IMPERIAL

Climate Change AI

ICLR

Improving Tropical Cyclone Forecasting With Video Diffusion Models

Zhibo Ren[1], Pritthijit Nath[2] and Pancham Shukla[3]

Tackling Climate Change with Machine Learning Workshop, ICLR 2025

Motivation

Accurate forecasting of tropical cyclones (TC) is critical to disaster preparedness and mitigation. Recent deep learning methods, although promising, typically forecast frame-by-frame, missing crucial temporal dynamics. This work introduces a novel video diffusion model (VDM) approach for TC forecasting, explicitly modeling temporal dependencies and improving prediction accuracy and coherence.

Inspired from previous methods, this study aims to make the following contributions:

1. Application of video diffusion models (VDM) tailored specifically for TC forecasting.
2. A novel two-stage training scheme improving spatial accuracy and temporal coherence, even with limited data.
3. Extending reliable forecasting lead times from 36 to 50 hours.
4. Using Fréchet Video Distance (FVD) as a robust metric for evaluating temporal consistency.

Data

Data Acquisition

1. **Satellite Data:** Infrared (IR) 10.8µm satellite images for a total of 51 significant tropical cyclones with major landfall impacts, covering six major oceanic basins from January 2019 to March 2023.
2. **Atmospheric Data:** Hourly ERA5 reanalysis data for four key atmospheric variables - 10m zonal wind (u10), 10m meridional wind (v10), surface pressure (sp) and 2m air temperature (t2m) acquired from the Copernicus Climate Data Store, covering each cyclonic period from formation to dissipation.

Data Processing

1. **Data Cleaning and Alignment:** Removed corrupted data points and aligned satellite and ERA5 data both temporally and spatially.
2. **Normalization and Augmentation:** Variables standardised to zero mean and unit variance. Data augmentation applied to enhance model robustness.
3. **Training-Test Split:** Dataset divided into 1,092 sequences for training and 335 for testing.

Methodology

Video Diffusion Models

Our approach applies a 3D UNet-based video diffusion model architecture, explicitly modelling temporal dependencies through additional temporal convolutional and attention layers. Our VDM:

- Generates 10 consecutive frames simultaneously.
- Conditions predictions on initial satellite imagery and ERA5 meteorological data.
- Uses classifier-free guidance and dynamic thresholding techniques for stable generation.

Two-Stage Training Strategy

1. **Single-Frame Training (first stage):** Establish spatial features, stabilise model parameters.
2. **Multi-Frame Training (second stage):** Fine-tune model on sequences, explicitly capturing temporal dynamics.

In low-data regime, this strategy is shown to improve FID scores significantly (from 1.26 to 0.49), ensuring higher-quality frame-level predictions even with limited training samples.

Evaluation Metrics

- **Single-frame quality:** Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Fréchet Inception Distance (FID).
- **Temporal coherence:** Fréchet Video Distance (FVD).

Affiliations

[1] Department of Computing, Imperial College London

[2] Department of Applied Mathematics and Theoretical Physics, University of Cambridge

Results

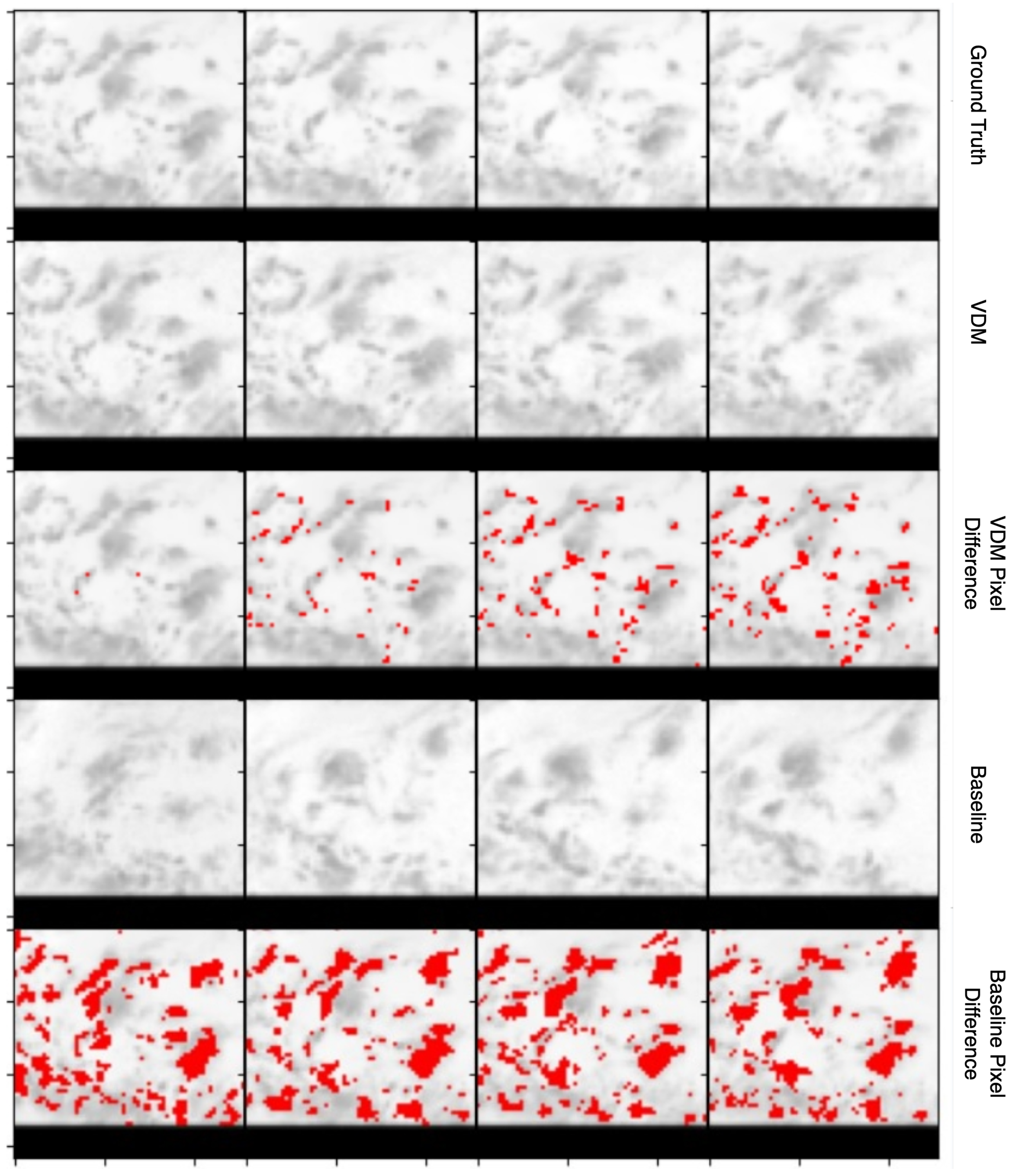

Figure 2: Visual comparison between our model (VDM) and the baseline - Nath et al. [1]. Row 1: ground truth, Row 2: VDM output, Row 3: pixel-wise difference between VDM and ground truth (red), Row 4: baseline output and Row 5: pixel-wise difference between baseline and ground truth (red).

Table 1: Performance comparison (over several evaluation metrics) between VDM and the baseline

Metric	Baseline	VDM	Gain
MAE ↓	0.221	**0.178**	19.3%
FID ↓	**0.229**	0.234	-2.4%
FVD ↓	445.8	**242.4**	45.6%
PSNR ↑	22.49	**26.13**	16.2%
SSIM ↑	0.524	**0.712**	36.1%

References

[1] Nath P, Shukla P, Wang S, Quilodrán-Casas C. Forecasting Tropical Cyclones with Cascaded Diffusion Models. arXiv; 2024. ArXiv:2310.01690. Available from: http://arxiv.org/abs/2310.01690.

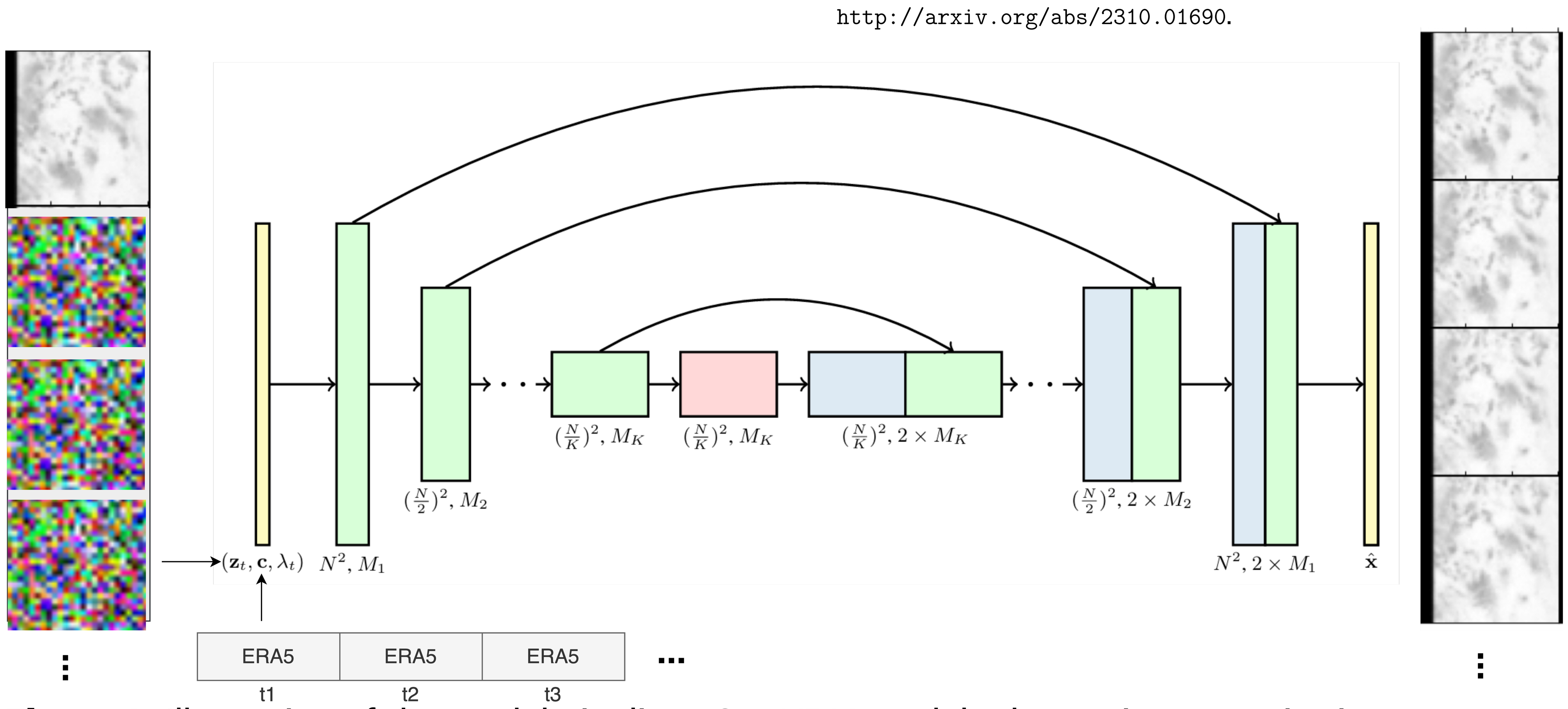

Figure 1: Illustration of the model pipeline. Our VDM model takes as input a noisy image $\mathbf{z}_t$, conditioning variables $\mathbf{c}$, and a timestep embedding λ_t, and progressively denoises the sample using a U-Net-based architecture with skip connections. ERA5 data from multiple timesteps (t_1, t_2, t_3) is used as conditioning information. Output $\hat{\mathbf{x}}$ represents the denoised prediction.