DINOv3 as a Frozen Encoder for CRPS-Oriented Probabilistic Rainfall Nowcasting

Weather4Cast 2025 Competition

Luciano Araujo Dourado Filho[1,2] Almir Moreira da Silva Neto[1] **Anthony Miyaguchi**[3] Rodrigo Pereira David[2] Rodrigo Tripodi Calumby[1] Lukáš Picek[4]

[1]Advanced Data Analysis and Management, University of Feira de Santana
[2]National Institute of Metrology, Quality and Technology
[3]Georgia Institute of Technology
[4]University of West Bohemia in Pilsen

NeurIPS 2025

Outline

The Weather4Cast 2025 Challenge

The competition frames nowcasting as a multi-modal super-resolution task:

Inputs (Satellite):

- 11 bands (VIS, IR, WV).
- Coarse resolution ($\approx 6 \times 6$ pixels).
- 4 context frames (past hour).

Targets (Radar):

- Ground-truth rain rate.
- High resolution (2km, 32×32 pixels).
- **Task:** Predict cumulative rainfall over 4 hours.

The Core Difficulty

Bridging the **super-resolution gap** while handling **spatial distribution shifts** across different European regions.

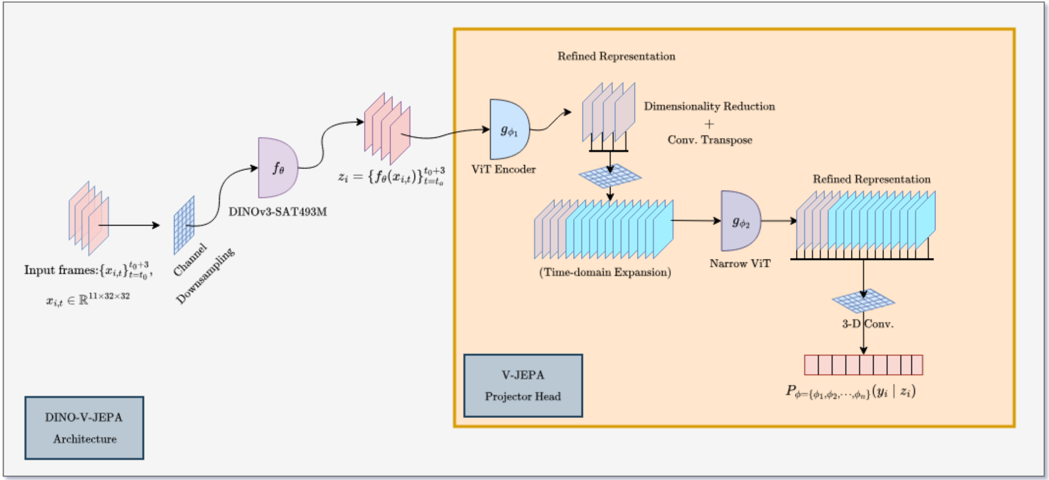
Problem Statement

Current limitations in Deep Learning Nowcasting:

1. **Compute Costs:** End-to-end training of large ViTs is prohibitive.
2. **Data Hunger:** Training large backbones requires massive datasets to generalize.
3. **Calibration:** Models often minimize RMSE but fail to capture tails (extreme events).

Our Hypothesis: *A frozen, pre-trained "World Model" (Satellite Encoder) can be reused with a lightweight projector to solve these issues efficiently.*

Proposed Architecture: DINO-V-JEPA



Deep Dive: The Projector

How we map frozen tokens to rainfall distributions:

- **Modality Gap:** A trainable downsampling layer adapts 11-channel satellite inputs to the RGB-trained DINOv3.
- **Latent Processing:**
  - **Encoder (V-JEPA ViT):** Maps DINOv3 tokens ($4 \times 196 \times 1024$) to a unified latent space.
  - **Decoder (V-JEPA ViT):** Compresses dim to 384, performs **Time-Expansion** (interpolating 4 input frames to 16 output slots) and input into another ViT for **representation refinement** before predictions.
- **Output Head:** 3D Convolution collapses features into K discrete rainfall bins.

Comparison Baselines: 3D-UNet

To prove the efficacy of the frozen encoder, we trained full spatiotemporal baselines:

3D-UNet Variants

- **Backbone:** Standard 3D-UNet with (2+1)D convolutions.
- **Head A (Aggregate):** Predicts discrete RPS directly (similar to our DINO approach).
- **Head B (Per-Pixel):** A **Gamma-Hurdle** model.
  - *Hurdle:* Binary classification (Rain / No-Rain) using Focal Loss.
  - *Gamma:* Regresses intensity (α, β) for rainy pixels.

Training Objective: Discrete CRPS

We optimize the **Rank Probability Score (RPS)** to align directly with the competition metric.

$$\mathcal{L}_{RPS} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \left(F_k(P_\phi(y_i|z_i)) - \mathbb{1}(k \geq y_i)\right)^2 \qquad (1)$$

- P_ϕ: Predicted categorical distribution over K bins.
- F_k: Empirical Cumulative Distribution Function (eCDF).
- $\mathbb{1}$: Indicator function (1 if bin k exceeds truth y_i).

Note: This bypasses assumptions about parametric distributions (e.g., Gaussian/Gamma).

Dataset Characteristics

Table: Summary Statistics of the Validation Data Split

Statistic	**Mean**	**Min**	**p50**	**p95**	**Max**
Masking (%)	0.0000	0.0000	0.0000	0.0000	0.0000
Zero-Inflation (%)	88.8463	0.0916	99.8718	100.0000	100.0000
Aggregate Target (mm)	0.3434	0.0000	0.0012	2.4084	4.1663
Non-Zero Mean (mm)	0.3073	0.0000	0.2354	0.9859	1.6961
Non-Zero Max (mm)	1.7536	0.0000	0.2950	7.9700	10.5800

Dataset Characteristics

Data: 2 years of data, OPERA radar ground truth.

The Zero-Inflation Challenge

- **Sparsity:** 88% of validation samples have **zero** rain.
- **Pixel Sparsity:** Median sample has $>$ 99.9% zero-valued pixels.
- **Implication:** Models can achieve low error by predicting "zero" everywhere, but fail the CRPS metric which penalizes uncalibrated uncertainty.

Data Pipeline: Super-Resolution

Challenge: Radar targets are $6\times$ higher resolution than satellite inputs.

- **Synchronized Cropping:** We implement `BaseSuperResCrop` to calculate Low-Res (Satellite) coordinates based on High-Res (Radar) crop targets.
- **Mask Handling:** Geometric augmentations (H/V Flips) are synchronized across the entire stack:
  - Input (11-channel Satellite)
  - Target (Rain Rate)
  - **Metadata Mask** (Valid/Invalid pixels)

Addressing Sparsity: Rain-Biased Sampling

The Sparsity Problem

Random cropping leads to empty targets and zero-inflation.

Solution: `RandomSuperResCrop`

- **Logic:** With probability p_{rain}, the sampler scans for pixels > 0.1 mm/hr.
- **Action:** Forces the crop to center on these "active" weather events.
- **Result:** drastically up-samples convective events during training while preserving valid spatial context.

Main Results: Leaderboard

Backbone	Loss Objective	Bin Max (mm/hr)	Bins (K)	CRPS $\downarrow$
DINO-VJEPA	**RPS (Aggregate)**	**128**	**25,601**	**3.5102**
UNET3D (v4)	Gamma-Hurdle (Per-Pixel)	512	129	4.7637
DINO-VJEPA	RPS (Aggregate)	128	129	5.5894
UNET3D (v10)	Gamma-Hurdle (Per-Pixel)	64	6,401	6.3634
UNET3D (v10)	RPS (Aggregate)	64	6,401	7.0249
UNET3D (v10)	RPS + Gamma-Hurdle	64	6,401	7.1057

Table: Test set performance. Our frozen approach outperforms trainable baselines by $\approx 26\%$.

Key Finding: Binning Granularity

A surprising finding was the impact of the number of bins (K) on the DINO-V-JEPA performance.

Coarse Bins ($K = 129$)

- $R_{max} = 128$ mm/hr.
- CRPS stalls at ≈ 5.6.
- *Issue:* Discretization bias masks small, meaningful changes in accumulation.

Fine Bins ($K = 25,601$)

- Same R_{max}.
- CRPS reaches **3.51**.
- *Benefit:* Sharpens the learning signal; the RPS gradient builds on finer residuals.

Training Dynamics

Trade-off: Fine-grained bins improve convergence but introduce instability.

Epoch	**Coarse** $(K = 129)$	**Fine** $(K = 25,601)$
1	5.6870	4.9539
3	6.2564	4.5317
5	6.2577	3.6404
6	**6.2810**	**3.5102**
7	5.8653	6.6455 (Instability)

Result: Early stopping is mandatory when using fine discretizations.

Training Configuration (DINO-V-JEPA)

Parameter	Value
Epochs	30
Learning Rate	1.0e-5, 5.0e-4 (init, effective)
Weight Decay	0.2, 0.5 (init, end)
Batch Size	4096 (via gradient accumulation)
Rain Sampling Probability	0.75
Rain Sampling Threshold	0.3

DINO-V-JEPA: Training configuration parameters.

Results Analysis: Baselines

Why did the 3D-UNet underperform?

- **Gamma-Hurdle Head:** Performed decently (CRPS 4.76) by explicitly modeling zeros, but struggled with the aggregate tail.
- **Multi-Task Learning:** Attempting to combine pixel-level Gamma loss with aggregate RPS failed.
- **Reason:** Numerical instability. Fitting Gamma parameters to zero-inflated data caused exploding gradients, and the Gamma CDF is numerically fragile in PyTorch.

Discussion: Frozen World Models

Why did the frozen encoder win?

1. **Sample Efficiency:** The DINOv3 backbone already "knows" spatial features (clouds, textures). The projector only needs to learn **temporal dynamics**.
2. **Focus on Calibration:** The lightweight head allocates capacity to getting the distribution (eCDF) right, rather than relearning feature extraction.

Limitations & Future Work

Limitations:

- **Instability:** Fine-grained binning requires careful monitoring (early stopping).
- **Single Task:** Only tested on 4-hour accumulation, not per-frame video prediction.

Future Directions:

- Explore **Parameter-Efficient Fine-Tuning (PEFT)** (e.g., LoRA) instead of fully frozen or fully trained.
- Investigate better numerical approximations for Multi-Task Probabilistic losses.

Conclusion

1. **Foundation Models Work:** Freezing DINOv3 + V-JEPA projector is a competitive strategy for rainfall nowcasting.
2. **Resolution Matters:** High-fidelity discretization ($K \approx 25k$) is critical for optimizing CRPS.
3. **Simplicity Wins:** Direct alignment with the evaluation metric (RPS) beat complex multi-task regularizers.

Code available at:

`https://github.com/acmiyaguchi/weather4cast-2025`

`https://github.com/FalsoMoralista/Weather-4-Cast`

References I

- Siméoni et al. DINOv3. arXiv:2508.10104, 2025.
- Assran et al. V-JEPA 2. arXiv:2506.09985, 2025.
- Zamo, M. & Naveau, P. Estimation of CRPS with limited information. Math. Geosciences, 2018.
- Gruca et al. Weather4cast 2024. NeurIPS Benchmark, 2024.
- Kendall et al. Multi-Task Learning Using Uncertainty. CVPR 2018.