FluxFlow: Conservative Flow-Matching for Astronomical Image Super-Resolution

Shuhong Liu[1,2,✒☾], **Xining Ge**[2,✒], **Quanfeng Xu**[3,✒], **Ziteng Cui**[1], **Liuzhuozheng Li**[1]
Gengjia Chang[2], **Jun Liu**[2], **Ziying Gu**[1], **Dong Li**[2], **Xuangeng Chu**[1,2], **Lin Gu**[4]
Tatsuya Harada[1,5]

[1]The University of Tokyo [2]I2WM [3]SHAO [4]Tohoku University [5]RIKEN

Abstract

Ground-to-space astronomical super-resolution requires recovering space-quality images from ground-based observations that are simultaneously limited by pixel sampling resolution and atmospheric seeing, which imposes a stochastic, spatially varying PSF that cannot be resolved through upsampling alone. Existing methods rely on synthetic training pairs that fail to capture real atmospheric statistics and are prone to either over-smoothed reconstructions or hallucination sources with no physical counterpart in the observed sky. We propose FluxFlow, a conservative pixel-space flow-matching framework that incorporates observation uncertainty and source-region importance weights during training, and a training-free Wiener-regularized test-time correction to suppress hallucination sources while preserving recovered detail. We further construct the DESI–HST Dataset, the large-scale real-world benchmark comprising 19,500 real co-registered ground-to-space image pairs with real atmospheric PSF variation. Experiments demonstrate that FluxFlow consistently outperforms existing baseline methods in both photometric and scientific accuracy. Project webpage: https://i2wm.github.io/fluxflow/.

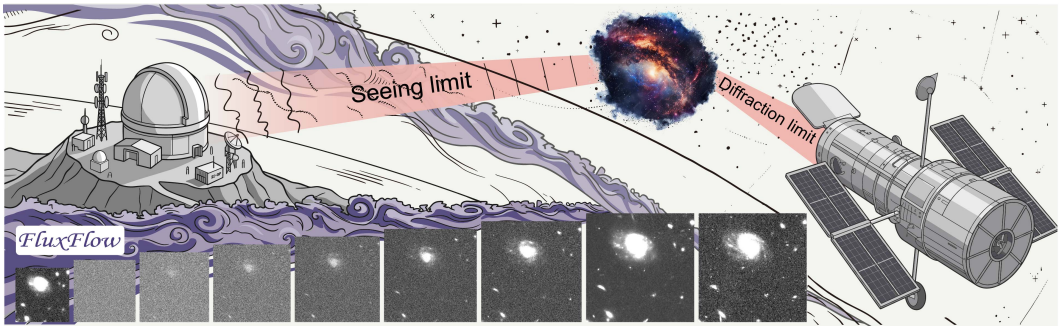


Figure 1: FluxFlow reconstructs HST-quality high-resolution images from seeing-limited low-resolution ground-based telescope observations.

1 Introduction

Image super-resolution (SR) has made substantial advances on natural images Liang et al. [2021], Zamir et al. [2022], Lipman et al. [2022], Chen et al. [2025b], Ren et al. [2026], where the degradation

✒These authors contribute equally to this work. ☾ Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

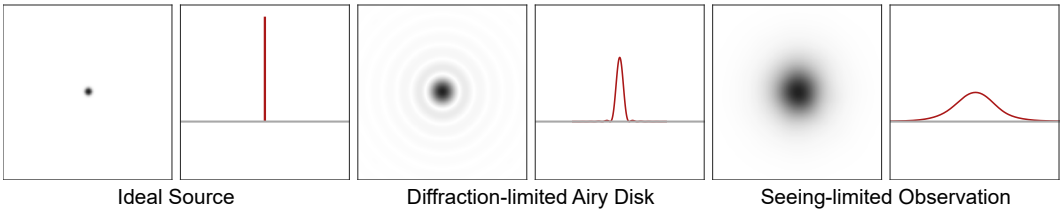


Figure 2: Illustration of point source degradation from ideal (left) to diffraction-limited (center) to seeing-limited (right), with 2D intensity images and 1D cross-sectional profiles.

model is dominated by limited sampling density on the image plane. In astronomical imaging, however, effective resolution is jointly governed by pixel sampling and angular resolution, where the latter is determined by the optical point spread function (PSF) and further degraded by atmospheric turbulence Sirianni et al. [2005], Dey et al. [2019]. Consequently, super-resolution that merely upsamples a PSF-blurred observation cannot recover the underlying finer physical structure.

This distinction is especially important in ground-to-space image super-resolution. As illustrated in Figure 2, Ground-based observations are typically seeing-limited, where atmospheric turbulence broadens the PSF and imposes an angular-resolution floor that often dominates over the detector sampling scale. In contrast, space-based telescopes such as the Hubble Space Telescope (HST) are free from atmospheric seeing and can approach the diffraction limit of their optics, yielding substantially sharper images. However, its small field of view and heavily oversubscribed observing time restrict such image quality to a small fraction of the sky, creating a persistent gap between wide-field ground-based surveys and high-resolution space-based observations.

Bridging this gap would enable high-resolution analysis over survey scales that are inaccessible to space telescopes alone. Wide-field surveys such as DESI Dey et al. [2019] image hundreds of millions of galaxies across thousands of square degrees, yet their seeing-limited image quality restricts measurements that depend on resolved structure, including weak-lensing shear estimation Mandelbaum et al. [2015], galaxy morphology Conselice [2014], and accurate PSF-aware photometry. A reliable mapping from ground-based observations to space-quality images would extend the reach of such analyses without requiring comparable space-based coverage.

Despite this potential, existing astronomical SR methods Miao et al. [2024], Shan et al. [2025], Wu et al. [2025], Zhang et al. [2026] are typically trained on synthetically degraded image pairs that do not fully reproduce the statistics of real observations, including atmospheric PSF variation, instrument-specific noise Liu et al. [2026], and spatially correlated artifacts introduced by frame co-addition Bertin [2010], Gonzaga et al. [2012]. This mismatch leads to a substantial domain gap when models are applied to genuine cross-instrument data. Regression-based methods further tend to produce overly smooth reconstructions that suppress fine morphological structure, while generative models, although better at recovering perceptual detail Saharia et al. [2022b], Wang et al. [2024], can introduce compact hallucination sources that can directly bias downstream scientific products, including source catalogs Bertin and Arnouts [1996], flux measurements Conselice [2014], photometric redshifts Salvato et al. [2019], and lensing analyses Bernstein and Jarvis [2002].

To address these challenges, we introduce FluxFlow, a flow-matching framework for ground-to-space astronomical super-resolution. Our method improves physical fidelity at both training and inference time. During training, it uses spatial reliability and source-importance maps to focus learning on scientifically focused regions while suppressing unstable gradients from artifacts. During inference, it enforces measurement consistency with the observed low-resolution image by back-projecting residuals through a Wiener-regularized inverse of the forward operator. Together, these designs suppress the generation of hallucination sources while preserving the fine structure. To overcome the scarcity of real observational data for ground-to-space SR, we further construct the DESI–HST Dataset, the large-scale astronomical SR dataset comprising real-world co-registered ground-to-space image pairs with authentic atmospheric PSF variation and observation uncertainty maps.

Our contributions are summarized as follows:

- We propose FluxFlow, a conservative pixel-space flow-matching framework with a physically informed training objective and a training-free Wiener-regularized inference-time correction that suppresses hallucination sources while preserving restored detail.
- We construct the DESI–HST Dataset, a real paired ground-to-space super-resolution benchmark containing 19,500 paired images at $\times 2$ and $\times 4$ scales, with real atmospheric PSF variation and per-pixel uncertainty weight maps.
- Experiments on our curated dataset demonstrate that FluxFlow achieves superior photometric fidelity and scientific accuracy over both recent regression-based and generative approaches.

2 Related Work

2.1 Regression-Based Image Restoration

Deep learning has driven substantial progress in single-image super-resolution and image restoration. CNN backbones such as MPRNet Zamir et al. [2021] and NAFNet Chen et al. [2022], Transformer variants including SwinIR Liang et al. [2021], Restormer Zamir et al. [2022], HAT Chen et al. [2023a, 2025b], SRFormer Zhou et al. [2023b], DAT Chen et al. [2023b], and HiT-SR Zhang et al. [2024a], and state-space models such as MambaIR Guo et al. [2024] and MaIR Li et al. [2025a] have steadily raised restoration quality through larger receptive fields and richer feature interaction. A parallel line pursues all-in-one and condition-aware restoration via degradation-adaptive Liang et al. [2022], prompt-based Potlapalli et al. [2023], Wang et al. [2023], Conde et al. [2024], and scale-adaptive He et al. [2025] designs, alongside efficiency-oriented work on quantization Qin et al. [2023], Fourier shifting Zhou [2024], and computational rebalancing Yu et al. [2024]. In astronomy, supervised pipelines address noise calibration Liu et al. [2026], X-ray data Sweere et al. [2022], galaxies Miao et al. [2024], star fields Wu et al. [2025], and cross-survey enhancement Luo et al. [2025], Li et al. [2022]. All these methods yield a deterministic output from the degraded input and cannot capture the ambiguity of severely ill-posed one-to-many inverse problems.

2.2 Diffusion Models for Conditional Generation

Diffusion models Song et al. [2020] serve as generative priors for restoration, with classifier Dhariwal and Nichol [2021], classifier-free Ho and Salimans [2022], latent Rombach et al. [2022], and cascaded Ho et al. [2022] variants enabling controllable synthesis Zheng et al. [2023b]. Restoration is cast as iterative refinement Saharia et al. [2022b,a], posterior sampling with measurement operators Chung et al. [2022], Song et al. [2023], or spatially conditioned generation Zhang et al. [2023], and is accelerated or constrained through residual shifting Yue et al. [2023], single-step generation Wang et al. [2024], Wu et al. [2024], Li et al. [2025b], Dong et al. [2025], consistency distillation He et al. [2024], Gong and Ma [2024], diffusion inversion Yue et al. [2025], faithfulness constraints Matsunaga et al. [2022], Zhang et al. [2025], Chen et al. [2025a], Polowczyk et al. [2026], and Schrödinger or Brownian bridge transport Liu et al. [2023], Li et al. [2023], Zhou et al. [2023a]. In astronomy, diffusion-based SR Ruan et al. [2025], Guo et al. [2026] targets gravitational lensing Reddy et al. [2024], solar imaging Song et al. [2024], galaxy reconstruction Shan et al. [2025], intensity interferometry Rai et al. [2025], and cross-survey generation Zhang et al. [2026].

3 DESI-HST Cross-Survey Dataset

Instruments and Data Collection. Unlike previous astronomical SR datasets that either synthesize LR images via degradation kernels Wu et al. [2025] or pair two ground-based surveys Miao et al. [2024], Luo et al. [2025], our dataset comprises genuinely observed ground-to-space image pairs from two independent instruments whose resolution gap is dominated by atmospheric turbulence. The LR images are drawn from the DESI Legacy Imaging Surveys Dey et al. [2019], a ground-based program providing i-band imaging ($\sim$7600–8730 Å) at $0.262''$/pixel [1] which offers the closest spectral match to the HR targets. The HR images are obtained from HST ACS/WFC in the F814W filter

[1] The symbols $'$ and $''$ denote arcminutes and arcseconds, respectively. One degree equals 60 arcminutes or 3600 arcseconds.



Figure 3: Paired examples from our DESI–HST dataset. Top: DESI cutouts at 128×128 pixels; Bottom: HST observations at 512×512 pixels. The image pairs are spatially aligned using World Coordinate System metadata. The DESI images are visibly blurred due to coarser angular resolution and atmospheric turbulence.

($\sim$7000–9500 Å), delivering diffraction-limited imaging at $0.03''$/pixel after drizzle stacking. [2]Details are provided in Appendix A.

Our dataset is collected from three extragalactic deep fields with overlapping HST and DESI coverage: COSMOS, GOODS-S, and UDS. Image pairs are spatially aligned using World Coordinate System metadata, with DESI coadds retained at native resolution to preserve their noise properties and PSF. The corresponding HST images are reprojected and flux-conservingly downsampled to produce aligned HR targets at $\times 2$ and $\times 4$ scale factors (HR at 256×256 and 512×512). The resulting LR–HR pairs are cropped into patches and split into 17,800 training and 1,700 test pairs per scale. Visualizations are shown in Figure 3 and alignment procedures are detailed in Appendix B.

Atmospheric Turbulence. Beyond pixel scale, the resolution gap between DESI and HST is dominated by atmospheric turbulence. Ground-based DESI observations are seeing-limited, with stochastic refractive-index fluctuations imposing a time-varying, spatially extended PSF of roughly $1.0''$–$1.3''$ FWHM in the DESI Imaging Surveys Dey et al. [2019]. HST operates above the atmosphere and is diffraction-limited, reaching a PSF FWHM of $\sim 0.09''$ Sirianni et al. [2005]. The resulting 11–$14\times$ ratio makes DESI-to-HST super-resolution substantially harder than previous datasets relying on pixel-scale downsampling alone.

Table 1: Comparison of astronomical SR datasets. **G** and **S** denote ground and space observations.[3]

Dataset	Size	Real	Flux	SR-Type
Miao et al. [2024]	2,000	✓	×	G→G
Shan et al. [2025]	9,383	✓	×	G→G
Song et al. [2024]	1,597	✓	×	S→G
Reddy et al. [2024][4]	2,880	✓	×	G→S
Luo et al. [2025]	17,000	✓	×	G→G
Li et al. [2022]	14,604	✓	×	G→G
Wu et al. [2025]	54,738	✓	✓	S→S
Zhang et al. [2026]	120,000	×	✓	S→G
Ours	19,500	✓	✓	G→S

4 Preliminaries of Flow-Matching

Flow Matching Lipman et al. [2022] learns a continuous-time normalizing flow that transports samples from a source distribution p_0 to a target distribution p_1 via an ordinary differential equation (ODE):

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = v_\theta(\mathbf{x}_t, t), \quad t \in [0, 1], \tag{1}$$

[2]For details of drizzle stacking, see Fruchter and Hook [2002] and Gonzaga et al. [2012]. We select the DESI i band and HST F814W because they are the closest available bandpasses between the two observing programs.

[3]Size counts image pairs. Real denotes observed pairs. Flux indicates whether linear photometric flux is preserved.

[4]Difflense Reddy et al. [2024]'s dataset was presented in the sRGB domain and not publicly available.

where v_θ is a time-dependent velocity field parameterized by a neural network. Given an initial sample $\mathbf{x}_0 \sim p_0$, integrating Equation (1) from t=0 to t=1 yields a sample $\mathbf{x}_1 \sim p_1$.

The central idea of Flow Matching is to define a conditional probability path $p_t(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1)$ that interpolates between the source and target distributions, together with its generating conditional velocity field $u_t(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1)$. We adopt the optimal-transport (OT) conditional path:

$$p_t(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1) = \mathcal{N}\big(\mathbf{x} \;\big|\; t\,\mathbf{x}_1 + (1-t)\,\mathbf{x}_0,\ \sigma_{\min}^2 \mathbf{I}\big), \tag{2}$$

where $\sigma_{\min}$ is a small constant. The conditional velocity field that generates this path takes the form:

$$u_t(\mathbf{x} \mid \mathbf{x}_0, \mathbf{x}_1) = \mathbf{x}_1 - \mathbf{x}_0. \tag{3}$$

The network v_θ is trained by minimizing the Conditional Flow Matching (CFM) objective Lipman et al. [2022], Tong et al. [2023]:

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{t\sim\mathcal{U}[0,1],\,\mathbf{x}_0\sim p_0,\,\mathbf{x}_1\sim p_1} \left\| v_\theta(\mathbf{x}_t, t) - u_t(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_1) \right\|^2, \tag{4}$$

where $\mathbf{x}_t = t\,\mathbf{x}_1 + (1-t)\,\mathbf{x}_0$ is the linearly interpolated sample. The CFM loss regresses against the conditional velocity u_t rather than the intractable marginal velocity field, yet provably yields the same gradient in expectation Lipman et al. [2022].

At inference, samples are generated by solving Equation (1) with a numerical ODE solver, i.e., Euler or Heun Karras et al. [2022], using N discretization steps from t=0 to t=1.

5 FluxFlow

Pixel-Domain Flow Matching. Modern generative models for natural images typically operate in the latent space of a VAE Rombach et al. [2022], which is ill-suited to scientific imaging. Astronomical frames are single-band and span a high dynamic range, whereas standard VAEs are trained on three-channel 8 bit photographs and cannot be transferred off-the-shelf. Retraining a tokenizer on astronomical data is also unappealing, since the compression itself degrades high-frequency structure of the targets and corrupts flux measurements. Motivated by recent studies on pixel-space diffusion that achieve high-resolution generation without a latent tokenizer Hoogeboom et al. [2023, 2025], Li and He [2025], we learn the velocity field v_θ directly on pixel inputs.

5.1 Physically-Informed Velocity Learning

While the conditional flow-matching objective in Equation (4) encourages perceptual realism, it does not guarantee that the generated images are physically consistent. Two quantities drive most downstream analysis. The integrated flux of a source encodes its luminosity and underpins stellar-mass and star-formation-rate estimates Conselice [2014], and the PSF shape determines apparent size for weak lensing and morphological classification Bernstein and Jarvis [2002], Mandelbaum et al. [2015]. We enforce both through spatially varying weights in the velocity objective that separately encode data reliability and scientific importance.

Inverse-Variance Weighting. The drizzled HST mosaic is accompanied by a per-pixel weight map (WHT) $\mathbf{W} \in \mathbb{R}^{H_{\text{hr}} \times W_{\text{hr}}}$ encoding the inverse variance at each pixel. It reflects spatially varying exposure depth from the dither pattern together with statistical noise and residual instrumental artifacts such as chip gaps and cosmic-ray trails, as shown in Figure 4. Pixels affected by these signatures receive lower weight. We compress the dynamic range with a square-root transform $\widetilde{\mathbf{W}} = \sqrt{\mathbf{W}}$ to suppress dominance by a few pixels.

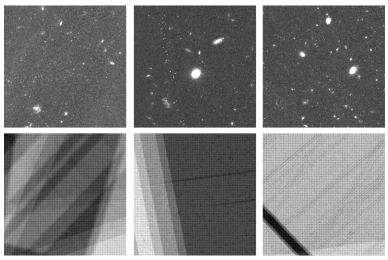


Figure 4: Visualization of HST science images (top) and WHT weight maps (bottom), showing structures from the drizzle stacking and local artifacts such as cosmic rays.

Source-Region Weighting. The inverse-variance map addresses data quality. Orthogonally, we introduce a scientific importance weight that flags which pixels matter most for analysis. Given a binary source mask $\mathbf{M}$ obtained from SExtractor Bertin and Arnouts [1996] run on the HST images,

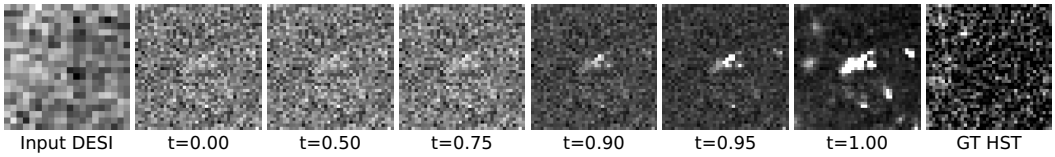


Figure 5: Sampling trajectory of a representative hallucination source across ODE integration steps. The spurious structure is already latent in the initial noise and progressively sharpened by the flow into a compact source-like morphology.

we construct $\mathbf{S} = 1 + \mathbf{M}$, which doubles the loss weight on source pixels relative to the background. We weight all pixels within a source uniformly because morphological and PSF-shape information resides primarily in the faint outer profile, which an intensity-dependent scheme would suppress.

Overall Objective. Multiplying the two maps yields a single spatially varying weight that encodes both reliability and importance, and is applied as a per-image normalized objective:

$$\mathcal{L}_{\mathrm{WFM}} \;=\; \mathbb{E}_{t,\,\mathbf{x}_0,\,\mathbf{x}_1}\left[\frac{\sum_p \big(\widetilde{\mathbf{W}} \odot \mathbf{S}\big)_p \,\big\|v_\theta(\mathbf{x}_t, t, \mathbf{c}) - u_t\big\|_p^2}{\sum_p \big(\widetilde{\mathbf{W}} \odot \mathbf{S}\big)_p}\right], \tag{5}$$

where $\mathbf{c}$ denotes the conditioning upsampled DESI observation specified in Appendix E, $\odot$ is element-wise multiplication, p indexes spatial pixel locations, and the denominator $\sum_p(\widetilde{\mathbf{W}} \odot \mathbf{S})_p$ is recomputed per training pair so that each pair contributes a gradient signal of comparable magnitude irrespective of the absolute scale of $\widetilde{\mathbf{W}} \odot \mathbf{S}$. In OT-CFM the velocity target $u_t = \mathbf{x}_1 - \mathbf{x}_0$ is independent of t and the interpolant $\mathbf{x}_t = (1-t)\,\mathbf{x}_0 + t\,\mathbf{x}_1$ is a pixelwise convex combination, so the mask $\mathbf{M}$ computed on $\mathbf{x}_1$ remains spatially aligned with source content at every t.

5.2 Measurement-Consistent Flow Sampling

The physics-informed weights improve velocity-field fidelity on source regions but leave the background unregulated. Unlike regression-based models Dong et al. [2015], Lim et al. [2017] that converge to the conditional mean, generative samplers can produce hallucinated sources with no counterpart in $\mathbf{y}$. We analyze their formation and propose a training-free, inference-time correction that enforces consistency with the observation.

Hallucination Sources. We track hallucinations along the full ODE trajectory. As shown in Figure 5, a hallucination is already present in the initial noise as a local intensity fluctuation, which the velocity field sharpens at late steps into a compact source-like morphology with surrounding ring artifacts. Figure 6 plots the mean brightness within a 2-pixel radius of 50 hallucination centroids as a function of t, nearly converging at t=1 that consistently exceed the HST . This behavior resists standard remedies such as classifier-free guidance Ho and Salimans [2022] and dense guidance Zheng et al. [2023a], Zhang et al. [2023]. The low SNR of ground-based observations makes hallucinations nearly indistinguishable from faint real sources in intermediate states $\mathbf{x}_t$[5], and once a hallucination forms, the velocity field self-reinforces it at every subsequent step.

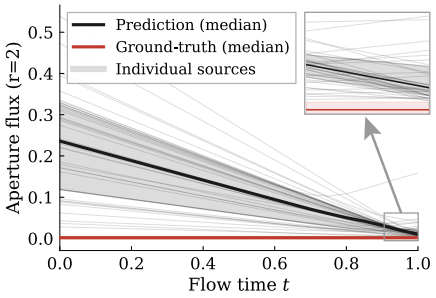


Figure 6: Aperture flux (sum) evolution of bright hallucinated sources along the ODE sampling trajectory. It starts from large-variance noise and converges to brightness levels exceeding the HST (red) at t=1.

Physical Forward-Model Consistency. Inspired by diffusion posterior sampling for inverse problems Chung et al. [2022], Song et al. [2023], we exploit

[5]We trained a U-Net to predict source masks from $\mathbf{x}_t$ on the DESI-HST dataset, obtaining an mIoU of only 11.3%.

Algorithm 1 MC-FS: Measurement-Consistent Flow Sampling

Require: v_θ, $\mathbf{y}$, h, N, η_0, SNR
$\mathbf{x}^{(0)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Precompute $W(\mathbf{k})$ via Equation (10)
for $i = 0, \ldots, N-1$ **do**
  $t_i \leftarrow i/N,\ \Delta t \leftarrow 1/N$
  $\tilde{\mathbf{x}}^{(i+1)} \leftarrow \mathbf{x}^{(i)} + v_\theta(\mathbf{x}^{(i)}, t_i, \mathbf{c})\,\Delta t$
  $\mathbf{r}^{(i)} \leftarrow \mathbf{A}(\tilde{\mathbf{x}}^{(i+1)}) - \mathbf{y}$
  $\eta_i \leftarrow \eta_0(1 - i/N)$
  $\mathbf{x}^{(i+1)} \leftarrow \tilde{\mathbf{x}}^{(i+1)} - \eta_i\,\mathcal{F}^{-1}\big(W \cdot \mathcal{F}[\mathbf{D}_s^\top \mathbf{r}^{(i)}]\big)$
end for
return $\mathbf{x}^{(N)}$

the known physical forward model of the imaging system to enforce measurement consistency at every step of the flow-matching ODE. The DESI observation $\mathbf{y}$ is related to the latent high-resolution image $\mathbf{x}$ by a degradation operator $\mathbf{A}$ consisting of convolution with the point spread function followed by spatial downsampling:

$$\mathbf{y} = \mathbf{A}(\mathbf{x}) + \boldsymbol{n}, \qquad \mathbf{A} = \mathbf{D}_s \circ \mathbf{H}_{\mathrm{PSF}}, \tag{6}$$

where $\mathbf{H}_{\mathrm{PSF}}$ denotes convolution with a constant Gaussian PSF kernel h of width σ_{PSF}, $\mathbf{D}_s$ is $s\times$ area downsampling, and $\boldsymbol{n}$ is observation noise. Re-projecting the prediction $\hat{\mathbf{x}}$ through $\mathbf{A}$ therefore helps suppress hallucinated sources by reducing residuals against $\mathbf{y}$. Concretely, at a discrete integration step i with current state $\mathbf{x}^{(i)}$, we first compute a candidate update $\tilde{\mathbf{x}}^{(i+1)}$ using the velocity field:

$$\tilde{\mathbf{x}}^{(i+1)} = \mathbf{x}^{(i)} + v_\theta\Big(\mathbf{x}^{(i)}, t_i, \mathbf{c}\Big)\,\Delta t, \tag{7}$$

and then measure the discrepancy between the candidate and the observation in the low-resolution pixel space:

$$\mathbf{r}^{(i)} = \mathbf{A}\Big(\tilde{\mathbf{x}}^{(i+1)}\Big) - \mathbf{y}. \tag{8}$$

This residual is then projected back to the high-resolution image via the adjoint operator $\mathbf{A}^\top$ and used to produce the corrected state $\mathbf{x}^{(i+1)}$ as:

$$\mathbf{x}^{(i+1)} = \tilde{\mathbf{x}}^{(i+1)} - \eta_i\,\mathbf{A}^\top\Big(\mathbf{r}^{(i)}\Big), \tag{9}$$

where η_i is a step-size schedule. Here, $\mathbf{D}_s^\top(\mathbf{y})[p,q] = \frac{1}{s^2}\,\mathbf{y}[\lfloor p/s \rfloor, \lfloor q/s \rfloor]$ is the strict inner-product adjoint of $s \times s$ area downsampling, and $\mathbf{H}_{\mathrm{PSF}}^\top = \mathbf{H}_{\mathrm{PSF}}$ because the Gaussian PSF is centro-symmetric. Both convolutions are realized as circular convolutions via FFT. However, naively applying Equation (9) in ODE sampling would accumulate PSF-scale smoothing across integration steps. This is because the adjoint $\mathbf{A}^\top = \mathbf{H}_{\mathrm{PSF}}^\top \circ \mathbf{D}_s^\top$ contains the same PSF convolution as the forward operator $\mathbf{A}$, so every correction step convolves the residual with the PSF kernel before adding it back to $\mathbf{x}^{(i+1)}$. With a finite number of sampling steps, slow recovery of high-frequency components can broaden the effective PSF of reconstructed sources. An example toy model is provided in Appendix F.

Wiener-Deconvolved Correction. To preserve the resolution of the high-resolution estimate while still enforcing measurement consistency, we instead back-project the residual through a Wiener correction for the rank-deficient $\mathbf{A}$, replacing the PSF adjoint in $\mathbf{A}^\top$ with a frequency-dependent reweighting. The correction signal is computed in the frequency domain using the 2D DFT $\mathcal{F}$ on the HR grid with circular boundary conditions, consistent with the FFT-based realization of $\mathbf{H}_{\mathrm{PSF}}$ above, and $\mathbf{k}$ indexes its discrete frequencies. Given the PSF transfer function $H(\mathbf{k}) = \mathcal{F}[h]$, we define the Wiener correction kernel:

$$W(\mathbf{k}) = \frac{H^*(\mathbf{k})}{|H(\mathbf{k})|^2 + \lambda_{\mathrm{SNR}}^{-1}}, \tag{10}$$

where H^* is the complex conjugate and λ_{SNR} is a regularization parameter controlling the aggressiveness of the deconvolution. When $\lambda_{\mathrm{SNR}} \to \infty$ the filter approaches the exact inverse H^{-1}; when $\lambda_{\mathrm{SNR}} \to 0$ it vanishes in magnitude but stays proportional to the matched filter $H^*(\mathbf{k})$ of Equation (9). The Wiener-corrected update then takes the form:

$$\mathbf{x}^{(i+1)} = \tilde{\mathbf{x}}^{(i+1)} - \eta_i\,\mathcal{F}^{-1}\Big(W(\mathbf{k}) \cdot \mathcal{F}\Big[\mathbf{D}_s^\top \mathbf{r}^{(i)}\Big]\Big), \tag{11}$$

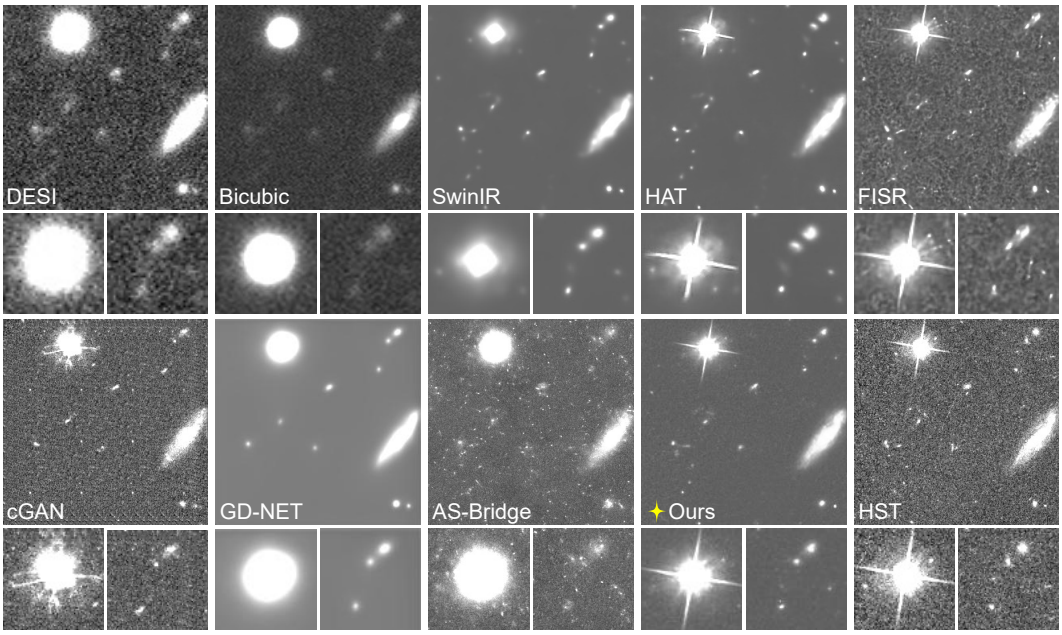


Figure 7: Qualitative comparison of baseline methods and FluxFlow on $\times 2$ super-resolution.

where W is defined on the HR grid and acts on the HR-lifted residual $\mathbf{D}_s^\top \mathbf{r}^{(i)}$ through the same transfer function $H(\mathbf{k})$. With these strict adjoints, Equation (9) becomes the exact gradient step on $\frac{1}{2}\|\mathbf{A}(\mathbf{x}) - \mathbf{y}\|^2$, and Equation (11) replaces $\mathbf{H}_{\mathrm{PSF}}^\top$ in this step with the Wiener kernel W. This helps reduce PSF broadening with a finite number of sampling steps while suppressing hallucinated sources. Structures in $\tilde{\mathbf{x}}^{(i+1)}$ that are inconsistent with $\mathbf{y}$ still produce a non-zero residual in Equation (8) and are therefore penalized. The step size η_i follows a linear decay schedule $\eta_i = \eta_0(1 - i/N)$, applying stronger corrections at early steps when the trajectory is far from convergence and relaxing the constraint at late steps to allow the flow to refine fine details. The full inference procedure is summarized in Algorithm 1.

6 Experiment

Implementation Details. We train our OT-CFM using the UNet backbone outlined in Appendix E. Optimization uses AdamW (lr = 1e-4, weight decay = 1e-4) for 300 epochs at batch size 32, with linear warm-up and cosine decay, and an EMA weights decay of 0.9999. In inference, we sample 10 Euler steps with MC-FS correction. Each $\hat{x}$ is forward-projected through a Gaussian PSF ($\sigma_{\mathrm{PSF}} = 2$ for both $\times 2$ and $\times 4$) and area-downsampling, and the residual against the DESI observation is Wiener-deconvolved ($\lambda_{\mathrm{SNR}} = 50$) and subtracted with $\eta_0 = 0.5$. The detailed model architecture of our flow model is provided in Appendix E. All methods are trained on a single A100-80G GPU.

Baseline Methods. We compare our method against conventional Bicubic upsampling, recent regression-based approaches, including SwinIR Liang et al. [2021], HAT Chen et al. [2025b], and FISR Wu et al. [2025], as well as generative approaches, including cGAN Rai et al. [2025], GD-Net Shan et al. [2025], and AS-Bridge Zhang et al. [2026]. These baselines span interpolation, direct regression, and generative reconstruction, allowing us to compare different approaches to recovering high-resolution source structure.

Evaluation Metric. We adopt PSNR and SSIM to evaluate photometric fidelity. To assess scientific accuracy, we extract source ellipsoid masks from the HST GT using SExtractor and compute the averaged Flux-L1 error, which measures the absolute flux deviation within source regions. PSNR and SSIM summarize image-level reconstruction quality, while Flux-L1 focuses on photometric errors within the detected sources. Additional experiments on AstroSR Miao et al. [2024] and STAR Wu et al. [2025] are provided in Appendix G.

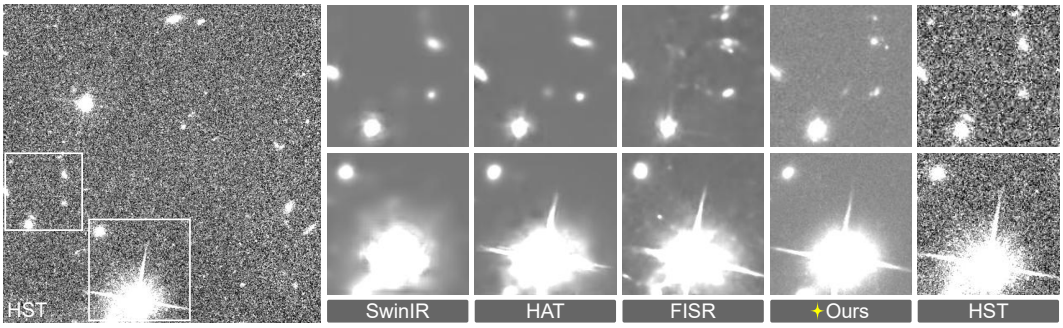


Figure 8: Qualitative comparison of baseline methods and FluxFlow on $\times 4$ super-resolution.

Table 2: Quantitative comparison at $\times 2$ and $\times 4$ on our DESI–HST super-resolution dataset.

Scale	Metric	Bicubic	SwinIR	HAT	FISR	cGAN	GD-Net	AS-Bridge	**Ours**
$\times 2$	PSNR↑	26.69	31.22	31.31	30.93	28.88	28.82	28.38	31.23
	SSIM↑	0.594	0.702	0.711	0.704	0.579	0.692	0.626	0.714
	Flux-L1↓	11.251	4.136	3.695	3.526	4.502	3.639	5.151	2.959
$\times 4$	PSNR↑	22.74	29.15	29.21	29.02	26.48	24.64	24.72	29.14
	SSIM↑	0.415	0.569	0.570	0.568	0.477	0.496	0.514	0.570
	Flux-L1↓	18.08	4.939	4.168	4.082	5.479	4.599	6.542	3.755

6.1 Quantitative and Qualitative Evaluation

Table 2 shows quantitative results at $\times 2$ and $\times 4$ on our DESI–HST dataset. FluxFlow obtains PSNR within 0.1 dB of the highest value at both scales, while achieving the best SSIM and Flux-L1 among baseline methods. PSNR and SSIM evaluate reconstruction fidelity, while Flux-L1 measures the absolute flux error within source regions. Figure 7 presents qualitative comparisons at $\times 2$. Regression methods tend to oversmooth source cores and extended galaxy structures, whereas FluxFlow preserves these details and suppresses hallucinated sources in regions with low signal to noise.

7 Ablation Study

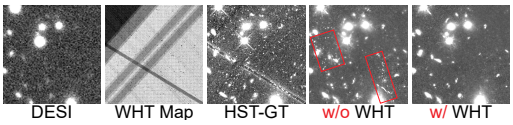


Figure 9: WHT weighting ablation. Defects (cosmic rays, drizzle gaps) in red boxes.

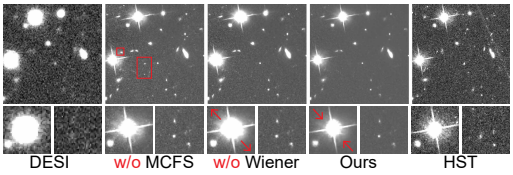


Figure 10: MC-FS and Wiener-deconvolved correction ablation. Hallucinations in red boxes.

Table 3: Ablation on photometry and detection metrics. *WHT* denotes inverse-variance weighting, *Src* denotes source-weighted loss, and MC-FS and Wiener correction are described in section 5.2.

Ablation	Photometry PSNR↑	SSIM↑	FluxL1↓	Detection Prec↑	Rec↑	F1↑
w/ VAE	28.90	0.593	5.075	0.859	0.211	0.3402
w/ Bridge	30.77	0.663	3.206	0.702	0.393	0.5019
w/o WHT	30.17	0.685	2.814	0.688	0.483	0.5272
w/o Src	31.16	0.711	3.098	0.729	0.464	0.5382
w/o MCFS	30.78	0.698	3.025	0.602	0.553	0.5256
w/o Wiener	30.79	0.699	3.031	0.708	0.448	0.5066
Ours	31.23	0.714	2.959	0.738	0.458	0.5412

Inverse-Variance and Source-Region Weighting. Figure 9 visualizes a region containing a cosmic-ray trail and a drizzle chip-gap in the HST . Without the WHT term defined in Equation (5), the velocity field treats these defects as valid targets and reproduces them in the reconstruction. The WHT-weighted variant suppresses their gradients and yields a clean output, and Table 3 reports the corresponding drops in PSNR, F1, and precision.The marginally lower Flux-L1 of w/o WHT is an acceptable trade-off, since it arises from fitting drizzle artifacts that overlap with source apertures.The w/o source-weight variant similarly degrades, as a uniform loss is dominated by background pixels and cannot preferentially preserve the faint outer galaxy profiles that carry morphological and PSF-shape

information. The two weights are complementary, with WHT filtering data quality and source-weight emphasizing scientifically informative regions.

Measurement-Consistent Sampling. We ablate MC-FS and the Wiener-regularization separately to isolate their roles. Without MC-FS, the sampler produces compact hallucination sources shown in Figure 10, and Table 3 reports the highest recall together with a precision collapse, as these spurious peaks inflate false positives and pull F1 below the full model. The w/o Wiener variant replaces the Wiener correction of Equation (11) with the raw adjoint of Equation (9) and visibly broadens the reconstructed PSF via the cumulative smoothing of Appendix F.2. The broadened response smears source flux beyond the target apertures into the background, and Table 3 reports a uniform photometric degradation across PSNR, SSIM, and Flux-L1. Our full model attains the strongest photometric scores together with the best detection F1, confirming that both inference-time components are necessary. Additional ablation studies on latent-space and bridge-based OT-CFM, hyperparameters, and visualizations of detection-outcome are provided in Appendix H and Appendix J.

8 Conclusion

We present FluxFlow, a conservative flow-matching framework for ground-to-space astronomical super-resolution. The training objective combines inverse-variance and source-region weighting, while measurement-consistent sampling applies Wiener correction under an approximate forward model. The DESI–HST dataset pairs independent ground- and space-based observations. Experiments show that FluxFlow delivers superior photometric and scientific accuracy over existing baselines.

Limitation. Hallucinations are suppressed rather than eliminated. The HST training targets contain residual noise, instrumental defects, optical diffraction patterns, and differences caused by the bandpasses, so reconstructions can inherit these features. The fixed isotropic Gaussian PSF does not capture variation across exposures, positions, or wavelengths, nor does it model anisotropy or extended wings. Future work will investigate generalization to other fields and instruments, source discovery, and calibrated uncertainty.

Acknowledgements and disclosures

This work was partially supported by JST Moonshot R&D Grant Number JPMJPS2011 and by the New Energy and Industrial Technology Development Organization (NEDO) under project JPNP26007. Ziteng Cui was supported by JSPS KAKENHI Grant Number JP26K21296. Shuhong Liu was supported by JST BOOST, Japan Grant Number JPMJBS2418.

References

sd-vae-ft-mse. https://huggingface.co/stabilityai/sd-vae-ft-mse, 2022.

Hiroaki Aihara, Nobuo Arimoto, Robert Armstrong, Stéphane Arnouts, Neta A Bahcall, Steven Bickerton, James Bosch, Kevin Bundy, Peter L Capak, James HH Chan, et al. The hyper suprime-cam ssp survey: overview and survey design. *Publications of the Astronomical Society of Japan*, 70(SP1):S4, 2018.

H. W. Babcock. The possibility of compensating astronomical seeing. *Publications of the Astronomical Society of the Pacific*, 65:229–236, 1953. doi: 10.1086/126606.

Kyle Barbary. Sep: Source extractor as a library. *Journal of Open Source Software*, 1(6):58, 2016.

Jacques M. Beckers. Adaptive optics for astronomy: Principles, performance, and applications. *Annual Review of Astronomy and Astrophysics*, 31:13–62, 1993. doi: 10.1146/annurev.aa.31.090193.000305.

GM Bernstein and M Jarvis. Shapes and shears, stars and smears: optimal measurements for weak lensing. *The Astronomical Journal*, 123(2):583–618, 2002.

Emmanuel Bertin. Swarp: resampling and co-adding fits images together. *Astrophysics Source Code Library*, pages ascl–1010, 2010.

Emmanuel Bertin and Stephane Arnouts. Sextractor: Software for source extraction. *Astronomy and Astrophysics Supplement Series*, 117(2):393–404, 1996.

Junyang Chen, Jinshan Pan, and Jiangxin Dong. Faithdiff: Unleashing diffusion priors for faithful image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28188–28197, 2025a.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33. Springer, 2022.

Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22367–22377, 2023a.

Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao Dong. Hat: Hybrid attention transformer for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025b.

Yubin Chen, Chengyue Gong, Weihang Guo, Xiaoyu Li, Zhenmei Shi, and Zhao Song. Provably efficient high-order flow matching in pixel space, 2025c. URL https://openreview.net/forum?id=wbIYlFCPPq.

Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12312–12321, 2023b.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2024.

Christopher J Conselice. The evolution of galaxy structure over cosmic time. *Annual Review of Astronomy and Astrophysics*, 52(1):291–337, 2014.

Arjun Dey, David J Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, et al. Overview of the desi legacy imaging surveys. *The Astronomical Journal*, 157(5):168, 2019.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

P Kingma Diederik and Welling Max. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2): 295–307, 2015.

Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23174–23184, 2025.

David L. Fried. Optical resolution through a randomly inhomogeneous medium for very long and very short exposures. *Journal of the Optical Society of America*, 56:1372–1379, 1966. doi: 10.1364/JOSA.56.001372.

Andrew S. Fruchter and Richard N. Hook. Drizzle: A method for the linear reconstruction of undersampled images. *Publications of the Astronomical Society of the Pacific*, 114:144–152, 2002. doi: 10.1086/338393.

Anne Gagneux, Ségolène Martin, Rémi Gribonval, and Mathurin Massias. Training flow matching: The role of weighting and parameterization. *arXiv preprint arXiv:2603.06454*, 2026.

Xiaoxuan Gong and Jie Ma. Ir-cm: The fast and general-purpose image restoration method based on consistency model. *Advances in Neural Information Processing Systems*, 37:75559–75584, 2024.

S Gonzaga, W Hack, A Fruchter, and J e Mack. The drizzlepac handbook. *The DrizzlePac Handbook*, 2012.

Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision*, pages 222–241. Springer, 2024.

Yuduo Guo, Peixin Weng, Hao Zhang, Chen Zhu, Jun Zhang, Jiamin Wu, and Qionghai Dai. Computational imaging towards next-generation optical observational astronomy. *Photonics Insights*, 5(1):R04–R04, 2026.

Guande He, Kaiwen Zheng, Jianfei Chen, Fan Bao, and Jun Zhu. Consistency diffusion bridge models. *Advances in Neural Information Processing Systems*, 37:23516–23548, 2024.

Xuyi He, Yuhui Quan, Ruotao Xu, and Hui Ji. A universal scale-adaptive deformable transformer for image restoration across diverse artifacts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12731–12741, 2025.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 23(47):1–33, 2022.

Emiel Hoogeboom, Jonathan Heek, and Tim Salimans. simple diffusion: End-to-end diffusion for high resolution images. In *International Conference on Machine Learning*, pages 13213–13232. PMLR, 2023.

Emiel Hoogeboom, Thomas Mensink, Jonathan Heek, Kay Lamerigts, Ruiqi Gao, and Tim Salimans. Simpler diffusion: 1.5 fid on imagenet512 with pixel-space diffusion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 18062–18071, 2025.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In *International Conference on Machine Learning*, pages 9483–9505. PMLR, 2022.

Ulugbek S Kamilov, Charles A Bouman, Gregery T Buzzard, and Brendt Wohlberg. Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1):85–97, 2023.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35:26565–26577, 2022.

Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. FlowDPS: Flow-driven posterior sampling for inverse problems. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025. URL https://arxiv.org/abs/2503.08136.

Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdm: Image-to-image translation with brownian bridge diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1952–1961, 2023.

Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, and Xi Peng. Mair: A locality-and continuity-preserving mamba for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7491–7501, 2025a.

Jianze Li, Jiezhang Cao, Yong Guo, Wenbo Li, and Yulun Zhang. One diffusion step to real-world super-resolution via flow trajectory distillation. *arXiv preprint arXiv:2502.01993*, 2025b.

Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. *arXiv preprint arXiv:2511.13720*, 2025.

Wanjun Li, Zhe Liu, and Hongtao Deng. A self-attention based srgan for super-resolution of astronomical image. In *2022 IEEE 8th International Conference on Computer and Communications*, pages 1977–1981. IEEE, 2022.

Jie Liang, Hui Zeng, and Lei Zhang. Efficient and degradation-adaptive network for real-world image super-resolution. In *European Conference on Computer Vision*, pages 574–591. Springer, 2022.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I^2SB: Image-to-image Schrödinger bridge. *arXiv preprint arXiv:2302.05872*, 2023.

Shuhong Liu, Xining Ge, Ziying Gu, Quanfeng Xu, Lin Gu, Ziteng Cui, Xuangeng Chu, Jun Liu, Dong Li, and Tatsuya Harada. Denoising the deep sky: physics-based ccd noise formation for astronomical imaging. *arXiv preprint arXiv:2601.23276*, 2026.

Zhijian Luo, Shaohua Zhang, Jianzhen Chen, Zhu Chen, Liping Fu, Hubing Xiao, Wei Du, and Chenggang Shu. Cross-survey image transformation: Enhancing sdss and decals images to near-hsc quality for advanced astronomical analysis. *The Astrophysical Journal Supplement Series*, 277(1): 22, 2025.

Rachel Mandelbaum, Barnaby Rowe, Robert Armstrong, Deborah Bard, Emmanuel Bertin, James Bosch, Dominique Boutigny, Frederic Courbin, William A Dawson, Annamaria Donnarumma, et al. Great3 results–i. systematic errors in shear estimation and the impact of real galaxy morphology. *Monthly Notices of the Royal Astronomical Society*, 450(3):2963–3007, 2015.

Naoki Matsunaga, Masato Ishii, Akio Hayakawa, Kenji Suzuki, and Takuya Narihira. Fine-grained image editing by pixel-wise guidance using diffusion models. *arXiv preprint arXiv:2212.02024*, 2022.

Jiawei Miao, Liangping Tu, Bin Jiang, Xiangru Li, and Bo Qiu. Astrosr: A data set of galaxy images for astronomical superresolution research. *The Astrophysical Journal Supplement Series*, 274(1):7, 2024.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Alicja Polowczyk, Agnieszka Polowczyk, Piotr Borycki, Joanna Waczyńska, Jacek Tabor, and Przemysław Spurek. Diamond: Directed inference for artifact mitigation in flow matching models. *arXiv preprint arXiv:2602.00883*, 2026.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023.

Mehrsa Pourya, Bassam El Rawas, and Michael Unser. Flower: A flow-matching solver for inverse problems. *arXiv preprint arXiv:2509.26287*, 2025.

Haotong Qin, Yulun Zhang, Yifu Ding, Xianglong Liu, Martin Danelljan, Fisher Yu, et al. Quantsr: accurate low-bit quantization for efficient image super-resolution. *Advances in Neural Information Processing Systems*, 36:56838–56848, 2023.

Km Nitu Rai, Yuri van der Burg, Soumen Basak, Prasenjit Saha, and Subrata Sarangi. Generative ai for image reconstruction in intensity interferometry: a first attempt. *arXiv preprint arXiv:2508.03398*, 2025.

Pranath Reddy, Michael W Toomey, Hanna Parul, and Sergei Gleyzer. Difflense: a conditional diffusion model for super-resolution of gravitational lensing data. *Machine Learning: Science and Technology*, 5(3):035076, 2024.

Bin Ren, Hang Guo, Yan Shu, Jiaqi Ma, Ziteng Cui, Shuhong Liu, Guofeng Mei, Lei Sun, Zongwei Wu, Fahad Shahbaz Khan, Salman Khan, Radu Timofte, Yawei Li, et al. The eleventh NTIRE 2026 efficient super-resolution challenge report. In *Proceedings of the Computer Vision and Pattern Recognition Conference Workshops*, 2026.

F. Roddier. The effects of atmospheric turbulence in optical astronomy. In *Progress in Optics*, volume 19, pages 281–376. Elsevier, 1981. doi: 10.1016/S0079-6638(08)70204-X.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM journal on imaging sciences*, 10(4):1804–1844, 2017.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

Hengxin Ruan, Qiufan Lin, Shupei Chen, Yang Wang, and Wei Zhang. Investigation on deep learning-based galaxy image translation models. *Astronomy & Astrophysics*, 701:A82, 2025.

Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022a.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022b.

Mara Salvato, Olivier Ilbert, and Ben Hoyle. The many flavours of photometric redshifts. *Nature Astronomy*, 3(3):212–222, 2019.

Matt L Sampson, Peter Melchior, Charlotte Ward, and Sufia Birmingham. Score-matching neural networks for improved multi-band source separation. *Astronomy and Computing*, 49:100875, 2024.

Qianqian Shan, Cuixiang Liu, Bo Qiu, Li Luo, Fuji Ren, Zhiren Pan, and Yicong Chen. Galaxy image super-resolution reconstruction using diffusion network. *Engineering Applications of Artificial Intelligence*, 142:109836, 2025.

M Sirianni, MJ Jee, N Benítez, JP Blakeslee, AR Martel, Gerhardt Meurer, M Clampin, G De Marchi, HC Ford, R Gilliland, et al. The photometric performance and calibration of the hubble space telescope advanced camera for surveys. *Publications of the Astronomical Society of the Pacific*, 117(836):1049–1112, 2005.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023.

Wei Song, Ying Ma, Haoying Sun, Xiaobing Zhao, and Ganghua Lin. Improving the spatial resolution of solar images using super-resolution diffusion generative adversarial networks. *Astronomy & Astrophysics*, 686:A272, 2024.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Robin Swanson, Esther Y. H. Lin, Masen Lamb, Suresh Sivanandam, and Kiriakos N. Kutulakos. Super resolved imaging with adaptive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025. URL https://www.cs.toronto.edu/~robin/aosr/.

Sam F Sweere, Ivan Valtchanov, Maggie Lieu, Antonia Vojtekova, Eva Verdugo, Maria Santos-Lleo, Florian Pacaud, Alexia Briassouli, and Daniel Cámpora Pérez. Deep learning-based super-resolution and de-noising for xmm-newton images. *Monthly Notices of the Royal Astronomical Society*, 517(3):4054–4069, 2022.

Yuan-Sen Ting. Deep learning in astrophysics. *Annual Review of Astronomy and Astrophysics*, 64, 2025.

A. Tokovinin. From differential image motion to seeing. *Publications of the Astronomical Society of the Pacific*, 114:1156–1166, 2002. doi: 10.1086/342683.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pages 945–948. IEEE, 2013.

Cong Wang, Jinshan Pan, Wei Wang, Jiangxin Dong, Mengzhu Wang, Yakun Ju, and Junyang Chen. Promptrestorer: A prompting image restoration method with degradation perception. *Advances in Neural Information Processing Systems*, 36:8898–8912, 2023.

Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. Sinsr: diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25796–25805, 2024.

Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. The MIT press, 1949.

Kuo-Cheng Wu, Guohang Zhuang, Jinyang Huang, Xiang Zhang, Wanli Ouyang, and Yan Lu. Star: A benchmark for astronomical star fields super-resolution. *arXiv preprint arXiv:2507.16385*, 2025.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 37: 92529–92553, 2024.

Donald G York, Jennifer Adelman, John E Anderson, Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, et al. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579–1587, 2000.

Wei Yu, Bowen Yang, Qinglin Liu, Jianing Li, Shengping Zhang, and Xiangyang Ji. Rethinking imbalance in image super-resolution for efficient inference. *Advances in Neural Information Processing Systems*, 37:6860–6883, 2024.

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36: 13294–13307, 2023.

Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23153–23163, 2025.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.

Xiangrui Zeng, Lingyu Zhu, Shiqi Wang, and Sam Kwong. Wienerflow: Wiener-adaptive flow matching for perception and fidelity trade-off in low-light image enhancement, 2025. URL https://openreview.net/forum?id=TwvnG6mobR.

Dichang Zhang, Yixuan Shao, Simon Birrer, and Dimitris Samaras. As-bridge: A bidirectional generative framework bridging next-generation astronomical surveys. *arXiv preprint arXiv:2603.11928*, 2026.

Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.

Leheng Zhang, Weiyi You, Kexuan Shi, and Shuhang Gu. Uncertainty-guided perturbation for image super-resolution diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17980–17989, 2025.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.

Xiang Zhang, Yulun Zhang, and Fisher Yu. Hit-sr: Hierarchical transformer for efficient image super-resolution. In *European Conference on Computer Vision*, pages 483–500. Springer, 2024a.

Yasi Zhang, Peiyu Yu, Yaxuan Zhu, Yingshan Chang, Feng Gao, Ying N Wu, and Oscar Leong. Flow priors for linear inverse problems via iterative corrupted trajectory matching. *Advances in Neural Information Processing Systems*, 37:57389–57417, 2024b.

Guangcong Zheng, Xianpan Zhou, Xuewei Li, Zhongang Qi, Ying Shan, and Xi Li. Layoutdiffusion: Controllable diffusion model for layout-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22490–22499, 2023a.

Qinqing Zheng, Matt Le, Neta Shaul, Yaron Lipman, Aditya Grover, and Ricky TQ Chen. Guided flows for generative modeling and decision making. *arXiv preprint arXiv:2311.13443*, 2023b.

Linqi Zhou, Aaron Lou, Samar Khanna, and Stefano Ermon. Denoising diffusion bridge models. *arXiv preprint arXiv:2309.16948*, 2023a.

Man Zhou. Improving the learning capability of small-size image restoration network by deep fourier shifting. In *Advances in Neural Information Processing Systems*, 2024.

Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. Srformer: Permuted self-attention for single image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12780–12791, 2023b.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1219–1229, 2023.

A Survey Data and Instrument Characteristics

To fully appreciate the complexity of mapping the ground-based low-resolution (LR) observations to the space-based high-resolution (HR) targets, it is necessary to detail the disparate instrument characteristics (Figure 11) and image co-addition pipelines of the DESI Legacy Imaging Surveys and the Hubble Space Telescope (HST), whose overlapping observational footprints are shown in Figure 12.

A.1 DESI Legacy Surveys: Ground-Based Seeing and Co-addition

The LR images from the DESI Legacy Imaging Surveys are fundamentally limited by atmospheric turbulence, which blurs the astronomical signal into a seeing-limited Point Spread Function (PSF). The i-band data ($\sim$7600–8730 Å) utilized in our benchmark, which offers the closest spectral match to the HR targets, is primarily acquired by the Dark Energy Camera (DECam), a 570-megapixel



Hubble Space Telescope (HST) Dark Energy Spectroscopic Instrument (DESI)

Figure 11: The left image shows HST, which provides the HR targets. The right image shows the Mayall telescope, which hosts the DESI spectrograph. The DECam i band LR images used here were obtained with the Blanco telescope. Image credits are NASA for HST and KPNO/NOIRLab/NSF/AURA/T. Slovinský for Mayall.

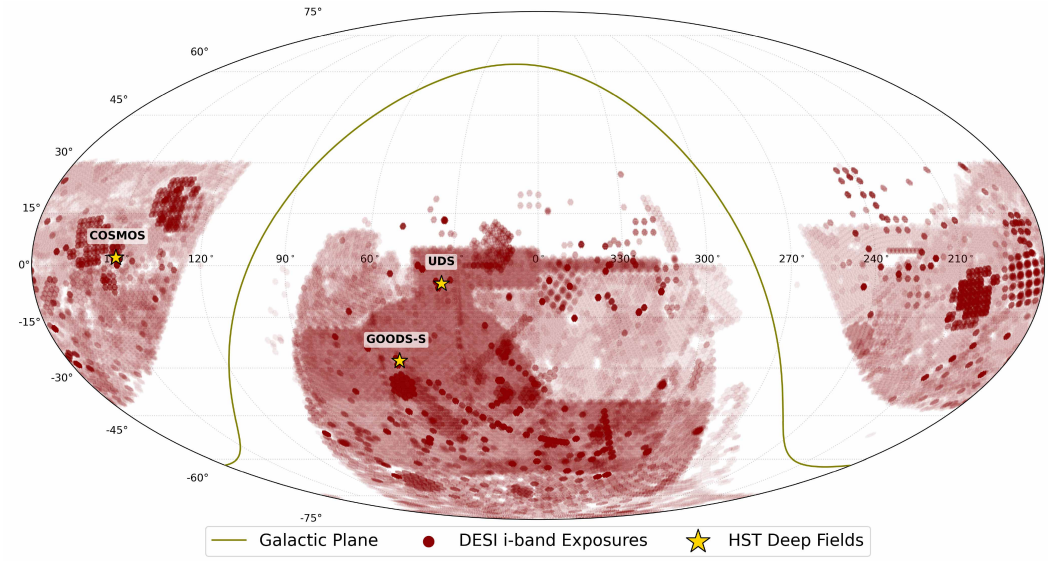


Figure 12: The footprint of the Legacy Surveys (DR10) i-band observations (dark red) overlaid with the locations of three equatorial and southern HST deep fields: COSMOS, UDS, and GOODS-S (gold stars). The olive green curve indicates the Galactic plane.

wide-field imager mounted at the prime focus of the Victor M. Blanco 4-meter telescope at the Cerro Tololo Inter-American Observatory (CTIO).

Despite the substantial light-gathering power of the 4-meter primary mirror, ground-based observations are subject to atmospheric degradation. With DECam's native pixel scale of $0.262''$/pixel, the typical seeing Full Width at Half Maximum (FWHM) for these observations ranges from $1.0''$ to $1.3''$. The final LR images provided in our benchmark are co-added mosaics. The data processing pipeline maps multiple individual dithered exposures onto a common predefined sky footprint (often referred to as "bricks"). This co-addition process includes sky-background subtraction, astrometric calibration, and standard inverse-variance weighting. While this improves the signal-to-noise ratio

(SNR) and depth, the resulting images still inherently suffer from the spatially varying PSF blur and elevated sky background noise intrinsic to ground-based optical astronomy.

Long exposures average short-timescale atmospheric speckles into an effective seeing PSF. Under Kolmogorov turbulence with $D/r_0 \gg 1$, the angular FWHM is approximately $0.98\lambda/r_0$, compared with the diffraction scale λ/D Fried [1966], Roddier [1981], Tokovinin [2002]. Here D is aperture diameter, r_0 the atmospheric coherence length, and λ the observing wavelength. This asymptotic relation does not identify the fitted Gaussian width measured in pixels. Conversion depends on plate scale, and the effective coadd response also includes instrumental and registration effects.

A.2 HST ACS/WFC: Space-Based Optics and Astrodrizzle

In contrast, the HR targets are obtained by the Hubble Space Telescope (HST), which orbits at an altitude of approximately 540 km. Operating entirely above the Earth's atmosphere, HST's 2.4-meter primary mirror delivers a highly stable, diffraction-limited PSF that is entirely immune to atmospheric seeing. The Advanced Camera for Surveys Wide Field Channel (ACS/WFC) in the F814W filter ($\sim$7000–9500 Å) serves as the primary instrument for our HR data. It consists of two 2048 $\times$ 4096 pixel charge-coupled devices (CCDs) separated by a physical chip gap, delivering a native pixel scale of approximately $0.05''$/pixel.

As introduced in the main text, raw HST exposures inherently suffer from this detector undersampling and transient cosmic-ray events. To mitigate these issues and achieve the final $0.03''$/pixel grid, the HST pipeline employs Astrodrizzle , based on Drizzle Fruchter and Hook [2002], Gonzaga et al. [2012], which combines multiple sub-pixel dithered exposures. Instead of a simple direct co-addition, the algorithm geometrically shrinks the footprint of the input pixels (controlled by the *pixfrac* parameter) and "drizzles" their flux onto a finer output grid. This sub-pixel interlacing effectively recovers high-frequency spatial structures that approach the optical diffraction limit of the 2.4-meter aperture, successfully bridging the gap between the native $0.05''$ detector resolution and the final $0.03''$ product.

A.3 Inverse-Variance Weight Maps for HST Observations

Concurrently, the generation of the HST inverse-variance weight maps (*wht*) is mathematically deterministic and crucial for our benchmark. Fundamentally, the value of each pixel in these maps represents the inverse variance ($1/\sigma^2$) of the measured flux. This variance (σ^2) is not an arbitrary metric but is rigorously derived from the intrinsic hardware characteristics of the ACS/WFC CCDs and the physics of photon counting. Specifically, the pipeline calculates the total noise per pixel by modeling the detector's read noise (electronic noise generated during signal readout), dark current (thermal electron accumulation), and the Poisson noise originating from the sky background.

During the Astrodrizzle process, as multiple dithered exposures are mapped onto the final output grid, these individual pixel variances are mathematically propagated. Because the final HR mosaic is constructed from these frames with varying overlap footprints, the effective signal-to-noise ratio (SNR) is intrinsically non-uniform across the field of view.

The weight maps encode the pipeline-estimated statistical reliability and masks, reducing the contribution of low-weight or invalid pixels to the training loss. They do not identify every defect and do not remove stable optical features such as diffraction spikes or flare. In particular, weighting the loss does not convert the HST target into an artifact-free sky image.

B Precision WCS-Based Image Registration

To learn reliable cross-device mappings, eliminating translational, rotational, and geometric distortions between the ground-based and space-based platforms is a prerequisite. We achieve absolute spatial anchoring by utilizing the World Coordinate System (WCS) metadata embedded within the Flexible Image Transport System (FITS) headers of both instruments.

The WCS framework provides a rigorous non-linear mathematical transformation mapping 2D detector coordinates (x, y) to absolute celestial coordinates (Right Ascension α and Declination δ in the J2000 epoch), while properly accounting for telescope-specific focal plane distortions. Because

our objective is to map DESI observations to higher-resolution HST counterparts at exact scaling factors ($\times 2$ and $\times 4$), we establish a common celestial grid anchored directly to the DESI footprint.

We establish the native DESI pixel grid as the absolute spatial reference, retaining the DESI coadds at native resolution to preserve their noise properties and PSF. Based on this anchor, we define a corresponding high-resolution grid that precisely inherits the DESI tangent plane center and orientation, with its pixel scale divided by the targeted super-resolution factor. The HST exposures are then mathematically reprojected and flux-conservingly downsampled onto this customized grid via their native WCS metadata, establishing a shared sampling grid. Residual astrometric errors may remain.

Finally, the resulting perfectly aligned LR–HR pairs are cropped into spatially corresponding patches to produce the final dataset elements, yielding HR targets structured at 256×256 and 512×512 resolutions for the $\times 2$ and $\times 4$ scales, respectively. This strict astrometric registration eliminates spurious spatial shifts, compelling the neural network to learn genuine morphological transformations rather than compensating for coordinate misalignments.

C Scope and Scientific Use

HST is an observational target, not artifact-free truth. Its limited overlap, noise, defects, bandpass, and optical structures constrain supervision and evaluation. WHT weighting reduces the learning of low-confidence defects such as chip gaps and cosmic-ray trails, but cannot remove every stable instrumental structure, including diffraction and flare patterns. The reported metrics evaluate photometric and morphological reconstruction accuracy against the HST observations. Scientific deployment requires independent validation.

FluxFlow cannot reliably recover a source absent from its HST supervision but present in another catalog, because such a source provides no positive HR training signal. MC-FS may preserve coarse evidence already present in the LR image, but cannot distinguish a real object from noise, transients, bandpass or depth differences, or model error. This dependence on available pairs limits discovery use. Unmatched reconstructed sources require independent catalog, multi-band, or multi-epoch validation.

We select the DESI i band and HST F814W because they are the closest available bandpasses between the two observing programs. The LR/HR band configuration is identical across training and testing. HST images, catalogs, and WHT maps are not required at inference. FluxFlow receives only the LR/DESI observation, and MC-FS compares the forward projection of the reconstruction with that same input.

Future work will investigate metadata-conditioned anisotropic Moffat profiles, spatially varying PSF fields, and per-source adaptive PSF estimation and reprojection, as well as multi-instrument and multi-epoch data.

D Additional Related Works

MC-FS draws on a line of research that injects measurement-consistency corrections into the sampling trajectories of generative models for inverse problems. We clarify its relationship to the most relevant prior approaches.

D.1 Diffusion Posterior Sampling

Diffusion Posterior Sampling (DPS) Chung et al. [2022] approximates the posterior score by adding to each reverse diffusion step a gradient of the log-likelihood evaluated at the Tweedie estimate $\hat{\mathbf{x}}_0(\mathbf{x}_t)$ using:

$$\mathbf{x}_{t-1} \leftarrow \mathbf{x}_{t-1}^{\text{prior}} - \zeta \nabla_{\mathbf{x}_t} \left\| \mathbf{y} - \mathbf{A}\big(\hat{\mathbf{x}}_0(\mathbf{x}_t)\big) \right\|^2 . \tag{12}$$

The gradient propagates through the denoiser via automatic differentiation, so DPS effectively composes the Jacobian of the score network with $\mathbf{A}^\top$. Pseudoinverse-Guided Diffusion Models (ΠGDM) Song et al. [2023] replace $\mathbf{A}^\top$ with the Moore–Penrose pseudoinverse $\mathbf{A}^+$ to obtain a more accurate posterior approximation, again applied to the Tweedie estimate.

Table 4: Conceptual comparison with the original diffusion-based posterior sampling formulations. The residual is evaluated at a one-step clean-image estimate for DPS and ΠGDM, and at the Euler proposal for MC-FS.

Method	Generative backbone	Residual point	Back-projection	Autograd through network
DPS Chung et al. [2022]	Diffusion	$\hat{\mathbf{x}}_0(\mathbf{x}_t)$	$\mathbf{A}^\top$ via Jacobian	Required
ΠGDM Song et al. [2023]	Diffusion	$\hat{\mathbf{x}}_0(\mathbf{x}_t)$	Pseudoinverse $\mathbf{A}^+$	Required
Ours	Flow matching	$\tilde{\mathbf{x}}^{(i+1)}$	Wiener-regularized inverse	Not required

FluxFlow learns a paired conditional transport toward $p_\theta(\mathbf{x}_{\mathrm{HST}} \mid \mathbf{y}_{\mathrm{DESI}})$, whereas conventional posterior solvers typically start from an unconditional image prior. Under the linear flow parameterization, the current velocity already gives a one-step destination estimate $\hat{\mathbf{x}}_{1|t} = \mathbf{x}_t + (1-t)v_\theta(\mathbf{x}_t, t, \mathbf{c})$. No additional clean-image predictor is required. MC-FS instead evaluates its residual on the Euler proposal $\tilde{\mathbf{x}}^{(i+1)}$, deliberately injecting measurement information early in the trajectory. Its Wiener correction avoids differentiation through the velocity network. This is a reconstruction heuristic supported by the stated spectral model and experiments, not an exact posterior score or probability-flow ODE. Table 4 contrasts the original diffusion formulations with MC-FS.

D.2 Plug-and-Play Frameworks for Inverse Problems

MC-FS is related to the alternating prior and data-fidelity structure of plug-and-play methods Venkatakrishnan et al. [2013], Romano et al. [2017], Zhang et al. [2021], Hurault et al. [2022], Kamilov et al. [2023], Zhu et al. [2023]. The learned conditional velocity proposes an update and the residual filtered by the Wiener kernel supplies a soft correction for measurement consistency. The relationship is an analogy because the blur-domain Wiener filter composed with area-downsampling adjoints is not, in general, the exact proximal operator of the full blur-and-downsampling data term.

MC-FS does not perform a hard projection onto $\mathcal{C}(\mathbf{y})$ and does not guarantee posterior samples. For a fixed initial Gaussian draw, its updates are deterministic. Variation across initial draws is not calibrated posterior uncertainty. Its purpose here is stable conditional reconstruction. Stochastic, uncertainty-calibrated sampling is left to future work. The design targets one conservative reconstruction per observation. Refreshing noise at every step could reintroduce compact fluctuations that the consistency correction is intended to attenuate. The initial Gaussian draw retains stochasticity, while the subsequent updates introduce no fresh noise.

D.3 Flow-Matching for Inverse Problems

Efficient SR approaches such as SinSR Wang et al. [2024], OSEDiff Wu et al. [2024], and TSD-SR Dong et al. [2025] are relevant to reducing sampling cost. Applying their natural-image pretrained models directly to single-band flux-calibrated observations would conflate efficiency with domain and degradation mismatch. A controlled comparison requires adaptation to the same paired data, normalization, uncertainty handling, and source-level evaluation.

FlowDPS Kim et al. [2025] decomposes flow updates into destination-image and noise components, enabling likelihood guidance and stochastic noise refresh. We adapt DPS, ΠGDM, FlowDPS, and FLOWER updates to a shared paired conditional FluxFlow checkpoint to isolate the effect of the sampling update.

Comparison of sampling methods. The comparison keeps the paired conditional velocity network fixed and changes the inference update. The DESI condition $\mathbf{c}$ remains present in every network call. These samplers share the following destination estimate and measurement residual.

$$\hat{\mathbf{x}}_{1|t} = \mathbf{x}_t + (1-t)v_\theta(\mathbf{x}_t, t, \mathbf{c}), \qquad \mathbf{r}_t = \mathbf{A}\hat{\mathbf{x}}_{1|t} - \mathbf{y}. \tag{13}$$

Writing $\mathbf{J}_{\hat{x}} = \partial \hat{\mathbf{x}}_{1|t} / \partial \mathbf{x}_t$, DPS differentiates the destination residual through the conditional network. ΠGDM replaces the adjoint residual with a regularized pseudoinverse residual before applying the destination Jacobian as follows.

$$\begin{aligned} \mathbf{g}_{\mathrm{DPS}} &= \nabla_{\mathbf{x}_t} \tfrac{1}{2}\|\mathbf{r}_t\|_2^2 = \mathbf{J}_{\hat{x}}^\top \mathbf{A}^\top \mathbf{r}_t, \\ \mathbf{g}_{\Pi\mathrm{GDM}} &= \mathbf{J}_{\hat{x}}^\top \mathbf{A}_\epsilon^\dagger \mathbf{r}_t. \end{aligned} \tag{14}$$

Here $\mathbf{A}_\epsilon^\dagger$ denotes the regularized pseudoinverse used in the adaptation. Its regularization is distinct from MC-FS's Wiener parameter λ_{SNR}.

FlowDPS refines the destination using the pseudoinverse and recombines it with a refreshed source-noise endpoint according to the flow schedule. FLOWER applies a data-fidelity proximal refinement to the destination and reprojects the refined estimate along the flow path.

MC-FS instead evaluates $\mathbf{A}\tilde{\mathbf{x}}^{(i+1)} - \mathbf{y}$ at the conditional Euler proposal and applies the Wiener correction in Equation (11). Thus it corrects the current trajectory directly, without network-Jacobian backpropagation or an iterative pseudoinverse/proximal solve. The linear MMSE and finite-step spectral analysis motivate this choice, but do not establish an exact posterior law for the corrected nonlinear trajectory.

A growing body of work studies flow matching directly in pixel space with various weighting and parameterization choices Chen et al. [2025c], Gagneux et al. [2026], and develops flow-matching solvers and priors specifically tailored to linear inverse problems Pourya et al. [2025], Zhang et al. [2024b]. WienerFlow Zeng et al. [2025] introduces Wiener-adaptive flow matching in the related setting of low-light image enhancement and is closest in spirit to MC-FS, but it changes the training formulation and is not a drop-in sampling update for the same fixed conditional checkpoint. Our correction instead targets an approximate ground-to-space super-resolution operator. In astronomy, score-based generative priors have been applied to multi-band source separation Sampson et al. [2024] and broader astrophysical applications Ting [2025]. FluxFlow complements these by focusing on the photometrically conservative ground-to-space mapping under realistic atmospheric PSFs.

Quantitative sampler comparison. We compare these sampling methods at $\times 2$ using the same FluxFlow checkpoint, loss, training strategy, 1,701-image test split, 10-step protocol, and calibrated forward model with $\sigma_{\mathrm{PSF}} = 2$. Table 5 reports the results. FLOWER achieves the highest PSNR and F1, while MC-FS achieves the best SSIM, Flux-L1, and precision and remains competitive in PSNR and F1. These results isolate the sampling update under a shared learned conditional distribution.

Table 5: Conditional sampler comparison on DESI–HST at $\times 2$ with a shared checkpoint and 10 sampling steps.

Method	PSNR ↑	SSIM ↑	Flux-L1 ↓	Precision ↑	Recall ↑	F1 ↑
DPS	30.56	0.653	3.197	0.621	**0.503**	0.5230
ΠGDM	30.75	0.686	3.060	0.711	0.456	0.5252
FlowDPS	31.26	0.705	2.986	0.654	0.490	0.5323
FLOWER	**31.32**	0.709	2.977	0.708	0.469	**0.5449**
Ours	31.23	**0.714**	**2.959**	**0.738**	0.458	0.5412

E Implementation Details

Source Labels on the HST. HST instance labels are produced by running SEP Bertin and Arnouts [1996], Barbary [2016] on each HST cutout, using the accompanying weight map to construct a validity mask in which pixels with non-positive weight are excluded from both the background statistics and the detection step. We estimate the local sky on a 32×32 mesh with no additional median filtering, subtract the resulting background plane from the science image, and obtain a per-pixel noise standard deviation from the same background model. Sources are then extracted by thresholding the background-subtracted image at $2\times$ the local RMS and retaining connected components of at least 5 pixels. SEP returns a segmentation map together with a structured catalog. We store the segmentation map as a 32-bit integer array in which the value 0 denotes background and each positive integer is a unique source identifier starting from one. The catalog records the centroid coordinates, the Kron ellipse parameters of semi-major axis, semi-minor axis, and position angle, the pixel count of each source, and the SEP extraction flag. Source identifiers in the catalog are aligned by construction with the labels of the segmentation map, so that selecting all pixels with a given label directly recovers the spatial support of the corresponding catalog entry. This labelling defines the set used by the detection metric in the main text and by the per-source flux comparison reported in the supplementary tables.

Flux-L1 Metric. Using the segmentation map and Kron-ellipse parameters produced above, we form an elliptical aperture mask M_k for each of the K HST sources and integrate the aperture flux on both the prediction and the HR HST with the SEP sum estimator. The per-image error is the absolute deviation of integrated flux per source, summed over all K HST sources, and aggregating across the test set then yields the macro-averaged Flux-L1 reported in Table 2:

$$\text{Flux-L1} = \sum_{k=1}^{K} \Big| \sum_{p \in M_k} \hat{x}_p - \sum_{p \in M_k} x_p^* \Big|, \tag{15}$$

where $\hat{x}$ denotes the prediction and x^* the HR HST image.

Source Detection on Prediction. To quantify hallucination in the predicted images, we run the same detection pipeline independently on the prediction and on the HR HST, match the two catalogues, and report the per-image macro-averaged precision, recall, and F1 in the detection columns of Table 3. We assess source recovery through a three-stage pipeline that compares each prediction against a HST label mask. For every predicted high-resolution image, we first replace invalid pixels with zeros and estimate a local sky background with SEP on a 64×64 grid. After subtracting this background plane we extract sources using a detection threshold of $2\times$ the local background RMS, requiring at least 5 connected pixels per candidate. Deblending operates over 32 contour levels with a flux-contrast threshold of 0.02, splitting overlapping blobs whose sub-peaks rise above this fraction of the parent flux. The reported sub-pixel centroids of the surviving detections form the predicted catalogue. On the HST side, the integer label mask assigns each positive value to one source. We discard labels covering fewer than 10 pixels to suppress fragmentary annotations, and the geometric centroid of every remaining label is computed as the unweighted mean of its pixel coordinates, with no flux weighting applied.

Predicted and HST centroids are then aligned through greedy nearest-first one-to-one matching. We build the pairwise squared-distance matrix between the two sets, retain only candidate pairs whose separation falls within a small tolerance of 10 pixels, and sort the survivors by ascending distance. Walking through this sorted list, the first pair whose prediction and HST source are both still unclaimed is registered as a true positive, after which both endpoints are marked used so that any later pair reusing either of them is skipped. Predictions left without a partner contribute false positives, and unmatched HST sources contribute false negatives. We compute precision, recall, and F1 per image and aggregate over all test samples by averaging the per-image scores, giving every image equal weight regardless of how many sources it contains and preventing dense scenes from masking failures on sparse ones.

Model Architecture. The velocity network v_θ is a conditional U-Net operating directly in pixel space, illustrated in Figure 13. The DESI observation $\mathbf{y}$ is first lifted to the HR grid via bicubic upsampling and concatenated channel-wise with the noisy state $\mathbf{x}_t$ to form the network input. The encoder is a stack of *Encoder Blocks*, each applying a Conv Block followed by 2×2 max-pooling that halves the spatial resolution and doubles the channel count from `base_dim`=48. A bottleneck Conv Block sits at the lowest resolution. The symmetric decoder mirrors this pyramid with *Decoder Blocks* that upsample via a transposed-convolution layer (*ConvTrans Block*) and fuse skip connections from the matching encoder level, producing the OT-CFM velocity prediction v_θ on the HR grid. Each Conv Block is a Conv2d–LeakyReLU–FiLM–Conv2d–LeakyReLU stack in which the FiLM layer Perez et al. [2018] applies a feature-wise affine modulation conditioned on the time-conditioning state τ, so that the time signal is injected at every resolution. The time step $t \in [0, 1]$ is mapped to τ by a sinusoidal positional embedding followed by a two-layer MLP with output dimension 256. The two scale settings ($\times2$ and $\times4$) share this architecture exactly and differ only in the spatial resolution of $\mathbf{x}_t$ and the upsampled condition.

Pixel Normalization. The DESI and HST cutouts have a wide intrinsic dynamic range in their native flux units (nanomaggies). Before training, we compute global intensity clip thresholds at the 0.01% tail percentiles of all pixels in the training split to suppress extreme values caused by cosmic rays and hot pixels Liu et al. [2026], which would otherwise dominate the normalization range. One threshold pair is computed for the LR set and another for the HR set, both estimated once on the training split and frozen across all experiments. The clipped pixels are then linearly min–max rescaled to $[0, 1]$ using the global per-modality flux extrema. Both training and inference operate

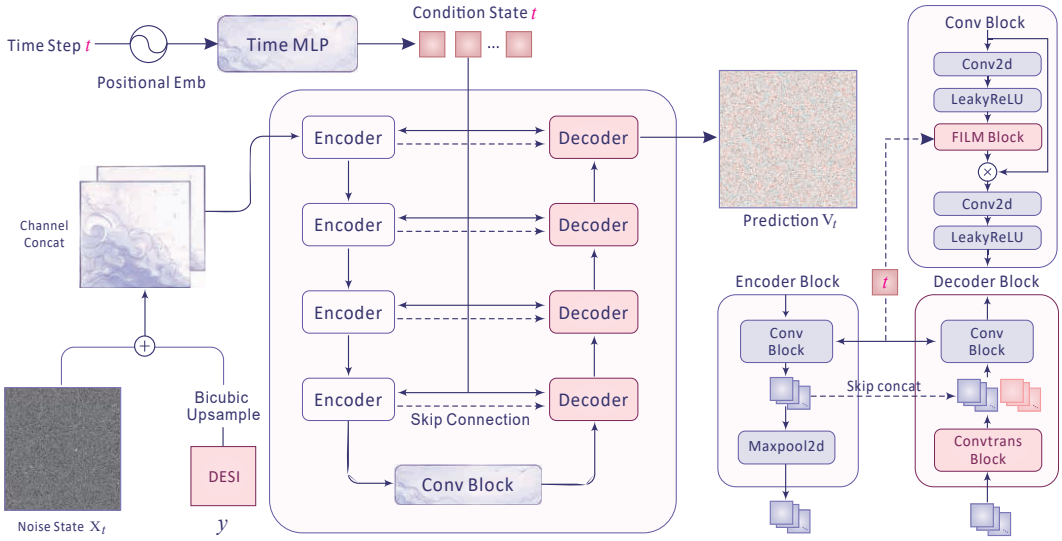


Figure 13: Architecture of the FluxFlow velocity network v_θ. A conditional U-Net takes the channel-wise concatenation of the noisy state $\mathbf{x}_t$ and the bicubic-upsampled DESI observation $\mathbf{y}$, and predicts the OT-CFM velocity v_θ on the HR grid. Time conditioning enters every Conv Block via FiLM modulation.

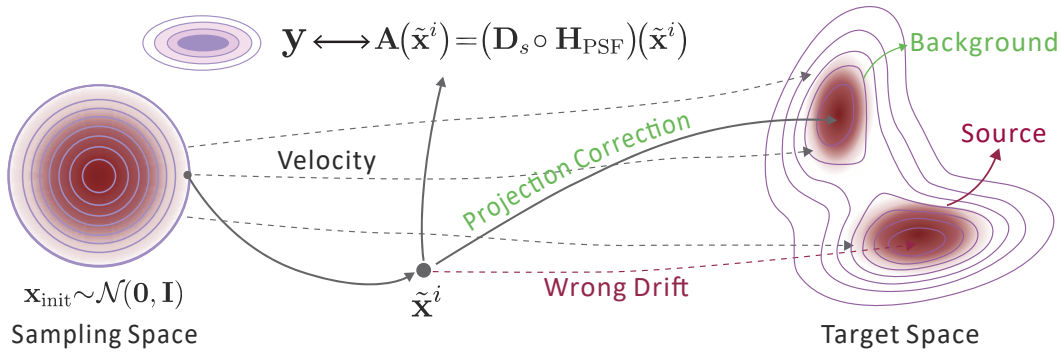


Figure 14: Conceptual illustration of MC-FS. The flow transports $\mathbf{x}_{\mathrm{init}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ toward the target space, but the learned velocity may drift the candidate $\tilde{\mathbf{x}}^i$ toward a hallucinated source (red). MC-FS compares $\mathbf{y}$ with $\mathbf{A}(\tilde{\mathbf{x}}^i)$ (Section 5.2) to expose this inconsistency and induce a projection correction (green) that pulls the trajectory back toward modes consistent with the observation.

in this normalized domain, and PSNR/SSIM are reported in the same domain so that the maximal value is exactly 1 and the metrics are directly comparable across methods. For the Flux-L1 metric, we denormalize the predictions back to nanomaggies before computing the per-source aperture flux deviation defined in Equation (15), so that the reported flux errors carry their original physical units.

Adaptive optics and reconstruction. Adaptive optics corrects wavefront distortions during acquisition Babcock [1953], Beckers [1993]; recent work combines controlled adaptive-optics distortions with computational super-resolution Swanson et al. [2025]. Its performance depends on the instrument, wavelength, guide-star geometry, and field of view. It also requires specialized hardware, real-time control, and demanding observing conditions. FluxFlow instead operates on existing seeing-limited survey coadds without adaptive-optics measurements or hardware control.

Forward-Model PSF Calibration. The Gaussian PSF width $\sigma_{\mathrm{PSF}} = 2$ used in the MC-FS forward operator defined in Equation (6) is calibrated directly on the paired DESI–HST training set. For each training pair we forward-project the HST high-resolution image through a candidate Gaussian-blur followed by $s \times s$ area-downsampling, and compute the validity-masked mean-squared error over

detected source regions against the corresponding DESI observation. Sweeping σ on a fine grid and taking the value that minimizes this error averaged across the training set yields the data-driven optimum $\sigma^* = 1.92$. We round this to the integer value $\sigma_{\mathrm{PSF}} = 2$ used throughout inference. Here σ is measured in HR-grid pixels. This is one dataset-level effective kernel, not a per-exposure or per-source PSF estimate, and it can absorb instrumental response, bandpass differences, and registration error as well as atmospheric blur. It should not be identified directly with a measured atmospheric coherence length.

Training Details. We train the velocity network v_θ of Figure 13 with the OT-CFM objective. For each sample we draw $x_0 \sim \mathcal{N}(0, I)$ and $t \sim \mathcal{U}(0, 1)$, form the linear interpolant $x_t = (1-t)\, x_0 + t\, x_1$ with target velocity $u_t = x_1 - x_0$, and regress $v_\theta(x_t, t, c)$ against u_t under the spatially-weighted MSE defined in Equation (5). On top of the optimizer setting reported in the main text, we apply gradient-norm clipping at 1.0 and a 1000-step linear warm-up before cosine decay to zero, and run training in fp32. The EMA model is evaluated every 20 epochs by Euler-sampling 10 steps from pure Gaussian noise, reporting MSE/PSNR in the normalized HST domain together with total-flux error in the original flux units.

Inference cost and profiling protocol. All samplers share the same UNet and introduce no additional trainable network. FluxFlow has 13.12M trainable parameters at both scales, with 50.04 MiB of FP32 parameter memory. Pixel-space inference is retained because of the VAE ablation in Table 3. We measure batch-size-one inference on one NVIDIA RTX 8000 GPU, using two warm-up images and the same 128-image subset, excluding data loading and synchronizing GPU execution. Table 6 reports the cost of 10 Euler steps. Network FLOPs count 10 network forward passes and exclude the FFT correction.

Table 6: FluxFlow inference cost on one NVIDIA RTX 8000 GPU with batch size one and 10 Euler steps.

Scale	HR shape	Time per image in ms	Peak memory in MiB	Network FLOPs
$\times 2$	256×256	60.80 ± 0.90	241.6	604.316 G
$\times 4$	512×512	219.94 ± 1.83	441.7	2.417 T

Table 7 compares the inference times of these samplers under the same timing protocol. At $\times 2$, MC-FS is $1.47\times$ faster than FLOWER, without inference-time automatic differentiation or conjugate-gradient solves.

Table 7: Inference time in milliseconds per image on one NVIDIA RTX 8000 GPU.

Method	$\times 2$	$\times 4$
DPS	114.41	405.01
ΠGDM	153.47	441.76
FlowDPS	97.93	252.63
FLOWER	89.52	254.96
Ours	**60.80**	**219.94**

Training Time. We report the end-to-end training time of FluxFlow and all baselines in Table 8, measured on a single NVIDIA A100-80GB GPU under an identical schedule budget. FluxFlow trains in 16h at $\times 2$ scale and 47h at $\times 4$ scale, substantially faster than the Transformer regression baseline HAT which needs over 540 hours and the bridge-based generative baseline AS-Bridge which needs 144 to 360 hours, while matching the photometric and detection metrics reported in Tables 2 and 3.

F Analysis of Measurement-Consistent Flow Sampling

This section explains the structure of MC-FS and analyzes simplified linear correction models. Section F.1 describes observable residuals and the limitations caused by the null space of the forward operator . Section F.2 derives a blur-domain Wiener filter under stationary signal and noise assumptions. Sections F.3 and F.4 study blur-only Fourier modes and a Gaussian-PSF toy problem to illustrate finite-iteration conditioning. These calculations do not prove posterior correctness, convergence of the nonlinear conditional sampler, or removal of structures in the measurement null space.

Table 8: End-to-end training time on a single NVIDIA A100-80GB GPU at $\times 2$ and $\times 4$ scales.

Method	SwinIR	HAT	FISR	cGAN	GD-Net	AS-Bridge	**Ours**
$\times 2$	94h	552h	94h	8h	5h	144h	**16h**
$\times 4$	98h	549h	176h	16h	184h	360h	**47h**

F.1 Forward-Model Consistency for Ill-Posed Image Super-Resolution

Astronomical super-resolution is a severely ill-posed inverse problem. Given the forward model:

$$\mathbf{y} = \mathbf{A}(\mathbf{x}) + \boldsymbol{n}, \qquad \mathbf{A} = \mathbf{D}_s \circ \mathbf{H}_{\mathrm{PSF}}, \tag{16}$$

the operator $\mathbf{A}$ is rank-deficient: spatial downsampling $\mathbf{D}_s$ collapses an $H_{\mathrm{hr}} \times W_{\mathrm{hr}}$ image into an $H_{\mathrm{lr}} \times W_{\mathrm{lr}}$ observation, discarding $(s^2-1)/s^2$ of the spatial degrees of freedom, while PSF convolution $\mathbf{H}_{\mathrm{PSF}}$ further attenuates high-frequency content. Consequently, the set of high-resolution images $\mathbf{x}$ consistent with a given DESI observation $\mathbf{y}$ forms a high-dimensional consistency set that allows for noise. The set is defined as

$$\mathcal{C}(\mathbf{y}) = \left\{ \mathbf{x} \in \mathbb{R}^{H_{\mathrm{hr}} \times W_{\mathrm{hr}}} : \|\mathbf{A}(\mathbf{x}) - \mathbf{y}\|^2 \leq \epsilon \right\}, \tag{17}$$

In the noiseless limit, exact equality defines an affine set. Here ϵ accounts for the noise level $\|\boldsymbol{n}\|^2$. The role of the flow matching prior v_θ is to select, from this enormous consistency set, samples that also lie on the data manifold of plausible HST-quality images.

A residual can expose structures whose forward projection disagrees with the observation under the chosen model. However, rank deficiency permits distinct HR structures with the same LR projection, so consistency is not sufficient to exclude hallucinations. Noise and forward-model mismatch can also cause genuine structure to have a nonzero residual.

Forward-model consistency correction directly addresses this failure mode by projecting the trajectory back toward $\mathcal{C}(\mathbf{y})$ at every integration step. Concretely, each ODE update is followed by a gradient step on the data-fidelity term:

$$\mathcal{D}(\mathbf{x}) = \tfrac{1}{2}\|\mathbf{A}(\mathbf{x}) - \mathbf{y}\|^2, \tag{18}$$

whose gradient is $\nabla_{\mathbf{x}}\mathcal{D}(\mathbf{x}) = \mathbf{A}^\top(\mathbf{A}(\mathbf{x}) - \mathbf{y})$. The Wiener variant replaces the PSF adjoint in this correction with the Wiener kernel rather than applying the exact gradient. Both depend on the approximate operator, and neither can resolve its null space without information from the learned conditional model.

F.2 Derivation of the Wiener Correction Kernel

We derive the Wiener correction kernel $W(\mathbf{k})$ of Equation (10) from the minimum mean-square error (MMSE) principle. Consider the deconvolution sub-problem at integration step i. Given an upsampled residual $\mathbf{r}_\uparrow^{(i)} = \mathbf{D}_s^\top(\mathbf{r}^{(i)})$ on the high-resolution grid, we seek a high-resolution correction signal $\boldsymbol{\rho}^{(i)}$ such that $\mathbf{H}_{\mathrm{PSF}}(\boldsymbol{\rho}^{(i)}) \approx \mathbf{r}_\uparrow^{(i)}$ in the presence of measurement noise. This order of operations matches the strict-adjoint update in Equation (11), with $H(\mathbf{k})$ and the residual it acts on both defined on the HR grid.

Working in the Fourier domain, let $H(\mathbf{k}) = \mathcal{F}[h]$ denote the optical transfer function of the PSF kernel h, and assume the correction signal admits a wide-sense stationary spectral representation with power spectral density $\Phi_\rho(\mathbf{k})$, while the noise has power σ_n^2. The Wiener filter is the linear estimator $\boldsymbol{\rho}^{(i)} = w * \mathbf{r}_\uparrow^{(i)}$ that minimizes the mean-square error:

$$\mathrm{MSE}(w) = \mathbb{E}\Big[\, \|\boldsymbol{\rho}^{(i)} - w * \mathbf{r}_\uparrow^{(i)}\|^2 \,\Big]. \tag{19}$$

Setting the variation $\partial\mathrm{MSE}/\partial w = 0$ and solving in the Fourier domain Wiener [1949] yields the optimal filter:

$$W(\mathbf{k}) \;=\; \frac{H^*(\mathbf{k})\,\Phi_\rho(\mathbf{k})}{|H(\mathbf{k})|^2\,\Phi_\rho(\mathbf{k}) + \sigma_n^2} \;=\; \frac{H^*(\mathbf{k})}{|H(\mathbf{k})|^2 + \lambda_{\mathrm{SNR}}^{-1}(\mathbf{k})}, \tag{20}$$

where $\lambda_{\mathrm{SNR}}(\mathbf{k}) = \Phi_\rho(\mathbf{k})/\sigma_n^2$ is the frequency-dependent signal-to-noise ratio. In practice we replace $\lambda_{\mathrm{SNR}}(\mathbf{k})$ by a single global constant $\lambda_{\mathrm{SNR}} = 50$, recovering the form in Equation (10). This simplification avoids estimating the full power spectrum of the residual and treats the regularization parameter as a tunable hyperparameter.

Limiting Behaviors. Two extreme regimes provide intuition. As $\lambda_{\mathrm{SNR}} \to \infty$, the noise term vanishes and Equation (20) reduces to the exact inverse filter:

$$W(\mathbf{k}) \;\longrightarrow\; \frac{H^*(\mathbf{k})}{|H(\mathbf{k})|^2} \;=\; \frac{1}{H(\mathbf{k})}, \tag{21}$$

This regime maximally amplifies high frequencies but is unstable wherever $H(\mathbf{k}) \approx 0$. Conversely, as $\lambda_{\mathrm{SNR}} \to 0$, the regularization dominates the denominator and the filter magnitude vanishes

$$W(\mathbf{k}) \;\longrightarrow\; \lambda_{\mathrm{SNR}}\, H^*(\mathbf{k}) \;\to\; 0, \tag{22}$$

whose frequency-response shape remains proportional to the matched filter $H^*(\mathbf{k})$, equivalent up to scaling to the adjoint $\mathbf{H}_{\mathrm{PSF}}^\top$ of Equation (9). The Wiener kernel thus interpolates continuously between the two extremes, and λ_{SNR} controls how aggressively the deconvolution restores high-frequency content suppressed by the PSF.

The observed robustness applies to the calibrated Gaussian setting.

F.3 Spectral Analysis of Adjoint Backprojection

The qualitative claim in Section 5.2 that naive adjoint back-projection accumulates PSF-scale smoothing across integration steps admits a precise frequency-domain characterization. For clarity we ignore the downsampling operator $\mathbf{D}_s$ in this analysis and focus on the PSF component $\mathbf{H}_{\mathrm{PSF}}$, which is responsible for the resolution loss. The downsampling step contributes additional aliasing but does not alter the qualitative conclusion.

For this illustrative calculation, consider a blur-only, periodic-convolution operator, omit the learned velocity forcing, and analyze a single Fourier mode at spatial frequency $\mathbf{k}$. The full operator $\mathbf{A} = \mathbf{D}_s\mathbf{H}$ couples frequencies through decimation, so the following diagonal calculation is not an exact analysis of the implemented SR sampler. After the velocity update of Equation (7), the candidate state has Fourier amplitude $\tilde{X}^{(i+1)}(\mathbf{k})$. The adjoint correction of Equation (9) can be written in the Fourier domain as:

$$X^{(i+1)}(\mathbf{k}) \;=\; \tilde{X}^{(i+1)}(\mathbf{k}) - \eta_i\, |H(\mathbf{k})|^2\, \tilde{X}^{(i+1)}(\mathbf{k}) + \eta_i\, H^*(\mathbf{k})\, Y(\mathbf{k}), \tag{23}$$

since $\mathbf{H}^\top\mathbf{H}$ acts in this blur-only calculation as multiplication by $|H(\mathbf{k})|^2$. Restricting attention to the homogeneous part gives the per-step transfer function:

$$G_i(\mathbf{k}) \;=\; 1 - \eta_i\, |H(\mathbf{k})|^2, \tag{24}$$

and after N integration steps, the cumulative attenuation factor on mode $\mathbf{k}$ is:

$$G_{\mathrm{total}}(\mathbf{k}) \;=\; \prod_{i=0}^{N-1} \left(1 - \eta_i\, |H(\mathbf{k})|^2\right). \tag{25}$$

For modes with $H(\mathbf{k}) \neq 0$, a constant-step blur-only iteration converges to $H^*(\mathbf{k})Y(\mathbf{k})/|H(\mathbf{k})|^2$ when $|1 - \eta|H(\mathbf{k})|^2| < 1$. The inverse filter amplifies observation-noise power by $|H(\mathbf{k})|^{-2}$. Convergence alone therefore does not imply denoising. At frequencies with $H(\mathbf{k}) = 0$, this correction leaves the mode unchanged. Finite iteration counts and small $|H(\mathbf{k})|$ can slow the recovery of high-frequency components in this simplified model. These observations motivate the toy comparison in Section F.4; they do not establish PSF broadening or convergence of the learned, downsampling-based sampler.

The Wiener correction has bounded per-step gain in this simplified model. Substituting $W(\mathbf{k})$ into the correction term replaces the multiplier $H^*(\mathbf{k})$ with $H^*(\mathbf{k})/(|H(\mathbf{k})|^2 + \lambda_{\mathrm{SNR}}^{-1})$, which is uniformly bounded by $\frac{1}{2}\lambda_{\mathrm{SNR}}^{1/2}$. The Wiener kernel still tracks the inverse filter $1/H(\mathbf{k})$ deep within the passband, but stays finite in the transition band and bounds the gain of each correction. This per-step bound does not guarantee asymptotic noise suppression or convergence of the full nonlinear sampler; the tradeoff with a finite number of iterations is examined in the toy model below.

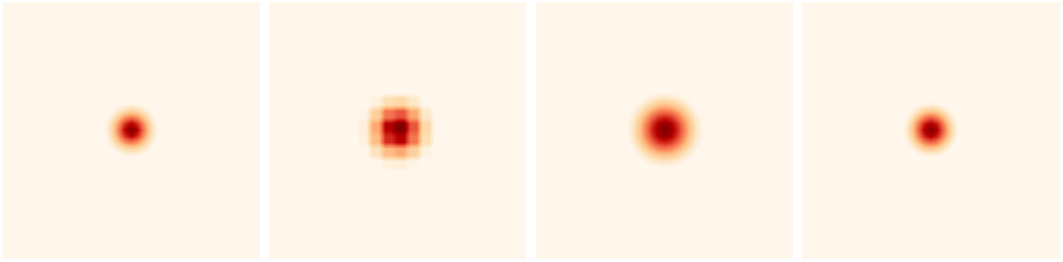


Figure 15: Toy reconstruction at $\sigma_n = 0$ with $N = 10$ Euler steps. The four panels show, from left to right, the HST-quality truth, the DESI observation, the naive adjoint back-projection, and the Wiener-corrected update. The naive recovery has fitted width $\sigma_{\rm fit} \approx 2.75$, visibly broadened relative to the truth despite the observation being noise-free. The Wiener-corrected update at the same N has $\sigma_{\rm fit} \approx 2.03$, restored to the truth.

F.4 Closed-Form Toy Model

The qualitative argument above identifies two distinct failure modes of unregularized adjoint back-projection. The first is *slow per-mode convergence* of high spatial frequencies, since the Gaussian PSF makes the per-step contraction rate $\eta|H(\mathbf{k})|^2$ exponentially small at high $|\mathbf{k}|$. The second is *noise amplification* of the unregularized inverse filter at the iteration fixed point. These two failure modes operate on different timescales. The first dominates at finite iteration count N, while the second is a $N \to \infty$ pathology. Because our sampling pipeline runs for a finite N, the first is the dominant cause of the apparent PSF widening reported in our ablation. The following closed-form analysis shows that the Wiener kernel accelerates the convergence of individual Fourier modes within the passband for a finite number of iterations.It does not change the asymptotic fixed point of the iteration, since the Wiener kernel is strictly positive over the passband and any fixed point of $X^{(i+1)} = X^{(i)} - \eta W(\mathbf{k})(H(\mathbf{k})X^{(i)} - Y(\mathbf{k}))$ still satisfies $X^{(\infty)}(\mathbf{k}) = Y(\mathbf{k})/H(\mathbf{k})$, but instead allows the iteration to be safely truncated at finite N before noise amplification substantially affects the reconstruction.

Setup. We take the PSF to be a normalized 2D Gaussian of width σ_h, the ground-truth PSF used to generate the simulated observation in the following toy analysis, so that

$$h(\mathbf{r}) \;=\; \frac{1}{2\pi\sigma_h^2}\exp\!\Big(-\tfrac{|\mathbf{r}|^2}{2\sigma_h^2}\Big), \qquad H(\mathbf{k}) \;=\; \exp\!\Big(-\tfrac{1}{2}\sigma_h^2|\mathbf{k}|^2\Big). \tag{26}$$

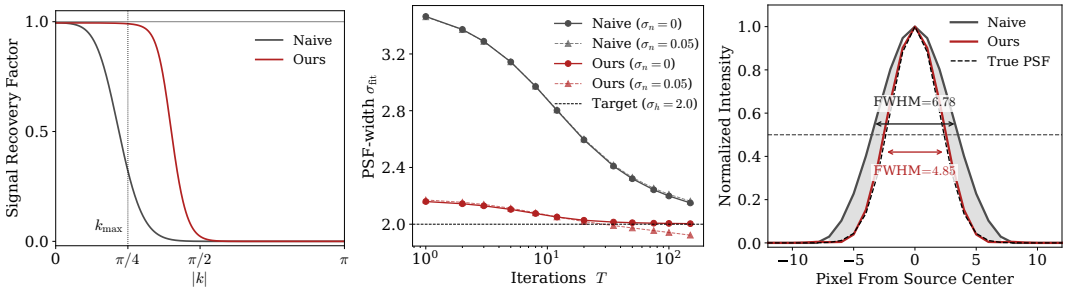
We probe the recovery operator with a single point source $x_{\rm src}(\mathbf{r}) = \Psi\,\delta(\mathbf{r})$ of amplitude Ψ, which acts as a clean impulse and isolates the recovery dynamics from any source structure. The observation noise is spatially white with power σ_n^2, so the forward observation in the Fourier domain becomes:

$$Y(\mathbf{k}) \;=\; H(\mathbf{k})\,\Psi + \Gamma(\mathbf{k}), \qquad \langle \Gamma(\mathbf{k})\,\overline{\Gamma(\mathbf{k}')}\rangle \;=\; \sigma_n^2\,\delta(\mathbf{k}-\mathbf{k}'). \tag{27}$$

We restrict the analysis to the band $|\mathbf{k}| \le k_{\max}$ supported by the LR observation. For $s\times s$ area downsampling we have $k_{\max} \approx \pi/s$ in HR frequency units. In contrast to a fixed-point analysis, our pipeline iterates Equation (9) for a finite number $N = 10$ of Euler steps, with constant step size η adopted purely as an analytical simplification for the closed-form toy-model derivation that follows. The actual sampling pipeline of Algorithm 1 uses the linear-decay schedule $\eta_i = \eta_0(1 - i/N)$ described in Section 5.2, and extending the closed-form expressions to that schedule requires the corresponding products of step-dependent factors. The closed-form analysis below is therefore parametric in N.

Per-mode Closed Form for Finite N. The adjoint update of Equation (9) acts independently on each Fourier mode. Writing the per-mode recursion $X^{(i+1)}(\mathbf{k}) = (1 - \eta|H(\mathbf{k})|^2)X^{(i)}(\mathbf{k}) + \eta H^*(\mathbf{k})Y(\mathbf{k})$ with $X^{(0)} = 0$ and summing the resulting geometric series yields the per-mode closed form:

$$X^{(N)}_{\rm naive}(\mathbf{k}) \;=\; g_N^{\rm naive}(\mathbf{k})\left[\,\Psi \;+\; \frac{\Gamma(\mathbf{k})}{H(\mathbf{k})}\,\right], \qquad g_N^{\rm naive}(\mathbf{k}) \;\equiv\; 1 - \big(1 - \eta|H(\mathbf{k})|^2\big)^N. \tag{28}$$



(a) Per-mode signal recovery $g_N(\mathbf{k})$ at $N = 10$. Naive drops sharply within the LR band $|\mathbf{k}| \le k_{\max}$, recovering only $\sim 40\%$ at the band edge, whereas ours persists.

(b) Fitted σ_{fit} vs. iteration N at $\sigma_n = 0$ and $\sigma_n = 0.05$. The two noise levels overlap, showing that the widening is a low-pass effect rather than noise amplification.

(c) 1D radial profile cuts at $N = 10$. Naive is visibly broader than the true PSF, with FWHM enlarged from 4.71 to 6.78 HR pixels, while Wiener has FWHM ≈ 4.85.

Figure 16: Numerical verification of the closed-form analysis at the same parameters as Figure 15. Panels (a)–(c) directly visualize Equations (28), (29) and (31).

The factor $g_N^{\text{naive}}(\mathbf{k})$ is the per-mode signal-recovery factor, gating both the signal term Ψ and the noise term Γ/H, and approaches unity only as $N \to \infty$. Per step, mode $\mathbf{k}$ contracts toward its fixed point at rate $\eta|H(\mathbf{k})|^2$, which is exponentially small at high $|\mathbf{k}|$ for a Gaussian PSF. For our setting $\sigma_h = 2$ and $k_{\max} = \pi/4$, $|H(k_{\max})|^2 = \exp(-\sigma_h^2 k_{\max}^2) = \exp(-\pi^2/4) \approx 0.085$, more than an order of magnitude below the deep-passband value $|H(\mathbf{0})|^2 = 1$, so the per-step contraction at the band edge is correspondingly slow. At any practical finite N the band edge therefore remains underconverged with $g_N^{\text{naive}}(k_{\max})$ bounded well below unity, while the deep passband saturates at $g_N^{\text{naive}} \approx 1$ within a few iterations.

Apparent PSF Widening at Finite N. Equation (28) attenuates both signal and noise by the same factor g_N^{naive}. The deterministic signal term is therefore a low-passed version of the true delta whenever $g_N^{\text{naive}} < 1$ at the band edge. Fitting a Gaussian profile to the recovered impulse returns a width:

$$\sigma_{\text{fit}}^{\text{naive}} \;>\; \sigma_\infty \qquad \text{for any finite } N, \tag{29}$$

where σ_∞ denotes the band-limited resolution floor set by $k_{\max}$, i.e. the spatial width of the inverse Fourier transform of the indicator of $|\mathbf{k}| \le k_{\max}$ (a 2D sinc-like profile), which is independent of the forward blur σ_h. As $N \to \infty$, $g_N^{\text{naive}}(\mathbf{k}) \to 1$ within the LR-supported band and $\sigma_{\text{fit}} \to \sigma_\infty$ rather than to σ_h. The apparent numerical agreement with $\sigma_h \approx 2$ in Figure 16b reflects the specific choice of σ_h and $k_{\max}$ used in this toy setup, not a genuine convergence to the optical blur scale. Figure 15 shows the qualitative outcome, with the naive recovery visibly broader than the truth even at the noise-free condition of $\sigma_n = 0$.

Wiener Correction. Substituting the Wiener kernel $W(\mathbf{k}) = H^*(\mathbf{k})/(|H(\mathbf{k})|^2 + \lambda_{\text{SNR}}^{-1})$ for $H^*(\mathbf{k})$ in the correction replaces the per-mode contraction rate $\eta|H(\mathbf{k})|^2$ with

$$\eta \, \frac{|H(\mathbf{k})|^2}{|H(\mathbf{k})|^2 + \lambda_{\text{SNR}}^{-1}}. \tag{30}$$

Within the LR-supported band, $|H(\mathbf{k})|^2 \gg \lambda_{\text{SNR}}^{-1}$ at our default $\lambda_{\text{SNR}} = 50$, so this ratio is close to unity rather than to $|H(\mathbf{k})|^2$. The per-step contraction rate is therefore approximately η uniformly across the band, an order of magnitude faster than the naive update at the band edge. The corresponding signal-recovery factor

$$g_N^{\text{Wiener}}(\mathbf{k}) \;=\; 1 - \left(1 - \eta \, \frac{|H(\mathbf{k})|^2}{|H(\mathbf{k})|^2 + \lambda_{\text{SNR}}^{-1}}\right)^N \tag{31}$$

saturates to unity across the entire band $|\mathbf{k}| \le k_{\max}$ within a small N (Figure 16a). The Wiener-corrected reconstruction is therefore essentially band-converged at modest N, and Equation (29) no longer applies. Figure 16b shows $\sigma_{\text{fit}}^{\text{Wiener}} \to \sigma_h$ within a few iterations, while the corresponding 1D profile cut (Figure 16c) tracks the true PSF within line width.

Table 9: Quantitative comparison at $\times 2$ and $\times 4$ on STAR dataset Wu et al. [2025].

Scale	Metric	Bicubic	EDSR	SRGAN	RCAN	SwinIR	HAT	FISR	**Ours**
$\times 2$	PSNR↑	28.71	33.50	32.25	33.89	34.05	34.38	34.40	34.66
	SSIM↑	0.6945	0.7729	0.7501	0.7808	0.7812	0.7836	0.7859	0.7918
	Flux-L1↓	125.4	48.78	51.35	43.44	42.41	41.27	41.05	41.84
$\times 4$	PSNR↑	26.21	30.73	30.33	32.13	32.24	32.67	32.51	32.44
	SSIM↑	0.5993	0.6549	0.6522	0.6838	0.6881	0.6936	0.6891	0.6867
	Flux-L1↓	239.2	67.47	74.72	60.52	55.35	48.49	50.94	57.20

Table 10: Quantitative comparison at $\times 2$ on AstroSR dataset Miao et al. [2024].

Scale	Metric	Bicubic	EDSR	SRGAN	RCAN	ENLCA	SwinIR	HAT	FISR	**Ours**
$\times 2$	PSNR↑	17.75	23.71	23.46	23.72	23.70	23.74	23.81	23.84	23.87
	SSIM↑	0.171	0.350	0.348	0.351	0.350	0.352	0.352	0.354	0.352
	Flux-L1↓	1916	213.6	165.9	226.3	216.1	223.1	193.3	192.5	198.7

Fixed-point Noise Amplification. For completeness, we record the $N \to \infty$ behaviour. In that limit, Equation (28) collapses to $X^{(\infty)}(\mathbf{k}) = \Psi + \Gamma(\mathbf{k})/H(\mathbf{k})$, whose noise spectrum $\Phi_{\text{rec}}(\mathbf{k}) = \sigma_n^2 \exp(\sigma_h^2 |\mathbf{k}|^2)$ diverges exponentially at high $|\mathbf{k}|$. Bandlimiting and integrating with the continuous angular-frequency Parseval factor $1/(2\pi)^2$ yields the closed-form per-pixel variance:

$$\sigma_{\text{rec}}^2 \;=\; \frac{\sigma_n^2}{4\pi\,\sigma_h^2}\big(\exp(\sigma_h^2 k_{\max}^2) - 1\big), \tag{32}$$

which evaluates to roughly $0.21\,\sigma_n^2$ at the DESI–HST setting ($\sigma_h = 2$, $k_{\max} = \pi/4$). While modest in absolute terms, this corresponds to a substantial amplification of the noise power that lives within the narrow LR-supported band $|\mathbf{k}| \le k_{\max}$, and the prefactor $\exp(\sigma_h^2 k_{\max}^2) - 1$ grows exponentially as $k_{\max}$ approaches the Nyquist limit, so this term remains a real concern for any iteration run to convergence. The practical N used in our pipeline keeps the recovery short of this fixed point. In our regime the dominant pathology is therefore the finite-N low-pass effect of Equation (29) rather than the asymptotic noise amplification of Equation (32).

G Additional Quantitative Results

We further evaluate FluxFlow on two existing astronomical super-resolution benchmarks. The AstroSR dataset Miao et al. [2024] pairs galaxy cutouts across two different ground-based optical surveys, drawing low-resolution images from the Sloan Digital Sky Survey (SDSS) York et al. [2000] and high-resolution targets from the Subaru Hyper Suprime-Cam survey (HSC) Aihara et al. [2018], and is distributed as three-channel RGB composites rather than flux-calibrated single-band photometry. The STAR dataset Wu et al. [2025] consists of space-based HST star-field cutouts in the F814W band Sirianni et al. [2005], whose low-resolution counterparts are produced by synthetic downsampling of the same HST images. Neither benchmark exposes the severe atmospheric PSF broadening that characterizes genuine ground-to-space observation pairs, as also summarized in Table 1, nor a calibrated atmospheric PSF kernel for the LR to HR transformation, so the LR to HR gap is much milder than the seeing-limited regime our full pipeline targets. We therefore set the PSF kernel h in the forward model of Equation (6) to the identity $\mathbf{H}_{\text{PSF}} = \mathbf{I}$ for these experiments, so that the inference-time MC-FS correction reduces to a residual back-projection through the area-downsampling adjoint $\mathbf{D}_s^\top$ alone and no Wiener deconvolution is applied. In addition, neither benchmark provides per-pixel inverse-variance weight maps or co-registered HR catalogs from which to derive source masks, so we omit both the inverse-variance term $\widetilde{\mathbf{W}}$ and the source-importance term $\mathbf{S}$ from the WFM objective in Equation (5), recovering the unweighted CFM loss of Equation (4). All other architectural and optimization choices follow Section 6.

The Flux-L1 metric in this section is defined consistently with Table 2 in the main text. We deliberately report the Flux-L1, defined as Equation (15), rather than the per-source averaged Flux Error (FE) defined in STAR Wu et al. [2025] for two reasons. First, our DESI to HST mapping operates in

calibrated nanomaggie units whose per-pixel values are extremely small. The sum-based formulation therefore preserves a meaningful range and remains consistent with the definition adopted in Table 2, while dividing by the number of recovered sources would compress the metric into a numerically uninformative regime. Second, individual sources span a very wide range of extent, brightness, and morphology, from compact point-like profiles to extended galaxies. Under such heterogeneity, a per-source mean weights a faint compact detection and a bright extended structure equally and obscures where the photometric error actually concentrates. Aggregating the residual as a sum over all pixels instead reflects the cumulative photometric fidelity of the field as a whole.

We further note an important difference in the evaluation protocol. The STAR benchmark Wu et al. [2025] adopts a per-image min-max normalization that depends on the ground-truth extrema of each test image, which is not reproducible at inference time when the ground truth is unavailable. To keep the evaluation usable in a deployment setting and consistent with the rest of the paper, we instead adopt the same global normalization protocol used for our DESI–HST experiments described in Appendix E, namely a 0.01% tail clip followed by global min-max rescaling fitted on the training split, and we report PSNR and SSIM in the normalized domain and Flux-L1 in the denormalized native units. For the AstroSR benchmark Miao et al. [2024], the data are distributed as 8-bit RGB composites with consistent scaling across SDSS and HSC, so no further normalization is required and we evaluate directly on the released values.

We evaluate the FluxFlow network on STAR and AstroSR, with results reported in Tables 9 and 10. As specified above, these configurations use an identity PSF and omit Wiener deconvolution, inverse-variance weighting, and source weighting. On STAR at $\times 2$, the reported configuration leads in PSNR and SSIM , while HAT and FISR have lower Flux-L1. It does not lead the $\times 4$ table. On AstroSR at $\times 2$, it leads in PSNR but not SSIM or Flux-L1. AstroSR 's RGB composites also limit the interpretation of Flux-L1 as physical photometry. These metric-dependent results support transfer of the conditional backbone to synthetic space-to-space and cross-instrument ground-to-ground settings. They do not establish zero-shot transfer of the full method or superiority across degradation types. The main results concern real ground-to-space reconstruction under atmospheric seeing. Comparing these different datasets alone does not isolate the cause of the performance differences.

H Additional Ablation Study

This section presents additional ablations beyond the main study of Table 3. We first examine two alternatives to pixel-space OT-CFM, a latent-space flow-matching variant in a frozen SD-VAE sdv [2022] feature space and a conditional-bridge variant that replaces the Gaussian source with a noisy bicubic-upsampled DESI input. We then sweep the three inference-time hyperparameters of MC-FS, the forward Gaussian PSF width σ_{PSF}, the Wiener regularization λ_{SNR}, and the correction step size η_0. We close with detection-overlay visualisations on six DESI–HST patches showing how MC-FS suppresses spurious sources relative to naive OT-CFM sampling.

Latent Space Flow-Matching. Many recent image generative models operate in the latent space of a pretrained VAE Diederik and Max [2019], which compresses input images and turns the diffusion process into a much lower-dimensional regression problem Rombach et al. [2022]. We test whether the same recipe transfers to astronomical super-resolution using SD-VAE sdv [2022]. The pixel-normalization protocol of Section E maps every cutout to $[0, 1]$ before encoding, so the SD-VAE sees inputs in its native operating range. The *w/ VAE* row of Table 3 reports the result. All photometric and detection metrics degrade substantially relative to pixel-space FluxFlow, with detection F1 collapsing primarily through a sharp drop in recall, indicating that compact sources are systematically washed out by the encoder rather than misplaced. We attribute this collapse to the VAE bottleneck smoothing out the compact, high-contrast point sources that dominate the HST signal. Figure 17 shows the visual signature. The latent-space variant blurs faint compact sources and oversmooths galaxy profiles, recovering only the low-frequency structure that survives the VAE bottleneck, while pixel-space FluxFlow recovers morphology and resolved structure that closely match the HST target. We therefore retain pixel-space training in the main model. For reproducibility, the variant encodes the bicubically upsampled DESI input and the HST target with the frozen SD-VAE encoder into latent codes z_x and z_y, then trains a conditional U-Net along the OT-CFM interpolant $z_t = (1 - t)\, z_0 + t\, z_y$ from a Gaussian prior $z_0 \sim \mathcal{N}(0, I)$, decoding back to pixel space at inference and averaging the three channels into a single-channel HR estimate. The latent U-Net uses 8 input and 4 output channels with

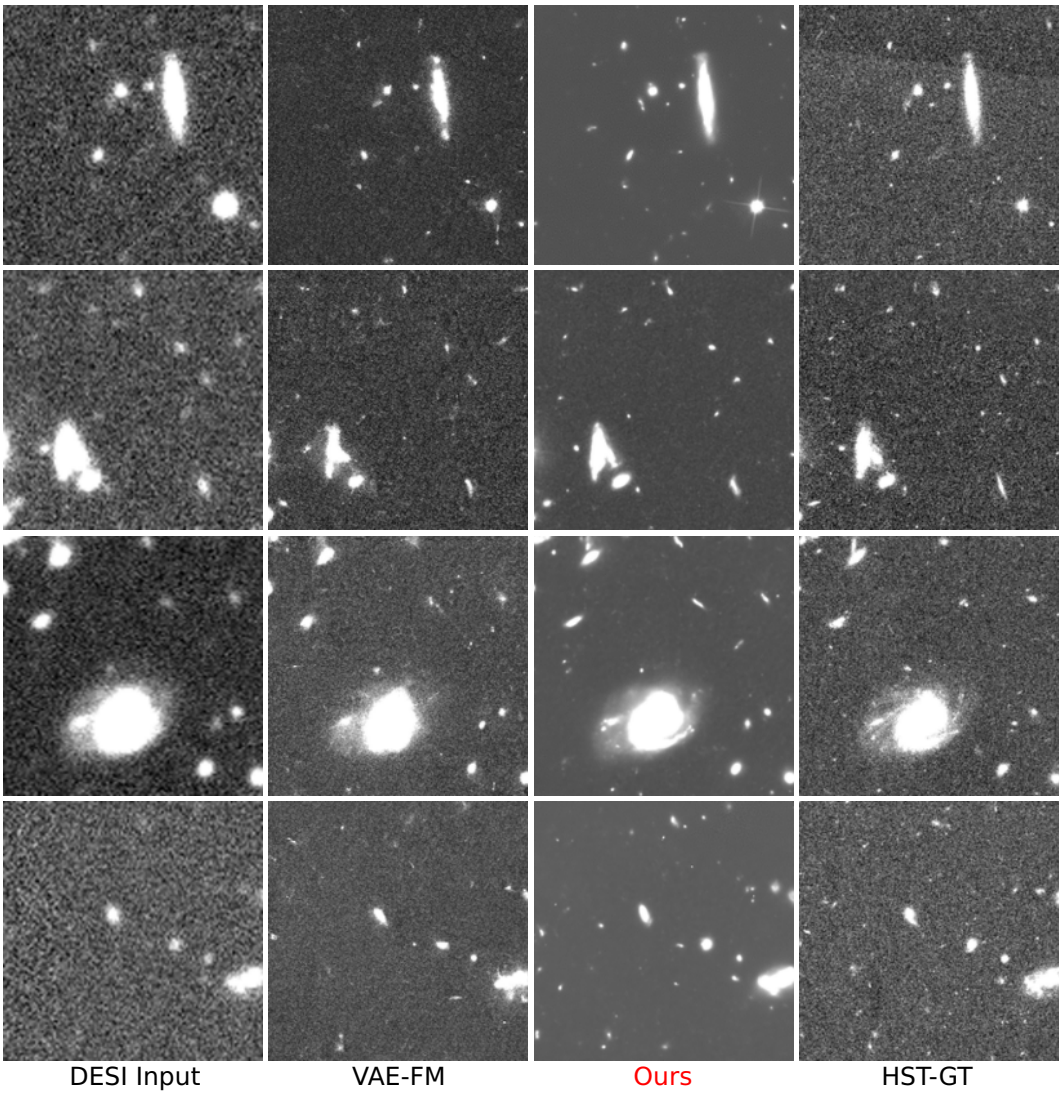
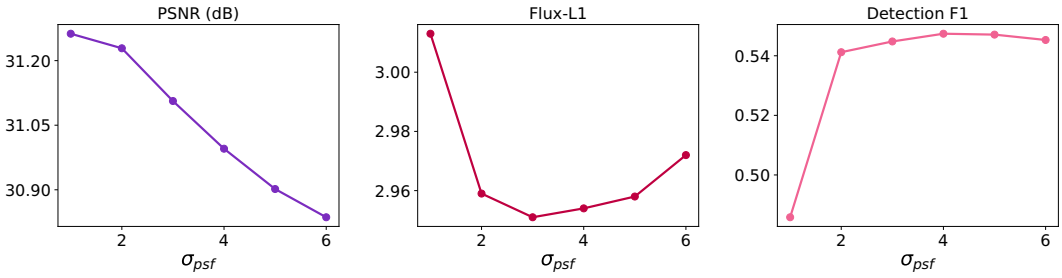


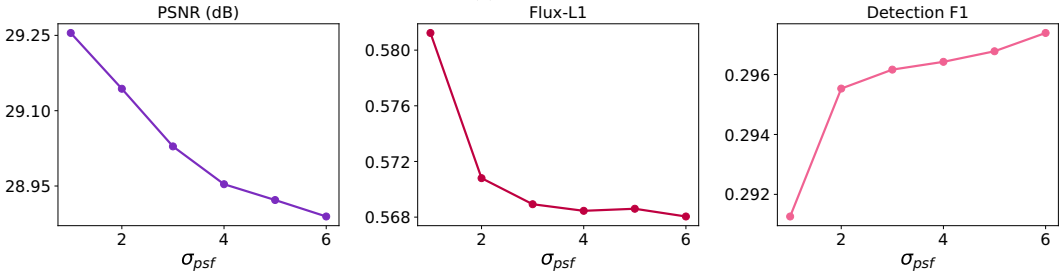
Figure 17: Qualitative comparison of the latent-space flow-matching ablation. Four representative DESI–HST test patches are shown from left to right as the DESI input, the latent-space FM variant operating on SD-VAE features, FluxFlow operating directly in pixel space, and the HST observations. The VAE-FM blurs faint compact sources and oversmooths the galaxy profiles, while pixel-space FluxFlow recovers source morphology and resolved structure that closely match the HST target.

base_dim $= 64$ and time_dim $= 256$, trained with AdamW for 300 epochs at batch size 88, learning rate 10^{-4}, weight decay 10^{-4}, 1000 warmup steps, gradient clipping at 1.0, and no EMA.

Conditional-Bridge Flow-Matching. The Gaussian source distribution of OT-CFM places the inference-time trajectory at pure noise, forcing the velocity network to recover the entire HR signal across certain Euler steps. A natural alternative is to start from a coarse condition-aligned image and refine the residual. We test this by replacing the Gaussian source with a noisy bicubic upsample of the DESI observation, $\mathbf{x}_0 = \mathrm{BicubicUp}(\mathbf{y}) + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, 0.4^2\mathbf{I})$, where the noise scale 0.4 was tuned on a small validation grid. The *w/ Bridge* row of Table 3 is intermediate. All photometric and detection metrics degrade modestly relative to the pixel-space OT-CFM model, with smaller gaps than the latent-space variant but consistent in direction across PSNR, Flux-L1, and detection F1. The interpretation is that the bridge starting point indeed helps the velocity field locally but couples the prior too tightly to the bicubic mean, suppressing the high-frequency content that pixel-space



(a) $\times 2$ scale.



(b) $\times 4$ scale.

Figure 18: Effect of the forward PSF width σ_{PSF} used inside MC-FS at $\times 2$ (top) and $\times 4$ (bottom). PSNR decreases monotonically with σ_{PSF}, while Flux-L1 and Detection F1 form broad plateaus over an intermediate range. We set $\sigma_{\mathrm{PSF}} = 2$ HR pixels at both scales in our experiments.

OT-CFM recovers under MC-FS guidance from a noise prior. We therefore retain the Gaussian source in the main model.

Effect of Forward PSF Kernel σ_{PSF}. The forward operator $\mathbf{A} = \mathbf{D}_s \circ \mathbf{H}_{\mathrm{PSF}}$ used inside MC-FS approximates the DESI seeing-limited PSF by a single isotropic Gaussian of width σ_{PSF}, which controls how aggressively the Wiener correction back-projects the LR residual onto the HR grid. Figure 18 sweeps σ_{PSF} from 1 to 6 HR pixels at both scales while the rest of the pipeline is held fixed. PSNR decreases monotonically with σ_{PSF} at both scales because a wider forward model relaxes the high-frequency agreement with the DESI observation. Flux-L1 and Detection F1 follow a different trade. At $\times 2$, Flux-L1 reaches its minimum near $\sigma_{\mathrm{PSF}} \in [3, 5]$ (~ 2.95) and Detection F1 forms a broad plateau across $\sigma_{\mathrm{PSF}} \in [2, 6]$ (between 0.541 and 0.547), dropping sharply only at the very narrow setting $\sigma_{\mathrm{PSF}} = 1$ (0.486) where over-aggressive deconvolution destroys faint sources. The $\times 4$ trend is qualitatively similar but flatter: PSNR decreases smoothly from 29.25 dB at $\sigma_{\mathrm{PSF}} = 1$ to 28.89 dB at $\sigma_{\mathrm{PSF}} = 6$, while Flux-L1 ($\sim 0.57$) and Detection F1 ($\sim 0.296$) sit on near-flat plateaus across $\sigma_{\mathrm{PSF}} \geq 2$. Our default $\sigma_{\mathrm{PSF}} = 2$ at both scales preserves the high-PSNR end while staying inside the Flux-L1 and F1 plateaus, which we found gave the most consistent reconstructions across patches with widely varying DESI seeing.

Effect of Wiener Regularization λ_{SNR}. The parameter λ_{SNR} in Equation (10) controls regularization of the correction. For small values the filter vanishes in magnitude while remaining proportional to $H^*(\mathbf{k})$. At large values it approaches an inverse where $H(\mathbf{k}) \neq 0$. Figure 19 reports a sweep from 10 to 100 at $\times 2$ with the other settings fixed. PSNR , Flux-L1 , and Detection F1 vary little over this tested range, which contains the default 50. This is empirical robustness under the calibrated Gaussian model, not a guarantee outside the tested interval or under PSF misspecification. The spectral interpretation depends on frequency. For frequencies satisfying $|H(\mathbf{k})|^2 \gg \lambda_{\mathrm{SNR}}^{-1}$, changing the regularizer has little effect on the filter denominator.This explains why a plateau can occur without making the parameter irrelevant. It is not a claim that every spatial frequency satisfies that inequality. Anisotropic or spatially varying PSFs can have directional attenuation and position-dependent spectra, so one global Fourier-domain filter and the observed robustness need not carry over to those settings.

Effect of Wiener Correction Step Size η_0. The base step size η_0 controls how aggressively each MC-FS iteration projects the candidate state toward the consistency set $\mathcal{C}(\mathbf{y})$, with the schedule

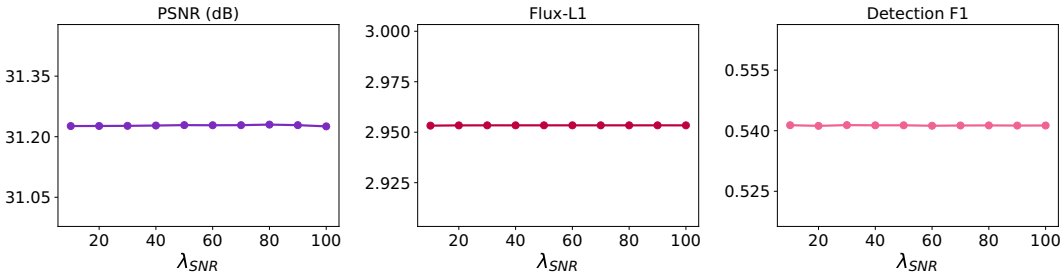


Figure 19: Effect of the Wiener regularization λ_{SNR} at $\times 2$. Sweeping λ_{SNR} from 10 to 100 leaves PSNR, Flux-L1, and Detection F1 essentially unchanged, under the calibrated Gaussian forward model. Our default $\lambda_{\mathrm{SNR}} = 50$ is the value used throughout Table 2.

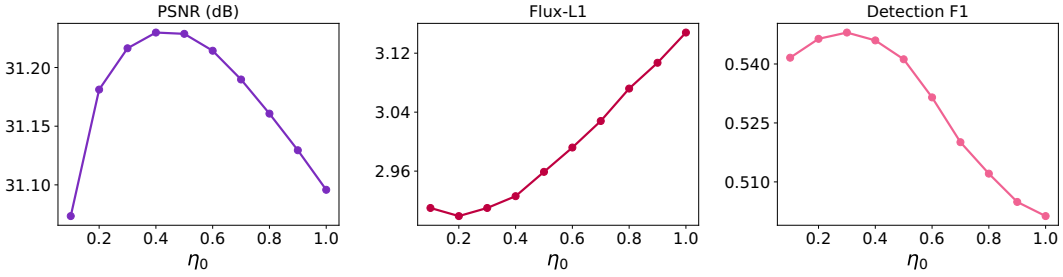


Figure 20: Effect of the base Wiener correction step size η_0 at $\times 2$, with the per-step decay $\eta_i = \eta_0(1 - i/N)$ otherwise unchanged. PSNR is concave with a peak near $\eta_0 \approx 0.4$, Flux-L1 reaches its minimum near $\eta_0 = 0.2$ and then rises, and Detection F1 peaks near $\eta_0 \approx 0.3$ before falling. Our default $\eta_0 = 0.5$ is the value used throughout Table 2 and sits just to the right of the joint optimum, biasing toward a slightly more conservative correction.

$\eta_i = \eta_0(1 - i/N)$ giving a final applied coefficient $\eta_{N-1} = \eta_0/N$. The continuous schedule reaches zero only at $t = 1$. Figure 20 sweeps η_0 from 0.1 to 1.0 at $\times 2$ with all other settings fixed. PSNR is concave, peaking at $\eta_0 = 0.4$ with 31.230 dB and degrading to 31.073 and 31.096 dB at $\eta_0 = 0.1$ and $\eta_0 = 1.0$. Too small a step under-corrects the residual and leaves hallucinated structure in place. Too large a step over-shoots into a non-physical region of $\mathcal{C}(\mathbf{y})$ and injects artificial flux. Flux-L1 confirms this over-shoot, rising from its minimum 2.899 at $\eta_0 = 0.2$ to 3.148 at $\eta_0 = 1.0$. Detection F1 peaks at $\eta_0 = 0.3$ (0.548) and falls steadily for $\eta_0 \geq 0.5$ as over-correction begins to suppress real low-SNR sources alongside hallucinations. Our default $\eta_0 = 0.5$ sits just to the right of the joint optimum, retaining the original reconstruction setting; this sweep does not establish robustness to arbitrary seeing variation.

Effect of Number of Sampling Steps T. The number of sampling steps T (denoted N in Algorithm 1) sets the discretization of the OT-CFM ODE and, jointly with the linear decay $\eta_i = \eta_0(1 - i/T)$, the per-step magnitude of the Wiener correction. Figure 21 sweeps T from 5 to 60 at $\times 2$ with all other settings fixed. PSNR is highest at $T = 5$ (31.26 dB) and decreases monotonically to 30.76 dB at $T = 60$, while Flux-L1 attains its minimum at $T = 10$ (2.96) and rises steadily to 3.40 at $T = 60$. Detection F1 peaks at $T = 10$ (0.541) and falls to 0.466 at $T = 60$. Changing T alters both ODE discretization and the number of corrections. At a fixed relative time i/T, the stated schedule has no $1/T$ factor; these measurements do not isolate the cause of the trend. Our default $T = 10$ sits at the joint optimum of Flux-L1 and Detection F1 with PSNR within a small margin of its peak Table 2.

Visualization of Source Detection. Figures 22 and 23 visualize how MC-FS suppresses hallucination sources across several representative DESI–HST test images. Each example uses four columns showing the DESI input, naive OT-CFM sampling without MC-FS, our MC-FS sampling, and the HST ground-truth, with the top row giving the reconstruction and the bottom row overlaying the SExtractor catalogue measured on the same image. Cyan circles mark ground-truth sources, green circles mark true-positive detections, and red dots mark spurious detections without a counterpart

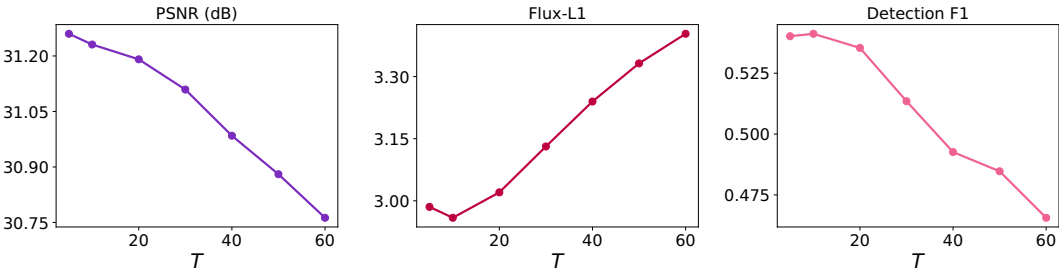


Figure 21: Effect of the number of sampling steps T at $\times 2$, with all other MC-FS hyperparameters held at their defaults. PSNR decreases monotonically with T. Flux-L1 and Detection F1 attain their joint optimum at $T = 10$ before degrading at larger T in this experiment. Our default $T = 10$ is the value used throughout Table 2.

in the ground truth. The naive column is densely populated with red dots in low-SNR background regions and around the wings of bright stars, the regimes in which the velocity field most readily sharpens noise into compact source-like morphology with no preimage in the DESI observation. The MC-FS column eliminates most of these red dots while preserving the green true-positive population, recovering precision without collapsing recall. The suppression is most pronounced on the saturated-star examples in Figure 22, where strong diffraction spikes induce many spurious sources under naive sampling but are correctly recognized as instrumental structure once the residual against the DESI observation is enforced. Figure 23 shows the same pattern on a bright extended galaxy and a faint background field, where naive sampling fragments the extended profile and hallucinates compact peaks in the surrounding noise. These visualizations are the per-image counterpart of the precision and F1 gains of the full model over the *w/o MCFS* ablation in Table 3.

I Downstream Scientific Application

We complement the field-aggregated metrics in Table 2 with a per-source evaluation that converts each reconstruction into the quantities that actually enter downstream catalog products. The relevant quantities are AB magnitudes for photometric calibration, a luminosity-proxy stellar mass for galaxy population statistics, and second-moment ellipticities for weak-lensing shape catalogs. Sources are read from a precomputed SEP Bertin and Arnouts [1996], Barbary [2016] segmentation map of the HST , with labels of fewer than 10 pixels discarded. Every reconstruction is measured under the same segmentation footprint $\mathcal{R}_i$ as the ground truth, so each prediction inherits a strictly one-to-one source correspondence with the GT and detector-side variability does not enter the comparison.

Local Background Subtraction. Before any per-source measurement, we subtract a local background mesh from both the ground truth and each reconstruction. The mesh is computed with SEP's `Background` routine on a 32×32 pixel grid with mesh-size parameters bw $=$ bh $= 32$ and mesh-filtering parameters fw $=$ fh $= 1$ that disable additional smoothing. The per-pixel weight map $\mathbf{W}$ acts as a validity mask in which pixels at $\mathbf{W}_p = 0$ are excluded from both the background fit and the subsequent measurement. The HST has already had its background removed during SEP preprocessing, but the residual background bias of each reconstruction is method-dependent and systematic, negative for bicubic upsampling and positive for several generative baselines, and would otherwise translate into a global flux offset that pushes a non-trivial fraction of $\hat{F}_i$ negative and out of the magnitude domain. All formulas below operate on the background-subtracted image.

Bright-Source Protocol. For the scientific tasks considered here the relevant population is the bright, resolved end where photometric, structural, and shear estimates are stable rather than dominated by sub-detection-limit fluctuations. We therefore retain only sources whose ground-truth integrated flux satisfies

$$F_i^{\mathrm{GT}} = \sum_{p \in \mathcal{R}_i} x_p^{\mathrm{GT}} \geq F_{\mathrm{thr}}, \tag{33}$$

on the background-subtracted GT, with $\mathcal{R}_i$ the SEP segmentation footprint of source i. We use the same threshold $F_{\mathrm{thr}} = 0.5\,\mathrm{nMgy}$ at both $\times 2$ and $\times 4$. Because the integrated flux of a fixed physical

Table 11: Per-source downstream-task accuracy at $\times 2$ on the matched bright-source set. Magnitude and shear are dimensionless, while the stellar-mass MAE is reported in $\Theta_\star$-scaled nanomaggies as defined in Equation (36). Best, second, and third are shaded with ■, ■, ■.

	$\Delta m \downarrow$		Mass MAE $\downarrow$		$\Delta\gamma \downarrow$	
Method	mean	median	mean	median	mean	median
Bicubic Keys [1981]	0.758	0.637	7.767	1.537	0.148	0.122
SwinIR Liang et al. [2021]	0.649	0.300	8.197	0.762	0.149	0.116
HAT Chen et al. [2025b]	0.542	0.192	7.962	0.524	0.151	0.112
FISR Wu et al. [2025]	0.561	0.230	8.014	0.608	0.165	0.133
cGAN Rai et al. [2025]	0.548	0.215	7.996	0.586	0.159	0.131
GD-Net Shan et al. [2025]	0.664	0.393	7.659	1.025	0.166	0.131
Ours	0.516	0.187	7.632	0.508	0.161	0.126

source is conserved across the two scales by construction of the dataset, the same threshold selects the same physical population, yielding $N_{\mathrm{src}} = 7437$ matched bright sources at both scales. All metrics below are reported as both the mean and the median across this set, the median being the relevant robust statistic for catalog products in which a small number of saturated or hallucinated cores would otherwise dominate the mean.

Magnitude Error. For each source the integrated flux on the reconstruction is $\hat{F}_i = \sum_{p\in\mathcal{R}_i} \hat{x}_p$, measured on the background-subtracted image and the same footprint as F_i^{GT}. The AB-style magnitude is computed with a calibrated zero point $Z_{\mathrm{p}} = 25$,

$$m_i \;=\; -2.5\,\log_{10}(F_i) + Z_{\mathrm{p}}, \tag{34}$$

and the per-source magnitude residual is

$$\Delta m_i \;=\; \left|\hat{m}_i - m_i^{\mathrm{GT}}\right| \;=\; 2.5\left|\log_{10}\hat{F}_i - \log_{10} F_i^{\mathrm{GT}}\right|, \tag{35}$$

evaluated only on sources with both $\hat{F}_i > 0$ and $F_i^{\mathrm{GT}} > 0$. This is the standard photometric residual used for source-catalog calibration and for color-magnitude diagrams. The logarithmic scaling makes Δm sensitive to multiplicative photometric bias, so a method that systematically biases integrated flux high or low is penalized regardless of the absolute scale.

Stellar-Mass Error. A bright resolved source contributes to galaxy stellar-mass functions and SFR-mass diagrams Conselice [2014] primarily through its luminosity. We adopt a single-band luminosity-proxy stellar mass with a fiducial mass-to-light ratio $\Theta_\star$ held common across reconstructions and ground truth,

$$\hat{M}_{\star,i} \;=\; \Theta_\star \cdot \hat{F}_i, \qquad \Theta_\star = 3, \tag{36}$$

on the same nanomaggie-scale flux as Equation (34). The mean absolute error is

$$\mathrm{Mass\,MAE} \;=\; \frac{1}{N_{\mathrm{src}}} \sum_{i=1}^{N_{\mathrm{src}}} \left|\,\hat{M}_{\star,i} - M_{\star,i}^{\mathrm{GT}}\,\right|, \tag{37}$$

reported in the same unit as $\Theta_\star \cdot F$, namely $\Theta_\star$-scaled nanomaggies. Unlike Δm, the linear form weights bright sources more heavily than faint ones, exposing methods that smear or hallucinate the high-luminosity end of the population, the regime that dominates galaxy stellar-mass-function inference.

Weak-Lensing Shear Error. Cosmic-shear analyses depend on the recovered shape of background galaxies Bernstein and Jarvis [2002], Mandelbaum et al. [2015]. For each source we compute flux-weighted second-order brightness moments on the background-subtracted reconstruction, with positivity-clipped weights to prevent residual negative-noise pixels from biasing the centroid,

$$\omega_p \;=\; \max(\hat{x}_p, 0), \qquad \bar{p}_{a,i} \;=\; \frac{\sum_{p\in\mathcal{R}_i} \omega_p\, p_a}{\sum_{p\in\mathcal{R}_i} \omega_p}, \qquad a \in \{x, y\}. \tag{38}$$

The second-moment tensor is then

$$Q_{ab,i} \;=\; \frac{\sum_{p\in\mathcal{R}_i} \omega_p\,(p_a - \bar{p}_{a,i})(p_b - \bar{p}_{b,i})}{\sum_{p\in\mathcal{R}_i} \omega_p}, \qquad a, b \in \{x, y\}, \tag{39}$$

Table 12: Per-source downstream-task accuracy at $\times 4$ on the same matched bright-source set as Table 11. Stellar-mass MAE is in $\Theta_\star$-scaled nanomaggies.

	$\Delta m \downarrow$		Mass MAE $\downarrow$		$\Delta\gamma \downarrow$	
Method	mean	median	mean	median	mean	median
Bicubic Keys [1981]	1.067	0.863	16.90	3.648	0.144	0.115
SwinIR Liang et al. [2021]	0.969	0.461	16.51	2.43	0.118	0.090
HAT Chen et al. [2025b]	0.886	0.397	16.19	2.207	0.108	0.081
FISR Wu et al. [2025]	0.876	0.417	16.25	2.246	0.124	0.097
cGAN Rai et al. [2025]	0.902	0.460	16.42	2.440	0.135	0.109
GD-Net Shan et al. [2025]	1.509	1.308	17.58	4.630	0.170	0.139
Ours	0.851	0.389	16.07	2.131	0.117	0.091

falling back to a uniform weight whenever $\sum_p \omega_p = 0$. The two ellipticity components Bernstein and Jarvis [2002] are

$$\hat{e}_{1,i} = \frac{Q_{xx,i} - Q_{yy,i}}{Q_{xx,i} + Q_{yy,i}}, \qquad \hat{e}_{2,i} = \frac{2\,Q_{xy,i}}{Q_{xx,i} + Q_{yy,i}}, \tag{40}$$

and the per-source shear error is the magnitude of the residual ellipticity vector

$$\Delta\gamma_i = \sqrt{\left(\hat{e}_{1,i} - e_{1,i}^{\mathrm{GT}}\right)^2 + \left(\hat{e}_{2,i} - e_{2,i}^{\mathrm{GT}}\right)^2}, \tag{41}$$

where $e_{a,i}^{\mathrm{GT}}$ is the same statistic measured on the HST observations. The median of $\{\Delta\gamma_i\}$ is the relevant statistic for cosmic-shear pipelines, where bulk biases on the bright end propagate directly into the inferred lensing power spectrum. Measuring the moments on the GT segmentation footprint, rather than re-running source detection on the reconstruction, isolates the per-source shape error from the orthogonal failure mode of catalog-level miscentering.

Discussion. At both $\times 2$ and $\times 4$, FluxFlow obtains the best mean and the best median on four of the six metrics, dominating the magnitude error and the luminosity-proxy stellar-mass MAE on the same matched bright-source set. The remaining shear advantage belongs to HAT, SwinIR, and bicubic upsampling, whose conservative mean-image behaviour preserves elliptical shapes through smoothing rather than recovering high-frequency structure. FluxFlow nevertheless places second on shear mean at $\times 4$ at 0.117 against the HAT best of 0.108 and third on shear median at 0.091 against the HAT best of 0.081, and the corresponding $\times 2$ gap is similarly small at roughly 8% of the best value. The systematic magnitude-error and stellar-mass advantages indicate that the inverse-variance and source-region weighting of Equation (5) together with the Wiener-corrected MC-FS sampling translate directly into the per-source flux fidelity that downstream catalog products require, while costing only a small amount of shape bias relative to the smoothest regression baselines.

J Additional Visualization

We provide additional qualitative comparisons against all evaluated baselines on the DESI–HST test set. Figures 24 and 25 present $\times 2$ super-resolution results, the former on a bright central star with an extended companion and the latter on a saturated star surrounded by closely packed compact sources. Figures 26 and 27 extend the comparison to $\times 4$, covering a faint background-galaxy field with a saturated upper-left source and a binary-star field with resolved diffraction spikes. Across all four cases, FluxFlow recovers diffraction patterns, faint background structures, and source morphology that regression baselines oversmooth and that generative baselines either speckle or oversharpen, consistent with the trends reported in Table 2.

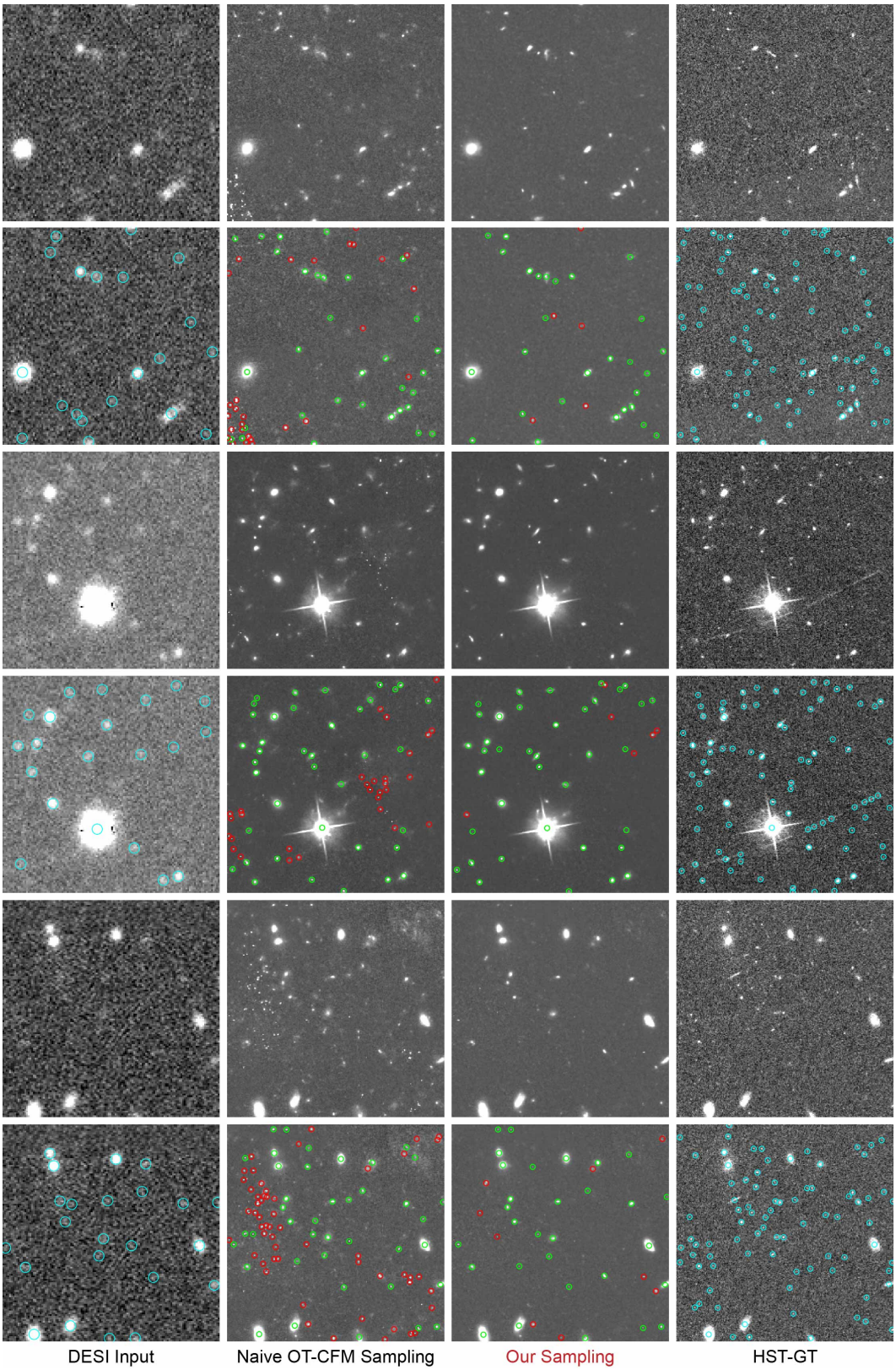


Figure 22: MC-FS hallucination suppression on three DESI–HST examples. Columns show the DESI input, naive OT-CFM, MC-FS, and HST . The second row overlays SExtractor detections, with Cyan for ground-truth sources, Green for true positives, and Red for false positives.

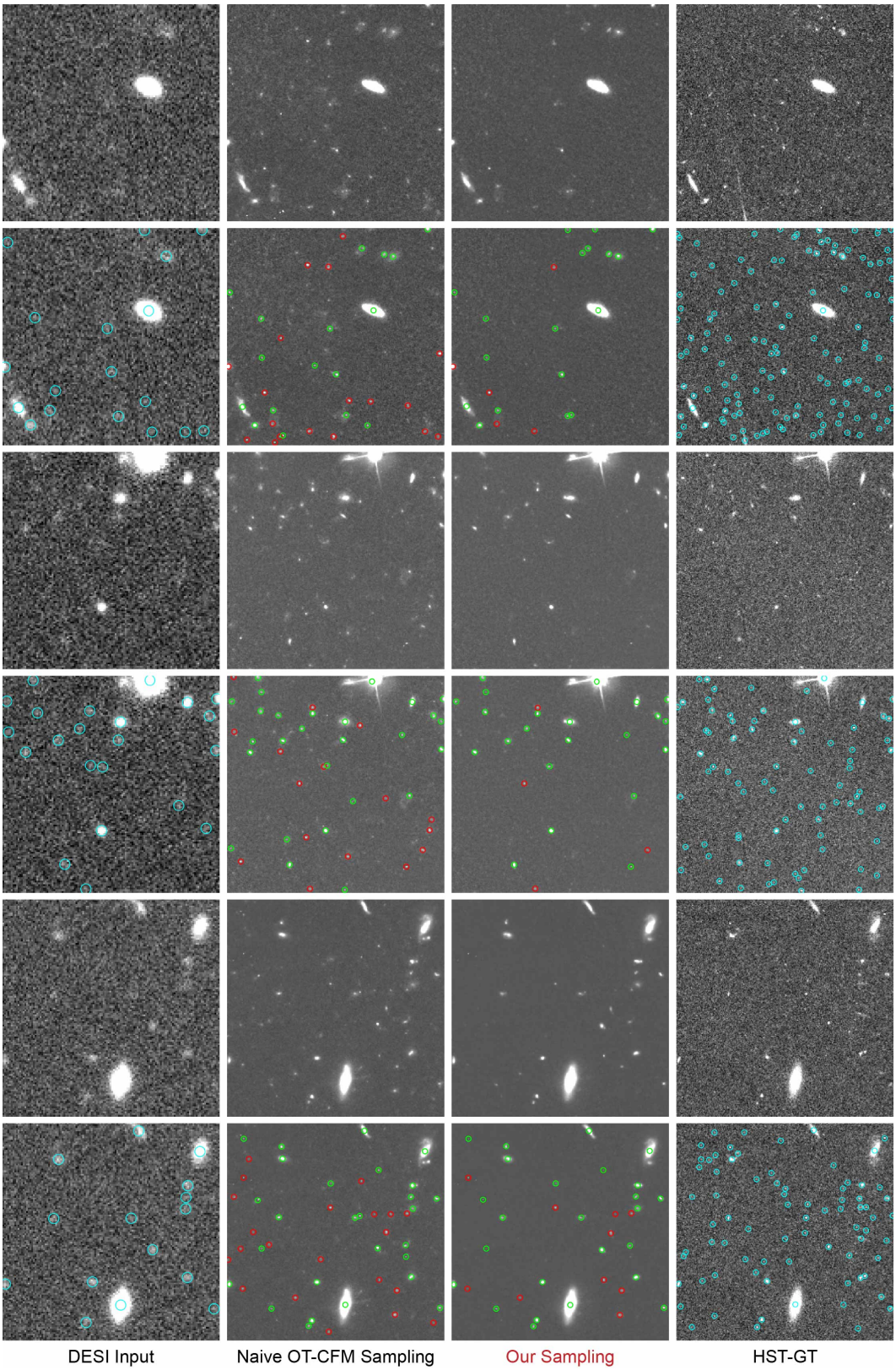


Figure 23: Additional MC-FS detection comparisons under the protocol of Figure 22, covering a faint background field, a saturated bright star with diffraction spikes, and a moderately crowded region. Cyan for ground-truth sources, Green for true positives, and Red for false positives.

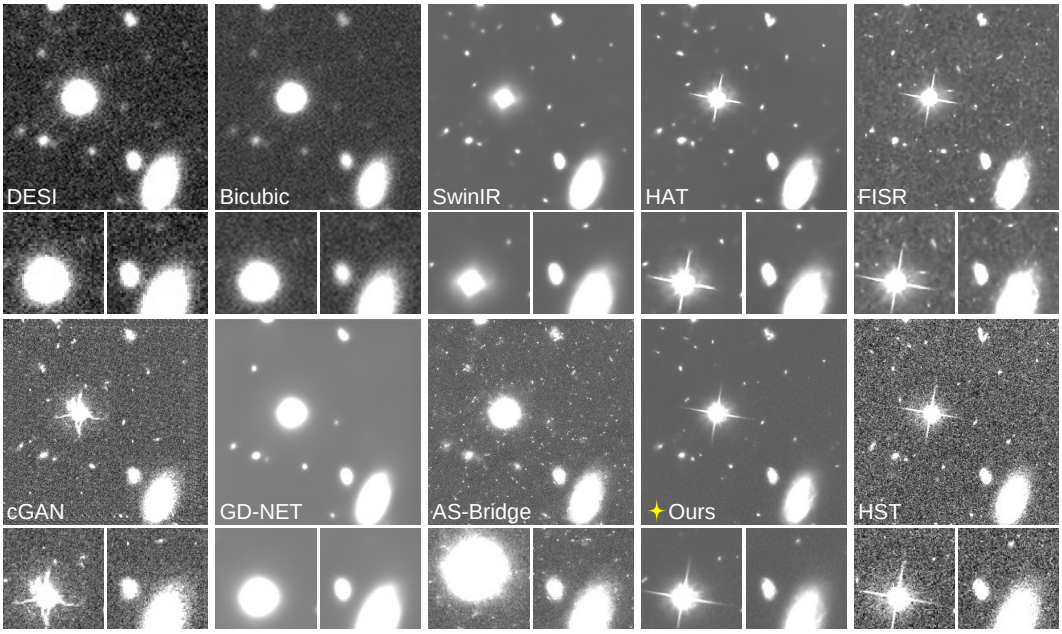


Figure 24: Qualitative DESI–HST $\times 2$ comparison of all evaluated methods, with two zoom-in panels per method. FluxFlow recovers the diffraction spikes of the central star and the morphology of the extended source while avoiding the speckled background of AS-Bridge.

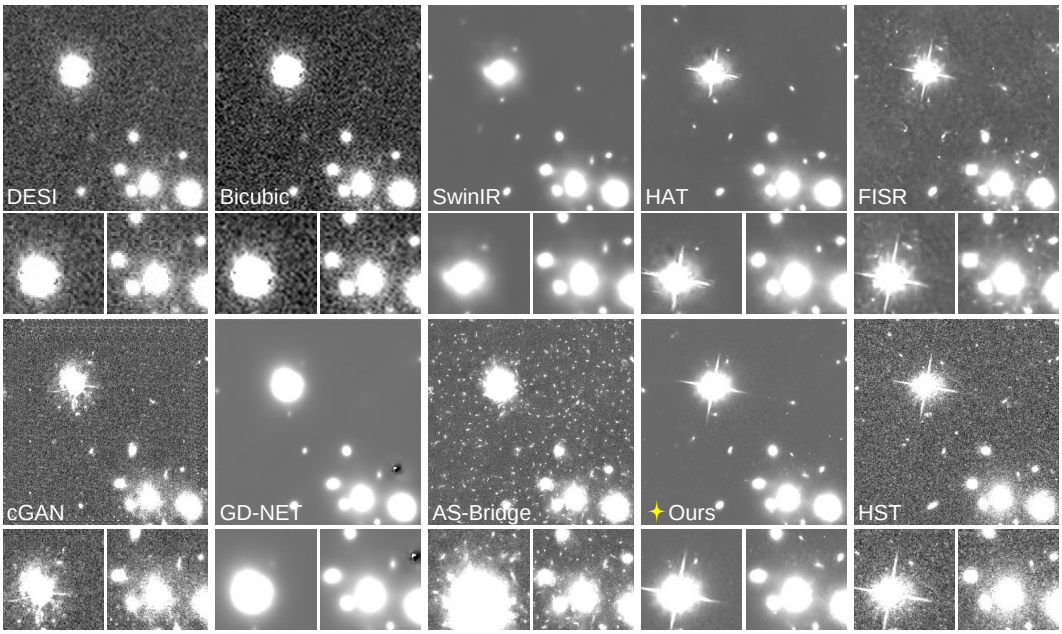


Figure 25: Additional DESI–HST $\times 2$ comparison, on a bright central star surrounded by closely packed compact sources. FluxFlow reproduces the spike pattern and resolves the neighboring sources without the merging or smearing seen in the baselines.

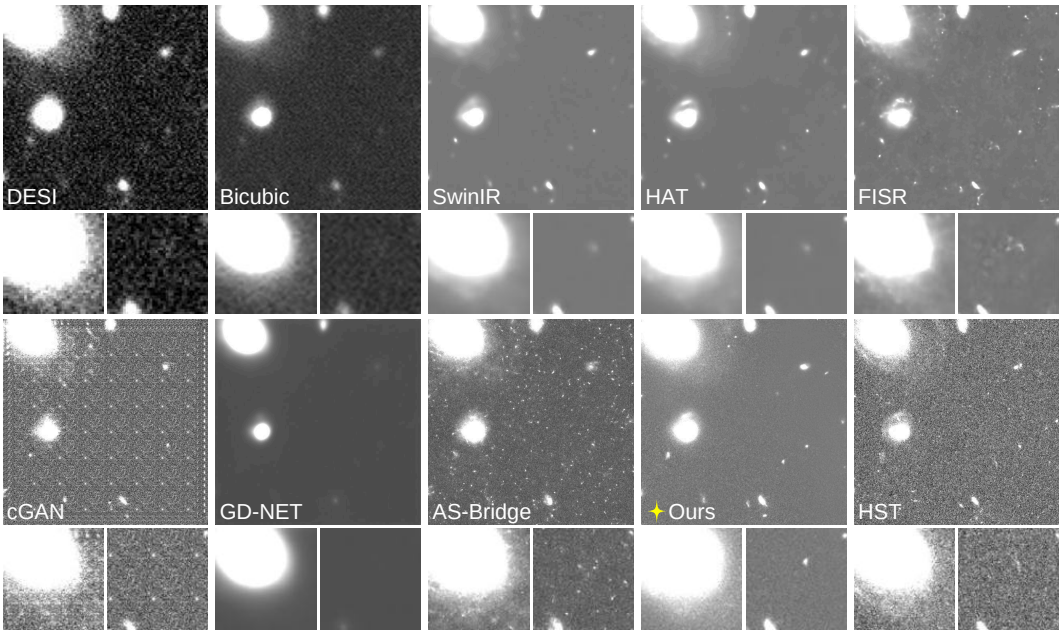


Figure 26: DESI–HST $\times 4$ comparison on a field with a saturated source on the upper-left corner and several low-SNR background galaxies. FluxFlow preserves faint background galaxies that Bicubic and GD-Net erase while reconstructing the saturated halo more cleanly than the speckled output of AS-Bridge.

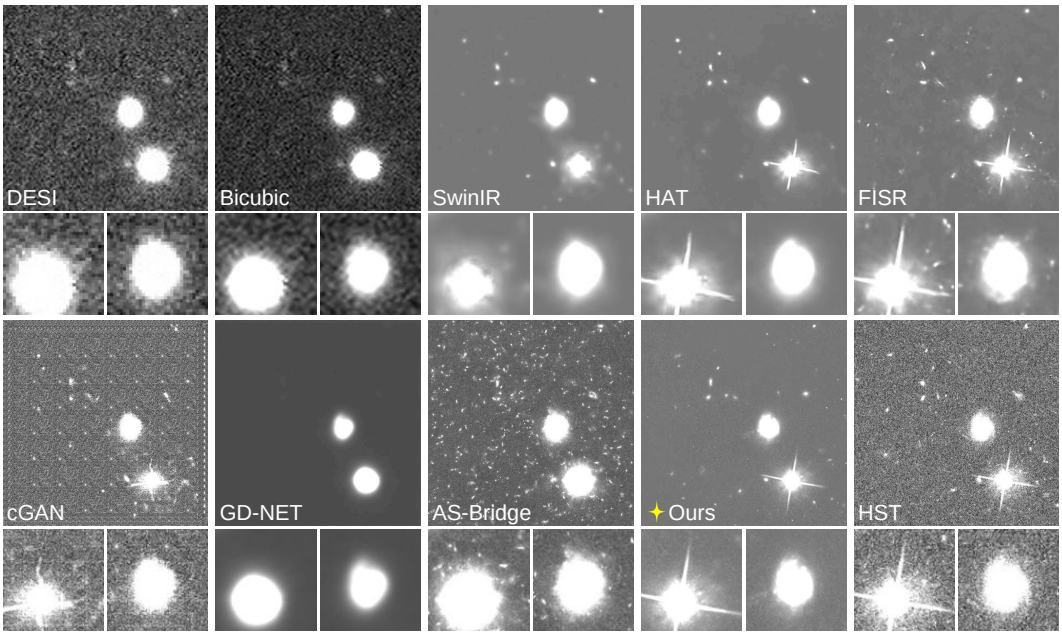


Figure 27: DESI–HST $\times 4$ comparison on a binary-star field. FluxFlow reproduces the four-pointed diffraction pattern on the lower companion together with surrounding faint sources, whereas regression baselines suppress the spike pattern and the generative baselines either oversmooth or oversharpen the background.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction restrict the claims to the evaluated benchmark and HST-referenced metrics; limitations distinguish these from independent sky truth.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The conclusion and Appendix C discuss PSF misspecification, HST artifacts, bandpass differences, transfer, and uncertainty.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [No]

   Justification: Appendix F provides derivations for the Wiener correction and simplified spectral models; a general convergence proof for the full sampler is not provided.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Implementation details included.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Code and usage instructions are available at `https://github.com/I2WM/FluxFlow`. The full reproduction package is not yet publicly available.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Included.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The reconstruction comparisons do not report across-seed confidence intervals or statistical significance tests. Point estimates should not be interpreted as statistically established superiority.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

Justification: Provided.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research uses astronomical survey images and does not involve human subjects or personal data.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Appendix C discusses access to survey analysis and the risk of unsupported scientific interpretation of generated sources, together with the need for independent validation.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: No risk

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: cited and follow the license.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: Provided

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: No crowdsourcing experiment.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: No risk.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLM assistance is used for writing, editing, and formatting this revision, not as a component of the proposed core method.