

DOI: 10.13382/j.jemi.B2205896

融合 Transformer 与残差通道注意力的恶劣场景水位智能检测方法^{*}

李欣宇^{1,2} 孙传猛^{1,2} 魏 宇^{1,2} 原 玥^{1,2} 武志博^{1,2} 李 勇³

(1. 中北大学省部共建动态测试技术国家重点实验室 太原 030051; 2. 中北大学电气与控制工程学院 太原 030051; 3. 重庆大学煤矿灾害动力学与控制国家重点实验室 重庆 400044)

摘 要:精准感知水位信息变化是实现精细水务管控和洪涝灾害的关键环节之一,而低照度、雾霾、雨雪、冰冻、波浪、镜头抖动等恶劣场景给水位检测带来极大挑战。针对现有方法中难以实现水位精准检测难题,构建一种融合 Transformer 与残差通道注意力机制的 Unet 模型(TRCAM-Unet),进而提出基于 TRCAM-Unet 的恶劣场景水位智能检测方法。关键技术包括通过全尺度连接结构实现多层次特征融合,通过 Transformer 模块强化区域特征的关联性,通过残差通道注意力模块强化有用信息的表达并削弱无用信息的干扰。相关试验和实践表明,TRCAM-Unet 取得了 98.84%MIOU 评分与 99.42%的 MPA 评分,在约 150 m 距离外水位检测最大误差不超过 0.08 m,水位偏差均值(MLD)仅有 1.609×10^{-2} m,优于 Deeplab、PSPNet 等主流语义分割算法。研究结果对解决恶劣场景下水位精准检测难题及洪涝灾害预警具有重要应用价值。

关键词:水位检测;深度学习;语义分割;Transformer;注意力机制

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6

Water level intelligent detection method based on fuse Transformer residual channel attention mechanism in harsh environments

Li Xinyu^{1,2} Sun Chuanmeng^{1,2} Wei Yu^{1,2} Yuan Yue^{1,2} Wu Zhibo^{1,2} Li Yong³

(1. State Key Laboratory of Dynamic Measurement Technology, North University of China, Taiyuan 030051, China; 2. School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China; 3. State Key Laboratory of Coal Mine Disaster Dynamics and Control, Chongqing University, Chongqing 400044, China)

Abstract: Accurate perception of water level changes is one of the key segments to achieve precision water affairs control and flood disaster, but harsh scenes such as low illumination, haze, rain and snow, freezing, lighting, and waves bring a great challenge to water level accurate detection. Aiming at the problem of accurate detection of water level in existing methods, this paper constructs a Unet model fused with transformer residual channel attention mechanism (called “TRCAM-Unet”), then, a water lever intelligent detection method in harsh environments based on TRCAM-Unet is proposed. The key technologies include that: Multi-level feature fusion is achieved by full scale connection structure. The relevance of regional feature is strengthened by transformer module. Strengthening the extraction ability of useful information and weakening the interference of useless information by residual channel attention module. The experiments and practices of water level semantic segmentation in harsh scenes shows that TRCAM-Unet achieved 98.84% MIOU scores and 99.42% MPA scores, the maximum error of water level detection outside 150 meters was not above 0.08 m, mean water level deviation (MLD) had only 1.609×10^{-2} meters, it is much better than the mainstream semantic segmentation models such as Deeplab, PSPNet, Unet. This study has important application value for water level accurate detection in harsh scenes and flood disaster early warning.

Keywords: water level detection; deep learning; semantic segmentation; Transformer; attention mechanism

收稿日期: 2022-10-09 Received Date: 2022-10-09

^{*} 基金项目:国家重点研发计划青年科学家项目(2022YFC2905700)、山西省高等学校科技创新项目(2020L0294)、山西省科技成果转化引导专项(202104021301061)项目资助

0 引言

近年来,特大暴雨、城市内涝时有发生,严重危害人民生命财产安全。精准检测并提供江河湖泊水位数据,是实现水资源精细管控、预防洪涝灾害、提高防洪工作可靠性与高效性的关键环节^[1]。现有的水位检测方法,主要依靠传感器或人眼检测等方法进行检测。付波霖等^[2]利用 Jason-3/Sentinel-3A 雷达高度计测量数据与改进的和改进的阈值法对测高水位进行波形重定改正获取滨海湿地水位,但该方法仍然存在 0.28 m 的平均均方误差,无法保证水位的精准识别;王旺等^[3]研制了一套液介式超声水位测量装置,并研究影响其测量能力的关键技术问题,研究发现波束倾角是现场测量的误差来源之一,实际检测中,受恶劣天气影响较大,该环境中易出现较大误差;薛瑄等^[4]提出了一种同轴电缆电容式水位测量方法,实验表明,该方法适用于存在恶劣水质影响的环境的测量,但该方法受水深影响精度较大,无法适用于所有环境。综上所述,采用接触式或非接触式检测方法受到雨雪等极端天气影响较大,存在着数据共享程度受限、站点在空间上不够连续等弊端^[5-6],无法满足现代水资源管理所迫切需求的自动化、数字化、智能化监测需求。

目前,深度学习技术得到了迅速发展,以 YOLO 系列^[7]、RetainNet^[8]、SSD^[9]等为代表的目标检测算法以及以 FCN^[10]、Segnet^[11]、Unet^[12]为代表的语义分割算法在计算机视觉领域展现了卓越性能。这使得基于深度学习的水位自动识别成为可能。一些学者引入目标检测算法进行水位识别研究;肖卓等^[13]利用 SSD 开展了视频水位检测研究,对水位的检测精度在 95% 左右,无法满足实际应用中,对水位精度的需求;廖赞等^[14]设计了一种卷积神经网络结构进行水位线检测,测试结果中表明存在大量误差样本;王磊等^[15]提出一种由 YOLO-v3 对象检测和 ResNet 刻度识别模型构成的实时水位识别算法,该识别方法具有准确率高、泛化能力强、识别速度快、环境适应能力强等特点,但对训练样本需求量大,少量样本可能会极大影响检测结果;夏平等^[16]针对水尺污渍和倾斜造成观读数据误差问题,提出基于改进超像素与图割算法的智能视觉水位识别算法,该方法对应用于特征不明显或动态图像检测效果较差,实际应用中无法应对恶劣环境;徐志康等^[17]采用图像处理技术处理采集的水尺图像实现水尺定位及分割,然后基于模糊 C 聚类方法分割水尺字符并输入到训练好的卷积神经网络中进行识别。目标检测网络存在的问题在于,识别主要方向为识别水面边界,当存在强风干扰、雨水等恶劣天气影响时,对检测精度影响较大,拟合的水面曲线容易出现较大的偏差。基于语义分割网络模型的水位识别技术也得到了关注;程

淑红等^[18]利用 U-net 语义分割网络检测水位线,该方法受图像质量影响较大,无法应对实际应用中的复杂环境;傅启凡等^[19]采用改进的 DeepLabv3+ 算法并结合空间注意力机制、通道注意力机制和边缘细化模块对水位标尺图像进行分割,以实现水位的智能识别。

然而在实践中,露天环境下如夜晚、雨雪、雾霾、冰冻等各类恶劣场景严重干扰了水位线的准确定位,导致上述方法难以做到水位精准监测。有效实现恶劣场景下的水位精准检测成为当前亟待解决的难题。鉴于此,本文结合恶劣场景下水位图像特点,通过全尺度连接结构实现多层次特征融合,通过 Transformer^[20] 模块强化区域特征的关联性,通过残差通道注意力模块强化对有用信息的提取能力并削弱无用信息的干扰,构建一种融合 Transformer 与残差通道注意力机制的 Unet 模型 (unet model fusion of Transformer and residual channel attention mechanism, TRCAM-Unet),进而提出基于 TRCAM-Unet 的恶劣场景水位智能检测方法。

1 恶劣环境的水位特征图像分析

水位监测区域通常位于露天环境,受季节、气候、天气等因素影响严重,这给水位精准检测带来严重挑战。如在夜晚、阴影等场景下的低照度挑战,灯光、阴雨、烈日等场景下的光影变换挑战,雾、霾、大风抖动镜头等场景下的图像模糊挑战,雨、雪、冰冻、波浪等场景下的干扰与水位模糊挑战等等。图 1 列举了部分恶劣场景下的水位图像,这些图像中水位线(水面与岸堤的交界线)受到综合不利因素的干扰,具有鲜明的弱边缘特征,严重制约了深度学习视觉识别方法的特征捕捉能力。传统检测算法中,通常对弱边缘特征的识别造成较大的识别错误,从而导致在恶劣环境中无法实现水面的精准检测。

2 基于 TRCAM-Unet 的无水尺水位检测技术

2.1 TRCAM-Unet 模型主体结构

在水位检测中,基于目标检测算法的技术方案通常以水位线(水面与岸堤的交界线)相邻区域为识别目标,通过多个识别结果进行特定线型拟合以获得实际水位线的估计。然而,水位线多呈自然曲线,简单的拟合成直线或其他线型,会造成精度损失。因此,本文采用语义分割方法进行水位检测。

在恶劣场景下,水位图像中“水面”、“背景(案堤)”语义特征是没有变化的,但恶劣场景下的综合不利因素影响了深度学习模型对语义特征的准确表达。同时,在深度学习中,堆叠“深度”有利于语义特征的学习,但往往会造成位置特征信息的丢失。因此,恶劣场景下的水

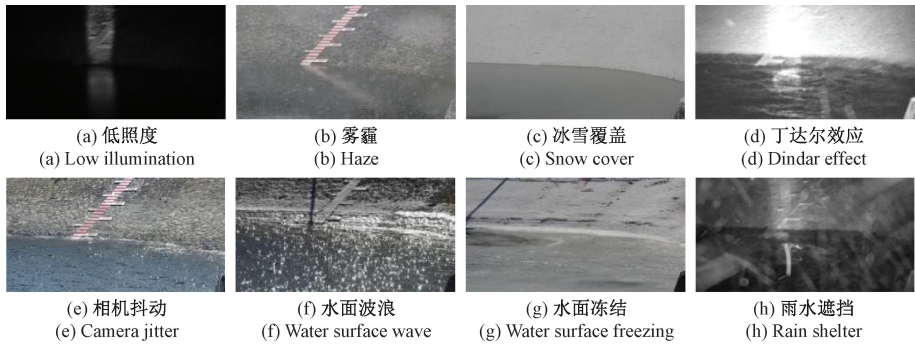


图 1 恶劣环境下影响的水位检测图像

Fig. 1 Water level detection image affected by harsh environments

位识别难点在于如何构建一种能有效表征“水面”、“背景(案堤)”语义特征的像素级分割模型,以及如何实现“深度”堆叠与误差反向传播之间的平衡。

针对上述现有难题,本文提出了一种融合 Transformer 与残差通道注意力机制的 Unet 模型 (TRCAM-Unet),通过全尺度连接结构实现多层次特征融合,通过 Transformer 模块增强网络对于水面图像区域

特征的关联性,通过残差通道注意力模块增强有用特征信息与降低无用噪声影响,降低恶劣场景下水面遮挡所造成的影响,整体结构如图 2 所示。该网络模型在 Unet 网络的基础上,将跳接结构替换为全尺度连接结构,将编码器的结构中最深维度的卷积结构替换为 Transformer 结构,在编码器与解码器的连接处增加残差通道注意力结构,从而构成本文提出的 TRCAM-Unet 网络模型。

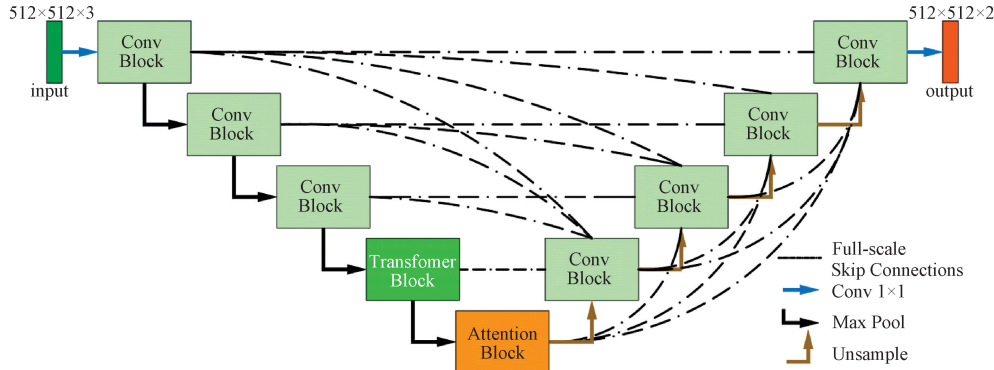


图 2 TRCAM-Unet 模型

Fig. 2 TRCAM-Unet module

2.2 全尺度连接结构

为了降低特征提取过程中网络深度造成的浅层特征的损失,Unet 网络基于残差连接模块设计了一种跳接结构 (skip connection structure)。这种结构将编码器的输出特征图直接拼接和解码器所对应结构,从而保证浅层特征的保留,增加了语义分割模型对区域特征的精准识别。

然而,受到恶劣场景下综合不利因素的影响,水位线呈现弱边缘特征,致使简单跳接结构无法保证网络能够全尺度的收集浅层信息,这对水位线位置与边界的精准定位是十分不利的。为最大限度的捕捉各尺度下的特征信息,特别是浅层的位置信息,在 TRCAM-Unet 使用全尺度连接结构。全尺度连接结构中解码器的每一层模块都融合了编码器所提取的全部尺度的特征图,以及来自于

解码器较低层的特征图,如图 3 所示。

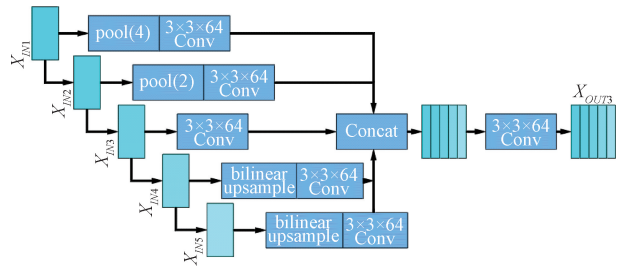


图 3 全尺度连接结构

Fig. 3 Full-scale Skip Connection structure

TRCAM-Unet 的编码器共输出 5 个尺度的特征图,解码器通过全尺度连接结构接受并融合所有尺度的特征图

输出。设编码器与解码器在各尺度特征的输出分别记为 X_{INi}, X_{OUTi} , 其中 $i \in [1, 5]$, 则有:

$$X_{OUTi} = \begin{cases} X_{INi}, i = 5 \\ H\left(\underbrace{C(D(X_{INk}))}_{Scales: 1^{th} \sim i^{th}}^{i-1}, \underbrace{C(X_{INi})}_{Scales: i+1^{th} \sim N^{th}}^N, C(U(X_{OUTi}))_{K=i+1}^N\right), \\ i = 1, \dots, 4 \end{cases} \quad (1)$$

式中: $H(\cdot)$ 表示经由卷积操作与拼接操作所实现的多尺度的特征融合; $C(\cdot)$ 表示卷积操作; $D(\cdot)$ 与 $U(\cdot)$ 分别表示上采样与下采样操作。

由于编码器在各尺度上的输出 X_{INi} 尺寸并不相同, 解码器在各尺度上进行特征融合时, 必须先调整 X_{INi} 使其尺寸一致, 再进行拼接与融合工作: 1) 对 $X_{IN1} \sim X_{IN(i-1)}$, 利用池化操作将其降维成与 X_{OUTi} 同样尺寸; 2) 对 $X_{IN(i+1)} \sim X_{IN5}$, 利用双线性插值上采样将其升维成相同尺寸; 3) 对 X_{INi} , 利用卷积操作进一步提取特征。通过将不同尺度的特征图调整为 64 通道与 X_{INi} 同尺度的特征图后, 将其拼接为 320 通道的特征图, 经由 $3 \times 3 \times 64$ 的卷积核卷积融合后得到最终的输出特征图。

由于全尺度连接结构中, 编码器中的每一尺度的特征图相较于 Unet 的跳接结构采用了更少的卷积层运算便拼接到了解码器中, 使得编码器特征图中的不同深度的特征能够得到更为有效的保留。

2.3 Transformer 模块

卷积神经网络 (convolutional neural networks, CNN) 利用卷积操作, 有效捕捉了输入图像的局部时空信息。然而, 在水位识别任务中, 通常“水面”绵延于图像特定区域, 且位置变化较小, 其特征具有很强的区域关联性, 显著区别于常规分割或者检测任务中目标离散、位置分布随机等特点。因此, 若能充分利用待分割目标特征的区域关联性, 将显著提升模型对语义特征的捕捉和描述能力, 有利于提升模型对恶劣场景综合不利因素的适应性和泛化性能。

近年来, Transformer 在计算机视觉^[21] 领域得到迅速发展。Transformer 自带长依赖特性, 利用自注意力机制来捕获全局信息, 可抽取更强有力的特征, 从而具有学习长距离依赖能力强、多模态特征融合能力强等优势。相比 CNN, Transformer 的自注意力机制不受局部相互作用的限制, 既能挖掘长距离的依赖关系又能并行计算, 可以根据不同的任务目标学习最合适的归纳偏置。

综上所述, 本文在 TRCAM-Unet 中引入 Transformer 多头自注意力机制 (multi-head self attention, MHSA), 提升模型对目标语义特征的表达能力。在 MHSA 模块中, 每个头都应用了独立的自注意力机制, 这使得模型可以针对水位图像特性, 在不同的表示子空间里挖掘、学习相

关的高阶特征, 从而极大强化了 TRCAM-Unet 对不同区域的区分能力。

在 MHSA 中, 单一注意力头的计算式如下:

$$h_i = \text{softmax}(qk + qr)v, i = 1, \dots, 8 \quad (2)$$

式中: h_i 表示第 i 个注意力头; $q = xW_q, k = xW_k, v = xW_v$, x 为输入特征图, W_q, W_k, W_v 为卷积层权重; r 为位置编码矩阵, $r = R_H \times R_W, R_H$ 与 R_W 为可以训练的纵向与横向的位置特征向量, 为自注意力机制增加不同位置特征的关联性; $\text{Softmax}(\cdot)$ 为激活函数, 将所得到的特征图通道方向上转换为和为 1 的权重:

$$\text{Softmax}(x) = \frac{e^{x_j}}{\sum_j e^{x_j}} \quad (3)$$

其中, x_j 表示特征图中的对应通道的像素; c 表示特征图通道数。单一注意力头的特征提取机制如图 4 所示。图中将 q 矩阵与 k 的转置相乘得到输入的相关性矩阵, 与编码矩阵相乘, 便得到了原特征图的权值矩阵, 最后与 v 矩阵相乘, 得到了加权后的输出特征图。

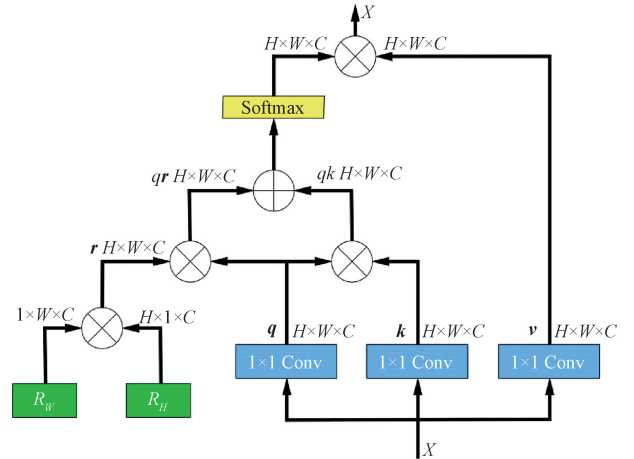


图 4 单一注意力头的特征提取机制

Fig. 4 Feature extraction mechanism for single attention head

这样, 多头自注意力机制可表达如下:

$$\text{MHSA}(q, k, v) = \text{concat}(h_1, \dots, h_8) W_0 \quad (4)$$

将 MHSA 得到的特征图, 经多层感知机 (multilayer perceptron, MLP) 进行特征融合, 即为 TRCAM-Unet 中 Transformer 模块的完整结构, 如图 5 所示。其输出可表达为:

$$T = \text{MLP}(\text{BN}(\text{MHSA}(\text{BN}(x)) + x)) + (\text{MHSA}(\text{BN}(x)) + x) \quad (5)$$

其中, $\text{BN}(\cdot)$ 表示正则化操作, 通过平移和缩放处理, 达到加快网络的训练速度、防止梯度消失或爆炸的目的, 从而降低网络发生过拟合的可能。

2.4 残差通道注意力模块

除增强对区域特征提取能力以外, 抑制水位图像中的干扰信号也是增强 TRCAM-Unet 性能的重要手段。在

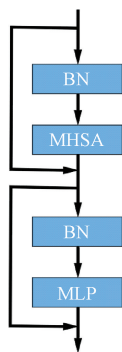


图5 Transformer 模块

Fig. 5 Transformer module

深度学习中,注意力机制(attention mechanism)可以使模型在众多的输入信息中聚焦于对当前任务更为关键的信息,降低或者过滤掉对其他信息的关注度,从而提高任务处理的效率和准确性。因此,本文在 TRCAM-Unet 中引入 SEblock(squeeze and excitation block)^[22],以强化模型对水面区域的关注,而忽略掉恶劣场景下的诸多干扰信息。SEblock 是一种基于通道维度的注意力模块,从特征图各通道中提取对任务有益的特征信息,实现对全局无用信息的抑制作用。同时,类似于跳接结构,SEblock 在改变通道的权重时,为了不丢失原本特征图中存在的特征信息,增加残差^[23]结构以实现浅层特征(区域特征)的保留。残差通道注意力模块如图6所示。

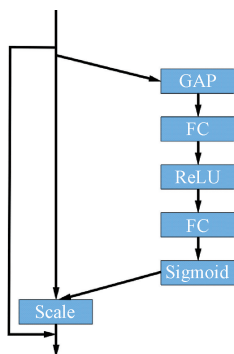


图6 残差通道注意力模块

Fig. 6 Residual channel attention module

首先,利用全局平均池化(global avgpooling, GAP)实现对全局特征的提取,并将其转换为一个能够代表全局特征的实数 $X_c \in R^C$,如式(6)所示。

$$X_c = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H x_{(i,j)} \quad (6)$$

其中, W, H 表示特征图的宽与高; $x_{(i,j)}$ 表示特征图中对应位置的灰度值。

然后,将得到的全局特征经两个全连接层进行特征。特别的,第1个全连接层会将特征通道数调整为原通道数的 $1/16$,第2个全连接层将其还原为原本通道数。这样缩放的好处是:在降低网络参数计算量的同时,融合不

同通道间的特征。第1个全连接层采用的激活函数为 ReLU,如式(7)所示;第2个全连接层采用的激活函数为 Sigmoid,如式(8)所示。

$$\text{ReLU}(x) = \max(0, x) \quad (7)$$

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

最后,增加残差结构,将原本的特征图与 SEblock 输出相加,即为本文所使用的残差通道注意力模块,其输出式(9)所示。

$$F_c = X_c(1 + \text{SE}(X_c)) \quad (9)$$

2.5 基于 TRCAM-Unet 的无水尺水位检测方案

TRCAM-Unet 模型有效地解决了恶劣场景下水位识别难题,在此基础上形成无水尺水位检测方案:

1) 在前期,利用水位图像中明显的、固定的标志物获得真实距离与像素距离的比例,如式(10)所示:

$$k = \frac{y}{y'} \quad (10)$$

其中, y 表示图像中水位线的像素距离; y' 表示图像中标志物的像素距离。

2) 通过 TRCAM-Unet 模型在恶劣场景下获得精准的水面分割图像。

3) 利用边缘检测算法处理水面分割图像获得水位线像素高程,利用上述比例关系即可转换为真是高程信息。

最终,基于 TRCAM-Unet 的无水尺水位检测方案流程如图7所示。

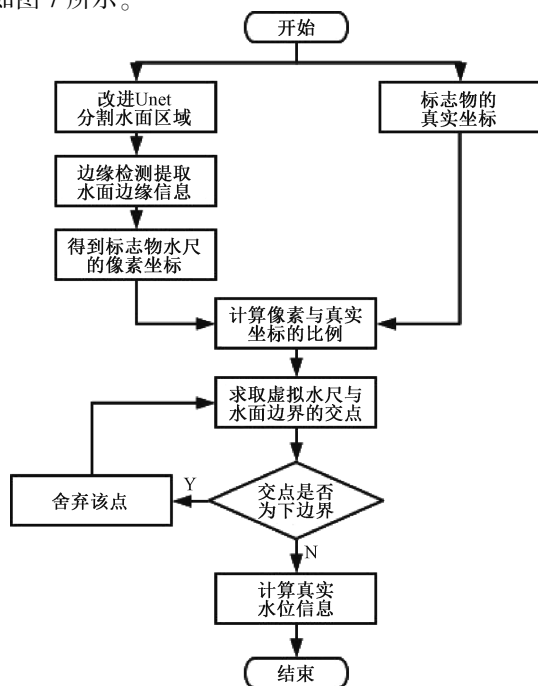


图7 恶劣场景下无水尺水位检测方案流程

Fig. 7 Process of water level detection scheme without water gauge in harsh scenarios

3 实验及结果分析

3.1 实验过程

利用采集的某水库水位图像验证本文所述 TRCAM-Unet 模型的有效性和优越性。对比模型为 Deeplab^[24]、PSPNet^[25]、Unet 等主流语义分割网络模型。为保证不同模型的测试结果不受除模型差异以外的影响,所有网络模型均采用相同的实验软硬件环境(如表 1 所示),且参数设置保持一致。本次实验中,batch size 设置为 4;epoch 设置为 50;动量设置为 0.9;Adam 优化器中的 β_1 、 β_2 、 ε , 设为 0.9、0.999、 10^{-8} ;初始学习率设置为 0.01;权重衰减系数设置为 0.000 5。

表 1 硬件环境和软件环境

Table 1 Hardware environment and software environment

环境类型	参数
CPU	AMD Ryzen 7 4800H
GPU	GeForce GTX 1650 Ti
内存/GB	4
显存/GB	16
操作系统	Windows 10
Cuda	10.1
Cudnn	7.4.1
深度学习框架	Tensorflow-gpu 2.2.0
语言	Python 3.6
其他关键库	Numpy 1.92.2、Pillow 8.2.0

本次实验共涉及 6 028 张水位图像,包含白天、夜晚、冰雪覆盖、灯光、阴影、波浪、阴雨、雾霾等正常与各类恶劣场景。本文采用 Labelme 为图像标注标签(如图 8 所示),并进行数据集增强。

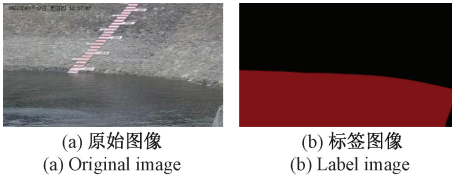


图 8 数据集标签

Fig. 8 Dataset label

采用交叉熵损失函数与 Dice 损失函数的和作为模型损失函数。交叉熵损失表示图像真实概率分布与预测概率分布之间的差异,其表达式为:

$$l_{CE} = - \sum_{i=1}^N p(x_i) \log_a q(x_i) \quad (11)$$

其中, N 为分类类别的数量; i 为类别序列号; $p(x_i)$ 为实际真实值的分类目标,如果是对应目标分类则取 1, 否则取 0; $q(x_i)$ 为预测的概率值;对数底 a 在无特殊申明下,可以取 e 。

Dice 损失表示正确分类信息再不正确分类信息中所占的比例,可以解决正负样本在图像中占比不平衡的问

题,其表达式为:

$$l_{Dice} = 1 - \frac{2 \sum_{i=1}^P p_i p_i^*}{\sum_{i=1}^P p_i + \sum_{i=1}^P p_i^*} \quad (12)$$

其中, p_i 与 p_i^* 表示像素 p 的预测值与真实值; P 为像素总数。

基于上述损失函数对 TRCAM-Unet 模型训练 50 轮次,其损失值变化如图 9 所示。

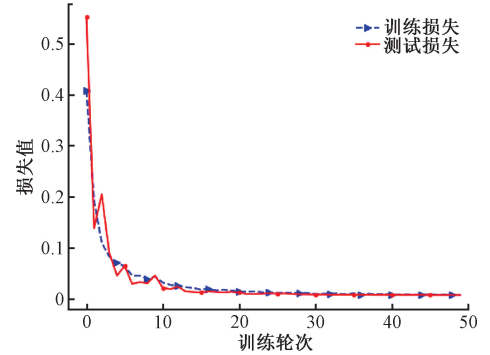


图 9 网络模型损失函数训练效果

Fig. 9 Training effect of network model loss function

3.2 实验结果及分析

1) 评估指标

在语义分割模型训练中,通常采用均交并比(mean intersection over union, MIOU)、类别平均像素准确率(mean pixel accuracy, MPA)作为评估指标。IOU 表示真实值与预测值的交集与并集的比值,可以有效地表示预测值与真实值的重叠情况,MIOU 则为全局内每一个分类的 IOU 值的平均值:

$$MIOU = \frac{1}{N+1} \sum_{i=1}^N \frac{p_{ii}}{\sum_{j=1}^N p_{ij} + \sum_{j=1}^N p_{ji} - p_{ii}} \quad (13)$$

其中, N 表示分类的类别个数; p_{ii} 表示正确分类的像素; p_{ij} 表示本属于 i 类却被分类于 j 类的像素; p_{ji} 表示本属于 j 类却被分类于 i 类的像素。

PA 值表示每一类中正确分类的像素占比,MPA 值则表示全局内各分类 PA 值的平均值:

$$MPA = \frac{1}{N+1} \frac{\sum_{i=1}^N p_{ii}}{\sum_{i=1}^N p_i} \quad (14)$$

其中, p_i 表示对应分类的全部像素点。

由于实验数据源自同一采集点,可采用水位偏差均值(mean level deviation, MLD)作为水位识别准确性指标:

$$MLD = \frac{\sum_{i=1}^L |l_i - l_i^*|}{L} \quad (15)$$

其中, L 表示检测图像总数; l_i 与 l_i^* 表示水位真实值与网络预测值。

2) 消融实验

为了证明本文所提到的不同改进方案对原语义分割模型的改进效果,本文进行了消融实验,实验结果如表 2 所示。

表 2 消融实验结果对比

Table 2 Comparison of ablation experimental results					
结构	1	2	3	4	5
Unet	⊙	⊙	⊙	⊙	⊙
全尺度连接					⊙
Transformer		⊙		⊙	⊙
残差注意力			⊙	⊙	⊙
MPA/%	98.12	99.21	99.32	99.39	99.42
MIOU/%	96.09	98.45	98.64	98.79	98.84
参数量/10 ⁶	23.84	19.32	23.88	19.35	16.48

该实验如表 2 所示,所有实验组皆以 Unet 语义分割模型为基础模型,进行不同改进。实验组 2 相较于实验组 1,将编码器中的一个尺度的卷积层全部替换为

Transformer 结构,结果在参数量大量下降的同时,准确率得到了提升,表示 Transformer 结构有效地增强局部特征的关联性。实验组 3 相较于实验组 1,在编码器与解码器连接处增加残差通道注意力模块,在以参数微量增加的前提下有效地增加了网络准确度。实验组 4 将实验组 3 与实验组 2 的两种改进融合,最终准确率同时优于单独改进方案。实验组 5 在实验组 4 的改进前提下,将基础跳跃结构替换为全尺度连接结构,从而实现了参数量大量下降的同时增加了模型检测准确率。综上所述,本文所提出的改进,有效地增强了网络模型检测性能。

3) 主观分析

测试集总计 600 张水位图像,包含各类恶劣场景。受篇幅限制,本文选择 8 幅有代表性的图像进行展示。如图 10 所示,从左至右分别为原始图像、本文方法识别

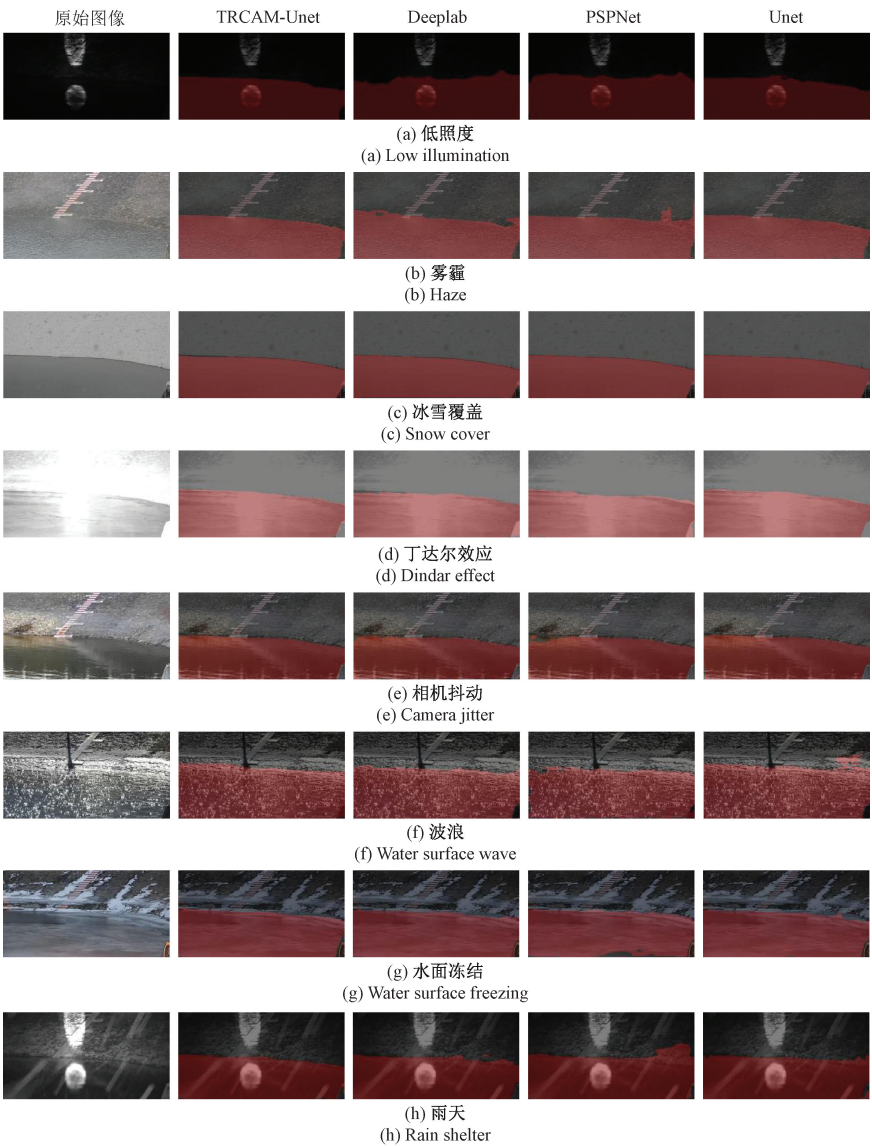


图 10 不同网络模型预测效果对比

Fig. 10 Comparison of prediction effects of different network models

效果、Deeplab 识别效果、PSPNet 识别效果、Unet 识别效果。为了便于展示,分割效果图将水面与水尺相交处放大展示,将分割的水面区域用标注展示,其余位置为案堤背景区域展示。将图 10(a)、(h)低照度和雨水遮挡的水位区域放大,如图 11 所示。由图 10 和 11 可知,本文的提出

的 TRCAM-Unet 模型,在低照度、雾霾、冰雪覆盖、相机抖动、波浪、水面冻结、雨天等恶劣场景下,分割得到的水面边界更加贴近于真实水面边界,而其他网络模型则会出现分割边界波动较大,存在错误区块识别。由此可见,本文模型可以更好的完成的在恶劣场景下的水位检测任务。

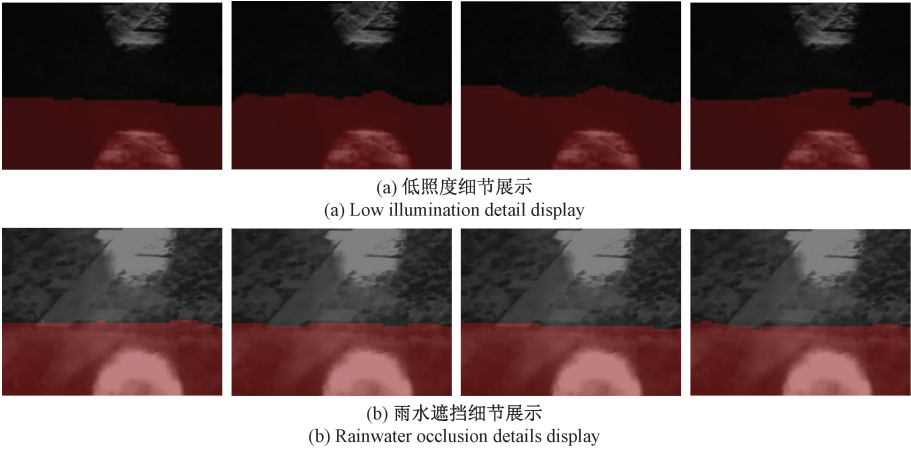


图 11 分割效果细节展示

Fig. 11 Semantic segmentation effect details display

4) 客观分析

以 MIOU、MPA 以及 MLD 作为评价指标,对比 TRCAM-Unet 模型与 Deeplab、PSPNet 与 Unet 主流语义分割网络模型性能,结果如表 3 所示。由表 3 可知,本文提出的 TRCAM-Unet 模型在 MIOU 评价指标中获得了 98.84% 的评分,以及在 MPA 评价指标中获得了 99.42

的评分,优于其他主流语义分割网络,有着较好的分割准确率;在水位高度检测任务中,TRCAM-Unet 模型仅存在 9.717×10^{-3} m 的平均水位误差,远远低于其他网络模型。可见,相比其他语义分割模型,TRCAM-Unet 模型可以更好的适应各类恶劣场景下的水位检测任务。

表 3 语义分割效果与水位监测效果评估分数

Table 3 Evaluation score of semantic segmentation effect and water level monitoring effect

	TRCAM-Unet	Deeplabv3	Unet	Pspnet
MIOU	98.84	97.35	96.09	97.87
MPA	99.42	98.67	98.12	98.96
MLD/m	9.717×10^{-3}	4.915×10^{-2}	4.405×10^{-2}	4.279×10^{-2}
参数量/ 10^6	16.48	2.76	23.84	2.45

4 恶劣场景下的水位检测实践

将本技术运用于某水库水位监测点实践。受地形地貌限制,该监测点相机镜头距水位线约 150 m。本次监测

了 2022 年 3 月 19 日~2022 年 4 月 6 日共计 18 日水位信息,监测时每隔 1 h 进行一次检测,共检测了 432 次,检测结果以虚拟水尺显示。在监测期间经历了白天、夜晚、低照度、灯光、雾霾、雨雪、大风天气。图 12 展示了部分恶劣场景下的检测结果,图中水面中曲线为语义分割结果

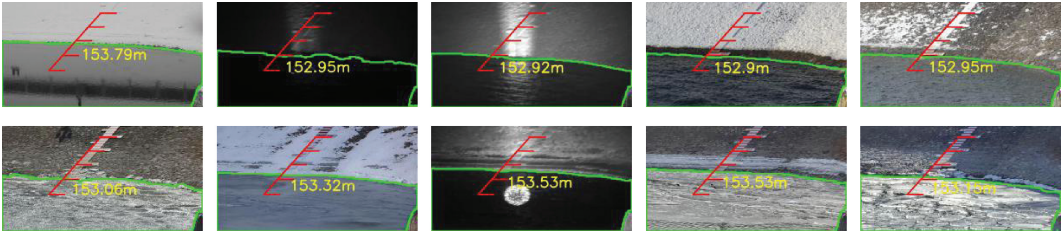


图 12 水位检测效果

Fig. 12 Water level detection effect

所产生的水位边界,折线线段为虚拟水尺,图中数字为本文方法检测的水位高程信息。检测水位与真实水位结果如 13 图所示,其差值如图 14 所示。由图 12~14 可知,在整个检测过程中,本文方法检测的水位与真实水位偏差很

小,最大偏差不超过 0.08 m,其 MLD 仅有 1.609×10^{-2} m。可见,本文方法可以实现恶劣场景下的无水尺水位精准检测。

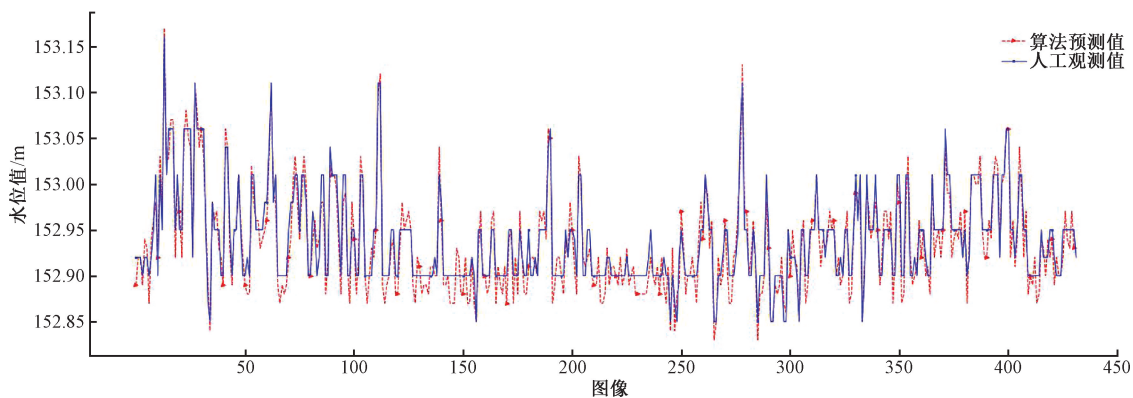


图 13 检测水位值与真实水位值对比

Fig. 13 Comparison of detected water level value and real water level value

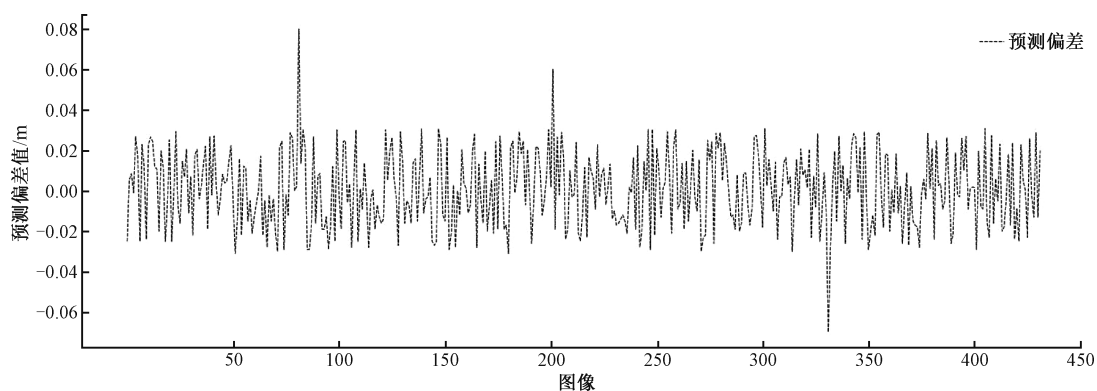


图 14 检测水位与真实水位的差值

Fig. 14 Detect the difference between the water level and the real water level

5 结 论

针对恶劣场景下的水位识别难题,为了实现在不同环境下均可以实现水位的精准识别,本文提出了 TRCAM-Unet 模型,通过全尺度连接结构实现多层次特征融合,通过 Transformer 模块增强网络对于水面图像区域特征的关联性,通过残差通道注意力模块增强有用特征信息与降低无用噪声影响,降低恶劣场景下水面遮挡所造成的影响;对真实恶劣场景的水位图像测试试验和实践表明,TRCAM-Unet 模型取得了 98.84% MIOU 评分与 99.42% 的 MPA 评分,在 150 m 距离外水位检测最大误差不超过 0.08 m,水位偏差均值 (MLD) 仅有 1.609×10^{-2} m,远优于 Deeplab、PSPNet、Unet 等主流语义分割网络模型。

在未来,本文会进一步增强语义分割网络模型对边界特征的提取能力以及改进注意力机制来降低恶劣环境对网络精准的干扰,处理图像特征中的不确定性、不精确信息。

参考文献

- [1] 孔锋. 透视变化环境下的中国城市暴雨内涝灾害: 形势、原因与政策建议 [J]. 水利水电技术, 2019, 50(10): 42-52.
KONG F. Perspective on urban rainstorm waterlogging disaster in China under changing environment: Situation, causation and policy suggestion [J]. Water Resources and Hydropower Engineering, 2019, 50(10): 42-52.
- [2] 付波霖, 覃娇玲, 何宏昌, 等. 利用 Jason-3/Sentinel-3A 雷达高度计监测北部湾滨海湿地水位变化 [J]. 农业工程学报, 2021, 37(5): 184-190.

- FU B L, QIN J L, HE H CH, et al. Monitoring the water level of coastal wetland in Beibu Gulf using Jason-3/Sentinel-3A altimetry satellites[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(5): 184-190.
- [3] 王旺,刘源,胡鹤鸣,等. 基于超声传播时间的水位精确测试[J]. 中国测试, 2022, 48(4): 35-41, 52.
- WANG W, LIU Y, HU H M, et al. Research on precision measurement method of water level based on ultrasonic propagation time[J]. China Measurement & Test, 2022, 48(4): 35-41, 52.
- [4] 薛瑄,王宇,王东,等. 同轴电缆电容感测式水位传感器设计[J]. 仪表技术与传感器, 2019(2): 9-13, 17.
- XUE X, WANG Y, WANG D, et al. Design of water level sensor base on capacitance sensing mechanism of coaxial cable[J]. Instrument Technology and Sensors, 2019(2): 9-13, 17.
- [5] 王红,孙福宝,杨涛,等. Jason_2 卫星测高数据在长江中游水位监测中的应用[J]. 三峡生态环境监测, 2018, 3(3): 48-54.
- WANG H, SUN F B, YANG T, et al. Application of Jason_2 satellite altimetry data to water level monitoring in the middle reaches of the Yangtze River[J]. Ecology and Environmental Monitoring of Three Gorges, 2018, 3(3): 48-54.
- [6] 文京川,赵红莉,蒋云钟,等. 卫星测高数据筛选方法研究——以 Jason-3 数据和洪泽湖为例[J]. 南水北调与水利科技, 2018, 16(3): 194-200, 208.
- WEN J CH, ZHAO H L, JIANG Y ZH, et al. Research on the quality screening method for satellite altimetry data—take Jason-3 data and Hongze Lake as an example[J]. South-to-North Water Transfers and Water Science & Technology, 2018, 16(3): 194-200, 208.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [10] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [11] BADRINARAYANAN V, HANDA A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling[J]. arXiv preprint arXiv:1505.07293, 2015.
- [12] 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. Springer, Cham, 2015: 234-241.
- [13] 肖卓,陶青川,沈建军. 基于 SSD 目标检测的视频水位检测算法[J]. 现代计算机(专业版), 2019(9): 60-64.
- XIAO ZH, TAO Q CH, SHEN J J. A video water-level recognition based on SSD object detect network[J]. Modern Computer (Professional Edition), 2019(9): 60-64.
- [14] 廖赞,段清,刘俊晖,等. 基于深度学习的水位线检测算法[J]. 计算机应用, 2020, 40(S1): 274-278.
- LIAO Y, DUAN Q, LIU J H, et al. Water line detection algorithm based on deep learning[J]. Journal of Computer Applications, 2020, 40(S1): 274-278.
- [15] 王磊,陈明恩,孟凯凯,等. 基于深度学习算法的水位识别方法研究[J]. 水利信息化, 2020(3): 39-43, 56.
- WANG L, CHEN M EN, MENG K K, et al. Research on water level recognition method based on deep learning algorithms[J]. Water Resources Informatization, 2020(3): 39-43, 56.
- [16] 夏平,王峰,雷帮军,等. 基于超像素和图割算法的智能视觉水位识别[J]. 计算机仿真, 2021, 38(3): 430-436, 441.
- XIA P, WANG F, LEI B J, et al. Intelligent visual water level recognition algorithm based on super-pixel and graph cut segmentation[J]. Computer Simulation, 2021, 38(3): 430-436, 441.
- [17] 徐志康,冯径,张之正,等. 一种结合卷积神经网络的水深自动测量方法[J]. 小型微型计算机系统, 2019, 40(4): 793-797.
- XU ZH K, FENG J, ZHANG ZH ZH, et al. Water level estimation combines with convolutional neural network[J]. Journal of Chinese Computer System, 2019, 40(4): 793-797.
- [18] 程淑红,赵考鹏,张仕军,等. 基于 U-net 的水位线检测[J]. 计量学报, 2019, 40(3): 361-366.
- CHENG SH H, ZHAO K P, ZHANG SH J, et al. Water level detection based on U-net[J]. Acta Metrologica Sinica, 2019, 40(3): 361-366.
- [19] 傅启凡,路茗,张质懿,等. 基于语义分割的水位监测方法研究[J]. 激光与光电子学进展, 2022, 59(4): 89-100.

- FU Q F, LU M, ZHANG ZH Y, et al. Research on water level monitoring method based on semantic segmentation [J]. Laser & Optoelectronics Progress, 2022, 59 (4): 89-100.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30.
- [21] ALEXEY D, LUCAS B, ALEXANDER K, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [23] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [24] CHEN L, 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 and Machine Intelligence, 2018, 40 (4): 834-848.
- [25] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.

作者简介



李欣宇, 2019 年于石家庄铁道大学获得学士学位, 现为中北大学在读硕士研究生, 主要研究方向为深度学习与语义分割。
E-mail: 546941034@qq.com

Li Xinyu, received his B. Sc. degree from Shijiazhuang Tiedao University in 2019.

Now he is a M. Sc. candidate in North University of China. His main research interests include deep learning and semantic segmentation.



孙传猛(通信作者) 2010 年于重庆大学获得学士学位, 2015 年于重庆大学获得博士学位, 现为中北大学讲师, 主要研究方向为深度学习、机器视觉, 以及高压、高冲击、高温、高速、强电磁干扰等恶劣环境下各种瞬态过程的信息获取。

E-mail: sun_c_m@163.com

Sun Chuanmeng (Corresponding author) received his B. Sc. degree from Chongqing University in 2010 and Ph. D. degree from Chongqing University in 2015, respectively. Now he is a lecturer in North University of China. His main research interests include deep learning, computer vision, and information acquisition of various transient processes in harsh environments such as high voltage, high shock, high temperature, high speed, and strong electromagnetic interference.



魏宇, 2019 年获得佳木斯大学学士学位, 现为中北大学电气与控制工程学院硕士研究生, 主要研究方向为深度学习与图像处理。

E-mail: 3069690@qq.com

Wei Yu received his B. Sc. degree from Jiamusi University in 2019. Now he is a M. Sc. candidate of School of Electrical and Control Engineering, North University of China. His main research interests include deep learning and image processing.



原玥, 2020 年于晋中学院获得学士学位, 现中北大学硕士研究生, 主要研究方向为深度学习与信号处理。

E-mail: 441972199@qq.com

Yuan Yue received her B. Sc. degree from Jinzhong College in 2020. Now she is a M. Sc. candidate at North University of China. Her main research interests include deep learning and signal processing.



武志博, 2011 年于中北大学获得学士学位, 2020 年于中北大学获得博士学位, 现为中北大学讲师, 主要研究方向为动态测试与智能仪器。

E-mail: wuzhibo@nuc.edu.cn

Wu Zhibo received his B. Sc. degree from North University of China in 2011 and Ph. D. degree from North University of China in 2020, respectively. Now he is a lecturer in North University of China. His main research interests include dynamic test and intelligent instrument.



李勇, 2007 年于重庆大学获得学士学位, 2013 年于博洛尼亚大学获得博士学位, 现为重庆大学副教授, 主要研究方向为结构工程与水力学。

E-mail: yong.li@cqu.edu.cn

Li Yong received his B. Sc. degree from Chongqing University in 2007 and Ph. D. degree from University of Bologna in 2013, respectively. Now he is an associate professor in Chongqing University. His main research interests include structural engineering and hydraulics.