

jh230022

深層学習を用いた医用画像処理の信頼性向上に関する研究

小田昌宏（名古屋大学）

概要

コンピュータによる医療支援を実現する上でコアとなるのは、患者の状態や疾患に関する情報を豊富に含んだ医用画像の自動処理技術である。近年は医用画像処理において深層学習が広く活用され、自動認識精度の向上が多数報告されている。しかし 3D 画像である CT 画像などを処理する深層学習モデル構築は計算コストが高く、さらにその判断根拠を人間が解釈しにくいという問題がある。コンピュータ医療支援の普及のためには、深層学習を用いたモデル構築の高速化と、モデルの判断根拠を人間にとって解釈可能な形とする必要がある。そこで本研究では、COVID-19 診断支援と術中画像のシーン認識の 2 つのアプリケーションを対象とし、スーパーコンピュータを用いた画像処理モデル構築の高速化と、モデルの判断根拠の可視化処理高速化を行った。TensorFlow を用いた実装で評価したところ、CT 画像処理モデル構築を最大 2.43 倍高速化することができ、判断根拠可視化に関しては、モデルの判断に貢献した画素を可視化する重要度マップの作成処理を平均 1.64 倍高速化できた。

1. 共同研究に関する情報

ン適用

(1) 共同利用・共同研究を実施している拠点名

名古屋大学 情報基盤センター

(2) 課題分野

大規模計算科学課題分野

(3) 共同研究分野 (HPCI 資源利用課題のみ)

超大規模数値計算系応用分野

(4) 参加研究者の役割分担

小田昌宏 (名大・代表) : 医用画像処理コード開発

片桐孝洋 (名大・副代表) : プログラム最適化、スパコン適用

森健策 (名大) : 医用画像処理コード開発

羽生達郎 (名大) : プログラム最適化、不確実性解析の実装

湯浅義尚 (名大) : プログラム最適化、スパコ

2. 研究の目的と意義

深層学習を用いることで、医師に匹敵する判断能力を持つ医療支援システムが実現している一方、深層学習を用いた処理の判断根拠が人間にとって解釈しにくい（ブラックボックス的である）ことが問題となり、医師や患者が支援システムを信頼せず利用を避ける場合がある。コンピュータによる医療支援を普及させ、健康寿命の延伸と医療費削減を実現するためには、深層学習を用いた医用画像処理の判断根拠を人間にとって解釈可能な形とする必要がある。

深層学習モデルの判断根拠を人間が解釈可能な形で表現するため、不確実性解析や判断寄与領域可視化が行われている。これらは説明可能 AI 研究に含まれる手法である。これらを用いることで、画像上で深層学習モデルの判断が曖昧となる領域や、判断に強く貢献した領域を可視化することが

可能である。深層学習モデルの判断結果と共に可視化結果を提供することで、医師や患者がシステムを利用する際の理解や信頼性を向上させることができると考えられる。しかし、判断根拠可視化を行うためには、深層学習モデルによる推定処理を多数回実行する必要があり時間を要する。また、医用画像処理で使用される 3 次元 CT 画像や動画は深層学習の計算過程で多くのメモリを使用する。深層学習の演算では GPU を活用する機会が多いが、GPU の数や搭載メモリ量の制限から、これまで判断根拠可視化の研究は十分に行われてこなかった。

本研究では、深層学習を用いた医用画像処理において、スーパーコンピュータを活用した画像処理モデル構築とモデルの判断根拠可視化を行った。具体的には以下 2 つアプリケーションに取り組んだ。

2.1. COVID-19 診断支援

COVID-19 の診断には PCR 検査などが用いられるが、誤った検査結果を出すことがあること、肺で発生する炎症の状況を正確に評価できないことなど問題がある。胸部 CT 画像を用いることで、肺の状態を観察可能となり詳細な状態評価ができる。多数の患者が発生する状況において全ての CT 画像を医師が確認するのは負担となることから、コンピュータによる診断支援の実現が期待される。そこで本アプリケーションでは、胸部 CT 画像からの疾患領域の自動セグメンテーション（画像上での領域切り出し）と疾患自動分類を対象として研究を実施した。

2.2. 術中画像のシーン認識

外科手術においては、外科医を支援する手術ナビゲーションシステム実現や手術支援ロボットの自動化を目指して、術中に撮影される画像や動画からのシーン自動認識を画像処理により実現することが求められる。術中画像のシーン認識では手術器具や臓器の認識が主な目的となる。手術においては一度の判断誤りが即座に医療事故につな

がる可能性があるため、認識精度を高めることと同時に、医師が解釈可能な形で認識根拠を提示し、医師の判断を仰ぐことが重要である。そこで本アプリケーションでは、術中画像からのシーン認識を行う深層学習モデルの判断根拠可視化に関する研究を実施した。

3. 当拠点の公募型共同研究として実施した意義

医用画像処理では、術中画像、内視鏡画像、超音波画像、CT 像、X 線画像など様々な画像を扱う。CT 画像などの 3D 画像や大規模な画像を扱う場合、データ量が非常に大きいため、研究グループ内の計算資源を用いると、深層学習モデル構築や判断根拠可視化に非常に長い時間を要するなど問題が生じていた。

上記の状況に対し、拠点公募型共同研究として研究を実施することで、十分な計算資源が利用可能となった。これにより、高精細な医用画像を用いた深層学習モデル構築、タスクに応じたモデル構築の高速化、判断根拠可視化の高速化が新たに可能となった。

医用画像処理分野ではスーパーコンピュータを活用した例がまだ少なく、本研究の成果が当該分野におけるスーパーコンピュータの利用拡大に結び付くと考えられる。

4. 前年度までに得られた研究成果の概要

該当なし

5. 今年度の研究成果の詳細

本研究の 2 つのアプリケーションである COVID-19 診断支援と術中画像のシーン認識それぞれについて研究成果を述べる。

5.1. COVID-19 診断支援

胸部 CT 画像を用いた COVID-19 診断支援では、画像からの(1) 疾患領域自動セグメンテーションと、(2) 疾患自動分類が主に用いられる。セグメンテーションによって肺内部の異常陰影領域を切り出し、CT 画像とセグメンテーション結果を併用し

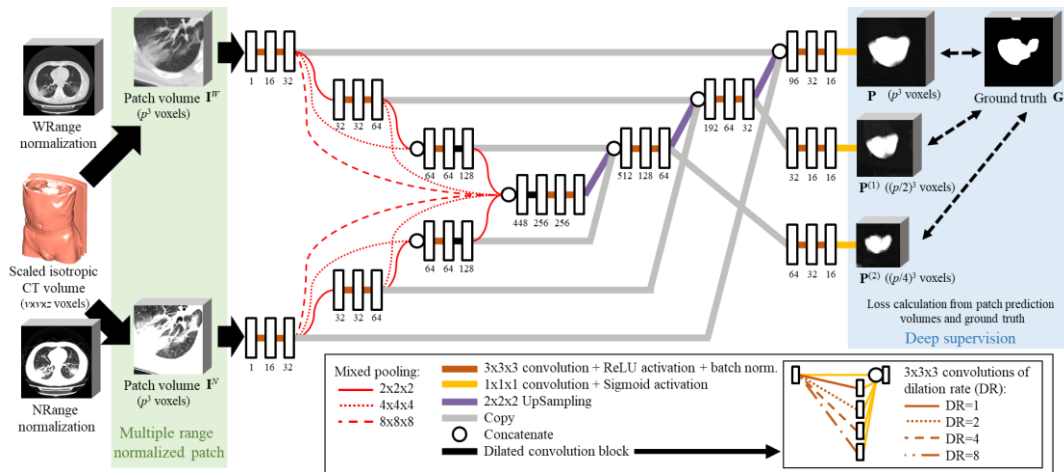


図 1 使用した COVID-19 症例の異常陰影領域セグメンテーション用深層学習モデル (M. Oda, et al., MICCAI CLIP Workshop, 2021) の構造

て疾患自動分類を行う。今年度は、上記 2 つのタスクそれぞれに対して研究開発を行った。

(1) 疾患領域自動セグメンテーション

セグメンテーション処理では、CT 画像から肺内部の異常陰影領域を推定する。推定においては COVID-19 症例の異常陰影領域セグメンテーション用に構築された深層学習モデル (M. Oda, et al., MICCAI CLIP Workshop, 2021) を使用した。モデルの構造を図 1 に示す。このモデルの入力、出力は共に 3D 画像であり、モデル内部に 3D カーネルを用いた畳み込み層などを多く持つ。そのため GPU を用いた学習処理において、処理に長い時間を要し、GPU メモリを多く使用する。

そこで、以下の観点での GPU 利用の並列化のプロトタイピングを行った。

● パラメタサーベイ用の基盤環境のプロトタイピング

以上の深層学習で高い学習性能を得るためには、多くのハイパーパラメタチューニングが必要であることが知られている。この処理は、パラメタサーベイとよばれ、簡単に並列化できることも知られている。

しかしながら、スーパーコンピュータシステムへの多数のジョブの投入が必須である。この作業は煩雑である。そこで、スーパーコンピュータのジョブスケジューラと連携し、パラメタサーベイ

を簡便に行うプログラム環境である Xcrypt を使い効率的に行う基盤環境の整備を行った (研究業績 [3])。

● Mirrored Strategy による複数 GPU 実行による深層学習の高速化

本アプリケーションの深層学習を用いた処理は、TensorFlow により実装されている。そこで TensorFlow の分散学習 API を利用し、簡便にノード内の複数 GPU を活用して高速化する実装を行った。

具体的には、tf.distribute.Strategy を活用した。これは、TensorFlow の処理を、複数マシン、複数 GPU や TPU などに分散させる API に含まれている、Mirrored Strategy を利用した。以下の実装部分を対象にした。

```
model =
network_multiin_multiscaleout_unet3d_dilated_den
sepooling3()
model.compile()
```

以上の対象を、Mirrored Strategy を用いて、以下のように書き換えた。

```
import tensorflow as tf
mirrored_strategy = tf.distribute.MirroredStrategy()
#mirrored_strategy =
```

表 1 「不老」 TypeII における複数 GPU 利用による深層学習の高速化

	1 GPU	2 GPU	3 GPU	4 GPU
実行時間 [秒]	13,440	8,339	6,845	5,528
速度向上 [倍]	1	1.61	1.96	2.43

`tf.distribute.MirroredStrategy(devices = ["/gpu:0"])`

with mirrored_strategy.scope():

model =

network_multiin_multiscaleout_unet3d_dilated_d
ensepooling3()

model.compile()

この実装により、名古屋大学のスーパーコンピュータ「不老」 TypeII サブシステムにある 1 ノードの複数 GPU (NVIDIA V100×4 基) において、表 1 のように最大で 2.43 倍の高速化を達成した。

(2) 疾患自動分類

疾患自動分類処理では、CT 画像とセグメンテーション結果を用い、COVID-19 感染の可能性が高いか低いかの判別を行う。ここでは 3D Convolutional Neural Network (CNN) を用いた画像分類を使用する。深層学習モデルを用いた処理の中でも、画像分類は画像からダイレクトにクラス分類結果を推定しており、モデルの判断根拠は解釈困難である。COVID-19 感染の可能性の高低は医師や患者にとって重要な情報であり、モデルの判断根拠を明確化できることが望ましい。そこで、疾患自動分類処理に用いる 3D CNN の判断根拠可視化を行った。

判断根拠可視化のため、説明可能 AI の一手法であり、深層学習を用いた画像分類モデルの判断に強く貢献した画素を画像上に提示可能な Integrated Gradients (M. Sundararajan, et al., ICML'17, 2017) を使用した。Integrated Gradients は、学習済み画像分類モデルに与える入力画像に対して少しずつ変化を加え、それによるモデルの分類結果の

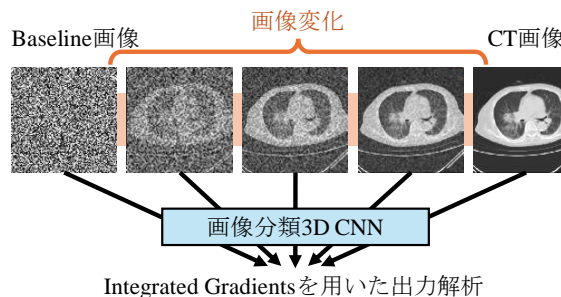


図 2 Integrated Gradients 適用時の画像変化のイメージ

変化から、画像上の各画素の分類への貢献度を評価し、重要度マップを作成する。本手法は判断根拠可視化に有用であるものの、入力画像に対する変化付与と画像分類モデルによる推定処理を数百回実行する必要があるため、処理に時間を要する。医療支援での利用を考えると、AI による診断支援情報推定はすぐに得られるものの、判断根拠に関する情報が得られるまで長い時間を要し実用的でない。そこで、複数 GPU での実行による判断根拠可視化処理の高速化を行った。

高速化のために、セグメンテーションと同様に TensorFlow の `tf.distribute.Strategy` を使い、複数 GPU に処理を分散させる `Mirrored Strategy` を使用した。これにより、重要度マップ作成に必要な、画像分類モデルを用いた多数の推定処理実行を並列化した。

Integrated Gradients による重要度マップ作成処理においては、Baseline 画像として一様分布に従うノイズから構成された 3D 画像を使用し、Baseline 画像と CT 画像をアルファブレンディングした画像をモデルに与えた。モデルへの入力画像の変化として、アルファブレンディングの割合を 200 段階変化させ (図 2)、それによるモデルの分類結果の変化から重要度マップ作成を行った。

上記の高速化のための実装を行い、1 ノード内の複数 GPU を使用することで、表 2 のように重要度マップ作成処理を 1.64 倍高速化することができた。この処理時間は CT 画像 11 例に適用した結果の平均値である。

表 2 CT 画像分類モデルの判断に関する重要度マップ作成処理における、複数 GPU 利用時の実行時間。CT 画像 11 例に適用した結果の 1 例あたりの平均値である。

	1 GPU	4 GPU
実行時間 [秒]	84.7	51.5
速度向上 [倍]	1	1.64

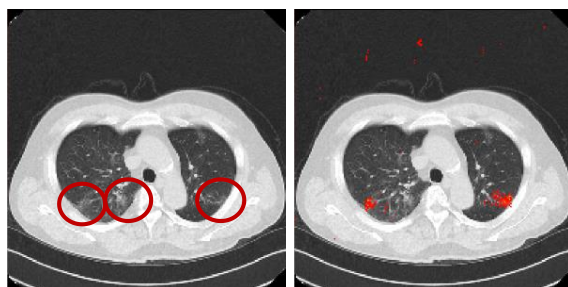
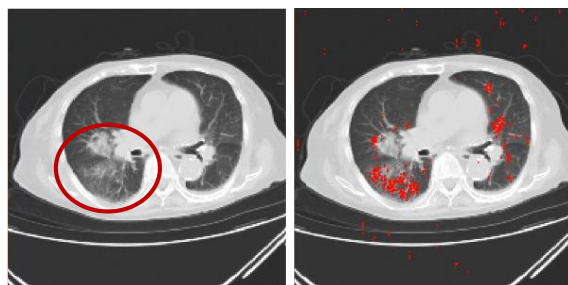
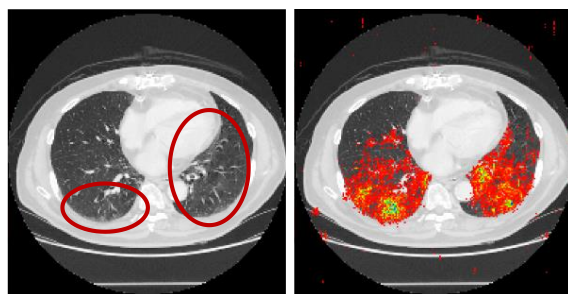
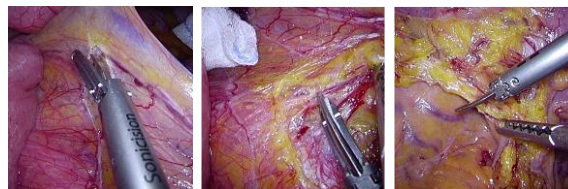


図 3 (左側) COVID-19 症例の CT 画像、(右側) 疾患自動分類 3D CNN から生成した重要度マップ。左側の図では異常陰影領域を丸で示す。右側の図では、色のついた領域が重要度マップを示す。赤から緑色がマップでの低値から高値に対応する。

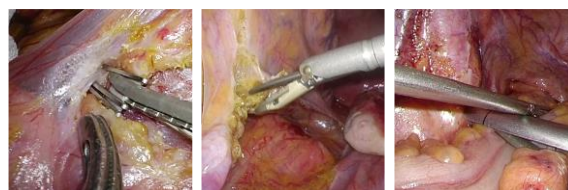
本処理により COVID-19 症例の CT 画像から生成された重要度マップを図 3 に示す。図の左側は、3D の CT 画像に含まれる断面画像 (Axial 断面)、右側が断面画像に重要度マップを重畳表示したものである。左側の図では、COVID-19 に起因すると



(a)



(b)



(c)

図 4 画像分類タスクで使用した腹腔鏡画像の例。(a) 血管シーリング装置を含む画像、(b) 超音波凝固切開装置を含む画像、(c) どちらも含まない画像。

と思われる肺内部の異常陰影領域を丸で示す。右側の図において色の付いている領域は、疾患自動分類 3D CNN の判断への貢献度が高い領域である。重要度マップにおいては、異常陰影が存在する領域に反応が現れており、COVID-19 の重要な特徴である肺内部の異常陰影を基にして 3D CNN が判断を行っていることが可視化された。

5.2. 術中画像のシーン認識

術中画像のシーン認識には、術具の認識、臓器のセグメンテーション、手術フェーズ分類など様々なタスクが含まれる。本研究では、画像に含まれる術具の種類に応じた画像分類タスクを対象とする。画像を基にして術具の種類を自動判別できれば、手術で行われている操作をコンピュータで自動認識して適切な支援情報を外科医に提供することが可能となる。このとき、分類結果だけでなく画像分類モデルの判断に貢献した領域を可視化することができれば、外科医にとって有用である。可視化には COVID-19 診断支援研究と同様に

表 3 腹腔鏡画像分類モデルの判断に関する重要度マップ作成処理における、複数 GPU 利用時の実行時間。画像 1296 枚に適用した結果の 1 枚あたりの平均値である。

	1 GPU	4 GPU
実行時間 [秒]	0.31	0.43

Integrated Gradients などが利用できるが、重要度マップ作成処理には時間を要する。外科手術においては、刻一刻と変化する手術状況に対して遅延なく支援情報を提示する必要がある。そのため重要度マップ作成をできるだけ高速に実行することが求められる。そこで、複数 GPU 利用による重要度マップ作成処理の高速化が可能か検証した。

本研究で扱うタスクでは、腹腔鏡を用いた外科手術中に撮影された腹腔鏡画像（第 3 回 JSCAS AI Challenge のデータセット利用）を使用し、その中に写っている術具の種類に応じて画像分類を行う。術具には血管シーリング装置、超音波凝固切開装置が含まれ、画像分類では、(a) 血管シーリング装置を含む画像、(b) 超音波凝固切開装置を含む画像、(c) どちらも含まない画像、の 3 クラスに分類を行う。3 クラスの画像の例を図 4 に示す。画像分類には CNN を使用する。COVID-19 診断支援の場合と同様に Integrated Gradients を使い、画像分類モデルの判断に強く貢献した画素を重要度マップとして可視化する。

重要度マップ作成処理に TensorFlow の tf.distribute.Strategy を導入し、複数 GPU に処理を分散させる Mirrored Strategy を使用した。これにより、重要度マップ作成に必要となる、画像分類モデルを用いた多数の推定処理実行を並列化した。1 ノード内で GPU 1 台を使用する場合、GPU 4 台を使用する場合に、画像 1 枚あたりの重要度マップ作成処理時間を計測したところ表 3 のようになった。この処理時間は画像 1296 枚に適用した結果の平均値である。

結果から、GPU 4 台を使用する場合に実行時間が長くなった。1 枚あたりの処理時間が短い状況では並列処理のためのオーバーヘッドの影響が大

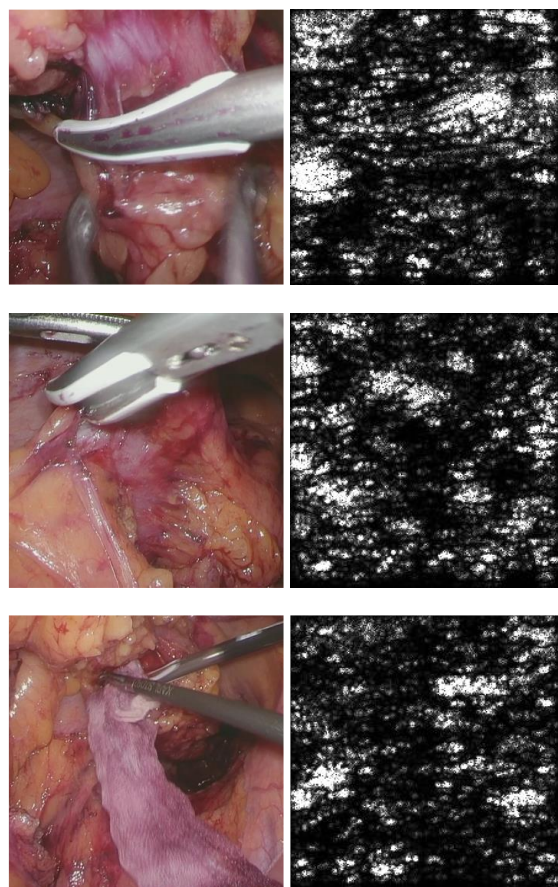


図 5 術中画像分類モデルから作成した重要度マップ。(左側) 入力した術中画像、(右側) 重要度マップ。重要度マップでは画素値が明るいほど高い重要度を表す。

きくなったと考えられる。本タスクでの重要度マップ生成の高速化にはアルゴリズム改善などで対応することが考えられる。

作成した重要度マップを図 5 に示す。術具付近に高い値を示す白色領域が多く存在する。その他にも画像中の様々な箇所がモデルの判断に用いられていることが分かる。

6. 進捗状況の自己評価と今後の展望

本研究で実施を予定していた 2 つのアプリケーションそれぞれにおいて、モデル構築の高速化や判断根拠可視化処理の高速化といった、計算資源を活用した重要な研究成果が得られた。ただ、術中画像のシーン認識では期待していた高速化は達成できなかった。また、本研究で得られた成果を国内および国際会議において積極的に発表した。

研究全体としては有用な成果が多数得られたことから、本研究の達成度は 90%と自己評価する。

COVID-19 診断支援においては、疾患領域自動セグメンテーション、疾患自動分類の 2 つのタスクに取り組んだ。疾患領域自動セグメンテーションでは、パラメタサーベイ用の基盤環境のプロトタイピングを行い、その成果を国内会議で発表した。また、複数 GPU を用いた深層学習モデル構築の高速化を行い、学習処理を最大 2.43 倍高速化することができた。疾患自動分類では、画像分類モデルの判断に貢献した画素を可視化する重要度マップ作成処理の高速化を行った。複数 GPU を用いた処理の導入により平均 1.64 倍の高速化が可能であった。これらの成果は医療現場における AI 活用の一助となるものであり、今後様々な対象へ応用されると考える。今回は単一ノード利用での高速化に留まったため、今後の課題として、複数ノードの GPU を活用した分散処理のための実装とアルゴリズム開発が挙げられる。

術中画像のシーン認識においては、術中画像分類モデルからの重要度マップ作成処理に対して、複数 GPU 利用を試した。しかし並列処理のためのオーバーヘッドの影響により、単一 GPU 利用時の方が処理が高速であるという結果になった。今後は、高速化を行う適切な対象の選定、重要度マップ作成アルゴリズムの改良などが必要である。

本研究の申請時には、深層学習モデルの不確実性解析を実施する予定であったが、モデル判断の曖昧さを示す不確実性よりも、モデル判断に貢献した領域を提示する方が判断過程の理解に有用であるため、重要度マップ生成を実施した。不確実性解析で得られる不確実性マップと今回得られた重要度マップの計算過程には類似する部分があり、重要度マップ作成処理の高速化の成果は不確実性マップ作成処理にも応用可能である。

7. 研究業績

(1) 学術論文 (査読あり)

(2) 国際会議プロシーディングス (査読あり)

[1] Masahiro Oda, Zhou Zheng, Jie Qiu, Yuichiro Hayashi, Kensaku Mori, Hirotsugu Takabatake, Masaki Mori, Hiroshi Natori, “What is Trustworthy AI and How It Contributes in Diagnosis,” RSNA2023, PHEE-12, 2023/11/26 (ポスター発表、アブストラクト査読)

(3) 国際会議発表 (査読なし)

(4) 国内会議発表 (査読なし)

[2] 小田昌宏, “医療支援における画像処理研究の動向と展望,” 第 29 回画像センシングシンポジウム, オーガナイズドセッション 2「安全・安心のための実用 CV 技術」, 2023 年 6 月 15 日 (口頭発表)

[3] T. Hanyu, M. Kawai, T. Katagiri, T. Hiraishi, T. Hoshino, T. Nagai, “Auto-tuning of Hyperparameters by Parallel Search Using Xcrypt,” HPC Asia 2024, 2024/01/26 (ポスター発表, アブストラクト査読)

[4] 楊 天予, 小田昌宏, 鄭 通, 林雄一郎, 森 健策, “不確実性を用いた胸部 CT 画像からの COVID-19 病変セグメンテーション手法,” 電子情報通信学会技術研究報告(MI) (123), no. 411, MI2023-74, pp. 137-140, 2024/03/04 (口頭発表)

(5) 公開したライブラリなど

(6) その他 (特許, プレスリリース, 著書等)