FNP: Fourier Neural Processes for Arbitrary-Resolution Data Assimilation

Kun Chen [1,2]   **Peng Ye** [1,2]   **Hao Chen** [2]   **Kang Chen** [2]   **Tao Han** [2]

Wanli Ouyang [2]   **Tao Chen** [1,*]   **Lei Bai** [2,*]

[1] Fudan University   [2] Shanghai Artificial Intelligence Laboratory

kunchen22@m.fudan.edu.cn   eetchen@fudan.edu.cn   baisanshi@gmail.com

Abstract

Data assimilation is a vital component in modern global medium-range weather forecasting systems to obtain the best estimation of the atmospheric state by combining the short-term forecast and observations. Recently, AI-based data assimilation approaches have attracted increasing attention for their significant advantages over traditional techniques in terms of computational consumption. However, existing AI-based data assimilation methods can only handle observations with a specific resolution, lacking the compatibility and generalization ability to assimilate observations with other resolutions. Considering that complex real-world observations often have different resolutions, we propose the **F**ourier **N**eural **P**rocesses (FNP) for *arbitrary-resolution data assimilation* in this paper. Leveraging the efficiency of the designed modules and flexible structure of neural processes, FNP achieves state-of-the-art results in assimilating observations with varying resolutions, and also exhibits increasing advantages over the counterparts as the resolution and the amount of observations increase. Moreover, our FNP trained on a fixed resolution can directly handle the assimilation of observations with out-of-distribution resolutions and the observational information reconstruction task without additional fine-tuning, demonstrating its excellent generalization ability across data resolutions as well as across tasks. Code is available at https://github.com/OpenEarthLab/FNP.

1 Introduction

Accurately estimating the true state of complex and chaotic Earth systems is an important and challenging task, which can contribute to a better understanding of nature and improve forecasting by reducing the error of initial conditions. The most accurate human knowledge of the Earth's state comes from observations, which are inherently limited in their scopes due to practical constraints. Data assimilation, based on limited observational information and short-term forecasts (referred to as the *background*), serves as the primary approach for state estimation [38, 21, 32]. Traditional data assimilation methods employed in operational systems include Kalman filters based on minimum variance estimation and variational methods based on maximum likelihood estimation [4, 8, 49]. Taking 3D variational (3D-Var) data assimilation as an example, data assimilation is regarded as an optimization problem under given conditions, aiming to find the *analysis* x_a that minimizes the

---

*Corresponding authors

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

objective function $J(x)$. It can be formulated as

$$x_a = x^* = \arg\min_x J(x) \tag{1}$$

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx) \tag{2}$$

where B and R correspond to the error covariance matrix of the background x_b and observation y, respectively, and H is observation operator that maps state variables to observational space, aligning the background and observations with different modalities (for example, satellites do not directly observe state variables such as wind speed) and resolutions.

With the significant achievements of machine learning in medium-range weather prediction [44, 6, 31, 9, 11, 5, 35, 61], data assimilation has gained increasing attention as one of the core components in building end-to-end global weather forecasting systems. Compared to traditional methods, machine learning-based data assimilation models offer the potential for competitive results with significantly reduced resource consumption and execution time [10, 28, 62], making it a promising research direction with practical applications. Chen et al. [10] proposed a data assimilation model for weather variables based on the idea of gated masks, and combined it with FengWu [9], an advanced AI-based weather prediction model, to build the first end-to-end AI-based global weather forecasting system. Subsequently, data assimilation models integrated with other AI-based weather prediction models were propsed [28, 62]. These methods have demonstrated performance and efficiency improvements through various experiments, but *all of them can only assimilate observations with the same resolution as the forecasting model*. Therefore, they need to interpolate the observations onto grids of corresponding resolution through pre-processing in advance, and the pre-trained models do not have the flexibility and out-of-domain generalization to assimilate observations with other resolutions. The pre-processing step implements part of the function of the observation operator H to perform resolution alignment, and it will introduce additional errors inevitably, thereby affecting the performance and generality of data assimilation methods.

Neural processes [17, 18, 30, 20, 46, 42] offer a promising and universal data assimilation framework for addressing the aforementioned challenges. Neural processes are a series of conditional generative models that continuously model the distribution of functions and fields based on paired coordinate-value conditions, and generate values at arbitrary target locations based on coordinate indices. Their flexible features that allow for grid or off-grid data are well-suited for assimilating observational data with diverse forms, without requiring any prior interpolation or mapping [55, 2, 53, 16, 54]. *In this context, data assimilation is defined as the process of generating the analysis given both background conditions and observational information conditions.* The network models the comprehensive functional representation based on the two conditional inputs and decodes it to obtain the posterior distribution of the target. Compared to deterministic data assimilation, the modeling of distribution by neural processes can provide uncertainty estimates and further be used for ensemble data assimilation [23]. Moreover, data assimilation task degrades to observational information reconstruction when the background condition is missing. These two tasks can be broadly categorized as conditional generation, enabling their straightforward integration into a unified framework for direct application through simple fine-tuning.

In this paper, we propose the Fourier Neural Processes (FNP) for data assimilation with arbitrary-resolution observations. FNP is flexible to adapt to varying resolutions and can be extended to any conditional generation task. Leveraging the efficiency of the designed modules and flexible structure of neural processes, *FNP achieves state-of-the-art (SOTA) results in data assimilation experiments with different resolutions*, and demonstrates increasing advantages over other models as the resolution and amount of observational information increase. The visualization of the analysis showcases the promising performance of FNP in capturing high-frequency information. Importantly, the FNP trained at a fixed resolution can be directly applied to data assimilation with other resolutions and observational information reconstruction task without fine-tuning, highlighting its excellent out-of-domain generalization. Additionally, ablation study for different modules and experimental settings validate the effectiveness and robustness of our approach.

2 Related Work

Machine learning for data assimilation. There exist strong mathematical similarities between machine learning and data assimilation, enabling their integration within a unified Bayesian frame-

2

work [19, 3]. With their powerful nonlinear fitting capabilities and low computational cost, machine learning techniques can both enhance traditional data assimilation methods and provide alternative algorithms [13, 7, 22, 56]. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are often employed as surrogate models to replace computationally expensive components in data assimilation, such as tangent-linear and adjoint models in 4D variational (4D-Var) data assimilation [25], localization functions in ensemble Kalman filters (EnKF) [58], and error covariance matrices [12, 45]. Implicit neural representations (INRs) [33] and various autoencoders (AEs) [47, 41, 1] can offer efficient order reduction frameworks for latent assimilation to address the challenges of high-dimensional data. More recently, algorithms based on diffusion models have also provided new solutions for data assimilation driven by the advancements and maturity of AIGC [28, 52, 48]. However, all these studies are aimed at assimilating fixed-resolution observations, and *we are the first to focus on arbitrary-resolution data assimilation*.

Machine learning for observational information reconstruction. Observational information reconstruction is the process of recovering missing values and obtaining complete field information from limited sparse observations. Traditional reconstruction methods primarily rely on kriging interpolation and principal component analysis-based infilling [29]. As Kadow et al. [29] successfully applied the image inpainting techniques in computer vision to reconstruct the global temperature data, deep learning has been widely used in various reconstruction tasks [40, 59, 15, 57]. We associate the reconstruction with the data assimilation task here, and *the flexibility of our method allows the FNP pre-trained on data assimilation task to be directly applied to the observational information reconstruction without fine-tuning and achieve promising performance.*

Neural processes family and its application in geoscience. Neural processes combine the advantages of neural networks and Gaussian processes, and have demonstrated excellent performance in function regression, image completion and classification tasks [17, 18]. Attentive neural processes (ANP) [30] enable the network to learn location-relevant representations by introducing the attention mechanism, which improves the accuracy of predictions and broadens the scope of modeling. Convolutional conditional neural processes (ConvCNP) [20] model translation equivariance in the data, adding an important inductive bias to the model and enabling zero-shot generalization to out-of-domain tasks. Evidential conditional neural processes (ECNP) [43] replace the standard Gaussian distribution with a hierarchical Bayesian structure through evidence learning to achieve the decomposition of epistemic-aleatoric uncertainty. Neural processes and their variants, leveraging their unique advantages, have been successfully applied in geoscience tasks such as climate downscaling [55], sensor placement [2] and observational information reconstruction [53], and have shown promising performance. Here *we apply the neural processes to arbitrary-resolution data assimilation for the first time and achieve SOTA performance*, further demonstrating its huge application potential.

3 Methods

3.1 Model Overview

The overall framework and model details of FNP are depicted in Figure 1. Initially, the background and observations undergo a unified coordinate transformation to obtain the coordinates x^c and values y^c of the conditional points when they are input into the network. This ensures spatial alignment of them, even in the presence of disparate resolutions, modalities, and data formats. Subsequently, FNP models the two components of the conditional information globally to get their respective spatial-variable functional representation. The dynamic alignment and merge (DAM) module integrates and aligns these functional representation into the target domain, resulting in a comprehensive functional representation over the target space. Finally, multi-layer perceptrons (MLPs) are employed to decode the functional representation and output the mean and variance of the analysis based on the coordinates x^t of the target points. In the following subsections, we provide a detailed description of the process for modeling the functional representation and the internal structure of the DAM module.

3.2 Spatial-Variable Functional Representation

Modeling the functional representation involves two main steps: embedding the data sets into an infinite-dimensional function space and performing deep feature extraction. The former is accomplished through the *SetConv* layer [20], which is a generalized form of the standard convolutional

3
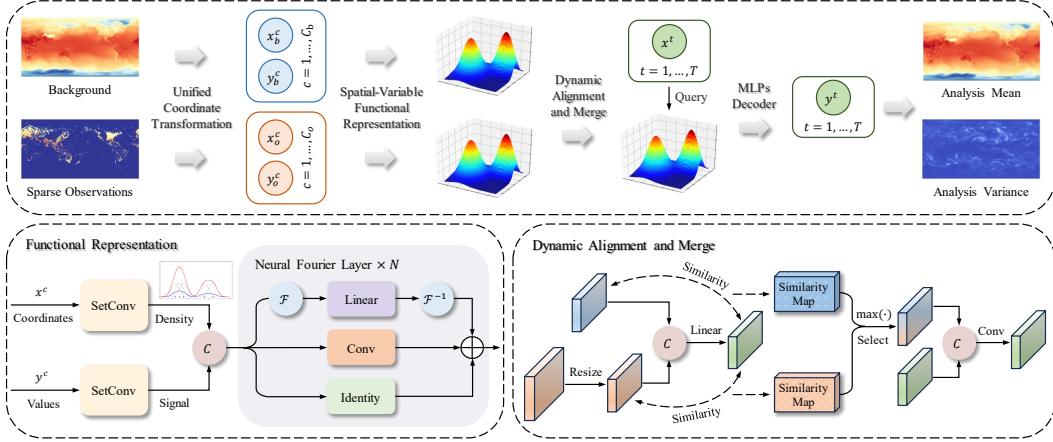
Figure 1: Overview of the network architecture of FNP. Unified coordinate transformation ensures spatial alignment of the background and observations, and extracts the coordinates and values of the conditional points. FNP models the two components of the conditional information globally to get their respective spatial-variable functional representation through *SetConv* for data embedding and stacking of neural Fourier layers for deep feature extraction. The dynamic alignment and merge module integrates these functional representation based on similarity to shared features and aligns them into the target domain, resulting in a comprehensive functional representation over the target space. MLPs are finally employed to decode the functional representation and output the mean and variance of the analysis based on the coordinates of the target points.

layer extended to operate on sets. It takes a set of continuous coordinate-value pairs as input and outputs a function that can be queried at continuous positions. The *SetConv* operation is permutation-invariant and includes an additional channel to estimate the density of the conditional points. When the input coordinates are discrete, *SetConv* essentially degenerates into a standard convolutional layer, simplifying the model into the on-the-grid version [14]. We strongly recommend readers to refer to the theoretical proofs and derivations of ConvCNP [20], and the project homepage of neural processes family [14] for a more detailed explanation.

The original deep feature extraction module is implemented using a standard CNN with residual structures. We choose to replace the basic convolutional layer with a more efficient neural Fourier layer (NFL) in FNP. Additionally, to address the multi-variable optimization problem in weather modeling tasks, we decouple the representations in spatial and variable dimensions to reduce the difficulty of network training. Below we provide further explanations on the motivations and implementation details of these design choices.

Spatial-variable decoupled representation. Data for different weather variables are usually stacked in the channel dimension, and direct data embedding will mix the spatial auto-correlation within variables with the inter-correlation among variables. An intuitive understanding is that explicitly separating the information in the spatial and variable dimensions allows for a clearer learning objective for each block, thereby reducing the difficulty of network training and fully unleashing the network's potential. In terms of implementation, we model a spatial functional representation separately for each meteorological variable, such as geopotential and temperature (surface variables are treated together as one variable), and model a variable functional representation that encompasses all variables, which are then concatenated together. The benefits of this approach have been confirmed in our experiments. We found that the spatial-variable decoupled (SVD) representation achieves better performance with fewer parameters and faster convergence speed. The detailed comparison of performance can be seen in Table 3.

Neural Fourier layer. The smoothness of neural network outputs poses a challenging drawback in weather modeling tasks, and the background generated by AI-based forecasting models also tends to be smoother. In our experiments, we find that neural processes also struggle to overcome the issue of smoothness. To address the desire for high-frequency information, we choose to introduce the

4

Fourier neural operator [34]. Besides, operations in the frequency domain can also bring additional advantages in terms of global receptive fields for models based on CNN. Therefore, in addition to the convolutional operation, each neural Fourier layer consists of a branch for linear operation in the frequency domain and a branch for identity mapping to preserve high-frequency details as much as possible [26].

3.3 Dynamic Alignment and Merge

The DAM module aligns functional representations from two conditional domains to the target domain for obtaining outputs at the target locations. In data assimilation tasks, the analysis typically shares the same resolution and modalities as the background, and the previous data embedding has already mapped inputs that may have different modalities into the same feature space. Therefore, it is only necessary to align the functional representation of the observation in the spatial resolution. We choose to use interpolation to adjust the spatial dimensions' size as it can accommodate inputs of arbitrary resolutions, thereby enhancing the dynamics and generality of the model. *Interpolation in the feature space differs fundamentally from that in the original observational space because the former has already extracted helpful information and contains redundancy to support dimensionality reduction, while the latter compresses valid information and missing values to the same extent.* The performance of data assimilation with different resolutions in Table 1 provides proof for this. As the amount of observational information increases, our model achieves significant improvements, while other models do not. A linear layer extracts shared features $\dot{y}$ from both parts after alignment, which are then used to calculate similarities with their respective feature components. The similarity calculation is performed in the channel dimension as the spatial distribution of information differs significantly between them, with the background having a more uniform spatial distribution while the observation exhibits greater spatial variability. In our implementation, the feature similarity is represented by the Euclidean distance between the two features, i.e.,

$$Sim_{h,w} = \sqrt{\sum_{i=1}^{k}(y_{h,w,i} - \dot{y}_{h,w,i})^2} \tag{3}$$

where h and w denote the indices for each grid point along the longitudinal and latitudinal directions, respectively, and k is the dimension of data embedding. The relative values of the similarity map then determines the selection of features. Specifically, features that are more similar to shared features will be retained, while features that are less similar will be discarded, that is,

$$y_{h,w} = \begin{cases} y_{h,w}^b, & Sim_{h,w}^b \geq Sim_{h,w}^o \\ y_{h,w}^o, & Sim_{h,w}^b < Sim_{h,w}^o \end{cases} \tag{4}$$

The dynamically filtered features will be spliced together with the shared features and sent to a convolutional layer for spatial smoothing, and the result will be used as the functional representation in the target domain for decoding and output.

4 Experiments

4.1 Experimental Settings and Implementation

Data preparation. We demonstrate the effectiveness of our methodology on the ERA5 dataset [27], a global atmospheric reanalysis archive containing hourly weather variables such as geopotential, temperature, wind speed, humidity, etc. We choose to conduct experiments on a total of 69 variables, including five upper-air variables with 13 pressure levels (i.e., 50hPa, 100hPa, 150hPa, 200hPa, 250hPa, 300hPa, 400hPa, 500hPa, 600hPa, 700hPa, 850hPa, 925hPa, and 1000hPa), and four surface variables. Specifically, the upper-air variables are geopotential (z), temperature (t), specific humidity (q), zonal component of wind (u) and meridional component of wind (v), whose 13 sub-variables at different vertical level are presented by abbreviating their short name and pressure levels (e.g., z500 denotes the geopotential at a pressure level of 500 hPa), and the surface variables are 10-meter zonal component of wind (u10), 10-meter meridional component of wind (v10), 2-meter temperature (t2m) and mean sea level pressure (msl). A subset of ERA5 dataset for 40 years, from 1979 to 2018, is chosen to train and evaluate the model.

5

Experimental settings. The advanced AI-based weather forecasting model, FengWu [9], is used as the surrogate model to generate the background. The observations are simulated by adding a proportional mask to ERA5, and the default setting corresponds to 24-hour forecast lead time and 10% observations. In other words, the background used for data assimilation is produced by FengWu (with 6-hour interval) through four auto-regressive iterative predictions based on ERA5 data from one day ago. The observational space usually has higher spatial resolution than the state space in actual operation systems. Therefore, the resolution of the forecasting model and background is set to 1.40625° (128 × 256 grid points) so that we can conduct experiments using observations with different resolutions such as 0.25° (721 × 1440 grid points) to verify the assimilation performance with arbitrary resolution.

Model training and evaluation. The FNP model is implemented based on the open-source code of the neural processes family project [14], and trained for 20 epochs using the AdamW optimizer [39] with a learning rate of 1e-4. We divide the ERA5 data from 1979-2015 as the training set, 2016-2017 as validation set, and 2018 as test set. The training is run on 4 NVIDIA Tesla A100 GPUs with a global batch size of 16, and takes approximately 2.5 days. The inference only needs a few minutes to perform data assimilation for a whole year on single A100 GPU. The dimension of data embedding for default setting is 128 and the number N of NFLs is 4, and a Gaussian likelihood is used with a negative log-likelihood (NLL) loss. We evaluate the performance of models by calculating the overall mean square error (MSE), mean absolute error (MAE), and the latitude-weighted root mean square error (WRMSE) which is a statistical metric widely used in geospatial analysis and atmospheric science [50, 51]. Given the estimate $\hat{x}_{h,w,c}$ and its ground truth $x_{h,w,c}$ for the c-th channel, the WRMSE is defined as

$$\text{WRMSE}(c) = \sqrt{\frac{1}{H \cdot W} \sum_{h,w} H \frac{\cos(\alpha_{h,w})}{\sum_{h'=1}^{H} \cos(\alpha_{h',w})} (x_{h,w,c} - \hat{x}_{h,w,c})^2} \tag{5}$$

where H and W represent the number of grid points in the longitudinal and latitudinal directions, respectively, and $\alpha_{h,w}$ is the latitude of point (h, w).

4.2 Arbitrary-Resolution Data Assimilation

We validate the performance of models by assimilating 10% observations with resolutions of 1.40625°, 0.703125°, and 0.25°, respectively, onto a 24-hour forecast background with 1.40625° resolution. Table 1 provides a quantitative comparison of the analysis errors between FNP and other models. The first row corresponds to the error level of the background. When assimilating observations with the same resolution as the forecasting model, *FNP achieves SOTA results (indicated in bold) in terms of overall MSE, MAE, and WRMSE metrics for the majority of variables.* Since Adas [10] is not flexible enough to support inputs with different resolutions, we follow its common practice to interpolate the observations and average the observations falling within the corresponding grid range when assimilating higher-resolution observations. Despite this, Adas still produces results with significantly high errors, so we only present the performance after fine-tuning to the interpolated observations. This indicates that Adas lacks the ability of out-of-domain generalization. In contrast, FNP and ConvCNP [20], with their flexible structures, can assimilate observations with different resolutions directly without interpolation. Therefore, the table presents the results for both cases with and without fine-tuning for FNP and ConvCNP.

In order to better understand and explain the performance of different models, we add different colors to represent the variations in results compared to that with 1.40625° resolution (blue indicating worse results, i.e., increased errors, and red indicating improved results). For the q700 variable, we additionally annotate the percentage of error increase or decrease. It is worth noting that with increasing resolution, the same ratio of observations implies a greater number of absolute observations and a larger amount of information. However, during the interpolation process, averaging the observation values within a region does not guarantee a reflection of the overall conditions unless there are observations at all points within the region. Therefore, interpolation inevitably leads to information loss, and the amount of lost information is negatively correlated with the number of observations within the region. Based on the balance between these two factors, the fine-tuned Adas exhibits different trends with two different resolutions: the results generally improve when assimilating observations with 0.703125° resolution, while all the results worsen when assimilating observations with 0.25° resolution. This indicates that as the resolution increases gradually, the impact of information loss due to interpolation becomes more significant and surpasses the positive effect of

6

Table 1: Quantitative performance comparison for arbitrary-resolution data assimilation. The best performance are shown in bold while the second best is underscored. Red color indicates the improved assimilation results compared to that with 1.40625° resolution, and blue color indicates worse results.

Model	Resolution	MSE	MAE	WRMSE							
				z500	t850	t2m	u10	v10	u500	v500	q700 (10^{-4})
Background	1.40625°	0.0288	0.0861	45.455	0.7200	0.7790	0.9336	0.9645	1.7278	1.7535	6.7220
Adas [10]	1.40625°	0.0221	0.0705	**22.930**	**0.6323**	0.7198	0.8198	0.8369	1.4673	1.4489	6.4955
ConvCNP [20]	1.40625°	0.0252	0.0795	31.253	0.6831	0.7662	0.8334	0.8553	1.5770	1.5876	6.5717
FNP (ours)	1.40625°	**0.0202**	**0.0664**	26.910	0.6600	**0.7049**	**0.7558**	**0.7665**	**1.4263**	**1.4450**	**6.4691**
Adas [10]	0.703125°	0.0144	0.0614	26.423	0.5981	0.6696	0.6631	0.6552	1.2935	1.1975	5.3798 ↓17.2%
ConvCNP [20] w/o fine-tuning	0.703125°	0.0378	0.1134	142.38	0.9106	0.8471	1.2264	1.1115	2.9765	2.6264	5.7900 ↓11.9%
ConvCNP [20]	0.703125°	0.0202	0.0712	25.452	0.5923	0.7861	0.5812	0.6339	1.3795	1.4331	5.7562 ↓12.4%
FNP (ours) w/o fine-tuning	0.703125°	0.0150	0.0618	52.075	0.6021	0.7385	0.7052	0.6833	1.3723	1.3036	5.4333 ↓16.0%
FNP (ours)	0.703125°	**0.0085**	**0.0419**	**20.710**	**0.5805**	**0.5310**	**0.4457**	**0.4406**	**0.9032**	**0.9180**	**4.5296 ↓30.0%**
Adas [10]	0.25°	0.0231	0.0766	34.694	0.6979	0.7728	0.8455	0.8537	1.5864	1.5265	6.5019 ↑0.10%
ConvCNP [20] w/o fine-tuning	0.25°	0.0651	0.1514	234.50	1.3208	0.8705	1.7134	1.5122	4.1316	3.6655	6.6619 ↑1.37%
ConvCNP [20]	0.25°	0.0280	0.0831	26.437	0.6326	0.7429	0.9283	0.9640	1.7265	1.7533	6.3612 ↓3.20%
FNP (ours) w/o fine-tuning	0.25°	0.0196	0.0729	84.516	0.6992	0.6912	0.9264	0.8687	1.8315	1.6723	6.1761 ↓4.53%
FNP (ours)	0.25°	**0.0058**	**0.0339**	**19.996**	**0.4549**	**0.4999**	**0.3839**	**0.3753**	**0.6308**	**0.5320**	**2.9614 ↓54.2%**

increased observation quantity. In practice, the observational data used in operational systems usually have resolutions of 0.1° or even higher, while the resolution of the most commonly used forecasting models is 0.25°. Therefore, the information loss caused by interpolation in existing methods is a very common and urgently addressed issue.

In contrast, *the fine-tuned FNP not only achieves SOTA performance in all metrics (including the z500 and t850 variables, in which FNP does not reach the optimum results with 1.40625° resolution), but also improves all assimilation results with the largest magnitude of error reduction.* The performance differences between FNP and other models increases significantly with increasing resolution, and the WRMSE decreases on some variables such as v10, u500, v500 and q700 are even more than 50% when assimilating observations with 0.25° resolution. Furthermore, as the absolute number of observations increases, providing more information, FNP is the only model that consistently improves the performance of data assimilation. This plays a crucial role in practical applications, as it means that all deployed observation instruments can be fully utilized, reducing the waste of human, material and financial resources as much as possible.

The results without fine-tuning reflect the out-of-domain generalization capability of models, as they have not encountered observational data with other resolutions during training. As the resolution increases, the discrepancy between the samples used for testing and the visible samples during training becomes larger, leading to a gradual decline in performance. This regular pattern can be observed in the assimilation results of both FNP and ConvCNP. However, the increased quantity of observations and information will bring the benefits, resulting in improved performance for some variables in out-of-domain settings, although the number of such variables decreases as the resolution increases. FNP consistently outperforms ConvCNP in terms of the number of variables showing improvement and even exhibits superior performance to fine-tuned versions of other models in some variables. *This demonstrates the excellent out-of-domain generalization capability of our method in adapting to changes in resolution*, and this capability is also applicable to changes in background resolution theoretically.

Figure 2 presents the visualization of assimilation results by different models for q700, with the visualization date-time randomly selected at 2018-04-02 06:00 UTC. The first row displays the ERA5 (ground truth), background, background error (background minus ERA5), and raw observations with a resolution of 0.25°. Other rows show the analysis, analysis increment (analysis minus background), and analysis error (analysis minus ERA5) obtained through data assimilation by different models, as well as the interpolated observations for Adas and analysis variances for ConvCNP and FNP. The background is smoother compared with ERA5, and the background error shows high spatial variability. It can be observed that *FNP accurately captures the distribution pattern of the background error, leading to analysis with rich high-frequency information and significantly reduced analysis error*. The comparison with ConvCNP, which also assimilates raw observations directly, confirms that FNP's excellent ability to capture high-frequency features does not sorely rely on higher-resolution observations. Furthermore, the smaller analysis variance also indicates a lower uncertainty in state

7

estimation achieved by FNP. More visualizations with different variables and resolutions are shown in Appendix B.
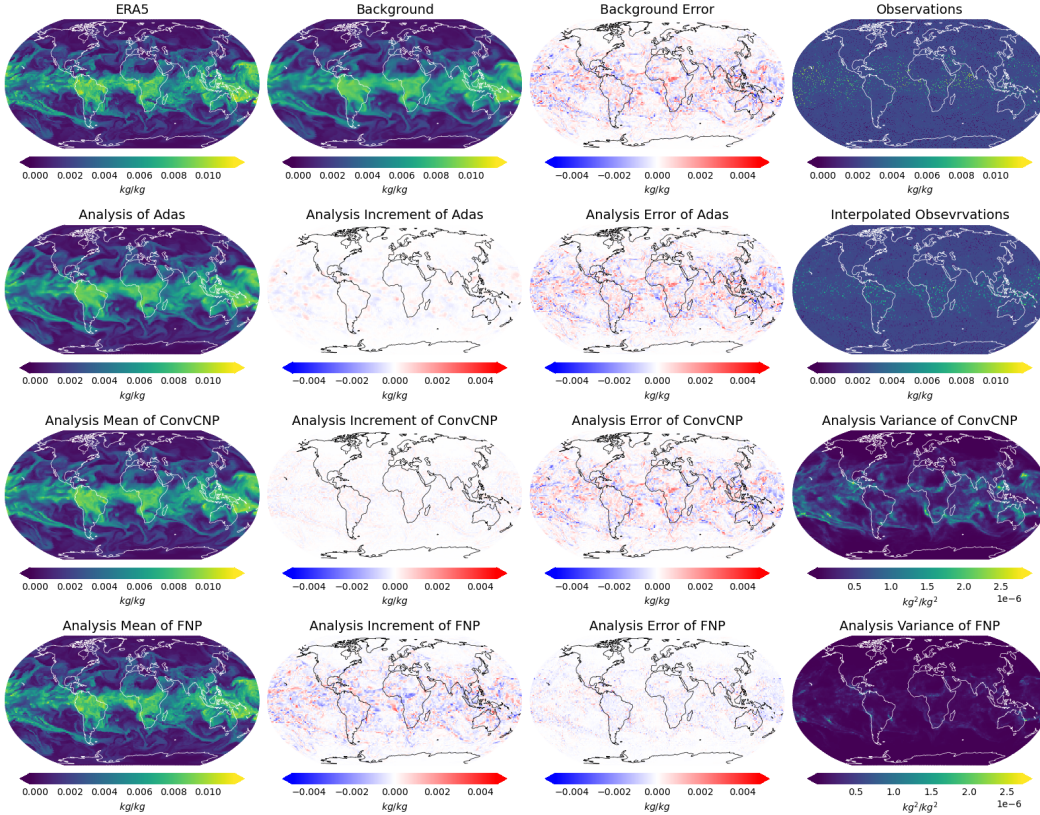
Figure 2: Visualization of assimilation results by different models for q700. The visualization date-time is randomly selected at 2018-04-02 06:00 UTC. The first row shows the ERA5 (ground truth), background, background error and observations with 0.25° resolution. Other rows show the assimilation results of different models.

Data assimilation aims to improve weather forecasting results by reducing initial errors. Therefore, it is crucial to explore the impact of different methods on forecast error reduction. Figure 3 provides results on the forecast WRMSE improvement of z500 variable over the next ten days through data assimilation, where lead time 0 corresponds to the reduction of initial errors. Darker colors indicate stronger improvements, meaning a greater reduction in forecast errors compared to not using data assimilation. Similar to the results of data assimilation, FNP consistently achieves state-of-the-art results in most cases, with its advantage becoming more pronounced as the resolution increases. Moreover, FNP is the only model that strictly enhances forecast improvement with increasing resolution and observational information. Additionally, apart from the accuracy of initial values affecting forecast errors, the physical characteristics of the initial states (e.g., physical balance) also influence the rate of forecast error growth. FNP demonstrates greater improvements in forecast errors at all lead times compared to improvements in initial errors, indicating that FNP not only reduces forecast errors but also slows down the growth rate of forecast errors. Other models do not exhibit the same trend, further highlighting the superior characteristics of the initial states produced by FNP.

4.3 Generalization to Observational Information Reconstruction

Theoretically, the functional representation learned based on observational conditions can be directly decoded through MLPs and output reconstruction results for the observational information without fine-tuning. Therefore, we evaluated the reconstruction performance of different models in the absence of the background conditions, as shown in Table 2. Similarly, Adas pre-trained on data assimilation task, cannot be directly used for information reconstruction. Hence, the table only

8

z500 WRMSE improvement of forecasts (m^2/s^2)

Adas (1.40625°)	22.52	25.11	29.46	34.72	39.62	44.75	49.74	54.21	55.79	55.75	54.16	51.23	48.12	43.95	39.49	35.50	31.03	27.41	24.79	21.96	18.18
ConvCNP (1.40625°)	14.20	17.55	22.57	28.25	33.24	36.91	40.09	43.75	44.46	44.50	44.15	42.92	40.33	36.45	32.47	29.40	26.85	25.35	24.68	23.24	19.41
FNP (1.40625°)	18.55	24.32	29.15	35.01	40.79	45.78	50.06	54.19	54.90	54.15	52.86	51.84	50.92	48.75	45.71	43.11	41.02	39.46	36.95	32.85	27.37
Adas (0.703125°)	19.03	24.69	29.32	34.83	40.42	45.36	49.71	54.29	55.72	55.70	54.71	52.50	49.34	44.78	39.99	35.86	32.38	29.63	26.68	22.73	17.35
ConvCNP (0.703125°)	20.00	24.06	28.74	33.65	39.02	43.31	46.77	50.47	51.13	50.70	50.39	48.92	46.52	42.64	38.10	34.12	30.62	28.21	26.08	23.17	18.56
FNP (0.703125°)	24.75	29.88	36.18	43.31	50.49	56.95	62.92	69.38	71.61	72.34	72.26	70.18	66.99	63.04	58.73	53.73	48.91	45.09	41.30	35.99	29.54
Adas (0.25°)	10.76	16.84	20.89	25.32	28.88	32.48	37.08	41.18	42.20	42.21	41.86	39.66	36.37	31.53	26.89	23.62	20.65	18.23	15.82	13.47	10.73
ConvCNP (0.25°)	19.02	9.67	10.57	12.28	14.54	16.69	18.57	20.33	20.57	19.72	17.78	16.59	16.01	14.98	13.84	12.94	11.71	11.25	11.19	11.06	9.99
FNP (0.25°)	25.46	30.23	36.55	44.07	51.66	58.76	66.56	74.85	78.71	80.60	80.62	78.38	74.37	69.88	64.59	59.99	55.28	52.28	50.14	45.64	38.06

Lead time (day): 0 — 1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Figure 3: Quantitative comparison of different data assimilation methods on the improvement of forecast errors for the next ten days.

presents the performance of retrained Adas, while both FNP and ConvCNP show results with and without fine-tuning. *The fine-tuned FNP achieves SOTA performance across all metrics, while FNP without fine-tuning also demonstrates good generalization.*

Table 2: Performance comparison in observational information reconstruction with 10% observations.

Model	Ratio	MSE	MAE	WRMSE							
				z500	t850	t2m	u10	v10	u500	v500	q700 (10^{-4})
Adas [10]	10%	0.0370	0.1006	46.592	1.0482	1.3715	<u>1.1539</u>	<u>1.1711</u>	2.0015	1.9932	9.5399
ConvCNP [20] w/o fine-tuning	10%	0.0376	0.1042	67.098	1.0630	1.4864	1.2482	1.2756	1.9896	2.0312	9.3894
ConvCNP [20]	10%	<u>0.0350</u>	<u>0.0933</u>	59.864	<u>0.8789</u>	<u>0.8219</u>	1.2165	1.2145	1.9715	2.0069	<u>6.8643</u>
FNP (ours) w/o fine-tuning	10%	0.0355	0.1014	<u>45.835</u>	1.0378	1.3382	1.1723	1.2024	<u>1.9559</u>	<u>1.9075</u>	9.1517
FNP (ours)	10%	**0.0291**	**0.0876**	**35.491**	**0.7298**	**0.8027**	**0.9411**	**0.9716**	**1.7464**	**1.7564**	**6.7258**

4.4 Ablation Study

We conduct ablation experiments on both the designed modules employed in FNP and experimental settings. Table 3 presents the quantitative performance comparison of FNP when different components are replaced. The overall MSE, MAE, and WRMSE metrics of all variables exhibit varying degrees of performance degradation when a specific module in FNP is replaced. FNP achieves the best performance when these designed components work in synergy and mutually reinforce each other.

Table 3: Ablation study of different components in FNP for data assimilation with 1.40625° resolution.

Model	Resolution	MSE	MAE	WRMSE							
				z500	t850	t2m	u10	v10	u500	v500	q700 (10^{-4})
Background	1.40625°	0.0288	0.0861	45.455	0.7200	0.7790	0.9336	0.9645	1.7278	1.7535	6.7220
FNP w/o NFL	1.40625°	0.0230	0.0749	30.040	0.6733	0.7504	0.8212	0.8254	1.5459	1.5566	6.4941
FNP w/o DAM	1.40625°	0.0214	0.0702	30.189	0.6816	0.7138	0.8248	0.8090	1.4674	1.4831	6.5631
FNP w/o SVD	1.40625°	0.0239	0.0757	27.588	0.6683	0.7555	0.7805	0.8430	1.5743	1.5789	6.5538
FNP	1.40625°	**0.0202**	**0.0664**	**26.910**	**0.6600**	**0.7049**	**0.7558**	**0.7665**	**1.4263**	**1.4450**	**6.4691**

Ablation study on the experimental settings is conducted by changing the forecast lead time of the background and the ratio of observations while keeping a fixed resolution of 1.40625°. Table 4 provides a quantitative performance comparison of different models when the observation proportion is reduced to 1% and when the forecast lead time of the background is extended to 48 hours. In these scenarios, all the models exhibit robustness and consistently improve the background. When the number of observations decreases or the background error increases, the amount of conditional

9

information they can provide becomes less, so it is reasonable to observe an increase in analysis error compared to Table 1. Similarly, *FNP achieves SOTA performance in terms of overall MSE, MAE, and WRMSE for the majority of variables.*

Table 4: Ablation study on the forecast lead time of the background and the ratio of observations.

Model	Ratio	MSE	MAE	WRMSE							
				z500	t850	t2m	u10	v10	u500	v500	q700 (10^{-4})
24h Background	-	0.0288	0.0861	45.455	0.7200	0.7790	0.9336	0.9645	1.7278	1.7535	6.7220
Adas [10]	1%	0.0272	0.0835	**30.595**	0.7178	0.7771	0.9115	0.9406	1.6881	1.7141	**6.7074**
ConvCNP [20]	1%	0.0283	0.0856	43.272	0.7186	0.7785	0.9286	0.9574	1.7203	1.7428	6.7173
FNP (ours)	1%	**0.0269**	**0.0812**	30.678	**0.6971**	**0.7708**	**0.9069**	**0.9351**	**1.6586**	**1.6778**	6.7084
48h Background	-	0.0467	0.1148	85.017	0.9103	0.9283	1.2525	1.2882	2.3583	2.3856	8.2664
Adas [10]	10%	0.0305	0.0840	30.473	**0.7355**	0.8134	1.0047	1.0176	1.7747	1.7217	7.8209
ConvCNP [20]	10%	0.0334	0.0918	40.378	0.7774	0.8619	0.9251	0.9422	1.8112	1.8160	7.7774
FNP (ours)	10%	**0.0252**	**0.0743**	**28.883**	0.7459	**0.7810**	**0.8307**	**0.8517**	**1.6285**	**1.6348**	**7.6078**

5   Conclusions

In summary, we present FNP that can assimilate observations with arbitry resolution. The outstanding performance and out-of-domain generalization of FNP in data assimilation and observational information reconstruction demonstrate its significant potential and broad application prospects. It not only contributes to the field of data assimilation but also makes meaningful explorations for AI-based end-to-end weather forecasting systems. Technically and theoretically, this methodology can also be applied to more tasks and scenarios, such as downscaling [37], station-scale state estimation and weather prediction [36, 60, 24].

Our work has certain limitations. Firstly, the observational data used in our experiments are generated through simulations rather than real-world observations. This may lead to differences in model performance when applied in actual scenarios, thus discounting its value for practical application. In fact, due to the complex and diverse nature of real observational data, the data assimilation community lacks relevant benchmarks and large-scale datasets. The establishment of such benchmarks and datasets would be a highly meaningful endeavor, enabling fair comparisons among different models and fostering rapid advancements in the field. Secondly, FNP inherently performs 3D data assimilation without the temporal dimension. Considering the flexibility of the FNP architecture, incorporating the temporal dimension is not challenging and is expected to produce additional benefits. In the future, we will further explore the broader possibilities of data assimilation and end-to-end weather forecasting.

Acknowledgments and Disclosure of Funding

This work is supported by National Natural Science Foundation of China (No. 62071127 and 62101137), National Key Research and Development Program of China (No. 2022ZD0160101), Shanghai Natural Science Foundation (No. 23ZR1402900), Shanghai Municipal Science and Technology Major Project (No.2021SHZDZX0103) and Shanghai Artificial Intelligence Laboratory.

References

[1] Maddalena Amendola, Rossella Arcucci, Laetitia Mottet, César Quilodrán Casas, Shiwei Fan, Christopher Pain, Paul Linden, and Yi-Ke Guo. Data assimilation in the latent space of a convolutional autoencoder. In *International Conference on Computational Science*, 2021.

[2] Tom R Andersson, Wessel P Bruinsma, Stratis Markou, James Requeima, Alejandro Coca-Castro, Anna Vaughan, Anna-Louise Ellis, Matthew A Lazzara, Dani Jones, Scott Hosking, et al. Environmental sensor placement with convolutional gaussian neural processes. *Environmental Data Science*, 2023.

[3] Rossella Arcucci, Jiangcheng Zhu, Shuang Hu, and Yi-Ke Guo. Deep data assimilation: integrating deep learning with data assimilation. *Applied Sciences*, 2021.

10

[4] Mark Asch, Marc Bocquet, and Maëlle Nodet. *Data assimilation: methods, algorithms, and applications*. 2016.

[5] Zied Ben Bouallègue, Mariana CA Clare, Linus Magnusson, Estibaliz Gascon, Michael Maier-Gerber, Martin Janoušek, Mark Rodwell, Florian Pinault, Jesper S Dramsch, Simon TK Lang, et al. The rise of data-driven weather forecasting: A first statistical assessment of machine learning-based weather forecasts in an operational-like context. *Bulletin of the American Meteorological Society*, 2024.

[6] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 2023.

[7] Caterina Buizza, César Quilodrán Casas, Philip Nadler, Julian Mack, Stefano Marrone, Zainab Titus, Clémence Le Cornec, Evelyn Heylen, Tolga Dur, Luis Baca Ruiz, et al. Data learning: Integrating data assimilation and machine learning. *Journal of Computational Science*, 2022.

[8] Alberto Carrassi, Marc Bocquet, Laurent Bertino, and Geir Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 2018.

[9] Kang Chen, Tao Han, Junchao Gong, Lei Bai, Fenghua Ling, Jing-Jia Luo, Xi Chen, Leiming Ma, Tianning Zhang, Rui Su, et al. Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv preprint arXiv:2304.02948*, 2023.

[10] Kun Chen, Lei Bai, Fenghua Ling, Peng Ye, Tao Chen, Kang Chen, Tao Han, and Wanli Ouyang. Towards an end-to-end artificial intelligence driven global weather forecasting system. *arXiv preprint arXiv:2312.12462*, 2023.

[11] Lei Chen, Xiaohui Zhong, Feng Zhang, Yuan Cheng, Yinghui Xu, Yuan Qi, and Hao Li. Fuxi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, 2023.

[12] Sibo Cheng and Mingming Qiu. Observation error covariance specification in dynamical systems for data assimilation using recurrent neural networks. *Neural Computing and Applications*, 2022.

[13] Sibo Cheng, César Quilodrán-Casas, Said Ouala, Alban Farchi, Che Liu, Pierre Tandeo, Ronan Fablet, Didier Lucor, Bertrand Iooss, Julien Brajard, et al. Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. *IEEE/CAA Journal of Automatica Sinica*, 2023.

[14] Yann Dubois, Jonathan Gordon, and Andrew YK Foong. Neural process family. `http://yanndubs.github.io/Neural-Process-Family/`, 2020.

[15] Marius Egli, Sebastian Sippel, Angeline G Pendergrass, Iris de Vries, and Reto Knutti. Reconstruction of zonal precipitation from sparse historical observations using climate model information and statistical learning. *Geophysical Research Letters*, 2022.

[16] Andrew Foong, Wessel Bruinsma, Jonathan Gordon, Yann Dubois, James Requeima, and Richard Turner. Meta-learning stationary stochastic process prediction with convolutional neural processes. *Advances in Neural Information Processing Systems*, 33:8284–8295, 2020.

[17] Marta Garnelo, Dan Rosenbaum, Christopher Maddison, Tiago Ramalho, David Saxton, Murray Shanahan, Yee Whye Teh, Danilo Rezende, and SM Ali Eslami. Conditional neural processes. In *International conference on machine learning*, 2018.

[18] Marta Garnelo, Jonathan Schwarz, Dan Rosenbaum, Fabio Viola, Danilo J Rezende, SM Eslami, and Yee Whye Teh. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018.

[19] Alan J Geer. Learning earth system models from observations: machine learning or data assimilation? *Philosophical Transactions of the Royal Society A*, 2021.

[20] Jonathan Gordon, Wessel P Bruinsma, Andrew YK Foong, James Requeima, Yann Dubois, and Richard E Turner. Convolutional conditional neural processes. In *International Conference on Learning Representations*, 2019.

[21] Nils Gustafsson, Tijana Janjić, Christoph Schraff, Daniel Leuenberger, Martin Weissmann, Hendrik Reich, Pierre Brousseau, Thibaut Montmerle, Eric Wattrelot, Antonín Bučánek, et al. Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 2018.

[22] Yoo-Geun Ham, Yong-Sik Joo, Jeong-Hwan Kim, Kang-Min Kim, and Jeong-Gil Lee. Partial-convolution-implemented generative adversarial network (gan) for global oceanic data assimilation. 2022.

[23] Thomas M Hamill. Ensemble-based atmospheric data assimilation. *Predictability of weather and climate*, 2006.

[24] Tao Han, Song Guo, Zhenghao Chen, Wanghan Xu, and Lei Bai. Weather-5k: A large-scale global station weather dataset towards comprehensive time-series forecasting benchmark. *arXiv preprint arXiv:2406.14399*, 2024.

[25] Sam Hatfield, Matthew Chantry, Peter Dueben, Philippe Lopez, Alan Geer, and Tim Palmer. Building tangent-linear and adjoint models for data assimilation with neural networks. *Journal of Advances in Modeling Earth Systems*, 2021.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[27] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 2020.

[28] Langwen Huang, Lukas Gianinazzi, Yuejiang Yu, Peter D Dueben, and Torsten Hoefler. Diffda: a diffusion model for weather-scale data assimilation. *arXiv preprint arXiv:2401.05932*, 2024.

[29] Christopher Kadow, David Matthew Hall, and Uwe Ulbrich. Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, 2020.

[30] Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, and Yee Whye Teh. Attentive neural processes. In *International Conference on Learning Representations*, 2018.

[31] Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 2023.

[32] François-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, 1986.

[33] Zhuoyuan Li, Bin Dong, and Pingwen Zhang. Latent assimilation with implicit neural representations for unknown dynamics. *Journal of Computational Physics*, 2024.

[34] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[35] Fenghua Ling, Lin Ouyang, Boufeniza Redouane Larbi, Jing-Jia Luo, Tao Han, Xiaohui Zhong, and Lei Bai. Is artificial intelligence providing the second revolution for weather forecasting? *arXiv preprint arXiv:2401.16669*, 2024.

[36] Zili Liu, Hao Chen, Lei Bai, Wenyuan Li, Keyan Chen, Zhengyi Wang, Wanli Ouyang, Zhengxia Zou, and Zhenwei Shi. Deriving accurate surface meteorological states at arbitrary locations via observation-guided continous neural field modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[37] Zili Liu, Hao Chen, Lei Bai, Wenyuan Li, Wanli Ouyang, Zhengxia Zou, and Zhenwei Shi. Mambads: Near-surface meteorological field downscaling with topography constrained selective state space modeling. *arXiv preprint arXiv:2408.10854*, 2024.

[38] Andrew C Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 1986.

[39] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in adam. 2018.

[40] Ziqi Ma, Jianbin Huang, Xiangdong Zhang, Yong Luo, Minghu Ding, Jun Wen, Weixin Jin, Chen Qiao, and Yifu Yin. Newly reconstructed arctic surface air temperatures for 1979–2021 with deep learning method. *Scientific Data*, 2023.

[41] Boštjan Melinc and Žiga Zaplotnik. Neural-network data assimilation using variational autoencoder. *arXiv preprint arXiv:2308.16073*, 2023.

12

[42] Peiman Mohseni and Nick Duffield. Spectral convolutional conditional neural processes. *arXiv preprint arXiv:2404.13182*, 2024.

[43] Deep Shankar Pandey and Qi Yu. Evidential conditional neural processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[44] Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

[45] Stephen G Penny, Timothy A Smith, T-C Chen, Jason A Platt, H-Y Lin, Michael Goodliff, and Henry DI Abarbanel. Integrating recurrent neural networks with data assimilation for scalable data-driven state estimation. *Journal of Advances in Modeling Earth Systems*, 2022.

[46] Jens Petersen, Gregor Köhler, David Zimmerer, Fabian Isensee, Paul F Jäger, and Klaus H Maier-Hein. Gp-convcnp: Better generalization for convolutional conditional neural processes on time series data. *arXiv preprint arXiv:2106.04967*, 2021.

[47] Mathis Peyron, Anthony Fillion, Selime Gürol, Victor Marchais, Serge Gratton, Pierre Boudier, and Gael Goret. Latent space data assimilation by using deep learning. *Quarterly Journal of the Royal Meteorological Society*, 2021.

[48] Yongquan Qu, Juan Nathaniel, Shuolin Li, and Pierre Gentine. Deep generative data assimilation in multimodal setting. *arXiv preprint arXiv:2404.06665*, 2024.

[49] Florence Rabier and Zhiquan Liu. Variational data assimilation: theory and overview. In *Proc. ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, Reading, UK, September 8–12*, 2003.

[50] Stephan Rasp, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 2020.

[51] Stephan Rasp, Stephan Hoyer, Alexander Merose, Ian Langmore, Peter Battaglia, Tyler Russel, Alvaro Sanchez-Gonzalez, Vivian Yang, Rob Carver, Shreya Agrawal, et al. Weatherbench 2: A benchmark for the next generation of data-driven global weather models. *arXiv preprint arXiv:2308.15560*, 2023.

[52] François Rozet and Gilles Louppe. Score-based data assimilation. *Advances in Neural Information Processing Systems*, 2023.

[53] Jonas Scholz, Tom R Andersson, Anna Vaughan, James Requeima, and Richard E Turner. Sim2real for environmental neural processes. *arXiv preprint arXiv:2310.19932*, 2023.

[54] Anna Vaughan, Stratis Markou, Will Tebbutt, James Requeima, Wessel P Bruinsma, Tom R Andersson, Michael Herzog, Nicholas D Lane, J Scott Hosking, and Richard E Turner. Aardvark weather: end-to-end data-driven weather forecasting. *arXiv preprint arXiv:2404.00411*, 2024.

[55] Anna Vaughan, Will Tebbutt, J Scott Hosking, and Richard E Turner. Convolutional conditional neural processes for local climate downscaling. *Geoscientific Model Development*, 2022.

[56] Wuxin Wang, Weicheng Ni, Tao Han, Lei Bai, Boheng Duan, and Kaijun Ren. Dabench: A benchmark dataset for data-driven weather data assimilation. *arXiv preprint arXiv:2408.11438*, 2024.

[57] Yueya Wang, Xiaoming Shi, Lili Lei, and Jimmy Chi-Hung Fung. Deep learning augmented data assimilation: Reconstructing missing information with convolutional autoencoders. *Monthly Weather Review*, 2022.

[58] Zhongrui Wang, Lili Lei, Jeffrey L Anderson, Zhe-Min Tan, and Yi Zhang. Convolutional neural network-based adaptive localization for an ensemble kalman filter. *Journal of Advances in Modeling Earth Systems*, 2023.

[59] Martin Wegmann and Fernando Jaume-Santero. Artificial intelligence achieves easy-to-adapt nonlinear global temperature reconstructions using minimal local data. *Communications Earth & Environment*, 2023.

[60] Haixu Wu, Hang Zhou, Mingsheng Long, and Jianmin Wang. Interpretable weather forecasting for worldwide stations with a unified deep model. *Nature Machine Intelligence*, 5(6):602–611, 2023.

13

[61] Wanghan Xu, Kang Chen, Tao Han, Hao Chen, Wanli Ouyang, and Lei Bai. Extremecast: Boosting extreme value prediction for global weather forecast. *arXiv preprint arXiv:2402.01295*, 2024.

[62] Xiaoze Xu, Xiuyu Sun, Wei Han, Xiaohui Zhong, Lei Chen, and Hao Li. Fuxi-da: A generalized deep learning data assimilation framework for assimilating satellite observations. *arXiv preprint arXiv:2404.08522*, 2024.

A Data Availability

The ERA5 dataset can be downloaded from the official website of Climate Data Store (CDS) at `https://cds.climate.copernicus.eu`.

B More Visualizations for Data Assimilation
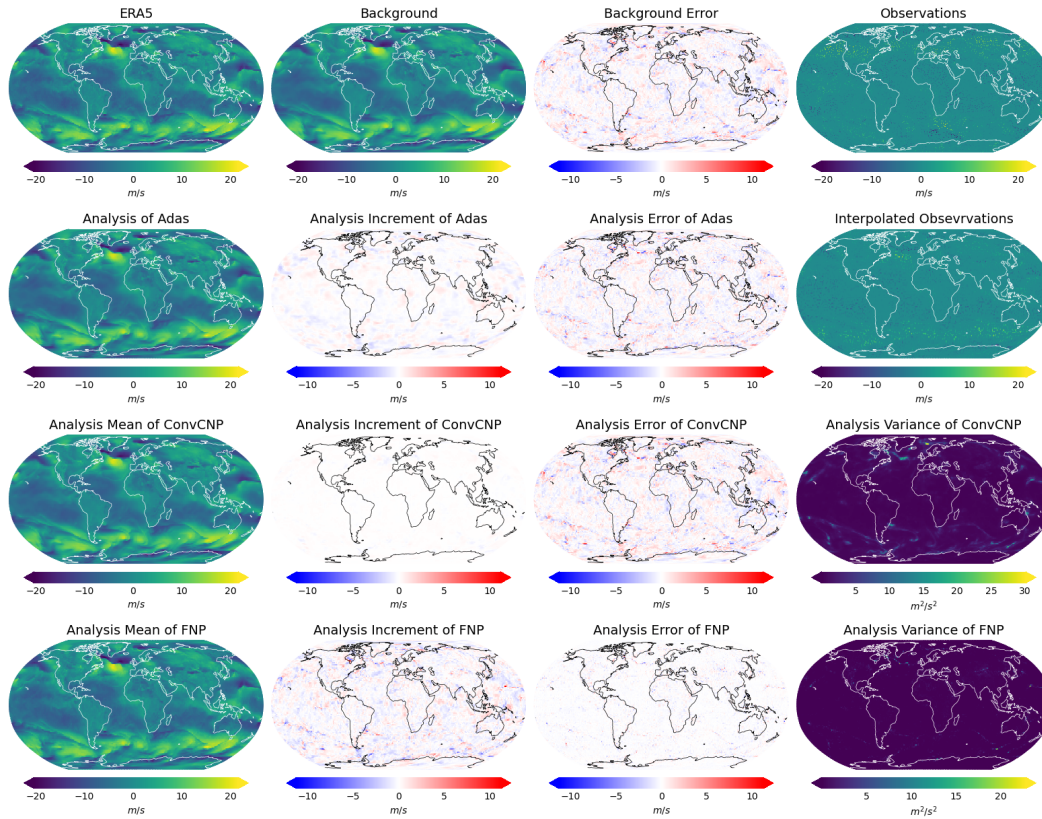
Figure 4: Visualization of assimilation results with 0.25° resolution for u10.

15
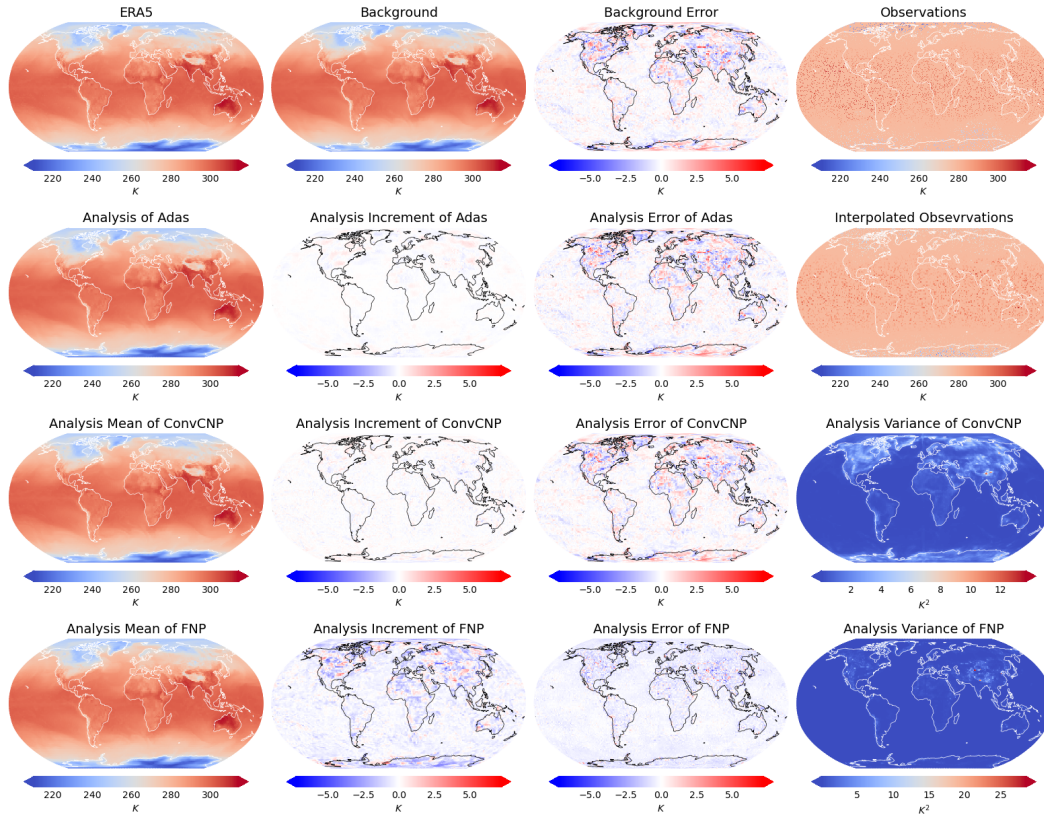
Figure 5: Visualization of assimilation results with 0.25° resolution for t2m.

16
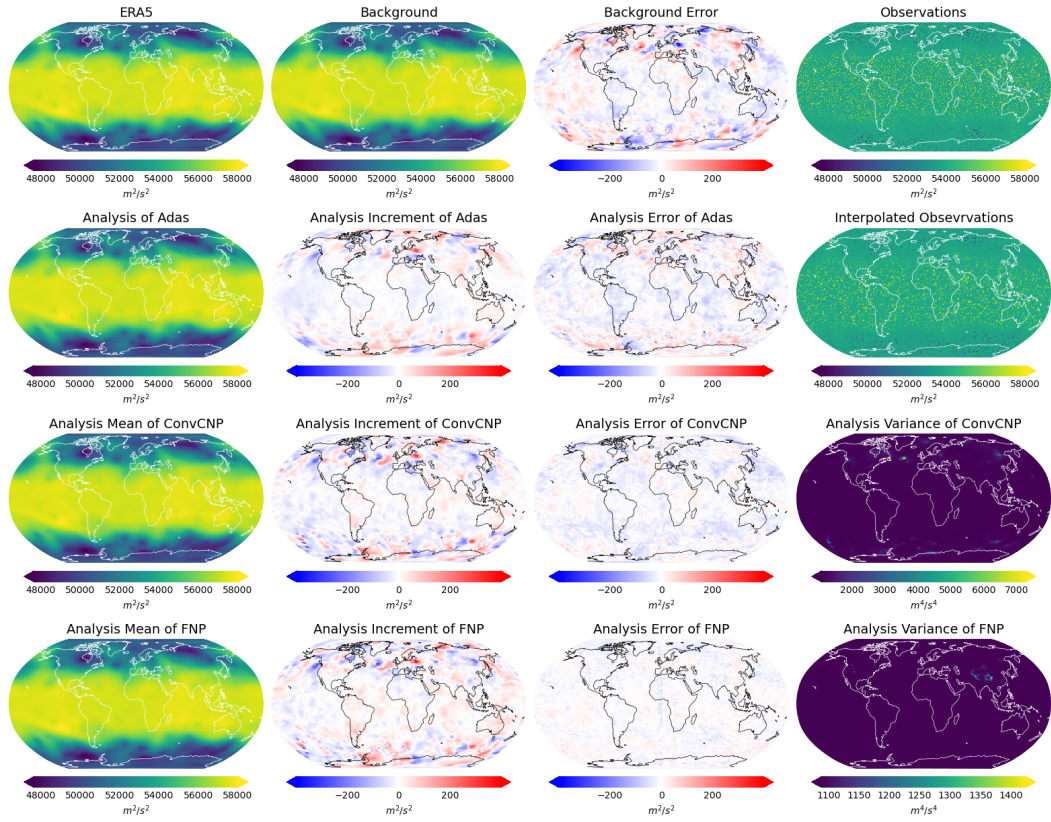
Figure 6: Visualization of assimilation results with 0.25° resolution for z500.

17
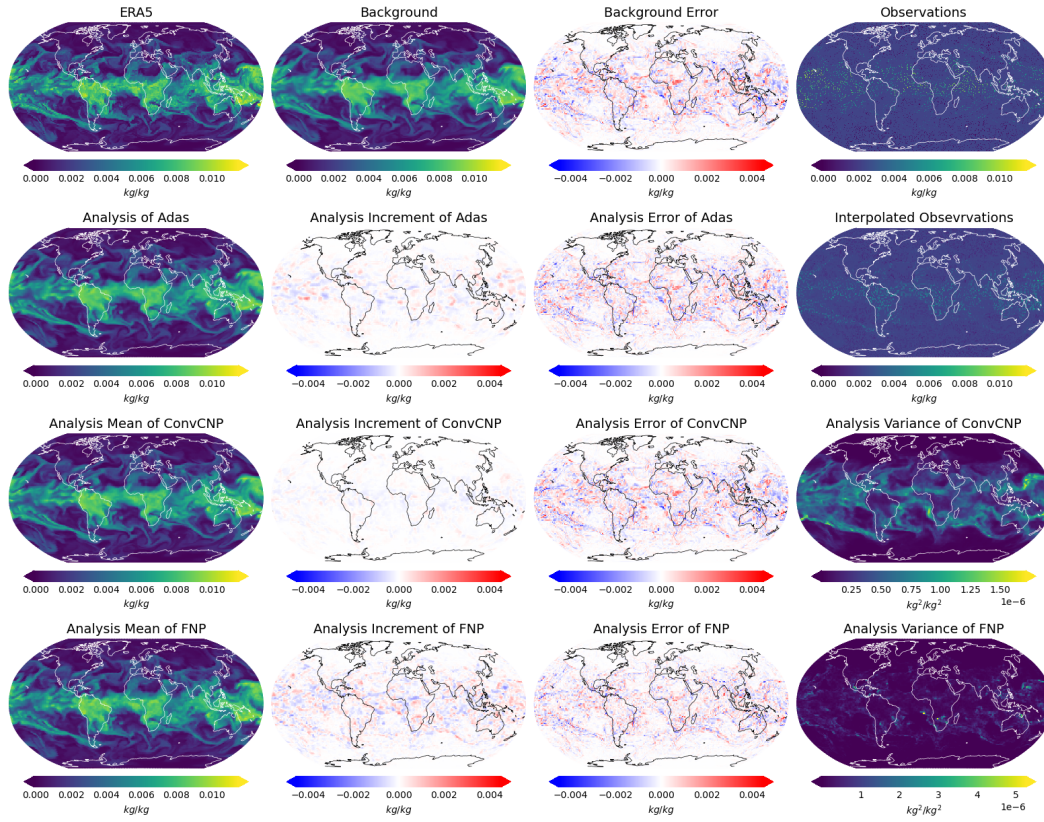
Figure 7: Visualization of assimilation results with 0.703125° resolution for q700.

18
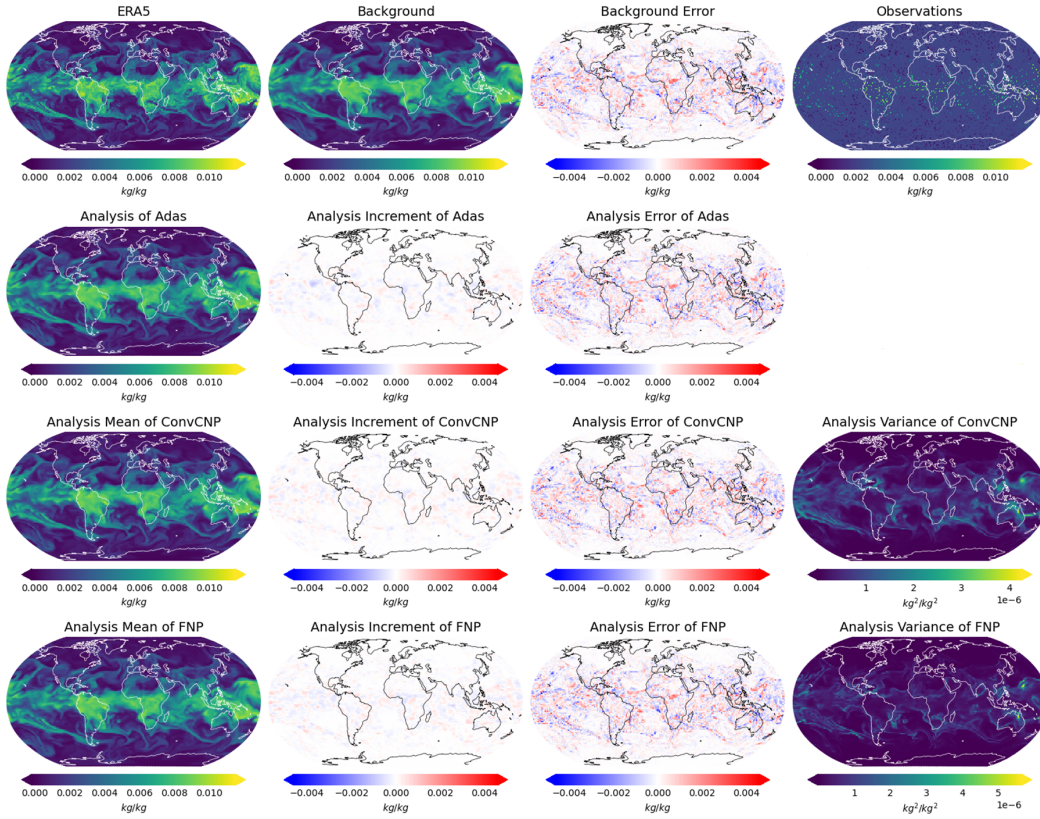
Figure 8: Visualization of assimilation results with 1.40625° resolution for q700.

19

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses the limitations of the work performed by the authors in Section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

20

Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper fully discloses all the information needed to reproduce the main experimental results in Section 4 and released code.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

21

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The paper provides open access to the code in Appendix **??**, and the data used in the paper is an open-source dataset.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The paper specifies all the training and test details necessary to understand the results in Section 4 and released code.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: The paper does not report error bars, like other related work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

22

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper provides sufficient information on the computer resources needed to reproduce the experiments in Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research conducted in the paper conforms with the NeurIPS Code of Ethics in every respect.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper discusses both potential positive societal impacts and negative societal impacts of the work performed in Section 5.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

23

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We cite the original papers or websites that produced the code or dataset, and the data used in the paper is an open-source dataset.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

24

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The code in the paper is well documented and the documentation is provided alongside the code.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

25

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

26