



· 文化遗产数字化保护 ·

基于 Transformer 两阶段策略的 古代服饰线图提取

周蓬勃¹, 冯 龙², 武浩东², 寇宇帆²

(1. 北京师范大学艺术与传媒学院, 北京 100032;

2. 西北大学 文化遗产数字化国家地方联合工程研究中心, 陕西 西安 710127)

摘要 古代服饰线图提取旨在精确获取轮廓与形状信息, 以助于再创作和传统服饰保护。但现有方法增加网络以提高泛化性, 导致参数量大增。为此, 提出了基于 Transformer 的两阶段边缘检测方法, 旨在解决图像局部信息丢失以及模型参数量大的问题。第一阶段将图像分割成 16×16 粗粒度补丁, 利用编码器进行全局自注意力计算以捕获补丁间依赖; 第二阶段采用 8×8 细粒度无重叠滑动窗口覆盖图像, 通过局部编码器计算窗口内注意力有效捕捉细微边缘且降低成本。设计了轻量特征融合模块, 支持全局与局部特征的高效整合。实验结果表明, 该方法在古代服饰和公共数据集上边缘轮廓信息提取效果优于现有方法, ODS 指标平均提升 15.9%。虽然 OIS 和 AP 未超过 Informative Drawing, 但在模型体量和耗时方面具有明显优势。

关键词 边缘检测; Transformer; 轻量特征融合模块

中图分类号: TP391.41 **DOI**: 10.16152/j.cnki.xdxbr.2025-01-006

Ancient clothing line drawing extraction based on Transformer two-stage strategy

ZHOU Pengbo¹, FENG Long², WU Haodong², KOU Yufan²

(1. School of Art and Media, Beijing Normal University, Beijing 100032, China;

2. National and Local Joint Engineering Research Center for Cultural Heritage Digitization,
Northwest University, Xi'an 710127, China)

Abstract The extraction of ancient costume line drawings aims to precisely obtain contour and shape information to aid in re-creation and traditional preservation. However, existing methods increase network depth to improve generalization, leading to a significant increase in the number of model parameters. Therefore, this paper proposes a two-stage edge detection method based on Transformer, aiming to solve the problems of local information loss in images and large model parameter sizes. The first stage divides the image into 16×16 coarse-grained patches and uses an encoder to perform global self-attention calculations to capture dependencies between patches; the second stage covers the image with an 8×8 fine-grained non-overlapping sliding window and calculates the attention within the window through a local encoder to effectively capture subtle edges and reduce costs. A lightweight feature fusion module is designed to support efficient integration of global and local

收稿日期: 2024-04-20

基金项目: 国家自然科学基金(62271393); 国博文旅部重点实验室开放课题(1222000812, CRRT2021K01)。

第一作者: 周蓬勃, 男, 博士, 高级工程师, 从事数字艺术与虚拟现实研究, zhoupengbo@bnu.edu.cn。

features. Experimental results show that this method outperforms existing methods in extracting edge contour information on ancient costume and public datasets, with an average improvement of 15.9% in the ODS metric. Although OIS and AP does not surpass Informative Drawing, this method shows obvious advantages in model size and time consumption.

Keywords edge detection; Transformer; lightweight feature fusion module

中国古代服饰文化是优秀传统文化的重要组成部分,承载着华夏五千年的文化内涵和符号,既是中华物质文明的载体,也承载了中华精神文明。然而,由于服饰是有机质文物,极易受到光照、温度和湿度的影响,因此多数情况下,这类文物只能长期珍藏于博物馆,难以用于进一步的研究和创作。然而,古代服饰上的纹样等元素,却是中华文化的重要组成部分。本文以古代服饰为对象,利用边缘检测方法获取古代服饰的线画图,用于二次开发和创作,为实现中华古代传统服饰文化的数字化保护和传承奠定深厚的基础。

边缘检测是图像处理和计算机视觉领域的基础技术,识别图像中的物体轮廓和边界^[1]。边缘指图像中亮度变化显著的地方,这些变化往往代表了物体的轮廓、场景分割线或其他重要的视觉特征。边缘检测对于物体识别^[2]、场景理解、图像分割和视觉效果增强等都具有重要作用。

传统的边缘检测方法如 Canny 边缘检测器^[3]等,具有低错误率、单一响应、强边缘和弱边缘通过双阈值划分等优点。但是对噪声较为敏感,一些算法(如 Canny)^[4-6]需要人工设定参数^[7](例如阈值高低),可能会导致边缘模糊或边缘信息不连续情况下难以定位边缘结构。

相较于传统方法,深度学习方法具有更强的细节捕捉能力和抗噪性。近年来,诸多学者基于深度学习方法对二维图像的边缘检测展开研究^[8-9]。EDTER^[10]使用全局编码器捕获粗粒度图像全局上下文信息,使用局部编码器挖掘细粒度图像短距离的局部线索,在 BSDS500、NYUDv2 等多个边缘检测上展示优越的性能^[11-12]; SwinNet^[13]基于 Swin Transformer 作为骨干网络,同时结合 CNN 的局部检测优势,类似的还有 EGCTNet^[14]; ECT^[15]使用 Transformer 捕获全局上下文信息,此外,引入了边缘聚合和对齐损失确保边缘一致性;为了捕获图像表面反射率、亮度、表面法线和深度信息不连续(RIND)造成的边缘,SWIN-RIND^[16]提出一种基于注意力机制和 Swin Transformer 的新型边缘检测方法^[17],能够对单一输入

图像进行边缘检测。通过自上而下和自下而上的多级特征聚合块作为解码器,将不同层次的特征融合成富含上下文信息的共享特征,显著提高了准确性和视觉性能。古代服饰具有服饰面料柔软、纹理特征丰富、形状复杂等特点。Transformer 基于自注意力机制能够捕获位置之间的长程依赖关系,从而高效提取图像全局上下文信息。Swin Transformer^[17]是一种全新的 Transformer 架构,涵盖了包括图像分类和语义分割等多个领域,可以作为计算机视觉的通用骨干网络。

DexiNed^[18]是一种基于 Transformer 的端到端边缘检测方法^[19-20]。其主要目标是通过 Transformer 的全局感知能力来捕捉图像中的长距离依赖关系,从而提高边缘检测的准确性。LDC^[21]旨在通过结合局部和全局上下文信息来提高边缘检测的准确性。该方法利用 Transformer 的多头注意力机制来捕捉图像中的多层次特征。MangaLine^[22]采用了一种轻量级的 Transformer 架构,并结合了特定的卷积模块来处理漫画图像中的独特纹理和色彩特征,在漫画风格图像的边缘检测任务上取得了优异的成绩,特别是在处理线条粗细变化较大的图像时。Informative Drawing^[23]结合了 Transformer 结构和绘图知识,旨在增强边缘检测的鲁棒性和细节保留能力。该方法通过引入绘图规则和约束,指导 Transformer 更好地捕捉图像中的边缘信息。上述方法在线稿连续的图像上效果显著,然而服饰文物的线稿并不连续。

针对上述问题,本文提出基于 Transformer 两阶段策略的古代服饰线图提取方法。该方法采用 Swin Transformer 中的移位窗口来替代传统的非重叠等大图像补丁窗口,充分提取图像全局与局部特征,此外,引入了轻量特征融合模块,用于融合全局和局部特征,解决线稿图不连续的问题。将本文方法在 PASCAL VOC^[24]、NYUDv2 和古代服饰等多个数据集上进行测试,并与 DexiNed、LDC、MangaLine 以及 Informative Drawing 等多个基线方法进行对比。实验结果表明,该模型在图像边缘检测任务上的性能较好,具有较强的泛化

性和有效性。

1 基于 Transformer 两阶段策略古代服饰边缘检测方法

卷积神经网络存在因下采样导致的局部信息丢失问题^[6]以及随感受野扩大局部细节逐渐被抑制等问题,因此,本文提出了基于 Transformer 的边缘检测方法,Transformer 可以有效地捕获图像的全局信息,并且不受到局部感受野的限制,使其在处理高分辨率的图像时能更好地理解图像结构和语义信息,而且在输入尺寸方面具有极高的灵活性。此外,采用端到端的方式进行学习,能够实现从原始图像到边缘检测结果的映射。

本文提出一种特征融合模型,旨在有效地学习和整合图像的全局和局部特征,涵盖了粗粒度和细粒度的补丁信息。此模型具备针对不同尺度特征操作的能力,能够捕获多样的上下文信息。该方法通过 2 个阶段分别提取全局和局部特征,并在此基础上进行特征融合。在特征融合阶段,引入了轻量特征融合模块(light feature fusion module, LFFM),用于融合全局和局部线索,以及

使用相同的双向特征聚合解码器(bi-directional multi-level aggregation, BiMLA)。另外,采用 Swin Transformer 中的移位窗口来替代传统的非重叠等大图像补丁窗口。

1.1 网络结构

假设用 A 和 B 分别表示源图像和预测图像,例如秦腔服饰原图和秦腔服饰线图。使用 f_g 表示全局特征,使用 f_r 表示局部细化特征。本节的目标是通过基于全局特征和局部特征的 Transformer 方法融合来推断 A 和 B 之间的有效映射,因此,本章的方法共包含 2 个转换模型: $A \rightarrow f_g, A \rightarrow f_r, f_g, f_r \rightarrow B$,其中, f_g 与 f_r 处理过程较为相似,均是使用双向聚合特征 BiMLA 解码器得到,但是在图像的处理细节及粒度上有区别。其中,由 A 图像域转换到 f_g 的处理过程如图 1 中 Stage 1 所示,主要目标是提取出全局特征,这个过程也可以得到一个较为粗糙的结果线图。由 A 图像域转换到 f_r 的处理过程如图 1 中 Stage 2 阶段 LocalRefinement,这个过程将图像分成 64 份,每组 16 份,在进行线性映射之后会嵌入位置信息,然后通过 BiMLA 解码器处理特征图。

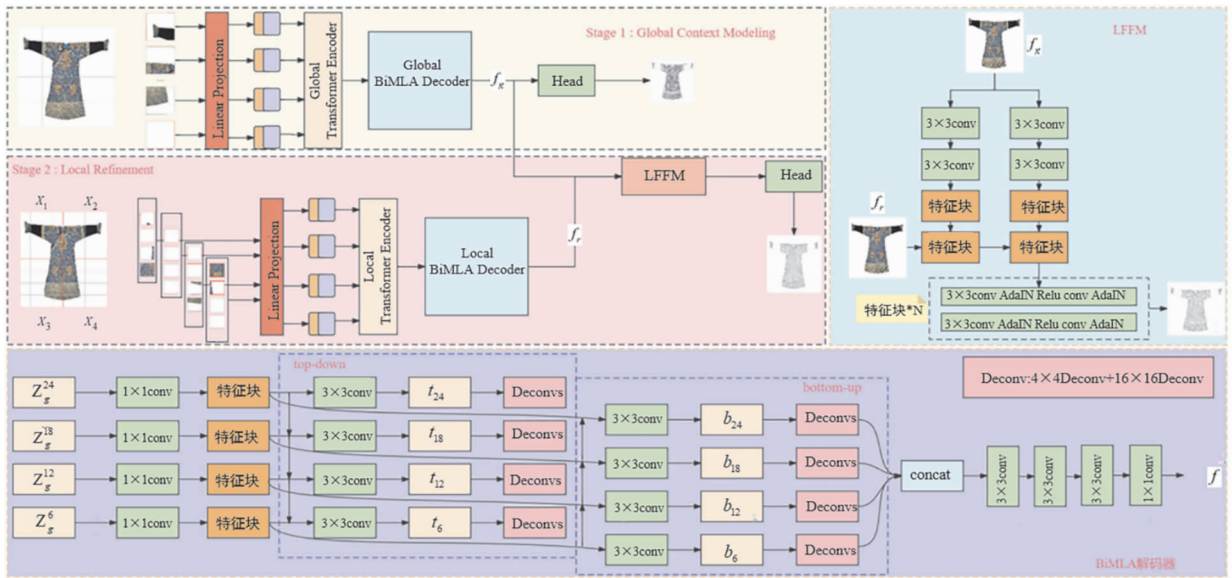


图 1 两阶段特征提取及融合示意图

Fig. 1 Schematic diagram of two stage feature extraction and fusion

对于 BiMLA 解码器,采用 2 条路径,一个自顶向下,一个自底向上,通过对特征块进行卷积、反卷积操作,然后将 2 条路径特征图连接后再次进行 3 次 3×3 卷积和 1×1 卷积得到特征 $f(f_g$ 和 $f_r)$ 。其中的 LFFM 模块取消了原来的特征自连

接,并且通过新的 2 条路径得到新的特征图随后进行融合,最终得到预测的服饰线图。

具体来说,首先需要对图像进行分区处理。将原始图像使用补丁分割成一系列大小与补丁相等的扁平图像块。每个 Transformer 编码器包含

若干 Transformer Block 块,其中又包含多头自注意操作(MSA)、1 个多层感知器(MLP)和 2 个层规范(LN)。最后并行执行 M 个自注意力操作,并将其连接输出进行投影。

该研究所提出方法可处理任意输入图像完成边缘检测任务,解决了 CNN 中局部信息丢失、依赖大量训练数据以及过拟合等问题。该方法在预测线图质量、训练速度及保留全局和局部细节特征方面表现出良好的平衡性。

1.2 两阶段特征提取

图像的结构主要依赖于低维特征,而高维特征在维持图像的纹理和颜色方面起到了关键作用。尽管高维特征中蕴含了丰富的细节,但其语义信息相对不足。此外,在深度学习的网络中,随着网络深度的增加,梯度消失问题变得更加突出,导致低维特征难以被充分利用。采用两阶段特征提取方法能够有效处理低维特征和高维特征。

第一阶段,采用全局 Transformer 编码器和解码器来探索粗粒度上下文特征。将图像分割成大小为 16×16 像素的粗粒度补丁,随后生成嵌入输入,编码器对全局进行自注意力计算捕获补丁之间的依赖。具体见公式(1),

$$Z_g = \{z_g^1, z_g^2, z_g^3, \dots, z_g^{L_g}\} = G_E(z_g^0) \quad (1)$$

式中: $z_g^1, z_g^2, \dots, z_g^{L_g} \in R^{\frac{HW}{256} \times C}$ 表示 G_E 中连续块地输出, L_g 是 G_E 中 Transformer Block 的个数。

BiMLA 解码器的设计促进了边缘感知任务中的注意力计算,它采用可学习的上采样策略处理注意力。该方法首先将 L_g Transformer Block 划分成 4 组,并以每组最后一个的嵌入特征 $\{z_g^6, z_g^{12}, z_g^{18}, z_g^{24}\}$ 作为输入,重新构造特定尺寸为 $\frac{H}{16} \times \frac{W}{16} \times C$ 的 3D 特征。在自上而下的路径中,对重构特征进行相同的操作(包括 1 个 1×1 卷积层和 1 个 3×3 卷积层),获得 4 个输出特征 $\{t^6, t^{12}, t^{18}, t^{24}\}$ 。同时,从最低层(z_g^6)开始,自下而上的路径逐渐接近最高层(z_g^{24}),并在多层特征上附加 1 个 3×3 的卷积层,最终产生另外 4 个输出特征 $\{b^6, b^{12}, b^{18}, b^{24}\}$ 。除此之外,BiMLA 在将每个聚合特征传递至下一层时,运用了反卷积块^[25]。该块包含 2 个反卷积层,分别使用 4×4 和 16×16 核,随后是批归一化和 ReLU 操作。通过这种方式,BiMLA 将双向路径采样得到的 8 个特征进行连接,形成一个统一的张量。为了增强特征的平滑性,BiMLA 还引入了额外的卷积堆栈来处理这些连接特征。

BiMLA 解码器过程见公式(2),

$$f_g = G_D(z_g^6, z_g^{12}, z_g^{18}, z_g^{24}) \quad (2)$$

式中: f_g 代表像素级全局特征; G_D 是全局 BiMLA 解码器。

边缘的精细度对于许多应用领域至关重要,因此探索像素级预测的细粒度上下文特征是十分必要的。然而,考虑到理想的边缘宽度通常仅 1 个像素,使用较大的 16×16 的补丁进行特征提取在捕获细微边缘时存在局限性。同时,若直接使用像素大小的补丁会导致计算成本急剧上升,在实际操作中不可取。方法是采用无重叠滑动窗口($H/2 \times W/2$)在图像 $X \in R^{H \times W \times 3}$ 上进行覆盖,将输入图像 X 分解为序列 $\{X^1, X^2, X^3, X^4\}$ 。

第二阶段,以 8×8 的细粒度补丁为单位,通过共享的局部编码器 R_E 计算窗口内的注意力,并将所有窗口的注意力特征连接在一起,获得 $z_r = \{z_r^1, \dots, z_r^{L_r}\} \in R^{\frac{HW}{64} \times C}$ 。为了降低计算量,设置局部 Transformer 编码器数量为 12 个。与全局 BiMLA 类似,采用二阶段方法,从 z_r 中均匀选取 $\{z_r^3, z_r^6, z_r^9, z_r^{12}\}$ 并将其输入到局部 BiMLA 中,生成高分辨率局部特征。

1.3 轻量级特征融合模块

采用改进的轻量特征融合模块(LFFM)结合 2 个级别的上下文信息,通过局部决策头来推断边缘结果图。相较于之前的 FFM 模块,LFFM 模块由空间特征块及 2 个 3×3 卷积层构成,并经过批归一化和 ReLU 操作,前者实现调制,后者完成平滑工作。值得一提的是,LFFM 模块采用基于 FDC^[26]结构的设计思路进行实现。

通道空间注意力机制^[27](channel-spatial attention mechanism, CSAM)融合通道和空间注意力^[28],以更好地捕获图像特征之间的通道相关性和空间位置的重要性,从而提高模型性能。首先计算通道权重因子并应用于特征图的不同通道特征上,对其进行加权累加得到权重特征图。其次,利用空间注意力对加权特征图的每个位置进行权重评估,以确定该位置在特征图中的重要程度。最后,通过将这些权重应用于已加权的特征图,对各个空间位置的特征进行加权求和,从而生成最终的特征表示。这一过程相当繁琐,为简化这一过程,提出了轻量特征融合模块 LFFM。

为了减少网络训练过拟合的可能性,进行了对密集卷积的移除操作。在 LFFM 模块中,首先对全局特征图进行 2 次分支计算,每个分支包括

2个 3×3 的卷积操作。然后,左侧分支与右侧分支进行加权后连接至右侧分支,接着依次进行卷积、归一化、ReLU、卷积、归一化等操作。

通过改进的通道注意力机制,有效减少了特征融合时的计算参数,从而提高了训练速度。同时,引入了自注意力机制,在提取的线图结果上也取得了良好效果。

1.4 损失函数

总损失由全局损失和局部损失两部分组成。给定线图 E ,对应源图像 Y ,定义总损耗为

$$l(E, Y) = - \sum_{i,j} (Y_{i,j} \alpha \log(E_{i,j}) + (1 - Y_{i,j})(1 - \alpha) \log(1 - E_{i,j})) \quad (3)$$

式中: $E_{i,j}$ 和 $Y_{i,j}$ 分别指代矩阵 E 和 Y 的第 (i, j) 元素。而 $\alpha = \frac{|Y^-|}{|Y^- + Y^+|}$ 则用于表示总样本中负像素所占的百分比。实际操作中,首先将数据集中的多个标签归一化为边缘概率图,其取值范围 $[0, 1]$ 。随后,应用一个预设的阈值 η 来筛选像素点,若某像素点的概率值超过 η ,则将其标记为正样点,反之,将其标记为负样点。

在训练阶段1,首先将全局决策头合并到全局特征映射上,表示为公式(4),

$$E_g = G_H(f_g) \quad (4)$$

式中:引入一个全局决策头,用 G_H 表示,它由 1×1 卷积层和Sigmoid函数构成。

另外,针对全局BiMLA提取的中间特征 t^6 、 t^{12} 、 t^{18} 、 t^{24} 和 b^6 、 b^{12} 、 b^{18} 、 b^{24} 执行相似操作,即使用 4×4 反卷积层和 16×16 反卷积层,逐步增强编码器在边缘感知注意力。

通过最小化每个边缘地图和真实源图像之间的损失来优化参数,损失函数可以表示为

$$L_g = L_g^E + \lambda_1 L_g^{\text{side}} = l(E_g, Y) + \lambda_1 \sum_{k=1}^8 l(S_g^k, Y) \quad (5)$$

式中: l 是损失函数; Y 是真实源图像; S_g^k 是边缘地图; λ_1 是权衡参数。

在第一阶段之后,固定第一阶段的参数,然后进入第二阶段。与一阶段训练相似,对从局部BiMLA解码器提取的中间特征执行反卷积操作,生成侧输出 $S_r^1, S_r^2, \dots, S_r^8$ 。损失函数定义如下(类似阶段一),

$$L_r = L_r^E + \lambda_2 L_r^{\text{side}} = l(E_r, Y) + \lambda_2 \sum_{k=1}^8 l(S_r^k, Y) \quad (6)$$

式中: L_r^E 和 L_r^{side} 是 E_r 和侧输出的损失;设置 λ_2 为0.4。

2 实验与结果分析

本文针对边缘检测图像,采用了经典的Canny^[29](手动调节阈值)、DexiNed^[18]、LDC^[21]、MangaLine^[22]、Informative Drawing^[23]算法为基线,并利用ODS(overall dataset scale)、OIS(optimal dataset scale)和AP(average precision)作为评估边缘检测算法性能的标准。其中,ODS反映了整体性能水平,得分越高越好。OIS衡量在最优阈值下的边缘检测结果的性能。AP代表算法在计算精度-召回率曲线并计算面积时,用于确定算法在多组数据集下的平均性能表现和稳定性。

2.1 实验设置

训练选用Adam^[30]优化器,参数 β_1 为0.5, β_2 为0.99。初始学习率为0.0002,训练总阶段数为2,批量设置为16。此外,所处理的图像最大尺寸为 512×512 ,裁剪尺寸统一为 256×256 。所使用的编码器补丁采用 16×16 像素、 8×8 像素和 4×4 像素的无重叠移动窗口,在处理中间卷积特征时,引入DropOut层且参数设置为0.2,以优化模型并解决训练过程中的过拟合问题,提高模型的泛化性能。模型在NVIDIA 3090显卡下进行训练。

2.2 数据集

PASCAL VOC作为一种应用于视觉模型分类、检测、分割等算法设计的数据集,被广泛应用于挑战赛中,如ImageNet Large Scale Visual Recognition Challenge等,旨在对视觉算法的性能进行全面评估和提升。包含10103个原图与批量二值化得到的边缘检测图。

NYUDv2数据集作为广泛应用于深度估计和图像分割等视觉任务的标准化数据集,包含家庭、办公室和实验室等多种复杂的室内环境。其中包括3397对原图与线图结果及871对测试数据。

古代服饰数据集选取了一部分由个人收集的古代服饰线图,解决复杂纹理线图提取问题。

在3个数据集中分别使用100张用于验证,100张用于测试,而剩余的300张用于训练。这些图像均经专业标注,未经处理,可信度高。

2.3 定性实验结果与分析

为保证实验结果的准确性与可靠性,本文选

用了 DexiNed、LDC、MangaLine 以及 Informative Drawing 等模型所提供的官方代码,并用其对应的代码环境实验。其中,DexiNed、LDC 和 Informative Drawing 使用源域和线图结果一一对应的数据集训练,而 MangaLine 仅使用了 BSDS500 数据集的图像(共 500 张,来源于自然场景,其中 200 张用于训练,100 张用于验证)。Informative Drawing 和本文方法一致,使用源域和线图域对齐的数据集进行训练。

图 2 是本文方法在 PASCAL 与 NYUDv2 数据集与基线模型的定性对比结果。图 3 是在古代服饰数据集上与基线模型的定性对比结果,包括源图像和提取的线图结果图像。Canny 代表使用手动设置阈值的处理结果,DexiNed、LDC 分别是密集网络、轻量网络提取出的线图,MangaLine 代表简单卷积(U-Net 简化块数、无跳跃连接)的提取结果,Informative Drawing 是使用 GAN 提取出的线图。可以发现,本文方法在整体性能上优于 DexiNed、LDC、MangaLine 和 Canny 基线方法,部分超越 Informative Drawing(其使用完整数据集训练)。

Canny 算子在古代服饰数据集中效果并不理

想,存在轮廓不连贯、清晰度不足等问题(见图 3)。DexiNed 难以完全捕捉到细微的边缘细节。可能无法检测到所有的内部边缘或者小细节。LDC 是一种轻量级的计算架构,采用更小的滤波器和更紧凑的模块,限制了模型学习复杂特征的能力。在房间局部细节边缘检测结果模糊。MangaLine 无法完全捕捉到细节丰富的古代服饰的边缘细节(见图 3)。Informative Drawing 引入几何损失从线条图像特征中预测深度信息,无法精确定位边缘,在 PASCAL 数据集存在边缘模糊问题。

而为了突出显示线图结果,由于源图像较大,将本文的对比方法单独列出(见图 3 中箭头所指示图像)。可以发现上述方法在 PASCAL 和 NYUDv2 效果显著,但在线稿不连续的古代服饰数据集中难以捕获图像细节纹理信息。可以看出本文模型方法提取的轮廓线不仅清晰,而且将古代服饰的针线导致的不连续性突出显示。不同于 DexiNed、Informative Drawing 轮廓线较粗的问题,本文解决了针对图像(高分辨率图像)轮廓线提取不清晰和线稿不连续的问题。

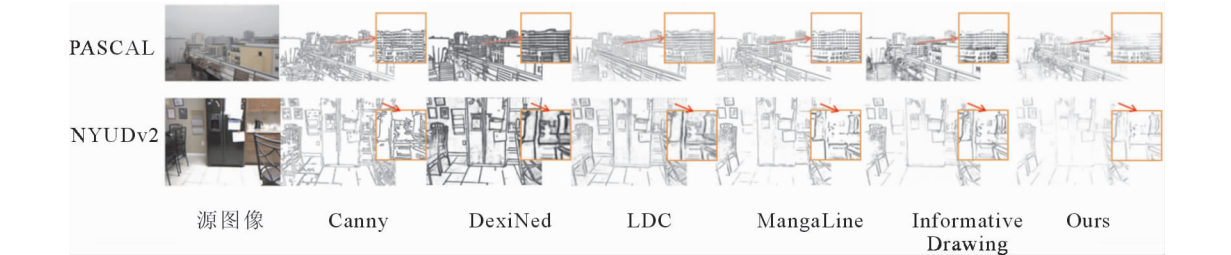


图 2 与基线模型在 PASCAL 与 NYUDv2 数据集的定性对比图

Fig. 2 Qualitative comparison with the baseline model on the PASCAL and NYUDv2 datasets

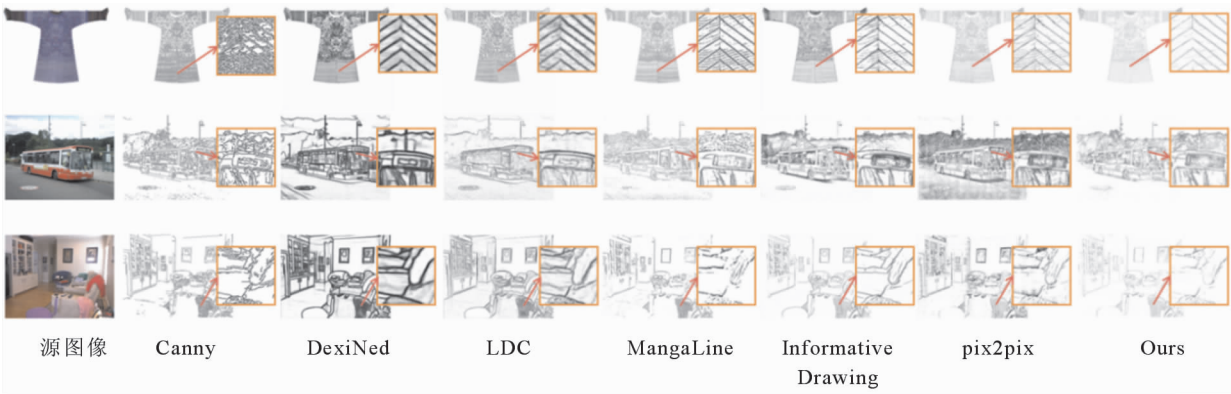


图 3 与基线模型在古代服饰数据集的定性对比图

Fig. 3 Qualitative comparison with the baseline model in the ancient clothing dataset

2.4 定量实验结果与分析

定量实验结果如表 1 所示,基于 ODS、OIS 和 AP 图像边缘检测评价指标,对 Canny、DexiNed、LDC、MangaLine、Informative Drawing 以及本节所提出方法在 PASCAL VOC 和 NYUDv2 数据集上的结果进行了详细的比较和分析。

可以发现,相较于 DexiNed、LDC、MangaLine 等使用完整数据集进行训练的 CNN 方法,本文使用的方法在图像边缘检测任务上具备更优秀的评价指标表现,这是因为 CNN 方法对训练数据量的要求较高,导致其泛化性方面不如使用相同数据量进行训练的 Transformer 方法。因此,本文提出的方法所生成的线图图像在质量和泛性上都有较大的提升。本文方法在 ODS 和 OIS 评分上表现良好,特别是在 PASCAL VOC 数据集上,本文模型的 ODS 得分比第二名高出 1.6%,比最低名次高出 35.7%。在 NYUDv2 数据集上,ODS 得分比第二名高出约 0.5%,同时训练时间缩短了 15 倍,比最低名次高出 38.8%。这些数据表明,本文方法在学习源图像和目标图像分布特征方面非常

成功,生成的线图质量高、完整且边缘轮廓清晰。

相比于需要大量训练数据的 CNN 方法,本文使用的 Transformer 方法在泛化性方面表现更佳。此外,与 GAN 方法相比,本文方法在训练次数和综合性能上也显示出竞争力。这些优势使得本文提出的方法在实际应用中具有潜在的价值,尤其是在需要快速、高效边缘检测的场景中。

本文方法在 PASCAL VOC 和 NYUDv2 数据集上的综合效果虽然未超过 Informative Drawing,但在模型体量和耗时方面展现了显著的优势。具体而言:Canny 方法因无法有效处理复杂纹理而表现较差;DexiNed 因多尺度特性设计在一定程度上提升了准确性;LDC 在复杂背景中鲁棒性不足,导致性能下降;MangaLine 更适合漫画风格图像,与服饰纹理差异较大;Informative Drawing 在某些指标上表现优异,但特征提取策略需要优化。而本文的方法基于 Transformer 的两阶段策略,实现了更轻量级和高效的边缘检测模型。这不仅降低了计算资源的消耗,还提高了实际应用中的处理速度。

表 1 实验结果质量评估表
Tab. 1 Quality assessment form for experimental results

Method	PASCAL VOC			NYUDv2		
	ODS	OIS	AP	ODS	OIS	AP
Canny	0.610	0.630	0.575	0.600	0.650	0.560
DexiNed	0.729	0.745	0.583	0.719	0.735	0.580
LDC	0.788	0.808	0.840	0.813	0.813	0.866
MangaLine	0.753	0.772	0.807	0.620	0.632	0.550
Informative Drawing	0.815	0.833	0.889	0.828	0.844	0.850
Ours	0.828	0.821	0.880	0.832	0.825	0.836

2.5 模型参数评估

在评估实验中,通常会采用默认配置来训练模型,并在结果中列出模型的参数总量以及在特定任务上的平均训练时间,来验证本文模型的效率。

图 4 列出 DexiNed、LDC、MangaLine、Informative Drawing 及本文模型的参数总量以及在图像线图提取任务上的平均训练时间。由图 4 可见,本文提出的模型在图像边缘检测方面表现优于传统的基于卷积神经网络(CNN)和生成对抗网络(GAN)方法。此外,本文模型的参数量仅仅包含 2.5 M,显著少于一些表现优秀的 CNN 和常见的

GAN 模型,这可能意味着模型更加高效,需要更少的计算资源。

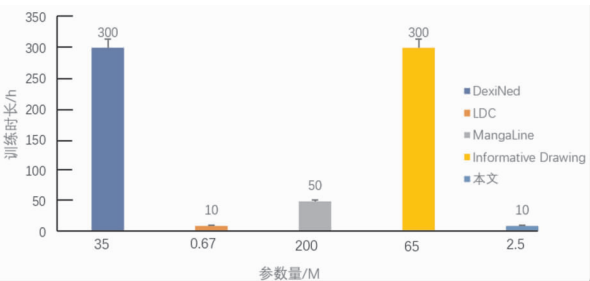


图 4 模型参数量及训练时长对比图
Fig. 4 Comparison of model parameters and training time

此外,相比于 GAN 方法,本文模型在训练时间上也有所缩短,这表明本文方法在提高模型收敛速度方面也是有效的。减少模型的参数量和训练时间对于实际应用来说是非常重要的,因为这可以使模型更快地部署到实际环境中,同时也减少了计算资源的消耗,参数量较少的模型通常需要更少的内存和存储空间,这使得它们更适合在内存受限的设备上运行,例如移动设备和嵌入式系统等。

2.6 消融实验

为验证本文方法中不同模块的有效性,基于 PASCAL VOC、NYUDv2 数据集设计了 5 个消融实验。(I) 改变 BiMLA 解码器卷积块为残差块;

(II) 去除 Transformer 中的层归一化;(III) 移除位置编码;(IV) 恢复传统窗口操作;(V) 移除轻量特征融合模块;(VI) 完整模型。

定性结果如图 5 所示。当(I)改用残差块时增加模型复杂度,导致特征利用不当,预测图像出现噪点和线条不连续;(II)删除 Transformer 的层归一化,编码器无法有效传递图像特征,训练时间增加;(III)去除位置编码,全局特征信息利用不佳,结果受输入图像分辨率影响;(IV)移除移动窗口机制,模型泛化性下降;(V)去除轻量级特征融合模块后,生成结果仍然存在噪声,例如生成结果的中心部分。这些消融实验揭示了各组件对模型性能的影响程度。

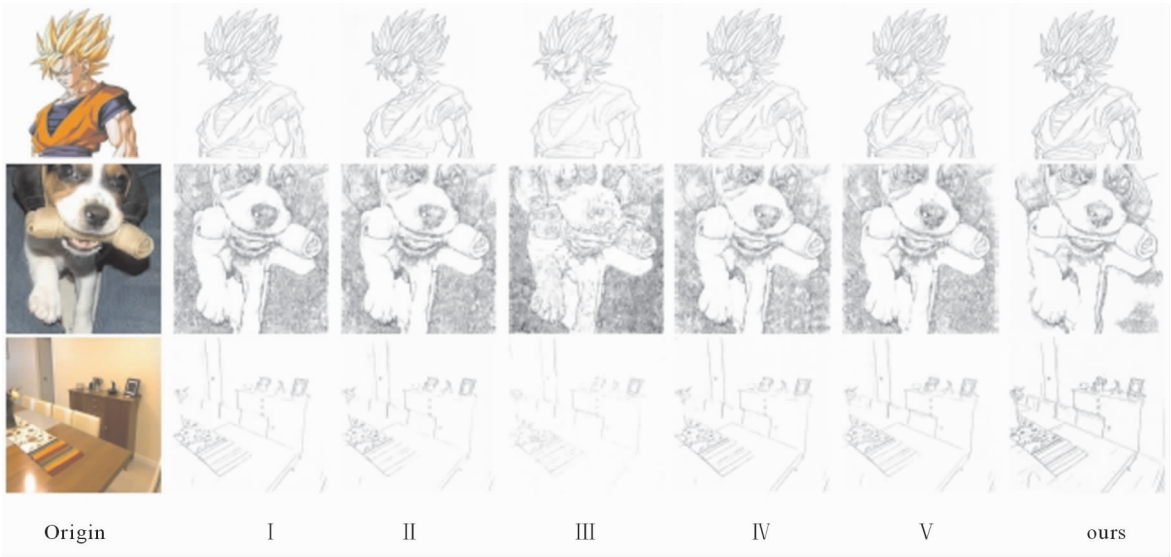


图 5 消融实验结果比较图
Fig. 5 Comparison of ablation experimental results

定量结果如表 2 所示,采用 3 个评价指标 ODS、OIS、AP 计算了本模型的不同消融实验的定量结果。值得注意的是,所提出的完整模型在 3

项图像处理质量和性能最佳的评价指标上均达到了最优,这为本节提供的边缘检测方法的有效性提供了实证依据。

表 2 消融实验结果质量评估表
Tab. 2 Ablation experiment result quality assessment form

Ablation	PASCAL VOC			NYUDv2		
	ODS	OIS	AP	ODS	OIS	AP
I	0. 810	0. 820	0. 860	0. 815	0. 821	0. 816
II	0. 815	0. 817	0. 830	0. 812	0. 815	0. 806
III	0. 820	0. 818	0. 875	0. 822	0. 810	0. 817
IV	0. 823	0. 809	0. 860	0. 819	0. 815	0. 817
V	0. 821	0. 795	0. 788	0. 790	0. 807	0. 803
Ours	0. 828	0. 821	0. 880	0. 832	0. 825	0. 836

3 结语

针对目前方法难以提取纹理复杂的中国古代服饰线图等问题,提出了一种基于 Transformer 两阶段策略的边缘检测模型。采用较大的补丁窗口对图像进行划分,全局编码器处理线性化操作后的划分窗口,之后传入全局 BiMLA 解码器。第一阶段,应用双向路径(包括自顶向下和自底向上)对特征信息进行处理,用来学习图像的全局语义信息。第二阶段,采用更细致的补丁窗口划分图像,嵌入位置编码,有效解决了线条模糊的问题,从而提高了预测图像的质量。最后,引入了轻量级特征融合模块,通过采用端到端的训练方式,加快了网络收敛速度。

实验结果表明,本文方法能够生成具有轮廓清晰、线条连贯和噪声最少的高质量图像,且具有强泛化性。在对 PASCAL VOC、NYUDv2 数据集进行图像预测时,均显著优于常规的 CNN 方法,且可接近于参数量较为庞大的 GAN 方法,其中 ODS 平均提高了 7.3%,OIS 平均提高了 4.1%,AP 平均提高了 12.8%。此外,可以用于提取纹理线条等复杂的古代服饰的线图,对保护和传承中华优秀传统文化至关重要。

该方法可以加快线图提取,但针对服饰图像如何选择最优的补丁窗口仍待进一步研究,因此,在实际应用中提取效果并不理想,无法达到实时性需求。为实现真正的实时图像边缘检测,未来的研究可以探索更高效的算法和网络结构(如采取动态特征融合策略)以提高训练和推理的效率,同时也需要考虑如何更好地融合图像多尺度语义特征,从而实现更准确和可控的边缘检测。

参考文献

- [1] SUN R, LEI T, CHEN Q, et al. Survey of image edge detection[J]. *Frontiers in Signal Processing*, 2022, 2: 826967.
- [2] SIVAPRIYA M S, SURESH S. ViT-DexiNet: A vision transformer-based edge detection operator for small object detection in SAR images[J]. *International Journal of Remote Sensing*, 2023, 44(22): 7057-7084.
- [3] AKBARI SEKEHRAVANI E, BABULAK E, MASOODI M. Implementing canny edge detection algorithm for noisy image[J]. *Bulletin of Electrical Engineering and Informatics*, 2020, 9(4): 1404-1410.
- [4] MCILHAGGA W. The canny edge detector revisited[J]. *International Journal of Computer Vision*, 2011, 91(3): 251-261.
- [5] LI Y B, LIU B L. Improved edge detection algorithm for canny operator[C]//2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference. Chongqing, China: IEEE, 2022: 1-5.
- [6] UY J N, VILLAVERDE J F. A durian variety identifier using canny edge and CNN[C]//2021 IEEE 7th International Conference on Control Science and Systems Engineering. Qingdao, China: IEEE, 2021: 293-297.
- [7] OJASHWINI R N, GANGADHAR REDDY R, RANI R N, et al. Edge detection canny algorithm using adaptive threshold technique[C]//Intelligent Data Engineering and Analytics. Singapore: Springer, 2020: 469-477.
- [8] 肖扬,周军. 图像边缘检测综述[J]. *计算机工程与应用*, 2023, 59(5): 40-54.
XIAO Y, ZHOU J. Overview of image edge detection[J]. *Computer Engineering and Applications*, 2023, 59(5): 40-54.
- [9] YE Y F, YI R J, GAO Z R, et al. Delving into crispness: Guided label refinement for crisp edge detection[J]. *IEEE Transactions on Image Processing*, 2023, 32: 4199-4211.
- [10] PU M Y, HUANG Y P, LIU Y M, et al. EDTER: Edge detection with transformer[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 1392-1402.
- [11] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//2021 International Conference on Machine Learning. ICML, 2021: 8748-8763.
- [12] ARBELÁEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(5): 898-916.
- [13] LIU Z Y, TAN Y C, HE Q, et al. SwinNet: Swin transformer drives edge-aware RGB-D and RGB-T salient object detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4486-4497.
- [14] XIA L G, CHEN J, LUO J C, et al. Building change detection based on an edge-guided convolutional neural network combined with a transformer[J]. *Remote*

- Sensing, 2022, 14(18): 4524.
- [15] XU S C, CHEN X X, ZHENG Y H, et al. ECT: Fine-grained edge detection with learned cause tokens [J]. Image and Vision Computing, 2024, 143: 104947.
- [16] MIAO, L, TAKESHI O, RYOICHI I. SWIN-RIND: Edge detection for reflectance, illumination, normal and depth discontinuity with Swin Transformer [C] // The 34th British Machine Vision Conference. Aberdeen, UK: BMVA Press, 2023: 1-10.
- [17] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C] // 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021: 9992-10002.
- [18] SORIA X, RIBA E, SAPPA A. Dense extreme inception network: Towards a robust CNN model for edge detection [C] // 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass Village, USA: IEEE, 2020: 1923-1932.
- [19] HAN K, WANG Y H, CHEN H T, et al. A survey on vision transformer [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 87-110.
- [20] SUBAKAN C, RAVANELLI M, CORNELL S, et al. Attention is all you need in speech separation [C] // 2021 IEEE International Conference on Acoustics. Toronto, Canada: IEEE, 2021: 21-25.
- [21] SORIA X, POMBOZA-JUNEZ G, SAPPA A D. LDC: Lightweight dense CNN for edge detection [J]. IEEE Access, 2022, 10: 68281-68290.
- [22] LI C Z, LIU X T, WONG T T. Deep extraction of manga structural lines [J]. ACM Transactions on Graphics, 2017, 36(4): 1-12.
- [23] CHAN C, DURAND F, ISOLA P. Learning to generate line drawings that convey geometry and semantics [C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 7905-7915.
- [24] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [25] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 3431-3440.
- [26] YIN Z Y, WANG Z S, FAN C, et al. Edge detection via fusion difference convolution [J]. Sensors, 2023, 23(15): 6883.
- [27] LE M, KAYAL S. Revisiting edge detection in convolutional neural networks [C] // 2021 International Joint Conference on Neural Network. Shenzhen, China: IEEE, 2021: 1-9.
- [28] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 7132-7141.
- [29] XIE S N, TU Z W. Holistically-nested edge detection [C] // 2015 IEEE International Conference on Computer Vision. Santiago, USA: IEEE, 2015: 1395-1403.
- [30] KINGMA D P, JIMMY B Adam: A method for stochastic optimization [J]. International Conference on Learning Representations, 2014: 6628106.

(编辑 李波)