A 3D Attention U-Net GAN for Global Precipitation Nowcasting

Dang Do[1,3], Shigenori Otsuka[1,2]

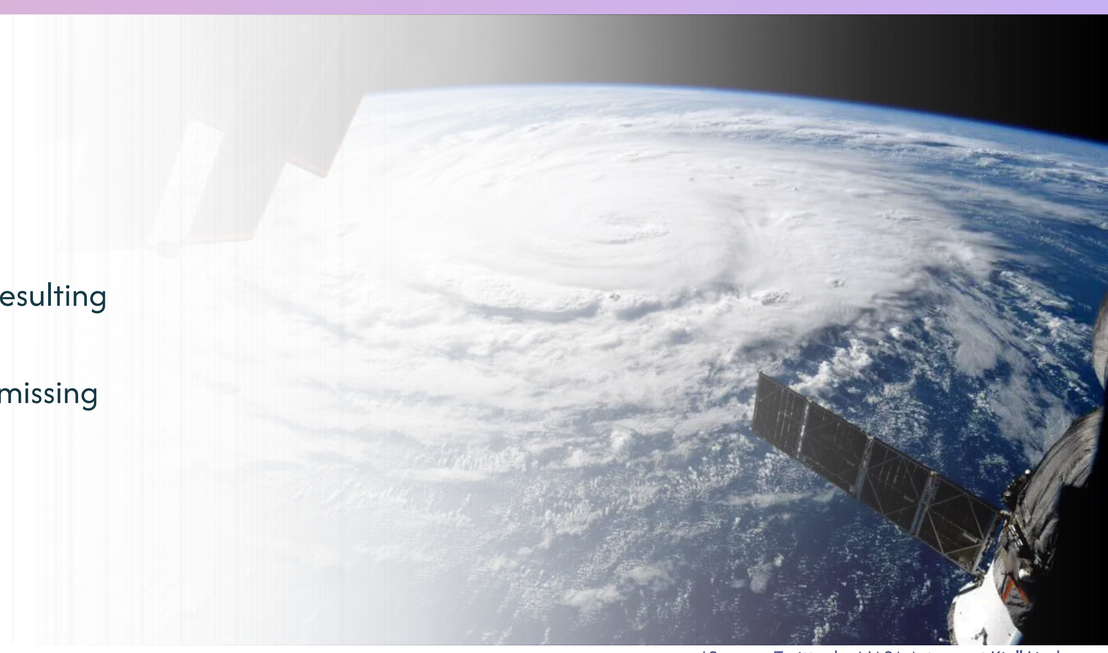
[1]RIKEN Center for Computational Science, [2]RIKEN iTHEMS,

[3]University of Yamanashi

RIKEN Center for Computational Science

Data Assimilation Team

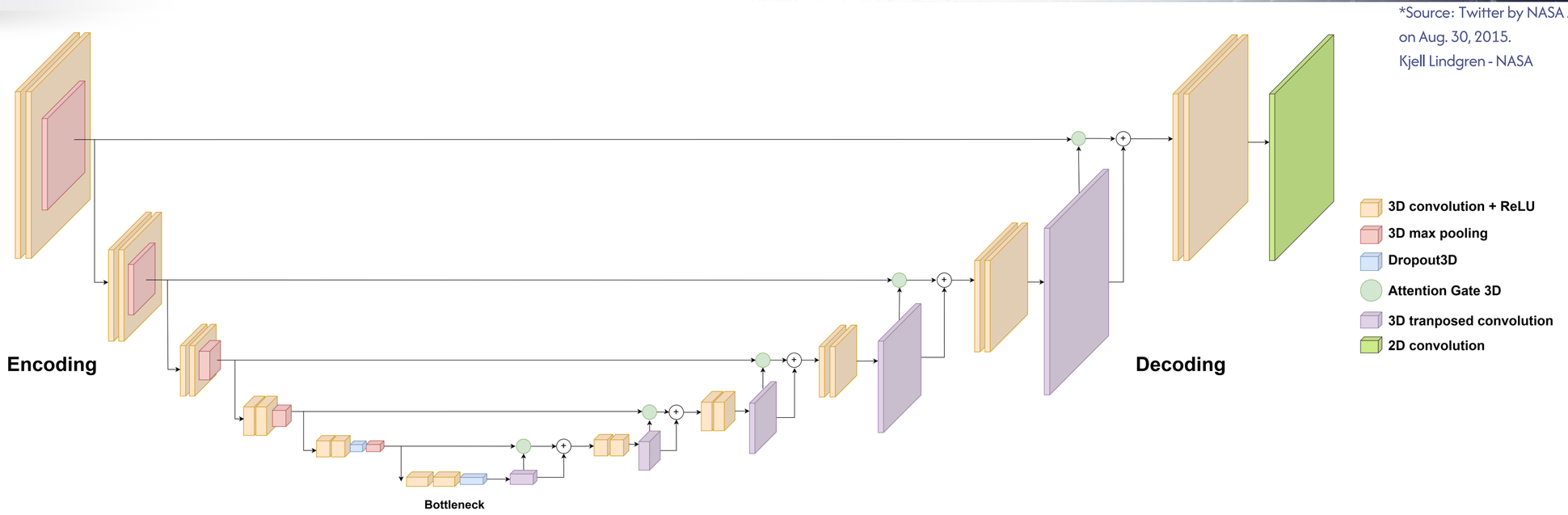
INTRODUCTION



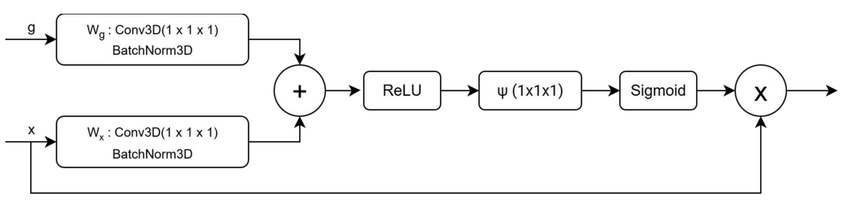
*Source: Twitter by NASA Astronaut Kjell Lindgren on Aug. 30, 2015.
Kjell Lindgren - NASA

- **Goal**: Accurate, high-resolution precipitation nowcasting is critical for disaster management.
- **Current Challenges**:
  - **Blurry Outputs**: Traditional models with L2 (MSE) loss fail to capture high-intensity rain, resulting in blurry outputs.
  - **2D Limitations**: Previous models (e.g., SmaAt-UNet) treat time as a 2D channel, thereby missing complex 3D weather dynamics.
  - **Computational Cost**: Recurrent models (e.g., ConvLSTM) are computationally intensive.

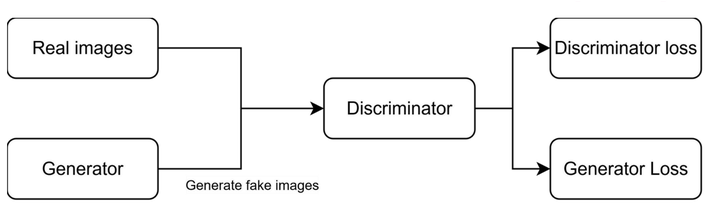
METHODOLOGY



Attention Gate 3D



Generative Adversarial Network (GAN)



Loss function

L1 Loss (MAE) + Generator Loss + Gradient Difference Loss (GDL)

Experimental Setup:

- Dataset: Global Satellite Mapping of Precipitation (GSMaP).
- Resolution: Half of the original (512 x 1536).
- Metric: Critical Success Index (CSI).

Training Infrastructure:

- Compute Nodes: 8x NVIDIA A100 80GB GPUs.
- Host CPU: AMD EPYC™ 9684X (96 Cores).
- Memory: 2TB RAM.
- Strategy: Distributed Data Parallel (DDP) for multi-GPU scaling.

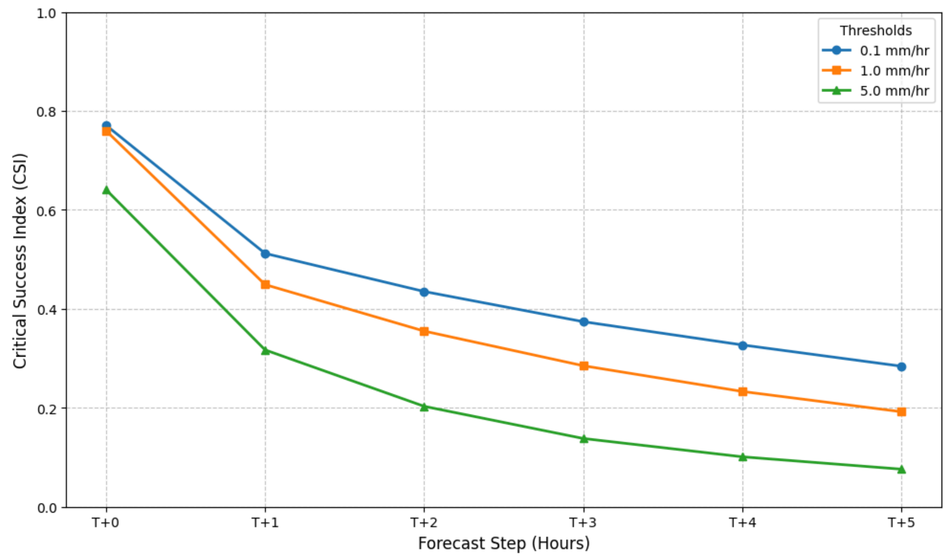
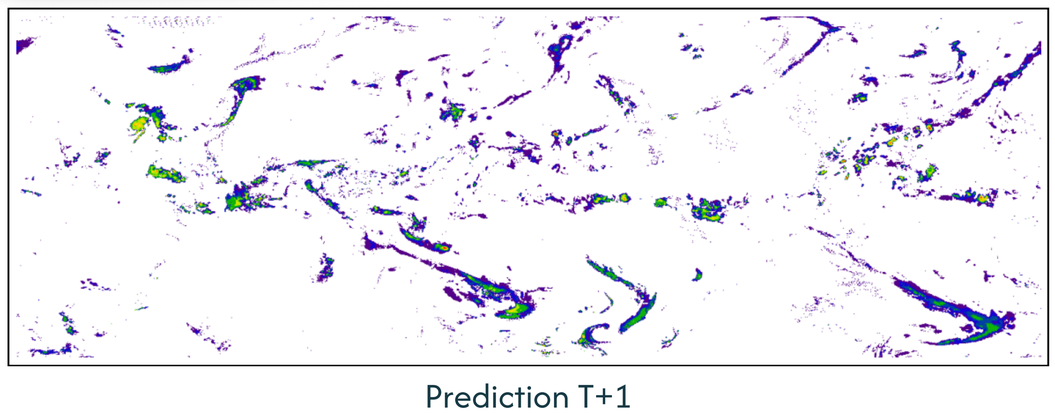
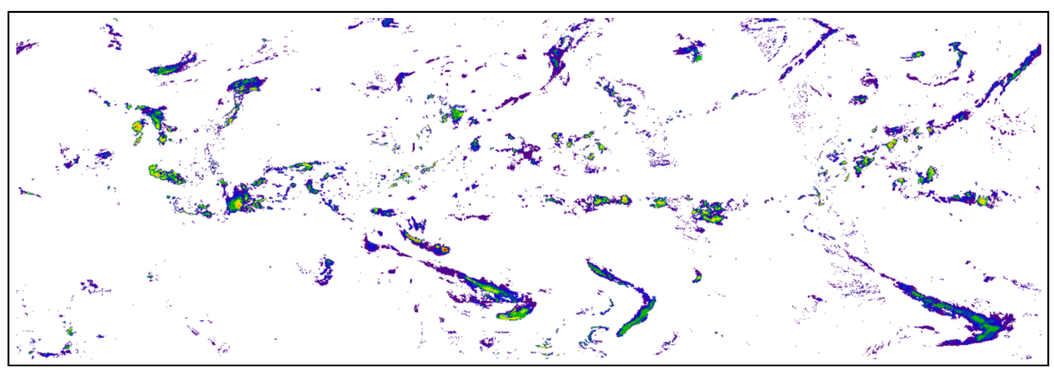


Table 1: 5-Step Validation CSI for January 2024 (higher is better)

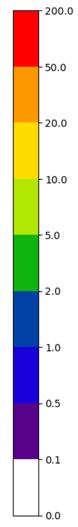
RESULTS



Prediction T+1



Ground truth T+1



CONCLUSION

The model delivers sharp, high-resolution forecasts for disaster response without the latency penalties of sequential processing

FUTURE WORK

Leverage Fugaku's massive node parallelism to train on full-resolution global data

HPC CHALLENGES

- I/O Bottleneck: We identified that the data loader could not keep up with the model's high throughput, leading to GPU starvation.
- Memory Optimization: Standard DDP replication consumed excessive VRAM. → Solution: Transitioning to Fully Sharded Data Parallel (FSDP2).