

基于改进 DeepLabV3+的铁谱图像语义分割

程亮,王静秋,刘信良,孙康

(南京航空航天大学 机电学院,江苏 南京 210016)

摘 要:针对传统铁谱图像分割方法需要人工设计特征、存在特征冗余及误差累积等缺点,提出一种改进的 DeepLabV3+方法,实现了对铁谱图像中严重滑动、疲劳、切削等 6 种磨粒的语义分割。该方法在 DeepLabV3+基础上,引入密集连接的空洞空间池化金字塔模块,以增大感受野;通过改进密集金字塔模块,采用互质膨胀率的空洞卷积,避免了栅格效应;采用全连接条件随机场以改善分割结果。实验结果表明:该方法在铁谱图像测试集上的平均交占比为 87.1%。

关键词:铁谱图像;语义分割;密集空间池化金字塔;条件随机场

中图分类号:TP391.41 **文献标志码:**B **文章编号:**1671-5276(2021)06-0103-03

Semantic Segmentation of Ferrograph Images Based on Improved DeepLabV3+

CHENG Liang, WANG Jingqiu, LIU Xinliang, SUN Kang

(College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: To overcome the shortcomings of traditional ferrograph images segmentation methods such as artificial design features, feature redundancy and error accumulation, an improved DeepLabV3+ method is proposed to realize the semantics of six kinds of wear particles like severe sliding, fatigue and cutting particles in ferrograph images segmentation. The proposed method based on DeepLabV3+ introduces densely connected spatial pooling pyramid module to increase the receptive field. By improving the dense spatial pooling pyramid module, atrous convolution with co-prime dilation is adopted to avoid grid effect. Conditional random field fully connected is applied to improve the segmentation results. Experimental results show that the mean intersection over union of the improved method on the test dataset of ferrograph images is 87.1%.

Keywords: ferrograph image; semantic segmentation; dense spatial pooling pyramid; conditional random field

0 引言

铁谱分析是在高梯度强度的磁场作用下,将摩擦副中产生的金属磨粒从润滑油中分离出来,按照一定的方法制成谱片和铁谱图像,然后通过相关软件对其进行定性和定量分析的方法^[1]。由于铁谱图像中磨粒尺寸不一,相同类型磨粒之间形态各异以及不同类型磨粒之间存在一定的相似性等问题,使用计算机图像处理技术对其进行准确分割较为困难。图像语义分割是对图像中每个像素进行分类,并用颜色索引表示其类别的方法。随着全卷积网络(fully connectional networks, FCN)^[2]的出现,诸如 DeepLab 系列、基于编码-解码、基于注意力机制以及基于概率图模型等语义分割方法被相继提出,极大地推进了图像语义分割的发展^[3]。

针对 FCN 对细节不敏感、缺乏空间一致性等缺点,2016 年,CHEN L C 等人提出了 DeepLab 网络^[4],该网络将分割结果输入至全连接条件随机场,解决了深度网络的定位效果差分割精度不足的问题。2017 年,CHEN L C 等人在 DeepLab 的基础上提出了 DeepLabV2 网络^[5]。该网络采用空洞卷积,可以在不降低图像分辨率情况下增大感

受野;其次,增加空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块,利用多个不同膨胀率的空洞卷积捕获多尺度特征。之后,CHEN L C 等人又提出了 DeeplabV3 网络^[6],该网络在 ASPP 中加入全局平均池化分支并在 ASPP 模块后使用批量归一化层,有效捕获全局信息。2018 年,该团队针对 DeepLabV3 池化和带步长卷积会造成一些物体边界细节信息的丢失并且扩张卷积计算代价过大的问题,提出了 DeepLabV3+网络。该网络在 DeepLabV3 的基础上增加了编码-解码模块,将 DeepLabV3 作为网络的编码器,并在此基础上增加了解码器模块用于恢复目标边界细节^[7-9]。

本文通过改进 DeepLabV3+网络,提出一种适用于铁谱图像语义分割的模型,如图 1 所示。该模型改进 DeepLabV3+ 的 ASPP 模块为密集连接的 ASPP 结构(Dense ASPP, D-ASPP),以更好地提取空间信息;然后引入全连接条件随机场(fully connected conditional random filed, FCCRF)改善铁谱图像中磨粒过分割问题。

1 改进 DeepLabV3+网络

在 DeepLabV3+网络中,基于空洞卷积的 ASPP 模块

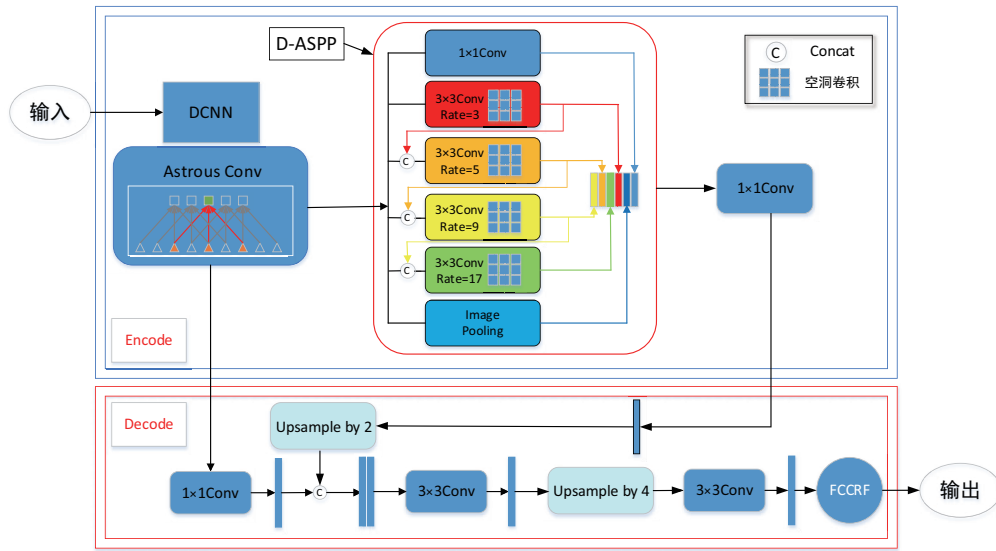


图1 改进 DeepLabV3+网络示意图

虽然可以在不降低图像分辨率的情况下提取特征,从而提升网络的分割精度,但是对于尺度不一物体的分割,模型可能会忽略较小的物体,而对于尺度较大的物体,又有可能超过感受野范围,无法获取物体全局信息^[3]。为了解决上述问题,本文引入并改进 D-ASPP 结构,以获取更丰富的语义特征。

1.1 D-ASPP

D-ASPP 结构是在原 ASPP 结构中加入级联的连接方式,使其在增大模型整体感受野的同时还能共享各分支信息。另外,本文改进了 D-ASPP 结构,增加了 $R_{\text{ate}}=3$ 的分支,用于小范围局部特征的提取;其次,将 D-ASPP 结构中各分支膨胀率设为互质数,防止出现栅格效应。改进后的 D-ASPP 具有更多的像素参与计算和更大的感受野的优点。

1)更多的像素参与计算。图2为在级联方式下,空洞卷积的计算过程示意图,其中灰色结点为参与计算的像素点,黑色结点为不参与计算的像素点。

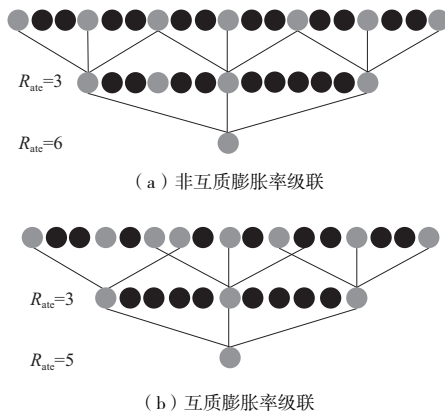


图2 空洞卷积级联示意图

从图2(a)中可以看出,如果以非互质膨胀率级联,参与计算的像素都集中在某些固定的像素点上,即栅格效应。因

此为了让更多的像素点参与计算,本文在 ASPP 结构中采用了互质膨胀率,即假设上层的膨胀率为 N ,则下层采用膨胀率为 $(2 \times N - 1)$,依次递推,其示意图如图2(b)所示。

2)更大的感受野。为了解决下采样时图像分辨率降低的问题,DeepLabV2 引入了空洞卷积。通过改变膨胀率,即可在不改变图像分辨率的情况下增大卷积核的感受野。空洞卷积中感受野计算公式如下:

$$R = (D - 1) \cdot (K - 1) + K \quad (1)$$

其中: K 表示卷积核尺寸; D 表示空洞卷积膨胀率。而当两个空洞卷积级联时,下层空洞卷积的感受野为

$$R = R_1 + R_2 - 1 \quad (2)$$

其中 R_1 和 R_2 为相邻两个空洞卷积的感受野。表1为感受野对比。

表1 感受野对比

ASPP 结构	普通连接	密集连接
(3,5)	11	17
(3,5,9)	19	35
(3,5,9,17)	35	69
(3,5,9,17,33)	67	135

从表1可以看出,相对于普通的 ASPP 结构,D-ASPP 结构的感受野近乎两倍于前者,能够更好地获取全局信息,这更有利于解决局部模糊性问题。

1.2 全连接的条件随机场

在图像的语义分割网络中,目标是单个像素点,因此会忽略考虑像素之间的关系。本文将全连接条件随机场融入改进的 DeepLabV3+模型,利用像素之间的相似性,改善分割结果。

为了将单个像素观测信息与周围环境进行建模,并考虑相邻像素之间的一致性^[10],在 FCCRF 模型中分别以 $\theta_i(x_i) = -\log P(x_i)$ 和 $\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K \omega_m \cdot k_m(f_i, f_j)$ 作为一阶和二阶势函数。其中随机变量 x_i 为某个

类别的代价, $P(x_i)$ 是经过 softmax 函数得到的概率值, ω_m 为高斯核权值, k_m 为特征核函数。 $\mu(x_i, x_j)$ 为一致性参数。如果 $x_i \neq x_j$, 则 $\mu(x_i, x_j) = 1$, 否则为 0, 这意味着只惩罚分配着不同标签的像素点。

对于多分类任务, 利用颜色和位置构建特征核函数, 即

$$k(f_i, f_j) = \omega_1 \exp\left(-\frac{|P_i - P_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right) + \omega_2 \exp\left(-\frac{|P_i - P_j|^2}{2\theta_\gamma^2}\right) \quad (3)$$

其中: I 为 3 维颜色向量; P 为 2 维像素点坐标向量; θ_α 、 θ_β 、 θ_γ 为高斯核尺度参数^[11]。

如图 3 所示, 在 FCCRF 模型中, 通过比较相邻像素的一致性, 鼓励相似性大的像素分配相同的标签, 而相似性小的像素分配不同的标签, 因此对于较小的空洞或者是像素分配错误的小区域都可以通过相邻区域的标签进行填充或者修正。

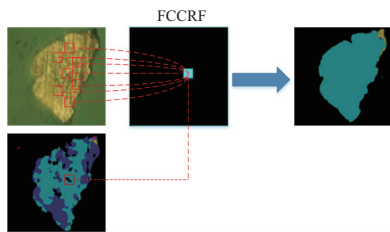


图 3 FCCRF 示意图

2 实验过程及结果分析

2.1 实验设备和数据

本文算法所采用的编程语言为 Python3.6, 网络搭建基于 Tensorflow 开源框架, 版本为 1.14。硬件配置如下: CPU 型号为 i7-8700k, 显卡型号为 NVIDIA RTX2070、显存为 8 GB, 操作系统为 Windows10。本文所采用的实验数据来自某公司油液, 经取样、制谱、拍摄得到的实际应用图像。图像总数为 200 张, 其中训练集 160 张, 验证集 20 张, 测试集 20 张, 其对应标签通过 LabelMe 标注工具进行标注。本文数据集中包含链状磨粒、球形磨粒、切削磨粒、疲劳磨粒、氧化物磨粒以及严重滑动磨粒共 6 类。

2.2 评价指标

本文所采用评价指标为平均较占比 (mean intersection over union, MIoU) 以及平均像素准确率 (mean pixel accuracy, MPA)。假设图像中共有 k 个类别 (不考虑背景), p_{ij} 表示真实值为 i 、预测值为 j ; p_{ji} 表示真实值为 j 、预测值为 i ; 而 p_{ii} 表示真实值和预测值均为 i 。 M_{IoU} 和 M_{PA} 的计算公式分别为:

$$M_{IoU} = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{\sum_{j=1}^k p_{ij} + \sum_{j=1}^k p_{ji} - p_{ii}} \quad (4)$$

$$M_{PA} = \frac{1}{k} \sum_{i=1}^k \frac{p_{ii}}{\sum_{j=1}^k p_{ij}} \quad (5)$$

2.3 实验过程与结果分析

1) ASPP 模块实验

为验证不同膨胀率组合的效果, 本文设计多组 ASPP 结构, 分别以是否级联、是否加入 $R_{ate}=3$ 的分支以及是否采用互质膨胀率为对比项, 表 2 为对比实验结果。

表 2 不同膨胀率组合对比实验

序号	ASPP 结构	是否级联	M_{PA}	M_{IoU}
1	(6, 12, 18)	否	80.9	74.9
2	(3, 6, 12, 18)	否	84.4	77.8
3	(5, 9, 17)	否	83.0	76.9
4	(5, 9, 17)	是	84.1	77.4
5	(3, 5, 9, 17)	否	84.5	78.3
6	(3, 5, 9, 17)	是	85.4	78.8

从表 2 中的结果可以看出, 序号 6 中采取 (3, 5, 9, 17) 的互质膨胀率组合、级联、加入膨胀率为 3 的分支结果最好, M_{PA} 为 85.4, M_{IoU} 为 78.8。

2) FCCRF 后处理

由于语义分割是针对像素分类, 忽略了像素之间的关系, 因此分割结果中可能存在过分割。为了解决这一问题, 本文在改进的 DeepLabV3+ 基础上, 将原始图像和预测结果作为输入, 采用 FCCRF 进行后处理, 以颜色信息和位置信息为约束修正像素分类 (图 4)。

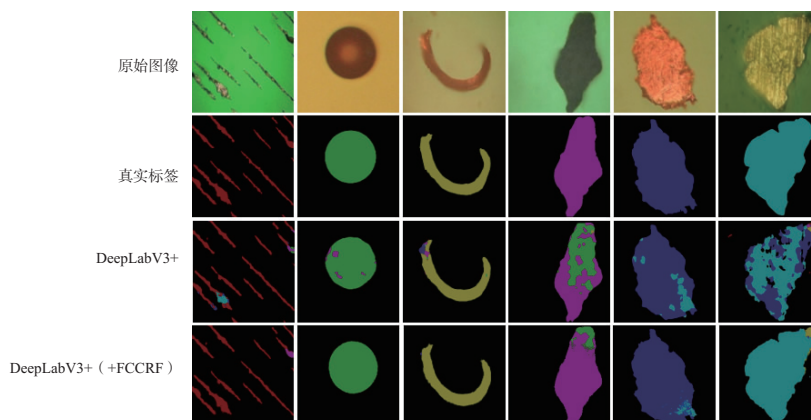


图 4 语义分割结果

(下转第 122 页)

真结果的误差较小,基本吻合,可为挖掘机的疲劳分析研究提供一定依据。

参考文献:

- [1] 王敏. 液压挖掘机工作装置的动态仿真及有限元分析[D]. 西安:长安大学,2015.
- [2] 王其军. 基于真实载荷驱动的液压挖掘机工作装置动力学仿真与优化[D]. 西安:长安大学,2018.
- [3] 刘厚茂. 液压挖掘机反铲工作装置动态仿真研究[D]. 石家庄:石家庄铁道大学,2015.
- [4] 同济大学. 单斗液压挖掘机[M]. 2版. 北京:中国建筑工业出版社,1986.

- [5] 马治宁. 单斗液压挖掘机工作装置疲劳寿命分析研究[D]. 太原:太原科技大学,2017.
- [6] 董芑羽. 挖掘机工作装置的动力仿真及有限元分析[D]. 天津:河北工业大学,2011.
- [7] 李志勇. 挖掘机工作装置力学分析系统开发与斗杆载荷谱编制[D]. 西安:长安大学,2016.
- [8] 崔跃玉. 挖掘机动臂载荷谱试验研究及其疲劳寿命估算[D]. 西安:长安大学,2016.

收稿日期:2020-10-26

(上接第105页)

从图4中可以看出,经过 FCCRF 后处理过的结果,磨粒内部的孔被正确填充,而且分类错误的像素也大多得到修正,很大程度地改善了过分割现象。

3) 实验结果对比

为了验证本文算法的有效性,在本文实验数据集上,将实验结果与其他几种经典分割网络 DeepLabV3、RefineNet^[12]、PSPNet^[13]进行对比,对比结果如表3所示。

表3 实验结果对比

方法	M_{PA}	M_{IoU}
RefineNet	72.9	68.3
PSPNet	76.2	69.0
DeepLabV3	77.5	68.6
DeepLabV3+	80.4	75.2
改进 DeepLabV3+	85.4	78.8
改进 DeepLabV3+(+FCCRF)	90.3	87.1

从表3结果可以看出,在本文数据集上,改进后的 DeepLabV3+的结果优于其他网络,并且加入 FCCRF 可以获取更加精确的结果, M_{PA} 为90.3, M_{IoU} 为87.1。

3 结语

本文在 DeepLabV3+的基础上,引入并改进了D-ASPP结构,在增大感受野的同时避免了栅格效应。其次,利用全连接条件随机场,以颜色信息和位置信息为约束,优化了语义分割忽略像素之间相关性的问题,提高了分割精度。实验结果表明,本文所用方法能够获得较为精确的分割结果,具有一定的实用性。

参考文献:

- [1] 赵畅畅,杨智宏,何伟楚,等. 铁谱定性分析技术及其在水泥行业油液监测中的应用[J]. 水泥,2020(7):23-25.
- [2] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4):

640-651.

- [3] 戴伟达. 基于全卷积神经网络的语义分割算法研究[D]. 南京:南京邮电大学,2019.
- [4] 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, 2018, 40(4): 834-848.
- [5] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[EB/OL]. 2016;arXiv: 1606.00915[cs.CV].
<https://arxiv.org/abs/1606.00915>.
- [6] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. 2017;arXiv:1706.05587[cs.CV].
<https://arxiv.org/abs/1706.05587>.
- [7] 景庄伟,管海燕,彭代峰,等. 基于深度神经网络的图像语义分割研究综述[J]. 计算机工程,2020,46(10):1-17.
- [8] 刘致臻,孙韶媛,任正云,等. 基于改进 DeepLabv3+的无人车夜间红外图像语义分割[J]. 应用光学,2020,41(1): 180-185.
- [9] 罗会兰,张云. 基于深度网络的图像语义分割综述[J]. 电子学报,2019,47(10):2211-2220.
- [10] 孙莹莹. 基于多尺度 FCN-CRF 和强化学习的高分辨 SAR 图像语义分割[D]. 西安:西安电子科技大学,2018.
- [11] 刘卫铭,辛月兰,姜星宇. 一种融合 FC-CRF 的密集网络语义分割方法[J]. 电子设计工程,2020,28(15):43-47.
- [12] LIN G S, MILAN A, SHEN C H, et al. RefineNet: multi-path refinement networks for high-resolution semantic segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017:5168-5177.
- [13] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017:6230-6239.

收稿日期:2020-10-28