融合模型在 GUI 设计风格生成的研究与应用

严武军[1] 张晓丽[1] 王 程[1] 景 莹[1]
YAN Wujun ZHANG Xiaoli WANG Cheng JING Ying

摘 要 风格生成方法作为“设计助手”的核心模块，为人机协同创作模式奠定技术基础。文本到图像个性化和风格化的目标是指导一个预先训练好的扩散模型来分析用户引入的新概念，并将其融入期望的风格中。现有设计风格生成方法面临部署成本高、准确率低以及同质化倾向等问题。为此，文章提出了一种融合 LoRA 与 MoE 机制的高效设计风格生成模型 LMSD（lora-moe-stable diffusion），在 Stable Diffusion 的图像生成过程中，通过插入 LoRA 层实现轻量化微调并且融合 MoE 模块以提高模型精度。同时，采用风格一致性损失函数，解决生成结果在风格维度上与输入 Prompt 保持一致。实验结果表明，该模型在优逸客提供的 GUI 作品集上风格匹配准确率提升至 89.7%，验证了其有效性和优越性，为生成式设计工具的智能化发展提供了有效路径，减少了人工设计过程中的重复劳动和时间成本，为 UI 设计师们开辟了广阔的创作空间与灵感源泉。

关键词 模型微调；风格生成；生成式人工智能；Stable Diffusion

doi：10.3969/j.issn.1672-9528.2026.03.038

0 引言

随着 UI 设计师对个性化时尚需求的不断增长，设计作品在风格上趋于极简主义、3D 动态、暗黑模式等流行趋势，呈现出一定的同质化倾向，缺乏差异化与个性表达，使得设计周期延长、创作成本升高 [1]。在此背景下，开发具备风格适应性、创意生成能力和轻量部署能力的智能设计模型，成为提升 UI 设计生产力的关键突破口 [2]。

随着人工智能技术的迅猛发展，生成式人工智能（artificial intelligence generated content, AIGC）在设计领域的应用呈现出爆发式增长，成为推动人机协同设计与创意生成的重要引擎。风格生成方法经历了从传统图像迁移到深度生成模型的演进。其中，Gatys 等 [3] 提出的基于卷积神经网络的风格迁移技术，通过最小化内容损失与风格损失，实现图像视觉风格的转换，但其在高层次语义理解和结构保持方面仍存在不足。在 2021 年，Hu 等 [4] 提出低秩适应（LoRA）技术用于大语言模型的微调，这是该技术的起源，为后续应用于 Stable Diffusion 奠定了理论基础。LoRA 通过引入低秩矩阵插入到原始模型的权重更新中，仅训练少量新增参数，冻结原模型参数，大幅降低微调成本。比如在 BERT 或 GPT 类模型中，仅需微调参数小于 1%。在资源受限的环境下（如边缘设备、科研实验环境），LoRA 的低参数训练和快速收敛特性非常适用，能显著节省训练时间与算力。但受限于难以应对“任务分布差异很大”的新任务，设置过小会造成欠拟合，泛化能力受限 [5]。

近年来，生成对抗网络（GAN）[6] 与扩散模型（diffusion model）的引入为风格生成带来了新的突破。尤其是在 UI 设计领域，研究者开始关注如何从结构化信息（如组件层级、交互逻辑）中提取风格特征，并生成富有创意和多样性的界面设计。例如，StyleGAN 系列在人脸风格建模中表现卓越，并被尝试迁移至产品展示、图标生成等任务中以减少设计师的工作负担。然而，对抗生成网络存在训练难度较大的问题，使得该网络的实际应用出现了瓶颈。

为应对上述挑战，本文提出一种融合 LoRA 与 MoE 机制的高效 UI 风格生成模型 LMSD（lora-moe-stable diffusion），以 Stable Diffusion 为生成基础，通过插入 LoRA 层实现轻量化微调，并在 U-Net 解码阶段引入多个风格专家子网络 $\{E_1,E_2,\cdots,E_n\}$，结合门控机制动态激活专家路径，提升模型的风格表达能力与可扩展性。同时，采用风格一致性损失函数，约束生成结果与用户输入的文本或图像风格描述在语义空间中的对齐程度，确保模型生成的 UI 或图像在风格上准确还原用户指定的描述或参考图。

本文的主要贡献如下：

（1）提出 LMSD 模型结构，面向 UI 设计风格生成任务，具备轻量化调参、高效风格迁移与模块化专家扩展能力。

1. 太原师范学院 山西晋中 030619

（2）提出 LoRA-Attninject 插入策略，通过在 Self-Attention 模块中注入低秩矩阵，实现不修改原始结构条件下的快速微调，显著降低模型训练成本。

（3）引入风格一致性损失函数，显式约束生成图像与目标风格向量在语义空间中的对齐程度。

（4）LMSD 模型在优逸客提供的 GUI 作品集上，其 FID、SSIM、LPIPS 及风格匹配准确率等多个指标表现优异，风格匹配准确率最高达 89.7%，验证了模型在真实场景下的实用性与通用性。

1 算法介绍

1.1 模型结构

1.1.1 LMSD

LMSD 模型整体由 Prompt 输入模块、LoRA 插入层、MoE 专家模块和扩散模型（Stable Diffusion）四个关键模块组成，整体架构如图 1 所示。

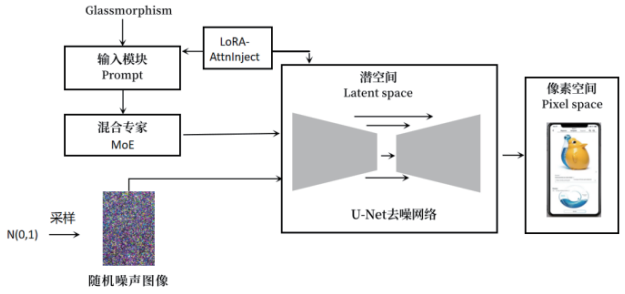

图 1 LMSD 模型图

模型首先接收用户输入的风格描述（文本）或示例图像，并编码为风格向量。在 U-Net 结构中关键层插入低秩微调模块，用于训练极少参数，构建多个风格专家网络，在生成过程中动态选择激活的子网络。Stable Diffusion 模型通过 U-Net 网络对输入的噪声图像逐步去噪，最终输出去噪后的图像。最终，根据输入风格信息选择合适的专家路径，引入风格一致性损失，引导模型保持所生成内容与指定风格的一致性。

1.1.2 Prompt

Prompt 输入模块是 LMSD 模型的起始模块，主要负责接收用户的风格指令（文本描述或图像示例），并将其编码为可用于生成过程的风格向量（Style Embedding），从而引导后续模型的风格选择与图像生成。输入形式支持文本输入，如“极简风格”“3D 未来感 UI”“日系插画风格”等自然语言描述，通过文本编码器（如 CLIP/Text Encoder）将语义信息转化为风格向量；图像输入（Style Reference Image），如用户上传的已有风格示例（APP 界面截图、插画、网页风格图），可通过图像编码器（如 CLIP/Image Encoder 或 VAE 编码器）提取其风格特征向量。

在模型内部，该模块将输入的文本或图像统一映射到相同的风格嵌入空间，记为：

$$\boldsymbol{s} = \mathrm{Encode}_{\mathrm{style}}(\mathrm{Prompt}) \tag{1}$$

式中：$\boldsymbol{s}$ 为生成图像的风格引导向量，作为 MoE 路由模块与 LoRA 插入层的核心参考依据。

1.1.3 Stable Diffusion

扩散模型（Diffusion models）由 Sohl-Dickstein 等 [7] 在 2015 年提出，起初 Sohl-Dickstein 等受到非平衡统计物理学（non-equilibrium statistical physics）理论启发，使用马尔可夫链来将一个分布逐渐转换为另一个分布。扩散模型自出现以来未受到如 GAN 一般的重视（二者提出的时间相近），直到 2020 年，Ho 等 [8] 给出去噪扩散概率模型（denoising diffusion probabilistic models, DDPM）并成功完成扩散模型的应用，此后扩散模型的研究逐渐进入增长期，涌现了大量的研究产出，并在 2021 年于 CIFAR-10（Canadian Institute for Advanced Research）战胜 GAN 成为最高水平（state of the art, SOTA）的模型 [9]。扩散模型是一个参数化的马尔可夫链，可以被解释为层次 VAEs 使用变分推理进行训练，在有限时间后产生与数据相匹配的样本。扩散模型作为近年来图像生成领域的前沿技术，在此后引发了生成领域指数级增长的兴趣。凭借其对图像细节的高度还原能力在多种任务中表现出色。Stable Diffusion 具备开放性强、模块化灵活的特点，支持通过微调实现个性化风格的迁移。特别是在设计风格生成任务中，Stable Diffusion 能够根据风格提示词或示例图，合成具有一致视觉语言的 UI 方案。Stable Diffusion 模型通过 U-Net 网络对输入的噪声图像逐步去噪，最终输出去噪后的图像。在本文中，通过为 SD 的不同块构建块式 LoRA 适配器，从而提高生成任务的个性化和风格化性能，如图 2 所示。

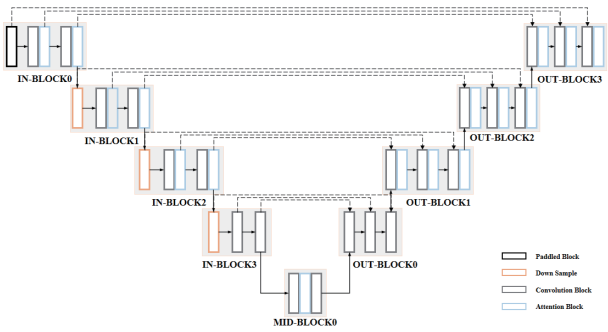

图 2 SD 中 U-Net 的块划分

1.1.4 LoRA-AttnInject

在此基础上，LoRA 提供了一个更为专业的途径，专门为稳定扩散模型量身定制，以提高涉及扩散过程的生成任务的适应性和性能，提出了不同的后续方法以持续改善 LoRA 的性能。LoRA 算法最初用于微调大语言模型，通过低秩分解的方法，将大模型需要更新的权重矩阵 $\Delta \boldsymbol{W}$ 分解为可训

练的矩阵 $\boldsymbol{A}$ 与矩阵 $\boldsymbol{B}$ 的乘积，从而显著地减少下游任务所需要的可训练参数的规模。随着模型越来越大，微调大模型的成本和资源代价也越来越高，为了快速、高效地将大模型基于特定任务进行微调，出现了许多微调模型的方法，如 Adapt Tuning、Prefix Tuning、LoRA 等，其中 LoRA 是 LLM 中低资源微调大模型的方法。在实际应用中，采用 Stable Diffusion v1.5 作为生成模型的预训练模型。参照 LoRA 的方法，冻结 Stable Diffusion，并对 Prompt 编码器及 U-Net 的 Self-Attention 的全连接层旁插入可训练权重 [10]。LoRA（Low-Rank Adaptation）通过在 Self-Attention 模块中的关键投影矩阵——查询（$\boldsymbol{Q}$）、键（$\boldsymbol{K}$）、值（$\boldsymbol{V}$）与输出（$\boldsymbol{O}$）中插入一组可训练的低秩矩阵模块，实现对模型的高效微调，将该机制命名为 LoRA-AttnInject 策略。该策略无需修改原始模型结构，而是通过模拟参数更新路径，引导原始权重的偏置方向，从而在保持性能的同时显著减少微调所需的参数量与计算开销。LoRA-AttnInject 具备良好的可插拔性与部署灵活性，可无缝集成至扩散模型中的 Attention 路径，构建轻量、可扩展的个性化风格微调机制，尤其适用于大规模模型在资源受限环境下的快速适配与落地应用。LoRA-AttnInject 的策略图如图 3 所示。

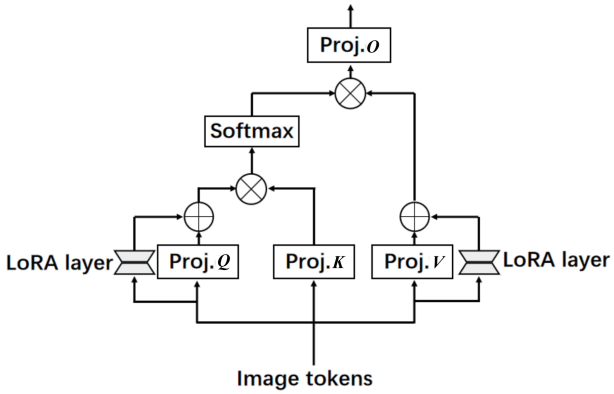

图 3 LoRA-AttnInject 策略图

1.1.5 MoE

MoE 是一种结构灵活、容量可扩展的神经网络架构。通过构建多个“专家子模型”（Experts），在每次前向计算时根据输入信息，仅激活其中一部分专家，从而实现更大的模型容量与更少的推理成本。近年来，MoE 结构也被引入图像和多模态任务中，用于处理多样风格、复杂情境下的特征选择问题 [11]。MoE 的核心模块包括专家网络组（多个并行子网络）、门控网络（Gating Function）以及 Sparse Routing。根据输入选择若干活跃专家，避免全专家激活，提升效率。MoE 能够在风格生成中根据不同提示或输入语义，动态选择最适合的专家模型，从而实现高质量、多样化的生成输出。

为了处理风格多样性与复杂性，本文引入专家混合模块（MoE），在 U-Net 解码阶段加入多个风格专家子网络 $\{E_1, E_2, \cdots, E_n\}$，每个专家可专注于一类风格建模。

由门控网络 G 接收输入风格向量 $\boldsymbol{s}$，生成每个专家的激活概率：

$$\boldsymbol{g} = \mathrm{Softmax}(W_g \cdot \boldsymbol{s}) \tag{2}$$

式中：$\boldsymbol{g} \in \mathbf{R}^n$ 为专家权重；W_g 为可训练参数。

根据 Top-k 策略，仅选择激活前 k 个专家参与前向传播。输出为选中专家的加权和：

$$\mathrm{Output} = \sum_{i=1}^{k} g_i \cdot E_i(x) \tag{3}$$

MoE 的引入使得模型具备按需分配“风格专长”的能力，同时通过稀疏激活控制计算资源，保持高效。

1.1.6 LoRA 和 MoE 结合的方法

基于以上 2 种风格定制生成的方法，本文提出一种以上 2 种方法相结合的方法。首先，参照 1.1.3 节中的 LoRA 方式，以 Stable Diffusion v1.5 为预训练模型，冻结预训练模型，对 Prompt 编码器及 U-Net 的 Self-Attention 的全连接层旁插入可训练权重。$\boldsymbol{A}$ 和 $\boldsymbol{B}$（风格强化模块），仅训练参数 $\boldsymbol{A}$ 和 $\boldsymbol{B}$。推理时，文本示例为：“［U］，《提示词》。”

其次，利用 1.1.4 节中 MoE 方法，以 Stable Diffusion v1.5 为预训练模型，将 LoRA 模块插入各专家网络中，使得每个专家具备独立学习风格的能力，同时共享主模型的大规模预训练能力。这种方式不仅保留了 MoE 的模型扩展能力，还能利用 LoRA 在低资源场景下的高适应性。推理时，文本示例为：“一幅画，［V］风格，《提示词》。”

两种方法采用同一个预训练模型进行微调，网络结构和层数一致，在前 2 步完成后，将第 2 步中训练完成的参数 $\boldsymbol{A}$ 和 $\boldsymbol{B}$（风格强化模块）插入 M 对应层的旁路。前向推理时，文本描述采用“特殊标识词，画面语义描述”组合的方式生成对应风格的画面，示例如下：“［U］，《提示词》。”如图 4 所示。

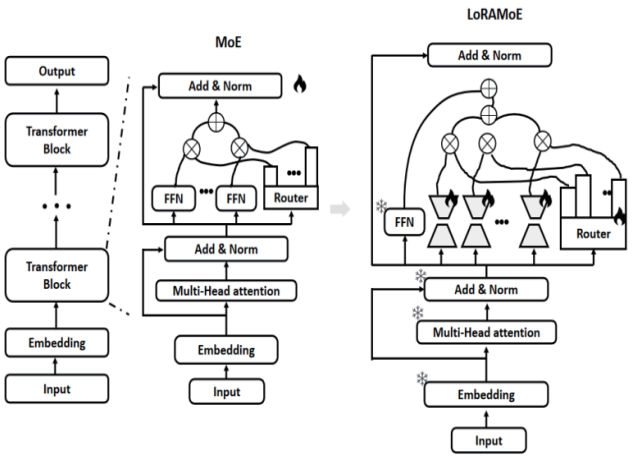

图 4 本文方法示意图

1.1.7 风格监督机制与损失函数

在设计风格生成任务中，仅靠模型的图像质量指标（如重建损失、感知损失）无法确保生成结果在风格维度上与输入 Prompt 保持一致。因此，本文引入风格一致性损失函数，用于显示约束生成图像与目标风格之间的语义对齐程度，确保模型生成的 UI 或图像在风格上准确还原用户指定的描述或参考图。

风格一致性损失函数（Style Consistency Loss）计算公式具体表示为：

$$\boldsymbol{L}_{\text{style}}=\|f_{\text{style}}(\hat{x})-\boldsymbol{v}_{\text{style}}\|^2 \tag{4}$$

式中：f_{style} 是风格判别网络；$\hat{x}$是生成图像；$\boldsymbol{v}_{\text{style}}$ 是目标风格向量。

最终训练损失为多项加权组合：

$$L=\lambda_1\cdot L_{\text{diffusion}}+\lambda_2\cdot L_{\text{style}}+\lambda_3\cdot L_{\text{recon}} \tag{5}$$

式中：$L_{\text{diffusion}}$ 是标准扩散重建损失；L_{style} 是风格保持损失；L_{recon} 是图像内容一致性重建损失。

2 实验结果与分析

2.1 数据集与实验环境

为验证 LMSD 模型的有效性，实验使用合作企业山西省优逸客科技有限公司提供的 GUI 作品集作为实验数据来源，该数据集涵盖超过 35 717 张高质量 UI 截图，并配有详细的界面层级结构与控件类型信息。细分为训练集 28 573 张、验证集 3 572 张和测试集 3 572 张。数据集中包含了 6 种不同的风格类别分别用数字 0 到 5 标注。0 代表极简风（Minimalist）；1 代表拟物风（Skeuomorphic）；2 代表卡通风（Cartoon）；3 代表扁平化风格（Flat）；4 代表暗黑模式（Dark Mode）；5 代表玻璃拟态（Glassmorphism）。具体示例如图 5 所示。

图 5 数据集示例图

2.2 模型训练与实验结果

在所有实验中，模型训练与微调均基于 Stable Diffusion v1.5 版本完成，默认冻结主干权重，训练新增模块参数（如 LoRA 层与 MoE 子网络）。为保证实验结果的稳定性与可复现性，以下是实验中的一些基础设置：batch_size 设置为 16，epoch 设置为 100。若验证集上的评估结果连续 5 个 epoch 没有增加，训练将提前停止。并规定了初始学习率为 0.000 1。若模型在连续两个训练周期（epochs）中的验证准确率未见增长，学习率将通过预设的学习率调度器自动缩减至原先值的 1/10。此策略旨在优化训练过程，以防止潜在的过拟合现象，并确保模型有效收敛。LoRA rank 默认设为 r=4，激活专家 k=2。为综合评估模型性能，采用表 1 评估指标进行评估。

表 1 评估指标

指标名称	说明
FID	评估生成图像与真实样本的分布差异（越低越好）
LPIPS	评估风格一致性（越低越好）
Style Accuracy	使用预训练风格分类器评估生成图像风格分类准确率
PSNR	峰值信噪比
SSIM	衡量两张图片的相似程度

2.3 模型对比实验

为了验证本文方法的有效性，将所提模型与以下几种对比方法进行性能对比，如表 2 所示。

表 2 不同模型在同一数据集的指标

方法	FID	LPIPS	PSNR	SSIM	SA/%
Stable Diffusion（全参数）	21.4	0.31	32.0	0.331	82.6
LoRA+Stable Diffusion	18.6	0.25	40.0	0.621	86.0
Ours (LoRA + MoE)	17.1	0.21	45.0	0.721	89.7

从表 2 可以看出，本文所提模型在多个指标上均取得了最佳性能：在图像分布距离（FID）与感知相似度指标（LPIPS）上，LMSD 分别达到 17.1 与 0.21，优于所有对比方法，说明生成图像更加真实与风格一致；在语义一致性评分（SA）上，LMSD 模型达到 89.7%，显著高于 Stable Diffusion 基线（82.6%）与 DreamBooth（84.2%），验证了专家模块对风格建模的强化作用；LMSD 模型在峰值信噪比和测量图片的失真程度上，分别达到了 45.0 和 0.721。综上，本文提出的融合 LoRA 与 MoE 架构在保持模型轻量化的同时实现了更高的生成质量与语义控制能力，为多风格 UI 生成任务提供了更优的解决方案。

2.4 消融实验

为分析所提模型中 LoRA 与 MoE 两个模块的独立贡献，本文设计了基于 Stable Diffusion 的四组模型变体进行消融

实验，分别为：原始模型（不含 LoRA 与 MoE）、仅引入 LoRA、仅引入 MoE，以及同时引入 LoRA 与 MoE 的完整模型。通过对比各组在 SSIM、FID 和 LPIPS 等指标上的表现，评估每个模块对图像生成质量的具体提升效果，从而验证两模块在结构还原与风格适应方面的有效性及协同优势。

实验结果如表 3 所示，通过实验发现，引入 LoRA 模块可有效提升结构还原度（SSIM）、降低感知误差（LPIPS）与分布偏差（FID），说明其对生成效果具有积极作用；增加 MoE 模块则在保持参数可控的基础上进一步提升风格适应能力；将二者结合使用（LoRA+MoE）表现最优，说明两种机制具有互补性与协同效应。LoRA 与 MoE 均对模型性能提升具有显著贡献，二者联合可实现最优性能表现。

表 3 消融实验

模型变体	SSIM	FID	LPIPS
无 LoRA，无 MoE（原模型）	0.826	21.4	0.31
+LoRA	0.860	18.6	0.25
+MoE	0.872	18.1	0.23
+LoRA+MoE（完整模型）	0.897	17.1	0.21

3 总结

本研究提出了一种融合 LoRA 与 MoE 机制的高效生成模型 LMSD，在 Stable Diffusion 的图像生成过程中，通过插入 LoRA 层实现轻量化微调并且融合 MoE 模块以提高模型精度。同时，采用风格一致性损失函数，解决生成结果在风格维度上与输入 Prompt 保持一致。尽管本研究在设计风格生成这一智能设计领域取得了一定的成果，但仍有许多值得探索的方向。未来的研究可以集中在如何进一步优化模型结构，探索通过风格聚类，自注意力等机制自动识别风格类型，减少人工标签依赖。将图像、文本、图标、动效等多模态风格要素共同建模，构建更全面的风格表达能力。

参考文献：

[1] 孟星彤 , 钟佩均 , 边卓 . 基于情境感知的 UI 交互设计研究 [J]. 工业设计 ,2023(11):96-99.

[2] 雷菁 . 互联网软件产品移动端 UI 设计方法研究 [J]. 科学技术创新 ,2024(5):86-89.

[3]GATYS L A, ECKER A S, BETHGE M. A neural algorithm of artistic style[J]. Journal of vision,2016,16(12):326.

[4]HU E J, SHEN Y L, WALLIS P, et al. LoRA: low-rank adaptation of large language models[EB/OL].(2021-10-16)[2026-01-22]. https://doi.org/10.48550/arXiv.2106.09685.

[5] 郭宇轩 , 孙林 .LoRA 模型微调 Stable Diffusion 的设计师风格服装生成方法 [J]. 北京服装学院学报 (自然科学版),2024,44(3):58-69.

[6] 金德虎 . 基于多粒度语义信息融合的文本到图像生成方法 [D]. 济南 : 山东师范大学 ,2024.

[7]SOHL-DICKSTEIN J,WEISS E A, MAHESWARANATHAN N, et al.Deep unsupervised learning using nonequilibrium thermodynamics[EB/OL].(2015-11-18)[2026-01-22].https://doi.org/10.48550/arXiv.1503.03585.

[8]HO J, JAIN A, ABBEEL P.Denoising diffusion probabilistic models[J].Advances in neural information processing systems, 2020,33:6840- 6851.

[9]RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin:Springer,2015: 234 - 241.

[10] 马诗洁 , 徐华艺 , 李聪聪 , 等 . 基于大模型微调范式的绘画风格模拟方法 [J]. 计算机应用 ,2024,44(S1):268-272.

[11] 史宏志 , 赵健 , 赵雅倩 , 等 . 大模型时代的混合专家系统优化综述 [J]. 计算机研究与发展 ,2025,62(5):1164-1189.

【作者简介】

严武军（1973—），男，山西临汾人，硕士，副教授，研究方向：Java EE 企业级开发、大数据、移动应用开发、人机交互工程、人工智能。

张晓丽（2001—），女，山西运城人，硕士研究生，研究方向：大模型、图像处理。

王程（2002—），女，山西长治人，硕士研究生，研究方向：大数据、人工智能。

景莹（1999—），女，山西运城人，硕士研究生，研究方向：深度学习、医学影像处理。

（收稿日期：2025-08-14 修回日期：2026-02-26）