A. Minkowski functionals foundations

Let $K \subset \mathbb{R}^2$ be a compact, convex set. The Steiner formula (Schneider and Weil, 2008) states that the area of the set of points within a distance r from K (for sufficiently small $r \geq 0$) is given by:

$$A_{K \oplus r} = A_K + P_K r + \chi_K \pi r^2 \tag{3}$$

where A_K is the area, P_K is the perimeter, and χ_K is the Euler characteristic of K. This formula holds not only for convex sets, but for any set with positive reach (Thäle, 2008), a condition satisfied by sufficiently regular boundaries. Equation (3) then holds for all r between 0 and the reach of K.

Equation (3) is a special case of the general Steiner formula in dimension $n \in \mathbb{N}$, where $K \subset \mathbb{R}^n$ is compact and convex (Schneider and Weil, 2008, Equation 1.2). In the general formulation, the n-volume of the dilated set $K \oplus r$ may be expressed as an n-degree polynomial in $r > 0$, and the coefficients of this polynomial define the $n + 1$ Minkowski functionals of K.

B. Topological Filtering via Persistent Homology

B.1. Stability and Feature Significance

While the Euler characteristic χ captures connectivity, the raw count of connected components is sensitive to noise; a single high-intensity pixel can create a spurious component, introducing discontinuities. To rigorously separate "signal" (stable storm cells) from "noise" (transient artifacts), we employ Persistent Homology.

Let X be a smooth function on domain D. We examine the excursion sets $E_u = \{s \in D : X(s) > u\}$. Components are "born" at local maxima and "die" when they merge with older components at saddle points (as u decreases). This lifecycle is captured by persistence pairs $(t_{\text{birth}}, t_{\text{death}})$.

Proposition 1. *For any threshold u, the number of connected components of the excursion set above level u equals the number of 0-dimensional persistence pairs such that $t_{\text{death}} \leq u < t_{\text{birth}}$.*

Proof of Proposition 1. The claim follows almost immediately from the fundamental lemma of persistent homology (Edelsbrunner and Harer, 2010, p. 181). The rank of the 0^{th} persistent homology group of the excursion set $\mathsf{H}_0(E_u)$ is given by the Betti number $\beta_0^{u,u} = \operatorname{rank} \mathsf{H}_0(E_u) = \{\#\,\text{connected components of } E_u\}$. Now, as a consequence of the fundamental lemma, $\beta_0^{u,u}$, which is the number of 0-dimensional homology classes alive at u, may be expressed as the number of 0-dimensional homology classes that are born before u (at some $t_{\text{birth}} > u$) and die after u (at some $t_{\text{death}} \leq u$). □

To stabilize the learning signal, we filter the persistence diagram by discarding pairs with a lifetime $t_{\text{birth}} - t_{\text{death}} < \epsilon$. This excludes points close to the diagonal of the persistence diagram, which correspond to low-persistence features indistinguishable from noise. By training on the filtered counts, we ensure the neural network focuses on learning robust, global topological structures rather than fitting numerical irregularities.

B.2. Numerical Implementation and Target Generation

We implement this filtration using the `GUDHI` library. Since standard algorithms compute sublevel set persistence, we input the negated precipitation field $-X$ to recover the superlevel set features required for precipitation clusters.

We fix the noise threshold at $\epsilon = 0.05\,\text{mm}\,\text{h}^{-1}$. This value was determined by analyzing the distribution of feature lifetimes on a validation set. The persistence histogram exhibits a clear separation between high-frequency numerical noise (clustering near the diagonal of the persistence diagram) and robust topological features (the heavy tail). The chosen threshold ϵ corresponds to the "elbow" point of this distribution, ensuring that we filter out transient artifacts while retaining physically significant convective structures.

The counting process is fully vectorized: for any given quantile level u, we sum the significant persistence pairs that satisfy the condition in Proposition 1. Additionally, we explicitly handle the infinite-persistence feature (representing the global background component); this feature is included in the count only for low-intensity thresholds ($u \leq 0.01\,\text{mm}\,\text{h}^{-1}$) to ensure consistent topology at the field boundaries.

To generate the complete γ-matrix for the entire dataset, we execute a rigorous offline target generation pipeline. This process extracts a four-dimensional geometric summary for every sample: area, perimeter, number of connected components (β_0), and number of topological holes (β_1).

First, we establish global physical intensity thresholds to ensure the evaluation levels are representative of the observed climatology. We compute an empirical cumulative distribution function by randomly sampling up to 5×10^7 precipitating pixels (exceeding the drizzle threshold of $0.1\,\text{mm}\,\text{h}^{-1}$) strictly from the training split. The user-defined quantile levels are then mapped to these empirical physical thresholds.

The persistent homology computations are fully vectorized across these threshold domains. Using the persistence diagram obtained from the cubical complex, we isolate both 0-dimensional and 1-dimensional persistence pairs. For a given threshold u,

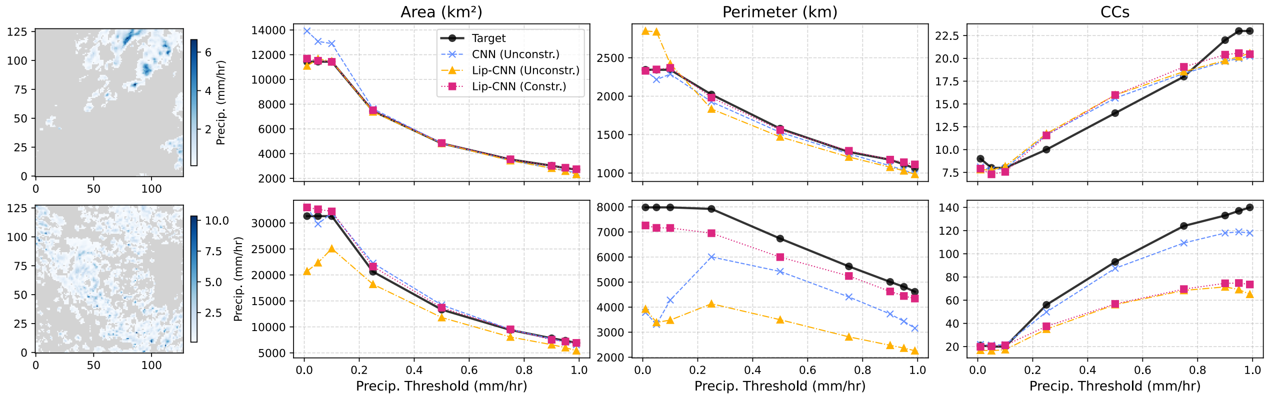

Figure 5. Qualitative prediction analysis. *Comparison of ground truth (solid) and predicted (dashed) Minkowski functionals.* ***(Top)*** *A well-behaved storm where all models converge.* ***(Bottom)*** *A challenging, fragmented case. Unconstrained models (blue/orange) predict geometrically impossible perimeters (violating monotonicity), whereas the Constrained model (magenta) maintains consistency by design.*

a topological feature is counted if its birth is greater than or equal to u, its death is strictly less than u, and its total persistence (lifetime) exceeds the empirically defined noise thresholds (which can be tuned independently for β_0 and β_1).

Concurrently, the extensive geometric measures are computed. Given a physical radar resolution of $2.0 \times 2.0\,\mathrm{km}^2$, the area at each threshold u is calculated simply as the sum of the binary excursion set multiplied by the pixel area ($4.0\,\mathrm{km}^2$). To calculate a mathematically robust perimeter, we apply the marching squares algorithm to the binary masks at a 0.5 contour level. The total perimeter is derived by integrating the Euclidean lengths of the resulting sub-pixel contour segments and scaling by the spatial resolution.

Finally, to manage the substantial combinatorial overhead of these exact topological and geometric evaluations, the target generation is executed offline using block-wise multiprocessing. The resulting multidimensional $\boldsymbol{\gamma}$-vectors are stored persistently in compressed Zarr arrays, which allows for highly efficient, chunked data loading within the downstream super-resolution optimization loop without bottlenecking the GPU accelerators.

C. Analytical Approximation of Euler Characteristic

The pronounced performance gap in predicting the number of connected components (R^2_{CC}) stems from a fundamental theoretical distinction in what the two surrogates actually evaluate. The analytical approximation computes the two-dimensional Euler characteristic (χ), which is topologically defined as the number of connected components (β_0) minus the number of topological holes (β_1). Consequently, the analytical method yields a net structural scalar ($\chi = \mathrm{CC} - \mathrm{H}$), whereas the neural emulator is trained to exclusively regress the pure connected component count (CC) derived from the exact ground truth. In complex precipitation fields, the prevalence of topological holes heavily skews the Euler characteristic, directly causing the severe degradation observed in the R^2_{CC} metric. Crucially, because the analytical formulation relies on localized morphological operations and continuous integrations, it is mathematically intractable to untangle the CC and H variables from the aggregated scalar χ. This structural entanglement represents a significant limitation of the analytical surrogate, underscoring the critical advantage of the neural emulator, which can isolate and map specific geometric invariants without topological interference.

D. Evaluation of Minkowski Functionals Emulations

D.1. Qualitative analysis of Emulated Minkowski Functionals

Figure 5 illustrates the emulation behaviors on specific samples. For standard storm structures (top row), all models capture the geometry with high fidelity. However, divergences appear in complex cases (bottom row). Notably, while the *vanilla CNN* predicts the area accurately, it fails to model the perimeter relationship correctly. The *constrained* model, by virtue of the conditioning $\hat{P} = f(\hat{A}, \hat{r})$, maintains a consistent perimeter curve. Interestingly, for this specific high-fragmentation case, the vanilla CNN slightly outperforms the Lipschitz models on the Connected Components (CC) metric. This local discrepancy suggests that the smoothing induced by spectral normalization may occasionally suppress the detection of extremely small, grid-scale fragments, a known trade-off of Lipschitz regularization.

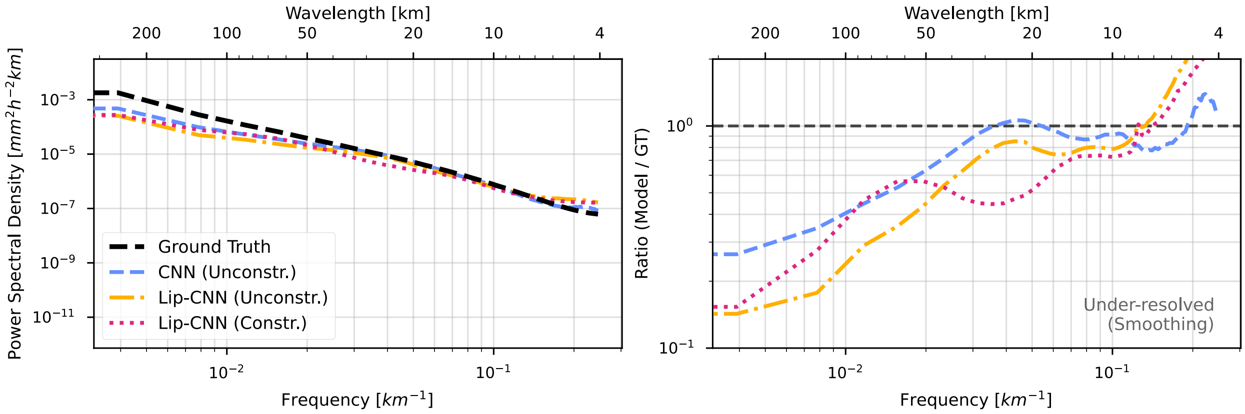

Figure 6. Spectral Validation. *(Left) Radially Averaged Power Spectral Density comparing the energy cascade of the ground truth (black) against the reconstructions. (Right) The spectral ratio $S_{model}(k)/S_{ref}(k)$. Ideal performance corresponds to $y = 1$.*

D.2. Diagnostic Inversion and Spectral Analysis via Radially Averaged Power Spectral Density

To verify that the emulator provides informative gradients, we perform a diagnostic inversion, steering a Gaussian noise vector $\mathbf{x_0}$ toward a target $\boldsymbol{\gamma}_{\text{true}}$ by minimizing the geometric discrepancy:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left(|| \log(\hat{\boldsymbol{\gamma}}_\theta(\mathbf{x}) + \mathbf{1}) - \log(\boldsymbol{\gamma}_{\text{true}} + \mathbf{1}) ||_2^2 + \lambda_{\text{TV}} \mathcal{R}_{\text{TV}}(\mathbf{x}) + \lambda_{L^2} ||\mathbf{x}||_2^2 \right) \tag{4}$$

where $\mathcal{R}_{\text{TV}}$ is the anisotropic total variation norm ($\lambda_{\text{TV}} = 10^{-5}$ and $\lambda_{L^2} = 10^{-6}$).

To ensure the gradients do not introduce nonphysical noise, we validate the spectral fidelity of the reconstructions via the radially averaged power spectral density. To quantify the multiscale structural fidelity of the reconstructed fields, we compute the Radially Averaged Power Spectral Density (RAPSD). Given a physical precipitation field $\mathbf{x} \in \mathbb{R}^{H\times W}$, we first apply a 2D Hanning window w to mitigate spectral leakage at the boundaries. We then compute the 2D Discrete Fourier Transform (DFT), denoted as $\hat{\mathbf{x}}(k_x, k_y)$. The 2D Power Spectral Density (PSD) is defined as:

$$P(k_x, k_y) = \frac{1}{(HW)^2} |\mathcal{F}\{w \odot \mathbf{x}\}(k_x, k_y)|^2 \tag{5}$$

Assuming statistical isotropy in the precipitation features, we reduce the 2D spectrum to a 1D profile by averaging the energy within annular bins of constant wavenumber magnitude $k = \sqrt{k_x^2 + k_y^2}$. The RAPSD, $S(k)$, is calculated as the azimuthal mean:

$$S(k) = \frac{1}{N_k} \sum_{(k_x,k_y)\in\mathcal{R}(k)} P(k_x, k_y) \tag{6}$$

where $\mathcal{R}(k)$ represents the set of discrete Fourier modes falling within the ring $k - \frac{1}{2} \leq \sqrt{k_x^2 + k_y^2} < k + \frac{1}{2}$ and $N_k = |\mathcal{R}(k)|$ is the number of modes in that bin. This metric allows us to verify if the emulator reconstructs the correct power-law decay (energy cascade) characteristic of atmospheric turbulence.

Figure 6 shows that all three architectures reproduce the fundamental power-law decay characteristic of three-dimensional, moist atmospheric turbulence, with no single architecture emerging as a definitive winner. Remarkably, the vanilla CNN achieves spectral fidelity comparable to the constrained models across the inertial subrange. This suggests that the Minkowski image loss is sufficiently informative to enforce structural realism even in a standard unconstrained network, effectively steering it away from the "checkerboard" artifacts typically associated with pixel-wise optimization.

However, a common deficiency is observed at the largest scales: all architectures under-resolve the lower frequencies (wavelengths >50 km), exhibiting a spectral ratio below 1.0. This indicates that while the Minkowski image loss successfully constrains local texture and object shapes (high-frequency topology), recovering the total energy of large-scale synoptic features remains a challenge for optimization-based inversion starting from noise. This difficulty stems directly from the localized inductive bias of standard convolutional operators, which inherently struggle to coordinate coherent mass transport over large spatial domains. To address this fundamental limitation and fully capture large-scale spectral energy, future work must explore global integration

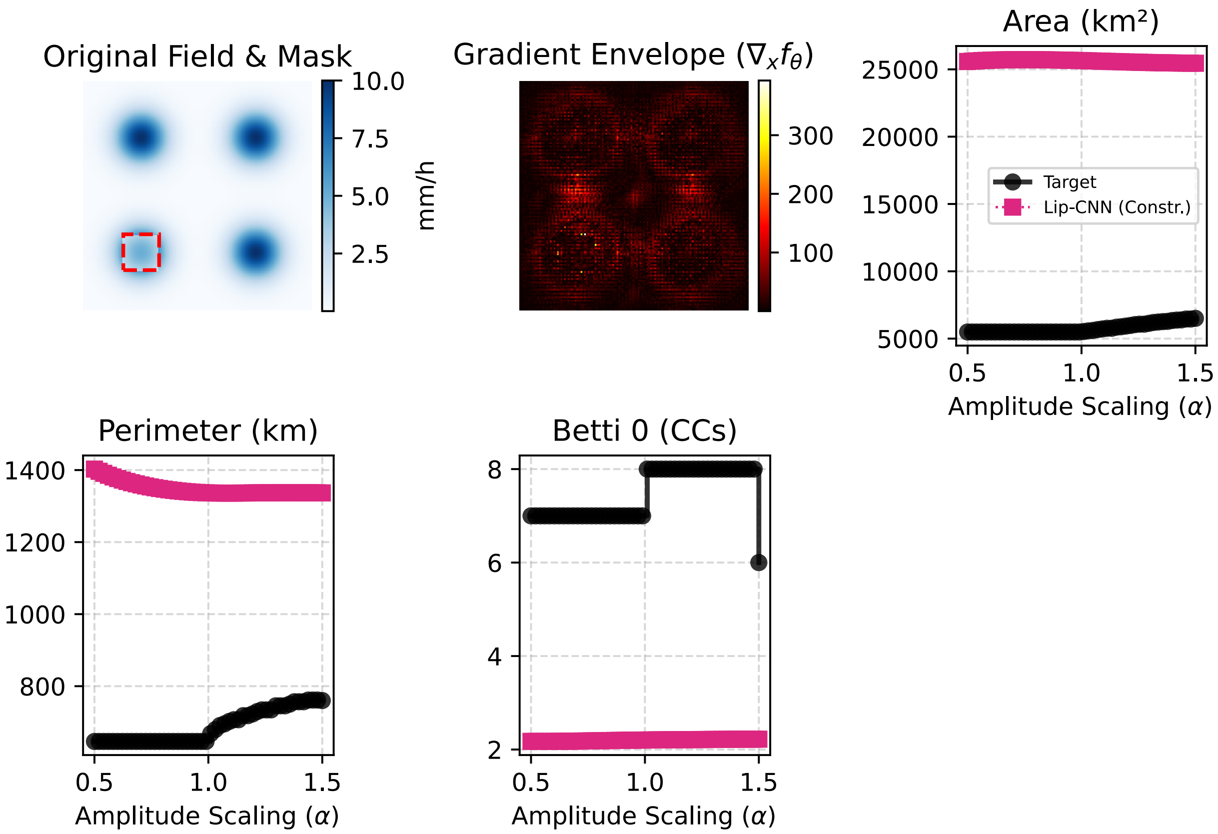

Figure 7. *Mechanistic evaluation of the surrogate's sensitivity to localized physical perturbations.* ***(Left)*** *A synthetic multi-peak Gaussian precipitation field. The red dashed box denotes the spatial mask applied to a single convective cell, where the intensity is systematically scaled by an amplitude factor α.* ***(Middle)*** *The spatial attribution map representing the surrogate's gradient envelope ($\nabla_{\mathbf{x}} f_\theta$), revealing a fragmented and noisy receptive field that lacks cohesive internal structure.* ***(Right)*** *The one-dimensional amplitude perturbation landscape for area, perimeter, and connected components (Betti 0). As the targeted cell's amplitude increases, the exact topological metrics (black) exhibit discrete step transitions when crossing the evaluation threshold. In contrast, the predictions from the constrained Lip-CNN (magenta) remain entirely static across the perturbation range, demonstrating that aggressive Lipschitz regularization over-smooths the latent space and renders the emulator blind to localized structural variations..*

architectures, such as Fourier neural operators (FNOs), as a necessary pathway to overcome the spatial constraints of conventional CNNs.

D.3. Mechanistic Evaluation on Synthetic Fields

To rigorously evaluate the emulator's sensitivity to localized physical changes, we conduct a mechanistic proof using a synthetic multi-peak Gaussian field. As shown in Figure 7, a spatial mask is applied to a single convective cell, and its intensity is systematically perturbed by an amplitude scaling factor α. As illustrated in the spatial attribution map, the surrogate's gradient envelope ($\nabla_{\mathbf{x}} f_\theta$) loosely identifies the active convective regions but exhibits a highly fragmented and noisy internal structure. More critically, the one-dimensional amplitude perturbation landscape exposes a fundamental failure in the surrogate's forward mapping. As the amplitude of the targeted cell varies, the exact integral-geometric targets (area, perimeter, and connected components) exhibit the expected discrete step transitions when the local intensity crosses the evaluation threshold. However, the constrained Lip-CNN predictions remain entirely static across the entire perturbation range. This extreme insensitivity demonstrates that the current regularization strategy—likely the interaction between aggressive spectral normalization and global pooling—severely over-smooths the latent representation. Consequently, the emulator becomes completely blind to isolated, localized intensity variations, rendering it incapable of providing the precise gradient signals required to optimize fine-scale precipitation structures.

E. Super-resolution Training Protocols

E.1. Preprocessing

To facilitate stable optimization across heavily skewed atmospheric variables and disparate physical units, we apply rigorous normalization protocols to the input tensors. The coarse-resolution precipitation inputs and high-resolution targets first undergo a log-linear transformation: variables are shifted via $x \mapsto \log(1 + x)$ and subsequently divided by the global maximum of the log-transformed training set, constraining the values to a strictly positive $[0, 1]$ interval. For the stochastic diffusion baseline (DDPM), these normalized tensors are further linearly mapped to the $[-1, 1]$ domain to satisfy the standard Gaussian noise assumptions inherent to the reverse diffusion process. The static orographic forcing, represented by a high-resolution digital elevation model (DEM), is standardized using the domain-wide mean and standard deviation to yield zero mean and unit variance. The normalized DEM is concatenated channel-wise with the low-resolution precipitation field to explicitly condition the generative spatial downscaling on underlying topography. The target integral-geometric metrics (γ-vectors) are exclusively log-transformed to compress their dynamic range prior to loss computation.

During training, the dataset is dynamically augmented using random horizontal and vertical flips (each applied with a 50% probability) to promote rotational and reflectional invariance. To address the severe class imbalance typical of meteorological radar data, where dry or drizzling pixels disproportionately dominate the spatial domain, the deterministic UNet architectures are trained using a stratified sampling strategy. We compute independent sample weights based on a binary wet/dry classification—defined by a localized maximum intensity threshold of 0.1 mm/h—and employ a weighted random sampler. This ensures the optimization landscape is adequately exposed to structurally complex, high-intensity convective events, preventing the network from collapsing to a climatological dry mean.

E.2. Dynamic Loss Weighting and Trust Mechanism

While the baseline log-space residual UNet is optimized solely via the MSE between the predicted and target log-transformed fields, the topologically constrained configurations augment this objective with the Minkowski image loss. The total loss function is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + w_{\text{geom}}(t) \cdot w_{\text{trust}} \cdot \mathcal{L}_{\text{Minkowski}} \tag{7}$$

where $w_{\text{geom}}(t)$ is a dynamic scaling factor governed by a cosine annealing schedule over the training epochs t. This allows the model to first learn large-scale mass conservation before focusing on fine-scale topological adjustments. Crucially, when utilizing the neural emulator, we introduce a dynamic trust mechanism $w_{\text{trust}} = \exp(-\tau \|\hat{\gamma}_{\text{true}} - \gamma_{\text{true}}\|^2)$ that penalizes the geometric loss component when the emulator's localized approximation error on the ground truth is high, parameterized by $\tau = 0.005725$. This ensures the generator is not penalized by out-of-distribution emulator hallucinations.

E.3. Trust coefficient computation

To compute the penalty parameter τ, we perform an offline calibration pass over the validation dataset to characterize the neural emulator's baseline error distribution. High-resolution ground truth precipitation fields are denormalized from the scaled training domain back to physical intensity values (mm h^{-1}) and masked by the $0.1\,\text{mm h}^{-1}$ drizzle threshold. These physical fields are processed by the frozen emulator to predict the multidimensional integral-geometric components. To match the scale of the objective function, the predictions undergo a logarithmic transformation, $\log(1 + x)$, before computing the mean squared error against the exact ground truth $\boldsymbol{\gamma}$-vectors. We extract the 90th percentile of this error distribution across the true climatology. The coefficient τ is algebraically derived to aggressively attenuate the geometric loss signal for out-of-distribution structures, specifically calibrating the trust factor w_{trust} to decay to a minimal weight of 0.1 when the emulator's approximation error reaches this 90th percentile threshold.

F. List of hyperparameters

Table 3. *Hyperparameters and configuration settings for the emulator and feature inversion experiments. Parameters with $^{(*)}$ are calibrated using the `Optuna` package (Akiba et al., 2019).*

Category	Parameter	Value
Data & Preprocessing	Patch Size	128×128
	Pixel Resolution	2.0 km
	Quantile Levels (Q)	$0.01, 0.05, 0.1, \ldots$
	Drizzle Threshold	0.1 mm/h
	Persistence Threshold	0.05 mm/h
Architecture (Emulator CNN)	Base Channel Width	32
	Max Channel Width	256
	Network Depth	5 layers
	Latent Dimension	256
	Normalization	GroupNorm ($G = 8$)
	Activation	GELU
Training (Emulator)	Optimizer	Adam
	Learning Rate$^{(*)}$	7.75×10^{-5}
	Weight Decay$^{(*)}$	2.49×10^{-6}
	Batch Size	128
	Max Epochs	50
	Early Stopping	15 epochs
	LR Scheduler	ReduceLROnPlateau
Loss (Emulator)	Weight λ_{Area}	3.0
	Weight $\lambda_{\text{Perimeter}}$	1.0
	Weight λ_{CC}	1.5
Inversion / Analysis	Optimization Steps	200
	Inversion LR	0.1
	Reg. λ_{TV}	1×10^{-5}
	Reg. λ_{L^2}	1×10^{-6}

Table 4. *Hyperparameters and configuration settings for the super-resolution experiments. Parameters with $^{(*)}$ are calibrated using the `Optuna` package (Akiba et al., 2019).*

Category	Parameter	Value
Architecture (UNet SR)	Model	LogSpaceResidualUNet
	Input Channels	2
	Output Channels	1
	Downscaling Factor	12.5×
Architecture (DDPM)	Model	Conditional ContextUNet
	Condition Channels	2
	Fwd Diff. Timesteps	1000
	DDIM Inference Steps	50
Training (SR — shared)	Optimizer	AdamW
	Batch Size	128
	Max Epochs	25
	Early Stopping	5 epochs
	LR Scheduler	ReduceLROnPlateau
	UNet LR$^{(*)}$	Optuna-tuned
	UNet WD$^{(*)}$	Optuna-tuned
	DDPM LR / WD	1.0×10^{-5}
Loss (UNet SR)	Base Loss	MSE (log space)
	Geo Loss Schedule	Cosine annealing
	Geo Warmup	5 epochs
	Gradient Clip	Max norm 1.0
	Trust Coeff. τ	5.725×10^{-3}