

基于改进全卷积网络的小麦图像分割^①

万园洁¹, 卿粼波¹, 何小海¹, 董德良², 石恒²

¹(四川大学 电子信息学院, 成都 610065)

²(中储粮成都粮食储藏科学研究所, 成都 610091)

摘 要: 针对漫水填充结合模板匹配的双面联合分割方法对小麦图像进行分割存在过分割以及欠分割现象, 提出基于改进的全卷积网络的图像语义分割方法. 该方法融入前二个池化层的输出信息作为 Softmax 层的输入, 探讨并得出了只融入第二个池化层的输出信息的网络模型优于同时融入前两个池化层的网络模型, 引入 Batch Normalization 层到网络层中, 并且针对小麦图像的需要将原来的 21 类网络输出类别更换为 2 类输出. 实验采用建立的小麦图像数据库, 结果表明改进后的网络使得过分割和欠分割现象明显减少, 分割效果得到了显著提升, 并且使用 F-measure 定量分析了模型的有效性.

关键词: 分割; 漫水填充; 模板匹配; 全卷积网络; Batch Normalization

引用格式: 万园洁, 卿粼波, 何小海, 董德良, 石恒. 基于改进全卷积网络的小麦图像分割. 计算机系统应用, 2018, 27(3): 221–227. <http://www.c-s-a.org.cn/1003-3254/6272.html>

Wheat Grain Image Segmentation Based on Improved Fully Convolutional Network

WAN Yuan-Jie¹, QING Lin-Bo¹, HE Xiao-Hai¹, DONG De-Liang², SHI Heng²

¹(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

²(Sinograin Chengdu Grain Storage Research Institute, Chengdu 610091, China)

Abstract: Because the result of method used by the project group currently which is a combination of floodfill and template matching is poor, and there is also under-segmentation or over-segmentation, this paper proposes the application of fully convolutional networks in semantic segmentation for wheat images. Firstly, the output information of the second pool layer is integrated as the input of the Softmax layer. Then, the Batch Normalization layer is introduced into the network layer, and 21 classes of output of the network are changed into the output of the 2 classes because of the characteristics of wheat. And the paper uses the F-measure to evaluate the result. The experimental results show that the proposed network can improve the segmentation result.

Key words: segmentation; floodfill; template matching; fully convolutional networks; Batch Normalization

1 引言

小麦是我国的重要粮食作物之一, 也是世界三大谷物之一. 每年我国都会大量收购小麦, 在收购过程中其品质的好坏会直接影响小麦的价格, 因此对小麦进行评级显得尤为重要. 目前主要依靠专业的质检员对小麦品质进行评估, 不仅工作量大, 而且由于没有统一

的评估标准, 在对小麦定级时就会存在不同程度上的分歧. 这些问题, 运用机器学习中的识别分类方法可以得到有效解决. 分割作为预处理是其中的一个重要环节, 成为本文的研究重点.

在许多的图像和视频的应用中, 分割都已成为不可或缺的一部分^[1]. 目前比较常用的传统图像分割方法

① 基金项目: 成都市科技惠民项目 (2015-HM01-00293-SF); 四川大学研究生课程建设项目 (2016KCJS5113)

收稿时间: 2017-06-29; 修改时间: 2017-07-17; 采用时间: 2017-07-26; csa 在线出版时间: 2018-02-09

主要有: ① 基于阈值的分割方法^[2]; ② 基于区域的分割方法: 基于区域分割的经典算法是区域生长^[3]. 然而, 区域生长的固有缺点是其往往导致过度的分割, 结果易受噪声影响; ③ 基于边缘的分割方法^[4]: 边缘是图像的最基本的特征之一, 广泛存在于目标与背景、目标与目标之间. 这种分割方法首要做的就是检测边缘, 之后根据边缘信息把前景与背景分离开来. 但是由于本文研究的小麦包括正常粒、破碎粒、生病粒、生虫粒这些不同种类, 特征也比较复杂多样, 这些都使得上述方法并不适合本文小麦图像分割. 因为不仅不同品种的小麦之间的颜色、纹理大不相同, 即使是同一类别但属不同批次的小麦的颜色、纹理等特征也不尽相同. 例如同样是正常的麦粒, 在某一批次中颜色偏黄, 体型相对偏瘦长但在另一批次中可能会偏白, 体型饱满一些. 特别是生病粒, 由于麦粒会有赤霉病、青霉病、白霉病等, 有些麦粒不止存在一种病症, 因此直接用阈值分割、区域生长都会出现分割不准确的现象, 加之采集图片清晰度不够 (图片会存在部分小麦阴影), 这些都造成了使用边缘流分割大多也会出现过分割的现象. 目前, 项目组所使用的基于模板匹配结合漫水填充的双面联合分割方法在一定程度上减少了过分割与欠分割现象, 但是这种现象仍然存在.

近年来, 深度学习已经成为机器学习领域的研究热点, 解决了许多计算机视觉方面的难题, 基于此, 本文提出将改进后的全卷积网络分割方法应用于我们的分割模块, 实验结果表明分割效果得到了明显提升.

本文后续部分的结构如下: 第2节介绍模板匹配结合漫水填充的双面联合分割方法; 第3节阐述改进的全卷积网络的分割方法; 第4节运用本文建立的小麦图像库进行实验结果比较, 并展示部分麦粒分割效果图, 第5节是结论.

2 基于模板匹配结合漫水填充的双面联合分割方法

传统机器学习中小麦不完善粒分类的关键在于提取比较有效的特征, 由于麦粒的特征不会只存在于某一面, 为了能够更加全面的提取特征, 项目组对一颗小麦采取上下两面扫描的方案同时得到上下两张扫描图像, 并对这两幅扫描图像同时处理最终得到一颗小麦的上下两幅扫描图像, 如图1所示.

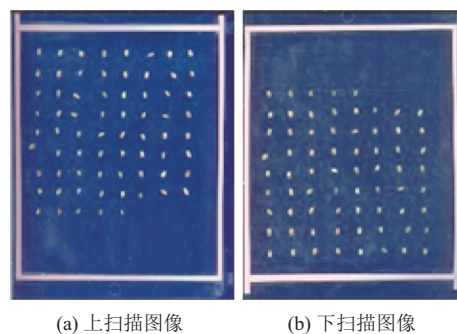


图1 上下扫描图像

基于模板匹配及漫水填充的双面联合分割方法主要包括以下几个步骤: 1) 采集同一版小麦的上下扫描图像; 2) 将下扫描图像翻转 180 度, 为模板匹配^[5]做准备; 3) 将下扫描图像沿着粉色框外 50 个像素的位置, 利用漫水填充^[6]进行分割, 同时将上扫描图像沿着粉色边框外 100 个像素的位置, 利用漫水填充算法进行分割; 4) 对图像进行滤波去噪, 并使用形态学方法再次去噪; 5) 再次利用漫水填充得到每颗小麦图像的掩码; 6) 得到一颗小麦的上下两幅掩码之后取其并集并反赋给原上下掩码, 进行分割, 简略分割流程图如图2所示.

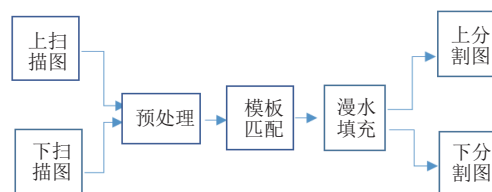


图2 模板匹配结合漫水填充的双面联合分割流程图

3 改进的全卷积网络的小麦图像分割方法

目前项目组所使用的基于模板匹配结合漫水填充的双面联合分割方法在效果上虽然取得了一定的提升, 但是仍然有部分小麦分割效果不好. 而且基于上述方法的分割过程比较复杂, 推广性不高, 对光照的鲁棒性不强. 现今, 深度学习在图像上的应用越来越成熟^[7]. 因此本文提出一种运用改进的全卷积网络对小麦图像进行分割的方法.

3.1 全卷积神经网络

Fully convolution neutral network 即为全卷积神经网络, 以下简称 FCN. FCN 最初是由 Jonathan Long 等人在文献^[8]提出. 该网络主要基于现有的 AlexNet, VGGNet 和 GoogleNet 微调得来. FCN 把分割归结为

一个像素回归问题, 其目标函数为^[8]:

$$\varepsilon(\theta) = \sum_p e(X_e(p), l(p)) \tag{1}$$

其中 p 是图像的像素点, $X_e(p)$ 是关于像素点 p 的回归函数, 损失函数 $e(X_e(p), l(p))$ 用来衡量输出 $X_e(p)$ 与 ground-truth $l(p)$ 的误差大小, 理论上模型效果越好, $\varepsilon(\theta)$ 值越小. FCN 主要由以下子网络组成^[9]:

① 卷积层: 卷积层的操作和传统的滑动窗很相似, 把卷积核作用于不同的区域并产生特征图, 即提取图像特征.

② 激活层 (ReLU Layers): ReLU 称为激活函数, 是对输入的一种非线性映射, 其函数表达式为: $f(x) = \max(0, x)$. 这种非线性的处理可以加快训练收敛的速度, 有助于得到更好的结果.

③ 池化层: 池化一般有两种:最大池化和平均池化. FCN 使用的是最大池化. 池化层可以理解为下采样, 这样做一方面可以缩小图像尺寸、减少网络参数, 另一方面增强了旋转不变性. 即一个 $N * N$ 的区域在特征图谱上滑动, 当 N 取值适当时可以认为该区域的最大值 (最大池化) 或者平均值 (平均池化) 可以近似表示原来的信息. 同时池化层还可以改善结果, 防止过拟合.

④ 反卷积层: 卷积层的反操作, 即是对上一层的输入进行上采样, 以得到与池化层之后进行下采样的图片同样的尺寸, 是整个网络的重要部分.

⑤ Softmax 层: 用来计算网络的输出和 ground-truth 之间的误差, 如式 (2).

上述网络层的所有权重都是使用随机梯度下降法^[10]通过 BP(反馈) 来学习得到.

3.2 用于小麦图像分割的全卷积网络的改进

全卷积网络原应用于具有 20 类目标的 Pascal VOC^[11]数据集. 而本文所需要的是把小麦作为目标, 从背景中提取出来, 如图 3 所示.

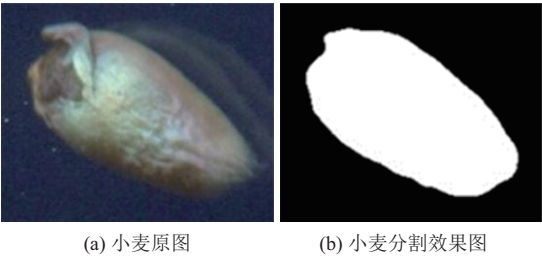


图3 小麦原图及其分割效果图

为了适合小麦数据集的需求, 尽量少引入噪声, 因此本算法将网络输出改为背景和前景两类.

具体操作过程如下: 首先根据待分割图像设置像素待分成的类别数量 N (本文为 2), 标签记为 $label_1, label_2 \cdots label_N$, 这个参数在训练模型之前已经设置好; 假设图片大小为 $H * W$, 则网络最终会生成 N 个 $H * W$ 大小的 map, 而 map 保存的值为每一个像素点属于 $label_i$ ($1 \leq i \leq N$) 的概率值, 最终通过比较得到每一个像素点在 N 个 map 里的最大值, 该最大值属于哪一个 label, 这个像素点就属于哪一类别, 具体如图 4 所示.

另外在可视化时由于每一种 label 指定的物体都有对应的具体的颜色, 因此在开始需要最先生成一个颜色与标签相对应的二类映射表, 如表 1 所示.

表1 颜色与标签的映射表	
RGB值	类别
(0, 0, 0)	0
(255, 255, 255)	1

像素 label 为 0 即为黑色部分, label 为 1 即为白色目标区域, 最终得到如图 3 中 B 图所示结果.

如图 4 所示, 假如每一个方格代表一个像素点, 方格内的值为该像素点属于某 label 的概率值 (例如左上角深灰色部分), 因为 $0.9 > 0.1$, 所以这个像素点被判定为 label 0, 对照颜色标签映射表可知, 可视化时该点为黑色部分; 类似地可以得到图中其他位置的映射结果.

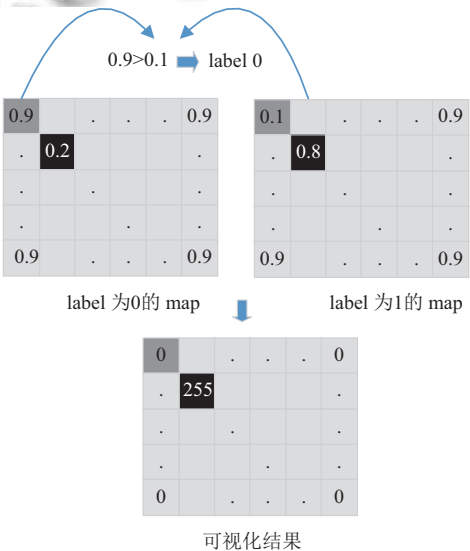


图4 图像的可视化过程

为了方便理解,在本文中称用自己建立的数据集对 FCN-8s 网络进行微调过的网络为 WheatFCN.

即使 WheatFCN 取得了一定的分割效果,但是仍然存在边界部分不光滑、内部有空洞等欠分割问题.对于卷积来讲,越是底层的卷积层,感知域越小,提取的特征越局部,越详细,包括一些重要的边缘信息以及点、线、面等简单的几何形状,而高层卷积的卷积核尺寸较大,通常是提取一些比较抽象的特征.因此全卷积网络采用 skip-net 多尺度特征提取方法,将网络中多个中间层的特征合并,但做信息融合时并未考虑前两层池化的输出信息,从而造成一些重要细节信息的丢失.为了使取得的特征更加全面,本文提出把第二层、第一层池化的输出也融入到后面的输入中,并且进一步实验比较得出了只融入第二个池化层的输出信息的网络模型优于同时融入前两个池化层的网络模型,具体的网络结构如图5所示.

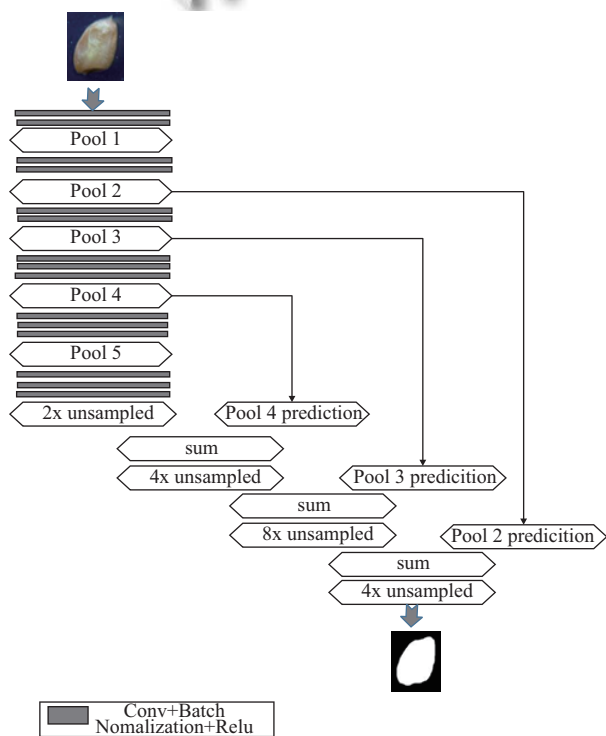


图5 WheatFCN+1 网络结构

此外,为了使得到的网络泛化能力更强,模型精度更加准确,本文还提出将 Batch Normalization 引入到网络中去. BN 最初是由 Sergey Ioffe、Christian Szegedy 提出来的.文献[12]中指出, BN 层的优势在于:能够加快收敛速度,不需要再慢慢调整学习率,即使选择的初

始学习率较大也没关系;不用考虑过拟合中 dropout、L2 正则项参数的选择问题,因为 BN 具有提高网络泛化能力的特性,另外也可以提升整个模型的精度[12].

经过改进的网络如图5所示,便于区分,在本文中将最终使用的改进的网络称为 WheatFCN+1,将同时融入第二池化层和第一池化层的输出信息的网络称为 WheatFCN+2.

4 数据集与实验结果

4.1 准备数据集

由于没有公开的小麦图像数据集,因此本文自己建立了数据集.首先,本文采集了 288 张如图2所示的扫描图像;然后设置一个固定大小的感兴趣区域在扫描图像上滑动将其分割成图3中 A 图所示的图像.图像的标签,即 ground-truth,是使用标签制作工具 lableMe 勾画出麦粒轮廓,然后再编写代码查找轮廓、填充轮廓,再经过微调校正得到最终的较为精准的小麦图像标签.为了得到一个对不同种类的小麦诸如正常、生病、破碎、生虫泛化能力都很强的模型,我们所建立的数据集要尽可能丰富,目前包括正常粒、生病粒、破碎粒、生虫粒共 10 000 粒,其中 8000 粒作为训练集,2000 粒作为测试集.

4.2 实验结果

本文实验使用深度学习框架 Caffe,实验环境是四核 Inter(R)Core(TM)i5-3475、Nvidia GTX960Ti 4G、ubuntu14.04.

首先本文用传统的分割方法对不同种类的小麦进行分割,但是实验效果出现严重的过分割或者欠分割现象,然后使用模板匹配结合漫水填充的双面联合分割方法进行实验,效果虽有明显提升,但仍然存在过分割或者欠分割现象,具体效果图如图6所示.

图6中小麦种类从上到下分别是正常、破碎、生病、生虫.从图6中可以看出阈值分割主要是出现欠分割,边缘存在轻微的过分割,而区域生长、边缘流大都会出现过分割现象,有些地方如边缘等也会伴随着欠分割.而项目组提出来的模板匹配结合漫水填充的双面联合分割效果明显较好,但是仍然存在过分割和欠分割现象.

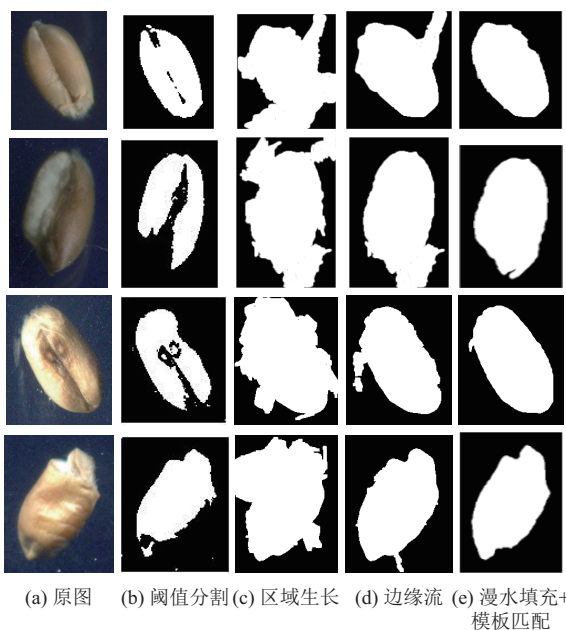


图6 传统方法与模板匹配结合漫水填充方法比较结果

然后我们用改进的 FCN 网络进行实验, 与改进前的网络相比分割效果有明显提升. 实验结果如图 7 所示.

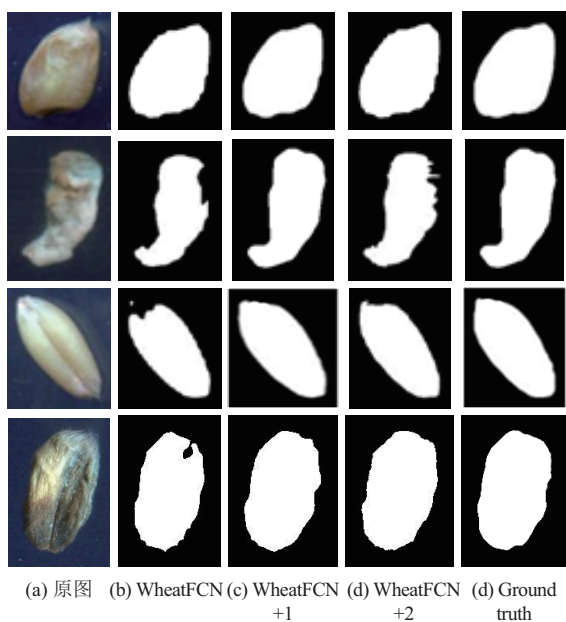


图7 改进前后的 FCN 网络的分割效果对比

图 7 中小麦种类从上到下分别是正常、破碎、生病、生虫. WheatFCN+1 网络为只融入池化层 2 的输出信息, WheatFCN+2 网络为同时融入池化层 2 和池化层 1 的输出信息. 由图 7 可以看出, WheatFCN 网络得到的结果不仅在内部有空洞, 在边界也有欠分割现

象, 而且会有明显的锯齿状, 不光滑. 但是 WheatFCN+1 网络得到的分割图不仅不存在欠分割现象而且边界相对而言也比较光滑. WheatFCN+2 分割效果虽然比 WheatFCN 要好, 但是相较于 WheatFCN+1 网络而言, 不仅提升不大, 反而出现了轻微的过分割现象, 边界也变得略粗糙. 实验结果表明我们改进过的网络 WheatFCN+1 的分割效果最好.

而模板匹配结合漫水填充的双面联合分割方法虽然效果有所改善, 但是依然会存在分割不理想的情况, 相比较而言我们改进的网络 WheatFCN+1 的分割效果更佳, 基本不会出现过分割或者欠分割现象, 具体实验结果如图 8 所示.

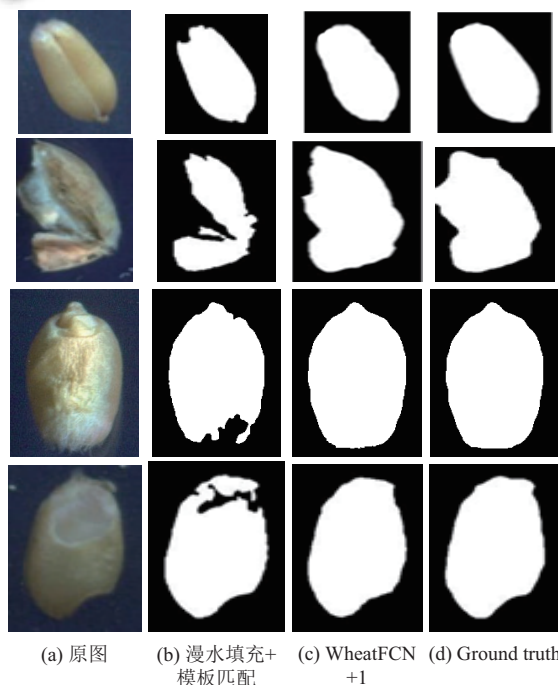


图8 模板匹配+漫水填充方法与 WheatFCN+1 方法结果比较

图 8 中小麦种类从上到下分别是正常、破碎、生病、生虫. 实验结果表明在模板匹配结合漫水填充的双面联合分割方法出现欠分割或者过分割的时候, WheatFCN+1 却总是可以得到比较好的分割效果.

为了更加客观地说明本文的网络分割效果最佳, 我们采取了文献[13]中所使用的 F-measure 评估方法对结果进行评估比较. 计算方法如下:

$$F = PR / (\partial R + (1 - \partial)P) \quad (2)$$

其中 P 是精确率, R 是召回率.

在图片分类、物体识别、信息检索、自动翻译等领域, P 、 R 是两个最基本的指标. 假设把分类中预测的标签结果分为以下 4 部分:

- (1) 存在且被正确判定的标签, 简称 A 类;
- (2) 不存在且被正确判定的标签, 简称 B 类;
- (3) 不存在但是被错误判定为存在的标签, 简称 C 类;
- (4) 存在但是被错误判定为不存在的标签, 简称 D 类.

根据以上 4 种分类, 可以由以下公式定义精确度和召回率:

$$P = A / (A + C) * 100\% \tag{3}$$

$$R = A / (A + B) * 100\% \tag{4}$$

而本文的分割问题完全可以归类为像素分类问题, 因此也可以用这两个值进行评估. 但是在实际应用中这两者在某些情况下往往是相互矛盾的, 需要综合考虑 P 与 R , 最常见的衡量指标即为 F-measure, F 值越大, 说明分割效果越好. 其中 δ 是权重, 根据文献[13]的说明, 我们将该参数选取为 0.5. 本文使用测试集小麦图像计算了各方法的 F-measure 的平均值, 结果如表 2 所示.

表 2 各方法的 F-measure 值

小麦种类	正常	破碎	生病	生虫
$F_{WheatFCN}$	0.875	0.814	0.784	0.861
$F_{WheatFCN+1}$	0.887	0.903	0.803	0.916
$F_{WheatFCN+2}$	0.878	0.823	0.802	0.902
$F_{模板匹配+漫水填充}$	0.826	0.691	0.798	0.841

由表 2 可以看出 WheatFCN+1 得到的 F 值对于四种不同的小麦均显示最高, 充分表明本文改进后的网络 WheatFCN+1 分割效果最佳.

此外, 由于我们用网络 WheatFCN+1 训练出来的模型相当于一个“黑匣子”, 即只要输入图 3 中 A 样小麦图像, 就可以得到图 3 中 B 图分割结果. 而且只要能保证输入的是同一粒小麦的上下扫描图像, 那么输出就一定是同一粒小麦的上下分割结果, 无需再像模板匹配结合漫水填充的双面联合分割方法一样进行上下配准、漫水填充的操作, 过程更加简便, 简略分割流程如图 9 所示.

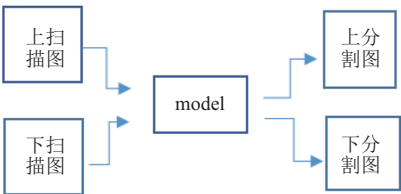


图 9 运用全卷积网络分割的流程

5 结论

对于小麦图像的分割, 项目组所使用的基于模板匹配结合漫水填充的双面联合分割方法与传统的分割方法诸如基于阈值分割、基于区域分割、基于边缘分割相比较, 效果有明显提升, 但是由于小麦情况比较复杂, 因此仍然存在分割效果不佳的现象. 为了进一步改善分割效果, 本文提出基于改进的全卷积网络的分割方法, 融入第二个池化层的输出信息使得到的特征更加全面, 将原来的 FCN 网络的 21 类输出目标改为适用于小麦的 2 类目标, 本文还引入了 Batch normalization 以提高网络的收敛速度和泛化能力. 实验结果表明, 本文提出的基于改进的全卷积网络的语义分割方法在效果上优于传统的分割方法以及模板匹配结合漫水填充的双面联合分割方法, 得到了更好的分割效果.

参考文献

- 1 Felzenszwalb PF, Huttenlocher DP. Efficient graph-based image segmentation. International Journal of Computer Vision, 2004, 59(2): 167–181. [doi: 10.1023/B:VISI.0000022288.19776.77]
- 2 Chen YB, Chen OTC. Image segmentation method using thresholds automatically determined from picture contents. EURASIP Journal on Image and Video Processing, 2009, (2009): 140492.
- 3 Adams R, Bischof L. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641–647. [doi: 10.1109/34.295913]
- 4 王倩, 王正勇, 范艳军, 等. 基于边缘流和区域合并的岩屑颗粒图像分割. 四川大学学报 (自然科学版), 2014, 51(1): 111–118.
- 5 Olson CF. Maximum-likelihood template matching. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head Island, SC, USA. 2000, 2: 52–57.
- 6 陈佳鑫, 贾英民. 一种基于漫水填充法的实时彩色目标识别方法. 计算机仿真, 2012, 29(3): 4–9.

- 7 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for scene segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481–2495. [doi: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615)]
- 8 Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. [doi: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683)]
- 9 Shen XY, Hertzmann A, Jia JY, *et al.* Automatic portrait segmentation for image stylization. *Computer Graphics Forum*, 2016, 35(2): 93–102. [doi: [10.1111/cgf.12814](https://doi.org/10.1111/cgf.12814)]
- 10 Konečný J, Liu J, Richtárik P, *et al.* Mini-batch semi-stochastic gradient descent in the proximal setting. *IEEE Journal of Selected Topics in Signal Processing*, 2016, 10(2): 242–255. [doi: [10.1109/JSTSP.2015.2505682](https://doi.org/10.1109/JSTSP.2015.2505682)]
- 11 Everingham M, Van Gool L, Williams CKI, *et al.* The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 2010, 88(2): 303–338. [doi: [10.1007/s11263-009-0275-4](https://doi.org/10.1007/s11263-009-0275-4)]
- 12 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*: 1502.03167v3, 2015.
- 13 Martin DR, Fowlkes CC, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(5): 530–549. [doi: [10.1109/TPAMI.2004.1273918](https://doi.org/10.1109/TPAMI.2004.1273918)]