**摘要**

随着地震勘探技术的不断发展，逐渐成为油气勘探的重要手段。同时，三维高分辨率地震勘探的需求也不断浮现，使得地震数据量飞速增长。数据量的爆炸性增长对数据存储，数据处理和数据传输三个方面带来了新的挑战。这使得对地震数据压缩变成了急需解决的问题。同时，数据量的增长使得噪声压制的难度上升。传统的噪声压制方法需要人工设定参数进行多次实验以找到最佳参数组合，数据量的上升使得人工操作变得十分艰难。所以寻找一种不需人工操作的地震数据压缩和噪声压制方法一直是地震数据处理领域的研究热点。本文利用中国海洋大学地球探测软件实验室的机群，针对地震数据，利用神经网络中的自编码器算法,设计了卷积自编码器和U-net自编码器进行了数据压缩和去噪实验，取得了良好的效果，而且得出了对地震数据压缩和去噪有指导意义的结论与建议。

**关键词：地震勘探；人工神经网络；卷积神经网络；自编码器；数据压缩；噪声压制**

**ABSTRACT**

With the development of seismic exploration，it has gradually become an important method of geophysical exploration. Meanwhile, the demand for 3-dimensional high-resolution seismic exploration has also emerged, making the size of seismic data increase rapidly. The explosive growth of data brought new challenges to data storage, data processing and data transmission, which makes the compression of seismic data an urgent problem. At the same time, the increase of the amount of data makes noise suppression more difficult. The traditional denoising method need scientists to set the parameters manually and perform experiments to find the optimal parameter combination. The manual operation is becoming more difficult. Therefore, finding a compression and denoising algorithm that does not require manual operation has always been a research hotspot in the field of seismic data processing. The thesis design a convolutional autoencoder and U-net autoenocder, based on autoencoder algorithm in artificial neural network, and experiment on these autoencoders, with the help of the computer cluster of the Earth Exploration Software Laboratory, Ocean University of China. The data compression and denoising experiment not only carried out good results but also gave instructive conclusion and recommendations for the compression and denoising of seismic data.

Key words: Seismic exploration; Artificial neural network; Convolutional neural network; Autoencoder; Data compression; Denoising

目录

1. 绪论
   1. 地震数据压缩与去噪研究背景和现状

一、地震数据压缩

地震勘探是目前使用范围最广而且最有效的地球物理勘探手段。使用地震勘探可以高效地探测地下岩层结构，再加上近年来发展的地震属性分析，可以有效地探查地下油气储存情况。为了提高地震勘探的精度与分辨率，地球物理学界发展了超万道地震仪，来实施两宽一高（宽频宽方位高密度）和多波多分量勘探，这必然使得地震数据量的爆炸增长，数据容量达到TB级别。海量的地震数据给数据传输，处理和储存提出了严峻的挑战，主要问题是需要极大的储存空间增加了传输带宽的需求和传输时间。若不对数据进行压缩处理的话，在地震数据处理和解释过程中会产生极大的不便。所以，对地震数据的压缩是一项十分有意义的研究。

目前数据压缩方法按照有无信息丢失可以分为无损压缩和有损压缩两大类。无损压缩，顾名思义，能保证解压数据和原始数据完全一样，但其压缩效率低，一般压缩倍数在两倍左右。有损压缩即是在允许有少量信息损失的情况下可以获得高倍压缩。有损压缩的压缩倍数较高，但存在信息丢失。

在过去几十年的时间里，地球物理学家们提出了各种地震数据压缩方法。1985年，日本东京大学地震研究所的笠原顺三等利用ADPCM（Adaptive Differential Pulse Code Modulation）编码方法[1]，实现了数据压缩。1986年，郭洪升，陈俊良提出了基于ADPCM方法的自适应频率压缩方法[2]，根据地震信号的特点，对数据进行了自适应频率压缩，压缩效果显著，波形基本无失真，且方法简单，计算量小，易于汇编语言实现。1992年，陈志文提出了应用分形维进行地震数据压缩和恢复[3],通过应用地震数据的自相似特征，给定度量尺度范围内的值，应用其他相同尺度范围内数据所表征的图形与给定度量尺度范围数据所表征图形相似比例因子，就可以很好地重建或恢复地震道数据。1994年，陈明伟，李世雄提出了在I. Daubechies提出的紧支集正交小波对地震道数据进行压缩及重建[4]；其次在KLT（卡努里变换）的原理上采用小波包对压缩后的数据再压缩。1994年，王嘉松，贾国平提出了基于分形几何理论的压缩方法[5]，利用Peano曲线的特性，用Peano曲线扫描实质上是将图像的高频成分转化成低频成分，从而使得扫描重拍的波形条件变好，得到良好的压缩效果。2004年，王培茂，薛亚雪等提出利用离散余弦变换和游程编码技术相结合进行提振数据体的压缩[6]，其效果与压缩数据体的松散程度，信噪比有一定关系，应用了相邻波形间的相似性。2005年，孙寰宇，武文波等在实验基础上，针对地震信号的特性，基于JPEG2000标准提出了新的地震数据混合编码方案[7]，不仅取得了较大的压缩率，而且大大提高了网络传输、实时处理的性能。2008年，耿瑜，吴帮玉等提出了基于curvelet变换的地震数据压缩[8]，重建数据和原始数据保持了较好的相似性，实现了良好压缩比和低能量损失下的数据压缩。2015年，徐锋涛，张正炳等提出了基于小波变化的零树编码与算数编码相结合的地震数据压缩算法[9]，结果表明，与单一使用基于小波变换的零树编码算法相比，该方法所得地震数据压缩倍数较高，数据解压效果较好。

随着计算机科学技术的发展，人工智能AI技术越来越普及，越来越容易实现，使得人工智能技术在各个行业展现出强大的力量，其中在地震数据处理也不例外，应用于去噪，直达波拾取，反演和插值等方面。Yaniv Hollander, Amin Merouane等人提出了利用卷积神经网络来提高初至波拾取的准确性[10]。Yue Ma1, Xu Ji等人实现了卷积神经网络的自动速度拾取[11]。Sara Mandelli, Federico Borra等人提出了利用卷积自编码器进行地震数据插值的方法[12]。但是学界上还没有利用神经网络进行地震数据压缩的工作，尤其是利用自编码器进行压缩，这一方法具有十分美好的前景。在自编码器的帮助下，地震数据可以高效压缩，速度快，效果好。将神经网络应用到地震数据的压缩是本文的主要目的之一。

二、噪声压制

地震数据去噪是地震数据处理领域的主要问题之一。我们接收到的地震道集其实是混杂着随机噪声和有效信号的波场记录，包含了十分丰富的信息。地震数据去噪的主要目的是：将波场中的噪声与有效信号分离，尽量保留来自地下的原始信息。数据中的噪声分为两种：规则噪声和随机噪声。随机噪声主要是由于环境对检波器的扰动造成，具有不规则性，随机分布在地震记录中。规则噪声则是在地震记录中可以看出有规则的时距曲线的噪声，通常会对有效信号进行很严重的压制，如面波，多次波等。为了提高地震成像的质量，需要对地震记录的噪声进行最大程度的压制，获取高信噪比的地震资料。

地震噪声压制一直是全世界学者的研究热点，在过去几十年，许多不同的方法被应用于地震信号的噪声压制上，取得了令人满意的结果。Hampson 提出了抛物线型 Radon 变换，应用用于消除 CMP道集中的多次波[13]。2004 年Ethan J N 等人提出了加权Radon 变换，并实验证明了该算法能在有效消除多次波的同时，可以很好的保持处理后的信号的振幅[14]。2004 年，Hermann F J 将 Curvelet 变换用于多次波的压制上[15]，取得了较好的效果。2008年Neelamani 等人将曲波变换应用于地震相干噪声和随机噪声的压制上[16]。张华等人将二维曲波变换运用于处理三维地震资料[17]，取得良好的去噪效果。Xu 等提出了基于尺度相关性的小波空域相关去噪算法[18]。

传统的地震噪声消除方法需要人为设定参数，需要不断反复的试验以选出噪声压制效果最好的参数，过程十分繁琐。如果可以有机器自动学习地震记录中噪声的特征，并将之与有效信号区分开来，从而实现对噪声进行识别压制，将大大方便地球物理学家们的工作。这时候就需借助人工智能的帮助。

随着机器学习技术和计算机硬件的飞速发展，运用人工智能进行地震噪声压制成为可能。目前学界中许多科学家都在尝试将计算机科学中图像处理算法应用到地震噪声的压制上，如神经网络算法，字典学习等。其中卷积神经网络是最为流行的算法之一。Yuchen Jin, Xuqing Wu等人提出了利用深残余网络进行地震数据去噪，取得了良好的效果[19]。Zhang et al在2017年提出了一种基于残差学习的全卷积去噪网络[20]，能够同时去除多种噪声。Peigen Xie等人提出了将生成对抗网络应用到陆地地震资料的噪声压制上[21]。Xu Si等人基于深度卷积神经网络残差学习的随机噪声衰减[22]。

目前，学界上应用自编码器算法进行噪声压制的案例还比较少，多数是单纯的神经网络。运用自编码器算法进行噪声压制，不仅可以压制噪声，同时由于自编码器是一个数据收缩-扩张的过程，所以还会有压缩的效果。运用自编码器进行地震噪声压制是本文的主要内容之一。

* 1. 本文研究内容

通过上述介绍，由于真实地质环境的复杂，再加上近年来地震数据量爆炸性增长，地震数据的压缩和噪声压制变得非常之重要并且十分具有挑战性。前文已介绍，学界已提出了很多数据压缩和去噪方法。但由于人工智能是一门新兴学科，人工智能在这两方面的应用还不多。人工智能，特别是神经网络中的自编码器算法，在地震资料处理领域具有十分光明的前景。

本文基于地球探测软件技术实验室的计算机群，利用python的keras神经网络库搭建全连接自编码器和卷积自编码器及其改进——U-net编码-解码结构网络，对地震数据进行压缩和去噪。

本文的内容进行如下安排：

第一章为绪论，重点讨论了地震数据压缩与噪声压制领域的研究发展和现状。

第二章为相关基础知识。本章重点叙述了神经网络的基本概念，介绍了常见的神经网络模型，及神经网络的延伸——自编码器，并对本文实验中用到的几个自编码器结构进行了详细的论述。

第三章介绍了本文用来对地震数据进行压缩和去噪的两种自编码器——卷积自编码器和U-net自编码器的结构。

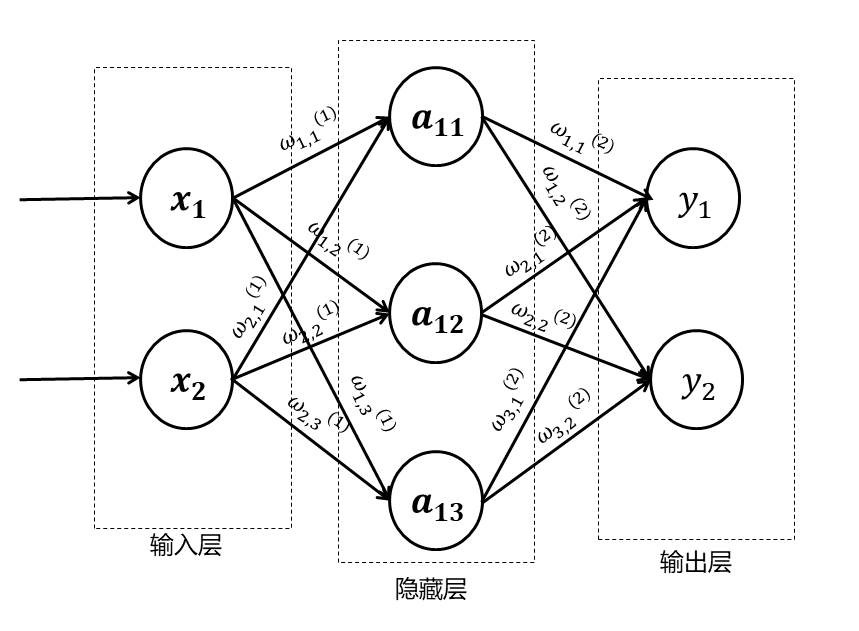
第四章是针对三种不同结构的自编码器在地震数据压缩和去噪的性能分析，包括自编码器的训练效率和压缩性能，以及去噪性能如信噪比等技术指标。

第五章是对文章所有内容的总结，并提出了算法存在的问题和对自编码器技术在地震资料处理的前景展望。

1. 相关基础知识

2.1 神经网络

人工神经网络（Artificial Neural Network）是对人脑中由神经元细胞组成的神经元网络的数学抽象，心理学家Warren McCulloch与数学家Walter Pitts于1943年提出了神经网络的原型模型。神经网络通过学习大量的数据来调整网络中各层所拥有的权值，从而达到获得一个效果理想的信息处理模型。根据输入数据的不同，神经网络可以分为两大类：监督学习和无监督学习。监督学习指的是用已知特性的数据样本作为网络的训练集，通过学习训练集的特征以建立一个数学模型，此模型可用于预测未知样本。有监督学习通常用于判别模型和图像分割等领域。而无监督学习指的是输入训练集没有被标记，需要网络自身学习数据的特性或数据之间的相似性，常用于数据降维等方面。下图为简单的神经网络基本结构。



注：为第层第个神经元与层神经元之间的权值，为输入值，为网络输出值

图1 简单神经网络示意图

神经网络通过前向传播和反向误差传播来进行计算和网络优化。数据在隐藏层中向前传递，称为前向传播。数据在隐藏层中的计算公式为：

用矩阵表示则为：

类似，输出层的计算公式可以表示为：

隐藏层的输出结果需要通过激活函数的规范，作为隐藏层的最终输出结果。

激活函数作为神经元的最终输出，使用不同的激活函数，往往会导致输出结果的不一，同时对网络收敛的影响也十分重大。所以在建立网络时需要选取适合的激活函数。常见的激活函数主要有：

（1）Sigmoid函数

Sigmoid函数是常用的激活函数，是一个非线性函数，数学表达式如下：

Sigmoid函数的图像如下：

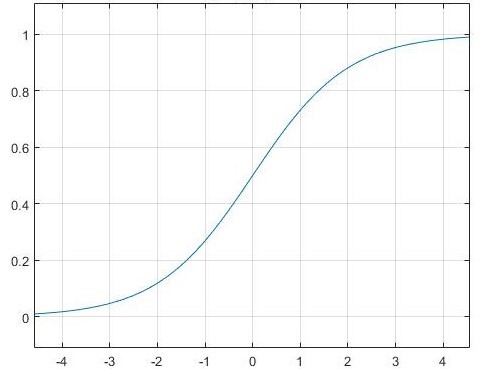


图2 Sigmoid函数图像

该函数能把神经元实际输出的值规范为0-1之间的连续输出。

（2） tanh函数

tanh函数的数学表达式为：

几何图像如下：

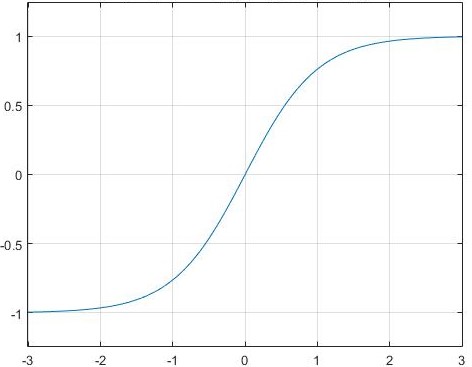


图3 tanh函数图像

tanh函数将输出值规范到-1~1之间。

（3） Relu函数

Relu函数，又称修正线性单元（Rectified Linear Unit, Relu），是神经网络中最为常见的激活函数，函数的数学表达式为：

或者

图像为：

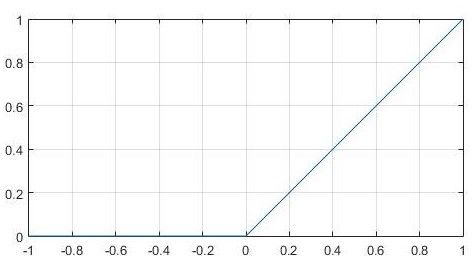


图4 Relu函数图像

函数实际为求最大值函数。使用Relu函数作为激活函数时网络的收敛会比Sigmoid函数快。

（4）Leaky Relu函数

Leak Relu函数是Relu函数的一个变种，数学表达式为：

图像为：

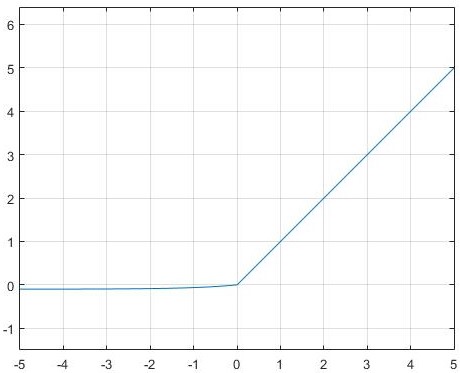


图5 Leaky Relu函数图像

与Relu函数不同的是，当输入值为负时，Leaky Relu函数的梯度为一个常数λ∈（0,1）。

神经网络的输出结果不可能与原始数据完全一样。根据网络的输出值与原始数据之间的差值，将误差逐层向后传递，修改每一层神经元的权值参数，不断反复这一过程，直到输出值与原始值的误差达到相当小的范围，即可视为训练完成。此过程称为反向传播算法（Back Propagation）。具体过程为：

（1） 将原始数据输入神经网络中的输入层。

（2） 对数据输入到输入层进行前向传播计算，并作为下一层神经元的输入值，直至数据流到达输出层输出结果。

（3） 计算输出值与原始数据的误差，将误差与预先设定的阈值对比，若误差大于阈值，则进行下一步，反之，则视为训练结束。

（4） 将误差进行反向传播，对每一层神经元的权值进行修正，直至到输入层为止。

（5） 不断重复步骤（2）～（4），直到达到设定的迭代最大次数为止。

经过几十年的发展，计算机科学家们发展出了许多不同结构不同功能的神经网络。常见的神经网络有如下两种：

（1） 全连接神经网络

全连接神经网络（Fully connected neural network），是人工神经网络中最简单的一种，是目前用途最广泛的神经网络模型之一。前馈神经网络中是一个单向多层结构，每层包含若干个神经元，且层内神经元不相互连通而与相邻层的神经元相互连接。各层神经元接收上一层神经元的输出，运算后把结果输出到下层神经元，直至到输出层输出最终结构。

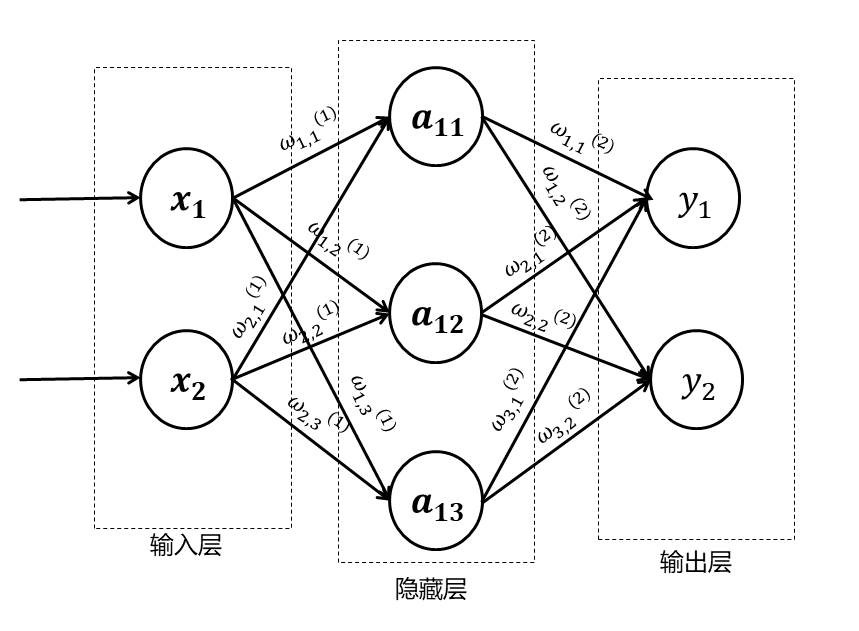


图6 简单全连接神经网络结构图

（2） 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）是一种包含卷积运算的神经网络。上个世纪60年代，Hubel和Wiesel等人通过研究脑皮层神经细胞发现有可以有效降低神经网络复杂程度的神经元，并在此基础上提出了今天的卷积神经网络。卷积神经网络主要用于图像处理，在图像分类，自然语言处理等方面取得很好的处理效果。

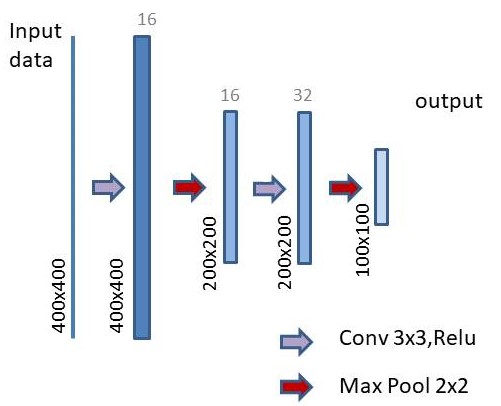


图7 简单卷积神经网络结构图，其中紫色箭头为卷积层，红色箭头为池化层

卷积神经网络主要包含两部分：卷积层与池化层。卷积层中包含着有若干个数值不一，大小相同的矩阵，称为卷积核，其用途为在输入的数据矩阵中作为一个窗口，平移滑动得到窗口内数据的加权平均，最终提取出原始数据的特征。通过卷积层提取原始数据特征的这个过程称为特征映射（feature map）。所以，卷积核也被称为神经元的感受野。

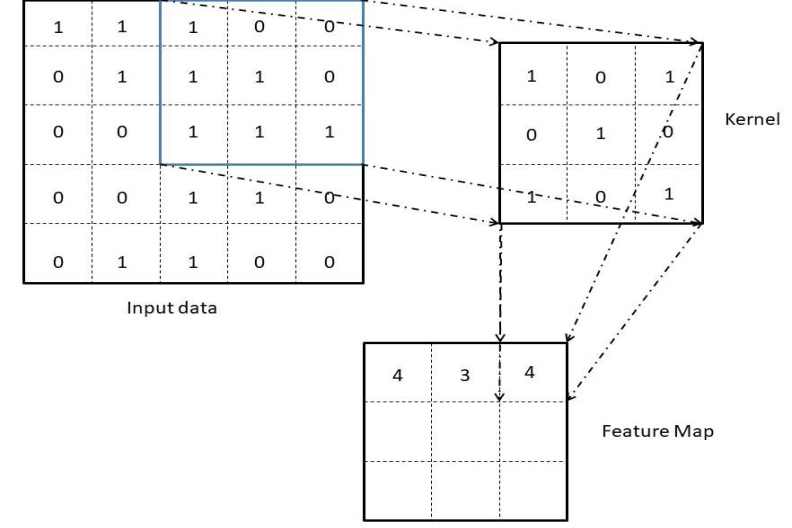


图8 卷积过程示意图，卷积核与滑动窗口的元素对应相乘后相加，即为卷积的结果

卷积核通常不需要人为设定初始值，卷积核的初始化可通过随机矩阵来设定。

卷积核提取特征过后，需减少特征向量的维度和再进一步提取数据特征，这个过程要借助池化（pooling）来解决。常用的池化方法有：最大池化（Max Pooling），平均池化（Mean Pooling）等。

最大池化：最大池化指的是按照一定大小的窗口在特征矩阵上平移滑动，取特征矩阵内各个相同尺度的局部的最大值，组成新的特征矩阵。如图9。

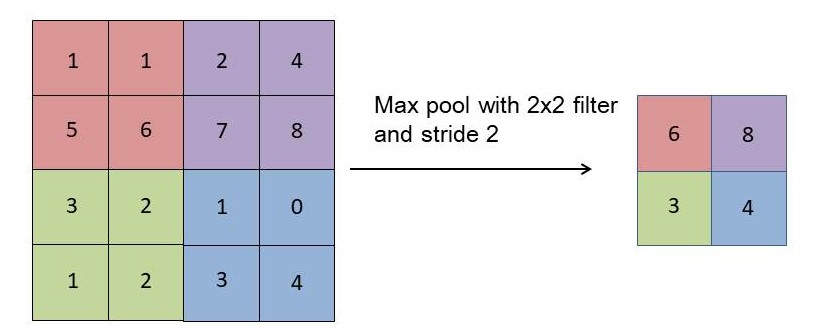


图9 最大池化示意图

平均池化：平均池化指的是按照一定大小的窗口在特征矩阵上平移滑动，取窗口内元素的平均值，组成新的特征矩阵,如图10。

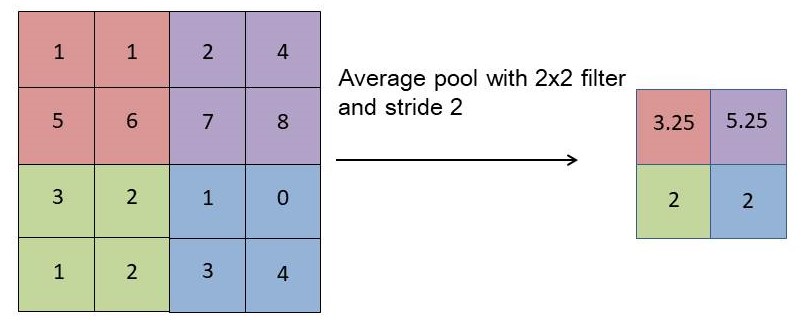


图10 平均池化示意图

2.2自编码器

自编码器（Auto-Encoder, AE）是新兴的无监督深度学习算法，是一种基于神经网络的数据压缩算法，最先由Rumelhart等[23]提出。自编码器的基本结构与普通神经网络一致，为输入层、隐藏层、输出层。

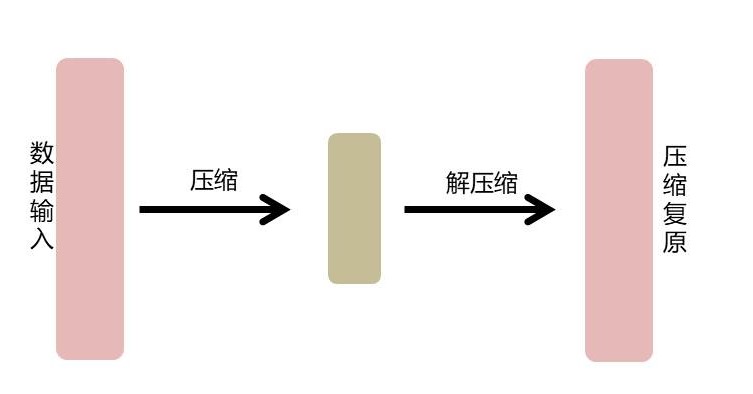


图11 自编码器简单示意图

主要流程可以简单概括为编码阶段和解码阶段：

编码阶段：使用编码器实现数据压缩。将原始数据输入到编码结构，经过向前传播提取到高层次特征，实现高维数据压缩为低维数据；

对于输入数据，经过编码器后得到编码，实现数据压缩，其中为激活函数：

解码阶段：使用解码器实现解压缩。将编码器输出的压缩结果继续向前传播到编码结构，实现低维数据恢复为原始高维数据。

对于编码，经过解码器后映射到高维空间，从而对编码进行解编码获得。其中为激活函数。

自编码器输出结果不可能与原始数据完全一样。根据网络的输出值与原始数据之间的差值，将误差逐层向后传递，修改每一层神经元的权值参数，直到输出值与原始值的误差达到相当小的范围，即可视为训练完成。要对解压缩后的数据与原始数据进行比较时，需要一个指标来衡量二者之间的误差。常用的误差函数有两种：平方误差函数（MSE）和交叉熵函数（cross-entropy）。

1. 平方误差函数

其中为样本个数，为输入网络的原始数据，为网络输出的预测值。

1. 交叉熵函数

近年来，自编码器不断发展，应用范围也不仅仅局限于数据降维，而被拓宽到图像分割，去噪等等其他领域，取得了优良的效果，是一种具有美好前景的神经网络算法。常见自编码器种类有如下几种：

（1） 全连接自编码器

神经元层为全连接神经网络层的自编码器称为全连接自编码器，是一种最为传统的自编码器。

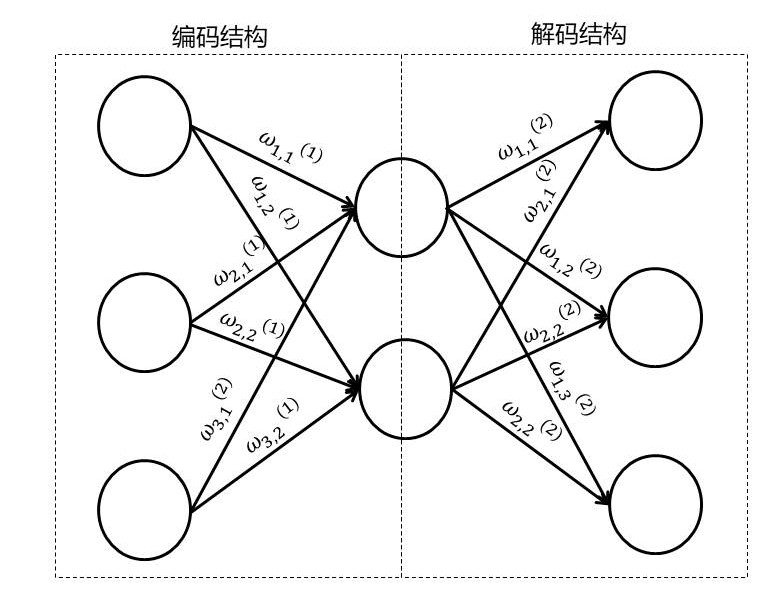


图12 全连接自编码器示意图

其优点为结构简单，直观，搭建网络容易。但参数数量过多，训练速度慢。

（2） 卷积自编码器

神经网络层为卷积神经网络层的自编码器称为卷积自编码器，利用传统自编码器的无监督学习，结合卷积神经网络的卷积、池化和上采样等操作，从而实现特征提取和数据解压缩。

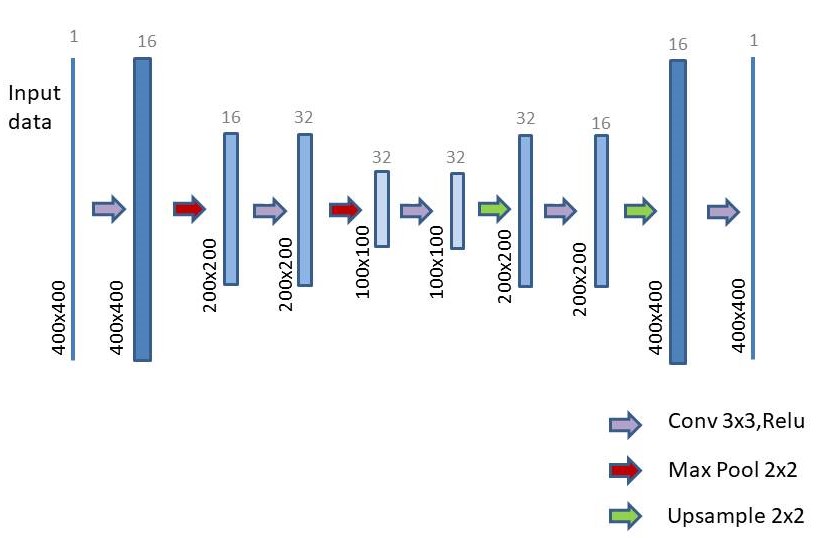


图13 卷积自编码器示意图

对传统的卷积自编码器的结构优化，可以产生新的卷积自编码器。其中U-net编码-解码结构就是一种十分成功的变式卷积自编码器。

U-net编码-解码结构（以下简称U-net结构）在2015年首先被Olaf Ronneberger等人[24]提出，在医疗影像分割领域中取得了良好的效果，近年来发展到了图像生成和自然图像处理等领域。因论文中的结构示意图像英语字母“U”而取名为U-net。

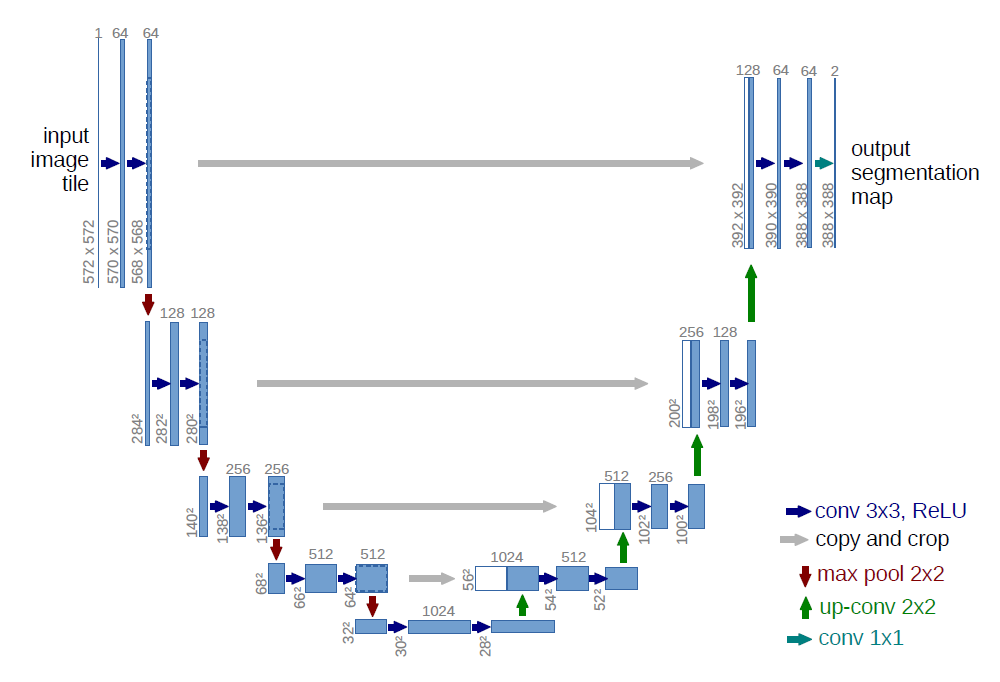


图14 Olaf Ronneberger等人提出的U-net结构

U-net结构与传统的卷积自编码器相比，多了一个特征融合的操作----Concatenate（连接）。Concatenate的作用是将卷积自编码器中维度相同的特征图链接起来，相当于打破了上采样过程和降采样过程相互隔离的情况，为两个阶段中相互对应的卷积层提供了反向误差传播的路径，更有利于网络的训练。同时，长连接可以联合浅层结构提取的高分辨率特征和深层结构提取的低分辨率特征来对图像进行分割，恢复等处理，更好地完成网络的任务。

1. 自编码器设计

自编码器的搭建可以选择任意类型的神经网络层，如全连接层，卷积层等，既可以是单一种类的神经网络层，也可以是多种神经网络层的组合，搭建者根据神经网络所需完成的任务选取合适的网络层来搭建网络。如图像分类领域中常常用卷积神经网络层来提取图像对象的特征，然后在卷积层后接全连接层来进行分类学习。但不同的单一神经网络中常常会有其优劣性。全连接神经网络层结构简单，常用于声音处理和分类学习，但其在图像处理方面缺乏理论上的直观性和效率，一般不直接用全连接神经网络进行图像处理，而是作为辅助来完成网络的任务。而卷积神经网络是模仿人脑视觉相关的神经细胞工作原理而建立的，因而卷积网络天生就是用来进行图像处理的。从二十世纪九十年代开始，卷积神经网络就开始蓬勃发展，如今发展诸如U-net，VGG-16等优秀卷积神经网络模型，广泛应用于自然语言处理，图像分割等方面，具有强大的生命力。

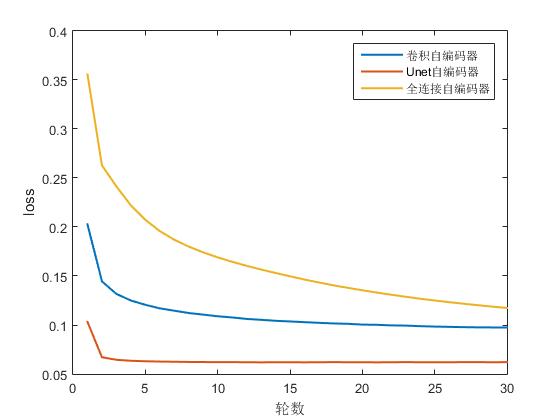
从广义上来讲，地震记录也是一副图像，是包含着地下反射界面信息的图画。因此，本文选择卷积神经网络来搭建卷积自编码器和U-net自编码器来进行地震数据压缩和去噪实验。本文不使用全连接自编码器的原因有三：

1. 全连接自编码器的参数过多。

卷积神经网络层的另外一个优势是可以大大降低网络的参数数量。搭建本实验所需深度的自编码器，全连接自编码器所需的参数量是卷积自编码器和U-net自编码器参数量的百倍。参数量的巨大不利于计算机储存和网络训练。

1. 相比于其他两个自编码器，全连接自编码器的训练效率低。

面对相同的目标，全连接自编码器的训练效率会比其他两个自编码器低很多。例如针对MNIST手写数字数据集的压缩试验，无论是收敛速度和损失量，卷积和U-net自编码器都比全连接自编码器优秀很多。



注：卷积自编码器训练到第10轮便开始收敛，U-net自编码器训练到第5轮开始收敛，而全连接自编码器一直到30轮结束还没开始收敛，且损失量比其他两个都高

图15 三种自编码器针对MNIST数据集的训练效率

1. 全连接自编码器的预测效果较差。

对MNIST数据集来说，由于样本都是28×28大小的图片，一般训练较少轮数就可以取得不错的效果。以下为训练30轮后各个网络的结果。

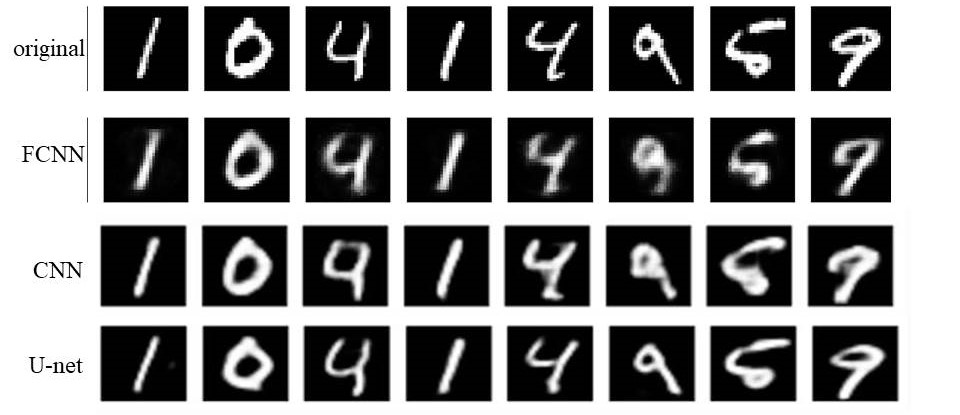


图16 训练30轮后各个网络的效果。第一排为原始图像，第二排为全连接自编码器的输出，第三排为卷积自编码器的输出，第四排为U-net的输出

可以看出，全连接自编码器解压缩恢复的图像分辨率和形状恢复远远不如其他两个自编码器的效果。

综合上述原因，本文不使用全连接自编码器进行实验。

本文针对地震数据压缩和去噪，根据地震数据集的大小，设计了两个网络深度和参数数量相当的自编码器：

1. 卷积自编码器

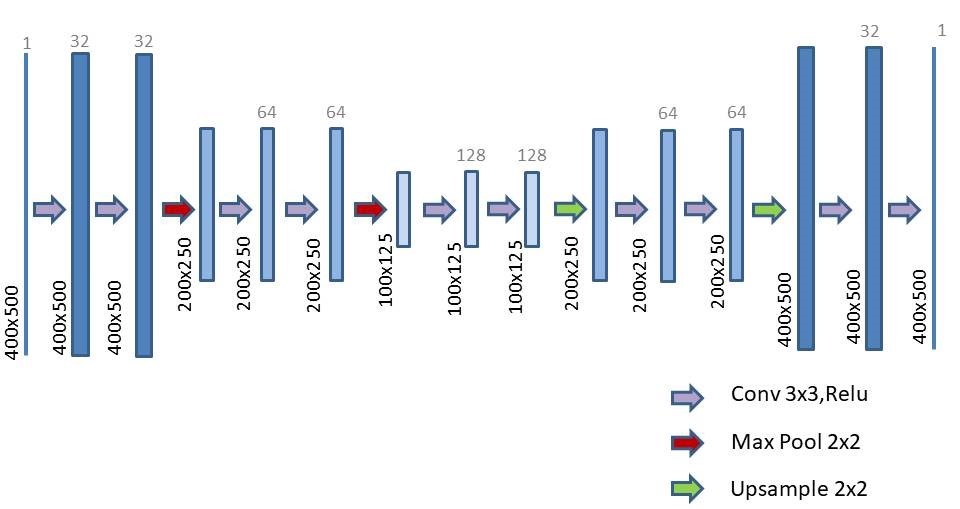


图17 卷积自编码器

1. U-net自编码器

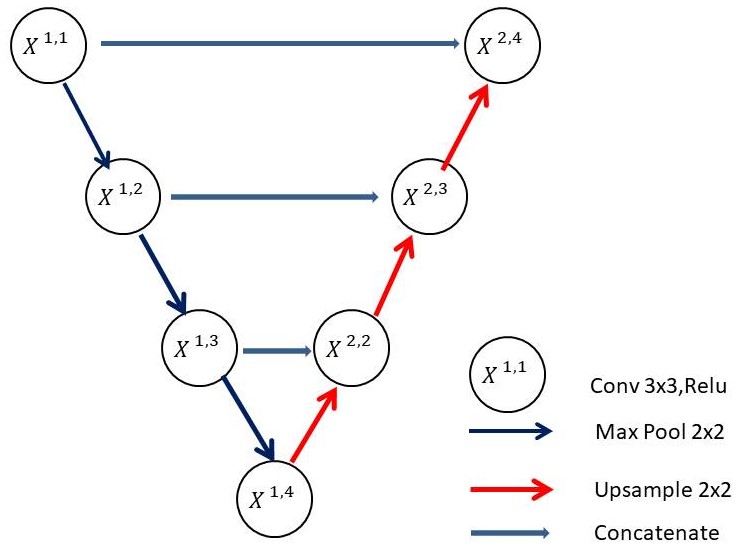


图18 U-net自编码器

由于神经网络模型的泛用性，同一结构的神经网络可以用于不同的用途，输入不同的训练集便可训练出作用不同的神经网络模型，唯一的区别只在于神经网络中的参数差异。所以本文只需设计出相同的卷积自编码器和U-net自编码器拓扑结构，输入压缩训练集和去噪训练集，从而训练出压缩自编码器和去噪自编码器。

两个自编码器的训练效果对比均在下一章详细论述。

1. 实验设计

4.1实验环境

本文所进行的实验依托于中国海洋大学地球探测软件实验室的机群，利用python中的keras神经网络库，使用jupyter编辑器和Matlab进行实验与分析。压缩和去噪实验所使用的训练集为数量为2000，裁剪后大小为400×500的地震记录集，验证集为一个道集数为461，大小为400×500的多炮记录。另外，去噪训练集需要人为添加随机噪声。

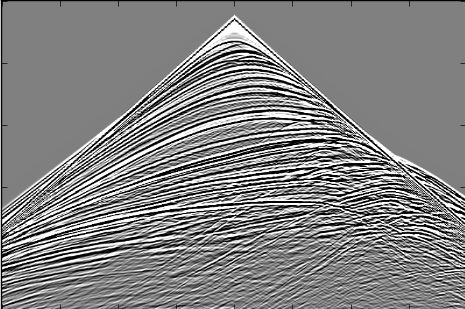
4.2 模型分析评判标准

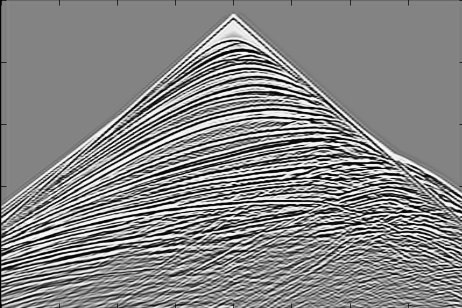
训练自编码器是为了得到效果良好的压缩模型和去噪模型。完美恢复出原图像是模型的最理想结果，但事实往往不能如愿以偿，不能完全恢复原始图像，需要设计出能评判模型效果的参数。而这种参数就是均方误差（MSE）和峰值信噪比（PSNR）。

其中255为灰度图像所能显示的最大像素值，为重建图像的像素值，为原始图像的像素值。下文全部使用MSE和PSNR作为重建和去噪的效果评价参数。MSE值越小，则表明模型的重构质量越高。PSNR值越大，则表明模型的去噪效果好，记录的信噪比高。

4.3 压缩模型实验和分析

实验所输入的训练集为炮集数为2000的地震记录集，训练轮数为30，验证集为炮集数为461的一个多炮记录集。下图为训练30轮过后的网络模型的压缩效果。





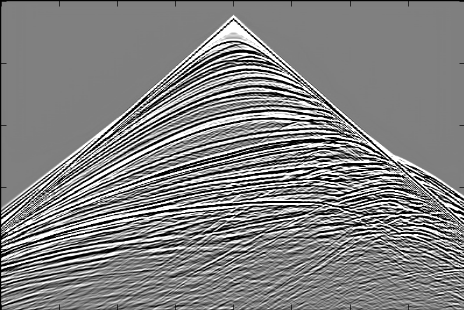
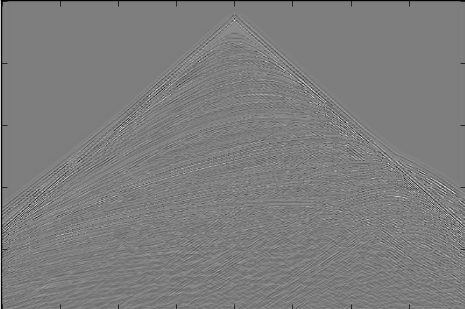


图19卷积自编码器和U-net自编码器30轮训练后的压缩效果。其中，第一个图为原始记录，第二个为卷积自编码器压缩复原后的记录，第三个为U-net自编码器压缩复原后的记录。



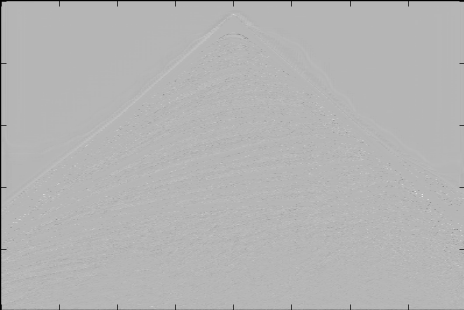


图20 卷积自编码器和U-net自编码器压缩-解压缩后的残差。上图为卷积自编码器的残差，下图为U-net自编码器的残差

下图为两个模型训练过程的损失量。

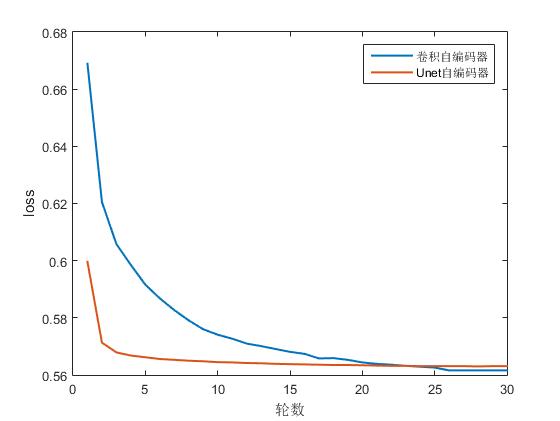


图21卷积自编码器和U-net自编码器训练过程的损失量

网络模型的压缩效果需要借助MSE来表征。下图为验证集中461个地震记录计算出的MSE所构成的曲线。

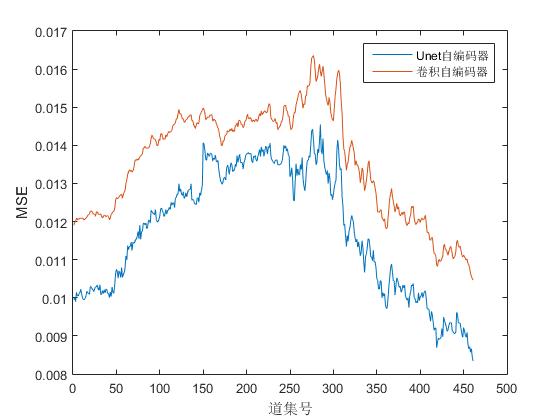


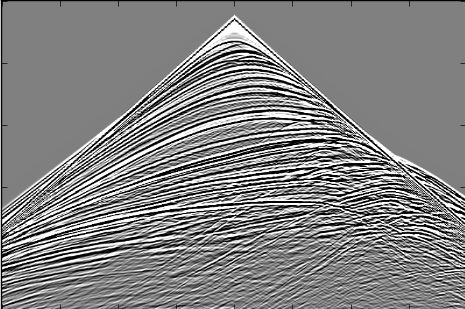
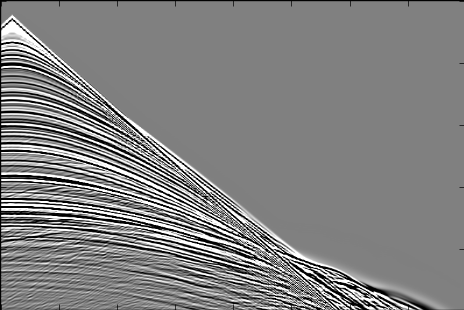
图22卷积自编码器和U-net自编码器的MSE曲线

由损失量曲线，可以看出U-net自编码器收敛速度较快，在第15轮训练时已收敛，收敛梯度逐渐减少。而卷积自编码器的收敛速度较U-net来说比较慢，训练到25轮时开始收敛。两者的误差量都在可接受的范围之内。

经过30轮的训练之后，卷积自编码器和U-net自编码器都取得了不错的效果，二者都能恢复出原始记录的同相轴。U-net自编码器的压缩复原效果分辨率较高，而卷积自编码器的分辨率比U-net自编码器略低一点。从两者的残差图来讲，U-net自编码器重构后的信息损失比较小，而卷积自编码器的损失较U-net来说，比较大。由MSE曲线，也可以看出U-net自编码器的重构质量比卷积自编码器的重构质量高。

总体来说，在地震记录压缩上，二者都能高质量地完成压缩任务，U-net自编码器的效果比卷积自编码器的效果好。

同时MSE曲线呈两边低中间高的形状，可能是验证集各个地震记录的数据量有关。由于验证集是一个多炮记录集，炮点一开始在检波点序列的前部，此时检波点序列接收到的数据量较少；随着炮点的移动，炮点逐渐向序列的中部移动，此时检波点序列接收到的数据量较大；炮点逐渐向序列的尾部移动，此时检波点序列接收到的数据量较少，从而地震记录的数据量从第一个记录开始逐渐递增，递增到某一个记录时，数据量达到最大，后面的记录数据量又降低。



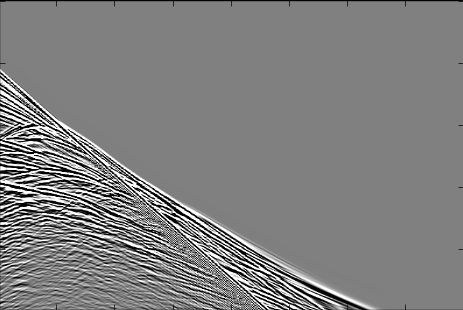


图23 验证集中第10,200,450个地震记录。第一个图是道集号为10的记录，第二个图是道集号为200的记录，第三个图是道集号为450的记录

信息量是影响压缩模型的一个重要因素，地震记录的数据量在一定程度上可以由地震记录的稀疏度来表示。

稀疏度越大，地震记录的数据量就越大，反之，则数据量就越小。

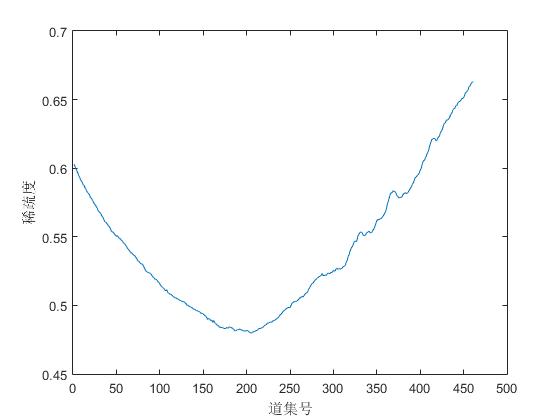


图24 训练集中每一个地震记录的稀疏度

一般来说，数据量越大，代表着信息越复杂。记录的数据量越大，经过网络模型压缩-解压缩之后，复原的效果越差，MSE也就越大。

对于实际工作情况，程序的运行时间对工作效率的影响是非常大的。压缩模型的压缩时间也是评价模型的重要指标。下图为卷积压缩自编码器和U-net压缩自编码器的平均工作时间。

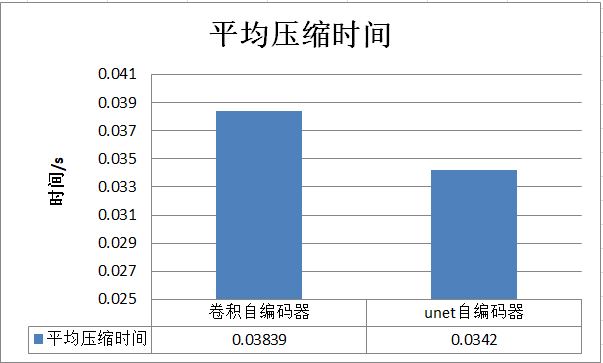


图25 卷积自编码器和U-net自编码器的平均压缩时间

可以看出，两个压缩模型的平均运行时间十分相近，只相差0.004秒，工作效率非常高，达到了十毫秒级别，具有一定实用性。

4.4 去噪模型实验和分析

去噪实验所输入训练集与压缩实验的训练集大致相同，区别的是去噪模型实验的训练集是经过人为添加随机噪声的。下图为经过训练30轮后的两个自编码器的去噪效果。

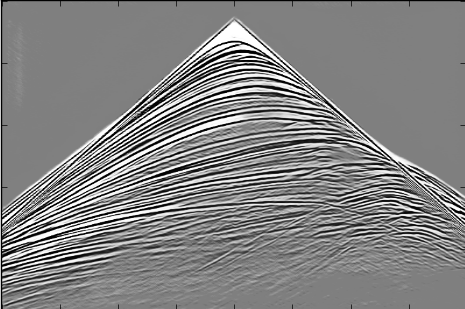
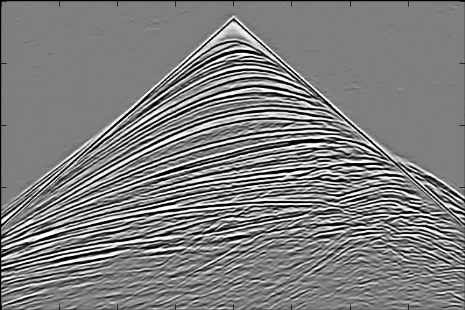
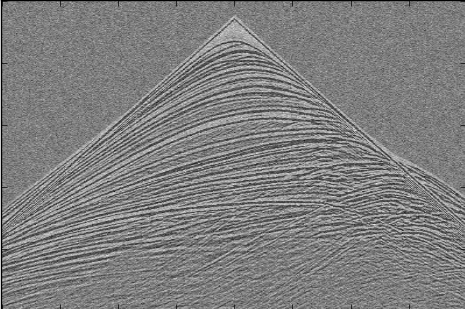


图26卷积自编码器和U-net自编码器去噪效果。其中第一个图是含有噪声的原始记录，第二个图是卷积自编码器去噪后的记录，第三个图是U-net自编码器去噪后的记录

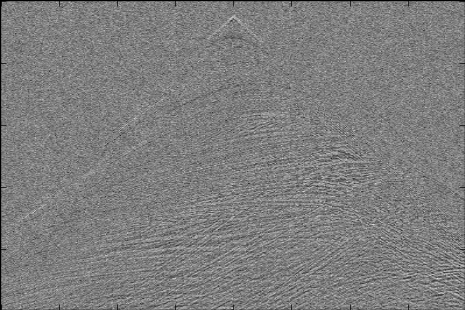
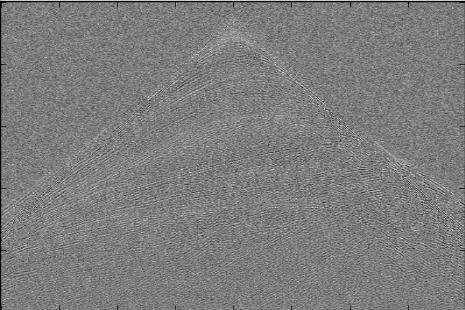


图27卷积自编码器和U-net自编码器去噪后的残差。上图为卷积自编码器的残差，下图为U-net自编码器的残差

下图为训练过程的损失量。

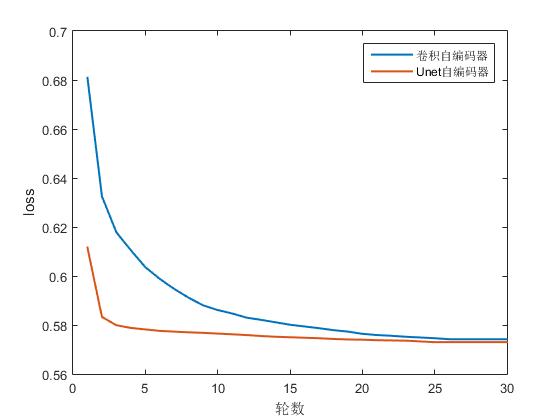


图28卷积自编码器和U-net自编码器训练过程的损失量

去噪模型的去噪效果，需要借助峰值信噪比（PSNR）来评判。

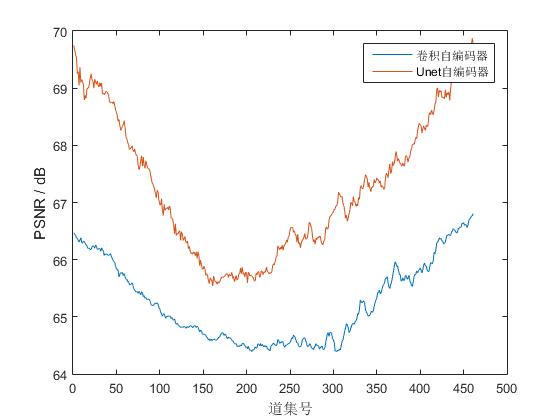


图29卷积自编码器和U-net自编码器的PSNR曲线

与压缩模型训练的情况相似，U-net自编码器收敛速度比较快，在第15轮就开始收敛；而卷积自编码器在第25轮开始收敛。两者的误差都控制在可接受的范围之中。

从图26和图27可以看出，卷积自编码器和U-net自编码器可以将随机噪声完全将随机噪声完全压制住，在噪声压制方面取得良好的效果。但是两者都存在同一个问题：对原始记录中的同相轴的恢复效果比较差，基本上只能恢复出原始记录中比较明显的同相轴。尤其是U-net自编码器，原始记录下部的记录识别为随机噪声被压制，不能完全恢复为原始记录。这可能是由于U-net结构的自编码器学习能力太强，以至于将下部的记录都学习成噪声并将之压制。

由PSNR曲线可以看出，U-net自编码器和卷积自编码器都可以取得信噪比较高的效果，其中U-net自编码器的信噪比比卷积自编码器的信噪比高，但二者相近，差距不是特别的大。

综合信噪比和损失情况，在深度和参数数量相当的情况下，卷积自编码器的去噪效果比U-net自编码器的去噪效果好。

另外，PSNR曲线成两边高中间低的形状，说明数据量小的去噪效果比数据量大的记录的去噪效果好。

传统的去噪方法需要工作人员人工设定参数，进行多次试验才能确定最佳的去噪参数组合，操作过程可能需要数十分钟之久。下图为卷积去噪自编码器和U-net去噪自编码器的平均运行时间。

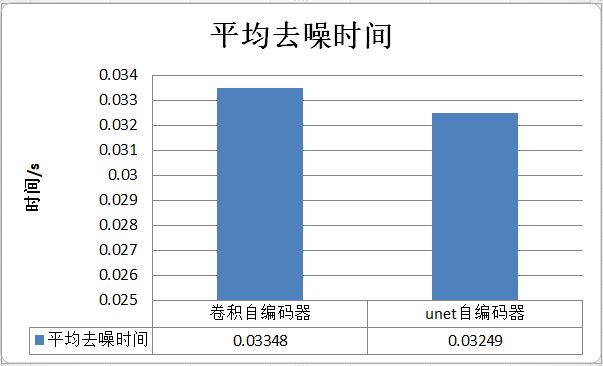


图30 卷积自编码器和U-net自编码器的平均去噪时间

可以看出，卷积自编码器和U-net自编码器的平均压缩时间十分的相近，达到了十毫秒级别。相比于传统的去噪方法，基于自编码器的地震数据去噪算法可以大大提高地震勘探工作的效率。

4.5 实验结论

本文建立卷积自编码器和U-net自编码器应用于地震数据压缩和去噪。首先输入训练集训练卷积自编码器和U-net自编码器的压缩与去噪模型，验证模型的可行性，借助MSE和PSNR来衡量模型训练的效果。

实验表明卷积自编码器和U-net自编码器在地震数据压缩与去噪上都能取得高质量的效果，二者各有千秋。在地震数据压缩方面上，总体上U-net自编码器的效果比卷积自编码器的效果好。在地震记录随机噪声压制方面上，总体上卷积自编码器的效果优于U-net自编码器。

经过上述实验，充分证明了将自编码器算法应用于地震数据压缩和去噪是可行的，自编码器算法能简单快捷地压缩-解压缩和去噪，且失真率较低，工作效率高。

另外，由于压缩后的数据已经与原始数据完全没有相似之处，且在解压缩结构不外泄的情况下，自编码器还会有数据加密的作用，对于传输或运送机密地区地震记录是一个可靠安全的保障。

1. 总结与展望

5.1总结

地震勘探是探测地球内部的最有效方法，也是最重要的手段之一。近年来，关于地震数据的压缩的研究逐渐火热起来。从1983年笠原顺三提出的ADPCM编码方法实现地震数据压缩之后，国内外学者对地震数据压缩技术的研究取得了丰厚的成果。地震数据压缩的研究十分的重要，特别是三维高分辨率地震勘探的发展和无线传送的需求的产生，使得地震数据压缩变得愈发重要。地震资料的噪声压制一直是这个领域的研究热点，各位学者都提出了各自的方法取得了十分理想的效果。本文利用中国海洋大学地球探测软件实验室的机群，针对地震数据压缩与去噪，对自编码器神经网络的进行了测试。此方法框架简单，理解直观，而且压缩比和信噪比都不俗。

本文的主要内容总结如下：

（1）对已提出的地震数据压缩和去噪研究进行了总结，得到自编码器神经网络也可应用于地震数据压缩与去噪领域上的结论。

（2）设计了参数数量相当的卷积自编码器和U-net自编码器，应用于地震数据压缩和去噪。

（3）利用实验室的机群针对地震数据，对模型进行训练，并对比各个自编码器的压缩与去噪结果，取得了良好的结果。其中，在地震数据压缩方面上，总体上U-net自编码器的效果比卷积自编码器的效果好。在地震记录随机噪声压制方面上，总体上卷积自编码器的效果优于U-net自编码器。

5.2展望

随着机器学习技术的不断发展，将以神经网络为代表的人工智能应用到地球物理领域已不再是遥远的事情，而会在未来相当短的时间内迅速发展起来，成为地球探测新的有力工具，需要地球物理学家们孜孜不倦的拓宽人工智能在地球物理的道路。

（1）相比于现实状况的地震记录，本文所使用的训练集样本量太少。若想进一步提升模型压缩与去噪效果，需要搭建数量更大的训练集。另外，训练集的尺寸大小为400×500，而实际地震记录往往比这个尺寸大得多，所以，要真正应用到实际处理上，需要增大训练集的尺寸以适用于实际地震记录。而且，需建立一个可以应用于任意尺寸地震记录的网络模型。

（2）应用于实际情况的神经网络往往是达到足够深的深度的。神经网络的层数越深，参数的数量越多，模型的抽象性越强，能提取更深层次的特征。但是神经网络层数如果过深的话，会出现网络退化的情况——深度过深的神经网络效果很有可能会比深度浅的神经网络效果差。另外深度过深，参数数量过多，会造成网络训练难度上升，效率减小，使得网络难以训练。参数量过大会造成计算机额外的内存负担，影响计算机运行速度。除此之外，深度过深的神经网络会造成过拟合的情况，降低模型应用于训练集以外数据的效果。所以，研究人员需要仔细验证，以找到最合适模型结构。

（3）机器学习技术在地震勘探领域尚大有作为的空间。未来可以考虑将多种功能结合到一个网络上，如训练出一个可以同时去除随机噪声和规则噪声如多次波等的网络，大大提高工作人员的工作效率。

（4）地震记录与数字图像有着异曲同工之处，机器学习已在图像处理领域取得了相当惊人的成果。地球物理学家可以向计算机科学家们取经，将图像处理领域中发展的机器学习技术应用到地球物理领域上，这也符合当今时代学科融合的趋势。

参考文献

1. Kanjo, Kenji,Kasahara, Junzo,Takahashi, Masayoshi. An Application of the Adaptive Differential PCM(ADPCM) Method to the Seismic Wave Signal Compression.[J] Bulletin of the Earthquake Research Institute, University of Tokyo, 1984:3
2. 郭洪升. 地震数据的实时自适应压缩方法研究[A]. 中国地震学会.中国地震学会第三次全国地震科学学术讨论会论文摘要汇编[C].中国地震学会:,1986:2.
3. 陈志文. 应用分形维进行地震数据的压缩与恢复[J]. 石油物探,1992,(01):1-7
4. 陈明伟. 地震数据压缩[A]. 中国地球物理学会.1994年中国地球物理学会第十届学术年会论文集[C].中国地球物理学会:,1994:1.
5. 王嘉松,贾国平. 压缩地震数据的新方法[J]. 石油物探,1994,(03):26-32+39
6. 刘财,王培茂,薛亚丽,刘洋,张智,邵志刚.离散余弦变换(DCT)编码在地震勘探数据压缩中的应用[J].吉林大学学报(地球科学版),2004(02):277-282.
7. 孙寰宇,武文波,杨志高,秦前清. 基于JPEG2000的地震数据准无损压缩[J]. 计算机工程与应用,2005,(16):185-188.
8. 耿瑜. 地震数据的Curvelet压缩方法[A]. 中国地球物理学会.中国地球物理学会第二十三届年会论文集[C].中国地球物理学会:,2007:1
9. 徐锋涛,张正炳,桂志先.基于EZW算法的地震数据压缩[J].石油地球物理勘探,2015,50(05):881-889+803.
10. Yaniv Hollander, Amin Merouane, and Orhan Yilmaz, Emerson-Paradigm.Using a deep convolutional neural network to enhance the accuracy of first break picking[C] segam2018-2982650.1.Page 4628
11. Yue Ma1, Xu Ji, Tong W. Fei, Yi Luo. Automatic velocity picking with convolutional neural networks.[C] segam2018-2987088.1.Page 2066.
12. Sara Mandelli, Federico Borra, Vincenzo Lipari\*, Paolo Bestagini, Augusto Sarti and Stefano Tubaro. Seismic Data Interpolation Through Convolutional Autoencoder.[C] segam2018-2995428.1.Page 4101
13. Hampson D. Inverse velocity stacking and multiple elimination[J].Journal of the Canadian Society of Exploration Geophysicists,1986(22):44-55
14. Ethan J N, Matthias GI, Amplitude Preservation of Radon-based multiple-removal filters.[J]Geophysics,71(5):123-126
15. Herrmann F J, Wang D, Hennenfent G, et al. Curvelet-based seismic data processing: A multiscaleand nonlinear approach [J]. Geophysics, 2008b, 73(1):A1-A5.
16. Neelamani R, Baumstein A, Ross W. Complex curvelet based adaptive substraction of several noise Templates[C].Expanded Abstracts of 78th Annual International SEG Meeting, 2008: 3650- 3655.
17. 张华,陈小宏,李红星,黄光南,陈晓.曲波变换三维地震数据去噪技术[J].石油地球物勘探,2017,52(02):226-232+192.
18. Xu Y, Weaver J B, Healy D M, et al. Wavelet transform domain filters: a spatially selective noise Filtration technique.[J]IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society,1994,3(6):747-58.
19. Yuchen Jin, Xuqing Wu, Jiefu Chen, Zhu Han, Wenyi Hu.Seismic Data Denoising By Deep Residual Networks.[C] segam2018-2998619.1.Page 4593.
20. Zhang K，ZuoW，Chen Y，et al．2017．Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising.[J] IEEE Transactions on Image Processing. 26(7):3142-3155.
21. Peigen XIE, Jean-Luc BOELLE, Henri PUNTOUS. Generative Adversarial Network Based Fast Noise Removal on Land Seismic Data.[C]10.1190/segam2018-2995310.1 Page 2171
22. Xu Si, Yijun Yuan. Random noise attenuation based on residual learning of deep convolutional neural network.[C] 10.1190/segam2018-2985176.1 Page 1986
23. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature, 1986,323(6088): 533–536. [doi: 10.1038/323533a0]
24. Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation.[C] Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234--241, 2015

致谢