引文格式：龚良雄,谢仁平,王红根,等. 顾及交叉注意力的高分辨率遥感影像果园分割算法[J]. 地理空间信息, 2024, 22(12): 60-64. doi:10. 3969/j.issn. 1672-4623.2024.12.014

顾及交叉注意力的高分辨率遥感影像果园分割算法

龚良雄[1]，谢仁平[2*]，王红根[1]，刘传瑞[1]
（1. 南昌市测绘勘察研究院有限公司，江西 南昌 330013；
2. 东莞理工学院 计算机科学与技术学院，广东 东莞 523808）

摘　要：针对高分辨率遥感影像中果园类内特征丰富、果园与其他地类类间特征相似导致的果园提取效果不理想、分割精度低等问题，以DeepLabV3+语义分割模型为基础，在果园特征提取阶段参考特征金字塔结构，将ResNetSt作为模型骨干网络，并引入CCNet的交叉注意力机制，以加强模型跨特征提取及捕获全图像上下文信息的能力，从而改善果园的边缘提取效果并提高分割精度。利用南昌市部分县区年度变更调查数据库中的果园数据集，对所提出ResNetSt-DeepLabV3+网络模型进行了果园提取训练和验证实验。实验结果表明，基于该模型的果园分割精度优于原DeepLabV3+和其他常用语义分割模型，分割精度达到了78.63%。选取东湖区扬子洲镇和新建区金桥乡作为现实场景果园提取实验区域，模型的分割精度分别为77.16%和76.97%，与U-Net、PSPNet和DeepLabV3+语义分割模型相比，最小分别增长了5.86%、6.95%和4.51%。本研究所提出的ResNetSt-DeepLabV3+模型能够更准确地从高分辨率遥感影像中获取果园区域信息，有效提高了果园的分割精度。

关键词：DeepLabV3+；ResNetSt；交叉注意力；高分辨率遥感影像
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）12-0060-05

High-resolution Remote Sensing Image Orchard Segmentation Algorithm Considering Criss-cross Attention

GONG Liangxiong[1], XIE Renping[2], WANG Honggen[1], LIU Chuanrui[1]
(1. Nanchang Institute of Surveying and Mapping Co., Ltd., Nanchang 330013, China;
2. School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, China)

Abstract: Aiming at the problems of poor extraction effect and low segmentation accuracy caused by rich features within orchard and similar features between orchard and other land-use types from high-resolution remote sensing images, based on DeepLabV3+ semantic segmentation model, referring to the feature pyramid structure, we took ResNetSt as the backbone network of model, and introduced the criss-cross attention mechanism of CCNet in the orchard feature extraction stage, thereby enhancing the model's ability of extracting cross-feature and capturing full image context information, so as to improve the edge segmentation effect and segmentation accuracy of orchard. Based on the orchard dataset from the annual change survey database of some counties in Nanchang City, we trained and validated ResNetSt-DeepLabV3+ network model for orchard extraction. The experimental results show that the orchard extraction accuracy based on this method is superior to the original DeepLabV3+ and other common semantic segmentation models, and the segmentation accuracy reaches 78.63%. Selecting Yangzizhou Town in Donghu District and Jinqiao Township in Xinjian District as experimental areas to extract orchards in real scenes, the segmentation accuracy is 77.16% and 76.97%, respectively. Compared with U-Net, PSPNet, and DeepLabV3+ semantic segmentation models, the minimum increases are 5.86%, 6.95% and 4.51%, respectively. The ResNetSt-DeepLabV3+ model can obtain the orchard region information from high-resolution remote sensing images more accurately, and effectively improve the segmentation accuracy of orchard.

Key words: DeepLabV3+, ResNetSt, criss-cross attention, high-resolution remote sensing image

果园是农产品生产的重要要素，关系到国计民生和社会稳定，同时也兼具水土资源保持、保护生物多样性、维持生态平衡等环境保护功能[1]。目前，主要通过影像目视判读结合实地调查方式查清果园的土地利用分布情况（位置、面积），其缺点是时效性差、成本高、数据获取效率低，而遥感技术具有覆盖率广、探测周期短、手段丰富、条件限制少和成本低等特点[2]。如何利用高分辨率遥感影像准确地、快速地

收稿日期：2023-08-19。
项目来源：国家自然科学基金资助项目（62001113）；广东省基础与应用基础研究基金资助项目（2019A1515110800）。
第一作者简介：龚良雄（1991—），硕士，注册测绘师，主要从事遥感智能解译、地理信息技术方面的工作，E-mail：1021386774@qq.com。
通信作者：谢仁平（1989—），博士，副研究员，主要从事桥梁图像分割、遥感图像检测方面的工作，E-mail：xierenping@dgut.edu.cn。

提取果园信息，对果园的精准化动态监测、区域农业经济的可持续发展及环境保护具有重要意义[3-8]。

目前研究在智慧农业领域取得了一定成效，但基于深度学习和高分辨率遥感影像的果园语义分割方法及应用较少[9-15]。为改善果园提取效果和提高分割精度，针对基于深度学习的高分辨率遥感图像果园信息提取，本文以DeepLabV3+模型为基础，提出了一种顾及交叉注意力的ResNetSt-DeepLabV3+高分辨率遥感影像果园分割算法。

1　DeepLabV3+网络简介

DeepLabV3+是一个端到端的语义分割模型，其目的是识别物体的类别和位置，并找到属于该类别的所有像素，它可以直接输出语义分割后的预测结果。DeepLabV3+采用了包含空洞卷积的编码器-解码器结构，其整体网络框架如图1所示。

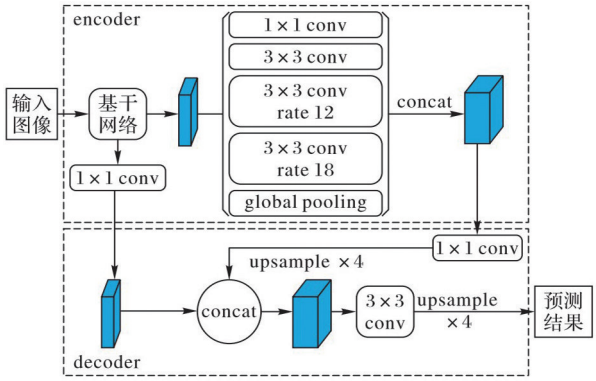


图1　DeepLabV3+网络结构示意图

2　研究方法

2.1　骨干网络的改进

传统的ResNet网络往往只能使用相对单一的局部特征进行分割，无法获得图像中的全局特征进行分割，导致分割效果不太理想，而ResNetSt网络包含一种可以获得不同特征交互信息的分离注意力机制（Split Attention）。因此本文在DeepLabV3+基础上将骨干特征提取网络Xception替换为ResNetSt，以此来增强模型区分果园与其他地物特征的能力。采用特征金字塔结构对深层特征图进行上采样，在保留深层图像语义信息的同时，获得更多浅层图像的位置信息，有助于网络模型在处理果园周边狭窄、细小地物时获得更好的分割效果。

首先，ResNetSt模块将输入图像（$h\times w\times c$）的特征图分解为k个特征图组，其次将每个特征图组沿通道维度划分为r个特征图小组，再对每个特征图小组进行1×1卷积，将通道数减少至（c'/k）/r个，然后使用3×3卷积将通道数增加到c'/k个。最后各特征图小组进入分离注意力机制模块，先按公式（1）将每个特征图组内的特征图小组进行加和，从而获得每个特征图组的融合特征。

$$\hat{U}^k=\sum_{i=r(k-1)+1}^{rk}U_i \tag{1}$$

式中，U_i为第k个特征图组下第i个特征图小组输出。

其次对具有c个通道的第k个特征图组$\hat{U}_c^k$进行全局平均池化，可收集跨空间维度的全局上下文信息，如公式（2）所示。

$$S_c^k=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\hat{U}_c^k(i,j) \tag{2}$$

式中，S_c^k为具有c个通道的第k个特征图组的特征图输出。然后使用S_c^k计算第k个特征图组下每个特征图小组的权重比例，如公式（3）所示。

$$a_i^k(c)=\begin{cases}\dfrac{\exp(G_i^c(S^k))}{\sum_{j=1}^{r}\exp(G_j^c(S^k))} & \text{if } r>1,\\ \dfrac{1}{1+\exp(-G_i^c(S^k))} & \text{if } r=1,\end{cases} \tag{3}$$

式中，$G_i^c(S^k)$为第k个特征图组下第i个特征图小组的权重。将特征图组下所有的特征图小组进行权重融合，得到每个特征图组下的融合特征图输出，如公式（4）所示。

$$V_c^k=\sum_{i=1}^{r}a_i^k(c)U_{r(k-1)+i} \tag{4}$$

最后，将各个特征图组得到的融合特征图按通道维度进行拼接，通过累加残差块的输入可以得到残差块的输出，如公式（5）所示。

$$Y=X+\text{Concat}\{V^1,V^2,V^3,\ldots,V^k\} \tag{5}$$

式中，Y为残差块的输；X为残差块的输入。

2.2　交叉注意力机制

在现实场景果园分割任务中，果园内种植了各种各样的果树（如橘子树、葡萄树等），从而会在遥感图像上产生各种各样的果园样本，仅通过局部特征信息无法准确地对果园进行分割。因此本文引入CCNet中的交叉注意力机制，加强了特征图中像素点的直接相关性，可有效解决在复杂样本情况下果园提取效果不理想的问题。

传统的自注意力机制可由公式（6）表示。

$$y_i=\frac{1}{C(x)}\Sigma_{\forall j}f(x_i,x_j)g(x_j) \tag{6}$$

式中，x_j 为输入图像的每个像素；$f(x_i,x_j)$ 为2个像素之间的相关性；$g(x_j)$ 为对 x_j 进行像素增强，这种机制的缺点是计算量相对较大。针对此问题，本文使用了CCNet模块中的交叉注意力模块，只需使用交叉模块2次即可获得全局空间的自注意力特征信息。

在第一次计算注意力机制时，特征图上每个位置的交叉特征信息（以蓝色点为例）$K(i,j)$ 由公式（7）计算得到 $\hat{K}(i,j)$。

$$\hat{K}(i,j)=K(i,j)\sum_{\forall h}\sum_{\forall w}f(K(i,j),K(h,w))K(i,j) \tag{7}$$

式中，h 和 w 为图像的高和宽。经过了第一次交叉注意力计算后，可得到每个像素与特征图中同行同列像素之间的自注意力关系。第二次交叉注意力 $\bar{\bar{K}}(i,j)$ 是以 $\hat{K}(i,j)$ 为基础，通过公式（8）计算得到。经过了第二次交叉注意力计算后，可得到每个像素与特征图中所有像素之间的自注意力关系。

$$\bar{\bar{K}}(i,j)=\hat{K}(i,j)\sum_{\forall h}\sum_{\forall w}f(\hat{K}(i,j),\hat{K}(h,w))\hat{K}(i,j) \tag{8}$$

2.3 ResNetSt-DeepLabV3+分割算法

本文提出的ResNetSt-DeepLabV3+分割算法整体流程为：首先输入图像经过骨干特征提取网络ResNetSt获得不同特征的交互信息，利用4个ResNetSt块进行4次下采样获得图像的多尺度特征，将多尺度特征与特征金字塔上采样得到的特征图进行通道拼接，使得各尺度上的全局特征与局部特征学习到不同的依赖权重；然后将拼接后的特征图进行双分支的特征提取，其中一个分支利用交叉注意力机制提取全局特征，另一个分支则利用不同采样率的空洞卷积提取不同感受野的局部特征；在综合考虑全局和局部特征信息的条件下，网络通过解码器对每个像素进行背景与果园的分类，算法的解码器直接采用了DeepLabV3+的解码器；最后，将融合得到的特征图上采样4倍得到最终的预测图。

3 实验和应用

3.1 数据集和预训练

为验证本文算法的有效性和准确性，利用部分县区年度变更调查数据库中的果园矢量数据集及相应的亚米级高分辨率遥感影像成果，制作了620张512×512分辨率的果园标注图像，再对图像进行随机的旋转、平移、拉伸、添加噪声等数据扩充，最终获得2 480张果园标注图像。模型的验证集和训练集按照2∶8比例随机划分，验证集和训练集的数量分别为496张、1 984张。

实验的硬件环境为Intel(R) Core(TM) i7-12700H CPU、NVIDIA GeForce RTX 3060 Ti GPU，深度学习框架采用PyTorch 1.10.2，网络参数配置如下：权重衰减系数为0.001，动量为0.9，训练次数为100，样本批处理数量为20，学习率初始值设为0.001，采用分步学习策略，使用交叉熵和Lovasz hinge作为损失函数，且损失函数的权重相同。

3.2 评价指标

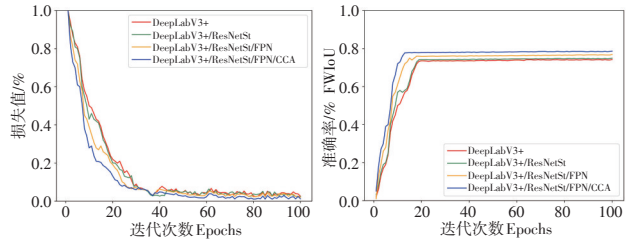
采用的评价指标为频权交并比FWIoU，根据每一个类别出现的频率设置其权重。

$$\mathrm{FWIoU}=\frac{1}{\sum_{i=0}^{N}\sum_{j=0}^{N}S_{ij}}\sum_{i=0}^{N}\frac{\sum_{j=0}^{N}S_{ij}S_{ii}}{\sum_{j=0}^{N}S_{ij}+\sum_{j=0}^{N}S_{ji}-S_{ii}} \tag{9}$$

式中，N 为样本的类别数；S_{ij} 为第 i 类样本预测为第 j 类的样本数。

3.3 消融实验

为验证DeepLabV3+模型所引入的ResNetSt、特征金字塔（FPN）以及交叉注意力（CCA）对高分遥感影像果园信息提取的促进作用，采用了逐次添加上述模块的方式进行消融实验。各模型的训练损失曲线和FWIoU曲线变化如图2所示，可以看出随着迭代次数的增加，DeepLabV3+/ResNetSt/FPN/CCA模型不仅训练损失值减小最快，而且准确率增加最快。说明了DeepLabV3+模型引入上述模块后能加快模型的收敛速度，同时也提高了模型的准确率。各模型消融实验结果如表1所示。



a 损失曲线　　b FWIoU曲线

图2 各模型训练损失曲线和FWIoU曲线

表1 消融实验结果/%

模型	分辨率	FWIoU
DeepLabV3+	512×512	74.26
DeepLabV3+/ResNetSt	512×512	74.91
DeepLabV3+/ResNetSt/FPN	512×512	76.85
DeepLabV3+/ResNetSt/FPN/CCA	512×512	78.63

消融实验结果表明，引入特征金字塔结构后的模型最大改进可达到1.94%，说明引入特征金字塔得到

的多尺度特征对果园提取是有效的。最后通过引入交叉注意力机制，模型的最终FWIoU可提高到78.63%，验证了各模块在高分辨率遥感影像果园提取任务中的有效性。

3.4　模型比对分析

为进一步验证本文算法对高分辨率遥感图像果园信息提取的准确性，在相同的训练参数、验证集条件下，将其与经典语义分割模型U-Net、PSPNet、DeepLabV3+进行比对实验。不同语义分割模型的评价指标比对结果如表2所示。

表2　预测比对结果

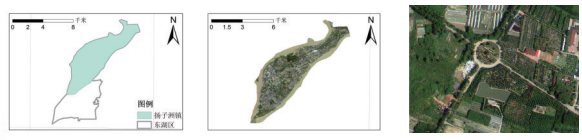
模型	分辨率	FWIoU/%
U-Net	512×512	72.34
PSPNet	512×512	71.21
DeepLabV3+	512×512	74.26
本文方法	512×512	78.63

对于高分辨率遥感影像果园提取任务，U-Net、PSPNet和DeepLabV3+模型的FWIoU分别只有72.34%、71.21%和74.26%，而ResNetSt-DeepLabV3+模型的FWIoU为78.63%，相较于DeepLabV3+、PSPNet和U-Net模型分别提升了5.88%、10.42%和8.70%。实验结果表明，在多样本特征及狭窄、小块地物的干扰下，本文算法能有效区分果园特征及非果园特征，所提取的果园信息更加全面、准确。

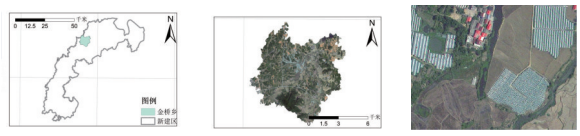
3.5　模型应用

经消融实验和模型比对分析可知，ResNetSt-DeepLabV3+模型相较于其他网络模型取得了更好的果园提取效果。为进一步验证该模型在高分遥感影像果园场景中的实际分割效果，选取了果园与其他地类类间特征相似的东湖区扬子洲镇、果园类内特征丰富的新建区金桥乡作为现实场景果园提取实验区域，扬子洲镇、金桥乡的地理位置、高分遥感影像及果园局部影像如图3所示，然后将该模型与其他语义分割模型进行面积提取精度与评价指标FWIoU比对分析，基于ResNetSt-DeepLabV3+模型提取的扬子洲镇、金桥乡果园空间分布如图4所示，面积提取精度与评价指标比对分析如表3所示。

从表3结果可得出，因扬子洲镇果园与其他地类类间特征相似，故各模型将林地、蔬菜大棚提取出来，导致扬子洲镇的果园提取面积均比目视解译面积大。随着ResNetSt、FPN模块的引入，模型的提取面积减幅最大，说明了在果园与其他地类类间特征相似的情况下，ResNetSt、FPN可提高果园提取精度。与之相反，金桥乡各模型的提取面积均比目视解译面

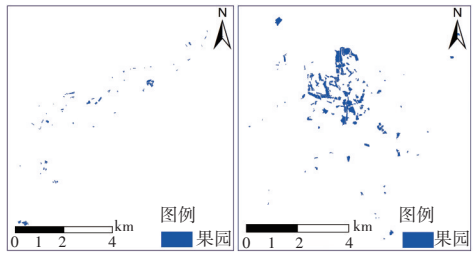


a 扬子洲镇地理位置　b 扬子洲镇高分遥感影像　c 果园局部影像



d 金桥乡地理位置　e 金桥乡高分遥感影像　f 果园局部影像

图3　扬子洲镇、金桥乡地理位置、高分影像及局部影像图



a 东湖区扬子洲镇　　b 新建区金桥乡

图4　果园提取结果

积小，是由于金桥乡果园类内特征比较丰富，导致各模型漏提了部分种植西瓜、橘子的果园。但随着CCA模块的引入，模型提取面积增幅最大，说明了在果园类内特征丰富的情况下，CCA模块可提高果园提取精度。

表3　面积提取精度与评价指标比对表

模型	扬子洲镇			金桥乡		
	提取面积/hm²	目视解译面积/hm²	FWIoU/%	提取面积/hm²	目视解译面积/hm²	FWIoU/%
U-Net	26.84	22.96	72.89	150.34	170.16	71.36
PSPNet	27.31		72.14	148.72		71.93
DeepLabV3+	26.71		73.22	152.64		73.65
DeepLabV3+/ResNetSt	25.05		75.36	152.92		73.97
DeepLabV3+/ResNetSt/FPN	23.82		77.15	153.56		74.48
DeepLabV3+/ResNetSt/FPN/CCA	23.80		77.16	159.61		76.97

4　结　语

为改善高分辨率遥感影像果园信息提取效果，本文提出了一种顾及交叉注意力的ResNetSt-DeepLabV3+高分辨率遥感影像果园分割算法。该算法将ResNetSt作为骨干特征网络，增强了模型区分果园与其他地物特征的能力，利用特征金字塔结构捕获特征图像的多尺度特征信息，改善了传统DeepLabV3+网络难以识别果园周边狭窄、小块地物的缺陷；引入CCNet的交叉注意力机制，使网络能够捕获全图像的上下文信息，而不是单纯依靠局部特征来判断是否是果园，有效解决了

实际场景中存在果园多样本特征的问题，从而提高果园提取的准确率。实验结果表明，在果园类内特征丰富、果园与其他地类特征相似的情况下，本文算法在评价指标FWIoU分数上较其他语义分割模型有所改进，提取的果园结构更加完整、边缘更加准确，为实现果园动态监测、管理和智慧生产提供了技术支撑。

参考文献

[1] 陈学森，王艳廷. 力推果园生草 建设生态文明[J]. 烟台果树，2014(1):1-4

[2] 宋荣杰. 基于高分辨率遥感影像的果园与设施农业用地提取方法研究[D]. 咸阳：西北农林科技大学，2019

[3] 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

[4] 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, Cham, 2015:234-241

[5] Zhao H S, Shi J P, QI X J, et al. Pyramid Scene Parsing Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:6 230-6 239

[6] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018:801-818

[7] Liu Z, Lin Y T, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021

[8] Liu Z, Hu H, Lin Y T, et al. Swin Transformer V2: Scaling up Capacity and Resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022

[9] 张蕊，李锦涛. 基于深度学习的场景分割算法研究综述[J]. 计算机研究与发展，2020，57(4):859-875

[10] 胡金梅，董张玉，杨学志. 一种面向对象的高分辨率遥感影像信息提取方法[J]. 地理空间信息，2021，19(9):10-13

[11] 苏腾飞，李瑞平. 面向对象的农田信息遥感影像分割算法[J]. 测绘科学，2016，41(3):49-53

[12] 苏腾飞，张圣微，李洪玉. 基于超像素MRF的农田地区高分遥感影像分割[J]. 国土资源遥感，2018，30(1):37-44

[13] 陈玲玲，施政，廖凯涛，等. 基于卷积神经网络的高分遥感影像耕地提取研究[J]. 农业机械学报，2022，53(9):168-177

[14] 刘尚旺，张杨杨，蔡同波，等. 基于改进PSPnet的无人机农田场景语义分割[J]. 灌溉排水学报，2022，41(4):101-108

[15] 任鸿杰，刘萍，岱超，等. 改进DeepLabV3+网络的遥感影像农作物分割方法[J]. 计算机工程与应用，2022，58(11): 215-223

（上接第45页）

参考文献

[1] 石雪. 基于层次化混合模型的高分辨率遥感影像分割方法研究[J]. 测绘学报，2023，52(1):168

[2] 吴亚楠，郭长恩，于东平，等. 基于不确定性分析的遥感分类空间分层及评估方法[J]. 地球信息科学学报，2022，24(9):1 803-1 816

[3] 李轶鲲，杨洋，杨树文，等. 耦合模糊C均值聚类和贝叶斯网络的遥感影像后验概率空间变化向量分析[J]. 自然资源遥感，2021，33(4):82-88

[4] 李玉，李天惠，赵泉华. 遥感影像直觉模糊集分割法[J]. 测绘学报，2023，52(3):405-418

[5] 张成坤，韩敏. 基于边缘保持滤波的高光谱影像光谱-空间联合分类[J]. 自动化学报，2018，44(2):280-288

[6] Lin H, Wang Y, Chen Y, et al. PSO Kernel Fuzzy Clustering Based Cerebrospinal Fluid Evaluation for Multiple Sclerosis Algorithm[J]. Journal of Medical Imaging and Health Informatics, 2020, 10(1):76-80

[7] 曹海春. 边缘保持滤波的遥感影像多特征联合分类[J]. 遥感信息，2021，36(1):50-55

[8] Kang X, Li S, Benediktsson J A. Spectral-spatial Hyperspectral Image Classification With Edge-preserving Filtering[J]. IEEETransactions on Geoscience & Remote Sensing, 2014(5):2 666-2 677

[9] 魏晓琴，刘海军. 改进布谷鸟优化模糊C均值的遥感影像分类算法[J]. 工程勘察，2022，50(8):58-62

[10] 曹浩，韩贤权，黎建洲，等. 基于EMP与混合核SVM的高光谱遥感影像分类[J]. 地理空间信息，2021，19(11):14-18

[11] 张婷慧，宇洁，叶张林，等. 结合粒子群寻优的混合核函数极限学习机分类模型[J]. 测绘科学技术学报，2019，36(1): 56-61

[12] 虞瑶，杨锦，邱洁. 随机子空间和MRF的高光谱遥感集成学习算法[J]. 测绘科学，2022，47(10):161-167

[13] 陈应霞，陈艳，刘丛. 遥感影像融合AIHS转换与粒子群优化算法[J]. 测绘学报，2019，48(10):1 296-1 304

[14] 李楠，朱秀芳，潘耀忠，等. 人工蜂群算法优化的SVM遥感影像分类[J]. 遥感学报，2018，22(4):559-569