**中图法分类号**:TP18;TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2026)04-1227-14

**论文引用格式**:Lu F B, Luo W C, Qiao Y Y, Ge X Y, He P and Wang M L. 2026. A high-fidelity line drawing extraction model based on cross-layer fusion and joint loss optimization. Journal of Image and Graphics, 31(4):1227-1240(鲁方博, 罗万闯, 乔永源, 葛贤钰, 贺鹏, 王美丽. 2026. 跨层级响应融合与联合损失优化的高保真线稿提取模型. 中国图象图形学报, 31(4):1227-1240)[DOI:10.11834/jig.250223]

# 跨层级响应融合与联合损失优化的高保真线稿提取模型

鲁方博[1],罗万闯[1],乔永源[1],葛贤钰[2],贺鹏[2],王美丽[1*]

1. 西北农林科技大学信息工程学院,咸阳　712100; 2. 西安纽扣软件科技有限公司,西安　710075

**摘　要**:**目的**　线稿提取是指利用边缘检测技术从原始图像中提取出具有语义连续性的轮廓和边缘信息,为动漫上色、风格迁移等下游任务提供结构化输入。针对现有线稿提取模型在复杂纹理场景下,提取结果仍存在线条不纯净、背景伪影等问题,提出基于跨层级响应融合与联合损失优化的高保真线稿提取模型 CLEAR-Net(cross-level edge aggregation response network)。**方法**　引入反卷积改进 $U^2$-Net 提取图像不同层级的响应;提出动态侧边聚合模块实现跨层级响应优化整合;针对复杂纹理场景下所产生的背景伪影问题,提出了一种新的监督机制——背景抑制损失,对背景伪影进行像素级动态惩罚;设计联合损失函数,联合背景抑制损失与改进的交叉熵损失,在抑制背景伪影的同时优化生成线条的质量。为构建可靠的评估基准,联合专业艺术团队构建首个高精度手绘线稿数据集 ArtLine-2K,包含 2 000 组渲染图—线稿对,并经数据增强扩充到 10 000 对,解决了当前线稿提取任务中高质量标注数据缺乏的问题。实验在 ArtLine-2K 数据集上与先进方法进行了比较。**结果**　实验结果表明,CLEAR-Net 生成结果与真实标注的差异肉眼难以区分,其核心精度指标:均方误差(mean squared error,MSE)和平均绝对误差(mean absolute error,MAE)分别为 0.000 247 和 0.004 810,与真实标注的误差达到亚像素精度(MAE < 0.005),在 ArtLine-2K 上取得了突破性能。生成结果经专业画师评估,可以直接进行二次创作,同时也在 ArtLine-2K 上进行了消融实验以验证提出方法的有效性。**结论**　CLEAR-Net 模型不仅优化整合了原始图像各层级的响应,而且提出了新的监督机制,解决了线稿提取任务中线条不纯净、背景伪影等问题。

**关键词**:边缘检测;线稿提取;背景伪影抑制;层级响应融合;动态特征聚合

# A high-fidelity line drawing extraction model based on cross-layer fusion and joint loss optimization

Lu Fangbo[1], Luo Wanchuang[1], Qiao Yongyuan[1], Ge Xianyu[2], He Peng[2], Wang Meili[1*]

*1. College of Information Engineering, Northwest A&F University, Xianyang　712100, China;*

*2. Xi'an Button Software Technology Co., Ltd., Xi'an　710075, China*

**Abstract**: **Objective** Line drawing extraction is one of the key tasks in the fields of computer vision and image processing. It aims to extract contour and edge information automatically with semantic continuity and structural consistency from the original image by using edge detection and feature learning techniques. In this way, a high-quality structured input for downstream tasks, such as animation coloring, style transfer, image generation, and illustration restoration, is provided. This task not only requires the model to identify the main outline of the object accurately but also needs to maintain the con-

**收稿日期**:2025-06-26;**修回日期**:2025-11-10;**预印本日期**:2025-11-17

*通信作者:王美丽　wml@nwsuaf.edu.cn

tinuity of the lines and the rationality of the overall structure while suppressing the interference of irrelevant background and texture details. When facing complex textures and rich background images, the existing line drawing extraction methods can obtain relatively clear lines in regular scenes; however, balancing the detail fidelity of the lines and the purity of the background is difficult, and problems such as line breakage, blurred contours, loss of local details, and background artifacts are prone to occur. These problems cause the extraction results to lack semantic integrity and artistic consistency, thereby reducing the input quality of downstream tasks and making the demands of actual creation and industrial applications for high-precision line drawings difficult to meet. In response to the above problems, this study proposes a high-fidelity line drawing extraction model, namely, cross-level enhanced aggregation and refinement network (CLEAR-Net), based on cross-level response fusion and joint loss optimization. This model fully utilizes multiscale semantic information by integrating feature responses at different levels and introduces a joint optimization strategy, thereby effectively improving the quality of line extraction. It suppresses background artifacts while ensuring structural consistency, thereby obtaining pure line drawing results. **Method** This study made structural improvements based on $U^2$-Net and introduced a deconvolution module to enhance the model's response ability to features at different levels. The model can fully restore the spatial detail information of the deep response by adding deconvolution operations in the upsampling stage; thus, delicate edge structure extraction is achieved in the multiscale feature fusion process. Subsequently, a dynamic side aggregation module was proposed to achieve dynamic fusion and optimization of cross-level features. This module can automatically allocate aggregation weights based on the correlation between features of different layers, strike a balance between global structural information and local texture details, and effectively enhance the coherence and integrity of the line structure. A background suppression supervision mechanism is proposed for the common background artifact problem in complex texture scenes. This mechanism enables the model to penalize the pseudo-responses in the background area dynamically. It also effectively reduces the interference of background noise. As a result, the purity and robustness of the results are enhanced. A joint loss function combining the background suppression loss with the improved cross-entropy loss is designed to enhance the quality of the generated results further. As a result, the background artifacts are suppressed, and the foreground lines are optimized, thereby achieving a dual improvement in line quality and background purity. Finally, this study was conducted in collaboration with a professional art team to build the first high-precision hand-drawn dataset, ArtLine-2K, which contains 2 000 pairs of high-quality rendered line drawing samples covering various painting styles and complex scenes. This dataset was expanded to 10 000 pairs of samples through data augmentation. Thus, the problem of scarce high-quality labeled data in the line drawing extraction task is effectively alleviated. A systematic comparison was conducted with multiple advanced methods on the ArtLine-2K dataset. **Result** Experimental results show that the differences between the generated results of CLEAR-Net and the real annotations are difficult to distinguish with the naked eye. The errors of its core accuracy indicators, MSE (0.000 247) and MAE (0.004 810), and the real annotations reach subpixel accuracy (MAE < 0.005), thereby achieving the breakthrough performance on ArtLine-2K. The generated results were evaluated by professional painters and could be directly used for secondary creation. The ablation experiments were also conducted on ArtLine-2K to verify the effectiveness of the proposed method. **Conclusion** Experimental results show that CLEAR-Net achieved a breakthrough performance on ArtLine-2K. The generated results are almost indistinguishable from the real annotations. The precision index MSE is 0.000 247, and MAE is 0.004 810. Moreover, the error of the proposed model reaches the subpixel level (MAE < 0.005), which is significantly better than that of the existing methods. Compared with other models, Clear-Net performs outstandingly in detail restoration, line continuity, and background purity. The generated line drawings have clear and natural lines with smooth edges and no artifacts. They can be directly used for secondary creation after being evaluated by professional artists. A systematic ablation experiment was carried out on ArtLine-2K to verify the effectiveness of the model structure design and loss function. Results show that the introduction of the feature extraction module, side fusion mechanism, background suppression loss, and smooth heating cross-entropy can synergistically and significantly reduce the error. Compared with the benchmark model, the proposed model achieves more than 95% improvement in overall performance. Furthermore, CLEAR-Net still maintains stable performance on low-quality datasets, such as Anime Sketch Colorization Pair, thereby demonstrating excellent cross-domain generalization ability and robustness.

**Key words**: edge detection; line draft extraction; background artifact suppression; hierarchical response fusion; dynamic feature fusion

## 0 引　言

线稿提取作为计算机视觉与数字艺术创作的交叉领域，其核心目标是从原始图像中提取具有语义连续性的轮廓信息，为动漫上色、风格迁移等下游任务提供结构化输入。该领域的发展经历了从传统手工特征工程到数据驱动范式的演变，近年来在模型架构设计与监督机制方面取得进展。

早期线稿提取算法依赖人工设计的梯度算子，如通过高斯滤波与梯度阈值化实现边缘定位，并通过数学形式定义了检测和定位标准，然而，此类方法对噪声敏感且难以区分语义边缘与纹理干扰（Canny，1986）。Arbeláez 等人（2011）提出了一个基于频谱聚类的轮廓检测框架，结合局部线索以全局优化方式进行轮廓检测，可以消减高频纹理干扰。Sert 和 Avsi（2019）提出了一种基于中性集结构的边缘检测方法。刘丹和王运宏（2020）提出了一种限制型自适应 SUSAN（smallest univalue segment assimilating nucleus）边缘检测算法，通过改进经典 SUSAN 算法，设计了基于像素值的动态门限和异侧噪声容忍机制，在保留经典算法计算效率的同时提升抗噪性能，实验表明其在强噪声干扰下的 FSIM（feature similarity）指标优于 Canny 等传统算子，但存在边缘定位精度与计算耗时的权衡问题。

随着深度学习的发展，相关研究者基于卷积神经网络提出了若干检测架构。Xie 和 Tu（2015）首次提出端到端全卷积网络架构，通过多层级侧边输出融合提升边缘连续性。Liu 等人（2017）通过融合 VGG16（Visual Geometry Group）网络所有卷积层特征增强检测能力，但其特征融合策略可能引入定位模糊问题；He 等人（2019）提出双向级联结构与尺度增强模块，通过分层监督策略提升多尺度检测性能，但模型复杂度较高。Chollet（2017）重新解构了卷积神经网络中的 Inception 模块，提出其本质是常规卷积与深度可分离卷积的中间形态。基于这一理论，Chollet（2017）进一步设计了一种新型架构 Xception，通过将 Inception 模块完全替换为深度可分离卷积，实现了更高的计算效率与特征表达能力的平衡。Guo 等人（2019）提出了一种基于深度可分离卷积的多领域学习架构，旨在通过通用参数化捕获不同视觉领域的共享结构。王素琴等人（2021）提出一种基于循环生成对抗网络的线稿自动提取模型，用以解决非对称数据训练问题。朱威等人（2021）提出一种多深度特征增强与顶层信息引导的边缘检测网络。该网络通过融合 UNet++的多深度特征增强机制，结合空洞卷积扩展感受野和高层语义引导模块，在 BSDS500（Berkeley segmentation dataset and benchmark）数据集上取得了 0.869 的平均精度（average precision，AP）。然而，其依赖特征叠加的策略导致边缘响应过粗，在高精度线条提取任务中难以满足需求。李文轩等人（2025）提出了一种边缘—区域特征金字塔融合方法，通过融合纹理特征与几何分析实现多尺度特征增强。该方法创新性地设计了模拟专业观察范式的数据扩充模块，为髁突骨微结构分割与预测提供了新思路，但其性能受限于训练数据多样性的不足。Soria 等人（2023）采用深度可分离卷积实现超轻量化设计，但其单一监督机制在复杂场景下对伪响应抑制不足。针对监督机制问题，Huan 等人（2022）揭示了卷积神经网络中特征混合与侧边融合模糊的问题，提出上下文感知追踪策略以提升边缘定位精度。Deng 和 Liu（2020）设计了一种新颖的损失函数，能够惩罚预测和真实轮廓之间的结构差异，同时提出了一种卷积编码器—解码器网络，并引入了超级模块来捕获高级特征的密集连接和语义信息。Xuan 等人（2022）引入精细化校正学习机制，利用注意力模块增强细粒度边缘定位能力。Yang 等人（2024）提出了一种基于纹理感知损失函数和高效编码—解码结构的边缘检测方法 TANet（textureness-aware neural network），有效抑制了高频纹理区域的伪响应，提升了边缘检测精度。Ye 等人（2024）提出了一种基于扩散概率模型的边缘检测方法 DiffusionEdge，通过在潜空间中引入不确定性蒸馏策略与自适应傅里叶滤波器，实现了无需后处理即可同时获得高准确性与高清晰度的边缘图，但由于扩散过程的迭代特性，推理效率仍有待进一步提升。Li 等人（2025）提出了一种双重解耦网络 DDN（doubly decoupled network），通过数据解耦缓解

模型过拟合问题，并通过特征解耦压缩浅层特征中的冗余计算，从而在边缘检测任务中实现了更高的准确性和更低的计算成本。然而，DDN在不同任务中需要调整目标函数，增加了模型迁移的复杂性。

当前边缘检测方法面临两大挑战（Jing等，2022）：1）缺乏针对特定任务的边缘检测模型，现有主流方法（如TEED（tiny and efficient edge detector）等）基于通用数据集（如BSDS500）训练，虽然在自然场景表现良好，但在特定领域（如线稿提取）存在显著局限；2）现有模型对复杂纹理或高密度边缘场景的相邻边缘区分度不足，易出现边缘粘连、背景伪影等伪现象。如动漫图像的线条具有高度抽象性与语义连续性，传统自然场景边缘检测模型在此类数据上表现受限（Soria等，2023）。尽管图像翻译方法可实现风格迁移（Zhu等，2017），但其未显式建模边缘结构特征，导致生成效果不佳。

针对上述挑战，本文提出一种基于跨层级响应融合与联合损失优化的高保真线稿提取模型，称为CLEAR-Net（cross-level edge aggregation response network）。首先，CLEAR-Net改进了U$^2$-Net的嵌套U型结构（Qin等，2020），使用并行卷积路径捕获不同感受野下的边缘响应，解决了传统模型在跨尺度特征融合中的信息融合问题，并改进邻层特征的融合方式，减少边缘混淆的同时，充分利用不同层级响应图的优势；然后，CLEAR-Net引入通道注意力机制（SE（squeeze-and-excitation）模块）并提出双通道深度可分离卷积层（Hu等，2018），动态调整跨层级响应权重，增强对主体轮廓的拓扑建模能力；最后，模型联合像素级改进交叉熵损失与结构感知损失，通过标签平滑（Szegedy等，2016）与温度缩放（Kull等，2019）策略优化线条质量，提升模型收敛速度。CLEAR-Net通过引入结构感知损失函数，有效抑制了背景伪影的干扰，解决了现有方法因伪影导致的线条污染问题。

为解决线稿提取中动漫领域高质量标注数据缺乏问题，本文联合专业艺术团队构建首个高精度手绘线稿数据集ArtLine-2K，包含2 000组渲染图—线稿对，并经数据增强扩充到10 000对。实验表明，本文方法在保持轻量化优势的同时，显著提升了对动漫风格图像的边缘检测性能，像素级误差相较于当前最优模型下降两个数量级，在ArtLine-2K上取得了突破性能。

## 1 网络模型（CLEAR-Net）

### 1.1 网络结构

如图1所示，CLEAR-Net模型包含3个核心模块：特征提取模块、侧边聚合模块和联合监督机制。CLEAR-Net的特征提取模块改自U$^2$-Net。改进的特征提取模块结构简单，通过分层卷积和密集的跳跃连接提取了丰富的层级响应，并且支持模型从头训练；侧边聚合模块在不同模式上动态感知不同层级响应的优势，从而得到精度更高的侧边融合结果；联合损失函数强化了对背景伪影的抑制，引入平滑因子和模型升温，加快模型收敛并减少模型预测结果对极端值的惩罚，从而得到质量更好的边缘信息。

### 1.2 特征提取模块

本文旨在提取动漫图像的线稿，线稿本质上属于图像的低阶视觉特征（如边缘轮廓、线条走向和局部纹理等），这类特征主要分布在网络浅层结构中。

因此，在构建特征提取模块时需控制网络深度，避免过深的层级结构导致浅层特征在传递过程中被过度抽象化或产生信息衰减，从而增加不必要的计算开支。同时特征提取模块应具备足够的表征能力，最好能够使数据不经过骨干网络直接计算。而U$^2$-Net作为典型的编码器—解码器结构，通过嵌套的U型层级架构实现了多尺度特征提取，其独特的跳跃连接机制能够绕过传统骨干网络的深度特征提取过程，直接从编码器各阶段捕获包含丰富空间细节的浅层特征。

如图2所示，特征提取模块能够提取出各侧边的响应，各侧边响应表示为

$$\{\boldsymbol{F}^{(k)}\}_{k=1}^{K} \subseteq \mathbf{R}^{H\times W\times C} \tag{1}$$

式中，$k$是层级的索引，表示不同的层次（或尺度）特征，$\boldsymbol{F}^{(k)}$表示第$k$层的响应图，大小为$H\times W\times C$。低阶响应图（如第1层和第2层输出）在边缘定位精度上具有像素级敏感性，可清晰反映线条转折点与曲率变化，但存在高频噪声弥散现象；高阶特征图（如第3层输出）则展现出对主体轮廓拓扑关系的强建模能力，能够有效关联离散边缘段形成闭合结构，但伴随空间细节的渐进式丢失。

当网络深度达到第4层级时，浅层解码器输出的特征图虽然能保留基础结构信息，但深层卷积核的响应效果明显衰减。前3层网络对线条走向、交

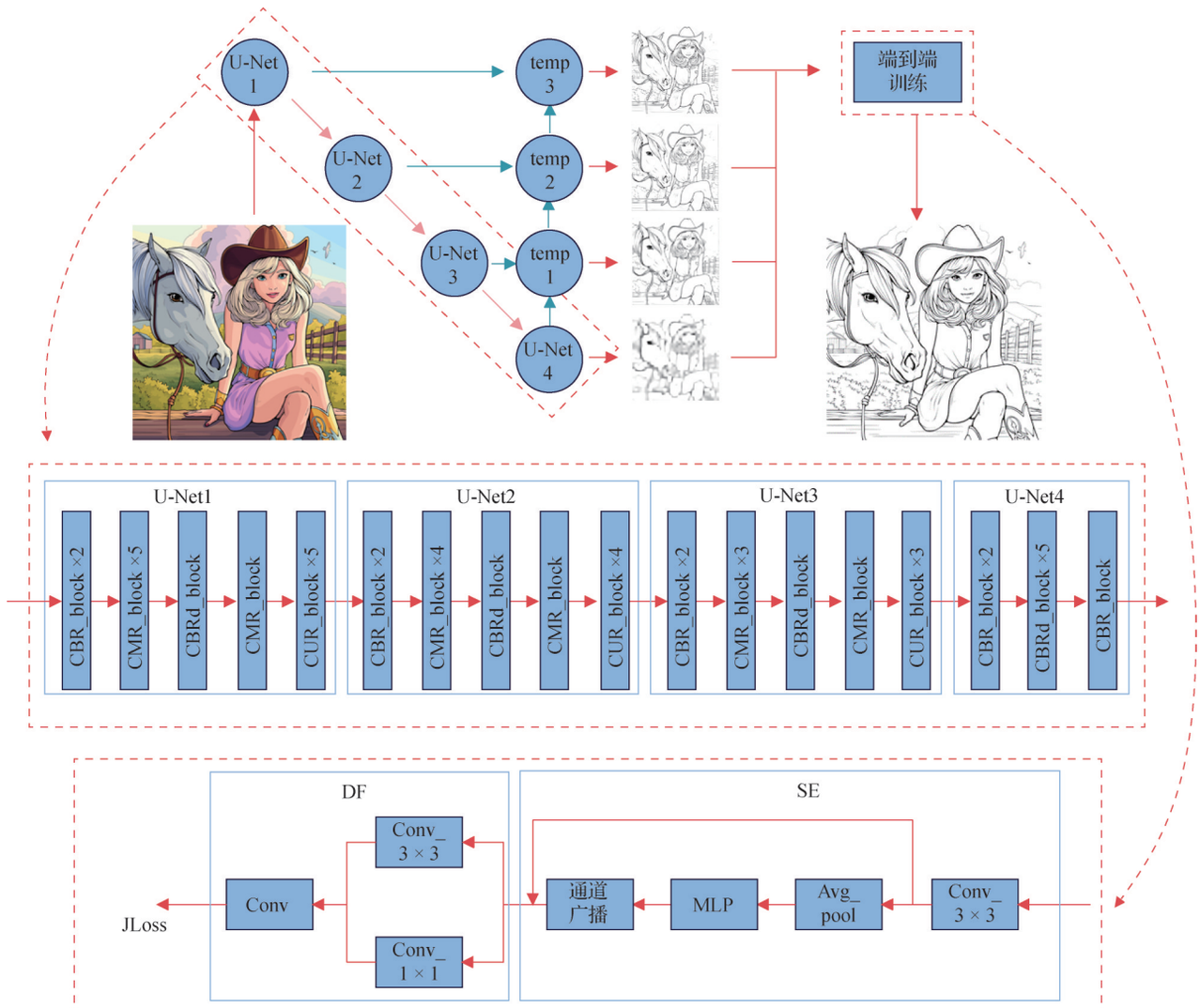



图 1　网络结构

Fig. 1　Network structure

图 2　各侧边响应

Fig. 2　Responses of each side

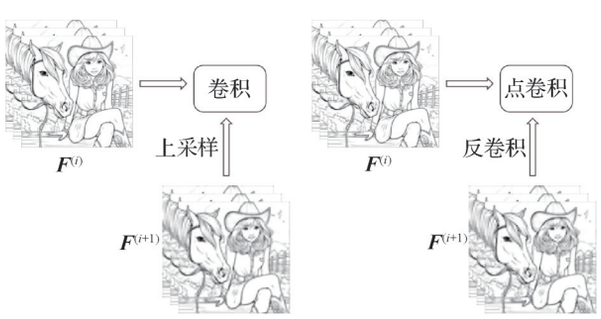
叉点等关键特征的响应强度保持稳定，而第 4 层特征图出现响应模糊和细节丢失现象，是由于深层网络感受野扩大导致局部细节被过度平滑。为优化模型效率，尝试移除第 4 层卷积结构以降低参数规模（约减少 14% 的参数量）和计算复杂度。

图 3 实验表明，3 层网络响应精度显著降低，模型因网络深度不足导致模型特征提取能力减弱。基于此，决定网络深度为 4 层。 针对不同层级响应的表征差异，本文提出临级特征动态融合策略，如图 4 所示，首先对高阶特征实施反卷积操作，卷积核大

小4×4、步长为2、填充为1、输出填充为0，通道数与高阶特征持平，以确保上采样后特征图空间分辨率扩展至原来的2倍并在通道维度上与临级低阶特征保持一致，便于后续的通道拼接。通过增加非线性，在保留高阶信息的同时尽可能少地引入边缘误差；再将重建后的高阶特征与相邻跳跃连接传递的低阶特征进行通道拼接，并通过点卷积实现跨层级响应的像素级融合。这种改进方案在保持多尺度优势的同时，利用低阶特征的几何约束修正高阶特征的边缘失真，形成互补增强的特征表达。

图3　可视化实验（网络深度为3）

Fig. 3　Visualization experiment (with a network depth of three)




(a) 原始临边连接　　(b) 改进临边连接

图4　临边连接改进

Fig. 4　Improvement of adjacent edge connection ((a) original adjacent edge connection; (b) improved adjacent edge connection)

### 1.3　侧边聚合模块

低阶特征与高阶特征在线稿提取任务中呈现显著的互补特性。低阶特征源自网络浅层卷积核的局部响应，其高空间分辨率特性能够精确捕获像素级的高频纹理信息，然而，这些特征易受相似纹理干扰，在复杂背景区域会产生大量伪边缘响应。相较之下，高阶特征通过深层卷积核的抽象处理，形成了对图像语义结构的全局理解，能够有效识别主干轮廓的拓扑关系并抑制孤立噪声点，但其在多次下采样过程中损失了边缘的亚像素级定位精度。针对这种层级响应的表征差异，CLEAR-Net设计动态融合机制，具体为

$$\boldsymbol{F}_{\mathrm{DynFusion}} = SE\left(\mathop{\Vert}_{k=1}^{K} \boldsymbol{F}^{(k)}\right) \tag{2}$$

$$\boldsymbol{F}_{\mathrm{fusion}} = \sigma\left(\underbrace{DWConv_{3\times 3}(\boldsymbol{F}_{\mathrm{DynFusion}})}_{\text{局部上下文建模}} + \underbrace{DWConv_{1\times 1}(\boldsymbol{F}_{\mathrm{DynFusion}})}_{\text{跨通道关联}}\right) \tag{3}$$

式中，侧边聚合模块先在各侧边融合路径嵌入SE注意力模块，“$\Vert$”表示向量拼接。如图5所示，SE模块通过全局平均池化层压缩空间维度信息，生成通道描述向量后，经两层全连接层与激活函数构建通道注意力权重，实现基于图像语义内容的自适应特征重标定，然后将重标定后的特征图沿通道维度拼接，$\boldsymbol{F}_{\mathrm{DynFusion}}$表示动态拼接后的特征，拼接后的特征图形状为$H\times W\times(K\cdot C)$。接着，$\boldsymbol{F}_{\mathrm{DynFusion}}$经过双流深度可分离卷积，首层3×3卷积进行局部特征交互，次层1×1逐点卷积则建立跨通道关联，实现多尺度特征的像素级信息整合；最终通过端到端训练，使网络自主建立层级响应的最优融合策略，其优化过程同时考虑局部细节保真度与全局结构连贯性，形成判别性更强的混合特征表达。

### 1.4　联合损失函数

采用交叉熵损失作为基础监督信号，但在初步实验中发现模型输出存在两类缺陷：1）背景区域出

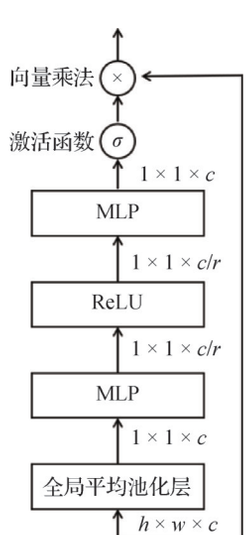



图5　SE模块

Fig. 5　Squeeze-and-excitation module

现与主体结构无关的伪边缘碎片，表现为离散点状噪声或短线段聚集现象，尤其在纹理复杂的服饰褶皱或环境阴影区域；2）线条存在颜色不均等质量问题，在边缘区域尤为显著。现有方法，如 Huan 等人（2022）提出的边界追踪强化学习算法，虽然能优化边缘连续性，但是对于背景伪影缺乏抑制。针对背景伪影问题，提出基于响应分布分析的抑制性损失函数，建立像素级虚假响应识别与惩罚体系：逐像素统计预测结果中的虚假响应，并根据其偏离程度进行不同程度的惩罚，从而抑制背景伪影。具体为

$$\mathcal{L}_{\text{sup}} = -\sum_{\substack{i=1 \\ \mathbb{I}_{\text{bg}}(i)=1}}^{H\times W} \omega(p_i, y_i)\cdot \mathcal{L}_{\text{CE}}(y_i, p_i) \tag{4}$$

式中，$\mathcal{L}_{\text{sup}}$ 是背景抑制损失函数，对背景伪影进行惩罚。$\mathbb{I}_{\text{bg}(\cdot)}$ 是指示函数，标记像素是否属于背景区域。$\mathcal{L}_{\text{CE}}(y_i, p_i)$ 是衡量模型预测值 $p_i$ 与真实标签 $y_i$ 之间差异的交叉熵损失函数。$\omega(p_i, y_i)$ 是像素级动态惩罚因子，$\mathbb{E}[y_{\text{bg}}]$ 用于调整每个像素的损失贡献，其计算为

$$\omega(p, y) = 1 + \gamma \cdot \left| p - \mathbb{E}[y_{\text{bg}}] \right| \tag{5}$$

式中，惩罚系数 $\gamma$ 赋予预测值和响应对比真实值偏离程度值的权重，可以根据实验过程进行动态调节；$\mathbb{E}[y_{\text{bg}}]$ 是标签中背景像素的真实值；式（5）先计算背景像素预测值和真实值的差异，再根据偏离程度计算出惩罚因子。

如图 6 所示，在单损失函数训练框架下，线稿生成模型对图像中的背景干扰要素表现出显著的敏感性。具体而言，当输入图像包含水渍痕迹、不规则阴影分布或复杂纹理背景时，模型的特征提取网络会将此类干扰要素误判为有效轮廓信息。进而，图像中的水渍、阴影等线条会被模型错误地响应，导致输出线稿中出现背景伪影。画师需要手工处理这些背景伪影，费时费力。而使用背景损失函数对背景伪影进行惩罚后，背景伪影消失，这种改进显著降低了人工修正的工作强度，提升了生成线稿的质量。

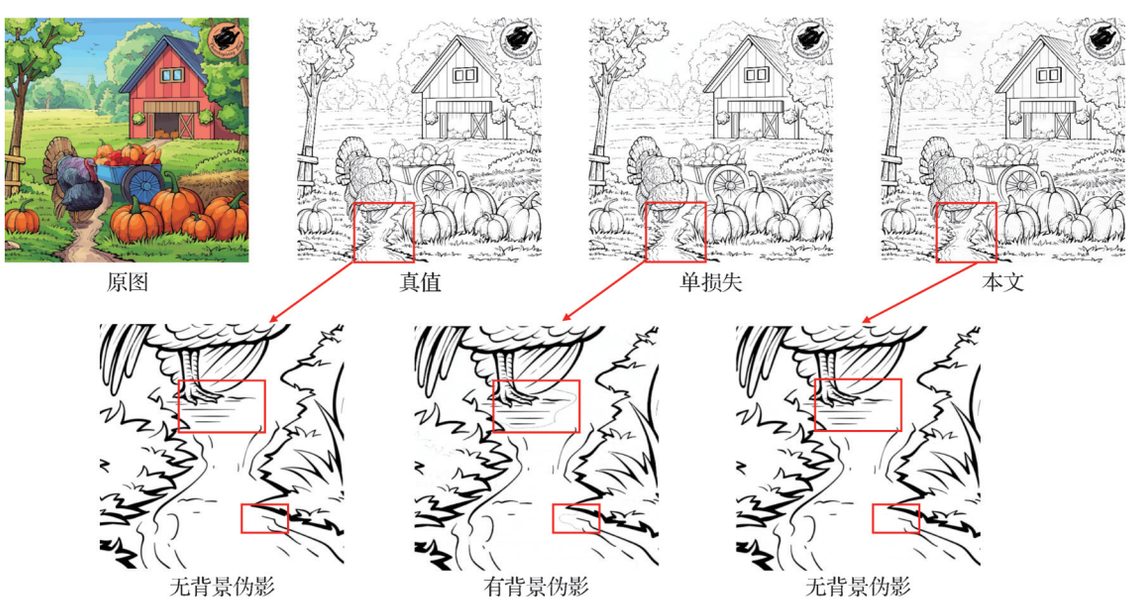



图 6　背景伪影

Fig. 6　Background artifacts

针对线条质量问题，传统交叉熵损失要求标签严格二值化（0/1），导致模型预测结果难以与极端真实值完全重合，进而产生以下问题：降低了模型对亚像素级位置偏差的容忍度。模型预测值在接近真实值时的梯度相对于其他梯度较小，从而停止优化。本工作的改进方案采用温度缩放与标签平滑的协同优化，平滑温度交叉熵为

$$\mathcal{L}_{\text{STCE}} = \mathcal{L}_{\text{CE}}(\tilde{y}, \sigma(z/\tau)) \tag{6}$$

$$\tilde{y} = \epsilon + (1 - 2\epsilon)y \tag{7}$$

$$\tau(t) = \max\left(0.8, 1.2 - 0.04 \cdot \lfloor t/10 \rfloor\right) \tag{8}$$

式中，首先对真实标签施加均匀分布平滑处理，将硬标签中的极端值 $y \in \{0, 1\}$ 转换为软标签 $\tilde{y} \in \{0 + \epsilon, 1 - \epsilon\}$，其中 $\epsilon$ 是一个极小数，用以控制平滑强度，平滑后的标签能够改善概率校准，增强模型对正确

极端预测值的支持，从而提升生成线稿的质量；其次在预测端引入温度系数$\tau = 0.8$缩放逻辑输出，加大概率分布$q = \sigma(z/\tau)$的差异，从而增强模型对极端预测结果的自信心。$z$为融合后的特征向量，为平衡训练稳定性与收敛速度，本工作设计分阶段升温策略：初始阶段设置高温$\tau = 1.2$，以拓宽参数搜索空间，每10个epoch线性降温至目标值$\tau = 0.8$，逐渐增强模型对极端预测值的支持。

训练初期，高温策略能够显著拓宽模型参数的搜索空间，使模型参数在优化过程中更容易向全局最优区域收敛。这一阶段，模型对细粒度特征的捕捉较为灵敏，能够建立较为全面的初始特征表示。随着训练的进行，模型参数逐渐逼近最优解，但参数梯度变小停止优化。温度平滑损失此时推动参数向极端值(0/1)偏移，以强化线条质量。这种对于极端值的支持，能够使背景抑制损失更好地工作，使高频噪声的预测值趋向纯白色，从而改善最终的视觉效果。最终模型的联合损失函数为

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{sup}} + \lambda_2 \mathcal{L}_{\text{STCE}} \tag{9}$$

式中，$\lambda_1$和$\lambda_2$是动态平衡因子，且$\lambda_1 + \lambda_2 = 1$。

## 2 实验与结果分析

实验硬件采用Intel Core i9-10980XE处理器(18核，3.00 GHz)和64 GB DDR4内存，搭配NVIDIA RTX 3090显卡(24 GB显存)。软件环境基于Ubuntu 20.04系统，构建PyTorch 1.12.1框架(CUDA 12.2)，主要依赖库包括Python 3.7、TorchVision 0.13.1、NumPy1.21.6、OpenCV4.8.1.78及scikit-image0.19.3。模型训练采用batch size = 5、learning rate = 0.001(Adam优化器，参数设置为betas =(0.9, 0.999)，weight decay = 0)、最大迭代次数 = 1 000的训练策略，同时在训练过程中设置断点重训机制与定期模型保存，以保证实验的稳定性与可重复性。

### 2.1 ArtLine-2K

针对线稿提取任务中高质量标注数据缺乏的现状，联合专业动画团队构建了首个亚像素级精度的手绘线稿数据集ArtLine-2K。该数据集包含2 000组严格配对的渲染图—线稿对，并经数据增强扩充到10 000对，数据集可联系作者获取。将ArtLine-2K按照6∶3∶1进行训练、验证和测试，评估不同模型的性能。

### 2.2 对比实验

本文采用多维评价体系对模型性能进行全面评估，具体包含以下维度：在精度评估方面，如图7所示，由于本文方法生成结果与真实标注(ground truth，GT)的视觉差异难以辨识(图像放大10倍后仅有细节差异)，故采用均方误差(mean squared error，MSE)和平均绝对误差(mean absolute error，MAE)作为核心量化指标，并统一将像素值归一化至[0, 1]区间进行精确计算。

同时，采用ODS(optimal dataset scale)和OIS(optimal image scale)衡量模型在数据集上单一尺度和不同尺度下的最佳表现。在速度性能方面，通过帧率(frames per second，FPS)指标衡量模型的实时检测能力。在模型复杂度方面，通过参数量指标反映模型的计算资源需求。这种多角度的评估既能客观反映生成质量与GT的差异性，又能系统评估模型在运算速度、训练效率和部署成本等实际应用场景中的综合表现。

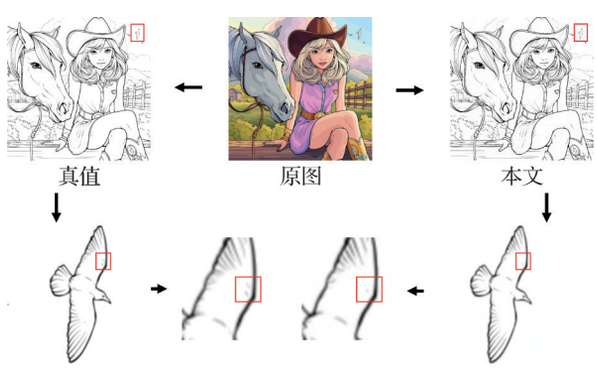



图7 细节对比

Fig. 7 Detail comparison

图8和图9通过ArtLine-2K数据集中的6组样本，对比了CLEAR-Net与当前主流模型TEED、DexiNed(dense extreme inception network for edge detection)、PiDiNet(pixel difference networks)、TANet、DiffusionEdge和DDN在线稿提取任务中的性能差异。

实验结果表明，在线稿提取任务中，本文模型具有显著优势。各模型虽然均能实现基础线稿提取功能，但在细节处理层面存在显著差异：TEED模型采用多尺度特征融合机制，其生成的线稿在局部纹理细节呈现能力突出，但受限于全局上下文感知不足，

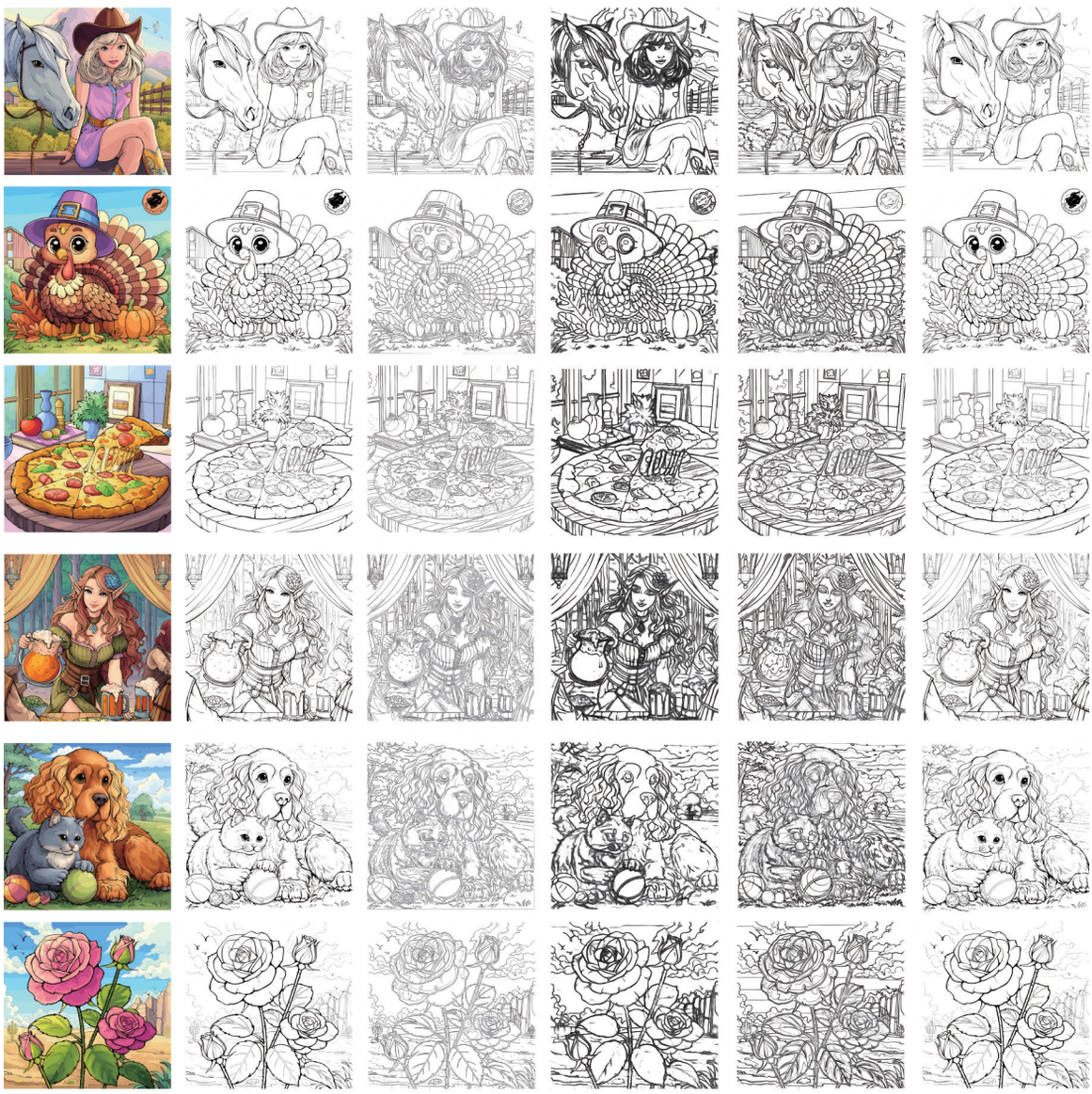



图 8　不同模型推理结果对比-1

Fig. 8　Comparison of the inference results of different models-1

导致图像主体结构出现不连续现象；DexiNed 主干信息保留较好，然而该模型在复杂背景场景中易产生伪边缘响应，显著影响线稿的纯净度；PiDiNet 对线条宽度缺乏有效约束机制，同时模型对背景（如水渍、阴影渐变）的抑制能力不足；其余模型均存在不同程度的上述问题。相较之下，本文方法在线条边缘信息的处理、背景伪影的控制、线条宽度的把控以及生成线条质量等方面均占优势。经专业画师评估，完全可以直接进行二次创作。

不同模型的性能对比如表 1 所示。实验数据表明，本文方法在 ArtLine-2K 数据集上的精度指标 MSE（0. 000 247）和 MAE（0. 004 81）实现了突破性提升，与真实标注的误差达到亚像素精度（MAE < 0. 005）。传统 Canny 方法虽然在实时性（390. 6 帧/s）上具有绝对优势，但其语义理解的缺失导致 MSE（0. 123 306）和 MAE（0. 132 221）显著高于深度学习模型，尤其在复杂纹理场景中产生大量虚假响应。在深度学习模型中，本文方法在保持轻量化的同时，达到了最好的 MAE 和 MSE，在量化指标上显著优于其他方法。

### 2. 3　消融实验

为了验证提出方法的有效性，使用 $U^2$-NetP 作为基准网络进行消融实验，实验结果如表 2 所示。实验通过逐步引入改进的特征提取模块、侧边融合机

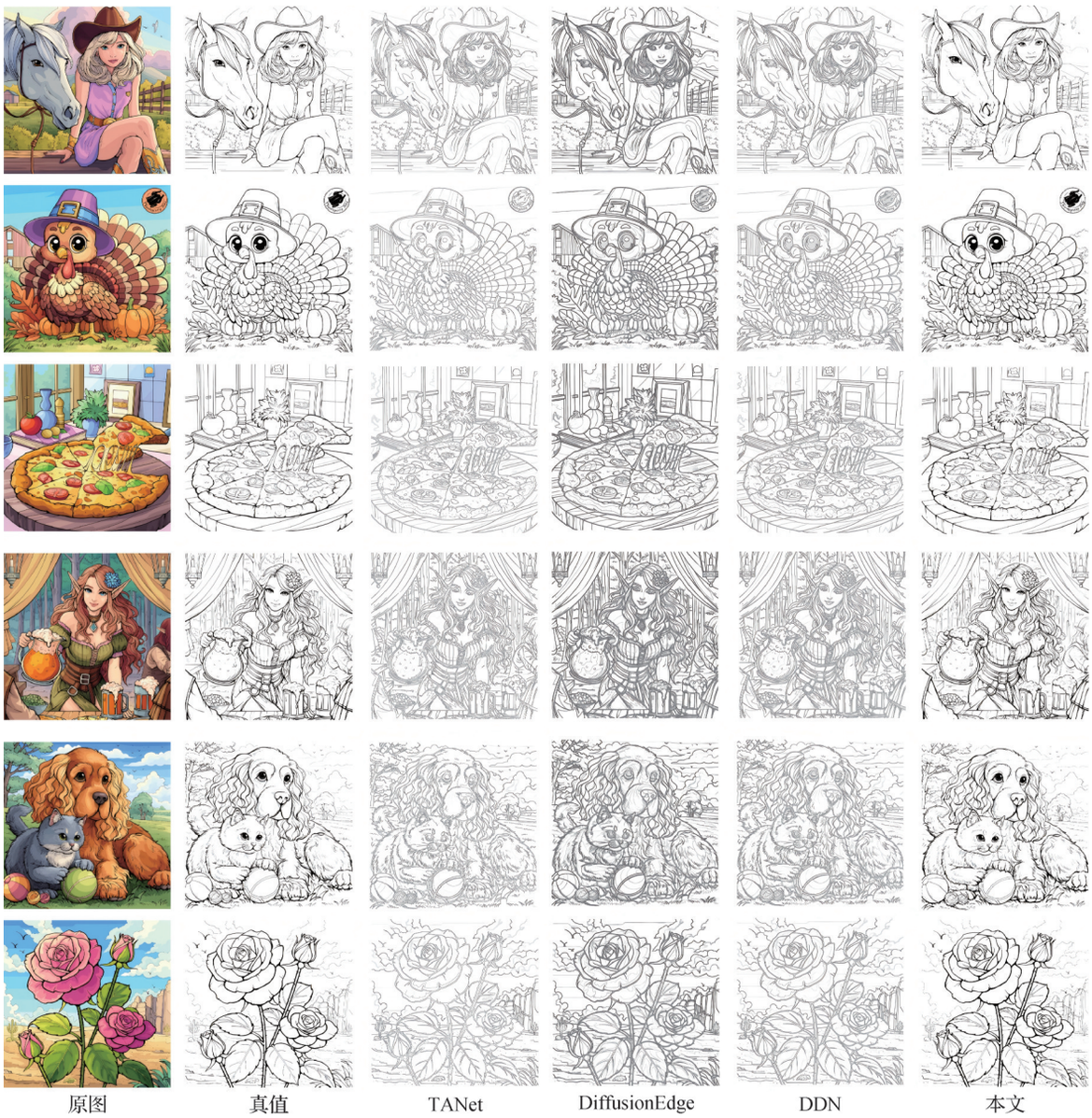



图9　不同模型推理结果对比-2

Fig. 9　Comparison of the inference results of different models-2

**表1　不同模型在ArtLine-2K数据集中的性能比较**

**Table 1　Performance comparison of different models on the ArtLine-2K dataset**

| 模型 | MSE ↓ | MAE ↓ | OIS ↑ | ODS ↑ | 帧率 ↑ | 参数量 ↓ |
|---|---|---|---|---|---|---|
| Canny(Canny, 1986) | 0.123 306 | 0.132 221 | 0.953 1 | 0.941 4 | **390.6** | – |
| TEED(Soria等, 2023) | 0.088 518 | 0.114 254 | 0.946 5 | 0.946 0 | 3.2 | **58.0 K** |
| DexiNed(Soria等, 2020) | 0.095 840 | 0.175 835 | 0.949 8 | 0.949 0 | 4.5 | 35.0 M |
| PiDiNet(Su等, 2021) | 0.107 059 | 0.177 215 | 0.942 6 | 0.941 2 | 3.1 | 710.0 K |
| TANet(Yang等, 2024) | 0.138 663 | 0.102 780 | 0.942 3 | 0.941 4 | 3.4 | 60.34 M |
| DiffusionEdge (Ye等, 2024) | 0.140 800 | 0.060 476 | 0.951 0 | 0.952 2 | 0.8 | 2.25 M |
| DDN(Li等, 2025) | 0.128 721 | 0.077 803 | 0.966 3 | 0.966 0 | 5.8 | 41.2 M |
| 本文 | **0.000 247** | **0.004 810** | **0.998 5** | **0.998 1** | 2.2 | 5.9 M |

注：加粗字体表示各列最优结果，“–”表示传统方法无参数。“↑”表示值越高越好，“↓”表示值越低越好。

制、背景抑制损失和平滑升温损失等4种方法，系统验证了各模块对模型性能的贡献。实验数据表明：基准模型 U²-NetP 在未引入任何改进时，MSE为0.042 171，MAE为0.101 493。当引入特征提取模块改进后，MSE和MAE分别下降30.0%和31.5%（0.029 528和0.069 548），证明该模块显著提升了特征表达能力。进一步引入侧边融合机制后，误差指标呈现断崖式下降（MSE为0.007 921，MAE为0.032 119），降幅达73.2%和53.8%，表明多尺度特征融合对边界保持具有关键作用。引入背景抑制损失后，MSE和MAE分别降至0.001 283和0.011 741，较前一阶段降幅达83.8%和63.4%，验证了该损失函数对背景噪声抑制的有效性。最终引入平滑升温交叉熵时，模型达到最优性能（MSE为0.000 247，MAE为0.004 810），较基准模型整体误差降低99.4%和95.3%，说明各组件的有效性。总体而言，各模块的协同作用呈现明显的累加效应，其中侧边融合和背景抑制损失对精度提升贡献最为显著。

**表2　消融实验**

**Table 2　Ablation experiment**

| 架构 | 特征提取模块改进 | 侧边融合 | 背景抑制损失 | 平滑升温交叉熵 | MSE↓ | MAE↓ |
|---|---|---|---|---|---|---|
| U²-NetP | – | – | – | – | 0.042 171 | 0.101 493 |
| U²-NetP | √ | – | – | – | 0.029 528 | 0.069 548 |
| U²-NetP | √ | √ | – | – | 0.007 921 | 0.032 119 |
| U²-NetP | √ | √ | √ | – | 0.001 283 | 0.011 741 |
| U²-NetP | √ | √ | √ | √ | **0.000 247** | **0.004 810** |

注：加粗字体表示各列最优结果，“√”表示采用，“–”表示未采用。“↓”表示值越低越好。

消融实验的可视化结果如图10所示。在仅使用基础网络（U²-NetP）进行线稿提取时，可以看到提取出的线稿分辨率不足、线条质量差，且在高频背景区域（水渍、阴影等）产生了背景伪影；改进特征提取网络后，引入反卷积增加网络的非线性，充分利用浅层特征的高频纹理信息和深层特征的骨干信息，线稿图线条质量变好，但是线条边缘噪声增大；接着对各侧边引入动态聚合模块，通过权重重标定和双流聚合，消除了线条边缘的噪声，但背景伪影问题仍未解决；然后创新性地提出了背景抑制监督机制，动态地对背景伪影进行像素级惩罚，背景伪影变淡。但在网络进行端到端的训练时，预测值在非常接近真实值时由于梯度较小会停止优化，故背景伪影未完全消失；最后针对仍存在的问题，对交叉熵损失函数

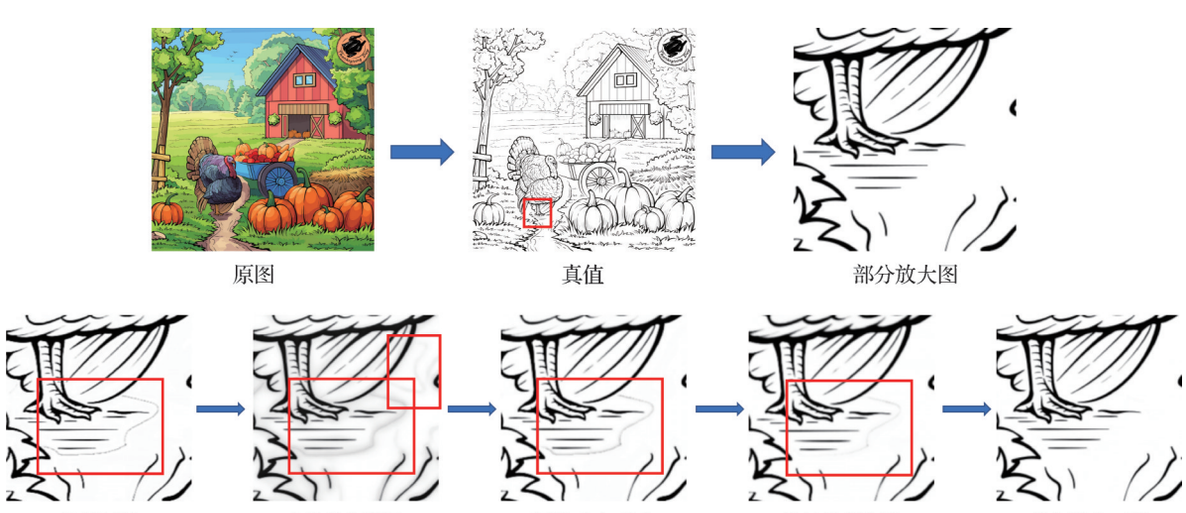



图10　消融实验结果可视化

Fig. 10　Visualization of the ablation experiment results

进行改进，并联合背景抑制损失共同监督，增强网络对极端真实值的信心，背景伪影完全消失，线条质量变优，网络达到最优。

### 2.4 泛化实验

为了验证CLEAR-Net的泛化能力，进一步引入了标注质量相对较低的Anime Sketch Colorization Pair数据集进行泛化性实验。该数据集的线稿与上文使用的ArtLine-2K数据集在标注风格与质量上存在显著差异，尤其表现为线条粗细不均、手绘规则不一致以及噪声较多等问题。实验严格按照数据集官方的划分进行训练与验证，训练策略和超参数设置均保持与前文相同。

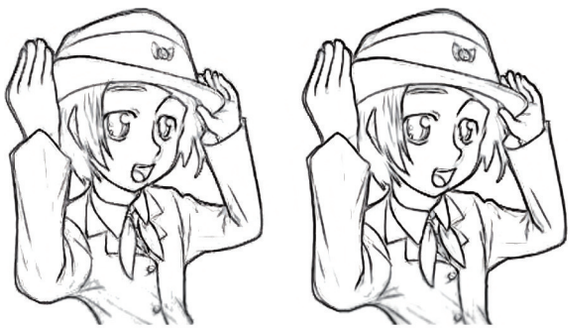
实验结果如图11所示，CLEAR-Net在Anime Sketch Colorization Pair数据集上依然能够较为准确地学习到该数据集的线稿手绘规律，显示出模型在不同数据域之间的良好迁移能力。尽管由于原始标注质量的限制，提取线稿的质量低于在ArtLine-2K上提取线稿的质量，但CLEAR-Net仍能够捕捉到线条结构、局部细节。

值得注意的是，如图12所示，由于CLEAR-Net优秀的像素级特征处理能力和噪声惩罚机制，使生成的图像不仅在视觉上线条更加平滑，且整体噪声水平低于真实标注，如人物五官线条更平滑、服饰褶皱处噪点更少，体现了模型在处理低质量数据时的鲁棒性。

原图　　真值　　本文

图11　泛化实验结果

Fig. 11　Generalization experiment results



真值　　本文

图12　泛化实验细节对比

Fig. 12　Comparison of generalization experiment details

## 3 结　论

本文针对动漫线稿提取任务，提出了基于跨层级响应融合与联合损失优化的CLEAR-Net模型。通过改进$U^2$-Net构建特征提取模块，在控制网络深度的同时实现浅层细节与深层语义的协同表达；设计的特征精修策略有效缓解了传统跳跃连接导致的边缘模糊问题；侧边聚合模块通过SE注意力机制与双流深度可分离卷积的级联结构，实现了多层级响应的自适应融合；联合损失函数创新性地结合背景抑制惩罚与平滑温度交叉熵，在抑制背景伪影的同时提升了线条质量。实验表明，该模型在保持较高推理速度的前提下，在ArtLine-2K数据集上取得了显著优于现有方法的性能指标。

然而，本研究仍存在局限性：本文模型主要针对颜色简单的平涂风格图像，对具有复杂材质的厚涂类动漫图像适应性不足，其颜色多样性易导致边缘响应混淆。未来工作将重点围绕复杂材质线稿提取展开：1）构建颜色不变性特征提取模块，通过解耦色彩信息与结构特征降低复杂涂色干扰；2）设计材质感知损失函数，结合材质语义分割信息强化对纹理边缘的判别能力；3）探索多模态特征融合机制，联合分析线条走向与材质渐变规律，提升复杂光影场景下的边缘定位精度。

**致谢**：本研究得到“AI涂色内容生成及线稿绘制项目”（项目编号：NKRJ-YW-202507001）支持。衷心感谢西安纽扣软件科技有限公司对本研究的大力支持。特别感谢该公司协助构建了高精度手绘线稿数据集ArtLine-2K，其专业艺术团队的高质量渲

染与标注工作为本研究提供了坚实的数据基础，保障了实验的顺利开展。

## 参考文献(Reference)

Arbeláez P, Maire M, Fowlkes C and Malik J. 2011. Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(5): 898-916 [DOI: 10.1109/TPAMI.2010.161]

Canny J. 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6): 679-698 [DOI: 10.1109/TPAMI.1986.4767851]

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1800-1807 [DOI: 10.1109/CVPR.2017.195]

Deng R X and Liu S J. 2020. Deep structural contour detection//Proceedings of the 28th ACM International Conference on Multimedia. Seattle, USA: ACM: 304-312 [DOI: 10.1145/3394171.3413750]

Guo Y H, Li Y D, Wang L Q and Rosing T. 2019. Depthwise convolution is all you need for learning multiple visual domains//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI: 8368-8375 [DOI: 10.1609/aaai.v33i01.33018368]

He J Z, Zhang S L, Yang M, Shan Y H and Huang T J. 2019. Bi-directional cascade network for perceptual edge detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3828-3837 [DOI: 10.1109/TPAMI.2020.3007074]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huan L X, Xue N, Zheng X Q, He W, Gong J Y and Xia G S. 2022. Unmixing convolutional features for crisp edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10): 6602-6609 [DOI: 10.1109/TPAMI.2021.3084197]

Jing J F, Liu S J, Wang G, Zhang W C and Sun C M. 2022. Recent advances on image edge detection: a comprehensive review. Neurocomputing, 503: 259-271 [DOI: 10.1016/j.neucom.2022.06.083]

Kull M, Perello-Nieto M, Kängsepp M, Silva Filho T, Song H and Flach P. 2019. Beyond temperature scaling: obtaining well-calibrated multiclass probabilities with Dirichlet calibration//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #1103 [DOI: 10.5555/3454287.3455390]

Li Y C, Poma X S, Xi Y K, Li G L, Yang C Z, Xiao Q, et al. 2025. A doubly Decoupled Network for edge detection. Neurocomputing, 624: #129442 [DOI: 10.1016/j.neucom.2025.129442]

Liu D and Wang Y H. 2020. Constraint self-adaptive SUSAN algorithm for edge detection. Journal of Computer-Aided Design and Computer Graphics, 32(6): 971-978 (刘丹, 王运宏. 2020. 限制型自适应SUSAN边缘检测算法. 计算机辅助设计与图形学学报, 32(6): 971-978) [DOI: 10.3724/SP.J.1089.2020.17996]

Liu Y, Cheng M M, Hu X W, Wang K and Bai X. 2017. Richer convolutional features for edge detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5872-5881 [DOI: 10.1109/CVPR.2017.622]

Qin X B, Zhang Z C, Huang C Y, Dehghan M, Zaiane O R and Jagersand M. 2020. $U^2$-Net: going deeper with nested U-structure for salient object detection. Pattern Recognition, 106: #107404 [DOI: 10.1016/j.patcog.2020.107404]

Sert E and Avci D. 2019. A new edge detection approach via neutrosophy based on maximum norm entropy. Expert Systems with Applications, 115: 499-511 [DOI: 10.1016/j.eswa.2018.08.019]

Soria X, Li Y C, Rouhani M and Sappa A D. 2023. Tiny and efficient model for the edge detection generalization//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 1356-1365 [DOI: 10.1109/ICCVW60793.2023.00147]

Soria X, Riba E and Sappa A. 2020. Dense extreme inception network: towards a robust CNN model for edge detection//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, USA: IEEE: 1912-1921 [DOI: 10.1109/WACV45572.2020.9093290]

Su Z, Liu W Z, Yu Z T, Hu D W, Liao Q, Tian Q, et al. 2021. Pixel difference networks for efficient edge detection//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 5097-5107 [DOI: 10.1109/ICCV48922.2021.00507]

Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z. 2016. Rethinking the inception architecture for computer vision//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2818-2826 [DOI: 10.1109/CVPR.2016.308]

Wang S Q, Zhang J Q, Shi M and Zhao Y J. 2021. Image extraction of cartoon line art based on cycle-consistent adversarial networks. Journal of Image and Graphics, 26(5): 1117-1127 (王素琴, 张加其, 石敏, 赵银君. 2021. 循环生成对抗网络的线稿图像自动提取. 中国图象图形学报, 26(5): 1117-1127) [DOI: 10.11834/jig.200465]

Xie S N and Tu Z W. 2015. Holistically-nested edge detection//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1395-1403 [DOI: 10.1109/ICCV.2015.164]

Xuan W J, Huang S L, Liu J H and Du B. 2022. FCL-Net: towards accurate edge detection via fine-scale corrective learning. Neural Networks, 145: 248-259 [DOI: 10.1016/j.neunet.2021.10.022]

Yang X, Cheng L F, Yuan G W and Wu H. 2024. Textureness-aware

neural network for edge detection//Proceedings of the 7th Chinese Conference on Pattern Recognition and Computer Vision. Urumqi, China: Springer: 254-268 [DOI: 10.1007/978-981-97-8505-6_18]

Ye Y F, Xu K, Huang Y H, Yi R J and Cai Z P. 2024. DiffusionEdge: diffusion probabilistic model for crisp edge detection//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 6675-6683 [DOI: 10.1609/aaai.v38i7.28490]

Zhu J Y, Park T, Isola P and Efros A A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2242-2251 [DOI: 10.1109/ICCV.2017.244]

Zhu W, Cen K, Xu X Z and He D F. 2021. Edge detection network with multi-depth feature enhancement and top-level information guidance. Journal of Computer-Aided Design and Computer Graphics, 33(11): 1705-1714 (朱威, 岑宽, 徐希舟, 何德峰. 2021. 多深度特征增强与顶层信息引导的边缘检测网络. 计算机辅助设计与图形学学报, 33(11): 1705-1714) [DOI: 10.3724/SP.J.1089.2021.18752]

## 作者简介

鲁方博,男,硕士研究生,主要研究方向为图像处理和计算机视觉。E-mail:599385027@nwafu.edu.cn

王美丽,通信作者,女,教授,博士生导师,主要研究方向为计算机图形学和计算机视觉。E-mail:wml@nwsuaf.edu.cn

罗万闯,男,硕士研究生,主要研究方向为图像处理和计算机视觉。E-mail:luowanchuang@nwafu.edu.cn

乔永源,男,本科生,主要研究方向为计算机视觉。E-mail:qiaoyongyuan@nwafu.edu.cn

葛贤钰,男,工程师,主要研究方向为图像处理和计算机视觉。E-mail:eric@buttontech.com

贺鹏,男,工程师,主要研究方向为计算机视觉。E-mail:johnson.he@buttontech.com