

# 扩散模型、图像超分辨率以及一切：综述

Brian B. Moser<sup>1</sup>, Arundhati S. Shanbhag<sup>2</sup>, Federico Raue<sup>3</sup>, Stanislav Frolov<sup>4</sup>,  
Sebastian Palacio<sup>5</sup> 和 Andreas Dengel<sup>6</sup>

**Abstract**— 扩散模型 (DM) 颠覆了图像超分辨率 (SR) 领域，进一步缩小了图像质量与人类感知偏好之间的差距。它们易于训练，可以生成非常高质量的样本，其真实性超过了以前的生成方法所生成的样本。尽管它们取得了令人鼓舞的结果，但它们也带来了需要进一步研究的新挑战：高计算要求、可比性、缺乏可解释性、颜色偏移等。不幸的是，由于出版物过多，进入该领域的人太多了。为了解决这个问题，我们对应用于图像 SR 的 DM 背后的理论基础进行了统一的叙述，并提供了详细的分析，强调了该领域的独特特征和方法，不同于该领域现有的更广泛的评论。本文阐明了对 DM 原理的连贯理解，并探讨了当前的研究途径，包括替代输入域、条件技术、引导机制、损坏空间和零样本学习方法。通过从 DM 的视角详细研究图像 SR 的演变和当前趋势，本文阐明了现有的挑战并规划了未来的潜在方向，旨在激发这一快速发展的领域的进一步创新。

**Index Terms**— 扩散模型 (DM)、超分辨率 (SR)、调查。

## 一、引言

在不断发展的计算机视觉领域，图像超分辨率 (SR) 任务（将低分辨率 (LR) 图像增强为高分辨率 (HR) 图像）由于其不适定性而长期存在挑战。对于任何给定的 LR 图像，可能存在多张 HR 图像，这些图像在亮度和颜色等方面有所不同 [1]。其应用范围广泛，从日常摄影到优化卫星 [2] 和医学图像 [3]。尽管之前的生成式 SR 模型取得了显著成就，但每种模型都有自己的局限性。例如，自回归模型 (ARM) 的计算需求往往超过其效用，而正则化流 (NF) 或变分自编码器 (VAE) 难以满足质量期望 [4]、[5]、[6]。尽管功能强大，但生成对抗网络 (GAN)

需要仔细的正则化和优化策略来克服不稳定问题 [7]。

扩散模型 (DM) 的出现标志着图像生成任务（包括 SR）的重大转变，挑战了 GAN 的长期主导地位 [8]、[9]、[10]、[11]。Dall-E 和稳定扩散等应用表明，DM 已在各个方面超越了 GAN [12]、[13]、[14]。它们能够从 LR 输入生成高质量图像，与人类评估者的定性判断高度一致，在 SR 中显示出巨大的前景 [15]。换句话说，人类评分者认为 DM 生成的 SR 图像比 GAN 等其他生成模型生成的图像更逼真。

然而，随着出版物数量的增加，及时了解最新发展变得越来越困难，尤其是对于刚进入该领域的人来说。DM 与之前的生成模型有着根本的不同，在解决早期模型的局限性的同时，也带来了新的挑战。尽管发展迅速，但确定连贯的趋势和潜在的研究方向仍然具有挑战性。本文旨在揭开 DM 的神秘面纱，提供全面的概述，将基础概念与图像 SR 的前沿联系起来，并批判性地分析当前的优势和劣势。

本文以先前的论文 *Hitchhiker's Guide to SR* [16] 为基础，该论文对图像 SR 领域进行了概述。Li 等人的综述 [17] 与其类似，该综述回顾了更一般的图像修复任务（如修复和去雾）中的 DM。两者都有重叠的主题，例如 DM 的基础和类型，即去噪扩散概率模型 (DDPM) [8]、基于分数的生成模型 (SGM) [11] 和随机微分方程 (SDE) [10]。此外，这两项综述都强调了条件策略和零样本扩散的引入，并展示了潜在的研究方向。但是，本文涵盖了与图像 SR 相关的所有主题，因此更详细地介绍了专门为图像 SR 开发的最新进展。此外，我们还解释了与 SR 相关的挑战，例如色移和级联图像 SR。我们还重点介绍了 DM 与其他生成式 SR 模型（即 VAE、GAN 和基于流的方法）的关系。此外，我们还回顾了基于频率的 DM、替代损坏空间和基于扩散的图像 SR 应用。

最后，本文讨论了新兴趋势及其重塑 SR 和 DM 发展的潜力，为未来的研究奠定了基础。通过在快速发展的 DM 领域提供清晰度和方向，我们的目标是

Received 6 February 2024; revised 23 June 2024 and 12 August 2024; accepted 1 October 2024. This work was supported in part by the BMBF Project XAINES under Grant 01IW20005, in part by SustainML (Horizon Europe) under Grant 101070408, and in part by the BMBF Project Albatross under Grant 01IW24002. (Corresponding author: Brian B. Moser.)

Brian B. Moser, Arundhati S. Shanbhag, Stanislav Frolov, and Andreas Dengel are with the German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany, and also with Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, 67663 Kaiserslautern, Germany (e-mail: Brian.Moser@dfki.de).

Federico Raue and Sebastian Palacio are with the German Research Center for Artificial Intelligence (DFKI), 67663 Kaiserslautern, Germany.

Digital Object Identifier 10.1109/TNNLS.2024.3476671

启发和指导下一波研究，促进不断突破 DM 在图像 SR 方面可能性界限的进步。

本文的结构如下。

*Section II—SR Basics:* 本节提供基本定义并介绍图像 SR 出版物中常用的用于评估图像质量的标准数据集、方法和指标。

*Section III—DMs Basics:* 本节介绍 DM 的原理和各种公式，包括 DDPM、SGM 和 SDE。本节还探讨了 DM 与其他生成模型的关系。

*Section IV—Improvements for DMs:* 增强 DM 的常见做法，重点关注高效的采样技术和改进的可能性估计。

*Section V—DMs for Image SR:* 本节介绍了 SR 中 DM 的具体实现，探索了替代领域（潜在空间和小波域），讨论了架构设计和具有零空间模型的多项任务，并研究了替代损坏空间。

*Section VII—Domain-Specific Applications:* 基于 DM 的 SR 应用，即医学成像、盲人面部恢复 (BFR)、面部 SR 中的大气湍流 (AT) 和遥感。

*Section VIII—Discussion and Future Work:* 图像 SR 的 DM 的常见问题以及特定于图像 SR 的 DM 值得注意的研究途径。

*Section IX—Conclusion:* 本节对本文进行总结。

## 图像 SR

图像 SR 的目标是将一张或多张 LR 图像转换为 HR 图像。该领域大致可分为两个领域 [16]：单图像超分辨率 (SISR) 和多图像超分辨率 (MISR)。在 SISR 中，一张 LR 图像产生一张 HR 图像。相比之下，MISR 方法使用多张 LR 图像来生成一个或多个 HR 输出。本节重点介绍 SISR，并探讨相关数据集、已建立的 SR 模型以及评估图像质量的技术。

给定一个 LR 图像  $x \in \mathbb{R}^{\bar{w} \times \bar{h} \times c}$ ，目标是用  $\bar{w} < w$  和  $\bar{h} < h$  生成一个 HR 对应图像  $y \in \mathbb{R}^{w \times h \times c}$ 。该关系由退化映射表示

$$x = \mathcal{D}(y; \Theta) = ((y \otimes k) \downarrow_s + n)_{\text{JPEG}_q} \quad (1)$$

其中  $\mathcal{D}$  是退化图  $\mathcal{D} : \mathbb{R}^{w \times h \times c} \rightarrow \mathbb{R}^{\bar{w} \times \bar{h} \times c}$ ， $\Theta$  包含退化参数，包括模糊  $k$ 、噪声  $n$ 、缩放  $s$  和压缩质量  $q$  [18] 等方面。退化通常是未知的，这在确定  $\mathcal{D}$  与参数  $\theta$  的逆映射时构成了主要挑战，通常体现为 SR 模型 [19]、[20]。这导致了一项优化任务，旨在最小化预测的 SR 图像  $\hat{y}$  与原始 HR 图像  $y$  之间的差异

$$\theta^* = \operatorname{argmin}_{\theta} \mathcal{L}(\hat{y}, y) + \lambda \phi(\theta) \quad (2)$$

其中  $\mathcal{L}$  表示预测的 SR 图像与实际 HR 图像之间的损失。这里， $\lambda$  是一个平衡参数，而  $\phi(\theta)$  则作为正则化项引入。

固有的复杂性源于预测  $\theta$  的不适定性，因为对于任何给定的 LR 图像，可能有多个 SR 图像有效，即，它们与真实图像相比可能具有相似的损失值，但由于亮度和颜色等许多方面，主观感知会有所不同 [1]、[15]、[21]。传统的回归技术，如标准卷积神经网络 (CNN)，通常适用于较低的放大倍数，但难以复制较高放大倍数所需的高频细节（例如  $s > 4$ ）。为了解决这个问题，SR 模型必须幻化出插值以外的真实细节，这通常属于生成模型的总体主题，而 DM 现在处于最前沿。

## A. Datasets

一些数据集提供了各种图像、分辨率和内容类型。通常，这些数据集由 LR 和 HR 图像对组成。但是，有些数据集仅包含 HR 图像，其中 LR 图像是通过带抗锯齿的双三次下采样创建的 - 这是 MATLAB 中 *imresize* 的默认设置 [22]。一个著名的通用 SR 训练集是多样化 2K 分辨率 (DIV2K) 数据集 [23]，其中包括专为图像 SR 设计的不同分辨率的各种逼真图像。在 DIV2K 上训练的 SR 模型的经典测试数据集是 Set5 [24]、Set14 [25]、BSDS100 [26]、Urban100 [27] 和 Manga109 [28]，涵盖了各种场景和图像内容，例如建筑物和漫画画。Flickr2K [23] 和 Flickr-Faces-HQ (FFHQ) [29] 分别提供了来自 Flickr 的以人为中心和以场景为中心的各种图像集。虽然 FFHQ 通常用于训练面部 SR 任务的模型，但 Flickr2K 通常与 DIV2K 结合用作训练数据扩展。面部 SR 的另一个数据集是 CelebA-HQ [30]，它提供高质量的名人图像，通常用于评估 FFHQ 训练的 SR 模型。对于 CV 中的更广泛应用，ImageNet [31] 和 Visual Object Classes (VOC2012) [32] 等数据集受到青睐。ImageNet 提供了广泛的图像，可帮助在各种对象类别上训练模型，而 VOC2012 对于对象检测和分割至关重要。两者对于涉及 SR 的多任务学习都很有价值。可以在 *Hitchhiker's Guide to SR* [16] 中找到更多数据集。

## B. SR Models

主要目标是设计一个 SR 模型  $\mathcal{M} : \mathbb{R}^{\bar{w} \times \bar{h} \times c} \rightarrow \mathbb{R}^{w \times h \times c}$ ，使其逆 (1)

$$\hat{y} = \mathcal{M}(x; \theta) \quad (3)$$

其中  $\hat{y}$  是 LR 图像的预测 HR 近似值， $x$  和  $\theta$  是  $\mathcal{M}$  的参数。参数  $\theta$  使用 (2) 进行优化，即最小化估计值  $\hat{y}$  和真实 HR 图像  $y$  之间的损失函数  $\mathcal{L}$ 。在我们详细研究 DM 如何履行这一职责之前，II-B1-II-B4 部分重点介绍了设计 SR 模型的标准方法，尤其是深度学习方法。

1) *Traditional Methods*: 传统的图像 SR 方法定义了一系列方法, 例如统计 [33]、基于边缘的 [34]、[35]、基于块的 [36]、[37]、基于预测的 [38]、[39] 和稀疏表示技术 [40]。它们从根本上依赖于图像统计数据和现有像素中固有的信息来生成 HR 图像。尽管这些方法很实用, 但它们的一个值得注意的缺点是可能会引入噪声、模糊和视觉伪影 [16]。

2) *Regression-Based Deep Learning*: 随着深度学习和计算能力的进步, 图像 SR 得到了显著的发展。通常, 它们使用 CNN 进行从 LR 到 HR 的端到端映射。初始模型 (例如 SRCNN [41]、FSRCNN [42] 和 ESPCNN [43]) 使用了具有不同深度和特征图大小的简单 CNN。后来的模型将来自更广泛 CV 领域的概念改编成 SR 模型, 例如 ResNet 导致了 SRResNet, 其中残差信息被传播到连续的网络层 [44]。同样, DenseNet [45] 改编为 SRDenseNet [46]。它们使用密集块, 其中每一层都会额外接收所有先前层中生成的特征。递归使用相同模块来学习表示的递归 CNN 也受到了其他 CV 方法的启发, 用于 DRCN [47]、DRRN [48] 和 CARN [49] 中的基于回归的 SR 方法。最近, 注意力机制已被纳入其中, 用于关注图像中的感兴趣区域, 主要是通过通道和空间注意力机制 [16]、[50]、[51]、[52]。所有这些方法的共同点都是基于回归。常用的损失函数是 L1 和 L2 损失。如上所述, 它们通常在较低放大倍数下产生令人满意的结果, 但难以复制更高放大倍数下所需的高频细节 (例如  $s > 4$ )。出现这些限制是因为这些模型主要学习从 LR 到 HR 图像的平均映射 (由于 L1 和 L2 损失), 这往往会产生过于光滑而缺乏细节的纹理, 在较大的上采样因子中尤其明显 [16]。为了解决这个问题, SR 模型必须超越简单的插值来幻化真实的细节, 这是通常由生成模型解决的挑战。

3) *Generative Adversarial Networks*: 最突出的生成模型之一是 GAN。它使用两个 CNN: 一个生成器 G 和一个鉴别器 D, 它们同时进行训练。生成器旨在生成尽可能接近原始样本的 HR 样本, 以欺骗鉴别器, 鉴别器会尝试区分生成的样本和真实样本。该框架 (例如在 SRGAN [44] 或 ESRGAN [53] 中) 使用对抗性损失和内容损失的组合进行优化, 以生成不太平滑的图像。最先进的 GAN 生成的图像更清晰、更详细。由于它们能够生成高质量和多样化的图像, 因此最近受到了广泛关注。然而, 它们容易受到模式崩溃的影响, 占用大量计算资源, 有时无法收敛, 并且存在稳定性问题 [7]。

4) *Flow-Based Methods*: 基于流的方法采用光流算法来生成 SR 图像 [54]。引入这些算法是为了通过学习给定 LR 输入的合理 HR 图像的条件分布来对抗图像 SR 的不适定性。它们引入了一个条件

归一化流架构通过计算 LR 和 HR 图像之间的位移场来对齐它们, 然后使用此信息恢复 SR 图像。它们采用完全可逆编码器, 能够将任何输入 HR 图像映射到潜在流空间并确保精确重建。该框架使 SR 模型能够使用基于精确对数似然的训练来学习丰富的分布 [54]。这有助于基于流的方法规避训练不稳定性, 但会产生大量计算成本。

### C. Image Quality Assessment

图像质量是一个多方面的概念, 涉及清晰度、对比度和无噪声等属性。因此, 基于生成的图像质量对 SR 模型进行公平评估并非易事。本节介绍了在图像 SR 背景下评估图像质量的基本方法 (尤其是对于 DM 而言), 这属于图像质量评估 (IQA) 的总称。<sup>1</sup> 本质上, IQA 是指任何类似于人类观察者的感知评估的指标, 具体而言, 是指应用 SR 技术后在图像中感知到的真实度。在本节中, 我们将使用以下符号:  $N_x = w \cdot h \cdot c$ , 它定义图像  $x$  的像素数;  $\in \mathbb{R}^{w \times h \times c}$  和  $\Omega_x = \{(i, j, k) \in \mathbb{N}_1^3 | i \leq h, j \leq w, k \leq c\}$ , 它定义  $x$  中所有有效位置的集合。

1) *Peak Signal-to-Noise Ratio*: 峰值信噪比 (PSNR) 是评估 SISR 重建质量最广泛使用的技术之一。它表示最大像素值  $L$  与 SR 图像  $\hat{y}$  和 HR 图像  $y$  之间的均方误差 (MSE) 之比

$$\text{PSNR}(y, \hat{y}) = 10 \cdot \log_{10} \left( \frac{L^2}{\frac{1}{N} \sum_{i=1}^N [y - \hat{y}]^2} \right). \quad (4)$$

尽管它是最流行的 IQA 方法之一, 但它并不能准确匹配人类感知 [15]。它关注的是像素差异, 而这些差异通常与主观感知的质量不一致: 像素的细微变化都可能导致更差的 PSNR 值, 但不会影响人类感知质量。由于其像素级计算, 使用相关像素损失训练的模型往往会实现较高的 PSNR 值 [16], 而生成模型往往会产生较低的 PSNR 值 [15]。

2) *SSIM Index*: 结构相似性 (SSIM) 与 PSNR 一样, 是一种流行的评估方法, 侧重于图像之间结构特征的差异。它通过比较亮度、对比度和结构来独立捕获 SSIM。SSIM 将图像  $y$  的亮度  $\mu_y$  估计为强度的平均值, 同时将对比度  $\sigma_y$  估计为其标准差

$$\mu_y = \frac{1}{N_y} \sum_{p \in \Omega_y} y_p \quad (5)$$

$$\sigma_y = \frac{1}{N_y - 1} \sum_{p \in \Omega_y} [y_p - \mu_y]^2. \quad (6)$$

<sup>1</sup>更多与SR相关的IQA方法可参见[16]。

为了捕捉计算实体之间的相似性，作者引入了一个比较函数  $S$

$$S(x, y, c) = \frac{2 \cdot x \cdot y + c}{x^2 + y^2 + c} \quad (7)$$

其中  $x$  和  $y$  是被比较的标量变量，而  $c = (k \cdot L)^2$ ,  $0 < k \ll 1$  是数值稳定性的常数。对于 HR 图像  $y$  及其近似值  $\hat{y}$ ，亮度 ( $C_l$ ) 和对比度 ( $C_c$ ) 比较使用  $C_l(y, \hat{y}) = S(\mu_y, \mu_{\hat{y}}, c_1)$  和  $C_c(y, \hat{y}) = S(\sigma_y, \sigma_{\hat{y}}, c_2)$  计算，其中  $c_1, c_2 > 0$ 。经验协方差

$$\sigma_{y, \hat{y}} = \frac{1}{N_y - 1} \sum_{p \in \Omega_y} (y_p - \mu_y) \cdot (\hat{y}_p - \mu_{\hat{y}}) \quad (8)$$

定义结构比较 ( $C_s$ )，即  $y$  与  $\hat{y}$  之间的相关系数

$$C_s(y, \hat{y}) = \frac{\sigma_{y, \hat{y}} + c_3}{\sigma_y \cdot \sigma_{\hat{y}} + c_3} \quad (9)$$

其中  $c_3 > 0$ 。最后，SSIM 定义为

$$\text{SSIM}(y, \hat{y}) = [C_l(y, \hat{y})]^\alpha \cdot [C_c(y, \hat{y})]^\beta \cdot [C_s(y, \hat{y})]^\gamma \quad (10)$$

其中  $\alpha > 0$ ,  $\beta > 0$  和  $\gamma > 0$  是可以调整的参数，可以调整组件的相对重要性。

3) *Mean Opinion Score*: 平均意见得分 (MOS) 是一种主观测量方法，利用人类感知质量来评估生成的 SR 图像。向人类观看者展示 SR 图像，并要求他们用质量得分对其进行评分，然后将这些得分映射到数值并取平均值。通常，这些分数的范围从 1 (差) 到 5 (好)，但可能会有所不同 [15]。虽然这种方法是对人类感知的直接评估，但与客观指标相比，它更耗时且更繁琐。此外，由于该指标的高度主观性，它很容易受到偏见的影响。

4) *Consistency*: 一致性衡量非确定性 SR 方法 (例如 GA N 或 DM 等生成模型) 的稳定性程度。与基于流的方法一样，生成方法也是有意设计为针对相同输入生成一系列合理输出。但是，一致性低下是不可取的。微小的变化会减轻相对一致的方法对输入的影响。然而，一致性可能会根据要求而变化。用于量化一致性的一个常用指标是 MSE

5) *Learned Perceptual Image Patch Similarity*: 与基于像素的 PSNR 和 SSIM 评估相反，学习感知图像块相似度 (LPIPS) 利用预训练的 CNN  $\varphi$ ，例如 VGG [55] 或 AlexNet [56]，并从 SR 和 HR 图像生成  $L$  特征图，然后计算它们之间的相似度。给定  $h_l$  和  $w_l$  分别作为第  $l$  个特征图的高度和宽度，以及缩放向量  $\alpha_l \in \mathbb{R}^{C_l}$ ，LPIPS 指标公式如下：

$$\text{LPIPS}(y, \hat{y}) = \sum_{l=1}^L \sum_p \frac{\|\alpha_l \odot (\varphi^l(\hat{y}) - \varphi^l(y))_p\|_2^2}{h_l \cdot w_l} \quad (11)$$

LPIPS 的运行方式是，通过  $\varphi$  将图像投影到感知特征空间，并评估 SR 和 HR 图像中相应块之间的差异 (按  $\alpha_l$  缩放)。

与 PSNR 和 SSIM [16] 等传统指标相比，该方法更符合人类感知，因此可以进行更加以人为本的评估。

6) *No-Reference Metrics*: 到目前为止讨论的所有 IQA 指标都需要参考 (真实) 图像。但是，在某些情况下没有可用的参考图像，例如在无监督设置中。幸运的是，我们可以通过测量统计特征与从相似领域的高质量图像集合 (即自然图像) 中获得的统计特征的距离来评估图像。这可以是像 BRISQUE [57] 那样具有观点和失真感知能力的，也可以是像 NIQE [58] 那样不具有观点和失真感知能力的。另一种评估无参考图像质量的有趣方法是利用视觉语言预训练的 CLIP 模型 [59]。一个例子是 CLIP-IQA，它使用两个含义相反的提示 (即 “好摄影” 和 “坏摄影” [60]) 计算编码图像的余弦相似度。由此得出的一个或另一个提示的相对相似度度量决定了图像质量。CLIP-IQA 的结果与不使用手工特征的 BRISQUE 的结果相当，并且超越了 NIQE 等其他无参考 IQA 方法。利用深度学习模型的另一方法是使用 TID2013 [61] 等 IQA 数据集训练它们预测主观分数。示例包括 DeepQA [62]、NIMA [63] 或 MUSIQ [64]。其他内容可以在 *Hitchhiker's Guide to SR* [16] 的基于学习的感知质量部分中找到。

### III. DMS 基础知识

DM 深刻影响了生成式 AI 领域，许多属于 DM 这一总称的方法应运而生。DM 与早期生成式模型的不同之处在于，DM 的执行时间是迭代时间步长的，既向前又向后，用  $t$  表示，如图 1 所示。正向和反向扩散过程的区别如下

- 1) *Forward*  $q$ : 使用噪声迭代地降低输入数据的质量，随时间向前变化 (即， $t$  增加)。
- 2) *Backward*  $p$ : 对退化的数据进行去噪，从而随时间向后迭代地逆转噪声 (即， $t$  减少)。

在正向扩散过程中，时间步长  $t$  会增加，而在反向扩散过程中，它会向 0 传播。让  $\mathcal{D} = \{x_i, y_i\}_{i=1}^N$  成为 LR-HR 图像对的数据集。对于每个时间步长  $t$ ，随机变量  $z_t$  描述当前状态，即图像和损坏空间之间的状态。在文献中，正向扩散中的  $z_t$  和反向扩散中的  $z_t$  没有明显区别。在正向扩散期间，我们假设  $z_t \sim q(z_t | z_{t-1})$ 。相反，在反向扩散中，我们假设  $z_{t-1} \sim p(z_{t-1} | z_t)$ 。我们将用  $0 < t \leq T$  表示  $T$ ，作为有限情况下的最大时间步长。初始数据分布 ( $t = 0$ ) 用  $z_0 \sim q(x)$  表示，然后缓慢注入噪声 (加法)。反之亦然，DM 通过在反向时间方向上运行参数化模型  $p_\theta(z_{t-1} | z_t)$  来消除其中的噪声，该模型近似于理想 (但无法实现) 的去噪分布  $p(z_{t-1} | z_t)$ 。

前向扩散  $q$  和后向扩散  $p$  的显式实现，由  $p_\theta$  近似，定义为

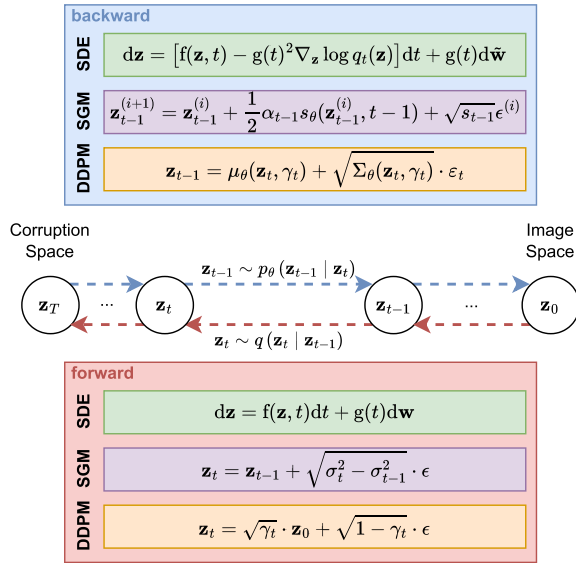


图 1. DM 的原理。前向扩散迭代地添加噪声（红色），将图像从图像空间转换到损坏空间。后向扩散，即迭代细化过程，将过程（蓝色）恢复回图像空间。图中显示了 DM 的三种不同实现，即 DDPM、SGM 和 SDE，以及它们各自的前向和后向扩散公式。

具体来说，DM 有三种类型：DDPM 和 SGM 两种离散形式，以及 SDE 的连续形式，如图 1 所示，我们将在下文讨论。

#### A. Denoising Diffusion Probabilistic Models

DDPM [8] 使用两个马尔可夫链在有限数量的离散时间步骤中实现前向和后向扩散。

1) *Forward Diffusion*: 它将数据分布转换为先验分布，通常是手动设计的（例如高斯分布），如下所示

$$q(z_t | z_{t-1}) = \mathcal{N}(z_t | \sqrt{1 - \alpha_t} z_{t-1}, \alpha_t \mathbf{I}) \quad (12)$$

其中超参数  $0 < \alpha_{1:T} < 1$  表示每个时间步骤中纳入的噪声方差。虽然通常采用高斯核，但也可以采用其他核类型。此公式可以简化为单步计算，如下所示

$$q(z_t | z_0) = \mathcal{N}(z_t | \sqrt{\gamma_t} z_0, (1 - \gamma_t) \mathbf{I}) \quad (13)$$

其中  $\gamma_t = \prod_{i=1}^t (1 - \alpha_i)$  [66]。因此， $z_t$  可以直接采样，而不管之前的时间步骤应该发生什么，方法是

$$z_t = \sqrt{\gamma_t} \cdot z_0 + \sqrt{1 - \gamma_t} \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (14)$$

2) *Backward Diffusion*: 目标是直接学习前向扩散的逆，并生成类似于先验  $z_0$ （通常是 SR 中的 HR 图像）的分布。在实践中，我们使用 CNN 来学习  $p$  的参数化形式。由于前向过程近似于  $q(z_T) \approx \mathcal{N}(0, \mathbf{I})$ ，可学习过渡核的公式变为

$$p_\theta(z_{t-1} | z_t) = \mathcal{N}(z_{t-1} | \mu_\theta(z_t, \gamma_t), \Sigma_\theta(z_t, \gamma_t)) \quad (15)$$

其中  $\mu_\theta$  和  $\Sigma_\theta$  是可学习的。类似地，以  $x$ （例如，LR 图像）为条件的条件公式  $p_\theta(z_{t-1} | z_t, x)$  则改用  $\mu_\theta(z_t, x, \gamma_t)$  和  $\Sigma_\theta(z_t, x, \gamma_t)$ 。

3) *Optimization*: 为了引导后向扩散学习前向过程，我们最小化前向和反向序列联合分布的 Kullback-Leibler (KL) 散度

$$p_\theta(z_0, \dots, z_T) = p(z_T) \prod_{t=1}^T p_\theta(z_{t-1} | z_t), \text{ and} \quad (16)$$

$$q(z_0, \dots, z_T) = q(z_0) \prod_{t=1}^T q(z_t | z_{t-1}) \quad (17)$$

这导致最小化

$$\begin{aligned} & \text{KL}(q(z_0, \dots, z_T) \| p_\theta(z_0, \dots, z_T)) \\ &= -\mathbb{E}_{q(z_0, \dots, z_T)} [\log p_\theta(z_0, \dots, z_T)] + c \\ &\stackrel{(i)}{=} \mathbb{E}_{q(z_0, \dots, z_T)} \left[ -\log p(z_T) - \sum_{t=1}^T \log \frac{p_\theta(z_{t-1} | z_t)}{q(z_t | z_{t-1})} \right] + c \\ &\stackrel{(ii)}{\geq} \mathbb{E}[-\log p_\theta(z_0)] + c \end{aligned} \quad (18)$$

其中 (i) 是可能的，因为两个项都是分布的乘积，而 (ii) 是 Jensen 不等式的乘积。常数  $c$  不受影响，因此与优化  $\theta$  无关。请注意，不带  $c$  的 (18) 是数据  $z_0$  对数似然的变分下限 (VLB)，通常由 DDPM 最大化。

#### B. Score-Based Generative Models

SGM 与 DDPM 非常相似，利用离散扩散过程，但采用另一种数学基础。Song 和 Ermon [10] 建议不直接使用概率密度函数  $p(z)$ ，而是使用其 (Stein) 得分函数，该函数定义为对数概率密度  $\nabla_z \log p(z)$  的梯度。从数学上讲，得分函数保留了有关密度函数的所有信息，但从计算上讲，它更易于使用。此外，将模型训练与采样过程分离，可以更灵活地定义采样方法和训练目标。

1) *Forward Diffusion*: 令  $0 < \sigma_1 < \dots < \sigma_T$  为噪声水平的有限序列。与 DDPM 一样，前向扩散通常分配给高斯噪声分布，即

$$q(z_t | z_0) = \mathcal{N}(z_t | z_0, \sigma_t^2 \mathbf{I}). \quad (19)$$

该方程产生一系列噪声数据密度  $q(z_1), \dots, q(z_T)$  和  $q(z_t) = \int q(z_t) q(z_0) dz_0$ 。因此，中间步骤  $z_t = z_0 + \sigma_t \cdot \epsilon$  和  $\epsilon \sim \mathcal{N}(0, \mathbf{I})$  可以在单个步骤中从先前的时间步骤中采样。

2) *Backward Diffusion*: 为了在反向扩散过程中恢复噪声，我们需要近似  $\nabla_{z_t} \log q(z_t)$ ，并选择一种方法从该近似值中估计中间状态  $z_{t_0}$ 。对于每个时间步骤  $t$  的梯度近似，我们使用一个经过训练的预测器，表示为  $s_\theta$ ，称为噪声条件得分网络 (NCSN)，这样  $s_\theta(z_t, t) \approx \nabla_{z_t} \log q(z_t)$  [11]。

NCSN 的训练将在下一节中介绍；现在，我们重点介绍使用 NCSN 的采样过程。使用 NCSN 进行采样涉及通过迭代方法使用  $s_\theta(\mathbf{z}_t, t)$  生成中间状态  $\mathbf{z}_{t_0}$ 。请注意，此迭代过程不同于扩散过程中进行的迭代，因为它仅解决  $\mathbf{z}_t$  的生成。这是与 DDPM 的一个关键区别，因为  $\mathbf{z}_t$  需要迭代采样，而 DDPM 直接从  $\mathbf{z}_{t+1}$  预测  $\mathbf{z}_{t_0}$ 。

有多种方法可以执行这种迭代生成，但我们将集中讨论一种称为退火朗之万动力学 (ALD) 的特定方法，该方法由 Song 和 Ermon [10] 提出。令  $N$  为时间步长  $t$  处  $\mathbf{z}_t$  的估计迭代次数， $\alpha_t > 0$  为相应的步长，它决定了估计从一个估计  $\mathbf{z}_{t-1}^{(i)}$  向  $\mathbf{z}_{t-1}^{(i+1)}$  移动的幅度。初始状态为  $\mathbf{z}_T^{(N)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 。对于每个  $0 < t \leq T$ ，我们初始化  $\mathbf{z}_{t-1}^{(0)} = \mathbf{z}_t^{(N)} \approx \mathbf{z}_t$ ，它是前一个中间状态的最新估计。为了迭代地获得  $\mathbf{z}_{t-1}^{(N)} \approx \mathbf{z}_{t-1}$ ，ALD 对  $i = 0, \dots, N-1$  使用以下更新规则：

$$\epsilon^{(i)} \leftarrow \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (20)$$

$$\mathbf{z}_{t-1}^{(i+1)} \leftarrow \mathbf{z}_{t-1}^{(i)} + \frac{1}{2}\alpha_{t-1}s_\theta(\mathbf{z}_{t-1}^{(i)}, t-1) + \sqrt{s_{t-1}}\epsilon^{(i)}. \quad (21)$$

该更新规则保证当  $\alpha_t \rightarrow 0$  和  $N \rightarrow \infty$  时， $\mathbf{z}_0^{(N)}$  收敛到  $q(\mathbf{z}_0)$  [67]。

与 DDPM 类似，我们可以通过将条件  $x$  (例如 LR 图像) 集成到  $s_\theta(\mathbf{z}_t, x, t) \approx \nabla_{\mathbf{z}_t} \log q(\mathbf{z}_t|x)$  中，将 SGM 转变为条件 SGM。

3) *Optimization*: 无需专门制定反向扩散公式，我们就可以训练一个 NCSN，使得  $s_\theta(\mathbf{z}_t, t) \approx \nabla_{\mathbf{z}_t} \log q(\mathbf{z}_t)$ 。可以使用去噪得分匹配方法估算得分 [68]

$$\begin{aligned} & \mathbb{E}_{\substack{t \sim \mathcal{U}(1,T) \\ \mathbf{z}_0 \sim q(\mathbf{z}_0) \\ \mathbf{z}_t \sim q(\mathbf{z}_t|\mathbf{z}_0)}} [\lambda(t)\sigma_t^2 \|\nabla_{\mathbf{z}_t} \log q(\mathbf{z}_t) - s_\theta(\mathbf{z}_t, t)\|^2] \\ & \stackrel{(i)}{=} \mathbb{E}_{\substack{t \sim \mathcal{U}(1,T) \\ \mathbf{z}_0 \sim q(\mathbf{z}_0) \\ \mathbf{z}_t \sim q(\mathbf{z}_t|\mathbf{z}_0)}} [\lambda(t)\sigma_t^2 \|\nabla_{\mathbf{z}_t} \log q(\mathbf{z}_t|\mathbf{z}_0) - s_\theta(\mathbf{z}_t, t)\|^2] + c \\ & \stackrel{(ii)}{=} \mathbb{E}_{\substack{t \sim \mathcal{U}(1,T) \\ \mathbf{z}_0 \sim q(\mathbf{z}_0) \\ \mathbf{z}_t \sim q(\mathbf{z}_t|\mathbf{z}_0)}} \left[ \lambda(t) \left\| -\frac{\mathbf{z}_t - \mathbf{z}_0}{\sigma_t} - \sigma_t s_\theta(\mathbf{z}_t, t) \right\|^2 \right] + c \\ & \stackrel{(iii)}{=} \mathbb{E}_{\substack{t \sim \mathcal{U}(1,T) \\ \mathbf{z}_0 \sim q(\mathbf{z}_0) \\ \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})}} [\lambda(t) \|\epsilon + \sigma_t s_\theta(\mathbf{z}_t, t)\|^2] + c \end{aligned} \quad (22)$$

其中  $\lambda(t) > 0$  是权重函数， $\sigma_t$  是在时间步  $t$  处添加的噪声水平，(i) 由 Vincent [68] 推导而来，(ii) 来自 (19)，(iii) 来自  $\mathbf{z}_t = \mathbf{z}_0 + \sigma_t \epsilon$ ，并且  $c$  再次是常数，在  $\theta$  的优化中不受影响。请注意，还有其他方法可以估计得分，例如基于得分匹配 [69] 或切片得分匹配 [70]。

### C. Stochastic Differential Equations

到目前为止，我们已经讨论了处理有限时间步长的 DM。通过将这些 DM 公式化为 SDE (也称为 Score SDE) 的解，可以将其推广到无限连续时间步长 [11]。事实上，我们可以将 SGM 和 DDPM 视为连续时间 SDE 的离散化。SDE 不是

完全与 DM 绑定，因为它们是描述随机过程的数学概念。因此，它们非常适合描述我们想要在 DM 中模拟的过程。与之前一样，数据在一般扩散过程中受到扰动，但可以推广到无限数量的噪声尺度。

1) *Forward Diffusion*: 我们可以用以下 SDE 表示前向扩散

$$d\mathbf{z} = f(\mathbf{z}, t)dt + g(t)d\mathbf{w} \quad (23)$$

其中  $f$  和  $g$  分别是漂移和扩散函数， $\mathbf{w}$  是标准维纳过程 (也称为布朗运动)。此广义公式允许统一表示 DDPM 和 SGM。DDPM 的 SDE 定义为

$$d\mathbf{z} = -\frac{1}{2}\alpha(t)\mathbf{z}dt + \sqrt{\alpha(t)}d\mathbf{w} \quad (24)$$

其中  $\alpha((t/T)) = T\alpha_t$  为  $T \rightarrow \infty$ 。对于 SGM，SDE 为

$$d\mathbf{z} = \sqrt{\frac{d[\sigma(t)^2]}{dt}}d\mathbf{w} \quad (25)$$

其中  $\sigma((t/T)) = \sigma_t$  表示  $T \rightarrow \infty$ 。从现在开始，我们用  $q_t(\mathbf{z})$  表示扩散过程中  $\mathbf{z}_t$  的分布。

2) *Backward Diffusion*: Anderson [71] 提出了逆时间随机微分方程

$$d\mathbf{z} = [f(\mathbf{z}, t) - g(t)^2 \nabla_{\mathbf{z}} \log q_t(\mathbf{z})]dt + g(t)d\tilde{\mathbf{w}} \quad (26)$$

其中， $\tilde{\mathbf{w}}$  是时间向后流动时的标准维纳过程， $dt$  是无穷小的负时间步长。(26) 式的解可以看作是将噪声逐渐转换为数据的扩散过程。Song 等人 [11] 证明了存在相应的概率流常微分方程 (ODE)，其轨迹具有与逆时 SDE 相同的边际，并且是

$$d\mathbf{z} = \left[ f(\mathbf{z}, t) - \frac{1}{2}g(t)^2 \nabla_{\mathbf{z}} \log q_t(\mathbf{z}) \right]dt. \quad (27)$$

因此，逆时间 SDE 和概率流 ODE 可以从相同的数据分布中进行采样。

3) *Optimization*: 与 SGM 中的方法类似，我们定义一个得分模型，使得  $s_\theta(\mathbf{z}_t, t) \approx \nabla_{\mathbf{z}} \log q_t(\mathbf{z})$ 。此外，我们将 (22) 扩展到连续时间，如下所示：

$$\mathbb{E}_{\substack{t \sim \mathcal{U}(0,T) \\ \mathbf{z}_0 \sim q(\mathbf{z}_0) \\ \mathbf{z}_t \sim q(\mathbf{z}_t|\mathbf{z}_0)}} [\lambda(t) \|s_\theta(\mathbf{z}_t, t) - \nabla_{\mathbf{z}_t} \log q_t(\mathbf{z}_t | \mathbf{z}_0)\|^2] \quad (28)$$

其中  $\lambda(t) > 0$  是加权函数。

### D. Relation Between DMs

正如在 SDE 部分中强调的那样，我们可以用 SDE 描述两种变体，即 SGM 和 DDPM。我们还可以通过重新表述优化目标来展示这种密切的关系。对于 DDPM，我们在 (18) 中看到

$$\begin{aligned} & \text{KL}(q(\mathbf{z}_0, \dots, \mathbf{z}_T) \| p_\theta(\mathbf{z}_0, \dots, \mathbf{z}_T)) \\ & \stackrel{(ii)}{\geq} \mathbb{E}[-\log p_\theta(\mathbf{z}_0)] + c \end{aligned}$$

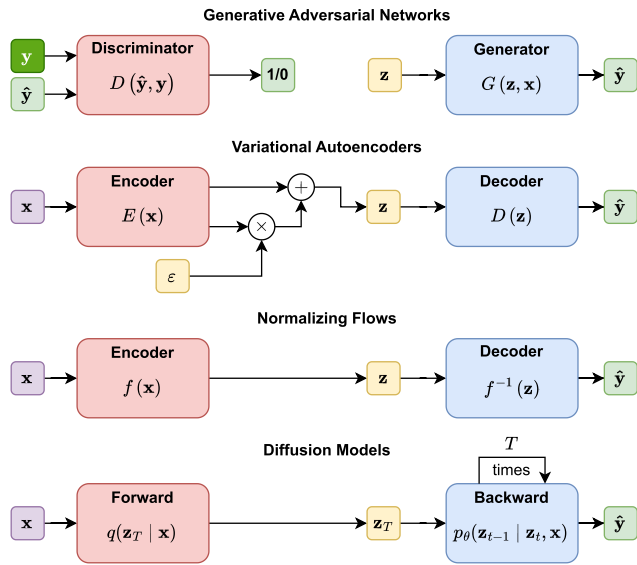


图 2. 生成模型 (GAN、VAE、NF 和 DM) 的概念概述。

最小化。通过重新加权 VLB，正如 Ho 等人 [8] 为提高样本质量所建议的那样，我们可以进一步得出

$$\mathbb{E}_{\substack{t \sim \mathcal{U}(1, T) \\ z_0 \sim q(z_0) \\ \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})}} [\lambda(t) \|\epsilon - \epsilon_\theta(z_t, t)\|^2]$$

其中  $\lambda(t) > 0$  是权重函数。如果我们现在取 SGM 中 (22) 式中的优化目标，即

$$\mathbb{E}_{\substack{t \sim \mathcal{U}(1, T) \\ z_0 \sim q(z_0) \\ \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})}} [\lambda(t) \|\epsilon + \sigma_t s_\theta(z_t, t)\|^2] + c$$

一旦我们设置了  $\epsilon_\theta(z_t, t) = -\sigma_t s_\theta(z_t, t)$ ，DDPM 和 SGM 之间的联系就变得清晰起来。由于常数  $c$  与优化无关，我们可以再次看到 DDPM 和 SGM 之间存在数学联系。

#### E. Relation to Other Image SR Generative Models

图像 SR 中的生成模型主要在处理从 LR 输入生成 HR 图像的任务方面有所不同，如图 2 所示。这些差异源于底层架构和训练目标。虽然它们具有显著的优势，但它们也面临着一系列挑战，例如训练稳定性和计算成本。

1) *GAN*: 一类著名的生成模型是 GAN [72]，它在各种视觉相关任务中都展现出了最佳性能，包括文本到图像 (T2I) 合成 [7] 和图像 SR [44]。GAN 以对抗性训练而闻名，其中生成器与鉴别器竞争。虽然 DM 不使用鉴别器，但它们利用类似的对抗性训练策略，通过迭代添加和去除噪声来实现逼真的数据生成。然而，使用 GAN 的方法通常存在不收敛、训练不稳定和计算成本高的问题。由于生成器和鉴别器之间的相互作用，它们需要仔细调整超参数。

2) *VAE*: VAE [73] 被设计为具有变分潜在空间的自动编码器，这在解决图像 SR 的不适定性方面尤其有用。VAE 的核心目标围绕建立对数数据似然性的 VLB，类似于 DM 的基本原理。在比较的背景下，我们可以将 DM 视为 VAE 的变体，但具有固定的 VAE 编码器，负责扰动输入数据，而 VAE 解码器类似于 DM 中的后向扩散过程。不过，与将输入压缩到潜在空间中较小维度的 VAE 不同，DM 通常保持相同的空间大小。

3) *ARM*: ARM 将图像视为像素序列，并根据先前生成的像素值按顺序生成每个像素 [6]。整个图像的概率是每个单独像素的条件概率分布的乘积。这使得 ARM 在 HR 图像生成方面的计算成本很高。相反，DM 通过将噪声逐渐扩散到初始数据样本中然后逆转此过程来生成数据。噪声会同时扩散到整个图像中，而不是按顺序扩散。

4) *NF*: NF [74] 是一类独特的生成模型，以将数据表示为错综复杂的分布而闻名。与 DM 和 VAE 一样，这些模型也是基于其生成数据的对数似然进行优化的。然而，NF 的与众不同之处在于其学习可逆参数化变换的独特能力。重要的是，这种变换具有易处理的雅可比行列式，使其易于计算。DiffFlow [75] 的概念作为一种创新算法进入人们的视野，它将 DM 的原理与 NF 的原理结合在一起。这种结合有望增强生成建模能力。然而，尽管前景光明，但 NF 通常被认为难以训练，并且计算要求很高 [76]。

#### IV. DMS 的改进

在更广泛的研究界，有几种方法可以改进用于图像生成的 DM，例如 Karras 等人所提出的 [77]。然而，本节重点介绍对图像 SR 特别有趣的增强：高效采样和增强似然估计。

##### A. Efficient Sampling

高效采样是指能够更快地从噪声中生成样本的策略，即在更少的时间步长内，而不会显著影响所生成图像的质量。例如，DDPM 需要大约 20 小时来采样 50 000 张  $32 \times 32$  图像，而 GAN 在 Nvidia 2080 Ti GPU 上只需不到一分钟；对于更大的  $256 \times 256$  图像，这需要近 1000 小时 [78]。幸运的是，在图像 SR 中通常会利用训练和推理计划之间的独立性。例如，一个模型可能以 1000 个时间步长进行训练，但后续的推理阶段可能只需要一小部分，即 200 [15]、[79]。然而，更广泛的 DM 研究界已经做出了进一步的尝试，专注于基于训练或无训练的采样。

*Training-based sampling* 方法使用训练有素的采样器来近似后向扩散过程，而不是传统的数值求解器，从而加速数据生成。这个过程可能是完整的，也可能是部分的。例如，Watson 等人 [80] 开发了一种动态规划算法，该算法使用固定数量的细化步骤来确定最佳推理路径，从而显著减少了所需的计算量。扩散采样器搜索 [81] 提供了另一种方法，通过调整内核起始距离来优化预训练 DM 的快速采样器。另一种技术是截断扩散，它通过提前结束前向扩散过程来提高速度 [82], [83]。这种提前终止会导致输出不是纯高斯噪声，从而带来计算挑战。这些挑战是使用预训练 VAE 或 GAN 的代理分布来解决的，它们与扩散的数据分布相匹配并促进有效的后向扩散。最后，知识蒸馏也用于加速采样。它涉及将知识从复杂、较慢的采样器（教师模型）转移到更简单、更快的模型（学生模型）[84], [85]。正如 Salimans 和 Ho [86] 所证明的，这种方法逐步减少了采样步骤的数量，以样本质量的略微下降来换取速度的提高。同样，Xiao 等人 [87] 解决了去噪步骤中与高斯假设相关的采样慢问题，这通常只对小步长有效。他们提出了去噪扩散 GAN，使用条件 GAN 进行去噪步骤，从而允许更大的步长和更快的采样。对于图像 SR，可以在 AddSR [88] 中找到利用知识蒸馏的应用程序。类似地，YONOS-SR [89] 使用知识蒸馏，但它们不是训练更快的采样器，而是传输不同的缩放任务知识，并使用无需训练的去噪扩散隐式模型 (DDIM) 进行有效采样，这将在下一节中介绍。

*Training-free sampling* 方法旨在通过在求解 SDE 或概率流 ODE 时最小化离散化步骤数来加快采样速度 [90], [91]。Song 等人 [90] 提出的 DDIM 将 DDPM 的马尔可夫前向扩散推广到非马尔可夫前向扩散。这种推广使 DDIM 能够学习马尔可夫链来逆转非马尔可夫前向扩散，从而实现更高的采样速度，同时将样本质量损失降至最低。Jolicœur-Martineau 等人 [91] 设计了一种具有自适应步长的高效 SDE 求解器，用于加速基于分数的模型的生成。结果表明，该方法比欧拉-丸山方法生成样本的速度更快，而且不影响样本质量。在 DDIM 和 Jolicœur-Martineau 等人的基础上，他们开发了一种基于分数的模型生成方法。[91], DDPM 求解器 [92] 受到 AnalyticalDDPM [93] 的启发，通过泰勒展开近似误差预测，从而通过解析解析 ODE 解的线性分量而不是依赖通用的黑盒 ODE 求解器来实现高效采样。该方法将采样步骤显著减少到 10–20。在后来的研究中，Lu 等人 [94] 引入了一个改进版本，其中 DDPM 求解器++ 本质上近似预测图像而不是误差。最近，一个更

UniPC [95] 引入了 DDPM 求解器++ 的通用公式和扩展。

### B. Improved Likelihood

对数似然的改进与各种应用和方法的性能增强直接相关，包括但不限于压缩 [96]、半监督学习 [97] 和图像 SR。鉴于 DM 不直接优化对数似然（例如，SGM 利用分数匹配损失的加权组合），因此需要优化形成负对数似然上限的目标。Song 等人 [98] 提出了一种名为 *likelihood weighting* 的方法来满足这一需求。此方法可最小化基于分数的 DM 的分数匹配损失的加权组合。精心选择的加权函数为加权分数匹配目标中的负对数似然设置了上限。最小化后，会导致对数似然升高。Kingma 等人 [99] 探索了同时训练噪声计划和扩散参数以最大化变分 DM 中的 VLB 的方法。此外，Nichol 和 Dhariwal [100] 提出的改进型去噪扩散概率模型 (iDDPM) 实现了余弦噪声调度。这会逐渐将噪声引入输入，而线性调度往往会更快地降低信息质量。使用余弦噪声调度可获得更好的对数似然，并有助于加快采样速度。

## V. 图像 SR 的 DMS

到目前为止，我们介绍了 DM 的理论框架。本节回顾了图像 SR 的实际应用和最新进展。我们将讨论 DM 的具体实现，主要是 DDPM。然后，我们讨论指导策略以增强条件使用，在 DDPM 的替代状态域中表示条件信息，并结合各种条件方法。此外，我们还探索了 SR 特定的研究领域，包括损坏空间、颜色变换和架构设计。图 3 提供了本节的拓扑概览。

### A. Concrete Realization of DMS

虽然 SGM 提供了相当大的设计灵活性，但图像 SR 趋势倾向于 DDPM。DDPM 的优势在于其实施简单，从而降低了进入门槛。这是一个显著的优势，因为它可以缩短开发周期并复制结果。此外，虽然 SGM 的灵活性有利于创建定制解决方案，但由于需要考虑大量设计变量，因此引入了设计复杂性。这在研究环境中带来了挑战，因为严格评估每个变量（例如，不同的采样算法）的影响变得很麻烦。此外，越来越多的 DDPM 文献也促进了它们的流行。随着越来越多的研究采用 DDPM，形成了一个良性循环，熟悉度和经过验证的有效性鼓励进一步采用。

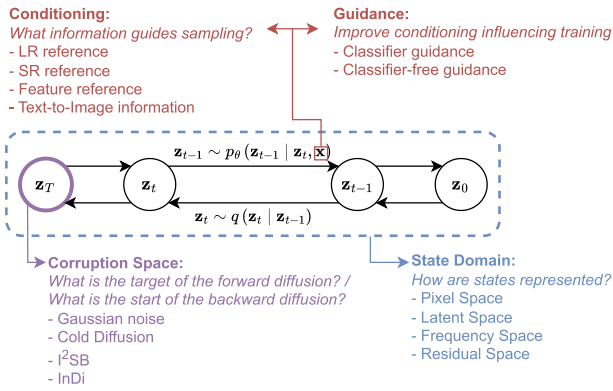


图 3. 本研究的拓扑结构。条件作用（第 V-D 节）引导反向扩散，而指导（第 V-B 节）是一种训练策略，旨在改善条件作用与 DM 的结合。状态域（第 V-C 节）描述了状态  $z_t$  的表示。腐败空间（第 V-E 节）描述了正向扩散过程的目标或反向扩散的起点。

先驱 DM 成果之一是 SR3 [15]，它具体实现了图像 SR 的 DDPM。与典型的 DDPM 一样，它将高斯噪声添加到 LR 图像中，直到  $z_T \sim \mathcal{N}(0, I)$ ，并在  $T$  细化步骤中迭代生成目标 HR 图像  $z_0$ 。SR3 使用去噪模型来预测噪声  $\epsilon_t$ 。去噪模型  $\varphi_\theta(x, z_t, \gamma_t)$  将 LR 图像  $x$ 、噪声方差  $\gamma_t$  和嘈杂的目标图像  $z_t$  作为输入。借助  $\varphi_\theta$  提供的  $\epsilon_t$  预测，我们可以重新表述 (14) 以近似  $z_0$ ，如下所示：

$$z_t = \sqrt{\gamma_t} \cdot \hat{z}_0 + \sqrt{1 - \gamma_t} \cdot \varphi_\theta(x, z_t, \gamma_t) \\ \iff \hat{z}_0 = \frac{1}{\sqrt{\gamma_t}} \cdot \left( z_t - \sqrt{1 - \gamma_t} \cdot \varphi_\theta(x, z_t, \gamma_t) \right). \quad (29)$$

将  $\hat{z}_0$  代入后验分布，以参数化  $p_\theta(z_{t-1} | z_t, x)$  的平均值，可得

$$\mu_\theta(x, z_t, \gamma_t) = \frac{1}{\sqrt{\alpha_t}} \left[ z_t - \frac{1 - \alpha_t}{\sqrt{1 - \gamma_t}} \cdot \varphi_\theta(x, z_t, \gamma_t) \right]. \quad (30)$$

在 SR3 中，作者将方差  $\Sigma_\theta$  简化为  $(1 - \alpha_t)$ ，以便于计算。因此，每个细化步骤（ $\epsilon_t \sim \mathcal{N}(0, I)$ ）可以表示为

$$z_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left[ z_t - \frac{1 - \alpha_t}{\sqrt{1 - \gamma_t}} \cdot \varphi_\theta(x, z_t, \gamma_t) \right] + \sqrt{1 - \alpha_t} \cdot \epsilon_t. \quad (31)$$

同时开展的工作重点是 SR3 的类似实现，但展示了实现去噪模型  $\varphi_\theta(x, z_t, \gamma_t)$  的不同变体，我们将在后面讨论。值得一提的是 SRDiff [79]，它大约在同一时间发布，并遵循了 SR3 的紧密实现。SRDiff 和 SR3 之间的主要区别在于 SR3 直接预测 HR 图像，而 SRDiff 预测 LR 和 HR 图像之间的残差信息，即差异。因此，它有一个替代状态域，接下来将讨论。

### B. Guidance in Training

基于扩散的图像 SR 的基础是条件分布的学习 [15], [101]。因此，条件  $x$ （例如 LR 图像）被集成到后向

扩散，即 DDPM 中的  $p_\theta(z_{t-1} | z_t, x)$  或 SGM/SDE 中的  $s_\theta(z_t, x, t)$ 。然而，这种简单的公式可能会导致模型忽略条件。一种称为指导的原理可以通过控制条件信息的权重来缓解这个问题，但代价是样本多样性。它可以分为分类器指导和无分类器指导。据作者所知，虽然它们可以有效地用于改进 DM，但它们还没有应用于图像 SR。

1) *Classifier Guidance*: 分类器引导使用分类器来引导扩散过程，方法是将 DM 的分数估计与采样期间分类器的梯度合并 [12]。此过程类似于 BigGAN [102] 中的低温或截断采样，有助于在模式覆盖率和样本保真度之间进行权衡。分类器与 DM 同时训练，以根据  $z_t$  预测条件信息  $x$ 。对于条件信息的加权，得分函数变为

$$\nabla_{z_t} \log q(z_t | x) = \nabla_{z_t} \log q(z_t) + \lambda \nabla_{z_t} \log q(x | z_t) \quad (32)$$

其中  $\lambda \in \mathbb{R}^+$  是用于控制权重的超参数。这种方法的缺点是它依赖于可以处理任意噪声输入的学习分类器，而大多数现有的预训练图像分类模型都缺乏这种能力。

2) *Classifier-Free Guidance*: 无分类器指导的目标是在没有分类器的情况下实现类似的结果[103]。它将(32)修改为

$$\nabla_{z_t} \log q(z_t | x) = (1 - \lambda) \nabla_{z_t} \log q(z_t) + \lambda \nabla_{z_t} \log q(z_t | x). \quad (33)$$

因此，我们有一个标准的无条件决策模型和一个条件决策模型，其分数估计值为  $\nabla_{z_t} \log q(z_t | x)$ 。当  $\lambda = 0$  时，无条件决策模型保持不变，而当  $\lambda = 1$  时，它与条件决策模型的 vanilla 公式一致。当  $\lambda > 1$  时，会出现有趣的情况，决策模型优先考虑条件信息并远离无条件得分函数，从而降低了生成无视条件信息的样本的可能性。然而，这种方法的主要缺点是训练两个独立决策模型的计算成本很高。可以通过训练单个条件模型并在无条件得分函数中用空值替换条件信息来缓解这一问题 [104]。

### C. State Domains

到目前为止，我们已经讨论了直接在像素空间上操作的方法。本节介绍将输入映射到替代状态域的不同方法：潜在、频率和残差空间。除了替代状态域带来的特殊挑战外，这些方法还涉及将像素域映射到其自身域的额外步骤，如图 4 所示。

1) *Latent Space*: SR3 [15] 和 SRDiff [79] 等模型通过在像素域中操作实现了高质量的 SR 结果。然而，由于这些模型的迭代性质和 RGB 空间中的高维计算，这些模型的计算量非常大。为了减少计算需求，

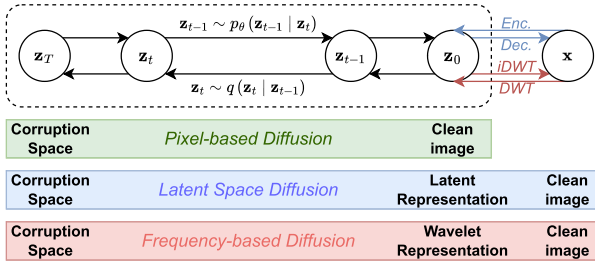


图 4. 状态域概览。绿色条显示原始 DM 在像素空间中的运行。蓝色条显示通过自动编码器对潜在空间域的利用。红色条显示 DM 在小波域中的应用。

可以将扩散过程移到自动编码器的潜在空间中 [105]。第一种此类模型是 Vahdat 等人 [106] 提出的基于潜在分数的生成模型 (LSGM)。它是一种在 VAE 的潜在空间中运行的常规 SGM，通过预训练 VAE，可实现更快的采样速度。它可产生与在像素域中运行的 DM 相当且更好的结果，同时速度更快。在 LSGM 的基础上，Rombach 等人 [13] 引入了潜在扩散模型 (LDM) [107]，它也在自动编码器的低维潜在空间中执行扩散。与 LSGM 相比，LDM 使用 DDPM 和预训练的自动编码器，例如 VQ-GAN [5]，并且不与去噪网络联合训练。这种方法显著降低了资源需求，同时不影响性能。由于训练过程解耦，它几乎不需要对潜在空间进行正则化，并允许在多个模型中重用潜在表示。Luo 等人 [108] 提出的 REFUSION（使用扩散模型进行图像恢复）对 LDM 进行了改进，其不同之处在于两个方面。首先，它使用一个包含从编码器到解码器的跳过连接的 U-Net，这为解码器提供了额外的细节。此外，它引入了非线性无激活块 (NAFBlocks) [109]，用元素级操作替换所有非线性激活，该操作将特征通道分成两部分并将它们相乘以产生一个输出。其次，他们使用潜在在替换训练策略训练 U-Net，该策略用编码的 LR 或 HR 图像部分替换潜在表示以进行重建训练。同样，Chen 等人 [110] 改进了 LDM 的架构方面，并提出了一种称为分层集成扩散模型 (HI-Diff) 的两阶段策略。在第一阶段，编码器将真实图像压缩为高度紧凑的潜在空间表示，其压缩率远高于 LDM。因此，DM（用于细化多尺度潜在表示）的计算负担大大减少。第二阶段是基于视觉变换器 (ViT) 的自动编码器，它在下采样过程中通过分层集成模块 (HIM)（一种交叉注意融合模块）整合第一阶段的潜在表示。

2) *Frequency Space*: 小波为 SR 提供了一种新颖的视角 [16], [111]。从空间域到小波域的转变是无损的，并且具有显著的优势，因为图像的空间尺寸可以缩小四倍，从而允许在训练期间更快地扩散

和推理阶段。此外，转换将高频细节分离到不同的通道中，有助于更集中、更有意识地关注高频信息，提供更高程度的控制 [112]。此外，它可以方便地作为插件功能合并到现有的 DM 中。扩散过程可以直接与所有小波交互，如 DiWa [113] 中提出的那样，或者专门针对某些波段，而其余波段则通过标准 CNN 进行预测。例如，WaveDM [114] 修改了低频带，而 WSGM [115] 或 ResDiff [116] 则相对于 LR 图像调节高频带。总之，小波域为未来的研究提供了一条有希望的途径。它有可能显著提高性能，同时保持（甚至提高）SR 结果的质量。

3) *Residual Space*: SRDiff [79] 是第一篇提倡将生成过程转移到残差空间（即上采样 LR 和 HR 图像之间的差异）的研究。这使 DM 能够专注于残差细节，加快收敛速度并稳定训练 [16], [111]。Whang 等人 [117] 也将残差预测作为其图像去模糊预测和细化方法的基本组成部分。然而，与 SRDiff 不同的是，他们使用 CNN 而不是双线性上采样 LR 提供 SR 预测，并使用他们的 DM 预测 SR 预测和 HR 基本事实之间的残差。ResDiff [116] 提出了一种改进方法，它在后向扩散期间还结合了 SR 预测及其高频信息，以提供更好的指导。Yue 等人 [118] 则以另一种方式提出了 ResShift。该技术通过操纵 HR 和 LR 图像之间的残差，在 HR 和 LR 图像之间构建马尔可夫变换链。因此，在前向过程中，残差不仅会添加均值为零的高斯噪声，还会作为训练期间噪声采样的均值添加。这种新方法大大提高了采样效率，即仅需 15 个采样步骤。

#### D. Conditioning DMs

DM 依靠条件信息来指导采样过程实现合理的 HR 预测。一种常见策略是在反向扩散期间使用 LR 图像。本节回顾了将条件信息整合到反向扩散中的各种替代方法。

1) *Low-Resolution Reference*: 通过简单的通道连接 [119] 可实现高质量 SR 预测。LR 图像与时间步  $t-1$  的去噪结果连接，并作为时间步  $t$  处噪声预测的条件输入。相比之下，Choi 等人 [120] 提出的迭代隐变量细化 (ILVR) 则对无条件 LDM [13] 的生成过程进行了条件化。这种方法利用了预先训练好的 DM，因此具有训练时间更短的优势。为了整合条件信息，去噪输出的低频分量被 LR 图像中相应的低频分量所替换。因此，在每个生成过程阶段，隐变量都与提供的参考图像对齐，从而确保在生成过程中进行精确的控制和调整。

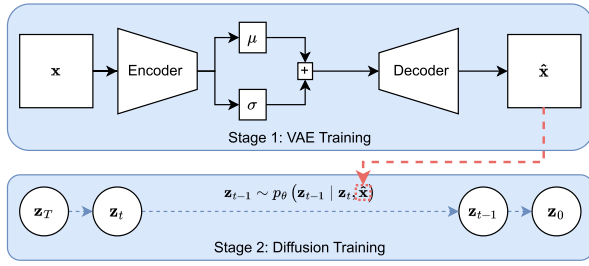


图 5. DiffuseVAE 概述。两阶段方法采用 VAE（第一阶段），生成变分预测作为 DM（第二阶段）的条件。

表 I DIV2K VAL 上一般图像的  $4\times$  SR 结果。请注意，EDSR、FxsR-PD、CAR 和 RRDB 是基于回归的方法，通常可以产生更好的 PSNR 和 SSIM

| SCORES THAN GENERATIVE APPROACHES [15] |                 |                 |                    |
|----------------------------------------|-----------------|-----------------|--------------------|
| Methods                                | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ |
| Bicubic                                | 26.70           | 0.77            | 0.409              |
| EDSR                                   | 28.98           | 0.83            | 0.270              |
| FxsR-PD                                | 29.24           | 0.84            | 0.239              |
| RRDB                                   | 29.44           | 0.84            | 0.253              |
| CAR                                    | 32.82           | 0.88            | -                  |
| RankSRGAN                              | 26.55           | 0.75            | 0.128              |
| ESRGAN                                 | 26.22           | 0.75            | 0.124              |
| SRFlow                                 | 27.09           | 0.76            | 0.120              |
| SRDiff                                 | 27.41           | 0.79            | 0.136              |
| IDM                                    | 27.59           | 0.78            | -                  |
| DiWa                                   | 28.09           | 0.78            | 0.104              |

表 II CELEBA-HQ 人脸  $16\times 16 \rightarrow 128 \times 128$  上的 PSNR 和 SSIM 比较。LR 输入与下采样 SR 输出之间的一致性测量 MSE ( $\times 10^{-5}$ )

| Methods          | PSNR $\uparrow$ | SSIM $\uparrow$ | Consistency $\downarrow$ |
|------------------|-----------------|-----------------|--------------------------|
| PULSE            | 16.88           | 0.44            | 161.1                    |
| FSRGAN           | 23.01           | 0.62            | 33.8                     |
| SR3 (regression) | 23.96           | 0.69            | 2.71                     |
| SR3 (diffusion)  | 23.04           | 0.65            | 2.68                     |
| DiWa             | 23.34           | 0.67            | -                        |
| IDM              | 24.01           | 0.71            | 2.14                     |

2) *Super-Resolved Reference*: 另一种在 LR 图像上调节去噪的方法是从预训练的 SR 模型中学习先验知识来预测参考图像。例如，CDPMSR [121] 使用现有和标准 SR 模型获得的预测 SR 参考图像来调节去噪过程。另一方面，ResDiff [116] 利用预训练的 CNN 来预测包含部分高频成分的低频、内容丰富的图像。该图像将噪声引导至残差空间，提供了一种调节生成过程的替代方法。

Pandey 等人 [127] 提出了一个令人兴奋的想法，即使用 DiffuseVAE 改变预测条件，如图 5 所示。这种方法将 VAE 生成的随机预测集成为 DM 的条件信息，充分利用了两种模型的优势。他们使用一种称为 *generator-refiner* 框架的两阶段方法。在第一阶段，VAE 在训练数据上进行训练。在第二阶段，VAE 在训练数据上进行训练。

在后续阶段，DM 使用由 VAE 生成的变化的、通常模糊的重构进行调节。该方法的主要优势在于生成的样本的多样性，该多样性在 VAE 的低维潜在空间中定义。此特性在采样速度和样本质量之间创造了更有利的平衡。它在需要多个预测的场景中是有利的，类似于 NF 的用例。

3) *Feature Reference*: 另一种调节方法涉及从预训练网络中提取的相关特征。SRDiff [79] 利用预训练编码器在后向扩散的每个步骤中对 LR 图像特征进行编码。这些特征可作为指导，帮助生成更高分辨率的输出。隐式 DM (IDM) [100] 采用不同的方法，通过用神经表征调节其去噪网络，从而能够学习各种尺度的连续表征。它们将图像编码为连续空间内的函数，并将其无缝集成到 DM 中。这些提取的特征适用于多种尺度，并在 DM 中的多个层中使用。为了全面了解这些方法之间的性能差异，可以在表 I 和表 II 中找到比较结果。最近，DeeDSR 被引入 [128]，它结合了从 LR 图像中提取的退化感知特征来指导 LDM [107] 的扩散过程。

4) *T2I Information*: 通过加入超出 LR 图像的条件信息（例如，其 SR 预测、LR 图像的直接连接或其特征表示），可以添加 T2I 信息。加入 T2I 信息被证明是有利的，因为它允许使用预先训练的 T2I 模型。这些模型可以通过添加针对 SR 任务量身定制的特定层或编码器进行微调，从而促进将文本描述集成到图像生成过程中。这种方法可以提供更丰富的指导来源，有可能改善 SR 任务中的图像合成和解释。Wang 等人 [107] 已将这一概念付诸实践，即 StableSR。StableSR 的核心是一个时间感知编码器，与冻结的稳定 DM（本质上是 LDM）一起训练。此设置无缝集成了可训练的空间特征变换层，从而能够根据输入图像进行条件调节。为了进一步增强 StableSR 的灵活性并在真实感和保真度之间实现微妙的平衡，他们引入了一个可选的可控特征包装模块。该模块可适应用户偏好，可根据个人需求进行微调。此功能的灵感来自 CodeFormer [129] 中引入的方法，该方法增强了 StableSR 的多功能性，可满足不同的用户需求和偏好。同样，Yang 等人 [130] 引入了一种称为像素感知稳定扩散 (PASD) 的方法。PASD 通过使用 CLIP 文本编码器 [59] 及其特征表示来合并 LR 输入的文本文本嵌入，使条件作用更进一步。这种方法通过合并文本信息增强了模型生成图像的能力，从而可以实现更精确和更具有上下文感知的图像合成。PASD 与其他方法的比较可以在表 III 中找到，展示了这种基于文本的条件作用对图像 SR 结果的影响。可以找到类似的并行工作

表三 调整大小后的 DIV2K VAL 上一般图像 4× SR 的结果 (128 × 128 → 512 × 512)

| Methods     | PSNR ↑ | SSIM ↑ | LPIPS ↓ |
|-------------|--------|--------|---------|
| BSRGAN      | 23.41  | 0.61   | 0.426   |
| Real-ESRGAN | 23.15  | 0.62   | 0.403   |
| LDL         | 22.74  | 0.62   | 0.416   |
| FeMaSR      | 21.86  | 0.54   | 0.410   |
| SwinIR-GAN  | 22.65  | 0.61   | 0.406   |
| LDM         | 21.48  | 0.56   | 0.450   |
| SD Upscaler | 21.21  | 0.55   | 0.430   |
| StableSR    | 20.88  | 0.53   | 0.438   |
| PASD        | 21.85  | 0.52   | 0.403   |

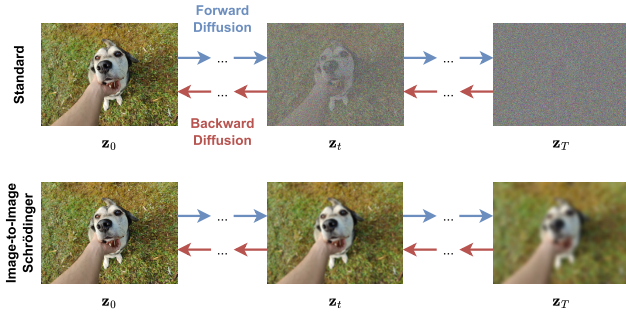


图 6. 标准损坏空间与  $I^2SB$  的比较。最终状态  $z_T$  是退化图像，而不是向干净图像（初始状态  $z_0$ ）注入噪声。

SeeSR [131] 提出了一种基于语义的文本编码方法。XPSR [132] 通过融合不同层次的语义文本编码（高级：图像的内容；低级：对 LR 图像的整体质量、清晰度、噪声水平和其他失真的感知）扩展了这一思想。

### E. Corruption Space

Karras 等人 [77] 确定了 DM 的三大支柱：噪声计划、网络参数化和采样算法。最近，许多作者主张考虑不同类型的损坏，而不是前向扩散期间使用的纯高斯噪声，例如软得分匹配 [135]，即后向扩散的起点或前向扩散的目标  $z_T$ 。软得分匹配直接将滤波过程纳入 SGM 中，训练模型预测干净的图像。损坏后，此预测图像与扩散观察结果对齐。注意，由于替代状态域（例如潜在、频率或残差）， $z_T$  的表示可能不同。冷扩散 [136] 提出了另一种修改 DDPM 损坏空间的巧妙方法。它表明生成能力并不强烈依赖于图像退化的选择。它表明，除了高斯噪声之外，还可以有效地使用新的实验类型的扩散，例如动物变身（即，人脸迭代退化为动物脸）。图像到图像的薛定谔桥（ $I^2SB$ ）朝着类似的方向发展，但不会对底层先验分布施加任何假设 [137]。在其扩散过程中，干净图像代表初始状态，而退化图像是前向和后向扩散过程中的最终状态。值得注意的是，它能够透明且



图 7. 当使用较小的批次大小（8 而不是 256）进行训练时，vanilla SR3 在  $64 \times 64 \rightarrow 256 \times 256$  设置中产生的颜色偏移示例。

如图 6 所示，冷扩散提供了一种从退化图像到其干净版本的可靠追踪路径。因此，它提供了增强的可解释性，因为直接解决了退化图像和干净图像之间的过程，而这在许多 DM 中并不常见。另一个好处是它在后向扩散中的效率更高，因为它需要更少的步骤（通常在 2 到 10 个之间）即可实现相当的性能。然而，它的条件性将其使用限制在训练期间的配对数据上，这不适合无监督 SR。虽然冷扩散和  $I^2SB$  在图像恢复方面显示出有希望的结果，但对图像 SR 的不同损坏类型进行广泛和更详细的定量分析仍然是一个令人兴奋且开放的研究途径。另一种替代损坏空间的途径是通过直接迭代反演 (InDI) [138] 提出。InDI 描述了一种直接映射策略，有效地弥合了两个质量空间之间的差距，而无需传统扩散过程通常需要的迭代细化。InDI 的内在灵活性和直接映射能力为增强图像质量提供了令人着迷的可能性，为研究探索提供了一条强大的途径。InDI 的原理与条件 DM 的原理的潜在整合可能会在图像 SR 领域带来重大进步。在基于扩散的图像增强的更广泛范围内对 InDI 进行详细研究和讨论可能会产生有价值的见解，并为图像处理中生成模型的持续发展做出重大贡献。

### F. Color Shifting

由于计算成本高，当有限的硬件需要较小的批量或较短的学习周期时，DM 偶尔会出现颜色偏移的情况 [139]。图 7 显示了 SR3 的示例。正如 StableSR 所介绍的那样，一个简单的修改可以通过执行颜色归一化来解决此问题，方法是在生成的图像上调整平均值和方差以适应 LR 图像的平均值和方差 [107]。从数学上讲，它给出以下方程：

$$\hat{z}_0 = \frac{z_0^c - \mu_{z_0}^c}{\sigma_{z_0}^c} \cdot \sigma_x^c + \mu_x^c \quad (34)$$

其中  $c \in \{r, g, b\}$  表示颜色通道， $\mu_{z_0}^c$  和  $\sigma_{z_0}^c$ （或  $\mu_x^c$  和  $\sigma_x^c$ ）分别是预测图像  $z_0$ （或输入图像  $x$ ）的第  $c$  个通道的均值和标准方差。YODA [140] 仅扩散区域，它通过使用 DINO [141] 生成的时间相关掩模更频繁地针对重要图像区域进行扩散，也可以减轻图像 SR 的色彩偏移效应。这表明

正确定义的架构和扩散设计对于消除这种影响至关重要。在未来的工作中必须进一步分析这种影响出现的原因。

### G. Architecture Designs for Denoising

DM 中的去噪模型设计提供了多种选择。正如大多数文献 [102] 所述, 大多数 DM 都采用 U-Net。例如, SR3 [15] 采用了 BigGAN [102] 的残差块, 并将跳过连接重新缩放为  $(1/(2)^{1/2})$  倍。SRDiff 采用了类似的方法 [79], 尽管它选择了原始残差块而不重新缩放跳过连接, 并使用 LR 编码器在后向扩散期间合并 LR 图像的信息。Whang 等人 [117] 利用初始预测器结合了确定性图像 SR 模型和 DM 的优势。它的优点是 DM 只需要学习确定性图像 SR 模型 (初始预测器) 无法预测的残差, 从而简化了学习目标。此外, 从 SR3 U-Net 中删除自注意力、位置编码和组规范化使其模型能够支持任意分辨率。基于小波的方法 DiWa [113] 也采用了初始预测器。此外, 小波 SR 模型 (例如 DWSR [112], 即深度为 10 的简单卷积层序列) 用于小波域中的去噪预测。在 WaveDM [114] 中, 确定性 U-Net 预测器用于高频带, 而扩散则应用于低频带。

Rombach 等人 [13] 提出的 LDM 在潜在空间中使用 VQ-GAN [5] 自动编码器。对于 DiffIR [142], 采用了多种最先进的 ViT 变体 [50]、[143]、[144]。另一种常见做法是预训练确定性组件, 如 SRDiff [79] 或 DiffIR [142] 等模型所示。总体而言, 设计去噪网络的潜在方法是无限的, 通常从通用计算机视觉的进步中汲取灵感。最佳去噪网络将根据任务而有所不同, 预计会开发出新的模型。

## VI. 基于扩散的零样本 SR

零样本图像 SR 旨在开发不依赖于先前图像示例或训练的方法 [16], [150]。通常, 这些方法利用单个图像中固有的冗余来改进。与前面讨论的其他条件方法不同, 它们通常利用预训练的 DM 进行生成, 在采样过程中将 LR 图像作为条件。此外, 它们不同于基于指导的方法, 在基于指导的方法中, 条件信息用于从头开始加权 DM 的训练。Li 等人最近的一项研究 [17] 将基于扩散的方法分为基于投影、基于分解和后验估计, 本节将介绍这些方法。表 IV 比较了所讨论的方法。

### A. Projection-Based

基于投影的方法旨在从 LR 图像中提取固有结构或纹理, 以补充生成的

在每个步骤中对图像进行扩散并确保数据一致性。修复任务领域中基于投影的方法的一个典型示例是 RePaint [152]。在 RePaint 中, 扩散过程被选择性地应用于需要修复的特定区域, 而其余图像部分则保持不变。受此概念的启发, YODA [140] 采用了类似的技术, 但针对的是图像 SR。YODA 结合了源自 DINO [141] 的重要性蒙版来定义每个时间步骤中要扩散的区域, 但它不是零样本方法。

一种零样本方法是 ILVR [120], 它将低频信息从 LR 图像投影到 HR 图像, 确保数据一致性并建立改进的 DM 条件。一种更复杂的方法是 come-closer-diffuse-faster (CCDF) [153], 它将统一投影方法修改为 SR, 如下所示:

$$\hat{z}_{t-1} = f(\mathbf{z}_t, t) + g(\mathbf{z}_t, t) \cdot \varepsilon_t \quad (35)$$

$$\mathbf{z}_{t-1} = (\mathbf{I} - \mathbf{P}) \cdot \hat{z}_{t-1} + \hat{x}, \quad \hat{x} \sim q(\mathbf{z}_t | \mathbf{z}_0 = \mathbf{x}) \quad (36)$$

其中  $f$  和  $g$  取决于 DM 的类型,  $\mathbf{P}$  是 LR 图像的退化过程,  $\hat{x}$  是添加了与时间相关的噪声的 LR 图像。

### B. Decomposition-Based

基于分解的方法将图像 SR 任务视为类似于 (1) 的线性逆问题

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{b} \quad (37)$$

其中  $\mathbf{A}$  是退化算子,  $\mathbf{b}$  是污染噪声。在最早的基于分解的方法中, 我们发现了 SNIPS [145] 及其后续工作 DDRM [146]。这些方法在谱域中采用扩散, 增强了 SR 结果。为了实现这一点, 它们将奇异值分解应用于退化算子  $\mathbf{A}$ , 从而促进了有助于改进 SR 结果的谱域变换。

去噪扩散零空间模型 (DDNM) 代表另一种基于分解的零样本方法, 适用于广泛的线性 IR 问题 [148], 从图像 SR 到着色、修复和去模糊等任务 [148]。它利用范围零空间分解方法 [154], [155] 有效地解决各种 IR 挑战。DDNM 通过将 (1) 重新配置为线性逆问题来解决问题, 但必须注意的是, 这种方法与 SNIPS 和 DDRM 的不同之处在于它在无噪声环境中运行

$$\mathbf{x} = \mathbf{A}\mathbf{y} \quad (38)$$

其中  $\mathbf{y} \in \mathbb{R}^{D \times 1}$  为线性化的 HR 图像,  $\mathbf{x} \in \mathbb{R}^{d \times 1}$  为线性化的退化图像。此外, 它还必须符合以下两个约束:

$$\text{Consistency : } \mathbf{A}\hat{\mathbf{y}} \equiv \mathbf{x}, \quad \text{Realness : } \hat{\mathbf{y}} \sim p(\mathbf{y}) \quad (39)$$

其中  $p(\mathbf{y})$  为真实图像分布,  $\hat{\mathbf{y}}$  为预测图像。范围零空间分解允许为  $\hat{\mathbf{y}}$  构建通解, 形式为

$$\hat{\mathbf{y}} = \mathbf{A}^\dagger \mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\bar{\mathbf{y}} \quad (40)$$

表 IV 零样本方法比较。粗体数据代表最佳性能。下划线表示第二佳性能。数值

来自 LI 等人。[17]

| Methods | ImageNet 1K     |                 |                    | CelebA 1K       |                 |                    | Time<br>[s/image] | Flops<br>[G] |
|---------|-----------------|-----------------|--------------------|-----------------|-----------------|--------------------|-------------------|--------------|
|         | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ |                   |              |
| Bicubic | 25.36           | 0.643           | 0.27               | 24.26           | 0.628           | 0.34               | -                 | -            |
| ILVR    | <u>27.40</u>    | <b>0.871</b>    | 0.21               | <u>31.59</u>    | 0.878           | 0.22               | 41.3              | 1113.75      |
| SNIPS   | 24.31           | 0.684           | 0.21               | 27.34           | 0.675           | 0.27               | 31.4              | -            |
| DDRM    | 27.38           | <u>0.869</u>    | 0.22               | <b>31.64</b>    | <b>0.946</b>    | 0.19               | <u>10.1</u>       | 1113.75      |
| DPS     | 25.88           | 0.814           | <u>0.15</u>        | 29.65           | 0.878           | 0.18               | 141.2             | 1113.75      |
| DDNM    | <b>27.46</b>    | <b>0.871</b>    | <u>0.15</u>        | <b>31.64</b>    | <u>0.945</u>    | <b>0.16</b>        | 15.5              | 1113.75      |
| GDP     | 26.51           | 0.832           | <b>0.14</b>        | 28.65           | 0.876           | <u>0.17</u>        | <b>3.1</b>        | 1113.76      |

其中  $A^\dagger \in \mathbb{R}^{D \times d}$  是满足  $AA^\dagger A \equiv A$  的伪逆。我们的目标是找到一个适当的  $\hat{y}$ ，使之生成零空间  $(I - A^\dagger A)\hat{y}$ ，并且与范围空间  $A^\dagger x$  一致，从而满足 (39) 中的实性。

DDNM 从时间步长  $t$  的  $z_0$  中导出范围零空间分解的干净中间状态，表示为  $z_{0|t}$ 。这是通过以下方程实现的

$$z_{0|t} = \frac{1}{\sqrt{\alpha_t}} \left( z_t - \epsilon_\theta(z_t, t) \sqrt{1 - \alpha_t} \right) \quad (41)$$

和  $\epsilon_t = \epsilon_\theta(z_t, t)$ 。为了生成满足方程  $Az_0 \equiv x$  的  $z_0$ ，模型保持零空间不变，同时将范围空间设置为  $A^\dagger y$ 。这将生成修正估计  $\hat{z}_{0|t}$ ，其定义为

$$\hat{z}_{0|t} = A^\dagger x + (I - A^\dagger A)z_{0|t}. \quad (42)$$

最后， $z_{t-1}$  是通过从  $p(z_{t-1}|z_t, \hat{z}_{0|t})$  中采样得出的。

$$\begin{aligned} z_{t-1} &= \frac{\sqrt{\alpha_{t-1}}\beta_t}{1 - \alpha_t} \hat{z}_{0|t} + \frac{\sqrt{\alpha_t}(1 - \alpha_{t-1})}{1 - \alpha_t} z_t + \sigma_t \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \end{aligned} \quad (43)$$

其中  $\alpha_t = 1 - \beta_t$  和  $\bar{\alpha}_t = \prod_{i=0}^t \alpha_i$ ，如图 8 所示。

项  $z_{t-1}$  表示  $\hat{z}_{0|t}$  的噪声版本。这种噪声有效地缓解了范围空间内容（由  $A^\dagger x$  表示）与零空间内容（由  $(I - A^\dagger A)z_{0|t}$  表示）之间的不一致。DDNM 的作者还表明  $\hat{z}_{0|t}$  符合一致性。

最后一步涉及定义  $A$  和  $A^\dagger$ ，其构造取决于手头的恢复任务。例如，在涉及按因子  $n$  缩放的 SR 任务中， $A$  可以定义为  $1 \times n^2$  矩阵，代表平均池化运算符。平均池化运算符表示为  $[(1/n^2) \dots (1/n^2)]$ ，用于将每个块平均为一个奇异值。类似地，我们可以将其伪逆构造为  $A^\dagger \in \mathbb{R}^{n^2 \times 1} = [1 \dots 1]^T$ 。原始作品提供了更多任务示例（例如着色、修复和恢复），说明了如何应用这些方法。此外，它还描述了由众多子操作组成的复合操作如何在这些环境中发挥作用。在他们的研究中，作者还引入了 DDNM<sup>+</sup> 来支持噪声图像的恢复。他们使用了一种类似于 RePaint [152] 中实现的“前后”策略的技术。利用这种方法可以进一步提高质量。

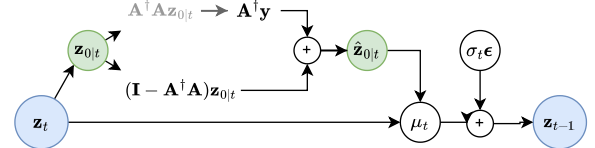


图 8. DDNM 概述 [148]。它利用范围零空间分解为多个任务构建通用解决方案，例如图像 SR、彩色化、修复和去模糊。

鉴于这种方法的新颖性，只有少数后续研究对其进行了扩展和构建，例如 CDPMSR [121] 中提出的研究。这个研究方向有望带来令人兴奋的可能性，尽管它需要进一步研究。例如，应该注意的是，与使用 DDPM 进行的任务特定训练相比，DDNM 方法引入了额外的计算开销。此外，退化算子  $A$  是手动设置的，这对于某些任务来说可能具有挑战性。另一个潜在的缺点是假设  $A$  充当线性退化算子，这可能并不总是成立，因此可能会限制模型在某些情况下的有效性。

### C. Posterior Estimation

大多数基于投影的方法通常解决无噪声逆问题。然而，这种假设可能会削弱数据一致性，因为投影过程可能会使样本路径偏离数据流形 [17]。为了解决这个问题并增强数据一致性，最近的一些研究 [147]、[156]、[157] 采取了不同的方法，旨在使用贝叶斯定理估计后验分布

$$p(z_t | x) = \frac{p(x | z_t) \cdot p(z_t)}{p(x)}. \quad (44)$$

这种贝叶斯方法为解决逆问题提供了更稳健、更概率的框架，最终改善了各种图像处理任务的结果。它得出相应的得分函数

$$\nabla_{z_t} \log p_t(z_t | x) = \nabla_{z_t} \log p_t(x | z_t) + s_\theta(x, t) \quad (45)$$

其中  $s_\theta(x, t)$  是从预训练模型中提取的，而  $p_t(x|z_t)$  是难以处理的。因此，目标是精确估计  $p_t(x|z_t)$ 。MCG [156] 和 DPS [147] 使用  $p_t(x|\hat{z}_0(z_t))$  近似后验  $p_t(x|z_t)$ ，其中  $\hat{z}_0(z_t)$  是根据 Tweedie 公式 [147] 给出的  $z_t$  为  $\hat{z}_0(z_t) = \mathbb{E}[z_0|z_t]$  的期望。虽然 MCG 也依赖于投影，但它可以

损害数据一致性, DPS 丢弃投影步骤, 估计后验概率为

$$\begin{aligned}\nabla_{z_t} \log p_t(\mathbf{x} | \mathbf{z}_t) &\approx \nabla_{z_t} \log p(\mathbf{x} | \hat{z}_0(\mathbf{z}_t)) \\ &\approx -\frac{1}{\sigma^2} \nabla_{z_t} \|\mathbf{x} - H(\hat{z}_0(\mathbf{z}_t))\|_2^2\end{aligned}\quad (46)$$

其中  $H$  是前向测量算子。在 IIGDM [157] 中可以找到该公式的进一步扩展, 即具有 Moore-Penrose 伪逆的线性、非线性、可微逆问题的统一形式。

GDP [149] 展示了一种估计  $p_t(\mathbf{x}|\mathbf{z}_t)$  的不同方法。作者指出,  $p_t(\mathbf{x}|\mathbf{z}_t)$  的条件概率越高, 退化模型  $\mathcal{D}(\mathbf{z}_t)$  的应用与  $\mathbf{x}$  之间的距离就越小。因此, 他们提出了一种启发式近似

$$p_t(\mathbf{x}|\mathbf{z}_t) \approx \frac{1}{Z} \exp(-[s\mathcal{L}(\mathcal{D}(\mathbf{z}_t), \mathbf{x})] + \lambda Q(\mathbf{z}_t)) \quad (47)$$

其中  $\mathcal{L}$  和  $Q$  分别表示距离和质量度量。术语  $Z$  用于规范化, 而  $s$  是控制指导权重的比例因子。然而, 由于  $\mathbf{z}_t$  和  $\mathbf{x}$  之间的噪声水平不同, 精确定义距离度量  $\mathcal{L}$  可能具有挑战性。为了克服这一挑战, GDP 在距离计算中用其干净估计  $\hat{z}_0$  代替  $\mathbf{z}_t$ , 为噪声差异问题提供了务实的解决方案。

## VII. 特定领域的应用

SR3 [15] 可在人脸和自然图像上生成照片般逼真且感知上最先进的图像, 但可能不适合遥感等其他任务。有些模型更适合某些任务, 因为它们可以解决特定领域的问题 [158]。本节重点介绍 DM 在特定领域的 SR 任务中的应用: 医学成像、面部 SR 的特殊情况 (BFR 和 AT) 以及遥感。

### A. Medical Imaging

磁共振成像 (MRI) 扫描被广泛用于辅助患者诊断, 但通常质量较低且受噪声影响。Chung 等人 [159] 提出了一种组合去噪和 SR 网络, 称为正则化反向扩散去噪器 + SR (R2D2+)。他们对 MRI 扫描进行去噪, 然后执行 SR 模块。受到 Chung 等人 [153] 的 CCDF (即零样本方法) 的启发, 他们从初始噪声图像而不是纯高斯噪声开始反向扩散。使用非参数、基于特征值的方法求解反向 SDE。此外, 他们通过低频正则化来限制 DM 的随机性。具体而言, 他们在校正高频信息的同时保留低频信息, 以产生清晰、超分辨率的 MRI 扫描。毛等人 [160] 解决了基于扩散的多对比 MRI SR 方法的缺乏问题。他们提出了一种解缠结条件扩散模型 (DisC-Diff), 以利用基于表示解缠结的多条件融合策略, 实现高质量的 HR 图像采样。具体来说, 他们采用具有多个编码器的解缠结 U-Net 来提取潜在表示, 并使用一种新颖的联合解缠结和 Charbonnier

损失函数来学习 MRI 对比中的表示。他们还实施课程学习, 并通过逐步增加训练图像的难度来改进其 MRI 模型, 以适应不同的解剖复杂性。Li 等人 [161] 使用 DiffMSR 通过将 DM 与转换器相结合引入了 DisC-Diff 的改进。

### B. Blind Face Restoration

之前讨论的大多数 SR 方法都是基于训练期间的固定退化过程, 例如双三次下采样。然而, 在实际应用中, 这些假设经常与实际退化过程不同, 并产生低于标准的结果。此外, 通常没有包含干净图像和真实世界失真图像对的数据集。这个问题在面部 SR (称为 BFR) 中得到了特别的研究, 其中数据集通常包含具有未知退化的监督样本  $(\mathbf{x}, \mathbf{y})$ 。

Yue 和 Loy [162] 使用 DifFace 提出了一种解决 BFR 问题的方法, 它利用了具有参数  $\theta$  的预训练 DM 的丰富生成先验, 这些 DM 被训练来近似  $p_\theta(\mathbf{z}_t|\mathbf{z}_{t-1})$ 。与在若干约束下学习从  $\mathbf{x}$  到  $\mathbf{y}$  的直接映射的现有方法 [163]、[164] 相比, DifFace 通过使用  $N < T$  生成所需 HR 图像  $\mathbf{y}$  的扩散版本  $\mathbf{z}_N$  来规避这个问题。它们通过过渡分布  $p(\mathbf{z}_N|\mathbf{x})$  预测起点, 即后验  $q(\mathbf{z}_N|\mathbf{x})$ 。过渡分布的公式类似于常规扩散过程, 即高斯分布, 但使用初始预测器  $\varphi(\mathbf{x})$  来生成均值, 即扩散估计量。由于他们的模型借用了预训练 DM 的逆马尔可夫链, 因此与 SR3 不同, DifFace 不需要对新的和未知的退化进行完全重新训练。

一种同时实现且性能更佳的方法是 Diff-BFR [165], 它采用两步方法实现 BFR: 身份恢复模块 (IRM), 采用两个条件 DDPM; 纹理增强模块 (TEM), 采用无条件 DDPM。在 IRM 的第一步中, 条件 DDPM 在与  $\mathbf{x}$  相同的 LR 空间中丰富面部细节。 $\mathbf{y}$  的下采样版本给出了目标目标。接下来, 它将输出调整为所需的  $\mathbf{y}$  空间大小, 并应用另一个条件 DDPM 来近似 HR 图像  $\mathbf{y}$ 。为了确保与实际图像的偏差最小, DiffBFR 采用了一种新颖的截断采样方法, 该方法在中间步骤开始去噪。TEM 通过图像纹理和锐化面部细节进一步增强了真实感。它施加了基于漫反射的面部先验, 其中无条件 DM 在 HR 图像上训练, 并从纯噪声开始向后扩散。但是, 它比 SR3 具有更多参数, 需要优化以加速采样。

另一种方法是 DR2E [166], 它采用两个阶段: 退化消除和增强模块。对于退化消除, 他们使用预训练的人脸 SR DDPM 从具有严重和未知退化的 LR 图像中去除退化。具体而言, 它们以  $T$  个时间步长扩散退化图像  $\mathbf{x}$  以获得  $\mathbf{x}_T = \mathbf{z}_T$ 。然后, 他们使用  $\mathbf{x}_t$  来引导后向扩散, 使得  $\mathbf{z}_t$  的低频部分被分布接近的  $\mathbf{x}_t$  的低频部分替换。理论上, 它可以产生视觉上干净的中间结果,

退化不变。在第二阶段，增强模块  $p_\theta(y | z_0)$ （一个任意的骨干 CNN，经过训练可使用简单的 L2 损失将 LR 图像映射到 HR）预测最终输出。对于有轻微退化的图像，DR 2E 可能比现有的基于扩散的 SR 模型更慢，甚至可以从输入中删除细节。

### C. AT in Face SR

AT 是由大气条件波动引起的，通过几何扭曲、空间变化模糊和噪声导致图像的感知退化。这些改变会对下游视觉任务（如跟踪或检测）产生负面影响。Wang 等 [167] 提出了一种变分推理框架，称为 AT-VarDiff，旨在纠正一般场景中的 AT。这种方法的显著特点是它依赖于从输入图像中提取的潜在任务特定先验信息派生出的条件信号来指导 DM。Nair 等 [168] 提出了另一种使用 SR 恢复受 AT 损害的面部图像的技术。该方法将类先验信息从在干净面部数据上训练的 SR 模型转移到旨在通过知识蒸馏抵消湍流退化的模型。最终模型在逼真的面部流形内运行，即使在严重扭曲的情况下也能生成逼真的面部输出。在推理过程中，该过程从噪声和湍流退化的图像开始，以确保恢复的图像与失真的图像非常相似。

### D. Remote Sensing

遥感超分辨率 (RSSR) 解决了从一张或多张 LR 图像进行 HR 重建的问题，以辅助卫星图像的物体检测和语义分割任务。RSSR 受限于 HR 图像中缺少具有复杂粒度的小目标 [169]。为了产生更精细的细节和纹理，刘等人 [170] 提出了具有细节补充 (DMDC) 机制的 DM。他们以类似于 SR3 [15] 的方式训练模型，并执行细节补充任务。为了生成高频信息，他们随机屏蔽了图像的几个部分以模拟密集物体。当模型学习小粒度信息时，SR 图像会恢复被遮挡的斑块。此外，他们引入了一种新颖的像素约束损失来限制 DMDC 的多样性并提高整体准确性。Ali 等人 [171] 为 RS 图像设计了一种新的架构，将 ViT 与 DM 集成，作为增强和超分辨率 (TESR) 的两阶段方法。在第一阶段（SR 阶段），SwinIR [50] 模型用于 RSSR。在第二阶段（增强阶段），通过使用 DM 重建更精细的细节，增强了噪声图像。Xu 等人 [172] 提出了一种基于双重条件 DDPM 的盲 SR 框架 (DDSR)。以 LR 图像编码为条件的内核预测器在第一阶段估计退化内核。接下来是一个 SR 模块，该模块由 U-Net 中的条件 DDPM 组成，以预测内核和 LR 编码为指导。RRDB 编码器从 LR 图像中提取编码。最近，Khanna 等人 [173] 引入了 DiffusionSat，它使用 LDM 进行 RSSR，并结合了额外的远程

感知条件信息（例如经度、纬度、云量等）。

## VIII. 讨论和未来工作

尽管相对较新，但 DM 正迅速成为一个有前途的研究领域，尤其是在图像 SR 领域。该领域正在进行多种研究，旨在提高 DM 的效率、加快计算速度并最大限度地减少内存占用，同时生成高质量、高保真度的图像。本节介绍了图像 SR 中 DM 的常见问题，并探讨了特定于图像 SR 的 DM 的值得关注的研究途径。

### A. Color Shifting

通常，最实用的进步来自于扎实的理论理解。如第 V-F 节所述，由于计算需求巨大，DM 在受到硬件限制（需要更小的批处理大小或更短的训练周期）的限制时，偶尔会出现颜色偏移 [139]。虽然定义明确的扩散方法 [140] 或颜色归一化 [107] 可能会缓解这个问题，但有必要从理论上理解它为什么会出现。

### B. Computational Costs

在 Ganguli 等人 [174] 进行的一项研究中，他们观察到大规模 AI 实验所需的计算能力在过去十年中激增了 30 多万倍。遗憾的是，这种资源强度的增加伴随着来自学术界的这些结果份额的急剧下降。DM 也不能幸免于此问题；它们的计算需求加剧了行业和学术界之间的差距。因此，迫切需要降低计算成本和内存占用，以实现实际应用和研究。减轻计算需求的一种策略是检查较小的空间大小域，如第 V-C 节所述。此类方法的示例包括 LDM [5]、[13] 和基于小波的模型 [113]、[115]。然而，LDM 能否以图像 SR 所需的高精度和细粒度精度重建数据仍有待商榷。因此，迫切需要进一步改进这些方法。另一方面，基于小波的模型在信息保存方面不存在瓶颈。这一优势表明它们应该成为更深入探索的主题。

### C. Efficient Sampling

DM 的一个好处是可以将训练和推理计划分离 [175]。这可以大大缩短实际应用中推理所需的时间，在现实场景中提供显著的效率优势。虽然减少推理过程中采取的步骤数量相对简单，但尚未开发出确定推理计划的系统方法 [176]。如第 IV-A 节所述，这个研究方向代表了一条有前途的途径。我们探索了基于训练的抽样

除了使用 AddSR [88] 和 YONOS-SR [89] 进行 SR 的方法之外, 还引入了需要更少采样步骤的高效 DM, 如 ResShift [118] 和 DiffIR [142]。如第 V-E 节所述, 使用不同损坏空间的方法给出了一种替代方案。与从纯高斯噪声中采样不同, 著名的研究如 Luo 等人 [108]、I<sup>2</sup>SB [137]、CCDF [153] 或 Cold Diffusion [136] 直接定义了从 LR 到 HR 图像的过程。其他减少计算时间的技术, 如知识蒸馏、替代噪声调度程序或截断扩散, 需要进一步研究图像 SR [84]、[85]、[87]、[177]。

#### D. Corruption Spaces

新的损坏空间方法允许更直接地将图像从 LR 上采样到 HR。探索不同损坏空间的意义在于解决当前 DM 框架中固有的局限性和假设, 例如在前向扩散过程中增加的多样性和模糊性。InDI 或 I<sup>2</sup>SB 等新方法所展示的适应性和效率, 尤其是在处理多样化和复杂的损坏模式方面, 凸显了未来研究的迫切需求。

#### E. Comparability

由于不同研究中使用的数据集各不相同, 因此比较 SR 中的 DM 非常复杂。它们的分辨率、内容多样性、颜色分布和噪声水平各不相同, 所有这些都会显著影响模型性能。一个模型可能在一个数据集上表现良好, 但在另一个数据集上表现不佳, 这使得评估其整体效果变得复杂。建立具有多样化、代表性数据集和统一评估指标的标准基准对于可比性至关重要。这种方法将有助于识别在不同条件和任务中始终表现良好的模型, 从而促进该领域的更快进步。此外, 评估生成模型的 SR 图像质量仍然存在问题。尽管 DM 通常可以生成更逼真的图像, 但它们在 PSNR 和 SSIM [16] 等标准指标上的得分通常较低。然而, 这些模型往往会从人类评估者那里得到更有利的评价 [15]。LPIPS [178] 的表现更好地反映了这种看法, 但图像 SR 领域必须适应更多样化的指标, 例如直接反映人类评分的预测因子 [179]、[180]。例如, 具有主观评分的数据集 (如 TID2013 [61]) 和神经网络 (如 DeepQA [62] 或 NIMA [63]) 可用于预测类似人类的图像评分, 值得进一步探索。

#### F. Image Manipulation

图像处理在多图像 SR 中特别有用, 可用于生成融合多种来源特征的 HR 图像, 从而有可能提高输出的质量和多样性 (例如, 用于灵活日光的 SR 预测的卫星图像)。SR Diff [79] 提出了两种潜在的扩展: 内容融合和潜在

空间插值。内容融合涉及两幅源图像内容的组合。例如, 在图像空间中进行扩散之前, 它们将一幅源图像中的眼睛替换为另一幅图像中的脸部, 就像 CutMix [181] 一样。反向扩散成功地在两幅图像之间创建了平滑的过渡。在潜在空间插值模型中, 两个 SR 预测的潜在空间被线性插值以生成新图像。虽然这些扩展已经产生了显著的结果, 但与其他生成模型 (如 VAE 或 GAN) 不同, DM 提供的潜在表示不太熟练 [182]。因此, 近期和正在进行的对 DM 中潜在表示操纵的研究既处于早期阶段, 也非常有必要 [183]、[184]、[185]。

#### G. Cascaded Image Generation

Saharia 等人 [15] 提出了级联图像 SR, 其中多个 DDPM 跨不同尺度链接。该策略应用于无条件类和条件生成, 将合成  $64 \times 64$  图像的模型与生成  $1024 \times 1024$  无条件人脸和  $256 \times 256$  类条件自然图像的 SR3 模型级联。级联方法允许同时训练几个更简单的模型, 由于训练时间更快、参数数量减少, 计算效率更高。此外, 他们实现了级联推理, 在较低分辨率下使用更多细化步骤, 在较高分辨率下使用更少步骤。他们发现这比直接生成 SR 图像更有效。尽管他们的方法在级联生成方面不如 BigGAN [102], 但它仍然代表着一个令人兴奋的研究机会。

### 九、结论

DM 通过增强技术图像质量和人类感知偏好彻底改变了图像 SR。虽然传统 SR 通常只关注像素级精度, 但 DM 可以生成美观逼真的 HR 图像。与以前的生成模型不同, 它们不会遇到典型的收敛问题。本文探讨了推动 DM 走上 SR 前沿的进展和各种方法。正如我们在应用程序部分所讨论的那样, 潜在的用例远远超出了我们之前的想象。我们介绍了它们的基本原理, 并将它们与其他生成模型进行了比较。我们探索了条件策略, 从 LR 图像引导到文本嵌入。零样本 SR 是一个特别有趣的范例, 也是一个主题, 以及损坏空间和图像 SR 特定的主题, 如色彩变换和建筑设计。总之, 本文提供了当前形势的全面指南, 以及对趋势、挑战和未来方向的宝贵见解。随着我们继续探索和完善这些模型, 图像 SR 的未来看起来比以往任何时候都更有希望。

#### 参考

- [1] W. Sun and Z. Chen, “使用内容自适应重采样器学习图像降尺度以实现升尺度”, *IEEE Trans. Image Process.*, 第 29 卷, 第 4027-4040 页, 2020 年。

- [2] D. Valsesia 和 E. Magli, “多时间图像超分辨率中的置换不变性和不确定性”, *IEEE Trans. Geosci. Remote Sens.*, 第 60 卷, 2022 年. [3] S. M. A. Bashir、Y. Wang、M. Khan 和 Y. Niu, “基于深度学习的单图像超分辨率综合回顾”, *PeerJ Comput. Sci.*, 第 7 卷, 第 e621 页, 2021 年 7 月. [4] D. Lee、C. Kim、S. Kim、M. Cho 和 W.-S. Han, “使用残差量化的自回归图像生成”, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 11513-11522 页. [5] P. Esser、R. Rombach 和 B. Ommer, “Taming transformers for high-resolution image synthesis”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021 年 6 月, 第 12868-12878 页. [6] B. Guo、X. Zhang、H. Wu、Y. Wang、Y. Zhang 和 Y.-F. Wang, “LAR-SR: 一种用于图像超分辨率的局部自回归模型”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 1899-1908 页. [7] S. Frolov、T. Hinz、F. Raue、J. Hees 和 A. Dengel, “对抗性文本到图像合成: 综述”, *Neural Netw.*, 第 144, 第 187-209 页, 2021 年 12 月. [8] J. Ho、A. Jain 和 P. Abbeel, “去噪扩散概率模型”, 载于 *Proc. NeurIPS*, 第 33 卷, 2020 年. [9] I. Goodfellow 等人, “生成对抗网络”, *Commun. ACM*, 第 63 卷, 第 11 期, 2020 年. [10] Y. Song 和 S. Ermon, “通过估计数据分布梯度进行生成建模”, 载于 *Proc. NeurIPS*, 第 33 卷, 2020 年. 32, 2019 年. [11] Y. Song、J. Sohl-Dickstein、D. P. Kingma、A. Kumar、S. Ermon 和 B. Poole, “通过随机微分方程进行基于分数的生成建模”, 2020 年, *arXiv:2011.13456*. [12] P. Dhariwal 和 A. Nichol, “扩散模型在图像合成方面击败 GAN”, 载于 *Proc. NeurIPS*, 第 34 卷, 2021 年. [13] R. Rombach、A. Blattmann、D. Lorenz、P. Esser 和 B. Ommer, “使用潜在扩散模型进行高分辨率图像合成”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 10674-10685 页. [14] A. Ramesh、P. Dhariwal、A. Nichol、C. Chu 和 M. Chen, “使用 CLIP 潜在特征的分层文本条件图像生成”, 2022 年, *arXiv:2204.06125*. [15] C. Saharia、J. Ho、W. Chan、T. Salimans、D. J. Fleet 和 M. Norouzi, “通过迭代细化实现图像超分辨率”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 第 45 卷, 第 4 期, 第 4713-4726 页, 2023 年 4 月. [16] B. B. Moser、F. Raue、S. Frolov、S. Palacio、J. Hees 和 A. Dengel, “超分辨率漫游指南: 简介和最新进展”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 第 45 卷, 第 8, 第 9862-9882 页, 2023 年 8 月. [17] X. Li 等人, “图像恢复和增强的扩散模型——综合调查”, 2023 年, *arXiv:2308.09388*. [18] A. Liu、Y. Liu、J. Gu、Y. Qiao 和 C. Dong, “盲图像超分辨率: 调查及其他”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 第 45 卷, 第 5 期, 第 5461-5480 页, 2023 年 5 月. [19] K. Zhang、J. Liang、L. Van Gool 和 R. Timofte, “设计一种实用的深度盲图像超分辨率退化模型”, *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021 年 10 月, 第 4771-4780 页. [20] X. Wang、L. Xie、C. Dong 和 Y. Shan, “Real-ESRGAN: 使用纯合成数据训练真实世界盲超分辨率”, *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021 年 10 月, 第 1905-1914 页. [21] S. Anwar 和 N. Barnes, “密集残差拉普拉斯超分辨率”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 第 44 卷, 第 1 期. 3, 第 1192-1204 页, 2022 年 3 月. [22] MATLAB, Mathworks, Inc. 美国马萨诸塞州纳蒂克, 2017 年. [23] E. Agustsson 和 R. Timofte, “NTIRE 2017 单图像超分辨率挑战赛: 数据集和研究”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017 年 7 月, 第 1122-1131 页. [24] M. Bevilacqua、A. Roumy、C. Guillemot 和 M. L. Alberi-Morel, *Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding*. 英国: 英国机器视觉协会 (BMVA) 出版社, 2012 年. [25] R. Zeyde、M. Elad 和 M. Protter, “论使用稀疏表示的单幅图像放大”, 载于 *Proc. Int. Conf. Curves Surf.* Cham, 瑞士: Springer, 2010 年. [26] D. Martin、C. Fowlkes、D. Tal 和 J. Malik, “人类分割自然图像数据库及其在评估分割算法和测量生态统计数据中的应用”, 载于 *Proc. 8th IEEE Int. Conf. Comput. Vision. ICCV*, 第 2 卷, 2001 年 7 月, 第 416-423 页. [27] J.-B. Huang、A. Singh 和 N. Ahuja, “从转换后的自我样本实现单幅图像超分辨率”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015 年 6 月, 第 5197-5206 页. [28] Y. Matsui 等人, “使用 Manga109 数据集进行基于草图的漫画检索”, *Multimedia Tools Appl.*, 第 76 卷, 第 20 期, 第 21811-21838 页, 2017 年 10 月. [29] T. Karras、S. Laine 和 T. Aila, “基于风格的生成对抗网络生成器架构”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019 年 6 月, 第 4396-4405 页. [30] T. Karras、T. Aila、S. Laine 和 J. Lehtinen, “逐步发展 GAN 以提高质量、稳定性和变化”, 2017 年, *arXiv:1710.10196*. [31] J. Deng、W. Dong、R. Socher、L.-J. Li、K. Li 和 L. Fei-Fei, “ImageNet: 一个大规模分层图像数据库”, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009 年 6 月, 第 248-255 页. [32] M. Everingham、L. Van Gool、C. K. I. Williams、J. Winn 和 A. Zisserman. (2012 年). *The PASCAL Voc2012 Results*. [在线]. 可访问网址: <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html> [33] K. I. Kim 和 Y. Kwon, “使用稀疏回归和自然图像先验的单图像超分辨率”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 第 32 卷, 第 6 期, 第 1127-1133 页, 2010 年 6 月. [34] G. Freedman 和 R. Fattal, “从局部自我示例进行图像和视频升级”, *ACM Trans. Graph.*, 第 30 卷, 第 2 期, 第 1-11 页, 2011 年 4 月. [35] J. Sun、Z. Xu 和 H.-Y. Shum, “使用梯度剖面先验的图像超分辨率”, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008 年 6 月, 第 1-8 页. [36] H. Chang、D.-Y. Yeung 和 Y. Xiong, “通过邻域嵌入实现超分辨率”, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 第 1 卷, 2004 年 7 月. [37] W. Freeman、T. Jones 和 E. Pasztor, “基于实例的超分辨率”, *IEEE Comput. Graph. Appl.*, 第 22 卷, 第 2 期, 第 56-65 页, 2002 年 3 月. [38] R. Keys, “用于数字图像处理的三次卷积插值”, *IEEE Trans. Acoust., Speech, Signal Process.*, 第 ASSP-29 卷, 第 6 期, 第 1153-1160 页, 1981 年 12 月. [39] M. Irani 和 S. Peleg, “通过图像配准提高分辨率”, *Graph. Models Image Process.*, 第 53 卷, 第 3, 第 231-239 页, 1991 年 5 月. [40] J. Yang、J. Wright、T. S. Huang 和 Y. Ma, “通过稀疏表示实现图像超分辨率”, *IEEE Trans. Image Process.*, 第 19 卷, 第 11 期, 第 2861-2873 页, 2010 年 11 月. [41] C. Dong、C. C. Loy、K. He 和 X. Tang, “使用深度卷积网络实现图像超分辨率”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 第 38 卷, 第 2 期, 第 295-307 页, 2016 年 2 月. [42] C. Dong、C. C. Loy 和 X. Tang, “加速超分辨率卷积神经网络”, 载于 *Proc. ECCV*. Cham, 瑞士: Springer, 2016 年. [43] W. Shi 等人, “使用高效亚像素卷积神经网络实现实时单图像和视频超分辨率”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016 年 6 月, 第 1874-1883 页. [44] C. Ledig 等人, “使用生成对抗网络实现照片级逼真的单图像超分辨率”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017 年 7 月, 第 105-114 页. [45] G. Huang、Z. Liu、L. Van Der Maaten 和 K. Q. Weinberger, “密集连接的卷积网络”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017 年 7 月, 第 2261-2269 页. [46] T. Tong、G. Li、X. Liu 和 Q. Gao, “使用密集跳过连接实现图像超分辨率”, 载于 *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017 年 10 月, 第 4809-4817 页. [47] J. Kim、J. K. Lee 和 K. M. Lee, “用于图像超分辨率的深度递归卷积网络”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016 年 6 月, 第 1637-1645 页. [48] Y. Tai、J. Yang 和 X. Liu, “通过深度递归残差网络实现图像超分辨率”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017 年 7 月, 第 2790-2798 页. [49] N. Ahn、B. Kang 和 K.-A. Sohn, “通过级联残差网络实现快速、准确、轻量级的超分辨率”, 载于 *Proc. ECCV*, 2018 年. [50] J. Liang、J. Cao、G. Sun、K. Zhang、L. Van Gool 和 R. Timofte, “SwinIR: 使用 swin 变压器进行图像恢复”, 载于 *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021 年 10 月, 第 1833-1844 页.

- [51] X. Chen, X. Wang, J. Zhou, Y. Qiao 和 C. Dong, “在图像超分辨率 Transformer 中激活更多像素”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023 年 6 月, 第 22367-22377 页。[52] C.-C. Hsu, C.-M. Lee 和 Y.-S. Chou, “DRCT: 使图像超分辨率远离信息瓶颈”, 2024 年, *arXiv:2404.00722*。[53] X. Wang 等人, “ESRGAN: 增强型超分辨率生成对抗网络”, 载于 *Proc. ECCVW*, 2018 年。[54] A. Lugmayr, M. Danelljan, L. Van Gool 和 R. Timofte, “SRFlow: 使用规范化流学习超分辨率空间”, 载于 *Proc. ECCV*, Cham, 瑞士: Springer, 2020 年, 第 715–732 页。[55] K. Simonyan 和 A. Zisserman, “用于大规模图像识别的超深卷积网络”, 2014 年, *arXiv:1409.1556*。[56] A. Krizhevsky, I. Sutskever 和 G. E. Hinton, “使用深度卷积神经网络进行 ImageNet 分类”, *Commun. ACM*, 第 60 卷, 第 6 期, 第 84–90 页, 2017 年 5 月。[57] A. Mittal, A. K. Moorthy 和 A. C. Bovik, “空间域中的无参考图像质量评估”, *IEEE Trans. Image Process.*, 第 21 卷, 第 6 期, 第 84–90 页, 2017 年 5 月。12, 第 4695-4708 页, 2012 年 12 月。[58] A. Mittal, R. Soundararajan 和 A. C. Bovik, “制作‘完全盲目的’图像质量分析仪”, *IEEE Signal Process. Lett.*, 第 20 卷, 第 3 期, 第 209-212 页, 2013 年 3 月。[59] A. Radford 等人, “从自然语言监督中学习可转移的视觉模型”, 载于 *Proc. ICML*, 2021 年。[60] J. Wang, K. C. Chan 和 C. C. Loy, “探索用于评估图像外观和感觉的剪辑”, 载于 *Proc. AAAI*, 2023 年, 第 37 卷, 第 2。[61] N. Ponomarenko 等人, “图像数据库 TID2013: 特点、结果和观点”, *Signal Process., Image Commun.*, 第 30 卷, 第 57-77 页, 2015 年 1 月。[62] J. Kim 和 S. Lee, “图像质量评估框架中人类视觉敏感度的深度学习”, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017 年 7 月, 第 1969-1977 页。[63] H. Talebi 和 P. Milanfar, “NIMA: 神经图像评估”, *IEEE Trans. Image Process.*, 第 27 卷, 第 8, 第 3998-4011 页, 2018 年 8 月。[64] J. Ke, Q. Wang, Y. Wang, P. Milanfar 和 F. Yang, “MUSIQ: 多尺度图像质量变换器”, 载于 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021 年 10 月, 第 5128-5137 页。[65] L. Yang 等人, “扩散模型: 方法和应用的综合调查”, *ACM Comput. Surv.*, 第 56 卷, 第 4, 第 1-39 页, 2024 年 4 月。[66] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan 和 S. Ganguli, “使用非平衡热力学的深度无监督学习”, 载于 *Proc. ICML*, 2015 年。[67] G. Parisi, “相关函数和计算机模拟”, *Nucl. Phys. B*, 第 180 卷, 第 3 期, 第 378-384 页, 1981 年 5 月。[68] P. Vincent, “分数匹配和去噪自动编码器之间的联系”, *Neural Comput.*, 第 23 卷, 第 7, 第 1661-1674 页, 2011 年 7 月。[69] A. Hyvärinen 和 P. Dayan, “通过分数匹配估计非归一化统计模型”, *J. Mach. Learn. Res.*, 第 6 卷, 第 4 期, 2005 年。[70] Y. Song, S. Garg, J. Shi 和 S. Ermon, “切片分数匹配: 一种可扩展的密度和分数估计方法”, 载于 *Proc. 35th Uncertainty Artif. Intell. Conf.*, 第 115 卷。P MLR, 2020 年, 第 574-584 页。[71] B. D. O. Anderson, “逆时间扩散方程模型”, *Stochastic Processes their Appl.*, 第 12 卷, 第 3, 第 313-326 页, 1982 年 5 月。[72] I. Goodfellow 等人, “生成对抗网络”, 载于 *Proc. NeurIPS*, 第 27 卷, 2014 年。[73] D. P. Kingma 和 M. Welling, “自编码变分贝叶斯”, 2013 年, *arXiv:1312.6114*。[74] D. Rezende 和 S. Mohamed, “具有规范化流的变分推理”, 载于 *Proc. ICML*, 2015 年。[75] Q. Zhang 和 Y. Chen, “扩散规范化流”, 载于 *Proc. NeurIPS*, 第 27 卷, 2014 年。34, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang 和 J. W. Vaughan 编, 2021 年。[76] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed 和 B. Lakshminarayanan, “概率建模和推理的规范化流程”, *J. Mach. Learn. Res.*, 第 22 卷, 第 1 期, 2021 年。[77] T. Karras, M. Aittala, T. Aila 和 S. Laine, “阐明基于扩散的生成模型的设计空间”, *Proc. NeurIPS*, 第 35 卷, 2022 年。[78] *Oxford VGGFace Implementation Using Keras Functional Framework V2+*。访问日期: 2024 年 10 月 17 日。[在线]。可访问网址: <https://github.com/rcmalli/keras-vggface> [79] H. Li 等人, “SRDiff: 具有扩散概率模型的单图像超分辨率”, *Neurocomputing*, 第 479 卷, 第 47-59 页, 2022 年 3 月。[80] D. Watson, J. Ho, M. Norouzi 和 W. Chan, “学习从扩散概率模型中有效采样”, 2021 年, *arXiv:2106.03802*。[81] D. Watson, W. Chan, J. Ho 和 M. Norouzi, “通过样本质量区分学习扩散模型的快速采样器”, 2022 年, *arXiv:2202.05830*。[82] Z. Lyu, X. Xu, C. Yang, D. Lin 和 B. Dai, “通过提前停止扩散过程加速扩散模型”, 2022 年, *arXiv:2205.12524*。[83] L. Zhang, A. Rao 和 M. Agrawala, “为文本到图像扩散模型添加条件控制”, 载于 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023 年 10 月, 第 3813-3824 页。[84] C. Meng 等人, “论引导扩散模型的蒸馏”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023 年 6 月, 第 14297-14306 页。[85] E. Luhman 和 T. Luhman, “迭代生成模型中的知识蒸馏以提高采样速度”, 2021 年, *arXiv:2101.02388*。[86] T. Salimans 和 J. Ho, “渐进式蒸馏以快速采样扩散模型”, 2022 年, *arXiv:2202.00512*。[87] Z. Xiao, K. Kreis 和 A. Vahdat, “使用去噪扩散 GAN 解决生成学习三难困境”, 2021 年, *arXiv:2112.07804*。[88] R. Xi 等人, “AddSR: 通过对抗性扩散蒸馏加速基于扩散的盲超分辨率”, 2024 年, *arXiv:2404.01717*。[89] M. Noroozi, I. Hadji, B. Martinez, A. Bulat 和 G. Tzimiropoulos, “只需一步: 通过规模蒸馏实现稳定扩散的快速超分辨率”, 2024 年, *arXiv:2401.17258*。[90] J. Song, C. Meng 和 S. Ermon, “去噪扩散隐式模型”, 2020 年, *arXiv:2010.02502*。[91] A. Jolicœur-Martineau, K. Li, R. Piché-Taillefer, T. Kachman 和 I. Mitliagkas, “使用基于分数的模型生成数据时必须快速行动”, 2021 年, *arXiv:2105.14080*。[92] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li 和 J. Zhu, “DPM-Solver: 一种大约 10 步的扩散概率模型采样快速解算器”, 载于 *Proc. Adv. Neural Inf. Process. Syst.*, 第 35 卷, 2022 年, 第 5775-5787 页。[93] F. Bao, C. Li, J. Zhu 和 B. Zhang, “Analytic-DPM: 扩散概率模型中最优逆方差的解析估计”, 2022 年, *arXiv:2201.06503*。[94] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li 和 J. Zhu, “DPM-Solver++: 扩散概率模型引导采样快速解算器”, 2022 年, *arXiv:2211.01095*。[95] W. Zhao, L. Bai, Y. Rao, J. Zhou 和 J. Lu, “UniPC: 用于快速采样扩散模型的统一预测校正框架”, 载于 *Proc. NeurIPS*, 第 36 卷, 2024 年。[96] J. Ho, E. Lohm 和 P. Abbeel, “通过局部位回编码对流进行压缩”, 载于 *Proc. NeurIPS*, 第 32 卷, 2019 年。[97] Z. Dai, Z. Yang, F. Yang, W. W. Cohen 和 R. R. Salakhutdinov, “需要糟糕的 GAN 才能实现良好的半监督学习”, 载于 *Proc.*, 第 30 卷, 2017 年。[98] Y. Song, C. Durkan, I. Murray 和 S. Ermon, “基于分数的扩散模型的最大似然训练”, 载于 *Proc. NeurIPS*, 第 33 卷, 20 34, 2021 年。[99] D. Kingma, T. Salimans, B. Poole 和 J. Ho, “变分扩散模型”, 载于 *Proc. NeurIPS*, 第 34 卷, 2021 年。[100] A. Q. Nichol 和 P. Dhariwal, “改进的去噪扩散概率模型”, 载于 *Proc. ICML*, 2021 年。[101] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi 和 T. Salimans, “用于高保真图像生成的级联扩散模型”, *J. Mach. Learn. Res.*, 第 23 卷, 第 47 期, 2022 年。[102] A. Brock, J. Donahue 和 K. Simonyan, “用于高保真自然图像合成的大规模 GAN 训练”, 2018 年, *arXiv:1809.11096*。[103] J. Ho 和 T. Salimans, “无分类器扩散指导”, 2022 年, *arXiv:2207.12598*。[104] C. Luo, “理解扩散模型: 统一的视角”, 2022 年, *arXiv:2208.11970*。[105] J. Kim 和 T.-K. Kim, “使用潜在扩散模型和隐式神经解码器进行任意尺度图像生成和上采样”, 2024 年, *arXiv:2403.10255*。[106] A. Vahdat 和 K. Kreis 和 J. Kautz, “基于分数的潜在空间生成建模”, *Proc. NeurIPS*, 第 34 卷, 2021 年。[107] J. Wang, Z. Yue, S. Zhou, K. C. Chan 和 C. Change Loy, “利用扩散先验实现现实世界图像超分辨率”, 2023 年, *arXiv:2305.07015*。[108] Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund 和 T. B. Schön, “利用均值回归随机微分方程进行图像恢复”, 2023 年, *arXiv:2301.11699*。

- [109] L. Chen, X. Chu, X. Zhang and J. Sun, “图像恢复的简单基线”, 载于 *Proc. ECCV*. 瑞士 Cham: Springer, 2022 年. [110] Z. Chen 等人, “用于逼真图像去模糊的分层积分扩散模型”, 2023 年, *arXiv:2305.12966*. [111] B.B. Moser, S.Frolov, F.Raue, S.Palacio 和 A. Dengel, “DWA: 用于图像超分辨率的差分小波放大器”, 载于 *Artificial Neural Networks and Machine Learning—ICANN 2023*, L.Iliadis, A.Papaleonidas, P.Angelov and C.Jayne 编, 瑞士 Cham: Springer, 2023 年. [112] T.Guo, H.S.Mousavi, T.H.Vu 和 V.Monga, “用于图像超分辨率的深度小波预测”, 载于 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017 年 7 月, 第 1100-1109 页. [113] B. B. Moser, S. Frolov, F. Raue, S. Palacio 和 A. Dengel, “向低分辨率挥手告别: 用于图像超分辨率的扩散小波方法”, 载于 *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2024 年 7 月, 第 1-8 页. [114] Y. Huang 等人, “WaveDM: 基于小波的图像恢复扩散模型”, 2023 年, *arXiv:2305.13819*. [115] F. Guth, S. Coste, V. De Bortoli 和 S. Mallat, “基于小波得分的生成建模”, 载于 *Proc. NeurIPS*, 第 35 卷, 2022 年. [116] S. Shang 等人, “ResDiff: 结合 CNN 和扩散模型实现图像超分辨率”, 2023 年, *arXiv:2303.08714*. [117] J. Whang, M. Delbracio, H. Talebi, C. Saharia, A. G. Dimakis 和 P. Milanfar, “通过随机细化进行去模糊”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 16272-16282 页. [118] Z. Yue, J. Wang 和 C. C. Loy, “ResShift: 通过残差移位实现图像超分辨率的有效扩散模型”, 载于 *Proc. Adv. Neural Inf. Process. Syst.*, 2024 年. [119] C. Saharia 等人, “具有深度语言理解的逼真的文本到图像扩散模型”, 载于 *Proc. NeurIPS*, 第 35 卷, 2022 年. [120] J. Choi, S. Kim, Y. Jeong, Y. Gwon 和 S. Yoon, “ILVR: 去噪扩散概率模型的条件方法”, 2021 年, *arXiv:2108.02938*. [121] A. Niu 等人, “CDPMSR: 用于单图像超分辨率的条件扩散概率模型”, 2023 年, *arXiv:2302.12831*. [122] B. Lim, S. Son, H. Kim, S. Nah 和 K. M. Lee, “用于单图像超分辨率的增强深度残差网络”, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017 年 7 月, 第 1132-1140 页. [123] S. H. Park, Y. S. Moon 和 N. I. Cho, “使用条件目标的灵活风格图像超分辨率”, *IEEE Access*, 第 10 卷, 第 9774-9792 页, 2022 年. [124] W. Zhang, Y. Liu, C. Dong 和 Y. Qiao, “RankSRGAN: 带有排序器的生成对抗网络, 用于图像超分辨率”, 载于 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019 年 10 月, 第 3096-3105 页. [125] S. Menon, A. Damian, S. Hu, N. Ravi 和 C. Rudin, “PULSE: 通过生成模型的潜在空间探索进行自监督照片上采样”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020 年 6 月, 第 2434-2442 页. [126] Y. Chen, Y. Tai, X. Liu, C. Shen 和 J. Yang, “FSRNet: 基于面部先验的端到端学习人脸超分辨率”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018 年 6 月, 第 2492-2501 页. [127] K. Pandey, A. Mukherjee, P. Rai 和 A. Kumar, “DiffuseVAE: 从低维潜在数据高效、可控、高保真生成”, 2022 年, *arXiv:2201.00308*. [128] C. Bi, X. Luo, S. Shen, M. Zhang, H. Yue 和 J. Yang, “DeeDSR: 通过感知退化的稳定扩散实现真实世界图像超分辨率”, 2024 年, *arXiv:2404.00661*. [129] S. Zhou, K. Chan, C. Li 和 C. C. Loy, “通过码本查找变换器实现稳健的盲人脸修复”, 载于 *Proc. NeurIPS*, 第 35 卷, 2022 年. [130] T. Yang, R. Wu, P. Ren, X. Xie 和 L. Zhang, “基于像素感知的稳定扩散以实现逼真的图像超分辨率和个性化风格化”, 2023 年, *arXiv:2308.14469*. [131] R. Wu, T. Yang, L. Sun, Z. Zhang, S. Li 和 L. Zhang, “SeeSR: 实现语义感知的真实世界图像超分辨率”, 2023 年, *arXiv:2311.16518*. [132] Y. Qu 等人, “XPSR: 基于扩散的图像超分辨率的跨模态先验”, 2024 年, *arXiv:2403.05049*. [133] J. Liang, H. Zeng 和 L. Zhang, “细节还是伪影: 一种实现逼真图像超分辨率的局部判别学习方法”, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 5647-5656 页. [134] C. Chen 等人, “通过隐式高分辨率先验的特征匹配实现现实世界的盲超分辨率”, 载于 *Proc. 30th ACM Int. Conf. Multimedia*. 美国纽约州纽约: ACM, 2022 年 10 月. [135] G. Daras, M. Delbracio, H. Talebi, A. G. Dimakis 和 P. Milanfar, “软扩散: 针对一般损坏的分数匹配”, 2022 年, *arXiv:2209.05442*. [136] A. Bansal 等人, “冷扩散: 无噪声反转任意图像变换”, 2022 年, *arXiv:2208.09392*. [137] G.-H. Liu, A. Vahdat, D.-A. Huang, E. A. Theodorou, W. Nie 和 A. Anandkumar, “I<sup>2</sup>SB: 图像到图像薛定谔桥”, 2023 年, *arXiv:2302.05872*. [138] M. Delbracio 和 P. Milanfar, “直接迭代反演: 图像恢复去噪扩散的替代方法”, 2023 年, *arXiv:2303.11435*. [139] J. Choi, J. Lee, C. Shin, S. Kim, H. Kim 和 S. Yoon, “感知优先训练扩散模型”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 11462-11471 页. [140] B. B. Moser, S. Frolov, F. Raue, S. Palacio 和 A. Dengel, “动态注意引导扩散用于图像超分辨率”, 2023 年, *arXiv:2308.07977*. [141] M. Caron 等人, “自监督视觉 Transformers 中的新兴特性”, 载于 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021 年 10 月, 第 9630-9640 页. [142] B. Xia 等人, “DiffIR: 用于图像恢复的有效扩散模型”, 2023 年, *arXiv:2303.09472*. [143] A. Vaswani 等人, “你所需要的就是注意力”, 载于 *Proc. NeurIPS*, 第 30 卷, 2017 年. [144] A. Dosovitskiy 等人, “一张图片胜过 16×16 个词: 用于大规模图像识别的 Transformers”, 2020 年, *arXiv:2010.11929*. [145] B. K. War, G. Vaksman 和 M. Elad, “SNIPS: 随机解决嘈杂的逆问题”, 载于 *Proc. NeurIPS*, 第 30 卷, 2017 年. 34, 2021 年. [146] B. K. War, M. Elad, S. Ermon 和 J. Song, “去噪扩散恢复模型”, 载于 *Proc. NeurIPS*, 第 35 卷, 2022 年. [147] H. Chung, J. Kim, M. T. McCan, M. L. Klasky 和 J. Chul Ye, “一般噪声逆问题的扩散后验采样”, 2022 年, *arXiv:2209.14687*. [148] Y. Wang, J. Yu 和 J. Zhang, “使用去噪扩散零空间模型进行零样本图像恢复”, 2022 年, *arXiv:2212.00490*. [149] B. Fei 等人, “用于统一图像恢复和增强的生成扩散先验”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023 年 6 月, 第 9935-9946 页. [150] A. Shocher, N. Cohen 和 M. Irani, “使用深度内部学习实现零样本超分辨率”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018 年 6 月, 第 3118-3126 页. [151] R. Li, X. Sheng, W. Li 和 J. Zhang, “OmniSSR: 使用稳定扩散模型实现零样本全向图像超分辨率”, 2024 年, *arXiv:2404.10312*. [152] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte 和 L. Van Gool, “RePaint: 使用去噪扩散概率模型进行修复”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 11451-11461 页. [153] H. Chung, B. Sim 和 J. C. Ye, “更接近扩散更快: 通过随机收缩加速逆问题的条件扩散模型”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 12403-12412 页. [154] J. Schwab, S. Antholzer 和 M. Haltmeier, “逆问题的深度零空间学习: 收敛分析和速率”, *Inverse Problems*, 第 35 卷, 第 2 期, 2019 年 2 月, 文章编号 025008. [155] Y. Wang, Y. Hu, J. Yu 和 J. Zhang, “基于 GAN 先验的零空间学习以实现一致的超分辨率”, 载于 *Proc. AAAI*, 2023 年, 第 37 卷, 第 3. [156] H. Chung, B. Sim, D. Ryu 和 J. C. Ye, “使用流形约束改进逆问题的扩散模型”, 载于 *Proc. NeurIPS*, 第 35 卷, 2022 年. [157] J. Song, A. Vahdat, M. Mardani 和 J. Kautz, “逆问题的伪逆引导扩散模型”, 载于 *Proc. ICLR*, 2022 年. [158] J. Lin 等人, “空间变异核细化与扩散模型的自适应多模态融合用于盲图像超分辨率”, 2024 年, *arXiv:2403.05808*. [159] H. Chung, E. S. Lee 和 J. C. Ye, “使用正则化逆扩散进行 MR 图像去噪和超分辨率”, *IEEE Trans. Med. Imag.*, 第 42 卷, 第 4 期, 第 922-934 页, 2023 年 4 月. [160] Y. Mao, L. Jiang, X. Chen 和 C. Li, “DisC-diff: 用于多对比 MRI 超分辨率的解缠结条件扩散模型”, 2023 年, *arXiv:2303.13933*.

[161] G. Li, C. Rao, J. Mo, Z. Zhang, W. Xing 和 L. Zhao, “重新思考多对比度 MRI 超分辨率的扩散模型”, 2024 年, *arXiv:2404.04785*. [162] Z. Yue 和 C. Change Loy, “DiffFace: 通过扩散误差收缩进行盲人脸修复”, 2022 年, *arXiv:2212.06512*. [163] X. Wang, Y. Li, H. Zhang 和 Y. Shan, “通过生成面部先验实现现实世界的盲人脸修复”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021 年 6 月, 第 9164-9174 页. [164] T. Yang, P. Ren, X. Xie 和 L. Zhang, “用于野外盲人脸修复的 GAN 先验嵌入网络”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021 年 6 月, 第 672-681 页. [165] X. Qiu, C. Han, Z. Zhang, B. Li, T. Guo 和 X. Nie, “DiffBFR: 面向盲人脸修复的引导扩散模型”, 2023 年, *arXiv:2305.04517*. [166] Z. Wang 等人, “DR2: 基于扩散的盲人脸修复鲁棒退化去除器”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023 年 6 月, 第 1704-1713 页. [167] X. Wang, S. López-Tapia 和 A. K. Katsaggelos, “通过变分深度扩散进行大气湍流校正”, 载于 *Proc. IEEE 6th Int. Conf. Multimedia Inf. Process. Retr. (MIPR)*, 2023 年 9 月, 第 1-4 页. [168] N. G. Nair, K. Mei 和 V. M. Patel, “AT-DDPM: 使用去噪扩散概率模型恢复因大气湍流而退化的人脸”, *Proc. WACV*, 2023 年. [169] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin 和 L. Zhang, “EDiffSR: 一种用于遥感图像超分辨率的有效扩散概率模型”, *IEEE Trans. Geosci. Remote Sens.*, 第 62 卷, 2024 年. [170] J. Li u, Z. Yuan, Z. Pan, Y. Fu, L. Liu 和 B. Lu, “带细节补充的遥感超分辨率扩散模型”, *Remote Sens.*, 第 14 卷, 第 19 期, 第 4834, 2022 年 9 月. [171] A. M. Ali, B. Benjdira, A. Koubaa, W. Boulila 和 W. El-Shafai, “TESR: 遥感图像增强和超分辨率的两阶段方法”, *Remote Sens.*, 第 15 卷, 第 9 期, 第 2346 页, 2023 年 4 月. [172] M. Xu, J. Ma 和 Y. Zhu, “双扩散: 用于 RSI 中盲超分辨率重建的双条件去噪扩散概率模型”, 2023 年, *arXiv:2305.12170*. [173] S. Khanna 等人, “DiffusionSat: 卫星图像的生成基础模型”, 载于 *Proc. ICLR*, 2024 年. [174] D. Ganguli 等人, “大型生成模型中的可预测性和意外性”, 载于 *Proc. ACM Conf. Fairness, Accountability, Transparency*, 2022 年 6 月. [175] N. Chen, Y. Zhang, H. Zen, R. J. Weiss, M. Norouzi 和 W. Chan, “WaveGrad: 估计波形生成的梯度”, 2020 年, *arXiv:2009.00713*. [176] Z. Cheng, “扩散模型的采样器调度程序”, 2023 年, *arXiv:2311.06845*. [177] T. Chen, “论噪声调度对扩散模型的重要性”, 2023 年, *arXiv:2301.10972*. [178] R. Zhang, P. Isola, A. A. Efros, E. Shechtman 和 O. Wang, “深度特征作为感知指标的不合理有效性”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018 年 6 月, 第 586-595 页. [179] X. Liu, J. Van De Weijer 和 A. D. Bagdanov, “RankIQ: 从排名中学习以实现无参考图像质量评估”, 载于 *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017 年 10 月, 第 1040-1049 页. [180] K. Ma, W. Liu, T. Liu, Z. Wang 和 D. Tao, “DipIQ: 通过学习对可区分图像对进行排名的盲图像质量评估”, *IEEE Trans. Image Process.*, 第 26 卷, 第 8, 第 3951-3964 页, 2017 年 8 月. [181] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo 和 J. Choe, “CutMix: 用于训练具有可本地化特征的强分类器的正则化策略”, 载于 *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019 年 10 月, 第 6022-6031 页. [182] B. Jing, G. Corso, R. Berlinghieri 和 T. Jaakkola, “子空间扩散生成模型”, 载于 *Proc. ECCV*. Cham, 瑞士: Springer, 2022 年. [183] M. Kwon, J. Jeong 和 Y. Uh, “扩散模型已经具有语义潜在空间”, 载于 *Proc. ICLR*, 2023 年. [184] Q. Wu 等人, “揭示文本到图像扩散模型中的解缠能力”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023 年 6 月, 第 1900-1910 页. [185] G. Kim, T. Kwon 和 J. C. Ye, “DiffusionCLIP: 用于稳健图像处理的文本引导扩散模型”, 载于 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022 年 6 月, 第 2416-2425 页.



Brian B. Moser 于 2021 年获得德国凯泽斯劳滕工业大学计算机科学硕士学位, 目前正在攻读博士学位。

他还是德国凯泽斯劳滕德国人工智能研究中心 (DFKI) 的研究助理。他的研究兴趣包括图像超分辨率和深度学习。



Arundhati S. Shanbhag 目前正在德国凯泽斯劳滕工业大学攻读硕士学位。她还是德国凯泽斯劳滕人工智能研究中心 (DFKI) 的研究助理。她的研究兴趣包括计算机视觉和深度学习。



费德里科·劳厄 (Federico Raue) 获得硕士学位 2005 年在比利时鲁汶天主教鲁汶大学获得人工智能学士学位, 并于 2005 年在比利时鲁汶大学获得博士学位。他目前是德国凯泽斯劳滕德国人工智能研究中心 (DFKI) 的高级研究员。他的研究兴趣包括元学习和多模态机器学习。



Stanislav Frolov 于 2017 年获得德国卡尔斯鲁厄理工学院电气工程硕士学位。他目前正在德国凯撒斯劳滕工业大学攻读博士学位。

他还是德国凯泽斯劳滕德国人工智能研究中心 (DFKI) 的研究助理。他的研究兴趣包括生成模型和深度学习。



Sebastian Palacio 目前是德国人工智能研究中心 (DFKI) 的机器学习研究员和多媒体分析与数据挖掘小组负责人, 该中心位于德国凯泽斯劳滕。他的博士论文主题是可解释的人工智能及其在计算机视觉中的应用。他的研究兴趣包括对抗性攻击、多任务、课程和自我监督学习。



Andreas Dengel 目前是德国凯泽斯劳滕工业大学计算机科学系的教授。他还是德国凯泽斯劳滕人工智能研究中心 (DFKI) 的执行主任, 担任智能数据和知识服务研究领域和 DFKI 深度学习能力中心的负责人。他的研究重点是机器学习、模式识别、量化学习、数据挖掘、语义技术和文档分析。