

Transformer 多模态特征提取下图像目标边缘检测方法

刘庆霞¹ 司贺杰¹ 王俊方¹ 徐恕贞¹
LIU Qingxia SI Hejie WANG Junfang XU Shuzhen

摘要

在对图像目标边缘检测时,复杂多变的光照条件会导致图像局部区域出现明暗不均、过曝或欠曝等现象。基于局部梯度或单一模态特征的方法由于缺乏对图像全局语义的理解,易受噪声影响,造成边缘模糊或断裂,导致检测精度较低。为解决这一问题,文章提出了一种基于 Transformer 多模态特征提取的图像目标边缘检测方法。首先,将采集的彩色图像通过灰度转换降低色彩干扰并采用高斯平滑抑制噪声,利用局部 Otsu 算法自适应计算最优分割阈值实现目标与背景的精确分离;其次,基于卷积神经网络(convolutional neural network, CNN)提取分割目标的多模态特征,结合 Transformer 的自注意力机制(Self-Attention)进行特征加权融合以增强对图像全局语义的理解,突出边缘特征,解决边缘模糊或断裂问题;最后,采用深度学习算法通过构建目标区域边缘概率图并优化损失函数实现边缘检测,再经掩码处理去除干扰得到最终边缘检测结果。实验结果表明,所提方法在实际应用中能够精准检测出目标区域的边缘,且边缘结构完整度数值为 0.945,检测精度较高。

关键词

Transformer; 多模态特征; 图像; 边缘检测; 目标边缘; 特征增强; 注意力机制

doi: 10.3969/j.issn.1672-9528.2026.03.019

0 引言

在复杂场景中,图像目标边缘检测能够精准提取物体的边缘轮廓,是图像分割、目标识别等多个领域的关键技术。在实际应用中,现有的检测方法会受到光照变化和纹理干扰的影响,且仅能捕捉图像的局部像素强度变化,缺乏对图像目标的语义理解,再通过局部梯度值识别边缘,导致方法的边缘检测精度不高。

在该研究背景下,针对目标检测领域,多位学者提出了多种图像目标边缘检测方法。张喜民等^[1]在双三次插值作用

下对目标图像进行映射,通过邻域选取、权重计算等合成插值,再利用 Zernike 矩在单位圆内展开图像,提取旋转不变特征,求解边缘参数、估计边缘方向、计算亚像素偏移值以实现边缘检测。然而,双三次插值主要基于局部邻域像素计算,在复杂光照下,局部区域明暗不均、过曝或欠曝会使邻域像素值异常,这些异常值参与插值计算会导致结果偏差,影响后续边缘检测准确性。王静静等^[2]先通过自适应中值滤波保留图像边缘结构信息,再对滤波后图像灰度化,计算前景与背景区域的类间方差确定最优阈值,接着用非极大值抑制方法沿梯度方向提取连续边缘,最后经形态学处理得到检测结果。但在复杂光照和噪声共同作用下,该方法检测出的

1. 郑州科技学院信息工程学院 河南郑州 450064

[15] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137 - 1149.

[16] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[EB/OL]. (2017-04-19)[2026-03-16]. <https://doi.org/10.48550/arXiv.1612.03144>.

电子信息科学与技术, email: 2799325027@qq.com。

蒋力畅(2004—),男,江苏徐州人,本科,研究方向:电子信息科学与技术, email: 3388718883@qq.com。

沈相同(2004—),女,山西运城人,本科,研究方向:电子信息科学与技术, email: 2969585715@qq.com。

孙敏(1981—),通信作者(email:sunmin@szcu.edu.cn),女,安徽滁州人,博士,副教授,研究方向:数字图像处理。

【作者简介】

王硕(2002—),男,江苏常州人,本科,研究方向:

(收稿日期: 2025-08-19 修回日期: 2026-03-17)

边缘可能不连续、不清晰甚至出现错误分割,且因缺乏全局语义理解,对于背景灰度相近的目标边缘,可能因阈值选择不当而无法准确检测。黄雪峰等^[3]采用 OpenMV4H7PLUS 模块实时获取微小机械零件图像,灰度化处理后用双边滤波消除噪声,再应用 Canny 算法计算图像梯度幅值和方向以提取边缘,最后经形态学处理实现边缘检测。然而, Canny 算法依赖图像梯度信息检测边缘,噪声会使梯度计算出现偏差,在复杂光照下,噪声与光照变化相互影响,导致梯度幅值和方向计算不准确,进而提取出错误边缘点。

本文设计了基于 Transformer 多模态特征提取的图像目标边缘检测方法,可提高目标边缘检测的准确性,增强对图像边缘的感知能力,还能拓展该方法在工业检测、自动驾驶等多个领域的应用,促进这些领域的创新发展与智能化转型。

1 图像目标边缘检测方法设计

1.1 基于灰度转换与局部 Otsu 算法的图像目标分割

在图像目标边缘检测中,原始彩色图像丰富的色彩信息会引入干扰,影响边缘检测的准确性与效率。灰度图像仅包含亮度信息,能简化数据、突出明暗变化特征,更利于边缘检测。并且,图像采集时可能受噪声干扰而质量下降,影响边缘检测效果。局部 Otsu 算法可依据图像局部特性自适应确定分割阈值,有效处理噪声与复杂背景,提高目标分割准确性。因此,本研究采用灰度转换与局部 Otsu 算法进行图像目标分割,通过将图像中的目标区域从背景中准确分离,为后续精确边缘检测提供清晰明确的目标轮廓基础。

基于收集到的图像,先对图像进行色彩转换,将其从原本的彩色图像转换成灰度图像,对转换后的图像进行平滑处理,其处理过程为:

$$L(x, y) = \sum_{i=1}^n \sum_{j=1}^m L_0(x+i, y+j) \frac{1}{2\pi a^2} e^{-\frac{i^2+j^2}{2a^2}} \quad (1)$$

式中: $L(x, y)$ 表示图像平滑处理结果; $L_0(x+i, y+j)$ 表示原始图像平移结果; a 表示原始图像的像素标准差; i, j 分别表示高斯坐标点; n 表示图像的宽度; m 表示图像的长度。

基于图像平滑处理结果,应用局部 Otsu 分割算法,将图像中的目标区域分割出来^[4]。在分割时,根据平滑处理后的灰度化图像,随机选择一个灰度值作为分割阈值 F ,将图像划分为目标区域和背景区域,再计算目标区域和背景区域的类间方差^[5]:

$$\sigma_t^2 = \eta_0 \eta_1 (\mu_0 - \mu_1)^2 \quad (2)$$

式中: σ_t 表示图像初始分割后目标区域与背景区域的类间方差; η_0 表示目标区域的像素占比; η_1 表示背景区域的像素占比; μ_0 表示目标区域的灰度均值; μ_1 表示背景区域的灰度均值。

根据上述公式,通过不断更换分割阈值 F ,计算对应的类间方差,选择类间方差最大值所对应的分割阈值作为最优分割阈值,由此将图像中的目标区域分割出来,用 M 表示。

1.2 基于 CNN 与 Transformer 的目标区域边缘特征增强

在完成图像目标分割后,虽然已经将目标区域从背景中分离出来,但此时目标区域的边缘特征可能并不突出。一方面,图像采集过程中不可避免会引入噪声,即便经过前面的处理,残留噪声仍可能干扰边缘特征,使得边缘模糊不清;另一方面,目标区域本身可能存在复杂的纹理、形状变化等情况,导致边缘特征不够明显,难以满足高精度边缘检测需求。因此,需要进一步对目标区域边缘特征进行增强。

基于分割的图像目标区域,应用 CNN 算法,提取目标区域的多模态特征,得到目标区域的特征图^[6]。其具体计算过程为:

$$D_k = p_d \times \varphi \left(R \left(\sum_{k=1}^K \gamma_k \times M + l_k \right) \right) + l_s \quad (3)$$

式中: D_k 表示基于 CNN 算法的目标区域多模态特征; γ_k 表示第 k 层卷积层的卷积函数; l_k 表示第 k 层卷积层的卷积参数; K 表示 CNN 算法中卷积层的数量; $R()$ 表示激活函数; $\varphi()$ 表示池化函数; p_d 表示输出层的权重矩阵; l_s 表示输出层的常数值。

将上述提取的多模态特征作为基础,在 Transformer 的作用下,引入自注意力机制,确定多模态特征对应的权重值,对多模态特征进行加权处理,实现对目标区域边缘特征的增强处理^[7]。注意力权重值的计算公式为:

$$Q_k = S \left(\frac{P_k V_k}{\sqrt{w_k}} \right) Z_k \quad (4)$$

式中: Q_k 表示图像目标区域中第 k 个模态特征的注意力权重值; P_k 表示第 k 个模态特征的查询向量; V_k 表示第 k 个模态特征的键向量; w_k 表示第 k 个模态特征的键向量维度; Z_k 表示第 k 个模态特征的值向量; $S()$ 表示注意力函数。

根据计算的多模态特征对应的注意力权重值,将多模态特征进行融合,实现对目标区域边缘特征的增强。其具体过程为:

$$D_z = \sum_{k=1}^K Q_k D_k \quad (5)$$

式中: D_z 表示目标区域边缘特征增强结果。

通过上述公式,在 Transformer 的作用下,实现对多模态特征的提取及边缘特征的增强,便于后续实现目标区域边缘的检测。

1.3 基于深度学习算法的图像目标区域边缘检测

在对图像目标区域边缘检测时,根据边缘特征的增强结

果 D_z ，基于深度学习算法，构建目标区域边缘概率图，其具体计算过程为：

$$S_b = \frac{1}{1 + e^{-w_b D_z + y_b}} \quad (6)$$

式中： D_z 表示目标区域边缘概率图； w_b 表示目标区域边缘概率预测参数； y_b 表示目标边缘点的概率误差。

根据构建的目标区域边缘概率图，对目标区域的边缘区域和非边缘区域进行分类。分类时，需设定损失函数，通过多次迭代，得到更为精准的目标区域分类结果^[8]。其具体计算过程为：

$$L_s = \varphi_s S_z + (1 - \varphi_s) S_c \quad (7)$$

$$S_l = D_z + \lambda L_s P_k \quad (8)$$

式中： L_s 表示设定的损失函数； S_z 表示边缘概率预测时的交叉熵损失函数； S_c 表示边缘检测的损失函数； φ_s 表示损失函数加权参数； S_l 表示图像目标区域的边缘结果； λ 表示预测边缘点映射参数； P_k 表示边缘点判断函数。

通过上述公式，得到边缘检测的初始结果，再对其进行掩码处理，去除边缘检测结果中的干扰因素。其掩码处理结果为：

$$Y = S_l \times M_l \quad (9)$$

式中： Y 表示掩码处理后的目标区域边缘检测结果； M_l 表示掩码参数。

基于上述过程，得到目标区域边缘检测结果，实现对图像边缘的检测。至此，Transformer 多模态特征提取下图像目标边缘检测方法的设计完成。

2 实验测试

2.1 实验图像集

本次实验从 BSDS500 图像集、NuScenes 图像集、COCO 2017 图像集等多个公开图像集随机选择实验图像，选择的实验图像类型为自然风景、动物、植物、建筑等 23 种类型，共选择图像数量为 15 256 张，构建实验图像集，用于本次实验。实验图像集的示例如图 1 所示。

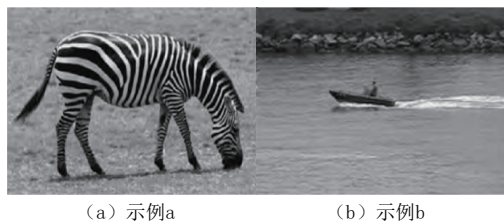


图 1 实验图像集示例

如图 1 所示，将随机选择的图像进行灰度化处理，减少后续设计中的计算量。同时，统一实验图像集中图像的尺寸，

将其固定为 1 600 px×900 px，便于后续对图像进行规范化处理。

2.2 图像目标区域边缘检测过程

应用本文方法对示例图像进行处理，设定本文方法的多个参数，其参数设定如下：目标区域边缘概率预测参数 w_b 为 1.45；加权参数 φ_s 为 0.55；边缘点映射参数 λ 为 1.66；边缘点概率预测误差 y_b 为 0.02。

基于上述设定的参数对实验图像集进行平滑处理，采用局部 Otsu 算法通过迭代计算类间方差确定最优分割阈值，实现目标区域的精准分割。随后基于 CNN 提取目标区域的多模态特征，并利用 Transformer 的自注意力机制计算特征权重，实现特征融合与增强。应用 Transformer 算法的具体参数如表 1 所示。

表 1 Transformer 的具体参数

序号	Transformer 参数	参数数值
1	注意力头数	15
2	注意力头维度	64
3	总参数量	2 362 368
4	输入变量维度	500
5	最大序列长度	512
6	隐藏层维度	3 072
7	编码层偏置系数	2.00
8	嵌入层维度	200

如表 1 所示，基于 Transformer 的参数，实现对图像目标区域边缘特征的增强。

基于此，将其输入深度学习模型中，计算目标区域边缘点的概率，通过设定损失函数，得到提取的边缘检测结果，再对其进行掩码处理，提高边缘检测结果的质量。

在验证本文方法的性能时，选择基于双三次插值与 Zernike 矩的图像亚像素边缘检测算法作为参照组 1，基于机器视觉模块 OpenMV 的微小机械零件图像边缘检测作为参照组 2，将这两种方法与本文设计的方法进行对比，验证本文方法的检测精度。

2.3 检测结果分析与讨论

应用三种方法图 1 中的示例 (a) 进行目标边缘检测，给出对应的检测结果，进而对比三种方法的检测精度。三种方法的检测结果如图 2 所示。本文方法能够精准检测出目标区域的边缘轮廓，仅存在较小的轮廓边缘不清晰现象。参照组 1 未能有效检测出目标区域的边缘轮廓，参照组 2 未能准确分割目标区域，导致检测结果存在大量伪影。本文方法的优势在于引入了 Transformer 机制，通过对目标区域边缘轮廓进行特征增强，实现了更清晰的边缘检测效果。实验结果表明，该方法在实际

应用中具有较高的检测精度。

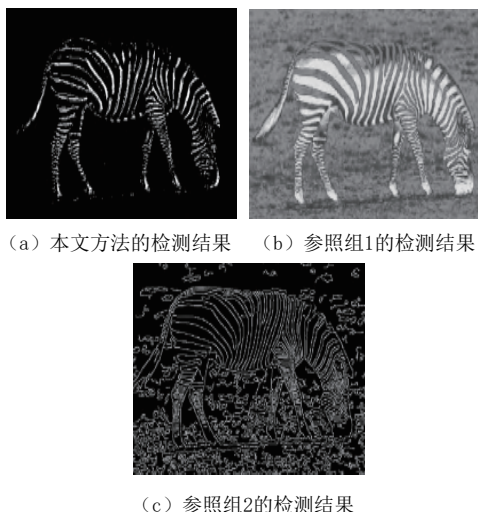


图2 三种方法的检测结果

在进一步验证本文方法的检测精度时，以检测到的边缘结构完整度为评价指标，能够衡量检测到的目标边缘的完整性，数值越高，方法的检测精度越高。

三种方法的边缘结构完整度如图3所示。在多次实验中，本文方法的检测的目标边缘结构完整度数值较高，其均值为0.945，参照组两种方法的边缘结构完整度数值较低，其中，参照组1的均值为0.658，参照组2的均值为0.591。这是因为本文方法在设计时引入Transformer算法，其通过计算图像目标区域的多模态特征注意力权重值，实现对目标边缘特征的增强，进而得到更为精准边缘检测结果。由此可见，本文设计的方法在实际应用中检测精度更高。

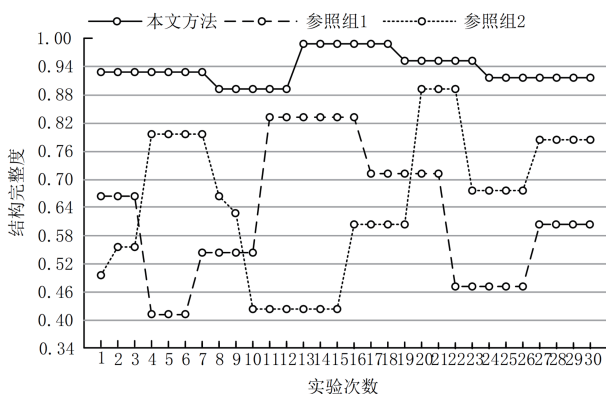


图3 三种方法的边缘结构完整度对比分析

3 结语

本文提出了一种基于Transformer多模态特征增强的图像目标边缘检测方法，通过灰度转换、局部Otsu分割、CNN特征提取和Transformer特征融合的协同处理，在BSDS500、NuScenes和COCO 2017等标准数据集上实现了显著的性能提升。实验结果表明，本方法在边缘结构完

整度指标上达到0.945，较传统基于Zernike矩(0.658)和OpenMV(0.591)的方法具有明显优势。通过局部Otsu算法的自适应阈值分割和CNN-Transformer的多模态特征融合，实现了对目标边缘的精准检测。本研究为工业检测和自动驾驶等领域的图像处理任务提供了可靠的技术方案，在工业质检中具有重要应用价值。

参考文献:

- [1] 张喜民, 詹海生. 基于双三次插值与Zernike矩的图像亚像素边缘检测算法[J]. 计量学报, 2025, 46(7): 989-995.
- [2] 王静静, 王洪君. 一种基于OTSU自适应阈值分割的含噪图像边缘检测算法[J]. 电子制作, 2025, 33(11): 42-45.
- [3] 黄雪峰, 康哲宇, 赵永华, 等. 基于机器视觉模块OpenMV的微小机械零件图像边缘检测[J]. 工业控制计算机, 2025, 38(5): 60-61.
- [4] 金义舒, 黄平, 郑福印, 等. 自适应阈值LOG与Canny算法结合的图像边缘检测研究[J]. 通信与信息技术, 2025(3): 122-124.
- [5] 印泽昊, 宁航, 高济辉, 等. 一种基于形态学的侧扫声呐图像边缘检测算法[J]. 科学技术创新, 2025(12): 71-74.
- [6] 唐泉, 王硕. 基于阈值分割和数学形态学的噪声图像边缘检测方法[J]. 山东师范大学学报(自然科学版), 2025, 40(1): 64-70.
- [7] 张喜民, 詹海生. 基于双三次插值的Canny-Devermay亚像素图像边缘检测算法[J]. 现代制造工程, 2025(3): 107-114.
- [8] 王军, 林宇航, 贾玉彤, 等. 一种改进Canny算子的图像边缘检测算法[J]. 小型微型计算机系统, 2024, 45(6): 1413-1417.

【作者简介】

刘庆霞(1998—), 女, 河南安阳人, 硕士, 助教, 研究方向: 计算机应用技术、数据分析、图像处理。

司贺杰(1993—), 女, 河南驻马店人, 硕士, 助教, 研究方向: 计算机应用技术、教学教改、知识图谱。

王俊方(1998—), 女, 河南南阳人, 硕士, 助教, 研究方向: 计算机应用技术、图像处理。

徐怒贞(1990—), 女, 河南南阳人, 硕士, 助教, 研究方向: 数据挖掘。

(收稿日期: 2025-08-22 修回日期: 2026-03-18)