



# AI研究WG EdgeTech+ 2024

## AI研究WGの紹介とエッジAIの現状

2024年11月22日

応用技術調査委員会 AI研究WG

中村 仁昭



- 株式会社Bee CTO 中村仁昭
  - 新大阪の組込ソフト会社です
  - JASAパビリオンで「ネットが無くても生成AI」を展示しています
- CQ出版のInterfaceで記事を書いています
  - 2024年9月号(7/25発売)の「OpenCVで体験 現場プロの画像処理50」特集でラズパイ+Python
  - 2024年5月号の「ラズベリーパイ5大研究」特集で生成AI(LLM)

- 研究会とセミナーの2本立てで開催
- 研究会
  - 今年で6年目になるDeep Learningを既に理解して開発できるメンバーが集り、様々なテーマでAI活用研究を行なう研究会
  - メンバーは現在8社 16名
- セミナー
  - 今年で8年目になる初学者向けのDeep Learningセミナー
- AI研究WG発表会
  - 年度末に研究会/セミナー別で発表会を実施



- エッジデバイス上でのDeep Learningの可能性や、様々なテーマで持続的に調査研究を行なう
- 1ヶ月に1度、定例会議を開きDeep Learning周辺の最近の動向の共有、メンバーの研究内容の進捗発表
- 全員でコンペに参加して実力を試したり
  - 個々のメンバーで興味のあるコンペに参加



- 1年間で3回の座学とグループでのDeep Learningデモ作成がゴール
- 講習にはGoogle Colaboratoryを利用
  - Colaboratoryはクラウドで実行されるJupyter ノートブック環境なのでお手軽
- フレームワークはTensorFlow+Keras
- グループ間の情報共有、全体連絡にSlackを活用



# 2023年度活動内容・研究会



- 個々の研究案件を継続
  - 個々の研究テーマにそって、グループに分かれて研究を進めている
- Deep Learningの最近の話題の共有
  - LLM以外の話題も積極的に共有
- 開催されている機械学習コンペの確認
  - SIGNATEやNishikaなど国内のコンペを中心に確認
- 発表会
  - 研究会のメンバーでオンラインで実施

# 共有した最近の動向



- 会話型AIが成長すると突然「新しい能力」を獲得すると判明！
- LightGBM予測モデル実装ハンドブック
- テーブルデータ用ニューラルネットワークは勾配ブースティング木にどこまで迫れるのか？（本編）
- Metaの「Llama 2」をベースとした商用利用可能な日本語LLM「ELYZA-japanese-Llama-2-7b」を公開しました | ELYZA, Inc.
- 時系列予測にTransformerを使うのは有効か？
- 2023 LLMサマースクール コンペの解法メモ | ドクセル
- 松尾研 LLM講座 講義コンテンツ | 東京大学松尾研究室 - Matsuo Lab
- etc.

# 個々の研究案件



- 組込環境で生成AI
- 低リソースデバイスAI
- 画像の異常検出
- エッジデバイス上で学習
- 強化学習
- コンペに挑戦



# 成果発表・研究会



- RaspberryPi4/5、Jetson Xavier NXでLLM (llama.cpp、ELYZA 7B)を動作させて速度を計測

## 動作速度

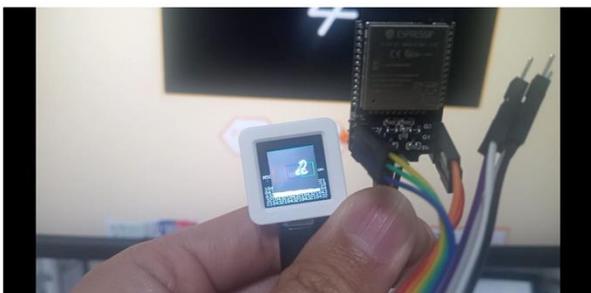
環境	Tokes/sec(tps)
Raspberry Pi 4	0.96
Raspberry Pi 5 without FP16	2.33
Raspberry Pi 5 with FP16	2.97
Jetson Xavier NX	10.97
GeForce RTX 4070Ti(12GB)	12.01



- 低リソースなエッジデバイスでフラッシュ暗算をさせてみる
  - ESP32で素早く切り替わる数字を認識して計算
  - MNIST(3層NN)を使用

## 実験

まずは、1桁のフラッシュ暗算を試してみた



## 実験結果

結果は

- 低リソースデバイスでAIを動かすことは可能だった
- 予測結果が“0, 3, 8”ばかりとなってしまった
- 推論速度が出せない

実験は以上になります。



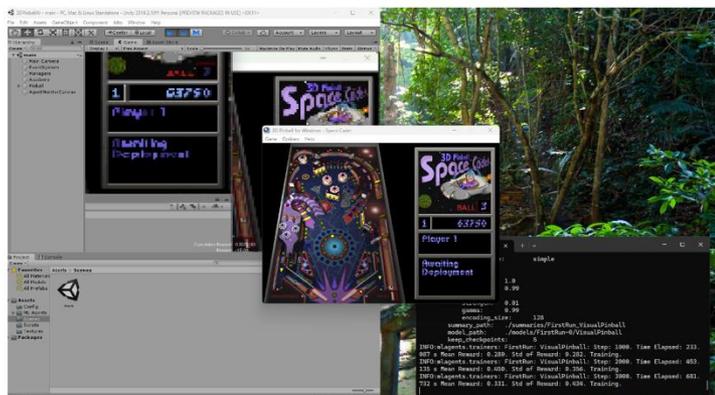
- 事前準備された環境以外で強化学習を行う
  - Windows標準ゲームだった3D Pinballを学習
  - 学習環境の構築にUnityのML-Agentsを使用

サンプルを触ってみる

SOHi

学習結果

SOHi



特定の操作を繰り返す状況に陥ることが原因と思われる。  
(ボール落下時のペナルティが影響?)



パラメータの調整、別モデルの使用などを試すも改善せず

- 木の健康状態を木の大きさ、抱える問題、周辺環境などの情報から予測するコンペ
  - 結果提出が期限切れだったが他の参加者の手法を参考に精度を向上させてみた

## 概要

- SMBC主催のコンペに参加
- コンペ目的
  - 木の健康状態を予測する
    - 良い/普通/悪い
- 使用可能なデータ
  - 木の大きさ、木の抱える問題、周辺環境などの情報

## 感想

- コンペ期間内に間に合わせる事が出来なかった
- モデルの選定によってかなりスコアが上昇した
  - catboosting
- ターゲットエンコーディングは適用するカテゴリ変数の選定次第でスコアの上昇に繋がる可能性あり
  - 実際にスコアが伸びている参加者もいた



# 2023年度活動内容・セミナー



## ■ 第1回 2023/6/14

- Deep Learningの説明
- Deep Learningの最近の話題
  - ー [速報]マイクロソフト、自然言語をプログラミング言語にAIで変換、新ノーコード機能をPower Appsに搭載。AI言語モデル「GPT-3」を採用。Microsoft Build 2021
  - ー 多層パーセプトロン(MLP)時代の到来と、トランスフォーマーの終焉
- Python基礎とMNISTデモをGoogle Colaboratoryで実施



## ■ 第2回 2023/8/2

- Neural Networkの解説
- Neural Networkの学習アルゴリズムの説明
- 課題発表に向けたグループ分け

## ■ 第3回 2023/11/8

- ハイパーパラメータなど学習にあたってのテクニックの解説
- CNNの解説
- ColabでKeras MNISTの学習結果の可視化デモ
- 課題進捗発表



- 第4回 2024/2/14
  - 課題進捗発表会
  - 課題推進
- セミナー成果報告会 2024/3/21
  - 成果発表



# 成果発表・セミナー

# 画像生成AIの作成

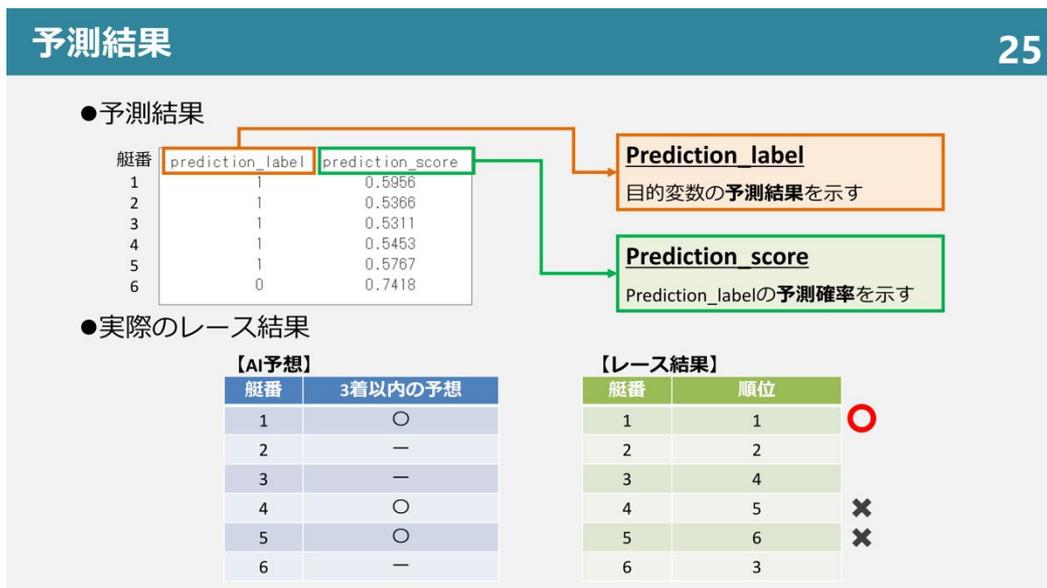


- GANを使用した生成AIを作成
- 色々な年齢の顔画像データを学習
- ある人物の写真を提示し、特定の年齢での顔写真を生成

生成した画像



- 3着以内に入賞するボートの予測モデルを作成
- PyCaratで複数モデルからlightgbmを選択
- Oputunaでパラメータチューニングを実施



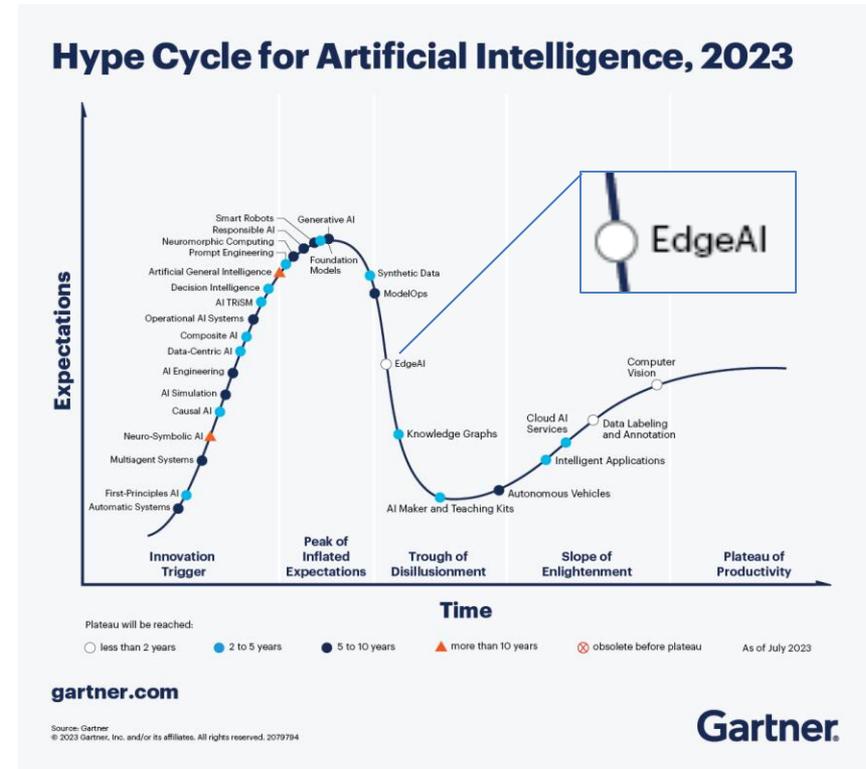


# 組込みAIの現状

# 組込みAI(EdgeAI)の位置付け



- GartnerのHype Cycle for Artificial Intelligence, 2023でようやく幻滅期に入り、ますます実装例は手に入りにくい状況に
  - 主流の採用までに要する年数は2年未満とされている



<https://www.gartner.com/en/articles/the-4-trends-that-prevail-on-the-gartner-hype-cycle-for-ai-2021>

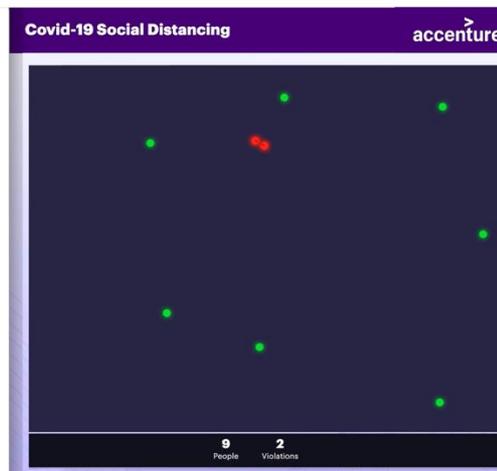
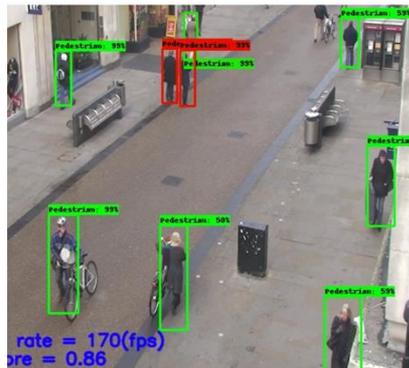


- スマホの音声認識やカメラ画像処理は採用されて久しい
- Intel coral の [Customer Stories](#) は実際に使われた実例がまとまっている
  - コンサルティング会社AccentureのAIを活用した [外観検査の取り組み](#)
  - くら寿司の [皿を数える装置](#)
  - ノルウェーの配電会社Pratexoの電力グリッドの変圧器の [異常検知](#)
  - など

# Accentureの外観検査の取り組み



- 知的財産と労働者を保護するためにプライバシーとセキュリティを強化する必要があり、ネットワークを使用しないローカルの分析にCoral Edge AIを選択



Coral

Safety Surveillance Analysis

Safety hats count	0
Safety vests count	1
Non safety hats count	0
Non safety vests count	0
No hats count	0

Sign out

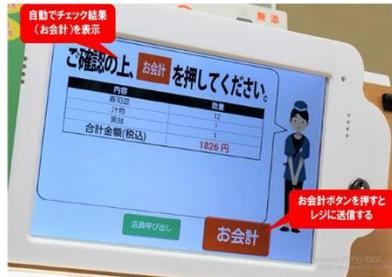
Inference Time: 12.26 ms (81.56 fps)  
Objects Detected: 1

This interface shows a video feed of a construction site. A worker in a yellow safety vest is highlighted with a bounding box and the label "100% safety vest". The interface includes a "Safety Surveillance Analysis" table with counts for safety hats, safety vests, and non-safety items. It also features a "Sign out" button and performance metrics like "Inference Time: 12.26 ms (81.56 fps)" and "Objects Detected: 1".

# くら寿司の皿を数える装置



- 回転レーンから取った皿の数をラズパイ4でQRコードの識別とTensorFlow(Coral USB Accelarator)を使った画像認識で皿の種類と数をカウント



<https://coral.ai/news/kura-sushi>



- Coral M.2アクセラレーターで各変圧器が発する音から機械学習モデルで問題が発生するかを予測し電力グリッドの信頼性を確保



# 組込みAI実装の普及に向けて



- 実装例が表に出てこない
  - 調査は継続する必要があるそう
- 幻滅期に入りメディアに注目されないため、自ら実装を進めないと情報が集まらない可能性が高い
  - Computer Visionなど啓発期に入っている技術を組込みに導入するのであれば低リスクで実装を進められる



「AI研究WG EdgeTech+ 2024」

2024/11/15 発行

発行者 一般社団法人 組込みシステム技術協会  
東京都 中央区 入船 1-5-11 弘報ビル5階  
TEL: 03 (6372) 0211 FAX: 03 (6372) 0212  
URL: <https://www.jasa.or.jp/>

本書の著作権は一般社団法人組込みシステム技術協会（以下、JASTA）が有します。  
JASTAの許可無く、本書の複製、再配布、譲渡、展示はできません。  
また本書の改変、翻案、翻訳の権利はJASTAが占有します。  
その他、JASTAが定めた著作権規程に準じます。