doi:10.1631/FITEE.1700789

题目：深度卷积神经网络高效计算研究进展

概要：近年来迅速发展的深度神经网络已成为许多智能系统的基础工具。同时，深度网络的计算复杂度和资源消耗也在持续增加，这给深度网络的部署带来了严峻挑战，尤其在实时应用中或应用设备资源有限时。因此，网络加速是深度学习领域的热门话题。为提升深度神经网络的硬件性能，最近几年涌现出一大批基于现场可编程门阵列（field-programmable gate array, FPGA）或专用集成电路（application-specific integrated circuit, ASIC）的加速器。本文针对网络加速、压缩、软硬件结合的加速器设计等方面的进展进行了详细而全面的总结。特别地，本文对网络剪枝、低秩估计、网络量化、拟合网络、紧凑网络设计以及硬件加速器进行了深入分析。最后，展望了该领域未来一些研究方向。

关键词：深度神经网络；加速；压缩；硬件加速器