



(12)发明专利申请

(10)申请公布号 CN 110111335 A

(43)申请公布日 2019.08.09

(21)申请号 201910379298.4

(22)申请日 2019.05.08

(71)申请人 南昌航空大学

地址 330000 江西省南昌市丰和南大道696号

(72)发明人 张桂梅 潘国峰 徐可

(74)专利代理机构 北京高沃律师事务所 11569

代理人 程华

(51)Int.Cl.

G06T 7/10(2017.01)

G06N 3/04(2006.01)

G06N 3/08(2006.01)

权利要求书3页 说明书10页 附图2页

(54)发明名称

一种自适应对抗学习的城市交通场景语义分割方法及系统

(57)摘要

本发明公开一种自适应对抗学习的城市交通场景语义分割方法及系统。获取语义分割模型的训练数据,对游戏合成数据集GTA5进行预处理,得到一个新的与城市场景真实数据集分布接近的合成数据集SG-GTA5;构建用于语义分割的生成对抗网络模型;基于训练数据集,对生成对抗网络模型进行自适应对抗学习,在不同特征层的对抗学习中使用自适应的学习率,通过该学习率调整各特征层的损失值,进而动态更新网络参数,得到优化的生成对抗网络模型;在城市场景真实数据集CityScapes上进行验证。本发明能够提高缺乏标注信息,存在较多尺度目标的复杂城市交通场景的语义分割精度,增强语义分割模型的泛化能力。



1. 一种自适应对抗学习的城市交通场景语义分割方法,其特征在于,所述方法包括:

获取语义分割模型的训练数据集,对原始的游戏合成数据集GTA5进行预处理,得到新的与城市场景真实数据集CityScapes分布接近的合成数据集SG-GTA5,作为源数据集, CityScapes训练数据集中不使用标注信息的图片作为目标数据集,源数据集和目标数据集组成语义分割模型的训练数据集;

构建弱监督的语义分割模型,即生成对抗网络模型,所述生成对抗网络包括生成器网络和判别器网络;

将所述新的源数据集SG-GTA5和目标数据集作为训练数据,对所述生成对抗网络进行自适应学习率的对抗训练,得到优化的生成器网络模型;

将城市场景数据集Cityscapes中的测试图像输入至所述优化的生成器网络模型中,得到分割结果;

对所述分割结果进行定量和定性评估,得到分割性能结果。

2. 根据权利要求1所述的自适应对抗学习的城市交通场景语义分割方法,其特征在于,所述获取语义分割模型的训练数据集,具体包括:

以CityScapes训练数据集中不使用标注信息的图片和带标注的游戏合成数据集GTA5作为输入,采用语义感知生成对抗网络算法对两个输入数据集进行对抗训练,得到一个训练完成的网络模型,将原始合成数据集GTA5输入到该训练完成的网络模型中,得到新的合成数据集SG-GTA5,并将其作为语义分割模型的源数据集, CityScapes训练数据集中不使用标注信息的图片作为目标数据集,源数据集和目标数据集组成了语义分割的训练数据集。

3. 根据权利要求1所述的自适应对抗学习的城市交通场景语义分割方法,其特征在于,所述构建弱监督的语义分割模型,具体包括:

构建生成器网络:

采用DeepLab v2作为网络基本框架;在最后一层采用多孔空间金字塔池化结构代替标准卷积作为最终的分类器;再将第四层和第五层卷积层的步长由2改成1;同时将第四层和第五层空洞卷积层的空洞数分别设置成2和4;最后,使用一个上采样层得到与输入图像大小相同的特征图;

构建判别器网络:

采用与传统卷积神经网络CNN类似的网络结构,使用全卷积层代替全连接层,网络由6个卷积层组成,前四个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为2,第五个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为1,它们的通道数分别为(64, 128, 256, 512, 1024, 1),除最后一层外,每个卷积层后面均连接一个非线性激活函数Leaky ReLU。

4. 根据权利要求1所述的自适应对抗学习的城市交通场景语义分割方法,其特征在于,所述将所述新的源数据集SG-GTA5和目标数据集作为训练数据,对所述生成对抗网络进行自适应学习率的对抗训练,得到优化的生成器网络模型,具体包括:

将带标注的源数据集中的图像和未带标注的目标数据集中的图像作为所述生成器网络的输入,通过所述生成器网络分别得到其对应的特征图;

将所述得到的特征图作为所述判别器的输入,采用自适应学习率调整各层特征的对抗性学习方法,通过反向传播的方式,对判别器网络和生成器网络中的参数进行不断的更新优化;

更新优化过程具体如下：

固定生成器网络参数，对所述判别器网络采用自适应的学习率进行训练，得到训练后的判别器网络模型；

固定所述训练后的判别器网络模型的参数，对所述生成器网络进行反向传播训练，得到优化的生成器网络模型，如此循环迭代，最后得到最优的生成对抗网络模型。

5. 一种自适应对抗学习的城市交通场景语义分割系统，其特征在于，所述系统包括：

训练数据集获取模块，用于获取语义分割模型的训练数据集，对原始的游戏合成数据集GTA5进行预处理，得到新的与城市场景真实数据集分布接近的合成数据集SG-GTA5，作为源数据集，CityScapes训练数据集中不使用标注信息的图片作为目标数据集，源数据集和目标数据集组成语义分割模型的训练数据集；

生成对抗网络构建模块，用于构建弱监督的语义分割模型，即生成对抗网络结构，所述生成对抗网络包括生成器网络和判别器网络；

生成对抗网络训练模块，用于将所述新的源数据集SG-GTA5和目标数据集作为训练数据，对所述生成对抗网络进行自适应学习率的对抗训练，得到优化的生成器网络模型；

分割结果确定模块，用于将城市场景数据集Cityscapes中的测试图像输入至所述优化的生成器网络模型中，得到分割结果；

评估模块，用于对所述分割结果进行定量和定性评估，得到分割性能结果。

6. 根据权利要求5所述的自适应对抗学习的城市交通场景语义分割系统，其特征在于，所述训练数据集获取模块，具体包括：

训练数据集获取单元，用于以未带标注信息的目标数据集和带标注的游戏合成数据集GTA5作为输入，采用语义感知生成对抗网络算法对两个输入数据集进行对抗训练，得到一个训练完成的网络模型，将原始合成数据集GTA5输入到该训练完成的网络模型中，得到新的合成数据集SG-GTA5，并将其作为语义分割模型的源数据集，CityScapes训练数据集中不带标注信息的图片作为目标数据集，源数据集和目标数据集组成了语义分割的训练数据集。

7. 根据权利要求5所述的自适应对抗学习的城市交通场景语义分割系统，其特征在于，所述生成对抗网络构建模块，具体包括：

生成器网络构建单元，用于构建生成器网络，采用DeepLab v2作为网络基本框架；在最后一层采用多孔空间金字塔池化结构代替标准卷积作为最终的分类器；再将第四层和第五层卷积层的步长由2改成1；同时将第四层和第五层空洞卷积层的空洞数分别设置成2和4；最后，使用一个上采样层得到与输入图像大小相同的特征图；

判别器网络构建单元，用于构建判别器网络，采用与传统卷积神经网络CNN类似的网络结构，使用全卷积层代替全连接层，网络由6个卷积层组成，前四个卷积层的卷积核大小设为 $4 \times 4$ ，步长设为2，第五个卷积层的卷积核大小设为 $4 \times 4$ ，步长设为1，它们的通道数分别为(64, 128, 256, 512, 1024, 1)，除最后一层外，每个卷积层后面均连接一个非线性激活函数Leaky ReLU。

8. 根据权利要求5所述的自适应对抗学习的城市交通场景语义分割系统，其特征在于，所述生成对抗网络训练模块，具体包括：

特征图提取单元，用于将带标注的源数据集中的图像和未带标注的目标数据集中的图

像作为所述生成器网络的输入,通过所述生成器网络分别得到其对应的特征图;

参数优化单元,用于将所述得到的特征图作为所述判别器的输入,采用自适应学习率调整各层特征的对抗性学习方法,通过反向传播的方式,对判别器网络和生成器网络中的参数进行不断的更新优化;

更新优化过程具体如下:

固定生成器网络参数,对所述判别器网络采用自适应的学习率进行训练,得到训练后的判别器网络模型;

固定所述训练后的判别器网络模型的参数,对所述生成器网络进行反向传播训练,得到优化的生成器网络模型,如此循环迭代,最后得到最优的生成对抗网络模型。

## 一种自适应对抗学习的城市交通场景语义分割方法及系统

### 技术领域

[0001] 本发明涉及基于弱/半监督的图像语义分割领域,特别是涉及一种自适应对抗学习的城市交通场景语义分割方法及系统。

### 背景技术

[0002] 语义分割是指将图像分割成几组具有特征语义的像素区域,并识别每个区域的类别,最终获得具有像素语义标注的图像。通过图像语义分割,使复杂的图像更容易理解和分析。语义分割是自动驾驶、视觉导航、图像检索、目标识别和检测跟踪等视觉分析与理解的基础。

[0003] 深度学习由于能够学习到高层语义特征,对复杂场景表现出强大的拟合能力,故深度学习在计算机视觉研究领域取得了突破性进展。在基于深度学习的网络框架中,具有代表性的卷积神经网络包括AlexNet、VGGNet、GoogleNet和ResNet,这些网络中的卷积层、池化层和全连接层是共有的相关网络层。深度学习应用在语义分割中的应用大部分都是基于上面的网络作为基本结构进行改进的。根据机器学习的方式将语义分割方法分为基于全监督和半(弱)监督的方法。

[0004] 针对图像语义分割算法研究,Long等提出了一种基于全卷积网络(Fully convolutional Neural,FCN)的语义分割方法,将图像分类网络中最后的全连接层用卷积层代替,从而保留了空间信息,然而该方法存在暴力池化问题,上下文相关性信息较差,因此分割结果不够精细。为了得到相关性更强的分割结果,Badrinarayanan等提出了SegNet分割网络用于道路、车辆的分割,SegNet中的池化层能记录池化层的值在原特征图中的空间位置,但是SegNet的分割精度仍有待提高。Noh等提出了DeConvNet分割网络,通过利用反卷积和上采样方法实现了图像的语义分割任务,在细小物体的语义分割问题上取得了较好的效果。在语义分割任务中,利用有效的上下文信息成为提升分割精度的关键。条件随机场能有效地学习到像素之间的相关性,它的典型应用代表是DeepLab系列,在DeepLab v1中引入了空洞卷积,根据分割物体的尺度来修改空洞的大小,从而调整感受野的大小,避免了上采样带来的信息丢失,提高了分割精度。在DeepLab v2中引用了不同特征层的分辨率信息,采用了多孔空间金字塔池化结构(ASPP),更有效地学习了不同尺度目标的信息。DeepLab v3采用级联或并行多个不同空洞卷积模块以捕获多尺度上下文信息,从而进一步提高了分割精度。但是基于全监督式分割方法需要大量标注样本,对于不同的场景或者同一场景在不同的光照和气候条件下需要重新标注,从而耗费大量的财力和物力。

[0005] 何凯明等提出了Boxsup方法,进一步释放了深层卷积网络的力量,得到了62.0%的分割精度(当时的全监督方法的精度为63.8%)。Noh H等提出了一种基于异构标注的半监督语义分割的深层网络模型,该方法能够基于图像级进行分类,同时也能基于像素级分类标签的训练数据进行分割。与其他半监督方法相比,该算法在Pascal Voc数据集测试中用到较强标注的训练图像较少,并且分割性能得到较好提升。A Khoreva等提出适合于语义标记和实例分割任务的弱监督分割方法,实验结果表明当给定精细的边界框输入标签时,

经过一轮训练就可以获得比以前较好的弱监督分割方法更好的分割结果。Papandreou G等提出了图像语义分割模型训练的期望最大化 (EM) 方法,该方法从弱标注的训练数据(如边界框或图像级标签),或从少数强标记和多个弱标记图像的组合中学习得到用于图像语义分割的深度卷积神经网络,取得了较好的分割效果,同时较大减少了所需的标注工作。以上方法都是基于现有数据集的弱监督方法,然而在大多数实际应用中,数据集的获取需要特定的环境,甚至很难获得弱标注信息,并且经过训练的模型并不一定能较好地泛化到没有真实标注训练过的目标图像域。此外,高质量的像素级标注通常很难以获得,需要耗费大量的人力和财力。因此寻求计算机自动生成合成数据集的方法成了一种较好的解决方案,并且合成数据集像素级的标注可以通过一个自动化的过程获得,所以数据收集成本低,将这些数据集与真实数据集同时使用,联合训练得到性能较高的模型。然而,在合成数据集上直接进行训练时,该模型并不能很好地泛化到真实数据集,主要原因是合成图像和真实世界图像之间的分布存在差异,所以用真实数据集训练出的预训练模型来直接训练虚拟数据集时,分割精度不够理想。

[0006] 不同数据集通过深度网络之后的输出往往存在区域偏移,Y Ganin等提出了一种新的域自适应学习方法,其中训练和测试时的数据均来自不同分布的同一场景。该方法在标准数据集上实现了最先进的域自适应性能。最近的研究表明,深层网络能够更好地学习到可转移的特征,并能较好地推广到新的区域自适应任务。然而,深度特征最终将沿网络由全局的特征向局部特征过渡,随着域间差异的增大,特征可转移性在较高层次上明显下降。针对该问题,M Long等提出了一种新的深度自适应网络,将深度卷积神经网络推广到域自适应场景中。尽管最近基于深度学习的语义分割取得了突破性的进展,但是将一个经过预处理的场景分割器部署到一个没有预训练的其它同一场景数据集中,由于数据集的偏差,无法获得令人满意的效果。Yi-Hsin Chen等提出一种弱监督学习方法,以适应不同城市的道路场景分割。该方法不需要采集大量感兴趣城市的标注图像,而是对分割器进行训练或微调,可以在不需要任何用户注释或交互的情况下,实现使用预训练的分割器对该城市自适应学习和分割。Yi-Hsuan Tsai等提出了一种基于语义分割的域自适应对抗性学习方法,即AdaptSegNet,该方法对合成数据集和真实数据集的输出特征采用了对抗性学习,并构建了一个多层次的对抗性网络,有效地实现了不同特征层的输出空间域自适应,通过在不同的数据集上进行测试,实验结果表明,所提出的方法在精度和视觉质量上均优于当时所有的方法。但是由于AdaptSegNet中的源域和目标域之间边缘与纹理分布存在差异和类漂移,此外该模型在不同特征层的对抗学习中加入了固定惩罚因子,从而导致分割结果并不理想。

[0007] 综上所述,国内外不少学者专家对基于深度学习的图像语义分割方法有了较多研究,但仍然在以下方面存在着问题,如:

[0008] (1) 基于全监督的图像语义分割方法需要大量的图像和对应的标注样本,并且对于不同的场景或者同一场景在不同的光照和气候条件下需要重新标注,耗时费力;而非监督式的图像语义分割则不需要额外进行标注,但是分割精度往往很低;

[0009] (2) 城市交通场景由于场景较复杂,并且易受光照和气候条件的影响,所以该类场景的语义分割通常存在类漂移和类感染,从而影响模型的分割的精度和模型的迁移性;

[0010] (3) 目前流行的基于域自适应对抗生成器网络的语义分割方法通常将源域(合成

数据集)和目标域(真实数据集)直接进行对抗训练,然而,直接在合成数据集进行训练时,该模型不能较好地泛化到真实数据集,主要原因是源域和目标域图像的颜色、纹理等特征分布相差较大,训练过程中的对抗损失值较大,模型在反向传播过程中容易出现梯度爆炸,从而影响分割精度。

[0011] (4)目前流行的基于域自适应对抗生成器网络的语义分割通常采用多个特征层融合的方法,但是在不同特征层的交叉熵损失值和对抗损失值加入固定了惩罚系数,并且通过多次实验手动调整得到,对复杂交通场景语义分割鲁棒性较差。实际上,随着网络训练次数的增加,模型各个输出特征层对分割结果的影响不同。

## 发明内容

[0012] 本发明的目的是提供一种自适应对抗学习的城市交通场景语义分割方法及系统,提高缺乏标注信息,存在多尺度目标的复杂城市交通场景的语义分割精度,增强语义分割模型的泛化能力。

[0013] 为实现上述目的,本发明提供了如下方案:

[0014] 一种自适应对抗学习的城市交通场景语义分割方法,包括:

[0015] 获取语义分割模型的训练数据集,对原始的游戏合成数据集GTA5进行预处理,得到新的与城市场景真实数据集分布接近的合成数据集SG-GTA5,作为源数据集,CityScapes训练数据集中不使用标注信息的图片作为目标数据集,源数据集和目标数据集组成语义分割模型的训练数据集;

[0016] 构建弱监督的语义分割模型,即生成对抗网络模型,所述生成对抗网络包括生成器网络和判别器网络;

[0017] 将所述新的源数据集SG-GTA5和目标数据集作为训练数据,对所述生成对抗网络进行自适应学习率的对抗训练,得到优化的生成器网络模型;

[0018] 将城市场景数据集Cityscapes中的测试图像输入至所述优化的生成器网络模型中,得到分割结果;

[0019] 对所述分割结果进行定量和定性评估,得到分割性能结果。

[0020] 可选的,所述获取语义分割模型的训练数据集,具体包括:

[0021] 以CityScapes训练数据集中不使用标注信息的图片和带标注信息的游戏合成数据集GTA5作为输入,采用语义感知生成对抗网络算法对两个输入数据集进行对抗训练,得到一个训练完成的网络模型,将原始合成数据集GTA5输入到该训练完成的网络模型中,得到新的合成数据集SG-GTA5,并将其作为语义分割模型的源数据集,CityScapes训练数据集中不使用标注信息的图片作为目标数据集,源数据集和目标数据集组成了语义分割的训练数据集。

[0022] 可选的,所述构建弱监督的语义分割模型,具体包括:

[0023] 构建生成器网络:

[0024] 采用DeepLab v2作为网络基本框架;在最后一层采用多孔空间金字塔池化结构代替标准卷积作为最终的分类器;再将第四层和第五层卷积层的步长由2改成1;同时将第四层和第五层空洞卷积层的空洞数分别设置成2和4;最后,使用一个上采样层得到与输入图像大小相同的特征图;

[0025] 构建判别器网络：

[0026] 采用与传统卷积神经网络CNN类似的网络结构，使用全卷积层代替全连接层，网络由6个卷积层组成，前四个卷积层的卷积核大小设为 $4 \times 4$ ，步长设为2，第五个卷积层的卷积核大小设为 $4 \times 4$ ，步长设为1，它们的通道数分别为(64, 128, 256, 512, 1024, 1)，除最后一层外，每个卷积层后面均连接一个非线性激活函数LeakyReLU。

[0027] 可选的，所述将所述新的源数据集SG-GTA5和目标数据集作为训练数据，对所述生成对抗网络进行自适应学习率的对抗训练，得到优化的生成器网络模型，具体包括：

[0028] 将带标注的源数据集中的图像和未带标注的目标数据集中的图像作为所述生成器网络的输入，通过所述生成器网络分别得到其对应的特征图；

[0029] 将所述得到的特征图作为所述判别器的输入，采用自适应学习率调整各层特征的对抗性学习方法，通过反向传播的方式，对判别器网络和生成器网络中的参数进行不断的更新优化；

[0030] 更新优化过程具体如下：

[0031] 固定生成器网络参数，对所述判别器网络采用自适应的学习率进行训练，得到训练后的判别器网络模型；

[0032] 固定所述训练后的判别器网络模型的参数，对所述生成器网络进行反向传播训练，得到优化的生成器网络模型，如此循环迭代，最后得到最优的生成对抗网络模型。

[0033] 一种自适应对抗学习的城市交通场景语义分割系统，包括：

[0034] 训练数据集获取模块，用于获取语义分割模型的训练数据集，对原始的游戏合成数据集GTA5进行预处理，得到新的与城市场景真实数据集分布接近的合成数据集SG-TGA5，作为源数据集，CityScapes训练数据集中不使用标注信息的图片作为目标数据集，源数据集和目标数据集组成语义分割模型的训练数据集；

[0035] 生成对抗网络构建模块，用于构建弱监督的语义分割模型，即生成对抗网络模型，所述生成对抗网络包括生成器网络和判别器网络；

[0036] 生成对抗网络训练模块，用于将所述新的源数据集SG-GTA5和目标数据集作为训练数据，对所述生成对抗网络进行自适应学习率的对抗训练，得到优化的生成器网络模型；

[0037] 分割结果确定模块，用于将城市场景数据集Cityscapes中的测试图像输入至所述优化的生成器网络模型中，得到分割结果；

[0038] 评估模块，用于对所述分割结果进行定量和定性评估，得到分割性能结果。

[0039] 可选的，所述训练数据集获取模块，具体包括：

[0040] 训练数据集获取单元，用于以CityScapes训练数据集中不使用标注信息的图片和带标注的游戏合成数据集GTA5作为输入，采用语义感知生成对抗网络算法对两个输入数据集进行对抗训练，得到一个训练完成的网络模型，将原始合成数据集GTA5输入到该训练完成的网络模型中，得到新的合成数据集SG-GTA5，并将其作为语义分割模型的源数据集，CityScapes训练数据集中不使用标注信息的图片作为目标数据集，源数据集和目标数据集组成了语义分割的训练数据集。

[0041] 可选的，所述生成对抗网络构建模块，具体包括：

[0042] 生成器网络构建单元，用于构建生成器网络，采用DeepLab v2作为网络基本框架；在最后一层采用多孔空间金字塔池化结构代替标准卷积作为最终的分类器；再将第四层和

第五层卷积层的步长由2改成1;同时将第四层和第五层空洞卷积层的空洞数分别设置成2和4;最后,使用一个上采样层得到与输入图像大小相同的特征图;

[0043] 判别器网络构建单元,用于构建判别器网络,采用与传统卷积神经网络CNN类似的网络结构,使用全卷积层代替全连接层,网络由6个卷积层组成,前四个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为2,第五个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为1,它们的通道数分别为(64,128,256,512,1024,1),除最后一层外,每个卷积层后面均连接一个非线性激活函数Leaky ReLU。

[0044] 可选的,所述生成对抗网络训练模块,具体包括:

[0045] 特征图提取单元,用于将带标注的源数据集中的图像和未带标注的目标数据集中的图像作为所述生成器网络的输入,通过所述生成器网络分别得到其对应的特征图;

[0046] 参数优化单元,用于将所述得到的特征图作为所述判别器的输入,采用自适应学习率调整各层特征的对抗性学习方法,通过反向传播的方式,对判别器网络和生成器网络中的参数进行不断的更新优化;

[0047] 更新优化过程具体如下:

[0048] 固定生成器网络参数,对所述判别器网络采用自适应的学习率进行训练,得到训练后的判别器网络模型;

[0049] 固定所述训练后的判别器网络模型的参数,对所述生成器网络进行反向传播训练,得到优化的生成器网络模型,如此循环迭代,最后得到最优的生成对抗网络模型。

[0050] 根据本发明提供的具体实施例,本发明公开了以下技术效果:本发明提供一种自适应对抗学习的城市交通场景语义分割方法,方法的特点为利用带标注的合成数据集训练网络模型,以实现无标注的真实城市交通场景的语义分割。方法包括:首先对游戏合成数据集GTA5进行预处理,得到一个新的与真实城市场景数据集Cityscapes分布接近的合成数据集SG-GTA5;然后构建用于语义分割的生成器网络模型和判别器网络模型,新建立的网络模型在不同特征层的对抗学习中使用自适应的学习率,通过该学习率调整各特征层的损失值,进而动态更新网络参数,得到优化的生成器网络模型;最后在城市场景真实数据集CityScapes上进行验证。本发明能够提高缺乏标注信息,存在较多尺度目标的复杂城市交通场景的语义分割精度,增强语义分割模型的泛化能力。

## 附图说明

[0051] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0052] 图1为本发明自适应对抗学习的城市交通场景语义分割方法流程图;

[0053] 图2为本发明自适应对抗学习的城市交通场景语义分割系统结构图;

[0054] 图3为本发明基于生成对抗网络的交通场景语义分割网络模型图。

## 具体实施方式

[0055] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完

整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0056] 本发明的目的是提供一种自适应对抗学习的城市交通场景语义分割方法及系统,提高缺乏标注信息,存在多尺度目标的复杂城市交通场景的语义分割精度,增强语义分割模型的泛化能力。

[0057] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0058] 训练数据集的获取、生成对抗网络的构建(包括生成器网络和判别器网络的构建)、生成对抗网络模型的训练、分割测试和验证。

[0059] 图1为本发明自适应对抗学习的城市交通场景语义分割方法流程图。如图1所示,一种自适应对抗学习的城市交通场景语义分割方法包括:

[0060] 步骤101:获取语义分割模型的训练数据集,对原始的游戏合成数据集GTA5进行预处理,得到新的与城市场景真实数据集分布接近的合成数据集SG-GTA5,作为源数据集, CityScapes训练数据集中不使用标注信息的图片作为目标数据集,源数据集和目标数据集组成语义分割模型的训练数据集;

[0061] 以CityScapes训练数据集中不使用标注信息的图片和带标注信息的游戏合成数据集GTA5作为输入,采用语义感知生成对抗网络(Semantic-aware Grad-GAN, SG-GAN)算法对两个输入数据集进行对抗训练,得到一个训练完成的网络模型,将原始合成数据集GTA5输入到训练完成的网络模型中,得到新的合成数据集SG-GTA5,并将其作为源数据集, CityScapes中的训练数据集作为目标数据集。

[0062] 具体实施步骤为:

[0063] 对合成数据集GTA5中的24966幅视频帧图像进行稀疏采样,每隔5帧抽取一张图片,不足5帧的从开始位置补齐,最后得到5000张图片,将其作为语义分割模型的源域;

[0064] 然后将CityScapes训练数据集中不使用标注信息的图片作为目标域,通过SG-GAN训练出一个从GTA5到CityScapes的转换模型;

[0065] 再将GTA5中的24966张图片输入到该转换模型中,最后输出的结果即为新的数据集SG-GTA5,并将其作为语义分割模型的源数据集。

[0066] 预处理后的合成数据集SG-GTA5在颜色、纹理上更接近真实数据集CityScapes,有效降低了对抗损失值,从而避免模型在反向传播过程中出现梯度爆炸。

[0067] 模型数据集包括新得到的合成数据集SG-GTA5和真实城市交通场景数据集CityScapes。其中SG-GTA5中的图片及其标注作为源数据集, CityScapes训练集中不使用标注信息的2975张图片作为目标数据集, CityScapes验证集中的500张带标注信息的图片作为模型的测试评估数据集。

[0068] 步骤102:构建弱监督的语义分割模型,即生成对抗网络结构,所述生成对抗网络包括生成器网络和判别器网络,具体包括:

[0069] 构建生成器网络:

[0070] 采用DeepLab v2作为生成器网络基本框架;在最后一层采用多孔空间金字塔池化结构(Atrous Spatial Pyramid Pooling, ASPP)代替标准卷积作为最终的分类器;再将第

四层和第五层卷积层的步长由2改成1;同时将第四层和第五层空洞卷积层的空洞数分别设置成2和4,扩大了模型的感受野;最后,使用一个上采样层得到与输入图像大小相同的特征图,不再使用批量归一化层;

[0071] 构建判断器网络:

[0072] 采用与传统卷积神经网络(Convolutional Neural Networks,CNN)类似的网络结构,使用全卷积层代替全连接层,网络由6个卷积层组成,为了提高网络的判别能力,在判别器网络中增加了一层卷积层;并在最后一层使用全卷积层代替全连接层,以更好地保留空间信息;前四个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为2,第五个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为1,六个卷积层的通道数分别为(64,128,256,512,1024,1)。除最后一层外,每个卷积层后面均连接一个非线性激化函数Leaky ReLU,其定义式为:

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ \alpha x_i & \text{if } x_i < 0 \end{cases}, \text{其中 } i=1,2,\dots,5, \alpha \text{ 是人为设定的在 } (0,1) \text{ 区间内的修正参数,本发明中}$$

设定 $\alpha=0.2$ 。

[0073] 步骤103:将所述新的源数据集SG-GTA5和目标数据集作为语义分割模型的训练数据集,对所述生成对抗网络进行自适应学习率的对抗训练,得到优化的生成器网络模型。步骤103具体包括:

[0074] 将带标注的所述新的合成数据集SG-GTA5中的图像和未带标注的Cityscapes训练数据中的图像作为所述生成器网络的输入,通过所述生成器网络分别得到对应的特征图;

[0075] 将所述得到的特征图作为所述判别器的输入,采用自适应学习率调整多层特征的对抗性学习方法,通过反向传播的方式,分别对判别器网络和生成器网络依次进行更新优化。

[0076] 对抗训练的具体实施步骤为:

[0077] 1、分别将来自源数据集SG-GTA5的图像 $I_S$ 和目标数据集CityScapes的图像 $I_T$ 输入到生成器网络中,以达到参数共享的目的;

[0078] 2、通过生成器网络分别得到源数据集SG-GTA5中的图像和目标数据集CityScapes中的图像对应的特征图 $Y_S, Y_T$ ;

[0079] 3、再使用步骤2得到的特征图作为判别器的输入,以判断输入的特征图是来自源域还是目标域,根据该判别结果计算判别器网络的损失值;

[0080] 对于判别器网络,损失函数如下式:

[0081]

$$L_d^i(P) = (1 - b_i) \cdot \left[ - \sum_{h,w} (1 - z) \log(D(P)^{(h,w,0)}) + z \log(D(P)^{(h,w,1)}) \right] + b_i \quad (1)$$

[0082] 其中,若判别来自于源域则取 $z=0$ 代入式(1)中,若判别器判断结果来自于目标域则取 $z=1$ 代入式(1)计算判别器网络的损失,其中 $b_i$ 指判别器每次训练所对应的学习率, $b_i$ 的计算如式(2), $h$ 和 $w$ 分别为图像高和宽, $i=0,1$ ,其中 $i=0$ 时对应第四层, $i=1$ 时对应第五层。 $D(P)^{(h,w,0)}$ 和 $D(P)^{(h,w,1)}$ 分别表示判别器网络对合成数据集和真实数据集的判断结果。

[0083] 4、先固定生成器网络的参数,根据步骤3计算得到的损失值 $L_D$ ,更新判别器网络的参数;

[0084] 5、将已知的合成图像的标注 $T_S$ 输入生成器网络得到对应的特征图,并与步骤2计

算得到的源数据集的特征图 $Y_S$ 计算交叉熵损失值 $L_{Seg}$ ;

[0085] 6、再固定判别器网络的参数,将更新后的判别器网络的对抗损失 $L_{adv}$ 与步骤5中得到的交叉熵损失 $L_{Seg}$ 进行融合,并加入自适应的学习率 $b_i$ 。通过该学习率调整对抗损失与交叉熵损失的权值,从而动态更新优化生成器网络,不断更新优化生成器网络,实际上是通过判别器网络输出的损失来优化生成器网络的参数,得到一个最优的生成器网络。

[0086] 步骤3和步骤6中的自适应学习率函数如下:

$$[0087] \quad b_i = \begin{cases} (1 - \frac{j}{n})^p, & \text{if } i = 0; \\ 50(1 - \frac{j}{n})^p, & \text{if } i = 1. \end{cases} \quad (2)$$

[0088] 其中 $i=0,1$ , $i=0$ 时对应第四层, $i=1$ 时对应第五层, $j$ 表示第 $j$ 次训练, $n$ 代表总训练次数, $p$ 为固定参数,其值设置为0.9。

[0089] 在网络训练的过程中,对于生成器网络的参数优化,采用牛顿快速随机梯度下降法,对于判别器网络的参数优化,使用自适应矩估计优化方法。

[0090] 步骤104:将城市场景数据集Cityscapes中的测试图像输入至所述优化的生成器网络模型中,得到分割结果。

[0091] 步骤105:对所述分割结果进行定量和定性评估,得到分割性能结果;具体包括:

[0092] 1、将Cityscapes中的测试图像输入步骤103中训练好的模型中,得到分割结果;

[0093] 2、再对上一步中的分割结果计算其平均交并比mIoU,与目前主流的性能较好的方法进行定量和定性的评估,验证本发明提出的分割模型的有效性。

[0094] 图2为本发明自适应对抗学习的城市交通场景语义分割系统结构图。如图2所示,一种自适应对抗学习的城市交通场景语义分割系统,包括:

[0095] 训练数据集获取模块201,用于获取语义分割模型的训练数据集,对原始的游戏合成数据集GTA5进行预处理,得到新的与城市场景真实数据集分布接近的合成数据集SG-GTA5,作为源数据集,CityScapes训练数据集中不使用标注信息的图片作为目标数据集,源数据集和目标数据集组成语义分割模型的训练数据集;

[0096] 生成对抗网络构建模块202,用于构建弱监督的语义分割模型,即生成对抗网络模型,所述生成对抗网络包括生成器网络和判别器网络;

[0097] 生成对抗网络训练模块203,用于将所述新的源数据集SG-GTA5和目标数据集作为训练数据,对所述生成对抗网络进行自适应学习率的对抗训练,得到优化的生成器网络模型;

[0098] 分割结果确定模块204,用于将城市场景数据集Cityscapes中的测试图像输入至所述优化的生成器网络模型中,得到分割结果;

[0099] 评估模块205,用于对所述分割结果进行定量和定性评估,得到分割性能结果。

[0100] 所述训练数据集获取模块201,具体包括:

[0101] 训练数据集获取单元,用于以CityScapes训练数据集中不使用标注信息的图片和带标注的游戏合成数据集GTA5作为输入,采用语义感知生成对抗网络算法对两个输入数据集进行对抗训练,得到一个训练完成的网络模型,将原始合成数据集GTA5输入到该训练完成的网络模型中,得到新的合成数据集SG-GTA5,并将其作为语义分割模型的源数据集,

CityScapes训练数据集中不使用标注信息的图片作为目标数据集,源数据集和目标数据集组成了语义分割的训练数据集。

[0102] 所述生成对抗网络构建模块202,具体包括:

[0103] 生成器网络构建单元,用于构建生成器网络,采用DeepLab v2作为网络基本框架;在最后一层采用多孔空间金字塔池化结构代替标准卷积作为最终的分类器;再将第四层和第五层卷积层的步长由2改成1;同时将第四层和第五层空洞卷积层的空洞数分别设置成2和4;最后,使用一个上采样层得到与输入图像大小相同的特征图;

[0104] 判别器网络构建单元,用于构建判别器网络,采用与传统卷积神经网络CNN类似的网络结构,使用全卷积层代替全连接层,网络由6个卷积层组成,前四个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为2,第五个卷积层的卷积核大小设为 $4 \times 4$ ,步长设为1,它们的通道数分别为(64,128,256,512,1024,1),除最后一层外,每个卷积层后面均连接一个非线性激活函数Leaky ReLU。

[0105] 所述生成对抗网络训练模块203,具体包括:

[0106] 特征图提取单元,用于将带标注的源数据集中的图像和未带标注的目标数据集中的图像作为所述生成器网络的输入,通过所述生成器网络分别得到其对应的特征图;

[0107] 参数优化单元,用于将所述得到的特征图作为所述判别器的输入,采用自适应学习率调整各层特征的对抗性学习方法,通过反向传播的方式,对判别器网络和生成器网络中的参数进行不断的更新优化;

[0108] 更新优化过程具体如下:

[0109] 固定生成器网络参数,对所述判别器网络采用自适应的学习率进行训练,得到训练后的判别器网络模型;

[0110] 固定所述训练后的判别器网络模型的参数,对所述生成器网络进行反向传播训练,得到优化的生成器网络模型,如此循环迭代,最后得到最优的生成对抗网络模型。

[0111] 本发明的自适应对抗学习的城市交通场景语义分割方法,具备以下优点:

[0112] (1) 构建了新的与真实数据集CityScapes分布接近的合成数据集SG-GTA5,提高了模型的泛化能力和分割精度。

[0113] 用不带标注信息的目标数据集和带标注信息的游戏合成数据集GTA5进行语义感知对抗性学习,得到新的数据集SG-GTA5,该数据集在颜色、纹理等分布上与真实数据集CityScapes更接近,从而能有效降低后续的对抗损失值,避免模型在反向传播过程中出现梯度爆炸,提高了模型的泛化能力和分割精度。

[0114] (2) 提出了一种学习率自适应调整各层特征的对抗性学习方法,以适应网络模型不同特征层的特点,从而提高了算法的性能。

[0115] 在不同特征层的对抗学习中使用自适应的学习率,通过该学习率调整各特征层的损失值,起到对损失值进行正则化的作用避免其出现较大的波动,并利用正则化的损失值自适应更新网络参数,优化了对抗网络中生成器网络和判别器网络的性能,从而进一步提高模型的分割精度。

[0116] (3) 在模型的判别器网络中新增加一层卷积层,步长设置为1,可以更好地学习物体类别的特征,有效缓解类漂移的问题。

[0117] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他

实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0118] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本发明的限制。

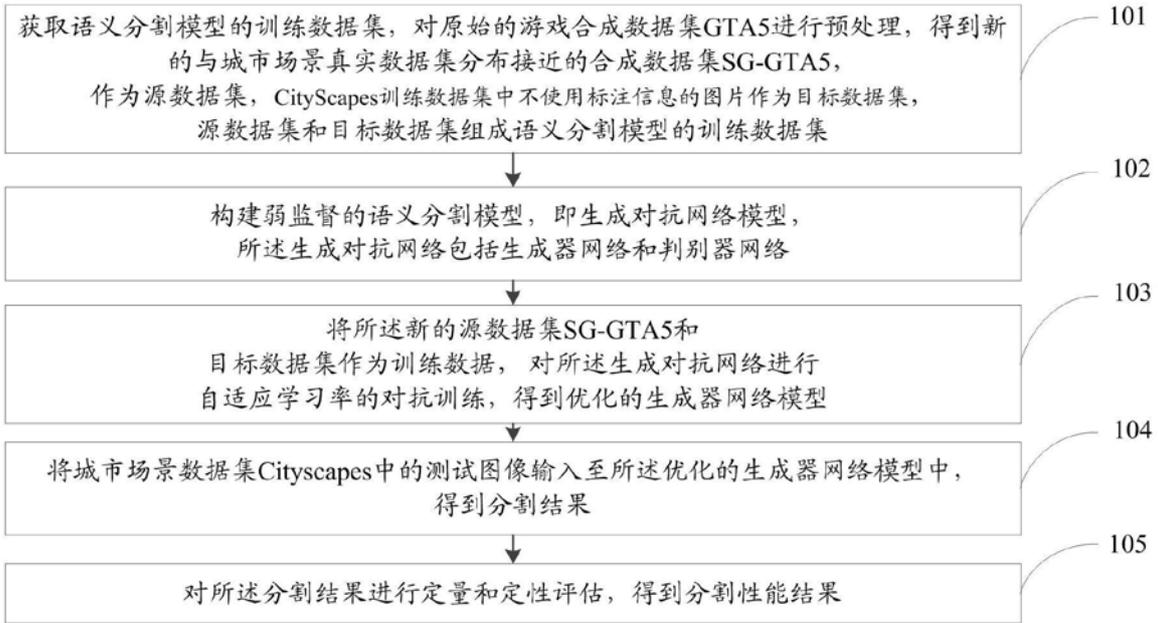


图1



图2

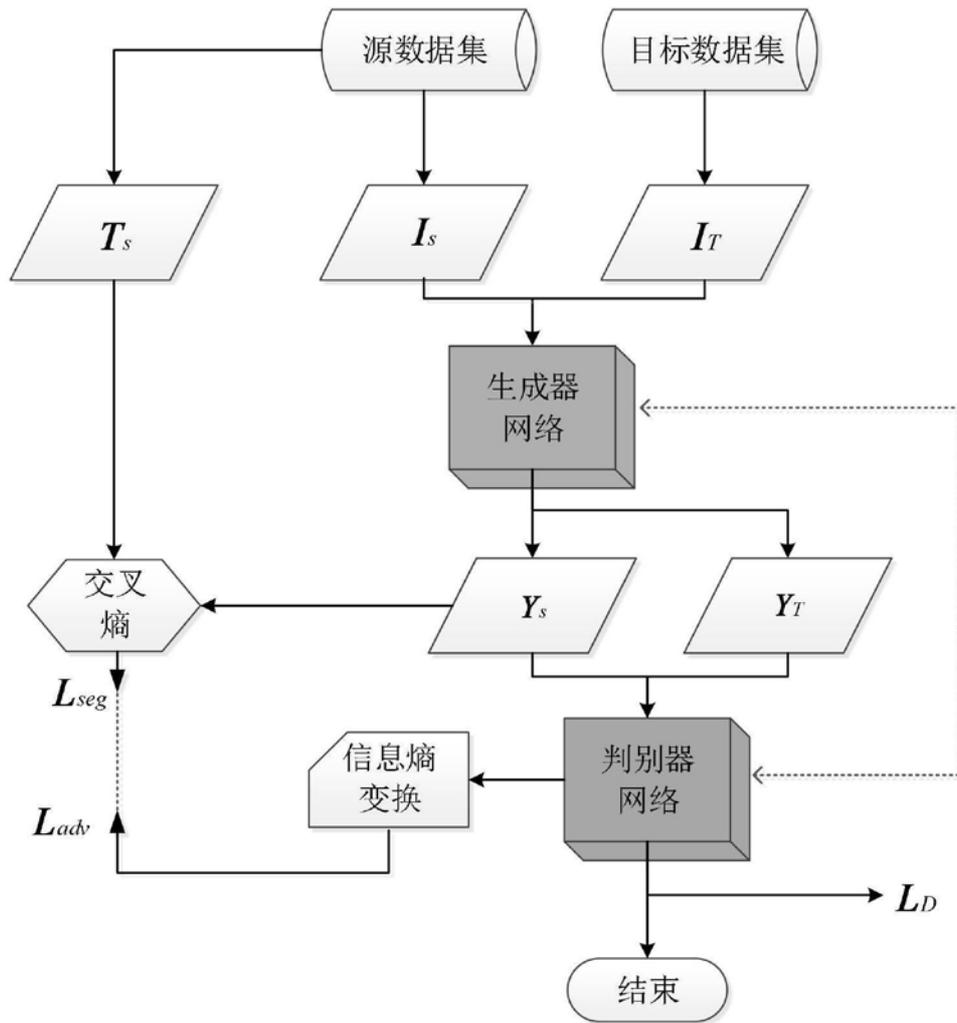


图3