用于无示例类增量学习的任务自适应显著性引导

刘夏雷 [2,1,*] 翟江天 [1,*] Andrew D. Bagdanov [3] 李珂 [4] 程明明 [2,1,†]

[1] VCIP, CS, 南开大学 [2] 南开国际先进研究院, 深圳福田 [3] MICC, 佛罗伦萨大学 [4] 腾讯优图实验室

{xialei,cmm}@nankai.edu.cn, {jtzhai30,tristanli.sh}@gmail.com, andrew.bagdanov@unifi.it

Abstract

无示例类增量学习（*EFCIL*）旨在仅通过当前任务数据顺序学习任务。*EFCIL* 具有重要意义，因为它缓解了有关数据隐私和长期数据存储的担忧，同时也缓解了增量学习中的灾难性遗忘问题。在这项工作中，我们引入了适用于 *EFCIL* 的任务自适应显著性，并提出了一种新框架，称之为任务自适应显著性监督（*TASS*），以减轻不同任务之间显著性漂移的负面影响。我们首先使用边界引导的显著性来保持模型注意力的任务适应性与可塑性。此外，我们引入了任务无关的低层信号作为辅助监督，以增加模型注意力的稳定性。最后，我们引入了一种用于注入和恢复显著性噪声的模块，以增强显著性保留的鲁棒性。我们的实验表明，我们的方法可以更好地保留跨任务的显著性图，并在 *CIFAR-100*、*Tiny-ImageNet* 和 *ImageNet-Subset* 的 *EFCIL* 基准测试中取得了最先进的结果。代码可在*https://github.com/scok30/tass*获取。

1. 引言

深度神经网络在许多计算机视觉任务上取得了最先进的性能。然而，大多数这些任务都只考虑一个静态世界，其中任务是明确定义且稳定的，并且所有训练数据都在单个训练会话中可用。真实世界由动态变化的环境和数据分布组成，尤其考虑到训练大型卷积神经网络的计算负担，这些因素重新引发了对增量学习新任务同时避免灾难性遗忘的研究兴趣 [14, 36]。

类增量学习（CIL）[1, 35] 是考虑向已训练的模型

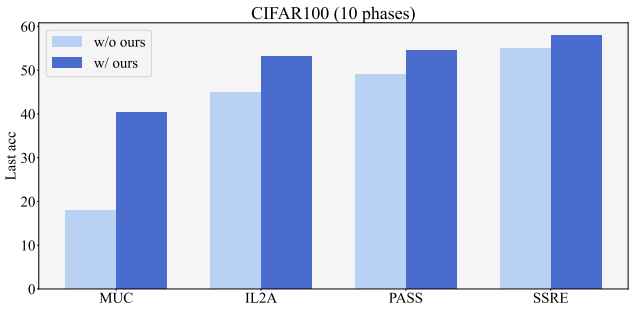

图 1. 我们提出了 TASS 方法，可以直接应用于许多最近的无示例类增量学习方法，从而显著提升 EFCIL 分类准确性，并减少灾难性遗忘。

中添加新类别的可能性的研究场景。大多数类增量学习方法依赖于一个内存缓冲区，用于存储来自过去任务的范例 [2, 10, 39, 44]。在本文中，我们考虑的是无示例类增量学习（EFCIL），这是一个更具挑战性的设置，其中不保留来自先前任务的任何数据。这是一种现实场景，由于隐私担忧或对数据长期存储的限制，这一情况备受关注。然而，无法保留来自过去任务的示例显著加剧了灾难性遗忘的问题。

有几项最近的研究工作考虑了 EFCIL 问题。DeepInversion [49] 反转训练的网络，从随机噪声生成图像作为范例，并将其与当前任务样本混合进行训练。SDC [50] 通过假设可以使用新数据近似和估计先前任务中类别的语义漂移，更新每个学习类别的原型。其他先前的研究工作提出了表示学习方法，用于克服灾难性遗忘 [55, 56]。如 IL2A [55] 指出，学习更好的表示可以减少在转移到新任务时的表示偏差。加入自监督学习任务，例如 Barlow Twins [38] 和旋转预测 [56]，也被提出以实现更稳定的表示并减轻遗忘。卷积神经网络（CNNs）天然地学习关注对其训练的任务具有区

*前两位作者贡献相同。

†通讯作者

分性的特征。在无示例类增量学习（EFCIL）中，灾难性遗忘也会发生，因为模型关注的显著特征漂移至新任务特定的特征。在学习新任务时，标准的正则化方法很少能防止这种显著性漂移。一种直接的约束显著性的方法是对样本的显著性图进行蒸馏 [11]。然而，在 EFCIL 的设置中，由于无法保存来自先前任务的样本，情况变得更加复杂。另一种方法是在当前任务样本和先前任务注意力之间应用显著性蒸馏 [9]。然而，在增强显著性一致性时，这种方法受到当前类别和旧类别之间的语义差距的影响。缺乏显著性约束可能导致注意力在未来任务中向背景漂移，从而导致遗忘。此外，仅仅在注意力上应用蒸馏 [9] 无法提供可塑性，容易受到注意力遗忘的影响，这是知识遗忘的一个关键因素。相比之下，我们的任务自适应显著性监督（TASS）方法旨在保持显著性集中于增量学习的任务上（更多详情请参见第 4节），同时保持其可塑性和稳定性。通过监督注意力，它提升了许多先前的 EFCIL 方法的性能，如图 1所示。

具体而言，TASS 整合了三个部分来解决这个问题。首先，我们使用膨胀的边界图以防止模型中间层跨越物体边界的显著性漂移。由于显著性漂移通常发生在跨任务时，通过膨胀边界监督鼓励模型集中于显著的前景区域，减少了显著性向背景转移的可能性，从而使模型能够自适应地选择与任务相关的前景内的注意力区域。其次，为了同时增强模型跨任务的注意力稳定性，我们在类增量框架中添加了一个与我们的核心 EFCIL 任务密切相关的与任务无关的低层辅助监督任务，这是因为图像分类已被证实有助于模型定位图像中最显著的区域。最后，我们提出了一种模块，将显著性噪声注入到特定的特征通道中，并训练网络对其进行去噪，帮助网络进一步抵抗跨任务的注意力漂移。

本工作的主要贡献为：**(i)** 我们为 EFCIL 设定下的任务自适应显著性监督提供了新的见解；我们还展示了缺乏或不足的显著性监督方法的负面影响，这说明了我们方法的优越性，并激发了在 EFCIL 中减轻显著性漂移的需求。**(ii)** 我们提出了由三个部分构成的任务自适应显著性监督（TASS），这些部分共同用于缓解显著性漂移问题。**(iii)** 我们展示了 TASS 可以轻松集成到其他最先进的方法中，如 MUC [30]，IL2A [55]，PASS [56]，SSRE [57]，从而实现显著的性能提升。**(iv)** 我们的实验表明，TASS 在 CIFAR-100、Tiny-ImageNet 和 ImageNet-Subset 的 EFCIL 基准测试中优于所有现有的 EFCIL 方法，甚至优于几种基于示例的方法。

2. 相关工作

我们首先讨论最近文献中关于增量学习的先前工作，然后描述关于 EFCIL 的工作。

2.1. 增量学习

在过去几年中，针对增量学习已经提出了多种方法 [1, 7]。最近的研究工作可以粗略地分为三类：基于回放的方法、基于正则化的方法和参数隔离方法。基于回放的方法通过保留先前任务的训练样本以减轻任务新近度偏差 [39, 52]。GEM [32]、AGEM [4] 和 MER [40] 通过调整当前训练样本的梯度以匹配旧样本从而利用过去任务的示例。回放可能会导致模型对存储样本过拟合。基于正则化的方法，如 LwF [24]，EWC，R-EWC [18, 29] 和 DMC [53]，提供了学习更好表示的方法，同时保留足够的可塑性以适应新任务。参数隔离方法 [34, 46] 使用针对每个任务具有不同计算图的模型。借助增长模型，新的模型分支以增加参数和计算成本为代价减轻灾难性遗忘。这也在其他领域得到了广泛研究，例如语义分割 [3, 26, 45, 51] 和目标检测 [13, 31]。

关于基于显著性引导的增量学习，LwM [9] 在当前任务数据上约束先前类别的显著性激活。然而，当前类别和旧类别之间存在语义差距，这导致在保留旧样本上的显著性激活时出现了不准确的蒸馏目标。RRR [11] 直接将每个样本的 Grad-CAM 显著性激活保存在回放缓冲区中，并应用蒸馏以记忆这些旧知识，这在增量学习过程中需要存储额外的样本。尽管在显著性引导的增量学习方面进行了这些初步研究，但显著性漂移问题仍然存在，并导致灾难性遗忘。

2.2. 无示例类增量学习

与传统的类增量学习相比，无示例类增量学习更适用于训练数据敏感且可能不会永久存储的应用场景。DAFL [5] 使用 GAN 从过去任务中生成合成样本作为存储实际数据的替代方案。DeepInversion 是另一种流行的 EFCIL 方法，它反转已训练的网络，使用随机噪声生成图像 [49]。Always Be Dreaming 在 EFCIL 领域进一步改进了 DeepInversion [43]。SDC 试图克服在旧类样本上训练新任务时由语义漂移引起的问题 [50]。它

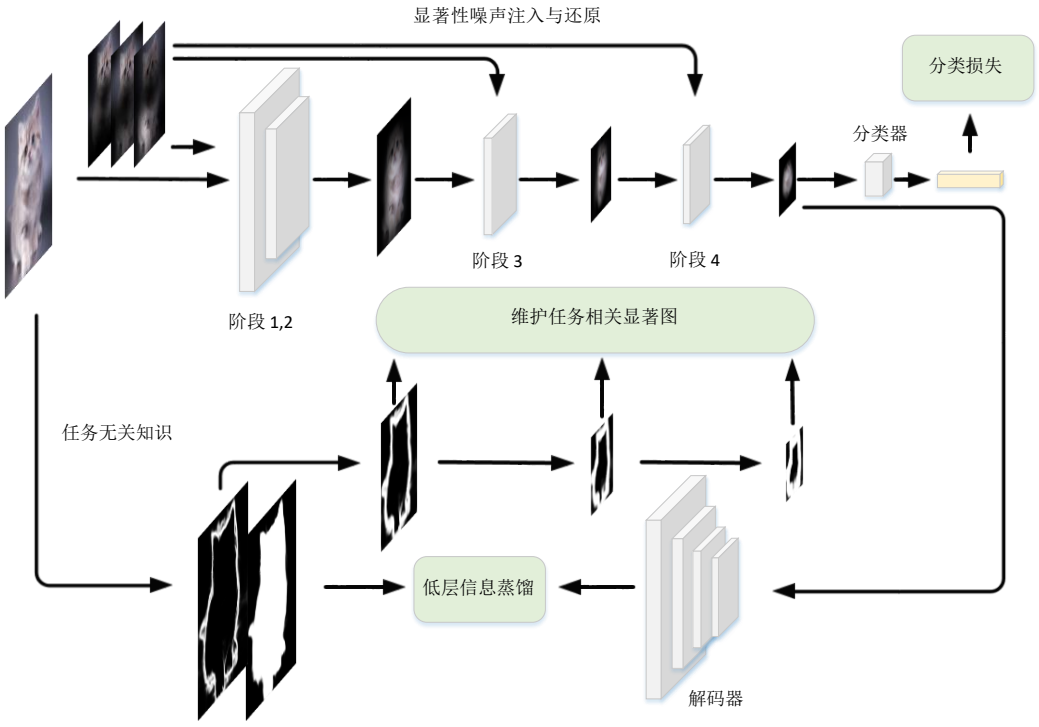

图 2. 任务自适应显著性监督（TASS）的整体框架。我们应用一个低层模型生成显著性图和边界图。边界图通过膨胀和下采样，以在编码器的不同阶段提供监督。在编码器之后附加一个解码器用于低层蒸馏，作为固定的、与任务无关的显著性引导。为了防止在后续训练阶段出现显著性漂移，我们在每个编码器阶段引入显著性噪声。模型被训练进行去噪，并减少当前数据在未来阶段的显著性漂移。TASS 可以集成到 EFCIL 方法中，为增量任务提供稳健的显著性引导。

直接估计每个学习类别的原型，以在最近类均值分类器中使用。PASS [56] 和 IL2A [55] 是基于原型的回放方法，用于高效且有效的 EFCIL。SSRE [57] 引入了一种重参数化方法，以在旧知识和新知识之间进行权衡。我们的任务自适应显著性监督（TASS）方法使用三个新部分以减少 EFCIL 中的显著性漂移，可以与上述几种方法互补。

3. 任务自适应显著性监督

我们首先定义了无示例类增量学习（EFCIL）场景。然后我们描述了我们的 TASS 方法，包括膨胀边界监督、辅助低层监督和显著性噪声注入。我们的整体框架如图 2 所示。

3.1. 无示例类增量学习

类增量学习旨在顺序学习由不相交类别样本组成的任务。记 $t \in 1, 2, \ldots, T$ 表示增量学习任务。每个任务的训练数据 D_t 包含 C_t 个类别，其中有 N_t 个训练样本 $(x_t^i, y_t^i)_{i=1}^{N_t}$，其中 x_t^i 是图像，$y_t^i \in C_t$ 是它们的标签。应用于类增量学习的大多数深度网络可以分为两个部分：一个特征提取器 F_θ 和一个通用分类器 G_ϕ，后者随着每个新任务 $t+1$ 的到来而增长，以包含类别 C_{t+1}。特征提取器 F_θ 首先将输入 x 映射到一个深度特征向量 $z = F_\theta(x) \in \mathbb{R}^d$，接着统一的分类器 $G_\phi(z) \in \mathbb{R}^{|C_t|}$ 产生类别 C_t 上的概率分布，用于对输入 x 进行预测。

类增量学习要求模型能够在任何训练任务中正确分类来自先前任务的所有样本——换言之，在学习任务 t 时，模型不能忘记如何对来自任务 $t' < t$ 的类别样本进行分类。无示例类增量学习进一步限制模型在学习每个新任务时不能访问先前任务的样本。一般而言，学习目标为最小化在当前训练数据 D_t 上定义的损失函数 $\mathcal{L}$：

$$\mathcal{L}_t^{\mathrm{CIL}}(x, y) = \mathcal{L}_{\mathrm{ce}}(G_{\phi_t}(F_{\theta_t}(x)), y) + \mathcal{L}_t^{\mathrm{M}}, \qquad (1)$$

其中 $\mathcal{L}_{\mathrm{ce}}$ 是标准的交叉熵分类损失，$\mathcal{L}_t^{\mathrm{M}}$ 是一种特定于方法的损失，用于在增量学习过程中减轻遗忘。注意，

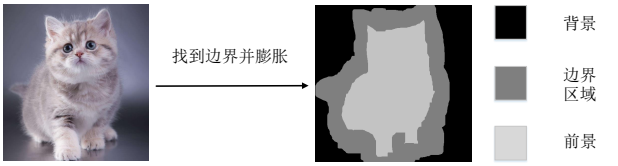

图 3. 我们对边界图进行膨胀操作，并在 CNN 主干的三个阶段应用二元交叉熵损失，以防止中层注意力漂移到边界区域。

若没有 $\mathcal{L}_t^{\mathrm{M}}$，方程 1 就会简化为对任务 t 进行微调。

3.2. 边界引导的中层显著性漂移正则化

简单地在任务之间蒸馏注意力并未考虑任务自适应的注意力。由一个模型生成每个输入图像 x 的低层表示（在我们的实验中是显著性图和边界图）。我们使用 CSNet [6] 生成显著区域和边界图，因为它轻量且高效，但在我们的框架中可以使用任何生成显著性图的模型（我们在补充材料中探讨了其他选项）。为了在主干网络的这些中间层引入注意力的可塑性，我们使用生成的边界图作为一种自适应监督，如图 3 所示。我们在物体边界上添加惩罚项，以避免注意力漂移到背景中。我们首先使用 0.5 作为阈值将生成的边界图二值化，然后通过以下方式膨胀边界图：

$$B_d(x) = \mathrm{Dilate}(A_b(x), d), \tag{2}$$

其中 $A_b(x)$ 是图像 x 的生成边界图，它由对显著性图使用拉普拉斯滤波器转换而来，d 表示应用在边界图上的膨胀半径，用于控制边界引导显著性的严格程度。

我们的模型生成中层显著性图时，并非像上述描述的低层显著性图那样在每一层使用解码器，而是在 CNN 主干的三个阶段使用 Grad-CAM [41][1]（详见图 2）。我们还尝试了几种其他用于生成学生显著性图的方法，并在补充材料中对它们进行了报告。为了与 Grad-CAM 生成的显著性边界图进行比较，通过下采样将生成的膨胀边界图 $B_d(x)$ 匹配到这三个阶段的特征图尺寸。我们在膨胀边界区域上使用二元交叉熵损失进行监督。该损失定义如下：

$$\mathcal{L}_t^{\mathrm{dbs}}(x) = -\frac{\sum_{j=1}^{N} B_d(x,j)\log(1-S(x,j))}{\sum_{j=1}^{N} B_d(x,j)}, \tag{3}$$

其中，$S(x,j)$ 表示我们模型在图像 x 上的像素 j 处的显著性图，$B_d(x,j)$ 是像素 j 处的膨胀生成的边界图，N 是图像 x 中的像素数。我们仅在膨胀边界区域内计算此损失（即 $B_d(x,j)=1$ 的地方）。这种损失有助于使学生显著性图避免与膨胀的教师边界区域相交。

[1] https://github.com/jacobgil/pytorch-grad-cam

3.3. 针对 EFCIL 的辅助低层监督

在类增量学习期间，我们提出从所有增量分类任务共享的低层静态任务中学习稳定特征。低层视觉任务，如显著目标检测，需要输入图像的有用表示。通过跨任务学习这些特征表示，模型可以专注于输入图像的关键区域，并利用学到的稳定特征减少表示漂移，因为低层特征在任务之间变化非常小。

显著性图预测与图像分类相关，因为前景在很大程度上决定了结果，而背景相对不重要。在学习具有新类别的新任务时，新类别图像的背景可能包含引入不必要噪声的新视觉概念，导致遗忘关键的先前知识。显著性引导训练 [17] 展示了显著性特征对学习分类任务的有效性。对显著区域边界的额外监督可以帮助显著目标检测任务的分割和定位 [12, 21, 25, 33, 54]。这两种任务之间的积极互动为与主分类任务相关的特征带来更丰富的注意力。它可以以静态知识的形式在类增量任务之间提供积极的指导。一些示例在图 5 中有所呈现。

我们将低层视觉任务作为网络的辅助监督，用于丰富与任务无关的注意力。边界图是通过对估计的显著性图使用拉普拉斯滤波器来计算。我们在主干网络 F_θ 后添加一个解码器 D_ψ [22]，用于预测输入图像的低层显著性图和边界图。将预测值和目标之间的平均 L2 距离作为低层显著性蒸馏损失：

$$\mathcal{L}_t^{\mathrm{lms}}(x) = \frac{||D_\psi(F_\theta(x)) - A(x)||_2}{\sqrt{N}}, \tag{4}$$

其中，$A(x)$ 表示输入 x 上的目标低层图，包括显著性图 $A_s(x)$ 和边界图 $A_b(x)$。$D_\psi(F_\theta(x))$ 是解码器生成的组合的显著性图和边界图，N 是显著性图中的像素数。

3.4. 显著性噪声注入

尽管我们应用低层蒸馏和膨胀边界监督在任务之间保持显著性表示，但模型仍可能在先前任务的样本中遗忘显著性。为解决这个问题，我们强制模型能够从注入的显著性噪声中恢复正确的显著性图。

在每个任务中，没有来自先前或未来任务的可用训练数据，因此我们无法直接知晓这些样本上的显著

Algorithm 1 TASS 伪代码

输入: 任务数目 T，任务 t 的训练样本 $D_t = \{(x_i, y_i)\}_{i=1}^{N}$，初始参数 $\Theta^0 = \{\theta_0, \phi_0, \psi_0\}$ 包含特征提取器 F_θ、分类器 G_ϕ 和低层解码器 D_ψ 参数。

输出: 模型 Θ^T

```
for t ∈ {1, 2, ..., T} do
    Θ^t ← Θ^{t−1}
    while 未收敛 do
        从 D_t 采样 (x, y)
        L_t^CIL ← 显著性噪声注入 (x, y)
        L_t^lms ← 低层多任务 (x, A(x))
        S ← 计算 GradCAM 显著性 (x, y)
        L_t^dbs ← 膨胀边缘监督 (S, A(x))
        通过公式 Eq. 5最小化 L_t^all 更新 Θ^t
    end while
end for
```

性漂移。我们没有使用真实显著性漂移信号来监督模型，而是在随机特征通道上引入显著性噪声。我们使用随机椭圆来近似未来任务中的潜在显著性漂移，并训练模型在每个阶段进行去噪。因此，模型可以学会有效减少实际的显著性漂移。

我们使用非常简单的方法生成椭圆形噪声。有六个参数维度：中心坐标 (x, y)、主轴和次轴长度 (a, b)、旋转角度 α 和掩码权重 w。这个过程的详细解释在补充材料中给出。借助膨胀边界监督，模型的每个阶段学会消除这种额外的显著性噪声，这有助于未来任务的泛化，并减轻先前任务中的显著性遗忘。

3.5. 学习目标与训练算法

总体学习目标结合了低层多任务学习、膨胀边界监督和随机显著性噪声注入模块：

$$\mathcal{L}_t^{\text{all}} = \mathcal{L}_t^{\text{CIL}} + \mathcal{L}_t^{\text{lms}} + \mathcal{L}_t^{\text{dbs}}. \tag{5}$$

将该损失与公式 1 进行比较，对于 TASS，$\mathcal{L}_t^{\text{M}} = \mathcal{L}_t^{\text{CIL}} + \mathcal{L}_t^{\text{lms}}$，因此将显著性感知监督与交叉熵损失结合在一起。整个过程详见算法 1。

4. 实验结果

在本节中，我们首先描述了我们的实验设置，然后将 TASS 与几个 EFCIL 基准方法进行比较。在第 4.3 节中，我们对 TASS 的各个部分进行了进一步分析。

4.1. 实验设置

我们遵循三个基准数据集上 EFCIL 的标准实验协议。

数据集。 我们在 CIFAR-100 [19]、Tiny-ImageNet [20] 和 ImageNet-Subset [8] 上进行实验。在大多数实验中，我们在第一个任务中训练模型学习一半的类别，然后将剩余的类别均匀分配给后续每个任务。我们使用的约定是：$F + C \times T$ 表示第一个任务包含 F 个类别，接下来的 T 个任务每个包含 C 个类别。这是 EFCIL 中常见的配置，它在 PASS [56] 和 SSRE [57] 中都有使用。

最先进方法。 由于我们专注于 EFCIL，我们主要与无示例的最先进方法进行比较：SSRE [57]、PASS [56]、IL2A [55]、EWC [18]、LwF-MC [39] 和 MUC [30]。为了展示我们方法的有效性，我们还将其性能与几种基于范例的方法进行比较，如 iCaRL（最近均值和 CNN）[39]、EEIL[2] 和 LUCIR [16]。我们还与集成了 SSRE 的 RRR [11] 进行比较，该方法专注于利用示例回放来保留显著性。

实现细节与性能指标。 我们使用 ResNet-18 [15] 作为特征提取的主干网络。与 SSRE [57] 和 PASS [56] 两种最先进的 EFCIL 方法使用相同的基础网络。我们使用 [22] 中的解码器来估计低层学生显著性图。所有实验都是从头开始使用 Adam 进行训练，共进行 100 个轮次，初始学习率为 0.001。学习率在第 45 和第 90 个轮次时按 10 的倍数减少。对于基于示例的方法，我们使用 herding [39] 来选择和存储每类 20 个样本，遵循常见的设置 [16, 39]。我们将 RRR [11] 与 SSRE 结合实现，以便与 TASS 进行公平比较。对于膨胀边界监督，我们将三个中层边界膨胀阶段中的 d 设定为图像尺寸的 5%、10% 和 15%。

我们报告了三种常见的类增量学习指标：平均和最终 top-1 准确率，以及截至任务 t 为止学习的所有类别的平均遗忘。记 Acc_i 为截至任务 i 为止学习的所有类别的准确率，平均准确率定义为 $Avg = \frac{\sum_{i=1}^{T} Acc_i}{T}$，最终准确率为 Acc_T。设 $a_{m,n}$ 表示在学习任务 m 后任务 n 的准确率，任务 i 在学习任务 k 后的遗忘度量 f_k^i 计算公式为 $f_k^i = \max_{t \in 1,2,\ldots,k-1}(a_{t,i} - a_{k,i})$。平均遗忘 F_k 定义为 $F_k = \frac{1}{k-1} \sum_{i=1}^{k-1} f_k^i$。

数据集	CIFAR-100			Tiny-ImageNet		
方法	5 任务	10 任务	20 任务	5 任务	10 任务	20 任务
MUC	38.45	19.57	15.65	18.95	15.47	9.14
+TASS	49.17 (+10.72)	40.34 (+20.77)	37.86 (+22.21)	32.47 (+13.46)	30.13 (+14.66)	27.70 (+18.56)
IL2A	55.13	45.32	45.24	36.77	34.53	28.68
+TASS	58.74 (+3.61)	53.24 (+7.92)	53.07 (+7.83)	42.49 (+5.72)	41.34 (+6.81)	40.59 (+11.91)
PASS	55.67	49.03	48.48	41.58	39.28	32.78
+TASS	59.10 (+3.43)	54.45 (+5.42)	52.37 (+3.89)	44.05 (+2.47)	43.06 (+3.78)	42.57 (+9.79)
SSRE	56.33	55.01	50.47	41.45	40.07	39.25
+TASS	59.26 (+2.93)	57.93 (+2.92)	53.78 (+3.31)	44.13 (+2.68)	43.86 (+3.79)	43.55 (+4.30)

表 1. 通过以即插即用的方式将 TASS 应用到其他 EFCIL 方法中，top-1 准确率的性能增益。绝对增益以 （红色） 表示。

数据集		CIFAR100									TinyImageNet								
设置		5 任务			10 任务			20 任务			5 任务			10 任务			20 任务		
方法		Avg↑	Last↑	$F\downarrow$	Avg↑	Last↑	$F\downarrow$	Avg↑	Last↑	$F\downarrow$	Avg↑	Last↑	$F\downarrow$	Avg↑	Last↑	$F\downarrow$	Avg↑	Last↑	$F\downarrow$
E=20	iCaRL-CNN†	51.07	40.12	42.13	48.66	39.65	45.69	44.43	35.47	43.54	34.64	22.31	36.89	31.15	21.10	36.70	27.90	20.46	45.12
	iCaRL-NCM†	58.56	49.74	24.90	54.19	45.13	28.32	50.51	40.68	35.53	45.86	33.45	27.15	43.29	33.75	28.89	38.04	28.89	37.40
	LUCIR†	63.78	55.06	21.00	62.39	50.14	25.12	59.07	48.78	28.65	49.15	37.09	20.61	48.52	36.80	22.25	42.83	32.55	33.74
	EEIL†	60.37	52.35	23.36	56.05	47.67	26.65	52.34	41.59	32.40	47.12	34.24	25.56	45.01	34.26	25.91	40.50	30.14	35.04
	RRR†	66.43	57.22	18.05	65.78	55.74	18.59	62.43	51.35	18.40	51.20	42.23	16.67	49.54	40.12	21.64	47.46	35.54	29.10
E=0	LwF_MC	45.93	36.17	44.23	27.43	50.47	17.04	20.07	15.88	55.46	29.12	17.12	54.26	23.10	12.33	54.37	17.43	8.75	63.54
	EWC	16.04	9.32	60.17	14.70	8.47	62.53	14.12	8.23	63.89	18.80	12.71	67.55	15.77	10.12	70.23	12.39	8.42	75.54
	MUC	49.42	38.45	40.28	30.19	19.57	47.56	21.27	15.65	52.65	32.58	17.98	51.46	26.61	14.54	50.21	21.95	12.70	58.00
	IL2A	63.22	55.13	23.78	57.65	45.32	30.41	54.90	45.24	30.84	48.17	36.14	25.43	42.10	35.23	28.32	36.79	28.74	35.46
	PASS	63.47	55.67	25.20	61.84	49.03	30.25	58.09	48.48	30.61	49.55	41.58	18.04	47.29	39.28	23.11	42.07	32.78	30.55
	SSRE	65.88	56.33	18.37	65.04	55.01	19.48	61.70	50.47	18.37	50.39	41.67	17.25	48.93	39.89	22.50	48.17	39.76	26.74
	TASS (Ours)	**68.75**	**59.26**	**16.42**	**67.42**	**57.93**	**17.78**	**62.76**	**53.78**	**17.78**	**55.12**	**44.13**	**15.40**	**54.21**	**43.86**	**18.47**	**52.79**	**43.55**	**22.51**

表 2. CIFAR-100 上不同任务数量下的平均 top-1 准确率、最终 top-1 准确率以及遗忘情况。基于回放的方法存储了每个先前类别的 20 个样本，用 † 标记。最佳整体结果用 **粗体**标出。我们所有实验均运行三次，并报告所有指标的平均值。

数据集	ImageNet-Subset								
设置	5 任务			10 任务			20 任务		
方法	Avg↑	Last↑	$F\downarrow$	Avg↑	Last↑	$F\downarrow$	Avg↑	Last↑	$F\downarrow$
LwF_MC	34.86	24.10	49.36	31.18	20.01	53.04	27.54	17.42	56.07
MUC	40.65	27.89	47.13	35.07	22.65	52.10	31.44	20.12	53.85
PASS	63.12	52.61	22.47	61.80	50.44	23.57	55.23	46.07	26.73
SSRE	69.54	58.46	17.22	67.69	57.51	18.60	61.23	50.05	23.22
TASS (Ours)	**74.32**	**63.14**	**14.37**	**72.60**	**57.93**	**16.09**	**68.79**	**57.60**	**18.41**

表 3. ImageNet-Subset 上不同任务数量下的平均 top-1 准确率、最终 top-1 准确率以及遗忘情况。我们所有实验均运行三次，并报告所有指标的平均值。

4.2. 与最先进方法的对比

我们在 CIFAR-100 上将 TASS 与最先进方法进行了比较，结果见表 2。TASS 的表现优于所有无示例方法。对于诸如 iCaRL [39]、EEIL [2] 和 LUCIR [16] 等基于示例的方法，我们的方法仍然具有明显更好的性能。在更长的序列（即 10 任务和 20 任务）上，与其他 EFCIL 方法相比，我们的方法在学习新类别时明显减少了遗忘。TASS 在最后一个任务上的表现比最佳方法 SSRE 高出约 3%。这种性能改进也可以从平均遗忘的角度观察到。

正如 Tiny-ImageNet 和 ImageNet-Subset 的表 3 和图 4 中所示，尽管我们的方法在图 4 中第一个任务的 top-1 准确率类似，但在大多数中间任务和最终任务上表现更好。在图 4 中更长序列的情况下，我们的方法与最佳基准方法之间的差距在很大程度上保持一致，表明我们的方法在减轻遗忘方面是有效的。与 CIFAR100 相比，表 3 中的性能提升在 Tiny-ImageNet 和 ImageNet-Subset 上更大，这表明我们的方法能推广到具有更大图像和物体尺度的数据集。值得一提的

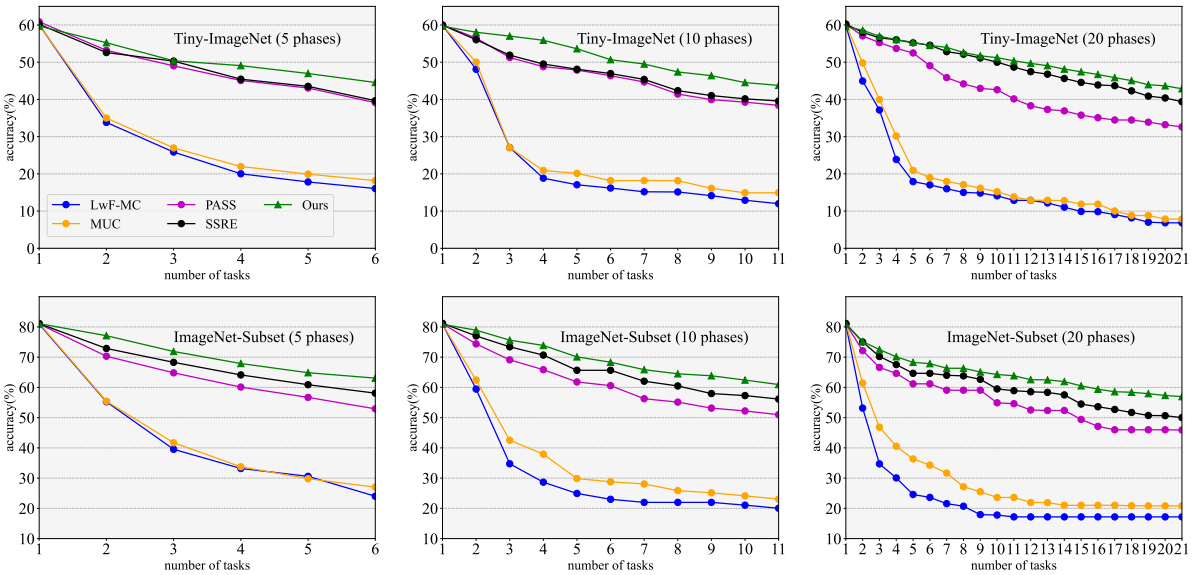

图 4. 针对不同任务数量在 Tiny-ImageNet 和 ImageNet-Subset 上的结果。我们的方法在效果上优于其他方法，特别是在更长的任务序列（即更多但更小的任务）上。

是，TASS 也产生具有较小方差的结果。我们认为这可能是由于 TASS 减少了对背景区域的显著性漂移，其中可能包括随机噪声。

与其他 EFCIL 方法即插即用。 一些现有的 EFCIL 方法，如 PASS [56], IL2A [55] 和 SSRE [57]，专注于通过嵌入正则化来减少遗忘。考虑到显著性对图像分类的重要性，自然会考虑是否可以将 TASS 整合到这些方法中。表1中的结果显示了这种集成带来的性能增益。在许多情况下，添加 TASS 可使 MUC 的性能翻倍，并显著提升 IL2A 和 PASS。当我们将其融入最佳基线 SSRE 时，它会产生约 3% 的持续增益。这些结果清楚地表明，通过显式地减轻显著性漂移，TASS 与其他缓解遗忘的方法是互补的。它们还证明了显著性漂移作为 EFCIL 灾难性遗忘的原因的重要性。

4.3. 其他分析

在本节中，我们更进一步研究我们提出的方法。若未特别指出，则使用集成到 SSRE [57] 中的 TASS 的结果。

消融实验。 我们使用 CIFAR-100 上的 10 任务设置执行消融（见表4）。我们在 PASS [56] 和 SSRE [57] 上进行消融。低层多任务监督是至关重要的，性能提升分别为 2.2% (PASS) 和 1.2% (SSRE)。膨胀边界监督进一步提升了约 1-2% 的性能。显著性噪声注入对两种方法都有帮助，并使 PASS 提升 1.5%。总体而言，TASS 相比于基线方法分别提升了 5.5% 和 2.9%。注意，SSRE 是先前最先进的方法，而 TASS 在很大程度上优于它。

方法 & 任务	$\mathcal{L}_{\text{lms}}$	$\mathcal{L}_{\text{dbs}}$	SNI	准确率
Baseline (PASS)				49.0
Variants	✓			51.2
	✓	✓		53.0
	✓	✓	✓	54.5
Baseline (SSRE)				55.0
Variants	✓			56.2
	✓	✓		57.3
	✓	✓	✓	57.9

表 4. 针对 TASS 各部分的消融。在 CIFAR-100 上进行了 10 任务设置的实验，我们报告了将 TASS 集成到 PASS 和 SSRE 中的 top-1 准确率（以百分比表示）。$\mathcal{L}_{\text{dbs}}$ (Eq. 3)、$\mathcal{L}_{\text{lms}}$ (Eq. 4) 和 SNI 分别表示 TASS 的三个组成部分：膨胀边界监督、低层多任务监督和显著性噪声注入。

低层多任务监督。 为分析我们提出的低层显著性监督的效果，我们在 5 任务和 10 任务设置下在 ImageNet-Subset 上进行了实验。我们首先在图5 (c) 中绘制了跨任务的损失。在第一个任务中学习预测边界和显著性

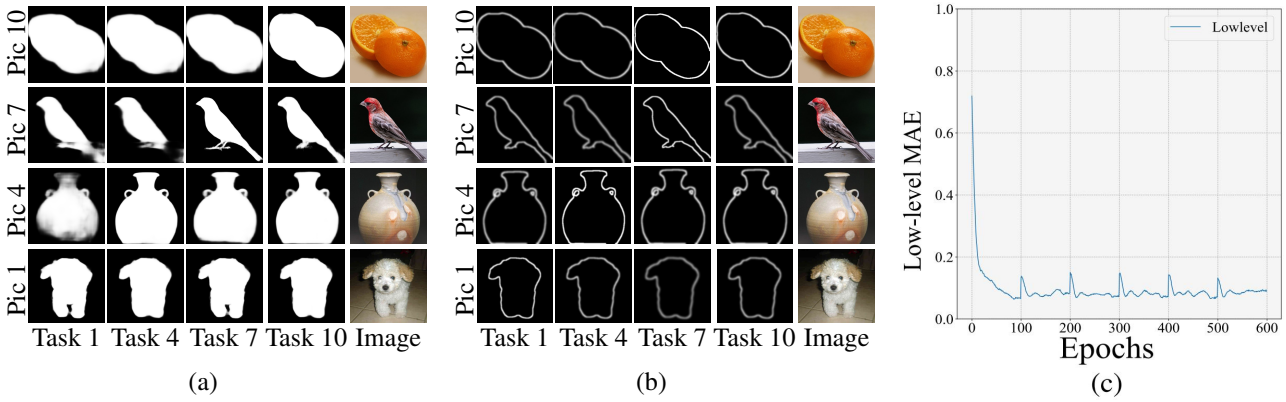

图 5. 可视化我们的学生编码器-解码器网络的显著性 (a) 和边界 (b) 图，其中包含增量学习不同阶段不同任务的原始图像。我们的方法产生了稳定的低层结果，同时减少了分类中的遗忘。(c) 跨任务的学生和教师网络之间的 MAE 损失。

图后，该网络在其余的 5 任务序列中保持了良好的性能。这表明在持续学习过程中，低层任务是稳定的。进一步，我们在图 5(a-b) 中可视化了增量学习过程中的显著性图和边界图预测结果。我们给出了一些在学习不同任务后预测的边界和显著性图的示例。尽管 CIL 涉及不同类别的样本，但我们发现低层输出是相对稳定并且与类别无关的。由于模型能够跨任务地稳定预测这些低层特征，它可以为持续学习保留有用的先验知识。

指标 & 方法	Avg↑	Last↑
FeTrIL [37]	65.20	56.34
SOPE [58]	65.84	56.80
PRAKA [42]	68.86	59.20
TASS (SSRE)	67.42	57.93
TASS (PRAKA)	**69.70**	**60.04**

表 5. 10 任务 CIFAR-100 设置的平均准确率和最终准确率。

指标 & 方法	Avg↑
SOPE [58]	60.20
FeTrIL [37]	65.00
TASS (FeTrIL)	**66.03**

表 6. ImageNet-Full 上 10 任务设置的平均准确率和最终准确率。

TASS 与更多的方法和基准。 由于我们的方法具有较强的泛化能力，我们还将我们的 TASS 范式应用于 PRAKA [42]，从而进一步提升了性能。我们在表5中给出了具体的实验对比。我们还在上表5和表6中与 FeTrIL [37] 进行了比较。上表6中在 ImageNet-Full 上的实验显示了一致的提升。

显著性漂移的定量分析。 我们衡量了 DINO 中最后一层的自注意力图与 SSRE 和 SSRE+TASS 的 Grad-CAM 显著性图之间的交并比 (IoU)。如表7所示，TASS 显著降低了学生模型中的显著性漂移，这再次显示了我们的方法在 CIL 期间保持显著性的有效性。

任务	1	4	7	10
SSRE	47.4	50.1	56.6	78.5
SSRE+TASS	75.2	82.3	88.5	90.1

表 7. 在 CIFAR-100 10 任务中，DINO 自注意力图与学生模型显著性之间的平均 IoU(%)。

5. 总结

本文提出了一种用于 EFCIL 的任务自适应显著性引导方法。TASS 背后的洞察是引导模型关注于显著区域并抑制显著性漂移。我们的研究表明，强健的显著性引导对于减轻跨任务遗忘至关重要。实验证明，TASS 是有效的，并且超越了最先进的方法。TASS 能够很容易地与其他方法结合使用，从而在基线上获得较大的性能提升。定性结果还表明，低层任务在不同的任务间是稳定的，从而导致更少的遗忘。

局限性与未来工作。 在这项工作中，我们引入了跨任务的辅助显著性知识来提升 CIL 性能，这可能会引起对于不公平比较的关注。然而，我们认为探索外部知识以使 CIL 系统适用于实际应用是很重要的。此外，我

们的方法可以很容易地与其他方法集成，低层辅助知识的成本可以忽略不计，并且易于获得。我们将在未来的工作中探索利用其他形式的知识，例如预训练的视觉和大语言模型。

鸣谢 该项目得到国家自然科学基金（NO. 62225604，62206135）、青年人才托举工程项目（2023QNRC00）以及中央高校基本科研业务费专项资金资助（南开大学，070-63233085）。算力由南开大学超算中心提供。

References

[1] Eden Belouadah, Adrian Popescu, and Ioannis Kanellos. A comprehensive study of class incremental learning algorithms for visual tasks. *Neural Networks*, 135: 38–54, 2021. 1, 2

[2] Francisco M Castro, Manuel J Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. In *ECCV*, 2018. 1, 5, 6

[3] Fabio Cermelli, Massimiliano Mancini, Samuel Rota Bulo, Elisa Ricci, and Barbara Caputo. Modeling the background for incremental learning in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9233–9242, 2020. 2

[4] Arslan Chaudhry, Marc' Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. In *ICLR*, 2019. 2

[5] Hanting Chen, Yunhe Wang, Chang Xu, Zhaohui Yang, Chuanjian Liu, Boxin Shi, Chunjing Xu, Chao Xu, and Qi Tian. Data-free learning of student networks. In *ICCV*, 2019. 2

[6] Ming-Ming Cheng, Shanghua Gao, Ali Borji, Yong-Qiang Tan, Zheng Lin, and Meng Wang. A highly efficient model to study the semantics of salient object detection. *IEEE TPAMI*, 2021. 4, 12

[7] Matthias Delange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Greg Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE TPAMI*, 2021. 2

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 5

[9] Prithviraj Dhar, Rajat Vikram Singh, Kuan-Chuan Peng, Ziyan Wu, and Rama Chellappa. Learning without memorizing. In *CVPR*, 2019. 2

[10] Arthur Douillard, Matthieu Cord, Charles Ollion, Thomas Robert, and Eduardo Valle. Podnet: Pooled outputs distillation for small-tasks incremental learning. In *ECCV*, 2020. 1

[11] Sayna Ebrahimi, Suzanne Petryk, Akash Gokul, William Gan, Joseph E. Gonzalez, Marcus Rohrbach, and trevor darrell. Remembering for the right reasons: Explanations reduce catastrophic forgetting. In *ICLR*, 2021. 2, 5

[12] Deng-Ping Fan, Ge-Peng Ji, Peng Xu, Ming-Ming Cheng, Christos Sakaridis, and Luc Van Gool. Advances in deep concealed scene understanding. *Visual Intelligence*, 1(1):16, 2023. 4

[13] Tao Feng, Mang Wang, and Hangjie Yuan. Overcoming catastrophic forgetting in incremental object detection via elastic response distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9427–9436, 2022. 2

[14] Ian J Goodfellow, Mehdi Mirza, Da Xiao, Aaron Courville, and Yoshua Bengio. An empirical investigation of catastrophic forgetting in gradient-based neural networks. *arXiv preprint arXiv:1312.6211*, 2013. 1

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 5

[16] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *CVPR*, 2019. 5, 6

[17] Aya Abdelsalam Ismail, Hector Corrada Bravo, and Soheil Feizi. Improving deep learning interpretability by saliency guided training. *NeurIPS*, 2021. 4

[18] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 2017. 2, 5

[19] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Technical report, 2009. 5

[20] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015. 5

[21] Hao Li, Dingwen Zhang, Nian Liu, Lechao Cheng, Yalun Dai, Chao Zhang, Xinggang Wang, and Junwei Han. Boosting low-data instance segmentation by unsupervised pre-training with saliency prompt. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15485–15494, 2023. 4

[22] Xiangtai Li, Ansheng You, Zhen Zhu, Houlong Zhao, Maoke Yang, Kuiyuan Yang, and Yunhai Tong. Semantic flow for fast and accurate scene parsing. In *ECCV*, 2020. 4, 5

[23] Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille. The secrets of salient object segmentation. In *CVPR*, 2014. 12

[24] Zhizhong Li and Derek Hoiem. Learning without forgetting. In *ECCV*, 2016. 2

[25] Liqiang Lin, Pengdi Huang, Chi-Wing Fu, Kai Xu, Hao Zhang, and Hui Huang. On learning the right attention point for feature enhancement. *Science China Information Sciences*, 66(1):112107, 2023. 4

[26] Zheng Lin, Zhao Zhang, Zi-Yue Zhu, Deng-Ping Fan, and Xia-Lei Liu. Sequential interactive image segmentation. *Computational Visual Media*, 9(4):753–765, 2023. 2

[27] Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang. A simple pooling-based design for real-time salient object detection. In *CVPR*, 2019. 12

[28] Jiang-Jiang Liu, Qibin Hou, and Ming-Ming Cheng. Dynamic feature integration for simultaneous detection of salient object, edge and skeleton. *IEEE TIP*, 2020. 12

[29] Xialei Liu, Marc Masana, Luis Herranz, Joost Van de Weijer, Antonio M Lopez, and Andrew D Bagdanov. Rotate your networks: Better weight consolidation and less catastrophic forgetting. In *ICPR*, 2018. 2

[30] Yu Liu, Sarah Parisot, Gregory Slabaugh, Xu Jia, Ales Leonardis, and Tinne Tuytelaars. More classifiers, less forgetting: A generic multi-classifier paradigm for incremental learning. In *ECCV*. Springer, 2020. 2, 5

[31] Yuyang Liu, Yang Cong, Dipam Goswami, Xialei Liu, and Joost van de Weijer. Augmented box replay: Overcoming foreground shift for incremental object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11367–11377, 2023. 2

[32] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *NeurIPS*, 2017. 2

[33] Zhouzhou Ma, Guanghua Gu, and Wenrui Zhao. Self-attention guidance based crowd localization and counting. *Machine Intelligence Research*, pages 1–17, 2024. 4

[34] Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *CVPR*, 2018. 2

[35] Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D Bagdanov, and Joost van de Weijer. Class-incremental learning: survey and performance evaluation on image classification. *IEEE TPAMI*, 2022. 1

[36] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, pages 109–165. Elsevier, 1989. 1

[37] Grégoire Petit, Adrian Popescu, Hugo Schindler, David Picard, and Bertrand Delezoide. Fetril: Feature translation for exemplar-free class-incremental learning. In *WACV*, 2023. 8

[38] Quang Pham, Chenghao Liu, and Steven Hoi. Dualnet: Continual learning, fast and slow. *NeurIPS*, 2021. 1

[39] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *CVPR*, 2017. 1, 2, 5, 6

[40] Matthew Riemer, Ignacio Cases, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro. Learning to learn without forgetting by maximizing transfer and minimizing interference. In *ICLR*, 2019. 2

[41] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *ICCV*, 2017. 4

[42] Wuxuan Shi and Mang Ye. Prototype reminiscence and augmented asymmetric knowledge aggregation for non-exemplar class-incremental learning. In *ICCV*, 2023. 8

[43] James Smith, Yen-Chang Hsu, Jonathan Balloch, Yilin Shen, Hongxia Jin, and Zsolt Kira. Always be dreaming: A new approach for data-free class-incremental learning. In *ICCV*, 2021. 2

[44] Yue Wu, Yinpeng Chen, Lijuan Wang, Yuancheng Ye, Zicheng Liu, Yandong Guo, and Yun Fu. Large scale incremental learning. In *CVPR*, 2019. 1

[45] Jia-Wen Xiao, Chang-Bin Zhang, Jiekang Feng, Xialei Liu, Joost van de Weijer, and Ming-Ming Cheng. Endpoints weight fusion for class incremental semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7204–7213, 2023. 2

[46] Ju Xu and Zhanxing Zhu. Reinforced continual learning. *NeurIPS*, 2018. 2

[47] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In *CVPR*, 2013. 12

[48] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013. 12

[49] Hongxu Yin, Pavlo Molchanov, Jose M Alvarez, Zhizhong Li, Arun Mallya, Derek Hoiem, Niraj K Jha, and Jan Kautz. Dreaming to distill: Data-free knowledge transfer via deepinversion. In *CVPR*, 2020. 1, 2

[50] Lu Yu, Bartlomiej Twardowski, Xialei Liu, Luis Herranz, Kai Wang, Yongmei Cheng, Shangling Jui, and Joost van de Weijer. Semantic drift compensation for class-incremental learning. In *CVPR*, 2020. 1, 2

[51] Lu Yu, Xialei Liu, and Joost Van de Weijer. Self-training for class-incremental semantic segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 2

[52] Jiang-Tian Zhai, Xialei Liu, Andrew D Bagdanov, Ke Li, and Ming-Ming Cheng. Masked autoencoders are efficient class incremental learners. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19104–19113, 2023. 2

[53] Junting Zhang, Jie Zhang, Shalini Ghosh, Dawei Li, Serafettin Tasci, Larry Heck, Heming Zhang, and C-C Jay Kuo. Class-incremental learning via deep model consolidation. In *WACV*, 2020. 2

[54] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *ICCV*, 2019. 4

[55] Fei Zhu, Zhen Cheng, Xu-yao Zhang, and Cheng-lin Liu. Class-incremental learning via dual augmentation. *NeurIPS*, 2021. 1, 2, 3, 5, 7

[56] Fei Zhu, Xu-Yao Zhang, Chuang Wang, Fei Yin, and Cheng-Lin Liu. Prototype augmentation and self-supervision for incremental learning. In *CVPR*, 2021. 1, 2, 3, 5, 7, 12

[57] Kai Zhu, Wei Zhai, Yang Cao, Jiebo Luo, and Zheng-Jun Zha. Self-sustaining representation expansion for non-exemplar class-incremental learning. In *CVPR*, 2022. 2, 3, 5, 7

[58] Kai Zhu, Kecheng Zheng, Ruili Feng, Deli Zhao, Yang Cao, and Zheng-Jun Zha. Self-organizing pathway expansion for non-exemplar class-incremental learning. In *ICCV*, 2023. 8

方法	5 任务	10 任务	20 任务
baseline (SSRE)	40.2	40.0	39.3
CAM	41.2	40.7	40.4
SmoothGrad	42.1	41.0	40.4
Grad-CAM	44.1	43.9	43.5

表 8. 在 Tiny-Imagenet 上生成显著性图方法的消融实验。

模型	参数量 (M)	FLOPS(G)
Ours	17.9	0.78
预训练显著性模型	0.0941	0.012

表 9. 预训练显著性模型的参数量和 FLOPs。FLOPs 的计算使用 $3 \times 32 \times 32$ 的图像。

A. 进一步的消融实验

显著性方法的消融实验。 为了展示 TASS 的泛化能力，我们使用了几种方法来计算显著性图，并在表 8 中报告了结果。尽管其他方法也带来了性能提升，Grad-CAM 的表现最佳，这证明了 TASS 的有效性。

针对低层目标图的消融实验。 在论文中，我们使用了 CSNet [6] 计算所有预训练的显著性图和边界图，因为它非常轻量化。与我们的主模型相比，预训练模型的参数量不到 1%，所需的浮点运算数（FLOPs）仅为 1.5%（如表 9 所示）。注意，我们在新的任务开始前离线计算所有低层图，因此额外的 FLOPs 应该分摊到各个训练轮次中。因此，低层模型所需的额外 FLOPs 仅占主模型的约 0.015%，在实际应用中可以忽略不计。

为了展示 TASS 的有效性，我们对低层图进行了消融实验。我们用 ResNet-152 网络生成的 Grad-CAM 替换了这些低层图。为了避免信息泄露，ResNet-152 是从头开始训练的。在每个新任务之前，我们首先只在任务数据上训练 ResNet-152，并使用 Grad-CAM 输出以监督增量模型中的显著性。根据表 10 的结果，我们可以看到 TASS 依然优于其他方法。此外，TASS 也适用于其他用于生成显著性图的模型（例如 DFI [28] 或 PoolNet [27]），并且在参数更多、更大的网络上表现更好。

低层信息源 (方法)	准确率 (%)
PASS	39.3
SSRE	40.0
ResNet152 (Ours)	42.1
CSNet (Ours)	43.9
PoolNet (Ours)	44.2
DFI (Ours)	44.4

表 10. 在 10 任务 Tiny-ImageNet 上针对低层显著性图的消融实验。

方法	参数量 (M)	准确率 (%)
PASS-Res18	14.5	50.4
PASS-Res32	21.7	51.2
SSRE-Res18	19.4	58.7
Ours-Res18	17.9	61.5

表 11. 不同方法网络架构的比较。Method-Res18 表示使用 ResNet18 作为主干的方法。

针对方法架构与显著性模型预训练的消融实验。 我们选择 PASS [56] 作为基线方法来应用 TASS（如表 11和表 12所示）。这两个表中的实验是在 ImageNet-Subset 上进行的，共有 5 个任务。由于有些方法使用了 ImageNet 预训练权重来更好地估计显著性图，因此我们在该数据集上从头开始训练了 CSNet [6]（有预训练和无预训练）用于显著目标检测 [23, 47, 48]。这使得我们能够证实，由于在 ImageNet 上对显著性网络进行预训练，信息泄漏不会发生。低层网络在没有预训练的情况下，效果几乎与在 ImageNet 上预训练显著性网络几乎一样好。我们还在表 11中比较了不同方法的参数量。结果表明，增加 PASS 的网络容量（从 ResNet-18 增加到参数更多的 ResNet-32）仅能带来微小的性能提升。我们基于 PASS 的方法使用 ResNet-18 实现了显著的性能提升，超越了参数更多的 SSRE。

我们的方法在非 DFCIL 场景中的应用。 我们在使用 PASS 的非 DFCIL 场景中应用了我们的显著性监督，每个类别包含 20 个样本，结果如表 13所示。TASS 在这里也能显著提升性能。

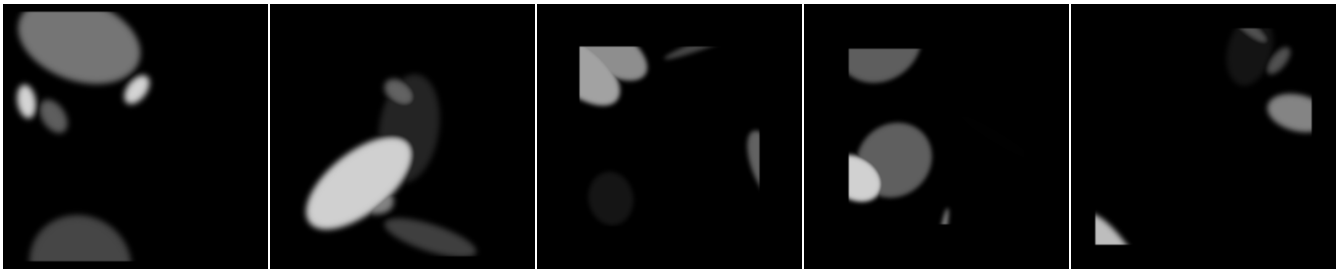

图 6. 一些生成的显著性噪声图的可视化

方法	准确率 (%)
无预训练	61.5
预训练显著性检测模型	62.0

表 12. 针对显著性检测网络预训练的消融实验。

方法	缓存大小	准确率 (%)
PASS	20	52.36
PASS+TASS	20	**55.75**

表 13. 非 DFCIL 场景中的 TASS。

多损失的超参数。 在公式 5 中，我们为所有损失项赋予相同的权重。根据建议，我们在表 14中探索了更多的选项。微调能略微提升性能，但为了方便，我们仍然保持 $\lambda_{\text{CIL}} = \lambda_{\text{lm}} = \lambda_{\text{dbs}} = 1.0$。公式 2 中的 $\sqrt{N}$，其中 N 是像素数量，用于归一化 L2 距离。

λ_{CIL}	λ_{lm}	λ_{dbs}	准确率 (%)
1	1	1	55.01
0.1	1	1	**55.27**
1	0.1	1	54.22
1	1	0.1	54.31

表 14. SSRE+TASS 的多损失超参数。

方法		L	D	S	LD	LS	DS	LDS
PASS	49.0	51.2	50.6	51.4	53.0	52.6	53.7	54.5
SSRE	55.0	56.2	55.8	56.7	57.3	57.0	57.6	57.9

表 15. LDS 分别代表低层多任务监督、膨胀边界监督和显著性噪声注入。

所有损失的排列消融。 我们在表 15中列出了全部三种损失项的所有可能组合。这些结果表明，每个组成部分都对最终性能有所贡献，并且它们的组合效果最佳。

F & 准确率 (%)	50	30	10	0
PASS	49.03±0.9	46.78±0.9	44.65±1.0	40.27±1.0
PASS+TASS	54.45±0.4	51.22±0.5	48.58±0.5	44.30±0.5

表 16. 针对 F 进行消融实验，其中 $T = 10$。

实验协议的类别划分。 我们遵循先前工作如 PASS 和 SSRE 的常规实验设置，将数据集的类别划分为 $F + C \times T$，其中 $F = 50$。根据建议，我们在表 16 中评估了不同的 F 选项。TASS 在所有设置下都表现出相对于基线的一致提升。

B. 更多对 TASS 的可视化

显著性噪声。 每个椭圆有 6 个维度：中心坐标 (x, y)、旋转角度 α、掩码权重 w，以及长轴和短轴 (a, b)。x、y、α 和 w 从以下范围的均匀分布中采样的：$x \in [0, H)$、$y \in [0, W)$、$\alpha \in [0, 2\pi)$、$w \in [0, 1]$。H 和 W 分别表示输入图像的高度和宽度。为了生成适当大小的椭圆，我们从高斯分布中采样长轴和短轴，其中 $\mu_a = \max(H, W)/2$，$\sigma_a = \max(H, W)/6$，$\mu_b = \min(H, W)/2$，$\sigma_b = \min(H, W)/6$。采样得到的 a 和 b 分别被限制在 $[0, \max(H, W)/2]$ 和 $[0, \min(H, W)/2]$ 范围内。对于每个椭圆，我们创建一个显著性图 S_i。我们重复这个随机生成过程 3-5 次，并对 S_i 进行逐元素的最大值操作，以获得一个单一的显著性图 S。然后，我们将 S 裁剪和调整为原始图像的大小，裁剪尺寸从 $[\min(H, W)/2, \min(H, W)]$ 的均匀分布中采样，引入中心感知的显著性噪声用于网络训练。最后，我们对 S 应用高斯模糊，以更好地模拟现实的显著性图。高斯模糊的核大小是最接近 $\min(H, W)/20$ 的奇数。对于每个编码器特征图，10% 的随机选定通道会直接用 S 进行掩码，其中每个选定的通道将有一个独立的 S。我们在

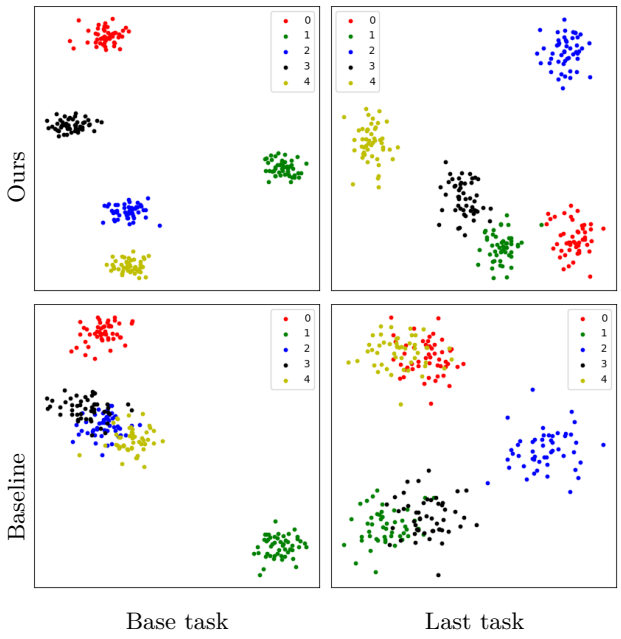

图 7. 使用和不使用 TASS 的嵌入 $F_\phi(x)$ 的可视化。与基线方法相比，我们的方法保留了更多具有判别性的表示。

图 6 中可视化了几个生成的样本。

嵌入可视化。 由于我们的方法帮助模型集中于前景，因此更多的类别特定像素会对嵌入产生贡献。因此，嵌入更加具有判别性，且背景信息的干扰较少。在图 7 中，我们使用 t-SNE 可视化在 ImageNet-Subset 上学习了基础任务和最后任务后的前五个类别的嵌入。在基础任务中，基线方法（SSRE）和我们的方法（SSRE+TASS）表现都很好。在最后的任务之后，显而易见，TASS 有助于在任务之间保持判别性特征，而基线方法的嵌入则出现了重叠。