doi:10.6041/j.issn.1000-1298.2021.09.022

基于卷积神经网络的生菜多光谱图像分割与配准

黄林生[1]　邵　松[1]　卢宪菊[2,3]　郭新宇[2,3]　樊江川[2,3]

(1.安徽大学农业生态大数据分析与应用技术国家地方联合工程研究中心,合肥 230601;
2.国家农业信息化工程技术研究中心,北京 100097;3.数字植物北京市重点实验室,北京 100097)

摘要:针对多光谱图像中由于多镜头多光谱相机各通道之间存在的偏差以及传统分割方法的不适用,图像分析处理过程往往会出现无法自动化分割或分割精度较低的问题,提出采用基于相位相关算法和基于 UNet 的语义分割模型对田间生菜多光谱图像进行各个通道的精确配准并实现前景分割。使用 Canny 算法对多光谱各通道图像进行边缘提取,进而使用相位相关算法对多光谱各通道图像进行配准,单幅图像平均处理时间 0.92 s,配准精度达到 99%,满足后续图像分割所需精度;以 VGG16 作为主干特征提取网络,直接采用两倍上采样,使最终输出图像和输入图像高宽相等,构建优化的 UNet 模型。实验结果表明:本文所提出的图像配准和图像分割网络,分割像素准确率达到 99.19%,平均 IoU 可以达到 94.98%,能够很好地对生菜多光谱图像进行前景分割,可以为后续研究作物精准表型的光谱分析提供参考。

关键词:生菜;多光谱图像;图像配准;图像分割;卷积神经网络

中图分类号:S636.2;TP391.41　　**文献标识码**:A　　**文章编号**:1000-1298(2021)09-0186-09　　**OSID**:

Segmentation and Registration of Lettuce Multispectral Image Based on Convolutional Neural Network

HUANG Linsheng[1]　SHAO Song[1]　LU Xianju[2,3]　GUO Xinyu[2,3]　FAN Jiangchuan[2,3]

(1. *National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Anhui University, Hefei* 230601, *China*
2. *National Engineering Research Center for Information Technology in Agriculture, Beijing* 100097, *China*
3. *Beijing Key Laboratory of Digital Plants, Beijing* 100097, *China*)

Abstract: In view of the deviations between the channels of multi-lens multi-spectral cameras and the inapplicability of traditional segmentation methods in multi-spectral images, the image analysis and processing process often has the problem of inability to automate segmentation or low segmentation accuracy, so a phase-based algorithm was proposed. And the semantic segmentation model based on UNet performs accurate registration of each channel of the field lettuce multispectral image and realizes foreground segmentation. The Canny algorithm was used to extract the edges of the multi-spectral channel images, and then the phase correlation algorithm was used to register the multi-spectral channel images. The average processing time of a single image was 0.92 s, efficiency was increased by 40%, and the registration accuracy reached 99%, which met the requirements of subsequent images and the required accuracy of segmentation. VGG16 was used as the backbone feature extraction network, and the double up sampling was directly used to make the final output image and the input image equal in height and width, and the optimized UNet model was constructed. The experimental results showed that the image registration and image segmentation network proposed achieved 99.19% pixel accuracy and an average IoU of 94.98%. It can perform foreground segmentation on lettuce multispectral images very well, which can be used for follow-up spectral analysis to study the precise phenotype of crops.

Key words: lettuce; multispectral image; image registration; image segmentation; convolution neural network

收稿日期:2021-04-18　修回日期:2021-06-19

基金项目:北京市农林科学院协同创新中心建设专项(KJCX201917)、国家自然科学基金面上项目(31871519)和北京市农林科学院科研创新平台建设项目(PT2021-31)

作者简介:黄林生(1977—),男,教授,主要从事遥感信息处理研究,E-mail:linsheng0808@163.com

通信作者:樊江川(1988—),男,工程师,主要从事植物表型组学和农业遥感研究,E-mail:fanjc@nercita.org.cn

0　引言

生菜是世界上广泛种植的一种重要的经济蔬菜作物，美国、欧洲和中国产量较高[1]。生菜富含维生素、类胡萝卜素、抗氧化剂和其他植物营养素[2]，是常见的蔬菜之一。生菜叶片是生物量累积的主要器官，其在营养生长过程中可多次收获[3]，因此如何无损地监测生菜生长状态和确定其采摘时期具有重要意义。

现阶段传感器技术和图像分析技术快速发展，在获取丰富数据的同时，可挖掘的图像信息也越来越丰富。拼接、配准、融合、语义分割和目标识别等技术是图像挖掘的基础。多光谱成像技术因其具有无损、快速、高效等优点，在对作物表型[4]以及生理状态的监测和分析中具有广阔的应用前景。其中，如何快速、精准地通过多光谱图像对作物进行感兴趣区域(ROI)提取是数据分析的前提，具有重要的意义。

光谱成像技术因其具备能够同时表征作物生长状态及生理生化信息的图像和光谱特征，近年来已被广泛应用于农业领域，如作物的品质检测、长势检测等领域。Pix4D 软件、ENVI 软件在处理离地面较近，如地面表型平台或者近地使用多光谱相机对作物进行拍摄分析表型(一般多光谱相机与地面距离 1 ~ 3 m 时)，并不能解决 5 通道引起的相位差，导致在分析光谱时产生较大误差。根据已有的研究，常用的匹配算法可以归纳为：像素匹配[5]、区域匹配[6]、特征匹配[7-9]、相位相关匹配[10]等。但采用单一的匹配算法进行图像匹配，往往难以达到理想的匹配结果。Canny 算法[11]是一种高效的边缘提取算法，对噪声具有较强的抗干扰能力，提取的边缘只有一个像素，准确度较高。因此使用 Canny 边缘提取算法与相位相关算法结合的方法对多光谱图像进行配准，可以大幅度降低图像配准时间。

从农业场景分割问题的大量研究成果来看，目前常用的一些图像分割算法主要可以分为 4 类：基于阈值的分割算法[12-13]，基于聚类的分割算法[14-16]，基于分类的分割算法[17-18]，以及基于图论的分割算法[19-20]。传统分割算法大多基于人工提取图像的灰度、颜色、纹理和空间等特征，利用增强前景和背景区域的差异性来实现图像前景的分割，因此在图像处理过程中往往存在大量欠分割和过分割的情况。随着深度学习技术的发展，其在自然语言处理[21]、图像识别[22]、视频跟踪[23]等领域的应用和推广均超越了传统的机器学习算法。2015 年美国笛卡尔实验室宣布开展深度学习在农业上的应用，此后卷积神经网络因其较强的图像特征提取能力开始广泛应用于农业领域，越来越多的学者开始将深度学习应用于作物表型信息获取分析[24]。文献[25]提出的全卷积神经网络(Fully convolutional neural networks，FCN)因其可以针对任意尺寸的输入图像并且实现端到端的像素分类，成为解决图像分割问题的重要方法。当前多数先进的语义分割框架均基于 FCN 实现且被应用于农业的各个领域。然而，FCN 的分割精度并不能满足细粒度表型提取的要求。UNet 是一种基于全卷积网络扩展的图像分割算法[26]，其通过对图像进行下采样后再进行上采样，跨连接层实现上下文信息融合，从而拥有更加丰富的语义特征和更加精准的分割效果。VGG16[27]卷积神经网络是牛津大学视觉几何组提出的深层特征提取网络，增加网络深度能够提升网络的最终性能，在图像领域获得了广泛应用[28-29]。UNet 语义分割模型因其网络结构在对图像进行分类时取得较好的效果得以在农业上广泛应用，但因其特征提取网络深度不够，无法提取图像更深层次的语义特征，不能满足精细化表型的提取要求。使用 VGG16 网络进行图像特征提取，加深网络深度，同时结合 UNet 语义分割模型网络结构可对多光谱生菜图像进行精细分割。

目前应用于农业领域的语义分割模型多因细粒度程度不够且分割精度无法满足需求而难以实现生菜多光谱图像精细化表型信息提取。针对这些问题，本文提出基于 VGG 特征提取网络的 UNet 语义分割模型来对生菜多光谱进行精准分割，使用结合 Canny 边缘提取算法的相位相关算法对生菜多光谱图像进行精确配准，以期为后期分析多光谱图像，提取生菜表型和生理信息提供参考。

1　实验材料

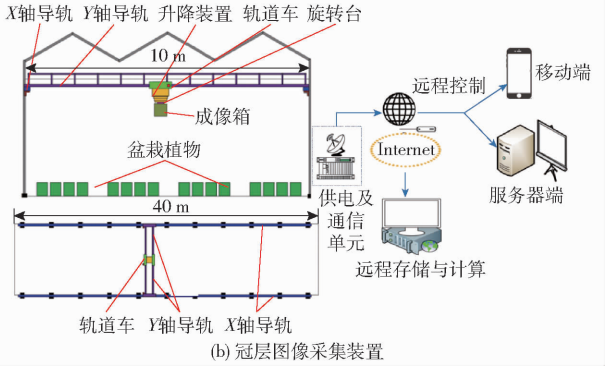
1.1　生菜多光谱图像采集装置

实验在北京市农林科学院联合大温室进行。生菜种植区尺寸为 10 m × 40 m，生菜采用盆栽，共计种植 271 个材料，每个材料有 4 株生菜，按 2 × 2 摆放，正常水肥管理，具体种植示意图如图 1a 所示。

种植区内部署了冠层图像采集装置，如图 1b 所示。冠层图像采集系统包括：植物表型平台基础结构、表型平台控制机构、传感器数据采集模块、数据传输系统。图像采集系统布设在种植区冠层顶部 4 m 高处，成像高度可以在 1 ~ 3 m 范围内自由调节。图像采集是由远程控制端(移动端手机或者服务器端)发出图像收集指令。图像采集装置收到远程指令后按照设定好的路线和采集方式进行采集，



(a) 种植示意图



(b) 冠层图像采集装置

图 1 生菜种植区概况

Fig. 1 Overview of lettuce planting area

采集到的图像可以通过无线路由端实时传输到服务器端。常用多镜头多光谱相机如图 2 所示,本研究采用 MicaSense Altum 5 通道多光谱相机,可以采集蓝(B)、绿(G)、红(R)、红边(Red edge)、近红外(NIR)5 个通道的多光谱图像,Altum 多光谱相机可以同时收集 5 个不连续的光谱波段,并且使用窄带滤光片提供针对单一波段图像大分辨率。相机参数见表 1。



(a) Sequoia 4通道多光谱相机



(b) RedEdge MX 5通道多光谱相机



(c) Airphen 6通道多光谱相机



(d) MicaSense Altum 5通道多光谱相机

图 2 多镜头多光谱相机

Fig. 2 Multi lens multispectral camera

1.2 多光谱图像获取及数据集构建

实验图像采集高度设定为 1.5 m,图像采集装置移动速度设置为 0.05 m/s,多光谱相机设定为 4 s 拍摄一次,拍摄完整各种植区共获得 1 200 幅多光谱图像。

经过筛选,剔除高重叠度图像(纵横重叠度在 30% 以上),共获得 400 幅有效多光谱图像,并形成原始图像数据集。为了减小图像尺寸、提高后期模型训练效率,将图像尺寸从 2 064 像素 × 1 544 像素缩放为 512 像素 × 512 像素。由于 CNN 卷积神经网络需要大量的训练数据以避免数据分析过程中出现过拟合现象,使用旋转、镜像、缩放、添加高斯噪声等方式进行标记样本数量扩充。最终数据集包含 2 310 幅图像,并以 7∶3 随机划分为训练集和测试集,因此共有 1 617 个训练样本输入 UNet 模型进行训练,其余 693 幅图像进行测试。

表 1 MicaSense Altum 相机参数

Tab. 1 MicaSense Altum parameters

参数	数值/型式
尺寸/(mm × mm × mm)	82 × 67 × 64.5
质量/g	357
分辨率/(像素 × 像素)	2 064 × 1 544
传感器尺寸/(mm × mm)	7.12 × 5.33
波段	绿、蓝、红、近红外和红边
数据接口	2 个 USB3.0 超级高速接口,支持 WiFi 或以太网和 USB3.0 存储
采样频率/Hz	1(所有波段,RAW 12-bit)
输出格式	DCM10bit(3B 每幅)

2 实验方法

多光谱图像配准与分割算法主要流程如图 3 所示,第 1 阶段对输入的多光谱图像使用 Canny 算法进行边缘提取,将得到的边缘提取图像使用相位相关算法得到各通道图像之间的平移参数,之后将各通道图像配准到相同位置。第 2 阶段将得到的配准后的生菜多光谱图像进行前景像素提取,利用 UNet 架构的优点,使用 VGG16 作为主特征提取网络,构造了生菜多光谱分割模型,将各通道的生菜多光谱图像从背景中精确分离出来。

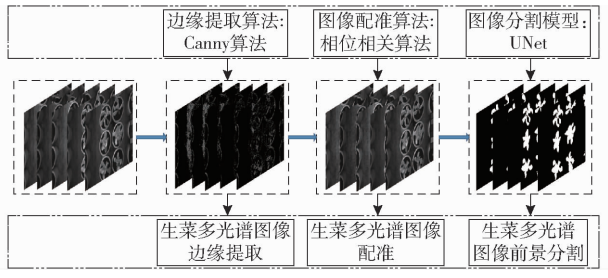


图 3 算法主要流程图

Fig. 3 Main flow chart of algorithm

2.1 改进相位相关算法对多光谱各通道图像配准

相位相关算法是将图像变换到频域,求得互功率谱,由峰值位置得到 2 幅图像的平移参数,由于需要估计整幅像素精度的平移参量,算法耗时较长。本文使用 Canny 算法对图像进行边缘特征提取,将得到的边缘图像作为后续算法的操作对象,可以大幅降低相位相关算法的图像配准时间。算法流程图和示意图如图 4 所示。

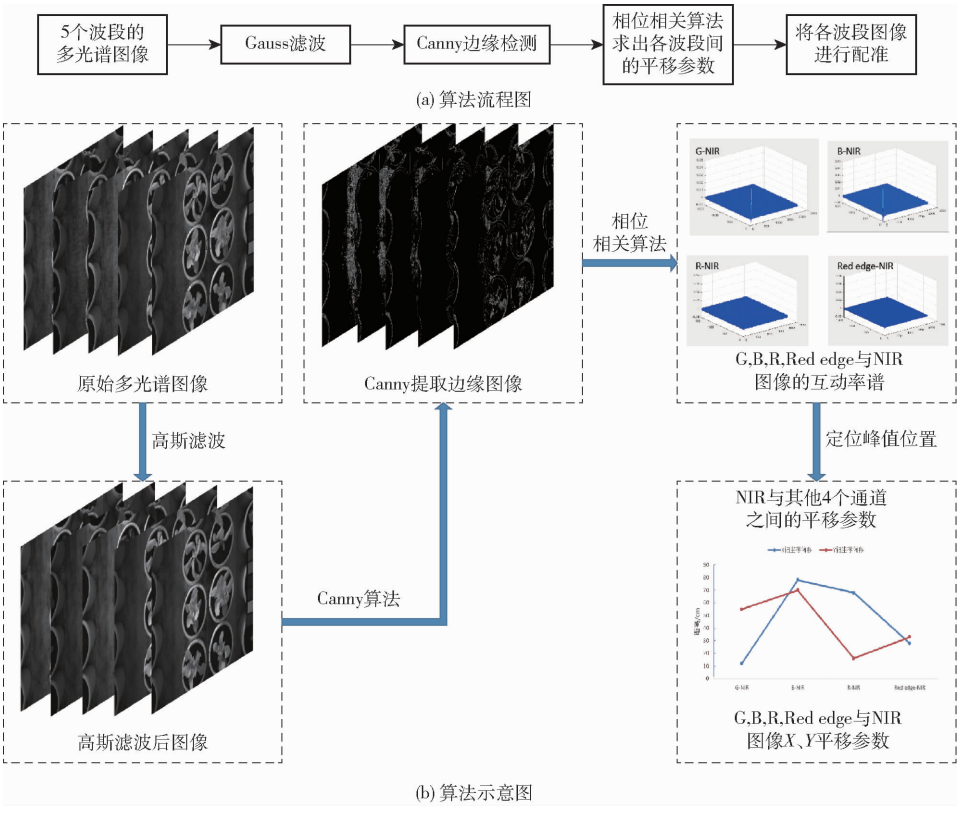


图 4　多光谱图像配准与分割算法

Fig. 4　Multispectral image registration and segmentation algorithm

2.1.1　Canny 边缘检测算法

本文 Canny 边缘检测算法的处理流程可以分为以下 4 个步骤：

(1)噪声去除：由于边缘检测很容易受到噪声影响，所以首先使用 5×5 的高斯滤波器来平滑和去除噪声。

(2)计算图像梯度：对平滑后的图像使用 Sobel 算子计算水平方向和竖直方向的一阶导数(图像梯度 G_x 和 G_y)，根据得到的这 2 幅梯度图(G_x 和 G_y)找到边界的梯度和方向，公式为

$$\text{Edge_Gradient}(G) = \sqrt{G_x^2 + G_y^2} \tag{1}$$

$$\text{Angle}(\theta) = \arctan\frac{G_x}{G_y} \tag{2}$$

式中　Edge_Gradient(G)——像素点的灰度函数

Angle(θ)——像素点的梯度函数

梯度的方向一般总是与边界垂直。梯度方向被归为 4 类：垂直、水平和 2 个对角线。

(3)使用非极大值抑制(即寻找像素点的局部最大值，将非极大值所对应的像素点灰度设置为 0)来消除边缘检测带来的杂散响应。

(4)滞后阈值：需要通过设定 2 个阈值来确定真正的边界，对于低于 minVal 阈值的像素点直接抛弃，高于 maxVal 阈值的像素点为真正的边界。如果处于阈值之间的像素点与真正的边界点相连便认为其也是边界点，否则抛弃。

如图 5 所示，D 线段像素都处于 minVal 阈值外面，因此 D 直接抛弃；A 是高于 maxVal 阈值的像素点，属于真正的边界点；B 像素虽然处于 2 个阈值之间，但是没有与真正的边界点相连，所以也被抛弃；C 像素与真正的边界点相连，因此也得以保留。因此选择合适的阈值点对于边缘检测非常重要。

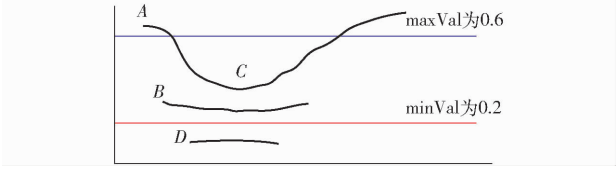


图 5　边界阈值选取

Fig. 5　Boundary threshold selection

本文中 Canny 算子的阈值设置 minVal 为 0.2，maxVal 为 0.6，用来计算图像梯度的 Sobel 卷积核的大小设置为 3，使用 L2gradient 参数作为求梯度大小的方程。通过 Canny 算法得到的边缘图像作为后续进行相位相关算法的操作对象。

2.1.2　相位相关匹配算法

图像配准是图像处理的基本任务之一，其主要用于将不同时间、不同传感器、不同视角或者不同拍摄场景获取的关于同一目标或者场景或者多幅图像进行最主要几何意义上的匹配过程。本文中室内冠层采集装置使用 MicaSense Altum 多光谱相机进行

多光谱数据采集,其采集到的数据是 5 幅不同波段的图像,5 幅图像来自 5 个镜头拍摄,图像之间具有水平偏移。图像的平移运动可以通过傅里叶变换到频域中相位的变化表现出来,所以本文采用基于相位相关的模板匹配方法精确计算不同波段之间图像的相对平移量。

假设 $f_1(x,y)$ 和 $f_2(x,y)$ 为 2 幅图像的时域信号,其满足的关系(即 $f_2(x,y)$ 由 $f_1(x,y)$ 经过简单的平移得到)为

$$f_2(x,y)=f_1(x-x_0,y-y_0) \tag{3}$$

根据傅里叶变换的性质可得

$$F_2(u,v)=F_1(u,v)e^{-j(ux_0+vy_0)} \tag{4}$$

式中 $F_1(u,v)$——$f_1(x,y)$ 的傅里叶变换

$F_2(u,v)$——$f_2(x,y)$ 的傅里叶变换

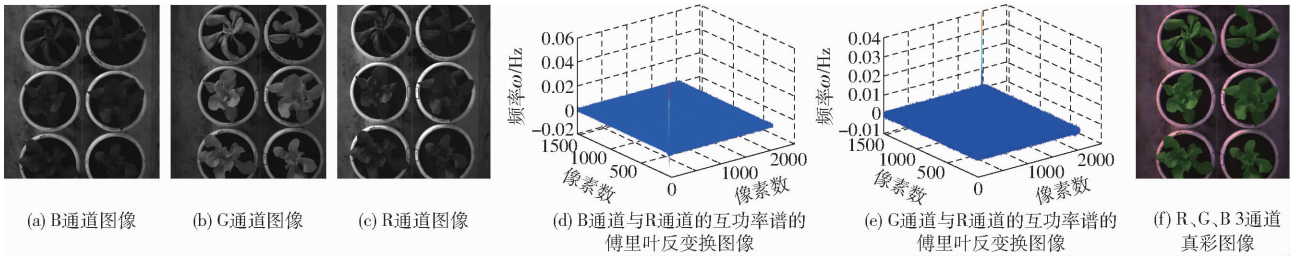
其互功率谱为

$$\frac{F_1^*(u,v)F_2(u,v)}{|F_1^*(u,v)F_2(u,v)|}=e^{-j(ux_0+vy_0)} \tag{5}$$

式中 $F_1^*(u,v)$——$F_1(u,v)$ 的复共轭

$e^{-j(ux_0+vy_0)}$ 的傅里叶反变换为一个二维脉冲函数 $\delta(x-x_0,y-y_0)$。求取式(4)的傅里叶反变换,然后找到最高峰的位置,最高位置的坐标即是平移参数 x_0、y_0。

图 6a ~6c 为多光谱 3 个通道图像,分别对 B 通道与 R 通道、G 通道与 R 通道进行相位相关算法计算求得互功率谱,再对其求傅里叶反变换,得到 2 个二维脉冲函数,对其定位峰值可分别得到 B 通道与 R 通道、G 通道与 R 通道的位移偏差。将其配准后进行叠加融合得到图 6f 的 3 通道真彩图像。图 7 为相位相关算法与本文算法在多光谱各通道上的匹配时间。



(a) B通道图像 (b) G通道图像 (c) R通道图像 (d) B通道与R通道的互功率谱的傅里叶反变换图像 (e) G通道与R通道的互功率谱的傅里叶反变换图像 (f) R、G、B 3通道真彩图像

图 6 多光谱图像配准过程

Fig. 6 Multispectral image registration process

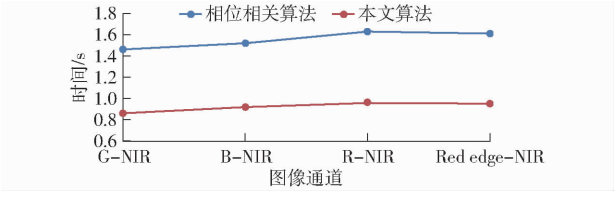


图 7 相位相关算法和本文算法在多光谱各通道上的匹配时间

Fig. 7 Matching time of phase correlation algorithm and proposed algorithm on each channel of multispectrum

本文采用 Canny 算法与相位相关算法相结合对多光谱各通道图像进行配准,如图 7 所示,相较于传统相位相关算法在各通道配准时间上均有大幅度提高。

为验证本文算法,选取 2 个基于特征点匹配的经典算法 SIFT(Scale-invariant feature transform)尺度不变特征变换算法和 SURF(Speeded up robust features)加速文件特征算法作为对比,用单幅图像的配准时间和匹配准确率作为评价指标,结果如表 2 所示。

由表 2 可以看到,基于特征点匹配的 SIFT 算法虽然匹配精度很高,但是匹配速度太慢,SURF 算法精度不高,本文算法相对于传统单一相位相关算法在匹配精度上提升 3 个百分点,匹配时间上减少了 0.61 s,效率提高了 40%,本文算法在配准精度和速度上都有比较明显的优势。

表 2 算法测试结果

Tab. 2 Algorithm test result

算法	匹配时间/s	匹配准确度/%
SIFT 算法	30	99
SURF 算法	2.3	71
相位相关算法	1.53	96
本文算法	0.92	99

2.2 基于 UNet 模型的生菜多光谱图像语义分割

2.2.1 UNet 网络结构

UNet 是目前比较流行的用于语义分割的深度学习模型,它是在全卷积网络(FCN)的基础上做了改进,提高了传统的 CNN 模型用于像素级图像分类的性能。UNet 的主要结构由卷积编码单元和卷积解码单元组成。重复的卷积运算在网络的 2 个部分被一个矫正的线性单元执行。在编码单元中,执行 2×2 的最大池化操作,将输入的样本降为 2 倍下采样。在每个降采样相位之后,特征数增加 1 倍以补偿分辨率的损失。解码单元由上采样块组成,通过使用卷积对输入进行上采样后,将上采样的特征映射与编码单元提取的相应特征映射连接起来,使网络在下采样过程中最大限度地保留最基本的特征信

息。重复上采样块，直到网络的输出与输入图像的尺寸匹配。一般情况下，编码单元主要用于捕获图像中的上下文信息，解码单元用于精确定位需要分割的部分。UNet 模型在语义分割方面有几个关键优势：首先，它结合了低层细节信息和高层语义信息，允许网络在大空间尺度上利用信息，而不会丢失有用的本地信息。其次，它只需要少量的训练样本就可以获得更好的分割性能。最后，端到端网络处理整个图像并直接产生分割图像。

2.2.2　VGG16 特征提取网络

目前研究人员开始应用 UNet 模型分割多通道遥感数据并提供了有效的解决方案。在本研究中，根据 UNet 模型的特点，采用 VGG16 作为主干特征提取网络。结构如图 8 所示，本文采用 Imagenet 上的预训练权重进行迁移学习，来提高模型的泛化性。

当输入的图像尺寸为512 × 512 × 1时，VGG16 网络的具体参数配置如表 3 所示。

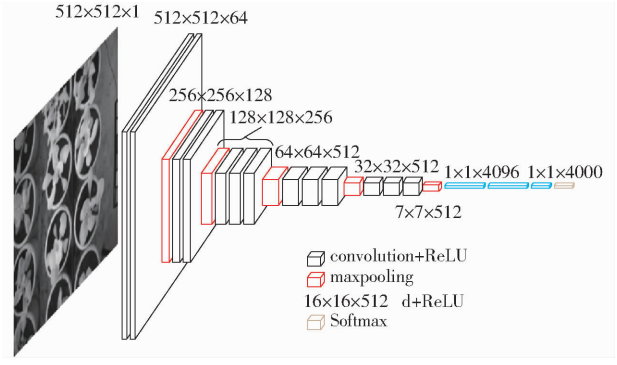


图 8　VGG 特征提取网络结构

Fig. 8　VGG feature extraction network structure

表 3　特征提取网络结构参数配置

Tab. 3　Configuration of network structure parameters for feature extraction

网络层		卷积核	通道数/步长	输出	结构数量
卷积模块 1	Conv	3 × 3	64	512 × 512 × 64	2
	maxpooling	2	2	256 × 256 × 64	1
卷积模块 2	Conv	3 × 3	128	256 × 256 × 128	2
	maxpooling	2	2	128 × 128 × 128	1
卷积模块 3	Conv	3 × 3	256	128 × 128 × 256	3
	maxpooling	2	2	64 × 64 × 256	1
卷积模块 4	Conv	3 × 3	512	64 × 64 × 512	3
	maxpooling	2	2	32 × 32 × 512	1
卷积模块 5	Conv	3 × 3	512	32 × 32 × 512	3
	maxpooling	2	2	16 × 16 × 512	1

2.2.3　整体网络结构

本文使用的 UNet 网络结构主要由主干特征提取网络（VGG16）和加强特征提取网络构成，其形状可以看作 U 型。利用 VGG16 主干特征提取网络可以获得 5 个初步的特征层，在加强特征提取网络里，利用这 5 个初步有效特征层进行特征融合，特征融合的方式就是对特征层进行上采样并且进行堆叠。为了方便网络的构建和更好的通透性，在加强特征提取网络里，在进行上采样时直接进行 2 倍上采样再进行特征融合，这样最终获得的特征层和输入图像的宽高相同。具体示意图如图 9 所示。

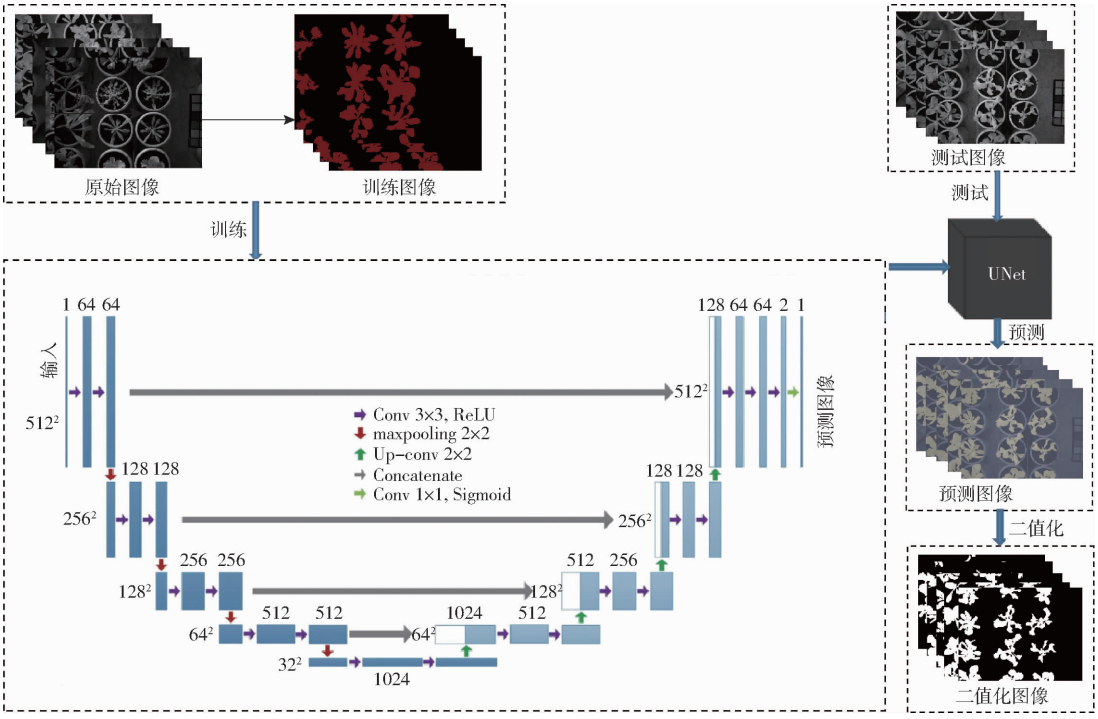


图 9　UNet 模型结构示意图

Fig. 9　Structure diagram of UNet model

2.2.4 模型训练

采用非线性函数 ReLU 作为模型隐藏层的激活函数,如图 10 所示,ReLU 由于非负区间的梯度为常数,因此不存在梯度消失问题(Vanishing gradient problem),使得模型的收敛速度维持在一个稳定状态。

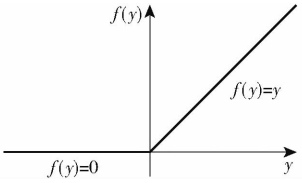


图 10 ReLU 激活函数

Fig. 10 ReLU activate function

$$\mathrm{ReLU}(x)=\begin{cases}x & (x>0)\\ 0 & (x\leqslant 0)\end{cases} \tag{6}$$

使用 UNet 全卷积神经网络进行图像分割,目的是为了输出一个分割精度较高,尺寸与输入图像相当的分割结果,本文所使用的损失函数主要由 Cross Entropy Loss 和 Dice Loss 组成。模型最终输出一个二分类图像,使用 Softmax 对像素点进行分类时使用 Cross Entropy Loss 作为 Loss 损失函数。交叉熵主要是用来表达神经网络输入输出的结果差异,一般来说,交叉熵越小,输入和输出的结果就越接近。分类的交叉熵损失函数为

$$\mathrm{loss}(\boldsymbol{x},C)=-\mathrm{lb}\frac{\exp(\boldsymbol{x}[C])}{\sum_j\exp(\boldsymbol{x}[j])}=-\boldsymbol{x}[C]+\mathrm{lb}\sum_j\exp(\boldsymbol{x}[j]) \tag{7}$$

式中 $\boldsymbol{x}$——输入向量

C——当前样本类别对应索引

$\boldsymbol{x}[C]$——当前样本预测函数输出

$\boldsymbol{x}[j]$——预测函数第 j 个输出

Dice Loss 将语义分割的评价指标作为 Loss,Dice 系数是一种集合相似度度量函数,通常用于计算 2 个样本的相似度,取值范围为[0,1]。越大表示预测结果和真实结果重合度越大,所以 Dice 系数越大越好,而 Loss 则越小越好,计算公式为

$$\mathrm{Dice_Loss}=1-\frac{2|\boldsymbol{X}\cap\boldsymbol{Y}|}{|\boldsymbol{X}|+|\boldsymbol{Y}|} \tag{8}$$

式中 Dice_Loss——损失函数

$\boldsymbol{X}$——预测值像素点集

$\boldsymbol{Y}$——真实值像素点集

模型训练过程就是对损失函数进行优化的过程。本文模型训练采用适应性矩估计优化算法(Adaptive torque estimation optimization algorithm, Adam),和一般传统的随机梯度下降不同。Adam 作为一阶优化算法可以在迭代训练数据时更新神经网络权重,并且其实现方式更直接,可以解决大规模数据和参数的优化问题,计算高效且所需内存少。

2.2.5 评价指标

为了量化本文语义分割方法对生菜冠层多光谱图像的分割效果以及对比不同方法的分割性能,在此引入平均像素准确率[30](Mean pixel accuracy, MPA)、平均交并比(Mean intersection over union, MIoU)、召回率(Recall)、精确率(Precesions)和 F1 值(F1-Score)来进行评价。

考虑到实际使用中需要处理高通量的冠层图像数据,本文利用平均处理时间来说明不同方法的时间性能。平均处理时间定义为某个分割方法对单幅图像分割所需时间的平均值。

3 结果与分析

3.1 实验参数设置及训练过程分析

在软件环境为 Windows 10 系统,开发语言环境为 anaconda 3 和 Python 3.6,硬件环境为 Intel(R) Xeon(R) CPU E5-2697 处理器和 NVIDIA GTX1070 Ti GPU 的计算机上进行实验,深度学习开发框架为 Tensorflow_Keras。

本文模型训练将初始学习率设置为 1×10^{-4},衰减率设置为 0.9,将训练集的 1 617 个训练样本每 2 幅图像作为一个批次(batch)输入到模型中进行训练,训练一共进行 300 个 epoch 迭代循环。为了加快训练速度和训练初期防止权值被破坏,本文在训练的前 10 个 epoch 冻结一部分神经网络进行训练,之后把所有神经网络进行解冻训练,同时将学习率调整为 1×10^{-5},衰减率依然设置为 0.9。

3.2 不同模型和网络结构精度比较

为了测试和对比本文所提出模型的性能,对测试集 693 幅生菜多光谱冠层图像采用传统的 UNet 方法、本文方法 UNet_VGG、一种编码解码语义分割网络方法 SegNet[31](A deep convolutional encoder-decoder architecture for image segmentation)、基于 VGG 特征提取网络的 SegNet 方法、金字塔场景解析网络方法 PSPNet[32](Pyramid scene parsing network)以及基于 VGG 特征提取网络的 PSPNet 方法进行语义分割并分析。

对比 6 个模型的精确率曲线和损失值曲线可以看出(图 11),UNet 模型比 SegNet 模型的收敛速度更快,并且分割的精度也更高。用 VGG 作为特征提取网络的模型要比传统语义分割模型的分割精度更高,随着迭代次数的增加,模型分割精度不断上升并趋于稳定,当迭代到 300 个 epoch 时,损失函数值基本收敛,表明模型达到了比较好的训练效果。

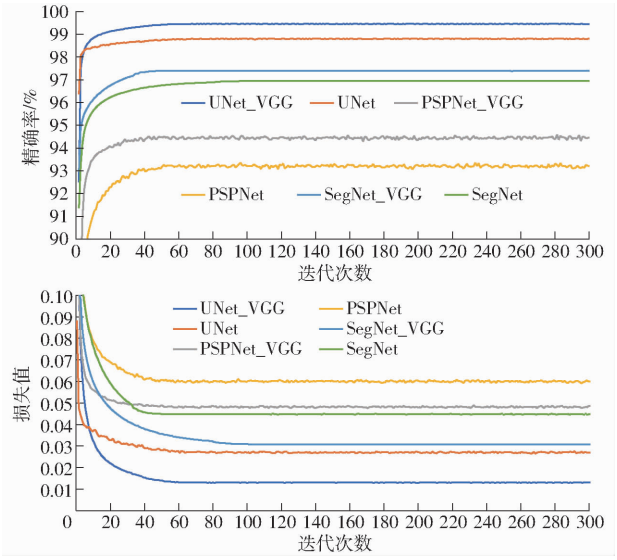


图 11　不同分割模型的精确率和损失曲线对比

Fig. 11　Precision and loss curves of different segmentation methods

表 4 显示了本文所提出的模型 UNet_VGG 与其他语义分割模型在测试集上的性能对比，可以看到，相较于传统的 UNet 模型，UNet_VGG 的平均像素准确率和 F1 值得分均有不同程度的提高。其中 PSPNet 模型在测试集上的表现最差，表明在处理二分类分割问题时，UNet 模型分割相较于其他语义分割模型具有很大的优势。相较于原有模型，在使用 VGG 作为下采样进行特征提取时，模型精度均有不同程度提高。本文模型 UNet_VGG 与其他模型的分割效果对比如图 12 所示。

表 4　不同分割方法测试集分类结果对比

Tab. 4　Comparison of classification results of samples for validation by different segmentation algorithms

分割模型	平均像素准确率/%	平均交并比/%	精确率/%	召回率/%	F1 值/%	单幅图处理时间/s
PSPNet	96.31	78.77	87.19	88.95	88.02	0.15
PSPNet_VGG	96.93	81.99	89.71	90.37	90.03	0.12
SegNet	97.87	91.23	95.24	95.82	95.81	0.12
SegNet_VGG	98.23	91.86	96.12	96.21	96.12	0.12
UNet	98.67	92.91	96.24	96.42	96.32	0.11
UNet_VGG	99.19	94.98	97.29	97.56	97.42	0.11

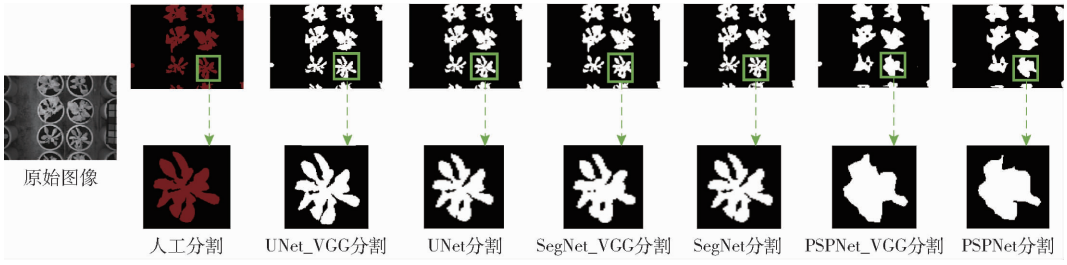


图 12　不同分割方法在测试集上的分割结果对比

Fig. 12　Comparison of segmentation results of different segmentation methods on test set

4　结论

（1）针对多镜头多光谱相机在近距离拍摄时各通道图像存在位移偏差，采用 Canny 算法与相位相关算法结合的方法对多光谱各通道图像进行配准，结果表明结合算法有更高的配准精度，且相对于单一算法图像配准时间减少 0.61 s，效率提高了 40%。

（2）为满足对作物精确表型提取的需求，以配准后的生菜多光谱图像构建图像数据集。在 UNet 模型基础上，特征提取网络部分使用 VGG 网络，构建了本文模型。该模型在测试集上的平均分割准确率为 99.19%，平均交并比为 94.98%，相较于传统 UNet 模型分别提高 0.52 个百分点和 2.07 个百分点。本文方法单幅图像平均处理时间为 0.11 s，可以满足对植物表型图像实时分割的要求。

（3）实验结果表明，本文方法可以对多光谱图像进行精确配准和生菜前景的精确提取。

参　考　文　献

[1]　ADHIKARI N D, SIMKO I, MOU B. Phenomic and physiological analysis of salinity effects on lettuce[J]. Sensors, 2019, 19(21):4814.

[2]　GRAHN C M, BENEDICT C, THORNTON T, et al. Production of baby-leaf salad greens in the spring and fall seasons of northwest Washington[J]. HortScience, 2015, 50(10):1467 - 1471.

[3]　SIMKO I, HAYES R J, FURBANK R T. Non-destructive phenotyping of lettuce plants in early stages of development with optical sensors[J]. Frontiers in Plant Science, 2016, 7:1985.

[4]　张慧春，周宏平，郑加强，等. 植物表型平台与图像分析技术研究进展与展望[J/OL]. 农业机械学报，2020，51(3)：1 - 17. ZHANG Huichun, ZHOU Hongping, ZHENG Jiaqiang, et al. Research progress and prospect in plant phenotyping platform and image analysis technology[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(3):1 - 17. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx? flag=1&file_no=20200301&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2020.03.001. (in Chinese)

[5] XU X, CAO Y, WANG Y, et al. A fast pixel matching method based on phase feature extraction in online phase-measuring profilometry[J]. Journal of Modern Optics, 2017, 64(2):1 - 8.
[6] HE A, QUAN C. An improved principal component analysis based region matching method for fringe direction estimation[J]. Optics Communications, 2018, 413:87 - 102.
[7] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91 - 110.
[8] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]// 2011 International Conference on Computer Vision. IEEE, 2011: 2564 - 2571.
[9] BAY H, TUYTELAARS T, GOOl L V. SURF: speeded up robust features[M]. Springer-Verlag, 2006.
[10] LEI W, ZHEN Z, ZHANG X, et al. Adaptive camera control method for efficient stereoscopic photography[C]// 2016 IEEE International Conference on Industrial Technology (ICIT). IEEE, 2016.
[11] CANNY J. A computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI - 8(6):184 - 203.
[12] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems Man & Cybernetics, 2007, 9(1):62 - 66.
[13] 刁智华,刁春迎,袁万宾,等. 基于改进型模糊边缘检测的小麦病斑阈值分割算法[J]. 农业工程学报,2018,34(10):147 - 152.
DIAO Zhihua, DIAO Chunying, YUAN Wanbin, et al. Segmentation algorithm with threshold for wheat lesion based on improved fuzzy edge detection[J]. Transactions of the CSAE, 2018, 34(10):147 - 152. (in Chinese)
[14] MENENDEZ H D, BARRERO D F, CAMACHO D. A genetic graph-based approach for partitional clustering[J]. International Journal of Neural Systems, 2014, 24(3):281 - 297.
[15] ACHANTA R, SHAJI A, SMITH K, et al. Slic superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(11):2274 - 2282.
[16] 杨信廷,孙文娟,李明,等. 基于 K 均值聚类和开闭交替滤波的黄瓜叶片水滴荧光图像分割[J]. 农业工程学报,2016, 32(17):136 - 143.
YANG Xinting, SUN Wenjuan, LI Ming, et al. Water droplets fluorescence image segmentation of cucumber leaves based on K-means clustering with opening and closing alternately filtering[J]. Transactions of the CSAE, 2016, 32(17):136 - 143. (in Chinese)
[17] WANG X Y, WU Z F, LIANG C, et al. Pixel classification based color image segmentation using quaternion exponent moments[J]. Neural Networks, 2016, 74(C):1 - 13.
[18] 杨长辉,王卓,熊龙烨,等. 基于 Mask R - CNN 的复杂背景下柑橘树枝干识别与重建[J/OL]. 农业机械学报,2019,50(8):22 - 30,69.
YANG Changhui, WANG Zhuo, XIONG Longye, et al. Identification and reconstruction of citrus branches under complex background based on Mask R - CNN[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(8): 22 - 30, 69. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 20190803&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2019.08.003. (in Chinese)
[19] FROYLON D G, PADBERG-GEHLE K. Normalized cuts and image segmentation[J]. Chaos an Interdisciplinary Journal of Nonlinear Science, 2015, 22(8):888 - 905.
[20] LI Z, CHEN J. Superpixel segmentation using linear spectral clustering[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
[21] ZHANG W N, ZHU Q, WANG Y, et al. Neural personalized response generation as domain adaptation[J]. World Wide Web, 2019, 22(4):1427 - 1446.
[22] REN S, HE K, GIRSHICK R, et al. Faster R - CNN: towards realtime object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137 - 1149.
[23] SEO E Y, CHOI Y J, KIM J H, et al. Training method of deep learning-based decoder for punctured polar codes[J]. The Journal of Korean Institute of Communications and Information Sciences, 2018, 43(7):1176 - 1181.
[24] BARBEDO J. Factors influencing the use of deep learning for plant disease recognition[J]. Biosystems Engineering, 2018, 172:84 - 91.
[25] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640 - 651.
[26] RONNEBERGER O, FISCHER P, BROX T. U - Net: convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.
[27] 戚超,左毅,陈哲琪,等. 基于改进 VGG16 的大米加工精度分级方法研究[J/OL]. 农业机械学报,2021,52(5):301 - 307.
QI Chao, ZUO Yi, CHEN Zheqi, et al. Rice processing accuracy classification method based on improved VGG16 convolution neural network[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5):301 - 307. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 20210533&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2021.05.033. (in Chinese)
[28] SHI J, MALIK J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(8):888 - 905.
[29] RANGARAJAN A K, PURUSHOTHAMAN R. Disease classification in eggplant using pre-trained VGG16 and MSVM[J]. Scientific Reports, 2020, 10(1):2322.
[30] DAI J, HE K, SUN J. Instance-aware semantic segmentation via multi-task network cascades[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
[31] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(12):2481 - 2495.
[32] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[J]. IEEE Computer Society, 2016, 660(1):6230 - 6239.