

采用混合域注意力机制的无人机识别方法

薛珊^{1,2}, 卫立炜¹, 顾宸瑜³, 吕琼莹¹

(1. 长春理工大学机电工程学院, 130022, 长春; 2. 长春理工大学重庆研究院, 401135, 重庆;
3. 西安交通大学信息与通信工程学院, 710049, 西安)

摘要: 针对在城市公园、广场和大型游乐场等公共环境中,雷达和无线电识别无人机易受到电子干扰、图像识别无人机易受到光线和遮挡物干扰的问题,提出了一种经济便捷、不易受到干扰的运用声音和采用通道空间混合域注意力机制多尺度分组卷积网络(ECSANet)的无人机识别方法。首先,建立民用的 9 大类无人机声音数据集,提取数据集的对数梅尔谱图及其动态特征;其次,为了网络参数量少,避免过拟合,设计了基于分组卷积、通道混洗和残差结构的通道混洗多尺度分组卷积网络(MSSGNet);然后,为了能更多、更有效地提取无人机声音特征,设计了通道空间混合域注意力机制模块(ECSA);最后,将 ECSA 模块插入 MSSGNet 网络构成改进的通道空间混合域注意力机制的多尺度分组卷积网络(ECSANet),形成新型声音识别无人机的方法。运用设计的 ECSANet 网络对自建的民用无人机声音数据集和 Urbansound8K 环境声音数据集进行了声音识别,识别结果表明:与 ResNet18、ResNet34、ResNeXt18 和 MobileNetV2 等基准网络相比,MSSGNet 网络参数更少,识别准确率更高,达到了 95.1%;ECSA 模块可以插入多种网络,在不增加很多参数的情况下令网络模型的识别准确率获得提升,在无人机等声音分类任务上具有很好的效果;与 MSSGNet 网络相比,改进的 ECSANet 网络识别准确率能达到 95.9%,提高了 0.8%,表明了该网络在识别小样本无人机方面的优越性和可行性。

关键词: 无人机;声音识别;对数梅尔谱图;神经网络;混合域注意力机制

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202210014 **文章编号:** 0253-987X(2022)10-0141-10

Drone Identification Method Based on Mixed Domain Attention Mechanism

XUE Shan^{1,2}, WEI Liwei¹, GU Chenyu³, LÜ Qiongying¹

(1. School of Mechanical and Electrical Engineering, Changchun University of Science and Technology, Changchun 130022, China; 2. Chongqing Research Institute, Changchun University of Science and Technology, Chongqing 401135, China; 3. School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: An economical, convenient and undisturbed drone detection method using sound and multiscale group convolution network with attention mechanism in mixed domain of channel space (ECSANet) is proposed in the context of susceptibility to electronic interference in identification of drones by radar and radio, and the interference of light and obstruction in identification of drones by images in public environments such as urban parks, squares and large amusement parks. Firstly, nine kinds of sound dataset of civil drones are established, and their logarithmic

收稿日期: 2022-04-14。 作者简介: 薛珊(1978—),女,教授,博士生导师。 基金项目: 吉林省重点科技研发资助项目(20180201058SF);吉林省教育厅科学技术研究资助项目(JJKH20210812KJ)。

网络出版时间: 2022-07-04

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.t.20220701.1417.004.html>

Mel spectra and dynamic characteristics are extracted. Secondly, based on packet convolution, channel shuffling and residual structure, a multi-scale group convolution network with channel shuffle (MSSGNet) is designed to reduce the network parameters and avoid over fitting. Then, the efficient channel and spatial attention (ECSA) is designed to extract more and more effective features of drone sounds. Finally, the ECSA is inserted into the MSSGNet to form an improved multiscale group convolution network with attention mechanism in mixed domain of channel space (ECSANet), offering a new method for sound recognition of drones. The designed ECSANet is used to identify the self-built civil drone sound dataset and environmental sound dataset urban-sound8k. The results reveal that when compared with benchmark networks such as ResNet18, ResNet34, ResNeXt18, and MobileNetV2, the MSSGNet has fewer network parameters but a higher identification accuracy (up to 95.1%). The ECSA can be inserted into a variety of networks to improve identification accuracy of network models without introducing too many parameters, and it works well for sound classification tasks like drones. As compared with the MSSGNet, the improved ECSANet has an identification accuracy of 95.9%, an increase of 0.8 percent, demonstrating the superiority and feasibility in identifying a small sample of drones.

Keywords: drone; voice recognition; log Mel-spectrogram; neural network; mixed domain attention mechanism

近年来,民用无人机黑飞的数量大大增加,给人们带来了极大威胁,如何应对无人机威胁并对其进行反制就成为了亟待解决的问题。在此背景下,对无人机进行反制的反无人机系统成为了研究热点。反无人机系统的关键之一,也是它的首要解决问题是对无人机进行探测识别。目前对无人机进行探测识别有多种方法,包括图像识别、无线电识别等,它们有各自的优缺点和适用范围,先进的趋势是采用多种识别方法结合的方式^[1-4]。在此情况下,经济便捷的声学识别方法得到了广泛关注,它不会对其他方法形成干扰,可以成为一种很好的辅助手段,如何运用声音识别无人机成为了研究的热点。

针对无人机声音识别问题,学者们已经对其进行了一系列的研究,目前大多采用卷积神经网络的方法识别。Seo等使用具有声学短时傅里叶变换特征的卷积神经网络来对无人机进行探测^[5];Casabianca等提取无人机的梅尔谱图特征,运用卷积神经网络对无人机进行识别^[6]。研究过程中使用的无人机以及特征等各不相同,并没有统一的标准,所以目前并没有统一且成熟的无人机公共声音数据集。现实中无人机音频样本数量较少,在运用大模型对其进行识别时,容易引起过拟合,导致识别准确率不高;除此之外,不同种类间的无人机声音音频特征相似,不易区别,这也影响了模型的识别准确率。如何运用深度学习中的先进方法提高对无人机声音的识别准确率,成为了人们研究的难点。

本文针对大模型识别无人机时准确率不高的问题,采用分组卷积^[7]的思想来减少网络参数,降低模型复杂度,设计小型网络通道混洗的多尺度分组卷积网络(multi-scale group convolution networks with channel shuffle, MSSGNet)来提高模型的识别准确率;针对特征相似不易区分而导致的准确率不高的问题,本文选择在网络中加入注意力,为此设计通道空间混合域注意力机制模块(efficient channel and spatial attention, ECSA),让模型重点关注有用信号而忽略无关噪声信号,增强信号特征区域的表征,使得模型的识别准确率获得进一步提高。

1 无人机数据集建立及特征提取

1.1 无人机声音数据集建立

采集城市公园、广场和大型游乐场等公共环境中的民用无人机声音数据,采样频率为44.1 kHz,其中包含着较高信噪比的背景噪声。实验采用的无人机包括DJI晓spark、DJI Phantom 4、DJI Mavic-mini、DJI Mavic Air、DJI Mavic Air 2、DJI Mavic 2和DJI Mavic Pro等7类无人机,作为已知无人机;将大疆悟1等一些小众无人机当作未知无人机,作为第8类,记作unknown类;将发动机、空调外机、钻孔、广场环境声、交通环境声、公园环境声等非无人机声音作为第9类,记作non-drone类。对获得的语音信号进行滤波、预加重、分帧和加窗等预处理,将长音频片段分割成1 s的短片段。建立的无人机声音数据集

的详细描述如表 1 所示,在数据集中随机选取 9 类音频各一帧信号的声谱图如图 1 所示。

表 1 自制无人机数据集的详细描述

Table 1 Detailed description of the self-made drone dataset

样本类别	数量
DJI Mavic 2	1 000
DJI Mavic Air	1 000
DJI Mavic Air 2	1 000
DJI Mavicmini	1 000
DJI Phantom 4	1 000
DJI 晓 spark	1 000
DJI Mavic Pro	1 000
unknown	1 040
non-drone	1 000
总计	9 040

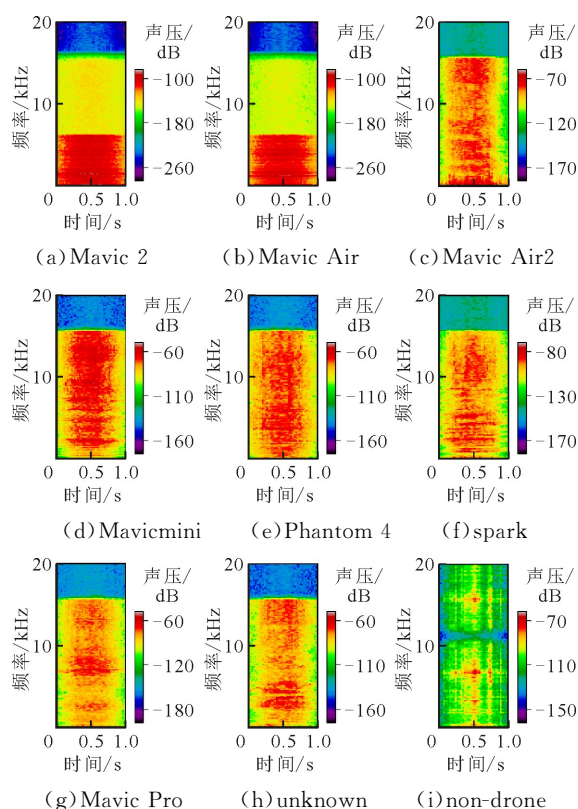


图 1 各类音频声谱图

Fig. 1 Spectrogram of various types of audio

1.2 特征提取

音频数据在输入神经网络之前,需要提取合适的特征参数来表征信号。常用的音频特征包括对数梅尔谱图(log Mel-spectrogram, log-Mel)^[8-9]、MF-CC^[10-11]等。本文采用 log-Mel 特征及其动态差分特征来表征无人机音频信号。

图 2 展示了特征生成的过程。采用 25 ms 窗口长度的短时傅里叶变换(STFT)和 34 个 Mel 滤波

器来提取 log-Mel 特征及其动态差分特征,其中 DJI Mavic 2 无人机的特征表示如图 3 所示。

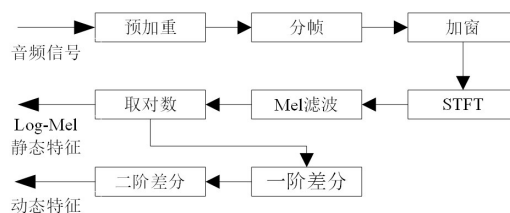


图 2 特征提取流程示意图

Fig. 2 Schematic diagram of the feature extraction process

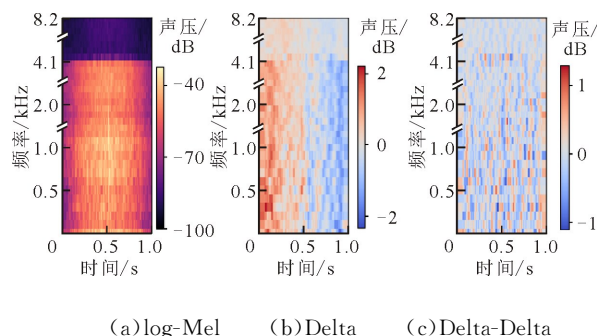


图 3 DJI Mavic 2 的 log-Mel 及差分特征谱图

Fig. 3 Log-Mel and differential feature spectrum of DJI Mavic 2

2 网络架构与混合域注意力设计

2.1 MSSGNet 网络设计

由于民用无人机声音数据集样本数量较少,当采用的分类网络较大时,其参数量过多,会产生过拟合现象,导致模型精度不高。所以,设计一种小型网络模型来减轻这种过拟合现象就显得很重要。基于此,基于分组卷积和通道混洗^[12]的思想,结合残差结构^[13],设计了小型网络 MSSGNet。

首先设计通道混洗的多尺度分组卷积模块(multi-scale group convolution module with channel shuffle, MSSG),步长(stride)为 1 时的结构示意图如图 4 所示,步长为 2 时结构相同,只是输出特征图的宽和高相应减半。采用分组卷积来减少模型参数,借鉴多尺度特征提取思想,分组后并行经过核大小为 3×3 的深度卷积^[14]以及 1×1 的卷积,再对两者结果进行拼接,其中选用小核的目的是为了在获得好的性能的同时能具有更少的参数;采用通道混洗来保证分组卷积(包括深度卷积)过程中不同组之间信息的交互;之后运用 1×1 的卷积对通道进行降维;结果通过捷径连接与输入特征图逐元素相加融合输出。其中卷积层、批归一化层(batch normalization, BN)以及整流线性单元(rectified linear

unit, ReLU)同时使用。MSSG 模块的结构示意图如图 4 所示,图中 C 、 H 、 W 为特征图的通道数、高度、宽度, $\oplus$ 表示矩阵相加。

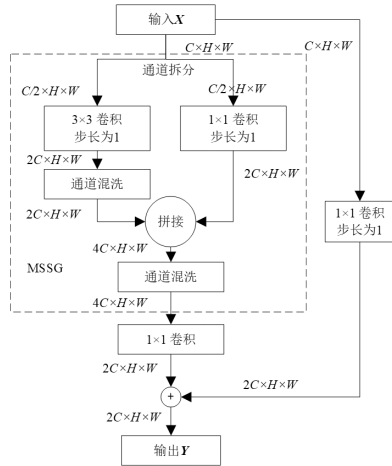


图 4 MSSG 模块的结构示意图

Fig. 4 Schematic diagram of structure of the MSSG block

为了充分提取特征, MSSGNet 网络首先经过核大小为 7×7 、步长为 1 的卷积核,并经过核大小为 2×2 、步长为 2 的最大池化操作进行降维,将它们记作 CONV1;之后经过顺序堆叠的若干 MSSG-X 模块,其中 MSSG-X 模块代表 X 个 MSSG 模块的顺序堆叠,其结构如图 5 所示;最后以全局平均池化层(GAP)和具有 Softmax 的全连接层(FC)结束。MSSGNet 网络的结构图如图 6 表示,图中每个特征图下方的标注均表示特征图的尺寸,即通道数 $C \times$ 高度 $H \times$ 宽度 W 。

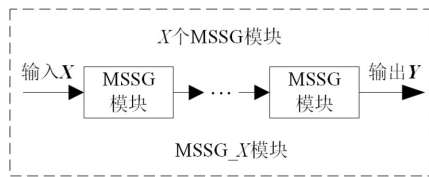


图 5 MSSG-X 模块的结构示意图

Fig. 5 Schematic diagram of structure of the MSSG-X

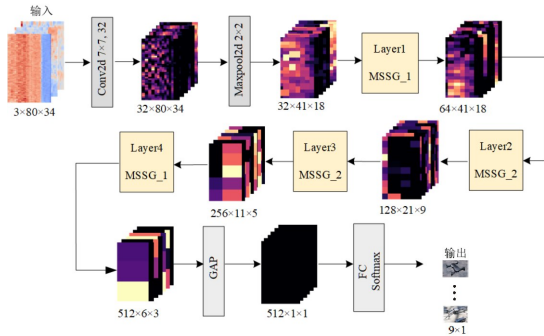


图 6 MSSGNet 网络结构示意图

Fig. 6 Structure diagram of the MSSGNet

2.2 通道空间混合域注意力机制模块(ECSA)设计

近年来,深度学习中的注意力机制在自然语言处理、图像识别和语音识别等各种任务中都获得了广泛应用,它可以增强表现力,对重要信息进行关注,弱化其他无关信息^[15-19]。为提高模型精度,选择加入注意力机制来增加模型对特征的提取能力。

2.2.1 通道注意力模块设计

为了让模型更多地关注那些能够提取无人机声音信号特征的通道,忽略提取背景噪声特征的通道,设计通道注意力模块(channel attention module, CAM)。

首先,对输入特征图 X 在空间维度进行压缩,其中 $X \in \mathbb{R}^{C \times H \times W}$,为了丰富特征,借鉴于 CBAM^[20],除了运用平均池化来聚合输入特征图 X 的空间信息外,还运用最大池化压缩空间,把两者聚合后获取的特征描述符先逐元素相加融合进行编码,再把编码后的结果通过多层感知器来学习一维通道注意力矩阵,其中加入了值为 16 的衰减率来减少参数;之后通过 h-sigmoid^[21] 函数进行归一化;最后与输入特征图 X 逐元素相乘获取加权后的输出特征图 $X' \in \mathbb{R}^{C \times H \times W}$,其结构示意图如图 7 所示,图中 FC 表示全连接层, $\otimes$ 表示矩阵乘法。对特征图 X 提取通道注意力矩阵 $A(X)$ 可表示为

$$A(X) = f(M(\text{Avgpool}(X) + \text{Maxpool}(X))) \quad (1)$$

$$X' = XA(X) \quad (2)$$

式中: $f(\cdot)$ 为 h-sigmoid 激活函数; $\text{Avgpool}(\cdot)$ 、 $\text{Maxpool}(\cdot)$ 分别为全局平均池化函数、最大池化函数; $M(\cdot)$ 表示经过中间处理模块处理后的矩阵。

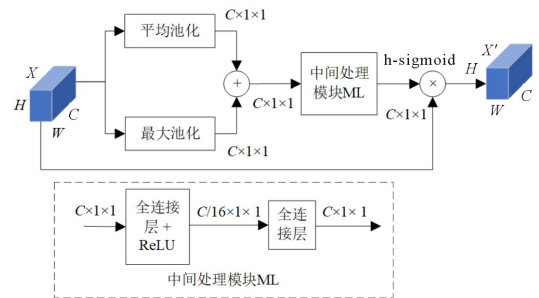


图 7 通道注意力模块 CAM 模块结构示意图

Fig. 7 Schematic diagram of structure of the CAM

2.2.2 空间注意力模块设计

通道注意力忽略了空间信息,通过空间注意力来弥补。输入网络的声音信号特征谱图,其水平方向是每帧信号特征所在的维度,代表帧信号的特征,而垂直方向是时间所在维度,表示时序信号在时间上帧与帧信号之间的联系,它们在各自维度上都占据着不同的比重。为了突出两个维度空间的信息,

设计了两条支路来分别对特征和时间两个维度提取注意力向量,让模型更多地关注信号的重要特征及时间区域,称其为空间注意力模块(spatial attention module, SAM)。

首先采用 1×1 卷积在通道维度对输入特征图 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ 进行压缩;之后分出两条支路,分别运用一维平均池化聚合特征图的水平和垂直方向上的信息,再通过 $7 \times 1, 1 \times 7$ 卷积分别获取一维垂直及水平注意力向量,并运用 sigmoid 函数进行归一化,其中选用 7×1 核大小的原因在于:在对不同卷积核大小的比较中,发现采用更大的核会产生更好的精度,这也意味着需要一个大的感兴趣区域来决定空间上重要的区域;最后把获得的两个一维向量与输入特征图逐元素相乘获取加权后的输出特征图 $\mathbf{Y} \in \mathbb{R}^{C \times H \times W}$ 。其结构示意图如图 8 所示,其中 θ, ϕ 和 λ 分别表示大小为 $3 \times 3, 7 \times 1$ 和 1×7 的卷积核; σ 表示 sigmoid 激活函数,表达式为

$$\mathbf{A}_H(\mathbf{X}) = \sigma(\text{Avgpool}(\mathbf{W}_\theta \mathbf{X}) \mathbf{W}_\varphi) \quad (3)$$

$$\mathbf{A}_W(\mathbf{X}) = \sigma(\text{Avgpool}(\mathbf{W}_\theta \mathbf{X}) \mathbf{W}_\lambda) \quad (4)$$

$$\mathbf{Y} = \mathbf{X} \mathbf{A}_W(\mathbf{X}) \mathbf{A}_H(\mathbf{X}) \quad (5)$$

式中: $\sigma(\cdot)$ 为 sigmoid 激活函数; $\mathbf{W}_\theta, \mathbf{W}_\varphi, \mathbf{W}_\lambda$ 为不同卷积核权重参数; $\mathbf{A}_H(\mathbf{X})$ 为对特征图 $\mathbf{X}$ 提取一维垂直注意力向量; $\mathbf{A}_W(\mathbf{X})$ 为对特征图 $\mathbf{X}$ 提取一维水平注意力向量。

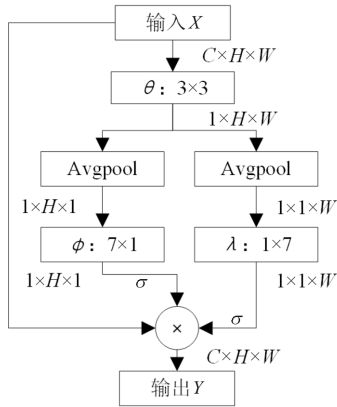


图 8 空间注意力模块 SAM 模块的结构示意图

Fig. 8 Schematic diagram of structure of the SAM

2.2.3 通道空间混合域注意力机制模块设计

类似 CBAM, 通道空间顺序连接: 输入特征图 $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$, 先经过 CAM 模块获取一维通道注意力图 $\mathbf{F}_C \in \mathbb{R}^{C \times 1 \times 1}$, 与 $\mathbf{X}$ 逐元素相乘得到中间特征图 $\mathbf{X}' \in \mathbb{R}^{C \times H \times W}$; 之后通过 SAM 模块获取一维垂直注意力图 $\mathbf{F}_H \in \mathbb{R}^{1 \times H \times 1}$ 和一维水平注意力图 $\mathbf{F}_W \in \mathbb{R}^{1 \times 1 \times W}$, 两者一同与 $\mathbf{X}'$ 逐元素相乘, 获得最终输出

特征图 $\mathbf{Y} \in \mathbb{R}^{C \times H \times W}$, 称为通道空间混合域注意力机制模块 (efficient channel and spatial attention, EC-SA)。其结构示意图如图 9 所示, 其值可表示为

$$\mathbf{X}' = \mathbf{F}_C(\mathbf{X}) \otimes \mathbf{X} \quad (6)$$

$$\mathbf{Y} = \mathbf{X}' \otimes \mathbf{F}_H(\mathbf{X}') \otimes \mathbf{F}_W(\mathbf{X}') \quad (7)$$

式中: $\otimes$ 表示逐元素相乘; $\mathbf{F}_C, \mathbf{F}_H, \mathbf{F}_W$ 表示对特征图提取相应注意力矩阵。

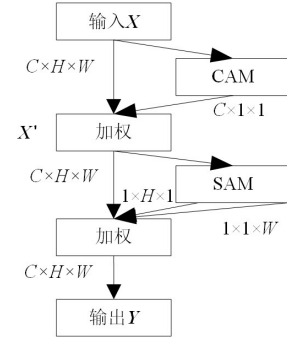


图 9 ECSA 模块结构示意图

Fig. 9 Schematic diagram of structure of the ECSA

2.3 通道空间混合域注意力机制的多尺度分组卷积网络设计

为了进一步提高网络模型对无人机的识别准确率, 在 MSSGNet 网络的 MSSG 模块中插入了混合域注意力机制模块 ECSA 模块, 位置处于 MSSG 模块中的通道降维后, 合称为 ECSAM 模块, 其结构示意图如图 10 所示。同样, 将 X 个 ECSAM 模块的顺序堆叠表示成 ECSAM_X 模块, 其结构图如图 11 所示。其余结构保持不变, 将 MSSGNet 网络中的 MSSG 模块替换为 ECSAM 模块后的网络称为通道空间混合域注意力机制的多尺度分组卷积网络 (Multiscale group convolution network with attention mechanism in mixed domain of channel space, ECSANet), 其网络结构示意图如图 12 所示。

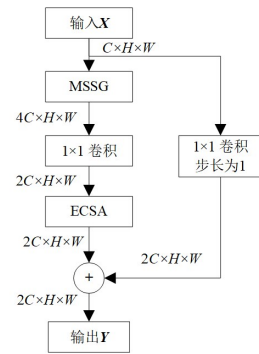


图 10 ECSAM 模块的结构示意图

Fig. 10 Schematic diagram of structure of the ECSAM block

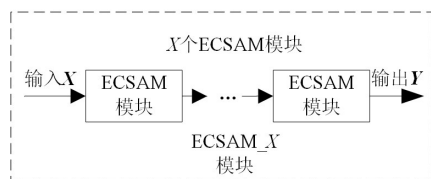


图 11 ECSAM_X 模块的结构示意图

Fig. 11 Schematic diagram of structure of the ECSAM_X

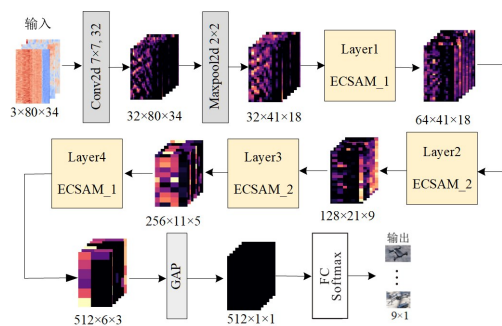


图 12 ECSANet 网络结构示意图

Fig. 12 Schematic diagram of structure of the ECSANet

3 实验

所有实验在 Python 3.7.6 环境、Windows10 平台上完成。运行设备 CPU 型号为 i7-9750H, 显卡为 GTX1660 Ti。在 GPU 环境下完成对网络模型的学习训练。

3.1 自制无人机数据集上的对比实验

该部分实验在自制无人机声音数据集上进行, 其中训练集和测试集的比例为 3:1。网络训练时采用带动量的 SGD 优化器, 动量的值为 0.9, 初始学习率为 0.01, 每经过 30 轮学习率下降为原来的 0.1 倍, 共训练 90 轮; 采用交叉熵损失函数; 批次大小设置为 128。

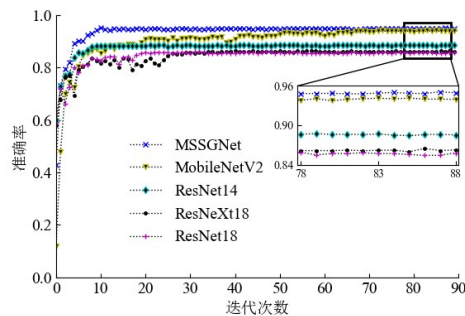
3.1.1 MSSGNet 网络与其他基准网络的对比实验

该实验是为了对所设计的 MSSGNet 网络进行评估。选择在自制无人机声音数据集上与其他基准网络之间进行比较, 其中基准网络有 ResNet18、ResNet34、ResNeXt18^[22] 和 MobileNetV2^[23]。运用准确率作为评价指标。其实验结果如表 2 和图 13 所示。网络的参数量可以体现模型的空间复杂度, 网络参数量越大, 训练网络时所需的数据量就越大。而当数据集样本数量太小时, 模型的训练更容易引起过拟合问题。

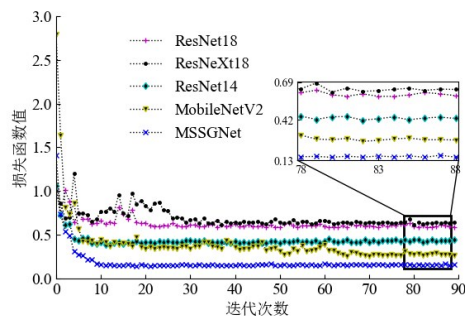
表 2 MSSGNet 与其他基准网络模型的实验结果比较

Table 2 Comparison of experimental results between MSSGNet and other baseline network models

网络模型	参数量/ 10^6	准确率/%
ResNet34	21.3	84.8
ResNet18	11.2	86.1
ResNeXt18	5.1	86.5
MobileNetV2	2.2	94.2
MSSGNet	1.4	95.1



(a) 准确率对比曲线



(b) 损失函数对比曲线

图 13 MSSGNet 网络与 4 种基准网络模型的准确率和损失函数对比曲线

Fig. 13 Accuracy and loss function comparison curves between MSSGNet and other baseline network models

从表 2 可以看出, MSSGNet 网络的参数量最少, 且识别准确率最高, 可以达到 95.1%。当运用其对无人机声音数据集进行识别时, 可以缓解过拟合现象。由图 13 可知, 与其他基准网络相比, MSSGNet 网络性能最好, MSSGNet 网络可以很好地应对无人机小样本识别任务, 且性能较好。

3.1.2 通道空间注意力机制模块 ECSA 模块的消融实验

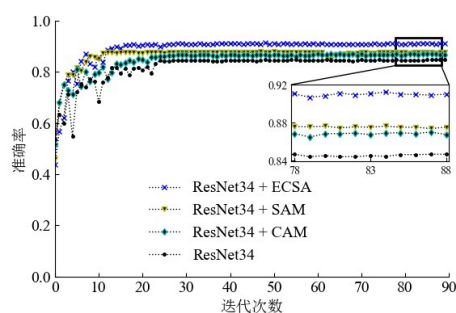
该实验是混合注意力 ECSA 模块的消融实验, 选择 ResNet34 作为基准网络, 实验时分别加入通道注意力 CAM 模块、空间注意力 SAM 模块以及混合注意力 ECSA 模块, 研究它们对模型识别准确率

的影响。将提取的特征参数分别传入网络进行训练,运用准确率作为模型的评价指标,最终实验结果如表 3 和图 14 所示。

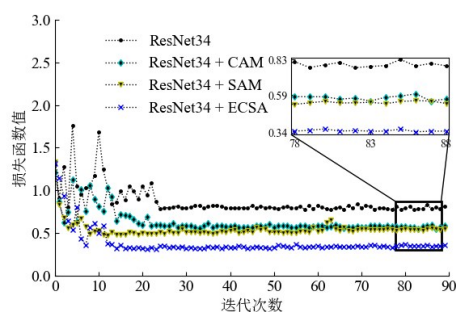
表 3 ResNet34 加入 3 种注意力模块后网络模型的相关实验结果对比表

Table 3 Comparison of related experimental results of ResNet34 after adding three attention models

网络模型	参数量/ 10^6	准确率/%
ResNet34 (Baseline)	21.29	86.4
ResNet34+CAM	21.45	88.9
ResNet34+SAM	21.33	89.6
ResNet34+ECSA	21.48	91.3



(a) 准确率对比曲线



(b) 损失函数对比曲线

图 14 ResNet34 加入注意力后的准确率和损失函数对比曲线

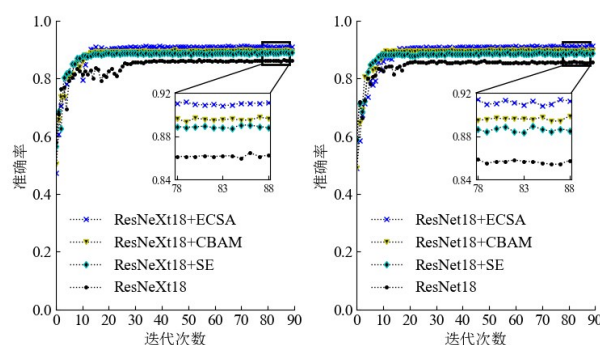
Fig. 14 Comparison curve of accuracy and loss function of ResNet34 after adding three kinds of attention

从表 3 和图 14 可以看出,基准网络 ResNet34 在加入 3 种注意力模块后,其识别准确率均获得了提高,在其中加入混合注意力 ECSA 模块后网络模型的识别准确率最高,相比于 ResNet34 提高了 4.9%,相比于单纯的通道、空间注意力分别提高了 2.5%、3.2%,并且网络模型增加的参数量基本上可以忽略不计。实验结果证明了混合注意力 ECSA 模块相比于单纯的通道和空间注意力,可以使模型的识别准确率变得更高;并且也表明了混合域注意

力 ECSA 模块是轻量的,在插入神经网络模型后带来的参数可以忽略不计。实验结果证明了设计的混合注意力 ECSA 模块的有效性。

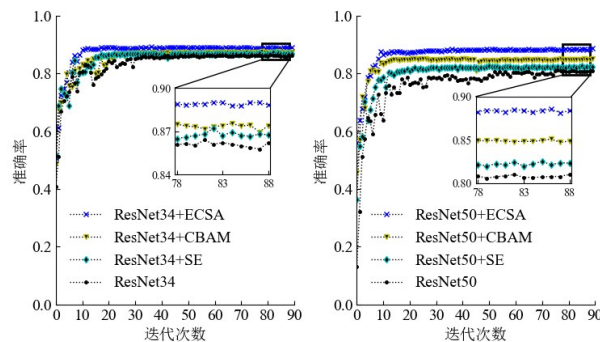
3.1.3 ECSA 模块与其他注意力机制模块的对比实验

该实验是对混合域注意力模块 ECSA 的可行性验证以及与其他诸如 SE (squeeze & excitation block)^[24]、CBAM (convolutional block attention module)等常用注意力机制模块的优越性对比。选择 ResNeXt18、ResNet18、ResNet34 和 ResNet50 等网络作为基准网络,将提取的特征参数分别传入各种网络进行训练。运用准确率作为模型的评价指标,最终各个模型的准确率对比曲线如图 15 所示。



(a) ResNeXt18

(b) ResNet18



(c) ResNet34

(d) ResNet50

图 15 加入不同注意力机制模块的基准网络模型的准确率对比曲线

Fig. 15 Accuracy comparison curve of baseline network models with different attention

从图 15 中可以看出,相比于各种基准网络,加入注意力后,基准网络模型的识别准确率均获得了提高,且加入设计的混合注意力 ECSA 模块后的基准网络模型识别准确率最高。实验结果证明了设计的混合注意力模块 ECSA 模块的可行性和优越性,表明了 ECSA 模块可以插入各种神经网络中来提高模型的识别准确率。研究表明,针对无人机声音

识别问题,设计的混合域注意力 ECSA 模块可以使得网络模型更准地识别无人机。

3.1.4 ECSANet 网络的对比实验

该实验是对 ECSANet 网络进行的可行性验证实验。选择 MSSGNet 网络为基准网络,把特征参数分别传入 MSSGNet、ECSANet 网络,最终模型的准确率对比曲线如图 16 所示,两者的混淆矩阵如图 17 所示。

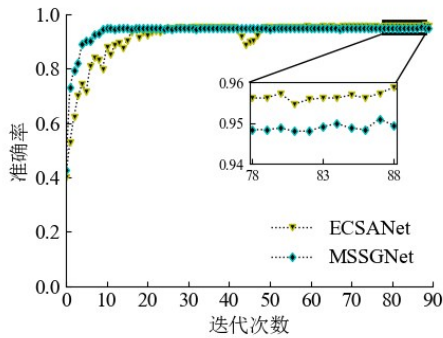


图 16 MSSGNet 网络与 ECSANet 网络模型的准确率对比曲线

Fig. 16 Accuracy comparison curve between MSSGNet and ECSANet models

预测标签	Mavic 2	Mavic Air	Mavic Air2	Mavic mini1	Phantom4	spark	Mavic Pro	unknown	non-drone
Mavic 2	280	7	0	0	0	0	0	0	0
Mavic Air	2	281	0	0	0	0	0	0	0
Mavic Air2	0	0	291	2	1	2	14	2	0
Mavic mini1	0	0	2	266	3	11	3	5	0
Phantom4	0	0	3	10	303	7	0	5	0
spark	0	0	1	1	2	282	0	4	0
Mavic Pro	0	0	15	2	1	0	267	9	0
unknown	0	0	2	1	1	7	6	297	0
non-drone	0	0	0	0	1	0	0	1	312
真实标签	Mavic 2	Mavic Air	Mavic Air2	Mavic mini1	Phantom4	spark	Mavic Pro	unknown	non-drone

(a) MSSGNet 网络

预测标签	Mavic 2	Mavic Air	Mavic Air2	Mavic mini1	Phantom4	spark	Mavic Pro	unknown	non-drone
Mavic 2	280	5	0	0	0	0	0	0	0
Mavic Air	2	283	0	0	0	0	0	0	0
Mavic Air2	0	0	291	0	0	1	10	1	0
Mavic mini1	0	0	5	270	4	6	0	8	0
Phantom4	0	0	1	6	301	4	1	1	0
spark	0	0	0	3	2	292	0	3	0
Mavic Pro	0	0	15	1	2	2	269	8	0
unknown	0	0	2	2	2	4	10	302	0
non-drone	0	0	0	0	1	0	0	0	312
真实标签	Mavic 2	Mavic Air	Mavic Air2	Mavic mini1	Phantom4	spark	Mavic Pro	unknown	non-drone

(b) ECSANet 网络

图 17 MSSGNet 网络与 ECSANet 网络的混淆矩阵图

Fig. 17 Confusion matrix diagram of MSSGNet and ECSANet

从图 16 可以看出,与 MSSGNet 网络相比,ECSANet 网络识别准确率获得了提高,由 95.1% 提高到了 95.9%。从图 17 可以看出,相比于 MSSGNet 网络,ECSANet 网络对于 DJI Mavic mini 和 DJI 晓 spark 这两类无人机的预测能力得到了增强,说明 MSSGNet 网络加入混合域注意力机制模块 ECSA 模块后,其对于信号的特征提取能力得到增强,从而使得网络模型对于复杂类别间的区分度得到增强,进而使得网络模型的识别准确率得到提高。实验证明了设计的网络 ECSANet 网络的有效性,它可以更准确地对无人机进行识别。

3.2 公共环境声音数据集 Urbansound8K 上的验证实验

公共环境声音数据集 Urbansound8K^[25] 是由 8 732 个带标签的声音片段组成的数据集,每个声音片段具有最大 4 s 的持续时间。8 732 段录音来自 10 个声音类别,即汽车喇叭、狗吠、发动机空转、风钻、空调、街头音乐、儿童玩耍、钻探、枪声和警笛。数据集由 Urbansound8K 的一部分构成,分别包括空调、儿童游戏、钻孔、发动机怠速以及手提钻共 5 类,每类包含 730 个片段。这部分实验主要是对 MSSGNet 网络以及 ECSANet 网络模型进行可行性验证。

同样,采用 log-Mel 及其动态差分特征作为表征音频信号的特征参数。实验时训练集和验证集的比例为 3:1,网络训练时采用带动量的 SGD 优化器,动量取值 0.9,初始学习率为 0.01,每经过 30 个 epoch 学习率下降为原来的 0.1 倍,共训练 90 个 epoch;损失函数采用的是交叉熵损失函数;每次传入网络数据的批次大小设置为 128。运用准确率作为模型的评价指标。

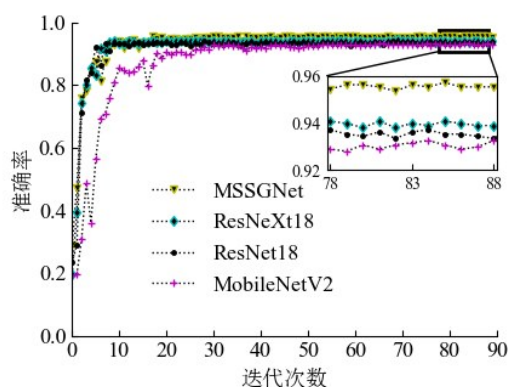
3.2.1 MSSGNet 网络的验证实验

该实验是为了验证所设计网络 MSSGNet 的可行性与优越性。采用 ResNet18、ResNeXt18 和 MobileNetV2 作为基准网络,提取特征参数输入几种网络,观察实验结果,实验结果如图 18 所示。相比于其他基准网络模型,MSSGNet 网络的识别准确率最高。实验结果证明了设计的 MSSGNet 网络的可行性和有效性,且它具有一定的优越性。

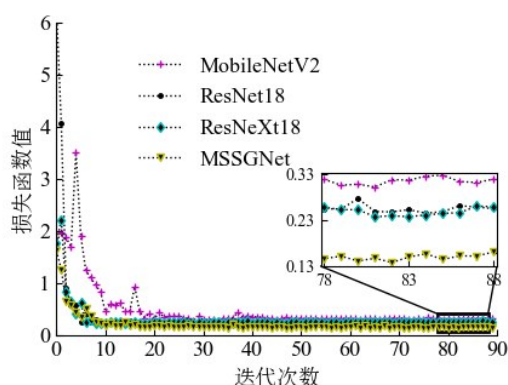
3.2.2 ECSANet 网络的验证实验

该实验是对 ECSANet 网络进行的可行性验证实验。将 MSSGNet 网络作为基准网络,把特征参数分别传入 MSSGNet、ECSANet 网络,最终模型的准确率对比如表 4 所示。可知相比于 MSSGNet 网络,ECSANet 网络识别准确率有了进一步提高。实

验证了 ECSANet 网络的可行性和有效性。



(a) 准确率对比曲线



(b) 损失函数对比曲线

图 18 MSSGNet 网络与基准网络模型的准确率和损失函数对比曲线

Fig. 18 Comparison curve of accuracy and loss function between the MSSGNet and other baseline network models

表 4 MSSGNet、ECSANet 网络的实验结果对比

Table 4 Comparison of network parameters and recognition accuracy between MSSGNet and ECSANet

方法	参数量/ 10^6	识别准确率/%
MSSGNet	1.4	95.8
ECSANet	1.5	96.1

4 结 论

(1) 为了提高反无人机系统声音识别小样本无人机的准确率,提出了一种融合通道空间混合域注意力的 ECSANet 网络。基于分组卷积、通道混洗和残差结构设计了 MSSGNet 网络,设计了通道空间混合域注意力模块 ECSA 模块,将 ECSA 模块插入 MSSGNet 网络构成基于通道空间混合域注意力的改进网络 ECSANet 网络,它参数量少,对无人机

声音信号特征具有很好的提取能力。

(2) 在自制无人机数据集和环境声音数据集 Urbansound8K 上进行实验验证。实验结果表明,与其他基准网络相比,MSSGNet 网络参数更少,识别准确率更高,在无人机等小样本识别任务上具有不错的性能;ECSA 模块可以插入多种网络,可以在不增加太多参数的情况下令网络模型的识别准确率获得提升,在无人机等声音分类任务上具有很好的效果;与 MSSGNet 网络相比,ECSANet 网络的识别准确率更高,表明了设计的网络在识别小样本无人机方面的优越性和可行性。

参考文献:

- [1] 罗俊海,王芝燕. 无人机探测与对抗技术发展及应用综述 [J]. 控制与决策, 2022, 37(3): 530-544.
LUO Junhai, WANG Zhiyan. A review of development and application of UAV detection and counter technology [J]. Control and Decision, 2022, 37(3): 530-544.
- [2] 陈唯实,黄毅峰,卢贤锋. 多传感器融合的无人机探测技术应用综述 [J]. 现代雷达, 2020, 42(6): 15-29.
CHEN Weishi, HUANG Yifeng, LU Xianfeng. Survey on application of multi-sensor fusion in UAV detection technology [J]. Modern Radar, 2020, 42(6): 15-29.
- [3] AL-EMADI S, AL-ALI A, AL-ALI A. Audio-based drone detection and identification using deep learning techniques with dataset enhancement through generative adversarial networks [J]. Sensors, 2021, 21(15): 4953.
- [4] ISMAIL M A A, BIERIG A. Identifying drone-related security risks by a laser vibrometer-based payload identification system [C]//Proceedings of Laser Radar Technology and Applications. Bellingham, WA, USA: SPIE, 2018: 1063603.
- [5] SEO Y, JANG B, IM S. Drone detection using convolutional neural networks with acoustic STFT features [C]//2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway, NJ, USA: IEEE, 2018: 1-6.
- [6] CASABIANCA P, ZHANG Yu. Acoustic-based UAV detection using late fusion of deep neural networks [J]. Drones, 2021, 5(3): 54.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Image net classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [8] 李云红,梁思程,贾凯莉,等. 一种改进的 DNN-

- HMM 的语音识别方法 [J]. 应用声学, 2019, 38(3): 371-377.
- LI Yunhong, LIANG Sicheng, JIA Kaili, et al. An improved speech recognition method based on DNN-HMM model [J]. Applied Acoustics, 2019, 38(3): 371-377.
- [9] DUMITRESCU C, MINEA M, COSTEA I M, et al. Development of an acoustic system for UAV detection [J]. Sensors, 2020, 20(17): 4870.
- [10] JIN Shuaipu, WANG Xiufeng, DU Leilei, et al. Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN [J]. Applied Acoustics, 2021, 172: 107562.
- [11] SHUKLA J, BARREDA-ÁNGELES M, OLIVER J, et al. Feature extraction and selection for emotion recognition from electrodermal activity [J]. IEEE Transactions on Affective Computing, 2021, 12(4): 857-869.
- [12] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 6848-6856.
- [13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [14] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2022-01-02]. <https://arxiv.org/abs/1704.04861>.
- [15] 余浩帅, 汤宝平, 张楷, 等. 小样本下混合自注意力原型网络的风电齿轮箱故障诊断方法 [J]. 中国机械工程, 2021, 32(20): 2475-2481.
- YU Haoshuai, TANG Baoping, ZHANG Kai, et al. Fault diagnosis method of wind turbine gearboxes mixed with attention prototype networks under small samples [J]. China Mechanical Engineering, 2021, 32(20): 2475-2481.
- [16] 郑作武, 邵斯琦, 高晓沅, 等. 基于社交圈层和注意力机制的信息热度预测 [J]. 计算机学报, 2021, 44(5): 921-936.
- ZHENG Zuowu, SHAO Siqu, GAO Xiaofeng, et al. Social circle and attention based information popularity prediction [J]. Chinese Journal of Computers, 2021, 44(5): 921-936.
- [17] 孙红帅, 王霞, 柳莹, 等. 频域注意力机制下的癫痫脑电信号分类 [J]. 西安交通大学学报, 2021, 55(2): 129-135.
- SUN Hongshuai, WANG Xia, LIU Xuan, et al. A classification method of epileptic electroencephalograms under frequency-domain attention mechanism [J]. Journal of Xi'an Jiaotong University, 2021, 55(2): 129-135.
- [18] 莫仁鹏, 李天梅, 司小胜, 等. 采用残差网络与卷积注意力机制的设备剩余使用寿命预测方法 [J]. 西安交通大学学报, 2022, 56(4): 194-202.
- MO Renpeng, LI Tianmei, SI Xiaosheng, et al. Remaining useful life prediction for equipment using residual network and convolutional attention mechanism [J]. Journal of Xi'an Jiaotong University, 2022, 56(4): 194-202.
- [19] 张连超, 乔瑞萍, 党祺玮, 等. 具有全局特征的空间注意力机制 [J]. 西安交通大学学报, 2020, 54(11): 129-138.
- ZHANG Lianchao, QIAO Ruiping, DANG Qiwei, et al. Spatial attention mechanism with global characteristics [J]. Journal of Xi'an Jiaotong University, 2020, 54(11): 129-138.
- [20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham, Germany: Springer International Publishing, 2018: 3-19.
- [21] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3 [C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2019: 1314-1324.
- [22] XIE Saining, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 5987-5995.
- [23] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 4510-4520.
- [24] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [25] 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, NY, USA: ACM, 2014: 1041-1044.

(编辑 赵炜 刘杨)