コンピュータ外科における画像ベース深層学習の利用 Deep Learning in Surgical Assistance

小田昌宏（正会員）*

名古屋大学大学院情報学研究科

Masahiro Oda

Graduate School of Informatics, Nagoya University

Key words

Deep learning, Image processing

この度は2017年度CAS Young Investigator Award ゴールド賞(日立賞)をいただきありがとうございました。ご指導ご鞭撻いただいた森健策先生、森研究室の皆様、ならびに研究においてご協力いただいた先生方に深く感謝致します。

深層学習を用いた医用画像処理研究の成果を提案している立場から、コンピュータ外科においても深層学習の利用がさらに活発となることを望んでいます。本稿ではコンピュータ外科での深層学習活用の難しさと対策を述べさせていただきます。

1. 画像ベース深層学習を用いた医療支援

コンピュータを用いた医療支援において、深層学習を用いた研究成果が多数報告されている。AIと呼ばれる場合もあるが、その実態は深層学習またはほかの機械学習を用いた技術である場合が多い。深層学習は画像、テキスト、ゲノム情報等の多様なデータに対して適用されているが、本稿では主に画像データについて述べる。これまで提案された深層学習ベースの医療支援手法にはCT像等の術前画像を用いたものが多く、外科で用いられる術中画像を用いたものはまだ少ない。これはなぜだろうか。

2. 深層学習の適用が多い・少ない画像モダリティ

深層学習を用いた医療支援において多く利用される画像は、検査用X線像、CT像、眼底画像である。CT像からの臓器セグメンテーション手法^{1),2)}、眼底画像からの血管セグメンテーション手法^{3),4)}等が提案されている。一方、利用が少ない画像は、術中X線像、超音波像、病理画像が挙げられる。これらの主な違いは、臓器や解剖構造の撮影位置や範囲が統一されているかどうかである。X線像やCT像は観察対象臓器や疾患が決まっていれば、撮影範囲の変化は少なくなる。臓器や解剖構造の撮影位置や範囲が統一されていれば、コンピュータによる画像解析において考慮すべきバリエーションが限定される。この点が深層学習を用いた医療支援において良好な成果を挙げるうえで重要となる。画像に含まれる情報には、臓器や疾患の形状、濃度値、患者の年齢、撮影位置や範囲、撮影機器等の様々なバリエーションが含まれ、そのバリエーションが多いほど画像からの支援情報抽出が難しくなる。そのため、研究の進度に応じてバリエーションの種類を適度に限定することが、研究開発で求められる。

3. 外科での深層学習利用の難しさ

外科で利用される画像はX線像、内視鏡像、超音波像等様々である。これらは術式、患者状態、医師の経験等によって撮影位置や範囲が大きく変化する。画像モダリティによっては、カメラパラメータ、照明条件、

*名古屋大学大学院情報学研究科
〒464-8601 愛知県名古屋市千種区不老町
moda@mori.m.is.nagoya-u.ac.jp

画像強調条件等が関わり、画像に含まれる情報のバリエーションが非常に大きい。このことが深層学習を用いたコンピュータ支援の研究開発を難しくしている。

一般画像を対象とした深層学習による画像認識理解の成果に目を向けると、カメラや照明条件のバリエーションに富む画像が与えられる場合でも、深層学習は高い成果を挙げている。例えばインターネットから画像を収集した ImageNet を用いて、深層学習を用いた高精度な画像分類が行われている。このような成果を挙げるためには、様々なカメラや照明条件で撮影された画像を収集し、画像に含まれる情報変化の範囲をなるべく網羅した学習データセットを用意する必要がある。可能な限り多くの画像を集めて学習データとし、深層学習アルゴリズムの学習に用いるべきである。

一方、外科で利用される画像は対象部位や疾患ごとに撮影位置や範囲、さらにモダリティが大きく異なり、それらをまとめて扱うことは困難である。しかし対象部位・疾患別に画像を収集すると、深層学習アルゴリズムの学習に十分な量を収集することが難しい。このように対象ごとに画像が細分化される点も、外科での深層学習を用いたコンピュータ支援の利用の難易度を上げている。

4. 深層学習を用いたコンピュータ外科支援の研究開発

ここまでで深層学習利用の困難さを挙げたが、それでは深層学習をどのように研究開発の中で利用するのが良いか。研究開発を行う上での利用法を考える。

4.1 研究進展に応じた対象の段階的拡大

外科で用いられる術中画像等を用いたコンピュータ支援手法を研究開発するうえでは、対象部位、疾患、術式、画像モダリティ、撮影機器を限定した状態で研究を開始すべきである。ある程度成果が得られたあとに限定範囲を徐々に拡大し、実用的なレベルへと近づけていく。これにより、研究成果が得られないことによる研究者のモチベーション低下の防止、問題発生時の原因発見の迅速化が期待できる。

4.2 深層学習を用いた多様な情報の横断的利用

コンピュータ支援手法の研究開発においては画像情報を用いるのみでなく、複数の画像モダリティ、術中ナビゲーションシステムから得られるセンサ情報、診療録を含むテキスト情報等の多様な情報を横断的に活用することも考慮されたい。従来は画像とテキストの

ように種類の異なる情報を組み合わせることが難しかった。一方、深層学習では画像、テキスト、センサ情報等をすべてベクトル化しネットワークに入力する。これらの情報はネットワーク内で結合や分離が可能である。そのため多様な情報を入力して情報横断的判断を行う手法の組み立てが可能である。深層学習を用いた情報横断的なコンピュータ支援手法の開発と、従来から一段進んだ術中支援の実現が期待される。

5. コンピュータ外科における深層学習利用研究の紹介

深層学習では様々な情報を利用することが可能であることを述べた。その例として、術中に取得される画像および位置情報を用いた深層学習の適用例を紹介する。情報の横断的利用によるコンピュータ外科支援を考えるうえでの参考とされたい。

5.1 画像データを用いた深層学習

腹腔鏡画像を対象とした研究が登場している。Twinnanda らは Convolutional Neural Network (CNN) を用い、腹腔鏡画像のシーン分類および術具検出を行った⁵⁾。Varytimidis らは腹腔鏡画像の類似シーン検索に CNN を用いている⁶⁾。Aksamentov らは内視鏡動画から残りの手術時間を推定する方法を提案している⁷⁾。動画に含まれる時系列情報を活用するため、Long Short-term Memory (LSTM) 層を含む CNN を用いた回帰を実施している。

5.2 位置・姿勢データを用いた深層学習

術具に位置センサを取り付けることで、術中の位置・姿勢およびそれらの時系列変化からなる軌跡を計測することができる。位置・姿勢データを用いた深層学習の例を紹介する。

Fawaz らは da Vinci のアームの軌跡データから操作者のスキルを3段階評価する CNN を提案した⁸⁾。CNN には座標値等を直接的に入力しており、座標値の集合から3層 CNN によってスキル評価を実現している。

我々は、LSTM 層をもつ CNN によって内視鏡挿入中の大腸形状を推定する方法を提案した⁹⁾。我々が開発している大腸内視鏡ナビゲーションシステムは、大腸内視鏡を用いた診断および治療時に大腸内の内視鏡位置を提示し、病変への到達を支援する (Fig.1)。本システムは、CT 像から取得した大腸形状および、位置センサを取り付けた内視鏡から得られる内視鏡形状を利用し、大腸形状と内視鏡形状をレジストレーションす

Fig. 1 大腸内視鏡ナビゲーションシステムの構成

Fig. 2 Shape Estimation Network⁹⁾

ることで、大腸内での内視鏡先端位置を算出する。しかし、内視鏡挿入中に大腸は大きく変形するため、レジストレーションのずれが生じていた。そこで、大腸形状は挿入されている内視鏡形状の影響を強く受けることに着目し、内視鏡形状の時系列変化から大腸形状を予測するネットワークである Shape Estimation Network を提案した (Fig.2)。本ネットワークは時間軸に沿った内視鏡形状の特徴量を入力すると、畳み込み層および LSTM 層を経由して、推定された大腸形状を出力する。大腸ファントムを用いた実験において、大腸形状が良好に推定されることを確認した (Fig.3)。

6. 公開データ整備によるコンピュータ外科での深層学習研究推進

外科でのコンピュータ支援の高性能化と研究の活性化を図るためにはどうすればよいだろうか。深層学習利用の観点から考える。

深層学習を研究開発に利用するうえで、最も時間と労力を要する部分はアノテーション付きデータの収集である。アノテーション付きデータの収集には以下の3段階のプロセスが存在する。

アノテーション付きデータ収集の3プロセス

1) データ収集プロトコルの策定

開発するコンピュータ支援手法を念頭に置いて、医

Fig. 3 大腸形状推定結果を用いた大腸内視鏡ナビゲーション画面

学および工学研究者の話し合いのなかで、収集データの詳細な種類や記録手順を決定する。画像を収集するのであれば、撮影機器、画像サイズ、撮影タイミング、保存形式等を決定する。

2) データ収集

1) で定めた手順に従いデータを収集する。前向きデータの収集は十分なデータを集めるのに時間を要するため、後ろ向き検討として既存のデータを利用する場合も多い。

3) データへのアノテーション付与

収集したデータへ手動でアノテーション情報を付与する。データが画像である場合は、アノテーション付与は以下3段階に分類される。

画像アノテーション付与のレベル分類

3-1) 画像単位アノテーション

画像1枚に対し1つのラベル値（クラス番号等の離散値、連続値で表された評価値等）を付与する。画像1枚に複数の値を付与する場合もある。

3-2) 物体単位アノテーション

画像に含まれる物体に対し、位置およびラベル値を付与する。位置は単一の座標値やバウンディングボックスで表現される場合が多い。

3-3) 画素単位アノテーション

画像に含まれる物体に対し、物体と背景を境界で切り分けた物体領域とラベル値を付与する。セマンティックアノテーションとも呼ばれる。

アノテーション付きデータ収集の3プロセスのなかで、2) および 3) が時間と労力を要する。特に 3-3) 画素単位アノテーションを行う場合は 3) に要する時間が大きく増加する。深層学習を用いた研究開発を行ううえでこの部分が障壁となり、成果が得られるまで

に長い時間を要する。1年から3年程度の期間の短い研究計画や研究資金に基づいて研究を実施するうえで、データ準備での時間の浪費は致命的である。

アノテーション付きデータ収集の負担を軽減するには、既存の公開データ利用が効果的である。データ準備の時間が短縮されるほか、公開データで深層学習モデルの学習を行い、その後自身のもつ少数のデータでモデルをファインチューニングするという転移学習のテクニックが利用可能となる。公開データ利用によって、自身がもつデータが少ない場合でも良好な成果を挙げることが可能となる。CT像や眼底画像を対象とした研究開発では深層学習が活発に利用されているが、そこには公開アノテーション付きデータの充実が貢献している。The Lung Image Database Consortium (LIDC) は1,018例の胸部CT像、National Institute of Health (NIH) は32,000例のアノテーション付きCT像を公開している¹⁰⁾。このような膨大なデータは深層学習を用いた研究開発において非常に有用である。

コンピュータ外科においてもアノテーション付きデータを整備し、公開することで深層学習を利用する研究が活性化すると考えられる。そしてデータを整備することが研究者の増加と分野の発展を招く。CT像等を用いた診断支援においては深層学習を利用した成果が多数発表され、レッドオーシャンとなっている。これと比較してコンピュータ外科では深層学習の応用例が少なく、ブルーオーシャンであるといえる。そのため診断支援の研究者が新たな応用先を求めてコンピュータ外科分野に参入してくる可能性がある。そのとき有用な公開データが用意されていれば、研究が一気に活性化すると考えられる。

コンピュータ外科に関連した公開アノテーション付きデータはいくつか登場している。JIGSAWS¹¹⁾はda Vinciのアーム動作軌跡データやカメラ画像を含むデータである。国際会議MICCAIで実施されるEndoscopic Vision (EndoVis) Challengeでは、da Vinciで撮影したカメラ画像と画素単位アノテーション結果を配布して、セグメンテーション精度を競うイベントを行っていた (Robotic Scene Segmentation Sub-Challenge)¹²⁾。公開データを整備するとともに、このようなChallengeを実施してデータ利用を促進することは有効である。コンピュータ外科では対象ごとに様々なデータが利用される。公開アノテーション付きデータ作成においては、利用価値の高い対象を選定してデータ整備を行うこととなるだろう。

7. おわりに

コンピュータ外科において深層学習を活用することが研究の活性化に結びつくと考える。関連分野である診断支援における深層学習研究の盛り上がりと対比させながら、コンピュータ外科における深層学習利用の難しさについて述べた。コンピュータ外科においては深層学習を活用することで情報横断的な研究の誕生が期待される。様々な情報を利用した深層学習の応用例として、術中画像および位置データを用いた研究を紹介した。そして、深層学習利用にはアノテーション付きデータ収集がボトルネックとなることを説明し、研究推進のためには公開データ整備が必要であることを述べた。

文献

- 1) Roth HR, Oda H, Zhou X, et al. An application of cascaded 3D fully convolutional networks for medical image segmentation. *Comput Med Imaging Graph* 2018;66:90-99.
- 2) Shen C, Milletari F, Roth HR, et al. Multi-class abdominal organs segmentation with improved V-Nets. *Proc. SPIE Med Imag* 2019;10949:109490B-1-7.
- 3) Zhang Y, Chung ACS. Deep supervision with additional labels for retinal vessel segmentation task. *MICCAI* 2018;11071:83-91.
- 4) Wu Y, Xia Y, Song Y, et al. Multiscale network followed network model for retinal vessel segmentation. *MICCAI* 2018;11071:119-126.
- 5) Twinanda AP, Shehata S, Mutter D, et al. EndoNet: A deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans. Med. Imag* 2016;36(1):86-97.
- 6) Varytimidis C, Rapantzikos K, Loukas C, et al. Surgical video retrieval using deep neural networks. *M2CAI workshop in MICCAI* 2016
- 7) Aksamentov I, Twinanda AP, Mutter D, et al. Deep neural networks predict remaining surgery duration from cholecystectomy videos *MICCAI* 2017;10434:586-593.
- 8) Fawaz HI, Forestier G, Weber J, et al. Evaluating surgical skills from kinematic data using convolutional neural networks. *MICCAI* 2018;11073:214-221.
- 9) Oda M, Roth HR, Kitasaka T, et al. Colon shape estimation method for colonoscope tracking using recurrent neural networks. *MICCAI* 2018;11073:176-184.
- 10) Yan K, Wang X, Lu L, et al. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of Med Imag* 2018;5(3):036501.
- 11) Gao Y, Vedula SS, Reiley CE, et al. The JHU-ISI Gesture and Skill

Assessment Working Set (JIGSAWS): A surgical activity dataset for human motion modeling. M2CAI- MICCAI Workshop 2014.

- 12) EndoVis 2018, Robotic Scene Segmentation Sub-Challenge
<https://endovissub2018-roboticscenesegmentation.grand-challenge.org/> (2019 年 10 月 4 日確認)