
微电子所在低功耗人工智能领域获进展

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/20478.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

微电子所在低功耗人工智能领域获进展

。近日，中国科学院微电子研究所感知中心低功耗智能技术与微系统团队在低功耗人工智能领域取得新进展。语音唤醒技术（Keyword Spotting, KWS）是人工智能领域的重要技术，是人声与低功耗设备或终端之间的典型且广泛使用的“触发器”，广泛应用于各种低功耗的智能芯片与微系统。通常高性能的深度卷积神经网络模型的语音唤醒模型复杂度高、计算量大、需要占用大量内存，难以将其部署到上述资源有限的硬件设备。

针对上述问题，科研团队提出了用于语音唤醒的极轻量化、高准确率的改进二值残差神经网络B-ResNet（Binary Residual Neural Network，图1），利用二值量化方法，将神经网络中的全精度权重、激活参数量化为1bit（+1, -1），显著降低内存占用，亦可将网络中存在的大量浮点卷积乘加运算简化为XNOR同或逻辑与popcount运算，大幅降低计算复杂度。为解决二值网络带来的精度下降问题，该团队在B-ResNet网络前向传播中提出了具有移位初始化且可学习的激活函数来优化网络各层激活值分布，降低信息损失（图2），并在反向传播过程中提出了具有可变周期性窗口的梯度修正近似方法，有效解决了梯度失配与消失问题（图3）。在GSCD（Google Speech Commands Dataset）标准语音数据集的12分类任务下，与基线网络Res8-narrow相比，该技术可降低33%参数量与72%计算量，实现更高的语音唤醒精度，为后续该功能的低功耗硬件实现奠定了良好基础。

相关研究成果（*Low-complex and Highly-performed Binary Residual Neural Network for Small-footprint Keyword Spotting*

，DOI：10.21437/Interspeech.2022-573）被国际语音通信协会（ISCA）组织语音领域会议Interspeech2022接收。

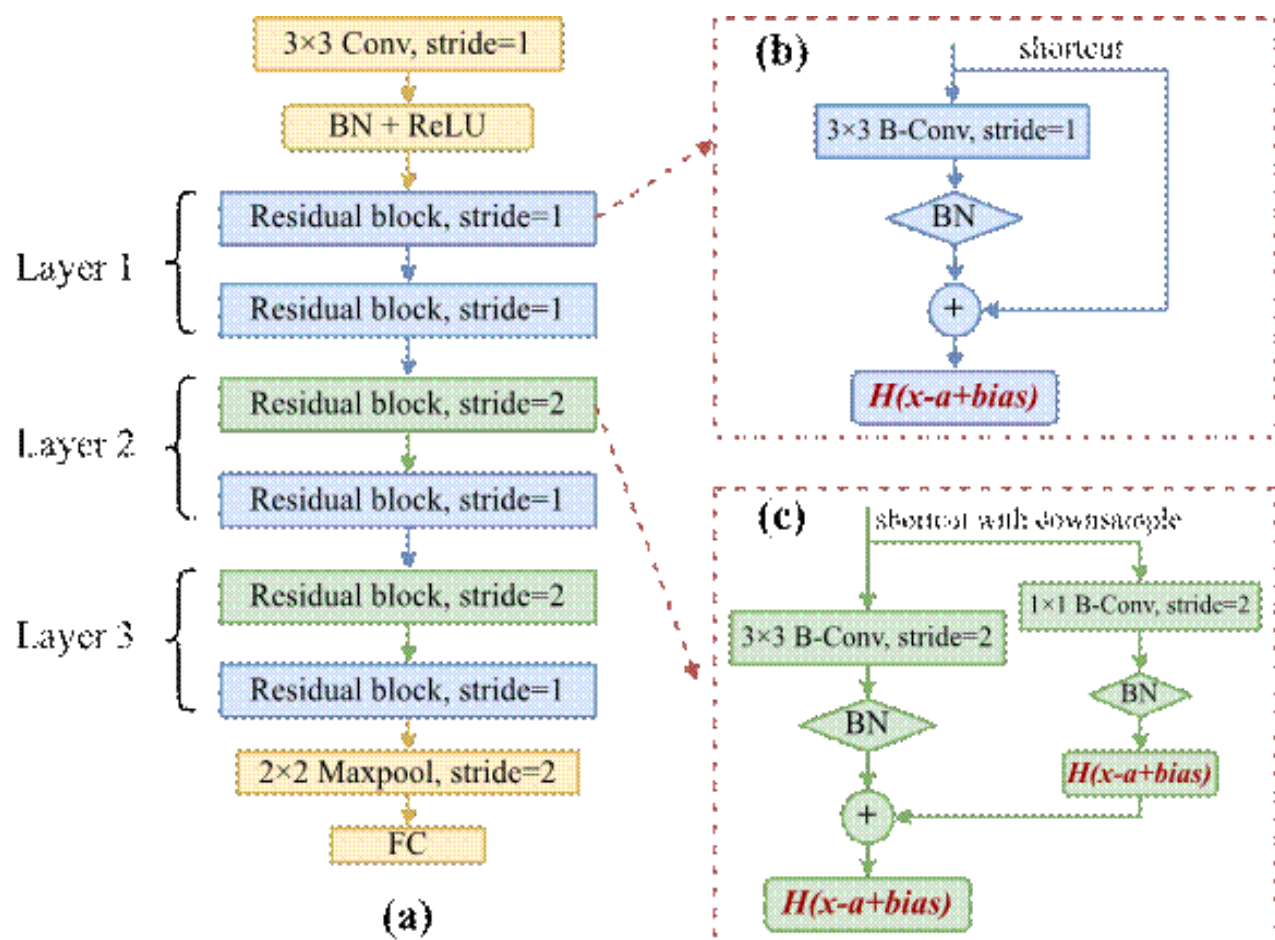


图1: B-ResNet的网络结构

$H(x-a+bias)$ 为提出的具有移位初始化并可学习的激活函数

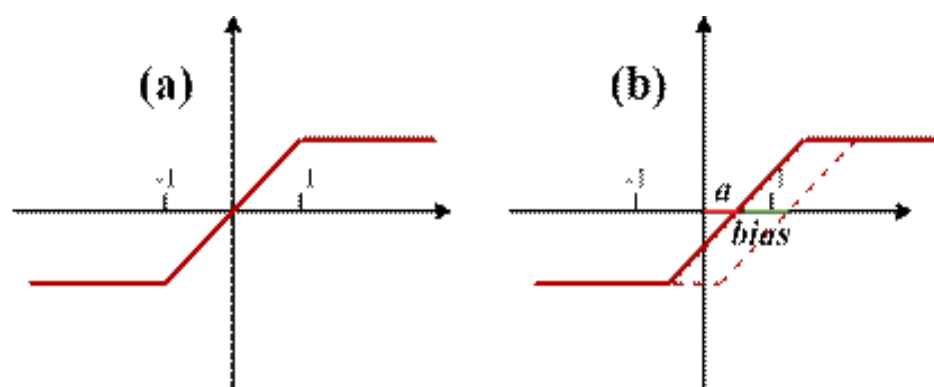


图2: 前向传播Hardtanh激活函数

(a)原始 (b)提出的移位初始化并可学习的激活函数 $H(x-a+bias)$

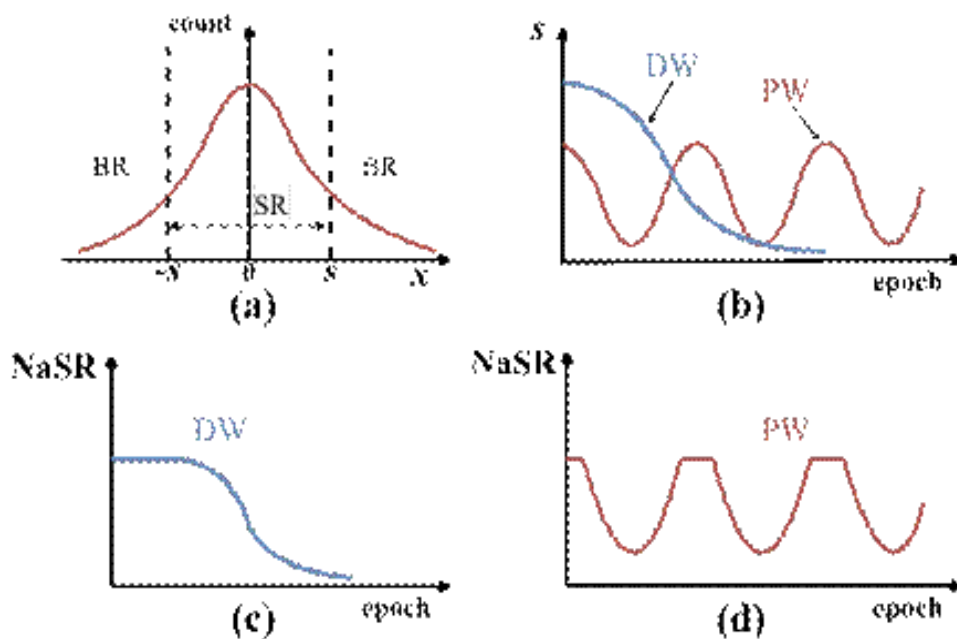


图3: 反向传播

提出的具有可变周期性窗口(PW)的梯度修正近似方法

研究团队单位：微电子研究所

更多科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发