

基于卷积神经网络的多维度分布式光纤振动传感事件识别

靳喜博^{1,2,3}, 刘琨^{1,2,3*}, 江俊峰^{1,2,3}, 王双^{1,2,3}, 徐天华^{1,2,3}, 黄悦朗^{1,2,3}, 胡鑫鑫^{1,2,3}, 张冬琦^{1,2,3}, 刘铁根^{1,2,3}

¹天津大学精密仪器与光电子工程学院, 天津 300072;

²天津大学光电信息技术教育部重点实验室, 天津 300072;

³天津大学光纤传感研究所, 天津 300072

摘要 基于双马赫-曾德尔干涉(DMZI)型分布式光纤振动传感系统与无人机(UAV)视频监测系统,通过卷积神经网络同步对光信号和无人机视频信号进行模式识别,从多维度对多类别扰动事件进行精准检测。与传统的模式识别方法相比,所提方案将两个不同维度上的信号有效结合,实现了不同维度上模式识别方法的优势互补,将识别信号的时间维度加入识别,解决了静态信号识别事件有限、准确率较低的问题。为了验证所提方案的可行性和有效性,对常见的9种传感行为(攀爬、轰击、剪切、脚踢、重敲击、轻敲击、拉扯、摇晃、无入侵)进行了实验测试和分析。实验结果表明,所提出的多维度模式识别方案可以对9种入侵事件达到99.58%的平均测试准确率,并且平均识别时间为0.16 s,短于系统的采样时间0.3 s,满足实际工程应用的需求。

关键词 光纤光学; 分布式光纤传感; 多维度传感; 模式识别; 卷积神经网络

中图分类号 TN253 **文献标志码** A

DOI: 10.3788/AOS230944

1 引言

分布式光纤振动传感系统(DOFVS)作为一种新型的分布式传感系统,由于其具有实时性好、准确度高、鲁棒性强等优点,在近些年的研究中被广泛关注。DOFVS在结构健康监测、管道泄漏检测、周界安防等领域^[1-2]有广阔的应用前景,对早期问题的发现和预防具有重要意义。DOFVS的分辨率、监测距离、监测精度等性能随着解调算法的改进和系统结构的优化而提升。同时,随着深度学习和人工智能等技术的发展,DOFVS也逐渐向着智能化发展。为了在实际工程应用中实现对振动信号的自动识别和分类,进一步提升DOFVS的稳定性和可靠性,为工程提供更加准确和有效的监测结果,越来越多的光信号模式识别方案被提出^[3]。

干涉光信号在时域上是典型的非稳定信号,因此对于一维光信号的模式识别来说,最关键的是光信号特征提取和分类器选取。Liu等^[4]提出了一种基于手动特征提取方案并与支持向量机(SVM)和径向基函数(RBF)结合的模式识别方案,该方案通过构建由光信号的过零率、样本熵、小波包能量熵、峰度值和多尺

度排列熵组成的特征向量作为特征值,使用支持向量机作为分类器,在1.01 s内对5种事件的识别准确率达到97.10%。对于手动特征提取方式来说,选取的特征向量维度越多,识别的结果也就越准确,然而数据预处理时间的增加也导致识别方案的实时性较差。Bai等^[5]使用卷积、长短期记忆、全连接的深度神经网络模型(CLDNN),对误报事件和真实的扰动事件进行了区分,实现了97%的识别准确率。然而,该方案对信号的处理仍然局限在时域范围内,可识别的事件种类仅为3类。Aktas等^[6]通过对相敏型光学时域反射系统的光信号做短时傅里叶变换(STFT)获取信号的时频信息并对其进行差分处理,使用短时傅里叶变换后的实部、虚部和幅值构建三通道RGB图像,使用5层的卷积神经网络进行分类识别。由于使用的卷积神经网络的卷积层数较少,该方案对于模式较为相近的事件识别准确率只有93.47%。薛康等^[7]提出了基于YOLOv5s(You Only Look Once version 5s)模型的时频图像与无人机图像的入侵事件检测方案,该方案将事件的时频图像和无人机二维图像同时送入深度学习模型中,对5种传感模式的平均识别准确率为96.6%。该方案是采用目标检测算法对无人机图像

收稿日期: 2023-05-08; 修回日期: 2023-06-01; 录用日期: 2023-06-12; 网络首发日期: 2023-06-22

基金项目: 国家自然科学基金(61922061, 61735011, 61775161)、国家重大科学仪器设备开发专项(2013YQ030915)、天津市自然科学基金杰出青年科学基金(19JCJQJC61400)

通信作者: *beiyangkl@tju.edu.cn

进行处理,丢失了时间维度的信息,相当于对静态的动作进行分类,因此对于行为动作的静态帧较为相近的模式无法进行准确识别,这限制了其可识别事件数量。

为了弥补上述方案存在的不足,本文提出了一种基于卷积神经网络的多维度分布式光纤振动传感识别方案,同时使用光信号和无人机的视频信号从多维度对入侵事件进行判别。首先,由 DOFVS 接收振动信息的光信号,从中解调出入侵的位置信息,将入侵位置的 GPS(Global Positioning System)坐标传输给无人机系统,无人机系统对入侵地点的行为进行视频拍摄。其次,对包含入侵信息的光信号进行端点检测,提取具有入侵特征的高频信号片段,消除冗余数据,降低数据量;并通过短时傅里叶变换的方式将一维时域光信号转换为二维时频图像。最后,将二维时频图像和带有时间维度的三维无人机视频信号同时送入卷积神经网络进行分类识别,将二者识别的结果相乘并进行归一

化,输出最终的识别结果。本文所提出的多维度识别方案是对光信号和视频信号的综合判断,实现了两路信号同时输入,共同对同一件入侵行为进行描述,最后输出一个识别结果。实验测试结果表明,提出的多维度识别方案能够准确地对攀爬、轰砸、剪切、脚踢、重敲击、轻敲击、拉扯、摇晃、无入侵这 9 类常见的传感事件进行精准识别。相较于已有的模式识别方案来说,所提方案拓宽了 DOFVS 的可识别事件种类,并极大提升了识别的准确率。

2 基本原理

2.1 双马赫-曾德尔传感系统

分布式光纤振动传感技术主要分为前向干涉型和背向散射型。本文采用基于双马赫-曾德尔干涉仪(DMZI)的前向干涉型分布式光纤振动传感结构,该系统具有结构简单、传感距离长、频率响应范围大等优点^[8]。其具体结构如图 1 所示。

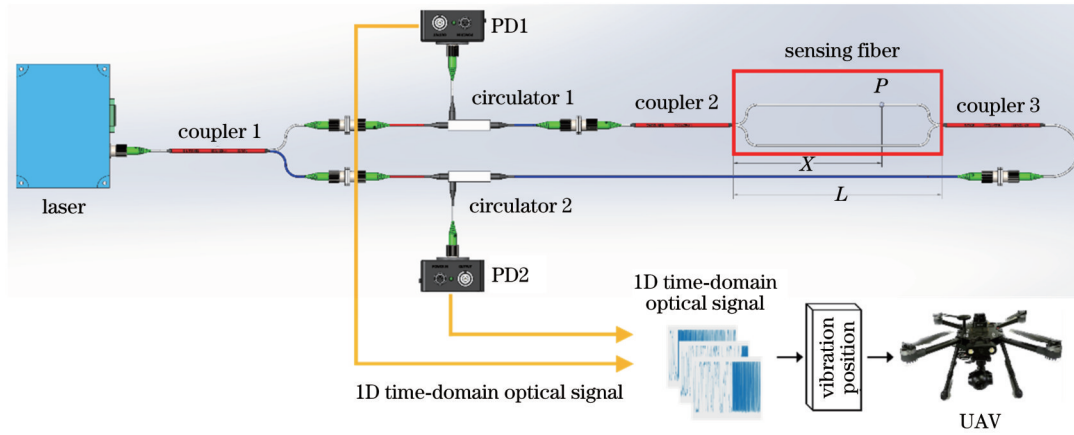


图 1 DMZI 系统结构原理图

Fig. 1 Schematic diagram of DMZI system

系统中采用的激光光源为分布反馈式激光光源,具有高单色性和长相干长度等优点,线宽小于 50 kHz,中心波长为 1550 nm。激光器输出的信号光由分光比为 50:50 的 3 dB 耦合器 coupler 1 分成两路传输:其中一路光信号通过环形器 circulator 1 传输并进入由耦合器 coupler 2 和 coupler 3 组成的 MZI,沿顺时针方向传输,最终被光电探测器(PD2)接收;另一路光信号则通过环形器 circulator 2 传输,同样进入耦合器 coupler 2 和 coupler 3 组成的 MZI 中,沿着逆时针方向进行传输,最终被 PD1 接收。高速采集卡 DAQ 对光电探测器输出的电信号进行采集,最后由工控机(IPC)对数据进行处理和分析。在两个 MZI 光纤干涉仪中,逆时针方向的干涉光和顺时针方向的干涉光都经过了振动信号的调制,但是由于传输方向不同,振动信号发生的相对位置也有所不同,因此两路信号之间存在着固定的时延差。假设传感光纤的总长度为 L ,振动信号作用在与 coupler 2 相对距

离为 x 的位置上,那么两个方向上的干涉光存在的时延差为 $\tau = t_{L-x} + t_L - t_x$,其中 t_{L-x} 、 t_L 和 t_x 分别为光在光纤中传输 $L-x$ 、 L 和 x 所需要的时间。根据时延差 τ ,可以计算出扰动位置的相对距离 x , x 可表示为

$$x = L - \frac{c \cdot \tau}{2n}, \quad (1)$$

式中: c 为光在真空中的传输速度, $c=3 \times 10^8$ m/s; n 为所用单模光纤的折射率。通过对两路振动信号进行互相关求得两者的时延差,进而得到振动的位置信息。利用光信号进行位置信息解调,利用干涉光信号进行传感模式判别。

2.2 无人机系统

城市监控视频信号的取得主要依赖于固定摄像头,固定摄像头可以在一定的范围内完成实时定点监控。在长距离分布式传感的应用背景下,无人机更适用于灵活机动地进行视频信号采集,并且可以自动对

异常事件实现干预。但是目前的无人机检测方案都是基于区域巡航实现的,面对突发事件时不能及时完成

点对点的探查。本文将其与 DMZI 相结合,具体的实现过程如图 2 所示。

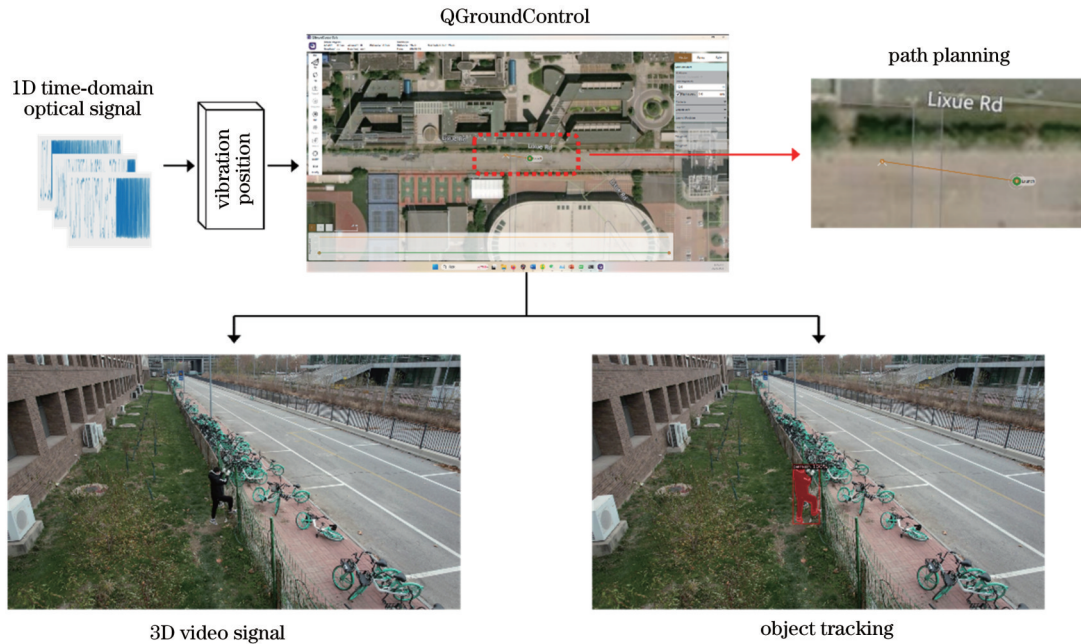


图 2 无人机系统的工作流程

Fig. 2 Workflow of unmanned aerial vehicle system

DMZI 的工控机与无人机地面站之间通过局域网进行连接。当振动事件发生时,工控机接收到存在一定时延的两路干涉信号,通过对光信号进行解调得到振动位置发生的相对位置 x ;再根据系统设定好的围栏布设位置,将 x 转换为无人机需前往的真实地理坐标;地面站读取坐标后自动规划路线,将任务自动上传至无人机;无人机到达目标地点后,使用其搭载的摄像头在振动位置对振动产生源进行视频拍摄,并将画面实时传输至地面站;若无人机发现非法入侵者,还可自动切换为目标跟踪模式,对入侵目标进行跟踪,直至危险消除;若并无异常事件,无人机将自动返航,继续执行巡航任务。本文所使用的地面站为 QGroundControl,该开源地面站经改写后,可以很好地与 DMZI 相结合,以实现无人机的自主控制。

2.3 光信号预处理

DMZI 系统接收到的干涉光信号是一维时域复杂信号,具有数据量大、噪声较多、变化速度快等特点。对于模式识别问题,目前主要有两大主流的解决方案:传统机器学习方案和深度学习方式。传统机器学习方案是通过手动构建特征向量的方式实现的,其分类准确度决定于构建的特征向量维度,维度越大则分类的准确率越高。然而,维度的增加也将导致预处理时间的大幅增加,在采样时间内无法完成对信号的识别^[9]。为了克服传统机器学习方案的不足,采用深度学习模型对信号进行自动特征提取的方案既可以很好地对特征进行保留,又可以提升检测效率。而对二维图像进

行分类的卷积神经网络近年来发展迅速,相对于一维卷积神经网络来说具有更好的特征提取能力和更高的鲁棒性^[10]。为此,需要将一维时域光信号转换成二维图像以进行分类识别。一维信号的二维化方法包括格拉姆角场变换^[11]、灰度图变换^[12]、短时傅里叶变换^[13]等。其中,短时傅里叶变换法可以将一维时域信号转换为二维时频信号,不仅完成了二维化的转换,还增加了信号的频域特征和功率分布特征,可以满足更多种类的信号识别需求。然而,短时傅里叶变换的时间取决于处理的信号数据量,为了能够提升检测效率,本文将选用文献[14]提出的基于中值滤波的端点检测方法提取振动信息特征丰富的高频片段。然后对高频片段进行短时傅里叶变换,具体实现过程如下所示:

$$F(\tau-f) = \int_{-\infty}^{+\infty} x(t) \omega(t-\tau_0) \exp(-i2\pi ft) dt, \quad (2)$$

式中: $x(t)$ 表示经端点检测后的高频信号片段; $\omega(t-\tau_0)$ 为窗函数; τ_0 为时移量; f 表示频率。图 3(a) 为一维时序光信号。本文选取汉明窗作为窗函数,窗函数长度为 2×10^4 ,重叠长度为 1.2×10^4 ,变换后的结果如图 3(b) 所示。由图 3(b) 可见,STFT 后的信号特征主要集中在功率较高的部分,功率较低的部分呈现出无规律的噪声,通过对无入侵信号进行 STFT 图像的分析,发现无入侵事件信号的功率主要分布在 30 dB 以下,因此将振动功率低于 30 dB 的部分设置为能量较低部分;当振动信号功率特别高时,由于其在整个二维时频矩阵中的占比较低,故将占比小于 0.01% 的部分设定

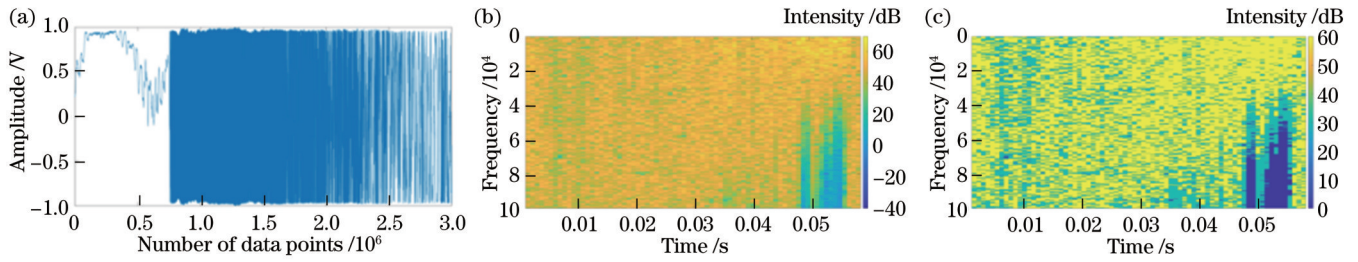


图3 重敲击信号图。(a)一维时序光信号;(b)STFT后二维时频信号图像;(c)功率分割后的STFT二维时频图像

Fig. 3 Knocking hard signals. (a) 1D time domain optical signal; (b) 2D time-frequency image after STFT; (c) 2D time-frequency image after STFT with power segmentation

为高能量部分(实验中选取 45 dB),其余为中能量部分,结果如图 3(c)所示,既保留了特征信息,也去除了噪声的影响。

3 多维度融合识别模型构建

经数据预处理后,光学信号转换为二维时频图像,而三维视频信号无需预处理步骤。由于两类信号的维度不同,因此也将选取不同维度的卷积神经网络模型进行检测。本节将分别介绍用于光信号分类的 Resnet 50 模型和用于视频信号识别的 SlowFast 模型,以及二者识别结果的融合方法和评价指标。

3.1 Resnet 50

在特征提取上,手动进行信号特征提取可能造成特征丢失等后果,相较于手动特征提取方法,深度学习算法的自动化特征提取方法可以从更多维度对信号进行特征提取,从而在入侵事件模式识别中扩充识别种类,增强系统的应用能力。对于需要识别的二维时频光信号来说,整张图片都包含特征信息。因此对其进行模式识别即对图像进行分类识别,由此模式识别问题转化为图像分类问题。在图像分类算法中,常用的模型有 VGGNet、GoogleNet、Resnet 等。因为卷积神经网络能够提取到低、中、高三种层次的特征,因此对应的网络层数越多,所提取到的特征信息也就越丰富,可识别的准确率越高。同时,越深的网络所提取的特征抽象性更强,具有更多的语义信息。为了避免线性结构的层数加深导致浅层参数无法更新的梯度消失^[15]问题,以及由深层结构的冗余导致过拟合的梯度退化^[16]问题,2015 年 He 等^[17]提出了 Resnet 残差结构。该模型已经广泛应用于医学图像诊断^[18]、图像分类^[19]、目标识别^[20]等领域。本文使用 Resnet 模型作为图像分类识别模型,对多入侵事件进行高准确度识别。

为了兼顾识别实时性和识别准确率,本文选择了具有 50 次卷积的 Resnet 50 模型,其具体结构如图 4 所示,其中: d_{Stride} 为步长; Conv1、Conv2D/3D 是卷积层,用于特征提取; BN 是批归一化层,可以加快模型收敛; ReLU 是激活函数,可增加模型的非线性程度; MaxPooling 是最大池化层,用于特征压缩; GAP 是全局平均池化,用于类别匹配; B1 和 B2 分别是带有下采

样步骤和不带有下采样步骤的 Bottleneck; $O_1 \sim O_9$ 分别表示 9 种传感事件(攀爬、轰砸、剪切、脚踢、重敲击、轻敲击、无入侵、拉扯、摇晃)的光信号识别置信度输出。

Bottleneck 是 Resnet 模型中最核心也是最基础的部分, Resnet 模型后的数字代表着卷积层与全连接层的层数之和,每个基本结构中的特征提取部分都包含三次卷积过程,如图 4(c)中的下支路所示。其中二维卷积核大小及通道数的具体参数如表 1 所示。对输入的特征图进行卷积运算后得到卷积输出结果,将卷积结果与输入特征图进行 add 运算。其特有的恒等映射非线性结构保留了卷积后的特征信息和原始特征信息,可以很好地解决梯度消失和梯度退化的问题,并且可以降低计算量、减少训练时间。由于输入的二维 STFT 图像在 Layer 1 中的输入特征通道数与经卷积后的输出特征通道数相同,均为 64 通道,因此在 Layer 1 中的 Bottleneck(B1)无须进行降采样,即步长 $d_{\text{Stride}} = 1$; 在 Layer 2、Layer 3 和 Layer 4 中,需要先使用 $d_{\text{Stride}} = 2$ 的 B1 进行降采样,保证输入特征通道数与输出特征通道数保持一致,这样才可以实现同通道数的 add 操作,以完成层与层之间的连接。利用卷积层完成对特征信息的提取后,通过全局平均池化输出一个具有 1000 类的全连接层。在本文的应用中,只对 9 种事件进行分类,而且入侵事件的二维时频图像与训练模型差别较大,因此在训练过程中不载入预训练参数,而是将整个网络进行训练。同时修改全连接层的输出个数为 9 个,实现种类的对应。

3.2 SlowFast

在近些年对视频行为检测的研究中,大量优秀视频检测网络涌现出来,主流的行为检测算法可以分为三种类型:行为识别、时序行为检测和时空行为检测。行为识别关注的是一段视频中的行为,即判断该段视频中发生的是什么事;时序行为检测是基于行为识别进一步确定在一段视频中事件发生的起始时间和结束时间;时空行为检测则不仅能够判断事件发生的类型和时间,还能够对产生该事件的物体位置进行定位。在无人机拍摄视频过程中,要求能够定位入侵者在视频中的位置并判断入侵者的入侵行为。因此,本文选择由 Facebook 团队提出的 SlowFast 的时空行为检测

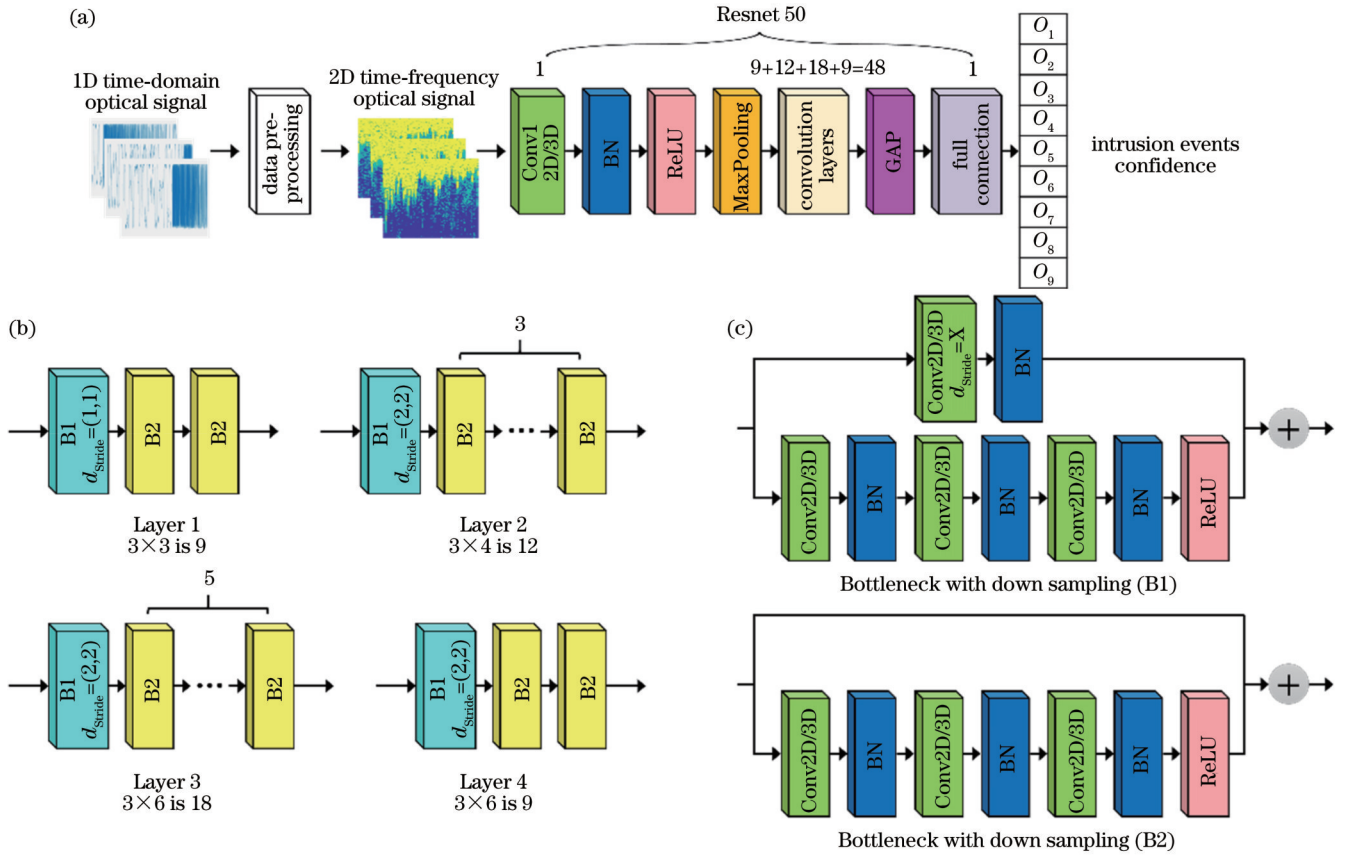


图 4 Resnet 50 模型。(a) 光信号识别总体结构; (b) 卷积层分布情况; (c) Bottleneck 结构

Fig. 4 Resnet 50 model. (a) Total structure of optical signal recognition; (b) distribution of convolution layers; (c) architecture of Bottleneck

表 1 Resnet 50 网络卷积层参数

Table 1 Parameters of each convolution layers in Resnet 50

Layer	Parameters $\{H \times W, C\}$
Conv1	$7 \times 7, 64$
Layer 1	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Layer 2	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Layer 3	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Layer 4	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$

模型对视频信号进行识别处理^[21]。该网络受人视觉原理的启发,在人眼的感知过程中,80%的视觉细胞用来感知环境中的静态环境信息,20%的视觉细胞用来

感知快速变化的动作信息。类似地,在 SlowFast 模型中也将视频信号的处理分为一条慢速路径(Slow)和一条快速路径(Fast)。Slow 路径用来对外界环境信息进行处理,Fast 路径则是对视频帧进行密集处理以获取视频中的运动信息。通过对两个通道的信号分别进行处理,可以在减少计算量的同时提升准确率。SlowFast 的具体结构如图 5 所示。图中 $V_1 \sim V_8$ 分别表示 8 种入侵事件(攀爬、轰砸、剪切、脚踢、重敲击、轻敲击、拉扯、摇晃)的视频信号识别置信度。

不同于二维卷积神经网络对图像特征的提取,视频流信息不仅包括空间尺寸 (x, y) , 还包括时间维度 t 的信息,因此卷积核也包含三个维度,即采用 3D 卷积核来进行卷积操作,所输出的卷积结果也是三个维度,3D 卷积核能够提取到更高维度上的特征。在 SlowFast 模型中,两条以不同速度处理的路径在特征维度上存在尺寸差异。假设图像的空间尺寸为 S^2 : 在 Slow 路径上,每隔 τ 帧给路径输入一次视频帧 ($\tau = 16$),其时间维度上的尺寸为 T ,特征通道数为 C ; 在 Fast 路径上对视频帧进行更为密集的采样,每隔 τ/α 帧给路径输入一次视频帧 ($\alpha = 8$),其时间维度上的尺寸为 αT ,由于 Fast 路径与 Slow 路径在时空维度上的尺寸不同,因此需要对 Fast 通道的特征通道数进行缩放,目的是使两路通道的运算速度保持一致,故

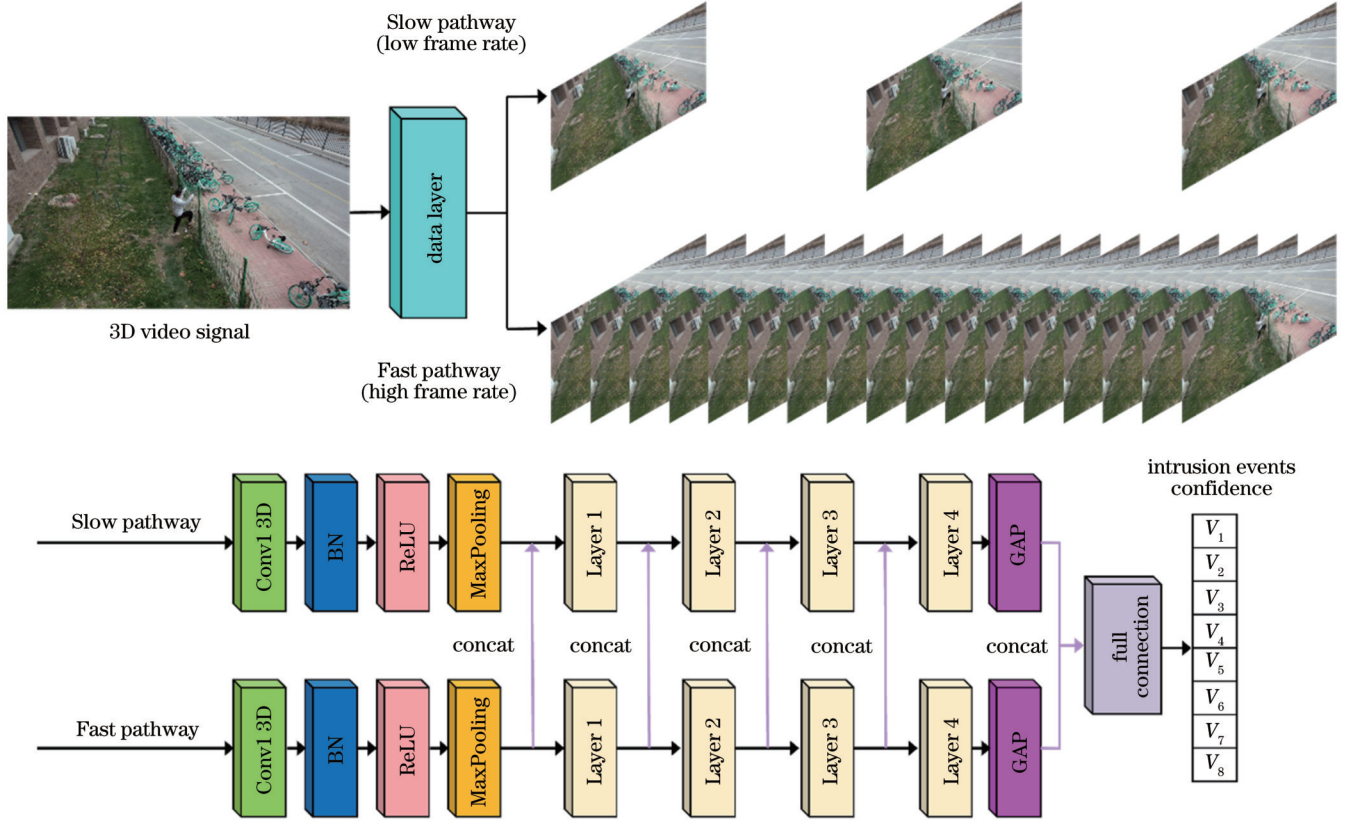


图 5 SlowFast 模型网络结构
Fig. 5 Schematic of SlowFast model structure

Fast通道的特征通道数设定为 βC ($\beta = 1/8$)。

SlowFast模型通过使用concat方式对两条路径的信息进行连接融合,由于两条路径的尺寸不同,Slow路径的尺寸为 $\{T, S^2, C\}$,Fast路径的尺寸为 $\{\alpha T, S^2, \beta C\}$,因此需要对两条路径进行尺寸匹配,在文献[21]中的实验选择对Fast通道进行时间维度卷积,使用卷积核为 5×1^2 的3D卷积进行尺寸匹配,输出通道为 $2\beta C$,步长为 α 。最后将融合后的特征图送入以Resnet3D-50为主干的特征提取网络,该网络结构与光信号处理的Resnet 50网络结构相同(图4),两网络的区别在于选取的卷积核维度不同,三维视频信号增加了时间维度,对应的卷积核维度也为三维,网络的具体参数如表2所示。

3.3 识别结果融合

光信号的分类结果准确性主要取决于事件时频特征分布情况,视频信号的分类结果准确性则主要取决于视频中入侵者的动作特征,因此两模型对于事件的识别能力也不相同。时频特征较为相近的事件(如轰砸和脚踢)中,光信号识别的准确率较低;而对于较为相近的动作(如攀爬和拉扯),视频识别的准确率较低。在两模型中,最终输出的结果是所有事件的置信度向量,若使用单一模型进行判别,则只会输出置信度最大的事件作为最终结论。为了实现两模型之间的优势互

表2 以Resnet3D-50为主干结构的SlowFast网络卷积层参数
Table 2 Parameters of each convolution layers in SlowFast with Resnet3D-50 backbone

Layer	Parameters $\{T \times S^2, C\}$	
	Slow pathway	Fast pathway
Data layer	$16, 1^2$	$2, 1^2$
Conv1	$1 \times 7^2, 64$	$5 \times 7^2, 8$
Layer 1	$\begin{bmatrix} 1 \times 1^2, 64 \\ 1 \times 3^2, 64 \\ 1 \times 1^2, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 1^2, 8 \\ 1 \times 3^2, 8 \\ 1 \times 1^2, 32 \end{bmatrix} \times 3$
Layer 2	$\begin{bmatrix} 1 \times 1^2, 128 \\ 1 \times 3^2, 128 \\ 1 \times 1^2, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 3 \times 1^2, 16 \\ 1 \times 3^2, 16 \\ 1 \times 1^2, 64 \end{bmatrix} \times 4$
Layer 3	$\begin{bmatrix} 3 \times 1^2, 256 \\ 1 \times 3^2, 256 \\ 1 \times 1^2, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 3 \times 1^2, 32 \\ 1 \times 3^2, 32 \\ 1 \times 1^2, 128 \end{bmatrix} \times 6$
Layer 4	$\begin{bmatrix} 3 \times 1^2, 512 \\ 1 \times 3^2, 512 \\ 1 \times 1^2, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 1^2, 64 \\ 1 \times 3^2, 64 \\ 1 \times 1^2, 256 \end{bmatrix} \times 3$

补,提出了置信度融合的方法。当光信号检测到入侵信息时,无人机对入侵视频进行拍摄,因此在置信度信

息融合过程中,只对 8 种入侵事件的置信度进行融合。假设在某一时刻下,光信号和视频信号识别结果的置信度向量分别为

$$C_{\text{opt}} = [O_1, O_2, O_3, O_4, O_5, O_6, O_7, O_8], \quad (3)$$

$$C_{\text{video}} = [V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8]. \quad (4)$$

为了将两个模型的优势结合起来,本文将两组置信度向量做内积,得到融合置信度向量:

$$C_{\text{mix}} = [O_1V_1, O_2V_2, O_3V_3, O_4V_4, O_5V_5, O_6V_6, O_7V_7, O_8V_8]. \quad (5)$$

使用线性归一化方法对融合置信度向量进行归一化处理:

$$O_iV_i' = \frac{O_iV_i - \min(C_{\text{mix}})}{\max(C_{\text{mix}}) - \min(C_{\text{mix}})}. \quad (6)$$

最后输出置信度最大的事件作为最终预测事件:

$$F_{\text{Event_pred}} = \text{index}[\max(C_{\text{mix}}')], \quad (7)$$

式中: C_{mix}' 为归一化后的融合置信度向量。该融合方案结合了两个神经网络的识别结果,从多个维度对事件进行描述,避免了识别特征单一所带来的局限性,可以进一步拓宽可识别事件种类,并且保持高准确度,实现智能精准感知。

3.4 评价指标

为了评估所提出的多维度融合分布式光纤振动传感事件识别方案,本文将采用精度(P)、召回率(R)、 F_1 值(F_1)和准确率(A)4个指标进行分析。其中:精度也被称为查准率,指被识别为正确的样本中实际正确样本的比例;召回率也被称为查全率,指实际正确样本中被识别为正确样本的比例; F_1 值代表对召回率和精度指标的调和平均,主要用于对神经网络模型的识别能力和分类能力进行评估;准确率则表示识别正确的样本在总样本中所占的比例。4个指标的具体定义^[22]为

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \times 100\%, \quad (8)$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \times 100\%, \quad (9)$$

$$F_1 = 2 \times \frac{P \times R}{P + R} \times 100\%, \quad (10)$$

$$A = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FN}} + N_{\text{FP}}} \times 100\%, \quad (11)$$

式中: N_{TP} 为实际为正样本、预测结果为正样本的数量; N_{TN} 为实际为负样本、预测结果为负样本的数量; N_{FP} 为实际为负样本、预测结果为正样本的数量; N_{FN} 为实际为正样本、预测结果为负样本的数量。

4 实验结果与分析

为了验证所提出方案的有效性和可行性,根据图 1 搭建了 DMZI 分布式光纤振动传感系统,其中光源的输出功率为 3.5 mW,采集卡的采样率为 10 MHz,对应的采样时间为 0.3 s。工控机选用 CPU: I9-13900, GPU: RTX 3090。无人机飞控选择 pixhawk FMUv5, GPS 为 m8n 模块,摄像头为 Q10F 型光电吊舱,飞行高度设置为 10 m。在测试过程中,2.25 km 待测光纤通过铠装光缆的形式固定在围栏上,分别选取不同的受试者在 50 m、110 m 和 1000 m 处对围栏进行 8 种入侵事件的实验,即攀爬、轰砸、剪切、脚踢、重敲击、轻敲击、拉扯、摇晃。同步采集 DMZI 系统的干涉光信号和无人机拍摄的视频信号,并利用工控机进行信号处理。

4.1 卷积神经网络训练

在光信号数据集构建过程中,需要将无入侵在内的 9 种事件进行采集,每种事件采集 1800 组光信号;在视频信号数据集构建过程中,无入侵事件时,无人机无需响应。因此对 8 种入侵事件进行了视频信号采集,并进行标注,针对每种入侵事件采集了 140 组视频信号,每组的采集时间为 20 s。训练集、验证集和测试集按照 8:1:1 的比例进行随机选取,具体的数据分布如表 3 所示。

表 3 光信号和视频信号数据分布

Table 3 Data distribution of optical signals and video signals

Label	Intrusion event	Optical signal			Video signal		
		Train	Validation	Test	Train	Validation	Test
1	Climbing	1440	180	180	112	14	14
2	Crashing	1440	180	180	112	14	14
3	Cutting	1440	180	180	112	14	14
4	Kicking	1440	180	180	112	14	14
5	Knocking hard	1440	180	180	112	14	14
6	Knocking lightly	1440	180	180	112	14	14
7	Pulling	1440	180	180	112	14	14
8	Wagging	1440	180	180	112	14	14
9	No intrusion	1440	180	180	0	0	0

光信号按照预处理步骤进行端点检测和短时傅里叶变换,端点检测后的高频片段长度为 6×10^6 ,短时傅

里叶变换后的功率分割区间设定为小于 30 dB、30 dB~45 dB,以及大于 45 dB,三个区间的值分别固

定为 0 dB、30 dB、60 dB。视频信号则是通过 Via 软件进行标注,构造时空行为检测所需的 AVA 格式数据集。通过对模型参数进行微调,最终选取 Resnet 50 模型学习率为 0.00001,每 5 个 epoch 更新一次,共训练 100 个 epoch;SlowFast 模型的学习率为 0.1125,每 5 个 epoch 更新一次,共训练 50 个 epoch。模型训练完成后,分别保留最佳训练权重文件。

4.2 结果测试与分析

为了验证训练模型的可行性和有效性,将得到的权重文件在测试集上进行验证,通过 3.4 节中的评价指标对模型效果进行评判。为了体现所提出的多维度传感事件识别方案的准确性,还与单独进行光信号识别和单独进行视频信号识别的方案进行了对比,测试结果如表 4 所示。

表 4 3 种不同识别方法的识别结果对比
Table 4 Comparison of recognition results of 3 different methods

Method	Event type	$P / \%$	$R / \%$	$F_1 / \%$
Using 2D time-frequency images ($A=96.22\%$)	Label 1	100.00	100.00	100.00
	Label 2	88.00	100.00	93.62
	Label 3	100.00	100.00	100.00
	Label 4	90.00	100.00	94.74
	Label 5	100.00	86.21	92.59
	Label 6	99.00	100.00	99.50
	Label 7	100.00	98.04	99.01
	Label 8	93.00	92.59	92.80
	Label 9	100.00	92.59	96.15
Using 3D video signals ($A=78.59\%$)	Label 1	30.35	22.91	26.11
	Label 2	100.00	100.00	100.00
	Label 3	100.00	82.42	90.36
	Label 4	100.00	100.00	100.00
	Label 5	23.68	100.00	38.29
	Label 6	100.00	100.00	100.00
	Label 7	75.26	60.79	67.26
	Label 8	100.00	100.00	100.00
Proposed method ($A=99.58\%$)	Label 1	100.00	100.00	100.00
	Label 2	100.00	100.00	100.00
	Label 3	100.00	100.00	100.00
	Label 4	100.00	100.00	100.00
	Label 5	100.00	100.00	100.00
	Label 6	100.00	100.00	100.00
	Label 7	100.00	100.00	100.00
	Label 8	100.00	100.00	100.00
	Label 9	100.00	92.59	96.15

由表 4 可见,光信号对于时频特征较为明显的事件的识别效果较好,攀爬、剪切、拉扯在测试集上可以达到 100% 的识别准确率,但是轰砸、脚踢、摇晃等具有相似时频特征的事件间会存在混叠;类似地,无人机视频信号关于攀爬、重敲击和拉扯事件的识别准确率较低,在动作上三者的混叠较大。由此可见,光信号识别和视频信号识别单独作用都不能取得较好的识别效果。而经置信度融合后,从两个维度同时对事件进行描述时,可在本文所选用的测试集上实现对 8 种入侵事件 100% 的识别准确率以及对 9 种传感事件 99.58% 的识别准确率。并且光信号识别和视频信号识别同步进行,识别信号总时间为 0.16 s,小于光信号

的采样时间 0.30 s,满足实时性检测的需求。

将本文提出的方法与已有的模式分类模型进行对比分析,对比结果如表 5 所示。由表 5 可知,本文提出的方法在实时检测的基础上,不仅可以极大提升可监测入侵事件数量,使得分布式光纤振动网络传感更加智能化,还可以提高识别准确率,实现精准监测。

5 结 论

根据实际应用的需要,提出了基于卷积神经网络的多维度分布式光纤振动传感事件识别方案,该方案由基于 Resnet 50 模型的二维时频信号识别和基于 SlowFast 模型的三维视频信号识别相结合实现。该

表 5 不同模式识别方案对比
Table 5 Comparison of different pattern recognition methods

Model	Response time /s	Number of intrusion events	Prediction average accuracy /%
SVM-RBF ^[4]	1.01	5	97.10
CLDNN ^[5]	4.00	3	97.00
5 layers CNN ^[6]	5.00	6	93.47
YOLOv5s ^[7]	0.30	5	96.60
Proposed method	0.30	9	99.58

方案不仅通过时频变换的方式拓展了光信号特征,而且使用卷积神经网络自动对特征进行提取和分类,降低了传统机器学习方案手动特征提取低鲁棒性的影响,同时还将三维视频信号识别与光信号识别相结合,实现对攀爬、轰砸、剪切、脚踢、重敲击、轻敲击、拉扯、摇晃、无入侵 9 种事件的检测,所得结果验证了方案的有效性。实验结果表明,9 种传感事件的平均识别准确率为 99.58%,并且识别时间保持在 0.16 s,实现事件变化实时同步响应。与传统单一光信号识别分类相比,所提方案极大地扩充了分布式光纤振动传感技术可识别事件的种类。因此,该事件识别方案在未来将进一步提升分布式光纤振动传感系统在实际工程中的稳定性和可靠性。

参 考 文 献

- [1] 于森,张耀鲁,何禹潼,等. 变分模态分解-排列熵方法用于分布式光纤振动传感系统去噪[J]. 光学学报, 2022, 42(7): 0706005.
Yu M, Zhang Y L, He Y T, et al. Variational mode decomposition and permutation entropy method for denoising of distributed optical fiber vibration sensing system[J]. Acta Optica Sinica, 2022, 42(7): 0706005.
- [2] 赵艳夺,王目光,张静,等. 高空间分辨率大带宽分布式光纤振动传感系统[J]. 光学学报, 2022, 42(19): 1906004.
Zhao Y D, Wang M G, Zhang J, et al. Distributed optical fiber vibration sensing system with high spatial resolution and large bandwidth[J]. Acta Optica Sinica, 2022, 42(19): 1906004.
- [3] 刘铁根,刘琨,戴林,等. 光电信息事件识别感知关键技术研究进展[J]. 光学学报, 2021, 41(1): 0106002.
Liu T G, Liu K, Dai L, et al. Research progress of key technologies in recognition sensing for opto-electronic information and event[J]. Acta Optica Sinica, 2021, 41(1): 0106002.
- [4] Liu K, Sun Z S, Jiang J F, et al. A combined events recognition scheme using hybrid features in distributed optical fiber vibration sensing system[J]. IEEE Access, 2019, 7: 105609-105616.
- [5] Bai Y, Xing J C, Xie F, et al. Detection and identification of external intrusion signals from 33 km optical fiber sensing system based on deep learning[J]. Optical Fiber Technology, 2019, 53: 102060.
- [6] Aktas M, Akgun T, Demircin M U, et al. Deep learning based multi-threat classification for phase-OTDR fiber optic distributed acoustic sensing applications[J]. Proceedings of SPIE, 2017, 10208: 102080G.
- [7] 薛康,刘琨,江俊峰,等. 基于YOLOv5s模型的光纤振动传感事件精准检测研究[J]. 光学学报, 2023, 43(2): 0228001.
Xue K, Liu K, Jiang J F, et al. Optical fiber vibration sensing detection with high accuracy based on YOLOv5s model[J]. Acta Optica Sinica, 2023, 43(2): 0228001.
- [8] Liu K, Jin X B, Jiang J F, et al. Interferometer-based distributed optical fiber sensors in long-distance vibration detection: a review[J]. IEEE Sensors Journal, 2022, 22(22): 21428-21444.
- [9] Li J C, Wang Y, Wang P F, et al. Pattern recognition for distributed optical fiber vibration sensing: a review[J]. IEEE Sensors Journal, 2021, 21(10): 11983-11998.
- [10] Shahid S M, Ko S, Kwon S. Performance comparison of 1D and 2D convolutional neural networks for real-time classification of time series sensor data[C]//2022 International Conference on Information Networking (ICOIN), January 12-15, 2022, Jeju-si, Republic of Korea. New York: IEEE Press, 2022: 507-511.
- [11] Lyu C G, Huo Z Q, Cheng X, et al. Distributed optical fiber sensing intrusion pattern recognition based on GAF and CNN[J]. Journal of Lightwave Technology, 2020, 38(15): 4174-4182.
- [12] 周子纯,刘琨,江俊峰,等. 基于卷积长短期记忆全连接深度神经网络的光纤振动传感事件识别[J]. 光学学报, 2021, 41(13): 1306019.
Zhou Z C, Liu K, Jiang J F, et al. Optical fiber vibration-sensing event recognition based on CLDNN[J]. Acta Optica Sinica, 2021, 41(13): 1306019.
- [13] Liu K, Ma P F, An J C, et al. Endpoint detection of distributed fiber sensing systems based on STFT algorithm[J]. Optics & Laser Technology, 2019, 114: 122-126.
- [14] Sun Z S, Liu K, Jiang J F, et al. High accuracy and real-time positioning using MODWT for long range asymmetric interferometer vibration sensors[J]. Journal of Lightwave Technology, 2021, 39(7): 2205-2214.
- [15] 吴婷婷,许晓东,吴云龙. 卷积神经网络中SPReLU激活函数的优化研究[J]. 计算机与数字工程, 2021, 49(8): 1637-1641.
Wu T T, Xu X D, Wu Y L. Research on optimization of SPReLU activation function in convolutional neural network[J]. Computer & Digital Engineering, 2021, 49(8): 1637-1641.
- [16] Zakaria N, Mohamed F, Abdelghani R, et al. Three ResNet deep learning architectures applied in pulmonary pathologies classification[C]//2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP), November 20-21, 2021, El Oued, Algeria. New York: IEEE Press, 2022.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [18] Sirco A, Almisreb A, Tahir N M, et al. Liver tumour segmentation based on ResNet technique[C]//2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), October 21-22, 2022, Penang, Malaysia. New York: IEEE Press, 2022: 203-208.
- [19] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.

- [20] Jiang K L, Xie T Y, Yan R, et al. An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation[J]. *Agriculture*, 2022, 12(10): 1659.
- [21] Feichtenhofer C, Fan H Q, Malik J, et al. SlowFast networks for video recognition[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 6201-6210.
- [22] 陈吉平. 基于 1-D CNN 的油气管道安全事件分布式光纤检测识别方法[D]. 成都: 电子科技大学, 2019: 40-41.
- Chen J P. Distributed optical fiber detection and identification method of oil and gas pipeline safety events based on 1-D CNN [D]. Chengdu: University of Electronic Science and Technology of China, 2019: 40-41.

Multi-Dimensional Distributed Optical Fiber Vibration Sensing Pattern Recognition Based on Convolutional Neural Network

Jin Xibo^{1,2,3}, Liu Kun^{1,2,3*}, Jiang Junfeng^{1,2,3}, Wang Shuang^{1,2,3}, Xu Tianhua^{1,2,3},
Huang Yuelang^{1,2,3}, Hu Xinxin^{1,2,3}, Zhang Dongqi^{1,2,3}, Liu Tiegeng^{1,2,3}

¹*School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China;*

²*Key Laboratory of Opto-Electronics Information Technology, Ministry of Education, Tianjin University, Tianjin 300072, China;*

³*Institute of Optical Fiber Sensing, Tianjin University, Tianjin 300072, China*

Abstract

Objective As a novel distributed sensing system, the distributed optical fiber vibration system (DOFVS) has been widely applied in recent years due to its advantages of real time, high accuracy, and strong robustness. DOFVS has many application fields, such as structural health monitoring, pipeline leak detection, and perimeter security. In recent years, DOFVS performances such as spatial resolution, monitoring distance, and accuracy have been improved with the demodulation algorithm development and system structure optimization. Meanwhile, with the development of technologies such as deep learning and artificial intelligence, DOFVS also gradually becomes intelligent. To achieve accurate automatic pattern recognition of vibration signals, we combine the DOFVS with an unmanned aerial vehicle (UAV) video monitoring system. The proposed system employs convolutional neural networks to realize pattern recognition in optical signals and video signals simultaneously. Our scheme increases the number of recognizable sensing events and improves recognition accuracy, expanding the intelligent application scenarios of DOFVS.

Methods We propose a multi-dimensional sensing event recognition scheme based on convolutional neural networks, combining the DMZI-based DOFVS and a UAV video monitoring system. The proposed scheme adopts Resnet 50 as the feature extraction backbone network to extract features of the optical signals and video signals. The optical signals are transformed from 1D time-domain signals to 2D time-frequency signals by short-time Fourier transform. The 2D time-frequency images are then segmented based on power distribution to reduce image noise, and the images are fed into a 2D Resnet 50 network to obtain the confidence of the recognized sensing events. The 3D video signals are fed into a SlowFast model with a 3D Resnet 50 as the feature extraction network to obtain the confidence of the recognized sensing events for video signals. Finally, the confidence vectors obtained from both optical and video signals are multiplied and normalized, and the event with the highest confidence is output as the final judgment event. To verify the feasibility of the proposed method, we conduct experiments to recognize nine types of sensing events, and the average recognition accuracy and system response time of the proposed scheme are obtained.

Results and Discussions The proposed scheme overcomes the limitation of recognizing multiple events when only recognizing optical signals. The employed dataset consists of two parts: one is the 2D time-frequency images corresponding to optical signals with 1800 images for each sensing event, and the other is video data obtained from UAV with 140 segments of 20 s videos for each intrusion event (Table 3). Both parts are divided into training, validation, and testing sets in an 8:1:1 ratio. To validate the feasibility and effectiveness of the proposed solution, we compare the results of recognizing optical signals alone, results of video signals alone, and the fused recognition results (Table 4). Optical signals achieve high recognition accuracy on events with more obvious time-frequency features, such as climbing, cutting, and pulling. However, the events with similar features have low accuracy, such as crashing, kicking, and wagging. Similarly, the accuracy of UAV video signals for events such as climbing, knocking hard, and pulling is low. When optical signal recognition and video signal recognition are applied separately, neither of them achieves sound pattern

recognition results. After confidence fusion, the proposed method achieves 99.58% recognition accuracy for nine sensing events in the testing set. Moreover, the recognition of optical signals and video signals can be performed simultaneously, and the system response time can meet the real-time detection needs.

Conclusions We propose a multi-dimensional DOFVS pattern recognition scheme based on convolutional neural networks (CNNs), which combines two models including a 2D time-frequency signal recognition model based on the Resnet 50 and a 3D video signal recognition model based on the SlowFast model. This scheme not only expands the features of the optical signal by time-frequency transformation but also automatically extracts and classifies features using CNNs. The impact of low robustness of manual feature extraction schemes can be reduced. Meanwhile, the 3D video signal recognition is combined with optical signal recognition to enable the detection of nine types of events including climbing, crashing, cutting, kicking, knocking hard, knocking lightly, pulling, wagging, and no intrusion. The effectiveness of the proposed scheme is verified via experiments, which demonstrate that the average accuracy of the nine events is 99.58% and the recognition time is 0.16 s to achieve real-time synchronous response to event changes. Compared with traditional single optical signal recognition, the proposed scheme greatly expands the event types that can be recognized in the DOFVS field. Therefore, this scheme will further improve the DOFVS stability and reliability in practical engineering applications in the future.

Key words fiber optics; distributed optical fiber sensing; multi-dimensional sensing; pattern recognition; convolutional neural network