

# 基于视频时空特征学习的近岸海浪周期检测

宋巍\*, 陈媛媛, 贺琪, 杜艳玲\*\*

上海海洋大学信息学院, 上海 201306

**摘要** 近岸海浪周期检测对于近岸精细化海洋预报至关重要。为此,提出一种新的基于视频时空特征学习的近岸海浪周期自动化检测方法。所提方法以连续海浪视频帧为输入,首先利用二维卷积神经网络(2D-CNN)提取视频帧的空间特征,将空间特征在时间维度上拼接成序列,再通过一维卷积神经网络(1D-CNN)提取时间维度特征,这种复合卷积神经网络(CNN-2D1D)能够实现海浪时空信息的有效融合,最后采用注意力机制对融合后的特征进行权重调整,并将所得结果线性映射为海浪周期。将所提方法与基于 VGG16 网络的单纯空间特征的检测方法和基于 ConvLSTM 和三维卷积(C3D)网络的时空特征融合的检测方法进行对比。实验结果表明,C3D 和 CNN-2D1D 的检测精度最高,平均绝对误差分别为 0.47 s 和 0.48 s,但 CNN-2D1D 比 C3D 的检测结果更稳定,均方根误差分别为 0.66 和 0.81,且 CNN-2D1D 需要的训练参数更少,这表明所提方法在波浪周期检测中更有效。

**关键词** 海洋光学; 波浪周期检测; 时空融合特征; 融合卷积神经网络; 近岸海浪监控视频; 深度学习

中图分类号 TP301.6

文献标志码 A

doi: 10.3788/LOP202158.2401001

## Nearshore Wave Period Detection Based on Video Spatiotemporal Feature Learning

Song Wei\*, Chen Yuanyuan, He Qi, Du Yanling\*\*

College of Information Technology, Shanghai Ocean University, Shanghai 201306, China

**Abstract** The detection of nearshore wave period is crucial for fine nearshore wave forecast. Thus, we propose a novel method to realize automatic detection of nearshore wave period by learning spatiotemporal features from nearshore wave surveillance videos. The method takes continuous ocean wave video frames as inputs. First, a two-dimensional convolutional neural network (2D-CNN) is used to extract spatial features of the video frame images, and the extracted spatial features are spliced into sequences in the time dimension. Then a one-dimensional convolutional neural network (1D-CNN) is used to extract temporal features. The composite convolutional neural network (CNN-2D1D) can realize the effective fusion of wave space-time information. Finally, the attention mechanism is used to adjust the weight of the fusion features and linearly maps the fusion features to wave period. The method in this paper is compared with the detection method only extracting spatial features based on VGG16 network and the detection method for spatiotemporal feature fusion based on the ConvLSTM and three-dimensional convolutional (C3D) network. The results of experiments show that C3D and CNN-2D1D achieve the best detection results, with an average absolute error of 0.47 s and 0.48 s, respectively, but CNN-2D1D is more stable than C3D, with a lower root-mean-square error (0.66) than C3D (0.81). And CNN-2D1D requires fewer training parameters. These results show that the proposed method is more effective in wave period detection.

**Key words** oceanic optics; wave period detection; spatiotemporal fusion features; fusion convolutional neural network; nearshore wave surveillance video; deep learning

**OCIS codes** 010.4450; 100.2000; 150.0155

收稿日期: 2020-11-30; 修回日期: 2021-01-16; 录用日期: 2021-02-17

基金项目: 国家自然科学基金(61972240, 41906179)、上海市科委部分地方院校能力建设项目(2005051900)

通信作者: \*wsong@shou.edu.cn; \*\*yldu@shou.edu.cn

# 1 引言

海浪是一种重要的海洋物理现象,按照其形成原因可分为风浪、涌浪和近岸海浪。近岸海浪是近岸海洋环境中最重要的动力因素之一,与近海岸地形地貌、工程设计、浅海岸生产作业密切相关,而要实现近岸海浪的精准监控,需要高精度检测浪高、周期、波长等海浪要素。海浪要素是能表现海浪状况的主要参数,其中海浪周期为两个连续波经过一个相同点所需的时间,是冲浪预报、海洋导航定位<sup>[1]</sup>和沿海岸的地理环境变迁检测<sup>[2]</sup>中的主要影响因素。海浪周期的长短影响波的折射,波的折射又影响海岸水深,进而影响破碎浪的波高和速度,较大的波高与速度对近岸港口的安全产生一定的威胁。而较大的周期却是冲浪的良好条件。对海浪周期的高精度检测及预报不仅可以有效避免灾害性海浪带来的经济损失,还可以为冲浪爱好者提供近岸海浪的实时信息。目前,现有的近岸海浪周期预报方法主要依赖于数值模型,如 WAM<sup>[3]</sup>和 WaveWatchIII<sup>[4]</sup>数值模型被广泛应用于全球海浪场预报,SWAN<sup>[5]</sup>、TOMAWAC<sup>[6]</sup>和 MIKE21 等数值模型则应用于近岸浅水区海浪预报。

现有的海浪周期检测方法主要有两种:1)通过浮标进行实时检测和人工观测等直接检测法;2)基于谱分析的非直接检测法。其中,浮标测量可以长期、全天候地提供较为准确的海浪参数的实时信息<sup>[7]</sup>,海洋浮标通过内置装载的多种传感器准确记录波浪的运动信息,通常被视为衡量其他海浪周期测量方法的标准。但浮标测量法仍然存在许多缺点,如高昂的生产、置放成本,有限的浮标置放点,因此不能获取大面积的海浪特征。基于谱分析的非直接检测法有不同的数据来源。随着遥测技术的发展,遥感卫星逐步被运用到海浪要素测量中<sup>[8-11]</sup>,分析遥感卫星得到的数据,利用经验函数对海浪要素(如有效浪高、周期等)进行反演计算。根据遥测数据来源的不同,反演方法可分为基于卫星高度计<sup>[8]</sup>和雷达数据<sup>[9-10,12]</sup>的反演方法。Wyatt 等<sup>[9]</sup>通过高频雷达来获取海浪的方向谱信息,从而反演计算海浪参数。相比于浮标测量法,此方法可在 120 km 的范围内对超过 1 m 的波高实现高精度测量,而周期和方向参数受雷达噪声的影响较大。Hwang 等<sup>[8]</sup>提出根据卫星高度计导出的海浪有效浪高和风速,使用半经验函数计算周期,该方法得到的谱峰周期和平均周期与波浪浮标得到的数据高度吻合,但

其计算量庞大,对设备的硬件要求高。Shao 等<sup>[10]</sup>则根据合成孔径雷达(SAR)影像,同样使用半经验函数并根据有效波高和截止波长推导平均周期,并将所得结果与浮标测量结果进行对比。该方法计算出的平均周期的均方根误差为 1.98 s。

近年来人们提出了计算机视觉与波浪检测相结合的方法,近岸相机可以为沿海岸管理<sup>[13-14]</sup>、海岸线检测和沿海地区监控提供可靠的数据来源<sup>[15-17]</sup>,其拍摄的海浪图像或视频中包含大量海浪特征的有效信息,可以用于进一步检测海浪要素。在文献<sup>[18]</sup>中,相机被用来拍摄近岸图像,通过分析图像中的相关信息进一步实现沙滩的管理。2013 年,李刚等<sup>[19]</sup>提出用灰度共生矩阵提取图像的纹理信息以识别水面波浪等级。Piepmeier 等<sup>[20]</sup>提出一种基于立体视觉检测波浪特征的方法,Shi 等<sup>[21]</sup>也提出基于立体视觉检测海浪周期和浪高的方法。2019 年,Osorio 等<sup>[22]</sup>提出一种将海浪堆叠图像中的像素强度在时间序列上的变化转化为频谱,进而通过频谱分析估算出谱峰周期的方法,他们建立了像素强度光谱与测得的海浪频谱之间的映射关系。然而,该方法对海浪能量较低(周期小于 4 s)或较高(周期大于 10 s)区域的周期检测误差较大。随着机器学习的不断发展,深度学习神经网络<sup>[23]</sup>具有可在高维数据中发现复杂结构的强大能力,被广泛应用于视频的时空特征提取中。2015 年,Tran 等<sup>[24]</sup>基于三维卷积神经网络(C3D)提取视频的时空特征以识别并分类人的行为,并证明了该网络相对二维卷积的优越性。三维卷积神经网络能够同时提取时间与空间信息,近年来一直被广泛应用于视频识别与检测<sup>[25-28]</sup>。Zhang 等<sup>[28]</sup>将 3D 卷积神经网络与卷积长短时神经网络(LSTM)相结合,通过从视频中学习时空特征进行手势识别。ConvLSTM<sup>[29]</sup>则可以在提取时间特征的同时完成空间特征学习,被应用于视频情感内容分析<sup>[30]</sup>和视频对象检测<sup>[31]</sup>。深度学习神经网络也逐渐应用于海浪检测,郑宗生等<sup>[32]</sup>将深度学习与波浪检测相结合,使用 CNN 提取波特征对波高进行分类,而宋巍等<sup>[33]</sup>提出了一种用 NIN 网络从视频图像中检测海浪浪高的方法,并在北海小麦岛近岸海域取得较好的检测结果,该方法针对二级以上的海浪检测的相对误差低于 20%,满足业务化检测需求。Buscombe 等<sup>[34]</sup>通过神经网络估算单个图像的波高和周期。

本文提出了一种时空协同分析的近岸视频海浪周期检测方法,其主要思想借鉴了视频中运动识别

的时间维度特征表达<sup>[34]</sup>,通过构建由两个不同维度的卷积构成的复合卷积网络,从而实现时间和空间维度融合的海浪周期分析。首先,采用二维卷积的方法分别从视频帧中提取图像的高维空间特征,再将这些空间特征组成时间序列并进行一维卷积,得到融合的时空特征,最后采用注意力机制<sup>[35]</sup>对融合的特征进行权重调整,并通过回归训练建立特征与海浪周期的映射。本文通过两个比较实验验证了该方法的有效性。这两个比较实验为:1)空间特征和时间-空间特征融合的对比如, VGG 网络只提取空间特征,将其与所提出的基于时空融合特征的复合卷

积神经网络(CNN-2D1D)进行对比;2)针对时空特征学习和融合的方法,并将其与 ConvLSTM 和 C3D 网络进行了对比实验。

## 2 基本原理

### 2.1 方法流程

波浪周期检测的整个过程如图 1 所示,分为三个部分:

1) 数据预处理。实验从海浪监控视频中获取图像帧并完成裁剪、降噪等处理,使其适合作为检测网络的输入。

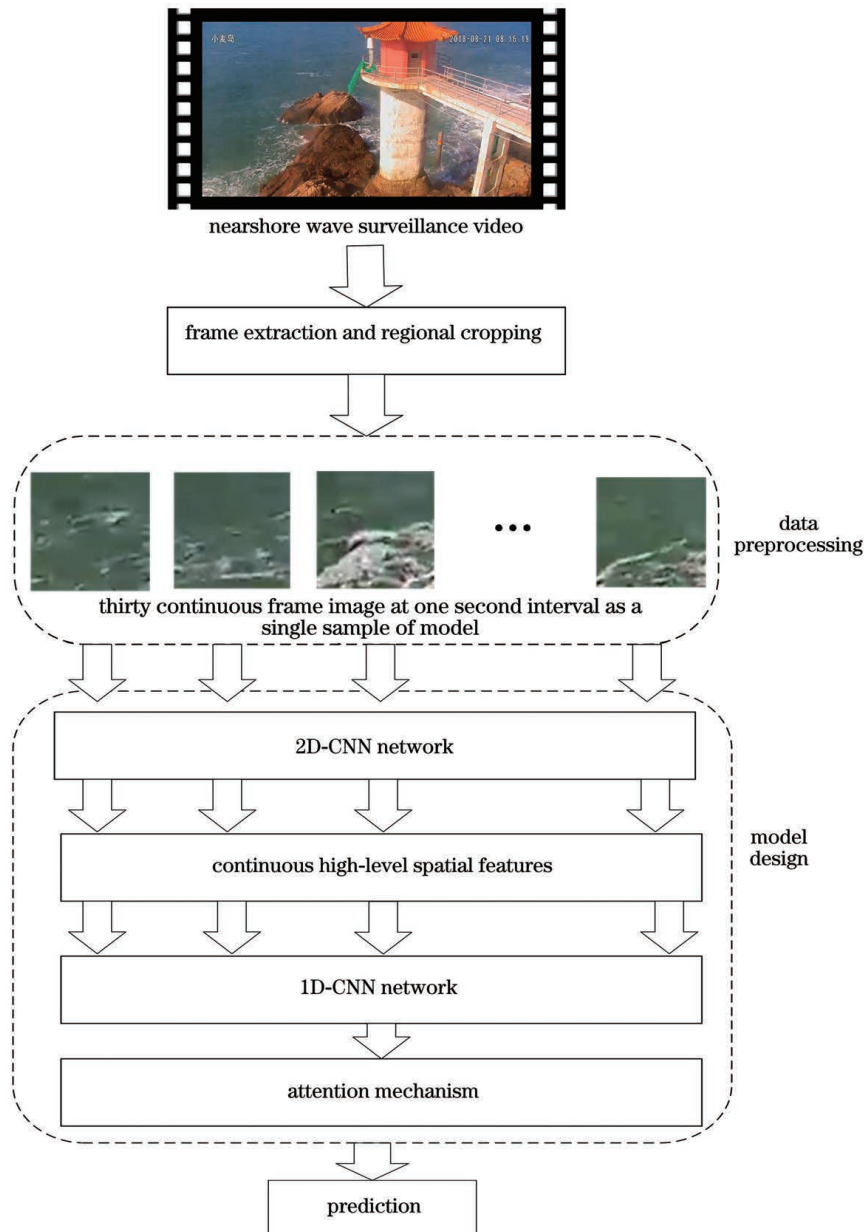


图 1 海浪周期检测模型框架  
Fig. 1 Framework of wave period detection model

2) 特征学习网络。从输入数据中提取出波浪周期时空特征的高层表达。

3) 周期预测。通过注意力机制对提取的高层特征进行权重调整,然后通过全连接层将所得结果线性映射为海浪周期,从而获得周期值。

### 2.2 数据预处理

实验使用的视频数据来源于中国黄海 X 岛监测站 2019 年 8 月 19 日至 23 日的监控视频,某些时段没有数据,视频数据包含 6:00—16:00 时间范围内的数据。每段视频的时长为 1 h,对应一个有效海浪周期,共收集了 49 个视频,这些视频能够匹配 30 个不同的有效海浪周期。视频所对应的周期真值来源于相同地点的浮标监测数据。

在进行周期检测时,由于周期的变化需要在较长的时间序列上体现,因此时间维度的特征表达通

过连续帧序列来表示波浪周期的变化细节。文中为了获取连续的波的细节信息,以 1 s 为间隔提取连续帧图像,分别以 10 帧和 30 帧组成一个样本,以检验不同长度序列对周期检测精度的影响。提取出的图像经过裁剪(避开障碍物,只保留海浪部分)后大小为 64 pixel×64 pixel。为保证波浪周期数据的合理分布,将数据集进行筛选,最终得到 39200 个图像样本,将其分成 29400 个训练样本、4900 个验证样本和 4900 个测试样本。

根据对比实验中不同模型对输入样本的要求,将这些样本分为三个不同的数据集,如表 1 所示。其中,数据集 I 是以单个图像为一个样本;数据集 II 是以 30 个连续图像序列为一个样本;数据集 III 是以 10 个连续图像序列为一个样本。这些样本所对应的周期为 4.3~12.7 s。

表 1 数据集构建与分布

Table 1 Distribution and construction of datasets

Dataset	Number of samples (training set/validation set/test set)	Size of single sample	Wave period /s
Dataset I	29400/4900/4900	64×64×3	
Dataset II	28959/4459/4459	30×64×64×3	4.3-12.7
Dataset III	27979/3479/3479	10×64×64×3	

### 2.3 模型设计

本文的模型采用融合时空特征的复合卷积神经网络(CNN-2D1D),如图 2 所示。该模型主要由两部分组成:二维卷积神经网络(2D-CNN)和一维卷积神经网络(1D-CNN),二者分别学习空间和时

间维度上的特征。2D-CNN 采用与 VGG 网络<sup>[19]</sup>类似的结构来学习每张输入的二维图像的空间特征。2D 卷积提取与周期相关的空间特征,1D 卷积提取与周期相关的时间特征,二者可以分别表示为

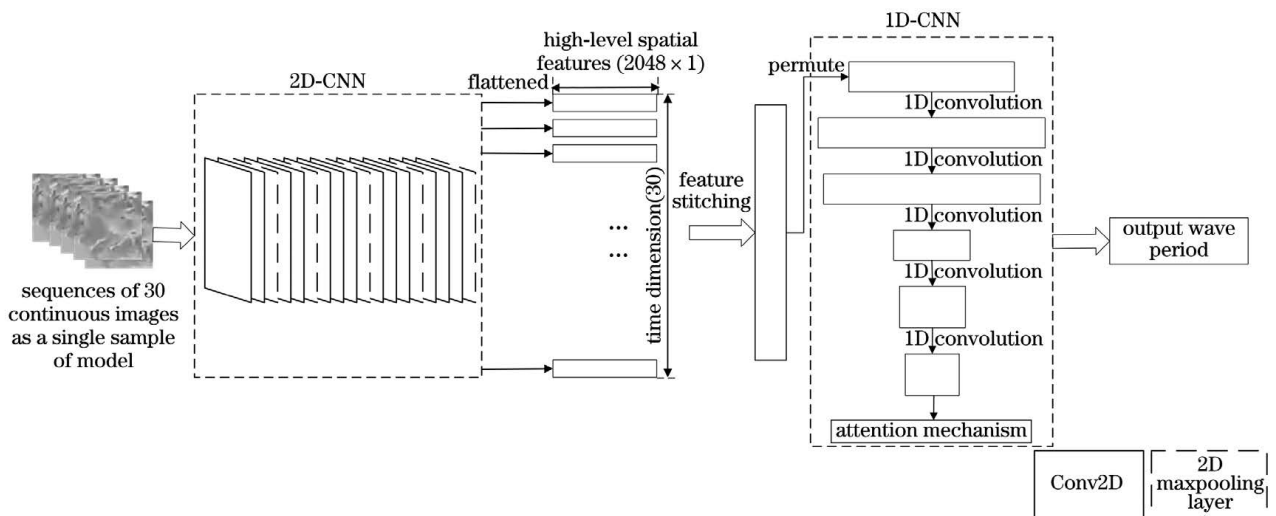


图 2 基于时空融合复合卷积神经网络的海浪周期检测模型

Fig. 2 Wave period detection model based on spatiotemporal fusion composite convolutional neural network

$$H(i, j) = F(i, j) * G(i, j) = \sum_m \sum_n F(m, n)G(i - m, j - n), \quad (1)$$

$$s(n) = f(n) * g(n) = \sum_{m=0}^{N-1} f(m)g(n-m), \quad (2)$$

式中： $F(i, j)$  表示像素点  $(i, j)$  处的输入特征； $G(i, j)$  表示大小为  $m \times n$  的卷积核； $H(i, j)$  表示卷积操作后的输出； $f(n)$  表示输入特征向量； $g(n)$  表示长度为  $n$  的一维卷积核； $N$  为特征向量  $f(n)$  的长度； $s(n)$  表示一维卷积操作后的输出。

2D-CNN 学习一个样本中的  $N$  ( $N = 10$  或  $30$ ) 张连续图像, 获得对应的  $N$  ( $N = 10$  或  $30$ ) 个海浪的空间特征, 随后将提取的空间特征在时间序列上合并, 再输入到 1D-CNN 中学习连续的时间特征。然后, 一维 CNN 在  $n$  个帧上执行卷积运算, 以实现空间和时间特征的融合。其特征提取过程为

$$\begin{cases} I = \{I_1, I_2, \dots, I_t, \dots, I_{30}\} \\ J_t = \text{Fla}[H(I_t)] \\ J = \{J_1, J_2, \dots, J_t, \dots, J_{30}\} \\ y = s(J^T) \end{cases}, \quad (3)$$

式中： $I$  表示大小为  $(30, 64, 64, 3)$  的单个输入样本； $I_t$  为样本中的第  $t$  张图像, 大小为  $(64, 64, 3)$ ； $H(I_t)$  表示 2D-CNN 提取的输入样本的第  $t$  张图像的高维特征, 是一个大小为  $(2, 2, 512)$  的特征矩阵,  $H(I_t)$  经过  $\text{Fla}(\cdot)$  操作被展平成大小为  $(2048, 1)$  的一维特征向量  $J_t$ , 单个样本中的 30 个一维特征向量  $J_t$  拼合成二维特征矩阵  $J$ , 经转置构成 30 行 2048 列的二维特征矩阵  $J^T$ ,  $y$  表示特征矩阵  $J^T$  经过一维卷积网络学习后输出的大小为  $(668, 128)$  的时空特征矩阵。

采用注意力机制对融合的时空特征进行权重分配, 最后通过全连接层建立时空特征与海浪周期的映射。用于空间特征学习的 2D-CNN 包含 5 个卷积段, 每个段有 2~3 个卷积层 (Conv2D)。每个段之间通过最大池化层连接 (2D Maxpooling)。所有卷积核的大小为  $3 \times 3$ , 池化窗口大小为  $2 \times 2$ 。5 个卷积段的卷积核个数分别为 64, 128, 256, 512, 512。用于提取时间特征的 1D-CNN 包含 4 个一维卷积层和 1 个池化层。前两个卷积层有 64 个卷积核, 其大小为 5, 两个卷积层后为最大池化层, 其池化窗口大小为 3, 最后两层同为卷积层, 卷积核个数为 128, 卷积核大小为 10。

在本文中使用时空注意力机制对与周期相关性较强的时空特征赋予较大权重, 对与周期相关性较弱的时空特征分配较小权重, 其过程可表示为

$$\begin{cases} w_i = \text{softmax}[\text{Dense}(y_i)] \\ y_{\text{out}} = y_i \odot w_i \end{cases}, \quad (4)$$

式中： $y_i$  为一维卷积神经网络提取的时空特征  $y$  的第  $i$  列特征向量。经过  $\text{Dense}(\cdot)$  全连接的操作后, 对应真值周期根据相关性赋予时空特征不同的注意力值。得到的注意力值经过  $\text{softmax}$  函数被转化为注意力分布  $w_i$ , 将  $y_i$  与  $w_i$  进行哈达玛积, 得到一个注意力向量  $y_{\text{out}}$ , 其可作为下一个全连接层的输入。通过注意力机制调整权重, 则 CNN-2D1D 能够更有效地学习时空特征。

模型采用平均绝对误差 (MAE) 作为损失函数, 采用 Adam 作为优化器, 学习率设置为 0.0001, 每轮迭代的衰减步长为 0.1。

### 2.4 对比模型设计

ConvLSTM 神经网络由 Shi 等<sup>[29]</sup> 在 2015 年提出, 文中首次将卷积神经网络 (CNN) 与 LSTM 相结合, 将传统的 LSTM 一维输入张量扩展为三维张量, 形成 ConvLSTM 层, 文中的网络通过叠加 4 个 ConvLSTM 层形成编码网络和预测网络。随后通过该网络提取雷达图像数据的时空特征, 实现局部地区降水量的预测。海浪周期是在时间上有较强相关性的海浪特征, 近岸海浪监控视频中同时包含了近岸海浪的时间、空间信息, 实验中用 ConvLSTM 搭建提取时空特征融合特征的网络进行回归训练, 从而检测海浪周期。其网络结构如图 3 所示, 用三层 ConvLSTM 层提取输入样本中的时空特征, ConvLSTM 层的卷积核大小分别为 32, 64, 64。为防止网络过拟合, 分别在第二层、第三层 ConvLSTM 后加入 Dropout 层, 将丢弃率设置为 0.5, 随后通过全连接层将时空特征与周期进行映射并进行回归训练。模型的初始学习率设置为 0.0001, 每轮迭代的衰减步长为 0.1。模型采用 Adam 优化器进行训练, 采用 MSE 作为损失函数。

VGGNet<sup>[36]</sup> 是由牛津大学 Visual Geometry Group (VGG) 研究组提出的卷积神经网络模型, 该网络用连续的小卷积核 ( $3 \times 3$ ) 代替传统卷积神经网络 (如 AlexNet) 的大卷积核 ( $5 \times 5, 7 \times 7, 11 \times 11$ ), 能够充分提取图像的空间细节信息, 并被逐步应用于视频的空间特征提取<sup>[37-38]</sup>。VGG16 包含 5 个卷积段共 13 个卷积层, 第 1、2 个卷积段分别包含 2 个卷积层, 2 个卷积段对应的卷积核个数分别为 64 和

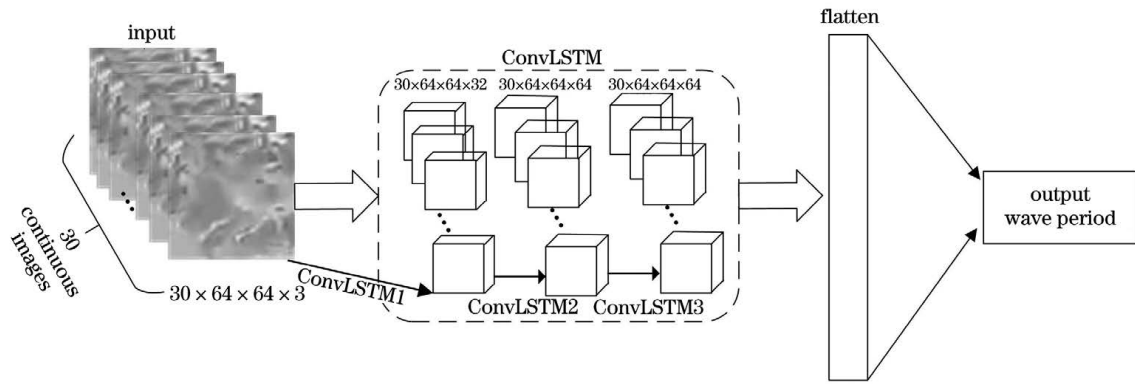


图 3 用于海浪周期检测的 ConvLSTM 模型

Fig. 3 ConvLSTM model for wave period detection

128,剩下 3 个卷积段都包含 3 个卷积层,对应的卷积核个数分别为 256,512,512,每个卷积段之间由最大池化层相连接,卷积层之后为 3 个全连接层,前两层全连接层的输出维度均为 4096,最后一层的输出维度为 1000,用于分类训练。鉴于 VGGNet 对空间细节特征的精确提取与良好的泛化能力,本文中 VGG16 最后一层的输出维度为 1,选用 ReLu 作为激活函数,用于提取海浪的空间细节特征,进行海浪周期检测。实验中将原网络的 sgd 优化器调整为 Adam,模型收敛至最优,初始学习率设置为 0.0001,每轮迭代的衰减步长为 0.1。将该网络与提取时空特征的复合卷积神经网络作对比。

近年来,三维卷积神经网络也被应用于视频的时空特征提取,并取得有效的成果。文献[24]中,作者提出一个简单的三维卷积神经网络(C3D)用于视频时空特征的提取、实现视频分类,并验证了 C3D 网络在时空特征提取上优于二维卷积神经网络,其结构简单,易于训练,取得了较高的分类精度。基于近岸海浪监控视频的海浪周期检测同样需要分析视频的时空特征,因此本文也研究了 C3D 网络结构,采取与文献[24]中相同的网络结构。由 8 个卷积层、5 个池化层和 3 个全连接层组成,卷积核大小均采用  $3 \times 3 \times 3$  的小卷积核,第一、二层卷积核数量分别为 64,128,二者之间用最大池化层连接,对于剩下的 6 层,每两层为一个卷积段,卷积段与卷积段之间由最大池化层连接,池化窗口的大小均为  $2 \times 2 \times 2$ ,池化步长均为 2。卷积层之后通过最大池化层与全连接层相连接,与 VGG16 和 ConvLSTM 相同,采用 Adam 优化器,初始学习率设置为 0.0001,每轮迭代衰减步长为 0.1。

### 3 实验结果与分析

通过一组对比实验,评估本文提出的 CNN-2D1D 模型针对近岸海浪周期检测的性能。由于本文方法基于时空特征学习来表示海浪周期的特征,因此在两种情况下将本文方法与不同的神经网络模型进行对比:

1) 验证时空融合特征是否为周期检测的重要特征。本文采用 VGG16 网络作为对比,该网络的输入为单帧图像,仅提取空间特征。

2) 验证所提出的时空特征学习网络在海浪周期检测中的有效性与实用性。将本文方法采用不同时空特征融合策略的 ConvLSTM 网络和 C3D 网络进行对比。此外,本文通过改变输入序列的长度(10 和 30),验证了不同输入序列长度对模型精度的影响。

#### 3.1 实验结果对比

2.2 节已经介绍了实验所使用的训练和测试数据,为检验时空特征对周期检测的有效性,本文分别设计了仅提取空间特征的 VGG 网络和提取时空特征的 ConvLSTM 网络、CNN-2D1D 网络和 C3D 网络。根据模型对输入要求的不同,此对比实验需用到数据集 I 和 III;为检验本文所提出的时空特征提取方法的有效性,采用同样是提取时空特征的 C3D 网络和 ConvLSTM 网络作为对比网络,此对比实验在数据集 III 上进行;此外,为研究时间序列长度对模型精度的影响,分别设计输入图像序列为 10 和 30 两种长度的模型进行训练和测试,该过程涉及数据集 II 和 III。模型性能的评价指标包括 MAE 和均方误差(RMSE),MAE 用于计算预测值与真值之间的绝对差值,RMSE 用于衡量预测值与真值之间的偏差。实验将本文所提出的复合网络

CNN-2D1D在两种情况下进行了分析:一是判断时空特征的融合是否对海浪周期的检测有效,对比网络是 VGG16<sup>[36]</sup>,因只针对空间特征进行学习,故输入数据选为单帧图像;二是检验时空融合的方法对海浪周期检测的影响,输入数据为连续的图像序列,对比网络是 ConvLSTM 和 C3D。ConvLSTM 与 C3D 网络的训练数据与 CNN-2D1D 复合网络一致。

表 2 给出了各模型在对应测试集上的周期检测的结果。首先,不考虑 ConvLSTM 与 CNN-2D1D 网络模型的差异性,对比两个模型受不同时间序列长度的影响,可以看出较长的时间序列(30 s)比短时间序列(10 s)的误差要小,检测精度更高。本实验中海浪周期的范围是 4~12 s,部分周期超过 10 s,10 s 长度的时间序列可能无法完全捕捉足够的周期信息,因而导致检测精度不高。30 s 长度的时间序列能够涵盖足够多的周期时间信息,同时也能避免时间序列长度太长导致的计算效率低的问题。因此,下文讨论的 CNN-2D1D 与 ConvLSTM 网络默认序列长度为 30。

表 2 不同模型在测试集上的海浪周期检测结果

Table 2 Wave period detection results of different models on test set

Model	Dataset	MAE	RMSE
VGG16	Dataset I	0.93	1.41
ConvLSTM10	Dataset II	1.08	1.57
CNN2D1D-10	Dataset II	1.21	1.66
ConvLSTM30	Dataset III	0.82	1.13
CNN2D1D-30	Dataset III	0.48	<b>0.66</b>
C3D	Dataset III	<b>0.47</b>	0.81

基于长度为 30 的序列的 ConvLSTM 网络(MAE 为 0.82, RMSE 为 1.13)的测试结果均优于仅基于单帧图像训练的 VGG16 模型(MAE 为 0.92, RMSE 为 1.98),这说明了采用时空特征融合的模式预测效果优于采用单一空间特征的模式。

在测试集上,C3D 网络(MAE 为 0.47, RMSE 为 0.81)和 CNN-2D1D 网络(MAE 为 0.48, RMSE 为 0.66)显示出比 ConvLSTM 网络(MAE 为 0.82, RMSE 为 1.13)更好的预测结果。所适用的数据集上的海浪周期真实值是通过计算特定时间段内经过某个区域的海浪个数,然后将该特定时间段除以海浪个数而获得的。结合测试结果可以推断出,海浪之间的长期相关性较小,在特定区域中每个波从产生到消失的过程是独立的。因此,LSTM 同时提取海浪的短期和长期的时间特征会降低其对海浪周期检测的准确性。

最后对比不同类型网络之间的差异,可以看出 C3D 网络的周期检测误差最小(MAE 为 0.47)。本文所提出的使用长度为 30 的序列的 CNN-2D1D 网络(MAE 为 0.48)的检测误差略高于 C3D,而本文模型的稳定性(RMSE 为 0.66)比 C3D 模型(RMSE 为 0.81)好。下文进一步评估了 C3D 网络与 CNN-2D1D 网络的稳定性、泛化性和模型的大小。

### 3.2 应用

为了进一步比较 C3D 模型和 CNN-2D1D 模型的泛化性能,分别选取 2018 年 8 月 22 日和 23 日 8:00—17:00 的连续视频,周期范围分别是 4.3~6.4 s 和 10.4~12.7 s。用训练好的模型对视频每隔 5 min 检测一次,将 1 h 的检测结果取平均,并将所得结果与该时间段的周期真值进行比较,检测结果如图 4 和表 3 所示。

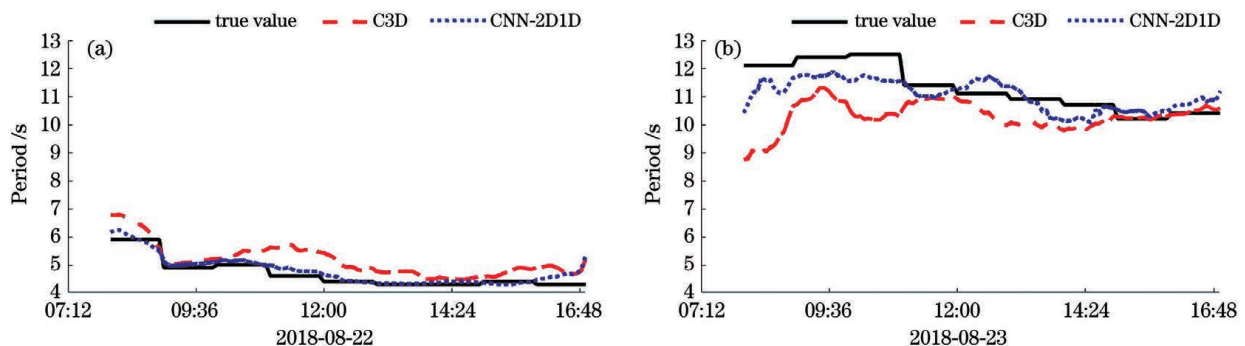


图 4 CNN-2D1D 模型和 C3D 模型在 2018-08-22 和 2018-08-23 的预测曲线图。(a)2018-08-22;(b)2018-08-23

Fig. 4 Prediction curves of CNN-2D1D model and C3D model on August 22, 2018 and August 23, 2018. (a) 2018-08-22; (b) 2018-08-23

表 3 C3D 模型和 CNN-2D1D 模型在 2018 年 8 月 22 日和 23 日的检测误差

Table 3 Detection errors of C3D and CNN-2D1D models on August 22, 2018 and August 23, 2018

Error	CNN-2D1D		C3D	
	August 22, 2018	August 23, 2018	August 22, 2018	August 23, 2018
MAE	<b>0.18</b>	<b>0.99</b>	0.46	1.01
RMSE	<b>0.26</b>	<b>1.21</b>	0.53	1.31

由图 4 可知, CNN-2D1D 模型检测值与周期真值的吻合度最高, 检测值拟合曲线较稳定, 波动较小。C3D 模型在 2018 年 8 月 22 日的检测值整体明显高于真值, 而在 2018 年 8 月 23 日的检测值又明显低于真值。从表 3 中给出的评价指标来看, CNN-2D1D 模型的 MAE 和 RMSE 的值均低于 C3D 模型, 证明 CNN-2D1D 模型的泛化性能更好, 更适用于近岸海浪周期检测。

在实际应用中, 计算量小且占内存小的模型更具实用性。比较 CNN-2D1D 模型和 C3D 模型的大小, 结果如表 4 所示。CNN-2D1D 模型训练所用的参数(15,019,202)远少于 C3D 模型训练所用的参数(52,830,593), 且 CNN-2D1D 模型的大小(173 MB)不足 C3D 模型(605 MB)的 1/3。因此, CNN-2D1D 模型在训练过程中能更快地收敛, 从而能较快地检测海浪周期值。

表 4 C3D 模型和 CNN-2D1D 模型的大小对比

Table 4 Comparison of model sizes of CNN-2D1D and C3D models

Model	CNN-2D1D	C3D
Parameters	15, 019, 202	52, 830, 593
Model size /MB	173	605

综上所述, 本文提出的 CNN-2D1D 对于波浪周期检测是有效的。CNN-2D1D 模型在使用较少的训练参数让模型更简单的同时, 得到了较好的检测结果和稳定性。

## 4 讨 论

通过将海浪的时间维度信息转化为海浪频谱来计算海浪周期的方法<sup>[22]</sup>, 在低能海浪区和高能海浪区的检测精度较低。而文献[39]中方法通过数值计算推导近岸海浪周期的预测公式, 其计算量大且复杂, 不易于实施。本文利用可以在时空维度上提供连续信息的近岸海浪监控视频, 通过 CNN-2D1D 网络从视频图像中自动提取与海浪周期相关的时空信息。三维卷积对时空特征有良好的学习能力, 且一

直被广泛应用于视频的时空特征提取<sup>[25-27]</sup>, 因此本文采用 C3D 网络作为对比网络, 分析 C3D 网络与 CNN-2D1D 网络在测试集上的检测结果, 以验证 CNN-2D1D 网络的合理性。

对比仅提取空间特征的 VGG 网络的实验结果, 采用时空融合特征的检测结果要好于仅采用空间特征的检测结果。其主要原因是海浪周期是与时间密切相关的海浪要素。对于用于提取时空特征的不同模型, C3D 网络结构简单且其在测试集上的平均绝对误差较小, 但其在训练过程中使用了更多参数, 并且泛化性较差。C3D 网络使用三维卷积核同时从时间、空间维度提取特征, 相比于二维卷积, C3D 网络保存了短时的时间维度特征信息, 网络模型简单, 易于学习, 因此在训练集上收敛较快, 测试集上效果也较好, 但同时产生了更多的训练参数。ConvLSTM 的检测结果不如 CNN-2D1D 和 C3D。与这两个网络相比, ConvLSTM 采用了一种更为复杂的方法来提取时间特征, 从而为模型添加了冗余信息。本文提出的 CNN-2D1D 模型首先使用二维卷积提取空间特征, 可以捕捉到更细节的空间特征, 然后使用简单的一维卷积提取时间维度特征, 避免了时间和空间特征融合时潜在的冲突和抵消。在最后一个全连接层之前添加了一个注意力机制, 以使模型更加关注有效特征。本文提出的模型在取得较小检测误差的同时, 保证了模型的泛化性能。因此, 本文提出的时空特征提取方法适用于波浪周期检测。

实验中的检测精度或受光照强度影响, 从 2018 年 8 月 22 日的检测结果可以看出, 16:00 之后的检测精度明显低于 7:00—16:00 之间的检测精度。分析得到光照强度影响视频的质量, 从而直接影响模型的输入结论。

## 5 结 论

针对近岸海浪周期检测, 从一个全新的角度提出了一套可行的研究方法, 建立了 CNN-2D1D 网络, 分别用二维卷积神经网络和一维卷积神经网络



提取近岸海浪监控视频的空间和时间特征,对海浪周期进行回归训练,以实现海浪周期检测,同时通过对比实验验证了本模型较好的泛化性能,证明该模型适用于近岸海浪周期检测。采用易于获取且包含了丰富的海浪运动信息的近岸海浪监控视频为数据来源,弥补了传统遥测技术成本高、覆盖面有限的缺点;同时采用训练深度学习模型进行周期自动检测,模型具有较好的实际应用性。

但是,采用 CNN-2D1D 模型进行海浪周期检测仍然存在一些问题,视频数据分布不均匀导致部分周期段(10s 以上)的视频数据量不足,因此出现了 CNN-2D1D 模型在部分周期的检测误差较大的现象;同时一天的不同时刻光照不同,视频质量受明暗度的影响,这为网络的特征提取增加了噪声,模型的测试结果也受到影响。

### 参 考 文 献

- [1] Seo S, Park S, Koo B. Effect of wave periods on added resistance and motions of a ship in head sea simulations[J]. *Ocean Engineering*, 2017, 137: 309-327.
- [2] Mortlock T R, Goodwin I D, McAneney J K, et al. The June 2016 Australian east coast low: importance of wave direction for coastal erosion assessment[J]. *Water*, 2017, 9(2): 121.
- [3] Group T W. The WAM model: a third generation ocean wave prediction model[J]. *Journal of Physical Oceanography*, 1988, 18(12): 1775-1810.
- [4] Tolman H L, Balasubramanian B, Burroughs L D, et al. Development and implementation of wind-generated ocean surface wavemodelsat NCEP [J]. *Weather and Forecasting*, 2002, 17(2): 311-333.
- [5] Booij N, Ris R C, Holthuijsen L H. A third-generation wave model for coastal regions: 1. Model description and validation[J]. *Journal of Geophysical Research: Oceans*, 1999, 104(C4): 7649-7666.
- [6] Benoit M, Marcos F, Becq F. Development of a third generation shallow-water wave model with unstructured spatial meshing[C]//25th International Conference on Coastal Engineering, September 2-6, 1996, Orlando, Florida, USA. Reston, VA, USA: American Society of Civil Engineers, 1996: 465-478.
- [7] Helmreich S. Reading a wave buoy [J]. *Science, Technology, & Human Values*, 2019, 44(5): 737-761.
- [8] Hwang P A, Teague W J, Jacobs G A, et al. A statistical comparison of wind speed, wave height, and wave period derived from satellite altimeters and ocean buoys in the Gulf of Mexico region[J]. *Journal of Geophysical Research: Oceans*, 1998, 103(C5): 10451-10468.
- [9] Wyatt L R, Green J J, Middleditch A, et al. Operational wave, current, and wind measurements with the Pisces HF radar[J]. *IEEE Journal of Oceanic Engineering*, 2006, 31(4): 819-834.
- [10] Shao W Z, Zhang Z, Li X F, et al. Ocean wave parameters retrieval from sentinel-1 SAR imagery [J]. *Remote Sensing*, 2016, 8(9): 707.
- [11] Mumby P J, Green E P, Edwards A J, et al. The cost-effectiveness of remote sensing for tropical coastal resources assessment and management [J]. *Journal of Environmental Management*, 1999, 55(3): 157-166.
- [12] Wang X C, He D X, Liu B X. Research on SAR sea significant wave height inversion method based on ELM model[J]. *Chinese Journal of Lasers*, 2020, 47(7): 0710002.  
王晓晨, 贺东旭, 刘冰宣. 基于 ELM 模型的 SAR 海浪有效波高反演方法研究 [J]. *中国激光*, 2020, 47(7): 0710002.
- [13] Osorio A F, Medina R, Gonzalez M. An algorithm for the measurement of shoreline and intertidal beach profiles using video imagery: PSDM[J]. *Computers & Geosciences*, 2012, 46: 196-207.
- [14] Múnera S, Osorio A F, Velásquez J D. Data-based methods and algorithms for the analysis of sandbar behavior with exogenous variables[J]. *Computers & Geosciences*, 2014, 72: 134-146.
- [15] Valentini N, Saponieri A, Damiani L. A new video monitoring system in support of Coastal Zone Management at Apulia Region, Italy[J]. *Ocean & Coastal Management*, 2017, 142: 122-135.
- [16] Pérez Muñoz J C, Ortiz Alarcón C A, Osorio A F, et al. Environmental applications of camera images calibrated by means of the Levenberg-Marquardt method[J]. *Computers & Geosciences*, 2013, 51: 74-82.
- [17] Nalamati M, Sharma N, Saqib M, et al. Automated monitoring in maritime video surveillance system [C]//2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), November 25-27, 2020, Wellington, New Zealand. New York: IEEE Press, 2020: 1-6.
- [18] Jiménez J A, Osorio A, Marino-Tapia I, et al. Beach recreation planning using video-derived coastal state indicators[J]. *Coastal Engineering*, 2007, 54(6/7): 507-521.

- [19] Li G, Xiong Y Z, Liu K K, et al. Water wave detection based on texture characteristics[J]. *Application Research of Computers*, 2013, 30(4): 1226-1229.  
李刚, 熊亚洲, 刘康克, 等. 一种基于图像纹理特征的波浪检测方法[J]. *计算机应用研究*, 2013, 30(4): 1226-1229.
- [20] Piepmeier J A, Waters J. Analysis of stereo vision-based measurements of laboratory water waves[C]// *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, September 20-24, 2004, Anchorage, AK, USA. New York: IEEE Press, 2004: 3588-3591.
- [21] Shi L, Yang L, Zhu H H, et al. Measurements of wave characteristics based on binocular vision and Zero-crossing method[C]// *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, October 12-14, 2018, Chongqing, China. New York: IEEE Press, 2018: 613-616.
- [22] Osorio A F, Montoya-Vargas S, Cartagena C A, et al. Virtual BUOY: a video-based approach for measuring near-shore wave peak period[J]. *Computers & Geosciences*, 2019, 133: 104302.
- [23] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [24] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks[C]// *2015 IEEE International Conference on Computer Vision (ICCV)*, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4489-4497.
- [25] Yang H, Yuan C F, Li B, et al. Asymmetric 3D convolutional neural networks for action recognition[J]. *Pattern Recognition*, 2019, 85: 1-12.
- [26] Li J N, Zhang S L, Huang T J. Multi-scale 3D convolution network for video based person re-identification[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 8618-8625.
- [27] Huang Y K, Guo Y C, Gao C. Efficient parallel inflated 3D convolution architecture for action recognition [J]. *IEEE Access*, 2020, 8: 45753-45765.
- [28] Zhang L, Zhu G M, Shen P Y, et al. Learning spatiotemporal features using 3DCNN and convolutional LSTM for gesture recognition [C] // *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3120-3128.
- [29] Shi X J, Chen Z R, Wang H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [C] // *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, December 7-12, 2015, Montreal, Quebec, Canada. [S.l.: s.n.], 2015: 802-810.
- [30] Niu J W, Li S J, Mo S S, et al. Affective analysis for video frames using ConvLSTM network [C] // *2018 IEEE International Conference on Communications (ICC)*, May 20-24, 2018, Kansas City, MO, USA. New York: IEEE Press, 2018: 1-6.
- [31] Tang Y, Zou W B, Jin Z, et al. Multi-scale spatiotemporal conv-LSTM network for video saliency detection[C]// *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, Yokohama Japan. New York: ACM, 2018: 362-369.
- [32] Zheng Z S, Hao J B, Huang D M, et al. Nearshore wave grade video monitoring based on deep learning [J]. *Marine Environmental Science*, 2017, 36(6): 934-940.  
郑宗生, 郝剑波, 黄冬梅, 等. 基于深度学习的近岸海浪等级视频监控 [J]. *海洋环境科学*, 2017, 36(6): 934-940.
- [33] Song W, Zhou X, Bi F, et al. Automatic wave height detection from nearshore wave videos[J]. *Journal of Image and Graphics*, 2020, 25(3): 507-519.  
宋巍, 周旭, 毕凡, 等. 近岸海浪视频浪高自动检测 [J]. *中国图象图形学报*, 2020, 25(3): 507-519.
- [34] Buscombe D, Carini R J, Harrison S R, et al. Optical wave gauging using deep neural networks[J]. *Coastal Engineering*, 2020, 155: 103593.
- [35] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention [EB/OL]. (2014-06-24) [2020-11-25]. <https://arxiv.org/abs/1406.6247v1>.
- [36] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2020-11-25]. <https://arxiv.org/abs/1409.1556>.
- [37] Zhu Z, Sang Q B, Zhang H. No reference video quality assessment based on spatio-temporal features and attention mechanism [J]. *Laser & Optoelectronics Progress*, 2020, 57(18): 181509.  
朱泽, 桑庆兵, 张浩. 基于空时特征和注意力机制的无参考视频质量评价 [J]. *激光与光电子学进展*, 2020, 57(18): 181509.
- [38] Liu M J, Cao Y Z, Zhu S Y, et al. Feature fusion video target tracking method based on convolutional neural network [J]. *Laser & Optoelectronics*

Progress, 2020, 57(4): 041502.

刘美菊, 曹永战, 朱树云, 等. 基于卷积神经网络的特征融合视频目标跟踪方法[J]. 激光与光电子学进展, 2020, 57(4): 041502.

[39] Hofland B, Chen X X, Altomare C, et al. Prediction formula for the spectral wave period  $T_{m-1, 0}$  on mildly sloping shallow foreshores[J]. Coastal Engineering, 2017, 123: 21-28.