

OKIアイディエス 会社紹介及び、AIの取組み

株式会社 **OKIアイディエス**

社 名 : 株式会社 O K I アイディエス (OKI IDS Co., Ltd. 略称 : OIDS)

所 在 地 : 群馬県高崎市双葉町 3 番 1 号 (沖電気工業内)

資 本 金 : 5, 0 0 0 万円

株 主 : 沖電気工業株式会社 (1 0 0 %)

設 立 : 2 0 1 4 年 3 月 1 0 日

代 表 者 : 代表取締役社長執行役員 滝澤 家信

従 業 員 数 : 8 2 名

業 務 内 容 : 映像・音声・通信のファームウェア、ハードウェア開発・販売・コンサル
 高速データ伝送のハードウェア開発・販売・コンサル
 通信機器、産業機器、医療用機器 等の製品開発

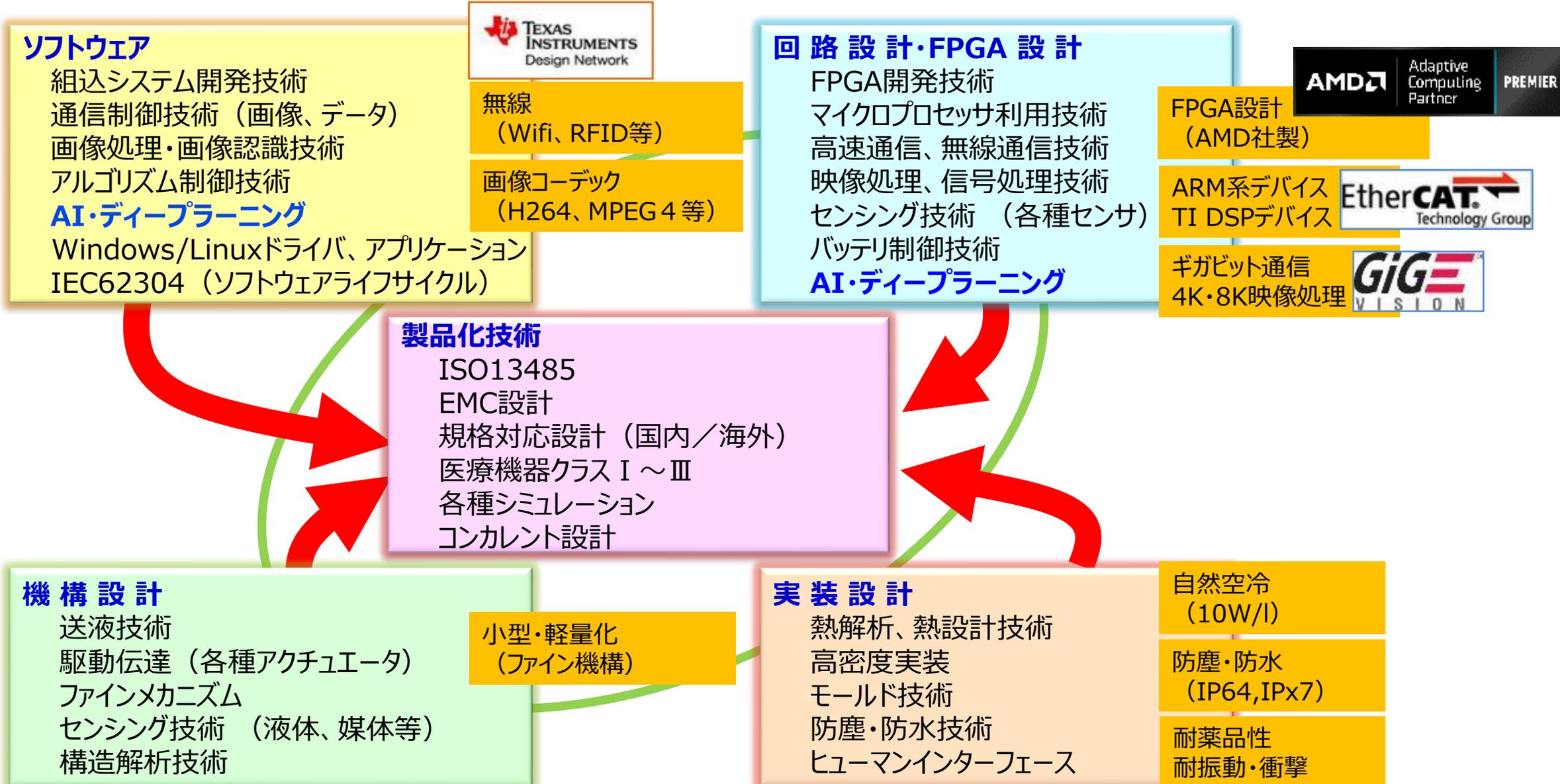


**グローバルトップの技術で、お客様にご満足いただける
 「新たな価値」をご提供し続けます。**



OKIアイディエスはAI、FPGA、ソフトウェア、回路、機構、実装の技術で
 設計・開発サービスをワンストップでご提供いたします

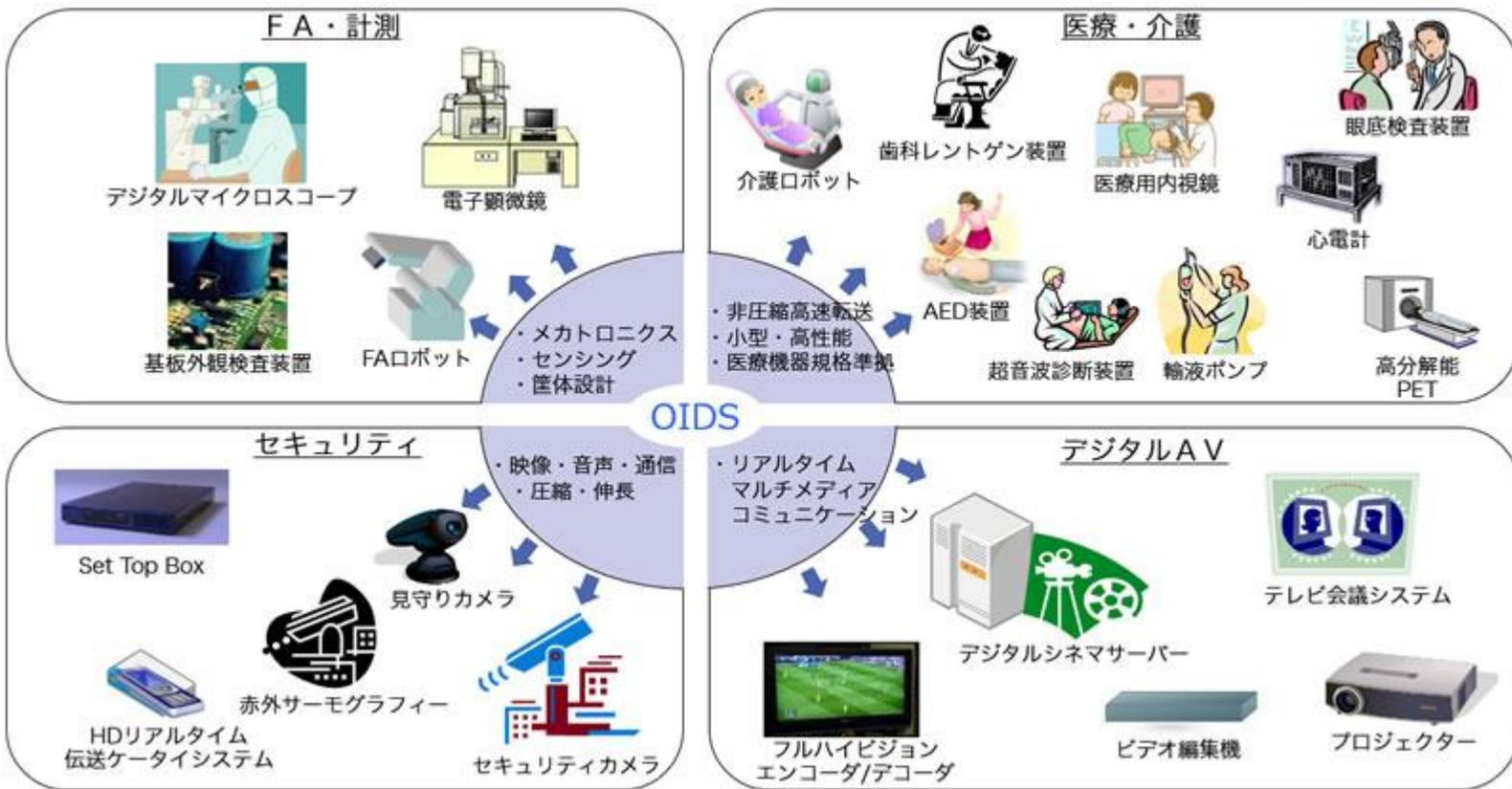
ソフトウェア、回路、機構、実装の総合技術力で、お客様へワンストップで製品サービスをご提供いたします。



映像・音声・通信、高速データ伝送など、OIDSのキーテクノロジーがお客様の製品に組み込まれています

FA・計測、医療・介護、セキュリティ、デジタルAV等、広範囲な領域のお客様製品の受託開発サービスを展開。

エッジデバイスにおけるAI・DLのデザインサービスを強化しております。



■ AI・DL受託サービスの内容

- ・お客様で開発したAIアルゴリズムの最適化
- ・AIモデルのプルーニング
- ・高位合成による、アルゴリズムの軽量化、ハードウェア化

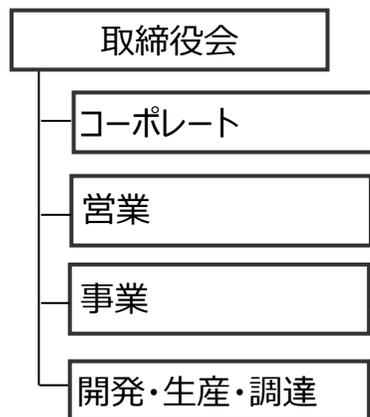
■ AI認定合格者の増強

日本DL協会主催 E資格合格者 5名
 日本DL協会主催 G検定合格者 29名



OKI会社概要

- ◇創業：1881年（明治14年）
- ◇売上高：3,691億円（2023年3月期）
- ◇従業員：14,653名（グループ連結）
- ◇資本金：440億円



取締役会長
鎌上 信也



代表取締役社長執行役員
兼 最高執行責任者
森 孝廣



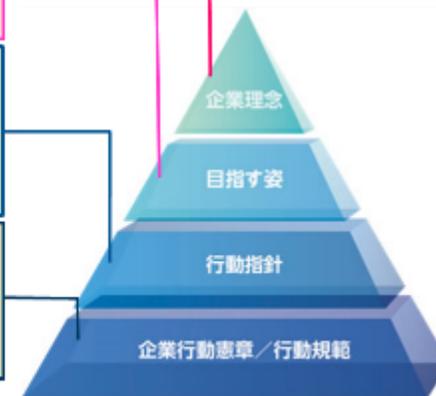
企業理念
OKIは「進取の精神」をもって、
情報社会の発展に寄与する商品を提供し、
世界の人々の快適で豊かな生活の実現に貢献する。

目指す姿
“モノづくり・コトづくり”を通して、
より安全で便利な社会のインフラを支える企業グループ

行動指針
誠実であれ
変革に挑戦する
迅速に行動する
勝ちにこだわる
チームOKI

OKIグループ企業行動憲章
OKIグループが企業理念に基づき果たすべき社会的責任

OKIグループ行動規範
OKIグループ企業行動憲章に則って役員および社員が
とるべき行動の規範

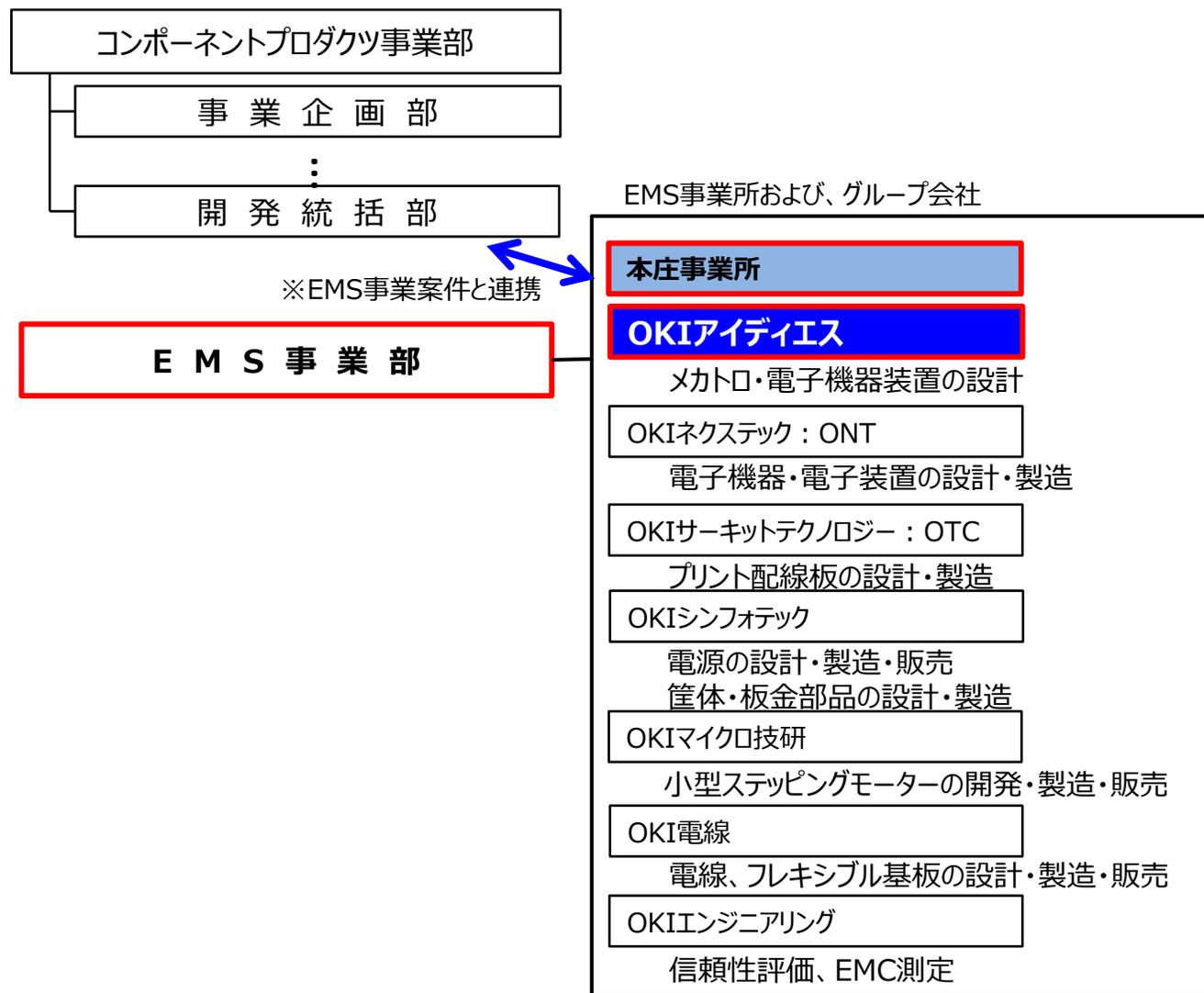


1980～
LEDプリンター
VoIPシステム
経軌調流型 ATM

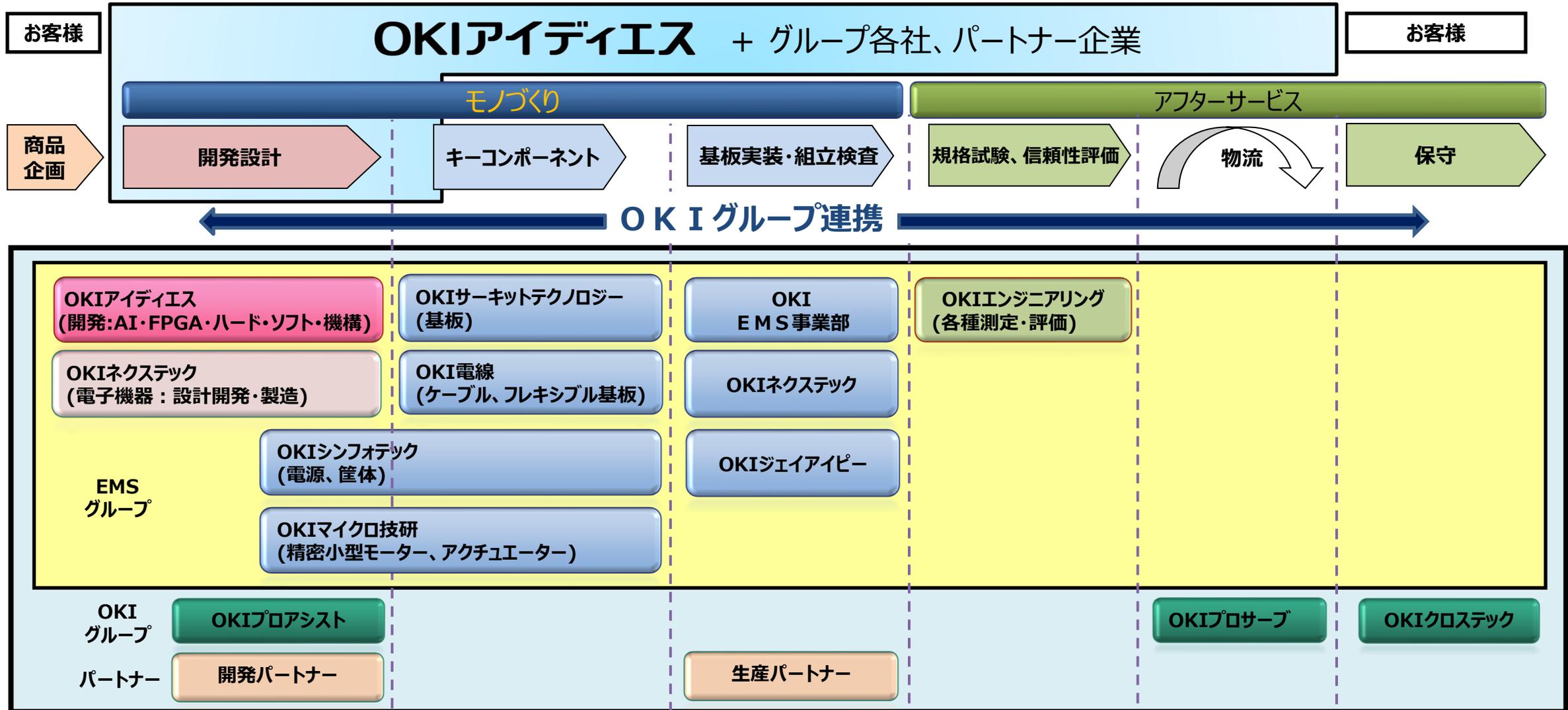
2000～
COREFIDO
ATM-Recycler G7
初期のEMS工場

2010～
AIエッジコンピューター AE2100
AIエッジロボット
ハイジニック タッチパネル

装置の方式設計から、部材調達、製造、組立、検査、信頼性評価まで対応します。



お客様にご満足いただける商品をご提供するために、仕様検討から製品出荷に至るまで
OKIグループ各社、パートナー企業と連携して対応いたします



- AMD社とのアライアンス プレミアメンバー締結の実力を保有
- デザインの品質確保を行うための開発フローの構築
- I P プラットフォームベースの開発による開発期間の短縮
- OKI製品を含めた多数の外部顧客様製品へのFPGA開発実績
- 開発デザインに対する動作保障の確保（サポート含む）

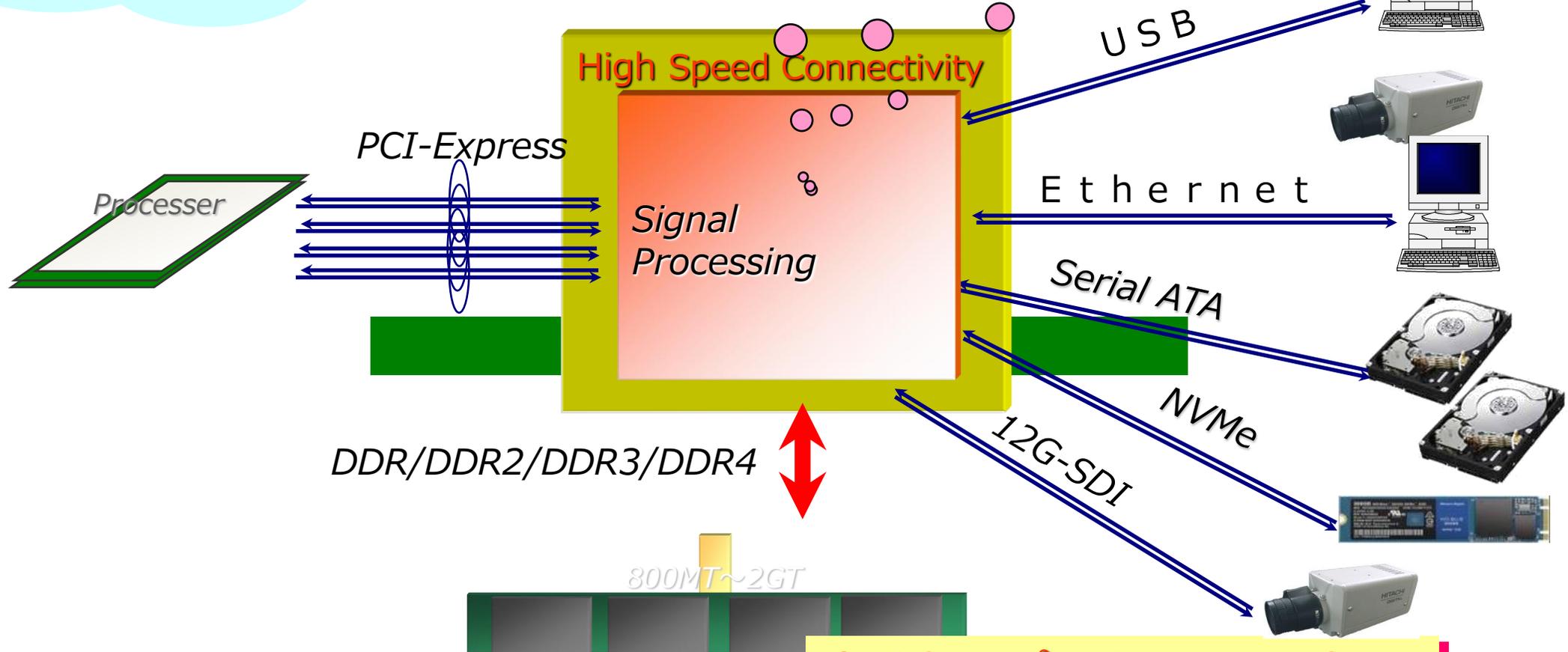


組込み系 F P G A デザイン実績からの品質保証

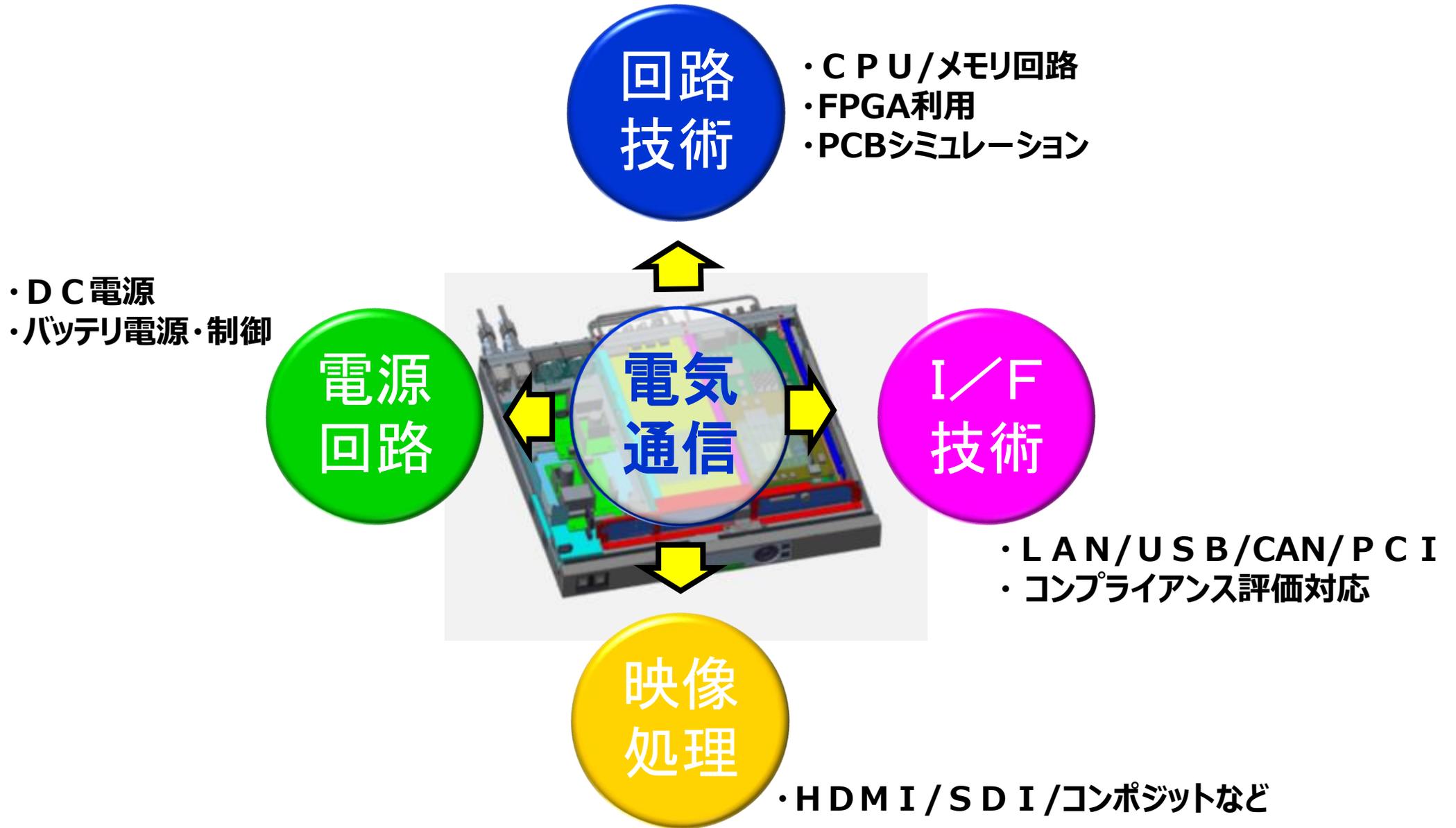
得意分野キーワード

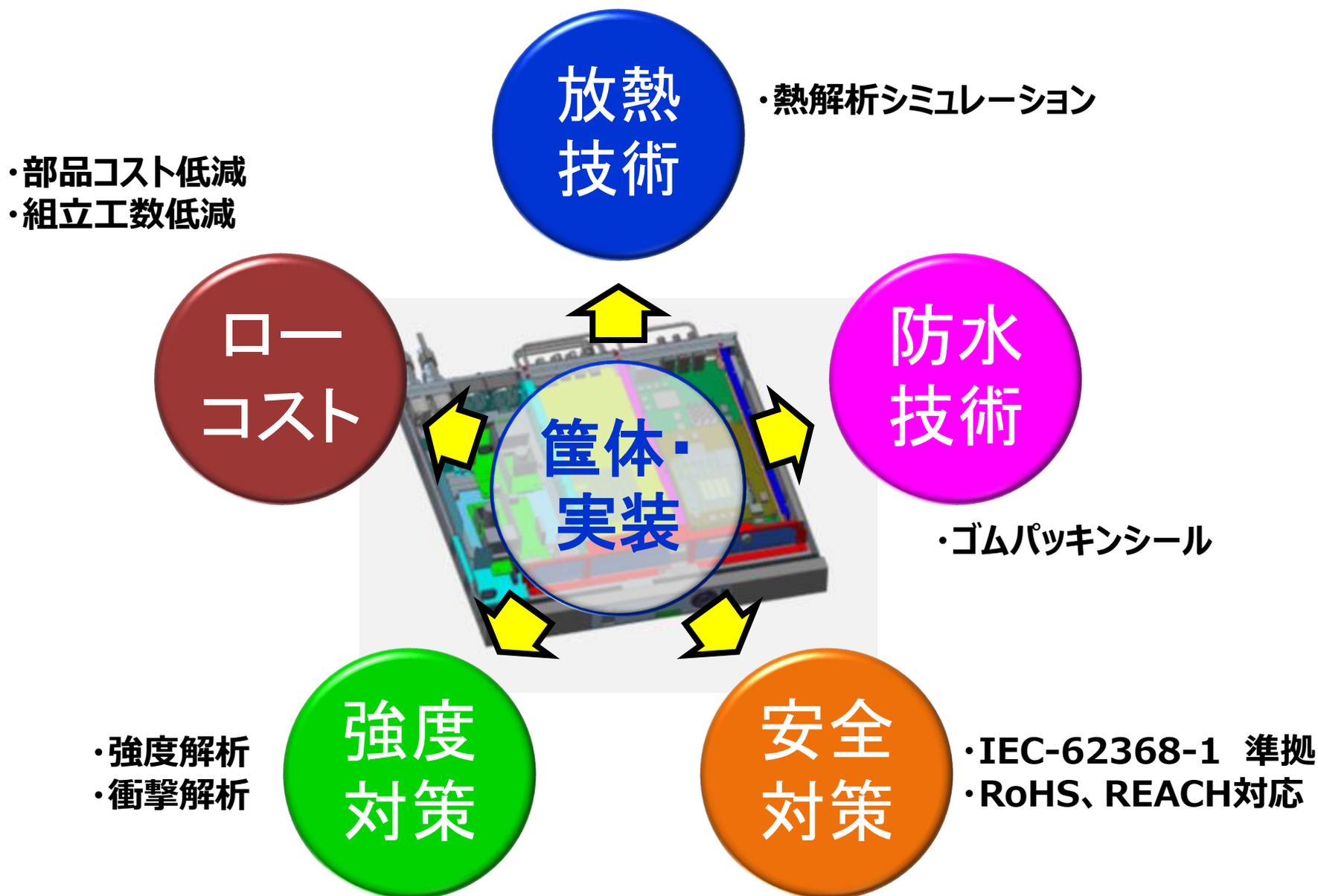
- | | |
|-------------------|-------------------------------------|
| ・I/Oコネクティビリティデザイン | <PCI、PCIe、USB、Ethernet、SATA、3G-SDI> |
| ・組込みデバイス制御デザイン | <外部MPU制御、各種Device制御> |
| ・組込プロセッサ制御 | <MicroBlaze、SoCプロセッサ制御> |
| ・映像処理技術 | <各種映像トリック処理> |
| ・信号処理技術 | <FFT、FIR、線形補完、スプライン（PIX）補完等> |

各種の画像／信号処理技術の高速化に伴い、さらに大容量のデータを高速に伝送する機能が必要。



高速転送プラットフォームの提供





AI関連

AIを取り巻く環境

■ 人工知能（AI）が身近になり、活用がはじまっている

- ✓ インターネット検索エンジン、応答アプリケーション
- ✓ 人型ロボット
- ✓ 画像解析、異常検知
- ✓ 掃除ロボット、自動運転、自立走行制御、監視カメラ



■ リアルタイムを求められる現場では「エッジAI」

✓ クラウドAI

商品（装置）から送信されたデータを、クラウド上のAI（データセンター等）が学習・推論する技術



クラウドAIは、ネットワーク経由でのやりとりが必要なため、リアルタイム性やセキュリティに課題有

✓ エッジAI

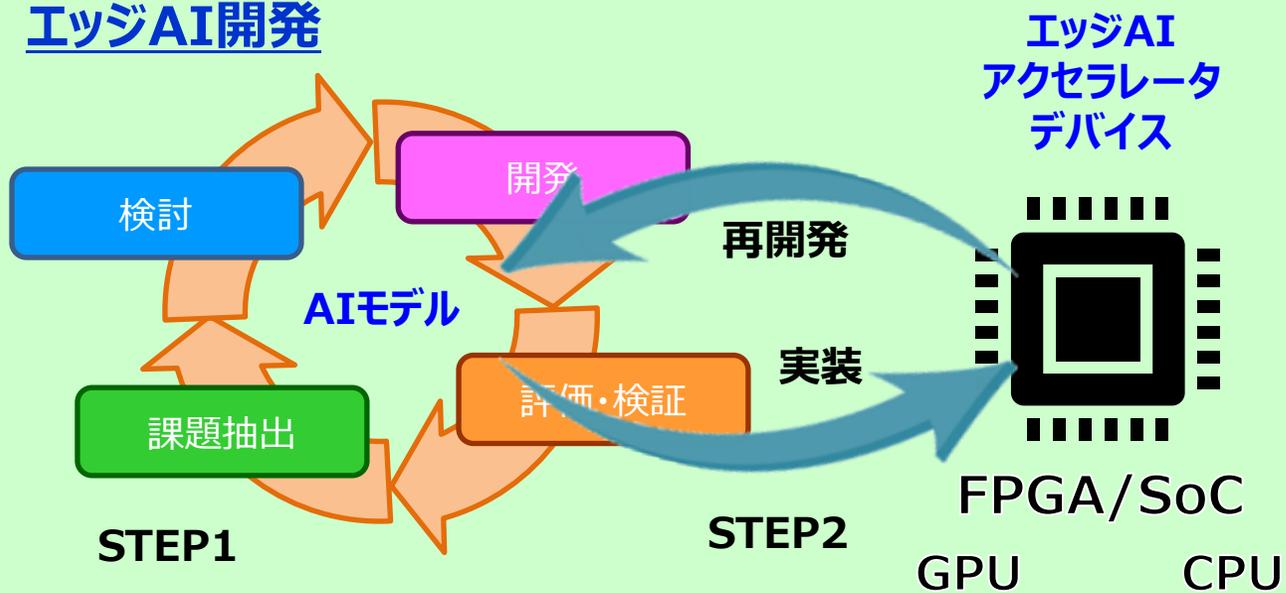
装置に実装されたAIが、学習・推論する技術
ネットワークが必須ではなく、処理の遅延が少ない



<求められる特性> 低消費電力、リアルタイム性、低遅延、コンパクト(サイズ)、耐久性

エッジAIの開発プロセスとアクセラレータ

エッジAI開発



- エッジAI開発は2 STEP
STEP1 : AIモデル開発
STEP2 : AIモデルの装置への実装
- 作成したモデルの評価検証を行い、課題を抽出。
- 目標品質を確保できるまで繰り返す (アジャイル開発)

エッジAIに求められる特性

- **低消費電力** : 長時間駆動、脱炭素
- **高効率** : 無駄の無い処理
- **低遅延** : 高速処理
- **リアルタイム性** : ハードウェア最適化
- **高柔軟性** : 容易なカスタマイズ性
- **ロバスト性** : 様々な環境で動作
- **小型・軽量** : 設置場所の自由度

FPGA/SoC

今後のエッジAIを支える
キーデバイス



OKIアイディエスのAI開発アプローチ

- AIモデルがこれからのお客様 (お持ちでないお客様) ⇒ AIモデルの作成・開発をサポート
- AIモデルが作成済みのお客様 ⇒ フィジビリティスタディを実施してハードへの実装をサポート
- AIモデルのハードウェアアクセラレーション

AI開発を組み込んだ商品開発プロセス

製品開発フロー



AIモデル開発

FPGA/SoCへの実装



- ・AIという新しい技術を取り込むことにより、商品(装置)開発は、今までよりも広範囲の検討が必要で、且つ難解に。
- ・AIの工程も分業化、細分化されており、それに特化した業態が主流で、進め方がわからない・・・

<お客様のご要望>

旧来通りのシステムインテグレータではなく、システムトータルでの課題を把握しながら、

適切にAIの組み込み、提案することができる知識、技術を持ったインテグレータが求められています。

AI/DL開発・設計サービスの特徴

AI/DL最適化設計サービス

➤ OIDSのFPGAによるAI/DL設計サービス

- AI技術 (AI認定の合格者)
- OKI PCAS、Mipsology社ZebraによるFPGA実装への開発時間の短縮・効率化を実現

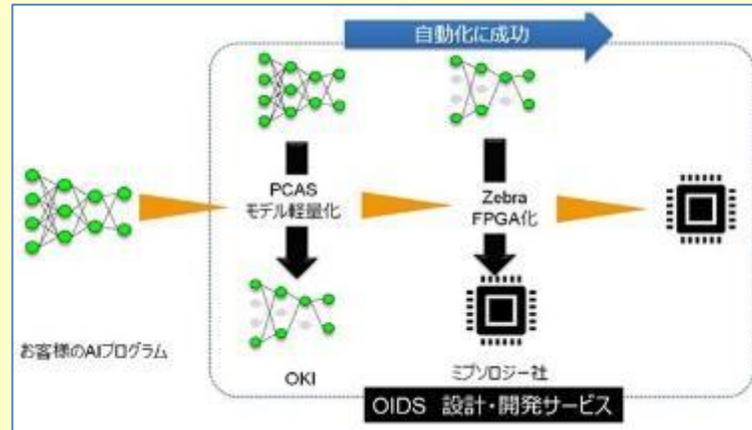


➤ OKI PCAS

- OKIのAI軽量化技術
- CNNモデルの圧縮

➤ Mipsology社 Zebra

- CPU/GPUで設計されたニューラルネットワーク (NN)をFPGAにシームレスに置換または補充可能



➤ 実装例

<p>CPU推論 未圧縮モデル 1fps</p> <p>①CPU(Original)</p>	<p>Zebra単体 未圧縮モデル 14fps (FPGAによる圧縮実施)</p> <p>②Zebra(Original)</p>	<p>Zebra+PCAS 低圧縮モデル 30fps以上 (FPGAによる圧縮実施)</p> <p>③PCAS(低圧縮率)</p>	<p>Zebra+PCAS 高圧縮モデル 60fps以上 (FPGAによる圧縮実施)</p> <p>④PCAS(高圧縮率)</p>
---	--	--	--

FPGA設計

➤ OIDSのFPGA開発の特徴

- AMD社の日本国内唯一のプレミアパートナー
- 高位合成・検証サービスやDFX等のツール、技術の導入により、開発装置の品質向上、開発期間の短縮に貢献

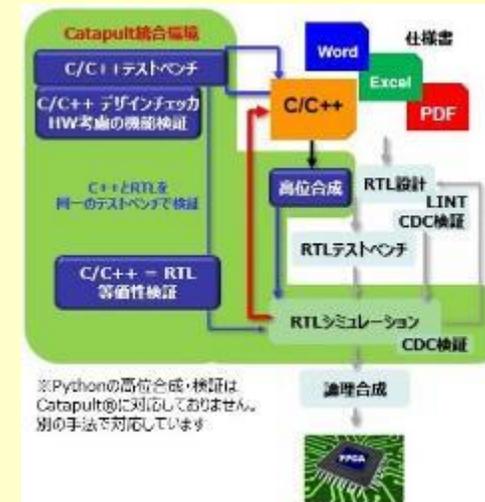


➤ 高位合成(HLS)開発・高位検証サービス

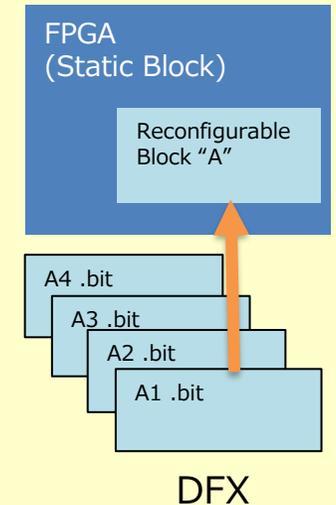
- 業界 (高位設計環境) 唯一のHLS/HLV統合環境 Catapult®
- 高位合成設計のQoR (Quality of Result) を等価性保証で確保

➤ DFX(Dynamic Function eXchange)

- ゼalinkスが提供している、システム動作中にロジックブロックの一部を動的に書き換えるための機能やデザインを含むソリューション
- 効率的な機能設計を行うことで、容量の小さなデバイスへの実装が可能に



※Pythonの高位合成・検証は、Catapult®に対応していません。別の手法で対応しています。



DFX

OKIアイディエス DNNアクセラレート ソリューション

AMD FPGA/SOC アクセラレート

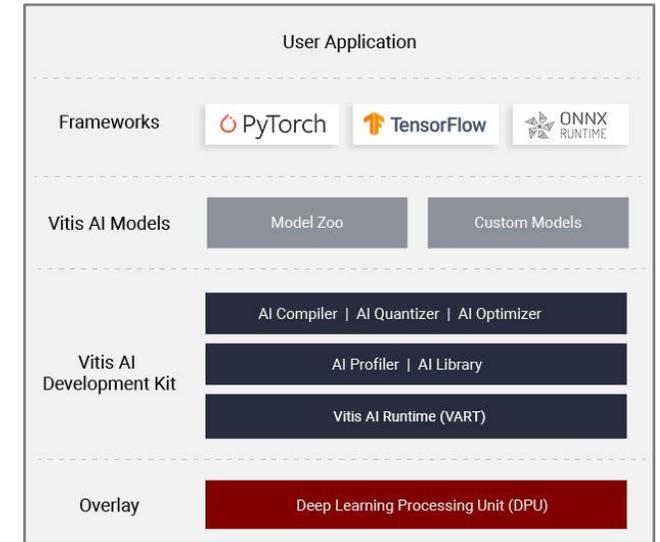
- **Vitis AI** : AI推論開発プラットフォーム
- **Zebra** : AI推論開発プラットフォーム

AMD社のAI推論開発プラットフォーム 最適化されたモデルやライブラリを使いFPGA上で高速推論

- 多様な深層学習タスク、CNN、RNN、NLP に対応できる主要フレームワークと最新モデルをサポート
- クオンタイザーおよびオプティマイザーツールを使用して、モデルの精度や処理効率を最適化
- シンプルなコンパイルフローと高レベルAPIを使用して、カスタムモデルの迅速な運用が可能
- 高効率でカスタマイズ可能な DPUコアを使用して、スループット、レイテンシ、消費電力など、エッジやクラウドのさまざまな要件に対応可

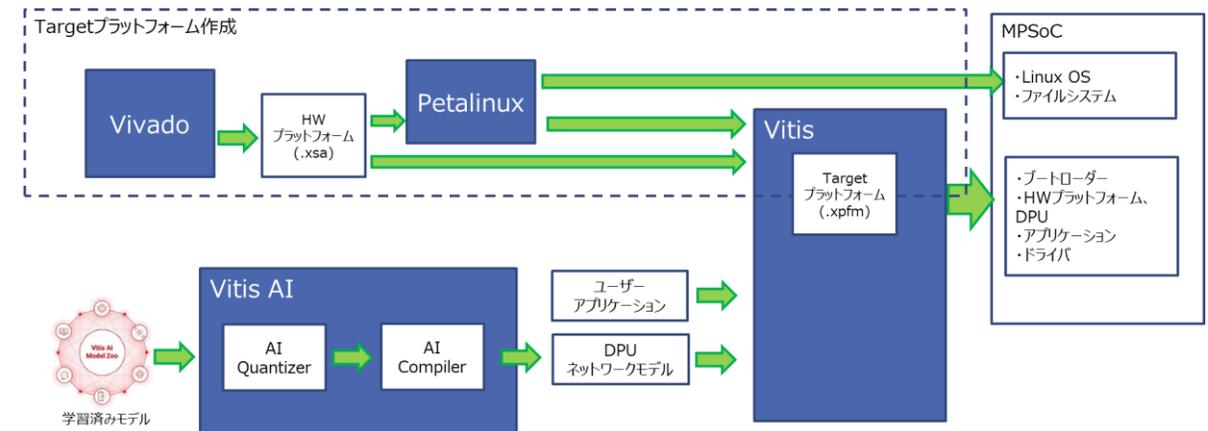
課題

- Vitis AI/DPUがサポートしていない機能は実装困難 (AMDのサポートや、FPGAのデザイン、CPUでの再検討要)
- フレームワークを直接サポートしていない(独自ファイルに変換)



Vitis AIの構造

Vitis and Vitis AI Design Flow



OKIアイディエス DNNアクセラレート ソリューション

AMD FPGA/SOC アクセラレート

- **Vitis AI** : AI推論開発プラットフォーム
- **Zebra** : AI推論開発プラットフォーム

GPUで学習したニューラルネットワークを1コマンドでFPGAで実行

libZebraによるプラグ&プレイ推論 (GPUは学習向けとして使用)

- 同一のニューラルネットワークを使用
- 同一のアプリケーションソフトウェアを使用
- 同一のフレームワークを維持
- 単一のLinuxコマンドを使用可能
- Zebraは自動的に量子化を実行

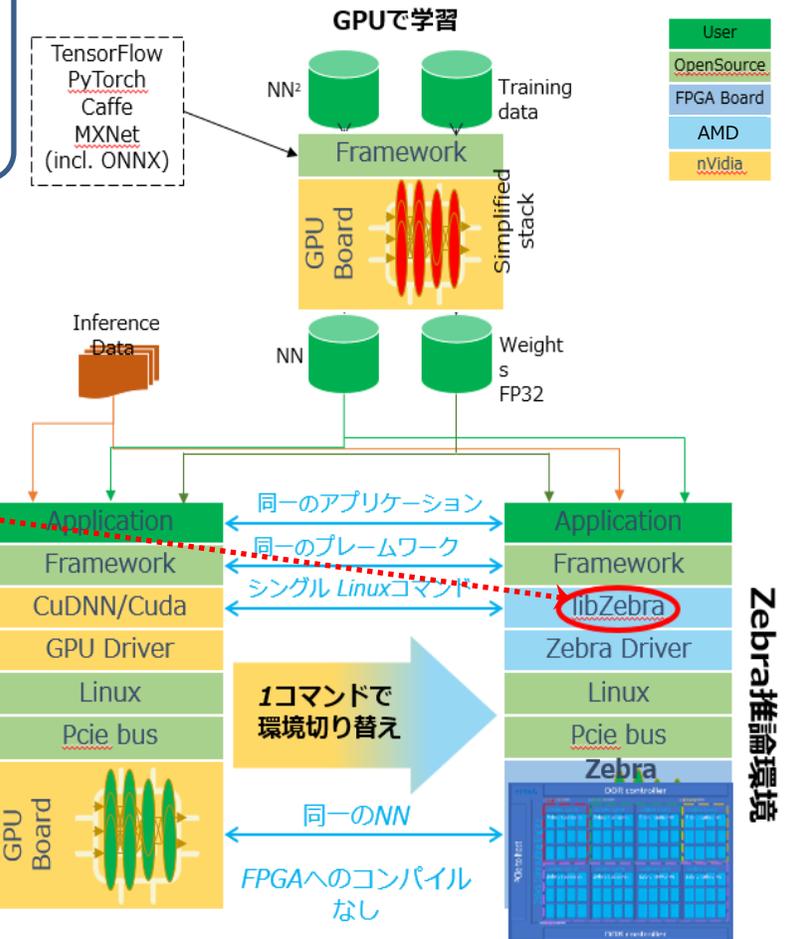
FP32で学習は行われ、Zebraはモデルをint8で推論演算

- Zebraの量子化は、新しいツールなしで完全に自動化
- Zebraは、学習データを用いた再トレーニングの必要なし

量子化

- 高速 (数分) 量子化
- 一般的な精度の低下は、FP32推論の1%未満

NN	量子化後の精度 Zebra vs. FP32
Inception-V2	-0.8 %
Inception-V3	-0.5 %
Inception-V4	-0.4 %
ResNet152	-0.7 %
ResNet50	-0.5 %
VGG16	-0.3 %
VGG19	-0.2 %
YOLO-V2	-0.8 %
YOLO-V3	-0.1 %
MobileNet-V2	-1.8 %



- User
- OpenSource
- FPGA Board
- AMD
- nVidia

TensorFlow
PyTorch
Caffe
MXNet
(incl. ONNX)

GPUで学習
NN² Training data
Framework
GPU Board
Simplified stack
NN Weight s FP32

CPU/GPU環境

Zebra推論環境

1コマンドで環境切り替え

同一のNN
FPGAへのコンパイルなし

FPGA/SoC高速化プラットフォーム「Zebra™」ソリューション

➤ Zebraスタンドアロン AIアクセラレータ

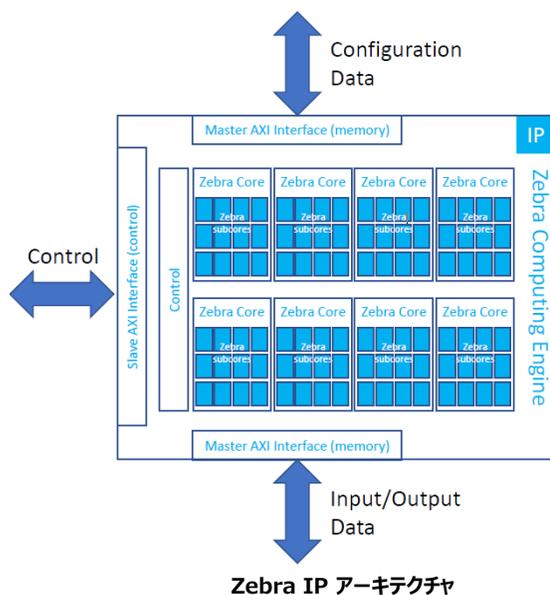
通常Zebraは、PCホストにプラグインされたFPGAカードで実行



➤ Zebra IP

MPSOC, VersalなどのFPGA上で他の機能と統合可能なマクロ

- ユーザ要求に基づいたカスタム仕様 (サイズ、パフォーマンス、レイテンシ等)
- AXIインターフェースで接続

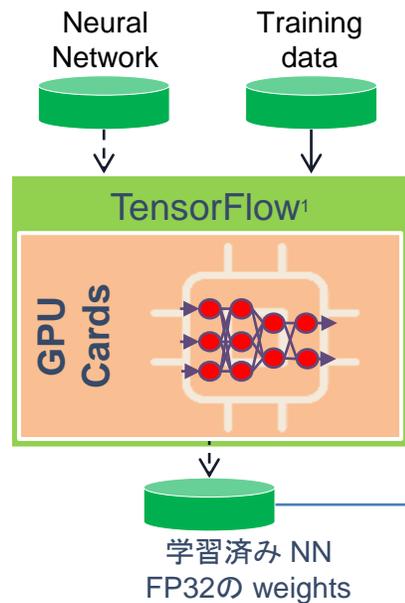


	Zebra スタンドアロン AIアクセラレータ	Zebra を IP として統合 (ユーザロジック含む)
概要相違	Zebra は PL 部分を 100% 使用	Zebra IP は PL の一部を使用
ML ス프레ームワーク	実行に必要ありません Zebra は ML フレームワークなしで推論を実行可能 ご要求によっては、TensorFlow、PyTorch、ONNX から実行	
組込ソフトウェア	Zebra 組み込みソフトウェアは Python 又は C ベース API で実行	
FPGA ビットストリーム	コンパイル済みビットストリームを提供	コンパイル済み Zebra IP (Vivado DCP netlist) をご提供 お客様は、ユーザロジックとともにフルチップのコンパイルを実行

2023年3月29日 OKIアイディエスWebセミナー「高位合成・高位検証プロセス、AIディープラーニング推論アクセラレーターMipsology Zebra」資料より抜粋

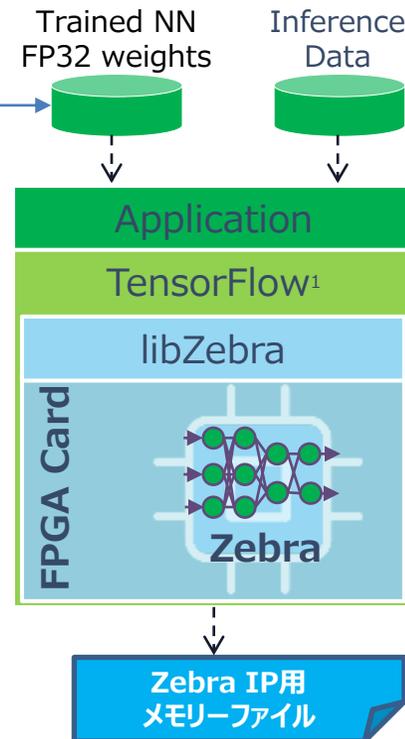
FPGA/SoC高速化プラットフォーム「Zebra™」フロー

1. GPUでのNN学習



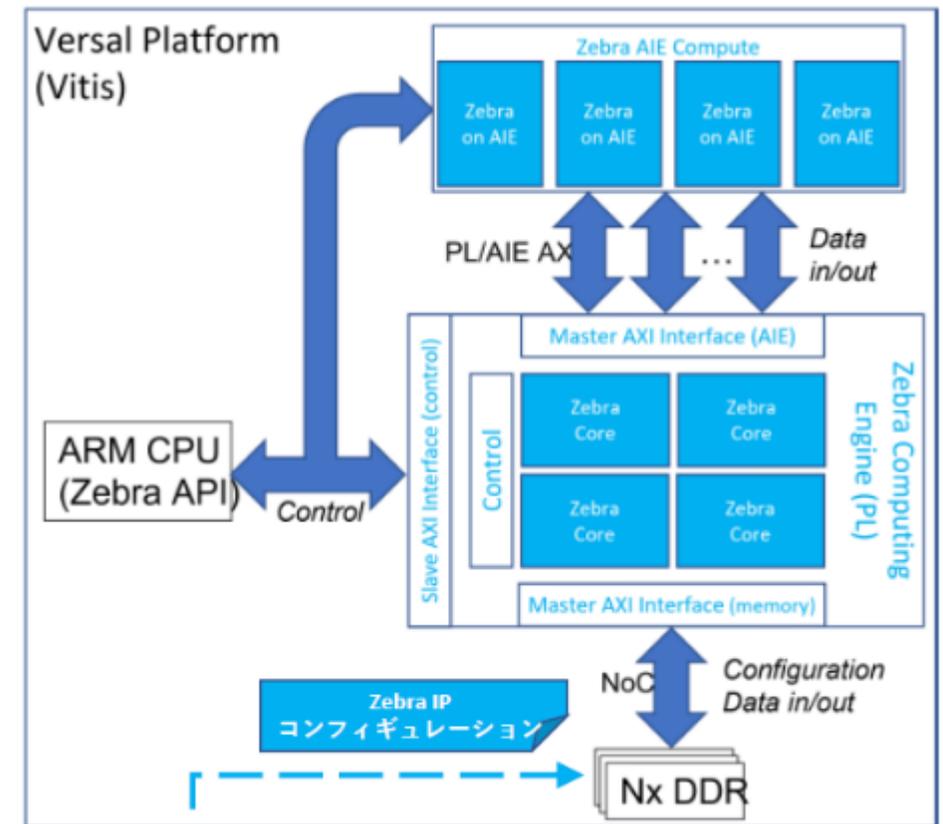
2. FPGA PCIeカードを使用し、NNフレームワークを使用したホストからのZebra実行確認

- ・量子化により、Zebra IP用のメモリアイルを生成



3. Zebra構成をお客様のターゲットシリコンに展開

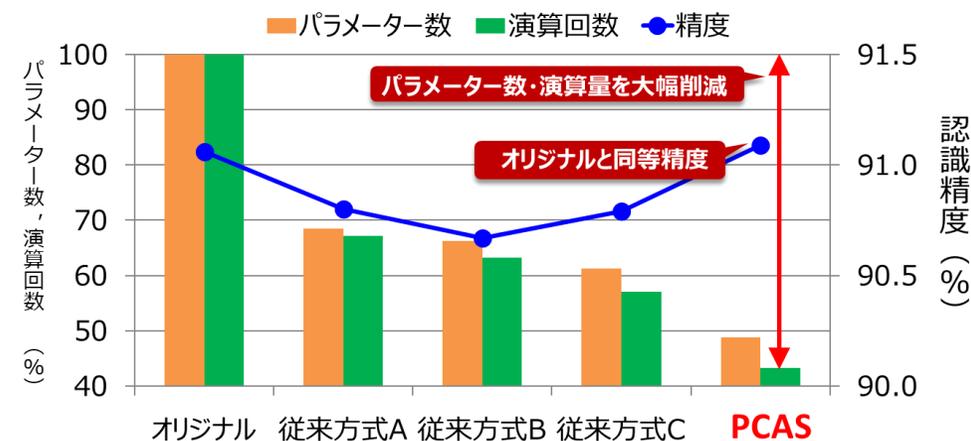
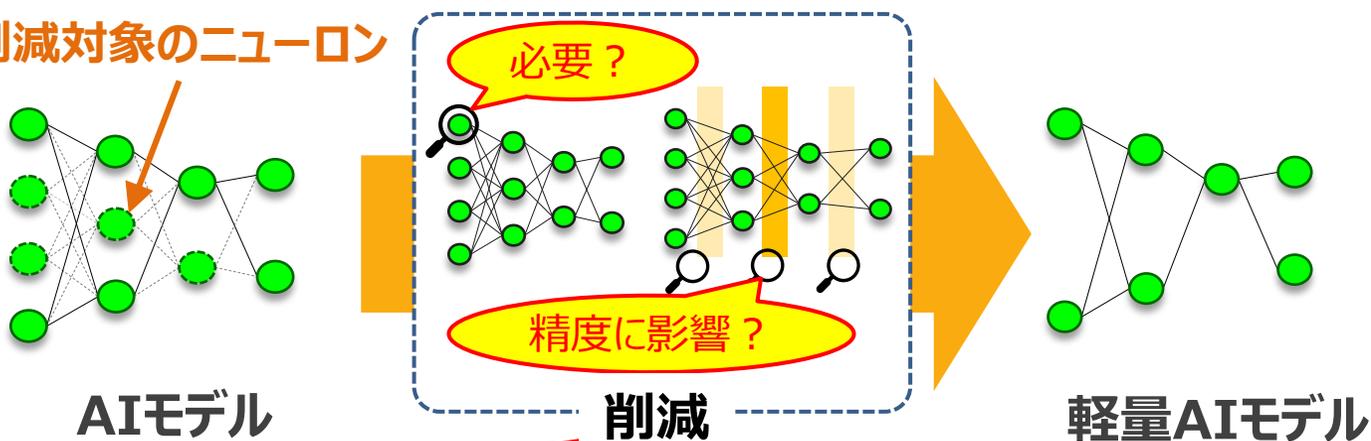
- ・シンプルなAPIを使用してARMからZebraにアクセス
- ・NNの演算にはPS側ARM性能を利用せず、バス帯域への影響も回避



OKI独自のAIモデル軽量化技術「PCAS」

- **OKI独自**のAIモデル軽量化技術（モデルプルーニング）
- AIモデルを構成するニューロンの中で重要でないものを**自動で特定・削減**

削減対象のニューロン



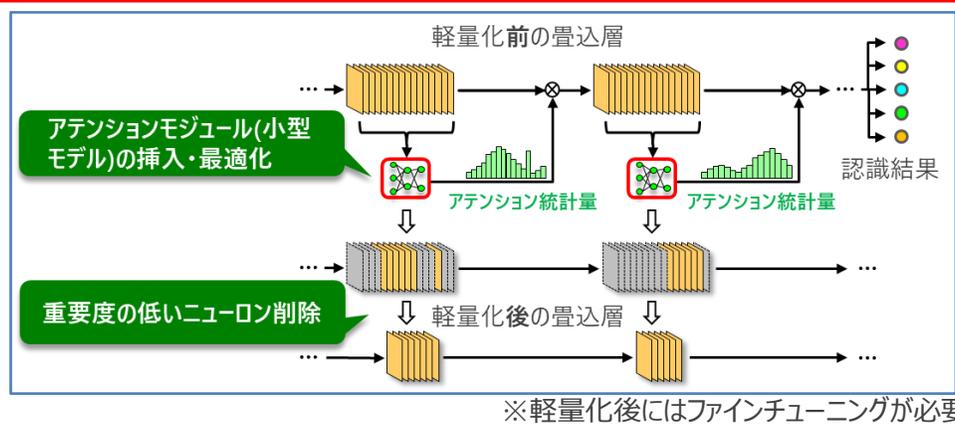
➤ プルーニング基準

アテンション統計量（最適化された小型モデルの出力平均）

➤ 層毎の削減率

統計量に基づく最適化で自動決定

= モデル分析作業が不要



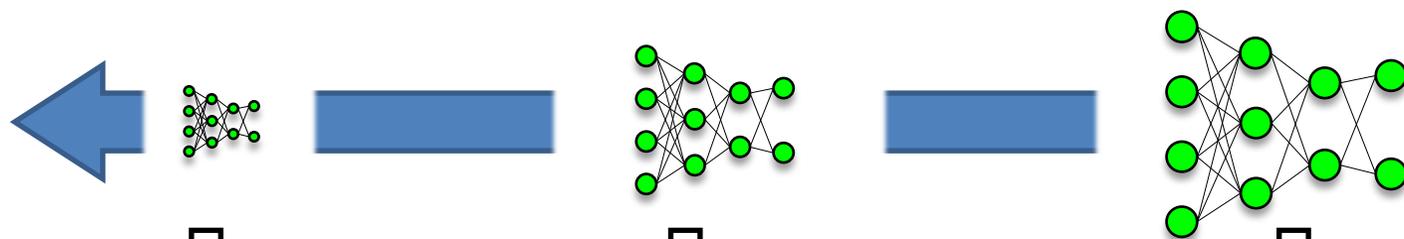
※軽量化後にはファインチューニングが必要

OKI独自のAIモデル軽量化技術「PCAS」

- 高度AIモデルを任意のレベルで軽量化可能
- お客様の**要求性能に応じた**柔軟なFPGA/SoCの選択が可能に

AIモデル

圧縮率：高
サイズ：小



圧縮率：低
サイズ：大

AIモデルの性能・サイズに応じたFPGAデバイスを選択

FPGAデバイス

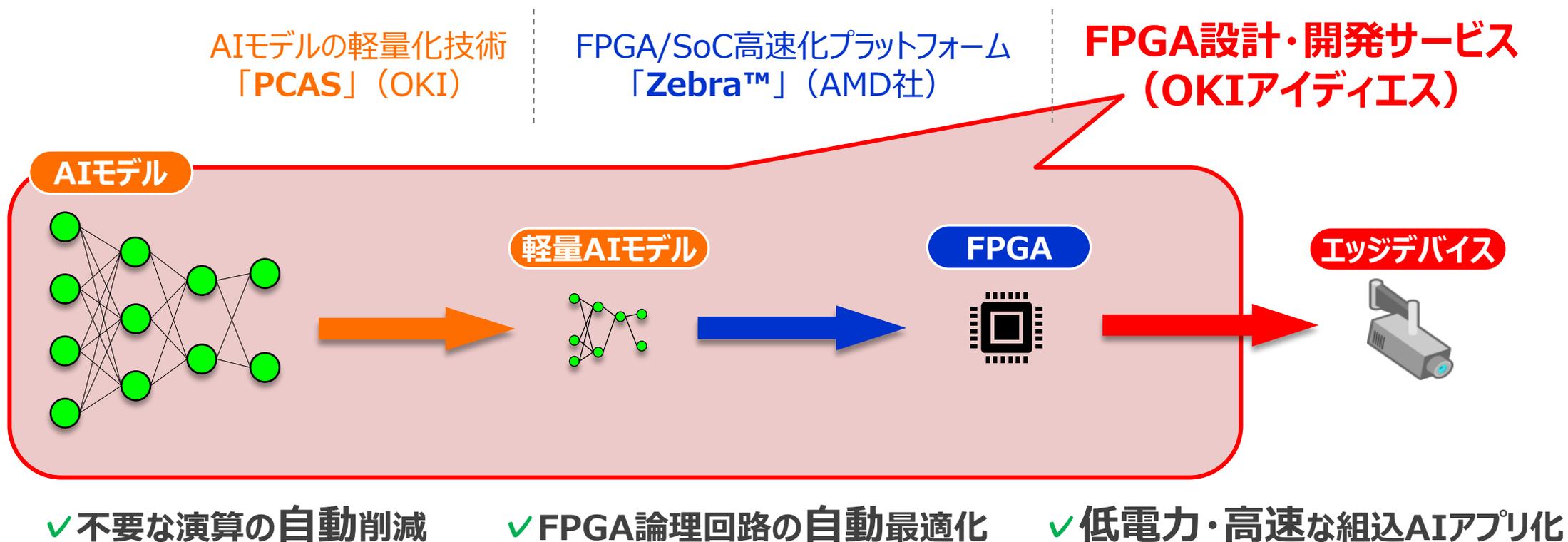
性能：低



性能：高

OKI独自のAIモデル軽量化技術「PCAS」

OKIアイディエスのFPGA/SoC設計開発サービスに「PCAS」を連携することで、お客様の目的に応じ **より柔軟に対応できるFPGA設計開発サービス** の提供を目指す



FPGA/SoC ベンチマーク

GPU



コード移植

FPGA/SoC



- ・CNNアルゴリズム実証
 - ・演算アルゴリズム開発
- 電力効率が悪い

- ・POC機開発
- ・エッジ機器開発

【求められる処理能力】

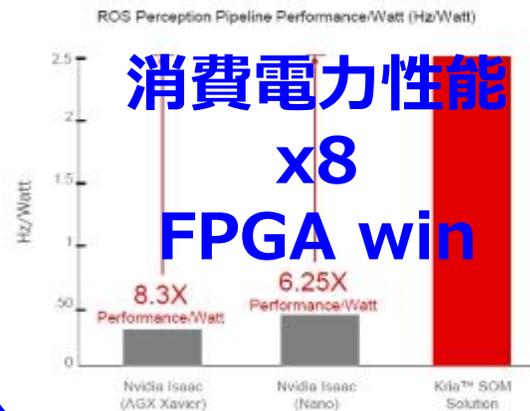
- ✓ 高効率処理：無駄の無い並列処理
- ✓ 低遅延処理：負荷依存の無い制御
- ✓ 低電力処理：必要機能のみでの動作
- ✓ 高柔軟処理：容易な機能入替



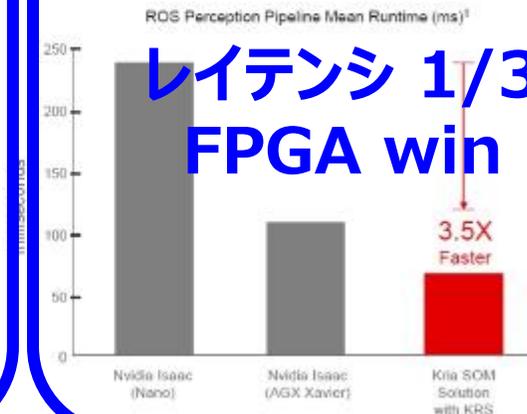
全てに対し、FPGAアドバンテージ

Adaptive Computing for Higher Performance/Watt, Lower Latency

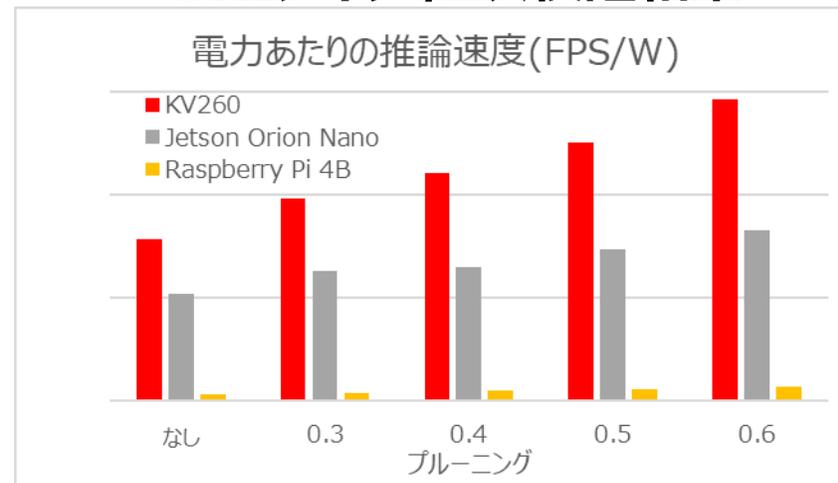
Up to 8.3X Performance/Watt¹
for Computationally Equivalent Workloads



Up to 3.5X Lower Latency²
(Faster Execution Time)



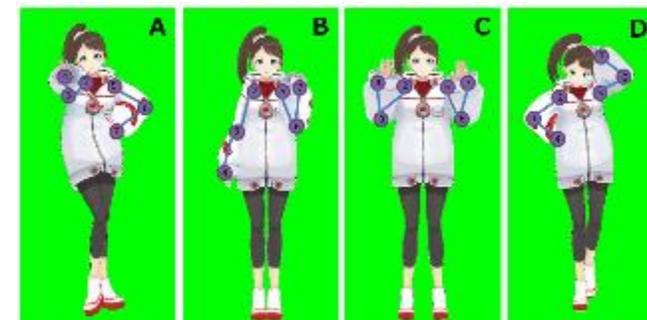
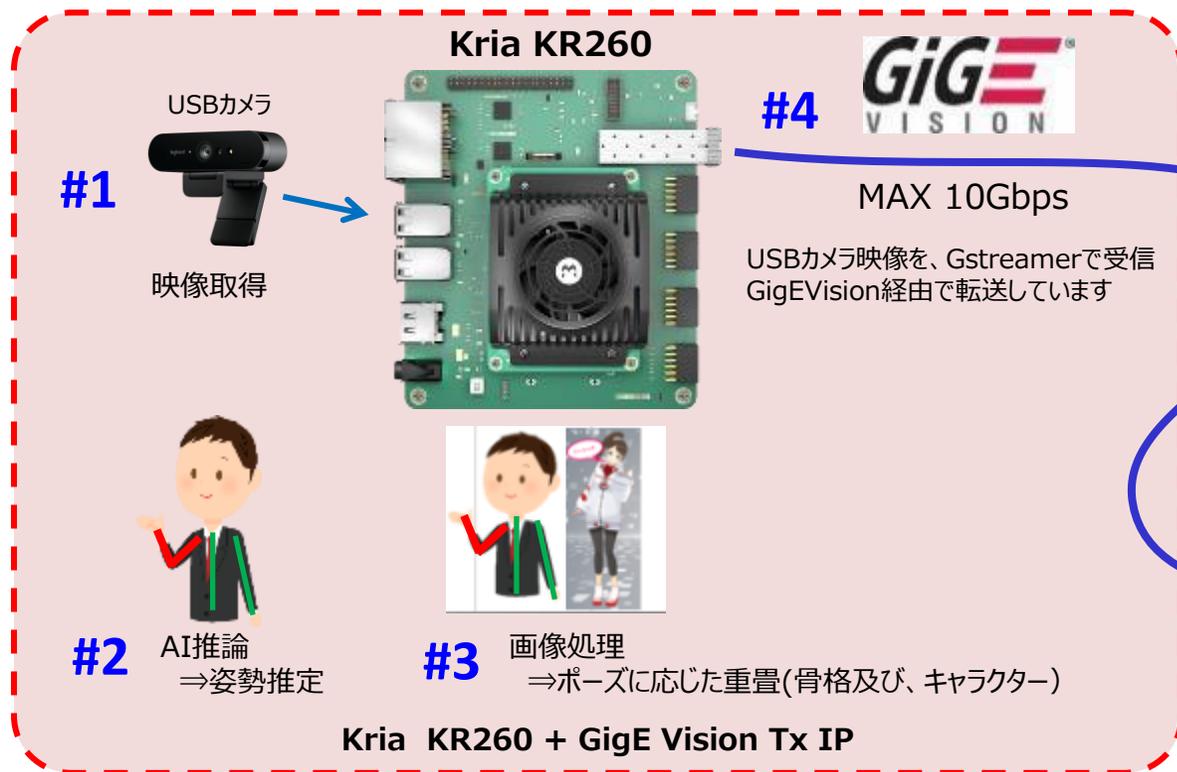
<OKIアイディエス検証結果>



AI関連デモ紹介

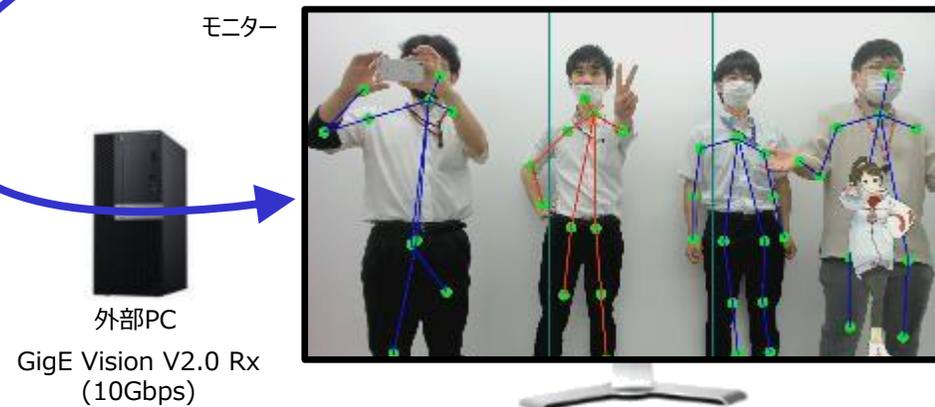
Kria™ KR26 SoM(AMD製)を活用した10GigE Vision/AI姿勢推定

<デモ構成>



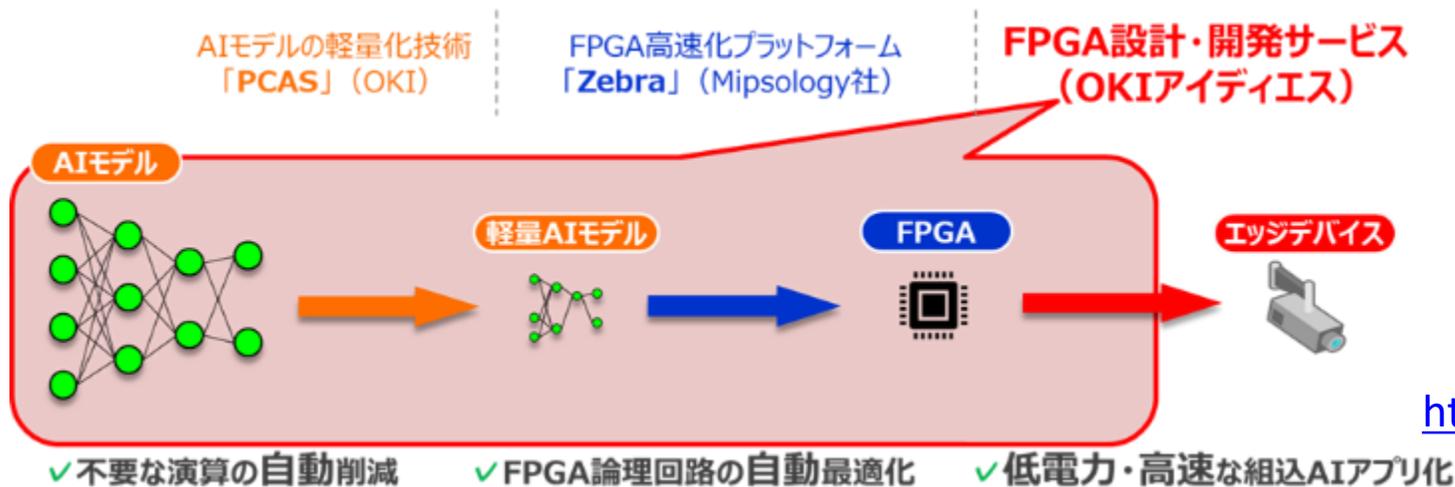
4種類のポーズを検出

※緒希ツナグ®(おきつなぐ)はOKIバーチャルキャラクターです



#1 映像取得 ⇒ #2 AI推論(姿勢推定) ⇒ #3 画像処理(重畳) ⇒ #4 GigE Visionで伝送 までの処理を
“Kria KR260”のみで実施 しています

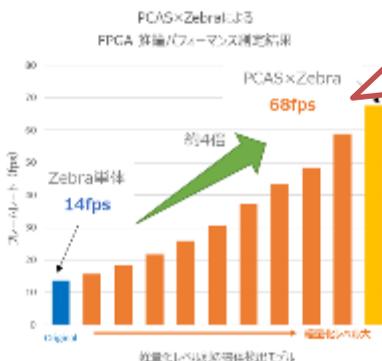
PCAS x Zebra により AIモデルのFPGA/SoCへの実装を自動化し、開発期間/コストを圧縮



<https://youtu.be/Vct11JNni4>

高度なAI処理を用いた評価実験で約4倍の高速化を確認

- モデル軽量化により、高度AI処理を高速化を実現します
 - ・ 80種の物体検出を行うモデル (Yolo v4) を用いた実験では Zebra単体と比較し4倍の高速化効果を確認



車載アプリケーションに求められる処理速度 (60fps) 以上を確認

高速化と認識精度はトレードオフ。各設定を簡単に試せるので、早期に性能を見極めることができます

[プレスリリースはこちら](#)



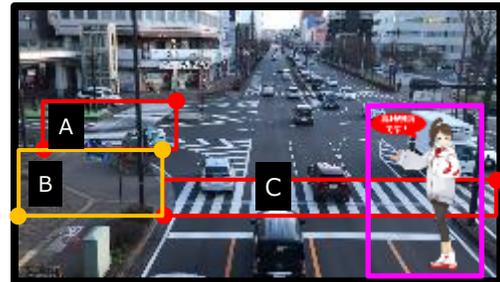
【デモ内容】設定を変えた4種類の方法の映像を同時に表示
会場では、カメラでのリアルタイム映像も表示します。

“特定の領域”に、“特定の物体”を検出した場合、“対応した画像”をリアルタイムに表示

PlanB.リアル (USB⇒SFP+)



リアル映像(展示会場)

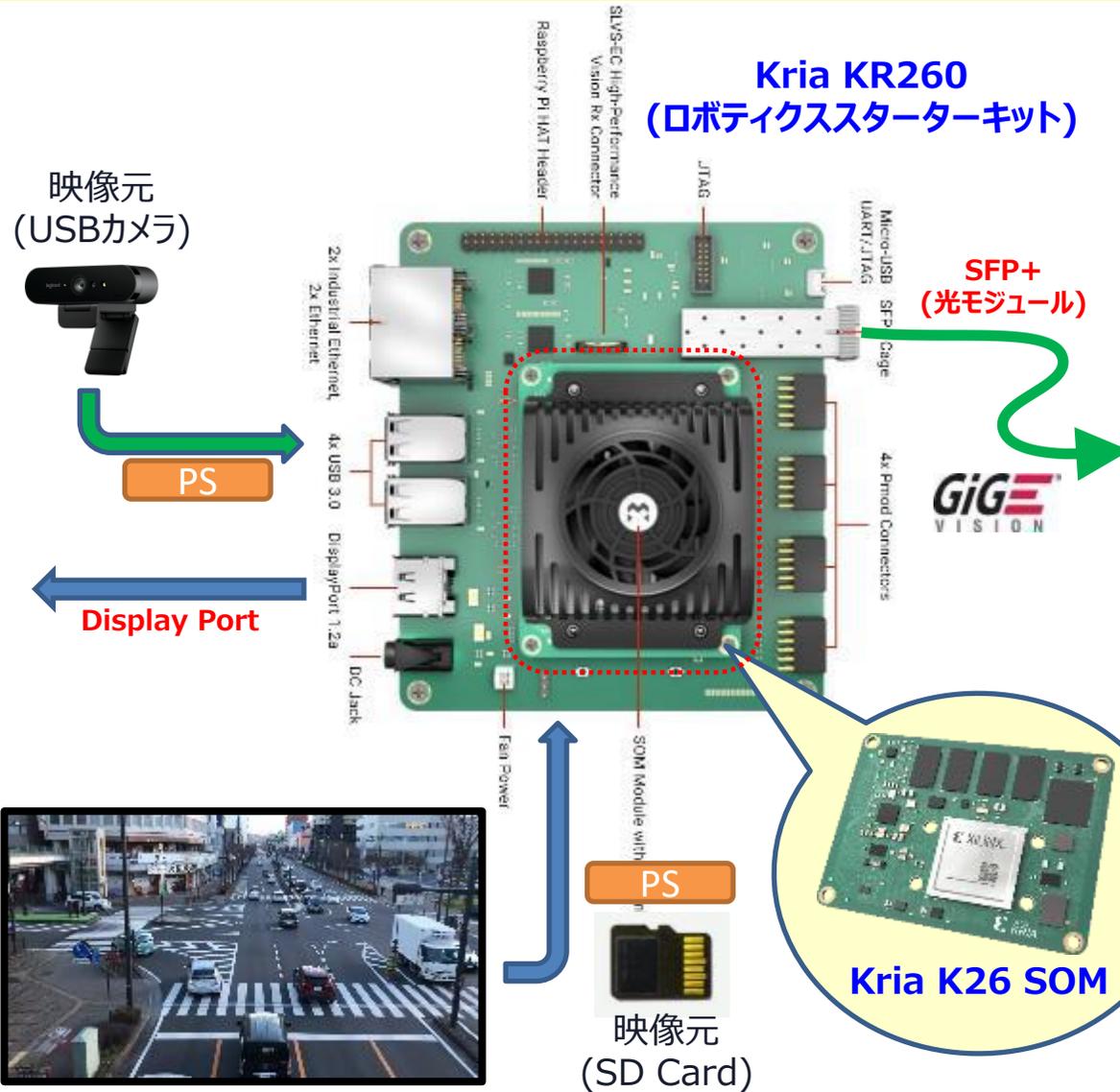


#4 モニター表示

PlanA.録画 (SD⇒DP)



録画再生 (車の走行映像)



https://youtu.be/f5KZ_kuY8Q0

<https://youtu.be/eSZBgNOc7-A>



モニター表示

- #1 映像取得
- #2 AI推論(物体検知)
- #3 画像処理(重畳)・高位合成
- #4 高速画像伝送(GigE Vision)

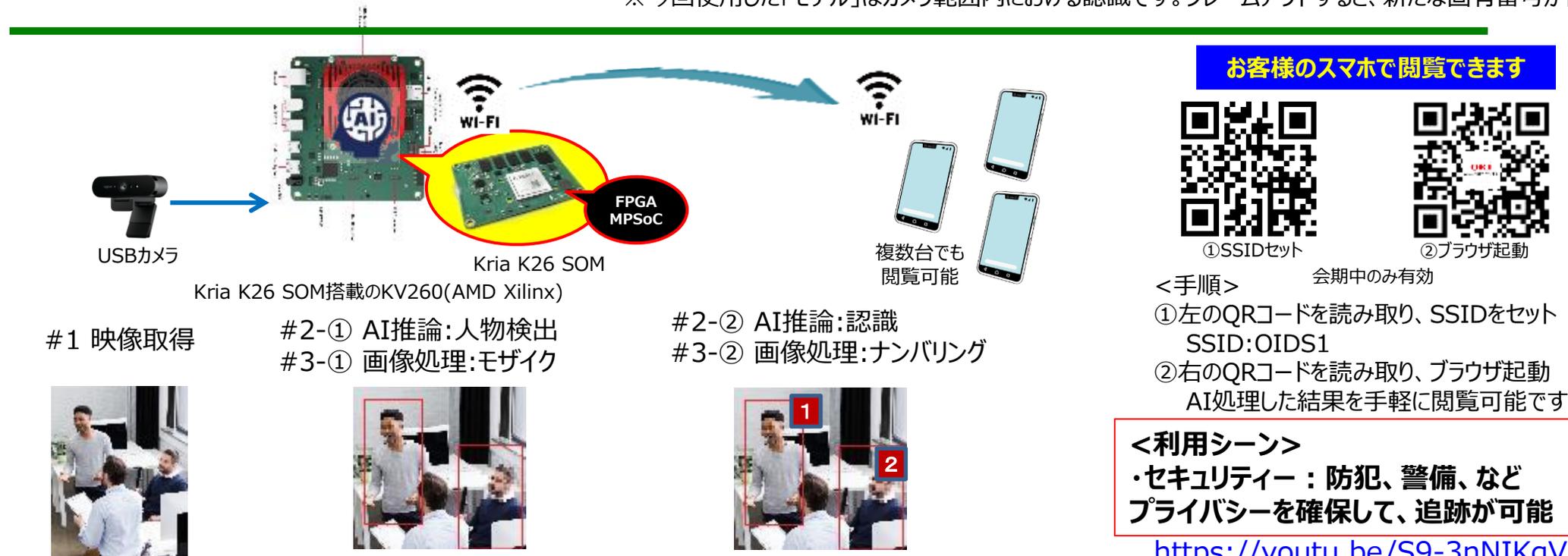
物体検知AI : YOLO v4
(Vitis AI ライブラリ)

小型エッジ端末のみで行うFPGAによるAI処理とストリーミング配信

FPGA/SoCへの最適実装技術で、プライバシーを配慮した人物特定・追跡できるスマホソリューション

- 「**人物検知AI(refinedet)**」：人物を**検出** ⇒「**モザイク処理**」>プライバシー配慮
- 「**人物認識AI(personreid-res18)**」：人物を**認識** ⇒「**ナンバリング**」>一定範囲内での行動追跡が可能

※今回使用した「モデル」はカメラ範囲内における認識です。フレームアウトすると、新たな固有番号が付与されます



#1 映像取得 ⇒ #2 AI推論(人物検出、人物認識)マルチ実装 ⇒ #3 画像処理(重畳)
⇒ #4 Wi-Fiでの画像伝送 一連の処理を、SOM搭載のKria KV260ボードに実装しました

1. デモのポイントと概要

<ポイント>

- AMD社の最新SOM “Kria™ K24 SOM” に最適実装
- ロボットハンドの制御処理を、ROS2で実行
- エッジ搭載を想定した軽量なCNN(畳み込みニューラルネットワーク)を初めて自社開発



Kria KD240

<デモ概要>

お客様とじゃんけん勝負をし、ロボットハンドが必ず勝つデモ

入力：USBカメラの前で、ハンドサイン「グー・チョキ・パー」の画像を取得

処理：KD240にて、AI推論「ハンドサイン分類」を実行

出力：ロボットハンドをお客様に勝つ手になるよう制御

2. デモ詳細

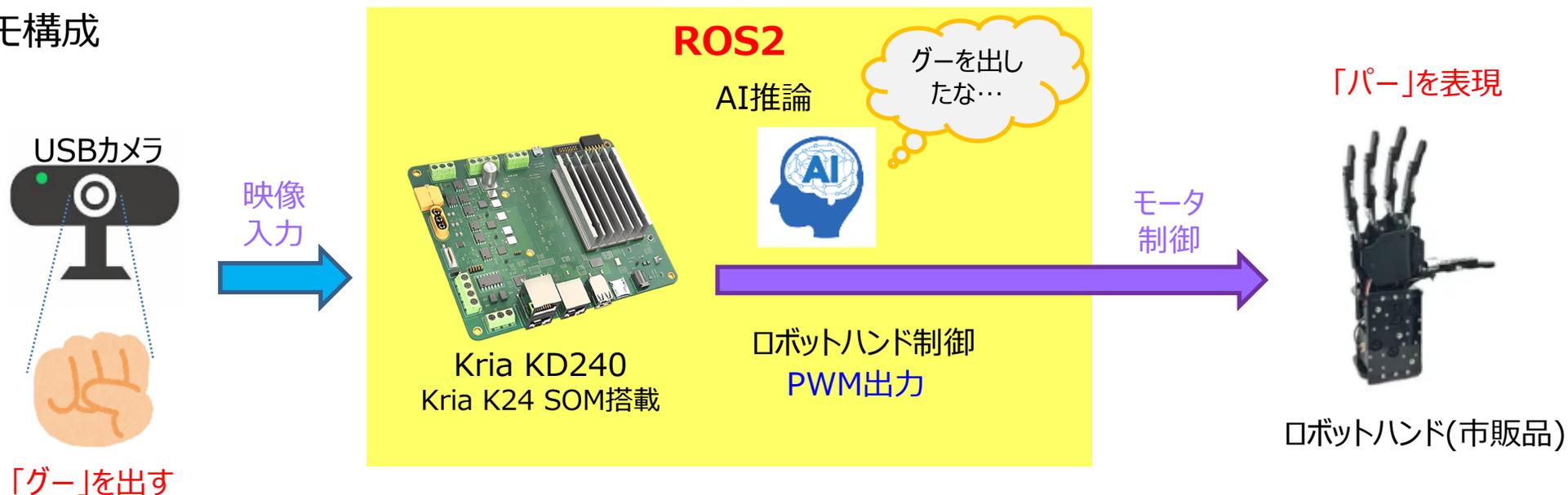
■ デモ内容

お客様と「じゃんけん」勝負するシンプルなデモです。

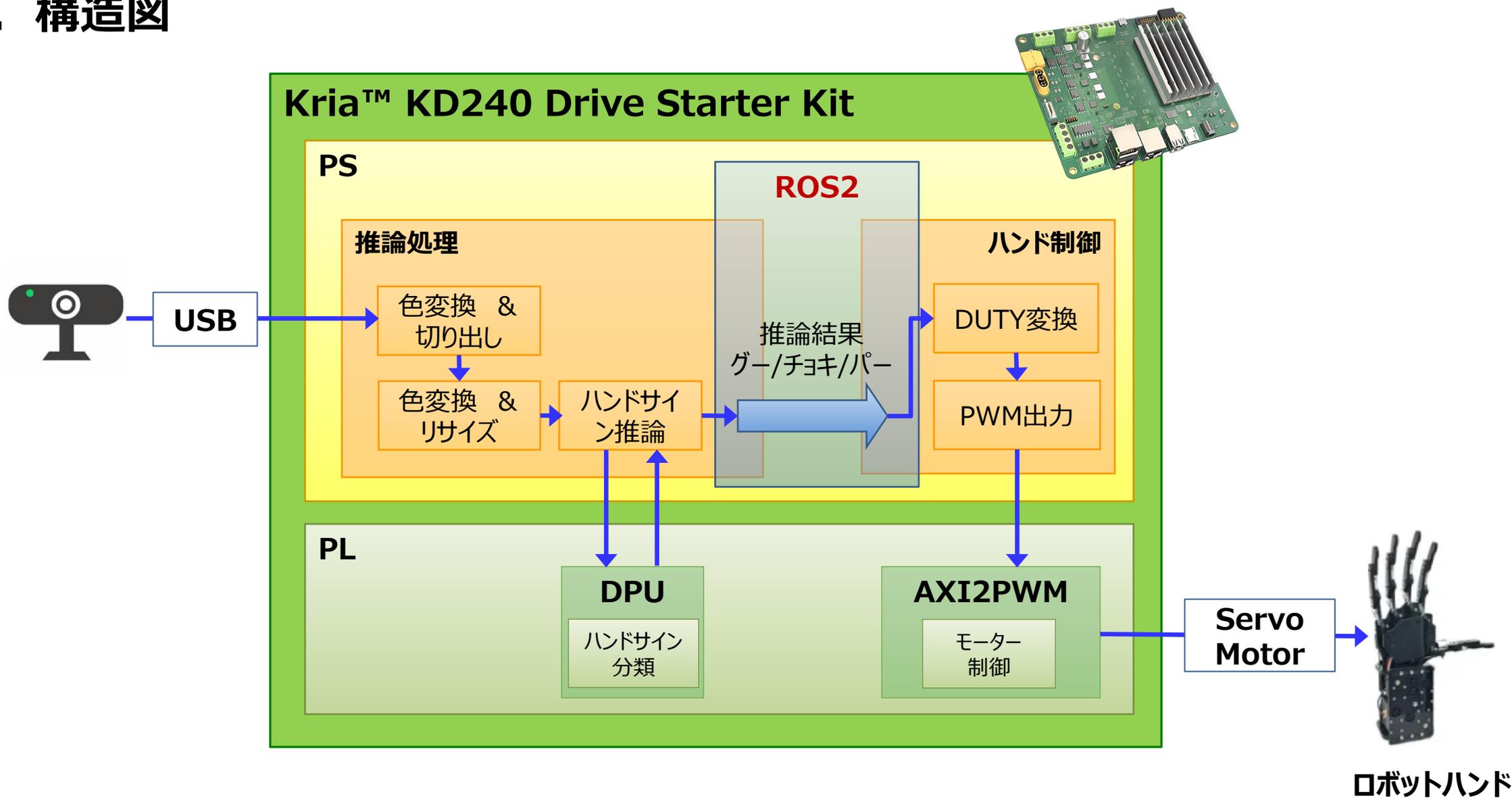
AMD社の最新SOM(System On Module)「Kria™ K24 SoM」搭載の「KD240 Drive Starter Kit」に実装

- 入力：USBカメラの前で、「グー、チョキ、パー」を出していただきます。
- 処理：ROS2(Robot Operating System)下で、AI推論処理部とロボットハンド制御部アプリケーションを分割
入力した前記画像の「ハンドサイン分類」を実行し、ロボットハンド制御するためのPWM信号を出力
- 出力：処理結果から、お客様に勝てるよう「ロボットハンド」に実装されているサーボモータを制御
⇒下記例では、映像入力が「グー」なので、勝つために「ロボットハンド」を「パー」の形にします。

■ デモ構成



3. 構造図



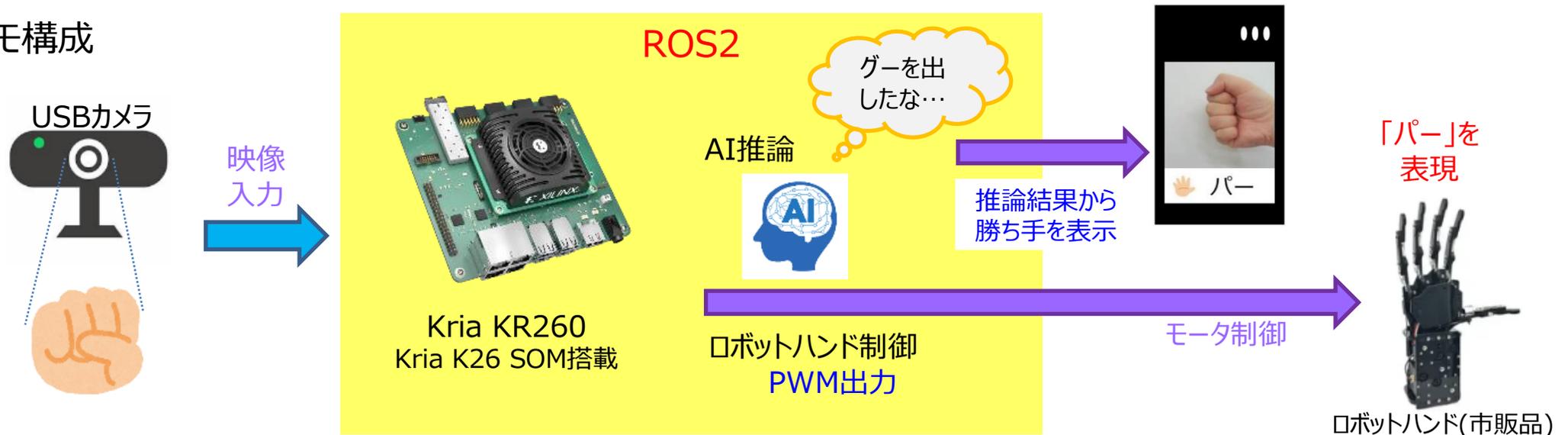
Kria™ K26 SOM へ実装の場合

■ デモ内容

AMD社のSOM(System On Module)「Kria™ K26 SoM」搭載の「KR260 Robotics Starter Kit」に実装

- 入力：USBカメラの前で、「グー、チョキ、パー」を出していただきます。
- 処理：ROS2(Robot Operating System)下で、AI推論処理部とロボットハンド制御部アプリケーションを分割
入力した前記画像の「ハンドサイン分類」を実行し、ロボットハンド制御するためのPWM信号を出力
- 出力：前記処理を受けて、下記2つの出力を出しています。
 - 出力1：「ハンドサイン分類」した結果をモニターに出力
⇒お客様が出した手(「グー、チョキ、パー」)をAI推論、つまりAIで予想した結果から勝ち手を表示
 - 出力2：処理結果から、お客様に勝てるよう「ロボットハンド」に実装されているサーボモータを制御
⇒下記例では、映像入力が「グー」なので、勝つために「ロボットハンド」を「パー」の形にします。

■ デモ構成



4. 技術のポイント

■ 技術のポイント

● ROS2に対応

ROS2(Robot Operating System)下で、AI推論処理部とロボットハンド制御部アプリケーションを分割。ROS2に則って構成されることで、他のROS2対応アプリケーションとの構成が可能に。本例では、ロボットハンドの応答性が格段に向上する効果を確認しました。

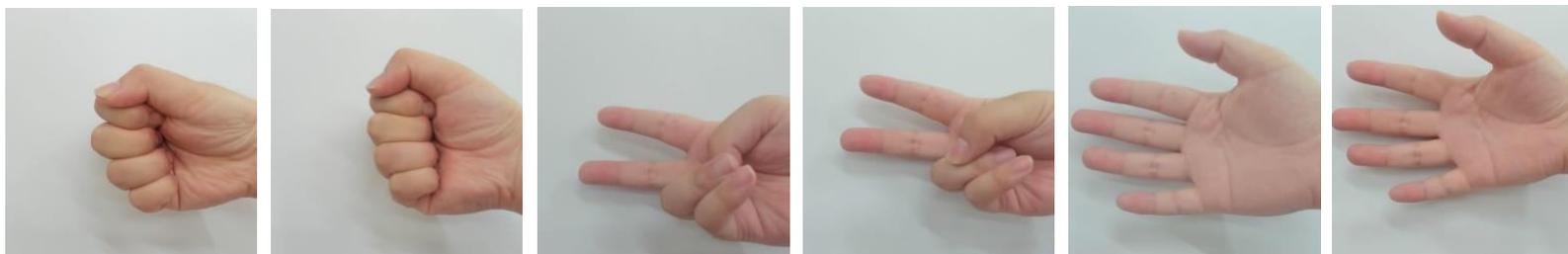
● Kria™ K26 SOM で先行開発し、Kria™ K24 SOMに移植

本デモ開発を短期間で行うために、既存モデルの「Kria K26 SOM」搭載の「KR 260 スターターキット」(右写真)で先行開発を実施しました。互換性があることも確認できましたので、お客様の選択肢として提案できます。

- エッジ搭載を想定した軽量な畳み込みニューラルネットワーク(CNN)をOKIアイディエスで開発（当社としては初）
学習用にグーチョキパーの画像を各200枚作成し、AI学習を社内サーバで実行
AIモデルの推論速度評価では、1127FPSを記録

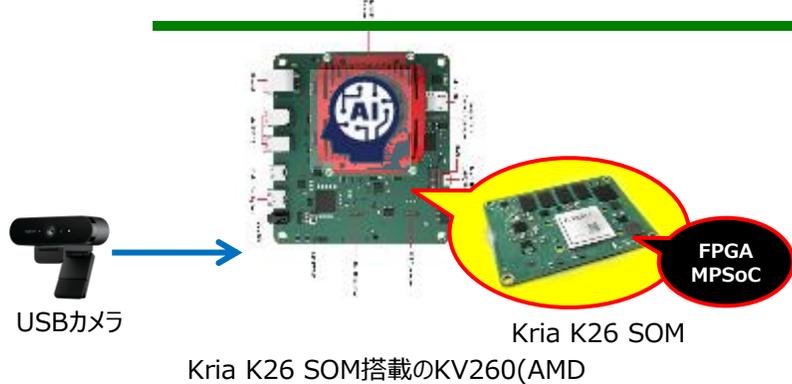


Kria KR260
Kria K26 SoM搭載



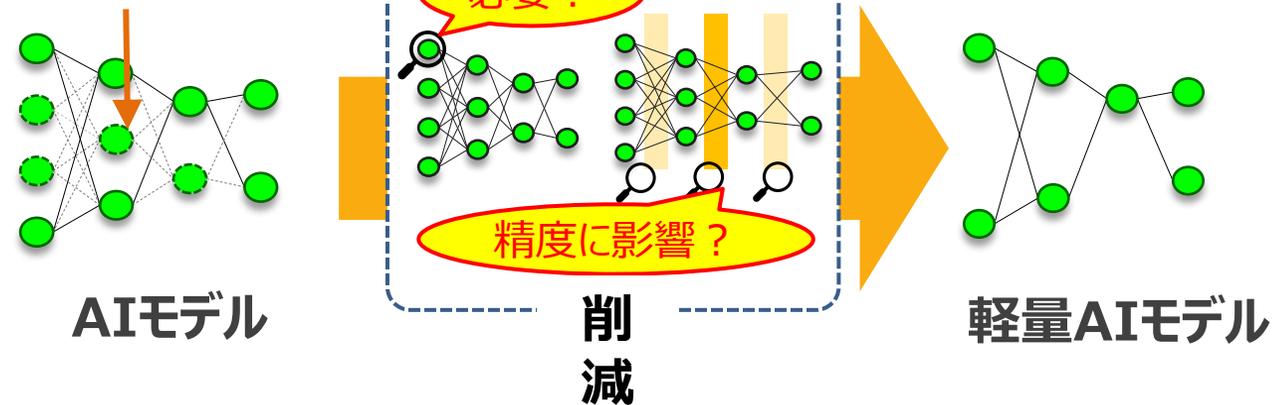
複数AIをエッジAIデバイスに実装。そのAIモデルを「PCAS」で軽量化。

- 複数AIを実装「**顔検知AI**」：人物の顔を**検出**。「**年齢分類AI**」：人物の年齢を**分類**
- OKI独自のAIモデル軽量化技術「PCAS」でAIモデルを軽量化



OKI独自のAIモデル軽量化技術「PCAS」を使って認識精度を最大限維持したままAIモデルの軽量化を実施

削減対象のニューロン



#1 映像取得

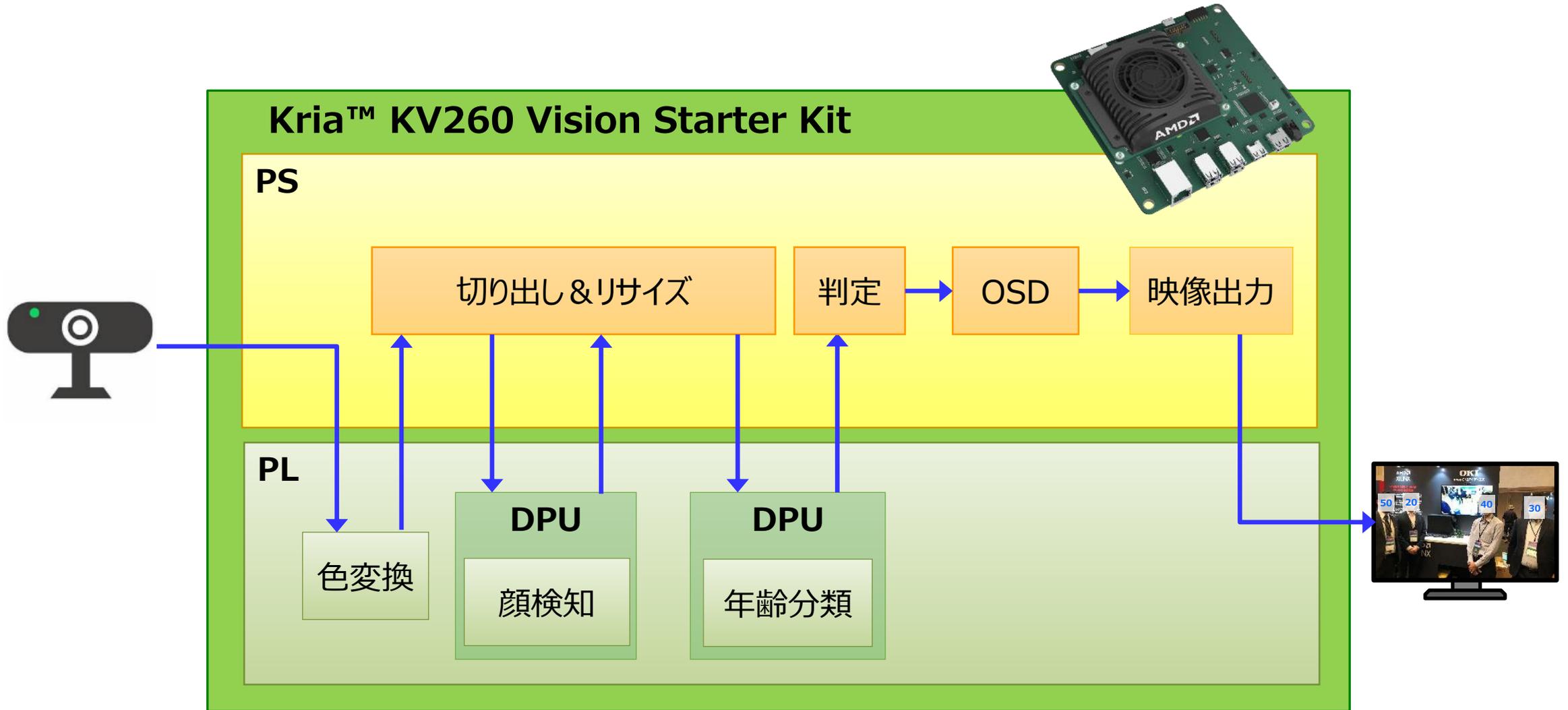


AI推論: 顔検出、年齢分類
画像処理: 顔部分に年齢表示



オリジナルモデルと、「PCAS」で軽量化を行ったモデルでの動作の差を体感ください

構造図



Kria™ 適応型システムオンモジュールの紹介

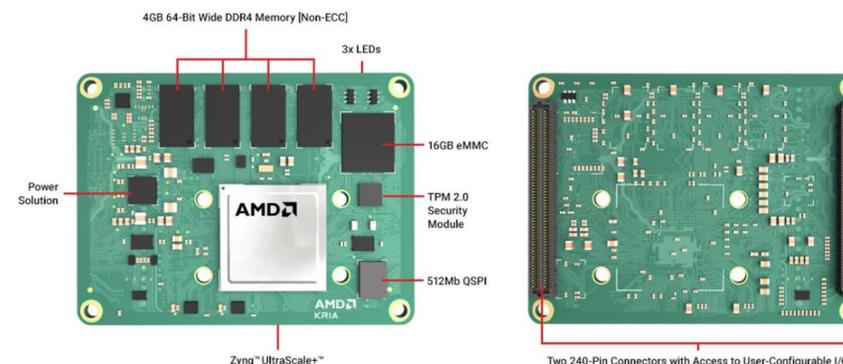
Kria™とは

- AMD社から登場したSOM(System on Module)シリーズ
- Zynq UltraScale+ MPSoC搭載
- 開発に使用したKria™ SOMをそのまま量産品に搭載することが可能



Kria™ K26 SOMの特徴

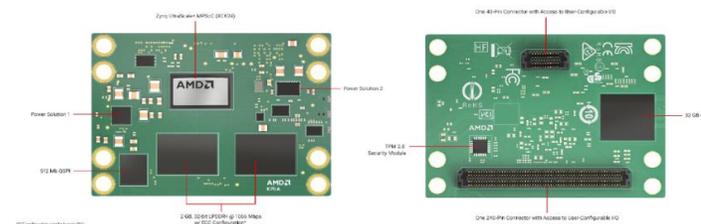
- ビジョンAIアプリケーション向け
- クレジットカード程度の小型サイズ (77 mm x 60 mm x 11 mm)



Kria™ K26 SOM

Kria™ K24 SOMの特徴

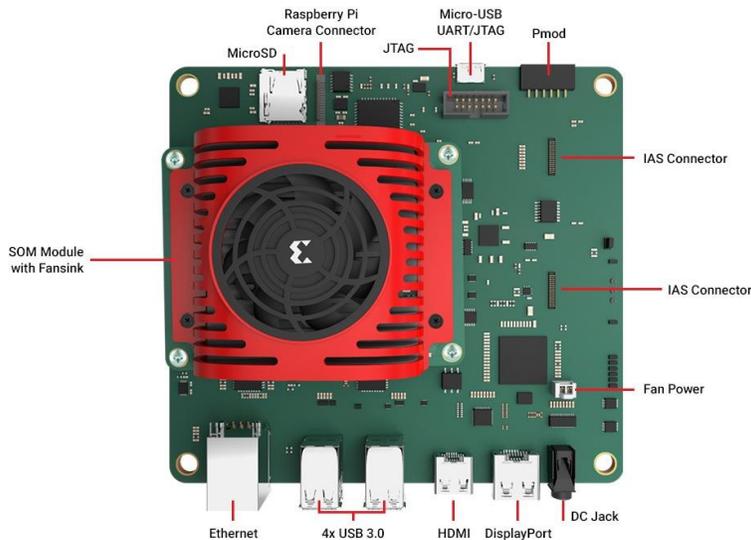
- モーター制御や DSP アプリケーション向け
- クレジットカードの半分程度の小型サイズ (60 mm x 42 mm x 11 mm)
- Kria™ K26 SOMと互換性あり



Kria™ K24 SOM

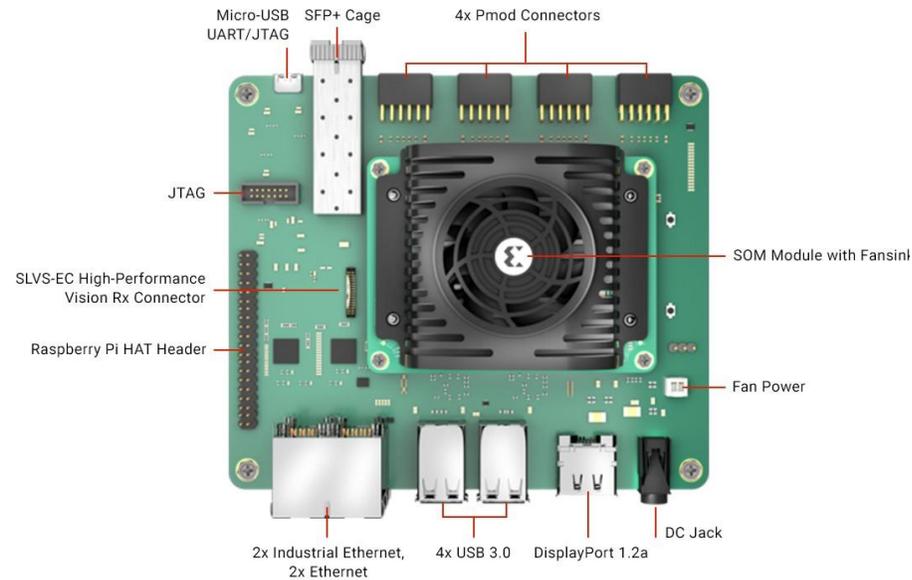
Kria™ スターターキットの紹介

KV260 ビジョンAI スターターキット



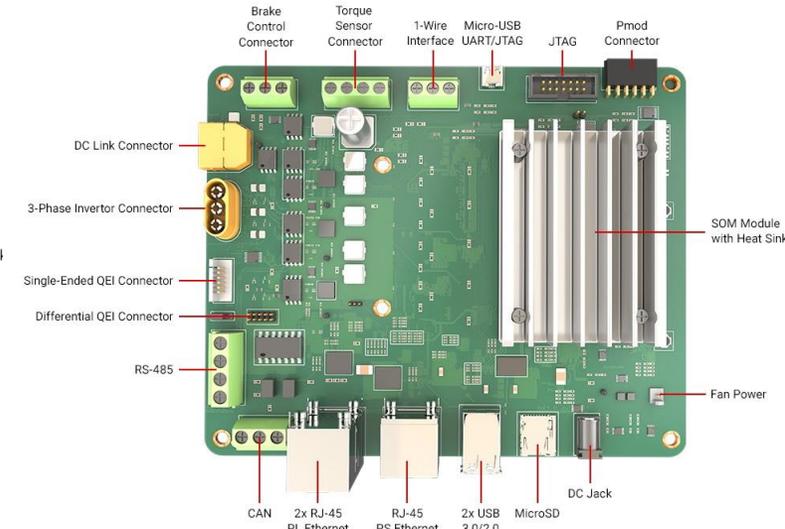
- ビジョンAIアプリケーションの開発向け
- LAN, USBコネクタやHDMI出力を搭載

KR260 ロボティクス スターターキット



- ロボティクスおよび産業機器アプリケーション開発向け
- LAN, USBコネクタやHDMI出力、SFP+ケージなどを搭載
- ROS2に対応

KD240 ドライブ スターターキット



- モーター制御や DSP アプリケーション向け
- LAN, USBコネクタや 3相インバーター用のコネクタ、イーサネットコネクタなどを搭載



Open up your dreams