CVPR JUNE 3-7, 2026 DENVER COLORADO

Low-Resolution Editing is All You Need for High-Resolution Editing

Junsung Lee[1*], Hyunsoo Lee[1*], Yong Jae Lee[2], Bohyung Han[1]

[1]Seoul National University

[2]University of Wisconsin-Madison

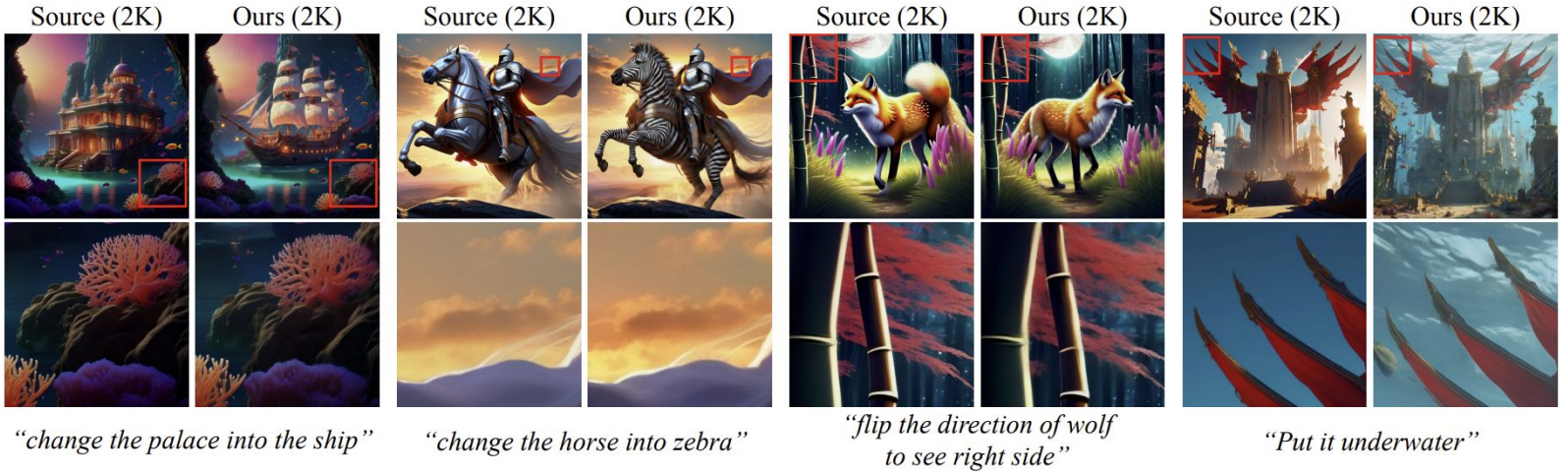
Introduction

- Importance & usage of high-resolution images are rapidly increasing in various domains

- Users seek precise and intentional modifications to an existing high-resolution image
  - Generating a new one from scratch losses the controllability
  - In contrast, image manipulation allows controlled adjustments that reflect the user’s intention
  - Editing serves as a reliable strategy to produce user-controlled high-resolution contents

Introduction

- Motivation: ***training-free high-resolution image editing***



  - Training a new model requires significant amount of computation
  - Existing I2I methods operate at low resolutions and cannot directly handle large-scale inputs
  - One possible method is to use Real-ISR baselines
    - But, it **fails** to recover micro-scale textures since the details of the source image are not conditioned

Problem Formulation

- We require an optimization that preserves both semantics and fine-grained details
- 1) **Preserving semantics**
  - Can be facilitated by leveraging existing low-resolution editing methods
  - Then, what is the most challenging part?
- 2) **Transfer the fine-scale details** of the high-resolution image into the edited one

- Input: 3 conditioning images
  - $I^{\text{high}}_{\text{src}}$: high-res source image
  - $I^{\text{low}}_{\text{src}}$: downsampled version of $I^{\text{high}}_{\text{src}}$
  - $I^{\text{low}}_{\text{ref}}$: low-res edited result of $I^{\text{low}}_{\text{src}}$
- Output: high-res edited image $I^{\text{high}}_{\text{ref}}$
  - Reflecting the overall semantics of $I^{\text{low}}_{\text{src}}$
  - Maintaining the fine-grained details of $I^{\text{high}}_{\text{src}}$

Related Works

- Low-resolution image editing methods
  - They can edit images only up to 1K resolution
  - Our approach leverages the generative priors of existing methods to enable editing even at 2K and higher resolutions

- Diffusion-based Real-ISR methods
  - They are not conditioned on the source image, losing the details of source image
  - Typically rely on training, our method leverages through test-time optimization without training

[Zhang25] Zhang et al. "In-context edit: Enabling instructional image editing with in-context generation in large scale diffusion transformer", NeurIPS, 2025.
[Zhu25] Zhu et al. "Kv-edit: Training-free image editing for precise background preservation", ICCV, 2025.
[Dong25] Dong et al. "Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution", CVPR, 2025.
[Sun25] Sun et al. "Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach", CVPR, 2025.
[Cheng25] Cheng et al. "Effective diffusion transformer architecture for image super-resolution", AAAI, 2025.
[Duan25] Duan et al. "Dit4sr: Taming diffusion transformer for real-world image super-resolution", ICCV, 2025.

Overview: ScaleEdit

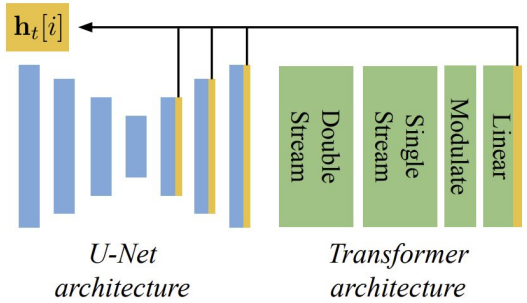
- Transfer fine-grained details with a **learnable transfer function**
  - Defined in the intermediate feature space of pretrained DM
- Patch-wise detail transfer strategy to fully inherit their strong generative priors
  - Divide the source image into model-native resolution regions
    - $I_{\text{src}}^{\text{high}} \rightarrow$ Generate *N x M* patches
    - Helps preserving micro-scale details while still benefiting from the priors of pretrained model
  - Then produce *N x M* **detail-enhanced** patches
    - By manipulating the intermediate features of pretrained generative model
- Patch **synchronization** strategy to ensure global consistency across patches
  - Combine *N x M* patches with synchronization $\rightarrow I_{\text{ref}}^{\text{high}}$

Method

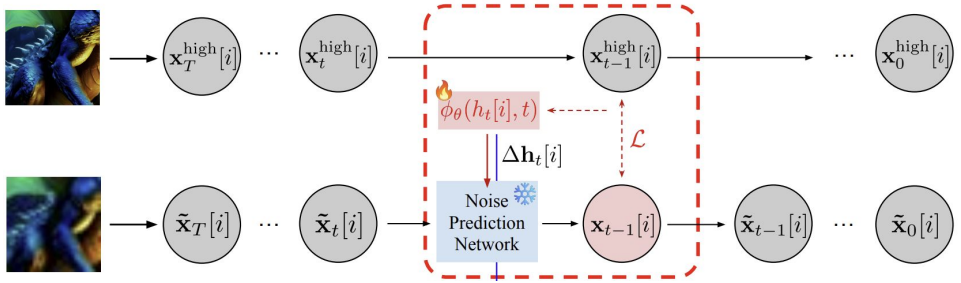
- Patch-wise diffusion trajectory extraction
  - Extract patch-wise diffusion trajectories for each of three input images

$$\{\mathbf{x}_t^{\text{high}}[i]\}_{t=0}^{T}, \{\mathbf{x}_t^{\text{low}}[i]\}_{t=0}^{T}, \{\mathbf{y}_t^{\text{low}}[i]\}_{t=0}^{T} \quad \text{with} \quad 1 \le i \le NM$$

- Transfer function $\phi(i,t)$
  - Goal: guide $\{\tilde{\mathbf{x}}_t[i]\}_{t=0}^{T}$ to follow $\{\mathbf{x}_t^{\text{high}}[i]\}_{t=0}^{T}$ (initial condition: $\tilde{\mathbf{x}}_T[i] = \mathbf{x}_T^{\text{low}}[i]$)
  - Idea: **adjust the intermediate feature** $\mathbf{h}_t[i]$ of the pretrained generative model
    - $\Delta\mathbf{h}_t[i] = \phi(i,t)$



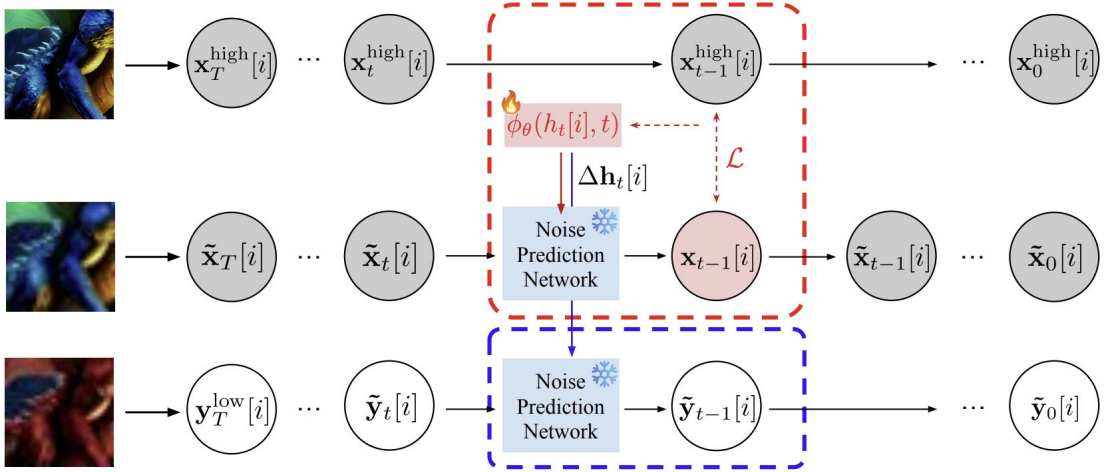
Method



- Optimization of $\Delta\mathbf{h}_t[i]$
  - $\mathcal{L} := ||\mathbf{x}_{t-1}^{\text{high}}[i] - f^{\text{rev}}(\tilde{\mathbf{x}}_t[i], t\,; \Delta\mathbf{h}_t[i])||_2^2$
    with $f^{\text{rev}}(\mathbf{x}_t^{\text{rev}}, t\,; \Delta\mathbf{h}_t[i]) := \sqrt{\alpha_{t-1}}\hat{\mathbf{x}}_0(\mathbf{x}_t^{\text{rev}}, t\,; \Delta\mathbf{h}_t[i]) + \sqrt{1-\alpha_{t-1}}\epsilon_\theta(\mathbf{x}_t^{\text{rev}}, t\,; \Delta\mathbf{h}_t[i])$

- Parameterization of $\Delta\mathbf{h}_t[i]$
  - Naive approach: $\phi(i,t) = \mathbf{c}_t[i]$ (const)
    - **FAILS** since it's not conditioned to source image; limited to model varying spatial structures
  - Solution: parameterize adaptive to $\mathbf{h}_t[i]$
    - $\Delta\mathbf{h}_t[i] := \phi_\theta(\mathbf{h}_t[i], t)$
    - Empirical choice – use $\text{conv}_{1\times1}(\mathbf{h}_t[i])$ & set $\Delta\mathbf{h}_t[i] = \mathbf{0}$, $^\forall t \geq \tau$

Method

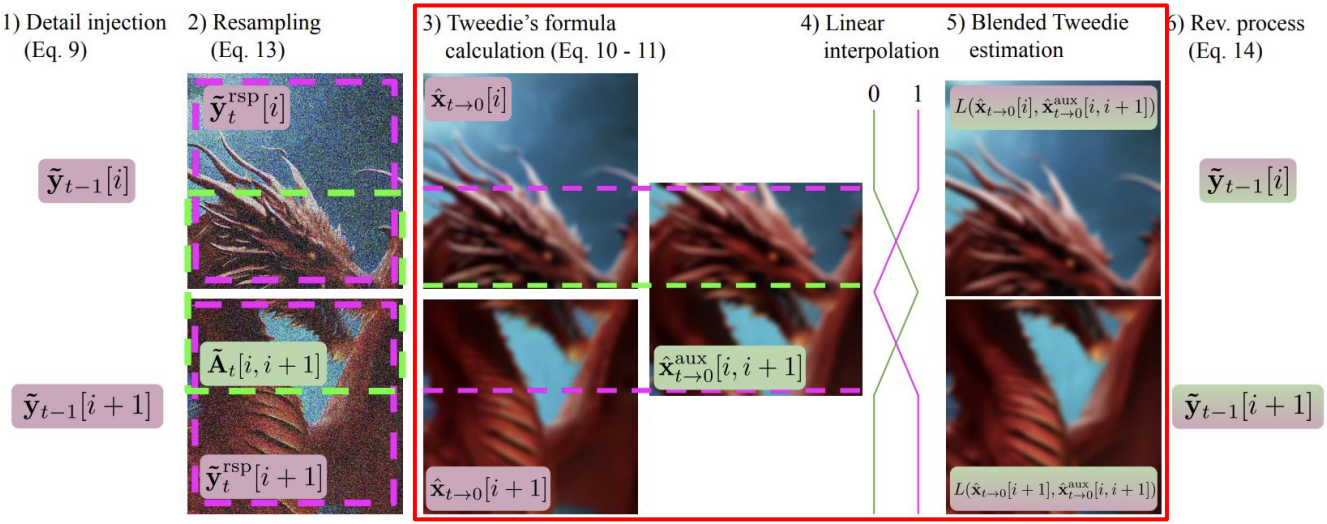


- Injection of $\Delta\mathbf{h}_t[i]$ during high-res image sampling

$$\tilde{\mathbf{y}}_{t-1}[i] = f^{\text{rev}}(\tilde{\mathbf{y}}_t[i], t\,; \Delta\mathbf{h}_t[i]) \quad \text{with} \quad \tilde{\mathbf{y}}_T[i] = \mathbf{y}_T^{\text{low}}[i]$$

Method

- Synchronization mechanism b/w patches for global coherence
  - Blended-Tweedie-based updates

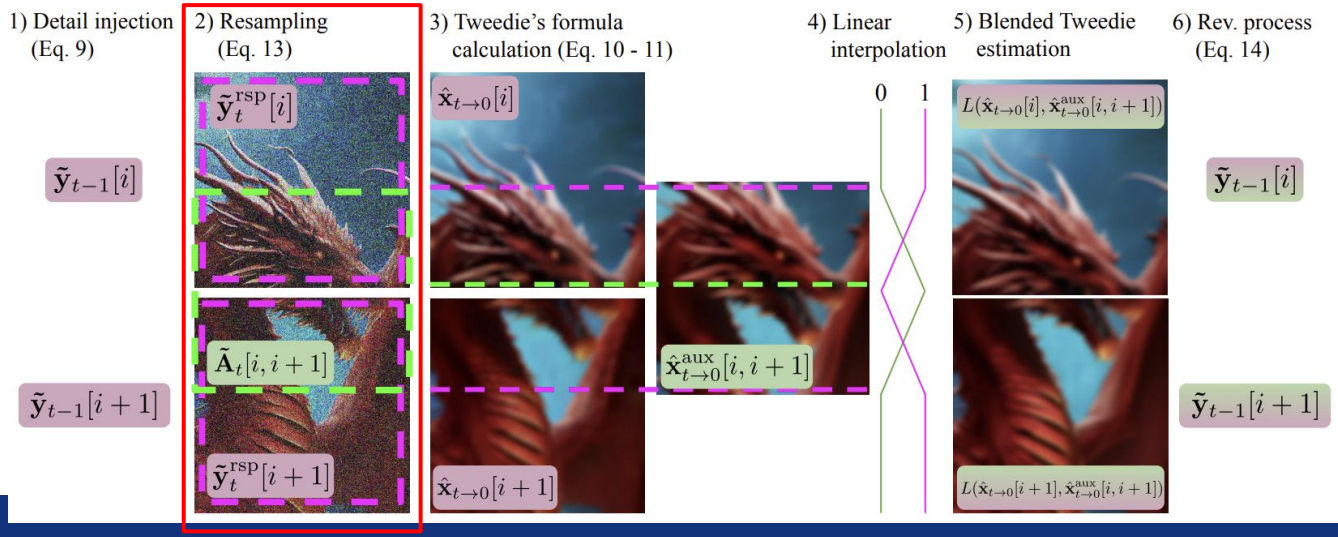


$$\hat{\mathbf{x}}^{\text{aux}}_{t\to0}[i, i+1] = \hat{\mathbf{x}}_0(\tilde{\mathbf{A}}_t[i, i+1], t)$$

$$\hat{\mathbf{x}}_{t\to0}[i] = \hat{\mathbf{x}}_0(\tilde{\mathbf{y}}_t[i], t)$$

Method

- Synchronization mechanism b/w patches for global coherence
  - Pitfalls: undefined $\Delta \mathbf{h}_t$ for $\tilde{\mathbf{A}}_t[i, i+1]$
    - Not trivial to do synchronization using $\tilde{\mathbf{y}}_t[i]$ and $\tilde{\mathbf{y}}_{t-1}[i+1]$
    - We need to **disentangle** the detail injection from synchronization
  - Solution: **resampling**



Method

- Resampling-based update
  - Given: detail-injected latent $\tilde{\mathbf{y}}_{t-1}[i]$
  - Apply DDIM forward step without detail injection

$$\tilde{\mathbf{y}}_t^{\mathrm{rsp}}[i] = f^{\mathrm{fwd}}(\tilde{\mathbf{y}}_{t-1}[i], t-1)$$
$$\tilde{\mathbf{y}}_t^{\mathrm{rsp}}[i+1] = f^{\mathrm{fwd}}(\tilde{\mathbf{y}}_{t-1}[i+1], t-1)$$

  - Calculated blended-Tweedie estimate using $\tilde{\mathbf{y}}_t^{\mathrm{rsp}}[i]$ and $\tilde{\mathbf{y}}_t^{\mathrm{rsp}}[i+1]$

- Update equation (blended-Tweedie + resample)

$$\tilde{\mathbf{y}}_{t-1}[i] = \sqrt{\alpha_{t-1}} L(\hat{\mathbf{x}}_{t\to 0}[i], \hat{\mathbf{x}}_{t\to 0}^{\mathrm{aux}}[i]) + \sqrt{1-\alpha_{t-1}}\epsilon_\theta(\tilde{\mathbf{y}}_t^{\mathrm{rsp}}[i], t)$$

Experiments

- Low-resolution editing method: Nano Banana
- Enhanced diffusion trajectory sampling using Null-text inversion
- Editing scenarios for evaluation
  - 1K-image editing / 2K-image editing
  - Object replacement / style transfer / background modification
- Assets
  - Generated 4K images with FreeScale, then downsample
- Comparison with Diffusion-based SR methods
  - DiT-SR, DiT4SR, PiSA-SR, and TSD-SR

[Comanici25] Comanici et al. “Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities”, 2025.
[Mokady23] Mokady et al. “Null-text inversion for editing real images using guided diffusion models”, CVPR, 2023.
[Qiu25] Qiu et al. “Freescale: Unleashing the resolution of diffusion models via tuning-free scale fusion”, ICCV, 2025.

Quantitative Evaluation

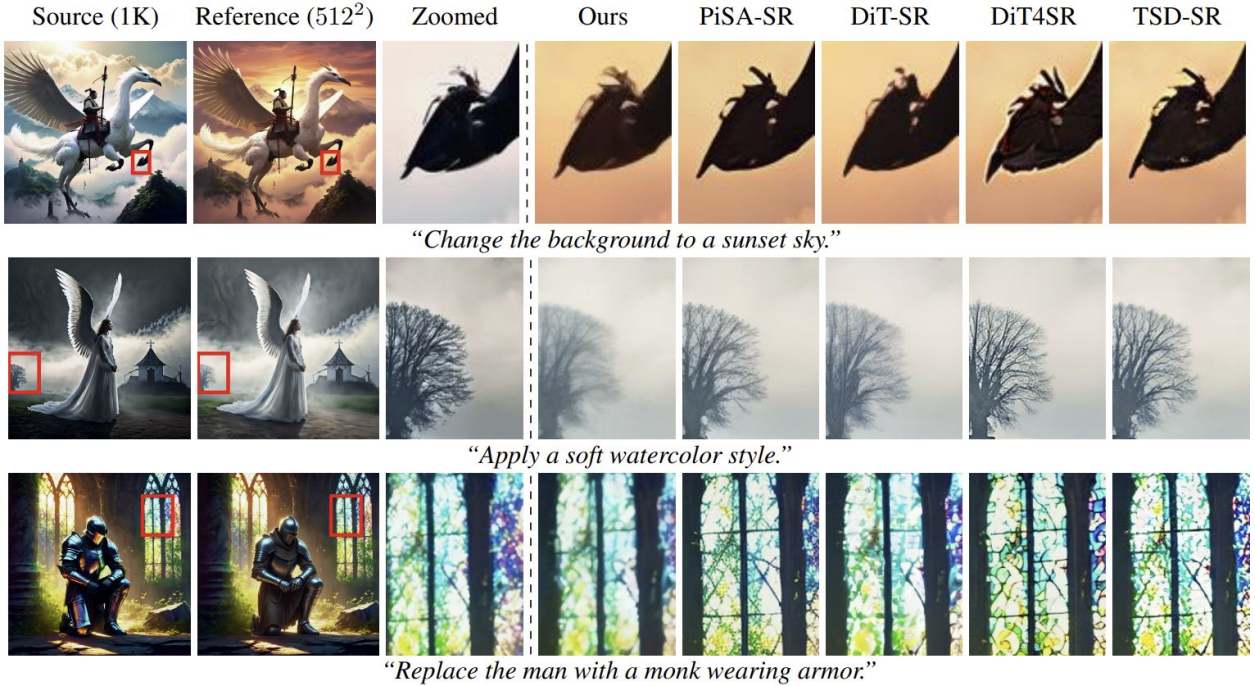
Table 1. Quantitative evaluation under 1K- and 2K-editing scenarios using the pretrained Stable Diffusion [41]. We compare the proposed method with diffusion-based super-resolution methods [7, 10, 12, 48]. Our method shows superior performance compared to the baselines.

Method	1K-editing					2K-editing				
	HaarPSI ↑	M-MSE ↓	M-SSIM ↑	M-PSNR ↑	LPIPS ↓	HaarPSI ↑	M-MSE ↓	M-SSIM ↑	M-PSNR ↑	LPIPS ↓
DiT-SR [7]	0.335	0.058	0.695	21.528	0.477	0.316	0.057	0.754	21.380	0.507
DiT4SR [12]	0.324	0.060	0.625	20.740	0.509	0.305	0.058	0.684	20.701	0.534
PiSA-SR [48]	0.328	0.058	0.668	21.273	0.465	0.312	0.056	0.755	21.320	**0.472**
TSD-SR [10]	0.329	0.061	0.649	20.766	0.489	0.312	0.059	0.715	20.796	0.514
ScaleEdit (Ours)	**0.342**	**0.054**	**0.739**	**22.132**	**0.460**	**0.331**	**0.053**	**0.806**	**21.955**	0.496

[Dong25] Dong et al. “Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution”, CVPR, 2025.
[Sun25] Sun et al. “Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach”, CVPR, 2025.
[Cheng25] Cheng et al. “Effective diffusion transformer architecture for image super-resolution”, AAAI, 2025.
[Duan25] Duan et al. “Dit4sr: Taming diffusion transformer for real-world image super-resolution”, ICCV, 2025.

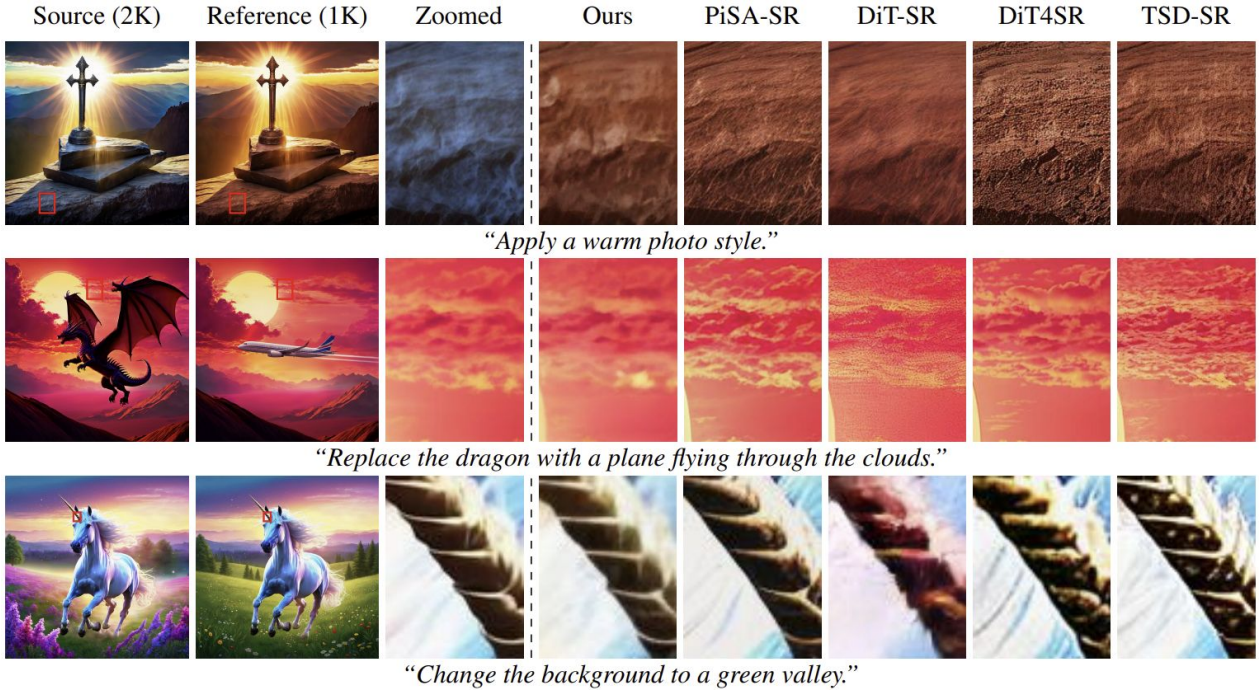
Qualitative Comparison

- 1K-Editing



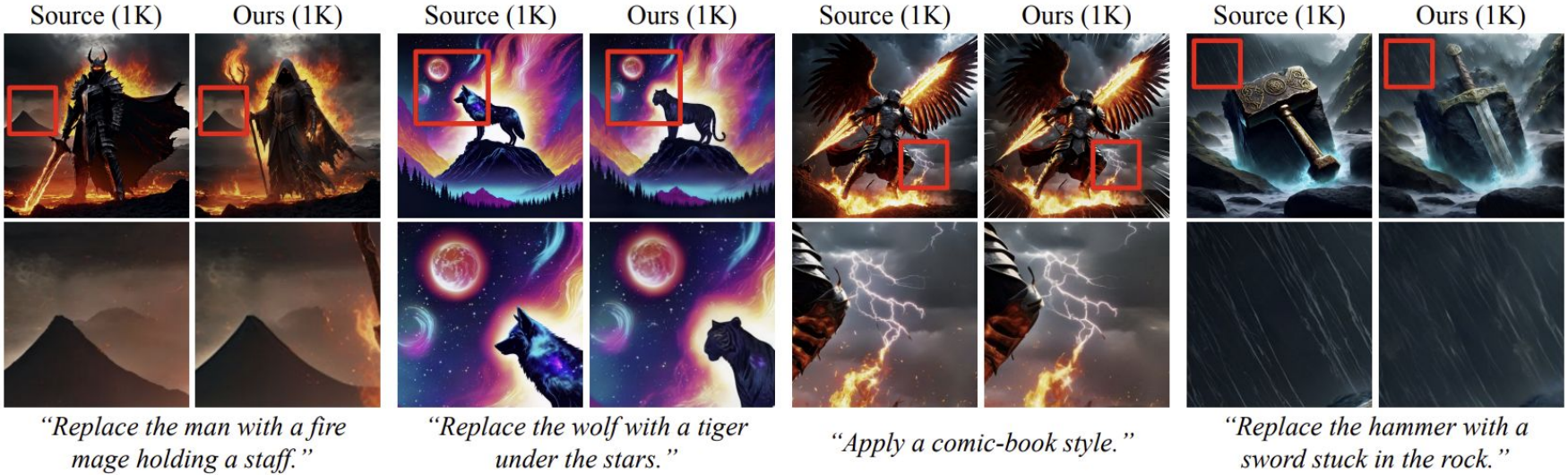
Qualitative Comparison

- 2K-Editing



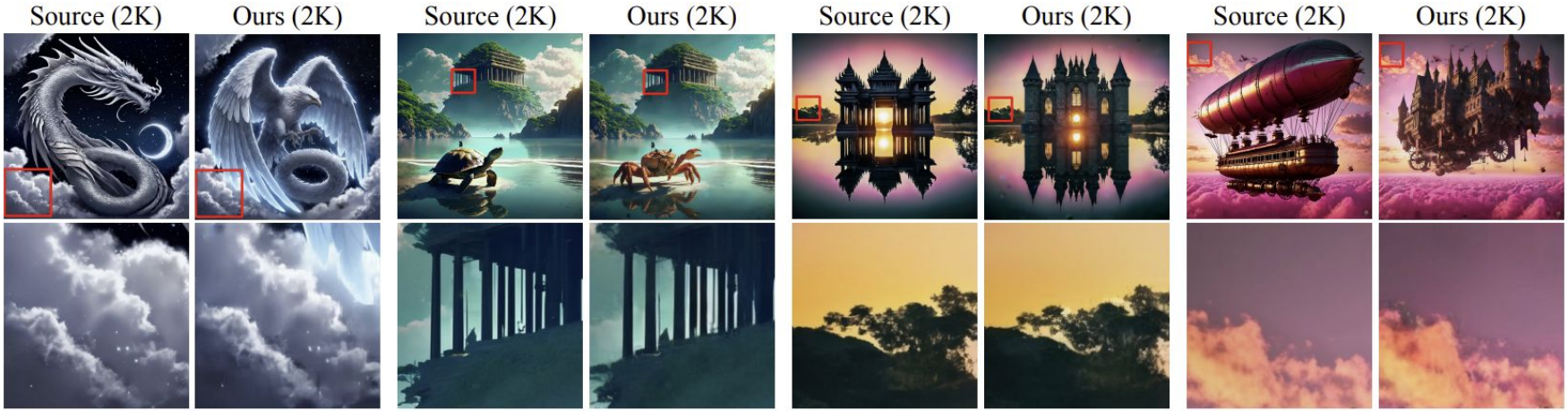
Additional Qualitative Results

- 1K-Editing



Additional Qualitative Results

- 2K-Editing

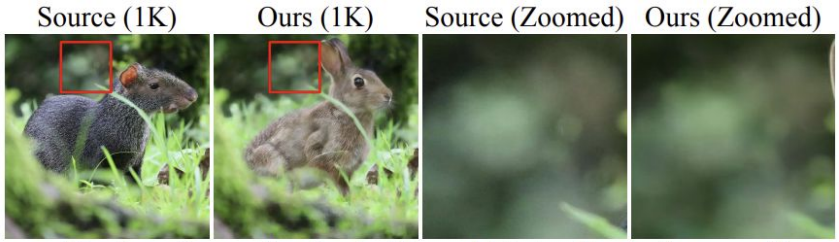


Additional Qualitative Results

- 8K-Editing

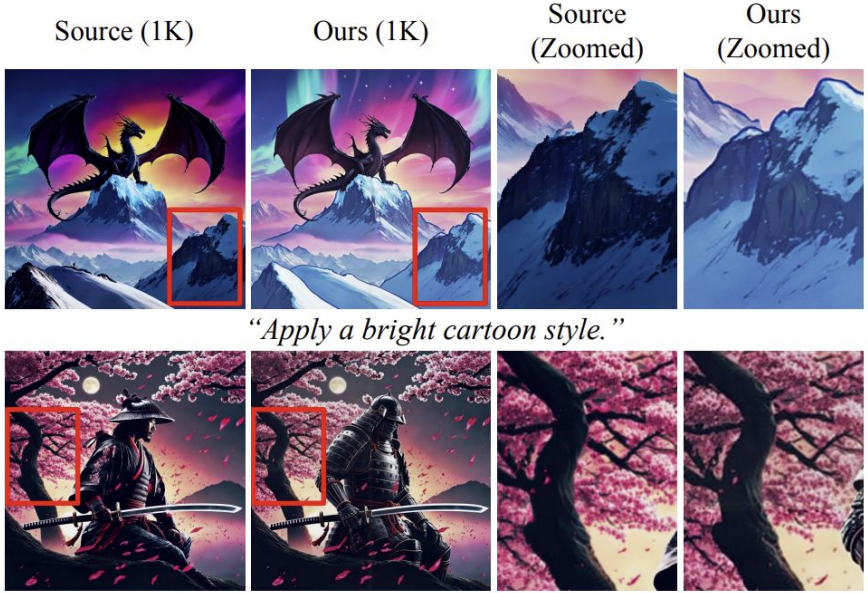
"Replace the man with a knight sitting on the bench."

- 1K Real Image Editing



"Replace the small animal with a rabbit standing in the grass."

- Results using the pretrained FLUX-1.dev



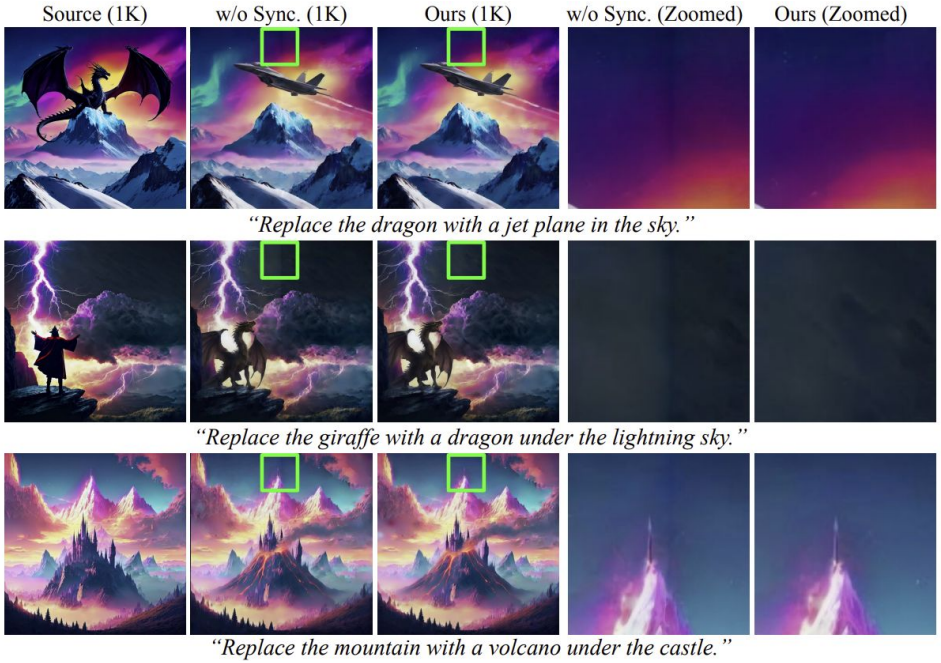
"Apply a bright cartoon style."

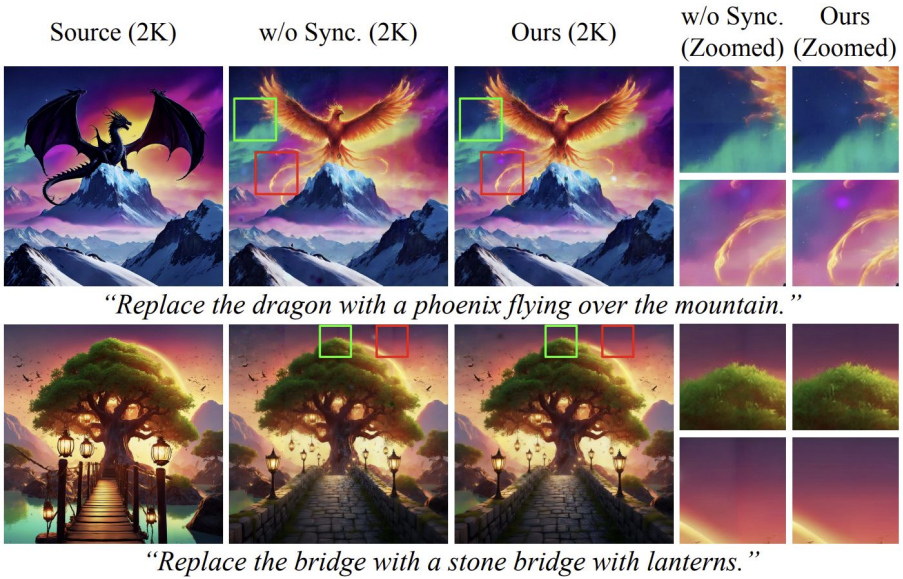
"Replace the samurai with a warrior in black armor."

Black Forest Labs, Flux, 2024.

Ablation Study

- Ablation on synchronization strategy





Conclusion

- Propose a **novel task**: high-resolution image editing
- Enable zero-training editing via patch-wise test-time optimization
- Develop a **feature transfer** function and a patch **synchronization** strategy
- Achieve **strong experimental results** on diverse editing scenarios