LOCUS: FEASIBLE-CENTERED NULL-SPACE FLOW FOR INVERSE PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Inverse problems combine incomplete measurements with prior knowledge, from structural assumptions in compressed sensing to learned generative priors. For noiseless linear measurements $y = Ax$, all measurement-consistent reconstructions lie in the affine fiber $A^\dagger y + \ker(A)$. Existing measurement-consistent constructions use this geometry to determine where generative transport may move, and raise a distinct design question: where within this feasible fiber should that transport be centered? We propose to predict an observation-dependent reconstruction, project it onto the feasible fiber, and use flow matching to generate the remaining null-space residual. This separates three roles: preserving what is measured, predicting what is inferable, and generating what remains after that prediction. The resulting dynamics preserve measurements throughout sampling. For independent projected-Gaussian sources and linear training paths, we show that excess center error relative to the conditional mean exactly equals excess displacement energy and ideal marginal-flow action. This gives prediction quality a direct interpretation in terms of residual transport. Across 17 large-cohort evaluations on FFHQ, CelebA-HQ, and ImageNet, spanning block and random-pixel inpainting and $4\times$–$16\times$ super-resolution, the proposed design improves both PSNR and LPIPS over the corresponding pixel-centered null-space flow. A controlled 512-image study separates the center's conditioning and source roles, while a separate 2,000-image random-pixel benchmark demonstrates strong reconstruction and perceptual quality at 20 flow evaluations.

1 INTRODUCTION

Image inpainting and super-resolution recover a clean image from observed pixels or a lower-resolution measurement. These problems are underdetermined: different images can agree with the same observation. Reconstruction therefore requires both fidelity to the measurement and an image prior to constrain the remaining ambiguity. Diffusion and flow models provide learned generative priors for this purpose (Chung et al., 2023; Wang et al., 2023; Kim et al., 2025; Martin et al., 2025).

For a known noiseless linear observation $y = Ax$, A describes how a clean image x is measured. Measurement consistency requires a reconstruction to reproduce y under the same operator. Two consistent reconstructions can therefore differ only by a component that A maps to zero, namely an element of its null space. Restricting reconstruction changes to these directions preserves the observation (Schwab et al., 2019; Wang et al., 2023; Shi et al., 2026).

Unobserved content can nevertheless be partly predictable from the observation. A learned predictor can provide initial image structure, leaving both conditional variation and prediction error. Measurement preservation alone does not specify how to exploit this prediction: the source distribution and conditioning of the generative model must also be chosen.

We introduce *LOCUS*, a conditional flow for residual generation around a learned feasible prediction (Figure 1). We project the prediction onto the measurement-consistent set and use the resulting center as both the source mean and a conditioning input to the velocity network. The flow learns the remaining residual, while projected source noise and velocities preserve the observations throughout sampling.

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

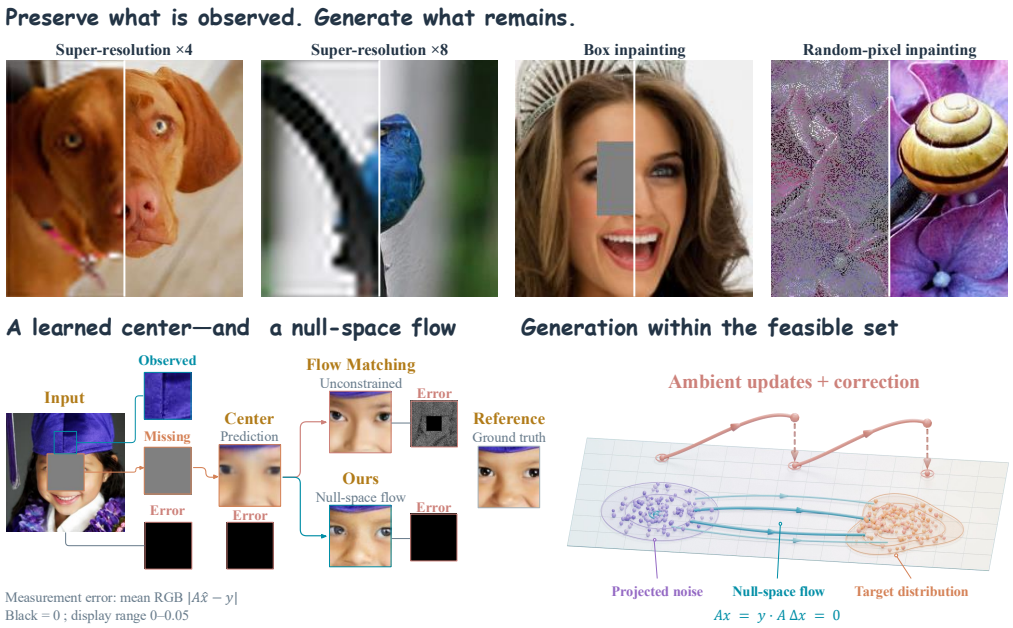

Figure 1: **Preserve, predict, then generate.** A feasible center predicts unobserved structure without changing the measurement; a null-space flow models the remaining residual. Examples span super-resolution and inpainting. The bottom row separates measured information, feasible prediction, and residual generation.

Our contributions are:

- A feasible-centered formulation that combines observation-dependent prediction with measurement-preserving residual flow.
- Exact identities equating reductions in expected squared center error with reductions in displacement energy and ideal population marginal-flow action, for independent projected-Gaussian sources at a fixed noise scale and linear training paths.
- Improvements in both PSNR and LPIPS over the corresponding pixel-centered null-space flow across 17 large-cohort settings on FFHQ, CelebA-HQ, and ImageNet at the same number of flow evaluations, complemented by a 2,000-image external benchmark and controlled studies of the center's predictive and transport roles.

2 Related Work

Generative priors for restoration. Incomplete measurements leave image structure to be inferred from a prior. Learned generators extend the structured-prior viewpoint of compressed sensing (Bora et al., 2017); diffusion and score-based models provide expressive image distributions (Ho et al., 2020; Song et al., 2021). Flow matching learns continuous transport by velocity regression (Lipman et al., 2023; Liu et al., 2023b), with generalized conditional flow matching allowing non-Gaussian source distributions (Tong et al., 2024). This flexibility motivates our use of a prediction-centered source: the flow can learn the residual correction after predictable structure has already been supplied.

Measurement-aware inverse solvers. DDRM and DDNM adapt pretrained diffusion priors to linear observation models (Kawar et al., 2022; Wang et al., 2023), while DPS guides sampling through an approximate likelihood gradient (Chung et al., 2023). Flow-based solvers apply related ideas through conditional transport in OT-ODE (Pokle et al., 2024), likelihood guidance in FlowDPS (Kim et al., 2025), alternating data-fidelity and denoising updates in PnP-Flow (Martin et al., 2025), and destination refinement in FLOWER (Pourya et al., 2026). D-Flow instead optimizes the source by differentiating through generation (Ben-Hamu et al., 2024). These methods offer reuse of pretrained

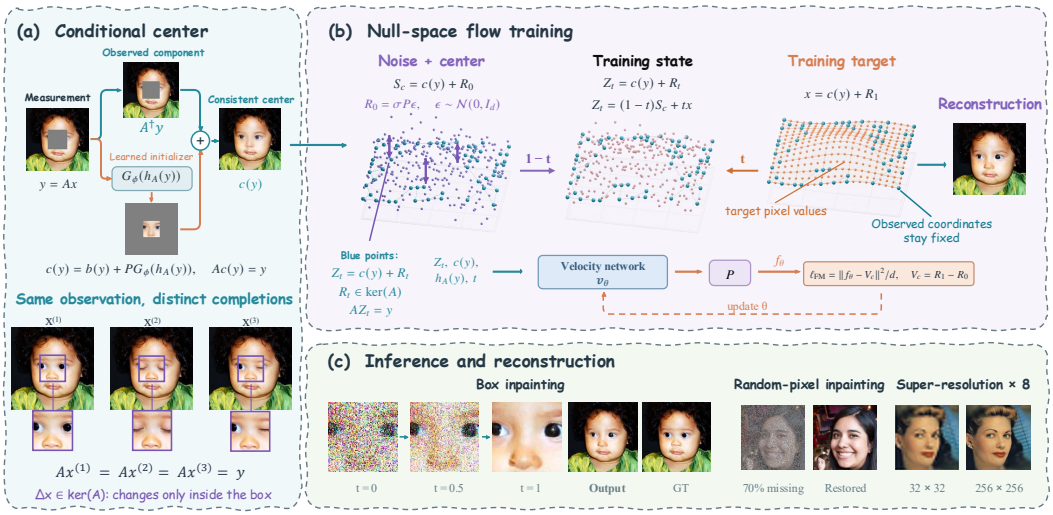

Figure 2: **Feasible prediction turns generation into residual transport.** (a) Projecting the predictor output gives a center $c(y)$ that preserves the observation. (b) Flow matching regresses a projected velocity field against the paired target $V_c = R_1 - R_0$. (c) Inference changes only unmeasured content; the same construction handles block and random-pixel inpainting and super-resolution.

priors across observation operators. LOCUS adopts task-trained conditional transport and builds measurement preservation into both its source and velocity field, so sampling requires no separate likelihood-guidance optimization. Operator-informed conditional velocities are also studied by Shoushtari et al. (2026); our focus is the complementary role of a learned feasible center in defining the residual transport problem.

Predictive sources and null-space generation. I^{2}SB bridges degraded and clean image distributions (Liu et al., 2023a), and ResShift designs restoration trajectories by shifting the image residual (Yue et al., 2023). PMRF transports an estimated posterior mean plus noise to the clean-image distribution (Ohayon et al., 2025). These approaches establish the value of informative sources, but source informativeness alone does not enforce measurement preservation along the subsequent dynamics. Deep null-space learning imposes data consistency on learned corrections (Schwab et al., 2019). NullFlow confines generation to the measurement-consistent set and uses MeanFlow's interval-average velocities for one-step sampling (Shi et al., 2026; Geng et al., 2025). LOCUS couples these two design principles: it projects a learned prediction into a feasible center, supplies the center as a conditioning input, and transports only the remaining null-space residual. Our contribution concerns this coupling and the relation between center error and the energy of the prescribed transport targets; our instantaneous-velocity objective retains ordinary flow matching.

3 Method

Problem setup. We consider noiseless linear observations $y = Ax$, where $x \in \mathbb{R}^d$ is a clean image and $A \in \mathbb{R}^{n \times d}$ is a known observation operator. For inpainting, A selects observed pixels; for super-resolution, it downsamples the image. A reconstruction z is feasible if $Az = y$. With the Moore–Penrose pseudoinverse $A^\dagger$, define

$$b(y) = A^\dagger y, \qquad P = I - A^\dagger A, \qquad \Pi_y(z) = b(y) + Pz. \tag{1}$$

Here $b(y)$ is a feasible reference, P is the orthogonal projector onto $\ker(A) = \{r : Ar = 0\}$, and Π_y projects onto the feasible set $\mathcal{C}_y = \{z : Az = y\} = b(y) + \ker(A)$. LOCUS first learns a feasible prediction and then models the residual left by that prediction (Figure 2).

3.1 Learning a Feasible Center

Let G_ϕ be an image predictor with trainable parameters ϕ. Its input $h_A(y)$ contains the image-sized reference $b(y)$ and task-specific operator information, including the observed-pixel mask for inpainting. We train its projected prediction on paired images and measurements:

$$c_\phi(y) = \Pi_y(G_\phi(h_A(y))), \qquad \mathcal{L}_{\text{center}}(\phi) = \frac{1}{d}\mathbb{E}\left\|c_\phi(y) - x\right\|^2. \tag{2}$$

The projection is fixed by A and guarantees $Ac_\phi(y) = y$. After training, we freeze the predictor's exponential-moving-average (EMA) parameters and denote its output by $c(y)$. This feasible center estimates predictable image structure; the subsequent flow models the residual distribution left by that prediction.

Assuming $\mathbb{E}\left\|x\right\|^2 < \infty$, let $m(y) = \mathbb{E}[x \mid y]$. The decomposition $x - c(y) = [x - m(y)] + [m(y) - c(y)]$ separates conditional variation from center prediction error; both terms lie in $\ker(A)$ because $Am(y) = y$.

3.2 Learning Null-Space Residual Flow

We use conditional flow matching with linear interpolation paths (Lipman et al., 2023; Liu et al., 2023b). For a training pair (x, y), draw $\epsilon \sim \mathcal{N}(0, I_d)$ and $t \sim \mathcal{U}[0, 1]$ independently of the data and each other. With source-noise scale $\sigma > 0$, define

$$\begin{aligned} R_0 &= \sigma P\epsilon, & R_1 &= x - c(y), & R_t &= (1-t)R_0 + tR_1, \\ S_c &= c(y) + R_0, & Z_t &= c(y) + R_t, & V_c &= R_1 - R_0. \end{aligned} \tag{3}$$

Here R_1 is the target image residual, whereas V_c is the velocity target for its interpolation from the noise residual R_0. Both endpoints lie in $\ker(A)$, giving $AZ_t = y$ and $AV_c = 0$. We train

$$f_\theta(z, y, t) = Pv_\theta(z, c(y), h_A(y), t), \qquad \mathcal{L}_{\text{FM}}(\theta) = \frac{1}{d}\mathbb{E}\left\|f_\theta(Z_t, y, t) - V_c\right\|^2. \tag{4}$$

The velocity U-Net v_θ receives the current image state, frozen center, observation condition, and time. It operates on full image tensors; P constrains its output directions. Flow gradients do not update the center predictor.

3.3 Measurement-Preserving Inference

Given y, we compute and cache $c(y)$ once, then integrate

$$z_0 = c(y) + \sigma P\epsilon, \qquad \dot{z}_t = f_\theta(z_t, y, t), \qquad \widehat{x} = z_1. \tag{5}$$

The center and observation condition remain fixed. Since $Az_0 = y$ and $Af_\theta = 0$, every state preserves the observation. Euler and Heun retain this property in exact arithmetic; numerical states are recomposed with Π_y to control roundoff. For inpainting, $P = I - M$ with observed-pixel mask $M = A^\dagger A$, so only missing pixels change. For super-resolution, updates preserve the downsampled measurement. We count flow-network evaluations (NFE) separately from the single center call; Appendix B specifies solvers and post-sampling image-range processing.

3.4 Center Error and Transport Energy

The conditional mean $m(y)$ minimizes expected squared prediction error. We quantify how center accuracy changes transport under our source family. Fix A, an exact projector P, a noise scale σ, and deterministic feasible centers with finite second moments. For any such center s, set $S_s = s(y) + \sigma P\epsilon$, $Z_t^s = (1-t)S_s + tx$, and $V_s = x - S_s$, using independent Gaussian noise. Define

$$\mathcal{E}_s = \mathbb{E}\left\|V_s\right\|^2 = \mathbb{E}\left\|x - s(y)\right\|^2 + \sigma^2 \operatorname{rank}(P), \tag{6}$$

and $u_s^*(z, y, t) = \mathbb{E}[V_s \mid Z_t^s = z, y, t]$. Its ideal marginal-flow action is $\mathcal{K}_s^* = \int_0^1 \mathbb{E}\left\|u_s^*(Z_t^s, y, t)\right\|^2 dt$, evaluated on the training-path marginals.

Proposition 1 (Center error and transport energy). *For every feasible center s under the construction above,*

$$\mathcal{E}_s - \mathcal{E}_m = \mathcal{K}_s^* - \mathcal{K}_m^* = \mathbb{E}\,\|s(y) - m(y)\|^2 . \tag{7}$$

Consequently, replacing b by c gives

$$\mathcal{K}_b^* - \mathcal{K}_c^* = \mathcal{E}_b - \mathcal{E}_c = \mathbb{E}\,\|x - b(y)\|^2 - \mathbb{E}\,\|x - c(y)\|^2 . \tag{8}$$

Proof sketch. Write $\delta(y) = s(y) - m(y)$. Coupling the paths with the same (x, y, ϵ) gives $Z_t^s = Z_t^m + (1-t)\delta(y)$ and $V_s = V_m - \delta(y)$. Conditional on y, the state translation is invertible, so the corresponding ideal velocities differ by $-\delta(y)$. Since $\mathbb{E}[V_m \mid y] = 0$, the cross term in their expected squared norms vanishes. Integration gives the action identity; expanding $\mathcal{E}_s$ gives the same excess. Subtracting the identities for b and c completes the result. □

Thus center MSE reduction exactly removes a predictable component of target and ideal-flow energy, with the conditional mean attaining their minimum in this source family. We evaluate finite-network reconstruction separately. Appendix A gives the full proof, projection properties, and the boundary extension.

4 Experiments

We examine whether a learned feasible center improves restoration across inverse problems, how LOCUS compares with external solvers, and which roles of the center explain the observed gains.

4.1 Evaluation setup

We evaluate block inpainting (80×80 missing regions), fixed random-pixel removal (70% missing), and bicubic super-resolution ($4\times$–$16\times$) on FFHQ, CelebA-HQ, and ImageNet (Karras et al., 2019; 2018; Russakovsky et al., 2015). The primary suite comprises 17 large-cohort settings and a separately identified 100-image FFHQ random-pixel pilot. Images have side 256, except $12\times$ SR and ImageNet $4 \times /8\times$ SR, which use 240.

The main comparison is with *Pixel-NSFM*, which uses the same null-space construction centered at $b(y)$. Within each LOCUS/Pixel-NSFM pair, test images, velocity backbone, operator, noise scale, current flow-stage updates, and Heun NFE40 sampler agree. Center training is additional, and inherited checkpoints are recorded in Appendix B. We also report *Feasible Center* without a flow, *Full-Space FM* from full-space Gaussian noise, and *Center + FM* with unprojected noise and velocities.

RGB PSNR and SSIM assess fidelity; AlexNet LPIPS and DISTS assess reference-based perceptual similarity (Wang et al., 2004; Zhang et al., 2018; Ding et al., 2022). Primary metrics use native float reconstructions. External protocols are specified separately, and image-bootstrap intervals describe test-image variability at fixed checkpoints (Appendix C.4).

4.2 Feasible centering improves cross-task restoration

LOCUS improves both PSNR and LPIPS over Pixel-NSFM in all 17 large-cohort settings (Table 1), with gains of 0.11–0.40 dB and LPIPS reductions of 1.9%–10.8%; the separate pilot agrees. The improvements persist as the observations become less informative: from $4\times$ to $16\times$ SR, the PSNR gain grows from 0.20 to 0.40 dB on FFHQ and from 0.11 to 0.28 dB on CelebA-HQ. ImageNet gains remain 0.30–0.36 dB across these scales.

Prediction and residual generation play complementary roles. The feasible predictor has strong pixel fidelity, while the flow improves reference-based perceptual similarity. On FFHQ $4\times$ SR, LOCUS lowers LPIPS from the center's 0.1160 to 0.0478, with PSNR changing from 32.97 to 30.59 dB. Against the corresponding generative Pixel-NSFM baseline, both metrics improve. Figure 3 illustrates the changes in facial structure and texture.

Table 1: **Feasible centering improves the corresponding null-space flow across tasks and datasets.** Mean PSNR (↑) and LPIPS (↓) use native float reconstructions. All flows use Heun NFE40; center-based flows require one additional center call. Bold identifies LOCUS for comparison with Pixel-NSFM, rather than the best of all five methods. The FFHQ random-pixel row is a 100-image pilot; other cohort sizes are shown in the table. Center training and inherited checkpoints are documented in Appendix B.

Method	Block		Random 70%		SR 4×		SR 8×		SR 12×		SR 16×	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
FFHQ, $n = 2,000$ *(random-pixel pilot:* $n = 100$*)*												
Feasible Center	33.24	0.0359	31.31	0.1183	32.97	0.1160	28.24	0.2485	25.40	0.3345	24.08	0.3861
Full-Space FM	29.49	0.0646	29.40	0.1070	29.50	0.0722	25.90	0.1351	23.32	0.2028	22.27	0.2491
Center + FM	29.47	0.0677	29.95	0.0903	27.15	0.2287	24.71	0.3366	22.96	0.4025	22.00	0.4640
Pixel-NSFM	31.01	0.0184	30.62	0.0740	30.39	0.0497	25.92	0.1333	23.43	0.1990	22.36	0.2447
LOCUS	**31.32**	**0.0174**	**31.31**	**0.0508**	**30.59**	**0.0478**	**26.17**	**0.1286**	**23.79**	**0.1878**	**22.76**	**0.2312**
CelebA-HQ, $n = 3,000$												
Feasible Center	33.94	0.0271	31.21	0.0917	33.33	0.1259	28.75	0.2434	26.01	0.3136	24.72	0.3626
Full-Space FM	31.00	0.0368	33.13	0.0292	29.80	0.0667	26.49	0.1221	24.01	0.1795	22.98	0.2230
Center + FM	31.21	0.0385	32.29	0.0438	27.31	0.2298	24.80	0.3262	23.37	0.3793	22.34	0.4326
Pixel-NSFM	32.15	0.0145	32.72	0.0355	30.59	0.0475	26.53	0.1187	24.10	0.1738	23.08	0.2158
LOCUS	**32.49**	**0.0136**	**33.01**	**0.0316**	**30.70**	**0.0466**	**26.70**	**0.1132**	**24.33**	**0.1628**	**23.36**	**0.2019**
ImageNet, $n = 1,000$												
Feasible Center	27.72	0.0837	29.52	0.0695	27.18	0.2463	23.67	0.4302	21.93	0.5402	20.76	0.6155
Full-Space FM	24.31	0.0737	25.94	0.0797	24.47	0.1397	21.71	0.2947	20.40	0.3969	19.63	0.4666
Center + FM	24.56	0.0742	26.35	0.0703	24.82	0.1356	22.07	0.2876	20.74	0.3848	19.93	0.4578
Pixel-NSFM	24.64	0.0590	26.23	0.0698	24.56	0.1356	21.73	0.2925	20.46	0.4061	19.59	0.4799
LOCUS	**24.87**	**0.0573**	**26.54**	**0.0637**	**24.86**	**0.1328**	**22.09**	**0.2867**	**20.76**	**0.3851**	**19.90**	**0.4594**

The updates retain the observations. Observed pixels remain exact in the recorded mask evaluations. For SR, the largest raw-trajectory measurement residual is 4.77×10^{-7} in model units. Appendix C.2 separates trajectory consistency from final image processing and visualizes residual maps.

4.3 Center conditioning and source centering

We separate the center's two roles in a controlled block-inpainting study with 512 reserved images and 20k matched flow updates. Three arms share initial flow weights, image order, noise, and sampled times: B uses (b, b), F uses (b, c), and C uses (c, c) for (source mean, center input). All use projected dynamics, and F/C share one frozen predictor.

Source centering reduces the endpoint residual energy by 89.5% and the expected velocity-target energy by 16.1%, as characterized by Proposition 1. At one Euler flow call, most of the 0.155 dB C-over-B PSNR gain comes from adding the center input (0.143 dB). The further source translation yields smaller improvements in held-out FM error and VGG LPIPS (Table 5). The combined design also reduces ground-truth-referenced seam RMSE relative to B by 11.3%, 5.2%, and 2.7% at flow NFE 1, 2, and 40. Appendix C.1 gives paired intervals, the complete budget comparison, and center-only results. These measurements distinguish the reduction in predictable displacement from its finite-network effects.

4.4 Comparison with external inverse solvers

Perceptual similarity across six FFHQ tasks. We evaluate perceptual similarity using AlexNet LPIPS and DISTS (Zhang et al., 2018; Ding et al., 2022), alongside PSNR and SSIM for complementary fidelity assessment. Table 2 covers block inpainting, random-pixel recovery, and four SR scales, with 2,000 FFHQ test images per task. Scores use saved RGB PNG reconstructions; the SR protocol

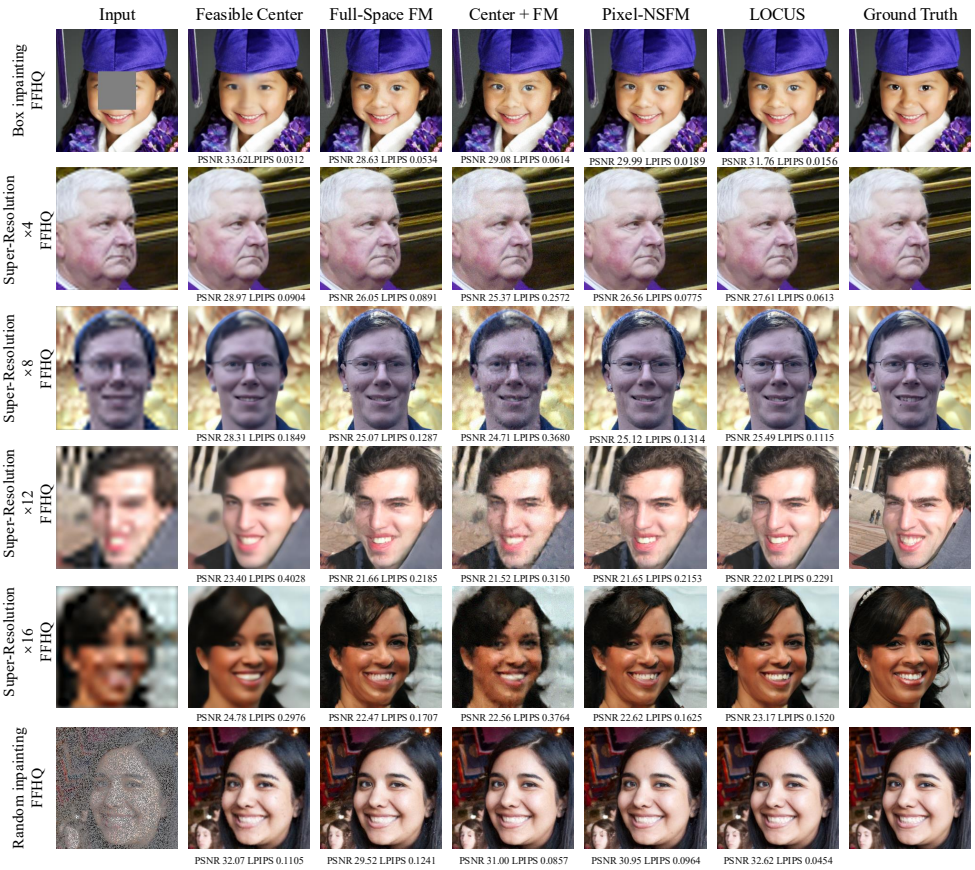

Figure 3: **A predictive center supplies structure, while residual generation changes local detail.** The center-only predictions retain facial layout but smooth fine texture, especially at larger SR factors. The displayed flow variants differ in the facial detail and texture they reconstruct. Columns and per-image scores identify the displayed configurations; Table 1 gives cohort means.

is separate from the separable-bicubic primary evaluation. Training and inference configurations remain method-specific (Appendix D.1).

LOCUS ranks first in both perceptual metrics across all six tasks, leading all 12 task–metric comparisons in Table 2. Relative to the lowest baseline score for each task and metric, LPIPS decreases by 3.3%–35.7% and DISTS by 7.5%–26.6% (computed from the displayed means). For $4\times$ SR, the reductions reach 35.7% and 26.6%, respectively; for random-pixel recovery, they are 23.2% and 26.0%. This consistency across different observation operators supports LOCUS's strength in recovering perceptually similar image structure and texture. The fidelity metrics provide a complementary view: FLOWER leads PSNR and SSIM, whereas LOCUS leads LPIPS and DISTS. Figure 4 highlights the visual benefit: LOCUS reconstructs finer textures and sharper boundaries, whereas FLOWER and PnP-Flow exhibit oversmoothing in the displayed examples. Table 1 separately assesses feasible centering within a common flow family.

Dedicated random-pixel benchmark. Table 3 reports a separate 2,000-image FFHQ random-pixel evaluation with clipped-float scoring and the training and sampling configuration specified below. Its results are reported separately from the cross-task PNG comparison in Table 2. LOCUS uses 100k task-training updates for each of its center and flow stages and 20 Euler flow evaluations. PMRF is trained on the same task using its official restoration and flow stages; architectures and training procedures remain method-specific. All results in this dedicated benchmark use the full prescribed test cohort and clipped-float scoring.

Table 2: **Perceptual similarity and fidelity across six FFHQ tasks.** Means over 2,000 images per task, scored on saved RGB PNGs. LPIPS/DISTS (↓) assess perceptual similarity; PSNR/SSIM (↑) assess fidelity. Best, second-best, and third-best values are bold, underlined, and italic for every method. Budgets differ across methods; *FlowDPS uses our matched-observation-size adaptation, and †DDNM block uses the legacy implementation. Dashes denote unreported results; protocols are in Appendix D.1.

Method	Block				SR 4×				SR 8×			
	LPIPS	DISTS	PSNR	SSIM	LPIPS	DISTS	PSNR	SSIM	LPIPS	DISTS	PSNR	SSIM
DDNM†	0.0266	*0.0192*	30.79	0.9618	0.0918	0.1208	29.39	*0.8571*	0.1577	0.1623	26.12	*0.7426*
FlowDPS*	0.0414	0.0341	28.64	0.9528	0.2654	0.2260	27.61	0.7834	0.3527	0.2580	25.78	0.7228
PnP-Flow	0.0533	0.0812	*31.04*	0.9540	0.1799	0.1684	30.70	0.8871	0.2956	0.2540	27.35	0.7890
FLOWER	*0.0270*	0.0176	**32.76**	**0.9708**	*0.1669*	*0.1551*	**30.82**	**0.8914**	*0.2635*	*0.2268*	**27.99**	**0.8100**
RePaint	0.0278	0.0303	29.41	0.9480	–	–	–	–	–	–	–	–
LOCUS	**0.0184**	**0.0140**	32.09	*0.9589*	**0.0590**	**0.0887**	*29.41*	0.8503	**0.1318**	**0.1390**	*26.17*	0.7295

Method	SR 12×				SR 16×				Random 70%			
	LPIPS	DISTS	PSNR	SSIM	LPIPS	DISTS	PSNR	SSIM	LPIPS	DISTS	PSNR	SSIM
DDNM	0.2094	0.1907	23.45	*0.6477*	0.2428	0.2053	22.46	*0.6189*	0.0419	0.0929	33.74	0.9386
FlowDPS*	0.4477	0.2981	23.20	0.6283	0.4957	0.3162	22.00	0.5844	0.2021	0.2336	28.54	0.8153
PnP-Flow	0.3966	0.3196	23.94	0.6785	0.4265	0.3499	*22.53*	0.6363	0.1063	0.1580	*33.07*	*0.9164*
FLOWER	*0.3446*	*0.2813*	**24.97**	**0.7191**	*0.3705*	*0.3041*	**23.90**	**0.6862**	*0.0464*	*0.1004*	**35.66**	**0.9576**
RePaint	–	–	–	–	–	–	–	–	–	–	–	–
LOCUS	**0.1934**	**0.1734**	*23.62*	0.6336	**0.2348**	**0.1899**	22.60	0.5957	**0.0322**	**0.0687**	31.53	0.8825

LOCUS achieves 35.22 dB PSNR, 0.9582 SSIM, 0.0176 LPIPS, and 0.0420 DISTS with one center call and 20 flow calls, while preserving the known pixels. Compared with FLOWER, PSNR increases by 0.68 dB, and LPIPS and DISTS decrease by 44.0% and 39.7%, respectively. The qualitative comparison in Figure 5 shows corresponding differences in local texture. Replacing PMRF's observed pixels at the output improves its results, with its best recorded LPIPS reaching 0.0359 (Appendix D).

Additional 100-image block and SR comparisons are reported in Appendix D.

Table 3: **LOCUS improves both perceptual and pixel metrics over the listed external configurations on the full random-pixel benchmark.** Means over 2,000 FFHQ images with 30% observed pixels; all methods receive identical observations and use clipped-float RGB scoring. Calls include the center where applicable, with costs varying by architecture; FlowDPS additionally uses 84 VAE-gradient calls. PMRF parentheses specify flow NFE. Known RMSE is measured in $[0, 1]$; DDNM uses the recorded simplified implementation.

Method	Calls	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓	Known RMSE
FlowDPS	56	27.76	0.7841	0.2485	0.2242	0.03944
PnP-Flow	100	33.11	0.9183	0.0894	0.1395	0.01404
FLOWER	100	34.54	0.9463	0.0314	0.0696	0
DDNM (simplified)	100	33.74	0.9379	0.0426	0.0948	≈ 0
PMRF (1)	2	32.38	0.9035	0.1078	0.1427	0.02140
PMRF (5)	6	31.86	0.8979	0.0704	0.1103	0.02206
PMRF (25)	26	31.20	0.8839	0.0510	0.0920	0.02368
LOCUS	21	**35.22**	**0.9582**	**0.0176**	**0.0420**	0



Figure 4: **Qualitative comparison across FFHQ restoration tasks.** Examples show block inpainting, $4\times$–$16\times$ super-resolution, and random-pixel recovery. The panels illustrate differences in reconstructed structure and texture among the displayed methods. Table 2 reports mean quantitative results over 2,000 test images per task. N/A denotes an unavailable task–method result.



Figure 5: **Recovering local texture from sparse observed pixels.** For FFHQ image 69686, LOCUS retains skin and fabric texture that is attenuated in the smoother reconstructions. Parentheses give flow NFE, excluding the center call. The full-cohort comparison is in Table 3; enlarged views are in Appendix E.1.

5 Conclusion

LOCUS uses a learned feasible prediction to initialize and condition null-space residual flow, preserving measurements throughout sampling under known noiseless linear observation models. For independent projected-Gaussian sources at a fixed noise scale and linear training paths, reductions in expected squared center error exactly equal those in displacement energy and ideal population marginal-flow action. Across 17 large-cohort settings, LOCUS improves PSNR and LPIPS over pixel-centered null-space flow; a separate 2,000-image random-pixel benchmark demonstrates strong fidelity and perceptual quality against external solvers. Controlled experiments link predictive conditioning to reconstruction gains and source centering to reduced transport-target energy.

REFERENCES

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through flows for controlled generation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 3462–3483. PMLR, 2024. URL https://proceedings.mlr.press/v235/ben-hamu24a.html.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 537–546. PMLR, 2017. URL https://proceedings.mlr.press/v70/bora17a.html.

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 (CVPR)*, pp. 22367–22377, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Chen_Activating_More_Pixels_in_Image_Super-Resolution_Transformer_CVPR_2023_paper.html.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022. doi: 10.1109/TPAMI.2020.3045810. URL https://doi.org/10.1109/TPAMI.2020.3045810.

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 (CVPR)*, pp. 23174–23184, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Dong_TSD-SR_One-Step_Diffusion_with_Target_Score_Distillation_for_Real-World_Image_CVPR_2025_paper.html.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, volume 38, pp. 75460–75482, 2025. doi: 10.52202/085713-2534. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/6d13e085b79d454da5910e4ca82a3d9d-Abstract-Conference.html.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=Hk99zCeAb.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.nips.cc/paper_files/paper/2022/hash/95504595b6169131b6ed6cd72eb05616-Abstract-Conference.html.

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 (ICCV)*, pp. 12328–12337, October 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Kim_FlowDPS__Flow-Driven_Posterior_Sampling_for_Inverse_Problems_ICCV_2025_paper.html.

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 (ICCV) Workshops*, pp. 1833–1844, 2021. URL https://openaccess.thecvf.com/content/ICCV2021W/AIM/html/Liang_SwinIR_Image_Restoration_Using_Swin_Transformer_ICCVW_2021_paper.html.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos Theodorou, Weili Nie, and Anima Anandkumar. I^2SB: Image-to-image schrödinger bridge. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22042–22062. PMLR, 2023a. URL https://proceedings.mlr.press/v202/liu23ai.html.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=XVjTT1nw5z.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11461–11471, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Lugmayr_RePaint_Inpainting_Using_Denoising_Diffusion_Probabilistic_Models_CVPR_2022_paper.html.

Ségolène Martin, Anne Gagneux, Paul Hagemann, and Gabriele Steidl. PnP-Flow: Plug-and-play image restoration with flow matching. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=5AtHrq3B5R.

Guy Ohayon, Tomer Michaeli, and Michael Elad. Posterior-mean rectified flow: Towards minimum MSE photo-realistic image restoration. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=hPOt3yUXii.

Ashwini Pokle, Matthew J. Muckley, Ricky T. Q. Chen, and Brian Karrer. Training-free linear image inverses via flows. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=PLIt3a4yTm.

Mehrsa Pourya, Bassam El Rawas, and Michael Unser. Flower: A flow-matching solver for inverse problems. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=QGd34p02mI.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y. URL https://doi.org/10.1007/s11263-015-0816-y.

Johannes Schwab, Stephan Antholzer, and Markus Haltmeier. Deep null space learning for inverse problems: Convergence analysis and rates. *Inverse Problems*, 35(2):025008, 2019. doi: 10.1088/1361-6420/aaf14a. URL https://arxiv.org/abs/1806.06137.

Xiao Shi, Edward P. Chandler, Chicago Y. Park, Shirin Shoushtari, and Ulugbek S. Kamilov. NullFlow: One-step generative reconstruction. *arXiv preprint arXiv:2606.22696*, 2026. URL https://arxiv.org/abs/2606.22696.

Shirin Shoushtari, Edward P. Chandler, Xiao Shi, and Ulugbek S. Kamilov. Fast and faithful: Principled conditional flow matching for inverse problems. *arXiv preprint arXiv:2609.12953*, 2026. URL https://arxiv.org/abs/2609.12953.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2011.13456.

Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2149–2159, 2022. URL https://openaccess.thecvf.com/content/WACV2022/html/Suvorov_Resolution-Robust_Large_Mask_Inpainting_With_Fourier_Convolutions_WACV_2022_paper.html.

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. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=mRieQgMtNTQ.

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 (CVPR)*, pp. 25796–25805, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wang_SinSR_Diffusion-Based_Image_Super-Resolution_in_a_Single_Step_CVPR_2024_paper.html.

Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/tip.2003.819861. URL https://doi.org/10.1109/tip.2003.819861.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/a8223b0ad64007423ffb308b0dd92298-Abstract-Conference.html.

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. ResShift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/2ac2eac5098dba08208807b65c5851cc-Abstract-Conference.html.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.

AI USE STATEMENT

An AI assistant assisted with organizing experiment records, developing and checking mathematical derivations and proofs, checking source references, interpreting recorded results, revising prose and LaTeX tables, and preparing and correcting labels in supplied artwork. It did not generate the reported experimental measurements. The authors are responsible for verifying the claims, references, and artifacts in the submitted version.

Appendix roadmap. Appendix A gives the theoretical analysis; Appendix B specifies implementation and evaluation protocols; Appendix C reports mechanism diagnostics and additional primary-suite results; Appendix D documents external comparisons; Appendix E provides qualitative evidence; and Appendix F discusses limitations and diagnostic evaluations.

A Theoretical analysis

A.1 Assumptions and three distinct energy quantities

Fix A and $Y = AX$, with $\mathbb{E}\|X\|^2 < \infty$ and deterministic, square-integrable feasible source means. Let $P = I - A^\dagger A$ be the orthogonal null-space projector and $Q = I - P$. Gaussian noise and time are independent of the data and each other. The kinetic-action convention here is $\int \|v\|^2 \, dt$, without a factor of $1/2$, on the unit interval. Pixel normalizations must be shared within every comparison.

Three different objects enter the analysis. First, $\mathcal{E}_s = \mathbb{E}\|X - S_s\|^2$ is the energy of the paired straight-path velocity target. Second, $\mathcal{K}_s^*$ is the Eulerian kinetic action of the ideal marginal FM field evaluated on the corresponding path marginals. Third, $\int \|f_\theta(z_t, y, t)\|^2 \, dt$ is the action of an actual learned rollout. The first two admit the exact centering identity below. The third requires evaluation of the trained field and its states; it is not obtained by relabeling target energy.

A.2 Projection and measurement preservation

Proposition 2 (Feasibility and removal of observable action)**.** *A feasible source and projected velocity preserve $Az_t = y$, including every explicit Runge–Kutta stage. For any square-integrable velocity v,*

$$\|Pv\|^2 = \|v\|^2 - \|(I - P)v\|^2 . \tag{9}$$

More generally, let $w : [0, 1] \to \mathbb{R}^d$ be absolutely continuous, with $\dot{w} \in L^2$ and feasible endpoints, and set $z_t = \Pi_y(w_t)$. With $\mathcal{K}(w) = \int_0^1 \|\dot{w}_t\|^2 \, dt$,

$$\mathcal{K}(z) = \mathcal{K}(w) - \int_0^1 \|(I - P)\dot{w}_t\|^2 \, dt \;\geq\; \|w_1 - w_0\|^2 . \tag{10}$$

Proof. The identities $P = P^\top = P^2$ and $PQ = 0$ give the orthogonal decomposition $v = Pv + Qv$. Pythagoras proves Equation (9). For a projected ODE, $A\dot{z}_t = APv_t = 0$ almost everywhere, so a feasible initial state remains feasible. An explicit Runge–Kutta stage or accepted state adds a linear combination of vectors Pv_j, and applying A to this combination gives zero. This proves the discrete claim at any step size.

For the path statement, $z_t = b(y) + Pw_t$ is absolutely continuous and $\dot{z}_t = P\dot{w}_t$ almost everywhere. Since w_0, w_1 are feasible, $z_0 = w_0$ and $z_1 = w_1$. Integrating Pythagoras gives the action identity. Finally, Cauchy–Schwarz on $[0, 1]$ gives

$$\|w_1 - w_0\|^2 = \left\| \int_0^1 \dot{z}_t dt \right\|^2 \leq \int_0^1 \|\dot{z}_t\|^2 \, dt.$$

If an action is infinite, the finite-energy comparison is understood only when its terms are defined; the stated energy decomposition applies to square-integrable derivatives. □

The lower bound is attained by a constant-velocity straight path between the two endpoints. More precisely, for $d_{01} = w_1 - w_0$,

$$\mathcal{K}(z) - \|d_{01}\|^2 = \int_0^1 \|\dot{z}_t - d_{01}\|^2 \, dt \geq 0.$$

Projection removes range-space excursions but need not remove nonconstant null-space motion. Also, the pointwise norm identity compares v and Pv at the same state. Integrating a separately trained ambient field usually produces different states, so this identity does not order the two learned rollout actions.

A.3 Feasible centering: proof of Proposition 1

Proof. Expand $V_s = X - s(y) - \sigma P\epsilon$. Independence and zero noise mean eliminate the cross term, and $\mathbb{E}\,\|P\epsilon\|^2 = \mathrm{tr}(P) = \mathrm{rank}(P)$. This proves Equation (6).

For fixed t, let $\mathcal{F}_s(t)$ be generated by (Z_t^s, Y) and set $u_s^*(Z_t^s, Y, t) = \mathbb{E}[V_s \mid \mathcal{F}_s(t)]$. Conditional expectation is an orthogonal projection in L^2, hence

$$\mathbb{E}\,\|V_s\|^2 = \mathbb{E}\,\|u_s^*\|^2 + \mathbb{E}\,\|V_s - u_s^*\|^2 . \tag{11}$$

For the two centers coupled with the same (X, Y, ϵ), put $\delta(Y) = c(y) - b(y)$. Then

$$Z_t^c = Z_t^b + (1-t)\delta(Y), \qquad V_c = V_b - \delta(Y).$$

These states are related by an invertible translation conditional on Y, so $\mathcal{F}_c(t) = \mathcal{F}_b(t)$. Since $\delta(Y)$ is measurable with respect to this common sigma-field,

$$V_c - \mathbb{E}[V_c \mid \mathcal{F}_c(t)] = V_b - \mathbb{E}[V_b \mid \mathcal{F}_b(t)].$$

Their irreducible squared residuals in Equation (11) are equal. Subtracting at each t and integrating over the unit interval proves

$$\mathcal{K}_b^* - \mathcal{K}_c^* = \mathcal{E}_b - \mathcal{E}_c.$$

Subtracting the common noise contribution gives Equation (8). To obtain the excess-energy form in Equation (7), apply the same identity to s and m. Conditional-mean orthogonality gives

$$\mathcal{E}_s - \mathcal{E}_m = \mathcal{K}_s^* - \mathcal{K}_m^* = \mathbb{E}\,\|X - s(Y)\|^2 - \mathbb{E}\,\|X - m(Y)\|^2 = \mathbb{E}\,\|s(Y) - m(Y)\|^2 .$$

The conditional mean therefore attains the minimum of both quantities within the specified source family; equality holds precisely when $s(Y) = m(Y)$ almost surely.

□

For completeness, $m(y) = \mathbb{E}[x \mid y]$ is feasible because $Am(y) = \mathbb{E}[AX \mid Y = y] = y$. The usual orthogonal conditional-mean decomposition gives

$$\mathbb{E}\,\|X - s(y)\|^2 = \mathbb{E}\,\mathrm{tr}\,\mathrm{Cov}(X \mid Y) + \mathbb{E}\,\|m(y) - s(y)\|^2 .$$

Thus m minimizes paired target energy and, by the action-difference identity, ideal marginal action within this independent projected-Gaussian source family at fixed σ. This is not minimization over arbitrary source distributions, couplings, or stochastic processes.

No existence assumption for a classical ODE is needed merely to define the Eulerian quantity $\mathcal{K}_s^*$ on the specified path marginals. Under the usual continuity-equation and ODE regularity conditions, the Bayes FM field realizes those marginals, and $\mathcal{K}_s^*$ is also its expected trajectory action. This extra interpretation does not make a finite trained field equal to that ideal one. In particular, a low-energy learned field could simply fail to move enough; action must be evaluated together with reconstruction quality.

A.4 Extension to fixed linear features

Let T be any fixed finite-dimensional linear map. The same argument yields

$$\mathbb{E}\,\|TV_s\|^2 = \mathbb{E}\,\|T(X - s(y))\|^2 + \sigma^2\,\|TP\|_{\mathrm{F}}^2 , \tag{12}$$

$$\int_0^1 \left(\mathbb{E}\,\|Tu_b^*\|^2 - \mathbb{E}\,\|Tu_c^*\|^2\right) dt = \mathbb{E}\,\|T(X - b(y))\|^2 - \mathbb{E}\,\|T(X - c(y))\|^2 . \tag{13}$$

Indeed, $Tu_s^* = \mathbb{E}[TV_s \mid \mathcal{F}_s(t)]$, and the equality of conditional-expectation residuals above remains true after applying T. The result also holds for bounded known $T(Y)$ measurable with respect to the observation, with the corresponding expected noise trace. We use a fixed mask and fixed T in the boundary statements below.

A.5 BOUNDARY FIDELITY AND TRANSPORT

Use a rectangular pixel grid with horizontal and vertical nearest-neighbor edges and unit pixel spacing. Let Ω be the missing set, M the observed-pixel mask, and $P = I - M$, shared across image channels. Let Γ contain each grid edge with exactly one endpoint in Ω, once only, oriented as (i, j) with $i \in \Omega$ and $j \notin \Omega$. The operator D_Γ stacks $x_i - x_j$ for all channels and edges. Define $T_\Gamma = D_\Gamma P$ and $N_\Gamma = n_{\mathrm{ch}}|\Gamma| > 0$.

Proposition 3 (Boundary accuracy and boundary-directed transport)**.** *For masking, orient each boundary edge $\gamma = (i, j)$ from a missing pixel i to an observed neighbor j. Let D_Γ compute these channel-wise differences, $T_\Gamma = D_\Gamma P$, and $N_\Gamma = n_{\mathrm{ch}}|\Gamma| > 0$. Define seam MSE by the gradient error below. For any feasible reconstruction $\widehat{x}$, it satisfies*

$$\mathcal{S}_\Gamma(\widehat{x}, x) := \frac{\|D_\Gamma(\widehat{x} - x)\|^2}{N_\Gamma} = \frac{1}{N_\Gamma} \sum_{(i,j)\in\Gamma} \|\widehat{x}_i - x_i\|^2 . \tag{14}$$

Define boundary target energy $\mathcal{E}_{s,\Gamma} = \mathbb{E}\,\|T_\Gamma V_s\|^2 / N_\Gamma$ and ideal boundary action $\mathcal{K}^*_{s,\Gamma} = \int_0^1 \mathbb{E}\,\|T_\Gamma u^*_s\|^2\, dt/N_\Gamma$. Then

$$\begin{aligned} \mathcal{E}_{s,\Gamma} &= \mathbb{E}\mathcal{S}_\Gamma(s(y), x) + \sigma^2, \\ \mathcal{K}^*_{b,\Gamma} - \mathcal{K}^*_{c,\Gamma} &= \mathcal{E}_{b,\Gamma} - \mathcal{E}_{c,\Gamma} = \mathbb{E}\mathcal{S}_\Gamma(b(y), X) - \mathbb{E}\mathcal{S}_\Gamma(c(y), X). \end{aligned} \tag{15}$$

Proof. Let $e = \widehat{x} - x$. Feasibility under noiseless masking gives $Me = 0$, so $e_j = 0$ for an observed endpoint. Thus each cross-boundary gradient error is

$$(\widehat{x}_i - \widehat{x}_j) - (x_i - x_j) = e_i - e_j = e_i.$$

Squaring and averaging proves Equation (14). If a missing pixel meets more than one observed neighbor, its error is counted once per boundary edge, as required by the definition.

Each row of T_Γ selects one missing pixel in one channel. It has norm one, even when another row selects the same coordinate. Hence $\|T_\Gamma P\|_{\mathrm{F}}^2 = \|T_\Gamma\|_{\mathrm{F}}^2 = N_\Gamma$. Applying Equation (12) and dividing by N_Γ gives the target-energy identity. For a feasible center $s(y)$, $X - s(y)$ vanishes on observed pixels, so its feature error is precisely the seam MSE. Equation (13) then gives the ideal boundary-action difference in Equation (15). □

This makes the sufficient condition for boundary energy improvement explicit: the learned center must reduce boundary-weighted MSE, not merely full-image MSE. A global error decrease can be concentrated far from the boundary and leave its behavior unchanged. The center's boundary error is directly evaluable on held-out target images.

A.6 BOUNDARY TRACE BOUNDS FOR ENDPOINT AND VELOCITY ERRORS

For each missing pixel, let $d_\Gamma(i)$ be the number of observed nearest neighbors. Since $T_\Gamma^\top T_\Gamma$ is diagonal with these counts on the corresponding channel coordinates,

$$\|T_\Gamma\|_{\mathrm{op}}^2 = \max_{i\in\Omega} d_\Gamma(i) =: r_\Gamma \le 4.$$

For an axis-aligned square hole with side length at least two pixels that does not touch the outer image boundary, $r_\Gamma = 2$ because each corner has two observed neighbors. In general,

$$\mathcal{S}_\Gamma(\widehat{x}, X) \le \frac{r_\Gamma}{N_\Gamma} \|\widehat{x} - x\|^2 \tag{16}$$

for any feasible reconstruction.

Let $z(t)$ and $r(t)$ be absolutely continuous feasible curves, with $r(1) = X$. Write $e_0 = z(0) - r(0)$ and $\delta v(t) = \dot{z}(t) - \dot{r}(t)$. By integration and linearity,

$$D_\Gamma(z(1) - X) = T_\Gamma e_0 + \int_0^1 T_\Gamma \delta v(t)\, dt. \tag{17}$$

In particular, for a shared source $e_0 = 0$, Cauchy–Schwarz implies

$$\mathcal{S}_\Gamma(z(1), X) \le \frac{1}{N_\Gamma} \int_0^1 \|T_\Gamma \delta v(t)\|^2 \, dt \le \frac{r_\Gamma}{N_\Gamma} \int_0^1 \|\delta v(t)\|^2 \, dt. \tag{18}$$

For a different source, Minkowski's inequality gives the transparent form

$$\sqrt{\mathcal{S}_\Gamma(z(1), X)} \le \frac{\|T_\Gamma e_0\|}{\sqrt{N_\Gamma}} + \left(\frac{1}{N_\Gamma} \int_0^1 \|T_\Gamma \delta v(t)\|^2 \, dt \right)^{1/2}.$$

These are pathwise control statements: accurate boundary initialization and small integrated boundary velocity discrepancy are sufficient to bound final seam error. Comparing upper bounds alone does not prove a ranking of the actual errors of two independently learned models.

One evaluable reference is the paired straight bridge $r(t) = (1-t)z(0) + tX$. Then $\dot{r} = X - z(0)$ and $\delta v(t) = f_\theta(z(t), y, t) - \{X - z(0)\}$. This is a discrepancy against that paired secant along a learned rollout. It is not the unknown Bayes velocity error at the rollout state and must not be labeled as one. Ordinary bridge-validation risk is evaluated on a different state distribution and cannot be substituted directly for this rollout integral.

For a discrete projected solver with updates $z_{n+1} = z_n + h_n \bar{v}_n$, $h_n > 0$, and $\sum_n h_n = 1$, the same relation holds with an exact finite sum:

$$T_\Gamma(z_K - X) = \sum_n h_n T_\Gamma \{\bar{v}_n - (X - z_0)\},$$

and Jensen's inequality bounds its squared norm by the corresponding weighted sum of squared discrepancies. For Euler, $\bar{v}_n$ is its projected field value; for Heun it is the average of its two projected field values. Floating-point reprojection corrections can be included in the measured increment rather than assumed absent.

A.7 What is and is not implied about visible seams

A seam metric must compare predicted gradients with the target gradients. Reducing raw gradient magnitude can suppress a genuine image edge and does not by itself indicate a natural seam. The stated metric is an objective local reconstruction error, not a proof of human-perceived realism.

It is also possible for final projection of an ambient prediction to increase this gradient error while decreasing image MSE. For an arbitrary ambient error e, the projected error is Pe, and

$$N_\Gamma \{\mathcal{S}_\Gamma(w, X) - \mathcal{S}_\Gamma(\Pi_y(w), X)\} = \sum_{(i,j)\in\Gamma} \big(\|e_j\|^2 - 2\langle e_i, e_j \rangle \big). \tag{19}$$

Thus projection improves this metric whenever the right side is nonnegative. In expectation, vanishing observed/missing error correlation is one sufficient condition. Conversely, if both endpoints have the same nonzero error, the original gradient error is zero and projection reveals the missing-side error. This counterexample is why projection's energy guarantee alone cannot certify better-looking boundaries than an ambient solver. In our construction the observed side is fixed throughout, while the center and flow are evaluated on how accurately they recover the remaining side.

For luminance metrics, a fixed channel map $L(R, G, B) = 0.299R + 0.587G + 0.114B$ can precede D_Γ. The general feature identities still apply. Its noise term is $\sigma^2 \|\ell\|^2$ per spatial boundary edge for independent RGB noise, rather than σ^2; RGB and luminance values must not be mixed. Mapping a model image from $[-1, 1]$ to $[0, 1]$ halves difference magnitudes and quarters squared energies. The reported seam RMSE uses the latter display units; all theoretical identities remain valid after consistent rescaling.

A.8 Ideal endpoint behavior and finite learned fields

Ordinary FM at one Euler call. Our velocity is instantaneous. At the independent source endpoint, the conditional-expectation version of its Bayes field satisfies

$$u_s^*(z, y, 0) = \mathbb{E}[X - S_s \mid S_s = z, Y = y] = m(y) - z.$$

Thus an ideal unit Euler update returns $S_s + u_s^*(S_s, Y, 0) = m(y)$. This endpoint identity is not implied by an almost-everywhere training loss in continuous time without an appropriate endpoint version or regularity. It also does not apply to MeanFlow's learned interval-average velocity or a distilled endpoint map. We therefore report one-Euler reconstruction as its own operating point rather than equating it with exact posterior sampling.

Finite learning and energy. Conditioning on the full observation makes a deterministic feasible center an invertible translation of each training path. Consequently the ideal conditional regression ambiguity is unchanged, as used in the action proof. For a finite learned field, the squared velocity risk decomposes into this common Bayes floor and approximation error. Center features and source parameterization can change approximation and optimization, which the controlled experiment measures. Reduced target energy does not by itself bound discretization error or guarantee a ranking of separately trained models; the supplementary early and alternate-operator experiments illustrate the need for these measured checks.

B Experimental setup and reproducibility

The primary suite, the controlled B/F/C study, and the external benchmarks use distinct evaluation protocols. This section documents the primary implementation and checkpoints; the study-specific details appear with their results in Appendices C.1 and D.

Reproducibility and code availability. We provide an anonymized ZIP archive of the LOCUS implementation as supplementary material accompanying this submission. This section documents the observation operators, model architecture, two-stage training procedure, sampling and output processing, evaluation cohorts, and checkpoint provenance. The theoretical assumptions and proofs are provided in Appendix A. Protocols for the controlled mechanism study and external comparisons are detailed in Appendices C.1 and D, respectively. The code will be made available in a public repository after the anonymous review process.

B.1 Operators and coordinate units

The central mask removes the 80×80 square from a 256×256 image. Its top and left coordinates are 88 under zero-based indexing. Masks are shared across RGB channels. The random-pixel operator removes 70% of spatial coordinates using the recorded fixed mask, shared across images and splits; it is not an independently resampled test mask. The ImageNet version removes 45,875 of 65,536 spatial coordinates with mask seed 20260831. Face- and server-specific mask tensors belong to their own protocol, and results from these protocols are not pooled.

The main SR suite uses `SeparableBicubicSROperator`. It constructs a separable downsampling matrix and its corresponding pseudoinverse, so $P = I - A^\dagger A$ is defined by that exact matrix. The alternate SVD protocol uses `OfficialSVDResizeOperator`, including its 0.03 singular-value threshold in both the forward operator and pseudoinverse. A nominal scale factor alone does not identify the same task. The $12\times$ evaluations use images of side 240 for an integer output size; ImageNet $4\times$ and $8\times$ in the primary suite also use 240.

Model coordinates are in $[-1, 1]$. Image-quality metrics use the evaluator's documented RGB image representation. A change from $[-1, 1]$ to $[0, 1]$ divides squared residuals by four, and the energy tables retain the model-coordinate normalization per null-space dimension. The seam table instead reports luminance-gradient RMSE in $[0, 1]$ units. Its channel map is $0.299R + 0.587G + 0.114B$.

B.2 Architecture and two-stage training

The primary suite uses the legacy conditional U-Net backbone with base width 64 and channel multipliers $(1, 2, 4, 4)$, attention, and a 16-image training batch. The center, current image, and operator condition are concatenated as image features; the time embedding enters the velocity network. The implementation represents full images, rather than compressed null-space coordinates. For LOCUS and Pixel-NSFM, projection restricts the evolution but does not itself reduce U-Net tensor dimensions or establish a FLOP reduction.

For LOCUS, the center and flow are trained on paired clean images and their synthetically computed measurements. At each center update, we compute $y = Ax$, predict $c_\phi(y) = b(y) + PG_\phi(h_A(y))$, and minimize the mean squared reconstruction error. We retain the center EMA and freeze it during flow training. Each flow update draws a clean image, independent Gaussian noise, and uniform time; constructs (Z_t, V_c) from Equation (3); and minimizes the projected velocity MSE in Equation (4). Flow EMA weights are used for evaluation. The primary suite and the separate FFHQ random-pixel LOCUS model use these squared objectives without endpoint, perceptual, frequency, or wavelet auxiliary losses. Pixel-NSFM uses the projected construction with $c(y) = b(y)$. Full-Space FM and Center + FM use unprojected source noise and velocities.

B.3 Discrete sampling and output processing

For LOCUS and Pixel-NSFM, Heun sampling with $h = 1/K$ computes

$$\begin{aligned} k_1 &= f_\theta(z_n, y, t_n), & \widetilde{z} &= \Pi_y(z_n + hk_1), \\ k_2 &= f_\theta(\widetilde{z}, y, t_n + h), & z_{n+1} &= \Pi_y\left(z_n + \tfrac{h}{2}(k_1 + k_2)\right). \end{aligned} \tag{20}$$

Euler instead uses $z_{n+1} = \Pi_y(z_n + hf_\theta(z_n, y, t_n))$. The center is computed once and cached. Heun20 therefore requires 40 flow calls and LOCUS requires one additional center call. The external FFHQ random-pixel result uses Euler20 with its distinct task-trained checkpoint.

For constrained outputs, the evaluator clips to the pixel box and reprojects onto the affine constraint. The SR evaluations use the recorded 100-cycle box/affine correction, ending with the affine operation. Mask evaluations require one such correction. The final affine step preserves the linear measurement up to roundoff; finite cycling can retain a small pixel-box violation. Saved PNG quantization and clipping can introduce additional SR measurement error. For this reason, raw float trajectory consistency and processed-image quality are reported separately.

B.4 Primary cohorts and checkpoint provenance

The primary table retains the pre-existing complete five-method cohorts at their recorded Heun40 working point, rather than selecting the best test LPIPS or the sign of a comparison. Its 17 large-cohort settings and one marked pilot account for all 18 tasks in that collection. The task index, original summary files, metric CSVs, and resolved training configurations are retained with their hashes in the evidence package.

FFHQ uses 2,000 test images except for the 100-image random-pixel pilot. CelebA-HQ uses 3,000 images. The six ImageNet primary tasks share a class-stratified 1,000-image test manifest, with one validation image per class; class coverage was checked against the dataset labels. Within every LOCUS/Pixel-NSFM pair, test identifiers and their order match exactly. The primary suite uses one recorded training seed per configuration. The split is an evaluation cohort, not 1,000 independent training runs.

The FFHQ $4\times$ stage is trained for 100k updates. Several subsequent face SR stages inherit their corresponding method-specific $4\times$ checkpoints, and some resume an interrupted run. FFHQ block80 and its random-pixel pilot each have a 5k adaptation stage after their method-specific 100k parents. ImageNet primary stages inherit their own recorded pilot or earlier SR checkpoints. These are matched *current flow-stage settings*, with separate method histories; they are not common-initialization or equal-total-training-compute experiments. The B/F/C experiment in Appendix C.1 supplies the stronger common-initialization comparison.

The Feasible Center column is the recorded standalone projected predictor. In two entries it differs from the exact center weights used by LOCUS: CelebA-HQ block uses a 30k Feasible Center baseline while the frozen best center comes from 26k; CelebA-HQ $12\times$ uses a 30k baseline while the frozen best center comes from 29k. EMA tensors differ in both cases. The CelebA-HQ random-pixel 30k and best aliases were checked to have identical EMA tensors. Thus the first column should not be read as an exact center-only ablation in the two exceptional rows.

Table 4: Recorded primary LOCUS/Pixel-NSFM flow-stage settings. Side is the image width/height. Steps refer to the current flow stage and exclude inherited training and the Feasible Center. The LOCUS and Pixel-NSFM configurations agree on these entries. *The 100-image FFHQ random-pixel row is a pilot.

Dataset	Task	Side	n	Steps	Batch	σ	Learning rate
FFHQ	Block	256	2000	5k	16	1	5×10^{-5}
FFHQ	Random 70%	256	100	5k	16	1	5×10^{-5}
FFHQ	SR 4×	256	2000	100k	16	0.075	2×10^{-4}
FFHQ	SR 8×	256	2000	60k	16	0.075	5×10^{-5}
FFHQ	SR 12×	240	2000	60k	16	0.075	5×10^{-5}
FFHQ	SR 16×	256	2000	60k	16	0.075	5×10^{-5}
CelebA-HQ	Block	256	3000	60k	16	1	5×10^{-5}
CelebA-HQ	Random 70%	256	3000	60k	16	1	5×10^{-5}
CelebA-HQ	SR 4×	256	3000	60k	16	0.075	5×10^{-5}
CelebA-HQ	SR 8×	256	3000	60k	16	0.075	5×10^{-5}
CelebA-HQ	SR 12×	240	3000	60k	16	0.075	5×10^{-5}
CelebA-HQ	SR 16×	256	3000	60k	16	0.075	5×10^{-5}
ImageNet	Block	256	1000	100k	16	1	5×10^{-5}
ImageNet	Random 70%	256	1000	100k	16	1	5×10^{-5}
ImageNet	SR 4×	240	1000	100k	16	0.075	5×10^{-5}
ImageNet	SR 8×	240	1000	100k	16	0.075	5×10^{-5}
ImageNet	SR 12×	240	1000	100k	16	0.075	5×10^{-5}
ImageNet	SR 16×	256	1000	100k	16	0.075	5×10^{-5}

C Additional quantitative and mechanism analysis

We first examine the contribution of feasible centering under controlled training conditions, then assess measurement preservation separately. Additional reconstruction metrics and paired-image intervals supplement the primary evaluation.

C.1 Controlled effects of center features and source centering

Design and paired evaluation. The B/F/C study uses one common initial flow state and matches training images, time draws, noise, minibatch order, and 20k flow updates. B uses source mean and center feature (b, b); F uses (b, c); C uses (c, c). Each arm uses the full observation. The center is fixed for F/C and is trained separately. This study is an auxiliary controlled experiment, not a subset of the primary table relabeled as a new ablation.

Its 512-image evaluation set is reserved from the known training and center-validation images, with two matched noise draws per image. The Feasible Center's historical training comprises 100k updates on a larger central mask and 5k mask80 adaptation; the center-validation prefix and earlier 64-image diagnostics are excluded from this reserved cohort. The original evidence retains the exact image list. We average repeated draws within an image before computing intervals. The quoted paired intervals use 5,000 image-bootstrap repetitions with seed 2026091405. All reported seam intervals were recomputed from the original per-image records. All three arms use projected dynamics. In particular, F changes the center feature while keeping the source mean at $b(y)$; it is not the unprojected Center + FM variant. These three arms support the stated sequential contrasts, rather than a complete factorial interaction analysis.

Target energy and FM diagnostics. The target energy in Table 5 is $\mathbb{E}\|X - s(y)\|^2/k + \sigma^2$ with $k = \mathrm{rank}(P)$ and $\sigma = 1$. It is measured from paired endpoints, not from the integrated action of a trained rollout. The FM diagnostic averages eight common time midpoints and two noise draws. The theorem identifies a population ideal-action difference; this experiment validates its directly measurable endpoint-energy component and independently measures finite-network FM error and output quality.

Table 5: **Center features and source translation contribute at different scales.** B/F/C share training randomness and flow initialization. Residual and expected target energies are per null-space coordinate; target energy analytically averages the unit-variance source noise. FM MSE averages eight common time midpoints. Output metrics use one Euler flow call, plus one frozen-center call for F/C; LPIPS-V uses VGG. Seam RMSE is reference-gradient error on edges crossing the mask.

Arm	Residual energy	Target energy	FM MSE	PSNR	LPIPS-V	Seam RMSE
Pixel-NSFM (B)	0.2198	1.2198	0.04072	33.591	0.02936	0.02448
Center feature (F)	0.2198	1.2198	0.04044	33.734	0.02777	0.02185
LOCUS (C)	0.0230	1.0230	0.04019	33.746	0.02759	0.02172

Table 6: Full same-flow-NFE seam comparison in the controlled study. The frozen center alone has seam RMSE 0.02075 and requires one call. C and F add this call to the listed flow NFE; B does not. The final column compares C with B.

Solver	Flow NFE	Pixel-NSFM (B)	Center feature (F)	LOCUS (C)	Reduction
Euler	1	0.02448	0.02185	0.02172	11.29%
Euler	2	0.02093	0.02021	0.01983	5.24%
Heun	40	0.02718	0.02691	0.02643	2.73%

Boundary metric. For the interior 80×80 square, there are 320 horizontal/vertical edges with exactly one missing endpoint. On each edge we compute the difference between predicted and ground-truth luminance differences, square it, average over edges, and take the square root. We then average these image-wise RMSEs across the cohort. This equals the square root of Equation (14) after the stated luminance map and unit conversion, when observed pixels are fixed. It measures boundary fidelity, rather than penalizing large genuine edges. Internal-hole gradients, raw unclipped seams, and out-of-range fractions are retained in the evidence.

Prediction and perceptual quality. At one Euler call, C improves PSNR over B by 0.1546 dB, with interval $[0.0911, 0.2208]$. Adding the center feature accounts for 0.1426 dB ($[0.0805, 0.2074]$), while the additional source translation gives 0.0120 dB ($[-0.0072, 0.0311]$). The source change also decreases VGG LPIPS by 1.73×10^{-4} ($[-2.89, -0.60] \times 10^{-4}$) and decreases the seam error. These measurements support complementary representation and source effects without attributing the complete LOCUS/Pixel-NSFM difference solely to a source shift.

The center has high reconstruction accuracy but does not have the lowest perceptual error. Its VGG LPIPS is 0.03237, compared with 0.02759 for C at one Euler call.

Sampling budgets and training cost. At two total neural calls, C with one flow call has higher PSNR than B with two flow calls, but B has lower seam RMSE (0.02093 versus 0.02172) and lower VGG LPIPS (0.02567 versus 0.02759). Thus the same-flow-NFE seam improvements in Table 6 should not be stated as dominance at every equal-total-call budget. Recorded flow-training GPU times are 1.595, 2.343, and 2.489 hours for B/F/C; they exclude the separately trained center's historical cost.

C.2 Measurement preservation: numerical and spatial evidence

Numerical trajectory consistency. We distinguish the raw endpoint residual, the maximum residual along the sampled trajectory for each image, and the residual after image-range processing. Masks preserve observed pixels exactly in these recorded constrained evaluations. Separable SR incurs floating-point projection error. These measurements concern the exact operator of each run; a residual under a different resize implementation is not equivalent. Table 7 summarizes the largest per-image trajectory maxima read from the stored CSVs.

Table 7: Largest LOCUS measurement residual across images in each primary cohort, in model units. Each trajectory value is the maximum over the recorded sampling states before final image-range processing. Mask values are exactly zero in the stored records.

Dataset	Task	Raw endpoint	Trajectory	Processed
FFHQ	Block	0	0	0
FFHQ	Random 70%	0	0	0
FFHQ	SR 4×	3.58×10^{-7}	4.17×10^{-7}	4.77×10^{-7}
FFHQ	SR 8×	3.58×10^{-7}	3.58×10^{-7}	3.58×10^{-7}
FFHQ	SR 12×	3.58×10^{-7}	4.17×10^{-7}	4.17×10^{-7}
FFHQ	SR 16×	3.58×10^{-7}	3.58×10^{-7}	5.36×10^{-7}
CelebA-HQ	Block	0	0	0
CelebA-HQ	Random 70%	0	0	0
CelebA-HQ	SR 4×	4.17×10^{-7}	4.17×10^{-7}	4.77×10^{-7}
CelebA-HQ	SR 8×	3.58×10^{-7}	3.58×10^{-7}	3.58×10^{-7}
CelebA-HQ	SR 12×	3.58×10^{-7}	4.17×10^{-7}	3.58×10^{-7}
CelebA-HQ	SR 16×	3.58×10^{-7}	4.77×10^{-7}	4.77×10^{-7}
ImageNet	Block	0	0	0
ImageNet	Random 70%	0	0	0
ImageNet	SR 4×	4.17×10^{-7}	4.17×10^{-7}	4.77×10^{-7}
ImageNet	SR 8×	3.58×10^{-7}	3.58×10^{-7}	3.58×10^{-7}
ImageNet	SR 12×	3.58×10^{-7}	4.17×10^{-7}	4.17×10^{-7}
ImageNet	SR 16×	3.58×10^{-7}	4.77×10^{-7}	5.36×10^{-7}

Spatial residual comparison. Figure 6 compares measurement preservation across the five internal variants on matched FFHQ examples. For each reconstruction $\hat{x}$, we reapply the original forward operator and visualize the absolute measurement residual $|A\hat{x} - y|$, averaged over RGB channels. These maps assess agreement with the measurements, rather than reconstruction error relative to the ground truth. For inpainting, missing locations contribute zero by construction, so consistency is assessed at observed locations. For super-resolution, residuals are evaluated in the low-resolution measurement space.

Roles of centering and null-space constraints. The Feasible Center provides an observation-dependent reference satisfying $Ac(y) = y$. However, Center + FM exhibits visible measurement residuals despite using a feasible center: unprojected source noise and velocities are not restricted to the measurement-consistent fiber. Full-Space FM likewise exhibits visible residuals. Both Pixel-NSFM and LOCUS preserve measurements through their null-space construction. Their shared consistency separates the role of the constraint from that of learned centering: the constraint determines where generation may move, while the feasible center determines the reference around which unresolved content is modeled. The reconstruction and transport-energy benefits of learned centering are evaluated separately in Tables 1 and 5.

Visualization and numerical precision. The maps are computed from saved 8-bit reconstructions and therefore include image-saving effects. Near-black regions indicate small residuals at the displayed scale and should not be interpreted as proof of exact numerical equality. Table 7 separately reports raw endpoint, trajectory, and processed residuals. In particular, the endpoint maps do not by themselves establish consistency throughout sampling.

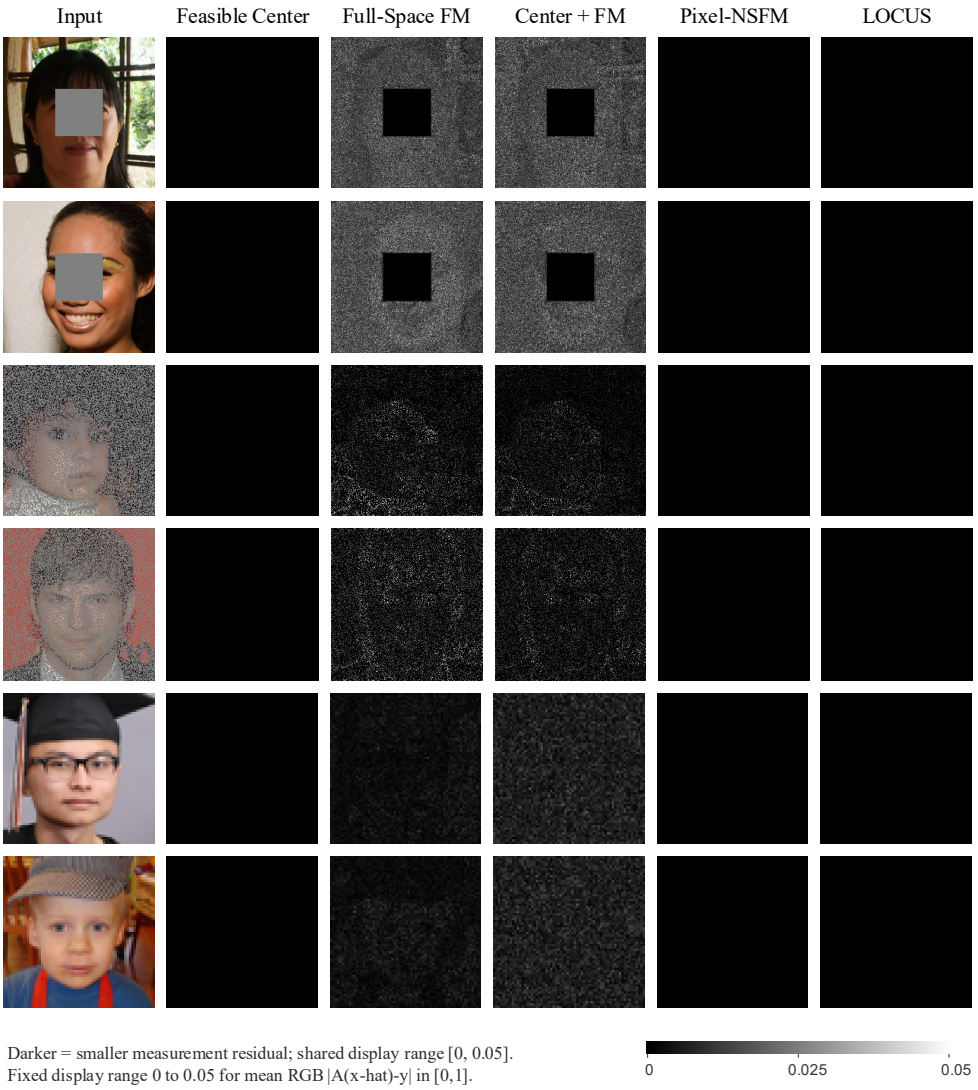

Figure 6: **Measurement preservation across internal variants.** Rows show two FFHQ examples each of block inpainting, random-pixel inpainting with 70% missing pixels, and $4\times$ super-resolution. After Input, columns show measurement-residual maps for Feasible Center, Full-Space FM, Center + FM, Pixel-NSFM, and LOCUS. Maps display mean absolute RGB measurement residuals computed from saved 8-bit reconstructions, in $[0, 1]$ image-intensity units, using a common display range of $[0, 0.05]$. Darker values indicate smaller residuals. For inpainting, missing locations are zero by construction. SR residuals are computed at the native measurement resolution and enlarged using nearest-neighbor interpolation for display.

C.3 Additional reconstruction metrics

Table 8 complements the PSNR and LPIPS results in Table 1 with SSIM and DISTS for the paired null-space flows. It uses the same native-float evaluation and primary cohorts, including the separately identified FFHQ random-pixel pilot. These reconstruction metrics are distinct from measurement residuals.

Table 8: Additional primary-suite metrics for the paired flows. SSIM is higher-is-better; DISTS is lower-is-better. Native float evaluation and cohorts match Table 1.

Dataset	Task	Pixel-NSFM SSIM	LOCUS SSIM	Pixel-NSFM DISTS	LOCUS DISTS
FFHQ	Block	0.9506	0.9521	0.0138	0.0132
FFHQ	Random 70%	0.8335	0.8541	0.1279	0.1053
FFHQ	SR 4×	0.8482	0.8516	0.0796	0.0775
FFHQ	SR 8×	0.6890	0.6994	0.1394	0.1356
FFHQ	SR 12×	0.5862	0.6062	0.1758	0.1687
FFHQ	SR 16×	0.5457	0.5694	0.1941	0.1870
CelebA-HQ	Block	0.9555	0.9568	0.0119	0.0115
CelebA-HQ	Random 70%	0.8768	0.8814	0.0814	0.0755
CelebA-HQ	SR 4×	0.8462	0.8479	0.0723	0.0712
CelebA-HQ	SR 8×	0.7098	0.7185	0.1247	0.1207
CelebA-HQ	SR 12×	0.6234	0.6359	0.1576	0.1505
CelebA-HQ	SR 16×	0.5865	0.6049	0.1761	0.1678
ImageNet	Block	0.9232	0.9243	0.0617	0.0602
ImageNet	Random 70%	0.7610	0.7697	0.1045	0.0941
ImageNet	SR 4×	0.6685	0.6800	0.1509	0.1485
ImageNet	SR 8×	0.4893	0.5099	0.2413	0.2347
ImageNet	SR 12×	0.4219	0.4384	0.2927	0.2819
ImageNet	SR 16×	0.3859	0.4024	0.3162	0.3066

C.4 Paired image uncertainty

We pair the LOCUS and Pixel-NSFM metric CSVs by their test image identifiers and bootstrap images with replacement 5,000 times (seed 20260922). Table 9 reports percentile 95% intervals for the difference of means, LOCUS minus Pixel-NSFM. These are pointwise intervals, without a simultaneous multiple-comparison adjustment. They describe variability over the fixed trained checkpoints' test-image population, and do not estimate training-seed uncertainty. Both metric differences have the favorable sign throughout their pointwise intervals in all listed pairs.

Table 9: Paired LOCUS minus Pixel-NSFM differences, with image-bootstrap 95% intervals. Positive PSNR and negative LPIPS favor LOCUS. FFHQ random70 uses its separately identified 100-image pilot.

Dataset	Task	Δ PSNR [95% interval]	Δ LPIPS [95% interval]
FFHQ	Block	+0.305 [+0.276, +0.332]	-0.0010 [-0.0011, -0.0009]
FFHQ	Random 70%	+0.687 [+0.620, +0.758]	-0.0232 [-0.0253, -0.0214]
FFHQ	SR 4×	+0.203 [+0.194, +0.213]	-0.0019 [-0.0021, -0.0017]
FFHQ	SR 8×	+0.254 [+0.246, +0.262]	-0.0047 [-0.0051, -0.0043]
FFHQ	SR 12×	+0.360 [+0.351, +0.370]	-0.0112 [-0.0118, -0.0106]
FFHQ	SR 16×	+0.404 [+0.394, +0.414]	-0.0134 [-0.0141, -0.0128]
CelebA-HQ	Block	+0.340 [+0.313, +0.369]	-0.0009 [-0.0009, -0.0008]
CelebA-HQ	Random 70%	+0.286 [+0.281, +0.291]	-0.0038 [-0.0039, -0.0038]
CelebA-HQ	SR 4×	+0.113 [+0.105, +0.120]	-0.0009 [-0.0010, -0.0008]
CelebA-HQ	SR 8×	+0.175 [+0.167, +0.184]	-0.0055 [-0.0058, -0.0052]
CelebA-HQ	SR 12×	+0.229 [+0.220, +0.238]	-0.0109 [-0.0114, -0.0105]
CelebA-HQ	SR 16×	+0.281 [+0.271, +0.291]	-0.0139 [-0.0145, -0.0134]
ImageNet	Block	+0.229 [+0.187, +0.271]	-0.0018 [-0.0021, -0.0014]
ImageNet	Random 70%	+0.306 [+0.295, +0.317]	-0.0061 [-0.0064, -0.0059]
ImageNet	SR 4×	+0.301 [+0.281, +0.323]	-0.0028 [-0.0038, -0.0017]
ImageNet	SR 8×	+0.364 [+0.344, +0.387]	-0.0058 [-0.0076, -0.0039]
ImageNet	SR 12×	+0.296 [+0.276, +0.318]	-0.0211 [-0.0230, -0.0191]
ImageNet	SR 16×	+0.313 [+0.292, +0.336]	-0.0205 [-0.0225, -0.0185]

D External comparisons

D.1 Evaluation conventions

The cross-task evaluation in Table 2 reports mean RGB PNG metrics over 2,000 FFHQ test images per task. Its SR results follow the stated SVD-based bicubic protocol, separately from the separable operator in Table 1; FlowDPS uses the high-resolution adaptation identified in the table.

The dedicated random-pixel evaluation in Table 3 also uses 2,000 FFHQ test images, with clipped-float scoring and the task-training and Euler20 configuration described below. These are separate evaluation protocols; their results are not pooled.

The 100-image block-inpainting and separable-SR comparisons in this appendix retain their own evaluation protocols. They score the exact saved RGB PNGs without a border crop, using AlexNet LPIPS; their results are not merged with either main-text external evaluation or the native-float primary comparison. DC100 denotes 100 box/affine correction cycles using the declared operator. It improves comparability of measurement fidelity but does not equalize prior training, backbone capacity, or neural computation. NFE entries retain the recorded flow/model evaluation counts. Feasible Center has zero flow calls but requires one predictor call; LOCUS's listed flow NFE excludes its additional cached center call. These counts do not imply equal computational cost across architectures.

D.2 FFHQ random-pixel benchmark

The full 2,000-image benchmark uses a fixed 70% missing mask at 256×256. The LOCUS center and flow each receive 100k task-training updates, with batch 16 and $\sigma = 1$ in model units. The implementation uses pure center MSE and ordinary velocity MSE, without auxiliary losses. Its evaluation retains the full prescribed test split and uses Euler20.

PMRF is trained on the same FFHQ restoration task with its official-code MMSE and flow stages, each for 100k updates. Its training IDs are 0–66,999. It is not an out-of-domain CelebA model. Its 121.5M-parameter architecture, full-space noise scale 0.1 in $[0, 1]$, AdamW hyperparameters, and learning rate 5×10^{-4} differ from LOCUS. Consequently, Table 3 is a task-level benchmark rather than a same-architecture causal test. Sampling calls also have different computational costs. The source evaluation records timing separately; mean LOCUS sampling time is 0.274 seconds/image in its recorded environment, including the first sample, so we do not convert call counts into a universal speed ratio.

Replacing only known pixels in PMRF's final output is a postprocessing diagnostic. At flow NFE 1/5/25, it produces PSNR 33.52/32.93/32.26 and LPIPS 0.0477/0.0364/0.0359, compared with the native entries in the main table. It requires no new learned model but does not constrain the preceding trajectory. A separate D-Flow (Ben-Hamu et al., 2024) run covers only 64 images (34.57 dB, 0.0348 LPIPS, mean 5,225.5 model evaluations), so it is not ranked in the full-cohort table.

D.3 Block inpainting

Block inpainting includes RePaint, a pretrained-diffusion sampling method, and LaMa, a supervised feed-forward inpainting model (Lugmayr et al., 2022; Suvorov et al., 2022). FLOWER and PnP-Flow can also solve the masking inverse problem; their use is valid when the observation operator is matched. We distinguish native runs, MC1 variants, legacy configurations, and the locally adapted FlowDPS implementation. A name alone does not identify identical inference settings.

CelebA-HQ. Table 10 shows a different outcome: DDNM SVD outperforms LOCUS in PSNR, LPIPS, and DISTS. The fidelity–perceptual-similarity tradeoff therefore does not explain every performance difference.

Table 10: CelebA-HQ block inpainting. All rows use the same 100-image operator protocol and RGB PNG scoring. DC100 denotes the shared 100-cycle box/affine correction. These pretrained baselines have different training data and inference costs.

Method	Flow/model NFE	PSNR ↑	LPIPS ↓	DISTS ↓
PnP-Flow MC1	100	30.95	0.0751	0.1149
FLOWER MC1	100	32.06	0.0223	0.0154
DDNM SVD	100	35.48	0.0094	0.0087
LOCUS	40	32.35	0.0142	0.0116
DDNM (legacy)	100	36.42	0.0144	0.0136
FlowDPS (custom)	–	28.91	0.0363	0.0293
PnP-Flow (legacy)	100	28.92	0.1930	0.2294
FLOWER (legacy)	100	32.86	0.0254	0.0160
RePaint	76	30.54	0.0215	0.0262
LaMa	–	28.15	0.0295	0.0227

D.4 Super-resolution with the separable operator

SR comparisons use public HAT, SwinIR, ResShift, SinSR, TSD-SR, and OSEDiff checkpoints (Chen et al., 2023; Liang et al., 2021; Yue et al., 2023; Wang et al., 2024; Dong et al., 2025; Wu et al., 2024). We retain both native and shared-correction versions where evaluated. The three tables below use $4\times$ downsampling with the declared separable operator. Selected $8\times$ and $12\times$ visual comparisons are reported separately in Appendix E.2.

FFHQ. In Table 11, LOCUS has lower LPIPS than ResShift (bicubic, 4) + DC100, but slightly higher DISTS and lower PSNR. The metrics therefore do not support a uniform ranking of these methods.

Table 11: FFHQ $4\times$ SR (separable). All rows use the same 100-image operator protocol and RGB PNG scoring. DC100 denotes the shared 100-cycle box/affine correction. These pretrained baselines have different training data and inference costs.

Method	Flow/model NFE	PSNR ↑	LPIPS ↓	DISTS ↓
HAT	1	31.69	0.1430	0.1479
OSEDiff	1	25.53	0.1090	0.1214
Feasible Center	0	31.55	0.1384	0.1376
ResShift (bicubic, 4)	4	29.53	0.0604	0.0824
ResShift (realSR, 15)	15	27.82	0.1004	0.1212
SinSR	1	27.31	0.0882	0.1129
SwinIR	1	31.57	0.1437	0.1498
TSD-SR (MSE)	1	26.79	0.0833	0.1046
HAT + DC100	1	31.71	0.1426	0.1453
OSEDiff + DC100	1	28.87	0.0756	0.0965
ResShift (bicubic, 4) + DC100	4	30.03	0.0573	0.0809
ResShift (realSR, 15) + DC100	15	28.96	0.0844	0.1044
SinSR + DC100	1	28.51	0.0728	0.1034
SwinIR + DC100	1	31.60	0.1433	0.1470
TSD-SR (MSE) + DC100	1	28.01	0.0659	0.0915
LOCUS (40)	40	29.54	0.0547	0.0823

CelebA-HQ. In Table 12, LOCUS obtains 30.57 dB, 0.0441 LPIPS, and 0.0705 DISTS, compared with 31.29 dB, 0.0508, and 0.0747 for ResShift (bicubic, 4) + DC100. Here, LOCUS reduces LPIPS and DISTS by 13.2% and 5.6%, respectively, at 0.72 dB lower PSNR.

Table 12: CelebA-HQ $4\times$ SR (separable). All rows use the same 100-image operator protocol and RGB PNG scoring. DC100 denotes the shared 100-cycle box/affine correction. These pretrained baselines have different training data and inference costs.

Method	Flow/model NFE	PSNR ↑	LPIPS ↓	DISTS ↓
HAT	1	32.93	0.1297	0.1444
OSEDiff	1	26.04	0.0986	0.1234
Feasible Center	0	33.20	0.1205	0.1303
ResShift (bicubic, 4)	4	30.76	0.0533	0.0765
ResShift (realSR, 15)	15	28.77	0.0882	0.1159
SinSR	1	28.11	0.0797	0.1078
SwinIR	1	32.84	0.1305	0.1459
TSD-SR (MSE)	1	27.03	0.0856	0.1120
HAT + DC100	1	32.95	0.1295	0.1410
OSEDiff + DC100	1	29.58	0.0664	0.0873
ResShift (bicubic, 4) + DC100	4	31.29	0.0508	0.0747
ResShift (realSR, 15) + DC100	15	30.02	0.0728	0.0954
SinSR + DC100	1	29.40	0.0655	0.0940
SwinIR + DC100	1	32.86	0.1302	0.1422
TSD-SR (MSE) + DC100	1	28.39	0.0705	0.0948
LOCUS (40)	40	30.57	0.0441	0.0705

ImageNet. In Table 13, ResShift (bicubic, 4) + DC100 outperforms LOCUS in PSNR, LPIPS, and DISTS. Together, these comparisons support perceptual advantages in specific settings rather than uniform superiority across methods and tasks.

Table 13: ImageNet $4\times$ SR (separable). All rows use the same 100-image operator protocol and RGB PNG scoring. DC100 denotes the shared 100-cycle box/affine correction. These pretrained baselines have different training data and inference costs.

Method	Flow/model NFE	PSNR ↑	LPIPS ↓	DISTS ↓
HAT	1	27.17	0.2345	0.1944
OSEDiff	1	23.22	0.1904	0.1739
Feasible Center	0	27.24	0.2576	0.1974
ResShift (bicubic, 4)	4	25.08	0.1189	0.1235
ResShift (realSR, 15)	15	24.15	0.1959	0.1800
SinSR	1	23.59	0.1737	0.1650
SwinIR	1	27.42	0.2393	0.1978
TSD-SR (MSE)	1	23.07	0.1592	0.1541
HAT + DC100	1	27.21	0.2341	0.1920
OSEDiff + DC100	1	24.91	0.1575	0.1471
ResShift (bicubic, 4) + DC100	4	25.41	0.1149	0.1215
ResShift (realSR, 15) + DC100	15	24.92	0.1792	0.1646
SinSR + DC100	1	24.39	0.1557	0.1566
SwinIR + DC100	1	27.47	0.2387	0.1950
TSD-SR (MSE) + DC100	1	23.93	0.1409	0.1383
LOCUS (40)	40	24.95	0.1411	0.1519

E Qualitative evaluation

The following comparisons examine local detail, large-factor super-resolution, and transfer across face-image datasets. Visible texture and reference-based perceptual similarity do not by themselves establish exact recovery of unobserved high-frequency content or certify realism. Aggregate comparisons are reported in Tables 1, 2, and 3, with additional evaluations in Appendix D.

E.1 Random-pixel inpainting: local detail

Figure 7 presents additional FFHQ examples with 70% of pixels missing and corresponding enlarged facial regions. The first example highlights lip contours and surrounding skin texture; the second and third focus on eyelid structure and periocular texture. In these selected examples, LOCUS retains more of the fine structures visible in the ground truth, whereas several competing reconstructions exhibit smoother local appearance and attenuated texture. The figure illustrates reconstruction quality of the complete method; measurement preservation is assessed in Appendix C.2, and the controlled centering study is reported in Appendix C.1.

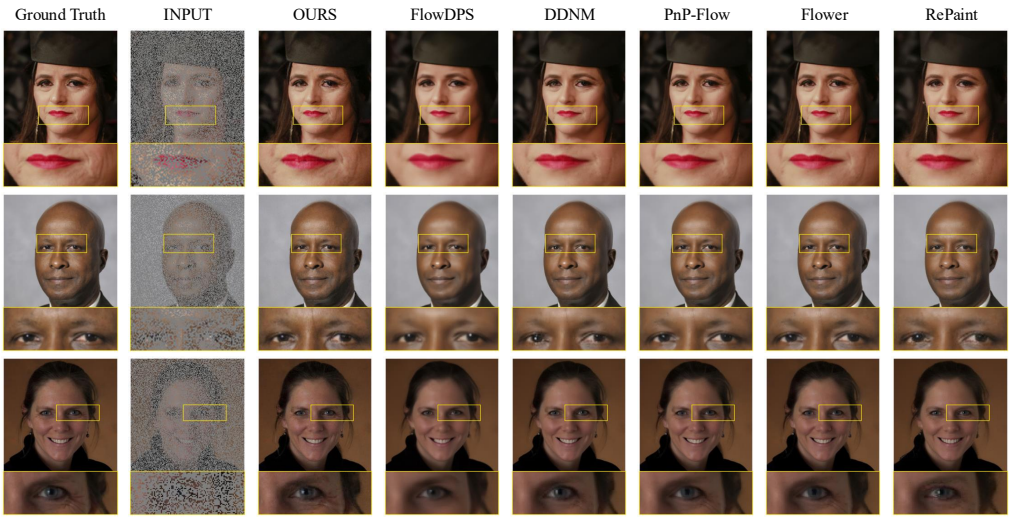

Figure 7: **Fine-detail recovery in random-pixel inpainting.** Each row presents one example. Columns show Ground Truth, Input, LOCUS, FlowDPS, DDNM, PnP-Flow, FLOWER, and RePaint. Yellow boxes mark corresponding regions, with enlarged views directly beneath each full image. The crops highlight differences in lip contours, eye structure, and local skin texture relative to the ground truth.

E.2 Super-resolution at large upscaling factors

Scope and example selection. Figures 8 and 9 present selected qualitative comparisons at $8\times$ and $12\times$ super-resolution. The $8\times$ examples are from FFHQ; the $12\times$ examples include both CelebA-HQ and FFHQ. The examples are selected from candidate pools ranked by LOCUS PSNR and illustrate reconstruction behavior; aggregate performance is assessed separately on the full evaluation cohorts. The two figures contain different images and do not constitute a paired comparison of degradation as the upscaling factor increases. In the separate primary evaluation (Table 1), LOCUS improves PSNR by 0.17–0.36 dB and reduces LPIPS by 0.0047–0.0112 relative to Pixel-NSFM across the four FFHQ and CelebA-HQ settings at $8\times$ and $12\times$. These results concern the paired null-space flows, not a full-cohort ranking of the external methods displayed here.

Selected $8\times$ examples. In Figure 8, LOCUS produces well-defined facial contours and visible skin and hair texture, including wrinkles and eyeglass boundaries. Several competing outputs appear smoother, although some baselines also preserve substantial detail in individual examples.



Figure 8: **Selected qualitative comparisons for $8\times$ super-resolution on FFHQ.** Each row presents one example, reconstructing a 256×256 image from a 32×32 observation. Columns show Ground Truth, Input, LOCUS, FlowDPS, DDNM, PnP-Flow, and FLOWER. Low-resolution inputs are enlarged for display. The examples highlight differences in facial contours, skin texture, hair, and eyeglass detail. The selection procedure is described in Appendix E.2.

Selected $12\times$ **examples.** In Figure 9, LOCUS retains relatively clear eye and mouth boundaries and local texture in the displayed cases. These observations describe selected reconstruction differences under severe downsampling; they do not establish that all generated fine structures match the ground truth or that LOCUS uniformly outperforms the external methods.



Figure 9: **Selected qualitative comparisons for** $12\times$ **super-resolution on CelebA-HQ and FFHQ.** Rows 1–4 show CelebA-HQ examples, and rows 5–6 show FFHQ examples. Each reconstruction has resolution 240×240, with a 20×20 observation. Columns show Ground Truth, Input, LOCUS, FlowDPS, DDNM, PnP-Flow, and FLOWER. Low-resolution inputs are enlarged for display. The examples expose differences in local facial structure and texture under severe downsampling. The selection procedure is described in Appendix E.2.

E.3 Cross-dataset transfer from FFHQ to CelebA-HQ

We evaluate FFHQ-trained LOCUS models on CelebA-HQ images to examine transfer across face-image datasets. Figure 10 presents qualitative results across block inpainting, super-resolution, and random-pixel recovery. This cross-dataset evaluation is reported separately from the aggregate comparisons in Table 1 and Appendix D. The training-to-test transfer statement refers to LOCUS; the figure does not establish a shared cross-dataset training protocol for all baselines. These examples provide qualitative transfer evidence rather than an aggregate estimate of robustness under distribution shift.



Figure 10: **FFHQ-to-CelebA-HQ generalization of LOCUS.** LOCUS is trained on FFHQ and evaluated on CelebA-HQ. Rows show block inpainting, $4\times$, $8\times$, $12\times$, and $16\times$ super-resolution, and random-pixel recovery (top to bottom). Columns show the reference, observation, LOCUS, and four external restoration methods. FlowDPS* denotes our adapted implementation.

F LIMITATIONS AND DIAGNOSTIC EVALUATIONS

The population identities and their finite-model interpretation are developed in Appendix A; the diagnostic evaluations below illustrate why those identities alone do not determine empirical rankings.

F.1 SCOPE OF THE OBSERVATION MODEL AND TRANSFER EVIDENCE

The method as analyzed assumes known noiseless measurements. For a full-column-rank operator, $\ker(A) = \{0\}$ and the hard-feasible construction has no unresolved coordinates. In particular, noisy denoising with $A = I$ is not addressed by imposing $Ax = y$ exactly. Known masks that vary across examples retain the same feasible construction when the operator is conditioned on and its projector is applied per example, but generalization beyond the fixed masks used here remains untested. Noisy measurements, uncertain operators, and nonlinear forward maps require extensions beyond the present noiseless linear formulation. Appendix E.3 provides qualitative evidence of FFHQ-to-CelebA-HQ transfer. Calibrated posterior coverage and broad robustness to distribution shifts remain unestablished.

F.2 EARLIER RANDOM-PIXEL CHECKPOINT

An ImageNet random70 evaluation after the earlier 20k training stage gives LOCUS 21.7195 dB and 0.4021 LPIPS, versus Pixel-NSFM 22.3995 dB and 0.3233 LPIPS. The primary table uses the later 100k stage and a class-stratified 1,000-image cohort, where LOCUS reaches 26.5366 dB and 0.0637 LPIPS. The old 1,000-image pilot contains five images from each of 200 classes, whereas the current test set contains one image from each of 1,000 classes. These are different evaluation cohorts and cannot be plotted as a matched learning curve or used to attribute the change solely to more training. The earlier record shows that a feasible parameterization does not determine finite-model quality by itself.

F.3 ALTERNATE SUPER-RESOLUTION PROTOCOL

The native-256 ImageNet $4\times$ SVD experiment at a 30k stage uses a different architecture, auxiliary losses, and the thresholded SVD operator. On its strict 100-image subset, LOCUS gives 23.3648 dB compared with Pixel-NSFM's 23.5030 dB. This experiment is not interchangeable with the primary 240-resolution separable SR row, and it is not part of the pure-FM matched-stage collection. Its configuration and complete source results are retained in the evidence package. Operator changes and additional training objectives require their own evaluation.

F.4 IMAGENET SUPER-RESOLUTION AND RANDOM-PIXEL RECOVERY

Figures 11 and 12 present selected ImageNet reconstructions for $4\times$ and $8\times$ super-resolution and random-pixel recovery with 70% missing pixels. These examples complement the face-image visualizations with natural scenes containing diverse objects and textures. The accompanying reconstruction-error maps show the remaining discrepancies relative to ground truth.

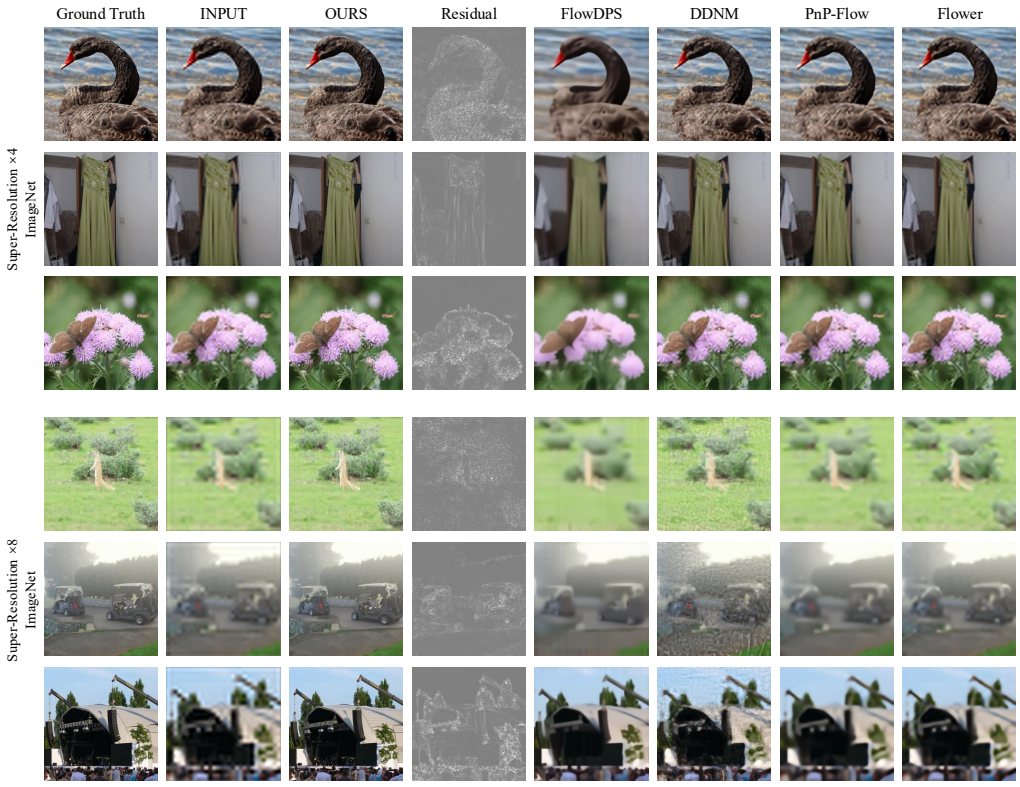

Figure 11: **Selected ImageNet super-resolution examples.** The top three rows show $4\times$ SR and the bottom three rows show $8\times$ SR, using 240×240 reference images. From left to right: ground truth, the displayed LOCUS input, LOCUS (labeled OURS), its reconstruction-error map (labeled Residual), FlowDPS, DDNM, PnP-Flow, and FLOWER. Residual denotes the channel-averaged absolute difference between the final LOCUS reconstruction and ground truth; mid-gray represents zero error and brighter pixels indicate larger errors. This differs from the residual relative to the feasible center modeled by the flow. LOCUS uses a separable bicubic operator, whereas the external methods use SRConv/SVD bicubic implementations, resulting in different observations. These panels provide qualitative references under the respective degradation models and sampling settings.

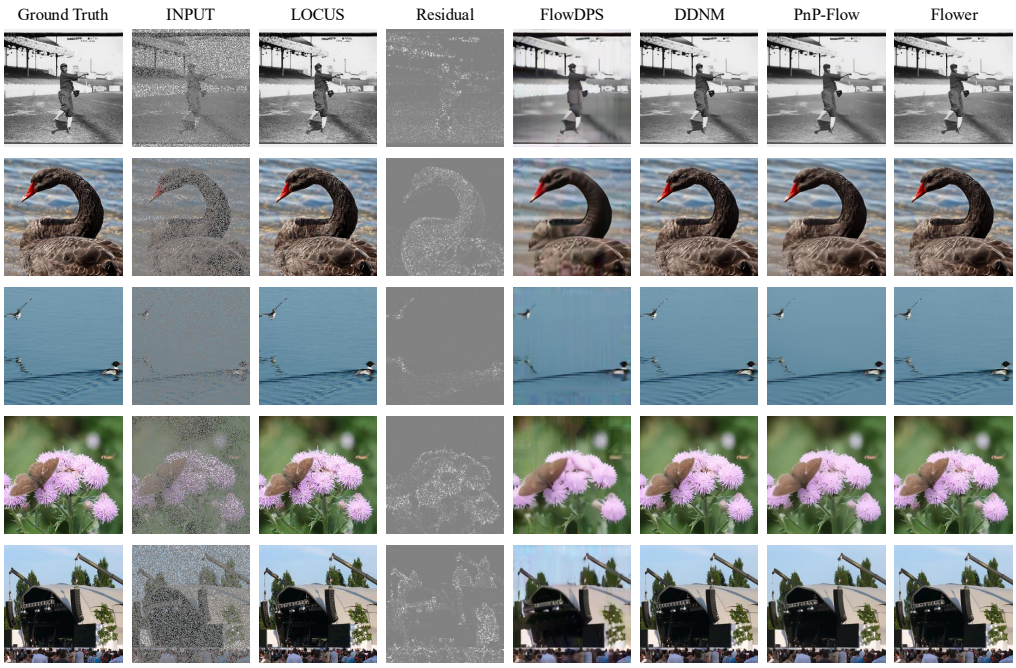

Figure 12: **Selected ImageNet random-pixel recovery examples.** Five 256×256 images are reconstructed from observations retaining 30% of the pixels, with the remaining 70% missing. All methods use the same observation for each example. Column order and reconstruction-error visualization follow Figure 11. OURS denotes LOCUS, and Residual visualizes its reconstruction error relative to ground truth. The examples illustrate recovery of object boundaries and fine textures across different natural scenes. Methods retain their recorded training and sampling settings; computational budgets are not matched.