

DOI:10.19651/j.cnki.emt.2210702

基于 ZYNQ 的枪声识别系统设计^{*}孙袖山^{1,2} 李剑^{1,2} 贺斌^{1,2} 庞润嘉^{1,2} 马翊翔^{1,2} 郭锦铭^{1,2}

(1. 中北大学省部共建动态测试技术国家重点实验室 太原 030051;

2. 中北大学信息探测与处理山西省重点实验室 太原 030051)

摘要: 枪声识别技术在军事环境下可以快速准确地提供战场信息,但是目前大部分枪声识别系统均部署在服务器端,实用性和可行性不高,针对这一问题,本文设计了一种基于 ZYNQ 的枪声识别系统。该系统以 ZYNQ7020 芯片为核心,充分利用 ZYNQ 芯片集 ARM 与 FPGA 于一体的特性,首先在芯片的 FPGA 部分设计了多通道数据传输链路和声场特征参数提取模块;其次在芯片的 ARM 部分部署经过 PC 端训练后的轻量化网络模型,对经过 FPGA 提取的特征参数进行处理,进而实现对枪声种类的识别;最后使用枪声数据集 NIJ Grant 2016-DN-BX-0183 中的 3 种枪声在外场进行试验。试验结果表明,该系统能够准确地对枪声进行分类,枪声的平均识别率达到 91.67%。该成果在枪声识别领域具有较强的应用价值。

关键词: ZYNQ;对数梅尔谱;枪声识别;神经网络

中图分类号: TJ206 **文献标识码:** A **国家标准学科分类代码:** 510.4030

Design of gun sound recognition system based on ZYNQ

Sun Xiushan^{1,2} Li Jian^{1,2} He Bin^{1,2} Pang Runjia^{1,2} Ma Yixiang^{1,2} Guo Jinming^{1,2}

(1. State Key Laboratory of Dynamic Testing Technology, North University of China, Taiyuan 030051, China;

2. Key Laboratory of Information Detection and Processing of Shanxi Province, North University of China, Taiyuan 030051, China)

Abstract: Gunshot recognition technology can quickly and accurately provide battlefield information in a military environment, but most gunshot recognition systems are currently deployed on the server side, which is not practical and feasible. To solve this problem, this paper designs a ZYNQ-based technology gun sound recognition system. The system takes ZYNQ7020 chip as the core, and makes full use of the ZYNQ chip that integrates ARM and FPGA. First, a multi-channel data transmission link and a sound field feature parameter extraction module are designed in the FPGA part of the chip. Secondly, the ARM part of the chip is deployed after the lightweight network model trained on the PC side processes the feature parameters extracted by the FPGA to identify the types of gunshots. Finally, the three gunshots in the gunshot dataset NIJ Grant 2016-DN-BX-0183 are used. Test in the field. The test results show that the system can accurately classify gunshots, and the average recognition rate of gunshots reaches 91.67%. This achievement has strong application value in the field of gunshot recognition.

Keywords: ZYNQ; log-mel spectrogram; gunshot classification; neural network

0 引言

目前,音频信号处理特别是音频识别技术已经广泛应用于人们生活中,如声纹密码,门禁系统等。枪声作为声音信号中尤为特殊的一种,一直备受关注。枪声识别技术在军事环境下可以快速准确地提供战场信息^[1],并且该技术在公共安全^[2]、自然保护^[3]等多个领域也都具有极其重要的意义。因此,如何将该技术应用于实际逐步成为人们关

注的问题。

而枪声识别系统就是枪声识别技术的实际应用,这种系统使用声传感器接收枪支在发射瞬间产生的音频信号,然后通过算法进行处理,进而实现对枪声种类的识别^[4]。针对枪声识别系统,国外进行的研究相对较早。2019年, Galangque 等^[5]使用 DAQ 系统以及 Matlab 工具实现了针对 M16 的枪声识别系统,该系统区分 M16 枪声与鞭炮声准确率达 99%;2021年, Tardif 等^[6]使用两套数据采集系

收稿日期:2022-07-17

^{*} 基金项目:国家自然科学基金(62101512)、山西省青年科学基金(20210302124031)项目资助

统采集枪声信号,在 PC 端完成了对 4 种枪声的识别,获得了 99% 的识别率。上述枪声识别系统大多集中于理论的研究以及算法精度的提升,在 PC 端对音频数据进行处理时,均能取得较好的效果,但是由于其过于依赖强大的算力支撑,在实际应用中的可行性缺乏考虑,因此很难应用于相关领域。

针对现有枪声识别系统缺乏实用性的问题,本文进行了基于 ZYNQ 的枪声识别系统设计。该系统以 ZYNQ7020 为主控芯片,由芯片的 FPGA 部分完成十六通道枪声信号的采集,并实现枪声信号 Log-Mel 特征参数的提取;在芯片内嵌的 ARM 部分搭载 Linux 操作系统,通过轻量化神经网络模型完成对枪声种类的判定,上述工作完成后,使用枪声数据集在外场进行试验,试验结果表明,该系统实现了多路音频信号的同步实时获取,并且完成了对于枪声信号种类的识别,弥补了现有枪声识别系统实用性不高的问题,为枪声识别技术的实际应用提供了强有力的技术支撑。

1 系统总体方案设计

1.1 枪声识别原理概述

枪声信号成分复杂,主要包括膛口波和马赫波^[7],枪口声波如图 1 所示。当枪击发时,膛内火药在狭窄的空间内燃烧产生膨胀的气流将子弹推出枪膛,子弹出膛后,膛口

喷出的气流形成的膛口波,其作用时间短,声音强度大,通常作为判断枪声种类和位置的重要依据;当子弹的速度超过音速时,弹头在空中飞行时与周围的空气摩擦会形成马赫波,从子弹的路径向外传播^[8],马赫波作为子弹的特征,是判断枪支口径的重要依据。

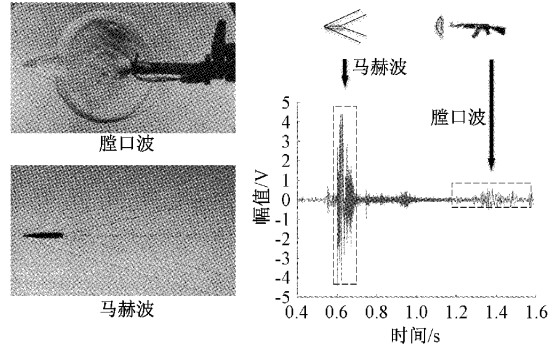


图 1 枪口声波示意图

在枪声识别时同时考虑膛口波、马赫波两种音频成分即可实现对枪支种类、口径的有效识别。

1.2 硬件系统总体设计

根据本系统需要实现的功能,将枪声识别系统分为阵列化声场信号获取、声场特征参数硬件提取和轻量化枪声识别网络 3 部分。系统整体设计框图如图 2 所示。

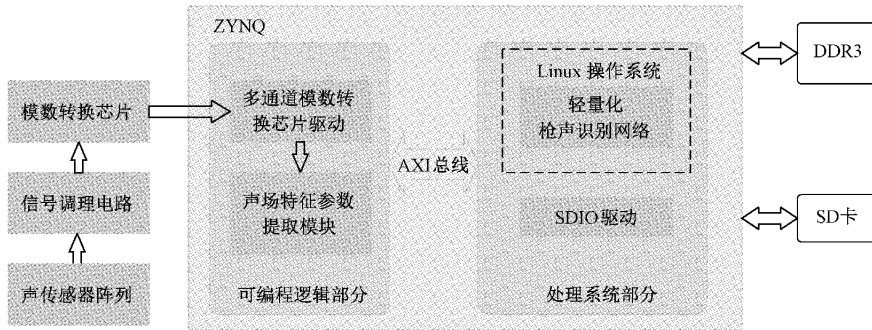


图 2 系统设计框图

如图 2 所示,系统使用声传感器阵列接收枪声的膛口波和马赫波信号,所采集的模拟信号经过多路 ADS8681 量化后,由可编程逻辑(programmable logic, PL)部分的模数转换芯片驱动模块接收,再送入声场特征参数提取模块经过分帧、加窗、快速傅里叶变换(fast Fourier transform, FFT)以及 Mel 滤波等操作提取特征参数即 Log-Mel 谱图,然后将谱图存入 DDR3 中;系统在识别枪声时,将 DDR3 中的谱图读出,输入到处理系统(processing system, PS)的 linux 系统中,通过经服务器训练后的轻量化网络模型输出枪声类别信息,实现枪声的高速、高精度识别。

2 系统设计

2.1 阵列化声场信号获取

1) 阵列化声场信号调理电路设计

为保证音频信号的精度,声传感器选用 PCB 公司的预

极化高灵敏度声传感器 376B02,该传感器灵敏度为 50 mV/Pa,频响范围达 20 kHz,满足音频采集的需求。针对该 ICP 型传感器,设计相应的适配电路以及调理电路以匹配模数转换芯片,调理电路如图 3 所示。

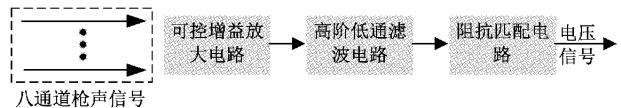


图 3 调理电路示意图

可控增益放大电路:采用高精度可编程运放 PGA281,根据前端信号的动态范围确定信号的放大倍数,使得放大后的信号与 ADC 可采集的电压范围相适应。

高阶低通滤波电路:使用声传感器采集枪声时,信号中会包含大量的风噪以及环境噪声,因此采用 MAX291 芯片搭配外围电路实现八阶巴特沃斯低通滤波器滤除高频

噪声;

阻抗匹配电路:使用轨对轨精密运放芯片 ADA4610 搭建电压跟随电路实现阻抗匹配,利用其高输入阻抗和低输出阻抗特性,提高带载能力。

2) 高速多通道信号采集

根据奈奎斯特采样定律^[9],模数转换芯片采样率应大于两倍声传感器响应频率 40 kHz,为确保信号质量,通常选用采样率为音频信号频率的 5~6 倍。综合考虑后续声源分类算法的精度要求以及数据处理的实时性要求,选用 ADS8681 模数转换芯片。芯片连接如图 4 所示。

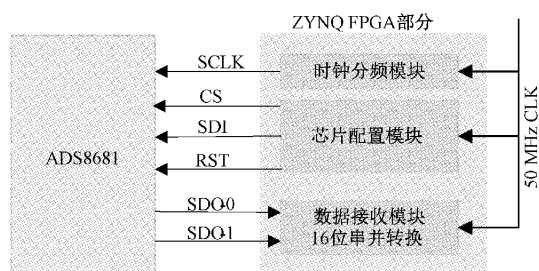


图 4 AD 接口设计

通过设计有限状态机 (finite-state machine, FSM) 实现

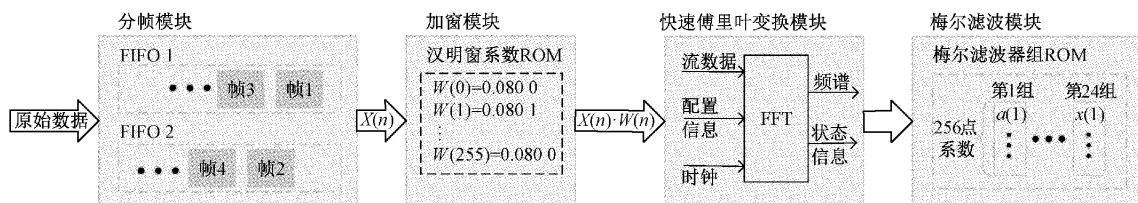


图 6 Log-Mel 特征提取流程

分帧模块:在 VIVADO 中调用两个 FIFO,设定 FIFO 位宽为 16,深度为一帧语音的长度 256,并设计 FSM 有限状态机控制 FIFO 的使能引脚,保证两 FIFO 输出的数据之间有半帧的帧移,两 FIFO 交替输出 256 位数据,进而实现对音频信号的分帧处理;

加窗模块:通过汉明窗公式即式(1)预先计算出各点的汉明窗系数,并调用 ROM 核存储该系数,音频帧各点数据与其相乘得到加窗后的数据;

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2 \cdot \pi \cdot n}{N}\right) \quad 1 \leq n \leq N, N = 256 \quad (1)$$

FFT 模块:通过在 Block Design 中调用 Fast Fourier transform IP 核实现快速傅里叶变换,本设计中设置 FFT 的位宽为 16,点数为 256 点;

Mel 滤波模块:FFT 输出为 32 bit 数据流,其高 16 位为虚部。低 16 位为实部,首先对复数取模获得其幅值。此次采样率 $f_s = 100 \text{ K}$,Mel 滤波器的最低频率为 $f_l = 0$,最高频率为 $f_h = f_l / 2 = 50 \text{ K}$,设置三角滤波器的数量为 24,FFT 点数为 256。将该滤波器组的系数以及频率点预先存入 ROM 中,在滤波时将 FFT 模块输出的频谱数据与

ADC 芯片对音频信号的采集,并将转换完成的数据封装为 128 位 AXI 型的数据,以适配后续调用的 AXI 型的 IP 核,ADC 状态转换图如图 5 所示。

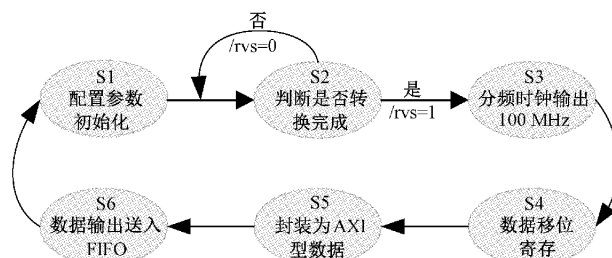


图 5 ADC 接口状态转换图

2.2 声场特征参数硬件提取

Mel 谱系数是基于人耳的听觉特性提出的一种音频特征,该系数计算的是音频短时功率谱的包络,能够很好地表征声道特性,目前已经广泛应用于音频识别与检测领域^[10]。基于 FPGA 的并行计算优势,Mel 特征提取可以利用 FPGA 内部丰富的数字信号处理器 (digital signal processor, DSP) 以及逻辑资源完成,可以大幅提高运算速度。其硬件实现的整体流程如图 6 所示。

ROM 中读出的数据相乘得到计算后的结果,最后对该结果进行对数运算以求得 Log-Mel 特征参数。

2.3 轻量化枪声识别网络

1) 枪声识别网络设计

ZYNQ7020 的 ARM Cortex-A9 处理器搭载 Linux 系统,可以成为深度学习的载体^[11],与前端的数据采集模块共同组成枪声识别系统。本文将 MobileNet 轻量化网络模型部署至 ZYNQ7020 的 ARM Cortex-A9 处理器中,将经过 PS 部分映射的 log-mel 谱图输入 MobileNet 神经网络中提取特征并获取枪声的类别信息。

MobileNet 网络采用深度可分离卷积模块替代原有的卷积层,极大减少了整体网络的参数数量,降低了网络对于硬件性能的要求^[12],网络整体结构如图 7 所示。

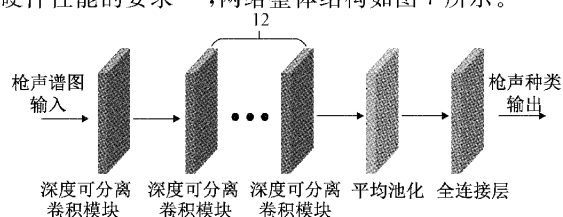


图 7 MobileNet 网络整体结构

深度可分离卷积^[13]模块将原先的卷积操作分为逐通道卷积 (depthwise convolution) 和逐点卷积 (pointwise convolution), 深度可分离卷积模块结构如图 8 所示。逐通道卷积的每个卷积核只负责一个通道, 因此这种卷积并不会增加特征图的通道数量。而逐点卷积将常规卷积的卷积核设置为 $1 \times 1 \times C$ (C 为通道数), 这种卷积运算将前一步的特征图在通道维度上进行加权组合。相比常规的卷积操作, 深度可分离卷积的计算量是常规卷积的约 1/3。

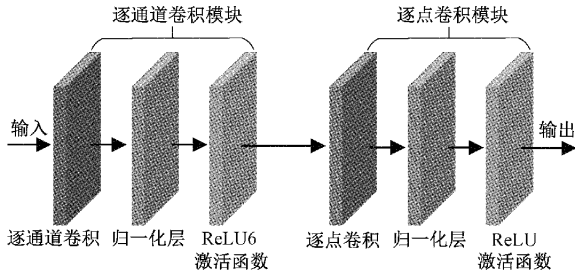


图 8 深度可分离卷积模块结构示意图

在本文的设计中, 逐通道卷积的卷积核的大小设置为 $3 \times 3 \times 1$, 而逐点卷积的卷积核设置为 $1 \times 1 \times N$ (N 为逐点卷积的通道数), 当卷积核移动步长设置为 1 时, 特征图经过逐通道卷积大小不变; 当卷积核移动步长设置为 2 时, 特征图每经过一次逐通道卷积, 其大小就会缩减至原先的 1/2。整体网络的具体参数如表 1 所示。

表 1 轻量化网络架构

卷积类别/移动步长	卷积核大小	输入维度大小
Conv/s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw/s1	$3 \times 3 \times 32dw$	$112 \times 112 \times 32$
Conv/s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw/s2	$3 \times 3 \times 64dw$	$112 \times 112 \times 64$
Conv/s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw/s1	$3 \times 3 \times 128dw$	$56 \times 56 \times 128$
Conv/s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw/s2	$3 \times 3 \times 128dw$	$56 \times 56 \times 128$
Conv/s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw/s1	$3 \times 3 \times 256dw$	$28 \times 28 \times 256$
Conv/s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw/s2	$3 \times 3 \times 256dw$	$28 \times 28 \times 256$
Conv/s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
$5 \times$ Conv dw/s1	$3 \times 3 \times 512dw$	$14 \times 14 \times 512$
$5 \times$ Conv/s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw/s2	$3 \times 3 \times 512dw$	$14 \times 14 \times 512$
Conv/s1	$1 \times 1 \times 512 \times 1\ 024$	$7 \times 7 \times 512$
Conv dw/s2	$3 \times 3 \times 1\ 024dw$	$7 \times 7 \times 1\ 024$
Conv/s1	$1 \times 1 \times 1\ 024 \times 1\ 024$	$7 \times 7 \times 1\ 024$
Conv dw/s1	Pool 7×7	$7 \times 7 \times 1\ 024$
FC/s1	$1\ 024 \times$ 类别数	$1 \times 1 \times 1\ 024$
Softmax/s1	—	$1 \times 1 \times$ 类别数

2) 神经网络训练

网络训练采用 NIJ Grant 2016-DN-BX-0183 项目枪声数据集^[14], 该数据集包含 6 000 个枪声样本, 18 种枪支型号, 选取每种类别的 80% 样本为训练集, 其余 20% 为验证集, 使用上述数据集在服务器端对轻量化神经网络进行训练。服务器端使用 2 个 1080Ti GPU, 结合并行计算架构 CUDA 对整个训练过程进行加速。训练完成后采用 5 折交叉验证^[15]的方式进行验证, 其混淆矩阵结果如图 9 所示。

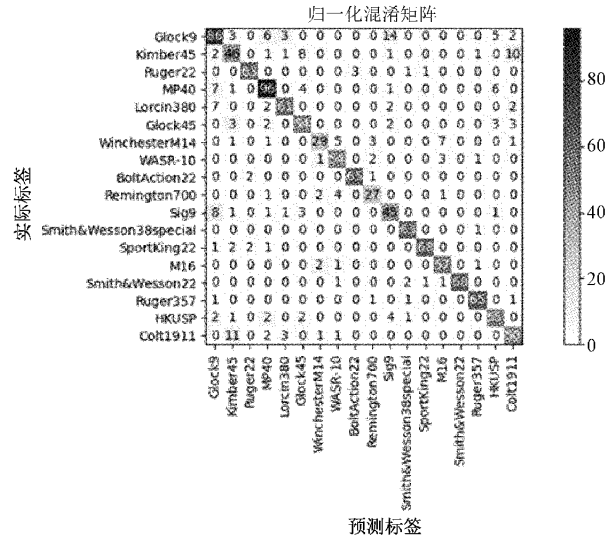


图 9 混淆矩阵结果

结果表明, 对于大部分枪声信号, 分类效果较好; 而在对 Kimber45 和 Colt1911 分类时, 结果容易混淆, 一方面是它们口径相似, 导致枪声信号相似, 另一方面, 这些手枪弹丸在飞行中没有产生马赫波, 缺少部分特征, 导致识别精度下降。

将服务器端训练所得的网络模型权重部署至 ZYNQ7020 的 PS 部分所搭载的 Linux 系统中, 实现在设备端对枪声信号的实时采集识别, 部署过程如图 10 所示。

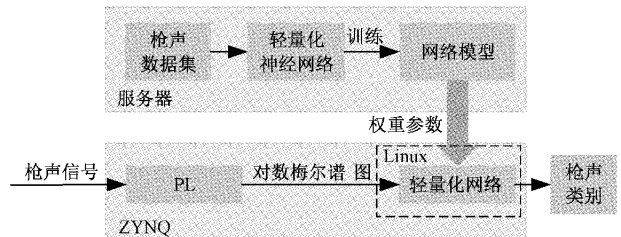


图 10 部署示意图

3 外场试验

为验证本方法的可行性, 搭建半实物仿真平台。以上述枪声数据集中的 M16、Remington700 和 Colt1911 三种枪声信号作为激励源, 将自研的八探头传感器阵列布置在声源周围预设的点位获取枪声信号, 最后使用枪声识别系

统对获取的枪声信号进行声源识别,识别完成后在显示器上输出枪声类别:

在设计现场布设方案时,需要综合考虑以下两个方面:首先,枪声信号的方向性极强^[16],因此需要考虑音响与传感器阵列的方向角;其次,传感器接收的音频信号的信噪比会受到两者间的距离影响,距离越近,信噪比越高。因此,在此次实验中,选取距离和角度各两种,共采用 4 种布置方案,即分别在与音响的距离和方向角为 10 m,0°;10 m,30°;20 m,0°;20 m,30° 的 4 个位置处对研制的枪声识别系统进行测试,试验现场如图 11 所示。



图 11 现场试验图

方案确定后,保持声源位置不动,模拟 3 种枪械射击各 5 次,在上述 4 个位置处布设传感器阵列采集信号,然后使用枪声识别系统识别枪声。

当使用 Remington700 音频信号作为激励时,系统采集的音频信号如图 12 所示,在该波形中 A 段为马赫波,B 段为膛口波。

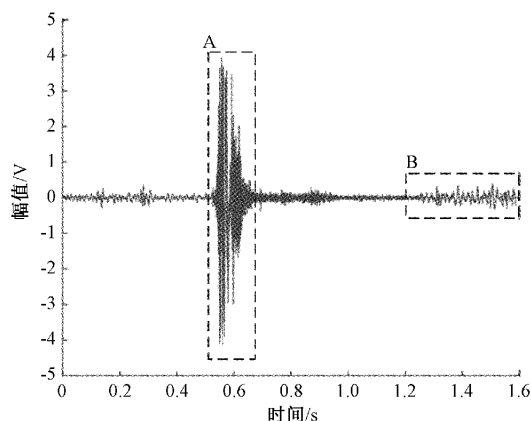


图 12 Remington700 枪声波形

试验完成后,最终的识别结果如表 2 所示。

本次试验共采集 60 次枪声,识别正确次数为 55 次,平均识别率 91.67%。对于距声源 10 m 处采集的信号,该系统的识别效果较好;而根据距声源 20 m 处的音频信号对枪声分类时,准确率下降,这是由于随着距离增加,音频信号的信噪比降低,导致识别精度下降。

试验结果表明:该枪声识别系统能够精确采集枪声信号,同时,枪声识别的识别精度较高,具备较高的应用价值。

表 2 枪声识别系统准确率表

枪声种类	位置	次数	正确次数	识别率
M16	10 m,0°	5	5	95%
	10 m,30°	5	5	
	20 m,0°	5	5	
	20 m,30°	5	4	
Remington700	10 m,0°	5	5	90%
	10 m,30°	5	5	
	20 m,0°	5	4	
	20 m,30°	5	4	
Colt1911	10 m,0°	5	5	90%
	10 m,30°	5	5	
	20 m,0°	5	4	
	20 m,30°	5	4	
平均识别率				91.67%

4 结 论

本文针对现有枪声识别系统缺乏实用性的问题,设计了基于 ZYNQ 的枪声识别系统,在 FPGA 端实现了枪声信号的 Log-Mel 特征提取,并在 ARM 端搭载 Linux 系统通过轻量化神经网络实现了枪声分类。相较于传统的枪声分类系统,本系统不依赖于强大的算力支撑,具备体积小,实用价值高的优点。试验表明,该系统实现了对于音频信号 Log-Mel 特征的提取以及对于枪声种类的识别,可以满足对枪声识别相关场景下的应用需求。

参考文献

- [1] 张克刚,叶湘滨. 基于短时能量和小波去噪的枪声信号检测方法[J]. 电测与仪表, 2015, 52(S1): 130-132, 138.
- [2] 李伟红,王伟冰,龚卫国. 低信噪比下公共场所异常声音声学特征提取[J]. 声学学报, 2019, 44(5): 934-944.
- [3] KOLUGURI, NITHIN R, MEENAKSHI, et al. Spectrogram enhancement using multiple window savitzky-golay MWSG filter for robust bird sound detection[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(6): 1183-1192.
- [4] 全璐,侯文,李秉臻,等. BP 神经网络的枪声识别[J]. 电声技术, 2018, 42(2): 13-16.
- [5] GALANGQUE C M J, GUIRNALDO S A. Gunshot classification and localization system using artificial neural network(ANN)[C]. 2019 12th International Conference on Information & Communication Technology and System (ICTS), 2019, DOI: 10.1109/ICTS.2019.8850937. x.

- [6] TARDIF B, LO D, GOUBRAN R. Gunshot sound measurement and analysis [C]. 2021 IEEE Sensors Applications Symposium (SAS)IEEE, 2021: 1-6.
- [7] ARSLAN Y. Impulsive sound detection by a novel energy formula and its usage for gunshot recognition[J]. ArXiv Preprint, 2017: ArXiv: 1706.08759.
- [8] NIMMY, RAJESH K R, NIMMY M, et al. Shock wave and muzzle blast identification techniques utilizing temporal and spectral aspects of gunshot signal[C]. 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2018, DOI: 10.1109/RAICS.2018.8635092.
- [9] 李冒金,李剑,刘宾,等. 基于 ZYNQ 的大动态冲击波超压测试系统设计[J]. 国外电子测量技术, 2022, 41(1):51-56.
- [10] GUO J, LI C, SUN Z, et al. A deep attention model for environmental sound classification from multi-feature data[J]. Applied Sciences, 2022, 12(12): 5988.
- [11] CORNA N, GARZETTI F, LUSARDI N, et al. Complete system-on-chip linux-based platform for measurement and generation of time domain signals[C]. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2019, DOI: 10.1109/NSS/MIC42101.2019.9059662.
- [12] 张凡,胡中秋,王能庆. 一种轻量化的高压线异物检测算法研究[J]. 电子测量技术, 2020, 43(23):97-103.
- [13] 万伟彤,李长峰,朱华波,等. 轻量化 CenterNet 网络的二维条码定位算法[J]. 电子测量与仪器学报, 2022, 36(5):128-135.
- [14] LI J, GUO J, SUN X, et al. A fast identification method of gunshot types based on knowledge distillation [J]. Applied Sciences, 2022, 12(11): 5526.
- [15] 徐桂芝,赵阳,郭苗苗,等. 基于深度分离卷积的情绪识别机器人即时交互研究[J]. 仪器仪表学报, 2019, 40(10):161-168.
- [16] BECK S D, NAKASONE H, MARR K W. Variations in recorded acoustic gunshot waveforms generated by small firearms [J]. The Journal of the Acoustical Society of America, 2011, 129(4): 1748-1759.

作者简介

孙袖山, 硕士研究生, 主要研究方向为 FPGA 数据采集与处理。

E-mail: S2005019@st.nuc.edu.cn

李剑, 教授, 硕士研究生导师, 主要研究方向为多维信号获取系统嵌入式开发、阵列化信号处理与重建。

E-mail: lijian@nuc.edu.cn