

基于自注意和对抗学习的道路场景水体检测方法

王臣毅, 王欢, 孟策

(南京理工大学计算机科学与技术学院, 江苏南京 210094)

摘要: 无人驾驶车辆近年来一直是研究的热点. 无人车运行环境复杂、不确定因素多, 尤其当其意外驶入水坑、泥潭等地形时可能直接导致抛锚, 造成不可估量的损失, 因此水体检测对无人车的运行有着重要意义. 借助深度网络的强大学习能力, 本文首先将反射注意力单元和自注意力机制相结合, 并在(U shape Network, U-Net)模型基础上添加残差卷积块和上采样卷积模块, 得到了新的道路场景水体区域检测模型(U shape Network with Attention for Road, URA-Net), 该模型能够更好地捕捉特征依赖关系, 提高水体语义特征的代表能力. 进一步, 本文提出了一种基于双生成器对抗学习的训练模型(Redundant With Dual Generative Adversarial Network, RWD-GAN), 它对 URA-Net 稍做修改, 拓展成两个生成器, 通过在对抗网络框架下让生成器与鉴别器、生成器与生成器之间实现对抗学习, 促进不同网络模型之间的信息传递. 在公开数据集上的大量实验表明 URA-Net 达到了 87.18% 的 F1 指标, 而 RWD-GAN 模型能够进一步提高水体检测的精度, 使提升到了 88.54%, URA-Net 和 RWD-GAN 均超出有深度网络水体检测方法的性能表现.

关键词: 水体检测; 自注意机制; 对抗学习; 深度学习

基金项目: 国家自然科学基金(No.61703209)

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2023)08-2213-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210743

Water Puddle Detection Method for Road Scene Based on Self-Attention and Adversarial Learning

WANG Chen-yi, WANG Huan, MENG Ce

(School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094, China)

Abstract: There has been much interest in the research of self-driving cars. Yet the detection of potentially dangerous obstacles on road makes this investigation more challenging. Water puddles, a typical obstacle of this kind, could trap a self-driving car or even cause serious accidents. Therefore, detecting water puddles is of great importance. To this end, this paper propose a novel water puddle detection model, URA-net(U shape Network with Attention for Road). Building its backbone on U-net(U shape Network) with residual and upsampling blocks added, URA-net combines both the reflection attention units and self-attention units, which can better characterize the dependence among image features so as to improve the representative capability to locate water puddles. Furthermore, a two-generator conditional adversarial network RWD-GAN (Redundant With Dual Generative Adversarial Network) is proposed, where two URA-Nets with a minor revision become the two generators to facilitate the information interaction in the adversarial learning process between the generators and the discriminator, as well as between the two generators themselves. Experiments on the public water puddle dataset demonstration that URA-net achieves 87.18% measure, while RWD-GAN can further improve the accuracy of URA-net, pushing F1-score to 88.54%. Both URA-net and RWD-GAN outperforms the state-of-the-arts.

Key words: water puddle detection; self-attention; Adversarial learning; deep learning

Foundation Item(s): National Natural Science Foundation of China (No.61703209)

1 引言

无人驾驶车辆近年来成为研究的热点, 但无人车运行环境复杂、不确定因素多, 尤其当其意外驶入水坑、泥

潭等地形时可能直接导致抛锚, 造成不可估量的损失, 因此水体的检测对无人车的安全行驶有着重要意义.

传统的水体检测方法设计与其所使用的传感器种

类具有密切联系,所提取的特征包括水体的各种物理特性. Matthies 等人^[1]使用激光雷达检测是否存在水体并测量其深度. Zhao 等人^[2]从饱和度与亮度的比值,以及另外四个从灰度共生矩阵中提取出的特征值,一起构成五维特征向量,而后训练支持向量机(Support Vector Machine, SVM)用于水体检测,取得了较好的识别准确率. 尽管如此,这种方法不能适用于所有的场景. Kim 等人^[3]利用立体视觉的方法得到场景深度信息. Yao 等人^[4]将颜色纹理特征和深度信息结合,制定一组融合规则来消除水体误检. Rankin 等人^[5]开发了一组互补的检测器:一个基于天空的反射,用于远距离的水体检测;另一个用于近距离的水体检测. 其检测原理基于以下物理现象:从水体的前缘到后缘,饱和度与亮度的比值的变化的变化是均匀的,且和其它类型地形的特性不同. 除了可见光图像, Matties 等人^[1]还探讨了利用短波红外(Short wave Infrared, SWIR)和中波红外(Mid-wave Infrared, MWIR)进行水体检测的可能性. 水体对近红外光特别是短波红外的吸收非常强,但由于短波红外入射角的有效范围大约是 60° ,会严重受到表面反射的限制. 相比而言,雪和冰对波长 $1.4\mu\text{m}$ 的红外光吸收也很强,因此短波红外同样也可以用来检测水体. 同时,通过对水库的24 h监测, Matties 等人^[1]发现大型水体在夜间往往比周围其他地形剪度高,这意味着中波红外也具有在夜间识别水体的能力.

水体还具有一个特殊的物理特性,即反射自水体表面的光是部分线性偏振的,而且这些反射光的偏振相位相似度比起周围场景的其他物体来说更高. Xie 等人^[6]通过比较偏振角和偏振相位的相似度来检测水体. Nguyen 等人^[7]建模反射光的偏振、颜色变化和反射、方向角之间的关系,说明来自天空的光对于水体偏振特性的影响.

Santana 等人^[8]提出一个通过计算跨越多帧的光流的熵来检测视频序列中水体的模型,该方法适用于带有波纹的水体. 他们还提出了一种基于外观的标签传播算法以解决平静水体中缺少纹理信息的问题. Mettes 等人^[9]引入了一种基于视频中水体表面时空表现的混合描述符. Yao 等人将传统的机器学习方法和图像分割结合在一起,通过主动学习的方式,避免大量标记训练样本的工作,并通过基于 mean-shift 的图像分割方法,细化最终的分割结果.

随着深度学习的兴起,不少学者开始使用深度网络来解决水体检测问题. Han 等人^[10]使用带有反射注意力单元(Reflection Attention Unit, RAU)的全卷积神经网络检测道路水体,网络的骨架为采用 FCN-8s^[11]. RAU 旨在捕获天空和附近场景与它们在水体中倒影之间的关系,改善特征表示. 为了丰富该领域的研究;

Han 等人^[10]还发布了名为“Puddle-1000”的道路水体图像数据集. Wang 等人^[12]利用条件生成对抗网络(conditional Generative Adversarial Network, cGAN)^[13]来处理道路图像水体区域检测问题,进一步提升了检测准确率. 同时还修正了“Puddle-1000”数据中部分标注错误. 这些深度学习方法提供了新的水体检测思路,并表现出了较高的检测性能.

分析可知, Han 等人^[10]提出的 RAU 本质上是一种块操作,利用对称差分实现一种特征的自注意力. 而这一操作非常类似于自注意力机制. 为此,是否可以用自注意力机制来替代 RAU 呢? 经过本文的研究,答案是:自注意力机制并不能直接替换 RAU,但是自注意力机制可以与 RAU 结合,提供更好的水体检测性能,其主要原因是自注意力机制并不能刻意关注 RAU 所关注的垂直的对称特性,它更注重提供场景上下文的全局特征依赖. 为此,我们设计了(U shape Network with Attention for Road, URA-net)模型,它在改进(U shape Network, U-net)基础上联合使用了反射注意力单元和双重注意力架构设计. 相比现有的水体检测模型,该模型已经达到了较好的性能.

但观察发现, URA-net 虚警较高,即 precision 指标仍不理想. 一种有效抑制虚警的方法是扩大网络的感受野,从宏观上识别虚警区域. 为此,我们尝试扩大其感受野,但扩大感受野可能会导致水体的漏检,因为虚警和漏检本身就是一对难以调和的矛盾,为此,受 Wang 等人^[12]工作的启发,我们利用条件生成对抗网络来进一步优化模型. 我们将一个未作修改的 URA-net 模型作为生成器 G_1 ,另外再设计一个同样的 URA-net 模型,通过在其适当的位置使用膨胀卷积来扩大感受野,作为生成器 G_2 . 并且,利用虚警和漏检相对抗的特点,我们设计了双生成器单鉴别器的条件对抗生成网络框架(Redundant With Dual Generative Adversarial Network, RWD-GAN)来共同优化 G_1 和 G_2 . RWD-GAN 提供了解决水体问题的一种新思路,即将一个大的优化任务分解为两个独立优化的子任务,每个生成器负责单一目标(减少虚警或者减少漏检),在训练过程中,鉴别器起到媒介作用,让两个生成器共享任务信息,更好地实现优化总目标(低虚警且低漏检). 同时,两个生成器联合实现水体的检测也是一种集成学习的思想,实验也表明其性能比单一模型更好.

本文基于深度学习的方法提出一种以 U-net^[14]为骨架的 URA-net 网络,它结合了反射注意力单元以及自注意力机制,提高了检测准确率. 在此基础上又提出了基于双生成器对抗学习的网络模型 RWD-GAN,两个生成器都遵循 URA-net 模型,一个负责减少漏检,一个负责降低误检. 实验证明,对抗学习的方式,能更有效地推动两

个生成器之间的信息流动,更好地完成水体检测任务.

本文内容安排如下:第2节介绍结合注意力反射机制和自注意力机制的URA-Net网络的设计思路和模型结构;第3节介绍基于双生成器单鉴别器对抗学习RWD-GAN的网络模型;第4节实验部分分别给出了URA-Net网络和RWD-GAN网络的详细实验和结论;第5节总结全文工作.

2 联合反射注意力单元和自注意机制的道路水体区域检测方法

2.1 反射注意力单元

由于水体表面具有较高的折射率,且水体的外观随着天气情况、视角变化、周围环境的变化^[15]而变化,

如果仅使用传统的卷积神经网络骨架从水体图像本身提取特征进行水体检测,学习出的模型并不具有较好的泛化性. Han等人^[10]将水体反射模型近似为平面镜模型,提出具有反射注意力单元(Reflection Attention Unit, RAU)的全卷积网络(Fully Convolutional Network, FCN),旨在利用水体表面反射天空和附近场景的物理特性.

RAU的结构如图1所示. 对于每个原始输入大小为 $[h, w, c]$ 的特征图 I ,先通过均值池化操作将其转换为大小为 $[n, w/2, c]$ (n 取8或16)的特征图 P ;然后在竖直方向上对其进行切片操作得到 n 个大小为 $[1, w/2, c]$ 的子特征图 $P_i(i=1, 2, \dots, n)$;对每个子特征图进行上采样操作,还原为原始特征图的大小,并与 I 相减,得到的特征图在通道维度上相连,最终组成大小为 $[h, w, n \times c]$ 的特征图 T ;再对特征图 T 做卷积操作,得到输出特征图 O .

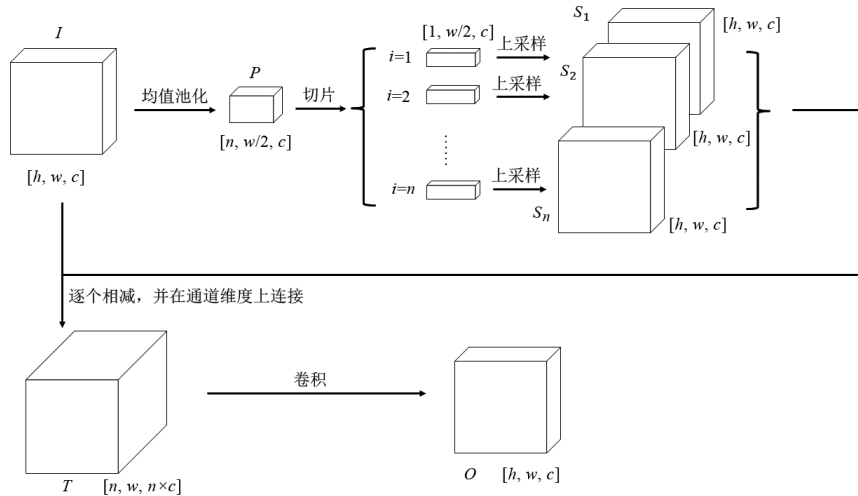


图1 RAU结构示意图

反射注意力单元的第一步是均值池化操作. 一方面,考虑到反射关系往往存在于竖直方向上,因此在竖直方向上将特征图划分为多个块状区域. 通过逐区域作差的方式,寻找特征之间的相似性. 另一方面,由于图像失真、车身颠簸等干扰因素,这种反射关系往往并非严格垂直,这种池化的操作也增强了对于误差的包容性. Han等人^[10]将RAU作为基本组件,插入FCN-8s的每卷积块之后,利用反射特性提取提取水体的鉴别性特征,取得了明显的效果.

2.2 自注意力机制

注意力机制最早被提出于20世纪90年代. 2014年, Mnih等人^[16]利用注意力机制进行图像分类,逐渐受到学者们的关注. 2017年, Vaswani等人^[19]提出了Transformer模型,在自然语言处理领域获得了巨大成功,其中的自注意力机制是其核心模块,为模型性能提升提供了重要保障.

在神经网络中,注意力机制通常是作为一个额外

的神经网络模块,能够选择输入的某些部分,或者给输入的不同部分分配不同的权重. 注意力机制能够从大量信息中筛选出重要的信息.

对于输入向量 $\mathbf{X} \in \mathbf{R}^{C \times N}$, 分别乘上权重矩阵 \mathbf{W}_f 和 \mathbf{W}_g , 对于 \mathbf{X} 的第 i 个列向量 \mathbf{x}_i , 分别得到 $f(\mathbf{x}_i) = \mathbf{W}_f \mathbf{x}_i$ 和 $g(\mathbf{x}_i) = \mathbf{W}_g \mathbf{x}_i$. 矩阵 \mathbf{S} 代表 $f(\mathbf{X})$ 和 $g(\mathbf{X})$ 之间的注意力图, 元素 s_{ji} 代表 j 个特征图上, 位置 i 的响应程度, s_{ji} 可以通过如下公式计算得到:

$$\beta_{ij} = f(\mathbf{x}_i)^T g(\mathbf{x}_j) \quad (1)$$

$$s_{ji} = \frac{\exp(\beta_{ij})}{\sum_{i=1}^N \exp(\beta_{ij})} \quad (2)$$

其中, C 代表特征通道维数, N 表示隐藏层的结点数, 中间向量 $\mathbf{O} = (o_1, o_2, \dots, o_n) \in \mathbf{R}^{C \times N}$, 其中第 j 个激活值 o_j 可以通过如下公式计算得到:

$$h(\mathbf{x}_i) = \mathbf{W}_h \mathbf{x}_i \quad (3)$$

$$v(x_i) = W_v x_i \quad (4)$$

$$o_j = v\left(\sum_{i=1}^N s_{j,i} h(x_i)\right) \quad (5)$$

以上公式中, $W_f, W_g, W_h \in \mathbf{R}^{\bar{C} \times C}, W_v \in \mathbf{R}^{C \times \bar{C}}$, 这四个权重矩阵均代表卷积核大小为 1×1 的卷积操作. 为了减少计算量, 将 \bar{C} 取为 $C/4$.

最后, 输出向量 $Y \in \mathbf{R}^{C \times N}$ 通过如下公式计算得到:

$$Y = \gamma O + X \quad (6)$$

其中, γ 是一个可学习的变量, 初始化为 0, 目的是先学习简单的任务, 然后逐步增加问题的复杂性.

2.3 URA-Net 网络

2.3.1 将自注意力机制用于水体检测

自注意力机制与本文所研究的道路场景图像水体区域检测问题的特性之间有密切的联系. 如图 2 所示, 在典型的道路水体图像中, 水体区域是不连续、不规则的, 分布在图像上的各个区域, 仅依靠堆叠卷积核扩大感受野的方法很难有效提取水体特征, 因此自注意力机制很适合用在道路场景图像水体区域检测问题中.

同时对于前述提到的水体区域的反射特性, Han 等



图 2 典型道路水体图像

人^[10]和 Wang 等人^[12]都针对这个特点做了大量的研究. 实验表明, 他们所使用的反射注意力单元(Reflection Attention Unit, RAU)使道路场景图像水体区域检测的效果有一定提升. 因此本文考虑将反射注意力单元和自注意力机制有机结合, 提高水体检测效果.

2.3.2 网络概述

本文将提出的模型命名为 URA-Net. 网络架构如图 3 所示, 其总体结构以 U-Net 为骨架, U-Net 架构的优点在于同时利用了来自底层和高层的图像特征, 保证恢复出来的分割结果较为精细. 网络由下采样分支、自注意力机制和上采样分支组成. 经典的 U-Net 网络中, 每个下采样卷积层仅由两个卷积核大小为 3×3 的卷积操作串联而成. 不同于经典的 U-Net 网络, URA-Net 将下采样卷积块改进为残差连接, 同时, 在每个下采样卷积块之后接上 RAU 以提取周围环境与水体反射之间的特征映射关系, 之后执行窗口大小为 2×2 的最大池化操作, 重复上述操作若干次, 由大小为 360×640 的输入图像得到维度为 $[23, 40, 1024]$ 的特征图 X. 对于特征 X, 通过自注意力模块提取特征之间的远程依赖性, 优化特征表示. 最后, 依次级联底层特征, 通过上采样块将中间特征表示恢复至原始输入大小.

2.3.3 残差块和上采样块

残差块和上采样块的结构如图 4 所示. 不同于传统 U-Net 网络将两个卷积核大小为 3×3 的卷积操作叠加在一起形成卷积块, 本文采用残差连接的方式重新组成卷积块, 对于任意输入特征 X, 输出特征 $H(X)$ 可以通过如下公式得到:

$$H(X) = F(X, W_i) + W_s X \quad (7)$$

其中, W_s 和 W_i 分别代表不同的权重矩阵, $F(\cdot, \cdot)$ 代表特征之间的某种映射关系. 残差连接的方式可以解决深

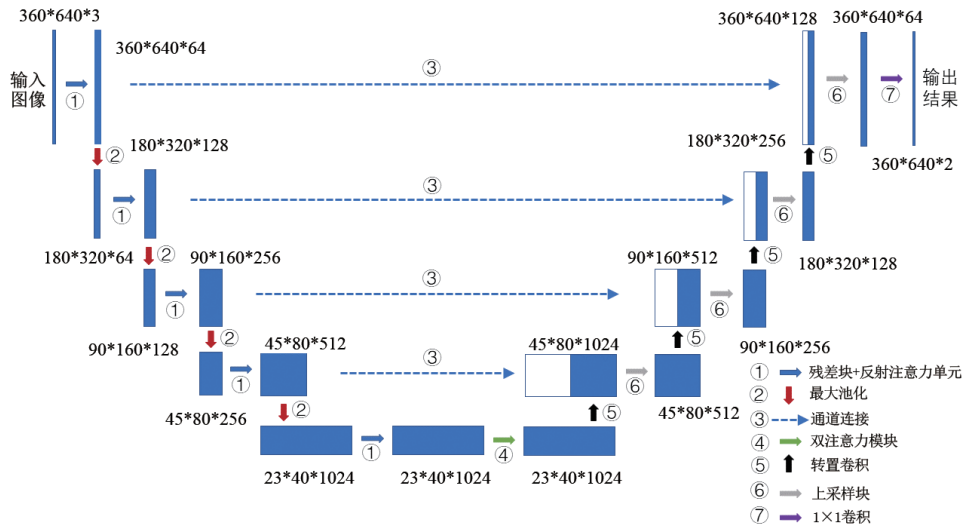


图 3 网络总体架构

度神经网络因为网络层数叠加所导致的模型“退化”问题^[18],对于潜在的期望映射 $H(\mathbf{X})$,与其直接学习,不如通过两个映射 \mathbf{W}_s 和 \mathbf{W}_i 拟合得到,这样避免了梯度“爆炸”的问题,降低了网络优化的复杂度.

对于上采样块而言,先通过转置卷积将特征图的长和宽还原为原特征图的两倍,然后通过“skip-connections”的方式级联底层特征,最后通过两个卷积核大小为 3×3 的卷积得到输出特征图.

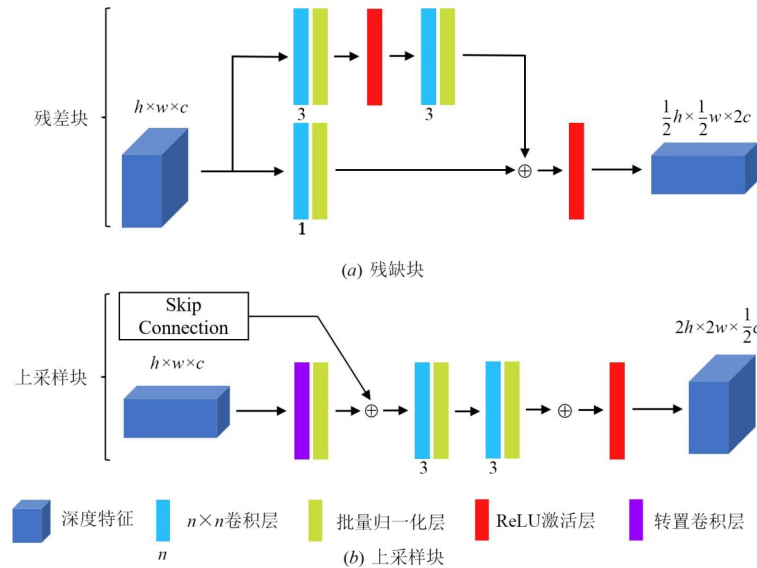


图4 残差块(上)和上采样块(下)

2.3.4 双重注意力模块

给定道路场景下的水体图像,由于周围环境、光照、距离、视角等方面的不同,水体所呈现的特征也具有不确定性.卷积操作的感受野往往较为局部,因此具有相同标签的水体像素所对应的特征可能会存在差异,这些差异增加了类内的一致性,进而影响水体检测的准确性.为此,本文引入注意力机制自适应地聚合远程上下文信息来改善分割结果^[19],它包含位置注意力和通道注意力两部分,下文称双重注意力模块.

双重注意力模块的结构如图5所示,由两部分组成,即位置注意力模块和通道注意力模块.二者采用并行的连接方式,在得到两个模块的输出特征后,执行特征融合.

对于位置注意力模块,通常的做法是,使用两个不同的 1×1 卷积核,生成新的特征图 $f(\mathbf{X})$ 和 $g(\mathbf{X})$,其中 $\{f(\mathbf{X}),g(\mathbf{X})\}\in \mathbf{R}^{C\times H\times W}$,然后将其重组为 $\mathbf{R}^{C\times N}$,其中 $N=H\times W$,为特征图像素数.随后,对 $f(\mathbf{X})$ 和 $g(\mathbf{X})^T$ 执行矩阵乘法,并应用softmax操作得到大小为 $[H\times W,H\times W]$ 的位置注意力矩阵 \mathbf{S} .同时,使用另一个 1×1 的卷积核生成特征图 $h(\mathbf{X})$.通过重组、矩阵乘法等操作,得到特征 \mathbf{O} ,最后和原特征图加权相加得到最终的特征 \mathbf{Y} .

高层特征图包含复杂的语义响应关系,通过利用通道映射之间的相互依赖关系,可以强调相互依赖的特征映射并且改善特定语义的特征表示.与位置注意力模块不同,输入特征 \mathbf{X} 被直接用来计算得到通道注

意力矩阵 \mathbf{S}' ,目的是保留不同通道特征图之间的关系.具体来说,将 \mathbf{X} 的维数重组为 $\mathbf{R}^{C\times N}$,然后在 \mathbf{X} 和 \mathbf{X}^T 之间执行矩阵乘法,并应用softmax操作得到通道注意力矩阵 \mathbf{S}' .特征图 \mathbf{X} 和通道注意力矩阵 \mathbf{S}' 执行矩阵乘法,并与原特征图加权相加得到最终的特征 \mathbf{Y}' .最后,用卷积模块转换两个注意力模块的输出,执行像素级别的特征融合.

3 双生成器对抗学习

为了能进一步提高水体检测模型的精度和稳定性,Wang等人^[12]研究了用cGAN学习框架来提升水体检测模型的效果,但性能提升较为有限.此外,一些学者也尝试使用传统的cGAN学习框架来提升语义分割问题,也只能获得少许的提升,或者无明显提升.分析其原因,这些方法都是将目标分割或语义分割模型直接作为生成器,模型架构不再变化,只是去设计合理的鉴别器即可.这本质上相当于只在生成器模型的损失函数中增加了对抗损失项,而相比交叉熵损失,这一对抗损失起不到主要作用.我们认为需要尝试新的cGAN框架以期获得更好的学习性能.此外,需要优化的水体分割模型在被当作生成器时,不一定一层不变,可能也需要在结构上做略微的调整.为此,我们提出了一种双生成器对抗学习框架.

3.1 双生成器条件生成对抗网络

生成对抗网络(Generative Adversarial Networks,

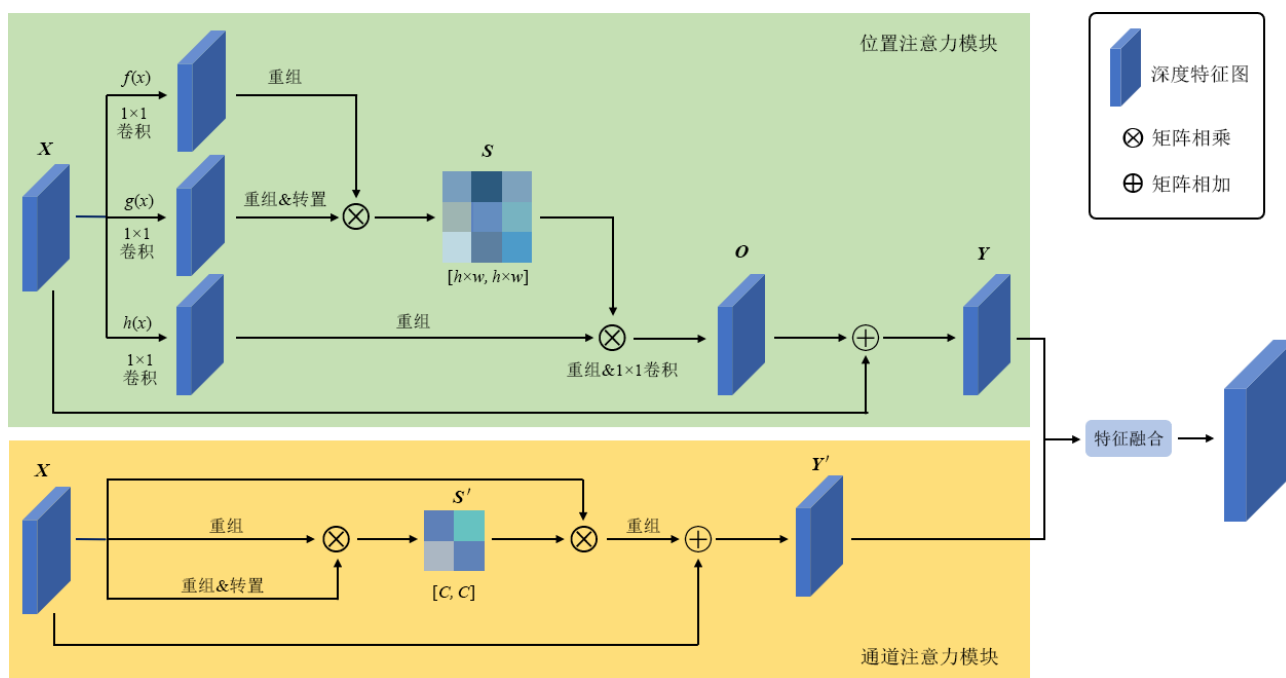


图5 双重注意力模块

GAN)于2014年由Goodfellow等人^[20]首次提出,并迅速发展,已经被广泛应用于图像生成、图像风格转换、图像识别、目标检测、文本生成图像、领域自适应、图像复原等领域。其主要思想是通过一个生成器与鉴别器的对抗来同步学习互相提升能力,最终的目标是获得一个高性能的生成器。之后文献[13]在GAN的基础上加上额外的约束条件提出了条件生成对抗网络(cGAN),使训练最终得到的生成能够满足特定的任务要求。

本节用cGAN的基本思想,针对第2.3节提出的道路水体检测模型,采用对抗学习的机制,提升水体检测效果。我们将本节提出的模型命名为RWD-GAN。其与一般的条件生成对抗网络的不同之处在于:

(1) 一般的条件生成对抗网络由一个生成器和一个鉴别器组成,二者相互对抗完成学习任务,达到平衡状态。然而,本节提出的网络具有两个生成器和一个鉴别器,在生成器和鉴别器进行对抗学习的同时,让两个生成器也进行对抗。生成器中的一支负责抑制误检,另一支负责降低漏检。两个生成器通过鉴别器传递信息,实现一种高检测率和低虚警率的对抗。

(2) 双生成器架构也可以看作一种集成学习的手段。集成学习通过构建多个学习器来完成学习任务,分为训练个体学习器和集成结果这两个步骤。深度神经网络模型复杂,参数众多,在实际应用中由于硬件设备的限制,难以使用多个网络模型直接进行集成学习。因此,我们通过对抗学习的方式,端到端地将两个独立且胜任的水体检测模型集成起来。

3.2 网络架构

3.2.1 网络概述

如图6所示,RWD-GAN由生成器和鉴别器组成,但是与cGAN不同的是,它具有两个生成器,分别表示为 G_1 和 G_2 ,以及一个鉴别器,表示为 D 。每个生成器都有自己独立的输出结果 O_1 和 O_2 。生成器 G_1 的映射可以表示为 $G_1(I; \theta_1)$,生成器 G_2 的映射可以表示为 $G_2(I; \theta_2)$,其中 I 表示输入图像。鉴别器 D 的输入由三部分组成,分别是 G_1 的输出 O_1 、 G_2 的输出 O_2 以及图像真实分割结果掩模 T 。这三个输入均以原图像 I 为条件,在通道维度上连接。此时,鉴别器 D 的任务不再是一个二分类问题,而是判断输入是生成器的 G_1 输出、生成器 G_2 的输出,还是真实结果掩模。

G_1 和 G_2 的基本结构仍然遵循本文在第2.3节提出的设置。不同的是,在 G_2 中,为了扩大下采样过程的感受野,保留更多细节,我们删除了最后一个下采样操作,并在最后两个下采样块中分别使用了膨胀率为2和4的空洞卷积。因此得到了尺寸更大的中间特征表示。对于水体精确细节的检测,倾向于使用小的感受野;相反,保持水体检测结果的一致性需要更多的上下文信息,倾向于使用较大的感受野。因此 G_1 的目标是使检测结果保留尽可能多的细节,减少漏检; G_2 的目标则是保持检测结果的一致性,降低误检。对于鉴别器网络,本文使用卷积神经网络(Convolutional Neural Network, CNN),它包括4个卷积核大小为 5×5 、步长大小为 2×2 的卷积操作和一个输出维度为3的全连接层,最后通过softmax函数得到激活值。

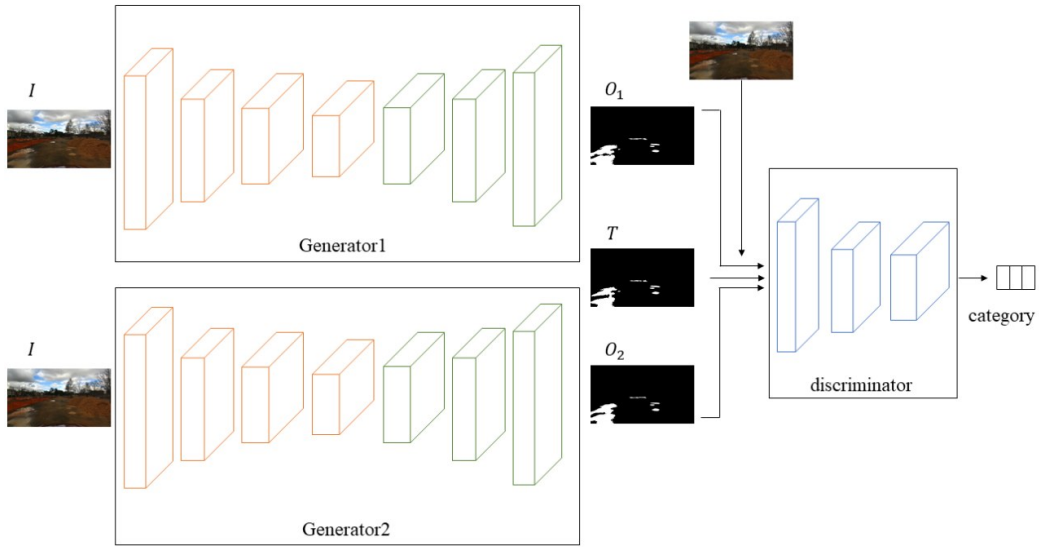


图6 网络架构示意图

直观地讲,我们可以分别训练两个生成器网络,然后以某种方式融合分割结果,比如 Nguyen 等人^[21]在处理阴影分割问题所采取的策略.但是,这种在训练完两个独立的网络之后融合结果的策略阻止了两个生成器网络之间的信息共享,这将会影响最终的分割结果.然而,在本文所提出的模型中,两个网络之间的信息流动是通过鉴别器实现的.具体来说,对抗机制的存在,迫使它们在拟合正确分布的同时,输出能够“欺骗”鉴别器的结果,具有三个类别的对抗损失增强了降低误检的能力,同时也增强了减少漏检的能力.通过这一过程,两个生成器将共同生成更加准确的分割结果.

3.2.2 损失函数设计

本文提出的 RWD-GAN 模型的损失函数包括以下三个部分:生成器 G_1 的数据损失、生成器 G_2 的数据损失以及对抗损失.它们与一般的 GAN 不同,本文使用了双生成器,因此对抗损失由三个目标项组成.

$$L_{GAN}(G_1, G_2, D, I) = E_{I,T} [\log D(I, T)] + E_I [\log(1 - D(I, G_1(I)))] + E_I [\log(1 - D(I, G_2(I)))] \quad (8)$$

其中, $E[\cdot]$ 表示期望操作,表示计算一个 mini-batch 中的所有样本对括号内函数值的平均值.第一个损失项对应于鉴别器网络,最后两个损失项对应于两个生成器网络.当固定鉴别器的参数时,最小化该损失函数将鼓励和生成与真实掩模相似的输出结果,当固定生成器的参数时,最小化该损失函数将提高鉴别器区分这些分割结果的能力.

本文使用交叉熵函数作为生成器输出的结果的损失函数.故生成器 G_1 和 G_2 的损失函数分别如式(9)和

式(10)所示:

$$L_{data}(G_1, T) = E_I [-\log(G_1(I))] + E_I [-\log(1 - G_1(I))] \quad (9)$$

$$L_{data}(G_2, T) = E_I [-\log(G_2(I))] + E_I [-\log(1 - G_2(I))] \quad (10)$$

本文所提出的 RWD-GAN 完整目标如式(11)所示:

$$(G_1^*, G_2^*, D^*) = \arg \min_{G_1, G_2} \max_D (\alpha_1 L_{data} + \alpha_2 L_{GAN}) \quad (11)$$

其中, α_1 和 α_2 是用于平衡数据损失和对抗损失关系的系数.

4 实验

4.1 实验数据集及设置

本文使用的数据集均为“Puddle-1000”.“Puddle-1000”数据集是由 Han 等人^[10]提供的道路水体图像数据集,共有 985 张图像,分为两个子集:柏油道路水体子集(OnRoad)和越野道路水体子集(OffRoad).OnRoad 数据集中,有 357 张图像,其中 272 张图像用于训练,85 张图像用于测试;OffRoad 数据集中,有 628 张图像,其中 530 张图像用于训练,98 张图像用于测试.所有图像都具有相应的标注掩模.Wang 等人^[12]修正了部分错误的标注,以保证不同图像之间的标注具有一致性.本文使用了 Wang 等人^[12]修正后的数据集.

OnRoad 数据集中,水体大多分布在柏油道路两侧且较为浑浊,反射主要来自天空,少部分来自路桩和指示牌;OffRoad 数据集中,水面反射天空、树木、围栏等景物.数据集的整体场景较为复杂,天空有较多云层分布,光照条件不佳,因此在该数据集上进行水体检测具有一定的挑战性.

本节的实验使用 tensorflow 框架实现,所使用的硬

件平台具有 3.7 GHz 的 CPU, 64 GB 的内存, 配备显存大小为 12 GB 的 Nvidia TITAN Xp 显卡. 本文使用交叉熵损失函数作为分割结果的损失函数, 使用 Adam 优化算法进行优化, 初始学习率为 2×10^{-4} , 一阶矩估计的指数衰减率为 0.9, 二阶矩估计的指数衰减率为 0.999. 由于显存大小的限制, 输入的批大小设置为 1. 输入图像的尺寸大小是 360×640 , 整个训练过程在 10^5 次迭代后终止.

4.2 URA-Net 网络实验

4.2.1 实验结果

目前现有的道路场景水体区域检测方法包括 Wang 等人^[10]提出的 FCN-8s-FL-RAU 模型和 Han 等人^[12]提出的 cGAN-mRAU 模型, 以及 Nguyen 等人^[7]提出的基于光的偏振的方法 GMM&polarisation 和 OFFSEG^[25]. 其中, OFFSEG 实现时分别使用了两个通用的骨架网络 HRnet^[28]和 BiSEnet^[29], 我们对两种不同骨架的 OFFSEG 模型均作了实验比较. 同时, 该方法是多类语义分割网络, 类别中除水体类别外, 还有其他类别. 故我们将其结果进行了合并, 将水体目标作为一类, 而其它类别(非水体类别)作为背景类别.

此外, 我们还增加了 DeepLab 模型^[22], 以及以 ResNet 模型为骨架的用在典型分割和检测任务上的方法的效果比较, 这些模型包括 res50_upernet^[26]和 res50_fpn^[27], 以及它们的结合膨胀卷积的改进模型 res50d_upernet 和 res50d_fpn. 同时, 我们尽量对这些模

型进行优化, 使其更好地检测水体, 主要包括: (1) 加入膨胀卷积和特征金字塔模块; (2) 设计模型的参数量与本文方法相当, 以保证同样数量的训练数据对其它模型也是足够的; (3) 实现这些模型时, 使用更适配的参数设置, 比如学习率.

本文将 URA-Net 模型与上述方法在数据集上的结果作为对比, 定量实验结果如表 1 所示. 部分水体检测可视化结果对比见图 7.

从表 1 中可以看出, 无论在 On Road, off Road 还是在二者的并集 Both Road 数据集上, 本文提出的方法在各项指标中都显示出了较优越的性能. 这得益于 RAU 在提取水体反射特征方面的强大能力以及注意力机制在捕捉特征远程依赖关系方面的有效性.

4.2.2 消融实验

为了验证基础架构、RAU 以及双重注意力模块等改进的有效性, 本文做了多个消融实验. 首先, 验证了 U-Net 结构和残差卷积块的有效性, 对比对象为 FCN-8s-FL-RAU 和 cGAN-mRAU. 其结果如表 2 所示, Backbone 代表本文的主干网络, 即去除了 RAU 和双重注意力模块后剩下的部分, 可以看到在没有添加其他任何组件的情况下, 依靠主干网络本身已经在 Off Road 数据集和 Both Road 数据集上得到了最优异的性能表现, 这证明了本文提出的主干网络架构的有效性.

遵循 Wang 等人^[10]和 Han 等人^[12]的实验设定, 本文使用 RAU 提取水体反射与周围环境之间的映射关系,

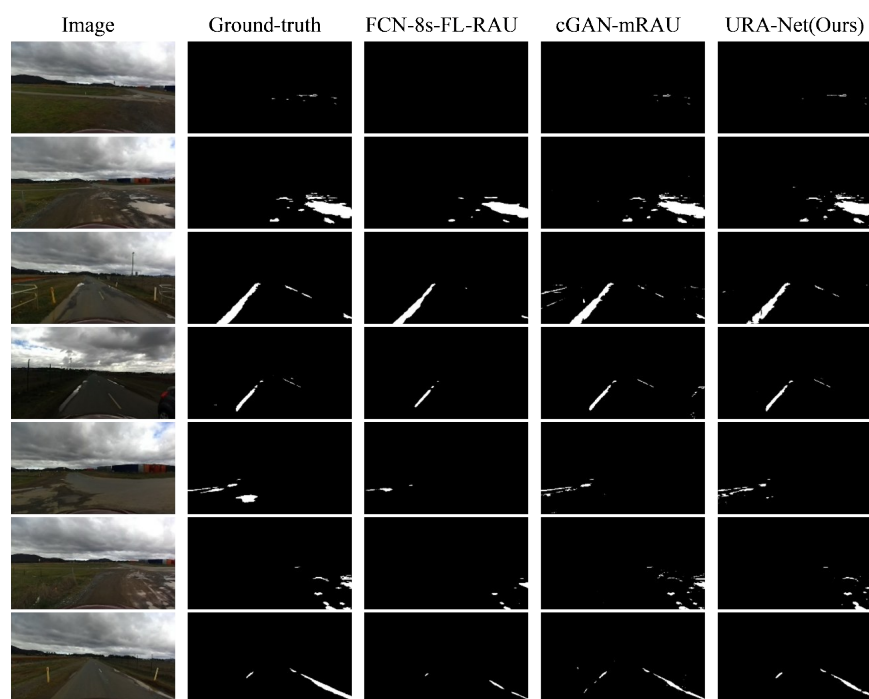


图7 不同深度学习网络的水体检测结果对比图

表 1 与现有道路场景水体区域检测方法的比较

Dataset	Model	Accuracy/%	F1-measure/%
On Road	GMM&polarisation	96.50	31.00
	DeepLab	98.83	21.97
	FCN-8s-FL-RAU	99.35	70.11
	cGAN-mRAU	99.72	87.33
	HRnet	99.56	87.08
	BiSEnet	99.49	83.67
	res50_upernet	99.52	85.77
	res50d_upernet	99.54	85.69
	res50_fpn	99.52	84.79
	res50d_fpn	99.55	85.93
	URA-Net(Ours)	99.73	87.83
	Off Road	GMM&polarisation	95.20
DeepLab		98.56	45.05
FCN-8s-FL-RAU		99.38	81.67
cGAN-mRAU		99.50	84.29
HRnet		99.54	86.24
BiSEnet		99.47	83.96
res50_upernet		99.54	86.27
res50d_upernet		99.53	85.48
res50_fpn		99.55	85.91
res50d_fpn		99.53	85.47
URA-Net(Ours)	99.56	86.80	
Both Road	GMM&polarisation	95.87	29.45
	DeepLab	98.59	40.36
	FCN-8s-FL-RAU	99.34	76.91
	cGAN-mRAU	99.54	82.82
	HRnet	99.61	86.78
	BiSEnet	99.58	84.36
	res50_upernet	99.62	86.05
	res50d_upernet	99.63	86.75
	res50_fpn	99.60	85.35
	res50d_fpn	99.61	85.79
	URA-Net(Ours)	99.64	87.18

为了验证RAU的有效性,在我们BackBone的基础上增加了RAU,其结果如表3所示.从表3中可以看出,无论是On Road, Off Road还是Both Road数据集上,本文提出的URA-Net在指标上的提升明显,均取得了最高的F1-Measure.在OnRoad数据集上,相比不添加任何组件的Backbone提升了0.99%,相比仅添加RAU的网络提升了0.87%,相比仅添加双重注意力模块的网络提升了1.34%.在Off Road数据集上,相比不添加任何组件的Backbone提升了1.07%,相比仅添加RAU的网络提升了0.2%,相比仅添加双重注意力模块的网络,提升了1.47%.在Both Road数据集上,相比不添加任何组件的Backbone提升了0.66%,相比仅添加RAU的网络提升

表 2 关于URA-Net主干网络的性能对比

Dataset	Methods	F1-measure/%
On Road	FCN-8s-FL-RAU	70.11
	cGAN-mRAU	87.33
	Unet	86.15
	BackBone(URA_net)	86.84
Off Road	FCN-8s-FL-RAU	81.67
	cGAN-mRAU	84.29
	Unet	85.11
	BackBone(URA_net)	85.73
Both Road	FCN-8s-FL-RAU	76.91
	cGAN-mRAU	82.82
	Unet	86.05
	BackBone(URA_net)	86.52

表 3 RAU 消融实验

Dataset	RAU	Attention	F1-measure/%
On Road	×	×	86.84
	√	×	86.96
	×	√	86.49
	√	√	87.83
Off Road	×	×	85.73
	√	×	86.60
	×	√	85.33
	√	√	86.80
BothR Road	×	×	86.52
	√	×	86.86
	×	√	86.13
	√	√	87.18

了0.32%,相比仅添加双重注意力模块的网络提升了1.05%.

从结果来看,RAU对道路场景图像水体区域检测问题具有一定的作用,其原因在于:RAU将原始特征图按照区域采样,每个区域的特征均与其他区域特征作差,将这些特征图在通道维度上连接,最后通过卷积将信息融合,这种方式寻找到了周围景物的反射和水体表面的映射关系,避免了光照、拍摄角度、相机畸变、反射周围物体等导致的特征不适宜问题.图8给出了RAU操作的可视化结果.先将输入图像灰度化,然后将RAU的预处理部分(去掉最后的卷积操作)作用在输入图像上,参数 n 取为16.原图中的水体部分已经用红色方框圈出,可以看到,在编号为1到7的特征图中,与其余各区域作差的目标区域是天空.水体部分特征均呈现较暗的视觉特征,而周围的道路部分则呈现较亮的视觉特征,这是由于水体部分反射了天空中的云,作差得到的差值较小,因此亮度偏低;而原本周围较暗的路面,其与云的差值较

大,因此亮度偏高.在编号为8到16的特征图中,与各其余区域作差的目标区域是路面,因此此时水体呈现相反的视觉特征,即呈现较亮的视觉特征,而路面

则呈现较暗的视觉特征.由此可以看出RAU在增强局部对比度、定位水体位置、学习水体特征等方面的作用.

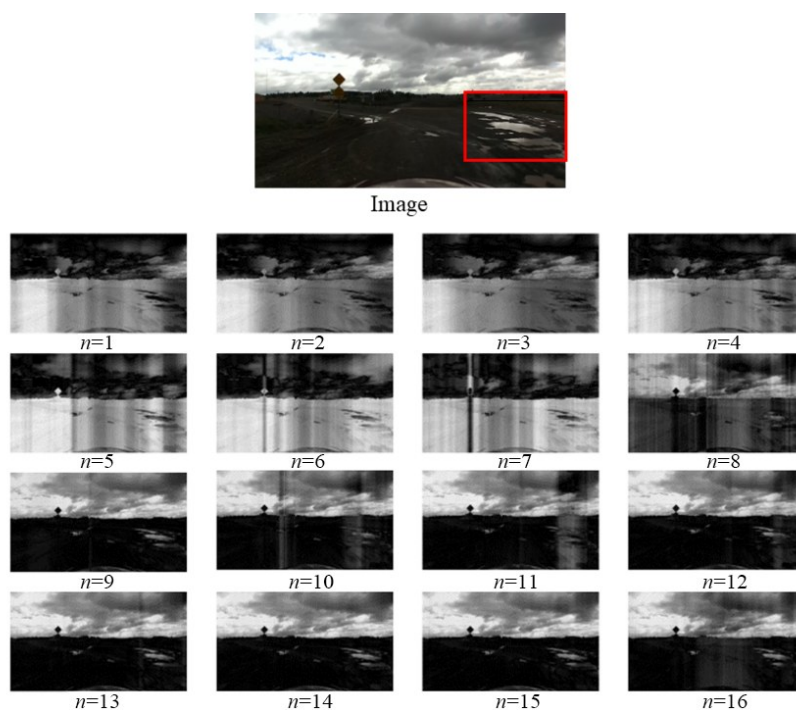


图8 RAU操作的可视化结果

然而,与直觉相反的是:直接在Backbone上使用双重注意力模块,其性能有一定的下降.这在三个数据集上都有所体现,只有在联合使用RAU和双重注意力模块时,才能体现出显著的性能提升.对于该现象,我们认为:如果单纯使用双重注意力机制,水体像素和与其特征相似的非水体像素之间会建立起远程依赖关系,这对道路水体检测任务会造成不良影响.相反,如果在双重注意力模块之间添加多个RAU处理单元,就可以准确定位水体像素,此时使用双重注意力模块即可自适应地聚合水体的语义特征,给出更精确的语义分割结^[23].

4.3 RWD-GAN网络实验

4.3.1 实验设置

本实验使用的数据集和第4.2节实验中使用的数据集相同,采用的评价指标仍然以F1-Measure为主,所有的对比试验也在相同的数据集下进行,因此所有的结果都具有可比性.

本节提出的方法使用tensorflow框架实现.由于本实验的网络模型较为复杂,原有的硬件设备显存容量有限,因此本实验在搭载了4块GeForce GTX 1080ti GPU的计算机上进行.我们使用tf.device函数指定每个计算的设备,同时在生成tensorflow会话时将allow_soft_placement参数指定为True.这样如果计算无

法由GPU执行时,tensorflow会将其放置于CPU上.

由于在水体检测问题中,水体像素数只占全部像素数的很小一部分,其分割损失通常很小,故为了平衡数据损失和对抗损失,本实验将 α_1 设置为 10^4 ,将 α_2 的初始值设置为1.同时,为了保证网络的收敛性,将按照 $(1-i/EPOCHS)$ 的方式进行线性衰减,其中 i 代表当前迭代epoch数,EPOCHS代表总的迭代epoch数.实验的批大小设置为1,两个生成器的初始学习率为 10^{-4} ,鉴别器的初始学习率为 10^{-5} ,使用Adam优化算法进行优化,一阶矩估计的指数衰减率为0.9,二阶矩估计的指数衰减率为0.999.整个训练过程在120个epoch后终止,约 10^5 次迭代.

需要注意的是,在训练阶段,两个生成器和一个鉴别器同时工作,然而在测试阶段,我们只需要两个生成器的结果,输出结果为它们的平均值.在第4.3.2节同样做了对比实验,讨论并对比了其他融合策略的结果.

4.3.2 实验结果

为了验证第3.2节提出的基于对抗学习的道路场景图像水体检测模型(RWD-GAN)的有效性,同时对比第2.3节中提出的URA-Net进行对比,结果如表4所示.可以看到,RWD-GAN在除了准确率以外的所有指标上都取得领先,相比第2.3节中提出的方法,在On-Road数据集上,F1-Measure提升了2.49%,在OffRoad数

据集上 F1-Measure 则提升了 4.45%, 在最大的 BothRoad 数据集上, F1-Measure 提升了 1.36%. 经观察, 模型的性能提升主要来自检测召回率的上升. 这表明了在 G_2 中设计的空洞卷积不仅有效地提升了模型提取上下文信息的能力, 抑制了误检, 而且在和 G_1 以及鉴别器 D 对抗学习的过程中提升了对细节特征的把握能力, 降低了漏检. 对于 G_1 来说亦是如此.

表 4 目前各种道路场景图像水体区域检测方法的结果对比

Dataset	Model	Accuracy/%	F1-measure/%
On Road	GMM&polarisation	96.50	31.00
	DeepLab	98.83	21.97
	FCN-8s-FL-RAU	99.35	70.11
	cGAN-mRAU	99.72	87.33
	URA-Net	99.73	87.83
	RWD-GAN	99.59	90.32
Off Road	GMM&polarisation	95.20	28.10
	DeepLab	98.56	45.05
	FCN-8s-FL-RAU	99.38	81.67
	cGAN-mRAU	99.50	84.29
	URA-Net	99.56	86.80
	RWD-GAN	99.56	91.25
Both	GMM&polarisation	95.10	30.34
	DeepLab	98.59	30.36
	FCN-8s-FL-RAU	99.34	76.91
	cGAN-mRAU	99.54	82.82
	URA-Net	99.64	87.18
	RWD-GAN	99.57	88.54

为了探究不同的结果融合策略对最终检测结果的影响, 本节做了对比试验. 生成器 G_1 的结果 O_1 和生成器 G_2 的结果 O_2 分别以求均值、求交集、求并集的运算方式, 计算得到最终的分割结果, 其公式如式 (12)~(14) 所示:

$$R = \frac{O_1 + O_2}{2} \quad (12)$$

$$R = O_1 \cap O_2 \quad (13)$$

$$R = O_1 \cup O_2 \quad (14)$$

不同融合策略的性能指标如表 5 所示. 选取的指标为最具代表性的 F1-measure, 从表中可以看出, 无论在哪个数据集上取均值的融合策略效果都最佳.

表 5 不同融合策略的性能指标 单位: %

Dataset	average	intersection	union
On Road	90.32	89.58	88.40
Off Road	91.25	90.31	90.98
Both Road	88.54	87.39	87.46

4.3.3 消融实验

为了探究第 3.2 节提出的两个生成器网络之间的对抗训练是否提升了模型的整体性能, 本节做了消融实验, 分别单独使用生成器 G_1 和生成器 G_2 在相同实验设置下训练, 对比三者的性能表现. 其结果如表 6 所示.

表 6 显示了单独使用生成器 G_1 、单独使用生成器 G_2 以及联合使用生成器 G_1 和 G_2 (即 RWD-GAN) 的性能比较. 其中, 生成器 G_1 选择局部感受野, 其目的是减少漏检; 生成器 G_2 使用膨胀卷积扩大感受野, 其目的就是降低误检. 根据消融实验的结果来看, 生成器的 G_1 的 Recall 是高于生成器 G_2 的, 生成器 G_2 的 Precision 是高于生成器 G_1 的, 这符合网络的设计初衷.

表 6 单独使用生成器 G_1 、单独使用生成器 G_2 和 RWD-GAN 的性能对比

Dataset	Model	Precision/%	Recall/%	F1-measure/%
On Road	G_1	91.08	84.81	87.83
	G_2	91.64	82.30	86.72
	G_1 -cGAN	91.23	84.93	87.97
	G_2 -cGAN	91.67	82.56	86.88
	Avg-cGAN	91.45	83.77	87.44
	RWD-GAN	92.50	88.24	90.32
	Off Road	G_1	91.02	82.97
G_2		89.15	80.43	84.57
G_1 -cGAN		91.06	82.97	86.83
G_2 -cGAN		90.02	80.89	85.21
Avg-cGAN		90.65	81.46	85.81
RWD-GAN		94.37	88.33	91.25
Both Road		G_1	91.04	83.64
	G_2	91.30	82.90	86.90
	G_1 -cGAN	91.12	83.71	87.26
	G_2 -cGAN	91.34	83.05	87.00
	Avg-cGAN	91.23	83.34	87.11
	RWD-GAN	91.89	85.43	88.54

与此同时, RWD-GAN 的 Precision, Recall 和 F1-measure 相比单独使用生成器 G_1 和单独使用生成器 G_2 均有所提高. 这说明了通过双生成器 cGAN 框架下的最小与最大博弈, 网络的整体性能得到了提升.

此外, 我们还增加了 RWD-GAN 与标准的 cGAN (单生成器单鉴别器) 的对比实验. 实验中, 我们将 RWD-GAN 中的 G_1 和 G_2 分别单独作为 cGAN 中的生成器进行训练, 得到 G_1 -cGAN 和 G_2 -cGAN 模型. 同时, 我们将 G_1 -cGAN 和 G_2 -cGAN 模型输出的结果的平均作为 Avg-cGAN 模型. 实验表明, 这些模型性能均不如 RWD-GAN.

5 总结

本文使用深度神经网络对道路场景图像中的水体区域检测问题展开研究.

首先提出了以 U-Net 为骨架的 URA-Net 网络模型, 同时结合了反射注意力机制和双重注意力模块. 反射注意力机制能够寻找垂直方向上的特征对应关系, 而双注意力机制中的位置注意力模块能捕捉特征之间的远程依赖关系, 通道注意力模块能加强相互依赖的特征映射并且改善特定语义的特征表示. 消融实验表明, 反射注意力机制和双注意力模块结合使用使水体检测结果有明显的提升. 最终实验表明, 我们的 URA-Net 网络在“Puddle-1000”数据集上的 F1 指标达到了 87.18%, 有望成为该领域的基线方法.

进一步, 提出基于双生成器单鉴别器的对抗学习道路场景水体检测方法 RWD-GAN, 生成器 G_1 遵循 URA-Net 的设置, 生成器 G_2 在 URA-Net 网络的最后两个残差卷积块加入膨胀卷积, 扩大感受野范围. 鉴别器 D 使用大的卷积核和卷积步长, 以适应道路场景图像水体区域检测问题. 实验结果表明, 通过对抗学习的方式, 推动了两个生成器之间的信息流动, 二者在执行各自子任务的同时, 通过对抗学习, 也受益于对方的子任务, 更好地完成了水体检测任务, RWD-GAN 网络最终在“Puddle-1000”数据集上达到了 88.54% 的 F1 指标, 取得了 SOTA 效果.

参考文献

- [1] MATTHIES L H, BELLUTA P, MCHENRY M. Detecting water hazards for autonomous off-road navigation[C]//Proceedings of SPIE, Unmanned Ground Vehicle Technology. Orlando: SPIE, 2003, 231-242.
- [2] ZHAO Y, DENG Y, PAN C, et al. Research of water hazard detection based on color and texture features[J]. Sensors & Transducers, 2013, 157(10): 428-433.
- [3] KIM J, BAEK J, CHOI H, et al. Wet area and puddle detection for advanced driver assistance systems (ADAS) using a stereo camera[J]. International Journal of Control, Automation and Systems, 2016, 14(1): 263-271.
- [4] YAO T, XIANG Z, LIU J, et al. Multi-feature fusion based outdoor water hazards detection[C]//2007 International Conference on Mechatronics and Automation. Harbin: IEEE, 2007: 652-656.
- [5] RANKIN A L, MATTHIES L H, BELLUTTA P. Daytime water detection based on sky reflections[C]//2011 IEEE International Conference on Robotics and Automation. Shanghai: IEEE, 2011: 5329-5336.
- [6] XIE B, PAN H, XIANG Z, et al. Polarization-based water hazards detection for autonomous off-road navigation[C]//2007 International Conference on Mechatronics and Automation. Harbin: IEEE, 2007: 1666-1670.
- [7] NGUYEN C V, MILFORD M, MAHONY R. 3D tracking of water hazards with polarized stereo cameras[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 5251-5257.
- [8] SANTANA P, MENDONÇA R, BARATA J. Water detection with segmentation guided dynamic texture recognition [C]//2012 IEEE International Conference on Robotics and Biomimetics (ROBIO). Guangzhou: IEEE, 2012: 1836-1841.
- [9] METTES P, TAN R T, VELTKAMP R. On the segmentation and classification of water in videos[C]//2014. International Conference on Computer Vision Theory and Applications (VISAPP). Lisbon: IEEE, 2014, 1: 283-292.
- [10] HAN X, NGUYEN C, YOU S, et al. Single image water hazard detection using fcn with reflection attention units [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: IEEE, 2018: 105-120.
- [11] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 3431-3440.
- [12] WANG L, WANG H. Water hazard detection using conditional generative adversarial network with mixture reflection attention units[J]. IEEE Access, 2019, 7: 167497-167506.
- [13] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL]. (2014-11-06)[2021-06-10]. <https://arxiv.org/abs/1411.1784>.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.
- [15] 马勇, 成谢锋, 唐振民, 等. 野外场景分析与水体识别新方法的研究[J]. 电子学报, 2013, 43(7): 1419-1424.
MA Y, CHENG X F, TANG Z M, et al. New methods for outdoor scene analysis and water body identification[J]. Acta Electronica Sinica, 2013, 43(7): 1419-1424. (in Chinese)
- [16] MNIH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//Advances in Neural Information Processing Systems. Montreal: MIT Press, 2014: 2204-2212.
- [17] 乔思波, 庞善臣, 王敏, 等. 基于残差混合注意力机制的

- 脑部 CT 图像分类卷积神经网络模型[J]. 电子学报, 2021, 49(5): 984-991.
- QIAO S B, PANG S C, WANG M, et al. A convolutional neural network for brain CT image classification based on residual hybrid attention mechanism[J]. Acta Electronica Sinica, 2021, 49(5): 984-991. (in Chinese)
- [18] 李志欣, 孙亚茹. 双路注意力引导图卷积网络的关系抽取[J]. 电子学报, 2021, 49(2): 315-323.
- LI Z X, SUN Y R. Dual attention guided graph convolutional networks for relation extraction[J]. Acta Electronica Sinica, 2021, 49(2): 315-323. (in Chinese)
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach: MIT Press, 2017: 5998-6008.
- [20] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. Montreal: MIT Press, 2014: 2672-2680.
- [21] NGUYEN V, YAGO VICENTE T F, ZHAO M, et al. Shadow detection with conditional generative adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Honolulu: IEEE, 2017: 4510-4518.
- [22] CHEN M H, HONG Y X, DENG C. Domain adaptation for semantic segmentation with maximum squares loss. Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE, 2019: 1476-1480
- [23] 冯建周, 宋沙沙, 王元卓, 等. 基于改进注意力机制的实体关系抽取方法[J]. 电子学报, 2019, 47(8): 1692-1700.
- FENG J Z, SONG S S, WANG Y Z, et al. Entity relation extraction based on improved attention mechanism[J]. Acta Electronica Sinica, 2019, 47(8): 1692-1700. (in Chinese)
- [24] Kasi V, Kartikeya S, Peng J. Offseg: A semantic segmentation framework for off-road driving[EB/OL]. (2021-03-23)[2021-06-10]. <https://arxiv.org/abs/2103.12417>.
- [25] ZHENG W, YAN L, WANG F Y, et al. Progressive knowledge-embedded unified perceptual parsing for scene understanding[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville: IEEE, 2021: 1633-1642.
- [26] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 1210-1214.
- [27] WANG J, SUN K, CHENG T, et al. Deep high-resolution representation learning for visual recognition [EB/OL]. (2019-08-20)[2021-06-10]. <https://arxiv.org/abs/1908.07919>.
- [28] YU C Q, WANG J B, PENG C, et al. BiSeNet: Bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018, 325-341.

作者简介



王臣毅 男, 1997年6月出生, 重庆人. 现为南京理工大学计算机科学与工程学院硕士研究生. 主要研究方向为计算机视觉、深度学习.
E-mail: wcyjerry@qq.com



王欢(通讯作者) 男, 1982年7月出生, 江苏句容人. 南京理工大学计算机科学与工程学院学院副教授. 主要研究方向为模式识别、图像处理、红外目标检测、目标跟踪、机器人视觉导航和人工智能.
E-mail: wanghuanphd@njust.edu.cn



孟策 男, 1997年11月出生, 江苏盐城人. 南京理工大学计算机科学与工程学院硕士研究生. 主要研究方向为计算机视觉、深度学习.