基于 YOLOv10 的 AIGC 图文生成优化与反馈修正方法

刘嗣久[1] 荆晓远[2] 任娟[1] 姚永芳[2] 孙其航[1]
LIU Sijiu JING Xiaoyuan REN Juan YAO Yongfang SUN Qihang

摘 要 近年来随着人工智能生成内容（AIGC）的发展，图文生成技术取得了显著的进步。然而由于现有大模型本身的技术问题，会存在复杂 prompt 描述下生成的图片与文本内容不一致且目标完整性不足的问题。为此，文章提出一种基于 YOLOv10 目标检测和反馈修正机制的 AIGC 图文生成优化方法。首先，对大模型输入文本指令提示词，让其根据文本指令生成初步图像；接着，应用 YOLOv10 对生成的初步图像进行目标检测，检测的对象即为文本内容中突出的几个对象；然后将检测到的问题图像和标注信息反馈给扩散模型，扩散模型根据这些反馈对图像进行局部修复，修正缺失的目标或调整目标的位置。实验结果表明，该方法能够显著提高生成图像的目标完整性和与文本描述的对象一致性，与传统方法相比，该机制能够有效地处理生成图像中的问题。其方法为 AIGC 领域中图像生成与优化提供了一种新的思路和方法。

关键词 YOLOv10；人工智能生成内容（AIGC）；扩散模型；目标检测；图像生成；反馈修正

doi：10.3969/j.issn.1672-9528.2025.02.036

0 引言

随着人工智能技术的迅猛发展，人工智能生成内容（artificial intelligence generated content，AIGC）已在各个领域广泛应用，并在图像生成、文本处理、视频制作、音频改进 [1] 等方面取得了长足进步。而大语言模型（large language models，LLM）作为 AIGC 的重要类型，其在图文生成能力方面取得了巨大的突破。例如，美国 OpenAI 公司开发的 ChatGPT-4、清华大学的 ChatGLM、百度公司的文心一言、科大的讯飞星火等，这些模型可以根据用户所输入的提示词描述在几秒钟内生成对应的图像。

然而，尽管现有的生成模型在生成图像和文本处理方面表现出色，但在图文一致性方面仍存在不足。例如在 prompt 文本描述涉及多个具体对象和场景关系的前提下，大模型生成的图像可能会遗漏重要对象，或者对象的位置和布局出现偏差，严重影响图文的适用性和实际效果。

针对模态失配这一问题，本文提出了一种基于 YOLOv10 目标检测与反馈修正机制的 AIGC 图像优化方法，旨在提高生成图像与文本提示的一致性。该方法先通过大语言模型根据 prompt 文本指令生成初步图像，然后用 YOLOv10 对初步图像进行目标检测，识别文本描述中的关键对象，从而筛选出有问题的图片。最后通过将有问题的图片反馈给扩散模型，从而有效修正生成过程中出现的目标缺失或位置不当的问题。接下来，本文将详细介绍该方法的实现过程、实验设置以及结果分析，探讨该机制的实际应用效果。

1 相关工作

1.1 AI 生成图像技术研究现状

在早期人工智能生成图像的研究中，主要集中于传统生成对抗网络（generative adversarial network，GAN）领域，比如 GAN-CLS、StyleGAN[2] 等算法在文本生成图片，图片风格多样化等方面展现出了优秀的性能。然而，GANs 网络生成图像也存在一些缺点，比如生成器和解码器之间存在对抗博弈，因此会导致训练过程中经常不稳定，容易出现模式崩溃且生成的多样性不足等问题。而随着 Huang 等人 [3] 提出 Transformer 架构以及近年来扩散模型的出现，一些基于扩散模型（diffusion models）生成图像的模型取得了显著的成就，如 DALLE·2[4]、GLIDE[5] 等，这些模型在图像生成的测试中效果明显优于传统 GAN。

扩散模型主要分成正向扩散过程和逆向去噪过程 [6] 两部分，基本结构如图 1 所示。正向扩散过程就是不断对原始数据加入高斯噪声，图 1 中从右向左 $X_0 \rightarrow X_T$ 的过程就是对图片逐渐加噪，让图片中原始数据的分布转换为一个简单的标准高斯分布。

1. 吉林化工学院信息与控制学院 吉林吉林 132000
2. 广东石油化工学院计算机学院和省市共建石化装备智能安全广东省重点实验室 广东茂名 525000
[基金项目] 国家自然科学基金项目“基于类不平衡深度特征学习的石化动设备故障信号分类研究”（62176069）

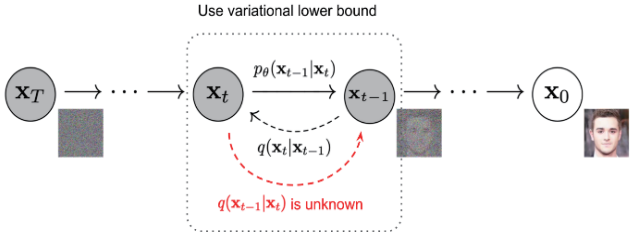


图 1 扩散模型的基本结构图

正向扩散过程本质上是一个马尔科夫过程，即每一步的噪声添加由预定义的噪声分布决定。公式表示为：

$$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{(1-\beta_t)}x_{t-1}, \beta I) \quad (1)$$

$$q\left(x_{1:T} \mid x_0\right) = \prod_{t=1}^{T} q(x_t \mid x_{t-1}) \quad (2)$$

式中：β_t 表示噪声添加系数；x_t 表示第 t 步的噪声数据；$\beta_t \in (0,1)$ 是人为设置的常数值，不同 t 的 β_t 是预先定义好的逐渐衰减。满足 $\beta_T > \cdots > \beta_2 > \beta_1$。

通过式（1）将小的高斯噪声加到 x_{t-1} 的分布中，得到 x_t。通过计算可得：

$$\begin{aligned} x_t &= \sqrt{\beta_t}x_{t-1} + \sqrt{1-\beta_{t-1}}z_{t-1} \\ &= \sqrt{\beta_t\beta_{t-1}}x_{t-2} + \sqrt{1-\beta_t\beta_{t-1}}z_{t-2} \\ &= \cdots \\ &= \sqrt{\overline{\beta}_t}x_0 + \sqrt{1-\overline{\beta}_t}z \end{aligned} \quad (3)$$

式中：$Z_{t-1}, Z_{t-2}, \cdots, \in \sim N(0,1)$ 为高斯噪声，推导出的结果为 $\overline{\beta}_t = \prod_{x=1}^{t}\beta_x$。根据式（3），结合式（1）和（2），可以得到 x_t 和 x_0 的关系为：

$$q(x_t \mid x_0) = N(x_t; \sqrt{\overline{\beta}_t}x_0, (1-\overline{\beta}_t)I) \quad (4)$$

由式（4）可以看到，在正向过程中，当 T 足够大时，x_t 可以完全收敛于高斯噪声。

逆扩散是一个逐步去噪的过程，从标准高斯分布中进行采样，每一步去噪将一个较小的高斯噪声数据转换为图像数据，进而得到其真实数据分布中的样本，以达到生成数据的目的 [7]。反向过程也是一个马尔科夫过程，通过 x_{t-1} 与给定的 x_t 分布，进行逐步去噪任意噪声以后，生成新的图像。当 $\beta_t \to 0$ 时，反向过程和正向过程具有相同的函数形式，无法直接得到 $p(x_{t-1}|x_t)$ 的分布，所以使用神经网络 θ 来拟合逆过程 $q(x_{t-1}|x_t)$。计算过程为：

$$p_\theta(x_0 : T) = p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t) \quad (5)$$

式（5）与式（2）同样是马尔科夫过程，因此二者公式形式相同。将式（5）通过 θ 拟合以后得到公式：

$$p_\theta(x_{t-1} \mid x_t) = N(x_{t-1}; \mu_\theta(x_t, t); \sum_\theta(x_t, t)) \quad (6)$$

虽然对 $q(x_{t-1}|x_t)$ 的结果未知，但是可以根据 $q(x_{t-1}|x_t)$ 来推断出 $q(x_{t-1}|x_t, x_0)$ 的值，对应的公式为：

$$q(x_{t-1} \mid x_t, x_0) = q(x_t \mid x_{t-1}) \times \frac{q(x_{t-1} \mid x_0)}{q(x_t \mid x_0)} \quad (7)$$

将式（1）和式（3）代入方程（7）后可得公式：

$$q(x_{t-1} \mid x_t, x_0) = N(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I) \quad (8)$$

扩散模型的特点是可以生成高质量的图片，与传统 GANs（生成对抗网络）相比，扩散模型在生成高分辨率、细节丰富的图像上具有优势，并能够对生成的图片进行控制 [8]。通过在生成过程中引入不同的条件，可以实现多样性和个性化。扩散模型还被应用于 3D 生成与建模、信号去噪、图像增强、语音合成等诸多领域。因此，本文在改进图像方面主要使用扩散模型方法对目标检测后的图像进行修改。

1.2 基于 YOLOv10 目标检测技术的研究现状

为了识别大模型初始生成图像中的文本标签词对象，需要对图像进行目标检测，且及时反馈给用户检测结果，提高实时检测精度与效率。由此，YOLOv10 作为开创性的实施目标检测方法被提上日程。

YOLO 系列算法最早是由 Joseph 等人 [9] 于 2016 年提出。它是一种单阶段目标检测算法，其核心思想是把目标检测任务简化为一个回归问题，直接从图像中预测目标的类别和位置，实现了较高的检测速度和较好的检测精度。截至 2024 年，YOLO 算法已经更新发展到了第 10 代，而本文采用最新的 YOLOv10 算法作为识别大模型生成的初始图像中目标对象的检测。

YOLOv10 网络结构如图 2 所示，主要是由输入层（Input）、主干网络（Backbone）、颈部网络（Neck）和头部网络（head）组成 [10]。

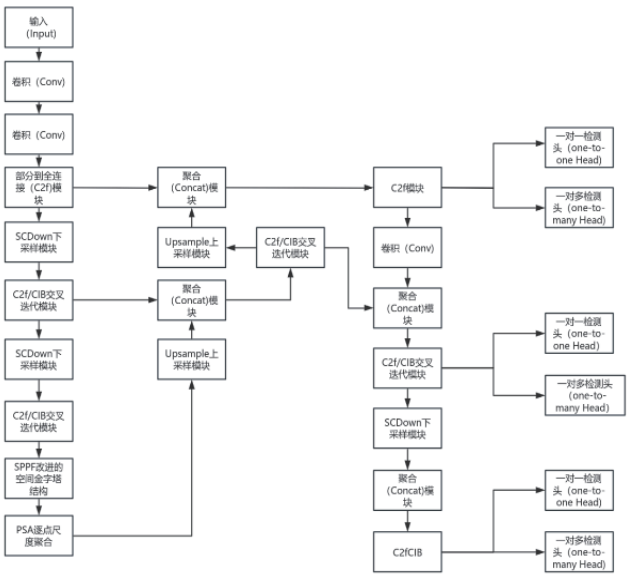


图 2 YOLOv10 网络结构图

其中，主干网络（Backbone）负责提取输入图像的特征；颈部网络（Neck）通过多尺度特征融合来增强检测能力，特

别是对于小目标、不同尺度目标的检测表现更好；头部网络（head）是生成最终的目标检测，包括边界框位置、类别以及置信度等。与传统 YOLO 相比，YOLOv10 引入了无须依赖非最大值抑制（NMS）的一致分配训练策略进行处理 [11]，可以轻量化设计更高效且适合实时检测的任务。

2 方法

本文方法通过 YOLOv10 实时目标检测与扩散模型（diffusion model）实时修正相结合，从而提高图片与文本内容的一致性。该机制可以称为 YOLO-DM，主要分为目标检测环节和图像修正环节。下面是对本文方法的具体介绍。

2.1 生成框架

YOLO-DM 对大模型生成的初始图片改进生成的整体框架如图 3，首先用户对大模型输入 prompt 文本指令，大模型根据文本指令生成初始图片，同时通过 LLM 解析器对 prompt 文本内容进行关键词提取 [12]。以图 3 为例，经过 LLM Parser[13] 对用户输入的文本内容分析，得出文本中指定的 3 个关键对象为“白色的狗”“蝴蝶”“猫”。经过 YOLOv10 对初始图片进行检测以后，得出的结果是图片中缺少蝴蝶，同时将缺少蝴蝶的情况反馈于扩散模型，扩散模型根据反馈问题进行修正，即对扩散模型输入增加一只蝴蝶的指令，对图片进行添加、删除、属性修改等操作，得到修正后的图片。

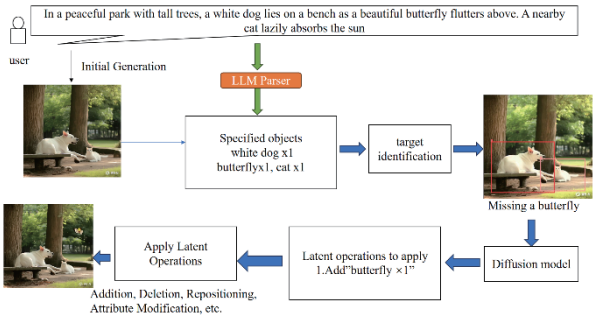


图 3 AIGC 图片检测与修正的过程结构图

2.2 文本条件下图像生成与 YOLOv10 检测的实时目标分析系统

本文所提出的实时目标分析系统可对大模型所生成的初始图片进行目标检测，及时检测出初始图片中的对象与用户所输入的提示词中关键对象标签词是否一致。而用户文本描述所说的关键对象标签词可通过 LLM 作为解析器，将用户所说的 prompt 文本描述传递到开放词汇表的对象检测器中进行筛选来实现。

本文对大语言模型输入的文本描述内容为：“在一个有高大树木的宁静的公园里，一只白色的狗躺在长凳上，一只美丽的蝴蝶在上面飞翔。附近的一只猫懒洋洋地晒着太阳。”LLM 解析器从用户提供的文本提示 prompt 中提取几个关键对象细节列表，表示为 O。解析器在文本指令和上下文示例的帮助下，可以很简单完成这一点，相关的分析结果如图 4 所示。

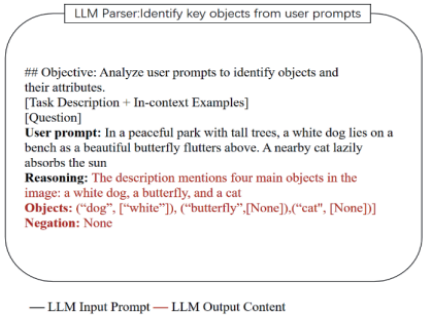
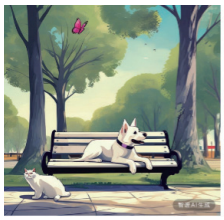


图 4 LLM 解析器根据 prompt 描述分析的结果

LLM 解析器通过解析用户提示符并输出与图像相关的关键短语，通过注意力机制选择文本中关键词，最终提取出包含“狗”“蝴蝶”“猫”的关键对象。同时，也分析出了 3 个关键对象的重要特征，比如狗需为白色，蝴蝶与猫没有特征要求。

2.2.1 数据的获取与预处理

本文选用的大语言模型为清华大学开发的智谱 AI，该大语言模型可以根据文本描述，在 3.5 S 内生成 200 张 256 px×256 px 的图像。首先调用了智谱 AI 的 API，接着将所想要的文本描述 prompt 输入智谱 AI 的对话框中，文本描述内容即 2.2 部分所说的 prompt，让 AI 生成 2000 张原始的图片，并保存到本地。图 5 展示了其中具有代表性的例子：图 5（a）展示了生成的图片基本符合提示描述，包含所有标签对象“狗”“猫”和“蝴蝶”；而图 5（b）大模型所生成的图片就存在标签对象“蝴蝶”的缺失。同时，生成的初始图片中存在多余的对象，比如文本描述中的“猫”“狗”“蝴蝶”均为一只，但是生成的初始图片中部分存在两只或者多只情况。因此需要对大模型生成的一系列初始图片进行标签对象的目标检测，筛选出目标对象缺失或者多余的图片。



（a）生成的拥有完整描述对象的图

（b）生成的缺失描述对象的图

图 5 智谱 AI 生成的部分代表性初始图

在对大模型生成的初始图片进行数据清洗训练后，将 2000 张图片按 8:1:1 的比例划分为训练集、验证集和测试集。其次，使用 Labellmg 工具对集中的图像进行标注，重点标注了图片中的“猫”“狗”和“蝴蝶”目标，标注信息被保存为与图像对应的格式文件。通过格式转换后，这些标注数据被用于整理模型训练。验证集和测试集则用于模型的性能评估和验证，确保训练过程中的有效监控和模型优化。

2.3 扩散模型依据反馈检测结果进行图像的修正

根据前面目标检测出的结果，可以将存在目标缺失或者多余的问题图片的情况反馈给扩散模型，扩散模型在每张图片中的特定边框中进行遮罩（mask image），以便后期根据修改指令对遮罩部分进行修改，未遮罩的部分保持原来的内容不变，通过向后扩散序列对遮罩部分进行处理，产生与遮罩相对应的掩蔽潜在层，与原始图片进行画布合并，从而实现图像的修正。扩散模型修正图片的流程示例如图 6 所示。

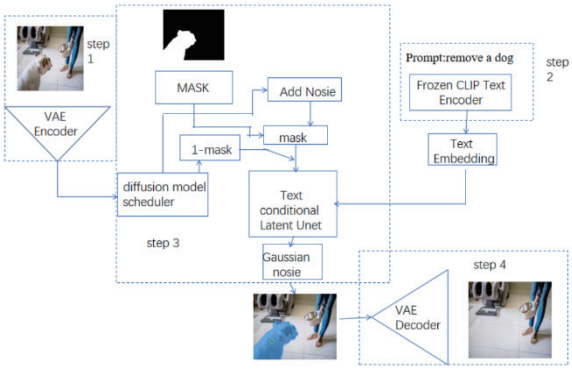


图 6 扩散模型根据反馈结果修改原始图片的流程图

3 实验结果与分析

3.1 相关参数

本实验所用的参数主要有：实例数量（Instances）、平均精度（mean Average Precision，mAP）、召回率（Recall）、准确率（Accuracy）。这些参数作为判定目标检测后图片优劣的指标。同时，本文采取 CLIP 相似度评分来判定修正以后的图片与初始文本描述的一致性。CLIP 得分越高，说明图像与文本描述匹配得更好。

3.1.1 目标检测结果分析

在模型训练中，设定训练 150 轮，每次训练 16 张图片，图像尺寸设定为 640 px×640 px，最终检测部分结果图如图 7 所示。

图 7 YOLOv10 对大模型生成的图片检测结果

从图 7 可以看出，YOLOv10 基本将每张图片中特定的对象全部检测，并且每张图像中狗均为白色，符合 LLM 解析器解析结果要求。虽然部分图像存在特定目标缺失或者多余的情况，但依照要求，多余的猫同时也被检测出。YOLOv10 训练好的验证集各项参数值表 1 所示。

表 1 验证集实验结果

类别	图像	数量	*R*	mAP50%	mAP50:95%
all	200	569	0.65	0.653	0.379
cat	200	213	0.61	0.545	0.310
butterfly	200	156	0.33	0.421	0.216
dog	200	200	1	0.995	0.611

依据上述实验结果，可以看出，根据 200 张图片数量，按照预期要求，每一张图片中都得有一只狗、一只蝴蝶，一只猫才符合要求，即 200 张图片中，3 个目标各自数量均为 200 才符合预期。而猫的数量是 213，蝴蝶的数量是 156。根据召回率 *R* 以及 mAP 值可以得知，狗的 *R* 值为 1，且 mAP 值远大于蝴蝶和猫的数值，这说明狗的数量表现良好，没有目标多余或者缺失的情况。而根据蝴蝶和猫的各项数值指标来看，猫的各项指标表现一般，而根据数量可以看出，猫检测出的数量远多于预测值，说明猫存在目标多余的情况。而蝴蝶存在目标缺失的情况。

基于此，本轮实验的混淆矩阵图如图 8 所示，根据混淆矩阵图，可以看出，蝴蝶的检测存在明显的误差，猫和狗的检测效果较好，说明大模型生成的图片有部分存在蝴蝶缺失的情况。

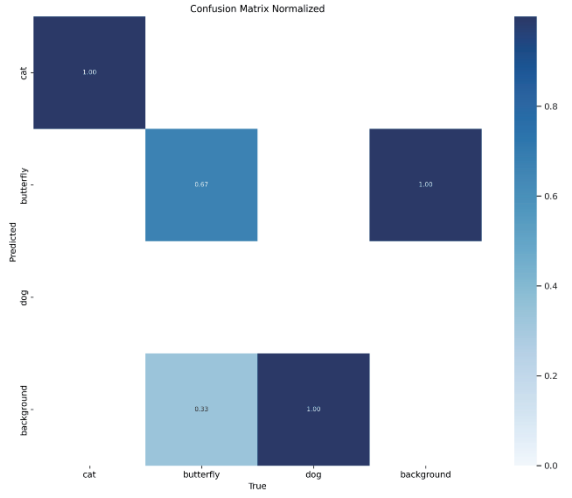


图 8 YOLOv10 混淆矩阵图

最终，根据检测结果进行筛选，共计筛选出存在问题的图像 57 张。其中缺失蝴蝶的图像共 44 张，多余猫数量的图片共 13 张。

3.1.2 图片的修正与图文评估

根据 YOLOv10 的检测结果，将存在问题的图片情况反

馈给扩散模型。本实验检测结果共存在两种问题，即部分图片缺失蝴蝶，部分图片猫咪多余。针对蝴蝶缺失的情况，对扩散模型输入修正指令，指令内容即在图片划定范围增加一只蝴蝶，最终扩散模型修正的实验结果示例图如图 9，图 9 是 44 张缺失蝴蝶图片中的其中一张示例图片，经过扩散模型修正以后的效果图。

图 9 Text prompt: "add a butterfly"

而对于存在多余猫的图片，可以对扩散模型输入移除指令，图 10 是 13 张存在多余猫咪图片中的其中一张示例，经过扩散模型移除多余对象后修正的效果图。可以看到，扩散模型接到反馈指令以后，对图片中其中一只猫框定了范围，并进行了移除。

图 10 Text prompt: "remove a cat"

经过扩散模型修正有问题的图片以后，对修正前后的图片与文本的一致性进行 CLIP score 评估，以验证修正后的图片与文本的一致性是否得以提高，文本即前文给大模型的 prompt 内容。表 2 是修正前后图片与文本相似度的评分对比。

表 2 修正前后 CLIP score 得分

	图片数量	CLIP score
存在问题的图片修前	57	0.292
存在问题的图片修后	57	0.702
修正前的所有图片	2000	0.507
修正后的所有图片	2000	0.741

CLIP 相似度得分范围通常为 0 到 1，由表 2 可以看到，根据前面 YOLOv10 目标检测出的 57 张存在多余猫和缺失蝴蝶的图片与 prompt 文本描述的 CLIP score 得分只有 0.292，说明这 57 张图片的内容与 prompt 文本描述内容整体一致性存在较大偏差，而通过扩散模型修正以后，57 张修正完以后的图片与 prompt 文本描述的 CLIP score 提高到了 0.702，图文一致性得到了大幅度提高。而再经过对智谱 AI 根据 prompt 描述生成的初始 2000 张图片进行 CLIP 相似度评分，可以看到，整体看来智谱 AI 根据文本描述生成的图像文本一致性表现一般，CLIP score 只有 0.507，主要原因还是因为生成初始的图片中存在目标缺失或者目标多余的情况，因此拉低了图文的一致性。经过对训练集和测试集的图片检测再反馈修正优化以后，整体的图文一致性也得到了提升。

3.2 对比实验

为了验证机制的有效性，将该机制应用到其他大语言模型，本文共测试文心一言、通义千问、讯飞星火以及 GPT4 还有文中之前提到的智谱 AI。将文中之前所提到的 prompt 描述分别输入 5 个不同大语言模型，让其各自根据文本描述生成 100 张图像，根据图像中是否含有 prompt 中描述的特定对象，即每张图片中必须有猫、狗和蝴蝶各一只，作为合格的标准，并将合格的图片在 100 张生成图片中的占比作为合格率（Pass Rate），将合格率以及 CLIP score 作为不同模型的评价指标。同时也将机制加入几个大模型中进行合格率和 CLIP score 的比较，以验证方法的有效性。经过实验，5 个大模型生成图像准确率见表 3。在表 3 中，机制为 YOLOv10-diffusion model，简称 YOLO-DM。

表 3 不同大模型生成图片的合格率和 CLIP score 比较

模型名称	合格图片数量	不合格图片数量	合格率 (pass rate)	CLIP Score
文心一言	12	88	12%	0.224
文心一言 +YOLO-DM	32	62	32%	0.413
通义千问	25	75	25%	0.302
通义千问 +YOLO-DM	88	12	88%	0.705
讯飞星火	17	83	17%	0.281
讯飞星火 +YOLO-DM	43	57	43%	0.478
GPT4	27	73	27%	0.303
GPT4+YOLO-DM	94	6	94%	0.732
智谱	71	29	71%	0.532
智谱 +YOLO-DM	100	100	100%	0.745

从表 3 可以看到，智谱根据 prompt 生成的图片合格率较高且图文一致性表现最好，而 GPT4、通义千问虽然初始生成的图片问题较多，但是经过反馈修正以后图片质量与文本一致性提高了较多，说明 GPT4 和通义千问生成的图片只存在目标缺失或者多余的小问题，较容易修正。而文心一言和讯飞星火无论是初始图片还是修正后的图片合格率和 CLIP score 得分均比较低，经过实验发现，两个大模型在应对相对复杂文本指令时，生成的图像质量较差，往往存在 3 个指定对象都缺失或者图片内容与文本完全不搭的情况，因此有些图片完全无法进行反馈修正，从而拉低了合格率以及 CLIP score 的得分。

总体看来，本文机制在其余大模型生成图片方面也有一定的成效，虽然对部分大模型表现一般，但是在图文一致性方面得到了显著的改进。说明该方法在 AIGC 图文生成与优化方面的应用具有广泛性。

4 结论

本文针对大语言模型根据提示词描述生成图像与文本内容不一致问题，通过使用 YOLOv10 对大模型生成的初始图片进行目标识别以及通过扩散模型对问题图像自校正的方法，实现了图像的局部修复，从而解决了图文之间的失配问题，提高了 AIGC 生成图片的准确性。通过实验结果表明，将目标检测以及扩散模型的自我校正机制相结合能显著提高生成图像中目标的完整性和准确性以及和文本内容的一致性。与传统的大模型图像生成方法相比，本方法不仅解决了图像生成中常见的问题，还提供了一种自动化的反馈修正机制，减少了人工干预的需求。同时该方法对计算机视觉和 AIGC 图像生成领域的进一步发展有着重要的意义。

参考文献：

[1] WU J Y, GAN W S, CHEN Z F, et al. AI-generated content (AIGC): a survey[DB/OL]. (2023-03-26)[2024-05-19].https://doi.org/10.48550/arXiv.2304.06632.

[2] KARRAS T, LAINE S, AITTALA M, et al. Analyzing and improving the image quality of stylegan[DB/OL].(2020-03-23)[2024-03-16].https://doi.org/10.48550/arXiv.1912.04958.

[3] HUANG Y P, XU J W, JIANG Z X, et al. Advancing transformer architecture in long-context large language models: a comprehensive survey[DB/OL].(2024-02-23)[2024-05-11].https://doi.org/10.48550/arXiv.2311.12351.

[4] RAMESH A, DHARIWAL P, NICHOL A, et al. Hierarchical text-conditional image generation with CLIP latents[DB/OL]. (2022-04-13)[2024-07-19].https://doi.org/10.48550/arXiv.2204.06125.

[5] NICHOL A, DHARIWAL P, RAMESH A, et al. GLIDE: towards photorealistic image generation and editing with text-guided diffusion models[DB/OL].(2022-03-08)[2024-06-11].https://doi.org/10.48550/arXiv.2112.10741.

[6] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems.NewYork:-CAI,2020:6840-6851.

[7] ZHOU Z, ZHU Y Q, NAKA N. Text to image generation in DCGAN and stable diffusion model[EB/OL].(2024-01-16)[2024-05-12].https://www.researchgate.net/publication/377383992_Text_To_Image_Generation_In_DCGAN_and_Stable_Diffusion_Model#:~:text=We%20explore%20t.

[8] CEN J, SI W W, LIU X, et al. Diffusion model and vision transformer for intelligent fault diagnosis under small samples[J]. Measurement science and technology, 2023, 35(3): 036204.

[9] 陈杨山，张传庆，赵曙光，等．基于 YOLOv7-Tiny 的交通标识检测算法研究 [J]. 计算机科学与应用，2023, 13(4): 737-744.

[10] 王博，柴锐．基于改进 YOLOv7 的变规模网络重叠区域多目标跟踪方法 [J]. 现代电子技术，2024,47(12):57-61.

[11] 黄毅，周纯，刘欣军，等．基于 YOLOv10 的无人机复杂背景下多尺度检测模型 [J/OL]. 光通信研究 :1-8 [2024-02-17]. http://kns.cnki.net/kcms/detail/42.1266.TN.20240822.1307.002.html.

[12] WU T H, LIAN L, GONZALEZ J E, et al. Self-correcting LLM-controlled diffusion models[C/OL]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE,2024[2024-04-02].https://ieeexplore.ieee.org/document/10657772.

[13] YU Q F, LI J C, YE W T, et al. Interactive data synthesis for systematic vision adaptation via LLMs-AIGCs collaboration[DB/OL]. (2023-05-22)[2024-04-18].https://doi.org/10.48550/arXiv.2305.12799.

【作者简介】

刘嗣久(1998—)，男，江苏苏州人，硕士研究生，研究方向：自然语言处理、大语言模型研究、计算机视觉。

荆晓远（1971—），通信作者（email: jingxy_2000@162.com），男，江苏南京人，博士，教授、博士生导师，研究方向：模式识别、计算机视觉、故障诊断。

任娟(1999—)，女，四川南充人，硕士研究生，研究方向：自然语言处理、大语言模型研究、故障诊断。

姚永芳(1975—)，女，湖南常德人，助理研究员，研究方向：模式识别、机器学习、人工智能与应用。

孙其航(1999—)，男，山东泰安人，硕士研究生，研究方向：自然语言处理、大语言模型、参数高效微调。

（收稿日期：2024-11-07）