



会社紹介資料

1. 会社紹介
2. 実現させたい未来
3. アラヤの強み
4. プロダクト概要のご紹介
5. 顧客例
6. エッジAIソリューションご紹介

 ARAYA

ation

Company Profile

認知神経科学の研究者が設立

脳技術を併せ持ったAI開発企業

社名 株式会社アラヤ

設立 2013年12月

所在地 東京都千代田区神田佐久間町1-11
産報佐久間ビル 6F

社員数 約80名

事業内容 AIプロダクト開発/
ニューロテックプロダクト開発事業



KANAI Ryota

金井 良太 (代表取締役)

Career

- 京都大学理学部卒業
- オランダ・ユトレヒト大学で実験心理学PhD取得
- 米国カリフォルニア工科大学にて、視覚経験と時間感覚の研究に従事
- 前英国サセックス大学准教授 (認知神経科学)

The Future We want to Realize

AI×ニューロテックで 人類の未来を圧倒的に面白く



独自のアルゴリズムで
高度な自動化を実現する



BMI*や人の脳状態センシング
で人類の能力を拡張する

*BMI: ブレイン・マシン・インターフェース

Product Introduction

人類の能力を飛躍的に拡張するアラヤのAI×ニューロテックソリューション

AIソリューション



画像認識AI

- 外観検査ソリューション
- カスタムAIソリューション

エッジAI

- モデル軽量化/エッジ実装支援
- エッジAIコンサルティング

自律AI

- 建機自動化/建設計画最適化
- 空調最適化

先端AI研究開発支援

- R&D支援
- 流体解析支援

ニューロテックソリューション



脳状態センシング

- 画像ベース脳波推定 (Face2Brain)
- 脳パフォーマンス推定 (Brain Performance Indicator)

BMI*

- BMI開発支援
- Araya Neuro Interface提供

神経科学を活用した研究開発支援

- 商品開発支援
- ニューロマーケティング

5. Some of our customers (自社サイト情報再掲)

大手製造/建設/物流業などの幅広い業界のソリューションを提供しています。



エッジAIソリューション ご紹介



エッジAIについて

クラウドコンピューティングとエッジコンピューティング

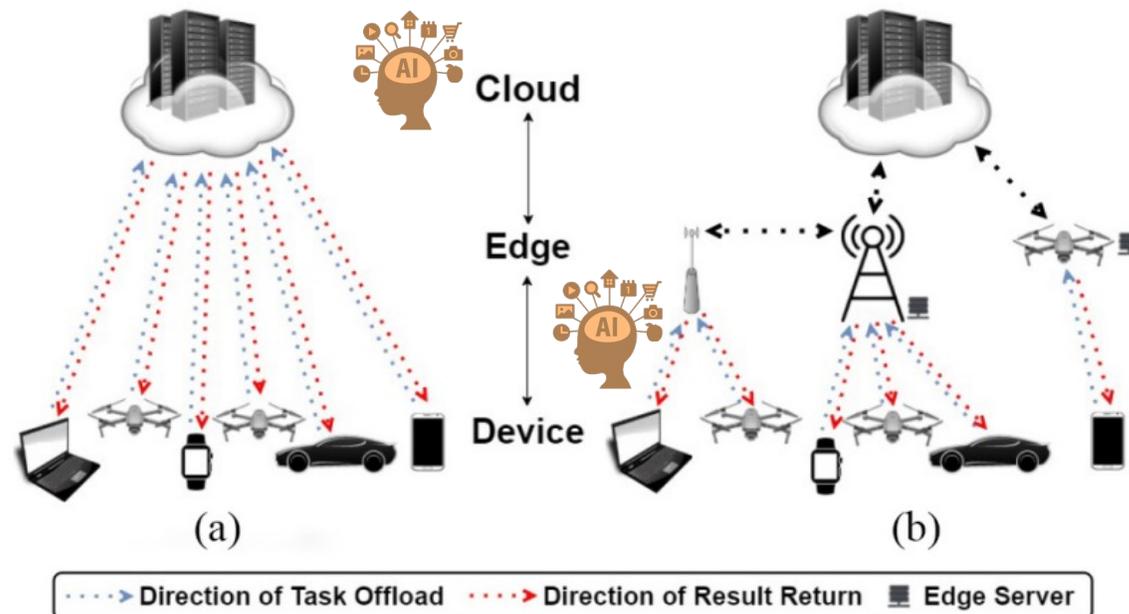
エッジAIの元となるエッジコンピューティングのアーキテクチャでは、処理するアプリケーションの中で、処理遅延が許されない、かつ、計算負荷の高い部分をエッジデバイスで実行し、エッジサーバー（の一部の）アプリケーションはエッジデバイスで処理するデータの同期のためにコアクラウドと通信する役割を担います。

アプリケーション例)

- 自動運転車同士のリアルタイムコミュニケーションにおけるend-to-endでの応答遅延10ms以下

この場合、クラウドコンピューティングでは、クラウドへend-to-endのアクセス遅延が80ms以上となり、要求を満たせないため、エッジコンピューティングを用いたアーキテクチャ検討が必要となります。

このエッジコンピューティング処理をAIとして行うものがエッジAIと言えます。



クラウド
コンピューティング/AI

エッジ
コンピューティング/AI

“A Survey on the Convergence of Edge Computing and AI for UAVs: Opportunities and Challenges”, Patrick McEnroe, 2022
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=9778241>

エッジAIのメリット・デメリット

クラウドAIとの対比から見たエッジAIのメリット・デメリット

メリット

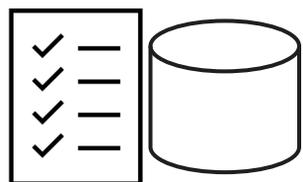
項目	クラウドAI	エッジAI
処理性能とストレージ量	大規模ストレージと強力な処理装置を備えたクラウドデータセンター	エッジにおける容量/重量/空間の制約による限定的な処理装置・ストレージ
セキュリティプライバシー	(潜在的な攻撃対象はそれほど大きくはないが) データが分散されていないため、DDoS等のの一度の攻撃で大きな混乱が生じる可能性がある	データが分散されるほどリスクは分散され攻撃が成功したときのインパクトは小さくなる。データの送信距離も短くなり傍受される可能性が低くなりデータを遠隔地の集中型クラウドサーバーに保存する必要もなくなる。その場でデータ処理して匿名化したデータを送信可能。
レイテンシー	エッジ端末とクラウドサーバ間のデータ送受信にともなう通信遅延が大きくなる可能性あり	遠隔地のリソースに依存する必要がなく、より短い距離でより少ないデータを送信可能
信頼性	中央のクラウドサーバーにダウンしてアクセスできない場合悲惨な結果になる可能性がある	複数のエッジサーバーはユーザーの近くに設置されていれば、1台がダウンしても別のサーバーがサービスを提供可能。エンドデバイスはローカルに保存されたデータで多くのリクエストを自分で処理可能
通信オーバーヘッド	大容量データを長距離送信する必要あり	デバイス上で前処理することでサーバー送信データ量を減らし、(遠隔地のクラウドサーバーではない) エッジサーバーに送ることによってデータ転送距離を短縮可能で、エネルギー消費量とコストを削減可能。

エッジAIソリューションご紹介

エッジ実装を想定したAI開発プロセス

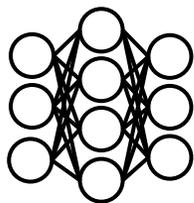
AIモデル開発から、エッジデバイス選定・モデル最適化・実装評価、システム実装まで、ご要望に応じた開発サービスをご提供

仕様/データセット



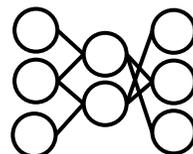
AIモデル
開発

AIモデル



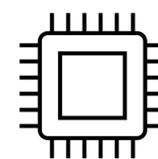
モデル
最適化

最適化モデル



エッジ
実装

エッジデバイス



システム
実装

製品/システム/アプリ



AIモデル選定
エッジデバイス
選定

データ収集
アノテーション

モデル学習
評価・改善

エッジ向け
モデル最適化

評価・改善

エッジデバイス
実装

評価・改善

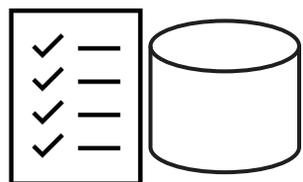
システム実装
評価・改善

エッジAIソリューションご紹介

エッジ実装を想定したAI開発プロセス

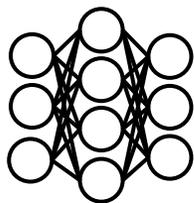
エッジデバイス毎の開発ツールと、汎用的な最適化手法とを適切に組み合わせてご要望のエッジAI性能実現を目指します。

仕様/データセット



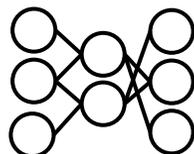
AIモデル
開発

AIモデル



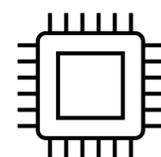
モデル
最適化

最適化モデル



エッジ
実装

エッジデバイス



システム
実装

製品/システム/アプリ



最適化手法

汎用
最適化

- 量子化
- 枝刈り(OFA)
- 蒸留

ツール
最適化

- TFLite
- TensorRT
- OpenVINO
- CoreML
- 各社NN用SDK

エッジデバイス

GPU

- Jetson ORIN (NVIDIA)
- Raspberry Pi4
- Edge TPU (Coral/Google)
- Myriad X (NCS2)

CPU/
Accel

スマホ

- iPhone(Neural Engine)
- Snapdragon(Qualcomm)

FPGA/
その他

- Kria K26(Xilinx FPGA)
- etc...

エッジAIソリューションご紹介

エッジデバイス対応実績例①

最もメジャーなNVIDIA社のJetsonシリーズの他、USBアクセラレータタイプ、FPGAの対応実績がございます

組み込み向けGPU

Jetson Nano



Jetson TX2



Jetson Xavier NX



Jetson AGX Xavier



<https://www.nvidia.com/ja-jp/autonomous-machines/embedded-systems/>

アクセラレータ

Intel NSC (Myriad2)



<https://ark.intel.com/content/www/jp/ja/ark/products/140109/intel-neural-compute-stick-2.html>

Coral (Edge TPU)



<https://coral.ai/products/>

FPGA (Xilinx, Intel)



<https://www.xilinx.com/products/app-store/kria.html>



<https://www.intel.co.jp/content/www/jp/ja/programmable/documentation/hhf1507759304946.html>

Jetson関連情報

組み込みGPUタイプ（NVIDIA Jetsonシリーズ）

NVIDIA社のAI推論用GPUアーキテクチャ（Tensorコア）を組み込み製品向けに最適化したデバイス。



Jetson	Nano	TX2	Xavier NX	AGX Xavier	Orin Series Nano/NX/AGX
発売日	2019年	2017年	2021年	2018年	2022年
GPU世代	Maxell	Pascal	Turing	Turing	Volta
TFLOPs	0.472	1.33※1	21	32※2	20(Nano)~275(AGX)
Cudaコア数	128	256	384	512	512(Nano)~2048(AGX)
Tensorコア数	—	—	48	64	16(Nano)~64(AGX)

※1：TX2iは1.26T、※2：Industrialは30T

Jetson関連情報

組み込みGPUタイプ (NVIDIA Jetsonシリーズ)

Turing世代のGPUアーキテクチャを備えるJetsonから、モデル最適化機能に演算精度 (量子化精度) を、浮動小数点 (FP32、FP16) の加え、整数8bit (INT8) での最適化に対応。

演算精度

- FP32
- FP16



Jetson Nano



Jetson TX2

演算精度

- FP32
- FP16
- INT8



Jetson NX



Jetson AGX Xavier

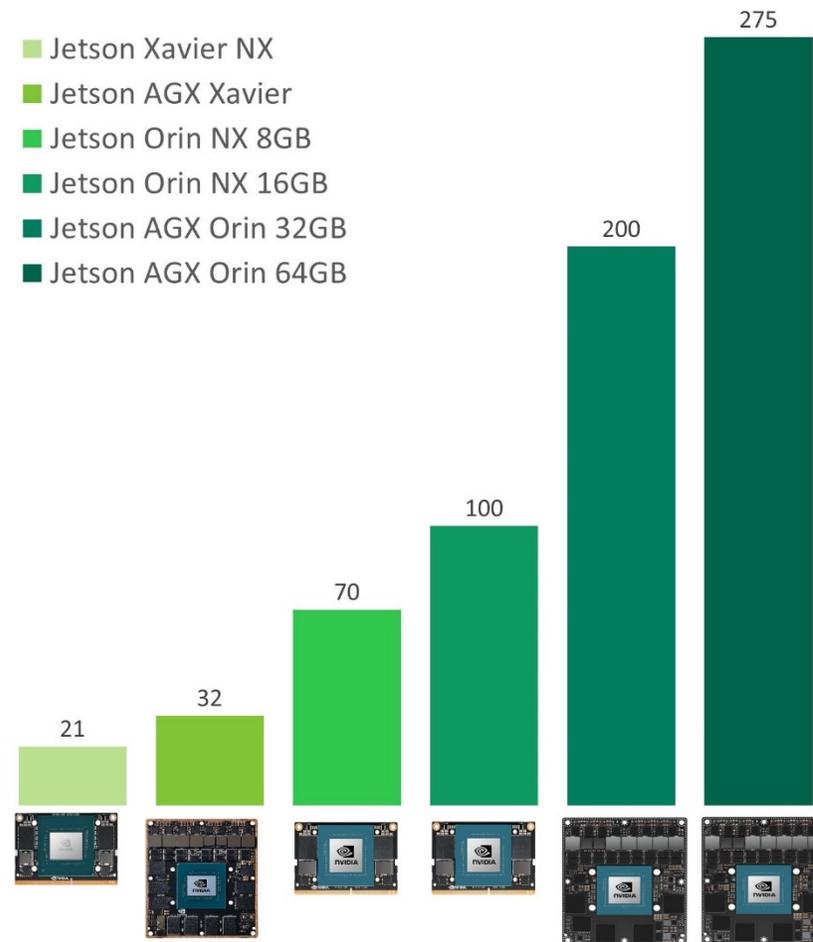


Jetson Orin Series

AI PERFORMANCE

- Jetson Xavier NX
- Jetson AGX Xavier
- Jetson Orin NX 8GB
- Jetson Orin NX 16GB
- Jetson AGX Orin 32GB
- Jetson AGX Orin 64GB

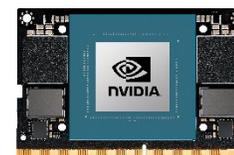
INT8 TOPs



<https://developer.nvidia.com/blog/delivering-server-class-performance-at-the-edge-with-nvidia-jetson-orin/>

Jetson関連情報

NVIDIA Jetson ベンチマーク結果 (Xavier NX v.s. Orin Nano, YOLOX※1)



Jetson	Xavier NX	Orin Nano
発売日/GPU世代	2021年/Turing	2022年/Volta
TFLOPs/ Cuda/Tensorコア数	21/ 384/48	20/ 512/16
YOLOX-X (CPU/GPU※2/TensorRT※3)	8.6/0.35/NA [s]	4.6/0.21/0.067 [s]
YOLOX-M (CPU/GPU/TensorRT)	2.7/0.11/0.036 [s]	1.6/0.074/0.026 [s]
YOLOX-S (CPU/GPU/TensorRT)	1.2/0.058/0.036 [s]	0.78/0.043/0.021 [s]

※1: <https://github.com/Megvii-BaseDetection/YOLOX>

※2: FP32

※3: FP16

エッジAIソリューションご紹介

エッジデバイス対応実績例②

近年注目を集めているAIチップ系のデバイス、スマホ・タブレットやマイコンでの実装経験もございます

AI/ML 専用チップ (AI Processor : AIチップ)

- Ambarella社車載用チップ
- Hailo社製車載用チップ
- etc ...



<https://hailo.ai/ja/product-hailo/hailo-8/>



<https://www.ambarella.com/products/automotive/>

カメラ・センサー体で様々な用途に 利用可能な汎用デバイス

- iOSデバイス (iPhone/iPad)
- Androidデバイス (スマホ)
- etc ...



<https://www.apple.com/jp>

小規模・特化したDL/ML処理を サイズ・リソース最小限に実行

- 各種Raspberry Pi
(+USB外付けアクセラレータ)
- ARM Cortex-m#系

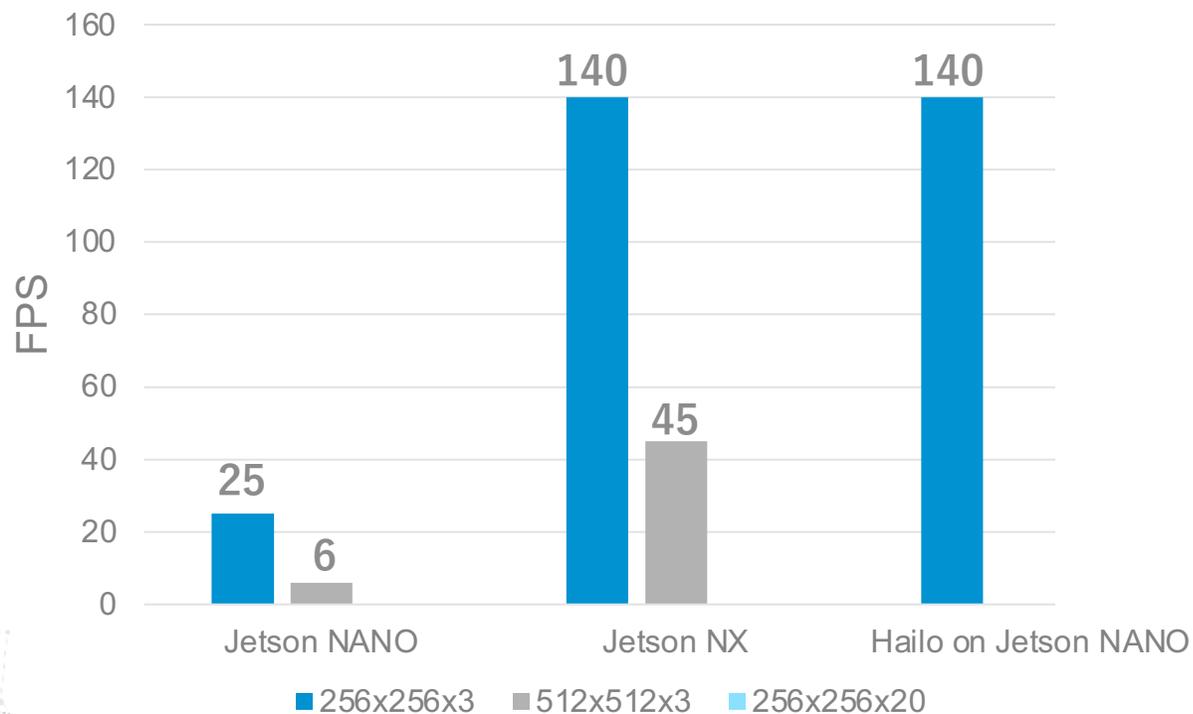


<https://www.raspberrypi.com/products/raspberry-pi-3-model-b/>

Jetson関連情報

AI専用チップタイプ

- Hailo社製AIチップ（Hailo-8 : 26TOPS）と Jetson（Nano/NX） との速度比較（FPS）
- 使用モデル（最適化）：DeepLabv3（Hailo-8 : INT8、Jetson : fp16）
- Hailo-8 は m.2モジュールで提供。Jetsonをマザーボードとして使用※
- Hailo-8 と Jetson NX とで同等性能を確認



※：マザーボードの帯域制約の考慮が必要



Hailo-8 on Jetson NANO
(m.2インタフェース接続)

エッジAIソリューションご紹介

エッジAI-BOXタイプの対応実績例

エッジデバイスを内蔵した完成品（エッジAI-BOX）の活用実績もあり、短TATなエッジAI開発に効果的です

組み込みGPUタイプ① Jetson Xavier

- AAEON社製
- 物体検出+トラッキング



<https://www.aaeon.com/en/p/nvidia-xavier-nx-embedded-box-pc-boxer-8251ai>

組み込みGPUタイプ② Jetson TX2

- WIBASE社製
- ポーズ推定



https://www.wibase.com/products_detail/en/ai/travis-v/

AIチップタイプ Hailo-8

- AnalogTech社製
- 物体検出



<https://www.analogtech.co.jp/products/AT-Hailo8-Multi/>

エッジAIソリューションご紹介

エッジAI開発事例紹介（抜粋）

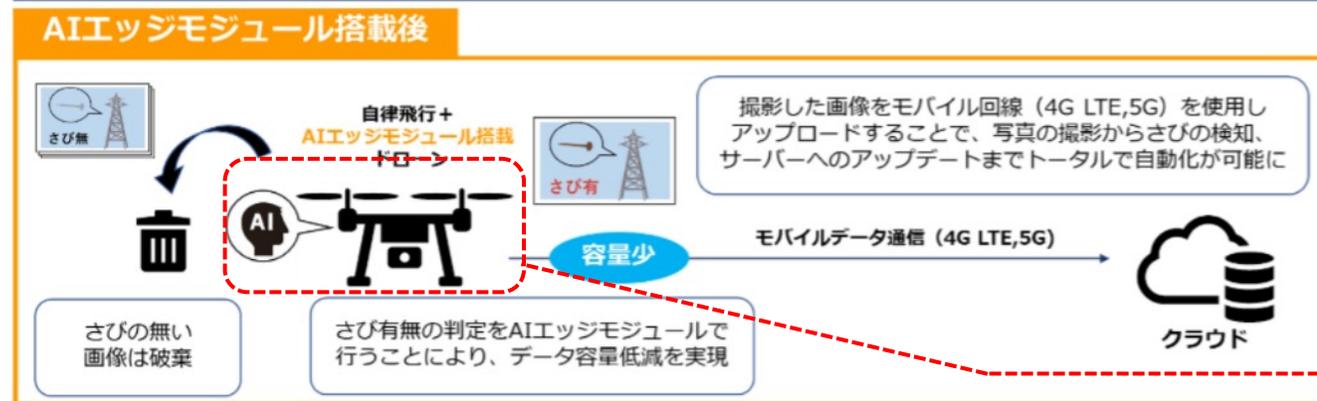
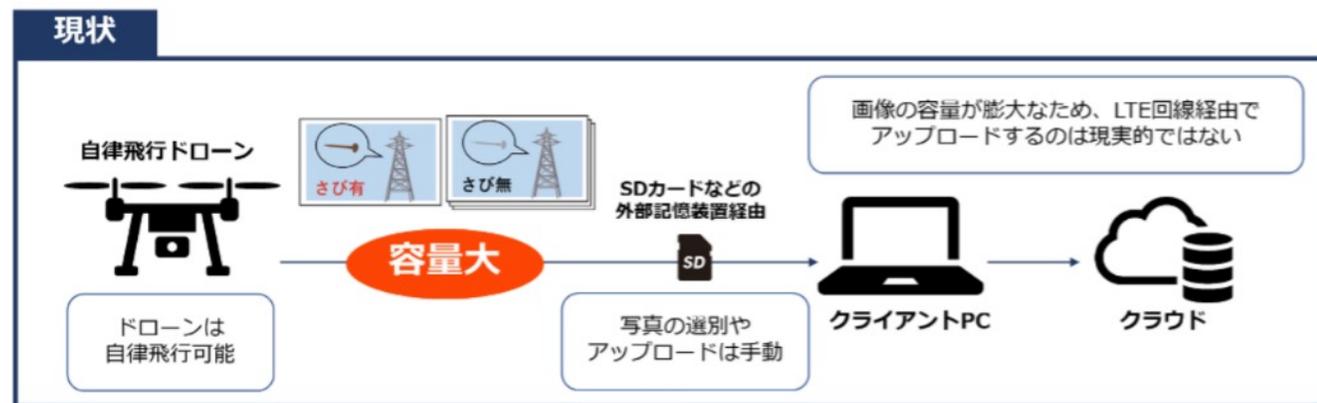
	事例	エッジ化手法	デバイス
エッジAI 対応	プルーニングと量子化による異常検知モデルの軽量化	量子化 枝刈	Jetson系
	中間層出力を用いた蒸留による異常検知モデルの軽量化	蒸留	Jetson系
	darknet用モデルのTensorRT化による棚卸システムの高速度化	蒸留	Jetson系
モデル開発 + エッジAI 対応	one-shot NASを用いた高効率な路上障害物検知モデルの開発	NAS 枝刈	AIチップ系
	FPGA上で動作する各種エッジAIモデル（分類/物体検出/セグメンテーション）の開発	量子化	FPGA
	荷物棚の充填率（積載率）推定処理のエッジ実装	量子化	Accelerator
	iOS上で動作する姿勢推定モデルの開発	蒸留	iPad系
	NEDO AIエッジコンピューティング 高効率・高速処理を可能とするAIチップ・次世代コンピューティングの技術開発	量子化 枝刈	FPGA/GPU/ AIチップ

エッジAIソリューションご紹介

エッジAI開発事例紹介 (NEDOプロジェクト)

ドローンフライトによるエッジAI実証※

- エッジAIを用いて、フライト撮影動画からサビ画像のみを検知してクラウド送信することでデータ通信量を削減



フライト実証時の推論動画(サビ検知箇所:赤)



使用したAIエッジモジュール



Jetson Nano (GPU系)



Hailo-8 (AIチップ系)

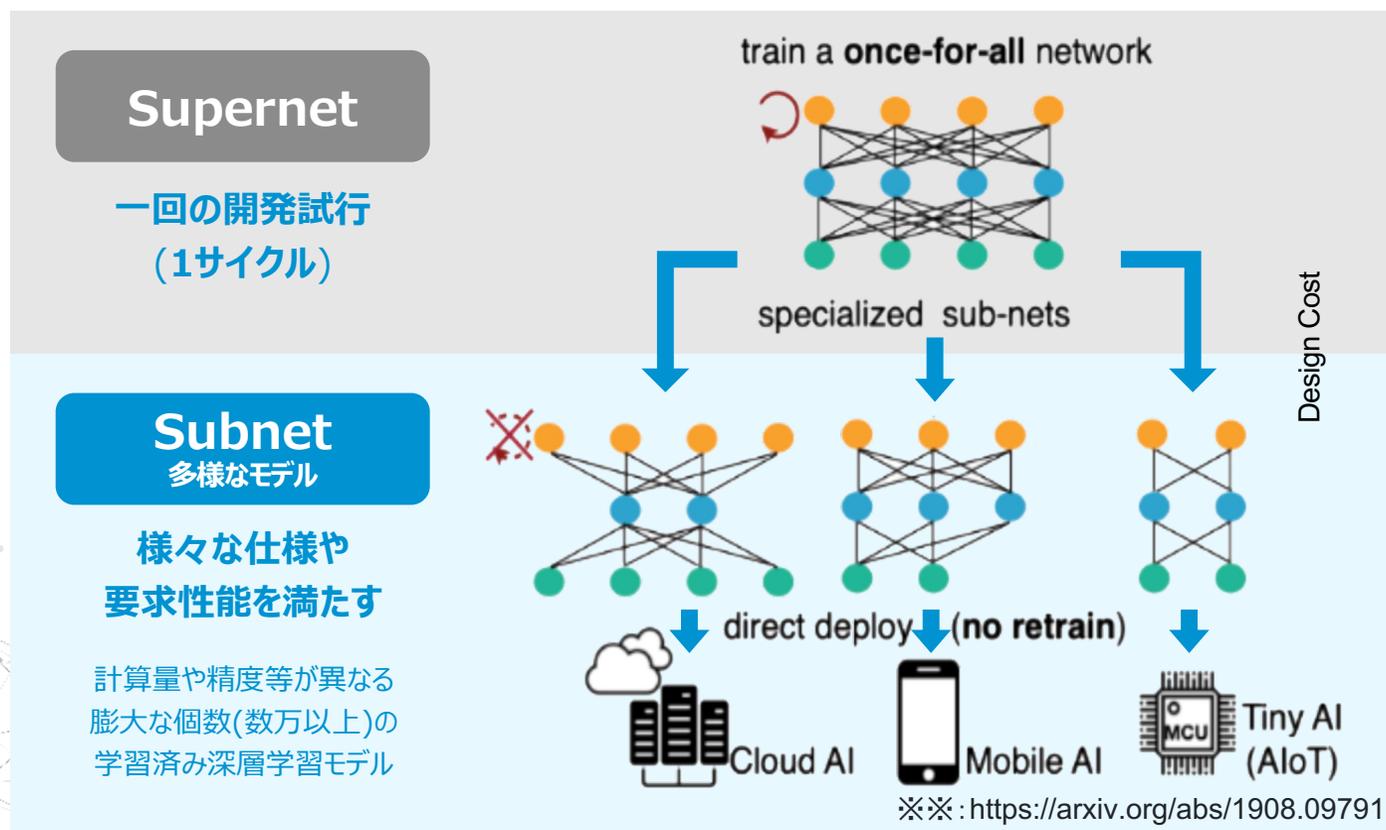


Intel Arria10 (FPGA系)

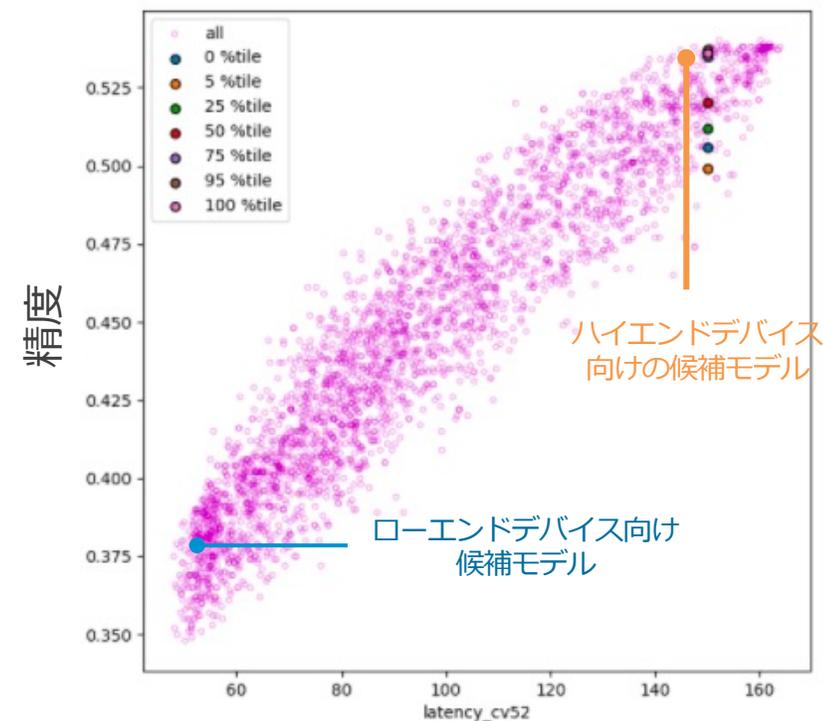
エッジAIソリューションご紹介

スケーラブルなモデル最適化技術のご紹介（サービス名: SubnetX※）

- 1つのAIモデル（Supernet：最大構成モデル）を一度学習するだけで多様なモデル（Subnet）を多数同時に取得できるOFA※※ベースの学習手法により、異なる性能のエッジデバイスへの効率的なモデル最適化技術を提供します。



Subnetの取得イメージ



処理レイテンシ