ECLIPSE: Efficient Continual Learning in Panoptic Segmentation with Visual Prompt Tuning

摘要

全景分割结合了语义和立场分割，是一项前沿的计算机视觉任务。尽管最近深度学习模型取得了进展，但现实世界应用的动态性要求持续学习，其中模型随着时间的推移自适应新类别（可塑性），而不会忘记旧类别（灾难性遗忘）。目前的连续分割方法通常依赖于蒸馏策略，如知识蒸馏和伪标记，这些策略是有效的，但会导致训练复杂度和计算开销的增加。本文提出一种新的、高效的基于视觉提示调优的连续全景分割方法 ECLIPSE。所提出方法包括冻结基本模型参数和只微调一小部分提示嵌入，解决了灾难性的获取和可塑性，并显著减少了可训练参数。为减轻持续分割中的错误传播和语义漂移等固有挑战，本文提出 logit 操纵，以有效利用各类的公共知识。在 ADE20K 连续全景分割基准数据集上的实验证明了 ECLIPSE 的优越性，特别是其对灾难性遗忘的鲁棒性和合理的可塑性，达到了新的先进水平。

关键词：全景分割；语义分割；视觉参数调整；持续学习

1 引言

图像分割是一项基本的计算机视觉任务，它涉及将图像分割为有意义的部分，以方便分析。全景分割是最先进的图像分割形式之一，它将语义分割（将像素分类为集合类别）与实例分割（在类别中识别单个对象）相结合。最近的全景分割研究取得了重大进展，尽管有这些进展，但现实世界的动态性质要求模型不仅要了解当前，而且要随着时间的推移而发展。连续图像分割解决了这一需求，使模型能够随着时间的推移逐步学习新类别，而不会忘记旧类别。如何在保持原有类别知识（避免获取[8]带来灾难性后果）的同时，高效地融合新的类别信息（可塑性），是一个极具挑战性的问题。

最近，各种连续分割方法[1,3,7,14,15,16,21,22,25]出现了，解决了关键挑战，并显示出显著的改进。大多数连续分割方法通常采用蒸馏策略，如知识蒸馏[1,7 ,22]和伪标记[3,7,15]。知识蒸馏可以通过将知识从旧模型迁移到新模型来缓解灾难性遗忘，而伪标记允许新模型使用先前学习的类的标签进行训练。尽管具有开创性，但这些方法涉及权衡，如需要双重网络转发和仔细调整超参数（例如，蒸馏损失权重和伪标签的阈值），这增加了训练的复杂性和计算开销。随着类的数量逐渐增加，维护可扩展和高效的蒸馏过程可能会变得具有挑战性。

本文提出一种新方法 ECLIPSE，用于全景分割中的高效持续学习，利用视觉提示调优（VPT）[9]的潜力，并免去了对传统蒸馏策略的需要。我们的方法从冻结基本模型的所有参数开始，并随着新类的出现反复微调一组新的提示嵌入。通过模型冻结解决了灾难性遗忘问题，并通过提示调

整提高了可塑性。所提出方法是第一个不需要蒸馏的连续全景分割，大大减少了可训练的参数，并简化了连续分割过程。尽管有这些优势，但在持续全景分割中面临着固有的挑战，需要进一步改进。虽然模型冻结保留了先验知识，但会同时传播先验误差。此外，在推理中区分输出掩码是否为非对象（no-obj）所需的非对象类的定义在每个持续学习步骤中都会发生变化，这被称为语义漂移问题。为了规避这些挑战，本文提出了一种简单有效的策略，称为对数操纵（Logit Manipulation）。它允许模型利用所有学习的类的类间知识，以更有意义地操作非对象对数。动态更新的非对象对数有助于抑制先验错误预测，并立即缓解语义漂移问题。

与之前的工作相比，ECLIPSE 对灾难性遗忘表现出出色的鲁棒性，特别是当连续步骤数量增加时，这比现有方法有了实质性的改进。同时，ECLIPSE 成功地将 VPT 集成到连续的全景分割中，有效缓解了灾难性遗忘，并有效扩展了模型的可扩展性。提出一种有效的对数操纵策略，消除了连续全景分割中的固有挑战：错误传播和语义漂移。

2 相关工作

2.1 全景分割

全景分割（Panoptic Segmentation）是计算机视觉中的前沿任务，结合了语义分割和实例分割的概念，旨在全面解析场景中的“物质”和“物体”。一些开创性研究（如文献 [4, 10, 11]）首次将语义分割与实例分割整合到统一框架中。随后，一些方法（如文献 [13, 17, 23]）通过在全卷积范式中引入动态卷积，显著提升了分割效果。最近，基于 Transformer 的架构（如文献 [5, 6, 12]）进一步推动了该领域的发展，利用注意力机制展现了强大的特征建模能力。然而，如何在不遗忘旧类别的同时动态适应新类别，仍然是连续全景分割领域的研究难题。

2.2 持续性分割

为了应对实际应用中不断变化的动态场景，连续分割成为一项重要的高级任务。一些开创性研究（如文献 [1]）揭示了连续分割中的独特挑战——语义漂移（Semantic Drift），其主要由背景类别引发。大多数方法（如文献 [1, 3, 7, 14, 15, 22, 25]）通过知识蒸馏和伪标签等蒸馏策略来缓解语义漂移问题。最近，Incrementer [16] 利用了基于 Transformer 模型的架构优势，通过增量类嵌入和多种蒸馏策略进一步优化。然而，大多数研究集中在连续语义分割领域，而更具挑战性的连续全景分割研究较少。CoMFormer [2] 是连续全景分割领域的开创性工作，其基于通用分割模型（即 Mask2Former [6]），通过查询机制的蒸馏策略同时执行全景分割和语义分割任务。然而，这类基于蒸馏的方法增加了训练复杂性和计算开销，并且需要对超参数（如损失权重、蒸馏温度以及伪标签阈值）进行精细调试。

作者的方法是首个适用于全景分割和语义分割的无蒸馏方法，大幅简化了连续学习过程并减少了训练计算量。

2.3 视觉提示调整

VPT [9] 提出了一种高效且有效的视觉 Transformer 模型微调方法。通过冻结预训练参数，仅微调一组可学习的提示，便能取得显著的性能提升。在连续图像分类领域，已有多个尝试将 VPT 应用于实践。例如，L2P [20] 和 DualPrompt [19] 通过冻结预训练模型，并利用键值机制从提示池中选择最相关的提示，取得了显著的性能表现，展示了在连续图像分类中应用 VPT 的潜力。

作者提出了首个基于 VPT 的连续分割方法，专门针对连续全景分割中的多个独特挑战进行了优化设计。

3 本文方法

3.1 基础网络架构

这篇论文提出了一种名为 ECLIPSE 的方法，用于在连续学习环境中进行高效的全景分割。ECLIPSE 基于 Mask2Former 架构，这是一种基于 Transformer 的通用分割模型，能够处理全景、实例和语义分割任务。

在文章当中作者使用 Mask2Former 作为基础网络框架。Mask2Former 作为一种使用 transformer 作为基础的通用分割模型，可以实现多种任务的分割，包括全景、实例和语义分割。而与之前的网络框架不同，Mask2Former 直接预测一系列掩码以及它们所属的类别。Mask2Former 包括三个主要的组件：图像编码器$\mathcal{M}_{enc}$、像素解码器$\mathcal{M}_{pixel}$和 transformer 解码器$\mathcal{M}_{trans}$，分别用于提取对应的图像嵌入$\mathcal{E}_{img}$、像素嵌入$\mathcal{E}_{pixel}$和对应的提示的掩码嵌入。在掩码嵌入生成之后，还会再加上一层 MLP 层用于将每个掩码对应的类别进行划分。最后，之前生成的像素嵌入$\mathcal{E}_{pixel}$和经过 MLP 层的掩码嵌入进行点积操作就能够得到最终的掩码图像。

在以上的操作当中，$\mathcal{M}$代表整体框架函数。它使用$\mathrm{x} \in \mathbb{R}^{3\times H\times W}$和$\mathrm{Q} \in \mathbb{R}^{N\times D}$作为输入，并且输出$\mathrm{m} \in \mathbb{R}^{N\times D\times W}$和$\mathrm{s} \in \mathbb{R}^{N\times C}$。其中 x 和 Q 分别表示图像和提示，m 和 s 分别表示输出的掩码和对应的类别分数。即：

$$\mathcal{E}_{img} = \mathcal{M}_{enc}(x), \mathcal{E}_{pixel} = \mathcal{M}_{pixel}(\mathcal{E}_{img}),$$

$$\mathcal{M}(\mathrm{x}, \mathrm{Q}) = \mathrm{MLP}\left(\mathcal{M}_{trans}(\mathcal{E}_{img}, \mathrm{Q})\right) \otimes \mathcal{E}_{pixel,}$$

$$(\mathrm{s}, \mathrm{m}) = \mathcal{M}(\mathrm{x}, \mathrm{Q})$$

3.2 使用提示调整进行持续分割

在 ECLIPSE 当中抛弃了传统持续学习当中使用知识蒸馏的方法，转而使用 VPT[9]策略进行持续性分割。文章当中的方法首先在基础类C^1上进行初始化训练模型（t=1），此时会对模型当中的所有参数进行更新：

$$(s^1, m^1) = \mathcal{M}(x, Q^1)$$

在初始化训练之后，反复应用冻结-调优策略。也就是说，当新类C^t到来（t>1）时，在训练时会冻结所有训练参数以保存之前的知识，并且只调整一小部分的可学习的提示嵌入$Q^t \in \mathbb{R}^{N^t\times D}$参数以及不共享的 MLP 层参数。即：

$$\mathcal{M}(\mathrm{x}, \mathrm{Q}^t) = \mathrm{MLP}^t\left(\mathcal{M}_{trans}(\mathcal{E}_{img}, \mathrm{Q}^t)\right) \otimes \mathcal{E}_{pixel,}$$

$$(s^t, m^t) = \mathcal{M}(x, Q^t)(t > 1)$$

在这些步骤当中，不对$\mathcal{M}_{enc}$、$\mathcal{M}_{pixel}$和$\mathcal{M}_{trans}$当中的参数进行更新，并且设置提示嵌入Q^t作为一个离散的特定任务模块，仅用于新类C^t的识别。随着步骤的不断推进，通过轻量级的特定任务提示集稳定地扩展了模型的可扩展性。如图 1 所示：

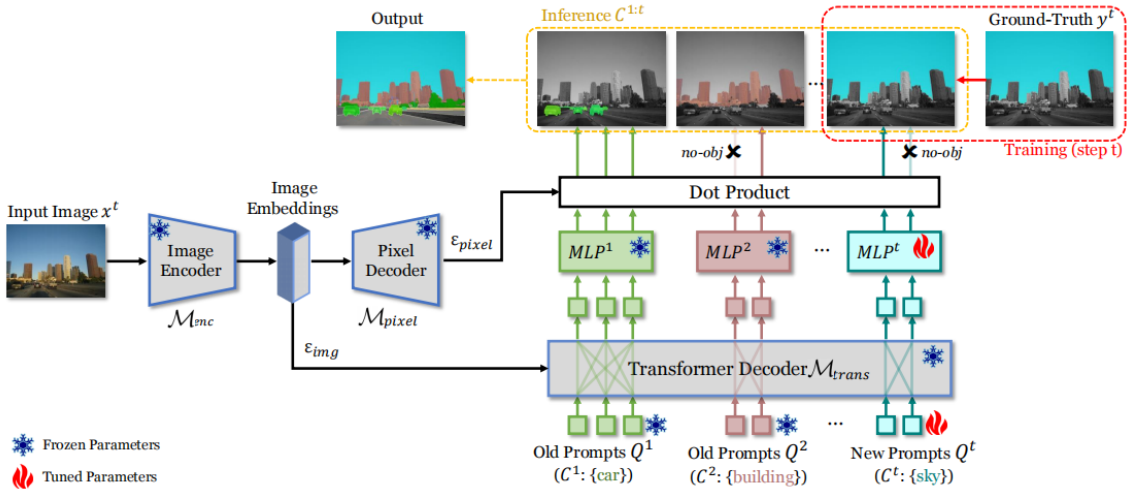

图 1. ECLIPSE 方法概述。文章当中的方法会冻结所有已训练的参数，仅微调一组提示嵌入Q^t和 MLP 层，用于识别当前任务的类别集合。在推理阶段，我们聚合所有提示集合$Q^{1:t}$的输出，以分割已经学习的全部类别$C^{1:t}$。其中红色火焰表示参数调整，蓝色雪花表示参数冻结。

3.3 解决语义混淆和漂移

为了语义混淆和漂移的问题，文章提出了一种简单但高效的方法——对数操作。首先利用旧类别与新类别之间的互信息，生成新的 no-obj logit，使其包含更有意义的信息。例如，如图 2 所示，提示Q^2的解码器输出被输入到其他的MLP^1和MLP^3层，随后通过聚合这些 MLP 层的对数操作生成新的非对象分数。在我们的方法中，由于Q^t的输出仅负责预测C^t类别的结果，因此其他不属于C^t的类别的对数可以被视为非对象类，从而更好地解决了语义混淆和语义漂移的问题。对数操作的具体过程为：

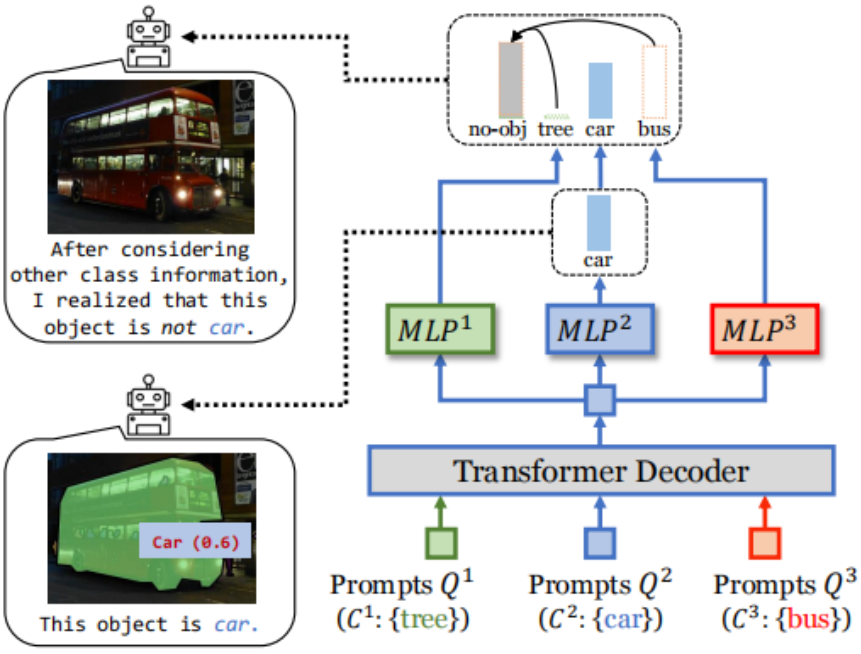

图 2. 为缓解非对象类的语义漂移问题，我们利用所有已学习类别之间的类间知识生成新的非对象分数。此外，通过对数操作，还可以修正由冻结的先前参数导致的语义混淆错误预测。

$$s_t^{C^{1:T}} = \mathrm{MLP}^{1:T}(Q_t),$$

$$s_t^{no-obj} = \delta \times \left(\sum_{k=1}^{t-1} s_t^{C^k} + \sum_{k=t+1}^{T} s_t^{C^k} \right),$$

$$c_t = argmax(s_t^{no-obj}, s_t^{C^t})$$

4 复现细节

4.1 与已有开源代码对比

无太明显的改进，主要是使用自己的数据集进行训练，在自己的数据集上得到分割结果的分数。

4.2 实验环境搭建

操作系统使用 linux 操作系统，Ubuntu 18.04。Python 版本为 3.8.20，cuda 版本为 10.2，pytorch 版本为 1.12.1。同时在 python 当中安装了 cython、scipy、shapely、timm、h5py、submitit、scikit-image 和 continuum 库。使用的显卡型号为 NVIDIA 2080ti，数量为 2。

5 复现实验

实验分别在两个数据集上进行，分别为原文章使用的数据集 ADE20K 和自己的医学图像数据集 CoNSeP。

5.1 文章原数据集实验复现

实验分别在两个不同的分割任务下进行，包括全景分割和语义分割。

表 1. 在不同的持续学习设置下，模型训练之后得到的结果。PQ 为评价指标

Task	Base-PQ	Novel-PQ	All-PQ
100-5 Panoptic Segmentation	40.8295	16.0238	32.5609
50-10 Panoptic Segmentation	45.9005	17.1521	28.7349

在作者使用的原始数据集当中，可以看出：无论是在持续学习设置为 100-5（预先学习 100 个类，然后分五步进行持续学习，每次学习 10 个类）或者是在设置 50-10（预先学习 50 个类，然后分五步进行持续学习，每次学习 10 个类）的情况之下，模型训练之后的分割效果都能够达到文章实验当中所呈现出来的分数。如图 3 所示。

Method	Backbone	Trainable Params	KD	**100-5** (11 tasks)			**100-10** (6 tasks)			**100-50** (2 tasks)		
				1-100	*101-150*	*all*	*1-100*	*101-150*	*all*	*1-100*	*101-150*	*all*
FT	R50	44.9M		0.0	25.8	8.6	0.0	2.9	1.0	0.0	1.3	0.4
MiB [2]	R50	44.9M	✓	24.0	6.5	18.1	27.1	10.0	21.4	35.1	19.3	29.8
PLOP [10]	R50	44.9M	✓	28.1	15.7	24.0	30.5	17.5	26.1	40.2	22.4	34.3
CoMFormer [4]	R50	44.9M	✓	34.4	15.9	28.2	36.0	17.1	29.7	40.2	23.5	34.6
ECLIPSE	R50	**0.60M**		**41.1**	**16.6**	**32.9**	**41.4**	**18.8**	**33.9**	**41.7**	**23.5**	**35.6**
	Swin-L	0.60M		48.0	20.6	38.9	48.6	25.5	40.9	48.2	29.8	42.0
joint	R50			43.2	32.1	39.5	43.2	32.1	39.5	43.2	32.1	39.5

(a)

Method	Backbone	Trainable Params	KD	**50-10** (11 tasks)			**50-20** (6 tasks)			**50-50** (3 tasks)		
				1-50	*51-150*	*all*	*1-50*	*51-150*	*all*	*1-50*	*51-150*	*all*
FT	R50	44.9M		0.0	1.7	1.1	0.0	4.4	2.9	0.0	12.0	8.1
MiB [2]	R50	44.9M	✓	34.9	7.7	16.8	38.8	10.9	20.2	42.4	15.5	24.4
PLOP [10]	R50	44.9M	✓	39.9	15.0	23.3	43.9	16.2	25.4	45.8	18.7	27.7
CoMFormer [4]	R50	44.9M	✓	38.5	15.6	23.2	42.7	17.2	25.7	45.0	19.3	27.9
ECLIPSE	R50	**0.60M**		**45.9**	**17.3**	**26.8**	**46.4**	**19.6**	**28.6**	**46.0**	**20.7**	**29.2**
	Swin-L	0.60M		52.8	22.9	32.9	53.2	25.7	34.8	53.0	25.3	34.5
joint	R50			50.2	34.1	39.5	50.2	34.1	39.5	50.2	34.1	39.5

(b)

图 3. 文章当中描述的在持续学习设置分别为 100-5 和 50-10 下的全景分割结果

之后又在语义分割领域设置了实验：

表 2. 在持续学习设置 100-5 下，模型训练之后得到的结果。mIoU 为评价指标

Task	**Base-mIoU**	**Novel-mIoU**	**All-mIoU**
100-5 Semantic Segmentation	43.467	18.0978	38.8859

同样可以看出：在持续学习设置为 100-5 的情况之下，模型训练之后的分割效果也能够达到

Method	Backbone	Trainable Params	KD	**100-5** (11 tasks)			**100-10** (6 tasks)			**100-50** (2 tasks)		
				1-100	*101-150*	*all*	*1-100*	*101-150*	*all*	*1-100*	*101-150*	*all*
SDR† [34]	R101	60.4M	✓	-	-	-	28.9	7.4	21.7	37.4	24.8	33.2
UCD† [44]	R101	60.4M	✓	-	-	-	40.8	15.2	32.3	42.1	15.8	33.3
SPPA† [30]	R101	60.4M	✓	-	-	-	41.0	12.5	31.5	42.9	19.9	35.2
RCIL† [45]	R101	58.0M	✓	38.5	11.5	29.6	39.3	17.6	32.1	42.3	18.8	34.5
SSUL† [5]	R101	1.78M	✓	39.9	17.4	32.5	40.2	18.8	33.1	41.3	18.0	33.6
REMINDER† [35]	R101	60.4M	✓	-	-	-	39.0	21.3	33.1	41.6	19.2	34.1
MiB [2]	R101	63.4M	✓	21.0	6.1	16.1	23.5	10.6	26.6	37.0	24.1	32.6
PLOP [10]	R101	63.4M	✓	33.6	14.1	27.1	34.8	15.9	28.5	43.4	25.7	37.4
CoMFormer [4]	R101	63.4M	✓	39.5	13.6	30.9	40.6	15.6	32.3	43.6	26.1	37.6
ECLIPSE	R101	**0.60M**		**43.3**	**16.3**	**34.2**	**43.4**	**17.4**	**34.6**	**45.0**	**21.7**	**37.1**
joint	R101			46.9	35.6	43.1	46.9	35.6	43.1	46.9	35.6	43.1

文章实验当中所呈现出来的分数。如图 4 所示。

图 4. 文章当中描述的在持续学习设置分别为 100-5 语义分割结果

5.2 使用医学图像数据集实验

在使用医学图像时，主要在数据集 CoNSeP 上进行语义分割实验，其中的持续学习设置分别为 1-1（预先学习 1 个类，然后分两步进行持续学习，每次学习 1 个类）、1-2（预先学习 1 个类，然后分 1 步进行持续学习，每次学习 2 个类）和 2-1（预先学习 2 个类，然后分 1 步进行持续学习，每次学习 1 个类）。最终的分割结果如表 3 所示。

表 3. 在不同的持续学习设置下，模型训练之后得到的结果。mIoU 为评价指标

Task	Base-mIoU	Novel-mIoU	All-mIoU
1-1 Semantic Segmentation	91.9741	61.6415	53.8143
1-2 Semantic Segmentation	91.8086	54.2998	50.1021
2-1 Semantic Segmentation	91.942	71.3054	63.7974

可以看出，模型在第一阶段的训练（即基础类别的学习）过程当中，能够得到较好的训练结果，当不断引入新的类别进行持续学习时，模型的分割效果有很大程度的降低。最终的分割结果还是强差人意。今后的过程中希望可以进行优化。

6 总结与展望

本文通过复现 ECLIPSE：Efficient Continual Learning in Panoptic Segmentation with Visual Prompt Tuning 的研究，验证了该方法在连续学习环境下对全景分割任务的有效性和优越性。ECLIPSE 方法通过冻结基础模型参数和微调一小部分提示嵌入的方式，显著减少了可训练参数，简化了训练过程，同时在保持对旧类别知识的记忆的同时，有效地学习新类别。实验结果表明，ECLIPSE 在 ADE20K 数据集上取得了新的最先进性能，尤其在灾难性遗忘的鲁棒性和新类别的可塑性方面表现出色。

在复现过程中，我们不仅在原文使用的数据集 ADE20K 上进行了实验，还在自己的医学图像数据集 CoNSeP 上进行了测试。在 ADE20K 数据集上的实验结果与原文报告的结果一致，证明了 ECLIPSE 方法的有效性。然而，在 CoNSeP 数据集上的实验结果显示，尽管模型在基础类别学习阶段表现良好，但在连续学习新类别时性能下降，这表明 ECLIPSE 方法在特定领域数据上可能需要进一步的调整和优化。

展望未来，我们认为 ECLIPSE 方法的研究和应用前景广阔。首先，随着类别数量的增加，如何优化因扩展提示集而导致的计算复杂性增加是一个值得研究的方向。其次，ECLIPSE 方法在不同领域数据集上的表现差异提示我们，模型可能需要针对特定领域的数据进行调整，以提高其泛化能力和适应性。此外，考虑到现实世界中类别的动态变化，研究如何使模型更加灵活地适应类别的增加、删除或变化也是一个重要的研究方向。

总之，ECLIPSE 方法在连续学习领域的全景分割任务中展现了巨大的潜力，但仍有许多挑战需要克服。我们期待未来的研究能够在理论和实践上进一步推动这一领域的发展，为计算机视觉任务提供更加高效和鲁棒的解决方案。

参考文献

[1] 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, 5, 6

[2] Fabio Cermelli, Matthieu Cord, and Arthur Douillard. Comformer: Continual learning in semantic and panoptic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3010–3020, 2023. 2, 3, 5, 6, 7, 8

[3] Sungmin Cha, YoungJoon Yoo, Taesup Moon, et al. Ssul: Semantic segmentation with unknown label for exemplarbased class-incremental learning. Advances in neural information processing systems, 34:10919–10930, 2021. 2, 4, 5, 6, 7

[4] Bowen Cheng, Maxwell D Collins, Yukun Zhu, Ting Liu, Thomas S Huang, Hartwig Adam, and Liang-Chieh Chen. Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 12475–12485, 2020. 1, 2

[5] Bowen Cheng, Alex Schwing, and Alexander Kirillov. Perpixel classification is not all you need for semantic segmentation. Advances in Neural Information Processing Systems, 34:17864–17875, 2021. 1, 2

[6] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 1290–1299, 2022. 1, 2, 3, 5, 6

[7] Arthur Douillard, Yifu Chen, Arnaud Dapogny, and Matthieu Cord. Plop: Learning without forgetting for continual semantic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 4040–4050, 2021. 2, 4, 5, 6

[8] Robert M French. Catastrophic forgetting in connectionist networks. Trends in cognitive sciences, 3(4):128–135, 1999. 1

[9] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In European Conference on Computer Vision, pages 709–727. Springer, 2022. 2, 3, 4

[10] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollar. Panoptic feature pyramid networks. In ´ Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 6399–6408, 2019. 1, 2

[11] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollar. Panoptic segmentation. In ´ Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 9404–9413, 2019. 2

[12] Feng Li, Hao Zhang, Huaizhe Xu, Shilong Liu, Lei Zhang, Lionel M Ni, and Heung-Yeung Shum. Mask dino: Towards a unified transformer-based framework for object detection and segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3041–3050, 2023. 1, 2

[13] Yanwei Li, Hengshuang Zhao, Xiaojuan Qi, Liwei Wang, Zeming Li, Jian Sun, and Jiaya Jia. Fully convolutional networks for panoptic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 214–223, 2021. 2

[14] Umberto Michieli and Pietro Zanuttigh. Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 1114–1124, 2021. 2, 6

[15] Minh Hieu Phan, Son Lam Phung, Long Tran-Thanh, Abdesselam Bouzerdoum, et al. Class similarity weighted knowledge distillation for continual semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 16866–16875, 2022. 2, 6

[16] Chao Shang, Hongliang Li, Fanman Meng, Qingbo Wu, Heqian Qiu, and Lanxiao Wang. Incrementer: Transformer for class-incremental semantic segmentation with knowledge distillation focusing on old class. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7214–7224, 2023. 2

[17] Xinlong Wang, Rufeng Zhang, Tao Kong, Lei Li, and Chunhua Shen. Solov2: Dynamic and fast instance segmentation. Advances in Neural information processing systems, 33:17721–17732, 2020. 1, 2

[18] Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun,

[19] Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, et al. Dualprompt:

Complementary prompting for rehearsal-free continual learning. In European Conference on Computer Vision, pages 631–648. Springer, 2022. 3

[20] Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 139–149, 2022. 3

[21] Guanglei Yang, Enrico Fini, Dan Xu, Paolo Rota, Mingli Ding, Moin Nabi, Xavier Alameda-Pineda, and Elisa Ricci. Uncertainty-aware contrastive distillation for incremental semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2):2567–2581, 2022. 2, 6

[22] Chang-Bin Zhang, Jia-Wen Xiao, Xialei Liu, Ying-Cong Chen, and Ming-Ming Cheng. Representation compensation networks for continual semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7053–7064, 2022. 2, 6

[23] Wenwei Zhang, Jiangmiao Pang, Kai Chen, and Chen Change Loy. K-net: Towards unified image segmentation. Advances in Neural Information Processing Systems, 34:10326–10338, 2021. 1, 2

[24] Hanbin Zhao, Fengyu Yang, Xinghe Fu, and Xi Li. Rbc: Rectifying the biased context in continual semantic segmentation. In European Conference on Computer Vision, pages 55–72. Springer, 2022. 2