

神经结构的进化优化  
在遥感分类问题中

丹尼尔·科奎林<sup>1</sup>， 罗科塞多纳<sup>2,3</sup>， 莫里斯·里德尔<sup>2,3</sup>， Markus Götz<sup>1</sup>

<sup>1\*</sup> 德国卡尔斯鲁厄理工学院 (KIT) Steinbuch 计算中心 (SCC)  
<sup>2\*</sup> Juelich 超级计算中心 (JSC),Forschungszentrum Jülich (FZJ),德国  
<sup>3\*</sup> 冰岛大学 (UoI) 工程与自然科学学院 (SENS),冰岛

抽象的

BigEarthNet 是标准的大型遥感数据集之一。之前已经表明,神经网络是对这些数据中的图像块进行分类的有效工具。

然而,找到最佳网络超参数和架构以准确分类 BigEarthNet 中的图像块仍然是一个挑战。手动搜索更准确的模型非常耗时耗力。因此,系统的方法是可取的。

一种可能性是自动进化神经架构搜索 (NAS)。使用此 NAS,许多常用的网络超参数 (例如损失函数)被消除,并确定了更准确的网络。

索引词 神经架构搜索、NAS、Evo

算法,遥感,分类

一、简介

在过去十年中,深度学习在许多领域取得了前所未有的成果,遥感 (RS) 也不例外。由于深度神经网络需要大量数据进行训练,来自星载传感器的持续测量流使 RS 成为一个常见的用例。

近年来,已发布可公开访问、标记的 RS 特定数据集 [1,2],允许 RS 社区训练针对特定研究问题量身定制的神经网络。

ResNet-50 是一种具有跳跃连接的卷积神经网络 (CNN),已被证明对计算机视觉任务有效 [3]。它已经过广泛测试,是许多不同领域 (包括遥感)中使用的许多 CNN 的基础。

有多种方法可以评估神经网络的有效性,其中许多方法使用准确率和召回率的思想。F1分数就是这样一种衡量标准。它是**精确率**和**召回率**的调和平均值,常用于多类多标签的情况。结果可以全局计算,即micro-F1 score,也可以是每个class的F1 score的未加权平均,即macro-F1

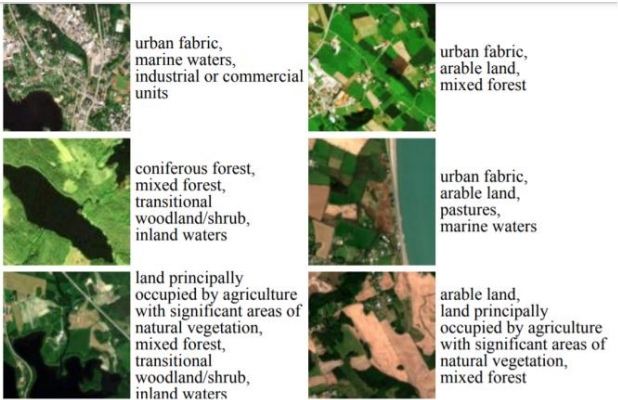


图1. Sentinel-2 tiles [6] 的示例补丁和标签。

得分[4]。还有其他方法,但它们与本文无关。

大地网<sup>1</sup>，一个大型遥感数据集,包含从 2017 年 6 月至 2018 年 5 月获取的 125 个 Sentinel-2 tiles (Level2A) 中提取的补丁 [1]。该档案包含 590,326 个补丁,每个补丁都分配有 19 个可用标签中的一个或多个。标签命名法是对 CORINE Land Cover [5] 的改编,它由 2018 年更新的来自 10 个欧洲国家的标签组成 [6]。每个补丁有 12 个不同分辨率的光谱波段:3 个 RGB 波段和 10 m 分辨率 (120 px × 120 px)的波段 8;波段 5、6、7、8a、11 和 12,分辨率为 20 m (60 px × 60 px);波段 1 和波段 9 分辨率为 60 m (20 px × 20 px)。卷云敏感波段 10 以及被雪或云覆盖的斑块都被省略了 [7]。示例补丁如图 1 所示。1.

在 [6] 中,针对 BigEarthNet 数据集对多种网络类型进行了不同程度的成功训练。这些实验排除了波段 1 和波段 9,使用 Adam 优化器、S 型交叉熵损失、0.001 的初始学习率,并训练了 100 个时期。结果发现,ResNet-50 是最准确的网络架构,微观 F1 得分为 77.11,宏观 F1 得分为 67.33。

<sup>1</sup><http://bigearth.net/>

由于部分数据被排除在这些实验中（波段 1 和 9），包含这些数据可能会提高网络的有效性。这与不同的网络架构或训练配置相结合，有可能大大提高训练时间和准确性。

2. 进化优化

神经网络的结构由其超参数决定。这些包括优化器、训练时期的数量、网络架构、损失函数等等。由于网络的超参数组合数量可能变得非常大，因此通常无法手动确定有效组合。因此，它通常通过选定的超参数搜索和神经架构搜索 (NAS) 自动完成。NAS 通常需要大量时间才能完成，因为它必须为每组超参数训练网络以确定该组的有效性或质量，在 NAS 中寻找最佳超参数的算法差异很大 [8]，但是这项工作利用进化 NAS 为 [6] 中使用的 ResNet-50 网络搜索最佳配置。

要使用的进化 NAS 始于一组随机超参数集，称为后代。这些网络经过训练后，会被分类到交配种群中。当将后代添加到交配种群时，性能最差的网络将被移除。新一代的后代是通过繁殖交配种群的两个随机成员来创建的，即从每个集合中随机抽样超参数。在交配过程中可能会发生突变，即修改少量后代的集合。然后对这些新网络进行训练，并重复该过程指定的代数或直到外部因素停止为止。

3. 实验

超参数搜索空间分为六组：优化器、学习率 (LR) 调度器、激活函数、损失函数、每个卷积块中的过滤器数量和激活顺序。搜索空间如表 1 所示，不包括以指数方式跨越 2 和 256 之间空间的过滤器数量。对于该网络，卷积块中的过滤器数量取决于与第一个块的过滤器数量的比率。

学习率调度器和优化器的参数也包含在搜索空间中。

[13] 表明，激活顺序、批量归一化 (BN) 和残差构建块中的卷积层会影响网络的准确性。测试配置如图 1 所示。2 将在引用激活顺序时使用。

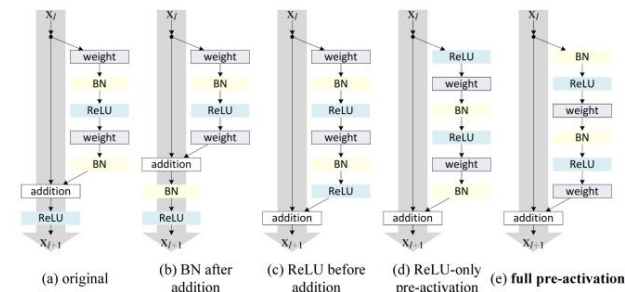


图2. 网络中使用的残差构建块中的激活层、批量归一化 (BN) 和卷积层的各种顺序。ReLU 是一个示例激活函数，在 NAS 中，激活函数由网络超参数定义 [13]。

数据的准备类似于 [6]。尽管如此，没有对图像执行图像增强。网络在 TensorFlow [17] 中实现，NAS 由开源包 propulate [18] 控制。propulate 通过 MPI [19] 维持种群的状态。这些实验是在 KIT 的 ForHLR 2 上对不同数量的 Nvidia A100 GPU [20] 进行的。如果在验证集上测量的损失在十个时期内不再增加，则网络使用提前停止来退出训练。

包含优化器 Adam、Adamax、Nadam 和 RMSprop 的超参数组合比其他优化器更频繁地失败。如果损失函数、优化器及其参数之间的组合不利，这些优化器将返回 NaN 值，这很可能是由于它们的自适应算法。这种行为通常称为不稳定。重要的是要注意，优化器的稳定性通常不能很好地衡量其有效性。由于其他优化器更稳定，更不稳定的优化器很快就会被排除在搜索空间之外。为了补偿这种影响，对 Adam、Adamax、Nadam 和 RMSprop 进行了单独的 NAS。

随着性能不佳的网络从总体中移除，超参数选择的相对频率可以用作具有该超参数的网络的稳定性和准确性的代理。这种测量是在所有已完成的搜索中进行的，以避免排除那些在不太稳定的优化器下表现更好的指标。通过这种衡量，最成功的损失函数是二元交叉熵和分类铰链。最不有效的损失函数是铰链、Kullback-Leibler 散度和平方铰链。指数衰减、反时限衰减和多项式衰减的使用频率大致相同。这表明，当给定合适的参数时，这三个都是学习率调度器的稳定和有效的选择。使用这个度量来确定激活函数的有效性是有缺陷的，因为激活函数

表1. 神经结构和超参数搜索空间。激活列标题显示激活函数。  
激活订单详情如图 1 所示。 2. ELU是指数线性单元,ReLU是整流线性单元,SELU是缩放指数线性单元,KL散度是Kullback-Leibler散度 ,tanh是双曲正切。

优化器[9, 10] LR 调度器[11]激活[12]激活顺序[13]			损失[14, 15, 16]
阿达德尔塔	指数衰减 ELU	原版的	二元交叉熵
阿达格拉德	反时限衰减指数	添加后的BN	分类交叉熵
亚当	多项式衰减硬 sigmoid	添加前激活	绝对铰链
高级管理硕士		仅激活预激活铰链完全预激活	
阿达玛克斯			KL散度
筛选器			方形铰链
那达慕			
RMSprop			
新元			

有效性被发现与网络优化器相关。即 Adamax 对 ELU 最有效,而 Adadelat 对 Softmax 最有效。尽管如此,最常选择的激活函数是线性函数。从理论上讲,这是由于它在许多优化器中的稳定性。因此,它在 NAS 的早期阶段被选中,并逐渐被忽视,因为发现了其他更有效的激活函数。与线性激活函数的使用类似,在 NAS 的早期阶段选择的过滤器数量主要是四个。随着 NAS 的发展,过滤器的数量主要选择 32 或 64。最有效的激活顺序是完全预激活 (如图 2 所示)。

如前所述,一些优化器被隔离到单独的搜索中,因此它们的频率度量不太有用。因此,它们是根据优化器产生的F1分数来判断的。 Ftrl 并没有比其他的更频繁地失败,但是它并没有继续产生有竞争力的结果,并且慢慢地从一般 NAS 中被淘汰。使用 Adamax,RMSprop 和 Adam 进行的单独搜索都产生了与集体 NAS 的竞争网络。表明因稳定性问题而被排除在外的网络对人口来说是不公平的。

最有效的优化器是 Adadelat 和 SGD,它们都产生了多个配置,其结果大于或等于目标 micro-F1分数 77.11。最准确的配置在 10 个时期内收敛到77.25 的微观F1分数和 69.57的宏观 F1分数。

这两者都是对 [6] 中提出的网络的准确性改进,并且发现的网络在 90 个更少的时期内收敛。最终的配置如下。优化器是 Adadelat,rho 为 0.8831,epsilon 为 3.4771e - 06;学习率调度器是多项式 de

初始学习率为 0.8666,衰减步数为 10585,结束学习率为 0.0035,幂为 0.8693,并且没有循环; 128个过滤器; softmax 激活函数;二元交叉熵损失 ;并具有完整的预激活功能。

在类别级别,所发现的网络在某些类别中明显优于 [6] 中使用的网络。即滨海湿地;沼泽、荒地和硬叶植被 ;和农林地区的准确率分别比提出的 ResNet50 实施高 9.60%、5.05% 和 6.86%。然而,“海滩、沙丘和沙滩”类别的表现却差了 17.4%。由于整体性能非常相似,类精度的巨大差异归因于训练数据中包含更多光谱带。

4.结论

最近大型 RS 数据集可用性的增加使研究人员能够在土地覆盖分类等任务中取得前所未有的成果。然而,这是以花费大量时间手动搜索最佳 DL 架构和超参数为代价的。

如果使用得当,NAS 可以确定网络超参数,从而加快收敛速度和/或提高准确性。但是,如果 NAS 的搜索空间包含具有不同稳定性的函数的超参数,则可能会不公平地排除某些超参数。为避免这种情况,需要仔细选择搜索空间超参数或多个 NAS。各种优化器就是例证,这些优化器需要单独的 NAS 才能产生有竞争力的结果。

已经表明,进化的NAS可以提高准确性并大大提高目标网络的收敛速度。 Adadelat、多项式衰减、Soft max 和二元交叉熵的使用已被证明略微超出

执行初始网络。

通过包含更多光谱带,一些类别显示出大大提高的准确性。然而,其中一个类别(海滩、沙丘和沙滩)的准确度明显低于之前的结果。因此,应该根据训练后网络的预期用例来包含更多频段。需要更多的研究来确定每个波段对图像准确分类的有效性。

这个结果,以及任何 NAS 的结果,都受到其搜索空间的限制,并且由于搜索空间内的功能是根据它们的共性选择的,因此有许多未包括在内的尖端功能。为了大大超越给定的网络,需要结合实施更高级的网络架构和选择更有用的损失函数。

致谢

这项工作是在巴登 - 符腾堡州科学、研究和艺术部以及联邦教育和研究部资助的超级计算机 ForHLR 上进行的。这项工作得到了亥姆霍兹协会在亥姆霍兹人工智能平台资助和亥姆霍兹分析框架合作下的倡议和网络基金的支持。

5. 参考文献

[1] Gencer Sumbul,Marcela Charfuelan,Begüm Demir 和 Volker Markl,“BigEarthNet:用于遥感图像理解的大规模基准档案”,arXiv 预印本 1902.06148,2019。

[2] Xiao Xiang Zhu, Jingliang Hu, Chunping Qiu, Yilei Shi, Jian Kang, et al., “So2Sat LCZ42: A Benchmark Data Set for the Classification of Global Local Climate Zones [Software and Data Sets],” IEEE Geoscience and遥感杂志,第一卷。 8,没有。 3,第 76-89 页,2020 年。

[3] Kaiming He,Xiangyu Zhang,Shaoqing Ren 和 Jian Sun,“用于图像识别的深度残差学习”,2016 年 IEEE 计算机视觉和模式识别 (CVPR) 会议。 2016 年 6 月,第 770-778 页,IEEE。

[4] Zachary Chase Lipton,Charles Elkan 和 Balakrish nan Narayanaswamy,“阈值分类器最大化 F1 分数”,2014 年。

[5] M. Bossard,J. Feranec 和 J. Otahel,“CORINE 土地覆盖技术指南 – 附录 2000”,Tech.代表,欧洲环境署,2000 年。

[6] Gencer Sumbul,Jian Kang,Tristan Kreuziger,Filipe Marcelino、Hugo Costa 等,“BigEarthNet 数据集

与遥感图像理解的新类别命名法”,2020 年。

[7] “去除多云和多雪斑块脚本”。

[8] Thomas Elsken、Jan Hendrik Metzen 和 Frank Hutter,“神经架构搜索:一项调查”,2019 年。

[9] Shiliang Sun,Zehui Cao,Han Zhu 和 Jing Zhao,“机器学习视角下的优化方法综述”,arXiv 预印本 1906.06821,2019。

[10] H. Brendan McMahan,“自适应在线学习分析技术”,arXiv 预印本 1403.3465,2014 年。

[11] Yanzhao Wu,Ling Liu,Juhyun Bae 等人,“Demystify ing Learning Rate Policies for High Accuracy Training of Deep Neural Networks”,arXiv 预印本 1908.06477,2019。

[12] Chigozie Nwankpa,Winifred Ijomah,Anthony Gacha gan 和 Stephen Marshall,“激活函数:深度学习实践和研究趋势的比较”,arXiv 预印本 1811.03378,2018 年。

[13] Kaiming He,Xiangyu Zhang,Shaoqing Ren 和 Jian Sun,“深度残差网络中的身份映射”,arXiv 预印本 1603.05027,2016 年。

[14] S. Kullback 和 RA Leibler,“关于信息和充分性”,Ann.数学。统计学家,卷。 22,没有。 1,第 79-86 页,1951 年 3 月。

[15] Qi Wang,Yue Ma,Kun Zhao 和 Yingjie Tian,“机器学习中损失函数的综合调查”,数据科学年鉴,2020 年 4 月。

[16] Shruti Jadon,“语义分割损失函数调查”,2020 年 IEEE 生物信息学和计算生物学计算智能会议 (CIBCB),2020 年 10 月。

[17] Martín Abadi,Ashish Agarwal,Paul Barham 等人,“TensorFlow: Het 异构系统上的大规模机器学习”,2015 年,软件可从 tensorflow.org 获得。

[18] Oskar Taubert,“Propulate”,https://github. com/oskar-taubert/propulate,2020 年。

[19] 消息传递论坛,“MPI:消息传递接口标准”,Tech.美国田纳西大学众议院,1994 年。

[20] Yuhsiang Mike Tsai,Terry Cojean 和 Hartwig Anzt,“评估 NVIDIA A100 Am pere GPU 用于稀疏线性代数计算的性能”,2020 年。