李健,王晓明,张英海等. 2020. 基于深度卷积神经网络的地震震相拾取方法研究. 地球物理学报,63(4):1591-1606,doi:10.6038/cig2020N0057.

Li J, Wang X M, Zhang Y H, et al. 2020. Research on the seismic phase picking method based on the deep convolution neural network. *Chinese J. Geophys.* (in Chinese),63(4):1591-1606, doi:10.6038/cjg2020N0057.

基于深度卷积神经网络的地震震相拾取方法研究

李健1,2, 王晓明2, 张英海1, 王卫东1*, 商杰2, 盖磊2

- 1 北京邮电大学电子工程学院,北京 100876
- 2 禁核试北京国家数据中心和北京放射性核素试验室,北京 100085

摘要 地震震相拾取是地震数据自动处理的首要环节,包括了信号检测、到时估计和震相识别等过程,震相拾取的准确性直接影响到后续事件关联处理的性能,影响观测报告的质量.为了提高震相拾取的准确性,进而提高观测报告质量,本文采用深度卷积神经网络方法来解决震相拾取问题,构建了多任务卷积神经网络模型,设计了分类和回归的联合损失函数,定义了基于加权的分类损失函数,以三分量地震台站的波形数据作为输入,同时实现对震相的检测识别和到时的精确估计.利用美国南加州地震台网的 200 万条震相和噪声数据对模型进行训练、验证和测试,对于测试集中直达波 P、S 震相识别的查全率达到 98%以上,到时估计的标准偏差分别为 0.067 s,0.082 s.利用迁移学习和数据增强,将模型用于对我国东北地区台网的 6 个台站 13000 条数据的训练、验证和测试中,对该数据集P、S 震相查全率分别达到 91.21%、85.65%.基于迁移训练后的模型,设计了用于连续数据的震相拾取方法,利用连续的地震数据对该算法进行了实际应用测试,并与国家数据中心和中国地震局的观测报告进行比对,该方法的震相检测识别率平均可达 84.5%,验证了该方法在实际应用中的有效性.本文所提出的方法展示了深度神经网络在地震震相拾取中的优异性能,为地震震相和事件的检测识别提供了新的思路.

关键词 多任务卷积神经网络;震相拾取;联合损失函数;迁移学习

doi:10.6038/cjg2020N0057

中图分类号 P631, P315

收稿日期 2019-02-20, 2020-03-02 收修定稿

Research on the seismic phase picking method based on the deep convolution neural network

LI Jian^{1,2}, WANG XiaoMing², ZHANG YingHai¹, WANG WeiDong^{1*}, SHANG Jie², GAI Lei²

1 Information & Electronics Technology Lab, School of Electronic Engineering,
 Beijing University of Posts and Telecommunication, Beijing 100876, China
 2 CTBT Beijing National Data Center, Beijing 100085, China

Abstract Picking seismic phase is the primary step of automatic seismic data processing, including signal detection, arrival time estimation and phase identification. The accuracy of seismic phase picking affects the performance of subsequent phase association processing and the quality of event bulletin. Scholars have carried out a lot of research on this issue, while further improvement is needed. In this work, the deep convolution neural network method was used to solve the seismic phase picking problem. A multi-task convolution neural network model was constructed. The joint loss function of classification and regression was designed. The weighted classification loss function was defined. The method took the waveform data of 3-component

基金项目 国防装备预研项目(41425070403)资助.

第一作者简介 李健,男,1980年生,博士在读,副研究员,主要从事机器学习、深度学习及核查地震数据处理等. E-mail: lijian1980@bupt. edu. cn *通讯作者 王卫东,男,1967年生,教授,博导,主要从事卫星移动通信、物联网及空天地海一体化等. E-mail: wangweidong@bupt. edu. cn

seismic stations as input, and realized the detection and identification of the seismic phase and the estimation of phase onset. The model was trained, verified and tested using 2 million phase and noise data from the Southern California Seismic Network (SCSN) of American. For the test set, the recall rate of direct P and S phase identification of direct waves reached over 98%, and the standard deviations of arrival time estimation for P and S waves were 0.067 and 0.082, respectively. Using transfer learning and data augmentation methods, the model was applied to a small dataset of some stations in the seismic network of northeast China, and the recall rates of direct P and S phases identification reached 91.21% and 85.65%, respectively. Based on the model after transfer learning, we designed a phase picking method for continuous data. The method was tested in practical application using continuous real waveform data. The results were compared with the National Data Center bulletin and the China Earthquake Administration bulletin. The phase detection rate of this method reached 84.5%, and the effectiveness of the method was verified. The method proposed in this paper demonstrates the excellent performance of deep neural network in seismic phase picking and provides a new idea for the detection and identification of seismic phases and events.

Keywords Multi-task Convolutional neural network; Phase picking; Joint loss function; Transfer learning

0 引言

神经网络方法通过构建一个非线性的映射函 数,将输入数据映射为需要的输出数据,实现分类或 回归任务. 早在 20 世纪 90 年代,人工神经网络就普 遍用于解决各种分类和预测问题,由于早期的神经 网络均为全连接层,受限于当时的计算水平,很难构 建深层网络,实际中的应用效果不是很理想.随着云 计算、大数据时代的到来,计算能力的发展提高了训 练的效率,训练数据的大幅度增加则可降低过拟合风 险,因此以深度学习为代表的复杂模型开始受到人们 的关注. 2010 年以来,以卷积神经网络(Convolutional Neural Network, CNN)为代表的深度神经网络已 成为计算机视觉任务的主要机器学习方法,同时 也在感知任务、自然语言处理、医药发现、灾难气 候发现甚至围棋比赛程序中表现出了优异的性能 (Krizhevsky et al., 2012; LeCun et al., 2015; He et al., 2016; Simard et al., 2003; Taigman et al., 2014; Peters et al. ,2018).

卷积神经网络一般由多个卷积层、池化层和全连接层组成,映射函数的系数通过大量已标注的历史数据进行训练得到.卷积层不同于传统的全连接神经网络,每相邻两层之间只有部分节点相连,数据被输入到局部连接的卷积层和池化层,每个卷积层包含一系列可学习的滤波器,与前一层的输出进行卷积,以识别数据集中任何感兴趣的模式.池化层用

于对卷积层的输出进行下采样,缩小最后全连接层中节点个数,达到减少整个神经网络中参数的目的. 卷积神经网络前一部分可以认为是系统的特征提取部分,从输出数据中提取相关信息,然后将其输送给标准的全连接层.

地震数据自动处理一般包括检测、识别、关联和 定位等过程(靳平等,2014). 震相拾取(包括信号检 测、震相识别以及到时估计等部分)是首要的一步, 其准确性直接关系到最后观测报告的质量. 国内外 学者针对 P、S 震相的自动检测、到时估计和识别开 展了大量研究,提出了多种方法.反映信号能量瞬间 变化的短时能量与长时能量比(STA/LTA)方法 (Allen, 1978, 1982; Earle and Shearer, 1994; 刘晗和 张建中,2014)由于其实现简单、高效的特点,常被用 于实际检测系统中,但存在阈值不好确定等问题,不 利于微弱信号的检测;基于自回归赤池信息准则 (AR-AIC)的震相到时估算方法(Akaike, 1974; Maeda, 1985; Leonard, 2000), 将自回归模型得到的 信号、噪声序列的拟合误差作为 AIC 参数,通过寻 求 AIC 曲线极小点得到震相到时,常用于实际检测 系统,但对于初动不明显、信噪比低的信号,效果不 佳;基于偏度和峰度等高阶统计量的检测方法 (Küperkoch et al., 2010; Baillard et al., 2014; 田优 平等,2016)、基于极化分析的震相拾取方法(Ross and Ben-Zion, 2014)以及基于地震波整体包络特征 的震相检测方法(赵鸿儒等,1990; Moore et al., 2012; 靳平等, 2014)等, 利用优化的特征函数作为检

验统计量进一步改善检测和识别性能,但计算量较 大,且与人工分析结果相比依然存在一定差距.由于 从地震震源到接收台站,地震波形受到多种因素的 影响,使得台站记录到的地震信号十分复杂,传统方 法都是从地震震相某一个或几个特征出发实现震相 检测和识别,未能涵盖波形数据中包含的所有特征, 另外采用这些方法还需要对数据进行大量预处理工 作,提取所需要特征,仔细设定检测阈值,这些都对 检测结果影响较大,全面禁止核试验条约组织国际 数据中心(International Data Centre, IDC)实时处 理来自国际监测系统(Internet Monitoring System, IMS)网络的地震、水声、次声和核素数据,实现对全 球可能核试验的监测. 其中地震震相拾取采用 STA/LTA 方法实现信号检测、采用 AR-AIC 方法 实现震相到时估计、采用基于规则和传统神经网 络相结合的方法实现震相的识别(Le Bras and Wuster, 2002). 目前 IDC 自动处理结果存在误检 事件和漏检事件都较高的问题(Draelos et al., 2012),震相拾取的不准确是导致这一问题主要原 因之一.

早期的神经网络方法也应用于解决地震震相拾 取问题(Dai and MacBeth, 1995; Gentili and Michelini, 2006),但效果有限.深度神经网络的出现,为解决地 震震相拾取、事件检测等问题提供了新的思路和途 径,许多学者开展了这方面的研究. Perol 等(2018) 利用卷积神经网络对俄克拉荷马地区地震事件检测 和定位,该方法检测出比原有目录中更多的事件. Ross 等(2018)利用卷积神经网络实现震相和噪声 识别,模型具有较高敏感性和良好的稳健性. Zhu 和 Beroza(2019)采用 U-net 实现 P、S 震相和噪声检 测,利用输出的概率分布估计到时. Wu 等(2018)设 计了一种级联卷积神经网络可检测不同尺度的地震 事件. Titos 等(2018)设计了深度卷积网络以识别火 山地震事件. 赵明等(2019a,2019b)将卷积神经网络 和U型神经网络用于汶川余震事件和首都圈地震 台网的检测识别等.

本文针对地震震相的拾取问题,包括震相识别 (P、S、N类)和到时估计,构建了多任务卷积神经网络模型,设计了分类和回归的联合损失函数,定义了基于加权的分类损失函数,以三分量波形数据作为输入,实现对震相的检测识别和到时的精确估计,利用美国南加州地震台网(SCSN)的 200 万条震相和噪声数据对模型进行训练、验证和测试.以我国东北地区几个台站震相拾取为例,利用少量台站数据对

上述模型进行迁移训练,在测试集上取得较好效果.并利用迁移后的模型,设计了用于连续数据的震相拾取方法,利用连续的真实波形数据对该算法进行了实际应用测试,并与国家数据中心观测目录和中国地震局的观测目录进行了比对,验证了该方法的实际应用效果.

1 数据和方法

1.1 数据集及预处理

本文所用的数据由 Ross 等(2018)进行了预处理,可从美国南加州地震数据中心获取(http://scedc. caltech. edu/research-tools/deeplearning. html).数据包括美国南加州地震台网 2000—2017 年的200 万条近震波形,观测仪器包括 692 个宽频带和短周期台站.数据为三分量事件波形,震相主要是直达 P波、S波,噪声数据为检测信号前 5 s 数据.数据采样率为100 Hz,以震相到时为中心,4 s 窗口长度截取数据.数据预处理包括去除线性趋势、0.1~20 Hz带通滤波,以每段数据最大值对数据进行归一化处理.

本研究还从中国地震局地球物理研究所获取了 我国东北地区包括牡丹江、延边、长白山、宽甸、沈阳 和长春等6个台站2013—2018年的震相数据和原 始波形数据.所用的震相数据来自1938个近震地震 事件,台站和事件位置如图1所示,事件震级分布如 图2所示.根据震相到时进行数据截取,对截取数据 进行预处理,包括去线性、滤波和归一化处理,以用 于后续的迁移学习.

1.2 卷积神经网络设计

本文主要解决震相的拾取问题,包括震相检测识别和到时估计,而这些任务都将直接通过三分量台站数据直接学习训练完成.此问题与图像识别中的目标检测和定位相似,首先检测识别是否存在某类震相,然后再给出震相的起始时间,可以通过多任务卷积神经网络来实现.输入采用三分量台站数据,输出分为两个:一个用于震相分类,实现震相的检测和识别(P、S、N),采用分类的方法;一个用于预测到时,采用回归方法.因为震相类别和其到时之间具有相关性,建立一个卷积神经网络模型,可以学到丰富而又准确的表示,同时用来识别震相和预测到时.多任务神经网络结构如图 3 所示.

一般卷积神经网络可分为两部分,首先是一系

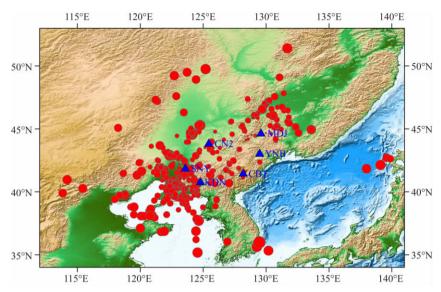


图 1 本文研究中所选取的东北地区 6 个台站及观测报告事件的位置分布蓝色三角形代表选取的台站;红色圆形代表选取的事件,圆形大小反映了震级大小.

Fig. 1 Map showing distribution of 6 stations and bulletin events in Northeast China selected in this paper

The blue triangles represent the selected stations; The red circles represent the selected events, and the size of the circle reflects the magnitude of the event.

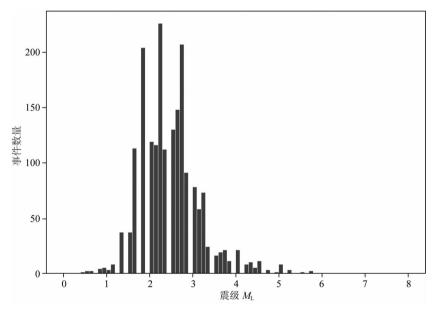


图 2 本文研究所选取的东北地区事件的震级分布

Fig. 2 Magnitude distribution of events in Northeast China selected in this paper

列卷积层和池化层,又被称为是卷积基,实现对数据的特征提取,也是多任务网络可以共享的部分,即共享隐藏层;然后是全连接层,可以根据输出任务的不同分别连接分类器和回归器. 我们构建的多任务模型前半部分由 5 个卷积层构成,每个卷积层都利用BN(batch normalization)实现正则化以防止过拟合,利用 ReLU 非线性函数作为激活函数,每个卷积层之间都采用池化层,对特征进行下采样,用于压

缩数据和参数的量,减少过拟合. 卷积层将从波形数据中自动学习并提取特征,因此不需要特征工程和更复杂的预处理. 输出层,即全连接层,分为两个分支:一个是由 softmax 构成的分类全连接层,实现对震相检测和识别(P、S、N),一个是由线性激活构成的回归全连接层,实现到时估计. 最终设计的模型形状如图 4 所示.

图中括号内容代表了每层输入和输出形状大

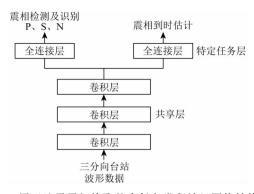


图 3 用于地震震相拾取的多任务卷积神经网络结构图 Fig. 3 Multi-task convolution neural network structure for seismic phase picking

小,由于处理的是一维的时间序列数据,输入形状为(None,400,3),即 400个时间序列数据,3 代表 3 个通道,第一层卷积网络采用 32 个卷积核,每个卷积核过滤器尺寸为 21,则输出形状变为(None,400,32),池化层(maxPooling)的尺寸为 2,经过下采样,输出大小变为(None,200,32).经过 5 个卷积层和池化层后的输出为(None,12,256),再经过 Flatten层,则输出矩阵大小为 12×256=3072.后续网络结构以震相识别分支为例,连接了两个全连接层,第一个全连接层输出个数为 256,最后一个全连接层输出个数为 3,采用 softmax 作为激活函数输出分类的概率.

Perol 等(2018)和 Ross 等(2018)分别采用卷积神经网络实现地震事件的检测和震相的分类,主要利用的是卷积神经网络的分类功能,且每个神经网络模型只实现单个任务.而本文采用了多任务神经网络结构,分别利用分类和回归功能同时实现震相的检测识别和到时的估计,各任务之间可以共享所学习到的信息,提高模型的泛化能力.

1.3 损失函数定义

损失函数用来度量预测值和真实值之间的差异,卷积神经网络根据训练任务的不同,采用不同的损失函数.分类、回归任务中常采用交叉熵损失函数和均方误差损失函数实现模型的训练,这些损失函数常独立用于单个任务的训练过程中.本文采用的多任务卷积神经网络模型,具有分类和回归两个分支,希望对两个任务同时进行训练,梯度下降算法对同一个标量进行优化,为此采用了联合损失函数,如公式(1):

$$L(t_{pi}, p_{pi}, t_{at}, p_{at}) = L_{cls}(t_{pi}, p_{pi}) + \lambda I\{t_{pi} = 1\}$$
 $\times L_{res}(t_{at}, p_{at}),$ (1)

公式中 L_{cls} 是分类损失函数, L_{reg} 是回归损失函数, t_{pi} 、 t_{ai} 分别是模型训练时的目标震相名称和到时值, p_{pi} 、 p_{ai} 是模型训练时的预测震相名称和到时值. $I_{A}(t_{pi})$ 是指示函数,其定义为

$$I_{A}(t_{pi}) = \begin{cases} 1 & t_{pi} = 0 \ \text{\'et} \ t_{pi} = 1 \\ 0 & t_{pi} = 2 \end{cases}$$
 (2)

表示只有当震相识别为 P 或 S 时,回归损失才起作用. λ 是权重参数,以控制分类和回归损失在联合损失函数中的比例. t_{pi} , p_{pi} 分别代表震相识别的真实值和预测值. 该损失函数实现对分类和回归任务的联合优化,训练过程中,两个任务计算的梯度进行累加,权重值 λ 用于平衡两个任务的比重,避免梯度过大导致的网络收敛不稳定. 根据采用的两种损失函数的取值范围,这里的权重设为 λ =0.4. 公式(1)中的分类损失 L_{cls} 定义为加权的 softmax 交叉熵损失函数,为不同震相的识别结果分配不同的权重,以体现不同的震相分类结果对损失函数的影响不同,达到最佳的分类性能,如公式(3).

$$L_{cls}(t_{pi}, p_{pi}) = -\alpha I(t_{pi} = 0) \log \left(e^{p_0} / \sum_{p_i = 0}^{2} e^{p_{pi}} \right) - \beta I(t_{pi} = 1) \log \left(e^{p_1} / \sum_{p_i = 0}^{2} e^{p_{pi}} \right) - (1 - \alpha - \beta) I(t_{pi} = 2) \log \left(e^{p_2} / \sum_{p_i = 1}^{2} e^{p_{pi}} \right),$$
(3)

其中 t_{pi} 、 p_{pi} 分别为目标及预测的震相名称,I(x)为指示函数,当 x 为真时,I(x)=1; $e^{p_x}/\sum_{pi=0}^2 e^{p_{pi}}$ 表示模型预测震相为 x 的概率; α , β 为小于 1 的权重参数,通过对不同的 α , β 值进行尝试,以模型训练的准确性作为标准,选择合适的权重参数值,最终确定为 $\alpha=0.4$, $\beta=0.4$.

公式(1)中的回归损失函数 Lreg 定义为均方误

差损失函数(MSE),如公式(4)所示,其中 $t_{\alpha i}$, $p_{\alpha i}$ 分别是模型训练时目标震相到时和预测震相到时,n 为一次训练一批数据(batch)的个数,MSE 是求一个 batch 的平均误差的函数:

$$MSE(t_{at}, p_{at}) = \frac{\sum_{i=1}^{n} (t_{at} - p_{at})^{2}}{n}.$$
 (4)

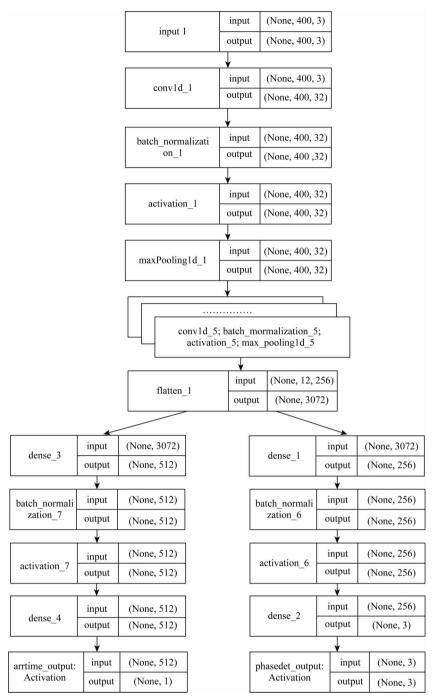


图 4 带有形状信息的多任务卷积神经网络模型图

Fig. 4 Multi-task convolution neural network model with shape information

在多任务模型训练中,采用加权的联合损失函数可以避免某个任务主导了整个损失,其余的任务没有机会影响共享层的学习过程的问题.同时,由于震相检测识别用于不同的应用场景,如用于微弱事件的监测或用于地震事件预警等,不同场景对震相的查全率和查准率有不同的侧重,对各震相的检测能力有不同侧重,而本文采用的多分类交叉熵损失函数为不同的震相识别结果分配不同权重,可实现

这一目的.

1.4 模型训练和测试

模型训练采用 Adam 随机优化算法(Kingma and Ba,2015),采用默认学习率 0.001,数据批量设置为 480 个,数据批量控制了每次迭代采用多少数据量.采用 1 块 NVIDIA TITANV 图形处理单元进行处理,学习过程的终止设定为验证集的损失值在先前 5 轮训练中不再下降,选择在验证集上具有

最低损失值的参数作为模型的最终参数.

选取了 200 万条三分量地震台站记录用于模型训练、验证和测试,其中 150 万用于训练和验证(训练集 80%,验证集 20%),50 万用于测试.在 150 万训练和验证集中包括 50 万条 P 波地震波形,50 万条 S 波地震波形,50 万条 S 波地震波形,50 万条噪声数据,选择数量基本相当的每类数据进行训练,确保训练不偏向任何一类.

震相识别的类别分为 3 类:类别 0 是 P 震相、类别 1 是 S 震相、类别 2 是噪声. 到时估计采用回归方法实现,选取的 P、S 震相数据是以其人工拾取的到时为窗口中心截取的数据,选择人工拾取的实际到时对应的秒数作为回归拟合值,对部分训练数据到时在[-0.5 s,0.5 s]范围内加上随机高斯噪声值,以模仿到时拾取中的误差.

2 结果

图 5 是模型训练时联合损失值随着迭代轮数下降情况,可以看到第 9 轮验证集的总体损失值达到最小值 0.20,此时验证集震相识别准确性达到 98.25%.在 1 块NVIDIA TITANV GPU 卡上训练模型,训练用时约 30 min.

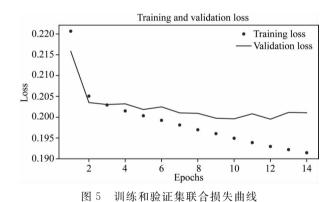


Fig. 5 Joint loss curve of training set and verification set

50 万条测试数据中包括约 16.6 万条 P 波波形、16.6 万条 S 波波形和 16.8 万条噪声.图 6 展示了部分震相的检测识别情况,其中 real:x[y]代表真实值,x代表震相类型,y代表到时位置,对于噪声数据,设置 y=0.相应的 pre:x[y]为模型的预测值.可以看到所有的震相类型都预测正确.到时估计预测大部分较为准确,子图 6j 到时预测误差较大.

为了进一步分析模型的性能,我们计算了测试结果的查准率、查全率指标.首先定义了混淆矩阵,如表1所示.

表 1 混淆矩阵定义

Table 1 Definition of confusion matrix

邓加古	真实值		
预测值	P	S	N
P	$TP_{ ext{PP}}$	ϵ_{SP}	$\epsilon_{ m NP}$
S	ϵ_{PS}	TP_{SS}	$\epsilon_{ m NS}$
N	ϵ_{PN}	ϵ_{SN}	$TP_{ m NN}$

P 震相的查准率 Prep 和查全率 Recp 分别定义为

$$Pre_{P} = \frac{TP_{PP}}{TP_{PP} + \epsilon_{SP} + \epsilon_{NP}} = 98.48\%,$$

$$Rec_{P} = \frac{TP_{PP}}{TP_{PP} + \epsilon_{PS} + \epsilon_{PN}} = 98.10\%,$$
(5)

同理,S 震相的查准率 Pre_s 和查全率 Rec_s 分别定义为

$$Pre_{s} = \frac{TP_{ss}}{TP_{ss} + \epsilon_{PS} + \epsilon_{NS}} = 98.38\%,$$

$$Rec_{s} = \frac{TP_{ss}}{TP_{ss} + \epsilon_{SP} + \epsilon_{SN}} = 98.40\%.$$
(6)

将本文中方法的测试结果与全面禁止核试验条约国际数据中心(IDC)系统中的震相识别结果进行比对,比较了两种方法对于 P类(Pg),S类(Lg)震相的查准率和查全率的性能,结果见表 2,可见本文中方法在查全率和查准率方面远超过 IDC,由于IDC 震相识别的类型较多、划分较细,即使我们放宽IDC 震相匹配标准,即 Pg 震相被识别为 Pn 都认为识别正确,Lg 震相被识别为 Sn 也认为识别正确,其 P、S 震相的查全率可达 84.01%、75.22%,但也低于本文中方法的性能.

表 2 本文方法与 IDC 方法震相识别的查全率、查准率比较 Table 2 Comparison of recall and precision of phase identification of our method and IDC method

	本文方法		IDC 方法	
	P震相	S震相	P 震相(Pg)	S 震相(Lg)
Precision/%	98.48	98.38	60.67	53.97
Recall/%	98.10	98.40	30.05	34.94

在到时估计方面,统计分析了到时估计值与真实值之间误差,P 震相的到时误差的均值和标准差分别为:-0.017 s,0.067 s,S 震相到时误差均值和标准差为:-0.024 s,0.082 s,如图 7 所示,可以看出 P、S 震相模型预测的到时整体上偏小于真实到时,即预测到时要偏早些.

将本文方法与 IDC 目前方法进行了比较,统计了 IDC一年的观测报告数据 Pg、Lg震相的到时误

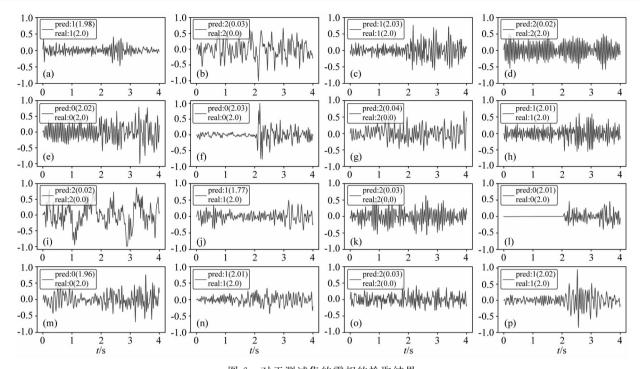


图 6 对于测试集的震相的拾取结果 Fig. 6 Results of phase picking for test set

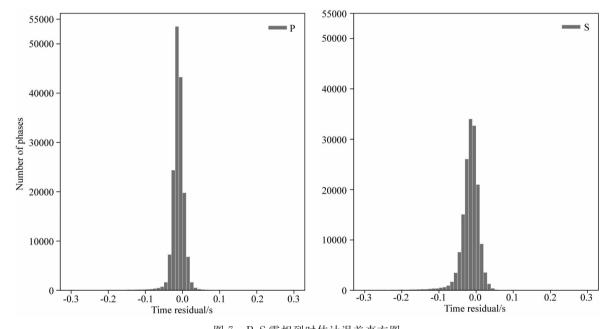


图 7 P、S 震相到时估计误差直方图 Fig. 7 P、S phase arrival time estimation error histogram

差情况,比较了到时误差的均值、标准差、50%的分位数和75%的分位数,具体情况见表3. IDC的人工交互分析是在自动处理结果的基础上进行,其结果会向自动处理结果倾斜,从统计分析结果中可以看到 IDC50%的震相到时估计结果与人工分析的结果一致,也说明了这一点,但同时也表明 IDC 目前采用的 AR-AIC 到时估计方法的有效性. 但对于经过人工校正的到时,其调整范围变化较大,统计分析结

果表现为 IDC 到时误差的均值、标准差以及 75%分位数都远大于本文方法. 总体上,本文方法的到时估计结果更稳定,具有较小的均值和标准差.

3 应用

3.1 迁移学习

要利用上百万条大型数据集进行模型训练一般

表 3 本文方法与 IDC 方法到时误差的比较
Table 3 Comparison of phase arrival time error
of our method and IDC method

	本文方法/s		IDC 方法/s	
	P 震相	S 震相	P 震相(Pg)	S 震相(Lg)
均值	-0.017	-0.024	0.263	0.303
标准差	0.067	0.082	0.607	0.798
50%分位数	0.012	0.016	0	0
75%分位数	0.019	0.028	0.670	1.070

不容易做到.对于地震震相拾取问题,不同台站间相同震相具有许多相似特征,模型之间应该具有一定通用型,但不同地区台网由于地质结构不同,台站记录到的震相又有各自的一些特点.我们采用迁移学习和数据增强方法,将上文中用 SCSN 区域地震网络训练得到的模型有效应用于我国东北区域部分台站小数据集的训练、验证和测试中,实现震相的可靠拾取.

迁移学习是一种将深度学习应用于小型数据集的非常高效的方法,它将一个在大数据集训练好的模型通过简单调整使其适用于新的任务(Tan et al.,2018).这种模型学到的特征在不同问题之间的可移植性也是深度学习与传统浅层学习方法相比的重要优势,使得深度学习对处理小数据问题非常有效.本文的迁移学习采用的是模型迁移,是对训练好的模型进行微调应用到目标领域,由于源领域与目标领域都是近震 P、S 震相数据,关联度比较高,我们迁移了模型的整个卷积层,仅对其全连接层的参数进行重新训练.按照这一思路我们保存上文中训

练好的模型的卷积层,冻结其参数,即在训练过程中 这些层的权重保持不变,在其顶部添加新的全连接 层,并在输入数据上端到端地运行整个模型,训练全 连接层参数.

为了防止训练样本过少引起的过拟合,我们进一步采用了数据增强措施,增强模型的泛化能力.数据增强是对现有的训练样本进行多种随机变换,以生成更多的训练数据.我们这里采用的数据增强方式包括平移和加噪.平移是类似之前为了进行到时检测对数据在时间窗口[一0.5 s,0.5 s]范围内移动;加噪是对波形每个采样点进行高斯随机扰动.同时为模型添加 Dropout 层,以进一步降低过拟合现象.在卷积神经网络的某层应用Dropout,可以将该层输出特征成比例的随机舍弃,减少模型参数,进而让卷积神经网络模型更加健壮.

目前禁核试北京国家数据中心(National Data Centre, NDC)仅获取了东北地区 6 个台站数据,而这些台站记录的近震事件数量有限,为了将利用美国南加州地震台网数据训练好的模型应用于这 6 个台站,我们采用了迁移学习方法.选取了该 6 个台站2013 年 1 月至 2018 年 1 月的 6500 条震相数据,其中 P 震相 2200 条、S 震相 2200 条、噪声数据 2100条(台站和事件信息在 2.1 节中进行了介绍),图 8 展示了这些数据的信噪比分布情况,采用 P 类震相到时前后 5 s 数据的均方根值分别作为噪声值和信号值,计算得到信噪比.可以看到样本中包含了大量的较低信噪比的信号,可以检验该方法对于小事件的震相拾取能力.

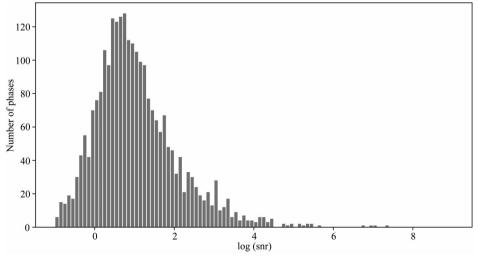


图 8 选取的东北地区台站数据信噪比分布情况

Fig. 8 Signal/noise ratio distribution of selected stations in northeast China

利用数据增强方法(平移和加噪),共得到13000个样本,并将这些样本顺序随机打乱,用其作为训练、验证和测试集.利用其中10000个样本对迁移模型的全连接层进行训练和验证(训练集80%,验证集20%),3000个样本对模型进行测试.将模型训练终止条件设定为验证集的损失值在先前6轮训练中不再下降.到第54轮时,验证集的总体损失值达到最小0.30,训练用时小于1 min.此时,测试集P、S震相识别的查全率分别为91.21%和85.65%,查准率分别为93.01%、98.91%,同样优于目前IDC方法的检测结果,具有不错的检测效果.图9展示了部分震相的拾取情况,可以看到除了子图(h)中的P震相被识别为噪声,其余震相识别都是正确的.

到时估计方面,P 震相的均值和标准差分别为:

-0.117 s, 0.264 s, S 震相的均值和标准差分别为: -0.077 s, 0.169 s, 相比于之前的结果,性能下降较大. 图 10 分别展示了 P、S 到时估计误差的直方图.

为了验证迁移学习的有效性,利用南加州数据训练得到的模型对上述3000个测试样本进行测试,得到的P、S 震相识别的查全率分别为88.45%、55.15%,查准率分别为69.76%、97.66%.可以看出,对东北地区台站的数据,利用经过迁移学习的模型进行震相拾取,结果的各项指标都有所改善.

3.2 连续数据应用测试

上述模型实现了对截取好的 4s 窗口的数据的 震相拾取,但在地震监测和核试验监测中,需要对连 续数据进行实时处理,得到震相和事件信息,为此, 我们设计了如下方法将迁移后的模型用于连续数据

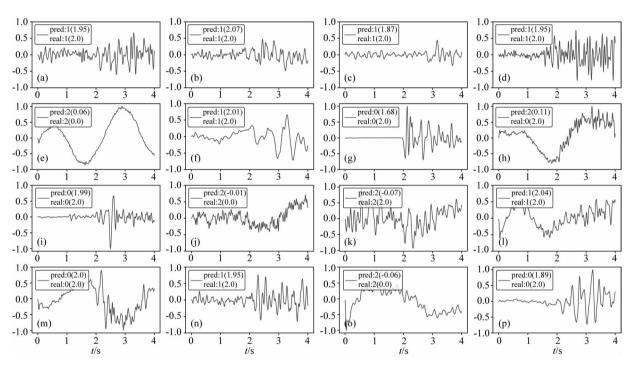


图 9 采用迁移学习后的模型得到的震相拾取结果

Fig. 9 The phase picking results obtained by using transfer learning model

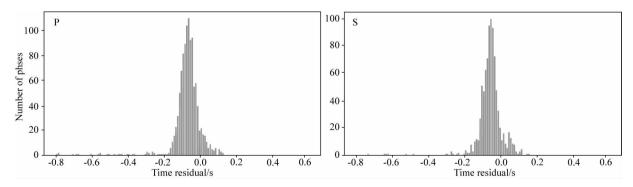


图 10 采用迁移学习后的模型得到的 P、S 震相到时估计误差直方图

Fig. 10 Error histograms of P and S phase arrival time estimation obtained by using transfer learning model

的检测处理中. 具体步骤如下:

- (1) 以 4 s 为窗口,0.1 s 为步长,滑动截取连续数据为若干待检测窗口;
- (2) 用卷积神经网络对每个窗口进行检测处理,得到震相识别概率和相应的到时拟合值;
- (3) 当识别为 P 或 S 概率超过 98%时,则识别为 P 或 S 震相,当识别概率都低于 98%时,则识别为噪声:
- (4) 为了降低误检率,设定识别为某类震相的 概率需要持续一定的时间,如 10 个窗口时间(即 1 s),同时在形成检测后 4 s 时间内不再形成新的检测.
- (5)由于后续尾波信号影响,一般识别为某个震相的概率会持续一段时间,在此时间窗口内确定到时估计值有一定难度.在概率持续时间的窗口内的到时拟合值为一系列离散值,我们的方法是,求这些离散值的极值,则满足到时拟合值大小要求(如大于 1.3)的首个极值即为此震相的到时估计值.

NDC 从中国地震局地球物理研究所获取了东北台网 6 个台站的数据,数据量有限,本文重点尝试如何将深度学习方法有效应用于小数据集台网的近震震相的拾取中.对于连续数据检测方法的实际测试,重点验证算法能否准确检测识别出真实事件的震相.

首先选取了 2018 年 8 月至 2018 年 12 月间 NDC 观测报告中 CBT 台站记录到的 9 个近震事件,为了直观反映检测效果,提高测试效率,截取出这 9 个事件的震相数据(每段波形从震相到时前 20 s 开始截取,持续 1 min),并将其合并成为一段连续波形数据,用于连续数据的测试.采用上述的检测方法对数据进行了处理,准确检测和识别出了全部的 9 个事件的 P、S 震相,并给出了较为精确的到时信息.如图 11 所示.

然后,为测试多个台站联合的震相拾取情况,选择了在 NDC 上述时间段观测报告中,有多个台站记录到近震事件的两天的数据进行处理,分别是2018 年 8 月 2 日和 8 月 5 日.8 月 2 日记录的 2 个事件分别发生在20180802 10:10:53 (GMT)和20180802 10:46:35 (GMT)两个时间段,两个事件分别被 CBT、KDN 和 YNB 台站记录到.图 12 展示了8月2日10—11 点之间3个台站的数据处理情况,可以看到 CBT、KDN 和 YNB 三个台站均检测识别出了这两个事件的 P、S 震相,并给出了到时信息,根据三个台站到时和震相信息可实现震相的关联和事件的定位.同时本文方法检测出了一些真实

的震相,但未关联形成事件,如 KDN 台站黄色框标 出的震相检测.同时,可以看到各个台站也存在许多 误检信号,经分析误检信号多数为噪声信号.

2018 年 8 月 5 日的事件时间为 2018-08-05 11:23:21(GMT),该事件被 CBT、YNB、KDN 和 MDJ 记录到.图 13 展示了 11:00—13:00 点之间 4 个台站的数据处理情况,可以看到在 CBT、YNB 和 MDJ 三个台站准确检测并识别出了该事件的 P、S 震相,在 YNB 台站准确检测识别了 S 震相,但将 P 震相误识别为 S 震相.同时 4 个台站都记录到了一个远震事件的 P 震相,但由于本文模型由近震震相数据训练得到,主要用于近震震相的检测识别,因此在 CBT、YNB 和 KDN 未形成检测,而 MDJ 台站在远震震相之前形成的 P 震相检测经分析是由该震相之前的一段高频噪声信号所引起,也不是由远震 程 M 所触发.

最后,为进一步测试算法在实际连续数据中的应用效果,又另外选取 2012 年 6 月至 12 月包含有这 6 个台站的东北地区事件的观测报告数据对该方法进行了测试,测试数据包括 138 个事件的 335 个震相(Pg、Sg). 利用本文算法对原始波形数据进行处理,其结果与观测报告中的震相进行比对,以测试算法的震相拾取性能. 震相匹配原则是台站、震相名称一致,到时差小于 4 s. 比对结果见表 4,震相的总体检测比例为 84.5%. 图 14 展示了匹配震相的到时误差分布情况. 同时,对于观测报告中的 1054 个P、S远震震相,本文方法未能检测出任何一个,再次说明本文方法仅适用于近震震相的检测识别. 实际应用测试中同样存在将较多的噪声信号检测识别为震相的情况,需要在后续研究中对噪声信号加以识别和剔除.

表 4 本文算法处理结果与 2012 年下半年观测报告 震相数据比对情况

Table 4 Comparison between the results of the algorithm proposed in this paper and the phase data of the second half of 2012 in CEA bulletin

	Pg 震相	Sg 震相	总计
震相数量	154	181	335
本文检测结果	127	156	283
检测比例	82.5%	86.2%	84.5%

5 讨论

通过本文的研究,可以看到深度卷积神经网络

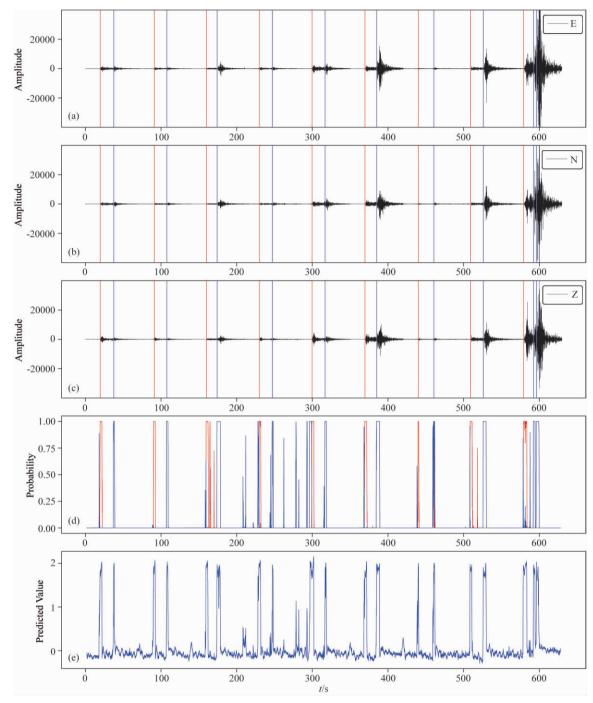


图 11 对合成的 CBT 台站的连续波形数据的震相拾取

(a) E 通道数据; (b) N 通道数据; (c) Z 通道数据; (d) 震相识别的概率曲线,红色曲线表示 P 震相识别概率, 蓝色曲线表示 S 震相识别概率; (e) 震相的到时估计曲线.

Fig. 11 Phase picking of synthetic CBT station continuous seismic waveforms

(a) Data of channel E; (b) Data of channel N; (c) Data of channel Z; (d) Phase identification probability curve. Red curve represents P phase identification probability; (e) Phase arrival time estimation curve.

在地震震相拾取方面具有很好的应用前景,相比于基于规则的和传统神经网络的识别方法,省去了特征工程这一复杂过程,取得了较好的效果.由于震相的检测识别和其到时之间存在很强的相关性,采用多任务卷积神经网络,可同时实现震相的检测识别

和到时的精确估计,一次完成震相拾取任务,提高了实时数据处理的效率.

地震监测由于不同地区地质结构差异,不同台 网记录到的近震震相和事件在具有一些相同的通用 特征外,又具有各自不同的特点,这是迁移学习应用

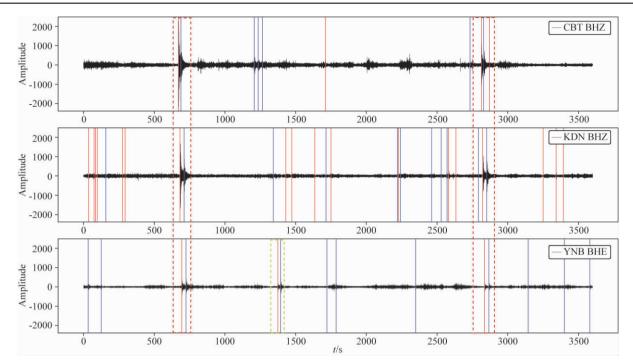


图 12 2018 年 8 月 2 日 10:00-11:00 CBT、KDN 和 YNB 台站数据处理情况

图中包含了 NDC 观测报告中的两个事件,发震时间分别为 2018-08-02 10:10:53 和 2018-08-02 10:46:35.事件在台站记录的 震相用红色虚线框标出,黄色虚线框标出了在 YNB 台站检测到的真实震相,但在 NDC 观测报告中未关联形成事件.

Fig. 12 Phase picking of stations CBT, KDN and YNB at 10:00—11:00, August 2, 2018

There are two real events occurred during this period in NDC bulletin. The origin times of events are 2018-08-02 10:10:53 and 2018-08-02 10:46:35, respectively. The phases of events recoded at stations are marked by red dashed boxes. The yellow dashed box marks the true phases detected at the YNB station, but are not associated with an event in NDC bulletin.

的很好的场景,本文将采用美国南加州地震台网数据集训练得到的卷积神经网络用于我国东部地区几个台站数据的训练和测试,取得较好的检测结果,验证了迁移学习在解决此类问题的可行性.

同时,迁移学习和数据增强方法,是实现将大规模数据集训练得到的网络有效应用于小规模数据集的有效方法,从测试结果来看,虽然不能达到大数据集完全从头训练的性能,但所需的训练时间和训练样本要远远小于训练完整模型所需要的.

在对直达波 P、S 震相检测识别性能方面,无论是针对大数据集训练的网络,还是迁移后的网络,都优于 IDC 目前的方法;在到时估计方面,迁移后模型的到时预测性能有所下降,整体上比实际到时值早些,相比于 AR-AIC 方法的优势不太明显,有待进一步研究改善.

对于连续波形数据的检测,利用训练好的模型,采用我们设计的连续数据检测方法可以较准确的检测出近震事件 P、S 震相,再根据多个台站检测的震相及到时信息,通过聚类关联,可以得到初步事件信息. 从实际的连续数据测试中可以发现,该方法在较准确检测出近震震相的同时,也存在一定的误检信

号,这些信号多数为噪声信号,同时该方法可以屏蔽对远震震相的检测.

对于用于监测全球核试验的地理上稀疏分布的 IMS 网络,其地震台站监测到的震相除了 Pg、Lg 震相外,还包括区域震震相(Pn、Sn等),远震震相(P、S、PKP等)以及噪声信号. 因此,要将该方法用于 IMS 网络中事件的检测,还需要进一步开展研究工作,以实现对噪声信号和其他震相信号的检测识别能力.

6 结论

本文针对地震震相拾取问题,建立了多任务 卷积神经网络模型,定义了基于加权的多分类交 叉熵损失函数,设计了分类和回归的联合损失函 数,利用 SCSN 大规模数据集对网络模型进行了训 练、验证和测试,同时实现对震相的检测识别和到 时的精确估算.接着又采用迁移学习和数据增强等 方法,将该模型应用于我国东北地区6个台站小型 数据集训练和测试中,取得较好效果.本文获得如下 主要认识:

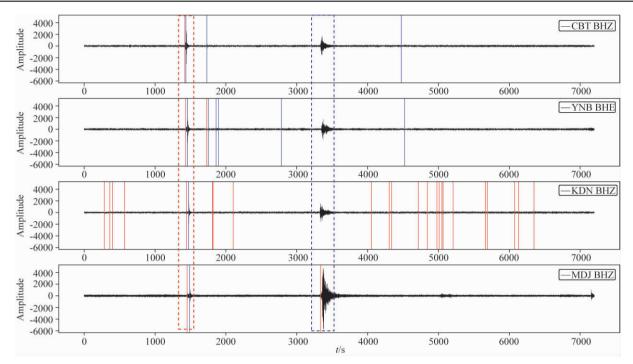


图 13 2018 年 8 月 5 日 11:00—13:00 CBT、YNB、KDN 和 MDJ 台站数据处理情况 图中包含了 NDC 观测报告中的一个事件,发震时间为 20180805 11:23:21. 事件在台站记录的震相用红色虚线框标出, 蓝色虚线框标出了台站记录的远震信号.

Fig. 13 Phase picking of stations CBT, YNB, KDN and MDJ at 11:00—13:00, August 5, 2018

There is one real event occurred during this period in NDC bulletin. The origin time of event is 20180805 11:23:21. The phases of events recorded at stations are marked with a red dashed box. The blue dashed box marks the a teleseismic phase recorded by stations.

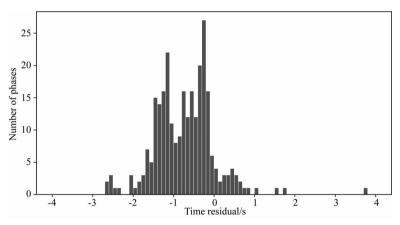


图 14 匹配震相的到时误差分布

Fig. 14 Distribution of arrival time estimation error of the matched phases

- (1)利用多任务卷积神经网络发展的地震震相 拾取方法,可有效实现对直达波 P、S 震相的拾取, 可将 P、S 震相识别的查全率提高至 98%以上,将 P、S 到时估计标准差减小至 0.067 s、0.082 s;
- (2)利用迁移学习和数据增强的方法,将利用某个区域台网大型数据集训练得到的模型有效应用到目标台网小型数据集的训练和检测中,测试结果优于 IDC 传统的震相拾取方法;
 - (3)利用迁移后的模型发展的连续波形检测方

法,可准确拾取出连续波形数据中的近震事件 P、S 震相,进而实现对事件的可靠检测.

致谢 采用的南加利福尼亚地震台网数据来自网址: http://scedc.caltech.edu/research-tools/deeplearning.html. 我国东北区域台站震相数据来自中国地震局地球物理研究所.文章中程序及图采用 Python3.6和 GMT(General Mapping Tool)4.0编制,卷积神经网络采用 Tensorflow训练.感谢审稿专家提出的修改意见!

References

- Akaike H. 1974. A new look at the statistical model identification.

 IEEE Transactions on Automatic Control, 19(6): 716-723,
 doi: 10.1109/TAC.1974.1100705.
- Allen R. 1982. Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B). S225-S242.
- Allen R V. 1978. Automatic earthquake recognition and timing from single traces. Bulletin of the Seismological Society of America, 68 (5): 1521-1532.
- Baillard C, Crawford W C, Ballu V, et al. 2014. An automatic kurtosis-based P- and S-phase picker designed for local seismic networks. Bulletin of the Seismological Society of America, 104(1): 394-409, doi: 10.1785/0120120347.
- Dai H, MacBeth C. 1995. Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophysical journal international*, 120(3): 758-774, doi: 10.1111/j.1365-246X.1995.tb01851.x.
- Draelos T J, Ballard S, Young C J, et al. 2012. Refinement and testing of the probabilistic event detection, association, and location algorithm. // Proceedings of the 2012 Monitoring Research Review:

 Ground-Based Nuclear Explosion Monitoring Technologies. New Mexico: National Nuclear Security Administration, 221-231.
- Earle P S, Shearer P M. 1994. Characterization of global seismograms using an automatic-picking algorithm. Bulletin of the Seismological Society of America, 84(2): 366-376.
- Gentili S, Michelini A. 2006. Automatic picking of P and S phases using a neural tree. *Journal of Seismology*, 10(1): 39-63, doi: 10.1007/s10950-006-2296-6.
- He KM, Zhang XY, Ren SQ, et al. 2016. Deep residual learning for image recognition. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 770-778, doi: 10.1109/CVPR.2016.90.
- Jin P, Zhang C L, Shen X F, et al. 2014. A novel technique for automatic seismic data processing using both integral and local features of seismograms. Acta Seismologica Sinica (in Chinese), 36 (3): 464-479, doi: 10.3969/j.issn.0253-3782.2014.03.012.
- Küperkoch L, Meier T, Lee J, et al. 2010. Automated determination of P-phase arrival times at regional and local distances using higher order statistics. Geophysical Journal International, 181(2): 1159-1170, doi: 10.1111/j.1365-246X.2010.04570.x.
- Kingma D, Ba J. 2015. Adam: A method for stochAstic optimmization.
 In: The International Conference on Learning Representations (ICLR).
- Krizhevsky A, Sutskever I, Hinton G E. 2012. ImageNet classification with deep convolutional neural networks. // Proceedings of the 25th International Conference on Neural Information Processing Systems, 1097-1105, doi: 10.1145/3065386.

- Le Bras R, Wuster J. 2002. IDC Processing of Seismic, Hydroacoustic and Infrasonic Data, Revision 1. IDC Documentation User Guides.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature*, 521 (7553); 436-444, doi: 10.1038/nature14539.
- Leonard M. 2000. Comparison of manual and automatic onset time picking. Bulletin of the Seismological Society of America, 90 (6): 1384-1390, doi: 10.1785/0120000026.
- Liu H, Zhang J Z. 2014. STA/LTA algorithm analysis and improvement of Microseismic signal automatic detection. *Progress in Geophysics* (in Chinese), 29(4): 1708–1714, doi: 10.6038/pg20140429.
- Maeda N. 1985. A method for reading and checking phase times in auto processing system of seismic wave data. *Zisin*, 38(3): 365-379, doi: 10.4294/zisin1948.38.3_365.
- Moore D A, Mayeda K M, Myers S M, et al. 2012. Processing in Signal-Based Bayesian Monitoring. In Proc. Monitoring Research Review (MRR 2012), Albuquerque, New Mexio.
- Perol T, Gharbi M, Denolle M. 2018. Convolutional neural network for earthquake detection and location. Science Advances, 4(2): e1700578, doi: 10.1126/sciadv.1700578.
- Ross Z E, Ben-Zion Y. 2014. Automatic picking of direct *P*, *S* seismic phases and fault zone head waves. *Geophysical Journal International*, 199(1): 368-381, doi: 10.1093/gji/ggu267.
- Ross Z E, Meier M A, Hauksson E, et al. 2018. Generalized seismic phase detection with deep learning. Bulletin of the Seismological Society of America, 108(5A): 2894-2901, doi: 10.1785/0120180080.
- Simard P Y, Steinkraus D, Platt J C. 2003. Best practices for convolutional neural networks applied to visual document analysis. // Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. Edinburgh, UK, UK: IEEE, doi: 10.1109/ICDAR.2003.1227801.
- Taigman Y, Yang M, Ranzato MA, et al. 2014. Deepface: Closing the gap to human-level performance in face verification. // The IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 1701-1708, doi: 10.1109/CVPR.2014.220.
- Tan C Q, Sun F C, Kong T, et al. 2018. A survey on deep transfer learning. // Kůrková V, Manolopoulos Y, Hammer B, et al eds. International Conference on Artificial Neural Networks. Cham: Springer, 270-279, doi: 10.1007/978-3-030-01424-7 27.
- Tian Y P, Zhao A H. 2016. Automatic identification of P-phase based on wavelet packet and Kurtosis-AIC method. Acta Seismologica Sinica (in Chinese), 38(1): 71-85, doi: 10.11939/jass. 2016. 01.007
- Titos M. Bueno A. García L. et al. 2018. A deep neural networks approach to automatic recognition systems for volcano-seismic events. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(5): 1533-1544, doi: 10. 1109/JSTARS. 2018. 2803198.
- Wu Y, Lin Y Z, Zhou Z, et al. 2018. DeepDetect: A cascaded region-based densely connected network for seismic event detection. IEEE Transactions on Geoscience and Remote Sensing, 57(1): 62-75, doi: 10.1109/TGRS.2018.2852302.

- Zhao H R, Sun J Z, Tang W B. 1990. The application of full wave phase analysis. *Acta Geophysica Sinica* (in Chinese), 33(1): 54-63.
- Zhao M, Chen S, Yuen D. 2019a. Waveform classification and seismic recognition by convolution neural network. *Chinese Journal of Geophysics* (in Chinese), 62(1): 374-382, doi: 10. 6038/cjg2019M0151.
- Zhao M, Chen S, Fang L H, et al. 2019b. Earthquake phase arrival auto-picking based on U-shaped convolutional neural network. *Chinese Journal of Geophysics* (in Chinese), 62(8): 3034-3042, doi: 10.6038/cjg2019M0495.
- Zhu W Q, Beroza G C. 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1): 261-273, doi: 10.1093/gji/ggy423.

附中文参考文献

靳平,张诚鎏,沈旭峰等. 2014. 基于信号整体与局部特征的地震

- 数据自动处理方法研究. 地震学报, 36(3): 464-479, doi: 10. 3969/j. issn. 0253-3782. 2014. 03. 012.
- 刘晗, 张建中. 2014. 微震信号自动检测的 STA/LTA 算法及其改进分析. 地球物理学进展, 29(4): 1708-1714, doi: 10.6038/pg20140429.
- 田优平, 赵爱华. 2016. 基于小波包和峰度赤池信息量准则的 P 波 震相自动识别方法. 地震学报, 38(1): 71-85, doi: 10.11939/ jass. 2016.01.007.
- 赵鸿儒, 孙进忠, 唐文榜. 1990. 全波震相分析的应用. 地球物理学报, 33(1): 54-63.
- 赵明, 陈石, Yuen D. 2019a. 基于深度学习卷积神经网络的地震波形自动分类与识别. 地球物理学报,62(1):374-382, doi:10.6038/cjg2019M0151.
- 赵明, 陈石, 房立华等. 2019b. 基于 U 形卷积神经网络的震相识别与到时拾取方法研究. 地球物理学报, 62(8): 3034-3042, doi: 10.6038/cjg2019M0495.

(本文编辑 汪海英)