CorrMatch: 基于相关性匹配的半监督语义分割标签传播方法*

孙博远 [1] 杨雨奇 [1] 张乐 [2] 程明明 [1] 侯淇彬 [1†]

[1] 媒体计算实验室, 计算机学院, 南开大学
[2] 信息与通信工程学院, 电子科技大学

摘要

本文提出了一种简单但高性能的半监督语义分割方法，称为 CorrMatch。以往的方法大多采用复杂的训练策略来利用未标注数据，但忽略了相关图在建模位置对关系中的作用。我们观察到，相关图不仅可以容易地聚类属于同一类别的像素，还包含丰富的形状信息，而这些信息在以往的研究中被忽略了。受此启发，我们的目标是通过设计两种新颖的标签传播策略，提高未标注数据的利用效率。首先，我们提出基于像素间的成对相似性建模进行像素传播，以扩展高置信度像素并挖掘更多有效信息。然后，我们进行区域传播，通过从相关图中提取的准确类别无关掩码来增强伪标签。CorrMatch 在多个流行的分割基准测试上均表现出色。以 DeepLabV3+ 结合 ResNet-101 作为我们的分割模型，仅使用 92 张标注图像，我们在 Pascal VOC 2012 数据集上获得了超过 76% 的 mIoU 得分。代码可在以下地址获取：https://github.com/BBBBchan/CorrMatch。

1. 引言

随着深度学习技术的发展，尤其是卷积神经网络（CNNs） [20, 14, 59, 67, 12] 的进步，许多重要的语义分割方法 [38, 69, 5, 17, 42] 已经取得了显著的成果。然而，基于深度学习的方法通常需要大规模的像素级标注数据集，即大量的标注图像。相比于图像分类和目标检测任务 [8, 36]，语义分割数据集的精确标注成本更高，且耗时较长。

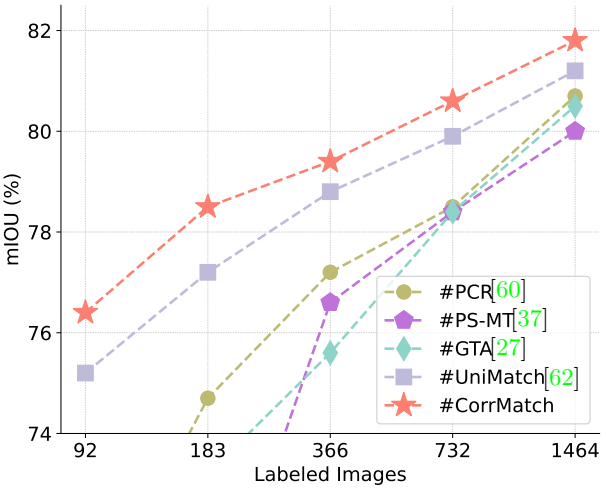


图 1: 在 Pascal VOC 数据集上与最新方法的比较。我们的 CorrMatch 在所有划分中均优于其他方法。

近年来，许多研究人员致力于通过弱监督 [56, 25, 54, 26]、半监督 [21, 22, 11, 40]，甚至无监督 [23, 13, 51, 18]的分割方法来降低对大规模精确标注数据的需求。在这些方案中，半监督语义分割仅需少量标注数据，并结合大量未标注数据进行训练，这种方式更接近真实世界的应用场景，因此吸引了越来越多学术界和工业界研究人员的关注。

在半监督语义分割的研究文献中，大多数工作采用 Mean Teacher 结构 [22, 37, 60, 27] 或自训练策略 [29, 63, 64] 来实现一致性正则化。如 表1 所示，这些方法通常需要额外的网络或训练阶段，使训练过程变得更加复杂。尽管最近的 UniMatch [62] 表

*本文为 CVPR 2024 论文 [48]的中文翻译版。
†通讯作者

表 1: 我们的 CorrMatch 与一些代表性方法的区别。SDA 代表强数据增强。

方法	多个网络	多阶段训练	多个 SDA 流	像素对相似性
PS-MT [37]	✓	✘	✘	✘
ST++ [63]	✘	✓	✘	✘
ELN [32]	✓	✓	✘	✘
UniMatch [62]	✘	✘	✓	✘
CorrMatch	✘	✘	✘	✓

明单阶段流程已经足够，但它仍然需要多个强数据增强流。不同于这些方法，我们的 CorrMatch 是一个更简洁的框架，不需要多个网络、额外的训练阶段或强数据增强流。

此外，在以往的研究中 [63, 37, 60]，利用未标注数据的最常见方式是设定固定阈值来筛选可靠的像素作为伪标签。然而，这些方法在利用未标注数据时往往面临困境，需要在伪标签的数量和准确性之间进行权衡。除此之外，受像素间相关性可以反映像素对相似性的启发，我们重新思考如何从标签传播的角度更准确地为未标注数据赋予伪标签。

首先，考虑到相关图蕴含全局像素对相似性，我们提出了像素传播策略。基于从提取特征构建的相关图，该策略将其传播到预测结果中，从而丰富预测信息，并提升语义一致性。同时，我们观察到，相关图的每一行都包含局部形状信息，因此可以获得一系列捕捉物体形状的二值图。因此，我们结合这些形状与高置信度区域的交集内最显著的预测类别，提出了区域传播策略，以准确地为这些形状分配类别标签，从而增强伪标签。通过将形状与高置信度区域的并集作为新的高置信度区域，我们可以扩大其范围，从而提高未标注数据的利用效率。如 图1 所示，我们的 CorrMatch 优于所有现有方法。

我们的主要贡献可总结如下：

- 我们展示了相关图在提高未标注数据利用效率方面的两大优势。
- 我们设计了一种简单但高性能的半监督语义分割框架，其中包含两种新颖的标签传播策略。
- 我们的 CorrMatch 在 Pascal VOC 2012 和 Cityscapes 数据集上达到了最新的最先进性能，同时在推理阶段无额外计算负担。

2. 相关工作

2.1. 半监督学习

半监督学习 [74, 44] 旨在探索如何利用标注数据和未标注数据构建模型，这一范式在深度学习时代到来之前就已经被广泛研究 [28, 3, 2]。随着深度学习和计算机视觉的进步，半监督学习受到了越来越多的关注 [35, 58, 15, 75, 4]。

自 Bachman et al. [1] 提出基于一致性正则化的方法以来，许多研究，如 Π-Model [34, 43]、Mean Teacher [49] 和 Dual Student [31]，已将其应用于半监督学习领域。最近，FixMatch [46] 提出了一种简单的弱到强一致性正则化框架，并成为许多相关方法的基准 [47, 16, 50, 62]。然而，后续的许多研究 [52, 66, 61] 指出，简单地设定手动固定阈值可能会导致性能下降和收敛速度变慢。其中，FreeMatch [52] 提出了一个与模型学习过程相关的动态阈值方案。然而，这些针对分类任务设计的策略并不适用于语义分割任务，因为每张图像通常包含多个类别。

2.2. 半监督语义分割

随着半监督学习在图像分类任务中取得显著成果 [35, 49, 46, 34]，许多研究也将类似的方法应用于语义分割任务 [21, 40, 57]。

其中一类方法 [11, 73, 22, 53, 37, 60, 68, 70] 采用了 Mean Teacher 结构。U^2PL [53] 试图通过对比学习更好地利用不可靠的预测结果。PS-MT [37] 通过 VAT [39] 技术构建了一个更严格的教师网络。ELN [32] 采用错误定位网络来缓解由于无效伪标签导致的确认偏差问题。所有这些方法都需要多个网络进行训练。与此同时，另一类方法——基于自训练的策略 [29, 63, 64, 9]，通常需要多个训练阶段。其中，ST++ [63] 提出了一个包含强数据增强的三阶段训练范式。SimpleBase [64] 采用了独立批归一化（Batch Normalization） [24] 来处理不同增强方式的图像。PC^2Seg [72] 在一致性训练的基础上引

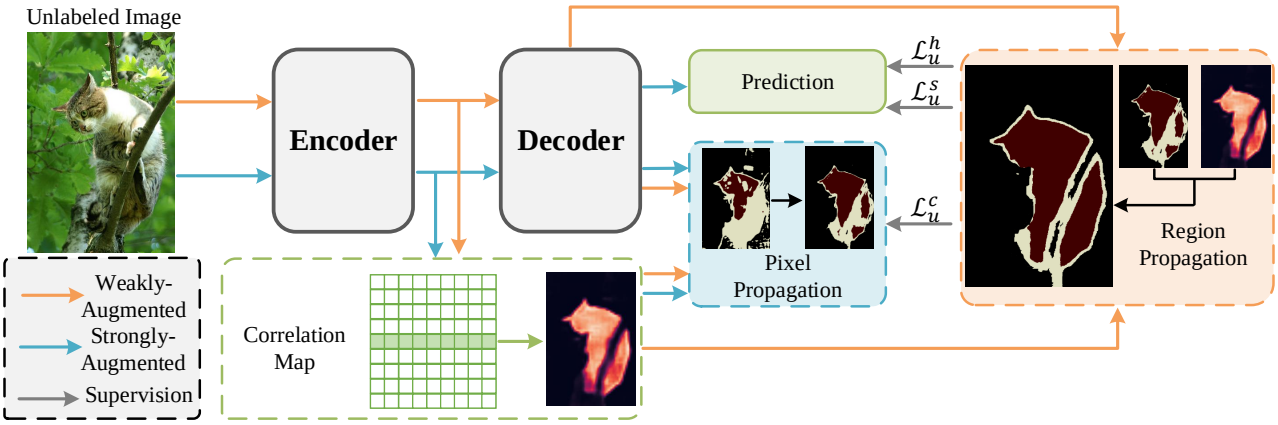


图 2: 我们 CorrMatch 处理未标注图像的流程示意图。我们基于 DeepLabv3+ 框架 [5] 进行构建。除了一致性正则化外，CorrMatch 还采用了基于相关匹配的两种标签传播策略。

入了特征空间对比学习。最近，UniMatch [62] 基于 FixMatch [46] 采用了一个单阶段框架，并结合多个强数据增强分支。不同于上述所有方法，CorrMatch 探索了如何更好地利用相关图来提高未标注数据的利用效率，并通过标签传播机制解决了这一问题，而以往的研究对此未予关注。

3. CorrMatch

半监督语义分割的目标是使用一个小规模的标注图像集和一个大规模的未标注图像集来训练语义分割网络 $\mathcal{F}$。我们提出了一种单阶段框架 CorrMatch，它利用像素对的相关性来实现两种标签传播策略。

3.1. 预备知识

CorrMatch 基于一个简单的框架 [62]，采用弱到强一致性正则化。对于标注图像 $\{x_i^l\}$ 及其对应的标签 $\{y_i^l\}$，我们使用标准的交叉熵损失进行训练。而未标注图像 $\{x_i^u\}$ 主要通过预测一致性进行利用。对于未标注图像，x_i^w 和 x_i^s 分别表示其经过弱增强和强增强的版本。一致性正则化将 x_i^w 的预测结果作为 x_i^s 的伪标签。我们在 图2 中展示了未标注图像的处理流程。

对于一个包含 N 张未标注图像的小批量样本，我们鼓励弱增强和强增强输入的预测结果保持一致，并通过强监督进行优化：

$$\mathcal{L}_u^h = \frac{1}{N}\sum_i^N \ell_c(\mathcal{F}(x_i^s), \mathcal{F}(x_i^w)) \odot \mathcal{M}_i, \tag{1}$$

其中 ℓ_c 是逐像素交叉熵损失函数，$\odot$ 代表逐元素乘法。$\mathcal{M}_i$ 是一个二值掩码，用于指示 $\mathcal{F}(x_i^w)$ 中高置信度的预测位置，其定义如下：

$$\mathcal{M}_i = \mathbb{1}(\max(\hat{\mathcal{F}}(x_i^w)) > \tau), \tag{2}$$

其中 $\hat{\mathcal{F}}(x_i^w) \in \mathbb{R}^{K\times HW}$ 为语义分割网络 $\mathcal{F}$ 生成的 logits 输出，K 为类别数。τ 为用于筛选高置信度预测像素的阈值。

然而，$\mathcal{L}_u^h$ 仅将 $\mathcal{F}(x_i^w)$ 视为硬伪标签，忽略了 logits $\hat{\mathcal{F}}(x_i^w)$ 中包含的额外信息。考虑到这一点，我们进一步在高置信度区域内约束弱增强和强增强图像的 logits 一致性。公式 式3 给出了软监督损失 $\mathcal{L}_u^s$ 的定义：

$$\mathcal{L}_u^s = \frac{1}{N}\sum_{i=1}^N \mathrm{KL}(\hat{\mathcal{F}}(x_i^s), \hat{\mathcal{F}}(x_i^w)) \odot \mathcal{M}_i, \tag{3}$$

其中 $\mathrm{KL}(\cdot)$ 为 Kullback-Leibler 散度损失函数。我们将上述框架视为我们的基线方法。

3.2. 像素传播

如 第1节所述，基于阈值选择的伪标签忽略了像素之间的语义相似性，限制了未标注数据的利用效

率。在本节中，我们提出像素传播策略，以增强模型对像素对相似性的整体感知，并提高未标注数据的利用率，该策略包括两个步骤：(1) 计算相关图；(2) 将相关图传播到预测结果中。

我们首先通过网络编码器后的线性层提取特征 w_1 和 $w_2 \in \mathbb{R}^{D\times HW}$，其中 D 为通道维度，HW 为特征向量的数量。这些提取的特征使相关匹配成为可能，从而量化像素对的相似程度。因此，我们通过特征向量的矩阵乘法计算相关图 $\mathcal{C}$：

$$\mathcal{C} = \text{Softmax}(w_1^\top \cdot w_2)/\sqrt{D}, \tag{4}$$

其中 $^\top$ 表示矩阵转置操作。相关图 $\mathcal{C} \in \mathbb{R}^{HW\times HW}$ 是一个二维矩阵，并通过 Softmax 函数进行激活，以获得像素对之间的相似性。$\mathcal{C}$ 能够准确地描绘属于同一目标的对应区域，如 图2 所示，这启发我们使用相关匹配将其传播到伪标签中。更多可视化结果见 图3。

为了增强模型对像素对相似性的感知，我们将相关图 $\mathcal{C}$ 传播到模型 logits 输出 $\hat{\mathcal{F}}(x_i^u)$，以通过标签传播获得另一种预测表示 $\mathrm{z}_i^u \in \mathbb{R}^{K\times HW}$：

$$\mathrm{z}_i^u = f_1(\hat{\mathcal{F}}(x_i^u)) \cdot \mathcal{C}, \tag{5}$$

其中 $f_1(\cdot)$ 是用于形状匹配的双线性插值函数。所得的 z_i^u 通过相关图强调同一目标的像素对相似性。

因此，我们计算 z_i^u 和高置信度伪标签之间的相关损失 $\mathcal{L}_u^c$ 作为监督：

$$\mathcal{L}_u^c = \frac{1}{|N|}\sum_{i=1}^{N}(\ell_c(\mathrm{z}_i^u, \mathcal{F}(x_i^w))) \odot \mathcal{M}_i. \tag{6}$$

对于标注图像 $\{x_i^l\}$，我们也计算 z_i^l 和 y_i^l 之间的交叉熵损失作为监督相关损失 $\mathcal{L}_s^c$，其中 z_i^l 可通过公式 式5 获得。至此，对于一个弱增强的未标注图像 x_i^w，其相关图 $\mathcal{C}_i^w$ 可有效建模像素对相似性。

3.3. 区域传播

在实验过程中，我们观察到相关图 $\mathcal{C}_i^w$ 中的每一行 c 表示单个特征向量与整个特征图中所有向量之间的相似性，这种信息隐含地包含了形状信息。基于这一观察，我们提出区域传播策略，以利用这些形状信息来增强伪标签。具体而言，我们首先对 c 进行归一化，并将其转换为二值图 $\hat{\mathrm{c}}$：

图 3: 我们提出的传播策略示意图。白色区域为低置信度忽略区域。结合形状信息和最显著类别，CorrMatch 可显著增强伪标签并扩展高置信度区域。

$$\hat{\mathrm{c}} = f_2(\mathbb{1}(\frac{\mathrm{c}-\min(\mathrm{c})}{\max(\mathrm{c})-\min(\mathrm{c})} > 0.5)), \tag{7}$$

其中 $f_2(\cdot)$ 是一个形状匹配函数，用于对齐 $\hat{\mathrm{c}}$ 和 $\mathcal{F}(x_i^w)$ 的形状。如 图3 所示，形状信息 $\hat{\mathrm{c}} \in \mathbb{R}^{H\times W}$ 显式地嵌入了类别无关的形状信息。对于每个 $\hat{\mathrm{c}}$，我们可以计算其与高置信度区域 $\mathcal{M}_i$ 之间的重叠比率 r_1。当 $\hat{\mathrm{c}}$ 与 $\mathcal{M}_i$ 具有较大的重叠度（即 $r_1 > \tau$），我们可以利用 $\hat{\mathrm{c}}$ 来调整伪标签 $\mathcal{F}(x_i^w)$。在当前伪标签 $\mathcal{F}(x_i^w)$ 下，我们可以计算在高置信度形状区域 $(\mathcal{F}(x_i^w) \odot \mathcal{M}_i \odot \hat{\mathrm{c}})$ 内每个类别 $l \in L$ 的数量，并使用以下公式定位最显著的类别 k^*：

$$k^* = \text{argmax}_{l\in L} G(l), \tag{8}$$

$$G(l) = \sum_{HW} \mathbb{1}[(\mathcal{F}(x_i^w) \odot \mathcal{M}_i \odot \hat{\mathrm{c}}) = l], \tag{9}$$

其中 L 是预测结果 $\mathcal{F}(x_i^w)$ 中出现的所有类别集合。基于最显著类别 k^*，我们可以计算其在高置信度形状区域内的占比 r_2。

当 k^* 高度符合高置信度形状区域（即 $r_2 > \tau$），我们可以将该类别 k^* 传播到增强的伪标签 $\mathcal{F}(x_i^w)$ 并扩展高置信度区域 $\mathcal{M}_i$：

$$\mathcal{F}(x_i^w) = \begin{cases} k^*, & \hat{\mathrm{c}} = 1 \\ \mathcal{F}(x_i^w), & \hat{\mathrm{c}} = 0 \end{cases}, \mathcal{M}_i = \mathcal{M}_i \cup \hat{\mathrm{c}} \tag{10}$$

表 2: 与最先进方法在 Pascal VOC 2012 验证集上的 mIoU (%) 结果对比。所有方法均在经典设置下训练，即标注图像来自原始 VOC 训练集，该训练集共包含 1,464 张图像。

方法	训练尺寸	1/16 (92)	1/8 (183)	1/4 (366)	1/2 (732)	完整 (1464)
Mean Teacher [49]	513 × 513	51.7	58.9	63.9	69.5	71.0
CutMix-Seg [11]	513 × 513	52.2	63.5	69.5	73.7	76.5
PseudoSeg [76]	513 × 513	57.6	65.5	69.1	72.4	73.2
CPS [6]	513 × 513	64.1	67.4	71.7	75.9	-
PC ^{2} Seg [72]	513 × 513	57.0	66.3	69.8	73.1	74.2
U^2 PL [53]	513 × 513	68.0	69.2	73.7	76.2	79.5
PS-MT [37]	513 × 513	65.8	69.6	76.6	78.4	80.0
GTA [27]	513 × 513	70.0	73.2	75.6	78.4	80.5
PCR [60]	513 × 513	70.1	74.7	77.2	78.5	80.7
RC^{2}L [68]	513 × 513	65.3	68.9	72.2	77.1	79.3
CCVC [55]	513 × 513	70.2	74.4	77.4	79.1	80.5
ST++ [63]	321 × 321	65.2	71.0	74.6	77.3	79.1
UniMatch [62]	321 × 321	75.2	77.2	78.8	79.9	81.2
CorrMatch	321 × 321	76.4	78.5	79.4	80.6	81.8

然而，考虑到相关图中每个形状的计算复杂度较高，且相邻区域往往具有相似的语义信息，导致相关图中存在相似的形状，从而对相关图中的每一行进行伪标签优化是冗余的。因此，我们采用随机采样方法加速标签传播。如 图3 所示，区域传播策略通过形状信息和最显著类别大幅扩展了高置信度区域。

值得一提的是，相关图的构建过程和标签传播仅在训练过程中参与计算，因此在推理过程中不会带来额外的计算开销。

3.4. 更多细节

动态阈值。如 FreeMatch [52] 所述，使用过于严格或过于宽松的固定阈值 τ 都会对模型收敛产生不利影响。同时，我们观察到，在不同实验设置下，最优的阈值是不同的（见 图5d）。因此，我们提供了一种与训练过程相关的动态阈值策略。

我们首先将阈值 τ 设为一个相对较小的初始值 (0.85)，然后基于 logits $\hat{\mathcal{F}}(x_i^w)$ 进行动态更新。我们使用指数移动平均（EMA） [41] 迭代更新 τ，每次更新的增量定义如下：

$$\Delta\tau = \frac{1}{|L|}\sum_{l\in L}\max[\mathbb{1}(\mathcal{F}(x_i^w)=l)\odot \overset{c}{\max}(\hat{\mathcal{F}}(x_i^w))], \tag{11}$$

其中 $\overset{c}{\max}(\cdot)$ 表示沿通道维度取最大值。该操作旨在提取 $\hat{\mathcal{F}}(x_i^w)$ 中所有预测类别的最大置信度，并使用其均值作为每次迭代的增量。我们发现，这种简单的阈值更新策略效果良好。我们将在 第4.3节进一步展示 τ 对初始化值的不敏感性。相关伪代码将在附录中提供。

损失函数。整体目标函数 $\mathcal{L}$ 由监督损失 $\mathcal{L}_s$ 和无监督损失 $\mathcal{L}_u$ 组成：$\mathcal{L}=\frac{1}{2}(\mathcal{L}_s+\mathcal{L}_u)$。与大多数方法相同，我们使用交叉熵损失 $\mathcal{L}_s^h$ 作为标注数据 $\mathcal{D}^l$ 的基本监督。因此，监督损失 $\mathcal{L}_s$ 定义为 $\mathcal{L}_s^h$ 和监督相关损失 $\mathcal{L}_s^c$ 的组合：$\mathcal{L}_s=\frac{1}{2}(\mathcal{L}_s^h+\mathcal{L}_s^c)$。对于未标注数据 $\mathcal{D}^u$，无监督损失 $\mathcal{L}_u$ 表达如下：

$$\mathcal{L}_u = \lambda_1\mathcal{L}_u^h + \lambda_2\mathcal{L}_u^s + \lambda_3\mathcal{L}_u^c, \tag{12}$$

其中 $\mathcal{L}_u^h, \mathcal{L}_u^s$ 和 $\mathcal{L}_u^c$ 分别表示无监督硬损失、软损失和相关损失。参数 $[\lambda_1,\lambda_2,\lambda_3]$ 默认为 $[0.5, 0.25, 0.25]$。

表 3: 与最先进方法在 Pascal VOC 2012 验证集上的 mIoU (%) 结果对比。所有方法均在增强设置(aug setting)下训练，即标注图像来自增强版 VOC 训练集，该训练集共包含 10,582 张图像。† 表示使用 U^{2}PL [53] 的数据划分。

方法	训练尺寸	1/16 (662)	1/8 (1323)	1/4 (2646)
Supervised	321 × 321	65.6	70.4	72.8
ST++ [63]	321 × 321	74.5	76.3	76.6
CAC [33]	321 × 321	72.4	74.6	76.3
UniMatch [62]	321 × 321	76.5	77.0	77.2
CorrMatch	321 × 321	77.6	77.8	78.3
U2PL† [53]	513 × 513	77.2	79.0	79.3
GTA† [27]	513 × 513	77.8	80.4	80.5
PCR† [60]	513 × 513	78.6	80.7	80.7
CCVC† [60]	513 × 513	76.8	79.4	79.6
AugSeg† [71]	513 × 513	79.3	81.5	80.5
CorrMatch†	513 × 513	81.3	81.9	80.9
CutMix-Seg [11]	513 × 513	71.7	75.5	77.3
CCT [40]	513 × 513	71.9	73.7	76.5
GCT [30]	513 × 513	70.9	73.3	76.7
CPS [6]	513 × 513	74.5	76.4	77.7
AEL [22]	513 × 513	77.2	77.6	78.1
FST [9]	513 × 513	73.9	76.1	78.1
ELN [32]	513 × 513	-	75.1	76.6
U^{2}PL [53]	513 × 513	74.4	77.6	78.7
PS-MT [37]	513 × 513	75.5	78.2	78.7
AugSeg [71]	513 × 513	77.0	77.3	78.8
CorrMatch	513 × 513	78.4	79.3	79.6

4. 实验

4.1. 实验配置

数据集。我们在 Pascal VOC 2012 和 Cityscapes 数据集上报告实验结果。Pascal VOC 2012 是一个包含 21 个类别的语义分割基准数据集，提供 1,464 张高质量标注图像用于训练，以及 1,449 张图像用于评估 [10]。我们还在扩展版 Pascal VOC 2012 数据集上进行实验，该数据集包含来自 Segmentation Boundary Dataset (SBD) [19] 的更多粗标注图像，总共包含 10,582 张训练图像。Cityscapes 是一个城市场景理解数据集，包括 2,975 张训练图像和 500 张验证图像，均带有精细标注 [7]。该数据集涵盖 19 个城市场景类别，所有图像的分辨率均为 1024×2048。

实现细节。遵循大多数先前的半监督语义分割方法，我们使用 DeepLabV3+ [5] 作为分割框架，并采用在 ImageNet [8] 预训练的 ResNet-101 [20] 作为主干网络。在 Pascal VOC 2012 训练过程中，我们使用 SGD 优化器，初始学习率设为 0.001，权重衰减设为 1e−4，裁剪尺寸设为 321×321 或 513×513，批量大小设为 16，训练 80 轮。在 Cityscapes 训练过程中，

表 4: 与最先进方法在 Cityscapes 验证集上的比较结果。所有实验均使用 ResNet-101 作为主干网络。

方法	1/16 (186)	1/8 (372)	1/4 (744)	1/2 (1488)
Supervised	65.7	72.5	74.4	77.8
CCT [40]	69.3	74.1	76.0	78.1
CPS [6]	69.8	74.3	74.6	76.8
AEL [22]	74.5	75.5	77.5	79.0
U^{2}PL [53]	70.3	74.4	76.5	79.1
PS-MT [37]	-	76.9	77.6	79.1
UniMatch[62]	76.6	77.9	79.2	79.5
PCR [60]	73.4	76.3	78.4	79.1
CorrMatch	77.3	78.5	79.4	80.4

按照 UniMatch [62] 的设置，我们使用 SGD 优化器，初始学习率设为 0.005，权重衰减设为 1e−4，裁剪尺寸设为 801×801，批量大小设为 16，训练 240 轮，并使用 4 × A40 GPUs 进行训练。

在评估指标方面，我们遵循以往研究 [6, 11, 37]，在 Pascal VOC 2012 数据集上报告基于原始图像的 mIoU（平均交并比）分数。对于 Cityscapes，我们遵循之前的方法 [6, 53, 62]，采用滑动窗口评估方式，

表 5: 不同组件的有效性消融实验，包括阈值 τ（Dyna. 表示动态策略）、硬损失 $\mathcal{L}_u^h$、软损失 $\mathcal{L}_u^s$、标签传播 $\mathcal{P}$。

τ	$\mathcal{L}_u^h$	$\mathcal{L}_u^s$	$\mathcal{P}$	92	1464
Dyna.	✓			73.6	80.0
Dyna.		✓		73.1	79.6
Dyna.	✓	✓		74.4	80.5
Dyna.	✓		✓	74.6	80.6
Dyna.	✓	✓	✓	76.4	81.8
固定	✓			73.1	79.9
固定	✓	✓		73.3	79.9
固定	✓		✓	74.3	80.1
固定	✓	✓	✓	75.5	80.8

表 6: 标签传播策略的消融实验。

方法	92	366	1464
无传播	74.4	78.5	80.5
仅像素传播	75.8	78.9	81.3
像素和区域传播	76.4	79.4	81.8

使用固定裁剪窗口计算 mIoU。所有结果均基于单尺度推理，在标准验证集上测量。

4.2. 与最先进方法的比较

Pascal VOC 2012 **经典设置的结果**。我们在 表2 中展示了我们的方法与最先进方法在 Pascal VOC 2012 经典数据集上的性能比较。我们的实验遵循 CPS [6] 设定的训练集划分方式。在完整训练集上，我们的方法取得了 81.8% 的 mIoU 分数。此外，CorrMatch 相较于现有最先进方法始终保持稳定的性能提升。特别地，CorrMatch 在所有数据划分中相较于 UniMatch 分别提升了 1.2%、1.3%、0.6%、0.7% 和 0.6%。

Pascal VOC 2012 **增强设置的结果**。在 表3 中，我们展示了我们的性能，并与现有方法在 Pascal VOC 2012 增强数据集上的表现进行了比较。可以明显看出，我们的结果始终显著优于现有最佳方法。我们的实验在 1/16、1/8 和 1/4 数据划分上进行。在 321×321 训练尺寸下，相较于监督基线，CorrMatch

表 7: 特征提取位置的消融实验。我们在 DeepLabV3+ 不同模块后提取特征来构建相关图，并采用标签传播策略。

位置	主干	ASPP	Fusion	Classifier
732	80.4	79.5	79.1	79.5
1464	81.8	80.6	80.1	80.8

表 8: 不同采样方法的消融实验。$\mathcal{R}$ 表示随机采样；$\mathcal{U}$ 表示均匀采样。

数量	16	32	64	128	256
$\mathcal{R}$	81.1	81.2	81.4	81.8	81.7
$\mathcal{U}$	81.0	81.1	81.2	81.4	81.0

分别提升了 12.0%、7.4% 和 5.5%。此外，我们的方法在各个数据划分上相较于 UniMatch 分别提升了 1.1%、0.8% 和 1.1%。对于 513×513 训练尺寸，我们的方法同样始终优于当前最先进方法，例如在 1/8 数据划分上，我们取得了 79.3% 的 mIoU，相较于 AugSeg [71] 提升了约 2%。我们还报告了在 U^2PL [53] 设定的数据划分下（513×513 训练尺寸）取得的结果，该划分方式包含更多精确标注的图像，因此期望的结果较高。相比于最佳方法 AugSeg [71]，我们的方法在 1/16 数据划分上提升了 2.0%。此外，与其他方法类似，我们观察到当数据划分从 1/8 增加到 1/4 时，性能在该设置下有所下降。这是因为在 1/8 数据划分中，几乎所有精确标注的图像都已包含，而大部分新增到更大划分的数据都是粗标注图像，导致性能未能进一步提升。

Cityscapes **结果**。在 表4 中，我们将 CorrMatch 的性能与 Cityscapes 数据集上的最先进方法进行了比较。我们遵循滑动窗口评估方法，并使用在线困难样本挖掘（OHEM）损失 [45]，这些技术已在之前的最新研究中广泛应用 [6, 53, 37, 62, 60, 22]。可以清楚地看到，我们的方法在所有数据划分上均优于其他方法。相比于 UniMatch [62]，我们的 CorrMatch 在 1/16、1/8、1/4 和 1/2 数据划分上分别提升了 0.7%、0.6%、0.2% 和 0.9%。

4.3. 消融实验

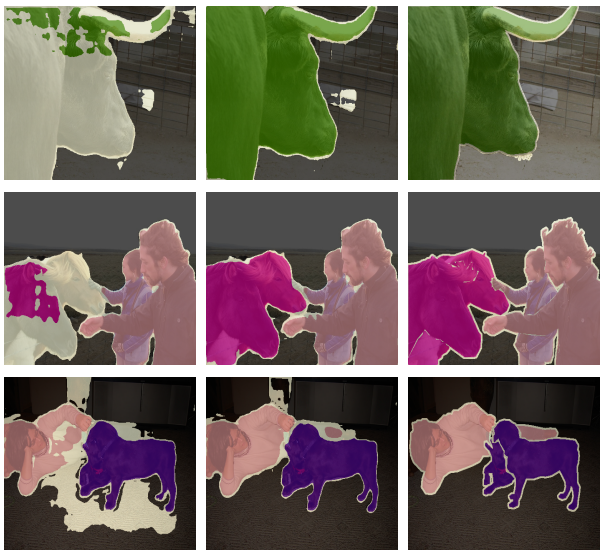
在本部分，我们对 CorrMatch 提出的策略进行一系列消融实验，以验证其设计的有效性。我们报告了在 Pascal VOC 2012 原始数据集上使用 ResNet-101 作为编码器的 DeepLabV3+ 网络的实验结果，训练尺寸设为 321 × 321。

各组件的有效性。我们首先对 CorrMatch 的不同组件进行消融实验，以证明其有效性，结果如 表5 所示。在使用无监督硬损失 $\mathcal{L}_u^h$ 和动态阈值策略的情况下，我们在 92 划分上取得 73.6%，在 1464 划分上取得 80.0% 的 mIoU 结果。在此基础上，加入软损失 $\mathcal{L}_u^s$ 作为基本框架，可分别带来 0.8% 和 0.5% 的提升。结合标签传播策略后，我们分别进一步提升了 2.0% 和 1.3%。这些结果表明，每个组件单独使用时均能有效提升性能。此外，将 $\mathcal{L}_u^h$ 替换为 $\mathcal{L}_u^s$ 会导致性能下降，说明 $\mathcal{L}_u^h$ 是必要的。最终，完整的 CorrMatch 在 92 和 1464 分割上分别达到了 76.4% 和 81.8% mIoU，相较于基线方法分别提升了 2.8% 和 1.8%。

我们还进行了固定阈值（0.95）的实验。可以观察到，相较于固定阈值的基线（73.1% 和 79.9%），改为动态阈值仅带来 +0.5% 和 +0.1% 的提升。然而，在加入所有组件后，相应的提升可达到 +0.9% 和 +1.0%。这证明了我们的阈值策略与标签传播策略能够很好地协同工作。

标签传播策略的影响。在 表6 中，我们对标签传播策略进行了消融实验。像素传播策略通过构建相关图并将其传播到预测中，结合相关损失 $\mathcal{L}^c$ 进行监督，带来了 1.4%、0.4% 和 0.8% 的提升。此外，结合区域传播策略后，可以挖掘更多局部形状信息，从而获得更准确的伪标签。该策略在 92、366 和 1464 划分上分别进一步提升了 0.6%、0.5% 和 0.5%。

特征提取位置的影响。在默认设置下，我们选择从主干网络提取特征，这使得所提出的策略更易于移植到其他分割网络。事实上，在特定网络结构下，特征提取的位置是灵活的。在 表7 中，我们展示了在 Deeplabv3+ 解码器的不同位置提取特征时的性能表现。结果表明，从主干网络提取特征始终优于其他选择。

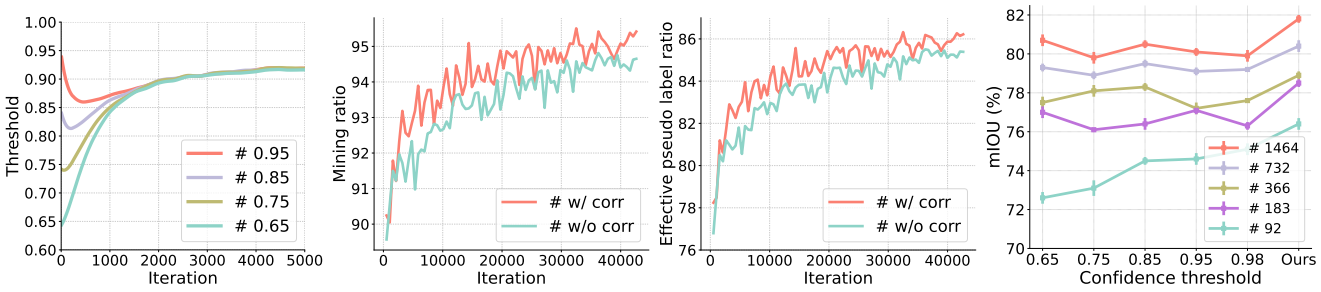


(a) 无传播　(b) 有传播　(c) 真实标签

图 4: Pascal VOC 2012 数据集上的定性实验结果。(a) 不使用标签传播的伪标签；(b) 采用 CorrMatch 进行标签传播后的伪标签；(c) 真实标签。(a) 和 (b) 中的白色区域表示由于置信度较低而被忽略的区域。

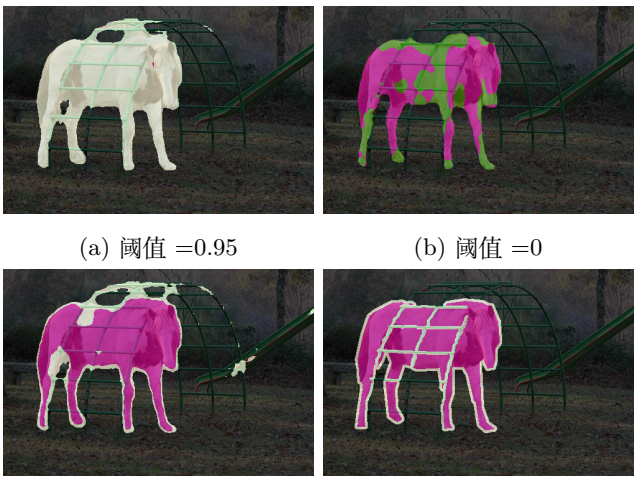
不同的采样策略。由于使用相关图中的所有形状来增强伪标签会带来巨大的计算负担，因此必须从中采样一部分形状。在 表8 中，我们对采样方法和采样数量进行了实验。我们在 1464 划分上对随机采样 $\mathcal{R}$ 和均匀采样 $\mathcal{U}$ 方法进行了实验，分别设置 16、32、64、128 和 256 的采样数量。实验结果表明，随机采样始终优于均匀采样。其中，使用 128 个样本的随机采样方法获得了最佳性能，与 256 样本策略相比，性能差异极小。因此，我们选择从相关图中随机采样 128 个形状，以在计算效率和性能之间取得平衡。

CorrMatch **的不同初始值影响**。由于我们基于 EMA 的阈值更新策略需要一个初始值 τ，因此我们在 图5a 中讨论了不同初始化值对 τ 的影响。实验结果表明，我们的阈值策略对不同的初始化值不敏感。即使使用不同的阈值初始化，在所有实验设置下，所有阈值都将在训练的早期阶段（约 40000 轮训练中的前 1500 轮）迅速趋向于相似的值。



(a) 不同的阈值初始化　(b) 挖掘比例　(c) 有效伪标签比例　(d) 不同的固定阈值

图 5: 关于标签传播和阈值策略的一些统计数据。(a)、(b) 和 (c) 的实验在 1464 划分上进行。



(a) 阈值 =0.95　(b) 阈值 =0

(c) 标签传播　(d) 真实标签

图 6: 不同策略下伪标签的比较。

5. 标签传播策略与阈值调整的对比讨论

传统的半监督语义分割方法主要依赖调整阈值来扩展高置信度区域 [62, 53]。然而，选择最合适的阈值可能是一项具有挑战性的任务。例如，我们在图5d 中的观察表明，最佳阈值可能会有显著变化。图6a 和 图6b 进一步说明，过于严格的阈值会限制未标注数据的利用，而过于宽松的阈值则会导致零散的错误像素预测。

不同于直接调整阈值的方法，标签传播不仅仅是扩展高置信度区域，而是利用相关图中的精确形状信息来为伪标签分配准确的预测。这种方式有助于在高置信度区域内保持更一致的语义结构，从而缓解预测的不连续性问题。在 图6c 和 图5d 的最后一列中，我们展示了 CorrMatch 生成的伪标签及其性能表现。这表明，CorrMatch 始终能够获得更准确、更完整的伪标签，并在所有数据划分上都实现了最佳结果，充分证明了所提出的标签传播策略的有效性。

6. 总结

我们提出了 CorrMatch，该方法利用相关匹配进行标签传播，以挖掘更准确的高置信度区域，从而提升半监督语义分割的效果。CorrMatch 的核心贡献在于重新审视相关图的作用，并设计了两种标签传播策略以丰富伪标签。在这些策略的加持下，CorrMatch 显著扩展了高置信度区域，从而更高效地利用未标注数据。实验结果表明，CorrMatch 相较于其他方法具有明显的优势。

A. 所提策略的伪代码

A.1. 区域传播策略的伪代码

在正文的 第3.3节部分，我们提出了区域传播策略。该策略结合从相关图中采样的形状信息和最显著类别，以增强伪标签并扩展高置信度区域。这里我们以 PyTorch 风格给出区域传播策略的伪代码。

Algorithm 1 PyTorch 风格的区域传播策略伪代码。

```
# shapes: Binary shape information sampled from correlation
    maps
# t: Confidence threshold
# hc_regions: Current high-confidence regions
# pseudo_label: Current pseudo label
def Region(shapes, t, hc_regions, pseudo_label):
    # Find the high-confidence shapes
    hc_shapes = shapes * hc_regions
    b, c, h, w = shapes.shape

    for i in range(b):
        for j in range(c):
            hc_shape = hc_shapes[i, j]
            shape = shapes[i, j]

            # Calculate the overlap between the high-
                confidence shape and original shape
            r1 = sum(hc_shape) / sum(shape)
            if r1 < t:
                continue

            # Find all unique classes and their counts in the
                pseudo label within the high-confidence shape
            unique_cls, cnt = unique(pseudo_label[i][hc_shape
                == 1])

            # Calculate the ratio of the most salient class
                within the high-confidence shape
            r2 = max(cnt) / sum(cnt)
            if r2 < t:
                continue

            # Assign the most salient class to the pseudo label
                with shape information
            top_cls = unique_cls[argmax(cnt)]
            pseudo_label[i][shape == 1] = top_cls

            # Update the new high-confidence regions with the
                current shape
            hc_regions[i] = hc_regions[i] | shape
```

A.2. 阈值更新策略的伪代码

在正文的 第3.4节部分，我们提出了阈值更新策略。其核心思想是维护一个与模型学习过程相关的动态全局阈值。具体而言，在优化过程中，我们通过对弱增强预测中所有类别的最大置信度的平均值进行逐步更新阈值。根据本文公式 (11) 中提出的增量 $\Delta\tau$，EMA 过程定义如下：

$$\tau = \lambda\tau + (1-\lambda)\Delta\tau, \tag{13}$$

其中，λ 为 EMA 的动量衰减因子。为了更清晰地展示该过程，我们在此提供 PyTorch 风格的阈值更新策略伪代码。

Algorithm 2 PyTorch 风格的阈值更新策略伪代码。

```
# pred: Logits of weak augmented images
# thresh_global: Current global threshold
# momentum: Coefficient of EMA
def update(pred, thresh_global, momentum):
    # initialize update value
    update_value = 0.0

    # get predicted mask and confidence from pred
    mask_pred = argmax(pred, dim=1)
    pred_conf = pred.softmax(dim=1).max(dim=1)

    # find all classes in the predicted mask
    unique_cls = unique(mask_pred)
    cls_num = len(unique_cls)

    for cls in unique_cls:
        # find the highest confidence score for each predicted
            class
        cls_map = (mask_pred == cls)
        pred_conf_cls_all = pred_conf[cls_map]
        cls_max_conf = pred_conf_cls_all.max()
        update_value += cls_max_conf

    # get the mean of all confidence scores
    update_value = update_value / cls_num

    # update thresh_global in EMA style
    thresh_global = momentum * thresh_global + (1 -
        momentum) * update_value
```

B. 更多实现细节

数据增强。我们遵循了先前研究的通用设置 [63, 62, 76]。对于弱数据增强，我们使用随机缩放（范围 [0.5, 2.0]）、随机水平翻转（概率为 0.5）以及随机裁剪（尺寸为 321、513 或 801）。对于强数据增强，我们使用 colorjitter 技术，调整图像的亮度、对比度、饱和度和色调，参数设置与先前研究一致 [63, 62, 76]。

表 9: CorrMatch 在 PASCAL VOC 2012 验证集上不同 EMA 动量衰减的比较，评测指标为 mIoU (%) ↑。

动量衰减	1 / 16(92)	完整 (1464)
0.99	75.6	79.8
0.999	76.4	81.8
0.999	75.7	80.3

此外，还应用了随机灰度转换和高斯模糊作为强数据增强方法。我们同样使用 CutMix [65] 技术，该方法已在许多先前研究中广泛采用 [63, 62, 76]。此外，为了学习更具鲁棒性的特征表示，我们采用了 UniMatch [62] 中的特征扰动策略，即在编码器特征中随机丢弃 50% 的通道。

特征提取器。如本文 第3.2节所述，我们从网络的编码器部分提取特征。具体来说，特征提取器由一个 3×3 卷积层、批归一化层 [24] 和一个激活层组成。然后，我们在提取的特征上应用两个独立的线性变换，以获得 w_1 和 w_2。

其他细节。我们使用带动量 $= 0.9$ 的 SGD 优化器，并采用多项式衰减策略 $(1 - \frac{\text{iter}}{\text{total iter}})^{0.9}$ 来在训练过程中逐步降低学习率。此外，在所提出的动态阈值更新策略中，我们将 EMA 的动量参数设为 0.999。此外，与 UniMatch [62] 一致，我们在 Cityscapes 数据集上将置信度阈值 τ 设为 0。

C. 更多消融实验

C.1. 动量衰减的影响

考虑到 CorrMatch 采用 EMA（指数移动平均）来迭代更新动态阈值，我们在 表9 中对 EMA 的动量衰减进行了消融实验。

C.2. 不同的软监督方法

如 第3.1节所述，我们在半监督语义分割任务中引入了软监督。在 表10 中，我们针对不同的软监督技术进行了实验。结果表明，不同的软监督方法

表 10: CorrMatch 在 PASCAL VOC 2012 验证集上不同软监督方法的比较，评测指标为 mIoU (%) ↑。

方法	1 / 16(92)	完整 (1464)
Kullback-Leibler divergence	76.4	81.8
Soft cross-entropy	76.2	81.6
Cosine similarity	76.1	81.5

具有相似的表现，表明 KL 散度（Kullback-Leibler Divergence）只是一个软度量，其他替代的软监督方法也可以实现可比的性能。

C.3. 不同损失权重

在 表11 中，我们进一步研究了不同损失权重的影响。当赋予未标注数据的权重过大时，模型性能受到显著影响，而更平衡的权重分配对模型性能的影响较小。实验结果表明，当 $[\lambda_1, \lambda_2, \lambda_3]$ 设置为 [0.5, 0.25, 0.25] 时，可获得最佳性能。

D. 关于标签传播的更多分析

D.1. 相关性模块不是注意力模块

相关图的构建与注意力机制存在本质区别：

- 在形式上，注意力机制中的键（Key, K）和值（Value, V）均来源于相同的特征表示，通常位于同一输入序列中。而我们的相关性机制首先计算提取特征之间的相关图，随后采用像素传播策略将其传播至模型输出，显然其来源不同于提取的特征。
- 在内容编码方面，相关图编码了来自不同区域的特征之间的两两相似性，而注意力图则是一组权重，决定了输入序列中不同位置的重要性。

综上所述，我们的相关性模块在形式和编码内容上均不同于注意力机制。此外，所提出的两种标签传播策略涉及将相关图传播至输出并利用形状信息增强伪标签，使得我们的相关性模块区别于注意力模块。

表 11: CorrMatch 在 PASCAL VOC 2012 验证集上不同损失权重的比较，评测指标为 mIoU (%) ↑。

$[\lambda_1, \lambda_2, \lambda_3]$	1 / 16(92)	完整 (1464)
$[0.5, 0.25, 0.25]$	76.4	81.8
$[0.25, 0.5, 0.25]$	75.6	81.2
$[0.25, 0.25, 0.5]$	75.9	81.1
$[0.3, 0.3, 0.3]$	75.4	80.2
$[0.5, 0.5, 0.5]$	73.4	79.5
$[1, 1, 1]$	70.6	78.0

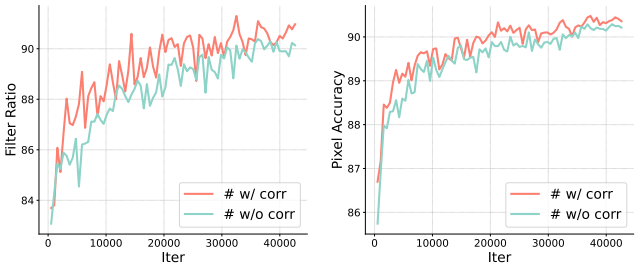
D.2. 更多统计分析

在 图7 中，我们展示了 PASCAL VOC 2012 验证集上的更多统计信息，以进一步证明基于相关性匹配的标签传播的有效性。我们统计了采用和不采用相关性匹配时的筛选率和像素准确率，分别见图7a 和 图7b。筛选率指被视为伪标签的高置信度像素占整幅图像的比例，反映了模型整体置信度水平。像素准确率则表示所有被正确预测的像素占整个图像的比例。所有实验均基于 1464 划分集，训练尺寸为 321×321。

可以清楚地看到，这三幅图的曲线趋势保持一致。即采用相关性匹配可以获得更优的结果。这表明，模型不仅倾向于做出更高置信度的预测，而且正确预测的高置信度像素数量也增加了。同时，采用相关性匹配后的更高像素准确率表明模型本身的性能得到了提升。这些统计结果进一步证明，所提出的 CorrMatch 结合标签传播策略，可以挖掘更准确的高置信度区域，从而提升模型对未标注数据的学习能力。

D.3. 掩码比例

在 图8 中，我们展示了训练过程中掩码比例（即被阈值筛选出的高置信度像素所占比例）的变化情况。我们分别对使用固定阈值与使用 CorrMatch 进行统计分析。显然，固定阈值越低，掩码比例越高。此外，训练早期的掩码比例过低会导致较少的伪标签，从而影响收敛速度。相反，训练后期掩码比例过高会包含更多错误预测，影响伪标签的准确性。这



(a) 筛选率　　(b) 像素准确率

图 7: 关于标签传播策略的更多统计分析。

两种情况均对模型收敛不利。然而，CorrMatch 通过在训练早期保持较高掩码比例，并在训练后期保持较低掩码比例，有效解决了这一问题。这一现象在 图8a、图8c、图8b 和 图8d 中均保持一致，进一步验证了我们方法的稳定性。

D.4. 更多可视化结果

在正文中，我们表明所提出的标签传播策略可以帮助挖掘可靠区域，并通过大量定量和定性实验验证了这一点。在 图9 中，我们进一步提供了更多的可视化结果，以进一步支持我们的结论。

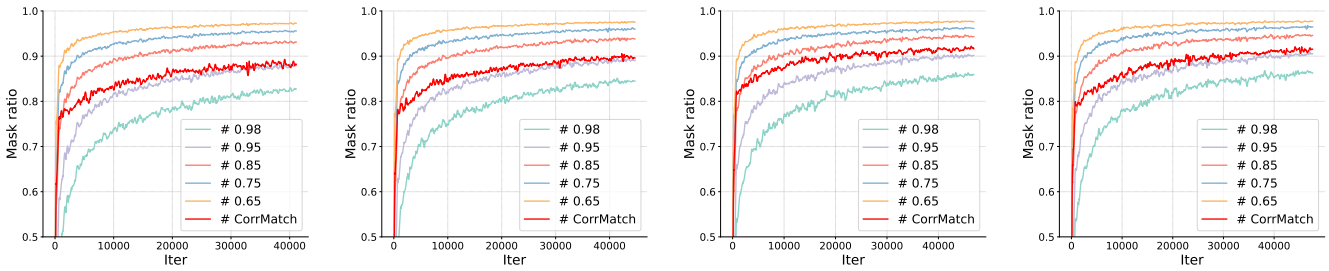
E. 动态阈值的更多分析

E.1. 为什么半监督语义分割需要特殊的动态阈值设计

在本文中，除了两种标签传播策略外，我们还提出了一种用于半监督语义分割的动态全局阈值。这里我们想要探讨一个问题：**既然动态阈值策略已在许多半监督学习工作中被广泛研究，为什么半监督语义分割仍然需要特殊的动态阈值设计？**

半监督学习与半监督语义分割任务存在本质上的不同。我们首先列出半监督学习与半监督语义分割之间的一些潜在区别。

1. **任务目标：** 在半监督学习中，目标是在图像级别进行预测，而半监督语义分割是一个密集预测任务，关注逐像素分类。其目标是对每个像素进行分类，并且一张图像可能包含多个类别。
2. **阈值的使用方式：** 在半监督学习中，阈值通常用于决定一张图像的预测是否可作为伪标签。而



(a) 1464 划分 (b) 732 划分 (c) 366 划分 (d) 92 划分

图 8: 不同划分比例下，使用不同固定阈值训练过程中掩码比例的变化情况。

表 12: 在 PASCAL VOC 2012 验证集上，不同阈值策略的对比，使用 mIoU (%) ↑ 作为评价指标。

方法	1 / 16(92)	完整 (1464)
CorrMatch	76.4	81.8
逐像素阈值更新	64.1	77.2
基于最大置信度更新	63.4	74.4
基于平均置信度更新	75.4	80.2

在半监督语义分割中，阈值则作用于单个像素，以筛选出高置信度区域，并将其作为伪标签。

3. **目标尺寸：**在半监督学习中，模型的任务是对整张图像进行分类。而在半监督语义分割中，模型的任务是将图像分割成不同语义对象的区域。由于图像中的目标往往大小不一，其特征分布可能存在显著差异，因此学习难度各不相同。

考虑到上述潜在差异，我们在 表12 中进行了一系列实验，证明简单地将半监督学习的策略扩展到逐像素的语义分割任务是不足够的，我们的方法在半监督语义分割任务中的设计是非平凡的。

1. **逐像素阈值更新：**首先，我们为每个像素设置一个独立的阈值，并根据其置信度单独更新。然而，由于不同语义的目标位置并不固定，其置信度分布也不会由像素位置决定，这种方法明显导致性能下降。
2. **基于最大置信度更新：**然后，我们尝试为每个类别使用全局阈值，并用全局最大置信度更新阈值。然而，由于某些像素更容易学习，并且它们的置信度值接近 1，这导致阈值迅速趋近于 1，使得大部分区域被视为低置信度区域，从而造成性能下降。
3. **基于平均置信度更新：**最后，我们尝试使用全局平均置信度进行阈值更新，使所有像素均等参与阈值更新。然而，不同类别占据的像素数量不同，且学习难度各异。例如，在 Pascal VOC 2012 数据集中，背景类通常占据图像的大部分，并且往往具有较高的置信度。因此，这种方法仍然会引入相对较高的阈值，导致性能下降。

总的来说，与半监督学习不同，设计半监督语义分割的策略需要考虑空间依赖性和逐像素预测，使得任务更为复杂和具有挑战性。我们的策略充分考虑了上述差异，通过计算预测中每个类别的最高置信度，并利用其平均值来维护动态阈值。实验结果表明，我们的阈值更新策略是非平凡的。此外，据我们所知，我们是第一个在半监督语义分割任务中引入动态阈值与标签传播的工作。

E.2. 为什么不采用按类别的阈值更新

鉴于我们提出的阈值策略最终仍然是更新一个全局阈值，可能会有人提出疑问：既然在半监督分类任务中，按类别的动态阈值更新策略已经取得了成功 [52, 66]，那么在语义分割任务中使用按类别的动态阈值更新策略是否会带来性能提升？然而，正如在 第E.1节中所讨论的，分类任务和语义分割任务具有不同的特性。因此，类似的策略可能并不适用于半监督语义分割任务。为了进一步说明这一点，我们进行了以下按类别阈值更新的实验。

我们首先初始化一个与类别数量相同大小的张

表 13: 在 PASCAL VOC 2012 验证集上，对比 CorrMatch 在使用和不使用按类别阈值更新策略下的 mIoU (%) ↑ 结果。* 表示采用了按类别的阈值更新策略。

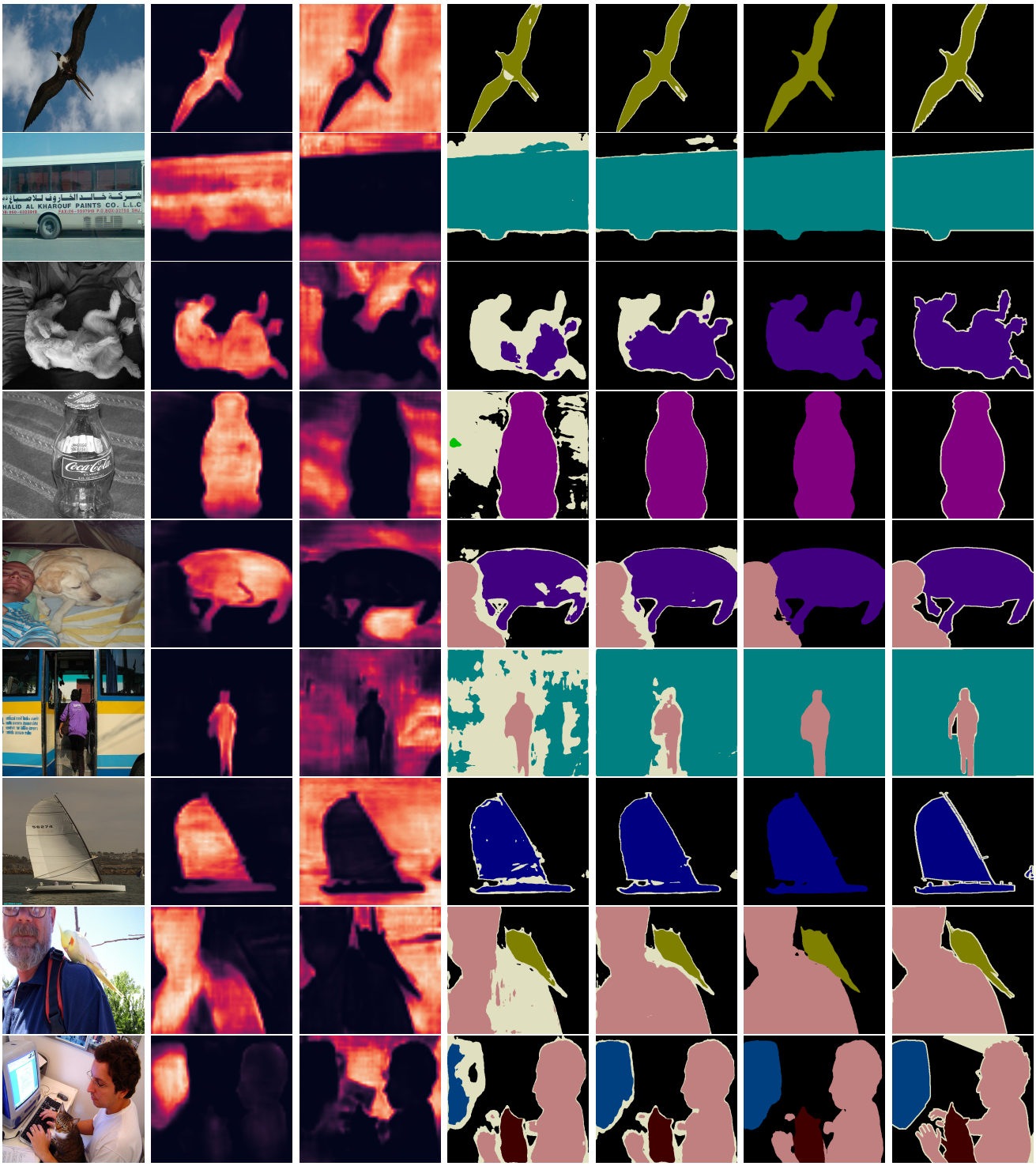
方法	1 / 16 (92)	1 / 8 (183)	1 / 4 (366)	1 / 2 (732)	完整 (1464)
CorrMatch	76.4	78.5	79.4	80.6	81.8
CorrMatch*	75.1	76.7	78.3	79.3	80.3

量，并将其初始值设为与全局初始化值相同。然后，我们使用与全局阈值更新相似的 EMA（指数移动平均）策略进行迭代更新。对于模型预测 $\mathcal{F}(x_i^w)$ 中的每个类别 l，每次迭代的更新过程如下：

$$\tau_l' = \max[\mathbb{1}(\mathcal{F}(x_i^w) = l) \circ \overset{c}{\max}(\hat{\mathcal{F}}(x_i^w))], \qquad (14)$$

其中，$\hat{\mathcal{F}}(x_i^w)$ 是对弱增强未标注图像的 logits 预测。该操作表示，我们取弱增强未标注图像中每个预测类别的最高置信度，并将其作为该类别阈值的更新增量。然后，类似于 FreeMatch [52]，我们使用最大值归一化操作来整合全局和局部阈值。

我们在原始 Pascal VOC 2012 数据集上进行了实验，采用 321×321 的训练尺寸，实验结果如 表13 所示。可以明显看出，相比于全局阈值更新策略，改为按类别的阈值更新方案会导致大约 1% 的性能下降。



(a) 图像 (b) 物体 (c) 背景 (d) 无相关匹配 (e) 有相关匹配 (f) 预测结果 (g) 真实标签

图 9: 来自 Pascal VOC 2012 数据集验证集的更多定性结果。(a) 输入图像；(b) 物体上的相关性图；(c) 背景上的相关性图；(d) 无相关匹配的伪标签；(e) 采用 CorrMatch 的伪标签；(f) CorrMatch 的预测结果；(g) 真实标签。(d) 和 (e) 中的白色区域表示由于置信度低而被忽略的区域。

参考文献

[1] P. Bachman, O. Alsharif, and D. Precup. Learning with pseudo-ensembles. NeurIPS, 27, 2014. 2

[2] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. Journal of machine learning research, 7(11), 2006. 2

[3] K. Bennett and A. Demiriz. Semi-supervised support vector machines. NeurIPS, 11, 1998. 2

[4] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel. Mixmatch: A holistic approach to semi-supervised learning. NeurIPS, 32, 2019. 2

[5] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In ECCV, pages 801–818, 2018. 1, 3, 6

[6] X. Chen, Y. Yuan, G. Zeng, and J. Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In CVPR, pages 2613–2622, 2021. 5, 6, 7

[7] 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 CVPR, pages 3213–3223, 2016. 6

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, pages 248–255. Ieee, 2009. 1, 6

[9] Y. Du, Y. Shen, H. Wang, J. Fei, W. Li, L. Wu, R. Zhao, Z. Fu, and Q. Liu. Learning from future: A novel self-training framework for semantic segmentation. arXiv preprint arXiv:2209.06993, 2022. 2, 6

[10] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, 88:303–308, 2009. 6

[11] G. French, S. Laine, T. Aila, M. Mackiewicz, and G. Finlayson. Semi-supervised semantic segmentation needs strong, varied perturbations. In Brit. Mach. Vis. Conf., 2020. 1, 2, 5, 6

[12] S. Gao, Z.-Y. Li, Q. Han, M.-M. Cheng, and L. Wang. Rf-next: Efficient receptive field search for convolutional neural networks. IEEE TPAMI, pages 1–19, 2022. 1

[13] S. Gao, Z.-Y. Li, M.-H. Yang, M.-M. Cheng, J. Han, and P. Torr. Large-scale unsupervised semantic segmentation. IEEE TPAMI, pages 1–20, 2022. 1

[14] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr. Res2net: A new multi-scale backbone architecture. IEEE TPAMI, 43(2):652–662, 2021. 1

[15] Y. Grandvalet and Y. Bengio. Semi-supervised learning by entropy minimization. In L. Saul, Y. Weiss, and L. Bottou, editors, Advances in Neural Information Processing Systems, volume 17. MIT Press, 2004. 2

[16] S. Grollmisch and E. Cano. Improving semi-supervised learning for audio classification with fixmatch. Electronics, 10(15):1807, 2021. 2

[17] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu. Segnext: Rethinking convolutional attention design for semantic segmentation. arXiv preprint arXiv:2209.08575, 2022. 1

[18] R. Harb and P. Knöbelreiter. Infoseg: Unsupervised semantic image segmentation with mutual information maximization. In Pattern Recognition: 43rd DAGM German Conference, DAGM GCPR 2021, Bonn, Germany, September 28–October 1, 2021, Proceedings, pages 18–32. Springer, 2022. 1

[19] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik. Semantic contours from inverse detectors. In ICCV, pages 991–998. IEEE, 2011. 6

[20] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016. 1, 6

[21] S. Hong, H. Noh, and B. Han. Decoupled deep neural network for semi-supervised semantic segmentation. NeurIPS, 28, 2015. 1, 2

[22] H. Hu, F. Wei, H. Hu, Q. Ye, J. Cui, and L. Wang. Semi-supervised semantic segmentation via adaptive equalization learning. NeurIPS, 34:22106–22118, 2021. 1, 2, 6, 7

[23] J.-J. Hwang, S. X. Yu, J. Shi, M. D. Collins, T.-J. Yang, X. Zhang, and L.-C. Chen. Segsort: Segmentation by discriminative sorting of segments. In ICCV, pages 7334–7344, 2019. 1

[24] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal

covariate shift. In ICML, pages 448–456. pmlr, 2015. 2, 11

[25] P.-T. Jiang, L.-H. Han, Q. Hou, M.-M. Cheng, and Y. Wei. Online attention accumulation for weakly supervised semantic segmentation. IEEE TPAMI, 44(10):7062–7077, 2022. 1

[26] P.-T. Jiang, Y. Yang, Q. Hou, and Y. Wei. L2g: A simple local-to-global knowledge transfer framework for weakly supervised semantic segmentation. In CVPR, 2022. 1

[27] Y. Jin, J. Wang, and D. Lin. Semi-supervised semantic segmentation via gentle teaching assistant. In NeurIPS, 2022. 1, 5, 6

[28] T. Joachims et al. Transductive inference for text classification using support vector machines. In ICML, volume 99, pages 200–209, 1999. 2

[29] R. Ke, A. I. Aviles-Rivero, S. Pandey, S. Reddy, and C.-B. Schönlieb. A three-stage self-training framework for semi-supervised semantic segmentation. IEEE Trans. Image Process., 31:1805–1815, 2022. 1, 2

[30] Z. Ke, D. Qiu, K. Li, Q. Yan, and R. W. Lau. Guided collaborative training for pixel-wise semi-supervised learning. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16, pages 429–445. Springer, 2020. 6

[31] Z. Ke, D. Wang, Q. Yan, J. Ren, and R. W. Lau. Dual student: Breaking the limits of the teacher in semi-supervised learning. In ICCV, pages 6728–6736, 2019. 2

[32] D. Kwon and S. Kwak. Semi-supervised semantic segmentation with error localization network. In CVPR, pages 9957–9967, 2022. 2, 6

[33] X. Lai, Z. Tian, L. Jiang, S. Liu, H. Zhao, L. Wang, and J. Jia. Semi-supervised semantic segmentation with directional context-aware consistency. In CVPR, pages 1205–1214, 2021. 6

[34] S. Laine and T. Aila. Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242, 2016. 2

[35] D.-H. Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML, volume 3, page 896, 2013. 2

[36] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In ECCV, pages 740–755. Springer, 2014. 1

[37] Y. Liu, Y. Tian, Y. Chen, F. Liu, V. Belagiannis, and G. Carneiro. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In CVPR, pages 4258–4267, 2022. 1, 2, 5, 6, 7

[38] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, pages 3431–3440, 2015. 1

[39] T. Miyato, S.-i. Maeda, M. Koyama, and S. Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE TPAMI, 41(8):1979–1993, 2018. 2

[40] Y. Ouali, C. Hudelot, and M. Tami. Semi-supervised semantic segmentation with cross-consistency training. In CVPR, pages 12674–12684, 2020. 1, 2, 6

[41] B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. SIAM journal on control and optimization, 30(4):838–855, 1992. 5

[42] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pages 234–241. Springer, 2015. 1

[43] M. Sajjadi, M. Javanmardi, and T. Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. NeurIPS, 29, 2016. 2

[44] M. Seeger. Learning with labeled and unlabeled data, 2000. 2

[45] A. Shrivastava, A. Gupta, and R. Girshick. Training region-based object detectors with online hard example mining. In CVPR, pages 761–769, 2016. 7

[46] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. NeurIPS, 33:596–608, 2020. 2, 3

[47] K. Sohn, Z. Zhang, C.-L. Li, H. Zhang, C.-Y. Lee, and T. Pfister. A simple semi-supervised learn-

ing framework for object detection. arXiv preprint arXiv:2005.04757, 2020. 2

[48] B. Sun, Y. Yang, L. Zhang, M.-M. Cheng, and Q. Hou. Corrmatch: Label propagation via correlation matching for semi-supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3097–3107, 2024. 1

[49] A. Tarvainen and H. Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. NeurIPS, 30, 2017. 2, 5

[50] P. Upretee and B. Khanal. Fixmatchseg: Fixing fixmatch for semi-supervised semantic segmentation. arXiv preprint arXiv:2208.00400, 2022. 2

[51] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, and L. Van Gool. Unsupervised semantic segmentation by contrasting object mask proposals. In ICCV, pages 10052–10062, 2021. 1

[52] Y. Wang, H. Chen, Q. Heng, W. Hou, M. Savvides, T. Shinozaki, B. Raj, Z. Wu, and J. Wang. Freematch: Self-adaptive thresholding for semi-supervised learning. arXiv preprint arXiv:2205.07246, 2022. 2, 5, 13, 14

[53] Y. Wang, H. Wang, Y. Shen, J. Fei, W. Li, G. Jin, L. Wu, R. Zhao, and X. Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In CVPR, pages 4248–4257, 2022. 2, 5, 6, 7, 9

[54] Y. Wang, J. Zhang, M. Kan, S. Shan, and X. Chen. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In CVPR, pages 12275–12284, 2020. 1

[55] Z. Wang, Z. Zhao, L. Zhou, D. Xu, X. Xing, and X. Kong. Conflict-based cross-view consistency for semi-supervised semantic segmentation. arXiv preprint arXiv:2303.01276, 2023. 5

[56] Y. Wei, H. Xiao, H. Shi, Z. Jie, J. Feng, and T. S. Huang. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In CVPR, pages 7268–7277, 2018. 1

[57] H. Xiao, D. Li, H. Xu, S. Fu, D. Yan, K. Song, and C. Peng. Semi-supervised semantic segmentation with cross teacher training. Neurocomputing, 508:36–46, 2022. 2

[58] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le. Self-training with noisy student improves imagenet classification. In CVPR, pages 10687–10698, 2020. 2

[59] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In CVPR, pages 1492–1500, 2017. 1

[60] H.-M. Xu, L. Liu, Q. Bian, and Z. Yang. Semi-supervised semantic segmentation with prototype-based consistency regularization. arXiv preprint arXiv:2210.04388, 2022. 1, 2, 5, 6, 7

[61] Y. Xu, L. Shang, J. Ye, Q. Qian, Y.-F. Li, B. Sun, H. Li, and R. Jin. Dash: Semi-supervised learning with dynamic thresholding. In ICML, pages 11525–11536. PMLR, 2021. 2

[62] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In CVPR, 2023. 1, 2, 3, 5, 6, 7, 9, 10, 11

[63] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In CVPR, pages 4268–4277, 2022. 1, 2, 5, 6, 10, 11

[64] J. Yuan, Y. Liu, C. Shen, Z. Wang, and H. Li. A simple baseline for semi-supervised semantic segmentation with strong data augmentation. In ICCV, pages 8229–8238, 2021. 1, 2

[65] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In ICCV, pages 6023–6032, 2019. 11

[66] B. Zhang, Y. Wang, W. Hou, H. Wu, J. Wang, M. Okumura, and T. Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. NeurIPS, 34:18408–18419, 2021. 2, 13

[67] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, et al. Resnest: Split-attention networks. In CVPR, pages 2736–2746, 2022. 1

[68] J. Zhang, T. Wu, C. Ding, H. Zhao, and G. Guo. Region-level contrastive and consistency learning for semi-supervised semantic segmentation. arXiv preprint arXiv:2204.13314, 2022. 2, 5

[69] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In CVPR, pages 2881–2890, 2017. 1

[70] Z. Zhao, S. Long, J. Pi, J. Wang, and L. Zhou. Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation. In CVPR, 2023. 2

[71] Z. Zhao, L. Yang, S. Long, J. Pi, L. Zhou, and J. Wang. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In CVPR, 2023. 6, 7

[72] Y. Zhong, B. Yuan, H. Wu, Z. Yuan, J. Peng, and Y.-X. Wang. Pixel contrastive-consistent semi-supervised semantic segmentation. In ICCV, pages 7273–7282, 2021. 2, 5

[73] Y. Zhou, H. Xu, W. Zhang, B. Gao, and P.-A. Heng. C3-semiseg: Contrastive semi-supervised segmentation via cross-set learning and dynamic class-balancing. In ICCV, pages 7036–7045, 2021. 2

[74] X. J. Zhu. Semi-supervised learning literature survey, 2005. 2

[75] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. Le. Rethinking pre-training and self-training. NeurIPS, 33:3833–3845, 2020. 2

[76] Y. Zou, Z. Zhang, H. Zhang, C.-L. Li, X. Bian, J.-B. Huang, and T. Pfister. Pseudoseg: Designing pseudo labels for semantic segmentation. arXiv preprint arXiv:2010.09713, 2020. 5, 10, 11