文章编号:1001-9014(2020)04-0473-10 DOI:10.11972/j.issn.1001-9014.2020.04.012

基于生成式对抗网络的遥感图像半监督语义分割

刘雨溪[1,2], 张 铂[1,2], 王 斌[1,2]
(1. 复旦大学 电磁波信息科学教育部重点实验室,上海 200433;
2. 复旦大学 信息学院智慧网络与系统研究中心,上海 200433)

摘要:高分辨率遥感图像的语义分割问题是目前遥感图像处理领域中的研究热点之一。传统的有监督分割方法需要大量的标记数据,而标记过程又较为困难和耗时。针对这一问题,提出一种基于生成式对抗网络的半监督高分辨率遥感图像语义分割方法,只需要少量样本标签即可得到较好的分割结果。该方法为分割网络添加全卷积形式的辅助对抗网络,以助于保持高分辨率遥感图像分割结果中的标签连续性;更进一步,提出一种新颖的能够进行注意力选择的对抗损失,以解决分割结果较好时判别器约束的分割网络更新过程中存在的难易样本不均衡问题。在ISPRS Vaihingen 2D语义标记挑战数据集上的实验结果表明,与现有其它语义分割方法相比,所提出方法能够较大幅度地提高遥感图像的语义分割精度。

关 键 词:高分辨率遥感图像;语义分割;深度学习;生成式对抗网络;损失函数
中图分类号:TP751 **文献标识码**:A

Semi-supervised semantic segmentation based on Generative Adversarial Networks for remote sensing images

LIU Yu-Xi[1,2], ZHANG Bo[1,2], WANG Bin[1,2]
(1. Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China;
2. Research Center of Smart Networks and Systems, School of Information Science and Technology, Fudan University, Shanghai 200433, China)

Abstract: Semantic segmentation of very high resolution (VHR) remote sensing images is one of the hot topics in the field of remote sensing image processing. Traditional supervised segmentation methods demand a huge mass of labeled data while the labeling process is very consuming. To solve this problem, a semi-supervised semantic segmentation method for VHR remote sensing images based on Generative Adversarial Networks (GANs) is proposed, and only a few labeled samples are needed to obtain pretty good segmentation results. A fully convolutional auxiliary adversarial network is added to the segmentation network, conducing to keeping the consistency of labels in the segmentation results of VHR remote sensing images. Furthermore, a novel adversarial loss with attention mechanism is proposed in the paper in order to solve the problem of easy sample over-whelming during the updating process of the segmentation network constrained by the discriminator when the segmentation results can confuse the discriminator. The experimental results on ISPRS Vaihingen 2D Semantic Labeling Challenge Dataset show that the proposed method can greatly improve the segmentation accuracy of remote sensing images compared with other state-of-the-art methods.

Key words: very high resolution remote sensing images, semantic segmentation, deep learning, generative adversarial networks, loss function
PACS:84.40.Xb

收稿日期:2019-10-08,**修回日期**:2019-12-06 **Received date**:2019-10-08,**Revised date**:2019-12-06
基金项目:国家自然科学基金(61971141, 61731021)
Foundation items:Supported by National Natural Science Foundation of China (61971141,61732021)
作者简介(Biography):刘雨溪(1994-),女,吉林延边人,硕士研究生,主要研究领域为高分辨率遥感图像的语义分割. E-mail: think6c@163.com
*通讯作者(Corresponding author): E-mail: wangbin@fudan.edu.cn

引言

语义分割是一种视觉场景解析任务，其目的在于对输入图像进行逐像素的标签预测。高分辨率遥感图像包含丰富的地物信息，对其进行语义分割能够实现对不同地物的识别，这在精准农业、环境监测、城市规划等领域都有重要的应用前景。然而，由于高分辨率遥感图像具有空间分辨率高、细节复杂等特点，现有的面向自然图像的语义分割技术无法直接应用于高分辨率遥感图像的语义分割中。如何实现高精度的高分辨率遥感图像语义分割，仍面临着诸多挑战，其困难主要表现在三个方面：第一，由于传感器角度、环境变化等因素的影响，高分辨率遥感图像具有类内差异大、类间差异小的特点，这会导致有效特征提取的困难；第二，高分辨率遥感图像尺寸较大，训练时需划分为较小的子图，这会破坏图像的连续性；第三，有监督的高分辨率遥感图像语义分割方法需要大量的像素级标签，而获取这些语义标签的过程则较为耗时耗力。

对于第一个挑战，传统的高分辨率遥感图像分割方法（包括基于像元、基于边缘检测、基于区域和基于物理模型等方法）难以有效地实现高分辨率遥感图像的特征提取，无法解决遥感图像类内差异大、类间差异小的问题，因此难以得到高精度的分割结果[1-3]。近年来，随着深度学习的发展，许多基于卷积神经网络的模型在高分辨率遥感图像的语义分割上取得了较好的效果[4-5]。卷积神经网络能够提取图像不同层级的空间-语义特征，且所提取特征具有具有尺度不变性、旋转不变性和亮度不变性，对不同环境、传感器角度等因素的影响更为鲁棒，因此对遥感图像有更好的特征表达能力[6]。目前，基于卷积神经网络的高分辨率遥感图像语义分割方法主要可分为三类：1）扩大感受野或结合多个不同尺度进行特征的提取，如使用孔洞卷积、金字塔池化模型、或构造网络来学习融合不同分辨率的图像等[6-8]；2）更多地考虑如何提取上下文信息，并通过跳级结构将上下文信息和空间信息进行融合[9-10]；3）使用集成学习的方法，联合多个不同初始化方式和不同结构的网络共同进行分割，或使用迁移学习，将其它数据类型的预训练模型迁移使用到目标数据上[11-13]。

然而，上述的基于卷积神经网络的遥感图像语义分割模型难以解决其它的两个难点问题。对于第二个挑战，现有的遥感图像语义分割模型多使用如条件随机场（CRFs）等后处理方法[14]，但这类方法只是使分割结果在局部上保持平滑，既无法从根本上改善遥感图像分割网络的分割效能，也难以学习整幅高分辨率遥感图像的上下文信息，还会在预测阶段导致额外的计算量。而对于第三个挑战，目前还没有针对高分辨率遥感图像语义分割的端到端半监督模型，然而，如何实现高精度的端到端高分辨率遥感图像半监督语义分割，在实际应用中具有十分重要的现实意义。

最近，在自然图像语义分割中的研究表明，将生成式对抗网络[15]引入语义分割，一方面能够学习到图像的结构信息，保持分割结果在空间上的标签连续性；另一方面，无标签样本能够用于训练模型，可减少对语义标注的需求[16-18]。高分辨率遥感图像的语义分割问题与自然图像的语义分割具有一定的相似性，为遥感图像语义分割网络添加辅助对抗网络，可以达到对遥感图像进行半监督语义分割的目的，解决第三个挑战的难题。然而，与自然图像不同，高分辨率遥感图像的尺寸很大，直接将其输入网络进行训练，会导致计算机显存不足的问题，因此必须要先将大尺寸遥感图像划分为尺寸较小的子图，利用子图来训练网络，但是，这一操作会导致原图像的空间连续性被破坏。引入生成式对抗网络固然有助于分割结果连续性的保持，但是，自然图像语义分割中所使用的分类形式辅助对抗网络只能学习到子图整体的结构信息（如某一类地物在子图中的位置），却无法学习到子图内部的局部结构信息（如两种地物之间的位置关系），而对遥感图像来说，地物在子图中的分布位置是不固定的，不同类别地物间的位置分布关系才更能够代表遥感图像的结构特征。基于这一考虑，我们拟对辅助对抗网络进行改进，将判别器由分类网络改为端到端的全卷积网络。这相当于将判别对象由子图大小的分割结果图推进到其中的每个像素点，判别器根据该像素周围对应感受野大小的区域内的上下文信息对其来源进行辨别，使判别器能够学习到子图内的局部结构信息，进而通过循环训练整个子图集学到划分前遥感图像的地物分布特点，从而可以维持大尺寸遥感图像分割结果的标签连续性。

但是，将判别器改为全卷积网络会导致新的较为严重的问题：由于相应的对抗损失变成了对分割结果和真实标签逐像素求交叉熵再求和的形式，则在生成器对输入图像大部分区域的分割结果较好

时,大量的足以迷惑判别器的像素点对应的损失函数的值较小,从而会拉低整体图像的损失,使判别器难以进一步对分割网络的梯度更新方向进行有效约束。为了解决这一问题,我们提出一种新颖的网络对抗损失,其作用在于自适应地调整难易样本损失值在整体损失中的权重,使模型在训练时能够进行注意力选择,更加专注于分类结果错误的部分,从而保证判别器始终能够对分割网络的更新方向进行有效的限制。

综上所述,本文提出一种基于生成式对抗网络的高分辨率遥感图像半监督语义分割模型,以便在较大程度地缓解对样本标签需求的同时,实现高精度的高分辨率遥感图像语义分割。针对高分辨率遥感图像的特点,构造出一种端到端的全卷积网络作为判别器,它能够学到图像的局部结构信息,有助于保持整体图像分割结果的标签连续性,优化分割结果;更进一步,本文提出一种新颖的基于注意力选择的对抗损失函数,它能够促使判别器始终对分割网络的梯度更新进行有效约束。相较于传统的非深度学习分割方法,所提议模型采用卷积神经网络,将有效地提取高分辨率遥感图像的特征;相较于其它基于深度学习的遥感图像语义分割模型,所提议方法将有效保持大尺寸遥感图像分割结果的空间连续性,同时所构造的基于注意力机制的损失函数能够使分割结果得到进一步优化。实验结果表明,所提出的模型在仅使用少量标签的情况下,能够充分利用无标签图像数据的信息,有效地实现对高分辨率遥感图像半监督语义分割,较好地解决有监督分割模型中由于标签数减少而带来的分割精度急剧下降的问题。

1　相关工作

1.1　生成式对抗网络

生成式对抗网络的目标在于学习真实数据的分布,从而能够生成近似于真实数据的伪数据。如图1所示,一个典型的生成式对抗网络包括两个子网络,分别为生成器G和判别器D。生成器用于得到非常近似于真实数据的伪数据,并希望这些伪数据能够迷惑判别器;而判别器的任务是对其输入的真实数据和伪数据进行辨别,希望能够尽可能地将伪数据从真实数据中区分出来。二者之间存在着最小-最大竞争关系。通过不断地交替优化生成器和判别器,最终二者间的博弈能够达到平衡状态。此时,判别器难以辨别其输入的来源,则可认为生成器能够生成非常近似于真实数据的虚假数据。

二者间的竞争关系可由以下损失函数表示:

$$\min_G \max_D V(G,D) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))], \quad (1)$$

其中,x代表真实数据,z代表生成器输入的随机向量,G代表生成器,D代表判别器。$D(x)$表示输入为真实数据时判别器的输出,$D(G(z))$表示输入为生成器生成的伪数据时判别器的输出。

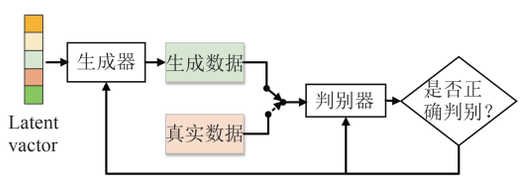

图1　生成式对抗网络框架

Fig. 1　The framework of Generative Adversarial Networks

1.2　基于生成式对抗网络的自然图像半监督语义分割

将生成式对抗网络引入自然图像语义分割中的半监督方法分为两种思路:一种是将分割网络作为判别器,通过生成器生成伪数据来扩充数据集,从而减少对有标签样本数量的需求[19];另一种是将分割网络作为生成器,将得到的分割结果和真实标签分别输入判别器进行判断,希望分割网络输出的分割结果能够尽可能地迷惑判别器,而同时判别器能够尽量辨别出当前输入的来源,从而实现网络的对抗[17]。第二种方法不仅能够实现自然图像的半监督分割,还能够保持分割结果的空间连续性,其损失函数可具体描述为如下。

设分割网络的输入自然图像为$\mathbf{X}_n$,$S(\cdot)$表示分割网络,则$S(\mathbf{X}_n)$为分割网络输出的分割结果。令$D(\cdot)$表示判别网络,$\mathbf{Y}_n$表示真实语义标签的one-hot编码,则$D(S(\mathbf{X}_n))$表示当判别网络的输入为分割结果时的输出,而$D(\mathbf{Y}_n)$表示当判别网络的输入为真实语义标签时的输出,则有

$$L_D = -\sum_{h,w}(1-y_n)\log(1-D(S(\mathbf{X}_n))^{(h,w)}) + y_n\log(D(\mathbf{Y}_n)^{(h,w)}), \quad (3)$$

$$L_{ce} = -\sum_{h,w}\sum_{c\in C}\mathbf{Y}_n^{(h,w,c)}\log(S(\mathbf{X}_n)^{(h,w,c)}), \quad (4)$$

$$L_{adv} = -\sum_{h,w}\log(D(S(\mathbf{X}_n))^{(h,w)}), \quad (5)$$

$$L_{semi} = -\sum_{h,w}\sum_{c\in C}(I(D(S(\mathbf{X}_n))^{(h,w)} > T_{semi}) \cdot \hat{\mathbf{Y}}_n^{(h,w,c)}\log(S(\mathbf{X}_n)^{(h,w,c)})), \quad (6)$$

其中，L_D为判别器损失函数，(h,w)表示像素的位置，y_n表示判别网络输入的来源，$y_n=0$表示输入为分割网络的分割结果，$y_n=1$表示输入为真实标签。生成器损失函数为L_{ce}、L_{adv}和L_{semi}三项的加权和，L_{ce}为语义分割中常用的多分类交叉熵损失，用于保证分割网络的基础分割能力；L_{adv}为对抗损失，极小化L_{adv}相当于极大化判别器损失函数L_D中的第一项，体现了判别器对于分割网络梯度更新方向的约束，即“对抗”的过程；L_{semi}为附加的自适应半监督损失，其中T_{semi}为选择阈值，$I(\cdot)$为指示函数，$\hat{\mathbf{Y}}_n$是对分割结果的one-hot编码，即对于$c^*=\arg\max_c S(\mathbf{X}_n)^{(h,w,c)}$，有$\hat{\mathbf{Y}}_n{}^{(h,w,c^*)}=1$，通过$I(D(S(\mathbf{X}_n))^{(h,w)}>T_{semi})$项，可以选择出分割结果能够迷惑判别器的区域，进而对这部分区域的分割结果做进一步的优化。由于在利用判别器约束分割网络梯度更新的过程中，只需要自然图像数据而不需要对应的真实标签，因此无标签样本也可以与用于训练对抗损失，从而实现了对自然图像的半监督语义分割。

由于高分辨率遥感图像的地物分布较为复杂，上述的第一种半监督语义分割方法难以生成高质量的遥感图像，并且其无法保持分割结果的空间连续性，因此本文借鉴第二种方法的思路来设计高分辨率遥感图像的半监督语义分割模型。

2 模型构建

2.1 网络结构

为了捕捉预测标签的空间邻近性，同时缓解对有标签样本数量的需求，本文将生成式对抗网络引入到高分辨率遥感图像语义分割任务中，提出一种基于生成式对抗网络的可实现高分辨率遥感图像半监督语义分割的端到端模型，其总体网络结构如图2所示。其中，分割网络作为生成器，实现从遥感图像到分割结果的转换；而判别网络作为辅助对抗网络，用来对分割网络的梯度更新方向做进一步约束。在二者的对抗中，分割网络的目的在于得到尽可能接近于真实标签、能够迷惑判别器的分割结果，而判别网络需要对分割结果和真实标签进行分辨，避免被生成的分割结果迷惑。判别网络对分割结果和真实标签图的辨别过程，实际上是在学习二者间的高阶差异，因此在利用辅助判别网络约束分割网络的更新时，这种高阶关系可被考虑到分割网络中，即分割网络被迫保持分割结果和真实标签的高阶一致性，从而能够维持分割结果在空间上的连续性。

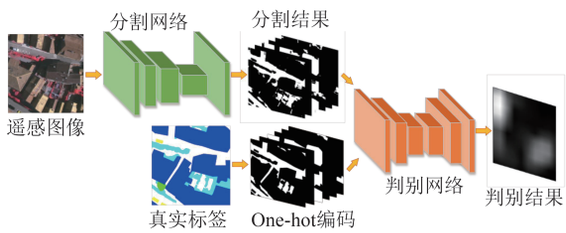

图2 所提议方法的总体框架

Fig. 2 The overall framework of the proposed method

由于高分辨率遥感图像中的不同类别地物分布较为分散，并且图像的空间分辨率相对较高，需划分为子图进行训练，基于以上特点，我们构建了一种端到端的全卷积网络，并用它来代替原始生成对抗网络中的分类形式的判别器。相比于其它方法，文中所提方法的输出是一张置信度图而不是一个标量值，输出的置信图中的每一像素表示输入中对应像素点来自于分割结果还是真实标签。使用全卷积网络的一个好处在于，由于没有全连接层的存在，输入图像的尺寸不再受到限制，从而进一步增加了所提议模型的普适性。更重要的是，相较于其它形式的判别网络对输入子图整体结构的关注，我们所构造的端到端的判别网络能够捕捉到高分辨率遥感图像局部区域的结构特征，从而通过利用整个子图集合的循环训练，学习到划分前的高分辨率遥感图像的地物分布特点。因此，所构建的端到端的判别网络更加适合于维持高分辨率遥感图像语义分割结果的标签连续性。

而所提议模型对半监督方法的实现，则体现在利用判别网络约束分割网络梯度更新方向的过程中。此时固定判别网络，将高分辨率遥感图像通过分割网络得到的分割结果输入判别器，得到判别器输出端的交叉熵损失为式(3)中的前一项。要使分割结果尽可能地迷惑判别器，也就是要极大化这一损失，进而可推导出对抗损失的表示形式如式(5)所示。通过极小化对抗损失，反向传播更新分割网络，就可以实现判别器对于分割网络梯度更新的约束。在这一过程中，只需要基于遥感图像数据得到分割结果，并不需要对应的真实标签图，因此没有真实标签的遥感图像数据也可以用于这一过程的训练中，从而实现对高分辨率遥感图像的半监督语义分割。

2.2　基于注意力选择的损失函数构建

由于判别器的输出为其输入尺寸大小的置信图,其损失函数为各个像素点交叉熵的和,因此,当训练进行到一定程度、分割网络对输入图像的大部分像素点都能够正确分类时,这些被正确分类的简单样本点会使分割结果经过判别器的总体对抗损失值降低,进而导致总体梯度被稀释,使得少量的不能迷惑判别器的困难样本点的分割结果难以得到进一步的修正。为了解决这一问题,我们在公式(5)的原始对抗损失的基础上,增加注意力选择机制,其数学表达为下式

$$L_{adv\_att} = -\sum_{h,w}(1 - D(S(\mathbf{X}_n))^{(h,w)})^{\gamma} \cdot \log(D(S(\mathbf{X}_n))^{(h,w)}) \quad . \quad (7)$$

上式相对于式(5)的对抗损失,增加了$(1 - D(S(\mathbf{X}_n))^{(h,w)})^{\gamma}$系数项,其目的在于可自适应地调整难易样本的损失值在整体损失中的权重,从而实现注意力选择。具体而言,当$D(S(\mathbf{X}_n))$较大时,说明分割网络的分割结果可以迷惑判别器,为简单样本,则乘以一个较小的权重系数;当$D(S(\mathbf{X}_n))$较小时,分割结果不能迷惑判别网络,则为困难样本,因此乘以一个较大的权重系数。由此,简单样本损失在总体损失中的占比被极大减小,相应地,困难样本的损失被放大,网络更加关注对困难样本的训练,从而确保判别器始终能够对分割网络的梯度更新方向进行有效的约束。

最终,所提出模型的分割网络损失函数如下所示:

$$L_{seg} = L_{ce} + \lambda_{adv} L_{adv\_att} + \lambda_{semi} L_{semi} \quad , \quad (8)$$

其中,L_{adv_att}为本文提出的基于注意力选择的对抗损失,L_{ce}和L_{semi}为交叉熵损失和自适应半监督损失,对应公式(4)和公式(6),λ_{adv}和λ_{semi}为对抗损失和半监督损失对应的权重。

3　实验结果与分析

3.1　实验数据

在ISPRS Vaihingen 2D语义标记挑战数据集[20]上评估所提出方法的性能。该数据集包含分辨率为9厘米的高分辨率正射影像(TOP)片以及相应的数字表面模型(DSM)。数据集中共包含33幅图像,如图3(a)所示,每幅图像的尺寸不完全相同,其中16幅图像提供了可用于训练的标签(1, 3, 5, 7, 11, 13, 15, 17, 21, 23, 26, 28, 30, 32, 34, 37),其余17幅图像为测试集。为进行消融实验,遵循相关文献12-13的划分方法,将原训练集中12幅(1, 3, 11, 13, 15, 17, 21, 26, 28, 32, 34, 37)图像作为本文实验中的训练集,其余4幅图像(5, 7, 23, 30)作为验证集。在实验中只使用TOP数据,它具有近红外、红光和绿光三个波段。图3(b)为其中的一幅遥感图像,图3(c)为其对应的标签图及图例,图例中的六种颜色对应Vaihingen数据集包含的六类地物,分别为不透水表面、建筑物、低植被、树木、汽车和背景。

由于遥感图像尺寸过大,直接输入网络会导致显存不足的问题。因此,首先将每幅遥感图像都划分为若干相同大小的子图,每个子图大小为321×321,相邻子图间有100个像素的重叠,在边缘处的子图,为了保持与前面图像相同的大小,重叠区域尺寸会有相应的变化。在处理边缘子图时,重叠区域的宽$b = 321 - (a - (321 - 100) \times (n - 1) - 100)$,其中$a$为待划分图像的长,$n$为根据边长$a$所能划分出子图的数目,如图4所示(在图4中,$n = 3$)。此外,为了进一步缓解过拟合问题,我们对数据集进行扩充,在训练时对每个子图进行随机镜像和翻转。

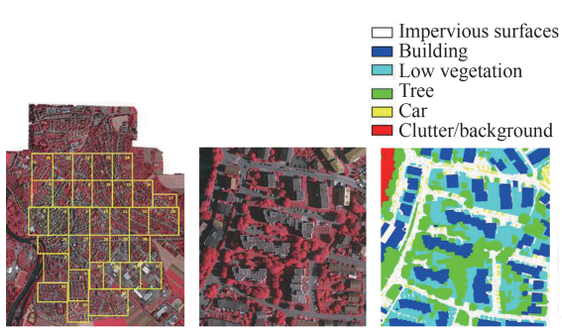

图3　ISPRS 2D Vaihingen数据集示意图 (a) 全部遥感图像,包含近红外、红光和绿光波段,(b) 编号为2的遥感图像,(c) 对应标签及图例

Fig.3　Illustration of the ISPRS 2D Vaihingen Labeling dataset (a) the entire remote sensing image, including near-infrared, red and green bands, (b) partial remote sensing image numbered 2, and (c) corresponding label map and its legend

3.2　评价指标

使用全局准确率OA和F1-Score作为评价指标来评估文中方法的有效性。全局准确率OA表示所有判断正确的结果占总体的比重,F1-Score为Precision和Recall的调和平均数,表示二者的综合结果。TP、TN、FP和FN的含义如图5所示。

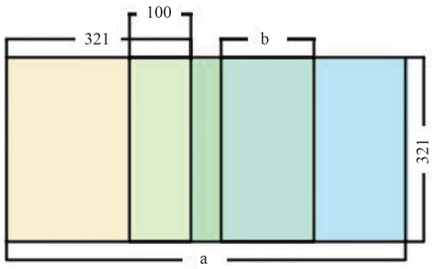

图4 子图划分示意图

Fig. 4 Illustration of cropping the entire image

全局准确率OA和F1-score的计算公式如下：

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad , \quad (9)$$

$$F_1 = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad , \quad (10)$$

其中

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad , \quad (11)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad . \quad (12)$$

混淆矩阵		真实值	
		Positive	Negative
预测值	Positive	TP	FP
	Negative	FN	TN

图5 混淆矩阵示意图

Fig. 5 Illustration of confusion matrix

3.3 实验细节

实验在Pytorch框架上构建模型，并使用NVIDIA TITAN XP的单个GPU对模型进行训练，该GPU具有12GB的内存。分割网络采用基于ResNet-101[21]的DeepLab-v2模型[22]，使用在ImageNet和MSCCOCO数据集上预训练的权重来初始化分割网络，在减少训练时间的同时能够缓解过拟合问题。分割网络的初始学习速率设置为2.5×10^{-4}，采用动量为0.9、权值衰减为10^{-4}的随机梯度下降(SGD)优化方法，并使用大小为0.9的poly策略进行更新。构造的判别网络为端到端的全卷积网络，分为编码器和解码器两个部分，编码器包含三个卷积核大小为3×3的卷积层，维度分别为32、64、128，步长为2，每个卷积层后附带LeakyReLU激活函数[23]以增加非线性；解码器各层首先对前一层的输出特征使用双线性插值的方法进行上采样，然后进行卷积操作，维度分别为64、32、1，即最终可得到与判别网络输入尺寸大小一致的判别置信图。在训练判别网络时，使用初始学习速率为10^{-4}的Adam优化器。此外，在训练过程中，将λ_{adv}设置为0.01，λ_{semi}为0.1，T_{semi}为0.2，γ为1。这些参数值的具体确定方法将在超参数分析实验中进行说明。

在进行半监督实验时，从训练集中随机选择对应比例的遥感图像样本及其标签作为有标签样本，其余所有样本(不包含标签)作为无标签样本来对模型进行训练。随机选择有标签训练样本而没有专门去挑选有代表性的训练图像的原因在于，前者能够使网络学习到更加鲁棒的特征，消减环境等因素对分割结果的影响。在实验中，首先使用有标签样本对分割网络单独训练5000个循环，再交替优化分割网络和判别网络，共训练20000次循环。先对分割网络进行单独训练的目的在于，使得模型在对抗训练时，分割网络输出的分割结果与真实标签相比具有一定的竞争性，从而能够有效的对判别器的参数进行更新。

3.4 实验结果

3.4.1 消融实验

为了检验所提出模型中的各个模块对结果带来的提升效果，我们分别在有监督和半监督的情况下进行消融实验。对于有监督模型，从基准分割模型(Baseline)开始，依次在模型中增加所构造的全卷积的辅助对抗网络($+L_{adv}$)和改进的对抗损失($+L_{adv\_att}$)，并使用全部的样本和标签来进行训练。而对于所提出的半监督模型，在训练基准模型(Baseline)时，我们仅使用随机选择的相应比例(1/4和1/8)的样本及其对应的标签。这是为了能够更好地与半监督方法的结果进行对比，以便确认半监督方法所能带来的提升效果。在训练依次添加各个模块(全卷积辅助对抗网络($+L_{adv}$)、改进的对抗损失($+L_{adv\_att}$)和自适应半监督损失($+L_{adv\_att}+L_{semi}$))的半监督模型时，除了随机选择的相应比例(1/4和1/8)的样本及其对应标签之外，训练集其余所有样本(未使用标签)也被用到模型的训练中。我们对不同的模型使用相同参数分别加以训练，具体的分割结果如表1所示。从表1中的实验结果可以得出以下结论。

1)在有监督模型中，可以看到，所构造的全卷积辅助对抗网络($+L_{adv}$)的引入使分割准确率提高了2.4%，平均F1值提高了3.7%。这是由于所构造的辅助对抗网络能够学习到遥感图像的局部结构信

息,有助于驱使分割结果和真实标签间维持高阶一致性,从而保持分割结果中像素间的空间连续性,改善分割结果。此外,所提出的基于注意力选择的对抗损失($+L_{adv\_att}$)也进一步提高了分割结果,证明所提出的对抗损失更加关注对困难样本的训练,能够使判别网络始终对分割网络的梯度更新方向进行有效的约束。

2)对于没有使用半监督方法的基准模型(Baseline),当所使用的有标签训练样本比例减少时,对应的分割准确率急剧下降,由85.3%(使用全部样本)下降到82.0%(使用1/4样本)和79.1%(使用1/8样本)。这说明对于有监督模型来说,使用的数据数量非常重要,即需求大量的训练样本,否则很难得到较好的结果;而所提出的基于生成式对抗网络的半监督语义分割模型,可以较大地缓解这种由于样本标签数目减少而导致的分割结果剧烈恶化问题。通过半监督对抗网络的引入($+L_{adv}$),分割准确度得到了较大幅度的提高,在1/4比例时提高了4.2%,而在1/8比例时提高了5.4%之多。而所提出的对抗损失($+L_{adv\_att}$)和附加的自适应半监督损失($+L_{adv\_att}+L_{semi}$)又进一步改善了分割结果,并且使用的样本标签的比例越小,改善效果越明显,如在1/4比例时,所提出的对抗损失使分割准确率提高了0.6%,自适应半监督损失提升了0.4%的准确率;而在1/8比例时,二者分别给分割准确率带来了1%和0.6%的提高。以上说明,所提议的模型能够充分利用无标签样本中所包含的信息来改善分割结果。

3.4.2　验证集结果对比

在实验中比较了所提议模型与在自然图像语义分割中性能优异的半监督语义分割模型SSGAN[19]和Semi-SegGAN[17]的分割结果,以此来验证本文所提议模型的有效性。SSGAN基于生成式对抗网络,使用少量有标签样本来生成伪图像、扩充数据集,从而实现半监督的语义分割。Semi-SegGAN则利用辅助对抗网络来对分割网络的更新方向进行约束,并可使用无标签样本来进行对抗训练,从而可实现半监督语义分割。由于所提议方法增加了注意力选择机制,可进一步解决分割结果较好时存在的梯度稀释问题。为了维持对比实验的公平性,本节实验中对不同模型均采用相同的基准网络,即不包含条件随机场的DeepLab-v2,并使用相同的样本(包括有标签样本和无标签样本)进行训练。

表2展示了不同方法分割准确率的结果对比。可以看出,SSGAN仅能够在基础网络(Baseline)上进行微小的提升,而Semi-SegGAN和本文模型都能使基准网络(Baseline)结果得到大幅度的改善。而所提出的方法通过在对抗损失中增加注意力选择机制,能够进一步提升最终的分割精度,这再一次证明,所提出的对抗损失函数能够有效缓解分割结果较好时简单样本过多稀释掉总体损失、从而导致判别器无法对分割网络的更新进行有效约束的问题。

3.4.3　超参数分析

在本小节中,我们分别对损失函数中的辅助对抗损失的权重λ_{adv}和自适应半监督损失的权重λ_{semi}、阈值T_{semi}和对抗损失中权值的指数γ的不同取值进行实验,从而确定合适的参数取值。

首先,根据相关文献的经验[24],将γ设置为1,然后分别调整λ_{adv}、λ_{semi}和T_{semi}进行实验,以确定三者

表1　不同标签样本比例下各部分提升效果比较

Table 1　Detailed performance comparison of each component with different proportions of labeled samples

类型	比例	Method	Imp Surf	Building	Low_veg	Tree	Car	OA	Mean F1
有监督	1	Baseline	86.3	92.6	70.7	85.6	69.9	85.3	81.0
		$+L_{adv}$	88.8	93.9	75.6	87.9	77.6	87.7	84.7
		$+L_{adv\_att}$	89.0	94.1	75.6	88.0	78.4	88.0	85.0
半监督	1/4	Baseline	83.0	90.7	62.7	83.7	62.0	82.0	76.4
		$+L_{adv}$	86.9	93.2	71.7	86.8	74.9	86.2	82.7
		$+L_{adv\_att}$	87.8	93.8	73.2	87.3	75.0	86.8	83.4
		$+L_{adv\_att}+L_{semi}$	87.8	93.9	73.8	87.4	76.2	87.2	83.8
	1/8	Baseline	80.0	87.9	60.1	81.0	49.1	79.1	71.8
		$+L_{adv}$	85.5	91.3	70.2	86.0	69.3	84.5	80.5
		$+L_{adv\_att}$	86.4	92.1	71.6	86.6	72.9	85.5	81.9
		$+L_{adv\_att}+L_{semi}$	86.7	92.1	74.2	87.0	73.1	86.1	82.6

表2 与其它半监督语义分割方法在验证集上的结果对比

Table 2 Accuracy comparison on validation dataset with other semi-supervised segmentation methods

比例	Method	Imp Surf	Building	Low_veg	Tree	Car	OA	Mean F1
1/4	Baseline	83.0	90.7	62.7	83.7	62.0	82.0	76.4
	SSGAN[19]	83.6	91.0	63.9	84.1	65.7	82.9	77.7
	Semi-SegGAN[17]	87.3	93.3	73.5	87.1	75.7	86.5	83.4
	本文方法	87.8	93.9	73.8	87.4	76.2	87.2	83.8
1/8	Baseline	80.0	87.9	60.1	81.0	49.1	79.1	71.8
	SSGAN[19]	81.1	88.3	62.5	82.0	54.3	81.6	73.6
	Semi-SegGAN[17]	85.9	91.5	70.8	86.1	72.3	84.9	81.3
	本文方法	86.7	92.1	74.2	87.0	73.1	86.1	82.6

的取值。本文采用控制变量法，并结合相关经验，进行了三组实验：

1）分别固定 λ_{adv} 为0.01、T_{semi} 为0.2，变化 λ_{semi} 的值；

2）分别固定 λ_{adv} 为0.01、λ_{semi} 为0.1，变化 T_{semi} 的值；

3）分别固定 λ_{semi} 为0.1、T_{semi} 为0.2，变化 λ_{adv} 的值。

具体实验结果如表3所示。由表3可见，当 λ_{adv} 为0.01、λ_{semi} 为0.1、T_{semi} 为0.2时，分割结果最好。此外，λ_{adv} 的取值对结果的影响要大于 λ_{semi} 和 T_{semi}，结合前面消融实验结果，可看出辅助对抗网络模块对结果的提升最为明显，这说明生成式对抗网络在遥感图像语义分割中的使用是所提出模型表现良好的主要原因，而改善后的对抗损失和自适应半监督损失能够进一步提高所提出模型的分割精度。

在确定了 λ_{adv}、λ_{semi} 和 T_{semi} 的取值后，进一步针对指数 γ 的取值进行实验。γ 的取值范围为0、0.5、1、2、5，其中 γ=0时即为原始的对抗损失，用来与所提出的对抗损失（γ>0）的分割结果作对比。如表4中的结果所示，当 γ>0时的分割结果均好于 γ=0时，这表明在对抗损失中增加的自适应权重系数确实能够缓解简单样本过多导致梯度稀释的问题，有助于分割准确性的提高。

3.4.4 测试集结果对比

为了进一步验证所提议方法的优越性，我们将所提议模型在测试集上的分割结果与其它用于高分辨率遥感图像语义分割模型的结果进行对比。对比方法均为基于卷积神经网络的深度模型，其中，UPB[9]、ITC_B2[7]和UFMG_3[28]均从融合多尺度特征的角度对全卷积网络加以改进，ETH_C[25]构造了一种高度结构化的基于哈尔小波的卷积神经网络模型，而CAS_Y3[6]使用PSPNet[29]的结构，VNU4使用FCN-8s网络[26]，CASZX1采用DeepLab-v3+模型[27]，这些网络均在自然图像的语义分割中性能表现优异。需要注意的是，上述对比方法均为有监督模型，只能使用有标签样本对网络进行训练；而本文所提出的模型可以进一步利用无标签样本来训练网络。因此，为了尽可能地利用数据集中所包含的信息，我们使用训练集中所有的样本及标签，以

表3 超参数 λ_{adv}、λ_{semi} 和 T_{semi} 取值分析

Table 3 Analysis on hyper-parameter λ_{adv}、λ_{semi} and T_{semi}

λ_{adv}	λ_{semi}	T_{semi}	OA	Mean F1
0.01	0	N/A	85.5	81.9
0.01	0.01	0.2	85.6	82.0
0.01	0.1	0.2	86.1	82.6
0.01	0.2	0.2	85.8	82.3
0.01	0.1	0.1	85.7	82.1
0.01	0.1	0.2	86.1	82.6
0.01	0.1	0.3	85.9	82.4
0.001	0.1	0.2	85.4	81.8
0.01	0.1	0.2	86.1	82.6
0.1	0.1	0.2	84.9	80.9

表4　超参数γ取值分析

Table 4　Analysis on hyper-parameter γ

γ	OA	Mean F1
0	85.1	81.6
0.5	85.3	81.8
1	86.1	82.6
2	85.5	82.0
5	85.5	81.8

及验证集和测试集中的所有样本(不包含标签)来对所提议模型进行训练,根据验证集分割正确率选择最优网络权重。测试集结果对比如表5所示,可视化的分割结果如图6所示。

由实验结果可看出,所提议模型的结果明显好于其它对比算法。我们考虑,这不仅是由于引入的全卷积辅助对抗网络能够迫使分割结果保持标签连续性,并且所提出的基于注意力选择的对抗损失可以使判别器在分割结果较好时仍能有效约束分割网络的更新方向;更是由于所提议的半监督模型在不使用任何额外标签的前提下,可以提取验证集和测试集中的图像样本的特征,从而更好的学习数据集的整体样本分布,同时能够更充分的利用数据集所包含的信息来改善分割结果。这体现了本文所提议的半监督模型在高分辨率遥感图像语义分割的实际应用价值,即只需要少量的样本标签即可得到令人满意的分割结果,从而可以缓解样本标注所带来的大量消耗。

表5　与其它性能优异方法的测试集结果对比

Table 5　Accuracy comparison on test dataset among the proposed method and other state-of-the-art methods

Method	Imp Surf	Building	Low_veg	Tree	Car	OA	Mean F1
UPB[9]	87.5	89.3	77.3	85.8	77.1	85.1	83.4
ETH_C[25]	87.2	92.0	77.5	87.1	54.4	85.9	79.6
CAS_Y3[6]	89.6	91.5	82.0	88.3	68.4	87.8	84.0
ITC_B2[7]	90.1	93.5	82.1	88.3	77.1	88.4	86.2
VNU4[26]	91.2	93.6	81.5	88.5	77.7	89.0	86.5
CASZX1[27]	91.3	93.9	81.9	88.3	77.6	89.0	86.6
UFMG_3[28]	90.7	94.3	82.5	88.5	77.4	89.0	86.7
所提议方法	92.7	95.1	84.3	89.4	86.2	90.6	89.5

4　结论

本文提出了一种基于生成式对抗网络的针对高空间分辨率遥感图像的半监督语义分割方法。

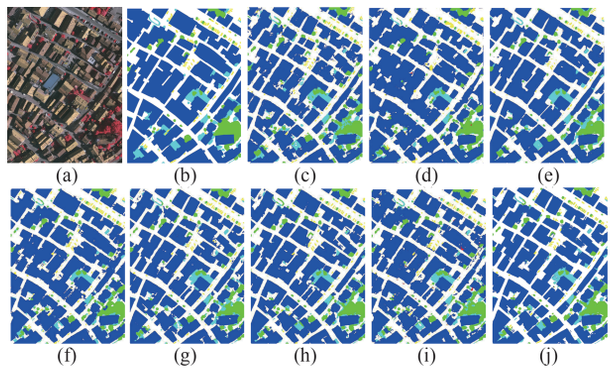

图6　所提议模型与其它性能优异的遥感图像语义分割模型的测试集分割结果可视化对比 (a) 待分割图像,(b) 真实标签图,(c) UPB,(d) ETH_C,(e) CAS_Y3,(f) ITC_B2,(g) VNU4,(h) CASZX1,(i) UFMG_3,(j) 所提议方法

Fig. 6　Visual comparison of segmentation results among the proposed method and other state-of-the-art models on test set: (a) image for segmentation, (b) ground truth label map, (c) UPB, (d) ETH_C, (e) CAS_Y3, (f) ITC_B2 (g) VNU4, (h) CASZX1, (i) UFMG_3, and (j) the proposed method

所提议方法能够充分学习和利用图像的空间特征,以缓解高分辨率遥感图像光谱特征类内差异大、类间差异小的问题。首先,通过构造全卷积的辅助对抗网络,一方面可以将无标签样本用于对模型的训练,实现半监督语义分割;另一方面,针对遥感图像特点设计的端到端全卷积对抗网络能够捕捉遥感图像的局部结构信息,有助于保持分割结果中的标签连续性,提高分割精度。更进一步,本文提出了一种新颖的对抗损失,通过自适应的调整难易样本损失在整体损失中的权重,可实现注意力选择,即在训练中更加关注困难样本部分,从而解决分割结果较好时判别器难以对分割网络的更新进行有效约束的问题。实验结果表明,所提出的方法能够更好地学习到数据整体的分布特性,充分利用无标签样本所包含的信息来提升分割精度;更进一步,本文方法中所提出的基于注意力选择的对抗损失能够使分割准确率得到进一步提高。与现有的其它高分辨率遥感图像语义分割方法相比,本文所提出的方法可大大降低对数据标注的需求,并且具有更好的半监督图像语义分割结果,这将在实际应用中具有较为重要的实际意义。

另外,由于本文所使用的基准模型网络层数仍较深,导致整体网络训练速度较为缓慢,在实际应用中存在效率低下的问题。最近的研究在提出一些能力较强且训练速度较快的轻量级网络,我们考虑将进一步改进所提议方法使用的基础模型,将轻

量级网络应用于我们的基准模型中,以降低计算耗时,使所提议模型可以高效率地应用于实际中。另外,我们也考虑进一步尝试将所提议方法应用于高光谱遥感图像的语义分割问题中。

References

[1] Mehmet S, Bulent S. Survey over image thresholding techniques and quantitative performance evaluation [J]. *Journal of Electronic Imaging*, 2004, **13**(1): 146-165.

[2] Pesaresi M, Benediktsson J A. A new approach for the morphological segmentation of high-resolution satellite imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, **39**(2): 309-320.

[3] Awad M M, Nasri A. Satellite image segmentation using Self-Organizing Maps and Fuzzy C-Means [C]. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, 2009.

[4] Schmidhuber J. Deep learning in neural networks: An overview [J]. *Neural Networks*, 2015, **61**: 85-117.

[5] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks [C]. International Conference on Neural Information Processing Systems, 2012.

[6] Bo Y, Lu Y, Fang C. Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018: **11**(9): 3252 - 3261.

[7] Bergado J R, Persello C, Stein A. Recurrent multiresolution convolutional networks for vhr image classification [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2018, **56**(11): 6361 - 6374.

[8] Chen G Z, Zhang X D, Wang Q, *et al*. Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, **11**(5):1633-1644.

[9] Marcu A, Leordeanu M. Dual Local-global contextual pathways for recognition in aerial imagery [J/OL]. 2016. https://arxiv.org/abs/1605.05462

[10] Pan X R, Gao L R, Andrea M, *et al*. Semantic labeling of high resolution aerial imagery and LiDAR data with fine segmentation network [J]. *Remote Sensing*, 2018, **10**(5): 743-766.

[11] Marmanis D, Datcu M, Esch T, *et al*. Deep learning earth observation classification using imagenet pretrained networks [J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, **13**(1): 105 - 109.

[12] Marmanis D, Wegner J D, Galliani S, *et al*. Semantic segmentation of aerial images with an ensemble of CNNs [J]. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, **3**: 473 - 480.

[13] Marmanis D, Schindler K, Wegner J D, *et al*. Classification with an edge: Improving semantic image segmentation with boundary detection[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, **135**: 158-172.

[14] Philipp K, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C]. International Conference on Neural Information Processing Systems, 2011.

[15] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets [C]. International Conference on Neural Information Processing Systems, 2014.

[16] Luc P, Couprie C, Chintala S, *et al*. Semantic segmentation using adversarial networks [C]. International Conference on Neural Information Processing Systems, 2016.

[17] Hung W C, Tsai Y H, Liou Y T, *et al*. Adversarial learning for semi-supervised semantic segmentation [C]. The British Machine Vision Conference(BMVC), 2018.

[18] Kohli P, L'ubor L, Torr P H S. Robust higher order potentials for enforcing label consistency [J]. *International Journal of Computer Vision*, 2009, **82**(3):302-324.

[19] Souly N, Spampinato C, Shah M. Semi supervised semantic segmentation using generative adversarial network [C], IEEE International Conference on Computer Vision (ICCV), 2017.

[20] 2D Semantic labeling contest, International Society for Photogrammetry and Remote Sensing (ISPRS) Working Group II/4, *Hannover*, *Germany*, 2014

[21] He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), Las Vegas, NV, 2016.

[22] Chen L C, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2016, **40**(4):834-848.

[23] Andrew L M, Awni Y H, Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models [C]. International Conference on Machine Learning (ICML), 2013.

[24] Lin T Y, Goyal P, Girshick R, *et al*. Focal loss for dense object detection [C], IEEE Conference on Computer Vision & Pattern Recognition(CVPR), 2017.

[25] Tschannen M, Cavigelli L, Mentzer F, *et al*. Deep structured features for semantic segmentation [C]. European Signal Processing Conference(EUSIPCO), Kos, 2017.

[26] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, **39**(4): 640-651.

[27] Chen L C, Zhu Y, Papandreou G, *et al*. Encoder-Decoder with atrous separable convolution for semantic image segmentation [C]. European Conference on Computer Vision(ECCV), 2018.

[28] Nogueira K, Mura M D, Chanussot J, *et al*. Dynamic multi-context segmentation of remote sensing images based on convolutional networks [J/OL]. 2018. https://arxiv.org/abs/1804.04020

[29] Zhao H, Shi J, Qi X, *et al*. Pyramid scene parsing networ [C]. IEEE Conference on Computer Vision & Pattern Recognition(CVPR), 2017.