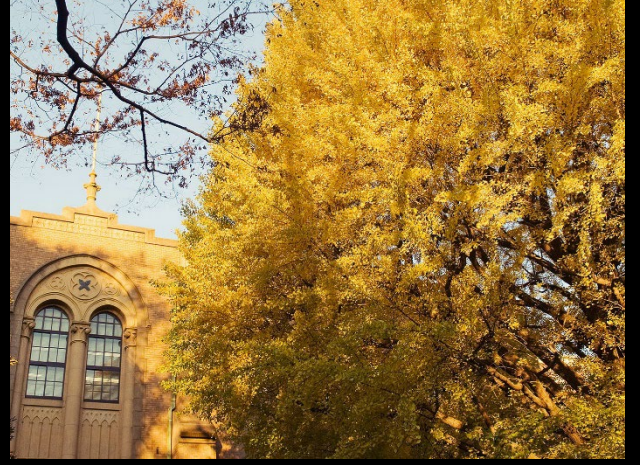


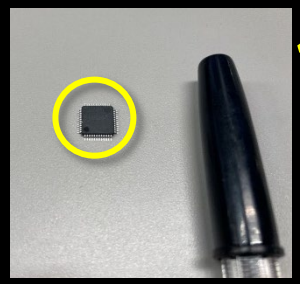
適応性と信頼性を両立するオンデバイス学習技術の確立

松谷宏紀（慶應大学）、渡邊竜司（パナソニック）、吉田康太（立命館大学）




AIの応用ドメイン：計算機の視点

IoT　　Mobile　　Edge server　　Cloud server

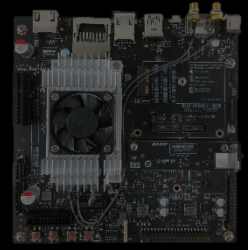


20-30mW

0.2-1W

2-10W

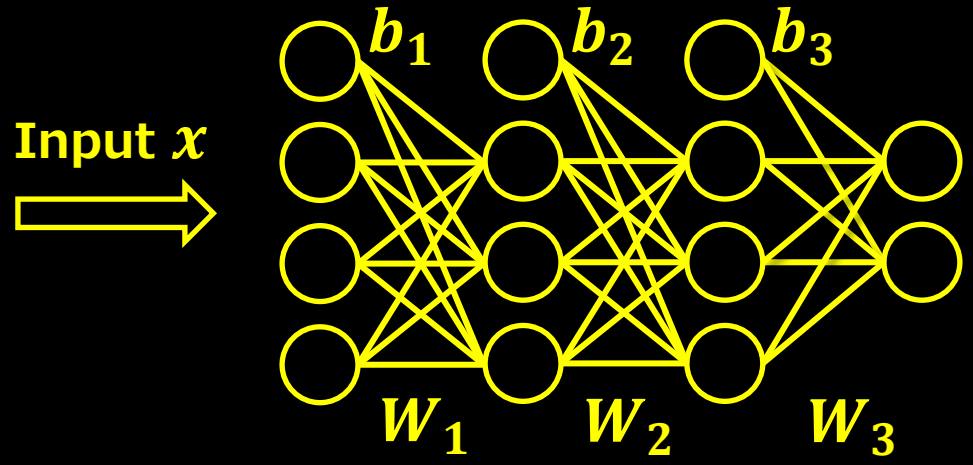
15W+



200W+

コントローラ　　組込みCPU　モバイルCPU　　組込みGPU　　高性能GPU

エッジAI：現実空間で動作する組込みAI



推論（Prediction）

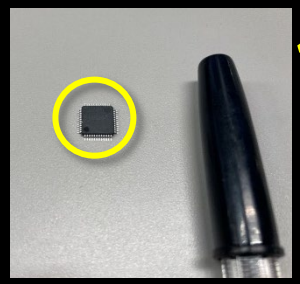
入力xと重み係数W, bをもとに出力を計算

学習（Training）

教師データをもとに重み係数W, bを計算

AIの応用ドメイン：計算機の視点

IoT　　Mobile　　Edge server　　Cloud server

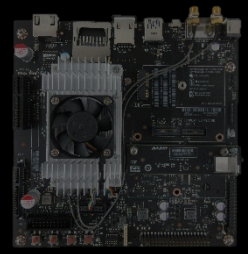


20-30mW

0.2-1W

2-10W

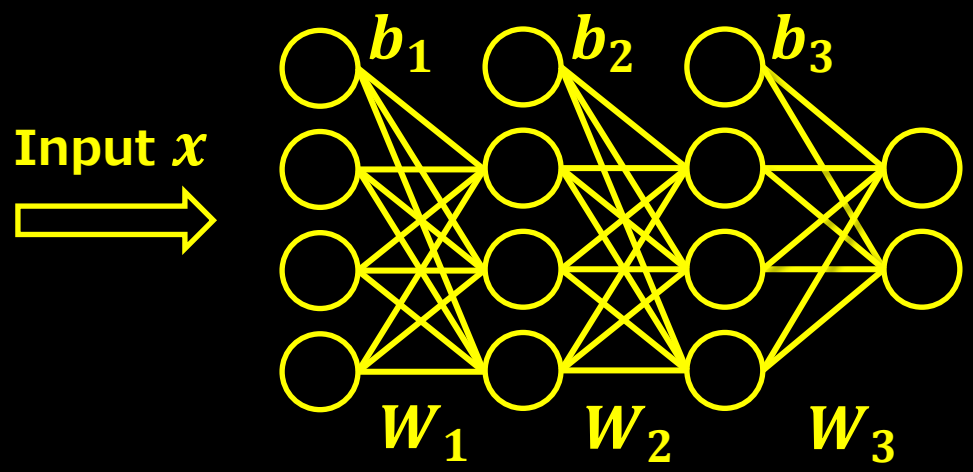
15W+



200W+

コントローラ　　組込みCPU　モバイルCPU　　組込みGPU　　高性能GPU

エッジAI：現実空間で動作する組込みAI



従来のエッジAI

現場では推論のみ（モデルはサーバで学習しておく）

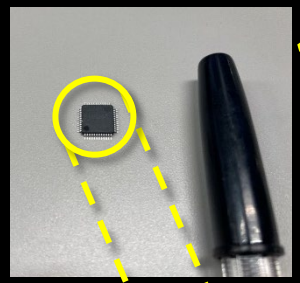
⇩

オンデバイス学習

現場で学習もできる（モデルを現場に合わせて更新）

さきがけ（2013~）、CREST（2017~）、AIP加速（2023~）

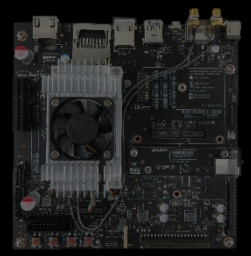
IoT　Mobile　Edge server　Cloud server


20-30mW

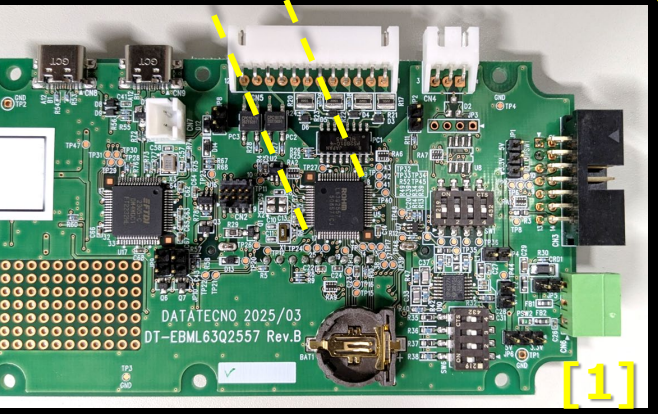

0.2-1W


2-10W

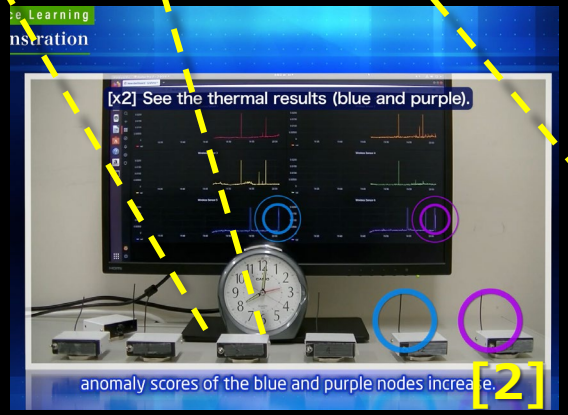

15W+


200W+

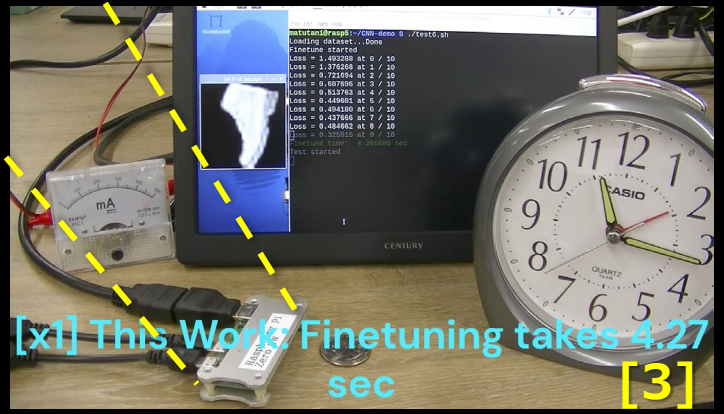
コントローラ　組込みCPU　モバイルCPU　組込みGPU　高性能GPU



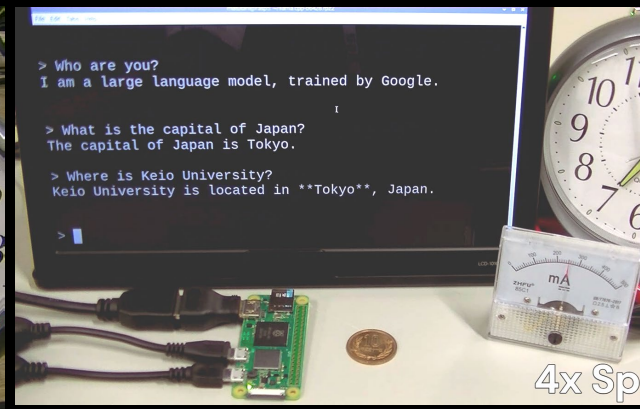
[1]



[2]



[3]

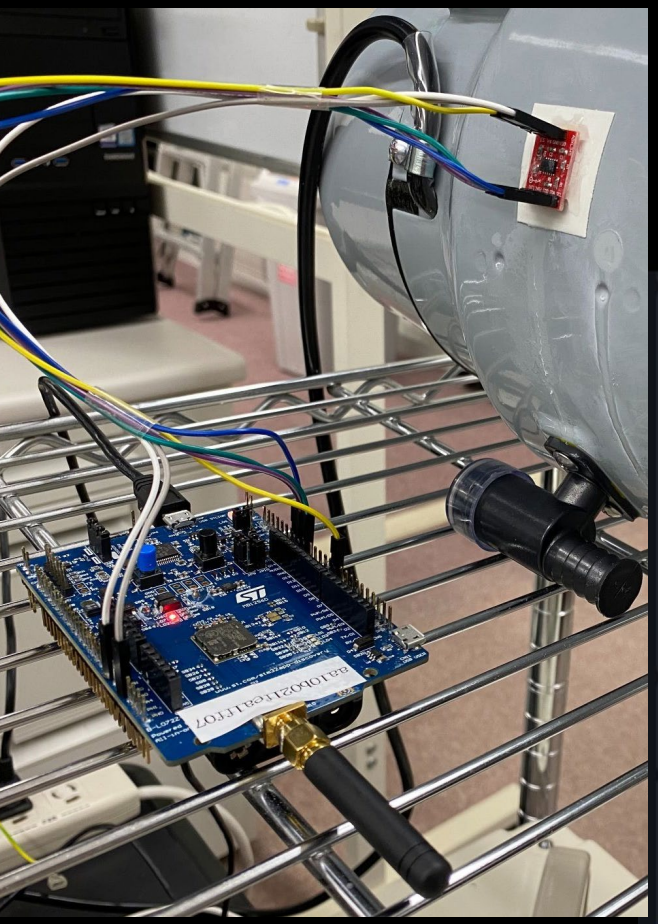


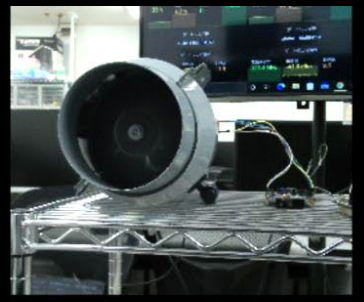
信号処理用NN　画像認識用CNN　エッジ用LLM

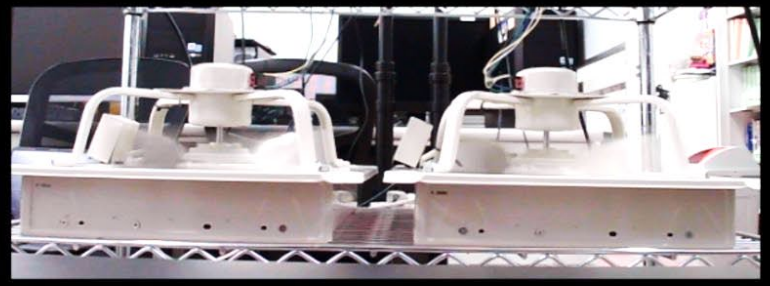
[1] ローム株式会社, "エッジコンピューティング向け完全自立型AIソリューション Solist-AI™", https://www.rohm.co.jp/support/solist-ai.
[2] Kazuki Sunaga et al., "Addressing Gap between Training Data and Deployed Environment by On-Device Learning", IEEE Micro (2023).
[3] Keisuke Sugiura et al., "InstantFT: An FPGA-Based Runtime Subsecond Fine-tuning of CNN Models", IEEE Micro (2026).

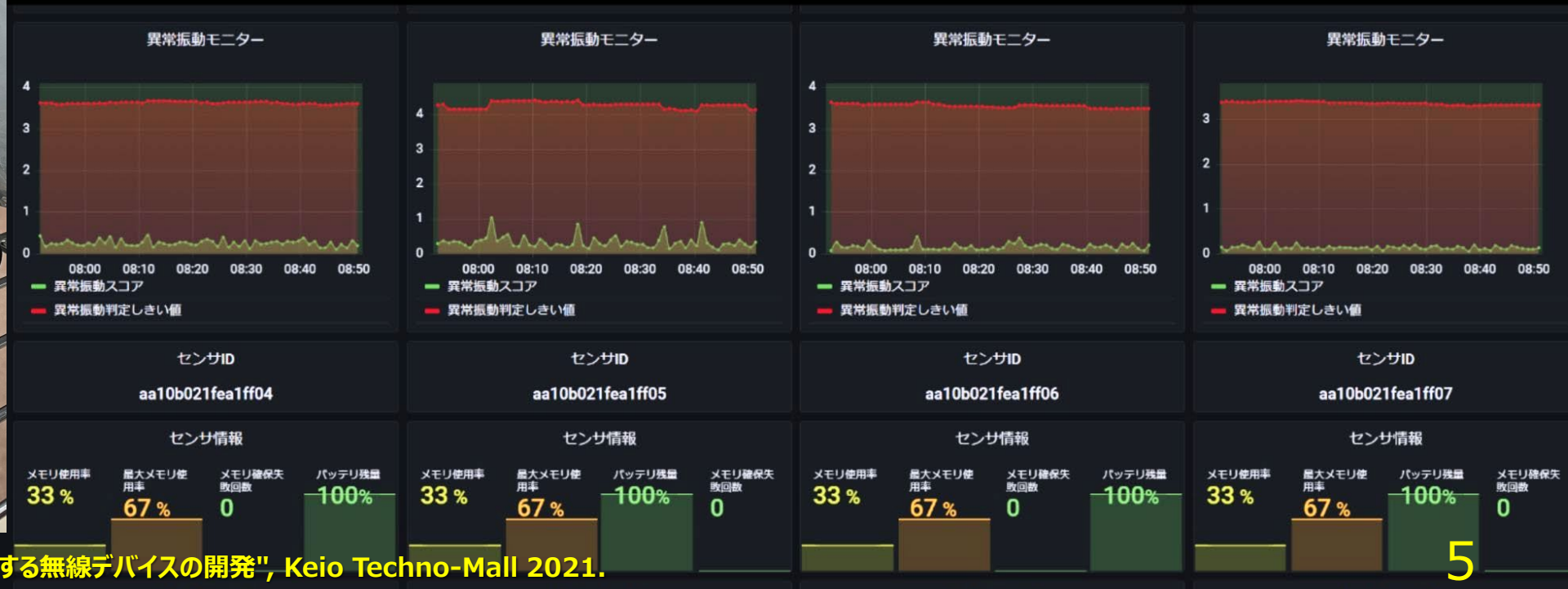
エッジAI：設備監視への応用例

- 無線センサノードによる空調システムの監視 [1]
  エッジ側でセンシング、前処理、推論（異常検知）、学習などの知的処理





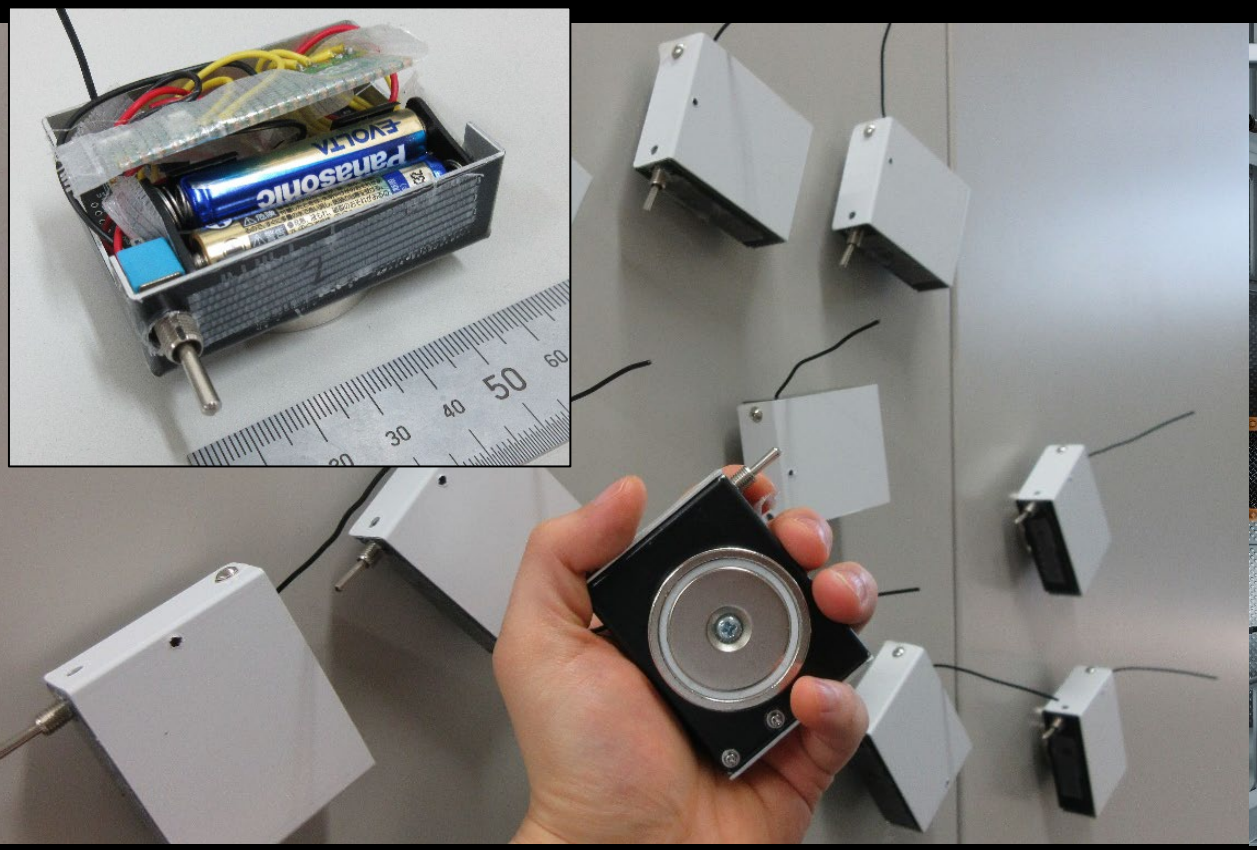


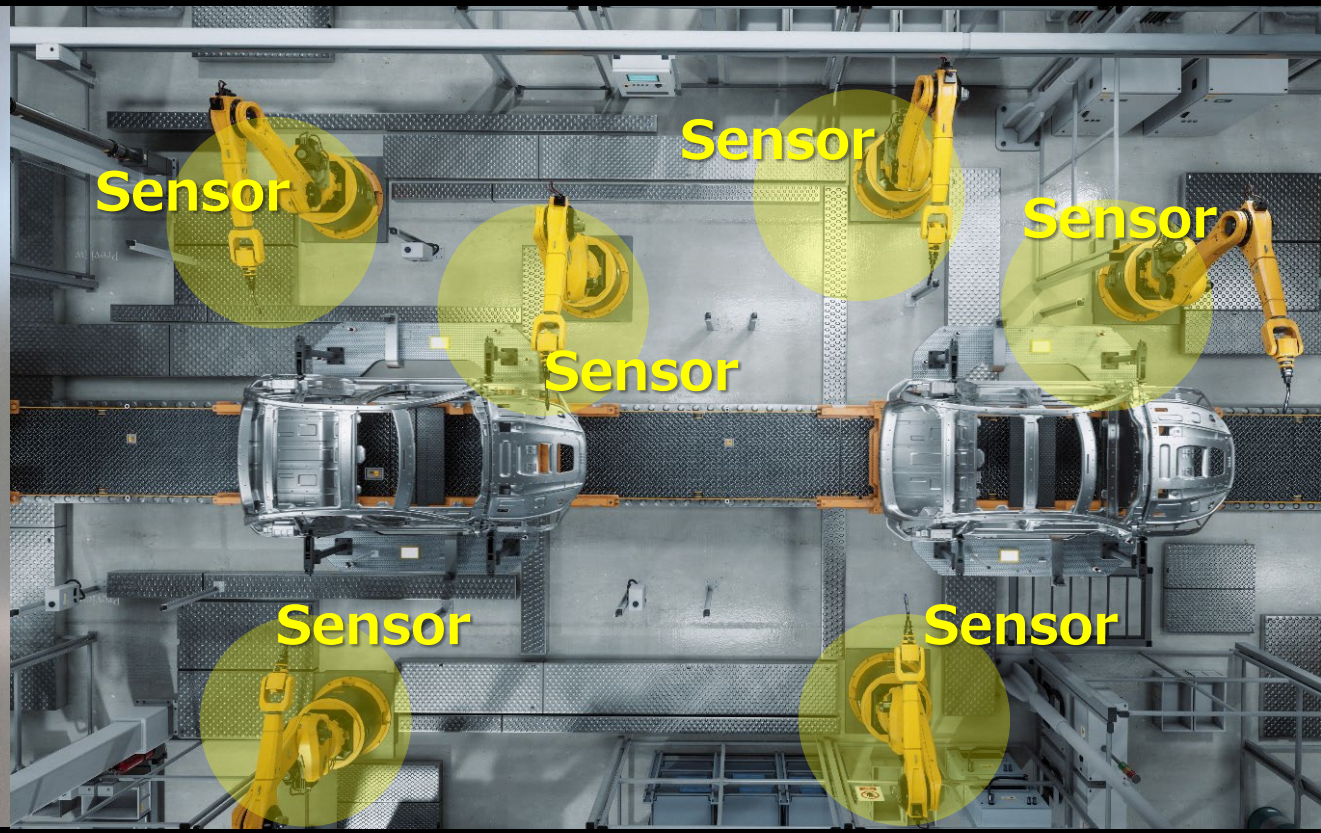


[1] 塚田 峰登, "回転機械の異常振動を検出する無線デバイスの開発", Keio Techno-Mall 2021.

オンデバイス学習：必要性

- 現場ごとにセンサの値の見え方が違う
  訓練データ（事前に集めたセンサの値）と現場で実際に得られる値がズレる
  原因：センサの角度（加速度センサのXYZ）、周囲のノイズ、個体差、…

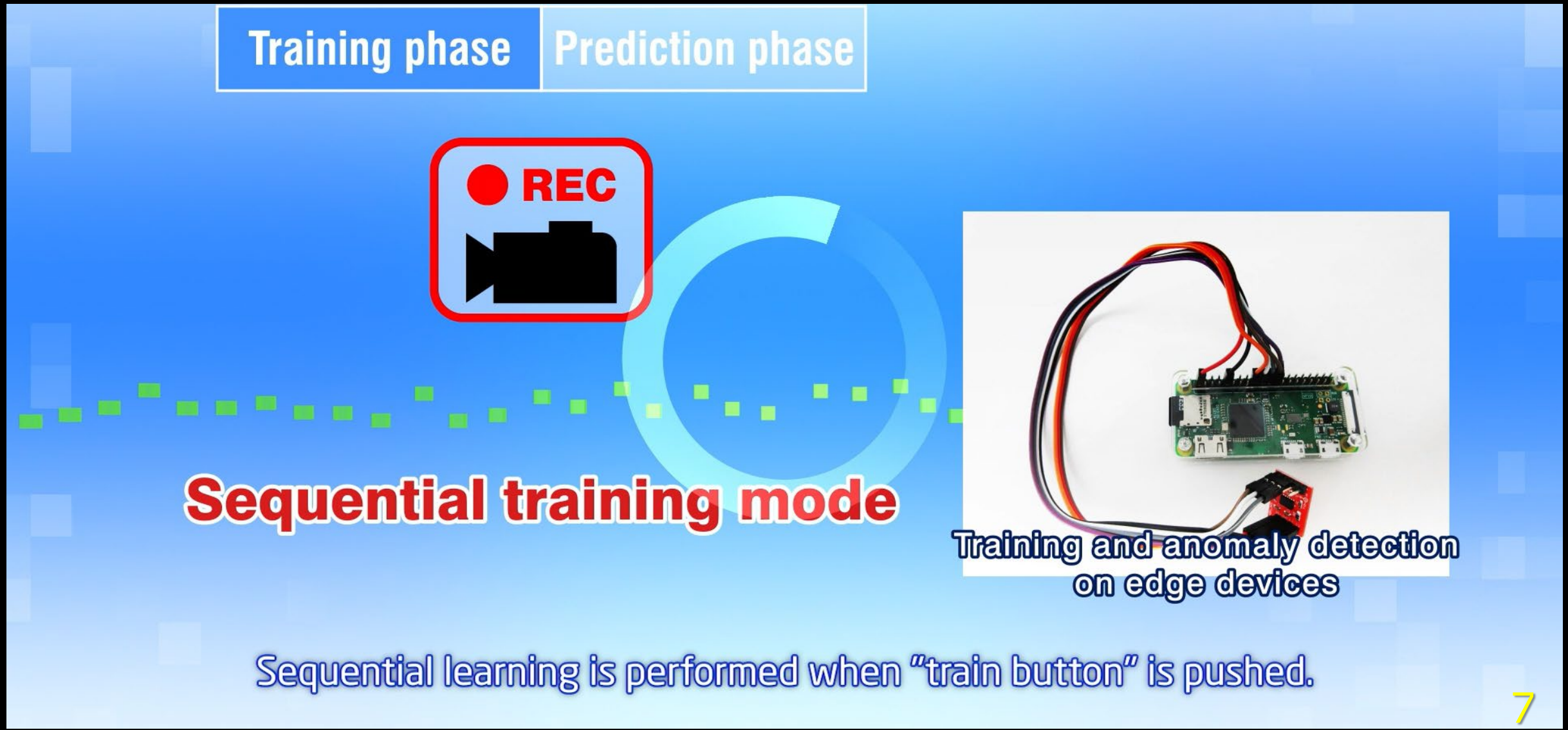




[1] Mineto Tsukada et al., "A Neural Network-Based On-device Learning Anomaly Detector for Edge Devices", IEEE Trans. on Computers (2020).

オンデバイス学習：必要性

- 現場ごとにセンサの値の見え方が違う → 現場でのオンデバイス学習



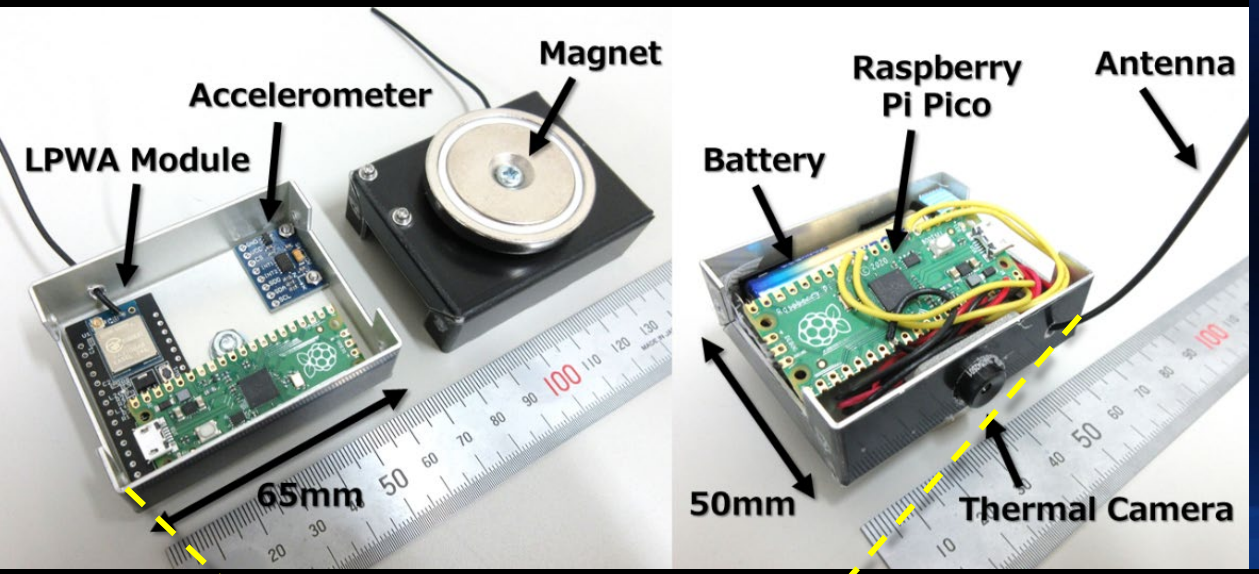
オンデバイス学習：基本的なアプローチ

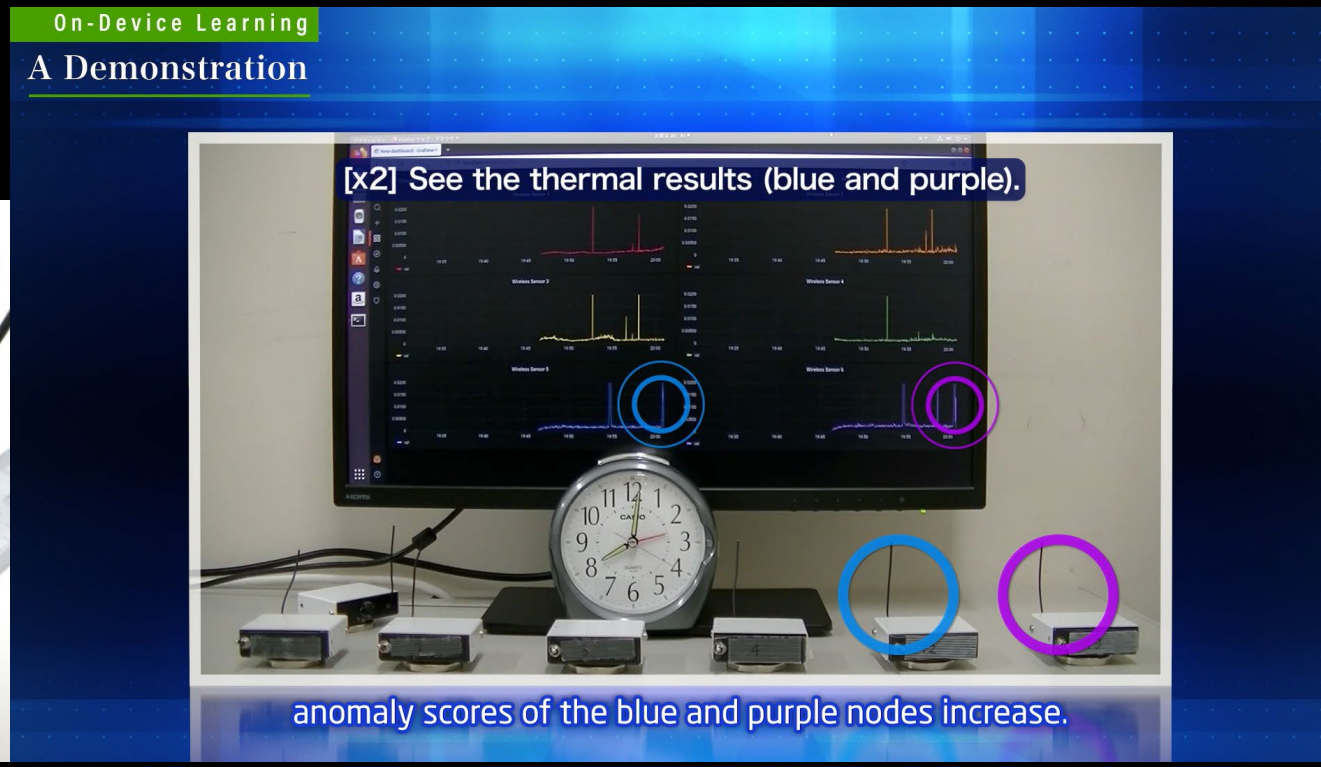
- **解析的重み計算** [1]
  ニューラルネットワークの逐次重み計算アルゴリズムを軽量化
- **パラメータ効率の良いファインチューニング** [2]
  誤差逆伝播＆勾配降下法をベースに順伝播と逆伝播を軽量化
- **同時摂動近似** [3]
  順伝播のみで重みパラメータを最適化
- **専用ハードウェア化**
  FPGA（Field-Programmable Gate Array）等で高速化

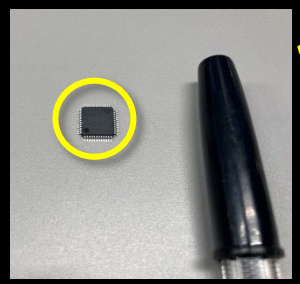
[1] N. Y. Liang et al., "A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks", IEEE Trans. on NN (2006).
[2] Edward J. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models", arXiv:2106.09685 (2021).
[3] Sadhika Malladi et al., "Fine-Tuning Language Models with Just Forward Passes", NeurIPS'23.

応用例：貼り付けるだけ異常検知器

- 置かれた現場でモデルを学習
  例：貼り付けるだけ異常検知器 [1]





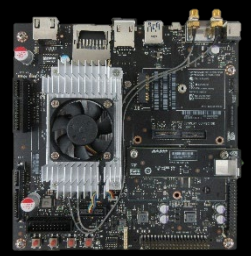


20-30mW

0.2-1W

2-10W

15W+

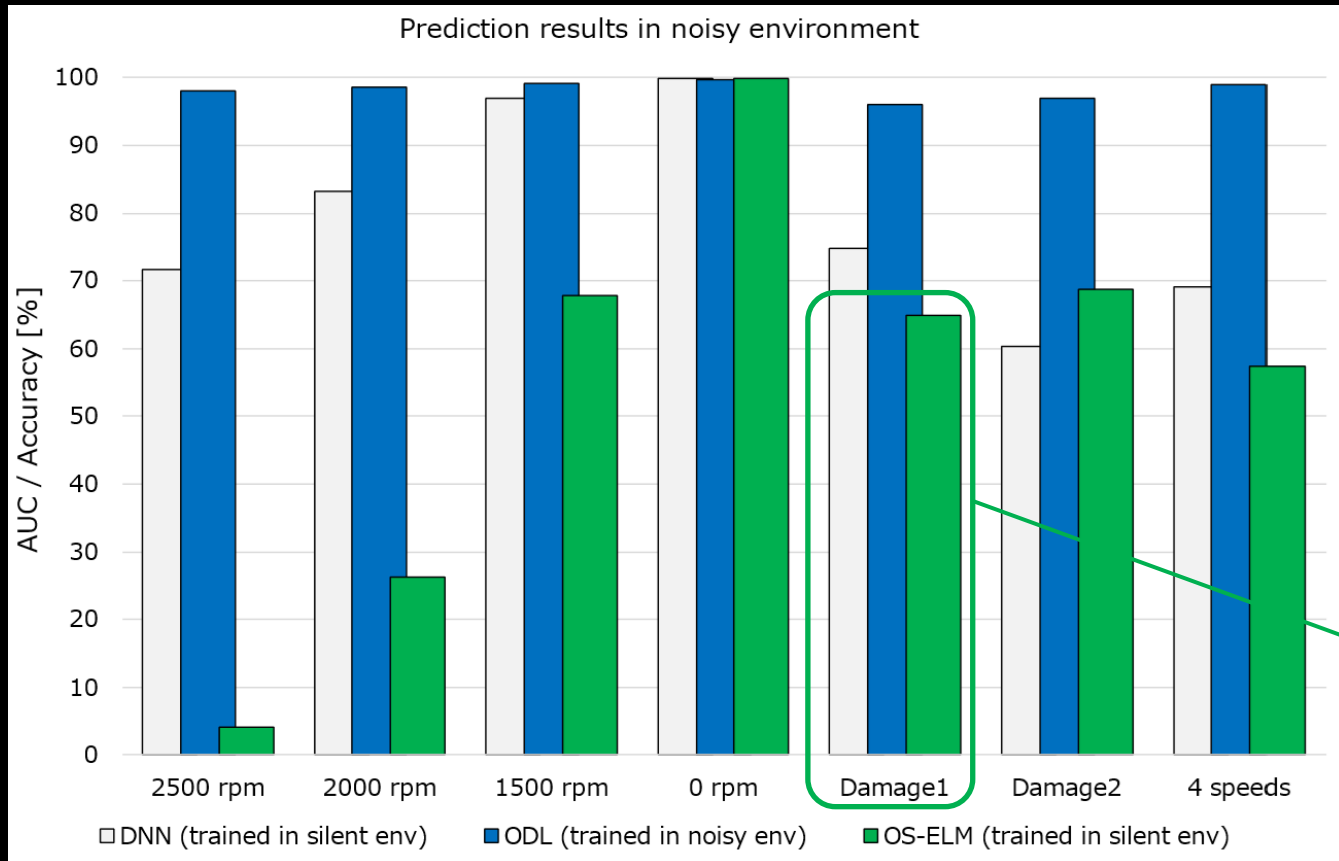


200W+

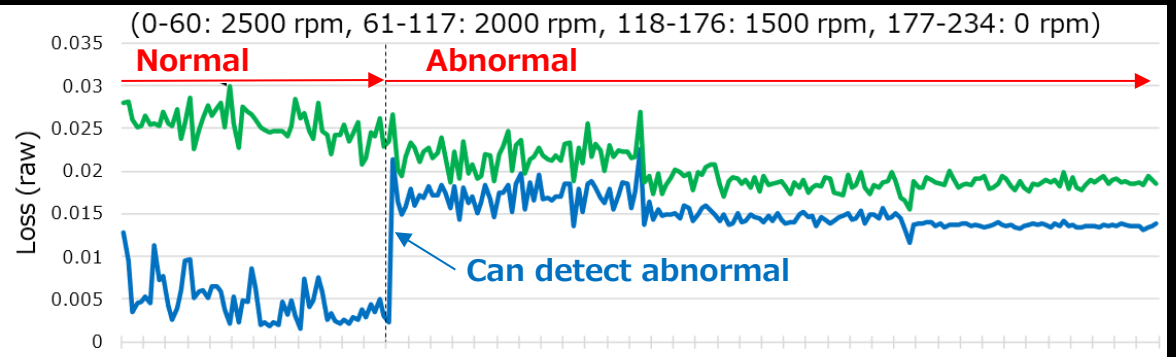
[1] Kazuki Sunaga et al., "Addressing Gap between Training Data and Deployed Environment by On-Device Learning", IEEE Micro (2023).

応用例：貼り付けるだけ異常検知器

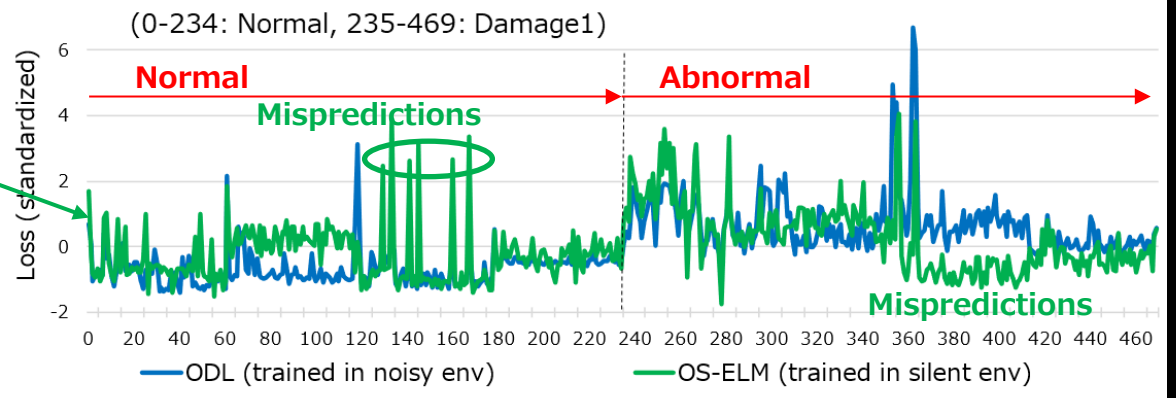
- 静音環境にて学習後、騒音環境に置かれて異常検知
  - 推論のみAI：騒音環境に適応できない → 精度が大幅に低下 ☹
  - オンデバイス学習：騒音環境に適応できる → 高い精度をキープ ☺



(a) Summary of seven tasks



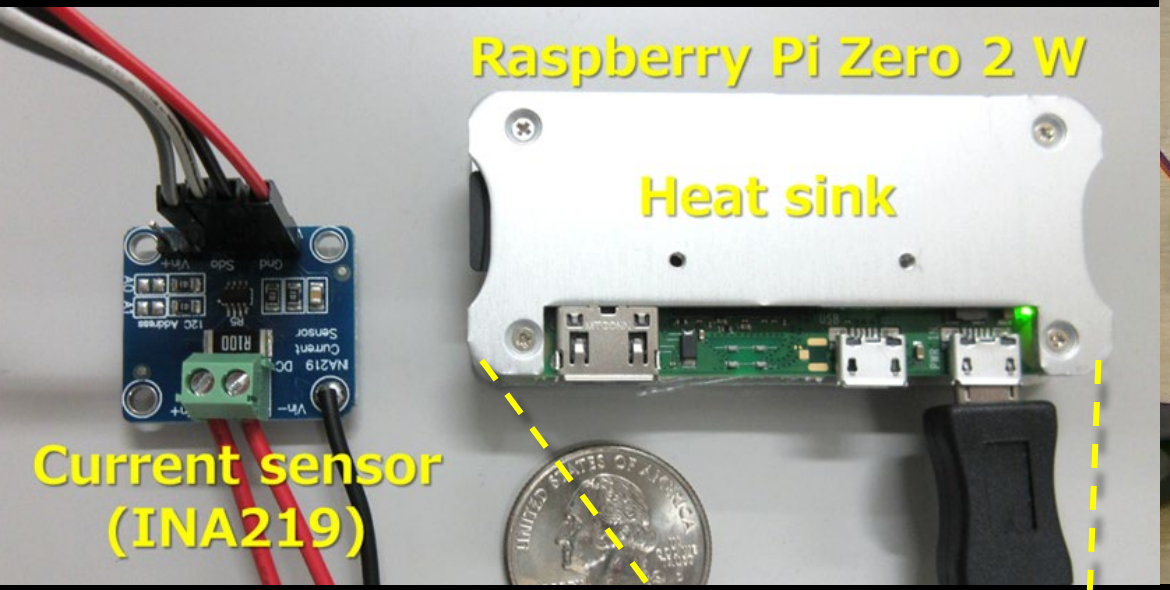
(b) Result of 2500 rpm task

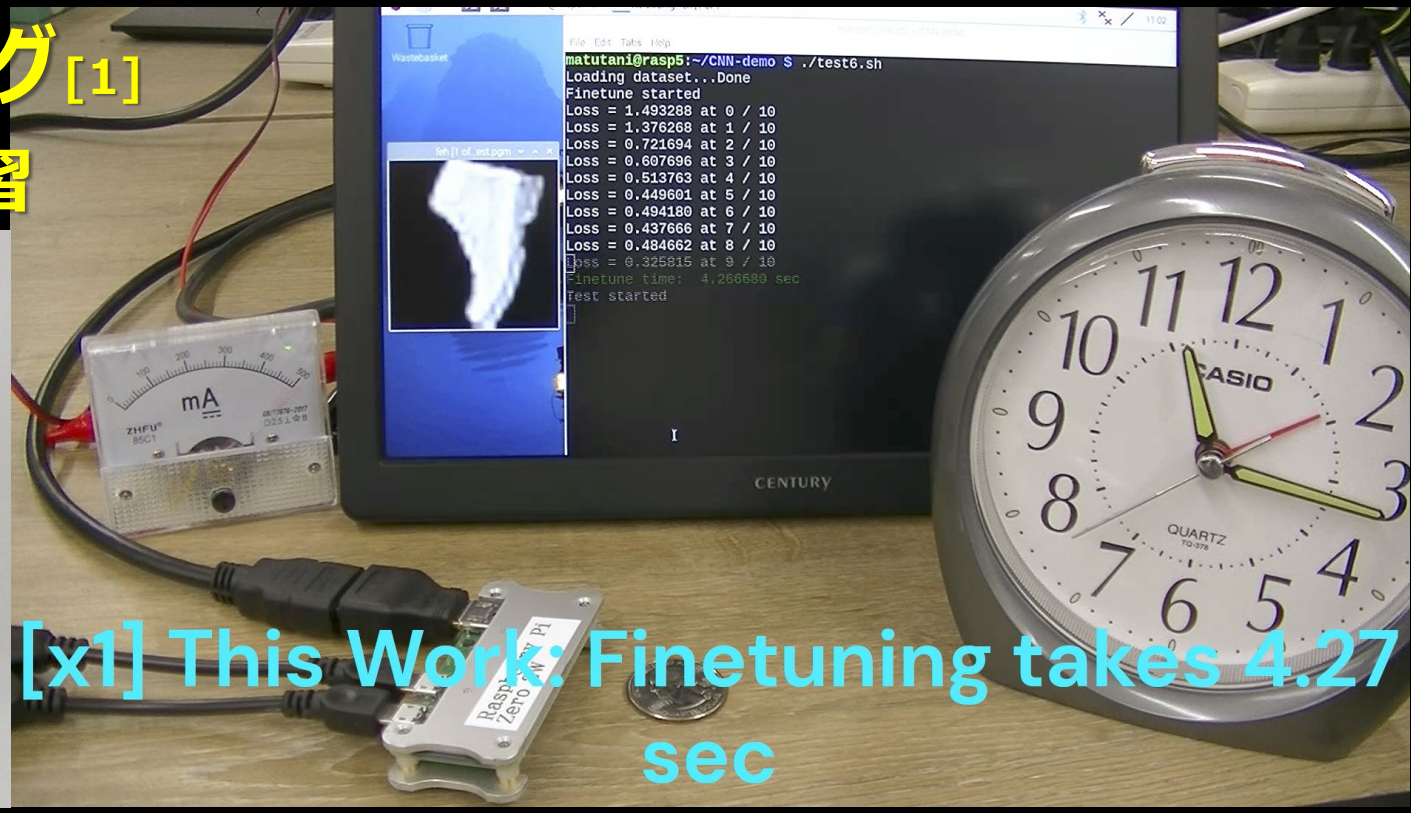


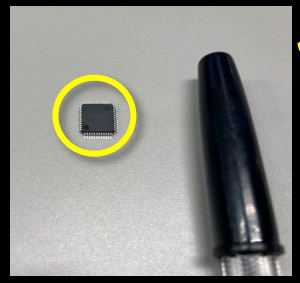
(c) Result of Damage1 task

応用例：画像認識AIのファインチューニング

- オンデバイスファインチューニング[1]
  エッジ向けDNNモデルの現場学習





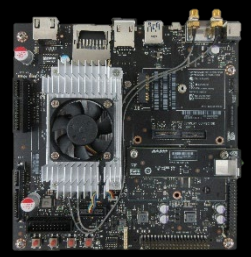


20-30mW

0.2-1W

2-10W

15W+

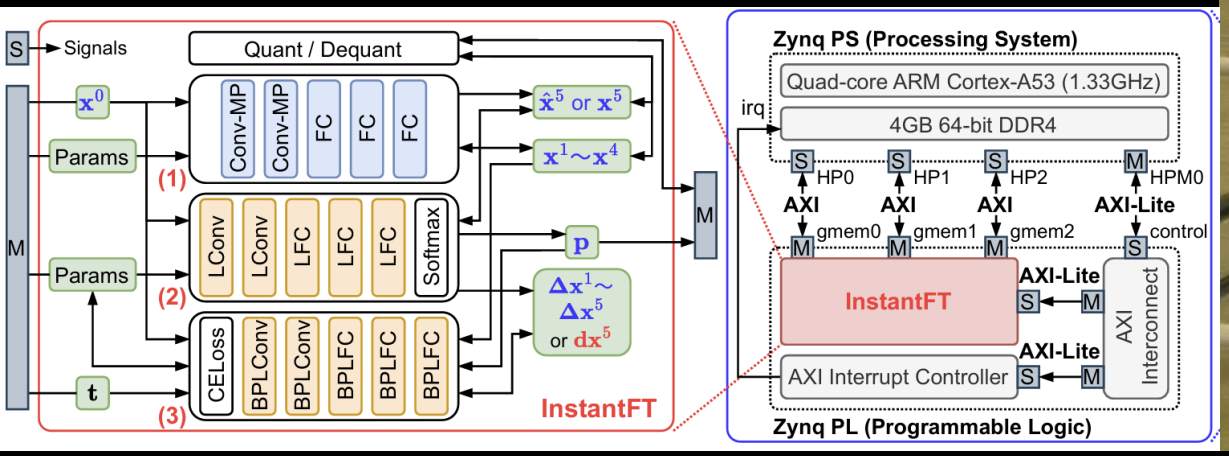


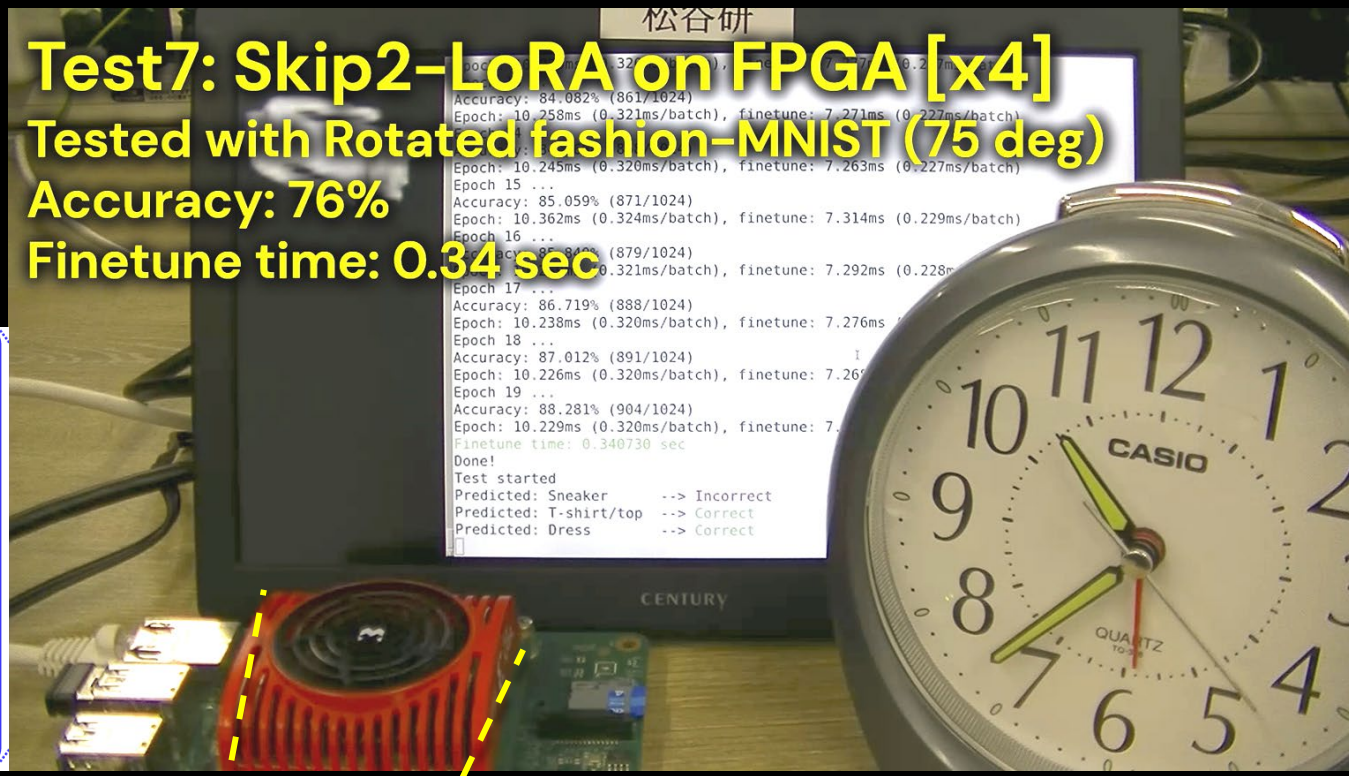
200W+

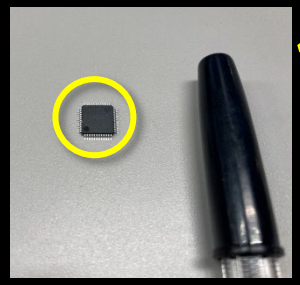
[1] Hiroki Matsutani et al., "Skip2-LoRA: A Lightweight On-device DNN Fine-tuning Method for Low-cost Edge Devices", ASP-DAC'25.

応用例：画像認識AIのファインチューニング

- 超高速ファインチューニング [1]
  FPGAによる専用回路化によって
  0.34秒でファインチューニング





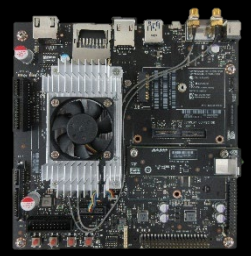


20-30mW

0.2-1W

2-10W

15W+

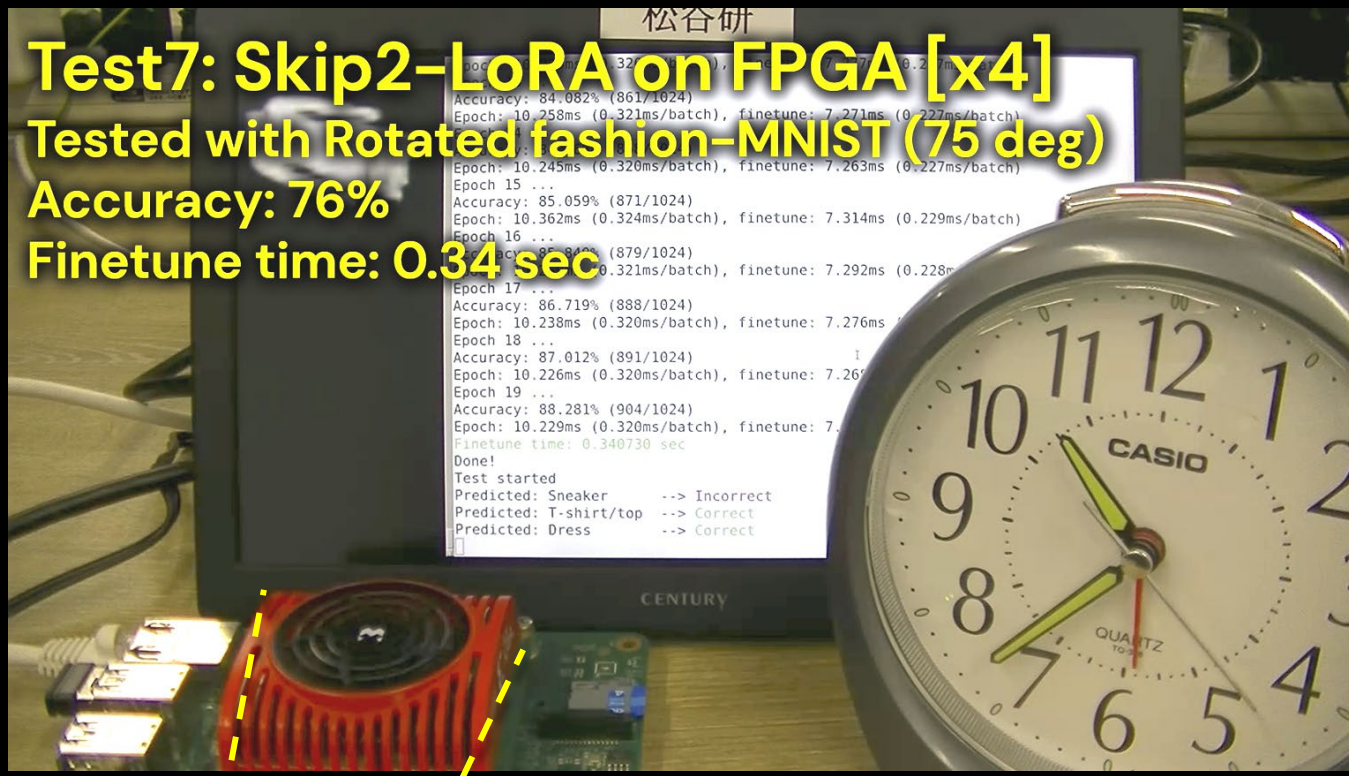


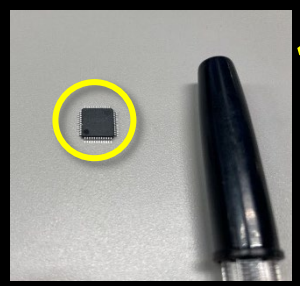
200W+

[1] Keisuke Sugiura et al., "InstantFT: An FPGA-Based Runtime Subsecond Fine-tuning of CNN Models", IEEE Micro (2026).

応用例：画像認識AIのファインチューニング

- 超高速ファインチューニング [1]
  FPGAによる専用回路化によって
  0.34秒でファインチューニング



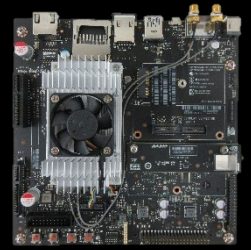


20-30mW

0.2-1W

2-10W

15W+



200W+

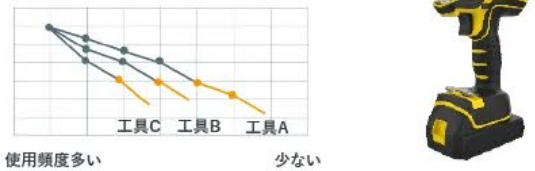
[1] Keisuke Sugiura et al., "InstantFT: An FPGA-Based Runtime Subsecond Fine-tuning of CNN Models", IEEE Micro (2026).

応用例：オンデバイス学習AIチップ

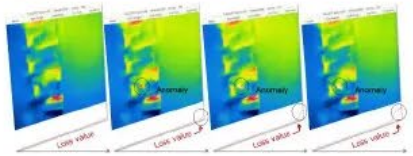
Solist-AI

※「Solist-AI™」はローム株式会社の商標または登録商標です。

定速運転での正常加速度波形を学習し、軸受損傷をパターン検知



個々のバッテリー状態を追加学習し、残量や劣化を高精度に推定



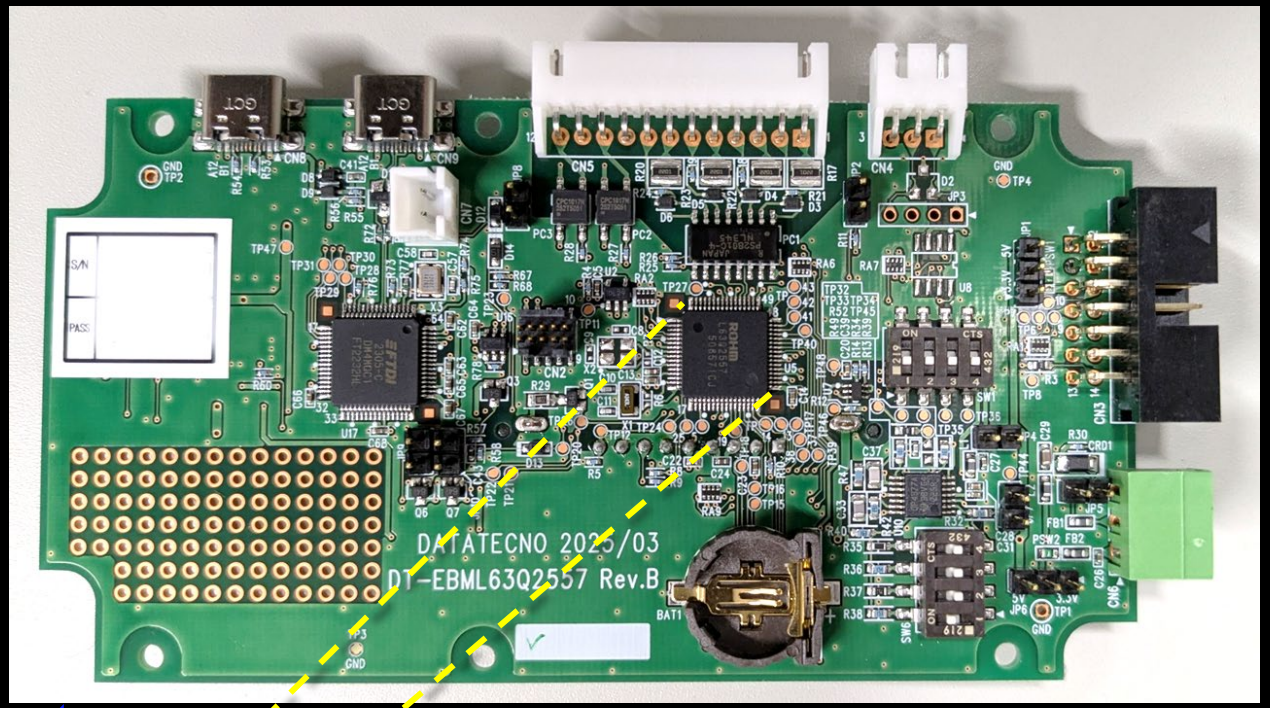
AIによって特異点での異常発熱を検知し、異常度の上昇を検出

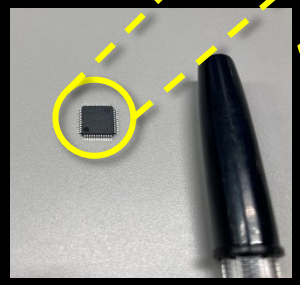
- 不良品選別
  人間の感覚による処理を自動化
- 測定環境へのマッチング
  最適なパラメータを現場で更新
- 劣化予知
  耐用年数の長いインフラ設備の保全
- 介護補助
  独居老人、寝たきりの人の見守り

[1] ローム株式会社, "エッジコンピューティング向け完全自立型AIソリューション Solist-AI™", https://www.rohm.co.jp/support/solist-ai.
[2] 日本経済新聞, "ローム、AI搭載マイコンを開発　ネット使わず自動学習", 日本経済新聞電子版（2025年3月18日）.

応用例：オンデバイス学習AIチップ

- オンデバイス学習機能付きAIチップ [1]



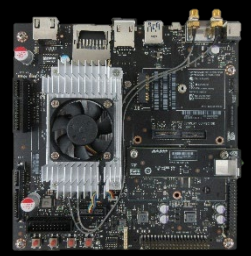


20-30mW

0.2-1W

2-10W

15W+

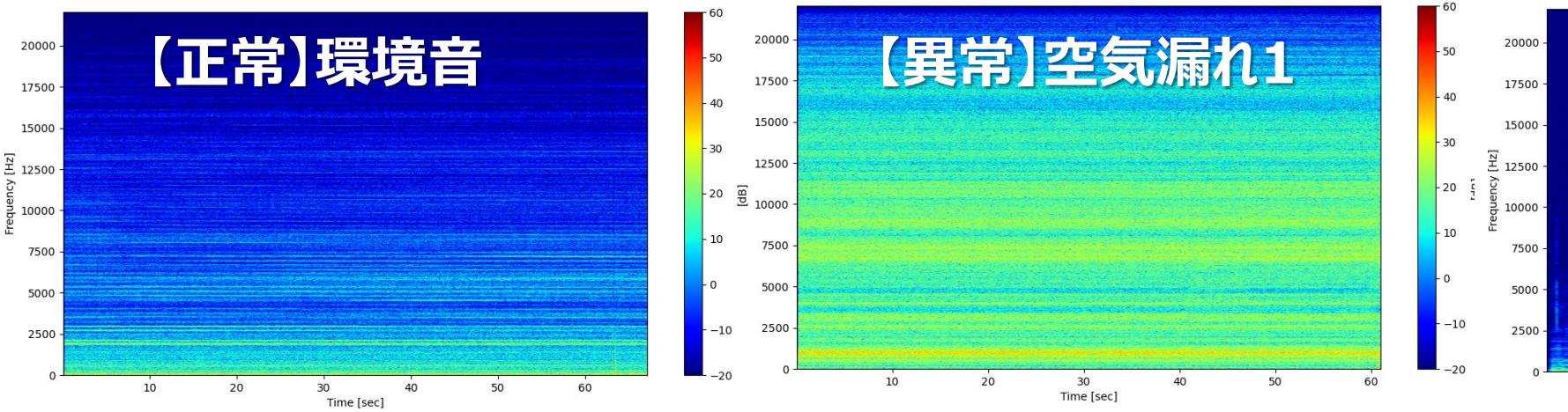


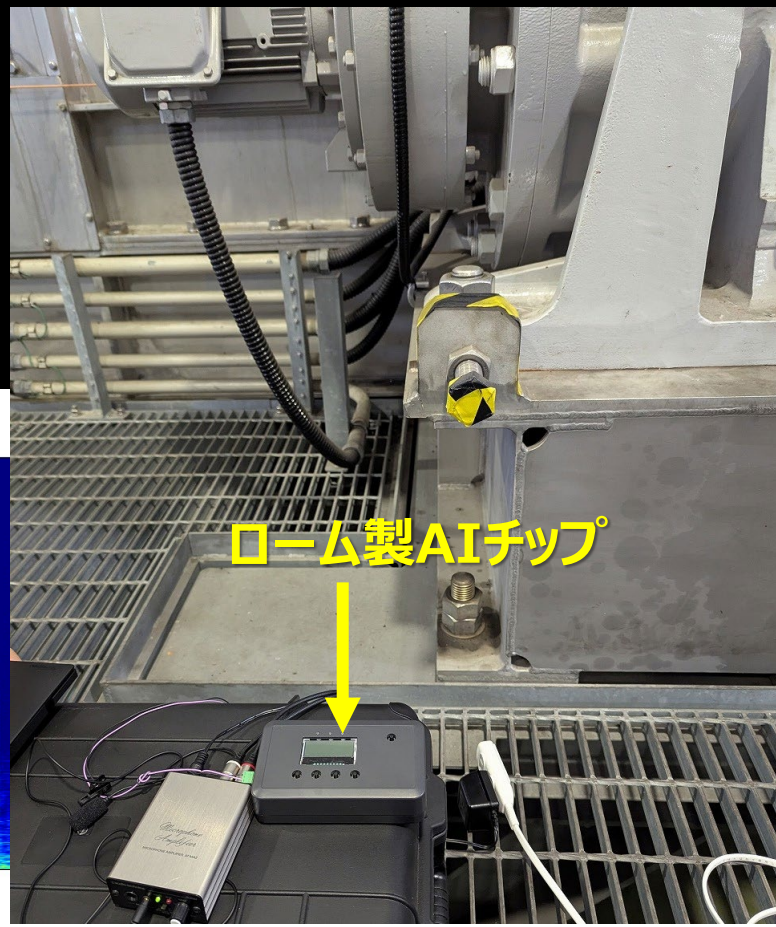
200W+

[1] ローム株式会社, "エッジコンピューティング向け完全自立型AIソリューション Solist-AI™", https://www.rohm.co.jp/support/solist-ai.

応用例：上下水道施設の異常検知

- オンデバイス学習を用いたポンプ室の異常検知
  - 雑音環境下において、空気漏れを検出
  - 正常音を雑音込みでオンデバイス学習 → 異常検知





正常音のみ学習（ベースライン）

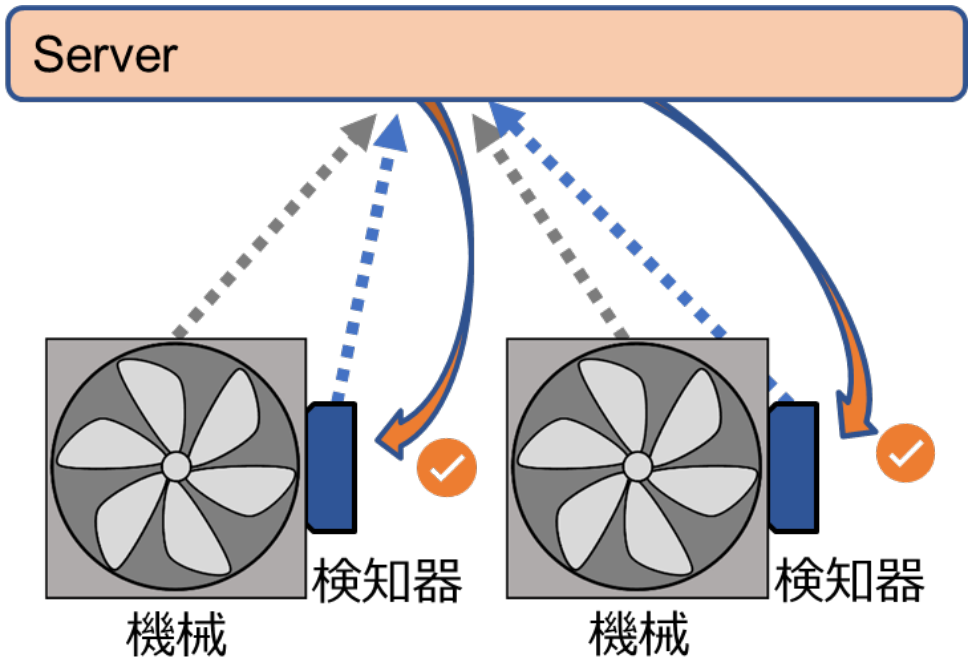
	正常音と判定	異常音と判定
実際の正常音	681	1691 誤検知
実際の異常音	2 見逃し	710

現場の雑音をオンデバイス学習（提案手法）

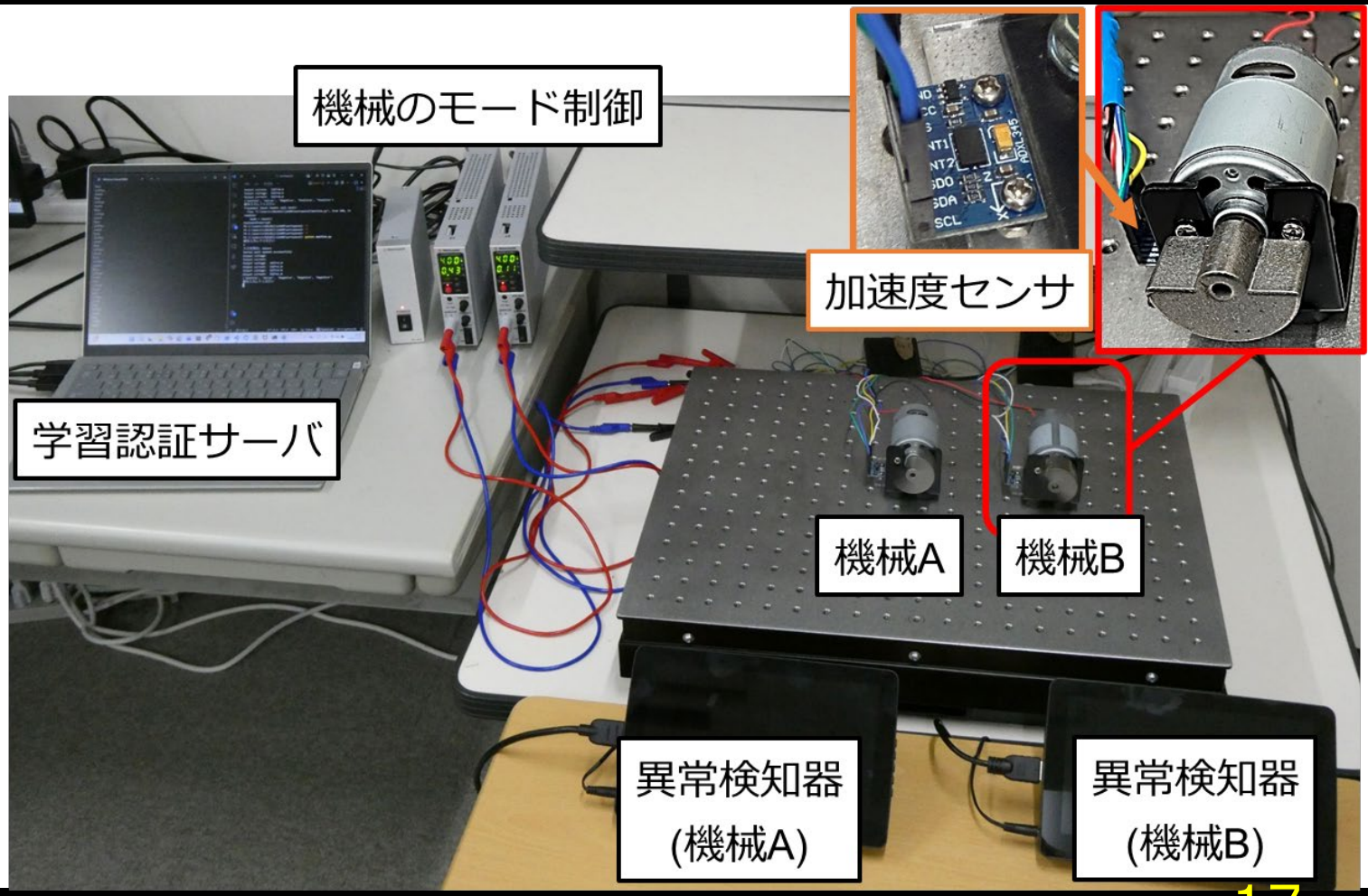
	正常音と判定	異常音と判定
実際の正常音	1016	58 誤検知
実際の異常音	5 見逃し	716

オンデバイス学習：意図せぬ追加学習への対策

- **分散合意：オンデバイス学習の「安価にバラまける」という特徴を活用**
  複数オンデバイス学習器の合意によって追加学習をトリガ [1]



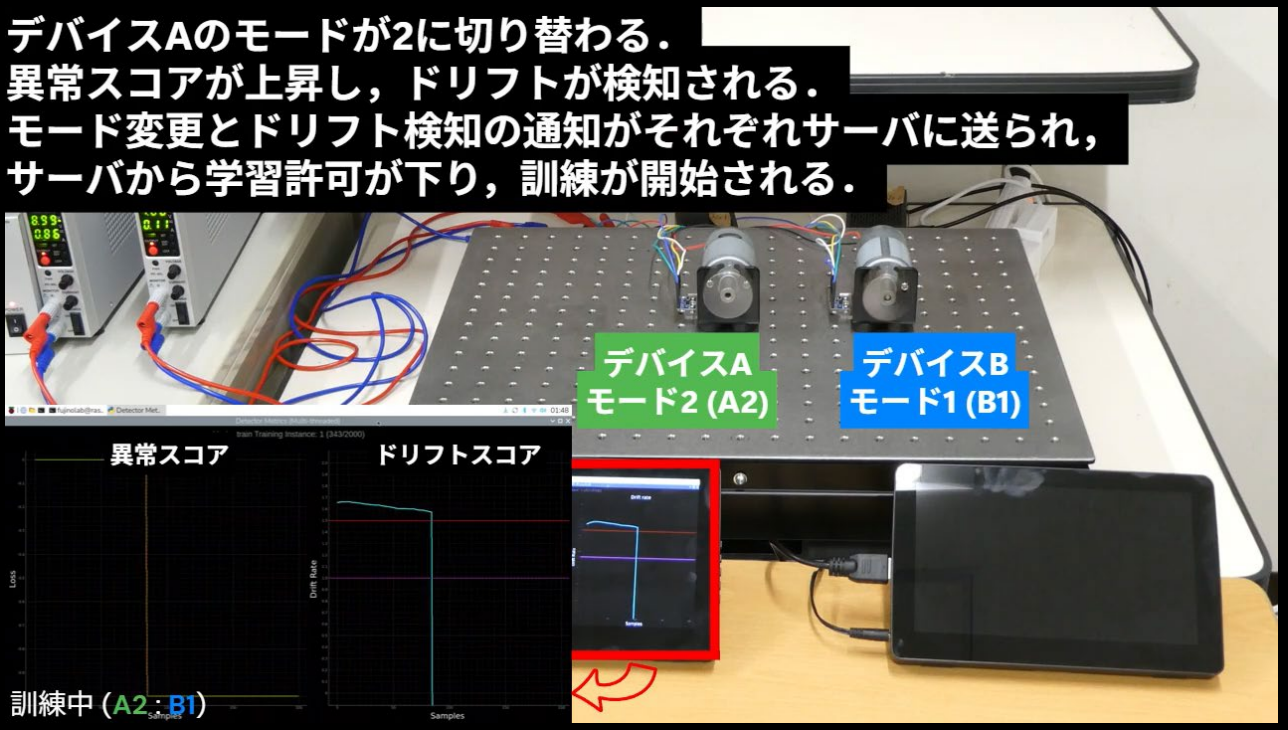
仮定：環境起因のドリフトなら、同一空間内の
他検知器においてもドリフトが検出されるはず

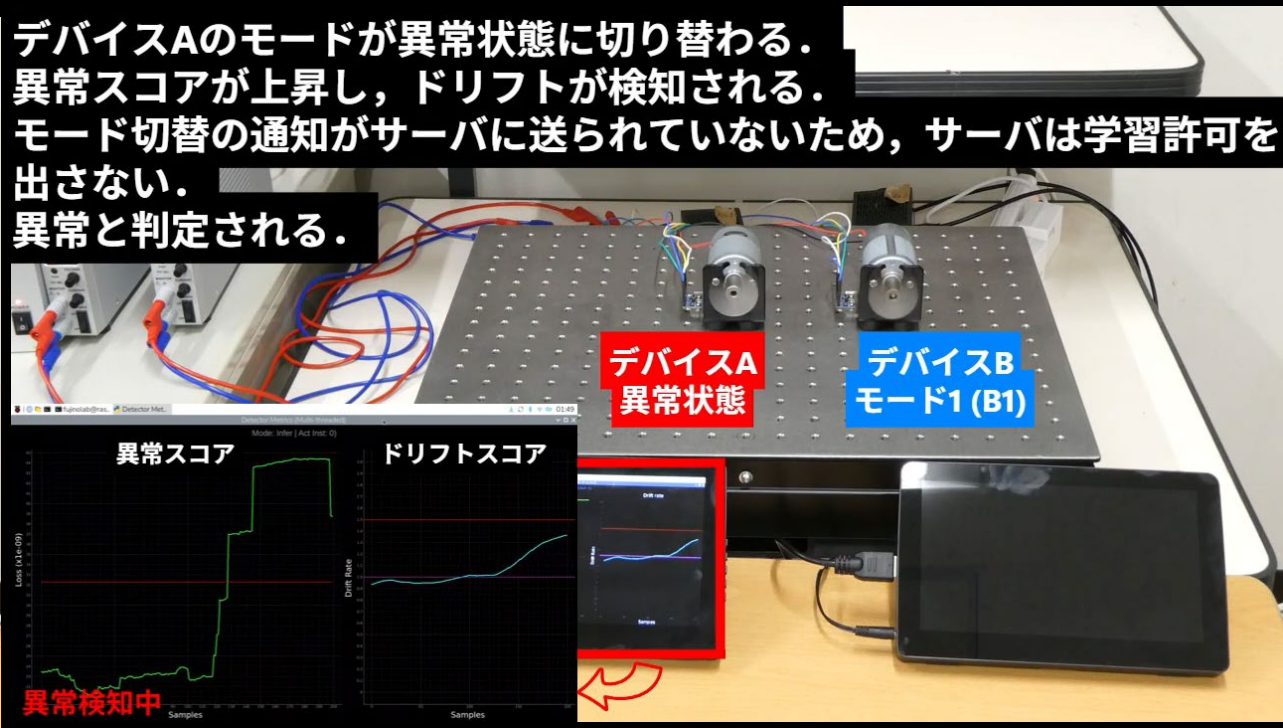


[1] Takahito Ino et al., "Mitigating Data Poisoning Attack in On-Device Learning Anomaly Detectors via Peer Consensus", AIcompS'25.

オンデバイス学習：意図せぬ追加学習への対策

- 分散合意：オンデバイス学習の「安価にバラまける」という特徴を活用
  複数オンデバイス学習器の合意によって追加学習をトリガ [1]





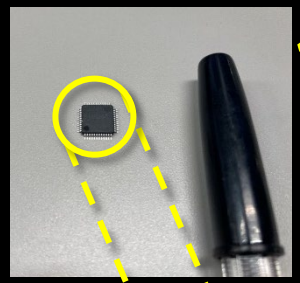
追加学習が許可されるケース（既知のドリフト）

追加学習が許可されないケース（異常）

[1] Takahito Ino et al., "Mitigating Data Poisoning Attack in On-Device Learning Anomaly Detectors via Peer Consensus", AIcompS'25.

オンデバイス学習：保有技術と応用ドメイン

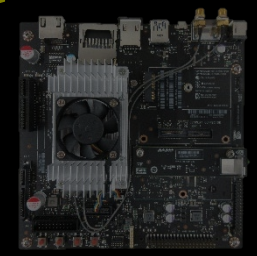
IoT　Mobile　Edge server　Cloud server


20-30mW

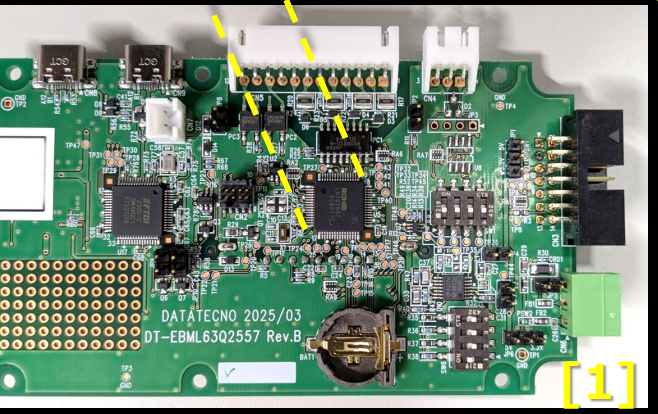

0.2-1W


2-10W

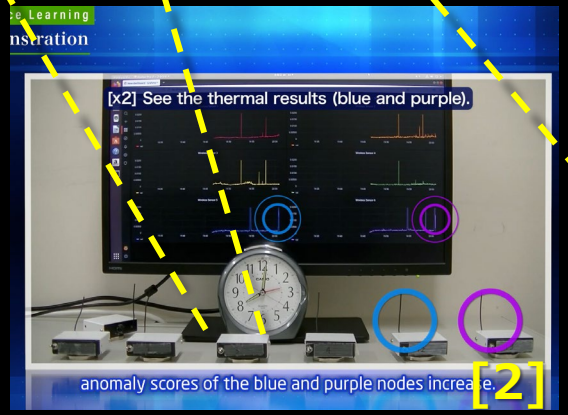

15W+


200W+

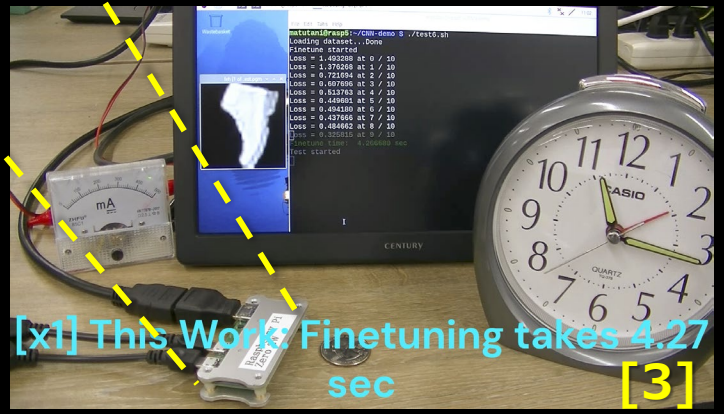
コントローラ　組込みCPU　モバイルCPU　組込みGPU　高性能GPU



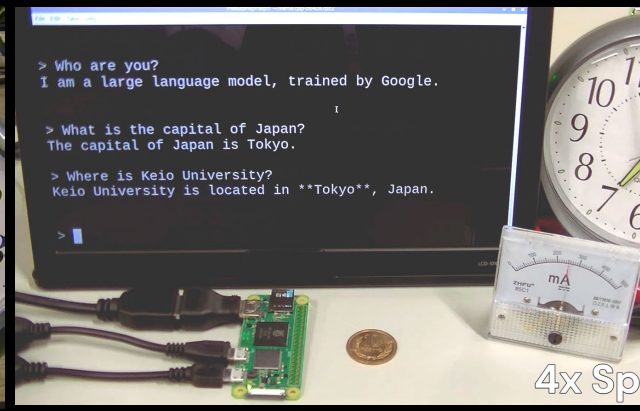
[1]



[2]



[3]

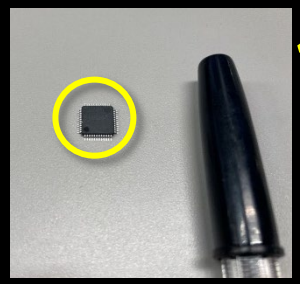


信号処理用NN　画像認識用CNN　エッジ用LLM

[1] ローム株式会社, "エッジコンピューティング向け完全自立型AIソリューション Solist-AI™", https://www.rohm.co.jp/support/solist-ai.
[2] Kazuki Sunaga et al., "Addressing Gap between Training Data and Deployed Environment by On-Device Learning", IEEE Micro (2023).
[3] Keisuke Sugiura et al., "InstantFT: An FPGA-Based Runtime Subsecond Fine-tuning of CNN Models", IEEE Micro (2026).

オンデバイス学習：まとめと今後の方向性

IoT　Mobile　Edge server　Cloud server


20-30mW


0.2-1W


2-10W


15W+


200W+

コントローラ　組込みCPU　モバイルCPU　組込みGPU　高性能GPU

エッジAI　　クラウドAI

クラウドへの一極集中

⇩

エッジに付加価値を持たせる

データや計算をデータセンタに集約

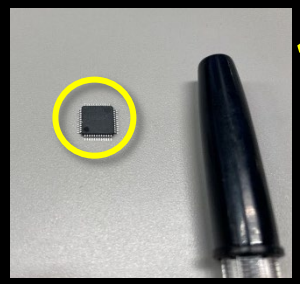
強力なGPU（最先端の半導体技術）

エッジはセンシングだけ

エッジの付加価値は小さい

オンデバイス学習：まとめと今後の方向性

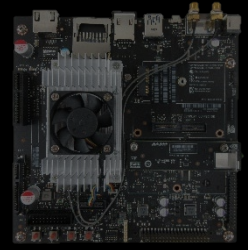
IoT　　Mobile　　Edge server　　Cloud server


20-30mW


0.2-1W


2-10W


15W+


200W+

コントローラ　　組込みCPU　モバイルCPU　　組込みGPU　　高性能GPU

エッジAI

クラウドへの一極集中

⇩

エッジに付加価値を持たせる

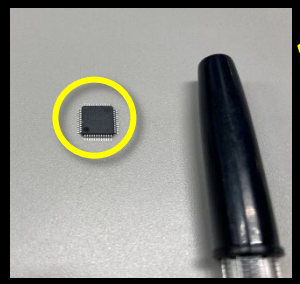
従来のエッジAI
現場では推論のみ（モデルはサーバで学習しておく）

オンデバイス学習
現場で学習もできる（モデルを現場で更新できる）

例：ユーザさんが自分の環境に合わせて、
ボタン１つでモデルの構築まで出来たら便利では？

最後に：本研究がもたらす未来イメージ

IoT　　Mobile　　Edge server　　Cloud server



20-30mW

0.2-1W

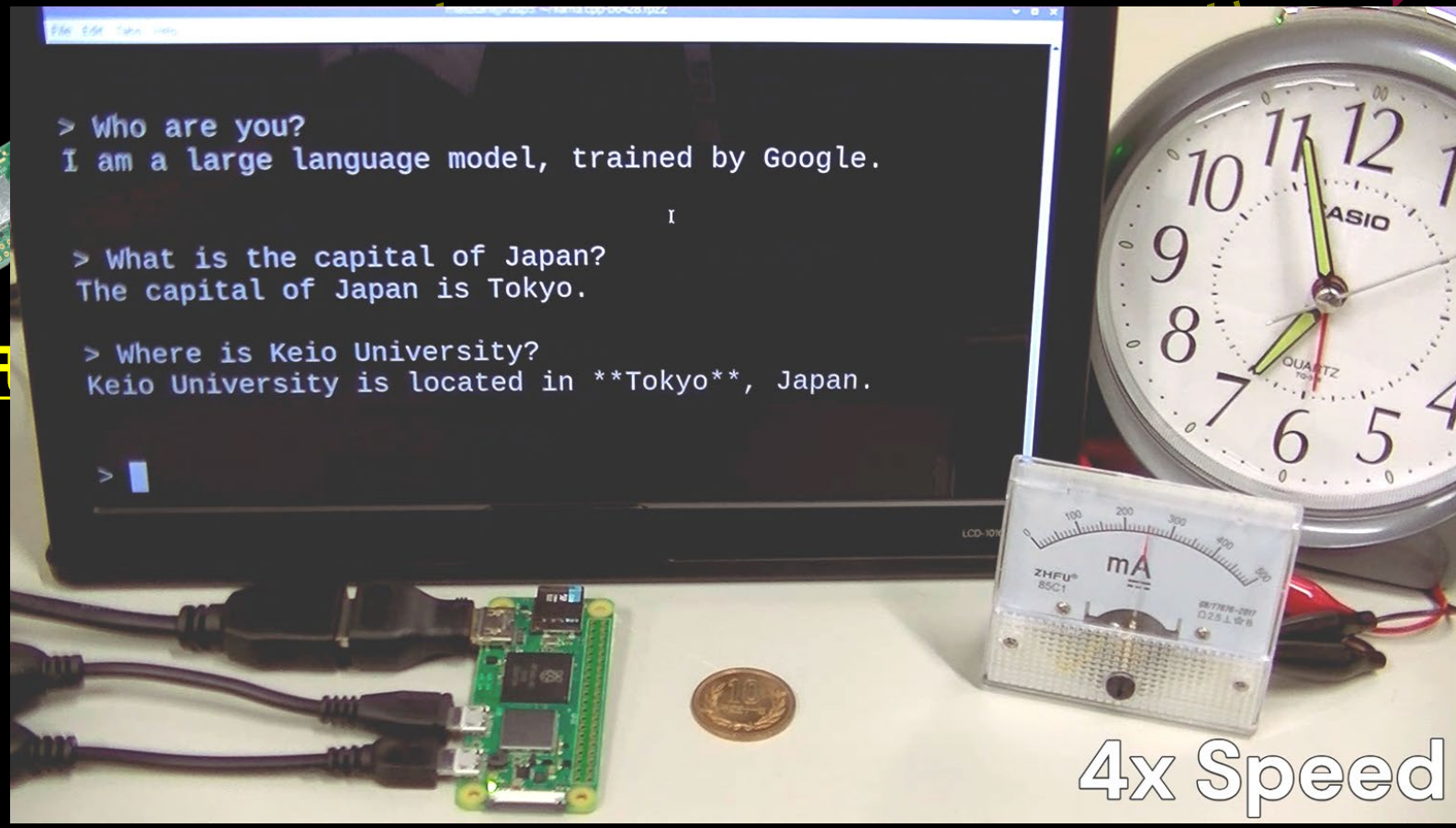
コントローラ　　組込みCPU　　モ

エッジAI

クラウドへの一極集中

⇩

エッジに付加価値を持たせる

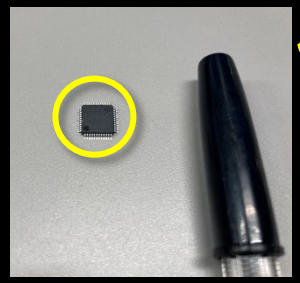


今後の方向性：
Physical AIに資する高度なAI技術開発

[1] Hiroki Matsutani et al., "Accelerating Local LLMs on Resource-Constrained Edge Devices via Distributed Prompt Caching", arXiv Preprint (2026).

最後に：本研究がもたらす未来イメージ

IoT　　Mobile　　Edge server　　Cloud server

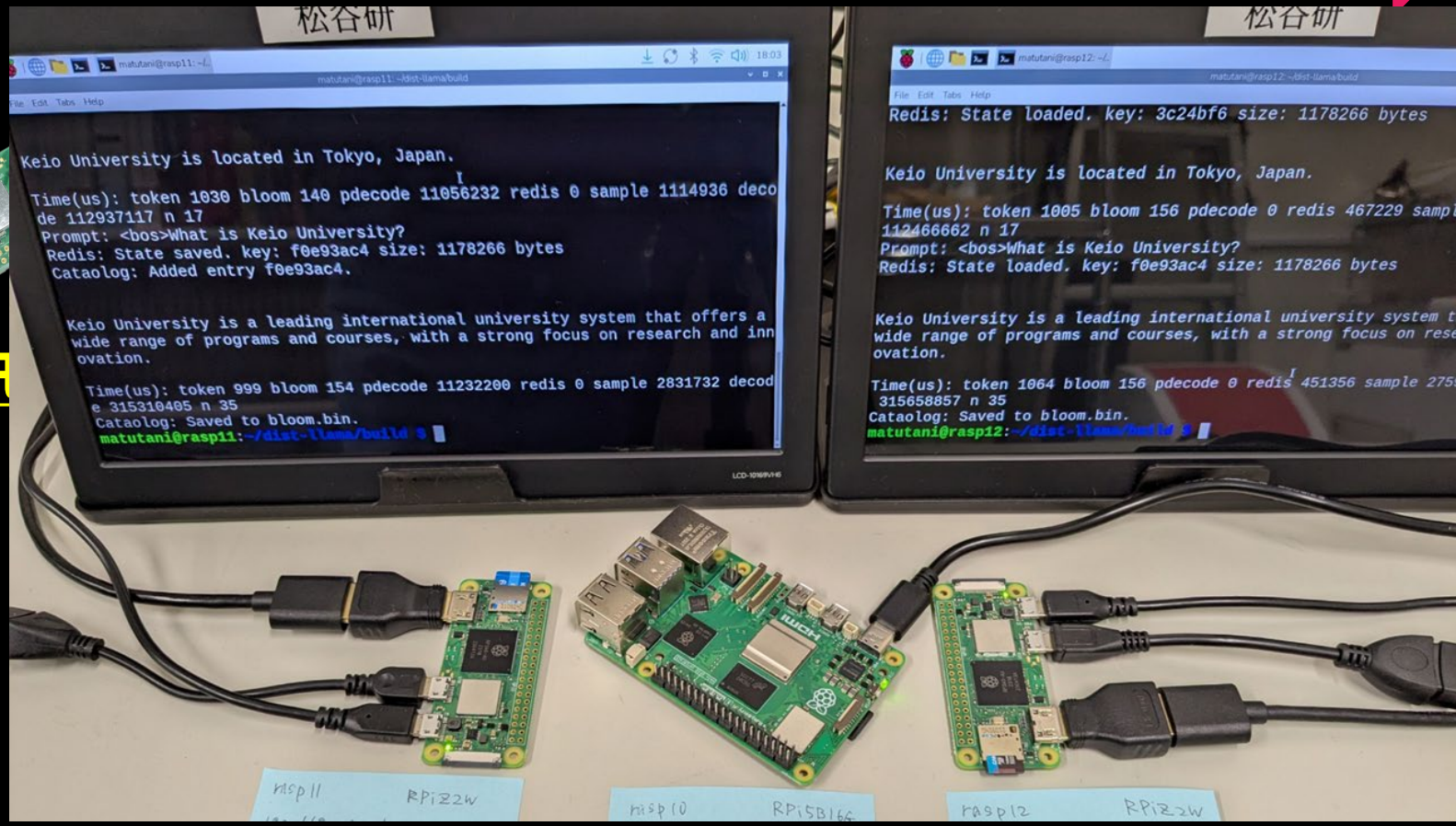


20-30mW

0.2-1W

コントローラ　　組込みCPU　　モ

エッジAI



クラウドへの一極集中

⇩

エッジに付加価値を持たせる

今後の方向性：
Physical AIに資する高度なAI技術開発

[1] Hiroki Matsutani et al., "Accelerating Local LLMs on Resource-Constrained Edge Devices via Distributed Prompt Caching", arXiv Preprint (2026).