

GPUかCPUか？量産を見据えた最適なエッジAIデバイス選定と実践ポイント

AI搭載製品の開発が加速する一方で、「PoC（概念実証）後のエッジAIデバイス選定の知見がない」「GPUとCPUどちらを選ぶべきか判断できない」といった声が増えています。本ウェビナでは、ルネサス社 RZ/VシリーズやNVIDIA社Jetson Orinを例に、GPUとCPUの特徴をユースケース別に比較しました。アラヤが実際の組み込みAI開発で培ってきた知見をもとに、性能やモデル軽量化にまで踏み込み、量産を見据えた最適なエッジAI導入の考え方と実践ポイントを解説しました。

ハードウェア特性と選定の評価基準

CPU（Intel/ARM）、GPU（Jetson）、Renesas RZ（DRP-AI）の3種のエッジAIデバイスを、得意領域・性能・開発環境・電力・供給性の観点から比較しました。CPUは扱いやすく、軽量モデルや制御用途向きな一方、性能は限定的です。GPUは高スループットや多カメラ・高FPS処理に優れ、大規模モデルにも対応する一方、電力消費や放熱対策が必要です。Renesas RZは低消費電力かつVision特化で、周辺処理との組み合わせが容易です。用途・要求性能・消費電力を踏まえ、自社のケースに最適なデバイスを選定することが重要です。

モデル軽量化技術が広げるデバイス選択の可能性

倉庫内安全監視システム向けに、RZ/V2H と Jetson Orin AGX の処理性能と適合性を比較しました。その結果、Jetson は最適化後に50FPS超と大きな余力を持つ一方、RZ/V2H はDRP-AI TVM最適化で2.3FPSを達成し、本ユースケース要件（2FPS以上）を満たすことを確認しました。本システムの「歩行帯 逸脱検知」は高FPSを必要とせず、数秒以内に検知できれば十分なため、性能・電力・コストのバランスを考慮すると、RZ/V2Hが適した選択肢になると判断できます。一方、Jetsonは高FPSが必要となる用途で適した選択肢となります。（例：フォークリフト走行検知）

エッジAI導入を成功させるためのロードマップ

エッジAI導入における成功ポイントを体系的に整理しました。エッジAIは低レイテンシ、リアルタイム性、通信費削減、プライバシー保護といった多くの利点を有しています。一方で、最適なモデル選定やデバイス選定、運用時のモデル管理が課題となります。導入を成功させるには、精度と速度のトレードオフを理解し、多様なモデルとエッジデバイスの組み合わせを実機で評価することが重要です。また、MLCommonsベンチマークの活用やデバイス毎のコスト構造・消費電力を比較し、最適な構成を検討する必要があります。アラヤでは、モデル開発・最適化・デバイス実装まで、一気通貫で支援するエッジAI開発サービスを提供しており、お客様が効率的にエッジAIを導入できるよう支援しています。

■ 本日の登壇者 ■



株式会社アラヤ
蓮井 樹生 様

取締役
Chief Engineering Officer

GPU vs CPU：ハードウェア特性と選定の評価基準

エッジAIハード比較（弊社観点）

項目	CPU (Intel/ARM)	GPU (Jetson)	Renesas RZ (DRP-AI)
得意領域	軽量モデル、制御/前後処理が多い(ビデオイン/出力)	高スループット、多カメラ/高FPS、モデル肥大化に強い	低消費電力でVision特化、カメラ/FPや周辺処理も対応しやすい
実測例	YOLOv4-Tiny INT8 (TF Lite, RPIS) 28.2 FPS (R35 FPS相当)	Jetson Orin Nano 0 YOLOv8n 16 ms, YOLOv8m 50 ms	RZ/V2MA, (DRP-AI) 0 Tiny YOLOv3 22.8 ms, YOLOv3 143.5 ms (前後処理含む/CPU側前後処理は別)
開発環境	汎用 (ONNX Runtime/TF Lite等)	CUDA/TensorRT等	DRP-AI TVM/Translator等
電力/熱	小～中 (高負荷で上がる)	中～高 (冷却/放熱設計が必要)	「省電力で推論」方向に寄せやすい (SoC統合)
供給/運用	選定が多岐にわたる/やすい	世代更新や供給の影響を受けやすいこと	産機向けに長期供給設計のケースが多い (部材は豊富)
ポイント	軽量モデルの最適化/ CPU側は設計しやすい	高FPS/マルチストリーミング/軽量モデル/TensorRT等で大規模に検証できる	「モデル/演算別に合わせた最適化」の柔軟性

エッジAIハード比較（アラヤ観点）

出所：投影資料より一部抜粋

倉庫内安全監視システムを例にしたデバイス比較

倉庫内安全監視システムに適したハードウェア（検討結果サマリ・考察）

- 性能：Jetson Orin AGX は最適化後に50FPS超と大きな余力。一方 RZ/V2H は DRP-AI TVM最適化で2.3FPSを達成し、本ユースケース要件を満たす (Jetson = 余力大 / RZ = 必要十分)。
- 最適化：両者とも ONNXそのままでは性能不足。Jetson は TensorRT で大幅加速しやすいため、RZ/V2H は 制約に合わせて変換・最適化が前提で、モデル選定と最適化設計の影響が大きい。
- ユースケース適合：今回検討した「歩行帯 逸脱検知」は高FPS不要・検勢スケールで検知できれば良いため、性能×電力(コスト)のバランスでは RZ/V2H は有利。Jetsonの50FPS超はオーバースペックだったが、他の高FPSが求められるユースケースでは選択候補になる (例：フォークリフト走行検知 etc.)

検討結果サマリ

エッジAIデバイス	RZ/V2H (CPU)	Jetson Orin AGX (IaaS GPU)
ユースケース	作業者の歩行帯 逸脱検知 (要求性能：2.0 FPS以上)	
処理速度 (ONNX)	0.4FPS	2.7FPS
処理速度 (最適化)	2.3FPS (TVM/コンパイラ/INT8)	53.1FPS (TensorRT/FP16)
評価結果	最適化により検知可	検知可

倉庫内安全監視システムを例にしたデバイス比較

出所：投影資料より一部抜粋

[他記事、ウェビナ情報はこちら](#)



エンジニアによるマガジンサイト