E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中图法分类号:TP18; TP391　　文献标识码: A　　文章编号: 1006-8961(XXXX)XX-0001-36

论文引用格式:Zuo Jialong, Deng Haoyou, Zuo Haotong, Zhou Hanyu, Zhu Jiaxin, Zhang Yicheng, Zhang Yiwei, Yan Yongxin, Huang Kaixing, Chen Weisen, Deng Yongtai, Jin Rui, Sang Nong, Gao Changxin. Comprehensive Evaluation of Nano Banana Pro Based on 14 Tasks and 40 Datasets [J/OL]. Journal of Image and Graphics, XXXX: 1-36. DOI: 10. 11834/jig. 260029.(左嘉龙, 邓昊优, 左皓童, 周寒宇, 祝家心, 张祎程, 张艺巍, 严永昕, 黄凯星, 陈伟森, 邓永泰, 金睿, 张锋, 桑农, 高常鑫. Nano Banana Pro基于14项任务与40个数据集的综合评估[J/OL]. 中国图

Nano Banana Pro基于14项任务与40个数据集的综合评估

左嘉龙[1],邓昊优[1],左皓童[1],周寒宇[1],祝家心[1],张祎程[1],张艺巍[1],严永昕[1],黄凯星[1],陈伟森[1],邓永泰[1],金睿[1],张锋[2],桑农[1],高常鑫[1*]

1. 华中科技大学人工智能与自动化学院,湖北省武汉市　430074; 2. 大疆创新科技有限公司,广东省深圳市　518057

摘　要: 文本到图像生成模型的快速发展彻底改变了视觉内容创作。虽然诸如Nano Banana Pro之类的商业产品已获得广泛关注,但其作为传统底层视觉任务通用解决方案的潜力仍未得到充分探索。本文致力于解答一个核心问题:Nano Banana Pro是否是底层视觉全能选手? 通过零样本评估的方式,在涵盖40个多样化数据集的14个底层视觉任务上进行了全面测试。仅使用简单文本提示而未进行微调的情况下,将Nano Banana Pro与最先进的专用模型进行对比。深入分析揭示了明显的性能分野:尽管Nano Banana Pro展现出卓越的主观视觉质量,其“幻觉生成”的高频细节常超越专用模型,但在传统基于参考的定量指标上表现欠佳。本文将这种差异归因于生成模型固有的随机性,即难以满足传统指标对像素级一致性的严苛要求。本文肯定了Nano Banana Pro作为底层视觉任务零样本解决方案的潜力,同时指出要达到领域专用模型的高保真度仍面临重大挑战。

关键词: 底层视觉;生成式模型;文生图模型;Nano Banana Pro;综合性评测

Comprehensive Evaluation of Nano Banana Pro Based on 14 Tasks and 40 Datasets

Zuo Jialong[1], Deng Haoyou[1], Zuo Haotong[1], Zhou Hanyu[1], Zhu Jiaxin[1], Zhang Yicheng[1], Zhang Yiwei[1], Yan Yongxin[1], Huang Kaixing[1], Chen Weisen[1], Deng Yongtai[1], Jin Rui[1], Sang Nong, Gao Changxin[1]

1. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract: The rapid evolution of large-scale text-to-image generation models has fundamentally transformed the landscape of visual content creation. Driven by advances in diffusion models, large multimodal pretraining, and scalable inference pipelines, modern generative systems have demonstrated unprecedented capabilities in synthesizing visually compelling images across a wide range of styles, scenes, and semantic conditions. Commercial models such as Nano Banana Pro have attracted significant attention due to their strong zero-shot generation ability, robust semantic understanding, and impressive perceptual quality. However, despite their success in creative image synthesis, a critical and largely underexplored question remains: Can such foundation generative models serve as general-purpose solvers for traditional low-level vision tasks? Low-level vision tasks—including dehazing, deblurring, super-resolution and so on—have historically been domi-

收稿日期:2026-01-13;修回日期:2026-05-02

*通信作者:高常鑫(cgao@hust.edu.cn)

基金项目:国家自然科学基金(62176097);湖北省自然科学基金(2022CFA055)

Supported by: National Natural Science Foundation of China(62176097);Natural Science Foundation of Hubei Province, China(2022CFA055)

nated by task-specific, regression-based models. These models are typically trained under strong supervision with paired data and optimized using pixel-aligned objectives such as PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index measure). While highly effective within their target domains, such specialist models lack flexibility, often require costly retraining for new tasks, and struggle to generalize beyond their training distributions. In contrast, foundation generative models promise a unified alternative: a single pretrained model capable of addressing diverse vision tasks through natural language prompts, without task-specific fine-tuning. In this work, we present the first large-scale, systematic zero-shot evaluation of Nano Banana Pro across a broad spectrum of low-level vision tasks. Specifically, we investigate whether Nano Banana Pro can function as a low-level vision all-rounder—a generalist model capable of producing high-quality results across heterogeneous restoration, enhancement, and fusion tasks. To this end, we conduct an extensive evaluation covering 14 distinct low-level vision tasks across 40 datasets, encompassing both synthetic and real-world degradations. The evaluated tasks include deblurring (motion, defocus), super-resolution, image denoising, deraining, shadow removal, reflection removal, flare removal, low-light image enhancement, underwater image enhancement, HDR (high dynamic range) reconstruction, multi-focus image fusion, and infrared–visible image fusion, among others. All experiments are conducted under a standard zero-shot protocol. Nano Banana Pro is queried exclusively through simple, task-oriented natural language prompts, without any model fine-tuning, parameter adaptation, or task-specific post-processing. This setting is deliberately chosen to reflect realistic deployment scenarios and to assess the intrinsic capability of the model as a foundation visual system. For each task, we compare Nano Banana Pro against state-of-the-art specialist methods specifically designed for the corresponding task. Our comprehensive evaluation reveals a consistent and striking performance dichotomy. On one hand, Nano Banana Pro frequently produces results with superior perceptual quality, characterized by enhanced clarity, vivid textures, improved contrast, and visually pleasing color distributions. In many challenging scenarios—such as severe noise, extreme low-light conditions, heavy underwater color distortion, or strong atmospheric degradation—the model is able to hallucinate plausible high-frequency details and recover semantically coherent structures that rival or even surpass those generated by domain-specific methods. Across multiple tasks, Nano Banana Pro achieves competitive or leading performance on no-reference perceptual metrics and consistently receives favorable qualitative assessments. On the other hand, when evaluated using traditional full-reference, pixel-aligned quantitative metrics, Nano Banana Pro systematically underperforms compared to specialist models. Metrics such as PSNR, SSIM, SCD (sum of correlations of differences), and VIF (visual information fidelity) consistently reveal notable gaps, particularly in tasks requiring strict structural alignment or physical signal fidelity. This discrepancy is especially pronounced in tasks like denoising, HDR reconstruction, and image fusion, where pixel-level consistency with the reference image is heavily rewarded. We attribute this behavior to the inherent stochastic and generative nature of diffusion-based models, which prioritize semantic plausibility and perceptual realism over deterministic pixel correspondence. As a result, even visually improved outputs may be penalized for global color shifts, localized texture synthesis, or subtle geometric deviations. Importantly, our analysis shows that these quantitative penalties do not necessarily indicate failure. In many datasets, the provided "ground-truth" images themselves contain residual noise, blur, or imperfect color balance. In such cases, Nano Banana Pro often generates cleaner, more visually appealing results that deviate from the reference but align better with human perception. This observation highlights a fundamental tension between regression-based evaluation paradigms and generative reconstruction behaviors, and suggests that current benchmarks may be insufficient for assessing foundation generative models. Beyond aggregate metrics, we conduct detailed task-wise and dataset-wise analyses to characterize the operational scope and limitations of Nano Banana Pro. The model excels in scenarios involving severe degradation, ambiguous structure, or incomplete information, where its strong semantic priors can compensate for missing signal. Conversely, it struggles in applications demanding strict physical accuracy, such as forensic analysis, scientific imaging, or safety-critical perception, where hallucinated details or slight structural inconsistencies may be unacceptable. Collectively, our findings position Nano Banana Pro as a powerful zero-shot contender for low-level vision, capable of delivering high perceptual quality across a remarkably diverse set of tasks without retraining. At the same time, achieving the pixel-level fidelity of domain specialists remains a significant challenge. Rather than framing this as a binary competition between generative and regression paradigms, our results suggest a more promising direction: strategic integration. Future robust vision systems may

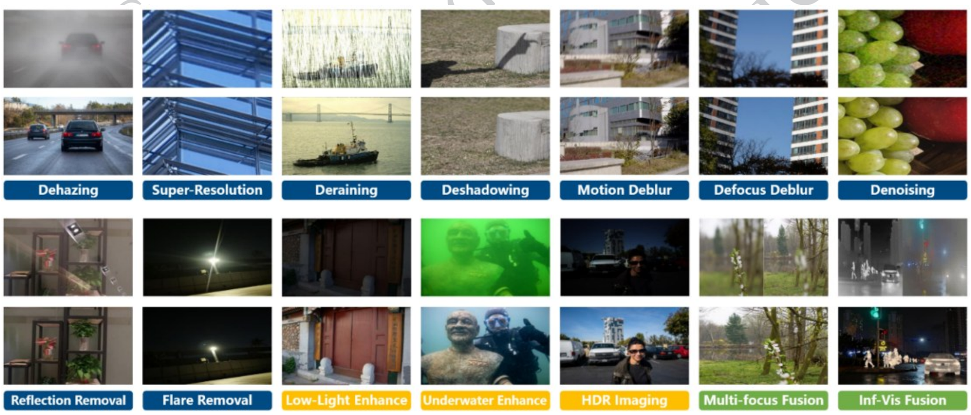
combine the semantic imagination of foundation generative models with the physical constraints and precision of task-specific networks, leveraging the strengths of both. In summary, this study provides the first comprehensive empirical answer to the question: Is Nano Banana Pro a low-level vision all-rounder? Our answer is nuanced. Nano Banana Pro substantially raises the upper bound of perceptual quality in zero-shot low-level vision, but has yet to establish a stable lower bound suitable for high-fidelity, safety-critical applications. By systematically documenting these strengths and limitations across 14 tasks and 40 datasets, this report offers a detailed reference point for future research on foundation models in low-level vision, and calls for the development of new evaluation frameworks that better reflect perceptual realism, semantic consistency, and downstream utility.

Key words: Low-Level Vision Tasks; Generative Models; Text-to-Image Models; Nano Banana Pro; Comprehensive Evaluation

论文引用格式:[DOI:10. 11834/jig. 260029]

0 引 言

近期,生成式人工智能显著改变了计算机视觉格局,Nano Banana Pro 等文生图模型在高保真内容创作中凭借多功能性脱颖而出,获得广泛关注。尽管其创造性合成能力广为人知,但此类大规模基础模型在传统底层视觉任务中的泛化能力尚未被充分探索。这一空白不仅揭示了能力层面的挑战,也提出了评估层面的难题,进而引出了核心研究问题:Nano Banana Pro是否堪称底层视觉全能选手?



本研究源于人类感知与传统评估指标间的根本矛盾。理论上,生成模型蕴含的丰富视觉先验使其能为图像修复、增强与融合生成合理的方案,但这可能与像素级精准保真度的指标要求相冲突。为此,我们仅使用简单提示词系统性地评估了 Nano Banana Pro 的零样本能力,并与需复杂微调的专用模型进行了全面的对比。

图 1 Nano Banana Pro 在 14 类底层视觉任务上的典型零样本结果示例。

Fig. 1 Exemplary zero-shot results of Nano Banana Pro across 14 low-level vision tasks.

注:对于每一种任务,第一行展示退化后的输入图像,第二行展示 Nano Banana Pro 在仅使用简单文本提示、无任何任务特定训练条件下生成的对应恢复结果。可视化结果表明,该模型在未进行专用训练的情况下,已展现出对多种底层视觉任务的初步泛化能力。蓝色边框表示图像恢复类任务,绿色边框表示图像增强类任务,黄色边框表示图像融合类任务。

我们的实证研究涵盖了3大领域、40个数据集上的14项底层视觉任务。如图1所示,Nano Banana Pro生成的图像感知质量卓越,在去雾、去雨等任务中展现出比专业模型更具美感的边缘与纹理。这一观察立即突显了一个关键挑战:模型可能主观表现优异,但在量化指标上处于劣势。因此,本工作旨在通过严谨的定量与定性分析,厘清其能力边界。需注意,由于未进行提示词调优或多轮推理筛选最优,本次评估仅是对该模型能力的保守估计。

研究结果揭示了预料之中的矛盾:Nano Banana Pro感知质量出色,但在PSNR、SSIM等基于参考的

指标上逊于领域专用模型。我们将这种差距归因于生成模型固有的随机性——模型更注重语义合理性,而非传统意义严格的像素级对齐。尽管 Nano Banana Pro 尚未成为完美的全能选手,但它促使我们重新思考底层视觉的定义。它为零样本图像修复建立了强力基准,并凸显了开发兼顾感知质量与像素精度的新型评估范式的紧迫性。

本报告余下部分将呈现图像修复、增强与融合领域的实验结果,并提供深入的对比分析。最后,我们将总结模型的局限性,并讨论包括为生成式底层视觉求解器开发感知特性评估方法在内的未来研究方向。

1 图像去雾

1.1 引言

真实世界图像去雾(real-world image dehazing,RID)旨在从实拍雾气观测图像中恢复清晰且高保真结果。不同于基于物理假设的合成去雾,真实雾气具有空间非均匀性、深度变化大、色彩偏移及噪声伪影等复杂特性,这使其成为极具挑战性的底层视觉难题。其目标不仅是优化视觉效果,更需保留精确的光度学和结构信息以支持检测、跟踪与分割等可靠的下游视觉任务。

早期去雾方法主要依赖暗通道(He 等,2010)、非局部先验(Berman 等,2016)等统计先验及其扩展(Qiu 等,2023;Song 等,2023),但泛化能力有限且易引入伪影。随着深度学习的发展,基于 CNN(convolutional neural networks)与 Transformer 的方法(Chen 等,2009;Shao 等,2020;Wu 等,2023)显著提升了去雾性能,但受限于大规模真实配对数据的缺乏,多数方法在真实场景中仍表现不足。为此,近期研究开始转向真实场景去雾,通过引入物理先验或改进推理策略提升泛化能力,但在浓雾等信息严重缺失场景下,传统方法仍难以有效恢复受损内容。

表 1 多数据集上去雾方法的定量对比结果

Table 1 Quantitative comparison of dehazing methods on multiple datasets

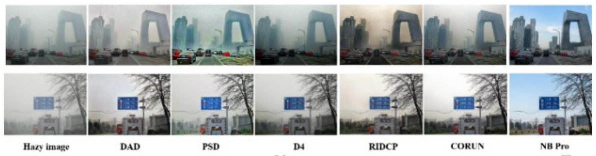
方法	RTTS			Fattal′s		
	FADE↓	BRISQUE↓	NIMA↑	FADE↓	BRISQUE↓	NIMA↑
MBSDN (Dong 等,2014)	1.363	27.67	4.53	0.579	**14.15**	5.43
Dehamer (Guo 等,2022)	1.895	33.24	4.52	0.698	15.53	5.16
DAD(Shao 等,2020)	1.130	32.24	4.31	0.484	29.64	**5.46**
PSD(Chen 等,2021)	0.920	27.71	4.60	0.416	23.61	4.99
D4(Yang 等,2022)	1.358	33.21	4.48	0.411	20.33	5.44
RIDCP(Wu 等,2023)	0.944	17.29	4.97	0.408	20.05	5.43
CORUN(Fang 等,2024)	**0.824**	**11.96**	**5.34**	**0.338**	14.82	5.39
NB Pro	0.986	27.21	4.95	0.683	22.16	5.44

1.2 定量与定性结果

本文评估 NB Pro 在 RTTS(Li 等,2018)和 Fattal′s(Fattal 等,2014)数据集上的性能,并与真实世界去雾任务中的最先进基线方法进行对比。表 1 列出了基于 FADE (fog aware density evaluator)、BRISQUE (blind/referenceless image spatial quality evaluator)和 NIMA(neural image assessment)指标的定量结果。整体而言,NB Pro 在主观视觉质量方面表现出色,在两个基准测试上均获得优异的 NIMA 分数,展现出优异的美学质量和感知效果。但在图像自然度方面表现欠佳,尤其在 Fattal′s 数据集上的 BRISQUE 分数较高,表明存在过度增强导致的伪影。

定性实验显示,NB Pro 能有效提升极端退化场景(如浓雾、强散射)的感知质量。图 2 展示了 NB Pro 与其他方法在 RTTS 数据集上的视觉对比,其借强大的生成能力恢复了远景建筑与自然环境被遮蔽的细节。

然而,NB Pro 也存在失败案例(图 3)。在中浓雾条件下,模型常产生失真输出,表现为不自然的色



可以观察到，NB Pro生成的结果在整体清晰度和视觉质量方面更具优势；但与此同时，其结果中也存在较为明显的过度增强伪影问题。

图2　NB Pro与其他方法的可视化结果对比

Fig. 2　A comparison of visual results between NB pro and other methods

彩偏移（如过饱和）以及产生幻觉元素（如将阴天场景强制渲染为蓝天）。这些失真破坏了大气真实性，导致结果偏离了真实去雾的预期。

综上所述，尽管NB Pro在不适定的真实世界去雾问题中展现了具有启发性的生成能力，展示了语义驱动先验在处理模糊退化中的潜力，但在色彩保真度、物理真实感及处理一致性上仍有局限。这表明它目前更适用于创意增强而非精确的底层复原。图3 NB Pro 在 RTTS 数据集上的去雾变形示例图像

Fig. 3 Anamorphic example images of Nano Pro in dehazing on the RTTS dataset

1.3　分析

由于训练分布与恢复目标不一致，NB Pro在去雾任务中难以保持光谱保真度与大气一致性，常将原本自然朦胧的色调增强为过饱和结果，甚至在阴天场景中生成蓝天等幻觉细节。其本质上并非求解逆物理散射过程，而是依赖生成先验对缺失信息进行语义补全。这种策略虽能提升主观视觉吸引力，却会破坏场景真实性与时间一致性。

因此，NB Pro更适合作为创意增强工具，而非强调严格物理约束的精确复原方法。尽管如此，在真实世界去雾这一高度不适定任务中，它仍展示了利用语义先验重建合理细节的潜力。

2　图像超分

2.1　引言

真实世界图像超分辨率旨在从复杂退化的低分辨率输入中恢复高保真、高分辨率内容。与基于简单下采样假设的合成超分辨率不同，真实场景通常伴随模糊、噪声、压缩伪影及相机参数变化等复合退化（Cai 等，2019；Wei 等，2020），使依赖逐像素优化的回归方法易出现纹理平滑和高频细节缺失（Dong 等，2014；Zhang 等，2020）。

近年来，生成式真实世界超分辨率发展迅速。GAN（generative adversarial network）方法如BSRGAN（blind super-resolution generative adversarial network）和Real-ESRGAN（real-enhanced super-resolution generative adversarial network）通过高阶退化建模提升了感知质量（Goodfellow 等，2014；Ledig 等，2017）；随后，DDPMs（denoising diffusion probabilistic models）（Ho 等，2020）及StableSR（Wang 等，2024）、DiffBIR（diffusion-based blind image restoration）（Lin 等，2023）等方法借助预训练文生图模型先验（Rombach 等，2022）进一步增强了纹理生成能力。为降低推理开销，又出现了SinSR（single-step super-resolution）等单步加速方法。然而，感知质量与保真度之间的权衡（Blau 等，2018）仍是该任务的核心问题。

基于此，本文对Nano Banana Pro进行了定量与定性评估，并与GAN方法、多步扩散方法及加速扩散方法进行比较。实验在DIV2K-Val（Agustsson 等，2017）、RealSR（Cai 等，2019）和DRealSR（Wei 等，2020）上开展，并结合全参考指标（Wang 等，2004；Zhang 等，2018）与无参考感知指标（Ke 等，2021；Mittal 等，2012；Wang 等，2023）系统评估其重建能力。

2.2　定量结果

本文将Nano Banana Pro与先进的基于GAN及基于扩散的图像超分辨率方法进行定量比较（见表2）。采用标准全参考指标：信号保真度（PSNR、SSIM）、感知相似度LPIPS（learned perceptual image patch similarity）。此外，还使用无参考指标NIQE（naturalness image quality evaluator）、MUSIQ（multi-scale image quality transformer）、CLIPIQA（clip image quality assessment）来量化生成图像的统计自然度和美学质量。结果如表2所示。

表 2 在合成数据集(DIV2K-Val)与真实场景数据集(RealSR 、DRealSR)上的定量对比结果

Table 2 Quantitative comparison on synthetic (DIV2K-Val) and real-world (RealSR, DRealSR) benchmarks.

测试数据集	方法	Full-Reference Metrics			No-Reference Metrics		
		PSNR↑	SSIM↑	LPIPS↓	NIQE↓	MUSIQ↑	CLIPIQA↑
DIV2K-Val	BSRGAN(Zhang 等,2021)	**24.58**	<u>0.6269</u>	0.3351	4.75	61.20	0.5071
	Real-ESRGAN (Wang 等,2021)	24.29	**0.6371**	**0.3112**	<u>4.68</u>	61.06	0.5501
	StableSR(Wang 等,2024)	23.26	0.5726	<u>0.3113</u>	4.76	**65.92**	<u>0.6192</u>
	DiffBIR(Lin 等,2023)	23.64	0.5647	0.3524	4.70	<u>65.81</u>	**0.6210**
	SinSR (Wang 等,2024)	<u>24.41</u>	0.6018	0.3240	6.02	62.82	0.5386
	Nano Banana Pro	20.29	0.4720	0.3645	**3.52**	65.40	0.5257
RealSR	BSRGAN	**26.39**	**0.7654**	**0.2670**	5.66	63.21	0.5001
	Real-ESRGAN	25.69	<u>0.7616</u>	<u>0.2727</u>	5.83	60.18	0.4449
	StableSR	24.70	0.7085	0.3018	5.91	**65.78**	<u>0.6221</u>
	DiffBIR	24.75	0.6567	0.3636	<u>5.53</u>	<u>64.98</u>	**0.6246**
	SinSR	<u>26.28</u>	0.7347	0.3188	6.29	60.80	0.5385
	Nano Banana Pro	23.56	0.6649	0.2978	**4.39**	60.18	0.5199
DRealSR	BSRGAN	**28.75**	<u>0.8031</u>	<u>0.2883</u>	6.52	57.14	0.4915
	Real-ESRGAN	<u>28.64</u>	**0.8053**	**0.2847**	6.69	54.18	0.4422
	StableSR	28.03	0.7536	0.3284	6.52	58.51	<u>0.5601</u>
	DiffBIR	26.71	0.6571	0.4557	<u>6.31</u>	**61.07**	**0.5930**
	SinSR	28.36	0.7515	0.3665	6.99	55.33	0.4884
	Nano Banana Pro	23.97	0.6323	0.3809	**5.03**	<u>59.00</u>	0.5145

注:表中黑色加粗和下划线分别表示最优和次优结果。

在传统保真度指标上,Nano Banana Pro 表现显著落后于其他方法。其在 DIV2K-Val 上的 PSNR 和 SSIM 显著低于最优方法,在 RealSR 和 DRealSR 上,其保真度分数也始终落后于基于 GAN 和基于扩散的基线模型。这表明在追求像素级重建的标准全参考框架下,其生成结果与真实参考图像存在系统性偏差。

Nano Banana Pro 与针对特定退化核优化的传统超分辨率模型有根本区别。后者通常经过端到端训练,以最小化像素级损失为目标,因此在 PSNR 和 SSIM 等指标上天生表现出色。相比之下,Nano Banana Pro 的生成过程优先考虑语义连贯性和视觉纯净度。其输出可视为对输入图像的合理重建,而非严格的像素到像素映射。因此,生成图像可能在局部纹理对齐和结构定位上与参考图像存在偏差,导致全参考指标得分全面下降。然而值得注意的是,在 NIQE 指标上,Nano Banana Pro 在全部三个数据集上始终取得了最佳分数(例如在 DIV2K-Val 上为 3.52,而 BSRGAN 为 4.75)。这表明其输出具有卓越的统计自然度,能有效抑制伪影。

2.3 定性结果

本节考察 Nano Banana Pro 在 DIV2K、RealSR 和 DRealSR 数据集上的生成表现。定性评估围绕四个关键场景组织,以突出模型的优势和失败模式:几何清晰度、视场角伪影、纹理保真度和字符重建。

在成功的结果中,Nano Banana Pro 在具有明显几何结构的场景中能够有效锐化模糊边缘、恢复线性纹理,并提升整体清晰度。

图 4 展示了 Nano Banana Pro 中观察到的一种显著结构性异常现象:非预期的视场(Field-of-View, FOV)扩展。通过对比可以发现,模型未能严格遵循输入图像的原始空间边界,而是在图像边缘区域错误地臆造了额外内容,从而导致视场范围的非真实扩展。

图4　在 Real-ISR 任务中出现的非预期图像边界扩展现象的可视化示例

Fig. 4　Visualization of unintended image boundary extension cases in Real-ISR tasks using Nano Banana Pro

生成结果中，还出现了石材表面的颗粒结构在重建过程中被赋予了不同的局部结构形态，而非被忠实还原的现象。体现 Nano Banana Pro 在处理复杂随机纹理区域时，倾向于合成看似合理但实际上并不存在的细节。而这在像素级别上引入了不一致性，导致保真度指标的下降。

图5进一步表明，该模型在文本重建上高度依赖输入中的语义可识别性；当原始字形仍保留一定结构线索时，模型通常能够较准确地恢复字符，但在严重退化、尤其是中文字符结构受损时，容易产生错误笔画或臆造字符，导致结果虽更清晰，却在语义上明显失真。

2.4　分析

本节的综合评估阐明了 Nano Banana Pro 在真实世界超分辨率领域内的运行特性。在定量分析中，我们发现该模型优先利用学习到的生成先验来合成纹理细节，而非严格遵循低分辨率输入。

在定性分析中，我们发现 Nano Banana Pro 缺乏与参考图像的精确像素级对齐。因此，我们观察到了非预期的视场扩展以及失败的文本重建现象。

综合这些发现，Nano Banana Pro 非常适合以感知为中心的应用，例如艺术放大、老照片修复或休闲摄影。在这些场景中，视觉纯净度和噪声消除至关重要。然而，由于其倾向于纹理幻觉、空间错位和语义改变，它不适用于对保真度要求苛刻的场景。

图5 文本重建效果可视化

Fig. 5 Visualization of text reconstruction results

3　图像去雨

3.1　引言

降降雨会遮蔽场景内容并破坏图像结构，显著影响交通、监控和自动驾驶等户外视觉系统的可靠性，因此单幅图像去雨是底层视觉中的重要任务。近年来，去雨算法和基准数据集不断发展，在多种降雨退化建模方面取得了明显进展。本文探讨 Nano Banana Pro 在单幅图像去雨任务中的应用潜力，重点评估其在多样场景下去除雨纹、恢复背景结构与保持视觉自然性方面的表现。

3.2　实验设置

为了全面评估 Nano Banana Pro 在单幅图像去雨任务上的性能和泛化能力，本节在三个广泛使用的基准数据集上进行了实验：两个合成数据集 Rain200L（Yang 等，2017）和 Rain200H（Yang 等，2017），以及一个真实世界数据集 SPA（spatial attentive）（Wang 等，2019）。这些数据集涵盖了广泛的雨水模式和场景复杂度，能够全面评估模型的复原能力。

3.3　定量与定性结果

为了公平比较，所有指标均在与 NeRD-Rain 完全相同的评估协议下计算，包括图像分辨率、色彩空间和 PSNR/SSIM 计算。根据表3报告的定量结果，我们系统地评估了 Nano Banana Pro 在两个合成数据集以及真实世界数据集上的图像去雨性能，并将其与多种代表性方法进行比较。对比方法涵盖了基于先验的方法、基于 CNN 的方法以及基于 Transformer 的方法。

表3　三个代表性基准上的定量对比结果

Table 3　Quantitative comparison results on three representative benchmarks

方法	Rain200L		Rain200H		SPA-Data	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DSC（Luo等，2015）	27.16	0.8663	14.73	0.3815	34.95	0.9416
GMM（Blau等，2018）	28.66	0.8652	14.50	0.4164	34.30	0.9428
DDN（Fu等，2017）	34.68	0.9671	26.05	0.8056	36.16	0.9457
RESCAN（Li等，2018）	36.09	0.9697	26.75	0.8353	38.11	0.9707
PReNet（Ren等，2019）	37.80	0.9814	29.04	0.8991	40.16	0.9816
MSPFN（Jiang等，2021）	38.58	0.9827	29.36	0.9034	43.43	0.9843
RCDNet（Wang等，2020）	39.17	0.9885	30.24	0.9048	43.36	0.9831
MPRNet（Zamir等，2021）	39.47	0.9825	30.67	0.9110	43.64	0.9844
DualGCN（Fu等，2017）	40.73	0.9886	31.15	0.9125	44.18	0.9902
SPDNet（Yi等，2021）	40.50	0.9875	31.28	0.9207	43.20	0.9871
Uformer（Wang等，2022）	40.20	0.9860	30.80	0.9105	46.13	0.9913
Restormer（Zamir等，2021）	40.99	0.9890	32.00	0.9329	47.98	0.9921
IDT（Xiao等，2022）	40.74	0.9884	32.10	0.9344	47.35	0.9930
DRSformer（Chen等，2023）	41.23	0.9894	32.17	0.9326	48.54	0.9924
NeRD-Rain（Chen等，2024）	**41.71**	**0.9903**	**32.40**	**0.9373**	**49.58**	**0.9940**
Nano Banana Pro	26.05	0.7954	21.10	0.6659	32.25	0.9142

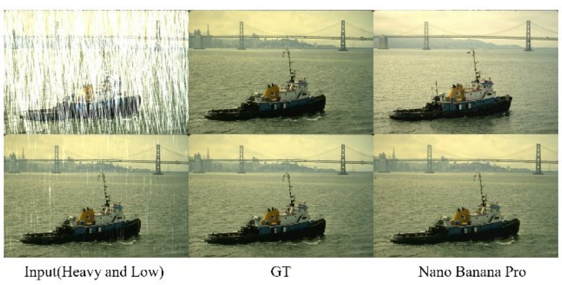


图6　在Rain200H数据集上的定性对比结果

Fig. 6　Qualitative comparison on the Rain200H dataset

定量结果显示（见表3），NB Pro在三个数据集上的性能均显著低于最先进的去雨模型。在Rain200L上，其取得了26.05 dB的PSNR和0.7954的SSIM，远低于基于监督学习的方法；在具有复杂雨水模式且更具挑战性的Rain200H上，它获得了21.10 dB的PSNR和0.6659的SSIM。尽管它优于传统的基于先验的方法，但与基于深度学习的方法相比，仍然存在相当大的性能差距。在真实世界数据SPA-Data，其达到32.25 dB的PSNR和0.9142的SSIM，仍然显著低于当前表现最好的方法。

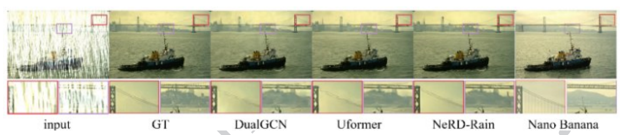


图7　不同降雨强度条件下的定性对比结果

Fig. 7　Qualitative comparison under different rain intensities

这种定量指标的下降与视觉结果（图7）一致：大而密集的合成雨线严重遮蔽背景内容，导致色彩偏差与细节的缺失或过度平滑影响了得分。由于NB Pro并非专门为图像去雨任务训练，局部结构通常侧重于隐式生成“幻觉”而非逐像素复原，因此其PSNR和SSIM受到负面影响。但在恢复全局结构方面，它展现出比监督基线更好的语义合理性。

此外，模型性能对雨量较为敏感：低雨量条件下，更多有效视觉信息得以保留，因此色彩和细节恢复相对较好；而强降雨带来的严重遮挡和模糊则更易引发色彩偏移与细节退化。总体来看，尽管生成式模型在零样本去雨中的像素级保真度较弱，但其语义与结构先验在信息严重缺失场景下仍具有一定潜力。

模型在去除雨纹的同时，消除了伴随存在的大气雾霾，尽管提升了图像清晰度，但由于结果与GT存在偏离，导致定量评价指标下降。

图8　Nano Banana Pro的失败案例

Fig. 8　Failure case of Nano Banana Pro

定性分析还表明，基于提示词的多模态生成模型在指令遵循上存在局限。即使提示明确，背景内容仍可能发生非预期变化。其原因之一是模型容易将雨线与薄雾混淆，在去雨过程中出现过度除雾，即细粒度语义理解的固有偏差（图8）。虽然这种结果在主观上更清晰，但会与真实场景产生较大偏差，从而拉低PSNR等量化指标。

4　图像去阴影

4.1　背景

阴影去除旨在消除投射阴影，恢复图像中阴影区与非阴影区的光照一致性（Guo等，2012）。由于阴影会带来强度不连续、色彩失真和纹理细节丢失，并影响检测与分割等下游任务，因此它是一项重要且具有挑战性的底层视觉任务。

早期方法（Qu等，2017）多依赖手工先验和统计特征，但在复杂光照、纹理背景和柔和阴影边界场景中泛化能力有限，容易产生伪影。随后，基于CNN的监督方法虽显著提升了性能，但仍依赖昂贵标注，且在真实场景中存在过拟合风险。本节系统评估Nano Banana Pro在单幅图像阴影去除任务中的表现，并分析其在真实应用中的优势与局限。

4.2　定量与定性结果

图9显示NB Pro在SRD（shadow removal dataset）数据集上表现良好的阴影去除结果，其能够有效去除阴影并高度保留原始元素。表4列出了其与SOTA（state of the art）方法的定量对比，使用PSNR和SSIM作为主要评估指标。

如表4显示，视觉感知与量化分数存在显著差异。ShadowDiffusion（Guo等，2023）和HomoFormer（Xiao等，2024）等领先方法获得了超过34dB的PSNR和0.97以上的SSIM值，而NB Pro的PSNR仅为20.67 dB，SSIM仅为0.682。这种差距可以归因于生成模型的特性：NB Pro优先考虑感知自然度而非像素级对齐，倾向于生成具有视觉吸引力但偏离真实像素值的图像，导致保真度指标得分较低。

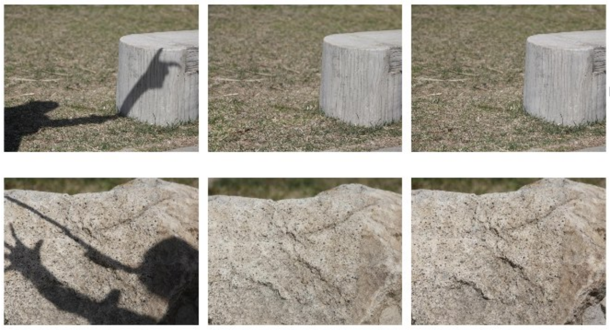


图9　在SRD数据集上，NB Pro在阴影去除任务中的若干表现良好的可视化示例

Fig. 9　Some well-performing visual examples of NB Pro on the SRD dataset for the shadow removal task

图10总结了NB Pro的两类典型失败模式。这些例子揭示了源于模型生成性质的两个核心且反复出现的局限性：过度生成的倾向和对细微阴影的“盲区”。

表4　在SRD数据集上的定量对比结果

Table 4　Quantitative comparisons on the SRD dataset

方法	PSNR ↑	SSIM ↑
DSC（Hu等，2018）	27.76	0.903
DHAN（Cun等，2020）	30.51	0.949
BMNet（Zhu等，2022）	31.69	0.956
ShadowFormer（Guo等，2023）	32.90	0.958
ShadowDiffusion（Guo等，2023）	34.73	0.970
HomoFormer（Xiao等，2024）	**35.37**	**0.972**
Nano Banana Pro	20.67	0.682

第一种失败模式源于模型固有的生成偏差。在生成视觉上“完整”场景的驱动力下，NB Pro常常优先考虑幻觉生成的内容而非保真度。如图10所示，其在去除阴影后错误合成新手部以填充空白，损害了语义完整性，凸显了创造性与底层视觉保真度之间的矛盾。

第二种失败模式暴露了模型在阴影检测敏感度不足,常忽略柔和或低对比度的微弱阴影。这表明训练数据或优化目标可能存在偏差,导致其对细微光照变化的权重过低。此外,模型在色彩保真度上也存在困难,常表现出色调与饱和度偏移。这些结构性幻觉、检测失效及色彩偏移共同导致了其量化表现不佳。

图 10　NB Pro 在 SRD 数据集上阴影去除任务中的若干可视化失败示例

Fig. 10　Some visual failure examples of NB Pro on the SRD dataset for the shadow removal task

5　运动模糊去除

5.1　背景

相机抖动或物体快速移动引起的运动模糊是普遍的图像退化现象。过去十年,深度学习彻底改变了动态场景去模糊领域:早期 CNN(如 DeepDeblur (Nah 等,2017)与 DeblurGAN(Kupyn 等,2018))为复杂架构奠定了基础,近期 Transformer 模型(如 Uformer (Wang 等,2022) 与 Restormer (Zamir 等,2022))则通过长程依赖建模取得了极高的 PSNR 分数,并占据了主导部分。然而,这些基于回归的方法常因使用 MSE 损失而导致结果过度平滑,为了最小化像素误差牺牲了高频纹理。

为恢复细节,研究焦点转向了 GAN(Wang 等,2026)和扩散模型(Chen 等,2023)。这些方法利用强大的生成先验合成合理纹理,显著提升了感知质量,但也带来了感知-失真权衡的挑战:在追求视觉清晰度时,模型可能偏离真值保真度,产生伪影或内容幻觉。

本报告对 Nano Banana Pro 进行了全面评估,揭示了该激进生成方法的根本性局限。尽管 NB Pro 擅长在静态环境或低光场景中合成锐利纹理,但存在严重的语义不稳定性。

5.2　定性结果

NB Pro 在合成数据集上展现了出色的去模糊能力,尤其在静态细节恢复上。如图 11 所示,模型有效抑制了 GoPro 与 HIDE(human-aware image deblurring)中的严重模糊,以高精度重建了建筑外立面及人行道纹理。此外,模型在文本保留也发挥出色,如"SEPHORA"被清晰渲染出来。这表明其在处理刚性运动和结构边缘处理方面展现出较强的空间适应性。

图 11　Nano Banana Pro 在合成模糊数据集(GoPro 与 HIDE)上的可视化结果。

Fig. 11　Visual results of Nano Banana Pro on synthetic blur datasets (GoPro and HIDE)

然而,处理涉及人体的高度动态场景时模型局限性明显。模型难以完全消除复杂的运动轨迹,导致衣物及头巾出现残留重影。同时,模型倾向于"幻觉生成"面部细节,虽主观视觉上清晰但存在明显的语义不一致,导致行人身份与真值严重不符,暴露出保真度缺失的缺点。

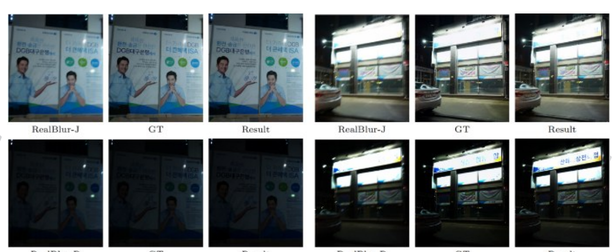


图 12　Nano Banana Pro 在真实场景模糊数据集(RealBlur-J 与 RealBlur-R)上的可视化结果

Fig. 12　Visual results of Nano Banana Pro on real-world blur dataset (RealBlur-J and RealBlur-R)

在真实数据集上,NB Pro 对低光及曝光过度展现出强鲁棒性。如图 12 所示,在 RealBlur-J 数据集中,模型成功提升了海报与招牌文本的可读性。其处理高动态范围场景的能力尤其值得注意,且能够

有效处理高对比度光照。但在RealBlur-R第二列中，其生成结果显著偏离真值：虽画面更洁净，但过度去噪及光照纹理的改变破坏了场景原有的氛围。由于依赖生成先验，模型产生了与真实情况之间明显的感知偏差。RealBlur-J海报示例中，恢复的人脸与原始图像不同；RealBlur-J第二列与RealBlur-R第二列中的文本字符也出现了偏离真值的生成错误。此外，其还原结果色彩保真度偶尔也会受损，如RealBlur-R第一列中出现了人物肤色的色彩偏移。这些现象表明，虽然NB Pro擅长产生令人愉悦的结果，但是却牺牲了对原始语义内容的忠实度。

5.3 定量结果

表5展示了NB Pro与最先进去模糊方法的定量比较，涵盖了Transformer模型及近期扩散模型。评估在GoPro（Nah等，2017）、HIDE（Shen等，2019）和RealBlur（Rim等，2020）这三个标准基准上进行，主要指标为PSNR和SSIM。

如表5所示，模型视觉清晰度与数值保真度之间存在显著差距。ID-CDM（image deblurring with consistency distillation models）和HI-Diff（hierarchical integration diffusion model）等表现最好的方法于GoPro上取得超33dB的PSNR分数、RealBlur-R上取得了超36dB的PSNR分数，而NBPro的PSNR较低，在GoPro上为21.41 dB，在HIDE上为21.35 dB。同样，NB Pro的SSIM分数处于0.645至0.778之间，也远低于竞争方法的0.90以上。这种定量差距源于模型对强生成先验的依赖，即优先考虑感知合理性而非与真值的严格像素级对齐。

较低的指标分值证实了定性分析中的局限性。PSNR对像素级偏差高度敏感，而NB Pro倾向于通过“幻觉”生成高频细节（如改变人脸身份或修改字符）来提升清晰度。这些生成的特征虽视觉连贯，但相对于参考图像来说，就是一种计算误差，会导致信噪比大幅下降。此外，模型生成过程中的语义不一致性，如色彩偏移及运动轨迹误解引起的“重影”效应，严重破坏了生成结果的结构相似性，导致SSIM下降。这进一步证明了模型生成的内容往往偏离底层的真实信号。

6 散焦模糊去除

6.1 引言

散焦模糊源于有限景深与光圈配置，是一种典型的空间变化退化。由于其点扩散函数会随场景深度变化，常导致高频细节和边缘信息的非均匀丢失，使散焦去模糊成为具有挑战性的逆问题。

近年来，单幅图像散焦去模糊取得了显著进展，相关方法通过双像素监督（Abuolaim等，2020）、迭代滤波与核预测（Lee等，2021；Son等，2021）、Transformer建模（Zamir等，2021；Mehri等，2021）以及更精细的核建模策略（Quan等，2024）不断提升复原性能。

基于此，本节评估Nano Banana Pro在散焦去模糊任务中的表现，并在DPDD（dual-pixel defocus deblurring）（Abuolaim 等，2022）和RealDOF（real depth of field）（Ruan等，2021）数据集上测试其在未经微调下处理该类逆问题的有效性，以分析其是真正恢复结构信息，还是主要依赖表面增强。

6.2 定量结果

表6展示了模型在DPDD和RealDOF数据集上的评估结果。结果表明，Nano Banana Pro与成熟的散焦去模糊方法差距显著：在DPDD上，其PSNR和SSIM落后SOTA模型GGKMNet（grouped gaussian kernel mixture network）6dB以上；类似的不足在RealDOF上也显而易见，甚至逊于DPDNet（dual-pixel defocus deblurring network）等早期基线。这些较低指标与我们的定性观察一致：低PSNR反映了像素级清晰度的缺失，而低SSIM则证实了结构幻觉及去模糊效果的不足。

6.3 定性结果

本节对DPDD和RealDOF数据集的视觉分析结果表明，NB Pro在从严重散焦中恢复高频细节的能力方面存在局限，其常常优先进行全局对比度增强而非有效除模糊。

在DPDD上，NB Pro更接近图像增强滤镜，而非专门的去模糊网络。如图13所示，模型对输入的主要修改是全局增加亮度和对比度，虽然增强了视觉效果，但未能解决根本退化。具体而言，案例2的严重失焦的前景依然模糊，案例1背景焦外成像范围仅在扩散程度上略微缩小。此外，模型在结构重建

表 5　Nano Banana Pro 与其他代表性方法在四个基准数据集上的定量对比结果

Table 5　Quantitative comparison results of Nano Banana Pro and other representative methods on four benchmarks

方法	GoPro		RealBlur-R		RealBlur-J		HIDE	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DeepDeblur（Nah 等，2017）	29.08	0.914	32.51	0.841	27.87	0.827	25.73	0.874
GAMD（Luan 等，2024）	33.14	0.9284	34.00	0.9265	–	–	–	–
DeblurGAN（Kupyn 等，2018）	28.70	0.858	33.79	0.903	27.97	0.834	24.51	0.871
DeblurGAN-v2（Kupyn 等，2019）	29.55	0.934	35.26	0.944	28.70	0.866	26.61	0.875
DBGAN（Zhang 等，2020）	31.10	0.942	33.78	0.909	24.93	0.745	28.94	0.915
Uformer-B（Wang 等，2022）	32.97	0.967	36.22	0.957	29.06	0.884	30.83	**0.952**
Stripformer（Tasi 等，2022）	33.08	0.962	36.08	0.954	28.82	0.876	31.03	0.940
Restormer（Zamir 等，2021）	32.92	0.961	36.19	0.957	28.96	0.879	31.22	0.942
IR-SDE（Luo 等，2023）	30.70	0.901	33.96	0.918	24.21	0.729	–	–
DiffIR（Xia 等，2023）	33.20	0.963	–	–	–	–	31.55	0.947
HI-Diff（Chen 等，2023）	**33.33**	0.964	36.28	**0.958**	**29.15**	**0.890**	31.46	0.945
ID-CDM（Wang 等，2026）	33.19	**0.970**	**36.34**	0.955	28.96	0.887	**31.53**	0.950
Nano Banana Pro	21.41	0.645	27.43	0.778	24.51	0.747	21.35	0.662

方面具有不稳定性：案例3显示模型产生了语义幻觉，尽管清晰度有所改善，但是却将“GE CANADA”误恢复为“OE CANADA”；案例4中，前景围栏得到充分锐化，但是却牺牲了几何保真度，导致背景车辆比例异常。

在RealDOF上，其局限性更明显，去模糊效果微乎其微。输出的结果会出现前景的散焦未能有效逆转，模糊区域在感知上与输入无异。虽然有些结果的锐度有所恢复，但却是以引入建筑物上的椒盐噪声为代价。有时模型也会受限于景深，而仅仅恢复最靠近镜头的地面纹理，而背景仍然处于散焦状态，说明NB Pro缺乏鲁棒的散焦去模糊能力。

6.4 分析

定量与视觉结果的差异表明，NB Pro在散焦去模糊中更依赖语义先验生成细节，而非严格反演散焦过程，因此在物理约束下稳定性有限。

一方面，在结构线索模糊、深度变化复杂的场景中，模型更多通过提升对比度增强观感，但这并不能真正恢复散焦信息；当场景结构较规则时，虽能重建部分细节，却也容易引入伪影和结构偏差。另一方面，当模糊更接近正常景深效果时，NB Pro往往倾向于保留背景模糊，仅锐化前景。

总体来看，在零样本且缺乏保约束的条件下，NB Pro更接近图像重合成模型，而非严格意义上的专用散焦复原工具，这也是其定量表现偏低的重要原因。

表6 在DPDD和RealDOF数据集上的定量对比结果

Table 6 Quantitative comparison on DPDD and RealDOF

方法	DPDD		RealDOF	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DPDNet（Abuolaim等，2020）	24.348	0.747	22.870	0.670
AIFNet（Ruan等，2021）	24.213	0.742	23.093	0.680
IFANet（Lee等，2021）	25.366	0.789	24.712	0.748
KPAC（Son等，2021）	25.221	0.774	23.975	0.762
GKMNet（Quan等，2021）	25.468	0.789	24.254	0.732
MDP（Abuolaim等，2022）	25.347	0.763	23.500	0.681
DRBNet（Ruan等，2022）	25.485	0.792	24.884	0.751
MPRNet（Mehri等，2021）	25.730	0.792	24.541	0.736
Restormer（Zamir，2022）	25.980	**0.811**	25.091	0.762
INIKNet（Quan等，2021）	26.055	0.803	25.231	0.765
NRKNet（Quan等，2023）	26.109	0.810	25.148	0.768
GGKMNet（Quan等，2024）	**26.272**	0.810	**25.355**	**0.770**
NB Pro	20.180	0.635	20.821	0.641

注：最佳结果以黑色加粗标

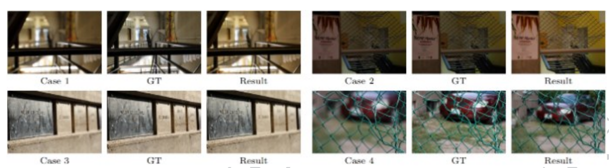


图13 Nano Banana Pro在DPDD数据集上的若干代表性定性结果

Fig. 13 Some representative qualitative results of Nano Banana Pro on the DPDD dataset

7 图像去噪

7.1 背景

视觉-语言模型的最新进展（Bai等，2025；Liu等，2023；Lu等，2025）标志着范式的转变，即朝着在单一框架内统一多模态与多任务的架构发展。值得注意的是，以Nano Banana Pro为代表的模型表明，通过协同训练理解与生成目标，可以激发新兴能力并提升跨任务的泛化性能。

在本技术报告中，我们系统评估了Nano Banana Pro在五个成熟基准上的去噪性能：McMaster（Zhang等，2011）、Kodak24（Rich Franzen，1999）、Urban100（Huang等，2015），以及面向真实传感器噪声的PolyU（Xu等，2018）和SIDD-small（Abdelhamed等，2018）。

7.2 实验设置

我们通过向Nano Banana Pro输入带噪图像及自然语言指令来评估其去噪性能，并将其与五种代表性的专用模型DnCNN、Restormer、MaskDenoising、HAT（hybrid adversarial training）和DIL（distortion invariant representation learning）进行比较。

数据集：我们在五个常用图像去噪基准上评估Nano Banana Pro，覆盖合成噪声与真实世界噪声场景，以反映模型在平滑纹理、自然场景以及复杂结构中的重建能力。对于合成噪声数据集McMaster、Kodak24和Urban100，在清晰图像上添加固定噪声水平σ=50的加性高斯白噪声。对于真实噪声数据集PolyU（polyu real-world noisy images dataset）和

SIDD-val（smartphone image denoising dataset validation set），直接使用其提供的带噪/清晰图像对，不额外引入合成噪声，以反映真实传感器噪声条件。

分辨率与失败案例处理：Nano Banana Pro的输出分辨率约为1K，具体尺寸随样本变化。我们采用双线性插值将输出调整至与对应真值图像一致的分辨率，并在此基础上计算评价指标。在评估过程中，模型偶尔生成与输入语义不符或去噪失败的结果。对此，我们重复提交相同输入与提示到API直至获得有效输出。该策略保证定量结果反映模型在成功生成情况下的性能，同时将失败现象视为统一生成模型在复原任务中的局限。

评估指标：采用PSNR衡量像素级保真度，SSIM评估结构相似性，二者均在RGB空间计算。

7.3 定量结果和定性结果

如表7所示，Nano Banana Pro在所有基准上均显著落后于专用模型。在合成数据集上，其PSNR相比DIL低5.04~7.42 dB，SSIM下降0.075~0.219，且这一差距在不同纹理复杂度下均持续存在。在真实噪声场景中性能进一步恶化：在SIDD Val上PSNR落后DIL 8.00dB，在PolyU上差距扩大至14.83dB，低MaskDenoising。这表明Nano Banana Pro在建模和去除复杂真实噪声方面存在明显不足。

表7　在合成噪声与自然噪声数据集上的性能定量对比结果

Table 7　Quantitative results of performance comparison on synthetic and natural noise datasets

噪声类型	数据集	DnCNN（Zhang等，2017）	Restormer（Zamir等，2022）	MaskDenoising（Chen等，2023）	HAT（Yan等，2022）	DIL（Li等，2023）	NB pro
高斯噪声 σ=50	McMaster（Zhang等，2011）	20.18/0.312	20.47/0.312	20.63/0.379	20.79/0.364	26.61/0.669	21.57/0.594
	Kodak24（Rich Franzen，1999）	19.78/0.301	20.12/0.321	20.72/0.368	21.04/0.390	27.46/0.736	20.04/0.517
	Urban100（Huang等，2025）	19.62/0.420	19.36/0.437	20.51/0.485	20.80/0.492	25.89/0.768	19.22/0.607
无噪声	SIDD Val（Abdelhamed等，2018）	–/–	–/–	33.14/0.913	28.58/0.570	34.76/0.848	26.76/0.681
	PolyU（Xu等，2018）	–/–	–/–	24.78/0.812	37.25/0.948	37.65/0.950	22.82/0.806

注：评价指标为PSNR和SSIM，数值越高表示性能越优

图14给出了Nano Banana Pro与先进基线方法的定性对比，揭示了生成模型在感知质量与像素级保真度之间的权衡。在包含文本的图像中，模型能够生成视觉上极为清晰的结果。然而，在纹理和颜色一致性方面，其表现明显不足：部分样例中高频细节被过度平滑或被幻觉生成的纹理取代，另一些样例中则出现明显色偏。总体而言，尽管Nano Banana Pro能生成感知上吸引人的图像，但难以满足高精度去噪任务对严格保真度的要求。

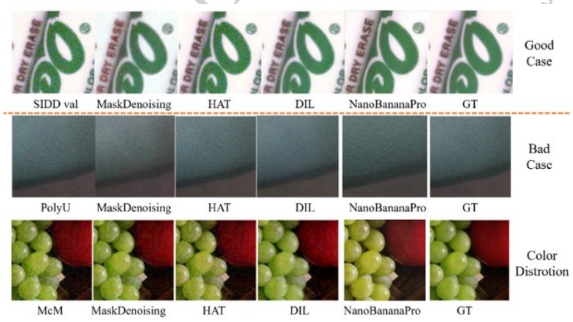


图14　去噪结果的可视化对比

Fig. 14　Visual comparison of denoising results

7.4 分析

结合实验结果与模型特性，Nano Banana Pro去噪性能不足主要源于两点：其一，该模型主要面向高级多模态理解与生成优化，并非为像素级复原任务设计，因此缺乏专用去噪模型所具备的结构偏

置和针对性约束；其二，作为统一生成模型，它更强调语义合理性与视觉连贯性，而非严格像素保真，这容易导致细节过度平滑，从而在定量指标上明显落后于监督训练的专用方法。

总体来看，Nano Banana Pro在合成高斯噪声和真实世界盲去噪基准上的表现均明显弱于先进专用模型，显示出其在直接去噪应用中的竞争力有限，因此不建议在未经调整的情况下直接用于去噪任务。

8 图像去反光

8.1 背景

在计算机视觉中，清晰、无干扰的图像是目标检测和语义分割等后续分析的基础。然而，玻璃、水面和金属等反射表面常引入干扰性反光层，造成细节模糊和信息遮挡。单幅图像去反光(single image reflection removal，SIRR)旨在从混合图像中分离传输层与反光层，以恢复真实场景，在自动驾驶、安防监控和消费电子等领域具有重要应用价值。

单图像去反光本质上是不适定逆问题。早期方法多依赖人工先验和简化线性模型，但在真实场景中受光照、视角和材质变化影响，泛化能力有限。尽管近年来相关算法和基准数据集不断发展，现有方法仍受制于高质量标注数据稀缺、强反光信息丢失以及传输层与反光层外观重叠等问题，难以同时兼顾反光去除与细节保留。基于此，本节聚焦生成模型 Nano Banana Pro，通过定量与定性分析其在真实去反光场景中的表现，以揭示其优势与局限。

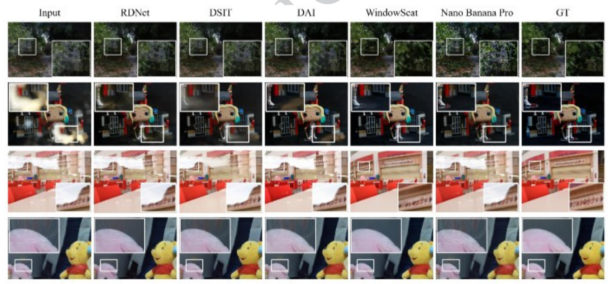


图 15 反射去除结果的定性对比

Fig. 15 Qualitative comparison of reflection removal results

8.2 定量结果

评估选用三个主流数据集：Real20(Zhang 等，2018)、Nature(Li 等，2020)以及 SIR^2(Wang 等，2017)(Objects、Postcard 和 Wild 子集)，并与 15 种最先进方法进行比较。像素级指标采用 PSNR 和 SSIM，结果汇总于表 8；感知质量进一步使用 SSIM 和 LPIPS，并与近期 SOTA 方法进行比较，如表 9 所示。

表 8 单幅图像反射去除方法的定量对比结果

Table 8 Quantitative Comparison of Single-Image Reflection Removal Methods

方法	Real20 (20)		Nature (20)		Objects (200)		Postcard (199)		Wild(55)		SIR^2 (454)		SIR^2(500)	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
ERRNet (Wei 等，2019)	22.89	0.803	–	–	24.87	0.896	22.04	0.876	24.25	0.853	23.55	0.882	–	–
IBCLN (Li 等，2020)	21.86	0.762	23.57	0.783	24.87	0.893	23.39	0.875	24.71	0.886	24.20	0.884	–	–
YTMT (Hu 等，2021)	23.26	0.806	–	–	24.87	0.896	22.91	0.884	25.48	0.890	24.08	0.890	–	–
Dong et al. † (Dong 等，2021)	23.34	0.812	23.45	0.808	24.36	0.898	23.72	0.903	25.73	0.902	24.25	0.901	–	–
DSRNet (w/o extra) (Hu 等，2023)	24.23	0.820	–	–	26.28	0.914	24.56	0.908	25.68	0.896	25.45	0.909		–
DSRNet (w. extra) (Hu 等，2023)	23.91	0.818	–	–	26.74	0.920	24.83	0.911	26.11	0.906	25.83	0.914		–
RAGNet (Li 等，2023)	22.95	0.793	–	–	26.15	0.903	23.67	0.879	25.52	0.880	24.99	0.890	–	–
RRW (Zhu 等，2024)	23.82	0.817	25.96	0.843	–	–	–	–	–	–	25.45	0.910	–	–

表8续表

方法	Real20 (20)		Nature (20)		Objects (200)		Postcard (199)		Wild(55)		SIR^2 (454)		SIR^2(500)	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DSIT(data I)(Hu等，2024)	25.06	0.836	–	–	26.81	0.919	25.63	0.924	27.06	0.910	26.32	0.920	–	–
DSIT (data II)(Hu等，2024)	25.22	0.836	–	–	27.27	**0.932**	25.58	0.922	27.40	0.918	26.54	0.926	–	–
RDNet (w/o nature)	24.43	0.835	–	–	25.76	0.905	25.95	0.920	27.20	0.910	26.02	0.912	–	–
RDNet (w. nature) (Zhao等，2025)	25.58	0.846	–	–	26.78	0.921	26.33	0.922	27.70	0.915	26.69	0.921	–	–
F2T2-HiT (Cai等，2025)	21.64	0.766	26.08	0.837	–	–	–	–	–	–	25.72	0.903	–	–
Huang et al. (Huang 等，2025)	25.12	0.828	27.03	0.853	27.07	0.930	26.43	0.931	27.96	**0.922**	26.90	0.929	–	–
L-DiffER* (Hong 等，2024)	23.77	0.821	23.95	0.831	–	–	–	–	–	–	–	–	25.18	0.911
DAI* (Hu 等，2025)	25.24	0.840	27.05	0.846	–	–	–	–	–	–	–	–	27.32	0.931
Lu et al.* (Lu 等，2025)	–	–	–	–	–	–	–	–	–	–	–	–	28.41	0.912
WindowSeat* (Zakarin 等，2025)	26.28	0.856	27.12	0.849	28.81	0.944	29.17	0.934	28.97	0.935	28.99	0.939	28.75	**0.940**
WindowSeat* (Qwen-IE)	**26.60**	**0.864**	**27.57**	**0.855**	**28.85**	0.938	28.70	0.933	**29.44**	**0.936**	28.84	0.936	28.60	0.937
Nano Banana Pro	20.26	0.655	21.48	0.723	21.95	0.751	19.29	0.675	23.58	0.798	20.98	0.724	21.11	0.730

注：最佳结果和次优结果分别以黑色加粗和下划线标出。† 表示训练数据包含 Nature 数据集；★ 表示基于生成式 AI 的方法。SIR^2 数据集包含 Objects、Postcard 和 Wild 三个子集；SIR^2 (454) 表示常用子集，而 SIR^2 (500) 表示完整的公开数据集。

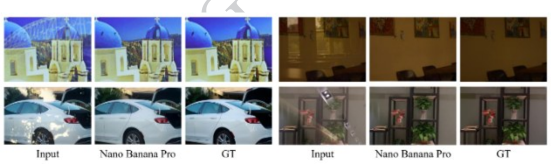


图 16　选取样本的可视化对比结果

Fig. 16　Visual comparison of selected samples

如表 8 所示，Nano Banana Pro 在所有数据集上的 PSNR 和 SSIM 均明显落后于专用模型。这是因为回归型方法直接最小化像素重建误差，而 NB Pro 更强调语义连贯性，易产生强度缩放或空间偏移，从而在像素指标上受到显著惩罚。表 9 的感知指标进一步表明，该模型的 LPIPS 明显高于 SOTA，反映出语义和风格层面的偏差。其生成过程更接近激进的“图像到图像转换”，往往改变光照、纹理或色彩分布，偏离真值的感知流形。

总体来看，较高的 LPIPS 和次优的 SSIM 说明通用生成模型在高保真去反光任务中存在固有限制。尽管生成结果在视觉上可能自然，但由于缺乏精确解耦反光与背景的机制，其在颜色、纹理和高层语义上与真值存在明显偏差，难以满足去反光任务对严

格保真度的要求。

8.3 定性结果

本节对 Nano Banana Pro 进行定性评估。我们首先在图 15 中将其结果与最先进方法进行对比，随后通过图 16 的代表性样本分析其复原能力。

如图 15 所示，Nano Banana Pro 的表现波动较大。个别样本中其结果可接近甚至优于现有方法，但整体稳定性明显不足。模型在高频细节保留方面较弱，常出现复杂纹理丢失，或因语义歧义将反光伪影误判为背景并加以增强，导致平均性能偏低。

图 16 展示了模型表现较好的案例。当反光层与传输层在语义或外观上差异明显时，Nano Banana Pro 能有效抑制反光并保持背景完整，显示出较高的性能上限。这一能力源于其强大的生成先验，但由于缺乏针对反光分离的监督，这些先验也可能被误用，导致解耦失败或引入生成性幻觉，从而产生次优结果。

表 9 主流数据集上的感知质量对比结果（MS-SSIM 与 LPIPS）

Table 9 Perceptual Quality Comparison（MS-SSIM and LPIPS）on Mainstream Datasets

方法	Real20 (20)		Nature (20)		Objects (200)		Postcard (199)		Wild (55)	
	MS-SSIM ↑	LPIPS ↓	MS-SSIM ↑	LPIPS ↓	MS-SSIM ↑	LPIPS ↓	MS-SSIM ↑	LPIPS ↓	MS-SSIM ↑	LPIPS ↓
DSRNet（Hu 等，2023）	0.8737	0.1831	0.9144	0.1478	0.9564	0.0847	0.9263	0.1260	0.9338	0.1096
DAI（Hu 等，2025）	0.9045	0.1790	0.9309	0.2161	0.9638	0.0689	0.9567	0.1029	0.9423	0.0941
RDNet（Zhao 等，2025）	0.9081	0.1442	0.9231	<u>0.1361</u>	0.9609	0.0836	0.9361	0.1121	0.9406	0.0992
DSIT（Hu 等，2024）	0.8934	0.1618	0.9223	0.1598	0.9586	0.0939	0.9441	0.1242	0.9447	0.0967
WindowSeat（Zakarin 等，2025）	<u>0.9296</u>	<u>0.1131</u>	<u>0.9435</u>	0.1368	**0.9759**	**0.0470**	**0.9693**	**0.0504**	<u>0.9625</u>	**0.0632**
WindowSeat (Qwen-IE)（Zakarin 等，2025）	**0.9396**	**0.1074**	**0.9494**	**0.1355**	<u>0.9661</u>	<u>0.0550</u>	<u>0.9664</u>	<u>0.0549</u>	**0.9655**	<u>0.0682</u>
Nano Banana Pro	0.8013	0.2411	0.8580	0.1851	0.8861	0.1552	0.8373	0.2513	0.8874	0.1578

注：↑ 表示数值越高性能越好，↓ 表示数值越低性能越优。最佳结果和次优结果分别以黑色加粗和下划线标出。基线方法结果引自 WindowSeat

9 图像去眩光

9.1 背景

镜头眩光由强入射光在镜头系统内散射和反射产生，通常表现为散射眩光和反射眩光，会降低图像质量并干扰立体匹配、光流估计和语义分割等下游任务，对自动驾驶和空中目标跟踪等安全关键应用构成风险（Hullin 等，2011；Li 等，2021；Wu 等，2021）。近年来，去眩光方法已由传统检测策略发展为依赖大规模数据集的深度学习框架。Flare7K++（Dai 等，2024）是该领域的重要基准，包含 7,000 个合成样本和 962 张真实眩光图像，覆盖多种复杂退化现象。

基于此，本文在 Flare7K++（Dai 等，2024）和 FlareReal600（Dai 等，2024）两个互补基准上评估 Nano Banana Pro 的去眩光能力，重点分析其在不同分辨率下消除夜间眩光伪影、保持场景语义与光度一致性的表现。

9.2 定量结果与定性结果

为全面评估 Nano Banana Pro 的去眩光能力，我们从定量与定性两个方面进行分析。

定量评估。我们首先在 Flare7K++数据集上，以约 1K 输出分辨率评估模型性能，采用 PSNR 和 SSIM 作为指标，并与 Restormer、Uformer 和 DeflareMamba 等专用方法进行比较（表 10）。

随后，在 FlareReal600 数据集上，以原始分辨率生成 2K 和 4K 结果，并在 PSNR、SSIM 基础上引入 LPIPS 评估感知质量，结果与 MIPI 2024 挑战赛冠军 MiAlgo AI 的公开验证集成绩进行对比（表 11）。实验表明两点趋势：其一，随着输出分辨率提高，性能指标整体下降；其二，在 FlareReal600 中，较高亮度场景会进一步拉低指标，而这一现象在低分辨率的 Flare7K++上并不明显。

表 10　Nano Banana Pro 与代表性专用方法在 Flare7K++ 数据集上的定量对比结果

Table 10　Quantitative comparisons of Nano Banana Pro and representative specialists on the Flare7K++ dataset

指标	输入	Restormer(Zamir 等,2022)	Uformer(Wang 等,2022)	Flare-level(Deng 等,2024)	DeflareMamba(Huang 等,2025)	NB Pro
PSNR ↑	22.56	27.60	27.63	27.05	26.06	24.92
SSIM ↑	0.857	0.897	0.894	0.901	0.898	0.844

表 11　Nano Banana Pro 与代表性专用方法在 Flare-Real600 数据集上的定量对比结果

Table 11　Quantitative comparisons of Nano Banana Pro and representative specialists on the FlareReal600 dataset

指标	PPDN(Dai 等,2024)	NB Pro(2K)	NB Pro(4K)
LPIPS ↓	0.143	0.287	0.361
PSNR ↑	22.15	19.07	18.32
SSIM ↑	0.708	0.496	0.424

注:一些低分样本在视觉上看起来仍较为令人满意,但由于在亮度或颜色方面与真实参考图像存在一定偏差,其定量评价指标得分相对较低。

图 17　在 Flare7K++ 数据集上的耀斑去除结果定性对比

Fig. 17　Qualitative comparison of flare removal results on the Flare7K++ dataset

定性分析。图 17 的对比结果揭示了 Nano Banana Pro 的双重特性。在理想条件下,模型能够显著抑制眩光并恢复细节,部分样本的感知质量甚至优于专用方法。

然而,扩散模型的随机性带来了明显不稳定性,易出现语义幻觉,如生成无关内容、错误抑制真实光源或虚构光源状态。提示词工程只能部分缓解这一问题,难以满足对确定性要求较高的应用场景。与此同时,如图 18 所示,定量指标与主观感知之间存在一定错位,部分低分样本在视觉上仍保持较高质量,说明像素级指标难以完全反映生成结果的感知优势。

总体来看,Nano Banana Pro 呈现出“高上限、低下限”的特征:其生成能力在感知质量上具备潜在优势,但当前仍缺乏稳定性和一致性,尚难满足高可靠性的去眩光需求。

图 18　Nano Banana Pro 在 Flare7K++ 数据集上部分高分与低分样本的示例

Fig. 18　Examples of some high-scoring and low-scoring samples from Nano Banana Pro on the Flare7K++ dataset

10　弱光图像增强

10.1　背景

弱光图像增强(Cai 等,2018;Wei 等,2018;Yang 等,2021)旨在从光照不足的图像中恢复视觉质量良好的结果,但同时面临亮度调整(Cai 等,2023)、噪声抑制(Hou 等,2023)和色彩复原等多重挑战,并需保持结构细节与语义一致性。传统方法依赖手工先验,或通过成对的低光/正常光数据进行监督训练。

在本章中,我们评估统一多模态模型 Nano Banana Pro 在弱光图像增强任务中的表现,覆盖 LOL1(low-light dataset version 1)(Wei 等,2018)、LOLv2-real(low-light dataset version 2, real subset)(Yang 等,2021)和 SICE(single image contrast enhancer)(Cai 等,2018)三个常用基准,并将其零样本结果与当前最先进的监督与无监督方法进行比较。

10.2　实验设置

数据集。我们在三个常用弱光增强基准上进行实验。LOLv1(Wei 等,2018)由真实拍摄的低光/正常光图像构成;LOLv2-real(Yang 等,2021)则涵盖更丰富的室内外场景;SICE(Cai 等,2018)规模更大,由

多曝光序列构建,其测试集包含更复杂的场景和光照。

评估指标:采用峰值信噪比(PSNR)和结构相似性指数(SSIM)作为全参考指标,数值越高表示相对于正常光真值的重建质量越好。

对比方法:在零样本设置下评估Nano Banana Pro,并将其与ZeroDCE(zero-reference deep curve estimation)(Guo等,2020)、RUAS(retinex-inspired unrolling with architecture search)(Liu等,2021)、LLFlow(low-light image enhancement with normalizing flow)(Wang等,2022)、LLFormer(low-light transformer)(Wang等,2019)、GSAD(global structure-aware diffusion process)(Hou等,2023)和Quadprior(Wang等,2024)方法进行比较。

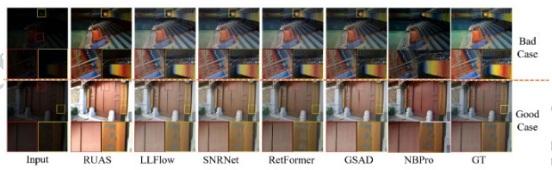


图19　Nano Banana Pro与多种代表性专用方法的可视化对比示例

Fig. 19　Visual comparison examples of Nano Banana Pro and several representative specialists.

10.3　定量结果和定性结果

表12给出了定量结果。在LOLv1和LOLv2-real上,Nano Banana Pro的零样本性能明显落后于最先进的监督方法,且在LOLv2-real上差距尤为显著,其PSNR和SSIM难以与真值参考对齐。这表明在缺乏任务特定训练的情况下,模型难以满足这些基准的增强要求。相比之下,在更具多样性和挑战性的SICE数据集上,Nano Banana Pro的指标略高于部分对比方法,显示出一定的零样本竞争力。

图19表明,Nano Banana Pro通常能够提亮暗区并恢复场景内容,但亮度控制仍不够稳定,部分样本存在过曝或增强不足现象,这可能与其依赖通用视觉先验而非显式光照建模有关。值得注意的是,模型较少引入颜色失真、光晕或结构破坏等常见伪影,整体纹理也能较好保留,说明其生成过程具有一定稳定性。

总体来看,Nano Banana Pro在弱光增强任务中展现出一定的零样本潜力,其优势在于输出自然、伪影较少,但这种相对保守的增强策略也限制了PSNR和SSIM等指标表现。与任务专用方法相比,其整体性能仍有明显差距,主要受限于缺乏显式光照建模、对提示词较敏感,以及难以适应不同基准中的真值定义。

11　水下图像增强

11.1　背景

水下图像退化是海洋环境中的典型视觉问题。由于光在水中传播时受到选择性吸收、多次散射和光照波动影响,图像常出现显著色偏、对比度下降和纹理模糊,从而影响目标识别、地形重建和行为分析等下游任务的可靠性,并对资源勘探、文化遗产考古、无人潜航器导航和海底设施维护等应用带来风险。传统方法多依赖像素域处理(Ancuti等,2012;Ancuti等,2018;Gao等,2019;Ghani等,2015;Iqbal等,2010)或物理建模(Drews等,2016;Galdran等,2015;He等,2010;Li等,2016;Li等,2016;Zhang等,2022),但在复杂场景下泛化能力有限。近年来,深度学习方法在色彩校正和细节恢复方面取得了显著进展,其中基于扩散模型的方法通过渐进式去噪较好地平衡了全局色调与局部纹理。

本节采用三个基准进行评估:UIEB(underwater image enhancement benchmark)(Li等,2019)挑战测试集用于复杂场景适应性评估;LSUI(large-scale underwater image dataset)(Peng等,2023)按WF-Diff(wavelet-based fourier information interaction with frequency diffusion adjustment)(Zhao等,2024)的测试划分选取400张图像以检验泛化能力;U45(He等,2020)为无参考真实水下图像数据,覆盖绿色偏、蓝色偏和雾化退化,更贴近实际应用。基于这些数据集,本文系统评估Nano Banana Pro的水下图像增强能力,重点考察其色偏校正、纹理恢复以及场景语义和亮度一致性保持表现。

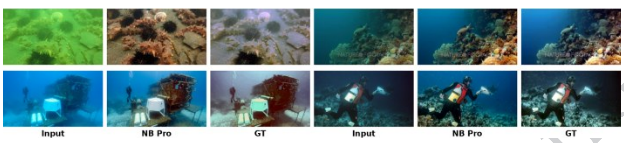
我们使用输出分辨率约为1K的Nano Banana Pro在三个数据集上进行实验,并采用无参考指标对水下图像增强性能进行评估。结果表明,Nano Banana Pro在UIQM(underwater image quality measure)和UCIQE(underwater color image quality evaluation)指标上整体具备与主流方法相当的竞争力。在UIEB数据集上,其表现与最优方法差距较小;在退化更复杂的LSUI数据集上,Nano Banana Pro在两项

表 12　在 LOL 和 SICE 数据集上的定量对比结果

Table 12　Quantitative comparisons on the LOL and SICE datasets. The best results are highlighted by black bold

方法	颜色模型	LOLv1（Wei 等，2018）			LOLv2-Real（Yang 等，2021）			SICE（Cai 等，2021）		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RetinexNet（Wei 等，2018）	Retinex	18.915	0.427	0.470	16.097	0.401	0.543	12.424	0.613	-
KinD（Zhang 等，2019）	Retinex	23.018	0.843	0.156	17.544	0.669	0.375	-	-	-
ZeroDCE（Guo 等，2020）	RGB	21.880	0.640	0.335	16.059	0.580	0.313	12.452	**0.639**	-
RUAS（Liu 等，2021）	Retinex	18.654	0.518	0.270	15.326	0.488	0.310	8.656	0.494	-
LLFlow（Wang 等，2022）	RGB	24.998	0.871	0.117	17.433	0.831	0.176	12.737	0.617	-
EnlightenGAN（Jiang 等，2021）	RGB	20.003	0.691	0.317	18.230	0.617	0.309	-	-	-
SNR-AW（Xu 等，2022）	SNR+RGB	26.716	0.851	0.152	21.480	**0.849**	0.163	-	-	-
Bread（Guo 等，2023）	YCbCr	25.299	0.847	0.155	20.830	0.847	0.174	-	-	-
PairLIE（Fu 等，2023）	Retinex	23.526	0.755	0.248	19.085	0.778	0.317	-	-	-
LLFormer（Wang 等，2023）	RGB	25.278	0.823	0.167	20.856	0.792	0.211	-	-	-
RetinexFormer（Cai 等，2023）	Retinex	27.140	0.850	0.129	**22.794**	0.840	0.171	-	-	-
GSAD（Hou 等，2023）	Retinex	**27.605**	**0.876**	**0.092**	20.153	0.846	**0.113**	-	-	-
QuadPrior（Wang 等，2024）	Kubelka-Munk	22.849	0.800	0.201	20.592	0.811	0.202	-	-	-
Nano Banana Pro	-	18.496	0.684	0.481	15.661	0.537	0.465	**14.081**	0.493	-

注：最佳结果以黑色加粗标出



11.2 定量结果

图 20　NB Pro 在 UIEB 和 LSUI 数据集上的水下图像增强可视化示例

Fig. 20　Visualization examples for underwater image enhancement using NB Pro on the UIEB and LSUI datasets

指标上均取得最佳结果，体现了较强的鲁棒性；在无参考的 U45 数据集上，Nano Banana Pro 获得最高 UIQM 分数，并在 UCIQE 中位列前三，显示出其在真实无参考场景中的应用潜力。

11.3　定性结果

为直观评估 Nano Banana Pro 的水下图像增强效果，我们在 UIEB、LSUI 和 U45 数据集上给出了可视化结果，并与代表性方法进行对比。图 20 展示了其在 UIEB 和 LSUI 上的典型成功案例，图 21 给出了相应的失败示例。模型在 U45 数据集上的结果不在此展示。

定性结果表明，在严重绿色或蓝色色偏、低照度和高浑浊度等强退化场景中，NB Pro 能够生成质量较高的增强结果，在多种退化叠加时其视觉效果甚至可优于参考图像（图 20）。但在色偏较弱的轻度退化场景中，模型对退化特征的识别能力有限，增强结果与输入差异较小，细节和对比度改善不足（图 21）。这一趋势在 U45 数据集中同样明显：其在高浑浊度或严重色偏场景中表现较好，而在轻度退化场景中的提升相对有限。

总体来看，NB Pro 在水下图像增强中表现出较强的复杂退化感知恢复能力，但在轻度退化条件下的像素保真度与增强稳定性仍然不足。

12　高动态范围图像重建

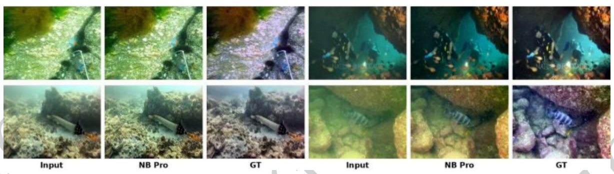
12.1　背景

高动态范围（HDR）重建旨在恢复传统成像系统难以完整记录的光强范围，以缓解 LDR 图像中的高光过曝和阴影欠曝问题，从而提升视觉质量及目

表 13 在 UIEB、LUSI 和 U45 数据集上的定量对比结果
Table 13 Quantitative comparisons on UIEB, LUSI and U45 datasets

方法	UIEB		LUSI		U45	
	UIQM ↑	UCIQE ↑	UIQM ↑	UCIQE ↑	UIQM ↑	UCIQE ↑
UWCNN(Li 等,2019)	3.8325	0.5552	4.1699	0.5453	4.387	0.5622
UIEC2-Net (Wang 等,2021)	3.327	0.609	3.9833	0.5888	**4.4293**	0.6104
U-Shape (Peng 等,2023)	3.332	0.5751	4.0334	0.574	4.3524	0.5856
PUGAN(Cong 等,2023)	3.2163	**0.6176**	4.0417	0.5834	4.3377	**0.6117**
DM-Water(Tang 等,2023)	**3.8925**	0.5994	4.0595	0.5883	4.1986	0.586
WF-Diff (Zhao 等,2024)	3.7388	0.5867	4.0308	0.5688	4.2193	0.5813
NB Pro	3.634	0.5899	4.2993	0.5961	4.3907	0.5899

注:最佳结果以黑色加粗标



增强失败案例可视化示例
图 21 NB Pro 在 UIEB 和 LSUI 数据集上的水下图像
Fig. 21 Visualization of failed cases in underwater image enhancement for NB Pro on the UIEB and LSUI datasets

标检测、场景理解和深度估计等下游任务的可靠性。

随着深度学习的发展,HDR 重建逐渐形成以大规模数据集为支撑的技术体系,其中 MIT FiveK(Bychkovsky 等,2011)和 HDR+(Vinker 等,2021)分别提供了高质量参考图像与真实拍摄条件下的大规模数据支持。

本研究将 Nano Banana Pro 与多种传统及深度学习方法进行定量比较。为保证公平性,所有图像统一下采样至 480p,并采用 PSNR、SSIM、LPIPS 和 ΔE 四项指标,分别评估像素保真度、结构相似性、感知相似性和色差表现,结果见表 14。

表 14 在 HDR+ 和 MIT-FiveK 数据集上的定量对比结果
Table 14 Quantitative comparison on HDR+ and MIT-FiveK datasets

方法	HDR+(480p)				MIT-FiveK(480p)			
	PSNR ↑	SSIM ↑	LPIPS ↓	ΔE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	ΔE ↓
UPE(Wang 等,2019)	23.33	0.852	0.150	7.68	21.82	0.839	0.136	9.16
HDRNet(Gharbi 等,2017)	24.15	0.845	0.110	7.15	23.31	0.881	0.075	7.73
CSRNet (He 等,2020)	23.72	0.862	0.104	6.67	25.31	0.909	**0.052**	6.17
DeepLPF(Moran 等,2020)	25.73	0.904	0.073	6.05	24.97	0.897	0.061	6.22
LUT(Zeng 等,2020)	23.29	0.855	0.117	7.16	25.10	0.902	0.059	6.10
sLUT(Wang 等,2021)	26.13	0.901	0.069	5.34	24.67	0.896	0.059	6.39
CLUT (Zhang 等,2022)	26.05	0.892	0.088	5.57	24.94	0.898	0.058	6.71
LLF-LUT++(Zhang 等,2025)	**28.43**	**0.924**	**0.056**	**4.54**	**26.06**	**0.912**	0.054	**4.93**
Nano Banana Pro	14.24	0.467	0.221	19.82	19.20	0.639	0.133	11.14

注:最佳结果以黑色加粗标出。

12.2 定量结果

整体来看,NB Pro 在两项基准上均明显落后于对比方法,说明其在强调像素级重建和色彩保真度的全参考评估下,与专业增强或调色参考图像仍存在系统性偏差。这一差异源于模型目标的不同:传统 HDR 或增强方法通常基于成对数据进行端到端

训练，直接优化像素级损失，因此在PSNR和SSIM上更具优势；而NB Pro更侧重语义连贯性和整体视觉效果，其输出更接近对输入的生成式重建，而非严格的像素映射，因此在亮度分布、局部对比度和色彩风格上更易偏离参考。

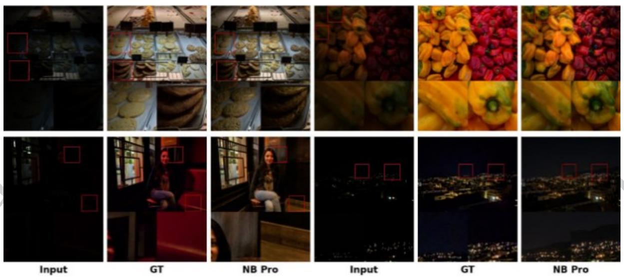


图22　NB Pro在HDR成像任务中出现的纹理细节丢失与人工添加现象的可视化示例

Fig. 22　Visualization of texture detail loss and artificial addition cases in HDR Imaging using NB Pro

值得注意的是，NB Pro在MIT-FiveK上的LPIPS相对差距较小，说明其在高层语义特征上仍与参考图像保持一定相似性。

12.3　定性结果

为直观分析NB Pro在HDR重建中的表现及其局限性，我们在HDR+和MIT-FiveK数据集上选取代表性场景进行定性可视化评估，实验涵盖常规光照、低光低细节以及密集纹理三类情形，这里仅展示低光低细节。

在常规光照场景中，NB Pro能较好地恢复动态范围和色彩层次，其整体视觉效果接近参考HDR图像，表明模型在基础HDR重建任务中具备有效性。

图22展示了低光、低细节场景下的表现。此类情况下，NB Pro易出现两类问题：一是细节丢失，如阴影区域中的纹理和轮廓未能恢复；二是冗余生成，如引入原场景中不存在的纹理、颜色偏差或虚构结构。这些现象源于低光条件下可用细节不足，导致模型在纹理感知和生成上产生偏差。

而当面对织物、植被等复杂纹理时，NB Pro难以在纹理保留与增强强度之间取得平衡，常表现为过度锐化，引入明显的边缘和局部伪影，削弱了纹理的自然层次。

总体而言，NB Pro在常规HDR场景中能够产生令人满意的视觉结果，但在低光低细节和高密度纹理条件下仍存在纹理缺失、冗余生成及过度锐化等问题。

综上，NB Pro更适用于对像素级精度要求较低、以视觉体验为主的非关键HDR应用，不适合需要严格细节与色彩还原的安全关键或专业场景。

13　多焦点图像融合

13.1　背景

在计算机视觉和数字摄影中，获取全清晰图像对后续分析十分重要。受景深限制，单次拍摄难以同时对不同深度物体成像，因此多焦点图像融合(Multi-Focus Image Fusion, MFIF)通过融合不同焦点图像生成全景深结果，已广泛应用于医疗成像和消费电子等领域。

近年来，深度学习被广泛用于MFIF，主要包括基于决策的方法(Li等，2020；Liu等，2017；Xiao等，2021)和基于重建的方法(Li等，2024；Lu等，2025；Zhang等，2020)。前者通过判别聚焦区域进行融合，但在焦点边界处容易失效；后者采用端到端重建策略，虽能整体建模，却常带来细节损失，两类方法均存在一定保真度不足。

本文在Lytro(Nejati等，2015)、MFFW(multi-focus image fusion in the wild)(Xu等，2020)、MFI-WHU(multi-focus image dataset of wuhan university)(Zhang等，2021)和SIMIF(Tsai等，2025)四个基准上，对Nanobanana与其他方法进行比较，并采用EN(entropy)(Jähne等，2005)、AG(average gradient)(Cui等，2015)、SF(spatial frequency)(Zheng等，2007)、NMI (normalized mutual information)(Hossny等，2008)、QY(Yang等，2018)和QCB(Chen等，2009)六项指标从多角度评估融合性能。

13.2　定量结果

由表15和表16可见，NB Pro在无参考指标上表现突出，取得了接近甚至超过当前最先进方法的结果，表明其生成图像具有较高的整体视觉质量。相对而言，在源参考指标上，NB Pro的表现明显较弱，与既有零样本和无监督方法面临的问题一致。这表明模型在融合过程中未能充分保持与源图像在梯度和结构层面的对应关系；它展现了过强的创造性，却牺牲了保真度。

13.3　定性结果

结果显示，不同样本间性能差异明显，既有成功

案例,也存在不足。在部分示例中,模型能够同时保留前景与背景的清晰区域,并在不同焦点区域之间实现自然过渡,几乎不引入伪影;部分区域甚至呈现出比源图像更丰富的细节,从而提升整体视觉质量。

然而,在其他案例中也观察到明显局限。例如,某些样本中前景花瓣或背景草地仍保持模糊状态,尽管这些区域在源图像中存在清晰对应。这表明模型在融合过程中未能始终准确识别或定位最优的合焦区域。

为进一步比较性能,图 23 将 NB Pro 与最先进的 MFIF 方法进行了对比。在“锁”场景中,NB Pro 能有效缓解散焦扩散引起的边界伪影和暗色鬼影,在区域过渡和平滑性方面优于多种对比方法,生成结果在视觉上更为自然。

表 15 在 Lytro 和 MFFW 数据集上的定量对比结果

Table 15 Quantitative comparison on the Lytro and MFFW datasets

方法	Lytro						MFFW					
	NMI	EN	AG	SF	QY	QCB	NMI	EN	AG	SF	QY	QCB
IFCNN (Zhang等,2020)	0.9388	7.5318	8.1473	19.3992	0.9518	0.7294	0.8206	7.1688	9.6849	22.0173	0.8715	0.6423
SESF (Ma等,2021)	1.6808	7.5322	8.1530	19.4251	0.9879	0.8064	1.0876	7.1850	9.8678	22.9291	0.9588	0.7418
GACN (Ma等,2022)	1.1723	7.5311	8.1245	19.3247	0.9878	0.8062	1.0820	7.1923	9.7328	22.2842	0.9273	0.7192
GRFusion(Li等,2023)	1.1879	7.5329	8.1823	19.4539	0.9863	0.8071	1.1426	7.1711	9.8826	22.6520	0.9334	0.7223
ZMFF(Hu等,2023)	0.8795	7.5223	7.7771	18.8184	0.9321	0.6731	0.7711	7.1546	9.2144	21.3405	0.8474	0.6722
MUFusion(Cheng等,2023)	0.8088	**7.6093**	8.1578	18.9240	0.8997	0.6819	0.7551	7.2026	8.9247	19.9974	0.8167	0.6205
DB-MFIF(Zhang等,2024)	1.0573	7.5386	8.2415	19.5290	0.9637	0.7770	0.8699	7.1935	10.1346	23.0305	0.8663	0.6647
MFFT(Zhai等,2024)	1.1533	7.5620	**8.6089**	**21.3759**	0.9523	0.7511	1.1310	**7.2281**	10.1971	**24.5709**	0.9336	0.6865
DMANet(Quan等,2025)	1.1897	7.5319	8.2059	19.5129	0.9853	0.8054	1.1513	7.1846	10.0948	23.2370	0.9506	0.7276
MCCSR (Zheng等,2025)	**1.1920**	7.5329	8.1757	19.4392	**0.9890**	**0.8084**	**1.1800**	7.1688	9.7407	22.2989	**0.9813**	**0.7517**
NB Pro	0.7476	7.5267	8.2977	20.1044	0.7755	0.6603	0.6319	7.1823	**10.2879**	23.0223	0.5537	0.5638

注:最佳结果以黑色加粗标出。

表 16 在 MFI-WHU 和 SIMIF 数据集上的定量对比结果

Table 16 Quantitative comparison on the MFI-WHU and SIMIF datasets

方法	MFI-WHU						SIMIF					
	NMI	EN	AG	SF	QY	QCB	NMI	EN	AG	SF	QY	QCB
IFCNN (Zhang等,2020)	0.8993	7.3285	11.3564	26.1798	0.9404	0.7367	1.0505	7.3897	6.9289	19.1344	0.9211	0.7364
SESF (Ma等,2021)	1.1878	7.3183	11.5399	26.5894	0.9855	0.8166	1.2995	7.3868	6.9931	19.4672	0.9807	0.8306
GACN (Ma等,2022)	1.2084	7.3127	11.4633	26.5089	0.9889	**0.8241**	1.3068	7.3802	6.9750	19.4241	**0.9845**	**0.8328**
GRFusion(Li等,2023)	0.7028	7.2763	10.6727	24.7632	0.8387	0.6512	0.8302	7.3942	7.0007	19.2939	0.8325	0.6259
ZMFF(Hu等,2023)	1.2134	7.3216	11.6609	26.8362	0.9848	0.8212	1.2898	7.3917	7.0215	19.4954	0.9771	0.8267
MUFusion(Cheng等,2023)	0.7449	**7.3744**	9.6015	21.4447	0.8608	0.6480	0.8876	7.3566	6.0633	16.3885	0.8387	0.6321
DB-MFIF(Zhang等,2024)	1.0693	7.3250	11.6956	26.8336	0.9466	0.7818	1.1510	7.4073	7.1849	19.7543	0.9157	0.7612
MFFT(Zhai等,2024)	1.1974	7.3358	**11.7636**	**27.5736**	0.9614	0.7648	1.2751	7.3866	7.1754	**20.5918**	0.9492	0.7724
DMANet(Quan等,2025)	1.2220	7.3158	11.6003	26.8034	0.9878	0.8222	1.3130	7.3834	7.0634	19.5735	0.9780	0.8270
MCCSR (Zheng等,2025)	**1.2246**	7.3120	11.4162	26.4025	**0.9891**	0.8227	**1.3151**	7.3782	6.9240	19.3575	0.9835	0.8298
NB Pro	0.5923	7.2986	10.5142	24.1274	0.5610	0.5745	0.8042	**7.4925**	**7.3561**	19.5545	0.5112	0.5775

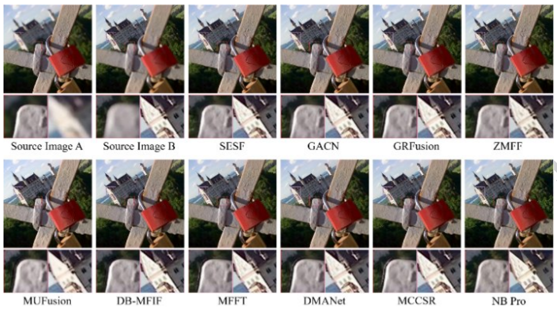


图 23　Lytro 数据集中 “lock” 样本的融合图像可视化结果
Fig. 23　Visualization of the fusion images of ‘lock’ sample from the Lytro dataset

14　红外与可见光图像融合

14.1　背景

红外与可见光图像融合(Infrared-Visible Image Fusion, IVIF)通过结合红外图像的热辐射信息与可见光图像的纹理细节,实现互补感知,可提升目标检测、自动驾驶和安防监控等任务的可靠性。传统方法如多尺度变换(Liu 等,2015)和稀疏表示(Li 等,2019)难以兼顾热目标显著性与背景纹理保真度,且易引入伪影。近年来,基于 CNN(Ma 等,2020;Tang 等,2022;Zhao 等,2021)、GAN(Li 等,2020;Ma 等,2019;Ma 等,2020)和 Transformer(Li 等,2022;Vs 等,2022;Wang 等,2022)的方法虽取得进展,但在跨模态对齐和细节保持方面仍有不足。

本文通过定量与定性分析,评估 NanoBanana 在 IVIF 任务中的融合质量、噪声控制和推理效率。

14.2　定量结果

参照文献(Zhao 等,2024),本文在 MSRS(multi-spectral road scenarios)(Tang 等,2022)、RoadScene (Xu 等,2020)和 M3FD(multi-scenario multi-modality benchmark to fuse infrared and visible for object detection)(Liu 等,2022)三个数据集上评估了 Nano Banana Pro 的 IVIF 性能,采用 EN、SD(standard deviation)、SF、AG 等无参考指标以及 SCD、VIF 等源参考指标,并与 10 种代表性方法进行比较。结果表明(表 17),NB Pro 在反映信息量和纹理细节的无参考指标上表现突出,尤其在 MSRS 和 RoadScene 上优势明显,说明其生成结果具有较高的清晰度和对比度,能够较好地重建高频边缘信息。相比之下,其在 SCD、VIF 等源参考指标上得分较低,反映出生成模型在提升感知质量的同时,也可能引入与源图像不完全一致的风格或像素偏差,而传统方法在像素级保真度上通常更稳定。

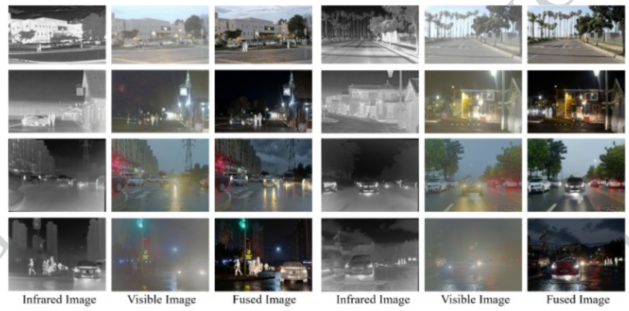


图 24　Nano Banana Pro 在 RoadScene 和 M3FD 数据集上的图像融合结果可视化示例
Fig. 24　Visualization examples of image fusion results by Nano Banana Pro in the RoadScene and M3FD datasets

此外,NB Pro 的性能具有一定场景依赖性:其在夜间或复杂光照条件下的 MSRS 和 RoadScene 上表现更优,而在 M3FD 上的整体优势有所减弱,说明其在特定多模态目标场景中仍有进一步优化空间。

14.3　定性结果

尽管 NB Pro 在多数场景中取得了较为理想的融合效果,但仍存在一定局限。在成功案例中,NB Pro 能在极端光照条件下有效融合红外与可见光信息:低光区域中的目标主体可以被清晰突出,热目标轮廓明确,背景照明得到增强,同时在过曝区域仍能恢复细节与纹理。

然而,也存在次优情况。在部分样本中,建筑结构出现过度锐化并伴随轻微过曝;另一些样本中,热目标周围产生明显光晕,表现为不自然的高亮边缘。图 24 展示了在其他数据集上的结果,整体而言,NB Pro 在目标突出和背景纹理重建方面表现良好,但在细粒度细节上偶尔引入幻觉或伪影,未能始终忠实保留源信息。

15　面向人类感知的实验

15.1　实验动机

尽管本文已基于 14 项底层视觉任务和 40 个数据集,从全参考、无参考及定性可视化等多个角度系统评估了 Nano Banana Pro 的性能,但实验表明,传统评价指标对生成式底层视觉模型仍存在局限。对

表 17 在 MSRS、RoadScene 和 M3FD 数据集上的定量对比结果
Table 17 Quantitative comparison on the MSRS, RoadScene, and M3FD datasets

方法	MSRS						RoadScene						M^{3}FD					
	EN	SD	SF	AG	SCD	VIF	EN	SD	SF	AG	SCD	VIF	EN	SD	SF	AG	SCD	VIF
SDN(Zhang等,2021)	5.25	17.35	8.67	2.67	0.99	0.50	7.30	44.06	14.58	5.80	1.37	0.61	6.87	36.22	15.32	5.61	1.41	0.55
TarD(Liu等,2022)	5.28	25.22	5.98	1.83	0.71	0.42	7.26	47.44	11.11	4.14	1.40	0.56	6.80	41.77	8.65	3.17	1.35	0.51
DeF (Liang等,2022)	6.46	37.63	8.60	2.80	1.35	0.77	7.36	47.03	10.99	4.38	1.62	0.63	6.90	36.81	9.85	3.65	1.42	0.58
Meta(Zhao等,2023)	5.65	24.97	9.99	3.40	1.14	0.31	6.88	31.97	14.38	5.57	0.92	0.55	6.73	30.56	16.48	6.02	1.31	0.65
CDDF (Zhao等,2023)	6.70	43.38	11.56	3.73	1.62	**1.05**	**7.52**	54.42	14.97	5.81	1.65	**0.66**	7.04	42.02	16.56	5.84	1.41	0.65
LRR (Li等,2023)	6.19	31.78	8.46	2.63	0.79	0.54	7.12	39.16	11.41	4.37	1.46	0.45	6.58	30.28	11.83	4.21	1.34	0.54
MURF (Xu等,2023)	5.04	16.37	8.31	2.67	0.86	0.40	6.91	33.34	13.88	5.37	1.04	0.52	6.59	28.89	11.82	4.81	1.21	0.39
DDFM(Zhao等,2023)	6.19	29.26	7.44	2.51	1.45	0.73	7.24	42.43	10.68	4.15	1.64	0.62	6.82	32.68	10.07	3.71	1.35	0.60
SegM(Liu等,2023)	5.95	37.28	11.10	3.47	1.57	0.88	7.29	46.14	14.47	5.57	1.61	0.65	6.88	36.20	16.19	5.83	1.38	**0.75**
EMMA (Zhao等,2024)	6.71	44.13	11.56	3.76	**1.63**	0.97	**7.52**	54.81	15.21	5.83	**1.69**	**0.66**	**7.12**	**44.01**	**16.92**	**6.23**	**1.48**	0.66
NB Pro	**6.85**	**44.95**	**14.39**	**4.56**	1.15	0.58	7.39	**56.98**	**21.81**	**7.07**	0.83	0.51	6.98	43.44	15.68	5.06	0.75	0.38

注:最佳结果以黑色加粗标出。

于以语义重建和视觉合理性为特征的基础生成模型而言,PSNR、SSIM 等强调像素对齐的指标难以充分反映其在自然度、可辨识性和整体观感上的实际表现。多个任务结果均显示,Nano Banana Pro 具有“像素保真度不足而感知质量较高”的双重特性,因此仅依赖传统客观指标难以完整刻画其真实价值。

基于此,本文进一步补充了面向人类感知的主观评价实验,从视觉体验角度验证 Nano Banana Pro 的输出质量。考虑到完整覆盖 14 项任务的主观测试成本较高,本文选取了运动模糊去除、单幅图像去雨、低光照增强、反射去除和真实图像超分辨率 5 个代表性任务作为补充验证对象,以较全面地反映其在底层视觉任务中的主观感知表现。

表 18 Nano Banana Pro 在五个基准任务进行人类感知评价方面的结果
Table 18 Human perceptual evaluation results of Nano Banana Pro on five benchmark tasks.

任务/指标	任务目标达成度	内容忠实度	关键信息辨识度	视觉自然性	伪影与局部一致性	任务平均分
去模糊	2.40	3.89	3.35	4.41	**4.46**	3.70
去雨	4.17	**4.52**	**4.56**	4.32	4.10	**4.33**
暗光增强	**4.26**	4.47	4.33	**4.35**	3.99	4.28
去眩光	3.76	4.18	4.25	4.17	3.74	4.02
超分辨率	3.02	2.90	3.25	3.38	3.08	3.13
指标平均分	3.52	3.99	3.95	4.13	3.87	3.89

15.2 人类感知评价维度与评分准则

本研究从人类最关注的五类感知属性出发,构建了一套通用主观评价框架,包括任务目标达成度、内容忠实度、关键信息辨识度、视觉自然性以及伪影与局部一致性。其分别用于衡量:模型是否完成核心任务、是否保持原始语义与结构、是否提升关键信息可辨识性、是否具备自然协调的整体观感,以及是否引入光晕、过锐化、纹理异常等副作用。所有维度均采用 1 – 5 Likert 量表评分,其中 1 分表示很差,5 分表示优秀。为便于统计,本文将评分不低于 4 记为“满意”,并进一步统计各任务在各维度上的满意率,从而同时反映模型在不同感知属性上的平均表

现与稳定性。

15.3 主观评价结果

表18显示，Nano Banana Pro在五个代表性任务上的主观表现存在明显差异：去雨最好，暗光增强次之，去眩光居中，而去模糊和超分辨率相对较弱；其综合主观感知得分为3.892，整体达到中上水平。

分维度看，视觉自然性得分最高，说明模型在整体观感、色彩协调和自然程度方面表现较好；任务目标达成度最低，表明其虽然“看起来较好”，但任务完成性仍不够稳定。

分任务来看，去雨在各维度上整体最强，暗光增强也表现较好；相比之下，去模糊和超分辨率波动更明显，尤其超分辨率在内容忠实度上表现较弱，反映出较明显的结构偏移和语义不稳定问题。

从满意率看，去雨、暗光增强和去眩光更容易获得主观认可，而去模糊和超分辨率满意率较低，说明Nano Banana Pro更适合以视觉改善和场景可用性为核心的任务，而在强调结构精确性和内容保持的任务中仍存在不足。

15.4 结果分析与启示

主观评价实验与前文大规模实验结论基本一致，进一步验证了Nano Banana Pro在底层视觉中的“双重特性”：其在视觉自然性、图像洁净度和整体观感方面具有优势，因此在去雨、暗光增强等任务中更容易获得较高主观评分；但在强调内容保真、结构一致性和任务精确完成的场景中，输出仍可能偏离原始内容，进而拉低任务目标达成度和内容忠实度。

综合以上结果，对生成式模型的评估不能仅仅依赖传统全参考指标，也不能单纯依据“看起来是否更清晰”做出判断。更合理的方式是同时结合客观保真度指标与主观感知评价，从“任务是否完成”“内容是否保持”“结果是否自然”三个层面进行综合分析。本文提出的五维人类感知评价框架，正是在这一背景下对现有评估范式的一种补充。它不仅有助于更准确地描述Nano Banana Pro这类生成模型的输出特征，也为后续研究建立更贴近人类视觉体验的评测协议提供了参考。

16 讨论

本研究通过对14项底层视觉任务的系统性零样本评估，揭示了Nano Banana Pro作为通用生成模型的双重特性：在感知质量上表现突出，但在传统像素保真度指标上明显落后。这一发现不仅量化了大规模生成模型在底层视觉中的能力边界，也促使我们反思任务定义、评估范式与模型演进路径。

16.1 生成式与回归范式

我们的研究突显了生成式与传统底层视觉模型之间的内在差异。传统方法主要遵循回归范式。它们通过像素级监督学习从退化输入到清晰参考的确定性映射，将其优化目标直接与PSNR和SSIM等指标对齐。相比之下，Nano Banana Pro体现了生成式范式。其训练核心涉及学习大规模图像数据的联合分布，并基于语义先验执行条件合成。其目标是生成看似合理且视觉上令人愉悦的图像，而不是与特定参考进行像素级对齐。因此，在信息严重丢失的区域，受限于输入信息的传统方法往往产生模糊或平淡的结果。然而，生成模型可以利用其强大的世界知识来幻觉生成看似合理的细节，产生主观上更优越的输出，但这构成了对标准真值的偏离。

16.2 传统指标的潜在误导性

我们的结果对全参考指标（如PSNR、SSIM）在底层视觉中的普遍适用性提出了挑战。这类基于像素差异的指标隐含地假设存在单一、像素级完美的真值，但这一前提并不适用于生成式修复：对于信息严重缺失区域，往往存在多种视觉上合理且语义正确的重建结果，而生成模型提供的这些合理解会被传统指标误判为错误。

此外，这类指标与人类感知并不总是一致。正如前文所示，Nano Banana Pro在无参考感知指标（如NIQE、NIMA）上表现优异，说明其输出在统计自然度和审美观感上更具优势；但合理的色彩调整、轻微去噪或细节增强，反而可能导致PSNR下降。与此同时，许多真实世界数据集中的真值图像本身仍含有残余噪声、轻微模糊或不理想的色彩平衡，使得生成模型输出更干净、更自然的结果时，反而被计为保真度损失。

因此，仅依赖传统全参考指标评价生成式底层视觉模型既不公平，也可能产生误导，这也进一步说明有必要建立新的评估框架。

16.3 Nano Banana Pro的操作范围与局限性

我们的评估清楚地界定了Nano Banana Pro的范围，这是由一个根本性的妥协所塑造的：它优先考虑语义合理性和视觉吸引力，而非精确的像素级保

真度。这使得该模型在创意和感知任务中非常有效，例如艺术图像增强、严重退化照片的修复，以及视觉效果比严格准确性更关键的场景。其无需专门训练即可执行这些任务的能力也使其成为快速原型设计的实用工具。

然而，这些能力伴随着固有的约束。该模型不适用于要求严格事实准确性的应用，包括法医检验、科学成像，或任何输出必须与原始场景数据精确对应的环境。其生成方法可能会引入改变，例如超分辨率中的边界柔化、去模糊中被改变的文本或非物理的色彩偏移，这些都优先考虑视觉完整性而非真实的再现。本质上，Nano Banana Pro 是用于常见视觉应用的强大语义重建器和增强器，但并非为严格保真度至关重要的高精度任务而设计。

16.4 未来研究方向

本研究表明，未来工作的重点主要集中在三个方面：生成与回归结合的混合架构、面向底层视觉的可控生成机制，以及适配生成模型的新评估框架。首先，未来的“全能选手”未必是纯粹的生成模型，更可能是生成—回归混合体：由前端回归模型先恢复基本结构、颜色和物理属性，再由后端生成模型在受约束条件下补充细节与提升感知质量，从而在保真度与视觉效果之间取得更好平衡。

其次，当前实验仅采用统一且简单的任务型提示词，虽然保证了公平性，但也意味着结果只是对模型能力的保守估计。未来应系统探索提示词工程、视觉线索引导和交互式细化，以减少色彩、结构和纹理上的不必要变异。

最后，生成模型的兴起也要求重新思考评估方式。传统依赖单一真值和像素级参考指标的框架，难以充分评价可产生多种合理重建结果的模型，因此亟需引入兼顾任务完成度、内容忠实度与感知质量的新型评估体系。

16.4.1 面向底层视觉的结构化约束提示

本文实验采用简单统一的任务型提示词，这保证了评估的公平性，但也限制了模型能力的进一步发挥。相比重新训练模型，构建面向底层视觉任务的结构化约束提示机制，可能是成本更低且更易见效的改进方向。

具体而言，提示词不应只是“去雾”“去雨”“超分”等单句命令，而应同时包含任务目标、不可改变项和允许增强项三部分：前者用于明确主任务，第二部分用于约束天气属性、文字语义、人物身份、场景边界和空间布局等关键内容保持不变，第三部分则限定模型只在局部可见性、纹理清晰度或雨纹抑制等方面进行增强，而不改变场景语义。

16.4.2 以保真度锚点为核心的混合复原架构

本文结果表明，未来鲁棒视觉系统的更优方向并非在生成范式与回归范式之间二选一，而是构建以保真度锚点为核心的混合复原架构：先由任务专用模型生成低风险、结构可信、物理属性稳定的初始结果，作为保真度锚点；再由生成模型在此基础上进行受约束的感知增强，仅在不确定区域补充细节和提升自然度，而非对整幅图像自由重建。

这样可以结合专用模型在像素对齐、几何边界和物理一致性上的稳定性，以及生成模型在严重退化条件下的语义补全能力，从而兼顾保真下限与感知上限。本文在去雾、超分、去雨和去阴影等任务中的实验现象均支持这一思路：专用模型更擅长维持结构与物理一致性，生成模型则在局部纹理恢复和严重退化区域补全方面更具优势。

16.4.3 三层式生成模型评估框架

本文已指出，依赖单一GT和像素级参考指标的传统评估方式，对生成式模型既不公平，也不充分。因此，未来评估可设计为三层结构：第一层为任务达成层，关注模型是否真正完成任务本身；第二层为内容忠实层，关注输出是否保持原始场景的关键内容与结构，可通过OCR一致性、边界保持率以及天气、光照和对象数量一致性等指标进行衡量；第三层为感知质量层，关注结果是否自然、清晰、协调、伪影少，并可结合本文提出的五维主观感知协议进行综合评价。

该框架的意义在于将“任务是否完成”“内容是否真实”“视觉是否自然”明确区分，从而避免生成模型仅凭主观观感掩盖内容失真，也避免专用模型仅凭像素指标优势掩盖感知质量不足。

17 结 论

对 Nano Banana Pro 的评估标志着一次范式转变：基础生成模型正在重新定义底层视觉的边界。该模型不再像传统复原工具那样运作，而更像是一个语义重建引擎，利用深度生成先验来合成视觉内容，而不仅仅是恢复像素。这一涌现挑战社区重新

思考成功的根本衡量标准:是绝对的像素保真度,还是感知合理性的最大化?

我们的实证结果证实,虽然 Nano Banana Pro 在零样本像素保真度上落后于特定领域的专家模型,但它在感知质量方面展现了非凡的潜力,尤其是在处理极端退化和跨任务泛化时。因此,该领域的发展轨迹不在于两种范式之间的二元选择,而在于战略整合。下一代鲁棒的视觉系统必须将生成模型的语义想象力与专用网络的物理约束和精度连接起来。

归根结底,Nano Banana Pro 是一把双刃剑。它成功提升了复杂视觉任务的感知质量上限,但尚未筑牢实现高精度所需的稳定下限

参考文献

A. Galdran, D. Pardo, A. Picon, and A. Alvarez-Gila. Automatic red-channel underwater image restoration. Journal of Visual Communication and Image Representation, 26:132 – 145, 2015.

A. S. A. Ghani and N. A. M. Isa. Underwater image quality enhancement through integrated color model with rayleigh distribution. Applied Soft Computing, 27:219 – 230, 2015.

Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In Proceedings of the IEEE conference on CVPR, pages 1692 – 1700, 2018.

Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In ECCV, pages 111 – 126. Springer, 2020.

Abdullah Abuolaim, Mahmoud Afifi, and Michael S Brown. Improving single-image defocus deblurring: How dual-pixel images help through multi-task learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 1231 – 1239, 2022.

Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. IEEE Signal processing letters, 20(3):209 – 212, 2012.

Armin Mehri, Parichehr B Ardakani, and Angel D Sappa. Mprnet: Multi-path residual network for lightweight image super resolution. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 2704 – 2713, 2021.

Bernd Jähne. Digital image processing. Springer, 2005.

Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In Proceedings of the IEEE/CVF ICCV, pages 13095 – 13105, 2023.

Bin Xiao, Haifeng Wu, and Xiuli Bi. Dtmnet: A discrete tchebichef moments-based deep neural network for multi-focus image fusion. In Proceedings of the IEEE/CVF ICCV, pages 43 – 51, 2021.

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. IEEE transactions on image processing, 28(1):492 – 505, 2018.

Boyuan Ma, Xiang Yin, Di Wu, Haokai Shen, Xiaojuan Ban, and Yu Wang. End-to-end learning for simultaneously generating decision map and multi-focus image fusion result. Neurocomputing, 470: 204 – 216, 2022.

Boyuan Ma, Yu Zhu, Xiang Yin, Xiaojuan Ban, Haiyou Huang, and Michele Mukeshimana. Sesf-fuse: An unsupervised deep model for multi-focus image fusion. Neural Computing and Applications, 33 (11):5793 – 5804, 2021.

C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert. Enhancing underwater images and videos by fusion. In Proc. IEEE/CVF CVPR, pages 81 – 88, 2012.

C. Li, J. Guo, B. Wang, R. Cong, Y. Zhang, and J. Wang. Single underwater image enhancement based on color cast removal and visibility restoration. Journal of Electronic Imaging, 25(3):033012, 2016.

C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert. Color balance and fusion for underwater image enhancement. IEEE Trans. Image Process., 27(1):379 – 393, 2018.

C. Zhao, W. Cai, C. Dong, and C. Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In 2024 IEEE/CVF Conference on CVPR, pages 8281 – 8291, Seattle, WA, USA, jun 2024. IEEE/CVF. doi: 10.1109/CVPR52729.2024.00813.

C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. IEEE Transactions on Image Processing, 25(12): 5664 – 5677, 2016.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. ECCV, 2014.

Chao Li, Yixiao Yang, Kun He, Stephen Lin, and John E Hopcroft. Single image reflection removal through cascaded refinement. In Proceedings of the IEEE/CVF conference on CVPR, pages 3565 – 3574, 2020.

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

Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based pseudo labeling and cooperative unfolding network. Advances in NeurIPS, 37: 97859 – 97883, 2024.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing, 29: 4376 - 4389, nov 2019. doi: 10.1109/TIP. 2019. 2955241.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Alykhan Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on CVPR, pages 4681 - 4690, 2017.

Chun-Chieh Tsai. Standard images for multifocus image fusion, 2025.

Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zeroreference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF conference on CVPR, pages 1780 - 1789, 2020.

Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In Proceedings of the IEEE/CVF conference on CVPR, pages 5812 - 5820, 2022.

Chunyang Cheng, Tianyang Xu, and Xiao-Jun Wu. Mufusion: A general unsupervised image fusion network based on memory unit. Information Fusion, 92:80 - 92, 2023.

Cui Yang, Jian-Qi Zhang, Xiao-Rui Wang, and Xin Liu. A novel similarity based quality metric for image fusion. Information Fusion, 9(2):156 - 160, 2008.

Dana Berman, Shai Avidan, et al. Non-local image dehazing. In Proceedings of the IEEE conference on CVPR, pages 1674 - 1682, 2016.

Daniyar Zakarin, Thiemo Wandel, Anton Obukhov, and Dengxin Dai. Reflection removal through efficient adaptation of diffusion transformers. arXiv preprint arXiv:2512.05000, 2025.

Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE/CVF conference on CVPR, pages 3937 - 3946, 2019.

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE conference on CVPRW, pages 126 - 135, 2017.

Feng Zhang, Haoyou Deng, Zhiqiang Li, Lida Li, Bin Xu, Qingbo Lu, Zisheng Cao, Minchen Wei, Changxin Gao, Nong Sang, et al. High-resolution photo enhancement in real-time: A laplacian pyramid network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025.

Fengyi Zhang, Hui Zeng, Tianjun Zhang, and Lin Zhang. Clut-net: Learning adaptively compressed representations of 3dluts for lightweight image enhancement. In Proceedings of the 30th ACM International Conference on Multimedia, pages 6493 - 6501. ACM, 2022.

Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In ECCV, pages 146 - 162. Springer, 2022.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. arXiv preprint arXiv: 2312.11805, 2023.

Guangmang Cui, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. Optics Communications, 341:199 - 209, 2015.

H. Li, J. Li, and W. Wang. A Fusion Adversarial Underwater Image Enhancement Network with a Public Test Dataset. arXiv preprint, 2019. doi: 10.48550/arXiv.1906.06819. arXiv:1906.06819.

Han Xu, Jiayi Ma, Zhuliang Le, Junjun Jiang, and Xiaojie Guo. Fusiondn: A unified densely connected network for image fusion. In Proceedings of the AAAI conference on artificial intelligence, volume 34, pages 12484 - 12491, 2020.

Han Xu, Jiteng Yuan, and Jiayi Ma. Murf: Mutually reinforcing multi-modal image registration and fusion. IEEE transactions on pattern analysis and machine intelligence, 45(10):12148 - 12166, 2023.

Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF conference on CVPR, pages 2157 - 2167, 2020.

Hanshu Yan, Jingfeng Zhang, Jiashi Feng, Masashi Sugiyama, and Vincent YF Tan. Towards adversarially robust deep image denoising. arXiv preprint arXiv:2201.04397, 2022.

Hao Tang, Chengcheng Yuan, Zechao Li, and Jinhui Tang. Learning attention-guided pyramidal features for few-shot fine-grained recognition. Pattern Recognition, 130:108792, 2022.

Hao Zhai, Wenyi Zheng, Yuncan Ouyang, Xin Pan, and Wanli Zhang. Multi-focus image fusion via interactive transformer and asymmetric soft sharing. Engineering Applications of Artificial Intelligence, 133:107967, 2024.

Hao Zhang and Jiayi Ma. Sdnet: A versatile squeeze-and-decomposition network for real-time image fusion.IJCV, 129(10):2761 - 2785, 2021.

Hao Zhang, Zhuliang Le, Zhenfeng Shao, Han Xu, and Jiayi Ma. Mff-gan: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion. Information Fusion, 66:40 - 53, 2021.

Hao Zhao, Mingjia Li, Qiming Hu, and Xiaojie Guo. Reversible decoupling network for single image reflection removal. In Proceedings of the CVPR Conference, pages 26430 - 26439, 2025.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. Advances in NeurIPS, 36: 34892 - 34916, 2023.

Haoyou Deng, Lida Li, Feng Zhang, Zhiqiang Li, Bin Xu, Qingbo Lu, Changxin Gao, and Nong Sang. Towards blind flare removal using knowledge-driven flare-level estimator. IEEE Transactions on Image Processing, 2024.

Haoyu Chen, Jinjin Gu, Yihao Liu, Salma Abdel Magid, Chao Dong, Qiong Wang, Hanspeter Pfister, and Lei Zhu. Masked image training for generalizable deep image denoising. In Proceedings of the IEEE/CVF Conference on CVPR, pages 1692 – 1703, 2023.

Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In Proceedings of the IEEE/CVF conference on CVPR, pages 3103 – 3112, 2020.

Huafeng Li, Dan Wang, Yuxin Huang, Yafei Zhang, and Zhengtao Yu. Generation and recombination for multifocus image fusion with free number of inputs. IEEE Transactions on Circuits and Systems for Video Technology, 34(7):6009 – 6023, 2023.

Huafeng Li, Yitang Wang, Zhao Yang, Ruxin Wang, Xiang Li, and Dapeng Tao. Discriminative dictionary learning-based multiple component decomposition for detail-preserving noisy image fusion. IEEE Transactions on Instrumentation and Measurement, 69(4): 1082 – 1102, 2019.

Hui Li, Tianyang Xu, Xiao-Jun Wu, Jiwen Lu, and Josef Kittler. Lrrnet: A novel representation learning guided fusion network for infrared and visible images. IEEE transactions on pattern analysis and machine intelligence, 45(9):11040 – 11052, 2023.

Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(4):2058 – 2073, 2020.

Hyeongseok Son, Junyong Lee, Sunghyun Cho, and Seungyong Lee. Single image defocus deblurring using kernel-sharing parallel atrous convolutions. In Proceedings of the IEEE/CVF ICCV, pages 2642 – 2650, 2021.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in NeurIPS, 2014.

Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In ECCV, pages 184 – 201. Springer, 2020.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed selfexemplars. In Proceedings of the IEEE conference on CVPR, pages 5197 – 5206, 2015.

Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image superresolution: A new benchmark and a new model. In Proceedings of the IEEE/CVF ICCV, pages 3086 – 3095, 2019.

Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing, 27(4):2049 – 2062, 2018.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pages 2555 – 2563, 2023.

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution.IJCV, pages 1 – 21, 2024.

Jiayi Ma, Han Xu, Junjun Jiang, Xiaoguang Mei, and Xiao-Ping Zhang. Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing, 29: 4980 – 4995, 2020.

Jiayi Ma, Linfeng Tang, Fan Fan, Jun Huang, Xiaoguang Mei, and Yong Ma. Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica, 9(7):1200 – 1217, 2022.

Jiayi Ma, Pengwei Liang, Wei Yu, Chen Chen, Xiaojie Guo, Jia Wu, and Junjun Jiang. Infrared and visible image fusion via detail preserving adversarial learning. Information Fusion, 54: 85 – 98, 2020.

Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. Fusiongan: A generative adversarial network for infrared and visible image fusion. Information fusion, 48:11 – 26, 2019.

Jichen Hu, Chen Yang, Zanwei Zhou, Jiemin Fang, Xiaokang Yang, Qi Tian, and Wei Shen. Dereflection any image with diffusion priors and diversified data. arXiv preprint arXiv:2503.17347, 2025.

Jie Cai, Kangning Yang, Ling Ouyang, Lan Fu, Jiaming Ding, Huiming Sun, Chiu Man Ho, and Zibo Meng. F2t2-hit: A u-shaped fft transformer and hierarchical transformer for reflection removal. arXiv preprint arXiv:2506.05489, 2025.

Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. IEEE transactions on pattern analysis and machine intelligence, 45(11):12978 – 12995, 2022.

Jie Xiao, Xueyang Fu, Yurui Zhu, Dong Li, Jie Huang, Kai Zhu, and Zheng-Jun Zha. Homoformer: Homogenized transformer for image shadow removal. In Proceedings of the IEEE/CVF conference on CVPR, pages 25617 – 25626, 2024.

Jie Yang, Dong Gong, Lingqiao Liu, and Qinfeng Shi. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In Proceedings of the ECCV, pages 654 – 669, 2018.

Jing Li, Hongtao Huo, Chang Li, Renhua Wang, and Qi Feng. Attentionfgan: Infrared and visible image fusion using attention-based generative adversarial networks. IEEE Transactions on Multimedia, 23:1383 – 1396, 2020.

Jing Li, Jianming Zhu, Chang Li, Xun Chen, and Bin Yang. Cgtf: Convolution-guided transformer for infrared and visible image fusion. IEEE Transactions on Instrumentation and Measurement, 71:1 – 14, 2022.

Jingwen He, Yihao Liu, Yu Qiao, and Chao Dong. Conditional sequential modulation for efficient global image retouching. In ECCV, pages 679 – 695. Springer, 2020.

Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and

Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. Advances in NeurIPS, 36: 79734 – 79747, 2023.

Jinxing Li, Xiaobao Guo, Guangming Lu, Bob Zhang, Yong Xu, Feng Wu, and David Zhang. Drpl: Deep regression pair learning for multi-focus image fusion. IEEE Transactions on Image Processing, 29:4816 – 4831, 2020.

Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In Proceedings of the IEEE/CVF conference on CVPR, pages 5802 – 5811, 2022.

Jinyuan Liu, Zhu Liu, Guanyao Wu, Long Ma, Risheng Liu, Wei Zhong, Zhongxuan Luo, and Xin Fan. Multiinteractive feature learning and a full-time multi-modality benchmark for image fusion and segmentation. In Proceedings of the IEEE/CVF ICCV, pages 8115 – 8124, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In Advances in NeurIPS, volume 33, pages 6840 – 6851, 2020.

Jun Xu, Hui Li, Zhetong Liang, David Zhang, and Lei Zhang. Real-world noisy image denoising: A new benchmark. arXiv preprint arXiv:1804.02603, 2018.

Juncheng Zhang, Qingmin Liao, Haoyu Ma, Jing-Hao Xue, Wenming Yang, and Shaojun Liu. Exploit the best of both end-to-end and map-based methods for multi-focus image fusion. IEEE Transactions on Multimedia, 26: 6411 – 6423, 2024.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In IEEE/CVF ICCV, pages 5148 – 5157, 2021.

Junyong Lee, Hyeongseok Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee. Iterative filter adaptive network for single image defocus deblurring. In Proceedings of the IEEE/CVF conference on CVPR, pages 2034 – 2042, 2021.

K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib. Enhancing the low quality images using unsupervised colour correction method. In IEEE International Conference on SMC, pages 1703 – 1709, 2010.

K. Panetta, C. Gao, and S. Agaian. Human-visual-system-inspired underwater image quality measures. IEEE Journal of Oceanic Engineering, 41 (3): 541 – 551, 2015. doi: 10.1109/JOE. 2015. 2410644.

Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In Proceedings of the IEEE/CVF ICCV, pages 4791 – 4800, 2021.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing, 26(7): 3142 – 3155, 2017.

Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In Proceedings of the IEEE/CVF conference on CVPR, pages 2737 – 2746, 2020.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence, 33(12):2341 – 2353, 2010.

Kaixuan Wei, Jiaolong Yang, Ying Fu, David Wipf, and Hua Huang. Single image reflection removal exploiting misaligned training data and network enhancements. In Proceedings of the IEEE/CVF Conference on CVPR, pages 8178 – 8187, 2019.

Kecheng Zheng, Juan Cheng, and Yu Liu. Unfolding coupled convolutional sparse representation for multi-focus image fusion. Information Fusion, 118:102974, 2025.

Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In Proceedings of the IEEE/CVF conference on CVPR, pages 8346 – 8355, 2020.

L. Peng, C. Zhu, and L. Bian. U-shape transformer for underwater image enhancement. IEEE Transactions on Image Processing, 32: 3066 – 3079, 2023. doi: 10.1109/TIP.2023.3276332.

Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In Proceedings of the IEEE/CVF conference on CVPR, pages 14049 – 14058, 2023.

Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: Global context helps shadow removal. In Proceedings of the AAAI conference on artificial intelligence, volume 37, pages 710 – 718, 2023.

Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. Journal of Electronic imaging, 20(2):023016 – 023016, 2011.

Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multicontext embedding deep network for shadow removal. In Proceedings of the IEEE conference on CVPR, pages 4067 – 4075, 2017.

Linfeng Tang, Jiteng Yuan, Hao Zhang, Xingyu Jiang, and Jiayi Ma. Piafusion: A progressive infrared and visible image fusion network based on illumination aware. Information Fusion, 83: 79 – 92, 2022.

Lingyan Ruan, Bin Chen, Jizhou Li, and Miu-Ling Lam. Aifnet: All-in-focus image restoration network using a light field-based dataset. IEEE Transactions on Computational Imaging, 7:675 – 688, 2021.

Lingyan Ruan, Bin Chen, Jizhou Li, and Miuling Lam. Learning to deblur using light field generated and real defocus images. In Proceedings of the IEEE/CVF conference on CVPR, pages 16304 –

16313, 2022.

M. Yang and A. Sowmya. An underwater color image quality evaluation metric. IEEE Transactions on Image Processing, 24(12):6062 - 6071, 2015. doi: 10.1109/TIP.2015.2480136.

Mansour Nejati, Shadrokh Samavi, and Shahram Shirani. Multi-focus image fusion using dictionary-based sparse representation. Information fusion, 25:72 - 84, 2015.

Matthias Hullin, Elmar Eisemann, Hans-Peter Seidel, and Sungkil Lee. Physically-based real-time lens flare rendering. ACM Trans. Graph., 30(4), 2011.

Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. ACM Transactions on Graphics (TOG), 36(4):1 - 12, 2017.

Mining Li, Ronghao Pei, Tianyou Zheng, Yang Zhang, and Weiwei Fu. Fusiondiff: Multi-focus image fusion using denoising diffusion probabilistic models. Expert Systems with Applications, 238: 121664, 2024.

Mohammed Hossny, Saeid Nahavandi, and Douglas Creighton. Comments on 'information measure for performance of image fusion'. Electronics letters, 44(18):1066 - 1067, 2008.

Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-ofmagnitude) faster and better. In Proceedings of the IEEE/CVF ICCV, pages 8878 - 8887, 2019.

Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE conference on CVPR, pages 8183 - 8192, 2018.

P. L. Drews, E. R. Nascimento, S. S. Botelho, and M. F. M. Campos. Underwater depth estimation and image restoration based on single images. IEEE Computer Graphics and Applications, 36(2):24 - 35, 2016.

Pengwei Liang, Junjun Jiang, Xianming Liu, and Jiayi Ma. Fusion from decomposition: A self-supervised decomposition approach for image fusion. In ECCV, pages 719 - 735. Springer, 2022.

Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In ECCV, pages 101 - 117, 2020.

Qiaosi Yi, Juncheng Li, Qinyan Dai, Faming Fang, Guixu Zhang, and Tieyong Zeng. Structure-preserving deraining with residue channel prior guidance. In Proceedings of the IEEE/CVF ICCV, pages 4238 - 4247, 2021.

Qiming Hu and Xiaojie Guo. Single image reflection separation via component synergy. In Proceedings of the IEEE/CVF ICCV, pages 13138 - 13147, 2023.

Qiming Hu and Xiaojie Guo. Trash or treasure? an interactive dual-stream strategy for single image reflection separation. Advances in NeurIPS, 34:24683 - 24694, 2021.

Qiming Hu, Hainuo Wang, and Xiaojie Guo. Single image reflection separation via dual-stream interactive transformers. Advances in NeurIPS, 37:55228 - 55248, 2024.

R. Cong, W. Yang, W. Zhang, C. Li, C.-L. Guo, Q. Huang, and S. Kwong. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. IEEE Transactions on Image Processing, 32: 4472 - 4485, 2023.

Raanan Fattal. Dehazing using color-lines. ACM transactions on graphics (TOG), 34(1):1 - 14, 2014.

Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Benchmarking single-image reflection removal algorithms. In Proceedings of the IEEE ICCV, pages 3922 - 3930, 2017.

Rich Franzen. Kodak lossless true color image suite, volume 5. https://r0k.us/graphics/kodak/, 1999.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on CVPR, pages 586 - 595, 2018.

Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF conference on CVPR, pages 10561 - 10570, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In IEEE/CVF Conference on CVPR, pages 10684 - 10695, 2022.

Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. IEEE transactions on pattern analysis and machine intelligence, 35(12):2956 - 2967, 2012.

Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In Proceedings of the IEEE/CVF conference on CVPR, pages 22282 - 22291, 2023.

Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In Proceedings of the IEEE/CVF Conference on CVPR, pages 6849 - 6857, 2019.

S.-B. Gao, M. Zhang, Q. Zhao, X.-S. Zhang, and Y.-J. Li. Underwater image enhancement using adaptive retinal mechanisms. IEEE Trans. Image Process., 28(11):5580 - 5595, 2019.

Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. Deeplpf: Deep local parametric filters for image enhancement. In Proceedings of the IEEE/CVF Conference on CVPR, pages 12826 - 12835, 2020.

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 CVPR, pages 3883 - 3891, 2017.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo

Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. arXiv preprint arXiv: 2502.13923, 2025.

Shuang Xu, Xiaoli Wei, Chunxia Zhang, Junmin Liu, and Jiangshe Zhang. Mffw: A new dataset for multi-focus image fusion. arXiv preprint arXiv:2002.04780, 2020.

Simin Luan, Cong Yang, Zeyd Boukhers, Xue Qin, Dongfeng Cheng, Wei Sui, and Zhijun Li. Gyroscope-assisted motion deblurring network. CoRR, 2024.

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

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

Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In Proceedings of the AAAI conference on artificial intelligence, volume 37, pages 2654 – 2662, 2023.

Tao Wang, Yong Li, Jingyang Peng, Yipeng Ma, Xian Wang, Fenglong Song, and Youliang Yan. Real-time image enhancer via learnable spatial-aware 3d lookup tables. In Proceedings of the IEEE/CVF ICCV, pages 2471 – 2480, 2021.

Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In The IEEE Conference on CVPR, June 2019.

Vibashan Vs, Jeya Maria Jose Valanarasu, Poojan Oza, and Vishal M Patel. Image fusion transformer. In 2022 IEEE International conference on ICIP, pages 3566 – 3570. IEEE, 2022.

Vineeth Murali and PV Sudeep. Image denoising using dncnn: An exploration study. In Advances in Communication Systems and Networks: Select Proceedings of ComNet 2019, pages 847 – 859. Springer, 2020.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Fredo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In CVPR 2011, pages 97 – 104, 2011. doi: 10.1109/ CVPR.2011.5995332.

W. Zhang, Y. Wang, and C. Li. Underwater image enhancement by attenuated color channel correction and detail preserved contrast enhancement. IEEE Journal of Oceanic Engineering, pages 1 – 18, 2022.

Wenda Zhao, Shigeng Xie, Fan Zhao, You He, and Huchuan Lu. Metafusion: Infrared and visible image fusion via meta-feature embedding from object detection. In Proceedings of the IEEE/CVF Conference on CVPR, pages 13955 – 13965, 2023.

Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In Proceedings of the IEEE conference on CVPR, pages 1357 – 1366, 2017.

Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing, 30:2072 – 2086, 2021.

Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In Proceedings of the IEEE/CVF conference on CVPR, pages 26057 – 26066, 2024.

Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In Proceedings of the ECCV, pages 254 – 269, 2018.

Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In Proceedings of the IEEE/CVF conference on CVPR, pages 5896 – 5905, 2023.

Xiang Chen, Jinshan Pan, and Jiangxin Dong. Bidirectional multi-scale implicit neural representations for image deraining. In Proceedings of the IEEE/CVF Conference on CVPR , June 2024.

Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In Proceedings of the AAAI conference on artificial intelligence, volume 34, pages 10680 – 10687, 2020.

Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In Proceedings of the IEEE/CVF conference on CVPR, pages 17714 – 17724, 2022.

Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. IJCV, 131(1):48 – 66, 2023.

Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In Proceedings of the IEEE conference on CVPR, pages 7454 – 7462, 2018.

Xiaoyu Li, Bo Zhang, Jing Liao, and Pedro V. Sander. Let's see clearly: Contaminant artifact removal for moving cameras. In 2021 IEEE/CVF, ICCV, pages 1991 – 2000, 2021. doi: 10.1109/ICCV48922.2021.00202.

Xin Li, Bingchen Li, Xin Jin, Cuiling Lan, and Zhibo Chen. Learning distortion invariant representation for image restoration from a causality perspective. In Proceedings of the IEEE/CVF Conference on CVPR, pages 1714 – 1724, 2023.

Xingyu Hu, Junjun Jiang, Xianming Liu, and Jiayi Ma. Zmff: Zero-shot multi-focus image fusion. Information Fusion, 92:127 – 138, 2023.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. In arXiv preprint arXiv:2308.15070, 2023.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF ICCV Workshops, pages 1905 – 1914, 2021.

Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In Proceedings of the IEEE conference on CVPR, pages 4786 – 4794, 2018.

Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In Proceedings of the IEEE conference on CVPR, pages 3855 – 3863, 2017.

Xueyang Fu, Qi Qi, Zheng-Jun Zha, Yurui Zhu, and Xinghao Ding. Rain streak removal via dual graph convolutional network. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 1352 – 1360, 2021.

Yael Vinker, Inbar Huberman-Spiegelglas, and Raanan Fattal. Unpaired learning for high dynamic range image tone mapping. In 2021 IEEE/CVF ICCV , pages 14637 – 14646, 2021. doi: 10.1109/ICCV48922.2021.01439.

Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In Proceedings of the IEEE/CVF conference on CVPR, pages 2037 – 2046, 2022.

Yanzuo Lu, Xin Xia, Manlin Zhang, Huafeng Kuang, Jianbin Zheng, Yuxi Ren, and Xuefeng Xiao. Hyper-bagel: A unified acceleration framework for multimodal understanding and generation. arXiv preprint arXiv:2509.18824, 2025.

Yi Tang, Hiroshi Kawasaki, and Takafumi Iwaguchi. Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In Proceedings of the 31st ACM International Conference on Multimedia (MM '23), pages 5419 – 5427, New York, NY, USA, 2023. Association for Computing Machinery. doi: 10.1145/3581783.3612475.

Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T. Barron. How to train neural networks for flare removal. In 2021 IEEE/CVF ICCV, pages 2219 – 2227, 2021. doi: 10.1109/ICCV48922.2021.00224.

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.

Yihang Huang, Yuanfei Huang, Junhui Lin, and Hua Huang. Deflaremamba: Hierarchical vision mamba for contextually consistent lens flare removal. In Proceedings of the 33rd ACM International Conference on Multimedia, page 8028 – 8037, 2025.

Yin Chen and Rick S Blum. A new automated quality assessment algorithm for image fusion. Image and vision computing, 27(10):1421 – 1432, 2009.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In Proceedings of the IEEE conference on CVPR, pages 6228 – 6237, 2018.

Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM international conference on multimedia, pages 1632 – 1640, 2019.

Yu Li, Ming Liu, Yaling Yi, Qince Li, Dongwei Ren, and Wangmeng Zuo. Two-stage single image reflection removal with reflection-aware guidance. Applied Intelligence, 53(16): 19433 – 19448, 2023.

Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In Proceedings of the IEEE conference on CVPR, pages 2736 – 2744, 2016.

Yu Liu, Shuping Liu, and Zengfu Wang. A general framework for image fusion based on multi-scale transform and sparse representation. Information fusion, 24:147 – 164, 2015.

Yu Liu, Xun Chen, Hu Peng, and Zengfu Wang. Multi-focus image fusion with a deep convolutional neural network. Information Fusion, 36:191 – 207, 2017.

Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In Proceedings of the IEEE ICCV, pages 3397 – 3405, 2015.

Yu Zhang, Yu Liu, Peng Sun, Han Yan, Xiaolin Zhao, and Li Zhang. Ifcnn: A general image fusion framework based on convolutional neural network. Information Fusion, 54:99 – 118, 2020.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In Proceedings of the IEEE/CVF ICCV, pages 12504 – 12513, 2023.

Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In Proceedings of the IEEE/CVF conference on CVPR, pages 2808 – 2817, 2020.

Yuchen Hong, Haofeng Zhong, Shuchen Weng, Jinxiu Liang, and Boxin Shi. L-differ: Single image reflection removal with language-based diffusion model. In ECCV, pages 58 – 76. Springer, 2024.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. IEEE Transactions on Image Processing, 32:1927 – 1941, 2023.

Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. Uiec2-net: Cnn-based underwater image enhancement using two color space. Signal Processing: Image Communication, 96: 116250, 2021. ISSN 0923-5965.

Yue Huang, Zi'ang Li, Tianle Hu, Jie Wen, Guanbin Li, Jinglin Zhang, Guoxu Zhou, and Xiaozhao Fang. Single image reflection removal via inter-layer complementarity. arXiv preprint arXiv:2505.12641, 2025.

Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Yihang Luo, and Chen Change Loy. Flare7k++: Mixing synthetic and real datasets for nighttime flare removal and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(11): 7041 - 7055, 2024. doi: 10.1109/TPAMI.2024.3406821.

Yuekun Dai, Dafeng Zhang, Xiaoming Li, Zongsheng Yue, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Peiqing Yang, Zhezhu Jin, Guanqun Liu, and Chen Change Loy. Mipi 2024 challenge on nighttime flare removal: Methods and results. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 1144 - 1152, 2024. doi: 10.1109/CVPRW63382.2024.00121.

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.

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

Yufeng Zheng, Edward A Essock, Bruce C Hansen, and Andrew M Haun. A new metric based on extended spatial frequency and its application to dwt based fusion algorithms. Information Fusion, 8(2):177 - 192, 2007.

Yuhui Quan, Xi Wan, Zitao Tang, Jinxiu Liang, and Hui Ji. Multi-focus image fusion via explicit defocus blur modelling. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, pages 6657 - 6665, 2025.

Yuhui Quan, Xin Yao, and Hui Ji. Single image defocus deblurring via implicit neural inverse kernels. In Proceedings of the IEEE/CVF ICCV, pages 12600 - 12610, 2023.

Yuhui Quan, Zicong Wu, and Hui Ji. Gaussian kernel mixture network for single image defocus deblurring. Advances in NeurIPS, 34: 20812 - 20824, 2021.

Yuhui Quan, Zicong Wu, and Hui Ji. Neumann network with recursive kernels for single image defocus deblurring. In Proceedings of the IEEE/CVF conference on CVPR, pages 5754 - 5763, 2023.

Yuhui Quan, Zicong Wu, Ruotao Xu, and Hui Ji. Deep single image defocus deblurring via gaussian kernel mixture learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In Proceedings of the ECCV, pages 286 - 301, 2018.

Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In Proceedings of the IEEE/CVF Conference on CVPR, pages 5627 - 5636, 2022.

Yurui Zhu, Xueyang Fu, Peng-Tao Jiang, Hao Zhang, Qibin Sun, Jinwei Chen, Zheng-Jun Zha, and Bo Li. Revisiting single image reflection removal in the wild. In Proceedings of the IEEE/CVF Conference on CVPR, pages 25468 - 25478, 2024.

Yuwei Qiu, Kaihao Zhang, Chenxi Wang, Wenhan Luo, Hongdong Li, and Zhi Jin. Mb-taylorformer: Multibranch efficient transformer expanded by taylor formula for image dehazing. In Proceedings of the IEEE/CVF ICCV, pages 12802 - 12813, 2023.

Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In Proceedings of the IEEE/CVF conference on CVPR, pages 7180 - 7189, 2021.

Zhaohan Wang, Chengjun Chen, and Chenggang Dai. Zero-shot realistic image deblurring with consistency model. Complex & Intelligent Systems, 12(1):29, 2026.

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 CVPR, pages 17683 - 17693, 2022.

Zheng Chen, Yulun Zhang, Ding Liu, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Hierarchical integration diffusion model for realistic image deblurring. Advances in NeurIPS, 36:29114 - 29125, 2023.

Zheng Dong, Ke Xu, Yin Yang, Hujun Bao, Weiwei Xu, and Rynson WH Lau. Location-aware single image reflection removal. In Proceedings of the IEEE/CVF ICCV, pages 5017 - 5026, 2021.

Zhengyang Lu, Weifan Wang, Tianhao Guo, and Feng Wang. Single-image reflection removal via self-supervised diffusion models. The Journal of Supercomputing, 81(1):338, 2025.

Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In Proceedings of the IEEE/CVF conference on CVPR, pages 22252 - 22261, 2023.

Zhishe Wang, Yanlin Chen, Wenyu Shao, Hui Li, and Lei Zhang. Swinfuse: A residual swin transformer fusion network for infrared and visible images. IEEE Transactions on Instrumentation and Measurement, 71:1 - 12, 2022.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4): 600 - 612, 2004.

Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Image restoration with mean-reverting stochastic differential equations. In Proceedings of the 40th International Conference on Machine Learning, pages 23045 - 23066, 2023.

Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Kai Zhang, Shuang Xu, Dongdong Chen, Radu Timofte, and Luc Van Gool. Equivariant multi-modality image fusion. In Proceedings of the IEEE/CVF conference on CVPR, pages 25912 - 25921, 2024.

Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, and Luc Van Gool. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In Proceedings of the IEEE/CVF conference on CVPR, pages 5906 – 5916, 2023.

Zixiang Zhao, Haowen Bai, Yuanzhi Zhu, Jiangshe Zhang, Shuang Xu, Yulun Zhang, Kai Zhang, Deyu Meng, Radu Timofte, and Luc Van Gool. Ddfm: denoising diffusion model for multi-modality image fusion. In Proceedings of the IEEE/CVF ICCV, pages 8082 – 8093, 2023.

Zixiang Zhao, Shuang Xu, Jiangshe Zhang, Chengyang Liang, Chunxia Zhang, and Junmin Liu. Efficient and model-based infrared and visible image fusion via algorithm unrolling. IEEE Transactions on Circuits and Systems for Video Technology, 32(3): 1186 – 1196, 2021.

Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Humanaware motion deblurring. In Proceedings of the IEEE/CVF ICCV, pages 5572 – 5581, 2019.

作者简介

左嘉龙,男,博士研究生,主要研究方向为计算机视觉、行人重识别和多模态理解/生成。Email: jlongzuo@hust.edu.cn

高常鑫,通信作者,男,教授,主要研究方向为计算机视觉和图像/视频理解。E-mail: cgao@hust.edu.cn

邓昊优,男,博士研究生,主要研究方向为计算机视觉和图像处理/生成。Email: haoyoudeng@hust.edu.cn

左皓童,男,本科生,主要研究方向为多智能体及其应用。Email: u202315328@hust.edu.cn

周寒宇,男,博士研究生,主要研究方向为计算机视觉和多模态对齐。Email: hanyzhou@hust.edu.cn

祝家心,女,硕士研究生,主要研究方向为计算机视觉和图像处理/生成。Email: Jiaxin_zhu@hust.edu.cn

张祎程,男,硕士研究生,主要研究方向为计算机视觉和图像处理/生成。Email: yicheng_zhang@hust.edu.cn

张艺巍,男,硕士研究生,主要研究方向为计算机视觉和图像处理/生成。Email: yiweiz@hust.edu.cn

严永昕,男,硕士研究生,主要研究方向为计算机视觉和图像处理/生成。Email: yongx_yan@hust.edu.cn

黄凯星,男,硕士研究生,主要研究方向为计算机视觉和红外图像处理。Email: huangkaixing@hust.edu.cn

陈伟森,男,硕士研究生,主要研究方向为计算机视觉与小样本学习。Email: wschen@hust.edu.cn

邓永泰,男,硕士研究生,主要研究方向为计算机视觉和行人重识别。Email: yongtai@hust.edu.cn

金睿,男,硕士研究生,主要研究方向为计算机视觉和长视频理解。Email: jrui@hust.edu.cn

张锋,男,博士,主要研究方向为计算摄影与计算机视觉。Email: keith.zhang@dji.com

桑农,男,教授,主要研究方向为模式识别和计算机视觉。E-mail:nsang@hust.edu.cn