

基于深度学习的半监督语义分割算法研究

李一彤, 张长伦*

北京建筑大学理学院, 北京

收稿日期: 2023年10月20日; 录用日期: 2023年11月10日; 发布日期: 2023年11月22日

摘要

语义分割是计算机视觉中最具挑战性的任务之一。然而, 在许多应用中, 由于像素级标记的高成本, 缺乏标记图像是一个常见的问题。这极大地阻碍了在一些注释大量图像成本高甚至不可行的领域的广泛应用。为了缓解这一问题, 半监督语义分割被提出并受到广泛的关注, 其中标记图像和未标记图像都被充分利用。本文首先介绍语义分割技术并引入半监督语义分割, 然后对深度学习在半监督语义分割中的常用方法分类, 随后重点对每种方法及其经典网络结构详细介绍并进行优劣对比, 对现有的方法提出了几种可行的改进策略, 最后对未来发展进行总结与展望。

关键词

深度学习, 语义分割, 半监督学习

Algorithm Research of Deep Learning in Semi-Supervised Semantic Segmentation

Yitong Li, Changlun Zhang*

School of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: Oct. 20th, 2023; accepted: Nov. 10th, 2023; published: Nov. 22nd, 2023

Abstract

Semantic segmentation is one of the most challenging tasks in computer vision. However, lack of marked images is a common problem in many applications due to the high cost of pixel-level marking. This greatly hinders widespread application in some fields where annotating large numbers of images is costly or even unfeasible. In order to alleviate this problem, semi-supervised semantic segmentation has been proposed and received widespread attention, in which both la-

*通讯作者。

beled and unlabeled images are fully utilized. This paper first introduces semantic segmentation technology and introduces semi-supervised semantic segmentation, then classifies common methods of deep learning in semi-supervised semantic segmentation, then focuses on each method and its classical network structure in detail and compares its advantages and disadvantages, puts forward several feasible improvement strategies for existing methods, and finally summarizes and looks forward to the future development.

Keywords

Deep Learning, Semantic Segmentation, Semi-Supervised Learning

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着计算机视觉和深度学习的快速发展, 图像语义分割成为了人工智能领域中一个备受关注的研究方向。语义分割用相应的语义类对每个像素进行分类, 从而对图像中属于该类的所有对象或区域给予相同的类标签, 旨在将图像分割成多个语义上有意义的区域。近年来语义分割在医学成像[1]、自动驾驶[2]、航拍场景分析[3]或金相图像[4]等不同领域都得到了重要应用[5] [6]。

图像语义分割离不开样本数据集, 根据地面真实标签(即被认为是正确的输出)的详细程度以及标记样本占所有可用图像总数的比例, 我们可能面临不同的场景: 完全监督场景、弱监督场景、半监督场景和无监督场景。由于在像素级别标记图像涉及的困难和工作量, 基于半监督学习(SSL)的方法[7]变得越来越重要, 其中我们有少量标记图像和大量未标记图像。这些半监督方法从有标签数据中以监督方式提取知识, 并从无标签数据中以无监督方式提取知识, 从而减少了在完全监督场景下所需的标记工作, 并获得明显优于无监督场景的结果。

传统的语义分割方法通常依赖于大量带有精确标签的训练数据, 这使得模型能够从大量已知信息中学习并进行预测。然而, 获得大规模高质量标记数据的过程往往耗费时间、金钱和人力资源, 限制了这些模型、方法的实际应用。为了克服这一限制, 半监督语义分割(SSS)应运而生, 它利用少量带有标签的数据与大量未标记数据相结合, 有标签数据可用于监督学习, 而无标签数据则用于提高模型的泛化能力, 以此来提高模型性能。这一领域的研究已经取得了显著的进展, 成为计算机视觉研究的一个热点话题。

半监督语义分割不仅仅是将监督和无监督学习方法简单地结合在一起, 它还涉及到许多复杂的技术挑战。例如, 标签的稀缺性、数据分布不一致性和有效地利用有限标签信息都是需要解决的问题。此外, 半监督语义分割的成功应用也需要考虑不同领域的实际需求和限制。

本文旨在全面探讨半监督语义分割的核心概念、现有方法、面临的挑战以及未来发展方向。我们将深入研究半监督学习的基本原理, 并详细介绍每一类技术和其中的经典网络框架。同时, 我们将探讨半监督语义分割中的技术挑战, 并对现有的方法及未来发展进行总结与展望。

2. 语义分割

半监督方法的性能在很大程度上取决于其所基于的监督模型的良好选择、拟合和训练。因此, 在本节中, 我们将介绍这些监督分割模型的背景。

最初, 在深度学习技术中, 通常是卷积神经网络(CNN) [8]被提出并应用于图像分类问题, 相对于之前使用的传统技术(如随机森林[9]、支持向量机[10]等), 获得了质的飞跃。由于获得了良好的效果, 这些技术被扩展到 CV 的其他领域, 试图解决越来越复杂和像素级的问题, 如目标检测[11]和分割[12] [13]。

2014 年, Long 等人[14]首次将卷积神经网络技术应用于图像语义分割任务, 提出了全卷积神经网络(FCN)。FCN 将图像级分类任务中的卷积神经网络[12] [15]的全连接层替换为卷积层, 并使用上采样操作将预测结果恢复至原图像大小, 从而得到整幅图像的像素级分类预测。

随着 FCN 的出现, 许多研究人员开始对基于全卷积神经网络的图像语义分割模型进行研究。首先是基于医学图像的语义分割模型, 2015 年, Ronneberger 等人[12]研究并提出了适用于医学场景的图像分割网络 UNet, 该网络采用编码器 - 解码器的网络架构, 编码器用于提取图像的特征, 解码器则用于恢复图像分辨率。Zhou 等人[16]提出了 UNet++, 通过集成不同深度的 UNet 和利用深度监督策略来解决网络深度难以确定的问题。2016 年, Milletari 等人[17]提出了适用于三维图像分割的网络 VNet。

基于自然图像的语义分割模型的研究重点在于对图像上下文特征的提取。2015 年, Yu 等人[18]提出了空洞卷积, 在普通卷积操作的基础上引入扩张率的概念, 从而有效提升了网络模型聚合上下文信息的能力。2017 年, Zhao 等人[19]提出了 PSPNet, 该网络使用了空间金字塔池化(SPP)技术, 通过聚集不同尺度的特征图以挖掘多层次的语义信息。2017 年至 2018 年, Google 研究员提出 DeepLab 系列[20] [21]的语义分割网络, 设计了空洞空间金字塔池化(ASPP)层, 结合了空洞卷积和金字塔结构, 进一步提高了网络的语义感知能力。

然而它们的成功密切依赖于具有高质量像素级注释的大型数据集。精细、密集的像素级标注成本高、耗时长, 在标注数据有限的情况下成为实际应用中的一个重大瓶颈。为此, 提出了半监督语义分割[22] [23], 并被应用于在较少标记但大量未标记的数据上训练模型。

3. 历史发展与现状

在本节中, 对半监督语义分割问题提出的技术进行了分类和详细解析。根据半监督语义分割文献中现有方法的性质和主要特点, 我们总结了一种分类方法, 将这些方法分为五类。

第一类为类似 GAN 结构和在两个网络之间进行对抗性训练的方法, 一个作为生成器, 另一个作为鉴别器。

第二类为一致性正则化方法。这些方法在损失函数中包含一个正则化项, 以最小化同一图像的不同预测之间的差异, 这些差异是通过图像或相关模型应用扰动获得的。

第三类为伪标记方法。一般而言, 这些方法依赖于先前对未标记数据所做的预测, 以及在标记数据上训练的模型以获得伪标签。

第四类为基于对比学习的方法。这种学习范式将相似元素分组, 并将它们与特定表示空间中的不同元素分开。

最后一类为混合方法, 即将一致性正则化、伪标记和对比学习等方法组合构成。

接下来我们将详细介绍每一类别的关键技术, 并对每一类别分析优劣。

3.1. 对抗性训练方法

半监督语义分割中的对抗性训练方法是一种利用对抗学习的思想, 将生成对抗网络(GANs) [24]或对抗性损失引入模型训练的技术。这种方法通过引入对抗性成分, 希望模型能够更好地利用未标记数据, 提高模型的泛化能力和性能。

一个典型的 GAN 框架由两个网络组成: 生成器和判别器。生成器的目的是学习目标数据的分布, 从

而允许从随机噪声中生成合成图像。判别器的目的是区分真实图像(属于真实分布)和假图像(由生成器生成)。这些网络的训练过程是以对抗的方式进行的。生成器试图混淆鉴别器, 生成与目标分布越来越相似的图像, 判别器试图提高其区分真实和虚假图像的能力。

3.1.1. 关键技术

AdvSemiSeg [25]是半监督语义分割领域最早的文章之一。与弱监督领域通常采用的分类级标签数据和分类级损失函数不同, 半监督学习更强调少量有标签数据与大量无标签数据的结合, 其核心在于如何通过有标签数据更好地挖掘无标签数据的监督信息, 从而达到提升模型性能而降低人力支出的目的。

AdvSemiSeg 在训练时同时使用了带有标注的图像和未标注的图像。如下图 1 所示, 当使用标注图像时, 分割网络同时受到基于地面真实标签的标准交叉熵损失 L_{ce} 和基于判别网络的对抗性损失 L_{adv} 约束, 注意, 训练判别网络仅仅使用标注的数据。当使用未标注图像时, 用分割网络得到初步分割结果, 然后将初步分割结果送入判别网络得到置信度图, 将置信度图作为监督信号, 对初步分割预测结果进行掩码操作后作为标签, 用自学习的方法通过 L_{semi} 训练分割网络。

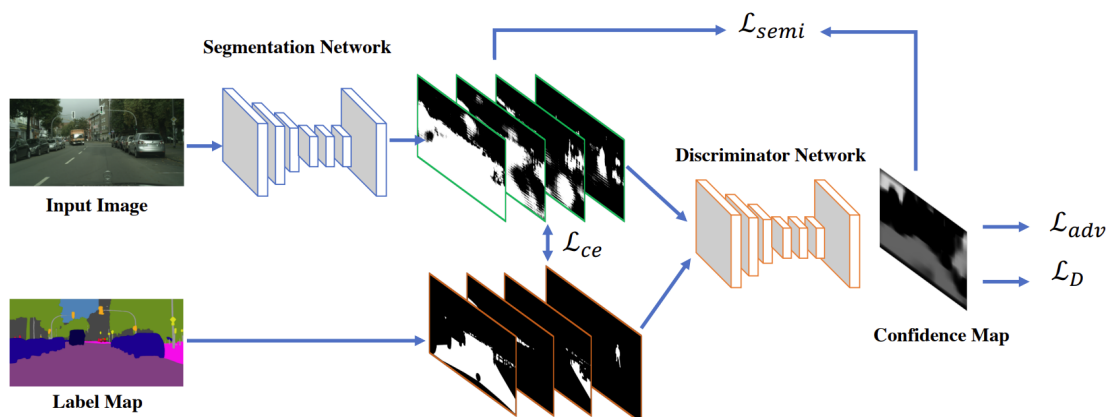


Figure 1. AdvSemiSeg network structure
图 1. AdvSemiSeg 网络结构

3.1.2. 优劣分析

对抗性训练方法的优势在于它可以通过生成对抗网络引入对抗性信号, 提高模型的鲁棒性。能够有效应对一些复杂和困难的场景, 如对抗攻击、不稳定的数据分布等。对于半监督学习来说, 最重要的就是更高效利用未标记数据, 而引入的对抗性信号恰恰可以解决这一问题。同时对抗性训练方法也适用于多模态和多领域的语义分割任务。

然而对抗性训练方法也存在一定的劣势。其训练过程较为复杂, 需要平衡生成器和鉴别器的性能, 这就需要更多计算资源, 对实验环境有了一定的限制。同时它也受限于生成器和鉴别器的性能, 可能对模型训练的稳定性产生挑战。

3.2. 一致性正则化方法

许多机器学习方法中, 尤其是深度学习中的神经网络, 都存在几个问题, 一个是模型容易过拟合, 另一个是模型在受到微小扰动、噪声后, 预测结果会受相当程度的影响。为了减少过拟合现象, 典型的监督学习中会添加一个新的损失项。在半监督学习中, 同样存在一种正则化方法, 即一致性正则化。

具体来说, 基于平滑假设和聚类假设, 具有不同标签的数据点在低密度区域分离, 并且相似的数据

点具有相似的输出。那么, 如果对一个未标记的数据应用实际的扰动, 其预测结果不应该发生显著变化, 也就是输出具有一致性。由于这种方法一般基于模型输出的预测向量, 不需要具体的标签, 所以其刚好能应用于半监督学习。通过在未标记数据上构造添加扰动后的预测结果 \hat{y} 与正常预测结果 y 之间的无监督正则化损失项, 提高模型的泛化能力。

3.2.1. 关键技术

基于一致性正则化的半监督语义分割方法之间的主要区别在于它们将扰动纳入数据的方式。基于此, 我们可以将这些方法分为四个子类别。

第一种为基于输入扰动的方法[26]-[31], 这也是最常见的一种一致性正则化方法。这些方法使用数据增强技术直接对输入图像施加扰动。对于给定的输入图像, 通过应用不同的数据增强或变换, 产生两个甚至多个扰动后的输入表示。这些变换可以包括旋转、翻转、缩放、裁剪、色彩变换等。使用原始图像和其对应的扰动图像作为模型输入, 产生对应的语义分割结果。然后引入一致性正则化, 鼓励模型在这些扰动后的输入上产生一致的输出, 即对于同一图像的不同扰动版本, 模型的输出应该保持一致。基于特征扰动的半监督语义分割方法充分利用了数据增强的思想, 通过对输入数据进行多种扰动, 鼓励模型产生稳定、一致的输出, 提高模型的泛化能力。故而这种方法常在输入的扰动上有所创新, 例如 CutMix [28]和 CutOut [29]技术, 其中 CutMix 的应用至今仍然十分广泛, 它的主要思想是将两张不同图像的一部分区域(矩形区域)切割并交换, 然后将这个混合的图像作为模型输入, 同时调整标签以反映混合图像中每个区域的贡献。原始预测和增强预测(调整后得到的新标签)之间的差异与之前使用正则化项的方法相同。除此之外, 近年来仍有很多文章提出了关于输入扰动上的创新, ComplexMix [26]提出了结合使用之前的数据增强技术, 即 CutMix 和 ClassMix [28]。

其次是基于特征扰动的方法, 将扰动纳入分割网络内部, 从而获得增强后的特征。交叉一致性训练(CCT) [32]被提出来解决半监督的语义分割问题。利用有标记的样本来训练编码器和解码器, 再利用这个训练的模型来训练未标记的样本, 从而产生伪标签。CCT 中很重要的一步就是在训练未标记样本的时候会对于编码器的输出加入多个扰动, 然后利用辅助的解码器使得输出结果与主编码器保持一致性。对于编码器的输出加入扰动这一技术就相当于对分割网络内部得到的特征进行了扰动, 如图 2 所示。

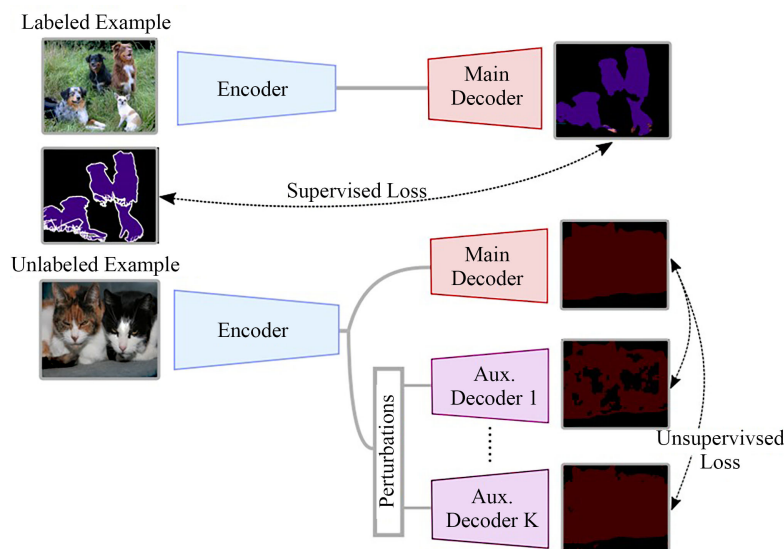


Figure 2. Cross consistency training network

图 2. 交叉一致性训练 CCT 网络

第三种为基于网络扰动的方法, 通过使用不同的网络获得扰动预测, 例如, 具有不同起始权值的网络[33] [34] [35]。网络之间的差异构成了最终预测中的扰动, 这就是交叉伪监督(CPS)方法的情况[34]。用两个相同结构、但是不同初始化的网络, 添加约束使得两个网络对同一样本的输出是相似的。具体来说, 当前网络产生的 one-hot 伪标签, 会作为另一路网络预测的目标, 这个过程可以用交叉熵损失监督, 就像传统的全监督语义分割任务的监督一样, 如图 3 所示。

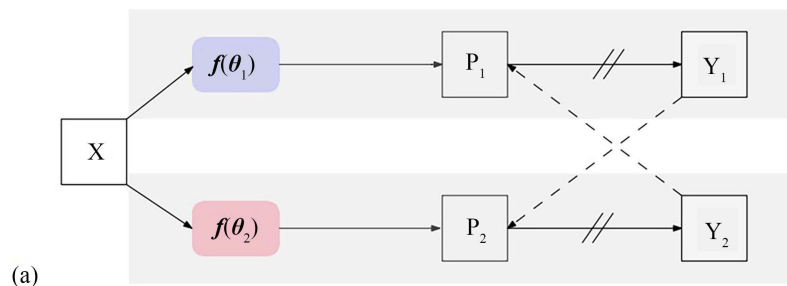


Figure 3. Cross pseudo supervision network
图 3. 交叉伪监督 CPS 网络

最后一个子类别结合了前面三种类型的扰动[36] [37]。半监督语义分割的扰动和严格平均教师 Mean Teacher [37]这一文章提出了一种结合输入、特征和网络扰动的方法。然而这种方法也存在一个问题, 即如果预测不够准确, 更大的扰动种类和强度可能会导致更多的问题。因此就需要定义额外的约束或损失项来防止模型的预测向偏离的方向发展。

3.2.2. 优劣分析

一致性正则化通过一致性损失来提高模型的鲁棒性, 使其对输入数据的变化更加稳健。它的优势在于训练过程相对简单, 与上述的对抗性训练方法相比它不需要额外的对抗网络。同时这种方法能简单有效利用未标签数据。

然而一致性正则化方法可能需要更多的标签样本来获得好的性能, 否则在约束未标签样本一致性的情况下, 容易出现过拟合的情况。它对不同任务需要不同的正则化项和参数设置, 需要进行调优来选择出最佳方案。

3.3. 伪标签方法

基于伪标签的半监督语义分割方法是半监督学习中常用的一种策略。是最广为人知的方法之一, 也是最早出现的半监督方法[38]。该方法通过使用模型对未标记数据生成的伪标签来进行训练。首先, 模型利用已标记的样本进行有监督训练, 然后利用模型对未标记数据的预测结果生成伪标签, 将这些伪标签与已标记数据一起构成一个扩充的训练集, 再进行模型更新和优化。

3.3.1. 关键技术

下面我们将对伪标签这一方法进行更详细的解析。

首先第一步进行有监督训练, 利用少量的已标记样本进行有监督训练, 训练初始模型。然后是伪标签的生成。使用初始模型对未标记样本进行预测, 生成伪标签, 即模型对未标记样本的预测结果。可以采用阈值策略或其他选择伪标签的方式, 如选择置信度高于某个阈值的预测结果作为伪标签。第三步为构建扩充训练集, 即将已标记样本和生成的伪标签样本合并, 构建一个扩充的训练集。这样, 原始有限的标记样本集就得到了扩充, 使得模型可以从未标记数据中学到更多信息。接着进行模型的更新与优化,

使用扩充训练集对模型进行更新和优化,采用常规的监督学习优化算法,如随机梯度下降(SGD)或其变体,以最小化损失函数。在这一步中模型会在包含有标签样本和伪标签样本的扩充训练集上进行训练,以调整模型参数,使其能够更好地拟合标签信息。最后就是迭代训练,重复上述步骤多次,迭代地生成伪标签、扩充训练集、更新模型,以逐步提高模型性能。

上述的流程称为自训练方法[39]-[45],它是最简单的伪标签和半监督方法,只基于一个监督基础模型,代表了最简单的伪标记形式,其中伪标签是由它们自己的高置信度预测生成的。

在自训练过程中整合数据增强技术也以不同的方式被提出。ST++ [39]方法在自训练过程中对未标记的图像应用数据增强技术。在自训练过程的每次迭代中,那些具有可靠伪标签的图像被优先考虑,而那些在伪标签中出现错误的概率更高的图像被丢弃。

3.3.2. 优劣分析

伪标签方法的优势是相比于其它方法可以更充分利用未标签数据,因为它通过生成伪标签来扩展标记数据集,整个方法都依赖于伪标签的生成,适用于小样本情况。能够在有限标签样本下获得良好性能。

然而它的劣势在于伪标记的质量受模型准确性的限制,可能引入噪声。对未标记数据的预测可能不稳定,导致不一致性伪标签。如果未标记数据的预测结果过于偏离,那么模型就面临着崩塌的风险。

3.4. 对比学习方法

对比学习是一种特征学习范式。在这种范式下,给定已知的样本分类,目标是 minimized 同一类样本之间的距离,同时最大化不同类样本之间的距离。常用于无监督学习和自监督学习任务。它学习了一个相似函数,使相同数据的视图在表示空间中更接近,同时将不同数据的视图分开。大多数最新的对比框架基于视图的全局表示来学习相似性分数,用单个向量参数化数据。另一方面,密集表示依赖于像素级表示,提供额外的监督,捕获对应的像素。基于密集表示的对比预训练最近得到了探索,并且在密集预测任务中表现出更好的性能,例如目标检测和关键点检测。

3.4.1. 关键技术

对比学习用于语义分割的研究近年来得到了广泛的应用,采用了多种不同的设计策略。部分方法基于生成的辅助标签和真值标签,通过预训练进行对比学习[46] [47],但代价是巨大的内存消耗。ReCo 方法[48]是最早提出的半监督对比学习方法之一。

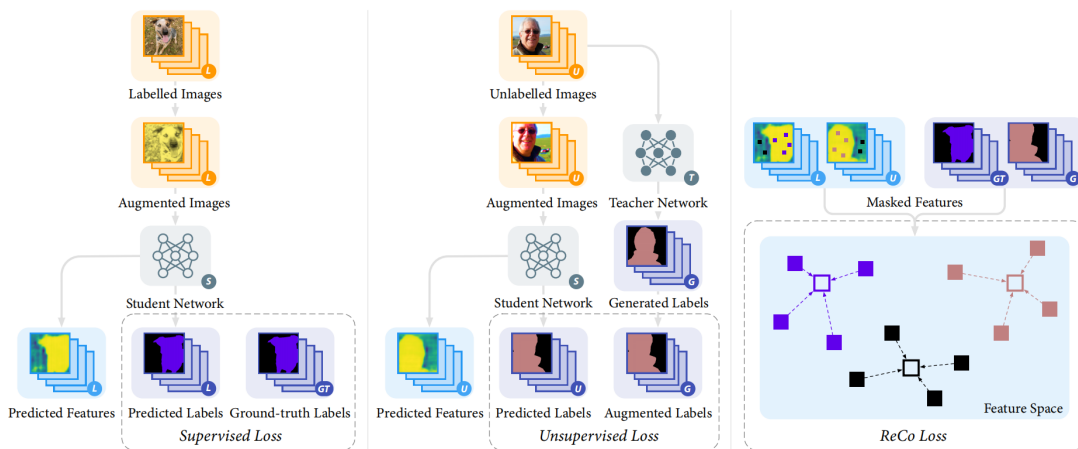


Figure 4. Regional contrast network
图 4. 区域对比 ReCo 网络

该方法包括在分割模型编码器之上链接一个辅助解码器, 该解码器将输入特征映射到更高维度的表示空间, 在该表示空间中进行查询目标和键的采样。通过所提出的对比损失函数, 强制查询像素在表示空间中靠近正键, 而远离负键。由于使用高维图像的所有像素来计算对比损失函数是不切实际的, ReCo 方法采用了一种主动采样策略, 该策略对图像中总像素的采样少于 5%。一方面, 该方法提出更高概率被选中为负键的像素是那些通常与查询类混淆的类别像素。另一方面, 它依靠预测置信度选择分割模型中较难分类的像素作为查询像素。

图4为应用于半监督语义分割和三损失训练的 ReCo 框架的可视化。监督损失是基于带有 GroundTruth 注释的标记数据计算的。对于生成伪标签的未标记数据, 计算无监督损失。最后, 根据标记和未标记图像预测的像素级密集表示计算 ReCo 损失。

3.4.2. 优劣分析

对比学习的优势在于它强调样本之间的差异性, 通过学习良好的数据表示或特征表示, 使得这些表示能够捕获数据样本的内在信息, 从而使相似样本的表示更加接近。这有助于提高模型的鲁棒性和泛化性能。与对抗性训练方法相同对比学习也适用于多领域和多模态的任务, 可以应用于各种类型的数据。

然而对比学习训练需要额外的对比样本对, 需要更多的计算资源。如果对整张图像的所有像素都做对比, 将会需要大量的内存和时间。这就制定合适的策略来选择查询样本和负样本。因此对比学习的模型设计和参数选择都具有挑战性。

3.5. 方法间的发展演变关系及其混合

3.5.1. 发展演变关系

早期, 一致性正则化和伪标记方法是首次应用于半监督语义分割的方法之一。它们主要关注如何有效地利用未标签数据, 通过一致性损失或生成伪标签来提高模型性能。这些方法的发展受到了半监督学习领域的启发, 特别是半监督分类。

之后随着生成对抗网络(GAN)的兴起, 对抗性训练方法开始应用于半监督语义分割。而对比学习方法起初主要应用于自监督学习领域, 它不需要标签来进行训练, 只需要在特征层面学习。这些方法随后逐渐被引入到半监督语义分割中, 强调如何通过最大化同类样本的相似性和不同类样本的差异性来提高性能。

随着半监督语义分割研究的深入, 研究人员开始将不同方法结合起来, 形成混合方法。这些混合方法旨在充分利用不同方法的优势, 解决单一方法的局限性。

3.5.2. 混合方法

最后一类为混合方法, 即将一致性正则化、伪标记和对比学习等方法组合构成。

将伪标签和一致性正则化这两种方法是这一类中最常见的一些做法。其中经典算法 FixMatch [49]就是将这两者结合起来的一种方法。

见图 5, Fixmatch 的方法比较简单, 它对当时现有的半监督语义分割方法进行了大幅简化。首先, FixMatch 对无标记图像的弱增强视图进行预测, 生成伪标签, 并且只有当模型产生高置信度预测时, 才会保留伪标签, 这一步可以看作是经典的伪标签方法。然后, FixMatch 对同一无标记图像的强增强视图进行预测, 并且将预测结果与上述伪标签进行匹配, 即利用伪标签来监督强增强视图这一分支进行训练。然后计算一致性损失, 以此来训练模型, 这一步就可以看作为一致性正则化方法。FixMatch 不但简单易行, 还结合了两种半监督学习方法的优点, 在各种半监督学习基准上都达到了最先进的性能。

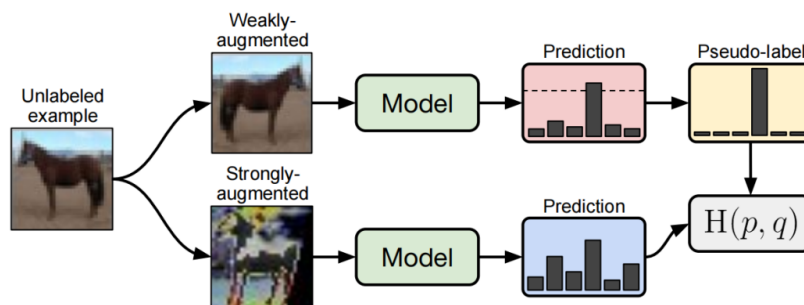


Figure 5. FixMatch network
图 5. FixMatch 网络

近年来对一致性正则化与对比学习相结合的方法的兴趣最近也有所增加。在这方面，已经提出了定向上下文感知(DCA) [50]等方法。在半监督学习中，模型容易过拟合有限的带标注数据，但是常用的分割网络 backbone 都是充分利用图像上下文信息的。在这种情况下，模型很容易过度依赖上下文进行预测，导致对训练期间未见过的场景的泛化能力较差。因此，为了防止模型滥用上下文信息，也提高模型的 self-awareness (即指对主体目标本身的关注)，作者引入了对比学习，目的是使得图像表征使对不断变化的环境更具有鲁棒性，如图 6 所示。

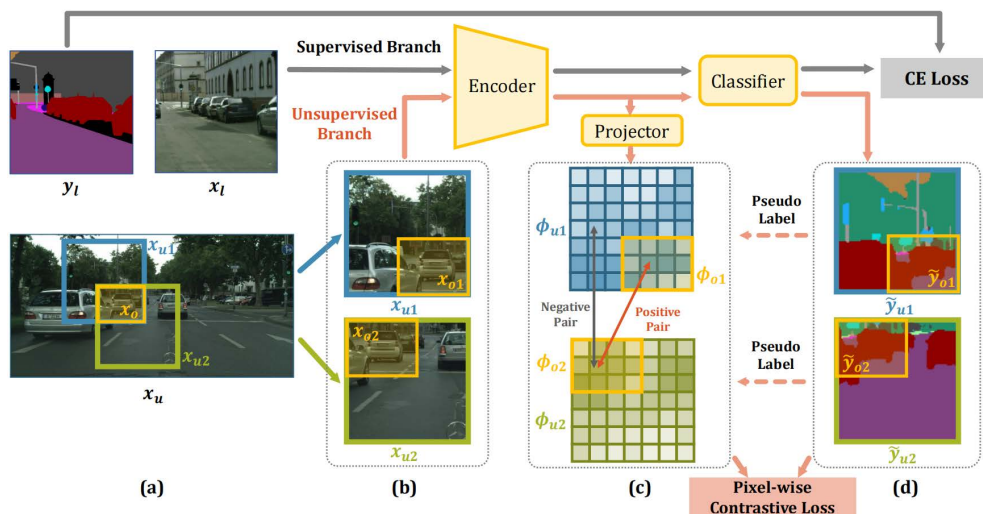


Figure 6. Directional context-aware consistency network
图 6. 定向上下文感知一致性 DCA 网络

如上图所示，从未标记的图像中裁剪两个随机补丁，它们被限制在一个重叠区域，可以认为重叠区域被放置在两个不同的环境中，即上下文增强。尽管真实标签是未知的，但仍然可以保持不同环境下高级特征的一致性。则目标就是保持重叠区域对应的特征映射之间像素到像素的一致性，这就利用了对比学习的核心思想。

4. 研究展望

半监督语义分割作为计算机视觉领域的研究热点之一，旨在充分利用有标签样本和大量未标签样本来改进语义分割模型的性能。随着深度学习的迅速发展和应用，半监督语义分割的研究也日益深入，展现出许多前景和挑战。其中关于模型的创新、实际场景的应用等方面都是具有挑战性的研究方向。

首先, 未来半监督语义分割领域将迎来更多模型创新。目前, 半监督语义分割中主要采用的方法包括生成对抗网络、对比学习、伪标签、一致性正则化等。未来的研究将探索更多新的模型架构、损失函数设计、数据扰动增强策略、特征提取方法等, 以进一步提高模型的性能和泛化能力。通过前文中的优劣对比可以看出, 没有一种方法适用于所有情况, 而是需要根据具体任务、数据和计算资源来选择适当的方法。通常, 深度学习中的实验和调优是必不可少的, 以确定哪种方法最适合解决特定问题。混合方法在一些情况下可能是最具前景的, 然而如何进行混合才能更有效也是现在研究的重点问题。

其次, 半监督语义分割将广泛应用于各种实际场景。半监督语义分割在自动驾驶、医疗图像分割、航空航天、遥感图像、农业等领域有着广泛的应用前景。例如, 在自动驾驶中, 半监督语义分割可以提高车辆对道路、行人、车辆等场景的理解, 提高自动驾驶的安全性和稳定性。在遥感图像分割领域上, 超高分辨率图像的标签标注十分耗时耗力, 故而也需要半监督学习思想的融入。

然而, 半监督语义分割依然面临诸多挑战。首先, 如何利用未标签样本的信息以及标签样本的信息之间的平衡是一个关键问题。需要深入研究在不同场景下的合适策略, 以确保模型的性能和泛化能力。其次, 数据分布的偏斜、数据不一致性、标签噪声等问题也是制约半监督语义分割性能的因素。未来的研究需要深入解决这些问题, 提高模型的鲁棒性。

最后, 我们根据前文中的发展演变关系可以看出, 这些方法的发展演变关系是相互影响和交叉的。半监督语义分割领域不断吸收来自半监督学习、生成对抗网络、对比学习等领域的创新, 以探索如何更好地利用未标签数据, 提高模型性能。随着研究的深入, 新的方法和混合方法不断涌现, 为半监督语义分割提供了更多的工具和技术。

综合而言, 半监督语义分割作为计算机视觉领域的前沿研究方向, 具有广泛的研究前景和实际应用价值。通过模型创新、应用场景拓展、实用性提高以及解决实际挑战, 我们可以预见半监督语义分割将在未来取得更多突破和进展, 为计算机视觉和人工智能领域的发展做出更大贡献。

参考文献

- [1] Medley, D.O., Santiago, C. and Nascimento, J.C. (2021) Cycoseg: A Cyclic Collaborative Framework for Automated Medical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 8167-8182. <https://doi.org/10.1109/TPAMI.2021.3113077>
- [2] Orsic, M., Kreso, I., Bevandic, P. and Segvic, S. (2019) In Defense of Pre-Trained Imagenet Architectures for Real-Time Semantic Segmentation of Road-Driving Images. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 12599-12608. <https://doi.org/10.1109/CVPR.2019.01289>
- [3] Mou, L., Hua, Y. and Zhu, X. (2019) A Relation-Augmented Fully Convolutional Network for Semantic Segmentation in Aerial Scenes. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 12408-12417. <https://doi.org/10.1109/CVPR.2019.01270>
- [4] Luengo, J., Moreno, R., Sevillano, I., Charte, D., Pelaez-Vegas, A., Fernández-Moreno, M., Mesejo, P. and Herrera, F. (2022) A Tutorial on the Segmentation of Metallographic Images: Taxonomy, New Metaldam Dataset, Deep Learning-Based Ensemble Model, Experimental Analysis and Challenges. *Information Fusion*, **78**, 232-253. <https://doi.org/10.1016/j.inffus.2021.09.018>
- [5] Katircioglu, I., Rhodin, H., Constantin, V., Sporri, J., Salzmann, M. and Fua, P. (2021) Self-Supervised Human Detection and Segmentation via Background Inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 9574-9588. <https://doi.org/10.1109/TPAMI.2021.3123902>
- [6] Sakaridis, C., Dai, D. and Van Gool, L. (2022) Map-Guided Curriculum Domain Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 3139-3153. <https://doi.org/10.1109/TPAMI.2020.3045882>
- [7] Chapelle, O., Schlkopf, B. and Zien, A. (2006) *Semi-Supervised Learning*. The MIT Press, Cambridge.
- [8] Rawat, W. and Wang, Z. (2017) Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, **29**, 2352-2449. https://doi.org/10.1162/neco_a_00990
- [9] Schroff, F., Criminisi, A. and Zisserman, A. (2008) Object Class Segmentation Using Random Forests. *The British*

- Machine Vision Conference*, 1-10. <https://doi.org/10.5244/C.22.54>
- [10] Felzenszwalb, P.F., Girshick, R.B., McAllester, D.A. and Ramanan, D. (2009) Object Detection with Discriminatively Trained Part Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 1627-1645. <https://doi.org/10.1109/TPAMI.2009.167>
- [11] Redmon, J., Divvala, S.K., Girshick, R.B. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [12] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention*, Vol. 9351, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [13] Chen, L.-C., Papandreou, G., Schroff, F. and Adam, H. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. *ArXiv*, **3**, 1-14.
- [14] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**, 1097-1105
- [15] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*, **6**, 1-14.
- [16] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., et al. (2019) UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, **39**, 1856-1867 <https://doi.org/10.1109/TMI.2019.2959609>
- [17] Milletari, F., Navab, N. and Ahmadi, S.A. (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 *Fourth International Conference on 3D Vision (3DV)*, Stanford, 25-28 October 2016, 565-571. <https://doi.org/10.1109/3DV.2016.79>
- [18] Yu, F. and Koltun, V. (2015) Multi-Scale Context Aggregation by Dilated Convolutions. *ArXiv*, **3**, 1-13.
- [19] Zhao, H., Shi, J., Qi, X., et al. (2017) Pyramid Scene Parsing Network. *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2881-2890. <https://doi.org/10.1109/CVPR.2017.660>
- [20] Chen, L.C., Papandreou, G., Kokkinos, I., et al. (2017) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [21] Chen, L.C., Papandreou, G., Schroff, F., et al. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. *ArXiv*, **3**, 1-14.
- [22] Hung, W.-C., Tsai, Y.-H., Liou, Y.-T., Lin, Y.-Y. and Yang, M.-H. (2018) Adversarial Learning for Semisupervised Semantic Segmentation. *BMVC*. <https://doi.org/10.48550/arXiv.1802.07934>
- [23] Mittal, S., Tatarchenko, M. and Brox, T. (2019) Semi-Supervised Semantic Segmentation with High- and Low-Level Consistency. *TPAMI*. <https://doi.org/10.48550/arXiv.1908.05724>
- [24] Li, D., Yang, J., Kreis, K., Torralba, A. and Fidler, S. (2021) Semantic Segmentation with Generative Models: Semi-supervised Learning and Strong Out-of-Domain Generalization. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 8296-8307. <https://doi.org/10.1109/CVPR46437.2021.00820>
- [25] Souly, N., Spampinato, C. and Shah, M. (2017) Semi Supervised Semantic Segmentation Using Generative Adversarial Network. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 5689-5697. <https://doi.org/10.1109/ICCV.2017.606>
- [26] Chen, Y., Ouyang, X., Zhu, K. and Agam, G. (2021) Complexmix: Semi-Supervised Semantic Segmentation via Mask-Based Data Augmentation. 2021 *IEEE International Conference on Image Processing (ICIP)*, Anchorage, 19-22 September 2021, 2264-2268. <https://doi.org/10.1109/ICIP42928.2021.9506602>
- [27] Grubišić, I., Oršić, M. and Šegvić, S. (2021) A Baseline for Semi-Supervised Learning of Efficient Semantic Segmentation Models. 2021 *17th International Conference on Machine Vision and Applications (MVA)*, Aichi, 25-27 July 2021, 1-5. <https://doi.org/10.23919/MVA51890.2021.9511402>
- [28] Olsson, V., Tranheden, W., Pinto, J. and Svensson, L. (2021) Classmix: Segmentation-Based Data Augmentation for Semi-Supervised Learning. 2021 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2021, 1368-1377. <https://doi.org/10.1109/WACV48630.2021.00141>
- [29] French, G., Laine, S., Aila, T., Mackiewicz, M. and Finlayson, G.D. (2020) Semi-Supervised Semantic Segmentation Needs Strong, Varied Perturbations. *BMVC*. <https://doi.org/10.48550/arXiv.1906.01916>
- [30] Li, X., He, Q., Dai, S., Wu, P. and Tong, W. (2020) Semi-Supervised Semantic Segmentation Constrained by Consistency Regularization. 2020 *IEEE International Conference on Multimedia and Expo (ICME)*, London, 6-10 July 2020, 1-6. <https://doi.org/10.1109/ICME46284.2020.9102851>

- [31] Kim, J., Jang, J. and Park, H. (2020) Structured Consistency Loss for Semi-Supervised Semantic Segmentation. *ArXiv*, **2**, 1-12.
- [32] Ouali, Y., Hudelot, C. and Tami, M. (2020) Semi-Supervised Semantic Segmentation with Cross-Consistency Training. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 12671-12681. <https://doi.org/10.1109/CVPR42600.2020.01269>
- [33] An, S., Zhu, H., Zhang, J., Ye, J., Wang, S., Yin, J. and Zhang, H. (2022) Deep Tri-Training for Semi-Supervised Image Segmentation. *IEEE Robotics and Automation Letters*, **7**, 10097-10104. <https://doi.org/10.1109/LRA.2022.3185768>
- [34] Chen, X., Yuan, Y., Zeng, G. and Wang, J. (2021) Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 2613-2622. <https://doi.org/10.1109/CVPR46437.2021.00264>
- [35] Peng, J., Estrada, G., Pedersoli, M. and Desrosiers, C. (2020) Deep Co-Training for Semi-Supervised Image Segmentation. *Pattern Recognition*, **107**, Article ID: 107269. <https://doi.org/10.1016/j.patcog.2020.107269>
- [36] Wu, Y., Liu, C., Chen, L., Zhao, D., Zheng, Q. and Zhou, H. (2022) Perturbation Consistency and Mutual Information Regularization for Semi-Supervised Semantic Segmentation. *Multimedia Systems*, **29**, 511-523. <https://doi.org/10.1007/s00530-022-00931-9>
- [37] Liu, Y., Tian, Y., Chen, Y., Liu, F., Belagiannis, V. and Carneiro, G. (2022) Perturbed and Strict Mean Teachers for Semi-Supervised Semantic Segmentation. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 4258-4267. <https://doi.org/10.1109/CVPR52688.2022.00422>
- [38] Zhu, X.J. (2008) Semi-Supervised Learning Literature Survey. *Computer Sciences TR*, **1530**, 52.
- [39] Yang, L., Zhuo, W., Qi, L., Shi, Y. and Gao, Y. (2022) ST++: Make Self-Training Work Better for Semi-Supervised Semantic Segmentation. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 4268-4277. <https://doi.org/10.1109/CVPR52688.2022.00423>
- [40] Teh, E.W., Devries, T., Duke, B., Jiang, R., Aarabi, P. and Taylor, G.W. (2022) The Gist and Rist of Iterative Self-Training for Semi-Supervised Segmentation. *2022 19th Conference on Robots and Vision (CRV)*, Toronto, 31 May-2 June 2022, 58-66. <https://doi.org/10.1109/CRV55824.2022.00016>
- [41] Li, H. and Zheng, H. (2021) A Residual Correction Approach for Semi-Supervised Semantic Segmentation. In: Ma, H., *et al.*, Eds., *Pattern Recognition and Computer Vision*, Vol. 13022. Springer, Cham, 90-102. https://doi.org/10.1007/978-3-030-88013-2_8
- [42] Yuan, J., Liu, Y., Shen, C., Wang, Z. and Li, H. (2021) A Simple Baseline for Semi-Supervised Semantic Segmentation with Strong Data Augmentation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 8209-8218. <https://doi.org/10.1109/ICCV48922.2021.00812>
- [43] He, R., Yang, J. and Qi, X. (2021) Re-Distributing Biased Pseudo Labels for Semi-Supervised Semantic Segmentation: A Baseline Investigation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 6930-6940. <https://doi.org/10.1109/ICCV48922.2021.00685>
- [44] Zhu, Y., Zhang, Z., Wu, C., Zhang, Z., He, T., Zhang, H., Manmatha, R., Li, M. and Smola, A.J. (2021) Improving Semantic Segmentation via Efficient Self-Training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**. <https://doi.org/10.1109/TPAMI.2021.3138337>
- [45] Chen, Z., Zhang, R., Zhang, G., Ma, Z. and Lei, T. (2020) Digging into Pseudo Label: A Low-Budget Approach for Semisupervised Semantic Segmentation. *IEEE Access*, **8**, 41830-41837. <https://doi.org/10.1109/ACCESS.2020.2975022>
- [46] Zhang, F.H., Torr, P., Ranftl, R. and Richter, S.R. (2021) Looking beyond Single Images for Contrastive Semantic Segmentation Learning. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., and Vaughan, J.W., Eds., *Advances in Neural Information Processing Systems (NeurIPS)*, Curran Associates, Inc., New York.
- [47] Zhao, X.Y., Vemulapalli, R., Mansfield, P.A., Gong, B.Q., Green, B., Shapira, L. and Wu, Y. (2021) Contrastive Learning for Label Efficient Semantic Segmentation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 10603-10613. <https://doi.org/10.1109/ICCV48922.2021.01045>
- [48] Liu, S., Zhi, S., Johns, E. and Davison, A.J. (2021) Bootstrapping Semantic Segmentation with Regional Contrast. *ArXiv*, **4**, 1-23.
- [49] Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A. and Li, C.-L. (2020) Fixmatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *Advances in Neural Information Processing Systems*, **33**, 596-608.
- [50] Lai, X., Tian, Z., Jiang, L., Liu, S., Zhao, H., Wang, L. and Jia, J. (2021) Semi-Supervised Semantic Segmentation with Directional Context-Aware Consistency. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 1205-1214. <https://doi.org/10.1109/CVPR46437.2021.00126>