采用扩散模型的图像融合研究综述

高志荣[a],熊承义[b]
(中南民族大学 a.计算机学院(人工智能学院); b.电子信息工程学院,湖北 武汉 430074)

摘 要 随着计算机视觉技术的不断进步,图像融合作为一种有效的信息获取和优化方式受到了广泛关注.近年来,扩散模型因其在生成与重建方面的突出优势,已成为图像融合领域的重点研究方向.对利用扩散模型进行图像融合的最新研究进展进行了系统综述.首先,详细阐述了扩散模型的基本原理,包括去噪扩散概率模型、噪声条件评分网络以及基于随机微分方程的扩散模型;介绍了前向扩散、反向过程及训练目标等关键要素,并剖析了该模型在图像融合中的应用潜力.接着,依据模型架构、训练策略和实际应用场景,深入分析了现有基于扩散模型的图像融合方法,着重比较了各类方法的特性与创新之处.同时,讨论了相关的融合数据集及评估指标,开展了对多种算法的定性和定量评估比较实验,清晰地展现了不同算法的优势与不足.最后,展望了未来的发展方向,提出应从提升融合性能、优化模型训练与应用策略入手,加强跨领域合作,以促进扩散模型在图像融合技术中的进一步应用与发展.

关键词 图像融合;扩散模型;综述
中图分类号 TP391.4;TN911.73 **文献标志码** A **文章编号** 1672-4321(2026)05-0658-16
doi:10.20056/j.cnki.ZNMDZK.20250855

Image fusion based on Diffusion Models : A review

Gao Zhirong[a], Xiong Chengyi[b]
(South-Central Minzu University, a. College of Computer Science(College of Artificial Intelligence); b. College of Electronic and Information Engineering, Wuhan 430074, Hubei China)

Abstract With the continuous advancement of computer vision technology, image fusion, as an effective way to obtain and optimize information, has received extensive attention. In recent years, diffusion models have become a key research direction in the field of image fusion due to their outstanding advantages in generation and reconstruction. This paper provides a systematic review of the latest research progress in using diffusion models for image fusion. Firstly, the basic principles of diffusion models are elaborated in detail, including the denoising diffusion probabilistic model, the noise-conditioned score network, and the diffusion model based on stochastic differential equations. Key elements such as the forward diffusion process, the reverse process, and the training objectives are introduced, and the application potential of this model in image fusion is analyzed. Secondly, based on the model architecture, training strategies, and practical application scenarios, the existing image fusion methods based on diffusion models are analyzed in depth, with a focus on comparing the characteristics and innovative aspects of various methods. At the same time, relevant fusion datasets and evaluation indicators are discussed, and qualitative and quantitative evaluation comparison experiments of multiple algorithms are carried out, clearly demonstrating the advantages and disadvantages of different algorithms. Finally, the future development directions are prospected. It is proposed that efforts should be made to improve the fusion performance, optimize the model training and application strategies, and strengthen cross-disciplinary cooperation to promote the further application and development of diffusion models in image fusion technology.

Keywords image fusion; diffusion model; review

收稿日期 2025-01-10
作者简介 高志荣(1972-),女,副教授,博士,研究方向:机器学习,E-mail: gaozhirong@mail.scuec.edu.cn
基金项目 多谱信息处理技术国家重点实验室基金资助项目(6142113210303);中央高校基本科研业务专项资金资助项目(CZY25010)

在当今数字化时代，图像作为一种重要的信息载体，广泛应用于众多领域．然而，图像在获取、传输和处理过程中，常常受到噪声干扰、模糊、分辨率降低、区域信息不完整及条件降质等问题的困扰，致使图像质量下降，信息传递受阻．图像融合[1]旨在通过算法手段，将来自不同传感器、不同时间或不同视角的多个图像进行集成，进而生成一个包含更丰富、更准确信息的融合图像．这一技术能够充分挖掘各图像的潜在价值，实现信息的互补与增强，从而显著提高图像信息的全面性、可靠性和可用性，为后续的图像分析、处理与理解提供坚实基础．

传统图像融合方法主要依赖数学模型和先验知识，如基于像素的融合[2-3]、基于特征的融合[4-5]、基于决策级的融合[6]等．尽管这些方法在一定程度上能够改善融合图像的质量，但它们往往受限于图像间的复杂关系和退化程度，难以达到理想的融合效果．随着深度学习技术的迅猛发展，卷积神经网络（Convolutional Neural Network，CNN）及其变体在图像特征提取和表示学习方面展现出卓越性能，为图像融合带来了新的机遇与突破．通过深度特征融合策略[7-9]，CNN 能够学习到不同图像间的深层关联，从而生成更加准确的融合图像．此外，残差网络[10-11]（Residual Network，ResNet）和生成对抗网络[12-14]（Generative Adversarial Network，GAN）等深度学习模型也为图像融合提供了全新的视角与解决方案．它们通过引入残差连接和对抗训练机制，能够学习到更加复杂和精细的图像特征，进一步提升融合图像的质量．

然而，深度学习方法在图像融合中也面临诸多挑战．首先，模型训练需要大量的标注数据，尤其是高质量的融合图像作为参考，而这在实际应用中往往难以获取．其次，深度学习模型的结构设计和训练过程相对复杂，特别是在处理多源异构图像数据时，模型的泛化能力和鲁棒性受到较大限制．此外，深度学习模型的可解释性较差，难以直观理解模型融合图像的内在机制和决策过程，这对于安全性和可靠性要求较高的应用领域而言，无疑是一个潜在的风险．扩散模型[15-16]（Diffusion Model，DM）作为一种新兴的生成模型，最近在图像生成领域崭露头角，展现出强大的应用潜力．其核心思想是通过模拟一个扩散过程，将数据分布逐渐转变为一个已知的简单分布（如高斯分布），然后再学习逆扩散过程从这个简单分布中采样生成数据．这种独特的生成方式使得扩散模型能够捕捉到数据的复杂分布特征和内在结构关系，为图像融合提供了全新的可能性．

近年来，国内外研究人员都开始积极探索将扩散模型应用于图像融合任务，并取得了一系列重要进展．扩散模型能够生成更加自然、真实且符合视觉感知的融合图像．与传统方法和其他深度学习方法相比，扩散模型在图像融合中展现出独特的优势，为图像融合技术的发展带来了新的契机和思路．本文旨在深入探究扩散模型在图像融合中的应用现状，力求全面、系统地呈现该领域的研究全貌，同时指出该领域研究存在的挑战与未来趋势，以有力促进该领域研究的未来发展．论文的主要贡献如下：

（1）全面、系统地综述了扩散模型在图像融合中的研究进展状况，从模型基础、方法分类、数据集与评估指标等多维度展开，构建了完整的知识体系．

（2）详细剖析了各模型的原理、创新思路及应用案例，涵盖医学、遥感、红外与可见光图像融合等多个领域，为相关研究提供了坚实的理论基础．

（3）针对不同的扩散模型框架，深入了比较它们在图像融合任务中的优缺点．通过对比，清晰呈现了各模型的特性，为实际应用提供了有效指导．

1　扩散模型基础理论

在生成模型领域，扩散概率模型（即扩散模型）带来了革新性变化，它通过马尔可夫链建模将复杂不稳定的生成过程转化为多个独立稳定的逆向过程．有三种通用的扩散模型框架被广泛使用，包括去噪扩散概率模型[17-19]（Denoising Diffusion Probability Model，DDPM）、噪声条件评分网络[20]（Noise Condition Scoring Network，NCSN）和基于随机微分方程[21]（Stochastic Differential Equation，SDE）的扩散模型．

1.1　去噪扩散概率模型

去噪扩散概率模型作为扩散模型的重要基石，是一种基于马尔可夫链的潜在变量模型，主要由前向过程和反向过程组成，如图 1 所示．

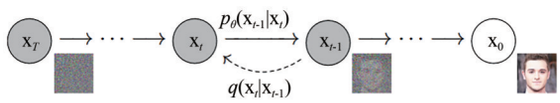


图 1　前向过程和反向过程

Fig. 1　Forward process and reverse process

1.1.1 前向过程

前向过程也被称为扩散过程或加噪过程,对应图1中从右往左,从真实数据x_0出发,通过逐步向其添加高斯噪声将其转变为一个高斯分布的噪声版本序列$x_{1:T}$.这个过程可以用一个联合分布$q(x_{1:T}|x_0)=\prod_{t=1}^{T}q(x_t|x_{t-1})$来表征,其中的转移核$q(x_t|x_{t-1})$是一个精心设计的高斯分布,即

$$q(x_t|x_{t-1})=N(x_t;\ \sqrt{1-\beta_t}\,x_{t-1},\beta_t I)\ , \qquad (1)$$

其中,x_t是第t步的噪声数据;x_{t-1}是第$t-1$步的噪声数据;β_t是第t步的噪声增加量,通常称为噪声调度,是一个在(0,1)区间内的固定值;I是单位矩阵;$N(x_t;\ \sqrt{1-\beta_t}\,x_{t-1},\beta_t I)$表示均值为$\sqrt{1-\beta_t}\,x_{t-1}$,协方差为$\beta_t I$的高斯分布.通过迭代地应用该公式,数据$x_0$会逐步被加入噪声,直到变成纯高斯噪声$x_T$.

为了简化计算,可以利用高斯分布的性质,将每一步的扩散过程简化为从x_0到x_t的一步计算,即:

$$q(x_t|x_0)=N(x_t;\ \sqrt{\overline{\alpha}_t}\,x_0,(1-\overline{\alpha}_t)I)\ , \qquad (2)$$

其中,$\alpha_t=1-\beta_t$,$\overline{\alpha}_t=\prod_{i=1}^{t}\alpha_i$,这意味着可以直接从原始数据$x_0$和时间步$t$计算出对应的噪声样本$x_t$,而不需要显式地模拟每个中间步骤.在实际应用中,通常设置$\beta_1<\beta_2<\cdots<\beta_T$,由于$\alpha_t=1-\beta_t$,致使$\alpha_1>\alpha_2>\ldots>\alpha_T\approx 0$,最终使得终端分布$q(x_T)$近似为:$q(x_T)\approx \mathrm{N}(x_T;\ 0,I)$,即经过$T$步前向扩散,数据已近乎化为纯噪声.

1.1.2 反向过程

反向过程也称为采样过程或去噪过程,对应图1中从左往右,是从噪声样本x_T开始,逐步去除噪声以恢复到原始数据x_0的过程.这个过程通过训练神经网络模型来近似逆向的条件概率分布.反向过程可以公式化描述为:

$$p_\theta(x_{t-1}|x_t)=N(x_{t-1};\ \mu_\theta(x_t,t),\ \textstyle\sum_\theta(x_t,t))\ , \qquad (3)$$

其中,$p_\theta(x_{t-1}|x_t)$是模型学习到的条件概率分布,表示在给定x_t的情况下生成x_{t-1}的概率;$\mu_\theta(x_t,t)$和$\sum_\theta(x_t,t)$是神经网络学习到的均值和协方差参数;θ是网络参数.为了简化模型训练,通常假设协方差矩阵$\sum_\theta(x_t,t)$是对角的或者直接设为固定值,重点在于学习逆向过程的均值$\mu_\theta(x_t,t)$.

在测试阶段,DDPM 通过逆向扩散过程从高斯噪声开始生成数据.具体步骤如下:

(1) 从高斯噪声$x_T\sim N(0,I)$开始;

(2) 使用训练好的模型$p_\theta(x_{t-1}|x_t)$来逐步生成$x_{t-1},x_{t-2},\cdots,x_0$;

(3) 最终得到的x_0就是生成的数据样本.

1.1.3 训练目标

DDPM的训练目标是最小化从正向过程到逆向过程的分布之间的差异.这个差异通常通过变分下界(Variational Lower Bound, VLB)来表示,具体形式为:

$$L=E_q[\textstyle\sum_{t=1}^{T}D_{KL}(q(x_{t-1}|x_t,x_0)\|p_\theta(x_{t-1}|x_t))]\ , \qquad (4)$$

其中,D_{KL}是 Kullback-Leibler散度,用于衡量真实后验分布$q(x_{t-1}|x_t,x_0)$和模型生成的分布$p_\theta(x_{t-1}|x_t)$之间的差异.

为了优化这个目标函数,DDPM作者提出了一种简化的训练目标,它基于去噪自编码器(Denoising Autoencoder, DA)的思想,将去噪过程的损失直接定义为预测噪声的误差.具体来说,模型会学习预测添加到数据中的噪声ε,并通过均方误差(Mean Square Error,MSE)来衡量预测噪声与真实噪声之间的差距:

$$L_{\text{simple}}=E_{x_0,\varepsilon,t}[\|\varepsilon-\varepsilon_\theta(x_t,t)\|^2]\ , \qquad (5)$$

其中,ε是正向扩散过程中添加的噪声;$\varepsilon_\theta(x_t,t)$是神经网络模型预测的噪声.

这个简化的目标函数非常直观,它训练模型去预测噪声,从而帮助模型更好地在逆向扩散过程中去噪.通过优化这个目标函数,DDPM能够学习到如何从噪声图像中生成逼真的原始图像.

1.2 噪声条件评分网络

尽管DDPM在扩散模型领域取得了重要突破,但在处理实际数据时,由于数据往往存在于低维流形上,DDPM在分数估计时存在不稳定现象,导致模型在恢复原始数据分布时出现偏差,影响生成样本的质量和多样性.为了克服这一局限性,噪声条件评分网络(NCSN)应运而生.NCSN引入了一种全新的思路,它不再局限于DDPM中单纯的噪声添加与预测机制,而是聚焦于通过训练一个共享神经网络,利用分数匹配技术来精确地估计不同噪声水平下扰动数据分布的分数函数.

在NCSN中,数据点x_0被高斯噪声分布$q(x_t|x_0)=N(x_t;\ x_0,\sigma_t^2 I)$扰动为$x_t$,生成一系列噪声数据密度.从一个随机样本$x_0$向数据高密度区域的样本$x_N$移动,其迭代更新公式为:

$$x_i=x_{i-1}+\frac{\gamma}{2}\nabla_x\log p(x)+\sqrt{\gamma}\,\omega_i\ , \qquad (6)$$

其中$i\in\{1,\cdots,N\}$,γ控制在分数方向上更新的幅度,x_0从先验分布采样,噪声$\omega_i\sim N(0,I)$用于避免

陷入局部最小值，该方法递归应用$N \to \infty$步.因此，一个生成模型可以在通过神经网络$s_\theta(x) \approx \nabla_x \log p(x)$估计分数后，使用上述方法从$p(x)$中采样.这个网络可以通过分数匹配进行训练，其目标是优化以下目标函数：

$$L_{\text{sm}} = E_{x \sim p(x)} \| s_\theta(x) - \nabla_x \log p(x) \|_2^2 . \tag{7}$$

训练过程可以利用分数匹配、去噪分数匹配和切片分数匹配等技术，从被扰动的数据点中学习.在样本生成阶段，NCSN 利用诸如淬火朗之万动力学（ALD）等迭代方法，依次从$\theta(x,t), \theta(x,t-1), \cdots, \theta(x,0)$生成样本.通过学习数据分布的梯度场，NCSN 能够在生成过程中实现更灵活的控制，以生成具有特定属性的样本.

1.3　基于随机微分方程的扩散模型

DDPM 模型存在采样效率偏低，生成样本时间长，难以满足时间敏感任务的需求；此外，在训练和采样的特定阶段（如$t = 0$时），数据方差趋近于 0 会引发数值不稳定现象，影响模型收敛效果和采样质量.基于随机微分方程（SDE）的扩散模型旨在通过构建连续扩散过程，实现更高效采样和更稳定训练，为生成模型带来新的突破与提升，其主要包括前向 SDE 数据扰动和反向时间 SDE 采样两个关键过程.

1.3.1　前向 SDE 数据扰动

在基于随机微分方程的扩散模型中，前向 SDE 用于数据扰动.这一过程通过向数据中逐步注入高斯噪声，将数据分布平滑地转换为已知的先验分布（如标准高斯分布）.前向 SDE 的数学形式可以表示为：

$$\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}w , \tag{8}$$

其中，$\mathrm{d}x$表示数据在时间步长dt内的变化量；$f(x,t)$是漂移项（drift term），描述了系统在没有随机扰动时的确定性变化；$g(t)$是扩散项（diffusion term），描述了系统因随机扰动产生的变化，即随机源；dw是一个维纳过程（Wiener process）或布朗运动的增量，表示随机噪声.

在实际应用中，前向 SDE 通过一系列的时间步长dt逐步将原始数据x_0转换为噪声数据x_T.这一过程可以看作是一个加噪过程，其中噪声的强度和类型由漂移项$f(x,t)$和扩散项$g(t)$决定.通过精心设计$f(x,t)$和$g(t)$，前向 SDE 可以将任何复杂的数据分布转换为易于处理的先验分布.

1.3.2　反向时间 SDE 采样

与前向 SDE 相对应，反向时间 SDE 用于从先验分布中采样，并逐步去除噪声，以恢复原始数据分布.反向时间 SDE 的数学形式可以表示为：

$$\mathrm{d}x = [f(x,t) - g(t)^2 \nabla_x \log p_t(x)]\mathrm{d}t + g(t)\mathrm{d}w , \tag{9}$$

其中，$\nabla_x \log p_t(x)$是得分函数（score function），表示对数数据分布梯度，即数据分布的增长率最大的方向.在反向时间 SDE 采样过程中，首先从先验分布中采样一个随机向量x_T，然后通过反向 SDE 逐步去除噪声，以生成新的数据样本x_0.这一过程依赖于得分函数的准确估计，通常通过训练深度神经网络来实现.

反向时间 SDE 采样过程可以看作是一个去噪过程，其中得分函数指导着噪声的去除方向.通过迭代地应用反向 SDE，可以逐步从噪声数据中恢复出原始数据分布.这一过程不仅实现了从先验分布到数据分布的转换，还提供了对不确定性进行量化的能力，这在标准回归设置中是不可能的.

1.3.3　用方差保持 SDE 解释 DDPM

在 SDE 的框架下，数据扰动过程可以建模为一个随机过程，其中数据的变化量由漂移项和扩散项共同决定.方差保持 SDE 是一种特殊的 SDE，它确保在扰动过程中，数据的方差保持在一个恒定的水平.这对于生成模型来说非常重要，因为它可以确保在生成过程中，生成的样本与原始数据具有相似的统计特性.

在 DDPM 中，数据通过一系列的时间步长逐渐添加噪声，这一过程可以建模为一个离散化的 SDE.具体来说，每一步的噪声添加可以表示为：

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \varepsilon_t , \tag{10}$$

其中，x_t是第t步的噪声数据；x_{t-1}是第$t-1$步的数据；α_t是一个衰减系数；ε_t是从标准正态分布中采样的噪声.

在 DDPM 中，通过精心设计参数α_t和β_t（其中$\alpha_t = 1 - \beta_t$）可维持方差不变，即当α_t和β_t满足特定条件时，每步添加噪声过程中的数据方差可保持恒定.方差保持 SDE 的作用十分关键，能确保生成样本与原始数据统计特性相似，从而稳定生成过程，减少不确定性，生成高质量样本；另外，因其保障了生成的稳定性，DDPM 可高效训练模型参数，加快训练速度并提升泛化能力，还能通过调整相关参数的值灵活控制噪声的添加强度和速度，以适应不同数据集和任务，实现更优生成效果.

2 基于扩散模型的图像融合方法分类

在本节中,对基于扩散模型的图像融合方法进行了概述.图2中定义了各种方法的分类并以图形方式对这些方法进行了说明.具体地,考虑到模型框架、实现方式与应用场景,基于扩散模型的图像融合方法可以分为以下几个主要类别(而每个类别可以进一步分为子类别,这将在本节后面讨论):

(1) 扩散模型框架:扩散模型框架构成图像融合技术的核心部分,其中融合算法的设计与应用至关重要,影响着图像融合效率与效果,为理解和优化提供关键切入点,成为推动图像融合技术发展的关键要素.

(2) 训练方式与模型构建:在融合算法基础上,训练方式与模型构建影响着图像融合技术的实现难度和成本,直接关系到融合效果的质量和稳定性,了解并掌握多种实现方式,对于图像融合技术的研发和应用具有重要意义.

(3) 应用场景:从医学诊断到遥感监测,从智能安防到日常生活,图像融合技术正在各个领域发挥着越来越重要的作用.这些应用场景展示了图像融合技术的广泛适用性,提供了不断推动技术发展的动力.

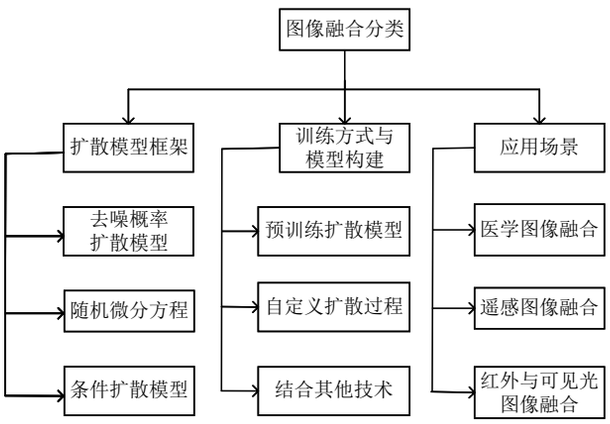


图2 基于扩散模型的图像融合方法分类

Fig. 2 Classification of image fusion methods based on diffusion models

作为对上述分类的进一步补充与阐释,表1以简洁明了的方式总结了最为关键的方法,同时清晰地标注了它们所属的类别及子类别,提供了一个全面而系统的概览.

表1 基于扩散模型的图像融合方法分类

Tab. 1 Classification of image fusion methods based on diffusion models

类别	子类	方法
扩散模型框架	去噪扩散概率模型	DDPM[15], FusionDiff[22], MFIF[23],AMND[24],TDMI[25], DDFM[26], MIF[27], Videlfusion[28], DifFUSER[29], DiffRecon[30]
	随机微分方程	SDE-NDE[31],VDM[32],PDE[33]
	条件扩散模型	STMMIF[34],DDPM-GAN[35], PanDiff[36], Diff-IF[37], Text-DiFuse[38],CCIF[39], CDP-SRFusion[40],HPIF[41], STFDiff[42]
训练方式与模型构建	预训练扩散模型	FSDC[43], MultiDiffusion[44], DRMF[45],AMMF[46], CCIF[39]
	自定义扩散模型	multi-focus[47],IVIF[48], nearest-neighbor[49], Hyperspectral[50],DPMMs[51]
	与其他技术结合	AMND[24], VDM[32], STMMIF[34],2Dmis[52], LFDT-Fusion[53], DGFusion[54], ColorFusion[55], DCAFuse[56], end-to-end[57], GLAD[58], satellite images[59]
应用场景	医学图像融合	PDE[33], STMMIF[34], DCAFuse[56],3Dmig[60], Fdiff-Fusion[61], SAD-MIF[62], DT-CWT[63], contourlet[64], pixel[65]
	遥感图像融合	VDM[32], nearest-neighbor[49], Hyperspectral[50], satellite images[59], remote[66], Semantic[67], Despeckling[68], Gaofen-1[69], Spatiotemporal[70], multispectral[71]
	红外与可见光图像融合	DGFusion[54],DCAFuse[56], end-to-end[57], Dif-fusion[72]

2.1 基于扩散模型框架

扩散模型框架构建了一套系统且具灵活性的理论与技术体系,它针对图像信息的扩散与重构流程进行建模,能有效应对图像噪声、不确定性,并整合多模态信息(例如图3所示的多模态图像融合案例),为高质量、精准且多样的图像融合奠定了坚实基础.依据其核心原理又衍生出了许多不同类别,此类别中更细粒度和更详细的子类有:

(1) 去噪扩散概率模型

去噪扩散概率模型(DDPM)在图像融合领域展现出了强大的应用潜力,众多研究人员从不同角度对其进行了探索并取得了丰富成果.

单模态图像融合.单模态图像融合针对同一模态的多幅图像进行融合处理,以获取质量更高、信息更丰富的图像.在这一范畴内,多聚焦图像融合是较为典型的类型.值得注意的是,LI M[22]、LI X[23]



图3　DDFM[26]在不同数据集上融合示意图

Fig. 3　Schematic diagrams of the fusion process of DDFM[26] on different datasets

以及 SINGH[24]等研究均围绕去噪扩散概率模型而展开，旨在借助该模型探索多聚焦图像融合的实现路径与优化策略.Li[22]提出的“FusionDiff”设计了轻量级 U-Net 架构，Li[23]的方法针对水下微气泡场景，Singh[24]在小波域利用平均法噪声扩散（AMND）进行多聚焦图像融合，但它们均难以应对复杂场景的融合需求，在融合精度与环境适应性方面有待进一步优化.

多模态图像融合.多模态图像融合是对不同成像模态图像的有机结合.在医学领域，计算机断层扫描（Computed Tomography，CT）与磁共振成像（Magnetic Resonance Imaging，MRI）的融合、遥感领域的光学遥感图像与合成孔径雷达（Synthetic Aperture Radar，SAR）图像融合；安防监控领域的红外与可见光图像融合等.Dorjsembe[25]首次尝试运用 DDPM 实现 3D 医学图像合成；Zhao[26]提出的 DDFM 将多模态图像融合任务表述为 DDPM 采样框架下的条件生成问题；Yang[27]提出的方法将多模态融合任务置于 DDPM 采样框架内，在热红外与可见光图像融合、近红外与可见光图像融合以及医学图像融合等多模态任务中均有应用和评价，展示了 DDPM 在多种多模态图像融合场景中的有效性.

在视频生成中，Luo[28]提出的 Videofusion 设计分解扩散过程及相应去噪管道，采用两个联合学习网络匹配噪声分解，以应对高维数据挑战；Le[29]介绍的 DifFUSER 在传感器故障时细化或合成传感器特征；Sun[30]提出的 DiffRecon 采用生成模型合成术中 CT 以消除尺寸差异，配准模型对齐合成 CT，在图像质量和剂量指标评估中表现出色.

去噪扩散概率模型（DDPM）存在生成效率低、模型复杂等问题，更多的探索思路是与其他深度学习技术结合，进一步优化采样过程、简化模型结构及改进训练方法以提升性能.

（2）　随机微分方程

随机微分方程（SDE）在图像融合领域有着独特的应用，为解决多种图像融合相关问题提供了新的思路与方法.

医学图像在采集、传输、显示等过程中会受到不同程度的噪声污染，传统的降噪方法往往需要依靠先验信息或者模型假设，导致降噪效果受到限制，而现有扩散模型降噪效率较低.针对这些问题，肖继锋[31]等人提出基于随机微分方程的医学图像去噪方法.首先对初始医学图像进行预处理得到目标医学图像，再通过基于随机微分方程的扩散模型对目标医学图像分析和求解，得到降噪后的医学图像.该方法能够在实现减小医学图像噪声，提高医学图像质量的同时，还可以提高医学图像的处理效率.

遥感图像融合.Zhang 等[32]针对遥感图像融合中的光学图像与合成孔径雷达（SAR）融合这一核心任务，提出了变分扩散方法，利用了基于 SDE 的扩散模型（即扩散去噪概率模型及其连续形式）的强大生成能力.

通用图像融合.Pop 等[33]提出了一种基于偏微分方程（Partial Differential Equation，PDE）的图像融合通用方法，虽然不是直接基于 SDE，但与 SDE 相关联，因为 PDE 是 SDE 的一种特殊形式.将像素级融合和扩散过程相结合，通过反转扩散过程在融合图像中插入源图像包含的相关信息，并添加正则化项解决逆扩散过程的不稳定问题.

基于随机微分方程的模型存在计算复杂度高，参数估计难，噪声敏感等问题，后期可围绕模型简化、算法优化、自适应参数调整等路径展开.

（3）　条件扩散模型

通过引入条件信息，条件扩散模型能够更精准地满足不同图像融合任务的多样化需求，有效提升融合图像的质量与针对性.

医学图像融合.Xu 等[34]提出的 TFS-Diff，Müller-Franzes G 等[35]的 Medfusion，均借助条件扩散模型，采用通道注意力模块、去伪影等方式，实现了医学图像融合与超分辨率.然而基于条件的医学图像融合可能存在数据依赖强、训练不稳定等问题.

遥感图像融合.Meng 等[36]提出的 PanDiff 方法通过全色锐化方法来提升遥感图像质量与信息丰富

度.前向扩散过程中,在多光谱图像数据上逐步添加高斯噪声,使得图像信息逐渐被随机化;而反向降噪阶段,依据全色图像和低分辨率多光谱图像经模态互校正模块校准后的信息作为条件,利用U-Net模型逐步预测并去除噪声,从而重构出高分辨率多光谱(HRMS)图像,实现了全色图像与多光谱图像的有效融合.

通用图像融合.Yi等[37]的Diff-IF将条件扩散模型和融合先验转化为特定图像融合任务的先验分布,用于红外与可见光图像及医学图像融合;Zhang等[38]提出的Text-DiFuse通过用户自定义文本控制提高融合性能并突出前景对象;Cao等[39]提出的条件可控融合(Conditional Controlled Fusion,CCF)框架用于一般图像融合任务,动态选择合适条件;李小松等人[40]提出基于条件扩散概率模型的多模态图像融合与超分辨率方法,能够在分辨率相对较低的源图像基础上生成分辨率相对较高的融合图像,以恢复源图像中被退化的细节和纹理信息.但基于条件扩散模型的通用图像融合可能存在条件信息依赖性强、容易过拟合等问题.

双条件扩散模型在图像融合中应用显著.如Li等[41]提出的DCDMF网络利用其迭代去噪减轻低分辨率高光谱图像(Low Resolution HyperSpectral Image,LRHSI)与全色(Panchromatic,PAN)图像融合的模糊效应,结合自动编码器进行了优化;Huang等人[42]提出了一种利用扩散模型的时空融合方法STFDiff,将预测日期的低分辨率图像与其他日期的低/高分辨率图像对进行融合,以生成预测日期的高分辨率图像,通过从初始化的高斯噪声出发,在不同日期的先验图像的引导下反复细化生成高分图像.

通过引入条件信息,能够根据不同任务需求灵活调整融合过程,然而,条件扩散模型在图像融合中的应用仍面临一些挑战,如条件信息的准确获取与有效利用、模型复杂度与计算资源消耗等问题,未来还需进一步深入研究与优化.

2.2 基于训练方式与模型构建

在图像融合技术中,训练方式与模型构建是决定其应用成效的关键因素.训练方式决定了模型从数据中学习和提取有效特征的能力,模型构建则影响着模型对图像融合任务的适应性和处理效果.例如,图4呈现了降质图像的融合情况,从中可直观体现出训练方式与模型构建的重要性.



(a) 照明退化 (b) 照度下降和低分辨率及噪声

图4 DRMF[45]用于退化图像融合示意图

Fig. 4 Schematic diagram of DRMF[45] for degraded image fusion

(1) 预训练扩散模型

预训练扩散模型是在大规模数据上预先训练,学习数据分布特征,得到预训练模型.该模型后续可用于少样本、自定义及退化图像数据集等相关任务,以解决数据量有限或图像质量不佳情况下的图像处理问题,提升模型在特定场景下的适应性与性能表现.

少样本数据集.Ran等[43]提出的基于预训练的去噪扩散概率模型,实现了从稀疏LiDAR数据生成密集深度图,为自动驾驶、3D点云生成等计算机视觉和摄影测量任务提供了灵活高效的深度图生成策略.

自定义的数据集.Bar-Tal[44]等人提出基于预训练扩散模型的方法,将元素放置在由文本引导的扩散模型生成图像中以获得更大的控制.通过多个不同扩散过程有机组合,旨在生成预定区域内连贯、平滑的图像.

退化图像数据集.Tang等[45]提出了退化鲁棒多模态图像融合(DRMF)方法,预先训练多个退化鲁棒条件扩散模型以处理不同模式和退化情况,如低光、低分辨率和噪声等.在红外-可见光和医学图像融合实验中,DRMF在复杂退化条件下仍能表现出色.

自适应图像融合.基于预训练扩散模型的自适应图像融合展现出显著优势.如在盲图像超分辨率任务中,Lin等[46]的框架借助预训练扩散模型,利用自适应多模态融合(AMF)提升深度信息准确性并对齐多模态信息,生成超分辨率图像;Cao等[39]的条

件可控融合（CCF）将特定约束注入预训练 DDPM，为通用图像融合提供了一种智能、自适应的解决方案，无需额外训练，拓宽了图像融合应用场景.

预训练扩散模型可能存在数据偏差大、特定任务微调难、过拟合与泛化受限等问题，但仍可进一步挖掘其潜在价值与优势，以拓展图像融合领域的应用场景和实践空间.

（2） 自定义扩散过程

自定义扩散过程是指用户根据特定需求，自行设定扩散模型中的参数、噪声添加方式、反向扩散步骤等，以实现个性化的图像生成或处理效果.

针对多焦点图像融合，Pan 等[47]提出的 DDMF，在去噪过程中增加了聚焦区域与散焦区域隐藏的特征差异表示，以精确分类像素，提升了多焦点图像融合效果；Ji 等[48]提出在各向异性扩散模型中引入迭代控制算子控制迭代次数，用于红外与可见光图像融合；Ducay 等[49-50]提出了基于最近邻的扩散（NNDiffuse）模型，实现从多光谱全色锐化扩展到高光谱-多光谱图像融合，无需训练数据与昂贵的计算资源，速度快且光谱失真少，适合于低分辨率 HSI-MSI 锐化等场景；而在通用图像融合领域，Yang 等[51]提出光谱扩散（SD）这一自定义设计用于轻量级图像融合，可解决扩散概率模型计算成本大等的问题.

但自定义扩散过程往往参数调整复杂、对使用者要求高、过程难复现，进一步的研究思路在于提升模型效率、优化采样与架构等方面，以拓展其应用边界与效果.

（3） 结合其他技术

扩散模型存在生成效率低、质量不稳定及复杂场景适应性差等问题.为解决这些问题，与其他技术相结合是可行策略，能够有效提升其性能表现，满足更多实际场景需求.

与 Transformer 结合.为捕捉图像中的长距离依赖关系，将 Transformer 的自注意力机制与扩散模型相结合，Pan 等[52]构建了基于 Swin Transformer 网络的扩散模型框架用于 2D 医学图像合成；Yang 等[53]的 LFDT-Fusion 模型，将像素空间自动编码器与基于 Transformer 的扩散网络组合，解决原始分辨率图像迭代特征映射效率低的难题.

与生成对抗网络（GAN）结合.利用扩散模型数据生成多样性及真实性的优势，借助 GAN 的快速生成和优化能力，可提高图像融合的整体性能.Yang 等[54]提出的 DGFusion 方法整合扩散模型与 GAN，协同完成了红外与可见光图像的高质量融合；Zhao 等[55]将 Y 形 GAN 和去噪扩散隐式模型联合实现了建筑草图的着色与融合.

与卷积神经网络（CNN）结合.借助通道注意力机制，Xu 等[34]运用条件扩散模型实现了三模态医学图像融合与超分辨率的同步处理；DCAFuse（Lu 等[56]）模型融合了扩散模型与卷积神经网络（CNN）的优势，通过双分支结构分别提取并聚合不同模态图像的互补特征，实现在多模态图像融合中的特征有效整合；

与传统技术结合.除了常见的深度学习框架，扩散模型还可与传统技术相结合.例如，用于遥感图像融合的变分扩散方法[32]，用于多焦点图像融合的平均方法噪声扩散（AMND）的小波方法[24]等，特别适合于对图像细节和整体质量有要求的应用场景.Han 等[57]提出基于语义区域引导的端到端红外和可见光图像融合方法，改进图像生成模块时基于扩散模型，避免复杂融合规则设计；Guo 等[58]的 GLAD 打造了基于全局注意力的扩散模型（GLAD）用于红外与可见光图像融合，聚焦图像关键区域，更好地融合两种模态图像的重要信息；Fiza 等[59]提出基于边缘判别扩散滤波器（EDDF）的多焦点卫星图像融合方法，结合各向异性扩散（AD）和引导滤光片（GF）技术，有效解决了卫星图像融合中边界锐边导致的视觉及空间一致性问题，提升了多焦点卫星图像融合质量.

扩散模型与其他技术结合，能提升生成效率与质量，增强对复杂任务的适应性；但仍可能存在模型不稳定、参数调优复杂、计算资源需求大及兼容性等问题，有待进一步提升.

2.3 基于应用场景

随着技术的发展，图像融合技术的应用场景愈发丰富多样.通过对若干典型应用场景的深入剖析（如参考文献[56]的方法在图 5 所示的两种不同场景中的应用情况），能够清晰地展现基于扩散模型的图像融合技术在推动相关领域进步方面所发挥的重要作用，并对其未来的应用潜力进行合理展望.

（1） 医学图像融合

计算机断层扫描（CT）.CT 图像主要提供骨骼等结构信息，对于组织生理功能和代谢方面成像能力相对较弱，且容易受到伪影干扰.针对此问题，Khader 等[60]，Sun 等[30]提出的 DiffRecon 模型，结合磁共振成像（MRI）和交叉注意力机制技术，合成了高质量医学数据，在隐私保护的人工智能及小数据集增强方面发挥关键作用.



(a) DCAFuse用于医学图像融合 (b) DCAFuse用于红外与可见光图像融合

图5 DCAFuse[56]用于不同场景示意图

Fig. 5 Schematic diagrams of DCAFuse[56] applied to different scenarios

与其他技术结合.Xu等[34]借助通道注意力模块整合多模态信息,实现了三模态医学图像融合与超分辨率同步处理;Lu等[56]的DCAFuse采用双分支扩散-CNN互补特征聚合网络用于多模态图像融合;Ding等[61]的FDiff-Fusion构建基于模糊学习的去噪扩散融合网络进一步用于3D医学图像分割中,提高了分割的准确性.

同步各向异性扩散方程.各向异性扩散模型在医学图像融合领域发挥着关键作用.Zhu等[62]、Aakaaram等[63]、Goyal等[64]以及Lepcha等[65]运用基于各向异性扩散技术展开研究,依据图像局部特征,自适应地调控扩散进程,有效区分噪声信息,基于扩散机制降低噪声对图像质量的影响.在多模态医学图像融合场景,特别是MRI与CT图像的融合应用中,高效整合了不同模态图像中的互补信息,提升了融合的精准度与效率.

(2) 遥感图像融合

变分扩散模型.为了有效应对模型缺乏可解释性及训练不稳定等问题等,Zhang[66]等、Zhang[67]等提出扩散后验采样技术,同时,Zhang[68],Zhang[32]等将变分扩散模型用于遥感图像融合,通过对不同来源图像信息的整合与优化,提升了图像质量并增强目标特征表现.Zhang等[68]提出的二元扩散方法、Zhang等[67]等的SAR-BM3D模型还可以用于光学图像与SAR图像融合去斑等任务中.

新型网络结构.为获取更全面的图像结构信息,Wei等[69],Ma等[70]提出了新的编码器-解码器网络结构,以减少噪声干扰导致的信息偏差,提升融合图像的稳定性与可靠性;Cao等[71],Fiza等[59]则围绕平衡全局特征和局部细节展开,设计了条件调制模块及各向异性扩散(AD)等方法.

最近邻扩散.基于深度学习的方法往往需要大量训练数据,为此,Ducay等[49-50]提出基于最近邻扩散的算法应用于高光谱-多光谱图像融合,减少了训练数据,保证了融合效率,同时提升了融合图像的质量与实用性.

(3) 红外与可见光图像融合

结合传统融合技术.为充分利用CNN、UNet++及语义区域引导等技术在图像特征提取中的优势,Lu等人[56]提出DCAFuse、Yang等人[54]提出DGFusion、Han等人[57]提出IQDM等模型,能够精准捕捉图像整体架构与主要特征分布,有效突出融合图像关键区域,使得下游任务例如图像识别中,机器能够更为迅速、准确地锁定目标区域,显著提升了图像识别准确率.

提升色彩保真度.色彩保真度对于图像融合至关重要,因为颜色在人类视觉感知中占据关键地位,能够精准反映物体的光谱特性.然而,当前大多数红外与可见光图像融合方法,存在色彩保真度不足的问题.针对这一困境,Yue等人[72]提出的Dif-Fusion方法,以及Guo等人[20]提出的GLAD方法,创新性地采用了相似的技术路径.首先,在潜空间运用去噪网络深入学习多通道数据的分布规律,以此为基础,充分结合全局注意力机制,对多源信息进行高效聚合,最终成功生成多通道融合图像.这一过程显著提升了颜色的保真度,使得融合后的图像在色彩呈现上更加精准、自然.

3 数据集与评估

3.1 数据集

3.1.1 医学图像融合数据集

Harvard医学图像数据集(http://www.med.harvard.edu/AANLIB/home.html)涵盖医学影像、电子健康记录、UCI数据和生物医学文献,提供大量CT-MRI、PET-MRI、SPECT-MRI、CT-SPECT图像对,可用于医学图像融合模型训练与评估及智慧医疗诊断,其丰富的多模态医学图像数据为医学图像融合研究提供了权威资源支持,包含多种类型医学图像数据,适用于多种医学影像相关研究.

3.1.2 遥感图像融合数据集

(1) QuickBird卫星数据集

QuickBird 卫星数据集（https://www.mcgill.ca/library/find/maps/quickbird），由美国 DigitalGlobe 公司发射的QuickBird卫星获取.其全色图像分辨率可达0.61~0.72 m，多光谱图像分辨率为2.44~2.88 m，为多光谱与全色图像融合研究提供了高分辨率的遥感图像数据，可用于分析不同地物在高分辨率下的光谱和空间特征融合情况.

（2）WorldView系列卫星数据集

WorldView 系列卫星数据集（https://earth.esa.int/eogateway/missions/worldview），例如 WorldView-2 卫星能提供0.5 m分辨率的全色图像和1.8 m分辨率的多光谱图像，WorldView -3卫星提供31 cm分辨率的全色图像和1.24 m分辨率的多光谱图像.这些数据可用于研究不同分辨率下的遥感影像融合，以及在高精度遥感应用中如何更好地结合光谱和空间信息.

（3）GeoEye系列卫星数据集

GeoEye 系列卫星数据集（https://earth.esa.int/eogateway/catalog/geoeye-1-esa-archive），以 GeoEye-1 卫星为例，它提供0.41 m分辨率的全色图像和1.65 m彩色分辨率的多光谱图像.该数据集有助于研究在特定地理区域和应用场景下的遥感影像融合技术，为土地利用监测、城市规划等提供数据支持.

（4）GaoFen系列卫星数据集

我国自主研发的民用光学遥感卫星GaoFen系列数据集（https://www.nature.com/articles/s41597-024-03009-5）也可用于遥感图像融合研究.如GaoFen 1号卫星拍摄的全色图像和多光谱图像分辨率分别为2 m和8 m，GaoFen2号卫星拍摄的全色图像和多光谱图像分辨率分别为1 m和4 m，这些数据能够满足不同尺度和精度要求的遥感影像融合实验与分析.

3.1.3　红外与可见光图像融合数据集

（1）TNO数据集

TNO数据集（https://doi.org/10.6084/m9.figshare.c.3860689.v1）在图像融合领域占据经典地位.它涵盖夜间图像（涉及可见光、近红外和长波红外波段）、图像序列（如皮划艇场景）以及动态监控场景运动序列等丰富内容.其数据来源于真实场景，拥有多种多波段图像，有很强的模拟性与可扩展性，在图像融合、颜色融合及目标识别等研究中发挥着重要作用.

（2）RoadScene数据集

RoadScene 数据集（https://github.com/jiayi-ma/RoadScene）专为图像融合研究打造，其从FLIR视频中精心挑选出具有代表性的场景构建而成，包含221对已配准的可见光和红外图像，涵盖丰富的道路、车辆以及行人场景，为实际交通等场景下的图像融合研究提供了极具价值的数据支撑.

（3）MSRS数据集

MSRS 数据集（https://github.com/Linfeng-Tang/MSRS）是基于MFNet数据集构建的多光谱道路场景数据集，主要用于红外和可见光图像融合.它包含1444 对高质量且已对齐的图像，有效改善了MFNet数据集图像对未对齐以及红外图像质量欠佳等状况，同时添加了语义分割标签，适用于多种图像融合相关研究.

（4）LLVIP数据集

LLVIP 数据集（https://bupt-ai-cz.github.io/LLVIP/）是专为微光视觉红外与可见光融合而设的数据集，包含30976张图像，涵盖丰富的夜晚场景及少量白天场景.这些图像由特定相机拍摄，并对行人进行了标注，具有场景多样性和精准的对齐性，在图像融合、行人检测、图像转换等任务方面展现出较高的实用价值.

（5）KAIST数据集

KAIST 数据集（https://soonminhwang.github.io/rgbt-ped-detection/data/）专注于多光谱图像分析，涵盖城市街道、校园等多样场景，提供大量可见光和红外图像，有力支撑目标检测、行人识别等研究及多光谱视觉算法的开发与评估.

（6）M3FD数据集

M3FD 数据集（https://github.com/dlut-dimt/TarDAL）包含了高分辨率的红外和可见光图像.可见光图像的分辨率为1024 × 768，成像范围较宽；红外图像的标准分辨率为640 × 512.标注了33000多个对象，包含人、车、公交车、汽车、摩托车、灯具、卡车等常见对象.图像对总数为4200（用于融合、检测和基于融合的检测），还有600个独立场景用于融合.

3.2　评估

3.2.1　评估指标

常用评估指标从多个维度对融合图像进行量化分析，为衡量融合效果提供了精确依据.不同客观评价指标相互补充，从不同侧面全面、精确地量化了融合图像的质量，为图像融合技术的研究、开发与应用提供了坚实的数据支撑，助力研究人员不断优化融合算法，提升图像融合效果.

(1) 信息熵(Entropy,EN)

信息熵用于度量融合图像包含信息量的多少,值越大表示图像包含的信息越丰富,其计算公式为:

$$\mathrm{EN} = -\sum_{l=0}^{L-1} p_l \log_2 p_l \,, \tag{11}$$

其中,L表示灰度级数,p_l为融合图像中对应灰度级的归一化直方图.

(2) 互信息(Mutual Information,MI)

互信息用于评估融合图像与源图像之间的共享信息,可量化融合图像从源图像中保留的信息量.对于融合图像F和源图像A、B,其互信息的具体定义为:

$$\mathrm{MI} = \mathrm{MI}_{A,F} + \mathrm{MI}_{B,F} \,, \tag{12}$$

其中,$\mathrm{MI}_{A,F}$和$\mathrm{MI}_{B,F}$分别表示从源图像A和源图像B中转移到融合图像F的信息.特别地,$\mathrm{MI}_{X,F}$(X代表源图像A或B)的具体定义为:

$$\mathrm{MI}_{X,F} = \sum_{x,f} P_{X,F}(x,f) \log \frac{P_{X,F}(x,f)}{P_X(x)P_F(f)} \,, \tag{13}$$

式中,$P_X(x)$和$P_F(f)$分别表示源图像X和融合图像F的边缘直方图,$P_{X,F}(x,f)$表示源图像与融合图像的联合直方图.融合图像的MI值越高,意味着相应的融合算法从源图像中转移到融合图像中的信息越多.

(3) 视觉信息保真度(Visual Information fidelity,VIF)

视觉信息保真度是一种基于人类视觉系统特性的图像质量评价指标,用于衡量融合图像与原始参考图像之间的视觉信息保持程度,能更准确地反映人类观察者对图像质量的感知.

VIF的计算基于自然场景统计模型,假设原始图像I和融合图像J都是零均值的平稳高斯随机场,其计算公式如下:

$$\mathrm{VIF}(I,J) = \sum_{l=1}^{L} \frac{\sigma^2_{i_l j_l}}{\sigma^2_{i_l j_l^2} + \sigma^2_{i_l j_l}} \,, \tag{14}$$

其中,L表示图像分解的层数,通常使用小波分解等方法将图像分解为不同尺度和方向的子带,$\sigma^2_{i_l}$和$\sigma^2_{i_l}$分别是原始图像I和融合图像J在第l层子带上的方差,$\sigma^2_{i_l j_l}$是原始图像I和融合图像J在第l层子带上的协方差.该公式计算了每一层子带的视觉信息保真度,并将它们进行求和,以得到整个图像的VIF值.

(4) 结构相似性度量(Structural Similarity Index Measure,SSIM)

结构相似性度量用于对融合过程中的信息损失和失真进行建模,以反映融合图像与源图像之间的结构相似性.对于融合图像F与源图像X之间的结构相似性$SSIM_{X,F}$,其定义为:

$$\mathrm{SSIM}_{X,F} = \sum_{x,f} \frac{2\mu_x\mu_f + C_1}{\mu_x^2 + \mu_f^2 + C_1} \times \frac{2\sigma_x\sigma_f + C_2}{\sigma_x^2 + \sigma_f^2 + C_2} \times \frac{\sigma_{xf} + C_3}{\sigma_x\sigma_f + C3} \,, \tag{15}$$

式中,x和f分别表示源图像和融合图像在一个滑动窗口内的图像块.σ_{xf}表示源图像与融合图像的协方差,σ_x和σ_f分别表示源图像和融合图像的标准差,μ_x和μ_f分别表示源图像和融合图像的均值;C_1、C_2和C_3是用来防止除数为零的常量.

SSIM指标的最终计算式为:

$$\mathrm{SSIM} = \frac{\mathrm{SSIM}_{A,F} + \mathrm{SSIM}_{B,F}}{2} \,, \tag{16}$$

其中A、B为源图像.SSIM越大,说明融合图像越接近源图像,即融合过程中的信息丢失和失真越小.

(5) 平均梯度(Average Gradient,AG)

平均梯度通过测量融合图像的梯度信息并以此表征融合图像纹理细节的指标,其具体定义为:

$$\mathrm{AG} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}} \,, \tag{17}$$

其中,$\nabla F_x(i,j) = F(i,j) - F(i+1,j)$,$\nabla F_y(i,j) = F(i,j) - F(i,j+1)$,$M$和$N$分别表示图像的行数和列数,$F(i,j)$表示融合图像在第$i$行第$j$列的像素值.AG值越高,意味着包含更加丰富的梯度信息(纹理细节),图像清晰度越高,融合质量越好.

(6) 标准差(Standard Deviation,SD)

标准差反映融合图像的对比度及分布的指标,其计算公式为:

$$\mathrm{SD} = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - \mu)^2} \,, \tag{18}$$

其中,$F(i,j)$表示融合图像在第i行第j列的像素值,μ表示融合图像的均值,M和N分别表示图像的行数和列数.人类视觉系统往往更能被具有高对比度的区域所吸引,因此具有更高SD的融合结果具有更好的对比度.

3.2.2 定性评估

本节以医学图像融合为例开展定性评估,以深入剖析该技术在医疗场景中的实际效能,也为其他类型的图像融合方法提供思路和参考.医学图像融

合通过整合CT与MRI等不同模态图像的优势信息，为疾病诊断、治疗规划等提供更全面的图像依据.选择DDFM、MIF、STMMIF、Diff-IF、Text-Difuse、SAD-MIF、Contourlet等代表性方法的融合结果进行定性比较分析，观察它们在结合CT与MRI图像特征方面的表现情况，如图6所示.

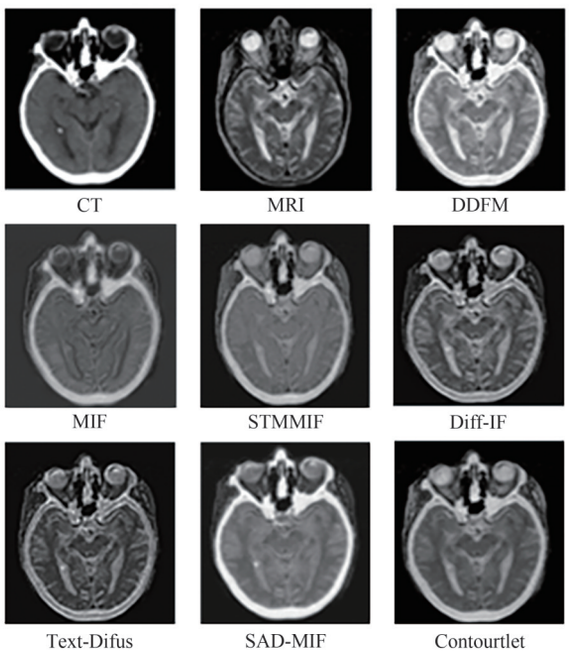


图6　图像融合定性对比结果

Fig. 6　Qualitative comparison results of image fusion

从图6可以看出，不同融合方法各有其特点.DDFM方法较好兼顾了解剖结构与软组织细节，边界清晰；MIF方法的融合效果均衡、自然，但细节方面却稍显不足.STMMIF方法的解剖结构突出，软组织细节表现欠佳.Diff－IF方法生成的融合图像具有良好的层次感和清晰度，其软组织与结构融合方面表现好.Text-Difuse方法融合图像的细节丰富，但部分区域融合协调性有待提升.SAD-MIF方法则具有融合稳定，结构与软组织融合平衡等特点.Contourlet方法在细节方面有独特的表现，但融合图像的整体一致性和自然程度不够.

3.2.3　定量评估

针对医学图像融合领域，基于哈佛（Harvard）数据集，仍旧以上述7种算法进行对比实验，它们分别采用了不同的扩散模型框架，包括DDFM、MIF、STMMIF、Diff-IF、Text-Difuse、SAD-MIF、Contourlet等经典及前沿的融合方法，并从多个维度进行定量评估.通过对几种不同方法在EN、MI、VIF、SSIM、AG、SD等指标上的表现对比，清晰了解各方法的优势与不足，为图像融合技术的发展提供参考.具体对比结果如表2所示.

表2　图像融合算法定量比较结果

Tab. 2　Quantitative comparison of image fusion methods

方法	EN	MI	VIF	SSIM	AG	SD
DDFM	4.64	1.99	**0.76**	**0.9**	5.00	63.11
MIF	5.41	1.50	0.66	0.80	5.992	56.99
STMMIF	5.50	1.20	0.57	0.82	**9.96**	50.00
Diff-IF	5.00	2.49	0.74	0.85	6.00	40.00
Text-Difuse	**6.44**	1.00	0.49	0.70	7.31	**80.19**
SAD-MIF	4.50	2.84	0.61	0.75	5.50	50.00
Contourlet	4.00	**4.79**	0.55	0.60	6.28	67.65

注：标注为粗体的：表现最佳；标注为下划线的：表现次佳.

综合各项指标来看，不同文献的图像融合效果各有优劣.首先，DDFM方法在视觉感知和结构相似性方面表现出色，VIF值0.76和SSIM值0.9相对较高，但清晰度和纹理细节稍显不足，AG值仅5.00；而MIF方法的各项指标表现则较为均衡.此外，从表中可以看出，STMMIF方法具有较高的融合图像清晰度，AG值达9.959，不过在视觉和结构保持方面表现一般.Diff-IF方法在信息重叠度和结构保持上有一定优势；Text-Difuse方法则具有丰富的融合图像信息，EN值达6.44，但视觉感知和结构相似性欠佳.SAD-MIF方法的各项指标处于中等水平；Contourlet方法在信息重叠上表现突出，MI值4.7864，但结构相似性较差.由此可见，每种融合方法都有其独特的优势和局限，在实际应用中需根据具体需求选择合适的方法.

4　总结与展望

图像融合技术在提升图像信息质量与可用性方面至关重要，扩散模型的引入为该领域带来新活力.本文全面综述了扩散模型在图像融合中的研究进展，涵盖模型基础、方法分类、数据集与评估指标等方面，以下是对现有研究的总结及对未来发展的展望.

4.1　总结

模型框架多样：去噪扩散概率模型（DDPM）在单模态和多模态图像融合中均有应用，如多聚焦图像融合、医学、遥感、红外与可见光图像融合等，但仍存在生成效率低和模型复杂等问题；基于随机微分方程的扩散模型（SDE）为医学、遥感等图像融合提供了新的思路，但面临计算复杂、参数估计困难和噪声敏感等挑战；条件扩散模型（CDM）通过引入

条件信息提升了融合的针对性,然而在条件信息处理和模型复杂度上仍需优化.

训练与构建多元:预训练扩散模型适用于少样本、自定义及退化图像数据集,增强了模型的适应性,但存在数据偏差和微调困难等问题;自定义扩散过程能够实现个性化图像处理,但参数调整复杂;与其他技术的结合虽可提升扩散模型性能,但可能导致模型不稳定和参数调优问题.

应用场景广泛:在医学图像融合中,有助于整合不同模态的信息,以提高诊断准确性;在遥感图像融合方面,能够提升图像质量和目标特征表现;在红外与可见光图像融合中,则改善了色彩保真度和融合效果.

数据集与指标丰富:常用的数据集涵盖医学、遥感、红外与可见光图像融合的各类专用数据,为研究提供了丰富的数据支持;评估指标包括主观和客观两类,客观指标如信息熵、互信息等可从多维度量化融合效果,而主观指标虽受人为因素影响,但在特定环境下依然具有重要价值.二者结合有助于全面评估图像融合质量.

4.2 展望

4.2.1 模型优化与性能提升

针对不同扩散模型框架的固有缺陷,开展优化研究,如改善DDPM的采样过程和结构、简化基于SDE模型的计算、降低条件扩散模型的复杂度等,同时增强模型对复杂场景和多模态数据的适应性,以提升融合图像质量和生成效率.此外,应着重增强模型的可解释性与可控性.鉴于扩散模型在图像融合中的工作机制复杂、决策过程难以理解,未来需探索增强可解释性的方法,使研究人员明确决策依据,从而有助于模型优化与应用.同时,研发可控性强的模型,使用户根据需求调整输出,实现个性化图像融合.

4.2.2 训练与融合策略改进

改进预训练模型的训练策略以减少数据偏差,并研发高效的微调方法;简化自定义扩散过程中的参数调整;探索与其他技术更为稳定和高效的融合方式,以平衡性能提升和资源消耗,从而增强模型在实际应用中的可行性.尤其要推进融合知识与数据驱动的混合方法,将传统图像处理知识与数据驱动的扩散模型相结合,实现优势互补.例如,可以利用传统的图像分割和特征提取为扩散模型提供先验信息,或者用扩散模型的特征来优化传统方法的参数,以提升图像融合的性能与鲁棒性.

4.2.3 应用领域拓展探索

在医学领域,进一步支持精准诊断和疾病监测;在遥感方面,满足高分辨率和多光谱融合需求,以服务于环境监测和资源勘探;在红外与可见光融合中,提升在安防监控和自动驾驶等场景的应用效果,同时挖掘扩散模型在新兴领域的应用潜力.此外,应积极探索新型图像融合任务,伴随图像处理与计算机视觉技术的发展,如视频融合、全景图像融合和多光谱图像融合等新任务不断涌现,未来应关注这些任务,发掘扩散模型在其中的应用潜力与优化策略.

4.2.4 跨领域合作与评估完善

加强跨领域合作与融合,图像融合技术与计算机视觉、人工智能和生物医学等多个领域密切相关.未来需促进跨领域交流,借鉴其他领域的成果与技术,为图像融合注入新活力与创新点.同时,应构建更科学、全面的评估体系,减少主观指标的不确定性,优化客观指标的权重与组合,开发新指标以准确衡量融合图像在复杂任务中的性能,为算法改进和技术发展提供可靠依据,推动图像融合技术的持续进步.

参 考 文 献

[1] Zhang Y. Understanding image fusion[J]. Photogrammetric engineering and remote sensing, 2004, 70(6): 657-661.

[2] Mitianoudis N, Stathaki T. Pixel-based and region-based image fusion schemes using ICA bases[J]. Information Fusion, 2007, 8(2): 131-142.

[3] Aslantas V, Toprak A N. A pixel based multi-focus image fusion method[J]. Optics Communications, 2014, 332: 350-358.

[4] Sreeja P, Hariharan S. An improved feature based image fusion technique for enhancement of liver lesions[J]. Biocybernetics and Biomedical Engineering, 2018, 38(3): 611-623.

[5] Hossny M, Nahavandi S, Crieghton D. Feature-based image fusion quality metrics[C]// Intelligent Robotics and Applications. Berlin: Springer Berlin Heidelberg, 2008: 469-478.

[6] Kausar N, Majid A. Random forest-based scheme using feature and decision levels information for multi-focus image fusion[J]. Pattern Analysis and Applications, 2016, 19(1): 221-236.

[7] Zhang Y, Liu Y, Sun P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99-118.

[8] Liu Y, Chen X, Peng H, et al. Multi-focus image fusion with a deep convolutional neural network[J]. Information Fusion, 2017, 36: 191-207.

[9] Rren X, Meng F, Hu T, et al. Infrared-visible image fusion based on convolutional neural networks (CNN)[C]// Intelligence Science and Big Data Engineering. Cham: Springer International Publishing, 2018: 301-307.

[10] Wang L J, Han J, Zhang Y, et al. Image fusion *via* feature residual and statistical matching [J]. IET Computer Vision, 2016, 10(6): 551-558.

[11] Wu Y, Huang M, Li Y, et al. A distributed fusion framework of multispectral and panchromatic images based on residual network[J]. Remote Sensing, 2021, 13(13): 2556.

[12] Xu H, Ma J, Zhang X P. MEF-GAN: Multi-exposure image fusion *via* generative adversarial networks [J]. IEEE Transactions on Image Processing, 2020, 29: 7203-7216.

[13] Zhang H, Le Z, Shao Z, et al. MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion[J]. Information Fusion, 2021, 66: 40-53.

[14] Zhang J, Jiao L, Ma W, et al. Transformer based conditional GAN for multimodal image fusion[J]. IEEE Transactions on Multimedia, 2023, 25: 8988-9001.

[15] Dhariwal P, Nichol A. Diffusion models beat gans on image synthesis [J]. Advances in Neural Information Processing Systems, 2021, 34: 8780-8794.

[16] Song J, Meng C, Ermon S. Denoising Diffusion Implicit Models [C]//International Conference on Learning Representations. Vienna: ISLR, 2021:1-20.

[17] Nichol A Q, Dhariwal P. Improved denoising diffusion probabilistic models [C]//International conference on machine learning. Vienna: PMLR, 2021, 139: 8162-8171.

[18] Austin J, Johnson D D, HO J, et al. Structured denoising diffusion models in discrete state-spaces[J]. Advances in Neural Information Processing Systems, 2021, 34: 17981-17993.

[19] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

[20] Croitoru F A, Hondru V, Ionescu R T, et al. Diffusion models in vision: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10850-10869.

[21] Song Y, Sohl-Dickstein J, Kingma D P, et al. Score-based generative modeling through stochastic differential equations[J]. arXiv:2020: 2011.13456.

[22] Li M, Pei R, Zheng T, et al. FusionDiff: Multi-focus image fusion using denoising diffusion probabilistic models [J]. Expert Systems with Applications, 2024, 238: 121664.

[23] Li X, Zong S, Duan Z, et al. A new generative method for multi-focus image fusion of underwater micro bubbles [J]. Scientific Reports, 2024, 14: 30280.

[24] Singh P, Diwakar M. Wavelet-based multi-focus image fusion using average method noise diffusion [J]. Recent Advances in Computer Science and Communications, 2021, 14(8): 2436-2448.

[25] Dorjsembe Z, Odonchimed S, Xiao F. Three-dimensional medical image synthesis with denoising diffusion probabilistic models[C]//Medical Imaging with Deep Learning. Zurich: MIDL, 2022:1-3.

[26] Zhao Z, Bai H, Zhu Y, et al. DDFM: denoising diffusion model for multi-modality image fusion[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE/CVF, 2023: 8082-8093.

[27] Yang B, Jiang Z, Pan D, et al. Multimodal image fusion based on diffusion model[C]//Proceedings of the 2024 International Conference on Advanced Robotics, Automation Engineering and Machine Learning. Hangzhou:ACM, 2024: 75-81.

[28] Luo Z, Chen D, Zhang Y, et al. Notice of removal: VideoFusion: Decomposed diffusion models for high-quality video generation[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver:IEEE, 2023: 10209-10218.

[29] Le D T, Shi H, Cai J, et al. Diffusion model for robust multi-sensor fusion in 3D object detection and BEV segmentation[C]//Computer Vision-ECCV 2024. Cham: Springer,2024: 232-249.

[30] Sun J, Cao N, Bi H, et al. DiffRecon: Diffusion-based CT reconstruction with cross-modal deformable fusion for DR-guided non-coplanar radiotherapy[J]. Computers in Biology and Medicine, 2024, 179: 108868.

[31] 肖继锋，蔡文培. 医学图像的处理方法、装置、计算机设备和存储介质[P]中国专利:CN117710727A. 2024-03-15.

[32] Zhang C, Han J, Zhu J, et al. Variational diffusion method for remote sensing image fusion [J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 4008405.

[33] Pop S, Lavialle O, Terebes R, et al. A PDE-based approach for image fusion [C]//Advanced Concepts for Intelligent Vision Systems. Berlin: Springer Berlin

Heidelberg, 2007: 121-131.

[34] Xu Y, Li X, Jie Y, et al. Simultaneous tri-modal medical image fusion andSuper-resolution using conditional diffusion model [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2024. Cham: Springer, 2024: 635-645.

[35] Muller-Franzes G, Niehues J M, Khader F, et al. A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis [J]. Scientific Reports, 2023, 13(1): 12098.

[36] Meng Q, Shi W, Li S, et al. PanDiff: A novel pansharpening method based on denoising diffusion probabilistic model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5611317.

[37] Yi X, Tang L, Zhang H, et al. Diff-IF: Multi-modality image fusion *via* diffusion model with fusion knowledge prior[J]. Information Fusion, 2024, 110: 102450.

[38] Zhang H, Cao L, Ma J. Text-DiFuse: An interactive multi-modal image fusion framework based on text-modulated diffusion model [C]//The Thirty-eighth Annual Conference on Neural Information Processing Systems. Vancouver: NeurIPS Foundation, 2024: 1-24.

[39] Cao B, Xu X, Zhu P, et al. Conditional controllable image fusion[C]//The Thirty-eighth Annual Conference on Neural Information Processing Systems. Vancouver: NeurIPS Foundation, 2024: 1-19.

[40] 李小松, 接玉婵, 程晓琦, 等. 基于条件扩散概率模型的多模态图像融合与超分辨率方法[P]中国专利: CN118314022B. 2024-11-19.

[41] Li S, Li S, Zhang L. Hyperspectral and panchromatic images fusion based on the dual conditional diffusion models [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5526315.

[42] Huang H, He W, Zhang H, et al. STFDiff: Remote sensing image spatiotemporal fusion with diffusion models[J]. Information Fusion, 2024, 111: 102505.

[43] Ran W, Yuan W, Shibasaki R. Few-shot depth completion using denoising diffusion probabilistic model [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver:IEEE, 2023: 6559-6567.

[44] Bar-Tal O, Yariv L, Lipman Y, et al. MultiDiffusion: Fusing diffusion paths for controlled image generation [J]. Proceedings of Machine Learning Research, 2023, 202: 1737-1752.

[45] Tang L, Deng Y, Yi X, et al. DRMF: Degradation-robust multi-modal image fusion *via* composable diffusion prior [C]//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne: ACM, 2024: 8546-8555.

[46] Lin J, Wang Y, Tao Z, et al. Adaptive multi-modal fusion of spatially rariant kernel refinement with diffusion model for blind image super-resolution [C]//Computer Vision-ECCV. Cham: Springer, 2024: 363-380.

[47] Pan X, Zhai H, Yang Y, et al. Improving multi-focus image fusion through noisy image and feature difference network[J]. Image and Vision Computing, 2024, 142: 104891.

[48] Ji J, Zhang Y, Lin Z, et al. Infrared and visible image fusion based on iterative control of anisotropic diffusion and regional gradient structure [J]. J Sensors, 2022, 2022: 1-10.

[49] Ducay R, Messinger D W. Image fusion of hyperspectral and multispectral imagery using nearest-neighbor diffusion [J]. Journal of Applied Remote Sensing, 2023, 17(2): 1-14.

[50] Ducay R, Messinger D W. Hyperspectral-multispectral image fusion using nearest-neighbor diffusion-based sharpening algorithm [C]//Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXVIII. Orlando:SPIE, 2022: 22.

[51] Yang X, Zhou D, Feng J, et al. Diffusion probabilistic model made slim [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver:IEEE, 2023: 22552-22562.

[52] Pan S, Wang T, Qiu R L J, et al. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model [J]. Physics in Medicine and Biology, 2023, 68(10): 105004.

[53] Yang B, Jiang Z, Pan D, et al. LFDT-Fusion: A latent feature-guided diffusion Transformer model for general image fusion [J]. Information Fusion, 2025, 113: 102639.

[54] Yang Z, Qin H, Zeng S, et al. DGFusion: A novel infrared and visible image fusion method based on diffusion and generative adversarial networks[J]. IEEE Access, 2024, 12: 147051-147064.

[55] Zhao L, Song D, Chen W, et al. Coloring and fusing architectural sketches by combining a Y-shaped generative adversarial network and a denoising diffusion implicit model [J]. Computer-Aided Civil and Infrastructure Engineering, 2024, 39(7): 1003-1018.

[56] Lu X, Jiang Y, Hong H, et al. DCAFuse: Dual-branch diffusion-CNN complementary feature aggregation network for multi-modality image fusion[C]//Proceedings

of the 32nd ACM International Conference on Multimedia. Melbourne:ACM, 2024: 1524-1533.

[57] Han G, Zhang X, Huang Y. An end-to-end based on semantic region guidance for infrared and visible image fusion[J]. Signal, Image and Video Processing, 2024, 18(1): 295-303.

[58] Guo H, Chen M, Li K, et al. GLAD: A global-attention-based diffusion model for infrared and visible image fusion[C]//Advanced Intelligent Computing Technology and Applications. Singapore:Springer, 2024: 345-356.

[59] Fiza S, Safinaz S. Multi-focus image fusion using edge discriminative diffusion filter for satellite images [J]. Multimedia Tools and Applications, 2024, 83(25): 66087-66106.

[60] Khader F, Muller-Franzes G, Tayebi Arasteh S, et al. Denoising diffusion probabilistic models for 3D medical image generation[J]. Scientific Reports, 2023, 13(1): 7303.

[61] Ding W, Geng S, Wwang H, et al. FDiff-Fusion: Denoising diffusion fusion network based on fuzzy learning for 3D medical image segmentation [J]. Information Fusion, 2024, 112: 102540.

[62] Zhu R, Li X, Zhang X, et al. MRI and CT medical image fusion based on synchronized-anisotropic diffusion model[J]. IEEE Access, 2020, 8: 91336-91350.

[63] Aakaaram V, Bachu S. MRI and CT image fusion using synchronized anisotropic diffusion equation with DT-CWT decomposition [C]//2022 Smart Technologies, Communication and Robotics (STCR). Sathyamangalam: IEEE, 2022: 1-5.

[64] Goyal B, Dogra A, Khoond R, et al. Medical image fusion based on anisotropic diffusion and non-subsampled contourlet transform [J]. Computers, Materials & Continua, 2023, 76(1): 311-327.

[65] Lepcha D C, Dogra A, Goyal B, et al. Multimodal medical image fusion based on pixel significance using anisotropic diffusion and cross bilateral filter [J]. Human-Centered Computing and Information Sciences, 2022, 12:1-20.

[66] Zhang C, Wang Z, Han J. Diffusion posterior sampling for remote sensing image fusion [C]//7th International Conference on Vision, Image and Signal Processing (ICVISP 2023). London: IET, 2023: 28-33.

[67] Zhang C, Chang Y, Wu Y, et al. Semantic information guided diffusion posterior sampling for remote sensing image fusion [J]. Scientific Reports, 2024, 14(1): 27259.

[68] Zhang C, Chang Y, Wu Y, et al. Binary diffusion method for image fusion with despeckling[C]//2024 2nd International Conference on Algorithm, Image Processing and Machine Vision (AIPMV). Zhenjiang: IEEE, 2024: 243-248.

[69] Wei J, Gan L, Tang W, et al. Diffusion models for spatio-temporal-spectral fusion of homogeneous Gaofen-1 satellite platforms[J]. International Journal of Applied Earth Observation and Geoinformation, 2024, 128: 103752.

[70] Ma Y, Wang Q, Wei J. Spatiotemporal fusion *via* conditional diffusion model [J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 5002405.

[71] Cao Z, Cao S, Deng L J, et al. Diffusion model with disentangled modulations for sharpening multispectral and hyperspectral images [J]. Information Fusion, 2024, 104: 102158.

[72] Yue J, Fang L, Xia S, et al. Dif-fusion: Toward high color fidelity in infrared and visible image fusion with diffusion models [J]. IEEE Transactions on Image Processing, 2023, 32: 5705-5720.

(责编&校对　雷建云)