计算机科学
COMPUTER SCIENCE

基于双流卷积神经网络的稻米缺陷分割

吴一博, 郝应光, 王洪玉

引用本文

吴一博, 郝应光, 王洪玉. 基于双流卷积神经网络的稻米缺陷分割[J]. 计算机科学, 2024, 51(6A): 230600107-8.

WU Yibo, HAO Yingguang, WANG Hongyu. Rice Defect Segmentation Based on Dual-stream Convolutional Neural Networks [J]. Computer Science, 2024, 51(6A): 230600107-8.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230600107

基于双流卷积神经网络的稻米缺陷分割

吴一博　郝应光　王洪玉
大连理工大学信息与通信工程学院　辽宁 大连 116024
(wyb1302030@163.com)

摘　要　目前水稻质量精细化评估因为没有水稻缺陷精细化检测相关工作而无法实现,传统的水稻质量评估都是基于粗略的缺陷有无分类而实现的。针对水稻缺陷像素级分类问题,提出了一种基于深度学习的水稻缺陷分割模型,该模型使用了一个改进的 DoubleU-Net 网络作为主要架构,分为 NETWORK1 和 NETWORK2 两部分,其中 NETWORK1 是基于 VGG-19 修改的 U 型网络结构,而 NETWORK2 是基于 Swin Transformer 修改的 U 型网络结构,将这两部分串联起来,同时融合 CNN 局部信息提取和 Transformer 全局信息提取的优势,可以更好地捕捉图像的上下文信息。同时,使用了多重损失函数,包括加权的二元交叉熵损失、加权的交并比损失和一个无需训练的智能损失网络,在提高模型训练稳定性的同时进一步提高了模型分割的精度。在制作的密集水稻缺陷数据集上进行训练测试,该模型均取得了较其他方法更好的分割性能,具有鲁棒性和较好的泛化能力。

关键词:稻米质量评估;语义分割;深度学习;卷积神经网络;Transformer
中图分类号　TP181

Rice Defect Segmentation Based on Dual-stream Convolutional Neural Networks

WU Yibo,HAO Yingguang and WANG Hongyu
Department of Information and Communication,Dalian University of Technology,Dalian,Liaoning 116024,China

Abstract　Currently,fine-grained assessment of rice quality cannot be achieved due to the lack of related work on fine-grained detection of rice defects. Traditional rice quality assessment is based on rough classification of defect presence or absence. To address the problem of pixel-level classification of rice defects,a deep learning-based rice defect segmentation model is proposed. The model uses an improved DoubleU-Net network as the main architecture,which consists of two parts,NETWORK1 and NETWORK2. NETWORK1 is based on a modified U-shaped network structure of VGG-19,while NETWORK2 is based on a modified U-shaped network structure of Swin Transformer. The two parts are concatenated,and the advantages of CNN local information extraction and Transformer global information extraction are integrated to better capture the contextual information of images. In addition,multiple loss functions are used,including weighted binary cross-entropy loss,weighted intersection-over-union loss,and an intelligent loss network that does not require training,to improve the stability of the model training and further improve the accuracy of model segmentation. The proposed model is trained and tested on a densely annotated rice defect dataset,and achieves better segmentation performance than other methods,with robustness and good generalization ability.

Keywords　Rice quality assessment,Semantic segmentation,Deep learning,Convolutional neural network,Transformer

1　引言

稻米是世界上最重要的粮食作物之一[1],超过 1/3 的世界人口将其视为主食,在全球范围内广泛消费和交易。然而,稻米的市场价值受其外观质量的影响很大。其中一个关键问题是稻米质量不均匀,存在各种缺陷,如发黄、变质等。这些缺陷会影响稻米的外观、口感和营养品质等,降低稻米的市场竞争力和价值。因此,需要通过稻米的质量分级进行品质把控[2]。现有的稻米分级方法主要依靠操作人员的肉眼检查,耗时且主观,分类结果极易受操作人员的经验和状态影响[3]。随着计算机科学、机械和自动化工程的发展,基于机器学习的图像处理方法在食品质量评估方面存在巨大的应用空间[4]。其中在图像处理方面,语义分割技术在许多领域中都得到了广泛的应用,使用深度学习的方法可以快速准确地分割出想要的结果,因此可以通过这种方法在稻米质量分级中更快、更客观地帮助人类。

目前的一些研究使用传统机器学习的方法来检测稻米外观的质量。Wan 等[5]设计了一种自动检测系统,通过提取形

基金项目:中央高校基本科研业务费专项基金(DUT21YG110)
This work was supported by the Fundamental Research Funds for the Central Universities of Ministry of Education of China(DUT21YG110).
通信作者:郝应光(yghao@dlut.edu.cn)

状轮廓、颜色和背光特征来快速分类稻米质量。但该方法对不同质量类别的识别精度差异较大，对于某些质量类别，如不成熟、开裂和破损等，检测准确率相当低。Payman 等[6]开发了一种基于专家视觉的系统来提取稻米的颜色和形状特征，并设置了一系列阈值来检测破碎、红斑和黑斑稻粒。Kaur 和 Singh[7]使用多类 SVM(支持向量机)来识别白垩和破碎稻米，但识别准确率并不高。Zareiforoush 等[8]分别利用灰度共生矩阵和米粒长度特征计算稻米的碾磨度(DOM)和破粒百分比(PBK)，并基于模糊逻辑输出稻米品质等级。然而，光照系统的噪声和稻米色度的不均匀性对算法的精度影响较大，导致算法的鲁棒性较差。Chen 等[9]提出了一种基于红外图像的方法对稻米进行快速质量检测。该方法也只是简单地将不同色度的稻米进行了分类，并不能准确分割出稻米的缺陷部分。

但上述方法均是针对少量分散稻米的检测手段，检测的稻米均是清晰的、分散的，并且背景十分简单，没有什么噪声干扰。而在实际工业应用上，稻米的样本并没有那么稀少分散，并且得到的稻米数据的图片也并不会很高清，因此上述方法并不能很好地适用于工业上的稻米缺陷检测分割。并且，对于稻米质量的精细化分类也需要得到稻米缺陷的像素级分类结果，而不是简单的分类。然而像素级的分类需要大量数据集来支撑训练，而在实际的工作中并不能提供足够数量的训练数据。针对这个问题，Ronneberger 等[10]提出了经典的 U 型网络 U-Net，其由一个带有跳跃连接的对称编码器-解码器组成。通过这种优雅的结构设计，U-Net 被应用于许多轮廓不规则的小目标分割任务中，并取得了巨大成功。按照这一技术路线，许多算法如 3D U-Net[11]，Res-UNet[12]，U-Net++[13]和 UNet3+[14]已经被开发用于各种类型的图像分割。上述方法均基于 CNN，虽然取得了优异的性能，但仍不能完全满足稻米缺陷分割的精度要求。由于卷积运算的固有局域性，基于 CNN 的方法很难学习到全局的语义信息[15]。一些研究试图通过使用空洞卷积[16]、自注意力机制[17]和图像金字塔[18]来解决这个问题。但是，这些方法在建立长距离依赖关系时仍有局限性。最近，受 Transformer 在自然语言处理(NLP)领域[19]的巨大成功的启发，研究人员将 Transformer 引入视觉领域[20]，Dosovitskiy 等[21]提出了 Vision Transformer(ViT)来执行图像识别任务。ViT 以具有位置嵌入的二维图像块作为输入，在大数据集上进行预训练，性能与基于 CNN 的方法相当。此外，Touvron 等[22]提出了高效处理数据的 Transformer 模型 DeiT，这表明 Transformer 可以在中等规模的数据集上进行训练，并且，将其与蒸馏方法相结合可以获得鲁棒性更好的 Transformer 结构。Liu 等[23]开发了一个基于窗口多头注意力结构的 Swin Transformer 模型，以 Swin Transformer 为骨干网络在图像分类、目标检测和语义分割等方面取得了先进的性能。ViT，DeiT 和 Swin Transformer 在图像识别任务中的成功展示了 Transformer 在视觉领域应用的潜力。

基于 Transformer 在视觉领域的巨大成功，许多研究人员提出了将 Transformer 和 U-Net 结合的方法。Chen 等[15]提出了结合 Transformer 和 CNN 的 TransUnet 模型，该方法同时具有两者的优点。Valanarasu 等[24]提出的 MedT 模型基于一种局部-全局的训练策略(LoGo)，进一步提高了性能，Cao 等[25]提出的 Swin-Unet 是第一个纯基于 Transformer 的 U-Net 架构，取得了较好的结果。

尽管 Swin-Unet 的分割精度已经足够优异，但是仍无法满足稻米缺陷分割这一任务。因此本文提出了一种新的 CNN 和 Transformer 结合的方法，可以准确高效地进行稻米缺陷的像素级语义预测，并且针对密集粘连的稻米数据，也可以做到很好的缺陷分割。本文提出的模型基于 DoubleU-net[26]的网络框架，该网络架构由两个 U-Net 结构串联而成，而本文使用的两个 U 型结构分别为一个基于 CNN 的 VGG-19[27]，以及一个基于 Transformer 的 Swin Transformer[23]，之后再通过特殊的解码器将 CNN 和 Transformer 相结合。本文的创新点在于：

1)针对稻米缺陷分割这一问题，提出了一种新的结合 CNN 和 Transformer 优点的双 U 型架构。CNN 在卷积运算中擅长提取局部特征，但在捕获全局特征表示方面还有一定的局限性。而 Transformer 则可以捕获全局特征，但是由于其更多的是针对全局的信息，对于图像的一些局部细节信息关注度并不高。这个架构可以充分利用 CNN 的局部性和平移不变性以及 Transformer 的全局性，从而提高模型对稻米缺陷分割的精确性。

2)针对稻米缺陷这种边缘轮廓不规则的特征，结合了多种 Loss 函数，其中包括加权的 BCE Loss、加权的 IoU Loss 和一种无需训练的 LossNet 智能损失函数，能够从边缘轮廓细节优化特征图的分割结果，从而提高最终的分割精度。

3)本文针对精细化的稻米质量分级，制作了相应的像素级分类预测的稻米缺陷数据集，并取得了较好的分割结果。

2 本文方法

本文提出了一个基于深度学习的稻米缺陷分割模型，整体模型由两个 U 型结构组成。第一个 U 型结构的编码器使用一个在 ImageNet[28]上训练过的 VGG-19[27]作为特征提取的主干网络，并且在后面加了一个空洞空间卷积池化金字塔层(ASPP)[29]，可以提取多尺度特征从而更好地捕获上下文信息。解码器则是通过跳跃连接将每层编码器得到的特征图与通过上采样得到的特征图拼接起来，这样可以更好地用到浅层特征和深层语义特征。第二个 U 型结构的编码器使用的是多层 Swin Transformer[23]作为特征提取的主干网络，而解码器则是同时融合了第一个 U 型结构中 VGG-19 所提取的特征图、第二个 U 型结构中 Swin Transformer 提取的特征图，以及通过上采样得到的特征图这 3 种特征，所有特征信息通过一个像素级的分类器，生成最终的稻米缺陷分割结果，整体网络框架如图 1 所示。

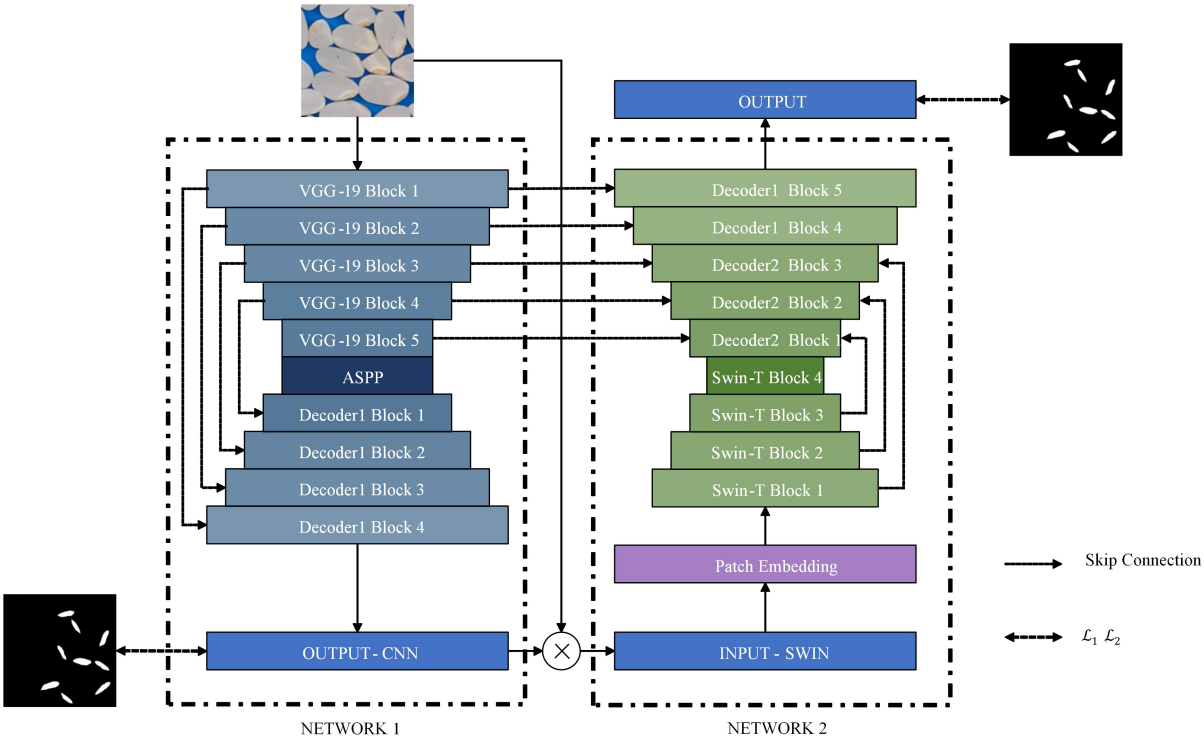


图 1　整体网络框架图

Fig. 1　Overall network block diagram

2.1　NETWORK1

NETWORK1 是一个基于纯 CNN 的 U 型网络结构,分为编码器和解码器两个部分,并通过跳跃连接将两部分结合起来。整体结构如图 2 所示。

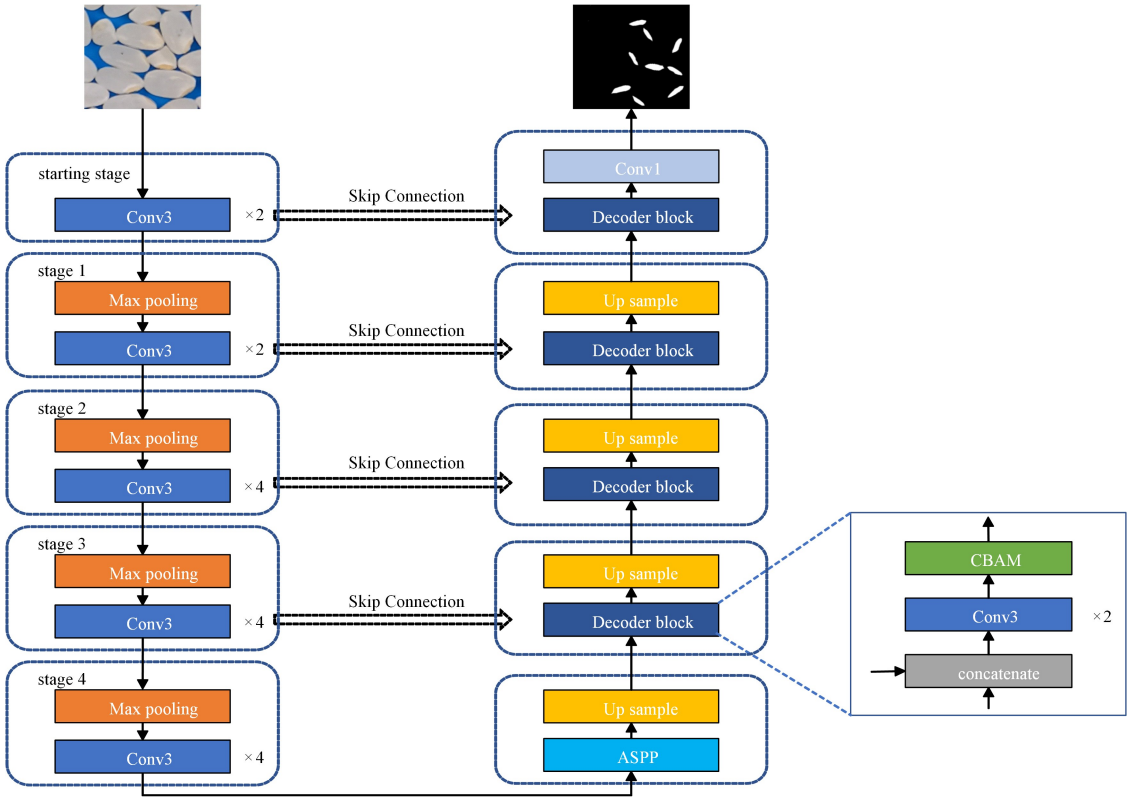


图 2　网络 1 框架图

Fig. 2　Network1 block diagram

NETWORK1 的编码器使用的是一个在 ImageNet 上预训练过的轻量化 VGG-19,这部分总共分为五阶段来进行不同尺寸和深度的特征提取,这 5 个阶段可以表示为 starting stage,stage1,stage2,stage3,stage4,每层依次进行下采样操作得到不同尺寸不同深度的特征图。为了更好地获取多尺度的上下文信息,在 VGG-19 的最后一层之后添加了一个空洞空间卷积池化金字塔模块(ASPP)。该模块使用具有不同采样率的多个并行空洞卷积层,并将每个采样层提取的特征在

单独的分支中进一步处理，最后融合每个采样层生成最终结果。该模块通过不同的空洞率构建不同尺寸感受野的卷积核，用来获取多尺度特征信息。

NETWORK1 的解码器分为 4 个阶段，每个阶段都要对输入的特征进行 2 倍的双线性上采样，使输入特征图的大小加倍。这样在空间维度上就满足和对应编码器生成的特征图跳跃连接的条件，之后再将上采样后得到的特征图与编码器中相同尺寸的特征图相融合，增强网络学习的语义信息。将融合后的特征执行两个 3×3 卷积运算，每个卷积运算之后都要加一个批量归一化和一个 ReLU 激活函数。后面还添加了一个 CBAM 模块，这是一个轻量级的卷积神经网络注意力模块，该模块同时结合了通道注意力机制和空间注意力机制，通过引入注意力机制可以使网络更准确地聚焦于目标对象，抑制无关背景的干扰，进而提高模型的性能。经过 4 个阶段的解码器之后，通过一个 1×1 卷积运算，进行稻米缺陷的像素级预测。

2.2 NETWORK2

NETWORK 2 是一个基于 Swin Transformer[23] 改进的 U 型网络结构，同样分为编码器和解码器两个部分，其中解码器分为两种：一种和 NETWORK 1 中的解码器相同；另一种使用了两个跳跃连接，同时结合了 NETWORK1 和 NETWORK2 的编码器所提取的特征。整体结构如图 3 所示。

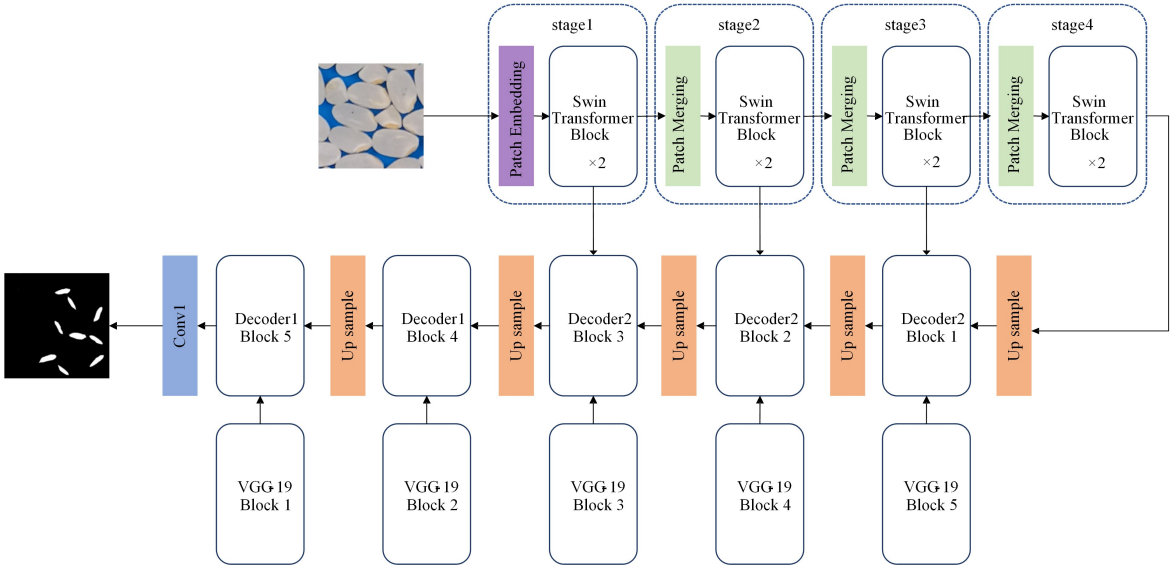


图 3 网络 2 框架图

Fig. 3 Network2 block diagram

NETWORK2 的编码器总共分为 4 部分，这 4 部分可以表示为 stage1，stage2，stage3，stage4。其中除了 stage1 是先通过一个 Patch Embedding 层之后，再经过两个 Swin Transformer Block，其余 3 个 stage 都是先通过一个 Patch Merging 层进行下采样，然后再通过 2 个 Swin Transformer Block。Patch Embedding 分为 Patch Partition 和 Linear Embedding 两部分，输入的图片首先在 Patch Partition 模块中进行分块，即将每 4×4 的相邻像素分为一个 Patch，并将其在通道维度进行展平。然后通过 Linear Embedding 模块对每个 Patch 在通道维度进行线性映射。通过这两个模块后，输入图像的长宽都将变为原来长宽的 1/4，通道数也将从原来 RGB 图像的 3 通道，变为设定好的通道数 C。而每个 stage 中叠加的 Swin Transformer Block[23]（如图 4 所示）与传统的多头自注意力（MSA）[19] 模块不同的是，Swin Transformer Block 是由两个基于滑动窗口的 Transformer 模块串联而成的，每个模块都是由归一化层 LayerNorm、多头注意力模块、残差连接以及 MLP 层构成。这两个模块的不同之处仅在于一个使用了基于窗口的多头注意力（W-MSA）结构，一个使用了基于滑动窗口的多头注意力（SW-MSA）结构。传统的 MSA 计算的是全局注意力，所有的像素都需要进行计算，随着输入图像尺寸的增大，计算量也将飞速上涨。通过使用 W-MSA 模块，将自注意力的计算限制在一个窗口之中，可以极大降低计算的复杂度，提高训练速度。但是 W-MSA 模块中的每个窗口是互相不重叠的，所以窗口与窗口之间是无法进行信息传递的。为了解决这个问题，Liu 等[23] 引入了 SW-MSA 模块，通过滑动窗口来实现不同窗口之间的信息交流。这样既降低了计算复杂度，又同时兼顾了全局信息的提取能力。MSA 复杂度和 W-MAS 复杂度的计算公式如式(1)、式(2)所示：

$$\Omega(MSA)=4HWC^2+2(HW)^2 \tag{1}$$

$$\Omega(W\text{-}MSA)=4HWC^2+2M^2HWC \tag{2}$$

其中，H 为图片的高度；W 为图片的宽度；C 为图片的深度；M 为窗口的大小。

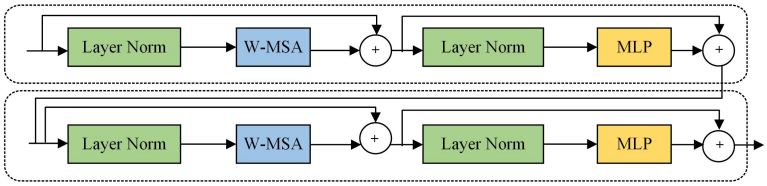


图 4 Swin transformer 框架图

Fig. 4 Swin transformer block diagram

在后 3 个 stage 中都会经过 Patch Merging 层进行下采样,经过 Patch Merging 层后,特征图的高和宽会减半,通道深度则会翻倍。

NETWORK2 的解码器分为 5 个阶段,每个阶段都要对输入的特征进行 2 倍的双线性上采样,使输入特征图的大小加倍。前 3 个阶段使用的是 Decoder2 Block,后两个阶段使用的是 Decoder1 Block。Decoder2 Block 同时融合了通过上采样得到的特征图,以及 NETWORK2 相应编码器生成的特征图和 NETWORK1 相应编码器生成的特征图这 3 个特征。而 Decoder1 Block 则与 NETWORK1 中的解码器结构相同,融合的只是上采样得到的特征图和 NETWORK1 编码器生成的特征图。通过这 5 个阶段,图像恢复到了原图的大小,最后再通过一个 1×1 的卷积层进行线性映射,输出像素级的预测结果。

2.3 损失函数

本文整体的损失函数可以表示为式(3):

$$\mathcal{L}=\alpha\mathcal{L}_1+\beta\mathcal{L}_2 \tag{3}$$

其中,α 和 β 都表示一个指定的超参数,本文将 α 设为 0.3,β 设为 0.7;$\mathcal{L}_1$ 和 $\mathcal{L}_2$ 分别表示 NETWORK1 和 NETWORK2 的损失函数。$\mathcal{L}_2$ 计算的是 NETWORK2 输出的像素级的预测结果 P_2 和真值图像 G 之间的损失,具体表示为式(4):

$$\mathcal{L}_2=\mathcal{L}^{w}_{\mathrm{IoU}}(P_2,G)+\mathcal{L}^{w}_{\mathrm{BCE}}(P_2,G)+L_f \tag{4}$$

$\mathcal{L}_1$ 计算的是 NETWORK1 输出的像素级预测结果 P_1 和真值图像 G 之间的损失,具体表示为式(5):

$$\mathcal{L}_1=\mathcal{L}^{w}_{\mathrm{IoU}}(P_1,G)+\mathcal{L}^{w}_{\mathrm{BCE}}(P_1,G)+L_f \tag{5}$$

其中,$\mathcal{L}^{w}_{\mathrm{IoU}}(\cdot)$和 $\mathcal{L}^{w}_{\mathrm{BCE}}(\cdot)$分别是加权交并比(IoU)损失[30]和加权二元交叉熵(BCE)损失[30]。IoU 损失用于计算模型预测结果与真实标签之间的差距,即像素级别的重叠程度。BCE 损失通常用于二元语义分割任务,其中每个像素的标签只有两个可能的取值。BCE 损失计算每个像素点的预测结果与真实标签之间的交叉熵,将这些交叉熵的均值作为损失值。但是传统的 IoU 损失和 BCE 损失在面对数据集中两种类别的样本数量差异较大时,计算结果可能会被影响。针对稻米缺陷分割这个问题来说,背景区域远大于缺陷区域,传统的 IoU 损失和 BCE 损失并不合适,而加权 IoU 损失和加权 BCE 损失通过为每个样本分配不同的权重来解决这个问题,可以为较少的类别样本赋予更高的权重,以平衡两个类别的影响。故加权 IoU 损失和加权 BCE 损失比较适合这种小目标的分割任务。

此外,针对小目标而言,边缘细节信息是十分重要的,因此额外添加了一个 LossNet 来进行细节信息的优化。具体来讲就是使用一个经过 ImageNet 预训练的分类网络 VGG-16,分别提取预测图和真值图的多尺度特征,然后计算相同尺度特征之间的欧几里得距离,进行像素级别的监督。计算过程如式(6)、式(7)所示:

$$\mathcal{L}_f=l_f^1+l_f^2+l_f^3+l_f^4 \tag{6}$$

$$l_f^i=\|F_P^i-F_G^i\|_2,i=1,2,3,4 \tag{7}$$

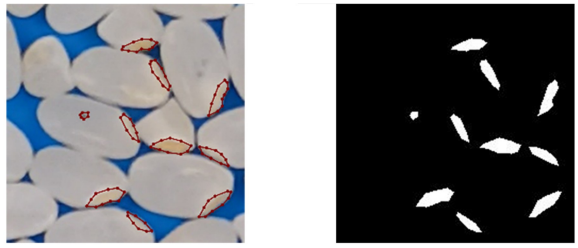
其中,F_P^i 和 F_G^i 分别表示预测图和真值图通过 LossNet 得到的第 i 级特征图。

3 实验结果及分析

3.1 实验配置

3.1.1 数据集

由于此前并没有精细化的稻米缺陷分割的相关工作,因此需要自行采集并制作数据集。由于实际工业应用上,稻米样本比较密集,采集到的图像也不会像实验室中使用高清显微镜拍摄,因此将稻米密集地铺洒在蓝色的托盘上,并通过普通相机进行拍照采集,所拍照片为三通道的 RGB 图像。考虑到图像尺寸过大会导致模型难以训练,将采集到的图像裁剪到长宽均为 224 的 RGB 图像。针对裁剪后的数据,使用麻省理工的开源软件 Labelme 对稻米的缺陷样本进行像素级标注。缺陷区域标记如图 5 所示。



(a)标注图　　(b)真值图

图 5　数据集示例图

Fig. 5　Sample graph of dataset

3.1.2 训练环境及训练参数

本文模型代码基于 Pytorch1.8 的框架实现,在 Ubuntu18.04 系统中使用 python 语言进行实验。训练时初始学习率设为 0.05,batch size 设为 24,训练 1 500 个 epoch。在这种设置下使用 GeForce GTX3090Ti GPU 完成加速训练。

3.2 性能指标

混淆矩阵是一种常用的分类评估工具,用于衡量模型对样本分类的准确性。混淆矩阵是一个二维矩阵,其中行代表真实标签,列代表预测标签。混淆矩阵的元素包括:True Positive(TP),即实际为正样本,预测为正样本的数量;False Positive(FP),即实际为负样本,预测为正样本的数量;True Negative(TN),即实际为负样本,预测为负样本的数量;False Negative(FN),即实际为正样本,预测为负样本的数量。

3.2.1 平均交并比 MIoU

计算每个类别的交并比(IoU)得分,然后对所有类别的 IoU 得分求平均,即 MIoU。IoU 用于计算模型预测结果与真实标签之间的重叠程度。具体来说,对于每个类别,IoU 的计算公式如下:

$$IoU=\frac{TP}{(TP+FP+FN)} \tag{8}$$

而对于二元分割来说 MIoU 和 IoU 是等价的。

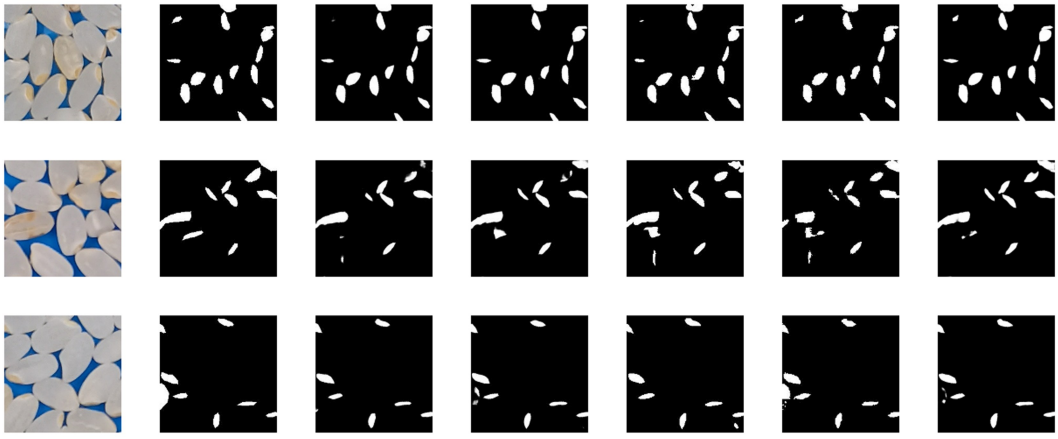
3.2.2 Dice 系数

Dice 指标是计算预测结果和真实结果之间的重叠面积与两者之和的比值。具体的计算公式如下:

$$Dice=\frac{2\times TP}{(2\times TP+FP+FN)} \tag{9}$$

3.3 模型测试结果

为了验证本文提出模型的有效性，将其与目前较好的分割方法进行了比较，包括 U-Net[10]，DoubleU-net[26]，TransU-net[15]，Swin-Unet[25]。图 6 依次展示了原图、真值图以及不同模型的预测结果。表 1 列出了不同模型测试结果的性能指标对比。



(a)原图 (b)真值 (c)U-net (d)DoubleU-net (e)TransUnet (f)Swin-Unet (g)本文方法

图 6 测试结果

Fig. 6 Test results

表 1 不同模型的性能指标

Table 1 Performance indicators of different models

模型	*MIoU*	*Dice*
U-Net	0.681	0.802
TransUnet	0.653	0.790
Swin-Unet	0.661	0.792
DoubleU-net	0.690	0.810
本文方法	**0.701**	**0.821**

3.4 消融实验

为了验证基于 DoubleU-net 框架来融合 CNN 和 Transformer、加入 LossNet 和在解码器上加入 CBAM 模块的有效性，本文分别做了 3 组消融实验，分别表示为消融实验 1、消融实验 2 和消融实验 3，具体结果如表 2—表 4 所列。

表 2 消融实验 1

Table 2 Ablation experiment 1

	MIoU	*Dice*
U-net	0.681	0.802
DoubleU-net	0.690	0.810
DoubleU-net-swin	**0.701**	**0.821**

表 3 消融实验 2

Table 3 Ablation experiment 2

	MIoU	*Dice*
$\mathcal{L}_{\mathrm{IoU}}^{w}(\cdot)+\mathcal{L}_{\mathrm{BCE}}^{w}(\cdot)$	0.691	0.819
$\mathcal{L}_{\mathrm{IoU}}^{w}(\cdot)+\mathcal{L}_{\mathrm{BCE}}^{w}(\cdot)+\mathcal{L}_{f}$	**0.701**	**0.821**

表 4 消融实验 3

Table 4 Ablation experiment 3

	MIoU	*Dice*
Decoder	0.697	0.820
Decoder +CBAM	**0.701**	**0.821**

表 2 列出了验证基于 DoubleU-net 框架来融合 CNN 和 Transformer 有效性的实验结果，在本组消融实验中设置 3 组对照实验，分别是基础 U-Net 网络结构、DoubleU-net 网络结构和本文设计的基于 DoubleU-net 框架来融合 CNN 和 Transformer 网络结构。其中 DoubleU-net 和本文方法的 NETWORK1 均使用 VGG-19 作为编码器，不同的是本文 NETWORK2 的编码器使用的是 Swin Transformer 架构。实验结果表明本文方法在 MIoU 和 Dice 两个性能指标上均获得了最好的结果，由此证明了本文网络框架设计的有效性。

表 3 列出了验证 LossNet 有效性的实验结果，图 7 展示了预测结果对比图。本组消融实验设置了两个对照组，分别是使用加权 IoU 损失和加权 BCE 损失，以及使用加权 IoU 损失、加权 BCE 损失和智能损失函数 LossNet。实验结果表明加入智能损失函数 LossNet 后，网络的分割性能得到提升，通过对比图可以发现，加入 LossNet 可以让模型更好地学习图像的边缘轮廓信息。

(a)真值图 (b)使用 LossNet (c)不使用 LossNet

图 7 LossNet 有效性对比图

Fig. 7 LossNet effectiveness comparison

表 4 列出了验证 CBAM 模块有效性的实验结果，实验结果表明加入 CBAM 模块后，网络的分割性能得到提升，CBAM 模块有效性得以证明。

结束语 稻米缺陷分割是一种密集小目标分割的实际应用，目标具有重叠、密集、像素占比较小等问题。本文针对这些难点，提出了一种基于深度学习的水稻缺陷分割模型。传统的基于 CNN 的深度学习模型擅长提取局部特征，但在捕获全局特征表示方面还有一定的局限性。而基于 Transformer 的方法可以捕获全局特征，但是由于其更多的是针对全局的信息，对于图像的一些局部细节信息关注度并不高。因此本文模型分为 NETWORK1 和 NETWORK2 两部分，同时结合了 CNN 的局部性和平移不变性以及 Transformer 的全局性。其中 NETWORK1 是基于 VGG-19 修改的 U 型网络结构，编码器部分使用经过 ImageNet 预训练的 VGG-19，分为 5 个阶段，分别提取了不同尺度不同深度的特征图。为

了更好地获取多尺度的上下文信息,在最后阶段的后面添加了一个 ASPP 模块,使用具有不同采样率的多个并行空洞卷积层来获取多尺度的特征信息。随后将编码器的输出送入解码器进行解码,解码分 4 个阶段,每个阶段都将输入的特征图进行 2 倍的双线性上采样,使输入特征图的大小加倍,同时通过跳跃连接与编码器生成的相同尺寸的特征图相融合来增强网络学习的语义信息。将融合后的特征执行两个 3×3 卷积运算,每个卷积运算之后都加一个批量归一化和一个 ReLU 激活函数。之后引入了一个注意力 CBAM 模块,结合了通道注意力机制及空间注意力机制,引入注意力机制可以使网络更准确地聚焦于目标对象,抑制无关背景的干扰,进而提高模型的性能。经过 4 个阶段的解码器之后,通过一个 1×1 卷积运算,进行像素级预测。将 NETWORK1 的输出和原图像相融合作为 NETWORK2 的输入,而 NETWORK2 的编码器使用多层 Swin Transformer 作为主干网络来提取不同尺寸的特征,并使用了两种不同的解码器进行解码,分为 5 个阶段,前三阶段和 NETWORK1 的解码器不同的是使用了两个跳跃连接,同时融合了通过上采样得到的特征图,以及 NETWORK1 相应编码器生成的特征图和 NETWORK2 相应编码器生成的特征图这 3 个特征,后两个阶段和 NETWORK1 的解码器相同,通过跳跃连接融合上采样得到的特征图和 NETWORK1 编码器生成的特征图。通过这 5 个阶段,图像恢复到了原图的大小,最后再通过一个 1×1 的卷积层进行线性映射,输出像素级的预测结果。训练时结合了多种 Loss 函数,其中包括加权的 BCE Loss、加权的 IoU Loss 和一种无需训练的 LossNet 智能损失函数,在提高模型训练稳定性的同时也提高了模型的分割精度,并在制作的稻米缺陷数据集上取得了较好的分割结果。本文提出的方法仍存在不足,当背景较为复杂时,噪声的干扰较大,分割精度会降低,后续将针对如何抑制背景噪声干扰进行相关研究。

参 考 文 献

[1] KIM M J,LIM J,KWON S W,et al. Geographical origin discrimination of white rice based on image pixel size using hyperspectral fluorescence imaging analysis [J]. Applied Sciences,2020,10(17):5794.

[2] DING C,LIU Q,LI P,et al. Distribution and quantitative analysis of phenolic compounds in fractions of Japonica andIndica rice [J]. Food Chemistry,2019,274:384-391.

[3] LIN Z,ZHANG X,WANG Z,et al. Metabolomic analysis of pathways related to rice grain chalkiness by a notched-belly mutant with high occurrence of white-belly grains[J]. BMC Plant Biology,2017,17:1-15.

[4] HARALICK R M,STERNBERG S R,ZHUANG X. Image analysis using mathematical morphology[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1987(4):532-550.

[5] WAN Y N,LIN C M,CHIOU J F. Rice quality classification using an automatic grain quality inspection system[J]. Transactions of the ASAE,2002,45(2):379.

[6] PAYMAN S H,BAKHSHIPOUR A,ZAREIFOROUSH H. Development of an expert vision-based system for inspecting rice quality indices[J]. Quality Assurance and Safety of Crops & Foods,2018,10(1):103-114.

[7] KAUR H,SINGH B. Classification and grading rice using multi-class SVM[J]. International Journal of Scientific and Research Publications,2013,3(4):1-5.

[8] ZAREIFOROUSH H,MINAEI S,ALIZADEH M R,et al. A hybrid intelligent approach based on computer vision and fuzzy logic for quality measurementof milled rice[J]. Measurement,2015,66:26-34.

[9] CHEN S,XIONG J,GUO W,et al. Colored rice quality inspection system using machine vision[J]. Journal of Cereal Science,2019,88:87-95.

[10] RONNEBERGERO,FISCHER P,BROX T. U-net:Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015:18th International Conference,Munich,Germany,October 5-9,2015,Proceedings,Part III 18. Springer International Publishing,2015:234-241.

[11] ÇIÇEK Ö,ABDULKADIR A,IENKAMP S S,et al. 3D U-Net:learning dense volumetric segmentation from sparse annotation [C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016:19th International Conference,Athens,Greece,October 17-21,2016,Proceedings,Part II 19. Springer International Publishing,2016:424-432.

[12] XIAO X,LIAN S,LUO Z,et al. Weighted res-unet for high-quality retina vessel segmentation[C]// 2018 9th International Conference on Information Technology in Medicine and Dducation(ITME). IEEE,2018:327-331.

[13] ZHOU Z,RAHMAN SIDDIQUEE M M,TAJBAKHSH N,et al. Unet++:A nested u-net architecture for medical image segmentation[C]// Deep Learning in Medical Image Analysis and Multi modal Learning for Clinical Decision Support:4th International Workshop, DLMIA 2018, and 8th International Workshop,ML-CDS 2018,Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer International Publishing,2018:3-11.

[14] HUANG H,LIN L,TONG R,et al. Unet 3+:A full-scale connected unet for medical image segmentation[C]// ICASSP 2020-2020 IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP). IEEE,2020:1055-1059.

[15] CHEN J,LU Y,YU Q,et al. Transunet:Transformers make strong encoders for medical image segmentation[J]. arXiv:2102.04306,2021.

[16] CHEN L C,PAPANDREOU G,KOKKINOS I,et al. Deeplab:Semantic image segmentation with deep convolutional nets,atrous convolution,and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,40(4):834-848.

[17] SCHLEMPER J,OKTAY O,SCHAAP M,et al. Attention gated networks:Learning to leverage salient regions in medical images[J]. Medical Image Analysis,2019,53:197-207.

[18] ZHAO H,SHI J,QI X,et al. Pyramid scene parsing network [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:2881-2890.

[19] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[J]. arXiv:1706.03762,2017.

[20] CARION N,MASSA F,SYNNAEVE G,et al. End-to-end object detection with transformers[C]// Computer Vision-ECCV

2020:16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I 16. Springer International Publishing, 2020:213-229.

[21] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv:2010.11929, 2020.

[22] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[C]// International Conference on Machine Learning. PMLR, 2021: 10347-10357.

[23] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012-10022.

[24] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: Gated axial-attention for medical image segmentation[C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27-October 1, 2021, Proceedings, Part I 24. Springer International Publishing, 2021: 36-46.

[25] CAO H, WANG Y, CHEN J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]// Computer Vision-ECCV 2022 Workshops: Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part III. Cham: Springer Nature Switzerland, 2023:205-218.

[26] JHA D, RIEGLER M A, JOHANSEN D, et al. Doubleu-net: A deep convolutional neural network for medical image segmentation[C]// 2020 IEEE 33rd International Symposium on Computer-based Medical Systems(CBMS). IEEE, 2020:558-564.

[27] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556, 2014.

[28] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248-255.

[29] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinkingatrous convolution for semantic image segmentation[J]. arXiv: 1706.05587, 2017.

[30] WEI J, WANG S, HUANG Q. F^3Net: fusion, feedback and focus for salient object detection[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020:12321-12328.

WU Yibo, born in 1998, postgraduate. His main research interests include dense small target segmentation based on deep learning and so on.

HAO Yingguang, born in 1968, associate professor. His main research interests include modeling complex time-varying systems and image processing algorithm.