolution WRF simulation outputs.

4 Summary

In summary, we show our preliminary investigation of the super-resolved 3D wind structures based on a newly developing SR network that utilizes a self-attention network and a generative model. It enables to incorporating of 3D dynamics of weather systems that are essential to reconstructing physically representative 3D wind fields, and then achieve to generate high-fidelity outputs. This work is a preliminary step toward building a tracing methodology that is capable of simulating the trajectory of GHGs combined with limited observation and reduces the computational overhead. Further algorithmic advancements should improve the accuracy of the methodology.

Acknowledgments and Disclosure of Funding

Supports for this work come from summer internship program at IBM Research, Yorktown Heights. The authors thank NeurIPS 2023 Tackling Climate Change with Machine Learning workshop committees for having opportunities to present our research outcomes.

References

- [1] Nicolaas J Annau, Alex J Cannon, and Adam H Monahan. Algorithmic hallucinations of near-surface winds: Statistical downscaling with generative adversarial networks to convection-permitting scales. *Artificial Intelligence for the Earth Systems*, 2023.
- [2] Wim C De Rooy, Peter Bechtold, Kristina Fröhlich, Cathy Hohenegger, Harm Jonker, Dmitrii Mironov, A Pier Siebesma, Joao Teixeira, and Jun-Ichi Yano. Entrainment and detrainment in cumulus convection: An overview. *Quarterly Journal of the Royal Meteorological Society*, 139(670):1–19, 2013.
- [3] Ruth DeFries, Frédéric Achard, Sandra Brown, Martin Herold, Daniel Murdiyarso, Bernhard Schlamadinger, and Carlos de Souza Jr. Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental science & policy*, 10(4):385–394, 2007.
- [4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [5] Zhang Jin-Yu, Chen Yan, and Huang Xian-Xiang. Edge detection of images based on improved sobel operator and genetic algorithms. In *2009 International Conference on Image Analysis and Signal Processing*, pages 31–35, 2009.
- [6] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [7] Rupa Kurinchi-Vendhan, Björn Lütjens, Ritwik Gupta, Lucien Werner, and Dava Newman. Wisosuper: Benchmarking super-resolution methods on wind and solar data. *arXiv preprint arXiv:2109.08770*, 2021.
- [8] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.
- [9] Jia Liu, Yongjian Sun, Kaijun Ren, Yanlai Zhao, Kefeng Deng, and Lizhe Wang. A spatial downscaling approach for windsat satellite sea surface wind based on generative adversarial networks and dual learning scheme. *Remote Sensing*, 14(3):769, 2022.
- [10] Salma Abdel Magid, Yulun Zhang, Donglai Wei, Won-Dong Jang, Zudi Lin, Yun Fu, and Hanspeter Pfister. Dynamic high-pass filtering and multi-spectral attention for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4288–4297, 2021.
- [11] William C Skamarock, Joseph B Klemp, Jimy Dudhia, David O Gill, Dale M Barker, Michael G Duda, Xiang-Yu Huang, Wei Wang, Jordan G Powers, et al. A description of the advanced research wrf version 3. *NCAR technical note*, 475:113, 2008.
- [12] Karen Stengel, Andrew Glaws, Dylan Hettinger, and Ryan N King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815, 2020.
- [13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [14] You Xie, Erik Franz, Mengyu Chu, and Nils Thuerey. tempogan: A temporally coherent, volumetric gan for super-resolution fluid flow. *ACM Transactions on Graphics (TOG)*, 37(4):1–15, 2018.
- [15] Mykhaylo Zayats, Małgorzata J. Zimoń, Kyongmin Yeo, and Sergiy Zhuk. Super resolution for turbulent flows in 2d: Stabilized physics informed neural networks. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 3377–3382, 2022.