

# Qumicoを使ったEdge AIの最新情報。AutoML、LSTM対応

～ ColabやLabelboxを使ったリモート時代の組込AI開発、人材育成。

---



2020/08/21

株式会社パソナテック  
DX戦略本部 デジタルテクノロジーグループ  
Qumico Product Manager  
夏谷

# 目次

- 自己紹介
- はじめに
- エッジAIの動向
  - ONNX, AutoML
  - Qumicoの紹介、デモ
- これからのエッジAI開発
  - リモート開発の実例
- ワークショップの紹介
- 会社紹介

# | 自己紹介

# 自己紹介

株式会社パソナテック  
DX戦略本部 デジタルテクノロジーグループ  
Qumico Product Manager  
夏谷 実

FPGAやLSIの画像処理から、Deep Learningの世界に入る。プログラミングが好き  
一昨年よりQumicoのプロダクトマネージャに就任。  
大阪と東京と半々くらいの生活を続けていたが、コロナの影響により完全リモートワークに移行。



Hatena id: natsutan

2019/9/25  
<http://natsutan.hatenablog.com/>



Qiita id: natsutan

2019/9/25  
<https://qiita.com/natsutan>



TensorFlow Users Group KANSAI

2019/9/25  
<https://tfug-kansai.connpass.com/event/146442/>

はじめに

# はじめに

ハンズオン資料プレゼントについて

Google  
Colaboratory

Google  
Cloud  
AutoML

Qumico

Deep Learningの学習から、C言語の変換  
までをブラウザで完結



ラズパイ上で画像  
の識別Deep  
Learningが動く

今日の発表の最後にアンケートのURLを記載しています。

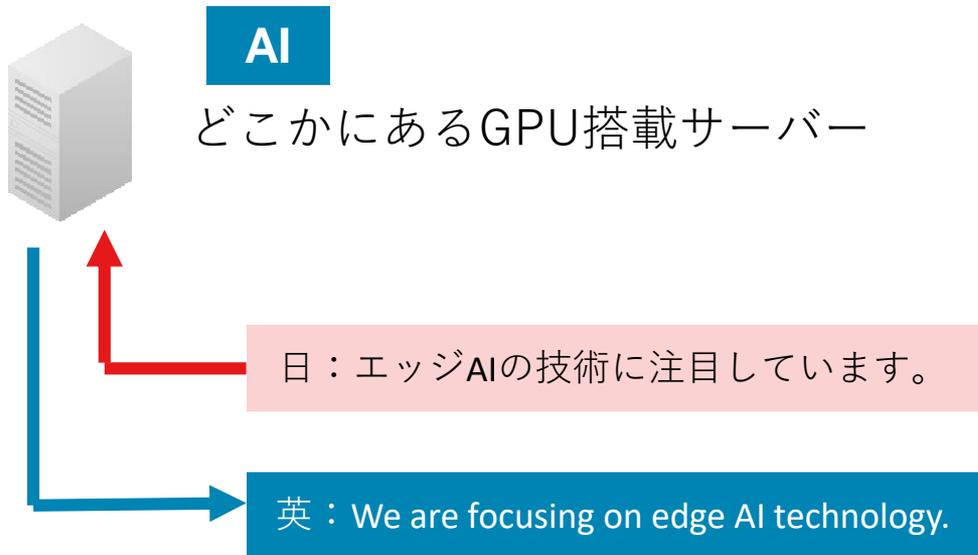
アンケートにご回答いただいた方に、もれなくハンズオン資料をプレゼントいたします。

# | エッジAIの動向

ONNX, AutoML

# エッジAIとは

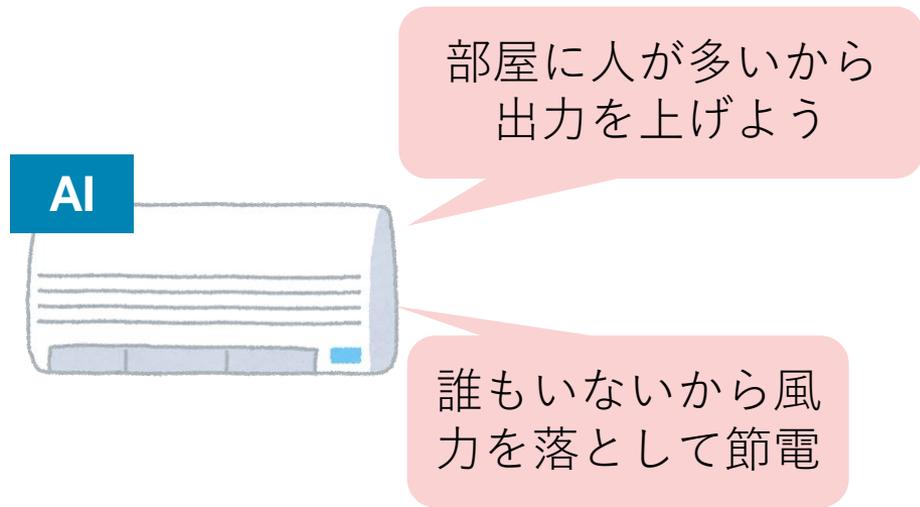
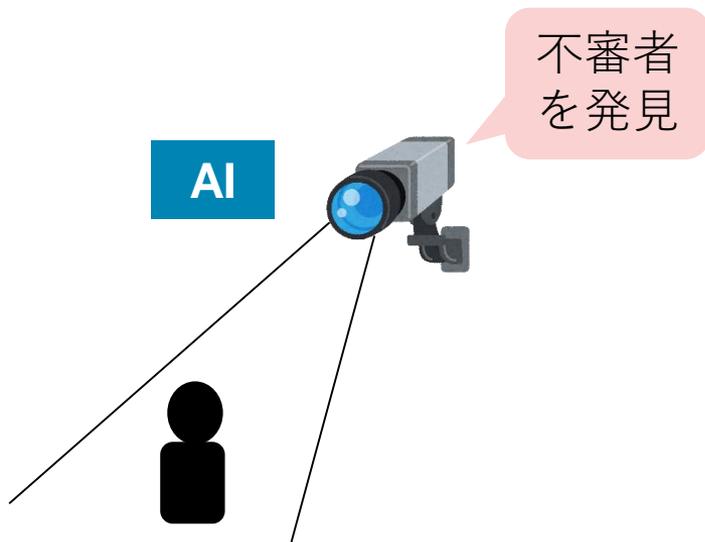
よくあるAIを使ったサービス例



- 同じようなアーキテクチャ
- ・ 画像の仕分け
  - ・ 商品の推薦
  - ・ 買い取り価格の見積もり
  - ・ 商品のカテゴリー分け

# エッジAIとは

エッジAIのイメージ



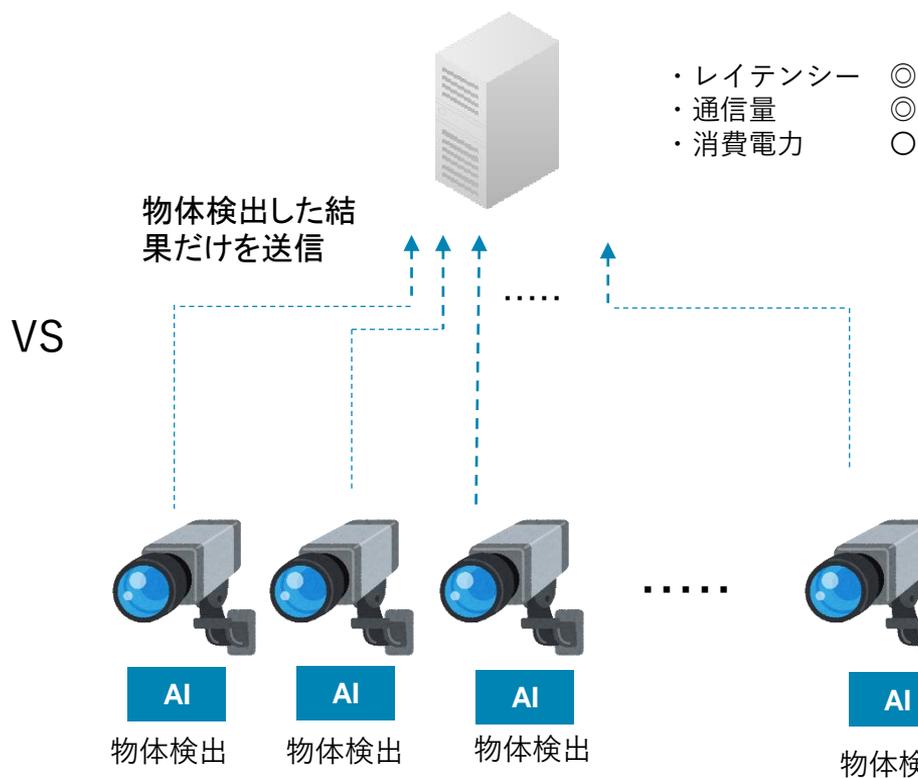
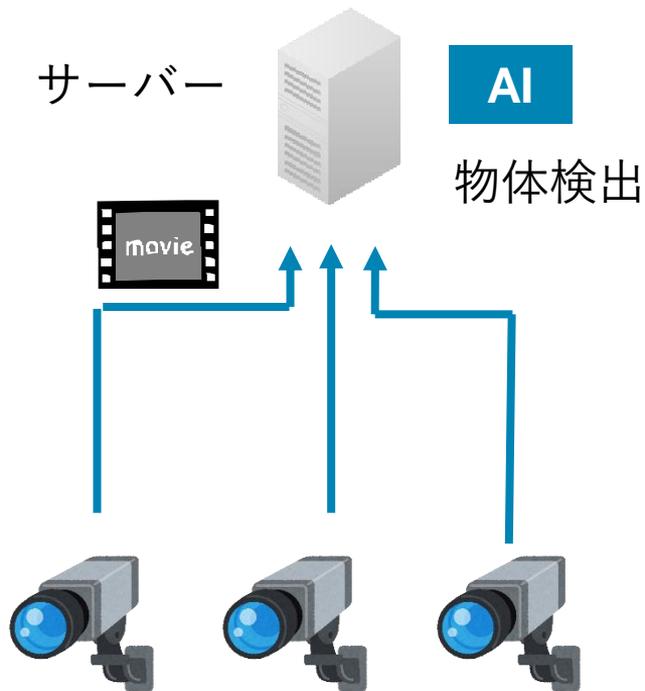
[https://www.irasutoya.com/2014/10/blog-post\\_185.html](https://www.irasutoya.com/2014/10/blog-post_185.html)  
[https://www.irasutoya.com/2013/01/blog-post\\_7571.html](https://www.irasutoya.com/2013/01/blog-post_7571.html)

AIを処理するサーバーが無く、それぞれの装置がAIを搭載する。

# なぜエッジAIが必要なのか

全ての映像をサーバーに送るアーキテクチャ

エッジAIを使ったアーキテクチャ



## エッジAIを使う理由

- ・ 自動運転等、応答速度が非常に重要なケース
- ・ コストが重要なケース
- ・ 通信環境が悪くても動く必要があるケース
- ・ プライバシーを守りたいケース

世の中にAIが普及するにつれ、エッジAIの重要性も増えていく

# 業界標準フォーマット ONNX

Open Neural Network Exchange (ONNX)

Microsoftの資料から引用



それぞれのフレームワークが、ONNXファイルの入出力をサポート

ONNXの実行（推論）環境  
ONNX runtime

2019/9/25

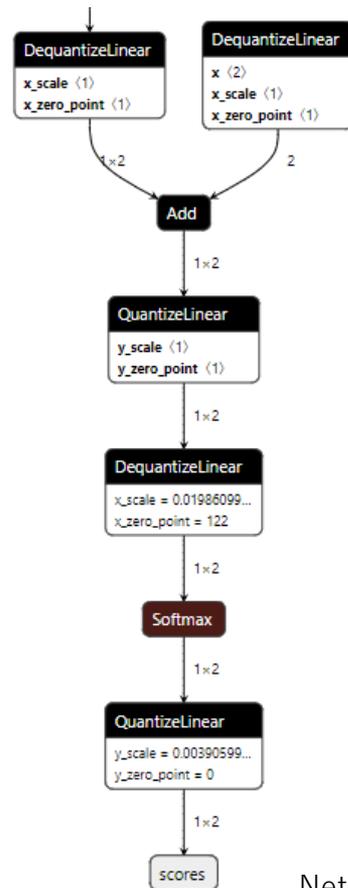
[https://static.sched.com/hosted\\_files/linaroconnectsandiego/83/ONNX%20%26%20ONNX%20Runtime%20-%20SAN19-211.pdf](https://static.sched.com/hosted_files/linaroconnectsandiego/83/ONNX%20%26%20ONNX%20Runtime%20-%20SAN19-211.pdf) <sup>12</sup>

# ONNXフォーマット

- Model
  - Version info
  - Metadata
  - Graph
- Graph
  - Inputs and Outputs
  - Graph name
  - Computational nodes
- Computational Graph
  - Operator
  - Operator Parameter
  - Inputs and Outputs

中身はProtocol Buffers

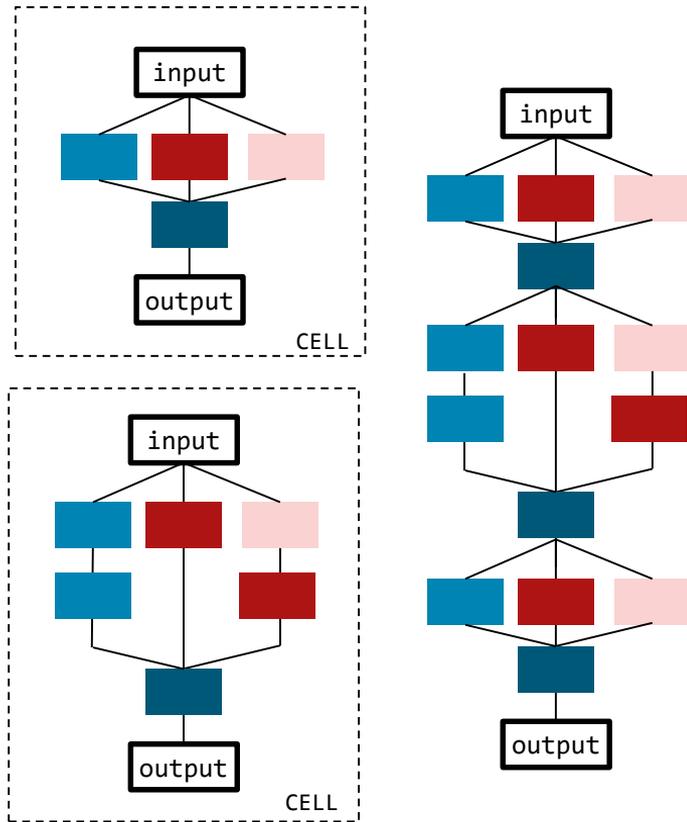
APIが用意されているPythonだけでなく Rust等でも操作可能



Netronで可視化

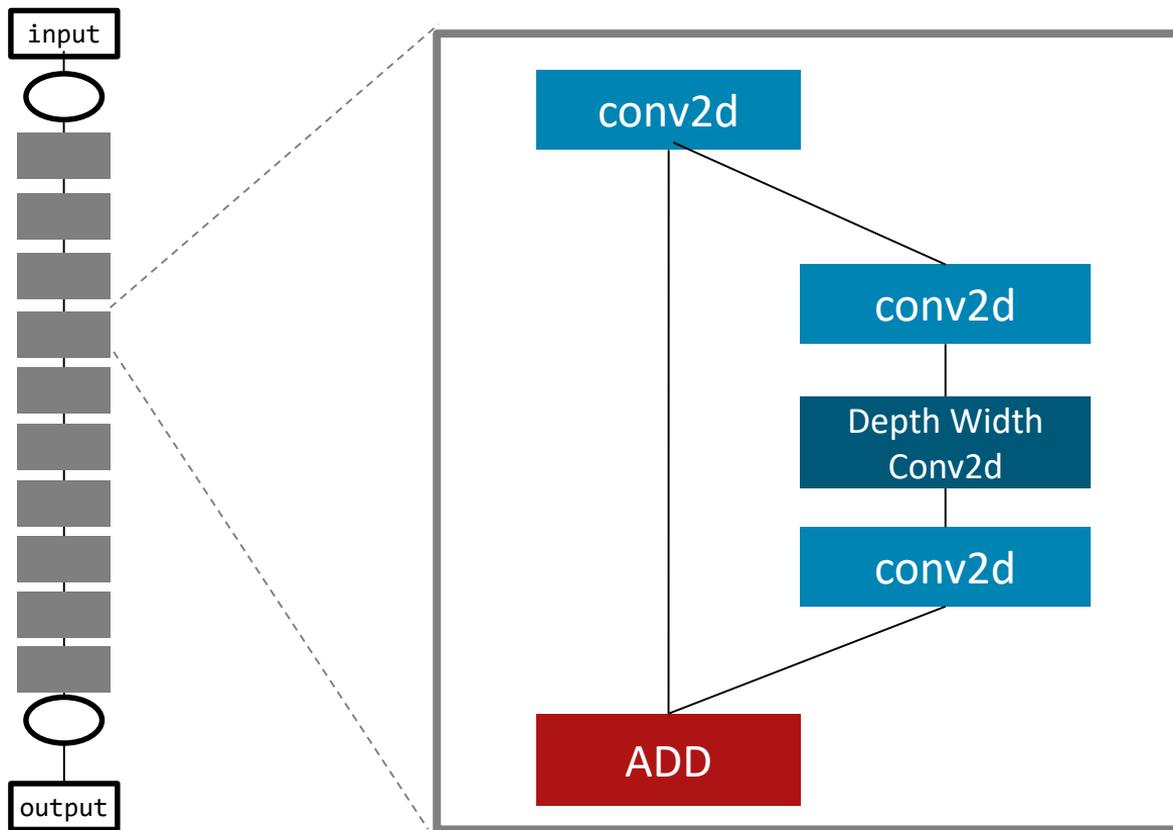
# Auto MLとは

- Automated Machine Learningの略
- 自動的に機械学習のハイパーパラメータを設定しパフォーマンスを上げる技術。NAS (Neural Architecture Search) と呼ばれるNNアーキテクチャの探索も含む。
- GoogleがCloud上でAuto MLを使えるようにしたのが「Cloud AutoML」と呼ばれるサービス。Auto MLと言った時にCloud AutoMLを指すことが多い。



NASの例

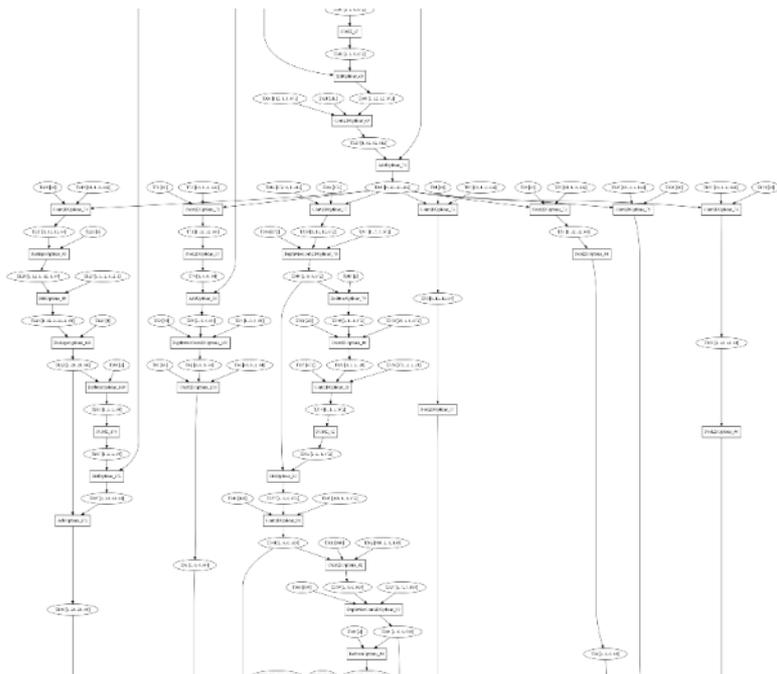
# AutoMLが出力するNN例（識別）



# AutoMLが出力するNN（物体検出）

2019-09-23

AutoML Vison 物体検出 edge版のニューラルネットワークを見てみた



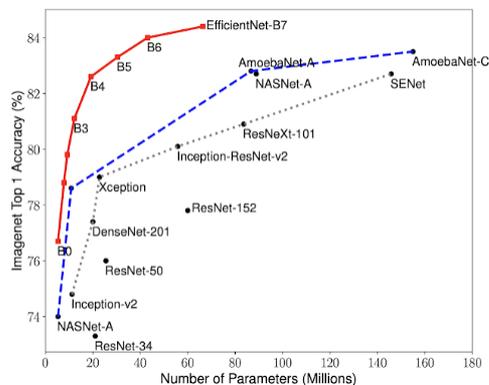
NAS (Neural Architecture Search) をしているようで、非常に複雑な構成になっています。

2019/10/4

<http://natsutan.hatenablog.com/entry/2019/09/23/125835><sub>16</sub>

# Could AutoMLの実力

## 識別



画像の識別に関しては、素人では絶対に勝てない

<https://developers-jp.googleblog.com/2019/07/efficientnet-automl-google.html>

## 物体検出

VOC\_20190515115012



Average precision ?

at 0.5 IoU

0.737

Precision\* ?

91.2%

Recall\* ?

50.15%

\* Using a score threshold of 0.501

⋮

だいたい6万円くらいで、2016年末の最先端 (YOLO v2)に近い数字

Cloud AutoMLが対応しているタスクならAutoMLを使いたい。

# Cloud AutoMLとAzure Custom Visionの比較

犬と猫の識別を、Google Cloud AutoMLと、マイクロソフトのAzure Custom Visionを使ってモデルを作成しました。画像データを100枚から2000枚まで増やして精度（accuracy）を比較。

| 枚数   | Google Cloud AutoML | Azure Custom Vision |
|------|---------------------|---------------------|
| 100  | 0.777               | 0.918               |
| 200  | 0.900               | 0.942               |
| 1000 | 0.947               | 0.914               |
| 2000 | 0.947               | 0.952               |

<https://natsutan.hatenablog.com/entry/2020/03/18/214112>

# AutoML

## PFN Optuna

Optuna はハイパーパラメータの最適化を自動化するためのソフトウェアフレームワークです。ハイパーパラメータの値に関する試行錯誤を自動的にを行いながら、優れた性能を発揮するハイパーパラメータの値を自動的に発見します。現在は Python で利用できます。

2019/9/30

<https://research.preferred.jp/2018/12/optuna-release/>

## Sony Prediction One

Prediction Oneは、データさえ用意すれば、数クリックで高度な予測分析を自動的に実行できるソフトウェアです。

2019/9/30

<https://predictionone.sony.biz/about/feature/>

## Amazon Forecast

Amazon Forecast は、Amazon.com と同じテクノロジーをベースとし、機械学習を使って時系列データを付加的な変数に結びつけて予測を立てます。Amazon Forecast を使用する際に、機械学習の経験は必要ありません。必要なのは過去のデータと、予測に影響を与える可能性があるその他の追加データだけです。たとえば、シャツの特定のカラーの需要は、季節や店舗の所在地によって変わることがあります。こうした複雑な関係性は、過去のデータのみに基づいて判断することは困難で、それを認識することに理想的に適しているのが機械学習です。Amazon Forecast は、ユーザーがデータを提供すると、それを自動的に精査し、何が重要かを識別して、予測を立てるための予測モデルを作成します。このモデルの精度は、時系列データのみに基づく場合と比べ、最大で 50% 高くなります。

2019/9/30

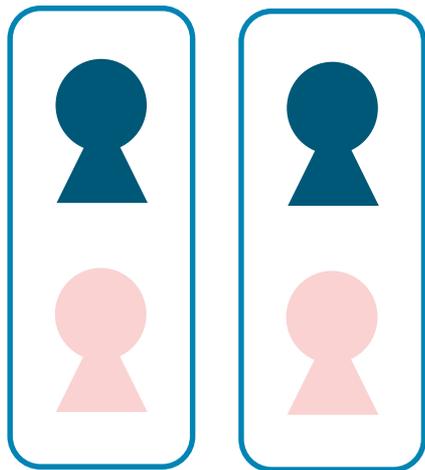
<https://aws.amazon.com/jp/forecast/>

# AutoMLが作るこれからの開発体制予測

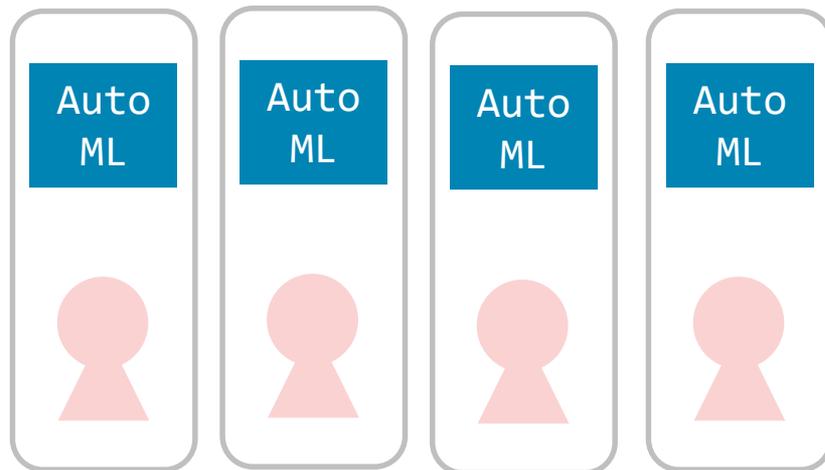
AIプロジェクトの数が、データサイエンティストに依存しないでスケールする。

データサイエンティスト

データエンジニア



AIプロジェクト AIプロジェクト  
AutoMLが対応していないタスクはデータサイエンティストが担当



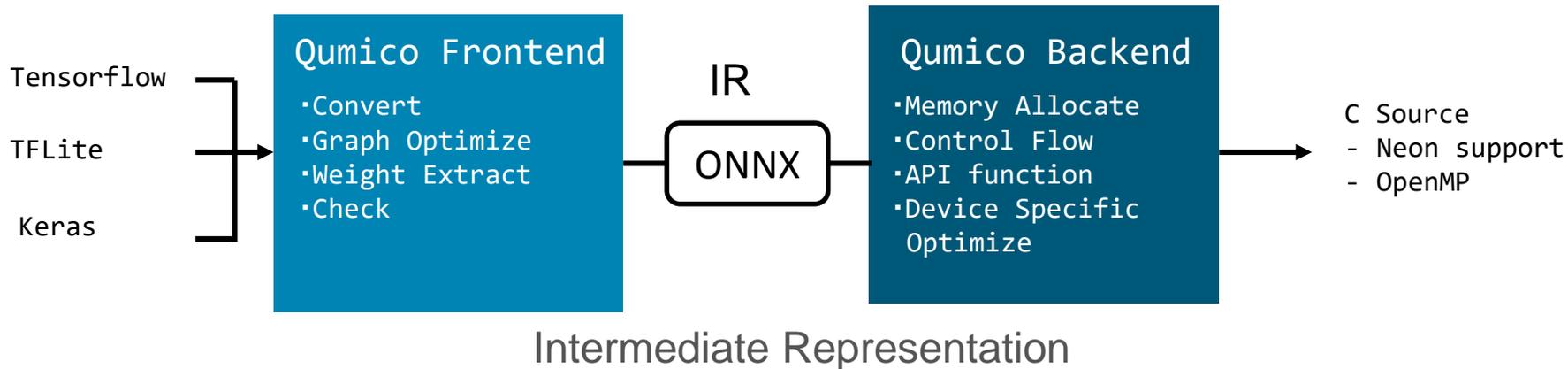
AIプロジェクト AIプロジェクト AIプロジェクト AIプロジェクト  
AutoMLが対応しているタスクは、AutoMLに入力するデータを用意するデータエンジニアがいれば良い

# | Qumicoの紹介

# Qumicoとは

- 目的
  - ディープラーニングの学習結果をエッジデバイス向けにC言語変換
    - OSレス環境での利用も想定
  - 出力：標準C(その他依存のライブラリなし)
  - IR：ONNXに準拠
- 弊社内製
- オープンソースとして公開(MITライセンス)
  - <https://github.com/PasonaTech-Inc/Qumico>

# Qumico Architecture

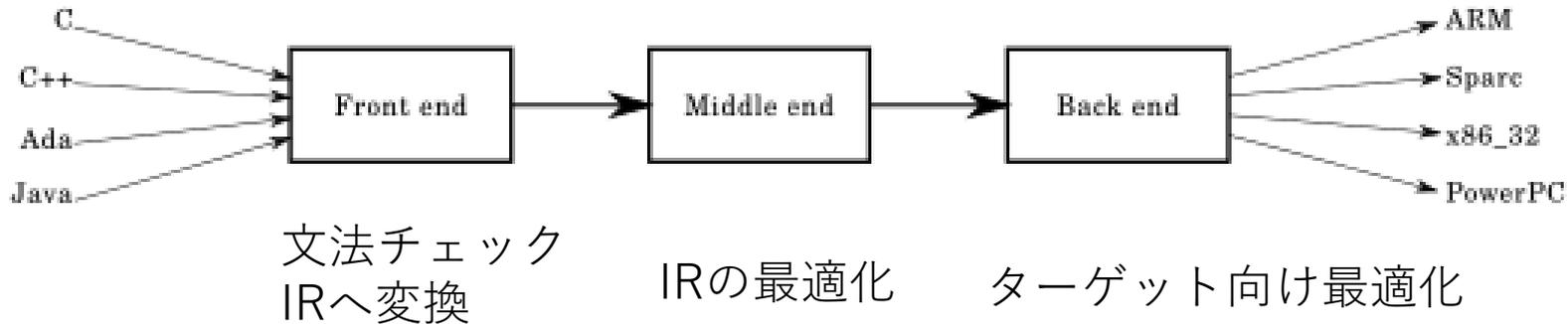


- フロントエンド
  - 各種DLフレームワークの学習モデルをONNXに変換する
  - 例：TensorFlowの学習済みモデルをONNXに変換
- バックエンド
  - ONNXをC言語に変換する

# Three Stage Compiler

スタンダードな3stage コンパイラの構成を踏襲

Wikipediaより



IR  
Intermediate Representation

2019/10/5

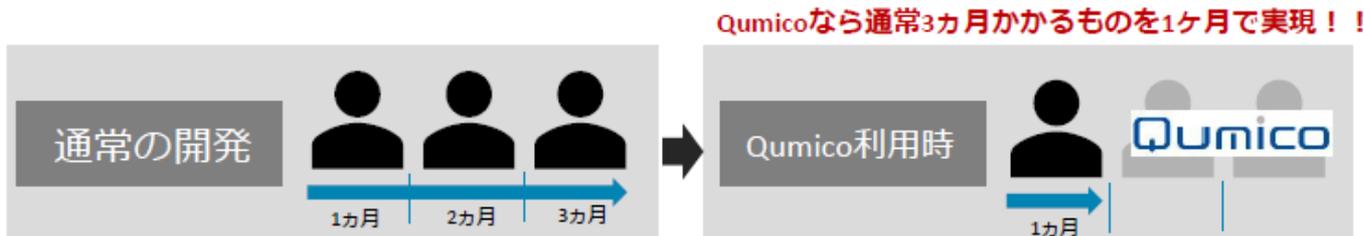
<https://en.wikipedia.org/wiki/Compiler> 24

# Qumicoの特徴

## 「Qumico」特徴

- 01 ディープラーニングの学習には、TensorFlow、Keras等の業界標準フレームワークを使用可能
- 02 学習結果を早い段階で組込機器へ実装できるため、実機でのスピーディーな性能評価が可能
- 03 SoCやカスタムCPUといった業界標準フレームワークが対応していないボードでも、エッジAIを動作させることが可能

●VGG16の実装の最適化の場合



# Qumicoを使った開発フロー

TensorFlow

Google Cloud AutoML  
(識別)

1.学習

2.変換

ONNX

3.チューニング

4.デプロイ



エッジデバイス

C

チューニング

- ・ CPUコア
- ・ SIMD
- ・ GPU
- ・ メモリアライメント等

計算グラフ最適化  
量子化

Qumicoバックエンド

Qumicoフロントエンド

学習済みモデル  
TFlite

ハンズオン資料に具体的な  
手順の説明あり

# Qumico の特徴

| インポート     | イメージ   | トレーニング | 評価 | テストと使用 |
|-----------|--------|--------|----|--------|
| すべての画像    | 14,558 |        |    |        |
| ラベル付き     | 14,558 |        |    |        |
| ラベルなし     | 0      |        |    |        |
| ラベルをフィルタ  |        |        |    |        |
| bottle    | 1,266  |        |    |        |
| person    | 13,369 |        |    |        |
| tvmonitor | 1,129  |        |    |        |

| tmmonitor   | イメージのフィルタリング  |
|---|---|
|  |  |
| tmmonitor(1)  | tmmonitor(1)  |

Cloud AutoML

Qumico



カメラ + ラズベリーパイ

AutoMLで学習したモデルが、Qumicoを使ってエッジで動く

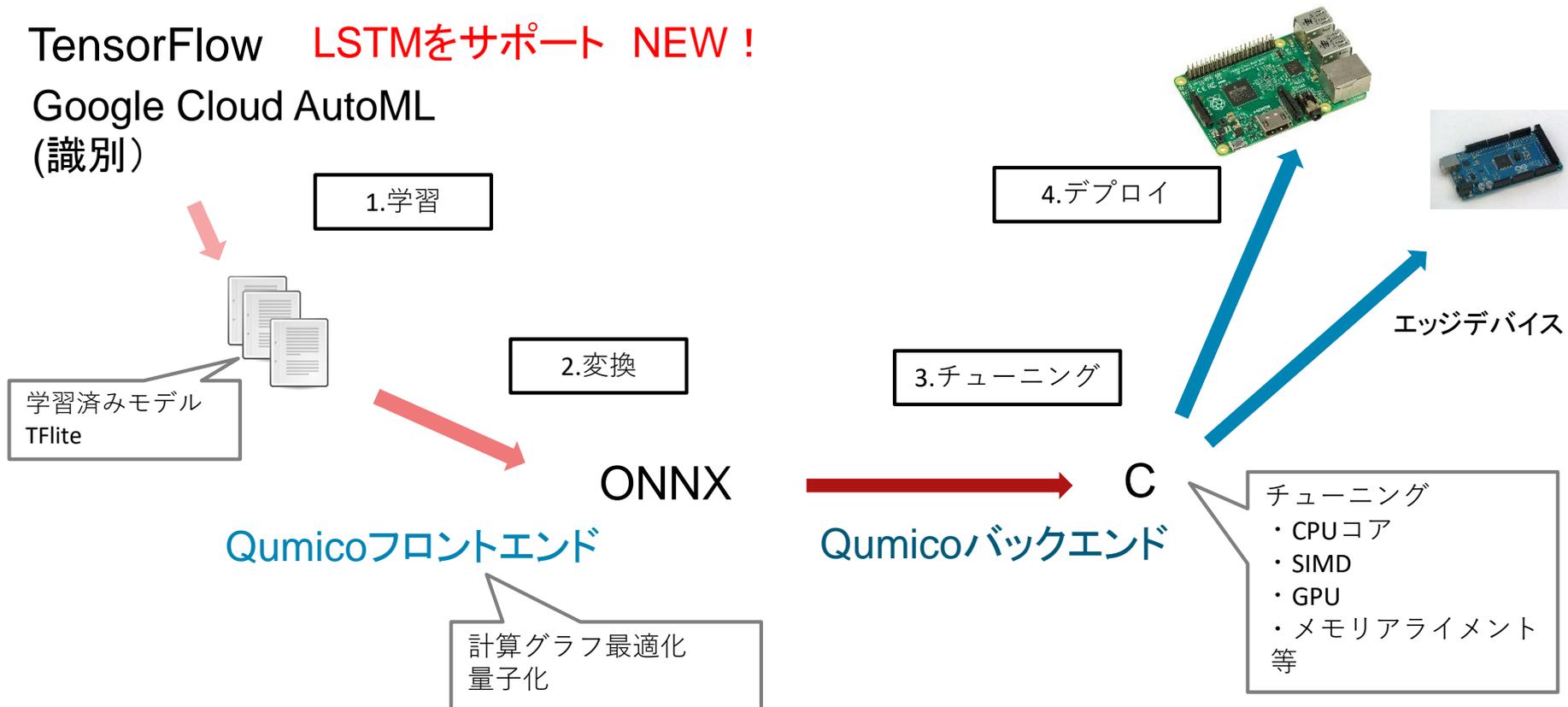
- ニューラルネットワークの知識不要
- 組込の知識不要

# | Qumicoデモ

# Qumico新機能

TensorFlow LSTMをサポート NEW!

Google Cloud AutoML  
(識別)



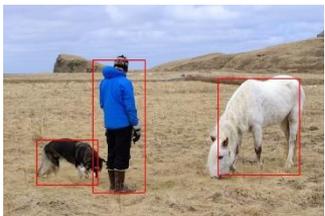
# 時系列データとLSTM

## 従来のQumico



分類タスク  
犬 or 猫

VGG16



物体検出

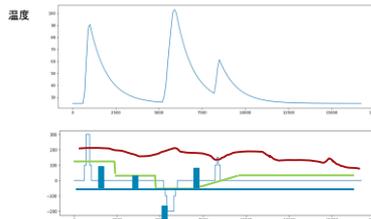
YOLO

画像

[https://ja.wikipedia.org/wiki/%E3%82%A4%E3%83%8C#/media/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Shiba\\_Inu.jpg](https://ja.wikipedia.org/wiki/%E3%82%A4%E3%83%8C#/media/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Shiba_Inu.jpg)

<https://github.com/pjreddie/darknet/blob/master/data/person.jpg>

## 新しいQumico



LSTM ← NEW !

- 株価の予測
- テキスト分析
- 翻訳
- センサーデータの異常検知

時系列



# | これからのエッジAI開発

# コロナ前



夏谷

滋賀県



週2~3日出張ベース

PM

経営陣



上司



開発メンバー



東京本社

# With コロナの体制図



リモートでどうやって開発を行うのか

<https://www.freemap.jp/item/japan/japan1.html>

<https://www.irasutoya.com/> 33

# やってみたことの紹介

- LabelBox
- Google Colab
- 組込機器の画面共有
- Azure IoT

# Labelbox

オンラインアノテーションサービス

<https://labelbox.com/>

Computer Vision

- Image Classification
- Object Detection
- Image Segmentation

← 今回はこの機能を評価しました

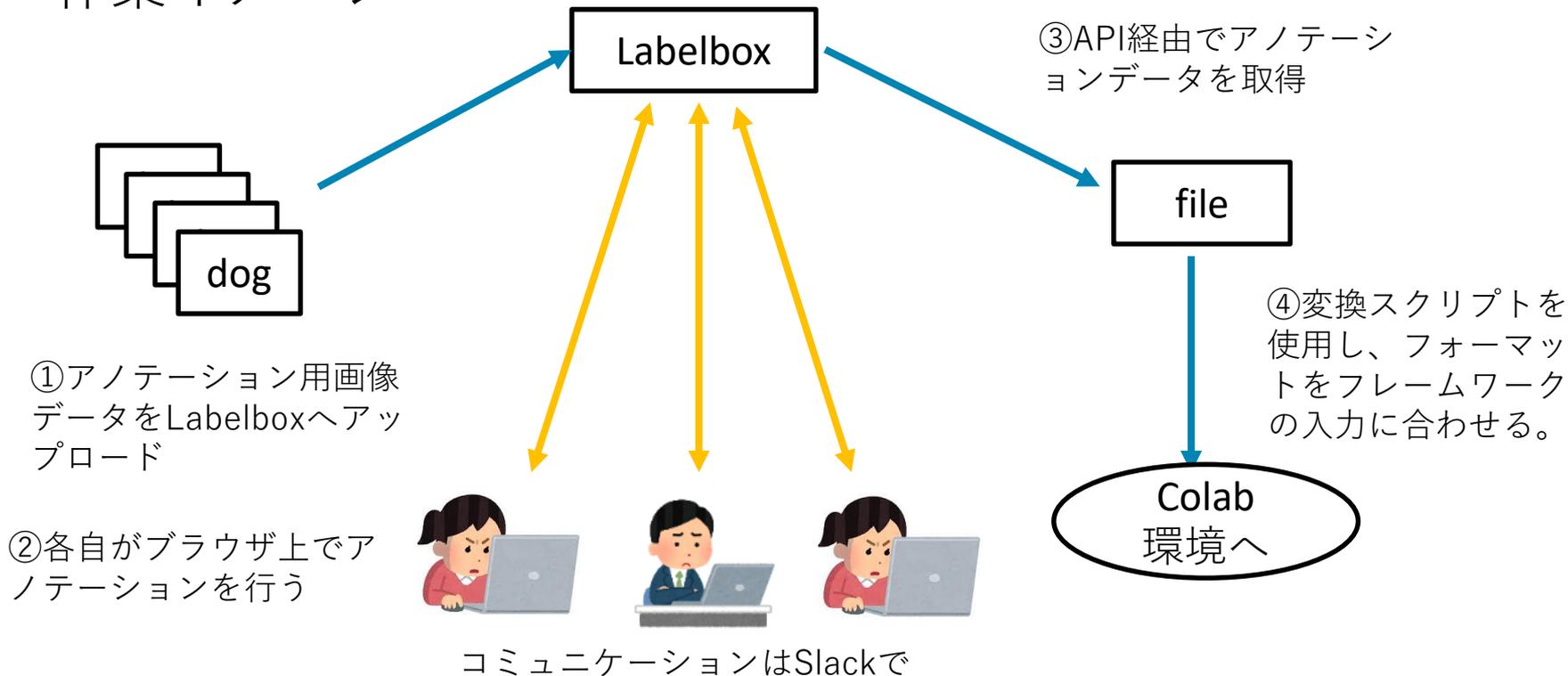
NLP & Others

|  | Developer | Pro                           | Enterprise |
|--|-----------|-------------------------------|------------|
| Usage & management                         |           |                               |            |
| Labeled Assets / year                      | 2,500     | 2,500+                        | Custom     |
| Users                                      | 5 Users   | Unlimited                     | Unlimited  |
| Role based access control                  | ■         | ■                             | ■          |
| Workforce (Fully managed labeling service) |           | Starts at \$6 / labeling hour | Custom     |

<https://labelbox.com/pricing>

# Labelbox

## 作業イメージ



# Labelbox

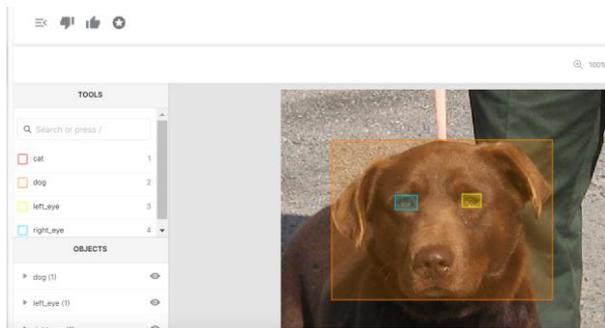
期間：2020年5月11日～6月10日

メンバー：四名（全員リモート）

アノテーション：犬の顔、右目、左目

元データ：Kaggle dogs vs cats

データ数：1849枚



無料のlimitの扱い  
が難しく、2500枚  
まで行かず。

<https://www.kaggle.com/c/dogs-vs-cats>

# Labelbox

- ・良かったところ

- アノテーションがブラウザのみで行える。

Windows、Mac、Ubuntuの全てをサポートしているツールが見つからなかった。

- 会社支給のセキュリティが厳しいPCでも作業が出来る。

- 画像を担当者に割り振る必要が無い。

割り振りはLabelboxでやってくれる。

- ・上手く行かなかったところ

- フィードバックが1日、2日という単位になる。

# Google Colab

## Google Colaboratory

- Jupyter notebookベースのクラウドサービス
- 無償でGPUも使用できる。
- 時間制限あり

<https://colab.research.google.com/notebooks/welcome.ipynb?hl=ja>

# Google Colabの使い方

## よく使うパターン

### ①Google Driveのmount

```
▶ from google.colab import drive  
drive.mount('/content/gdrive', force_remount=True)
```

↳ Mounted at /content/gdrive

アノテーション済み  
データはGoogle Drive  
の中

### ②git clone

```
▶ !git clone -b feature/natu https://github.com/natsutan/dream "/content/gdrive/My Drive/dream"
```

### ③ライブラリのインストール

```
▶ # 必要なライブラリをインストールする  
! pip install -r requirements.txt  
! pip install tensorflow-gpu==1.13.2
```

ハンズオン資料に具体的な  
手順の説明あり

作業の度にこの手順を実行する

# Google Colabの使い方

## リモート開発時に一番役に立ったこと

```
▶ from google.colab import drive
  drive.mount('/content/gdrive', force_remount=True)

🔗 Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client\_id=947318989803-6bn6qk8qdgf4n4g3pfee649
Enter your authorization code:
.....
Mounted at /content/gdrive

▶ !git clone -b feature/natu https://github.com/natsutan/dream "/content/gdrive/My Drive/dream"

🔗 fatal: destination path '/content/gdrive/My Drive/dream' already exists and is not an empty directory.
```

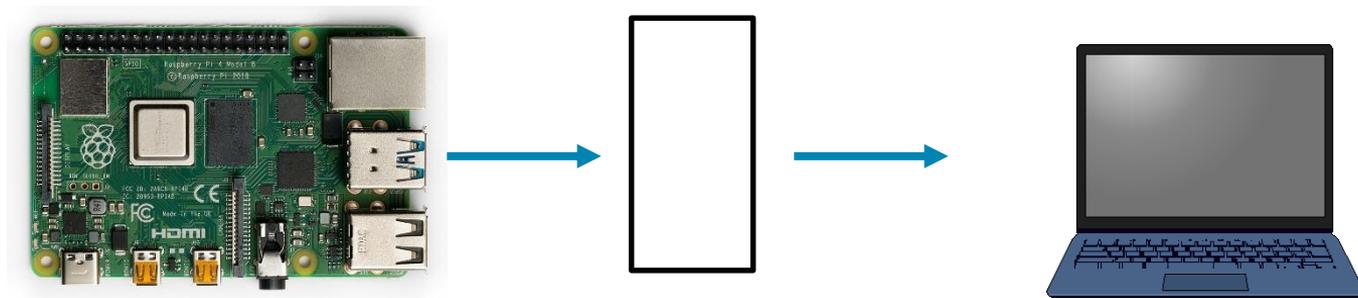
- ・上手く動かないColabファイルをそのまま送ってもらうと、やったことの履歴やエラーメッセージが全部残っている。
- ・手元での再現性が非常に高い。
- ・多くの問題はGoogle Drive内のファイルが違う事が原因。

手順やエラーメッセージの確認のためのやりとりが無し

# Google Colab

- ・良かったところ
  - 全員が同じ環境になる。OS、Driver、コンパイラ
  - 環境をファイル一つで共有できる
  
- ・上手く行かなかったところ
  - 長時間の学習中に接続が切れる。対策をするための手間が増える。
  - gitとは相性が良くない
  - Google Driveの内容を同じに保ちにくい

# 組込機器の画面をTeams/Zoomに出す



HDMI capture

I-O DATA GV-USB3HD/E

[https://ja.wikipedia.org/wiki/Raspberry\\_Pi#/media/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Raspberry\\_Pi\\_4\\_Model\\_B\\_-\\_Top.jpg](https://ja.wikipedia.org/wiki/Raspberry_Pi#/media/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Raspberry_Pi_4_Model_B_-_Top.jpg)

# Azure IoTの取り組み

現在の取り組み：

Azure IoT、IoT Hubの機能を使って、リモート組込開発を効率化できないか

# | ワークショップの紹介

# ワークショップの趣旨と現状

ワークショップの趣旨：  
情報系以外の学科を履修している大学生にも、Deep Learningを簡単に体験できないか？  
ハンズオンを使ったワークショップをしたい

物理、化学、  
建築など

ワークショップの現状：  
テキストは出来たがコロナで開催の目処が立たず。  
趣旨が素晴らしいだけになんとかしたい

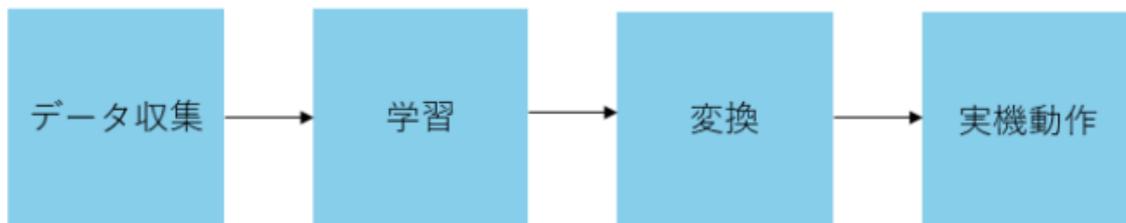


資料を自習ベースに修正

# ワークショップの内容

## ワークショップアウトライン

Ver2.0



1

©2018 Pasona Tech, Inc. All Rights Reserved.

データ収集から、学習、ラズパイ動作までを2日間かけて体験します。  
扱うタスクは画像の識別  
二～三名のチーム構成を想定していますが一人でも大丈夫です。

# ワークショップに必要なアイテム

- Google アカウント
- インターネットに繋がるPC
- ラズパイ + カメラ + モニターの環境

# ワークショップ 1日目

| 時刻（一日目）     |             |
|-------------|-------------|
| 10:00～10:30 | 導入・流れ説明     |
| 10:30～11:00 | パソナテックとは    |
| 11:00～11:15 | アイスブレイク     |
| 11:15       | ワークショップ開始   |
| 12:30～13:15 | 昼食          |
| 13:15       | ワークショップ続きから |
| 18:00       | 終了          |

Deep Learningで使う  
データを集める。

AutoMLを使って学習、  
モデルのダウンロード  
まで

# ワークショップ2日目

| 時刻（二日目）     |           |
|-------------|-----------|
| 10:00～10:10 | 挨拶・流れ説明   |
| 10:10       | ワークショップ開始 |
| 10:15       | 中間発表      |
| 12:30～13:15 | 昼食        |
| 13:15       | ワークショップ続き |
| 16:30       | 最終発表      |
| 17:15       | 総評・挨拶     |
| 18:30       | 終了・解散     |

ラズベリーパイでの  
エッジAI動作

結果発表

# ワークショップ開催希望の方へ

## 大学関係者

- ・新卒採用のチームが担当

## 企業関係者

- ・営業のチームが担当

最後のアンケートで、ワークショップ開催希望と記入ください。ただし、ご希望に添えない場合もございます。

# | 会社紹介

# パソナテックの取り組み（2019年～）

## データエンジニアの育成

### データエンジニアとは？

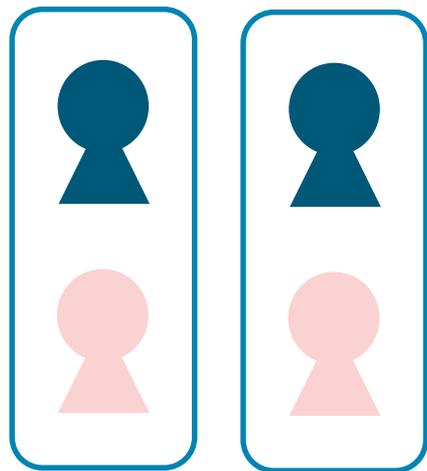
データの前処理を中心に、データサイエンティストの指示で、データそのものを扱うエンジニア

# AutoMLが作るこれからの開発体制予測

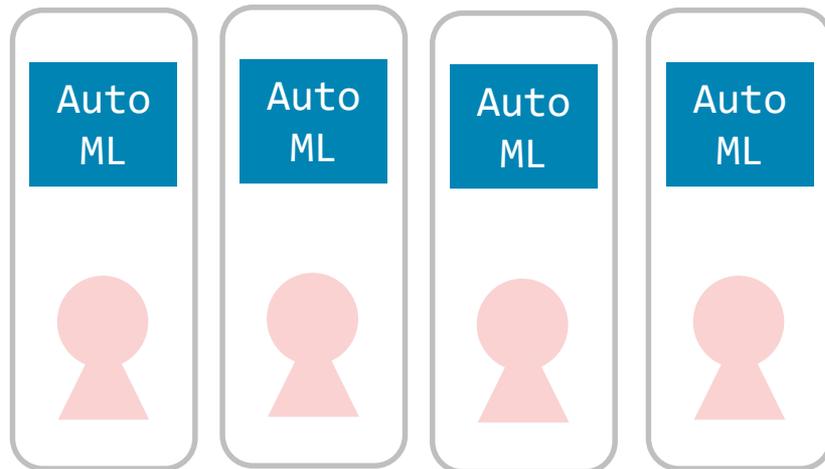
AIプロジェクトの数が、データサイエンティストに依存しないでスケールする。

データサイエンティスト

データエンジニア



AIプロジェクト AIプロジェクト  
AutoMLが対応していないタスクはデータサイエンティストが担当



AIプロジェクト AIプロジェクト AIプロジェクト AIプロジェクト  
AutoMLが対応しているタスクは、AutoMLに入力するデータを用意するデータエンジニアがいれば良い

# パソナテックの取り組み（2019年～）

## データエンジニアのスキルセットの標準化と育成コースの開発

データエンジニア 詳細

|              | level | skill                 | description   |
|--------------|-------|-----------------------|---|
| データ分析初級      | 1     | Python,Excel,R        | データの基本統計量（平均、分散、分散係数）を算出できる。                            |
| 可視化初級        | 1     | Python,Excel,R        | データの基本統計量を可視化できる。                                       |
| 前処理初級        | 1     | Python,Excel,R,C++    | 事前に集めたデータの基本的なデータの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。 |
| アノテーション      | 1     | Excel                 | 数値に代り、データに分類ラベルなどのアノテーションを付与できる。                        |
| 環境インストール     | 2     | Python,Excel,R,C++    | 実際に必要な開発環境のインストールと動作確認ができる。                             |
| 前処理中級        | 2     | Python,Excel,R,C++,統計 | 数値のデータを結合し、データの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。    |
| データ分析中級      | 3     | Python,Excel,R,統計     | 数値のデータを結合し、データの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。    |
| 可視化中級        | 3     | Python,Excel,R,統計     | 数値のデータを結合し、データの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。    |
| データ分析上級      | 4     | Python,R,統計           | 数値のデータを結合し、データの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。    |
| 前処理上級        | 4     | Python,R,統計           | 数値のデータを結合し、データの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。    |
| アノテーションツール作成 | 4     | Python,C++            | 数値のデータを結合し、データの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。    |
| データデプロイ      | 4     | Python,Excel,R,C++    | 数値のデータを結合し、データの操作ができる。数値による異なるデータの結合、数値のソート、クレンジング等。    |

物体検出用データオーギュメンテーション  
XML、乱数処理、回転、拡大、縮小、画像の合成

データエンジニア研修 基礎編：2019年6月開始  
データエンジニア研修 実践編：2019年9月開始  
データエンジニア研修 応用編：2019年9月開始

```
/filename>
width>
>
size>
```

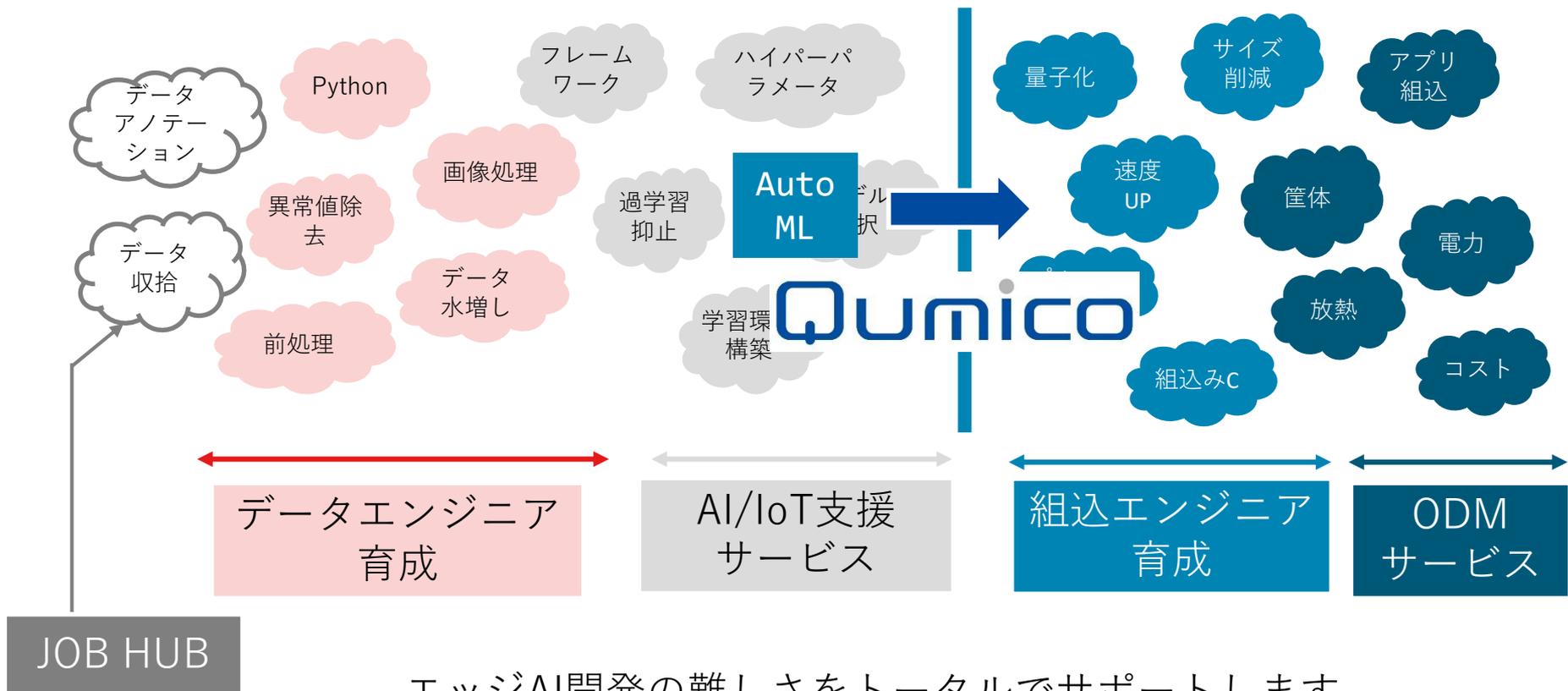


画像用NNの基礎  
画像フォーマットの基礎  
画像処理ライブラリ、アルゴリズム



```
<name>pencilcase</name>
<bndbox> <xmin>100</xmin>
<ymin>100</ymin>
<xmax>261</xmax>
<ymax>186</ymax>
</bndbox>
</object>
</annotation>
```

# エッジAIの難しいところ



エッジAI開発の難しさをトータルでサポートします。

# アンケート



上手く行かないときは、  
[mnatsutani@pasona.tech](mailto:mnatsutani@pasona.tech)  
までメールください。

もしくはtwitter  
@natsutan  
まで

