

文章编号: 2095-2163(2023)02-0092-07

中图分类号: TP391

文献标志码: A

# 基于深度迁移学习的煤岩显微组分识别算法

季菁菁<sup>1</sup>, 奚峥皓<sup>1</sup>, 李忠峰<sup>2</sup>

(1 上海工程技术大学 电子电气工程学院, 上海 201620; 2 营口理工学院, 电气工程学院, 辽宁 营口 115000)

**摘要:** 针对 VGG-16 网络模型对小数量的煤岩显微组分图像数据集识别准确度低、模型参数量过多的问题, 提出一种基于迁移学习的小样本煤岩显微组分图像识别算法。算法首先改进了 VGG-16 分类网络模型, 将 Res2Net 模块与 VGG-16 相结合, 并用深度可分离卷积层代替 VGG-16 和 Res2Net 模块的原始卷积层, 以此作为深度迁移识别网络预训练模型; 然后迁移预训练模型中部分网络结构和参数, 并结合优化的分类器模块完成网络的学习与优化。实验结果表明, 在样本数据不充足的条件下, 基于深度迁移学习识别网络模型识别准确率为 96.33%, 模型参数量为 11.79 M。与其它网络模型相比, 该方法在小数量的煤岩显微组分图像识别中具有明显的优越性。

**关键词:** 煤岩显微组分; 小样本; 深度可分离卷积; Res2Net; 迁移学习

## Coal rock microfraction identification method based on deep transfer learning

JI Jingjing<sup>1</sup>, XI Zhenghao<sup>1</sup>, LI Zhongfeng<sup>2</sup>

(1 School of Electronic and Electrical Engineering, Shanghai University of Engineering and Technology, Shanghai 201620, China;

2 School of Electrical Engineering, Yingkou Institute of Technology, Yingkou Liaoning 115000, China)

**[Abstract]** To solve the problems of low accuracy of the VGG-16 network model in recognizing small number of coal rock microfraction images and excessive number of model parameters, an algorithm for recognizing small sample coal rock microfraction images based on transfer learning is proposed. Firstly, the VGG-16 network model is improved. The Res2Net module is combined with VGG-16 and the original convolutional layers of the VGG-16 and Res2Net modules are replaced with depth-separable convolutional layers as a pre-training model for the deep transfer recognition network. Then some of the network structures and parameters in the pre-training model are transferred and combined with the optimized classification module to complete the learning and optimization of the network. The experimental results show that, under the condition of insufficient sample data, the recognition accuracy of the deep migration learning based recognition network model is 96.33% and the number of model parameters is 11.79 M. Compared with other network models, the proposed method has obvious advantages in the recognition of small number of microscopic components of coal rocks.

**[Key words]** coal rock microfraction; small samples; depthwise separable convolution; Res2Net; transfer learning

## 0 引言

煤岩质量决定着煤岩的燃烧效率, 高质量煤岩燃烧可以有效减少碳排放, 有利于节能减排发展战略的进一步实施。对煤岩显微组分分析, 是对煤岩质量进行评估的有效途径之一<sup>[1-2]</sup>。但煤岩显微图像组分的复杂多样性, 导致在对煤岩组分进行识别时需要人工的参与, 使判断结果具有主观性。在工业生产应用中, 通常用到的煤岩显微组分分析系统主要由分光光度计和偏光显微镜等组成, 其识别分

类准确度表现优异。但是, 该系统的操作与分析过程均需要专业人员参与其中, 普适性较差。

随着计算机视觉技术的发展, 针对上述问题, 已有学者采用将计算机视觉与图像处理结合的煤岩显微图像的分析识别方法。如: 文献[3]提出了基于多重分形谱法的煤矿质分类方法; 文献[4]提出了基于轮廓波变换的煤岩壳质组显微组分的分类方法; 文献[5]基于 PCA 与 RBF-SVM 对煤岩显微镜质组组分进行分类; 文献[6]提出一种基于改进极限学习机的焦煤惰质组的分类方法; 文献[7]基于

**基金项目:** 国家自然科学基金(61801286)。

**作者简介:** 季菁菁(1996-), 女, 硕士研究生, 主要研究方向: 深度学习图像处理; 奚峥皓(1981-), 男, 博士, 副教授, 硕士生导师, 主要研究方向: 光电信息处理研究。

**通讯作者:** 奚峥皓 Email: zhenghaoxi@sues.edu.cn

**收稿日期:** 2022-04-17

空间域处理方法,根据特征亮度值对组分进行分析等等。虽然上述方法对识别分类煤岩组分有一定的效果,但对煤岩组分特征提取工作中仍需要人工参与,无法实现真正的自动化识别分类。

随着深度学习的发展,越来越多的研究尝试将深度学习用于图像处理任务中。相较于传统算法,深度学习可以避免在对图像特征进行提取过程中的人工参与。如:文献[8]中提出的 AlexNet 网络在 ImageNet 上的亮眼表现,证明了在机器视觉领域里深度学习的不可或缺性;文献[9]提出一种基于全卷积神经网络的花岗岩组分分布特征的分析模型;文献[10]提出一种用于大规模图像识别的 VGG-16 卷积网络等等。虽然这些方法在进行深度学习图像处理时都表现出了良好的性能,但其训练时需要大量的数据集。由于煤岩显微组分图像数据集制取方式复杂,获取成本高,现有的数据集数量无法满足上述网络的训练要求,使得网络很难达到预期识别准确度的要求。而文献[11]提出了一种基于迁移学习的煤岩壳质显微组分识别模型,其适用于小样本数据集的训练。迁移学习可以将大规模数据集上训练好的网络结构与参数迁移至目标数据集网络中,有效解决对小样本数据集进行识别问题。

本文在文献[11]的基础上,对预训练 VGG-16 分类网络模型进行了改进,并在 ImageNet 上完成预训练,将训练好的部分改进网络模型与参数,结合优化的分类器共同组成 TR-VGG-16 深度迁移网络模型,实现对小样本煤岩显微组分识别。

## 1 基于 TR-VGG-16 深度迁移网络模型的构建

### 1.1 Res2Net 模块

为了提高 VGG-16 分类网络的特征提取能力,引入 Res2Net 模块,实现对跳跃连接层传输的下采样特征图信息的进一步特征提取。Res2Net 模型如图 1 所示。

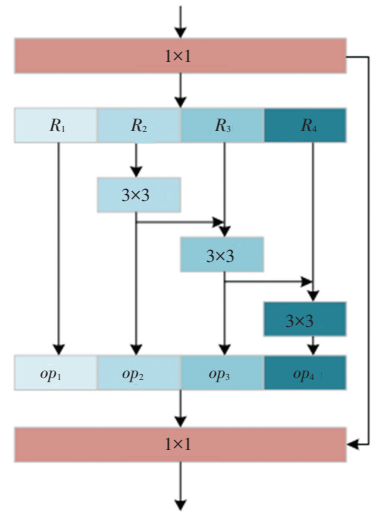


图 1 Res2Net 模块

Fig. 1 Res2Net module

经过第一层卷积输出特征后,Res2Net 网络根据通道数将特征平均分为  $S$  组,每组特征为  $R_\varphi$ ,随后对除  $R_1$  外的每组特征进行卷积操作,其计算公式如式(1)所示:

$$Op_\varphi = \begin{cases} R_\varphi, \varphi = 1 \\ K_\varphi(R_\varphi + Op_{\varphi-1}), 1 < \varphi \leq s \end{cases} \quad (1)$$

其中,  $K_\varphi()$  为对特征进行卷积操作。由图 1 可知,从第 2 组特征计算开始,每次  $K_\varphi()$  计算的输入都是由前一组  $Op_\varphi$  残差连接本组特征  $R_\varphi$  组成。最后将得到的所有多尺度特征  $Op_\varphi$  进行拼接,输入到下一层卷积层,最后得到 Res2Net 模块的输出。

### 1.2 深度可分离卷积

深度可分离卷积 (Depthwise Separable Convolution) 是 MobileNet 模型的核心组成部分,是一种因式分解卷积<sup>[12]</sup>。与传统卷积运算作用于图像所有通道不同,深度可分离卷积在进行卷积运算时,将普通卷积运算分解为 Depthwise 运算和 Pointwise 运算,运用不同的卷积核作用于图像不同的通道。深度可分离卷积步骤如图 2 所示。

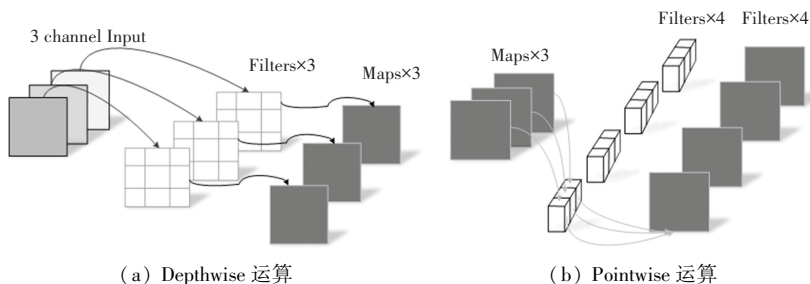


图 2 深度可分离卷积

Fig. 2 Depthwise separable convolution

首先,进行 Depthwise 运算,对输入特征的每个通道内的数据进行深度卷积运算;之后进行 Pointwise 运算。利用大小为  $1 \times 1 \times M$  的卷积核进行逐点卷积运算, $M$  为上一层特征通道数,对不同通道的输出加以组合。输出结果如式(2)、式(3)所示:

$$H' = \frac{(H - h + 2P_h)}{S_h} + 1 \quad (2)$$

$$W' = \frac{(W - w + 2P_w)}{S_w} + 1 \quad (3)$$

式中: $H'$  和  $W'$  为输出图像的高和宽, $H$  和  $W$  为输入图像的高和宽, $P_h$  和  $P_w$  为输入的垂直与水平的填充, $S_h$  和  $S_w$  为垂直与水平的步幅。

### 1.3 改进的 VGG-16 分类网络模型构建

传统 VGG-16<sup>[13]</sup> 分类网络模型由 13 个卷积层,5 个最大池化层,3 个全连接层和 1 个 Softmax 层组成。为了提高网络的特征提取能力,在网络中添加了 Res2Net 块。随着 Res2Net 块的加入,网络模型层数加深、参数增多。为减少参数量,提高运算效率,将 VGG-16 的所有卷积层调整为积核大小为  $3 \times 3$  的深度可分离卷积层。改进的 VGG-16 网络模型如图 3 所示。

实现特征提取。

$D^i$  表示进行深度可分离卷积运算, $i$  为卷积层级。将数据从以  $D^1$  输入,即从第 1 层开始卷积操作,依次在经过  $i$  层时进行如式(4)的卷积运算。

$$D^i = \delta(D^{i-1}), i = \{1, 2, \dots, 10\} \quad (4)$$

而后经过 Res2Net\_block 进行深度可分离卷积运算。输出如式(5)所示:

$$O_p = \sum_{\varphi=1}^s O_{p_\varphi} \quad (5)$$

改进的 VGG-16 分类网络结构通过卷积层和池化层的连接实现对图像特征的提取,保留分类器前的部分,将其作为 TR-VGG-16 网络特征提取网络结构,以实现图像特征提取和输出。

### 1.4 TR-VGG-16 识别网络模型的构建

已知 ImageNet 数据集具有规模大、质量优和多样性高的特点,在此用其对预训练模型进行训练。本文对煤岩显微图像的 3 种组分进行分类识别,在分类器中设置 2 个全连接层,且第二层全连接层设为 3 个神经元。经过全连接层的计算过程如式(6)所示:

$$P = (I + 1) \times O \quad (6)$$

式中: $I$  为输入的神经元,第一层  $I$  为改进的 VGG-16 特征提取网络输出的特征映射向量, $O$  为偏差量。

由于网络的加深,会导致训练过程中,输入数据的分布发生偏移,使其模型收敛速度变慢。因此,在网络中加入批量归一化层(Batch Normalization, BN),降低模型对参数初始化的依赖程度,提高初始学习率,进而提高网络训练的速度。小批量样本  $x$  的集合  $B = \{x_{1 \dots l}\}, l \in \{1, \dots, m\}, m$  为小批量的尺寸。BN 层的输出结果为  $y_l$ , 其计算公式如式(7)所示:

$$y_l = \gamma \cdot \hat{x}_l + \eta \quad (7)$$

其中, $\gamma, \eta$  分别为网络学习得到的放缩系数和平移系数, $\hat{x}_l$  为批量样本中每个样本归一化的结果,其计算公式如式(8)所示:

$$\hat{x}_l = \frac{x_l - v_l}{\sqrt{\sigma_l^2 + \tau}} \quad (8)$$

其中, $x_l$  表示第  $l$  个小批量的训练数据; $\tau$  为避免分母为 0 的因子( $\tau \rightarrow 0$ ); $v_l$  为批量样本的均值; $\sigma_l$  为样本的方差。 $v_l, \sigma_l$  分别依据式(9)和式(10)计算得到:

$$v_l = \frac{1}{m} \sum x_l \quad (9)$$

$$\sigma_l = \sqrt{\frac{1}{m} \sum (x_l - v_l)^2} \quad (10)$$

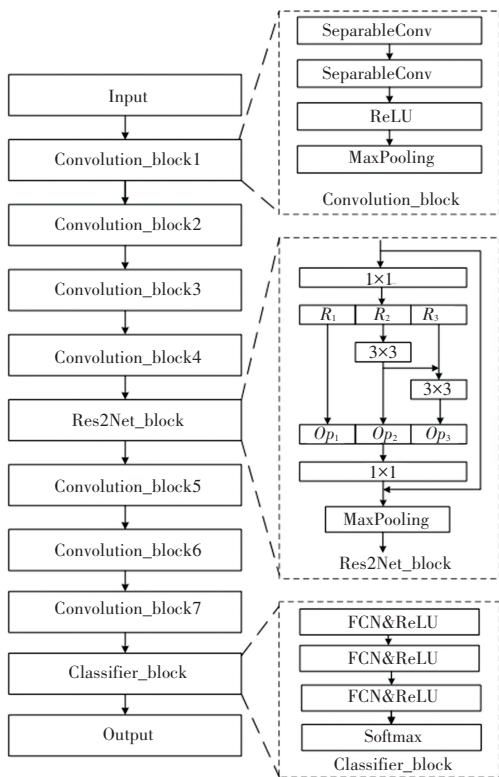


图 3 改进的 VGG-16 分类网络结构

Fig. 3 Improved VGG-16 classification network structure

设:煤岩显微组分图像集合为  $X$ , 对应标签集合为  $Y$ , 由  $X$  和  $Y$  构成一个集合  $A = \{X, Y\}$ 。将  $A$  输入,

由于煤岩显微组分数据集数量有限, 而过少的数据集在训练中会导致过拟合现象出现, 即网络的损失函数较小, 预测验证集的准确率偏高。为防止这一现象, 在网络中引入 Dropout 层, 其模型如图 4 所示。

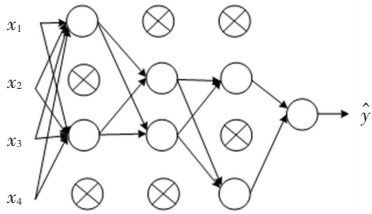


图 4 Dropout 示意图

Fig. 4 Dropout diagram

其中,  $x_1, x_2, x_3, x_4$  为输入,  $\hat{y}$  为输出。

Dropout 层的主要功能, 是在每个批次的模型训练中, 随机让网络某些隐含层节点的权重不工作, 减轻训练模型的复杂程度, 防止训练过拟合。

由于本文任务是对煤岩显微组分图像数据集进行多分类, 所以选用 Softmax 函数, 求得所属类别的概率分布值  $\text{Softmax}(\alpha_\beta)$ , 其计算如式(11) 所示:

$$\text{Softmax}(\alpha_\beta) = \frac{e^{\alpha_\beta}}{\sum_{j=1}^n e^{\alpha_n}} \quad (11)$$

其中,  $\alpha_\beta$  为第  $\beta$  个类别的输出值,  $n$  为数据集总类别数。

综上所述, 基于 TR-VGG-16 煤岩组分识别网络结构如图 5 所示。

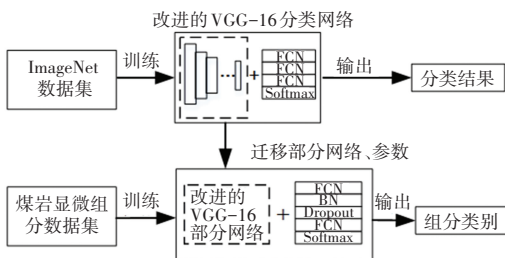


图 5 基于深度迁移学习的煤岩组分识别网络示意图

Fig. 5 Deep transfer learning coal rock component identification network

为了计算网络每次迭代的前向结果与真实值的差距, 进而达到指导下一步的训练向正确方向进行的目的, 引入了损失函数。本文选用交叉熵作为该网络的损失函数, 定义如下:

$$\text{Loss} = - \sum_s \sum_v P_v(s) \log(Q_v(s)) \quad (12)$$

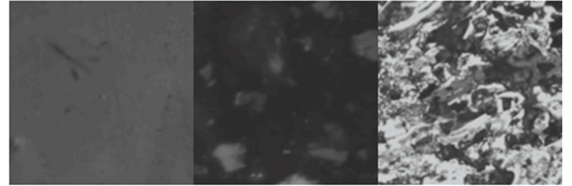
式中:  $P_v(s)$  为第  $s$  个训练样本在  $v$  类下的真实值,  $Q_v(s)$  为第  $s$  个训练样本在  $v$  类下类别输出值。

## 2 实验结果分析

### 2.1 实验数据

#### 2.1.1 煤岩显微组分数据集

实验选用产自甘肃一带的新生代煤, 用来制作煤岩显微组分数据集的煤样。其主要组分分为: 镜质组、丝质组和壳质组。如图 6 所示。



(a) 镜质组 (b) 丝质组 (c) 壳质组

图 6 煤岩显微组分展示

Fig. 6 Microfraction of coal rocks

采用国际规定的煤岩显微组分制备方法 (GB/T 16773-2013) 制作粉煤光片<sup>[14]</sup>, 用图像采集装置, 采集偏反光显微镜下的煤岩组分图像。由于实验条件所限, 本次实验数据集中仅包含 3 种煤岩组分, 尺寸为  $2\ 264 \times 2\ 264$  的 RGB 显微图像各 15 张。为了减少网络计算量, 对得到的数据进行预处理, 即将原 RGB 图像转换成灰度图像。

#### 2.1.2 数据增强

较少的数据集进行网络模型训练时会导致模型的过拟合, 但由于采集到的数据集具有数量少但图像尺寸较大的特点, 所以本文对煤岩组分数据集进行数据增强处理。

(1) 图像剪裁: 将原有图像剪裁成  $256 \times 256$  大小的图像, 不足尺寸的则舍弃。将剪裁后的图像用于后续的数据增强。

(2) 均衡化增强: 使用直方图均衡化方法处理煤岩灰度图像, 根据灰度图像的概率密度确定映射函数。概率密度公式如式(13) 所示:

$$\lambda_s = \sum_{h=0}^s \frac{m_h}{M}, s = 0, 1, \dots, l - 1 \quad (13)$$

其中,  $\lambda_s$  为灰度的累积概率密度;  $m_h$  为灰度级  $h$  出现的个数;  $M$  为总像素数;  $l$  为总数灰度级。

这一处理方式使图像的直方图均匀分布, 提高了图像的对比度, 使各组分特征的呈现更加明显, 从而提高网络模型的识别能力。

(3) 噪声添加: 在图像上随机添加一些离散孤立的像素点, 模拟在真实获取图像过程中可能会遇到的随机干扰情况。本文选择添加高斯噪声、泊松噪声和椒盐噪声的处理方法。



概率密度函数如式(14)~式(16)所示:

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right] \quad (14)$$

$$p(z) = \frac{\lambda^2}{z!} e^{-\lambda} \quad (15)$$

$$p(z) = \begin{cases} p_{az} = a \\ p_{bz} = b \\ 0 \text{ 其他} \end{cases} \quad (16)$$

(4) 图像缩放: 将图像沿  $X$  轴和  $Y$  轴方向进行增缩。设缩放系数为  $Z_x$  和  $Z_y$ , 缩放前图像像素点坐标为  $c(x, y)$ , 缩放后像素点坐标为  $c'(x', y')$ 。图像缩放的变换矩阵如下列矩阵所示:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (17)$$

综上所述, 对均衡化增强均值、噪声方差、泊松系数、椒盐噪声比和缩放因子进行设定。其中, 缩放因子为 0.5~2 之间随机缩放, 均衡化增强为 50, 高斯噪声方差为 0.01, 泊松系数为 1, 椒盐噪声比为 0.9。

筛选后将数据集扩充至 1 500 张, 将其中 1 200 张图像作为实验的训练集, 300 张作为验证集。

## 2.2 实验环境

本文网络在 GPU 环境下运行, 具体配置见表 1。

表 1 平台硬件与软件配置

Tab. 1 Hardware and software configuration

项目	参数	项目	参数
显存	32 G	深度学习框架	TensorFlow
CPU	Intel core i7-7500	Cuda	10.0
内存	2T	语言	Python3.6
GPU	NVIDIA Tian(12 GBs)	操作系统	Windows10

对网络进行训练时的训练参数设置: 初始学习率为 0.001, Dropout 概率为 0.75, 迭代次数为 60 次, 选取 Adam 自适应学习率算法进行参数优化。

## 2.3 实验结果

### 2.3.1 识别准确率

训练完成后, 用测试集对模型进行识别性能的评估。本文选用识别准确率 (Accuracy,  $A$ ) 对网络模型识别能力进行评价。定义公式如下:

$$A = \frac{G_r}{G} \quad (18)$$

其中,  $G_r$  为正确识别的个数,  $G$  为测试集总样本的个数。

为了证明 TR-VGG-16 模型的有效性, 将其分别与不同网络模型的准确率进行对比, 对比结果见表 2。

表 2 不同模型对测试样本识别准确率对比

Tab. 2 Comparison of the recognition accuracy of different models for test samples

网络模型	识别准确率 $A$ / %
文献[10]	81.33
文献[11]	89.33
Ours	96.33

由表 2 知, 文献[10]所提网络模型的识别准确率仅有 81.33%, 文献[11]的识别准确率虽高于文献[10], 但也仅为 89.33% 以上, 而本文提出 TR-VGG-16 网络的识别准确率显著高于前两种网络, 其识别准确率可以达到 96.33%。

### 2.3.2 混淆矩阵分析

为了细化体现各组分的识别准确度, 选用混淆矩阵对模型识别结果进行对比, 如图 7 所示。

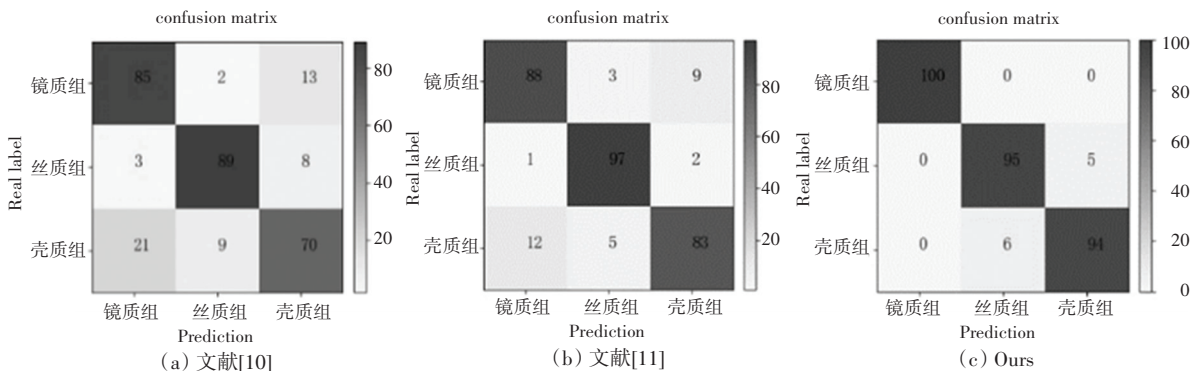


图 7 3 种网络模型的混淆矩阵

Fig. 7 Confusion matrix for the three network models

如图 7(a)所示,行为实际类别,列为识别结果。从文献[10]方法的混淆矩阵可以看出:其中 13 个镜质组被识别成壳质组,21 个镜质组被识别成壳质组,同时也有少量的其它组分间出现误判情况。由图 7(b)可知,文献[11]也存在着多个镜质组与壳质组样本间相互误判的情况。由图 7(c)可知,TR-VGG-16 对镜质组的识别全部正确,仅存在少量丝质组和壳质组组分间的误判。因此表明,本文提出的 TR-VGG-16 网络模型具有优异的煤岩显微组分识别能力。

### 2.3.3 收敛速度

将对文献[10]、文献[11]的网络模型与本文方法的收敛性进行可视化对比分析,分析结果如图 8 所示。由图中可见,TR-VGG-16 网络模型在第 10 个周期时几近收敛,且其识别准确率达到 90% 以上,并逐渐稳定;而文献[10]、文献[11]的网络模型较本文网络模型相比,同周期准确率较低且收敛性较差。由此可见,在同等条件下,本文提出的网络模型的收敛性和稳定性明显高于其它两种网络模型。

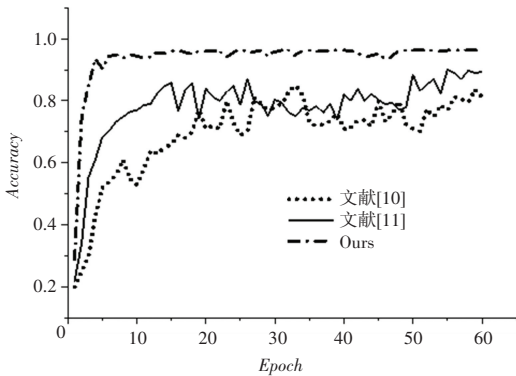


图 8 3 种模型收敛性对比

Fig. 8 Comparison of the convergence of the three models

### 2.3.4 消融实验

由于 TR-VGG-16 网络是在 VGG-16 网络基础上,结合迁移学习提出的改进,其中不仅将 VGG-16 中传统卷积层替换成深度可分离卷积,还结合了 Res2Net 模块。为了探究这些模块的加入对网络功能的影响,本文进行了消融实验。

实验设定:A1 为传统 VGG-16 网络模型;A2 为在 A1 基础上进行迁移学习的 VGG-16 网络模型;A3 为将 A2 预训练模型 VGG-16 中的传统卷积层替换成深度可分离卷积层的网络模型;A4 为以 A3 为基础在预训练模型中加入 1 个 Res2Net 模块的网络模型,即本文的 TR-VGG-16 模型;A5 为在 A4 中多加 1 个 Res2Net 模块的网络模型;A6 为以 A5 为

基础再多添加 1 个 Res2Net 模块的网络模型。不同模型准确率和参数减少量对比结果见表 3。

表 3 不同模型准确率和参数减少量对比

Tab. 3 Comparison of accuracy and parameter reduction of different models

模型	准确率/%	参数量/M
A1	70	138.95
A2	89.33	38.27
A3	90	12.65
A4	96.33	11.79
A5	94.33	37.28
A6	87.21	69.43

由表 3 可知,将 A1 与 A2、A3 网络准确率进行对比,引入迁移学习的 A2、A3 网络的识别准确率分别为 89.33% 和 90%,均高于 A1 的 70%;由 A3 与 A1、A2 的参数量对比可得,随着迁移学习和深度可分离卷积的引入,网络参数量由原来的 138.95 M 分别减少至 38.27 M 和 12.65 M。由此可知,迁移学习和深度可分离卷积的使用,不仅可以提高网络识别准确率,相较于传统 VGG-16 网络还能明显减少网络结构的参数量。

为进一步证明 Res2Net 模块的加入对网络识别准确率的影响,对加入不同数量的 Res2Net 模块的网络进行准确率验证。由表 3 的 A3、A4、A5、A6 模型可以看出,加入 1 个 Res2Net 模块时,其网络的识别准确率最高,达到 96.33%。而随着加入的 Res2Net 模块增多,其识别率反而大幅度下降。究其原因,由于 Res2Net 模块的不断叠加,使网络模型深度加深,网络参数量剧增,训练过程出现过拟合现象,最终导致网络识别准确率退减。

## 3 结束语

本文基于改进的 VGG-16 分类网络提出一种煤岩显微组分图像识别网络模型 TR-VGG-16。旨在通过该模型提升对小样本煤岩显微图像组分识别的准确度。在实验中,引入迁移学习,解决小样本数据集训练问题。通过用深度可分离卷积代替传统卷积,用以减少网络参数量。并且将加入 Res2Net 模块的数量对模型识别准确率的影响进行验证分析。实验表明,本文方法可以在训练集数量较小的情况下对组分进行准确识别,同时还具有良好的训练稳定性。本文方法识别准确率为 96.33%,其收敛速度较快,且具有稳定的识别准确率。尽管本文网络模

(下转第 102 页)