

文章编号:1671-6833(2023)02-0061-07

## 基于多传感器融合的无人机生命搜寻方法

李明辉<sup>1</sup>, 马文凯<sup>1</sup>, 周翊民<sup>2</sup>, 叶玲见<sup>2</sup>

(1. 陕西科技大学 机电工程学院, 陕西 西安 710016; 2. 中国科学院 深圳先进技术研究院, 广东 深圳 518055)

**摘要:**为应对单个生命探测传感器在野外、灾区生命搜寻时的不稳定状况,提出了一种基于多传感器信息融合的无人机(UVA)生命搜寻方法。首先,构建不同结构的 ResNeXt 网络以提取不同维度信息的特征,利用一维 ResNeXt 网络提取音频梅尔频谱系数的深层特征,利用二维 ResNeXt 网络提取红外图像的深层特征;其次,使用判别相关分析(DCA)对2种高维特征进行降维融合,兼顾不同特征的相关性和类别性,以获得更丰富的环境信息,从而提高生命搜寻的准确性;最后,将融合特征输入支持向量机分类器以进行生命识别的决策,建立具有相关性的音频和图像双模态数据集,并将所提方法在该数据集上进行实验比较和分析,对其性能进行有效评估。实验结果表明:所提方法在特征提取和特征融合方面效果优于其他传统方法,且多传感器融合识别准确率可达98.7%,证明该方法能有效提高特殊场景下人体检测的准确性,多传感器融合检测效果优于单传感器。

**关键词:**数据融合; 红外图像特征; 音频特征; 判别相关分析(DCA); 无人机生命搜寻方法

**中图分类号:** TP391.4

**文献标志码:** A

**doi:**10.13705/j.issn.1671-6833.2023.02.003

在野外、灾区等场景下开展救援任务,需要考虑安全和效率问题,快速、准确的生命搜索技术对救援任务而言显得尤为重要<sup>[1]</sup>。在传统的生命探测任务中,可以使用各种类型的传感器来搜索目标区域,例如,超宽带脉冲雷达、热红外探测器和声波传感器。超宽带脉冲雷达<sup>[2]</sup>可以探测到人类的呼吸和心跳,它不会受到环境反射的干扰,以非接触形式进行工作。热红外探测器<sup>[3]</sup>主要通过感应温度变化来判断目标是否存在,可以在恶劣的环境中工作,如浓烟、大雾和自然灾害期间的昏暗环境。声学检测<sup>[4]</sup>可以识别目标的声音信号,通过收集环境声波信号来执行救援任务。生命探测传感器在不同的场景下发挥着重要的作用,但它们的探测能力和环境适应性是有限的。由于无人机拥有重量轻、体积小、机动性好等特点,可以利用它同时携带不同类型的传感器来实现生命识别。可见光相机<sup>[5]</sup>可以收集信息来识别人体的皮肤颜色。雷达<sup>[6]</sup>可用于收集图像,以便在紧急情况下在户外搜寻受伤和失踪人员。热释电红外传感器(PIR)<sup>[7]</sup>是另一种可用于人体搜索的设备。此外,音频传感器也可应用于灾害

现场的人体探测<sup>[8]</sup>。

上述用于人体检测的传感器可以在短时间内收集大量的传感器数据,这些数据包含了目标人体的信息,但数以百计的图像或音频信号会增加计算负担,减缓处理器的处理速度。在深度学习中,深度网络有能力从包含大量样本对象的训练数据中学习特征,从而进行检测和决策。目前,深度学习和图像处理相结合可用于目标搜索检测。卷积神经网络可用于目标检测算法<sup>[9]</sup>,用低成本的热成像仪采集数据,在复杂背景下利用鸟瞰视角检测目标。同时,在无人机生命搜索任务中,可以利用 Faster-RCNN 网络对人体目标进行检测<sup>[10]</sup>。然而,红外探测仪虽能在恶劣环境中探测生命,但受环境温度分布影响较大,音频传感器虽能检测到人的声音,但受环境噪声影响较大,任何一种传感器均受工作场景的限制,不能满足生命搜索等任务的严格要求。因此,有必要研究一种基于多传感器融合的生命探测方法来提高探测精度。

本文提出了一种基于判别相关分析(DCA)<sup>[11]</sup>多传感器特征融合的无人机生命检测方法。将音频

收稿日期:2022-09-09;修订日期:2022-10-18

基金项目:国家自然科学基金资助项目(61973296);深圳市科创委资助项目(JCYJ20170818153635759);咸阳市重点研发计划(S2021ZDYF-GY-0244)

通信作者:周翊民(1975—),女,天津人,中国科学院深圳先进技术研究院研究员,博士,博士生导师,主要从事控制理论、机器人、智能电网等研究,E-mail:ym.zhou@siat.ac.cn。

引用本文:李明辉,马文凯,周翊民,等. 基于多传感器融合的无人机生命搜寻方法[J]. 郑州大学学报(工学版), 2023, 44(2): 61-67. (LI M H, MA W K, ZHOU Y M, et al J. UAV life search method based on multi-sensor fusion[J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(2): 61-67.)

传感器采集到的声音的梅尔尺度频谱系数(MFCC)特征和红外相机采集到的红外图像特征进行融合,能够有效提高人体探测精度,该方法融合多传感器信息,具有较强的实施性和移植性,对于搜寻任务具有重要意义。

## 1 理论基础

### 1.1 ResNeXt

在深度学习识别和分类任务中,最关键的任务就是特征提取。卷积神经网络 LeNet<sup>[12]</sup>被引入用于手写数字的识别,但受限于图像处理器的性能。随着 ReLU 等机制的引入,AlexNet<sup>[13]</sup>引起了广泛的关注。更多的神经网络不断发展,如 VGGNet<sup>[14]</sup>和 GoodLeNet<sup>[15]</sup>。

随着神经网络深度的增加,图像中包含的语义信息也会增加,但网络的性能会下降。主要原因是梯度分散或爆炸,导致训练过程无法进一步收敛。残差网络<sup>[16]</sup>可以在保持模型精度的前提下,通过加深和拓宽网络来有效解决退化问题。但是,超参数数量的增加也会增加网络的计算成本。ResNeXt<sup>[17]</sup>网络模型在和其他网络模型具有同等参数量的情况下有着更好的准确度。因此,本文使用这样的残差网络来提取样本的特征。

ResNeXt 结合 Inception 网络架构对 ResNet 进行了改进,采用了拆分—转换—合并的思想,可以表示为

$$R(x) = \sum_{i=1}^C T_i(x)。$$

(1)

式中:  $T_i$  为相同的拓扑结构;  $C$  为模块中相同分支的数量,这是除分支结构的深度和宽度之外,用来描述模型复杂性的第 3 个指标,通常被称为基数,在本文中设定为 32。图 1 描述了 ResNet 网络和 ResNeXt 网络的基本模块。

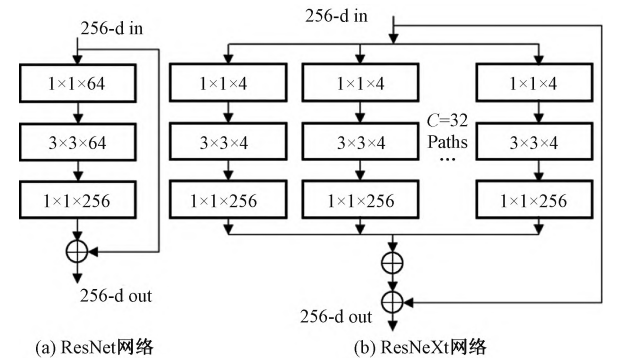


图 1 ResNet 网络和 ResNeXt 网络的基本模块

Figure 1 Block of ResNet and block of ResNeXt

ResNeXt 有 32 个具有相同结构的通道,这些不同的通道可以从传感器数据中学习各种特征。与 ResNet 相比,ResNeXt 可以充分利用特征,大大降低了过拟合的风险。

### 1.2 判别相关分析法

判别相关分析(DCA)是由典型相关分析(CCA)<sup>[18]</sup>延伸出来的一种特征融合技术,它可以最大限度地兼顾相关性和类别性,并且能对特征进行显著地降维。因此,DCA 适用于多模态的异构特征融合,本文将引入到多传感器生命搜寻任务中。

$X$  和  $Y$  分别代表不同的特征矩阵,通过 DCA 方法最终得到的是 2 个转换矩阵  $W_x$  和  $W_y$ ,而  $\hat{X}$  和  $\hat{Y}$  则是降维后用于融合的特征矩阵,如下所示:

$$\hat{X} = W_x X;$$

(2)

$$\hat{Y} = W_y Y。$$

(3)

新的 2 个特征集具有很好的分类效果和对应相关性。最后采用串联或求和的方式得出融合特征,如下式:

$$Z_1 = \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} = \begin{bmatrix} W_x^T X \\ W_y^T Y \end{bmatrix} = \begin{bmatrix} W_x & 0 \\ 0 & W_y \end{bmatrix}^T \begin{bmatrix} X \\ Y \end{bmatrix};$$

(4)

$$Z_2 = \hat{X} + \hat{Y} = W_x^T X + W_y^T Y = \begin{bmatrix} W_x \\ W_y \end{bmatrix}^T \begin{bmatrix} X \\ Y \end{bmatrix}。$$

(5)

### 1.3 梅尔频率倒谱系数

梅尔频率倒谱系数(mel-scale frequency cepstral coefficients,MFCCs)是语音信号处理中最常用的特征之一,具有较高的识别性能和噪声抑制能力。它与频率的关系可以近似地用下式表示:

$$Mel(f) = 2595 \lg(1 + f/700)。$$

(6)

式中:  $f$  为频率,Hz。从物理的角度来看,这个过程将声音信号的频谱映射到梅尔频谱,并提取梅尔频谱的包(频域),这对人类的语音识别具有重要意义。

## 2 无人机生命搜寻方法

本文提出的无人机生命搜寻方法是首先利用无人机运用音频传感器和红外相机采集数据,构建不同结构的 ResNeXt 网络,从不同类型的传感器中提取红外图像特征和音频特征;其次,使用判别相关分析(DCA)来融合深度特征,2 个高维特征向量被结合起来,以获得一个比单一输入更具辨别力的向量:最后,融合后的特征用来做人类识别的决策,其整体框架如图 2 所示。该方法是一个基于特征融合的端到端架构,主要由 4 个基本步骤组成:数据采集、特

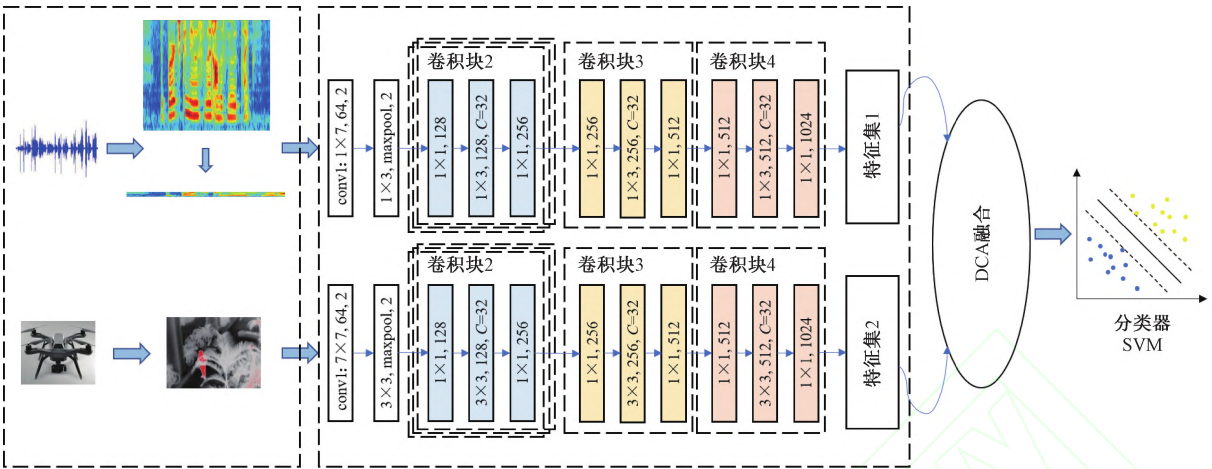


图 2 生命探测总体框架

Figure 2 Overall framework for life detection

征提取、特征融合和目标分类。

2.1 数据采集

在本文研究中,截至目前并没有合适的公共数据集,所以有必要去创建一个具有相关性的多传感器数据集。使用 F450 无人机搭载艾睿天眼 T2 红外热成像仪和 USB 麦克风采集数据。

第 1 种数据是 640 张图像数据,无人机在距离地面 3~5 m 的室外场景进行拍摄,利用小部分遮挡模拟真实场景。其中,正样本为包含 1~5 个目标人体的图像,负样本为无人场景下的环境图像。

第 2 种数据是 640 个音频数据,无人机在飞行过程中采集上述场景中的声音。其中,正样本包含人的喊叫声、哭泣声、咳嗽声及笑声各 80 个,负样本包含引擎转动声、风声、警笛声及钻孔声<sup>[19]</sup> 各 80 个。图 3 和图 4 为数据集部分样本,包含正样本和负样本。

由于神经网络的训练需要两类传感器大量的样本,且为防止数据集样本数量小导致过拟合,并让模型学习到更多有力的特征,对数据集进行扩充。数据集的第一部分通过改变亮度、对比度、随机旋转和翻转这些参数进行扩展。数据集的第 2 部分通过添加噪声、波形位移、波形拉伸和音调变化而进一步扩展。

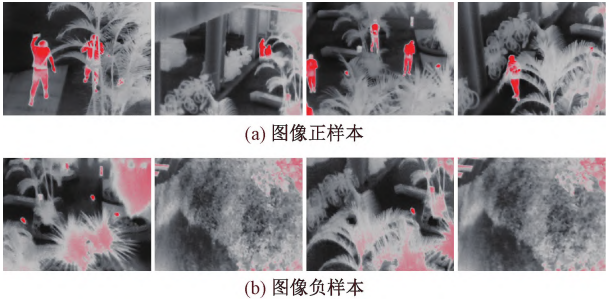


图 3 红外图像数据集部分样本

Figure 3 Partial sample of infrared image dataset

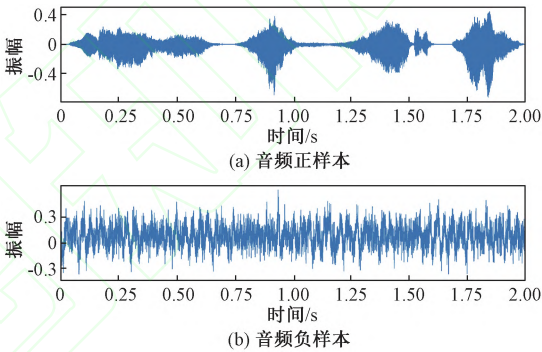


图 4 音频信号数据集部分样本

Figure 4 Partial sample of audio dataset

最终构建的数据集包含 2 560 个具有 2 种模态的数据样本。

2.2 特征提取

卷积神经网络 (CNN) 可以提取复杂数据的深层特征,而残差网络可以解决网络深度增加导致的网络性能下降的问题。ResNeXt 可以在不增加参数复杂性的情况下提高网络精度,同时减少超参数的数量。

然而,现有的深层 ResNeXt 模型结构复杂、参数多,需要很长的训练时间。因此,为了提高网络模型的运行速度,本文减少了相同残差块的数量,对网络进行轻量化处理。在提取深层特征后,为了进一步促进融合,去掉了原网络的最后一层、卷积层、池化层和全连接层,仅保留中间的高层特征。根据经验,设定基数  $C$  为 32。

针对热红外相机和声波传感器的数据结构不同的问题,本文对经典的卷积核进行优化,通过使用 2 个不同结构的卷积核来构建网络:一维卷积核用来处理一维数据,二维卷积核用来处理二维数据。

一般来说,图像的输入是一个二维图像结构,因此,卷积核通常是一个二维矩阵:



$$\omega_{M \times N} = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1N} \\ \vdots & \vdots & \vdots \\ \omega_{M1} & \cdots & \omega_{MN} \end{bmatrix}.$$

(7)

式中： $M \times N$  表示卷积核的大小， $M = N = n$  ( $n$  为整数)。

如果输入数据是一维序列信号,传统方法是将一维输入信号转换成二维输入,即把时间序列信号转换成频域图像。虽然这符合原始网络结构,但整个过程可能导致某些关键信息丢失。考虑本研究的实际应用问题,提取音频数据中的 MFCC 特征序列,将其转换成一维输入,能够更好地识别人声。因此卷积核大小为  $1 \times W$ ,即

$$\omega_{1 \times W} = [\omega_{11}, \omega_{12}, \cdots, \omega_{1W}].$$

(8)

将具有不同卷积核的网络分别记为 1D-ResNeXt 和 2D-ResNeXt。

1D-ResNeXt 用来提取一维输入的深层特征,2D-ResNeXt 用来提取二维输入的深层特征。这两层网络共同构成目标网络架构,具体参数见表 1。

表 1 网络结构参数

Table 1 Specific parameters of the model

	一维 ResNeXt-50	二维 ResNeXt-50
卷积块 1	1×7, 64, 步长 2	7×7, 64, 步长 2
	1×3, 最大池化, 步长 2	3×3, 最大池化, 步长 2
卷积块 2	$\begin{bmatrix} 1 \times 1, 128 \\ 1 \times 3, 128, C = 32 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128, C = 32 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
卷积块 3	$\begin{bmatrix} 1 \times 1, 256 \\ 1 \times 3, 256, C = 32 \\ 1 \times 1, 512 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256, C = 32 \\ 1 \times 1, 512 \end{bmatrix} \times 1$
卷积块 4	$\begin{bmatrix} 1 \times 1, 512 \\ 1 \times 3, 512, C = 32 \\ 1 \times 1, 1\,024 \end{bmatrix} \times 1$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512, C = 32 \\ 1 \times 1, 1\,024 \end{bmatrix} \times 1$

2.3 特征融合

上述网络输入的大小为  $224 \times 224$  和  $1 \times 1\,024$ 。通过上述网络,能够提取红外图像和音频 MFCC 信号的深层特征。经过第 4 个卷积块,1D-ResNeXt 提取的声音特征被映射为  $1 \times 64 \times 1\,024$  维度,而 2D-ResNeXt 提取的图像特征为  $14 \times 14 \times 1\,024$  维度。

所有样本的图像特征映射矩阵为  $\mathbf{X}$ , 音频特征映射矩阵为  $\mathbf{Y}$ , 使用判别相关分析方法计算出最终的转换矩阵  $\hat{\mathbf{X}}$  和  $\hat{\mathbf{Y}}$ 。最终通过式(5)所示的方式进行融合,将转换矩阵相加得到融合特征矩阵。不同模型数据的融合能够以较低的维度为模型决策提供更多的信息,并保留较强的判别能力特征,从而提升

检测效果<sup>[20]</sup>。

2.4 目标分类

传感器数据经过特征提取与特征融合,在不丢失特征信息的情况下,极大地压缩了数据的维度。为了进一步准确地识别特殊场景下是否存在人体,需要根据融合后的特征构造分类模型。

支持向量机(SVM)能够有效处理线性和非线性分类问题,且适合处理小样本分类问题<sup>[21]</sup>,在本任务中融合特征均为低维度特征,因此采用 SVM 作为分类器进行有监督训练。

在训练过程中,将 2 560 个融合特征样本以 4:1 的比例划分为训练集和测试集,经过测试,确定惩罚系数为 180,核函数为高斯核。分类器最终输出结果为有人或无人。

3 实验验证

为了验证所提方法在生命搜寻任务上的有效性,率先在计算机上进行验证。计算机使用的是 Ubuntu16.04 操作系统,CPU 为 Intel(R) Xeon(R) Silver 4214R CPU @ 2.40 GHz,显卡为英伟达 GeForce RTX 3090,在此硬件上使用 Pytorch 深度学习框架来完成验证。

3.1 数据预处理

为了防止模型提取的特征维度过大不便保存且能应用于移动无人机设备,需要将原始数据进行处理后再进行训练。

对于搜寻过程采集的红外图像,能提取到人的轮廓特征完成分类即可,因此将图像调整为  $224 \times 224$ ,并进行归一化。提取原始音频数据的 MFCC 系数特征,将其一维化并补充大小至  $1 \times 1\,024$ ,便于深层特征融合。

数据集被随机分为训练集、验证集和测试集,比例为 8:1:1。

3.2 基于红外图像的生命搜寻实验

将设计的网络用来训练图像数据以提取图像特征。在训练过程中,使用的具体设置参数如下:使用随机梯度下降法(SGD)优化检测模型,图像批处理量大小为 128,动量设为 0.5,初始学习率为 0.001,训练 150 个周期。

为了定量评估目标检测方法的性能,引入准确率 Acc 进行评价:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}.$$

(9)

式中:TP 代表正确预测的阳性样本;TN 代表正确预测的阴性样本;FP 代表错误预测的阳性样本;

$FN$  代表错误预测的阴性样本。

经过 150 次迭代,模型训练完成,训练过程曲线如图 5 所示。结果表明,在训练初期,经过 45 次迭代,损失值急剧下降,并逐渐趋于收敛。 $Acc$  的收敛性达到 96.0%。

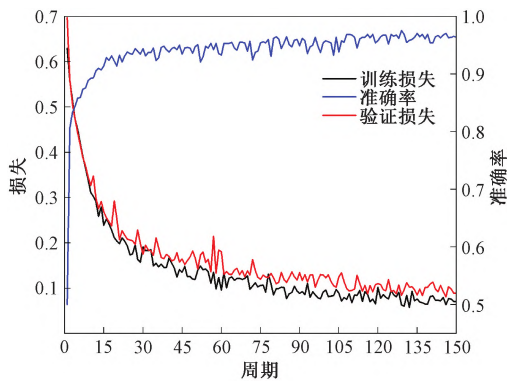


图 5 基于红外图像的训练过程曲线

Figure 5 Training process curves based on image data

3.3 基于声音信号的生命搜寻实验

音频采样率被设定为 16 000 Hz。在训练过程中,由于声音数据需要提前加载,设置线程数为 4,网络训练参数与第一组实验中的参数相同。

经过 150 次迭代训练,模型充分学习了音频信号的特征,得到的训练过程曲线如图 6 所示。结果显示,在训练的早期阶段,损失值急剧下降。虽然有振荡,但在 30 次迭代后损失值趋于收敛,准确率达到 84.1%。

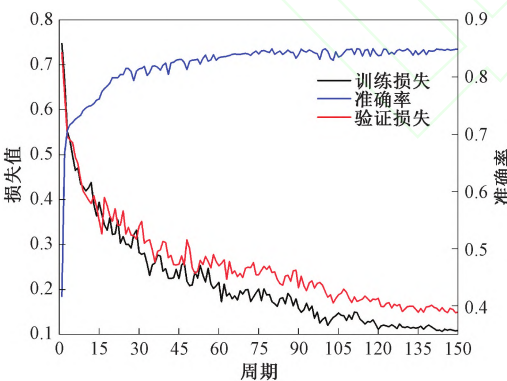


图 6 基于声音信号的训练过程曲线

Figure 6 Training process curves based on audio data

3.4 本文方法生命搜寻实验

首先,利用以上已训练好的单类传感器网络模型分别提取音频信号特征和红外图像特征,输出经过最后一个卷积块提取出的深层特征作为融合对象,音频样本提取出的特征命名为  $X$ , 图像样本提取出的特征命名为  $Y$ 。

其次,由于提取出的深层特征是包含冗余信息的高维数据,通过特征级融合能极大地提取输入数

据的信息。通过判别相关分析法获得转换后的低维特征向量  $\hat{X}$  和  $\hat{Y}$ , 并按照求和的方式进行融合。

最后,通过所有样本的融合特征训练 SVM 分类器,选用高斯核函数,惩罚因子为 180。测试样本的识别准确率为 98.7%。为了进一步评估所提出方法的分类效果,引入混淆矩阵作为判别分类精度的标准,如图 7 所示。

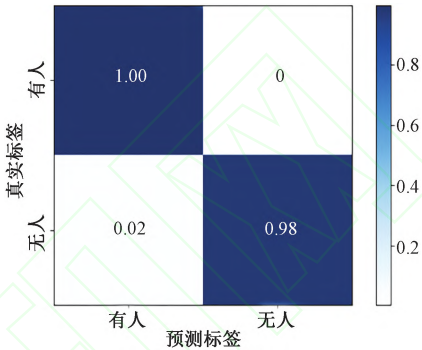


图 7 DCA 融合的混淆矩阵

Figure 7 Confusion matrix for DCA fusion

3.5 对比分析

为了验证所提方法在生命搜寻任务中的有效性,将本方法与其他方法进行对比。在特征提取方面,轻量化 ResNeXt 网络与 VGG-16、ResNet-34 和 ResNeXt-50 进行了比较;为了证明该方面对于音频分类的有效性,与主流的语音识别模型进行对比,如 LSTM、GRU 及 TCN,但由于分类所需特征与语音识别所需特征不同,以上网络显然不适用于搜寻任务。实验结果如表 2 所示,目标模型在任务分类中表现较好,其中红外图像数据集的识别准确率为 96.0%,音频数据的识别准确率为 84.1%。

表 2 不同网络结构对比

Table 2 Comparison of different network structures		
网络	样本	准确率/%
VGG-16	图像	91.2
	音频	80.5
LSTM	音频	83.0
GRU	音频	79.3
TCN	音频	83.6
ResNet-34	图像	94.4
	音频	83.8
ResNeXt-50	图像	95.9
	音频	84.2
轻量化 ResNeXt	图像	96.0
	音频	84.1

在特征融合方面,采用文中目标网络结构提取深层特征,将提取出的特征采用不同的策略进行融合,将目标融合方法与以往的串联、求和以及典型相

关分析法进行对比,并将音频转换为二维输入与图像串联在本数据集上进行测试,结果如表 3 所示。

表 3 融合方法对比

Table 3 Comparison of fusion methods

融合方法	准确率/%
串联	96.3
求和	94.6
CCA	96.6
两种图像特征串联	95.5
DCA	98.7

本文方法进行生命搜寻所得识别准确率为 98.7%,这与传统方法相比最少提高 2.1 个百分点。CCA 融合方法混淆矩阵如图 8 所示,识别准确率较高,但存在较多分类错误情况。

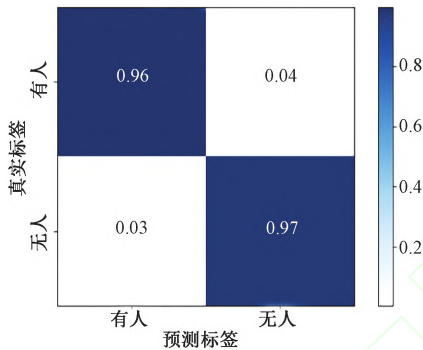


图 8 CCA 融合的混淆矩阵

Figure 8 Confusion matrix for CCA fusion

从整体来讲,本文方法在特征提取、特征融合的过程中,不仅最大化音频和图像之间对应特征的相关性,保留有效特征,还对特征进行显著降维,具有较低的计算复杂度,适用于实时生命搜寻。

### 4 结论

针对单一生命探测传感器在救援搜寻任务中的局限性,本文提出了一种多传感器融合的无人机生命搜寻方法,将音频传感器和红外相机信息进行融合。在特征提取阶段,将一维的音频 MFCC 特征和二维的红外图像特征输入轻量化 ResNeXt 神经网络模型;在特征融合阶段,用判别相关分析法对高维特征进行降维、融合;在分类阶段,采用预训练的 SVM 分类器进行实时检测。利用自建双模态数据集对该方法进行评估,实验结果表明,本文的模型在特征提取、特征融合方面均优于其他方法,识别准确率达到 98.7%,在无人生命搜寻任务中具有广阔的应用前景。

### 参考文献:

[1] KEYWORD S, WOLFE S. UAVs for land use applica-

tions: UAVs in the civilian airspace institution of engineering and technology [C]//IET Seminar on UAVs in the Civilian Airspace. London:IET, 2013: 1-13.

[2] MA Y Y, LIANG F L, WANG P F, et al. Research on identifying different life states based on the changes of vital signs of rabbit under water and food deprivation by UWB radar measurement [C]//2019 Photonics & Electromagnetics Research Symposium-Fall ( PIERS-Fall ). Piscataway:IEEE, 2019: 397-403.

[3] SEKI M, FUJIWARA H, SUMI K. A robust background subtraction method for changing background [C]//Proceedings Fifth IEEE Workshop on Applications of Computer Vision. Piscataway:IEEE, 2000: 207-213.

[4] LIU X C, FENG X L. Research on weak signal detection for downhole acoustic telemetry system [C]//2010 3rd International Congress on Image and Signal Processing. Piscataway:IEEE, 2010: 4432-4435.

[5] ZACHARIE M, FUJI S, MINORI S. Rapid human body detection in disaster sites using image processing from unmanned aerial vehicle (UAV) cameras [C]//2018 International Conference on Intelligent Informatics and Biomedical Sciences ( ICHBMS ). Piscataway:IEEE, 2018: 230-235.

[6] MCCLURE J, SAHIN F. A low-cost search-and-rescue drone for near real-time detection of missing persons [C]//2019 14th Annual Conference System of Systems Engineering ( SoSE ). Piscataway:IEEE, 2019: 13-18.

[7] VALSAN A, B P, G H V D, et al. Unmanned aerial vehicle for search and rescue mission [C]//2020 4th International Conference on Trends in Electronics and Informatics ( ICOEI ) (48184). Piscataway:IEEE, 2020: 684-687.

[8] YAMAZAKI Y, PREMACHANDRA C, PEREA C J. Audio-processing-based human detection at disaster sites with unmanned aerial vehicle [J]. IEEE Access, 2020, 8: 101398-101405.

[9] PERDANA M I, RISNUMAWAN A, SULISTIJONO I A. Automatic aerial victim detection on low-cost thermal camera using convolutional neural network [C]//2020 International Symposium on Community-centric Systems ( CcS ). Piscataway:IEEE, 2020: 1-5.

[10] MARUŠIĆ Ž, ZELENKA D, MARUŠIĆ T, et al. Visual search on aerial imagery as support for finding lost persons [C]//2019 8th Mediterranean Conference on Embedded Computing ( MECO ). Piscataway:IEEE, 2019: 1-4.

[11] HAGHIGHAT M, ABDEL-MOTTALEB M, ALHALABI W. Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition [J]. IEEE Transactions on Information Forensics and Security,

- 2016, 11(9): 1984–1996.
- [12] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324.
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25: 1097–1105.
- [14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2021-11-12]. <https://arxiv.org/abs/1409.1556>.
- [15] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//*Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway:IEEE, 2015: 1–9.
- [16] 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. Piscataway:IEEE, 2016: 770–778.
- [17] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2017: 5987–5995.
- [18] 龚瑞昆, 刘佳. 图像处理的玉米病害识别研究[J]. *现代电子技术*, 2021, 44(24): 149–152.
- GONG R K, LIU J. Research on maize disease recognition based on image processing[J]. *Modern Electronics Technique*, 2021, 44(24): 149–152.
- [19] SALAMON J, JACOBY C, BELLO J P. A dataset and taxonomy for urban sound research [C]//*Proceedings of the 22nd ACM International Conference on Multimedia*. New York: ACM, 2014: 1041–1044.
- [20] 张坚鑫, 郭四稳, 张国兰, 等. 基于多尺度特征融合的火灾检测模型[J]. *郑州大学学报(工学版)*, 2021, 42(5): 13–18.
- ZHANG J X, GUO S W, ZHANG G L, et al. Fire detection model based on multi-scale feature fusion[J]. *Journal of Zhengzhou University (Engineering Science)*, 2021, 42(5): 13–18.
- [21] 穆晓敏, 张嗣思, 齐林. 基于 2D-FrFT 多阶次特征融合的人脸表情识别方法[J]. *郑州大学学报(工学版)*, 2012, 33(1): 109–112.
- MU X M, ZHANG S S, QI L. Human emotion recognition using fused 2D-fractional Fourier transform features based on CCA[J]. *Journal of Zhengzhou University (Engineering Science)*, 2012, 33(1): 109–112.

## UAV Life Search Method Based on Multi-sensor Fusion

LI Minghui<sup>1</sup>, MA Wenkai<sup>1</sup>, ZHOU Yimin<sup>2</sup>, YE Lingjian<sup>2</sup>

(1. School of Electrical and Mechanical Engineering, Shanxi University of Science and Technology, Xi'an 710016, China; 2. Shenzhen Institute of Advanced Technology, Chinese Academy of Science, Shenzhen 518055, China)

**Abstract:** In order to cope with the unstable conditions of single sensor for life detection in the fields and disaster areas, a life search method was proposed based on multi-sensor information fusion. Firstly, ResNeXt networks with different structures were constructed to extract features with different dimensional information. Deep features of audio Mel-scale Frequency Cepstral Coefficients were extracted using a one-dimensional ResNeXt network, and deep features of the infrared images were extracted using a two-dimensional ResNeXt network. Secondly, the two high-dimensional features were fused by dimensionality reduction using discriminant correlation analysis (DCA) to take into account the correlation and category of different features in order to obtain richer environmental information, thus improving the life search accuracy. Finally, the fused features were fed into a support vector machine classifier for decision making in life recognition. A bimodal dataset of audio and images with correlation was created and the proposed method was experimentally compared and analyzed in this dataset to evaluate the search performance of the proposed method. The experimental results demonstrated that the proposed method could outperform other traditional methods in feature extraction and feature fusion, and the multi-sensor fusion recognition accuracy could reach 98.7%, which proved that the method could effectively improve the accuracy of human detection in special scenes, and the performance of multi-sensor fusion based human detection was higher than that with single sensor.

**Keywords:** data fusion; infrared image features; audio features; discriminant correlation analysis (DCA); UAV life search methods