

基于均值教师模型联合 多级扰动的半监督遥感影像变化检测

于松岩^{1,2}, 翟钰杰^{1,2}, 许叶彤^{1,2}, 赵伟强³, 雷涛^{1,2†}

(1. 陕西科技大学 电子信息与人工智能学院, 陕西 西安 710021; 2. 陕西科技大学 陕西省人工智能联合实验室,
陕西 西安 710021; 3. 中电科西北集团有限公司西安分公司, 陕西 西安 710065)

摘 要:目前深度学习方法在遥感影像变化检测方面取得了较大的进步,然而现有的遥感影像变化检测方法仍然以全监督网络为主,其网络性能严重依赖标签数据的数量和质量。为此,提出了一种基于均值教师模型联合多级扰动的半监督遥感影像变化检测网络(UniMTCD-Net)。首先,将不同性质的强扰动分离到不同的分支分别进行学习并约束一致性,形成多样化的扰动空间,避免了单分支学习困难的问题,从而有效提升对无标签数据的利用效率;其次,采用均值教师模型,不仅扩展了教师模型生成的伪标签和学生模型输出的强预测之间的差异,同时教师模型参数通过指数移动平均(EMA)更新的方式,使得伪标签的生成更加准确。实验结果表明,与主流半监督方法相比,UniMTCD-Net 具有更好的检测性能,尤其在 5% 的标签训练数据下检测性能更加优秀,进一步验证了 UniMTCD-Net 在遥感影像变化检测中的有效性和优越性。

关键词:变化检测;半监督;一致性;均值教师模型

中图分类号:TP3

文献标识码:A

Semi-supervised Remote Sensing Image Change Detection Based on Mean Teacher Model with Multi-level Perturbations

YU Songyan^{1,2}, ZHAI Yujie^{1,2}, XU Yetong^{1,2}, ZHAO Weiqiang³, LEI Tao^{1,2†}

(1. School of Electronic Information and Artificial Intelligence, Shaanxi University of science and technology, Xi'an, Shaanxi 710021, China; 2. Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China; 3. Xi'an Branch, China Electronics Technology Group Corporation Northwest Group Corporation, Xi'an, Shaanxi 710065, China)

Abstract:Currently, deep learning methods have made significant progress in remote sensing image change detection. However, the existing approaches in remote sensing image change detection primarily rely on fully supervised networks, which heavily depend on the quantity and quality of label data. In order to solve these problems, this paper proposes UniMTCD-Net, a semi-supervised remote sensing image change detection network that combines the mean teacher model with multi-level perturbation. Firstly, different types of strong perturbations are separated into different branches for learning and consistency is constrained to form a diversified perturbation space, avoiding the difficulty of single branch learning and effectively improving the utilization efficiency of unlabeled data. Secondly, by using the mean teacher model, not only the difference between the pseudo labels generated by the teacher model and the strong predictions output by the student model are extended, but also the teacher model parameters are updated by exponential moving average (EMA), making the generation of pseudo labels more accurate. Experimental results demonstrate that compared with mainstream semi-supervised methods, UniMTCD-Net has better detection performance, especially under 5% labeled training data, the detection performance is more superior, further verifying the effectiveness and superiority of UniMTCD-Net in remote sensing image

收稿日期:2023-06-27

基金项目:国家自然科学基金资助项目(61871259;61861024),陕西省重点研发计划重点产业创新链项目(2021ZDLGY08-07)

作者简介:于松岩(1998—),男,陕西西安人,硕士研究生,研究方向:计算机视觉。

† 通信联系人, E-mail: haloosy@163.com

change detection.

Key words: change detection; semi-supervised; consistency; mean teacher model

遥感影像变化检测是遥感技术中一个重要应用领域。它通过对不同时间或不同传感器采集的数据进行比较分析,从而确定地表特征或景观在短时间内的变化情况。其在诸多领域都有广泛的应用,例如城市扩张监测^[1]、环境监测^[2]和灾害评估^[3]等。随着成像技术的快速发展,传统的遥感影像变化检测方法耗时耗力且检测精度不高,效果差。近年来,深度学习的快速发展为遥感影像变化检测注入新的活力,其精度和自动化水平远超传统方法,现已成为遥感影像变化检测领域的主流方法。然而,对于双时相遥感影像的标注存在一些问题。通常需要由专家逐像素进行比对,这个过程耗时且昂贵。另外,随着对地观测项目的不断增加,我们经常能够获得许多未标记的多时相遥感影像,如何对这些未标记图像进行有效利用则成为一大挑战。为解决上述问题,由于基于半监督的方法只需要少量标注样本即可达到较为优异的检测性能,因此更适用于遥感影像变化检测。

1 相关工作

随着深度学习的快速发展,卷积神经网络(Convolutional Neural Networks, CNN)在遥感影像变化检测方面做出了巨大的贡献。Daudt 等^[4]提出了一种用于变化检测的全卷积孪生网络,通过将 CNN 方法应用于遥感影像变化检测领域,极大地改变了传统方法提取特征质量差的缺点。

随后,针对 CNN 的方法,研究人员探索了更多的技术以提高网络性能。Chen 等提出通过注意力机制^[5]来获得更多区分特征,Fang 等^[6]则利用密集网络和多尺度信息融合来获取更丰富的语义信息和更准确的定位信息。近两年来,随着 Transformer 模型在自然语言处理任务展现出优异的性能,研究人员将其用于计算机视觉任务,如图像分类^[7]、目标检测^[8]和语义分割^[9]等,并表现了出色的性能。同时,Transformer 也逐渐被引入到遥感影像变化检测任务中,以更好地捕获上下文建模中的长程依赖关系。BiT^[10]首次将 Transformer 与 CNN 相结合,MSTDSNet^[11]则应用了

Swin Transformer,同时纯 Transformer 方法^[12]也被用于变化检测。

上述的监督学习方法都需要大量标签数据才能实现更好的泛化性能。然而,对地观测设备获取的大量遥感影像都是未标注的。因此,为了充分利用现有的未标注数据来解决变化检测问题,并减少对大量的标签训练样本的依赖,越来越多的研究人员开始探索将半监督学习方法应用于变化检测任务。Peng 等人^[13]利用生成对抗网络进行半监督训练。而一致性正则化的发展^[14-16]也为变化检测领域带来了 SiameseSSL^[17]和 Semi CD^[18]等方法。这些方法使用不同的网络结构或扰动来计算不同输出的一致性。尽管这些半监督变化检测方法已经取得了较好的结果,但目前还没有方法能够综合考虑各种扰动因素,并且充分利用伪标签的思想来研究模型在大量未标记数据中的泛化性能。

针对上述问题,设计了一个名为 UniMTCD-Net 的半监督变化检测网络,该框架统一了输入扰动、特征扰动以及模型扰动,以联合各种扰动输出一致性。首先,通过多级扰动的联合,促使模型更好地适应遥感影像的复杂场景,提升网络的鲁棒性和泛化性;其次,为了提高伪标签的质量和可靠性,本文采用均值教师模型进行伪标签的生成,并加入预定义的置信度,从而有效增强模型的训练效果。

2 UniMTCD-Net

2.1 变化检测框架

典型的遥感影像变化检测框架如图 1 所示。通常给定一对双时相遥感图像,然后对其进行特征提取后,将提取到的特征在随后的网络层中进行组合和比较以生成变化图。它可以被简单地视为一个二分类问题。

变化检测网络由以下三个主要部分组成:

编码器:利用孪生网络架构中的编码器^[19],其中编码器 f_e 的权重共享,以获得其隐藏特征表示 F_1 和 F_2 。可表示为:

$$F_1 = f_e(I_{T_1}) \quad (1)$$

$$F_2 = f_e(I_{T_2}) \quad (2)$$

其中 I_{T_1} 和 I_{T_2} 代表 T_1 和 T_2 时刻的图像。

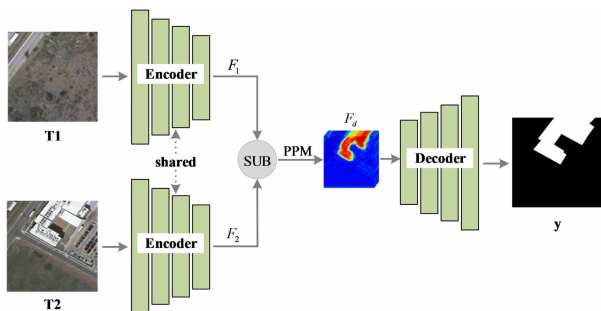


图 1 变化检测网络结构图

特征差异模块:当从编码器获得隐藏特征表示 F_1 和 F_2 时,通过计算 F_1 和 F_2 之间的绝对差来获得隐藏特征的差异,并且随后通过金字塔池化模块(PPM)^[20] f_{PPM} 对其进行处理以有效地获得不同尺度的变化信息。将上述过程的输出表示为双时

相图像的隐藏特征差 F_d 。特征差异模块内部过程表示为:

$$F_d = f_{PPM}(|F_1 - F_2|) \quad (3)$$

解码器:解码器 f_d 的目的是从隐藏特征差异 F_d 估计输出变化概率图,使用上采样操作将其恢复到与输入图像具有相同的空间分辨率。可表示为:

$$y = f_d(F_d) \quad (4)$$

2.2 网络整体结构

UniMTCD-Net 结构如图 2 所示。网络训练主要分为四个步骤:(1)从少量标记数据中学习一个初步的学生模型。(2)使用弱扰动的图像通过教师模型来获取变化区域的伪标签。(3)对于学生模型施加不同的强扰动,使用由强扰动生成的复杂图像来获得预测。(4)使用伪标签来惩罚复杂图像的结果。

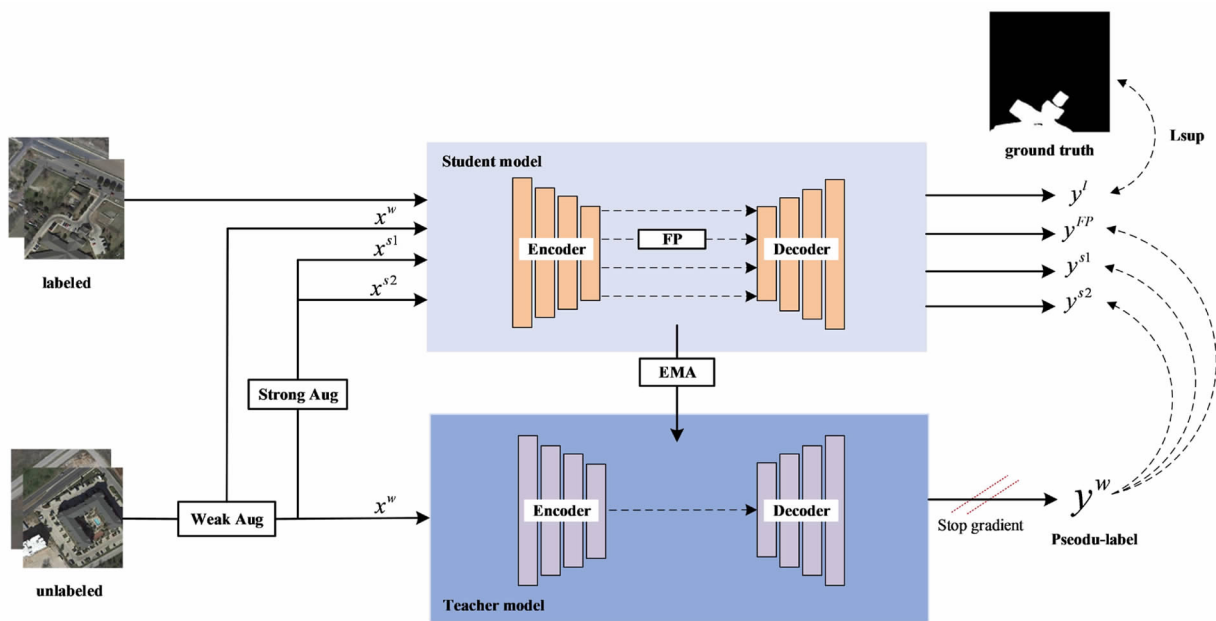


图 2 UniMTCD-Net 整体结构图

2.3 均值教师模型

尽管基于伪标签的半监督学习方法展现出了良好的性能,但将其应用于遥感影像变化检测时,由于地物轮廓模糊的边界区域通常会生成不准确的预测结果,进而在训练过程中导致错误信息积累,产生较差的结果。因此,在使用伪标签进行训练时,需要仔细考虑数据增强的选择和模型的可靠性。

针对上述问题,提出了一种半监督训练方法,使用均值教师模型构建伪标签以增加用少量标签数据监督训练模型的泛化能力。首先,使用少量标

签数据监督学生变化检测模型的训练。其次,根据均值教师(Mean Teacher, MT)方法^[14]通过指数移动平均(EMA)将每个训练轮的学生模型参数迁移到具有相同结构的教师模型。该方法改善了所有层输出而不仅是顶层输出,从而提高了教师模型的中间表示能力,产生比直接使用最终权重更好的预测结果,解决了伪标签生成质量差的问题。通过教师模型生成的伪标签对学生模型输出的强预测进行约束,有效地降低了伪标签带来的误差,提高模型的训练效果。EMA 的具体过程如下所示:

$$\theta_i = \alpha \theta_{i-1}^t + (1 - \alpha) \theta_i^s \quad (5)$$

其中 θ 表示教师模型参数, θ^s 表示学生模型参数, i 表示 epoch 的迭代次数, α 是可学习的参数, 被设置为 0.996。

2.4 联合多级扰动一致性

基于一致性正则化的半监督学习方法已成功应用于遥感影像变化检测。一致性正则化方法主要基于平滑假设^[21], 即对于输入空间中的两个相邻点, 它们的标签必须相同。因此基于该假设通过对未标记数据施加扰动, 并且训练不受这些扰动影响的模型来利用无标签数据。基于一致性正则化方法的区别主要在于对输入数据添加扰动的方式, 主要分为四个类别, 第一个是输入扰动, 采用数据增强技术对输入图像添加扰动; 第二个是特征扰动, 该扰动添加于分割网络编码器输出的特征图; 第三个是模型扰动, 通过使用不同的网络获得扰动实现预测一致, 例如具有不同起始权重的网络^[22]; 第四个是综合扰动, 其结合了前面三种扰动类型。

PS-MT^[23] 将三个级别的扰动添加到同一个分支中, 导致学习难度过大。UniMatch^[24] 则提出了一种双流扰动技术, 通过使两个强视图同时引导一个共同的弱视图, 提高模型的泛化能力, 验证了将图像级强扰动和特征级强扰动分离的必要性。本文提出将不同的强扰动分离成多个独立的前馈流以降低模型的学习难度。同时, 均值教师模型被视作为模型扰动的一种形式, 促进学生模型学习更加准确的知识表示, 从而提高检测的准确率和模型的鲁棒性。

各种扰动添加过程表示如下:

$$x^w = A^w(x^u) \quad (6)$$

$$e^w = g(x^w) \quad (7)$$

$$y^{FP} = h(P(e^w)) \quad (8)$$

其中 x^u 为无标签图像, A^w 为输入弱扰动, g 为编码器, h 为解码器, e^w 是 x^w 提取的特征图, P 表示特征扰动。

$$y^w = \hat{F}(x^w) \quad (9)$$

$$y^s = F(A^s(x^w)) \quad (10)$$

其中 A^s 是输入强扰动, $\hat{F}$ 为教师模型对弱扰动图像生成伪标签, F 为学生模型对强扰动图像进行预测。

在监督阶段, 将标签数据输入学生模型, 仅预测标签数据中感兴趣的变化区域, 并使用交叉熵 (CE) 损失作为监督损失 L_{sup} 来衡量实际变化图 $\hat{y}^i$ 与预测变化概率图 y^i 之间的性能差异。监督损失函数 L_{sup} 如下:

$$y^i = F(x^w) \quad (11)$$

$$L_{sup} = CE(y^i, \hat{y}^i) \quad (12)$$

在无监督阶段, 将无标签数据经过弱增强后送入教师模型中生成伪标签, 并在弱增强的基础上添加不同类型的强扰动, 在学生模型上对其进行预测。通过应用一致性正则化来促使模型在对同样样本的不同视图或扰动上产生一致的预测, 从而提升模型的鲁棒性和泛化能力。定义无监督损失函数 L_{unsup} 如下:

$$L_{unsup} = Y \cdot \left\{ \lambda CE(y^w, y^{FP}) + \frac{\mu}{2} [CE(y^w, y^{S1}) + CE(y^w, y^{S2})] \right\} \quad (13)$$

其中 $Y = \mathbb{I}[\max(y^w) \geq \tau]$, 损失权重 λ 和 μ 设置为 0.5, 置信度阈值 τ 设置为 0.95。总损失函数如下:

$$L = \frac{1}{2} (L_{sup} + L_{unsup}) \quad (14)$$

3 实验

3.1 实验设置

实验训练中使用的开发环境为 PyCharm 2023.1, 深度学习框架为 PyTorch 1.11.0, 并在具有 24GB VRAM 的 GeForce RTX 3090Ti GPU 上进行训练。在训练阶段, 将来自 ST++^[25] 和 CutMix^[26] 的颜色变换组合作为强输入扰动 A^s , 将对原始图像进行调整大小, 裁剪并翻转作为弱输入扰动 A^w , 将对编码器输出特征图进行通道维度上的 Dropout 作为特征扰动。学生模型采用 SGD 优化器, 将学习率设置为 0.02, 权重衰减设置为 0.0001, 动量设置为 0.9, 批量大小为 4, 迭代次数为 80。

3.2 数据集介绍

本文在公开的大型建筑物变化检测数据集 LEVIR-CD 上进行了实验验证。LEVIR-CD 数据集包含 637 对大小为 1024×1024 的高分辨率遥感图像, 在本文实验中, 通过随机裁剪将其裁剪成 256×256 大小的图像对, 并从原始训练集中随机选择实验所需数量的标签样本, 同时通过丢弃相应的变化标签将训练集中的其余图像视为无标签样本。训练集按标签样本和无标签样本划分为 [5%, 95%], [10%, 90%], [20%, 80%] 和 [40%, 60%]。裁剪后的标签样本示例如图 3 所示。



图 3 LEVIR-CD 裁剪数据集

3.3 实验结果与分析

为了证明所提方法的优越性,使用 s4GAN^[27]、SemiCDNet^[13]、SemiCD^[28] 和 UniMatch^[24] 作为对比实验。在实验中,LEVIR-CD 建筑物数据集的标记率被设置为[5%,10%,20%,40%],以总准确度(OA)和交并比(IoU^c)的类别变化区域(IoU^c)作为评估指标。实验结果如表 1 和表 2 所示。

表 1 检测模型在不同标记率下的 OA 性能指标/%				
模型	5%	10%	20%	40%
s4GAN	97.89	98.11	98.51	98.62
SemiCDNet	98.17	98.42	98.58	98.63
SemiCD	98.47	98.63	98.68	98.72
UniMatch	98.62	98.83	98.84	98.79
UniMTCD-Net	98.89	98.96	98.98	98.99

表 2 检测模型在不同标记率下的 IoU ^c 性能指标/%				
模型	5%	10%	20%	40%
s4GAN	64.0	67.0	73.4	75.4
SemiCDNet	67.6	71.5	74.3	75.5
SemiCD	72.5	75.5	76.2	77.2
UniMatch	75.6	79.0	79.0	78.2
UniMTCD-Net	79.8	81.1	81.3	81.6

表 1 和表 2 根据不同的评估指标展示了不同检测模型的性能。结果表明,UniMTCD-Net 在 5%、10%、20% 和 40% 四个分区中均获得了最佳结果。尤其在仅有 5% 的真实标签数据的情况下,UniMTCD-Net 实现了 79.8% 的 IoU^c 和 98.89% 的 OA,比第二名高出 4.2% 和 0.27%,这进一步验证了 UniMTCD-Net 的可行性以及其在变化检测具有更加优异的检测精度。

为了更直观地体现出 UniMTCD-Net 的优越性,图 4 展示了不同检测模型的变化检测结果可视化。从图中可以看出,UniMTCD-Net 的检测结果与标签图像非常相近,并且边界更加明确,线条更

加流畅,明显优于其他变化检测网络。

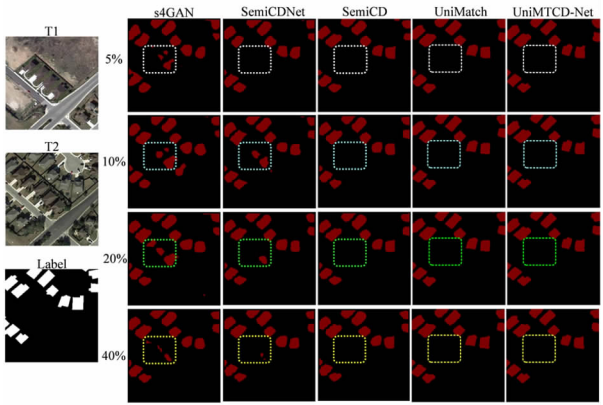


图 4 不同检测模型的结果图对比

4 结 论

UniMTCD-Net 为半监督遥感影像变化检测任务中存在的问题提供了一种有效且可行的解决方案。这项研究的主要贡献在于提出通过联合多级扰动一致性和均值教师模型的方法,以生成更可靠的伪标签,解决模型对标签数据依赖的问题。在 LEVIR-CD 数据集上的实验证明了本文所提方法的有效性,并与目前最先进的半监督变化检测方法(即 UniMatch)进行比较,结果表明,UniMTCD-Net 在性能上具有显著优势,取得了令人满意的结果。但仍存在需要改进的部分,本文在生成伪标签时设置的置信度阈值来抑制可能不正确的伪标签。因此在未来工作中,将重点研究如何充分利用这些不确定的像素,进一步完善算法。

参考文献

[1] HAERYUN L, KYUNGSU L, HEE J K, et al. Local similarity siamese network for urban land change detection on remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2021,14.

[2] 曹飞,穆宝慧,徐丹,等. 遥感技术在环境变化监测中的应用进展[J]. 环境与可持续发展,2020,45(2):96—99. DOI:10.19758/j.cnki.issn1673—288x.202002096.

[3] LEI T, ZHANG Y X, LV Z Y, et al. Landslide inventory mapping from bitemporal images using deep convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters,2019,16(6).

[4] DAUDT R C, LE SAUX B, BOULCH A. Fully convolutional siamese networks for change detection[C]//2018 25th IEEE International Conference on Image Processing (ICIP).

- IEEE, 2018: 4063–4067.
- [5] CHEN H, SHI Z W. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection[J]. *Remote Sensing*, 2020, 12(10): 1662.
 - [6] FANG S, LI K Y, SHAO J Y, et al. SNUNet-CD: a densely connected Siamese network for change detection of VHR images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 1–5.
 - [7] CHEN C F R, FAN Q, PANDA R. Crossvit: cross-attention multi-scale vision transformer for image classification [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 357–366.
 - [8] HAN G X, MA J W, HUANG S Y, et al. Few-shot object detection with fully cross-transformer [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 5321–5330.
 - [9] WANG L B, LI R, ZHANG C, et al. UNetFormer: a UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 190: 196–214.
 - [10] CHEN H, QI Z P, SHI Z W. Remote sensing image change detection with transformers[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1–14.
 - [11] SONG F, ZHANG S X, LEI T, et al. MSTDSNet-CD: multiscale swin transformer and deeply supervised network for change detection of the fast-growing urban regions[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1–5.
 - [12] BANDARA W G C, PATEL V M. A transformer-based siamese network for change detection [C]//*IGARSS 2022 – 2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022: 207–210.
 - [13] PENG D F, BRUZZONE L, ZHANG Y J, et al. SemiCD-Net: a semisupervised convolutional neural network for change detection in high resolution remote-sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 59(7): 5891–5906.
 - [14] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. *Advances in Neural Information Processing Systems*, 2017, 30.
 - [15] SOHN K, BERTHELOT D, CARLINI N, et al. Fixmatch: simplifying semi-supervised learning with consistency and confidence[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 596–608.
 - [16] OUALI Y, HUDELOT C, TAMI M. Semi-supervised semantic segmentation with cross-consistency training [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020: 12674–12684.
 - [17] HAFNER S, BAN Y, NASCETTI A. Urban change detection using a dual-task siamese network and semi-supervised learning [C]//*IGARSS 2022 – 2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022: 1071–1074.
 - [18] BANDARA W G C, PATEL V M. Revisiting consistency regularization for semi-supervised change detection in remote sensing images[J]. *arXiv preprint arXiv:2204.08454*, 2022.
 - [19] BROMLEY J, GUYON I, LECUN Y, et al. Signature verification using a " siamese " time delay neural network[J]. *Advances in Neural Information Processing Systems*, 1993, 6.
 - [20] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 2881–2890.
 - [21] CHAPELLE O, SCHOLKOPF B, ZIEN A. Semi-supervised learning (chapelle, o. et al., eds.; 2006) [book reviews] [J]. *IEEE Transactions on Neural Networks*, 2009, 20(3): 542–542.
 - [22] CHEN X K, YUAN Y H, ZENG G, et al. Semi-supervised semantic segmentation with cross pseudo supervision [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021: 2613–2622.
 - [23] LIU Y Y, TIAN Y, CHEN Y H, et al. Perturbed and strict mean teachers for semi-supervised semantic segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 4258–4267.
 - [24] YANG L H, QI L, FENG L T, et al. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 7236–7246.
 - [25] YANG L H, ZHUO W, QI L, et al. St++: Make self-training work better for semi-supervised semantic segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 4268–4277.
 - [26] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019: 6023–6032.
 - [27] MITTAL S, TATARCHENKO M, BROX T. Semi-supervised semantic segmentation with high-and low-level consistency[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 43(4): 1369–1379.