arXiv:2603.01069v1 [cs.AR] 1 Mar 2026

SHIELD8-UAV：面向低能耗无人机声学检测与时序跟踪的精度感知型一维特征驱动卷积神经网络（1D-F-CNN）顺序式 8 位硬件实现

Susmita Ghanta电气工程系 印度理工学院查谟分校Jagti 181221，印度
2024pvl0105@iitjammu.ac.in

Karan Nathwani电气工程系 印度理工学院查谟分校Jagti 181221，印度
karan.nathwani@iitjammu.ac.in

Rohit Chaurasiya电气工程系 印度理工学院查谟分校Jagti 181221，印度
rohit.chaurasiya@iitjammu.ac.in

摘要—边缘端的实时无人机（UAV）声学检测要求在严格的功耗和硬件限制下实现低延迟推理。本文提出 SHIELD8-UAV，这是一种面向连续声学监测的精度感知型一维特征驱动卷积神经网络（1D-F-CNN）加速器的顺序式 8 位硬件实现。该设计在共享多精度数据通路上逐层执行，无需复制处理单元。层敏感度量化框架支持 FP32、BF16、INT8 和 FXP8 模式；结构化通道剪枝则将展平特征维度从 35,072 降至 8,704（75%），从而减少串行全连接层周期数。模型在 FP32 模式下的检测准确率达到 89.91%，而在 8 位模式下的性能下降不到 2.5%。

该加速器在 Pynq-Z2 FPGA 上使用 2,268 个查找表，功耗为 0.94 W，端到端延迟为 116 ms；与 QuantMAC [1] 和 LPRE [2]，相比，延迟分别降低了 37.8% 和 49.6%，逻辑资源占用也比并行设计低 5–9%。采用 UMC 40 纳米工艺进行 ASIC 综合后，结果显示其最高工作频率为 1.56 GHz，核心面积为 3.29 mm^2，总功耗为 1.65 W。这些结果表明，顺序执行结合精度感知量化和面向串行化的剪枝，无需依赖大规模并行，即可实现实用的低能耗边缘推理。

索引词—无人机声学检测、一维卷积神经网络、边缘人工智能加速器、顺序执行、精度感知量化、剪枝、现场可编程门阵列、专用集成电路

I. 引言

边缘人工智能（AI）的迅速发展，加剧了对节能型神经网络加速器的需求；这类加速器须能在受限的功耗和硬件预算下进行实时推理。在新兴的边缘工作负载中，无人机（UAV）声学检测在监视、安防和自主监测应用中备受关注，因为这些应用要求设备在低能见度和非视距条件下可靠运行。传统卷积神经网络加速器主要依靠复制处理单元实现空间并行，以获得高吞吐量。尽管这类架构在数据中心环境中有效，却会带来显著的面积、内存带宽和功耗开销，因此不适用于边缘部署。此外，全连接层和高精度计算会进一步增加延迟与能耗，限制其在资源受限平台上的可扩展性。

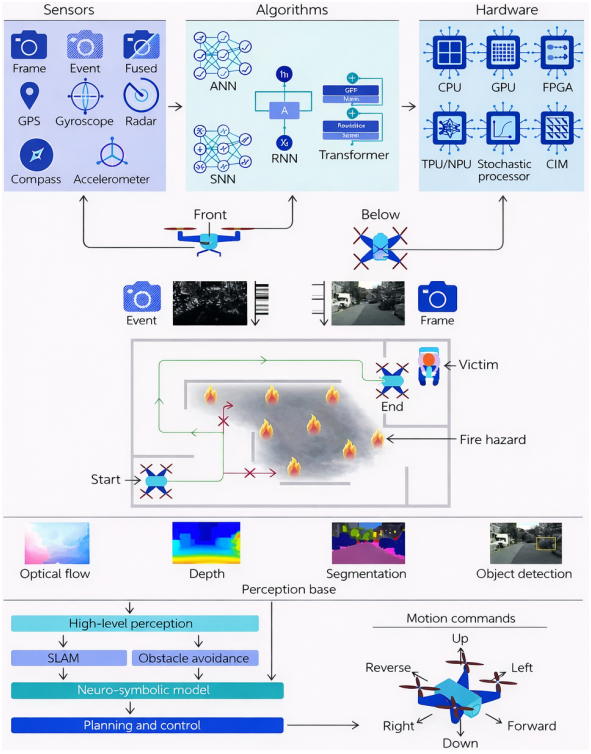


图1：面向边缘平台无人机声学检测与时序跟踪的算法—硬件协同设计栈概览。

近期研究探索了模型压缩和低精确率推理，以提升硬件效率；然而，许多现有设计仍依赖部分并行计算结构，缺乏统一的硬件—算法协同优化。具体而言，冗余的数据通路复制和非结构化计算模式会导致顺序层执行期间片上资源利用效率低下。图 1 展示了提出的协同设计架构，涵盖感知、特征提取、学习和硬件映射。声学信号被转换为紧凑的特征向量，由轻量级 1D-F-CNN 处理，并在采用共享数据通路复用的顺序多精度加速器上执行。该

图中凸显了算法设计与硬件执行之间的系统级集成，这构成了 SHIELD8-UAV 的核心贡献。

为解决这些局限，本文提出了一种面向边缘无人机声学检测的硬件感知型顺序式卷积神经网络加速器。该设计采用共享的可重构数据通路，以串行化方式执行卷积层和全连接层，无需复制硬件，同时保持功能灵活性。此外，精度感知量化策略支持多种数值格式下的自适应计算，并结合结构化剪枝来降低全连接层串行化开销。本文的主要贡献如下：

- 一种可复用的顺序式逐层执行卷积神经网络加速器，将卷积层和全连接层映射至共享计算结构，从而免去数据通路复制，并将现场可编程门阵列的逻辑资源占用降至 2,268 个查找表，约比具有代表性的并行 CNN 加速器小 5–9× 。
- 一种由层敏感度驱动的多精度推理框架，支持 FP32、BF16、INT8 和 FXP8 数值格式，检测准确率达到 89.91%，且在 8 位运算下的准确率下降不到 2.5%。
- 一种考虑串行化的结构化通道剪枝策略，可将展平特征维度从 35,072 降至 8,704（减少 75%），从而直接减少全连接层的执行周期和验证复杂度。
- 该算法–硬件协同设计已在现场可编程门阵列和 40纳米专用集成电路上得到验证，实现了 0.94 瓦功耗下 116 毫秒的实时推理延迟；与此前可复用的边缘加速器（包括 QuantMAC [1] 和 LPRE [2] ）相比，延迟分别降低了 37.8% 和 49.6%。

实验结果表明，所提出的协同设计框架能够以较低的计算复杂度和硬件利用率实现高效实时推理，因此非常适用于低功耗边缘人工智能系统。

II. 背景与动机

A. 无人机声学检测与特征驱动学习

无人机检测已采用射频、雷达、视觉和声学传感等模态进行探索。其中，声学传感具有传感器成本低、被动工作以及在非视距和低照度条件下鲁棒性强等优势，因此非常适合边缘部署。早期方法依赖手工设计的声学特征和浅层分类器，而近期方法则采用基于时频表示的深度学习模型，以提高鲁棒性。

小波散射变换与一维卷积神经网络架构相结合，在无人机声学分类中表现出优异性能 [3]，，大规模基于音频的卷积神经网络模型也进一步验证了学习型表征的有效性 [4]。更广泛地说，一维卷积神经网络适用于环境声音分析，因为它们能够捕捉时间依赖关系，同时保持紧凑的参数规模 [5]。然而，连续边缘监测对功耗、存储器，

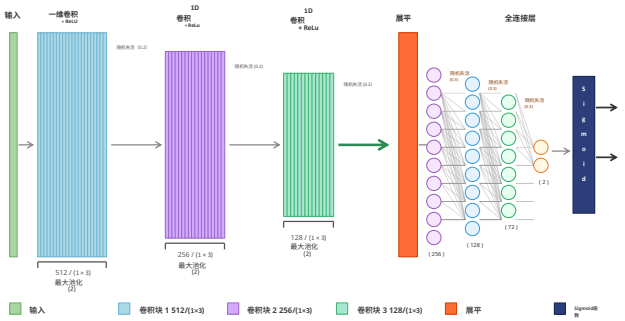


图 2：用于无人机声学检测的提出的 1D-F-CNN 架构。该网络处理紧凑的特征序列，并专为高效映射至顺序式硬件执行而设计。

和延迟提出了严格限制。在这类场景中，重复的特征图传输和全连接层串行化可能成为执行成本的主要部分，因此需要采用轻量级的特征驱动网络，在准确率与硬件效率之间取得平衡。

B. 顺序式精度感知边缘硬件

边缘推理工作负载主要由乘累加运算和存储器数据移动构成，因此需要采用针对计算效率和带宽利用率进行优化的架构。传统加速器通过复制处理单元实现空间并行，导致资源受限平台上的面积、布线和功耗开销较高。相比之下，顺序式层复用方法会将所有层调度到共享计算结构上，从而提高硬件利用率，同时避免数据通路复制。通过精度缩放还能进一步提升能效：降低操作数位宽可减少带宽需求和开关活动。已有研究表明，精度感知量化和多精度乘加单元能够在保持模型准确率的同时提升能效 [6]，，近期加速器的发展趋势则强调算法与硬件协同设计以及面向利用率的优化，而非追求峰值吞吐量 [7]。

尽管取得了这些进展，密集层输入序列化仍是顺序边缘加速器的一大瓶颈，会增加延迟和验证开销。因此，本研究致力于实现两个目标：（i）采用统一多精度计算数据通路，实现可复用顺序推理，并使激活延迟与MAC数据加载重叠；（ii）协同设计量化与剪枝，直接降低连续边缘监测工作负载的序列化开销和调度成本。

III. 提出的 SHIELD8-UAV 方法

SHIELD8-UAV采用算法—硬件协同设计框架，针对顺序式边缘推理对网络架构、量化方案和加速器设计进行联合优化。提出的系统集成了四个关键组件：(i) 紧凑的特征驱动型1D-F-CNN，(ii)

(iii) 逐层精度感知量化，(iii) 面向序列化感知优化的结构化剪枝，以及(iv) 可复用的多精度顺序式加速器架构。

A. 面向无人机检测的特征驱动型一维*F-CNN*

所提出的模型处理从短时长音频片段中提取的紧凑型声学特征向量，避免了计算开销较大的二维频谱图处理。这种设计可减少内存占用、特征图存储和数据移动，从而实现高效的硬件映射。

给定输入特征向量 $\mathbf{x} \in \mathbb{R}^{1\times M}$，每个卷积块都会依次执行一维卷积、激活函数、池化和正则化：

$$\mathbf{o} = \mathcal{D}_{0.2}\Big(\mathcal{M}_{1\times 2}\big(\sigma_R(\mathcal{C}_{1\times 3}(\mathbf{x}))\big)\Big), \tag{1}$$

其中，$\mathcal{C}_{1\times 3}$ 表示一维卷积，$\sigma_R(\cdot)$表示ReLU，$\mathcal{M}_{1\times 2}$ 表示最大池化，$\mathcal{D}_{0.2}$ 表示随机失活。三个卷积阶段可捕获旋翼谐波和周期性特征等时序声学模式，随后通过全连接层进行二分类。与二维卷积神经网络相比，一维架构在保留判别性时序表示 [3]的同时，显著减少了参数量和中间特征尺寸。

B. 结合层敏感性的精度感知量化

为平衡准确率与硬件效率，本文根据量化敏感性采用逐层精度分配策略。对于每一层 l，定义其敏感性分数为

$$s_{l,\mathrm{sc},k} = \frac{\Big(\big\|Q^{\mathrm{PwQ}}(\mathbf{w}_l) - \mathbf{w}_l\big\| - \big\|Q^{\mathrm{PwQ}}_{\mathrm{sc},k}(\mathbf{w}_l) - \mathbf{w}_l\big\|\Big)\cdot \|\nabla \mathcal{L}_{\mathbf{w}_l}\|}{n_l} \tag{2}$$

$$s_l = \max\left(s_{l,\mathrm{sc},16},\, s_{l,\mathrm{sc},8}\right), \tag{3}$$

其中，$\mathbf{w}_l$ 和 n_l 分别表示第 l 层的权重和大小。敏感度较高的层分配较高精度（FP32/BF16），而敏感度较低的层则采用 INT8 或 FXP8，以降低算术成本。

使用学习得到的裁剪边界 W_l和 W_h 对权重进行量化。对于全精度权重 W 和位宽 n，缩放因子为

$$\mathrm{scale}(k) = \mathrm{mean}(|W|)\cdot \frac{2^n-1}{2^{n-1}}, \tag{4}$$

整数化权重为

$$\widehat{W} = \mathrm{round}\left(\left(\mathrm{clip}\left(\frac{W}{k}, W_l, W_h\right) - W_l\right)\cdot \frac{2^n-1}{W_h - W_l}\right), \tag{5}$$

重构结果为

$$Q^{\mathrm{PwQ}}(W) = \widehat{W}\cdot \frac{W_h - W_l}{2^n-1} + W_l. \tag{6}$$

激活函数量化遵循PACT公式，并采用可学习裁剪参数 α:

$$y = \mathrm{PACT}(x) = 0.5\left(|x| - |x-\alpha| + \alpha\right), \tag{7}$$

表 I：全连接层特征缩减与硬件收益

指标	剪枝前 [3]	剪枝后
展平尺寸	35,072	8,704
尺寸缩减	75%	
全连接乘加运算次数	非常高	降低≈75%
串行化周期数	35,072	8,704
设计调试	High	减少
时序与功耗	具有挑战性	已改进

$$x^q = \mathrm{round}\left(y\cdot \frac{2^n-1}{\alpha}\right)\cdot \frac{\alpha}{2^n-1}. \tag{8}$$

该公式支持在FP32、BF16、INT8和FXP8模式下进行灵活的多精度推理，同时相较于 [1]保持检测准确率。

C. 面向序列化感知优化的结构化剪枝

在顺序加速器中，由于数据通过共享数据通路串行传输，展平到全连接层接口成为主要的延迟瓶颈。为解决这一问题，在展平之前应用结构化通道剪枝，将特征维度从 35,072 减少至 8,704。这直接减少了全连接层乘累加运算以及串行化硬件流水线中的执行周期。与传统的面向压缩的剪枝不同，此处的目标是进行调度感知优化，从而提升可复用加速器设计的延迟表现、仿真效率和验证可行性。表 I 详细讨论了相关影响。

D. 顺序式多精度加速器架构

提出的 POLARON 加速器采用可复用的共享计算数据通路，由输入与权重缓冲区、可配置乘累加器阵列、累加与激活函数单元以及基于有限状态机的控制引擎组成（图 3）。所有神经网络层均在同一计算架构上顺序执行，无需为每一层配置专用处理单元，从而减小面积并降低控制复杂度。特征图和权重从片上缓冲区流入乘累加器阵列，中间激活值则写回本地存储器，以供后续层复用。数据通路通过可编程对齐与缩放逻辑支持 FP32、BF16、INT8 和 FXP8 计算，同时扩展精度的累加器可在低精度运算期间保持数值稳定性。控制器在统一执行流水线中协调卷积、池化和全连接运算，以实现端到端网络推理。

与主机系统的通信通过 AXI 接口实现，其中 AXI-Lite 用于配置，AXI-DMA/Stream 用于数据传输。输入数据在计算前暂存于片上特征存储器中。逐层精度选择可根据工作负载需求调整数值格式，从而平衡能耗与准确率。累加后，结果依次经过归一化、缩放与平移以及溢出处理，然后进入基于 CORDIC 的激活单元，该单元支持 Swish、SoftMax、SeLU、GELU、Sigmoid函数、Tanh 和 ReLU。配置预取器解析层

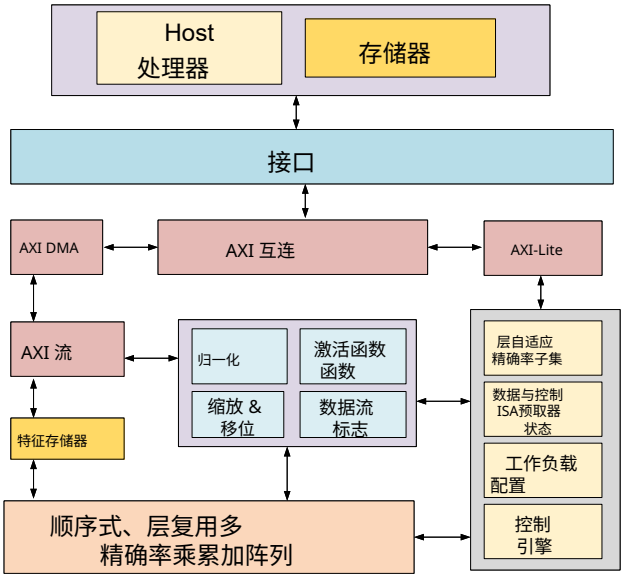


图 3：提出的精确率感知顺序式加速器架构，采用共享计算并支持多精度。

元数据，并在运行时更新执行参数。由此产生的数据流可最大限度地减少存储器流量和延迟，同时支持面向边缘人工智能工作负载的精度感知推理。

IV. 方法与实验设置

A. 数据集准备与增强

我们整理了一个由无人机和非无人机环境声音组成的异构声学数据集，以模拟真实部署条件。无人机录音采集自多个商用四旋翼平台，涵盖不同距离和朝向，以捕捉旋翼谐波、启动瞬态和飞行状态变化。背景样本包括可代表开阔场地和机场场景的环境声音及航空器相关音频。

为提高鲁棒性，我们在受控的信噪比范围内为录音添加了高斯噪声。每段音频流被切分为 0.8 秒的窗口，以平衡时序特征捕捉与推理延迟，随后进行幅度归一化，确保各样本的缩放一致。我们使用 librosa 进行特征提取，包括梅尔频率倒谱系数、池化梅尔频谱图系数、功率谱密度（PSD）和过零率（ZCR），从而表征周期性旋翼特征和宽带声学活动。此外，我们还纳入了公共音频库（例如 AudioSet 和 Pixabay），以提升背景多样性并减少环境特定偏差 [3]。

B. 训练与评估指标

数据集被划分为训练集、验证集和测试集，以进行无偏评估。模型训练采用 Adam优化器和交叉熵损失，并根据验证集表现进行早停。检测性能通过准确率、精确率、召回率和 F1分数进行报告。鉴于

表II：无人机检测模型的评估指标

模型	特征	准确率（%）	精确率（%）	召回率 (%)	F1 (%)
支持向量机 [8]	梅尔 (128)	87.88	88.41	87.44	87.92
	梅尔频率倒谱系数 (20)	84.39	85.62	83.17	84.37
	$\log_{10}$(功率谱密度)	83.10	80.23	88.43	84.13
一维残差卷积神经网络 [3]	原始音频	88.90	92.78	84.54	88.47
二维卷积神经网络 [9]	梅尔频率倒谱系数 (20)	88.96	88.85	89.32	89.09
K近邻算法 [8]	梅尔频率倒谱系数 (20)	88.73	87.87	90.21	89.02
集成 [10]	梅尔 (128)	86.50	86.88	85.55	86.21
一维多通道卷积神经网络 [11]	梅尔频率倒谱系数 (128)	83.26	86.11	85.52	85.81
基线 (FP32)	梅尔 (128)	89.13	91.64	86.48	88.99
	梅尔频率倒谱系数 (20)	89.91	89.00	91.39	90.18
	$\log_{10}$(功率谱密度)	87.87	88.88	86.92	87.89
	ZCR	60.64	58.61	77.58	66.77
提出的方法 (BF16)	梅尔 (128)	88.94	91.42	86.21	88.71
	梅尔频率倒谱系数 (20)	89.80	88.89	91.28	90.07
	$\log_{10}$(功率谱密度)	87.65	88.65	86.68	87.66
	ZCR	60.14	58.11	77.08	66.27
提出的方法 (INT8)	梅尔 (128)	88.02	90.31	85.06	87.73
	梅尔频率倒谱系数 (20)	89.14	88.23	90.62	89.41
	$\log_{10}$(功率谱密度)	86.55	87.56	85.60	86.57
	ZCR	57.34	55.31	74.28	63.47
提出的方法 (FXP8)	梅尔 (128)	87.65	89.88	84.72	87.02
	梅尔频率倒谱系数 (20)	88.97	88.06	90.45	89.24
	$\log_{10}$(功率谱密度)	86.02	87.03	85.07	86.04
	ZCR	55.40	53.37	72.34	61.53

连续监测系统的要求，还评估了虚警率和漏检率，以衡量实际可靠性。此外，通过测量不同信噪比水平下的推理准确率，进一步分析了鲁棒性，从而反映现实的噪声部署场景。

C. 算法-硬件协同设计评估

所提出的框架通过算法级仿真和硬件实现进行了验证。在 RTL 实现之前，我们开发了基于 Python 的算术仿真模型，用于分析采用 FP32、BF16、INT8 和 FXP8 精度格式时的推理行为。加速器采用 Verilog HDL 实现，并通过周期精确仿真进行验证。我们使用 AMD Vivado 工具链在 Pynq-Z2 平台上进行 FPGA 综合，同时使用 Cadence Genus，针对 UMC 40nm CMOS 工艺库并在统一的时序约束下进行 ASIC 综合。为确保公平比较，在必要时，我们采用归一化假设评估了具有代表性的基线加速器。硬件指标包括 LUT/DSP 利用率、工作频率、功耗和估算面积；系统级评估则考量了连续运行期间的端到端推理延迟和能耗。

V. 结果与讨论

A. 检测性能和精度影响

表II总结了所提出的1D-F-CNN与经典及深度学习基线模型的检测性能。使用FP32模式下的MFCC特征时，该模型达到了89.91%的准确率和90.18%的F1分数，表明其在采用紧凑特征表示的同时仍具备具有竞争力的性能。BF16推理的准确率与FP32几乎相同，说明其对尾数缩减不敏感。INT8和FXP8的性能有所下降，但仍适用于

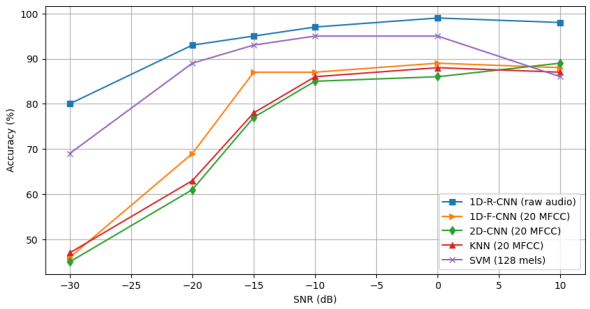
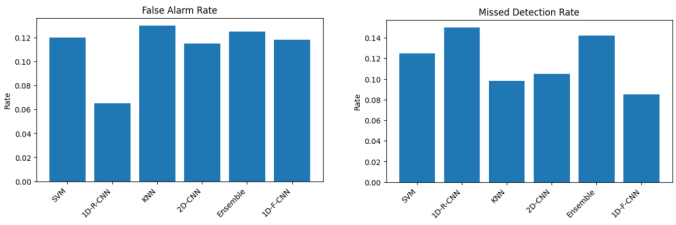


图 4：检测准确率与信噪比（SNR）的关系。



（a）虚警率　　（b）漏检率

图 5：噪声条件下的误差分析：（a）虚警率和（b）漏检率。当信噪比较低时，虚警率仍然较低；而背景噪声影响无人机声学特征时，漏检率会升高。

实际监测场景。这些结果验证了逐层精度感知量化策略能够在降低数值精度的情况下保持准确率。

图 4 展示了不同信噪比下的准确率变化趋势，图 5 展示了虚警和漏检情况。模型在中等噪声条件下保持稳定的性能，体现了其有效捕捉旋翼声学模式的能力。在大多数信噪比范围内，虚警率都较低；只有在信噪比极低时，由于谐波特征受到掩蔽，漏检率才会升高。总体而言，特征驱动型一维架构在持续的现实环境监测中表现出良好的鲁棒性。表 I 显示，结构化剪枝将展平特征尺寸从 35,072 降至 8,704（75%），从而按比例减少了顺序共享数据通路中的全连接层周期数。因此，剪枝不仅提高了模型紧凑性，还降低了推理延迟并提升了硬件验证效率。

B. 硬件利用率

表III将提出的架构与具有代表性的 FPGA 加速器设计风格进行了比较。全并行设计以大量消耗查找表、寄存器和数字信号处理器为代价，实现了高吞吐量。相比之下，提出的顺序架构采用统一计算数据通路，在保留功能灵活性的同时最大限度地减少资源重复。因此，与并行或部分可复用的加速器相比，该设计的硬件利用率更低。表IV进一步证实，通过计算复用和精度自适应，该设计在保持竞争力的性能的同时降低了资源开销。

表III：与具有代表性的架构基线相比的 FPGA 资源利用率

架构	LUTs	寄存器/触发器	块随机存取存储器/数字信号处理器	功率（瓦）
全并行 [13]	20790	30684	53	2.2
硬件复用 [1]	14428	15582	23	1.28
层复用 [14]	13956	16323	24	1.24
层多路复用 [15]	11265	11348	32	0.73
提出的方法 (SHIELD8-UAV)	2,268	3250	8	0.94

表 IV：与既有架构工作的 FPGA 定量比较

参数	Lu 等人 [16]	Aimar 等人 [17]	Mian 等人 [18]	RAMAN [19]	提出的方法
平台	Zynq-7100	VC707	ZCU102	Efinix Ti60	VC707
模型	CNN	CNN	一维卷积神经网络	深度可分离卷积神经网络	1D-F-CNN
查找表 (K)	22.9	23.9	39	37.2	2.2
寄存器/触发器（K）	10.7	20.1	27.8	8.6	3.25
功率（瓦）	1.1	2.2	1.54	0.15	0.94
频率（MHz）	60	170	200	75	100

40纳米工艺下的 ASIC 综合结果（表V）表明，该设计在频率、面积和功耗之间实现了均衡取舍。与优先追求峰值吞吐量的大规模并行加速器不同，提出的设计着眼于持续推理效率，适用于连续感知工作负载。共享的多精度数据通路降低了开关活动，并支持根据不同的能耗预算进行可扩展部署。

C. 系统级延迟与能耗

我们根据公式 9 和 10 建立了周期精确时序模型，用于分析可复用并行加速器架构的延迟。完整系统部署在 Pynq-Z2 平台上进行端到端评估。所提出的加速器在受限功耗（≤5W）下实现了 116 ms 的推理延迟，优于此前的可复用架构和边缘平台，包括 Flex-PE [12]（186.4 ms）、GR-ACMTr [13]（772 ms）、LPRE [2]（184 ms）、QuantMAC [1]（163.7 ms）、Jetson Nano（226 ms）和树莓派（555 ms）。

$$T_P = T_{\mathrm{MAC}} + T_{\mathrm{AF}},$$
$$T_R = T_{\mathrm{MAC}} + T_{\mathrm{Serial}} + K \cdot T_{\mathrm{AF}},$$

其中 T_{MAC}，　T_{AF}，　T_{PISO}分别是乘累加和串行激活函数的延迟。

(9)

$$Total_T_P = \sum_{l=1}^{L-1} n(l) + L - 1,$$
$$Total_T_R = \sum_{l=1}^{L} n(l) + 2L - 3,$$

(10)

其中 L 和 $n(l)$分别表示层数和该层中的乘累加次数。

与以往可复用加速器相比，提出的系统延迟更低，能效也具有竞争力。这一改进主要归功于减少全连接层的串行化、采用精度感知运算以及统一的计算复用，这些措施共同最大限度地减少了存储器传输和执行开销。图6总结了系统整体

表 V：采用 40纳米工艺的各类人工智能加速器综合后ASIC评估指标

设计	频率（GHz）	面积（mm^2）	功率（瓦）
JSSC'25 [20]	1.25	2.12	1.22
TVLSI'25 [21]	2.05	3.67	1.08
TVLSI'25 [12]	0.53	4.85	0.47
ISCAS'25 [14]	1.93	4.73	5.71
TCAS-I'22 [22]	1.46	10.80	1.02
TRETS'23 [13]	1.18	4.77	1.82
提出的方法	**1.56**	**3.29**	**1.65**

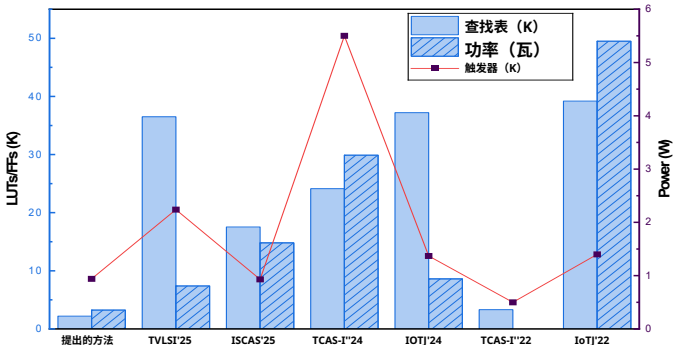


图6：系统级性能汇总。

性能，表明顺序执行与硬件感知协同设计相结合，可实现实用的实时边缘推理。

VI. 结论与未来工作

本文提出了 SHIELD8-UAV，这是一种面向资源受限边缘平台、用于连续无人机声学监测的精度感知顺序式一维 F-CNN 加速器。通过结合可复用的逐层执行、多精度计算和序列化感知优化，提出的方法在 8 位推理（INT8/FXP8）下实现了 89.91% 的检测准确率，准确率下降低于 2.5%。结构化通道剪枝将展平特征维度从 35,072 降至 8,704（减少 75%），从而直接降低全连接层序列化周期和验证开销。

硬件结果验证了实际边缘部署的可行性：FPGA 实现仅消耗 2,268 个查找表和 0.94 W 功耗，同时实现 116 ms 的端到端延迟，性能优于先前的加速器（QuantMAC、LPRE）和嵌入式平台（Jetson Nano、树莓派）。40纳米工艺下的 ASIC 综合达到 1.56 GHz，面积为 3.29mm^2，功耗为 1.65W，证实该设计能够扩展至 FPGA 原型验证之外。未来工作将探索运行时自适应精度控制，并将其扩展至多类声学场景识别。

参考文献

[1] N. Ashar, G. Raut 等，“QuantMAC：通过支持量化的乘累加单元提升深度神经网络的硬件性能”，IEEE Access，第 12 卷，第 43600–43614 页，2024 年。[2] O. Kokane, M. Lokhande, G. Raut, A. Teman 和 S. K. Vishvakarma，“LPRE：基于对数Posit的可重构边缘人工智能引擎”，载于 IEEE 国际电路与系统研讨会，第 1–5 页，2025 年。[3] M. Ali 和 K. Nathwani，“利用小波散射变换和一维卷积神经网络检测无人机”，IEEE Signal Processing Letters，第 31 卷，第 1790–1794 页，2024 年。[4] Y. Wang, Z. Chu, I. Ku, E. C. Smith 和 E. T. Matson，“大规模无人机音频数据集及基于音频和卷积神经网络的无人机分类”，载于 IEEE 国际机器人与计算会议，第 186–189 页，2022 年。[5] S. Abdoli, P. Cardinal 和 A. L. Koerich，“使用一维卷积神经网络进行端到端环境声音分类”，Expert Systems with Applications，第 136 卷，第 252–263 页，2019 年。[6] M. Lokhande, A. Jain 和 S. K. Vishvakarma，“精度感知的片上学习与自适应运行时可配置人工智能加速”，IEEE 国际超大规模集成电路设计与测试研讨会，2025 年 8 月。[7] T. Norrie, N. Patil 和 D. H. Yoon，“Google TPUv2 和 TPUv3 的设计流程”，IEEE Micro，第 41 卷，第 56–63 页，2022 年。[8] M. Anwar, Z. Kaleem 和 A. Jamalipour，“面向公共安全应用的机器学习启发式基于声音的业余无人机检测”，载于 IEEE Transactions on Vehicular Technology，第 68 卷，第 2526–2534 页，2019 年。[9] G. Raj, R. I. S. Mathew 和 L. R. Cenkeramaddi，“使用定制轻量级卷积神经网络进行空中飞行器分类：一种深度学习方法”，载于 2025 年 IEEE 电子、计算与通信技术国际会议（CONECCT），第 1–6 页，2025 年。[10] J. McCoy, A. Rawal, D. B. Rawat 和 B. M. Sadler，“用于可持续多模态无人机分类的集成深度学习”，IEEE Transactions on Intelligent Transportation Systems，第 24 卷，第 12 期，第 15425–15434 页，2023 年。[11] V. K. Puduru, B. Pardhasaradhi, S. Koorapati 和 L. R. Cenkeramaddi，“用于空中目标分类的轻量级深度学习模型”，载于 2025 年 IEEE 第 14 届通信系统与网络技术国际会议（CSNT），第 230–234 页，2025 年。[12] M. Lokhande, G. Raut 和 S. K. Vishvakarma，“Flex-PE：面向人工智能工作负载的灵活型单指令多数据多精度处理单元”，IEEE Transactions on VLSI Systems，第 33 卷，第 6 期，第 1610–1623 页，2025 年。[13] G. Raut, S. Karkun 和 S. K. Vishvakarma，“提升基于可扩展 CORDIC 的深度神经网络性能的实证方法”，ACM Transactions on Reconfigurable Technology and Systems，第 16 卷，2023 年 6 月。[14] O. Kokane, G. Raut, S. Ullah 等，“回顾：面向神经网络应用的基于 CORDIC 的可配置激活函数”，IEEE 计算机学会年度超大规模集成电路研讨会（ISVLSI），2025 年。[15] G. Raut, A. Biasizzo, N. Dhakad, N. Gupta, G. Papa 和 S. K. Vishvakarma，“面向深度神经网络加速器的数据复用与硬件重用架构”，Neurocomputing，第 486 卷，第 147–159 页，2022 年。[16] J. Lu, D. Liu, X. Cheng 等，“面向可穿戴心电分类设备的高效非结构化稀疏卷积神经网络加速器”，IEEE Transactions on Circuits and Systems I，第 69 卷，第 11 期，第 4572–4582 页，2022 年。[17] A. Aimar, H. Mostafa 等，“Nullhop：基于特征图稀疏表示的灵活卷积神经网络加速器”，IEEE Transactions on Neural Networks and Learning Systems，第 30 卷，第 3 期，第 644–656 页，2019 年。[18] F. A. Mian 和 S. Zafar，“基于片上系统的 1-D 卷积神经网络实现：通过 HLS4ML 进行三通道心电心律失常分类”，IEEE Embedded Systems Letters，第 16 卷，第 4 期，第 429–432 页，2024 年。[19] A. Krishna, S. Rohit Nudurupati 等，“RAMAN：用于边缘推理的可重构稀疏微型机器学习加速器”，IEEE Internet of Things Journal，第 11 卷，第 14 期，第 24831–24845 页，2024 年。[20] K. Li, M. Huang, A. Li 等，“采用 NAS 优化深度神经网络的 28 纳米互补金属氧化物半导体 29.12 TOPS/W 向量脉动阵列加速器”，IEEE Journal of Solid-State Circuits，第 60 卷，第 10 期，第 3790–3801 页，2025 年。[21] S. M. Cherati, M. Barzegar 和 L. Sousa，“基于最高有效数字优先的节能神经网络乘累加单元”，IEEE Transactions on Very Large Scale Integration Systems，第 1–12 页，2025 年 2 月。[22] R. Pilipovi´c、P. Buli´c 和 U. Lotri˘c，“一种两阶段操作数裁剪近似对数乘法器”，IEEE Transactions on Circuits and Systems I: Regular Papers，第 68 卷，第 2535–2545 页，2021 年 6 月。