PlaSo: Unsupervised Residual Plug and Play for Image Super-Resolution

Xuan-Hieu Le, Denis Kouamé, Duong-Hung Pham

PlaSo: Unsupervised Residual **Pl**ug and Pl**a**y for Image **S**uper-Res**o**lution

Xuan Hieu Le, Denis Kouamé and Duong Hung Pham

***Abstract*—This letter addresses the problem of single image super resolution using Plug-and-play (PnP) frameworks incorporating with residual learning. We propose a novel PnP algorithm, PlaSo, which integrates residual learning mechanisms into diffusion processes to mitigate common challenges such as hallucination artifacts and over-smoothing. Furthermore, by initializing the reverse sampling process from the degraded input rather than random noise, PlaSo achieves faster convergence. Experiments demonstrate that PlaSo consistently outperforms state-of-the-art methods, both qualitatively and quantitatively. Our code is available at: https://gitlab.irit.fr/minds/theseeric/plaso.**

***Index Terms*—Plug-and-play, diffusion, super resolution**

I. Introduction

Single image super-resolution (SISR) remains a long-standing challenge in image processing, focusing on recovering high-resolution (HR) images from their degraded low-resolution (LR) counterparts. SISR methods can be categorized into model-based and deep learning (DL)-based methods. Model-based methods rely on handcrafted regularization terms [1], [2], which are insufficient to capture complex structures of natural image. DL approaches [3], [4], [5] typically achieve superior reconstruction quality by leveraging strong data-driven priors, but require paired training data and often generalize poorly to unseen degradations [6].

In this context, PnP framework has emerged as a promising solution for SISR, effectively combining the complementary strengths of both classical model-based and DL approaches while addressing their respective limitations [7]. By decoupling the data fidelity from the prior term, the PnP framework leverages powerful trained DL denoisers, thus achieving superior performance over classical priors [8] while maintaining stronger robustness to degradation variations than DL methods [6] and remains training-free. Notably, several recent PnP-based methods [9], [10], [11] have adopted diffusion models (DMs) into their frameworks, thanks to DMs strong generative capabilities over conventional denoisers, with DiffPIR [9] as a notable example. However, they often require hundreds of reverse steps, leading to computational inefficiency. Moreover, starting the reverse process from pure Gaussian noise lacks structural cues from the degraded input, hindering interpretability and alignment with SISR objectives.

To address these limitations, we draw inspiration from the recent success of residual DMs [12], [13], [14]. Unlike standard approaches that initiate sampling from pure Gaussian noise with stochastic perturbations, the residual DMs a directional transform from the target output back to the conditional input. This formulation enables a more efficient sampling trajectory. However, current state-of-the-art (SOTA) residual DMs are still predominantly supervised and can exhibit issues such as over-smoothing or hallucination [15], [16], which limit their generalization and adaptability. Such shortcomings also heighten risks of data infidelity, which are especially problematic in domains where data fidelity is paramount.

In response, we propose PlaSo, a novel unsupervised DM-based PnP method that extends and improves upon DiffPIR [9] through a residual diffusion process utilizing ResShift [12]. Particularly, it introduces a restructured deterministic sampling mechanism to improve reconstruction quality and better mitigate hallucinations. Our contributions are thereby twofolds:

- PlaSo leverages the advantages of both PnP and residual DMs while simultaneously addressing their aforementioned limitations. To our knowledge, we are the first to jointly incoperate residual DMs and PnP for SISR.
- Empirically, PlaSo consistently provides overall better qualitative and comparable, if not better quantitative performance than other SOTA methods across key metrics. PlaSo also delivers the highest perceptual quality while demonstrating notable additional advantages over other works. Specifically, compared to DiffPIR and other PnP frameworks, PlaSo significantly achieves a 6× reduction in sampling steps while enhancing interpretability. Compared to other DMs, PlaSo maintains higher fidelity.

This research was supported by French National Research Agency (ANR), SONATINE project (ANR-23-CE45-0002-01).

All authors are with the IRIT Lab, Université de Toulouse, and CNRS, Toulouse, France. (Corresponding author: hieu.le-xuan@irit.fr)

II. METHODOLOGY

A. Denoising Diffusion Probabilistic Models (DDPM)

DM have recently gained significant attention, with DDPM [17] standing out as one of the most studied approaches. DDPM has two stages: a forward process and a reverse sampling process. The forward process is defined by a Markov chain which incrementally corrupts the data with a small amount of noise. Formally, given the observation $\mathbf{x}_t$ at timestep t, the forward process is:

$$\mathbf{x}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\epsilon_{t-1}, \epsilon_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (1)$$

where $\{\beta_t \in (0,1)\}_{t=1}^{T}$ schedule the rate of diffusion. By further letting $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, we can derive $\mathbf{x}_t$ directly from original clean image $\mathbf{x}_0$ as:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_{t-1} \quad (2)$$

On the other hand, the reverse process uses a network to gradually predict and remove added noise to recover the high quality image. This process is modeled as:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta\left(\mathbf{x}_t, t\right)\right) + \sqrt{\beta_t}\epsilon_t, \quad (3)$$

where $\epsilon_\theta\left(\mathbf{x}_t, t\right)$is the approximator of the noise ϵ_t. DDPM requires a lengthy reverse process including hundreds to a thousand steps, making it computationally demanding, and often lacks input or task cues.

B. Residual Shifting for SISR

Residual learning has been proven to effectively simplify the reverse process [18]. By directly incorporating the LR image and estimating the LR-HR residual, these models, pioneered by ResShift [12] reduced the reverse process from thousands of steps to just a few, while simultaneously enhancing the quality of SISR outcomes [12], [13], [14]. Specifically, the forward process is

$$x_t = x_{t-1} + \alpha_t(y_0 - x_0) + \kappa\sqrt{\alpha_t}\cdot\epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \quad (4)$$

where y_0 is the LR observation, which we can assume to match the spatial resolution of x_0 by pre-upsampling, η_t is a shifting sequence of hyperparameters that increases monotonically with the timestep t, and satisfies $\eta_T \to 1$, and κ is a scaling factor for noise variance and $\alpha_t = \eta_t - \eta_{t-1}$. And the reverse sampling process is thereby defined as:

$$x_{t-1} = \frac{\eta_{t-1}}{\eta_t}x_t + \frac{\alpha_t}{\eta_t}f_\theta(x_t, y_0, t) + \kappa\sqrt{\frac{\eta_{t-1}}{\eta_t}\alpha_t}\cdot\epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \quad (5)$$

where $f_\theta(x_t, y_0, t)$ is a trained model aiming to predict x_0 based on the original observation input and timestep information. While being more advanced, these models remains supervised, relies on large datasets, and experimentally showed significant oversmooth, which limited its applicability to complex real-world SISR scenarios.

C. Our proposed method: PlaSo

We begin by formulating the stated SISR problem using a commonly adopted degradation model in literature, where the observed low-resolution image $\mathbf{y}$ is defined as:

$$\mathbf{y} = \mathbf{DHx} + \mathbf{n}, \quad (6)$$

where $\mathbf{x}$ is the estimated HR image, $\mathbf{H}$, $\mathbf{D}$ are blurring and decimation operators, and $\mathbf{n}$ is the additive noise.

Under this degradation model, the SISR task can be posed as the following inverse problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma_n^2}\|\mathbf{y} - \mathbf{DHx}\|^2 + \lambda\mathcal{P}(\mathbf{x}), \quad (7)$$

where σ_n^2 is the noise variance, $\mathcal{P}(\mathbf{x})$ represents a prior, and λ is a regularization parameter that balances the data fidelity term and the prior term. By using Half-quadratic splitting (HQS) [19] and introducing an auxiliary variable $\mathbf{z}$, the optimization problem in (7) can be addressed by iteratively solving the following subproblems:

$$\begin{cases} \mathbf{x}_0^{(t)} = \arg\min_{\mathbf{z}} \frac{1}{2\sigma_t^2}\left\|\mathbf{z} - \mathbf{x}_t\right\|^2 + \mathcal{P}(\mathbf{z}) & \text{(8a)} \\ \hat{\mathbf{x}}_0^{(t)} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{DHx}\|^2 + \rho_t\left\|\mathbf{x} - \mathbf{x}_0^{(t)}\right\|^2 & \text{(8b)} \\ \mathbf{x}_{t-1} = \hat{\mathbf{x}}_0^{(t)} & \text{(8c)}, \end{cases}$$

where $\sigma_t = \sqrt{\frac{\lambda}{\tau}}$, $\rho_t = \sigma_n^2\tau$, with τ being the penalty parameter introduced by HQS.

Eq. (8a) is a denoising problem, with new noise level governed by σ_t. While DDPM can be applied to solve (8a) by replacing it with a sampling step in DDPM's reverse process, as in DiffPIR [9], we argue that this is inefficient due to its lengthy reverse process, and further lacks structural and task-specific alignment cues. To address this, we propose starting from the LR input itself, thus enabling a more directional mapping between LR and HR image pairs, faster sampling speed and better interpretability. Accordingly, we propose to replace Eq. (8a) with one sampling step from the residual-based models in Eq. (5). This can be feasibly achieved by rewriting the Eq. (4) as:

$$\boldsymbol{x}_t = \boldsymbol{x}_0 + \eta_t(\boldsymbol{y}_0 - \boldsymbol{x}_0) + \kappa\sqrt{\eta_t}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \quad (9)$$

Recalling that the added noise level in the above equation is proportional to $\sqrt{\eta_t}$ scaled by a factor κ. This reformulation thereby enables a possible integration of residual diffusion-based sampling into the Eq. (8a).

Eq. (8b) involves the effects of decimation and blurring, which can be efficiently handled in the frequency domain. This enables a closed-form solution by leveraging the linearity and convolution properties [1].

Eq. (8c) corresponds to the sampling refinement step[1], which includes a deterministic estimated noise term and

[1]This equation is the same as Eq. (15) in DiffPIR [9].

a stochastic random noise term, as defined in Eq. (10). This step re-injects noise into the current output before the next step, which is intended to improve reconstruction quality [20], [9]:

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\,\hat{\mathbf{x}}_0^{(t)} + \underbrace{\sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{1-\zeta}\,\hat{\epsilon}_t}_{\text{Deterministic term}} + \underbrace{\sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\zeta}\,\epsilon_t}_{\text{Stochastic term}}, \quad \epsilon \sim \mathcal{N}(0,\mathbf{I}), \tag{10}$$

where the parameter ζ controls stochasticity, while $\bar{\alpha}_{t-1}$ controls the signal-noise ratio of the refinement. This step has two limitations. First, Eq. (10) and its variants in recent PnP studies [11], [10] rely on $\bar{\alpha}_{t-1}$ a formulation specifically designed for DDPM [17]. In contrast, ResShift utilizes different noise governance. Consequently, directly plugging ResShift into DiffPIR will yield suboptimal results; for further details, please refer to our ablation study in Section III-B2. Second, the parameter ζ is advantageous for tasks such as inpainting, where output diversity is desired [21]. However, stochasticity accumulates variability that eventually lead to artifacts or information loss, making it less reliable for tasks such as SISR [22], [23], [24]. Additionally, maintaining ζ complicates tuning efforts.

Based on these insights, we first restructure PlaSo's sampling process by introducing a residual-adapted variant of Eq. (10) to address the first limitation. Next, we eliminate ζ and perform refinement using only deterministic noise $\hat{\epsilon}_t$, thereby resolving the second limitation. Accordingly, we derive the following equations for sampling refinement:

$$\hat{\boldsymbol{\epsilon}}_t = \left(\mathbf{x}_t - \hat{\mathbf{x}}_0^t - \eta_t \mathbf{e}_0\right) \cdot \left(\frac{1}{\kappa\sqrt{\eta_t}}\right) \tag{11}$$

$$\mathbf{x}_{t-1} = \hat{\mathbf{x}}_0^{(t)} + \kappa\sqrt{\alpha_{t-1}} \cdot \hat{\boldsymbol{\epsilon}}_t, \tag{12}$$

where e_0 is the residual estimation between the LR and HR images, for example: $e_0 = y_0 - x_0$.

Considering all these aspects, the pseudo-code for our method PlaSo is presented in Algorithm 1.

Algorithm 1: Pseudo-code of PlaSo

Require a trained model f_θ, σ_n.
Define: T, λ, κ, η_t, α_t, ρ_t.
While t=T to 1 do:

Step 1: $\mathbf{x}_0^{(t)} = \frac{\eta_{t-1}}{\eta_t}\mathbf{x}_t + \frac{\alpha_t}{\eta_t} f_\theta(\mathbf{x}_t, \mathbf{y}_0, t)$ *# sampling from denoiser*

Step 2: $\hat{\mathbf{x}}_0^{(t)} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{H}\mathbf{x}\|^2 + \rho_t \left\|\mathbf{x} - \mathbf{x}_0^{(t)}\right\|^2$ *# balancing fidelity and regularization strength*

Step 3: $\hat{\boldsymbol{\epsilon}}_t = \left(\mathbf{x}_t - \hat{\mathbf{x}}_0^t - \eta_t \mathbf{e}_0\right) \cdot \left(\frac{1}{\kappa\sqrt{\eta_t}}\right)$ *# estimating noise*

Step 4: $\mathbf{x}_{t-1} = \hat{\mathbf{x}}_0^{(t)} + \kappa\sqrt{\alpha_{t-1}} \cdot \hat{\boldsymbol{\epsilon}}_t$ *# estimation correction*

Return: $\mathbf{x}_0$

III. Experimental results

A. Dataset and metrics

We compare PlaSo against several approaches including SOTA PnP works such as DiffPIR [9], DDNM+ [10] and DKP-DIP[11], and two residual DMs tailored for super resolution: ResShift [12], SinSR [13]. We adopted two prevailing SISR testsets, Set5 and Set14, to facilitate future comparabilityto ensure consistency and facilitate future comparisons. To further assess the robustness and generalizability of our approach across diverse image domains, we also conduct evaluations on a holdout subset of 200 randomly selected images from the ImageNet validation set [25]. We followed the degradation pipeline described in Eq. (6).

To comprehensively evaluate the performance of all studied methods, we employ both reference metrics such as Peak signal-to-noise ratio (PSNR), Multiscale structural similarity (MS-SSIM), Learned Perceptual Image Patch Similarity (LPIPS) [26], and no-reference metric including CLIP-IQA [27]. Computational complexity is assessed via the total cost of restoring a single image for each method, measured in measured FLOPs (T) [28], inference time (seconds), and NFE [29].

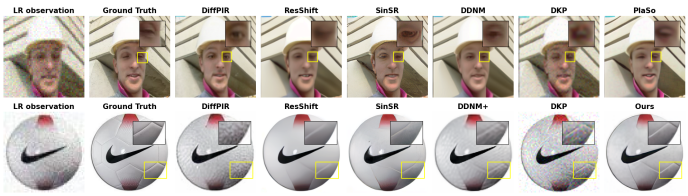

Fig. 1: Qualitative comparison of SISR methods for 4× SR.

B. Experimental Results

1) Results and discussions: Visual results from all methods are shown in Figure 1. One can remark that PlaSo demonstrates overall better fine-grained textures (e.g., ball's surface), delivers more realistic and faithful results (e.g., man's eyes) compared to others. In contrast, ResShift and DiffPIR and DKP-DIP tend to suffer from oversmoothing or artifacts in certain cases, whereas SinSR, and DDNM+, although less prone to blurring, introduces hallucinated details (e.g., man's eyes).

These observations align with quantitative results presented in Table I. On all test sets, PlaSo ranks first on most metrics, including MS-SSIM, LPIPS, CLIPIQA, while securing second-best positions in PSNR and NFE. Notably, compared to DiffPIR and other works, PlaSo delivers substantially better human-level perceptual quality, as evidenced by its higher CLIPIQA scores [27], while maintaining comparable PSNR to the best PSNR provider, DiffPIR, and even slightly better performance on other evaluation metrics. Importantly, whereas other methods sacrifice fidelity for perceptual quality (e.g.,

TABLE I: Quantitative comparison of methods on 4X SISR. Best values are in bold, second-best in underlined.

Dataset	Metric	DiffPIR [9]	ResShift [12]	SinSR [13]	DDNM+ [10]	DKP-DIP [11]	Ours
ImageNet	PSNR (↑)	**23.9701**	22.2619	21.2773	23.3075	22.5063	<u>23.7244</u>
	MS-SSIM (↑)	<u>0.8683</u>	0.8257	0.7986	0.8614	0.8339	**0.8752**
	CLIPIQA (↑)	0.4174	0.4876	0.5144	<u>0.5194</u>	0.5034	**0.6081**
	LPIPS (↓)	0.3838	0.4375	<u>0.3821</u>	0.4189	0.4219	**0.3799**
Set5	PSNR (↑)	**26.0871**	23.3704	22.7312	24.7409	25.0981	<u>25.2753</u>
	MS-SSIM (↑)	<u>0.9175</u>	0.8742	0.8627	0.9049	0.9112	**0.9185**
	CLIPIQA (↑)	0.6299	0.6664	<u>0.7350</u>	0.6281	0.5180	**0.7829**
	LPIPS (↓)	<u>0.2927</u>	0.3308	0.3070	0.3332	0.3240	**0.2904**
Set14	PSNR (↑)	**23.2452**	21.3721	20.6243	22.4547	21.4998	<u>22.7229</u>
	MS-SSIM (↑)	<u>0.8673</u>	0.8123	0.7843	0.8501	0.8212	**0.8676**
	CLIPIQA (↑)	0.6244	0.6463	<u>0.6722</u>	0.5730	0.4901	**0.6925**
	LPIPS (↓)	<u>0.3664</u>	0.4175	0.3850	0.3829	0.4146	**0.3592**

SinSR) or perceptuality for fidelity (e.g., ResShift, DiffPIR, DDNM+, DKP-DIP), our method strikes a better fidelity - realism balance with significantly shortened sampling (See Table III).

We argue that our improvement is not incidental. Maintaining an appropriate trade-off between prior terms and data fidelity has long been a challenge in SISR [30], [31]. Excessive dominance of the prior can result in unstable, biased estimation and oversmoothing [32] or failing to encompass the true image distribution, thereby reducing reconstruction fidelity [33]. ResShift and SinSR rely on fully diffusion-based processes and thus lack an explicit mechanism to control prior dominance. In contrast, PlaSo employs a mechanism to balance prior influence and data fidelity. DiffPIR, DKP-DIP and DDNM+ initiate the restoration process from pure noise, thereby limiting the utilization of cues from the degraded observation and tasks. In contrast, we employ a unified and more interpretable image-to-image distribution mapping which explicitly guide the restoration process. This design explains our more efficient sampling process and enhanced performance, which are consistent with prior literature [18], [12].

TABLE II: Ablation study on the impact of various sampling refinement strategies

Refinement Strategy	PSNR	LPIPS	CLIPIQA
(a) No refinement	22.81	0.3411	0.5215
(b) DiffPIR's stochastic (ζ = 0.25)	21.76	0.3506	0.4312
(c) DiffPIR's stochastic (ζ = 0.5)	21.45	0.3564	0.5729
(d) Our stochastic (ζ = 0.25)	23.54	0.3239	0.7415
(e) Our stochastic (ζ = 0.5)	22.67	0.3308	0.6970
(f) Ours deterministic refinement	**25.28**	**0.2904**	**0.7829**

TABLE III: Comparison of the computational complexity between methods on a RTX6000 GPU.

	Inference time (s)	NFEs	FLOPs (T)
DiffPIR [9]	19.02	100	18.63
ResShift [12]	3.02	15	2.92
SinSR [13]	0.78	1	0.06
DDNM+ [10]	17.49	100	18.56
DKP-DIP [11]	34.28	200	4.89
Ours	3.15	15	3.61

2) Ablation study and computation complexity: We evaluate the performance of PlaSo on Set5 dataset without using any refinement strategy at all (a), or using default DiffPIR refinement strategy (Eq. 10) with ζ (b-c), or using our Eq. (11) but with added stochastic noise using $\sqrt{\zeta}$ as scaling factor (d-e), and finally using our deterministic sampling strategy with Eq. (11) (f). We adopted commonly used values of ζ from [9].

Table II reveals that naive combinations of both DiffPIR and ResShift, without refinement or with DiffPIR's stochastic refinement result in lowest performance, even worse than using our stochastic strategy (d-e) or the model Resshift alone. Prior studies [34], [35] suggested that jointly lowered PSNR and LPIPS metrics correlate with reduced hallucinations. Importantly, using deterministic sampling allows us to simultaneously reduce hallucinations more effectively, proven by an improvement of up to 3.8 dB PSNR and 6% LPIPS over DiffPIR's refinement (c) and 1.74 dB in PSNR, 3% in LPIPS over best stochastic baseline (d). Notably, these gains are achieved without compromising perceptual quality.

We evaluate the computational complexity of all methods in Table III. Notably, PlaSo achieves significantly lowered computation cost and 6 times faster inference time than other PnP methods (DDNM, DKP-DIP, DiffPIR) while delivering better quality (Table I). The only method with significantly lower computation cost than us is SinSR, however, it requires training a distillation model and exhibits reduced fidelity.

IV. Conclusion

This letter has introduced PlaSo, a novel solution for SISR that harnesses the advantages of both residual DLs and Plug-and-Play frameworks, through a DM. PlaSo achieves SOTA performance across several datasets with a much more efficient sampling strategy than DiffPIR, while offering enhanced image realism without compromising data fidelity, a limitation observed in methods such as SinSR and ResShift. Future work will explore integrating more SOTA architectures into PlaSo as well as shortening reverse process.

REFERENCES

[1] Ningning Zhao, Qi Wei, Adrian Basarab, Nicolas Dobigeon, Denis Kouame, and Jean-Yves Tourneret, "Fast single image super-resolution using a new analytical solution for l2-l2 problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, Aug. 2016.

[2] Wenkun Zhang, Hanming Zhang, Linyuan Wang, Ailong Cai, Lei Li, and Bin Yan, "Limited angle ct reconstruction by simultaneous spatial and radon domain regularization based on tv and data-driven tight frame," *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 880, Feb. 2018.

[3] Huafeng Li, Moyuan Yang, and Zhengtao Yu, "Joint image fusion and super-resolution for enhanced visualization via semi-coupled discriminative dictionary learning and advantage embedding," *Neurocomputing*, vol. 422, pp. 62–84, 2021.

[4] Huafeng Li, Yueliang Cen, Yu Liu, Xun Chen, and Zhengtao Yu, "Different input resolutions and arbitrary output resolution: A meta learning-based deep framework for infrared and visible image fusion," *IEEE Transactions on Image Processing*, vol. 30, pp. 4070–4083, 2021.

[5] Wanxin Xiao, Yafei Zhang, Hongbin Wang, Fan Li, and Hua Jin, "Heterogeneous knowledge distillation for simultaneous infrared-visible image fusion and super-resolution," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–15, 2022.

[6] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6360–6376, 2022.

[7] Ulugbek S. Kamilov, Charles A. Bouman, Gregery T. Buzzard, and Brendt Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 85–97, Jan. 2023.

[8] Tianlong Chen, Xiaohan Chen, Wuyang Chen, Zhangyang Wang, Howard Heaton, Jialin Liu, and Wotao Yin, "Learning to optimize: A primer and a benchmark," *J. Mach. Learn. Res.*, vol. 23, no. 1, pp. 1–59, Jan. 2022.

[9] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool, "Denoising diffusion models for plug-and-play image restoration," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2023, p. 1219–1229.

[10] Yinhuai Wang, Jiwen Yu, and Jian Zhang, "Zero-shot image restoration using denoising diffusion null-space model," in *The Eleventh International Conference on Learning Representations*, 2023.

[11] Zhixiong et al Yang, "A dynamic kernel prior model for unsupervised blind image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 26046–26056.

[12] Zongsheng Yue, Jianyi Wang, and Chen Change Loy, "Efficient diffusion model for image restoration by residual shifting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 1, 2025.

[13] Yufei et al Wang, "Sinsr: Diffusion-based image super-resolution in a single step," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[14] Leheng Zhang, Weiyi You, Kexuan Shi, and Shuhang Gu, "Uncertainty-guided perturbation for image super-resolution diffusion model," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025.

[15] Daniel Siromani, Di You, and Pier Luigi Dragotti, "A hallucination metric and correction scheme for diffusion-based image restoration," in *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept. 2024, p. 1–6.

[16] Sumukh K Aithal, Pratyush Maini, Zachary Chase Lipton, and J Zico Kolter, "Understanding hallucinations in diffusion models through mode interpolation," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[17] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2020, NIPS '20, Curran Associates Inc.

[18] Jiawei et al Liu, "Residual denoising diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024.

[19] D. Geman and Chengda Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[20] Hyungjin Chung et al, "Diffusion posterior sampling for general noisy inverse problems," in *The Eleventh International Conference on Learning Representations*, 2023.

[21] ," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11451–11461, 2022.

[22] Bohan Xiao, Peiyong Wang, Qisheng He, and Ming Dong, "Deterministic image-to-image translation via denoising brownian bridge models with dual approximators," in *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, June 2025, pp. 28232–28241.

[23] Gen Li, Yu Huang, Timofey Efimov, Yuting Wei, Yuejie Chi, and Yuxin Chen, "Accelerating convergence of score-based diffusion models, provably," 2024, ICML'24, JMLR.org.

[24] Chong et al Wang, "Reconciling stochastic and deterministic strategies for zero-shot image restoration using diffusion model in dual," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2025, pp. 23207–23216.

[25] Jia et al Deng, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[26] Richard et al Zhang, "The unreasonable effectiveness of deep features as a perceptual metric," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.

[27] Richard et al Zhang, "The unreasonable effectiveness of deep features as a perceptual metric," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.

[28] Wangbo Zhao, Yizeng Han, Jiasheng Tang, Kai Wang, Yibing Song, Gao Huang, Fan Wang, and Yang You, "Dynamic diffusion transformer," in *International Conference on Representation Learning*, Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, Eds., 2025, vol. 2025, pp. 65520–65552.

[29] Sungbin et al Lim, "Score-based generative modeling through stochastic evolution equations in hilbert spaces," Red Hook, NY, USA, 2023, NIPS '23, Curran Associates Inc.

[30] Moon Gi Kang et al, "General choice of the regularization functional in regularized image restoration," *IEEE Transactions on Image Processing*, vol. 4, no. 5, pp. 594–602, 1995.

[31] Antigoni Panagiotopoulou and Vassilis Anastassopoulos, "Super-resolution image reconstruction techniques: Trade-offs between the data-fidelity and regularization terms," *Information Fusion*, vol. 13, no. 3, pp. 185–195, 2012.

[32] Minmin Shen, Ping Xue, and Ci Wang, "The influence of regularization parameter on error bound in super-resolution reconstruction," in *Advances in Multimedia Information Processing - PCM 2008*, Yueh-Min Ray Huang, Changsheng Xu, Kuo-Sheng Cheng, Jar-Ferr Kevin Yang, M. N. S. Swamy, Shipeng Li, and Jen-Wen Ding, Eds., Berlin, Heidelberg, 2008, pp. 535–542, Springer Berlin Heidelberg.

[33] Marzieh Gheisari and Auguste Genovesio, "Super-resolution through stylegan regularized latent search: A realism-fidelity trade-off," *ArXiv*, November 2023.

[34] Seunghoi Kim et al, "Tackling hallucination from conditional models for medical image reconstruction with dynamicdps," in *Proceedings of the 2025 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2025, Accepted for presentation.

[35] Daniel Siromani, Di You, and Pier Luigi Dragotti, "A hallucination metric and correction scheme for diffusion-based image restoration," in *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2024, pp. 1–6.