基于卷积神经网络的岩石样品智能分类设计与实现

张钰
武汉大学，武汉 430072，中国

摘要：随着计算机硬件性能的大幅提升和人工智能研究的快速发展，人工智能识别技术已应用于医学、教育、工业、商业、勘探等众多行业专业领域。在地质勘探中，岩石样品的分类识别是地质分析的重要环节。然而，传统的人工分类方法面临着较高的经济和时间成本，且容易受到主观判断的影响而影响识别结果。为了避免这些问题，研究了基于计算机的人工智能识别技术在岩石图像识别中的具体应用以及不同参数造成的性能差异。实验中使用的样本数据为测井现场工业相机采集的岩屑和岩心样品，共计350幅图像，涵盖7种类型的岩石样品。本文将这350幅岩石图像按一定比例划分为训练集、验证集和测试集。由于数据集较小且各类岩石的数量各不相同，因此在训练之前采用数据增强技术对原始数据进行扩充，扩充过程中均衡了不同类型岩石样品的数量。在训练阶段，利用开源深度学习框架Pytorch构建多层卷积神经网络，对训练集进行学习和分类训练，测试中最佳准确率达到82.86%。然而，对于拉伸图像的识别率较低，需要进一步优化网络结构。基于实验结果总结了神经网络训练的优化方向，包括提升数据集质量、优化网络结构、调整训练策略等。为了尽可能直观地展示实验数据，实验中使用PlotNeuralNet、Python matplotlib等数据可视化工具进行具体的可视化工作，并从数据与图像相结合的角度对实验数据及模型效果进行分析。

关键词：深度学习；卷积神经网络；岩石样品图像识别

The Design and Realization of Intelligent Classification of Rock Samples Based on Convolutional Neural Network

Zhag Yu
Wuhan University, Wuhan 430072, China

Abstract: With substantial improvement of computer hardware performance and rapid development in artificial intelligence research, artificial intelligence recognition technology has been applied in many industries, such as medicine, education, industry, commerce, exploration and other professional fields. In geological exploration, the classification and identification of rock samples is an important link in geological analysis. However, the traditional manual classification faces high economic and time costs, and is vulnerable to subjective judgment which may affect the recognition result. In order to avoid these problems, research is carried out on the specific application of computer-based artificial intelligent recognition technology in recognition of rock images and the performance difference caused by different parameters. The sample data used in the experiment are lithic fragments and core samples capture by industrial cameras at the well logging site, with a total of 350 images which cover 7 types of rock samples. In this paper, the 350 rock images are divided into training set, validation set and test set based on certain proportion. Since the data set is too small and the quantity of each type of rocks varies, data augmentation technology is used to expand the

© By the Author(s) 2025

original data before training. During the expansion, the quantity of different types of rock samples are balanced. At the training stage, the open-source deep learning framework Pytorch is used to construct a multi-layer convolutional neural network for learning and classification training on the training set. The best accuracy rate in the tests reaches 82.86%. However, the recognition rate is poor for stretched images, which required further optimization of the network structure. Based on the experimental results, the optimization orientations of neural network training are summarized, including improvement of data set quality, optimization of network structure, and adjustment of training strategies. In order to display the experimental data as intuitively as possible, data visualization tools such as PlotNeuralNet and Python matplotlib are used in the experiment for specific visualization work, after which the experimental data and model effect are analyzed from the perspective with combination of data and images.

Keywords: In-depth learning; Convolutional neural network; Recognition of rock sample images

对于计算机而言，图片以非结构化形式存储，信息无法直接获取，因此信息检索的效率受到极大限制。人工处理数亿张图像非常困难。在大数据时代，数据处理速度尤为关键。深度学习是实现智能识别的一种方式。数据量和模型规模的不断增长使得深度学习技术越来越可靠。同时，现代卷积神经网络 (CNN) 在物体识别领域的发展也显著提高了计算机图像识别的效率。

岩石图像识别是应用图像识别技术的地质研究领域之一。岩石样品类型的识别在油气勘探中具有重要意义。然而，传统的人工岩石样品图像识别存在诸多缺陷，包括：

(1) 效率低、成本高；

(2) 难以提取更深层次、更复杂的抽象特征；

(3) 识别结果容易受到主观判断的影响。

基于计算机人工智能识别技术的程序可以高效地从图像中提取特征，有效避免主观识别对分类结果的负面影响，降低人工识别的各种成本，为岩石样品图像识别提供了一种可行的新方法。

一、神经网络结构设计

CNN 卷积层的核函数可以对给定图像进行空间特征提取，例如边缘和形状。本质上，这些核函数通过反向传播学习到的权重，学习从图像中捕捉空间特征。此外，堆叠的卷积层可用于在后续的每一层中从空间特征中检测复杂的空间形状。因此，它们可以提取人类难以感知的高度抽象的图像特征。

为了更好地提取抽象特征，实验中神经网络使用了4个卷积层。

函数第一个参数表示输入通道数，即cin。由于每次卷积都需要接收上一层神经元的输入，因此每层的输入通道数与上一层的输出相关。函数第二个参数表示输出通道数，即cout。第三个参数表示核函数的大小。初始使用较大的核函数大小以便于提取更多全局特征[1]，后续卷积层使用较小核函数大小，以便识别小的局部特征。第四个参数表示核函数的步长，用于减小输出图像的尺寸[2]。初始使用较大的步长，后续步长均设置为1。第五个参数表示填充大小。填充通常用于在卷积后添加列和行的零，以保持空间大小不变，从而由于保留了边界处的信息而提高性能。为了保证模型的输入输出正确，padding大小统一设置为2。

池化层用于减小互相关运算输出的大小，从而减少网络的计算量和需要更新的参数。池化层概括了互相关运算输出的特征图中的特征。

池化操作类似于核函数。池化操作包括在互相关运算输出的每个通道上滑动一个二维矩阵，并概括矩阵覆盖区域内的某些特征。对于大小 2×2 的池化层，它将每个特征图中的像素或值的数量减少到四分之一。

常用的池化操作有两种：平均池化和最大池化。

本文采用最大池化。

最大池化用于计算每个特征图中指定数据量的最大值。最大池化能够突出采样或池化后特征图中最重要的特征。

最大池化更适合提取显著特征，而平均池化有时无法提取出好的特征，因为它会考虑所有值并生成平均值，这对于图像分类等任务来说可能效率较低。此外，由于平均池化会考虑所有值并将其作为输出传递到下一层，因此所有值都会被用于特征映射和输出，从而导致计算效率降低[3]。如果不需要考虑卷积层的所有输入，那么平均池化就不太合适。

总而言之，最大池化在分类问题中通常优于平均池化。因此，最好将其用作池化层。在代码中，每个卷积层后接一个池化层，池化层的大小指定为 2×2，步长为 2。步长的概念与卷积层中的相同。

为了有效地训练多层神经网络，需要一个类似于线性函数的非线性函数作为激活函数。此外，它还必须更加谨慎地处理激活函数和输入函数，并抑制饱和。实现该激活函数的节点或单元称为修正线性激活单元（ReLU）。

对于大于零的值，ReLU 表现出线性函数的特性。当使用反向传播训练神经网络时，它展现出许多线性激活函数的理想性质。然而，它是一个非线性函数，当输入为负值时，输出为 0。与传统的机器学习激活函数相比，ReLU 具有一些适合深度神经网络训练的优点，包括：

(1) 更好的反向传播：由于当输入为负值时导数为 0，神经元不会被激活，也不会变得越来越小，从而更不容易出现梯度消失。

(2) 更高的计算效率：由于函数定义简单，ReLU 仅包含比较、加法和乘法。因此，编译器可以通过位移操作来优化计算。同时，由于只有部分神经元被激活，ReLU 的计算效率远高于 sigmoid 和 tanh。

(3) 加速梯度下降：由于ReLU函数的线性和非饱和性，神经网络可以加速梯度下降，直至收敛到损失函数的全局最小值。鉴于这些优势，实验中优先选择ReLU作为神经网络训练的激活函数。

批量归一化是一种协调神经网络中多层调整参数的方法。它可以重新调整前一层的输出，并对每个单元的均值和方差进行归一化，以稳定学习效率。对神经网络中前一层的输出进行归一化可以避免后续层的输出分布在参数调整时发生显著变化，从而提高模型的稳定性和多层模型的训练效率。此外，批量归一化还可以优化性能，减少训练次数，并平滑相应的优化过程。

其定义见公式 3。γ 和 β 是大小为 C 的可学习参数向量。如果没有特别指定，γ 设置为 1，β 设置为 0。表示标准化后的数据。标准差通过 torch.var(input, unbiased=False) 计算。它取决于输出卷积层的数量。

全连接层通常构成网络的最后几层，其输入是最终池化层或卷积层扁平化后的输出。卷积层的输出代表数据的抽象特征。虽然输出可以直接扁平化并连接到输出层，但采用全连接层更容易进行非线性组合学习这些特征。全连接层将输出映射到样本空间并完成最终的分类。

在Pytorch中，全连接层是通过torch.nn.Linear()函数实现的，该函数对输入数据进行线性变换。该函数的第一个参数表示输入样本的大小，第二个参数表示每个输出样本的大小。为了提高执行效率，实验中使用了双层全连接。

PlotNeuralNet是一个开源的网络结构可视化工具。用其生成的网络结构，网络第一层的输入为80×80、1通道的图像。第一层卷积操作之后进行一个最大池化，每个池化层的核大小为2，步幅数为2。池化

之后再次进行批量归一化，最终窗口形状为40×40，从图像中可以观察到窗口大小的变化。将第一层的输出作为第二层的输入，以此类推。经过两次全连接后，结果映射到类别数7。

在机器学习或深度学习中，损失函数是一种评估算法，用于计算模型对数据集的有效性。其输出称为损失值（Loss）[4]。一般而言，在智能识别领域，损失值越小，神经网络对数据集的拟合效果越好；损失值越大，模型对数据集的预测效果越差。在模型改进中，损失可以很好地反映改进是否有效。损失函数与模型精度相关，是神经网络调整模型参数的关键组成部分。对于模型的每次预测，损失函数明确地给出了神经网络预测值与实际值之间的绝对差值。

交叉熵是最流行的损失函数之一。它是信息论中的一种度量方法。通常，它计算两个概率分布之间的差异。由于它最小化了预测概率分布与实际概率分布之间的距离，因此在分类问题中表现出色。在训练神经网络时，模型会根据损失函数（loss）的大小调整网络中的各项参数，以寻求较小的交叉熵损失。交叉熵的定义请参考公式4。其中，n表示类别数；ti表示实际情况；pi表示第i个类别的概率。

在Pytorch中，交叉熵是通过torch.nn.CrossEntropyLoss()函数实现的。在反向传播过程中，神经网络通过损失函数更新模型的参数。

优化器是一种用于调整模型参数（例如权重和学习率）以降低损失的算法。它通过调整模型来降低损失函数的输出，充当损失函数和模型参数之间的桥梁。损失函数就像一个优化指南针，它告诉优化器是否应该朝着正确的方向优化权重。如果调整方向错误，它也会向优化器提供反馈。最初，我们无法知道模型的哪些参数是合适的。然而，通过损失函数的反馈，最终可以找到损失更低的参数组合。

常用的优化器有很多，包括最基本的梯度下降法、作为梯度下降法变体之一的随机梯度下降法（SGD）、AdaGrad、RMSProp 和 Adam 等。虽然这些算法的思想各不相同，但它们都旨在最小化损失。鉴于此，可以采用易于调试的任何算法。在本实验中，采用 Adam 作为神经网络训练的优化算法。

Adam 是一种梯度下降优化算法[5]，适用于海量数据和复杂参数的情况。它是自适应学习率的优化算法之一。与其他优化算法相比，它需要的内存更少，但效率更高。

Adam 需要 6 个主要参数，分别是步长 Y，即学习率；矩估计的指数衰减率 β1 和 β2；用于数值稳定性的小常数 E；初始参数 θ；权重衰减系数 λ。其中，β1 和 β2 默认分别为 0.9 和 0.999；E 默认为 1e-8；λ 默认为 0。

在算法初始化中，β1 和 β2 初始化为 0，分别代表第一动量和第二动量。“动量”一词源自物理学，定义为质量乘以速度。在神经网络中，第一动量也称为标准动量，主要用于处理高曲率梯度[6]，以加快训练速度。第二动量采用 Nesterov 动量，它是第一动量的变体，在标准动量的基础上添加了一个修正因子。

在算法流程中，首先计算相关变量的梯度。如果λ不为0，则涉及梯度计算。之后，分别更新第一动量和第二动量，并修正偏差。在更新参数时，amsgrad 指的是是否使用 Adam 的变体，但该变体并不一定更好，通常不采用。在计算更新时，会在分母中添加 E，以提高数值稳定性。

二、网络训练过程

由于神经网络超参数的不确定性，实验中使用了大量的超参数组合。部分数据如表1所示。其中，AOV表示验证集准确率；AOT表示测试集准确率；Dropout_p表示Dropout概率。对于编号为n的数据项，我们将其称为数据项n或参数组n。其中，验证集最高准确率为93.9%，测试集最高准确率为82.86%，具有良好的通用性。

三、结果对比分析

获取实验数据后，观察不同参数对模型训练效果的实际影响，并总结其背后的规律[2]，并尝试解释其原因。

实验中，我们考察了5种不同学习率的具体表现。

展示了5个不同学习率（LR）值时loss下降的变化情况。从图中可以看出，在使用Adam优化器的情况下，采用特别大的学习率（0.1）会导致loss完全无法收敛；而采用特别小的学习率（0.00001）会导致loss下降得非常缓慢。

学习率决定了神经网络模型的训练速度，它的变化会导致模型准确率的显著差异。可以观察到，LR越小，上升越平缓，LR越大，上升越陡峭。

需要注意的是，当LR为0.1时，loss图像的趋势比较特殊[1]，它并没有像其他参数那样呈现整体下降的趋势，而是在接近某个特定值后就不再有太大的变化。然而，其验证集准确率甚至在80次迭代后都不再变化，这表明无论输入如何，模型的输出始终是固定的。图4展示了在不同LR值下得到的模型内部参数值。可以看出，当LR为0.1时，该模型的内部参数值相对于其他模型而言明显偏大。通常，出现这种情况可以认为是神经网络在训练过程中损失函数陷入了局部最优解。

虽然基于Adam优化器的学习率可以自适应，但如果一开始就设置过大的学习率，神经网络仍然可能陷入局部最优解。一般来说，采用较小的学习率可以避免这种情况的发生。

在训练模型的过程中，寻求神经网络的全局最优解并非易事，可能需要大量的经验和练习。然而，一些局部最优解实际上已经非常接近全局最优解，差别不大。然而，当学习率为0.1时，loss最终在2.0左右，这比其他学习率下的loss要大得多。因此，这个结果距离全局最优解仍然有相当大的距离。

当batch size相对较小时，也更容易出现过拟合。

当batch size为32时，模型验证集的准确率在迭代50次后不再提升。

Dropout作为一种正则化技术，可以抵抗过拟合。[4] 这些模型的参数中，只有epoch不同，其中LR为0.0001，batch size为128。从中可以看出，未引入Dropout时，模型快速完成拟合。然而，随着迭代次数的增加，验证集和测试集的准确率未能呈现明显的上升趋势。当Dropout概率设置为0.2时，可以看出模型具有一定的抗过拟合能力。经过超过50次迭代后，测试集的准确率不再随着验证集的提升而提高，说明模型仍然陷入过拟合。当Dropout概率设置为0.5时，虽然会导致模型训练速度变慢，但很好地抑制了过拟合，从而获得了更好的泛化能力。然而，当Dropout概率设置为0.8时[3]，模型训练速度过慢，说明在这种情况下0.8的概率显然过高。

该数据集的主要问题是高质量数据稀缺和样本过于单一。

尽管岩石分布比例较为不均匀，但在光照、湿度等不同环境条件下，岩石的分布差异较大，特定环境下的数据也较为稀缺。如果能够在数据增强之前直接扩充原始数据集，相比现有数据集，效果会更好。

数据集中可能存在标注错误的数据，导致数据集中存在较大的噪声干扰，需要专业人员重新校准数据集。这些因素都限制了准确率的进一步提升。

本实验中，数据增强方法相对较少，主要集中在裁剪和旋转方面[5]。因此，可以考虑使用其他图像生成技术，例如液化、模糊、锐化、光晕等，以创建数据集更丰富的样本。

卷积神经网络的卷积层可以实现图像的特征提取。

对于CNN来说，层数越多，拟合效果越好。但同时也容易导致过拟合，计算时间显著增加。因此，简化网络结构可以减少过拟合。具体做法是减少网络层数和神经元数量，以限制网络的拟合程度。

此外，还可以组合多个模型，取其中出现频率最高的预测作为输出。例如，可以先用一个模型对图像进行大类分类，再用其他模型对每个子类进行预测。

为了抑制过拟合，可以在网络训练中添加更多正则化方法，例如增加Dropout层、批量归一化层、使用L1/L2正则化等。此外，实验中还采用了提前停止策略，以便在适当的时机停止模型训练。合理地组合这些正则化方法也是优化的一个方向。

四、结论

基于深度学习的人工智能识别技术是当前的研究热点之一。本文研究了卷积神经网络在岩石样品图像识别中的应用，模型最佳测试准确率达到82.86%。实验中，采用了Dropout、批量归一化和提前停止等正则化技术，有效抑制了过拟合。本文还探讨了学习率、批量大小和Dropout概率对模型训练的影响，总结出一些规律，并对不同参数下的模型训练进行了分析。基于已有研究成果，提出了一些优化方向，包括提升数据集质量、优化网络结构、调整训练策略等。

参考文献：

[1] Kingma D P. 一种随机优化方法. arXiv: 1412.6980，2014.

[2] Timothy D. 将 Nesterov 动量融入 Adam. 自然灾害，2016，3(2)：437-453.

[3] Reddi SJ, Kale S, Kumar S. 关于 Adam 的收敛性及其改进. arXiv: 1904.09237，2019.

[4] Srivastava N, Hinton G, Krizhevsky. A. Dropout：一种防止神经网络过拟合的简单方法. 机器学习研究杂志，2014，15(1)：1929-1958.

[5] 李岩. 基于深度学习的岩石图像识别. 北京林业大学，2020.

[6] 程刚，郭伟. 基于深度卷积神经网络的岩石图像分类. 2017, 887(1)：012089.