一种移动卷积神经网络的 FPGA 实现

李炳辰，黄　鲁

（中国科学技术大学 微电子学院，安徽 合肥 230026）

摘　要：卷积神经网络是深度学习的一种重要模型，广泛应用于图像处理等领域.常用的神经网络模型因结构复杂，参数众多，不适于放在移动端运行.本文基于模块化和硬件复用的思想，给出了一种基于 FPGA 的手写数字字符识别网络的硬件实现，基于 MobileNet 的原理改进结构，在实现了算法硬件加速的同时，有效地降低了网络的参数数量和整体运算量.基于 MNIST 数据集的实验结果表明，对比传统结构的神经网络，改进结构的参数量减少了 23.26%，计算量减少了 31.32%，在保持速度不变的前提下，用更少的资源和更低的功耗实现了整个网络.

关键词：FPGA；卷积神经网络；硬件加速；MobileNet；移动端

Hardware implementation of a convolutional neural network for mobile terminal based on FPGA

LI Bing-chen, HUANG Lu

(School of Microelectronics, University of Science and Technology of China, Hefei 230026, China)

Abstract：Convolutional neural networks are an important model of deep learning and are widely used in image processing and other fields. The commonly used neural network model is complex and has many parameters, which is not suitable for running on the mobile end. Based on the idea of modularization and hardware reuse, this paper presents a hardware implementation of handwritten digital character recognition network based on FPGA. Based on the principle of MobileNet, the structure is improved, and the algorithm hardware acceleration is realized, and the number of parameters of the network and the overall calculation amount are effectively reduced. The experimental results based on the MNIST dataset show that compared with the traditional neural network, the parameter size of the improved structure is reduced by 23.26%, and the calculation amount is reduced by 31.32%. The entire network is implemented with less resources and lower power consumption while maintaining the same speed.

Key words： FPGA; convolutional neural network; hardware acceleration; MobileNet; mobile terminal

作者简介：

李炳辰　男，（1995-），硕士.研究方向为 FPGA 算法加速.　E-mail:hnlibc@mail.ustc.edu.cn.

黄　鲁　男，（1961-），硕士，副教授.研究方向为数模混合高速接口集成电路设计.