

jh210011-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, 説明可能な異常検知手法, およびフローベース深層生成モデルでの高解像度架空画像の生成手法について検討を実施した。

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 CAD 開発, 異常検知手法確立, 架空画像生成手法確立

埜 敏博 GPU 間通信, モデル並列化, ノード間通信,

野村 行弘 CAD 開発, 異常検知手法確立
林 直人 臨床面での評価

花岡 昇平 CAD 開発, 架空画像生成手法確立, 臨床面での評価

中尾 貴祐 CAD 開発, 異常検知手法確立, 臨床面での評価

竹永 智美 CAD 開発

柴田 寿一 CAD 開発, 異常検知手法確立, 架空画像生成手法確立

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

鈴木 雄介 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 CAD 開発, 異常検知手法確立

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 2018 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 項目についても検討を行う。

- (a) 異常検知手法における，異常判定の根拠を空間的に可視化できるアルゴリズムを確立することで，異常検知手法の信頼性向上を目指す。
- (b) 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，入力画像サイズが大きい場合でもフローベース深層生成モデルを用いた異常検知手法が学習できるようにする。

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

2018 年度までに非同期並列型ベイズ的最適化 (Bayesian optimization, BO) [B]に基づくパラ

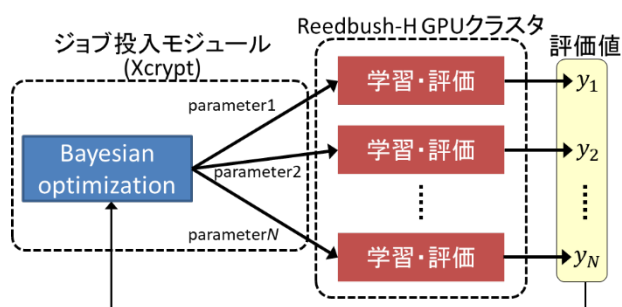


図 1 実装した DL の学習方法の流れ

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) 説明可能な異常検知手法の確立

病変を含む医用画像全体の解像度を病変が見えない程度に粗くし，正常症例を事前知識とするフローベース深層生成モデルである Glow [E]で正常な細かい構造を復元し，復元画像と元画像の差分を可視化することが当初の計画であった。しかし，

画像を粗くする操作には粗さの恣意性が残り、大小の病変を同時かつ統一的に検出することが困難である。最近、SwiecickiらはInpainting(欠損修復)の技術を採用し、マンモグラフィの一部を意図的に欠損させ、欠損箇所を正常な組織構造を事前知識とするGANで欠損修復し、修復誤差を可視化した[F]。この手法は大小の病変を同時かつ統一的に検出できると考えられる。これらを受けて、これまでにフローベース深層生成モデルによる欠損修復の定式化と実装を完了した。

(2) フローベース深層生成モデルでの高解像度架空画像の生成手法の確立 (研究業績[6])

当初の予定では大局的かつ局所的に高解像度画像を学習できるフローベース深層生成モデルを確立することを想定していた。しかし、生成対象の架空画像の解像度が高くなり(ここでは3次元で256×256×256程度かそれ以上を想定)、メモリ不足の問題が表面化する前に、ある程度まで解像度を高く(ここでは3次元で128×128×128程度を想定)すると、フローベース深層生成モデルによる画像の学習が進まなくなることが判明した。この課題を克服するため、プログレッシブ学習を提案した。プログレッシブ学習では学習する画像の諧調を1bitから徐々に増やしていくものである。我々は世界で初めてプログレッシブ学習をフローベース深層生成モデルに適用したが、このプログレッシブ学習は潜在的にはGANやVAEの学習にも適用できるものである。さらに、プログレッシブ学習によってフローベース深層生成モデルでは世界初となる高解像度(128×128×128)の3次元胸部CTモデルの構築とそのモデルによる高解像度架空画像の生成手法を確立することに成功した。

(3) 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、新たな成果が得られた結果を示す。

(3-a) EOB-MR 画像における肝結節病変の検出・分類 (研究業績[1])

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。現在、画像診断レポート自動生成に向けた肝臓領域の亜区域分類について検討中である。

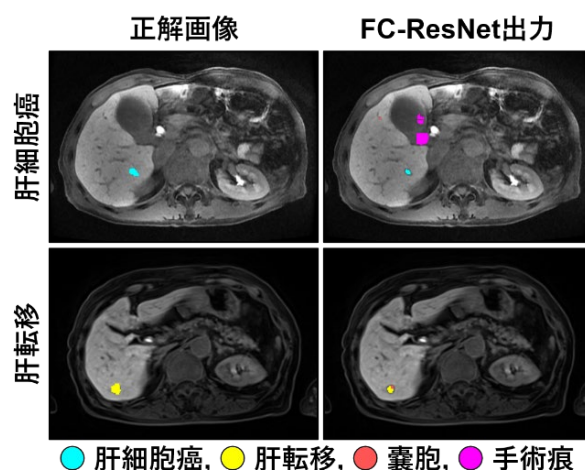


図2: 肝結節病変検出・分類結果

(2-b) 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

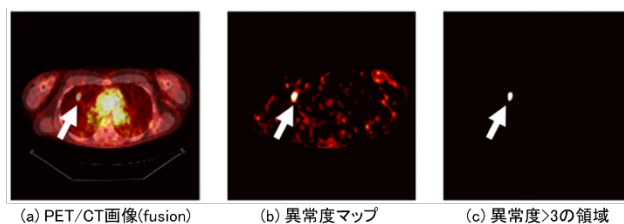


図3: 検出例 (矢印: 肺腫瘍による異常集積)。

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 82.2%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．

(3-c) フローベース深層生成モデル Glow を用いた異常検知 (研究業績[3])

フローベース深層生成モデルである Glow を用いた異常検知手法について検討した．胸部 X 線写真の異常検知では ROC 曲線の下面積 (ROC-AUC) が 0.783 であった．また，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．

(3-d) 架空画像を用いた胸部 X 線写真の病変検出 (研究業績[5])

正常な胸部 X 線写真を学習させた Glow より生成した，架空の胸部 X 線写真に対して架空の (生成された) 病変を埋め込んだ画像を用いて U-Net に基づく病変検出モデルを学習した．日本放射線技術学会より公開されている胸部 X 線写真のデータセットを用いた評価の結果，症例あたりの偽陽性数が 1 個以下で，感度 60%程度であった．

(3-e) 医用画像解析における Sim2Real 機械学習のためのベンチマークデータと学習手法の提案 (研究業績[7])

ロボットの制御分野においてシミュレーションを用いて学習を行い，実世界に適用する Sim2Real 機械学習が注目を集めており，この中でも Domain randomization (DR) と呼ばれる枠組みが近年注目を集めている．従来の Sim2Real はシミュレータによって生成された教師データから現実世界に転移させる際に転移学習やドメイン適応という方法がとられていたのに対して，DR ではできるだけ多くの多用なパターンのデータを生成することで様々なデータから学習し，深層学習がもつ表現力によって現実世界を内包する幅広いデータ空間に対応する識別器を学習することを目標と

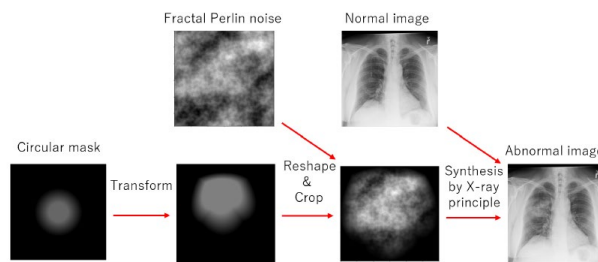


図 4: 疑似肺炎胸部 X 線写真の作成

している．本研究では，胸部 X 線写真の肺炎検出を対象に，シミュレーションにより作成された疑似病変データを用いて深層学習に基づく物体検出手法である Single shot multibox detector (SSD) [H]モデルを学習した．

ベンチマークデータセットとして，シミュレータを用いて放射線科医が見分けづらい 101 症例のデータを作成した．シミュレータは，正常な胸部 X 写真に対してフラクタルパーリンノイズを基に X 線の原理から病変を付加する (図 4)．シミュレーションからの学習が本質的には継続学習であり，さらに問題の難易度を自由に設定できるという点に着目し，カリキュラム学習と継続学習を DR に組み合わせることを考えた．さらに，カリキュラム学習として，人は中庸を好むというゴルディロックス原理に基づく手法 (ゴルディロックスカリキュラム) を提案した．

表 1 に実験結果を示す．UDR が DR で完全にランダムにシミュレータのパラメータを割り振りながら学習する手法である．BayRn が既存手法 [I] で実際のデータの性能が良くなるようにシミュレータのパラメータを最適化しながら学習する DR 手法である．GDR は本研究で提案したゴルディロックスカリキュラム DR となる．さらに，カリキュラムとして易しい問題から難しい問題へと変化させる Easy2Hard との比較を行っている．学習手法として SGD, Adam, NVRM-SGD [J] を用いている．NVRM-SGD は継続学習特有の問題である破滅的忘却を緩和する手法であるが詳細は後述する．表 1 より，GDR は基本的に性能が良く，特に NVRM-SGD を用いたものが最良である．free-response receiver operating characteristic curve (FAUC) は通常の AUC (Area Under Curve) とは全く異なる

表 1: 実験結果

Method	FAUC	CPM	TPR@FPI0.2
UDR (Adam)	0.324 ± 0.037	0.394 ± 0.026	0.245 ± 0.040
UDR (SGD)	0.257 ± 0.021	0.335 ± 0.032	0.197 ± 0.011
UDR (NVRM-SGD)	0.246 ± 0.023	0.361 ± 0.030	0.175 ± 0.029
BayRn (Adam)	0.320 ± 0.035	0.410 ± 0.044	0.251 ± 0.056
BayRn (SGD)	0.302 ± 0.044	0.388 ± 0.049	0.232 ± 0.038
BayRn (NVRM-SGD)	0.282 ± 0.026	0.384 ± 0.028	0.203 ± 0.028
GDR (Adam)	0.352 ± 0.028	0.452 ± 0.027	0.275 ± 0.039
GDR (SGD)	0.377 ± 0.032	0.474 ± 0.036	0.302 ± 0.035
GDR (NVRM-SGD)	0.378 ± 0.015	0.491 ± 0.009	0.331 ± 0.034
GDR (SAM)	0.354 ± 0.033	0.471 ± 0.022	0.291 ± 0.046
Easy2Hard-1 (SGD)	0.350 ± 0.033	0.457 ± 0.049	0.276 ± 0.050
Easy2Hard-1 (NVRM-SGD)	0.362 ± 0.024	0.470 ± 0.016	0.297 ± 0.038
Easy2Hard-2 (SGD)	0.365 ± 0.017	0.459 ± 0.032	0.289 ± 0.015
Easy2Hard-2 (NVRM-SGD)	0.363 ± 0.044	0.466 ± 0.046	0.320 ± 0.054

る指標である（具体的には 0.5 を下回ることが十分にあり）。カリキュラム学習のターゲット値を FAUC と設定しているため、Easy2Hard-1 では FAUC が 1 から段階的に 0.1 刻みで低くなっていくような設定をしているのに対して、Easy2Hard-2 では 0.75 から 0.3 に減っていくような設定になっている。GDR では FAUC の値を 0.6 に固定している。Easy2Hard-2 の方が Easy2Hard-1 よりも性能が良くなっているのは、難易度はある程度狭い範囲で固定したほうが良いからではないかと考えている。

表 2 は、GDR においてターゲット値である FAUC を変化させた場合の性能の変化を調べたものである。表 2 の結果より、FAUC=0.6 が最良であった。

本研究では、DR を本来の目的である多種多様なデータから学習することに対して、継続学習とカリキュラム学習を利用してある一定の制約のもと学習することで、より効率的な DR 手法を開発することを目的とした。しかし、実際に進めると、継続学習特有の問題である破滅的忘却が見られることが分かった。これは深層学習を継続的に学習するとある時点でそれまで学習していたことを忘却してしまうという性質である。これに対して、すでに我々は破滅的忘却を緩和する手法として NVRM-SGD という学習手法を提案していたため、この手法を GDR に応用した。実際に Adam などといった他の学習手法を用いると破滅的忘却が見られ性能が低くなるが、NVRM-SGD によって GDR を補うことで性能低下が軽減されていることがわかった。

本研究は現在論文投稿中であるが、arXiv にて

表 2: GDR における難易度設定の分析

Target	FAUC	CPM	TPR at FPs/Image=0.2
FAUC=0.3	0.335 ± 0.047	0.444 ± 0.047	0.277 ± 0.058
FAUC=0.4	0.344 ± 0.040	0.457 ± 0.039	0.307 ± 0.048
FAUC=0.5	0.365 ± 0.025	0.481 ± 0.020	0.302 ± 0.026
FAUC=0.6	0.377 ± 0.015	0.483 ± 0.017	0.312 ± 0.034
FAUC=0.7	0.357 ± 0.021	0.471 ± 0.019	0.283 ± 0.035

公開している（研究業績[7]）。作成したシミュレータおよびベンチマークデータセットについては、公開の準備を進めている。

- [E] Kingma DP, NeurIPS 2018:10236-10245
- [F] Swiecicki A, Sci. Rep. 2021;11, 10276
- [G] Takenaga T, IJCARS 2019;14(8):1259-1266
- [H] Liu W, ECCV2016
- [I] Muratore F, ICRA 2021
- [J] Xie Z, Neural computation 2021

6. 今年度の進捗状況と今後の展望

年度途中に使用するスーパーコンピュータが東大 Reedbush-H/L から東大 Wisteria/BDEC-01 へ変更となることに伴い、新たな研究倫理申請や研究データの移行などにより停滞した期間があったものの、実施した 3 項目（説明可能な異常検知手法の確立、フローベース深層生成モデルでの高解像度架空画像の生成手法の確立、DL を用いた CAD の開発）についてそれぞれ一定の成果が得られた。また、昨年度までの実施内容のうち、DL を用いた汎用的な病変形状ラベル推定方法の構築について論文誌に掲載された。

今年度で JHPCN 共同研究としての取り組みは一区切りとなるが、今後も東大 Wisteria/BDEC-01 の使用を継続するため、これまでに構築した学習フレームワークなどを使用した新たな CAD 開発を進めていく予定である。

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] Takenaga T, Hanaoka S, Nomura Y, Nakao T, Shibata H, Miki S, Yoshikawa T, Hayashi N, Abe O, Multichannel three-dimensional fully

convolutional residual network-based focal liver lesion detection and classification in Gd-EOB-DTPA-enhanced MRI. *Int J Comput Assist Radiol Surg.* 2021 Sep;16(9):1527-1536

[2] Nomura Y, Hanaoka S, Takenaga T, Nakao T, Shibata H, Miki S, Yoshikawa T, Watadani T, Hayashi N, Abe O, Preliminary study of generalized semiautomatic segmentation for 3D voxel labeling of lesions based on deep learning. *Int J Comput Assist Radiol Surg.* 2021 Nov;16(11):1901-1913.

[3] Shibata H, Hanaoka S, Nomura Y, Nakao T, Sato I, Sato D, Hayashi N, Abe O, Versatile anomaly detection method for medical images with semi-supervised flow-based generative models. *Int J Comput Assist Radiol Surg.* 2021 Dec;16(12):2261-2267.

[4] Nakao T, Hanaoka S, Nomura Y, Hayashi N, Abe O, Anomaly detection in chest ¹⁸F-FDG PET/CT by Bayesian deep learning. *Jpn J Radiol.* 2022 Jan 30 (Epub ahead of print).

Takenaga T, Hayashi N, Abe O, X2CT-FLOW: Maximum a posteriori reconstruction using a progressive flow-based deep generative model for ultra sparse-view computed tomography in ultra low-dose protocols. *arXiv preprint 2021; arXiv:2104.04179v2*

[7] Suzuki T, Hanaoka S, Sato I, Goldilocks-curriculum Domain Randomization and Fractal Perlin Noise with Application to Sim2Real Pneumonia Lesion Detection. *arXiv preprint 2022; arXiv:2204.13849*

(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)

[5]花岡昇平, 野村行弘, 林直人, 吉川健啓, 柴田寿一, 中尾貴祐, 阿部修, 架空症例および架空病変を学習に用いた胸部単純写真における結節性病変自動検知. 第34回電子情報研究会・第4回日本医用画像人工知能研究会合同学術集会, 2021.9.19

(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)

[6] Shibata H, Hanaoka S, Nomura Y, Nakao T,