BioMicroAgents: A Co-evolutionary Multi-Agent Framework for High-Fidelity Biomicroscopy Imaging

Yongwei Jiang
South China University of Technology
antontohafonti@gmail.com

Peilin Li
Tsinghua University
e1351227@u.nus.edu

Anfu Deng
South China University of Technology
denganfu0702@163.com

Jun Yin
Tsinghua University
yinj24@mails.tsinghua.edu.cn

Xiaoyu Chen*
South China University of Technology
xiaoyuchen@scut.edu.cn

Abstract

Biomicroscopy imaging involves the complex coordination of multiple procedures, such as focusing, illumination, and staining. Traditional unitary methods struggle with dynamical environments, while hybrid models pre-trained on natural images tend to introduce artifacts for higher perceptual metrics. Notably, biological imaging is not a one-off task. It is an iterative process similar to human experts observation, verification and adjustment. They can benefit from history, remembering which processing strategies are suitable and can also identify issues such as artifacts and mis-staining from verification. To this end, we propose BioMicroAgents, the first multi-agent framework with co-evolutionary capabilities designed to simulate and optimize whole biomicro imaging process including brightening, denoising, deblur, super-resolution, and virtual staining. Through reinforcement learning, the model can reflect based on historical experience to reduce artifacts and mis-staining, while enhancing edge and structural features to improve the accuracy of downstream tasks. Specifically, BioMicroAgents includes: (i) Multimodal Analytical Agents, which utilize multimodal RAG to provide template images and biomicro knowledge, and employ the memory mechanism to analyze optimal processing paths; (ii) We propose three biomicro fidelity metrics and utilize three agent groups (MicroNavigator, MicroProcessor, and MicroAuthenticator) to perform rigorous analysis, processing, and verification. (iii) Co-evolutionary ability, which achieves dynamic policy and self-correction through interactive feedback strategies and memory mechanisms, allowing the agent to evolve from history and other agents. We test model across 12 biomicro tasks, BioMicroAgents achieved improvements in perceptual quality while adhering to strict fidelity in sub-cellular features. It also enhances the accuracy of downstream tasks such as segmentation and cell counting, showcasing the potential of general-purpose visual agents in bioscience.

*Corresponding author.

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

1 Introduction

Biological microscopy image processing faces multiple challenges, including low illumination, noise, defocus blur, insufficient resolution, difficulty in cross-staining modality translation, and limited capability for continuously accumulating experience from historical failure cases after deployment. The main limitations of existing approaches are as follows. End-to-end microscopy image restoration models are usually trained for individual tasks, such as denoising, super-resolution, or virtual staining. They can achieve promising performance under fixed data distributions, but struggle to adapt to compound degradations and multi-stage requirements in real experimental workflows. General-purpose image enhancement or generative models, although equipped with strong visual restoration capabilities, often introduce non-biological artifacts due to their perceptual optimization objectives, thereby damaging sub-cellular structures and staining consistency [13]. Existing agentic image processing frameworks [51, 50] are mostly designed for natural image scenarios and lack microscopy-specific optical priors, histological morphology constraints, and historical experience memory, making it difficult to guarantee high-fidelity biomedical image processing. These limitations indicate that biological microscopy image processing should not be regarded as a one-shot image-to-image mapping. Instead, it should be modeled as a coordinated closed-loop process consisting of observation, planning, processing, verification, and experience accumulation [51]. To this end, we propose **BioMicroAgents**, the first self-evolving multi-agent framework for high-fidelity biological microscopy imaging. The framework simulates expert microscopy imaging workflows through three coordinated agent groups: MicroNavigator integrates multimodal RAG and historical memory to perform task understanding and processing-path planning; MicroProcessor dynamically schedules expert models for low-light enhancement, denoising, deblurring, super-resolution, and virtual staining; MicroAuthenticator verifies biological fidelity from cellular morphology, staining consistency, and histological texture authenticity, and triggers rollback when artifacts or mis-staining are detected. The main contributions of this work are summarized as follows:

- **Dynamic multi-agent framework for microscopy workflows:** We formulate biological microscopy image processing as a closed-loop process involving planning, execution, verification, and rollback, supporting dynamic composition of multiple microscopy operations, including low-light enhancement, denoising, deblurring, SR, and virtual staining.
- **Adaptive planning with microscopy knowledge and experience:** We retrieve microscopy templates, staining references, and biomedical knowledge through multimodal RAG, and combine them with memory of successful and failed trajectories, enabling the system to evolve according to sample conditions, user requirements, and historical experience to generate more reliable processing strategies.
- **Biomedical-fidelity-oriented co-evolution mechanism:** We explicitly incorporate biomedical fidelity constraints into the agent RL loop to facilitate co-evolution among agents for the first time.
- **Multi-task experimental validation:** We systematically evaluate the proposed framework on SR, virtual staining, low-light enhancement, denoising, and deblurring.

2 Related Work

The related work spans workflow-aware microscopy imaging, general restoration, and memory-augmented agents. Learning-based methods have advanced restoration, autofocus, virtual staining, super-resolution, modality translation, and diffusion-based reconstruction [45, 31, 42].

2.1 Workflow-Aware Microscopy Imaging

Microscopy spans acquisition control, focus and illumination adjustment, restoration, cross-modality prediction, virtual staining, and biological analysis. Optimizing these stages independently preserves physical interpretability but limits coordination because processing should reflect the degradation, biological target, modality, staining protocol, and downstream task. Learning-based restoration, autofocus, illumination adjustment, **virtual staining and cross-modality prediction**, super-resolution, and enhancement improve acquisition and support segmentation, detection, classification, and counting while reducing costly or destructive procedures. Yet most are fixed task-specific mappings that ignore cross-stage dependencies. Systems must balance contrast and denoising against boundary preservation, staining fidelity, and downstream reliability. Scarce ground truth

limits pixel-wise losses, while natural-image perceptual objectives may hallucinate textures or alter colors, compromising scientific validity. These risks motivate workflow-aware reasoning across stages.

2.2 General Image Restoration

General restoration has progressed from convolutional and adversarial models to Transformer- and diffusion-based frameworks for denoising, deblurring, super-resolution, low-light enhancement, and illumination correction. Newer architectures improve long-range modeling and texture reconstruction, while diffusion models recover fine structures, correct severe corruption, and translate across domains [45, 31, 42]. In microscopy, these techniques address low signal-to-noise ratio, uneven illumination, defocus, photobleaching, limited resolution, and modality-specific artifacts, improving sharpness, contrast, and texture fidelity. Direct transfer nevertheless remains difficult: unlike natural-image restoration, microscopy must preserve biological correctness, quantitative consistency, and modality-specific color or intensity distributions. Visually plausible outputs may distort morphology, fabricate structures, or alter staining, while scarce ground truth weakens reference-based metrics. BioMicroAgents therefore places restoration models within an agentic workflow guided by task objectives, historical experience, and downstream biological constraints.

2.3 Memory Mechanisms for Self-Evolving Agents

Memory lets agents reuse and reflect on prior interactions, improving planning, tool use, trajectory reuse, and failure avoidance beyond static inference [44, 46]. Yet existing frameworks mainly target language, navigation, embodied control, software tools, or general visual reasoning [38, 43, 5]; their textual records rarely capture microscopy-specific semantics. Microscopy memory must account for morphology, staining, texture, artifacts, and downstream analysis, since enhancement may damage boundaries, virtual staining may shift meaningful colors, and denoising may erase diagnostic structures. To the best of our knowledge, such memory has not been systematically studied for workflow-aware biological microscopy. BioMicroAgents fills this gap by recording and retrieving successful and failed trajectories to avoid repeated errors, reuse effective strategies, and adapt across tasks, imaging conditions, and biological objectives.

3 Method

BioMicroAgents comprises three agent groups: MicroNavigator, MicroProcessor, and MicroAuthenticator(Fig 1). Co-evolution is defined as a coupled policy update process where:(1) Processor policy π_p is optimized via rewards from Authenticator. (2) Navigator policy π_n is updated via memory retrieval shaped by Authenticator feedback. First, the **MicroNavigator** group utilizes a multimodal Retrieval-Augmented Generation (RAG), combining user inputs I_{ori}, D to retrieve microscopic priors K_{micro} and staining templates I_{temp}, which collectively constitute the prior context $\mathcal{C} = \{\{I_{ori}, I_{temp}\}; \{K_{micro}, D\}\}$. Leveraging the semantic understanding of a Vision-Language Model (VLM) and perception metrics, the system identifies the required atomic skill set $s \subseteq \mathcal{S}$. Concurrently, the MicroMem extracts successful and failed experiences E^+, E^- from historical trajectories M_{mem} according to $\mathcal{C}$. Ultimately, the MicroNavigator derives the optimal task path $\mathcal{P} = \{a_1, a_2, \ldots, a_n\}$. Next, **MicroProcessor** and **MicroAuthenticator** collaboratively perform the execution and evolution. Based on $\mathcal{P}$ and $\mathcal{C}$, the system dynamically schedules expert agents from $\mathcal{S}$. At each step k, the models perform a rigorous quality audit by the biomedical evaluation vector $\mathbf{H} = [H_{CMI}, H_{CC}, H_{HTA}]^\top$ with segmentation agent. This vector covers Cyto-Morphological Integrity (H_{CMI}), Chromatographic Consistency (H_{CC}), and Histological Texture Authenticity (H_{HTA}), supplemented by perceptual scores Ψ. If $\mathbf{H} < \tau_H$ or $\Psi < \tau_p$, failure trajectory E^- and reason r^- are stored in the MicroMem, triggering the rollback to return to MicroNavigator. Through carefully designed multi-reward, the evaluation experiences E^+, E^- from MicroAuthenticator can be transformed into rewards for MicroProcessor, enabling co-evolution of agents. Simultaneously, the information within E^+, E^- can be provided to MicroNavigator, enabling it to better organize $\mathcal{P}$. Compared with decoupled reward models, this approach better aligns with the interaction and feedback logic of human experts to maintain biological fidelity.

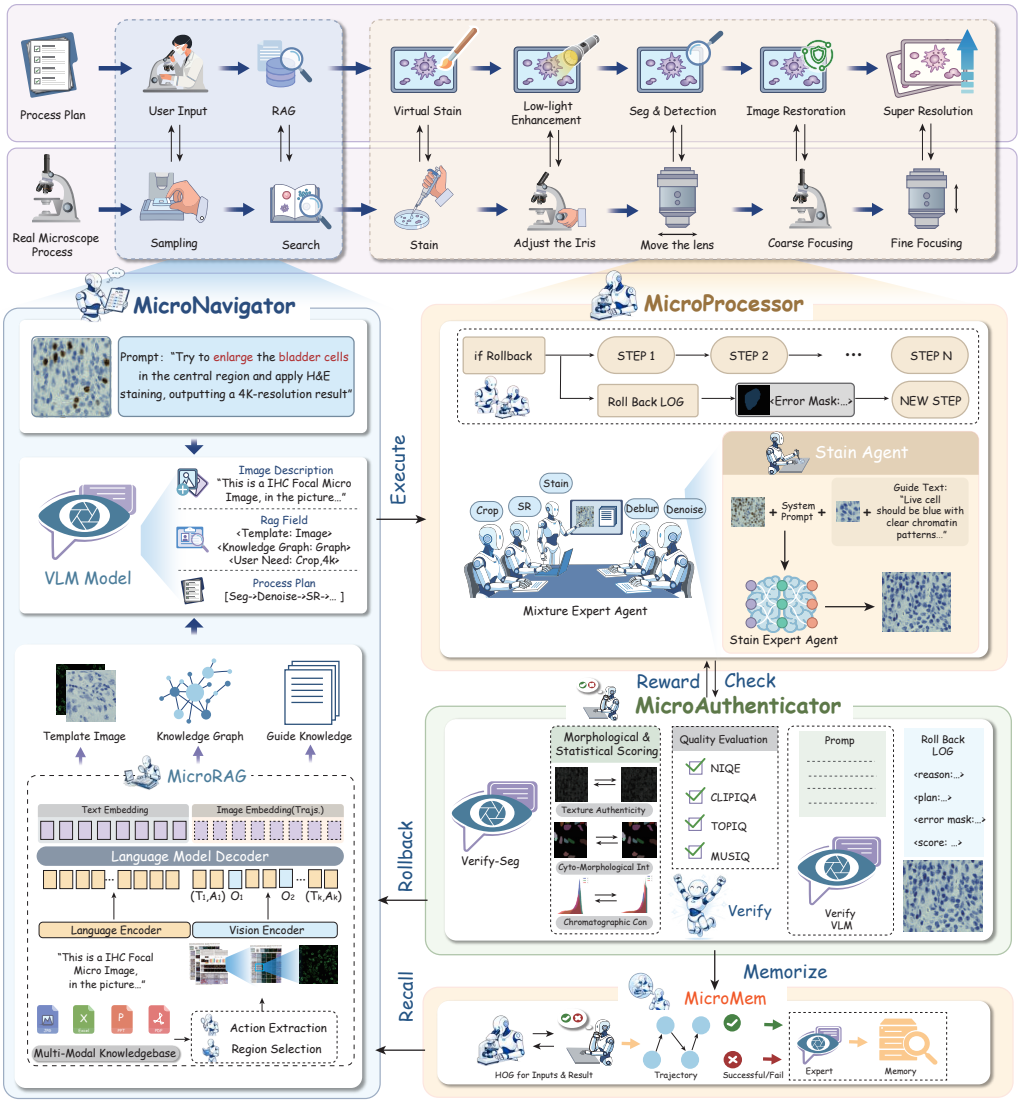


Figure 1: The Main Workflow of BioMicroAgents, include three main three agent groups: MicroNavigator (analysis and planning), MicroProcessor (execution), and MicroAuthenticator (verification). The evaluation metrics and results from MicroAuthenticator are fed back as reward signals and experience to MicroNavigator and MicroProcessor, enabling their co-evolution.

3.1 MicroNavigator

Traditional image reconstruction is inherently ill-posed, necessitating the incorporation of biomicro knowledge. Specifically, the introduction of template images provides color space and high-dimensional distribution information, ensuring consistency with biological logic.

(1) **NavigatorAgent** is a vision-language model (VLM) fine-tuned on biomicro image with corresponding descriptions. Beyond perceptual metrics, it considers the user input and invokes MicroRAG and MicroMem to retrieve$\{I_{temp}, K_{micro}\}$ and history$H$ for generating $\mathcal{P}$.

(2) **MicroRAG** builds a microscopy knowledge base from public literature, providing morphology and staining templates for conditional control. Literature is characterized by high information density. For better search, we adopt an Active Visual Perception framework that models retrieval as sequential **selection**, **cropping**, and **zooming** and generated Visual Perception Tokens(Fig 1,bottom-left), and the model is trained with GRPO using the following rewards:

$$r_\varphi = \alpha \cdot r_{\text{Ret}} + \beta \cdot r_{\text{Act}} + \gamma \cdot r_{\text{Pat}}. \tag{1}$$

where, $r_{\text{Pat}} \sim \text{Parse}(\mathcal{H})$ for rule-based pattern reward, $r_{\text{Act}} \sim \pi_{\text{RM}}(\cdot|\mathcal{Q}, \mathcal{A}_{\text{GT}}, \mathcal{A}_{\text{pred}})$ for action reward, $r_{\text{Ret}} = \frac{\text{DCG}(\mathcal{D}_{trj}, \mathcal{D}_{rel})}{\text{IDCG}(\mathcal{D}_{rel})}$ for retrieval efficiency reward.

(3) We enable a agentic memory to promote co-evolution between MicroNavigator and MicroAuthenticator. **MicroMem** is a memory repository co-maintained by MicroNavigator and MicroAuthenticator. It contains both successful and failed experience $E^+, E^- = \{q, r, g = \{\mathbf{H}, \Psi\}, p, b\}$ (q for query, g for score from MicroAuthenticator, p for process plan, b for error mask from MicroAuthenticator), enabling the system to draw lessons from historical data and logically organize processing plans.

3.2 MicroProcessor and MicroAuthenticator

(1) **Execution**: MicroProcessor supports five microscopic operations: ❶ low-light enhancement, ❷ deblurring, ❸ denoising, ❹ super-resolution, and ❺ virtual staining. These operations correspond to real-world processes such as light adjustment, focusing, stage stabilization, exposure control, objective magnification switching, and chemical staining.

We utilize:(i) **Brighten**: EnlightenGAN [21], KinD++ [48], LLflow [33], DiffLL [20], Zero-Dce [16], Retinex [35] (ii) **Deblur**: Uformer [34], Restormer [41], X-Restormer [6] (iii) **Denoise**: Noise2Noise [26], Restormer, X-Restormer, SwinIR [22] (iv) **SuperResolution**: DiffBIR [23], HAT [7], PFT [25], SwinIR, X-Restormer, MicroEdit-SR (v) **Virtual Stain**: Cycle-GAN [49], Stable Diffusion [28], Pixel-DiT [40], NextStep [29], MicroEdit-VS.

(2) **Evaluation**: To address non-biological artifacts and color distortions in perceptual-driven models, as well as the limitations of traditional fidelity metrics in improving downstream analysis and SR limits, we design the MicroAuthenticator. By combining segmentation models with perceptual metrics, it enforces sub-cellular structural verification and biochemical consistency on MicroProcessor outputs, ensuring morphological rigor during restoration and generation. To achieve this goal, we designed three fidelity metrics: **Cellular Morphological Integrity** (H_{CMI}), **Chromatographic Color Consistency** (H_{CC}), and **Histological Texture Authenticity** (H_{HTA}), denoted as $\mathbf{H} = [H_{CMI}, H_{CC}, H_{HTA}]^\top$.

(i) **Cyto-Morphological Integrity**: We measure this through the Dice and Average Surface Distance (ASD) after normalization:

$$H_{CMI} = \frac{1 + \beta^2}{\frac{1}{R_{Dice}} + \frac{\beta^2}{R_{ASD}}} \tag{2}$$

In this formulation, β is set to the mean value derived from the supervised fine-tuning (SFT) results.

(ii) **Chromatographic Consistency**: This metric is designed to ensure that the virtual staining model does not deviate from the authentic staining templates in terms of color space:

$$H_{CC} = \exp\left(-\lambda \cdot D_B(H_{out}, H_{ref})\right) \tag{3}$$

where $D_B(H_1, H_2) = -\ln\left(\sum_i \sqrt{P_1(i) \cdot P_2(i)}\right)$, and $P(i)$ represents the normalized color histogram frequency (Since fluctuations in luminance can simultaneously interfere with all three color channels and given that we employ a dedicated agent for illumination enhancement, we perform the calculation in the CIE Lab color space.).

(iii) **Histological Texture Authenticity**: Since algorithms based on conventional perceptual metrics are prone to introducing artifacts, we incorporate Histological Texture Authenticity (H_{HTA}) based on the Gram Matrix from pre-train Convolution neural network:

$$H_{HTA} = \exp\left(-\eta \cdot \sum_{l \in L} \mathcal{D}^l_{Gram}\right) \tag{4}$$

The model also considers perceptual metrics: NIQE and CLIPNIQA. If $\mathbf{H} < \tau_H \vee \psi < \tau_p$, a rollback mechanism is triggered. This process generates E^- in MicroMem, after which the MicroNavigator reschedules the $\mathcal{P}$.

3.3 Co-evolution between MicroProcessor and MicroAuthenticator by RL

Traditional interactions between verification and processing agents are often unidirectional or restricted to input-based control, making it difficult for the processing agent to learn the preferences of the verification agentparticularly in the case of **H**-based MicroAuthenticators.

To address this, we frame the interaction as a multi-reward reinforcement learning process. By feeding the evaluation metrics of the MicroAuthenticator back to the MicroProcessor as rewards, we mimic the evaluation-reflection loop of human experts, thereby achieving agent co-evolution. Although the designs of MicroAuthenticator and MicroProcessor are well established, achieving their co-evolution remains a major challenge. The diffusion component within MicroProcessor is flexible and powerful, making it difficult to optimize using multi-rule-based reward functions derived from **H**, and prone to reward collapse. To address this, inspired by [4, 37, 11, 24], we use the GDPO algorithm tailored for biomicro imaging tasks, enabling stable co-evolution between MicroAuthenticator and MicroProcessor.

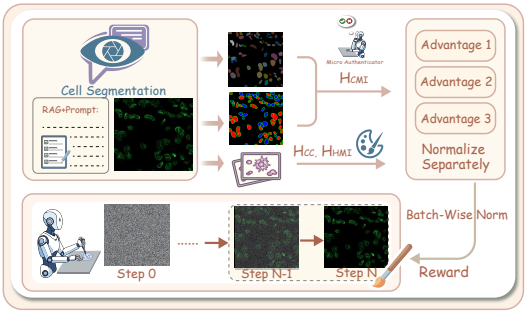


Figure 2: Computation of the reward signal

First, we formulate the denoising process of the diffusion model as a Markov Decision Process (MDP). To ensure semantic integrity and prevent reward hacking, we define the reward only at the final step (t = 0). Meanwhile, the deterministic rectified flow diffusion lacks stochastic exploration capability; following[2, 1, 12], we introduce an SDE formulation into its reverse process as follows:

$$dz_t = \left(u_t - \frac{1}{2}\epsilon_t^2 \nabla \log p_t(z_t) \right) dt + \epsilon_t dw \tag{5}$$

In 3.2, the reward has been regularized to converge within a stable bounded range. Meanwhile, inspired by [24, 17], to ensure stability under multiple rewards, we compute independent advantages A_{CMI}, A_{CC}, A_{HTA} for each reward dimension, sum them and then apply batch normalization.

$$\hat{A}_{\text{sum}}^{(i,j)} = \frac{A_{\text{sum}}^{(i,j)} - \text{mean}\left\{ A_{\text{sum}}^{(i',j')} \mid i' \in D_{\text{Batch}} \right\}}{\text{std}\left\{ A_{\text{sum}}^{(i',j')} \mid i' \in D_{\text{Batch}} \right\} + \epsilon}, A_{\text{sum}}^{(i,j)} = A_{CMI}^{(i,j)} + A_{CC}^{(i,j)} + A_{HTA}^{(i,j)} \tag{6}$$

4 Experiments

4.1 Comparison Experiment

We tested our method on various biological microscopy datasets and tasks including super-resolution (SR), virtual staining, deblurring, denoising, and brightening. We used several evaluation methods: similarity metrics (SSIM, PSNR) for visual quality, perceptual metrics (NIQE, CLIPIQA, FID) for realism, and fidelity metrics (ACC, ASD, DICE) for accuracy. Our model does not just focus on visual effects, but also considers the impact on downstream tasks like cell segmentation and counting.

We performed 4x, 6x, and 8x SR experiments on **bcSR** [19], **caco** [3], **IHC** [15], and **LAP** datasets, comparing SwinIR, X-Restormer, HAT, PFT, AgenticIR [50], DiffBIR, and 4KAgent [51]. Results in Tables 1 and 2 and Fig 3 , 8 and 9 show that current end-to-end methods are highly homogenized due to over-reliance on similarity constraints. Meanwhile, multi-agent methods focused on perception often create non-biological artifacts. Our method achieves a balance. In downstream segmentation and counting, our algorithm significantly improves boundary and structure clarity. This proves our method moves beyond basic enhancement to optimize semantic accuracy for biomedical analysis.

We evaluate our model on three virtual staining tasks: **IHC2HE**, **IHC2LAP**, and **IHC2BCL** [14], where the goal is to translate tissue images across staining modalities. Compared with Cycle-

Table 1: Quantitative comparison on different datasets for $4\times$, $6\times$, and $8\times$ SR. **Bold** and underline indicate the best and second-best performance.

Dataset	Method	4×				6×				8×			
		PSNR↑	SSIM↑	NIQE↓	CLIP↑	PSNR↑	SSIM↑	NIQE↓	CLIP↑	PSNR↑	SSIM↑	NIQE↓	CLIP↑
bcSR	SwinIR	28.13	0.73	4.95	0.258	24.16	0.52	4.98	0.231	22.48	0.36	5.02	0.205
	X-Restormer	28.58	0.75	4.88	0.294	24.62	0.55	4.95	0.272	22.85	0.41	4.95	0.246
	HAT	29.01	0.78	4.82	0.335	25.11	0.58	4.84	0.312	23.16	0.46	4.89	0.283
	PFT	28.84	0.77	4.84	0.312	24.95	0.57	4.87	0.298	23.02	0.44	4.91	0.265
	AgenticIR	29.81	0.82	4.75	0.417	25.99	0.65	4.62	0.389	23.75	0.52	4.81	0.359
	DiffBIR	30.15	0.84	4.71	0.449	26.31	0.68	4.76	0.415	24.11	0.55	4.78	0.392
	4KAgent	30.49	0.87	4.67	0.482	26.85	0.73	4.73	0.456	24.48	0.58	4.74	0.428
	Ours	**30.75**	**0.88**	**4.16**	**0.537**	**27.22**	**0.77**	**4.18**	**0.512**	**24.81**	**0.62**	**4.63**	**0.433**
caco	SwinIR	29.85	0.746	-	-	25.88	0.582	-	-	23.25	0.441	-	-
	X-Restormer	30.02	0.754	-	-	26.27	0.604	-	-	23.51	0.468	-	-
	HAT	30.24	0.768	-	-	26.42	0.621	-	-	23.78	0.485	-	-
	PFT	30.41	0.787	-	-	26.65	0.645	-	-	24.02	0.508	-	-
	AgenticIR	27.45	0.621	-	-	23.18	0.458	-	-	21.05	0.312	-	-
	DiffBIR	29.68	0.728	-	-	25.62	0.562	-	-	23.10	0.425	-	-
	4KAgent	30.76	0.805	-	-	27.05	0.702	-	-	24.36	0.572	-	-
	Ours	**30.95**	**0.818**	-	-	**27.32**	**0.745**	-	-	**24.97**	**0.613**	-	-
IHC	SwinIR	31.25	0.84	4.55	0.252	25.47	0.62	4.78	0.233	24.08	0.51	4.92	0.207
	X-Restormer	31.68	0.86	4.46	0.268	25.92	0.65	4.71	0.245	24.36	0.55	4.86	0.221
	HAT	32.12	0.89	4.42	0.281	26.41	0.68	4.67	0.263	24.76	0.60	4.76	0.243
	PFT	**32.56**	**0.91**	4.39	0.294	**26.84**	**0.71**	4.62	0.278	**25.03**	**0.64**	4.77	0.259
	AgenticIR	30.11	0.80	4.16	0.342	24.71	0.56	4.12	0.331	23.19	0.45	4.38	0.315
	DiffBIR	30.54	0.82	4.02	0.351	25.20	0.59	4.05	0.342	23.71	0.48	4.15	0.328
	4KAgent	30.35	0.81	3.89	0.364	24.95	0.57	3.92	0.358	23.45	0.46	3.98	0.346
	Ours	30.86	0.84	**3.74**	**0.371**	25.40	0.61	**3.76**	**0.369**	24.01	0.50	**3.82**	**0.355**
LAP	SwinIR	34.25	0.86	4.22	0.261	28.16	0.62	4.25	0.235	24.48	0.46	4.29	0.212
	X-Restormer	34.86	0.87	4.16	0.262	28.59	0.64	4.23	0.248	24.85	0.48	4.25	0.224
	HAT	35.31	0.89	4.15	0.274	29.14	0.66	4.27	0.259	25.37	0.51	4.26	0.235
	PFT	**35.94**	**0.91**	4.12	0.295	**29.54**	**0.69**	4.13	0.261	**25.43**	**0.54**	4.22	0.248
	AgenticIR	32.61	0.79	3.75	0.316	27.23	0.59	3.79	0.288	23.71	0.42	3.82	0.261
	DiffBIR	33.10	0.82	3.58	0.328	27.71	0.61	3.61	0.304	24.36	0.45	3.73	0.283
	4KAgent	32.84	0.85	3.32	0.345	27.46	0.60	3.34	0.317	23.95	0.43	3.39	0.297
	Ours	33.35	0.83	**3.15**	**0.362**	27.95	0.63	**3.18**	**0.338**	24.40	0.46	**3.36**	**0.315**

GAN, diffusion-based methods, and advanced image translation models (VQ-GAN [30], AR-Transformer [10], Pixel-DiT, NextStep, and QwenEdit [36]), our approach achieves the best overall performance (Table 3, Fig 10, 4 and 11). For fairness, SD, Pixel-DiT, and ImageEdit are all injected with annotated template images and knowledge during fine-tuning. BioMicroAgents not only performs accurate cross-stain translation but also preserves cellular structures and boundaries. Meanwhile, downstream metrics such as DICE demonstrate that co-evolution based on is beneficial for enhancing structural information and promoting downstream performance 3.3. This is especially evident in the IHC2BCL task 11, which requires distinguishing subtle differences between live and dead cells; unlike other methods that introduce artifacts, our model enables more precise translation, leading to improved counting and segmentation accuracy of dead cells.

We conduct denoising and low-light enhancement experiments on the ConfocalBPAE [47] and MicroBrighten datasets, where MicroBrighten is a microscopy low-light enhancement dataset that we specifically propose and describe in detail in the Appendix A. Blur in microscopy images is typically caused by vibration and long-term acquisition drift and rarely exhibits the severe degradation commonly seen in natural images; therefore, we simulate defocus and motion blur on the IHC dataset (see Appendix A for details). The results of the three experiments are shown in Tables 4, 5, 6 and Fig 14, 12, 13.

4.2 Ablation Study

We follow a standard agentic paradigm (Text-RAG analysis planning verification) and adopt an orthogonal ablation strategy to isolate and evaluate the contributions of different components, including multimodal RAG, RL-based co-evolution, memory-based co-evolution. Notably, virtual staining itself benefits cell counting and segmentation [13]; thus, we account for this in our experimental design: VS denotes inclusion of virtual staining, while NOVS indicates its absence. The results(Tab 7, Fig 6b, Fig 5) demonstrate that, through co-evolution, the model can simultaneously improve perceptual metrics and downstream task accuracy. Multimodal RAG is more effective than text-only RAG and removing either failure memory (NOFM) or success memory (NOSM) degrades performance across all metrics(meanwhile, strategy replication is more critical than risk aversion).

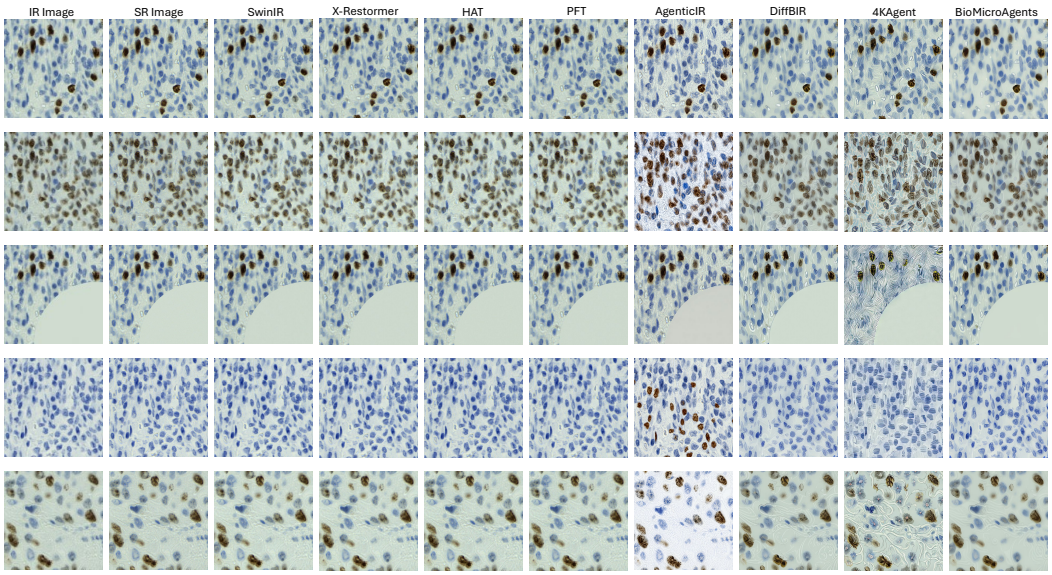


Figure 3: Visual comparisons on Micro Image SR task (Immunohistochemistry, 4X, Whole)

Table 2: Quantitative comparison on IHC and LAP datasets for $4\times$, $6\times$, and $8\times$ super-resolution.

Dataset	Method	$4\times$		$6\times$		$8\times$	
		DICE↑	ACC↑	DICE↑	ACC↑	DICE↑	ACC↑
IHC	SwinIR	0.682	0.761	0.634	0.724	0.582	0.688
	X-Restormer	0.685	0.772	0.638	0.738	0.586	0.705
	HAT	0.689	0.785	0.642	0.752	0.591	0.728
	PFT	0.692	0.796	0.645	0.765	0.595	0.741
	AgenticIR	0.641	0.716	0.592	0.673	0.548	0.639
	DiffBIR	0.684	0.821	0.635	0.804	0.584	0.785
	4KAgent	0.638	0.702	0.588	0.662	0.442	0.577
	Ours	0.724	0.826	0.678	0.815	0.621	0.782
LAP	SwinIR	0.681	0.765	0.632	0.732	0.579	0.695
	X-Restormer	0.684	0.778	0.636	0.745	0.583	0.712
	HAT	0.688	0.789	0.639	0.761	0.588	0.734
	PFT	0.691	0.801	0.644	0.778	0.592	0.751
	AgenticIR	0.639	0.782	0.589	0.758	0.545	0.714
	DiffBIR	0.683	0.834	0.633	0.815	0.582	0.771
	4KAgent	0.635	0.703	0.585	0.637	0.541	0.613
	Ours	0.722	0.829	0.675	0.818	0.618	0.778

Meanwhile, we visualize the reward variations and corresponding outputs during the co-evolution process of the MicroAuthenticator and MicroProcessor in Fig 6a. It can be observed that non-biological artifacts are reduced while structural information is enhanced. We conducted a memory-evolution study on a fixed held-out test set. An online memory stream was constructed only from the training split, and long-term memory snapshots were saved at different stream ratios. For each snapshot, all model parameters were frozen, and the same held-out test

Table 7: Ablation Study on Coev. and RAG

Task	Coev-RL.	RAG	NIQE↓	CLIPIQA↑	ACC↑	DICE↑
HE(VS)	✓	✓	2.82	0.435	0.895	0.782
	×	✓	3.21	0.363	0.883	0.772
Lap(VS)	✓	✓	2.81	0.394	0.892	0.781
	×	✓	3.23	0.341	0.879	0.772
BCL(VS)	✓	✓	2.93	0.257	0.825	0.685
	×	✓	3.28	0.253	0.811	0.662
Lap(NOVS)	✓	✓	3.15	0.362	0.829	0.722
	×	✓	3.39	0.235	0.821	0.717
Lap(VS)	✓	✓	2.81	0.394	0.892	0.781
	✓	×	2.74	0.403	0.871	0.738
Lap(NOVS)	✓	✓	3.15	0.362	0.829	0.722
	✓	×	3.12	0.338	0.853	0.715
Lap(NOFM)	✓	✓	3.05	0.352	0.882	0.772
Lap(NOSM)	✓	✓	3.18	0.331	0.878	0.768

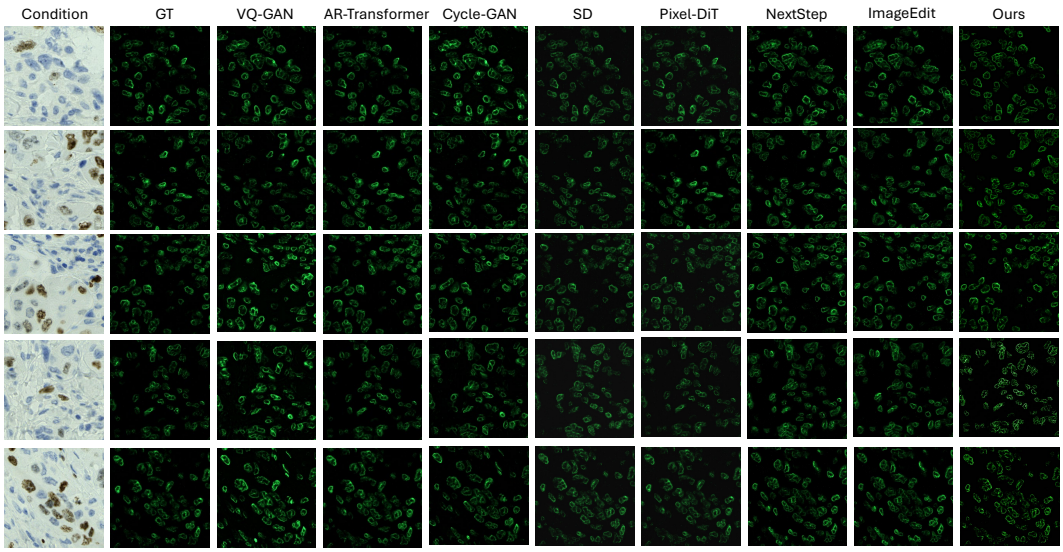


Figure 4: Visual comparisons on Virtual Staining(Immunohistochemistry2Lap)

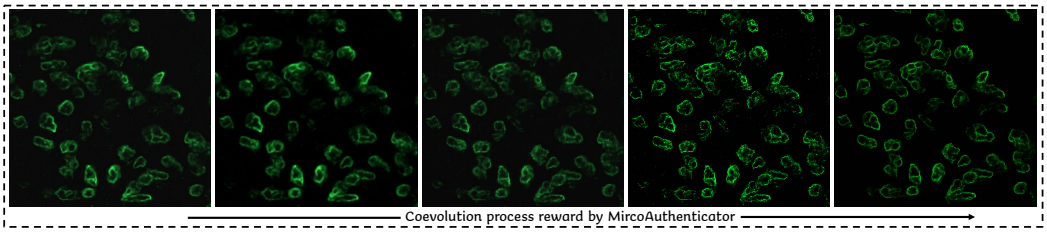
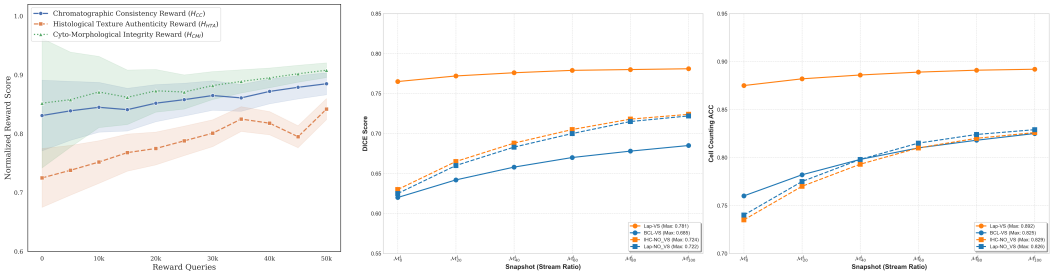


Figure 5: Visual comparisons on the coevolution process

set was evaluated with memory write-back disabled. This protocol isolates the contribution of retrieved historical experience from parameter updates and prevents test-time memory leakage.



(a) Micro-GDPO reward curve

(b) Memory-evolution analysis.

Figure 6: Visualization of reward evolution and memory mechanism.

5 Conclusion and Limitation

Despite achieving state-of-the-art results on current tasks and benchmarks, this work is limited by data availability and lacks validation on larger datasets and more diverse tasks. In practice, microscopy workflows naturally produce many degradedgold standard pairs, but failed or suboptimal samples are often discarded, resulting in data waste. Future work will focus on collecting these data, expanding dataset scale and task diversity, and improving the models generalization.

Table 3: Quantitative comparison on different datasets of Virtual Stain Task.

Dataset	Method	Fidelity Metrics				Perception Metrics		
		PSNR↑	Dice↑	ASD↓	ACC↑	SSIM↑	NIQE↓	CLIPIQA↑
HE	*Image Control*							
	VQ-GAN	34.05	0.682	3.12	0.813	0.802	3.78	0.214
	AR-Trans	34.17	0.715	2.85	0.825	0.828	3.72	0.245
	Cycle-GAN	34.36	0.751	2.46	0.862	0.840	3.65	0.268
	Image-Text Control							
	Diffusion	34.53	0.758	2.31	0.868	0.853	3.59	0.302
	Pixel-DiT	34.76	0.762	2.15	0.872	0.874	3.42	0.325
	NextStep	35.12	0.764	1.97	0.879	0.885	3.35	0.348
	QwenEdit	35.47	0.774	1.83	0.885	0.898	3.22	0.367
	Coevolution Control							
	Ours	**35.52**	**0.782**	**1.62**	**0.895**	**0.912**	**2.82**	**0.435**
Lap2	VQ-GAN	34.02	0.680	3.18	0.805	0.805	3.75	0.198
	AR-Trans	34.15	0.712	2.92	0.828	0.821	3.68	0.221
	Cycle-GAN	34.28	0.753	2.48	0.861	0.835	3.61	0.242
	Diffusion	34.50	0.756	2.38	0.865	0.849	3.52	0.275
	Pixel-DiT	34.82	0.761	2.22	0.871	0.864	3.45	0.298
	NextStep	35.05	0.766	2.05	0.875	0.880	3.32	0.315
	QwenEdit	34.96	0.768	1.88	0.882	0.907	3.25	0.342
	Ours	**35.12**	**0.781**	**1.58**	**0.892**	**0.912**	**2.81**	**0.394**
BCL	VQ-GAN	16.26	0.212	–	0.425	0.312	6.73	0.133
	AR-Trans	31.25	0.584	4.15	0.748	0.728	3.65	0.235
	Cycle-GAN	31.28	0.615	3.82	0.762	0.745	3.55	0.238
	Diffusion	31.45	0.628	3.65	0.775	0.765	3.49	0.244
	Pixel-DiT	31.52	0.643	3.42	0.782	0.772	3.42	0.246
	NextStep	31.60	0.655	3.18	0.793	0.805	3.36	0.247
	QwenEdit	31.75	0.658	2.95	0.808	0.824	3.25	0.254
	Ours	**31.95**	**0.685**	**2.62**	**0.825**	**0.842**	**2.93**	**0.257**

Table 4: DeBlur Task

Method	DefocusBlur		MotionBlur	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Uformer	32.125	0.8242	32.054	0.8315
Restormer	33.582	0.8915	33.421	0.8806
AgenticIR	34.014	0.9128	33.876	0.9054
4KAgent	34.652	0.9385	34.412	0.9247
Ours	**35.108**	**0.9562**	**34.985**	**0.9418**

Table 5: Brighten Task

Method	PSNR↑	SSIM↑	LPIPS↓
EnlightenGAN	16.28	0.61	0.35
KinD++	19.87	0.81	0.19
LLFlow	23.76	0.91	**0.12**
DiffLL	23.43	0.90	0.15
Zero-DCE	13.56	0.54	0.37
RetinexNet	15.43	0.52	0.48
Ours	**24.40**	**0.92**	0.13

Table 6: Denoise Task

Method	PSNR↑	SSIM↑	NIQE↓
SwinIR	32.94	0.788	3.53
N2N	33.54	0.813	3.38
Restormer	33.62	0.825	3.18
AgenticIR	35.16	0.871	2.75
4KAgent	**35.33**	**0.872**	2.73
Ours	34.29	0.843	**2.13**

6 Acknowledgements

This work was supported by the National Key Research and Development Program (2025YFC3410001). Computational resources and debugging support were provided by the Scientific Computing Platform of South China University of Technology. We sincerely appreciate all the support provided.

References

[1] Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

[2] Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

[3] Soufiane Belharbi, Mara K Whitford, Phuong Hoang, Shakeeb Murtaza, Luke McCaffrey, and Eric Granger. Sr-caco-2: A dataset for confocal fluorescence microscopy image super-resolution. *Advances in Neural Information Processing Systems*, 37:59948–59983, 2024.

[4] Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*, 2023.

[5] Tianyi Chen, Yinheng Li, Michael Solodko, Sen Wang, Nan Jiang, Tingyuan Cui, Junheng Hao, Jongwoo Ko, Sara Abdali, Leon Xu, et al. Cua-skill: Develop skills for computer using agent. *arXiv preprint arXiv:2601.21123*, 2026.

[6] Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *European Conference on Computer Vision*, pages 74–91. Springer, 2024.

[7] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023.

[8] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning, 2020.

[9] Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. Mem0: Building production-ready ai agents with scalable long-term memory, 2025.

[10] Patrick Esser, Robin Rombach, and Björn Ommer. Taming transformers for high-resolution image synthesis, 2020.

[11] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems*, 36:79858–79885, 2023.

[12] Ruiqi Gao, Emiel Hoogeboom, Jonathan Heek, Valentin De Bortoli, Kevin P Murphy, and Tim Salimans. Diffusion meets flow matching: Two sides of the same coin. 2024. *URL https://diffusionflow. github. io*, 2024.

[13] P. Ghahremani, Y. Li, A. Kaufman, et al. Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification. *Nature Machine Intelligence*, 4:401–412, 2022.

[14] Parmida Ghahremani, Yanyun Li, Arie Kaufman, Rami Vanguri, Noah Greenwald, Michael Angelo, Travis J Hollmann, and Saad Nadeem. Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification. *Nature machine intelligence*, 4(4):401–412, 2022.

[15] Parmida Ghahremani, Joseph Marino, Ricardo Dodds, and Saad Nadeem. Deepliif: An online platform for quantification of clinical pathology slides. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21399–21405, 2022.

[16] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1780–1789, 2020.

[17] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[18] Zirui Guo, Lianghao Xia, Yanhua Yu, Tu Ao, and Chao Huang. Lightrag: Simple and fast retrieval-augmented generation. 2024.

[19] Feiyang Jia, Li Tan, Ge Wang, Caiyan Jia, and Zhineng Chen. A super-resolution network using channel attention retention for pathology images. *PeerJ Computer Science*, 9:e1196, 2023.

[20] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (ToG)*, 42(6):1–14, 2023.

[21] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021.

[22] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.

[23] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *European conference on computer vision*, pages 430–448. Springer, 2024.

[24] Shih-Yang Liu, Xin Dong, Ximing Lu, Shizhe Diao, Peter Belcak, Mingjie Liu, Min-Hung Chen, Hongxu Yin, Yu-Chiang Frank Wang, Kwang-Ting Cheng, et al. Gdpo: Group reward-decoupled normalization policy optimization for multi-reward rl optimization. *arXiv preprint arXiv:2601.05242*, 2026.

[25] Wei Long, Xingyu Zhou, Leheng Zhang, and Shuhang Gu. Progressive focused transformer for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2279–2288, 2025.

[26] Youssef Mansour and Reinhard Heckel. Zero-shot noise2noise: Efficient image denoising without any data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14018–14027, 2023.

[27] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017.

[28] 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 (CVPR)*, pages 10684–10695, June 2022.

[29] NextStep Team, Chunrui Han, Guopeng Li, Jingwei Wu, Quan Sun, Yan Cai, Yuang Peng, Zheng Ge, Deyu Zhou, Haomiao Tang, Hongyu Zhou, Kenkun Liu, Ailin Huang, Bin Wang, Changxin Miao, Deshan Sun, En Yu, Fukun Yin, Gang Yu, Hao Nie, Haoran Lv, Hanpeng Hu, Jia Wang, Jian Zhou, Jianjian Sun, Kaijun Tan, Kang An, Kangheng Lin, Liang Zhao, Mei Chen, Peng Xing, Rui Wang, Shiyu Liu, Shutao Xia, Tianhao You, Wei Ji, Xianfang Zeng, Xin Han, Xuelin Zhang, Yana Wei, Yanming Xu, Yimin Jiang, Yingming Wang, Yu Zhou, Yucheng Han, Ziyang Meng, Binxing Jiao, Daxin Jiang, Xiangyu Zhang, and Yibo Zhu. Nextstep-1: Toward autoregressive image generation with continuous tokens at scale. *arXiv preprint arXiv:2508.10711*, 2025.

[30] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

[31] Jianhui Wang, Yangfan He, Kun Li, Sida Li, Lan Zhao, Jun Yin, Miao Zhang, Tianyu Shi, and Xueqian Wang. Mdanet: A multi-stage domain adaptation framework for generalizable low-light image enhancement. *Neurocomputing*, page 129572, 2025.

[32] Qiuchen Wang, Ruixue Ding, Yu Zeng, Zehui Chen, Lin Chen, Shihang Wang, Pengjun Xie, Fei Huang, and Feng Zhao. Vrag-rl: Empower vision-perception-based rag for visually rich information understanding via iterative reasoning with reinforcement learning. *arXiv preprint arXiv:2505.22019*, 2025.

[33] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 2604–2612, 2022.

[34] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022.

[35] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

[36] Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhengyi Wang, An Yang, Bowen Yu, Chen Cheng, Dayiheng Liu, Deqing Li, Hang Zhang, Hao Meng, Hu Wei, Jingyuan Ni, Kai Chen, Kuan Cao, Liang Peng, Lin Qu, Minggang Wu, Peng Wang, Shuting Yu, Tingkun Wen, Wensen Feng, Xiaoxiao Xu, Yi Wang, Yichang Zhang, Yongqiang Zhu, Yujia Wu, Yuxuan Cai, and Zenan Liu. Qwen-image technical report, 2025.

[37] Zeyue Xue, Jie Wu, Yu Gao, Fangyuan Kong, Lingting Zhu, Mengzhao Chen, Zhiheng Liu, Wei Liu, Qiushan Guo, Weilin Huang, et al. Dancegrpo: Unleashing grpo on visual generation. *arXiv preprint arXiv:2505.07818*, 2025.

[38] Yutao Yang, Junsong Li, Qianjun Pan, Bihao Zhan, Yuxuan Cai, Lin Du, Jie Zhou, Kai Chen, Qin Chen, Xin Li, et al. Autoskill: Experience-driven lifelong learning via skill self-evolution. *arXiv preprint arXiv:2603.01145*, 2026.

[39] Runpeng Yu, Xinyin Ma, and Xinchao Wang. Introducing visual perception token into multimodal large language model. *arXiv preprint arXiv:2502.17425*, 2025.

[40] Yongsheng Yu, Wei Xiong, Weili Nie, Yichen Sheng, Shiqiu Liu, and Jiebo Luo. Pixeldit: Pixel diffusion transformers for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

[41] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022.

[42] Tianyi Zeng, Tianyi Wang, Miao Zhang, Jun Yin, Zimo Zeng, Feiyang Zhang, Yangyang Wang, Junfeng Jiao, Yuantao Wang, Yangfan He, et al. Tcstnet: A text-driven color style transfer network for low-light image enhancement. *Expert Systems with Applications*, page 130012, 2025.

[43] Hanrong Zhang, Shicheng Fan, Henry Peng Zou, Yankai Chen, Zhenting Wang, Jiayu Zhou, Chengze Li, Wei-Chieh Huang, Yifei Yao, Kening Zheng, et al. Evoskills: Self-evolving agent skills via co-evolutionary verification. *arXiv preprint arXiv:2604.01687*, 2026.

[44] Haozhen Zhang, Quanyu Long, Jianzhu Bao, Tao Feng, Weizhi Zhang, Haodong Yue, and Wenya Wang. Memskill: Learning and evolving memory skills for self-evolving agents. *arXiv preprint arXiv:2602.02474*, 2026.

[45] Miao Zhang, Jun Yin, Pengyu Zeng, Yiqing Shen, Shuai Lu, and Xueqian Wang. Tscnet: A text-driven semantic-level controllable framework for customized low-light image enhancement. *Neurocomputing*, page 129509, 2025.

[46] Shengtao Zhang, Jiaqian Wang, Ruiwen Zhou, Junwei Liao, Yuchen Feng, Zhuo Li, Yujie Zheng, Weinan Zhang, Ying Wen, Zhiyu Li, et al. Memrl: Self-evolving agents via runtime reinforcement learning on episodic memory. *arXiv preprint arXiv:2601.03192*, 2026.

[47] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11710–11718, 2019.

[48] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International journal of computer vision*, 129(4):1013–1037, 2021.

[49] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[50] Kaiwen Zhu, Jinjin Gu, Zhiyuan You, Yu Qiao, and Chao Dong. An intelligent agentic system for complex image restoration problems. In *The Thirteenth International Conference on Learning Representations*, 2025.

[51] Yushen Zuo, Qi Zheng, Mingyang Wu, Xinrui Jiang, Renjie Li, Jian Wang, Yide Zhang, Gengchen Mai, Lihong V. Wang, James Zou, Xiaoyu Wang, Ming-Hsuan Yang, and Zhengzhong Tu. 4kagent: Agentic any image to 4k super-resolution. 2025.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract and introduction outline our key algorithmic contributions and empirical gains.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

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

   Answer: [Yes]

   Justification: We have dedicated Section 5 to discuss the limitations of our approach.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

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

   Answer: [N/A]

Justification: This paper does not include theoretical results.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

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

Answer: [Yes]

Justification: Section 4 and the Appendix provide sufficient details to ensure the reproducibility of this work, with the Appendix specifically covering component selection and hyperparameter settings. Furthermore, our model was trained and fine-tuned on publicly available models and datasets.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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

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

   Answer: [No]

   Justification: The newly proposed dataset has been submitted as part of the supplementary material. However, as the experimental code is currently unorganized and contains some internal logs, we need additional time to clean it up. That said, our framework is built entirely on existing public models and datasets, so we are confident that it can be easily reproduced.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: We include the experimental design and details in Section 4 and the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

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

   Answer: [No]

   Justification:There is no statistical significance experiment in this study.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: We provide detailed hardware configurations in the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: We have read the NeurIPS Code of Ethics and confirm that our study adheres to all its guidelines, particularly regarding data privacy and potential societal impacts.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

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

    Answer: [N/A]

    Justification: Our work is fundamental in nature, and it is currently difficult to foresee any negative applications.

    Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

    Answer: [N/A]

    Justification: This paper does not involve such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: Yes, we have explicitly mentioned and strictly adhered to its license and terms of use.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

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

    Answer: [Yes]

    Justification: Yes, we have provided a detailed description of the dataset in the Appendix.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

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

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

    Answer: [No]

    Justification: This paper does not involve these studies.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

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

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

    Answer: [N/A]

    Justification: This paper does not involve these studies.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

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

    Answer: [N/A]

    Justification: We only used Large Language Models (LLMs) for polishing and formatting.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.

A Dataset and Equipment

To systematically evaluate our model's performance and generalization across three super-resolution tasks, we utilized four cross-domain datasets: (1) DEEPLIIFIHC, which assesses the restoration of complex histopathological textures and fine biomarker structures; (2) DEEPLIIFIHCLAP, a LAP2-focused subset for evaluating these specific fine structures; (3) caco, used to evaluate single-cell microstructural reconstruction; and (4) bcSR, a benchmark dataset of 1,200 breast cancer lymph node tissue images designed for pathological medical image evaluation.

Additionally, we evaluated the model's virtual staining performance using the IHC2LAP, IHC2HE, and IHC2BCL subsets. By computationally translating medical images between staining styles, virtual staining eliminates cumbersome chemical procedures. This approach preserves scarce tissue samples, reduces diagnostic costs, and enables multimodal analysis on a single tissue section, providing a comprehensive foundation for precise computer-aided diagnosis.

To address the scarcity of low-light microscopy enhancement datasets, we propose the MicroBrighten dataset. This microscopic biology dataset systematically collects multiple categories of classic biological histology, cytology, and morphology teaching samples.

We tested the denoising task on the ConfocalBPAE dataset, which is a specialized denoising dataset for microscopy images covering the three primary modalities of fluorescence microscopy: confocal microscopy, two-photon microscopy, and wide-field microscopy. All images were acquired using high-quality commercial fluorescence microscopes and imaged from real biological specimens.

Defocus blur and motion blur in microscope images are often caused by slight tilting and trembling of the sample during the capture process, or by improper operation. Compared to blur in real-world scenarios, this type of blur is usually not particularly severe and can be uniformly modeled with super-resolution models. To verify the effectiveness of the model in deblurring, we simulated the mathematical processes of two types of blur and constructed deblurred image pairs on the IHC dataset. Specifically:

(1) Defocus Blur: The point spread function is simulated by constructing a circular mask with a radius of R and internal weights of $1/(\pi R^2)$, which is then convolved with the sharp image three times(We set $R = 25px$).

(2) Motion Blur: A linear kernel is constructed to simulate the directionality and displacement of motion blur, including hyperparameters length L and angle $\theta()$.

Table 8: The split of Dataset

Dataset	Caco	BcSR	IHC	Lap	IHC2BCL	IHC2HE	IHC2Lap	MicroBrighten	BPAE	IHC	IHC
Task	SR	SR	SR	SR	VS	VS	VS	Brighten	Denoise	Defocus	Demotion
Ratio	6:3:1	7:2:1	7:2:1	7:2:1	7:2:1	7:2:1	7:2:1	7:2:1	8:1:1	7:2:1	7:2:1
Open-access	1	1	1	1	1	1	1	1	1	1	1

All experiments were conducted on two primary platforms: a server equipped with four NVIDIA A100 GPUs (80GB) and 512GB RAM, running Ubuntu 22.04 with CUDA 12.8; and a workstation featuring an NVIDIA Pro 6000 (96GB, 256GB RAM), running Ubuntu 24.04 with CUDA 13.1. Additionally, an NVIDIA RTX 4090 (Windows 11) was utilized for preliminary validations to ensure computational efficiency.

B Results and Analysis

B.1 Micro Image Super-Resolution

From Fig. 7, we observe that traditional reconstruction-based methods are fundamentally constrained by the input resolution, making it difficult to recover fine-grained cellular structures beyond the original signal. In contrast, perceptual-driven approaches, while improving visual quality metrics, often introduce hallucinated artifacts that are not biologically grounded, potentially distorting cellular morphology and staining consistency(Fig 9, 3). For downstream biological analysis(Table 2), such as segmentation and cell counting, structural fidelity is more critical than strict pixel-wise alignment with ground truth. In particular, accurate boundary delineation is essential in

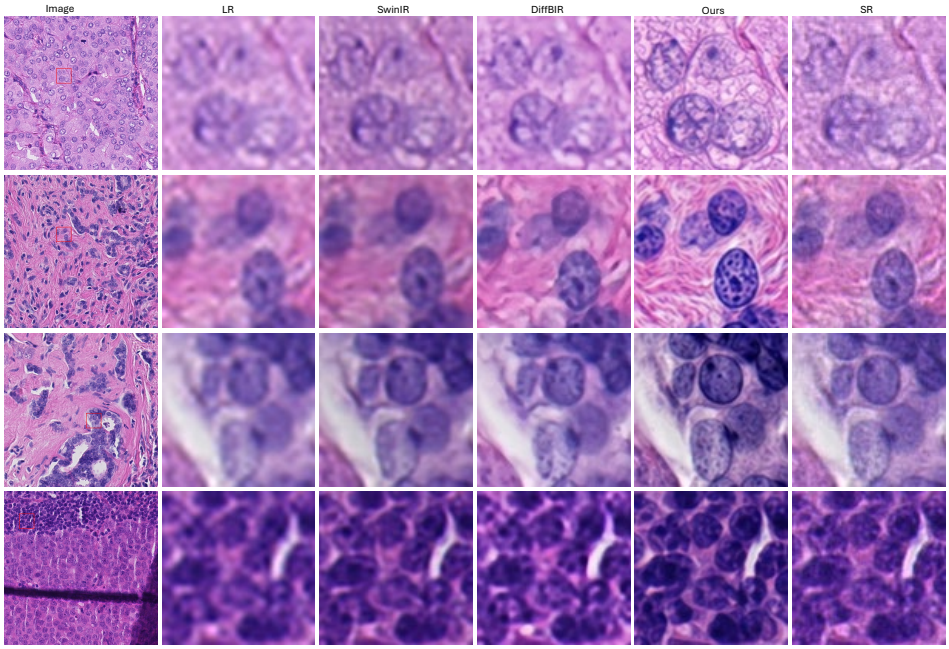


Figure 7: Visual comparisons on Micro Image SR task (bcSR, 4X, Zoom)

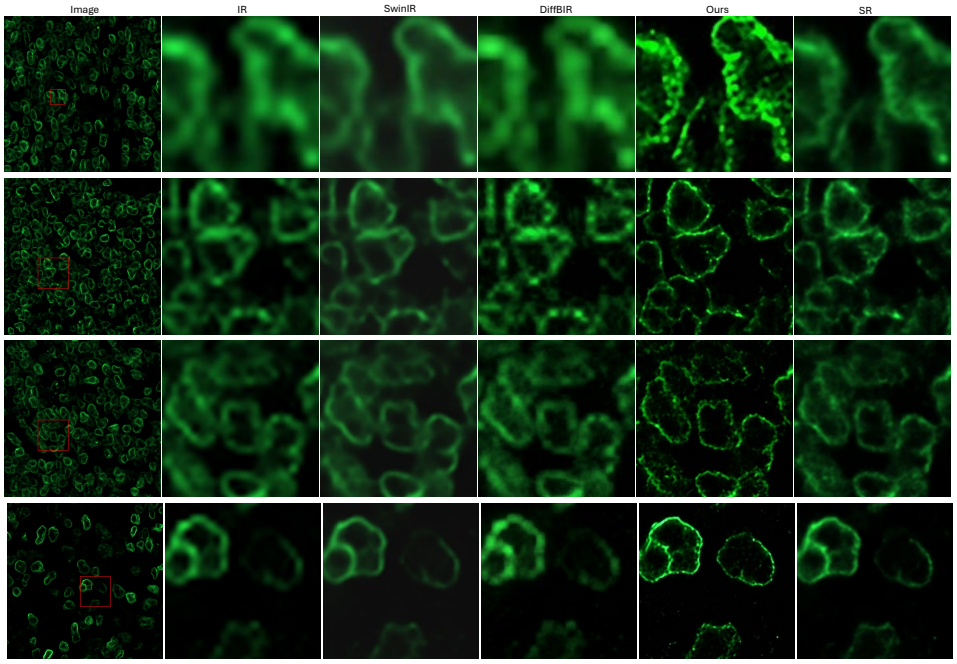


Figure 8: Visual comparisons on Micro Image SR task (Lap, 4X, Zoom)

challenging scenarios, including densely packed cells and regions with insufficient staining, where weak signals can easily be overwhelmed or misinterpreted. These issues become more evident in complex cases, where perceptual methods tend to generate spurious textures, while traditional methods fail to resolve ambiguous structures. In contrast, our method consistently enhances biologically meaningful structures, especially boundary and morphological consistency, leading to improved downstream performance.

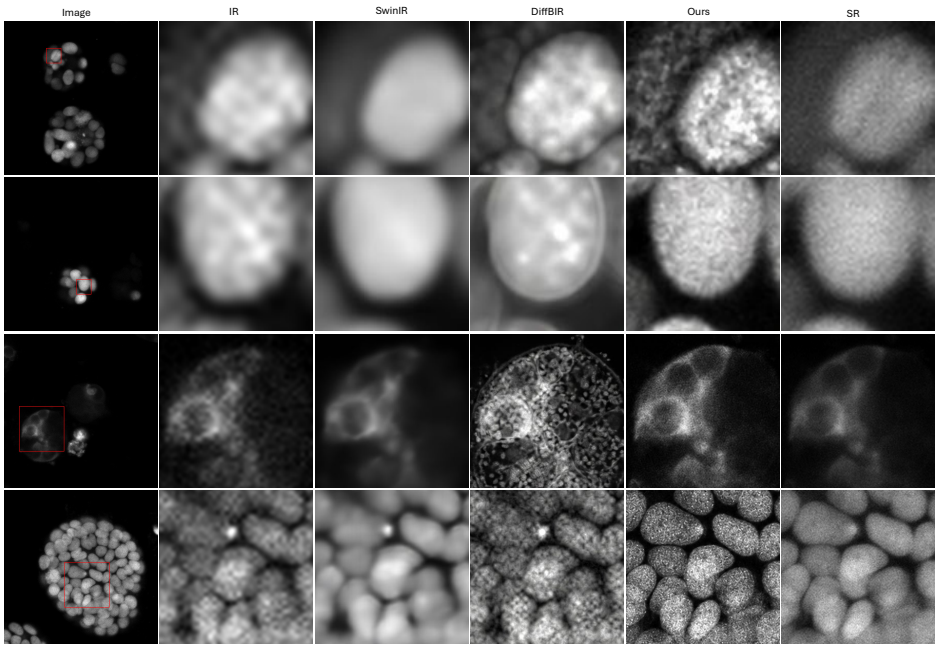


Figure 9: Visual comparisons on Micro Image SR task (caco, 4X, Zoom)

B.2 Micro Image Virtual Stain

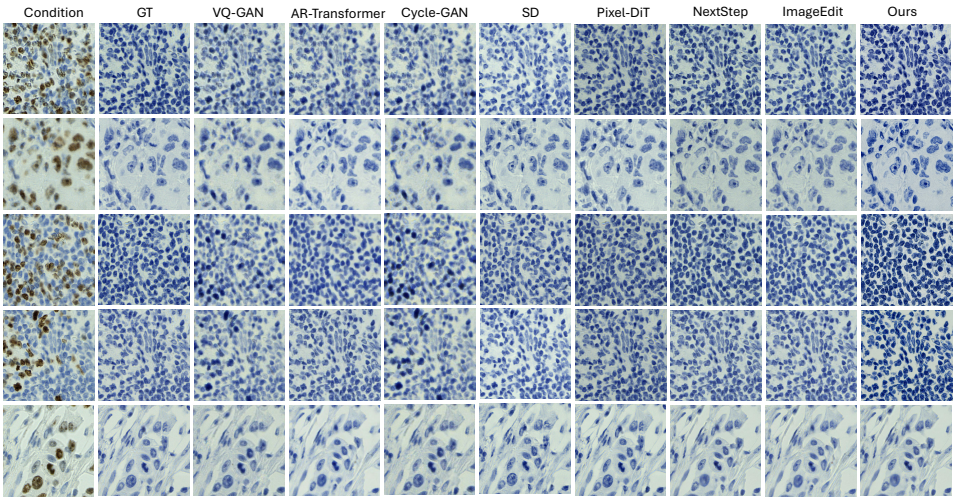


Figure 10: Visual comparisons on Virtual Staining(Immunohistochemistry2Hematoxylin)

Virtual staining requires the model to not only learn the transformation across staining domains, but also to capture the underlying semantic structure of cellular morphology. This involves preserving biologically meaningful patterns while ensuring consistency in color distribution and staining logic. However, existing methods often struggle to balance these two aspects. Traditional image-to-image translation approaches tend to focus on pixel-level mapping, which may lead to over-smoothing and loss of fine structural details. On the other hand, generative or perceptual-driven models can produce visually plausible results but often introduce incorrect staining patterns or non-biological textures, especially in regions with ambiguous or weak signals.For downstream biological analysis, such as cell classification and counting, accurate semantic alignment between structure and staining is more critical than purely visual realism. In particular, regions with heterogeneous cell populations

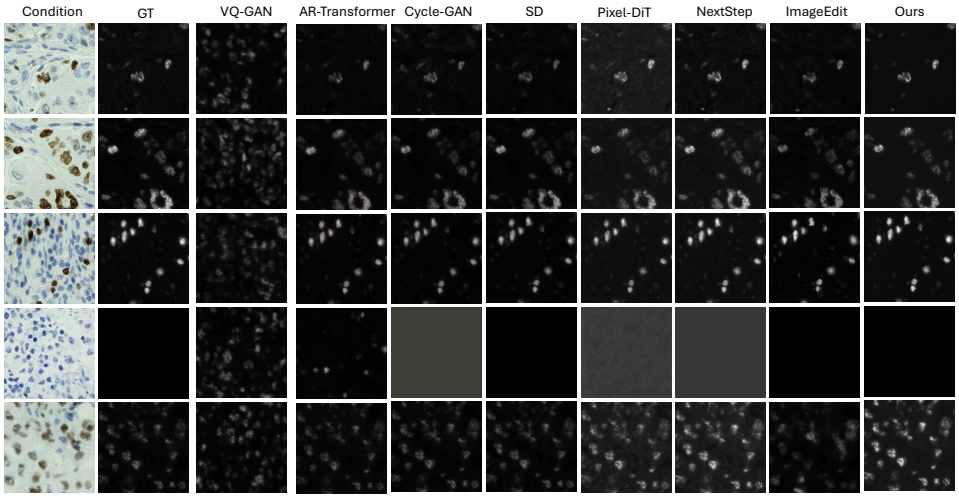


Figure 11: Visual comparisons on Virtual Staining(Immunohistochemistry2BCL)

or insufficient staining require the model to infer correct biological semantics rather than rely on superficial appearance.

B.3 Micro Image Brighten

Despite the significant distribution shift in the MicroBrighten dataset compared to other benchmarks, our model consistently delivers robust performance. The results(Fig.12, Tab. 5) demonstrate that our approach ensures structural consistency with the ground trutha factor that carries greater weight than perceptual metrics in this context.

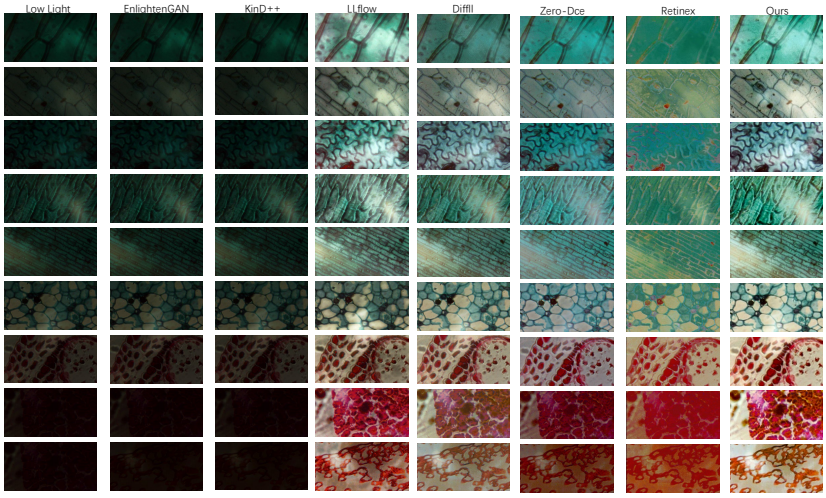


Figure 12: Visual comparisons on Low Light Enhance(Microscopic Section)

B.4 Micro Image Denoise

Over-reliance on ground truth (GT) is also a prevalent issue in denoising tasks, where existing data-driven end-to-end approaches often fail to achieve further performance gains. In contrast, our method maintains an optimal balance between perceptual and fidelity metrics. Furthermore, through consultation with biomedical experts (PhDs in Biomedical Science and Engineering), it was confirmed that the results in Fig. 13 demonstrate superior visual alignment without introducing artifacts or losing original information.

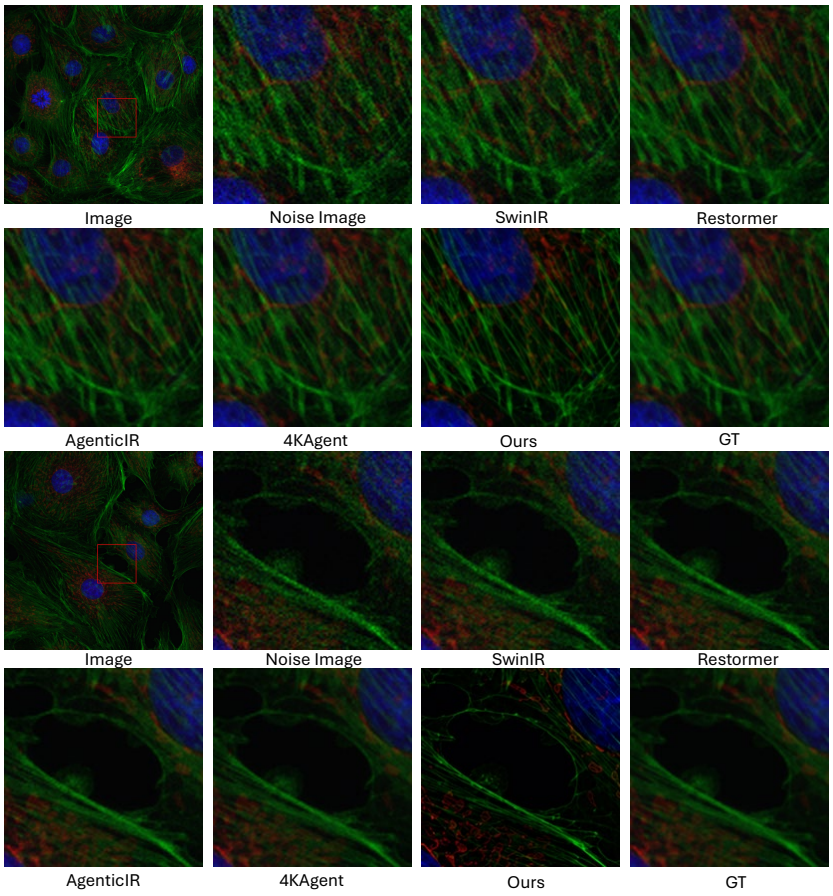


Figure 13: Visual comparisons on Virtual Staining(ConfocalBPAE, Zoom)

B.5 Micro Image Deblur

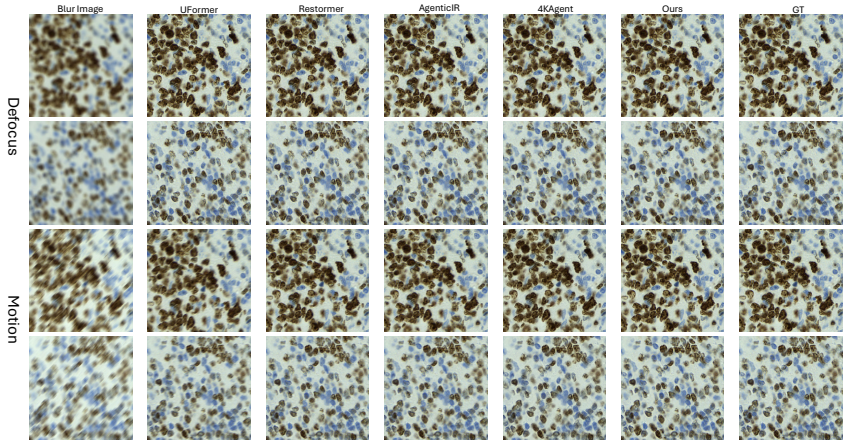


Figure 14: Visual comparisons on Deblur(Immunohistochemistry)

Comparatively, blur in microscopy is primarily driven by Poisson noise under low-photon conditions. When defocus or motion blur occurs, it typically stems from micro-vibrations, translation, or improper handling, and is generally less severe than the blur encountered in natural images. Consequently, we believe that simulated data can effectively support model training and enhance

generalization. Notably, many natural scene deblurring datasets, such as the GoPro [27], are also synthetically generated. Results in Fig 14 demonstrate the performance of our model.