CrossRefine: A Microservice for Cross-Domain Spatial Super-Resolution

Daniil Sukhorukov[1,2], **Andrei Zakharov**[1] and **Ilya Makarov**[1,3]
[1]AXXX, Moscow, Russia
[2]ITMO University, Saint Petersburg, Russia
[3]Trusted AI Center, RAS, Moscow, Russia
{suhorukov.dan0, andzkrv}@gmail.com, iamakarov@hse.ru

Abstract

High-resolution spatial fields are critical for local decision-making, yet many operational and scientific workflows produce coarse outputs due to computational limits. We present *CrossRefine*, a deployable microservice for cross-domain spatial super-resolution that enhances multi-channel spatial tiles without modifying upstream models. It is built around a unified, topography-conditioned adversarial UNet trained across geographically diverse regions to ensure robustness to heterogeneous terrains and domain shifts. Unlike region-specific enhancement models, the system generalizes across domains within a single architecture, balancing numerical fidelity and structural realism through a hybrid regression–adversarial objective. The service provides REST API endpoints for batch and streaming inference, supports mixed-precision, and offers per-tile diagnostics and confidence maps to promote safe deployment. In our demo, we show interactive refinement of coarse spatial inputs, side-by-side comparison with interpolation and non-adversarial baselines, and real-time profiling of latency and throughput on commodity hardware. CrossRefine illustrates how spatial super-resolution can be delivered as a practical AI microservice, enabling scalable refinement of existing computational workflows without requiring higher-resolution upstream simulations.

Demonstration video: https://shorturl.at/lz2un

1 Introduction

High-resolution spatial fields are required by many downstream applications (hydrology, urban planning, precision agriculture), yet common operational and research pipelines produce outputs at coarse grids because increasing physical-model resolution is computationally expensive. For example, global reanalyses and operational forecasts are often supplied at resolutions on the order of $0.25° - 1°$ [Hersbach *et al.*, 2020], which limits the representation of orography, coastlines, and urban structure. In practice, users either run costly regional dynamical nests or accept interpolated coarse fields; both options have clear trade-offs in cost, latency, or fidelity.

We formulate the problem as multi-channel spatial super-resolution [Wang *et al.*, 2020]. Let $X_{LR} \in R^{C\times n\times m}$ denote a stack of coarse channels (temperature, winds, pressure, precipitation, etc.) and $X_{HR} \in R^{C\times N\times M}$ the desired high-resolution product with $N = k \cdot n, M = k \cdot m$. The goal is to learn a parametric mapping $f_\theta : X_{\text{LR}} \mapsto X_{\text{HR}}$ that (i) reconstructs fine spatial structure, (ii) preserves large-scale consistency, and (iii) is robust across geographic domains. Key practical constraints include limited HR training data in many regions, the need to avoid creating physically implausible artifacts, and the requirement that the module be lightweight and deployable as a post-processing service.

Prior work in computer vision (SRCNN [Dong *et al.*, 2015], UNet [Ronneberger *et al.*, 2015]) and generative modeling (SRGAN [Ledig *et al.*, 2017]) provides building blocks for this task; however, three domain-specific gaps remain for operational use: (1) single-region models overfit local statistics and do not generalize across heterogeneous terrains, (2) pointwise losses (MAE/MSE) produce overly smooth fields that miss localized extremes, and (3) many proposals ignore integration and latency requirements for real-time pipelines. Global products such as ERA5 [Hersbach *et al.*, 2020] and operational forecasts (GFS [NOAA NCEP, 2024]) are attractive training sources but exhibit region-dependent error characteristics that a practical model must accommodate.

In this work we adopt a systems-oriented approach: treat enhancement as a modular, deployable microservice and design the learning objective and architecture to satisfy cross-region generalization and operational constraints. Concretely, we (i) train a unified topography-conditioned UNet generator across diverse regions, (ii) introduce an adversarial loss to recover high-frequency structure while retaining a channel-weighted regression term for numerical fidelity, and (iii) optimize for inference efficiency and streaming deployment on commodity hardware. The demonstration uses public reanalysis/forecast products (e.g., from NOAA) to illustrate generalization, latency/throughput trade-offs, and qualitative gains over interpolation and non-adversarial baselines.

2 System Architecture

2.1 Processing Pipeline

Ingestion: tiled NetCDF/zarr [Ambatipudi and Byna, 2023] inputs (or streamed tiles) are converted to GPU tensors by a scalable loader. Training uses public reanalysis/forecast sources (e.g., ERA5 and GFS) for geographic diversity.

Preprocessing: per-channel geo-aware normalization, alignment to a static topo channel, and LR simulation via controlled avg-pooling ($\times 2$, $\times 4$). Tile sizes are chosen to fit GPU memory (e.g., 200×200 HR).

Model family: interchangeable generators (residual SRCNN [Jha *et al.*, 2025], topography-conditioned UNet [Vandal *et al.*, 2017; Sha *et al.*, 2020; Chiang *et al.*, 2024], adversarial UNet [Schönfeld *et al.*, 2021]) accept C×H×W tensors and output residual HR corrections. Loss is channel-weighted regression plus optional adversarial/perceptual term; simple physics checks (e.g., non-negative precipitation) are applied.

Training & generalization: multi-region sharded batches, data augmentation, and channel weighting to avoid dominance by large-magnitude fields; held-out regions verify cross-domain generalization.

Inference & deployment: packaged as a microservice with batched GPU queues and container manifests for batch or edge deployment. Seam-aware stitching and reprojection utilities assemble tiles into mosaics.

Performance & safety: tunable knobs (batch size, tile size, precision) and per-tile diagnostics (latency, baseline RMSE, physics flags). A conservative confidence mask can suppress adversarial detail where reliability is low.

2.2 Frontend Interface

Live demo UI: side-by-side panels (LR input, bicubic, UNet, SRGAN, reference when available), tile selector, and toggles for scale factor and adversarial weight.

Operational controls: upload tile or point to Zarr, trigger batch jobs, view per-tile latency/throughput, and inspect diagnostics and confidence maps.

Integration hooks: example client snippets (REST API / xpublish), and short instructions to embed the microservice into existing pipelines or cloud workflows for on-site demos.

3 Technical Implementation

The implementation is engineered for reproducible experiments and real-time demoability while retaining fidelity to operational constraints.

Data & preprocessing. Training uses public reanalysis and operational forecast archives: ERA5 (2018–2021) for train/validation and GFS (2022) as a held-out test. Input tiles contain five dynamic channels (T2m, U10, V10, MSLP, TP) plus a static topography channel; files are stored as NetCDF/zarr and loaded via xarray into PyTorch tensors. We cut global fields into HR tiles (typical 200×200), simulate LR inputs using controlled avg-pooling ($\times 2$, $\times 4$), and apply geo-aware per-channel normalization together with rotations, small translations, and seasonal stratification to improve cross-region robustness. Tile metadata (geo bounds, time) is preserved to enable reprojection and seam-aware stitching at inference.

Model implementations. All models share C×H×W I/O schema and are implemented in PyTorch. WeatherSRCNN is a compact, residual SRCNN variant adapted for multichannel inputs with early concatenation of topo information. WeatherUNet is a three-level encoder–decoder with DoubleConv blocks and skip connections; it predicts a residual correction added to bilinear upsampling to preserve large-scale structure. WeatherSRGAN pairs the UNet generator with a purpose-built discriminator (8 conv blocks with progressive downsampling and two FC layers) so the generator learns to recover high-frequency textures while keeping numerical consistency.

Losses & optimization. We optimize a hybrid objective: a channel-weighted regression term (MAE) ensures numerical fidelity—T2m and TP receive higher weights to avoid domination by large-magnitude fields such as pressure, and a small adversarial term ($\lambda = 0.005$) plus optional perceptual loss recovers structural realism. The discriminator uses binary cross-entropy; training employs Adam with standard learning-rate scheduling, gradient clipping, and alternating update steps to stabilize GAN training. Simple physics checks (e.g., non-negative precipitation) are computed during training and can be turned into soft penalty terms for conservative deployments.

Training & generalization. To produce a single multi-region model we mix tiles from different geographic shards and years within minibatches, use augmentation and seasonal sampling, and validate on held-out regions to detect overfitting to local climatologies. GAN stability is managed by limiting adversarial weight, using occasional discriminator freezes, and monitoring RMSE/MAE alongside perceptual metrics; typical training runs span $\sim$30 epochs per data shard with early stopping based on cross-region validation.

Inference, deployment & performance. The model is packaged as a Docker microservice exposing REST/gRPC endpoints for tiled inference and optional S3/zarr triggers. Inference supports mixed-precision (FP16) and an int8 quantization path for edge/booth hardware; tunable parameters include tile size, batch size, and precision to trade throughput vs. latency. Outputs include HR tiles, optional per-pixel confidence maps, and per-tile diagnostics (latency, baseline RMSE, physics flags); seam-aware stitching and reprojection utilities produce mosaics suitable for visualization or downstream ingestion.

Tools & reproducibility. Python, xarray, and PyTorch; the release bundle ships training recipes, prebuilt containers, a small test dataset, and Kubernetes manifests for single-GPU or small-cluster reproduction.

4 Demonstration & Evaluation

The performance of WeatherSRCNN, WeatherUNet, and WeatherSRGAN was evaluated using ERA5 reanalysis data and the GFS025 forecast dataset. For both sources, two input degradation scenarios were considered: spatial downscaling by a factor of $\times 2$ and $\times 4$, followed by reconstruction to the original $0.25^\circ \times 0.25^\circ$ grid.

RMSE was used as evaluation metric for each of the five variables (T2M, U10, MSLP, TP), as well as aggregated values across all channels. The ERA5 dataset (2018–2021) was used for training and validation, while GFS025 (2022) served as an external test set to assess the generalization capability of the models.

Scale	Model	T2M	U10	MSLP	TP
×2	Bicubic	.286	.167	.105	.104
	SRCNN	.219	.142	.095	.128
	UNet	.162	.097	.044	.079
	SRGAN	**.157**	**.089**	**.043**	**.076**
×4	Bicubic	.618	.447	.250	.196
	SRCNN	.423	.343	.220	.195
	UNet	.346	.246	.104	**.159**
	SRGAN	**.330**	**.242**	**.101**	.163

Table 1: **ERA5:** RMSE for $\times 2/\times 4$ downscaling to $0.25°$; our multi-region, topography-conditioned adversarial UNet improves over bicubic/SRCNN and a non-adversarial UNet (lower is better).

Scale	Model	T2M	U10	MSLP	TP
×2	Bicubic	.454	.399	.084	.158
	SRCNN	.391	.364	.116	.163
	UNet	.389	.347	.077	.129
	SRGAN	**.388**	**.343**	**.076**	**.128**
×4	Bicubic	.784	.690	.169	.258
	SRCNN	.618	.602	.240	.254
	UNet	.577	.537	.169	.227
	SRGAN	**.574**	**.535**	**.152**	**.226**

Table 2: **GFS025:** RMSE for $\times 2/\times 4$ downscaling to $0.25°$. Our multi-region, topography-conditioned adversarial UNet is best/near-best across variables (lower is better).

4.1 Quantitative Evaluation

The comparison results on ERA5 are presented in Table 1, and those on GFS025 are shown in Table 2. In all scenarios, the deep learning models substantially outperform bicubic interpolation.

It is important to note that on the external GFS025 test set, the differences between UNet and SRGAN remain consistent, indicating good generalization capability of the proposed adversarial architecture.

4.2 Qualitative Demonstration

To illustrate the differences between the methods, we present reconstruction maps of the surface temperature field (T2M) over Europe and the zonal wind component (U10) over the northern part of South America.

As shown in Figure 1, bicubic interpolation results in smoothing of temperature gradients, particularly in regions of strong advection. WeatherSRCNN partially restores spatial structure; however, a noticeable loss of high-frequency components remains. WeatherSRGAN provides sharper reconstruction of frontal boundaries and localized anomalies, while the corresponding error map exhibits lower amplitude and more spatially organized residual patterns.

A similar pattern is observed for zonal wind speed (Figure 2). In areas with complex coastlines and orography, WeatherSRGAN more effectively reproduces fine-scale circulation structures, whereas interpolation-based methods tend to smooth them out. This is particularly important for local forecasting tasks and the assessment of extreme events.

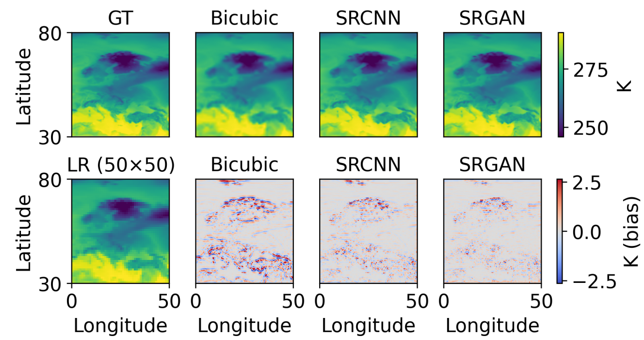

Figure 1: **2026-01-09/18:00.** Downscaling of the temperature field (T2M) by a factor of four. **Top row:** Ground Truth (GT), Bicubic, WeatherSRCNN, WeatherSRGAN. **Bottom row:** degraded input LR (AvgPool 50×50) and bias maps. SRGAN more effectively reconstructs small-scale gradients and frontal structures.

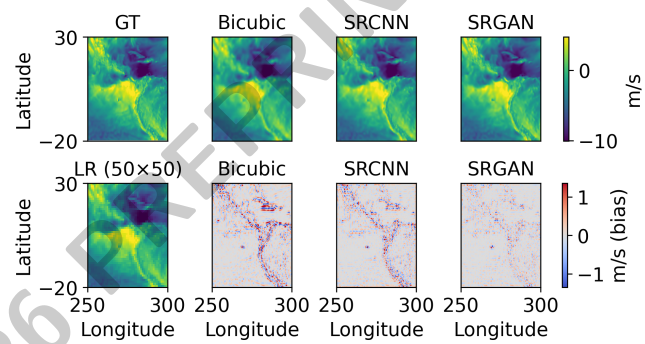

Figure 2: **2026-01-09/18:00.** Downscaling of the U10 field over northern South America. WeatherSRGAN reconstructs coastal and orographic effects more accurately, reducing local errors compared to interpolation and WeatherSRCNN.

5 Conclusion

This paper presents a deployable, topography-conditioned module that performs multi-channel spatial super-resolution across heterogeneous regions. Trained on reanalysis archives (e.g., ERA5) and evaluated on operational forecasts (e.g., GFS), the system demonstrates consistent numerical improvements over interpolation and non-adversarial baselines while recovering sharper local structure. The implementation is production-oriented: a Dockerized microservice with mixed-precision and quantization paths, tile stitching, and per-tile diagnostics for latency, error, and confidence. Future work will emphasize rigorous uncertainty quantification, stronger physics-aware constraints, and broader domain transfer; we also stress careful presentation of generated detail (confidence masks and physics checks) to avoid misinterpretation of synthesized high-frequency features.

Acknowledgements

The work was supported by a grant, provided by the Ministry of Economic Development of the Russian Federation (agreement dated June 20, 2025 No. 139-15-2025-011, identifier 000000C313925P4G0002).

References

[Ambatipudi and Byna, 2023] Sriniket Ambatipudi and Suren Byna. A comparison of hdf5, zarr, and netcdf4 in performing common i/o operations, 2023.

[Chiang *et al.*, 2024] Chia-Hao Chiang, Zheng-Han Huang, Liwen Liu, Hsin-Chien Liang, Yi-Chi Wang, Wan-Ling Tseng, Chao Wang, Che-Ta Chen, and Ko-Chih Wang. Climate downscaling: A deep-learning based super-resolution model of precipitation data with attention block and skip connections, 2024.

[Dong *et al.*, 2015] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks, 2015.

[Hersbach *et al.*, 2020] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049, 2020.

[Jha *et al.*, 2025] Shailesh Kumar Jha, Vivek Gupta, Priyank J Sharma, Anurag Mishra, and Saksham Joshi. Deep learning super-resolution for temperature data downscaling: a comprehensive study using residual networks. *Frontiers in Climate*, 7:1572428, 2025.

[Ledig *et al.*, 2017] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network, 2017.

[NOAA NCEP, 2024] NOAA NCEP. Global Forecast System (GFS). https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast, 2024. National Centers for Environmental Prediction; National Centers for Environmental Information (NCEI), Accessed: 2026-02-14.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[Schönfeld *et al.*, 2021] Edgar Schönfeld, Bernt Schiele, and Anna Khoreva. A u-net based discriminator for generative adversarial networks, 2021.

[Sha *et al.*, 2020] Yingkai Sha, David John Gagne II, Gregory West, and Roland Stull. Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. part i: Daily maximum and minimum 2-m temperature. *Journal of Applied Meteorology and Climatology*, 59(12):2057–2073, 2020.

[Vandal *et al.*, 2017] Thomas Vandal, Evan Kodra, Sangram Ganguly, Andrew Michaelis, Ramakrishna Nemani, and Auroop R Ganguly. Deepsd: Generating high resolution climate change projections through single image super-resolution, 2017.

[Wang *et al.*, 2020] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. Deep learning for image super-resolution: A survey, 2020.