

Trans-Precision Neural Network Deployment for Low-Power Embedded Systems

The methodology efficiently deploys and accelerates floating-point neural networks on embedded systems, optimizing performance, energy consumption, and hardware utilization.

Abstraction levels:

1. Model deployment

2. **System infrastructure**

3. Streamlined acceleration

4. Optimized processing

Industry
standard
framework

HW/SW co-design framework

