

高效映射与灵活互联：基于多 FPGA 加速三维卷积神经网络 肺结节分割与分类

曹壮^{1,2}, 张甜^{1,2}, 何小威^{1,2}, 刘胜^{1,2}, 沈俊忠^{1,2}

¹国防科技大学计算机学院, 中国长沙市, 410073

²先进微处理器芯片与系统重点实验室, 中国长沙市, 410073

摘要：三维卷积神经网络（3D CNN）在肺结节检测领域具备极高的应用价值与发展潜力。然而，由于其计算复杂度高，内存开销大，在单块现场可编程门阵列（FPGA）上的硬件加速面临严峻挑战。为此，本文面向多 FPGA 平台提出一套高效映射策略，充分利用系统级大规模并行能力，最大限度提升计算效率。本系统通过集成 6 片定制化 FPGA 板卡，实现了业内领先的性能指标：肺结节分割算力达 15.9 TOPS，分类任务算力达 3.8 TOPS。与 CPU 基线相比，系统加速比高达 128.2 倍；与 GPU 实现相比，其能效比提升 6.7 倍。同时，在真实临床标准数据集上，该系统达到 87.1% 的召回率，整体性能居于国际先进水平。

关键词：肺结节检测；三维卷积神经网络；现场可编程门阵列（FPGA）；多 FPGA 系统；硬件加速；并行映射

<https://doi.org/10.1631/ENG.ITEE.2025.0186>