

# IBMの最新AI関連研究

東京基礎研究所が取り組む

「大規模深層学習基盤」「画像異常検出」「音響解析」技術のご紹介

IBM東京基礎研究所では、AI用ハードウェアから、AIの理論、高速化、アプリケーションまで、さまざまな分野において、お客様の課題を意識したエンタープライズAIの研究を行っています。それらの研究には、成果を製品としてお客様にお届けするものも、お客様の課題にあてはめて応用し研究員が直接価値を提供するものもあります。本稿ではそれらの中から具体例として、「大規模深層学習を実現する実行基盤」「画像の注目領域を加味した異常検出」「AIを活用した音響解析」を紹介します。

## ▶▶ 1. はじめに

IBM東京基礎研究所では、今日のAIブームが訪れる前の“AIの冬”と呼ばれた時代にも、非構造データの処理や最適化、機械学習といった、今で言うAI関連技術を連綿と研究し続けてきました。AIの第3次ブームにある今日、それらの研究はWatson関連製品やお客様ソリューションの中で開花しています。東京基礎研究所で扱っているAI関連技術は、AIチップ、機械学習の理論、深層学習フレームワークの高速化、自然言語処理、音声認識、画像認識などの多岐にわたっています。それらの研究開発のステージはプロジェクトにより大きく異なっており、実用化まで時間のかかる基礎的な研究もあれば、翌年に製品としてのリリースを目指す研究開発、お客様の課題に直接研究員が取り組んでいるプロジェクトもあります。どの分野のどのプロジェクトにも共通して言えることは、お客様企業の課題を意識したエンタープライズAIの研究であるということです。具体的には、例えばお客様が所有しているビジネスデータを利用して業務効率化を行いたいがそのデータの分量が限られている、といった課題があります。

本稿では、東京基礎研究所のAI関連研究の中から以下

の3つを紹介します。

- ①通常の方法では学習が困難な大規模なモデルの深層学習を実現する実行基盤の開発
- ②起こりうる異常のパターン全てを予期することはできないが、典型的な異常パターンは既知であるという状況において、少量の既知異常データを用いて異常検出の精度を改善する画像異常検出技術
- ③高齢者の生活の見守り、溶接工程の異音検出、熟練工によって行われていた打音検査の自動化などに応用できる音響解析技術

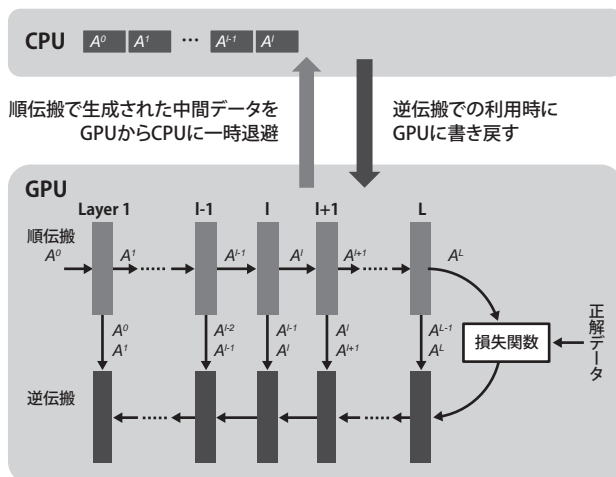


図1. データ・スワッピング手法

## 2. 大規模深層学習を実現する実行基盤

深層学習は画像認識、音声認識などさまざまな分野において、既存手法を圧倒する精度を達成し続けています。そのさらなる精度向上には、より深く大規模な学習モデルが必要不可欠となっており、大規模画像認識の競技会である「ImageNet Large Scale Visual Recognition Challenge (ILSVRC)」[1]においても、大規模なモデルを利用したチームが優勝しています。また実応用においても、医療画像処理での3次元画像や高解像度の衛星画像の利用などにより、今後も学習モデルの大規模化は継続するものと考えられます。

深層学習で広く用いられているGPUの演算能力は、新しいデバイスが出るたびに飛躍的に向上していますが、メモリー容量は演算能力ほど向上していません。GPUメモリーに利用されているHigh Bandwidth Memory (HBM)は高価なため、今後はメモリー容量の飛躍的な増加が困難と考えられています。そのため、大規模深層学習の実行にはGPUメモリー容量の制約がますます大きな課題となります。

IBM東京基礎研究所ではその課題を解決する手法として、「データ・スワッピング手法」に着目して研究開発を行っています。データ・スワッピング手法は、学習過程において生成された中間データを一時的にCPUメモリーに退避し、利用するときにGPUへ書き戻すことにより、GPUメモリーの効率的な利用を実現します(図1)。

われわれはこの手法をTensorFlowで実現するコ

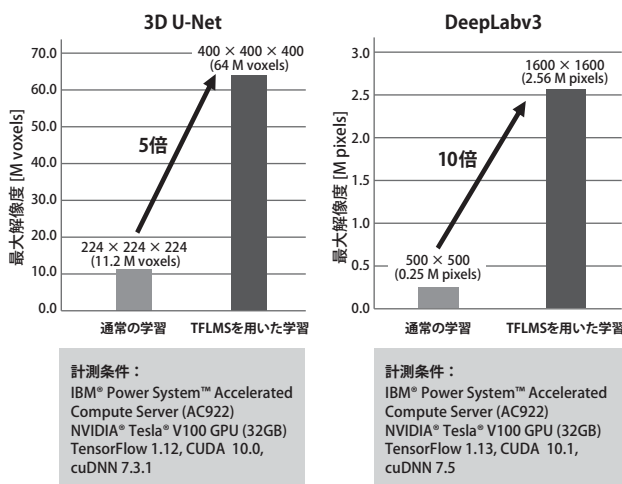


図2. 学習可能な解像度の向上

フトウェア「TensorFlow Large Model Support (TFLMS)」を開発しました[2]。この開発において課題はCPU-GPU間の通信による実行性能の低下でしたが、転送するメモリー量やタイミングを適切に管理する性能最適化パラメーターの自動チューニング機能を導入することで、実行性能の低下を回避し高速な学習を実現しています。

次に、TFLMSを利用した大規模深層学習モデルの実行事例として、高解像度画像を用いた画像のセグメンテーションの事例を示します。セグメンテーションとは画像の各ピクセルをカテゴリ分類することで画像に含まれる物体の種類、位置、形状を認識する技術であり、医療画像における病変の認識や自動運転時の道路状況の認識など、さまざまな分野で利用されています。

図2はTFLMSを用いることによって学習可能となる最大解像度を示しています。3D U-NetやDeepLabv3はそれぞれ3次元画像、2次元画像の高性能なセグメンテーションを実現するために広く用いられているモデルです。図に示しているように、TFLMSを用いることで、それぞれ5倍、10倍の解像度の画像が学習可能となります。このような高解像度の画像を用いることによって、より正確なセグメンテーションが実現可能となります。

図3は3D U-Netを用いた3次元の脳MRI画像の脳腫瘍のセグメンテーションの結果を示しています。通常は高解像度の3次元MRI画像を何枚かの画像に分割してから学習や推論に利用することで使用GPUメモリーを削減しますが、このような画像分割は精度の悪化を招きます。それに対し、TFLMSを用いれば元の画像を分割す

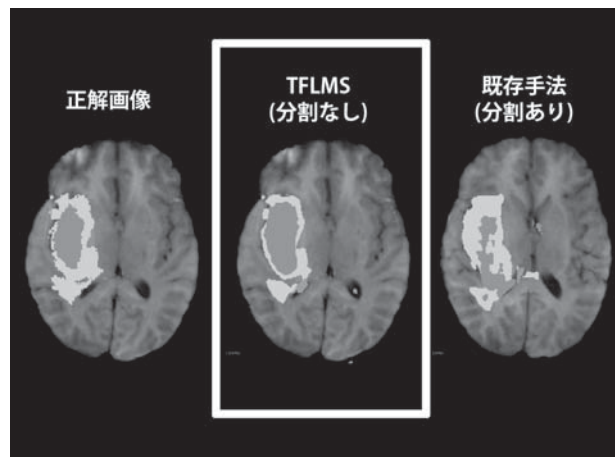


図3. 脳腫瘍のセグメンテーションの結果

ることなく学習することが可能になり、より良いセグメンテーションが実現されます[3]。

本章では、大規模深層学習を実現する実行基盤としてTFLMSを紹介しました。TFLMSはIBMが提供するディープ・ラーニング・プラットフォーム「IBM Watson Machine Learning Accelerator」で提供されています[4]。GPU搭載の高速サーバーである「IBM Power System AC922」は、CPU-GPU間に高速な接続(NVLink 2.0)を備えているためTFLMSのさらなる高速化が可能となります。IBM Watson Machine Learning Acceleratorには、ほかにも高速な分散深層学習を実現する「IBM Distributed Deep Learning (DDL)」など、深層学習を支える実行基盤を提供しています。われわれは今後もそれらと連携し、大規模深層学習を実現する実行基盤の開発を推進していきます。

### ▶▶ 3. 注目領域を加味した画像の異常検出

本章では、画像認識の一つの応用である異常検出に関する研究事例を紹介します。近年、画像情報を活用したさまざまな異常検出手法が提案されており、工場での不良品の自動検出、異物や危険人物の監視画像からの検出、前述のような医療画像における病変検知など、多様な分野へ応用されています[5]。以下ではその一例としてわれわれが行った研究を紹介します。

この研究では、認知症の簡易的なテストとして用いられる「山口キツネ・ハト模倣テスト」[6]の自動化の試みに異常検出を応用しました。このテストは、被験者にハトの

ポーズ(図4)を模倣するように指示し、その模倣の成否によって認知症かどうかを簡易的に調べるために医師が行っています。これを画像の異常検出の問題に当てはめると、正解となるハトのポーズの画像が正常となり、不正解となるさまざまなポーズの画像はすべて異常です。

異常検出の問題に一般的に適用される方法は、正常と異常の画像を大量に集め、それらを判別する識別器を学習する方法です。識別器には、深層学習を応用した手法[7]などが考えられます。しかしこれらの手法では、起こりうる全ての異常パターンを事前に知り得ていなくてはならず、本質的な未知の異常を検出することは不可能です。

その点を解決するために、自己符号化器(Auto-encoder)を利用する手法が提案されています。自己符号化器は、与えられた学習データを復元することを目的とする学習器です。正常画像のみから学習した自己符号化器には、正常画像を復元する際には誤差が小さいのに対し、異常画像の復元の際には誤差が大きくなるという性質があります。これを応用した異常検出では、テスト画像と復元画像の誤差が十分小さい場合に正常、大きい場合に異常と判定することができます。特に近年では、より高精度な変分自己符号化器(Variational Auto-encoder[8])を用いるのが一般的となっています。

一方で、異常状態の一部について事前に知り得ていることも、しばしばあります。例えば不良検査においては、よく起きる不良品が事前に特定できていることがあります。ハトのポーズの例でも、参考文献[6]では認知症患者が取る代表的なポーズを紹介しています。ところが、先

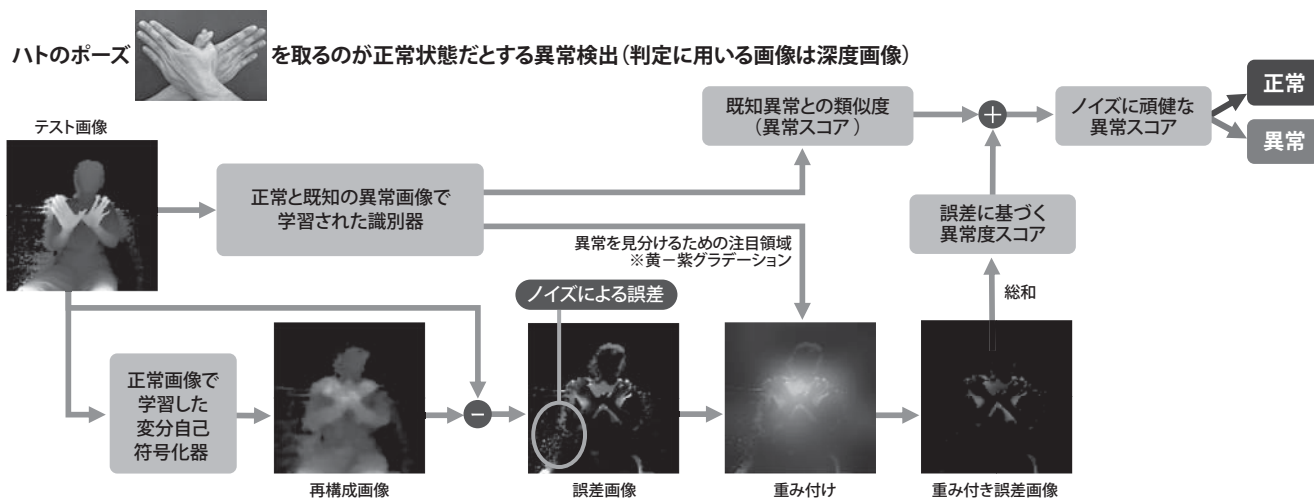


図4. 提案手法の判定の流れ

の自己符号化器を用いる手法には、それらを活用できないという問題があります。また別の問題として、再構成画像との類似度を計算する際に、テスト画像に周辺ノイズ（誤差）がある場合、そのノイズにより類似度の計算が誤ってしまうことがありました。

そこで東京基礎研究所では、これらの問題を解決する手法を提案しました[9][10]。本手法ではテスト画像が与えられた時に、まず正常画像と既知の異常画像を識別するように学習した識別器を用いて、既知の異常画像との類似性を表す異常スコアを計算し、それと同時に、異常を見分けるための注目領域を算出します。そして、変分自己符号化器の出力した誤差画像にこの領域情報で重みをかけることにより、求めた異常スコアと先の異常スコアを合算することでノイズに頑健な判定を実現します。注目領域の算出には、Grad-CAM[11]と呼ばれる学習済み識別器から画像のどの領域を用いて識別判定を行ったのかを可視化する手法を活用します。図4に、判別時の一連の流れを示しています。われわれは、この手法により世界最高レベルの精度を達成しました(表1)。例えばハトのポーズの判定では、異常検出で一般的に使用される精度尺度である受信者動作特性曲線(ROC)の曲線下面積(AUC)が、従来手法で0.91や0.95程度だったのに対し、提案手法では約0.99を達成しました(1.0が最大で、その場合は正常・異常ともに精度が100%を意味します)。このような識別精度の向上は、例えば医療画像の正確な判断を支援する基盤に、また、工場などにおいては検品作業時間の削減などにもつながると考えています。

表1. 提案手法と従来手法の精度

	ノイズを含む手書き数字画像	公開されている手のポーズ画像	山口ハトテストの深度画像
注目領域への重みなしの誤差に基づく判別[12]	0.63	0.82	0.95
既知異常で学習した識別器のみでの判別[7]	0.88	0.91	0.91
注目領域への重みありの誤差に基づく判別[10]	0.88	0.85	0.97
提案手法[9]	0.95	0.94	0.99

## ▶▶ 4. AIを活用した音響解析

近年スマート・スピーカーが、人間と家電製品や機械とのユーザー・インターフェースとして徐々に一般家庭に浸透してきています。スマート・スピーカーに搭載されているマイクアレイは、人間の声だけでなく、さまざまな生活音や環境音を認識するために使うことができます。例えば、健康に影響する日常生活動作の変化点を検出するためや、独居老人を見守るための活用も考えられます。音響データからその場の状況やイベントを検出するタスクは、機械学習においては多クラス分類問題とみなすことができます。

IEEEの音声信号および音響信号処理技術委員会が公認する「DCASEチャレンジ」という音響認識分野で最大の国際競技会があります。DCASEはDetection and Classification of Acoustic Scenes and Eventsの略称で、一般公開された共通の音響データセットを使い、音響による場面やイベントの解析方法を競い、その技術向上を支援する主旨の競技会です。競技会には課題が異なる5つのタスクがあります。

2018年、われわれは複数チャンネルの音響データから屋内の活動を検出する課題のタスク5に挑戦しました。タスク5では4つのマイクが1列に並んだマイクアレイを使用し、マイクアレイは所定の7カ所のいずれかに配置されます。ただし、どこに置かれたマイクアレイのデータなのかは競技会の参加者には知らされません。

これらのマイクアレイが設置された家で、1人の人が1週間休暇を過ごし、その生活音を録音して音響データセットを構築しています。音響データには「不在」「調理」「皿洗い」「食事」「ソーシャル」「掃除機」「テレビ鑑賞」「仕事」「その他」の9種類のラベルが付けられています。競技会では、最初に正解クラスが付いた開発用データセットと基準実装が公開され、その3カ月後に正解クラスが分からない評価用データセットが公開されました。競技会の参加者は、評価用データセットの音響クラスを推定し、推定結果を主催者に送り、主催者が推定結果を集計して分類性能を競うことになりました。

われわれはこのタスクに向けてデータ拡張(Data Augmentation)の手法を考案し、競技会に挑みました。

データ拡張とは、元々ある訓練用データを加工することによって人工的にデータ量を増やし、元の訓練用データが比較的少量であっても機械学習モデルの精度を向上させるテクニックです。例えば、画像データの場合、画像の拡大・縮小や平行移動、左右反転、回転などを行うことによってデータ拡張が行われます。データ拡張によって訓練用データの多様性を増やすことができ、機械学習モデルをより様々な入力データに対応できるようにします。

今回われわれはデータ拡張を行うにあたって2つの仮説を立てました。

- [仮説1] 音響クラスは音響イベントの順番に依存しない
- [仮説2] 同じ音響クラスの音を混ぜた音響データは元の音響クラスに属する

例えば「食事」クラスに属する音響データで「皿とフォークがなる音」の後に「皿とスプーンがなる音」があったとします。これは「皿とスプーン」の後に「皿とフォーク」の音がなっても食事風景であることには変わりません。これが仮説1になります。人間の音声データの場合、順番が変わると言葉の意味が変わってしまうのとは好対照です。また別の音響データとして「ナイフとフォークがなる音」が「食事」クラスに属していたとします。この場合、「皿とスプーン」の後に「ナイフとフォーク」の音を人工的につなげて合成した音響データも食事風景になります。これが仮説2になります。これら2つの仮説を基に、2つの音響データをそれぞれ複数の区間に分け、その順番をランダムにシャッフルして混ぜ合わせて1つの新しい音響データを合成するデータ拡張を行いました。今回の食事風景に限らず、音響イベントがランダムに発生することは世の中によくあるため、数多くの場面で2つの仮説が成り立ち、このデータ拡張手法が適用できます。また、データ拡張を行うのと同時に、深層学習に用いるニューラルネットワーク（以下、NN）にも工夫を施しました。

図5に開発用データセットにおけるクラス別分類性能評価値(F1スコア)を示します。一番左が全体の分類性能になっています。示された棒グラフはそれぞれ、「基準実装」「提案NN」「基準実装+提案データ拡張」「提案NN+提案データ拡張」の4つの分類性能を示しています。提案データ拡張、提案NNともに分類性能を改善しており、

2つを組み合わせた提案手法によって、全体で基準実装が84.50%だった性能値を89.95%まで向上させることができました。この提案手法を評価用データセットに適用し、推定結果を提出した結果、われわれのチームは第1位を獲得することができました[13]。

さらにわれわれは、製造業の現場においてもAIを活用した音響解析を適用できると考えています。現在、安川電機とのパートナーシップにより、工場の溶接工程における溶接音を使った溶接品質診断に取り組んでいます。従来、熟練工が溶接中の音を聞いて溶接の良し悪しを判断していた職人の感覚をデジタル世界に取り込んだサイバー・フィジカル・システム(CPS)を構築しています。また、ハンマーで軽く叩いて出た音の違いによって判断する打音検査のデモ(図6)を日本IBM箱崎本社7階のThinkLabにて展示しています。

## ▶▶ 5. まとめ

今日のAIブームの中で、人間を超える高性能を発揮して注目を集めるAI技術がある一方で、ビジネスの現場における実際の問題には、今のAI技術を使っても一朝一夕に解決できない問題がまだまだたくさんあります。問題を切り出してお膳立てをすれば適用できるAI技術があったとしても、ビジネスプロセスの中で日常的に利用するには、それとはまた別のアプローチが必要になる場合もあります。ビジネスの現場だからこそ生じる先進的な問題に対して、ビジネスで利用できる形で解決する新技術を発明する。私たちは、そのような研究に今後も邁進してまいります。

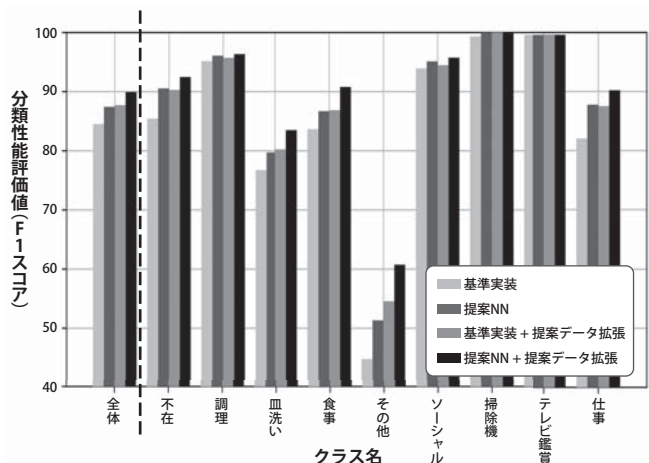


図5. 開発用データセットにおけるクラス別分類性能評価値

[参考文献]

- [1] ImageNet Large Scale Visual Recognition Challenge (ILSVRC): <http://www.image-net.org/challenges/LSVRC/>
- [2] Tung D. Le, Haruki Imai, Yasushi Negishi, Kiyokuni Kawachiya: TFLMS: Large Model Support in TensorFlow by Graph Rewriting, <https://arxiv.org/abs/1807.02037>
- [3] Haruki Imai, Samuel Matzek, Tung D. Le, Yasushi Negishi, Kiyokuni Kawachiya: Fast and Accurate 3D Medical Image Segmentation with Data-swapping Method, <https://arxiv.org/abs/1812.07816>
- [4] IBM: Watson Machine Learning Accelerator, <https://www.ibm.com/jp-ja/marketplace/deep-learning-platform>
- [5] Varun Chandola, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." ACM computing surveys (2009). [http://www.cs.umn.edu/sites/cs.umn.edu/files/tech\\_reports/07-017.pdf](http://www.cs.umn.edu/sites/cs.umn.edu/files/tech_reports/07-017.pdf)
- [6] Haruyasu Yamaguchi, Yohko Maki, and Tetsuya Yamagami. "Yamaguchi fox-pigeon imitation test: a rapid test for dementia." Dementia and geriatric cognitive disorders (2010)
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778 (2016). <https://arxiv.org/abs/1512.03385>
- [8] Diederik P. Kingma, and Max Welling. "Auto-encoding variational bayes." In Proceedings of the 2nd International Conference on Learning Representations (2014).
- [9] Daiki Kimura, Minori Narita, Asim Munawar, and Ryuki Tachibana. "Spatially-weighted anomaly detection with regression model". In Proceedings of the Meeting on Image Recognition and Understanding (2018).
- [10] Minori Narita, Daiki Kimura, and Ryuki Tachibana. "Spatially-weighted anomaly detection", In Proceedings of the Symposium on Sensing via Image Information (2018).
- [11] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-cam: Visual explanations from deep networks via gradient-based localization." In Proceedings of the IEEE International Conference on Computer Vision (2017).
- [12] Jinwon An, and Sungzoon Cho. "Variational autoencoder based anomaly detection using reconstruction probability." Special Lecture on IE (2015).
- [13] Monitoring of domestic activities based on multi-channel acoustics, <http://dcase.community/challenge2018/task-monitoring-domestic-activities-results>



日本アイ・ビー・エム株式会社  
東京基礎研究所  
AI担当シニア・マネージャー

**立花 隆輝**  
Ryuki Tachibana

1998年日本IBM入社。以来、東京基礎研究所にてマルチメディア信号処理や音声言語処理などに従事。現在は、さらに自然言語処理、画像処理、ロボット関連機械学習応用などを含めたAI関連プロジェクトのマネジメントを行う。



日本アイ・ビー・エム株式会社  
東京基礎研究所  
リサーチ・スタッフ・メンバー

**今井 晴基**  
Haruki Imai

2004年日本IBM入社。以来、東京基礎研究所にて、ハイパフォーマンス・コンピューティング分野等のアプリケーション高速化の研究開発等に従事。現在は、大規模深層学習の実行基盤の研究開発を行う。



日本アイ・ビー・エム株式会社  
東京基礎研究所  
リサーチ・サイエンティスト

**木村 大毅**  
Daiki Kimura

前職のIT企業研究所を経て、2016年日本IBM入社。修士・博士課程在学中、研究補佐員、日本学術振興会 特別研究員を兼任。優秀学術論文賞ほか、さまざまな賞を受賞。特に、画像処理や知能ロボットなどの研究に従事。



日本アイ・ビー・エム株式会社  
東京基礎研究所  
リサーチ・スタッフ・メンバー

**井上 忠宣**  
Tadanobu Inoue

1996年日本IBM入社。以来、東京基礎研究所にて、モバイル/ウェアラブル・コンピューティング、半導体リソグラフィ技術等の研究開発に従事。現在は、音響分析を中心に深層学習の研究開発を行う。

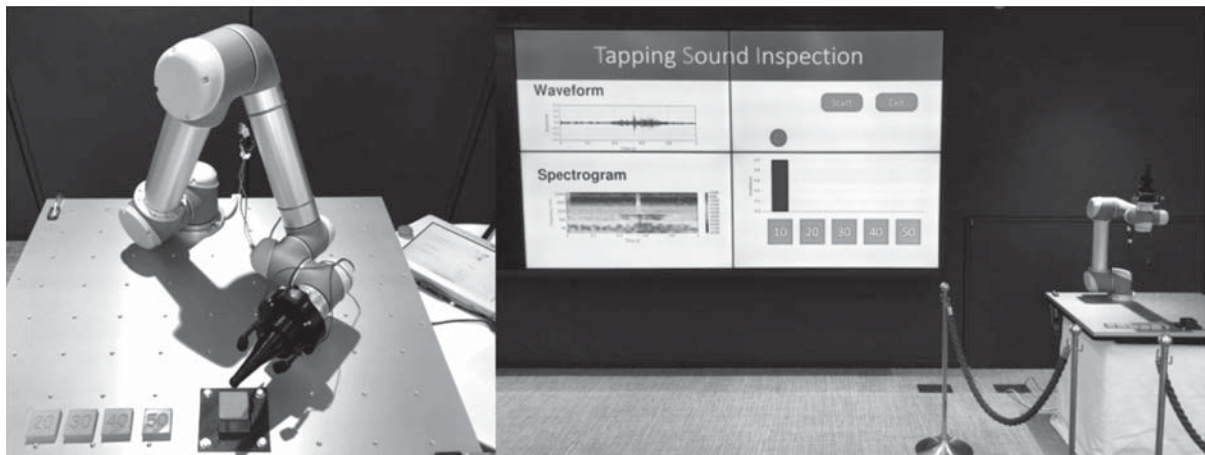


図6. ThinkLabに展示している打音検査デモ機