# 说明文档

## 项目概述

我们提出了一种通过对抗学习实现半监督语义分割的算法。生成对抗网络（GANs）为无监督和半监督学习提供了许多可能性。典型的 GAN 由两个子网络组成，即生成器和判别器，这两个子网络在训练过程中进行一个最小-最大博弈。生成器接收一个样本向量，并输出目标数据分布的样本，例如人脸，而判别器旨在将生成的样本与目标样本区分开来。然后，生成器通过反向传播来迷惑判别器，从而生成类似于目标分布的样本。我们应用了类似的方法，将分割网络视为 GAN 框架中的生成器。

与典型生成器在给定噪声向量时生成图像不同，我们的分割网络在给定输入图像时输出语义标签的概率图。在这种设置下，我们希望将分割网络的输出在空间上尽可能接近真实标签图。为此，我们采用了对抗学习方案，并提出了一种完全卷积的判别器，用于区分真实标签图和分割预测的概率图。

结合空间交叉熵损失，我们的方法使用对抗损失来鼓励分割网络生成的预测概率图在高阶结构上接近真实标签图。这个想法类似于概率图模型的使用，如条件随机场（CRFs），但在测试阶段不需要额外的后处理模块。此外，推理时不需要判别器，因此我们提出的框架在测试时不会增加任何计算负担。通过采用对抗学习，我们进一步在半监督设置下利用了所提出的完全卷积判别器。

让判别器利用未标记数据的一种方法是仅通过对抗学习训练分割网络，而不使用交叉熵损失。然而，根据我们的实验，这种方法无法提高性能，因为对抗损失会积极地鼓励预测结果接近真实分布，却忽略了分割的正确性。相反，我们利用由判别器网络生成的置信图作为监督信号，以“自学习”的方式指导交叉熵损失。置信图指示了预测分布中哪些区域接近真实标签分布，以便分割网络可以信任这些预测，从而可以通过掩码交叉熵损失进行训练。通过采用提出的框架，我们展示了通过在标注图像领域中添加没有任何标注的图像，可以进一步提高分割准确性。

总体而言，我们开发了一个对抗性框架，无需在推理过程中增加额外的计算负担，就能提高语义分割的准确性。我们通过利用未标记图像的判别器网络响应来辅助分割网络的训练，从而促进了半监督学习。

## 算法概述

此项目应用的半监督语义分割算法由两个主要部分组成：分割网络和判别器网络。分割网络可以是任何现有的语义分割网络，例如全卷积网络（FCN），DeepLab或DilatedNet。该网络接收尺寸为H × W × 3的输入图像，并输出尺寸为H × W × C的类别概率图，其中C表示目标数据集中语义类别的数量。分割网络通过三种损失函数进行优化：交叉熵损失（）、对抗损失（）和半监督损失（）。

判别器网络基于FCN架构，处理由分割网络生成的类别概率图或真实标签图。它输出尺寸为H × W × 1的空间概率图，其中每个像素表示该像素是否来自真实标签（p = 1）或分割网络（p = 0）。与典型的GAN判别器处理固定大小的输入图像并输出单一概率值不同，此网络判别器是一个全卷积网络，能够处理任意大小的输入。

在训练过程中，使用标注和未标注的图像，以半监督的方式进行。对于标注数据，分割网络通过与真实标签图的标准交叉熵损失和来自判别器网络的对抗损失进行训练。判别器网络仅使用标注数据进行训练。

对于未标注数据，分割网络采用半监督方法。首先，分割网络生成未标注图像的初始分割预测。然后，将该预测通过判别器网络以获得置信度图。置信度图作为监督信号，通过“自学”方案，用掩码交叉熵损失训练分割网络。置信度图指示预测分割的局部质量，从而指导分割网络在训练过程中信任哪些区域。判别器网络通过损失（）进行训练。图1显示了所提出算法的概述。

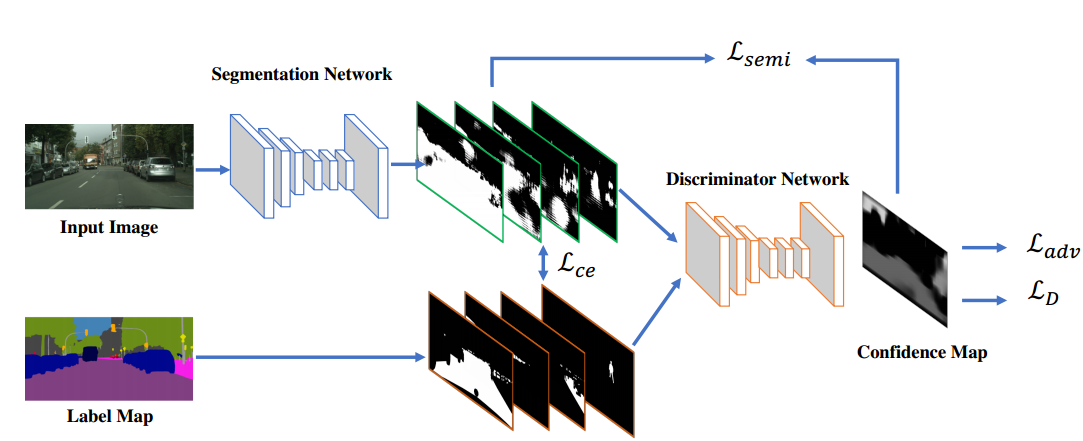


图 1 提出的半监督语义分割系统概述。在使用损失训练的全卷积判别器网络中，我们在训练过程中使用三个损失函数来优化分割网络:交叉熵损失，对抗损失和半监督损失。

## 半监督训练与对抗网络

#### 3.1 训练方案

给定一个大小为 H × W × 3 的输入图像 ，将分割网络表示为 S(·)，预测的概率图表示为 ，大小为 H × W × C，其中 C 是类别数。对于全卷积鉴别器，我们将其表示为 D(·)，它接受大小为 H × W × C 的概率图并输出大小为 H × W × 1 的置信图。在我们的方法中，鉴别器网络有两种可能的输入：分割预测 或独热编码的真实标签 。

**鉴别器网络训练。**为了训练鉴别器网络，我们最小化关于两类的空间交叉熵损失 LD。损失可以形式化地写为：

其中，若样本来自分割网络，则 = 0，若样本来自真实标签，则 = 1。注意，鉴别器网络接受一个 *C* 通道的概率图作为输入。为了将具有离散标签的真实标签图转换为 *C* 通道概率图，我们对真实标签图采用独热编码方案，其中当像素属于类别 *c* 时，，否则为0。

**分割网络训练。**我们提出通过最小化多任务损失函数来训练分割网络：

其中，、和分别表示空间多类别交叉熵损失、对抗损失和半监督损失。和是用于平衡多任务训练的两个常数。

首先考虑使用标注数据的场景。给定输入图像、独热编码的真实标签 和预测结果，交叉熵损失通过以下公式获得：

采用全卷积鉴别器网络 D(·) 通过对抗损失 进行对抗学习：

通过这个对抗损失，我们试图通过最大化分割预测被认为是真实标签分布的概率来训练分割网络欺骗鉴别器。

**使用未标记数据进行训练。**现在我们考虑在半监督设置下的对抗训练。对于未标记的数据，显然我们不能应用，因为没有可用的真实标注。我们提出使用“自我教导”策略利用未标记数据训练鉴别器。主要思想是，训练过的鉴别器可以生成一个置信图，即 ，这推断出预测结果接近真实分布的区域。然后，我们使用阈值二值化这个置信图，以突出可信赖的区域。因此，我们定义自我教导的真实标签为使用这个二值化置信图的掩蔽分割预测。由此产生的半监督损失定义为：

其中，I(·) 是指示函数， 是控制自我教导过程敏感性的阈值。注意，在训练过程中，我们将自我教导的目标 和指示函数的值都视为常数，因此 (5) 可以简单地视为一个掩蔽的空间交叉熵损失。这种策略在 在 0.1 到 0.3 之间变化时非常稳健。

#### 3.2 网络架构

**分割网络。**我们采用在 ImageNet 数据集上预训练的 ResNet-101模型的 DeepLab-v2框架作为我们的分割基线网络。遵循最近关于语义分割的研究实践，移除了最后的分类层，并将最后两个卷积层的步长从 2 改为 1，使得输出特征图的分辨率有效地是输入图像大小的 1/8。为了扩大感受野，在 conv4 和 conv5 层使用了扩张卷积，步长分别为 2 和 4。在最后一层之后，我们采用Atrous空间金字塔池化（ASPP）作为最终分类器。最后，我们应用一个上采样层以及 softmax 输出以匹配输入图像的大小。

**鉴别器网络。**对于鉴别器网络，它由 5 个卷积层组成，核大小为 4 × 4，通道数为 {64, 128, 256, 512, 1}，步长为 2。每个卷积层后面跟一个由 0.2 参数化的 Leaky-ReLU，除了最后一层。为了将网络转换为全卷积网络，我们在最后一层添加了一个上采样层，以将输出调整到输入图的大小。请注意，我们没有使用批量归一化层。我们发现批量归一化层非常不稳定，因为系统只能用小批量大小进行训练。

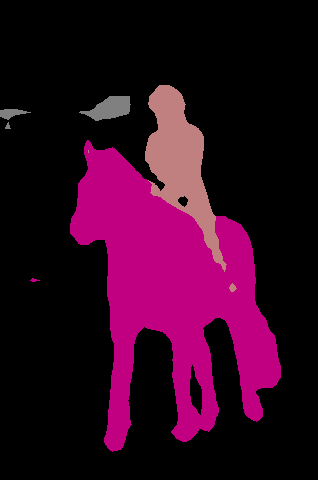
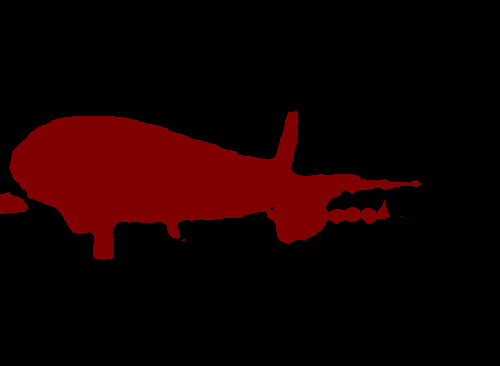
#### 3.3 实现细节

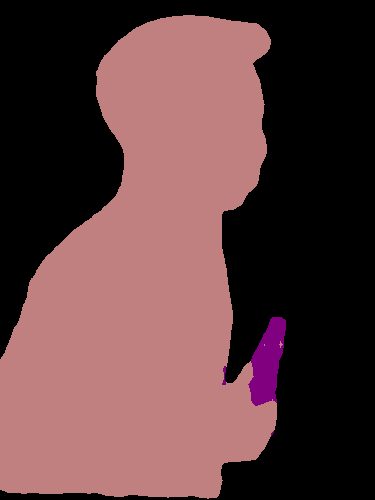
我们使用 PyTorch 框架实现我们的网络。在GPU 上训练我们的系统。为了训练分割网络，我们使用带有 Nesterov 加速的随机梯度下降（SGD）作为优化器，其中动量为 0.9，权重衰减为 10^-4。初始学习率设置为 2.5 × 10^-4，并使用幂次衰减的方式减少，幂次为 0.9。对于训练鉴别器，我们采用 Adam 优化器，学习率为 10^-4，并且使用与分割网络相同的多项式衰减。动量设置为 0.9 和 0.999。

对于半监督训练，我们随机交替迭代有标签数据和无标签数据，并相应地应用第 3.1 节描述的训练方案。请注意，为了防止模型受到初始噪声掩码和预测的影响，我们在使用有标签数据训练了 5000 次迭代后开始半监督学习。我们联合更新分割网络和鉴别器网络。在每次迭代中，只有包含真实数据的批次用于训练鉴别器。

## 方案总结

在这项工作中，我们采用了一种用于半监督语义分割的对抗学习方案。我们训练一个全卷积鉴别器网络，以利用有标签和无标签数据增强分割网络。对于有标签数据，分割网络的对抗性损失旨在学习更高阶的结构信息，而无需后处理。对于无标签数据，鉴别器网络生成的置信度图作为自我教导信号，用于完善分割网络。在 PASCAL VOC 2012 数据集和 Cityscapes 数据集上进行了实验，以验证所提算法的有效性。图2显示了所提方法生成的分割结果的视觉效果。





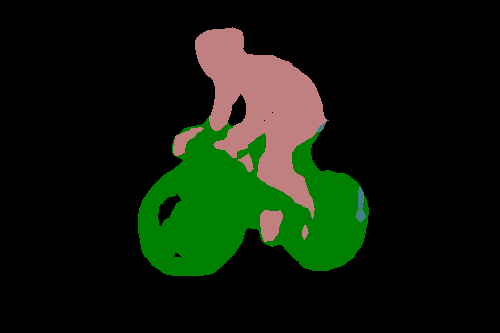


图 2 通过半监督语义分割的对抗学习生成的分割效果图