基于改进 U-Net 的皮革表面缺陷精细化分割方法

王愉锦[1,2)], 黄惠玲[2)], 付磊[1,2)], 韩军[2)*]

[1)] (福州大学先进制造学院 泉州 362000)
[2)] (中国科学院福建物质结构研究所泉州装备制造研究中心 泉州 362000)
(junhan@fjirsm.ac.cn)

摘 要: 针对皮革缺陷形态多变、局部相似程度高导致细节信息难以提取、类型错分的问题, 提出一种基于改进 U-Net 结构的表面缺陷精细化分割方法. 编码端, 在保留原始图像细节特征的同时嵌入级联扩张卷积模块获取全局特征, 同时在跳跃连接中添加特征融合模块, 改善因高低特征张量直接拼接造成的局部信息丢失; 解码端使用基于通道注意力机制的解码模块代替原始卷积层, 自适应地指导网络关注缺陷区域; 为进一步整合高层信息, 还嵌入全局平均池化模块, 将输出结果作为解码端的语义指导增强网络对相似缺陷的分辨能力. 在包含 7 种缺陷的皮革数据集上进行实验的结果表明, 所提方法在 PA, MPA, FWIoU 和 MIoU 上分别达到 99.17%, 93.27%, 98.39%和 88.88%, 对比 U-Net 分别提升 0.28, 2.78, 0.53 和 4.03 个百分点; 定性分析和定量分析结果表明, 该方法对于皮革缺陷能得到更加精细的分割结果.

关键词: 皮革缺陷; 语义分割; U-Net; 特征融合; 注意力机制; 扩张卷积
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024.19832

Refined Segmentation Network for Leather Surface Defect Detection Based on Improved U-Net

Wang Yujin[1,2)], Huang Huiling[2)], Fu Lei[1,2)], and Han Jun[2)*]

[1)] (*School of Advanced Manufacturing, Fuzhou University, Quanzhou* 362000)
[2)] (*Quanzhou Institute of Equipment Manufacturing, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Quanzhou* 362000)

Abstract: Leather defects with variable morphology and high local similarity are of difficulty in extracting features comprehensively and accurately. In this work, a refined surface defect segmentation method based on improved U-Net network is proposed. On the encoder side, a cascaded dilated convolution module is embedded to obtain the global features while preserving the detail information of the original image, and a feature fusion module is added to the jump connection to reduce local features loss caused by directly splicing of the high-level and low-level feature tensor; on the decoder side, a decoding module based on the channel attention mechanism, which can guide the network to adaptively focus on defective regions, is used to replace the original convolutional layer; to further integrate high-level information, a global average pooling module is embedded as the semantic guide to improve the discrimination capability of the network from similar defects at the decoding end. The experimental results conducted on a leather dataset containing 7 kinds of defects show that the proposed method achieves 99.17%, 93.27%, 98.39%, and 88.88% in PA, MPA, FWIoU, and MIoU, which is 0.28, 2.78,

收稿日期: 2022-07-09; 修回日期: 2023-02-17. **基金项目**: 福建省科技计划(2021T3060, 2021T3032); 福建省闽都实验室主任基金(2021ZR107); 泉州市科技计划(2021C063L). **王愉锦**(1999—), 女, 硕士研究生, 主要研究方向为深度学习、语义分割; **黄惠玲**(1990—), 女, 硕士, 工程师, 主要研究方向为光学检测、算法研究; **付磊**(1998—), 男, 硕士研究生, 主要研究方向为机器学习; **韩军**(1983—), 男, 博士, 研究员, 硕士生导师, 论文通信作者, 主要研究方向为光学成像与检测技术、无损检测.

0.53, and 4.03 percentage points better than that of U-Net. The qualitative and quantitative analysis results demonstrate that the algorithm proposed has remarkable ability to refine the segmentation in leather defect recognition.

Key words: leather defects; semantic segmentation; U-Net; feature fusion; attention mechanism; dilated convolution

皮革是日用品的必需材料, 在皮革原料加工前后, 蝇虫叮咬、烙铁标记、机械拉扯等多种因素会导致皮革表面出现缺陷[1]. 表面缺陷会严重影响后续的生产加工, 进而影响皮革制品的质量和使用寿命. 长期以来, 皮革缺陷采用人工方式检测, 易受情绪、体力等因素影响, 无法保证产品质量. 因此, 设计一种自动化的皮革表面缺陷检测分割方法, 对实现我国制革业自动化和产业化具有重大意义[2].

目前, 传统的皮革缺陷分割方法包括阈值法[3]、边缘法[4]、频谱法[5]和聚类法[6]等. Kasi 等[4]提出一种自适应边缘检测方法, 用于清晰地提取不规则的缺陷边界; 贺福强[6]结合粒子群优化和模糊聚类方法实现缺陷纹理信息聚类: 这些方法通常针对具体类型进行设计, 需要手动调整阈值, 且对光照和颜色变化非常敏感, 难以适应实际复杂工业场景中的缺陷检测.

近年来, 深度学习技术发展迅速, 语义分割通过给每个像素分配类型标签实现自动分割. 与传统方法相比, 语义分割的自学习特性能显著地降低外界因素干扰, 避免人工设计特征提取器的烦琐工序, 给相关工作提供了新的思路. 2015 年, Long 等[7]提出像素级、端到端的全卷积网络(fully convolutional networks, FCN), 奠定了语义分割的基石, 但其中的粗糙上采样结构导致大量细节特征丢失. 为改善该状况, 语义分割领域涌现一批新技术, 如编码-解码技术、基于感受野的技术、特征融合技术、基于注意力机制的技术和多尺度技术等. U-Net[8]是经典的编码-解码结构, 它由对称的收缩路径和扩张路径构成, 在医学影像等纹理丰富的图像中展示了优秀分割的效果. 为扩大感受野, Chen 等[9]提出 DeepLab 系列, 通过在卷积核中插孔的方式平衡全局和局部信息; 而 Peng 等[10]在全局卷积网络(global convolutional network, GCN)中引入大卷积核, 与堆叠小卷积核相比, 该方法能获得更完整的高级语义. 在特征融合方面, Zhang 等[11]利用交叉融合弥补高低层信息间的差距; Liu 等[12]通过反池化后的全局特征与局部特征串联融合, 以此添加上下文信息. Fu 等[13]提出包含空间域和通道域的双注意网络(dual attention network, DANet), 通过加强图像空间和通道间的依赖关系提升模型精度. Li 等[14]在金字塔注意力网络(pyramid attention network, PAN)中混合多种技术, 将多尺度与注意力机制结合获取像素级注意力, 增强局部的一致性. 以上方法根据像素间依赖性对全局上下文建模, 提高对象内部一致性, 实现多种场景下的分割任务.

目前, 将语义分割应用于皮革缺陷分割的工作相对较少, 且已有方法[15-16]往往针对区分度较大、类型单一的目标, 缺乏对纹理丰富及形态多变的缺陷进行分割; 特别是在天然环境下, 根据损伤程度不同, 缺陷会表现出局部形态差异, 极容易造成类间错分和类内漏分, 增加任务难度.

为了保留更多缺陷形态信息, 克服类间相似度高、局部区分度低导致的边界模糊、分割错误等问题, 本文提出一种基于 U-Net 编码-解码结构的精细化分割方法. 通过在编码路径嵌入级联扩张卷积模块(cascaded dilated convolution module, CDCM), 提升全局语义的提取能力; 为解决语义鸿沟带来的局部信息损失, 在跳跃连接中加入特征融合模块(feature fusion module, FFM); 最后, 提出一种基于通道注意力机制的解码模块(channel attention mechanism module, CAMM), 用于自适应地关联缺陷相关通道, 解决类间错分和类内漏分, 提升网络整体分割性能.

1 本文方法

1.1 网络结构

图 1 所示为本文方法框架. 输入图像首先经过 5 组卷积层, 其中, 前 3 组每组卷积完成后对图像进行最大池化下采样操作, 后 2 组用重叠池化的方式代替普通下采样, 并在第 5 组卷积层中引入 CDCM 进行密集预测, 得到特征图尺寸为原始图像的 1/8; 然后将编码端每一层级的输出结果经过 FFM, 与解码端对应层级的特征向量共同作为输入, 在 CAMM 中完成通道加权, 在第 5 组卷积层顶端嵌入全局平均池化模块(global average pooling module, GAPM), 作为该层 CAMM 中的高层语义

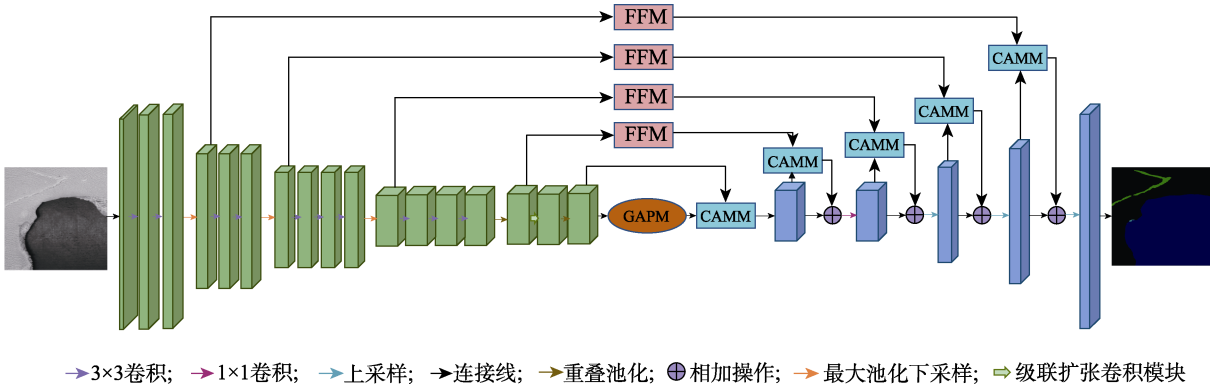


图 1　本文方法框架

输入；最后将解码端输出结果依次进行叠加与上采样，得到分割结果图.

1.2　CDCM

皮革缺陷包含丰富的形态特征，低层分辨率信息对于正确分割至关重要. 为保留原始图像的空间细节，取消编码路径后 2 层的下采样操作. 感受野计算公式为

$$R_l = R_{l-1} + (k_l - 1) \times \prod_{i=1}^{l-1} s_i \tag{1}$$

其中，s_i 表示步长. 可以看出，取消下采样会使步长 s_i 减小，从而感受野 R_l 减小；而感受野是获取语义特征的关键，若感受野缺失，则导致网络失去全局性判断，无法识别多尺度缺陷.

为了保留细节特征的同时获取全局语义信息，在第 5 层级引入 CDCM，其结构如图 2 所示. 该模块由 3 个串联的子模块组成，每个子模块中都包含一个 D_{conv} 和 ReLU 激活函数，其中，$D_{\text{conv}}\left[(3,3),d\right]$ 表示扩张卷积运算，$(3,3)$ 和 d 分别表示卷积核大小和扩张系数. 新的卷积核尺寸为

$$k_{\text{new}} = k_{\text{ori}} + (k_{\text{ori}} - 1)(d - 1) \tag{2}$$

其中，k_{ori} 表示原始卷积核大小；d 表示扩张系数.

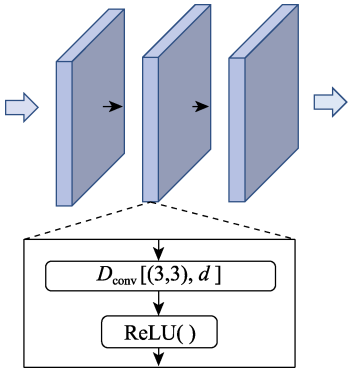


图 2　CDCM 结构

通过改变扩张系数 d，能任意调节输入和输出卷积核间的对应关系，进而灵活地控制感受野大小. CDCM 能在一定程度上弥补下采样次数减少带来的视野不足，保留局部细节的同时获取更多语义信息，提升网络识别多尺度缺陷的能力.

1.3　FFM

为了弥补收缩路径中损失的局部信息，U-Net 加入跳跃连接结构，但该结构仅将张量直接进行拼接，不仅使信息间流通不畅，还造成语义鸿沟，导致实际得到的局部特征十分有限.

为了克服这一问题，受到 ExFuse[11]工作的启发，本文在跳跃连接中增添 FFM，该模块通过将高层信息引入低层张量，使其编码更多高级语义，减少高低特征图之间的差异. 如图 3 所示，FFM 包含融合模块(fusion module，FM)和上采样模块(upsampling module, UM)这 2 部分.

UM 用于聚集高层语义信息. 对于第 l 层特征图，将 $l+1 \sim L$ 层特征图上采样后在通道维度上进行拼接，为得到的中间语义信息

$$\begin{aligned} \boldsymbol{S}_l &= \text{UM}\left(\boldsymbol{X}_{l+1}, \boldsymbol{X}_{l+2}, \cdots, \boldsymbol{X}_L\right) = \\ &\text{concat}\left(u\left(\boldsymbol{X}_{l+1}\right), u\left(\boldsymbol{X}_{l+2}\right), \cdots, u\left(\boldsymbol{X}_L\right)\right) \end{aligned} \tag{3}$$

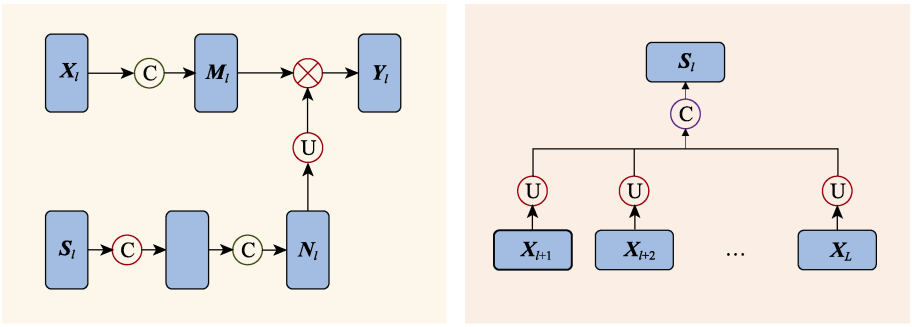
其中，$u(\cdot)$ 表示双线性插值上采样操作；$\text{concat}(\cdot)$ 表示张量拼接. .

FM 则用于将高级语义融入低层特征，加强全局依赖性. 该模块共分为 2 路进行操作. 对于上方分支，将 $\boldsymbol{X}_l$ 输入到一个 3×3 大小的卷积进行特征抽取，得到低维张量

$$\boldsymbol{M}_l = c\left(\boldsymbol{X}_l, 3\right) \tag{4}$$

其中，$c(\boldsymbol{A}, b)$ 表示卷积运算；$\boldsymbol{A}$ 为输入特征图；b 为卷积核尺寸. 本文中，$\boldsymbol{A} = \boldsymbol{X}_l$，$b = 3$.

对于下方分支，输入端 $\boldsymbol{S}_l$ 依次经过一个 1×1 卷积和一个 3×3 卷积，这 2 组卷积起到降维和提取



a. FM b. UM

⊗ 点乘; U 上采样; C 张量拼接; C 1×1卷积; C 3×3卷积

图 3 FFM 结构

全局信息的作用，得到高维特征张量

$$\boldsymbol{N}_l = c\left(c\left(\boldsymbol{S}_l,1\right),3\right) \tag{5}$$

将 2 路结果进行融合，取下路分支结果 $\boldsymbol{N}_l$ 进行双线性插值上采样，与上路结果 $\boldsymbol{M}_l$ 逐像素相乘，得到输出特征图

$$\boldsymbol{Y}_l = \boldsymbol{M}_l \otimes u\left(\boldsymbol{N}_l\right) \tag{6}$$

其中，$\otimes$ 表示点乘运算. $\boldsymbol{Y}_l$ 即为包含更多语义信息的特征张量，将其代替原始特征图 $\boldsymbol{X}_l$，作为该层编码端最终输出结果.

1.4 CAMM

皮革缺陷具有类间相似度高的特性，且损伤程度决定了同一缺陷内部存在局部差异. 针对这一问题，本文提出 CAMM，该模块融合高低层特征信息作为指导，对低层特征张量进行加权，从而自适应地关联缺陷相关通道，细化网络分割结果.

图 4 所示为 CAMM 结构图. 首先将低层张量 $\boldsymbol{L}\in\mathbb{R}^{H\times W\times C_\mathrm{l}}$ 与高层张量 $\boldsymbol{H}\in\mathbb{R}^{H\times W\times C_\mathrm{h}}$ 拼接，得到包含二者信息的 $\boldsymbol{D}\in\mathbb{R}^{H\times W\times\left(C_\mathrm{l}+C_\mathrm{h}\right)}$. 其中，高层张量部分用作语义指导，旨在增强全局一致性，而低层张量部分则用于提供特征的不同辨别信息.

对张量 $\boldsymbol{D}$ 进行全局最大池化和全局平均池化，生成 $\boldsymbol{M}\in\mathbb{R}^{1\times1\times\left(C_\mathrm{l}+C_\mathrm{h}\right)}$ 和 $\boldsymbol{N}\in\mathbb{R}^{1\times1\times\left(C_\mathrm{l}+C_\mathrm{h}\right)}$，其中的每个向量值 $\boldsymbol{M}_\mathrm{c}$ 和 $\boldsymbol{N}_\mathrm{c}$ 分别表示为

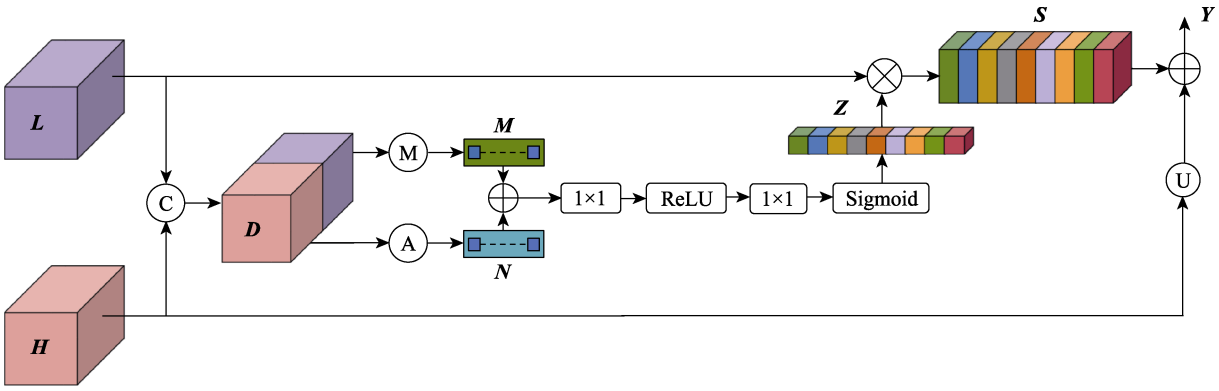
$$\boldsymbol{M}_\mathrm{c} = \max\left(\boldsymbol{D}_\mathrm{c}\right) \tag{7}$$

$$\boldsymbol{N}_\mathrm{c} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{D}_\mathrm{c}\left(i,j\right) \tag{8}$$

将 $\boldsymbol{M}$ 和 $\boldsymbol{N}$ 逐像素相加，得到融合上下文最大语义信息和平均语义信息的向量 $\boldsymbol{P}\in\mathbb{R}^{1\times1\times\left(C_\mathrm{l}+C_\mathrm{h}\right)}$；将其依次经过1×1卷积 φ，ReLU 函数，1×1卷积 ω 和 Sigmoid 激活函数，实现通道间信息交流和整合，生成权值向量 $\boldsymbol{Z}\in\mathbb{R}^{1\times1\times C_\mathrm{l}}$. $\boldsymbol{Z}$ 包含 $[0,1]$ 对 $\boldsymbol{L}$ 每个通道的选择权重，其值越接近 1，表示该通道与缺陷信息的关联程度越大；其计算公式为

$$\boldsymbol{Z} = \mathrm{Sigmoid}\left(\omega\left(\mathrm{ReLU}\left(\varphi\left(\boldsymbol{P}\right)\right)\right)\right) \tag{9}$$

将张量 $\boldsymbol{Z}$ 和低层输入 $\boldsymbol{L}$ 逐像素相乘，得到 $\boldsymbol{S}\in\mathbb{R}^{H\times W\times C_\mathrm{l}}$，完成对低层特征 $\boldsymbol{L}$ 的加权. 加权过程可以表述为



C 张量拼接; ⊗ 点乘; U 上采样; ⊕ 点加; M 全局最大池化; A 全局平均池化

图 4 CAMM 结构

$$\boldsymbol{S}=\boldsymbol{Z}\otimes\boldsymbol{L} \tag{10}$$

最后对高层特征进行上采样，结果与 $\boldsymbol{S}$ 进行像素相加，得到

$$\boldsymbol{Y}=\boldsymbol{S}\oplus u(\boldsymbol{H}) \tag{11}$$

其中，$\boldsymbol{Y}\in\mathbb{R}^{H\times W\times C_1}$ 即为该层编码端最终输出.

此外，本文嵌入 GAPM，将其作为第 5 组卷积层对应 CAMM 的高层语义输入端.

GAPM 的原理如图 5 所示. 对于输入特征图，首先进行通道维度的全局平均池化[17]，整合空间信息；然后对得到向量值进行像素复制，直至恢复原有尺寸. 这一操作能够学习缺陷的判别性特征，提供更强的全局一致信息，为正确识别缺陷像素提供语义特征.

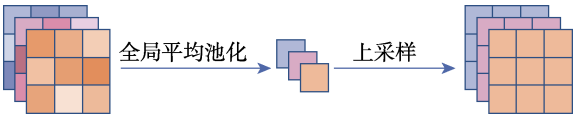


图 5　GAPM

2　实验设置

2.1　实验数据

本文实验中使用的皮革缺陷数据由峰安皮业股份有限公司提供，包括开创伤、刺刮伤、烙印、破洞、皮肤癣、烂面和刺猴 7 种缺陷类型，以及一类无缺陷样本. 图 6 所示为每种缺陷类型和无缺陷类型的具体形态，其中还包括一类无缺陷样本.

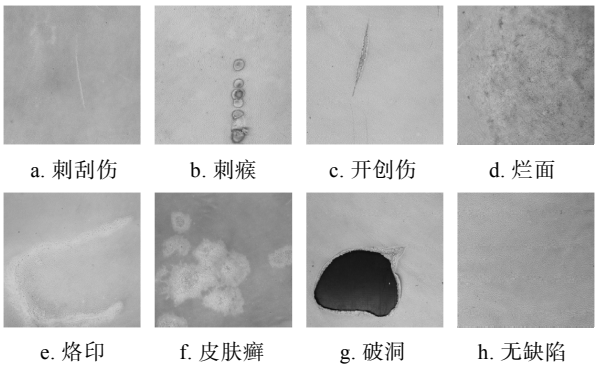


图 6　皮革缺陷类型和无缺陷类型图像

原始数据集中有 131 幅 3072×2048 像素的图像，为扩充小样本数据集，需对原始图像进行切割. 如图 7 所示，使用 768×768 大小的目标窗在原始图像上进行顺序裁剪，可以得到 15 幅带有部分重叠区域的图像，尺寸与目标窗相同.

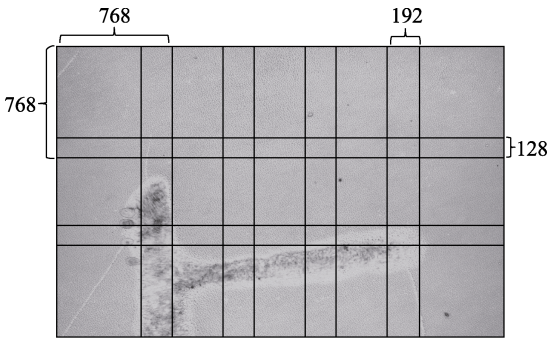


图 7　图像裁剪示意图

通过统计每种缺陷类型图像在扩充后数据集中的数量可以看出，破洞、皮肤癣、烂面、刺猴的图像数量远小于其他数据类型，为避免类型不均衡导致的分割误差，需对这 4 类缺陷样本进行增强；分别在破洞、皮肤癣、烂面、刺猴中选取适量图像，去除重复样本后，通过左旋 90°、左右镜像翻转、添加高斯噪声和模糊的方式进行增强，使得各类图像数量不低于 250 幅. 采用这种方式，可以使每种类型的图像在总缺陷图像中的占比约为 1/7. 表 1 所示为 7 种类型缺陷增强前和增强后的图像数量.

表 1　各类缺陷数量　　幅

类型	原始图像数量	增强后图像数量
开创伤	298	298
刺刮伤	353	353
烙印	253	253
破洞	99	263
皮肤癣	73	259
烂面	60	260
刺猴	59	255

将缺陷图像和无缺陷图像进行混合，共得到 2341 幅 768×768 像素的图像；对其按照 7∶1∶2 的比例随机划分，可以得到 1638 幅训练图像、235 幅验证图像和 468 幅测试图像.

2.2　评价指标

为了评估本文方法性能，对分割结果进行量化，选用语义分割中常用的像素精度(pixel accracy, PA)、平均像素精度(mean pixel accuracy, MPA)、平均交并比(mean intersection over union, MIoU)和加权交并比(frequency weighted intersection over union, FWIoU)作为评价指标[18]. 其中，MIoU 是语义分割领域中衡量分割精度的通用度量，作为主要性能指标. PA 表示正确标记的像素点占总像素点的比例，计算公式为

$$\mathrm{PA}=\frac{\sum_{i=0}^{k}p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}} \tag{12}$$

其中, p_{ii} 和 p_{ij} 表示真正值和假正值; k 表示缺陷类.

MPA 表示对每类分类正确的像素点数占该类像素的比值求平均后的结果, 计算公式为

$$\mathrm{MPA}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}} \tag{13}$$

MIoU 表示预测值和真实值交集与并集的比值, 然后对所有类型求均值后的结果, 计算公式为

$$\mathrm{MIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{14}$$

其中, p_{ji} 表示真负值.

FWIoU 根据每类出现的频率, 对各类的 IoU 值进行加权求和, 计算公式为

$$\mathrm{FWIoU}=\frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}}\sum_{i=0}^{k}\frac{\sum_{j=0}^{k}p_{ij}p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{15}$$

其中, 共有 $k+1$ 种类型数(包括 k 种缺陷类和 1 种正常类); p_{jj} 表示假负值; p_{ii}, p_{ij}, p_{ji} 和 p_{jj} 的含义如表 2 所示.

表 2 混淆矩阵

预测值\实际值	正样本	负样本
正样本	p_{ii}	p_{ij}
负样本	p_{ji}	p_{jj}

2.3 参数配置及实验环境

本文采用 768×768 像素的图像作为网络输入, 所有主干网络均使用在 ImageNet 数据集[19]上的预训练结果作为初始化权重. 训练的最大迭代次数为 230 次, 批量大小为 8, 初始学习率为 0.0001, 且每经过 50 个 epochs 学习率下降为原来的 0.5; 优化器选用 Adam 算法, 设置除偏置外的其他参数权值衰减为 10^{-6}, 损失函数采用像素级交叉熵损失函数. 所有实验均采用早停策略防止过拟合, 在损失函数连续 15 个 epochs 未下降时停止训练.

本文实验采用 Intel(R) Core(TM) i9-10900X, 2 张 NVIDIA GeForce RTX 3090 显卡, Ubuntu 18.04 操作系统, CUDA 版本为 11.4, 用 Pytorch 1.8.0 作为深度学习框架, Python 3.9.5 为编程语言.

3 结果与分析

3.1 消融实验

为了说明本文方法各模块的有效性, 设计 5 组实验进行对比, 结果如表 3 所示, 其中, 第 1 组为参照组. 可以看出, 第 2 组和第 3 组实验分别验证了主干网络和 CDCM 的有效性, 改进后的网络在分割精度上有较大提升, 尤其是 MPA 和 MIoU 分别提升 2.11 和 2.24 个百分点, 表明改进后的编码端能够获取更加丰富的上下文信息; 添加 FFM 后, 主要分割指标获得进一步提升, 说明该模块能在一定程度上消除语义差异, 细化分割效果; 继续添加 CAMM 和 GAPM, MIoU 值较前一组实验提升 1.37 个百分点, 说明该模块对聚焦缺陷形态、消除类间错分是有效的; 添加所有模块后精度达到最优, PA, MPA, FWIoU 和 MIoU 分别较基准提升 0.28, 2.78, 0.53 和 4.03 个百分点, 说明各模块对缺陷像素的提取和分割具有适用性.

3.1.1 CDCM 有效性的验证

为了说明在不同层级使用 CDCM 对结果的影响, 设置 3 组实验, 记作(5), (4,5), (3,4,5), 括号内的数字表示使用 CDCM 的层级, 初始化膨胀系数为 2. 实验结果如表 4 所示. 可以看出, 随着使用 CDCM 的层级增多, 准确率逐渐下降, 说明较早使用扩张卷积会使感受野增大速度过快, 削弱了

表 3 不同模块组合下的融合实验分割指标对比

实验	模块						分割指标/%			
	U-Net	VGG16	CDCM	GAPM	FFM	CAMM	PA	MPA	FWIoU	MIoU
第 1 组	√						98.89	90.49	97.86	84.85
第 2 组		√					99.02	91.48	98.11	85.75
第 3 组		√	√				99.07	92.60	98.22	87.09
第 4 组		√	√		√		99.11	92.66	98.25	87.51
第 5 组		√	√	√	√	√	**99.17**	**93.27**	**98.39**	**88.88**

注. 粗体表示最优值.

表 4　CDCM 在不同层级的分割指标对比　%

层级	PA	MPA	FWIoU	MIoU
(5)	**99.07**	**91.95**	**98.19**	**86.87**
(4,5)	99.03	90.28	98.13	85.01
(3,4,5)	98.61	87.41	97.34	80.24

注. 粗体表示最优值.

底层网络提取细节特征的能力；而皮革缺陷包含丰富的多尺度信息，过大的感受野会使网络对于小缺陷群的细节信息感知能力下降，从而分割精度降低.

进一步，对 CDCM 中的最优膨胀系数设置 3 组实验，结果如表 5 所示，括号内的数字表示 3 个卷积的膨胀系数值. 由式(1)(2)可知，膨胀系数的选取间接影响感受野的大小，且膨胀系数越大，感受野越大. 可以看出，在一定范围内，选取更大的膨胀系数有助于网络捕获更多信息，当 3 个膨胀系数都选取 4 时，MIoU 值达到最优值 87.09%；此外，选用递增的膨胀系数并不能提升精度.

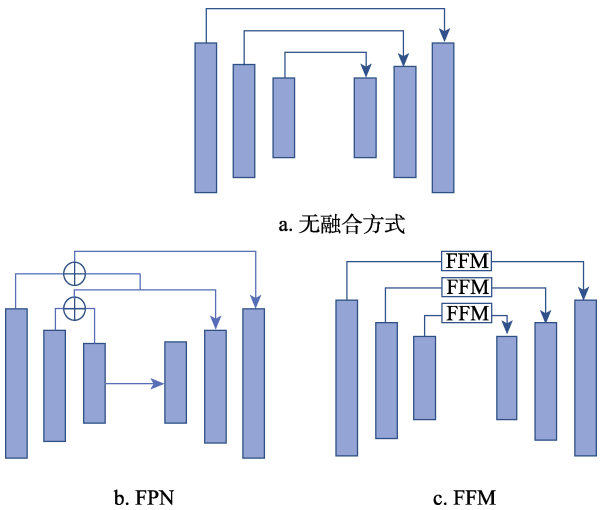
表 5　不同膨胀系数的分割指标对比　%

膨胀系数	PA	MPA	FWIoU	MIoU
(2,2,2)	**99.07**	91.95	98.19	86.87
(4,4,4)	**99.07**	**92.60**	**98.22**	**87.09**
(2,4,6)	99.02	91.51	98.11	85.46

注. 粗体表示最优值.

3.1.2　FFM 有效性的验证

选取图 8 中 3 种不同的融合结构对 FFM 的有效性进行验证. 图 8b 所示为特征金字塔(feature pyramid networks, FPN)[20]融合方式，具体过程如下：经过上采样的深层特征直接与浅层特征进行相加，结果作为输入，继续与上层特征进行融合，直到融合所有特征层为止. 3 种特征融合结构的实验结果如表 6 所示. 可以看出，未采用特征融合的方法在 4 个评价指标中表现最差，而采用 FPN 和本文特征融合方法则均有提升，且本文特征融合方法提升更多，说明直接融合的方法忽视了特征间的相异性，难以得到最佳表现结果，而包含更多语义信息的低级特征与高级特征融合相对容易，验证了 FFM 的有效性.



a. 无融合方式　b. FPN　c. FFM

图 8　3 种融合结构

表 6　不同特征融合结构下实验分割指标对比　%

特征融合结构	PA	MPA	FWIoU	MIoU
无融合方式	99.07	92.60	98.22	87.09
FPN[20]	**99.12**	92.63	98.23	87.22
FFM	99.11	**92.66**	**98.25**	**87.51**

注. 粗体表示最优值.

3.1.3　CAMM 有效性的验证

为了验证 CAMM 的有效性，在 U-Net 解码端引入 3 种通道注意力机制：压缩和激励模块(squeeze and excitation, SE)[21]，DANet[13]的通道注意力模块(channel attention module, CAM)和 PAN[14]的全局注意力上采样模块(global attention upsample, GAU). 将未添加注意力模块的实验结果作为参照实验，实验结果如表 7 所示. 可以看出，添加了 SE 和 CAM 的网络在 MIoU 精度上分别下降 1.21 和 7.47 个百分点，而采用 GAU 和 CAMM 的分割性能提升了 0.11 和 1.37 个百分点，说明注意力机制在皮革缺陷的分割结果中呈不稳定性，过强或弱地关注图像某一区域都会导致精度的下降；CAMM 在 MIoU 上达到 88.88%，明显优于未添加注意力机制的网络，说明该模块能够使网络关注缺陷的重要像素特征，有效地提取重要信息.

表 7　不同通道注意力模块的实验分割指标对比　%

注意力模块	PA	MPA	FWIoU	MIoU
未添加注意力模块	99.11	92.66	98.25	87.51
SE[21]	99.04	92.72	98.18	86.30
CAM[13]	98.77	88.38	97.72	80.04
GAU[14]	99.13	92.56	98.26	87.62
CAMM	**99.17**	**93.27**	**98.39**	**88.88**

注. 粗体表示最优值.

3.2　定性分析

为了直观地分析本文方法在皮革缺陷分割上的有效性，选取 FCN[7]，Deeplabv3+[22]，ExFuse[11]，GCN[10]，DANet[13]和 U-Net[8]这 6 种在语义分割领

域中表现优异的方法作为对比方法. 针对皮革缺陷分布, 分别对包含单类缺陷和多类缺陷的图像进行分析.

3.2.1 包含单类缺陷的分割结果

图 9 所示为 7 种方法在 7 种缺陷类型的分割结果, 从上至下依次为刺刮伤、刺猴、皮肤癣、破洞、烙印、烂面和开创伤, 颜色表示具体的缺陷类型. 可以看出, 对于皮肤癣类缺陷, 损伤程度造成其局部纹理存在特征差异, 导致 FCN, Deeplabv3+, ExFuse, GCN 和 U-Net 将其错分为烙印, 本文方法对该缺陷中的大部分像素能够正确分类, 验证了该方法对缺陷类内差异性的自适应能力; 开创伤附近的粗糙背景纹理存在与缺陷特征相似的特点, 容易造成错分割现象, 在处理这类场景时, 本文方法表现出较强的抗干扰能力; 由于刺刮伤、刺猴、烙印等缺陷的边缘部分损伤程度较轻, 细节难以提取, 导致 Deeplabv3+, ExFuse, GCN, DANet 和 U-Net 在分割刺猴的边界时均出现粘连现象, 刺刮伤也有类似情况, 而本文方法在边界的分割中较为精细, 在细节提取上表现了更优秀的能力.

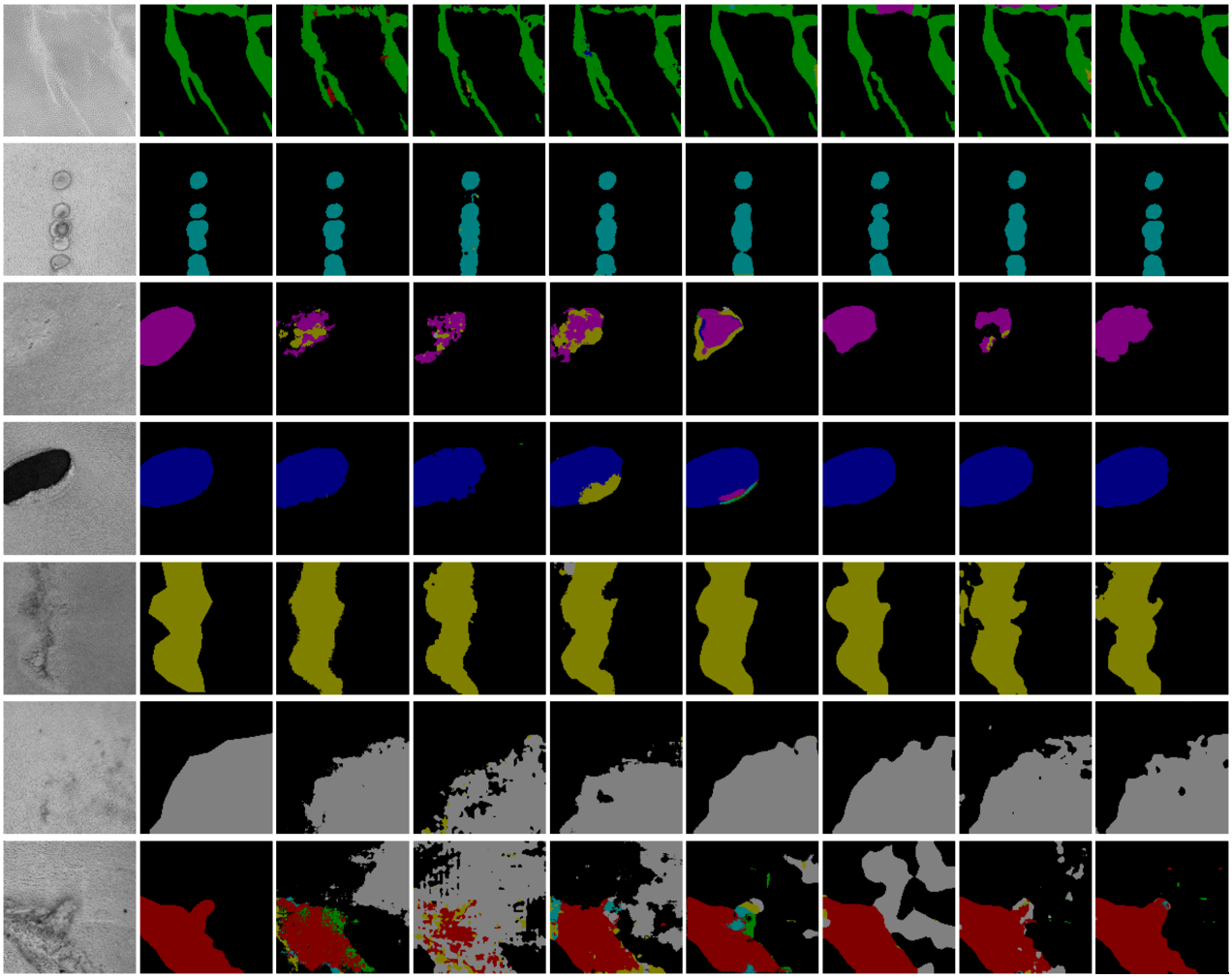
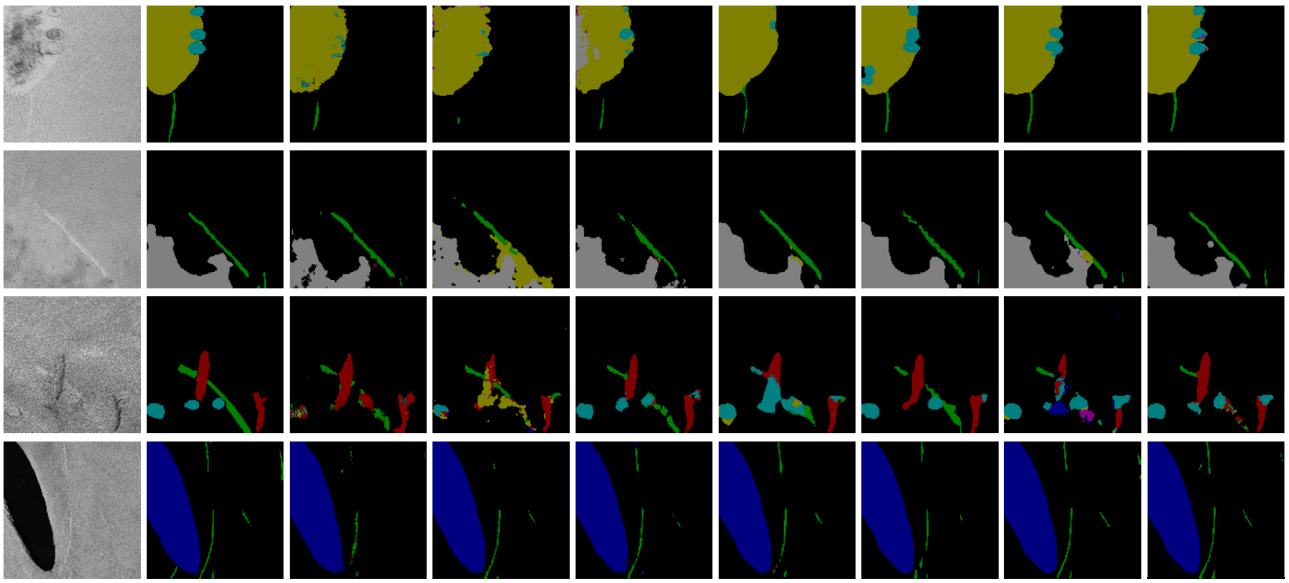


图 9 包含单类缺陷的分割结果

实验结果表明, 本文方法在包含部分差异的同一缺陷、具有相似特征的不同类型以及边缘处理中展示了更佳的分割效果, 验证了该方法具有更强的细节提取能力和识别缺陷的能力.

3.2.2 包含多类缺陷的分割结果

图 10 所示为 7 种方法在包含多种缺陷类型图像中的分割结果. 可以看出, 第 1 行中, 刺猴的位置距烙印边缘较近, 且烙印受损程度较重, 表现出和部分烙印纹理的相似性, 造成 FCN, Deeplabv3+ 和 GCN 几乎将刺猴与烙印融为一体, 表明具有局部相似性的不同缺陷极易造成像素错分, 与其他方法相比, 本文方法识别结果较为精细, 表现出对类间相似缺陷的较强区分能力; 第 3 行包含开创伤、刺刮伤和刺猴等尺寸不一的缺陷, 本文方法对



a. 原图　b. 标签　c. FCN[7]　d. Deeplabv3+[22]　e. ExFuse[11]　f. GCN[10]　g. DANet[13]　h. U-Net[8]　i. 本文

图 10　混合缺陷的分割结果

图像大部分像素都能正确识别，对于其他方法出现的较为严重的错分、漏分现象得到明显缓解，说明本文方法在多尺度分割中的有效性；对于具有丰富纹理的刺刮伤、损伤程度较轻的缺陷类型效果普遍较差，而本文方法基本能够正确地分割，说明该方法能有效地保留图像的空间信息，还原缺陷原始分辨率.

实验结果表明，本文方法对于包含相似信息的不同缺陷、多尺度缺陷和微小细节的分割效果更加精细，再次验证该方法对精细化和自适应缺陷特征提取的有效性.

3.3　定量分析

从分割精度和速度 2 个维度对 7 种语义分割方法进行定量比较，采用 PA, MPA, FWIoU 和 MIoU 作为精度指标，模型参数量和检测速度作为效率的衡量依据，结果如表 8 所示. 其中, FCN 和 U-Net 是语义分割编码-解码结构的代表；Deeplabv3+和 GCN 是将大感受野应用于多尺度分割的代表；ExFuse 是特征融合技术的代表；DANet 是注意力机制技术的代表. 可以看出，本文方法在 PA, MPA, FWIoU 和 MIoU 上达到最优，分别为 99.17%, 93.27%, 98.39%和 88.88%, 其中，PA 和 MPA 均超过 90%, 表明大部分像素点都能被正确分割；在检测效率方面, FCN 的模型参数量和检测速度分别为 19.12 MB 和 37.42 帧/s, 在所有方法中表现最佳，但由于其粗糙的上采样结构，导致分割精度低于其他方法；与基准方法 U-Net 相比，本文方法在更小参数量和更快检测速度的前提下，MIoU 提高了 4.03 个百分点，说明该方法能显著地改善 U-Net 上采样不精细、语义鸿沟等问题；与 Deeplabv3+, ExFuse, GCN 和 DANet 相比，本文方法在 MIoU 上分别提升 13.75, 9.92, 7.35 和 5.16 个百分点，说明在多尺度分割、缺陷的自适应等方面优于对比方法；虽然 Deeplabv3+, GCN 和 DANet 在推理速度上略高于本文方法，但复杂的模型结构导致参数量较大，且在 PA, MPA, FWIoU 和 MIoU 上的分割结果较差，无法满足精细化分割的要求.

表 8　7 种方法的定量结果对比

方法	PA/%	MPA/%	FWIoU/%	MIoU/%	模型参数量/MB	检测速度/(帧·s^{-1})
FCN[7]	98.23	82.41	96.67	75.06	**19.12**	**37.42**
Deeplabv3+[22]	98.03	81.22	96.33	75.13	54.61	26.61
ExFuse[11]	98.32	85.69	96.82	78.96	110.23	23.30
GCN[10]	98.87	88.98	97.86	81.53	59.08	29.19
DANet[13]	98.83	89.51	97.77	83.72	66.43	26.91
U-Net[8]	98.89	90.49	97.86	84.85	31.38	19.40
本文	**99.17**	**93.27**	**98.39**	**88.88**	24.09	25.53

注. 粗体表示最优值.

实验结果表明，本文方法可以在保证较低模型参数和计算复杂度的同时，有效地提升皮革缺陷数据集的精细化检测结果，比对比方法更加符合实际工业缺陷分割的检测需求.

4 结 语

本文从工业缺陷实际场景出发，分析导致皮革缺陷分割面临的难点问题，针对其丰富纹理特性和局部差异性导致的错误分割、边缘模糊等问题，提出改进 U-Net 结构的皮革表面缺陷分割方法. 首先，采取 2 种策略提升细节提取：一方面，嵌入 CDCM 增大感受野，获取全局信息；另一方面，添加 FFM 消除语义鸿沟，缓解局部信息缺失. 其次，使用通道注意力机制改善局部差异性，让网络自适应地关注缺陷相关通道，聚焦缺陷细节形态. 实验结果表明，本文方法对于形态多变、局部相似程度高的缺陷分割效果良好，MIoU 值较基准方法提升了 4.03 个百分点，表明本文方法在精细化分割中的有效性. 但是，由于本文方法采用对称的编码-解码结构，推理速度较慢，因此，轻量化模型结构并将这种分割框架应用于实际工业生产任务，是未来工作的重点.

参考文献(References):

[1] Aslam M, Khan T M, Naqvi S S, *et al*. On the application of automated machine vision for leather defect inspection and grading: a survey[J]. IEEE Access, 2019, 7: 176065-176086

[2] Tao Xian, Hou Wei, Xu De. A survey of surface defect detection methods based on deep learning[J]. Acta Automatica Sinica, 2021, 47(5): 1017-1034(in Chinese)
(陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, 47(5): 1017-1034)

[3] Kumar M P, Ashok S D. A multi-level colour thresholding based segmentation approach for improved identification of the defective region in leather surfaces[J]. Engineering Journal, 2020, 24(2): 101-108

[4] Kasi M K, Rao J B, Sahu V K. Identification of leather defects using an autoadaptive edge detection image processing algorithm[C] //Proceedings of the International Conference on High Performance Computing and Applications. Los Alamitos: IEEE Computer Society Press, 2014: 1-4

[5] Sobral J L. Leather inspection based on wavelets[C] //Proceedings of the 2nd Iberian Conference on Pattern Recognition and Image Analysis. Heidelberg: Springer, 2005: 682- 688

[6] He Fuqiang. Research and application of visual inspection technology for the large area of leather surface[D]. Hangzhou: Zhejiang University, 2012(in Chinese)
(贺福强. 大面积皮革表面的视觉检测技术与应用研究[D]. 杭州: 浙江大学, 2012)

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 3431-3440

[8] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] //Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241

[9] Chen L C, Papandreou G, Kokkinos I, *et al*. Semantic image segmentation with deep convolutional nets and fully connected CRFs[OL]. [2022-07-09]. https://arxiv.org/abs/1412.7062

[10] Peng C, Zhang X Y, Yu G, *et al*. Large kernel matters — improve semantic segmentation by global convolutional network[C] //Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1743-1751

[11] Zhang Z L, Zhang X Y, Peng C, *et al*. ExFuse: enhancing feature fusion for semantic segmentation[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 273-288

[12] Liu W, Rabinovich A, Berg A C. ParseNet: looking wider to see better[OL]. [2022-07-09]. https://arxiv.org/abs/1506.04579

[13] Fu J, Liu J, Tian H J, *et al*. Dual attention network for scene segmentation[C] ///Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 3141-3149

[14] Li H C, Xiong P F, An J, *et al*. Pyramid attention network for semantic segmentation[OL]. [2022-07-09]. https://arxiv.org/abs/1805.10180v3

[15] Liong S T, Gan Y S, Huang Y C, *et al*. Automatic defect segmentation on leather with deep learning[OL]. [2022-07-09]. https://arxiv.org/abs/1903.12139

[16] Liong S T, Zheng D N, Huang Y C, *et al*. Leather defect classification and segmentation using deep learning architecture[J]. International Journal of Computer Integrated Manufacturing, 2020, 33(10-11): 1105-1117

[17] Lin M, Chen Q, Yan S C. Network in network[OL]. [2022-07-09]. https://arxiv.org/abs/1312.4400v1

[18] Garcia-Garcia A, Orts-Escolano S, Oprea S, *et al*. A review on deep learning techniques applied to semantic segmentation[OL]. [2022-07-09]. https://arxiv.org/abs/1704.06857

[19] Deng J, Dong W, Socher R, *et al*. ImageNet: a large-scale hierarchical image database[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2009: 248-255

[20] Lin T Y, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection[C] //Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 936-944

[21] Hu J, Shen L, Albanie S, *et al*. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023

[22] Chen L C, Zhu Y K, Papandreou G, *et al*. Encoder-decoder with atrous separable convolution for semantic image segmentation[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 833-851