

瑞萨电子为AI芯片提供全新 计算存储一体化技术

瑞萨电子株式会社推出全新 AI 加速器，可高速且低功耗地执行 CNN（卷积神经网络）处理，向下一代瑞萨嵌入式人工智能（e-AI）迈进，加速端点设备智能化。采用该加速器的测试芯片可实现效率达到 8.8 TOPS/W（每秒每瓦万亿次操作），达到业界极高水平。瑞萨加速器基于计算存储一体化(PIM)架构，即当读取存储器数据时，在存储器电路中执行乘法和累加运算。

为构建全新 AI 加速器，瑞萨推出了以下三种技术。一是可执行大规模 CNN 计算的三进制 (-1,0,1) SRAM 结构 PIM 技术。二是与比较器配合使用的 SRAM 电路，可在低功耗下读取存储器数据。三是能够防止在制造过程中因工艺变化而导致的计算错误。将以上技术结合，既能缩短深度学习处理中的存储器访问时间，又可降低乘法和累加运算所需的功率。因此，当通过手写字符识别测试 (MNIST) 进行评估时，新加速器在保持 99% 以上准确率的同时，达到了业界最高能效等级。

下一代AI芯片技术的关键特性：

可根据所需精度调整计算位数的三进制 (-1,0,1) SRAM 结构 PIM 技术

三进制 (-1,0,1) SRAM 结构 PIM 技术采用三元存储器与简单数字计算块相组合，保证硬件数量增加的同时将计算误差降至最低。同时，允许根据所需精度进行位数切换（如 1.5 位（三进制）和 4 位计算之间）。可针对不同用户需求支持不同的精度及运算规模，用户可优化精度与功耗间的平衡。

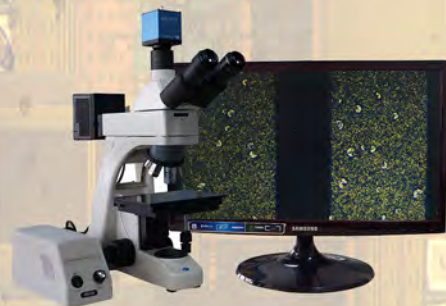
结合比较器和复制单元的高精度 / 低功耗存储器数据读取电路

当采用 PIM 架构时，通过检测 SRAM 结构中的位线电流值以读取存储器数据。尽管使用 A/D 转换器进行高精度位线电流检测是有效的，但这种方法功耗高且芯片面积较大。因此，瑞萨将比较器（1 位感测放大器）与复制单元相结合，方便灵活控制电流，从而开发出高精度的存储器数据读取电路。此外，由于神经网络操作所激活节点（神经元）数量非常少（约 1%），通过关闭未激活节点（神经元）读取电路以实现更低的运行功率。

消除制造过程中因工艺变化导致计算误差的规避技术

由于制造过程中的工艺变化，导致 SRAM 结构中位线电流值产生误差，从而造成存储器读取数据时出现错误。为解决这个问题，瑞萨在芯片内部覆盖了多个 SRAM 计算电路模块，由制造工艺变化最小的模块执行计算任务。由于激活节点只是所有节点中的一小部分，因此激活节点被有选择地分配给制造过程变化最小的 SRAM 计算电路模块执行计算。从而将计算误差降至几乎可忽略的水平。

现在，瑞萨正立足于此开发并推出新型尖端技术，致力于通过在物联网边缘及端点采用 AI 技术来增强智能，为实现智能社会做出贡献。

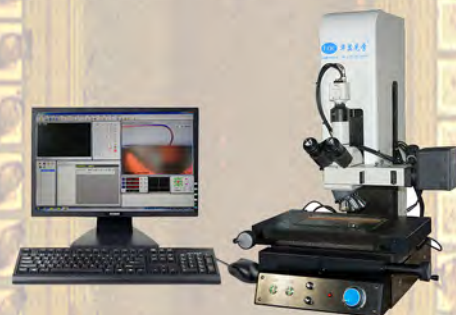


（支持15" 工作台

2000万有效像素

支持扫码追溯，链接MES系统）

工具金相显微镜系统



（邦定弧高测量

三轴精密量测晶圆智能检测）



深圳市华显光学仪器有限公司

0755-81753034

cy@china-eoc.com

光学解决方案，我们帮您做的更好！

http://www.china-eoc.com