Representation learning of near-surface atmospheric fields using SHViT modules

10 Sep 2025

EMS Annual Meeting, Ljubljana, 7–12 Sep 2025

Martin Vozár
martin.vozar@iblsoft.com
martin.vozar@fmph.uniba.sk

IBL Software Engineering
FMPH, Comenius University, Bratislava

Introduction

Short abstract:

- CERRA dataset (Schimanke et al., 2021)
- SHViT architecture adaptations (Yun a Ro, 2024)
- Comparing optimization strategies.
- Domain generalization.

Study as a part of NWP postprocessing framework development at IBL Software Engineering.

Data: Subdomains

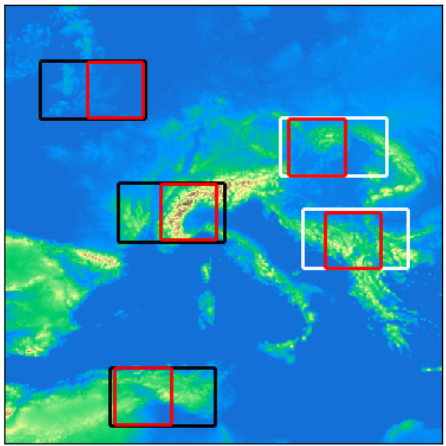
Selected regions:

- non-standard shapes
  (from previous experiments)

Square random crop:

- 64x64 pixels $\rightarrow$ $\approx$350x350 km

black - train/valid regions
white - test regions
red - random crop size



Data: Variables

During previous work and from related research:

GANs are tricky:

- in general
- when you use many variables
- when you use different variables/domain

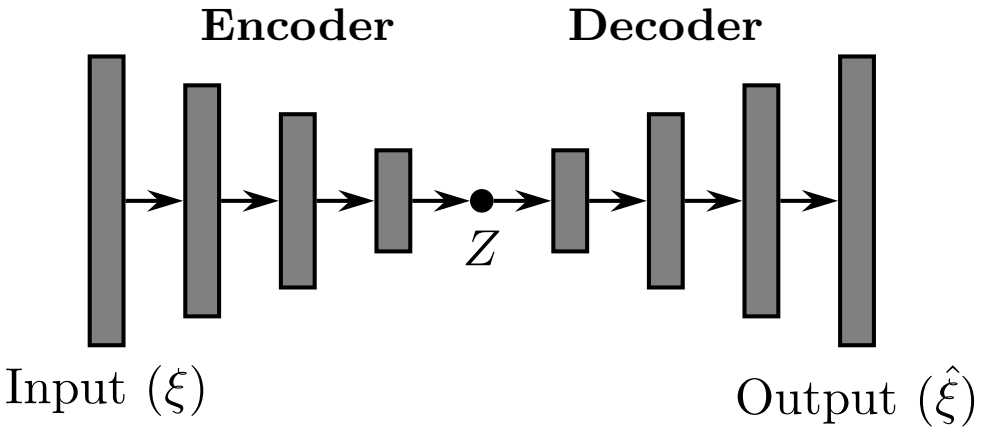
Selection of near-surface atmospheric fields:

- 2m temperature
- 2m relative humidity
- mean sea level pressure
- 10m wind speed

Sample-wise robust scaling

$$\xi_{bchw} \leftarrow \frac{x_{bchw} - \mathcal{Q}_{0.5}(x_{bc})}{\mathrm{IQR}(x_{bc})}$$

$$\mathrm{IQR}(\cdot) = \mathcal{Q}_{0.75}(\cdot) - \mathcal{Q}_{0.25}(\cdot)$$


Encoder
Decoder
z
Input (ξ)
Output $(\hat{\xi})$

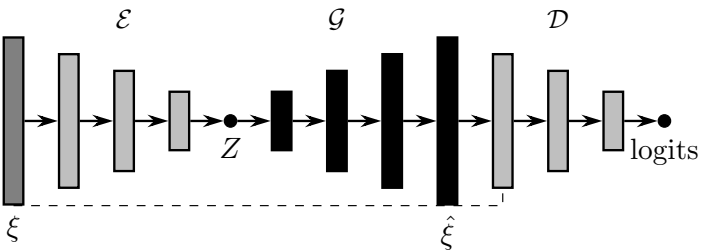
$$\xi \xrightarrow{\mathcal{E}} \mathbf{Z} \xrightarrow{\mathcal{G}} \hat{\xi} \xRightarrow{\mathcal{D}} \text{logits}$$

ξ : data ndarray
$\mathbf{Z}$: latent repr. vector

$\mathcal{E}$: Encoder
$\mathcal{G}$: Generator (Decoder)
$\mathcal{D}$: Discriminator

Representation learning

- Autoencoders (AE) $\mathcal{E},\ \mathcal{G}$
- Adversarial autoencoders (AAE) $\mathcal{E},\ \mathcal{G},\ \mathcal{D}$
- Generative Adversarial Networks (GAN) $\mathcal{G},\ \mathcal{D}$



Representation learning: some details

AE, AAE $\quad \mathbf{Z} \leftarrow \mathcal{E}(\xi)$

GAN $\quad \mathbf{Z} \leftarrow N(0, \mathbf{I})$

$$\hat{\xi} \leftarrow \mathcal{G}(\mathbf{Z})$$

$$[\text{logits}_\xi, \text{logits}_\mathcal{G}] \leftarrow \text{logits} \leftarrow \mathcal{D}([\xi, \hat{\xi}]_b)$$

AE $\quad \mathcal{L}_{\mathcal{E},\mathcal{G}} \leftarrow ||\hat{\xi} - \xi||_2$

AAE $\quad \mathcal{L}_{\mathcal{E},\mathcal{G}} \leftarrow ||\hat{\xi} - \xi||_2 + \text{sofplus}(-\text{logits}_\mathcal{G})$

$\mathcal{L}_\mathcal{D} \leftarrow \text{sofplus}(-\text{logits}_\xi) + \text{sofplus}(\text{logits}_\mathcal{G})$

GAN $\quad \mathcal{L}_\mathcal{G} \leftarrow \text{sofplus}(-\text{logits}_\mathcal{G})$

$\mathcal{L}_\mathcal{D} \leftarrow \text{sofplus}(-\text{logits}_\xi) + \text{sofplus}(\text{logits}_\mathcal{G})$

Representation learning: some details

For the networks using Discriminator, we alternate between AAE and GAN optimization schemes.

So far, it seems to stabilize the early optimization, but further research is needed.

SHViT modules: Encoder, idea

Single-Head Vision Transformer
SHViT (Yun a Ro, 2024)
CNN with Attention op.

Why this architecture?

- The paper presents strong claims about computational efficiency.
- The codebase is readable and easy to work with.

Encoder → Generator

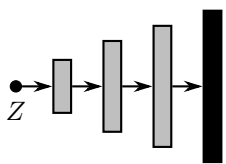
Most straightforward solution:
Inverse, symmetrical Generator.

More sophisticated ideas:
Spectral-hierarchical Generator
(Brochet et al., 2025; Karras et al., 2020)

SHViT modules: Generators

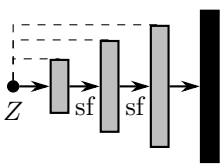
Inverse (symmetrical)

SHViTAE, SHViTGAN



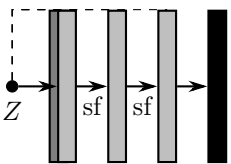
Spectral-hierarchical, increasing resolution between stages

SHhGAN 1



Spectral-hierarchical, full resolution in each stage

SHhGAN 2

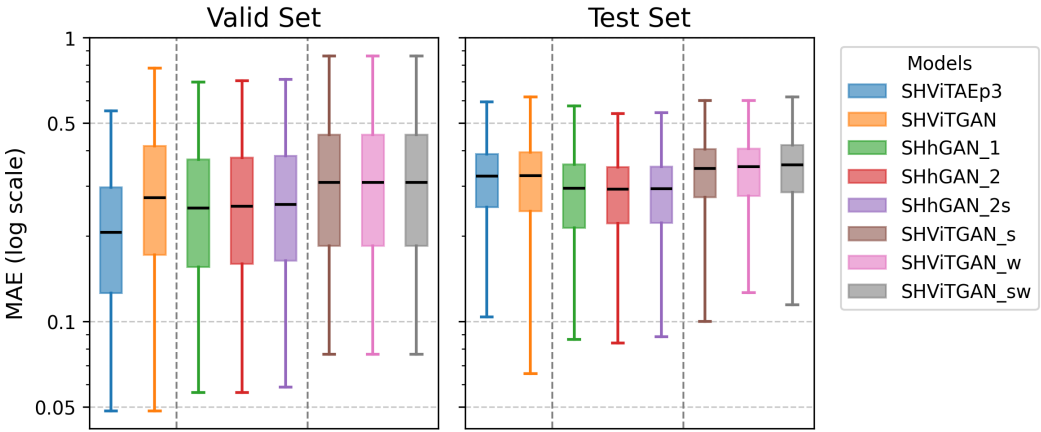


SHViT modules: Discriminator

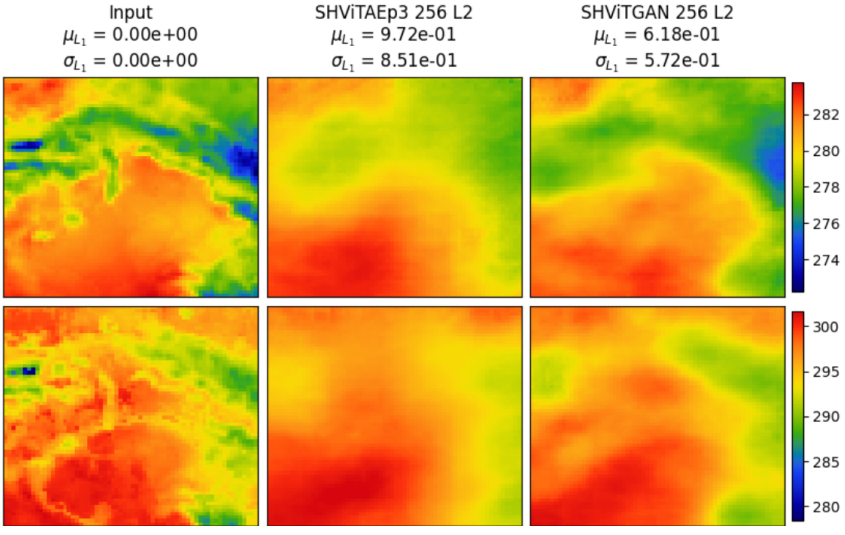
We use essentially the same module as for the Encoder.

We add the option to include
power spectrum and discrete wavelet transform
in the Discriminator channels.
(in fact, it did not help)

Results


Valid Set
Test Set
MAE (log scale)
1
0.5
0.1
0.05
Models
SHViTAEp3
SHViTGAN
SHhGAN_1
SHhGAN_2
SHhGAN_2s
SHViTGAN_s
SHViTGAN_w
SHViTGAN_sw

Results: Example (test set)



Results

Region	SHViTAEp3				SHViTGAN			
	t2m	mslp	rh2m	10mws	t2m	mslp	rh2m	10mws
$alps_0$	2.44	1458.82	41.35	0.93	2.44	1458.82	41.35	0.93
$brit_0$	0.57	272.95	8.90	0.65	0.57	272.95	8.90	0.65
$nafr_1$	1.23	488.50	20.35	0.84	1.23	488.50	20.35	0.84
$slvk_0$	1.03	29.37	3.97	0.55	0.74	20.91	3.05	0.42
$dalm_0$	1.64	48.66	5.75	0.68	1.30	36.62	4.67	0.56

Model and variable MAE grouped by region.

Discussion

Normalization:

- Method chosen for most direct comparison of "difficultyöf variables.
- Not necessarily optimal for all.

The experiments were primarily tuned for SHViTAE/GAN, while SHhGAN architectures are still experimental.

More thorough evaluation of the outputs characteristics is required.

Future work

GAN stability tuning
on different configurations.

Tuning the experiment on higher resolution,
preferably ensemble data,
using random crop.

Planned/considered datasets:

- UKV (2km) AWS rolling archive
- COSMO REA2

Development:

- Framework (pipelines, evaluation).
- SHhGAN architecture/s.

References i

Schimanke, S. et al. (2021). **CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present.** Accessed on 26-03-2025. DOI: 10.24381/cds.622a565a. URL: https://doi.org/10.24381/cds.622a565a.

Yun, Seokju a Youngmin Ro (2024). **SHViT: Single-Head Vision Transformer with Memory Efficient Macro Design.** arXiv: 2401.16456 [cs.CV]. URL: https://arxiv.org/abs/2401.16456.

Brochet, Clément et al. (2025). **"Enriching Operational High-Resolution Ensemble Forecasts with Generative Deep Learning".** In: *Artificial Intelligence for the Earth Systems* 4.1. DOI: 10.1175/AIES-D-24-0058.1. URL: https://doi.org/10.1175/AIES-D-24-0058.1.

Karras, Tero et al. (2020). **Analyzing and Improving the Image Quality of StyleGAN.** arXiv: 1912.04958 [cs.CV]. URL: https://arxiv.org/abs/1912.04958.