SRViT: Vision Transformers for Estimating Radar Reflectivity from Satellite Observations at Scale

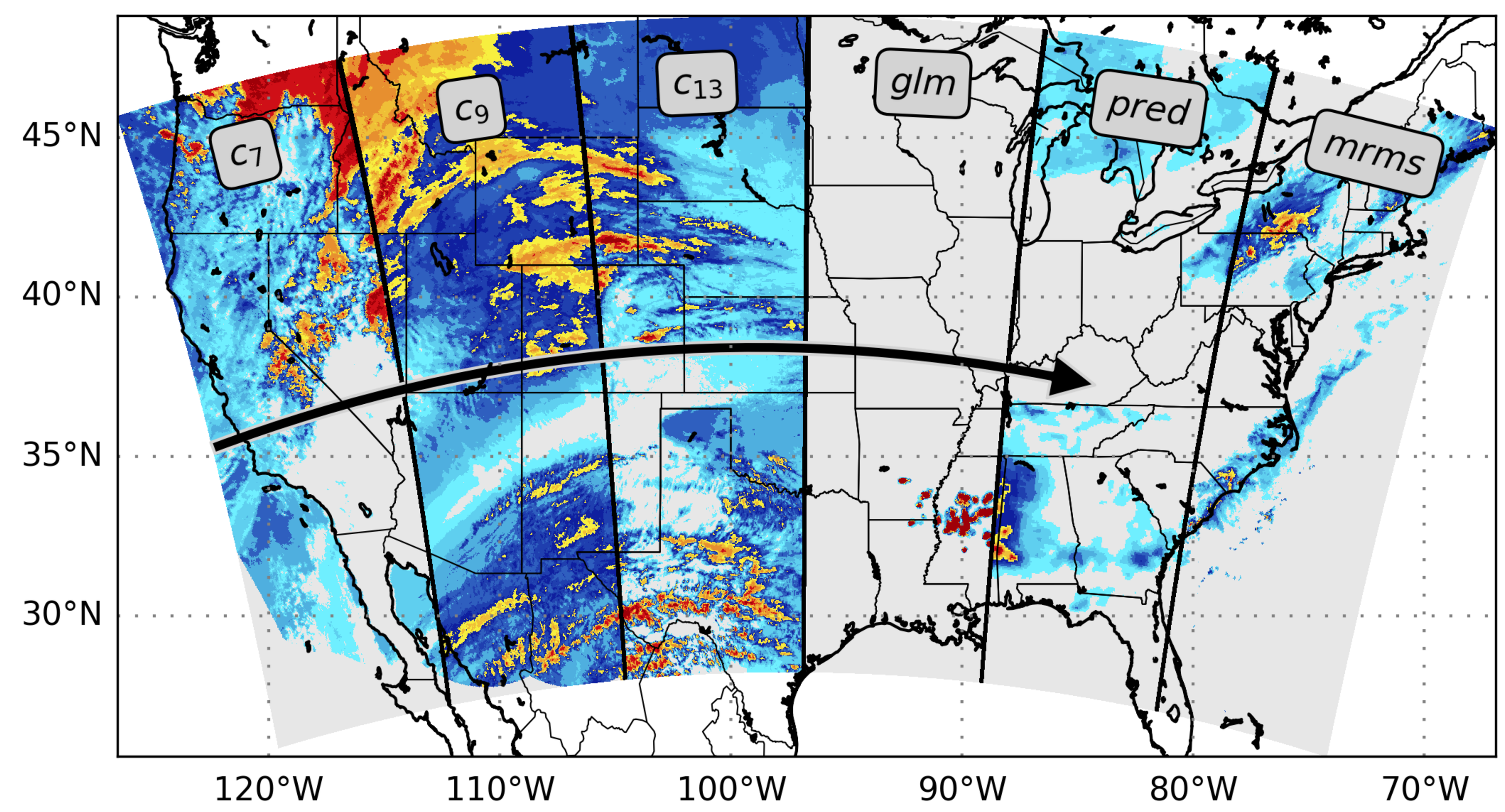
Jason Stock [1] Kyle Hilburn [2] Imme Ebert-Uphoff [2,3] Charles Anderson [1]

[1] CS, Colorado State University [2] Cooperative Institute for Research in the Atmosphere [3] ECE, Colorado State University

Connecting Models and Observations

Overview

A transformer to estimate high-resolution (3 km) radar reflectivity fields from geostationary satellite imagery, accurately capturing the complex atmospheric phenomena both locally and across larger domains.



Introduction

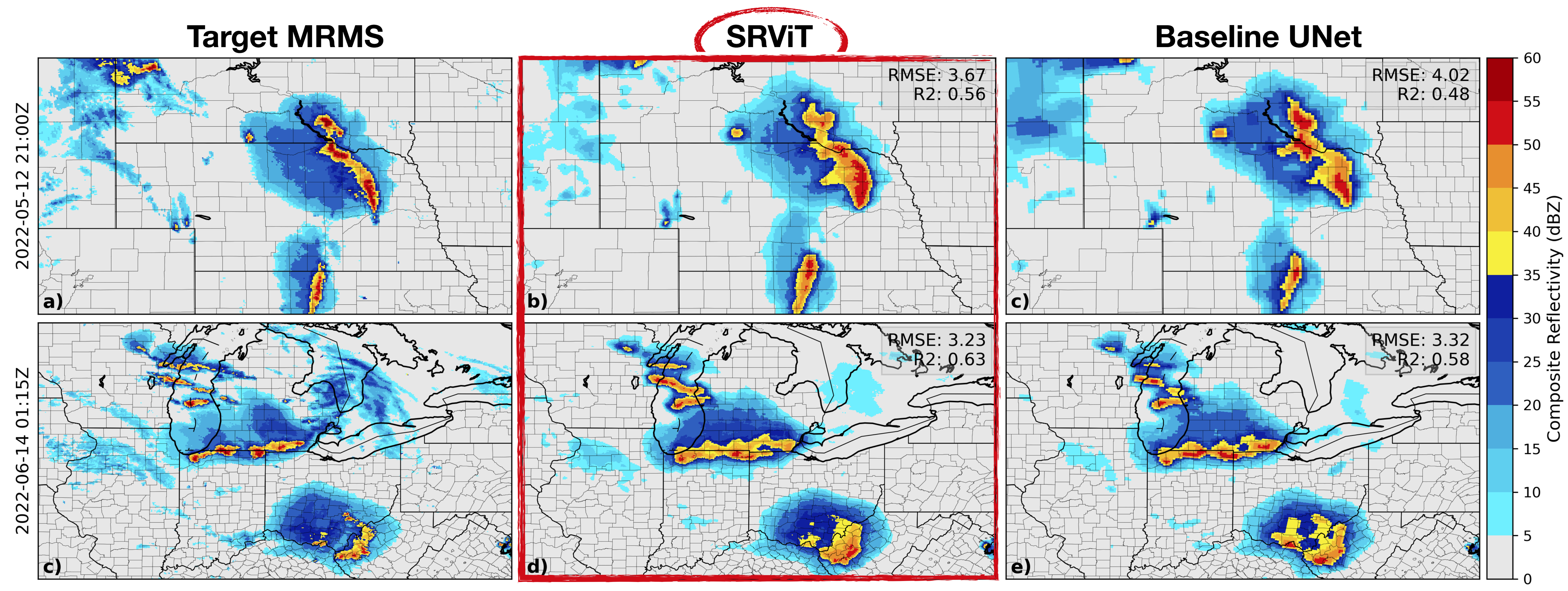
Motivation: (a) radar is useful to issue warnings and integrate into numerical weather prediction models; but (b) is limited to sparse ground stations; and (c) convolutional approaches can have narrow receptive fields and blurry output.

Question: will a deterministic, transformer-based network that contextualizes synoptic observations over the United States outperform a convolutional model?

Dataset Details

Input Data: GOES-16 Advanced Baseline Imager (ABI) (Level-L1b; infrared channels 7 / 9 / 13) and Geostationary Lightning Mapper (GLM) observations.
Target Data: Multi-Radar Multi-Sensor (MRMS) composite reflectivity.

Spatial Coverage: follows a 3 km HRRR mass grid, 768 × 1536-pixel images.
Temporal Range: restricted to the warm season (i.e., Apr-Sep) for years 2018-2022, sampled on 6h periods with a 15 min refresh (96 samples/day).



Methodology (🛰 → 📡)

SRViT: transformer for image-to-image translation, reconstructing patches with $\phi : \mathbf{X}^{l} \rightarrow \mathbf{X}^{l+1} \in \mathbb{R}^{n \times d}$ for $l = 1 \ldots L$ followed by a linear decoder and CNN.

Weighted loss: balance the rare, high radar reflectivity values with the small, common values with $\mathscr{L}_e = \frac{1}{m}\sum_{i=1}^{m} \exp(w_0 t_i^{w_1}) \cdot (y_i - t_i)^2$, trained end-to-end.

Comparisons: evaluate against a fully-convolutional network and Base-ViT.

Guiding Domain Experts

Token (Re)Distribution: explains the redistribution of input tokens $\mathbf{X} \in \mathbb{R}^{n \times d}$, as a result of self-attention, to the value of an intermediate token $\mathbf{z}_1, \ldots, \mathbf{z}_n \in \mathbb{R}^{d}$

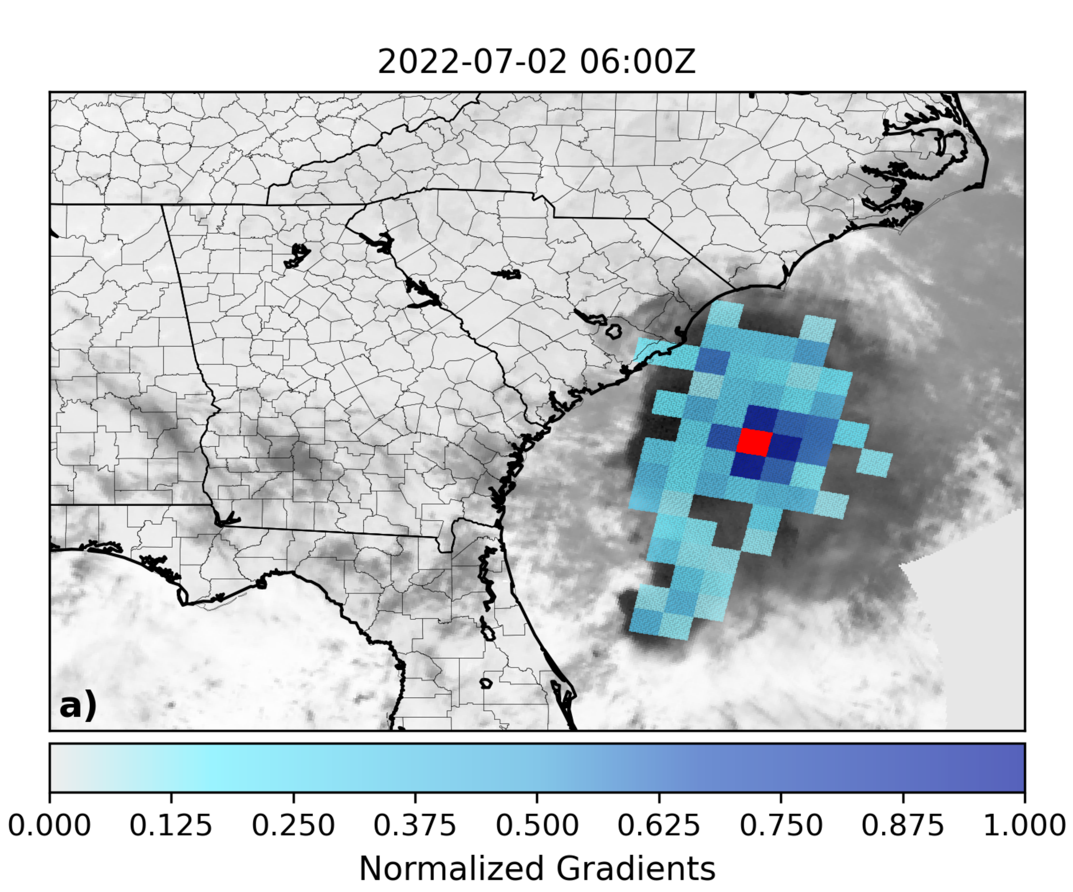
① compute the vector-Jacobian product for each intermediate token $\mathbf{z}_i$

$$\mathbf{g}_i = \mathbf{1} \cdot \frac{\partial \mathbf{z}_i}{\partial \mathbf{X}} = \sum_{k=1}^{d} \frac{\partial (\mathbf{z}_i)_k}{\partial \mathbf{X}} \in \mathbb{R}^{n \times d}$$

② construct the matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ with a reducing function $f : \mathbb{R}^d \rightarrow \mathbb{R}$

$$\mathbf{U} = \left[f(\mathbf{g}_1), f(\mathbf{g}_2), \ldots, f(\mathbf{g}_n) \right]^{\top}$$

③ visualize a token from the mean over the network, $\bar{\mathbf{U}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{U}^{(l)}$

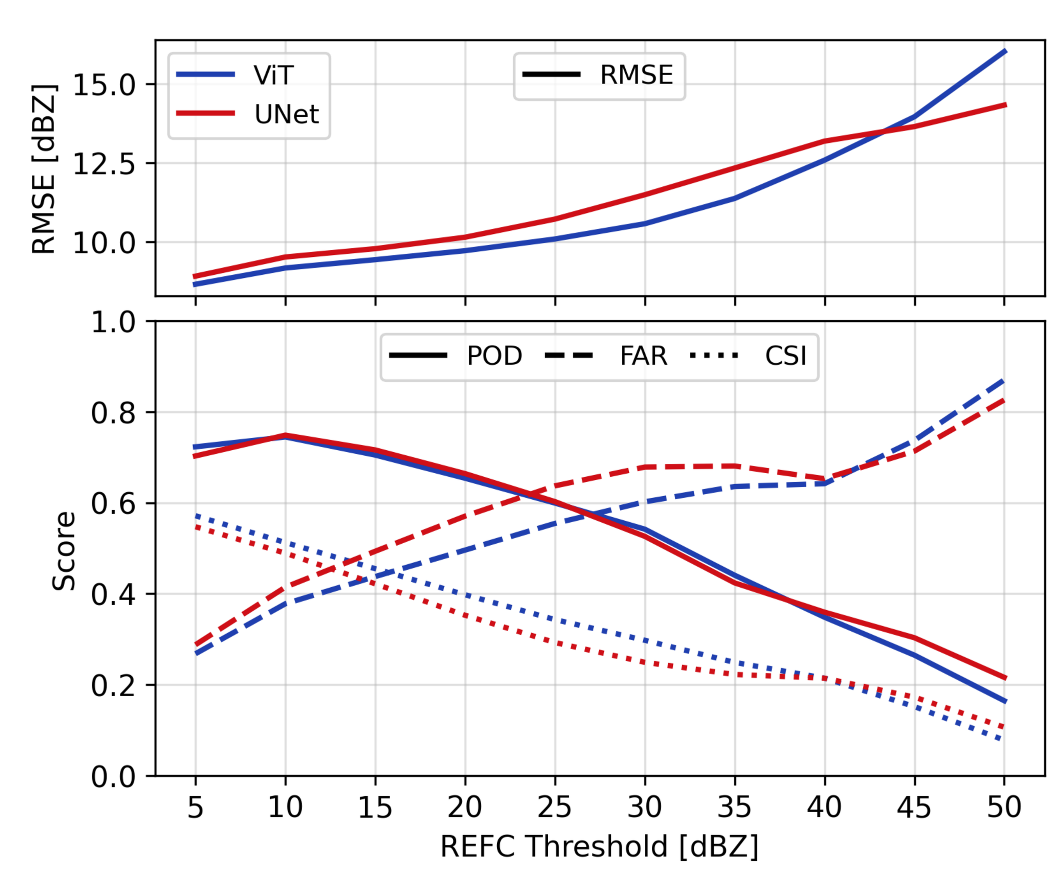


Experimental Results

Standard: mean statistics over the entire test set

Better overall pixel-wise performance

Model	↓ RMSE (dBZ)	↑ R^2	↑ Sharpness (g)
MRMS	–	–	0.48 ± 0.16
UNet	3.21	0.488	0.21 ± 0.09
Base-ViT	3.05	0.487	0.21 ± 0.09
SRViT	3.09	0.572	0.24 ± 0.11



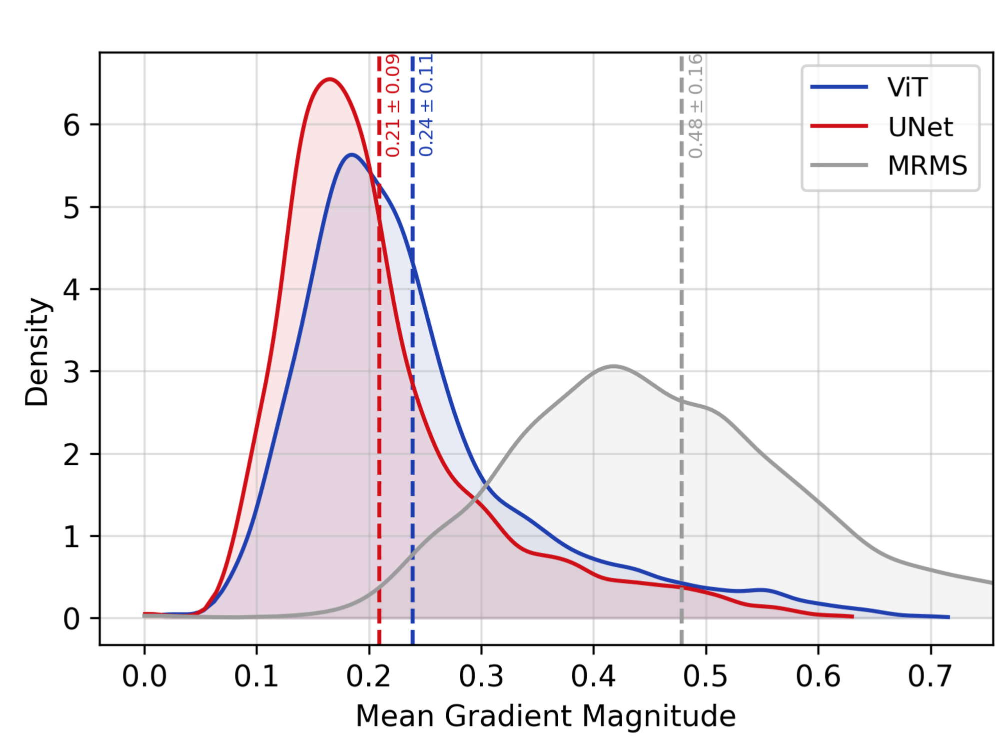
Categorical: probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) at varying composite reflectivity thresholds

Improves low- and mid-value estimates of reflectivity, < 40 dBZ

Sharpness: mean magnitude of image gradients, i.e., convolution of a Sobel filter,

$$g = \frac{1}{m}\sum_{i=1}^{m} (G_{x_i}^2 + G_{y_i}^2)^{\frac{1}{2}}$$

Sharper predictions over convolutional approaches



ICML 2024 ML4ESM Workshop — stock@colostate.edu | cs.colostate.edu/~stock