

本文引用格式: 郭瑞桁,郭瑞洲,刘夏丽.改进 SAM 的皮革缺陷分割与自适应裁切方法[J].自动化与信息工程,2026,47(4):55-62.

GUO Ruiheng, GUO Ruizhou, LIU Xiali. An improved SAM for leather defect segmentation and adaptive cropping [J]. Automation & Information Engineering, 2026,47(4):55-62.

改进 SAM 的皮革缺陷分割与自适应裁切方法

郭瑞桁¹ 郭瑞洲¹ 刘夏丽²

(1.广东瑞洲科技有限公司, 广东 佛山 528200

2.广东工业大学, 广东 广州 510006)

摘要: 针对皮革表面缺陷区域定位及柔性裁片形变引起的 CAD 路径偏移问题, 提出一种改进 SAM 的皮革缺陷分割与自适应裁切方法。该方法基于 SAM-Med2D ViT-B 图像编码器、通道-空间双分支适配器构建 Leather-SAM, 结合 ORB 特征匹配与全局仿射变换矩阵、多横梁裁切安全控制等, 构建基于 Leather-SAM 的缺陷识别与裁切优化算法。实验结果表明, 该方法的 IoU、Dice 和轮廓 F1 分别为 54.58%、65.22%和 88.08%, 推理速度为 12.2 f/s, 为皮革智能排版与裁切提供了可复现的技术流程。

关键词: 皮革缺陷; SAM-Med2D; 二值分割; 仿射校正; 自适应裁切

中图分类号: TP391.41

文献标志码: A

文章编号: 1674-2605(2026)04-0008-08

DOI: 10.12475/aie.20260408

开放获取

An Improved SAM for Leather Defect Segmentation and Adaptive Cropping

GUO Ruiheng¹ GUO Ruizhou¹ LIU Xiali²

(1.Guangdong Ruizhou Technology Co., Ltd., Foshan 528200, China

2.Guangdong University of Technology, Guangzhou 510006, China)

Abstract: To address the issues of leather surface defect localization and CAD path deviations caused by flexible cutting piece deformation, we propose an improved SAM-based method for leather defect segmentation and adaptive cropping. This method constructs Leather-SAM based on the SAM-Med2D ViT-B image encoder and a channel-spatial dual-branch adapter, and combines ORB feature matching, global affine transformation matrix, and multi-beam cutting safety control to build a defect recognition and cutting optimization algorithm based on Leather-SAM. Experimental results show that the method achieves IoU, Dice, and contour F1 scores of 54.58%, 65.22%, and 88.08%, respectively, with an inference speed of 12.2 f/s, providing a reproducible technical pipeline for intelligent leather layout and cutting.

Keywords: leather defect; SAM-Med2D; binary segmentation; affine correction; adaptive cropping

0 引言

天然皮革由动物原皮鞣制而成, 受生长环境、饲养条件、鞣制工艺等影响, 皮面可能存在生长纹、刮伤、破洞、色斑等天然瑕疵。这些瑕疵直接影响后续排版裁切以及成品品质, 因此皮革原料入厂阶段的缺陷识别与精准分级是管控成品良品率的首要环节。现

阶段国内大量中小型皮具加工厂仍采用人工质检(肉眼挑拣缺陷)、粉笔标记裁片轮廓、单刀固定路径裁切的传统生产模式。其中, 人工质检受视觉疲劳、个体经验差异、主观判断标准不统一等因素影响, 存在检测效率较低、结果一致性不足等问题。优质皮革原料因缺陷漏检而导致裁切报废损耗长期维持在较高水平。

采用传统数字图像处理算法（如边缘算子检测、固定阈值分割、小波多尺度变换、图像显著性提取等）进行皮革表面缺陷检测，需反复调试算法参数，泛化性能较差，无法适配多品类皮革混线生产工况^[1]。通过支持向量机、决策树等传统机器学习分类模型进行皮革表面缺陷种类判别存在以下问题：1）需人工提取方向梯度直方图、尺度不变特征变换等特征；2）缺少统一的行业测试基准，目前行业主流分级标准为鞋类及相关行业研究协会（shoe and allied trades research association, SATRA）规范^[2]；3）仅能实现有无缺陷的二元判定，缺少缺陷像素级定位能力，无法获取缺陷实际尺寸、位置数据，不能为后续裁切避让提供坐标信息^[3-5]。利用经典卷积神经网络（ResNet^[6]、VGG^[7]、DenseNet^[8]、AlexNet^[9]、SqueezeNet^[10]、ShuffleNet^[11]、GoogleNet^[12]等）进行皮革表面缺陷分类，需大量标注样本以完成训练，而占比较小的微小缺陷因样本不均衡，易导致模型偏向大类缺陷识别。近年来，以分割一切模型（segment anything model, SAM）为代表的通用视觉大模型，凭借少样本或零样本分割能力广泛应用于各类工业视觉检测场景，其基于视觉提示（点、框、掩码）的交互机制，为皮革多品类缺陷智能检测提供新的技术路径。但 SAM 基于通用自然图像预训练，直接应用于皮革表面缺陷识别时，易受皮革复杂天然纹理干扰，同时其编码器缺少针对性特征筛选模块，对浅痕、褪色等微小缺陷提取能力不足，且现有 SAM 改进方案大多聚焦于医学影像分割，缺少与后端裁切控制系统的算法联动，无法同步完成缺陷识别与切割路径自适应修正。

皮革普遍具有低刚度、高弹性、厚度不均、表面摩擦系数离散度大等物理特性，且其原料在裁床台面平铺作业过程中，受环境温湿度变化、空气流动、自重挤压、台面摩擦力、静电吸附等多重因素干扰，易出现局部拉伸、收缩、褶皱、翘边等不规则形变，导致标准 CAD 图纸与实物轮廓产生位置偏差。若继续沿用预设固定切割路径，易出现裁片缺料、边角破损、

原料浪费等问题。传统柔性裁床采用单横梁单刀头结构，设备结构简单、控制逻辑稳定，但单次仅能沿单一路径裁切，加工效率较低，难以满足大批量流水线生产需求。而多横梁并行裁切设备早期仅依靠固定机械限位划分裁切区域，缺少基于实物形态的动态避让算法，且相邻裁切刀组在加工临近裁片时，易发生机头磕碰、停机现象。为此，研究人员基于图像特征匹配技术开展柔软材料形变补偿研究，通过标准 CAD 图纸与实拍图像关键点匹配，实现切割路径微调。但现有路径优化研究普遍局限于算法仿真，缺少与多刀避让控制逻辑的结合，无法部署于实体设备。

针对上述问题，本文提出改进 SAM 的皮革缺陷分割与自适应裁切方法。首先，以 SAM-Med2D 为基础，在 ViT-B 图像编码器中引入通道-空间双分支适配器，通过提示编码器和掩膜解码器生成类别无关的二值缺陷掩膜；然后，对皮革图像进行预处理和模板匹配，并利用定向 FAST 和旋转 BRIEF（oriented FAST and rotated BRIEF, ORB）特征、汉明距离及随机抽样一致性（random sample consensus, RANSAC）算法估计全局仿射矩阵，实现裁片定位和 CAD 裁切路径校正；最后，根据裁片轮廓总长度将任务分配至左、中、右 3 个裁切区域，结合相邻刀头安全距离校验规则和分级避让逻辑，实现多横梁协同裁切。

1 相关研究

1.1 Leather-SAM 网络架构

本文利用 Leather-SAM 实现多种皮革表面缺陷的类别无关二值分割。Leather-SAM 沿用了 SAM-Med2D 三段式架构，主要由 ViT-B 图像编码器、提示编码器和双向 Transformer 掩膜解码器组成。Leather-SAM 网络架构及掩膜生成流程如图 1 所示。

1）将皮革图像变换为（256×256）像素，经步长为 16 的图像块嵌入后，形成 16×16 个 768 维的视觉 token。

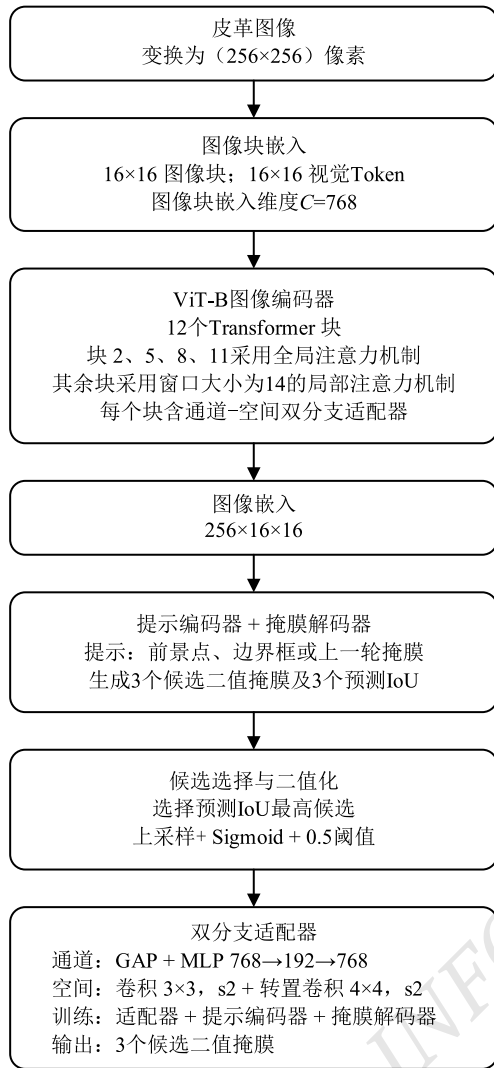


图1 Leather-SAM 网络架构及掩膜生成流程

2) ViT-B 图像编码器包含 12 个 Transformer 块。

其中，第 2、5、8、11 个块采用全局注意力机制，其余块采用窗口大小为 14 的局部注意力机制。在每个块的第二个残差阶段，将通道-空间双分支适配器与原多层感知机 (multilayer perceptron, MLP) 并行连接。其中，通道分支采用全局平均池化及 $768 \rightarrow 192 \rightarrow 768$ 映射进行通道重标定；空间分支采用步长为 2 的 3×3 卷积和 4×4 转置卷积提取并恢复空间特征，最后与输入残差相加并执行层归一化操作。

3) 经过 12 个 Transformer 块处理后，ViT-B 图像编码器输出尺寸为 $B \times 16 \times 16 \times 768$ 的视觉特征。首先，将该视觉特征尺寸调整为 $B \times 768 \times 16 \times 16$ ；然后，

先通过颈部模块中的 1×1 卷积将通道数由 768 降至 256，再经过 3×3 卷积融合局部空间位置信息；最后，两层卷积后均进行二维层归一化操作，最终得到尺寸为 $B \times 256 \times 16 \times 16$ 的图像嵌入。该图像嵌入保留了皮革图像的缺陷语义信息与空间位置信息，并作为掩膜解码器的图像特征输入。

4) 提示编码器将前景点、边界框或上一轮掩膜编码为提示特征。掩膜解码器通过双向 Transformer 融合图像嵌入与提示特征，并为同一提示生成 3 个候选二值掩膜及 3 个预测 IoU。

5) 3 个输出通道各对应 1 个候选二值掩膜而非类别，推理时选取预测 IoU 最高的候选，将其上采样至原图尺寸后，通过 Sigmoid 激活函数并以 0.5 为阈值进行二值化，从而得到二值缺陷掩膜。

6) 训练时，冻结 ViT-B 图像编码器中除通道-空间双分支适配器之外的参数，仅对 12 个通道-空间双分支适配器、提示编码器和掩膜解码器进行更新。优化器采用 Adam，其初始学习率、动量参数、权重衰减分别设为 1×10^{-4} 、(0.9, 0.999)、0。损失函数为

$$L = 20 L_{\text{focal}} + L_{\text{dice}} + L_{\text{iou}}$$

式中： L_{focal} 为像素级二分类 Focal 损失，其正负样本平衡系数 $\alpha = 0.25$ ，难易样本调节系数 $\gamma = 2$ ，用于增强模型对前景中较小或难以区分缺陷像素的学习； L_{dice} 为软 Dice 损失，根据 Sigmoid 激活函数处理后的连续掩膜概率与人工标注二值掩膜之间的重叠程度计算，用于缓解缺陷前景与背景之间的面积不平衡； L_{iou} 为 IoU 回归损失，采用均方误差计算 IoU Head 预测的掩膜质量分数与软 IoU 监督值之间的差异，其中软 IoU 监督值由候选掩膜的连续前景概率与人工标注二值掩膜计算得到，并在参与回归损失计算时停止其梯度传播。上述所有损失项均先按样本单独计算，再在批次维度上取平均值。

1.2 掩膜插值恢复和二值化处理

首先，掩膜解码器生成低分辨率的掩膜概率图；然后，通过双线性插值将其上采样至输入图像尺寸，

进一步映射为原始图像分辨率；最后，采用固定阈值对连续掩膜概率图进行二值化处理，得到二值掩膜概率图。

1.3 ORB 特征匹配与全局仿射变换矩阵

首先，根据相机标定比例将 CAD 轮廓栅格化为模板图；然后，分别从模板图与实拍图像中提取 FAST 关键点，即由 FAST 算法检测得到局部灰度变化显著的关键点，并计算关键点主方向，生成具有旋转不变性的 BRIEF 二进制描述符；最后，根据描述符的汉明距离建立初始匹配关系，采用比值检验筛选相似度不足的匹配点，再通过 RANSAC 算法剔除误匹配点对，以避免由任意 3 组匹配点直接求解造成全局仿射变换矩阵不稳定。

全局仿射变换矩阵含 6 个自由参数(a_{11} 、 a_{12} 、 a_{21} 、 a_{22} 、 t_x 、 t_y)。为求解该矩阵的未知参数，至少需要 3 组非共线点构造线性方程组。所得 2×3 变换矩阵用于将标准 CAD 坐标映射至实物坐标，从而在整体线性形变条件下对 CAD 路径进行坐标校正。

1.4 多横梁裁切安全控制

将裁床工作台划分为左、中、右 3 个裁切分区，分别对应左裁切横梁、中裁切横梁、右裁切横梁。各裁切分区内所有裁片总长度近似相等，以均衡三组横梁的工作量。裁切分区规则如表 1 所示。

表 1 裁切分区规则

分区	对应执行机构	裁切顺序	特殊规则
左裁切	左裁切横梁	自左至右 自上而下	不可切区 暂停等待
中裁切	中裁切横梁	自左至右 自上而下	优先裁切，完工主动避让
右裁切	右裁切横梁	自左至右 自上而下	不可切区，可划为代加工区

皮革裁切前，逐裁片进行相邻刀头安全距离校验，并根据校验结果执行分级避让逻辑。定义 S_1 为中裁切横梁与左裁切横梁的安全间距； S_2 为中裁切横梁与右裁切横梁的安全间距； D 为相邻刀头执行下一段裁切路径时的预测最小间距。相邻刀头处于临界安全距离

时的几何关系如图 2 所示。

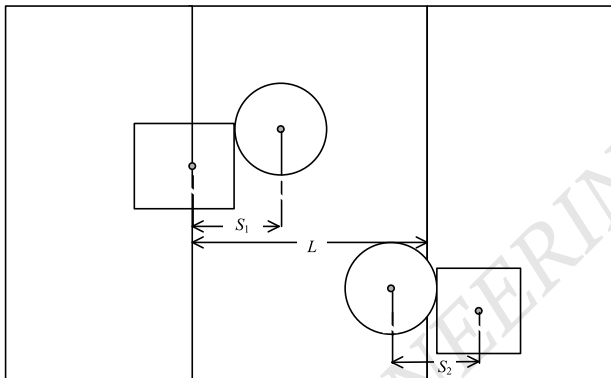


图 2 相邻刀头处于临界安全距离时的几何关系

图 2 中：对于左、中裁切横梁， $D=S_1$ ；对于右、中裁切横梁， $D=S_2$ 。

安全距离校验规则：1) $D > S$ ，相邻裁切横梁同步正常裁切；2) $D \leq S$ ，触发分级避让逻辑。

分级避让逻辑：当相邻刀头均满足安全距离校验规则时，各横梁并行裁切；当任一安全距离不足时，优先在侧横梁任务队列中搜索无冲突裁片；若无法通过任务换序消除冲突，则暂停对应侧横梁，由中裁切横梁完成当前裁片后重新校验安全距离。

2 基于 Leather-SAM 的缺陷识别与裁切优化算法

本文基于 Leather-SAM 网络架构、ORB 特征匹配与全局仿射变换矩阵、多横梁裁切安全控制等，构建基于 Leather-SAM 的缺陷识别与裁切优化算法。该算法主要包括图像预处理、基准模板匹配定位裁片、Leather-SAM 缺陷分割流程、提取 ORB 关键点与构建双特征数组、特征匹配与仿射矩阵求解、CAD 路径坐标变换、路径排序分区分配、多横梁安全距离校验与避让控制等环节。

2.1 图像预处理

针对皮革原料光照不均、纹理杂乱、成像噪声大等问题，本文采用图像多级预处理算法进行色彩空间转换、灰度化、自适应二值化、高斯模糊滤波。

1) 色彩空间转换：将皮革原料 RGB 图像转换到 HSV 空间，在 OpenCV 中置设 H 为 $65 \sim 90$ 、 S 为

50~160、 V 为 15~170; 通过 inRange 获得背景掩膜并通过取反, 得到皮革图像候选前景; 采用形态学处理和连通域面积筛选, 去除台面孔洞与零散噪声, 并恢复至 RGB 格式供后续模型处理。

2) 灰度化: 采用工业通用灰度换算公式 $Gray=0.299R+0.587G+0.114B$, 将 RGB 三通道图像转换为单通道灰度图; 该灰度图用于裁片轮廓提取、ORB 关键点检测和特征匹配。

3) 自适应二值化: 动态滑动阈值分割, 高于阈值时像素置 255, 低于置 0, 以分离裁片与背景。

4) 高斯模糊滤波: 采用高斯卷积核进行邻域加权平滑, 以抑制二值图像中的孤立噪声和轮廓像素毛刺, 减少局部灰度突变对轮廓提取的干扰。滤波后的二值图像用于提取裁片外轮廓, 并为后续基准模板制作和 ORB 特征匹配提供图像输入。

2.2 基准模板匹配定位裁片

基准模板是皮革全幅图像裁片匹配的参考标准, 其制作步骤如下:

1) 采集标准、无变形的基准裁片图像, 并利用图像多级预处理算法提取其外轮廓;

2) 采用逐层下采样的方式构建多尺度图像金字塔, 以降低图像分辨率差异对基准模板匹配的影响, 尺度参数需根据皮革全幅图像和基准模板的原始分辨率、最小模板尺寸和验证集上的定位准确率、误匹配率和运行时间综合确定;

3) 在验证集确定的旋转角度搜索范围、角度步长和尺度范围内生成旋转缩放模板, 并记录各模板的尺度、角度和轮廓特征;

4) 将全部特征参数(包括模板编号、原始图像尺寸、裁片外轮廓点集、多尺度图像金字塔层级、旋转角度、尺度系数、ORB 关键点坐标及其旋转 BRIEF 描述符等)打包生成 YAML 格式标准模板文件, 存入模板库, 用于后续皮革全幅图像批量匹配。

对待定位皮革全幅图像进行匹配时, 在预设尺度和角度范围内计算模板匹配得分, 候选结果经非极大值抑制后输出位置与偏转角。

2.3 Leather-SAM 缺陷分割流程

首先, 将预处理后的 RGB 图像调整为(256×256)像素, 并将前景点、边界框或上一轮掩膜作为提示信息输入 Leather-SAM; 然后, ViT-B 图像编码器、提示编码器分别生成图像嵌入、提示特征; 接着, 掩膜解码器据此输出 3 个候选二值掩膜与 3 个预测 IoU; 最后, 选择预测 IoU 最高的候选, 经双线性插值、Sigmoid 映射和 0.5 阈值处理, 得到二值缺陷掩膜。

二值缺陷掩膜中的前景像素表示提示区域内的疑似缺陷, 背景像素表示正常区域。进一步统计各连通缺陷区域的像素面积、外接矩形、中心坐标和边界轮廓, 并将其转换为后续排版和裁切避让所需的缺陷区域数据。若需要获得缺陷实际物理尺寸, 还需根据相机标定比例将像素坐标转换为裁床坐标。

2.4 提取 ORB 关键点与构建双特征数组

皮革等柔软材料形变通常表现为整体平移、旋转、均匀缩放及各向同性或异性拉伸。对于形变量较小(即线性度良好)的材料(如牛皮、较厚合成革), 可采用全局仿射变换实现裁切路径自适应修正。本文实验所用的样本中, 90%以上裁片的整体形变率≤15%(通过匹配点对残差统计得到), 全局仿射变换可接受在此范围内的误差。

根据 ORB 特征匹配方法, 构建双特征数组(目标特征数组和基准特征数组)。双特征数组均采用相同的 ORB 特征提取流程, 具体构建过程如下:

1) 目标特征数组 P_{tar} : 对待定位裁片的实拍图像进行轮廓提取, 确定裁片区域, 并在该区域内执行上述 ORB 特征提取流程, 得到 P_{tar} ;

2) 基准特征数组 P_{std} : 依据相机标定比例将 CAD 裁切轮廓栅格化, 生成基准模板图, 并在模板图中执行上述 ORB 特征提取流程; 同时保存模板关键点的像素坐标及其对应的 CAD 坐标, 组成 P_{std} 。

2.5 特征匹配与仿射矩阵求解

利用汉明距离完成目标特征数组 P_{tar} 与基准特征数组 P_{std} 特征的配对。筛选 3 组以上非共线匹配点, 构建方程组:

$$\begin{cases} x' = a_{11}x + a_{12}y + t_x \\ y' = a_{21}x + a_{22}y + t_y \end{cases}$$

式中： a_{11} 、 a_{12} 、 a_{21} 、 a_{22} 为缩放旋转系数， t_x 、 t_y 为平移系数。

利用方程组求解 a_{11} 、 a_{12} 、 a_{21} 、 a_{22} 、 t_x 、 t_y 6 个系数，生成 2×3 全局仿射变换矩阵 M 。

2.6 CAD 路径坐标变换

利用全局仿射变换矩阵 M 对基准切割路径数组 $Path_{std}$ 逐坐标映射，生成适配实际形变的目标切割路径数组 $Path_{tar}$ ，从而完成 CAD 图纸随皮革原料变形的动态修正。

2.7 路径排序分区分配

本文利用贪心算法优化所有修正后的裁切路径执行顺序：以当前刀头位置为起点，每次选择距离最近的未裁切路径起点作为下一个目标，生成近似最短的空行程移动序列。该算法计算量小，可满足裁切实时性要求。裁切路径排序后分配至对应横梁，最终输出各横梁的有序裁切路径序列。

2.8 多横梁安全距离校验与避让控制

修正后的 CAD 路径需进行多横梁裁切安全控制。首先，修正后的 CAD 路径按裁片位置和轮廓总长度分配至左、中、右 3 个任务队列；然后，每条路径执行前，根据相邻刀头当前位置及下一段裁切路径计算预测最小间距，将其与安全距离比较并执行分级避让逻辑。

3 实验与结果分析

3.1 实验数据集

本实验数据集取自广东省广州市、佛山市某皮鞋制造企业生产现场。利用工业相机采集分辨率为 7700×1589 的皮革原图，经翻转、缩放、亮度扰动等数据增强后，共获得 2870 幅有效样本。该数据集包含羊皮、牛皮两类基材；缺陷尺寸跨度大，微小划痕像素占比不足 0.5%，大块破损占比超过 30%，缺陷分布接近真实产线。该数据集根据 SATRA 分级标准划分为 6 个品质等级，用于缺陷识别、品级判定联合测试。

3.2 实验软硬件环境

硬件：Intel i7-7700 CPU，双路 NVIDIA RTX 4070 Ti SUPER GPU（合计 40 GB 显存）。

软件：Windows 11 操作系统，Python 3.10、PyTorch 2.5.1 深度学习框架，OpenCV 图像处理库。

评价指标：分割任务采用 IoU、Dice；裁切任务采用裁切效率、刀头碰撞故障率。

3.3 Leather-SAM 运行性能与双分支结构分析

为验证 Leather-SAM 权重加载的完整性和模型推理性能，基于 SAM-Med2D ViT-B 及通道-空间双分支适配器构建网络，并以严格模式加载已训练好的模型。首先，核对权重文件中的参数名称和张量形状，统计缺失键、多余键及形状冲突情况；然后，将模型切换至评价模式并关闭梯度计算，在 NVIDIA RTX 4070 Ti SUPER GPU 上对测试目录中的 239 张图像逐幅推理；接着，将输入图像统一调整为 (256×256) 像素，并采用测试脚本中相同的提示配置和候选二值掩膜选择规则；最后，推理完成后，统计模型前向计算时间、空掩膜数量以及 IoU Head 输出的候选二值掩膜质量分数预测值。Leather-SAM 实测运行性能如表 2 所示。

表 2 Leather-SAM 实测运行性能

测试项目	数值
权重严格加载	缺失键：0 多余键：0
GPU 推理图像	239 张
平均吞吐量	约 12.2 f/s
空掩膜数量	8 张 (3.35%)
IoU Head 平均预测质量分数	86.666%

由表 2 可知：Leather-SAM 模型中 386 个参数张量均能够与 SAM-Med2D ViT-B、通道-空间双分支适配器匹配，未出现缺失键、多余键或形状冲突；在 GPU 上完成 239 张图像的前向推理，平均单张图像的推理时间为 0.0817 s，对应模型推理速度约为 12.2 f/s，该时间不包括图像读取、图像尺寸调整、提示框生成和结果保存所用时间，仅反映模型前向计算性能；239 张

图像中有 8 张图像（约占 3.35%）输出空掩膜，表示模型未在提示区域内分割出缺陷，该现象既可能对应无缺陷样本，也可能由于微小缺陷在缩放过程中丢失、提示框不准确或模型漏检等原因导致，因而不能直接用于评价分割准确率；模型内部 IoU Head 输出的平均预测质量分数为 86.666%，该数值是模型对自身掩膜质量的估计，并非由预测掩膜与人工真值掩膜计算得到的真实 IoU。

3.4 裁切性能对比实验

选取传统单刀固定裁切、三刀固定分区无避让裁切与本文方法（三刀安全距离避让裁切）进行裁切性能对比实验，连续 72 h 的实验结果如表 3 所示。

表 3 裁切性能对比实验结果

裁切方案	平均裁切效率/(片/h)	刀头碰撞故障率/%	原料报废率/%
单刀固定裁切	126.0	0	4.82%
三刀固定分区无避让裁切	218.0	12.76%	3.95%
本文方法	219.7	0.28%	1.17%

由表 3 可知：本文方法的平均裁切效率为 219.7 片/h，相较于单刀固定裁切提高 74.4%，相较于三刀固定分区无避让裁切提高 0.8%；本文方法的刀头碰撞故障率由 12.76%降低至 0.28%，原料报废率由 3.95%降低至 1.17%，表明多横梁裁切安全控制能够在保持多横梁并行裁切效率的同时，降低横梁运动冲突和异常裁切造成的生产损失。

3.5 消融实验

为分析通道分支、空间分支在已训练 Leather-SAM 中的作用，在不重新训练模型的条件开展固定权重的分支消融实验。该实验统一加载已训练好的模型的权重，并设置无分支、仅通道分支、仅空间分支和双分支 4 种推理配置。4 种推理配置均采用相同的 231 对可评测图像与人工标注掩膜，输入图像统一调整为 (256×256) 像素，并使用由真值掩膜外接矩形生成的相同边界框提示。推理过程中保持权重文件、候选二值掩膜选择规则、双线性插值方式和 0.5 二值

化阈值一致，分别计算预测掩膜与人工标注掩膜之间的 IoU、Dice 和轮廓 F1。实验过程中不更新网络参数，以确保结果差异仅来自推理阶段通道分支和空间分支的启用状态。实验结果如表 4 所示。

表 4 通道-空间双分支适配器匹配消融实验 %

推理配置	IoU	Dice	轮廓 F1
无分支	4.78	7.74	15.65
仅通道分支	1.58	3.02	5.80
仅空间分支	8.21	13.25	27.23
双分支	54.58	65.22	88.08

由表 4 可知，双分支配置的 IoU、Dice 和轮廓 F1 分别达到 54.58%、65.22%和 88.08%，均高于其余 3 种配置，表明已训练模型的通道与空间特征处理需要协同工作，单独关闭任一分支均会破坏权重训练时形成的特征分布。

4 工程应用分析

本方法已在广东某皮革裁床设备完成样机部署。在实际运行中，单张整皮图像采集、缺陷识别、路径优化全流程耗时≤1.2 s（图像采集与预处理≈50 ms，模板匹配≈30 ms，Leather-SAM 推理≈950 ms（GPU 加速），路径生成与优化≈170 ms），可适配流水线连续裁切。

5 结论

本文构建了基于 SAM-Med2D ViT-B 和通道-空间双分支适配器的皮革表面缺陷分割模型 Leather-SAM，明确了适配器插入位置、训练参数、组合损失和 3 个候选二值掩膜输出方式。该方法的 IoU、Dice 和轮廓 F1 分别为 54.58%、65.22%和 88.08%；固定权重的分支消融实验结果表明，双分支适配器在皮革表面缺陷识别与智能裁切流程中的可行性。进一步研究可考虑：

1) 建立原始图像、皮料、采集批次、缺陷类别和增强样本之间的数据血缘，按照皮料或采集批次实施组级训练集、验证集、测试集划分，并仅对训练集进行数据增强；

2) 扩展不同旋转、缩放、剪切和局部拉伸条件下的路径补偿试验,系统分析重投影、轮廓、路径与实际切边误差,并完善全局仿射残差超限时的 TPS、分片仿射和自动拒绝机制;

3) 进一步优化 Adapter 的参数效率,在统一数据和提示条件下完成各结构独立训练消融,并分别记录分割任务、路径优化与多横梁控制的评价指标、运行日志和代码版本。

©The author(s) 2024. This is an open access article under the CC BY-NC-ND 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

参考文献

- [1] SERAFIM A F L. Multiresolution pyramids for segmentation of natural images based on autoregressive models: Application to calf leather classification[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 5562-5570.
- [2] LYTHGOE L. Effective leather grading. Accessed: Aug. 2011. [Online]. Available, [2026-08-17]. <https://www.satara.com/bulletin/article.php?id=708>.
- [3] LIU G, CAI N, XIAO P, et al. Leather defect detection based on photometric stereo and saliency object detection[J]. Computer Engineering and Application, 2019,55(8):215-219.
- [4] SANTOS FILHO E Q S, SOUSA P H F, REBOUAS FILHO P P, et al. Evaluation of goat leather quality based on computational vision techniques[J]. Circuits Systems and Signal Processing, 2020,39(2):651-673.

作者简介:

郭瑞桁,男,2001年生,硕士研究生,主要研究方向:机械制造。E-mail: 1017011691@qq.com

郭瑞洲,男,1991年生,硕士研究生,高级工程师,主要研究方向:软件开发。E-mail: ryan.guo@vip.163.com

刘夏丽,女,1992年生,硕士研究生,工程师,主要研究方向:智能制造、设备可靠性建模。E-mail: lxl@gdut.edu.cn

声明

本刊从 2024 年 7 月 1 日开始为开放获取期刊,所有录用和发表的论文可以立即并永久供读者免费阅读、下载、链接。本刊采用知识共享-署名-非商业-禁止演绎许可协议(Creative Commons Attribution-NonCommercial-No-Derivs),允许第三方用户按照(CC-BY-NC-ND 4.0),即署名(BY)、非商业用途(NC)、禁止演绎(ND)的方式共享,前提是原作未以任何方式进行更改,并且被适当引用。详情请访问 <https://creativecommons.org/licenses/by-nc-nd/4.0/>。

- [5] GAN Y S, CHEE S S, HUANG Y C, et al. Automated leather defect inspection using statistical approach on image intensity[J]. Journal of Ambient Intelligence and Humanized Computing, 2021,12(10):9269-9285.
- [6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.
- [8] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261-2269.
- [9] KRIZHEVSKY A, SUTSKEVER I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25.
- [10] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. arXiv Preprint arXiv:1602.07360, 2016.
- [11] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018:6848-6856.
- [12] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:1-9.
- [13] KIRILLOV A, MINTUN E, RAVI, et al. Segment anything[C]. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023:4015-4026).