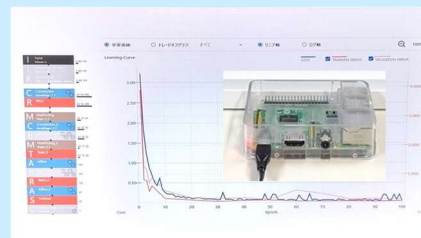


現場で使える低価格なAI実装モデルの構築

生産技術部



概要

安価なエッジのデバイス(以下、デバイス)で、AI処理を行うためのAI実装モデルの構築を行いました。処理能力の低いデバイスでAI処理を行うためのAI圧縮による軽量化や使用するデバイスの検証を行い、各条件でのAI処理時間、精度を比較評価しました。

[AI圧縮による軽量化]

量子化と枝刈りの圧縮手法を用いて、AI処理の軽量化及び計算効率の向上を試みました。

- ・量子化：AIの演算や重みの値を32bit浮動小数点から桁の小さな固定小数点に置換
- ・枝刈り：ニューラルネットワーク内で、重みの小さい接続や影響の少ないノードを削除

[AI実装のデバイス]

Raspberry Piを対象にAI実装を行いました。デバイス性能による比較のため、Raspberry Pi3B+とRaspberry Pi4Bへ実装を行いました(図1)。



	<table border="1"><tr><td>SoC</td><td>Broadcom BCM2837B0</td></tr><tr><td>CPU</td><td>ARM Cortex-A53 (ARM v7) 64-bit 1.4GHz</td></tr><tr><td>メモリ</td><td>1GB LPDDR2 SDRAM</td></tr></table>	SoC	Broadcom BCM2837B0	CPU	ARM Cortex-A53 (ARM v7) 64-bit 1.4GHz	メモリ	1GB LPDDR2 SDRAM		<table border="1"><tr><td>SoC</td><td>Broadcom BCM2711</td></tr><tr><td>CPU</td><td>Quad core Cortex-A72 (ARM v8) 64-bit 1.5GHz</td></tr><tr><td>メモリ</td><td>4GB LPDDR4 SDRAM</td></tr></table>	SoC	Broadcom BCM2711	CPU	Quad core Cortex-A72 (ARM v8) 64-bit 1.5GHz	メモリ	4GB LPDDR4 SDRAM
SoC	Broadcom BCM2837B0														
CPU	ARM Cortex-A53 (ARM v7) 64-bit 1.4GHz														
メモリ	1GB LPDDR2 SDRAM														
SoC	Broadcom BCM2711														
CPU	Quad core Cortex-A72 (ARM v8) 64-bit 1.5GHz														
メモリ	4GB LPDDR4 SDRAM														

図1 実装デバイス

[実装するAI]

実装するAIとして以下の2種のAIを用いました。

- ・実装AI①：CIFAR-10データセットによる10種の画像分類
- ・実装AI②：ねじ頭部の傷の有無を判別(図2)

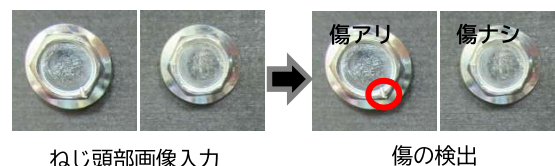


図2 実装AI②ねじ頭部の傷判別

[AI実装の比較評価]

軽量化の有無と実装デバイスによる比較評価を行いました(図3)。比較項目は、精度と処理時間。

- ・処理時間：[推論用AIの読み込み]→[テスト画像1枚の入力]→[結果出力]の一連の流れを完了するまでの時間
- ・精度：テスト画像入力時の正答率
(実装AI①のテスト画像1000枚)
(実装AI②のテスト画像100枚)

比較評価の結果、性能の高いデバイス及び軽量化手法を用いることで精度を落とさず、処理時間を短縮することができました。

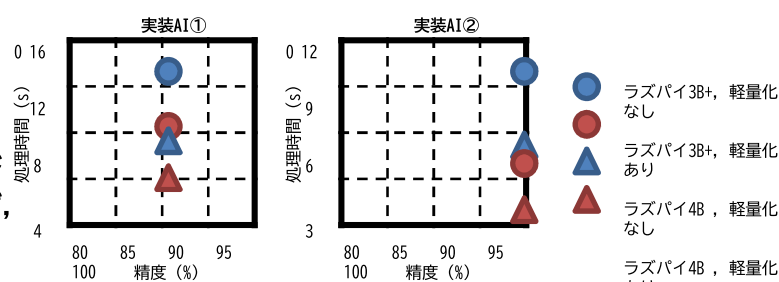


図3 AI実装の比較評価



いちおし

安価なデバイスを用いたAI実装を行うことができます。より、低コストでのAI活用を検討できます。



キーワード

エッジAI, AI実装, AI軽量化, RaspberryPi, 画像分類

