DYNAFILL: DYNAMIC MANIFOLD-CONSISTENT FLOW SAMPLING FOR LOCALIZED IMAGE INPAINTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Image inpainting using pre-trained generative models typically relies on inversion, which is computationally expensive and often accumulates errors. Recent inversion-free methods struggle to preserve unmasked content while generating coherent structures in masked regions. In this paper, we introduce DynaFill, a novel inversion-free framework for localized image inpainting. Unlike existing fixed-anchor inversion-free editing methods, DynaFill is driven by a Dynamic Manifold-Consistent Sampling strategy. Specifically, we decouple the sampling trajectory from the static source prior by dynamically anchoring the noise perturbation to the current generation state rather than the corrupted original image. This allows the generative path to adaptively slide along the data manifold, ensuring that the filled content remains structurally faithful to the global context. To further enhance efficiency and locality, we integrate an enhanced ODE solver that accelerates sampling by $2.2\times$, alongside a localized adaptive manifold blending mechanism that strictly preserves unmasked pixels while enabling precise boundary continuity. Extensive experimental results demonstrate that DynaFill achieves state-of-the-art performance, validating the effectiveness of the proposed method for inversion-free inpainting.

1 INTRODUCTION

Image inpaintingłthe task of plausibly filling corrupted regions in an imagełis a long-standing challenge in computer vision with wide-ranging applications, including photo restoration, object removal, and creative content editing. Recent advances in generative modeling, particularly diffusion models (DMs) (Ho et al., 2020; Song et al., 2020) and rectified flow models (Lipman et al., 2023; Liu et al., 2023; Esser et al., 2024), have substantially advanced synthesis and editing.

It is crucial to recognize that image inpainting is fundamentally different from global image generation. Global generation synthesizes an entire image from scratch, where the model enjoys complete freedom over every pixel. Inpainting, by contrast, is a localized generation problem: only a fraction of pixels are unknown, while the vast majority must be preserved with pixel-level exactitude. This imposes two critical constraints: first, the generated content must seamlessly integrate with the surrounding context, requiring precise boundary continuity and color consistency with known pixels; second, the generative process must be strictly spatially confinedłany modification to unmasked pixels constitutes a failure. These constraints make inpainting inherently more challenging than global generation. A method that excels at global editing may fail catastrophically when required to operate locally, precisely because it lacks the mechanisms to decouple the generative dynamics of masked and unmasked regions.

Existing approaches for image inpainting can be broadly grouped according to whether they require additional training. One family of methods involves training or fine-tuning. For example, early CNN-based (Iizuka et al., 2017; Liu et al., 2018) and GAN-based (Pathak et al., 2016; Yu et al., 2018) methods learn a direct mapping from masked images to completed outputs, while more recent approaches fine-tune pre-trained DMs (Saharia et al., 2021; Nichol et al., 2021)or train dedicated conditional models (Lugmayr et al., 2022) to achieve better quality. However, these methods incur substantial computational costs: training a diffusion model from scratch or even fine-tuning it requires significant GPU resources and large-scale datasets, making them impractical for many real-world scenarios.

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.

This high training cost has motivated a line of research toward training-free methods that leverage pre-trained generative models as powerful priors without additional optimization or fine-tuning. Among these, the inversion-based paradigm has emerged as the most widely adopted approach. Several representative works have demonstrated its effectiveness: Null-text Inversion(Mokady et al., 2023) introduced an optimization-based inversion technique to improve edit fidelity; EDICT (Wallace et al., 2023) proposed exact inversion via coupled transformations to minimize discretization errors. However,(Huberman-Spiegelglas et al., 2024) revealed that even exact inversion fails to preserve structure, motivating alternative approaches. Inversion-based methods suffer from two critical drawbacks: (i) the inversion step is computationally expensive and introduces discretization errors that accumulate during re-sampling; (ii) even when inversion is exact, the generated content often fails to preserve the global structure of the source image. Importantly, these methods were primarily designed for global generation tasksłwhere the entire image can be modifiedłand their adaptation to localized inpainting remains an afterthought rather than a principled design. To deal with inpainting, (Avrahami et al., 2023) proposed a hard replacement strategy using the source masked latent and the resulting latent. However, this strategy is commonly used to enforce locality introduces visible discontinuities at mask boundaries.

Recently, a new training-free paradigm has emerged that completely bypasses the inversion step. Instead of projecting the source image into noise space and back, inversion-free flow editing (Kulikov et al., 2025; Deng et al., 2025; Xu et al., 2023) constructs a direct ODE path between the source and target distributions, achieving strong structure preservation without the computational overhead of inversion. Despite their effectiveness in global image editing, existing inversion-free methods face several challenges when applied to localized inpainting. First, they typically rely on explicit Euler integration, which requires a large number of sampling steps to maintain generation quality. Second, these methods generally lack explicit mechanisms for strictly preserving unmasked regions, as the ODE updates the entire latent representation. In addition, static anchoring strategies tie noisy source samples to a fixed source latent throughout sampling. While such a design is effective for global editing, it can become restrictive for localized inpainting, where the masked region requires sufficient generative freedom to synthesize novel content. These limitations reveal a mismatch between global inversion-free editing and localized inpainting. The core challenge lies in spatially decoupling the generative dynamics: the masked region must explore the data manifold to synthesize novel content, while the unmasked region must remain strictly anchored to the source. Current inversion-free methods treat all pixels uniformly and lack the mechanisms to enforce this dichotomy. To the best of our knowledge, existing inversion-free flow-based editing methods have not explicitly addressed strict consistency of local generation and boundary continuity for localized image inpainting.

To bridge this gap, we propose **DynaFill**, a novel framework that redefines how rectified flow models can be leveraged for localized image inpainting within the inversion-free paradigm. Specifically, we introduce three synergistic innovations: (1) **an enhanced ODE integrator** based on the predictor-corrector method that achieves second-order accuracy while reducing sampling steps; (2) **dynamic manifold anchoring**, a time-adaptive noise sampling strategy that replaces the static source reference with a dynamically evolving anchor; and (3) **adaptive manifold blending**, a locally consistent generation mechanism that fuses static and dynamic anchors at the noise construction stage to strictly preserve unmasked pixels while eliminating boundary discontinuities. Collectively, these designs enable DynaFill to effectively extend inversion-free rectified-flow editing to localized image inpainting. Extensive experiments demonstrate that DynaFill achieves state-of-the-art performance both quantitatively and qualitatively. Our code and models will be made publicly available to facilitate further research.

2 METHODOLOGY

In this section, we present DynaFill, a novel inversion-free framework for localized image inpainting based on pre-trained rectified flow models.

2.1 PRELIMINARIES: INVERSION-FREE EDITING

Flow matching learns a transportation between two distributions X_0 (data) and X_1 (noise, typically $\mathcal{N}(0, \mathbf{I})$):

$$\mathrm{d}Z_t = V(Z_t, t)\mathrm{d}t, \quad t \in [0, 1] \tag{1}$$

where $V(Z_t, t)$ is a time-dependent velocity field parameterized by a neural network. Rectified Flow reduces discretization errors during ordinary differential equation (ODE) solving by straightening the trajectory, thus enabling generation to be completed in fewer steps. The training objective ensures that the marginal distribution of Z_t follows a linear interpolation between X_0 and X_1:

$$Z_t \sim (1-t)X_0 + tX_1, \quad t \in [0, 1] \tag{2}$$

Sampling is performed by solving an ODE backward in time from $t = 1$ (pure noise) to $t = 0$ (generated image). Together with inversion, this approach has been applied to various image generation tasks. Inversion first inverts the source image into the model's noise space via ODE integration, then resamples it under a target condition.

However, inversion-free methods are becoming hot, as inversion usually accumulates errors from velocity and numerical integration. Training-free and inversion-free approaches typically compute an editing field by differencing the drifts conditioned on source and target prompts. One quite flexible inversion-free editing method constructs a direct ODE path from the source to the target distribution without traversing the Gaussian noise space. The path is governed by:

$$dZ_t^{\text{FE}} = \mathbb{E}\left[V^{\text{tar}}(\hat{Z}_t^{\text{tar}}, t) - V^{\text{src}}(\hat{Z}_t^{\text{src}}, t)\right] dt \tag{3}$$

The noisy samples $\hat{Z}_t^{\text{src}}$ and $\hat{Z}_t^{\text{tar}}$ are respectively constructed as:

$$\hat{Z}_t^{src} = (1-t) \cdot Z_0^{src} + t \cdot N_t, \quad N_t \sim \mathcal{N}(0, \mathbf{I}) \tag{4}$$

$$\hat{Z}_t^{tar} = Z_t^{FE} + \hat{Z}_t^{src} - Z_0^{src} \tag{5}$$

This formulation enables inversion-free editing: the ODE is solved directly from $t = 1$ to $t = 0$, starting from the source image and arriving at the target image, without requiring the expensive inversion step. The source path and the target path share the same noise sample. The noise in the speed difference V_{tar} and V_{src} is completely cancelled out, leaving only the semantic editing direction. The parallelogram displacement $Z_t^{FE} - Z_{t-1}^{FE}$ preserves the path information from the "previous time" to the "current time". This method is well suited for tasks that require preserving the overall structure and layout of the source image, while only modifying colors, textures, or styles.

2.2 DYNAMIC MANIFOLD ANCHORING

The above inversion-free method performs with high efficacy in global editing. The noisy sample $\hat{Z}_t^{src}$ in equation 4 is anchored to the fixed source latent Z_0^{src}. This assumes that the entire image content is trustworthy and should be preserved throughout the process. Therefore, it is necessary to use "shared noise" to ensure that the editing direction is pure and does not introduce unnecessary structure distortion. For global editing, this is a reasonable inductive bias.

However, for localized editing such as image inpainting, rigidly anchoring to Z_0^{src} forces the generative trajectory to continuously regress toward a corrupted prior, suppressing the model's ability to synthesize novel, context-aware content. The failure manifests as severe object distortion and a crucial loss of consistency in unedited areas, leading to disintegration of the background and surrounding structures.

To overcome this limitation, we propose Dynamic Manifold Anchoring as shown in Figure 1: instead of anchoring to the fixed source latent, we dynamically update the anchor to the current generation state Z_t^{DA}:

$$\hat{Z}_t^{tar} = (1-t) \cdot Z_t^{DA} + t \cdot N_t, \quad N_t \sim \mathcal{N}(0, \mathbf{I}) \tag{6}$$

This simplificationła direct consequence of replacing Z_0^{src} with Z_t^{DA}łmeans that both velocity fields now operate on different noisy input:

$$V_t^{\Delta} = V^{tar}(\hat{Z}_t^{tar}, t) - V^{src}(\hat{Z}_t^{src}, t) \tag{7}$$

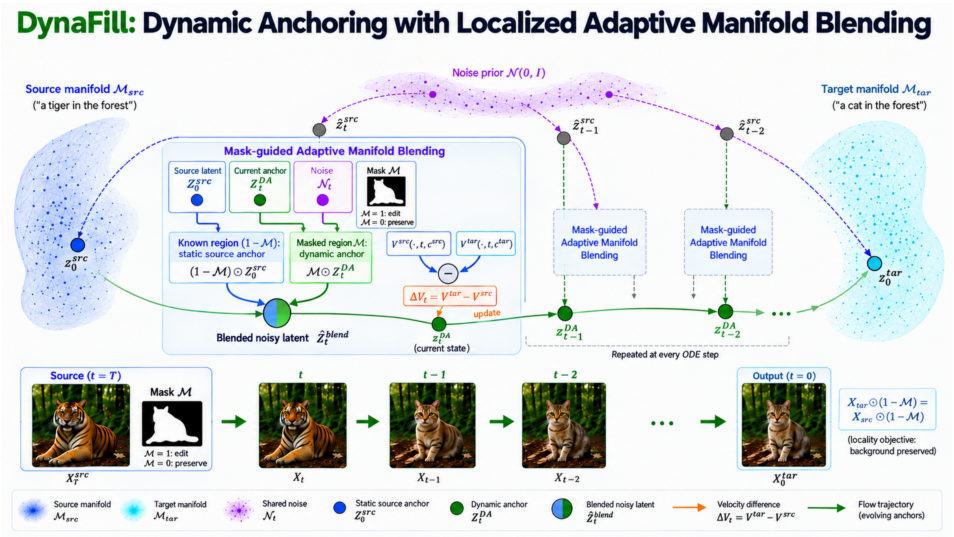

Figure 1: DynaFill dynamically updates the anchor with the current flow state Z_t^{DA} at each generation step. By reconstructing the noisy source point from the evolving anchor and estimating the local editing direction, DynaFill maintains manifold consistency throughout the flow-based editing process.

Manifold-Consistent Interpretation. Why do we term this strategy manifold-consistent? Consider the following geometric interpretation:

Static Anchoring: The noisy samples are constrained to lie on a fixed linear interpolation path between Z_0^{src} and the noise distribution. This path is independent of the generative trajectoryłit remains fixed regardless of how the trajectory evolves.

Dynamic Anchoring (DynaFill): As shown in Figure 1, the anchor point Z_t^{DA} slides along the ODE solution trajectory. At each timestep, the noisy samples are regenerated from this new anchor, ensuring that the distribution is always centered at the current solution. This creates a self-correcting feedback loop: the noisy samples used to compute the velocity difference are always consistent with the current state of the generative process. Mathematically, this corresponds to an online manifold alignment strategy.

Figure 1 provides a conceptual comparison between static and dynamic anchoring. The dynamic anchoring encourages the perturbation direction at each step to remain locally aligned with the geometry of the evolving generative trajectory. Consequently, the ODE solver remains locally aligned with the manifold geometry, avoiding the accumulation of drift errors that arise from static anchoring. The masked area is invalid information, and we do not trust the structure. At this point, we instead hope that the model introduces a certain degree of ”exploratory” in the masked region. The ’noise residue’ from shared noise is no longer purely interference here, but acts as a random perturbation probe in the masked area, helping the model jump out of local optima and generate more realistic and diverse content.

2.3 Enhanced ODE Integration

In standard Rectified Flow, the ODE is discretized using the explicit Euler solver:

$$Z_{t_{i-1}} = Z_{t_i} + \Delta t_i \cdot V_{t_i}^{\Delta}, \quad \Delta t_i = t_{i-1} - t_i < 0 \tag{8}$$

While computationally simple, this first-order method has $O(h^2)$ local truncation error and requires a large number of small steps to achieve satisfactory generation quality.

Predictor-Corrector Sampling. To improve the accuracy-efficiency trade-off, we replace the explicit Euler method with the predictor-corrector Euler method, which achieves second-order accuracy ($O(h^3)$ local truncation error) with only one additional velocity evaluation per step.

Step 1: Predictor. Given the current state Z_{t_i} and the velocity difference $V^{\Delta}_{t_i}$, we first compute a coarse prediction of the next state using the explicit Euler step as in equation 8.

Step 2: Corrector. We then re-evaluate the velocity at the predicted state $\tilde{Z}_{t_{i-1}}$.

$$\tilde{V}^{\Delta}_{t_{i-1}} = V^{tar}(\tilde{Z}_{tar}, t_{i-1}) - V^{src}(\tilde{Z}_{src}, t_{i-1}) \tag{9}$$

where $\tilde{Z}_{tar}$ is got by $\tilde{Z}_{t_{i-1}}$ with equation 6.

Step 3: Trapezoidal Update. The final update is computed by averaging the velocities at the start and end of the interval.

$$Z_{t_{i-1}} = Z_{t_i} + \frac{\Delta t_i}{2} \cdot \left(V^{\Delta}_{t_i} + \tilde{V}^{\Delta}_{t_{i-1}} \right) \tag{10}$$

This is equivalent to approximating the integral using the trapezoidal rule.

For the same step size h, our predictor-corrector method achieves one order higher convergence accuracy than explicit Euler. In contrast, we can use substantially larger step sizes or fewer total steps while maintaining comparable quality.

2.4 Localized Adaptive Manifold Blending

For image inpainting, given a binary mask $M \in \{0,1\}^{H \times W}$ where $M = 1$ indicates the masked region and $M = 0$ indicates the known region, the generated image X_{tar} must satisfy:

$$X^{tar} \odot (1 - M) = X^{src} \odot (1 - M) \tag{11}$$

This indicates that unmasked pixels must remain strictly unchanged. Without explicit locality control, the ODE updates would inevitably modify the entire latent representation, violating this requirement.

Limitation of Hard Replacement. Existing diffusion-based inpainting methods (Avrahami et al., 2023) enforce locality via hard replacement:

$$\hat{Z}_{t-1} = Z^{\text{pred}}_{t-1} \odot M + Z^{\text{src}}_{0} \odot (1 - M) \tag{12}$$

This forcefully overwrites the unmasked regions with the source latent at each denoising step. While this enforces strict locality, it introduces discontinuities at the mask boundary, as the ODE trajectory is abruptly interrupted and reset, causing visible seams and color shifts.

Our Solution: Localized Adaptive Manifold Blending. We propose an Adaptive Manifold Blending mechanism that operates at the noise construction stage rather than the prediction stage. This allows the ODE trajectory to evolve smoothly across the mask boundary while still ensuring locality. The key idea is to construct the noisy sample using a mask-guided mixture of two anchoring strategies:

$$\hat{Z}^{\text{blend}}_{t} = \underbrace{((1-t) \cdot Z^{\text{src}}_{0} + t \cdot N_t)}_{\text{Unmasked: static anchor}} \odot (1 - M) + \underbrace{((1-t) \cdot Z^{DA}_{t} + t \cdot N_t)}_{\text{Masked: dynamic anchor}} \odot M \tag{13}$$

This formulation ensures: The noisy sample of the unmasked regions is constructed using the fixed source latent Z^{src}_0, guaranteeing that the unmasked content remains tied to the original image throughout the ODE integration. The noisy sample of the Masked regions follows the dynamic anchoring strategy section 2.2, adapting to Z^{DA}_t and enabling diverse, context-aware synthesis. Boundary continuity: The Adaptive Manifold Blending allows the ODE solver to propagate gradient information naturally across the boundary, avoiding the abrupt discontinuities of hard replacement.

3 Experiments and Results

In our implementation, we adopt the official checkpoint of SD3 Medium provided by Hugging Face[1]. We use 10 sampling steps and set the source guidance scale and target guidance scale to 3 and 12, respectively. We construct a unified evaluation set of 1,000 image–mask pairs from PIE-Bench (Ju

[1] https://huggingface.co/stabilityai/stable-diffusion-3-medium-diffusers



Figure 2: Qualitative comparison with representative state-of-the-art methods. Across diverse inpainting scenarios, DynaFill achieves more accurate semantics and coherent structures.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-Score ↑	FID ↓
Blended Diffusion	15.8385	0.5809	0.3445	29.29	77.7636
SD3 Inpainting	20.7444	0.7303	0.2085	**29.61**	69.2681
BrushNet	15.6506	0.6988	0.3142	29.30	72.6002
PowerPaint	15.4678	0.5859	0.3564	**29.61**	77.2671
Ours	**23.9203**	**0.8268**	**0.1275**	29.48	**68.0394**

Table 1: Quantitative comparison on our localized inpainting benchmark.

et al., 2024b), EditBench (Wang et al., 2023), and COCO (Lin et al., 2014). The resulting benchmark covers diverse object categories, spatial layouts, and editing regions, enabling evaluation of both target-content generation and source-region preservation. All images are resized to (512×512) to ensure evaluation consistency. All experiments are conducted on a Linux workstation equipped with an NVIDIA RTX A6000 GPU.

3.1 Evaluation Protocol

Following prior image editing and inpainting studies, we adopt five widely used metrics: PSNR, SSIM, LPIPS, CLIP Score, and FID. We benchmark our method against six representative approaches, including SD v1.5 Inpaint (Rombach et al., 2022), Blended Latent Diffusion (Avrahami et al., 2023), PowerPaint (Zhuang et al., 2024), SD3-Inpainting (Zhang et al., 2024), BrushNet (Ju et al., 2024a). All experiments are conducted using publicly available implementations with their default settings. All metrics are computed directly on the generated outputs without additional optimization or post-processing.

3.2 Comparisons

As shown in Table 1, DynaFill achieves better numerical results. It indicates that DynaFill faithfully executes local modifications without introducing unnecessary global changes to the background, revealing stronger global consistency and Local vector alignment in the direction of change. The



Figure 3: Intermediate results during sampling. The proposed predictor–corrector integration progressively refines the masked region and produces coherent inpainting results within 10 sampling steps.

Method	Step	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-Score ↑	FID ↓	Time ↓
Euler ODE Solver	40	20.1054	0.7922	0.1837	**30.35**	68.9912	6.8s
Enhanced ODE Solver	10	**23.9203**	**0.8268**	**0.1275**	29.48	**68.0394**	**3.1s**

Table 2: Quantitative comparison between Euler ODE Solver and Enhanced ODE Solver.

lower FID shows that DynaFill prioritizes visual authenticity and achieves more remarkable quality.

Figure 2 presents visual comparisons between our DynaFill and baselines. The test cases cover diverse local editing tasks. As shown in the Figure, Blended Diffusion produces results with incoherent structure and semantic misalignment. SD3 Inpaint and FlowChef often fail to generate the target content entirelyłmissing the hat and the bird in their respective tasks, indicating a bias toward source preservation over prompt adherence. PowerPaint and BrushNet achieve better text alignment but exhibit artifacts such as unnatural colors, misaligned structures, distorted proportions, and tiling artifacts. In contrast, DynaFill consistently generates semantically accurate content with plausible structures.

3.3 Ablation Studies

Effect of Enhanced ODE Integration. Table 2 evaluates the effectiveness of the enhanced ODE solver. Although the enhanced Euler solver introduces additional computation at each step through predictor–corrector integration, its higher-order approximation enables accurate sampling with substantially fewer steps and time. As shown in Table 2, enhanced Euler reduces the number of sampling steps and time from 40 to 10 and 6.8s to 3.1s while improving most quality metrics. This improvement can be attributed to the second-order approximation of the enhanced Euler solver, which provides higher numerical accuracy than the first-order Euler. Meanwhile, the inference efficiency is boosted to approximately $2.2\times$ acceleration. These results demonstrate that the proposed enhanced ODE integration effectively accelerates the sampling process while preserving semantic editing ability and visual quality, making our inversion-free inpainting more practical.

Effect of Dynamic Manifold Anchoring. Qualitatively, Figure 4 illustrates the influence of the two anchoring strategies. The fixed source anchor tends to preserve excessive information from the original latent, which limits the semantic transition in the masked region. For example, in the pizza-to-cake and sofa-to-chair editing tasks, the generated results retain residual structures from the

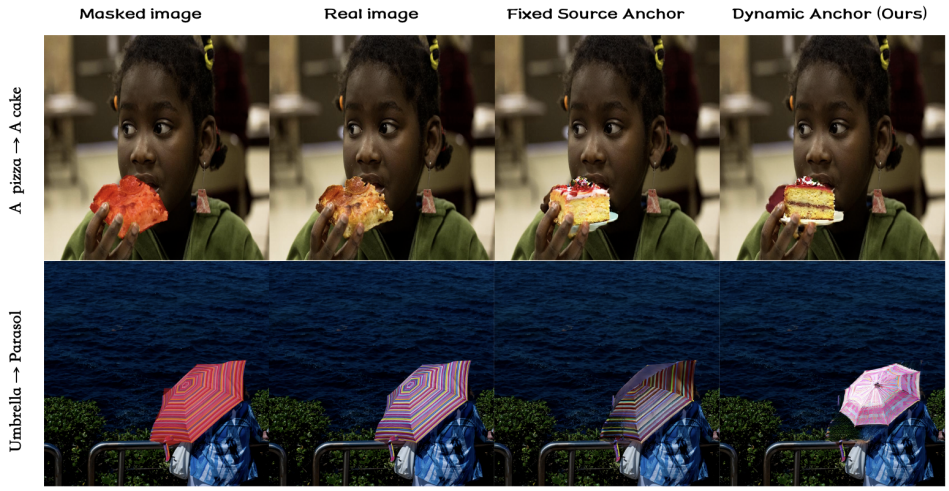

Figure 4: Qualitative comparison between fixed and dynamic anchoring strategies. The fixed source anchor preserves excessive information from the original content, causing residual structures from the source object in the edited regions. By updating the anchor along the generation trajectory, DynaFill enables more effective target-oriented synthesis and produces visually coherent results with better semantic alignment.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-Score ↑	FID ↓
Fixed Anchor	21.6995	0.8145	0.1474	**29.77**	68.1321
Dynamic Anchor	**23.9203**	**0.8268**	**0.1275**	29.48	**68.0394**

Table 3: Quantitative comparison between fixed and dynamic anchoring strategies.

source objects, leading to incomplete target synthesis. These qualitative observations are consistent with the quantitative results in Table 3. Table 3 show that the dynamic anchor variant improves most evaluation metrics, achieving higher PSNR (23.9203 vs. 21.6995), better SSIM (0.8268 vs. 0.8145), lower LPIPS (0.1275 vs. 0.1474), and lower FID (68.0394 vs. 68.1321).

These improvements indicate that dynamically updating the anchor enhances structural preservation and perceptual fidelity while maintaining comparable semantic alignment. In contrast, the fixed source anchor constrains the noisy latent around the original source latent throughout sampling, biasing the generation process toward the corrupted source prior, particularly in masked regions. By using the current flow state as the anchor, DynaFill allows the trajectory to adapt to the evolving generative state, resulting in more effective editing.

Effect of Adaptive Manifold Blending. We further investigate the influence of Adaptive Manifold Blending (AMB). Table 4 compares the proposed soft latent blending strategy with the conventional hard replacement strategy. The dynamic blending variant consistently improves all evaluation metrics, achieving higher PSNR (23.9203 vs. 22.3008), better SSIM (0.8268 vs. 0.8184), lower LPIPS (0.1275 vs. 0.1498), higher CLIP Score (29.48 vs. 29.30), and lower FID (68.0394 vs. 70.4033).

Figure 5 demonstrates that AMB provides more accurate semantic transitions and coherent structure during localized editing. As shown in the figure, for the wolf cow example, hard replacement produces a noticeable error, with unnatural color transition and unrealistic structures. In contrast, AMB generates a seamlessly integrated cake with fine-grained texture details, natural color harmonization, and consistent structure, resulting in a visually convincing completion.

4 Conclusion

This paper presents DynaFill, a novel inversion-free framework for localized image inpainting based on rectified flow models. Unlike inversion-based methods that require costly ODE inversion and may

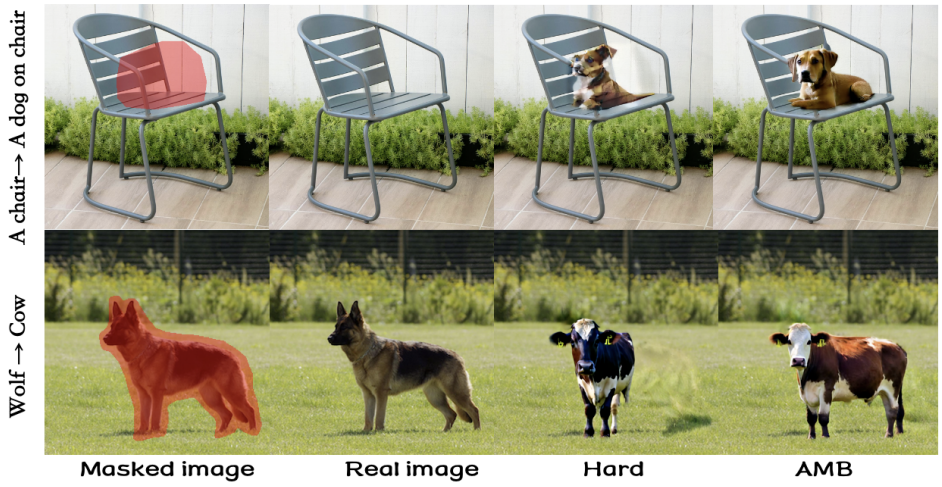

Figure 5: Comparison of hard replacement and AMB. Hard replacement directly overwrites the known regions and may introduce boundary discontinuities, whereas AMB performs latent fusion at the noise construction stage, enabling smoother transitions and better preservation of known content.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	CLIP-Score ↑	FID ↓
Hard	22.3008	0.8184	0.1498	29.30	70.4033
AMB	**23.9203**	**0.8268**	**0.1275**	**29.48**	**68.0394**

Table 4: Quantitative comparison between hard blending and AMB.

accumulate errors, and inversion-free approaches designed for global editing, DynaFill decouples the generative dynamics of masked and unmasked regions. It introduces three key components: (1) a predictor–corrector ODE integrator that reduces sampling steps and time from 40 to 10 and 6.8s to 3.1s respectively, which achieves $2.2\times$ boostment; (2) dynamic manifold anchoring that replaces the static source reference with a dynamically evolving anchor, which enabls full generative freedom in masked holes and coherent editing; and (3) localized adaptive manifold blending, which fuses static and dynamic anchors at the noise construction stage to preserve unmasked pixels and reduce boundary discontinuities. Extensive experiments demonstrate the effectiveness of DynaFill for high-quality localized inpainting.

AI USE STATEMENT

We did not use generative AI tools in the research, execution, writing, or preparation of this work.

REPRODUCIBILITY STATEMENT

Implementation details, experimental settings, evaluation protocols, and additional results needed to reproduce the main findings are provided in the main paper and appendix.

REFERENCES

Omri Avrahami, Ohad Fried, and Dani Lischinski. Blended latent diffusion. *ACM Transactions on Graphics (TOG)*, 42(4):1–11, 2023.

Yingying Deng, Xiangyu He, Fan Tang, Weiming Dong, and Xucheng Yin. Inversion-free style transfer with dual rectified flows. *arXiv preprint arXiv:2511.20986*, 2025.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning*, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Inbar Huberman-Spiegelglas, Vladimir Kulikov, and Tomer Michaeli. An edit friendly DDPM noise space: Inversion and manipulations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12469–12478, 2024.

Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (TOG)*, 36(4):1–14, 2017.

Xuan Ju, Xian Liu, Xintao Wang, Yuxuan Bian, Ying Shan, and Qiang Xu. BrushNet: A plug-and-play image inpainting model with decomposed dual-branch diffusion. In *European Conference on Computer Vision*, pp. 150–168. Springer, 2024a.

Xuan Ju, Ailing Zeng, Yuxuan Bian, Shaoteng Liu, and Qiang Xu. Pnp inversion: Boosting diffusion-based editing with 3 lines of code. In *International Conference on Learning Representations*, 2024b.

Vladimir Kulikov, Matan Kleiner, Inbar Huberman-Spiegelglas, and Tomer Michaeli. FlowEdit: Inversion-free text-based editing using pre-trained flow models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19721–19730, 2025.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll
'ar, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pp. 740–755. Springer, 2014.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 85–100, 2018.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

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*, pp. 11461–11471, 2022.

Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6038–6047, 2023.

Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.

Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2536–2544, 2016.

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

Chitwan Saharia, William Chan, Huiwen Chang, Chris A Lee, Jonathan Ho, Tim Salimans, David J Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. *arXiv preprint arXiv:2111.05826*, 2021.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Bram Wallace, Akash Gokul, and Nikhil Naik. Edict: Exact diffusion inversion via coupled transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22532–22541, 2023.

Su Wang, Chitwan Saharia, Ceslee Montgomery, Jordi Pont-Tuset, Shai Noy, Stefano Pellegrini, Yasumasa Onoe, Sarah Laszlo, David J. Fleet, Radu Soricut, Jason Baldridge, Mohammad Norouzi, Peter Anderson, and William Chan. Imagen editor and editbench: Advancing and evaluating text-guided image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18359–18369, 2023.

Sihan Xu, Yidong Huang, Jiayi Pan, Ziqiao Ma, and Joyce Chai. Inversion-free image editing with natural language. *arXiv preprint arXiv:2312.04965*, 2023.

Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5505–5514, 2018.

Xulu Zhang, Xiao-Yong Wei, Jinlin Wu, Tianyi Zhang, Zhaoxiang Zhang, Zhen Lei, and Qing Li. Compositional inversion for stable diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 7350–7358, 2024.

Junhao Zhuang, Yanhong Zeng, Wenran Liu, Chun Yuan, and Kai Chen. A task is worth one word: Learning with task prompts for high-quality versatile image inpainting. In *European Conference on Computer Vision*, pp. 195–211. Springer, 2024.