

文章编号: 2095-2163(2023)03-0175-07

中图分类号: TP391.41

文献标志码: A

# 基于深度学习的荧光图像成分分类

王爽<sup>1</sup>, 何小海<sup>1</sup>, 滕奇志<sup>1</sup>, 龚剑<sup>2</sup>

(1 四川大学电子信息学院, 成都 610065; 2 成都西图科技有限公司, 成都 610065)

**摘要:** 在石油地质图像分析工作中, 荧光图像分析技术是获取岩石包含的油藏信息的重要方式。目前提取荧光成分的方法是先对荧光图像进行 RGB 阈值分割, 再根据一定的聚类算法利用每个像素点的颜色信息进行成分分类, 此过程需要大量人工交互并且分类结果不够准确。针对这些问题, 本文提出了一种基于深度学习的荧光图像分类算法, 使用 VGG16 结合 UNet 的方法来构建网络, VGG16 网络作为编码器部分进行主干特征提取, UNet 作为解码器部分进行加强特征提取, 并增加注意力机制来防止像素级信息的缺失, 提高特征提取的准确性。通过对多组图像进行实验表明, 该方法可以获得良好的分类效果。

**关键词:** 荧光图像; 图像分类; 深度学习; 注意力机制; VGG16-UNet

## Fluorescence image component classification based on deep learning

WANG Shuang<sup>1</sup>, HE Xiaohai<sup>1</sup>, TENG Qizhi<sup>1</sup>, GONG Jian<sup>2</sup>

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

2 Chengdu Xitu Technology Co., Ltd., Chengdu 610065, China)

**[Abstract]** In the work of petroleum geological image analysis, fluorescence image analysis technology is an important way to obtain reservoir information contained in rocks. At present, the method of extracting fluorescent components is to first perform RGB threshold segmentation on the fluorescent image, and then use the color information of each pixel to classify the components according to a certain clustering algorithm. This process requires a lot of manual interaction and the classification results are not accurate enough. To solve these problems, this paper proposes a fluorescence image classification algorithm based on deep learning, which uses the method of VGG16 combined with UNet to construct the network, VGG16 network as the encoder part for backbone feature extraction, UNet as the decoder part for enhanced feature extraction, and an attention mechanism is added to prevent the loss of pixel-level information and improve the accuracy of feature extraction. Experiments on multiple sets of images show that this method can achieve good classification results.

**[Key words]** fluorescence image; image classification; deep learning; attention mechanism; VGG16-UNet

## 0 引言

在荧光显微技术中, 将赋存于烃源岩及储集岩中的石油统称为发光沥青, 发光沥青可分为油质沥青、胶质沥青及沥青质沥青。其中, 油质沥青发出黄、绿、蓝色的荧光, 胶质沥青的荧光以橙色为主, 沥青质沥青以褐色为主<sup>[1]</sup>。沥青组分的荧光色彩特性对于分析研究不同沥青组分的类别、含量以及发光颜色等数据至关重要。在实际的工作环境中, 荧光产生的机理是通过紫外光或蓝光等激发光源的照射, 岩石中石油的烃类会发出波长较长的可见光。故而对于不同的激发光源, 得到的荧光图像不同, 如

图 1 所示。同时由于不同地区地质结构的多样性, 不同地区、不同层位的油发出的荧光颜色也会有所差异。为了得到沥青的种类和含量, 目前的做法是用阈值分割方法先把荧光部分提取出来, 再根据 ISODATA 聚类算法<sup>[2]</sup>利用每个像素点的颜色信息进行成分分类, 再将分类结果保存为划分文件, 此过程速度较慢, 并且不同荧光的色彩差异性会导致分析结果的不准确。图 2 是用 ISODATA 聚类算法生成的同一划分文件对不同光源激发的荧光图像进行分类的结果。由图 2 可以看到由于激发光源不同, 同一荧光图像的分类结果是不同的, 这显然是不对的。为了得到正确的结果, 需要对不同色彩表现的

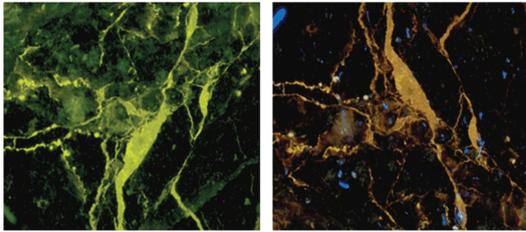
**作者简介:** 王爽(1997-), 女, 硕士研究生, 主要研究方向: 数字图像处理; 何小海(1964-), 男, 博士, 教授, 博士生导师, 主要研究方向: 图像处理、图像/视频编码通信、机器视觉与智能系统; 滕奇志(1961-), 女, 博士, 教授, 主要研究方向: 图像处理、模式识别。

**通讯作者:** 何小海 Email: hxh@scu.edu.cn

收稿日期: 2022-10-11

哈尔滨工业大学主办 ◆ 专题设计与应用

荧光图片分别聚类,生成不同的划分文件,随后在实际分析过程中,选择对应的划分文件进行分析,但此过程较为繁琐。



(a) 蓝光激发荧光图像 (b) 紫外光激发荧光图像

图1 不同激发光源下的荧光图像

Fig. 1 Fluorescence image under different excitation light sources

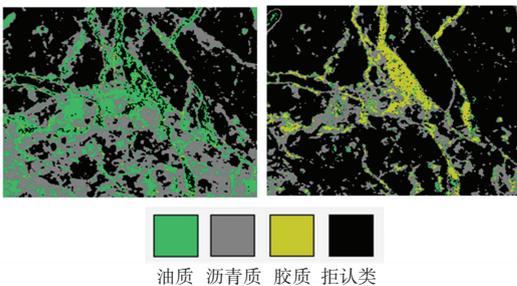


图2 同一划分标准对不同激发光源的荧光图像(图1)下的分割结果  
Fig. 2 Segmentation results of fluorescence images (Fig. 1) of different excitation light sources with the same division standard

此前,用于解决色彩差异问题的方法是对荧光图像进行色彩迁移。Reinhard等学者<sup>[3]</sup>提出了一种适用于大部分彩色图像的色彩迁移算法,该方法的缺点是迁移结果中损失了较多的颗粒以及孔隙细节。庞战<sup>[4]</sup>提出的基于FCM的局部迁移算法,相比于全局彩色迁移算法,虽然细节更为丰富,但存在匹配关系不够准确的问题,偶尔会产生错误的色彩迁移结果。王杰<sup>[5]</sup>提出的改进的松弛化最优传输荧光图像色彩迁移方法虽然在整体色彩准确性和稳定性等方面比前述的两者都好,但是也会出现局部色彩误迁移的问题。

近些年随着深度学习的迅速发展,图像语义分割结合了图像分类、目标检测和图像分割技术,通过一定的方法逐像素点进行分类,再进行区域划分,最终得到一幅具有像素级的分割图像。2012年,卷积神经网络(Convolutional neural network, CNN)、尤其是Krizhevsky等学者<sup>[6]</sup>提出的AlexNet模型,在计算机视觉领域取得了重大突破,成为图像分类、对象检测、目标跟踪和图像分割等计算机视觉领域任务中性能优异、且获广泛应用的深度神经网络学习模型<sup>[7]</sup>。卷积神经网络主要由4个部分组成,分别是:

卷积层、池化层、全连接层以及 $softmax$ 分类器<sup>[8]</sup>,通过对卷积层和池化层进行一系列的叠加,能够获得更深层次的网络结构,如VGGNet<sup>[9]</sup>、ResNet<sup>[10]</sup>等网络。这些深度学习方法能够自动学习图像中的有效特征从而进行分类,免除了传统荧光图像分析方法中人工交互以及需要对不同激发光源分开聚类产生划分标准的繁杂过程。因此,本文提出了基于深度学习的荧光图像分类技术。

## 1 本文方法

VGG模型是由Simonyan等学者<sup>[9]</sup>于2014年提出的一种通过增加卷积神经网络的深度来提高识别性能的网络,其中最佳网络包含16个权值层。VGG模型由于其基于微小滤波器尺寸的设计,已广泛用于分类和定位任务,基于VGG架构的网络则已在检测和分割<sup>[11-12]</sup>等方面进行探索与研究。一般来说,网络越深,特征提取能力越强。然而,下采样可能会导致低层特征丢失。UNet使用跳跃连接将低层特征与高层特征进行融合,能够保留高分辨细节信息。因此本文采用结合VGG16与UNet的方法来构建网络,在VGG16模型的基础上,将最后的3个全连接层替换为UNet解码器部分的结构。用VGG16网络作为编码器部分进行主干特征提取,产生低分辨率的图像;UNet作为解码部分进行加强特征提取,将低分辨率图像映射到像素级别的预测上去;同时增加注意力机制来防止像素级信息的缺失,提高特征提取的准确性。

### 1.1 编码器模块

VGG的核心思想是利用多个小卷积核代替大卷积核,既可以保证感受视野,又能够在减少卷积层参数的同时提升网络的表达能力。近年来,VGG网络衍生出了A~E七种不同的层次结构,本文使用的是其中的D结构,也就是VGG16,主要包含了13个卷积层、5个池化层和3个全连接层。一般来说,池化层、也即下采样层,分为最大池化和平均池化两种,主要位于连续的卷积层之间,用来压缩图像的特征尺寸。而在VGG16网络中使用的是最大池化方法,在几个卷积层之后续接一个最大池化形成了一个block。卷积层、池化层用来将原始图像数据映射至特征空间中,最后的3个全连接层以及 $softmax$ 分类器则用来将映射后的特征进行分类。VGG16网络结构如图3所示,由图3可知,该结构共包括5个block,每个block的通道数一致,最大池化层用来减少特征图的尺寸<sup>[13]</sup>。

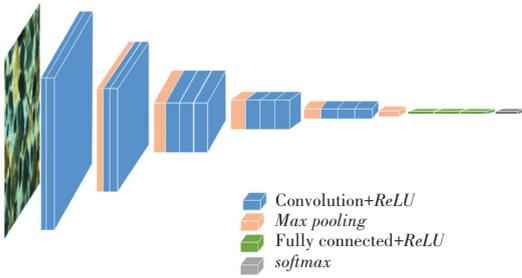


图3 VGG16 网络结构

Fig. 3 VGG16 network structure

1.2 解码器模块

UNet<sup>[14]</sup>是2015年首次提出用于生物医学图像分割的功能强大的卷积神经网络架构之一。这是一个“U”型对称网络结构,由编码器和解码器两个通用部分组成。编码器能够捕捉上下文信息,完成图片的下采样。编码部分遵循交替卷积和池化操作的卷积网络的典型架构,并逐步向下采样进行特征映射,同时增加每层特征映射的数量。每层通过对2个3×3的卷积核、ReLU函数,包括一个2×2的最大池化核进行操作,一共经过4次相应的下采样操作后,特征图的宽度得以减小,通道数翻倍,从而能更好地完成对像素点的特征提取。解码器部分能够实现精确定位,主要用于特征图的上采样。通过对特征图的上采样以及卷积操作处理后,又融合了编码部分对应的特征图,进一步增加了输出的分辨率,从而逐步恢复特征图像的像素信息及图像精度。UNet将特征图从编码器的每一层传递到解码器的类似层,就使得分类器可以依据不同规模和复杂性的特征来做出决策。这种架构已被证明对于数据量有限的分割问题非常有用,网络结构如图4所示。

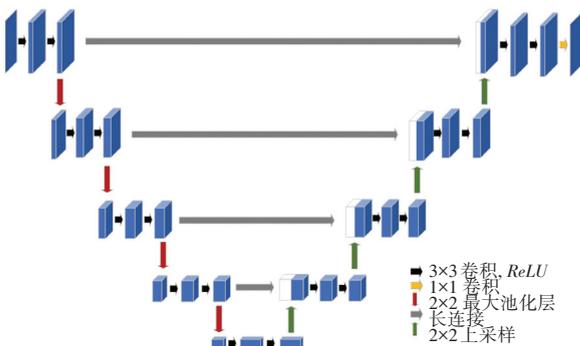


图4 UNet 网络结构

Fig. 4 UNet network structure

1.3 注意力机制

随着深度学习的迅速发展,注意机制在图像分类以及语义分割中也得到了大范围的应用。卷积运算的原理是将空间层次特性图的局部特性与通道层

次特性图的局部特性加以相互融合,进而得到多维多尺寸的空间结构信息。但是,各个通道所含信息内容的重要性也因图像分类任务的差异而有所不同,如果不加以区别地整合每个通道的信息内容,那么在某种程度上会丢失许多重要的信息。因此,Woo等学者<sup>[15]</sup>于2018年提出了一种轻量的注意力模块(Convolutional Block Attention Module, CBAM),可以在通道和空间维度上增加注意力,并且在ResNet和MobileNet等经典结构上添加了CBAM模块并进行对比分析,同时也进行了可视化,发现CBAM更关注目标物体识别,这也使得CBAM具有更好的解释性。

CBAM由通道注意力模块和空间注意力模块组成<sup>[16]</sup>。其中,通道注意力模块是一种考虑特征图通道之间关系的注意力机制,突出输入图像提供的代表性信息。空间注意力模型能够帮助神经网络系统,使其更注重图片中那些对图像分类起着决定性意义的关键像素区域而忽略那些无关紧要的区域。实验证明,输入特征先指向通道注意力模块、再指向空间注意力模块效果会更好。

给定一个中间特征映射  $F \in R^{C \times H \times W}$  作为输入,CBAM按顺序推断出一个1D的通道注意映射  $M_c \in R^{C \times 1 \times 1}$  和一个2D的空间注意映射  $M_s \in R^{1 \times H \times W}$ ,总的作用公式<sup>[17]</sup>可描述为:

$$F_1 = M_c(F) \otimes F \quad (1)$$

$$F_2 = M_s(F_1) \otimes F_1 \quad (2)$$

其中,“ $\otimes$ ”表示元素级乘法; $F$ 表示输入特征图; $F_1$ 表示经过通道注意力加权得到的特征图; $F_2$ 表示经过空间注意力加权得到的特征图; $M_c(F)$ 表示通道注意力输出权值; $M_s(F_1)$ 为空间注意力输出权值。整体结构如图5所示。

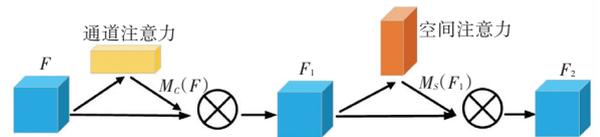


图5 CBAM 模块结构

Fig. 5 Structure of CBAM module

1.4 整体 VGG16-UNet 网络模型

本文所用的整体网络的结构框架如图6所示。图6中,左半部分是编码器部分,有5个子模块。设计上整合了原始VGG16的13个卷积层以及4个池化层,实现下采样。网络的右半部分是解码器部分,也有5个块。每个块后面是一次上采样操作,用于恢复图像的尺寸。此外,每层之间有一组跳跃连接

用于恢复图像的尺寸,这些跳跃连接是使用连接操作去组合相应的特征映射来实现的。由于这是用于语义分割的全卷积神经网络 FCN 的变体,需要保留图像的空间维度信息,因此需要使用跳跃连接。最后一个卷积层只有一个滤波器,这与大多数其他神经网络的最后一个密集层相似,并给出二进制掩码预测。对于网络的瓶颈中心部分,本文去掉了 VGG16 最后的 3 个全连接层,取而代之的是通过上采样直接把特征图放大成一个与编码器同层部分具有相同尺寸的图像。此外,本文还在编码器部分提取出来的所有有效特征层上增加了注意力机制。

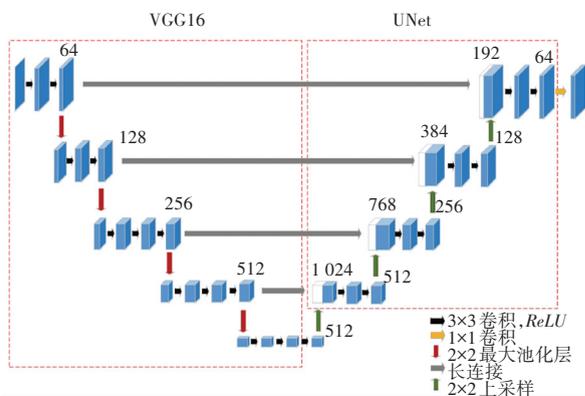


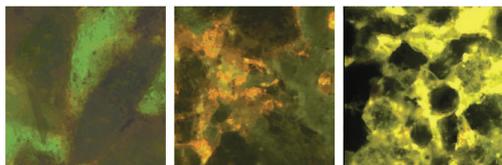
图6 VGG16-UNet 整体网络结构

Fig. 6 VGG16-UNet overall network structure

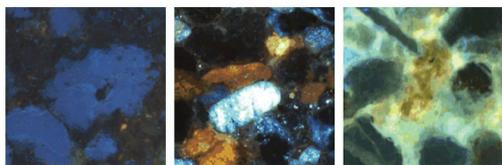
## 2 实验

### 2.1 数据集

本文使用的数据集均是通过高清相机在显微镜下对不同石油地质部门的荧光薄片进行拍摄而得的,随后对采集到的图像进行裁剪,裁剪后数据集中的原始图像大小均为  $512 \times 512$ 。本文实验所用数据集共有 1 199 张,包括蓝光激发的荧光图像和紫外光激发的荧光图像,如图 7 所示。训练集和测试集按照 9:1 的比例,即用作训练数据 1 079 张,测试数据 120 张。



(a) 蓝光激发的荧光图像



(b) 紫外光激发的荧光图像

图7 数据集图示

Fig. 7 Dataset images

实际研究中,考虑到在显微镜下采集到的荧光图像的尺寸通常是远大于  $512 \times 512$  的,所以在预测时会对原图进行预处理,首先将原图裁剪为  $512 \times 512$  的图片,然后用训练后的模型分别对每一个裁剪后的图进行预测,接着将预测后的图加以拼接,最后可以得到一整张荧光图像分类结果图。

### 2.2 实验设置

在本次实验中,硬件环境为 Ubuntu 20.04 LTS 操作系统,CPU 的型号为 i7-9700,GPU 型号为 GeForce RTX 2080 Ti 以及 32 GB 内存。开发环境采用的是 Pytorch 深度学习框架。在训练过程中,初始学习率为  $1 \times 10^{-4}$ ,BatchSize 为 4,共训练 100 epoches,每个 epoch 总共训练 1 199 张图。同时,使用 Adam 优化器,因为 Adam 算法能够自适应调节学习效率。

卷积层使用 ReLU 激活函数,该函数是从底部开始半修正的一种函数,数学公式具体如下:

$$f(x) = \max(0, x) \quad (3)$$

其中,  $x$  表示输入。

对岩石荧光薄片图像的石油成分进行提取,是一个多分类问题。对于分类问题,最常用的损失函数是交叉熵损失函数(Cross Entropy Loss)。

在二分类的情况下,网络预测的结果最终只有 2 类,假设对于每个类别预测的概率分别为  $p$  和  $1 - p$ ,那么此时交叉熵损失函数的计算公式为:

$$J_{CE} = - [y \cdot \log(p) + (1 - y) \cdot \log(1 - p)] \quad (4)$$

其中,  $y$  是样本标签,正样本标签为 1,负样本标签为 0;  $p$  表示预测为正样品的概率。

本文是一个多分类问题,多分类的交叉熵损失可表示为:

$$J_{CE} = - \sum_{i=1}^K y_i \log(p_i) \quad (5)$$

其中,  $K$  是种类数量;  $y_i$  是一个 One-hot 向量,除了目标类为 1 之外其他类别上的输出都为 0;  $p_i$  是神经网络的输出,也就是类别  $i$  的概率。

### 2.3 评价指标

为了评估系统的有效性,本实验采用了语义分割中最常用的评价指标,分别是精确率(Precision)、像素准确率平均值(mean Pixel Accuracy,  $mPA$ )和平均交并比(Mean intersection over union,  $mIoU$ )。这些指标可以衡量每个像素值的分类对整个分类任务的性能。对于多分类问题,假设总共有  $k + 1$  个类别(包括背景), $P_{ij}$  是类  $i$  被推

断为属于类  $j$  的像素数量。也就是说,  $P_{ii}$  表示分类正确的真正数(真正例  $TP$  + 真负例  $TN$ ), 而  $P_{ij}$  和  $P_{ji}$  通常分别被解释为假正例  $FP$  和假负例  $FN$ 。对于本次实验中采用的评价指标, 可做阐释分述如下。

(1) 精确率: 在全部预测为正的结果中, 被预测正确的正样本所占的比例。推得的数学定义公式为:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

(2)  $mPA$ : 对每个类计算正确分类的像素数量与所有像素数量的比值, 再对类总数取平均。推得的数学定义公式为:

$$mPA = \frac{1}{k + 1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij}} \quad (7)$$

(3)  $mIoU$ : 这是用于图像分割问题的标准评价指标。具体就是 2 个集合(原始图和预测图)的交集和并集之间的比例。在本实验中, 则表示真实分割图和系统预测的分割图之间的交并比。该比例可以看作交集(真正数)与总数(包括真正数、假正例、假负例)的比值。最后也是按类别取平均。由此推得的数学定义公式为:

$$mIoU = \frac{1}{k + 1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (8)$$

### 2.4 组件消融实验

为证明本文提出的用 UNet 作为编码部分以及注意力机制模块对网络的有效性, 分别进行了实验比较上述模块对结果的影响。实验结果见表 1。从 3 组实验可以看出, VGG16+UNet 模型相较于原始 VGG16 网络, 各个分类评价指标都有了大幅度的提升; 同时添加了注意力机制的模型能取得最佳效果, 也说明了注意力机制对整个模型框架的有效性。

表 1 不同模块对实验结果的影响

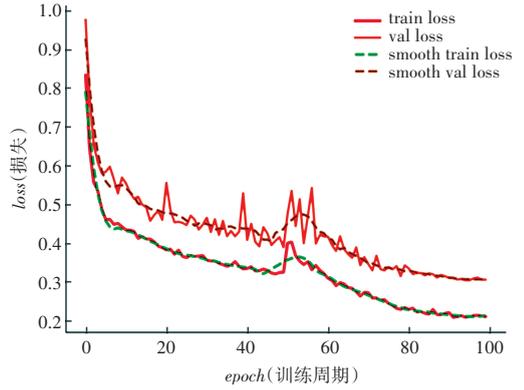
Tab. 1 Effect of different modules on experimental results

网络模型	Precision	mPA	mIoU
VGG16	71.51	69.37	58.24
VGG16+UNet	86.35	86.20	75.11
本文模型	88.39	86.44	77.67

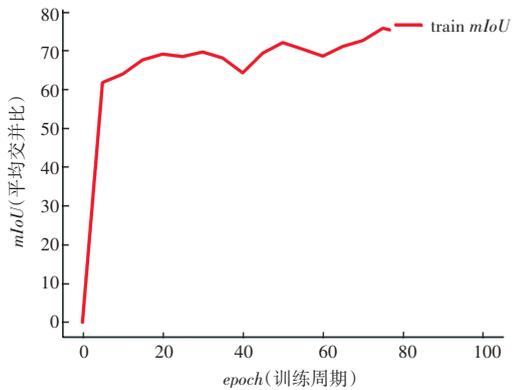
### 2.5 结果分析

图 8 展示了训练过程中的一组训练参数曲线图。图 8(a) 是损失函数曲线图, 图 8(b) 是  $mIoU$  曲线图。横坐标表示训练的次数, 纵坐标分别表示训练损失值以及平均交并比。由图 8 可以看出, 损失值随着训练次数的增长而逐步收敛到一个下界, 而

平均交并比则随着训练次数的增长而逐步增大。损失函数中间凸起的原因是训练过程分为 2 个阶段, 分别是冻结阶段和解冻阶段。设置冻结阶段是为了满足机器性能不足时的训练需求, 此过程需要的显存空间较小, 可以根据显卡情况动态设置参数, 解冻阶段占用的显存空间较大, 网络的参数也会随之发生改变。



(a) 损失函数曲线



(b)  $mIoU$  曲线

图 8 训练参数曲线

Fig. 8 Training parameters curve

为了客观验证本文所用网络的分类性能, 使用了相同的数据集分别在 ResNet、PspNet 等主流分割网络上进行了对比实验, 表 2 为不同网络模型在 3 个评价指标上的结果。由表 2 可以看到, 本文所使用的方法在各个指标上都有一定的优势。

表 2 不同网络模型下荧光成分分类结果评价指标

Tab. 2 Evaluation indicators for classification results of fluorescent components under different network models

网络模型	Precision	mPA	mIoU
UNet	77.37	74.83	64.59
PspNet	74.06	71.68	56.81
ResNet50	75.50	71.60	57.87
FCN_ResNet101	83.90	80.32	71.40
本文方法	88.39	86.44	77.67

同时,本文还选取了不同尺寸大小的荧光图片,在分类时间上与传统 ISODATA 聚类方法做了对比,结果见表3。由表3可以看到,在不同尺寸的图像下,本文的运行时间都优于传统算法,尤其是在大图下,其时间性能上的优势则体现得更加明显。

表3 荧光成分分类运行时间对比

Tab. 3 Comparison of fluorescent component classification operation time

图像	图像大小	ISODATA 聚类算法	本文算法
图像1	701×507	266	194
图像2	768×576	282	196
图像3	1 728×1 296	1 047	244
图像4	2 736×1 824	4 563	376

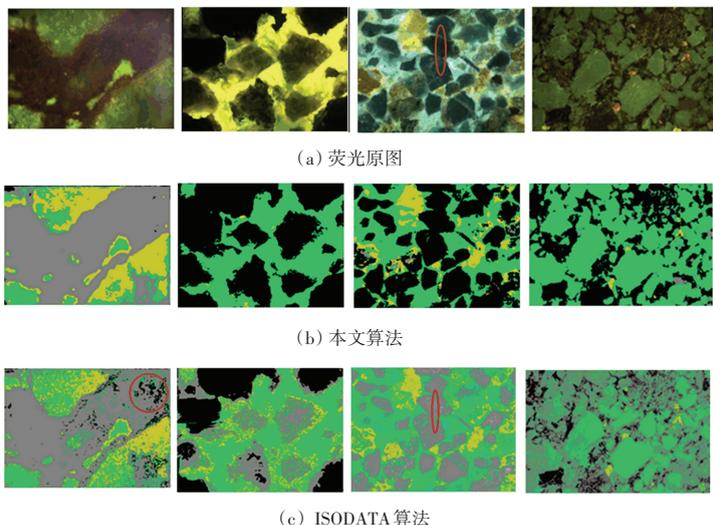


图9 不同分类方法下的分类结果对比

Fig. 9 Comparison of classification results under different classification methods

### 3 结束语

本文针对岩石荧光薄片图像在荧光成分提取上需要大量人工交互,并且对于不同色彩差异的荧光图像需要分开聚类并产生不同的划分文件的问题,提出了一种基于深度学习的荧光图像成分分类技术。该方法在 VGG16-UNet 的网络技术基础上融入了注意力机制,UNet 框架实现了端到端的自动分割;VGG16 用较小的卷积核既有利于提高感受视野,又有利于在减小卷积层参数的同时提高网络的表现力;同时融入注意力机制模块,增加特征信息内容的学习权重,抑制无关背景信息内容的影响。通过一系列实验数据比对分析,分类结果不论是在主观效果上,还是在客观指标上都有一定的优势。对多个不同油田采集的荧光薄片图像进行实验,均取得了较好的分类结果,实验结果也说明,该模型具有

最后,为了直观地感受本文的分类效果,图9给出了表3中的4张荧光图像的分类结果,包括了不同激发光源下的岩石薄片荧光图像。图9(a)是原始荧光图像,图9(b)是用本文的深度学习方法分类的结果图,图9(c)是用 ISODATA 聚类算法分类的结果。从结果可以看到,本文分类的结果在整体上没有出现误分类的情况,并且相较于 ISODATA 聚类算法不会出现太多孤立的噪点,见图9(c)中第一列标出来的区域。在对整张荧光图像进行分类时,ISODATA 聚类算法有时会把不发光的部分划分为沥青质,而本文算法的分类结果则不会出现此问题,见图9中第三列图标记的部分。

良好的性能,并且减少了传统荧光成分分析过程中的人工交互过程。后期将考虑在保证有效性的前提下,优化网络模型,进一步减少参数量,将算法应用于实际工程中。

### 参考文献

- [1] 林硕,张燕,贾彦勇,等. 荧光图像分析法在河南油田的应用[J]. 西部探矿工程,2016,28(10):39-41,44.
- [2] 何霖,滕奇志,罗代升,等. 一种改进的 ISODATA 算法及在彩色荧光图像中的应用[J]. 四川大学学报(自然科学版),2007,44(03):563-568.
- [3] REINHARD E, ADHIKMIN M, Gooch B, et al. Color transfer between images[J]. IEEE Computer Graphics and Applications, 2001, 21(5): 34-41.
- [4] 庞战. 岩石薄片高分辨率荧光图像分析系统的设计与实现[D]. 成都:四川大学,2017.
- [5] 王杰. 岩石荧光薄片图像分析相关技术研究[D]. 成都:四川大学,2021.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet

- classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems. Denver, USA: NIPS Foundation, 2012, 25:1106-1114.
- [7] 李梦怡,朱定局. 基于全卷积网络的图像语义分割方法综述[J]. 计算机系统应用,2021,30(09):41-52.
- [8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA:IEEE,2015: 3431-3440.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
- [10] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016:770-778.
- [11] IGLOVNIKOV V, SHVETS A. TeraNet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation[J]. arXiv preprint arXiv:1801.05746, 2018.
- [12] PRAVITASARI A A, IRIAWAN N, ALMUHAYAR M, et al. UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation [J]. TELKOMNIKA (Telecommunication Computing Electronics and Control), 2020, 18(3): 1310-1318.
- [13] 吴碧巧,邢永鑫,王天一. 基于VGG16和迁移学习的高分辨率掌纹图像识别[J]. 智能计算机与应用,2021,11(05):37-42.
- [14] 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. Cham: Springer,2015: 234-241.
- [15] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany:dblp,2018: 3-19.
- [16] KHAN R U, WONG W S, ULLAH I, et al. Evaluating the efficiency of CBAM-Resnet using Malaysian sign language[J]. CMC-Computers Materials & Continua, 2022, 71(2): 2755-2772.
- [17] 李红艳,李春庚,安居白,等. 注意力机制改进卷积神经网络的遥感图像目标检测[J]. 中国图象图形学报, 2019, 24(08): 1400-1408.

(上接第174页)

- [5] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015:1440-1448.
- [6] 邢宇驰,李大军,叶茂茂. 基于YOLOv5的遥感图像目标检测[J]. 江西科学,2021,39(04):725-732.
- [7] CHEN Chenyi, LIU Mingyu, TUZEL O, et al. R-CNN for small object detection [C]//Computer Vision - ACCV2016. Cham: Springer International Publishing,2017:214-230.
- [8] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015,37(9):1904-1916.
- [9] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [10] REDMON J, FARHADI A. YOLO900. Better, faster, stronger [J]. arXiv preprint arXiv:1804.02767,2018.
- [11] REDMON J, FARHADI A. YOLOv3: An incremental improvement[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City:IEEE,2018: 89-95.
- [12] 南晓虎,丁雷. 深度学习的典型目标检测算法综述[J]. 计算机应用研究,2020,37(S2):15-21.
- [13] 杜传明. 百度地图API在小型地理信息系统中的应用[J]. 测绘与空间地理信息,2011,34(02):152-153,156.
- [14] 卢卫中,钟安元. 基于HTML与CSS网页设计[M]. 重庆:重庆大学电子音像出版社,2020.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE, 2016:779-788.
- [16] 李一鸣,王潇. 基于YOLOv5s模型的轧钢表面缺陷检测[J]. 制造业自动化,2021,43(11):117-119.
- [17] ZHENG Zhaohui, WANG Ping, REN Dongwei, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics,2022,52(5):8574 - 8586.
- [18] JOCHER G. YOLOv5 [EB/OL]. [2020]. [https://github.com/ultralytics/yolov5\(2020\)](https://github.com/ultralytics/yolov5(2020)).
- [19] 余平,张淑芳. SQL Server数据库基础[M]. 北京:北京邮电大学出版社,2018.
- [20] CHEN Xinlei, FANG Hao, LIN T Y, et al. Microsoft COCO captions: Data collection and evaluation server[EB/OL]. [2020-05-13]. <https://arxiv.org/abs/1504.00325>.