



TITLE:

摘出臓器を対象とした単一カメラ 画像からの鉗子圧の推定

AUTHOR(S):

戸田, 輝; 中尾, 恵; 増井, 仁彦; 桑, 直人; 松田, 哲也

CITATION:

戸田, 輝 ...[et al]. 摘出臓器を対象とした単一カメラ画像からの鉗子圧の推定. 電子情報通信学会技術研究報告 (信学技報) 2021, 121(231): 7-12: MI2021-26.

ISSUE DATE:

2021-11

URL:

<http://hdl.handle.net/2433/265902>

RIGHT:

©2021 by IEICE

摘出臓器を対象とした単一カメラ画像からの鉗子圧の推定

戸田 輝[†] 中尾 恵[†] 増井 仁彦^{††} 桑 直人^{††} 松田 哲也[†]

[†] 京都大学大学院情報学研究科 〒606-8501 京都市左京区吉田本町

^{††} 京都大学医学研究科 〒606-8507 京都市左京区聖護院川原町 54

E-mail: †hika-toda@sys.i.kyoto-u.ac.jp

あらまし ロボット支援手術をはじめとする鏡視下手術では臓器に及ぼされる外力を正確には知ることはできず、また計測も困難である。本研究では、単一カメラ画像に基づく生体臓器に対する鉗子圧の推定を目指している。3軸力覚センサを軸内に組み込んだ鉗子を製作し、摘出臓器に対する押込操作時の鉗子圧とカメラ画像を取得した。複数操作を対象に計測された鉗子圧データと時系列カメラ画像について同期を取った学習用データベースを構築し、深層学習による学習と推定を行って鉗子圧の推定誤差を確認したので報告する。

キーワード 外力推定, 深層学習, コンピュータビジョン, ロボット支援手術

Force estimation in forceps manipulation of ex-vivo organs from a single-viewpoint camera image

Hlikaru TODA[†], Megumi NAKAO[†], Kimihiko MASUI^{††}, Naoto KUME^{††}, and Tetsuya MATSUDA[†]

[†] Graduate School of informatics, Kyoto University

Yoshida Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

^{††} Graduate School of Medicine, Kyoto University Hospital

Kawahara-chou, Shougoin, Sakyo-ku, Kyoto, 606-8507, Japan

E-mail: †hika-toda@sys.i.kyoto-u.ac.jp

Abstract In laparoscopic surgery including robotic surgery, it is not possible to accurately measure the contact force applied to organs. The purpose of this study is to estimate the forceps pressure applied to an organ based on a single-viewpoint camera image. Using forceps with a three-axis pressure sensor, the forceps pressure and camera images during the pushing operation to ex-vivo organs were acquired. Synchronized dataset of forceps pressures and time-series camera images were created for multiple operations, and deep-learning was applied to confirm the estimation error of forceps pressure.

Key words Force estimation, Deep learning, Computer vision, Robotic surgery

1. はじめに

近年、ロボット支援手術をはじめとする鏡視下手術が広く実施されている。特にロボット手術では関節を有することから、鏡視下に複雑な操作を行うことが可能であることや、マニピュレータに手振れ防止やスケーリングなどの機能があり微細な操作を行えることが挙げられる。一方、執刀医は術野を直接視認できず、視覚情報のみに基づいてすべての操作を行うため、臓器に対して加わっている力を認識することが難しい。臓器に直接接触するロボットアームの先端は消毒や交換が必要となるため、鉗子による操作によって発生する力の直接的な計測は難しい。

これまでに手術鉗子に力覚センサを取り付け、臓器に加わる力を直接測定する試み [1] がなされているが、上述の通り実用化には至っていない。臓器に及ぼされる力を直接的に計測するアプローチとは別に、臓器の弾性体モデルを構築し、手術時に発生する臓器の変形や応力の可視化を目指した研究 [2] [3] がなされている。また、観測された弾性体の一部の形状から外力や弾性率の推定を目指した研究 [4] [5] も報告されている。しかし、力学モデルには操作対象の弾性率分布に加えて境界条件の設定も必要であり、手術時における患者個人の生体臓器を対象とした推定には課題が残る。

近年では機械学習を臓器状態の推定に応用する試みがなされている。深層学習に基づく弾性体の変形推定 [6] [7] や、操作

対象を撮像したカメラ画像から外力を直接的に推定する枠組み [8] [9] [10] [11] が報告されている。時系列データに対応した深層学習モデルを用いて、操作対象の外観の変化から外力を推定する試み [12] もなされている。しかし、これら研究はシミュレーションモデルや人工ファントムを用いた検証がなされているに過ぎず、生体臓器を対象とした推定精度は確認されていない。生体臓器や手術時の操作を対象とした機械学習モデルの構築には、カメラ画像に加えて鉗子圧の計測が必要不可欠である。

本研究では、生体臓器に対する鉗子操作を対象として、単一カメラ画像のみから深層学習に基づいて鉗子圧を推定する枠組みの構築を目指している。本稿では、その最初の試みである動物の摘出臓器を対象とした計測実験と学習用データベースの構築に主眼をおく。3軸力覚センサを組み込んだ鉗子を製作し、摘出臓器に対する押込操作時の鉗子圧とカメラ画像を取得した。複数操作を対象に同期済みの学習データを作成し、深層学習による学習と推定を行って鉗子圧の推定誤差を確認したので報告する。

2. 摘出臓器を対象とした鉗子圧の計測

深層学習に基づいて単一カメラ画像と鉗子圧の関係を学習するためには、学習に用いる教師データが必要となる。本研究ではロボット支援手術において手術鉗子を臓器に接触させる状況を想定した実験環境を構築し、鉗子操作時における鉗子圧とカメラ画像の同時計測を行う実験を計画した。臓器操作時の鉗子圧を計測するために力覚センサを組み込んだ鉗子を製作し、動物の摘出臓器に対して押込操作を加えた際の変形を動画として撮影することとした。本節では、これら計測実験の詳細について説明する。

2.1 力覚センサ付き鉗子と計測環境

図 1(a) は製作した力覚センサ付き鉗子の先端周辺の外観を示している。鉗子の軸部分を一部切断し、切断部に株式会社テック技販製の3軸力覚センサを組み込むことによって鉗子操作時に軸内に生じる力を3自由度で計測可能とした。力覚センサの定格容量は $\pm 5\text{ N}$ であり、各軸方向の応力(F_x, F_y, F_z)を100Hzのリフレッシュレートで計測可能である。

本研究では、製作した力覚センサ付き鉗子によって計測が可能な臓器操作を考慮し、計測対象を臓器表面に対する押込操作とした。鉗子先端の把持部は閉じた状態で固定し、把持力の計測は対象としない。また、解析を容易なものとするため、推定対象の鉗子圧を鉗子の姿勢や軸回り回転に依存しない値とする。すなわち、式(1)に基づいて得られる3軸力覚センサの計測値のユークリッドノルムを鉗子圧 F と定義する。

$$F = \sqrt{F_x^2 + F_y^2 + F_z^2} \quad (1)$$

図 1 (b)(c) は、本実験において製作した鉗子圧の計測環境と動物の摘出臓器をそれぞれ示している。臓器には市販のブタの脾臓を採用し、プラスチック製のボックストレーナ内に設置した。臓器全体の位置ずれを防ぐために、画鋲によって脾臓の一部を固定した。ボックストレーナ上部に設置された

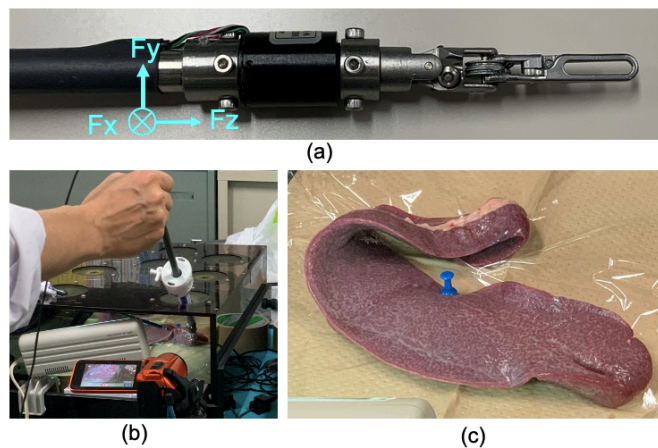


図 1 力覚センサ付き鉗子を用いた摘出臓器に対する押込操作の計測、(a) 力覚センサ付き鉗子の先端付近の外観、(b) 実験環境、(c) 摘出臓器 (脾臓)

孔から力覚センサ付き鉗子1本を挿入して腹腔鏡手術を模した鉗子操作を行うことが可能である。ボックストレーナと操作者の間にハンディカメラ1台と臓器表面を照明するための小型のライト1個を設置した。腹腔鏡手術に近い外観で押込操作時の鉗子先端の動きと脾臓表面の変形の様子が撮影できるようにカメラの方向と拡大率を調整した。

2.2 計測の流れ

本計測実験では、ロボット支援手術時において臓器を変形、圧排する状況を想定し、摘出臓器に対して類似の操作を行うこととした。想定される押込操作のパリエーションを反映した学習データとするために、設置した脾臓に対する押込操作を、押込の条件を変更しつつ複数回試行するものとした。計測の流れは以下のSTEP1から3の通りである。なお、すべての鉗子操作は鏡視下手術の執刀経験のある外科医により行われた。

- STEP1 計測開始直後にボックストレーナの床部を強く押し込むことによって、鉗子圧とカメラ映像の同期用データを得る。
- STEP2 摘出臓器に対する押込操作を繰り返し行い、STEP1と合わせて計60秒間鉗子圧とカメラ映像を計測する。
- STEP3 STEP1, STEP2による計測を臓器に対する鉗子の角度と押込箇所、押込の強さを変更しつつ、計6回分繰り返す。

各STEPにおける実験設定を説明する。本実験では、鉗子圧とカメラ映像はそれぞれ独立に計測がなされるため、カメラ画像から鉗子圧を推定する機械学習モデルの学習データとして利用するためには互いのデータの同期が必要である。STEP1では計測開始から約3秒の間に剛体に対する強い押込操作を行い、鉗子先端と剛体が接触状態から非接触状態へ移行する時刻をカメラ映像と鉗子圧データの同期時刻として用いることとした。本時刻はカメラ映像と鉗子圧データの両方において事後に同定可能であり、両データを厳密に同期可能と考えた。

STEP2はSTEP1の同期操作の後に計測を中断することな

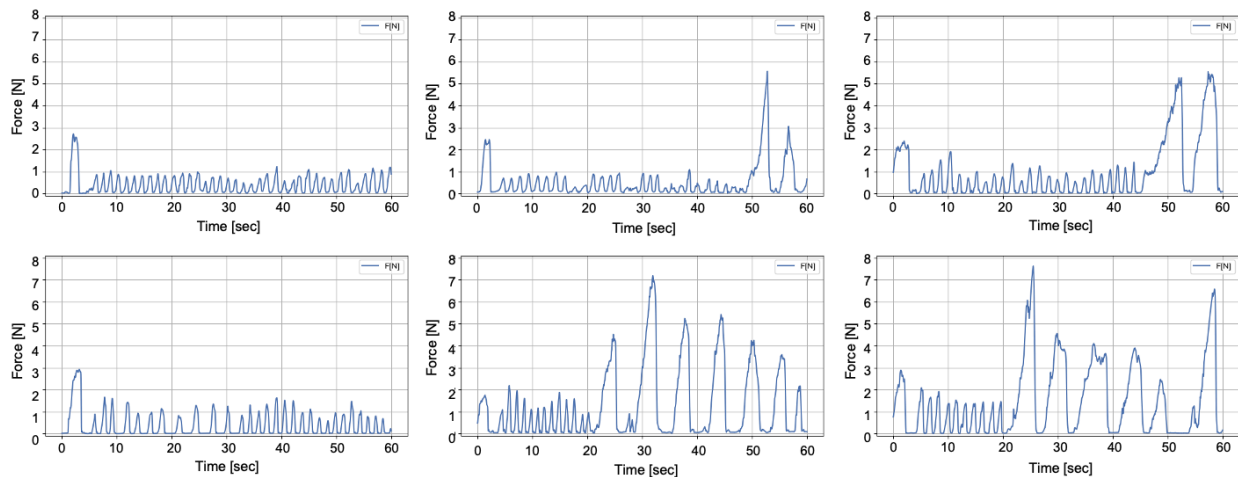


図2 各 Trial において計測された鉗子圧の推移, 上段: 左から順に Trial1, 2, 3, 下段: 左から順に Trial4, 5, 6

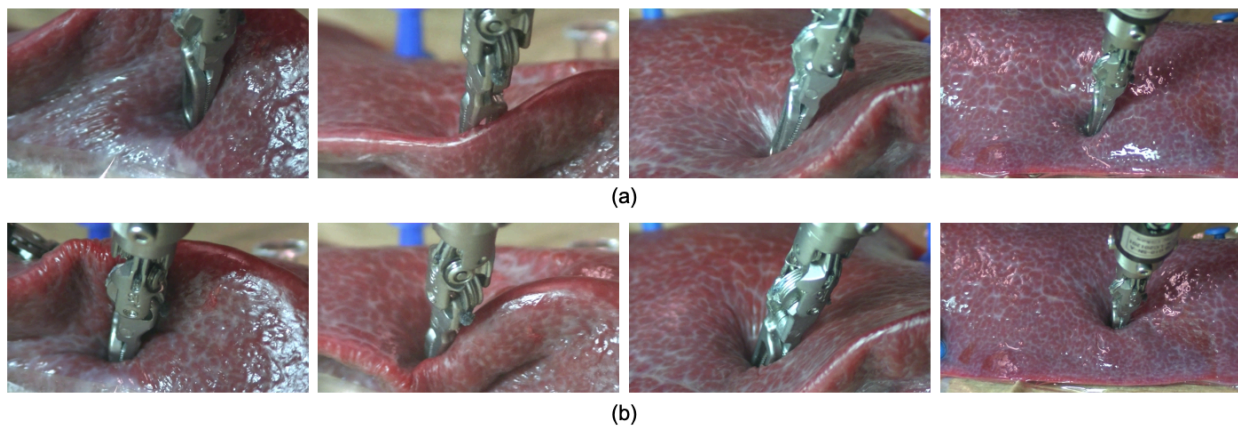


図3 臓器表面に対する押込操作, (a) 通常押込操作の例, (b) 強い押込操作の例

く引き続き実施された。ロボット支援手術において想定される臓器に対する押込の大きさを踏まえ、適切とみなせる力で摘出臓器に対する押込操作を STEP1 と合わせて時刻 60 秒まで反復して実施した。このとき、鉗子の角度及び押込方向はできるだけ一定に保ちつつ、押込量や押込速度については医師の判断にゆだねることとした。

手術時に想定される押込操作のバリエーションを獲得するために、押込操作の条件を変更しつつ STEP1, STEP2 からなる 60 秒間の計測を計 6 回 (以後、これらを Trial 1, 2, ..., 6 と呼ぶ) 行った。Trial ごとに摘出臓器の表面に対する鉗子の角度と押込箇所を変更した。また、一部の Trial では計測後半に手術時においては通常用いられない強い押込を摘出臓器に対して加えることとした。提案する推定モデルの開発動機の一つに、臓器に対する過度の押込を検出することがあり、異常とみなされる押込操作に関する計測データを得ることを目的に実施した。

2.3 計測データ

図 2 は各 Trial において得られた鉗子圧の推移を示している。計測開始直後に確認される大きな鉗子圧はカメラ映像との同期のために行なった剛体に対する強い押込み操作による

ものである。それ以降は、摘出臓器の表面に対する押込操作によって得られた鉗子圧であり、ピーク値は 1N 前後、一回の押込は 1-2 秒で繰り返し実施されていることが確認できる。Trial 2, 3, 5, 6 では、後半で摘出臓器に対して与えられた手術時においては通常用いられない程度の強い押込操作に対して得られた鉗子圧が記録されている。なお、強い押込によって破断したり表面の構造が局所的に変化している可能性があるが、各 Trial で摘出臓器に対する鉗子の押込位置を変更しており、全体の変形特性への影響は微小とみなせる。

図 3 はハンディカメラで撮影された摘出臓器に対する一連の押込操作の様子を例示している。計測中、ハンディカメラの位置は固定された状態で、焦点を押込箇所周辺に合わせて撮像されている。動画のサイズは 1920 × 1080 pixel であり、フレームレート 30 frame per second (FPS) であるため、各 Trial の全フレーム数は 60 秒間で 1800 である。図 3(a) は手術時に適切とみなせる力で行われた 1N 前後の押込操作の様子であり、力覚センサ付き鉗子によって摘出臓器の表面が押し込まれ、それに伴って変形が生じている様子が確認される。図 3(b) は手術時において通常用いられない強い押込を加えた場合の様子である。(a) と比較して、臓器表面がより大きく変

形していることが確認される。

2.4 鉗子圧とカメラ画像の同期

本節では、独立に計測された鉗子圧データとカメラ映像に対して実施した同期処理について説明する。各 Trial の STEP1 において取得した鉗子圧データとカメラ映像を実験者が確認し、対応する時刻の同定が容易なタイミングとして、鉗子先端が押込対象から離れる時刻を採用した。この時刻は鉗子圧は押込操作後に初めてゼロに近い値を取り、カメラ映像においても視覚的に確認可能である。動画内における時刻の指定は、Adobe Premiere Pro を用いて、取得した実験動画を 1 フレーム単位で再生しつつ厳密に行った。

次に STEP1 における同期処理用の操作に関するデータを学習対象から除外した。除外対象は、各 Trial において時刻 0 から 3.6 秒間に相当する 108 frame とした。鉗子圧データについてはリフレッシュレートをカメラ映像 30FPS に合わせてリサンプリングを行った。その結果、各 Trial において 1692 frame 分の画像と対応する鉗子圧の値を取得し、学習データセットとした。全ての Trial 分を合わせて 10152 frame のデータとなった。

本データベースの作成後、同期が正確に行えているかどうか、各 Trial の後半部分における 1) 押込操作の開始時刻、2) 押込のピークとなる時刻、3) 押込の終了時刻に着目し、鉗子圧データとカメラ画像がフレーム単位で一致していることを確認した。図 4 は Trial 2 における同期済みのデータのうち時刻 20 秒から 23 秒の鉗子圧データと、4 つの時刻に対応するカメラ画像を示している。A は鉗子が摘出臓器と接触し始める時刻、C は摘出臓器から離れる時刻であり対応する鉗子圧も小さい。B, D は、カメラ映像において周辺時刻で最も臓器変形が大きく変形しているタイミングであり、鉗子圧も同様に大きくなっている。鉗子圧データとカメラ画像間正しく対応が取れていることが確認できる。

3. 鉗子圧推定のための深層学習モデル

本節ではカメラ画像から鉗子圧を推定する際に用いる深層学習モデルについて説明する。ベースラインとなる推定性能を確認するため、深層学習モデルには単一の 2 次元カメラ画像 I を入力画像として受け付け、3 次元応力ベクトル (F_x, F_y, F_z) を出力する単純な畳み込みニューラルネットワーク (CNN) による回帰モデルを採用した。本 CNN は Convolution, Batch Normalization, Max pooling からなる畳み込みレイヤーと全結合レイヤーの計 13 層から構成される。畳み込みレイヤーには過学習を防ぐための Dropout を組み込むこととした。

本研究では得られた力ベクトルを式 (1) に基づいて鉗子圧のスカラー値に変換し、予測値とする。損失関数 \mathcal{L} には式 (2) によって定義される平均二乗誤差 (Mean Squared Error, MSE) を用いてネットワークの最適化を行う。

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n (F_i - \hat{F}_i)^2 \quad (2)$$

ここで、 F_i は真値となる鉗子圧の計測値であり、 \hat{F}_i は予測

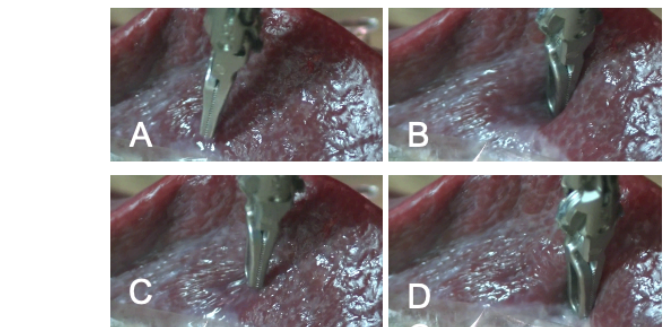
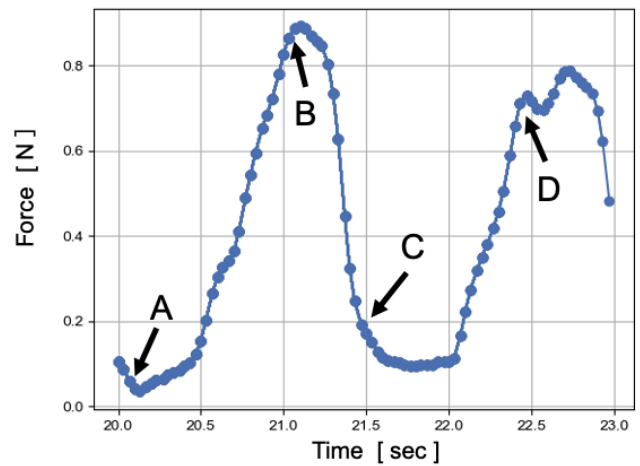


図 4 鉗子圧とカメラ画像の同期, 上段: 鉗子圧データ, 下段: 対応する時刻におけるカメラ画像

値、 n はトレーニングデータ数である。なお、入力画像内において予測に用いられた領域や画像特徴の視覚的な理解を可能とするために、Grad-CAM [13] の考え方に基づいて畳み込みレイヤー最下層の出力に関する勾配を求め、入力画像へ勾配マップを重畳可視化することとした。

上記 CNN は Python 3.9, Tensorflow 2.6 をバックエンドとした Keras ライブラリを用いて実装した。バッチサイズは 60, トレーニング epoch は 1000 とし、ネットワークの最適化には学習率を 1×10^{-4} とした Adam を用いた。入力画像のサイズは 320×180 とした。

4. 鉗子圧の推定実験

構築した鉗子圧-カメラ画像データを用い、前節に記述した CNN による鉗子圧の推定誤差を確認する実験を行った。データセットは各 Trial 1692 frame の前半 6 分の 1 に相当する 282 frame をテストデータとし、後半 1410 frame をトレーニングデータとして分割後、すべての Trial のデータを混合して、学習と推定に用いることとした。すなわち、トレーニングデータは 8460 frame, テストデータは 1692 frame とした。学習の際は、画像サイズを 1920×1080 pixel から、 320×180 pixel にリサイズ後にネットワークに入力した。また、過学習の有無の確認を行うため、訓練データの 10% を Validation data として学習を行った。本学習は CPU: Intel Core-i9, Memory: 64GB, GPU: NVidia TITAN RTX を搭載した PC 上で実施し、1000epoch のトレーニングに要した時間は約 158 分で

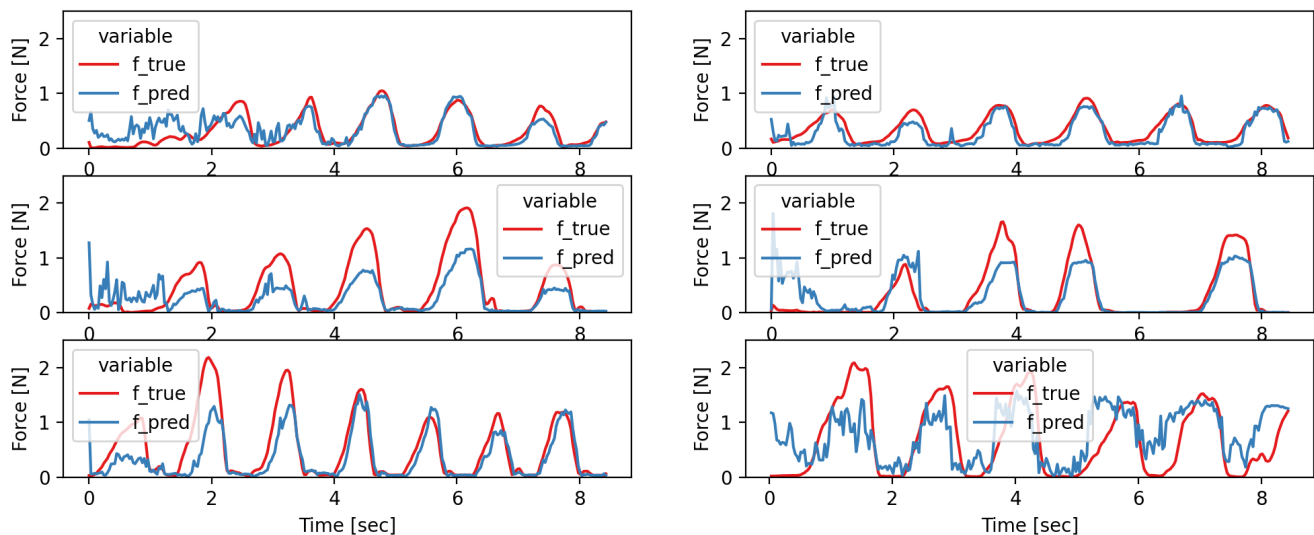


図5 鉗子圧の真値と予測値の推移, 赤線: 真値, 青線: 予測値, 上段: Trial1, 2, 中段: Trial3, 4, 下段: Trial5, 6

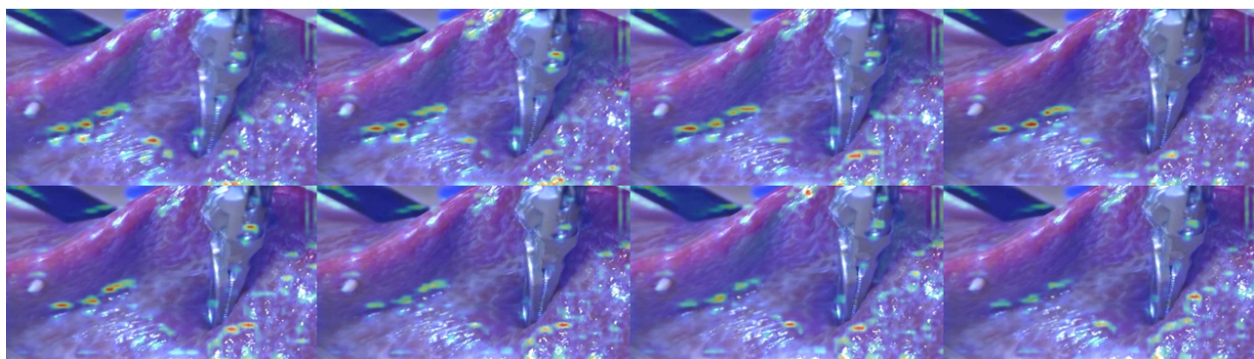


図6 Grad-CAMによる注目領域の重畳可視化

あった。

各 Trial のテストデータに対して求めた鉗子圧の予測値と真値の推移を図5に示す。図からいずれの Trial においても押込操作時の鉗子圧のピーク時刻に合わせて大きな推定値となっており、特に Trial1, 2 においては真値と予測値との間に良好な一致が見られた。一方、Trial 3, 4, 5, 6 における 1N を超える鉗子圧での押込に対しては予測値との間にずれが大きくなった。

全テストデータを対象に式 (3) によって定義される MAE (Mean Absolute Error) を推定精度として算出し、0.12N であることを確認した。

$$MAE = \frac{1}{m} \sum_{i=1}^m |F_i - \hat{F}_i| \quad (3)$$

ここで、 m はテストデータ数である。

最後に CNN が注目する入力画像内の領域を確認するため、予測誤差が小さかった Trial 1 の 5.5-6.5 秒 付近のフレームを対象に Grad-CAM [13] による可視化を行った。図6は算出された勾配の大きさをカラーマップに変換し、入力画像に重畳した結果である。一連の画像から、勾配は鉗子先端の金属部のみで大きな値を取っておらず、鉗子先端周辺の一部、臓器表

面の高輝度領域や臓器形状が大きく変化する領域などに広く分布していることが確認される。以上から、CNN は臓器表面の特徴を広範に捉えて鉗子圧を推定していると考えられる。

5. おわりに

本研究では、単一カメラ画像に基づく鉗子圧の推定性能を確認することを目的に、3軸力覚センサを軸内に組み込んだ鉗子を製作し、摘出臓器に対する複数の押込操作を対象とした計測を行い、鉗子圧データとカメラ映像の学習用データベースを構築した。CNN に基づく回帰によって鉗子圧を 0.12N の誤差で推定可能であることを確認した。手術鉗子の検出や時間依存関係を考慮した学習により、さらなる精度向上を見込むことができる。今後は、他の摘出臓器への適用や、in-vivo での計測など実際の手術に近い環境を対象とした計測と推定実験を予定している。

謝辞

本研究は日本学術振興会 科研費補助金 基盤研究 (B) 19H04484 及び 基盤研究 (C) 21K12724 の支援による。

文献

- [1] C. R. Wottawa, B. Genovese, B. N. Nowroozi, S. D. Hart, J.

- W. Bisley, W. S. Grundfest et al., "Evaluating tactile feedback in robotic surgery for potential clinical application using an animal model," *Surgical endoscopy*, vol. 30, no. 8, pp. 3198-3209, 2016.
- [2] M. Nakao and K. Minato, "Physics-based Interactive Volume Manipulation for Sharing Surgical Process", *IEEE Trans. on Information Technology in Biomedicine*, Vol.14, No. 3, pp. 809-816, 2010.
- [3] M. Tonutti, G. Gras, and G.-Z. Yang, "A machine learning approach for real-time modelling of tissue deformation in image-guided neurosurgery," *Artificial Intelligence in Medicine*, vol. 80, pp. 39-47, 2017.
- [4] R. Sakata, M. Nakao, T. Matsuda, "Estimation of external forces on the basis of local displacement Observations of an Elastic Body", *Advanced Biomedical Engineering*, Vol. 6, pp. 21-27, 2017.
- [5] M. Morita, M. Nakao, and T. Matsuda, "Elastic modulus estimation based on local displacement observation of elastic body," *Proc. IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2138-2141, 2017.
- [6] K. Morooka, T. Taguchi, X. Chen, R. Kurazume, M. Hashizume, T. Hasegawa, "A method for constructing real-time FEM-based simulator of stomach behavior with large-scale deformation by neural networks", In *Proc. SPIE*, Vol. 8316, 83160J, 2012.
- [7] U. Yamamoto, M. Nakao, M. Ohzeki, T. Matsuda, "Deformation estimation of an elastic object by partial observation using a neural network", *arXiv*, abs/1711.10157, 2017.
- [8] A. I. Aviles, S. M. Alsaleh, J. K. Hahn and A. Casals, "Towards retrieving force feedback in robotic-assisted surgery: A supervised neuro-recurrent-vision approach," *IEEE Transactions on Haptics*, vol. 10, no. 3, pp. 431-443, 2017
- [9] A. Marban, V. Srinivasan, W. Samek, J. Fernandez, and A. Casals, "Estimation of interaction forces in robotic surgery using a semi-supervised deep neural network model," *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 761-768, 2018.
- [10] A. Abiri, Y. Y. Juo, A. Tao, S. J. Askari, J. Pensa, J. W. Bisley et al., "Artificial palpation in robotic surgery using haptic feedback," *Surgical Endoscopy*, vol. 33, no. 4, pp. 1252-1259, 2019.
- [11] Cong Gao., et al. "Learning to See Forces: Surgical Force Prediction with RGB-Point Cloud Temporal Convolutional Networks," *arXiv:1808.00057*, 2018
- [12] H. Shin, H. Cho, D. Kim, D. K. Ko, S. C. Lim, and W. Hwang, "Sequential Image-Based Attention Network for Inferring Force Estimation Without Haptic Sensor," *IEEE Access*, vol. 7, pp. 150237-150246, 2019.
- [13] Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *The IEEE International Conference on Computer Vision (ICCV)*, pp. 618-626, 2017.