引文格式:虞欣,于家钰,郑肇葆,等.遥感影像场景识别的贝叶斯共轭批次归一化方法[J].武汉大学学报(信息科学版),2025,50(9):1838-1847.DOI:10.13203/j.whugis20220632

Citation: YU Xin, YU Jiayu, ZHENG Zhaobao, et al. Scene Recognition of Remotely Sensed Images Based on Bayes Adjoint Batch Normalization[J]. Geomatics and Information Science of Wuhan University, 2025, 50(9): 1838-1847. DOI: 10.13203/j.whugis20220632

遥感影像场景识别的贝叶斯共轭批次归一化方法

虞 欣[1] 于家钰[1] 郑肇葆[2] 孟令奎[2] 李林宜[2]

1 北京石油化工学院人工智能研究院,北京,102617

2 武汉大学遥感信息工程学院,湖北 武汉,430079

摘 要:归一化方法作为特征预处理的关键部分,在浅学习和深度学习中都是至关重要的。针对批次归一化方法中存在对批次样本容量依赖较大的问题,当前的优化思路主要是从样本信息的其他维度(如通道、层、时间等)来弥补批次样本容量较小的不足。从贝叶斯理论的角度出发,通过将总体信息、先验信息和样本信息以科学、合理的融合方式来弥补批次样本容量不足的缺陷,从而可以更加准确地估计样本均值和样本方差,使得归一化后的特征落入最佳的非饱和区域,以便更好地反映整个特征空间的原始表征,进而深度学习模型可以达到最佳的特征表达能力。实验与分析表明,所提的贝叶斯共轭批次归一化方法可行且有效,在NWPU-RESISC45数据集上,其分类精度比批次归一化方法高5.64%。得益于总体信息和先验信息,所提方法受样本容量的影响较小。

关键词:场景识别;批次归一化;共轭;贝叶斯

中图分类号:P237 **文献标识码**:A **收稿日期**:2024-06-21

DOI:10.13203/j.whugis20220632 **文章编号**:1671-8860(2025)09-1838-10

Scene Recognition of Remotely Sensed Images Based on Bayes Adjoint Batch Normalization

YU Xin[1] *YU Jiayu*[1] *ZHENG Zhaobao*[2] *MENG Lingkui*[2] *LI Linyi*[2]

1 Academy of Artificial Intelligence, Beijing Institute of Petrochemical Technology, Beijing 102617, China

2 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

Abstract: **Objective**: Normalization method plays an important role in feature preprocessing phase not only in conventional machine learning domain but also in contemporary deep learning domain. Batch normalization (BN) is very successful, but its performance very depends on the sample size. Therefore, many researchers try to improve it when the sample size is inadequate through adding the sample size merely in the sample information space. **Methods**: This paper utilizes Bayes theory to integrate general information, prior information and sample information, and to offset the inadequate sample information. In this way, the mean value and the variance of sample can be estimated more precisely and more robust especially when the sample size is small, and the normalized feature will better fall into non-saturating region of activation function, which enables deep learning model to better describe original feature space. **Results**: The Top-1 test classification accuracy in NWPU-RESISC45 dataset is improved by 5.64% than that of BN method. Moreover, with the help of general information and prior information, the Bayes adjoint batch normalization (BABN) method is not sensitive to the sample size. **Conclusions**: The experiment results show that the proposed BABN method is feasible and effective, and performs better than BN method and other variants in the scene recognition of remote sensing image.

Key words: scene recognition; batch normalization; adjoint; Bayes

基金项目:国家重点研发计划(2021YFB3900603);北京市科技新星计划(20250484800)。

第一作者:虞欣,博士,教授,主要研究方向为影像解译、人工智能和贝叶斯统计等。china_yuxin@163.com

通信作者:李林宜,博士,副教授。 lilinyi@whu.edu.cn

遥感影像分析与理解一直是遥感领域中研究的热点和难点[1-2]。在2012年之前,随机森林法、贝叶斯网络、马尔可夫链、支持向量机等传统的机器学习方法受到学者们较多关注[3-4]。由于这些方法都是事先人工设计特征,且特征提取与分类过程是相对独立的两个部分,通常又被称为浅学习。在2012年ImageNet大规模视觉识别挑战赛中,深度学习方法比浅学习方法高出约10%的分类精度,成为计算机视觉领域中的里程碑事件[5-6]。之后国内外掀起了深度学习的研究热潮,逐渐成为遥感领域学者们的关注热点。目前,第三代人工智能技术的兴起为遥感影像智能解译提供了方向,即将知识推理和数据学习结合起来,建立一个以数据自主学习为核心、以领域先验知识推理为引导的新一代遥感影像解译范式[7]。

归一化方法是一种简化计算的方式,将有量纲的表达式进行变换,化为无量纲的表达式,成为标量。常见的方法有批次归一化(batch normalization, BN)、层归一化(layer normalization, LN)、组归一化(group normalization, GN)和实例归一化(instance normalization, IN)等。无论是在浅学习,还是在当今流行的深度学习中,归一化方法在特征预处理的过程中都至关重要[8]。2012年,Google在AlexNet中提出局部响应归一化方法[9],其创意受到认知神经科学中的侧抑制现象[10-11]的启发,即由于相邻神经元的活动导致某个神经元活动的降低。在浅学习中,对输入的特征进行标准化处理可有效提高模型的精度。受此启发,既然在网络的输入层可以加入归一化,那么在网络的隐层为何不可以加入标准化处理呢?于是,笔者对每层进行归一化(包括加权和标准化),然后通过缩放平移来适度还原,做到了既不过分破坏输入信息,又抑制了各批次之间位置点像素分布的剧烈变化带来的学习难度。实际上,在深度学习中,各卷积层中的权重和偏置的更新会导致网络输出数据的分布发生变化,随着模型深度的增加,依据链式法则,数据分布的偏移现象会随着层数的增加而逐渐被放大,这种现象被称为内部协方差偏移(internal covariance shift, ICS)。于是,BN[12]因其能够有效提升网络训练效率而获得广泛应用与持续研究,成为目前主流的一种归一化方法,为深度学习模型的训练提供了一种有效的解决方案。

然而,BN方法对每个批次的样本容量(batch size, BS)比较敏感。在BS较小的条件下,其性能常被打破[12-14],这是因为此时带有一定的噪声,使得估计的该批次的样本均值和样本方差缺乏一定的鲁棒性,导致性能急剧下降。此外,深度学习因其模型的复杂度需要大量的计算,通常需要利用GPU或CPU进行并行计算来加快模型的学习[13]。考虑到硬件方面(主要是显存)的限制,在利用GPU进行并行计算时,每个批次的BS受到显存大小的限制[14]。因而,每个批次的BS较小会阻碍BN在耗费显存任务中的应用,如目标检测[15]、语义分割[16]和动作识别[17]等。

针对每个批次BS较小的问题,学者们提出了从批次样本大小以外的其他维度,如通道、层、时间等来间接弥补这个缺陷的一系列改良方法,如IN[18]、LN[19]、GN[14]、可切换归一化(switchable normalization, SN)[20]、位置归一化(positional normalization, PN)[21]、交叉迭代批量归一化(cross-iteration batch normalization, CBN)[22]和滤波器响应归一化(filter response normalization, FRN)[23]等。深度学习因其模型的复杂度,往往需要借助大量的训练样本进行模型的学习。一方面,这些方法忽略了整个训练样本集BS较大的优势,另一方面,因其他维度信息与样本容量信息在物理属性等方面存在明显的差异,所以用其他维度来间接弥补每个批次的BS较小的方法只是一种近似方法,所估计的样本均值和样本方差与总体均值和总体方差必然存在一定的差距。因而,ICS还存在进一步改善的空间。本文基于整个训练集具有大BS的优势,借助贝叶斯统计理论,充分发挥训练样本集中大样本的总体信息和先验信息,直接弥补BS较小的不足。所提方法既避免了BS所携带的统计信息与其他维度所携带统计信息在物理属性等方面的差异,又可以更加准确地估计BS较小条件下的样本均值和方差,使得在贝叶斯准则下最大限度地接近样本的总体均值和总体方差,从而在原有BN的基础上进一步减少ICS,使得深度学习模型能更好地表征原始特征空间,进而可以获得更高的模型精度。

随着深度学习模型的发展,模型的规模越来越大,模型的深度越来越深,这使得训练深度神经网络也变得越来越复杂,从而更加凸显归一化方法的作用和地位[24-25]。BN的缺陷如下:(1)依赖于每个BS大小,当批次的BS较小时,估计的样本均值和样本方差与总体均值和总体方差会

存在较大的偏差;(2)对于循环神经网络(recurrent neural network, RNN),其动态效果不明显,且当推理序列长度超过训练样本的长度时更容易出现问题;(3)当批次中的样本不满足独立分布时,其性能会显著下降。为了改善BS较小时的统计估计问题,国内外学者们针对不同任务提出了一系列改良的思路和方法[26-41]。LN方法[19]的思想与BN类似,只是BN在每个神经元对一个批次的样本进行归一化,而LN则在每一层对单个样本的所有神经元节点进行归一化,即在通道、宽度和高度这3个维度求均值和方差,并进行归一化,而与批次中BS大小无关。LN方法更适合处理不定长序列数据的神经网(如RNN),且可以更快、更好地帮助模型达到最优效果。针对图像风格化任务,IN方法[18]考虑到生成的风格结果主要依赖于某个图像实例本身,而与整个批次中的其他样本图像无关,所以在这个任务中BN是不合适的,仅在图像的宽度和高度两个维度进行归一化。LN虽然不依赖BS,但是在图像领域里使用的卷积神经网络中,直接对当前层所有通道数据进行归一化并不合适,这是因为各个通道获取的信息实际上是不同的电磁波段所获取的信息,它们之间存在物理属性上的根本差异,不同的通道之间也就必然存在物理信息的不一致性。而GN方法[14]先将通道分为许多组,再对每一组做归一化。这实际上是在物理信息不一致性与批次BS之间的一种平衡选择,弱化了LN中物理信息不一致性产生的影响,同时也弥补了BS过小的不足。PN方法[21]是一种跨通道的归一化方法,认为BN、IN和GN等方法在空间层面归一化信息的同时丢弃了重要的统计信息,如果有效利用统计信息,可以提高生成式对抗网络(generative adversarial network, GAN)和分类网络的性能。CBN方法[22]从时间维度对批次的BS进行弥补,即通过上一批次中估计的样本均值和方差来计算更新当前批次的样本均值和方差,从而实现间接扩大BS(即弥补小BS的不足),进而改善网络性能。由于梯度迭代的属性,网络权重在多轮的批次计算过程中缓慢改变,所以根据泰勒展开式可以近似估计出相邻批次的样本均值和方差。2021年德克萨斯大学提出的夹心归一化方法考虑到训练数据集通常较大,且每个类别变化也很大,不同类别的特征理应对应不同的均值和方差,其归一化、缩放、偏置也应该不同,若单独拆分特征层,会导致训练不平衡,降低收敛速度,阻碍整个网络训练。该方法在BN的基础上加入分类的条件信息,对BN后的每个特征层通过多层感知机进行仿射变换。不同于BN,其变换因子是由输入特征决定的,由训练样本学习得到。

总体而言,BN方法通常可以在BS较大时取得良好的性能。然而在BS较小时的性能会急剧下降。GN方法在不同的BS下具有较大的稳定性,而GN在BS较大时性能略差于BN。不同的归一化方法适用于不同的任务。如IN方法在图像风格迁移中表现较好,LN方法在RNN中表现较好,PN方法在生成网络中表现较好,跨批次归一化方法在目标检测任务中较好,但上述方法在其他视觉任务中泛化性能较差。

深度学习模型通常需要大量的训练样本对模型进行学习,而上述方法都忽略了整个训练样本集具有大BS的优势。此外,上述方法都是从独立于BS的其他维度,如通道、层、时间等来间接弥补BS较小的不足。然而,不同维度所携带的信息在物理属性方面存在着差异,将BS维度和其他维度的信息融合起来估计样本均值和方差,势必存在不同维度之间信息不一致的问题。虽然从数量角度来看,其他维度的信息在一定程度上弥补了BS的不足,但是两个不同维度信息的不一致性也在一定程度上影响了样本均值和样本方差估计的鲁棒性,本文通过贝叶斯共轭批次归一化方法(Bayes adjoint batch normalization, BABN)在一定程度上缓解这一问题。

1 BABN基本原理

1.1 归一化基本原理

目前,常见的归一化方法的计算式为:

$$\hat{x}_i=\frac{1}{\sigma_i}(x_i-\mu_i) \tag{1}$$

式中,x表示某一层中的特征;i为下标索引;μ和σ分别表示样本的均值和标准差。对于二维的图像,$i=i_N\times i_C\times i_H\times i_W$,用来表示在$N\times C\times H\times W$中的一个特征索引,其中$N$表示批次样本的维度,$C$表示通道(或波段)的维度,$H$和$W$分别表示图像空间的高度维度和宽度维度。

样本均值和标准差的计算式分别为:

$$\mu_i=\frac{1}{m}\sum_{k\in S_i}x_k \tag{2}$$

$$\sigma_i=\sqrt{\frac{1}{m}\sum_{k\in S_i}(x_k-\mu_i)^2+\varepsilon} \tag{3}$$

式中,ε表示一个很小的常数,通常设置为1×10^{-5};

S_i表示在计算样本均值和标准差时的像素集合;m表示该集合的势。对于不同类型的归一化方法,其主要差异在于计算样本均值和标准差的集合。

对于BN方法[12],集合S_i定义为:

$$S_i=\{k|i_C=k_C\} \tag{4}$$

式中,i_C和k_C分别表示沿着C轴的索引i和k在波段维度上的下标索引,以下表示类似。对具有相同波段索引的像素进行归一化,即BN方法对每个波段计算样本均值和方差($N\times H\times W$)。

对于LN方法[19],集合S_i定义为:

$$S_i=\{k|k_N=i_N\} \tag{5}$$

即LN方法是对每个样本计算样本均值和样本方差($C\times H\times W$),与批次维度N无关。

对于IN方法[18],集合S_i定义为:

$$S_i=\{k|k_N=i_N,k_C=i_C\} \tag{6}$$

即IN方法对每个样本和每个波段计算样本均值和方差($H\times W$)。

BN、LN和IN均学习了一个逐波段的线性变换,以补偿特征表示时可能的信息损失。线性变换为$y_i=\gamma\hat{x}_i+\beta$,其中$\gamma$和$\beta$是可训练的尺度和平移方面的参数。

对于GN方法[14],集合S_i定义为:

$$S_i=\left\{k|k_N=i_N,\left\lfloor\frac{k_C}{C/G}\right\rfloor=\left\lfloor\frac{i_C}{C/G}\right\rfloor\right\} \tag{7}$$

式中,G表示分组数量,是预定义的超参数;C/G表示每个组内的波段数;$\lfloor\ \rfloor$为取单精度浮点数操作;$\left\lfloor\frac{k_C}{C/G}\right\rfloor=\left\lfloor\frac{i_C}{C/G}\right\rfloor$表示索引$i$和$k$是相同的波段分组。GN方法对每个组内的$C/G$个波段数计算样本均值和方差(($C/G$)$\times H\times W$)。

1.2　贝叶斯共轭方法

前文提及的BN、LN、IN和GN方法估计样本均值和方差的主要差别在于所采用的样本空间不同,适用于不同的任务。从贝叶斯角度来看,这些方法的共同缺陷在于仅采用了样本信息,而忽略了总体信息和先验信息,从而使得估计的样本均值和方差不够准确、鲁棒,进而在一定程度上制约了归一化的实际效果。

采用密度函数形式,贝叶斯方法的计算式为:

$$\pi(\theta|x)=\frac{h(x,\theta)}{m(x)}=\frac{p(x|\theta)\pi(\theta)}{\int p(x|\theta)\pi(\theta)\mathrm{d}\theta} \tag{8}$$

式中,$m(x)$表示x的边缘密度函数;$\pi(\theta)$表示参数θ的先验密度函数,反映了在观测数据之前对参数θ的概率分布假设;$p(x|\theta)$表示在已知θ的条件下关于x的条件概率密度函数,即似然函数,描述了在不同参数θ下观测到数据x的概率密度;$h(x,\theta)$则是θ和x的联合概率密度函数,综合了先验和似然的信息。然而在样本x给定的条件下,θ的条件分布被称为θ的后验分布,记作$\pi(\theta|x)$,它集中体现了总体、样本和先验3种信息中与θ相关的所有内容,排除了无关信息,是进行统计推断的重要依据,基于此分布进行推断通常更为有效且合理。

在实际应用中,尤其是深度学习领域,训练样本的规模往往较大,通常达到上万甚至更多。假设样本的特征服从正态分布,在正态方差已知的条件下,其正态均值的共轭分布也是正态分布。假设现有一个完整的训练集,样本的每个特征服从正态分布$N(\mu_0,\sigma_0^2)$,其中μ_0和σ_0^2为总体均值和方差。另有一个批次的样本,BS为m,样本均值和方差分别为μ_s和σ_s^2,μ_1和σ_1^2为根据总体信息、样本信息和先验信息综合估计得到的最终的样本均值和方差,即后验均值和后验方差。按照贝叶斯方法对μ_1和σ_1^2进行估算,具体的推导过程参见文献[42],二者的计算式分别为:

$$\mu_1=\frac{\mu_s\sigma_s^{-2}+\mu_0\sigma_0^{-2}}{\sigma_0^{-2}+\sigma_s^{-2}}=\frac{\sigma_0^{-2}}{\sigma_0^{-2}+\sigma_s^{-2}}\mu_0+\frac{\sigma_s^{-2}}{\sigma_0^{-2}+\sigma_s^{-2}}\mu_s \tag{9}$$

$$\sigma_1^2=\frac{\sigma_0^2\sigma_s^2}{\sigma_0^2+\sigma_s^2} \tag{10}$$

实际上,方差的倒数发挥着重要作用,通常称其为精度。从式(9)来看,μ_1可以看作总体均值μ_0和样本均值μ_s的加权平均,并根据总体方差和样本方差的大小进行科学合理的融合。如果方差相对较大,方差的倒数就相对较小,所以其均值在最终估计中的权重较小,反之亦然。这表明后验均值是在先验均值与样本均值之间采取了一种科学、合理的折衷方案。从式(10)来看,后验分布的精度是样本均值分布的精度与先验分布精度之和,增加BS或减少先验分布方差都有利于提高后验分布的精度。这便从理论上说明当BS较小时,如何充分利用总体信息和先验信息来弥补BS较小时所携带样本信息的不足。

1.3　BABN方法

实际上,现有的BN方法在估算样本均值和方差时,其精度还存在一定的改善空间,特别是当批次的BS较小时,其估计值的精度明显降低。

而相比于LN、IN和GN方法等,虽然可以从其他角度对批次的BS进行弥补,但其效果的改善仍然存在一定的局限性。总的来说,这些方法仅仅利用了样本信息,却忽略了总体信息和先验信息。本文利用贝叶斯统计理论的优势,充分发挥总体信息和先验信息,将这两种信息与原有的样本信息进行了有机的融合,克服了BS较小的缺陷,并达到了较高的估计精度。在大样本容量的条件下,正态分布的均值和方差的共轭分布也是正态分布,则其样本均值和方差的估算公式分别同式(9)和式(10)。

对于BABN方法,集合S_i定义为:

$$S_i=\{k|k_M=i_M\} \tag{13}$$

为区别于批次样本N,本文用M来表示总体。类似于BN方法,BABN对于每个波段计算样本均值和方差($(M+N)\times H\times W$)。

由于深度学习模型在训练时通常数据量很大,样本容量少则上万,甚至上百万,如ImageNet数据集,其样本容量约为1.28 MB。因而,在现有的深度学习模型框架下,如PyTorch、Caffe、TensorFlow等,一般采用指数滑动平均法去进行近似计算,其优点是不占内存,计算方便。但是其结果必然不如使用整个训练集计算得到的样本均值和方差那么准确。然而,贝叶斯方法提供了一种可以分批进行计算整个训练集的均值和方差的方法,既不受限于内存的大小,又保证了估计的均值和方差的估算精度,在一定程度上也能够增加模型的鲁棒性,减少过拟合的可能性。

1.4 核心算法流程

由于PyTorch等深度学习框架提供了自动求解梯度的功能,BABN比较容易实现。实际上,BABN与BN方法的主要不同之处在于样本均值和方差的计算。输入总体均值μ_0和总体方差σ_0^2,某一批次的特征集合$B=\{X_1,X_2,\cdots,X_m\}$,其中m为该批次中的样本容量,可学习的尺度参数γ和平移参数β;输出$\{y_i=f_{\gamma\beta}(x_i)\}$。具体的算法流程如下:

1)依据该批次的样本和式(2),计算得到第i个特征的样本均值μ_{si};

2)依据该批次的样本和式(3),计算得到第i个特征的样本方差σ_{si}^2;

3)根据总体均值、总体方差、样本均值和样本方差,依据式(9)计算得到该批次下第i个特征的最终样本均值μ_{1i};

4)根据总体均值、总体方差、样本均值和样本方差,依据式(10)计算得到该批次下第i个特征的最终样本方差σ_{1i}^2;

5)依据$\hat{x}_i=\dfrac{x_i-\mu_{1i}}{\sqrt{\sigma_{1i}^2+\varepsilon}}$计算得到贝叶斯正则化后的特征;

6)依据$y_i\leftarrow\gamma\hat{x}_i+\beta\equiv f_{\gamma\beta}(x_i)$,加入平移和缩放参数,得到最终的贝叶斯正则化结果;

7)重复上述步骤,依次对该批次下所有的特征进行贝叶斯正则化处理。

2 实验与分析

本文实验基于PyTorch深度学习框架,实验平台搭载Ubuntu 18.04.6 LTS操作系统,配置Intel(R)Core i5-10400F @2.90 GHz 6核12线程CPU处理器,32 GB RAM和NVDIA RTX 3060显卡,显存为12 GB。

2.1 数据集

近年来,遥感领域利用深度学习技术支持场景理解、地物目标检测与土地覆盖分类等任务。随着遥感影像样本数据集的不断增加,其多样性的特点也日益显现,在尺度、传感器、时相等方面存在较大差异[43-45]。本文实验使用NWPU-RESISC45[16]、AID[46]、EuroSAT[47]和SAT-6[48]4种公开的遥感场景分类数据集和一种公开的自然场景分类数据集Mini-ImageNet[49],对BABN方法进行验证和比较分析。5种数据集的部分样本示例如图1所示,基本情况对比见表1。

1)NWPU-RESISC45数据集是西北工业大学发布的一个用于遥感图像场景分类的大规模公开数据集。该数据集包含45个场景类别,每个类别有700张图像,总计31 500张图像,每张图像的分辨率为256×256像素。

2)AID数据集是一个大规模航拍图像数据集,来自Google Earth影像,由武汉大学和华中科技大学于2017年共同发布。该数据集包含30种航空场景类型,每个场景类别约有220~420张图像,总计约10 000张图像,每张图像的大小约为600×600像素。

3)EuroSAT数据集由哥白尼地球观测计划提供的Sentinel-2卫星拍摄,覆盖13个光谱波段,包含10类场景共计27 000张图像,图像大小为64×64像素,空间分辨率达到10~30 m。本文实验仅使用该数据集的可见光波段影像数据。

4)SAT-6数据集的图像是从美国国家农业影像计划数据集中提取的,该数据集由跨越美国

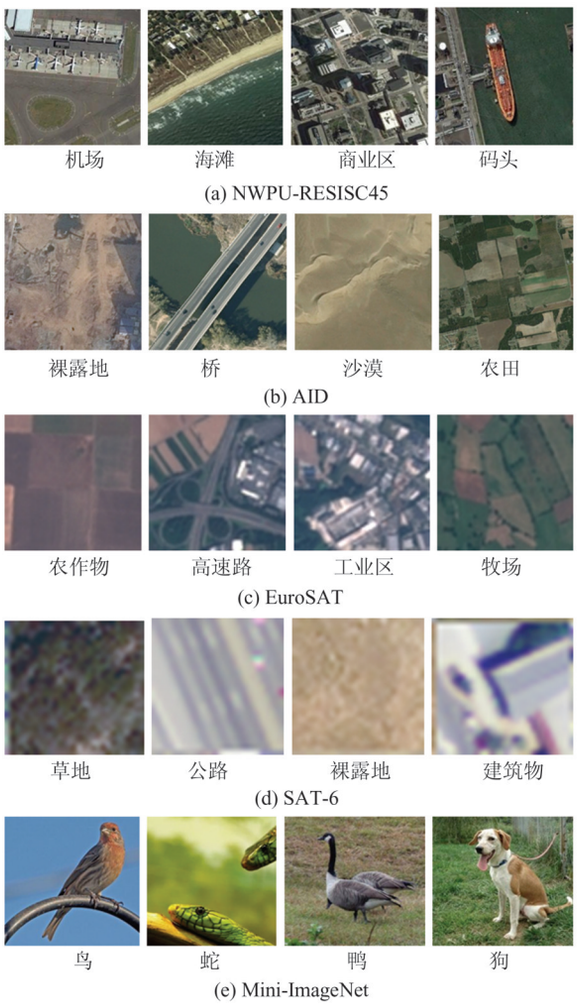


图 1　5种数据集样本示例

Fig. 1　Examples of Five Experimental Datasets

表 1　5种数据集的基本情况

Table 1　Basic Information of Five Datasets

数据集	场景数	总样本数	影像大小/像素
NWPU-RESISC45	45	31 500	256×256
AID	30	10 000	600×600
EuroSAT	10	27 000	64×64
SAT-6	6	405 000	28×28
Mini-ImageNet	100	60 000	不固定

大陆的共330 000个场景组成,图像由红色、绿色、蓝色和近红外4个波段组成,每个样本图像大小为28×28像素。为了保持整个数据集固有的高方差,从多个场景中采样1 500个图像块,这些场景覆盖了不同的景观,如农村地区、城市地区、茂密的森林、山区以及覆盖整个加利福尼亚州的水体、农业区等。

5) Mini-ImageNet数据集由Google于2016年从ImageNet数据集中抽取一小部分制作而成,约3 GB,共有100个类别,每个类别有600张图像,总计60 000张图像,图像大小不固定。考虑到ImageNet数据集规模庞大,约100 GB,对普通研究员或开发者而言,训练对硬件要求较高,通常需要多个高端显卡并行训练,即使如此,训练仍需数天时间,因此Mini-ImageNet数据集的制作有效降低了使用门槛。

实验过程中,除了公开数据集自身已经划分好的训练集和测试集之外,本文实验过程中对其他数据集进行随机划分,训练样本和测试样本的划分比例为6∶4,即60%的样本用于训练,40%的样本用于测试,并评定分类精度[50]。

2.2　网络设置

在本文的实验中,采用ResNet18的网络结构用于遥感场景的分类。在网络初始化方面,采用文献[15]的方法对所有的网络参数进行初始化。对于所有的归一化方法都采用了仿射变换,进一步增强网络表征特征的能力,其尺度参数的初始值设置为1,平移参数的初始值设置为0。使用随机梯度下降算法对所有的模型进行优化,基础学习率设置为0.1,权重衰减因子为1×10^{-4},动量值为0.9。在训练过程中,采用文献[28]的方法进行数据增强,同时轮数设置为100。测试过程中,影像输入为中心裁剪为224×224像素大小的影像块。实验最终结果取最后10次随机实验Top1测试分类精度的平均值和标准差(即波动大小)。其他设置细节与文献[27]保持一致。此外,本文实验的主要目的在于验证本文所提出方法的有效性,仅比较不同的归一化方法之间的差异性,而并不追求该方法达到的最优分类结果。

实验中的参考基准是用BABN训练ResNet18模型。为了与其他归一化方法比较,实验中BABN方法依次被其他归一化方法替换,而模型中的其他超参数都保持不变。此外,为了更好地分析BS对不同归一化方法的影响,批次的样本容量分别设置为2、4、8、16、32。

2.3　实验结果与分析

2.3.1　不同归一化方法Top1测试分类精度比较

首先,按照BN方法对批次的样本容量的常规设置,设置为32。考虑到在场景的类别数和每个场景的样本数量等方面都相对较大,实验中选用NWPU-RESISC45数据集。经过100次迭代,所有归一化方法均已收敛,5种归一化方法的分类精度见表2,其中精度表示Top1测试分类精度;标准差表示最后10次随机实验结果精度的标准差;Δ表示测试方法与BN方法的精度差值。

表2 5种归一化方法的分类精度

Table 2 Classification Accuracy of Five Normalization Methods

参数	BN	GN	IN	LN	BABN
精度	0.870 4	0.854 3	0.810 2	0.869 9	0.926 8
标准差	0.001 9	0.001 2	0.000 8	0.000 9	0.000 7
Δ	—	−0.016 1	−0.060 2	−0.000 5	0.056 4

由表2可以看出,BABN方法的精度最高,标准差最小,且BABN较BN方法的Top1测试分类精度高5.64%,验证了BABN方法的有效性。借助贝叶斯理论,将总体信息、样本信息和先验信息三者科学合理融合,可以更加准确地估计样本均值和方差,使得归一化后的特征落入最佳的非饱和区域,从而更好地反映整个特征空间的原始表征,进而深度学习模型可以达到最佳的特征表达能力。此外,其他归一化方法不同程度地将不同维度的信息强制融合在一起,表面上增加了样本的容量,实际上在一定程度上也增加了不同维度信息之间的不一致性,这对样本均值和方差的估计精度造成不同程度的影响。当批次的样本容量较大时,LN方法受到的影响相对较小,IN方法受到的影响相对较大,而GN方法介于LN和IN方法之间,该实验结果与理论层面是一致的。

2.3.2 批次BS对不同归一化方法的影响

BN方法的性能主要依赖于BS,当BS较小时,BN方法的性能会急剧下降,而BABN方法在理论上克服了BS较小的问题。为了分析不同的样本容量对不同归一化方法性能的影响,在NWPU-RESISC45数据集上将批次的样本容量分别设置为2、4、8、16、32,5种归一化方法在不同批次样本容量下的分类精度见表3,该精度是最后10次随机实验结果的平均值。

表3 5种归一化方法在不同批次样本容量下的分类精度

Table 3 Classification Accuracy of Five Normalization Methods with Different Batch Sizes

方法	BS				
	32	16	8	4	2
BABN	0.926 8	0.908 7	0.898 5	0.888 0	0.878 0
BN	0.870 4	0.885 7	0.873 2	0.837 0	0.681 7
GN	0.854 3	0.875 0	0.869 2	0.820 2	0.799 3
IN	0.810 2	0.861 6	0.868 3	0.872 2	0.813 0
LN	0.869 9	0.879 6	0.869 4	0.815 5	0.774 5

由表3可以看出,当BS为32时,BN方法的精度较高。随着BS由32逐渐减少至2时,BN方法的分类精度从原来的0.870 4下降到0.681 7;但对BABN来说影响却较小,分类精度从0.926 8下降到0.878 0,仍然保持较高的分类精度;而GN、IN和LN方法均有不同程度的下降,虽然下降幅度不大,但最终的分类精度都低于BABN方法BS为2时的0.878 0。

2.3.3 不同的网络骨架对归一化方法的影响

为了进一步验证BABN方法的有效性,将不同的网络骨架进行替换,其他实验参数保持不变,比较在不同网络骨架下BN与BABN在NWPU-RESISC45数据集上Top1的测试分类精度。设置BS为4,实验结果见表4,其中BABN−BN表示两种方法的精度差值。

表4 不同网络骨架下的分类精度(BS为4)

Table 4 Comparison of Classification Accuracy with Different Backbones(BS Is 4)

方法	ResNet18	VGG16	Movile-Net-v3	Incep-tion-v3	Dense-Net-121
BN	0.837 1	0.799 1	0.852 3	0.804 3	0.796 1
BABN	0.888 0	0.848 3	0.895 7	0.858 6	0.839 9
BABN−BN	0.050 9	0.049 2	0.043 4	0.054 3	0.043 8

由表4可以看出,BABN优于BN方法,分类精度平均提高了5%左右。当BS为相对较小值,如BS为4时,所携带的信息也相对有限,通过利用总体信息和先验信息来弥补BS较小时所携带样本信息的不足,便可以更加准确和鲁棒地估计样本均值和方差,从而更好地表征与学习特征,进而提高Top1测试的分类精度。实验结果表明,BABN方法对不同的网络架构都有显著的提升效果。

2.3.4 不同归一化方法在不同数据集上的差异

为了进一步验证BABN方法的有效性,对其在不同的数据集上进行了实验与分析,具体的实验结果见表5,其中BS为2。由表5可知,BABN在5个不同类型的数据集中分类精度最优。

表5 不同数据集的归一化方法的分类精度(BS为2)

Table 5 Comparison of Classification Accuracy on Different Datasets(BS Is 2)

方法	AID	EuroSAT	NWPU-RESISC45	Mini-Ima-geNet	SAT-6
BABN	0.938 7	0.973 5	0.878 0	0.886 9	0.990 7
BN	0.732 1	0.676 7	0.681 7	0.489 5	0.421 3
GN	0.857 7	0.926 5	0.799 3	0.494 9	0.976 2
IN	0.865 1	0.942 1	0.813 0	0.505 8	0.934 5
LN	0.793 3	0.922 2	0.774 5	0.455 8	0.977 2

2.3.5　不同归一化方法的时间复杂度比较

从算法的时间复杂度角度对5种归一化方法在不同的BS下进行了比较实验,结果见表6,其中,时间复杂度数值为除去第一轮迭代后,在剩余99次迭代中每次迭代所需时间的平均值。

由表6可以看出,随着BS的增加,由于GPU的并行计算的优势,其时间复杂度逐渐减少。在上述的5种归一化方法中,BN和BABN方法的时间复杂度最小,而且两者基本一致,这说明BABN方法既可以提高算法的精度,又保持了较好的时间复杂度。

表6　5种归一化方法的时间复杂度

Table 6　Computation Time of Five Normalization Methods

BS	BN/s	GN/s	IN/s	LN/s	BABN/s
2	139	130	136	212	139
4	86	86	90	133	86
8	66	67	72	89	67
16	56	57	63	66	57
32	50	51	57	54	51

3　结　语

当前的归一化方法主要是利用样本信息来估计样本均值和方差,当BS较小时,学者们提出从其他维度来弥补BS较小的不足。然而,在分析不同归一化方法的基础上,对于BN方法中对BS依赖这一问题,本文从贝叶斯理论的角度出发,考虑将总体信息、先验信息和样本信息科学、合理的融合方式来弥补BS的不足,从而更加准确地估计样本均值和方差,使得归一化后的特征落入最佳的非饱和区域,以便更好地反映整个特征空间的原始表征,进而深度学习模型可以达到最佳的特征表达能力。实验与分析表明,本文提出的BABN方法是可行、有效的,在NWPU-RESISC45数据集上,其分类精度比BN方法要高5.64%。而且得益于总体信息和先验信息的帮助,BABN受BS的影响较小,而BABN的时间复杂度与BN方法基本一致。此外,从理论角度来看,BABN方法也适用于影像语义分割、目标检测和动作识别等应用领域,这也是将来需要进一步深入研究的内容。

参　考　文　献

[1] 张永军, 万一, 史文中, 等. 多源卫星影像的摄影测量遥感智能处理技术框架与初步实践[J]. 测绘学报, 2021, 50(8): 1068-1083.
ZHANG Yongjun, WAN Yi, SHI Wenzhong, et al. Technical Framework and Preliminary Practices of Photogrammetric Remote Sensing Intelligent Processing of Multi-source Satellite Images[J]. *Acta Geodaetica et Cartographica Sinica*, 2021, 50(8): 1068-1083.

[2] 史文中, 张敏. 人工智能用于遥感目标可靠性识别: 总体框架设计、现状分析及展望[J]. 测绘学报, 2021, 50(8): 1049-1058.
SHI Wenzhong, ZHANG Min. Artificial Intelligence for Reliable Object Recognition from Remotely Sensed Data: Overall Framework Design, Review and Prospect[J]. *Acta Geodaetica et Cartographica Sinica*, 2021, 50(8): 1049-1058.

[3] 邵振峰, 孙悦鸣, 席江波, 等. 智能优化学习的高空间分辨率遥感影像语义分割[J]. 武汉大学学报(信息科学版), 2022, 47(2): 234-241.
SHAO Zhenfeng, SUN Yueming, XI Jiangbo, et al. Intelligent Optimization Learning for Semantic Segmentation of High Spatial Resolution Remote Sensing Images[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(2): 234-241.

[4] 龚健雅, 张展, 贾浩巍, 等. 面向多源数据地物提取的遥感知识感知与多尺度特征融合网络[J]. 武汉大学学报(信息科学版), 2022, 47(10): 1546-1554.
GONG Jianya, ZHANG Zhan, JIA Haowei, et al. Multi-source Data Ground Object Extraction Based on Knowledge-Aware and Multi-scale Feature Fusion Network[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(10): 1546-1554.

[5] 虞欣, 郑肇葆, 李林宜. 适用于训练样本选择的斜交因子模型研究[J]. 武汉大学学报(信息科学版), 2022, 47(11): 1870-1877.
YU Xin, ZHENG Zhaobao, LI Linyi. Oblique Factor Model for Selecting Training Samples[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(11): 1870-1877.

[6] 李彦胜, 张永军. 耦合知识图谱和深度学习的新一代遥感影像解译范式[J]. 武汉大学学报(信息科学版), 2022, 47(8): 1176-1190.
LI Yansheng, ZHANG Yongjun. A New Paradigm of Remote Sensing Image Interpretation by Coupling Knowledge Graph and Deep Learning[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(8): 1176-1190.

[7] 周培诚, 程塨, 姚西文, 等. 高分辨率遥感影像解译中的机器学习范式[J]. 遥感学报, 2021, 25(1): 182-197.
ZHOU Peicheng, CHENG Gong, YAO Xiwen, et

al. Machine Learning Paradigms in High-Resolution Remote Sensing Image Interpretation[J]. *National Remote Sensing Bulletin*, 2021, 25(1): 182-197.

[8] 刘建伟, 赵会丹, 罗雄麟, 等. 深度学习批归一化及其相关算法研究进展[J]. 自动化学报, 2020, 46(6): 1090-1120.
LIU Jianwei, ZHAO Huidan, LUO Xionglin, et al. Research Progress on Batch Normalization of Deep Learning and Its Related Algorithms[J]. *Acta Automatica Sinica*, 2020, 46(6): 1090-1120.

[9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.

[10] CARANDINI M, HEEGER D J. Normalization as a Canonical Neural Computation[J]. *Nature Reviews Neuroscience*, 2012, 13: 51-62.

[11] HEEGER D J. Normalization of Cell Responses in Cat Striate Cortex[J]. *Visual Neuroscience*, 1992, 9(2): 181-197.

[12] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. *CoRR*, 2015(1): 448 - 456.

[13] PENG C, XIAO T T, LI Z M, et al. MegDet: A Large Mini-Batch Object Detector[C]//Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.

[14] WU Y X, HE K M. Group Normalization[C]//European Conference on Computer Vision (ECCV), Munich, Germany, 2018.

[15] HE K M, ZHANG X Y, REN S Q, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[C]//IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015.

[16] CHENG G, HAN J W, LU X Q. Remote Sensing Image Scene Classification: Benchmark and State of the Art[J]. *IEEE*, 2017, 105(10): 1865-1883.

[17] SHARMA S, KULKARNI R, AJITHAPRASAD S, et al. Fringe Pattern Normalization Algorithm Using Kalman Filter[J]. *Results in Optics*, 2021, 5(1): 100152.

[18] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance Normalization: The Missing Ingredient for Fast Stylization[EB/OL]. [2016-05-17] https://arxiv.org/abs/1607.08022v3.

[19] BA J, KIROS J, HINTON G E. Layer Normalization[EB/OL]. [2016-06-14] https://arxiv.org/abs/1607.06450.

[20] SHAO W Q, LI J Y, REN J M, et al. SSN: Learning Sparse Switchable Normalization *via* SparsestMax[J]. *International Journal of Computer Vision*, 2020, 128(8): 2107-2125.

[21] LI B, WU F, WEINBERGER K Q, et al. Positional Normalization[C]//Neural Information Processing Systems, Vancouver, Canada, 2019.

[22] YAO Z, CAO Y, ZHENG S, et al. Cross-Iteration Batch Normalization[C]//Computer Vision and Pattern Recognition (CVPR), Beijing, China, 2021.

[23] SINGH S, KRISHNAN S. Filter Response Normalization Layer: Eliminating Batch Dependence in the Training of Deep Neural Networks[C]//Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020.

[24] IOFFE S. Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models [EB/OL]. [2022-12-21] https://arxiv.org/abs/1702.03275v2.

[25] SONG Y B, XIANG J, JIANG J W, et al. A Cross-Domain Change Detection Network Based on Instance Normalization[J]. *Remote Sensing*, 2023, 15(24): 5785.

[26] QIAO S Y, WANG H Y, LIU C X, et al. Weight Standardization[EB/OL]. [2009-01-23] https://arxiv.org/abs/1903.10520.

[27] CHEN Peng, WANG Benkang, GAO Sa, et al. Building Collapse Assessment with Residual Network[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(8): 1179-1184.

[28] SZEGEDY C, LIU W, JIA Y Q, et al. Going Deeper with Convolutions[C]//Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019.

[29] KAZIMI B, SANDFELD S. Enhancing Semantic Segmentation in High-Resolution TEM Images: A Comparative Study of Batch Normalization and Instance Normalization [J]. *Microscopy and Microanalysis*, 2024, 1: 1527-1546.

[30] HUANGI L, HUANGI L, YANG D W, et al. Decorrelated Batch Normalization[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.

[31] VRIES H D, STRUB F, JÉRÉMIE M, et al. Modulating Early Visual Processing by Language [EB/OL]. [2017-11-12] https://arxiv.org/abs/1707.00683.

[32] SINGH S, SHRIVASTAVA A. EvalNorm: Estimating Batch Normalization Statistics for Evaluation [C]//IEEE/CVF International Conference on Com-

puter Vision (ICCV), Seoul, Korea, 2019.

[33] JIA S H, CHEN D J, CHEN H T. Instance-Level Meta Normalization[C]//Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019.

[34] GÜLÇEHRE Ç, BENGIO Y. Knowledge Matters: Importance of Prior Information for Optimization [J]. *Journal of Machine Learning Research*, 2016, 17(8): 1-8.

[35] ARPIT D, ZHOU Y B, KOTA B U, et al. Normalization Propagation: A Parametric Technique for Removing Internal Covariate Shift in Deep Networks [EB/OL]. [2016-10-25] https://arxiv. org/abs/1603. 01431v6.

[36] REN M Y, LIAO R J, URTASUN R, et al. Normalizing the Normalizers: Comparing and Extending Network Normalization Schemes[EB/OL]. [2016-08-18] https://arxiv. org/abs/1611. 04520v2.

[37] GONG X Y, CHEN W Y, CHEN T L, et al. Sandwich Batch Normalization: A Drop-in Replacement for Feature Distribution Heterogeneity [C]// Conference on Applications of Computer Vision (WACV), Waikoloa, USA, 2022.

[38] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral Normalization for Generative Adversarial Networks[EB/OL]. [2018-12-09] https://arxiv. org/abs/1802. 05957v1.

[39] LIAO Q L, KAWAGUCHI K, POGGIO T. Streaming Normalization: Towards Simpler and More Biologically-Plausible Normalizations for Online and Recurrent Learning [EB/OL]. [2016-19-15] https://arxiv. org/abs/1610. 06160v1.

[40] LUO P, PENG Z L, REN J M, et al. Do Normalization Layers in a Deep ConvNet Really Need to Be Distinct [EB/OL]. [2018-07-23] https://arxiv. org/abs/1811. 07727v1.

[41] SALIMANS T, KINGMA D P. Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks [EB/OL]. [2016-05-19] https//arxiv. org/abs/1602. 07868.

[42] 茆诗松, 汤银才. 贝叶斯统计(第 2 版)[M]. 北京: 中国统计出版社, 2012.
MAO Shisong, TANG Yincai. Bayesian Statistics [M]. 2nd ed. Beijing: China Statistics Press, 2012.

[43] 冯权泷, 陈泊安, 李国庆, 等. 遥感影像样本数据集研究综述[J]. 遥感学报, 2022, 26(4): 589-605.
FENG Quanlong, CHEN Boan, LI Guoqing, et al. A Review for Sample Datasets of Remote Sensing Imagery [J]. *National Remote Sensing Bulletin*, 2022, 26(4): 589-605.

[44] 胡翔云, 张觅, 张祖勋, 等. 遥感影像智能解译系统 EasyFeature 的关键技术及应用[J]. 武汉大学学报(信息科学版), 2025, 50(3): 554-561.
HU Xiangyun, ZHANG Mi, ZHANG Zuxun, et al. Key Technologies and Applications of EasyFeature—An Intelligent Interpretation System for Remote Sensing Images [J]. *Geomatics and Information Science of Wuhan University*, 2025, 50(3): 554-561.

[45] 钟守熠, 肖青, 闻建光, 等. 测绘地物波谱本底数据库[J]. 遥感学报, 2020, 24(6): 701-716.
ZHONG Shouyi, XIAO Qing, WEN Jianguang, et al. Design and Realization of Ground Object Background Spectral Library for Surveying and Mapping [J]. *Journal of Remote Sensing*, 2020, 24(6): 701-716.

[46] XIA G S, HU J W, HU F, et al. AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 3965-3981.

[47] HELBER P, BISCHKE B, DENGEL A, et al. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(7): 2217-2226.

[48] BASU S, GANGULY S, MUKHOPADHYAY S, et al. DeepSat: A Learning Framework for Satellite Imagery[C]//The 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, USA, 2015.

[49] DENG J, DONG W, SOCHER R, et al. ImageNet: A Large-Scale Hierarchical Image Database [C]//Computer Vision and Pattern Recognition, Miami, USA, 2009.

[50] 陶超, 阴紫薇, 朱庆, 等. 遥感影像智能解译: 从监督学习到自监督学习[J]. 测绘学报, 2021, 50(8): 1122-1134.
TAO Chao, YIN Ziwei, ZHU Qing, et al. Remote Sensing Image Intelligent Interpretation: From Supervised Learning to Self-Supervised Learning[J]. *Acta Geodaetica et Cartographica Sinica*, 2021, 50(8): 1122-1134.