

ResShift: Efficient Diffusion Model for Image Super-resolution by Residual Shifting

Zongsheng Yue Jianyi Wang Chen Change Loy
S-Lab, Nanyang Technological University
{zongsheng.yue, jianyi001, ccloy}@ntu.edu.sg

Abstract

Diffusion-based image super-resolution (SR) methods are mainly limited by the low inference speed due to the requirements of hundreds or even thousands of sampling steps. Existing acceleration sampling techniques inevitably sacrifice performance to some extent, leading to over-blurry SR results. To address this issue, we propose a novel and efficient diffusion model for SR that significantly reduces the number of diffusion steps, thereby eliminating the need for post-acceleration during inference and its associated performance deterioration. Our method constructs a Markov chain that transfers between the high-resolution image and the low-resolution image by shifting the residual between them, substantially improving the transition efficiency. Additionally, an elaborate noise schedule is developed to flexibly control the shifting speed and the noise strength during the diffusion process. Extensive experiments demonstrate that the proposed method obtains superior or at least comparable performance to current state-of-the-art methods on both synthetic and real-world datasets, *even only with 15 sampling steps*. Our code and model are available at <https://github.com/zsy0AOA/ResShift>.

1 Introduction

Image super-resolution (SR) is a fundamental problem in low-level vision, aiming at recovering the high-resolution (HR) image given the low-resolution (LR) one. This problem is severely ill-posed due to the complexity and unknown nature of degradation models in real-world scenarios. Recently, diffusion model [1, 2], a newly emerged generative model, has achieved unprecedented success in image generation [3]. Furthermore, it has also demonstrated great potential in solving several downstream low-level vision tasks, including image editing [4, 5], image inpainting [6, 7], image colorization [8, 9]. There is also ongoing research exploring the potential of diffusion models to tackle the long-standing and challenging SR task.

One common approach [10, 11] involves inserting the LR image into the input of current diffusion model (e.g., DDPM [2]) and retraining the model from scratch on the training data for SR. Another popular way [7, 12, 13, 14] is to use an unconditional pre-trained diffusion model as a prior and modify its reverse path to generate the expected HR image. Unfortunately, both strategies inherit the Markov chain underlying DDPM, which can be inefficient in inference, often taking hundreds or even thousands of sampling steps. Although some acceleration techniques [15, 16, 17] have been developed to compress the sampling steps in inference, they inevitably lead to a significant drop in performance, resulting in over-smooth results as shown in Fig. 1, in which the DDIM [16] algorithm is employed to speed up the inference. Thus, there is a need to design a new diffusion model for SR that achieves both efficiency and performance, without sacrificing one for the other.

Let us revisit the diffusion model in the context of image generation. In the forward process, it builds up a Markov chain to gradually transform the observed data into a pre-specified prior distribution, typically a standard Gaussian distribution, over a large number of steps. Subsequently, image

ResShift: Mô hình khuếch tán hiệu quả cho hình ảnh
Siêu phân giải bằng dịch chuyển dư

Zongsheng Yue Jianyi Wang Chen Change Loy S-Lab, Đại
học Công nghệ Nanyang
{zongsheng.yue, jianyi001, ccloy}@ntu.edu.sg

trừu tượng

Các phương pháp siêu phân giải hình ảnh (SR) dựa trên khuếch tán chủ yếu bị hạn chế bởi tốc độ suy luận thấp do yêu cầu hàng trăm hoặc thậm chí hàng nghìn bước lấy mẫu. Các kỹ thuật lấy mẫu gia tốc hiện tại chắc chắn sẽ làm giảm hiệu suất ở một mức độ nào đó, dẫn đến kết quả SR quá mờ. Để giải quyết vấn đề này, chúng tôi đề xuất một mô hình khuếch tán mới và hiệu quả cho SR giúp giảm đáng kể số bước khuếch tán, do đó loại bỏ nhu cầu tăng tốc sau trong quá trình suy luận và sự suy giảm hiệu suất liên quan của nó. Phương pháp của chúng tôi xây dựng chuỗi Markov chuyển giữa hình ảnh có độ phân giải cao và hình ảnh có độ phân giải thấp bằng cách dịch chuyển phần dữ liệu giữa chúng, cải thiện đáng kể hiệu quả chuyển đổi. Ngoài ra, một lịch trình tiếng ồn phức tạp được phát triển để kiểm soát linh hoạt tốc độ chuyển đổi và cường độ tiếng ồn trong quá trình khuếch tán. Các thử nghiệm mở rộng chứng minh rằng phương pháp được đề xuất đạt được hiệu suất vượt trội hoặc ít nhất có thể so sánh với các phương pháp tiên tiến hiện tại trên cả bộ dữ liệu tổng hợp và thế giới thực, thậm chí chỉ với 15 bước lấy mẫu. Mã và mô hình của chúng tôi có sẵn tại <https://github.com/zsy0AOA/ResShift>.

1. Giới thiệu

Độ phân giải siêu cao của hình ảnh (SR) là một vấn đề cơ bản trong tầm nhìn ở mức độ thấp, nhằm mục đích khôi phục hình ảnh có độ phân giải cao (HR) cho hình ảnh có độ phân giải thấp (LR). Vấn đề này được đặt ra một cách nghiêm trọng do tính phức tạp và bản chất chưa được biết đến của các mô hình suy thoái trong các tình huống thực tế. Gần đây, mô hình khuếch tán [1, 2], một mô hình thế hệ mới xuất hiện, đã đạt được thành công chưa từng có trong việc tạo hình ảnh [3]. Hơn nữa, nó cũng đã chứng tỏ tiềm năng to lớn trong việc giải quyết một số nhiệm vụ thị giác cấp thấp ở hạ nguồn, bao gồm chỉnh sửa hình ảnh [4, 5], inpainting hình ảnh [6, 7], tô màu hình ảnh [8, 9]. Ngoài ra còn có nghiên cứu đang diễn ra khám phá tiềm năng của các mô hình khuếch tán để giải quyết nhiệm vụ SR lâu dài và đầy thách thức.

Một cách tiếp cận phổ biến [10, 11] liên quan đến việc chèn hình ảnh LR vào đầu vào của mô hình khuếch tán hiện tại (ví dụ: DDPM [2]) và đào tạo lại mô hình từ đầu trên dữ liệu huấn luyện cho SR. Một cách phổ biến khác [7, 12, 13, 14] là sử dụng mô hình khuếch tán được đào tạo trước và điều kiện làm mô hình trước và sửa đổi đường dẫn ngược của nó để tạo ra hình ảnh nhân sự mong đợi. Thật không may, cả hai chiến lược đều kế thừa chuỗi Markov dựa trên DDPM, có thể suy luận kém hiệu quả, thường thực hiện hàng trăm hoặc thậm chí hàng nghìn bước lấy mẫu. Mặc dù một số kỹ thuật tăng tốc [15, 16, 17] đã được phát triển để nén các bước lấy mẫu trong suy luận, nhưng chúng chắc chắn sẽ dẫn đến hiệu suất giảm đáng kể, dẫn đến kết quả quá mịn như trong Hình 1, trong đó DDIM [16] thuật toán được sử dụng để tăng tốc độ suy luận. Vì vậy, cần phải thiết kế một mô hình khuếch tán mới cho SR đạt được cả hiệu suất và hiệu suất mà không phải hy sinh cái này cho cái kia.

Chúng ta hãy xem lại mô hình khuếch tán trong bối cảnh tạo ra hình ảnh. Trong quá trình chuyển tiếp, nó xây dựng chuỗi Markov để dần dần chuyển đổi dữ liệu được quan sát thành phân phối trước được chỉ định trước, diễn hình là phân phối Gaussian tiêu chuẩn, qua một số lượng lớn các bước. Sau đó, hình ảnh

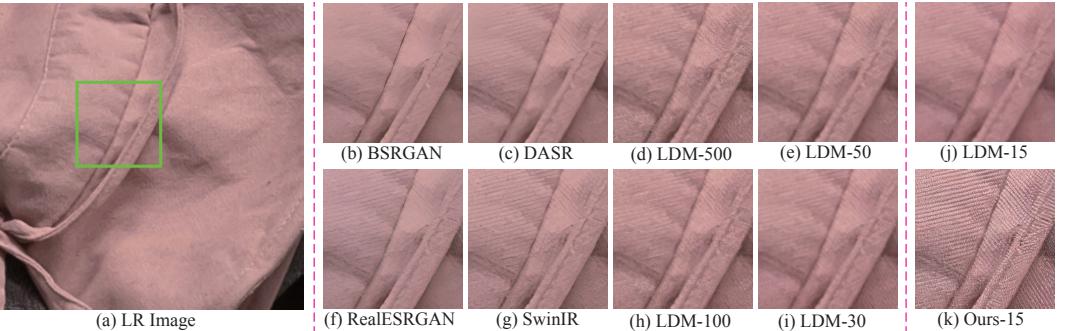


Figure 1: Qualitative comparisons on one typical real-world example of the proposed method and recent state of the arts, including BSRGAN [18], RealESRGAN [19], SwinIR [20], DASR [21], and LDM [11]. As for LDM and our method, we mark the number of sampling steps with the format of “LDM (or Ours)-A” for more intuitive visualization, where “A” is the number of sampling steps. Note that LDM contains 1000 diffusion steps in training and is accelerated to “A” steps using DDIM [16] during inference. Please zoom in for a better view.

generation can be achieved by sampling a noise map from the prior distribution and feeding it into the reverse path of the Markov chain. While the Gaussian prior is well-suited for the task of image generation, it may not be optimal for SR, where the LR image is available. In this paper, we argue that the reasonable diffusion model for SR should start from a prior distribution based on the LR image, enabling an iterative recovery of the HR image from its LR counterpart instead of Gaussian white noise. Additionally, such a design can reduce the number of diffusion steps required for sampling, thereby improving inference efficiency.

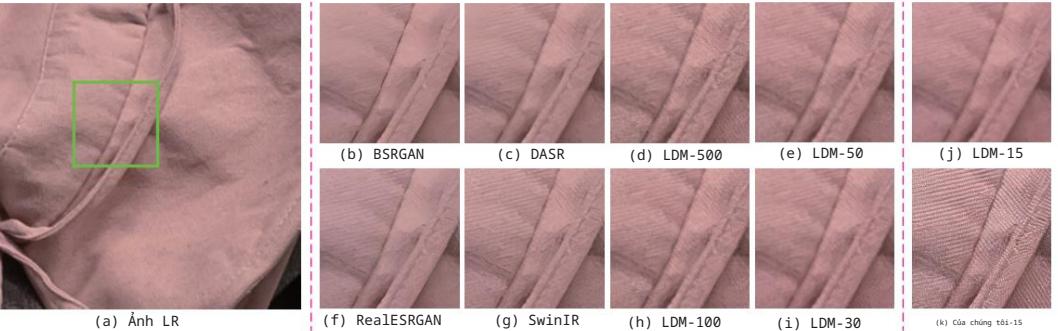
Following the aforementioned motivation, we propose an efficient diffusion model involving a shorter Markov chain for transitioning between the HR image and its corresponding LR one. The initial state of the Markov chain converges to an approximate distribution of the HR image, while the final state converges to an approximate distribution of the LR image. To achieve this, we carefully design a transition kernel that shifts the residual between them step by step. This approach is more efficient than existing diffusion-based SR methods since the residual information can be quickly transferred in dozens of steps. Moreover, our design also allows for an analytical and concise expression for the evidence lower bound, easing the induction of the optimization objective for training. Based on this constructed diffusion kernel, we further develop a highly flexible noise schedule that controls the shifting speed of the residual and the noise strength in each step. This schedule facilitates a fidelity-realism trade-off of the recovered results by tuning its hyper-parameters.

In summary, the main contributions of this work are as follows:

- We present an efficient diffusion model for SR, which renders an iterative sampling procedure from the LR image to the desirable HR one by shifting the residual between them during inference. Extensive experiments demonstrate the superiority of our approach in terms of efficiency, as it requires only 15 sampling steps to achieve appealing results, outperforming or at least being comparable to current diffusion-based SR methods that require a long sampling process. A preview of our recovered results compared with existing methods is shown in Fig. 1.
- We formulate a highly flexible noise schedule for the proposed diffusion model, enabling more precise control of the shifting of residual and noise levels during the transition.

2 Methodology

In this section, we present a diffusion model, *ResShift*, which is tailored for SR. For ease of presentation, the LR and HR images are denoted as y_0 and x_0 , respectively. Furthermore, we assume y_0 and x_0 have identical spatial resolution, which can be easily achieved through pre-upampling the LR image y_0 using nearest neighbor interpolation if necessary.



Hình 1: So sánh định tính trên một ví dụ thực tế diễn hình về phương pháp được đề xuất và trạng thái nghệ thuật gần đây, bao gồm BSRGAN [18], RealESRGAN [19], SwinIR [20], DASR [21] và LDM [11]. Đối với LDM và phương pháp của chúng tôi, chúng tôi đánh dấu số bước lấy mẫu bằng định dạng “LDM (hoặc của chúng tôi)-A” để trực quan hóa hơn, trong đó “A” là số bước lấy mẫu. Lưu ý rằng LDM chứa 1000 bước khuếch tán trong quá trình huấn luyện và được tăng tốc lên các bước “A” bằng cách sử dụng DDIM [16] trong quá trình suy luận. Vui lòng phóng to để nhìn rõ hơn.

có thể đạt được việc tạo ra bằng cách lấy mẫu bẩn đồ nhiều từ phân phối trước đó và đưa nó vào đường dẫn ngược của chuỗi Markov. Mặc dù ưu tiên Gaussian rất phù hợp cho nhiệm vụ tạo hình ảnh nhưng nó có thể không tối ưu cho SR, nơi có sẵn hình ảnh LR. Trong bài báo này, chúng tôi lập luận rằng mô hình khuếch tán hợp lý cho SR nên bắt đầu từ phân phối trước dựa trên hình ảnh LR, cho phép khôi phục lặp lại hình ảnh HR từ đối tác LR của nó thay vì nhiễu trắng Gaussian. Ngoài ra, thiết kế như vậy có thể giảm số bước khuếch tán cần thiết để lấy mẫu, từ đó cải thiện hiệu quả suy luận.

Theo động lực đã nói ở trên, chúng tôi đề xuất một mô hình khuếch tán hiệu quả liên quan đến chuỗi Markov ngắn hơn để chuyển đổi giữa hình ảnh HR và hình ảnh LR tương ứng của nó. Trạng thái ban đầu của chuỗi Markov hội tụ về phân bố gần đúng của hình ảnh HR, trong khi trạng thái cuối cùng hội tụ về phân bố gần đúng của hình ảnh LR. Để đạt được điều này, chúng tôi thiết kế cẩn thận một hạt nhân chuyển tiếp để dịch chuyển phần dư giữa chúng từng bước một. Cách tiếp cận này hiệu quả hơn các phương pháp SR dựa trên khuếch tán hiện có vì thông tin còn lại có thể được truyền nhanh chóng qua hàng chục bước. Hơn nữa, thiết kế của chúng tôi cũng cho phép trình bày ngắn gọn và mang tính phân tích về giới hạn dưới của bằng chứng, giúp giảm bớt việc đưa ra mục tiêu tối ưu hóa cho hoạt động đào tạo. Dựa trên hạt nhân khuếch tán được xây dựng này, chúng tôi tiếp tục phát triển một lịch trình tiếng ồn rất linh hoạt để kiểm soát tốc độ dịch chuyển của phần dư và cường độ tiếng ồn trong mỗi bước. Lịch trình này tạo điều kiện cho sự đánh đổi giữa độ trung thực và hiện thực của các kết quả được khôi phục bằng cách điều chỉnh các siêu tham số của nó.

Tóm lại, những đóng góp chính của công việc này như sau:

- Chúng tôi trình bày một mô hình khuếch tán hiệu quả cho SR, mô hình này thể hiện quy trình lấy mẫu lặp từ hình ảnh LR sang hình ảnh HR mong muốn bằng cách dịch chuyển phần dư giữa chúng trong quá trình suy luận. Các thử nghiệm mở rộng chứng minh tính ưu việt của phương pháp của chúng tôi về mặt hiệu quả, vì nó chỉ cần 15 bước lấy mẫu để đạt được kết quả hấp dẫn, vượt trội hoặc ít nhất có thể so sánh với các phương pháp SR dựa trên khuếch tán hiện tại đòi hỏi quá trình lấy mẫu lâu dài. Bản xem trước kết quả đã khôi phục của chúng tôi so với các phương pháp hiện có được hiển thị trong Hình 1.
- Chúng tôi xây dựng một lịch trình tiếng ồn rất linh hoạt cho mô hình khuếch tán được đề xuất, cho phép nhiều hơn kiểm soát chính xác sự dịch chuyển của mức dư và tiếng ồn trong quá trình chuyển đổi.

2 Phương pháp luận

Trong phần này, chúng tôi trình bày một mô hình khuếch tán, *ResShift*, được thiết kế riêng cho SR. Để dễ trình bày, hình ảnh LR và HR được ký hiệu lần lượt là y_0 và x_0 . Hơn nữa, chúng tôi giả sử y_0 và x_0 có độ phân giải không giống nhau, có thể dễ dàng đạt được điều này thông qua việc lấy mẫu trước hình ảnh LR y_0 bằng cách sử dụng phép nội suy lân cận gần nhất nếu cần.

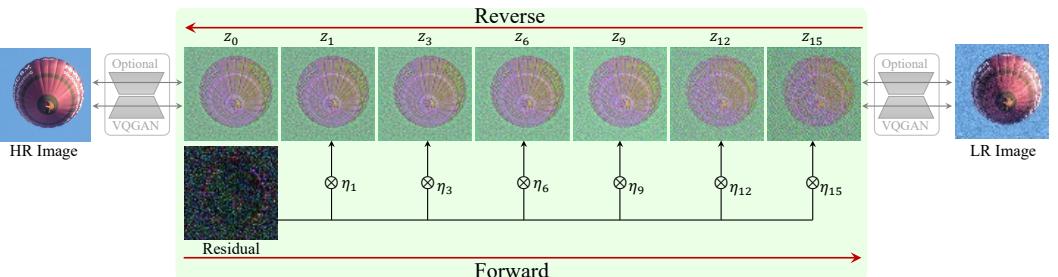


Figure 2: Overview of the proposed method. It builds up a Markov chain between the HR/LR image pair by shifting their residual.

2.1 Model Design

The iterative generation paradigm of diffusion models has proven highly effective at capturing complex distributions, inspiring us to approach the SR problem iteratively as well. Our proposed method constructs a Markov chain that serves as a bridge between the HR and LR images as shown in Fig. 2. This way, the SR task can be accomplished by reverse sampling from this Markov chain given any LR image. Next, we will detail the process of building such a Markov chain specifically for SR.

Forward Process. Let's denote the residual between the LR and HR images as e_0 , i.e., $e_0 = y_0 - x_0$. Our core idea is to transit from x_0 to y_0 by gradually shifting their residual e_0 through a Markov chain with length T . A shifting sequence $\{\eta_t\}_{t=1}^T$ is first introduced, which monotonically increases with the timestep t and satisfies $\eta_1 \rightarrow 0$ and $\eta_T \rightarrow 1$. The transition distribution is then formulated based on this shifting sequence as follows:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{y}_0) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1} + \alpha_t e_0, \kappa^2 \alpha_t \mathbf{I}), \quad t = 1, 2, \dots, T, \quad (1)$$

where $\alpha_t = \eta_t - \eta_{t-1}$ for $t > 1$ and $\alpha_1 = \eta_1$, κ is a hyper-parameter controlling the noise variance, \mathbf{I} is the identity matrix. Notably, we show that the marginal distribution at any timestep t is analytically integrable, namely,

$$q(\mathbf{x}_t | \mathbf{x}_0, \mathbf{y}_0) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_0 + \eta_t e_0, \kappa^2 \eta_t \mathbf{I}), \quad t = 1, 2, \dots, T. \quad (2)$$

The design of the transition distribution presented in Eq. (1) is based on two primary principles. The first principle concerns the standard deviation, i.e., $\kappa \sqrt{\alpha_t}$, which aims to facilitate a smooth transition between \mathbf{x}_t and \mathbf{x}_{t-1} . This is because the expected distance between \mathbf{x}_t and \mathbf{x}_{t-1} can be bounded by $\sqrt{\alpha_t}$, given that the image data falls within the range of $[0, 1]$, i.e.,

$$\max[(\mathbf{x}_0 + \eta_t e_0) - (\mathbf{x}_0 + \eta_{t-1} e_0)] = \max[\alpha_t e_0] < \alpha_t < \sqrt{\alpha_t}, \quad (3)$$

where $\max[\cdot]$ represents the pixel-wise maximizing operation. The hyper-parameter κ is introduced to increase the flexibility of this design. The second principle pertains to the mean parameter, i.e., $\mathbf{x}_0 + \alpha_t e_0$, which induces the marginal distribution in Eq. (2). Furthermore, the marginal distributions of \mathbf{x}_1 and \mathbf{x}_T converges to $\delta_{\mathbf{x}_0}(\cdot)^1$ and $\mathcal{N}(\cdot; \mathbf{y}_0, \kappa^2 \mathbf{I})$, which act as two approximate distributions for the HR image and the LR image, respectively. By constructing the Markov chain in such a thoughtful way, it is possible to handle the SR task by inversely sampling from it given the LR image \mathbf{y}_0 .

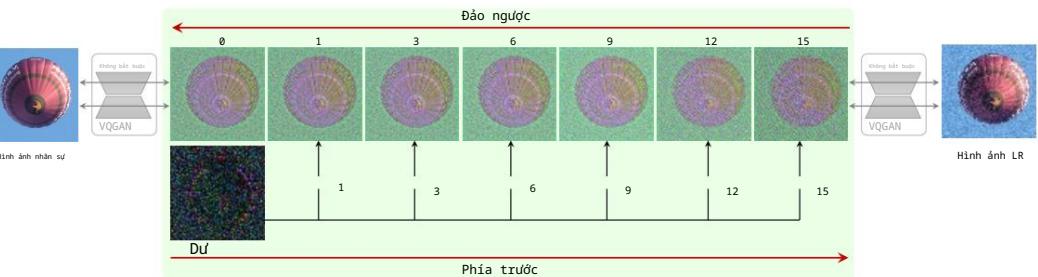
Reverse Process. The reverse process aims to estimate the posterior distribution $p(\mathbf{x}_0 | \mathbf{y}_0)$ via the following formulation:

$$p(\mathbf{x}_0 | \mathbf{y}_0) = \int p(\mathbf{x}_T | \mathbf{y}_0) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_0) d\mathbf{x}_{1:T}, \quad (4)$$

where $p(\mathbf{x}_T | \mathbf{y}_0) \approx \mathcal{N}(\mathbf{x}_T | \mathbf{y}_0, \kappa^2 \mathbf{I})$, $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_0)$ is the inverse transition kernel from \mathbf{x}_t to \mathbf{x}_{t-1} with a learnable parameter θ . Following most of the literature in diffusion model [1, 2, 8], we adopt the assumption of $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, \mathbf{y}_0, t), \Sigma_\theta(\mathbf{x}_t, \mathbf{y}_0, t))$. The optimization for θ is achieved by minimizing the negative evidence lower bound, namely,

$$\min_\theta \sum_t D_{KL}[q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_0) \| p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{y}_0)], \quad (5)$$

¹ $\delta_\mu(\cdot)$ denotes the Dirac distribution centered at μ .



Hình 2: Tổng quan về phương pháp đề xuất. Nó xây dựng chuỗi Markov giữa hình ảnh HR/LR. ghép đôi bằng cách dịch chuyển phần dư của chúng.

2.1 Thiết kế mô hình

Mô hình tạo lặp của các mô hình khuếch tán đã được chứng minh là có hiệu quả cao trong việc nắm bắt các bản phân phối phức tạp, truyền cảm hứng cho chúng tôi tiếp cận vấn đề SR một cách lặp đi lặp lại. Đề xuất của chúng tôi phương pháp này xây dựng chuỗi Markov đóng vai trò là cầu nối giữa hình ảnh HR và LR như trong Hình 2. Bằng cách này, nhiệm vụ SR có thể được thực hiện bằng cách lấy mẫu ngược từ chuỗi Markov đã cho bất kỳ hình ảnh LR nào. Tiếp theo, chúng tôi sẽ trình bày chi tiết quy trình xây dựng chuỗi Markov dành riêng cho SR.

Quá trình chuyển tiếp. Hãy biểu thị phần dư giữa ảnh LR và HR là e_0 , tức là $e_0 = y_0 - x_0$. Ý tưởng cốt lõi của chúng tôi là chuyển từ x_0 sang y_0 bằng cách dịch chuyển dần e_0 còn lại của chúng thông qua Markov xích có chiều dài T . Một dãy dịch chuyển $\{\eta_t\}_{t=1}^T$ được giới thiệu lần đầu tiên, điều này làm tăng lên một cách đơn giản với dấu thời gian t và thỏa mãn $\eta_1 = 0$ và $\eta_T = 1$. Sau đó, phân bố chuyển tiếp được xây dựng dựa trên trình tự dịch chuyển này như sau:

$$q(x_t | x_{t-1}, y_0) = N(x_t; x_{t-1} + \alpha_t e_0, \kappa^2 \alpha_t I), \quad t = 1, 2, \dots, T, \quad (1)$$

trong đó $\alpha_t = \eta_t - \eta_{t-1}$ với $t > 1$ và $\alpha_1 = \eta_1$, κ là siêu tham số kiểm soát phương sai nhiễu, I là ma trận nhận dạng. Đáng chú ý, chúng tôi chỉ ra rằng phân bố cận biên ở bất kỳ bước thời gian t nào là phù hợp về mặt phân tích. có thể tích hợp, cụ thể là

$$q(x_t | x_0, y_0) = N(x_t; x_0 + \eta_t e_0, \kappa^2 \eta_t I), \quad t = 1, 2, \dots, T. \quad (2)$$

Thiết kế phân phối chuyển tiếp được trình bày trong biểu thức. (1) dựa trên hai nguyên tắc chính. Các nguyên tắc đầu tiên liên quan đến độ lệch chuẩn, tức là $\kappa \sqrt{\alpha_t}$, nhằm mục đích tạo điều kiện thuận lợi cho quá trình chuyển đổi suôn sẻ giữa x_t và x_{t-1} . Điều này là do khoảng cách dự kiến giữa x_t và x_{t-1} có thể bị giới hạn bằng $\sqrt{\alpha_t}$, với điều kiện là dữ liệu hình ảnh nằm trong phạm vi $[0, 1]$, nghĩa là

$$\max[(x_0 + \eta_t e_0) - (x_0 + \eta_{t-1} e_0)] = \max[\alpha_t e_0] < \alpha_t < \sqrt{\alpha_t}, \quad (3)$$

trong đó $\max[\cdot]$ thể hiện hoạt động tối đa hóa pixel. Siêu tham số κ được giới thiệu để tăng tính linh hoạt của thiết kế này. Nguyên tắc thứ hai liên quan đến tham số trung bình, tức là $x_0 + \alpha_t e_0$, tạo ra sự phân bố cận biên trong biểu thức. (2). Hơn nữa, phân phối cận biên của x_1 và x_T hội tụ về $\delta_{x_0}(\cdot)$ và $N(\cdot; y_0, \kappa^2 I)$, đóng vai trò là hai phân bố gần đúng cho ảnh HR và ảnh LR tương ứng. Bằng cách xây dựng chuỗi Markov một cách chu đáo như vậy theo cách này, có thể xử lý tác vụ SR bằng cách lấy mẫu nghịch đảo từ nó cho hình ảnh LR y_0 .

Quá trình ngược lại. Quá trình ngược lại nhằm mục đích ước tính phân phối sau $p(x_0 | y_0)$ thông qua công thức sau:

$$p(x_0 | y_0) = p(x_T | y_0) \prod_{t=1}^T p_\theta(x_{t-1} | x_t, y_0) dx_{1:T}, \quad (4)$$

trong đó $p(x_T | y_0) \approx N(x_T | y_0, \kappa^2 I)$, $p_\theta(x_{t-1} | x_t, y_0)$ là hạt nhân chuyển đổi nghịch đảo từ x_t sang x_{t-1} với tham số có thể học được θ . Theo hầu hết các tài liệu về mô hình khuếch tán [1, 2, 8], chúng tôi áp dụng giả định $p_\theta(x_{t-1} | x_t, y_0) = N(x_{t-1}; \mu_\theta(x_t, y_0, t), \Sigma_\theta(x_t, y_0, t))$. Việc tối ưu hóa cho θ đạt được bằng cách giảm thiểu giới hạn dưới của bằng chứng tiêu cực, cụ thể là,

$$\min_\theta \sum_t D_{KL}[q(x_{t-1} | x_t, y_0) \| p_\theta(x_{t-1} | x_t, y_0)], \quad (5)$$

¹ $\delta_\mu(\cdot)$ biểu thị phân bố Dirac có tâm tại μ .

where $D_{KL}[\cdot \parallel \cdot]$ denotes the Kullback-Leibler (KL) divergence. More mathematical details can be found in Sohl-Dickstein et al. [1] or Ho et al. [2].

Combining Eq. (1) and Eq. (2), the targeted distribution $q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0, \mathbf{y}_0)$ in Eq. (5) can be rendered tractable and expressed in an explicit form given below:

$$q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0, \mathbf{y}_0) = \mathcal{N}\left(\mathbf{x}_{t-1} \middle| \frac{\eta_{t-1}}{\eta_t} \mathbf{x}_t + \frac{\alpha_t}{\eta_t} \mathbf{x}_0, \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t \mathbf{I}\right). \quad (6)$$

The detailed calculation of this derivation is presented in the supplementary material. Considering that the variance parameter is independent of \mathbf{x}_t and \mathbf{y}_0 , we thus set $\Sigma_\theta(\mathbf{x}_t, \mathbf{y}_0, t) = \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t \mathbf{I}$. As for the mean parameter $\mu_\theta(\mathbf{x}_t, \mathbf{y}_0, t)$, it is reparameterized as follows:

$$\mu_\theta(\mathbf{x}_t, \mathbf{y}_0, t) = \frac{\eta_{t-1}}{\eta_t} \mathbf{x}_t + \frac{\alpha_t}{\eta_t} f_\theta(\mathbf{x}_t, \mathbf{y}_0, t), \quad (7)$$

where f_θ is a deep neural network with parameter θ , aiming to predict \mathbf{x}_0 . We explored different parameterization forms on μ_θ and found that Eq. (7) exhibits superior stability and performance.

Based on Eq. (7), we simplify the objective function in Eq. (5) as follows,

$$\min_{\theta} \sum_t w_t \|f_\theta(\mathbf{x}_t, \mathbf{y}_0, t) - \mathbf{x}_0\|_2^2, \quad (8)$$

where $w_t = \frac{\alpha_t}{2\kappa^2 \eta_t \eta_{t-1}}$. In practice, we empirically find that the omission of weight w_t results in an evident improvement in performance, which aligns with the conclusion in Ho et al. [2].

Extension to Latent Space. To alleviate the computational overhead in training, we move the aforementioned model into the latent space of VQGAN [22], where the original image is compressed by a factor of four in spatial dimensions. This does not require any modifications on our model other than substituting \mathbf{x}_0 and \mathbf{y}_0 with their latent codes. An intuitive illustration is shown in Fig. 2.

2.2 Noise Schedule

The proposed method employs a hyper-parameter κ and a shifting sequence $\{\eta_t\}_{t=1}^T$ to determine the noise schedule in the diffusion process. Specifically, the hyper-parameter κ regulates the overall noise intensity during the transition, and its impact on performance is empirically discussed in Sec. 4.2. The subsequent exposition mainly revolves around the construction of the shifting sequence $\{\eta_t\}_{t=1}^T$.

Equation (2) implies that the noise level in state \mathbf{x}_t is proportional to $\sqrt{\eta_t}$ with a scaling factor κ . This observation motivates us to focus on designing $\sqrt{\eta_t}$ instead of η_t . Song and Ermon [23] show that $\kappa\sqrt{\eta_1}$ should be sufficiently small (e.g., 0.04 in LDM [11]) to ensure that $q(\mathbf{x}_1 | \mathbf{x}_0, \mathbf{y}_0) \approx q(\mathbf{x}_0)$. Combining with the additional constraint of $\eta_1 \rightarrow 0$, we set η_1 to be the minimum value between $(0.04/\kappa)^2$ and 0.001. For the final step T , we set η_T as 0.999 ensuring $\eta_T \rightarrow 1$. For the intermediate timesteps, i.e., $t \in [2, T-1]$, we propose a non-uniform geometric schedule for $\sqrt{\eta_t}$ as follows:

$$\sqrt{\eta_t} = \sqrt{\eta_1} \times b_0^{\beta_t}, \quad t = 2, \dots, T-1, \quad (9)$$

where

$$\beta_t = \left(\frac{t-1}{T-1}\right)^p \times (T-1), \quad b_0 = \exp\left[\frac{1}{2(T-1)} \log \frac{\eta_T}{\eta_1}\right]. \quad (10)$$

Note that the choice of β_t and b_0 is based on the assumption of $\beta_1 = 0$, $\beta_T = T-1$, and $\sqrt{\eta_T} = \sqrt{\eta_1} \times b_0^{T-1}$. The hyper-parameter p controls the growth rate of $\sqrt{\eta_t}$ as shown in Fig. 3(h).

The proposed noise schedule exhibits high flexibility in three key aspects. First, for small values of κ , the final state \mathbf{x}_T converges to a perturbation around the LR image as depicted in Fig. 3(c)-(d). Compared to the corruption ended at Gaussian noise, this design considerably shortens the length of the Markov chain, thereby improving the inference efficiency. Second, the hyper-parameter p provides precise control over the shifting speed, enabling a fidelity-realism trade-off in the SR results as analyzed in Sec. 4.2. Third, by setting $\kappa = 40$ and $p = 0.8$, our method achieves a diffusion process remarkably similar to LDM [11]. This is clearly demonstrated by the visual results during the diffusion process presented in Fig. 3(e)-(f), and further supported by the comparisons on the relative noise strength as shown in Fig. 3(g).

trong đó $D_{KL}[\cdot \parallel \cdot]$ biểu thị sự phân kỳ Kullback-Leibler (KL). Nhiều chi tiết toán học hơn có thể được tìm thấy trong Sohl-Dickstein et al. [1] hoặc Ho và cộng sự. [2].

Kết hợp phương trình (1) và phương trình (2), phân phối mục tiêu $q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0, \mathbf{y}_0)$ trong biểu thức (5) có thể được biểu diễn dễ dàng và được thể hiện dưới dạng rõ ràng dưới đây:

$$q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0, \mathbf{y}_0) = N \frac{\eta_{t-1}}{\eta_t} \mathbf{x}_t + \frac{\eta_t}{\eta_{t-1}} \mathbf{x}_0, \quad \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t \mathbf{I}. \quad (6)$$

Việc tính toán chi tiết đạo hàm này được trình bày trong tài liệu bổ sung. Xét rằng tham số phương sai độc lập với \mathbf{x}_t và \mathbf{y}_0 , do đó chúng ta đặt $\Sigma_\theta(\mathbf{x}_t, \mathbf{y}_0, t) = \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t \mathbf{I}$. Đổi với tham số trung bình $\mu_\theta(\mathbf{x}_t, \mathbf{y}_0, t)$ được tham số hóa lại như sau:

$$\mu_\theta(\mathbf{x}_t, \mathbf{y}_0, t) = \mathbf{x}_t + \frac{\eta_t}{\eta_{t-1}} f_\theta(\mathbf{x}_t, \mathbf{y}_0, t), \quad (7)$$

trong đó f_θ là mạng nơ ron sâu có tham số θ , nhằm dự đoán \mathbf{x}_0 . Chúng tôi đã khám phá các dạng tham số hóa khác nhau trên μ_θ và nhận thấy rằng phương trình (7) thể hiện sự ổn định và hiệu suất vượt trội.

Dựa trên phương trình (7), chúng tôi đơn giản hóa hàm mục tiêu trong biểu thức (5) như sau,

$$\text{phút}_\theta = \frac{1}{t} \sum_{t=1}^T w_t \|f_\theta(\mathbf{x}_t, \mathbf{y}_0, t) - \mathbf{x}_0\|_2^2. \quad (\text{số } 8)$$

Ở đây $w_t = \frac{\alpha_t}{2\kappa^2 \eta_t \eta_{t-1}}$. Trong thực tế, theo kinh nghiệm, chúng tôi nhận thấy việc bỏ qua trọng số w_t sẽ dẫn đến một sự cải thiện rõ ràng về hiệu suất, phù hợp với kết luận của Ho et al. [2].

Mở rộng đến không gian tiềm ẩn. Để giảm bớt chi phí tính toán trong quá trình đào tạo, chúng tôi chuyển mô hình nói trên vào không gian tiềm ẩn của VQGAN [22], trong đó hình ảnh gốc được nén theo hệ số bốn theo chiều không gian. Điều này không yêu cầu bất kỳ sửa đổi nào trên mô hình của chúng tôi ngoài việc thay thế \mathbf{x}_0 và \mathbf{y}_0 bằng mã tiềm ẩn của chúng. Một minh họa trực quan được hiển thị trong Hình 2.

2.2 Lịch trình tiếng ồn

Phương pháp được đề xuất sử dụng siêu tham số κ và lịch trình nhiều chuỗi $\{\eta_t\}$ dịch chuyển $\overset{T}{\underset{t=1}{\longrightarrow}}$ để xác định $t=1$ trong quá trình khuếch tán. Cụ thể, siêu tham số κ điều chỉnh cường độ nhiễu tổng thể trong quá trình chuyển đổi và tác động của nó đến hiệu suất sẽ được thảo luận thực nghiệm trong Phần 4.2.

Phản trình bày tiếp theo chủ yếu xoay quanh việc xây dựng chuỗi dịch chuyển $\{\eta_t\}$ $\overset{T}{\underset{t=1}{\longrightarrow}}$.

Phương trình (2) ngụ ý rằng mức nhiễu ở trạng thái \mathbf{x}_t tỷ lệ với $\sqrt{\eta_t}$ với hệ số tỷ lệ κ .

Quan sát này thúc đẩy chúng tôi tập trung vào việc thiết kế $\sqrt{\eta_t}$ thay vì η_t . Song và Ermon [23] cho thấy rằng $\kappa\sqrt{\eta_1}$ phải đủ nhỏ (ví dụ, 0.04 trong LDM [11]) để đảm bảo rằng $q(\mathbf{x}_1 | \mathbf{x}_0, \mathbf{y}_0) \approx q(\mathbf{x}_0)$.

Kết hợp với ràng buộc bổ sung $\eta_1 = 0$, chúng tôi đặt η_1 là giá trị tối thiểu trong khoảng $(0, 0.04/\kappa)$ và 0.001. Đổi với bước cuối cùng T , chúng tôi đặt $\eta_T = 0.999$ để đảm bảo $\eta_T \rightarrow 1$. Đổi với các dấu thời gian trung gian, tức là $t \in [2, T-1]$, chúng tôi đề xuất một lịch trình hình học không đồng nhất cho $\sqrt{\eta_t}$ như sau:

$$\sqrt{\eta_t} = \sqrt{\eta_1} \times b_0^{\beta_t}, \quad t = 2, \dots, T-1, \quad (9)$$

Ở đây

$$\beta_t = \frac{t-1}{T-1} \times (T-1), \quad b_0 = \exp\left[\frac{1}{2(T-1)} \log \frac{\eta_T}{\eta_1}\right]. \quad (10)$$

Lưu ý rằng việc lựa chọn β_t và b_0 dựa trên giả định $\beta_1 = 0$, $\beta_T = T-1$ và $\sqrt{\eta_T} = \sqrt{\eta_1} \times b_0^{T-1}$. Siêu tham số p kiểm soát tốc độ tăng trưởng của $\sqrt{\eta_t}$ như trong Hình 3(h).

Lịch trình tiếng ồn được đề xuất thể hiện tính linh hoạt cao ở ba khía cạnh chính. Đầu tiên, đổi với các giá trị κ nhỏ, trạng thái cuối cùng \mathbf{x}_T hội tụ thành nhiễu loạn xung quanh ảnh LR như được mô tả trong Hình 3(c)-(d). So với sự tham nhũng kết thúc ở nhiễu Gaussian, thiết kế này rút ngắn đáng kể độ dài của chuỗi Markov, từ đó cải thiện hiệu quả suy luận. Thứ hai, siêu tham số p cung cấp khả năng kiểm soát chính xác tốc độ dịch chuyển, cho phép cân bằng giữa độ trung thực và hiện thực trong kết quả SR như được phân tích trong Phần 4.2. Thứ ba, bằng cách đặt $\kappa = 40$ và $p = 0.8$, phương pháp của chúng tôi đạt được quá trình khuếch tán tương tự như LDM [11]. Điều này được thể hiện rõ ràng bằng các kết quả trực quan trong quá trình khuếch tán được trình bày trong Hình 3(e)-(f) và được hỗ trợ thêm bằng các so sánh về cường độ nhiễu tương đối như trong Hình 3(g).

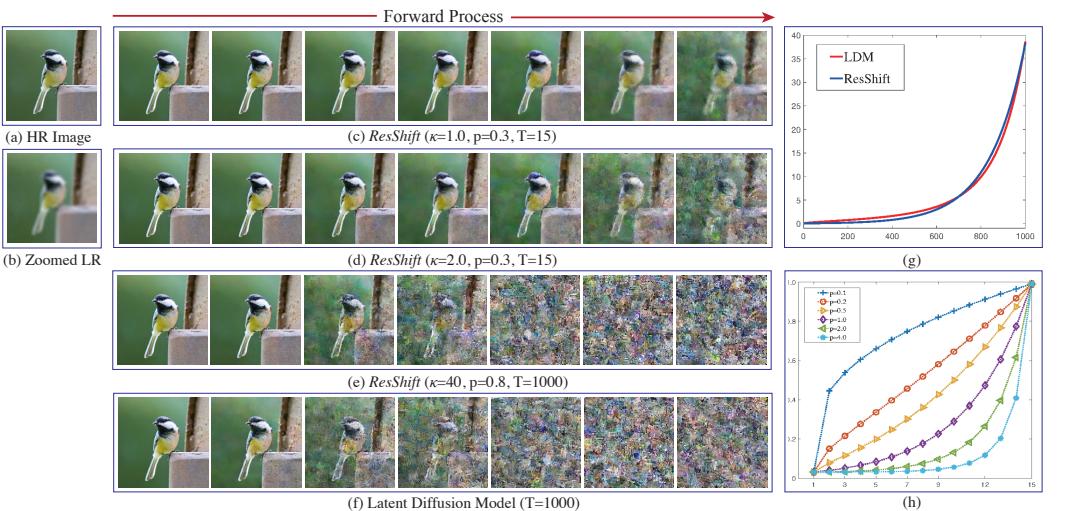


Figure 3: Illustration of the proposed noise schedule. (a) HR image. (b) Zoomed LR image. (c)-(d) Diffused images of *ResShift* in timesteps of 1, 3, 5, 7, 9, 12, and 15 under different values of κ by fixing $p = 0.3$ and $T = 15$. (e)-(f) Diffused images of *ResShift* with a specified configuration of $\kappa = 40$, $p = 0.8$, $T = 1000$ and LDM [11] in timesteps of 100, 200, 400, 600, 800, 900, and 1000. (g) The relative noise intensity (vertical axes, measured by $\sqrt{1/\lambda_{\text{snr}}}$, where λ_{snr} denotes the signal-to-noise ratio) of the schedules in (d) and (e) w.r.t. the timesteps (horizontal axes). (h) The shifting speed $\sqrt{\eta_t}$ (vertical axes) w.r.t. to the timesteps (horizontal axes) across various configurations of p . Note that the diffusion processes in this figure are implemented in the latent space, but we display the intermediate results after decoding back to the image space for the purpose of easy visualization.

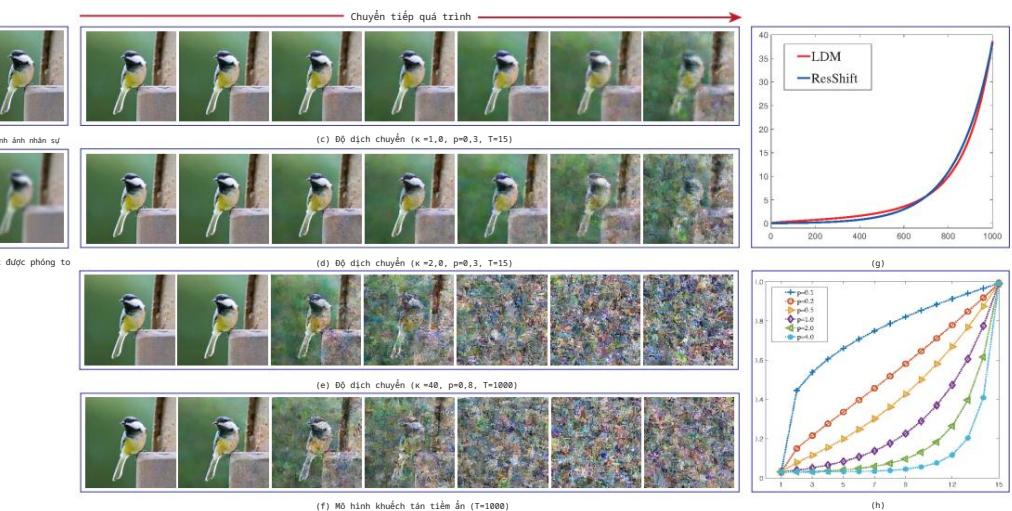
3 Related Work

Diffusion Model. Inspired by the non-equilibrium statistical physics, Sohl-Dickstein et al. [1] firstly proposed the diffusion model to fit complex distributions. Ho et al. [2] established a novel connection between the diffusion model and the denoising scoring matching. Later, Song et al. [8] proposed a unified framework to formulate the diffusion model from the perspective of the stochastic differential equation (SDE). Attributed to its robust theoretical foundation, the diffusion model has achieved impressive success in the generation of images [3, 11], audio [24], graph [25] and shapes [26].

Image Super-Resolution. Traditional image SR methods primarily focus on designing more rational image priors based on our subjective knowledge, such as non-local similarity [27], low-rankness [28], sparsity [29, 30], and so on. With the development of deep learning (DL), Dong et al. [31] proposed the seminal work SRCNN to solve the SR task using a deep neural network. Then DL-based SR methods rapidly dominated the research field. Various SR technologies were explored from different perspectives, including network architecture [32, 33, 34, 35], image prior [36, 37, 38, 39], deep unfolding [40, 41, 42], degradation model [18, 19, 43, 44].

Recently, some works have investigated the application of diffusion models in SR. A prevalent approach is to concatenate the LR image with the noise in each step and retrain the diffusion model from scratch [10, 11, 45]. Another popular way is to utilize an unconditional pre-trained diffusion model as a prior and incorporate additional constraints to guide the reverse process [7, 12, 13, 46]. Both strategies often require hundreds or thousands of sampling steps to generate a realistic HR image. While several acceleration algorithms [15, 16, 17] have been proposed, they typically sacrifice the performance and result in blurry outputs. This work designs a more efficient diffusion model that overcomes this trade-off between efficiency and performance, as detailed in Sec. 2.

Remark. Several parallel works [47, 48, 49] also exploit such an iterative restoration paradigm in SR. Despite a similar motivation, our work and others have adopted different mathematical formulations to achieve this goal. Delbracio and Milanfar [47] employed the Inversion by Direct Iteration (InDI) to model this process, while Luo et al. [48] and Liu et al. [49] attempted to formulate it as a SDE. In this paper, we design a discrete Markov chain to depict the transition between the HR and LR images, offering a more intuitive and efficient solution to this problem.



Hình 3: Minh họa kế hoạch tiếng ồn đề xuất. (a) Hình ảnh nhân sự. (b) Ảnh LR được phóng to. (đi a CD) Hình ảnh khuếch tán của ResShift theo các dấu thời gian 1, 3, 5, 7, 9, 12 và 15 dưới các giá trị khác nhau của κ bằng cách sửa $p = 0.3$ và $T = 15$. (e)-(f) Hình ảnh khuếch tán của ResShift với một giá trị xác định cấu hình $\kappa = 40$, $p = 0.8$, $T = 1000$ và LDM [11] theo các bước thời gian 100, 200, 400, 600, 800, 900 và 1000. (g) Cường độ nhiễu tương đối (trục dọc, được đo bằng $1/\lambda_{\text{snr}}$, trong đó λ_{snr} biểu thị tỷ lệ tín hiệu trên nhiễu) của các lịch trình trong (d) và (e) ghi các dấu thời gian (trục ngang). (h) Tốc độ dịch chuyển $\sqrt{\eta_t}$ (trục dọc) ghi theo dấu thời gian (trục ngang) trên các cấu hình khác nhau của p . Lưu ý rằng các quá trình khuếch tán trong hình này được thực hiện trong không gian tiềm ẩn, nhưng chúng tôi hiển thị kết quả trung gian sau khi giải mã trở lại không gian ảnh nhằm mục đích dễ hình dung.

3 công việc liên quan

Mô hình khuếch tán. Lấy cảm hứng từ vật lý thống kê không cân bằng, Sohl-Dickstein et al. [1] trước hết đề xuất mô hình khuếch tán để phù hợp với các phân bố phức tạp. Họ và cộng sự. [2] đã thiết lập một kết nối giữa mô hình khuếch tán và kết hợp tính điểm khử nhiễu. Sau đó, Song và cộng sự. [8] đã đề xuất một khuôn khổ thống nhất để xây dựng mô hình khuếch tán từ góc độ phương trình vi phân ngẫu nhiên (SDE). Nhờ nền tảng lý thuyết vững chắc, mô hình khuếch tán đã đạt được thành công ấn tượng trong việc tạo ra hình ảnh [3, 11], âm thanh [24], đồ thị [25] và hình dạng [26].

Hình ảnh siêu phân giải. Các phương pháp SR hình ảnh truyền thống chủ yếu tập trung vào việc thiết kế các ưu tiên hình ảnh hợp lý hơn dựa trên kiến thức chủ quan của chúng ta, chẳng hạn như độ tương tự không cục bộ [27], thứ hạng thấp [28], độ thưa thớt [29, 30], v.v. Với sự phát triển của deep learning (DL), Dong et al. [31] đã đề xuất công việc quan trọng SRCNN để giải quyết nhiệm vụ SR bằng cách sử dụng mạng lưới thần kinh sâu. Sau đó, các phương pháp SR dựa trên DL nhanh chóng thống trị lĩnh vực nghiên cứu. Các công nghệ SR khác nhau đã được khám phá từ các góc độ khác nhau, bao gồm kiến trúc mạng [32, 33, 34, 35], hình ảnh trước [36, 37, 38, 39], mở rộng sâu [40, 41, 42], mô hình xuồng cấp [18, 19, 43, 44].

Gần đây, một số công trình đã nghiên cứu ứng dụng mô hình khuếch tán trong SR. Một cách tiếp cận phổ biến là ghép ảnh LR với nhiều trong mỗi bước và huấn luyện lại mô hình khuếch tán từ đầu [10, 11, 45]. Một cách phổ biến khác là sử dụng mô hình khuếch tán được huấn luyện trước vô điều kiện làm mô hình trước và kết hợp các ràng buộc bổ sung để hướng dẫn quá trình ngược lại [7, 12, 13, 46].

Cả hai chiến lược thường yêu cầu hàng trăm hoặc hàng nghìn bước lấy mẫu để tạo ra hình ảnh nhân sự thực tế. Mặc dù một số thuật toán tăng tốc [15, 16, 17] đã được đề xuất nhưng chúng thường làm giảm hiệu suất và dẫn đến kết quả đầu ra bị mờ. Công việc này thiết kế một mô hình khuếch tán hiệu quả hơn nhằm khắc phục sự cân bằng giữa hiệu quả và hiệu suất, như được trình bày chi tiết trong Phần 2.

Nhận xét. Một số công trình song song [47, 48, 49] cũng khai thác mô hình khôi phục lặp lại như vậy trong SR. Mặc dù có động cơ tương tự, công việc của chúng tôi và những người khác đã áp dụng các công thức toán học khác nhau để đạt được mục tiêu này. Delbracio và Milanfar [47] đã sử dụng Phép đảo ngược bằng phép lặp trực tiếp (InDI) để mô hình hóa quá trình này, trong khi Luo et al. [48] và Liu và cộng sự. [49] đã cố gắng xây dựng nó dưới dạng SDE. Trong bài báo này, chúng tôi thiết kế chuỗi Markov rực rỡ để mô tả quá trình chuyển đổi giữa hình ảnh HR và LR, đưa ra giải pháp trực quan và hiệu quả hơn cho vấn đề này.

Table 1: Performance comparison of *ResShift* on the *ImageNet-Test* under different configurations.

Configurations			Metrics				
T	p	κ	PSNR↑	SSIM↑	LPIPS↓	CLIPQA↑	MUSIQ↑
10	0.3	2.0	25.20	0.6828	0.2517	0.5492	50.6617
			25.01	0.6769	0.2312	0.5922	53.6596
			24.52	0.6585	0.2253	0.6273	55.7904
			24.29	0.6513	0.2225	0.6468	56.8482
			24.22	0.6483	0.2212	0.6489	56.8463
15	0.3	2.0	25.01	0.6769	0.2312	0.5922	53.6596
			25.05	0.6745	0.2387	0.5816	52.4475
			25.12	0.6780	0.2613	0.5314	48.4964
			25.32	0.6827	0.3050	0.4601	43.3060
			25.39	0.5813	0.3432	0.4041	38.5324
15	0.3	2.0	24.90	0.6709	0.2437	0.5700	50.6101
			24.84	0.6699	0.2354	0.5914	52.9933
			25.01	0.6769	0.2312	0.5922	53.6596
			25.31	0.6858	0.2592	0.5231	49.3182
			24.46	0.6891	0.2772	0.4898	46.9794

Figure 4: Qualitative comparisons of *ResShift* under different combinations of (T, p, κ) . For example, “(15, 0.3, 2.0)” represents the recovered result with $T = 15$, $p = 0.3$, and $\kappa = 2.0$. Please zoom in for a better view.

4 Experiments

This section presents an empirical analysis of the proposed *ResShift* and provides extensive experimental results to verify its effectiveness on one synthetic dataset and three real-world datasets. Following [18, 19], our investigation specifically focuses on the more challenging $\times 4$ SR task. Due to page limitation, some experimental results are put in the supplementary material.

4.1 Experimental Setup

Training Details. HR images with a resolution of 256×256 in our training data are randomly cropped from the training set of ImageNet [50] following LDM [11]. We synthesize the LR images using the degradation pipeline of Realesrgan [19]. The Adam [51] algorithm with the default settings of PyTorch [52] and a mini-batch size of 64 is used to train *ResShift*. During training, we use a fixed learning rate of $5e-5$ and update the weight parameters for 500K iterations. As for the network architecture, we employ the UNet structure in DDPM [2]. To increase the robustness of

Bảng 1: So sánh hiệu suất của *ResShift* trên *ImageNet-Test* dưới các cấu hình khác nhau.

Cấu hình			Số liệu				
T	p	κ	PSNR	SSIM	LPIPS	CLIPQA	MUSIQ
10	0.3	2.0	25,20	0,6828	0,2517	0,5492	50,6617
			25,01	0,6769	0,2312	0,5922	53,6596
			24,52	0,6585	0,2253	0,6273	55,7904
			24,29	0,6513	0,2225	0,6468	56,8482
			24,22	0,6483	0,2212	0,6489	56,8463
15	0.3	2.0	25,01	0,6769	0,2312	0,5922	53,6596
			25,05	0,6745	0,2387	0,5816	52,4475
			25,12	0,6780	0,2613	0,5314	48,4964
			25,32	0,6827	0,3050	0,4601	43,3060
			25,39	0,5813	0,3432	0,4041	38,5324
15	0.3	2.0	24,90	0,6709	0,2437	0,5700	50,6101
			24,84	0,6699	0,2354	0,5914	52,9933
			25,01	0,6769	0,2312	0,5922	53,6596
			25,31	0,6858	0,2592	0,5231	49,3182
			24,46	0,6891	0,2772	0,4898	46,9794

Hình 4: So sánh định tính của *ResShift* dưới các kết hợp khác nhau của (T, p, κ) . Ví dụ, “(15, 0.3, 2.0)” biểu thị kết quả được khôi phục với $T = 15$, $p = 0.3$ và $\kappa = 2.0$. Vui lòng phóng to để có cái nhìn tốt hơn.

4 thí nghiệm

Phần này trình bày phân tích thực nghiệm về *ResShift* được đề xuất và cung cấp kết quả thử nghiệm sâu rộng để xác minh tính hiệu quả của nó trên một tập dữ liệu tổng hợp và ba tập dữ liệu trong thế giới thực.

Theo [18, 19], cuộc điều tra của chúng tôi đặc biệt tập trung vào nhiệm vụ $\times 4$ SR đầy thách thức hơn. Quá hạn để hạn chế số trang, một số kết quả thực nghiệm được đưa vào tài liệu bổ sung.

4.1 Thiết lập thử nghiệm

Chi tiết đào tạo. Hình ảnh nhân sự có độ phân giải 256×256 trong dữ liệu đào tạo của chúng tôi được chọn ngẫu nhiên được cắt từ tập huấn luyện của ImageNet [50] theo LDM [11]. Chúng tôi tổng hợp các hình ảnh LR sử dụng đường dẫn xuống cấp của Realesrgan [19]. Thuật toán Adam [51] với giá trị mặc định cài đặt của PyTorch [52] và kích thước lô nhỏ 64 được sử dụng để huấn luyện *ResShift*. Trong quá trình đào tạo, chúng tôi sử dụng tốc độ học cố định là $5e-5$ và cập nhật các tham số trọng số cho các lần lặp 500K. Đối với kiến trúc mạng, chúng tôi sử dụng cấu trúc UNet trong DDPM [2]. Để tăng cường độ vững chắc của

Table 2: Efficiency and performance comparisons of *ResShift* to other methods on the dataset of *ImageNet-Test*. “LDM-A” represents the results achieved by accelerated the sampling steps of LDM [11] to “A”. Running time is tested on NVIDIA Tesla V100 GPU on the x4 (64→256) SR task.

Metrics	Methods						
	BSRGAN	RealESRGAN	SwinIR	LDM-15	LDM-30	LDM-100	<i>ResShift</i>
PSNR↑	24.42	24.04	23.99	24.89	24.49	23.90	25.01
LPIPS↓	0.259	0.254	0.238	0.269	0.248	0.244	0.231
CLIPQA↑	0.581	0.523	0.564	0.512	0.572	0.620	0.592
Runtime (s)	0.012	0.013	0.046	0.102	0.184	0.413	0.105
# Parameters (M)	16.70	16.70	28.01	113.60	118.59		

ResShift to arbitrary image resolution, we replace the self-attention layer in UNet with the Swin Transformer [53] block.

Testing Datasets. We synthesize a testing dataset that contains 3000 images randomly selected from the validation set of ImageNet [50] based on the commonly-used degradation model, i.e., $y = (x * k) \downarrow + n$, where k is the blurring kernel, n is the noise, y and x denote the LR image and HR image, respectively. To comprehensively evaluate the performance of *ResShift*, we consider more complicated types of blurring kernels, downsampling operators, and noise types. The detailed settings on them can be found in the supplementary material. It should be noted that we selected the HR images from ImageNet [50] instead of the prevailing datasets in SR such as *Set5* [54], *Set14* [55], and *Urban100* [56]. The rationale behind this setting is rooted in the fact that these datasets only contain very few source images, which fails to thoroughly evaluate the performance of various methods under different degradation types. We name this dataset as *ImageNet-Test* for convenience.

Two real-world datasets are adopted to evaluate the efficacy of *ResShift*. The first is *RealSR* [57], containing 100 real images captured by Canon 5D3 and Nikon D810 cameras. Additionally, we collect another real-world dataset named *RealSet65*. It comprises 35 LR images widely used in recent literature [19, 58, 59, 60, 61]. The remaining 30 images were obtained from the internet by ourselves.

Compared Methods. We evaluate the effectiveness of *ResShift* in comparison to seven recent SR methods, namely ESRGAN [62], RealSR-JPEG [63], BSRGAN [18], Realesrgan [19], SwinIR [20], DASR [21], and LDM [11]. Note that LDM is a diffusion-based method with 1,000 diffusion steps. For a fair comparison, we accelerate LDM to the same number of steps with *ResShift* using DDIM [16] and denote it as “LDM-A”, where “A” indicates the number of inference steps. The hyper-parameter η in DDIM is set to be 1 as this value yields the most realistic recovered images.

Metrics. The performance of various methods was assessed using five metrics, including PSNR, SSIM [64], LPIPS [65], MUSIQ [66], and CLIPQA [67]. It is worth noting that the latter two are non-reference metrics specifically designed to assess the realism of images. CLIPQA, in particular, leverages the CLIP [68] model that is pre-trained on a massive dataset (i.e., Laion400M [69]) and thus demonstrates strong generalization ability. On the real-world datasets, we mainly rely on CLIPQA and MUSIQ as evaluation metrics to compare the performance of different methods.

4.2 Model Analysis

We analyze the performance of *ResShift* under different settings on the number of diffusion steps T and the hyper-parameters p in Eq. (10) and κ in Eq. (1).

Diffusion Steps T and Hyper-parameter p . The proposed transition distribution in Eq. (1) significantly reduces the diffusion steps T in the Markov chain. The hyper-parameter p allows for flexible control over the speed of residual shifting during the transition. Table 1 summarizes the performance of *ResShift* on *ImageNet-Test* under different configurations of T and p . We can see that both of T and p render a trade-off between the fidelity, measured by the reference metrics such as PSNR, SSIM, and LPIPS, and the realism, measured by the non-reference metrics, including CLIPQA and MUSIQ, of the super-resolved results. Taking p as an example, when it increases, the reference metrics improve while the non-reference metrics deteriorate. Furthermore, the visual comparison in Fig. 4 shows that a large value of p will suppress the model’s ability to hallucinate more image details and result in blurry outputs.

Hyper-parameter κ . Equation (2) reveals that κ dominates the noise strength in state x_t . We report the influence of κ to the performance of *ResShift* in Table 1. Combining with the visualization in

Bảng 2: So sánh hiệu quả và hiệu suất của *ResShift* với các phương pháp khác trên tập dữ liệu của *ImageNet-Test*. “LDM-A” thể hiện kết quả đạt được bằng cách đẩy nhanh các bước lấy mẫu của LDM [11] tới “A”. Thời gian chạy được thử nghiệm trên GPU NVIDIA Tesla V100 trong tác vụ SR x4 (64→256).

Số liệu	phương pháp						
	BSRGAN	RealESRGAN	SwinIR	LDM-15	LDM-30	LDM-100	Độ phân giải
PSNR	24,42	24,04	24,89	0,283	0,9269	0,523	24,49
LPIPS	0,259	0,512	0,013	0,102238,70			0,248
CLIPQA	0,581			0,564			0,620
(Các) thời gian chạy	0,012			0,046			0,105
# Thông số (M)	16,70			28,01			118,59

ResShift sang độ phân giải hình ảnh tùy ý, chúng ta thay thế lớp tự chú ý trong UNet bằng lớp Swin Transformer [53].

Kiểm tra bộ dữ liệu. Chúng tôi tổng hợp tập dữ liệu thử nghiệm chứa 3000 hình ảnh được chọn ngẫu nhiên từ bộ xác thực của ImageNet [50] dựa trên mô hình suy thoái thường được sử dụng, nghĩa là $y = (x * k) \downarrow + n$, trong đó k là hạt nhân làm mờ, n là nhiễu, y và x biểu thị ảnh LR và ảnh ảnh nhân sự, tương ứng. Để đánh giá toàn diện hiệu suất của *ResShift*, chúng tôi xem xét thêm các loại hạt nhân làm mờ phức tạp, toán tử lấy mẫu xuống và các loại nhiễu. Các cài đặt chi tiết về chúng có thể được tìm thấy trong tài liệu bổ sung. Cần lưu ý rằng chúng tôi đã chọn nhân sự hình ảnh từ ImageNet [50] thay vì các bộ dữ liệu phổ biến trong SR như Set5 [54], Set14 [55] và Đô thị100 [56]. Lý do đầu tiên là các bộ dữ liệu phổ biến trong SR như Set5 [54], Set14 [55] và Đô thị100 [56]. Lý do đầu sau là bắt nguồn từ thực tế là những bộ dữ liệu này chỉ chứa rất ít hình ảnh nguồn, không thể đánh giá kỹ lưỡng hiệu suất của các phương pháp khác nhau theo các loại suy thoái khác nhau. Chúng tôi đặt tên tập dữ liệu này là *ImageNet-Test* để thuận tiện.

Hai bộ dữ liệu trong thế giới thực được sử dụng để đánh giá hiệu quả của *ResShift*. Đầu tiên là *RealSR* [57], chứa 100 ảnh thật được chụp bằng máy ảnh Canon 5D3 và Nikon D810. Ngoài ra, chúng tôi thu thập một tập dữ liệu thực tế khác có tên *RealSet65*. Nó bao gồm 35 hình ảnh LR được sử dụng rộng rãi trong thời gian gần đây vẫn học [19, 58, 59, 60, 61]. 30 hình ảnh còn lại là do chúng tôi lấy từ internet.

Các phương pháp so sánh *Chúng tôi đánh giá tính hiệu quả của *ResShift* so với bản gốc* Các phương pháp SR, cụ thể là ESRGAN [62], RealSR-JPEG [63], BSRGAN [18], Realesrgan [19], SwinIR [20], DASR [21] và LDM [11]. Lưu ý rằng LDM là phương pháp dựa trên sự khuếch tán với 1.000 các bước khuếch tán. Để so sánh công bằng, chúng tôi tăng tốc LDM lên cùng số bước bằng *ResShift* sử dụng DDIM [16] và ký hiệu là “LDM-A”, trong đó “A” biểu thị số bước suy luận. Các siêu tham số η trong DDIM được đặt thành 1 vì giá trị này mang lại hình ảnh được phục hồi chân thực nhất.

Số liệu. Hiệu suất của các phương pháp khác nhau được đánh giá bằng năm số liệu, bao gồm PSNR, SSIM [64], LPIPS [65], MUSIQ [66] và CLIPQA [67]. Điều đáng chú ý là hai cái sau là số liệu không tham chiếu được thiết kế đặc biệt để đánh giá tính chân thực của hình ảnh. CLIPQA, đặc biệt, tận dụng mô hình CLIP [68] được đào tạo trước trên tập dữ liệu lớn (ví dụ: Laion400M [69]) và do đó thể hiện khả năng khái quát hóa mạnh mẽ. Trên các bộ dữ liệu trong thế giới thực, chúng tôi chủ yếu dựa vào CLIPQA và MUSIQ làm thước đo đánh giá để so sánh hiệu suất của các phương pháp khác nhau.

4.2 Phân tích mô hình

Chúng tôi phân tích hiệu suất của *ResShift* trong các cài đặt khác nhau về số bước khuếch tán T và các siêu tham số p trong biểu thức. (10) và κ trong biểu thức. (1).

Các bước khuếch tán T và siêu tham số p . Phân phối chuyển tiếp được đề xuất trong biểu thức. (1) giảm đáng kể các bước khuếch tán T trong chuỗi Markov. Siêu tham số p cho phép linh hoạt kiểm soát tốc độ dịch chuyển dữ trong quá trình chuyển đổi. Bảng 1 tóm tắt hiệu suất của *ResShift* trên *ImageNet-Test* dưới các cấu hình khác nhau của T và p . Chúng ta có thể thấy rằng cả hai T và p thể hiện sự đánh đổi giữa độ trung thực, được đo bằng các số liệu tham chiếu như PSNR, SSIM và LPIPS cũng như tính hiện thực được đo bằng các số liệu không tham chiếu, bao gồm CLIPQA và MUSIQ, kết quả siêu giải quyết. Lấy p làm ví dụ, khi nó tăng thì tham chiếu số liệu cải thiện trong khi số liệu không tham chiếu xấu đi. Hơn nữa, việc so sánh trực quan trong Hình 4 cho thấy giá trị p lớn sẽ ngăn chặn khả năng tạo ảnh của mô hình hơn chi tiết và dẫn đến kết quả đậm đà.

Siêu tham số κ . Phương trình (2) cho thấy κ chiếm ưu thế về cường độ nhiễu ở trạng thái x_t . Chúng tôi báo cáo ảnh hưởng của κ đến hiệu suất của *ResShift* trong Bảng 1. Kết hợp với trực quan hóa trong

Table 3: Quantitative results of different methods on the dataset of *ImageNet-Test*. The best and second best results are highlighted in **bold** and underline.

Methods	Metrics				
	PSNR↑	SSIM↑	LPIPS↓	CLIPQA↑	MUSIQ↑
ESRGAN [62]	20.67	0.448	0.485	0.451	43.615
RealSR-JPEG [63]	23.11	0.591	0.326	0.537	46.981
BSRGAN [18]	24.42	0.659	0.259	<u>0.581</u>	54.697
SwinIR [20]	23.99	0.667	<u>0.238</u>	0.564	53.790
RealESRGAN [19]	24.04	0.665	0.254	0.523	52.538
DASR [21]	24.75	<u>0.675</u>	0.250	0.536	48.337
LDM-15 [11]	<u>24.89</u>	0.670	0.269	0.512	46.419
ResShift	25.01	0.677	0.231	0.592	53.660

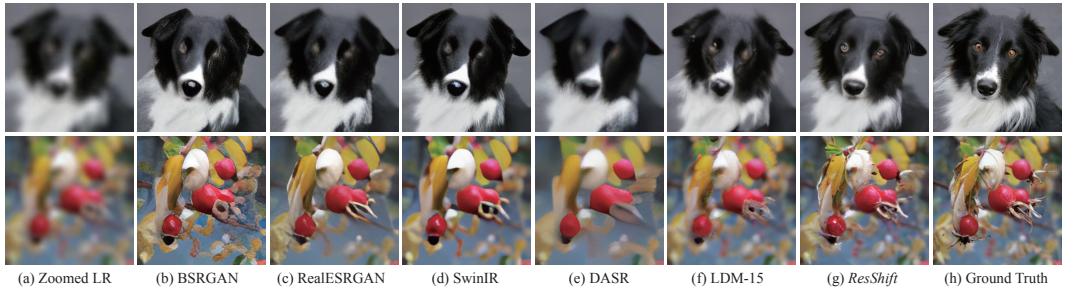


Figure 5: Qualitative comparisons of different methods on two synthetic examples of the *ImageNet-Test* dataset. Please zoom in for a better view.

Fig. 4, we can find that excessively large or small values of κ will smooth the recovered results, regardless of their favorable metrics of PSNR and SSIM. When κ is in the range of [1.0, 2.0], our method achieves the most realistic quality indicated by CLIPQA and MUSIQ, which is more desirable in real applications. We thus set κ to be 2.0 in this work.

Efficiency Comparison. To improve inference efficiency, it is desirable to limit the number of diffusion steps T . However, this causes a decrease in the realism of the restored HR images. To compromise, the hyper-parameter p can be set to a relatively small value. Therefore, we set $T = 15$ and $p = 0.3$, and yield our model named *ResShift*. Table 2 presents the efficiency and performance comparisons of *ResShift* to the state-of-the-art (SotA) approach LDM [11] and three other GAN-based methodologies on *ImageNet-Test* dataset. It is evident from the results that the proposed *ResShift* surpasses LDM [11] in terms of PSNR and LPIPS [65], and demonstrates a remarkable fourfold enhancement in computational efficiency when compared to LDM-100. Despite showing considerable potential in mitigating the efficiency bottleneck of the diffusion-based SR approaches, *ResShift* still lags behind current GAN-based methods in speed due to its iterative sampling mechanism. Therefore, it remains imperative to explore further optimizations of the proposed method to address this limitation, which we leave in our future work.

Perception-Distortion Trade-off. There exists a well-known phenomenon called perception-distortion trade-off [70] in the field of SR. In particular, the augmentation of the generative capability of a restoration model, such as elevating the sampling steps for a diffusion-based method or amplifying the weight of the adversarial loss for a GAN-based method, will result in a deterioration in fidelity preservation while concurrently enhancing the authenticity of restored images. That is mainly because the restoration model with powerful generation capability tends to hallucinate more high-frequency image structures, thereby deviating from the underlying ground truth. To facilitate a comprehensive comparison between our *ResShift* and current SotA diffusion-based method LDM, we plotted the perception-distortion curves of them in Fig. 7, wherein the perception and distortion are mea-

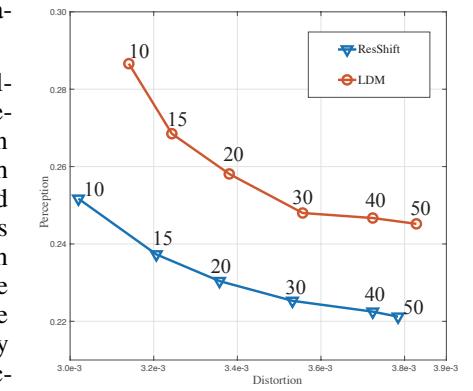
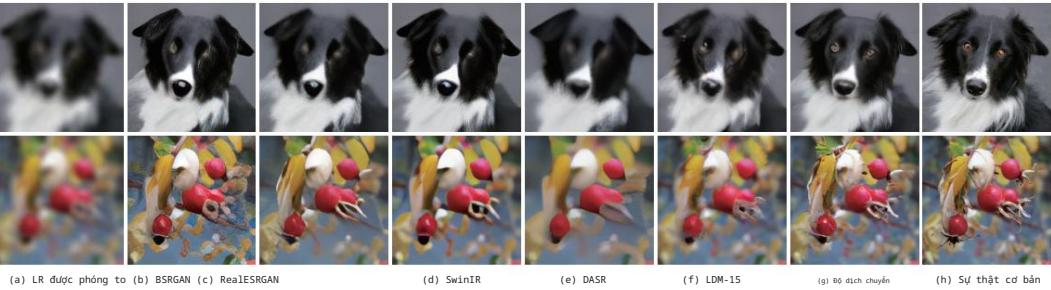


Figure 7: Perception-distortion trade-off of *ResShift* and LDM. The vertical and horizontal axes represent the strength of the perception and distortion, measured by LPIPS and MSE, respectively.

Bảng 3: Kết quả định lượng của các phương pháp khác nhau trên tập dữ liệu của *ImageNet-Test*. Tốt nhất và kết quả tốt thứ hai được đánh dấu bằng chữ in đậm và gạch chân.

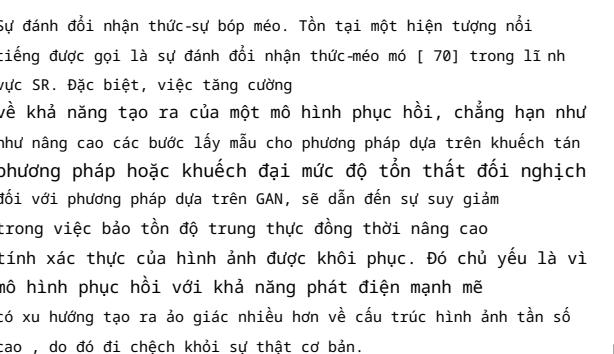
phương pháp	Số liệu			
	PSNR	SSIM	LPIPS	CLIPQA
ESRGAN [62]	20,67	0,448	0,485	0,451
RealSR-JPEG [63]	23,11	0,591	0,326	0,537
BSRGAN [18]	24,42	0,659	0,259	<u>0,581</u>
SwinIR [20]	23,99	0,667	<u>0,238</u>	0,564
RealESRGAN [19]	24,04	0,665	0,254	0,523
DASR [21]	24,75	<u>0,675</u>	0,250	0,536
LDM-15 [11]	<u>24,89</u>	0,670	0,269	0,512
ResShift	25,01	0,677	0,231	0,592



Hình 5: So sánh định tính của các phương pháp khác nhau trên hai ví dụ tổng hợp của bộ dữ liệu *ImageNet-Test*. Vui lòng phóng to để nhìn rõ hơn.

Hình 4, chúng ta có thể thấy rằng các giá trị κ quá lớn hoặc quá nhỏ sẽ làm mịn các kết quả được khôi phục, bất kể số liệu thuận lợi của họ về PSNR và SSIM. Khi κ nằm trong khoảng [1.0, 2.0], phương pháp đạt được chất lượng thực tế nhất được chỉ ra bởi CLIPQA và MUSIQ, cao hơn mong muốn trong các ứng dụng thực tế. Do đó chúng tôi đặt κ là 2,0 trong công việc này.

So sánh hiệu quả. Để nâng cao hiệu quả suy luận, người ta mong muốn hạn chế số lượng bước khuếch tán T . Tuy nhiên, điều này làm giảm tính chân thực của hình ảnh nhân sự được khôi phục. ĐẾN thỏa hiệp, siêu tham số p có thể được đặt thành một giá trị tương đối nhỏ. Vì vậy ta đặt $T = 15$ và $p = 0.3$ và mang lại mô hình của chúng tôi có tên *ResShift*. Bảng 2 trình bày hiệu quả và hiệu suất so sánh *ResShift* với phương pháp tiếp cận hiện đại (SotA) LDM [11] và ba phương pháp dựa trên GAN khác các phương pháp trên tập dữ liệu *ImageNet-Test*. Rõ ràng từ kết quả là *ResShift* được đề xuất vượt qua LDM [11] về PSNR và LPIPS [65], và chứng minh hiệu quả gấp bốn lần đáng chú ý nâng cao hiệu quả tính toán khi so sánh với LDM-100. Mặc dù thẻ hiện đáng kể tiềm năng trong việc giảm thiểu nút thất hiệu quả của các phương pháp SR dựa trên khuếch tán, *ResShift* vẫn tụt hậu so với các phương pháp dựa trên GAN hiện tại về tốc độ do cơ chế lấy mẫu lặp lại của nó. Do đó, điều bắt buộc là phải khám phá những tối ưu hóa hơn nữa của phương pháp được đề xuất để giải quyết hạn chế này mà chúng tôi sẽ đề lại trong công việc tương lai của mình.



Hình 7: Sự đánh đổi giữa nhận thức và b López mỏ của *ResShift* và LDM. Chiều dọc và trực ngang thể hiện sức mạnh của nhận thức và sự biến dạng, được đo lường bởi LPIPS và MSE tương ứng.

Table 4: Quantitative results of different methods on two real-world datasets. The best and second best results are highlighted in **bold** and underline.

Methods	Datasets			
	<i>RealSR</i>		<i>RealSet65</i>	
	CLIPQA↑	MUSIQ↑	CLIPQA↑	MUSIQ↑
ESRGAN [62]	0.2362	29.048	0.3739	42.369
RealSR-JPEG [63]	0.3615	36.076	0.5282	50.539
BSRGAN [18]	<u>0.5439</u>	63.586	0.6163	65.582
SwinIR [20]	0.4654	59.636	0.5782	<u>63.822</u>
RealESRGAN [19]	0.4898	59.678	0.5995	63.220
DASR [21]	0.3629	45.825	0.4965	55.708
LDM-15 [11]	0.3836	49.317	0.4274	47.488
<i>ResShift</i>	0.5958	<u>59.873</u>	0.6537	61.330

sured by LPIPS and mean square-error (MSE), respectively. This plot reflects the perception quality and the reconstruction fidelity of *ResShift* and LDM across varying numbers of diffusion steps, i.e., 10, 15, 20, 30, 40, and 50. As can be observed, the perception-distortion curve of our *ResShift* consistently resides beneath that of the LDM, indicating its superior capacity in balancing perception and distortion.

4.3 Evaluation on Synthetic Data

We present a comparative analysis of the proposed method with recent SotA approaches on the *ImageNet-Test* dataset, as summarized in Table 3 and Fig. 5. Based on this evaluation, several significant conclusions can be drawn as follows: i) *ResShift* exhibits superior or at least comparable performance across all five metrics, affirming the effectiveness and superiority of the proposed method. ii) The notably higher PSNR and SSIM values attained by *ResShift* indicate its capacity to better preserve fidelity to ground truth images. This advantage primarily arises from our well-designed diffusion model, which starts from a subtle disturbance of the LR image, rather than the conventional assumption of white Gaussian noise in LDM. iii) Considering the metrics of LPIPS and CLIPQA, which gauge the perceptual quality and realism of the recovered image, *ResShift* also demonstrates evident superiority over existing methods. Furthermore, in terms of MUSIQ, our approach achieves comparable performance with recent SotA methods. In summary, the proposed *ResShift* exhibits remarkable capabilities in generating more realistic results while preserving fidelity. This is of paramount importance for the task of SR.

4.4 Evaluation on Real-World Data

Table 4 lists the comparative evaluation using CLIPQA [67] and MUSIQ [66] of various methods on two real-world datasets. Note that CLIPQA, benefiting from the powerful representative capability inherited from CLIP, performs stably and robustly in assessing the perceptual quality of natural images. The results in Table 4 show that the proposed *ResShift* evidently surpasses existing methods in CLIPQA, meaning that the restored outputs of *ResShift* better align with human visual and perceptive systems. In the case of MUSIQ evaluation, *ResShift* achieves the competitive performance when compared to current SotA methods, namely BSRGAN [18], SwinIR [20], and RealESRGAN [19]. Collectively, our method shows promising capability in addressing the real-world SR problem.

We display four real-world examples in Fig. 6. More examples can be found in the supplementary material. We consider diverse scenarios, including comic, text, face, and natural images to ensure a comprehensive evaluation. A noticeable observation is that *ResShift* produces more naturalistic image structures, as evidenced by the patterns on the beam in the third example and the eyes of a person in the fourth example. We note that the recovered results of LDM are excessively smooth when compressing the inference steps to match with the proposed *ResShift*, specifically utilizing 15 steps, largely deviating from the training procedure's 1,000 steps. Even though other GAN-based methods may also succeed in hallucinating plausible structures to some extent, they are often accompanied by obvious artifacts.

5 Conclusion

In this work, we have introduced an efficient diffusion model named *ResShift* for SR. Unlike existing diffusion-based SR methods that require a large number of iterations to achieve satisfactory results,

Bảng 4: Kết quả định lượng của các phương pháp khác nhau trên hai bộ dữ liệu trong thế giới thực. Tốt nhất và thứ hai kết quả tốt nhất được đánh dấu bằng chữ in đậm và gạch chân.

phương pháp	Bộ dữ liệu			
	<i>RealSR</i>		<i>RealSet65</i>	
	CLIPQA	MUSIQ	CLIPQA	MUSIQ
ESRGAN [62]	0,2362	29.048	0,3739	42.369
RealSR-JPEG [63]	0,3615	36.076	0,5282	50.539
BSRGAN [18]	<u>0,5439</u>	63.586	0,6163	65.582
SwinIR [20]	0,4654	59.636	0,5782	63.822
RealESRGAN [19]	0,4898	59.678	0,5995	63.220
DASR [21]	0,3629	45.825	0,4965	55.708
LDM-15 [11]	0,3836	49.317	0,4274	47.488
<i>ResShift</i>	0,5958	<u>59.873</u>	0.6537	61.330

được đảm bảo bởi LPIPS và sai số bình phương trung bình (MSE). Cốt truyện này phản ánh chất lượng nhận thức và độ trung thực tái cấu trúc của *ResShift* và LDM qua nhiều bước khuếch tán khác nhau, nghĩa là 10, 15, 20, 30, 40, và 50. Có thể thấy, đường cong nhận thức-biến dạng của *ResShift* của chúng tôi luôn nằm dưới LDM, cho thấy khả năng vượt trội của nó trong việc cân bằng nhận thức và sự biến dạng.

4.3 Đánh giá dữ liệu tổng hợp

Chúng tôi trình bày một phân tích so sánh của phương pháp được đề xuất với các phương pháp tiếp cận SotA gần đây trên Tập dữ liệu *ImageNet-Test*, như được tóm tắt trong Bảng 3 và Hình 5. Dựa trên đánh giá này, một số có thể rút ra những kết luận quan trọng như sau: i) *ResShift* thể hiện tính năng vượt trội hoặc ít nhất có thể so sánh được trên cả 5 chỉ số, khẳng định tính hiệu quả và ưu việt của phương pháp đề xuất. ii) Giá trị PSNR và SSIM cao hơn đáng kể mà *ResShift* đạt được cho thấy khả năng của nó tốt hơn duy trì độ trung thực của hình ảnh chân thực. Lợi thế này chủ yếu phát sinh từ việc chúng tôi thiết kế tốt mô hình khuếch tán, bắt đầu từ sự nhiễu loạn tinh tế của hình ảnh LR, thay vì mô hình thông thường giả định nhiễu Gauss trắng trong LDM. iii) Xem xét các số liệu LPIPS và CLIPQA, đánh giá chất lượng cảm nhận và tính chân thực của hình ảnh được khôi phục, *ResShift* cũng chứng minh sự vượt trội rõ ràng so với các phương pháp hiện có. Hơn nữa, về mặt MUSIQ, cách tiếp cận của chúng tôi đạt được hiệu suất tương đương với các phương pháp SotA gần đây. Tóm lại, *ResShift* đề xuất thể hiện khả năng vượt trội trong việc tạo ra kết quả thực tế hơn trong khi vẫn giữ được độ trung thực. Đây là của tầm quan trọng hàng đầu đối với nhiệm vụ của SR.

4.4 Đánh giá dữ liệu thực tế

Bảng 4 liệt kê đánh giá so sánh sử dụng CLIPQA [67] và MUSIQ [66] của các phương pháp khác nhau về hai bộ dữ liệu trong thế giới thực. Lưu ý rằng CLIPQA, được hưởng lợi từ khả năng đại diện mạnh mẽ kế thừa từ CLIP, hoạt động ổn định và mạnh mẽ trong việc đánh giá chất lượng cảm nhận của thiên nhiên hình ảnh. Kết quả trong Bảng 4 cho thấy *ResShift* được đề xuất rõ ràng vượt trội hơn các phương pháp hiện có trong CLIPQA, nghĩa là các điều kiện khôi phục của *ResShift* phù hợp hơn với thị giác và khả năng nhận thức của con người hệ thống. Trong trường hợp đánh giá MUSIQ, *ResShift* đạt được hiệu suất cạnh tranh khi so với các phương pháp SotA hiện tại, cụ thể là BSRGAN [18], SwinIR [20] và RealESRGAN [19].

Nói chung, phương pháp của chúng tôi cho thấy khả năng đầy hứa hẹn trong việc giải quyết vấn đề SR trong thế giới thực.

Chúng tôi hiển thị bốn ví dụ thực tế trong Hình 6. Bạn có thể tìm thấy nhiều ví dụ khác trong phần bổ sung vật liệu. Chúng tôi xem xét các kịch bản đa dạng, bao gồm truyện tranh, văn bản, khuôn mặt và hình ảnh tự nhiên để đảm bảo đánh giá toàn diện. Một quan sát đáng chú ý là *ResShift* tạo ra hình ảnh tự nhiên hơn cấu trúc, được chứng minh bằng các mẫu trên chùm tia trong ví dụ thứ ba và đôi mắt của một người trong ví dụ thứ tư. Chúng tôi lưu ý rằng kết quả thu hồi của LDM quá tròn trịa khi nên các bước suy luận để phù hợp với *ResShift* được đề xuất, cụ thể sử dụng 15 bước, phần lớn đi chệch khỏi 1.000 bước của quy trình đào tạo. Mặc dù các phương pháp dựa trên GAN khác cũng có thể thành công trong việc tạo ra ảo giác về các cấu trúc hợp lý ở một mức độ nào đó, chúng thường đi kèm với hiện vật rõ ràng.

5 Kết luận

Trong công việc này, chúng tôi đã giới thiệu một mô hình khuếch tán hiệu quả có tên *ResShift* cho SR. Không giống như hiện có, phương pháp SR dựa trên khuếch tán yêu cầu số lần lặp lớn để đạt được kết quả khả quan,



Figure 6: Qualitative comparisons on four real-world examples. Please zoom in for a better view.

our proposed method constructs a diffusion model with only 15 sampling steps, thereby significantly improving inference efficiency. The core idea is to corrupt the HR image toward the LR image instead of the Gaussian white noise, which can effectively cut off the length of the diffusion model. Extensive experiments on both synthetic and real-world datasets have demonstrated the superiority of our proposed method. We believe that our work will pave the way for the development of more efficient and effective diffusion models to address the SR problem.

Acknowledgement. This study is supported under the RIE2020 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from the industry partner(s).



Hình 6: So sánh định tính trên bốn ví dụ trong thế giới thực. Vui lòng phóng to để nhìn rõ hơn.

phương pháp đề xuất của chúng tôi xây dựng một mô hình khuếch tán chỉ với 15 bước lấy mẫu, từ đó cải thiện đáng kể hiệu quả suy luận. Ý tưởng cốt lõi là làm hỏng hình ảnh HR đối với hình ảnh LR thay vì nhiễu trắng Gaussian, điều này có thể cắt giảm độ dài của mô hình khuếch tán một cách hiệu quả. Các thử nghiệm mở rộng trên cả bộ dữ liệu tổng hợp và thế giới thực đã chứng minh tính ưu việt của phương pháp được đề xuất của chúng tôi. Chúng tôi tin rằng công việc của chúng tôi sẽ mở đường cho việc phát triển các mô hình khuếch tán hiệu quả và hiệu quả hơn để giải quyết vấn đề SR.

Nhìn nhận. Nghiên cứu này được hỗ trợ trong khuôn khổ Sáng kiến tài trợ của Quỹ liên kết ngành RIE2020 - Dự án hợp tác ngành (IAF-ICP), cũng như đóng góp bằng tiền mặt và hiện vật từ (các) đối tác trong ngành.

References

- [1] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning (ICML)*, pages 2256–2265. PMLR, 2015.
- [2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 6840–6851, 2020.
- [3] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 8780–8794, 2021.
- [4] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2021.
- [5] Omri Avrahami, Dani Lischinski, and Ohad Fried. Blended diffusion for text-driven editing of natural images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18208–18218, 2022.
- [6] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11461–11471, June 2022.
- [7] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12413–12422, 2022.
- [8] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2021.
- [9] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *Proceedings of ACM SIGGRAPH Conference*, pages 1–10, 2022.
- [10] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2022.
- [11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022.
- [12] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14367–14376, 2021.
- [13] Zongsheng Yue and Chen Change Loy. Difface: Blind face restoration with diffused error contraction. *arXiv preprint arXiv:2212.06512*, 2022.
- [14] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C.K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023.
- [15] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning (ICML)*, pages 8162–8171. PMLR, 2021.
- [16] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2021.
- [17] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [18] 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 International Conference on Computer Vision (ICCV)*, pages 4791–4800, 2021.

Người giới thiệu

- [1] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan và Surya Ganguli. Học tập không giám sát sâu bằng cách sử dụng nhiệt động lực học không cân bằng. Trong Hội nghị quốc tế về học máy (ICML), trang 2256–2265. PMLR, 2015.
- [2] Jonathan Ho, Ajay Jain và Pieter Abbeel. Mô hình xác suất khuếch tán khử nhiễu. Trong Ký yếu về những tiến bộ trong hệ thống xử lý thông tin thần kinh (NeurIPS), tập 33, trang 6840–6851, 2020.
- [3] Prafulla Dhariwal và Alexander Nichol. Các mô hình khuếch tán đánh bại các tổ chức về tổng hợp hình ảnh. Trong Ký yếu về những tiến bộ trong hệ thống xử lý thông tin thần kinh (NeurIPS), tập 34, trang 8780–8794, 2021.
- [4] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu và Stefano Ermon. Sdedit: Tổng hợp và chỉnh sửa hình ảnh có hướng dẫn với các phương trình vi phân ngẫu nhiên. Trong Ký yếu của Hội nghị Quốc tế về Đại diện Học tập (ICLR), 2021.
- [5] Omri Avrahami, Dani Lischinski và Ohad Fried. Khuếch tán hỗn hợp để chỉnh sửa hình ảnh tự nhiên theo hướng văn bản. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 18208–18218, 2022.
- [6] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte và Luc Van Gool. Sơn lại: Sơn lại bằng cách sử dụng mô hình xác suất khuếch tán khử nhiễu. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 11461–11471, tháng 6 năm 2022.
- [7] Hyungjin Chung, Byeongsu Sim và Jong Chul Ye. Đến gần hơn-khuếch tán-nhanh hơn: Tăng tốc các mô hình khuếch tán có điều kiện cho các vấn đề nghịch đảo thông qua sự co lại ngẫu nhiên. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 12413–12422, 2022.
- [8] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon và Ben Poole. Mô hình tổng quát dựa trên điểm số thông qua các phương trình vi phân ngẫu nhiên. Trong Ký yếu của Hội nghị Quốc tế về Đại diện Học tập (ICLR), 2021.
- [9] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet và Mohammad Norouzi. Bảng màu: Mô hình khuếch tán hình ảnh sang hình ảnh. Trong Ký yếu của Hội nghị ACM SIGGRAPH, trang 1–10, 2022.
- [10] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet và Mohammad Norouzi. Hình ảnh siêu phân giải thông qua tinh chỉnh lặp đi lặp lại. Giao dịch của IEEE về Phân tích Mẫu và Trí thông minh Máy (TPAMI), 2022.
- [11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser và Björn Ommer. Tổng hợp hình ảnh có độ phân giải cao với các mô hình khuếch tán tiềm ẩn. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 10684–10695, 2022.
- [12] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon và Sungroh Yoon. Ilvr: Phương pháp điều hòa để khử nhiễu các mô hình xác suất khuếch tán. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), trang 14367–14376, 2021.
- [13] Zongsheng Yue và Chen Change Loy. Difface: Phục hồi mặt mù với sự co lại lỗi khuếch tán. Bản in trước arXiv arXiv:2212.06512, 2022.
- [14] Jianyi Wang, Zongsheng Yue, Shangchen Chu, Kelvin CK Chan và Chen Change Loy. Khai thác khả năng khuếch tán trước để có độ phân giải siêu cao của hình ảnh trong thế giới thực. Bản in trước arXiv arXiv:2305.07015, 2023.
- [15] Alexander Quinn Nichol và Prafulla Dhariwal. Cải thiện mô hình xác suất khuếch tán khử nhiễu. Trong Hội nghị Quốc tế về học máy (ICML), trang 8162–8171. PMLR, 2021.
- [16] Jiaming Song, Chenlin Meng và Stefano Ermon. Các mô hình tiềm ẩn khuếch tán khử nhiễu. Trong Ký yếu của Hội nghị Quốc tế về đại diện học tập (ICLR), 2021.
- [17] Cheng Lu, Yuhao Chu, Fan Bao, Jianfei Chen, Chongxuan Li và Jun Zhu. Bộ giải DPM: Bộ giải ODE nhanh để lấy mẫu mô hình xác suất khuếch tán trong khoảng 10 bước. Trong Ký yếu về những tiến bộ trong hệ thống xử lý thông tin thần kinh (NeurIPS), 2022.
- [18] Kai Zhang, Jingyun Liang, Luc Van Gool và Radu Timofte. Thiết kế mô hình suy giảm thực tế cho hình ảnh siêu phân giải mù sâu. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), trang 4791–4800, 2021.

- [19] 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 International Conference on Computer Vision Workshops (ICCV-W)*, pages 1905–1914, 2021.
- [20] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCV-W)*, pages 1833–1844, 2021.
- [21] Jie Liang, Hui Zeng, and Lei Zhang. Efficient and degradation-adaptive network for real-world image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 574–591, 2022.
- [22] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12873–12883, 2021.
- [23] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.
- [24] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi, and William Chan. WaveGrad: estimating gradients for waveform generation. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2020.
- [25] Chenhao Niu, Yang Song, Jiaming Song, Shengjia Zhao, Aditya Grover, and Stefano Ermon. Permutation invariant graph generation via score-based generative modeling. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108, pages 4474–4484, 2020.
- [26] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan. Learning gradient fields for shape generation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 364–381, 2020.
- [27] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing (TIP)*, 22(4):1620–1630, 2012.
- [28] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision (IJCV)*, 121:183–208, 2017.
- [29] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing (TIP)*, 20(7):1838–1857, 2011.
- [30] Shuhang Gu, Wangmeng Zuo, Qi Xie, Deyu Meng, Xiangchu Feng, and Lei Zhang. Convolutional sparse coding for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1823–1831, 2015.
- [31] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 38(2):295–307, 2015.
- [32] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883, 2016.
- [33] 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 (TIP)*, 26(7):3142–3155, 2017.
- [34] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 624–632, 2017.
- [35] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1664–1673, 2018.
- [19] Xintao Wang, Liangbin Xie, Chao Dong và Ying Shan. Real-esrgan: Đào tạo siêu phân giải mù trong thế giới thực bằng dữ liệu tổng hợp thuần túy. Trong Kỷ yếu của Hội nghị Quốc tế IEEE/CVF về Hội thảo Thị giác Máy tính (ICCV-W), trang 1905-1914, 2021.
- [20] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, và Radu Timofte. Swinir: Phục hồi ảnh bằng biến áp swin. Trong Kỷ yếu của Hội nghị Quốc tế IEEE/CVF về Hội thảo Thị giác Máy tính (ICCV-W), trang 1833-1844, 2021.
- [21] Jie Liang, Hui Zeng và Lei Zhang. Mạng hiệu quả và có khả năng thích ứng với suy thoái cho hình ảnh có độ phân giải siêu cao trong thế giới thực . Trong Kỷ yếu của Hội nghị Châu Âu về Thị giác Máy tính (ECCV), trang 574-591, 2022.
- [22] Patrick Esser, Robin Rombach và Bjorn Ommer. Biến áp thuần hóa để tổng hợp hình ảnh có độ phân giải cao. Trong Kỷ yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 12873-12883, 2021.
- [23] Yang Song và Stefano Ermon. Mô hình hóa tổng quát bằng cách ước tính độ dốc của phân phối dữ liệu. Trong Kỷ yếu về những tiến bộ trong hệ thống xử lý thông tin thần kinh (NeurIPS), tập 32, 2019.
- [24] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi và William Chan. WaveGrad: ước tính độ dốc để tạo dạng sóng. Trong Kỷ yếu của Hội nghị Quốc tế về Đại diện Học tập (ICLR), 2020.
- [25] Chenhao Niu, Yang Song, Jiaming Song, Shengjia Zhao, Aditya Grover và Stefano Ermon. Tạo đồ thị bắt biến hoàn vị thông qua mô hình tổng quát dựa trên điểm số. Trong Kỷ yếu của Hội nghị quốc tế về trí tuệ nhân tạo và thống kê (AISTATS), tập 108, trang 4474-4484, 2020.
- [26] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely và Bharath Hariharan. Học các trường gradient để tạo hình. Trong Kỷ yếu của Hội nghị Châu Âu về Thị giác Máy tính (ECCV), trang 364-381, 2020.
- [27] Weisheng Dong, Lei Zhang, Quang Minh Shi, và Xin Li. Biểu diễn thua thoát tập trung không cục bộ để khôi phục hình ảnh. Giao dịch của IEEE về Xử lý Hình ảnh (TIP), 22(4):1620-1630, 2012.
- [28] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng và Lei Zhang. Giảm thiểu định mức hạt nhân có trọng số và các ứng dụng của nó đối với tầm nhìn ở mức độ thấp. Tập chí Quốc tế về Thị giác Máy tính (IJCV), 121:183-208, 2017.
- [29] Weisheng Dong, Lei Zhang, Quang Minh Shi, và Xiaolin Wu. Làm mờ hình ảnh và siêu phân giải bằng cách chọn miền thua thoát thích ứng và chính quy hóa thích ứng. Giao dịch của IEEE về Xử lý Hình ảnh (TIP), 20(7):1838-1857, 2011.
- [30] Shuhang Gu, Wangmeng Zuo, Qi Xie, Deyu Meng, Xiangchu Feng và Lei Zhang. Mã hóa thua thoát cho hình ảnh có độ phân giải siêu cao. Trong Kỷ yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), trang 1823-1831, 2015.
- [31] Chao Dong, Chen Change Loy, Kaiming He và Xiaoou Tang. Hình ảnh siêu phân giải sử dụng mạng tích chập sâu. Giao dịch của IEEE về Phân tích Mẫu và Trí thông minh Máy (TPAMI), 38(2): 295-307, 2015.
- [32] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert và Zehan Wang. Siêu phân giải hình ảnh và video đơn thời gian thực bằng cách sử dụng mạng thần kinh tích chập pixel phụ hiệu quả. Trong Kỷ yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 1874-1883, 2016.
- [33] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng và Lei Zhang. Ngoài bộ khử nhiễu gaussian: Học phần còn lại của CNN sâu để khử nhiễu hình ảnh. Giao dịch của IEEE về xử lý hình ảnh (TIP), 26(7): 3142-3155, 2017.
- [34] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, và Ming-Hsuan Yang. Mạng kim tự tháp laplacian sâu cho độ phân giải siêu nhanh và chính xác. Trong Kỷ yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 624-632, 2017.
- [35] Muhammad Haris, Gregory Shakhnarovich, và Norimichi Ukita. Mạng chiều ngược sâu cho độ phân giải siêu cao. Trong Kỷ yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 1664-1673, 2018.

- [36] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte. Flow-based kernel prior with application to blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10601–10610, 2021.
- [37] Kelvin CK Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, and Chen Change Loy. GLEAN: Generative latent bank for large-factor image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14245–14254, 2021.
- [38] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 44(11):7474–7489, 2021.
- [39] Zongsheng Yue, Qian Zhao, Jianwen Xie, Lei Zhang, Deyu Meng, and Kwan-Yee K. Wong. Blind image super-resolution with elaborate degradation modeling on noise and kernel. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2128–2138, 2022.
- [40] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1671–1681, 2019.
- [41] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3217–3226, 2020.
- [42] Jiahong Fu, Hong Wang, Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kxnet: A model-driven deep neural network for blind super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 235–253, 2022.
- [43] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3262–3271, 2018.
- [44] Chong Mou, Yanze Wu, Xintao Wang, Chao Dong, Jian Zhang, and Ying Shan. Metric learning based interactive modulation for real-world super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 723–740, 2022.
- [45] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479:47–59, 2022.
- [46] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [47] Mauricio Delbracio and Peyman Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *arXiv preprint arXiv:2303.11435*, 2023.
- [48] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Image restoration with mean-reverting stochastic differential equations. *arXiv preprint arXiv:2301.11699*, 2023.
- [49] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I²SB: Image-to-image schrodinger bridge. In *International Conference on Machine Learning (ICML)*, 2023.
- [50] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.
- [51] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2015.
- [52] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.
- [53] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10012–10022, 2021.
- [36] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool và Radu Timofte. Hạt nhân dựa trên dòng chảy trước với ứng dụng có độ phân giải siêu mù. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 10601-10610, 2021.
- [37] Kelvin CK Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu và Chen Change Loy. GLEAN: Ngân hàng tiềm ẩn tạo cho hình ảnh có độ phân giải siêu lớn. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 14245-14254, 2021.
- [38] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy và Ping Luo. Khai thác thế hệ sâu trước để phục hồi và xử lý hình ảnh linh hoạt. Giao dịch của IEEE về Phân tích Mẫu và Trí thông minh Máy (TPAMI), 44(11):7474-7489, 2021.
- [39] Zongsheng Yue, Qian Zhao, Jianwen Xie, Lei Zhang, Deyu Meng và Kwan-Yee K. Wong. Hình ảnh siêu phân giải mù với mô hình suy giảm phức tạp về nhiễu và hạt nhân. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 2128-2138, 2022.
- [40] Kai Zhang, Wangmeng Zuo và Lei Zhang. Độ phân giải siêu cao plug-and-play sâu cho các hạt nhân mờ tùy ý. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 1671-1681, 2019.
- [41] Kai Zhang, Luc Van Gool, và Radu Timofte. Mạng mở rộng sâu cho hình ảnh siêu phân giải. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 3217-3226, 2020.
- [42] Jiahong Fu, Hong Wang, Qi Xie, Qian Zhao, Deyu Meng và Zongben Xu. Kxnet: Mạng lưới thần kinh sâu được điều khiển bằng mô hình cho độ phân giải siêu mù. Trong Ký yếu của Hội nghị Châu Âu về Thị giác Máy tính (ECCV), trang 235-253, 2022.
- [43] Kai Zhang, Wangmeng Zuo và Lei Zhang. Học một mạng siêu phân giải tích chập duy nhất cho nhiều lần xuống cấp. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 3262-3271, 2018.
- [44] Chong Mou, Yanze Wu, Xintao Wang, Chao Dong, Jian Zhang và Ying Shan. Điều chế tương tác dựa trên học tập số liệu cho siêu phân giải trong thế giới thực. Trong Ký yếu của Hội nghị Châu Âu về Thị giác Máy tính (ECCV), trang 723-740, 2022.
- [45] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, và Yueting Chen. Srdiff: Độ phân giải siêu cao của một hình ảnh với các mô hình xác suất khuếch tán. Điện toán thần kinh, 479:47-59, 2022.
- [46] Bahjat Kawar, Michael Elad, Stefano Ermon và Jiaming Song. Mô hình phục hồi khuếch tán khử nhiễu. Trong Ký yếu về những tiến bộ trong hệ thống xử lý thông tin thần kinh (NeurIPS), 2022.
- [47] Mauricio Delbracio và Peyman Milanfar. Đảo ngược bằng phép lặp trực tiếp: Một giải pháp thay thế cho việc khử nhiễu khuếch tán để phục hồi hình ảnh. bản in trước arXiv arXiv:2303.11435, 2023.
- [48] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund và Thomas B Schön. Phục hồi hình ảnh với các phương trình vi phân ngẫu nhiên đảo ngược trung bình. bản in trước arXiv arXiv:2301.11699, 2023.
- [49] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, và Anima Anandkumar. I2 SB: Cầu schrodinger từ ảnh tới ảnh. Trong Hội nghị quốc tế về học máy (ICML), 2023.
- [50] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, và Li Fei-Fei. Imagenet: Cơ sở dữ liệu hình ảnh phân cấp quy mô lớn. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 248-255, 2009.
- [51] Diederik P. Kingma và Jimmy Ba. Adam: Một phương pháp tối ưu hóa ngẫu nhiên. Trong Ký yếu của Hội nghị quốc tế về đại diện học tập (ICLR), 2015.
- [52] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, và những người khác. Pytorch: Một thư viện deep learning có phong cách bắt buộc, hiệu suất cao . Trong Ký yếu về những tiến bộ trong hệ thống xử lý thông tin thần kinh (NeurIPS), tập 32, 2019.
- [53] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin và Baining Guo. Máy biến áp Swin : Máy biến áp tầm nhìn phân cấp sử dụng các cửa sổ được dịch chuyển. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), trang 10012-10022, 2021.

- [54] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.
- [55] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International Conference on Curves and Surfaces*, pages 711–730. Springer, 2012.
- [56] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015.
- [57] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3086–3095, 2019.
- [58] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, volume 2, pages 416–423, 2001.
- [59] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76: 21811–21838, 2017.
- [60] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017.
- [61] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing (TIP)*, 27(9):4608–4622, 2018.
- [62] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision Workshops (ECCV-W)*, 2018.
- [63] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (CVPR-W)*, pages 466–467, 2020.
- [64] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004.
- [65] 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.
- [66] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5148–5157, 2021.
- [67] 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*, 2023.
- [68] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, pages 8748–8763, 2021.
- [69] Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, and Aran Komatsuzaki. Laion-400m: Open dataset of clip-filtered 400 million image-text pairs. In *Proceedings of Advances in Neural Information Processing Systems Workshops (NeurIPS-W)*, 2021.
- [70] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [54] Marco Bevilacqua, Aline Roumy, Christine Guillemot và Marie Line Alberi-Morel. Độ phức tạp thấp của phân giải hình ảnh đơn dựa trên việc nhúng hàng xóm không âm. 2012.
- [55] Roman Zeyde, Michael Elad và Matan Protter. Mở rộng quy mô hình ảnh bằng cách sử dụng các biểu diễn thưa thoát. Trong Hội nghị quốc tế về đường cong và bề mặt, trang 711–730. Mùa xuân, 2012.
- [56] Jia-Bin Huang, Abhishek Singh, và Narendra Ahuja. Độ phân giải siêu cao của một hình ảnh từ các mẫu tự chuyển đổi. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 5197–5206, 2015.
- [57] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, và Lei Zhang. Hướng tới độ phân giải siêu cao cho một hình ảnh trong thế giới thực: Chuẩn mực mới và mô hình mới. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), trang 3086–3095, 2019.
- [58] David Martin, Charless Fowlkes, Doron Tal, và Jitendra Malik. Cơ sở dữ liệu về các hình ảnh tự nhiên được phân đoạn của con người và ứng dụng của nó để đánh giá các thuật toán phân đoạn và do lường số liệu thống kê sinh thái. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), tập 2, trang 416–423, 2001.
- [59] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki và Kiyoharu Aizawa. Truy xuất manga dựa trên bản phác thảo bằng cách sử dụng bộ dữ liệu manga109. Công cụ và ứng dụng đa phương tiện, 76: 21811–21838, 2017.
- [60] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, và Luc Van Gool. Ảnh chất lượng Dslr trên thiết bị di động có mạng tích hợp sâu. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), 2017.
- [61] Kai Zhang, Wangmeng Zuo và Lei Zhang. Ffdnet: Hướng tới giải pháp khử nhiễu hình ảnh dựa trên CNN nhanh chóng và linh hoạt. Giao dịch của IEEE về xử lý hình ảnh (TIP), 27(9):4608–4622, 2018.
- [62] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao và Chen Change Loy. Esrgan: Mạng đối thủ tạo ra siêu độ phân giải nâng cao. Trong Ký yếu của Hội nghị Châu Âu về Hội thảo Thị giác Máy tính (ECCV-W), 2018.
- [63] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, và Feiyue Huang. Độ phân giải siêu cao trong thế giới thực thông qua ước tính hạt nhân và chèn nhiễu. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Hội thảo Thị giác Máy tính (CVPR-W), trang 466–467, 2020.
- [64] Chu Vương, AC Bovik, HR Sheikh và EP Simoncelli. Đánh giá chất lượng hình ảnh: từ khả năng hiển thị lỗi đến sự tương đồng về cấu trúc. Giao dịch của IEEE về Xử lý Hình ảnh (TIP), 13(4):600–612, 2004.
- [65] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, và Oliver Wang. Hiệu quả phi lý của các tính năng sâu sắc như một thước đo nhận thức. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), trang 586–595, 2018.
- [66] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar và Feng Yang. Musiq: Máy biến áp chất lượng hình ảnh đa tỷ lệ. Trong Ký yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính (ICCV), trang 5148–5157, 2021.
- [67] Jianyi Wang, Kelvin CK Chan, và Chen Change Loy. Khám phá clip để đánh giá giao diện của hình ảnh. Trong Ký yếu của Hội nghị AAAI về Trí tuệ nhân tạo, 2023.
- [68] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, và những người khác. Học các mô hình trực quan có thể chuyển đổi từ giám sát ngôn ngữ tự nhiên. Trong Hội nghị Quốc tế về học máy (ICML), trang 8748–8763, 2021.
- [69] Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev và Aran Komatsuzaki. Laion-400m: Tập dữ liệu mở gồm 400 triệu cặp văn bản-hình ảnh được lọc bằng clip. Trong Ký yếu hội thảo về những tiến bộ trong hệ thống xử lý thông tin thần kinh (NeurIPS-W), 2021.
- [70] Yochai Blau và Tomer Michaeli. Sự đánh đổi nhận thức-sự bóp méo. Trong Ký yếu của Hội nghị IEEE/CVF về Thị giác máy tính và Nhận dạng mẫu (CVPR), tháng 6 năm 2018.