

## 基于门控机制与重放策略的持续语义分割方法

杨静<sup>①②</sup> 何瑶<sup>\*①②</sup> 李斌<sup>①②</sup> 李少波<sup>①②</sup> 胡建军<sup>③</sup> 溥江<sup>④</sup>

<sup>①</sup>(贵州大学机械工程学院 贵阳 520025)

<sup>②</sup>(贵州大学公共大数据国家重点实验室 贵阳 550025)

<sup>③</sup>(美国南卡罗莱纳州大学计算机科学与工程系 哥伦比亚 29208)

<sup>④</sup>(贵州民族大学机械电子工程学院 贵阳 550025)

**摘要:** 基于深度神经网络的语义分割模型在增量更新知识时由于新旧任务参数之间的干扰加之背景漂移现象, 会加剧灾难性遗忘。此外, 数据常常由于隐私、安全等因素无法被存储导致模型失效。为此, 该文提出基于门控机制与重放策略的持续语义分割方法。首先, 在不存储旧数据的情况下, 通过生成对抗网络生成及网页抓取作为数据来源, 使用标签评估模块解决无监督问题、背景自绘模块解决背景漂移问题; 接着, 使用重放策略缓解灾难性遗忘; 最后, 将门控变量作为一种正则化手段增加模型稀疏性, 研究了门控变量与持续学习重放策略结合的特殊情况。在Pascal VOC2012数据集上的评估结果表明, 在复杂场景10-2, 生成对抗网络 (GAN)、Web的设置中, 该文在全部增量步骤结束后的旧任务性能比基线分别提升了3.8%, 3.7%, 在场景10-1中, 相比于基线分别提升了2.7%, 1.3%。

**关键词:** 持续学习; 语义分割; 重放策略; 门控变量

**中图分类号:** TP183; TN919.81

**文献标识码:** A

**文章编号:** 1009-5896(2022)YU-0001-10

**DOI:** [10.11999/JEIT230803](https://doi.org/10.11999/JEIT230803)

## A Continual Semantic Segmentation Method Based on Gating Mechanism and Replay Strategy

YANG Jing<sup>①②</sup> HE Yao<sup>\*①②</sup> LI Bin<sup>①②</sup> LI Shaobo<sup>①②</sup> HU Jianjun<sup>③</sup> PU Jiang<sup>④</sup>

<sup>①</sup>(School of Mechanical Engineering, Guizhou University Guiyang 520025, China)

<sup>②</sup>(State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China)

<sup>③</sup>(Department of Computer Science and Engineering, University of South Carolina, Columbia 29208, USA)

<sup>④</sup>(School of Mechatronics Engineering, Guizhou Minzu University, Guiyang 550025, China)

**Abstract:** Due to the interference and background drift between new and old task parameters, semantic segmentation model based on deep neural networks promotes catastrophic forgetting of old knowledge. Furthermore, information frequently cannot be stored owing to privacy concerns, security concerns, and other issues, which leads to model failure. Therefore, a continual semantic segmentation method based on gating mechanism and replay strategy is proposed. First, without storing old data, generative adversarial network and webpage crawling are used as data sources, the label evaluation module is used to solve the unsupervised problem and the background self-drawing module is used to solve background drift problem. Then, catastrophic forgetting is mitigated by replay strategy; Finally, gated variables are used as a regularization means to increase the sparsity of the module and study the special case of gated variables combined with continual learning replay strategy. Our evaluation results on the Pascal VOC2012 dataset show that in the settings of complex scenario 10-2, Generative Adversarial Networks (GAN) and Web, the performance of the old task after all incremental steps are improved by 3.8% and 3.7% compared with the baseline, and in scenario 10-1, they are

收稿日期: 2023-08-01; 改回日期: 2023-12-02; 网络出版: 2023-12-14

\*通信作者: 何瑶 [gs.yaohe21@gzu.edu.cn](mailto:gs.yaohe21@gzu.edu.cn)

基金项目: 国家自然科学基金(62166005), 贵阳市科技人才培养对象及培养项目(筑科合同[2023]48-8号), 国家重点研发计划(2018AAA010800)

Foundation Items: The National Natural Science Foundation of China (62166005), Developing Objects and Projects of Scientific and Technological Talents in Guiyang City (ZKH[2023]48-8), The National Key R&D Program of China (2018AAA010800)

improved by 2.7% and 1.3% compared with the baseline, respectively.

**Key words:** Continual learning; Semantic segmentation; Replay strategies; Gating variables

## 1 引言

得益于更大的数据集、更强的计算能力以及网络结构创新,深度学习在图像分类<sup>[1,2]</sup>、语义分割<sup>[3,4]</sup>以及目标检测<sup>[5,6]</sup>等任务上几乎超越了人类的性能。然而传统神经网络只能在预先设定的批量学习任务下进行训练,当有新数据出现时,为适应新数据变化会出现“灾难性遗忘”(Catastrophic Forgetting, CF),持续学习(Continual Learning, CL)技术使得神经网络能在不重新训练的情况下学习新数据<sup>[7]</sup>。

为了解决灾难性遗忘问题,目前已有研究人员提出使用正则化方法<sup>[8-10]</sup>。文献<sup>[8]</sup>提出弹性权重整合(Elastic Weight Consolidation, EWC),通过计算先前任务的重要权重和其相关的Fisher信息矩阵进行约束以保持其在新任务上的稳定性。文献<sup>[9]</sup>提出智能突触(Synaptic Intelligence, SI),为每个单独的突触建立一个对过去任务局部重要性评估指标,并在学习新任务时对这些参数进行惩罚防止遗忘。这些正则化方法在使用多头分类器和任务标识符的场景中确实是有效的。然而,正如文献<sup>[11]</sup>所注意到的,正则化方法通过限制网络参数的变化控制过拟合,会导致网络的可塑性降低。

重放记忆方法是解决学习效率的主要手段之一<sup>[12-14]</sup>,旨在持续学习中保持旧任务的知识,减少任务之间的耦合并提高学习速度。在学习新任务以缓解遗忘的同时,对旧任务样本进行重放。文献<sup>[12]</sup>提出存储部分旧数据联合当前数据训练模型缓解灾难性遗忘。由于内存限制或隐私问题,数据不总是能被存储。针对此,文献<sup>[13]</sup>提出使用生成模型生成伪样本,通过使用生成模型生成一些与过去任务相关的伪样本,并与当前任务的真实样本混合输入模型。文献<sup>[14]</sup>提出生成负重放去处理不完美的生成数据,并在增量训练分类器时尝试充分利用它们。

除了灾难性遗忘问题,任务级过拟合现象也是持续学习中的另一个重要障碍。由于网络在持续学习过程中通过反复训练和微调适应新任务,导致参数在训练任务和新任务之间发生过拟合。任务级过拟合会带来一些问题。首先,当网络在新任务上过度拟合时,它可能会丧失对旧任务的准确性。其次,任务级过拟合可能会导致对新任务的过度自信,使网络误分类新的样本或对新的样本产生过度自信预测。

针对上述问题,本文提出将重放策略与门控变量结合,重放策略使模型可以重训练先前的任务样本,平衡新任务和旧任务之间的影响,保持对旧任务的记忆。门控变量调整任务的权重,提高模型的稳定性和泛化性能,减轻持续学习中的遗忘问题。具体而言,首先,在学习新任务时,本文使用生成对抗网络重放部分旧数据以达到持续语义分割目的,从过去任务的样本池中随机选择一部分样本进行重新训练,这样可以使网络保持对过去任务的记忆,并避免忘记之前学到的知识;其次,为了增加数据的多样性,使用网络抓取派生样本的方法作为第2个重放源,通过在Flickr网站上进行搜索和抓取图像获取数据;最后,为了进一步缓解灾难性遗忘提高持续学习语义分割的精度,本文在模型中引入门控变量机制,并将其与持续学习重放策略结合,本方法不仅可以解决遗忘问题,还可以作为一种正则化手段,增加模型的稀疏性,减少持续学习过程中新旧参数之间的干扰。

## 2 具有随机门控的持续语义分割重放方法

为了缓解语义分割模型在持续学习过程中的灾难性遗忘现象,提升持续学习语义分割精度,本文基于重放策略,并在网络模型中引入门控0-1伯努利变量,将其作为一种正则化方法,通过增加网络稀疏性减少数据重放过程中参数间干扰。

### 2.1 持续语义分割重放方法

本文重放的数据来源分为两种:生成对抗网络(Generative Adversarial Networks, GAN)生成以及网络抓取,训练时通过重放部分旧数据以达到持续语义分割目的。两种重放源按数据集Pascal VOC2012里首字母排序后选取前20个类进行数据生成与抓取,表1给出了20个类的中英文类名。

本文使用的第1个重放源是基于GAN生成的数据,在持续学习过程中将生成的旧任务数据输入网络以缓解灾难性遗忘。为了达到这个目的,使用高保真自然图像合成的大规模生成对抗网络为以前训练的类重新创建不再可用的样本并混合新任务数据一起输入网络。为了使生成网络生成训练过程中所需要的图像样本,GAN需要在一个通用的大规模可视化数据集上进行预训练。为此,本文选择基于图像网络数据集(ImageNet)的预训练。图1展示了部分GAN生成的图像。

考虑到互联网丰富的图像资源,本文的第2个重放源是使用网络抓取派生样本进行重放。本文从

表 1 20个类的中英文类名

中文类名	飞机	自行车	鸟	船	瓶子	巴士汽车	汽车	猫	椅子	牛
英文类名	airplane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow
中文类名	餐桌	狗	马	摩托车	人	盆栽	羊	沙发	火车	监视器
英文类名	dining table	dog	horse	motorbike	person	potted plant	sheep	sofa	train	monitor

Flickr网站抓取给定的20个类的图像，每个类抓取1000张图片。图2展示了部分使用网页抓取的图片。

为了使两种重放源的图像数据可以用于模型训练，使用文献[15]的标签评估模块以及背景自绘模块对图像进行处理以符合训练要求，其中标签评估模块在训练过程中对图像注释解决无监督问题，背景自绘模块解决持续学习语义分割的背景漂移问题。由于模型的数据来源于GAN生成和网络抓取，因此对于数据由于隐私等问题无法被存储、采集，本文方法依然可以进行模型训练。

为了处理背景漂移，文献[15]提出了一种简单而有效的自绘机制，将知识从以前的模型转移到当前的模型中。当重放源重新创建旧类样本时，图像背景使用自绘模块作用于当前样本的背景区域，减少背景偏移，同时带来类似于知识蒸馏[16]的正则化效果。在训练集 $T_k$ 的每一步 $k$ 中，自绘模块取每个标签的背景区域，用之前模型 $M_{k-1}$ 的相关预测对其进行标记。因为标签中的背景区域是根据自教学方案而改变的基于对旧模型的预测，重绘模块将步骤 $k > 0$ 中可用的每个原始标签映射 $Y$ 替换为其内绘制版本 $Y^{bi}$ ，如式(1)所示

$$Y^{bi}[H, W] = \begin{cases} Y[H, W], & Y[H, W] \in C_k \\ \operatorname{argmax}_{c \in C_{0 \rightarrow k-1}} M_{k-1}(X)[H, W][c], & \text{其他} \end{cases} \quad (1)$$

其中， $X, Y \in T_k$ 表示像素坐标。步骤 $k=0$ 中的标签没有被绘制，因为在那个阶段缺乏对过去类的任何先验知识。当执行背景自绘时，每个集合 $T_k^{bi} \subset \chi \times y_{C_{0 \rightarrow k}} (k > 0)$ 包含绘制后的所有样本。

## 2.2 基于随机门控的网络训练方法

由于引入了门控0-1伯努利变量，网络的训练模式将会产生变化。当 $\delta_i^l$ 为0时，表示该神经元不参与参数更新。训练时，每次随机将 $\delta_i^l$ 赋值为0，这相当于随机从大网络中采样选择小网络进行训练，使得网络产生稀疏性，参数之间的干扰变小进而提升网络持续学习能力，如图3所示。

图3(a)为引入门控变量的持续语义分割模型，图3(b)为引入门控变量进行训练的网络。当图3(a)中的 $\delta_i^l$ 为0时就产生了图3(b)中的虚线部分，这导

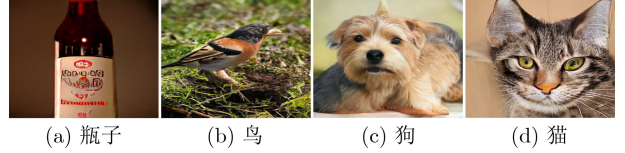


图 1 部分使用GAN生成的图片



图 2 部分网页抓取的图片

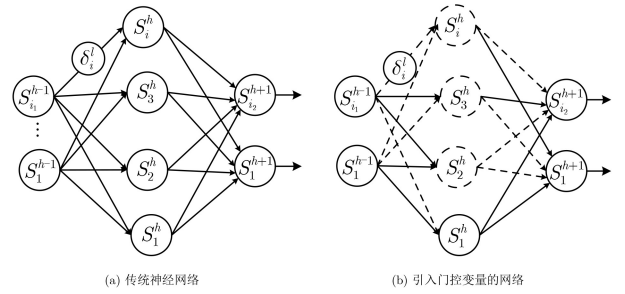


图 3 网络引入门控0-1伯努利变量的训练模式变化

致了每次训练都是从大网络中采样一个小网络进行训练。在网络引入门控0-1伯努利变量后，训练时可用的子网络数 $N$ 如式(2)所示

$$N = C_m^n, m = i, n = i \times p \quad (2)$$

其中， $p = P(\delta_j^l = 1)$ ，即 $p$ 是 $\delta_j^l = 1$ 时的概率，也可以被称为子网采样率； $m, i$ 代表 $l$ 层神经元总数； $n$ 代表被采样参与训练的神经元数量。由式(2)可知，当子网络采样率 $p$ 达到0.5时， $N$ 的值达到最大，此时每次训练可选择的子网络数量最多，网络稀疏性与稳定性达到最强，参数间干扰最小。

在网络的多个隐藏层引入伯努利变量时，网络的子网络数量 $N$ 如式(3)所示

$$N = C_{m_1}^{m_1} \times C_{m_2}^{m_2} \times \dots \times C_{m_l}^{m_l} \quad (3)$$

其中， $m_l = i, n_l = i \times p_l$ ， $l$ 表示在所有应用门控0-1伯努利变量的网络层中的第 $l$ 层。由式(3)可知，在网络的多个隐藏层中引入门控0-1伯努利变量时， $p$ 取0.5时子网络数量最大，网络稀疏性达与稳定性到最强。

### 2.3 基于随机门控的神经元稀疏激活方法

引入门控0-1伯努利变量后,网络训练模式转换为每次随机从网络中选择一个子网络进行训练,神经元在这个过程中也是以随机方式被激活进行参数更新,从而减少了参数间干扰,缓解灾难性遗忘。

文献[17]的研究结果表明,对于人工神经网络中第 $h$ 层的神经元 $i$ 在引入门控0-1伯努利变量后的活动,如式(4)所示

$$S_i^h = \sum_{l < h} \sum_j \omega_{ij}^{hl} S_j^l \delta_j^l, S_j^0 = I_j \quad (4)$$

其中,变量 $\omega$ 表示模型权重, $I$ 表示模型输入向量, $\delta_i^h$ 表示应用于模型第 $h$ 层神经元 $i$ 的门控0-1伯努利变量, $P(\delta_j^l = 1) = p$ (即: $h$ 层神经元 $i$ 被激活,参与训练的概率为 $p$ )。

假设仅在 $h$ 层中应用 $\delta_i^l$ ,其中神经元的输出形式为 $O_i^h = \sigma(S_i^h)$ 且 $S_i^h = \sum_{l < h} \omega_{ij}^{hl} \delta_j^l O_j^l$ 。对于一个固定的输入, $O_j^l$ 是一个常数,因为 $\delta_i^l$ 没有应用于第 $l$ 层。因此每个神经元的激活方差如式(5)所示

$$\text{Var}(S_i^h) = \sum_{l < h} (\omega_{ij}^{hl})^2 (O_j^l)^2 p(1-p) \quad (5)$$

其中,神经元的激活方差 $\text{Var}(S_i^h)$ 在 $p$ 接近0或1时变小,取0.5时最大,当子网络概率 $p$ 取0.5时,神经元以最稀疏模式被激活参与训练,此时模型稀疏性与稳定性达到最强,与式(2)、式(3)的结论相同。由于 $p$ 取值在接近0时与接近1时有相同的正则化效果,但是会丢失过多神经元,所以一般不取0.5以下的值。当 $p=0.5$ 时,网络具有最强的稀疏性与稳定性,但是会失去部分可塑性导致收敛速度变慢。

使用门控0-1伯努利变量作为正则化手段会推动神经元活跃度向着0或1发展,这意味着神经元在当前训练过程中只有更新或者不更新两个状态。当网络完成任务 $t$ 并开始学习任务 $t+1$ 时,门控0-1伯努利变量通过禁用或启用神经元来创建新的“门”,以此避免灾难性遗忘。具体来说,门控0-1伯努利变量可以通过随机地将一部分神经元的激活置为0,从而减少神经元之间的依赖关系。这样每个神经元都需要学习适应不同的子网络,从而增加了网络的鲁棒性和泛化能力。

在2.2节的基础上,基于已有专家们的研究,本文分析门控0-1伯努利变量结合重放策略时的情况,进一步完善了门控0-1伯努利变量结合持续学习技术时的理论。文献[18]在神经网络中引入门控0-1伯努利变量,取 $p=0.5$ 时网络稳定性最强,具有最缓慢的遗忘曲线,式(2)和式(3)的结论也印证了这一点。然而,持续学习重放策略在训练时,旧任

务的数据样本会重新输入网络,为了使网络保持对旧任务的性能,需要提升子网络采样率。

对于持续学习重放策略,从1层神经元 $j$ 到 $h$ 层神经元 $i$ 的前馈过程如式(6)所示

$$z_i^{(l+1)} = (\omega_{ij}^{hl})_{D_d + D_{\kappa < d}} \tilde{y}_i^l + b_{ij}^{hl} \quad (6)$$

其中, $D_{\kappa < d}$ 为旧任务的全部或部分数据样本。当 $p$ 取0.5时,网络的稀疏性最强,相应的,由于子网络数量最多,各子网络所含神经元数量也最少,这意味着网络的可塑性也相应降低。由于旧任务的数据样本会重新输入网络,对基于重放方法的持续语义分割而言,在 $p=0.5$ 的基础上,适当提高 $p$ 的取值,使网络具有一定稳定性的同时不失去过多可塑性,更有利于网络保持对旧任务的性能。

如图4所示,本文使用ResNet-101<sup>[19]</sup>作为主干网络,Deeplab-V2<sup>[20]</sup>作为分割结构。本文的重放策略结构是独立于特定的网络结构。

## 3 实验结果与分析

所有实验均在同一硬件环境中进行,具体为NVIDIA GeForce RTX 2080Ti显卡,inter Xeon(R) Silver 4210R CPU,华硕32G\*2运行内存。使用DeepLab-V2<sup>[20]</sup>作为语义分割网络,ResNet-101<sup>[19]</sup>作为骨干网络。在此基础上,在ResNet-101的随机3层全连接层中应用了门控0-1伯努利变量。重放的数据来源分为GAN,Web,所有的实验结果均为5次实验的平均交并比(mean Intersection over Union, mIoU) $\pm$ 标准差。

### 3.1 开放场景实验设置

实验场景共有6个:一共训练20个类,首先训练10个类,后续的10个类分为1个增量步骤(10-10)、2个增量步骤(10-5)、5个增量步骤(10-2)以及10个增量步骤(10-1)进行增量训练;首先训练19个类,再训练后续的1个类(19-1);首先训练15个类,后续的5个类执行1个增量步骤进行训练(15-5)。

本实验在Overlapped的场景中进行,在第1阶段,选择只有 $C_0$ 标记像素的训练图像的子集。然后,每个增量步骤的训练集包含所有从 $C_k$ 中标记像素的图像,即 $T_k^{\text{bi}} \subset \chi \times y_{C_0 \rightarrow k}, k > 0$ 。与初始步骤类似,标签被限制在 $C_k$ 中的语义类中,而剩余的像素被分配给 $b$ , $b$ 在每一步都经历了语义转换。每个训练步骤包含至少有1个新类的像素的所有图像,即训练集包含所有从 $C_k$ 中标记像素的图像。这种设置下,模型为了获得最多的知识恢复增益,需要采用更高的子网络采样率以获取更强的模型可塑性。

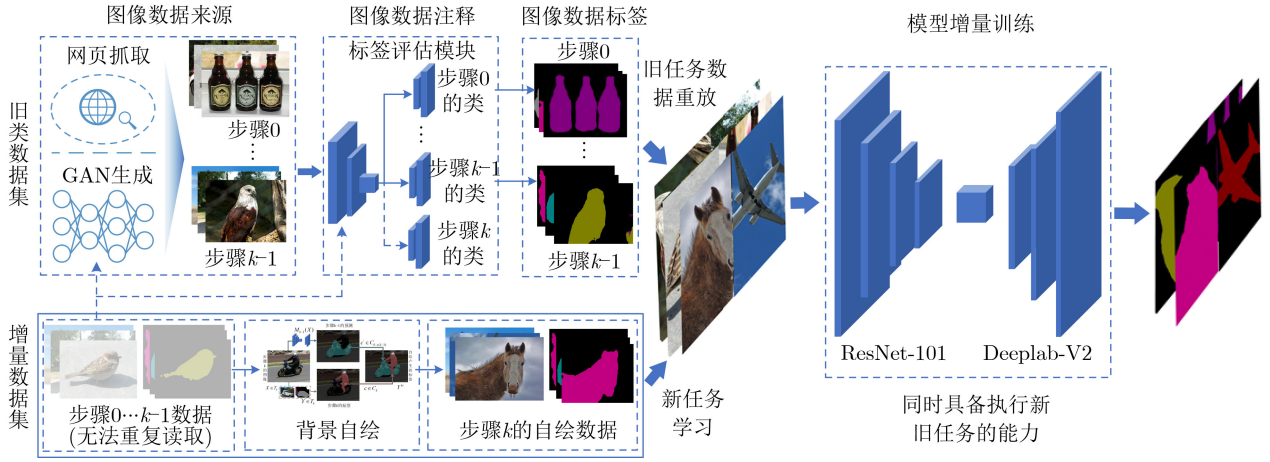


图 4 网络模型结构示意图

### 3.2 对比方法

(1)学而不忘记(Learning without Forgetting, LwF)<sup>[21]</sup>。

(2)增量分类器和表示学习 (iCaRL: incremental Classifier and Representation Learning, iCaRL)<sup>[22]</sup>。

(3)类增量语义分割方法(Incremental Learning Techniques for semantic segmentation, ILT)<sup>[23]</sup>。

(4)既不使用旧数据也不使用旧标签的类增量学习语义分割方法(Class-Incremental Learning for semantic segmentation re-using neither old data nor old labels, CIL)<sup>[24]</sup>。

(5)语义分割中增量学习的背景建模(Modeling the Background for incremental learning in semantic segmentation, MiB)<sup>[16]</sup>。

(6)基于稀疏和解纠缠潜在表征排斥-吸引的持续语义分割方法(continual semantic segmentation via repulsion-attraction of Sparse and Disentangled latent Representations, SDR)<sup>[25]</sup>。

(7)基于类增量学习网络的航空图像小目标语义分割增强(Class-Incremental Learning network for small Objects Enhancing of semantic segmentation in aerial imagery, CILOE)<sup>[26]</sup>。

(8)持续的语义分割, 不忘学习(learning without forgetting for continual semantic segmentation)<sup>[27]</sup>。

(9)继承与升华(Inherit with Distillation and Evolve with Contrast, IDEC), 演化与对比:无样例记忆类增量语义分割探索( exploring class incremental semantic segmentation without exemplar memory)<sup>[28]</sup>。

此外, 还包括了增量模型精度下限的微调(Fine-

Tuning approach, FT)以及使用完整数据集进行联合训练(Joint)作为增量模型精度上限, 上述这些方法的数据来源于文献[15]。GAN与Web的数据来源于使用GAN生成数据以及从网页抓取数据后, 使用文献[15]的标签评估模块与背景自绘模块进行实验的结果。

### 3.3 稳定性-可塑性实验

#### 3.3.1 模型稳定性-可塑性实验

为了测试引入门控0-1伯努利变量对模型稳定性-可塑性的影响, 我们分别测试了 $p=0.9$ ,  $p=0.8$ ,  $p=0.7$ ,  $p=0.6$ 以及 $p=0.5$ 这5种子网络采样率时旧类与新类的mIoU在每个增量步骤训练结束后的变化情况, GAN与Web分别表示只使用重放策略而未引入门控变量的实验结果, 如表2所示。

(1)GAN结合门控变量的设置。在这个设置中, 门控变量在19-1, 15-5, 10-10这3个简单场景中对模型稳定性-可塑性的提升较弱, 新任务的性能提升均在一个百分点以内, 在复杂的场景10-1, 10-2中, 在全部增量步骤结束后, 旧任务(0-18)性能相比于重放策略(GAN& $p$ ), 分别提升了2.7%, 3.8%。在模型可塑性变化上, 19-1, 15-5, 10-10这3个简单场景中, 新任务性能相比重放策略, 均只上升了不到一个百分点。在最复杂的场景10-2与10-1中, 与最优结果相比, 新任务性能在每个增量步骤后的平均提升数值分别为0.84%, 4.0%。

综合上述对实验结果的分析可以发现: 门控变量在overlapped, GAN的设置中, 对模型稳定性-可塑性的优化不够明显。这是因为在overlapped设置下, 每个训练步骤都会包含至少有一个新类像素的所有图像, 其灾难性干扰更加严重。

(2)Web结合门控变量的设置。以简单场景19-1为例, 与最优结果相比, 全部增量任务结束后模型对旧任务(0-19)的提升为1.0%; 在复杂的场景10-1,

表2 门控0-1伯努利变量对模型稳定性-可塑性的影响

场景	类	GAN	$p=0.9$	$p=0.8$	$p=0.7$	$p=0.6$	$p=0.5$	Web	$p=0.9$	$p=0.8$	$p=0.7$	$p=0.6$	$p=0.5$
19-1	0-19	68.2±0.8	67.9±0.3	67.0±0.8	68.2±0.4	67.6±0.3	68.4±0.3	67.7±1.3	67.2±0.7	67.1±1.0	68.3±0.7	67.7±0.3	68.4±0.3
	20	50.9±1.1	51.3±1.3	51.2±1.1	52.2±1.1	50.6±1.7	50.1±2.3	51.0±1.5	51.9±1.4	53.0±0.9	51.7±1.6	52.0±2.7	52.9±1.6
15-5	0-15	68.9±0.3	69.1±0.3	69.1±0.5	69.0±0.4	69.1±0.5	69.9±0.2	69.4±1.7	70.0±0.4	69.9±0.9	69.8±0.3	69.9±0.7	70.7±0.3
	16-20	51.5±0.5	51.3±0.6	51.3±0.4	51.6±0.6	51.5±0.8	51.8±0.5	54.2±0.4	54.2±0.6	54.6±0.5	54.3±0.3	55.0±0.5	55.0±0.4
10-10	0-10	66.8±0.5	67.0±0.2	66.4±0.3	67.3±0.3	66.9±0.5	66.6±0.7	68.4±0.3	68.6±0.2	67.6±0.6	68.4±0.2	68.1±0.1	68.0±0.7
	11-20	58.1±0.6	59.2±0.5	58.4±0.6	58.6±0.4	58.3±0.4	57.7±0.4	58.1±0.5	59.6±0.6	58.1±0.6	58.8±0.3	58.1±0.4	57.5±0.4
10-5	0-10	68.7±0.6	68.4±0.4	68.0±0.2	68.8±0.4	68.6±0.2	68.1±0.9	69.5±0.5	69.9±0.2	69.0±0.4	69.9±0.3	69.5±0.2	68.7±0.6
	11-15	63.1±0.3	63.8±1.2	63.0±0.7	63.3±0.7	63.2±0.9	63.0±0.5	63.7±0.8	65.0±0.7	63.9±0.4	64.3±0.3	63.9±1.1	63.7±1.1
10-2	0-15	61.3±0.8	61.7±0.8	61.5±0.5	61.6±0.7	62.0±0.5	61.6±0.4	63.1±0.5	62.5±1.3	64.0±0.6	62.2±0.8	63.9±0.6	63.9±1.0
	16-20	49.3±0.5	50.0±1.1	49.2±0.6	48.4±0.2	48.7±1.1	49.0±0.1	53.0±0.4	54.1±0.8	53.1±0.5	53.0±0.3	53.4±0.9	53.2±0.3
10-1	0-10	71.2±0.7	70.9±0.8	70.0±0.3	71.4±0.6	71.1±0.4	70.2±1.0	72.3±0.4	72.3±0.4	71.5±0.6	72.0±1.1	71.8±0.4	71.5±1.0
	11-12	57.0±1.0	56.2±1.5	56.2±1.1	57.1±0.2	57.7±1.2	56.5±1.0	60.5±0.9	60.1±1.7	60.8±0.7	59.8±1.2	60.7±0.4	60.8±0.5
10-2	0-12	65.6±1.3	65.9±0.7	65.4±0.6	65.8±0.6	65.5±0.3	64.7±0.5	67.3±0.9	68.3±1.1	67.4±0.2	68.0±1.0	66.9±0.6	67.1±0.6
	13-14	53.3±1.7	53.8±0.8	53.5±1.1	53.3±0.7	53.1±0.4	52.9±1.5	60.0±1.6	60.6±1.3	61.0±0.3	61.0±1.4	60.9±0.4	60.6±1.3
10-2	0-14	62.5±1.3	62.7±0.6	62.1±0.6	62.9±0.6	61.9±0.5	62.0±0.9	64.3±1.6	65.5±0.6	64.9±0.6	65.2±0.3	64.2±1.2	64.7±0.7
	15-16	51.4±1.8	53.6±1.5	53.4±1.6	53.1±1.2	53.4±2.7	53.2±1.4	52.8±1.6	54.7±1.8	54.9±1.2	54.2±0.9	53.9±1.7	53.6±1.3
10-2	0-16	56.6±0.9	58.0±1.1	56.5±0.8	57.2±0.3	56.5±0.7	56.4±0.6	59.0±1.1	61.7±0.9	59.7±0.2	60.3±0.3	59.4±0.8	60.1±0.9
	17-18	36.8±1.0	36.0±0.8	35.2±0.4	35.1±1	34.2±1.3	34.5±1.5	41.1±0.7	41.8±1.2	41.8±0.7	41.2±2.1	40.6±1.1	42.1±1.1
10-2	0-18	53.1±0.8	55.1±1.2	53.4±0.8	54.2±0.2	52.8±0.3	53.4±0.6	57.5±0.7	59.6±0.7	57.5±0.6	58.0±0.4	57.1±0.7	57.8±1.1
	19-20	55.1±1.1	54.1±2.0	54.8±0.7	55.1±1	53.9±1.1	54.2±0.2	60.2±1.1	60.0±1.2	60.2±1.0	59.7±0.7	60.5±0.7	60.1±1.0
10-1	0-10	73.7±0.6	74.0±0.7	72.9±0.5	74±0.7	73.2±0.8	72.4±0.6	73.9±0.3	74.3±0.4	73.1±0.7	74.6±0.2	73.7±0.8	72.8±0.7
	11	31.8±1.3	33.8±2.9	32.1±2.4	33.5±2.1	35.9±3.6	35.0±3.2	31.5±3.4	34.2±3.9	34.0±1.2	32.4±1.5	38.2±1.7	34.4±3.0
10-1	0-11	66.4±0.4	66.7±0.8	65.5±0.2	66.5±0.7	66.5±1.2	65.2±0.6	66.5±0.3	67.6±0.8	66.1±0.3	67.4±0.5	67.5±0.9	66.4±1.1
	12	60.1±2.6	59.1±3.5	60.5±1.0	57.2±2.6	60.1±2.4	57.5±0.4	63.8±1.8	65.2±2.7	65.2±3.1	62.2±1.2	66.0±1.7	65.8±2.9
10-1	0-12	65.1±0.7	65.4±0.9	64.2±0.4	65.7±0.5	64.3±1.0	64.1±0.6	65.9±0.2	66.6±1.3	65.1±0.6	66.6±0.6	65.9±1.0	65.0±1.0
	13	38.8±1.5	39.3±1.9	39.4±1.5	39.5±0.6	39.0±1.3	40.2±2.2	45.5±1.6	45.2±3.2	43.3±2.3	44.9±1.9	45.4±1.6	47.6±1.1
10-1	0-13	64.4±0.9	65.1±0.2	63.5±0.8	64.3±0.8	64.2±0.4	63.2±0.3	66.0±0.7	66.8±1.2	65.0±0.6	67.1±0.8	65.5±1.2	65.4±0.7
	14	57.2±1.3	58.2±1.4	58.9±0.9	59.4±1.5	57.9±0.4	56.4±2.2	62.8±2.7	64.7±1.5	64.6±0.3	65.6±1.9	64.7±2.9	62.3±2.5
10-1	0-14	62.3±0.9	62.9±0.4	61.3±0.6	62.4±0.8	61.9±0.9	60.6±0.8	63.8±0.5	65.0±0.9	63.4±0.3	65.4±0.8	63.4±1.2	63.8±0.8
	15	67.3±0.5	68.7±1.0	68.9±0.3	69.1±0.6	68.2±0.6	68.5±0.6	68.4±0.5	68.4±0.3	69.2±0.1	68.3±0.6	69.1±0.7	69.4±0.5
10-1	0-15	63.6±0.5	64.2±0.4	63.4±0.6	64.2±0.6	63.3±1.3	63.1±0.4	65.0±0.5	66.2±0.6	65.1±0.4	66.2±0.6	65.3±0.9	65.2±0.4
	16	26.5±4.0	28.2±1.2	27.7±3.2	26.4±1.6	26.2±1.3	25.6±1.9	37.1±1.8	35.6±1.6	37.2±2.7	36.4±1.6	36.5±1.7	38.2±1.6
10-1	0-16	57.9±0.4	58.6±0.5	57.2±0.6	58±0.6	57.2±0.5	56.9±0.7	60.4±0.2	61.4±1.4	60.3±0.2	61.2±1.1	60.5±1.5	60.8±0.6
	17	33.6±1.3	34.0±0.5	32.4±1.9	33±0.8	31.1±1.4	33.0±0.3	42.3±2.9	44.7±2.2	43.6±2.7	42.2±1.9	43.2±2.0	45.6±1.4
10-1	0-17	56.8±0.7	58.2±0.2	57.2±1.1	57.8±0.6	56.2±0.6	56.4±0.7	61.1±0.4	61.8±0.6	60.6±0.9	61.9±0.7	60.7±0.8	61.2±0.4
	18	24.1±0.9	23.7±0.9	25.3±0.6	24.7±1.7	23.7±1.7	23.9±1.2	23.9±0.9	24.4±0.5	25.2±0.4	24.2±0.6	25.4±2.0	24.8±0.6
10-1	0-18	54.3±0.8	55.6±0.5	54.5±0.5	54.8±1	54.6±0.5	53.9±0.4	59.1±0.5	59.1±0.6	58.8±1.0	59.5±1.1	58.8±0.7	58.5±0.6
	19	49.4±1.8	51.4±2.0	52.4±3.2	50.5±1.5	50.6±0.4	50.7±0.6	57.1±1.0	59.7±1.6	60.0±1.9	57.3±1.5	57.6±3.1	57.1±1.9
10-1	0-19	56.4±0.9	57.4±0.4	56.4±0.3	57.7±0.3	56.1±0.3	56.1±0.3	60.7±0.4	61.4±0.7	60.5±1.1	61.5±0.5	60.6±0.7	60.4±0.5
	20	45.1±2.2	42.5±2.3	43.9±1.1	43.9±2.6	44.0±2.7	43.9±1.6	51.4±1.5	50.6±1.0	49.5±2.0	50.2±1.6	48.4±2.2	48.8±1.4

10-2中, 在全部增量步骤结束后, 旧任务(0-18)性能相比于重放策略(Web&p), 分别提升了1.3%, 3.7%。在可塑性提升上, 这10-1和10-2场景的提升分别为

1.7%, 5.6%。整体上看, 门控变量在overlapped场景中, 对基于重放策略的持续学习语义分割模型的性能优化是有效的。

从GAN与Web两个设置的结果可以发现，在所有6个增量场景里，GAN与Web设置中，引入门控0-1伯努利变量对网络重放后的知识恢复效果以及可塑性均有不同程度的提升。这表明，在网络中引入门控变量以减少重放过程中的参数间干扰，有利于提升网络的稳定性-可塑性平衡点。

### 3.3.2 模型稳定性及重放策略特殊情况验证

为了进一步验证引入门控0-1伯努利变量对网络稳定性的影响，以及结合持续学习重放策略时的特殊情况，本文展示了在GAN与Web两种设置下，旧类在场景10-1, 10-2中的每一个增量步骤结束后，mIoU的变化情况，如图5所示。

通过观察图5的4个子图可以发现，在合适的子网络采样率下，网络保持旧类的能力相比于只使用重放方法有较大提升，具体来说，设置0.7的子网络采样率时，其增益效果在场景10-1以及10-2(GAN)设置中是较为明显的，在图5(a)—图5(c)3个子图中也以实线的方式标出。而在图5(d)中，增益效果最强的设置则是出现在 $p=0.9$ 的设置中。观察图5的4个子图可以发现，引入门控变量能够有效提升模型重放后的知识恢复效果，模型的遗忘曲线也比仅使用重放策略更缓慢，这说明门控变量在overlapped设置下也能有效提升模型性能。

图5的4个子图的结果也很好验证了门控变量

结合重放方法的特殊情况。对比图5(a)—图5(c)中 $p=0.5$ 与 $p=0.7$ 两条折线的变化趋势可以发现，在 $p=0.5$ ，网络达到最强稳定性时，其对初始类集与背景类的性能要弱于 $p=0.7$ 的设置，这是因为过高的稳定性影响了网络从数据流中提取旧任务信息的能力，造成旧知识的恢复不如 $p=0.7$ 的设置。通过提升子网络采样率至 $p=0.7$ ，适当降低网络稳定性的同时提升网络可塑性更有利于网络保持旧任务的性能，这也符合该文在2.2节的理论分析与预期。

为了进一步观察模型引入门控0-1伯努利变量后的稳定性变化，同时验证门控变量结合重放策略的特殊性，下文展示初始类集与背景类(0-10类)在场景10-2,10-1中，每一个增量步骤结束后的mIoU变化，如图6所示。

观察图6可以发现，相比与重放策略(浅蓝色虚线)的遗忘曲线，4个子图中的最优结果要缓慢的多。除了图6(c)的最优结果里， $p=0.7$ 与 $p=0.9$ 的结果较为接近，其余子图的最优结果与前文的结果一致。模型在持续学习过程中保持定量类性能的实验结果表明，使用门控0-1伯努利变量构建门控机制在这个实验设置下依然有效。门控变量不仅可以优化数据重放后模型对旧类的性能，也可以优化模型保持定量类的性能。针对结合重放策略，从图6可以看出，效果最好的情况并不会出现在 $p=0.5$ (黑色

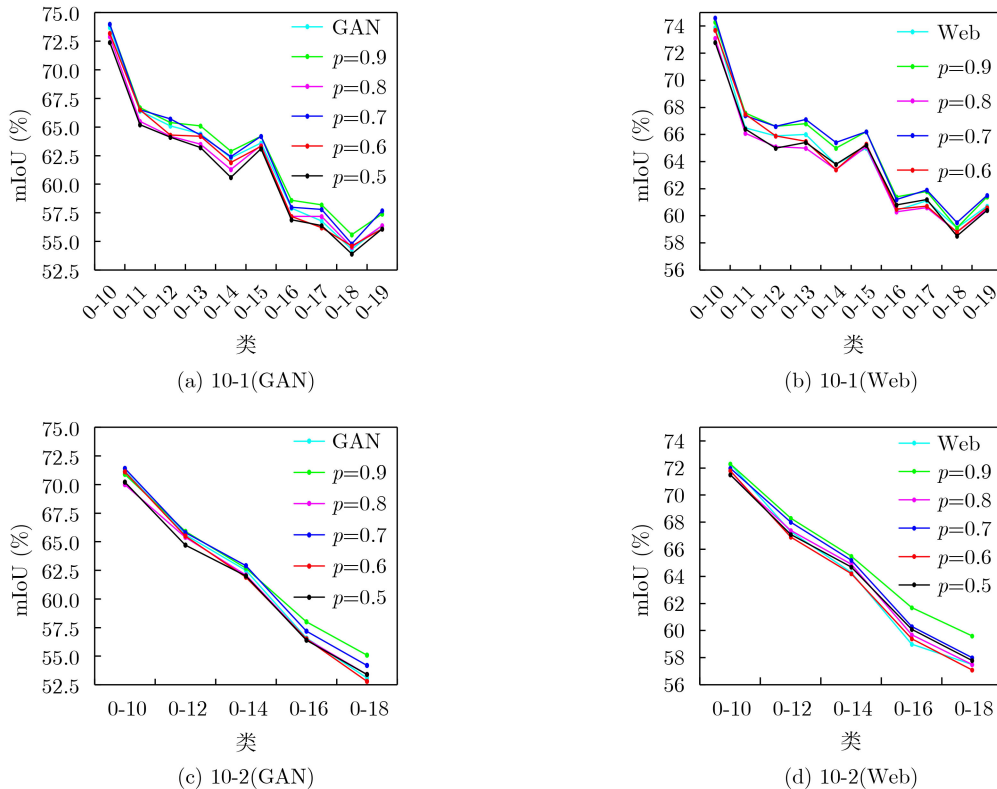


图 5 Overlapped场景中旧类mIoU的变化

虚线)的设置,此时适当提升网络可塑性更有利于知识恢复,这在图6(c)、图6(d)中可以更直观看出,与2.2节的理论预期一致。

### 3.4 开放场景实验对比分析

在Overlapped场景的设置中,每个训练步骤包含至少有一个新类像素所有图像,只对后者进行注释。需要注意的是,图像现在可能包含将来要学习类的像素,但被标记为背景。图7和图8展示了本文方法在overlapped设置下,增量场景10-5, 10-1下与其他方法的对比结果。

观察图7和图8可以发现,FT方法在增量训练后,由于灾难性遗忘导致表现不佳。在场景10-5中, $p=0.6$ 时,在基类和全部类中Web设置相比表现最好的IDEC的性能提升了4.9%, 1.4%,比SDR的性能提升了18.6%, 16.2%。在场景中10-1中 $p=0.7$ ,除了IDEC和本文所提的方法外,其余9种对比方法的性能均出现大幅度下降。Web设置在初始类集的mIoU中,性能比其他方法好,相比于PLOD提升21个百分点;增量类集11-20相比于表现最好的IDEC提升8个百分点,比CILOE提升32个百分

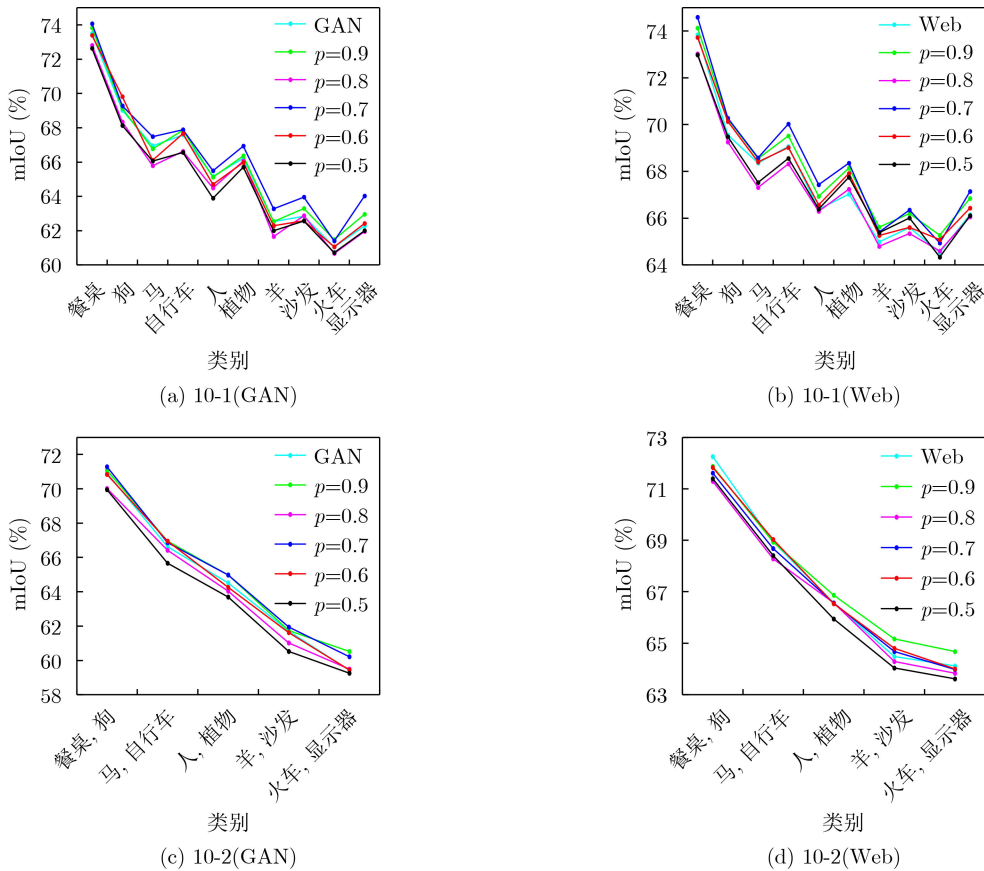


图6 overlapped场景中0-10类mIoU的变化

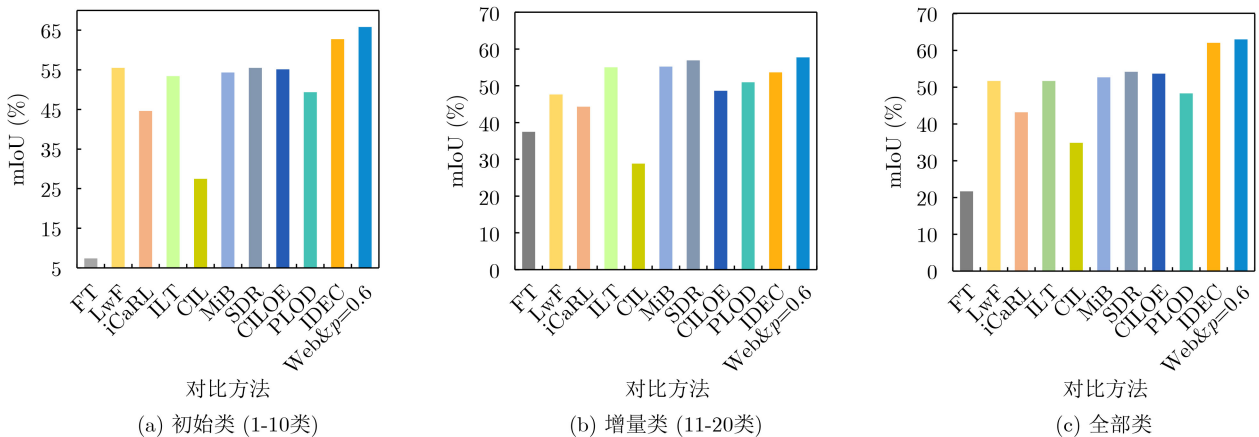
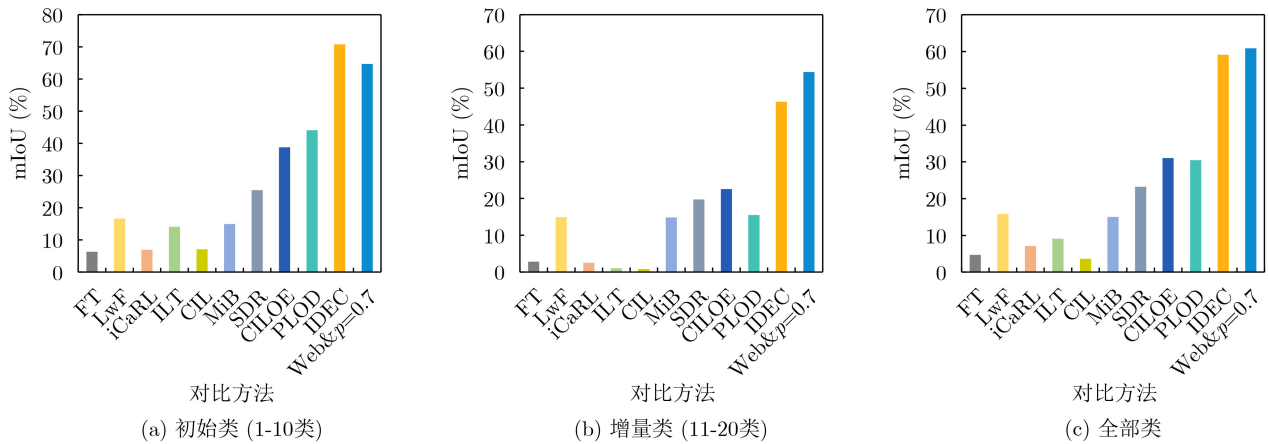


图7 overlapped设置下场景10-5的方法对比结果( $p=0.6$ )



图 8 overlapped设置下场景10-1的方法对比结果( $p=0.7$ )

点；而全部类的mIoU，在 $p=0.7$ 时，Web设置相比于表现最好的IDEC提升不太明显，但是比CILOE方法提升了30个百分点。此外，除了在场景10-1下的基类中，本文方法对旧任务的性能依然保持较高的水平。

#### 4 结束语

针对持续学习语义分割模型在持续学习过程中的灾难性遗忘、数据由于隐私等因素无法被存储导致模型无法训练的问题，本文提出一种基于门控机制与重放策略的持续语义分割方法。通过使用GAN及网页抓取两种数据重放源解决了数据无法被存储的问题，同时构建门控机制以优化模型性能。所提方法在Pascal VOC2012数据集上的评估结果证明了在增量场景10-5,10-1下相比于其他方法具有较大优势，且进一步验证了本文方法在提高模型的稳定性-可塑性上的有效性。在未来的工作中，将致力于改进生成模型以生成质量更好的样本以及考虑融合正则化方法提高模型性能。

#### 参考文献

- [1] GONG Xuan, XIA Xin, ZHU Wentao, *et al.* Deformable Gabor feature networks for biomedical image classification[C]. 2021 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, USA, 2021: 4003–4011. doi: [10.1109/WACV48630.2021.00405](https://doi.org/10.1109/WACV48630.2021.00405).
- [2] NING Xin, TIAN Weijuan, YU Zaiyang, *et al.* HCFNN: High-order coverage function neural network for image classification[J]. *Pattern Recognition*, 2022, 131: 108873. doi: [10.1016/j.patcog.2022.108873](https://doi.org/10.1016/j.patcog.2022.108873).
- [3] HE Junjun, DENG Zhongying, ZHOU Lei, *et al.* Adaptive pyramid context network for semantic segmentation[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 7511–7520. doi: [10.1109/CVPR.2019.00770](https://doi.org/10.1109/CVPR.2019.00770).
- [4] YANG Jing, LI Shaobo, WANG Zheng, *et al.* Using deep learning to detect defects in manufacturing: A comprehensive survey and current challenges[J]. *Materials*, 2020, 13(24): 5755. doi: [10.3390/ma13245755](https://doi.org/10.3390/ma13245755).
- [5] CHEN Pengfei, YU Xuehui, HAN Xumeng, *et al.* Point-to-box network for accurate object detection via single point supervision[C]. 17th European Conference on Computer Vision, Tel Aviv, Israel, 2022: 51–67. doi: [10.1007/978-3-031-20077-9\\_4](https://doi.org/10.1007/978-3-031-20077-9_4).
- [6] SHENG Hualian, CAI Sijia, ZHAO Na, *et al.* Rethinking IoU-based optimization for single-stage 3D object detection[C]. 17th European Conference on Computer Vision, Tel Aviv, Israel, 2022: 544–561. doi: [10.1007/978-3-031-20077-9\\_32](https://doi.org/10.1007/978-3-031-20077-9_32).
- [7] CHAUDHRY A, ROHRBACH M, ELHOSEINY M, *et al.* Continual learning with tiny episodic memories[EB/OL]. <https://arxiv.org/abs/1902.10486v1>, 2019.
- [8] KIRKPATRICK J, PASCANU R, RABINOWITZ N, *et al.* Overcoming catastrophic forgetting in neural networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(13): 3521–3526. doi: [10.1073/pnas.1611835114](https://doi.org/10.1073/pnas.1611835114).
- [9] ZENKE F, POOLE B, and GANGULI S. Continual learning through synaptic intelligence[C]. The 34th International Conference on Machine Learning, Sydney, Australia, 2017: 3987–3995.
- [10] ALJUNDI R, BABILONI F, ELHOSEINY M, *et al.* Memory aware synapses: Learning what (not) to forget[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 144–161. doi: [10.1007/978-3-030-01219-9\\_9](https://doi.org/10.1007/978-3-030-01219-9_9).
- [11] VAN DE VEN G M and TOLIAS A S. Three scenarios for continual learning[EB/OL]. <https://arxiv.org/abs/1904.07734>, 2019.
- [12] WU Yue, CHEN Yinpeng, WANG Lijuan, *et al.* Large scale

- incremental learning[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 374–382. doi: [10.1109/CVPR.2019.00046](https://doi.org/10.1109/CVPR.2019.00046).
- [13] ZHAI Mengyao, CHEN Lei, and MORI G. Hyper-LifelongGAN: Scalable lifelong learning for image conditioned generation[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021: 2246–2255. doi: [10.1109/CVPR46437.2021.00228](https://doi.org/10.1109/CVPR46437.2021.00228).
- [14] GRAFFIETI G, MALTONI D, PELLEGRINI L, *et al.* Generative negative replay for continual learning[J]. *Neural Networks*, 2023, 162: 369–383. doi: [10.1016/j.neumet.2023.03.006](https://doi.org/10.1016/j.neumet.2023.03.006).
- [15] MARACANI A, MICHIELI U, TOLDO M, *et al.* RECALL: Replay-based continual learning in semantic segmentation[C]. 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 7006–7015. doi: [10.1109/ICCV48922.2021.00694](https://doi.org/10.1109/ICCV48922.2021.00694).
- [16] CERMELLI F, MANCINI M, BULÒ S R, *et al.* Modeling the background for incremental learning in semantic segmentation[C]. x2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 9230–9239. doi: [10.1109/CVPR42600.2020.00925](https://doi.org/10.1109/CVPR42600.2020.00925).
- [17] BALDI P and SADOWSKI P J. Understanding dropout[C]. The 26th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2013: 2814–2822.
- [18] MIRZADEH S I, FARAJTABAR M, and GHASEMZADEH H. Dropout as an implicit gating mechanism for continual learning[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, USA, 2020: 945–951. doi: [10.1109/CVPRW50498.2020.00124](https://doi.org/10.1109/CVPRW50498.2020.00124).
- [19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [20] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al.* DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848. doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184).
- [21] LI Zhizhong and HOIEM D. Learning without forgetting[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(12): 2935–2947. doi: [10.1109/TPAMI.2017.2773081](https://doi.org/10.1109/TPAMI.2017.2773081).
- [22] REBUFFI S A, KOLESNIKOV A, SPERL G, *et al.* iCaRL: Incremental classifier and representation learning[C]. 2017 IEEE conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 5533–5542. doi: [10.1109/CVPR.2017.587](https://doi.org/10.1109/CVPR.2017.587).
- [23] MICHIELI U and ZANUTTIGH P. Incremental learning techniques for semantic segmentation[C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop, Seoul, Korea (South), 2019: 3205–3212. doi: [10.1109/iccvw.2019.00400](https://doi.org/10.1109/iccvw.2019.00400).
- [24] KLINGNER M, BÄR A, DONN P, *et al.* Class-incremental learning for semantic segmentation re-using neither old data nor old labels[C]. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 2020: 1–8. doi: [10.1109/ITSC45102.2020.9294483](https://doi.org/10.1109/ITSC45102.2020.9294483).
- [25] MICHIELI U and ZANUTTIGH P. Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 1114–1124. doi: [10.1109/CVPR46437.2021.00117](https://doi.org/10.1109/CVPR46437.2021.00117).
- [26] LI Junxi, SUN Xian, DIAO Wenhui, *et al.* Class-incremental learning network for small objects enhancing of semantic segmentation in aerial imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5612920. doi: [10.1109/TGRS.2021.3124303](https://doi.org/10.1109/TGRS.2021.3124303).
- [27] DOUILLARD A, CHEN Yifu, DAPOGNY A, *et al.* PLOP: Learning without forgetting for continual semantic segmentation[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021: 4039–4049. doi: [10.1109/CVPR46437.2021.00403](https://doi.org/10.1109/CVPR46437.2021.00403).
- [28] ZHAO Danpei, YUAN Bo, and SHI Zhenwei. Inherit with distillation and evolve with contrast: Exploring class incremental semantic segmentation without exemplar memory[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(10): 11932–11947. doi: [10.1109/TPAMI.2023.3273574](https://doi.org/10.1109/TPAMI.2023.3273574).

杨 静: 男, 副教授, 研究方向为视觉计算.

何 瑶: 女, 硕士生, 研究方向为持续学习.

李 斌: 男, 硕士, 研究方向为持续学习.

李少波: 男, 二级教授, 研究方向为大数据.

胡建军: 男, 美国终身教授, 研究方向为深度学习.

溥 江: 女, 副教授, 研究方向为先进制造技术.

责任编辑: 余 蓉