Overview

• Applying RISC-V core with custom instructions to mobile AI
  - Conv layers: >90% computation in CNNs

• Profiling RISC-V program with additional custom instructions

• Evaluating hardware performance for multiple configurations
Research Outcomes

**Platform: Andes N25, COPILOT**

- **(SI, SO):** input/output channel partition
  - (4, 4): 9.8x speed up, 32% more area;
  - (8, 8): 24.1x speed up, 153% more area

We thank Andes Technology and TSRI for technical support.

---

2D Convolution

<table>
<thead>
<tr>
<th>Cycles (x10^12)</th>
<th>Energy (mJ)</th>
<th>Area (x10^5 μm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI, SO=4,4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI, SO=8,8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Evaluated in TSMC 40nm technology