Probabilistic Weather Forecasting with Deterministic Guidance-based Diffusion Model

- Supplementary Material -

1 Introduction

In this Supplementary material, we provide the following details:

- The limitation of high-diversity probabilistic forecasting model.
- Detailed information about the PNW-typhoon dataset.
- Additional qualitative results.
- Architecture details of DGDM.

2 Ensemble using previous diffusion model

Not only DGDM, but also probabilistic models such as PreDiff [1] provide diverse results that can be used for ensemble methods. Fig. 1 shows the visualization comparison of ensemble results of probabilistic models and deterministic model [5] on the Moving MNIST dataset. The near future (step 1) is easy to predict from the given information, so all models predict accurately. However, since the far future (step 10) is harder to predict, the deterministic model produces blurry results, while the probabilistic model generates clear results, but does not predict the location and shape of the digits. In particular, the digit 8 in the last frame is hard to recognize because it has changed shape, and the digit 3 is misplaced a lot. Ensemble with these 10 samples, simply averaged from sampling these results, shows blurrier results than TAU. DGDM is able to use ensemble methods without degradation because it controls the possible future with deterministic branches.

3 Detailed in the PNW-Typhoon dataset

The PNW dataset serves as a valuable resource for forecasting extreme weather events, particularly in the Pacific region, through the analysis of satellite observations. This dataset provides insights into atmospheric conditions during typhoon occurrences using data obtained from the GK-2A satellite. All data preprocessing strictly adhered to the official GK2A user manual[1] to maintain the integrity of the GK2A physical value data. The spatial resolution of the dataset is 768 $\times$ 768, covering latitudes and longitudes from (90, 100) to (0,

[1] https://nmsc.kma.go.kr/enhome/html/base/cmm/selectPage.do?page=satellite.gk2a.intro

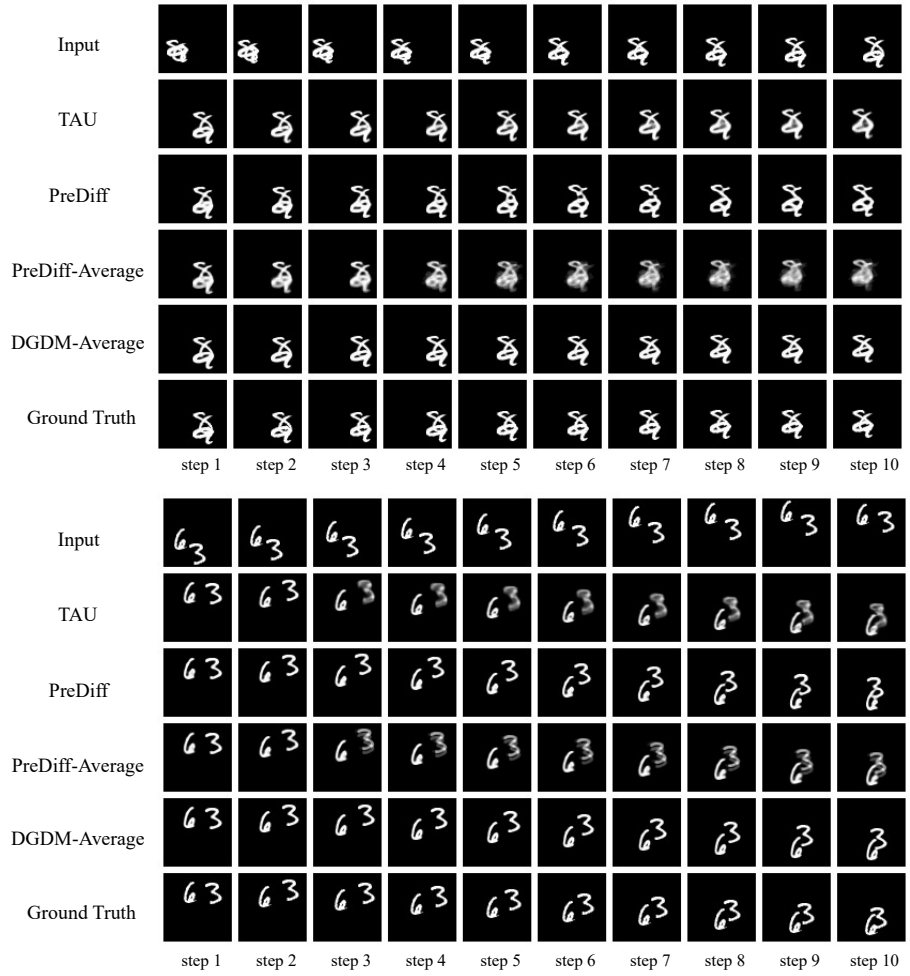

Fig. 1: Visualization comparison of ensemble results of probabilistic models and deterministic model results.

155). The PWN dataset span from January 2019 to October 2023, covering a diverse range of typhoon events. Throughout this period, we meticulously identified and selected dates corresponding to days when typhoons occurred. The distribution of typhoons across the years is as follows: 30 in 2019, 26 in 2020, 29 in 2021, 29 in 2022, and 14 in 2023. There are a total of 460 typhoon days, with an average of about 3 days per typhoon. Observations are recorded at hourly intervals, providing a high-resolution temporal view of atmospheric conditions during typhoon events. Consequently, resulting in a dataset comprising 11,040 unique data. This comprehensive coverage allows for robust analysis and forecast modeling in uncertain regional weather events. The dataset is structured

based on three main spectral channels: Infrared Red (IR), Short Wave (SW), and Water Vapor (WV). Each channel provides distinct information about the atmospheric composition and dynamics during typhoon occurrences.

4 Additional Qualitative Results

In this section, we provide additional qualitative results that could not be included in the main text. Fig. 2 shows the qualitative analysis of DGDM-Best (20 samples) from the PNW-Typhoon dataset. As seen in the figure, DGDM-Deterministic shows blurry results, but DGDM-Probability yields relatively clear results. These qualitative experimental results demonstrate that DGDM can be useful not only in predicting large-scale climate phenomena but also smaller-scale ones. Fig. 3 shows the qualitative experimental results of DGDM in the WeatherBench dataset. As can be seen in the figure, DGDM demonstrates its effectiveness in global climate modeling, such as with WeatherBench.

5 Detailed architecture of DGDM

We present the detailed architecture of the deterministic and probability branches of the DGDM. As shown in Tab. 1, the Deterministic branch adopts an encoder-translator-decoder structure, a leading approach in recent deterministic video frame prediction [3, 5]. Tab. 2 presents the architecture details of the probability branch. The probability branch of DGDM employs a 3D-UNet architecture, which is identical to that used in video diffusion models.

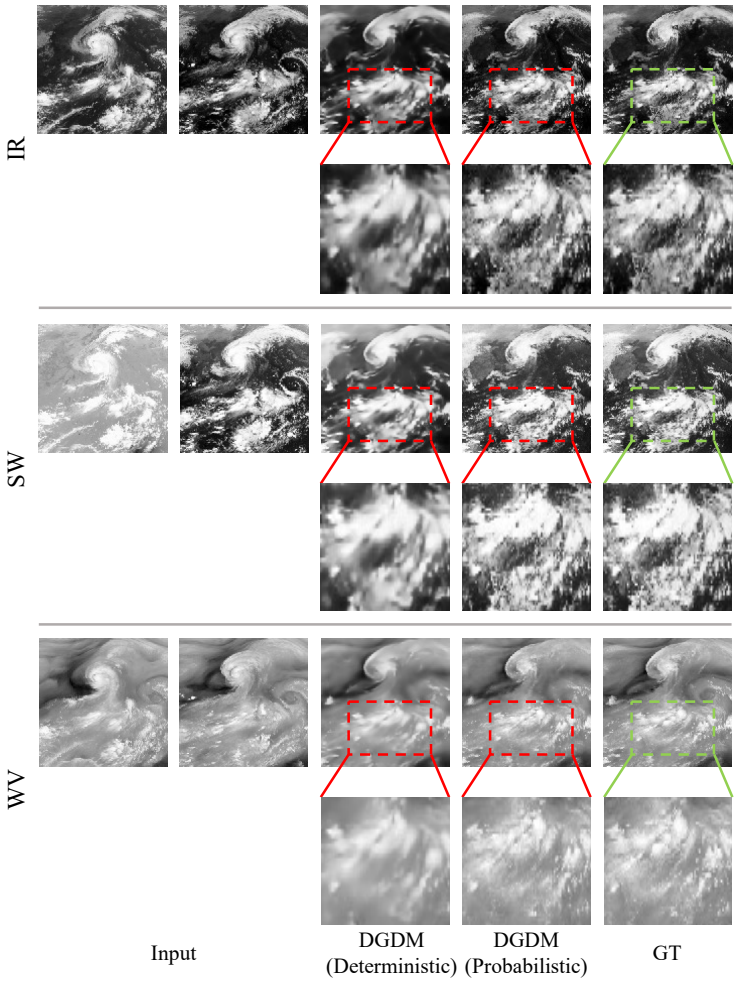

Fig. 2: Qualitative results from the PNW-Typhoon dataset with DGDM. The two samples on the left are the 0th and 9th samples, and a total of 10 frames from 0 to 9 are inputted. Also, the results all display the 19th frame, which is 10 hours later.

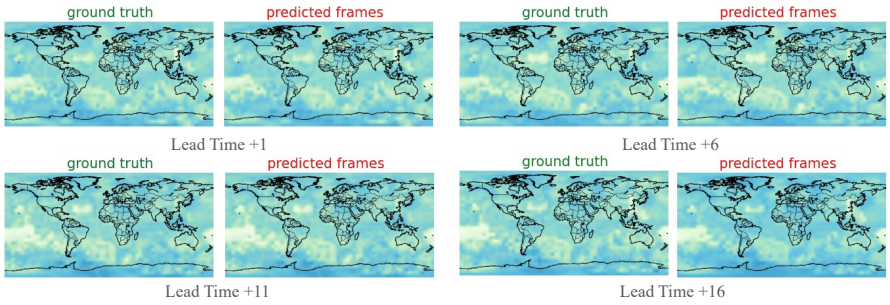

Fig. 3: Qualitative results of DGDM from the WeatherBench dataset.

Table 1: Detailed architecture of the deterministic branch. **Conv3x3** denotes a 2D convolution layer with a 3×3 kernel size, while **LayerNorm** refers to layer normalization. **SiLU** is the Sigmoid Linear Unit activation layer [2], and **Pixelshuffle** is a upsampling layer [4] that rearrange channel dimension into the spatial dimension. Throughout the encoder and decoder, the time dimension T is treated as a batch, while in the Translator and the final convolutional layer, it is reshaped into the channel dimension.

Block	Layer	Resolution	Channels
Encoder			
2D CNN	Conv3×3	$H \times W$	$C \to 64$
	LayerNorm		64
	SiLU		64
2D CNN	Conv3×3	$H \times W \to H/2 \times W/2$	64
	LayerNorm		64
	SiLU		64
2D CNN	Conv3×3	$H/2 \times W/2$	64
	LayerNorm		64
	SiLU		64
2D CNN	Conv3×3	$H/2 \times W/2 \to H/4 \times W/4$	64
	LayerNorm		64
	SiLU		64
Translator			$64 \to 64{\times}T$
ConvNext×8	Conv7×7	$H/4 \times W/4$	$64{\times}T$
	LayerNorm		$64{\times}T$
	Linear		$64{\times}T \to 256{\times}T$
	GELU		$256{\times}T$
	Linear		$256{\times}T \to 64{\times}T$
Decoder			$64{\times}T \to 64$
2D CNN	Conv3×3	$H/4 \times W/4$	$64 \to 256$
	LayerNorm		256
	SiLU		256
Upsample	Pixelshuffle	$H/4 \times W/4 \to H/2 \times W/2$	$256 \to 64$
2D CNN	Conv3×3	$H/2 \times W/2$	64
	LayerNorm		64
	SiLU		64
2D CNN	Conv3×3	$H/2 \times W/2$	$64 \to 256$
	LayerNorm		256
	SiLU		256
Upsample	Pixelshuffle	$H/2 \times W/2 \to H \times W$	$256 \to 64$
2D CNN	Conv3×3	$H \times W$	$64 \to 64$
	LayerNorm		64
	SiLU		64
Readout	Reshape		$64 \to 64{\times}T$
	Conv3×3	$H \times W$	$64{\times}T \to C \times T$

Table 2: Detailed architecture of the probabilistic branch. **Conv7×7**, **Conv3×3**, and **Conv1×1** are 3D convolutional layers with kernel sizes of 1×7×7, 1×3×3, and 1×1×1, respectively. **GroupNorm8** is the Group Normalization layer with 8 groups. **Time MLP** is the block used for embeded the denoising step t. **SiLU** and **GeLU** are the Sigmoid Linear Unit and Gaussian Error Linear Unit activation layers [2], respectively. **ResNetBlock** consists of a sequence with a Linear layer addressing the denoising step t followed by a SiLU activation, two blocks of Conv3×3, GroupNorm8, and SiLU activation. **SpAttention** is a spatial attention layer that considers only the spatial dimension with deterministic features. **TeAttention** is a self-attention layer that focuses on the temporal dimension of sequential data. Downsampling is performed by **Conv4×4**, a 3D convolutional layer with a stride of 2 and a 1×4×4 kernel size and upsampling is performed through **Deconv4×4**, a 3D transposed convolutional layer with the same stride and kernel size. **Up block** receives concatenated inputs from the features of previous block and down block, which have the same resolution.

Block	Layer	Resolution	Channels
Initial Block	Conv7×7	$H \times W$	$C \rightarrow 64$
	TeAttention		64
Time MLP	Linear		$64 \rightarrow 256$
	GELU		256
	Linear		256
Cond Block	Conv3×3	$H/4 \times W/4$	64
	GroupNorm8		64
	SiLU		64
	Conv3×3	$H/4 \times W/4$	64
	GroupNorm8		64
	SiLU		64
Down Block	ResBlock	$H \times W$	64
	ResBlock	$H \times W$	64
	SpAttention		64
	TeAttention		64
	Conv4×4	$H \times W \rightarrow H/2 \times W/2$	64
Down Block	ResBlock	$H/2 \times W/2$	$64 \rightarrow 128$
	ResBlock	$H/2 \times W/2$	128
	SpAttention		128
	TeAttention		128
	Conv4×4	$H/2 \times W/2 \rightarrow H/4 \times W/4$	128
Down Block	ResBlock	$H/4 \times W/4$	$128 \rightarrow 256$
	ResBlock	$H/4 \times W/4$	256
	SpAttention		256
	TeAttention		256
	Conv4×4	$H/4 \times W/4 \rightarrow H/8 \times W/8$	256
Down Block	ResBlock	$H/8 \times W/8$	$256 \rightarrow 512$
	ResBlock	$H/8 \times W/8$	512
	SpAttention		512
	TeAttention		512
Mid Block	ResBlock	$H/8 \times W/8$	512
	SpAttention		512
	TeAttention		512
	ResBlock	$H/8 \times W/8$	512
Up Block	ResBlock	$H/8 \times W/8$	$1024 \rightarrow 256$
	ResBlock	$H/8 \times W/8$	256
	SpAttention		256
	TeAttention		256
	Deconv4×4	$H/8 \times W/8 \rightarrow H/4 \times W/4$	256
Up Block	ResBlock	$H/4 \times W/4$	$512 \rightarrow 128$
	ResBlock	$H/4 \times W/4$	128
	SpAttention		128
	TeAttention		128
	Deconv4×4	$H/4 \times W/4 \rightarrow H/2 \times W/2$	128
Up Block	ResBlock	$H/2 \times W/2$	$256 \rightarrow 64$
	ResBlock	$H/2 \times W/2$	64
	PoAttention		64
	TeAttention		64
	Deconv4×4	$H/2 \times W/2 \rightarrow H \times W$	64
Up Block	ResBlock	$H \times W$	$128 \rightarrow 64$
	ResBlock	$H \times W$	64
	SpAttention		64
	TeAttention		64
Out Block	ResBlock	$H \times W$	$192 \rightarrow 64$
	Conv1×1	$H \times W$	$64 \rightarrow C$

References

1. Gao, Z., Shi, X., Han, B., Wang, H., Jin, X., Maddix, D., Zhu, Y., Li, M., Wang, Y.: Prediff: Precipitation nowcasting with latent diffusion models. arXiv preprint arXiv:2307.10422 (2023)
2. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
3. Seo, M., Lee, H., Kim, D., Seo, J.: Implicit stacked autoregressive model for video prediction. arXiv preprint arXiv:2303.07849 (2023)
4. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR. pp. 1874–1883 (2016)
5. Tan, C., Gao, Z., Wu, L., Xu, Y., Xia, J., Li, S., Li, S.Z.: Temporal attention unit: Towards efficient spatiotemporal predictive learning. In: CVPR. pp. 18770–18782 (2023)