多模态信息交互的零样本分类及其在火星探测场景的应用

檀晓萌[1,2]，席博博[1,3]，薛长斌[1]，李云松[3]
（1.中国科学院国家空间科学中心复杂航天系统电子信息技术重点实验室，北京 100190；2.中国科学院大学，北京 100190；
3.西安电子科技大学通信工程学院，陕西 西安 710071）

摘　要：为应对传统基于图像的火星场景分类算法面对火星复杂未知环境、未见类别频发的情况时表现不佳的问题，研究了多模态信息交互的零样本分类算法，并探索了该算法向深空探测领域的迁移及优化。研究内容主要分为数据集构建和算法研究2个方面。在数据集构建方面，整合并重构了火星探测零样本分类数据集。在算法研究方面，提出了一种基于多模态特征交互的零样本场景分类算法，然后结合知识蒸馏技术，对其进行了模型压缩优化，使其可在保证零样本分类性能的同时，大幅降低模型的参数量和计算复杂度。为验证算法的有效性，进行了大量对比实验和可视化分析，结果证明了所提算法的可行性和有效性。

关键词：火星图像分类；多模态信息；零样本分类；知识蒸馏
中图分类号：TP391.4；V47
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025076

Zero-shot classification with multi-model information interaction and its application in Mars exploration scenarios

TAN Xiaomeng[1,2], XI Bobo[1,3], XUE Changbin[1], LI Yunsong[3]
1. National Space Science Center, Key Laboratory of Electronic Information Technology for Complex Aerospace Systems, Beijing 100190, China
2. University of Chinese Academy of Sciences, Beijing 100190, China
3. School of Cyber Engineering, Xi'dian University, Xi'an 710071, China

Abstract: To address the issue that traditional Mars image classification methods based on images do not perform well in the face of the complex and unknown environment of Mars and the frequent occurrence of unseen categories, a zero-shot classification algorithm of multi-modal information interaction was studied and the migration and optimization of the algorithm to the field of deep space exploration was explored. The research content was mainly divided into two aspects: dataset construction and algorithm research. In terms of dataset construction, the zero-shot classification dataset for Mars exploration was integrated and reconstructed. For algorithm research, a zero-shot scene classification algorithm based on multi-modal feature interaction was proposed. Then combined with knowledge distillation, the model was compressed and optimized, which greatly reduced the number of parameters and complexity of the model while ensuring the performance of zero-shot classification. To verify the effectiveness of the algorithm, a large number of comparative experiments and visualization analysis are carried out, and the results prove the feasibility and effectiveness of the proposed algorithm.

Keywords: Mars image classification, multi-modal information, zero-shot classification, knowledge distillation

收稿日期：2024–12–31；修回日期：2025–04–14
通信作者：席博博，xibobo1301@foxmail.com
基金项目：国家自然科学基金资助项目（No.62401434）
Foundation Item：The National Natural Science Foundation of China (No.62401434)

0 引言

深空探测是指对月球以及远的地外天体进行空间探测的活动[1]，其科学目标包括寻找地外宜居环境和生命信号，预防太阳活动和小天体撞击对地球的灾害性影响，探究太阳系及其行星的起源和演化历史等。具体的任务有探索物质（如水、有机物、氧气等）、识别天体形态和运行轨道、探测物质成分及矿产资源[2]等。火星是太阳系中与地球最相似的行星，可能保存着太阳系生命起源的时间和行星演化过程中灾难性变化的最好纪录，对研究地球起源与演化具有非常重要的比较意义。因此，火星成为探寻地外生命、探索生命起源与演化等重大科学问题最有价值的行星之一。此外，火星距离地球较近，是人类最有可能登陆的地外行星。这些因素促使火星成为国际行星探测的重点目标，也使其成为除月球外人类探索最多的地外天体[3-4]。人类对火星的探索活动包括环绕火星的遥感探测器和着陆火星的原位探测器。

当前的深空探测模式需要探测器在接近目标过程中进行探测并传回数据，随后地面人员开展数据提取、建模和分析，并生成控制指令回传给探测器。这种“地面测控站+航天器”的模式存在一些不足，如传输数据量有限。以火星探测为例，其表面地形多变、环境恶劣和沙尘暴频发，这意味着目标类别外观多变，存在随时出现的未见类别。这些因素使基于图像的火星探测任务面临着科学目标识别难、探测效率低的问题，同时也对能够适应复杂环境和任务的先进视觉处理算法提出了更迫切的需求。

随着人工智能技术的发展，以机器学习/深度学习为代表的算法在视觉处理上展现出了优异性能[5-9]。但这些先进的视觉处理算法具有较大的参数量，如当前常用的视觉编码器（ViT, vision Transformer）[10]，其各变体的参数量从5.7百万到632百万不等，即使简化了参数的DeiT模型[11]，其参数量依旧在百万级（5~86百万）。在深空探测领域，探测器上搭载计算机的存储空间十分有限，如火星探测车上的内存至多为256 MB，相关信息如表1所示。由此可见，当出现未见类别时，上述模型在传统探测模式下，训练数据的下传和神经网络模型的上注都将面临巨大挑战。

表1 火星探测车嵌入式计算机系统比较

探测车	芯片	内存/MB	闪存/MB
勇气号和机遇号[12]	20 MHz BAE RAD6000	128	0.25
好奇号[13]	20 MHz BAE RAD750	256	2
毅力号[14]	20 MHz BAE RAD750	256	2

为缓解上述问题，本文创新性地将零样本学习算法引入火星探测领域，不需要重新训练模型即可识别未见类别图像，有效规避了传统方法中神经网络模型反复上注的烦琐流程。作为图像探测技术的前沿热点之一，零样本学习算法[15]通过深度挖掘已见类别数据，赋予模型推理未见类别的强大能力。尤为关键的是，在测试阶段，该算法仅需上注文本知识即可实现未见类别图像的识别，极大地简化了模型的部署与应用流程，为火星探测任务的高效推进提供了一种高效、灵活的解决方案。

本文主要的研究工作如下。

1) 构建了火星探测零样本分类数据集，重新定义了以“岩石土壤”场景分类为主题的图像类别，并为其注释了对应的语义信息，为后续零样本学习研究提供了数据基础。

2) 提出了一种基于多模态特征交互的零样本场景分类算法，开创性地将零样本学习算法应用于火星场景分类任务。该算法通过特征提取与对齐集成于一体的框架，利用CNN-Transformer级联结构解决了火星图像局部-全局特征同时提取的难题。同时，通过跨模态注意力融合模块，有效实现了火星图文间的特征对齐，显著提升了零样本分类性能。

3) 提出了一种结合知识蒸馏的轻量化零样本场景分类算法，创新性地采用“轻量化模型设计+知识蒸馏”的模型压缩策略。通过上下文卷积实现图像编码器的轻量化设计，并结合知识蒸馏技术，完成教师-学生模型之间的知识迁移，在保证零样本分类性能的同时，大幅降低了模型参数量和计算复杂度。

1 相关工作

1.1 数据集调研

本文调研了国内外公开发布的火星表面图像分类数据集，其中常用的数据集有MSL[16]和MSL-v2[17]，如图1所示。

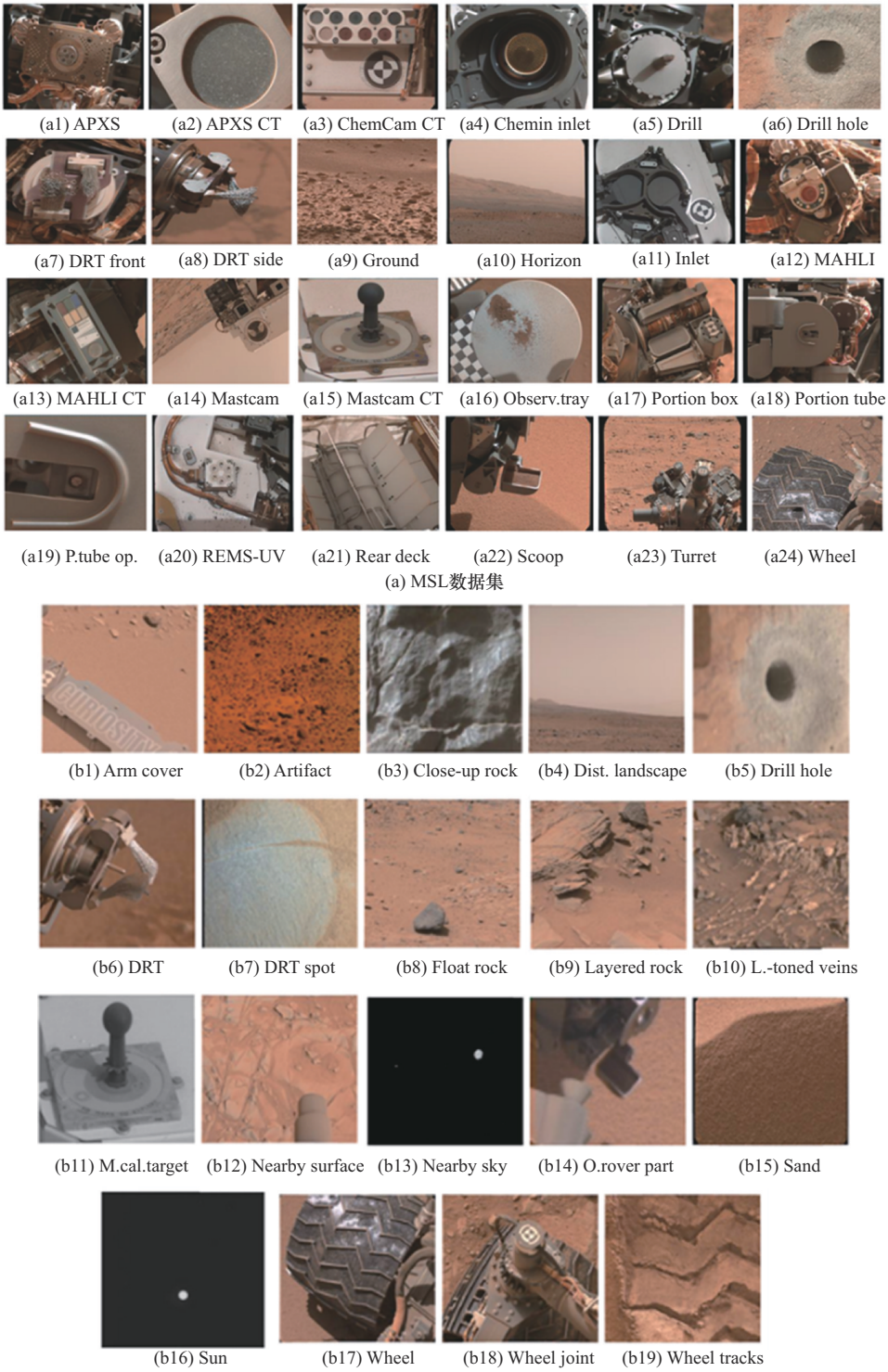


图1　火星表面图像分类数据集

这些数据集存在以下局限性。首先，数据集包含的类别间存在较大重叠，且多数类别集中在火星探测车及其部件上，缺乏多样性和代表性。其次，图像的类型、尺寸和分辨率等参数不统一，增加了数据预处理的复杂性。此外，这些数据集均为单模态数据集，仅包含图像信息和支持单模态的图像分类任务验证，无法直接用于多模态学习的零样本分类任务。

1.2　算法调研

算法调研主要围绕以下 2 个核心领域展开，一

是火星图像分类算法，二是零样本分类算法。下面将分别对这2个方面的调研内容展开详细介绍。

1) 火星图像分类算法

当前，将基于深度学习的图像处理方法用于火星图像分类领域已有较多探索。例如，Auld等[18]针对火星高分辨率成像科学实验图像研究了火星沟壑分类问题。Wagstaff等[16]使用迁移学习构建了基于AlexNet的火星图像分类网络，验证了在地球图像上训练的卷积神经网络可以成功地进行微调以适应火星图像分类任务。为提高分类器的准确性，Lu等[19]在Wagstaff等[16]的基础上，修改了卷积神经网络的微调方法，分析和分类了每个分类器在训练集和验证集中的常见错误，并采用分类器校准方法提高了分类器的可靠性。Wagstaff等[17]在此前基础上，创建了新的MSL-v2数据集训练和评估最新版本的火星图像分类器。为实现有效的监督学习多类别火星图像分类，Nandi等[20]结合迁移学习和集成方法构建了一个动态路由模块，设计了一种新颖的轮询算法，在MSL数据集中达到了88%的测试精度。Lyu等[21]提出了一种基于视觉的高精度火星地形分类方法，通过分析火星地形特征，提取专门针对地形分类的图像特征，使火星探测车地形分类精度超过了90%。Vincent等[22]为解决行星图像训练数据缺乏以及当前模型由于领域迁移的归纳偏差问题，提出了一种自监督学习框架，利用对比学习技术提升了分类器的性能。

此外，针对现有火星分类模型在火星数据不平衡和失真的情况下表现不佳的问题，Wang等[23]设计了基于半监督对比学习的火星图像分类新框架，通过表示学习开发鲁棒的视觉表示，并通过改进对比学习方法实现有监督的类间对比学习和无监督的相似性学习。这种范式有效提高了算法的性能。

2) 零样本分类算法

零样本分类旨在使深度学习模型能够识别没有训练过的新类别，通过对已有标注样本训练加以其他辅助信息（通常为语义信息）来完成新类别或新概念的识别[24]。当前零样本分类算法主要可分为3种，分别是基于嵌入模型、基于生成网络和基于预训练模型的方法。基于嵌入模型的方法研究最为广泛，也最为经典[25]，这类方法具有简单有效、计算效率高和可解释性强的优势。基于生成网络的方法可处理复杂数据，应对类别不平衡问题，但训练通常较为困难，生成样本质量不稳定。基于预训练模型的方法泛化能力强，微调效率高，但计算资源需求高，模型复杂度高，存在领域偏见。综上所述，在火星场景分类任务中引入零样本学习机理，最适合的方法为基于嵌入模型的方法。因此，本节将重点介绍这类方法的研究进展。

根据嵌入空间的不同，基于嵌入模型的方法可进一步分为基于语义空间嵌入、基于视觉空间嵌入和基于公共空间嵌入。Socher等[26]引入了跨模态迁移思想，提出了基于空间嵌入的模型，极大地促进了零样本分类算法的发展。Frome等[27]提出了深度视觉语义嵌入模型，利用已标注图像和未标注文本信息训练并识别图像，有效提升了零样本分类性能。Akata等[28]提出了属性标签嵌入模型，将每一类都嵌入属性空间中，算法性能的提升充分证明了属性的重要性。Romera-Paredes等[29]提出了一种简单高效的嵌入对齐策略，将特征、属性和类之间的关系建模为2个线性层网络，有效简化了零样本分类框架。Hou等[30]受视觉状态空间模型的启发，提出了一种参数高效的框架。该框架通过语义感知局部映射模块来集成语义嵌入，将视觉特征映射为局部语义表示，利用全局表示学习模块激励模型学习全局语义表示，通过语义融合模块实现2种语义表示的结合，增强语义特征的可辨别性。

上述基于语义空间嵌入的算法使零样本学习分类算法性能得到了较大的提高，但在距离度量过程中可能会出现多个语义向量与映射后的图像距离最近的情况，即“枢纽点”问题。为解决该问题，有学者提出了将映射子空间改为视觉空间的方法，即基于视觉空间嵌入的算法。例如，Zhang等[31]提出了一种多模态融合方法，选取卷积神经网络输出的视觉特征空间为嵌入空间，将属性和词向量等语义特征嵌入视觉空间，并采用循环神经网络实现语义空间表示的端到端学习。Sung等[32]在此基础上提出了关系网络，包含嵌入模块和关系模块两部分，嵌入模块将属性和词向量等语义特征嵌入视觉空间，然后通过关系模块来判断类别。

此外，还有学者提出了将图像和类别标签同时映射到一个公共空间，即基于公共空间嵌入的算法。例如，Reed等[33]提出了结构联合嵌入模型，通过端到端的训练，在公共空间完成对图像视觉特

征和语义信息的匹配辅助分类。为增强图像、语义与分类的关联性，Tao 等[34]提出了语义保留局部嵌入的算法，将图像和语义数据同时嵌入派生空间进行分类，通过跨域匹配来执行子空间学习以提高算法性能。为提高零样本细粒度分类性能，Chen 等[35]设计了一种基于对比学习的端到端零样本分类方法，通过自监督多模态学习集成来自预训练语言模型的潜在语义知识，采用属性级对比学习策略，进一步增强模型对细粒度视觉特征的辨别能力。

2　算法设计

为实现零样本分类算法向深空探测领域的迁移，本节研究主要分为数据集构建和算法研究 2 个方面。在数据集构建方面，以火星探测为应用背景，构建用于零样本分类的火星探测数据集。在算法研究方面，进行了基于多模态特征交互的零样本场景分类算法及其轻量化设计研究。

2.1　数据集构建

首先，本文整理分析了火星探测数据集 MSL[16]和 MSL-v2[17]的相关信息，如 1.1 节所述。通过去掉特征增强的重复场景和灰度图像，筛选适于“岩石土壤”场景主题的类别图像。在数据集类别定义方面，确定了如图 2 所示的 10 个类别，其类别名称和对应的语义信息如表 2 所示。

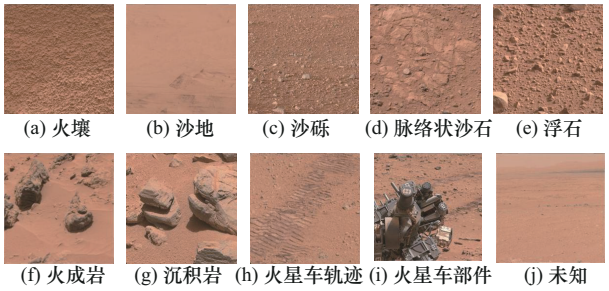


图 2　本文构建的零样本火星场景分类数据集 ZSMars

2.2　算法研究

在零样本分类算法向火星探测领域迁移的过程中，首先研究了基于多模态特征交互的零样本场景分类算法。考虑到火星探测航天器存储空间和计算资源受限的实际情况，进一步研究了结合知识蒸馏的轻量化零样本场景分类算法。

1) 基于多模态特征交互网络（MFINet, multi-modal feature interaction network）的零样本场景分类算法

如图 3 所示，零样本图像场景分类算法 MFINet 主要由 3 个部分构成，分别为用于提取视觉特征的局部-全局特征提取（Res-DeiT, ResNet-data efficient image Transformer）模块、用于解释语义的语义信息提取（Core-BERT, core-bidirectional encoder representation from Transformer）模块和进行多模态交互的跨模态特征融合（CMFF, cross modal feature fusion）模块。

表 2　零样本火星场景分类数据集 ZSMars 中的类别名称和对应的语义信息

类别索引	类别名称	语义信息
0	火壤	火星上未固结或固结较差的风化物质，颜色偏红，受地形的影响不会形成脊，流动性相对较低
1	沙地	火星上细小岩石构成的颗粒物质，一般由岩石经风化和剥蚀而形成，通常由迎风坡和背风坡构成一个脊，流动性相对较强
2	沙砾	火星上的颗粒物质，一般由岩石经风化和剥蚀而形成，比沙地粗糙并且比岩石颗粒小，流动性一般
3	脉络状沙石	火星上由成片的砂砾和岩石构成的一种地貌，具备深浅不一的脉络状/延展性条纹，通常以岩石居多，砂砾覆盖其上
4	浮石	火星表面从大岩石中脱落下来的小型岩石，呈块状或椭圆形，质地坚固且硬脆，比沙砾的岩石颗粒大，基本没有砂砾
5	火成岩	火星上一种由岩浆喷出地表或侵入地壳冷却凝固所形成的岩石，有明显的矿物晶体颗粒或气孔，岩石颜色偏黑，形状不规则；岩石体积大小不一，不具备流动性
6	沉积岩	火星上一种由成层堆积的松散沉积物固结而成的岩石，通常具有一定的层状纹理；不含或含有少量的沙土；岩石体积较大，不具备流动性；沉积物一般指松散碎屑物，如砾石、砂、粘土、灰泥等
7	车辙	火星表面探测车车轮的花纹在松软的火壤或者沙地经过后留下的痕迹，近镜头图片包含明显的车轮花纹，远镜头图片车轮花纹形状模糊，显示出重物拖拉的长条痕迹
8	火星探测车部件	火星探测车或火星探测车上存在的一部分结构
9	未知	火星上的远镜头图片，以松软的土壤为主体，也包括部分天空，或者山丘，但难以分辨地面中具体的成分组成，区域之间有明显的明暗分层

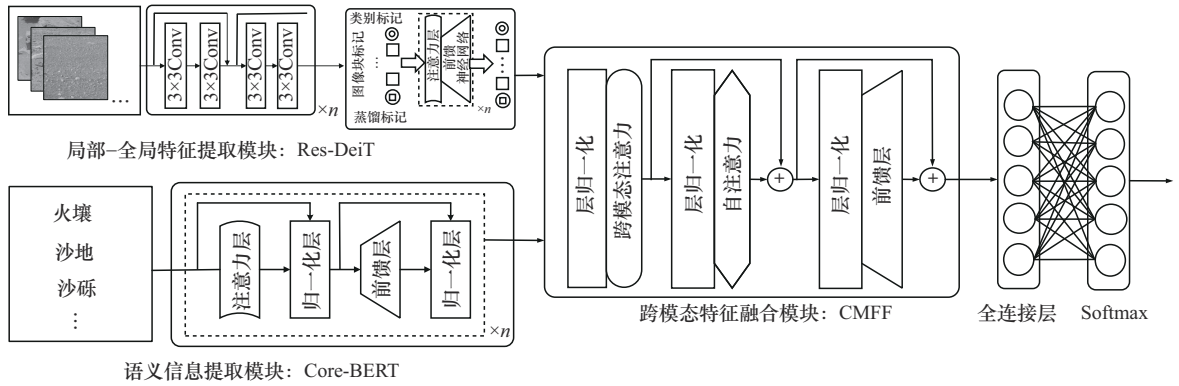


图3　基于多模态特征交互的零样本场景分类算法MFINet

与多数现有的零样本嵌入模型相比，MFINet的不同之处在于它将视觉特征提取网络、自然语言处理网络和视觉语义对齐网络集成于端到端的训练过程中。这使MFINet框架能够在视觉语义对齐网络的指导下提取更多语义上有意义的视觉特征和更多视觉信息丰富的语义信息。因此算法的输入有视觉和语义2种模态，视觉输入为图像，如图2所示；语义输入为文字，如表2所示。具体而言，在Res-DeiT模块中，采用级联卷积神经网络和Transformer结构获取图像的局部–全局联合视觉特征。在语义提取方面，构建了Core-BERT网络作为轻量化语义编码器，它继承了原始BERT网络的核心架构以实现语义信息提取。在CMFF模块中，跨模态注意力层可以促进视觉模态和语义模态之间的特征交互，有效地缓解了2种模态间的差距。

2) 结合知识蒸馏的轻量化零样本火星场景分类（KDMSC，knowledge distillation-based lightweight zero-shot Mars scene classification）算法

本文所提KDMSC算法如图4所示。该算法引入了知识蒸馏损失函数，由复杂教师模型和轻量化学生模型组成，利用教师模型最佳的预测结果来有效指导学生模型的优化过程。

教师模型沿用了MFINet算法。学生模型的语义编码器与教师模型一致，只是在图像特征提取网

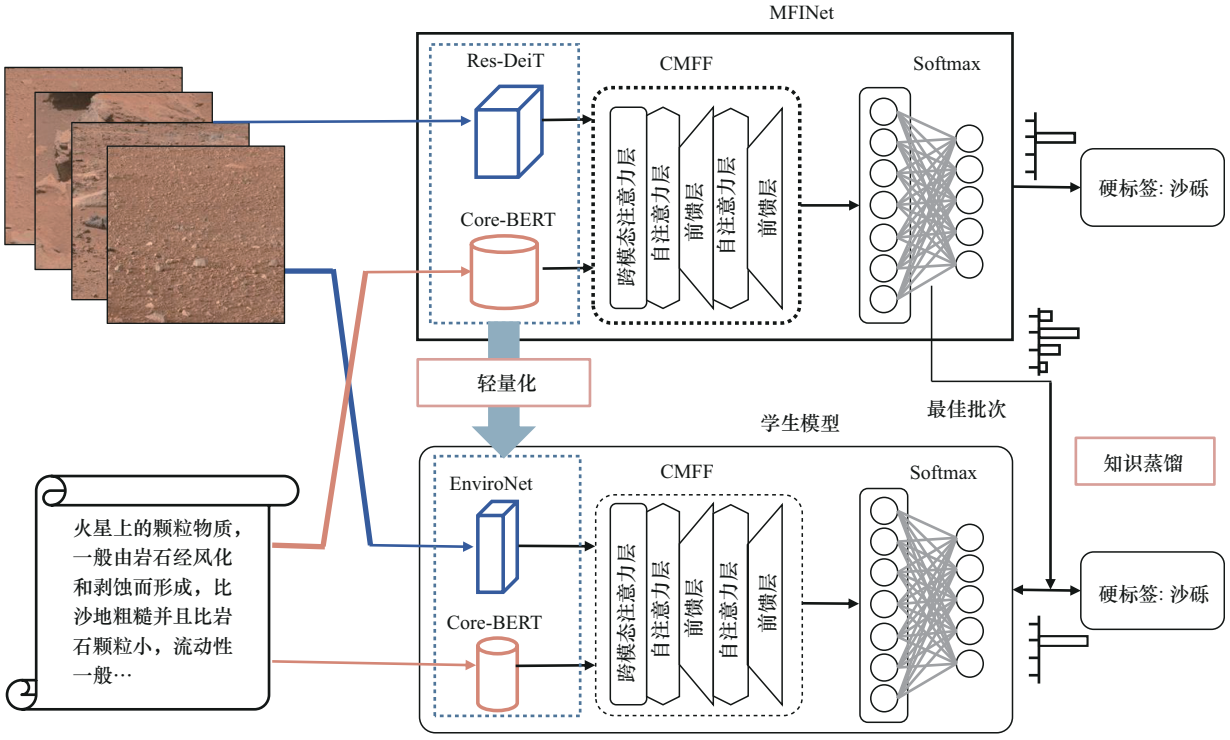


图4　结合知识蒸馏的轻量化零样本火星场景分类算法

络方面，学生模型通过引入上下文卷积构建了EnviroNet网络，可有效捕获图像的局部特征和全局特征，其详细结构如表3所示。该网络中采用的上下文卷积如图5所示。随后，同样利用CMFF模块将2个模态的特征投影到同一潜在空间，实现视觉特征和语义特征间的匹配。

表3　　EnviroNet的详细结构

阶段	EnviroNet	输出
res1	7×7 Conv, 64	112×112
	max pool	
res2	[1 × 1Conv,64 Cotlayer 1 × 1Conv,256]	56×56
res3	[1 × 1Conv,128 Cotlayer 1 × 1Conv,512]	28×28
res4	[1 × 1Conv,256 Cotlayer 1 × 1Conv,1 024] × 2	14×14
res5	[1 × 1Conv,512 Cotlayer 1 × 1Conv,2 048]	7×7
—	1 000 FC, Softmax	1×1

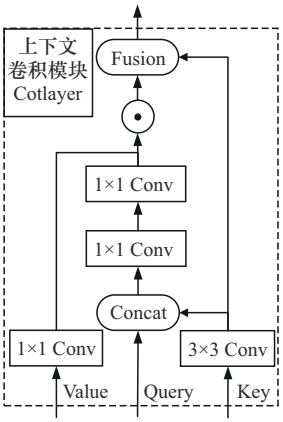


图5　轻量化图像特征提取网络EnviroNet中上下文卷积示意

3) 损失函数的构建

MFINet算法的损失函数L_{MF}由类回归损失L_{c}和交叉熵损失L_{r}两部分构成，计算式为

$$L_{\mathrm{MF}} = L_{\mathrm{c}} + \lambda L_{\mathrm{r}} \tag{1}$$

其中，λ为交叉熵损失L_{r}的权重。类回归损失L_{c}和交叉熵损失L_{r}的计算式分别为

$$L_{\mathrm{c}} = \left\| \boldsymbol{X} - \boldsymbol{y} \right\|_2^2 \tag{2}$$

$$L_{\mathrm{r}} = -\log\left(\frac{\mathrm{e}^{\boldsymbol{X}\cdot\boldsymbol{R}}}{\sum\limits_{\hat{c}\in l^s} \mathrm{e}^{\boldsymbol{X}\cdot\boldsymbol{R}^{\hat{c}}}} \right) \tag{3}$$

其中，$\boldsymbol{X}$和$\boldsymbol{R}$分别为由CMFF模块得到图像和文本特征，$\boldsymbol{y}$为图像对应的文本向量，$\hat{c}$为类别对应的文本信息，l^s为图像类别对应的未见类别文本描述信息。

KDMSC算法中教师模型和学生模型均为MFINet算法的损失函数。此外，KDMSC算法的总体损失函数额外添加了蒸馏损失函数，计算式为

$$L_{\mathrm{KD}} = L_{\mathrm{stu}} + \lambda_{\mathrm{kd}} L_{\mathrm{kd}} \tag{4}$$

其中，L_{stu}与L_{MF}函数一致，λ_{kd}为蒸馏损失L_{kd}的权重。L_{kd}的计算式为

$$L_{\mathrm{kd}} = \mathrm{KL}\left(F_{\mathrm{t}}\left(Z\left(\boldsymbol{X}_1, \boldsymbol{Y} \right) \right), F_{\mathrm{s}}\left(Z\left(\boldsymbol{X}_1, \boldsymbol{Y} \right) \right) \right) \tag{5}$$

其中，KL表示Kullback-Leibler散度函数，$F_{\mathrm{t}}(\cdot)$和$F_{\mathrm{s}}(\cdot)$分别表示教师模型和学生模型的输出结果，$Z\left(\boldsymbol{X}_1, \boldsymbol{Y}\right)$表示通过CMFF模块得到的特征，$\boldsymbol{X}_1$和$\boldsymbol{Y}$分别表示输入CMFF模块的图像向量和文本向量。

3　实验分析

为充分证明本文算法的有效性，利用词向量和语义向量2种语义信息，在3种可见/不可见比率下进行了对比实验，并对算法进行了复杂度分析，同时对实验结果展开了可视化分析。

3.1　实验设置

1) 设置细节

对零样本火星场景分类数据集的类别进行了3种可见/未见类别的划分，如表4所示。在6/4可见/不可见比率中，可见类别包含火壤、沙砾、浮石、火星探测车轨迹、火星探测车部件和未知，未见类别包含沙地、脉络状沙石、火成岩和沉积岩。在7/3可见/不可见比率中，可见类别包含火壤、沙砾、浮石、火星探测车轨迹、火星探测车部件、火成岩和未知，未见类别包含沙地、脉络状沙石和沉积岩。在8/2可见/不可见比率中，可见类别包含火壤、沙砾、浮石、火星探测车轨迹、火星探测车部件、火成岩、脉络状沙石和未知，未见类别包含沙地和沉积岩。需要说明的是，表4中可见/不可见比率后括号中内容为可见/未见类别图像数量比率，类别后括号中内容为该类别图像数量。

表4 数据集可见/未见类别的划分

类别	可见/不可见比率为6/4（2 869/1 978）	可见/不可见比率为7/3（3 059/1 788）	可见/不可见比率为8/2（3 491/1 356）
可见类	火壤（69）、沙砾（1 253）、浮石（715）、火星探测车轨迹（108）、火星探测车部件（154）和未知（570）	火壤（69）、沙砾（1 253）、浮石（715）、火星探测车轨迹（108）、火星探测车部件（154）、火成岩（190）和未知（570）	火壤（69）、沙砾（1 253）、浮石（715）、火星探测车轨迹（108）、火星探测车部件（154）、火成岩（190）、脉络状沙石（432）和未知（570）
未见类	沙地（196）、脉络状沙石（432）、火成岩（190）和沉积岩（1 160）	沙地（196）、脉络状沙石（432）和沉积岩（1 160）	沙地（196）和沉积岩（1 160）

2) 评价指标

为充分说明本文算法的可行性和有效性，采用2种评价指标作为衡量标准，分别是总体精度（OA, overall accuracy）和平均精度（AA, average accuracy）计算式分别为

$$\mathrm{OA}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}+\mathrm{TN}} \tag{6}$$

$$\mathrm{AA}=\frac{\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}+\dfrac{\mathrm{TN}}{\mathrm{FP}+\mathrm{TN}}}{2} \tag{7}$$

其中，TP表示分类器预测结果为正样本、实际也为正样本，即正样本被正确识别的数量；FP表示分类器预测结果为正样本、实际为负样本，即误报的负样本数量；TN表示分类器预测结果为负样本、实际也为负样本，即负样本被正确识别的数量；FN表示分类器预测结果为负样本、实际为正样本，即漏报的正样本数量。

3.2 对比实验

为评估本文算法的性能，本节选取了DeViSE[27]、ALE[28]、ESZSL[29]、SJE[33]和DUET[35]算法来进行对比实验。所有算法的OA和AA结果分别如表5和表6所示，其中，粗体为实验的最佳结果，下划线为次佳结果。

表5和表6的实验结果表明，KDMSC算法在不同语义向量（词向量和句子向量）以及不同可见/不可见比率下的分类性能都表现出了显著优势。具体来看，当输入为词向量时，KDMSC算法在6/4可见/不可见比率下实现了最佳的OA，达到了46.8%，而在7/3和8/2可见/不可见比率下表现次佳，分别为62.7%和97.4%。这表明其在较高可见类别比例下能够更好地平衡可见类别和未见类别的分类性能。相比之下，MFINet在7/3和8/2可见/不可见比率下实现了最佳OA，分别达到了64.8%和99.1%，说明其在处理更高比例的未见类别时具有一定优势。

表5 对比实验的OA结果

算法	词向量可见/不可见比率			句子向量可见/不可见比率		
	6/4	7/3	8/2	6/4	7/3	8/2
DeViSE	36.6%	47.5%	65.3%	17.3%	34.0%	71.7%
ALE	37.9%	46.1%	60.6%	17.8%	32.7%	67.9%
ESZSL	34.0%	46.8%	71.6%	21.9%	27.2%	50.3%
SJE	34.8%	45.1%	56.9%	18.0%	38.4%	69.7%
DUET	37.9%	61.9%	82.9%	32.4%	43.2%	57.9%
MFINet	45.1%	**64.8%**	**99.1%**	39.1%	54.9%	73.6%
学生模型	44.8%	59.1%	95.9%	37.0%	50.6%	78.9%
KDMSC	**46.8%**	62.7%	97.4%	**40.2%**	**62.8%**	**80.1%**

表6 对比实验的AA结果

算法	词向量可见/不可见比率			句子向量可见/不可见比率		
	6/4	7/3	8/2	6/4	7/3	8/2
DeViSE	34.8%	54.5%	72.7%	38.1%	43.9%	77.1%
ALE	37.1%	51.1%	70.8%	38.6%	43.9%	74.9%
ESZSL	30.7%	42.7%	61.9%	33.9%	41.8%	67.8%
SJE	36.3%	50.2%	66.2%	38.5%	45.4%	75.7%
DUET	38.5%	61.9%	94.8%	30.9%	41.7%	58.9%
MFINet	41.1%	**64.7%**	**96.9%**	34.2%	44.9%	67.9%
学生模型	44.8%	63.7%	96.5%	33.2%	48.2%	66.8%
KDMSC	**45.3%**	63.7%	96.7%	**39.6%**	**54.9%**	**77.9%**

然而，当输入为句子向量时，KDMSC算法在3种可见/不可见比率下均取得了最佳的OA和AA，显示出其在零样本分类方面的强大能力，能够更好地利用句子向量的语义丰富性来提升分类性能。进

一步对比发现，仅进行轻量化设计的学生模型在所有情况下均出现了精度下降，KDMSC 通过结合知识蒸馏技术，不仅有效遏制了模型压缩带来的精度下降趋势，还在输入句子向量时进一步提升了算法性能，这表明知识蒸馏在零样本分类任务下能够更好地保留和传递关键信息。此外，OA 结果也显示，KDMSC 在输入句子向量时依然保持了最佳的分类性能，而 MFINet 在 6/4 和 7/3 可见/不可见比率下表现次佳。值得注意的是，ALE 算法在 6/4 可见/不可见比率下实现了次佳的 AA，DeViSE 算法在 8/2 可见/不可见比率下实现了次佳的 AA，这表明这些算法在特定条件下具有一定的优势，但总体性能仍不如 KDMSC 和 MFINet 算法。

综合来看，KDMSC 算法在不同语义向量和可见/不可见比率下均表现出优越的性能，通过知识蒸馏技术整合教师模型和学生模型的优势，不仅有效避免了轻量化带来的精度损失，还进一步提升了算法性能，显示出其在火星探测领域零样本分类任务中的实用性和优越性，能够为火星探测任务提供更高效、更准确的分类支持。

3.3　复杂度分析

在词向量和句子向量 2 种语义信息输入的情况下，MFINet 和 KDMSC 的参数量及计算复杂度信息如表 7 所示。

从表 7 的实验结果可以看出，KDMSC 在不牺牲算法性能的前提下，显著降低了模型的参数量和计算复杂度。具体而言，KDMSC 的模型大小和参数量不到 MFINet 的 $\frac{1}{2}$，其计算复杂度更是降低至 MFINet 的 $\frac{1}{10}$ 以内。

进一步分析发现，输入语义向量的类型（词向量和句子向量）对模型的计算复杂度没有明显影响，这表明模型压缩方案在不同输入条件下均能保持稳定的计算效率。然而，在模型大小和参数量方面，输入句子向量时模型的参数量相对略高，这可能是因为句子向量蕴含更丰富的语义信息，需要更多的参数来有效捕捉和融合这些信息，从而实现更优的分类性能。相比之下，输入词向量时模型的参数量相对较少，但依然能够达到较高的分类精度，这说明模型压缩方案在不同输入条件下均能有效平衡语义信息的复杂性和模型的轻量化需求。

综上所述，KDMSC 不仅在整体上降低了模型的参数量和计算复杂度，还在不同输入条件下表现出良好的适应性和稳定性。这一结果表明，该方案能够在保证算法性能的前提下，有效优化模型的资源占用，使其更适合在火星探测等资源受限的场景中应用。

3.4　可视化结果分析

本节将从图像特征学习和最终分类结果 2 个方面对所提算法的零样本分类效果进行可视化展示及分析。在图像特征学习方面，采用 t-SNE 算法对图像流的特征进行可视化对比分析。在最终分类结果方面，绘制分类结果混淆矩阵的热力图对其进行研究分析。

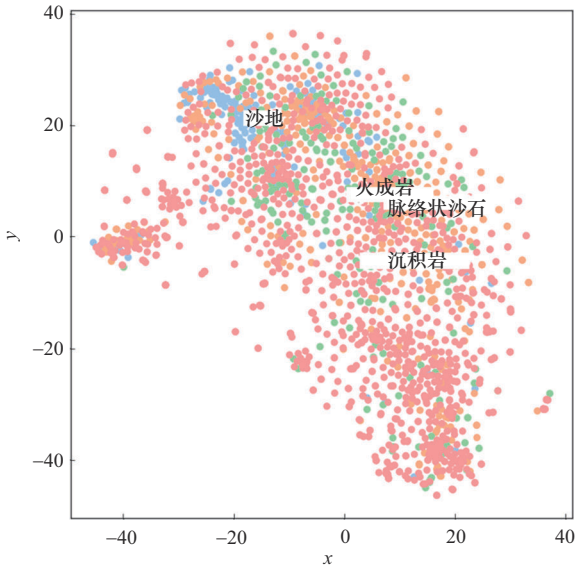
1) 图像特征学习

为直观展示本文算法在图像特征学习方面的效果，选取了在 6/4 可见/不可见比率下输入的 4 种未见类别图像，并将其在 KDMSC 算法不同阶段学习到的图像特征进行可视化，结果如图 6 所示。其中，图 6(a)~图 6(c) 分别为输入的原始图像特征、EnviroNet 提取的图像特征和 CMFF 提取的图像特征的可视化结果。

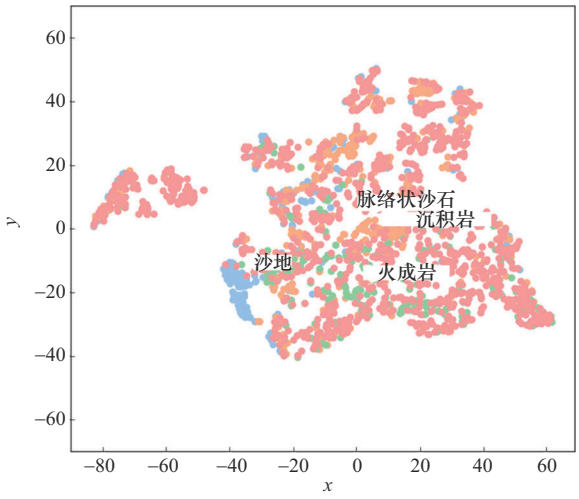
对比图 6 可以看到，原始图像特征各类别分布比较混杂，尤其是火成岩和脉络状沙石这两类。经 EnviroNet 提取之后的图像特征类内聚集度明显提高很多，但各个类别间的簇间边界依旧不明显。结合了语义辅助的 CMFF 模块所提取的图像特征不仅类内聚集度更高，且簇间边界明显，簇间得到了有效分离。这不仅有效证明了 EnviroNet 和 CMFF 模块设置的有效性，同时充分说明了本文算法能够理解类别间的语义关系，而不仅是基于外观特征的分类。

表 7　参数量及计算复杂度信息

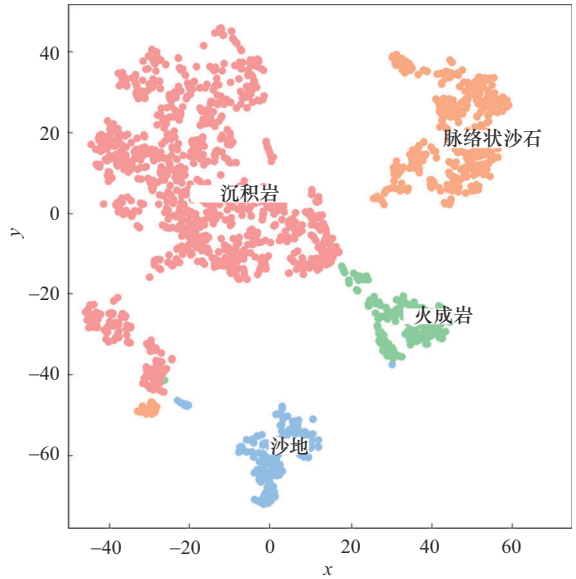
模型	词向量			句子向量		
	计算复杂度/GFLOPS	模型大小/MB	参数量	计算复杂度/GFLOPS	模型大小/MB	参数量
MFINet	34.33	564.04	147.07×10^6	34.34	562.42	147.43×10^6
KDMSC	3.33	269.85	70.74×10^6	3.33	271.23	71.10×10^6



(a) 原始图像特征可视化结果



(b) EnviroNet提取的图像特征可视化结果

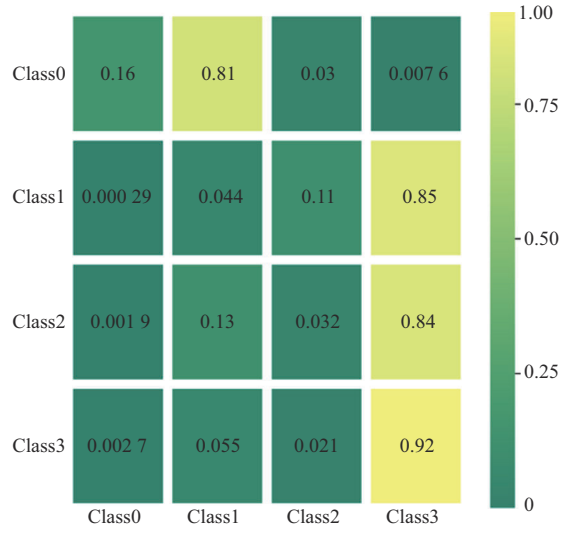


(c) CMFF提取的图像特征可视化结果

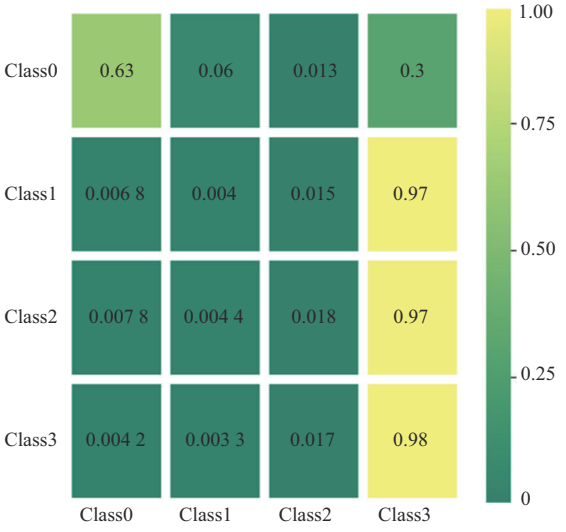
图 6　各阶段图像特征的可视化结果

2) 最终分类结果

为充分展示本文算法在零样本分类任务中的性能，选取了在 6/4 可见/不可见比率下输入词向量时，MFINet 和 KDMSC 算法对 4 种未见类别的分类结果混淆矩阵，对其进行热力图展示，如图 7 所示。其中，图 7(a)和图 7(b)分别为 MFINet 和 KDMSC 算法的分类结果混淆矩阵热力图。图 7 中的 Class0~Class3 对应的图像类别名称如表 8 所示。



(a) MFINet分类结果的混淆矩阵热力图



(b) KDMSC分类结果的混淆矩阵热力图

图 7　分类结果的混淆矩阵热力图

由图 7 可知，MFINet 算法仅能正确识别沉积岩，对沙地是次高的预测；KDMSC 算法则实现了沙地和沉积岩的有效识别。可见 KDMSC 算法可有效整合教师模型和学生模型的优势信息，进一步提高本文算法的零样本分类性能。

表 8　　图 7 中类别对应的名称

类别	名称
Class0	沙地
Class1	脉络状沙石
Class2	火成岩
Class3	沉积岩

4　结束语

为高效应对火星复杂未知场景下的图像感知挑战，本文在火星探测领域率先开展了一系列研究。首先，构建了火星探测领域的零样本分类数据集，为后续研究奠定了坚实基础。其次，提出了一种基于多模态特征交互的零样本场景分类算法，通过对齐图像与文本模态信息，显著提升了算法对未见类别的识别能力。进一步地，结合知识蒸馏和轻量化模型设计技术，对算法进行了模型压缩优化，在不损失零样本分类性能的前提下，大幅降低了模型的参数量和计算复杂度，使其更契合火星探测设备资源受限的实际场景。为验证本文算法的有效性，开展了大量对比实验，并对结果进行了深入的可视化分析。实验结果表明，本文算法在零样本分类任务中展现出显著的性能优势，为火星探测领域的图像感知技术发展提供了新的思路和方法。

参考文献：

[1] 吴伟仁, 于登云. 深空探测发展与未来关键技术[J]. 深空探测学报, 2014, 1(1): 5-17.
WU W R, YU D Y. Development of deep space exploration and its future key technologies[J]. Journal of Deep Space Exploration, 2014, 1(1): 5-17.

[2] 叶培建, 孟林智, 马继楠, 等. 深空探测人工智能技术应用及发展建议[J]. 深空探测学报, 2019, 6(4): 303-316, 383.
YE P J, MENG L Z, MA J N, et al. Suggestions on artificial intelligence technology application and development in deep space exploration[J]. Journal of Deep Space Exploration, 2019, 6(4): 303-316, 383.

[3] SIMON J I, HICKMAN-LEWIS K, COHEN B A, et al. Samples collected from the floor of jezero crater with the Mars 2020 perseverance rover[J]. Journal of Geophysical Research: Planets, 2023, 128(6): 1-42.

[4] HERD C D K, BOSAK T, HAUSRATH E M, et al. Sampling Mars: geologic context and preliminary characterization of samples collected by the NASA Mars 2020 perseverance rover mission[J]. Proceedings of the National Academy of Sciences of the United States of America, 2025, 122(2): 1-12.

[5] 宋亚飞, 李乐民, 权文, 等. 时序数据图像化: 战术意图识别及可移植框架[J]. 通信学报, 2024, 45(8): 149-165.
SONG Y F, LI L M, QUAN W, et al. Timing data visualization: tactical intent recognition and portable framework[J]. Journal on Communications, 2024, 45(8): 149-165.

[6] 高红民, 曹雪莹, 陈忠昊, 等. 基于多尺度近端特征拼接网络的高光谱图像分类方法[J]. 通信学报, 2021, 42(2): 92-102.
GAO H M, CAO X Y, CHEN Z H, et al. Hyperspectral image classification method based on multi-scale proximal feature concatenate network[J]. Journal on Communications, 2021, 42(2): 92-102.

[7] 廖育荣, 王海宁, 林存宝, 等. 基于深度学习的光学遥感图像目标检测研究进展[J]. 通信学报, 2022, 43(5): 190-203.
LIAO Y R, WANG H N, LIN C B, et al. Research progress of deep learning-based object detection of optical remote sensing image[J]. Journal on Communications, 2022, 43(5): 190-203.

[8] 王万良, 李卓蓉. 生成式对抗网络研究进展[J]. 通信学报, 2018, 39(2): 135-148.
WANG W L, LI Z R. Advances in generative adversarial network[J]. Journal on Communications, 2018, 39(2): 135-148.

[9] 吴健, 盛胜利, 赵朋朋, 等. 最小差异采样的主动学习图像分类方法[J]. 通信学报, 2014, 35(1): 107-114.
WU J, SHENG S L, ZHAO P P, et al. Minimal difference sampling for active learning image classification[J]. Journal on Communications, 2014, 35(1): 107-114.

[10] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J]. arXiv Preprint, arXiv: 2010.11929, 2020.

[11] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[C]//Proceedings of the 38th International Conference on Machine Learning. New York: PMLR, 2021: 10347-10357.

[12] BAJRACHARYA M, MAIMONE M W, HELMICK D. Autonomy for Mars rovers: past, present, and future[J]. Computer, 2008, 41(12): 44-50.

[13] Mars science laboratory: mission: rover: brains. Curiosity rover-NASA science[R]. 2009.

[14] BAE systems computers to manage data processing and command for upcoming satellite missions[R]. 2008.

[15] CAO W P, WU Y H, SUN Y X, et al. A review on multimodal zero-shot learning[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2023, 13(2): e1488.

[16] WAGSTAFF K, LU Y, STANBOLI A, et al. Deep mars: CNN classification of mars imagery for the PDS imaging atlas[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 7867-7872.

[17] WAGSTAFF K, LU S, DUNKEL E, et al. Mars image content classification: three years of NASA deployment and recent advances[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35

(17): 15204-15213.

[18] AULD K S, DIXON J C. A classification of Martian gullies from HiRISE imagery[J]. Planetary and Space Science, 2016, 131: 88-101.

[19] LU S, WAGSTAFF K, CAI J, et al. Content-based classification of Mars imagery for the PDS image atlas[J]. AGU Fall Meeting Abstracts, 2019, 2019: 1-6.

[20] NANDI A, MALLICK A, DE A, et al. Mars-TRP: classification of Mars imagery using dynamic polling between transferred features[J]. Engineering Applications of Artificial Intelligence, 2022, 114: 105014.

[21] LYU F T, LI N, LIU C K, et al. Highly accurate visual method of Mars terrain classification for rovers based on novel image features[J]. Entropy, 2022, 24(9): 1304.

[22] VINCENT G M, WARD I R, MOORE C, et al. CLOVER: contrastive learning for onboard vision-enabled robotics[J]. Journal of Spacecraft and Rockets, 2023, 61(3): 728-740.

[23] WANG W J, LIN L L, FAN Z J, et al. Semi-supervised learning for Mars imagery classification[C]//Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2021: 499-503.

[24] 贾霄, 郭顺心, 赵红. 基于图像属性的零样本分类方法综述[J]. 南京大学学报(自然科学), 2021, 57(4): 531-543.
JIA X, GUO S X, ZHAO H. A review of zero-shot learning classification methods based on image attributes[J]. Journal of Nanjing University (Natural Science), 2021, 57(4): 531-543.

[25] TAN X M, XI B B, LI J J, et al. Review of zero-shot remote sensing image scene classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 11274-11289.

[26] SOCHER R, GANJOO M, SRIDHAR H, et al. Zero-shot learning through cross-modal transfer[J]. arXiv Preprint, arXiv: 1301.3666, 2013.

[27] FROME A, CORRADO G S, SHLENS J, et al. DeViSE: a deep visual-semantic embedding model[J]. Advances in Neural Information Processing Systems, 2013: 26.

[28] AKATA Z, PERRONNIN F, HARCHAOUI Z, et al. Label-embedding for attribute-based classification[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2013: 819-826.

[29] ROMERA-PAREDES B, TORR P H S. An embarrassingly simple approach to zero-shot learning[C]//Proceedings of the 32nd International Conference on Machine Learning. Berlin: Springer, 2017: 11-30.

[30] HOU W J, FU D J, LI K, et al. ZeroMamba: exploring visual state space model for zero-shot learning[J]. arXiv Preprint, arXiv: 2408.14868, 2024.

[31] ZHANG L, XIANG T, GONG S G. Learning a deep embedding model for zero-shot learning[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 3010-3019.

[32] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: relation network for few-shot learning[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 1199-1208.

[33] REED S, AKATA Z, LEE H, et al. Learning deep representations of fine-grained visual descriptions[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 49-58.

[34] TAO S Y, YEH Y R, WANG Y F. Semantics-preserving locality embedding for zero-shot learning[C]//Proceedings of the British Machine Vision Conference 2017. Piscataway: IEEE Press, 2017: 3.

[35] CHEN Z, HUANG Y F, CHEN J Y, et al. DUET: cross-modal semantic grounding for contrastive zero-shot learning[C]//Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 405-413.

[作者简介]

檀晓萌（1997– ），女，河北石家庄人，中国科学院大学博士生，主要研究方向为深度学习、零样本学习等。

席博博（1994– ），男，河南洛阳人，博士，西安电子科技大学讲师，主要研究方向为深度网络、高光谱遥感图像处理、基于模型的系统工程、数字孪生等。

薛长斌（1972– ），男，辽宁锦州人，中国科学院国家空间科学中心研究员、博士生导师，主要研究方向为深度学习、深空探测科学载荷全流程设计与仿真技术等。

李云松（1973– ），男，辽宁葫芦岛人，博士，西安电子科技大学教授、博士生导师，主要研究方向为深度学习图像/视频处理编码及传输、芯片设计和高性能计算等。