

文章编号: 1001-1749(2023)04-0484-13

# Transformer 网络在大地电磁反演成像中的应用

刘高村, 王绪本, 袁崇鑫, 李德伟, 谢卓良

(成都理工大学 地球勘探与信息技术教育部重点实验室, 成都 610059)

**摘要:** 传统大地电磁反演通常是基于确定性梯度的迭代求解, 不仅需要大量时间计算雅可比矩阵, 还依赖于初始模型的输入和正则化因子等参数的设置。近年来学者们不断引入机器学习方法以试图改善大地电磁反演, 该方法不需要计算雅可比矩阵, 不用输入初始模型, 训练好的网络仅需几毫秒就可实现反演成像。这里利用 Google 团队提出的 Transformer 神经网络经典框架搭建大地电磁数据和模型之间的映射网络, 以 9 240 组正演数据为样本, 对 Transformer 网络参数进行训练。采用南非开源大地电磁数据, 实现了由视电阻率图像到电阻率模型的反演成像。研究表明: ①经训练后的 Transformer 网络可以较准确的反映出异常体位置和大小; ②网络实现了简单的矩阵并行化运算, 大幅度提高训练的效率和成像效率高于传统的反演。

**关键词:** 大地电磁测深; 有限元正演; Transformer; 多头注意力机制; 反演成像

**中图分类号:** P 631.2 **文献标志码:** A **DOI:** 10.3969/j.issn.1001-1749.2023.04.09

## 0 引言

大地电磁测深法是一种利用天然电磁场测量, 研究地球深部电性结构特征的一种地球物理勘探方法, 普遍运用于矿产普查、区域构造研究、深部地球动力学等方面。该方法在近 70 年的发展历程中, 大地电磁反演一直是学者比较关注的研究方向。Constable 等<sup>[1]</sup>提出了 Occam 反演, 随后由 Siripunvaraporn 等完善了 Occam 反演法, 在 Occam 反演法的基础上, Smith 等<sup>[3]</sup>提出了快速松弛反演, 这类方法需要添加正则化因子或者计算雅可比矩阵, 且对初始模型依赖性强, 较好的初始模型会使迭代过程平稳, 容易收敛, 取得很好的反演效果。根据目标函数梯度信息, 学者提出了 NLCG 反演、最小二乘法和高斯牛顿法等梯度方法<sup>[4]</sup>。这类方法避免了计算雅可比矩阵, 且计算效率高, 收敛性强; 该方法最

大的局限依赖于初值, 不能保证收敛到全局极值, 以上传统反演方法理论已经十分完善, 但均需要设置初始模型, 对初始模型十分依赖<sup>[5]</sup>。于是学者提出了一些新的非线性反演方法, 如: 蒙特卡洛法<sup>[6]</sup>、遗传算法<sup>[7]</sup>、模拟退火<sup>[8]</sup>、粒子群算法<sup>[9]</sup>、神经网络<sup>[10]</sup>等算法, 这些反演方法对初始模型依赖性小, 不用计算灵敏度矩阵, 整体思路为寻找全局最优解<sup>[11-12]</sup>。其中人工神经网络反演不需要多次的迭代, 消耗反演时间, 训练好的网络可以快速成像, 受到学者的喜爱。

起初人工神经网络主要使用 BP 神经网络, 其仿生学独有的学习记忆功能, 对数据进行自组织和自适应能力的自主学习, 实现非线性反演, 如李创世等<sup>[13]</sup>将 TEM 曲线转化为 BP 神经网络反演应用于瞬变电磁勘探; 徐江<sup>[14]</sup>把 BP 神经网络运用到半航空瞬变电磁法, 验证了 BP 神经网络反演的准确性和有效性; 胡祖志等<sup>[15]</sup>总结了 BP 神经网络

在三维中的应用情况;徐海浪等<sup>[16]</sup>采用 BP 网络实现了高密度电法的二维电阻率反演。为了完善 BP 神经网络反演,学者提出用粒子群优化算法等优化 BP 神经网络<sup>[17-19]</sup>,提高反演的精确度;敬荣中<sup>[20]</sup>将模糊系统和神经网络结合,在 VES、TEM 和 MT 均取得了较好的效果,以上 BP 神经网络属于浅层神经网络,对于复杂的实际大地电磁数据,需要建立更加复杂的映射关系,于是学者提出深度神经网络<sup>[21]</sup>。卷积神经网络是深度学习在图像中运用较为广泛的一种,众多学者把卷积神经网络(CNN)运用在电磁法二维及三维数据处理<sup>[22-26]</sup>取得了较好的效果。除此之外,还有学者提出 KNN 神经网络和长短记忆循环神经网络等深度神经网络,均取得了较为明显的效果<sup>[27-28]</sup>。对比传统反演方法,机器学习方法在海量数据处理、大地电磁噪声模式识别和特征提取、数据预测等方面具有巨大的优势<sup>[29-30]</sup>,然而获取训练网络的样本需要正演产生,大量的样本会消耗大量时间,目前仅仅适合 1D 或者小规模 2D 数中<sup>[5,31]</sup>。笔者将在前人的研究上继续探讨神经网络方法在大地电磁法中的应用,将自然语言翻译中最为受到关注的 Transformer 网络运用至大地电磁反演成像中。该网络框架实现了并行输入,但有别于 CNN 和 RNN 读取信息,节省了大量的时间。

Transformer 网络由 Google 团队<sup>[32]</sup>提出,放弃了传统卷积神经网络和循环神经网络的模型架构,成为了自然语言处理任务中的实际标准。Parmar 等<sup>[33]</sup>将 Transformer 架构迁移到计算机视觉领域。后基于该网络的出色表现,周哲韬等<sup>[34]</sup>利用振动信息进行预测,提高了预测的准确度;何林远等<sup>[35]</sup>提出稀疏 Transformer,能很好地适应遥感图像中各类旋转目标的形状与分布特性。以上研究表明,该网络框架在自然语言翻译和图像处理、数据分析处理上均有着较好的效果,在训练速度上远远地超过了循环神经网络和卷积神经网络。于是笔者将探索其在 MT 数据处理中应用的可能性,并提出针对 MT 数据处理时的注意事项及改进方向。结果表明,该方法具有较快的训练速度,网络具有一定的泛化能力,该方法是切实可行的。

## 1 方法原理

### 1.1 Transformer

Transformer 网络框架基础是利用并行的多个

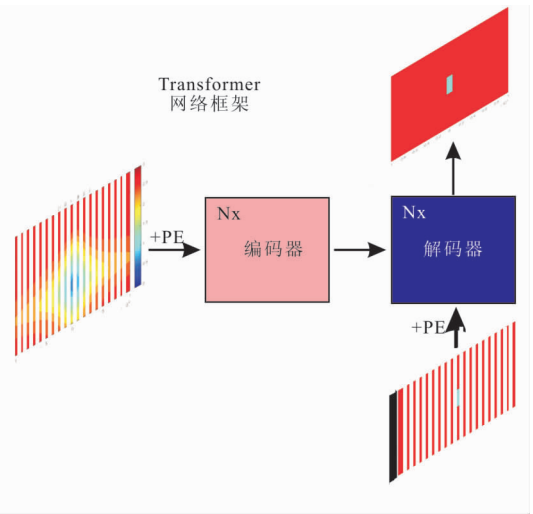


图 1 Transformer 框架

Fig. 1 Transformer framework

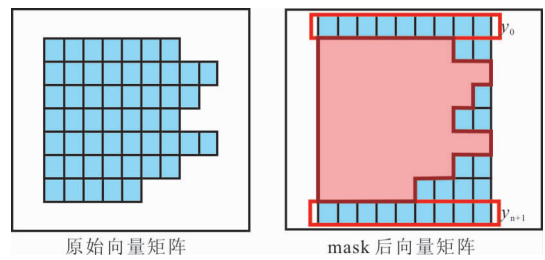


图 2 序列掩码

Fig. 2 Padding mask

注意力机制获得不同的网络参数,提取不同的信息,再通过实现编码和解码的方式预测出所需要的结果。

一个完整的 Transformer 如图 1 所示,包括输入编码、位置编码(Positional Encoding,图中用 PE 表示)、编码器、解码器、全连接层。灵活运用多头注意力机制是该网络的亮点,实现了并行化处理,加快了网络的处理速度以及处理效果。

### 1.2 输入编码和掩码

Transformer 网络输入信息时需要编码,使语言文字变为机器可识别的维度矩阵,同时添加掩码。其中,包括两种掩码方式,填充掩码和序列掩码。填充掩码主要是针对语言长短不一的问题提出的,如图 2 所示方式补全维度信息,输出矩阵还需要增加  $y_0$  和  $y_{(n+1)}$  向量,添加两个向量序列作为额外信息输入;序列掩码是运用在解码器中用于掩饰未来信息中的编码。

### 1.3 位置编码

Transformer 有别于卷积神经网络和循环神经网络的并行网络结构,输入信息里无时间顺序位置

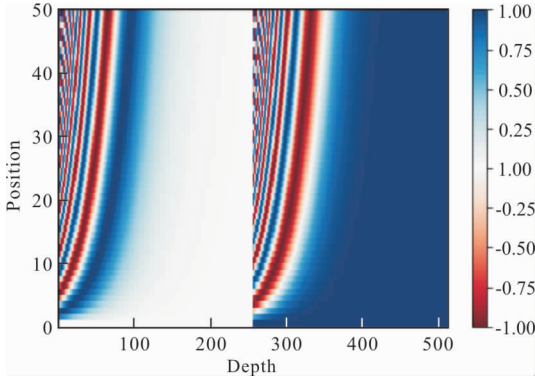


图3 位置编码

Fig. 3 Positional Encoding

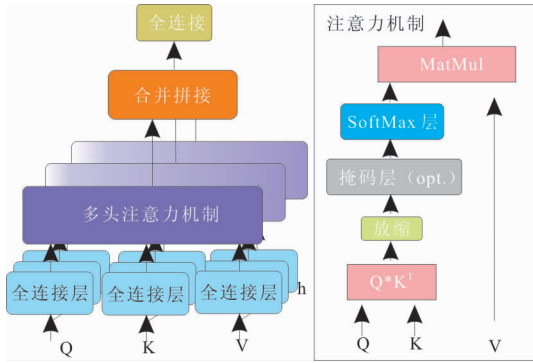


图4 多头注意力机制示意图

Fig. 4 Schematic diagram of Multi-Head attention

信息,为解决位置关系,引入位置编码,构成公式(1)所示的位置关系。

$$\begin{cases} PE_{(l,2i)} = \sin \varphi \\ PE_{(l,2i+1)} = \cos \varphi \end{cases} \quad (1)$$

式中: $PE$ 为编码矩阵; $\varphi = l/1000^{2i/d}$ ;  $l$ 代表测点所对应的位置; $i$ 为矩阵的奇偶位置,取值范围在 $[0, 0.5d-1]$ ;  $d$ 为频点数。为更好展示位置编码,绘制如图3所示,左半部分的值由正弦函数生成,右半部分由余弦函数生成。

对任意向量 $X_l$ 进行位置编码,由式(2)生成一个全新向量 $a_l$ :

$$a_l = X_l + PE_l \quad (2)$$

式中: $PE_l = [\sin(l/f_1), \cos(l/f_1), \sin(l/f_2), \cos(l/f_2), \dots, \sin(l/f_{(d/2)}), \cos(l/f_{(d/2)})]$ ,  $PE_l$ 是 $X_l$ 所对应的位置编码, $f_i = 1000^{2i/d}$ 。 $X_l$ 经过位置编码处理后能使Transformer神经网络在向前传播和梯度训练中将位置信息保存下来,进一步将输入向量之间的相对位置和绝对位置保存下来。

#### 1.4 多头注意力机制

多头注意力机制(Multi-Head Attention)是

Transformer中的核心关键,如图4所示,将单头注意力机制复制为多个,提取了不同信息后再把再多个头拼接通过一个全连接层控制输入和输出的维度保持不变,得到最终的输出。

注意力机制将输入矩阵 $X$ 矩阵通过3个待训练的全连接层,经过三个 $W$ 矩阵,线性变换后生成 $Q$ (query)、 $K$ (key)、 $V$ (value)矩阵:

$$Q = W^Q X \quad (3)$$

$$K = W^K X \quad (4)$$

$$V = W^V X \quad (5)$$

计算 $Q$ 和 $K$ 的相似性或者相关性,获得 $K$ 对应 $V$ 的权重矩阵,为避免在网络训练中致使训练参数过大,还需除以 $\sqrt{d_k}$ ,通过Softmax归一化后获得最终的权重因子,对权重和 $V$ 进行加权求和,即得到了最终的Attention数值。即:

$$Attention = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

考虑大地电磁与自然语言的差异,在进行线性变换生成 $Q, K, V$ 矩阵时,将原本的低维信息扩展到一个较高维度后,我们不仅要关注高阻信息,也需要注意低阻异常,而原来的注意力机制会被高阻吸引,于是引入了投影概念作为注意力机制的一部分,  $\|K^T\|$ 为向量 $K^T$ 的范数:

$$Attention = \text{Softmax}\left(\frac{QK^T}{\|K^T\| \sqrt{d_k}}\right)V \quad (7)$$

#### 1.5 编码器

编码器(Encoders)是由 $N$ 个Encoder层组成,每个Encoder包括一个多头注意力机制(Multi-Head Attention)和一个前馈神经网络(Feed Forward Network),它们依靠Layer Normalization进行残差拼接而成,如图5所示。

残差连接<sup>[36]</sup>表示为:

$$layre\_output = x + Layer(x) \quad (8)$$

那么任意 $N$ 层与 $n$ 层之间( $N > n$ )的关系可以表示为:

$$x_N = x_n + \sum_{i=n}^{N-1} F(x_i) \quad (9)$$

于是在网络的传播过程中,不仅可以保存原始信息,又可提取特征信息。

在梯度训练中损失函数 $f$ 与 $x_l$ 梯度表达为:

$$\frac{\partial f}{\partial x_n} = \frac{\partial f}{\partial x_N} \frac{\partial x_N}{\partial x_n} = \frac{\partial f}{\partial x_N} \left(1 + \frac{\partial}{\partial x_n} \sum_{i=1}^{N-1} F(x_i)\right) \quad (10)$$

最终化简为:

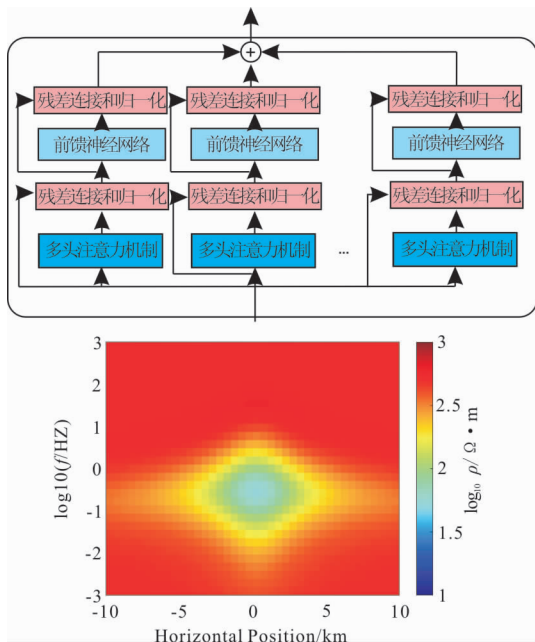


图 5 编码器流程图

Fig. 5 Encoder flow chart

$$\frac{\partial f}{\partial x_n} = \frac{\partial f}{\partial x_N} + \frac{\partial f}{\partial x_N} \frac{\partial}{\partial x_n} \sum_{i=1}^{N-1} F(x_i) \quad (11)$$

于是在训练过程中比传统多出一个常式项

$\frac{\partial f}{\partial x_n}$ , 而  $\frac{\partial}{\partial x_n} \sum_{i=1}^{N-1} F(x_i)$  不会恒等于 -1, 那么便避免了梯度消失的问题, 同时加入残差打破了原本的线性表示解决了网络模型的退化问题。

前馈神经网络由两个全连接层通过维度变换提取信息, 并引入 drop 层和激活函数 ReLu。

$$Layer_1 = \text{linear}(x, \text{ReLu}) \quad (12)$$

$$Layer_2 = \text{linear}(\text{drop}(Layer_1), \text{ReLu}) \quad (13)$$

两次空间维度的变化使得原有的信息更为丰富, 加入 drop 层和 ReLu 函数增加网络的非线性, 有效的预防梯度爆炸与梯度消失等问题。

## 1.6 解码器

解码器 (Decoders) 结构形式与编码器结构组成相似, 不同的是多了遮掩多头注意力机制 (mask multi-Head Attention) 层, 相对于普通的多头注意力机制, 该层多加了序列掩码层, 添加在  $Q$  和  $K$  矩阵内积除以放缩因子和 Softmax 之间, 用来掩盖未来数据, 使之不提前被泄露。然后经由线性变化生成矩阵作为多头注意力机制的  $K$  矩阵输入, 而编码器的最终输出经由线性变化生成  $Q$  与  $V$  矩阵输入 (图 6)。需要注意的是, 解码器层数需要和编码器层数一样。

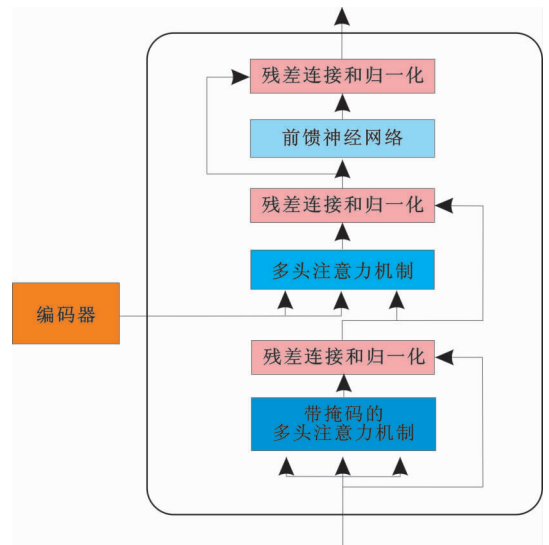


图 6 解码器流程图

Fig. 6 Decoder flow chart

## 2 Transformer 大地电磁反演成像

### 2.1 程序编码与网络训练

笔者借助了由 Shewchuk J R<sup>[37]</sup> 提出的自适应三角网格剖分方法, 完成了 9 240 组网格剖分与模型建立, 正演频点设为大地电磁常用范围  $10^{-3}$  Hz ~  $10^3$  Hz, 共 32 个频点, 利用 tensorflow2.0 库搭建了 Transformer 神经网络的自定义层。

将均方损失函数 (MSE) 用于评价深度学习网络训练标准的损失函数, 其值作为整个网络的预测误差, 记为:

$$MSE = \frac{1}{N} \sum_{n=1}^N (P - u)^2 \quad (14)$$

式中:  $P$  为预测输出;  $u$  为期望输出, 即大地电磁测深反演成像中的理论模型电阻率参数。

这里使用的激活函数:

$$\text{ReLU}(x) = \max(x, 0) \quad (15)$$

$$\text{Tanh}(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \quad (16)$$

更新网络权重和阈值采用 Adam 优化器, 该方法是结合了 AdaGrad 优化算法和 RMSProp 优化算法, 计算速度快, 且对内存的需求小, 能自动的调整学习速度, 训练效果好。

### 2.2 大地电磁正演模型

本次正演选用自适应三角网格剖分有限元方法, 共创建了 9 240 组正演模型, 如图 7 所示, 建立地下高阻异常体和地下低阻异常体各 4 620 组, 在

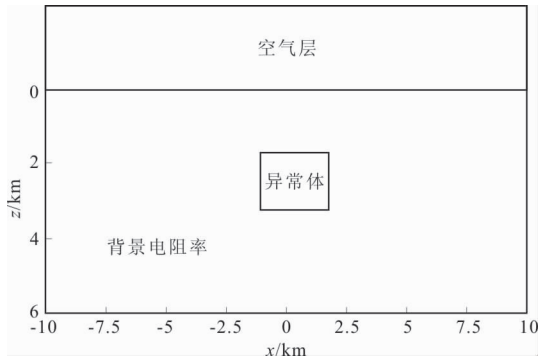


图7 模型示意图

Fig. 7 Schematic diagram of model

有限元正演中将区域划分为地质模型区域和向外延区域。地质模型区域  $x$  设为  $[-1\ 000, 1\ 000]$ , 埋深  $h$  范围设为  $[0, 6\ 000]$ , 考虑响应影响区域将测点采集区域设置在  $[-10\ 000, 10\ 000]$ , 测点间距为 250 m。整体区域向左延展至  $-10^5$  m, 向右延展至  $10^5$  m, 底部延伸至  $2 \times 10^5$  m。

异常体在浅部容易探测, 深部异常体不容易被探测。为了选取具有代表性的模型, 采用控制变量法, 将模型分为浅层异常体和深部异常体。异常体设为正方形, 表 1 和表 2 中顶点为矩形左上角控制顶点, 宽为矩形的边长, 其控制右上角顶点, 埋深为异常体顶界面的深度。由于设置地质模型  $x$  范围在  $[-1\ 000, 1\ 000]$ , 故而筛选后, 浅部异常体模型 5 400 个, 深部异常体模型 3 840 个, 共创建样本 9 240 个。

表 1 浅部异常体参数表

Tab. 1 Parameters of shallow abnormal body

顶点	$[-1\ 000, 1\ 000]$ 按 100 间隔取样
宽	$[100, 1\ 000]$ 按 200 等间隔取样
埋深	$[500, 700, 900, 1\ 100, 1\ 300, 1\ 500]$
异常体电阻	$[0.1, 1, 10, 1\ 000, 2\ 000, 5\ 000]$
背景电阻	$[100, 500]$

表 2 深部异常体参数表

Tab. 2 parameters of deep abnormal body

顶点	$[-1\ 000, 1\ 000]$ , 按 200 等间距取样
宽	$[500, 750, 1\ 000]$
埋深	$[1\ 500, 5\ 000]$ , 按 500 等间距取样
异常体电阻	$[0.1, 1, 10, 1\ 000, 2\ 000, 5\ 000]$
背景电阻	$[100, 500]$

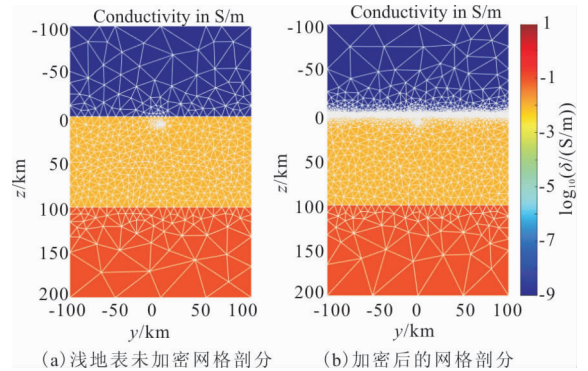


图8 地质模型三角网格剖分

Fig. 8 Triangulation of geological model

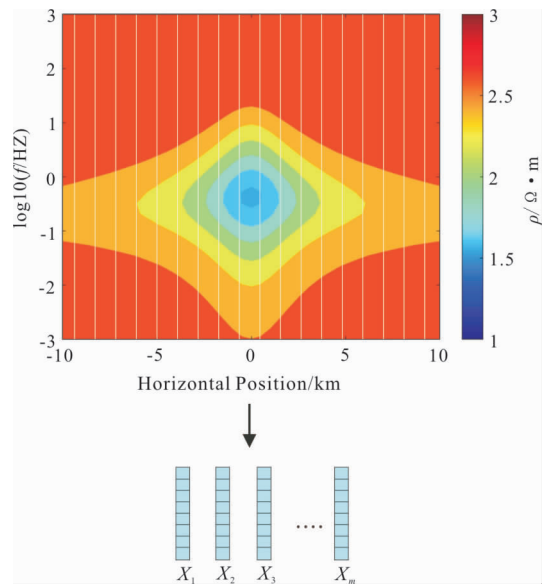


图9 视电阻率断面预处理图

Fig. 9 Pretreatment of apparent resistivity section

正演生成大量样本需要时间, 在空气层以及深部扩展层选用较大的网格, 而在异常体附近和中间层选用较为小的网格。地质模型网格剖分情况如图 8(a)所示, 0 m 以上蓝色部分为空气层, 电导率设为  $10^{-9}$  (s/m), 中部为地层区域, 电导率为  $10^{-2}$  (s/m), 最底部红色为网络扩展部分, 电导率为  $10^{-1}$  (s/m)。为加密地表浅部的节点数量, 新增一层 100 m 厚度的地层。加密网格剖分后效果如图 8(b)所示。图中白色为网格剖分线, 而异常体和 100 m 厚度层处因为网格剖分过密, 图中呈白色部分。

有限元正演计算时频点范围设为大地电磁测深中常用的  $10^{-3}$  Hz  $\sim$   $10^3$  Hz, 共 32 个频点, 取对数等间距分布, 采样点设置为 41 个, 间距为 500 m。

### 2.3 Transformer 大地电磁反演成像

在进行大地电磁反演成像时把 TE 模式和 TM

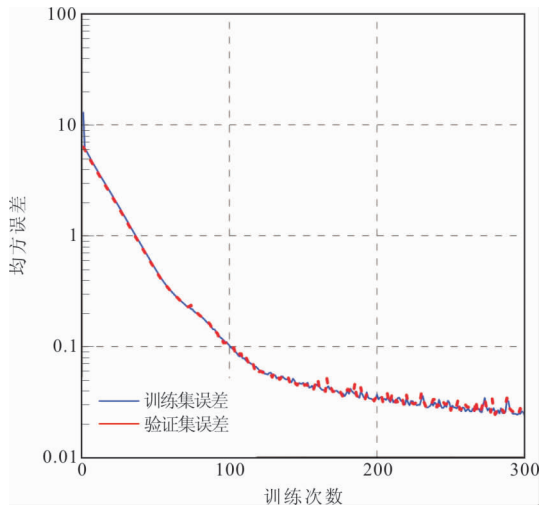


图 10 训练误差收敛曲线

Fig. 10 Training error convergence curve

模式的视电阻率视作自然图像,如图 9 所示,进行预处理切割,以避免个别频点缺失的情况。将图像分割为 80 等份,通过图像信息中的 RGB 像素信息为一维电阻率的向量,取对数归一化后作为输入信号输入至编码器,即为  $80 \times 40 \times 2$  数组矩阵,再将地电模型矩阵输入至解码器,即  $20 \times 70$  数组矩阵,作为监督学习,实现双向输入,总计输入 7 800 个,输出矩阵为  $20 \times 70$ ,即输出参数 1 400 个。

采用 6 个头,编码器和解码器各 6 层,由于视电阻率拟断面图像切割与模型网格所设置的矩阵大小略有不同,需要设置一个编码器到解码器的转换层,这里采用两个全连接层将编码器输入矩阵转换为与解码器同样维度的矩阵,与 FFN 不一样的是此处使用的激活函数为 Tanh。

将 9 240 组使用随机种子打乱顺序后,选取 20% 样本作为测试集,使用批量样本训练的方式, batch size 为 300,共进行 200 个 epochs,训练所用时间 15min25s,绘制网络训练误差收敛曲线图,从图 10 可知该网络具有较好的稳定性。

为了解该网络的泛化能力,了解该网络在大地电磁中使用效果,用未知的矩阵异常体正演得到的 TE 与 TM 模式下视电阻率拟断面。所选用的高阻模型异常体为  $1\ 000\ \Omega \cdot \text{m}$ ,背景场电阻率为  $100\ \Omega \cdot \text{m}$ ;低阻模型异常体为  $1\ \Omega \cdot \text{m}$ ,背景场电阻率为  $100\ \Omega \cdot \text{m}$ 。电阻率模型测试中所用的低阻异常体和高阻异常体的 TE 模式和 TM 模式下的视电阻率拟断面图如图 11 和图 12 所示,完成一次预测仅需要几毫秒,远远快于传统的反演方法。

图 13 所示是高阻异常体和低阻异常体电阻模

型,以及其神经网络的反演成像结果,用红色框在反演成像结果示意图中标注出对应的理论异常体位置。通过对比,①无论是高阻异常体模型还是低阻体模型反演出来的视电阻率模型,与所对应的原始模型设置的异常部分稍微有些许差异,在可容许的范围之类,能够较为准确地反映出异常体在地下赋存的情况位置和电阻率情况,只是在边界处存在偏差;②背景场整体基本一致,部分存在一定的误差,所差在 0.1 左右。高阻异常体模型电阻率和反演成像视电阻率均方误差为 0.025 7,低阻异常体模型电阻率和预测电阻率均方误差为 0.030 3。综合损失函数曲线、拟合模型对比分析以及反演成像模型的均方误差,该方法能较为准确地反映出异常体的位置分布与异常体电阻信息,体现了较好的准确性。

上述对比进一步验证网络的泛化能力,新增了平行四边形异常体和双矩阵异常体,相比于上述的高阻异常体和低阻异常体,网络对这两种模型更为陌生。图 14(c)为异常体模型展示,图 14(b)、图 14(d)是反演成像结果,可以看到双异常体的预测较为准确,可能是因为训练采用的是正方形作为训练样本;而平行四边形异常体具有一定误差,从反演成像结果来看,在实际异常体范围存在明显的异常,呈类似的倾斜高阻异常,准确度较为一般,背景场较为准确,基本符合实际情况。

### 3 Transformer 网络讨论与分析

#### 3.1 编码器和解码器的选择

为分析研究该网络的训练效果和训练效率,以及了解编码器和解码器对网络的影响,控制 Head 不变,选用 6 个 Head,讨论不同层解码器和编码器的组合方式,对样本进行多次训练,取得平均值, Epoch 和 Batch\_size 均为 200 和 300,训练样本 9 240 组,分别得到训练的时间和最终的均方误差值,以及对未知样本测试的均方误差值,笔者将 10 次测试结果绘制成图。

由图 15 可知,该网络训练速度较快,较为复杂的 6 层组合方式,训练 200 轮仅耗时不到 16 min,而相对简单的 4 层结构只需约 10 min。

由图 16~图 18 可知,多层耗时更多,但训练误差和测试误差来看更为稳定。对于少层的网络结构,对异常体电阻率的预测效果不够稳定,异常体分布位置预测存在偏差,少层网络泛化能力较弱。

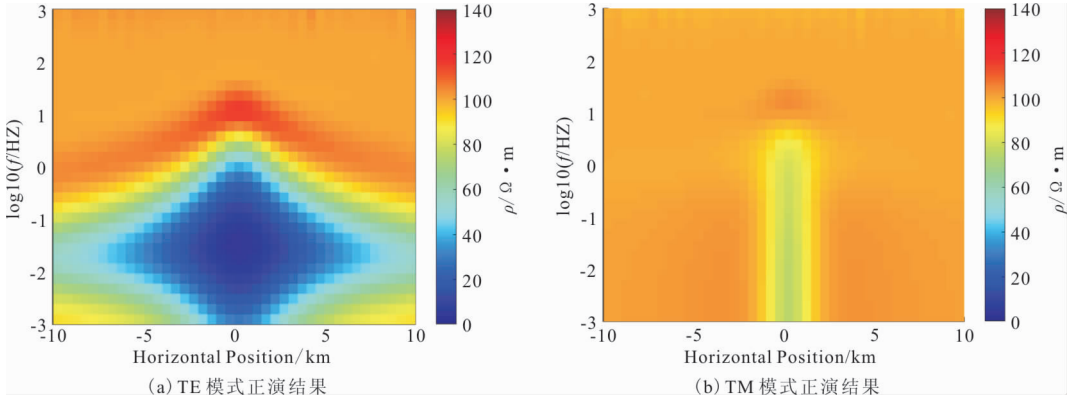


图 11 低阻异常体的 TE 和 TM 模式的正演电阻率拟断面图

Fig. 11 Forward resistivity pseudo section of TE and TM models of low resistivity anomaly

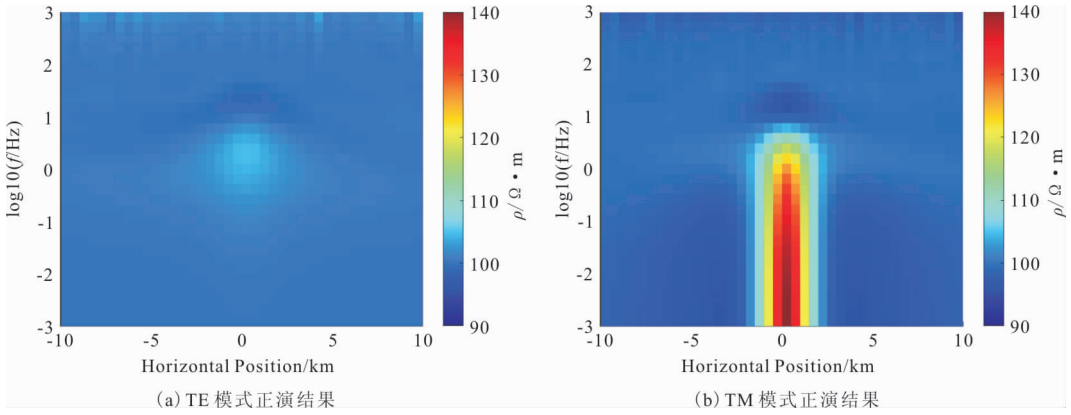


图 12 高阻异常体的 TE 和 TM 模式的正演电阻率拟断面图

Fig. 12 Forward resistivity pseudo section of TE and TM models of high resistivity anomaly

### 3.2 多头注意力机制与位置编码

为研究影响网络泛化能力的因素,假设视电阻率拟断面图分割重组后的矩阵  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_n]^T$ , 即  $\mathbf{A}$  由一系列图像转换而来的向量组成, 每个向量又可由一系列图像特征值表示, 即  $\mathbf{A}_1 = a_1^1, a_1^2, a_1^3, \dots, a_1^m, \mathbf{A}_2 = a_2^1, a_2^2, a_2^3, \dots, a_2^m$ 。

在多头注意力中:

$$\mathbf{Q} = \mathbf{A} * \mathbf{W}_Q \quad (17)$$

$$\mathbf{K} = \mathbf{A} * \mathbf{W}_K \quad (18)$$

式中:  $\mathbf{W}_Q, \mathbf{W}_K$  为权重矩阵:

$$\begin{cases} \mathbf{W}_Q = \mathbf{W}_Q^1, \mathbf{W}_Q^2, \dots, \mathbf{W}_Q^Z \\ \mathbf{W}_K = \mathbf{W}_K^1, \mathbf{W}_K^2, \dots, \mathbf{W}_K^Z \end{cases} \quad (19)$$

$\mathbf{Q}, \mathbf{K}$  也可以表示为式(20)。

$$\begin{cases} \mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3, \dots, \mathbf{Q}_n]^T \\ \mathbf{K} = [\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \dots, \mathbf{K}_n]^T \end{cases} \quad (20)$$

将式(19)带入式(17), 在转为式(20)。

$$\begin{cases} \mathbf{Q}_1 = [\mathbf{A}_1^T * \mathbf{W}_Q^1, \mathbf{A}_1^T * \mathbf{W}_Q^2, \dots, \mathbf{A}_1^T * \mathbf{W}_Q^Z] \\ \mathbf{Q}_2 = [\mathbf{A}_2^T * \mathbf{W}_Q^1, \mathbf{A}_2^T * \mathbf{W}_Q^2, \dots, \mathbf{A}_2^T * \mathbf{W}_Q^Z] \\ \dots\dots \\ \mathbf{Q}_n = [\mathbf{A}_n^T * \mathbf{W}_Q^1, \mathbf{A}_n^T * \mathbf{W}_Q^2, \dots, \mathbf{A}_n^T * \mathbf{W}_Q^Z] \end{cases} \quad (21)$$

同理:

$$\begin{cases} \mathbf{K}_1 = [\mathbf{A}_1^T * \mathbf{W}_K^1, \mathbf{A}_1^T * \mathbf{W}_K^2, \dots, \mathbf{A}_1^T * \mathbf{W}_K^Z] \\ \mathbf{K}_2 = [\mathbf{A}_2^T * \mathbf{W}_K^1, \mathbf{A}_2^T * \mathbf{W}_K^2, \dots, \mathbf{A}_2^T * \mathbf{W}_K^Z] \\ \dots\dots \\ \mathbf{K}_n = [\mathbf{A}_n^T * \mathbf{W}_K^1, \mathbf{A}_n^T * \mathbf{W}_K^2, \dots, \mathbf{A}_n^T * \mathbf{W}_K^Z] \end{cases} \quad (22)$$

注意力权重因子  $\mathbf{W}_{qk} = \mathbf{Q} * \mathbf{K}^T$ , 则  $\mathbf{W}_{qk}$  为  $n * n$  的矩阵, 同样可以表示成为由向量组成的表示  $\mathbf{W}_{qk} = [\mathbf{W}_{qk}^1, \mathbf{W}_{qk}^2, \dots, \mathbf{W}_{qk}^n]^T$ , 选取任意向量分析, 均为  $\mathbf{Z}$  维向量, 例如  $\mathbf{W}_{qk}^1 = [\omega_1^1, \omega_1^2, \dots, \omega_1^n]$ , 于是  $\mathbf{W}_{qk}^1$  中的向量可表示为:

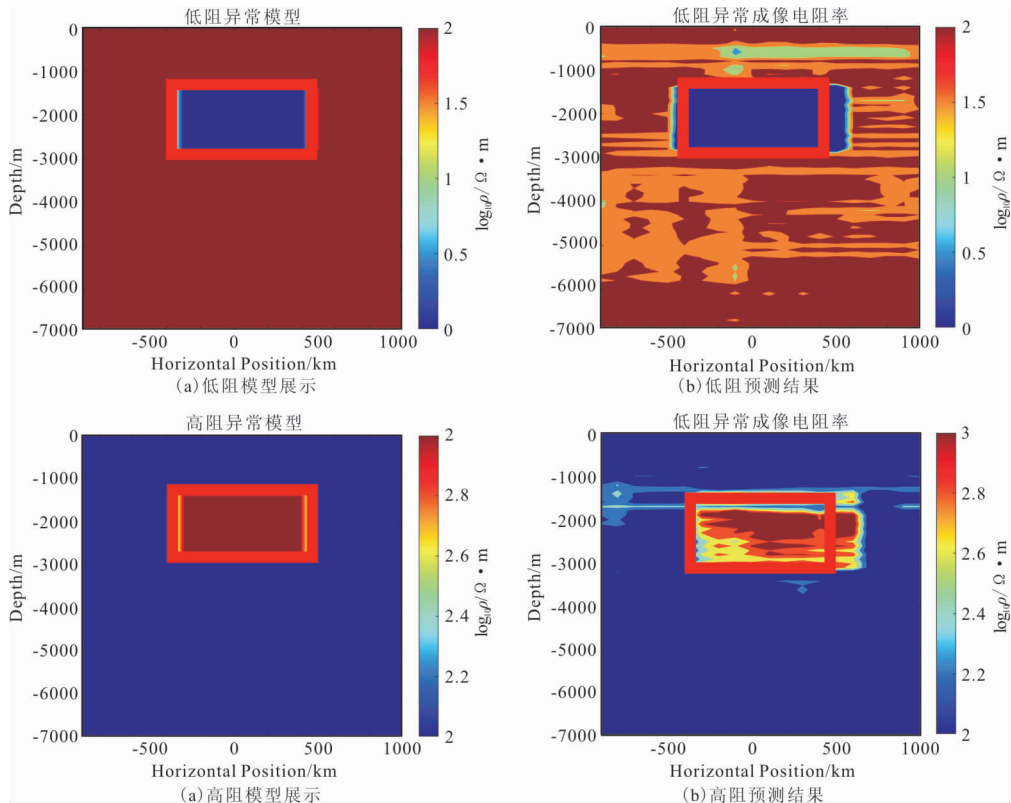


图 13 单一矩阵异常体电阻率模型展示和反演成像结果

Fig. 13 Display of Single matrix abnormal volume resistivity model and inversion imaging results

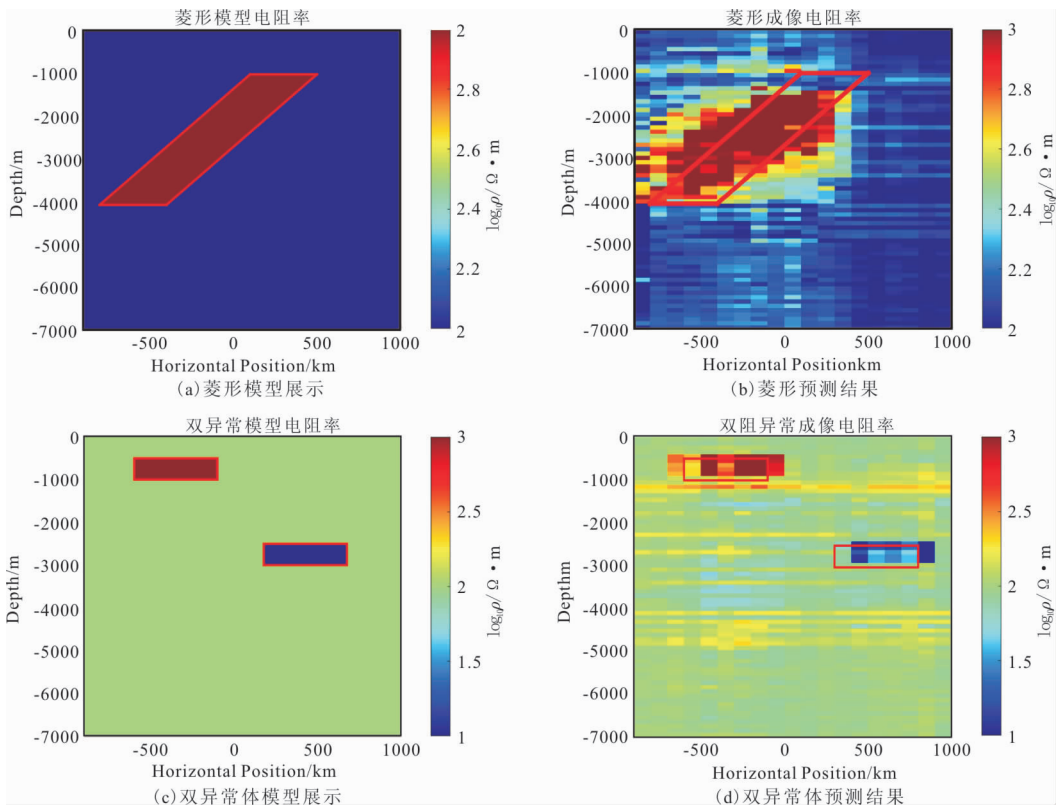


图 14 复杂异常体电阻率模型展示和反演成像结果

Fig. 14 Complex display of abnormal volume resistivity model and inversion imaging results

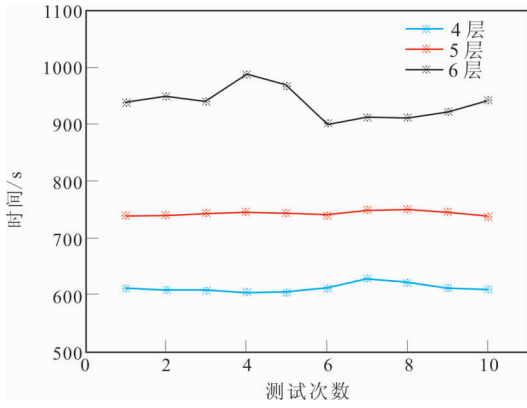


图 15 三种的组合方式的训练时间

Fig. 15 Training time of three combination modes

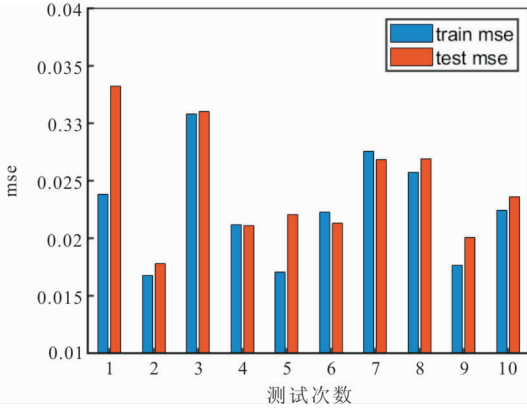


图 16 4层组合的训练误差和测试误差

Fig. 16 Training error and test error of 4-layer combination

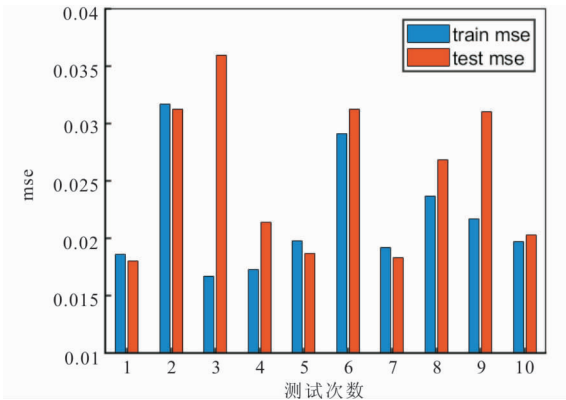


图 17 5层组合的训练误差和测试误差

Fig. 17 Training error and test error of 5-layer combination

$$\begin{cases} \omega_1^1 = \sum_{i=1}^Z \mathbf{A}_1^T * \mathbf{W}_Q^i * \mathbf{A}_1^T * \mathbf{W}_K^i \\ \omega_1^2 = \sum_{i=1}^Z \mathbf{A}_1^T * \mathbf{W}_Q^i * \mathbf{A}_2^T * \mathbf{W}_K^i \\ \dots \\ \omega_1^n = \sum_{i=1}^Z \mathbf{A}_1^T * \mathbf{W}_Q^i * \mathbf{A}_n^T * \mathbf{W}_K^i \end{cases} \quad (23)$$

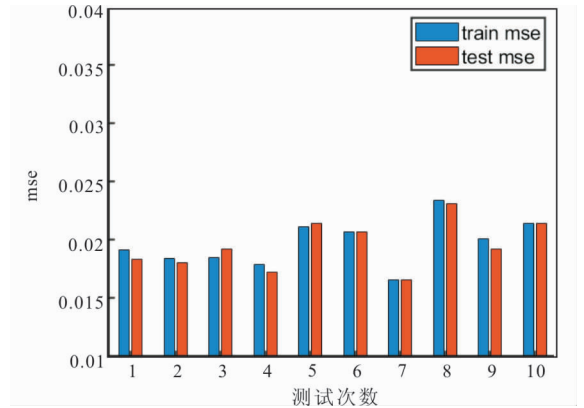


图 18 6层组合的训练误差和测试误差

Fig. 18 training error and test error of 6-layer combination

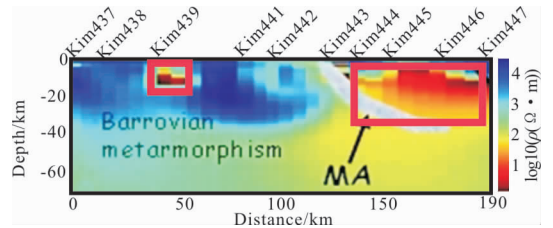


图 19 NLGC 反演结果

Fig. 19 NLGC inversion results

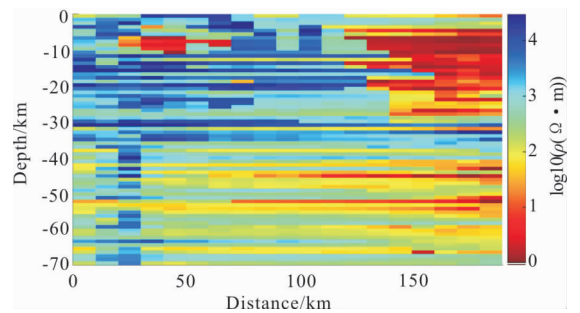


图 20 Transformer 反演成像色块图

Fig. 20 Transformer inversion imaging color block diagram

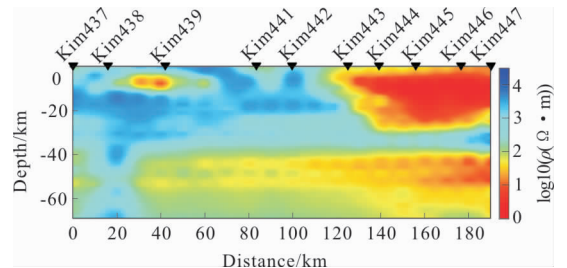


图 21 Transformer 反演成像结果图

Fig. 21 Transformer inversion image diagram

从式(23)中可以看出  $w_1^i, w_2^i, \dots, w_n^i$  值主要由所对应的  $\mathbf{A}_i (1 \leq i \leq n)$  向量决定。而  $\mathbf{A}_i = a_i^1, a_i^2, a_i^3, \dots, a_i^m$ , 在模型正演中, 对于任意的  $\mathbf{A}_i$ , 在远离异常体区域  $a_i^1, a_i^2, a_i^3, \dots, a_i^m$  均近似相等, 而在受到异常体影响的区域,  $a_i^1, a_i^2, a_i^3, \dots, a_i^m$  数值两极分化, 一部分不受异常体影响或者影响较小, 数值接近模型背景场电阻率, 另一部分数值因异常体影响会大于或者小于背景电阻率值。因此, 较大值或者较小值会集中在受到异常体影响的区域  $\tau$ , 同理推出  $\mathbf{W}_{\text{opt}}$  矩阵的较大值或者较小值也会集中在受到异常体影响的区域。然而注意力机制需要全局注意, 而不能只是局限于某个区域中, 于是笔者采用了向量投影的思路, 使得注意力的关注点更为分散, 更为合理的注意力机制将有效的提高网络的效果。

引入位置编码, 需要对输入的矩阵  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_n]^T$ , 添加维度一样的位置矩阵  $\mathbf{PE}$ 。经过多次测试, 发现未添加位置编码的均方误差比添加大于 0.004, 甚至会导致均方误差无穷。而在 Positional Encoding 中采用的是正弦函数和余弦函数, 则矩阵中每个位置大小的范围在  $(-1, 1)$  之间; 大地电磁正演中所选用的电阻率均远远大于 1, 会覆盖掉位置信息, 需对输入矩阵先归一化; 常用的归一化处理, Z-core 会使得异常体位置相同(异常体电阻率不同)的样本产生同样的结果; min-max 归一化使得数值不是趋近 0 就是趋近 1 的两极分化。于是, 合理位置编码和归一化方法对注意力机制有着很大的影响, 进而影响着网络的训练效果。笔者选用对数归一化, 将视电阻率归一在  $0 \sim 4$ , 满足归一化的同时又考虑位置编码信息。

## 4 实际数据

经过对比, 最终笔者选用了非洲南部实测大地电磁数据来检验该网络的训练效果, 数据来源于 MTNet 的开源大地电磁数据 SAMTEX ([http://www.mtnet.info/data/download\\_data.html](http://www.mtnet.info/data/download_data.html)), 并将反演成像结果与 Khoza 等<sup>[38]</sup> 采用正则化大地电磁反演的结果对比, 正则化反演剖面了用时 69 h, 如图 19 所示。笔者只选用了质量较为好的 Kim 测线, 一共有 10 个测点, 删除周期大于 3S 的频点, 设置深度设置为 70 km, 范围与图 15 红色区域对应, 反演成像色块图如图 20 所示。由于该测线剖面较为长, 测点间隔稀疏, 又预设网格设置比较少, 导致色块图看较为混乱, 在里程 20 km ~ 40 km, 120 km

~ 190 km 存在一大一小两个低阻体与图 19 红色框标注大体一致, 为了更好地展示 Transformer 网络的效果, 我们对色块图插值滤波成图, 见图 21。

将 Transformer 反演成像和 Khoza 等人采用正则化大地电磁反演的结果对比, 虽然存在一定差异, 但是对于低阻体异常位置反映大致相同, Transformer 预测在埋深 8 km ~ 10 km 里程 39 km 处存在较小的低阻, 里程 120 km ~ 190 km 存在较大的低阻范围, 均与正则化反演结果相似; 对于深部的 Transformer 表现稍微较差, 呈现出横向上的薄层现象, 还需要完善算法, 考虑加密设置的网络或者添加正则化等方法。从两种方法的用时来看, Transformer 仅需要 2.9 ms, 而传统反演根据剖面长度往往需要几分钟甚至更长时间。

## 5 结论

1) 笔者提出的基于 Transformer 网络的大地电磁反演成像能准确的预测处异常体的真实电阻率和范围大小, 验证了该方法的可行性和有效性。

2) 利用 adam 对 9 000 多组异常体训练网络, 训练后的网络体现了具有较好的泛化能力, 仅需几毫秒便能预测地下模型电阻率, 相比传统的反演方法更为快速。

3) Transformer 网络的多头注意力机制, 能高效的实现网络训练, 实现并行化处理数据, 即使是复杂组合的编码器(解码器)也能在极短的时间内对几千组数据进行训练。

4) 该方法采用矩阵并行计算, 在进行 MT 二维反演成像时, 需要合理的位置编码和归一化的处理, 建立各个位置之间的联系, 使得均方误差更小。

### 参考文献:

- [1] CONSTABLE S C, PARKER R L, CONSTABLE C G. Occam's inversion: a practical inversion algorithm for generating smooth models from electromagnetic sounding data[J]. Geophysics, 1987, 52(3): 289 - 300.
- [2] SIRIPUNVARAPORN W, EGBERT G. An efficient data-subspace inversion method for 2-D magnetotelluric data[J]. Geophysics, 2000, 65(3): 791 - 803.
- [3] SMITH J T, BOOKER J R. Rapid inversion of two- and three-dimensional magnetotelluric data[J]. Journal of Geophysical Research, 1991, 96(B3): 3905 -

- 3922.
- [4] 朱培民,王家映.地球物理资料非线性反演方法讲座(六)共轭梯度法[J].工程地球物理学报,2008(04):381-386.  
ZHU P M, WANG J Y. Lecture on nonlinear inversion methods of geophysical data (VI) conjugate gradient method[J]. Chinese Journal of Engineering Geophysics, 2008(04): 381-386. (In Chinese)
- [5] 柳建新,赵然,郭振威.电磁法在金属矿勘查中的研究进展[J].地球物理学进展,2019,34(1):0151-0160.  
LIU J X, ZHAO R, GUO Z W. Research progress of electromagnetic methods in the exploration of metal deposits[J]. Progress in Geophysics, 2019, 34(1): 0151-0160. (In Chinese)
- [6] 王家映.地球物理资料非线性反演方法讲座(二)蒙特卡洛法[J].工程地球物理学报,2007(02):81-85.  
WANG J Y. Lecture on nonlinear inversion methods of geophysical data (II) monte carlo method[J]. Chinese Journal of Engineering Geophysics, 2007(02): 81-85. (In Chinese)
- [7] 张荣峰.采用生物遗传算法的大地电磁测深资料反演[J].物探化探计算技术,1996(1):67-70.  
ZHANG R F. Inversion of magnetotelluric sounding data using biological genetic algorithm[J]. Computing Techniques for Geophysical and Geochemical Exploration, 1996(1): 67-70. (In Chinese)
- [8] SEND M K, BHATTACHARYA B B, STOFFA P L. Nonlinear inversion of resistivity sounding data[J]. Geophysics 1993, 58(4): 496-507.
- [9] 肖敏.二维大地电磁粒子群优化算法反演方法研究[D].武汉:中国地质大学,2010.  
XIAO M. Research on inversion method of two-dimensional magnetotelluric Particle swarm optimization algorithm[D]. Wuhan: China University Of Geosciences, 2010. (In Chinese)
- [10] SPICHAK V., POPOVA I. Artificial neural network inversion of magnetotelluric data in terms of three-dimensional earth macroparameters [J]. Geophysical Journal International, 2000, 142(1):15-26.
- [11] 师学明,王家映.层状介质大地电磁模拟退火反演法[J].地球科学,1998,23(5):542-546.  
SHI X M, WANG J Y. One dimensional magnetotelluric sounding inversion using simulated annealing[J]. Earth Science, 1998, 23(5): 542-546. (In Chinese)
- [12] 尹彬,胡祥云.非线性反演的贝叶斯方法研究综述[J].地球物理学进展,2016,31(3):1027-1032.  
YIN B, HU X Y. Overview of nonlinear inversion using Bayesian method [J]. Progress in Geophysics, 2016, 31(3): 1027-1032. (In Chinese)
- [13] 李创世,张彦鹏,李实,等.瞬变电磁勘探中的人工神经网络反演法[J].地球物理学进展,2001,35(6):604-615.  
LI C S, ZHANG Y P, LI S, et al. New inversion method of artificial neural network in transient electromagnetic inversion[J]. Journal of Xian Jiaotong University, 2001, 35(6): 604-615. (In Chinese)
- [14] 徐江.基于人工神经网络电性源半航空视电阻率反演方法研究[D].长春:吉林大学,2014.  
XU J. Apparent resistivity inversion research of Electrical source of Semiairborne transient electromagnetic based on neural network method[D]. Changchun: Jilin University, 2014. (In Chinese)
- [15] 胡祖志,胡祥云.大地电磁三维反演方法综述[J].地球物理学进展,2005,20(1):214-220.  
HU Z Z, HU X Y. Review of three dimensional magnetotelluric inversion methods [J]. Progress in Geophysics, 2005, 20(1): 214-220. (In Chinese)
- [16] 徐海浪,吴小平.电阻率二维神经网络反演[J].地球物理学报,2006,49(2):584-589.  
XU H L, WU X P. 2-D resistivity inversion using the neural network method[J]. Chinese Journal of Geophysics, 2006, 49(2): 584-589. (In Chinese)
- [17] 戴前伟,江沸波.基于混沌振荡 PSO-BP 算法的电阻率层析成像非线性反演[J].中国有色金属学报,2013,23(10):2897-2904.  
DAI Q W, JIANG F B. Nonlinear inversion for electrical resistivity tomography based on chaotic oscillation PSO-BP algorithm[J]. The Chinese Journal of Nonferrous Metals, 2013, 23(10): 2897-2904. (In Chinese)
- [18] 高明亮.直流电法地电模型参数辨识的非线性方法研究[D].长春:吉林大学,2016.  
GAO M L. Research of Nonlinear Parameter Identification for Geoelectric Model based on DC Resistivity Method[D]. Changchun: Jilin University, 2016. (In Chinese)
- [19] 王鹤,刘梦琳,席振铎,等.基于遗传神经网络的大地电磁反演[J].地球物理学报,2018,61(4):1563-1575.  
WANG H, LIU M L, XI Z Z, et al. Magnetotelluric

- inversion based on BP neural network optimized by genetic algorithm[J]. Chinese Journal of Geophys, 2018, 061(4): 1563—1575. (In Chinese)
- [20] 敬荣中. 地球物理非线性联合反演方法研究[D]. 长沙:中南大学, 2002.  
JING R Z. Study on geophysical non-linear integration inversion[D]. Changsha: Central South University, 2002. (In Chinese)
- [21] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436—444.
- [22] PUZYREV V. Deep learning electromagnetic inversion with convolutional neural networks [J]. Geophysical Journal International, 2019, 218(2): 817—832.
- [23] LIU B, GUO Q, LI S, et al. Deep Learning Inversion of Electrical Resistivity Data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, PP (99): 1—14.
- [24] 廖晓龙, 张志厚, 范祥泰, 等. 基于改进粒子群深度神经网络的频率域航空电磁反演[J]. 中南大学学报(自然科学版), 2020, 51(8): 121—132.  
LIAO X L, ZHANG H H, FAN X T, et al. Frequency domain airborne EM inversion based on improved particle swarm depth neural network[J]. Journal of Central South University (Science and Technology), 2020, 51 (08): 121—132. (In Chinese)
- [25] XIAO H H, ZHANG D B, WANG W, et al. Non-detection text recognition of certificate image based on Transformer [J]. Information Technology (In Chinese), 2021, (6): 78—85, 90.
- [26] 王丹鹤. 基于深度学习的大地电磁正反演研究[D]. 桂林: 桂林理工大学, 2021.  
WANG D H. Forward and inversion of magnetotelluric based on deep learning[D]. Guilin: Guilin University of Technology, 2021. (In Chinese)
- [27] 韩晓冰, 柳庆瑜, 范梦宁, 等. 基于疏密交叉网格建模的KNN大地电磁反演方法[J]. 采矿技术, 2020, 20(3): 155—159.  
HAN X B, LIU Q Y, FAN M N, et al. KNN magnetotelluric inversion method based on dense cross grid modeling[J]. Mining Technology, 2020, 20(3): 155—159. (In Chinese)
- [28] 许滔滔, 王中兴, 肖卓伟, 等. 基于LSTM循环神经网络的大地电磁工频干扰压制[J]. 地球物理学进展, 2020, 35(5): 2016—2022.  
XU T T, WANG Z X, XIAO Z W, et al. Magnetotelluric power frequency interference suppression based on LSTM recurrent neural network [J]. Progress in Geophysics, 2020, 35(5): 2016—2022. (In Chinese)
- [29] 王昊, 严加永, 付光明, 等. 深度学习在地球物理中的应用现状与前景[J]. 地球物理学进展, 2020, 35(2): 0642—0655.  
WANG H, YAN J Y, FU G M, et al. Current status and application prospect of deep learning in geophysics [J]. Progress in Geophysics, 2020, 35(2): 0642—0655. (In Chinese)
- [30] 韩盈, 安志国, 屈文章. 基于机器学习的大地电磁时域数据处理的研究现状[J]. 地球物理学进展, 2021, 36(5): 1975—1987.  
HAN Y, AN Z G, QU W Z. Research status of magnetotelluric time domain data processing based on machine learning [J]. Progress in Geophysics, 2021, 36(5): 1975—1987. (In Chinese)
- [31] 汤井田, 任政勇, 周聪等. 浅部频率域电磁勘探方法综述[J]. 地球物理学报, 2015, 58(8): 2681—2705.  
TANG J T, REN Z Y, ZHOU C, et al. Frequency-domain electromagnetic methods for exploration of the shallow subsurface: A review [J]. Chinese J. Geophys, 2015, 58(8): 2681—2705. (In Chinese)
- [32] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] Proceedings of the NIPS, 2017: 5998—6008.
- [33] PARMAR N, VASWANI A, USZKOREIT J, et al. Image transformer [C]. International Conference on Machine Learning. PMLR, 2018: 4055—4064.
- [34] 周哲韬, 刘路, 宋晓, 等. 基于Transformer模型的滚动轴承剩余使用寿命预测方法[J/OL]. 北京航空航天大学学报: 1—17.  
ZHOU Z T, LIU L, SONG X, et al. Remaining useful life prediction method of rolling bearing based on Transformer Model [J/OL]. Journal of Beijing University of Aeronautics and Astronautics: 1—17. (In Chinese)
- [35] 何林远, 白俊强, 贺旭, 等. 基于稀疏Transformer的遥感旋转目标检测[J/OL]. 激光与光电子学进展: 1—17.  
HE L Y, BAI J Q, HE X, et al. Remote sensing rotated object detection based on sparse Transformer [J/OL]. Laser & Optoelectronics Progress: 1—17. (In Chinese)
- [36] HE, K., ZHANG, X., REN, S, et al. Deep residu-

al learning for image recognition[C]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770—778.

- [37] SHEWCHUK J R. Delaunay refinement algorithms for triangular mesh generation[J]. Computational Geometry, 2002, (22): 21—74.
- [38] KHOZA T D, JONES A G, MULLER M R, et al.

Lithospheric structure of an Archean craton and adjacent mobile belt revealed from 2-D and 3-D inversion of magnetotelluric data; example from southern Congo craton in northern Namibia[J]. Journal of Geophysical Research Solid Earth, 2013, 118(8): 4378—4397.

## Application of transformer network in magnetotelluric inversion imaging

LIU Gaocun, WANG Xuben, YUAN Chongxin, LI Dewei, XIE Zhuoliang

(The Ministry of Education Key Laboratory of Earth Exploration And Information Technology  
( Chengdu University of Technology ), Chengdu 610059, China)

**Abstract:** The traditional magnetotelluric inversion is usually based on the iterative solution of the deterministic gradient. It not only needs much time to calculate the Jacobian matrix but also depends on the input of the initial model and the setting of regularization factors and other parameters, which has a strong multi-solution. In recent years, scholars have continued introducing machine learning methods to improve magnetotelluric inversion's shortcomings. In this paper, the classical framework of the transformer neural network proposed by the Google team is used to build the mapping network between magnetotelluric data and the model, 9240 groups of forward data are used as samples to train the parameters of the transformer network, and the South African open source magnetotelluric data is used to realize the rapid imaging from apparent resistivity image to resistivity model. The research shows that the trained transformer network can accurately reflect the location and size of the abnormal body; The network realizes simple matrix parallel operation, dramatically improves the training efficiency, and the imaging efficiency is higher than the traditional inversion.

**Keywords:** MT; finite element forward modeling; transformer; Multi-Head attention; inversion imaging