彭晓旺, 刘伟, 陶家俊, 等. 基于 Transformer 的手绘与标准印刷化学分子结构图像双向循环生成方法[J]. 智能计算机与应用,2025,15(4):77-85. DOI:10.20169/j.issn.2095-2163.25030303

基于 Transformer 的手绘与标准印刷化学分子结构图像双向循环生成方法

彭晓旺, 刘 伟, 陶家俊, 贺先域
(湖南中医药大学 信息科学与工程学院, 长沙 410208)

摘 要: 手绘化学分子结构图像生成是一项富有意义且有挑战性的任务。目前,基于深度学习的手绘化学分子结构识别研究取得了一些进展,然而训练数据集的匮乏严重制约了模型的性能提升。此外,针对标准印刷体分子结构的识别模型已趋于成熟并展现出优异的识别效果,然而这些工具却没有被充分用于手绘分子式的识别,如何有效利用现有技术成果实现高精度的手绘分子识别,成为当前研究的重要方向之一。针对上述问题,本文提出了 BiMIGAN 模型,该模型整体是一个类循环生成网络架构,其中的生成器使用了引入 Transformer 的 U 型卷积神经网络,判别器采用了 PatchGAN 判别器,在训练的过程中还采用了预训练与梯度惩罚的训练策略保证了模型生成图像的质量。在使用自建手绘分子数据集 HNUCM-HDM 的实验中,模型生成图像的 *FID* 与 *KID* 与其他方法相比均是最优的,分别为 20.809 和 0.006 5(±0.000 3);在后续识别实验中,使用模型将手绘分子标准化后再进行识别,能大幅提升识别准确性,最多可达 40%。

关键词: 手绘化学分子结构图像生成; 手绘化学分子结构识别; 循环生成对抗网络; Transformer; 梯度惩罚
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2025)04-0077-09

Bidirectional cyclic translation between printed and hand-drawn molecule structure images based on Transformer

PENG Xiaowang, LIU Wei, TAO Jiajun, HE Xianyu
(School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, China)

Abstract: The generation of hand-drawn chemical molecular images is a meaningful and challenging task. Currently, some progress has been made in deep learning-based recognition of hand-drawn chemical molecular structures. However, the scarcity of training datasets severely limits the performance improvement of models. Additionally, recognition models for standard printed molecular structures have matured and demonstrated excellent performance, yet these tools have not been fully utilized for hand-drawn molecular recognition. To address the aforementioned issues, this paper proposes the BiMIGAN model, a cyclic generative network with a Transformer-enhanced U-Net generator and PatchGAN discriminator, trained using pre-training and gradient penalty for high-quality generation. In experiments using the self-built hand-drawn molecular dataset HNUCM-HDM, the model achieves the best *FID* and *KID* scores compared to other methods, with values of 20.809 and 0.006 5 (±0.000 3), respectively. In subsequent recognition experiments, standardizing hand-drawn molecules using the model before recognition could significantly improve accuracy, with a maximum increase of up to 40%.

Key words: hand-drawn molecule generation; hand-drawn molecule recognition; CycleGAN; Tranformer; gradient penalty

0 引 言

化学分子结构识别系统已广泛应用于各个领域,包括药物研发、生物化学、教育和有机合成等[1-2]。特别是在新型药物研发的过程中,这些识别系统至关重要,因为针对于药物研发中数据缺乏的问题最有效的办法就是从已出版的文档中提取化学信息[3]。此外,手绘化学分子结构式也是化学界

基金项目: 湖南省重点研发计划项目(2024JK2130); 湖南省自然科学基金面上项目(2022JJ30438); 长沙市自然科学基金项目(kq2202260); 湖南省中医药科研课题(B2023039)。

作者简介: 彭晓旺(1995—),男,硕士研究生,主要研究方向:计算机视觉,光学化学识别。
通信作者: 刘 伟(1982—),男,博士,教授,主要研究方向:人工智能,化学信息学。Email:weiliu@hnucm.edu.cn。

收稿日期: 2025-03-03

学生和研究人员常用的信息传递方式,如何从这些文档中便捷有效地提取蕴含大量化学信息的分子是一个亟待解决的问题。

自动化地从文档中提取化学结构的过程被称为光学化学结构识别(Optical Chemical Structure Recognition, OCSR)。近年来,快速发展的深度学习方法取得了可观成就。这些数据驱动方法可以显著减轻对领域知识的依赖,并且具有出色的灵活性和可扩展性,能够应用于各个领域[4-5]。因此,许多研究着眼于将深度神经网络应用于 OCSR[6-7]。

在深度学习领域,大数据驱动的方法能显著提升识别模型的识别精确度[8],这在面向标准印刷体分子结构识别的研究中[9]和面向手绘分子的识别研究中[10]都得到了证实。然而,目前可用于手绘分子结构识别研究的公开数据集匮乏,Brinkhaus 等学者[11]提供的手绘分子结构数据集仅包含约 5 000 个图片,这对于手绘化学分子结构识别的研究来说杯水车薪。合成数据是许多数据驱动识别方法中用来提供训练数据的重要手段[12],但仍需指出的是目前基于深度学习的生成仿真手绘分子图像的方法与工具却不多见。

另一方面,标准印刷体分子式识别模型研究成果在近年来陆续涌现,部分识别方法在识别印刷体分子图像时能实现高达 98% 的完全匹配率[13]。然而,这些模型在面对复杂的手绘分子结构时,识别能力通常会显著下降。手绘分子结构因线条粗细不均、形态不规则、等特点,难以被现有模型准确识别,这导致大量已有优秀模型的潜力未能得到充分利用。如果不对手绘分子结构图像进行预处理,就无法发挥这些模型的优势。因此,结合图像生成技术合理利用现有识别模型,是提高手绘分子识别准确率和模型复用率的关键点。

基于以上 2 种情况,本文结合循环生成对抗网络、Tranformer 以及卷积神经网络提出了一种新型的非配对双向循环生成化学分子图像网络 BiMIGAN。本文的主要贡献有 3 点:

(1)本文创新性地将像素级 Transformer 网络作为中间层引入到 U-Net 型生成器中,利用其强大的捕捉数据全局特征依赖关系的能力,结合 U 型卷积网络结构实现多尺度的特征提取,显著提升了模型生成复杂手绘图像的能力。

(2)本文在训练中使用了自监督预训练和梯度惩罚等训练手段实现了生成对抗网络的高质量,高稳定性的图像生成。

(3)本文提出了一种新的手绘化学分子结构识别的研究模式,即利用生成模型将手绘化学分子结构图像标准化再使用现有高效模型进行识别,从而避免构建和训练参数量巨大的手绘分子识别模型。

1 相关工作

手绘化学分子结构图像生成是一个前沿的研究方向,目前相关的工作相对较少,但是与其类似的其他手写字符生成早已取得了长足的发展。因此,本节将结合手绘化学分子的生成研究和手写字符生成的研究展开介绍。

1.1 基于化学信息工具库的生成方法

Weir 等学者[14]在研究手绘化学分子结构识别的过程中,受限于数据的匮乏,开发了一种基于 RDKit 化学信息工具库的图像生成方法。RDKit 原本是一个开源的化学信息学工具包,广泛用于分子建模、药物设计、化学信息分析和机器学习等领域。可以通过解析一个化学分子的 SMILES(Simplified Molecular Input Line Entry System)表示[15]来还原分子的 2D 结构图形。而他的具体方法就是修改了库中有关图像生成的代码,通过随机化绘图参数以及添加各种噪声背景和图像增强的方式来实现仿真手绘分子图像的生成。受此启发 Brinkhaus 等学者[16]开发了更全面的生成工具,该方法生成图像时首先随机地在 3 种化学信息开发工具库 CDK[17]、RDKit[18]和 Indigo[19]中选择一个,然后通过伪随机设置绘图参数产生初始图像,接着随机地添加各种图像增强,包括标签和弯曲的箭头等,为了更好地实现不同生成参数的随机化,开发者设计了一个描述符指纹组,并通过 RDKit 的 MaxMin 算法实现[20]用于从所有有效指纹中挑选不同的组合样本来生成图像。

1.2 基于生成对抗网络的生成方法

Fogel 等学者[21]提出了 ScrabbleGAN 模型,是一种基于卷积神经网络的半监督的合成手写文字的方法,模型整体架构遵循 GAN 构架的基本范式。不同的是,除了判别器 D 之外,模型中还引入了一个文本识别器 R,将对生成的文本图像进行评估。一方面判别器 D 会有助于生成器 G 生成更为逼真的图像,另一方面识别器 R 则可用于保证结果可读且真实,该架构最大限度地减少了 2 个网络的联合损失。Luo 等学者[22]提出了一种新方法,可以基于生成对抗网络为任意长度和陌生文本生成参数化且可控的手写风格,称为 SLOGAN。具体来说,研究提出

了一个风格库，将特定的手写风格参数化为潜在向量，这些向量作为风格先验输入到生成器中，以实现相应的手写风格。风格库的训练只需用到具有特定作者风格的源图像，而不需要属性注释。此外，本次研究通过提供易于获得的印刷样式图像来嵌入文本内容，从而就能通过改变输入的印刷图像来灵活地生成多样化的内容。最后，生成器在双重判别器的指导下生成单字或者是一串字符的仿真草书风格图像，与之前的工作相比，该方法在生成网络的各项通用评估指数上都取得了显著提升。

2　方法

在本研究中，考虑到要充分利用生成网络双向生成图像的能力，以及受限于 2 个图像域数量不均衡的问题，本文以循环生成网络为基础网络，将 Tranformer 引入网络中的生成器，将 patchGAN 引入网络中的判别器，提出了 BiMIGAN 生成网络，同时利用预训练以及梯度惩罚的方法实现模型的快速稳定训练，以及模型性能的提升。

2.1　BiMIGAN 网络模型

BiMIGAN 网络模型使用类循环生成对抗网络的结构。模型整体结构设计如图 1 所示。BiMIGAN 模型是一个类 CycleGAN 的生成模型，模型由 2 个生成器和 2 个判别器组成，通过循环生成以及对比学习实现 2 个非配队图像域之间的相互转换。设 2 个不同的图像域分别为 X 和 Y，生成器 $G_{X\to Y}$ 尝试将图像从 X 域转换到 Y 域中，使其看起来像 Y，而判别器 D_B 则需要分辨出转换后的图像和真实图像。类似地，另一个生成器 $G_{Y\to X}$ 用于将域 Y 中的图像转换为域 X 中的图像，判别器 D_A 用于区分 X 中的真实图像与生成器 $G_{Y\to X}$ 伪造的假图像。

模型中的生成器和判别器分别是一个神经网络，BiMIGAN 模型与传统循环生成的不同之处是使用了卷积神经网络与 Transformer 混合结构的生成器和 PatchGAN 判别器，并且采用了更先进的梯度惩罚方机制来优化模型训练。

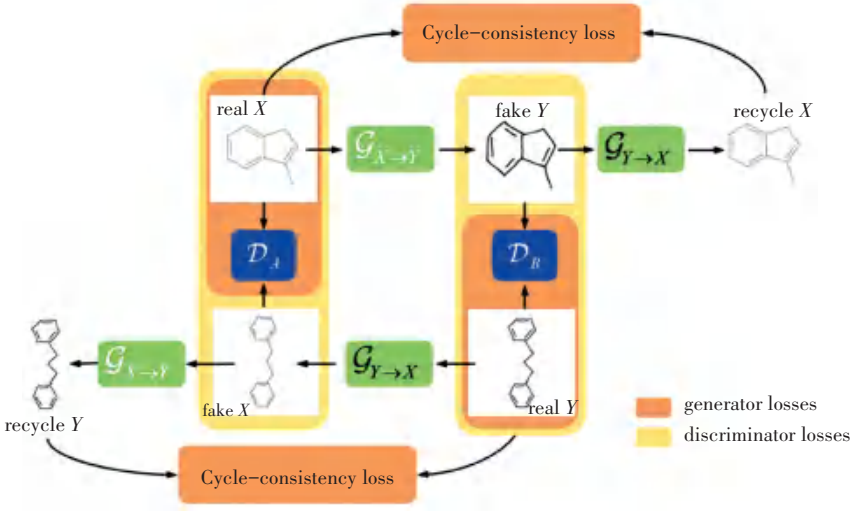

图 1　BiMIGAN 整体网络结构

Fig. 1　BiMIGAN network structure

2.2　像素级 Tranformer 生成器

模型中的生成器采用了一种基于 U-Net 架构的设计，但其瓶颈部分的中间层被替换为多个像素级 Vision Transformer（ViT）模块，如图 2 所示。U-Net 编码器通过 4 个卷积层和下采样层提取图像特征，每一层的特征图通过跳跃连接传递到对应的解码器层。最低层次的特征被输入到 ViT 模块中。

在 U-Net 编码过程中，原始图像被转换为维度为 (w_0, h_0, f_0) 的特征张量。每个卷积层使用核大小为 $k=3$、$p=1$ 的卷积核，下采样则使用核大小为 $k=2$、$s=2$ 的卷积核。因此，随着每一层的特征提取，特征图的宽度和高度减半，而特征维度在第一层保持不变，在接下来的 3 层中翻倍。最终，编码器输出一个维度为 $(w, h, f)=(\frac{w_0}{16}, \frac{h_0}{16}, 8f_0)$ 的特征张量，作为 ViT 的输入维度。

ViT 由一系列 Transformer 编码器模块组成。为了适应 ViT 的输入，U-Net 编码器输出的特征张量被展平为一个 token 序列。每个 token 序列的长度为 $w\times h$，每个 token 是一个长度为 f 的向量。为了增强 Transformer 的收敛性，研究中采用了 ReZero 正则化技术，并引入了一个可训练的缩放参数 α，用于

调整残差块中非主干分支的幅度。Transformer 的输出被投影回维度 f，并重新调整为宽度 w 和高度 h。

在本研究中，共使用了 12 个 Transformer 编码器模块，并设置 $f = 384$。

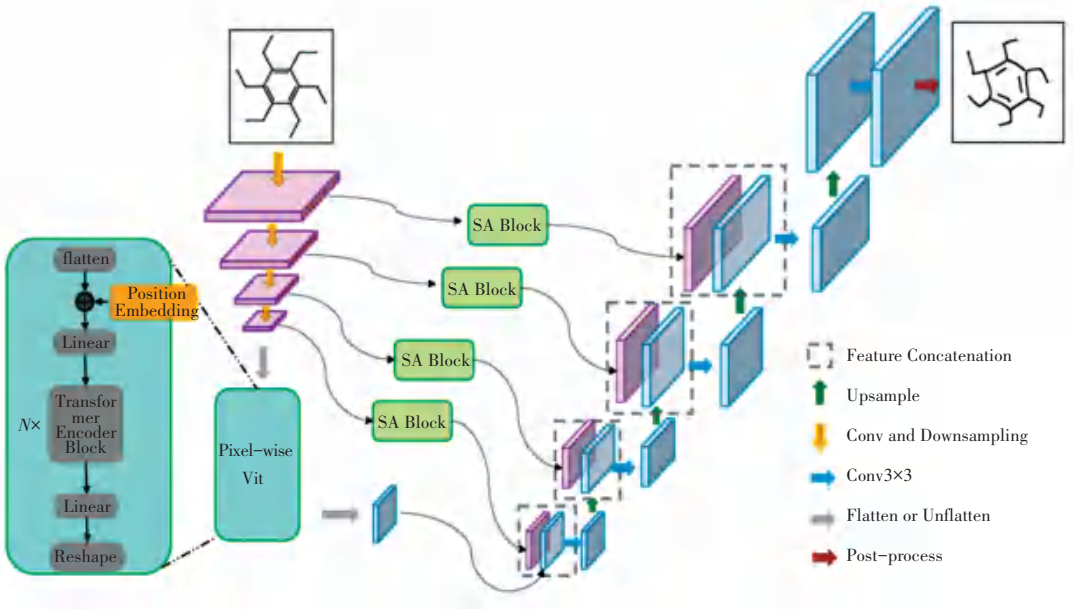

图 2 BiMIGAN 网络中的生成器结构

Fig. 2 Generator of BiMIGAN

2.3 判别器与梯度惩罚机制

模型中的判别器则是采用基础的 PatchGAN 判别器[23]，将输入图像分割成多个重叠或不重叠的小块，然后对每个小块进行真实性判断。这允许模型学习更加细粒度的细节，有助于提高生成图像的质量和逼真度。PatchGAN 判别器采用全卷积神经网络结构，这种设计不仅使网络能够处理任意大小的输入图像，而且通过共享权重减少了参数量，提高了效率。

本研究使用最小平方损失以及梯度惩罚（Gradient Penalty，GP）[24] 来加强判别器的性能。判别器通过最大化判别真实图像的概率以及最小化判错假图像的概率来更新损失函数。对于 X 图像域来说，损失函数可以定义为：

$$\mathcal{L}_{D_X} = \frac{1}{2}E_{x\sim p_{\text{data}}(x)}[(D_X(x)-1)^2] + \frac{1}{2}E_{\hat{x}\sim p_{\text{fake}}(\hat{x})}[(D_X(\hat{x}))^2] \tag{1}$$

梯度惩罚是指通过在判别器损失函数中添加一个正则项直接约束判别器的梯度范数，满足 Lipschitz 约束[25] 的限制，从而防止判别器过度优化（即梯度爆炸）以及避免生成器梯度消失（即判别器过于强大，导致生成器无法学习）。本模型中，采用的 GP 可以定义为：

$$\mathcal{L}_{D_X}^{GP} = \mathcal{L}_{D_X} + \lambda_{GP}E\left(\frac{\| \nabla_x D_A(x) \|_2 - \gamma}{\gamma}\right)^2 \tag{2}$$

其中，λ_{GP} 表示超参数，用于控制梯度惩罚的程度；E 表示 x 的期望分布；$\| \nabla_x D_A(x) \|_2$ 表示输入 x 时判别器梯度的 L_2 范数；γ 是一个超参数，表示目标范数。

2.4 自监督预训练

预训练是一种为大型网络在下游任务训练奠定良好的基础的有效方法[26]，使用预训练初始化要训练的模型并进行微调通常结果会显著优于随机初始化。在本研究中，对 BiMIGAN 的生成器进行了图像修复任务的预训练。具体来说，将图像分割为不重叠的 32×32 大小的 *patch*，并随机用掩码覆盖将其中 40%的 *patch* 的像素值设为零。生成器通过预测原始未掩码图像，并使用 L_1 损失函数来进行训练。本研究中考虑了 2 种预训练模型：

（1）在标准印刷体分子图像数据集上进行预训练；

（2）在 ImageNet 数据集[27] 上进行预训练。

在第 5 节中，本研究对使用这 2 种预训练模型以及针对无预训练的情况进行了消融实验。

3 实验设计

3.1 数据集

本研究使用的数据集在不同训练阶段也有所不

同。在预训练阶段使用的数据集有 2 个。一个是 ChEBML 标准分子数据集，数据来自 ChEMBL 化学小分子数据库，共计 190 万条小分子 SMILES 字符串，研究选用工具对这些字符串进行解析并生成标准化学分子图像；另一个是 ImageNet 预训练数据集，这是一个公开的图像预训练数据集，在图像相关任务中获得广泛使用。在微调阶段，使用的是研究团队自建数据集 HNUCM-HDM。该数据集收集了来自真人手绘的 9134 张手绘分子图像，这些图像中的分子类型包含碳氢两种元素、双键三键等多种化学键形式，苯环、稠环、桥环等多种类型的分子环状结构，数据类型分布广泛具有实用研究价值。数据集中的真实手绘分子图像按照 8∶2 的比例划分为训练集和测试集，同时手绘分子对应的标准印刷分子将作为标准图像域的训练和测试数据，由于标准分子图像与手绘分子图像之间是一对多的映射关系，因此通过图像增强的手段将标准图像域的数据集扩充至与手绘分子图像数量相等。

3.2　实验环境、超参数及评价指标

本文所有实验的运行环境见表 1，超参数见表 2。

表 1　实验环境

Table 1　Experimental environment

CPU	GPU	Python version	CUDA version	Torch version
AMD Ryzen 7 5800	NVIDIA RTX4090	3.9	12.0	2.0.0

表 2　超参数

Table 2　Hyperparameters

优化器	初始学习率	Epoch	Batch Size	λ_{GP}	γ
Adam	0.000 1	500	1	0.8	0.1

学习率的更新在预训练阶段采用线性递减的策略，在微调阶段，训练的前半段保持不变，后半段采用线性递减策略。

在分子图像生成任务中，为了全面评估模型的生成性能，本文使用了生成研究领域中常见的评价指标：*FID*（Fréchet Inception Distance）[28] 和 *KID*（Kernel Inception Distance）[29]。这 2 个指标衡量生成图像的特征相似性和多样性。从而反映模型的生成质量。

FID 通过比较生成图像和真实图像在使用 Inception V3 模型提取的深层特征上的统计分布，来衡量两者之间的相似性。设生成图像的特征向量的均值和协方差矩阵分别为 μ_g 和 Σ_g，真实图像的特征向量的均值和协方差分别为 μ_r 和 Σ_r。则可使用下式进行计算：

$$FID = \|\mu_g - \mu_r\|_2^2 + T_r(\Sigma_g + \Sigma_r - 2(\Sigma_g\Sigma_r)^{\frac{1}{2}}) \quad (3)$$

其中，$\|\mu_g - \mu_r\|_2^2$ 表示均值向量差的平方的 L_2 范数；T_r 表示矩阵的迹，$(\Sigma_g\Sigma_r)^{\frac{1}{2}}$ 表示协方差矩阵 Σ_g 和 Σ_r 的乘积的平方根。

KID 使用核方法（如多项式核或高斯核）计算生成图像特征分布与真实图像特征分布之间的距离，这是一种无偏估计的方法，因此更加贴近人眼感知的结果。设 X 为生成图像特征，Y 为真实图像特征，核函数用 k 表示，则数学计算公式具体如下：

$$KID(X,Y) = \frac{1}{n(n-1)}\sum_{i\neq j}k(x_i,x_j) + \frac{1}{m(m-1)}\sum_{i\neq j}k(y_i,y_j) - \frac{2}{nm}\sum_{i,j}k(x_i,y_j) \quad (4)$$

此外，为了更加准确评估模型生成图像的仿真性能，本研究进行了真人测试图灵实验。测试中，志愿者被要求观察一系列图像，这些图像由机器生成的分子图像和真实手绘分子图像随机组成。参与者需要在 6 s 的观察时间内，凭直觉判断每张图像是模型生成的、还是真人手绘。

为了方便量化统计图灵测试的结果，研究中定义了辩真率（True Classification Rates，*TCR*）来计算不同方法 i 生成的图像被认定为真的频率，其计算公式为：

$$TCR_i = \frac{T_i}{N_i} \quad (5)$$

其中，N_i 表示测试者针对生成类别 i 的测试样本总数，T_i 表示测试者针对生成类别 i 的回答为真的样本数。

这些统计数据能够直观地展示与了解测试者在区分生成图像与真实图像时的迷惑程度，从而反映生成方法在模拟真人手绘方面的表现差异。

在第二阶段的识别测试中，研究中采用了 OCSR 领域最常用的评估参数：完全匹配率（Exact Match Rate，*EMR*）和 Tanimoto 指数[11]，用来评估生成图像对于识别模型性能的影响。

OCSR 任务的输出结果通常是一个 SMILES 字符串，因此，完全匹配是根据模型预测的 SMILES 是否与真实的 SMILES 编码完全一致来判定。如果模型生成的 SMILES 与 GT 完全相同，则 *EM* 值为 1，反之则为 0。完全匹配率是完全匹配结果的均值，可通过下式计算求出：

$$EMR = \frac{Num_{EM}}{Num_{\text{total}}} \tag{6}$$

其中，Num_{EM} 表示测试结果完全匹配的数量，Num_{total} 表示测试集中图像数量。

EMR 作为一种严格的评价标准，能直观地反映识别模型是否精确识别了分子图像。

Tanimoto 指数是一种常用的相似性衡量方法，通过比较模型生成的 SMILES 编码与真实的 SMILES 编码之间的共同特征来计算得到。具体而言，Tanimoto 系数的计算基于 2 个集合的交集与并集，其计算公式如下：

$$Tanimoto(S_{\text{pred}}, S_{\text{gt}}) = \frac{| S_{\text{pred}} \cap S_{\text{gt}} |}{| S_{\text{pred}} \cup S_{\text{gt}} |} \tag{7}$$

其中，S_{pred} 表示模型生成的 SMILES 编码所对应的分子特征集合，S_{gt} 表示真实 SMILES 编码的分子特征集合。*Tanimoto* 系数的值介于 0 到 1 之间，当 *Tanimoto* 系数为 1 时，说明模型生成的 SMILES 编码与真实编码完全相同。由于 *Tanimoto* 系数能够反映分子结构的相似程度，即使模型未识别出完全正确的编码，该指标也能用于评估模型在识别复杂分子结构时的性能。

4 实验结果分析

4.1 手绘分子图像生成实验

在第一节生成相关实验中，本研究与目前主流的手绘分子结构图像生成工具 RDKit[18] 和 RanDepict[17] 以及原始的 CycleGAN[30] 进行了对比。这些方法或工具使用了各不相同的生成策略，包含图像增强方法和深度学习方法。通过与现有的广泛使用的生成工具以及更早的深度学习生成方法比较，能够从多方面评估模型在生成仿真手绘分子图像的表现，同时体现不同方法的差异性与实用性。4 种不同方法的对比结果见表 3。本研究中提出的 BiMIGAN 模型的 FID 值和 KID 值远低于其他的方法，分别为 20.809 和 0.006 5(±0.000 3)。

表 3 不同生成方法的机器评估指标对比

Table 3 Comparison of machine evaluation metrics for different generation methods

方法	评价指标	
	FID ↓	*KID* ↓
RDKit tool	45.252	1.532 1(±0.002 8)
RanDepict	58.765	1.743 3(±0.001 1)
CycleGAN	131.674	2.574 8(±0.119 2)
BiMIGAN(ours)	**20.809**	**0.006 5(±0.000 3)**

4 种方法的生成结果样例展示在图 3 中，从上至下，分子的结构逐渐复杂，最左边的一列是标准打印体分子，最右边一列则是真人手绘的分子图像。

图 3 中除了 CycleGAN 方法生成的图像与真实手绘分子差距较大之外，其他工具生成的手绘分子图像均在一定程度上与真人手绘分子相似，这与机器评估指标的结果一致。CycleGAN 作为深度学习的方法并没有取得更好的生成结果，究其原因可能在于是训练过程中损失动荡难以收敛，最终在相同的训练轮次下产生了模式崩溃，与 BiMIGAN 模型结果相比印证了梯度惩罚的重要性以及 Tranformer 生成器的优异性。

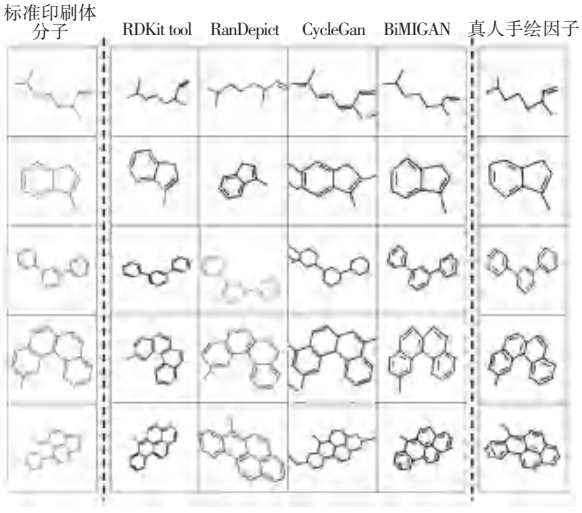

图 3 不同方法生成手绘分子图像结果对比

Fig. 3 Comparison of different methods for generating hand-drawn molecules

值得一提的是，BiMIGAN 模型是一个双向图生图模型，这意味着模型不仅可以将标准分子转化为手绘分子，还可以将手绘分子逆向还原为标准分子，具体生成样例展示在图 4 中。这种手绘转标准的方法是后续识别研究新方法的基础。

真人图灵测试结果如图 5 所示，实验结果表明 BiMIGAN 模型在 *TCR* 数据上最接近真实手绘数据结果。在所有真实手绘图像样本中，有 80.14% 的测试者认为真实手绘图像是真人手绘图像，在 BiMIGAN 图像测试样本中有 76.45% 的测试者认为模型生成的仿真图像是真人手绘图像，与真实图像相比差距仅 3% 左右，这一比例远远超过其他生成工具。真实手绘图像的 *TCR* 不足 100% 是因为真实图像和生成图像混合在一起进行的测试，同时测试者被告知了其中有假图像，这使得部分测试者会误把写得过于标准的图像认定为仿真图像或者是把书写得过于潦草的图像认定为工具生成出错的图像。

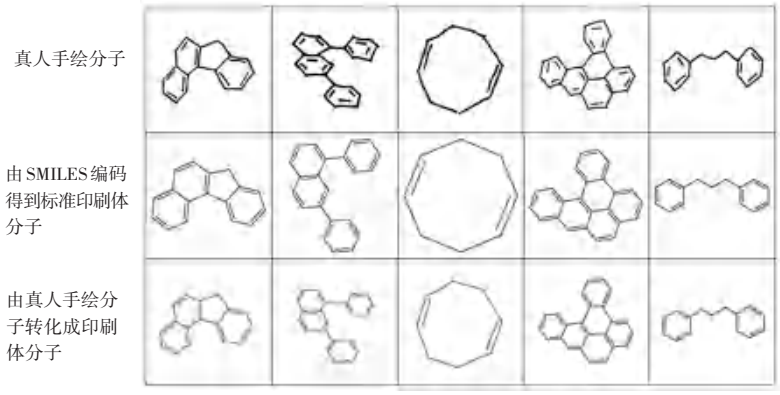

图 4　BiMIGAN 标准化手绘分子图像实例

Fig. 4　Samples of standardized hand-drawn molecules by BiMIGAN

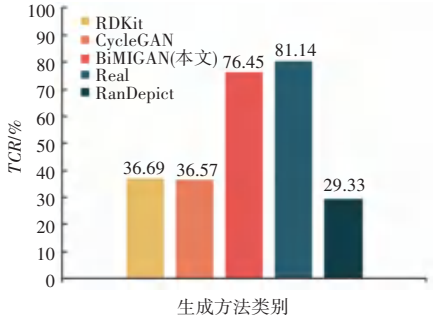

图 5　真人图灵测试结果对比图

Fig. 5　Comparison of Turing test results

4.2　手绘分子识别实验

考虑到将手绘分子转化为标准分子再利用现有的高效识别模型进行识别是一种省时省力的方法。因此，本研究还设计了仿真实验，测试了 4 种不同的化学分子识别模型分别识别用 BiMIGAN 转化前和转化后的图像的准确率以及 Tanimoto 指数，具体的测试结果见表 4。

无论是用于识别标准分子图像的模型（如 MolScribe），还是用于识别手绘分子图像的模型（如 DECIMER2.0），先将手绘图像经 BiMIGAN 模型转换为标准分子图像后再识别，模型的识别结果都比直接识别原始手绘图像时有显著提升。对于专门用于识别标准分子图像的模型，这一提升尤为明显，最高提升接近 40%。这也突显了所提出方法的优势，即能够将原本无法识别的手绘图像转换为现有模型可以成功识别的格式。

表 4　手绘图像标准化前后 OCSR 模型性能比较

Table 4　Performance comparison of OCSR model before and after image standardization

模型名称	*EMR* /% ↑		*Tanimoto* ↑	
	原始图像	标准化后图像	原始图像	标准化后图像
MolScribe[12]	46.45	70.54	0.607 1	0.808 7
SwinOCSR[31]	27.39	73.19	0.552 9	0.880 1
DECIMER 2.0[10]	71.66	84.37	0.816 9	0.909 3
MolecularDet[32]	79.35	81.98	0.848 4	0.878 0

本文的研究结果证明了通过标准化手绘分子图像、并充分利用现有的高性能模型的方式实现高精度的手绘分子图像识别是一个解决高精度识别手绘化学分子模型的切实可行、并具潜力的研究方向。

5　消融实验

为了进一步验证模型的有效性和可靠性，设计并进行了消融实验，消融实验结果见表 5。

表 5　消融实验结果

Table 5　Results of ablation experiment

预训练模型	*GP*	*FID* ↓	*KID* ↓
印刷体分子图像	√	20.809	0.007(±0.000 3)
ImageNet	√	25.226	0.011(±0.000 4)
None	√	55.465	0.018(±0.000 9)
印刷体分子图像	×	58.080	0.019(±0.002 4)
ImageNet	×	63.130	0.020(±0.001 7)
None	×	93.980	0.056(±0.001 5)

消融实验的结果表明 *GP* 结合印刷体分子图像

预训练的方式能取得最佳的生成性能,2 种预训练模型无论在有梯度惩罚、还是没有梯度惩罚的情况下,都能对模型生成性能产生好的影响,但是 ImageNet 通用图像预训练模型的效果明显逊色于和手绘分子相似的印刷体分子预训练模型的效果好,但差距不大。从是否使用 *GP* 产生的巨大差距可以推断,在使用预训练网络时,梯度惩罚是获得最佳性能的必要条件,因为使用预训练模型初始化生成器为图像生成任务提供了一个良好训练起点。然而,判别器在初始化时使用的是随机初始化,这意味着在训练初始阶段,判别器可能向生成器传递无意义的随机信号,这种随机信号会使生成器偏离良好的起点,从而削弱预训练的优势,梯度惩罚则能较好地控制判别器损失在一个范围内波动,从而传递更少无意义的随机信号。

6 结束语

本文提出了 BiMIGAN 生成模型,用于进行手绘分子结构图像和标准印刷体分子图像的双向生成。模型中的生成器采用了引入像素级 Transformer 模块作为中间层的 U-Net 生成网络,判别器采用了使用全卷积神经网络的 PatchGAN 判别器,通过采用自监督的预训练以及梯度惩罚的训练手段,使得模型的训练更加稳定。实验结果表明,文中提出的方法有效地提升了生成手绘分子图像与真实手绘图像的相似程度,真人测试的结果则进一步说明了模型生成图像以假乱真的程度。此外,通过下游的手绘分子识别任务,验证了本模型生成的手绘分子图像在提升识别模型性能上具有与真实分子具有相近的效果,还提出并论证了通过标准化手绘分子图像再使用现有模型识别的新型方法的可行性,为未来的手绘分子识别与研究提供了新的研究方向。

参考文献

[1] RAJAN K, BRINKHAUS H O, ZIELESNY A, et al. A review of optical chemical structure recognition tools [J]. Journal of Cheminformatics, 2020, 12(1):60.

[2] LI Qingliang, CHENG Tiejun, WANG Yanli, et al. PubChem as a public resource for drug discovery[J]. Drug Discovery Today, 2010, 15(23-24): 1052-1057.

[3] GAULTON A, OVERINGTON J P. Role of open chemical data in aiding drug discovery and design [J]. Future Medicinal Chemistry, 2010, 2(6): 903-907.

[4] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge[J]. Nature, 2017, 550(7676): 354-359.

[5] 施思齐, 涂章伟, 邹欣欣, 等. 数据驱动的机器学习在电化学储能材料研究中的应用[J]. 储能科学与技术, 2022, 11(3): 739.

[6] CLEVERT D A, LE Tuan, WINTER R, et al. Img2Mol - accurate SMILES recognition from molecular graphical depictions [J]. Chemical Science, 2021, 12(42): 14174-14181.

[7] STAKER J, MARSHALL K, ABEL R, et al. Molecular structure extraction from documents using deep learning [J]. Journal of Chemical Information and Modeling, 2019, 59(3): 1017-1029.

[8] SUN Chen, SHRIVASTAVA A, SINGH S, et al. Revisiting unreasonable effectiveness of data in deep learning era [C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 843-852.

[9] RAJAN K, ZIELESNY A, STEINBECK C. DECIMER 1.0: deep learning for chemical image recognition using transformers [J]. Journal of Cheminformatics, 2021, 13(1): 16.

[10] RAJAN K, BRINKHAUS H O, ZIELESNY A, et al. Advancements in hand-drawn chemical structure recognition through an enhanced DECIMER architecture [J]. Journal of Cheminformatics, 2024, 16(1): 78.

[11] BRINKHAUS H O, ZIELESNY A, STEINBECK C, et al. DECIMER-hand-drawn molecule images dataset[J]. Journal of Cheminformatics, 2022, 14(1): 36.

[12] QIAN Yujie, GUO Jiang, TU Zhengkai, et al. MolScribe: Robust molecular structure recognition with image-to-graph generation[J]. Journal of Chemical Information and Modeling, 2023, 63(7): 1925-1934.

[13] ZHANG Xiaochen, YI Jiacai, YANG Guoping, et al. ABC-Net: A divide-and-conquer based deep learning architecture for SMILES recognition from molecular images [J]. Briefings in Bioinformatics, 2022, 23(2): bbac033.

[14] WEIR H, THOMPSON K, WOODWARD A, et al. ChemPix: automated recognition of hand-drawn hydrocarbon structures using deep learning[J]. Chemical Science, 2021, 12(31): 10622-10633.

[15] WEININGER D. SMILES, a chemical language and information system [J]. Journal of Chemical Information and Computer Sciences, 1988, 28(1): 31-36.

[16] BRINKHAUS H O, RAJAN K, ZIELESNY A, et al. RanDepict: Random chemical structure depiction generator [J]. Journal of Cheminformatics, 2022, 14: 31.

[17] STEINBECK C, HAN Yongquan, KUHN S, et al. The Chemistry Development Kit (CDK): An open-source Java library for chemo-and bioinformatics[J]. Journal of Chemical Information and Computer Sciences, 2003, 43(2): 493-500.

[18] BENTO A P, HERSEY A, FÉLIX E, et al. An open source chemical structurecuration pipeline using RDKit [J]. Journal of Cheminformatics, 2020, 12: 51.

[19] DRISCOLL M, BROCK B, ONG F, et al. Indigo: A domain-specific language for fast, portable image reconstruction [C]// 2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Piscataway, NJ: IEEE, 2018: 495-504.

[20] ASHTON M, BARNARD J, CASSET F, et al. Identification of diverse database subsets using property-based and fragment-based molecular descriptions [J]. Quantitative Structure-Activity Relationships, 2002, 21(6): 598-604.

[21] FOGEL S, AVERBUCH-ELOR H, COHEN S, et al. Scrabblegan: Semi-supervised varying length handwritten text generation [C]//Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 4324-4333.

[22] LUO Canjie, ZHU Yanzhi, JIN Lianwen, et al. SLOGAN: handwriting style synthesis for arbitrary-length and out-of-vocabulary text[J]. arXiv preprint arXiv, 2202.11456, 2022.

[23] CHEN Gang, ZHANG Guipeng, YANG Zhenguo, et al. Multi-scale patch-GAN with edge detection for image inpainting[J]. Applied intelligence, 2023, 53(4): 3917-3932.

[24] GAO Xin, DENG Fang, YUE Xianghu. Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty[J]. NeuroComputing, 2020, 396: 487-494.

[25] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM, 2017: 214-223.

[26] BAO Hangbo, DONG Li, PIAO Songhao, et al. Beit: Bert pre-training of image transformers[J]. arXiv preprint arXiv, 2106.08254, 2021.

[27] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248-255.

[28] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium[J]. arXiv preprint arXiv, 1706.08500, 2017.

[29] BIŃKOWSKI M, SUTHERLAND D J, ARBEL M, et al. Demystifying mmdgans[J]. arXiv preprint arXiv, 1801.01401, 2018.

[30] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 2223-2232.

[31] XU Zhangpeng, LI Jianghua, YANG Zhaopeng, et al. SwinOCSR: End-to-end optical chemical structure recognition using a Swin Transformer[J]. Journal of Cheminformatics, 2022, 14(1): 41.

[32] TAO Jiajun, LIU Wei, PENG Xiaowang, et al. Recognition of hand-drawn hydrocarbon structure formulas using anchor-free detector[C]//Pacific Rim International Conference on Artificial Intelligence. Cham: Springer, 2024: 98-110.