

jh190047-DAH

Deep Learning を用いた医用画像診断支援に関する研究

佐藤 一誠 (東京大学)

概要 本共同研究では、Deep Learning (DL) を用いた高性能なコンピュータ支援検出 (CAD) の開発を行うことを目的とする。昨年度までに構築した DL 学習環境を用いて複数の CAD (胸部 MR 画像の肺結節検出, 胸部 X 線写真の病変検出など) の開発を進めるとともに、開発用データ収集における医師のアノテーションの負担軽減を目的とした DL を用いた汎用的な病変形状ラベル推定方法を構築した。さらに、GPU メモリの制限が DL 学習のボトルネックとなる問題に対し、メニーコア型プロセッサを用いた大規模 DL 学習の検証を実施した。

1. 共同研究に関する情報

(1) 共同研究を実施した拠点名

東京大学

(2) 共同研究分野

- 超大規模数値計算系応用分野
- 超大規模データ処理系応用分野
- 超大容量ネットワーク技術分野
- 超大規模情報システム関連研究分野

(3) 参加研究者の役割分担

| | |
|-------|------------------------------------|
| 佐藤 一誠 | 病変形状ラベル推定方法の構築 CAD 開発 |
| 埴 敏博 | GPU 間通信, ノード間通信, KNL 最適化 |
| 林 直人 | 臨床面での評価 |
| 花岡 昇平 | 病変形状ラベル推定方法の構築, CAD 開発, 臨床面での評価 |
| 野村 行弘 | 病変形状ラベル推定方法の構築, CAD 開発. |
| 竹永 智美 | 病変形状ラベル推定方法の構築, CAD 開発 |
| 柴田 寿一 | CAD 開発 |
| 佐藤 大介 | CAD 開発, 臨床面での評価 |
| 中尾 貴祐 | CAD 開発, 臨床面での評価 |
| 中村 優太 | CAD 開発, 臨床面での評価 |
| 星野 哲也 | GPU 向け最適化 |
| 関谷 勇司 | ネットワーク運用技術 |
| 橋詰 明宗 | GPU 間通信, ノード間通信, KNL 最適化 |

2. 研究の目的と意義

CT や MRI をはじめとする医用画像を解析し、自動検出された病変の位置を提示する、コンピュータ支援検出 (computer-assisted detection, CAD) は画像診断における医師の見落とし低減を目的として古くから研究されている。しかし、日常的に臨床で使われている CAD は乳房 X 線写真に対する腫瘍検出、胸部 X 線写真もしくは胸部 CT における肺結節検出などに限られているのが現状である。その要因の 1 つとして、多くの CAD が病変検出性能を経験のある医師と同程度とした場合、医師に比べて拾いすぎ (偽陽性, false positive; FP) が多いことが挙げられる。このため、CAD のさらなる高性能化が必須である。

一方、CAD の主要技術である機械学習分野において、Deep Learning (DL) がこの数年注目を集めており、医用画像処理にも利用されている。DL の一手法である畳み込みニューラルネットワーク (convolutional neural network, CNN) は脳の視覚野における情報処理を模したものであり、画像診断は視覚処理である。このため、医師が画像診断に用いる有用な情報を適切に CNN に学習させることができれば、CAD のさらなる高性能化の実現が期待できる。

本共同研究全体の目的は、DL を用いた高性能な CAD の開発、および開発した CAD の臨床現場での評価である。今年度は前年度までに構築した学習フレームワークを用いた CAD の開発を継続

するとともに、以下の 2 項目について新たに検討を行う。

- (a) CAD 開発用症例データ収集の際に問題となる、病変形状ラベル入力における医師の負担軽減を目的とした、DL を用いた汎用的な病変形状ラベル推定方法を構築する。
- (b) GPU のメモリの制約により学習が困難である、入力画像サイズが大きい場合やネットワーク規模が大きな DL の学習について、東大 Oakforest-PACS システム (OFP) 搭載のメニーコア型プロセッサでの実現可能性を検証する。本共同研究で DL を用いた高性能な多種の CAD が開発されることにより、臨床医学における画像診断の質的向上ならびに医師の負担軽減が期待される。CAD の開発には医師が診断した医用画像データを用いることから、データを作成した国の診断基準が CAD に反映される。そのため、国内で使用される CAD は国内で開発することが非常に重要であると考え、CAD の研究・開発ならびに臨床使用に関して多くの経験を持つ 5 名の放射線科医師 (林, 花岡, 中尾, 佐藤大介, 中村), および 2 名の診療放射線技師 (野村, 竹永) が共同研究者として参画している。共同研究者間で密に連携を取ることで、CAD の開発から臨床での評価までを促進させる体制を整えている。さらに、多くの臨床施設で高性能な CAD が使われるようになれば、本共同研究に賛同する臨床施設の参画による拠点連携研究へ進展することが期待される。

本共同研究で構築する大規模データに対応した DL 学習自動化のフレームワークは本共同研究のみでなく、DL を用いた様々な研究の促進に寄与することが期待される。

3. 当拠点公募型研究として実施した意義

DL は医用画像処理を含む様々な領域で高い性能を発揮しているものの、パラメータ調整が難しいことが問題である。本共同研究で DL を用いた CAD 開発に向けて DL 学習の高速化およびパラメータ自動探索を用いた DL 学習方法の確立を目指している。前年度の成果により数百症例を用いた

DL 学習方法は確立しつつあるものの、より大規模 (千~万症例) の学習を短期間で行うためにはさらなる効率化が必要と考える。このフレームワークを構築するためには、CAD 開発に係わる各分野の専門家と計算機センターの緊密な協力が不可欠であり、当拠点公募型共同研究として取り組む必要がある。さらに、本研究成果を広く公開することによって、各基盤センターの GPU 搭載のスーパーコンピュータにおいても同様のことが容易に実現できるようになると考える。

本共同研究は東大単独で実施しているが、過去の研究[A]より CAD 開発用の症例データを多施設連携で収集することの重要性を認識しており、将来的には複数の臨床施設の参画による拠点連携研究へ発展させることを視野に入れている。

[A] 野村, MEDICAL IMAGING TECHNOLOGY, vol.32, no.2, pp.98-108, 2014.

4. 前年度までに得られた研究成果の概要

前年度までに非同期並列型ベイズ的最適化 (Bayesian optimization, BO) [C]に基づくパラメータ自動探索を用いた DL 学習フレームワークの構築した。その成果は今年度国際会議にて発表するとともに論文誌に先行公開された (研究業績 [1],[2])。図 1 に構築した DL の学習フレームワークでの処理の流れを示す。GPU クラスタ (RB-H) のログインノード上でジョブ並列スクリプト言語 (Xcrypt) [B]で記述したジョブ投入モジュールを実行し、非同期並列型 BO により生成されたパラメータを用いた学習ジョブを RB-H の計算ノードに繰り返し投入する。

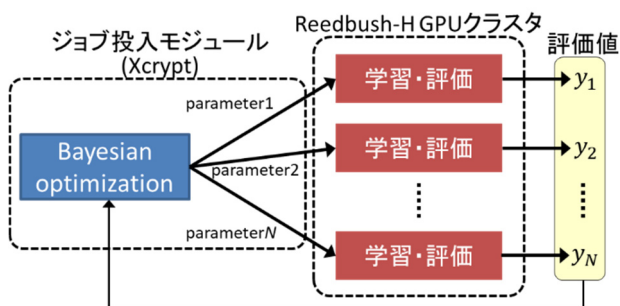


図 1 実装した DL の学習方法の流れ

構築した DL 学習フレームワークを用いて 6 種類の CAD を開発した。これらのうち、(a)および (b)は学習フレームワークの論文 (研究業績[1]) の実証例であり、(c)については学習フレームワークの論文 (による(c)は論文誌へ投稿中である。また、(d)~(f)については性能改善を継続している。

- (a) 胸部 CT 画像の肺結節検出
- (b) 胸部 X 線写真の肺腫瘍検出
- (c) 頭部造影 MR 画像の転移性脳腫瘍検出
- (d) 4D-CNN を用いた EOB-MR 画像における肝結節病変検出
- (e) 頭部救急 CT 画像の異常検知
- (f) FDG-PET/CT 画像における異常検知

[B] Hiraishi T, WHIST 2012, 2012.
 [C] Snoek J, Advances in Neural Information Processing Systems 2012;2951-2959.

5. 今年度の研究成果の詳細

(1) DL を用いた汎用的な病変形状ラベル推定方法の構築

単一モデルによる多様な病変 (濃淡・形状) の形状推定の初期検討として、周囲とのコントラストが異なる 2 種類の病変 (胸部 CT 画像の肺結節、肝臓造影 MR 画像の肝結節性病変 (図 2)) の領域抽出を 1 種類の病変のみで学習したモデルで実現可能かを検討した。DL モデルには 3D FC-ResNet [D]を使用し、元画像および病変の位置・サイズ情報より生成した球状マスク画像を入力とした。対

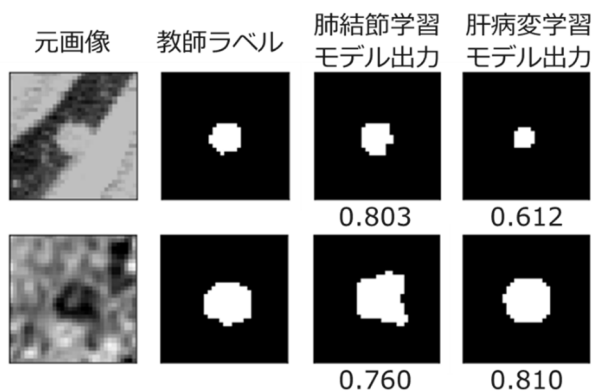


図 2: 汎用的な病変形状ラベル推定方法の結果例 (上段: 肺結節 (胸部 CT 画像), 下段: 肝結節性病変 (肝臓造影 MR 画像), モデル出力下部の数値は Dice 係数(最大:1))

表 1: 汎用的な病変形状ラベル推定結果 (数値は Dice 係数, 平均±標準偏差)

| 抽出対象 | 学習モデル | |
|--------|-------------|-------------|
| | 肺結節 | 肝結節性病変 |
| 肺結節 | 0.685±0.121 | 0.471±0.152 |
| 肝結節性病変 | 0.681±0.131 | 0.757±0.134 |

象とした 2 種類の病変は周囲の画素値との関係が異なるため、学習では白黒反転による data augmentation を併用した。図 2 および表 1 より良好な結果が得られており、1 種類の病変のみで学習した DL モデルによる他種類の病変領域抽出の実現可能性を示せた。今後、学習および抽出対象の病変を増やした検討を行うことで、汎用的な領域推定方法の確立を目指す。

(2) メニーコア型プロセッサを用いた大規模 DL 学習の検証

ONP に PyTorch および horovod をインストールし、複数ノードを用いた DL 学習環境を構築した。当初導入した PyTorch 1.4.0 では MKL-DNN に起因した性能不具合などにより調整が難航したものの、年度末に導入した PyTorch 1.5.0 により処理速度が大幅に改善された。

構築した環境において U-Net を用いた胸部 X 線写真の肺野領域抽出の学習を試みた (初期検証のため入力画像サイズを 256×256 にて実施)。Reedbush-H/L (1 ノード) では 473 秒 (1GPU), 300 秒 (2 GPUs) で完了した学習が ONP では最短で 321 秒であった (ノード数: 32, MPI プロセス: 32, スレッド数: 32)。また、ノード数を 16,8 と減らした場合の最短学習時間はそれぞれ、405 秒, 658 秒であった (ONP の処理時間は PyTorch 1.5.0 正式公開後に再測定した結果である)。

(3) 複数ノード・GPU を用いた大規模 DL 学習の検討 (研究業績[6],[8])

3D U-Net を用いた胸部 CT 画像の肺結節自動検出を対象に複数ノードによる大規模 DL 学習について検討した。Chainer で実装したモデルを使用し、データ並列による 2~64 GPUs での性能検

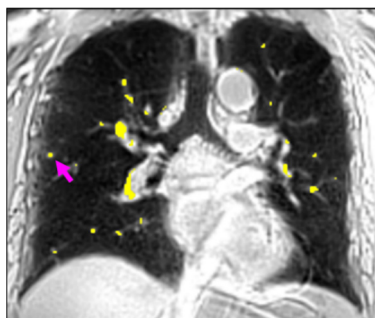


図 3: 胸部 MRI の肺結節病変候補検出結果
(黄色：病変候補領域，矢印：肺結節)

証を行った。1 GPU での学習における最適ハイパーパラメータを複数 GPU の学習に適用した場合、8 GPU 以下であれば性能が低下することなく、GPU 数の増加による学習時間の短縮が図ることができた。16~64 GPU の場合は 1 GPU での学習と同等の性能を得るためには各 GPU 数において BO (Optuna を使用) によるハイパーパラメータ探索が必要であり、64 GPU では 1 GPU に比べて学習時間を 1/49.6 倍に短縮できた。

(4) DL を用いた CAD の開発

DL を用いた CAD の開発を継続している。(4-a) は今年度新たに実施したものであり、(4-b) は昨年度より継続して実施しているものである。

(4-a) 胸部 MRI の肺結節検出 (研究業績[3],[8])

近年実用化された高速撮像法 (Ultra short TE) を用いて撮像された胸部 MR 画像を用いた肺結節検出の初期開発を行った。手法は 3 次元形状解析に基づく病変候補抽出 (図 3) および 3D CNN を用いた病変候補分類の 2 段階処理で構成されている。肺結節を有する 90 症例の 5-fold cross validation による初期評価を行った結果、病変候補検出処理は症例あたりの平均偽陽性数が 197.7 個のときの検出感度が 86.3% であった。3D CNN による病変候補分類を含む全体の性能は症例あたりの平均偽陽性数が 4.9 個のときの感度が 60.7% であった。さらに、3D FC-ResNet による肺野領域抽出の改良についても検討した。

(4-b) 胸部 X 線写真の異常検知・異常強調 (研究業績[4],[5],[7],[10])

胸部 X 線写真で診断可能な疾患や異常は多岐に

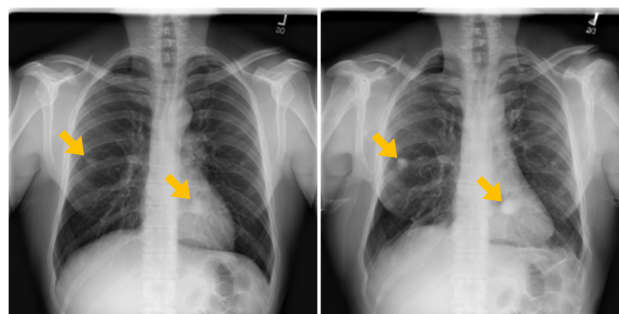


図 4: Glow を用いた胸部 X 線写真の異常強調
(左：元画像，右：強調結果，矢印：肺腫瘍)

わたることを鑑み、今年度は主に正常症例のみを学習した異常検知に基づく病変検出について検討した。昨年度より継続している α -GAN を用いた異常検知手法について各種病変の検出を試みた結果、ROC 曲線の下面積 (ROC-AUC) が 0.751 であった。この成果は論文誌に投稿中である。また、可逆変換が可能な Flow ベースの生成モデルである Glow [E] を用いた異常検知についても検討を行い、初期検討の結果では ROC-AUC が 0.783 であった。さらに、Glow を用いた病変の異常強調手法を構築した (図 4)。今後、有用性の評価のために複数の医師による観察実験を実施する予定である。

[D] Takenaga T, IJCARS 2019;14(8):1259-1266

[E] Kingma DP, NeurIPS 2018:10236-10245

6. 今年度の進捗状況と今後の展望

本共同研究全体では概ね順調に進捗した。今年度新たに実施した 3 項目 (DL を用いた汎用的な病変形状ラベル推定方法、複数ノード・GPU を用いた大規模 DL 学習の検討、メニーコア型プロセッサを用いた大規模 DL 学習の検証) については一定の成果が得られた。メニーコア型プロセッサを用いた大規模 DL 学習については、来年度は Oakbridge-CX での検討と併せてさらに進めていく予定である。

各 CAD 開発については、教師あり学習による病変検出のほかに、異常検知に基づく手法について検討した。その中でも生成モデルの Glow を用いた異常検知は良好な結果が得られていることから、対象画像・部位を変えた場合の検討や、他の

生成モデルを用いた場合の検討を進めていく予定である。

成果発表については、前年度まで構築した非同期並列型 BO による学習フレームワークが論文誌に先行公開された。各 CAD の開発については主に新たに実施したものについて発表を行った。現在、改良を進めている肝結節病変検出などについては、前年度までの発表内容より十分な性能改善が図られた段階で成果発表を行う予定である。また、今年度実施内容のうち未発表分については、今後の成果発表に向けた準備を進めている。

今年度までは多数症例による学習や大規模ネットワークの学習に着目してきたが、CAD の臨床使用を想定した場合、高性能かつ小規模な DL モデルが望ましい。このため、来年度は neural architecture search (NAS) に基づく DL 構造探索方法を RB-H/L 上に構築するなどの新たな検討を進めていく。

7. 研究業績一覧（発表予定も含む）

(1) 学術論文（査読あり）

[1] Nomura Y, Sato I, Hanawa T, Hanaoka S, Nakao T, Takenaga T, Hoshino T, Skiya Y, Miki S, Yoshikawa T, Hayashi N, Abe O, Development of training environment for deep learning with medical images on supercomputer system based on asynchronous parallel Bayesian optimization. J Supercomput 2020 Jan 20 (Epub ahead of print)

(2) 国際会議プロシーディングス（査読あり）

[2] Nomura Y, Hanawa T, Sato I, Hanaoka S, Nakao T, Murata M, Takenaga T, Hoshino T, Sekiya Y, Hayashi N, Abe O, Development of training environment for deep learning with medical images on supercomputer system based on asynchronous parallel Bayesian optimization, ISC High Performance 2019, Research Poster 20.

[3] Nomura Y, Hanaoka S, Yoshikawa T, Sato I, Nakao T, Murata M, Takenaga T, Koshino S, Miki S, Watadani T, Hayashi N, Abe O, Preliminary study

of automated detection of pulmonary nodule in ultrashort echo time MR images, Int J CARS 2019;14(Suppl.1): S82-S83.

[4] Nakao T, Hanaoka S, Nomura Y, Murata M, Takenaga T, Miki S, Watadani T, Yoshikawa T, Hayashi N, Abe O, Deep generative model-based unsupervised detection of inappropriate images in a chest X-ray dataset, Int J CARS 2019;14(Suppl.1): S170-S171.

(3) 国際会議発表（査読なし）

[5] Hanaoka S, Abnormality emphasize using a flow-based deep generative model for chest X-ray images, UTokyo-NTU Joint Conference 2019, 2019 年 12 月.

[6] Hashizume A, Hanawa T, DNN training using multiple GPUs for medical image recognition, HPC in Asia 2020 Poster, 2020 年 1 月.

(4) 国内会議発表（査読なし）

[7] 花岡昇平, 野村行弘, 佐藤一誠, 中尾貴祐, 竹永智美, 渡谷岳行, 三木聡一郎, 吉川健啓, 林直人, 阿部修, glow による胸部単純写真の異常強調と検知, 第 3 回人工知能応用医用画像研究会, 2019 年 6 月.

[8] 城間祐花, 竹永智美, 野村行弘, 吉川健啓, 林直人, 阿部慎司, 胸部 ultra short TE MR 画像における深層学習を用いた肺野セグメンテーションの初期検討, 日本放射線技術学会第 66 回関東支部研究発表大会, 2019 年 11 月.

[9] 橋詰明宗, 堀敏博, 医用画像診断支援に向けた複数 GPU による Deep Learning 学習の検討, 情報処理学会ハイパフォーマンスコンピューティング研究発表会, 2019 年 12 月.

(5) その他（特許, プレスリリース, 著書等）

[10] Shibata H, Hanaoka S, Nomura Y, Nakao T, Sato I, Hayashi N, Abe O, Anomaly detection in chest radiographs with a weakly supervised flow-based deep learning method, arXiv:2001.07847, Jan 2020