MISR-GS: TEST-TIME MULTI-IMAGE SUPER-RESOLUTION WITH EXPLICIT 2D GAUSSIAN FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-image super-resolution (MISR) reconstructs a high-resolution image from multiple low-resolution observations whose sub-pixel shifts provide complementary measurements of the latent high-resolution signal. Recent test-time optimization methods represent this shared signal with coordinate-based implicit neural representations (INRs), but their globally parameterized capacity and frequency-encoding hyperparameters motivate a broader question: what representation is best suited for an observation-constrained, test-time MISR inverse problem? We propose MISR-GS, which represents the latent high-resolution image as a shared explicit field of learnable anisotropic 2D Gaussian primitives. Because standard 2D Gaussian image fitting can allocate primitives using an observed target image whereas the high-resolution target is unavailable in MISR, MISR-GS derives both initialization and progressive capacity allocation solely from the low-resolution burst. A pseudo-HR reference is used only to initialize Gaussian positions and colors, while multi-frame reprojection residuals guide subsequent densification. The resulting field, frame-wise affine alignment, and photometric parameters are jointly optimized through a differentiable render-and-degrade model in the low-resolution observation space, with an LR-only ringing-aware regularizer stabilizing weakly constrained high-frequency solutions. Across the reported $\times 2/\times 4/\times 8$ settings, MISR-GS improves PSNR over the strongest evaluated optimization-based TTO baseline by 0.98–2.97 dB, while requiring no paired HR training data. We additionally evaluate computational cost and robustness to challenging alignment conditions to characterize the quality–efficiency trade-off of explicit Gaussian fields for test-time MISR.

1 INTRODUCTION

High-resolution imaging is often limited by sensor resolution and acquisition conditions. Image super-resolution (SR) aims to recover high-frequency signals from degraded observations. However, single-image super-resolution (SISR) methods typically rely on image priors learned from large-scale high-resolution datasets Ledig et al. (2017); Zhang et al. (2023). Although these priors can produce visually appealing results, the recovered details may not always be directly supported by observations, which is undesirable in applications requiring faithful reconstruction.

Multi-image super-resolution (MISR) provides a more observation-constrained setting by exploiting multiple low-resolution (LR) observations of the same scene. Due to sub-pixel shifts between frames, an LR burst contains complementary samples of the latent high-resolution (HR) image Tsai & Huang (1984); Irani & Peleg (1991); Elad & Feuer (1997). Therefore, the goal of MISR is not only to enhance visual sharpness, but to recover image structures that are supported by multiple observations.

Recent test-time optimization (TTO) approaches formulate MISR as an inverse reconstruction problem, where an image-specific HR representation is optimized to reproduce the observed LR burst without requiring paired LR-HR training data Nam et al. (2022); Jyhne et al. (2026). SuperF Jyhne et al. (2026) represents the unknown HR image using a shared implicit neural representation (INR) and jointly optimizes frame transformations with the image representation. This formulation demonstrates that a shared continuous representation can effectively exploit complementary infor-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

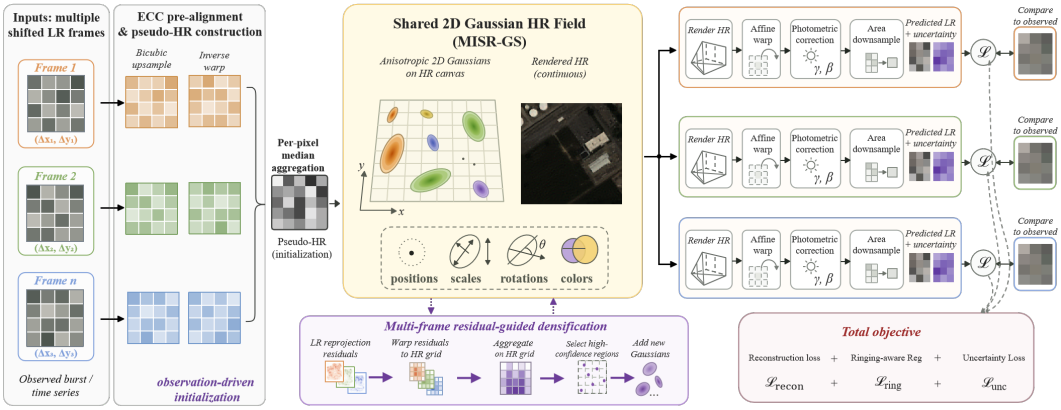

Figure 1: **Pipeline.** MISR-GS represents the unknown HR image as a shared explicit 2D Gaussian field and optimizes it through differentiable render-and-degrade supervision from multiple shifted LR observations. Gaussian initialization and progressive densification are driven solely by LR observations and multi-frame reprojection residuals.

mation from LR bursts. However, it also raises an important question: *Can an explicit, spatially adaptive representation provide an effective alternative for LR-supervised test-time MISR?*

Coordinate-based INRs provide continuous image representations through implicitly parameterized functions. However, their representation capacity is distributed through network parameters rather than explicitly assigned to spatial regions. In contrast, explicit Gaussian representations model images using spatially localized primitives, allowing representation capacity to be directly controlled and allocated. This difference suggests that explicit representations may provide a useful alternative for MISR, where the reconstruction process must infer both image content and representation capacity from LR observations.

Recent 2D Gaussian representations have demonstrated promising results for image modeling and rendering through explicit spatial primitives Zhang et al. (2024; 2025); Zhu et al. (2025). However, existing Gaussian-based image reconstruction methods typically assume access to target images during optimization or rely on learned predictors to initialize and guide the representation. Such settings are fundamentally different from LR-supervised TTO MISR, where the latent HR image is unavailable. Therefore, a key challenge remains: how to construct and optimize an explicit Gaussian field when the HR target is never observed.

In this work, we introduce *MISR-GS*, an explicit Gaussian representation framework for LR-only test-time MISR. MISR-GS models the latent HR image using learnable anisotropic 2D Gaussian primitives. Since the HR target is unavailable, Gaussian initialization and progressive capacity allocation are derived directly from LR observations. The resulting Gaussian field is jointly optimized with frame-specific geometric and photometric parameters through a differentiable render-and-degrade model constrained by LR observations. Our contributions are summarized as follows:

- We introduce MISR-GS, an LR-supervised test-time MISR framework based on an explicit 2D Gaussian field, demonstrating that explicit spatial primitives provide an effective alternative to implicit neural fields for multi-image reconstruction.
- We propose an observation-driven Gaussian allocation strategy that constructs and refines the explicit representation directly from LR bursts, without requiring access to HR targets.
- We develop an LR-constrained optimization framework that jointly optimizes the Gaussian field and imaging parameters through differentiable render-and-degrade supervision, enabling stable reconstruction from multiple LR observations.

2 RELATED WORK

2.1 MULTI-IMAGE SUPER-RESOLUTION (MISR)

MISR has been studied through classical reconstruction, learning-based burst fusion, and test-time optimization (TTO) approaches Khattab et al. (2020). Classical and optimization-based methods exploit complementary sub-pixel information from observed burst sequences. Wronski et al. (2019) use steerable kernel regression for handheld multi-frame super-resolution, while Lafenetre et al. (2023) extend this paradigm to multi-exposure satellite imagery.

Learning-based MISR methods learn burst alignment and fusion from data. Existing approaches include supervised Bhat et al. (2021); Cornebise et al. (2022) and self-supervised Nguyen et al. (2022) formulations. Recent methods further improve reconstruction through feature-level alignment and aggregation. BSRT uses flow-guided deformable alignment with a Swin Transformer Luo et al. (2022), while Burstormer models multi-scale local and non-local burst interactions for joint super-resolution and restoration Dudhane et al. (2023). These approaches rely on learned priors, whereas this work focuses on per-instance optimization from observed LR bursts.

Closer to our setting, INR-based TTO methods optimize a continuous HR representation directly from LR observations. SuperF Jyhne et al. (2026) shares an INR across multiple LR frames and jointly optimizes frame transformations with the implicit representation. However, explicit representation choices for LR-supervised TTO MISR remain less explored. We therefore investigate a 2D Gaussian representation as the shared HR field, where spatially localized primitives are optimized directly from LR observation consistency.

2.2 EXPLICIT GAUSSIAN REPRESENTATIONS

Explicit representations encode visual content using directly optimized primitives, providing an alternative to coordinate-based implicit neural representations. Gaussian splatting has become a representative explicit representation paradigm for 3D scene modeling and novel-view synthesis Fei et al. (2024); Kerbl et al. (2023), and has also been explored in applications such as satellite reconstruction Aira et al. (2025) and medical imaging Yang et al. (2026).

For image super-resolution, recent works extend Gaussian representations from feature modeling to image-space rendering. GaussianSR Hu et al. (2025) uses adaptive 2D Gaussian fields to represent LR latent features for arbitrary-scale SISR, while ContinuousSR Peng et al. (2026) predicts a continuous HR Gaussian field directly from LR images. GSMSR Cui et al. (2026) further applies 2D Gaussian representations to multi-modal MRI super-resolution with cross-modal interaction and Gaussian alignment.

However, existing Gaussian-based super-resolution methods mainly rely on offline training or HR supervision. Their Gaussian fields are initialized or guided by learned predictors or available HR targets. In contrast, MISR-GS optimizes the Gaussian representation directly from multiple LR observations in a per-burst TTO setting. Without access to HR supervision during optimization, the primitive initialization and capacity allocation must be inferred from the observed LR burst itself, leading to the observation-driven Gaussian allocation strategy proposed in this work.

3 METHOD

MISR-GS follows a representation-centered formulation for LR-supervised test-time reconstruction. We represent the latent HR image using a shared explicit 2D Gaussian field with spatially localized primitives that can be directly optimized and allocated. Since the HR target is unavailable, Gaussian initialization is derived solely from the observed LR burst. During optimization, additional primitives are allocated according to persistent multi-frame LR reprojection errors. The resulting Gaussian field is optimized through a differentiable render-and-degrade process under LR observation constraints. The following subsections describe each stage in detail.

3.1 Problem Formulation

Given a set of T low-resolution observations $\{I_{\mathrm{LR}}^{(t)}\}_{t=1}^{T}$ of the same scene with unknown sub-pixel misalignment across frames, we aim to recover a single shared high-resolution image I_{HR} while jointly solving for the geometric and photometric parameters of each frame. Let Ω_{HR} and Ω_{LR} denote the HR and LR image domains, respectively. No high-resolution ground truth is available during optimization. We assume that each LR frame is generated from the shared HR signal through frame-specific geometric misalignment, spatial averaging, photometric variation, and observation noise:

$$I_{\mathrm{LR}}^{(t)} \approx \mathcal{C}^{(t)}\left(\mathcal{D}_s\left(\mathcal{T}_{\mathbf{A}^{(t)}}(I_{\mathrm{HR}})\right)\right) + \boldsymbol{\epsilon}^{(t)}, \qquad t = 1, \ldots, T, \tag{1}$$

where $\mathcal{T}_{\mathbf{A}^{(t)}}$ denotes geometric warping under the frame-specific transformation $\mathbf{A}^{(t)}$, $\mathcal{D}_s$ denotes box-filtered area downsampling with scale factor s, $\mathcal{C}^{(t)}$ models frame-specific photometric variation, and $\boldsymbol{\epsilon}^{(t)}$ denotes observation noise. The reconstruction model treats $\mathcal{D}_s$ as its assumed inference-time degradation operator; it does not require access to the exact image-formation kernel used to generate or acquire the LR observations. The first frame is used as the reference frame, with $\mathbf{A}^{(1)} = \mathbf{I}$. The test-time MISR problem is to estimate the latent HR signal and the frame-wise nuisance parameters solely from the LR observations.

3.2 Shared Explicit 2D Gaussian Field

Gaussian Function. We represent the unknown HR image $I_{\mathrm{HR}} \in \mathbb{R}^{H\times W\times 3}$ using a finite set of learnable anisotropic 2D Gaussian primitives $\mathcal{G} = \{g_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \mathbf{c}_i)\}_{i=1}^{N}$, where $\boldsymbol{\mu}_i \in \mathbb{R}^2$ denotes the Gaussian center, $\boldsymbol{\Sigma}_i \in \mathbb{R}^{2\times 2}$ the covariance matrix, and $\mathbf{c}_i \in \mathbb{R}^3$ the RGB color. We build on the optimized 2D Gaussian rasterization implementation of Image-GS Zhang et al. (2025), while using the image-plane parameterization described below. Because all primitives lie in a single 2D image plane and no visibility ordering is required, opacity-based alpha compositing for occlusion handling is unnecessary; we instead use normalized kernel aggregation. Thus, the borrowed rasterization backend should not be interpreted as reusing the full Image-GS image-fitting objective or target-dependent allocation strategy. The spatial influence of the i-th Gaussian at pixel $\mathbf{x} = (x, y)$ is

$$\phi_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^{\top}\boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right). \tag{2}$$

Covariance Decomposition. To maintain the positive semi-definite constraint of $\boldsymbol{\Sigma}_i$, we employ the Rotation-Scaling (RS) decomposition from Fast2DGS Wang et al. (2026). For each Gaussian, we maintain a rotation angle $\theta_i \in [0, \pi]$ and a positive scaling vector $\boldsymbol{\sigma}_i = (\sigma_{x,i}, \sigma_{y,i}) \in \mathbb{R}_+^2$:

$$\boldsymbol{\Sigma}_i = \mathbf{R}(\theta_i)\mathbf{S}(\boldsymbol{\sigma}_i)\mathbf{S}(\boldsymbol{\sigma}_i)^{\top}\mathbf{R}(\theta_i)^{\top}, \tag{3}$$

where the rotation matrix is defined as $\mathbf{R}(\theta_i) = \begin{bmatrix}\cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i\end{bmatrix}$, and the scaling matrix is $\mathbf{S}(\boldsymbol{\sigma}_i) = \mathrm{diag}(\sigma_{x,i}, \sigma_{y,i})$.

Rendering Model. Since a 2D image representation does not require depth sorting, the alpha-blending operation used in 3DGS Kerbl et al. (2023) is replaced by normalized weighted aggregation:

$$\hat{I}_{\mathrm{HR}}(\mathbf{x}) = \frac{\sum_{i\in\mathcal{N}(\mathbf{x})}\phi_i(\mathbf{x})\,\mathbf{c}_i}{\sum_{i\in\mathcal{N}(\mathbf{x})}\phi_i(\mathbf{x}) + \varepsilon_{\mathrm{num}}}, \tag{4}$$

where $\mathcal{N}(\mathbf{x})$ denotes the set of Gaussians contributing to pixel $\mathbf{x}$, and $\varepsilon_{\mathrm{num}} > 0$ ensures numerical stability. For convenience, we denote the differentiable rasterization of the Gaussian field as $\hat{I}_{\mathrm{HR}} = \mathcal{R}_{\mathrm{GS}}(\mathcal{G})$.

3.3 GAUSSIAN-SPECIFIC LR-SPACE OPTIMIZATION

Render-and-Degrade Optimization. The shared Gaussian field is optimized by explaining all LR observations through a differentiable render-and-degrade process. For the t-th frame, the predicted LR observation is

$$\hat{I}_{\mathrm{LR}}^{(t)} = \mathcal{C}^{(t)}\left(\mathcal{D}_s\left(\mathcal{T}_{\mathbf{A}^{(t)}}\left(\mathcal{R}_{\mathrm{GS}}(\mathcal{G})\right)\right)\right). \tag{5}$$

We parameterize the frame-specific photometric correction as

$$\mathcal{C}^{(t)}(\mathbf{z}) = \boldsymbol{\gamma}^{(t)} \odot \mathbf{z} + \boldsymbol{\beta}^{(t)}, \tag{6}$$

where $\boldsymbol{\gamma}^{(t)}$ and $\boldsymbol{\beta}^{(t)}$ denote per-frame gain and bias, respectively, and $\odot$ denotes element-wise multiplication. For the reference frame, we fix $\mathbf{A}^{(1)} = \mathbf{I}$. All learnable variables are optimized in the LR observation space. With

$$\Theta = \left\{\mathcal{G}, \mathbf{A}^{(t)}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\beta}^{(t)}\right\}_{t=1}^{T}, \tag{7}$$

the basic LR reconstruction objective is

$$\min_{\Theta} \mathcal{L}_{\mathrm{recon}} = \frac{1}{T|\Omega_{\mathrm{LR}}|} \sum_{t=1}^{T} \sum_{\mathbf{w}\in\Omega_{\mathrm{LR}}} \left\| \hat{I}_{\mathrm{LR}}^{(t)}(\mathbf{w}) - I_{\mathrm{LR}}^{(t)}(\mathbf{w}) \right\|_1, \tag{8}$$

This objective constrains the HR Gaussian field only through its consistency with the observed LR frames. A practical challenge in LR-only test-time optimization is that the HR Gaussian field is supervised only after geometric warping and area downsampling. Consequently, high-frequency structures in the rendered HR image can be strongly attenuated by the degradation operator and remain only weakly constrained by the LR reprojection objective. In our preliminary experiments, unconstrained anisotropic Gaussians may degenerate into extremely thin splats, producing oscillatory or grid-like artifacts even when the LR reprojection error continues to decrease.

Ringing-Aware Regularization. LR-space supervision may leave high-frequency oscillations in the rendered HR image weakly constrained. To suppress such artifacts while preserving monotonic edges, we introduce a sign-consistency gated second-order regularizer. Let

$$d_x^-(\mathbf{x}) = \hat{I}_{\mathrm{HR}}(\mathbf{x}) - \hat{I}_{\mathrm{HR}}(\mathbf{x}-\mathbf{e}_x), \qquad d_x^+(\mathbf{x}) = \hat{I}_{\mathrm{HR}}(\mathbf{x}+\mathbf{e}_x) - \hat{I}_{\mathrm{HR}}(\mathbf{x}), \tag{9}$$

and define the horizontal sign-consistency gate as

$$g_x(\mathbf{x}) = \sigma\left(-\alpha \frac{d_x^-(\mathbf{x}) d_x^+(\mathbf{x})}{|d_x^-(\mathbf{x})||d_x^+(\mathbf{x})| + \epsilon}\right), \tag{10}$$

where $\sigma(\cdot)$ is the sigmoid function. The vertical gate $g_y(\mathbf{x})$ is defined analogously. Our regularizer is

$$\mathcal{L}_{\mathrm{ring}} = \mathrm{mean}_{\mathbf{x}}\left[g_x(\mathbf{x})\left|\nabla_{xx}^2 \hat{I}_{\mathrm{HR}}(\mathbf{x})\right|_1\right] + \mathrm{mean}_{\mathbf{x}}\left[g_y(\mathbf{x})\left|\nabla_{yy}^2 \hat{I}_{\mathrm{HR}}(\mathbf{x})\right|_1\right]. \tag{11}$$

Adjacent differences have the same sign along monotonic edges, yielding a small gate, whereas sign reversals at local extrema activate the penalty. We use $\alpha = 6$ and $\epsilon = 10^{-8}$. This regularizer depends only on the rendered prediction and requires no HR supervision.

Uncertainty-Aware Reprojection Loss. Multi-frame observations may contain noise, blur, local occlusions, or photometric inconsistencies that cannot be fully explained by a single shared HR signal. To reduce the influence of unreliable LR measurements, MISR-GS optionally introduces a learnable per-frame log-variance map. We denote the log-variance map of frame t by $\ell^{(t)} : \Omega_{\mathrm{LR}} \to \mathbb{R}$. For the uncertainty-aware variant, we replace the standard ℓ_1 reprojection loss with an uncertainty-weighted reprojection objective:

$$\mathcal{L}_{\mathrm{unc}} = \frac{1}{2T|\Omega_{\mathrm{LR}}|} \sum_{t=1}^{T} \sum_{\mathbf{w}\in\Omega_{\mathrm{LR}}} \left[\left\| \hat{I}_{\mathrm{LR}}^{(t)}(\mathbf{w}) - I_{\mathrm{LR}}^{(t)}(\mathbf{w}) \right\|_2^2 \exp\left(-\ell^{(t)}(\mathbf{w})\right) + \ell^{(t)}(\mathbf{w})\right]. \tag{12}$$

This uncertainty-aware reprojection allows inconsistent pixels to contribute less to the optimization while keeping the reconstruction constrained by the observed LR measurements.

3.4 LR-Only Gaussian Initialization and Densification

Existing 2D Gaussian image representations typically allocate primitives using the target image, which is unavailable in MISR. We therefore derive both Gaussian initialization and progressive densification exclusively from multi-frame LR observations.

Observation-Driven Initialization. We first construct a pseudo-HR reference from the LR burst. Each LR frame is bicubically upsampled and inverse-warped to the reference coordinate system using an ECC-initialized transformation $\tilde{\mathbf{A}}^{(t)}$. ECC is used only for initialization; the frame-wise transformations remain learnable during test-time optimization. The aligned observations are fused by pixel-wise median:

$$I_{\mathrm{pHR}} = \mathrm{Median}_{t=1}^{T}\left[\mathcal{T}_{(\tilde{\mathbf{A}}^{(t)})^{-1}}\left(\mathcal{U}_s(I_{\mathrm{LR}}^{(t)})\right)\right], \tag{13}$$

where $\mathcal{U}_s$ denotes bicubic upsampling to the HR grid. I_{pHR} is used only for Gaussian initialization and never as a reconstruction target.

To allocate more primitives to structurally informative regions, we compute the Sobel gradient magnitude $G_{\mathrm{sob}}(\mathbf{x})$ of I_{pHR}, normalize it as $\hat{G}_{\mathrm{sob}}(\mathbf{x}) = G_{\mathrm{sob}}(\mathbf{x})/\max_{\mathbf{x}'} G_{\mathrm{sob}}(\mathbf{x}')$, and define

$$p^{\mathrm{init}}(\mathbf{x}) = (1-\lambda)\frac{\hat{G}_{\mathrm{sob}}(\mathbf{x})^2}{\sum_{\mathbf{x}'\in\Omega_{\mathrm{HR}}}\hat{G}_{\mathrm{sob}}(\mathbf{x}')^2} + \frac{\lambda}{|\Omega_{\mathrm{HR}}|}, \tag{14}$$

where $\lambda \in [0,1]$ balances structure-aware and uniform sampling. Gaussian centers are sampled without replacement according to p^{init}, with colors initialized from I_{pHR}, constant initial scales, and zero rotations.

Residual-Guided Densification. During optimization, additional Gaussians are allocated to regions that remain poorly explained by the current shared HR field. For each frame, we compute the LR reprojection residual $R_{\mathrm{LR}}^{(t)} = I_{\mathrm{LR}}^{(t)} - \hat{I}_{\mathrm{LR}}^{(t)}$, and lift it to the shared HR coordinate system:

$$\widetilde{R}^{(t)} = \mathcal{T}_{(\mathbf{A}^{(t)})^{-1}}\left(\mathcal{U}_s(R_{\mathrm{LR}}^{(t)})\right). \tag{15}$$

Let $M^{(t)}(\mathbf{x}) \in \{0,1\}$ indicate whether $\mathbf{x}$ receives a valid lifted observation from frame t. We aggregate the channel-averaged residual magnitude as

$$E(\mathbf{x}) = \frac{\sum_{t=1}^{T} M^{(t)}(\mathbf{x})\|\widetilde{R}^{(t)}(\mathbf{x})\|_{1,c}}{\sum_{t=1}^{T} M^{(t)}(\mathbf{x})}, \tag{16}$$

for locations supported by at least one valid frame. New Gaussian centers are sampled without replacement according to

$$p^{\mathrm{den}}(\mathbf{x}) = (1-\lambda)\frac{E(\mathbf{x})^2}{\sum_{\mathbf{x}'\in\Omega_{\mathrm{HR}}}E(\mathbf{x}')^2} + \frac{\lambda}{|\Omega_{\mathrm{HR}}|}, \tag{17}$$

which preferentially allocates representation capacity to persistently under-explained regions while retaining uniform spatial coverage. For color initialization, we separately preserve the residual sign:

$$\bar{R}(\mathbf{x}) = \frac{\sum_{t=1}^{T} M^{(t)}(\mathbf{x})\widetilde{R}^{(t)}(\mathbf{x})}{\sum_{t=1}^{T} M^{(t)}(\mathbf{x})}, \qquad \mathbf{c}_{\mathrm{new}} = \hat{I}_{\mathrm{HR}}(\boldsymbol{\mu}_{\mathrm{new}}) + \bar{R}(\boldsymbol{\mu}_{\mathrm{new}}). \tag{18}$$

New Gaussians use the same initial scale and rotation parameterization as the initial set.

4 Experimental Results and Discussion

4.1 Datasets

We evaluate MISR-GS on two MISR benchmarks covering satellite and handheld imaging scenarios. SatSynthBurst is a synthetic satellite burst dataset constructed from 20 WorldStrat HR images Cornebise et al. (2022). Each sample contains 16 LR frames with upsampling factors of $\times 2$, $\times 4$, and $\times 8$. SyntheticBurst Bhat et al. (2021) contains synthetic ground-level bursts with 14 LR frames and is provided natively at $\times 8$. We follow the standard evaluation protocols of each dataset, with detailed preprocessing and metric computation described in Appendix A.

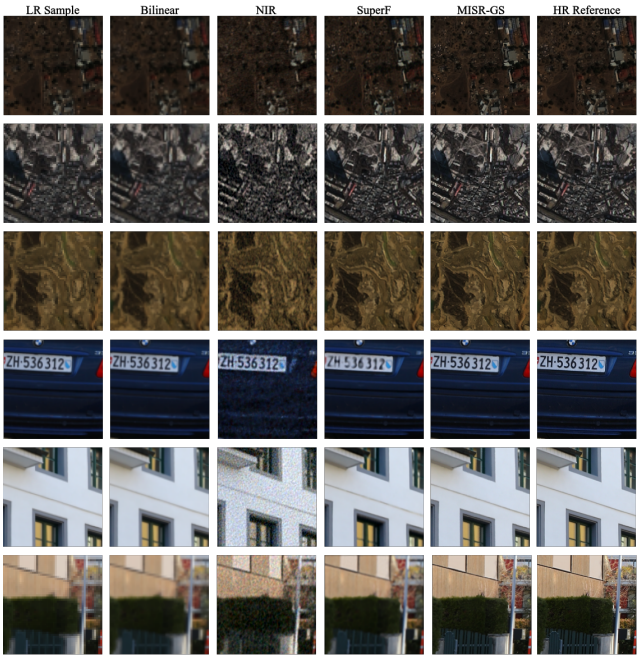

Figure 2: **Qualitative comparison with upsampling factor $\times$ 4**. From left to right, we show: one low-resolution (LR) frame, bilinear upsampling, NIR Nam et al. (2022), SuperF Jyhne et al. (2026), our MISR-GS approach, and the high-resolution (HR) reference. Samples are from SatSynthBurst (rows 1–3) and SyntheticBurst (rows 4–6). More results are provided in Appendix Figures 5 to 7.

4.2 Experimental Setup

We implement MISR-GS in PyTorch and conduct all experiments on a single NVIDIA RTX 4090 GPU with 24 GB of memory. We report PSNR, SSIM, and LPIPS against the HR ground truth, which is strictly reserved for evaluation and is never used during optimization or model selection. MISR-GS jointly optimizes the Gaussian field, frame-wise affine transformations, photometric parameters, and uncertainty maps using Adam. All available LR frames are used at each iteration, and each burst is optimized for at most $10{,}000$ iterations. We use a unified budget of $50{,}000$ Gaussians for all datasets, initialize 50% of them from an LR-derived pseudo-HR reference, and progressively allocate the remaining primitives using multi-frame LR reprojection residuals. Frame transformations are initialized using ECC pre-alignment and then jointly optimized without ground-truth transformation priors. The objective combines uncertainty-aware LR reprojection, an LR SSIM term, and the ringing-aware regularizer in Eq. 11. We use the same optimization and regularization hyperparameters for SatSynthBurst and SyntheticBurst. Learning-rate scheduling, early stopping, and checkpoint selection depend exclusively on full-burst LR reprojection metrics. Complete hyperparameters and stopping criteria are provided in Appendix B, while computational costs are reported in Appendix Table 5.

4.3 Comparison to existing test-time optimization approaches

Baseline approaches. Since MISR-GS follows the test-time optimization paradigm, we compare it with representative optimization-based MISR baselines. We first report the performance of simple interpolation methods, including bilinear or bicubic upsampling of the LR reference frame, which serve as non-learning lower bounds. We then compare with the steerable kernel regression method by Lafenetre et al. (2023), which adapts the handheld burst super-resolution approach described by Wronski et al. (2019) to multi-exposure satellite imagery and provides a strong learning-free TTO baseline. Most importantly, we compare MISR-GS with SuperF Jyhne et al. (2026), a recent INR-based test-time optimization method for MISR. SuperF represents the unknown HR image as a shared coordinate-based implicit neural field and jointly optimizes frame-wise affine alignment.

Table 1: Comparison with TTO baselines. PSNR (↑) for different upscaling factors. Standard deviations are given in parentheses. All methods use 16 LR frames on SatSynthBurst and 14 LR frames on SyntheticBurst.

Method	SatSynthBurst			SyntheticBurst		
	×2	×4	×8	×2	×4	×8
Bilinear	33.41 (3.41)	29.58 (3.62)	26.32 (3.73)	27.64 (3.55)	26.17 (3.74)	25.53 (3.82)
Lafenetre et al. (2023)	33.26 (3.58)	27.43 (3.84)	25.05 (3.71)	27.31 (3.36)	26.35 (3.15)	25.28 (3.02)
NIR Nam et al. (2022)	26.09 (3.87)	24.45 (4.32)	23.67 (3.94)	23.82 (4.45)	22.93 (4.41)	22.04 (4.38)
SuperF Jyhne et al. (2026)	35.36 (5.73)	31.72 (4.06)	26.39 (3.18)	26.18 (4.34)	25.64 (4.43)	24.67 (4.27)
MISR-GS (ours)	36.34 (1.76)	33.27 (1.95)	28.56 (2.34)	29.15 (3.48)	27.68 (3.92)	27.43 (3.97)

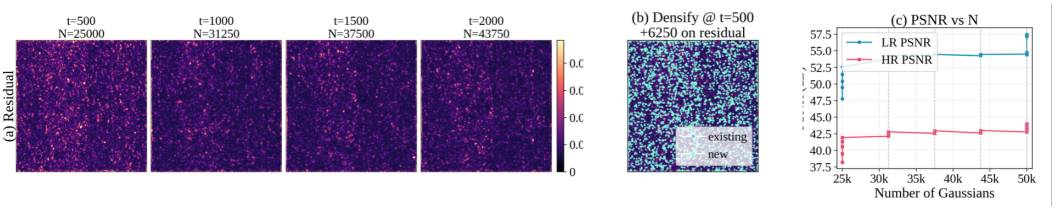

Figure 3: **Visualization of residual-guided Gaussian allocation.** The aggregated LR reprojection residuals and newly allocated Gaussian positions illustrate how the proposed rule directs additional capacity toward under-explained regions during optimization.

In contrast, MISR-GS uses a shared explicit 2D Gaussian field as the optimizable HR representation while keeping the reconstruction constrained by the same LR observation principle. This comparison evaluates MISR-GS as a complete explicit representation framework for LR-supervised test-time MISR against existing optimization-based approaches. For a fair comparison, all methods are evaluated on the same LR bursts, upsampling factors, and image regions. When running SuperF, we follow its official optimization setting as closely as possible, including frame-wise affine alignment and LR-space reconstruction supervision. For reference, we report both reconstruction quality and optimization efficiency whenever available.

Comparison results. We report PSNR results in Tab. 1 for upsampling factors ×2, ×4, and ×8. The baselines cover different MISR settings, including single-frame interpolation, learning-free optimization-based reconstruction Wronski et al. (2019); Lafenetre et al. (2023), INR-based image fusion Nam et al. (2022), and TTO-based MISR Jyhne et al. (2026). Across the evaluated datasets and scales, MISR-GS consistently achieves the highest PSNR among the reported methods, with gains of 0.98–2.97 dB over SuperF.

Qualitatively, optimization-based baselines can suppress noise but may also smooth fine structures in handheld and satellite bursts (Fig. 2 and Appendix Figs. 5–7). SuperF further exhibits color shifts and boundary blur in some examples. In contrast, MISR-GS better preserves local edges and repeated high-frequency patterns in the shown cases. These observations are provided only as qualitative evidence, while all quantitative comparisons are based on full-reference metrics.

4.4 Ablation Studies and Mechanism Analysis

We visualize the behavior of the observation-driven Gaussian allocation mechanism and quantify the contribution of the stabilization objectives. The visualization examines where additional Gaussian capacity is allocated, while the objective ablation evaluates the stabilization terms under a fixed representation pipeline.

Visualization of Residual-Guided Allocation. We visualize how the proposed allocation mechanism distributes newly introduced Gaussians during optimization. As shown in Fig. 3, new primitives are preferentially placed in regions with large aggregated LR reprojection residuals, consistent with the intended behavior of the allocation rule. This analysis illustrates the mechanism but does not isolate its contribution from that of progressive capacity growth.

Table 2: **Ablation of the reconstruction objective.** We compare the standard LR-space ℓ_1 reprojection loss with the uncertainty-aware reprojection loss $\mathcal{L}_{\text{unc}}$ and evaluate the contribution of the ringing-aware regularizer $\mathcal{L}_{\text{ring}}$. All remaining components are fixed. Results are reported at $\times 4$ using 16 LR frames for SatSynthBurst and 14 LR frames for SyntheticBurst. Standard deviations across samples are shown in parentheses.

Reprojection Loss	$\mathcal{L}_{\text{ring}}$	**SatSynthBurst**			**SyntheticBurst**		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
ℓ_1	✗	30.71 (2.56)	0.826 (0.218)	0.302 (0.067)	25.07 (3.81)	0.683 (0.121)	0.395 (0.047)
ℓ_1	✓	32.98 (1.57)	0.851 (0.126)	0.271 (0.052)	26.98 (3.81)	0.712 (0.121)	0.301 (0.070)
$\mathcal{L}_{\text{unc}}$	✓	33.27 (1.95)	0.864 (0.027)	0.257 (0.038)	27.68 (3.92)	0.776 (0.103)	0.286 (0.089)

Ablation of Stabilization Objectives. We study the effect of the optimization objectives while keeping the Gaussian representation, initialization, densification, alignment, and optimization settings fixed. Starting from the standard LR-space ℓ_1 reprojection loss, we first introduce the ringing-aware regularizer. We then replace the ℓ_1 reprojection loss with the uncertainty-aware reprojection loss $\mathcal{L}_{\text{unc}}$ while retaining the ringing-aware regularizer. The LR SSIM term is kept fixed across all variants. Table 2 reports the resulting reconstruction quality. A controlled comparison with fixed, uniform, and pseudo-HR-guided allocation strategies is provided in Appendix C.

4.5 Discussion

Limitations. MISR-GS has several limitations. First, as a test-time optimization method, it requires iterative optimization for each input burst, and its computational cost increases with the number of LR frames, Gaussian primitives, and optimization iterations. Second, the current formulation mainly assumes frame-wise affine motion and global photometric variation, which may be insufficient for strong parallax, non-rigid motion, occlusion, or rolling-shutter effects. Third, since the HR representation is supervised only through LR reprojection, some high-frequency structures remain weakly constrained by the render-and-degrade model despite our stabilization strategies. Finally, the Gaussian budget and several representation-related hyperparameters still need to be selected carefully, motivating future work on adaptive Gaussian allocation and more expressive motion and degradation models.

Impact on society. MISR-GS may broaden access to high-resolution image reconstruction by recovering fine spatial details from multiple low-resolution observations without requiring paired high-resolution training data. This capability can benefit applications where high-resolution acquisition is costly or difficult, including remote sensing, environmental observation, and handheld photography. Its test-time formulation also makes the method potentially adaptable to new imaging domains without collecting large task-specific training datasets. At the same time, super-resolved images should be interpreted with care, as reconstructed details may be over-interpreted as directly measured information, and improved imaging capability may also raise concerns related to privacy and surveillance.

5 Conclusion

We present MISR-GS, an explicit Gaussian representation framework for LR-supervised test-time MISR. By modeling the latent HR image with a shared anisotropic 2D Gaussian field, MISR-GS provides a spatially localized alternative to coordinate-based implicit representations while preserving the LR-only reconstruction principle. To enable explicit representation learning without HR supervision, MISR-GS derives Gaussian construction and refinement from the observed LR burst and optimizes the representation through a differentiable render-and-degrade model. Experiments on two MISR benchmarks demonstrate the effectiveness of explicit Gaussian fields for observation-constrained MISR, showing improved reconstruction quality over existing TTO approaches without paired LR-HR training data.

REFERENCES

Luca Savant Aira, Gabriele Facciolo, and Thibaud Ehret. Gaussian splatting for efficient satellite image photogrammetry. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5959–5969. IEEE, 2025.

Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9205–9214. IEEE, 2021.

Julien Cornebise, Ivan Oršolić, and Freddie Kalaitzis. Open high-resolution satellite imagery: The worldstrat dataset–with application to super-resolution. *Advances in Neural Information Processing Systems*, 35:25979–25991, 2022.

Yushuai Cui, Hongying Liu, Wenbin Zuo, Chaowei Fang, Fanhua Shang, Liang Wan, Lin Guo, and Ao Wei. Scale-aware transformer and efficient 2d gaussian splatting for arbitrary-scale super-resolution of multi-modal mri. *Pattern Recognition*, pp. 114696, 2026.

Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burstormer: Burst image restoration and enhancement transformer. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5703–5712. IEEE, 2023.

Michael Elad and Arie Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE transactions on image processing*, 6(12):1646–1658, 1997.

Ben Fei, Jingyi Xu, Rui Zhang, Qingyuan Zhou, Weidong Yang, and Ying He. 3d gaussian splatting as a new era: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 31(8): 4429–4449, 2024.

Jintong Hu, Bin Xia, Bin Chen, Wenming Yang, and Lei Zhang. Gaussiansr: High fidelity 2d gaussian splatting for arbitrary-scale image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3554–3562, 2025.

Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 53(3):231–239, 1991.

Sander R Jyhne, Christian Igel, Morten Goodwin, Per-Arne Andersen, Serge Belongie, and Nico Lang. Superf: Neural implicit fields for multi-image super-resolution. In *International Conference on Learning Representations*, volume 2026, pp. 99513–99536, 2026.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023. URL `https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/`.

Mahmoud M Khattab, Akram M Zeki, Ali A Alwan, and Ahmed S Badawy. Regularization-based multi-frame super-resolution: A systematic review. *Journal of King Saud University Computer and Information Sciences*, 32(7):755–762, 2020.

Jamy Lafenetre, Ngoc Long Nguyen, Gabriele Facciolo, and Thomas Eboli. Handheld burst super-resolution meets multi-exposure satellite imagery. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2056–2064. IEEE, 2023.

Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:305–319, 2018.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 105–114. IEEE, 2017.

Ziwei Luo, Youwei Li, Shen Cheng, Lei Yu, Qi Wu, Zhihong Wen, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Bsrt: Improving burst super-resolution with swin transformer and flow-guided deformable alignment. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 997–1007. IEEE, 2022.

Seonghyeon Nam, Marcus A Brubaker, and Michael S Brown. Neural image representations for multi-image fusion and layer separation. In *European conference on computer vision*, pp. 216–232. Springer, 2022.

Ngoc Long Nguyen, Jérémy Anger, Axel Davy, Pablo Arias, and Gabriele Facciolo. Self-supervised super-resolution for multi-exposure push-frame satellites. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1848–1858. IEEE, 2022.

Long Peng, Anran Wu, Wenbo Li, Xinjie Zhang, Xueyuan Dai, Xin Di, Haoze Sun, Renjing Pei, Yang Wang, Yang Cao, et al. Pixel to gaussian: Ultra-fast continuous super-resolution with 2d gaussian modeling. In *International Conference on Learning Representations*, volume 2026, pp. 92172–92199, 2026.

Roger Y Tsai and Thomas S Huang. Multiframe image restoration and registration. *Multiframe image restoration and registration*, 1:317–339, 1984.

Hao Wang, Ashish Bastola, Chaoyi Zhou, Wenhui Zhu, Xiwen Chen, Xuanzhao Dong, Siyu Huang, and Abolfazl Razi. Fast 2dgs: Efficient image representation with deep gaussian prior. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1184–1193, 2026.

Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM Transactions on Graphics (ToG)*, 38(4):1–18, 2019.

Guangpu Yang, Steffen Kieß, Hanxiang Luo, Xingyu Liu, and Sven Simon. Exact-gs: Mathematically rigorous and accurate 3d gaussian splatting for 3d x-ray reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4902–4911, 2026.

Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. In *European Conference on Computer Vision*, pp. 327–345. Springer, 2024.

Yulun Zhang, Kai Zhang, Zheng Chen, Yawei Li, Radu Timofte, Junpei Zhang, Kexin Zhang, Rui Peng, Yanbiao Ma, Licheng Jia, et al. Ntire 2023 challenge on image super-resolution ($\times$ 4): Methods and results. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1865–1884. IEEE, 2023.

Yunxiang Zhang, Bingxuan Li, Alexandr Kuznetsov, Akshay Jindal, Stavros Diolatzis, Kenneth Chen, Anton Sochenov, Anton Kaplanyan, and Qi Sun. Image-gs: Content-adaptive image representation via 2d gaussians. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2025.

Lingting Zhu, Guying Lin, Jinnan Chen, Xinjie Zhang, Zhenchao Jin, Zhao Wang, and Lequan Yu. Large images are gaussians: High-quality large image representation with levels of 2d gaussian splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10977–10985, 2025.

A Dataset Construction and Evaluation Protocol

A.1 Datasets

SatSynthBurst. We use SatSynthBurst for the satellite-image experiments and follow the construction protocol of SuperF Jyhne et al. (2026). The dataset is built from 20 HR images selected from WorldStrat Cornebise et al. (2022). Each HR image is converted into a burst of 16 LR observations at scale factors $\times 2$, $\times 4$, and $\times 8$, with the HR resolution kept fixed. The degradation model includes sub-pixel shifts, spectral variation, additive Gaussian noise, and MTF-based filtering following the Sentinel-2 setting described by Lanaras et al. Lanaras et al. (2018).

SyntheticBurst. For ground-level bursts, we use SyntheticBurst Bhat et al. (2021) with the same 50-sample subset adopted by SuperF Jyhne et al. (2026). Each sample contains 14 LR frames and is provided natively at $\times 8$. To evaluate $\times 2$ and $\times 4$, we downsample the HR reference to the corresponding target resolution and leave the LR frames unchanged. The same LR burst is therefore used across all three output scales.

A.2 Evaluation Protocol

We follow the SuperF evaluation protocol Jyhne et al. (2026) for spatial and spectral post-processing. SyntheticBurst does not provide an LR base frame that is spatially registered with the HR reference. After reconstruction, we therefore search for the best spatial alignment between the prediction and the HR image within a 4-pixel translation neighborhood and a rotation range of $[0^\circ, 4^\circ]$. This step is used only for metric computation and is applied in the same way to all compared methods. The HR reference is not used during test-time optimization.

Before computing PSNR, SSIM, and LPIPS, we crop 16 pixels from each image boundary. We also apply the same global spectral correction to all methods. This post-processing is separate from the frame-wise gain and bias parameters optimized by MISR-GS and is performed only after reconstruction.

B Implementation Details

Optimization. We use Adam with learning rates of 5×10^{-4}, 2×10^{-3}, 2×10^{-3}, and 5×10^{-3} for Gaussian positions, scales, rotations, and colors, respectively. The learning rates for the frame-wise affine parameters, photometric parameters, and log-variance maps are 10^{-2}, 10^{-3}, and 10^{-2}, respectively. All available LR frames are used at every iteration, with a maximum optimization budget of $10{,}000$ iterations per burst.

Initialization and Densification. The pseudo-HR reference is constructed by bicubically upsampling the LR frames, aligning them to the reference coordinate system using ECC, and applying pixel-wise median fusion. Gaussian positions are initialized using a mixture of pseudo-HR gradient-based sampling and 30% uniform sampling, while their colors are initialized from the pseudo-HR reference. We initialize $25{,}000$ of the final $50{,}000$ Gaussians and add the remaining primitives in four residual-guided densification stages at intervals of 500 iterations.

Loss Configuration. The LR SSIM loss is assigned a weight of 0.1. For the uncertainty-aware variant, $\mathcal{L}_{\text{unc}}$ replaces the standard ℓ_1 reprojection loss and is applied with unit weight, i.e., $\lambda_{\text{unc}} = 1.0$. The log-variance maps are initialized to -4.0 and clamped to $[-8, 4]$. We regularize them using an ℓ_2 penalty with weight 10^{-4} and a spatial smoothness penalty with weight 10^{-3}. The photometric gain and bias parameters are regularized with weight 10^{-4}. For the ringing-aware regularizer, we set $\alpha = 6$, $\epsilon = 10^{-8}$, and $\lambda_{\text{ring}} = 0.08$. The boundary smoothness term has weight 0.04. These values are shared across all datasets and upsampling factors. We do not use additional luminance or chromatic total-variation penalties, Gaussian scale or anisotropy constraints, or Gaussian prefiltering.

Learning-Rate Scheduling and Early Stopping. We evaluate full-burst LR reprojection PSNR and SSIM every 100 iterations. An update is considered a significant improvement only when LR PSNR increases by more than 0.1 dB and LR SSIM increases by more than 10^{-3} relative to their previously recorded best values. After $1{,}000$ iterations without such an improvement, all learning rates are divided by 10. Optimization terminates after a second $1{,}000$-iteration plateau or upon reaching the maximum iteration budget. The plateau schedule is activated only after the final Gaussian budget has been reached. The best checkpoint is selected solely by full-burst LR reprojection PSNR. HR ground truth and HR evaluation metrics are not used for learning-rate scheduling, early stopping, or checkpoint selection.

Table 3: **Ablation of Gaussian allocation strategies.** All variants use the same MISR-GS pipeline, optimization objective, training schedule, and final budget of $50{,}000$ Gaussians. Results are reported at $\times 4$ using 16 LR frames for SatSynthBurst and 14 LR frames for SyntheticBurst.

Allocation Strategy	SatSynthBurst			SyntheticBurst		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Fixed allocation	32.41	0.837	0.284	26.83	0.742	0.315
Progressive uniform	32.46	0.839	0.281	26.87	0.744	0.312
Pseudo-HR error-guided	32.84	0.852	0.268	27.24	0.761	0.297
Multi-frame LR residual (ours)	**33.27**	**0.864**	**0.257**	**27.68**	**0.776**	**0.286**

C Ablation of Gaussian Allocation

We conduct a controlled ablation to isolate the effect of Gaussian allocation. All variants use the same MISR-GS representation, render-and-degrade model, optimization objective, affine initialization, densification schedule, stopping criterion, and final budget of $50{,}000$ Gaussians. They differ only in how Gaussian capacity is introduced and where newly added primitives are placed.

The *fixed allocation* variant, inspired by the direct fixed-capacity optimization paradigm of GaussianImage Zhang et al. (2024), initializes all $50{,}000$ Gaussians before optimization and performs no subsequent densification. The *progressive uniform* variant starts from $25{,}000$ Gaussians and samples the positions of newly added primitives uniformly. The *pseudo-HR error-guided* variant adapts the target-error allocation principle of Image-GS Zhang et al. (2025) to our LR-only setting. Since the HR target used by Image-GS is unavailable, we compute its allocation score from the difference between the current reconstruction and the LR-derived pseudo-HR reference:

$$E_{\mathrm{pHR}}(\mathbf{x}) = \left\| I_{\mathrm{pHR}}(\mathbf{x}) - \hat{I}_{\mathrm{HR}}(\mathbf{x}) \right\|_{1,c}, \tag{19}$$

where $\|\cdot\|_{1,c}$ denotes the channel-averaged absolute error. The proposed *multi-frame LR residual* strategy instead aggregates reprojection residuals from all observed LR frames, as defined in Eq. 16. Both the pseudo-HR reference and the multi-frame residuals are derived exclusively from the LR burst; none of the allocation strategies uses the HR ground truth.

For all progressive variants, optimization starts with $25{,}000$ Gaussians. The remaining primitives are introduced in four stages at intervals of 500 iterations, with $6{,}250$ Gaussians added at each stage. This controlled setup separates the effect of progressive capacity growth from that of the spatial allocation signal.

Table 3 shows that progressive uniform allocation provides only marginal improvements over fixed allocation. On SatSynthBurst, it improves PSNR by 0.05 dB, while on SyntheticBurst the improvement is 0.04 dB. The similarly small changes in SSIM and LPIPS indicate that progressive capacity growth alone accounts for only a limited portion of the reconstruction gain.

Introducing a spatially informative allocation signal yields larger improvements. Compared with progressive uniform allocation, pseudo-HR error guidance improves PSNR by 0.38 dB on SatSynthBurst and 0.37 dB on SyntheticBurst, while also improving SSIM and LPIPS. This result indicates that allocating Gaussian capacity toward under-explained image regions is more effective than uniformly distributing newly introduced primitives.

The proposed multi-frame LR residual strategy achieves the best performance across all metrics and both datasets. Relative to pseudo-HR error guidance, it improves PSNR by 0.43 dB on SatSynthBurst and 0.44 dB on SyntheticBurst. It also improves SSIM by 0.012 and 0.015, and reduces LPIPS by 0.011 on both datasets. These results suggest that residuals aggregated directly from multiple LR observations provide a more informative allocation signal than the single fused pseudo-HR reference. Compared with fixed allocation, the complete strategy improves PSNR by 0.86 dB on SatSynthBurst and 0.85 dB on SyntheticBurst, supporting the effectiveness of observation-driven Gaussian capacity allocation for LR-supervised MISR.

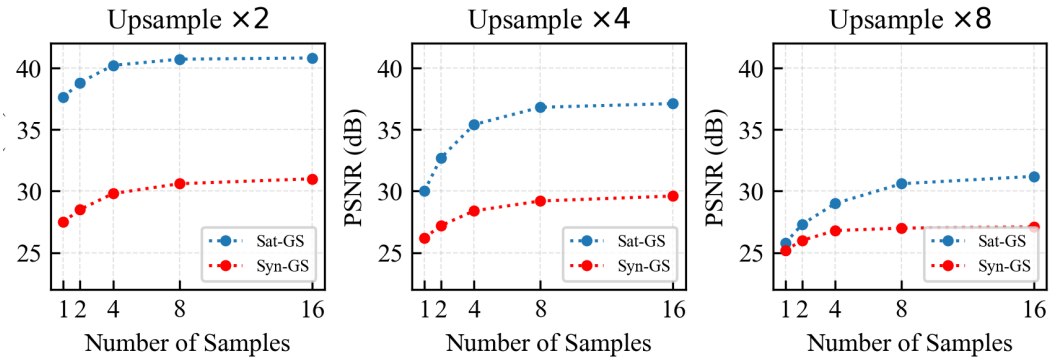

Figure 4: **Sensitivity to the number of LR frames.** From left to right, we report PSNR for upsampling factors 2, 4, and 8 by varying the number of LR frames on the horizontal axis.

D Additional Results

D.1 Qualitative results for different upsampling factors: $\times 2$, $\times 4$, $\times 8$

We provide additional qualitative comparisons for both satellite image and ground-level image bursts at upsampling factors $\times 2$ (Fig. 5), $\times 4$ (Fig. 6), $\times 8$ (Fig. 7).

D.2 Sensitivity to the number of LR frames

We study the sensitivity to the number of available LR frames in Fig. 4. This is a critical aspect for both application domains. Handheld bursts might be limited in the number of frames since the scene might change during long overall exposure times. Satellite imagery such as Sentinel-2 is captured with a revisit period of ≈ 5 days. Therefore, longer time windows are often required to obtain multi-frame satellite observations. However, longer time windows may introduce appearance changes due to ground activities or seasonal variations, which can hinder MISR. Furthermore, cloud-free images may be scarce depending on the geographic region.

For the SatSynthBurst dataset, we observe that the PSNR saturates with 8 samples for the $\times 2$ setting, while it continues to increase slightly when using 16 samples for larger upsampling factors. In contrast, PSNR on ground-level bursts continues improving with more frames even for the $\times 2$ setting. Additional frames may help reduce the high noise level in ground-level bursts.

D.3 Aligning large shifts without preprocessing

We further evaluate SuperF Jyhne et al. (2026) and MISR-GS under increasing inter-frame shifts without prealignment. In both methods, frame transformations are optimized jointly with the corresponding HR representation. As shown in Tab. 4, reconstruction quality decreases for both methods as the maximum shift increases from 1 to 4 LR pixels. MISR-GS exhibits a smaller PSNR degradation across all three upsampling factors, with the difference becoming most pronounced at $\times 8$, where the PSNR drops by 8.92 dB for MISR-GS compared with 11.78 dB for SuperF. Under the 4-pixel shift setting, MISR-GS outperforms SuperF by 1.29, 1.21, and 6.21 dB at $\times 2$, $\times 4$, and $\times 8$, respectively. These results indicate that MISR-GS is more robust to large inter-frame misalignment under the evaluated setting, although large shifts remain challenging at high upsampling factors.

E Use of Large Language Models

We used LLMs as search engines to support literature research and as coding assistants.

Table 4: **Comparing sub-pixel with large misalignments (PSNR).** We compare SuperF and MISR-GS on bursts with sub-pixel shifts, i.e. max shifts of 1.0 LR pixels, vs. bursts with large shifts up to 4.0 LR pixels. *Experimental setup*: upsampling factor $\times 2$, $\times 4$, $\times 8$; 16 LR frames. Standard deviation across samples shown in parentheses.

		SatSynthBurst		
		$\times 2$	$\times 4$	$\times 8$
	Bilinear	34.73 (3.65)	29.81 (3.88)	26.83 (4.03)
max shift: 1.0 LR pixels	SuperF	38.17 (2.01)	34.06 (2.90)	28.32 (3.14)
	MISR-GS	39.32 (1.63)	35.14 (2.47)	31.67 (2.82)
max shift: 4.0 LR pixels	SuperF	36.41 (2.93)	30.15 (3.57)	16.54 (7.22)
	MISR-GS	37.70 (2.48)	31.36 (3.12)	22.75 (6.81)

Table 5: **Computational cost.** We compare the per-instance optimization cost of SuperF and MISR-GS across different upsampling factors. Runtime and peak memory are measured on an NVIDIA RTX 4090 and averaged over five runs.

Factor	**Method**	**Time/Iter (ms)**	**Average Iters**	**Total Time (s)**	**Peak Memory (MB)**
$\times 2$	SuperF	6.32	2000	12.64	188.6
	MISR-GS	5.28	4066	21.47	652
$\times 4$	SuperF	9.95	2000	19.91	549
	MISR-GS	4.72	4139	19.54	613
$\times 8$	SuperF	24.93	2000	49.86	1863
	MISR-GS	4.61	3956	18.24	583
$\times 16$	SuperF	85.99	2000	171.99	7258
	MISR-GS	4.54	3966	18.01	568

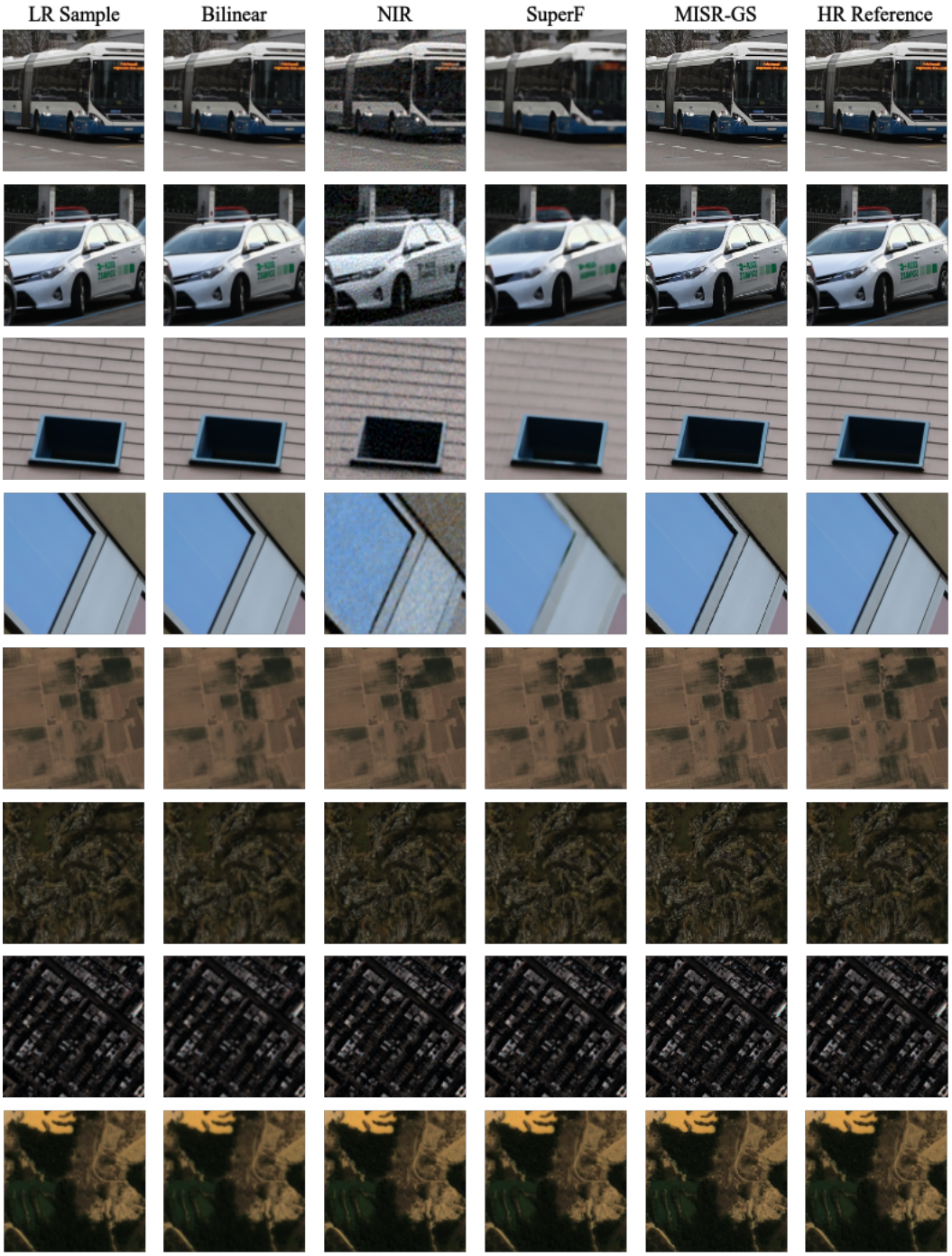

Figure 5: **Qualitative comparison with upsampling factor $\times$ 2**. From left to right, we show: one low-resolution (LR) frame, bilinear upsampling, steerable kernel regression NIR Nam et al. (2022), SuperF, our MISR-GS approach, and the high-resolution (HR) reference.

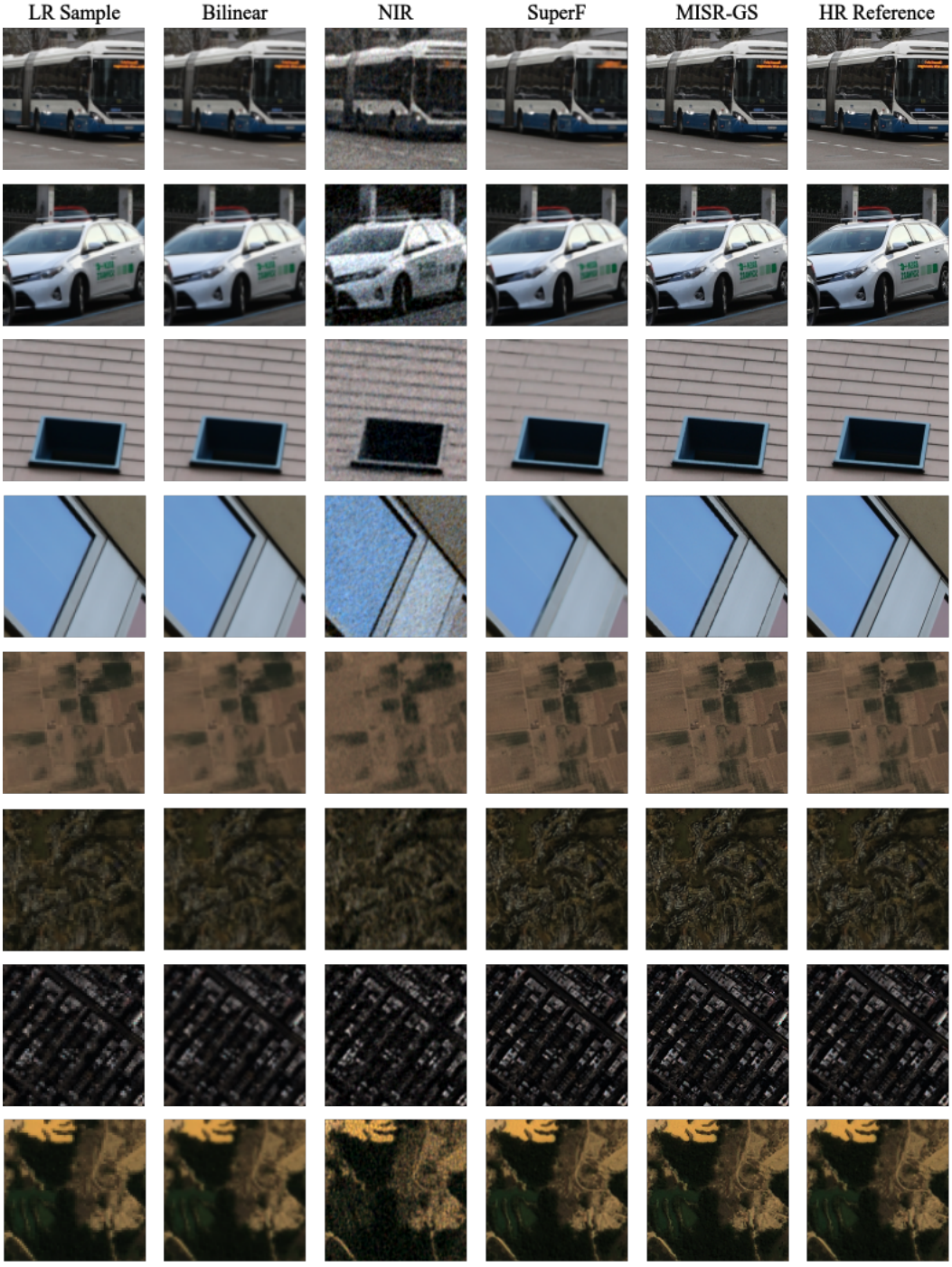

Figure 6: **Qualitative comparison with upsampling factor $\times$ 4**. From left to right, we show: one low-resolution (LR) frame, bilinear upsampling, steerable kernel regression NIR Nam et al. (2022), SuperF, our MISR-GS approach, and the high-resolution (HR) reference.

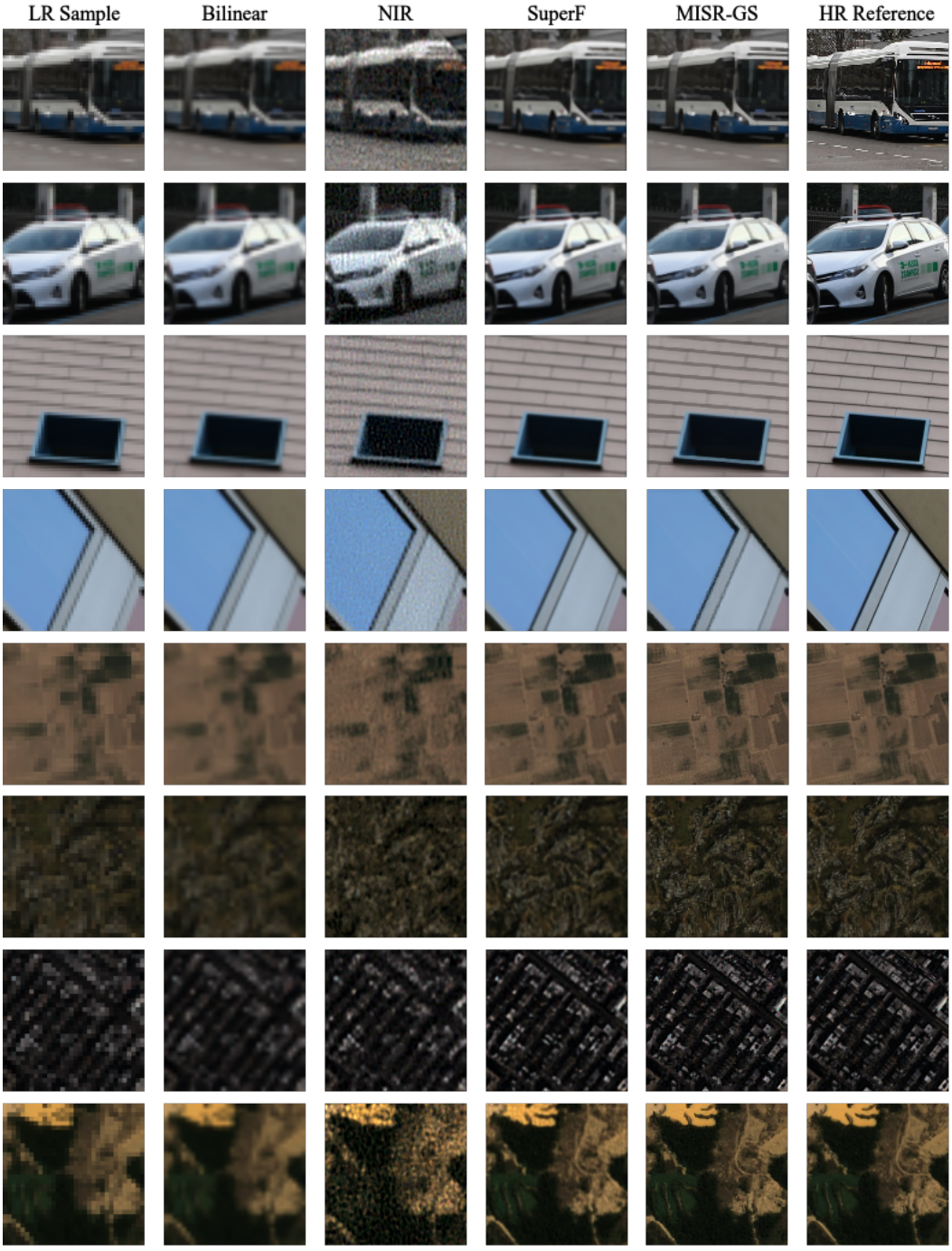

Figure 7: **Qualitative comparison with upsampling factor $\times$ 8**. From left to right, we show: one low-resolution (LR) frame, bilinear upsampling, steerable kernel regression NIR Nam et al. (2022), SuperF, our MISR-GS approach, and the high-resolution (HR) reference.