



# クラウド・ネットワークの高速化と AIサービスを広げていきましょう！

2019年11月29日

インテル株式会社  
コミュニケーションサービスプロバイダーセグメント  
アカウントソリューションマネージャー  
小鮎 亮介

# 法務情報および免責条項

- © 2019 Intel Corporation. 無断での引用、転載を禁じます。
- Intel、インテル、Intel ロゴ、Intel Inside ロゴ、Intel Atom、Xeon、Agilex、eASIC、Intel Nervana、Intel Optane、Movidius、Quartus は、アメリカ合衆国および / またはその他の国における Intel Corporation またはその子会社の商標です。
- \* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。
- インテル® テクノロジーの機能と利点はシステム構成によって異なり、対応するハードウェアやソフトウェア、またはサービスの有効化が必要となる場合があります。詳細については、各システムメーカーまたは販売店にお問い合わせいただくか、<http://www.intel.co.jp/> を参照してください。
- テストでは、特定のシステムでの個々のテストにおけるコンポーネントの性能を文書化しています。ハードウェア、ソフトウェア、システム構成などの違いにより、実際の性能は掲載された性能テストや評価とは異なる場合があります。購入を検討される場合は、ほかの情報も参考にして、パフォーマンスを総合的に評価することをお勧めします。性能やベンチマーク結果について、さらに詳しい情報をお知りになりたい場合は、<http://www.intel.com/benchmarks/> (英語) を参照してください。
- 記載されているコスト削減シナリオは、指定の状況と構成で、特定のインテル® プロセッサ搭載製品が今後のコストに及ぼす影響と、その製品によって実現される可能性のあるコスト削減の例を示すことを目的としています。状況はさまざまであると考えられます。インテルは、いかなるコストもコスト削減も保証いたしません。
- 本資料には、開発中の製品、サービス、プロセスに関する情報が記載されています。ここに記載されているすべての情報は、予告なく変更されることがあります。インテルの最新の予測、日程、製品仕様およびロードマップをご希望の方は、インテルの担当者までお問い合わせください。
- 本資料は、(明示されているか否かにかかわらず、また禁反言によるとよらずにかかわらず) いかなる知的財産権のライセンスも許諾するものではありません。
- インテルは、本資料で参照しているサードパーティーのベンチマーク・データまたは ウェブサイトについて管理や監査を行っていません。本資料で参照している ウェブサイトにアクセスし、本資料で参照しているデータが正確かどうかを確認してください。
- 絶対的なセキュリティを提供できるコンピューター・システムはありません。



# 業界のメガトレンド 『データ・セントリック社会』

クラウド・  
コンピューティング  
の普及拡大

AIと  
アナリティクス  
の発展

ネットワークと  
エッジ  
のクラウド化



# 業界のメガトレンド 『データ・セントリック社会』

クラウド・  
コンピューティング  
の普及拡大

AIと  
アナリティクス  
の発展

ネットワークと  
エッジ  
のクラウド化

- クラウド・ネットワークキングの高速化
- AIサービスを広げていきましょう!



# ネットワークのクラウド化に貢献

データセンター | クラウド

コア

アクセス | エッジ

デバイス | モノ



## ネットワークのIA化

2011

NFV  
を定義

2013

最初のNFV  
PoC(概念実証)

2015

20%  
の通信事業者が  
NFVを採用

2017

インテル®  
DPDK  
Linux\*  
Foundationに移管

2018

65%  
ネットワーク仮想化

2019

最初の完全な  
クラウド・ネイティブ  
ネットワーク



# 『データ・セントリック社会』向け インテル最新製品ポートフォリオ

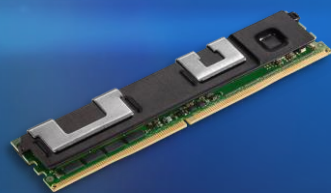
## 高速な移動

インテル®  
イーサネット  
800 シリーズ



## より多くを保存

インテル®  
Optane™ DC  
パーシステント・  
メモリー



インテル®  
SSD  
データセンター



第2世代  
インテル® Xeon®  
スケーラブル・  
プロセッサ



## あらゆるものを処理

インテル®  
Xeon®  
D-1600  
プロセッサ



インテル®  
Agilex™ FPGA



# 2019年4月発表



# 第2世代 インテル® Xeon® スケーラブル・プロセッサ

(Code name: Cascade Lake)

## プラットフォームの更なる最適化

- CPU間通信改善(Intel® Mesh Architecture<sup>(\*1)</sup>), CPUキャッシュの更なる最適化,メモリ通信幅の拡大など

(\*1) Intel® Mesh Architectureは、前世代Skylakeより対応

## 利用用途に応じた推奨モデル化

- NFV, 検索機能など利用用途に応じた推奨モデル(SKU) およびエコシステムソリューションを確立

## ソフトウェアでのコア数・クロック数の変更

- 特的モデル<sup>(\*2)</sup>では、コア数・クロック数変更がソフトウェアで可能(Intel® Speed Select Technology)

(\*2) YモデルCPU等



- DIMM型の大容量メモリーキャパシティ構成を実現。

## 暗号化通信, AI処理などの高速化

- Intel QAT<sup>(\*3)</sup>での暗号化通信, Intel® deep learning boostでのAI処理のアクセラレーションなど

(\*3) Intel® QuickAssist Technologyは前世代Skylakeより対応し、現世代Cascade Lakeで対応が拡大

## セキュリティー機能の強化

- いくつかの主要なセキュリティー・サイド・チャンネルを、ハードウェア・レベルで対応

# 対前世代比最大1.33倍のパフォーマンス向上 (平均)



# ハードウェア・オフロード選択肢

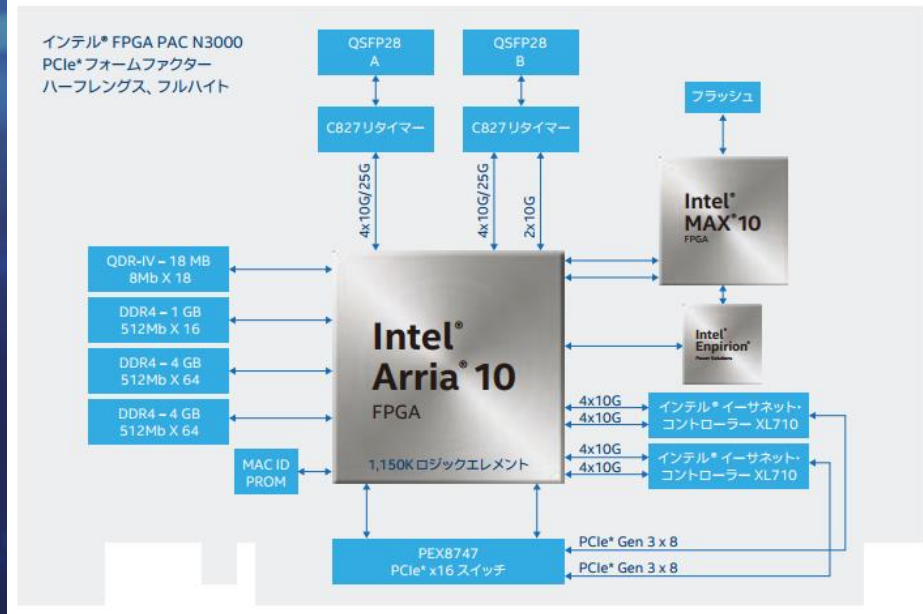
## インテル® FPGA PAC N3000



通常CPUで処理する仮想ネットワーク処理を、FPGA NICカード (プログラム可能なネットワークカード) で専用に対応させることで、5G通信需要を効率的に処理します。

### 適用例

- オープン・バーチャル・スイッチ (OVS)
- 階層型 QoS (H-QoS)
- インターネット・プロトコル・セキュリティ (IPSec)
- IPv6セグメント・ルーティング (SRv6)
- ベクトルパケット処理 (VPP)
- 仮想化無線アクセス・ネットワーク (vRAN)





# ネットワーク/エッジの強化

**BAREFOOT**

「ホワイトボックス  
スイッチ」向け技術

Growing Ecosystem



+

Barefoot Network Analytics



+

Barefoot SDE/Compiler



+

Barefoot Hardware



=

New Network  
Functions

Differentiation

Network Analytics

Rapid  
Innovation

## 1. ディスアグリゲート構成

- 目的・用途に応じた、豊富なハードウェア・ソフトウェアの組合せ

## 2. プログラマブル

- P4言語による、機能追加・カスタマイズに速やかに対応

## 3. 高パフォーマンス

- 必要機能のみをPipeline処理させることで、リソースを最大限活用

# エッジマイクロサービス提供における オーケストレーションツール



An **open source** software toolkit to enable **easy orchestration & management** of edge services **across diverse** network platform and access technologies in **multi-cloud environments**

## ABSTRACT NW & PLATFORM COMPLEXITY

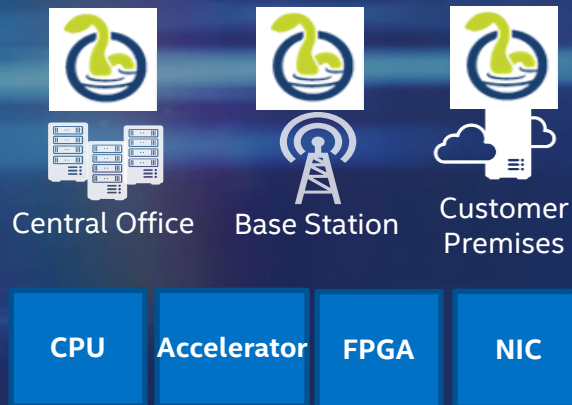
## ONBOARD/ORCHESTRATE APPS & SERVICES

## ACCELERATE SERVICES INNOVATION

### IOT Devices



### OPENNESS Edge Node SW



### OPENNESS Edge Controller SW



### Microservices APIs

- Application Agent
- Dataplane agent
- Lifecycle agent
- Virtualization agent
- Core NW Config Agent
- Controller API
- Edge auth agent

### App Ecosystem Frameworks Support



End to End Services Infrastructure

\*Other names and brands may be claimed as the property of others



# 業界のメガトレンド 『データ・セントリック社会』

クラウド・  
コンピューティング  
の普及拡大

AIと  
アナリティクス  
の発展

ネットワークと  
エッジ  
のクラウド化

- クラウド・ネットワークキングの高速化
- AIサービスを広げていきましょう!





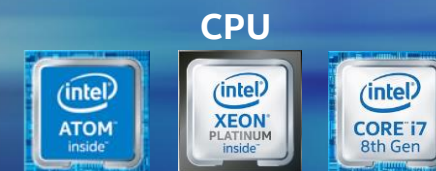
# エッジとAIへの注力: Why INTEL

インテルでは、エッジとAI処理を既存のコンピューティング処理の延長と捉え、エンドtoエンドでのソリューションを提供させて頂いています。

1. OpenVINOツールキットによる、インテル製品 (CPU, I-GPU, VPU, FPGA)各世代における、AI推論の最適化処理 (**Wright once, Run anywhere**)



2. 各種シリコンハードウェア製品での最適化処理



AI アクセラレーター



3. 様々なエッジ・ユースケースの開発と公開

(vBNG, vCDN, vRAN, OpenNESS等)



(無償)



# AIライフサイクル

## ~モデルの構築 / モデルの運用~

Intel works with customers across the entire AI lifecycle

### モデルの構築

(学習メイン、推論少々)

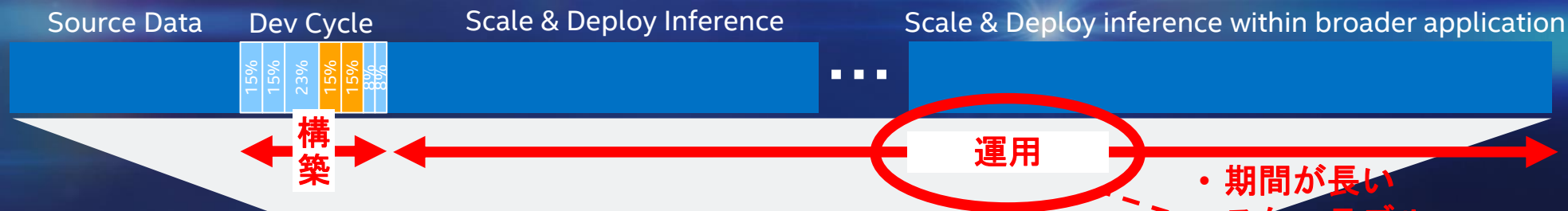
目標の精度を目指し、限られた期間内にひたすらモデルを作り続けるフェーズ

### モデルの運用

(推論メイン、学習少々)

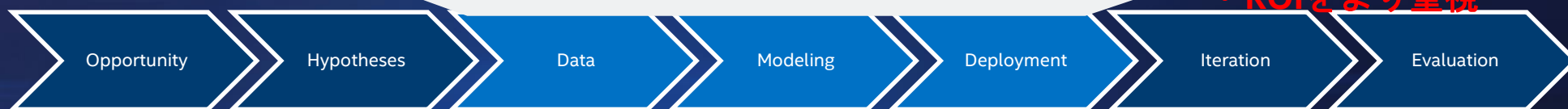
モデルを本番環境へ展開し新規ビジネスデータをひたすら推論するフェーズ

BUILD, DEPLOY  
& SCALE



- 期間が長い
- スケーラブル
- ROIをより重視

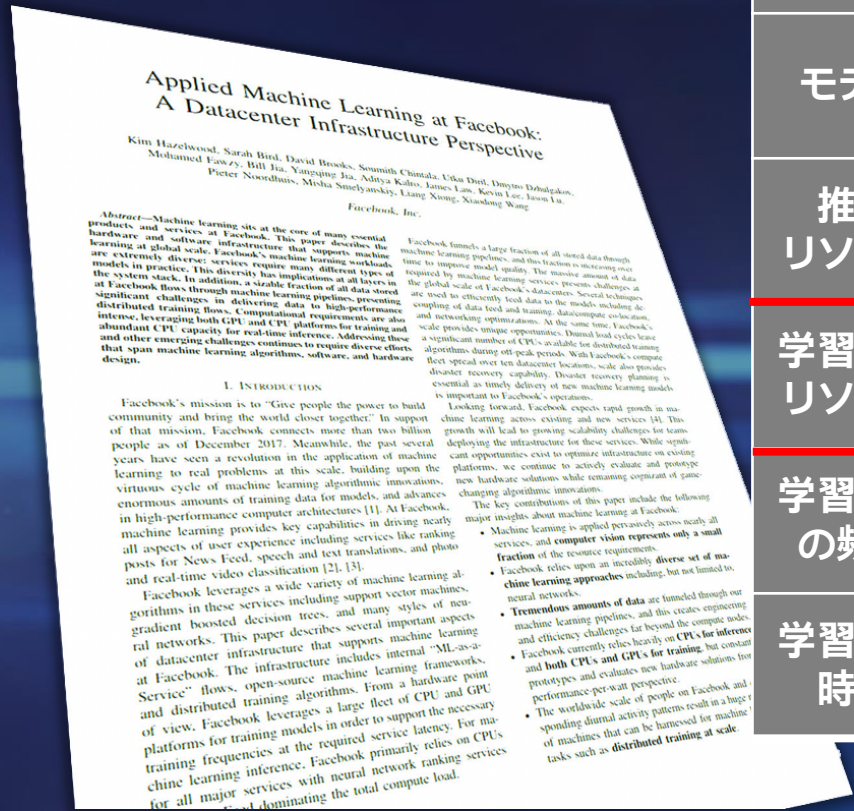
TIME-TO-  
SOLUTION



# CPU中心での大規模なDL

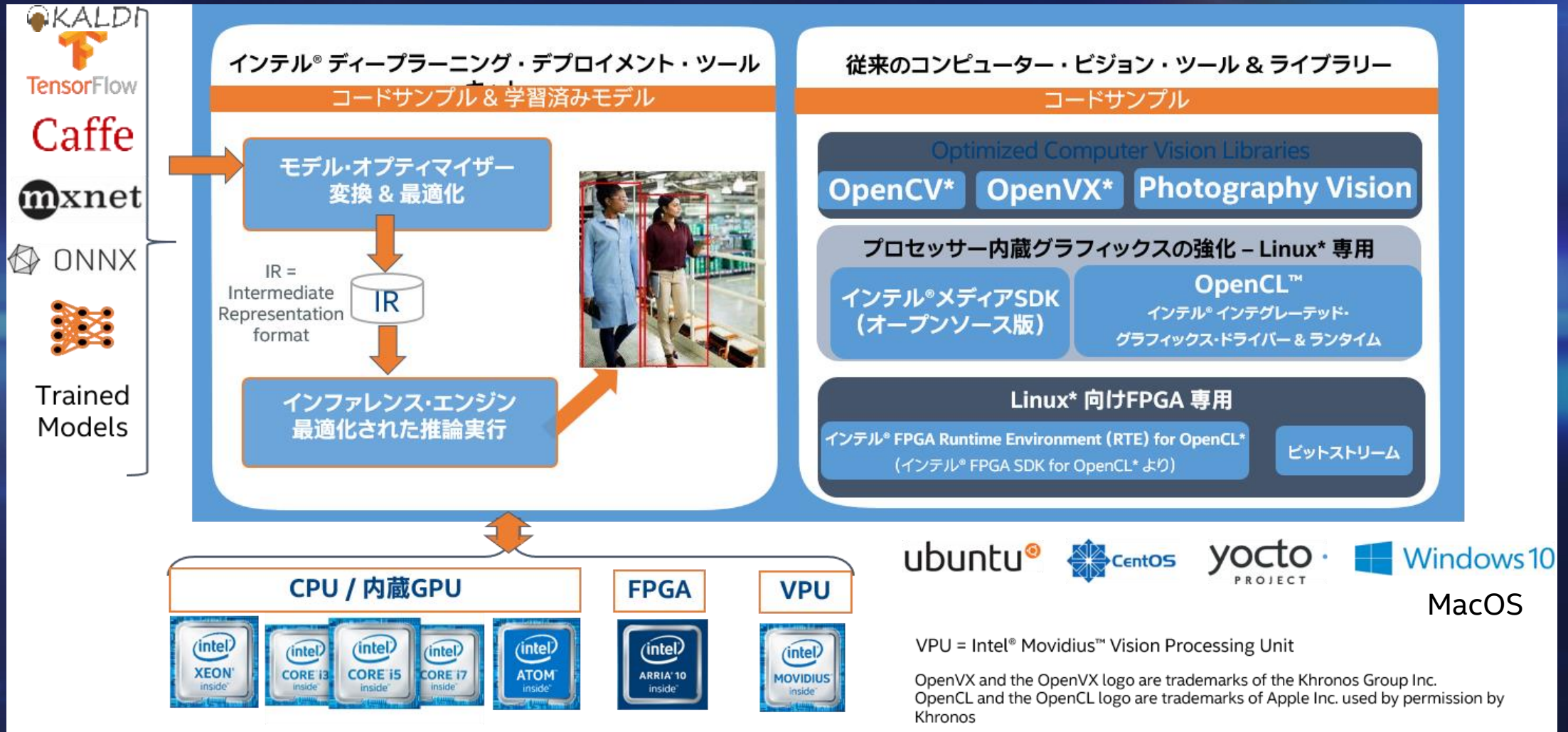
出典: [research.fb.com/wpcontent/uploads/2017/12/hpca-2018-facebook.pdf](https://research.fb.com/wpcontent/uploads/2017/12/hpca-2018-facebook.pdf) (英語)

サービス	ランキング・アルゴリズム	写真のタグ付け	写真のテキスト生成	検索	言語の翻訳	迷惑メールのフラグ設定	音声
モデル	MLP	SVM、CNN	CNN	MLP	RNN	GBDT	RNN
推論リソース	CPU	CPU	CPU	CPU	CPU	CPU	CPU
学習処理リソース	CPU	GPUとCPU	GPU	状況によって異なる	GPU	CPU	GPU
学習処理の頻度	毎日	写真N枚ごと	数カ月ごと	毎時間	毎週	毎日未満	毎週
学習処理時間	数時間	数秒	数時間	2~3時間	数日	2~3時間	数時間



多くのクラウドユーザーが、ディープラーニングにCPUを広く使用

共通のコード・汎用的なアルゴリズムをCPU, GPU, AIアクセラレーターで共有可能



# AIアクセラレーター製品

インテル® ビジョン・アクセラレーター・デザイン・プロダクト  
(インテル® Movidius™ VPU搭載版)

インテル® ビジョン・アクセラ  
レーター・デザイン・プロダクト  
(インテル® Arria® 10 FPGA 搭  
載版)

インテル® ビジョン・  
アクセラレーター・  
デザインに基づく  
アドインカード(例)



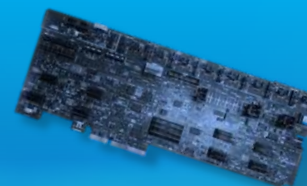
インテル® Movidius  
MA2485 VPU  
x 1

M.2, Key E



インテル® Movidius™  
MA2485 VPUs  
x2

miniPCIe\*



インテル® Movidius™  
MA2485 VPUs  
x8

PCIe\* x4



インテル® Arria® 10 FPGA  
1150GX/660GX

PCIe\* x8

インターフェイス

ソフトウェア  
開発ツール

OpenVINO™ ツールキット

エッジで実行するNN推論モデルの構築; インテル® CPU, GPU, VPU, FPGAで共通の開発環境



# OpenVINO™ツールキットに含まれる インテルの学習済みモデルの例(無償)



### 性別・年齢

IMAGE INPUT	RESULT
	Female, 18.97
	Male, 26.52
	Male, 33.41

### 顔認識



### 人物検出



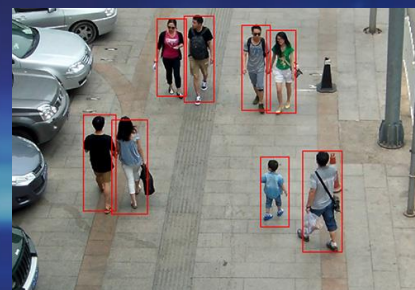
### 行動検出



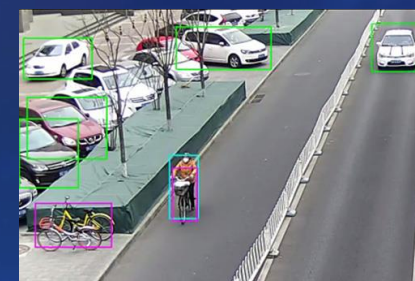
### テキスト認識



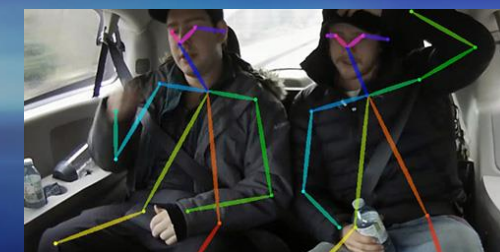
### 顔特徴点検出、 感情認識



### 人・車両検出



### 人のポーズ推論



### 超解像



オープンな開発者コミュニティで公開されているモデルに加えて、インテルが開発した学習済みモデル 50 種類以上を無償で提供

<https://software.intel.com/en-us/openvino-toolkit/documentation/pretrained-models>



# まとめ:クラウド・ネットワーキングの高速化、 AIサービスを広げていきましょう!



(無償)

OpenVINO™

OpenVINOとCPU, VPU, FPGAでのAI推論最適化処理、エッジソリューションの開発  
AI推論高速処理とエッジソリューションへの注力

OpenVINO セミナー

検索

FPGA PAC N3000



クラウド・ネットワーキングの高速化  
最新コンピューティング技術での大規模データ利活用

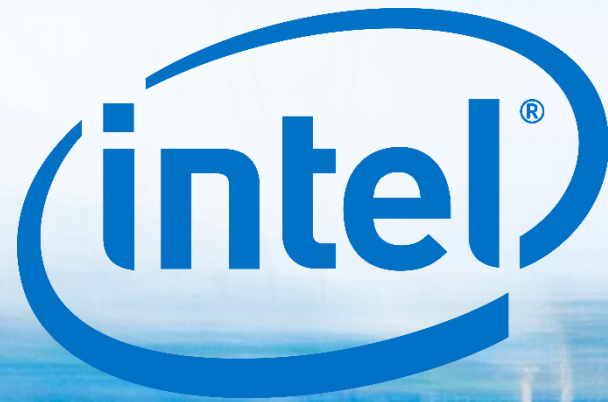


インテル® Optane™ テクノロジー  
メモリーおよびストレージのボトルネックを軽減



OEM&ISVエコシステム・パートナー様との連携  
データセンターからエッジまでの  
ソリューション提案





ご清聴ありがとうございました。