



基于改进样本驱动的高光谱矿物识别模型压缩方法

邓可望 赵慧洁 李娜 蔡辉

Improved data-driven compressing method for hyperspectral mineral identification models

Deng Kewang, Zhao Huijie, Li Na, Cai Hui

在线阅读 View online: <https://doi.org/10.3788/IRLA20210252>

您可能感兴趣的其他文章

Articles you may be interested in

基于全光衍射深度神经网络的矿物拉曼光谱识别方法

Raman mineral recognition method based on all-optical diffraction deep neural network

红外与激光工程. 2020, 49(10): 20200221-1-20200221-8 <https://doi.org/10.3788/IRLA20200221>

利用双神经网络的相机标定方法

Camera calibration method based on double neural network

红外与激光工程. 2021, 50(11): 20210071-1-20210071-9 <https://doi.org/10.3788/IRLA20210071>

基于神经网络的非线性大气修正实现红外目标辐射测量

Nonlinear atmospheric correction based on neural network for infrared target radiometry

红外与激光工程. 2020, 49(5): 20190413-20190413-8 <https://doi.org/10.3788/IRLA20190413>

基于卷积神经网络的反无人机系统图像识别方法

Image recognition method of anti UAV system based on convolutional neural network

红外与激光工程. 2020, 49(7): 20200154-1-20200154-8 <https://doi.org/10.3788/IRLA20200154>

高光谱-LiDAR 融合的条件随机场分类方法

Conditional random field classification method based on hyperspectral-LiDAR fusion

红外与激光工程. 2021, 50(12): 20210112-1-20210112-12 <https://doi.org/10.3788/IRLA20210112>

PCA特征提取和弹性BP神经网络的水下气泡识别

Underwater bubbles recognition based on PCA feature extraction and elastic BP neural network

红外与激光工程. 2021, 50(6): 20200352-1-20200352-7 <https://doi.org/10.3788/IRLA20200352>

基于改进样本驱动的高光谱矿物识别模型压缩方法

邓可望¹, 赵慧洁^{1,2}, 李娜^{1*}, 蔡辉³

- (1. 北京航空航天大学 仪器科学与光电工程学院, 北京 100191;
2. 北京航空航天大学 北航青岛研究院, 山东 青岛 266101;
3. 中国人民解放军 96901 部队, 北京 300140)

摘要: 针对机载成像高光谱遥感仪器获得的大批量高光谱数据很难实现高效快速的矿物信息提取和识别的问题, 提出了一种基于改进样本驱动的高光谱矿物识别模型压缩方法, 对神经网络中的冗余神经元进行剪枝, 从而获取高效的矿物识别模型。首先, 以验证数据集中的正确识别样本为数据驱动, 计算各神经元经激活函数后的输出零值频率, 并将其作为该神经元重要性判据, 探讨各神经元对神经网络正确识别样本的贡献; 其次, 通过设置重要性阈值对冗余神经元进行剪枝, 并对剪枝网络进行再训练, 在保留原网络正确识别特性的基础上, 提升压缩模型识别精度; 最终通过多次迭代剪枝获得高效的压缩矿物识别模型。利用基于改进样本驱动模型压缩方法对基于多层感知机的矿物识别模型进行压缩改进, 并以美国内华达州 Cuprite 矿区的机载可见光/红外成像光谱仪的高光谱数据作为测试数据, 获得了压缩比 3.33、矿物识别精度 94.35% 的高效矿物识别模型。

关键词: 神经网络; 网络剪枝; 高光谱; 矿物识别

中图分类号: TP751 **文献标志码:** A **DOI:** 10.3788/IRLA20210252

Improved data-driven compressing method for hyperspectral mineral identification models

Deng Kewang¹, Zhao Huijie^{1,2}, Li Na^{1*}, Cai Hui³

- (1. School of Instrumentation Science and Opto-Electronic Engineering, Beihang University, Beijing 100191, China;
2. Beihang University Qingdao Research Institute, Beihang University, Qingdao 266101, China;
3. Unit 96901 of the People's Liberation Army of China, Beijing 300140, China)

Abstract: It was difficult to extract mineral features efficiently and quickly from large quantities of hyperspectral data obtained by airborne imaging hyperspectral spectrometers. An improved data-driven compressing method for mineral identification models was proposed in this paper, which pruned redundant neurons in neural networks to obtain efficient mineral identification models. Firstly, the average percentage of zeros driven by correctly identified samples in the validation set (*C-APoZ*) of each neuron was calculated as a criterion of importance for the neuron, so as to explore the contribution of the neuron to the network for identifying samples correctly. Then, the redundant neurons were pruned by setting the importance threshold, and

收稿日期: 2021-12-10; 修订日期: 2022-01-25

基金项目: 国家重点研发计划 (2016YFB0500505, 2017YFC0602104); 国家自然科学基金 (61975004); 高分国土资源遥感应用示范系统 (二期)(04-Y30B01-9001-18/20); 青岛创业创新领军人才计划 (18-1-2-22-zhc)

作者简介: 邓可望, 男, 博士生, 主要从事高光谱图像识别方面的研究。

导师简介: 赵慧洁, 女, 教授, 博士生导师, 博士, 主要从事多/高光谱智能感知技术、多模智能感知与数据融合技术等方面的研究。

通讯作者: 李娜, 女, 副教授, 博士生导师, 博士, 主要从事遥感图像处理与识别方面的研究。

the pruned network was retrained to improve the identification accuracy while preserving the correct identification abilities of the original network. Finally, an efficient compressed model for mineral identification was obtained through multiple iterative pruning. In this paper, the improved data-driven compressing method was conducted on the mineral identification models based on multilayer perceptron (MLP) to promote their efficiency. The hyperspectral data of the Nevada mining area collected by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) were applied to evaluate the proposed method. The results show that the proposed method obtained an efficient model for mineral identification with the compression rate of 3.33 and the identification accuracy of 94.35%.

Key words: neural network; network pruning; hyperspectral; mineral identification

0 引言

随着高光谱技术的不断发展,光谱分辨率与空间分辨率的不断提高,为高光谱矿物识别提供了有效技术手段。传统的矿物识别方法主要基于矿物光谱的相似性或者光谱的诊断特征,较为依赖矿物光谱且容易受成像条件的影响^[1-3]。深度学习技术在图像分类和识别领域取得了有效应用,极大地提升了模型的鲁棒性,越来越多的学者将神经网络应用于高光谱矿物识别^[4-5]。但是,该类神经网络模型由于网络结构复杂,网络参数庞大,很难满足许多高光谱应用中对矿物的快速信息提取和高效识别的要求。尽管 Zhang 等^[6]将 GPU 技术应用于矿物识别应用中,提高了计算速度,但是,针对矿物识别模型效率提升的模型结构研究还比较缺乏。Denil^[7]等提出神经网络存在着大量的冗余参数,仅需要部分网络参数就能获得与原网络相同的效果。除此之外, Han^[8]等提出过多的网络参数会带来过拟合现象,导致识别精度的下降。为了获取高效的矿物识别模型,必须去除神经网络中的冗余单元,因此,对矿物识别模型进行模型压缩成为必要的手段。目前,神经网络模型压缩主要分为网络剪枝、参数量化、设计结构化矩阵、知识蒸馏这四类方法^[9]。其中,网络剪枝方法直接作用于神经网络中的冗余单元,在降低模型规模的同时,还能减少过拟合程度,显著提升模型效能,得到了广泛应用^[10]。

网络剪枝的核心思想是建立网络元素的重要性判别依据,对重要性低的冗余元素进行剪除,并通过再训练获得新的压缩模型,其主要包括非结构化网络剪枝和结构化网络剪枝^[11]。非结构化网络剪枝主要针对神经网络中不同网络层之间的冗余连接进行剪枝,以实现模型压缩^[12-14]。但是,该类网络剪枝方法

可能会对部分重要连接进行误剪枝,并缺少相应的连接恢复过程,从而导致网络不收敛或者精度严重下降。Guo 等^[15]提出动态网络外科手术方法,对误剪枝的重要连接进行恢复,在获得高压缩比的同时,几乎没有精度损失。然而,非结构化网络剪枝使得权值参数矩阵呈现稀疏化,不利于硬件的加速,并且需要专门的软件进行支持。结构化剪枝主要针对神经网络中的冗余神经元、卷积核或者通道进行剪枝,以实现网络压缩^[16-17]。结构化网络剪枝将网络单元进行整体剪除,不需要特殊的硬件或者软件支持,得到了更为广泛的应用。然而,现有的结构化剪枝方法通常将所有的验证样本作为神经网络单元重要性的数据驱动,并未考虑误识别样本可能造成的影响。

因此,针对上述问题,文中提出了一种基于改进样本驱动的高光谱矿物识别模型压缩方法,对神经网络中冗余神经元进行剪枝,以得到高效的矿物识别模型。该方法以验证数据集中正确识别样本为基础,首先计算各神经元经激活后输出零值频率,即正确识别样本驱动的激活输出零值率 (Average Percentage of Zeros driven by Correctly identified samples, C-APoZ),并将 C-APoZ 作为神经元重要性依据,然后根据 C-APoZ 值对神经网络进行迭代剪枝,实现对原网络的高效压缩。为验证方法的有效性,将提出的模型压缩方法应用于基于多层感知机的矿物识别模型,并以美国内华达州 Cuprite 矿区的 AVIRIS 高光谱数据作为测试数据。

1 基于改进样本驱动的高光谱矿物识别模型压缩方法

目前,深度学习神经网络模型已经在高光谱矿物识别方面得到了应用^[5],文中提出的基于改进样本驱

动的高光谱矿物识别模型压缩方法将这些模型视作预训练模型,在此基础上进行网络剪枝。该方法利用神经元的 $C-APoZ$ 作为重要性判断依据,对预训练神经网络逐层进行剪枝,并对剪枝网络进行再训练,通过多次迭代实现对高光谱矿物识别模型的高效压缩。

1.1 基于改进样本驱动的神元重要性判据

基于神经网络的高光谱矿物识别模型,由于网络层数多,每层的神元数目多,使得不同网络层之间的连接较多,导致整个矿物识别模型复杂,计算时间长。网络剪枝方法可以剪除神经网络中的冗余单元,实现对模型的压缩,其示意图如图 1 所示。虚线部分为网络剪枝方法去除的冗余元素。

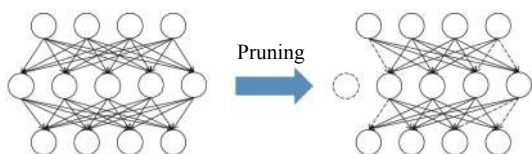


图 1 网络剪枝示意图

Fig.1 Schematic of network pruning

神经网络大部分的参数存在于全连接层中,故针对全连接层的网络剪枝成为重要的研究方向^[7]。若依据连接重要性进行非结构化剪枝,会造成权值矩阵稀疏化,需要专门的硬件和软件支持,故将其转化为对神经元的结构化剪枝。为了探究全连接神经网络中各神经元的重要性,研究相邻三层网络层组成的局部神经网络,分别记作 $L-1$ 层、 L 层和 $L+1$ 层。为了方便数据在神经网络中的表示,令 X_i 为第 i 层网络层的输出, W_i 和 b_i 分别为连接第 $i-1$ 层网络层和第 i 层网络层的权重矩阵和偏置向量。局部神经网络的各网络层输出之间的关系如下式所示:

$$X_L = \sigma(W_L X_{L-1} + b_L) \quad (1)$$

$$X_{L+1} = \sigma(W_{L+1} X_L + b_{L+1}) \quad (2)$$

式中: $\sigma(\cdot)$ 为神经网络的激活函数。为了克服网络的梯度消失问题,并且加速网络的收敛速度,通常选择 ReLU 函数作为全连接层的激活函数,其计算公式为:

$$\text{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3)$$

由公式 (1)~(3) 可得出,若 X_L 中第 j 个神经元的输出 X_L^j 为 0 时,则与其相连的权重在该次计算过程中不

再发生任何作用,且对最后的识别结果不产生任何影响,可视其为冗余神经元。通常采用验证样本集在神经网络模型上的表现来衡量该模型的效能,因此,可利用验证样本集作为数据驱动来探讨神经网络各神经元的冗余程度。当以整个验证样本集作为数据驱动时,若某一神经元的输出零值次数越多,则其对神经网络输出结果的影响越小,其重要性越低,越有可能成为冗余单元。Hu 等^[17] 提出以激活输出零值率 (Average Percentage of Zeros, APoZ) 作为神经元重要性判断依据,采用基于样本驱动的剪枝方法对神经网络进行剪枝,获得了较为高效的压缩模型。神经元的 APoZ 值越大,其重要性越低,被剪枝的优先级越高。第 L 层网络层的第 i 个神经元的 $APoZ_L^i$ 的计算公式为:

$$APoZ_L^i = \frac{\sum_{k=1}^N f(X_L^i(k))}{N} \quad (4)$$

式中: N 为验证样本的数量; $X_L^i(k)$ 为第 k 个验证样本在第 L 层网络层的第 i 个神经元的激活输出,若 $X_L^i(k)=0$, 则 $f(\cdot)$ 的值为 1, 否则, $f(\cdot)$ 的值为 0。

基于样本驱动的网络剪枝方法以整个验证样本集作为数据驱动,逐网络层对 APoZ 值高的神经元进行剪枝,实现网络压缩。由于被剪枝的神经元输出零值频率高,对神经网络的输出结果产生影响小,故剪枝后的网络与剪枝前的网络在验证样本集上具有相似的输出。因此,剪枝网络不仅保留了原网络对大部分样本正确识别的能力,还保留了对部分样本错误识别的能力,这可能导致识别精度下降。

针对这一问题,文中提出基于改进样本驱动的剪枝方法,以验证样本集中的正确识别样本作为数据驱动来判断神经元的重要性,即利用 $C-APoZ$ 作为神经元重要性判据。第 L 层网络层的第 i 个神经元的 $C-APoZ_L^i$ 的计算公式为:

$$C-APoZ_L^i = \frac{\sum_{k=1}^N f(X_L^i(k)) \cdot Flag_k}{\sum_{k=1}^N Flag_k} \quad (5)$$

式中: 当第 k 个样本为正确识别样本时, $Flag_k$ 为 1, 否则, $Flag_k$ 为 0。

对于一个能识别 K 类矿物的神经网络矿物识别

模型,当以验证样本作为数据驱动时,若识别正确,则识别模型的输出类别标签唯一;若识别错误,则输出类别标签可能存在 $K-1$ 种可能,故错误识别样本对于模型具有不可预测性,很难评价其对于模型神经元的影响。利用文中提出的 $C-APoZ$ 来衡量神经网络中各神经元的重要性时,对验证样本进行筛选,仅以正确识别验证样本作为驱动,忽略了错误识别样本可能带来的随机影响,最大程度保留了原模型对验证样本集正确识别的能力,是剪枝网络模型效能提升的重要保障。除此之外,基于 $C-APoZ$ 神经元判据相较于传统的基于 $APoZ$ 神经元判据,去除了错误识别样本对于剪枝网络的限制,降低了模型过拟合程度,有助于提升剪枝网络的识别精度。

1.2 网络剪枝和再训练

考虑到连接仅存在全连接神经网络的相邻网络层之间,且前网络层的输出才会对后网络层的输出产生影响,故必须逐网络层进行网络剪枝。为了确定神经元是否被剪枝,需要逐网络层设置合适的 $C-APoZ$ 阈值,对冗余神经元进行剪枝,降低网络的冗余性。文中选取第 L 层网络层 $C-APoZ$ 值的平均值与标准差之和作为该层神经元重要性阈值 T_L ,对于该层的第

i 个神经元,若 $C-APoZ_i > T_L$,则此神经元被剪除;否则,将其进行保留。在对神经网络进行剪枝后,剪枝网络的识别能力会呈现一定程度的下降,需要对其进行再训练,对模型参数进行微调,以获得稳定的压缩网络。另外,单次的网络剪枝很难获得高压缩比的压缩网络,需要进行多次迭代剪枝,才能获得高效的压缩矿物识别模型。

1.3 基于改进样本驱动的高光谱矿物识别模型压缩

全连接层能够对浅层特征进行特征重组,获得更为丰富的深层次特征,同时也能提升神经网络模型对特征的非线性表达程度,被广泛应用于高光谱矿物特征提取。但是,全连接层也带来了大量的参数,增加了模型的冗余程度和过拟合程度,成为高光谱矿物识别模型效能提升需要重点关注的对象。文中提出基于改进样本驱动的神经元重要性判据 $C-APoZ$, 定量化描述全连接层中各神经元的冗余程度,并通过设置合适的阈值对冗余神经元进行剪枝,多次迭代后获得高压缩比和高识别精度的压缩矿物识别模型,其流程图如图 2 所示。其中,Epoch 为设置的迭代总次数,Iteration 为当前迭代次数。

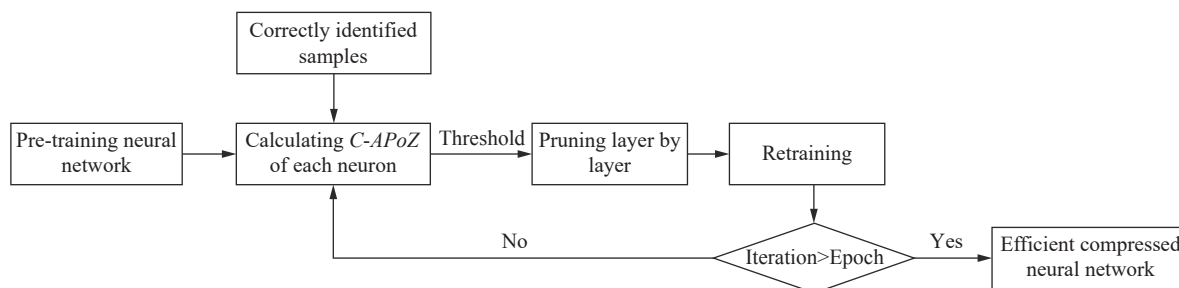


图 2 基于改进样本驱动的高光谱矿物识别模型压缩方法流程图

Fig.2 Flow chart of improved sample-driven compression method for hyperspectral mineral identification model

文中提出的基于改进样本驱动的高光谱矿物识别模型压缩方法具体实现步骤如下:

- (1) 加载预训练高光谱矿物识别模型;
- (2) 计算全连接网络层各神经元的重要性判据 $C-APoZ$, 并计算各层的 $C-APoZ$ 阈值;
- (3) 逐网络层对 $C-APoZ$ 高于阈值的冗余神经元进行剪枝;
- (4) 对剪枝网络进行再训练,使其达到稳定状态;

- (5) 重复步骤 (2)~步骤 (5), 直至达到迭代次数;
- (6) 获得高效能的压缩矿物识别模型。

2 算法性能验证

在验证过程中,选取基于改进光谱吸收指数向量的多层感知机矿物识别模型^[5]作为预训练神经网络模型。该模型以多层感知机为基础,包含输入层、隐藏层和输出层,各网络层之间为全连接层。其中,输

入层为改进光谱吸收指数向量;隐藏层为一层包含多个隐藏单元的网络层;输出层为矿物类别标签。该实验主要将基于改进样本驱动的网络剪枝方法应用于预训练神经网络以获得压缩矿物识别模型,并且探讨压缩程度和识别精度变化。

2.1 实验数据

该实验选取 AVIRIS 传感器采集的美国内华达

州 Cuprite 矿区的高光谱数据作为测试数据。该数据的大小为 400×350 pixel, 包含 50 个波段, 光谱范围为 1 990~2 480 nm。Clark 等^[18]利用 Teracorder 系统对该区域 AVIRIS 数据进行矿物填图, 获得了相应的地面真值图。内华达州 Cuprite 矿区的假彩色图和地面真值图如图 3 所示。

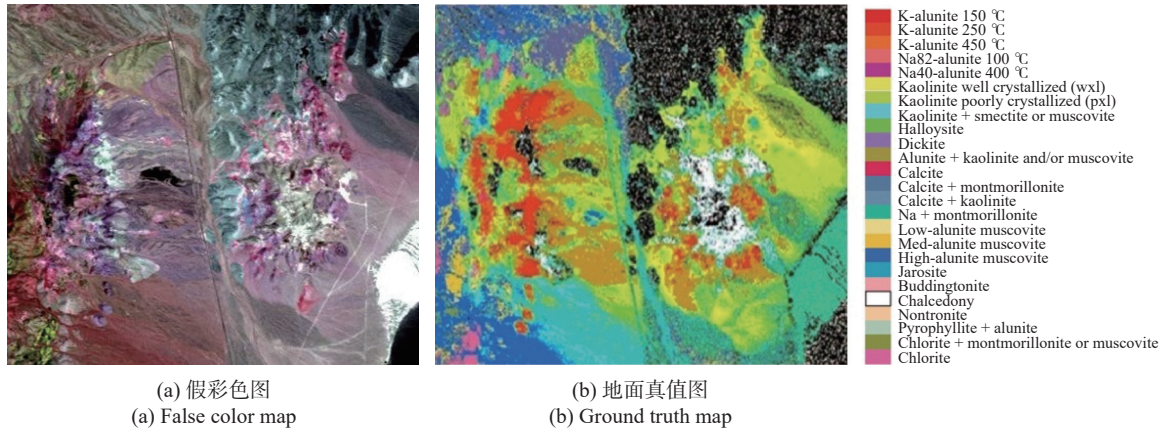


图 3 内华达州 Cuprite 矿区高光谱数据

Fig.3 Hyperspectral datasets of the Cuprite mine in Nevada

实验选取内华达州 Cuprite 矿区 7 类分布广泛的矿物进行识别, 它们分别是白云母、高岭石、多水高岭石、方解石、蒙脱石、明矾石和玉髓石。各类矿物随机选取 100 个样本作为训练样本, 白云母、高岭石、多水高岭石、蒙脱石 4 类矿物随机选取 400 个样本作为验证样本, 而多水高岭石、方解石、玉髓石 3 类矿物随机选取 240 个样本作为验证样本。7 类待识别矿物的信息如表 1 所示。

表 1 待识别矿物信息表

Tab.1 Information of identified minerals

Class name	Training samples	Testing samples	Diagnostic bands/nm
Muscovite	100	400	2 200, 2 350
Halloysite	100	240	2 170, 2 210
Calcite	100	240	2 160, 2 340
Kaolinite	100	400	2 170, 2 210
Montmorillonite	100	400	2 230
Alunite	100	400	2 170, 2 320
Chalcedony	100	240	2 250
Total	700	2 320	

2.2 预训练神经网络

该实验选取基于改进光谱吸收指数向量的多层感知机矿物识别模型作为预训练神经网络模型^[5]。该网络模型最重要的参数为隐藏层神经元个数 h , 为了方便后续剪枝操作, 需选取合适的 h 使预训练神经网络模型的识别精度达到局部极值点。选取不同的 h , 预训练神经网络模型的识别精度如表 2 所示。

表 2 不同隐藏层神经元数目对应的模型识别精度

Tab.2 Model identification accuracy corresponding to the different number of hidden units

h	Overall accuracy
10	91.98%
15	92.54%
20	93.01%
25	93.36%
30	93.62%
35	93.14%
40	92.76%
45	92.50%

由表 2 可知,当隐藏层神经元数目 h 为 30 时,预训练神经网络模型的识别精度达到极大值 93.62%,将此时的模型状态作为剪枝的初始状态。

3 结果与分析

3.1 网络剪枝结果

预训练神经网络包含三层网络层,其中,输入层与输出层不能进行网络剪枝,故仅针对隐藏层进行网络剪枝。选取验证样本集中的 $2320 \times 93.62\% \approx 2172$ 个正确识别样本作为文中提出的基于改进样本驱动的剪枝方法的数据驱动,计算隐藏层 30 个神经元的 $C-APoZ$ 值。为了与传统的基于样本驱动的网络剪枝方法进行比较,以整个验证样本集作为数据驱动,计算隐藏层各单元的 $APoZ$ 值。除此之外,分别计算隐藏层的 $C-APoZ$ 阈值和 $APoZ$ 阈值,结果如图 4 所示。

由图 4 可知,由于正确识别样本占据了验证样本中的绝大部分,使得隐藏层各单元的 $C-APoZ$ 值和 $APoZ$ 值变化不大。但是,它们的重要性阈值发生了变化,从而使得剪枝结果发生变化。分别以 $C-APoZ$ 和 $APoZ$ 作为神经元重要性判据对预训练网络进行网络剪枝,并对剪枝网络进行再训练,以压缩比和识别精度对压缩网络进行评价,结果如表 3 所示。其中,

压缩比为神经网络压缩前后所占存储空间的比值。

从表 3 可以看出,以文中提出的 $C-APoZ$ 作为神经元重要性判据进行剪枝时,能够获得压缩比和识别精度更高的压缩网络。为了验证被剪枝神经元的冗余性,图 5 给出了 Cuprite 矿区在被剪枝神经元的输出图像。

由图 5 可以直观发现,被剪枝的神经元部分所含信息较少,部分具有相似的输出,故可认为它们均为冗余单元。如第 1、12、17 神经元对于任何输入数据,其输出均为零值,其对网络几乎不产生影响;第 6、23、27 神经元具有较为相似的输出,其对网络有着相似的影响。分别以 $C-APoZ$ 和 $APoZ$ 作为神经元重要性判据进行多次迭代网络剪枝,表 4 中给出了 7 次迭代剪枝的结果。

在选取高效能的压缩矿物识别模型时,需要综合考虑其压缩比和识别精度,保证矿物识别模型总体识别精度的基础上,尽可能地提升模型的压缩比,去除模型中的冗余单元。图 6 给出了基于不同重要性依据的网络剪枝方法在迭代过程中获得的压缩识别模型对应的模型参数。

为了分析文中提出的剪枝方法对识别效果的影响,图 7 给出了原始神经网络模型和迭代剪枝 5 次获

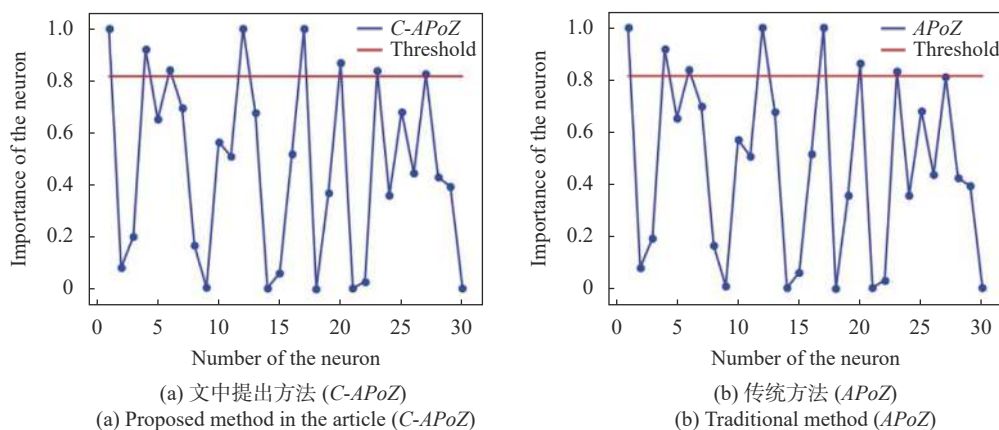


图 4 隐藏层神经元的重要性结果图

Fig.4 Importance diagram of neurons in hidden layer

表 3 不同网络剪枝方法结果表

Tab.3 Result chart of different pruning methods

Importance criteria	Sequence number of the pruned neuron	Number of pruned units	Compression rate	Identification accuracy after retraining
Proposed $C-APoZ$	1, 4, 6, 12, 17, 20, 23, 27	8	1.36	94.61%
$APoZ$	1, 4, 6, 12, 17, 20, 23	7	1.30	94.57%

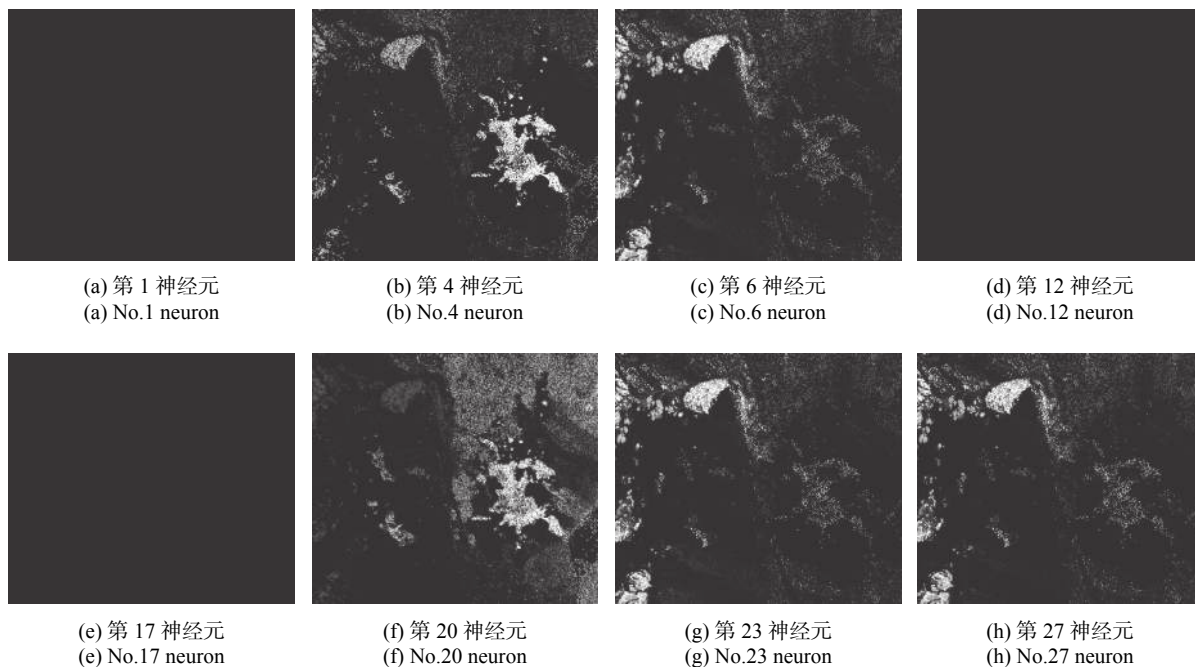


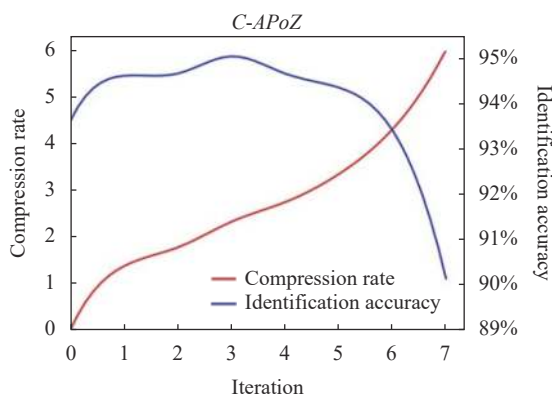
图 5 被剪枝神经元输出结果图

Fig.5 Output results of the pruned neurons

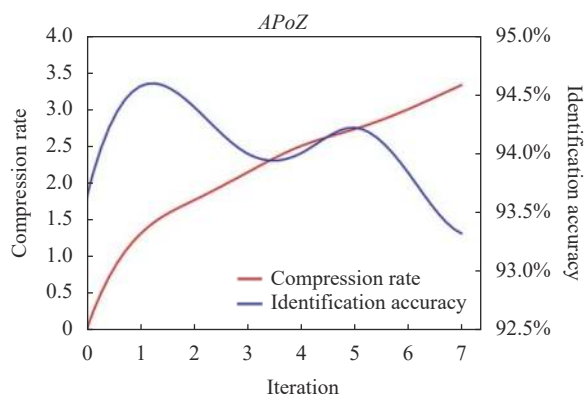
表 4 基于不同重要性判据的迭代剪枝结果表

Tab.4 Iterative pruning results based on different importance criteria

Iteration	<i>C-APoZ</i> (Proposed method)			<i>APoZ</i>		
	Compression rate	Threshold	Identification accuracy	Compression rate	Threshold	Identification accuracy
0	0	0.817	93.62%	0	0.814	93.62%
1	1.36	0.651	94.61%	1.30	0.643	94.57%
2	1.76	0.502	94.66%	1.76	0.597	94.40%
3	2.31	0.464	95.04%	2.14	0.469	94.00%
4	2.73	0.445	94.66%	2.50	0.426	94.00%
5	3.33	0.448	94.35%	2.73	0.444	94.22%
6	4.29	0.379	93.41%	3.00	0.435	93.84%
7	6.00	0.316	90.13%	3.33	0.412	93.32%



(a) 文中提出方法 (*C-APoZ*)
(a) Proposed method in the article (*C-APoZ*)



(b) 传统方法 (*APoZ*)
(b) Traditional method (*APoZ*)

图 6 迭代剪枝过程中压缩模型参数变化图

Fig.6 Parameters of the compressed identification models obtained in the iterative pruning process

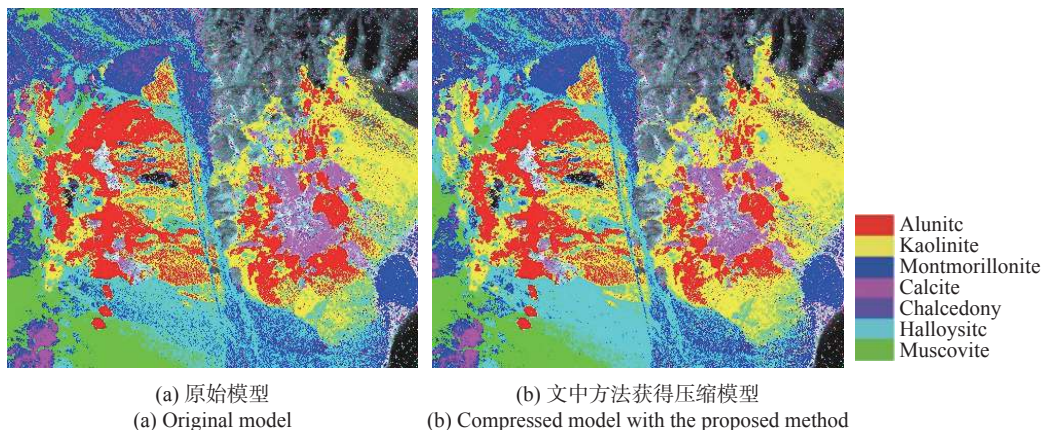


图 7 内华达州矿物识别结果图

Fig.7 Identification results of the Nevada mining area

得的压缩模型在实验数据区域的矿物识别结果图。

3.2 分析与讨论

文中利用基于改进样本驱动的网络剪枝方法对基于改进光谱吸收指数向量的多层感知机矿物识别模型进行网络剪枝,并且与传统的基于样本驱动的网络剪枝方法进行比较,剪枝结果如表 4 所示。两种剪枝方法获得的压缩模型在识别精度方面均得到一定程度的提高,说明预训练神经网络模型存在着过拟合现象,对其进行网络剪枝是十分必要的。随着迭代剪枝的进行,压缩模型的重要性阈值不断下降,说明神经元 $C-APoZ$ 值整体下降,神经网络模型的冗余性也随之下降。在相同压缩比的情况下,如压缩比为 1.76、2.73 和 3.33 时,尽管识别模型结构一样,但是压缩模型的形成过程不同,使得模型处于不同的稳定状态,参数达到了不同的局部最优值,而基于改进样本驱动的网络剪枝方法总能够获得识别精度更高的压缩模型,充分说明文中提出的剪枝方法能够尽可能保留原模型正确识别样本的能力。如图 6 所示,在 5 次迭代剪枝之前,压缩模型识别精度变化不大,压缩比进一步提升,模型效能不断提升;在迭代过程超过 5 次以后,图 6(a) 中压缩模型的压缩比尽管显著提升,其识别精度也随之明显下降,而图 6(b) 中压缩模型的压缩比提升缓慢,其识别精度也显著下降。因此,文中认为经过 5 次迭代剪枝后,两类剪枝方法所获得的压缩模型效能达到最高。此时,基于改进样本驱动的网络剪枝方法获得压缩比和识别精度分别为 3.33 和 94.35% 的高效压缩识别模型,在两种评价指标上均优于传统的基于样本驱动的网络剪枝方法所

获得的压缩模型,体现了文中提出方法在高光谱矿物识别模型压缩方面的优越性。另外,图 7 中,基于改进样本驱动的网络剪枝方法尽管去掉了大量的神经元,大大压缩了模型规模,但是仍然能获得与原模型相似的识别结果图,同样展示了文中提出方法在模型压缩方面的高效性。

4 结 论

文中提出基于改进样本驱动的网络剪枝方法,针对高光谱矿物识别模型中存在的冗余神经元进行剪枝,在降低神经网络冗余性和过拟合的同时,提升了矿物识别精度,获得了高效的压缩矿物识别模型。该方法相较于传统的基于样本驱动的网络剪枝方法,摒弃了误识别样本对神经元重要性的影响,只对原模型正确识别样本的能力进行保留。实验结果表明,利用文中提出的剪枝方法对已有高光谱矿物识别模型进行网络剪枝,获得了压缩比为 3.33 的高效压缩识别模型,对 Cuprite 矿区高光谱数据的识别精度也由 93.62% 提升到了 94.35%,充分体现了该剪枝方法的高效性。后续需要对误识别样本对神经元重要性带来的负面影响进行分析,以获得更加全面的神经元重要性判据。

参考文献:

[1] Balas C, Epitropou G, Tsapras A, et al. Hyperspectral imaging and spectral classification for pigment identification and mapping in paintings by El Greco and his workshop [J]. *Multimedia Tools and Applications*, 2018, 77: 9737-9751.

- [2] Amer R, Mezayen A A, Hasanein M. ASTER spectral analysis for alteration minerals associated with gold mineralization [J]. *Ore Geology Reviews*, 2016, 75: 239-251.
- [3] Hecker C, Ruitenbeck F, Werff H, et al. Spectral absorption feature analysis for finding ore: A tutorial on using the method in geological remote sensing [J]. *IEEE Geoscience and Remote Sensing Magazine*, 2019, 7(2): 51-71.
- [4] Okada N, Maekawa Y, Owada N, et al. Automated identification of mineral types and grain size using hyperspectral imaging and deep learning for mineral processing [J]. *Minerals*, 2020, 10(9): 809.
- [5] Deng Kewang, Zhao Huijie, Li Na, et al. Identification of minerals in hyperspectral imagery based on the attenuation spectral absorption index vector using a multilayer perceptron [J]. *Remote Sensing Letters*, 2021, 12(5): 449-458.
- [6] Zhang Minghua, Zou Yaqing, Song Wei, et al. GGCN: GPU-based hyperspectral image classification algorithm [J]. *Laser & Optoelectronics Progress*, 2020, 57(20): 231-237. (in Chinese)
- [7] Denil M, Shakibi B, Dinh L, et al. Predicting parameters in deep learning [J]. *Neural Information Processing Systems (NIPS)*, 2012: 1097-1105.
- [8] Han S, Mao H, Dally W. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding [J]. *Fiber*, 2015, 56(4): 3-7.
- [9] Cheng Yu, Wang Duo, Zhou Pan, et al. A survey of model compression and acceleration for deep neural networks [J]. *IEEE Signal Processing Magazine*, 2017, 35(1): 126-136.
- [10] Wang Yulong, Zhang Xiao, Xie Lingxi, et al. Pruning from scratch[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12273-12280.
- [11] Hassibi B, Stork D, Wolff G. Optimal brain surgeon and general network pruning[C]//IEEE International Conference on Neural Networks, 1993: 293-299.
- [12] Cun Y, Denker J, Solla S. Optimal brain damage[J]. *Advances in Neural Information Processing Systems*, 1990, 2: 598-605.
- [13] Hassibi B. Second order derivatives for network pruning: Optimal brain surgeon [J]. *Advances in Neural Information Processing Systems*, 1992, 5: 164-171.
- [14] Roy S, Panda P, Srinivasan G, et al. Pruning filters while training for efficiently optimizing deep learning networks[C]//2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-7.
- [15] Guo Yiwen, Yao Anbang, Chen Yurong. Dynamic network surgery for efficient DNNs[C]//Neural Information Processing Systems, 2016: 1387-1395.
- [16] Luo Jianhao, Wu Jianxin, Lin Weiyao. ThiNet: A Filter Level Pruning Method for Deep Neural Network Compression[C]//IEEE International Conference on Computer Vision (ICCV), 2017: 5068-5076.
- [17] Hu Hengyuan, Peng Rui, Tai Yuwing, et al. Network trimming: A data-driven neuron pruning approach towards efficient deep architectures [J]. *arXiv preprint arXiv*, 2016: 1607.03250.
- [18] Clark R N, Swayze G A, Livo K E, et al. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems [J]. *Journal of Geophysical Research*, 2003, 108(E12): 5131.