

基于改进的协作主动学习 高光谱图像的胶囊网络 分类

王恒、王立国

摘要：对于高光谱图像分类（HIC）任务，大多数基于不确定性的主动学习（AL）方法只考虑不确定性，而没有考虑主动选择样本的多样性和专家标记的预算。在本文中，我们提出了一个协作主动学习（CAL）框架来解决这个问题。该框架由两个设计良好的基本分类器和一个巧妙的CAL方案组成，该方案考虑了主动选择样本的不确定性和多样性以及专家注释的成本。具体来说，为了受益于胶囊网络准确识别和定位特征的能力，我们设计了两种改进的胶囊网络。对于这两个网络，我们称其为第一个CapsViT（胶囊视觉变压器），它将视觉变压器（ViT）引入胶囊网络（CapsNet），以了解胶囊特征之间的全局关系。我们称之为第二个CapsGLOM（胶囊GLOM），该网络的基本结构来自于杰弗里·辛顿提出的GLOM系统，我们从CapsNet构造主胶囊的方式来改进其实现细节。在提出的CAL框架中，采用CapsViT和CapsGLOM的两个基分类器，在充分考虑专家注释成本的前提下，根据CAL方案选择信息最丰富的样本。在四个基准高光谱图像数据集上的实验结果表明，我们提出的CAL框架能够获得令人满意的分类结果。同时，与其他先进的深度模型相比，我们提出的CapsViT和CapsGLOM在监督HIC任务中也具有竞争力。源代码可以在网上获得(<https://github.com/swiftest/CAL>)。

索引术语———高光谱图像分类、协作主动学习、CapsViT、CapsGLOM、视觉转换器。

I. 介绍

由安装在机载和空载飞机上的高光谱传感器获得的高光谱图像（HSIs）记录了具有数百个连续光谱波段的地面场景。虽然HSIs的空间分辨率通常较低，但这种多光谱成像特性使其广泛应用于精准农业、矿产检测和城市发展规划[1]–[3]。高光谱图像分类（HIC）作为遥感领域的一项基本和重要任务，一直是社区研究的一个活跃课题。大量的光谱波段意味着更多的分类信息，但这可能导致所谓的休斯现象

王恒和王立国在哈尔滨工程大学信息与通信工程学院工作，黑龙江，哈尔滨，150001中国邮箱：wanghenghrbeu@hrbeu.edu.cn；王立国@hrbeu.edu.cn。（通讯作者：王立国）
稿件于2023年5月12日收到。

以及更高的计算成本。为了减少频谱冗余，采用了传统的主成分分析（PCA）[4]、频带选择[5]和基于三维卷积神经网络（CNN）[6]的自动编码器等降维方法。这些方法都是有效的和成功的，但它们在一定程度上不可避免地丢失了一些信息。最近的一些研究表明，[7]–[10]保留了全条带在设计分类模型时提供的信息，因为这些模型具有足够大的模型容量。

HIC面临的两个最严重的挑战是：1)较高的类内差异和类间相似性；2)标记训练样本的缺乏。这里的第一个挑战是，随着深度学习（DL）在计算机视觉（CV）和自然语言处理（NLP）领域的快速发展和持续发展，许多旨在提取深度可辨别特征的优秀深度模型已经被开发出来。罗伊等人。[4]利用三维和二维CNN的顺序排列，构建了一个简单而高效的杂化SN（混合光谱CNN）网络。不久之后，罗伊等人。[10]设计了一种新的基于注意力的自适应谱空间核残差网络（A2S2K-RN），该模型的启发性部分是使用1 1 7和3 3 7个三维卷积核联合学习自适应谱空间特征。虽然cnn中的池化操作可以有效地检测某一特征的出现，以便更准确地建模新出现的特征之间的相对位置关系，帕莱蒂等人。[11]首次引入了胶囊网络（CapsNet）进入HIC，以实现更准确的分类。3D与Tang等人提出的空间光谱注意网络（3DOC-SSAN）进行八度卷积。[12]与查克拉博蒂等人提出的光谱网具有类似的优势。[13]. 前者使用4个三维八度卷积块在空间域内捕获高频和低频信息，后者对输入patch进行4级小波分解，探索高频和低频空间信息。从将光谱序列分组为不同时间步长的输入信号的角度，研究了许多基于递归神经网络（RNNs）的深度模型，并将其应用于HIC任务[8]、[14]、[15]。Hong等。[16]提出了一种小批图卷积网络（miniGCN），使GCN可以在大规模遥感数据集上以小批的方式进行训练。随着变压器在NLP领域的出现及其快速优势[17]，多索维茨基等人。[18]指出

2023-3043267©2023 IEEE

授权许可使用仅限于：沈阳航空航天大学。已于2023年9月10日12:14:07 UTC从IEEE Xplore下载。限制适用。

© 2023 IEEE。允许个人使用，但再版/再发需要IEEE的许可。看到<https://www.ieee.org/publications/rights/index.html>获取更多信息。

一个图像相当于16 16个字，并提出了视觉变压器（ViT）。受ViT对全球背景的强大学习能力的启发，他等人。[19]将从类似VGG的深度CNN中提取的底层特征发送到创新设计的密度变压器模块，以学习远程光谱依赖关系。Hong等。[20]设计了一个名为频谱器的主干网络和一个跨层自适应融合（CAF）模块，该模块用于在不同深度融合变压器的输出，以改善不同层之间的信息交换。张等人。[6]引入了MobileViT来构建一个轻量级的分类模型。赵等人。[21]建立了一个卷积变压器网络（CTN）来捕获HSI补丁的局部-全局特性。王等人。[22]提出了一种基于变压器编码器和解码器（DCN-T）的双上下文网络，该网络利用变压器有效编码基于相似聚类的同质区域内的区域适应和全局聚合空间上下文。太阳等。[23]构建了一个光谱-空间特征标记化变压器（SSFTT）模型来捕获光谱-空间特征和高级语义特征。拉维库马尔等人。[24]构建了一个基于深度矩阵胶囊网络（DMCapsNet）的概念和期望最大化（EM）路由算法的新模型。该模型首先通过将主胶囊乘以视点不变变换矩阵来计算卷积胶囊的投票，然后计算了卷积的高斯混合分布胶囊的分配概率和投票。最后，基于投票和分布的相似性，将主胶囊聚合成卷积胶囊，实现部分-整体关系的建模。EM算法[25]作为模型的核心算法，根据M步的激活、投票和分配概率更新高斯分布的参数，根据M步投票与更新高斯分布的相似性更新分配概率

E.

对于第二个挑战，即很少有标记的训练样本。虽然网格搜索可以用来训练一个相对最优的模型，但这个问题极大地限制了各种深度模型的存在，以充分发挥其潜力。半监督HIC作为一个研究热点被广泛研究，其核心问题是如何充分利用和利用大量未标记样本[26]–[29]的价值。吴等人。[30]提出了最经典的半监督分类方法使用伪标签，其中，一个深度卷积RNN（CRNN）模型预先训练使用所有的训练数据及其伪标签被定义为集群分配生成的约束狄利克雷过程混合模型（C-DPMM），然后微调使用有限的可用性标签数据。方等人。[31]提出了一个协作学习框架，首先训练一个轻量级的CNN模型标记样本，然后使用近似排序聚类（AROC）集群所有样本的深层特征，最后未标记样本的集群标签符合预测标签用于扩大训练集进行迭代训练。西德加等人。[32]开发了一个概率伪标签

半监督HIC的生成框架。该框架假设（某一类）标记样本的编码特征服从高斯分布，然后使用最大似然估计（MLE）来估计该分布并分配相应的类标签。当为未标记样本分配标签时，该框架使用一个置信区间来度量未标记样本的编码特征与标记样本的分布之间的相似性得分。曾等人。[33]提出了一种图聚类卷积网络，它利用双聚类构造图，帮助GCN找到可靠的依赖关系。贾等人。[34]提出了一种新的半监督暹罗网络，该网络集成了一个用于学习所有样本的潜在深度特征的三维自动编码器和一个用于修正无监督特征的暹罗网络，以减少类内距离和增加类间距离。尽管上述半监督的分类方法都是基于深度模型的计算效果好，计算复杂度可接受，分类精度往往不令人满意。

当有限的训练样本可用时，另一种解决方案是引入生成式对抗网络（GANs）[35]–[41]。朱等人。在基于辅助分类器GAN（AC-GAN）模型的HIC中使用GAN来缓解过拟合，提高分类性能[42]。詹等人。[43]为半监督HIC设计了一种名为HSGAN的一维GAN。利用未标记样本对一维GAN进行训练，整个模型训练完成后，将softmax层替换鉴别器的顶层，将网络转换为分类器。冯等人。[44]设计了一种基于协同学习和注意机制的启发性卷积GAN，利用鉴别器提取的真实样本的深度特征，引导生成器生成高质量的样本。为了提高样本的多样性和实现高质量的恢复变分自动编码器（VAE），许多组合模型结合这两个生成模型的优点已经提出，如半监督变分干（SSVGAN）[45]，条件VAE敌性训练过程（CVA2E）[46]，基于自我关注的条件VAE干（神圣）[47]，变分干与交叉空间和光谱相互作用（CSSVGAN）[48]，等等。王等人。[49]利用CapsNet联合卷积长短期记忆（ConvLSTM）模块设计了鉴别器，并构建了HIC的ACGAN模型。由于CapsNet具有准确定位和识别特征的能力，我们也可以在文献[50]和[51]中看到结合CapsNet和GAN的模型。其他一些有趣的基于GAN的分类模型也可以在[52]–[56]中找到。实际上，GAN模型中的生成器是一种有效的正则化方法。然而，GAN的训练是不稳定的，即难以达到纳什均衡，容易出现模型崩溃的[39]、[57]、[58]。在有限的训练样本、集成学习[59]、迁移学习[60]和领域自适应（或领域泛化）[61]，[62]也表现良好。传统的集成学习框架需要训练多个基础网络，这大大增加了计算成本。最近，徐等人。[63]提出了一种鲁棒的自集成网络（RSEN）

对于HIC，它包括两个子网：基本网络和结构相同的集成网络。此外，为了使集成网络预测的未标记样本的伪标签更加可信，作者还提出了一种一致性滤波策略。迁移学习的有效性在于，在辅助域学习到的先验信息可以用来帮助学习目标HSI域的新特征问题。例如，Hao等人。[64]使用预先训练好的ResNet（残差网络）知识，在他们提出的几何感知深度RNN（Geo-DRNN）中表示几何感知的ResNet损失，以便更好地编码复杂的几何。域自适应（DA）旨在减少特征级的域偏移，并学习域不变模型，如Zhang等人提出的TSTnet（拓扑结构和语义信息传递网络）。[65]，它使用图最优传输（GOT）来对齐源域（SD）和目标域（TD）子图之间的拓扑关系，以协助分布对齐。领域泛化（DG）比DA更具挑战性，因为它只在SD数据上学习模型，而在训练阶段不访问TD。最近，张等人。[66]提出了一个跨场景HIC的DG框架，称为SDEnet（单源域扩展网络）。该框架的生成器使用一个语义编码器来确保扩展域（ED）包含各种td的特定信息，并使用一个变形编码器来提取域不变特征。该框架的鉴别器使用监督对比学习来学习类之间的域不变表示SD、ED及其随机线性组合中间域（ID）。

主动学习（AL）从另一个角度考虑了只有少数标记样本时的HIC问题。与使用伪标签的方法不同，AL通过查询信息最丰富的样本并补充专家注释[67]–[69]来构建最优训练数据集。Cao等[2]提出了一种主动深度学习方法，选择具有高预测熵（EP）值或低优优（BvSB）值的候选样本进行主动标记，在迭代训练中利用马尔可夫随机场（MRF）来增强类标签平滑性。刘等人。[70]提出了一个基于加权增量字典学习方案，项目初始训练数据的最高输出的深度信念网络（DBN）（无监督学习阶段）作为初始字典原子，然后选择一些最好的样本标记在每个迭代根据代表测量基于稀疏表示和不确定性测量基于信息熵。豪特等人。[71]利用所开发的贝叶斯cnn（B-CNN，退出网络）预测的平均后验概率来测量候选池中样本的不确定性，并详细分析和比较了不同AL采集函数的性能。Patel等人。[72]提出了一种基于多样性和不确定性的AL选择方案，该方案结合了多样性标准和BvSB标准，积极选择信息样本。首先，他们使用K-Means算法对当前的训练样本进行聚类，然后计算与不确定样本的距离

由BvSB确定的样本到每个聚类中心，并以距离最小的簇中心作为这些样本的聚类标签。最后，选择到指定聚类质心的距离大于平均距离的样本作为具有多样性的样本。王等人。[73]引入了随机多图（RMG）算法，并利用锚定图算法（用于构造和学习图）对BvSB确定的不确定样本的标签进行预测，而不是专家标记。王等人。[74]提出了一个面向特征的对抗（FAAL）框架，使用干学习高水平的分布特征标记样本的卷积模块的分类器，然后使用一个训练有素的鉴别器作为收购启发式判断一个未标记的样本之前表示。Yoo等人。[75]提出了一种新的选择策略，即损失预测，它利用深度模型的多个中间层的知识来预测未标记输入的目标损失。雷等人。[76]以ResNet18作为基本学习器，以损失预测作为查询函数，研究HIC的主动深度学习。刘等人。[77]利用深度密集连接的CNN的中间层构建损失预测模块。

一般来说，许多标记的样本可能是冗余的和不必要的，因为可能有重复的样本提供相似或相同的信息[34]。虽然AL依赖于先验的专家知识或需要一定的人工注释成本，但它可能是解决基于深度模型的HIC问题的最实用、最有效、最有效的方法。基于不确定性的AL方法仍然是主流，因为它们可以在多次迭代中快速提高分类器的性能。但是，本阶段的这些方法大多没有考虑到除文档[72]之外需要主动标记的样本的多样性，除了文档[74]外，大多数都没有考虑到专家注释的预算。为了解决上述问题，本文开发了一个具有HIC实际应用意义的协作AL（CAL）框架。详细地说，该框架包括两个健壮的基本分类器

CAL和一个CAL计划。两个鲁棒的分类器之一是CapsNet与ViT（CapsViT）相结合，它利用CapsNet在有限标记样本时准确分类的能力，以及ViT学习特征的全局相关性的能力。另一个基本分类器的灵感来自于Hinton [78]提出的GLOM系统。这是一个强大的系统，但对于构建这个系统的一些想法并不能使它完美地工作，我们将其改进为应用于HIC，并将其命名为CapsGLOM，作为一个鲁棒的基本分类器，它可以通过构造一个解析树来学习每个位置的全局依赖性。此外，还提出了一种同时考虑不确定性、多样性以及专家注释预算的简单有效的CAL方案。

综上所述，本研究的动机是为了解决基于不确定性的AL方法没有考虑到积极选择样本的多样性和专家注释成本的问题这一问题，主要贡献体现在以下四个方面。

1) 一种新颖的协作主动学习（CAL）框架

是为标记样品很少的HIC而设计的。该框架由两个鲁棒的基分类器和一个CAL方案组成，该方案同时考虑了所选信息样品的多样性和专家注释的预算。

- 2) CapsNet与ViT的结合是该CAL框架中的两个基本分类器之一。它充分利用了CapsNet准确定位和识别特征的能力，以及ViT在全球相关性学习中无与伦比的能力。
- 3) 在改进后的GLOM系统CapsGLOM的基础上，提出了一种解释能力较强的模型作为另一种基分类器，并首次应用于HIC系统中。
- 4) 我们通过在四个基准的HSI数据集上进行严格的实验，验证了我们提出的框架的有效性和优势。同时，在具有足够训练样本的监督分类场景下，验证了所提出的CasViT和CapsGLOM与目前先进的深度模型相比，也获得了具有竞争力的性能。

后续文本组织如下。第二节简要介绍了相关方法，包括CapsNet、ViT和GLOM系统。第三节

编写我们提出的CAL框架，包括已开发的CapsViT、CapsGLOM和CAL方案。第四节验证了该框架在四个HSI数据集上的有效性和优越性。第五节讨论了已开发的CapsViT和CapsGLOM的有效性、灵敏度和复杂性。第六节总结了全文的工作，并对未来可能的研究方向提出了一些建议。

II. 相关工程

本节简要回顾了与AL方法、CapsNet、ViT和GLOM系统相关的工作，并为下一节中描述的推荐的CAL框架提供了基础。

A. AL方法

主动学习（AL）是一种提倡数据效率的强大技术，其目标是以最少的获取步骤（oracle查询）达到一定水平的预测精度。AL方法大致可以分为四种类型：基于不确定性的[67]、基于表示的[67]、基于性能的[67]和基于损失预测的AL [75]。不同的AL方法使用不同的查询策略，但这些启发式的共同目标是搜索信息最丰富的样本。

基于不确定性的方法选择深度模型在两个或两个以上类之间存在犹豫的样本，最常用的查询策略包括BvSB、EP和蒙特卡罗（MC）dropout等。尽管基于类型的方法提供了主要的优势，例如它对大数据集的可伸缩性和它少数的超参数，但如果数据集包含每个样本的许多（关闭的）副本，它们会不惜代价选择单个信息样本的所有副本

其他信息样本的效率，从而浪费了数据效率。这里应该提到的是，由Kirsch等人提出的BatchBALD（基于分歧的批处理贝叶斯主动学习）。[79]属于基于不确定性的方法，获得的点是联合信息，而不是单独信息（i.e. 在一个批处理内加强多样性）。然而，BatchBALD扩展到基于贝叶斯神经网络（BN）的BALD，而我们的工作集中在BP（反向传播）神经网络，其中参数被视为点估计而不是分布。

基于表示的方法试图查询点，以使所标记的点均匀地分布在整个数据空间中。这些方法中最具代表性的包括核心集方法[80]和变分对抗性主动学习（VAAL）[81]。其中，核心集方法等价于解决k-Center问题，即选择一批中心点（已标记点），使数据点（未标记点）与其最近的中心之间的最大距离最小化。k-Center问题是NPHard，但是利用贪婪算法可以有效地得到一个2-OPT解。贪婪算法的结果给出了很好的初始化，Sener等。[80]制定了一个混合整数程序（MIP）来改进2-OPT的解决方案。虽然核心集方法的结果令人印象深刻，但在使用大数据集时，计算每个候选点和每个训练点之间的距离太耗时，在实践中不有用。VAAL是一种新的方法，和关键VAAL VAE用于映射标记和未标记数据到一个潜在的特征空间和欺骗敌对网络（鉴别器）预测所有数据点，从标记和未标记池，标记池而鉴别器学习有效地估计的概率数据来自未标记的数据。VAAL的目标不再是明确地测量不确定性，而是查询将产生高不确定性的点，从而查询在标记集中没有很好地表示的点。虽然VAAL是有效的，但关于训练的稳定性和易用性仍然存在问题。

基于性能的方法试图通过可靠地估计分类器性能的改进来自动选择最佳查询策略。这些方法有两个主要的限制。首先，它们只能与现有的手工设计的启发式方法相结合，其次，从稀缺的注释数据中进行可靠的分类性能评估是有问题的。为了克服这些限制，科尼什科娃等人。[82]提出了LAL（学习主动学习）。LAL的本质是一个回归问题，也就是说，给定一个训练过的分类器和一个候选点，它估计了通过添加标签到该点可以期望的泛化误差的减少。具体来说，对于每个AL步骤t，LAL随机分割标记的点 L_t 进入一个训练集 L_{τ_i} 和一个测试集 U_{τ_i} 根据不同的分裂尺寸， $\tau_i = 2, \dots, T+2$ 。对于每个分裂，从U中采样M个不同的点 τ_i 并添加到 L_{τ_i} 然后，记录相应的分类器/数据点状态对。这个过程是重复的Q次，总共产生了QMT记录。接下来，训练一个回归因子，将测试性能的变化与分类器和新器的属性联系起来

添加数据点。最后，将得到的回归因子作为LAL策略，查询潜在误差减少幅度最大的数据点。虽然LAL方法的思想是好的，但它的复杂性往往是不可接受的，特别是当使用一个深度模型作为分类器时。

基于损失预测的AL方法简单但依赖于任务任务，并能有效地应用于深度模型。在这里，我们将简要介绍一下它，以便在接下来的实验部分中比较不同的查询策略。该查询策略通过一个损失预测模块估计未标记样本的损失，该模块连接到目标模型的几层，并考虑多层知识进行损失预测。Yoo等人。定位的损失预测模块的损失函数为

$$L_{\text{损失}}(\hat{l}^p, l^p) = \max(0, -\star(l_i, l_j) \cdot \hat{l}_i - \hat{l}_j + \xi) \quad (1)$$

其中 $l^p = (l_i, l_j)$ 表示训练样本的损失对 (x_i, x_j) ，由目标模型计算得出，以及 $\hat{l}^p = (\hat{l}_i, \hat{l}_j)$ 表示由损失预测模块计算出的损失， ξ 是一个预先定义的正边缘（通常为1），和 $\star(l_i, l_j)$ 是一个指示器函数

$$\mathbf{1}(l_i, l_j) = \begin{cases} +1, & \text{if } l_i > l_j \\ -1, & \text{otherwise} \end{cases} \quad (2)$$

当 $l_i > l_j$ ，对模块没有损失，只有当 \hat{l}_i 大于 $\hat{l}_j + \xi$ ，否则模块会有损失，迫使它增加 \hat{l}_i 和减少 \hat{l}_j 。对于这种启发式方法，一个样本的预测损失越大，分类器就越有可能对其进行错误分类。可以看出，该模块的损失的计算要求批大小应该为偶数。在接下来的实验部分中，我们将做一个小的修改，以满足批大小不是偶数的情况。

在早期的研究中，Demir等人。[83]结合了不确定性和多样性准则，研究了利用支持向量机进行遥感图像分类的不同AL技术。具体来说，他们首先根据不确定性准则过滤未标记的样本，然后使用一个设计来评估多样性的查询函数来选择样本进行标记。所提出的CAL框架也考虑了不确定性和多样性标准。但是，我们还没有定义任何多样性标准，最重要的是，我们强调了专家注释的预算。

B. 胶囊网络

在CNN中，虽然池化操作可以通过“平移不变性”来识别特定的特征，但并没有考虑这些特征之间的空间层次关系。与CNN的不变性相比，Hinton提出了“等方差”的概念，即在对一个对象的表示进行变换后，其结果等同于变换后的对象的表示。基于“等方差”的思想，CapsNet精确地准确定位和识别特征。图1为手写字符分类的经典字幕网络示意图。

首先，将输入的手写字符输入到前两个卷积层中，所采用的激活函数是整流线性单元（ReLU）。此时，输出是一个形状为256 6 6的特征立方体。每个胶囊是由一组标量神经元组合组成的“向量神经元”，因此，从构建尺寸为8的主胶囊（primarycap）的角度来看，第二卷积层的输出可以视为32个8 6 6形状的特征立方体的堆栈。每个特征立方体由36（6 6）个胶囊组成，因此32个特征立方体总共包含1152个主胶囊。

每个胶囊向量反映了一个特定实体的表示。向量的模具描述了实体存在的概率，其方向对应于实体的一般姿态。为了使胶囊向量的模具更好地表示实体的存在，有必要应用一个非线性挤压函数，以确保向量的方向保持不变，同时，将模具的值压缩到[0 1]区间。这是图中所示的南瓜函数的功能。1、其公式如下

$$\text{Squash}(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|} = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \hat{s}_j \quad (3)$$

在哪里 $\hat{s}_j = \frac{s_j}{\|s_j\|}$ 是输入向量 s 的单位向量 j 。

通过压缩函数的主胶囊输出表示为 s_i ($i = 0, 1, 2, \dots, 1151$)。

CapsNet使用许多主胶囊来表示所有可能的浅层实体，每个浅层实体都可以对类别胶囊层中的实体进行类别预测。胶囊网中的类别胶囊层统称为数字胶囊（数字胶囊）层。有10个胶囊单元，用 v 表示 j ($j = 0, 1, \dots, 9$)，正好对应于10个需要识别的数字。自然地，从主胶囊到数字胶囊的预测是通过使用全连通层来实现的，并存在一个耦合系数 c_{ij} 在每个主胶囊 u 之间 i 和每个数字胶囊 v_j ，这是一个概率分布，表示主胶囊对数字胶囊的支持度。对于每个主胶囊，其与所有数字胶囊的耦合度之和为1

$$\sum_i c_{ij} = 1 \quad (4)$$

数字胶囊 v_j 是通过将每个主胶囊相对于数字胶囊的预测向量乘以相应的耦合系数 c 得到的 c_{ij} ，然后把它们加在一起。当然，不要忘记使用压缩功能进行压缩激活。整个过程可以表示为

$$v_j = \sum_i c_{ij} (W_{ij} u_i) = \sum_i c_{ij} j | \hat{u}_i \quad (5)$$

在哪里 \hat{u}_i 表示主胶囊 u 的预测向量 i 相对于数字胶囊 v_j 和 c_{ij} 表示它们之间的耦合系数。有趣的是， c_{ij} 不是CapsNet的可训练参数，而是一个由基于前向传播一致性的动态路由协议迭代更新的参数。具体来说，它是

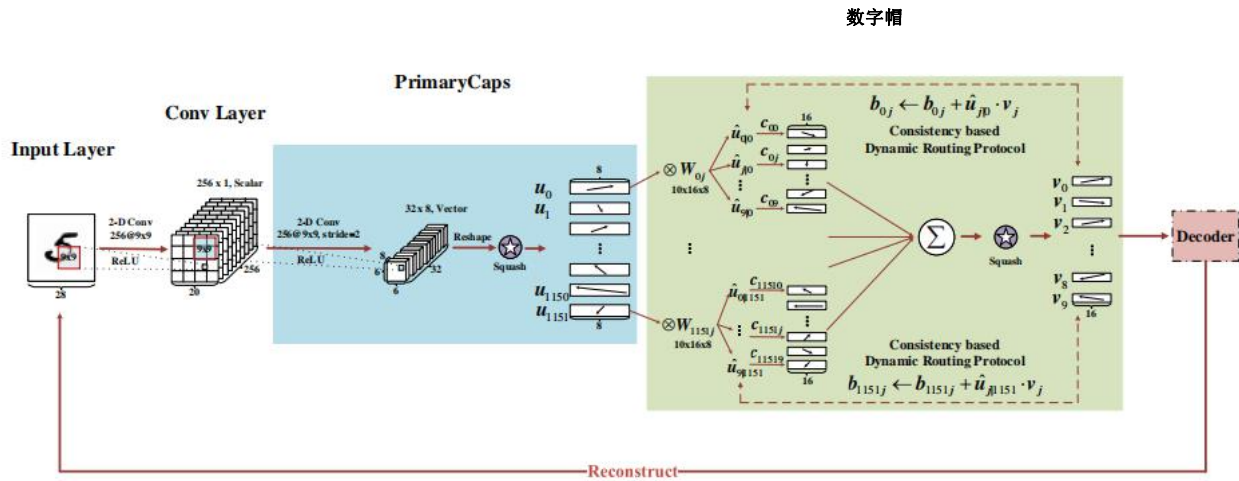


图1. 手写字分类的胶囊网络。

通过计算一致性 b 来确定 i, j 在数字胶囊 v 之间 j 和预测向量 \hat{u}_{ij} ，然后通过softmax函数，可以表示为

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (6)$$

式中， k 表示数字胶囊的类别数。在初始时间，将数字胶囊与预测向量的一致性初始化为零向量，即 $b_{ij} = 0$ 。然后，在每个路由寻址过程中，一致性 b_{ij} 是否以累积的方式进行更新

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{ij} \cdot v_j \quad (7)$$

在哪里 $\hat{u}_{ij} \cdot v_j$ 表示预测向量和数字胶囊之间的点积，即它们之间的所谓一致性。

为了控制学习过程并确定图像中的多个数字（或类别），CapsNet计算每个数字胶囊 v 的边际损失 j 并将其总结为模型优化的目标函数。对于任何数字胶囊 v_j ，其保证金损失可表示为

$$L_j = T_j \max(0, m^+ - |v_j|)^2 + \lambda (1 - T_j) \max(0, |v_j| - m^-)^2 \quad (8)$$

其中， T_j 是类别 j 的指示符函数，当且仅当类别 j 出现时， $T_j=1$ ，否则， $T_j=0$ ，和 $|v_j|$ 表示 v 的模长 j ，这实际上是类别 j 存在的概率。 m^+ 为上限边缘阈值（通常为0.9），当类别 j 出现时，如果其预测概率小于0（假阴性），则该模型将受到惩罚。同样， m^- 是较低的边缘阈值（通常为0.1），当第 j 类不出现时，如果其预测概率大于0.1（假阳性），该模型也将受到惩罚。 λ 是用于调整惩罚假阴性和假阳性的两个损失的比例的稀疏系数。为了防止在学习开始时没有出现的数字胶囊的过度压缩，通常取它为0.5。

图中的解码器。1可以是任何一个简单的网络，用于解码与中的标签对应的数字胶囊

将样本重建输入样本。因此，为了使CapsNet也具有高级的编码和解码功能，整个网络的损失函数可以写为

$$L = \frac{1}{N} \sum_{n=1}^N \left[\left(\sum_j L_{nj}^{\text{margin}} \right) + \alpha L_n^{\text{recon}} \right] \quad (9)$$

其中 L_{nj}^{margin} 表示训练样本 n 相对于数字胶囊 v 的裕度损失 j ， L_n^{recon} 表示重构损失，可以用均方根误差（RMSE）计算， α 表示正则化系数，由于网络专注于分类任务，通常取一个较小的值，如0.01， N 表示小批量的大小。

C. 视觉变压器

变压器[17]显著提高了神经元机器翻译（NMT）的当前水平，没有使用任何循环层或卷积层，只使用注意机制。鉴于变压器能够编码同一句子中每个单词和其他单词之间的关系，并更多地关注最相关的单词，多索维茨基等人。[18]提出的视觉变压器（ViT），模型如图所示。2.

标准变压器接收一个一维的令牌嵌入序列作为输入，以处理三维图像，ViT首先重塑输入图像 $x \in \mathbb{R}^{3 \times H \times W}$ （这里以自然图像为例）形成一系列扁平的补丁

$\mathbf{x}^n = (x_p^1, x_p^2, \dots, x_p^N)^T \in \mathbb{R}^{N \times (3 \cdot P^2)}$ ，其中 (H, W) 为原始图像的分辨率， (P, P) 为每个patch的分辨率， $x_p^n \in \mathbb{R}^{3 \cdot P^2}$ 是第 n 个扁平的补丁，和吗

$N = HW/P^2$ 是生成的补丁的数量。接下来，这些扁平的补丁 x ($n = 1, 2, \dots, N$)被馈入一个没有激活函数的线性层，以得到底层嵌入。这些底层嵌入需要在输入变压器编码器之前添加一个位置嵌入。这允许模型访问每个补丁的位置，这是必要的，因为编码器的“多头注意”层不考虑补丁的顺序或位置，而只考虑它们之间的关系。若要执行分类任务，则需要执行一个类别

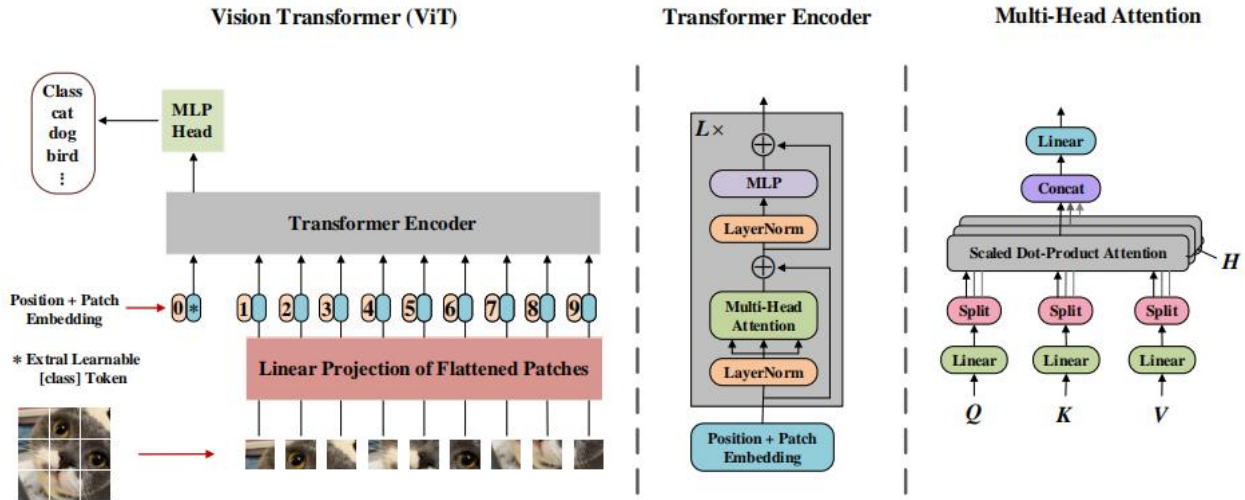


图2. 视觉变压器 (ViT)，用于自然图像分类。

需要在所有补丁的嵌入之前插入标记，并分配一个位置嵌入。

变压器编码器由多头自注意 (MSA) 层和小型多层感知器 (MLP) 组成，两者都是基于残差结构的。在每个块之前，应用Layernorm (LN) 对每个位置的嵌入特征进行规范化。ViT想要的是可学习的类标记的最佳表示。从变压器编码器的工作原理来看，得到最优表示的过程就像查找由它创建的字典一样。但实际上，编码器并没有一个特定的令牌来表示键，它有一个对这些概念的向量化表示。因此，用于搜索的键（称为查询，Q）与字典中的任何键(K) 都不完全匹配。解决方案是计算查询和字典中每个键之间的相似度，然后使用softmax函数将这些相似度分数转换为其之和为1的权重。然后，该模型可以计算出相应值的加权和 (V)

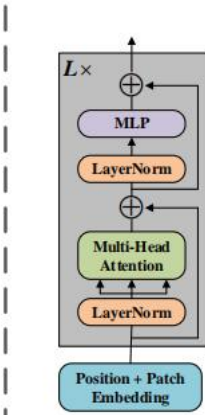
$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right) \mathbf{V} \quad (10)$$

其中，Q、K、V $\in \mathbb{R}^{(N+1) \times d}$ 分别表示查询、键和值，d表示每个查询的维度，每个键和每个值， $\mathbf{Q}\mathbf{K}^T \in \mathbb{R}^{(N+1) \times (N+1)}$ 表示每个查询/键对的相似度得分，是一个缩放因子。最终的输出形状是 $[N+1, d]$ ，这意味着每个查询都有一行，每一行表示查询结果。

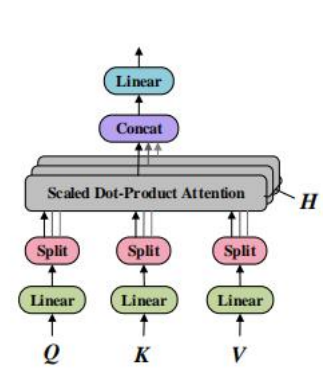
一个位置的最优表示是几个相关位置的表示的线性组合。此外，每个补丁的表示编码了许多不同的特征，如果只使用一个注意层，那么就只能一次查询所有的特征。MSA允许模型将许多不同的补丁特性投射到不同的子空间中，每个子空间都专注于补丁特性的一个子集。

在编码器顶部的小MLP包含两个线性层，使用GELU（高斯误差线性单位）作为

Transformer Encoder



Multi-Head Attention



激活函数GELU激活函数的表达式如下

$$\text{GELU}(x) = x(x) \frac{x}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt \quad (11)$$

其中(x)为标准高斯分布的累积分布函数。最后，ViT的整个计算过程可以用以下公式来总结

$$(12)$$

$$(13)$$

$$\mathbf{z}_0 = (\mathbf{x}_{class}, \mathbf{x}_p^1 \mathbf{W}, \mathbf{x}_p^2 \mathbf{W}, \dots, \mathbf{x}_p^N \mathbf{W})^T + \mathbf{W}_{pos} \quad (14)$$

$$\mathbf{z}_l' = \text{MSA}(\text{LN}(\mathbf{z}_{l-1})) + \mathbf{z}_{l-1} \quad (15)$$

$$\mathbf{z}_l = \text{MLP}(\text{LN}(\mathbf{z}_l')) + \mathbf{z}_l'$$

其中 $\mathbf{y} = \text{MLP}(\text{LN}(\mathbf{z}_L^0))$ 投影

层的权重，D为投影维数， $\mathbf{x} \in \mathbb{R}^D$ 表示可学习的类标记， $\mathbf{W}_{pos} \in \mathbb{R}^{(N+1) \times D}$ 表示可学习的位置嵌入，L表示共有L个编码器堆叠，L个= 1, 2, 3, ..., L表示编码器编号， \mathbf{z}_1 表示第1个编码器z的输出特性 \mathbf{z}_L^0 表示由顶层编码器学习到的最佳类标记表示，y表示最终的预测概率，用于计算交叉熵来训练整个模型。

D. GLOM系统

CapsNets的一个基本缺点是，它们使用一组预先分配的胶囊来建模一组可能的部分。这使得我们很难判断汽车的前灯和眼睛是否真的是不同的部分。如果它们被同一个胶囊模拟，胶囊就不能预测整体的身份。如果它们被不同的胶囊建模，它们相对于它们整体的相似性就无法被捕获。基于上述分析，Hinton提出了一个名为GLOM的虚拟系统来表示神经网络中的部分-整体层次结构。该系统通过预先分配一个普遍存在的位置，同时学习补丁在最低、子部分、部分、对象和场景级别的嵌入

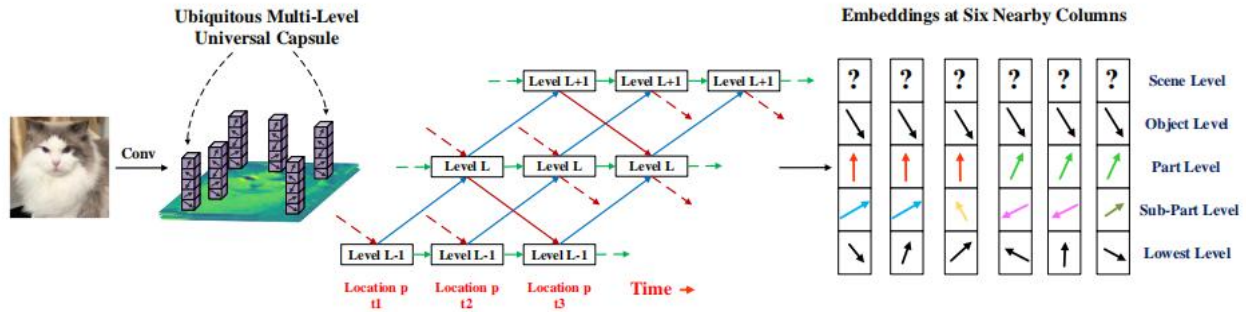


图3. GLOM系统的概述。

多层次的通用胶囊在每个图像补丁的位置。这种无处不在的多层次的通用胶囊可以被比作一个基因序列。每个细胞（这里的贴片）对每个基因都有不同的表达强度。GLOM系统利用上述思想，并使用由不同级别的相同向量组成的岛屿来表示解析树中的一个节点。图3为该系统的基本示意图。

GLOM系统的核心组件是一个基于时间步长的多层次嵌入学习模块。对于只有一个突出对象的静态图像¹，在一定水平上的嵌入应该随着时间的推移逐渐稳定，产生几乎相同向量的不同岛屿。随着嵌入级别的增加，这些岛应该逐渐变大，从而在解析树中形成一个节点。解析树的根节点对应于对象级的嵌入，并且所有的位置（属于同一对象）都应该学习完全相同的对象级向量。此外，顶级的嵌入，即。场景级嵌入，包含对象相对于摄像机的三维姿态信息。

在GLOM中，一个补丁的多级嵌入形成了一个列。对于每一列，在每个离散时间 t ，一个级别的嵌入通过四个贡献的平均值更新。其中三种如图3所示：1）自上而下神经网络得到的预测（红色箭头）；2）自下而上神经网络得到的预测（蓝色箭头）；3）同一水平的嵌入（绿色箭头）。自上而下和自下向上的网络是两个不同的网络，它们跨地点和时间共享。但它们可以在不同的级别之间有所不同。图4中为这两种网络的工作原理图。其中，自底向上的网络只是简单地预测了高级嵌入。然而，自顶向下的网络预测了低级别的嵌入，这与补丁的绝对位置或相对位置有关，因此应该添加关于该位置的信息。

除了图3中所示的三种贡献外，对于每个离散时间 t 和每一个列，每个级别的嵌入更新也会受到附近列中同一级别的嵌入的影响。对于每个级别，不同列中的嵌入之间的交互是

¹GLOM还可以解析具有多个对象的视频流和静态图像中的部分-整个层次结构。

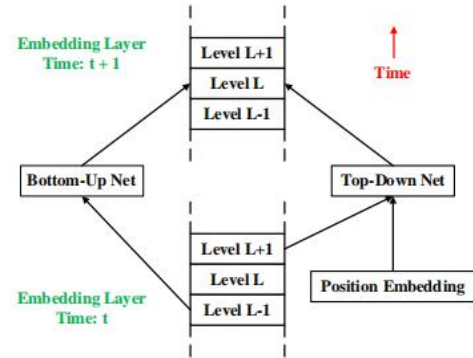


图4. 对单个列的两个相邻级别的GLOM之间的自上而下和自底向上的预测。

通过一个注意力加权的局部平滑器来实现。让 $E = (e_1, e_2, \dots, e_N)^T \in \mathbb{R}^{N \times d}$ 表示不同列中相同级别的嵌入，其中 N 表示列的总数（patch）， d 表示每个嵌入的维数。这些嵌入之间的相互作用可以表示为

$$\text{Interact}(E) = \text{softmax} \left\{ \frac{E [\text{norm}(E)]^T}{\sqrt{d}} \right\} E \quad (16)$$

其中，“norm”表示向量归一化操作。我们可以看到，在不同列中同一级别的嵌入之间的相互作用实际上计算了一个注意加权平均，这借鉴了变形器中的比例点积注意。因此，上述第四个贡献可以简单地总结为前一次在不同列中在同一层次上的嵌入的注意加权平均值。

III. 提出的框架

基于少数训练样本下的主动学习的HIC是本研究的基线。我们开发了一个协作主动学习（CAL）框架，其中包括两个新颖的和

优秀的基础分类器和一个CAL方案，考虑-其中包括积极选择的样本的多样性和专家注释的预算。第一个基分类器是CapsNet与ViT的结合，命名为CapsViT，第二个是在GLOM系统上改进的，命名为CapsGLOM。接下来，我们将详细介绍所提议的框架的组件。

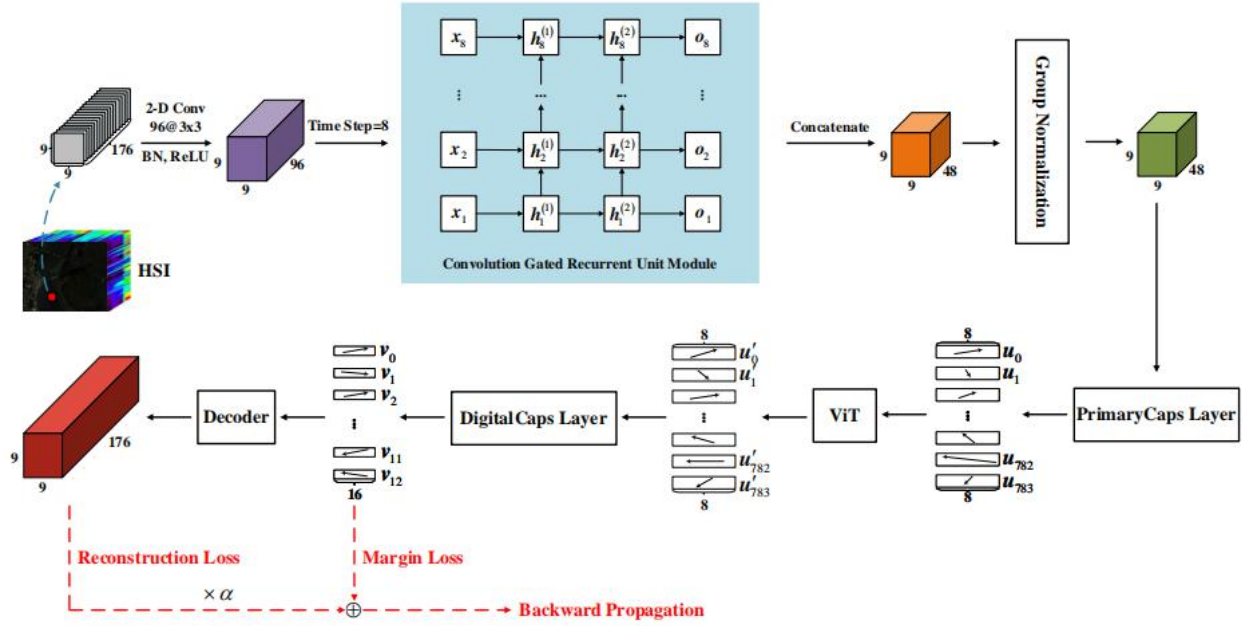


图5. 为HIC提出的CapsViT。实际的培训过程是小批量进行的，为了演示，这里只以一个样本为例。

A. CapsViT

CapsNet不仅可以检测特定特征的出现，还可以准确定位该特征在单元超球嵌入空间中的位置。然而，CapsNet并没有考虑到初级胶囊之间的全局相关性。学习最佳的“初级”胶囊²通过探索每个位置的主胶囊之间的全局相关性，这是我们将ViT引入胶囊网的动机。⁵在图中，我们以肯尼迪航天中心（KSC）的数据集为例，展示了所提出的CapsViT的完整图。

CapsViT是一个不需要降维预处理的端到端模型。如图5所示，为了充分利用光谱特性，在CapsViT的开始阶段，我们使用了一个卷积层和一个具有两个隐藏层的卷积门控递归单元（ConvGRU）模块来学习局部和非局部跨通道光谱依赖关系。以KSC数据集（有176个光谱波段）为例，假设训练样本x的空间大小为99，i.e., $x \in \mathbb{R}^{17699 \times \times}$ 。一个小批的训练样本x批量 $2 \times \mathbb{R}^{bs \times 17699 \times \times}$ （其中bs表示批处理大小）被发送到第一个卷积层

$$f_0 = \text{ReLU}(\text{BN}(W_0 * x_{\text{批量}} + b_0)) \quad (17)$$

其中，*表示二维卷积运算， W_0 和 b_0 分别为权重和偏差，BN表示批处理归一化。在这一层有96个卷积核，接受野（RF）的大小为3 3。 \times

第一个卷积层f的输出 $o_2 \in \mathbb{R}^{bs \times 9699 \times \times}$ 然后沿着通道尺寸均匀地分成8个部分，并在每个时间步长发送到ConvGRU模块作为输入到

²这里的初级胶囊指的是较深的初级胶囊，它已被ViT提炼。

学习局部和非局部的跨通道光谱依赖关系。⁶ ConvGRU模块由两层ConvGRU的结构如图第一层的ConvGRU，每个门控制器有12个卷积核。对于第二层的ConvGRU，每个门控制器中有6个卷积核。让 $x_1, x_2, \dots, x_8 \in \mathbb{R}^{bs \times 1299 \times \times}$ 分别表示每个时间步长的输入，然后第一层在每个时间步长的ConvGRU的输出可以用以下公式表示

$$\begin{aligned} z_t &= \sigma(W_{xz} * x_t + W_{hz} * h_{t-1}) \\ r_t &= \sigma(W_{xr} * x_t + W_{hr} * h_{t-1}) \\ g_t &= \tanh(W_{xg} * x_t + W_{hg} * (r_t \otimes h_{t-1})) \\ y_t &= h_t = z_t \otimes h_{t-1} + (1 - z_t) \otimes g_t \end{aligned} \quad (18)$$

其中 W_{xz} , W_{xr} 和 W_{xg} 是输入x的每个门控单元中的卷积核 m_t , W_{hz} , W_{hr} 和 W_{hg} 是状态h的卷积核 m_t , 表示元素级乘法, a表示s型激活函数, \tanh 表示 \tanh 激活函数。在每个时间步长y时，第一层的ConvGRU的输出 $y_1, y_2, \dots, y_8 \in \mathbb{R}^{bs \times 1299 \times \times}$ 用作第二层ConvGRU的输入。最后，我们在8个时间步长o中得到8个输出 $o_1, o_2, \dots, o_8 \in \mathbb{R}^{bs \times 699 \times \times}$ 。虽然 o_8 整合了o的信息 o_1, o_2, \dots, o_7 ，我们将它们连接起来，以充分利用浅光谱和深光谱依赖关系。然后，为了使学习过程快速、流畅，采用了分组参数为8个（对应8个时间步长）的组归一化方法。

接下来，将获得的特征发送到primarycap层，其中有128个RF大小为3 3的卷积核。 \times 为了调整所获得的初级胶囊的数量，在这一层中采用无填充的卷积。通过整合8个神经元到一个胶囊中

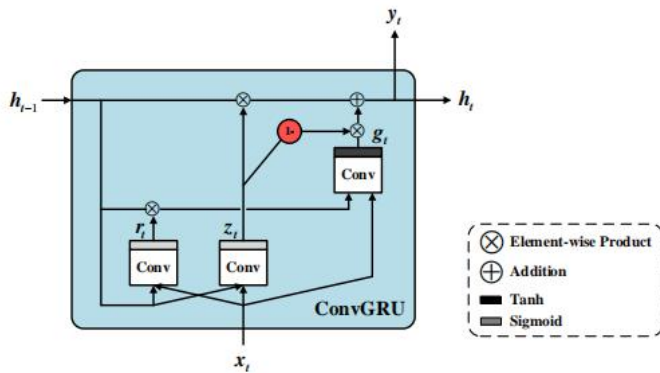


图6. ConvGRU的架构。

通道尺寸，我们可以得到784 (16 7 7) 主胶囊 $u \times 0; u_1; \dots; u_{783}$ 。主胶囊的模具代表某个实体特征（如猫眼）出现的概率，它的方向代表关于该实体的姿势信息（如猫眼的颜色或大小，等等）。每个实体特征都出现在单元超球体的特定区域内。因此，我们接下来将这些主胶囊喂入ViT，以了解主胶囊在每个“位置”的最佳表征，这考虑了所有主胶囊之间的全局相关性。我们将vit处理的初级胶囊表示为 $u; u; \dots; u_{783}$ ； u ，它们更具鉴别性，因此对分类有积极的影响。需要补充的是，在ViT模块中，查询、键和值的映射维度为16，并且在编码器顶部的MLP中的两个线性层的输出维度都为 8^3 。此外，有5个编码器堆叠，每个编码器使用4头注意。

最后，有13个数字胶囊 $v_0; v_1; \dots; v_{12}$ 在数字映射层中，对应于KSC数据集中的13种类型的地面对象。将训练样本标签对应的数字胶囊输入到解码器中，重构输入样本，解码器由一个全连接（FC）层和三个转置卷积层组成。在反向传播过程中，使用Eq作为损失函数来训练模型。9.9为了便于理解，表一列出了模型的参数配置和每个层的输入输出形状（Conv2d1代表第一个卷积层，ConvTrans表示转置卷积）。

B. CapsGLOM

GLOM系统是一个虚构而强大的系统。但是，构建这个系统的一些想法并不能使它运行完美，这主要体现在以下两个方面。

- 1) 属于同一对象的所有位置都应该被给予完全相同的对象级嵌入，但在实践中，很难区分哪些位置属于同一对象。

³MLP中的第二个线性层用于恢复原始维度。

- 2) 如果我们将普遍存在的多层次通用胶囊视为一个基因序列，那么属于不同对象的基因序列应该是不同的。

鉴于上述两个缺陷，我们开发了一个新的HIC深度模型，即CapsGLOM。在CapsGLOM中，我们在不同的位置学习不同的多层次嵌入，然后使用GLOM系统的多层次嵌入学习模块在每个层次上学习共识意见（即几乎相同的向量的岛）。7图中显示了该CapsGLOM的框架。

如图7所示，CapsGLOM的前半部分是相同的作为CapsViT，也就是说，我们使用一个两层的ConvGRU模块来学习局部和非局部的跨通道光谱依赖关系，然后进行组归一化。让 $f_{1/2} \quad R^{bs \ 4899 \times \times}$ 表示组标准化的输出。接下来，有两个分支，第一个分支使用一个完全连接的层来学习每个位置的底层特征，这些特征作为自底向上网络的输入，以预测最低级别的嵌入。需要明确的是，位置可以像像素一样细粒度，或者它们可以对应于更大的图像补丁。在CapsGLOM中，我们将位置定义为特征映射中每个像素的位置。对每个位置的底层特征的获取可以用以下公式来表示

$$f_{\text{在...下面}} = \text{GELU}(\text{注册表}(f_1)W_1 + b_1)) \quad (19)$$

其中，重塑 $(f_1) \quad R^{bs \ 8148 \times \times}$ 表示在 f 上的整形操作 1 ， $W_1 \quad R^{4832 \times}$ 和 $b_1 \quad R^{32}$ 分别表示全连通层的权重和偏差，和 $f_{\text{在...下面}} \quad R^{bs \ 8132 \times \times}$ 代表了学习的基础尺寸为32的特征。第二个分支使用二维结构通过卷积来学习5级嵌入 f 最低的， f 子部分， f 部分， f 目标， f 地点在每个位置。因为我们将每个“像素”定义为一个位置，所以像素级卷积， $i. \times e.$ ，这里使用卷积。每个位置的多级嵌入可以通过以下公式计算出来

$$(f_{\text{scene}}, f_{\text{object}}, f_{\text{part}}, f_{\text{sub-part}}, f_{\text{lowest}})^T = \text{LN}(\text{Reshape}(\text{GELU}(W_2 * f_1 + b_2))) \quad (20)$$

其中 W_2 和 b_2 分别表示像素级卷积的权重和偏差。因为每个位置有5个嵌入级别，每个嵌入级别的维数为32，所以像素级卷积的核数为160。将像素级卷积得到的特征图沿通道维度平均分为5个部分，然后重塑为所有位置的多层次嵌入。值得注意的是，虽然像素级卷积和全连接层是等价的，但上述两个分支中的GELU激活函数和LN操作的顺序是不相同的。换句话说，您可以将第二个分支中的GELU、Reshape和LN的顺序更为GELU、LN和GELU，这样就可以使用全连接层或像素级卷积来同时学习每个位置的底层特征和多层嵌入。但是我们构建多层次嵌入的方式借鉴了CapsNet的构建方式

表i

字幕的参数配置和各层的输入和输出形状

层 (类型)	核数	内核形状	步幅	填充	输入形状	输出形状
组2d1	96	3 3×	1	1	[- 1, 176, 9, 9]	[- 1, 96, 9, 9]
ConvGRU-2	12	3 3×	1	1	[- 1, 12, 9, 9]	[- 1, 12, 9, 9]
ConvGRU-3	6	3 3×	1	1	[- 1, 12, 9, 9]	[- 1, 6, 9, 9]
GN-4	-	-	-	-	[- 1, 48, 9, 9]	[- 1, 48, 9, 9]
PrimaryCaps-5	128	3 3×	1	0	[- 1, 48, 9, 9]	[- 1, 784, 8]
ViT-6	-	-	-	-	[- 1, 784, 8]	[- 1, 784, 8]
DigitalCaps-7	-	-	-	-	[- 1, 784, 8]	[- 1, 13, 16]
FC-8	-	-	-	-	[- 1, 16]	[- 1, 128]
ConvTrans2d-9	256	3 3×	1	0	[- 1, 128, 1, 1]	[- 1, 256, 3, 3]
ConvTran2d10	256	3 3×	2	1	[- 1, 256, 3, 3]	[- 1, 256, 5, 5]
ConvTran2d11	176	3 3×	2	1	[- 1, 256, 5, 5]	[- 1, 176, 9, 9]

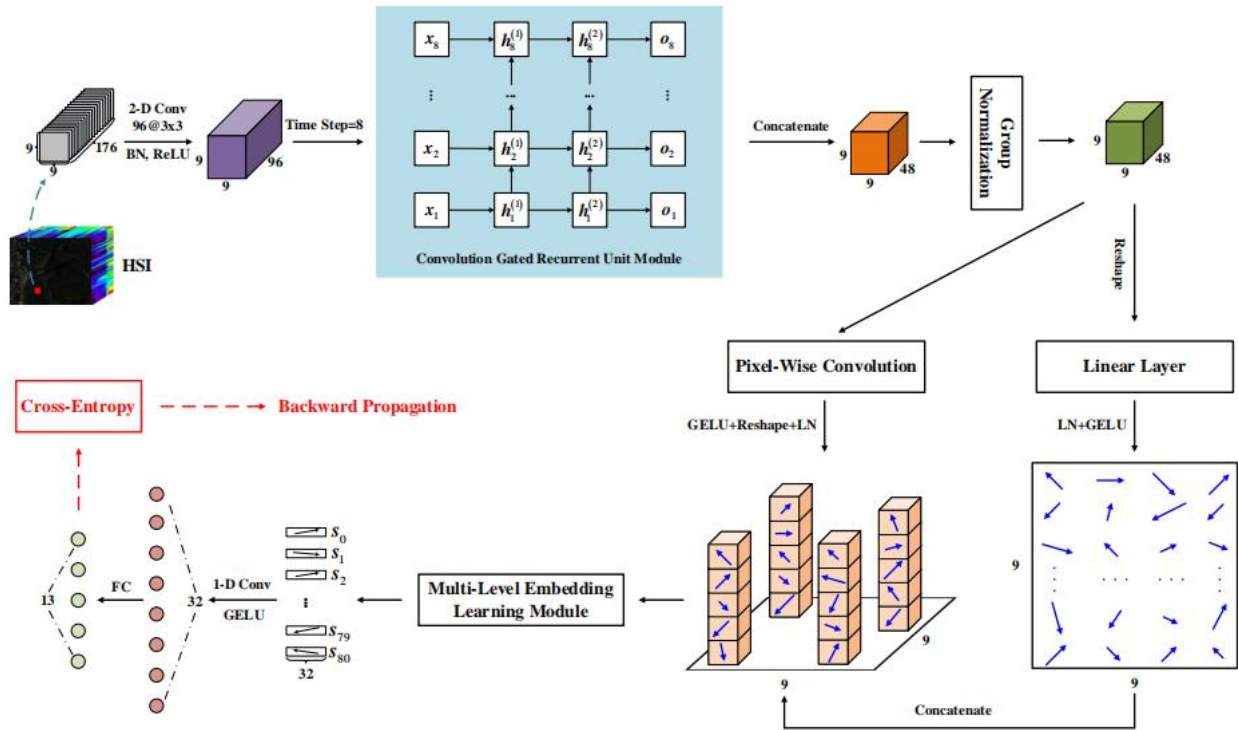


图7. 为HIC提议的CapsGLOM。实际的培训过程是小批量进行的，为了演示，这里只以一个样本为例。

一级胶囊⁴，当然，您也可以使用带有填充的步幅卷积而不是像素级卷积来构建多级嵌入。

然后，将每个位置的底层特征和多层次嵌入输入到GLOM系统的多层次嵌入学习模块中，以学习每个层次的共识意见。在多层次嵌入学习模块中，时间步长的设置会影响模型的学习速度和性能。在模型设计的过程中，许多实验表明，3个时间步长是最好的。此外，自底向上和自上而下的网络都是由GELU激活的两个线性层组成的，这两个线性层的输出维度分别为128和32。由这些预测的每个级别上的嵌入

⁴CapsGLOM中所有位置的多级嵌入没有被压缩，因为它们的模具并不表示实体特征出现的概率。

有两个网络可以用以下公式来表示

$$(f_{scene}, f_{object}, f_{part}, f_{sub-part}, f_{lowest})^T \leftarrow \text{bottom-up}(f_{object}, f_{part}, f_{sub-part}, f_{lowest}, f_{under}) \quad (21)$$

$$(f_{object}, f_{part}, f_{sub-part}, f_{lowest})^T \leftarrow \text{top-down}(f_{scene}, f_{object}, f_{part}, f_{sub-part}) \quad (22)$$

让 $E_t = (f_{地点}; f_{目标}; f_{部分}; f_{子部分}; f_{最低的})^T$ 预示在 t 时刻的所有位置的多级嵌入，然后在下一轮的 $t+1$ 时刻，它可以通过以下公式进行更新

$$+ \boldsymbol{E}_t + \text{Interact}(\boldsymbol{E}_t)] / (3, 4, 4, 4, 4)^T \quad (23)$$

其中Interact表示不同列中同一级别嵌入向量之间的交互，请参考等式16.

表二

顶棚的参数配置及每层的输入、输出形状

层（类型）	核数	内核形状	步幅	填充	输入形状	输出形状
组2d1	96	3 3×	1	1	[- 1, 176, 9, 9]	[- 1, 96, 9, 9]
ConvGRU-2	12	3 3×	1	1	[- 1, 12, 9, 9]	[- 1, 12, 9, 9]
ConvGRU-3	6	3 3×	1	1	[- 1, 12, 9, 9]	[- 1, 6, 9, 9]
GN-4	-	-	-	-	[- 1, 48, 9, 9]	[- 1, 48, 9, 9]
线性5	-	-	-	-	[- 1, 81, 48]	[- 1, 81, 32]
Conv2d-6	160	1 1×	1	0	[- 1, 48, 9, 9]	[- 1, 160, 9, 9]
自下而上的7	-	-	-	-	[- 1, 81, 5, 32]	[- 1, 81, 5, 32]
自上而下8	-	-	-	-	[- 1, 81, 4, 32]	[- 1, 81, 4, 32]
Conv1d-9	1	81	1	0	[- 1, 81, 32]	[- 1, 1, 32]
FC10	-	-	-	-	[- 1, 32]	[- 1, 13]

由于我们正在执行一个分类任务，并且每个位置（像素）可能代表一个背景像素或其他类型的地面对象，所以我们使用每个位置的场景级嵌入来执行分类，因为这些嵌入描述了包含整个场景的样本信息。它们被输入到一维卷积层，以学习最终分类的集成特征。最后，用交叉熵作为损失函数来训练整个模型。为了便于理解，表二中列出了模型的参数配置和每一层的输入和输出形状（Conv2d-6代表像素级卷积）。

C. CAL方案

目前，大多数基于不确定性的HIC深度AL方法只考虑所选样本的不确定性，而没有考虑专家注释的多样性和预算。在本小节中，我们提出了一个协作主动学习（CAL）方案，该方案同时考虑了主动选择的样本的不确定性和多样性，以及专家注释的成本。

我们提出的CAL方案属于一种基于不确定性的AL方法。BvSB和EP是两种常用的不确定性度量方法。对于任何未标记的样品 x_i ，我们可以使用一个训练过的模型来估计它的类属概率。让你 $i = (y_{i1}, y_{i2}, \dots, y_{iC})^T$ R^C 表示估计的后验概率，然后为未标记样本 x 的基于bvsvb的不确定性 i 可以表示为

$$BvSB(x_i) = PB(y_i) \quad P\text{-某人}(y_i) \quad (24)$$

其中 $PB(y_i)$ 表示 y 中的最高（概率）值 i 和 $P\text{某人}(y_i)$ 表示第二高的值。对于该准则，样本的BvSB值越小，分类器对样本类别的判断就越模糊。熵通常用来描述混杂度，因此也可以作为AL中的不确定度测度，可以用以下公式计算

$$H(x_i) = - \sum_{c=1}^C P(y_{ic}) \log P(y_{ic}) \quad (25)$$

其中，对数表示自然对数。对于任何未标记的样本，估计的后验概率的熵越高，分类器就越难判断其类别。对于上述两个不确定度指标，BvSB为

通常比预测的EP [2]更好，因此我们在我们提出的CAL方案中采用了基于bvsvb的查询策略。

在所提出的CAL方案中，使用CapsViT进行计算利用候选样本的不确定性和CapsGLOM协助专家注释，确保积极选择的样本的多样性。CAL的主要优点是，它可以充分利用从两个基分类器中学习到的知识，并确保主动选择的样本的多样性，而无需设计任何多样性标准。最重要的是，与传统的基于批处理的AL相比，CAL充分考虑了专家标签的成本

。让 X^L 表示已标记的样本集和 X^C 表示候选样本集。需要补充的是，对于样本数量较少的数据集，整个测试集都可以作为候选集。如果样本数量太大，为了减少计算时间和内存需求，通常会选择适当数量的未标记样本来构建候选集（池）。回到这一点，在每次AL迭代中，我们首先用 X 训练CapsViT和CapsGLOM^L，然后发送 X^C 到CapsViT，计算每个候选样本的BvSB值，并送往CapsGLOM进行分类。假设有 M 个候选样本，即 $X^C = \{x_1^C, x_2^C, \dots, x_M^C\}$ ，我们可以得到排序BvSB值 $\{BvSB(x_1^C), BvSB(x_2^C), \dots, BvSB(x_M^C)\}$ 按升序排列

$$\text{索引} = \text{目标排序}(BvSB(X^C)) \quad (26)$$

接下来，我们使用键0、1... -1 ， $C1$ ，每个键对应于 a 收集器（空列表）。让我们将这些指数表示为 $\{\text{ind } 0, \text{ind } 1, \dots, -1\}$ ，然后是候选人 sam 与索引对应的数组可以表示为 $\{X^C[\text{ind } 0], X^C[\text{ind } 1], \dots, X^C[\text{ind } (M-1)]\}$ 。以上内容重新排序的候选样本依次添加到收集器中以CapsGLOM估计的类别为关键。这样，每个收集器中样本的BvSB值保持在从最小到最大的顺序。最后，我们只选择每个收集器中的第一个样本来进行专家注释。如果专家难以在一个特定的收集器中标记第一个样本，他可以尝试标记第二个样本，以此类推。因此，所提出的CAL方案不仅考虑了主动选择的样本的不确定性和多样性，而且还考虑了专家标签的成本。作为结论，

算法1: 协作主动学习 (CAL) 框架

输入: 带标记的样本集: X^L ; 候选样本集: X^C ; AL迭代的次数: I ; 类别的数量: C 。
输出: 经过训练的CapsViT和CapsGLOM。

```

1 初始化  $i = 0$ ;
2. 当我做<时, 我就这样做了
3   列车CapsViT和与 $X^L$ ; 计算 $X$ 的BvSB值 $C$ 通过
4   CapsViT:
       候选人_bvsvb=BvSB( $X^C$ );
5    $X$ 分类 $C$ 通过CapsGLOM:
       pred_capsglom=CapsGLOM( $X^C$ );
6   获取指标: 指数=目标排序(候选人_bvsvb); dict =
7   {0: [ ], 1: [ ], ..., C1: [ ]};
8   整理当前信息最丰富的样本
       迭代:  $X^L=[ ]$ ;
9    $M = \text{len}(X^C)$ ;
10  对于  $j = 0$  到  $(M - 1)$  do—
11     ind=indices[
12     pred_label=pred_capsglom[ind];
13     dict[pred_标签附加 $X^C[ind]$ ];
14  最后部分
15  对于  $c = 0$  到  $(C - 1)$  do—
16     仅选择每个收集器中的第一个样本:
17      $x_c=\text{dict}[c][0]$ ;
18     让专家注释 $x_c$ ;
19     虽然这里有一个困难
20     请专家注释下一个示例
21     在当前收集器中:  $\text{dict}[c]$ ;
22     最后部分
23      $X^L.\text{append}(x_c)$ ;
24  最后部分
25  更新已标记的样本集:  $X^L$ 
26  更新候选样本集:  $X^C$ 
27   $i = i + 1$ ;
28  列车CapsViT和CapsGLOM与 $X^L$ 到最后
29  时间

```

我们在算法1中提出了我们提出的CAL框架的算法流程。

增值实验结果及分析

为了分析我们提出的CAL框架的有效性和优越性,我们在四个基准的HSI数据集上进行了广泛的实验: 肯尼迪航天中心(KSC)数据集、休斯顿大学2013(HU2013)数据集、帕维亚大学(PU)数据集和赤土数据集。

我们在每个数据集上进行了两个实验: 一个训练样本的AL实验和一个有足够样本的监督分类实验。在AL实验中,我们比较了CapsViT、CapsGLOM和稍加修改的A2S2K-RN [10]。在监督分类实验中,我们将所提出的CapsViT和CapsGLOM与一些最近提出的深度模型: 混合体[4], CAEVT

[6], A2S2K-RN [10], CapsNet [11], 光谱网[13]和光谱前[20]。所有的实验都使用了三个指标来评估每个模型的性能: 总体精度(OA)、平均精度(AA)和Kappa系数(κ)。最后,所有的实验都是在一个使用NVIDIA GeForce RTX 3090 GPU加速器(具有24-GB GPU内存)的系统上进行的。

用于构建CapsViT和CapsGLOM的一些超参数分别见表I和表2。

在开始介绍实验部分之前,我们在表三介绍了这两个模型的其他一些超参数设置,包括初始学习率(η_0)、学习速率调度、优化器、权值衰减系数和训练历元。这里的训练时期是指默认值,在实践中应该根据数据集的特征和初始训练样本的数量进行调整。

A. 实验1: 肯尼迪航天中心的数据集

KSC数据集是1996年由肯尼迪号的AVIRIS(机载可见/红外成像光谱仪)捕获的

佛罗里达的航天中心。AVIRIS采集了224个10 nm宽的波段,中心波长为400–2500 nm。KSC数据集的空间分辨率为18 m,在去除吸水率和低信噪比(信噪比)波段后,使用176个波段进行分析。该数据集的空间维度为512 614个像素,共包含13个类别中的5211个样本。图8a为该数据集的三波段(波段29、19、9)假彩色图像,以及图8b显示了用颜色编码的地面真相。

在该数据集上的AL实验中,每个类别中只随机抽取1个样本作为初始训练样本,即初始训练集由13个训练样本组成。图8中为所选择的训练样本,每个样本的空间大小为9 9。我们将讨论不同空间大小的训练样本对基于该数据集的分类结果的影响。但是,考虑到较大的空间尺寸通常意味着更大的计算量,在其他数据集上的实验都采用9 9的空间大小。我们将AL迭代次数设置为4,这意味着总共选择了52个额外的样本进行专家标记。此外,我们在比较A2S2K-RN时做了一个小的修改。原始的A2S2K-RN分别使用1 1 和7 7 的卷积核来提取光谱特征和空间特征,我们将这两个卷积核替换为1 1、3 3 和5 5 的三个卷积核。为了防止批归一化遇到一个小批只包含一个训练样本的情况,将所有比较模型用于训练的批大小设置为30。最后,有一件事是要明确的,为了重现我们的实验,我们使用一个随机种子(种子=0)来选择初始训练样本,为了消除参数初始化对实验结果的影响,我们也使用这个随机种子

表三

一些其他的超参数设置

模型	初始学习率 (η_0)	学习率调度	优化器	权重衰减	训练时代
CapsViT	5e-4	$\eta_0 \times 0.99$ 时代	亚当	1e-3	200
CapsGLOM	1e-3	$\eta_0 \times 0.99$ 时代	亚当	0.00	200

表iv

使用13个初始训练样本（学习样本的空间大小为9 9）在KSC数据集上进行主动学习实验。✕

课		主动学习迭代=0			主动学习迭代=2			主动学习迭代=4		
		A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM
准确度 (%)	C1	96.18	98.82	99.61	97.89	93.55	97.63	95.13	96.71	94.87
	C2	90.50	90.91	26.45	54.96	71.49	39.26	99.59	100.00	95.04
	C3	79.61	69.41	65.10	99.61	96.08	94.90	95.69	88.24	88.63
	C4	14.34	7.57	18.73	16.73	17.53	29.48	48.61	33.47	69.72
	C5	71.25	20.00	57.50	99.38	76.88	77.50	100.00	87.50	56.25
	C6	64.91	90.79	93.42	82.89	83.33	96.05	93.86	98.25	96.93
	C7	60.58	60.58	92.31	48.08	47.12	67.31	99.04	64.42	64.42
	C8	42.56	28.37	41.63	99.30	100.00	100.00	92.56	91.63	97.91
	C9	20.42	22.93	15.99	95.57	100.00	100.00	84.01	100.00	100.00
	C10	28.29	10.67	50.87	80.15	98.76	89.33	97.52	98.76	96.53
	C11	97.85	93.54	95.45	100.00	100.00	100.00	98.09	100.00	99.28
	C12	83.47	90.04	84.86	96.61	95.02	95.42	100.00	95.62	100.00
	C13	100.00	100.00	100.00	84.56	99.14	100.00	100.00	100.00	100.00
OA (%)		70.62	67.76	70.28	86.63	90.32	90.38	93.73	93.34	94.31
AA (%)		65.38	60.28	64.76	81.21	82.99	83.61	92.62	88.81	89.20
κ		0.6725	0.6400	0.6691	0.8513	0.8922	0.8928	0.9302	0.9259	0.9366

控制每个模型的初始化⁵。

除了初始训练集中的13个训练样本外，剩下的5198个样本作为测试样本和候选样本用于主动选择。需要注意的是，由于KSC数据集中的样本总数并不大，所以我们没有单独划分候选样本集。在表四中，我们报告了A2S2K-RN、CapsViT和CapsGLOM对初始训练集和经过2次和4次AL迭代后获得的训练集的分类结果。在这个实验中，A2S2K-RN也被训练了200个时代，我们假设专家能够在每个收集器中注释第一个未标记的样本。

如表四所示，当每个类别只有1个训练样本时，每个模型在测试集上的性能都不令人满意。除此之外，每个模型识别每个类别的能力也有很大的差异。但是，可以发现，对于第4类（Pine），这三种模型都难以正确识别，CapsGLOM的精度为18.73%。CapsGLOM也有其性能不佳的时刻，比如在分类类2（Willow）时，准确率仅为26.54%。对于CapsViT，其对第4类的识别精度最低，这也导致它在OA、AA和h方面落后于其他两种模型。虽然这三个模型可能会过拟合训练集，但A2S2K-RN在所有三个指标上都优于其他两个分类器。因此，我们欣赏其他优秀的分类器，并没有强调我们提出的CAL框架中的两个基本分类器在所有情况下都是最好的，事实上，

⁵所有比较的模型都由PyTorch DL框架实现，我们设置了以下命令：torch.用手的_种子(0)，火炬.cuda.用手的_种子(0)，火炬.backends.cudnn.deterministic=True和火炬.后端.cudnn.benchmark=False。

它们可以灵活地被其他优秀的分类器取代，只要这些分类器可以处理具有相同形式（维度）的样本。这里要注意的一件事是，我们比较的三个模型是轻量级模型，所以当替换任何基本分类器在汽车框架与一个轻量级的模型，我们建议设置早期停止基于训练精度或训练损失，以避免过拟合。

经过2次AL迭代后，CapsViT和CapsGLOM的改进效果均优于A2S2K-RN。此外，CapsGLOM在分类第4类方面继续保持着显著的优势，并在三个指标中达到了最大值。经过4次AL迭代后，所有的比较模型的性能都得到了显著提高。CapsGLOM对第4类样本的分类精度仍处于领先地位。此外，它在OA和h方面优于其他两种模型，这也验证了我们提出的CapsGLOM模型的优越性。CapsViT和A2S2K-RN在OA和h方面的差异不大，而A2S2K-RN在所有比较模型中AA最高，这也说明它在不同类别中具有更平衡的分类性能。最后，我们在图-8f中给出了每个模型经过4次AL迭代后得到的分类图。8

不同的训练样本空间大小会导致不同的分类结果。同时，不同的空间大小也意味着不同的计算量。为了比较该CAL框架在不同样本量下的性能，我们使用7 7 个训练样本在KSC数据集上重新进行了实验。✕✕需要注意的是，当使用77个训练样本时，CapsViT中的PrimaryCaps层将得到400（1655）个主胶囊，而不是784个（见表I）。✕✕在表V中，我们给出了比较结果

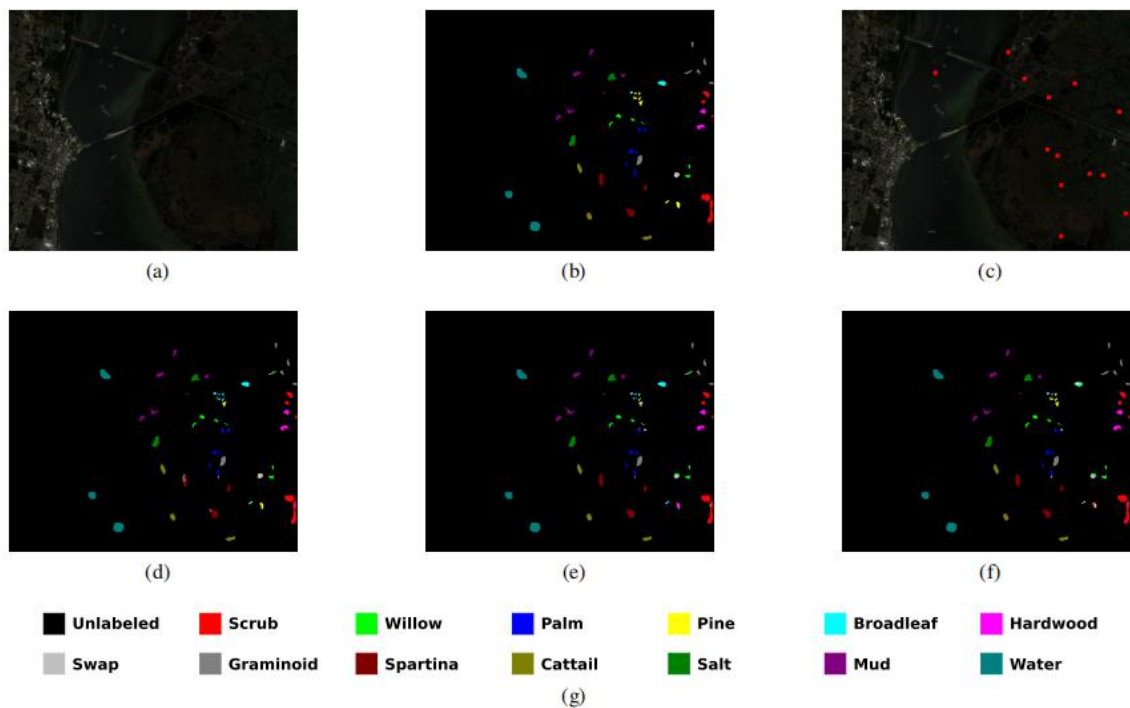


图8. 由不同模型在KSC数据集上获得的分类图的可视化。(a) 三波段假色复合材料。(b) 地面真相。(c) 培训样本。(d)–(f) 分别由A2S2K-RN、CapsViT和CapsGLOM经过4次AL迭代后得到的分类图。(g) 图例。

表v
在KSC数据集上进行主动学习实验（训练样本的空间大小为7 7）。×

课		主动学习迭代=0			主动学习迭代=2			主动学习迭代=4		
		A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM
准确度 (%)	C1	93.29	98.82	95.39	96.97	96.97	98.29	97.76	98.16	98.82
	C2	80.99	62.81	32.23	93.80	98.76	98.76	98.76	97.52	98.35
	C3	87.45	80.00	89.41	99.61	94.51	99.22	91.37	94.12	96.86
	C4	14.74	7.97	13.15	27.89	12.75	17.13	29.08	29.88	41.04
	C5	51.88	55.00	60.63	90.00	88.75	80.62	93.13	81.88	86.88
	C6	67.11	55.70	93.86	98.25	77.63	90.79	94.30	62.72	79.39
	C7	89.42	60.58	74.04	35.58	25.00	48.08	84.62	42.31	63.46
	C8	1.86	14.19	41.63	91.63	66.98	82.56	93.72	96.05	99.53
	C9	15.41	45.66	18.30	44.51	100.00	98.65	91.71	75.34	100.00
	C10	35.48	12.66	85.86	92.56	76.43	99.01	97.77	100.00	100.00
	C11	91.63	91.87	97.85	100.00	98.80	100.00	99.52	100.00	100.00
	C12	65.34	90.24	87.65	96.61	87.25	94.02	94.82	99.40	99.40
	C13	88.77	99.57	100.00	100.00	100.00	100.00	100.00	100.00	100.00
OA (%)		62.68	67.60	74.01	86.96	86.30	91.38	92.94	89.75	94.61
AA (%)		60.26	59.63	68.46	82.11	78.76	85.16	89.74	82.87	89.52
Kappa		0.5878	0.6379	0.7115	0.8550	0.8473	0.9040	0.9213	0.8856	0.9400

类似于表四。可以看出，当最初的13个开始时，×使用了7 7 个训练样本，CapsViT的整体性能变化不大，CapsGLOM的性能有一定程度的提高，而A2S2K-RN的性能下降了很大幅度。对于第8类（谷氨类），A2S2KRN的分类准确率（召回率）仅为1.86%，难以接受。经过4次AL迭代后，CapsGLOM在OA和h中仍然保持领先，而A2S2K-RN的AA值仍然最高。在本实验中，虽然A2S2K-RN最终仍优于CapsViT，但CapsViT不仅在初始训练集上优于A2S2K-RN，而且对不同空间规模的训练也表现出一定的鲁棒性。

样品

为了提高CapsViT的有效性和优越性，我们基于上述13 7 7个训练样本进行了额外的实验。×当我们从CapsViT中删除ViT模块时，它得到的OA、AA和h值分别为58.23%、45.34%和0.5309。这种急剧的性能衰减足以证明将ViT引入CapsNet的成功和重要性。造成这种明显差异的原因是，在引入ViT后，对于每个主胶囊，CapsViT可以学习其全局最优表示，这对分类很有意义。为了验证我们的声明，我们将一个回调函数传递给CapsViT。

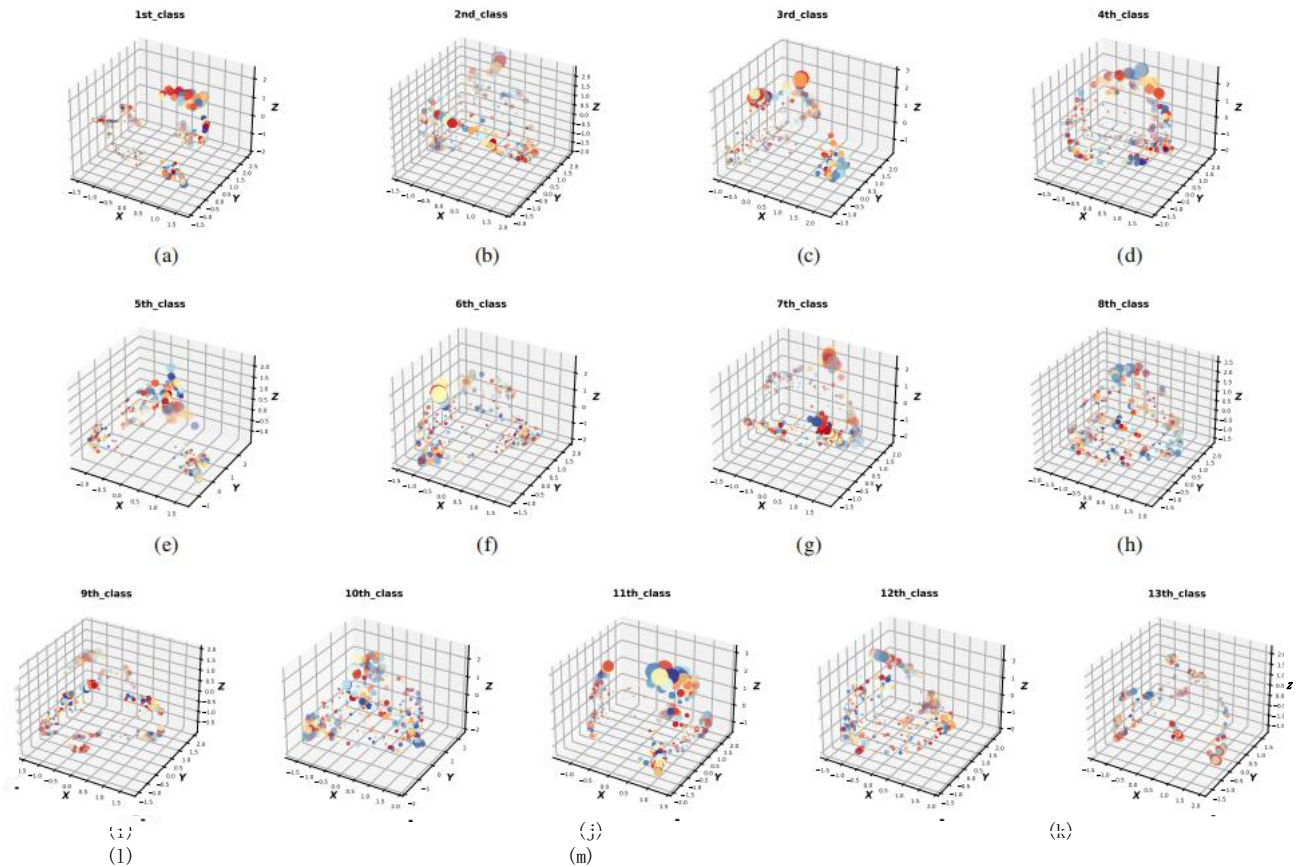


图9. 对KSC数据集的CapsViT中的主帽层输出的可视化。每个点代表一个胶囊特征，点的大小表示实体特征出现的概率。(a)–13 7 7 个训练样本的 (m) 特征。✕

观察主帽层和ViT模块的输出特性。对于每个训练样本，primary帽层和ViT模块的输出都是400个16维的胶囊向量。为了可视化每个训练样本的胶囊特征，我们首先需要降低这些特征的维数。在这里，对于每个训练样本，我们使用主成分分析（PCA）将400个16维的胶囊向量投影到一个在保持最大差异的方向上的三维空间中。虽然投影的胶囊向量不再具有被挤压的特性，但我们仍然可以使用它们的模来表示实体特征出现的概率。

13 7 7个训练样本的PrimaryCaps层和ViT模块的输出如图9和图中所示。✕分别为10。通过比较无花果。9和无花果。10，我们可以发现主要的输出层相对离散，只关注位置和概率发生的各种特性，而生活模块的输出更均匀，和一些特性将变得更加明显后考虑所有特性之间的全局相关性。

接下来，我们研究了不同的查询策略对我们提出的CAL框架的影响，并使用了基于EP和损失预测的两种查询策略进行了比较。在比较基于损失预测的查询策略时，我们使用CapsViT中的多层输出来构建损失预测模块。所使用的图层包括第一个图层

卷积层、GN层、主帽层和ViT模块。其中，primap层和ViT模块的输出需要首先重塑为特征立方体，然后发送到损失预测模块。需要注意的是，在我们的实验中，损失预测模块的梯度并没有反向传播到CapsViT。另外，在计算损失预测模块的损失时，要求每个小批的样本数必须是偶数，这通常不能保证在每个历元的最后一个小批。所以我们做了一个小的修改，当小批量样本的数量是一个奇数时，我们将第一个样本和中间的样本配对来计算一个损失值。当一切都准备好时，我们根据表6中不同的查询策略给出实验结果，训练样本的空间大小为9 9 （稍后不再声明）。✕

通过查看表VI，我们可以发现，虽然基于损失预测的启发式算法得到的结果是可以接受的，但其性能不如基于BvSB和EP算法得到的结果好。此外，在使用基于ep的查询策略时，CapsViT和CapsGLOM都获得了更好的分类结果，而A2S2K-RN则没有。同时，在使用基于ep的启发式算法时，CapsGLOM获得了最佳的AA，为90.41%，但该值仍远低于92.62%的AA值通过使用基于bvsvb的A2S2K-RN查询策略实现。因此，总的来说，虽然CAL框架在采用KSC数据集上获得了更好的分类结果

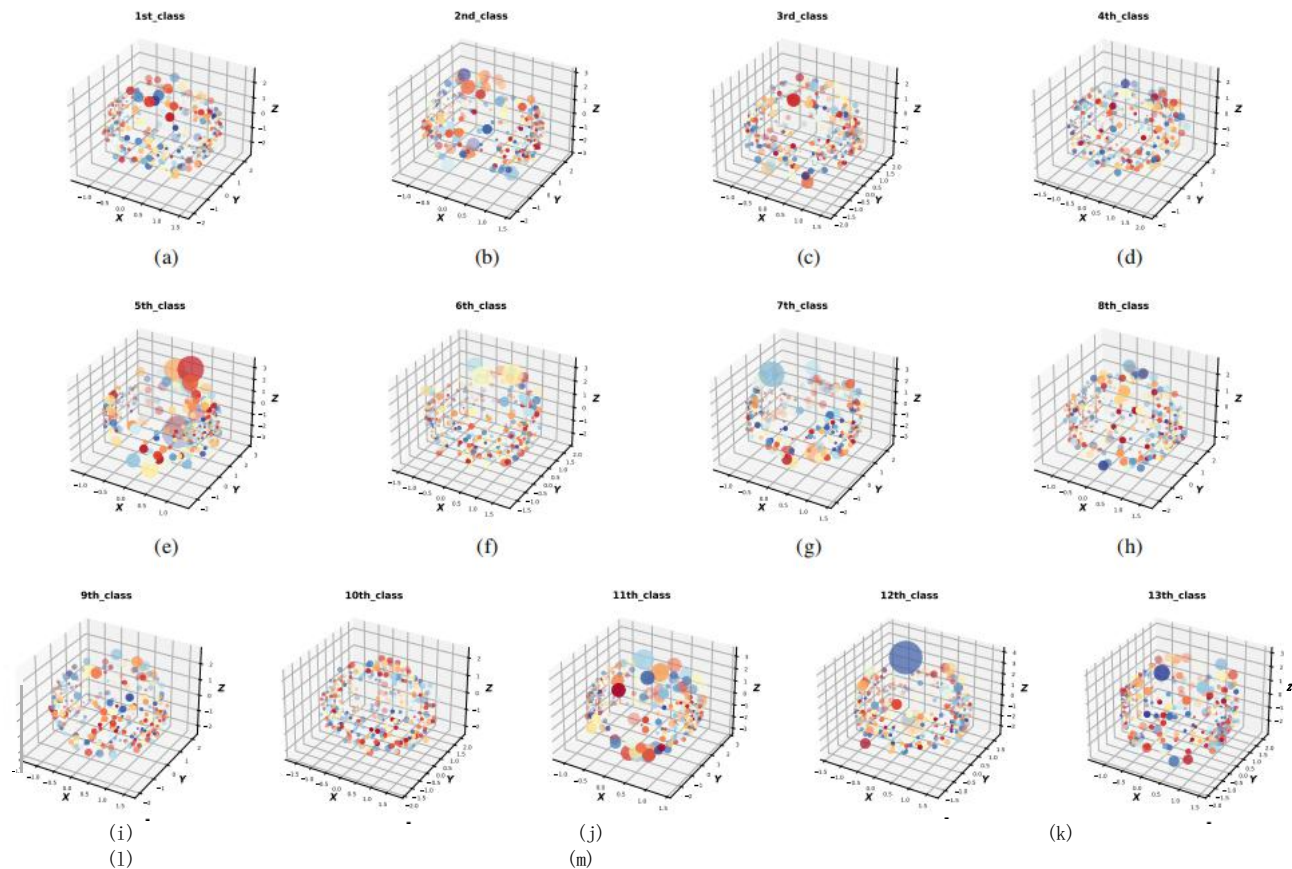


图10. 在CapsViT中可视化针对KSC数据集的ViT模块输出。每个点代表一个胶囊特征，点的大小表示实体特征出现的概率。(a)–13 7 7 个训练样本的 (m) 特征。

表VI

使用不同的查询策略对KSC数据集进行主动学习实验(训练样本的空间大小为

$\times 9 \quad 9$) .

课	基于BvSB			基于EP的			基于损失预测		
	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM
OA (%)	93.73	93.34	94.31	91.57	94.52	95.09	89.44	89.28	91.11
AA (%)	92.62	88.81	89.20	88.72	90.27	90.41	86.18	83.00	87.01
AC	0.9302	0.9259	0.9366	0.9061	0.9390	0.9454	0.8823	0.8805	0.9010

基于ep的查询策略，考虑到其中两个基本分类器的可替代性，基于bvsb的查询策略仍然是CAL框架的默认选择。

最后，为了进一步验证CapsViT和CapsGLOM的优越性，我们随机抽取每个类别中10%的标记样本，作为训练样本进行监督分类实验。需要比较的深度模型包括混合SN[4]、CAEVT[6]、A2S2KRN [10]、光谱网[13]、光谱前体[20]、SSFTT [23]和DMCapsNet [24]。在这一系列的实验中，我们没有设置随机种子来控制训练样本的选择和每个要比较的模型的初始化，所以每个模型重复实验10次，以报告平均精度和标准差。此外，还有几件事要说：

- 1) 在比较CAEVT时，我们没有使用三维卷积自动编码器对原始数据集进行预处理以进行降维，而是使用了

由我们设计的ConvGRU模块来学习所有波段的光谱特征。

- 2) 除了SSFTT（它被训练了100个时代），DMCapsNet（它在KSC和HU2013数据集上训练了100个时代，在PU和日本人数数据集上训练了50个时代）和频谱网（它被训练在KSC、PU和千圭数据集上的250个时代在HU2013数据集上的200个时代），所有的比较模型都训练了200个时代，批大小设置为32。

- 3) 我们没有将验证集分开划分，以获得每个模型的最优参数，因为所有模型在此之后都不会严重过拟合如果有足够的训练样本，就会有200个训练时代。

- 4) 公平，除了混合SN，DMCapsNet和谱网（分别使用2525, 2525个25和2424个训练样本），训练sam的空间大小

表VII

使用10%的训练样本对KSC数据集进行监督分类实验。

课		HybridSN	卡维特	A2S2K-RN	DMCapsNet	斯弗特	光谱网	光谱元件	CapsViT	CapsGLOM
准确度 (%)	C1	95. \pm 252. 05	99. \pm 440. 49	99. \pm 780. 3	97. \pm 721. 20	99. \pm 870. 12	99. \pm 710. 31	99. \pm 650. 42	99. \pm 850. 24	99. \pm 650. 31
	C2	82. 484. 90	94. \pm 635. 35	2	92. 065. 55	92. 808. 02	96. 653. 08	88. 589. 16 \pm	98. 811. 54	98. 211. 75
	C3	91. 703. 82	91. \pm 879. 15	99. 131. 97	99. 521. 16	95. 265. 61	96. 612. 54	88. 915. 18 \pm	99. 261. 40	98. 521. 47
	C4	75. 406. 14	74. \pm 3410. 40	98. 782. 18	85. 274. 53	82. 307. 26	90. 715. 73	\pm 63. 1913. 98	88. 505. 87	88. 504. 52
	C5	95. 972. 93	81. 465. 33 \pm	88. 639. 80	76. 186. 56	87. 224. 82	97. 154. 60	\pm 69. 1012. 46	94. 034. 96	92. 575. 99
	C6	84. 564. 96	90. \pm 975. 36	96. 465. 76	88. 885. 79	86. 845. 51	98. 791. 41	91. \pm 503. 58	97. 524. 22	96. 124. 99
	C7	87. 237. 96	97. \pm 556. 66	98. 791. 79	99. 361. 91	98. 512. 39	100. 000. 00	98. \pm 093. 49	99. 041. 93	99. 470. 86
	C8	94. 012. 30	99. \pm 590. 79	99. 890. 32	99. 201. 68	99. 920. 17	98. 791. 62	95. \pm 743. 03	100. 000. 00	100. 000. 00
	C9	92. 264. 26	99. \pm 940. 14	99. 950. 10	99. 441. 53	99. 910. 10	99. 870. 38	98. \pm 441. 81	100. 000. 00	100. 000. 00
	C10	92. 234. 67	99. \pm 810. 35	99. 940. 10	91. 874. 81	98. 601. 44	99. 940. 17	98. \pm 042. 44	100. 000. 00	99. 970. 08
	C11	97. 532. 80	99. \pm 760. 51	99. 780. 66	99. 280. 65	99. 660. 67	100. 000. 00	99. \pm 710. 34	100. 000. 00	99. 600. 65
	C12	94. 652. 87	98. \pm 981. 18	100. 000. 00	97. 810. 95	97. 461. 42	99. 600. 85	95. \pm 755. 35	99. 361. 28	99. 401. 04
	C13	100. 000. 00	100. 000. 00 \pm	99. 620. 92	100. 000. 00	100. 000. 00	100. 000. 00	100. 000. 00 \pm	100. 000. 00	100. 000. 00
		\pm 0.000000		100. 000. 00	\pm 0.000000	\pm 0.000000	\pm 0.000000	\pm 0.000000	\pm 0.000000	\pm 0.000000
	\pm 0.000000		\pm 0.000000	\pm 0.000000	\pm 0.000000	\pm 0.000000	\pm 0.000000	\pm 0.000000	\pm 0.000000	
OA (%)		93. \pm 251. 59	96. \pm 841. 23	99. \pm 090. 50	96. \pm 310. 60	97. \pm 160. 42	98. \pm 890. 22	94. \pm 642. 20	98. \pm 960. 42	98. \pm 740. 38
AA (%)		91. \pm 021. 70	94. \pm 492. 05	98. \pm 520. 93	94. \pm 350. 81	95. \pm 260. 80	98. \pm 290. 43	91. \pm 283. 30	98. \pm 180. 70	97. \pm 850. 63
AC		\pm 0. 92480. 018	\pm 0. 96480. 014	\pm 0. 98990. 006	\pm 0. 95890. 007	\pm 0. 96840. 005	\pm 0. 98760. 003	\pm 0. 94030. 025	\pm 0. 98840. 005	\pm 0. 98590. 004

所有模型使用的pe为9 9。x

由不同模型得到的分类结果见表7。显然，当使用10%的标记样本作为训练样本时，A2-2SK-RN在KSC数据集上表现最好。HybridSN的性能最差，可能是因为其结构简单。虽然HybridSN和SpeactralNET使用的训练样本的空间尺寸更大（这样的样本将包含更多的空间信息），但SpectralNET的性能明显优于HybridSN。特别是，对于第4类，其他模型的召回率都没有超过90%，而SpectralNET的召回率为90.71%，偏差相对较小。此外，光谱得到的性能并不令人满意，与A2S2K-RN得到的最高OA值99.09%相比，光谱得到得到的OA值仅为94.64%，因此，简单地将训练样本补丁的每个展开频带作为输入序列的一个元素提供给ViT的方法是值得怀疑的。对于CAEVT，尽管它使用的MobileViT块比ViT的计算量要少得多，但它的性能不是很好。至于SSFTT，它是一个轻量级的模型，其性能是中等的。此外，尽管DMCapsNet使用视点不变变换矩阵来提取视点不变特征，并通过计算选票和分布之间的相似性来聚合低级胶囊特征，但与CapsViT相比，其性能一般。我们提出的CapsViT在OA、AA和h中分别比DMCapsNet高2.65%、3.83%和0.0295，验证了将ViT引入CaspNet的有效性和合理性。最后，通过比较所有的分类结果，可以发现CapsViT和CapsGLOM分别获得了第二名和第四名的最佳结果，这足以表明它们的优越性。

B. 实验2: 休斯顿大学2013年的数据集

休斯顿大学2013 (HU2013) 数据集由国家机载激光测绘中心 (NCALM) 在休斯顿大学校园和邻近的城市地区获得, 并在2013年IEEE GRSS (地球科学和遥感学会) 数据融合中发布

比赛。×该图像在380 nm到1050 nm区域内有144个光谱波段，其中包括349 1905像素，空间分辨率为2.5 m。该数据集包含15个复杂的土地覆盖类别，假彩色图像和地面真实情况如图所示。11a和图。11c.

对于该数据集，有15029个样本，我们从每个类别中随机抽取5个样本形成初始训练集，并设置AL迭代次数为5次，即共有75个样本被专家主动标记。此外，虽然该数据集中的样本总数相对较大，但为了不遗漏信息最丰富的样本，我们仍然没有在该数据集上单独划分候选数据集。与KSC数据集上的实验相比，我们在HU2013数据集的实验中做了两个变化：1) 批大小调整为32，因为不会包含一个训练样本的小批；2) 考虑到CapsViT和CapsGLOM可能没有在该数据集上得到充分的训练，我们将这两个模型的训练期调整为250而不是默认的200。表八列出了A2S2K-RN、CapsViT和CapsGLOM对初始训练集的分类结果，以及经过3次和5次AL迭代得到的训练集的分类结果。

从表八可以看出，在初始训练集上，CapsViT获得的OA和h最高，A2S2KRN获得的AA最高。这一结果证明，当只有少数训练样本时，CapsViT可以获得相对理想的分类结果。虽然CapsGLOM的表现平平，但它在五个类别中召回率最高，这有助于它能够帮助专家进行注释。经过3次AL迭代，A2S2K-RN开始占上风，主要是因为它对第11类（Raiel）的识别精度优于CapsViT和CapsGLOM。经过5次AL迭代，我们可以发现A2S2KRN获得了最好的性能，其OA比CapsGLOM高1.81%，比CapsViT高1.69%。在这一点上，很明显，我们有理由将我们提出的CAL框架替换为A2S2K-RN中的CapsGLOM。从视觉上比较上述三个模型的性能

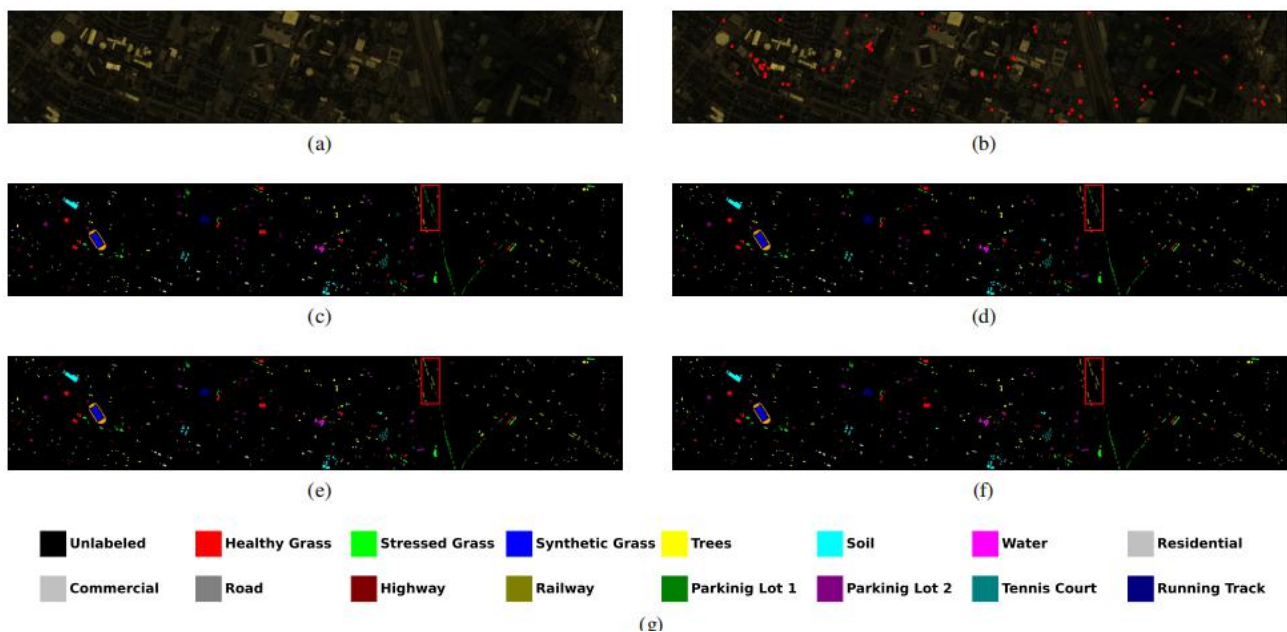


图11. 在HU2013数据集(a)三波段假彩色复合图上，由不同模型获得的分类图的可视化。.. (b)培训样本。(c)地面真相。(d)–(f)分别由A2S2K-RN、CapsViT和CapsGLOM经过5次AL迭代后得到的分类图。(g)图例。

表VIII
使用75个初始训练样本对HU2013数据集进行主动学习实验。

课		主动学习迭代=0			主动学习迭代=3			主动学习迭代=5		
		A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM
准确度 (%)	C1	94.54	93.58	97.75	95.67	85.96	90.69	98.48	94.54	98.64
	C2	79.02	91.67	70.62	90.56	95.20	98.32	96.08	99.68	99.92
	C3	97.98	96.68	95.09	100.00	99.57	100.00	100.00	100.00	100.00
	C4	94.67	94.92	92.01	99.11	99.19	99.35	97.82	98.55	92.33
	C5	91.67	100.00	99.84	100.00	99.92	99.27	99.19	99.84	98.71
	C6	93.75	88.75	85.00	98.13	91.56	88.44	97.81	94.69	95.63
	C7	63.90	39.67	46.95	95.49	87.17	89.31	94.38	90.26	92.72
	C8	46.81	45.76	49.80	66.67	71.19	68.60	78.93	84.75	90.40
	C9	72.89	76.18	79.31	83.24	79.15	74.66	80.51	78.99	73.62
	C10	49.10	49.35	42.23	76.68	67.51	79.62	99.35	88.54	90.75
	C11	50.08	48.13	63.41	80.57	56.67	66.26	86.34	66.67	64.72
	C12	56.03	68.97	52.20	84.45	93.24	88.19	86.16	96.17	95.28
	C13	94.61	92.03	94.40	95.04	94.83	92.24	93.75	92.67	92.89
	C14	100.00	94.09	99.53	100.00	99.76	100.00	100.00	100.00	99.53
	C15	97.56	100.00	100.00	100.00	100.00	99.69	100.00	100.00	98.63
OA (%)		74.57	75.01	73.93	89.26	86.00	87.44	92.90	91.21	91.09
AA (%)		78.84	78.65	77.88	91.04	88.06	88.98	93.92	92.36	92.25
Kappa		0.7255	0.7304	0.7184	0.8839	0.8487	0.8643	0.9232	0.9050	0.9037

因此，所得到的分类图如图所示。11d11f. 在无花果。11，从红色矩形突出显示的地方我们可以看到，A2S2K-RN在分类类11方面明显优于CapsViT和CapsGLOM。表八中的实验结果确实相同，对于第11类，三种模型的召回率分别为86.34%、66%、67%和64.72%。

接下来，我们在该数据集上进行监督分类实验，在每个类别中选择10%的样本作为训练样本（共选择1510个样本）。由不同模型得到的实验结果见表九。如表九所示，当训练样本足够时，SSFTT的性能在

HU2013数据集足够亮，其次是A2S2K-RN。与SSFTT获得的最高OA值98.99%相比，我们提出的CapsViT和CapsGLOM也具有竞争力，OA值分别为98.05%和98.。分别为16%。当然，轻量级和高性能的SSFTT可以指导我们在未来的工作中改进CapsViT和CapsGLOM。DMCapsNet在HU2013数据集上的性能与CapsViT非常接近，与DMCapsNet相比，由于引入了ViT模块，CapsViT具有更强的学习和表达能力，因此OA和h值比前者高0.02%和0.。0002然而，DMCapsNet实现了0. AA比CapsViT高13%，值得肯定。同时，通过CapsGLOM和

表ix

使用10%的训练样本对HU2013数据集进行监督分类实验。

课		HybridSN	卡维特	A2S2K-RN	DMCapsNet	斯弗特	光谱网	光谱元件	CapsViT	CapsGLOM
准确度 (%)	C1	99. ±320. 6	97. ±321. 5	99. ±650. 54	96. ±912. 63	99. ±371. 08	97. ±361. 6	97. ±091. 7	99. ±090. 82	98. ±781. 1
	C2	5	0	99. 231. 14	99. 170. 79	99. 490. 57	0	8	99. 570. 66	6
	C3	99. 800. 16	99. ±500. 4	99. 870. 10	99. 520. 30	99. 940. 11	97. ±931. 8	96. ±394. 8	99. 710. 39	99. ±120. 7
	C4	99. 590. 53	2	99. 740. 36	96. 413. 06	99. 630. 23	0	4	96. 092. 51	9
	C5	99. 320. 81	99. ±890. 1	99. 990. 03	99. 730. 42	100. 000. 00	99. ±890. 1	99. ±570. 7	99. 980. 04	99. ±780. 1
	C6	100. 000. 00	9	98. 252. 06	95. 553. 55	98. 492. 47	6	4	90. 924. 49	3
	C7	97. 163. 02	97. ±431. 1	98. 591. 16	97. 441. 65	98. 241. 28	97. ±811. 0	96. ±931. 3	98. 820. 51	97. ±552. 3
	C8	97. 471. 13	8	95. 692. 22	96. 871. 75	97. 321. 66	9	0	94. 982. 10	1
	C9	97. 471. 65	99. ±700. 2	97. 161. 68	94. 962. 51	97. 301. 20	99. ±830. 2	99. ±061. 0	96. 251. 41	99. ±960. 0
	C10	96. 041. 42	6	99. 950. 08	100. 000. 00	99. 680. 41	2	1	99. 410. 56	4
	C11	99. 930. 14	91. ±373. 0	99. 060. 97	99. 890. 10	99. 141. 03	90. ±893. 8	81. ±784. 7	97. 831. 35	95. ±074. 0
	C12	99. 531. 07	8	99. 140. 56	98. 791. 66	97. 130. 81	4	2	98. 400. 64	8
	C13	98. 681. 23	95. ±601. 6	98. 081. 72	94. 741. 73	97. 462. 14	95. ±091. 8	91. ±871. 8	96. 191. 49	97. ±511. 9
	C14	97. 162. 04	4	100. 000. 00	100. 000. 00	99. 710. 25	9	0	100. 000. 00	4
	C15	100. 000. 00	93. ±432. 0	100. 000. 00	99. 090. 36±	99. 950. 11±	93. ±852. 3	90. ±732. 5	99. 980. 05±	96. ±441. 2
		99. 980. 05±	7	±±±±±±±	±±±±±±±	±±±±±±±	7	6	±±±±±±±	8
		±±±±±±±	91. ±782. 0	±±	±	±	97. ±330. 8	91. ±152. 7	±	95. ±881. 5
		±	8				7	9		2
			97. ±461. 1				99. ±540. 5	97. ±371. 3		98. ±990. 9
			1				7	6		5
			96. ±320. 5				98. ±681. 0	94. ±591. 2		98. ±530. 9
			1				9	8		9
			95. ±891. 8				96. ±791. 9	96. ±651. 5		97. ±490. 9
			3				4	1		8
			94. ±382. 6				95. ±431. 2	66. ±285. 7		97. ±251. 6
			9				8	4		1
			99. ±321. 0				99. ±710. 4	96. ±682. 2		99. ±900. 3
			3				9	1		1
			99. ±950. 0				99. ±850. 3	98. ±591. 9		99. ±980. 0
			8				0	2		5
OA (%)		98. ±800. 25	96. ±660. 28	98. ±920. 25	98. ±030. 41	98. ±990. 24	97. ±500. 33	94. ±380. 55	98. ±050. 21	98. ±160. 35
AA (%)		98. ±760. 33	96. ±620. 38	98. ±960. 27	97. ±940. 50	98. ±990. 29	97. ±330. 42	92. ±980. 62	97. ±810. 34	98. ±150. 49
σ		±0. 98700. 003	±0. 96390. 003	±0. 98830. 003	±0. 97870. 004	±0. 98900. 003	±0. 97290. 004	±0. 93920. 006	±0. 97890. 002	±0. 98010. 004

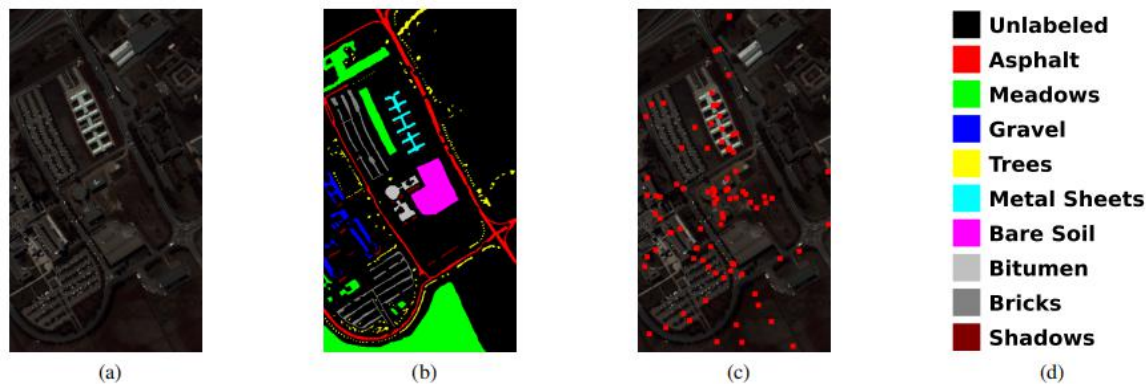


图12. 帕维亚大学的数据集。(a) 三波段假色复合材料。(b) 地面真相。(c) 培训样本。(d) Legend.

CapsViT很接近，它们之间的差异只有0.11%，这也证明了CapsGLOM可以很好地学习不同位置之间的全局相关性。最后，尽管HybridSN的成功可能是由于使用了具有更大空间大小的训练样本，但它在HU2013数据集上的性能值得称赞。OA值为94.38%，仍然相当平淡。

C. 实验3：帕维亚大学的数据集

帕维亚大学（PU）的数据集是由ROSIS（反射光学系统成像光谱仪）在2001年在帕维亚大学上空收集的。该数据集包

含103个光谱波段，波段范围为430~860nm，空间分辨率为1.3 m。有9个土地覆盖类，图像的空间维度为610 340。图12a和图12b表示假彩色图像和相应的地面真实值。

对于该数据集，有42776个样本，由于该数据集的高空间分辨率，我们从每个类别中随机选择10个样本，形成初始训练集。所选的初始训练样本如图所示。12c. 由于这个数据集只包含103个光谱波段，在AL中

实验中，所有比较模型的训练期均调整到150，以避免过拟合。考虑到专家注释的成本，将AL迭代次数设置为5次。此外，为了避免由于小批量中样本过少而导致的训练不稳定，我们将所有比较模型的批大小都设置为30。关键是对PU数据集，有大量的样本，为了模拟真实场景中的应用，从每个类别中选择10%（未标记）的样本来构建候选样本集。A2S2K-RN、CapsViT和CapsGLOM经过不同AL迭代后得到的分类结果如表X所示。

正如我们可以看到，在初始训练集上，A2S2K-RN表现最好，CapsGLOM也取得了类似的结果。虽然CapsViT的表现不如A2S2K-RN和CapsGLOM，但它是可以接受的。经过3次或5次AL迭代后，A2S2K-RN仍然最好，其次是CapsGLOM，CapsViT表现不佳。A2S2K-RN能够表现得如此之好，主要是因为它可以很好地学习光谱-空间特征的抽象表示。当然，在这种情况下，我们可以使用A2S2K-RN来代替CapsViT。最后，A2S2K-RN、CapsViT和CapsGLOM经过5次AL迭代得到的分类图如图所示。13.

表x

利用90个初始训练样本对pu数据集进行主动学习实验。

课		主动学习迭代=0			主动学习迭代=3			主动学习迭代=5		
		A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM
准确度 (%)	C1	87. 21	87. 24	89. 29	96. 60	91. 69	92. 07	92. 98	88. 69	93. 49
	C2	86. 35	84. 25	83. 93	94. 70	88. 72	94. 54	97. 97	94. 67	97. 45
	C3	94. 83	72. 52	86. 55	94. 02	79. 22	90. 33	96. 98	92. 10	80. 33
	C4	98. 00	93. 91	88. 47	98. 13	99. 28	96. 66	98. 59	96. 95	97. 02
	C5	99. 63	100. 00	99. 48	100. 00	100. 00	100. 00	100. 00	100. 00	100. 00
	C6	82. 92	82. 25	92. 19	87. 05	93. 33	88. 09	93. 52	77. 70	86. 93
	C7	93. 86	91. 14	93. 56	98. 03	92. 42	95. 76	99. 24	90. 30	96. 82
	C8	77. 59	86. 96	84. 80	92. 40	82. 76	94. 83	90. 47	92. 21	95. 51
	C9	99. 68	99. 79	98. 51	98. 93	99. 57	97. 65	98. 83	99. 89	96. 05
OA (%)		87. 52	85. 88	87. 37	94. 47	90. 21	93. 65	96. 15	91. 72	94. 59
AA (%)		91. 12	88. 67	90. 75	95. 54	91. 89	94. 44	96. 51	92. 50	93. 73
F1		0. 8376	0. 8166	0. 8371	0. 9270	0. 8727	0. 9161	0. 9490	0. 8903	0. 9282

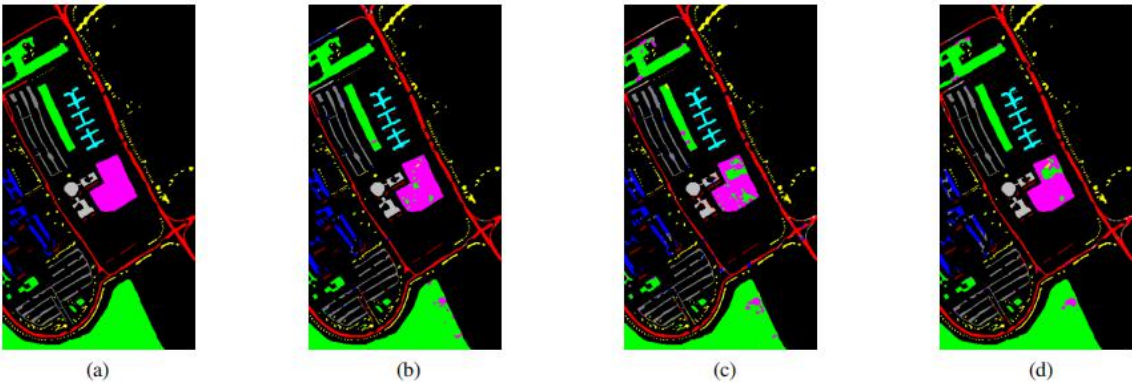


图13. 不同模型在PU数据集上进行5次AL迭代后得到的分类图。(a)地面真相。分别由A2S2K-RN、CapsViT和CapsGLOM获得的(b)-(d)分类图。

表xi

使用2%的训练样本对pu数据集进行监督分类实验。

课		HybridSN	卡维特	A2S2K-RN	DMCapsNet	斯弗特	光谱网	光谱元件	CapsViT	CapsGLOM
准确度 (%)	C1	99. ±120. 5	95. ±562. 9	99. ±120. 5	97. ±680. 7	98. ±970. 3	93. ±444. 4	89. 061. 71	98. ±840. 5	98. ±820. 6
	C2	8	7	7	0	3	±444. 4±641	98. ±580	6	8
	C3	99. ±910. 1	99. ±880. 0	99. ±770. 2	99. ±860. 0	99. ±730. 2	. 83	. 56	99. ±880. 1	99. ±860. 1
	C4	3	6	6	8	2	98. 850. 26	68. ±815. 9	6	0
	C5	93. ±122. 0	88. ±233. 1	94. ±902. 3	93. ±533. 2	85. ±383. 9	77. 375. 48	7	89. ±693. 4	92. ±633. 9
	C6	7	5	7	9	9	93. 632. 08	94. ±561. 5	3	9
	C7	95. ±161. 3	97. ±771. 1	98. ±750. 9	84. ±571. 6	97. ±541. 7	98. 811. 91	7	97. ±940. 5	97. ±480. 7
	C8	3	6	5	3	3	94. 632. 46	99. ±860. 1	6	3
	C9	99. ±810. 2	99. ±910. 1	99. ±950. 0	99. ±330. 5	99. ±990. 0	91. 421. 64	7	99. ±980. 0	99. ±580. 7
		9	4	8	9	2	93. 812. 21	86. ±423. 3	5	4
		99. ±970. 0	95. ±663. 1	99. ±670. 5	99. ±920. 1	98. ±021. 6	85. 5310. 47	9	99. ±710. 3	99. ±810. 2
		4	5	5	1	8	72. ±805. 3	8	7	7
		99. ±290. 8	89. ±553. 4	96. ±652. 3	99. ±310. 8	98. ±990. 9	5	95. ±732. 6	95. ±592. 8	
		3	9	4	2	1	79. ±612. 1	3	0	
		96. ±142. 0	94. ±851. 0	97. ±971. 6	95. ±061. 8	96. ±481. 4	2	98. ±141. 4	97. ±461. 4	
		1	1	6	4	4	96. ±831. 6	4	4	
		87. ±804. 2	99. ±890. 1	99. ±800. 2	81. ±774. 1	98. ±740. 9	6	99. ±820. 2	98. ±361. 9	
		6	6	3	8	6		1	9	
OA (%)	98. ±510. 30	97. ±240. 67	99. ±100. 29	97. ±270. 17	98. ±230. 41	95. ±160. 45	91. ±501. 00	98. ±790. 22	98. ±790. 24	
AA (%)	96. ±700. 58	95. ±700. 84	98. ±510. 41	94. ±560. 44	97. ±090. 64	91. ±971. 51	87. 391. 37±	97. ±750. 39	97. ±730. 58	
F1	±0. 98020. 004	±0. 96330. 009	±0. 98810. 004	±0. 96380. 002	±0. 97660. 005	±0. 93580. 006	±0. 88670. 013	±0. 98390. 003	±0. 98390. 003	

为了验证我们提出的CapsViT和CapsGLOM的优势，我们在PU数据集上进行了监督分类实验。每个类别中选取2%的样本作为训练样本（共选择858个样本），实验结果见表XI。可以发现，当训练样本足够时，虽然A2S2K-RN仍然是性能最好的模型，但我们提出的CapsViT和CapsGLOM是最具竞争力的

两个模型，它们与OA中的A2S2K-RN之间的差异仅为0. 31%。一个值得注意的结果是，CapsViT和CapsGLOM的平均OA均为98. 79%，但CapsViT达到了更高的AA值。

D. 实验4: 赤石数据集

赤城的数据集是2014年由头墙高透视vnir-c成像传感器在日本茨城市的农业和城市地区采集的。该数据集在363 nm到1018 nm的光谱范围内有128个波段。✕该场景由2517

2335个像素组成，空间分辨率为2.5米。19类的地面真相包括水、三种裸露的土壤、7种植被和8种人造物体。图14a和图。14b表示假彩色图像和相应的地面真实值。

该数据集有77592个样本和5个样本

授权许可使用仅限于：沈阳航空航天大学。已于2023年9月10日12:14:07 UTC从IEEE Xplore下载。限制适用。

© 2023 IEEE。允许个人使用，但再版/再分发需要IEEE的许可。看到<https://www.ieee.org/publications/rights/index.html>获取更多信息。

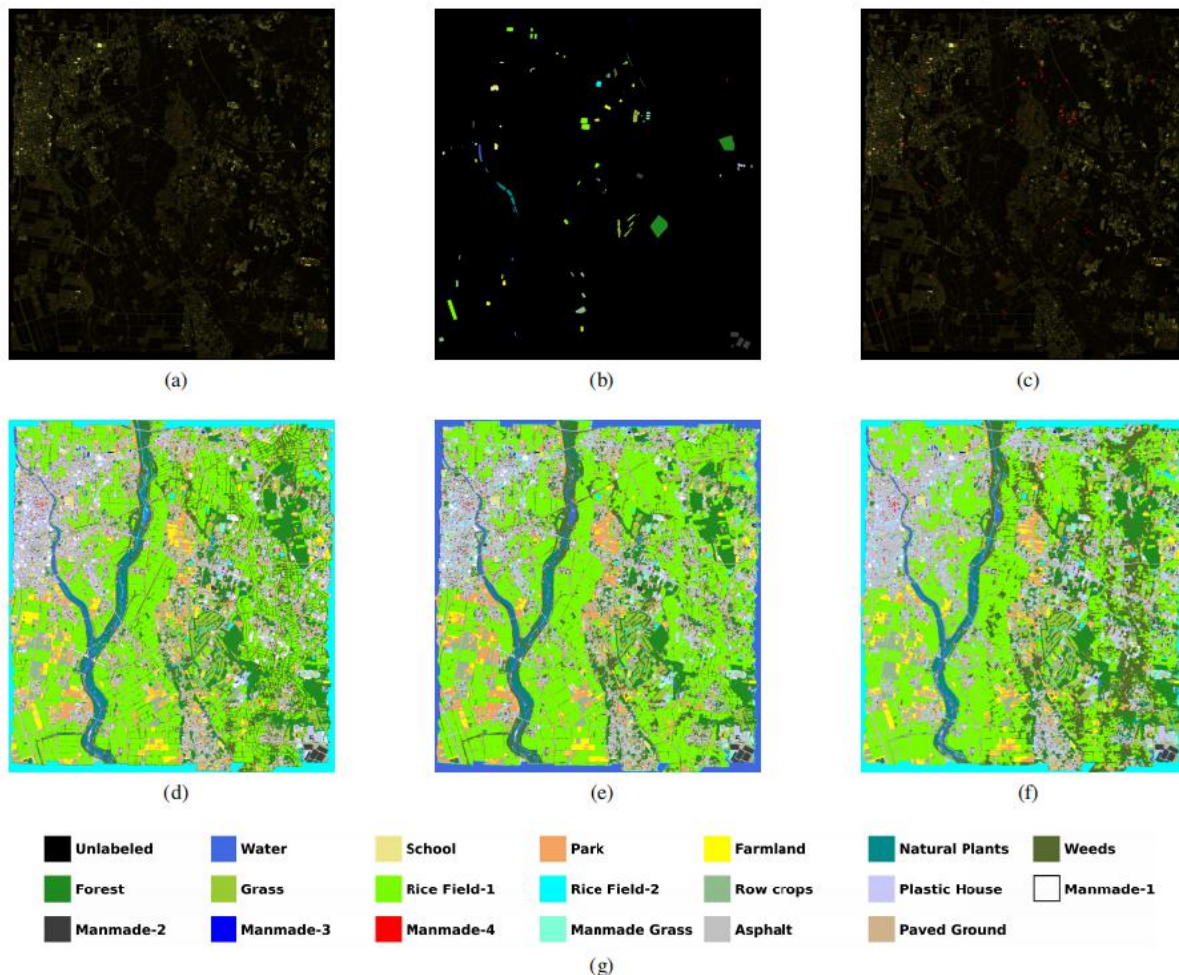


图14. 在赤木数据集上不同模型获得的分类图的可视化。(a)三波段假色复合材料。(b)地面真相。(c)培训样本。(d)-(f)分别由A2S2K-RN、CapsViT和CapsGLOM经过5次AL迭代后得到的分类图。(g)图例。

从每个类别中随机选择来构建初始训练集。此外，从每个类别中随机抽取10%的（未标记）样本来构建候选样本集。在AL实验中，为了避免过拟合，A2S2K-RN、CapsViT和CapsGLOM的训练轮分别设置为150、200和200。批量大小被设置为32。A2S2KRN、CapsViT和CapsGLOM经过不同AL迭代后得到的分类结果见表XII⁶。通过比较实验结果，我们可以发现CapsViT获得了最高的OA和Kappa，CapsGLOM在初始训练集上获得了最高的AA。这一发现再次表明，当训练样本稀缺时，CapsViT和CapsGLOM具有更好的分类性能。经过3个AL步骤后，CapsGLOM的性能改善最大，同时也表现最好。经过5个主动学习步骤，虽然CapsViT和CapsGLOM的性能低于A2S2K-RN，但得到了令人满意的分类结果，说明我们设计的CAL方案是成功的。值得注意的是，尽管所提议的CAL框架可以

⁶为了紧凑，我们没有给出每个类的准确性，感兴趣的读者可以自己重现我们的实验。

将CapsViT和CapsGLOM的OA分别从93.80%和93.01%提高到98.73%和98.77%，只需要5次AL迭代，最优的AL迭代取决于专家注释的预算。最后，我们在图中给出了由A2S2K-RN、CapsViT和CapsGLOM得到的全景分类图。14d-无花果。14f.全景分类图可以更好地描述场景的景观，这取决于分类器的泛化能力。通过看图。14d-无花果。在14f时，我们可以看到，由于A2S2K-RN具有更好的分类性能，因此通过它获得的全景分类图具有更好的纹理细节。由此我们可以得出结论，虽然CapsViT和CapsGLOM不是最好的分类器，但CAL主动选择的样本也可以用于训练其他优秀的分类器，这完全符合AL的核心思想，即数据效率。

为了充分衡量CapsViT和CapsGLOM的分类性能，我们在芯片数据集上进行了监督分类实验。从每个类别中随机抽取30个样本构建训练集，其余样本作为测试样本。所有比较模型的结果见表十三。通过观察结果，我们可以看到，而

表XII
利用95个初始训练样本对芯片数据集进行主动学习实验。

指标	主动学习迭代=0			主动学习迭代=3			主动学习迭代=5		
	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM	A2S2K-RN	CapsViT	CapsGLOM
OA (%)	93.03	93.80	93.01	96.16	95.89	98.14	99.55	98.73	98.77
AA (%)	93.19	92.95	93.85	96.30	96.08	96.54	99.56	97.08	98.50
κ	0.9196	0.9286	0.9194	0.9558	0.9526	0.9786	0.9948	0.9853	0.9858

表13
在芯片数据集上进行监督分类实验，每个类使用30个训练样本。

指标	HybridSN	CAEVT	A2S2K	RN	DMCapsNet	SSFTT	SpectralNET	SpectralFormer	CapsViT	CapsGLOM
OA (%)	98.140.63	980.47	99.120.40	98.850.58	98.600.56	94.481.04			98.±760.31	98.±770.25
AA (%)	96.750.70±.45±±±±±±±±	980.34	99.220.38	98.900.62	98.810.38	94.900.88			98.±690.52	98.±720.21
κ	95.990.48±.42±±±±±±±±	0.98210.005	0.98980.005	0.98680.007	0.98390.006	0.93650.012			±0.98560.004	±0.98580.003

A2S2K-RN保持领先，其后的DMCapsNet排名第二，CapsViT和CapsGLOM是A2S2K-RN和DMCapsNet之后表现最好的模型。此外，虽然我们使用了更多的训练样本来训练CapsGLOM，但它得到的平均OA和h值，即98.77%和0.9858，与5次AL迭代后的CAL框架得到的值一致，证明了我们设计的CAL框架的有效性。

最后，为了清楚地了解被比较的每个模型的复杂性，我们还在表十三中给出了它们的平均训练和测试时间。从这里，我们可以看到，自从SSFTT的模型构建是最轻的，消耗的时间最少。此外，HybridSN的结构相对简单，也不需要太多的时间。由于DMCapsNet需要训练视点不变变换矩阵，以及是否通过EM迭代算法激活胶囊的成本，因此也需要大量的时间来训练和测试。由于ViT模块的引入，CapsViT需要更多的时间来训练和预测。此外，尽管CapsGLOM在性能上非常接近CapsViT，但在所有模型中，所需的时间都是适中的。顺便说一下，当CapsViT应用于具有较大空间尺寸的训练样本时，最好先使用步幅卷积来减小空间尺寸，否则会构建太多的主胶囊，从而增加模型的复杂性。

V. 讨论

为了清楚地了解已开发的CapsViT和CapsGLOM的有效性、灵敏度和复杂性，我们进行了广泛的分析实验。所有的分析实验都是在PU数据集上进行的，因为它是遥感界最经典和最广泛使用的数据集之一。

A. 消融研究

为了进一步证实CapsViT和CapsGLOM中各成分的有效性，我们进行了消融研究。在这一部分中，为了消除不同的训练样本对每种消融方法的影响，我们使用了一个随机的种子（种子=0）来选择固定的训练样本。除此之外，

消除了模型初始化对实验结果的影响，所有的消融实验都进行了10次，以报告更稳定和可靠的平均结果。与在PU数据集上的监督分类实验一样，我们从地面真相中随机抽取2%的样本作为训练样本，所有的消融方法都训练了200个epoch。

表十四为消融实验的结果，其中OA、AA、h均为10次实验的平均值。对于CapsViT，当我们同时删除ConvGRU和ViT时，它的性能显著恶化。去除ViT后，该模型对应于传统的CapsNet和ConvGRU模块。我们可以看到，它的性能与原始模型相差不远，与同时去除ConvGRU和ViT的方法相比，OA的性能提高了0.27%，证明了其有效性的ConvGRU模块。当我们删除ConvGRU时，它的性能没有下降太大，OA提高了与同时去除ConvGRU和ViT的方法相比，其差异为0.32%，从而证明了ViT模块的有效性。此外，我们可以观察到ViT比ConvGRU对CapsViT带来了更多的改进。对于CapsGLOM，当删除ConvGRU时，性能下降，但不显著，这证明了所提出的CapsGLOM是成功的，而ConvGRU模块是有效的。

为了更好地观察CapsViT和CapsGLOM及其组件的复杂性，我们还在表十四中给出了每种消融方法的可训练参数数、浮点操作（FLOPs）和平均训练时间。结果表明，ViT占据了CapsViT的主要计算量。因此，我们建议在将CapsViT应用于大空间大小的训练样本时，最好使用步长卷积来减小空间大小，否则会构建太多的主胶囊，从而增加模型的复杂性。相比之下，CapsGLOM相对轻量级，这也暗示了其无限的潜力和魅力。

B. 训练样本数量的影响

为了研究不同数量的训练样本对CapsViT和CapsGLOM的影响，我们进行了扩展

表XIV
在pu数据集上进行的烧蚀实验。

方法	ConvGRU	ViT	可训练参数	×成本 (106FLOP)	培训时间(s)	OA	AA	κ
没有ConvGRU和其他版本的版本			1, 103, 168	12. 658	70. 64	98. 53	97. 41	0. 9805
CapsViT		/	4, 804, 616	50. 943	286. 53	98. 85	98. 10	0. 9848
	/		1, 133, 936	16. 877	162. 43	98. 80	98. 09	0. 9841
	/	/	4, 835, 384	55. 162	340. 40	98. 89	98. 07	0. 9852
CapsGLOM w/o ConvGRU		-	336, 507	26. 713	81. 60	98. 82	98. 03	0. 9844
CapsGLOM	/	-	413, 355	32. 895	192. 44	98. 85	98. 03	0. 9848

在PU数据集上的实验。具体来说，我们从每个类中随机选择10个100个训练样本来观察CapsViT和CapsGLOM的性能变化。为了控制实验中的变量，并简单地比较CapsViT和CapsGLOM，我们使用随机种子（种子=0）来选择训练样本，并控制CapsViT和CapsGLOM的初始化。结果如图所示。15。

从图中可以看出，随着训练样本数量的增加，CapsViT和CapsGLOM的性能总体上呈上升趋势。当各类别的训练样本数量达到60个时，上升趋势相对稳定。这是因为，尽管每个类别的训练样本数量增加了，但额外的样本并不是真正提供信息，可能包含大量重复的信息。正如我们所看到的AL实验数据集，虽然每个类别的初始样本只有10, 5步骤后，也就是说，45样本积极标记，OACapsViT和CapsGLOM可以达到91. 72%和94. 59%，此时，训练样本的比例只有0. 3256%。但是，要超过91. 72%的OA，CapsViT可能需要从每个类别中随机选择18个样本（0. 38%）；要超过94. 59%的OA，CapsGLOM可能需要从每个类别中随机选择20个样本（0. 42%）。

最后，重要的是要注意，尽管默认的训练轮CapsViT和CapsGLOM200，在实际监督分类应用程序，我们建议设置一个更大的训练轮，然后决定是否尽早停止，以确保最优模型收购基于准确性或损失的验证集。

C. CapsViT重建减重量的影响

从(9)中可以看出，将重建损失权重 α 作为正则化系数来约束CapsViT，使其具有一定的重建能力。权值过大会使模型更加关注其编码能力，而权值过小会削弱模型的重建能力，失去正则化约束。为了研究不同的重构损失权值对CapsViT性能的影响，我们在PU数据集上进行了比较实验。我们从{0. 001、0. 005、0. 01、0. 05、0的集合中选择了 α 的值。1}。此外，为了控制实验变量，我们使用随机种子来控制训练样本的选择（2%）和模型的初始化。图16显示了在PU数据集上使用不同的 α 值时CapsViT的性能。

通过观察柱状图，可以发现不同的 α 值会导致不同的性能，当将其设置为0. 01时，CapsViT获得了最好的性能。虽然大多数性能差异小于0. 2%，但考虑到在许多实际应用中，实例的高级表示具有重要意义，但重构损失作为一个重要超参数的权重需要仔细调整。当然，0. 01只是 α 的默认值，当CapsViT被部署在一个真实的应用程序中时，我们建议使用网格搜索来为 α 分配一个适当的值。

VI . 结论

在本文中，我们提出了一个基于改进的HIC胶囊网络的CAL框架。所提出的CAL框架包括两个优秀的基分类器和一个CAL方案

这同时考虑了主动选择的样本的多样性和专家注释的成本。第一个基分类器是CapsNet与ViT的结合，它被命名为CapsViT。通过引入ViT，CapsViT可以更好地了解不同初级胶囊之间的全局相关性，提高分类性能。第二个基分类器是基于改进的GLOM系统设计的，命名为CapsGLOM。CapsGLOM没有使用预先分配的通用胶囊来学习不同位置的多级嵌入，而是借鉴了CapsNet构建主胶囊的方式，并让网络自己学习这些嵌入。在四个基准HSI数据集上进行的AL实验表明，我们提出的CapsViT和CapsGLOM可以在初始训练集上获得良好的甚至最好的分类性能，这为后续主动选择信息最丰富的样本提供了道路。同时，所提出的CAL框架是灵活的，任何优秀的深度模型都可以用来取代我们设计的两个基本分类器中的任何一个，只要它可以处理具有相同形式（维度）的样本。此外，与先进的深度模型相比，所提出的CapsViT和CapsGLOM可以在四个数据集上的监督分类实验中获得非常有竞争力的结果。

最后，我们未来的工作主要体现在以下两个方面：1) CapsViT在原始CapsNet的基础上引入了ViT，导致了相对较高的复杂度。如何降低CapsViT的复杂性将是一项有意义的工作；2) 鉴于HU2013数据集上轻量级SSFTT和芯片数据集上DMCapsNet上的优异性能，它们的优势可用于提高CapsViT和CapsGLOM；3) 提高性能

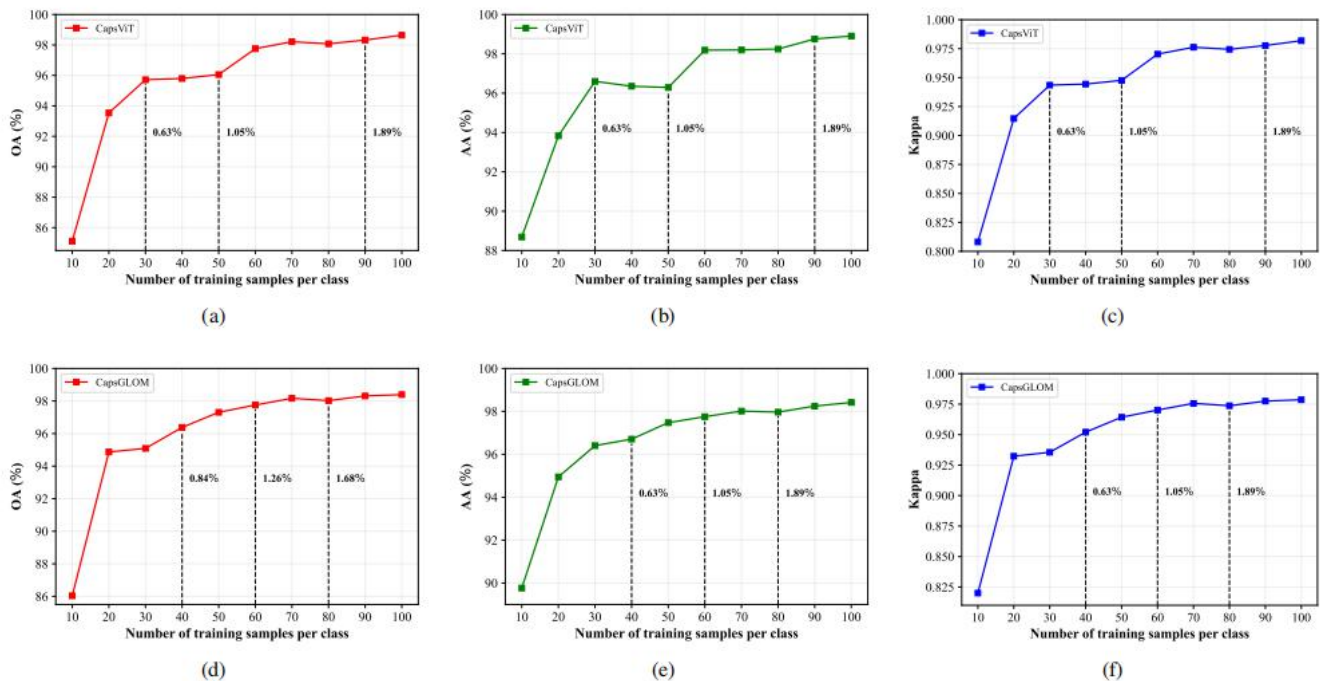


图15. 不同训练样本数下的CapsViT和CapsGLOM在PU数据集上的性能。CapsViT的(a)–(c)性能。(d)–(f)CapsGLOM的性能。虚线附近的百分比是训练样本的百分比。

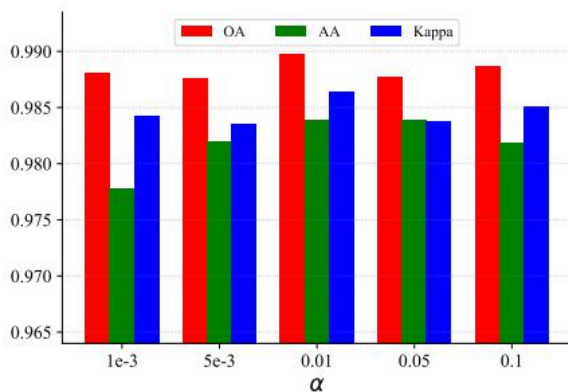


图16. CapsViT中重建损失权重 α 对PU数据集的敏感性。

从可解释性的CapsGLOM将有助于HIC的未来研究。

确认

国家自然科学基金资助项目62071084，国家民族事务委员会领导人才项目。

参考文献

- [1] D. 达塔, P. K. 马利克, K. Bhoi, M. F. 伊贾兹, J. Shafi和J. 崔, “高光谱图像分类: 潜力、挑战和未来方向”, 《计算智能与神经科学》, 第1卷。2022, 2022.

- [2] X. 曹, J. 姚, Z. Xu和D. 孟, “高光谱图像分类”与卷积神经网络和主动学习, “IEEE地球科学与遥感学报”, 卷。58岁, 没有。7, pp. 4604 – 4616, 2020.
- [3] Y. 陈, L. 黄, L. 朱, N. 横须贺和X. 贾, “细粒度”基于深度学习的高光谱图像分类, “遥感”, 卷。11日, 没有。22, 2019.
- [4] S. K. 罗伊, G. 克里希纳, S. R. Dubey和B. B. “Hybridsnn: 探索高光谱图像分类的3D-2D cnn特征层次结构”, “IEEE地球科学与遥感字母”, 第1卷。17日, 没有。2, pp. 277 – 281, 2020.
- [5] Y. 蔡, X. 刘和Z. 蔡, “b网: 乐队的端到端框架”《高光谱图像的选择》, IEEE地球科学与遥感学报, 卷。58岁, 没有。3, pp. 1969 – 1984, 2020.
- [6] Z. 张, T. 李, X. 唐, X. 胡和Y. 彭, “CAEVT: 卷积自动编码器满足轻量级视觉变压器的高光谱图像分类”, “传感器”, 卷。22日, 没有。10, 2022.
- [7] Z. 郑, Y. 钟, A. 妈妈和L. 张, 《FPGA: 快速无补丁》全端到端高光谱图像的全局学习框架分类, “IEEE地球科学与遥感学报”, 第1卷。58岁, 没有。8, pp. 5612 – 5626, 2020.
- [8] Q. 朱, W. 邓, Z. 郑, Y. 钟, 问。关, W. 林, L. 张, 和D. “光谱高光谱图像分类的光谱空间依赖全局学习框架”, 《控制学报》, 第1卷。PP, pp. 1 – 15, 2021.
- [9] L. 王, H. 王, L. 王, X. 王, Y. Shi和Y. Cui, “RSSGL: 统计损失正则化三维凸出的高光谱图像分类”, “IEEE地球科学与遥感学报”, 第1卷。60, pp. 1 – 20, 2022.
- [10] S. K. 罗伊, S. 曼娜, T. 宋和L. “基于注意力的广告-高光谱图像分类的选择性光谱空间核研究”, “IEEE地球科学与遥感学报”, 第1卷。59岁, 没有。9, pp. 7831 – 7843, 2021.
- [11] M. E. Paoletti, J. M. 蒙特, R. 费尔南德斯-贝尔特兰, J. 广场, A. 广场, J. 列支敦士登和F. “高光谱图像分类的胶囊网络”, 《IEEE地球科学与遥感学报》, 第1卷。57岁, 没有。4, pp. 2145 – 2160, 2019.
- [12] X. 唐, F. 孟, X. 张, Y. M. 一张, J. 妈妈, F. 刘和L. 焦, “基于空间光谱注意网络的三维八度卷积的高光谱图像分类”, 《IEEE地球科学与遥感学报》, 第1卷。59岁, 没有。3, pp. 2430 – 2447, 2021.
- [13] T. 查克拉伯蒂和美国。 “光谱网: 探索空间光谱”

- 小波网络为高光谱图像分类，“arXiv电子打印，4月。2021。
- [14] W. 胡，H. 李，L. 平移，W. 李，R. 道和Q. 《空间光谱特征》利用深度凸神经网络提取高光谱图像分类，“IEEE地球科学与遥感学报”，第1卷。58岁，没有。6，pp. 4237–4250，2020。
- [15] W. 胡，H. 李，Y. 邓，X. 太阳，Q. 杜和A. 广场，轻量级张量注意驱动的凸神经网络，“IEEE《信号处理选择主题杂志》，第1卷。15日，没有。3，pp. 734–745，2021。
- [16] D. 红，L. 高，J. 姚，B. 张，A. 广场和J. 查努索，“图高光谱图像分类的卷积网络”，“IEEE地球科学与遥感学报”，第1卷。59岁，没有。7，pp. 5966–5978，2021。
- [17] A. Vaswani, N. Shazeer, N. 帕尔马，J. Uszkoreit, L. 琼斯。N. 戈麦斯 L. 凯泽和我。“注意是你所需要的”，《神经信息处理系统的进展》，卷。30。柯伦联合公司，2017年，第3页。6000–6010。
- [18] A. Dosovitskiy, L. 拜尔，A. 科尔斯尼科夫，D. 魏森伯恩，X. 翟，T. Unterthiner, M. 德哈尼，M. 经理，G. Heigold, S. Gelly等。，“一幅图像值16x16字：用于大规模图像识别的变形金刚”，arXiv预印本 arXiv: 2010.11929, 2020年。
- [19] X. 他，Y. 陈和Z. 林，“空间光谱变压器
窦图像分类，“遥感，卷。13日，没有。3，2021。
- [20] D. 香港，Z. 韩，J. 姚，L. 高，B. 张，A. 广场和J. Chanus-“光谱形成者：用变压器重新思考高光谱图像分类》，《IEEE地球科学与遥感学报》，第1卷。60，pp. 1–15，2022。
- [21] Z. 赵，D. 胡，H. 王和X. 余，“卷积变压器”高光谱图像分类网络，“IEEE地球科学与遥感字母”，第1卷。19，pp. 1–5，2022。
- [22] D. 王，J. 张，B. 杜，L. 张和D. 道，“DCN-T：双重上下文高光谱图像分类，“IEEE图像处理事务，vol. 32，pp. 2536–2551，2023。
- [23] L. 太阳，G. 赵，Y. 郑和Z. 吴，“光谱-空间特征”高光谱图像分类的标记化变压器，“IEEE地球科学与遥感学报，卷。60，pp. 1–14，2022。
- [24] A. 拉维库马尔。N. Rohit, M. K. Nair和V. Bhatia，高光谱使用深度矩阵胶囊进行图像分类”，2022年国际数据科学、代理和人工智能国际会议（ICDAAI），卷。01，2022，pp. 1–7。
- [25] G. E. 辛顿，S. Sabour和N. “矩阵胶囊与他们”在2018年的学习代表国际会议上。
- [26] S. S. 萨万特和M. “基于半监督技术”超光谱图像分类：一项调查，“2017年功率和先进计算技术的创新（i-PACT），页。1–8，2017。
- [27] C. 丁，Y. 李，Y. 温，M. 郑，L. 张，W. 魏和Y. 张，“利用伪标签学习提高少镜头高光谱图像分类”，遥感，卷。13日，没有。17，2021。
- [28] R. 悬挂，F. 周，问。刘和P. “超分类”光谱图像通过多任务生成对抗网络，“IEEE地球科学与遥感学报”，第1卷。59岁，没有。2，pp. 1424–1436，2021。
- [29] H. 梁，W. 鲍，X. 沈和X. 张教授，“光谱-空间注意”《基于生成对抗网络的高光谱图像分类特征提取》，《应用地球观测与遥感精选专题期刊》，第1卷。14，pp. 10 017–10 032，2021。
- [30] H. 吴和S. “半监督的深度学习使用伪高光谱图像分类的标签，“IEEE图像处理事务，vol. 27日，没有。3，pp. 1259–1270，2018。
- [31] B. 方，Y. 李，H. 张和J. cw. -陈，“协作学习”轻量级卷积神经网络和深度聚类的高光谱图像半监督分类与有限的训练样本，“ISPRS摄影测量和遥感杂志，卷。161，pp. 164–178，2020。
- [32] M. Seydgar, S. Rahnamayan, P. Ghamisi和A. Bidgoli, “Semisu-使用概率伪标签生成框架的监督高光谱图像分类，“IEEE地球科学与遥感交易，卷。60，pp. 1–18，2022。
- [33] H. 曾，问。刘，M. 张，X. 韩和Y. 王，“半监督利用图聚类卷积网络的高光谱图像分类，”arXiv预印本arXiv: 2012.10932, 2020。
- [34] S. 贾，S. 江，Z. 林，M. 徐，W. 太阳，Q. 黄，J. 朱和X. 贾，“A用于高光谱图像分类的半监督西亚罗网络，”
- 《IEEE《地球科学与遥感学报》，第1卷。60，pp. 1–17，2022。
- [35] J. 伙伴伊恩，宾夕法尼亚州。-琼，M. Mehdi, X. 必应，WF. -大卫 O. 谢吉尔和C. 亚伦，“生成对抗网络”第27届神经信息处理系统国际会议论文集，第1卷。2，2014，pp. 2672–2680。
- [36] M. Arjovsky, S. 钦塔拉和L. 波图，瓦瑟斯坦甘，arXiv预印本，arXiv: 1701.07875v3, 2017年。
- [37] I. Gulrajani, F. 艾哈迈德，M. Arjovsky, V. 杜莫林和A. C. 考维尔《神经信息处理系统的进展》，第1卷。30，2017。
- [38] A. 克雷斯基尔，T. 白色，V. 杜莫林，K. 阿鲁库马兰，B. 森古普塔和A. A. “生成对抗网络：概述”，IEEE信号处理杂志，第1卷。35岁，没有。1，pp. 53–65，2018。
- [39] Z. 王，问。她和T. E. 生成对抗网络
计算机视觉：一个调查和分类法，“ACM计算调查（CSUR），卷。54岁，没有。2，pp. 1–38，2021。
- [40] X. 陈，Y. 段，R. Houthoof, J. 舒尔曼，我。萨特斯克利夫和P. “英福根：通过最大化生成对抗网络的信息学习可解释表示”，神经信息处理系统的进展，卷。29，2016。
- [41] T. Miyato, T. 金冈，M. 山和Y. 吉田，“光谱生成式对抗网络的标准化，”arXiv预印本arXiv: 1802.05957, 2018。
- [42] L. 朱，Y. 陈，P. Ghamisi和J. A. “生成广告”关于高光谱图像分类的讽刺网络，“IEEE地球科学与遥感学报”，第1卷。56岁，没有。9，pp. 5046–5063，2018。
- [43] Y. 詹，D. 胡，Y. 王和X. 半监督高光谱研究基于生成对抗网络的图像分类，“IEEE地球科学与遥感字母”，第1卷。15日，没有。2，pp. 212–216，2018。
- [44] J. 冯，X. 冯，J. 陈，X. 曹，X. 张，L. 焦和T. 南斯拉夫“基于协同学习和注意机制的生成对抗网络用于高光谱图像分类”，遥感，卷。12日，没有。7，pp. 212–216，2020。
- [45] C. 陶，H. 王，J. 齐和H. 李，“半监督变分基因-“高光谱图像分类的有效对抗性网络”，《应用地球观测与遥感精选专题杂志》，第1卷。13，pp. 914–927，2020。
- [46] X. 王，K. 谭，Q. 杜，Y. 陈和P. Du，“CVA2E：一个条件变分自编码器与对抗性训练过程的高光谱图像分类，“IEEE地球科学与遥感学报，卷。58岁，没有。8，pp. 5676–5692，2020。
- [47] Z. 陈，L. 童，B. 钱，J. Yu和C. 肖，“以自我关注为基础条件变分自动编码器生成对抗网络，“遥感，卷。13日，没有。16，2021。
- [48] Z. 李，X. 朱，Z. 新，F. 郭，X. 崔和L. 王，“变分基因-具有交叉空间和光谱相互作用的对抗网络用于高光谱图像分类，“遥感，卷。13日，没有。16，2021。
- [49] WY. -王，HC. -李，YJ. -邓，我。-邵，XQ. -卢和Q. 杜，“高光谱图像分类的生成对抗胶囊网络”，IEEE地球科学与遥感字母，卷。18日，没有。3，pp. 523–527，2021。
- [50] X. 王，K. 谭，Q. 杜，Y. 陈和P. 杜，“帽子-三人组：甘-高光谱图像分类，IEEE地球科学与遥感学报，卷。57岁，没有。9，pp. 7232–7245，2019。
- [51] J. 王，S. 郭，R. 黄，L. 李，X. 张和L. 焦，“双高光谱图像分类的通道胶囊生成对抗网络，“IEEE地球科学与遥感学报”，第1卷。60，pp. 1–16，2022。
- [52] W. 刘，J. 你和我。“HSIGAN：一个条件高光谱图像《利用辅助分类器的合成方法》，《应用地球观测与遥感精选专题杂志》，第1卷。14，pp. 3330–3344，2021。
- [53] Y. 曹，Y. 王，J. 彭，C. 邱，L. 丁和X. X. 朱，“SDFL-FC：高光谱图像分类的半监督深度特征学习，“IEEE地球科学与遥感学报》，卷。59岁，没有。12，pp. 10 488–10 502，2021。
- [54] W. 赵，X. 陈，J. 陈和Y. 曲，“样品生成与自注意生成对抗性适应网络（sagaan），“遥感》，卷。12日，没有。5，p. 843，2020。
- [55] J. 冯，H. 余，L. 王，X. 曹，X. 张和L. 焦，“分类”基于多类空间光谱生成的高光谱图像

- 《地球科学与遥感学报》，第1卷。57岁，没有。8，pp.5329 - 5343，2019。
- [56] S. 张，X. 张，T. 李，H. 孟，X. 曹和L. 王，“对抗”用小标记的高光谱图像分类的表示学习，“遥感，卷。14日，没有。11，2022。
- [57] J. 王，F. 高，J. Dong和Q. “自适应下降块增强高光谱图像分类的生成对抗网络”，“IEEE地球科学与遥感学报”，第1卷。59岁，没有。6，pp.5040 - 5053，2021。
- [58] F. 张，J. 白，J. 张，Z. 肖和C. 裴，“优化基于高光谱图像分类的训练方法”，“IEEE地球科学与遥感字母”，第1卷。18日，没有。10，pp.1791 - 1795，2021。
- [59] J. 施，T. 邵，X. 刘，X. 张，Z. 张和Y. 雷，“进化”“高光谱图像分类的多任务集成学习模型”，《应用地球观测与遥感精选专题杂志》，第1卷。14，pp.936 - 950，2021。
- [60] S. Hao, W. 王，Y. 叶，E. 李和L. Bruzzone，“一个深网络《超分辨率辅助高光谱图像分类与分类损失》，“IEEE地球科学与遥感学报，vol.56岁，没有。8，pp.4650 - 4663，2018。
- [61] Y. 张，W. 李，M. 张，S. 王，R. 道和Q. 杜，“图高光谱图像分类的信息聚合跨域少镜头学习，IEEE神经网络与学习系统IEEE学报，pp.1 - 14，2022。
- [62] Y. 张，M. 张，W. 李，S. 王和R. 道，“语言感知”《跨场景高光谱图像分类的领域泛化网络》，《IEEE地球科学与遥感学报》，第1卷。61，pp.1 - 12，2023。
- [63] Y. 徐，B. 杜和L. 张，“强大的自集成网络《光谱图像分类》，《神经网络与学习系统学报》，第3页。1 - 14，2022。
- [64] S. Hao, W. 王和M. “几何意识的深度循环”高光谱图像分类的神经网络，“IEEE地球科学与遥感学报，卷。59岁，没有。3，pp.2448 - 2460，2021。
- [65] Y. 张，W. 李，M. 张，Y. 曲，R. 陶和H. 齐，“拓扑”跨场景高光谱图像分类的结构和语义信息传递网络，《神经网络与学习系统学报》，第1卷。34岁，没有。6，pp.2817 - 2830，2023。
- [66] Y. 张，W. 李，W. 太阳，R. 道和Q. 杜，“单源域跨场景高光谱图像分类扩展网络”，“IEEE图像处理学报”，第1卷。32，pp.1498 - 1512，2023。
- [67] R. 梭罗，五。ACHARD, L. RisserB. 贝塞洛特和X. 布里奥特特“高光谱图像分类的主动学习：比较评论”，IEEE地球科学与遥感杂志，页。2 - 24，2022。
- [68] F. 施，Z. 王，M. 胡和G. 翟教授，“主动学习加上深度学习”可以建立经济效益和鲁棒的多通道图像模型：高光谱图像分类的一个案例，“传感器，vol.20岁，没有。17，pp.2 - 24，2020。
- [69] M. 徐，问。赵和S. 贾佳，“多视角空间光谱主动学习”关于高光谱图像分类，《IEEE地球科学与遥感学报》，第1卷。60，pp.1 - 15，2022。
- [70] P. 刘，H. 张和K. B. “分类的主动深度学习”IEEE《应用地球观测和遥感精选专题杂志》，第1卷。10，没有。2，pp.712 - 724，2017。
- [71] J. M. 蒙特，M. E. Paoletti, J. 广场，J. 李和A. 广场，“主动学习”利用一种新的贝叶斯方法，利用卷积神经网络进行高光谱图像分类。56岁，没有。11，pp.6440 - 6461，2018。
- [72] U. Patel, H. 戴夫和V. 《高光谱图像分类》使用基于不确定性和多样性的主动学习，“可扩展计算：实践与经验，vol.22日，没有。3，pp.283 - 293，2021。
- [73] Q. 王，M. 陈，J. 张，S. 康和Y. 王，“改善活跃高光谱图像半监督分类的深度学习”，“遥感”，第1卷。14日，没有。1，2022。
- [74] G. 王和P. Ren，“高光谱图像分类与特征-”面向对抗性主动学习，“遥感，卷。12日，没有。23，2020。
- [75] D. Yoo和我。S. Kweon，《主动学习的学习损失》，2019年IEEE/CVF计算机视觉和模式识别会议（CVPR），2019年，第3页。93 - 102。
- [76] Z. Lei, Y. 曾，P. 刘和X. “超规范的主动深度学习-不确定性学习的图像分类，IEEE地球科学与遥感字母，卷。19，pp.1 - 5，2022。
- [77] B. 刘，A. Yu, P. 张，L. 丁，W. 郭，K. 高和X. 左，“活跃”高光谱图像分类的深度密集连接卷积网络，“国际遥感杂志，第1卷。42岁，没有。15，pp.5915 - 5934，2021。
- [78] G. “如何在一个神经网络中表示部分-完整的层次结构-*arXiv预印本arXiv: 2102.12627, 2021年。*
- [79] A. 基尔希，J. V.阿默斯福特和Y. BatchBALD：高效以及针对深度贝叶斯主动学习的多样化批量获取，”arXiv: 1906.08158v2, 2019。
- [80] O. 发件人和S. “卷积神经的主动学习”“网络：一种核心集的方法，”arXiv: 1708.00489v4, 2017。
- [81] S. Sinha。以卜拉希米和T. 达雷尔，“变分对抗性是活跃的”2019年IEEE/CVF计算机视觉国际会议（ICCV），2019年，页。5971 - 5980。
- [82] K. Konyushkova, R. 施尼特曼和P. “学习主动学习”从数据来看，“2017。
- [83] B. Demir, C. 佩尔塞洛和L. 批处理模式主动学习遥感图像交互分类方法，“IEEE地球科学与遥感学报，卷。49岁，没有。3，pp.1014 - 1031，2011。



王恒收到了理学学士学位。2015年毕业于中国郑州轻工业大学计算机和通信工程学位。S. 2019年毕业于中国哈尔滨工程大学信息与通信工程学位。他正在攻读博士学位。D. 毕业于中国哈尔滨工程大学信息与通信工程学院。他的研究兴趣包括计算机视觉、深度学习、遥感和深度生成模型。



王立国（IEEE成员）获得M学位。S. 自然科学学位和博士学位。D. 2002年和2006年分别获得哈尔滨工业大学（HIT）工程学学位，2006年至2008年在哈尔滨工程大学信息与通信工程学院担任博士后研究职位，现任教授。

他的研究兴趣是遥感图像处理和机器学习。他已经出版了4本书，35项专利，以及230多篇论文