

jh200042-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 (EOB-MR 画像における肝結節病変検出/鑑別, 胸部 X 線写真の病変検出, 胸部 FDG-PET/CT 画像における異常検知など) の開発を進めるとともに, CAD の臨床使用を想定した高性能かつ小規模な DL 構造探索方法について検討した。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

超大規模データ処理系応用分野

(3) 参加研究者の役割分担

佐藤 一誠	DL 構造探索, CAD 開発
埴 敏博	GPU 間通信, ノード間通信, KNL 最適化
野村 行弘	DL 構造探索, KNL 最適化 CAD 開発
林 直人	臨床面での評価
花岡 昇平	DL 構造探索, CAD 開発 臨床面での評価
竹永 智美	CAD 開発
中尾 貴祐	CAD 開発, 臨床面での評価
柴田 寿一	DL 構造探索, CAD 開発
佐藤 大介	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
鈴木 陽大	DL 構造探索
鈴木 雄介	CAD 開発, 臨床面での評価

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 項目について併せて検討を行う。

(a) 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)および Oakbridge-CX システム(OBCX)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。

(b) CAD の臨床使用を想定した高性能かつ小規模な DL 構造探索方法を構築する。

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）、1 名の乳腺外科医（鈴木雄介）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

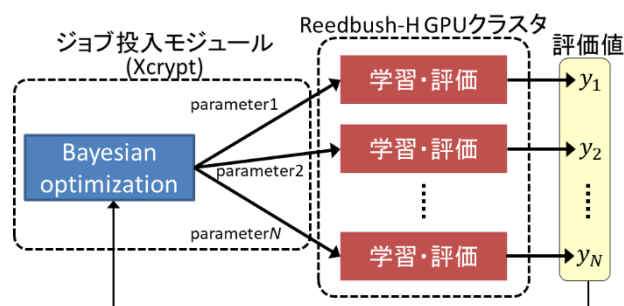


図 1 実装した DL の学習方法の流れ

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

2018 年度までに非同期並列型ベイズの最適化 (Bayesian optimization, BO) [B]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[C]. 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 (Reedbush-H/L) のログインノード上でジョブ並列スクリプト言語 (Xcrypt) [D]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 Reedbush-H/L の計算ノードに繰り返し投入する。

構築した DL 学習フレームワークを用いて 6 種類の CAD の開発を進めている。これらのうち、(b)～(g)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) 胸部 MR 画像の肺結節検出
- (d) 頭部造影 MR 画像の転移性脳腫瘍検出
- (e) EOB-MR 画像における肝結節病変検出
- (f) 頭部救急 CT 画像の異常検知
- (g) FDG-PET/CT 画像における異常検知

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

[C] Nomura Y, J. Supercomputing 2020;76:7315-7332

[D] Hiraishi T, WHIST 2012, 2012.

5. 今年度の研究成果の詳細

(1) DL 構造探索方法の構築

DL モデルの構造探索方法として Neural architecture search (NAS) [E]と呼ばれる方法が注目されているが、他施設での CAD の運用を考えた場合、施設ごとに計算機環境が異なるため、施設ごとの計算機環境に合わせたサイズの構造を模索する必要があり、モデルや学習の維持コストがかかり過ぎるという結論に達した。そのため、計算機環境の十分な施設で 1 度大きなモデルを学習し、その後モデルサイズを学習無しでモデル構造を可変にする手法の開発を行った。具体的には、DL モデルの汎化性能を上げるための Dropout と呼ばれる手法に着目し、ミニバッチごとに構造的な Dropout を用いて、大きいサイズのモデルを小さいサイズのモデルにランダムに変換しながら学習を行った。実験として 3 段,4 段,5 段の U-Net に変換するような構造的 Dropout を作り、CVC-ClinicDB [F]と呼ばれる大腸内視鏡検査ビデオで学習しポリープ検知のための U-Net を構築した。結果としては、個別に学習した場合が 3 段の Dice 係数が 66.6%、4 段で 74.3%、5 段で 87.2%であったのに対して、5 段の U-Net を 1 度学習させたのち 3 段、4 段と可変にした場合の精度が 3 段 60.0%、4 段 77.9%、5 段 88.13%となった。したがって、1 度の学習でその後学習無しに構造を変化させても個別にその構造で学習した場合と同等

の結果が得られている。今後は、段数だけでなく特徴マップ数も学習後に可変にすることをを行うのと、最も小さいサイズの性能が下がる現象が他の実験でも見られたためこの原因解明と解決策を模索する予定である。さらに最も大きいサイズの出力を模倣学習することで小さいサイズの性能を上げる方法も考慮する。また、構造的 Dropout による学習は通常の学習に比べて収束が遅いため、並列化による高速化も考える必要がある。今回の方法では、計算資源の豊富な施設で可能な限り大きいサイズのモデルを 1 度学習すればその後は実行する施設の計算資源に合わせてサイズを変えられるため、大規模情報基盤の共同利用に適したアプローチであると考えられる。

[E] Elsken T, J. Mach. Learn. Res. 2019;20:1-21

[F] Bernal J, Comput Med Imag Grap. 2015;43:99-111

(2) DL を用いた CAD の開発

2018 年度までに構築した学習環境において DL を用いた CAD の開発を継続している。以下に昨年度までに開発を開始していたもののうち、大幅な性能改善が図れたもしくは新たな成果が得られた 4 種類の CAD の結果を示す。

(2-a) FC-ResNet を用いた EOB-MR 画像における肝結節病変の検出・分類

2018 年度に実施した EOB-MR 画像における肝臓領域抽出で用いた FC-ResNet [G]を改変し、単

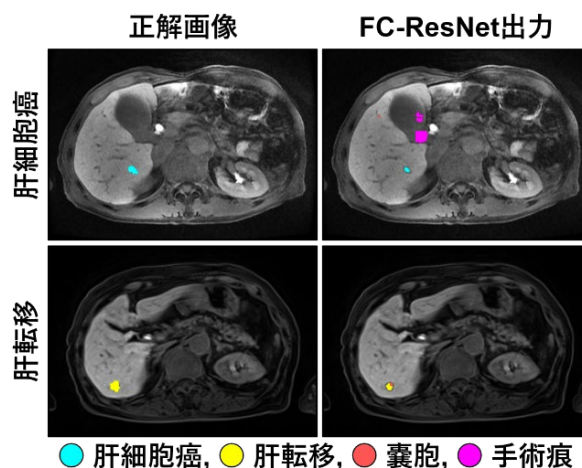


図 2 : 肝結節病変検出・分類結果

一ネットワークで肝結節病変（肝細胞癌、転移、血管腫、嚢胞、手術痕）の検出および分類を同時に行った。この結果、検出感度 60%における症例あたりの FP 数はそれぞれ 25 個であり、病変の分類精度は 0.79 であった。図 2 に検出・分類結果例を示す。この成果は論文誌へ投稿中である。

(2-b) 胸部 X 線写真の異常検知・異常強調 (研究業績[1],[2],[5])

胸部 X 線写真で診断可能な疾患や異常は多岐にわたることを鑑み、昨年度より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を進めている。まず、 α -GAN を用いた異常検知手法について各種病変の検知を試みた結果、ROC 曲線の下面積 (ROC-AUC) が 0.752 であった[1]。また、可逆変換が可能なフローベース深層生成モデルである Glow [H]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 と性能が向上した[5]。さらに、Glow を用いた病変の異常強調手法を構築し (図 3) [3], 9 名の放射線科医師による観察実験を行った。その結果、強調画像を用いることで放射線科医の病変検出感度が向上した。

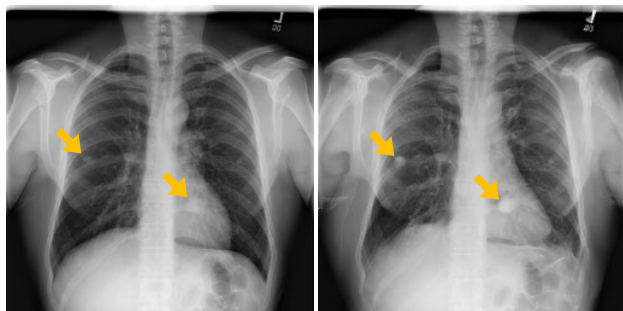


図 3: Glow を用いた胸部 X 線写真の異常強調 (左: 元画像, 右: 強調結果, 矢印: 肺腫瘍)

(2-c) 頭部救急 CT 画像の異常検知

胸部 X 線写真の異常検知で用いた Glow を 3 次元データ用に拡張した上で、頭部救急 CT 画像の異常検知に適用した。その結果、ROC-AUC が 0.866 であった。この成果は論文誌へ投稿中である。

(2-d) 胸部 FDG-PET/CT 画像における異常検知 (研究業績[4])

正常症例のみを学習に用い、テスト症例中の FDG-PET の異常集積を検出する異常検知方法を構築した。この方法では、CT 画像を入力すると正常と仮定した場合の PET の画素値 (SUV 値) の期待値と期待標準誤差を推定して出力する 2DU-Net を正常症例 1,800 例で学習した。検出では CT 画像から予測した SUV 値の確率分布と PET の SUV 値より異常度 (z-score) マップを計算し、3.0 を閾値として二値化して得られた連結領域を最終的な出力とした。異常集積が認められた胸部病変を含む 34 例で評価した結果、症例あたりの FP 数が 3 個のときの検出感度が 77.8%であった。図 4 に検出結果例を示す。現在、対象範囲を全身とした場合で検討を進めている。

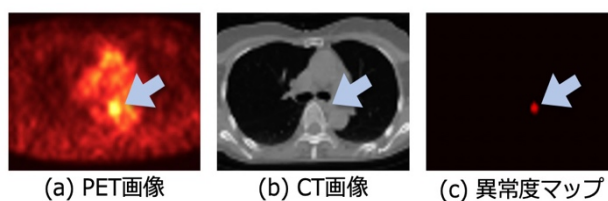


図 4: 検出例 (矢印: 縦隔結節による異常集積)。

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

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

(3) メニーコア型プロセッサを用いた大規模 DL 学習の検証

OFP および OBCX に PyTorch 1.5.0 および horovod をインストールした上で Reedbush-H/L で構築した DL 学習フレームワークを実装した。その際、投入ジョブに使用するノード数を可変に設定できるように調整を行った。

構築した環境において U-Net を用いたマンモグラフィ画像 (2048×1280) の腫瘍検出の学習を試みた。Reedbush-L (1 ノード) では 16,338 秒 (1GPU), 1,678 秒 (4 GPUs) で完了した学習が OFP では 23,090 秒 (ノード数: 4, 64 コア/ノード), OBCX では 15,759 秒であった (ノード数: 4, 54 コア/ノード)。

6. 今年度の進捗状況と今後の展望

今年度は新型コロナウイルス感染拡大の影響により、研究の遂行が停滞していた期間があったものの、実施した3項目(DL構造探索方法の構築、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)についてそれぞれ一定の成果が得られた。

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おり、論文を投稿中である。また、肝結節病変については、病変の検出だけではなく分類も同時に実現できた。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。

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、DL学習フレームワークを導入できたものの、OFPならびにOBCXの混雑状況の影響により初期的な検討のみに留まり、計算資源を十分に利用できなかった。

フローベース深層生成モデルによる異常検知は良好な性能が得られているものの、学習画像のサイズが大きい場合(2次元画像で1024×1024以上、3次元画像で64×64×64以上を想定)に勾配情報の保持に伴うGPUメモリ量の制約から学習が困難となる。そのため、モデル並列化の導入や使用GPUメモリ量を抑制可能なアルゴリズムの導入により、利用メモリ量の問題を克服していく予定である。

2021年11月末にReedbush-H/Lの運用終了となることに伴い、東大Wisteria/BDEC-01にDL学習環境を構築した上で研究を継続していく予定である。

7. 研究業績一覧(発表予定も含む。投稿中・投稿予定は含まない)

(1) 学術論文 (査読あり)

[1] T. Nakao, S. Hanaoka, Y. Nomura, M. Murata, T. Takenaga, S. Miki, T. Watadani, T. Yoshikawa, N. Hayashi, O. Abe, 'Unsupervised deep anomaly

detection in chest radiographs', Journal of Digital Imaging 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)

該当無し

(3) 国際会議発表 (査読なし)

該当無し

(4) 国内会議発表 (査読なし)

[2] 野村行弘, 花岡昇平, 竹永智美, 中尾貴祐, 柴田寿一, 三木聡一郎, 吉川健啓, 渡谷岳行, 林直人, 阿部修, ディープラーニングに基づく3次元汎用半自動病変セグメンテーションの初期検討, 第39回日本医用画像工学会大会, 2020年9月.

[3] 花岡昇平, 野村行弘, 柴田寿一, 三木聡一郎, 中尾貴祐, 林直人, 吉川健啓, 渡谷岳行, 阿部修, Glowによる胸部単純写真における教師なし異常強調の試み, 第33回電子情報研究会・第3回日本医用画像人工知能研究会合同研究会, 2020年11月.

[4] 中尾貴祐, 花岡昇平, 野村行弘, 林直人, 阿部修, ベイズ深層学習を用いた全身FDG-PET/CTにおける胸部異常検知, 第80回日本医学放射線学会, 2021年4月.

(5) 公開したライブラリなど

該当無し

(6) その他(特許, プレスリリース, 著書等)

[5] H. Shibata, S. Hanaoka, Y. Nomura, T. Nakao, I. Sato, N. Hayashi, O. Abe, Anomaly detection in chest radiographs with a weakly supervised flow-based deep learning method, arXiv:2001.07847v3, Oct 2020