# 基于表征补偿的连续语义分割*

Chang-Bin Zhang [1*] Jia-Wen Xiao [1†] Xialei Liu [1‡] Ying-Cong Chen [2] Ming-Ming Cheng [1]
[1] TMCC, CS, Nankai University [2] Hong Kong University of Science and Technology

## 摘要

在这项工作中，本文研究了连续语义分割问题。在这个问题上，深度神经网络需要不断地纳入新的类别，而不会产生灾难性遗忘的问题。本文提出使用一种结构化的重参数化机制，称作表征补偿（Representation Compensation, RC）模块，用以解耦新旧知识的表示学习。RC 模块由两个动态演变的分支构成，其中一个分支是冻结的，另一个分支是可训练的。此外，本文在空间和通道两个维度上设计了一个池化立体知识蒸馏策略以进一步提高模型的可塑性和稳定性。本文对两个具有挑战性的连续语义分割场景——连续类分割和连续域分割进行了实验。本文的方法超越了最先进的性能，并且在推理过程中没有引入任何额外的计算开销和参数。本文代码发布于 https://github.com/zhangchbin/RCIL。

## 1. 引言

数据驱动的深度神经网络 [72, 96, 64, 109] 已经取得了很多里程碑式的成果。然而，这些全监督的模型 [23, 93, 16] 只能处理固定数量的类别。在现实世界的应用场景更希望模型可以动态扩展以识别新的类别。一个直接的方法是重建训练集，用所有可用的数据重新训练模型，即所谓的 联合训练. 然而，考虑到重新训练模型的开销、算法的持续发展以及隐私问题，只用当前的数据来更新模型用以识别新类别和旧类别是特别关键的。不仅如此，简单地用新数据对训练好的模型进行微调会导致灾难性遗忘 [48]。因此，本文寻求找到一种连续学习的方式，这可以使一个模型在不发生灾难性遗忘的情况下识别出新的类别。

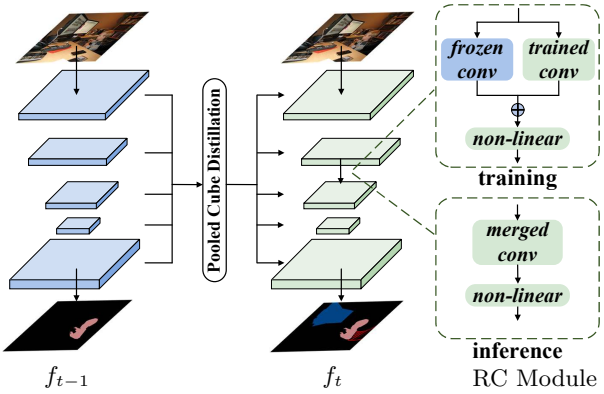



图 1: 本文为避免灾难性遗忘提出的针对连续语义分割的训练框架插图。在本文的方法中设计了两种机制，表征补偿（Representation Compensation, RC）模块以及池化立体蒸馏（PCD）。

在连续语义分割场景中 [62, 8, 27, 63]，给出先前训练好的模型以及新类别的训练数据，该模型应该区分所有看到的类别，包括以前的类别（旧类别）和新类别。然而，为了减少标注带来的开销，新的训练数据通常只包含新类别的标签，同时把旧的类别作为背景处理。直接通过新数据进行学习而不加以额外的设计非常具有挑战性，这很容易引起灾难性遗忘 [48]。

如论文 [51, 48, 28]所述，在新数据上对模型进行微调很可能导致灾难性遗忘，i.e.模型迅速拟合了新类别的数据分布，同时失去了对于旧类别的辨别能力。一些方法 [48, 66, 43, 56, 80, 95, 67] 对模型

*本文为 CVPR'22 论文 [101]的中文翻译版。
†前两位作者具有同等贡献。
‡通讯作者：刘夏雷 (xialei@nankai.edu.cn)。

参数添加正则化以提高其稳定性。然而，所有的参数都会根据新类别的训练数据更新。这非常具有挑战性，因为新旧知识在模型参数中耦合在一起。因此，要保持学习新知识和保留旧知识的脆弱平衡是非常困难的。一些其它的方法 [57, 82, 75, 91, 76, 45] 选择增加模型的容量以更好地权衡稳定性和可塑性，但这以增加网络存储为代价。

在这项工作中，本文提出了一个易于使用的表征补偿模块，旨在记忆旧知识，同时留出额外的容量以学习新知识。受到结构重参数化的启发 [24, 25]，在训练时本文使用两个并行的分支替换网络中的卷积层，称为表征补偿模块。如 图1所示，在训练时，两个并行卷积的输出结果将在非线性激活层之前进行融合。在每一个连续学习步骤的开始，本文等价地将两个并行卷积的参数合并为一个，这个合并后卷积层的参数将被冻结以保存旧的知识。另一个分支是可训练的，并且它继承了前一步中相应分支的参数。表征补偿策略是通过冻结的分支从而记忆旧知识，同时使用可训练的分支为新知识提供额外的容量。重要的是，在推理时该模块并没有带来额外的参数和计算开销。

为了进一步减缓灾难性遗忘的现象 [48]，本文引入了一种知识蒸馏机制 [70]，其在隐藏层之间（如图1所示），称作池化立体蒸馏，它可以抑制局部特征图中误差和噪声带来的负面影响。这篇论文的主要贡献是：

- 本文提出了一个训练期间使用的双分支表征补偿模块，一个分支用于保留旧知识，另一个分支用于适应新的数据。随着任务数量的增加，模型在推理过程中始终保持相同的计算和内存开销。
- 本文分别在 *连续类分割*以及 *连续域分割*两项任务上进行了实验。实验结果表明本文的方法在三个不同的数据集上均取得了最高的精度。

## 2. 相关工作

**语义分割。** 早期的方法侧重于对上下文关系进行建模 [49, 104, 2]。目前的方法更关注多尺度特征聚合 [59, 34, 65, 52, 53, 3, 68, 81]。一些方法 [55, 50, 22, 14, 37, 32, 38] 受到 Non-local [85]的启发，利用注意力机制在图像上下文之间建立联系。另一个研究方向 [94, 15, 61] 旨在融合来自不同感受野的特征。最近，Transformer [7, 97, 26, 110, 86, 105] 在语义分割任务中大放异彩，其专注于多尺度特征融合 [89, 12, 84, 102] 和上下文特征聚合 [79, 58]。

**连续学习。** 连续学习侧重于减轻灾难性遗忘，同时对新学习的类别具有辨别能力。为了解决这个问题，许多工作 [47, 4, 11, 77, 5] 通过基于回放的机制对学到的知识进行回顾。知识可以用多种形式进行存储，例如范例 [83, 9, 4, 73, 11, 6]，原型 [107, 108, 35]或是生成网络 [60]等。虽然这些基于回放的方法通常能取得优异的表现，但它们需要存储空间和存储权限。在更具挑战性的没有回放的场景中，许多方法对正则化进行探索以保留学到的知识，包括知识蒸馏 [10, 51, 69, 74, 18, 28, 21, 100]，对抗训练 [29, 88]，普通的正则化 [98, 48, 66, 43, 56, 80, 95, 67]等。而其它方法关注的是神经网络的容量。其中一个研究方向 [57, 82, 75, 91, 76, 45] 是在学习新知识的同时扩展网络架构。另一个研究方向 [44, 1] 探讨了网络参数的稀疏正则化，其目的是为每个任务激活尽可能少的神经元。这种稀疏正则化减少了网络中的冗余，但也限制了每项任务的学习能力。一些工作提出通过将自监督学习结合到特征提取器来学习更好的表示 [87, 9]，同时解决类别不平衡的问题 [99, 54, 46, 103, 39]。

**连续语义分割。** 连续语义分割仍然是一个亟待解决的问题，主要关注/ 语义分割中的灾难性遗忘 [48]。在这个领域，连续类分割是一个经典的设置，之前的许多工作取得了很大的进展：[92, 41] 探讨了基于回放的方法以回顾旧的知识；MiB [8] 对潜在的类别进行建模以解决背景类歧义的问题；PLOP [27] 对中间层应用知识蒸馏的策略；SDR [63] 利用原型匹配在隐空间表示中执行一致性约束。其它的方法 [78, 31, 95] 利用高维信息、自训练以及模型调整来克服这个问题。此外，连续域分割是 PLOP [27] 提出的一种新颖设置，旨在整合新领域而非新类。与先前方法不同，本文关注的是动态扩展网络，解耦旧类和新类的表示学习。

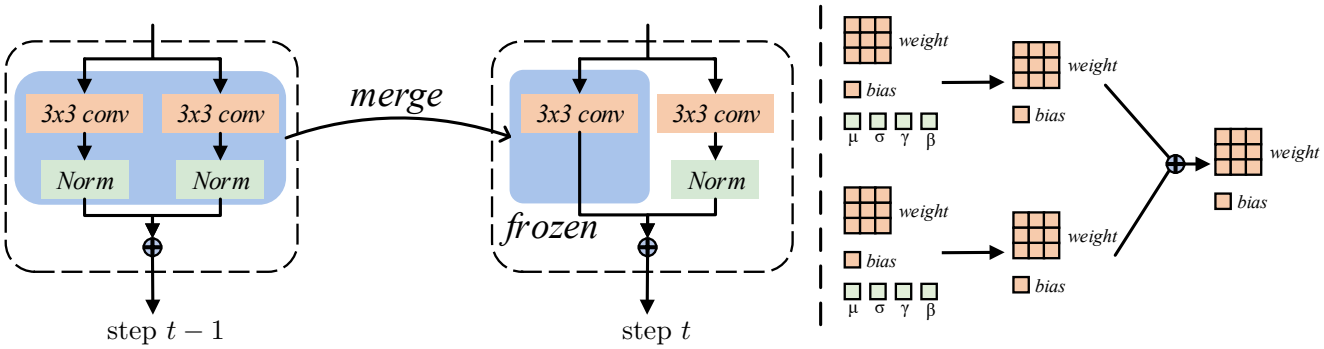



图 2: 表征补偿机制的说明。本文修改 $3 \times 3$ 卷积为两个并行的卷积。两个分支的特征在激活层之前进行聚合。因此，在步骤 $t$ 开始时，在步骤 $t-1$ 训练的两个并行分支可以等效合并为一个卷积层，该卷积层将被冻结并作为步骤 $t$ 的一个分支。步骤 $t$ 中的另一个分支根据第 $t-1$ 步中对应的分支进行初始化。本文在图的右侧演示了合并操作。

## 3. 方法

### 3.1. 准备工作

令 $\mathcal{D} = \{x_i, y_i\}$ 表示训练集，$x_i$ 表示输入图像，$y_i$ 是对应的分割真值图。在具有挑战性的连续学习场景中，本文称每一次在新加入的数据集上的训练为一步。在第 $t$ 步，给定一个模型 $f_{t-1}$、参数 $\theta_{t-1}$ 以及 $\{C_0, C_1 ... C_{t-1}\}$ 个类别，在 $\{\mathcal{D}_0, \mathcal{D}_1 ... \mathcal{D}_{t-1}\}$ 连续训练，当模型遇到新加入的数据集 $\mathcal{D}_t$ 和额外的 $C_t$ 个新类别时，它应该学习区分 $\sum_{n=0}^{t} C_n$ 个类别。当在 $\mathcal{D}_t$ 训练时，之前类别的训练数据是不可见的。此外，为了节省训练开销，$\mathcal{D}_t$ 中的真值图只包含 $C_t$ 个新类别，而之前的类别被标记为 背景。因此，这里有一个紧迫的问题，就是灾难性遗忘。为了验证不同方法的有效性，通常需要多次进行持续学习 e.g.$N$ 步。

### 3.2. 表征补偿网络

为了解耦旧知识的保留与新知识的学习，如 图2所示，本文介绍了提出的表征补偿机制。在绝大多数的深度神经网络中，一个 $3 \times 3$ 的卷积后接归一化层和非线性激活层是一个常见的组件。本文将该结构加以修改，为每个组件添加了一个并行的其后跟有归一化层的 $3 \times 3$ 卷积。两个并行的卷积——归一化层的输出将进行融合，之后由非线性激活层进行校正。更形式化地描述，该结构包括两个并行的卷积层，具有权重 $\{W^0, W^1\}$ 以及偏差 $\{b^0, b^1\}$，其后分别接有两个独立的归一化层。令 $Norm^0 = \{\mu^0, \sigma^0, \gamma^0, \beta^0\}$ 以及 $Norm^1 = \{\mu^1, \sigma^1, \gamma^1, \beta^1\}$ 表示两个归一化层 $Norm^0$ 和 $Norm^1$ 的均值、方差、权重以及偏差。因此，在非线性激活层之前对于输入 $x$ 的计算可以表示为

$$\begin{aligned} \hat{x} &= \sum_{i=0}^{1} Norm_i(W_i x + b_i) \\ &= \sum_{i=0}^{1} (\gamma_i \frac{W_i x + b_i - \mu_i}{\sigma_i} + \beta_i) \\ &= (\sum_{i=0}^{1} \frac{\gamma_i W_i}{\sigma_i}) x + \sum_{i=0}^{1} (\frac{\gamma_i b_i - \gamma_i \mu_i}{\sigma_i} + \beta_i) \\ &= \hat{W} x + \hat{b}. \end{aligned} \tag{1}$$

该公式表明两个并行的分支可以被等价表示为一个带有权重和偏置的分支。本文同样在 图2的右侧展示了该变换。因此，对于该修改后的结构，本文可以等价地将两个分支的参数合并到一个卷积中。更准确地说，在步骤 0 中，所有参数都可以用于训练一个可以区分 $\mathcal{C}_0$ 个类别的模型。对于后续的学习步骤，模型应当对新添加的类别进行分割。在这些连续学习步骤中，网络将通过在之前步骤训练的参数进行初始化，这有利于知识迁移 [8]。在第 $t$ 步开始，因为模型应当避免遗忘之前学到的知识，本文将在上一步训练的两个并行分支的对应卷积层合并到一个卷积层中。合并后分支的参数将被冻结以记忆之前学到的知识，如 图2所示。另一个分支是可训练的以学习新的知识，该分支通过对应的前一个步骤的

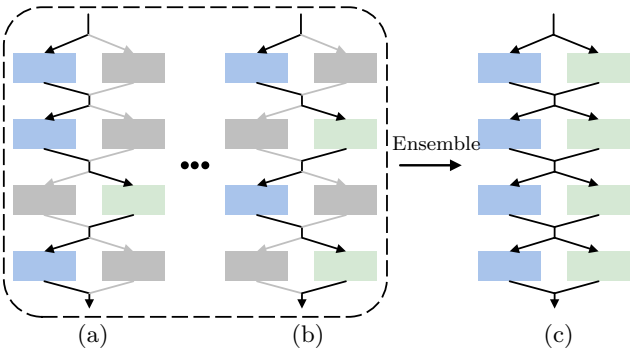



图 3: 本文提出的表征补偿网络的说明。本文的结构 (c) 可以被视作大量子网络的隐式集合 (a), (b), etc. 。蓝色 表示继承自合并后教师网络的被冻结的层。绿色 表示可训练的层。灰色 表示在子网络中忽略的层。

分支进行初始化。此外，本文设计了一种随机路径丢弃策略，该策略被用于聚合来自两个分支的输出 $x_1$ 和 $x_2$。在训练期间，在非线性激活层之前的输出可以表示为

$$\hat{x} = \eta \cdot x_1 + (1 - \eta) \cdot x_2, \tag{2}$$

其中 $\eta$ 是随机通道加权向量，且从集合 $\{0, 0.5, 1\}$ 均匀采样。在推理期间，向量 $\eta$ 中的元素设置为 0.5。实验结果表明，该策略可以带来微小的改进。

RC-**模块有效性分析**。如 图3所示，并行的卷积结构可以被视为许多子网络的隐式集成 [40, 36]。在这些子网络中部分层的参数继承自合并后的教师网络（在上一步训练完毕），且会被冻结。在训练期间，如 [90, 33]，这些被冻结的教师层将对可训练的参数施加正则化，使可训练的层表现如教师模型。在子网络中只有一层可训练的特殊情况下，如 图3(a) 所示，在训练期间，该层将兼顾冻结层的表示和新知识的学习。因此，该机制可以减轻可训练层灾难性遗忘的现象。本文进一步将这种效应推广到一般的子网络，如 图3(b)，其将会使可训练层适应被冻结层的表达。此外，所有子网络均集成在一起，将来自不同子网络的知识整合到一个网络中，如 图3(c) 所示。

### 3.3. 池化立体知识蒸馏

为更进一步减轻对于旧知识的灾难性遗忘，在 PLOP [27] 之后，本文还探索了特征蒸馏。如 图4(a) 所示，PLOP [27] 引入了条带池化 [38] 以整合特征。池化操作在迁移学习中起到了关键的作用。在本文的方法中，本文设计了空间维度上基于平均池化的知识蒸馏。此外，本文在每个位置针对通道维度使用平均池化来保证它们各自的激活强度。总体来说，如 图4(b) 所示，本文在空间维度和通道维度使用平均池化。形式上，对于所有 $L$ 个阶段，本文选择在最后一个非线性激活层前的特征图 $\{X^1, X^2, ..., X^L\}$，包括解码器和骨干网络的所有阶段。对于教师模型和学生模型的特征，本文首先计算每个像素值的平方以保留负信息。之后，本文同时对于空间维度和通道维度施加多尺度平均池化。教师模型和学生模型的特征 $\hat{X}_T^l, \hat{X}_S^l$ 可以通过平均池化操作计算：

$$\begin{aligned} \hat{X}_T^{l,m} &= M \odot [(X_{T,ij}^l)^2] \\ \hat{X}_S^{l,m} &= M \odot [(X_{S,ij}^l)^2], \end{aligned} \tag{3}$$

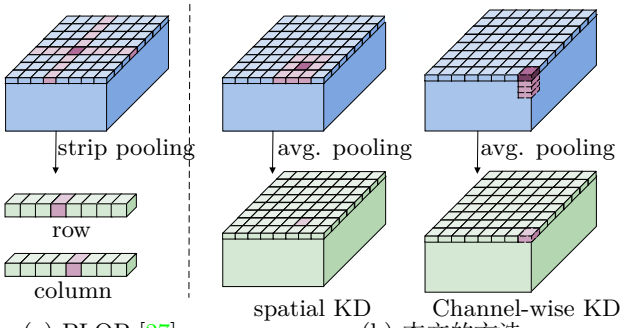
其中 $M$ 表示第 $m$ 个平均池化核，$l$ 表示第 $l$ 个阶段。对于空间维度的平均池化，本文使用多尺度窗口来建模局部区域中像素之间的关系。核 $M$ 的尺寸属于 $\mathcal{M} = \{4, 8, 12, 16, 20, 24\}$ 且步长大小设置为 1。同时对于通道维度的平均池化，本文简单的设置窗口的大小为 3。之后，对于中间层的空间知识蒸馏损失函数 $L_{skd}$ 可以表示为

$$L_{skd} = \frac{1}{L} \frac{1}{|\mathcal{M}|} \sum_{l=1}^{L} \sum_{m=1}^{|\mathcal{M}|} \sqrt{\sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{d=1}^{D} [(\hat{X}_{T,ijd}^{l,m} - \hat{X}_{S,ijd}^{l,m})^2]}, \tag{4}$$

其中 $H, W, D$ 分别表示高度、宽度以及通道数。同样的等式可以应用到 $\mathcal{M} = \{3\}$ 的通道维度以表示 $L_{ckd}$。总的来说，蒸馏目标可以表示为：

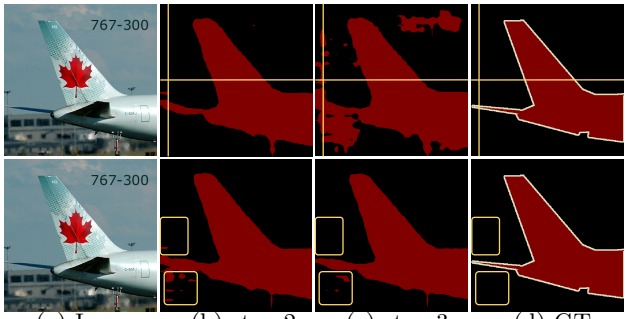
$$L = L_{skd} + L_{ckd} \tag{5}$$

**平均池化** vs. **条带池化**。得益于其强大的特征整合和建模长程依赖的能力，条带池化在许多全监督语义分割模型 [42, 38]中大放异彩。连续分割的性能相对于全监督分割而言仍然有巨大的差距。相




(a) PLOP [27] (b) 本文的方法

图 4: PLOP [27] 与本文提出的池化立体知识蒸馏机制的比较。




(a) Image (b) step 2 (c) step 3 (d) GT

图 5: PLOP [27]中使用的条带池化（第一行）以及本文提出的方法中的平均池化（第二行）的影响。

比于全监督分割，在连续分割场景中，预测结果通常会有更多的噪声或错误。因此，在蒸馏过程中，当使用条带池化对特征进行聚合时，这种长程依赖对于交叉点会引入某些不相关的噪声，引起噪声扩散。这将会引起学生模型预测结果的进一步恶化。在本文的方法中，本文在局部区域使用平均池化来抑制噪声的负面影响。具体来说，因为局部区域的语义信息通常是相近的，当前关键点可以通过聚合局部区域的特征来找到更多的相邻像素以进行决策。因此，当前关键点在局部区域中受到噪声的负面影响会更少。

如 图5(b) 顶部所示的示例，条带池化将噪声和错误引入了教师模型的交叉点。在蒸馏过程中，噪声将进一步传播到学生模型中，引起噪声扩散。对于 图5底部所示的平均池化，关键点将考虑许多相邻的点，从而产生对噪声更鲁棒的聚合特征。

## 4. 实验

在这一节，本文首先展示实验设置的细节，如数据集、约定以及训练细节。之后本文从定量和定性实验中说明本文方法的有效性。

### 4.1. 实验设置

#### 4.1.1 数据集

PASCAL VOC 2012 [30] 是一个常用的数据集，其包含 10582 张训练图像以及 1449 张验证图像，具有 20 个对象类和背景类。ADE20K [106]是一个包含日常生活场景的语义分割数据集。其包含 20210 张训练图像以及 2000 张验证图像，具有 150 个类别。Cityscapes [19] 数据集包含 2975 张训练图像、500 张验证图像以及 1525 张测试图像。其具有来自 21 个城市的 19 个类别。

#### 4.1.2 约定

**连续类分割。** 在连续类分割中，模型经过训练可以在多个步骤中按顺序识别不同的类别。在每一步模型学习一个或几个类别。按照 [8, 27, 63]，本文假设前一步的训练数据不可用，即模型只能访问当前步骤的数据。此外，仅标记当前步骤中要学习的类。所有其他类都被视为 背景。 [8] 提出了两种常用的设置来进行连续类分割，即 不相交和 重叠。在不相交设置中，假定本文知道未来所有的类别，当前训练步骤中的图像不包含任何未来的类别。重叠设置则更加真实。它允许可能的未来类别出现在当前的训练图像中。

本文在 PASCAL VOC 2012 [30]以及 ADE20K [106]上进行了连续类分割的实验。按照 [8, 27, 63]，如 第3.1节所定义，本文称每一次在新加入数据集上的训练为一步。形式化地，$X$-$Y$ 表示本文实验中的连续设置，$X$ 表示在第一步中本文需要训练的类别数。在随后的每个学习步骤中，新加入的数据集包含 $Y$ 个类别。在 PASCAL VOC 2012 [30]，本文在三种设置上进行了实验，包括 15-5（2 步），15-1（6 步）以及 10-1（11 步）。例如，15-1 表示在第一步中本文在初始的 15 个目标

类别上训练模型。在随后的 5 步中，模型预计将在新数据集上进行训练，其中每个数据集都包含一个新添加的类。因此，该模型可以在最后一步区分 20 个对象类。在 ADE20K [106]上，本文使用了四种设置，分别是 100-50（2 步），50-50（3 步），100-10（6 步），以及 100-5（11 步）。

**连续域分割。** 该任务由 [27]提出。不同于连续类分割，这种设置是为了处理领域偏移现象而非聚合新的类。在现实世界场景中，领域偏移也可能会频繁发生。本文假定在不同的领域中的类别是相同的。旧域的训练数据在训练新领域的数据时是不可见的。本文在 Cityscapes [19]数据集上进行了连续域分割实验。按照 PLOP [27]，本文认为每个城市的训练数据为一个领域。本文同样应用了三种设置，11-5 （3 步），11-1 （11 步）以及 1-1 （21 步）。在这些实验设置中，本文使用与连续类分割相同的标记，但是每一步都会添加新的领域数据（城市）而不是类。

#### 4.1.3 实现细节

根据 [8, 63, 27]，本文使用 Deeplab-v3 [13] 架构，以 ResNet-101[36]作为骨干网络。DeepLab-v3 的输出步长设置为 16。如前述方法，本文同样在通过 ImageNet [20]预训练好的骨干网络上使用了原地激活批归一化 [71]。本文使用 MiB [8] 提出的损失函数来辅助训练过程。同时本文使用与 [8, 27, 63]相同的训练策略。具体来说，本文使用了相同的数据增强，e.g.水平翻转和随机裁剪。对于所有实验，批大小设置为 24。对于第一个训练步骤本文设置初始学习率为 0.02，对于随后的连续学习步骤设置为 0.001。学习率按照 poly 提出的策略进行调整。本文使用随机梯度下降优化器在每一个步骤训练模型，训练轮数为 30（PASCAL VOC 2012 [30]），50（Cityscapes [19]）以及 60（ADE20K [106]）。按照 [8, 63, 27]，本文也使用 20% 的训练数据进行验证。本文数据集原始的验证集上报告了平均交并比（mIoU）。

### 4.2. 连续类分割

PASCAL VOC 2012 **数据集。** 与之前的方法 [8, 27, 63]一样采用相同的设置，本文在不同的连续学习设置上进行了实验，如 15-5，15-1 以及 10-1。如 表1所示，本文报告了在最后一个步骤的实验结果。简单的微调方法会引起灾难性遗忘的现象。模型会很快遗忘旧知识，同时也不能很好地学习新知识。实验结果表明本文的方法本文的方法在重叠和不相交两种设置上均显著改善了分割表现。特别是在具有挑战性的 15-1 设置中，本文的方法在 mIoU 指标上分别提升了%6.0（不相交）和%4.8（重叠），取得了最好的表现。本文同样展示了不同方法在每一步的表现，如 图6a和图6b所示。这表明本文的方法能减少连续学习过程中对于旧知识的遗忘。在 表1中，本文同样报告了在旧类别和新类别上的表现。对于所有设置，旧类别的表现得到显著改善。这得益于表征补偿模块以及知识蒸馏机制，其可以有效保留旧知识。另一方面，本文提出的表征模块以及蒸馏机制为学习新知识流出了空间。在 第4.4节中，本文将进一步分析这两种机制的有效性。在如 图7所示的 15-1 重叠设置中，本文进一步展示了不同方法的定性结果。

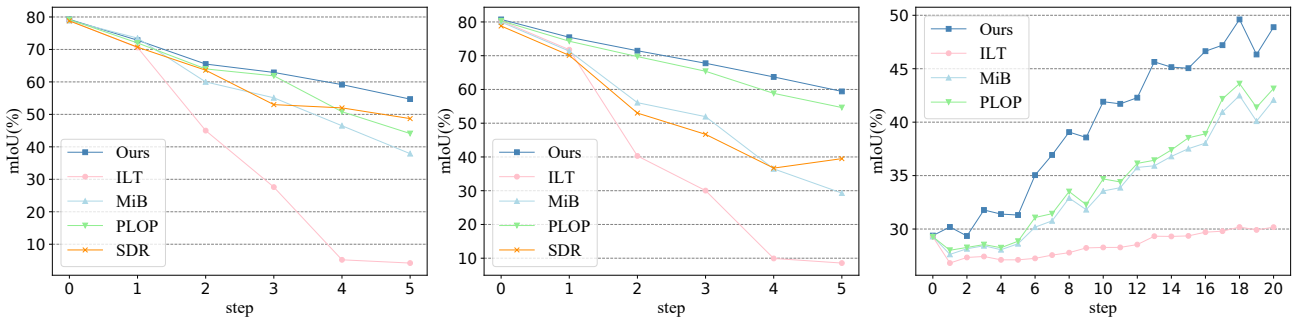
ADE20K **数据集。** 为了验证本文方法的有效性，本文在极具挑战性的语义分割数据集 ADE20K [106]上进行了实验。实验结果如 表2以及 表3所示。在不同的连续学习任务如 100-50、100-10 以及 50-50 上，本文的方法比目前最好的方法平均提升了约 1.4%。为了进一步验证本文的方法，本文同样在更具挑战性的场景，包含了 11 步的 100-5 上进行了实验。在该场景中，本文的方法同样达到了最好的性能，在 mIoU 这个指标上超过先前方法 0.9%，如 表3所示。取得的改进归功于本文提出的表征补偿模块以及池化立体蒸馏机制。

### 4.3. 连续域分割

在连续语义分割的背景下，除了需要细分新的类别，提高对于新域的处理能力也具有重要的意义。按照 [27]，本文在连续域语义分割数据集 Cityscapes [19]上进行了实验。Cityscapes [19]中的每一个城市可以被认做一个域，这在域适应语义分

| | 15-5 (2 步) | | | | | | 15-1 (6 步) | | | | | | 10-1 (11 步) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **不相交** | | | **重叠** | | | **不相交** | | | **重叠** | | | **不相交** | | | **重叠** | | |
| Method | 0-15 | 16-20 | 全部 | 0-15 | 16-20 | 全部 | 0-15 | 16-20 | 全部 | 0-15 | 16-20 | 全部 | 0-10 | 11-20 | 全部 | 0-10 | 11-20 | 全部 |
| Fine-tuning | 5.7 | 33.6 | 12.3 | 6.6 | 33.1 | 12.9 | 4.6 | 1.8 | 3.8 | 4.6 | 1.8 | 3.9 | 6.3 | 1.1 | 3.8 | 6.4 | 1.2 | 3.9 |
| Joint | 79.8 | 72.6 | 78.2 | 79.8 | 72.6 | 78.2 | 79.8 | 72.6 | 78.2 | 79.8 | 72.6 | 78.2 | 78.2 | 78.0 | 78.2 | 78.2 | 78.0 | 78.2 |
| LwF [51] | 60.4 | 37.4 | 54.9 | 60.8 | 36.6 | 55.0 | 5.8 | 3.6 | 5.3 | 6.0 | 3.9 | 5.5 | 7.2 | 1.2 | 4.3 | 8.0 | 2.0 | 4.8 |
| ILT [62] | 64.9 | 39.5 | 58.9 | 67.8 | 40.6 | 61.3 | 8.6 | 5.7 | 7.9 | 9.6 | 7.8 | 9.2 | 7.3 | 3.2 | 5.4 | 7.2 | 3.7 | 5.5 |
| MiB [8] | 73.0 | 43.3 | 65.9 | 76.4 | 49.4 | 70.0 | 48.4 | 12.9 | 39.9 | 38.0 | 13.5 | 32.2 | 9.5 | 4.1 | 6.9 | 20.0 | 20.1 | 20.1 |
| SDR [63] | 74.6 | 44.1 | 67.3 | 76.3 | 50.2 | 70.1 | 59.4 | 14.3 | 48.7 | 47.3 | 14.7 | 39.5 | 17.3 | 11.0 | 14.3 | 32.4 | 17.1 | 25.1 |
| PLOP [27] | 71.0 | 42.8 | 64.3 | 75.7 | 51.7 | 70.1 | 57.9 | 13.7 | 46.5 | 65.1 | 21.1 | 54.6 | 9.7 | 7.0 | 8.4 | 44.0 | 15.5 | 30.5 |
| Ours | 75.0 | 42.8 | 67.3 | 78.8 | 52.0 | 72.4 | 66.1 | 18.2 | 54.7 | 70.6 | 23.7 | 59.4 | 30.6 | 4.7 | 18.2 | 55.4 | 15.1 | 34.3 |

表 1: 对于不同连续类分割场景在 PASCAL VOC 2012 数据集上最后一步的平均交并比（%）。**红色** 表示最高的结果，**蓝色** 表示次高的结果。




(a) 15-1 不相交 PASCAL VOC 2012　(b) 15-1 重叠 PASCAL VOC 2012　(c) 1-1 Cityscapes

图 6: 三种实验设置中每一步的 mIoU (%)。(a)(b) 是连续类分割的设置。(c) 是连续域分割的设置。

| | 100-50 (2 步) | | | 100-10 (6 步) | | | | | | | 50-50 (3 步) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **方法** | 1-100 | 101-150 | 全部 | 1-100 | 101-110 | 111-120 | 121-130 | 131-140 | 141-150 | 全部 | 1-50 | 51-100 | 101-150 | 全部 |
| ILT [62] | 18.3 | 14.8 | 17.0 | 0.1 | 0.0 | 0.1 | 0.9 | 4.1 | 9.3 | 1.1 | 13.6 | 12.3 | 0.0 | 9.7 |
| MiB [8] | 40.7 | 17.7 | 32.8 | 38.3 | 12.6 | 10.6 | 8.7 | 9.5 | 15.1 | 29.2 | 45.3 | 26.1 | 17.1 | 29.3 |
| PLOP [27] | 41.9 | 14.9 | 32.9 | 40.6 | 15.2 | 16.9 | 18.7 | 11.9 | 7.9 | 31.6 | 48.6 | 30.0 | 13.1 | 30.4 |
| Ours | 42.3 | 18.8 | 34.5 | 39.3 | 14.6 | 26.3 | 23.2 | 12.1 | 11.8 | 32.1 | 48.3 | 31.3 | 18.7 | 32.5 |
| Joint | 44.3 | 28.2 | 38.9 | 44.3 | 26.1 | 42.8 | 26.7 | 28.1 | 17.3 | 38.9 | 51.1 | 38.3 | 28.2 | 38.9 |

表 2: 对于不同重叠的连续学习场景在 ADE20K 数据集上最后一步的平均交并比（%）。**红色** 表示最高的结果，**蓝色** 表示次高的结果。

割任务 [17] 中被广泛使用。在该场景中，本文不考虑不同域之间的类别差异。如 表4所示，实验结果表明在所有设置中本文的方法相比于之前的方法取得了更高的平均交并比。本文的方法在具有挑战性的包含 21 个学习步骤的 1-1 设置上相比于之前的方法提升了 3.7%，达到了最好的结果。对于该设置，在 图6c中本文展示了模型在每一步的表现。因为 MiB [8]旨在解决语义漂移的问题，而该问题并未在连续域分割中出现，因此 MiB [8]表现略差于微调。这些实验表明本文的方法在连续域语义分割上同样有效，其得益于模型在保持旧知识的同时能够学习新知识。

| 方法 | 1-100 | 101-150 | 全部 |
|---|---|---|---|
| ILT [62] | 0.1 | 1.3 | 0.5 |
| MiB [8] | 36.0 | 5.6 | 25.9 |
| PLOP [27] | 39.1 | 7.8 | 28.7 |
| 本文的方法 | 38.5 | 11.5 | 29.6 |

表 3: ADE20K 数据集上 100-5 重叠设置的最终平均交并比（%）。

| 方法 | 11-5 (3 步) | 11-1 (11 步) | 1-1 (21 步) |
|---|---|---|---|
| Fine-tuning | 61.7 | 60.4 | 42.9 |
| LwF [51] | 59.7 | 57.3 | 33.0 |
| LwF-MC [69] | 58.7 | 57.0 | 31.4 |
| ILT [62] | 59.1 | 57.8 | 30.1 |
| MiB [8] | 61.5 | 60.0 | 42.2 |
| PLOP [27] | 63.5 | 62.1 | 45.2 |
| 本文的方法 | 64.3 | 63.0 | 48.9 |

表 4: 在 Cityscapes [19]上连续域语义分割的最终平均交并比（%）。

| MiB‡[8] | RC | Strip [38] | S-KD | C-KD | 15-1 |
|---|---|---|---|---|---|
| ✓ | | | | | 36.1 |
| ✓ | ✓ | | | | 43.0 |
| ✓ | ✓ | | ✓ | | 58.3 |
| ✓ | ✓ | | | ✓ | 58.4 |
| ✓ | | | ✓ | ✓ | 57.8 |
| ✓ | ✓ | ✓ | | | 57.9 |
| ✓ | ✓ | | ✓ | ✓ | 59.4 |

表 5: 最终的表征补偿模块（RC）与空间维度（S-KD）和通道维度（C-KD）池化立体蒸馏机制消融实验的平均交并比（%）。实验在 15-1 重叠设置的 PASCAL VOC 2012 上进行。† 表明基线方法经自适应因子优化 [27]。

### 4.4. 消融实验

在这一节中，本文首先分析所提出的表征补偿模块以及池化立体蒸馏机制的有效性。随后本文将讨论在连续学习场景中类别学习顺序的鲁棒性。

**表征补偿**。本文在 PASCAL VOC 2012 [30]上进行了实验。

如 表5所示，本文提出的表征补偿模块相比于基线方法 MiB [8]取得了 7% 的提升。使用该模块后，本文的方法达到了最先进的性能。本文认为这种性能得益于本文方法具有记忆旧知识的同时允许学习新知识的能力。在本文的方法中，合并卷积层以及冻结参数的操作旨在减轻模型对于旧知识的遗忘。因此，在 表6中，本文进一步研究了这两种操作的有效性。具体来说，基于简单的并行卷积分支（Parallel-Conv），合并卷积层（Merge）以及冻结参数（Frozen）的操作能带来 2.7% 的提升。实验结果表明模型受益于之前步骤中冻结的知识。

| 并行卷积 | 合并 | 冻结 | Drop-path | 15-1 |
|---|---|---|---|---|
| ✓ | | | | 40.1 |
| ✓ | ✓ | | | 42.0 |
| ✓ | ✓ | ✓ | | 42.8 |
| ✓ | ✓ | ✓ | ✓ | 43.0 |

表 6: 表征补偿模块的消融实验。所有的实验均在不使用池化立体蒸馏的情况下于 PASCAL VOC 2012 上进行。

| 不使用池化 | GAP | 最大池化 | 条带池化 | 平均池化 |
|---|---|---|---|---|
| 52.0 | 36.1 | 48.0 | 54.6 | 56.1 |

表 7: 蒸馏机制中不同池化方法的比较。所有的实验均在 15-1 重叠设置的 PASCAL VOC 2012 进行，且使用了 PLOP 的框架。GAP 表示全局平均池化。

**蒸馏机制**。在 表5中，本文研究了在空间和通道维度上知识蒸馏机制的重要性。在空间和通道维度上，知识蒸馏的表现相似，在 mIoU 这个指标上能够超越基线方法 15.3%。结合表征补偿模块，同时使用上述两种知识蒸馏机制能够达到目前最高的精度。本文进一步比较了知识蒸馏机制中不同池化方式的有效性，如 表7所示。实验结果表明平均池化相比较条带池化而言能够在 mIoU 指标上提升 1.5%。

**类别顺序的鲁棒性**。在连续语义分割场景中，流水线中的类别顺序至关重要。为了验证类别顺序的鲁棒性，本文对五种不同的类别顺序进行实验，包括四个随机顺序和原始的升序顺序。在 图8中，本文展示了不同方法 [62, 8, 63, 27]的平均表现以及标准方差。实验结果表明本文的方法相较于之前方法而言在不同的类别顺序上更加鲁棒。

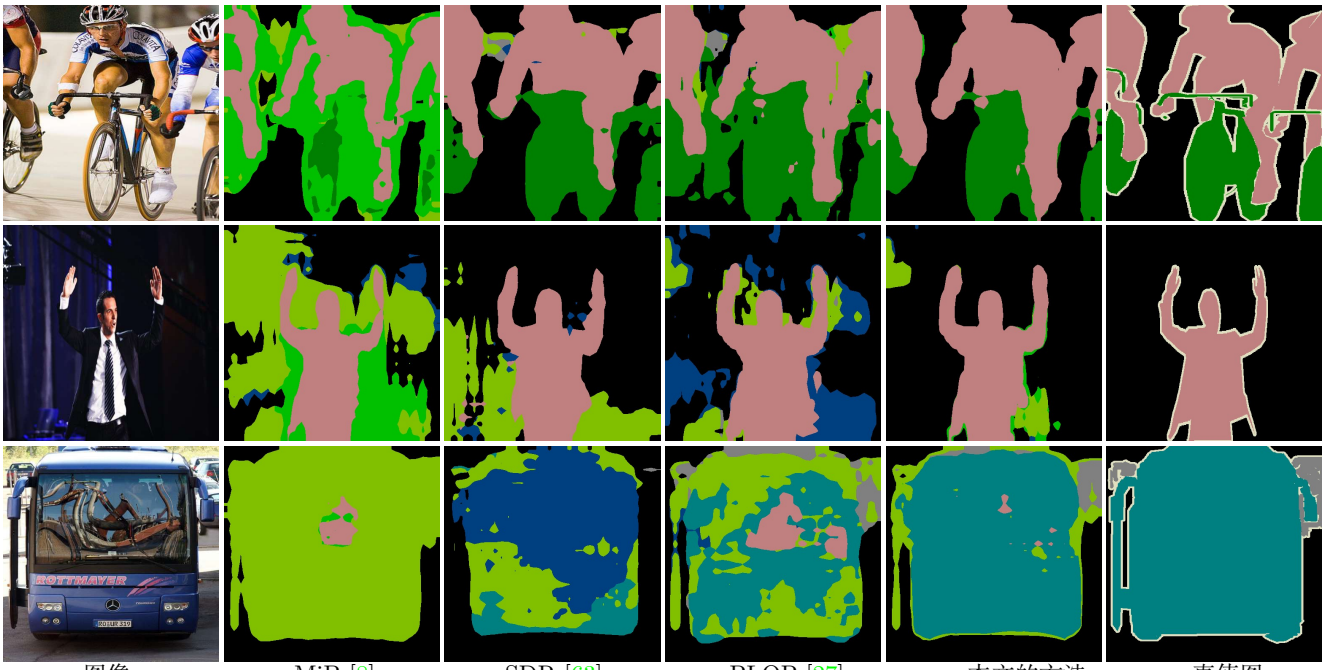
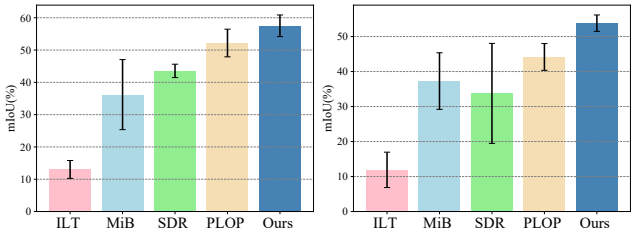



图 7: 不同方法之间的定性比较。所有预测结果均来自 15-1 重叠设置的最后一步。




(a) 15-1 重叠　　(b) 15-1 不相交

图 8: 不同连续学习类别顺序下的平均表现和标准方差。

## 5. 结论以及局限

在这项工作中，为了在记忆旧类知识的同时创造学习新类别的空间，本文提出了表征补偿模块，其可以在不增加额外推理开销的前提下动态扩展网络。同时，为了进一步减轻对于旧知识的遗忘，本文提出了在空间和通道维度的池化立体蒸馏机制。本文在两个常用的基准，连续类分割以及连续域分割上进行了实验，结果表明本文的方法达到了最好的性能。

虽然本文提出的两个部分能达到目前最好的效果，但其在包含多个步骤的连续学习过程中表现较差，如 表1 所示的 10-1 设置。在这些具有挑战性的场景中，如何提高模型的性能还有很长的路要走。同时，本文提出的方法在训练期间需要更多的计算开销。

**鸣谢**该项目得到中国国家重点研发计划（批准号：2018AAA0100400）和国家自然科学基金（NO.61922046）的支持。

## 参考文献

[1] D. Abati, J. Tomczak, T. Blankevoort, S. Calderara, R. Cucchiara, and B. E. Bejnordi. Conditional channel gated networks for task-aware continual learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 3931–3940, 2020. 410

[2] A. Arnab, S. Jayasumana, S. Zheng, and P. H. Torr. Higher order conditional random fields in deep neural networks. In Eur. Conf. Comput. Vis., 2016. 410

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell., 39(12):2481–2495, 2017. 410

[4] J. Bang, H. Kim, Y. Yoo, J.-W. Ha, and J. Choi. Rainbow memory: Continual learning with a mem-

ory of diverse samples. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[5] E. Belouadah and A. Popescu. Il2m: Class incremental learning with dual memory. In Int. Conf. Comput. Vis., pages 583–592, 2019. 410

[6] P. Buzzega, M. Boschini, A. Porrello, D. Abati, and S. Calderara. Dark experience for general continual learning: a strong, simple baseline. In Adv. Neural Inform. Process. Syst., 2020. 410

[7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In Eur. Conf. Comput. Vis., pages 213–229, 2020. 410

[8] F. Cermelli, M. Mancini, S. R. Bulo, E. Ricci, and B. Caputo. Modeling the background for incremental learning in semantic segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9233–9242, 2020. 409, 410, 411, 413, 414, 415, 416, 417

[9] H. Cha, J. Lee, and J. Shin. Co2l: Contrastive continual learning. In Int. Conf. Comput. Vis., pages 9516–9525, 2021. 410

[10] A. Chaudhry, P. K. Dokania, T. Ajanthan, and P. H. Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In Eur. Conf. Comput. Vis., 2018. 410

[11] A. Chaudhry, A. Gordo, P. K. Dokania, P. Torr, and D. Lopez-Paz. Using hindsight to anchor past knowledge in continual learning. In The National Conference on Artificial Intelligence (AAAI), 2021. 410

[12] C.-F. Chen, Q. Fan, and R. Panda. Crossvit: Cross-attention multi-scale vision transformer for image classification. In Int. Conf. Comput. Vis., 2021. 410

[13] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell., 40(4):834–848, 2017. 414

[14] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille. Attention to scale: Scale-aware semantic image segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., 2016. 410

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Eur. Conf. Comput. Vis., 2018. 410

[16] L.-Z. Chen, Z. Lin, Z. Wang, Y.-L. Yang, and M.-M. Cheng. Spatial information guided convolution for real-time rgbd semantic segmentation. IEEE Trans. Image Process., 30:2313–2324, 2021. 409

[17] Y.-H. Chen, W.-Y. Chen, Y.-T. Chen, B.-C. Tsai, Y.-C. Frank Wang, and M. Sun. No more discrimination: Cross city adaptation of road scene segmenters. In Int. Conf. Comput. Vis., pages 1992–2001, 2017. 415

[18] A. Cheraghian, S. Rahman, P. Fang, S. K. Roy, L. Petersson, and M. Harandi. Semantic-aware knowledge distillation for few-shot class-incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[19] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In IEEE Conf. Comput. Vis. Pattern Recog., pages 3213–3223, 2016. 413, 414, 416

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In IEEE Conf. Comput. Vis. Pattern Recog., pages 248–255, 2009. 414

[21] P. Dhar, R. V. Singh, K.-C. Peng, Z. Wu, and R. Chellappa. Learning without memorizing. In IEEE Conf. Comput. Vis. Pattern Recog., 2019. 410

[22] H. Ding, X. Jiang, B. Shuai, A. Q. Liu, and G. Wang. Context contrasted feature and gated multi-scale aggregation for scene segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., pages 2393–2402, 2018. 410

[23] H. Ding, X. Jiang, B. Shuai, A. Q. Liu, and G. Wang. Semantic segmentation with context encoding and multi-path decoding. IEEE Trans. Image Process., 29:3520–3533, 2020. 409

[24] X. Ding, Y. Guo, G. Ding, and J. Han. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In Int. Conf. Comput. Vis., October 2019. 410

[25] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun. Repvgg: Making vgg-style convnets great again. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[26] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani,

M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Int. Conf. Learn. Represent., 2021. 410

[27] A. Douillard, Y. Chen, A. Dapogny, and M. Cord. Plop: Learning without forgetting for continual semantic segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 409, 410, 412, 413, 414, 415, 416, 417

[28] A. Douillard, M. Cord, C. Ollion, T. Robert, and E. Valle. Podnet: Pooled outputs distillation for small-tasks incremental learning. In Eur. Conf. Comput. Vis., volume 12365, pages 86–102, 2020. 409, 410

[29] S. Ebrahimi, F. Meier, R. Calandra, T. Darrell, and M. Rohrbach. Adversarial continual learning. In Adv. Neural Inform. Process. Syst., 2020. 410

[30] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. 413, 414, 416

[31] J. Frey, H. Blum, F. Milano, R. Siegwart, and C. Cadena. Continual learning of semantic segmentation using complementary 2d-3d data representations. arXiv preprint arXiv:2111.02156, 2021. 410

[32] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu. Dual attention network for scene segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., pages 3146–3154, 2019. 410

[33] S. Fu, Z. Li, J. Xu, M.-M. Cheng, Z. Liu, and X. Yang. Interactive knowledge distillation. arXiv preprint arXiv:2007.01476, 2020. 412

[34] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. In IEEE Conf. Comput. Vis. Pattern Recog., pages 447–456, 2015. 410

[35] T. L. Hayes, K. Kafle, R. Shrestha, M. Acharya, and C. Kanan. Remind your neural network to prevent catastrophic forgetting. In Eur. Conf. Comput. Vis., pages 466–483, 2020. 410

[36] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In IEEE Conf. Comput. Vis. Pattern Recog., 2016. 412, 414

[37] S. Hong, J. Oh, H. Lee, and B. Han. Learning transferrable knowledge for semantic segmentation with deep convolutional neural network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 3204–3212, 2016. 410

[38] Q. Hou, L. Zhang, M.-M. Cheng, and J. Feng. Strip pooling: Rethinking spatial pooling for scene parsing. In IEEE Conf. Comput. Vis. Pattern Recog., pages 4003–4012, 2020. 410, 412, 416

[39] S. Hou, X. Pan, C. C. Loy, Z. Wang, and D. Lin. Learning a unified classifier incrementally via rebalancing. In IEEE Conf. Comput. Vis. Pattern Recog., pages 831–839, 2019. 410

[40] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger. Deep networks with stochastic depth. In Eur. Conf. Comput. Vis., pages 646–661, 2016. 412

[41] Z. Huang, W. Hao, X. Wang, M. Tao, J. Huang, W. Liu, and X.-S. Hua. Half-real half-fake distillation for class-incremental semantic segmentation. arXiv preprint arXiv:2104.00875, 2021. 410

[42] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu. Ccnet: Criss-cross attention for semantic segmentation. In Int. Conf. Comput. Vis., pages 603–612, 2019. 412

[43] A. Iscen, J. Zhang, S. Lazebnik, and C. Schmid. Memory-efficient incremental learning through feature adaptation. In Eur. Conf. Comput. Vis., pages 699–715, 2020. 409, 410

[44] S. Jung, H. Ahn, S. Cha, and T. Moon. Continual learning with node-importance based adaptive group sparse regularization. In Adv. Neural Inform. Process. Syst., 2020. 410

[45] M. Kanakis, D. Bruggemann, S. Saha, S. Georgoulis, A. Obukhov, and L. Van Gool. Reparameterizing convolutions for incremental multi-task learning without task interference. In Eur. Conf. Comput. Vis., pages 689–707, 2020. 410

[46] C. D. Kim, J. Jeong, and G. Kim. Imbalanced continual learning with partitioning reservoir sampling. In Eur. Conf. Comput. Vis., pages 411–428, 2020. 410

[47] C. D. Kim, J. Jeong, S. Moon, and G. Kim. Continual learning on noisy data streams via self-purified replay. In Int. Conf. Comput. Vis., pages 537–547, 2021. 410

[48] 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, 114(13):3521–3526, 2017. 409, 410

[49] V. Koltun et al. Efficient inference in fully connected crfs with gaussian edge potentials. In Adv. Neural Inform. Process. Syst., 2011. 410

[50] X. Li, Z. Zhong, J. Wu, Y. Yang, Z. Lin, and H. Liu. Expectation-maximization attention networks for semantic segmentation. In Int. Conf. Comput. Vis., pages 9167–9176, 2019. 410

[51] Z. Li and D. Hoiem. Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell., 40(12):2935–2947, 2017. 409, 410, 415, 416

[52] D. Lin, Y. Ji, D. Lischinski, D. Cohen-Or, and H. Huang. Multi-scale context intertwining for semantic segmentation. In Eur. Conf. Comput. Vis., pages 603–619, 2018. 410

[53] G. Lin, A. Milan, C. Shen, and I. Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., 2017. 410

[54] Q. Liu, O. Majumder, A. Achille, A. Ravichandran, R. Bhotika, and S. Soatto. Incremental meta-learning via indirect discriminant alignment. In Eur. Conf. Comput. Vis., 2020. 410

[55] S. Liu, S. De Mello, J. Gu, G. Zhong, M.-H. Yang, and J. Kautz. Learning affinity via spatial propagation networks. In Adv. Neural Inform. Process. Syst., 2017. 410

[56] Y. Liu, S. Parisot, G. Slabaugh, X. Jia, A. Leonardis, and T. Tuytelaars. More classifiers, less forgetting: A generic multi-classifier paradigm for incremental learning. In Eur. Conf. Comput. Vis., pages 699–716, 2020. 409, 410

[57] Y. Liu, B. Schiele, and Q. Sun. Adaptive aggregation networks for class-incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[58] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In Int. Conf. Comput. Vis., 2021. 410

[59] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., 2015. 410

[60] A. Maracani, U. Michieli, M. Toldo, and P. Zanuttigh. Recall: Replay-based continual learning in semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 7026–7035, 2021. 410

[61] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In Eur. Conf. Comput. Vis., pages 552–568, 2018. 410

[62] U. Michieli and P. Zanuttigh. Incremental learning techniques for semantic segmentation. In Int. Conf. Comput. Vis. Worksh., 2019. 409, 415, 416

[63] U. Michieli and P. Zanuttigh. Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 409, 410, 413, 414, 415, 416, 417

[64] Y. Nirkin, L. Wolf, and T. Hassner. Hyperseg: Patch-wise hypernetwork for real-time semantic segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., pages 4061–4070, 2021. 409

[65] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In Int. Conf. Comput. Vis., pages 1520–1528, 2015. 410

[66] P. Pan, S. Swaroop, A. Immer, R. Eschenhagen, R. E. Turner, and M. E. Khan. Continual deep learning by functional regularisation of memorable past. In Adv. Neural Inform. Process. Syst., 2020. 409, 410

[67] D. Park, S. Hong, B. Han, and K. M. Lee. Continual learning by asymmetric loss approximation with single-side overestimation. In Int. Conf. Comput. Vis., pages 3335–3344, 2019. 409, 410

[68] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun. Large kernel matters–improve semantic segmentation by global convolutional network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 4353–4361, 2017. 410

[69] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert. icarl: Incremental classifier and representation learning. In IEEE Conf. Comput. Vis. Pattern Recog., 2017. 410, 416

[70] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio. Fitnets: Hints for thin deep nets. In Int. Conf. Learn. Represent., 2015. 410

[71] S. Rota Bulò, L. Porzi, and P. Kontschieder. In-place activated batchnorm for memory-optimized training

of dnns. In IEEE Conf. Comput. Vis. Pattern Recog., 2018. 414

[72] M. Seyedhosseini and T. Tasdizen. Semantic image segmentation with contextual hierarchical models. IEEE Trans. Pattern Anal. Mach. Intell., 38(5):951–964, 2016. 409

[73] D. Shim, Z. Mai, J. Jeong, S. Sanner, H. Kim, and J. Jang. Online class-incremental continual learning with adversarial shapley value. In The National Conference on Artificial Intelligence (AAAI), 2021. 410

[74] C. Simon, P. Koniusz, and M. Harandi. On learning the geodesic path for incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[75] P. Singh, P. Mazumder, P. Rai, and V. P. Namboodiri. Rectification-based knowledge retention for continual learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 15282–15291, 2021. 410

[76] P. Singh, V. K. Verma, P. Mazumder, L. Carin, and P. Rai. Calibrating cnns for lifelong learning. In Adv. Neural Inform. Process. Syst., volume 33, 2020. 410

[77] J. Smith, Y.-C. Hsu, J. Balloch, Y. Shen, H. Jin, and Z. Kira. Always be dreaming: A new approach for data-free class-incremental learning. In Int. Conf. Comput. Vis., 2021. 410

[78] S. Stan and M. Rostami. Unsupervised model adaptation for continual semantic segmentation. In The National Conference on Artificial Intelligence (AAAI), 2022. 410

[79] R. Strudel, R. Garcia, I. Laptev, and C. Schmid. Segmenter: Transformer for semantic segmentation. In Int. Conf. Comput. Vis., 2021. 410

[80] X. Tao, X. Chang, X. Hong, X. Wei, and Y. Gong. Topology-preserving class-incremental learning. In Eur. Conf. Comput. Vis., pages 254–270, 2020. 409, 410

[81] Z. Tian, T. He, C. Shen, and Y. Yan. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation. In IEEE Conf. Comput. Vis. Pattern Recog., pages 3126–3135, 2019. 410

[82] V. K. Verma, K. J. Liang, N. Mehta, P. Rai, and L. Carin. Efficient feature transformations for discriminative and generative continual learning. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[83] E. Verwimp, M. De Lange, and T. Tuytelaars. Rehearsal revealed: The limits and merits of revisiting samples in continual learning. In Int. Conf. Comput. Vis., pages 9385–9394, 2021. 410

[84] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Int. Conf. Comput. Vis., 2021. 410

[85] X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. In IEEE Conf. Comput. Vis. Pattern Recog., pages 7794–7803, 2018. 410

[86] Y. Wang, Z. Xu, X. Wang, C. Shen, B. Cheng, H. Shen, and H. Xia. End-to-end video instance segmentation with transformers. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[87] G. Wu, S. Gong, and P. Li. Striking a balance between stability and plasticity for class-incremental learning. In Int. Conf. Comput. Vis., pages 1124–1133, 2021. 410

[88] Y. Xiang, Y. Fu, P. Ji, and H. Huang. Incremental learning using conditional adversarial networks. In Int. Conf. Comput. Vis., pages 6619–6628, 2019. 410

[89] E. Xie, W. Wang, Z. Yu, A. Anandkumar3, J. Alvarez, and P. Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In Int. Conf. Comput. Vis., 2021. 410

[90] C. Xu, W. Zhou, T. Ge, F. Wei, and M. Zhou. Bert-of-theseus: Compressing bert by progressive module replacing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 7859–7869, 2020. 412

[91] S. Yan, J. Xie, and X. He. Der: Dynamically expandable representation for class incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[92] S. Yan, J. Zhou, J. Xie, S. Zhang, and X. He. An em framework for online incremental learning of semantic segmentation. In ACM Int. Conf. Multimedia, 2021. 410

[93] K. Yang, X. Hu, and R. Stiefelhagen. Is context-aware cnn ready for the surroundings? panoramic semantic segmentation in the wild. IEEE Trans. Image Process., 30:1866–1881, 2021. 409

[94] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang. Denseaspp for semantic segmentation in street

scenes. In IEEE Conf. Comput. Vis. Pattern Recog., pages 3684–3692, 2018. 410

[95] L. Yu, B. Twardowski, X. Liu, L. Herranz, K. Wang, Y. Cheng, S. Jui, and J. v. d. Weijer. Semantic drift compensation for class-incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 6982–6991, 2020. 409, 410

[96] M. Zand, S. Doraisamy, A. Abdul Halin, and M. R. Mustaffa. Ontology-based semantic image segmentation using mixture models and multiple crfs. IEEE Trans. Image Process., 25(7):3233–3248, 2016. 409

[97] Y. Zeng, J. Fu, and H. Chao. Learning joint spatial-temporal transformations for video inpainting. In Eur. Conf. Comput. Vis., pages 528–543, 2020. 410

[98] F. Zenke, B. Poole, and S. Ganguli. Continual learning through synaptic intelligence. In Int. Conf. Mach. Learn., 2017. 410

[99] C. Zhang, N. Song, G. Lin, Y. Zheng, P. Pan, and Y. Xu. Few-shot incremental learning with continually evolved classifiers. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[100] C.-B. Zhang, P.-T. Jiang, Q. Hou, Y. Wei, Q. Han, Z. Li, and M.-M. Cheng. Delving deep into label smoothing. IEEE Trans. Image Process., 2021. 410

[101] C.-B. Zhang, J. Xiao, X. Liu, Y. Chen, and M.-M. Cheng. Representation compensation networks for continual semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, 2022. 409

[102] D. Zhang, H. Zhang, J. Tang, M. Wang, X. Hua, and Q. Sun. Feature pyramid transformer. In Eur. Conf. Comput. Vis., pages 323–339, 2020. 410

[103] B. Zhao, X. Xiao, G. Gan, B. Zhang, and S.-T. Xia. Maintaining discrimination and fairness in class incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 13208–13217, 2020. 410

[104] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr. Conditional random fields as recurrent neural networks. In Int. Conf. Comput. Vis., 2015. 410

[105] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In IEEE Conf. Comput. Vis. Pattern Recog., 2021. 410

[106] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba. Scene parsing through ade20k dataset. In IEEE Conf. Comput. Vis. Pattern Recog., 2017. 413, 414

[107] F. Zhu, X.-Y. Zhang, C. Wang, F. Yin, and C.-L. Liu. Prototype augmentation and self-supervision for incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 5871–5880, 2021. 410

[108] K. Zhu, Y. Cao, W. Zhai, J. Cheng, and Z.-J. Zha. Self-promoted prototype refinement for few-shot class-incremental learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 6801–6810, 2021. 410

[109] L. Zhu, D. Ji, S. Zhu, W. Gan, W. Wu, and J. Yan. Learning statistical texture for semantic segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., pages 12537–12546, 2021. 409

[110] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai. Deformable detr: Deformable transformers for end-to-end object detection. In Int. Conf. Learn. Represent., 2021. 410