

持续自我监督学习中的掩码自编码器在遥感中的应用

Lars Möllenbrok *Student Member, IEEE*, Behnood Rasti *Senior Member, IEEE*,
Begüm Demir, *Senior Member, IEEE*

摘要—连续学习 (CL) 方法的发展旨在从持续获取的训练数据中顺序地学习新任务，这在遥感 (RS) 领域引起了极大的关注。现有的 RS 中的 CL 方法，在学习新任务的同时，增强了对灾难性遗忘的鲁棒性。这是通过使用大量的标记训练样本实现的，但在 RS 中收集这些样本成本高昂且并不总是可行。为了解决这个问题，我们提出了一种基于掩码自编码器 (记作 CoSMAE) 的新颖连续自我监督学习方法。提出的 CoSMAE 由两个组件组成：i) 数据插值；ii) 模型插值知识蒸馏。数据插值与通过当前任务图像和先前任务图像的插值得到的信息保留有关。模型插值知识蒸馏则涉及同时从过去模型和当前模型中提取知识，通过插值它们的模型权重形成一个教师来进行知识蒸馏。这两个组件互为补充，在数据和模型层面对 MAE 进行正则化，以促进更好的跨任务泛化并减少灾难性遗忘的风险。实验结果显示，CoSMAE 在应用于 MAE 时比最先进的 CL 方法取得了高达 4.94% 的显著改进。我们的代码公开获取地址是：<https://git.tu-berlin.de/rsim/CoSMAE>。

Index Terms—持续学习，自监督学习，掩码自动编码器，深度学习，mixup，遥感。

I. 介绍

他卫星的重复通过操作和不断增长的地观测任务导致了新的遥感 (RS) 图像的持续收集以及新训练集的频繁发布。深度神经网络 (DNNs) 在从 RS 图像学习方面展现了显著性能 [1]。然而，每次有新数据可用时都从头开始重新训练 DNNs 会带来巨大的计算挑战，如高资源消耗和增加的训练时间。另一方面，当前模型难以顺序地从数据中学习。仅使用新数据更新模型可能会导致之前 (即旧) 数据的知识丧失。这种现象被称为灾难性遗忘 [2]，其发生在顺序训练过程中因

为数据可能非独立同分布。在 RS 领域，这些分布变化可能是由于不同的获取条件、光谱响应的变化或土地覆盖的改变而引起的。为了解决这些问题，可以使用持续学习 (CL)，旨在从顺序提供的数据 (即任务) 中进行学习。

几种连续学习方法在推荐系统中被开发出来以解决灾难性遗忘的问题。例如，在 [3] 中，提出了一种结合分类器和变分自编码器的方法。该自编码器学会从旧任务生成特征，并用于分类器的训练以保留关于旧数据分布的信息。在 [4] 中，使用了冻结的骨干网络与基于分析学习的线性分类器相结合的方法。基于分析学习的分类器能够递归地学习新数据同时记住旧数据。[5] 中提出的方法动态扩展了一个任务共用特征提取器。对于每个新任务，都会添加一个特定的任务分类器，并附带一个轻量级的特征转移模块，该模块学会将特征重新对齐到之前的分布上。上述方法是监督分类问题设计的，因此需要每项连续学习任务 [3]–[5] 都有标记训练图像可用，这既耗时又昂贵。相应地，在实际的连续学习应用中它们并不实用。解决这一问题的有效方法是利用自监督学习 (SSL)，它旨在无需标记训练数据即可学习通用图像特征。

根据我们的知识，在 RS [6] 中只有一种连续的 SSL (CSSL) 方法，该方法结合了对比学习方法与基于弹性权重固化 (EWC) [7] 的正则化。EWC 在学习任务后估计模型权重的重要性，并在学习新任务时阻止对重要权重进行更改。然而，随着正则项的累积，这种方法变得难以处理且无法扩展到大量任务。此外，EWC 仅依赖于旧模型的权重来保留旧任务的知识。然而，这些旧权重可能并不总是能够最优地捕捉旧任务的知识，导致性能不佳。

CSSL 方法的发展在计算机视觉社区中得到了广泛

L. Möllenbrok, B. Rasti and B. Demir are with the Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, 10623 Berlin, Germany, also with the BIFOLD - Berlin Institute for the Foundations of Learning and Data, 10623 Berlin, Germany. Email: lars.moellenbrok@tu-berlin.de, behnood.rasti@tu-berlin.de, demir@tu-berlin.de.

的扩展。作为一个示例，在 [8] 中，自监督学习与无监督数据混合策略相结合，通过插值旧任务和新任务的图像来减轻灾难性遗忘。由于该方法仅依赖于数据层面的正则化，因此性能提升有限。在 [9] 中，一个时间投影器将通过自监督学习获得的表示映射回其之前的状态。然后使用旧模型的知识蒸馏对投影后的表示进行正则化。虽然这有助于保留先前任务的知识，但旧模型的权重对于处理新数据来说是次优的，导致性能下降。

为了解决上述问题，本文提出了一种新的 CSSL 方法，该方法基于掩码自编码器 (MAEs) [10]，我们将其表示为 CoSMAE。CoSMAE 的主要贡献在于其两个关键组件：1) 在 RS 中引入的数据混合；2) 模型混合知识蒸馏，包括一种新的策略，该策略通过插值旧编码器与当前编码器的权重来进行知识蒸馏。

II. 提出的 CoSMAE 方法

设 $\mathcal{T}_{1:T} = (\mathcal{T}_1, \dots, \mathcal{T}_T)$ 是一个由 T 任务组成的序列，其中每个任务 $\mathcal{T}_t$ 都与一组 $\mathcal{X}_t = \{X_1^t, \dots, X_{n_t}^t\}$ 的 n_t 个未标记的 RS 图像相关联。第一项任务 $\mathcal{T}_1$ 与一个 RS 图像档案 $\mathcal{X}_1$ 相关联，而随后的任务 $\mathcal{T}_{2:T}$ 可以与新获取的 RS 数据或新发布的 RS 训练集相关联。我们提出了一种新颖的自监督学习方法（记为 CoSMAE），该方法定义在最近在 RS 图像分类中取得显著成功的 MAEs 背景下。

提出的 CoSMAE 旨在在顺序训练 $\mathcal{T}_{1:T}$ 期间正则化 MAE，并提高跨任务的泛化能力，同时降低灾难性遗忘的风险。

为此，它包含两个组件：1) 数据混合；2) 模型混合知识蒸馏。前者通过在内存缓冲区中使用新图像与旧图像插值来操作数据层面，而后者则通过将当前编码器的模型权重与旧编码器的权重插值来操作模型层面以进行知识蒸馏。CoSMAE 的概述如图 1 所示。由于 CoSMAE 的定义严格依赖于 MAE，在以下小节中，我们首先在持续学习背景下介绍 MAE 的基础知识，然后详细描述 CoSMAE 的两个关键组件。

A. 关于 CL 背景下 MAE 的基础知识

对于一个由编码器 e 和解码器 d 组成的 MAE 模型 $f = (e, d)$ ，令 $f_t = (e_t, d_t)$ 表示在任务 $\mathcal{T}_{1:t}$ 上训练的模型。为了学习任务 $\mathcal{T}_t$ ，我们考虑了基于补丁的掩码图像建模和均方误差重建损失，定义如下：

$$\mathcal{L}_{Recon} = \frac{1}{N_r} \sum_{k=1}^{N_r} \|X_{i_k}^t - \hat{X}_{i_k}^t\|^2, \quad (1)$$

其中 $X_{i_k}^t$ 表示被掩码的来自 X_i^t 的第 k 个原始补丁， $\hat{X}_{i_k}^t$ 是模型学习到的其重构。被遮罩的块数 N_r 取决于遮罩比率 r 。对于学习第一个任务，(1) 是唯一的目标。对于后续任务的学习（即 $t \geq 2$ ），我们假设可以访问一个小内存，该内存由一个内存缓冲区和前一个编码器 e_{t-1} 的存储权重组成。内存缓冲区 $\mathcal{M}_t = \{X_1^{\mathcal{M}_t}, \dots, X_M^{\mathcal{M}_t}\}$ 的大小为 M ，存储来自先前任务的图像 $\mathcal{T}_{1:t-1}$ 。我们假设根据持续学习文献 [11]，无法访问来自先前任务的图像。一旦 MAE 成功训练，可以通过迁移学习将其应用于不同的下游任务。

B. 数据混排

该组件旨在从数据层面规范模型，以减少灾难性遗忘并促进所学特征的更好泛化。受 [8] 的启发，在 RS 中首次，我们利用无监督的 mixup 将当前任务 $\mathcal{T}_t$ 的图像与内存缓冲区 $\mathcal{M}_t$ 的图像进行线性插值。详细来说，对于学习任务 $\mathcal{T}_t$ ，我们不直接在 $\mathcal{X}_t$ 上训练 f_t ，而是使用图像

$$\tilde{X}_i^t = \lambda_1 X_i^t + (1 - \lambda_1) X_j^{\mathcal{M}_t}, \quad (2)$$

进行训练，其中 $\lambda_1 \sim \text{Beta}(\alpha_1, \alpha_1)$ 为 $\alpha_1 \in (0, \infty)$ 和随机选择的 $X_j^{\mathcal{M}_t} \in \mathcal{M}_t$ 。值得注意的是，我们从任务中均匀选择图像以存储在内存缓冲区 $\mathcal{M}_t$ 中。内存缓冲区中的图像被替换，以便内存缓冲区包含每个任务的等量图像。

C. 模型混合知识蒸馏

该组件旨在规范当前编码器 e_t ，以降低灾难性遗忘的风险并促进所学特征的更好泛化。为此，我们引入了一种新的知识蒸馏策略，同时提炼先前任务 $\mathcal{T}_{1:t-1}$ 和当前任务 $\mathcal{T}_t$ 中学到的知识。这是通过插值前一个编码器 e_{t-1} 的权重 θ_{t-1} 和当前编码器 e_t 的权重 θ_t 来获得作为知识蒸馏教师的混合编码器 $\tilde{e}_t$ 的权重 $\tilde{\theta}_t$ 实现的。详细来说，权重 $\tilde{\theta}_t$ 是通过以下方式获得的：

$$\tilde{\theta}_t = \lambda_2 \theta_t + (1 - \lambda_2) \theta_{t-1}, \quad (3)$$

其中 $\lambda_2 \sim \text{Beta}(\alpha_2, \alpha_2)$ 对于 $\alpha_2 \in (0, \infty)$ 。混合编码器 $\tilde{e}_t$ 结合了旧知识和当前任务的知识。需要注意的是，混合编码器 $\tilde{e}_t$ 没有经过训练，也不需要任何梯度计算。

令 c_i^t 表示当前编码器 e_t 的输出特征，对于图像 $\tilde{X}_i^t$ 。我们不直接正则化 c_i^t ，而是引入一个额外的可学习投影器 g ，在应用正则化之前（参见图 1）。它确保当前编

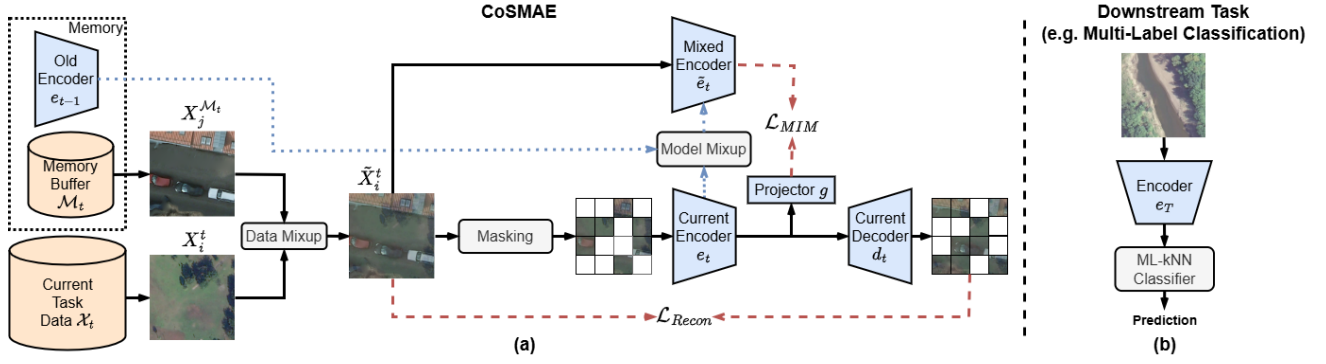


图 1. (a) 所提出的 CoSMAE 的概述。它采用数据混洗和模型权重知识蒸馏来正则化模型并保留来自先前任务的知识。黑色箭头显示数据流动，蓝色虚线箭头表示模型权重的流动，红色虚线箭头表示损失计算。(b) 我们下游任务评估的示意图。

码器 e_t 学习混合编码器 $\tilde{e}_t$ 的知识，而不强制当前编码器 e_t 的输出特征接近混合编码器 $\tilde{e}_t$ 的输出特征。迫使特征接近会降低模型适应新数据的能力。然后，在当前编码器的投影特征 $g(c_i^t)$ 和混合编码器的输出特征 $\tilde{c}_i^t$ 之间的知识蒸馏中，使用了互信息最大化 (MIM) 损失。MIM 损失基于归一化的温度缩放交叉熵 [12]。对于一批图像 $\mathcal{B}$ ，其定义如下：

$$\mathcal{L}_{MIM}(\mathcal{B}) = \frac{1}{2|\mathcal{B}|} \sum_{\tilde{X}_i^t \in \mathcal{B}} -\log\left(\frac{e^{S(g(c_i^t), \tilde{c}_i^t)/\tau}}{\sum_q \mathbf{1}_{\{i \neq q\}} e^{S(g(c_i^t), \tilde{c}_q^t)/\tau}}\right) -\log\left(\frac{e^{S(\tilde{c}_i^t, g(c_i^t))/\tau}}{\sum_q \mathbf{1}_{\{i \neq q\}} e^{S(\tilde{c}_i^t, g(c_q^t))/\tau}}\right) \quad (4)$$

其中， S 表示余弦相似度， τ 是温度参数，而 $\mathbf{1}$ 是指示函数。训练模型的损失由以下给出：

$$\mathcal{L} = \mathcal{L}_{Recon} + \beta \mathcal{L}_{MIM}, \quad (5)$$

其中超参数 β 决定了 MIM 损失对总损失的贡献。应注意到，对于混合编码器，未遮罩的图像被用作输入。 $\tilde{e}_t$

III. 实验结果

A. 数据集描述

在实验中，我们模拟了新训练集的顺序可用性，每个训练集对应一个 CL 任务。对于 CL，我们选择了以下数据集：1) US3D [13]; 2) UAVid [14]; 3) Potsdam [15]; 和 4) TreeSatAI [16]。作为下游任务，我们考虑了应用于 UCMerced 数据集 [17] 的多标签分类 (MLC)。US3D 数据集包含 WorldView-3 卫星在三个不同美国城市上空获取的图像。我们从内布拉斯加州的杰克逊维尔选择了图像子集。每张图像有八个波段，地面样本距离为 30 厘米。我们将图像分为大小为 256×256 像素的非重叠图像块。然后，我们随机选择原始训练集中

的一部分共 21081 个补丁用于训练，并从原始验证集选择了 3072 个补丁用于验证。UAVid 数据集包含从倾斜视角捕捉的不同视频序列中的 RGB 图像。每张图像大小为 3840×2160 像素。我们从原始训练和验证集中选取了这些图像，并将它们分成大小为 256×256 像素的非重叠图像块，结果得到 24560 个用于训练的补丁和 8640 个用于验证的补丁。波茨坦数据集包含 38 张大小为 6000×6000 像素的空中真正正射影像，具有四个波段。这些图像的空间分辨率为 5 厘米。对于实验，我们将图像分割成非重叠的 256×256 像素大小的图像块。我们随机选择了 15870 个补丁用于训练和 1058 个补丁用于验证。TreeSatAI 数据集包含 50380 张大小为 300×300 像素的空中图像，具有四个波段。这些图像的空间分辨率为 20 厘米。我们随机选择了一组 15000 张图像用于训练和 1000 张用于验证。UCMerced 数据集包含 2100 张大小为 256×256 像素的 RGB 图像，空间分辨率为 30 厘米。我们随机选择了 1000 张图像用于训练，300 张图像用于验证，以及 800 张图像用于测试。对于训练和评估，我们使用了在 [18] 中描述的图像多标签。

B. 试验设计

作为 MAE 架构的主体，我们为编码器和解码器都使用了 ViT 模型。编码器由 12 个块、12 个头、尺寸为 16 的 patch 以及 768 维的隐藏维度组成。解码器使用了 8 个块、16 个注意力头和 512 维的嵌入维度。遮罩比例 r 设置为 0.75。对于投影器 g ，我们使用了一个具有 128 维隐藏层和 ReLU 激活函数的两层 MLP。我们采用了随机水平翻转和随机缩放裁剪至 224×224 像素，最小裁剪比例为 0.8 的方法。对于每个任务，我们在批处理大小为 128 的情况下训练了 300 个周期。我们使用

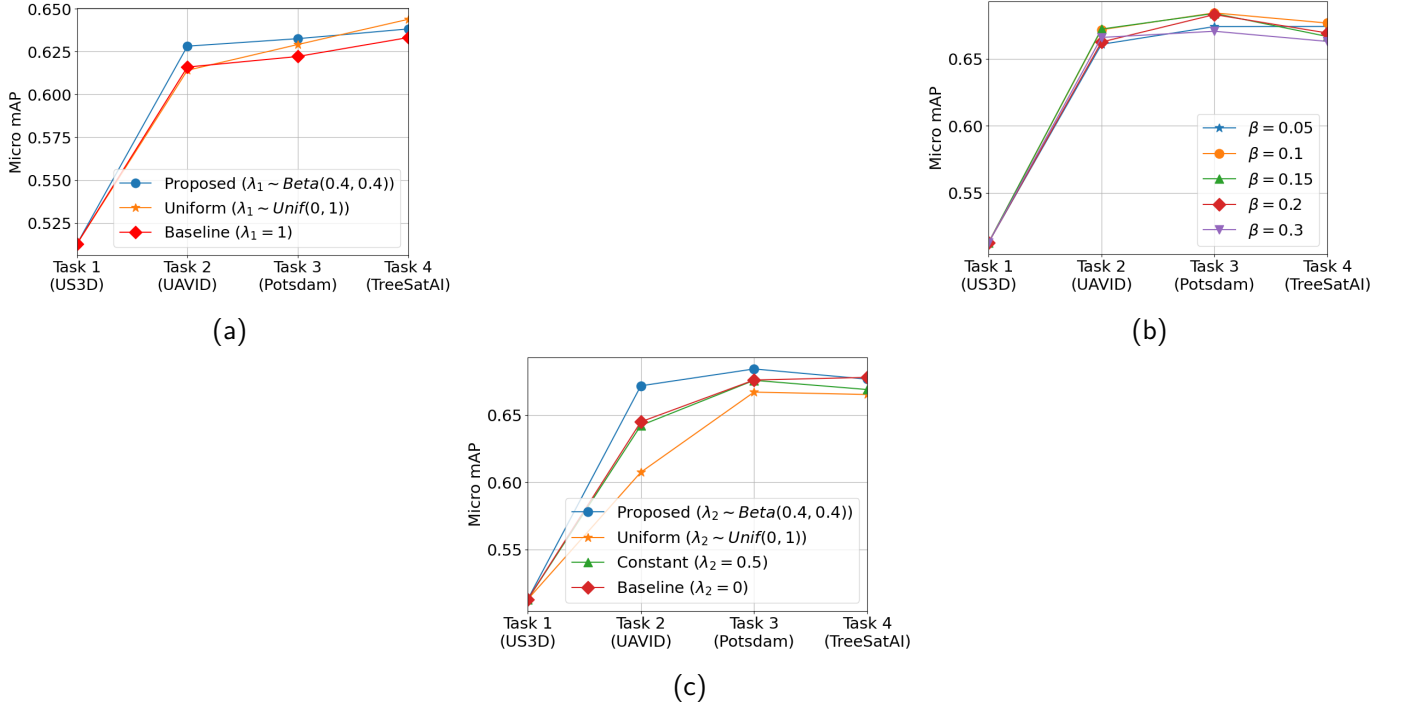


图 2. UCMerced 上依次训练任务所获得的微 mAP 结果，用于评估：(a) 所提出的数据混洗、均匀数据混洗和基准；(b) 不同值的 β ；以及 (c) 所提出的模型混合知识蒸馏、均匀模型混洗、常数模型混洗和基准。

AdamW 优化器，学习率为 $1e^{-3}$ ，并且进行了前 10 个周期的线性预热以及余弦退火。内存缓冲区 $\mathcal{M}_t$ 的大小 M 设置为 1000。MIM 损失中的温度 τ 设置为 0.5。为了评估，学习到的编码器在评估数据集中用作特征提取器，并使用带有 $k = 10$ 的多标签 k -最近邻 (ML- k NN) 分类器对提取的表示进行分类。结果以微观平均精度均值 (mAP) 提供。请注意，我们在宏观 mAP 方面观察到了类似的行为（由于篇幅限制未在论文中报告）。在我们的实验中，我们使用了每个数据集的 RGB 波段。所有实验结果都是基于三次试验中使用三个随机初始化模型获得的平均微观 mAP 得出的。所有实验均在一个 NVIDIA A100 GPU 上进行。

C. 消融和敏感性研究

在本小节中，首先我们展示了一个消融研究来分析所提出方法的两个组成部分的影响。为此，我们展示了通过排除所提方法个别组件而获得的结果。当不使用数据混洗时，模型仅在当前任务的数据上训练而不进行图像插值。当不使用模型混合知识蒸馏时，只使用重建损失进行训练，没有与混合编码器的知识蒸馏。表 I 显示了完成特定连续学习任务序列中的某个任务后的 UCMerced 分类获得的微平均 mAP。从表格中可以看出，当同时使用数据混洗和模型混合知识蒸馏时，获得

了最高的微平均 mAP。此外，我们观察到，与数据混洗相比，模型混合知识蒸馏对性能的影响更大。

表 I

消融研究：使用所提出的 CoSMAE 的不同组件在 UCMERCEd 上进行顺序训练的各个任务中获得的微小 MAP (%)。

Data Mixup	Model Mixup KD	Task 1 (US3D)	Task 2 (UAVID)	Task 3 (Potsdam)	Task 4 (TreeSatAI)
✗	✗		60.33	61.29	61.12
✓	✗		62.82	63.25	63.83
✗	✓	51.28	65.59	67.17	66.93
✓	✓		67.19	68.44	67.68

我们还分析了在不同的采样策略下，插值参数 λ_1 的数据混排效果，未应用模型混排知识蒸馏。图 2a 显示了完成特定任务序列中的连续学习任务后的微平均 mAP。从图中可以看出，所提出的数据混排策略 ($\lambda_1 \sim \text{Beta}(0.4, 0.4)$) 在所有任务上都获得了最高的微平均 mAP。基于均匀分布采样的数据混排 ($\lambda_1 \sim \mathcal{U}_{[0,1]}$) 在第二和第三个任务上的表现较低，但在最终任务上略微优于提出的混合方法。在提出的数据混排中， λ_1 被高概率选择接近 0 或 1。通过这种方式，数据混排可以同时作为旧数据信息回放和一种正则化模型的数据增强策略。这使得保留了旧数据的知识并提高了模型的泛化能力。当应用均匀采样策略时， λ_1 在 0 和 1 之

间的所有值都具有相同的概率。因此，对于那些不接近 0 或 1 的 λ_1 值，存在许多插值图像。这些图像遵循的数据分布与任务的数据分布非常不同，这是由于不同的形状和纹理重叠所造成的。因此，模型表现变差。在最后一个任务上的不同行为可能归因于 TreeSatAI 数据集与其他数据集相比的不同特性。我们还分析了模型混合知识蒸馏相对于 β 的效果，以确定 MIM 项与重构损失的贡献。从图 2b 可以看出，对于 $\beta = 0.1$ ，达到了最高的 mAP 值。最后，我们分析了在不同的采样策略下模型混合知识蒸馏对插值参数 λ_2 的影响（参见图 2c）。在这种情况下，对于所有模型我们使用数据混排、相同的投影器以及用于知识蒸馏的互信息最大化损失函数 $\beta = 0.1$ 。从图 2c 可以观察到，所提出的模型混排知识蒸馏 ($\lambda_2 \sim \text{Beta}(0.4, 0.4)$) 在所有任务中整体上实现了最高的微平均 mAP。使用一个常数插值参数 $\lambda_2 = 0.5$ 的性能与基线 ($\lambda_2 = 0$) 相似，而均匀采样 λ_2 ($\lambda_2 \sim \mathcal{U}_{[0,1]}$) 则会降低性能。在提出的模型 mixup 知识蒸馏中， λ_2 被选为接近 0 或 1 的概率很高。因此，混合编码器可以利用旧任务和当前任务的知识来规范化当前编码器。模型被迫学习能够泛化到两种正则化的特征，从而形成一个更稳健的模型。当对 λ_2 ($\lambda_2 \sim \mathcal{U}_{[0,1]}$) 应用均匀采样策略时，混合编码器在旧编码器和当前编码器之间的线性插值中以相等的概率采取任何权重组合。这导致模型过度受到正则化的约束，在所有任务上的表现都很差。

D. 与最先进的 CL 方法比较

在本小节中，我们将 CoSMAE 与最先进的连续学习策略（经验回放 [ER] [19]，EWC [7]，黑暗经验回放 [DER] [11]，终身无监督混合 [LUMP] [8]，投影函数正则化 [PFR] [9]）应用于 MAE 进行了比较，同时也与一个基线（该基线基于任务顺序训练 MAE 但不应用连续学习策略）进行了比较。DER 最初是为监督式连续学习问题提出的。我们通过用编码特征的正则化替换内存图像对数的正则化来将其适应为自监督式，如 [8] 中所提议的。表 II 显示了在完成连续学习任务序列中的特定任务后的微 mAP。此外，表 II 展示了完成整个任务序列所需的总训练时间。需要注意的是，在第一个任务中，所有方法的训练是相同的，因为在这一阶段没有采用任何连续学习策略。从表格中可以看出，CoSMAE 的表现显著优于其他方法。例如，在第二个任务中，CoSMAE 实现了 66.71% 的微 mAP，这比最先进的技术高出 4.94%，

表 II

在 UCMERGED 数据集上通过顺序训练的每个任务获得的微平均精确率 (%) 以及所提出方法、最新方法和基线方法的训练时间 (小时)。

Method	微小 mAP				Total training time
	Task 1 (US3D)	Task 2 (UAVID)	Task 3 (Potsdam)	Task 4 (TreeSatAI)	
Baseline		60.33	61.29	61.12	10.08
ER [19]		61.18	61.72	62.57	10.41
EWC [7]		60.93	61.68	63.06	10.9
DER [11]	49.11	61.75	61.69	61.26	11.92
LUMP [8]		61.65	61.53	62.37	10.16
PFR [9]		61.77	66.00	66.32	12.93
CoSMAE		66.71	68.04	67.26	13.77

比基线高出 6.38%。对于第三个任务，所提出的方法比最先进的技术高出 2.04%，比基线高出 6.75%。在上一项任务中，所提出的方法比最先进的技术高出 0.94%，比基线高出 6.14%。大多数最先进的技术仅比基线有微小的改进。从表中我们还可以看到，CoSMAE 的总训练时间为 13.77 小时，比基线多 3.69 小时，比第二好的方法 PFR 多 0.84 小时。CoSMAE、PFR 和 DER 的增加的训练时间归因于知识蒸馏，在训练期间需要额外的前向传递。我们希望指出的是，基线预计具有最低的训练时间，因为没有采用任何保留旧知识的附加策略。

IV. 结论

在信中，我们提出了一种用于遥感的持续自监督学习方法（记为 CoSMAE）。该方法由两个部分组成：i) 数据混合；ii) 模型混合知识蒸馏。这两个组件的联合考虑降低了灾难性遗忘的风险，同时提高了模型对下游任务的泛化能力。从实验结果来看，CoSMAE 在应用于 MAE 背景下的最先进的 CL 方法中表现优于其他方法，但训练时间略有增加。特别是，其新颖的组件（即模型混合知识蒸馏）通过形成基于插值模型权重的强大教师，使模型能够学习更具辨别力的特征，从而大大提高了性能。值得强调的是，虽然我们在 MLC 问题中使用 RGB 数据集评估了该方法，但它独立于所考虑的学习任务（因此可以通过使用特定于任务的头部与任何任务一起使用）以及数据模态。作为未来的工作，我们计划利用该方法对不同传感器（例如，多光谱、高光谱）在不同学习任务（例如，语义分割）下获取的多模态数据进行顺序训练。

参考文献

- [1] D. Tuia, K. Schindler, B. Demir, X. X. Zhu, M. Kochupillai, S. Džeroski, J. N. van Rijn, H. H. Hoos, F. Del Frate, M. Datcu, V. Markl, B. Le Saux, R. Schneider, and G. Camps-Valls, “Artificial intelligence to advance earth observation: A review of models, recent trends, and pathways forward,” *IEEE Geoscience and Remote Sensing Magazine*, pp. 2–25, 2024.
- [2] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars, “A continual learning survey: Defying forgetting in classification tasks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3366–3385, 2022.
- [3] N. Ammour, “Continual learning using data regeneration for remote sensing scene classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [4] H. Zhuang, Y. Yan, R. He, and Z. Zeng, “Class incremental learning with analytic learning for hyperspectral image classification,” *Journal of the Franklin Institute*, vol. 361, no. 18, p. 107285, 2024.
- [5] X. Lu, X. Sun, W. Diao, Y. Feng, P. Wang, and K. Fu, “LIL: Lightweight incremental learning approach through feature transfer for remote sensing image scene classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–20, 2022.
- [6] V. Marsocci and S. Scardapane, “Continual barlow twins: Continual self-supervised learning for remote sensing semantic segmentation,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 5049–5060, 2023.
- [7] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska et al., “Overcoming catastrophic forgetting in neural networks,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.
- [8] D. Madaan, J. Yoon, Y. Li, Y. Liu, and S. J. Hwang, “Representational continuity for unsupervised continual learning,” in *International Conference on Learning Representations*, 2022.
- [9] A. Gomez-Villa, B. Twardowski, L. Yu, A. D. Bagdanov, and J. Van de Weijer, “Continually learning self-supervised representations with projected functional regularization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 3867–3877.
- [10] J. Hackstein, G. Sumbul, K. Norman Clasen, and B. Demir, “Exploring masked autoencoders for sensor-agnostic image retrieval in remote sensing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–14, 2025.
- [11] P. Buzzega, M. Boschini, A. Porrello, D. Abati, and S. Calderara, “Dark experience for general continual learning: a strong, simple baseline,” *Advances in neural information processing systems*, vol. 33, pp. 15 920–15 930, 2020.
- [12] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” *International conference on machine learning*, pp. 1597–1607, 2020.
- [13] M. Bosch, K. Foster, G. Christie, S. Wang, G. D. Hager, and M. Brown, “Semantic stereo for incidental satellite images,” *IEEE Winter Conference on Applications of Computer Vision*, pp. 1524–1532, 2019.
- [14] Y. Lyu, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, “UAVid: A semantic segmentation dataset for UAV imagery,” *ISPRS journal of photogrammetry and remote sensing*, vol. 165, pp. 108–119, 2020.
- [15] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf, “The ISPRS benchmark on urban object classification and 3D building reconstruction,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences; I-3*, vol. 1, no. 1, pp. 293–298, 2012.
- [16] S. Ahlswede, C. Schulz, C. Gava, P. Helber, B. Bischke, M. Förster, F. Arias, J. Hees, B. Demir, and B. Kleinschmit, “TreeSatAI benchmark archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing,” *Earth System Science Data Discussions*, pp. 1–22, 2022.
- [17] Y. Yang and S. Newsam, “Geographic image retrieval using local invariant features,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 818–832, 2012.
- [18] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, “Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1144–1158, 2017.
- [19] A. Robins, “Catastrophic forgetting, rehearsal and pseudorehearsal,” *Connection Science*, vol. 7, no. 2, pp. 123–146, 1995.