

ロボット手術映像からの転移学習による顕微鏡縫合訓練における振り返り支援手法の提案

佐藤 響* 清藤 哲史† 宮藤 詩緒*

概要. 脳神経外科の顕微鏡縫合は熟達まで時間を要する一方、指導者による評価は主観的でばらつきがあり即時性に欠け、学習効率を損なっている。学習効率向上に向け客観的フィードバックを行うモデルが検討されているが、医療では大規模なラベル付き動画の収集が難しく、従来の教師あり学習は高精度化が難しい。本研究はこの制約に対し、少数データでも機能する振り返り支援を目的とし、転移学習を用いた時系列フィードバックを提示する。具体的には、ロボット手術映像(JIGSAWS)で学習した表現を顕微鏡縫合へ転移し、動画全体の熟練度推定に加えてスライディングウィンドウで作成したクリップに基づく時系列ヒートマップを提示する。医師レビューでは一部でフィードバック判断の一致が見られた一方で、非動作区間の誤評価や微細運動(針先角度・糸テンション等)への非感度が指摘され、現状の有効性は限定的であることが示唆された。これを踏まえ、フェーズ分類(針通し・結紮等)との連携、非動作区間の分離、および注視すべき局所領域の強調を進め、要改善箇所の提示に直結するUIを目指す。

1 はじめに

脳神経外科医の技能訓練、特に顕微鏡縫合は高い精度を要求するため数万針という相当な訓練が必要である[9]。一方、学習者が受け取る評価は指導者の主観性に左右されやすく、指導者への負担も大きい。これに対し、個人訓練を支援するための各種訓練支援手法が提案されてきた[1, 2, 4, 6, 7]。なかでも、スコアリングシステムや習熟度推定による訓練評価は、従来学習者個人の主観的評価しかなかった個人訓練に対し、客観的な評価基準を導入できる点で有用である[1, 5, 6]。ただし、総合点のみの提示では「どこを直すべきか」が分かりにくいという課題があり、学習者の改善点に直結する時系列フィードバックが求められる。

しかし、医療現場では大規模にラベル付けされた動画データが乏しいため、汎用的なモデルをそのまま適用するのは難しい。このため、少数データでも安定に機能し、かつ具体的なフィードバックを提供できる枠組みが必要である。そこで本研究では、ロボット手術映像からの転移学習を用い、顕微鏡縫合に対して動画全体の熟練度推定に加え時系列ヒートマップで良否区間を可視化する振り返り支援を提示する。

具体的には、公開されているロボット手術データセット JIGSAWS [3] で動画表現を事前学習し、顕微鏡縫合タスクへ転移させる。推論では動画をクリッ

プに分割し、単調校正により出力の分布とスケールを整えたうえで動画全体の熟練度を推定する。併せて、クリップ列の時系列スコアをヒートマップとして提示し、良否区間を可視化する。

2 提案手法

2.1 概要

本研究は、JIGSAWSを用いた事前学習と、対象タスク(顕微鏡縫合)に対する出力校正および時系列解析を統合した熟練度評価パイプラインを提案する。まず、JIGSAWSでCNNモデルを事前学習する。次に、顕微鏡縫合データを入力し、得られたスコアをアイソトニック回帰で校正し、閾値最適化を行って3クラスに分類する。推論時には動画全体のスコア分類、およびスライディングウィンドウで動画をクリップ化し、予測結果を時系列スコアのヒートマップとして可視化する。

2.2 事前学習ステージ

特徴抽出には $R(2+1)D-18$ モデル [8] を採用する。JIGSAWS データセットを用い、(1) クリップ単位のジェスチャ分類、(2) 動画単位のスコア回帰、(3) クリップの貢献度予測、の3つのタスクによるマルチタスク学習を行う。

2.3 校正 (Calibration) と離散化

転移学習先の顕微鏡縫合データにラベルスケール差や分布ずれを合わせるため、事前学習モデルの連続出力をアイソトニック回帰で単調校正する。これは、ドメイン間のスケール差を吸収しつつ、スコアの順序性を保持するためである。そして校正スコア

に対し、QWK(Quadratic Weighted Kappa)を最大化するように2つの閾値(t_0, t_1)を探索し、3クラス(初心者, 中級者, 上級者)に離散化する。

2.4 スコアの時系列可視化

訓練中の「うまくできている部分」と「改善すべき部分」を時間軸上で明示するため、動画全体にわたりスライディングウィンドウでクリップを生成し、各クリップの予測スコアを並べて時系列ヒートマップとして提示する。

3 実験

3.1 データセット

事前学習にはロボット手術動画データセット JIGSAWS (Suturing, Knot-Tying タスク, 計 75 本) を使用した。評価には、ガーゼ繊維を結ぶ顕微鏡縫合訓練タスクの映像 (14 名, 計 24 本) を使用した。ラベルは経験年数に基づき 3 クラス (初心者: 3 年未満, 中級者: 8 年未満, 上級者: 8 年以上) に分類されている。

3.2 動画の熟練度評価と精度検証

評価は動画単位で行う。各動画から均等に抽出した 8 クリップの校正スコアの平均を用い、最適化された閾値でクラス分類する。評価は、被験者一名を検証用とし、残り被験者で校正と閾値最適化を行う LOUO (Leave-One-User-Out) 交差検証で行った。

3.3 時系列ヒートマップ表示

時系列ヒートマップの妥当性と教育的有用性を検討するため、熟練医師によるレビューを実施した。

4 結果

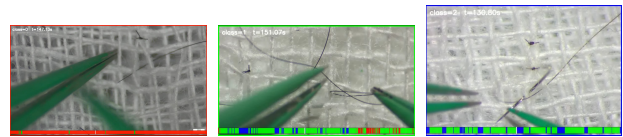
4.1 動画単位の推定結果 (Accuracy/QWK)

提案手法は正解率 (Accuracy) = 0.75, QWK (Quadratic Weighted Kappa) = 0.679 を達成した。これは、事前学習なしの顕微鏡縫合映像のみを用いて R(2+1)D-18 モデルを学習させた Accuracy = 0.583, QWK = 0.536 を上回る結果であった。

4.2 時系列ヒートマップと医師レビュー

時系列ヒートマップ (図 1) は、動画全体のおおまかな傾向 (例: 初心者例で赤が多い) はラベルと整合したが、上級者の動画でも青 (上級) が限定的で緑 (中級) が優勢となる場面が多く見られた。

熟練医師によるレビューでは、一部で判断が一致したものの多くでズレが報告され、現状の有効性は限定的であるという結果になった。具体的には、(i) 器具が映っていない非動作区間でも緑 (中級) と評価される、(ii) 医師が重視する微細運動 (針先の角度、糸のテンション等) への追従が不十分、といった課題が指摘された。



(a) 初心者映像 (b) 中級者映像 (c) 上級者映像

図 1. ヒートマップ例 (赤: 初心者, 緑: 中級者, 青: 上級者)



図 2. 振り返り UI イメージ

5 考察と今後の展望

JIGSAWS を用いた事前学習が有効であった理由として、ロボット手術と顕微鏡縫合はドメインこそ異なるものの、「縫合」というタスクに共通する器具操作や糸の取り回しが存在するため、転移に有効な時空間特徴量を事前学習できたと考えられる。その上で、転移先でのスコアのアイソトニック回帰による単調性の維持とスケール整合、さらに QWK 最大化による閾値の整合化によりラベル境界の解釈が安定化したことが有効に機能したためと解釈できる。

一方、時系列評価では医師評価との乖離が残った。これは、(1) 技能と無関係な視覚的变化 (位置直し) の誤認識、(2) 技能判断に本質的な微細運動が捉えられていない、(3) 非動作区間の扱いが未分離、という問題に起因すると考えられる。現状の可視化は教育支援の手掛かりとしては課題が残る、今後は医師の注視点や非動作区間を考慮したモデル改良が必要である。

そこで今後は、フェーズ分類を前段に導入し、その結果を用いて非動作区間や無関係区間を検出・除外するとともに、フェーズごとの学習・校正によって文脈に依存した基準でスコアを整合させる。併せて、特定領域 (ROI) への注目 (例: 器具・針先・結紮部位) を強化して微細な運動・形状特徴を捉えやすくする。UI は現行の全体時系列ヒートマップ中心から、図 2 に示すようにフェーズごとに良し悪しを提示する表示へと移行し、どこを直すべきかをより具体的に提示する。

謝辞

本研究は、JST さきがけ JPMJPR23I9 の支援を受けたものです。

参考文献

- [1] A. De Mauro, J. Raczowsky, M. E. Halatsch, Heinz W. Horn. Mixed Reality Neurosurgical Microscope for Training and Intra-operative Purposes. In R. Shumaker ed., *Virtual and Mixed Reality*, pp. 542–549, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- [2] S. Eom, S. Kim, S. Rahimpour, and M. Gorlatova. AR-Assisted Surgical Guidance System for Ventriculostomy. 02 2022.
- [3] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh, et al. Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling. In *MICCAI workshop: M2cai*, Vol. 3, p. 3, 2014.
- [4] Y. Kojima, S. Miyafuji, Y. Tashiro, S. Kiyofuji, and H. Koike. MR MANE: MR Microsurgical Suturing Skill Acquisition for Novice Using Imitation of Example. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces*, AVI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [5] R. Sharma and A. Suri. Microsurgical suturing assessment scores: a systematic review. *Neurosurgical Review*, 45(1):119–124, 2022.
- [6] Y. Tashiro, S. Miyafuji, D.-H. Hwang, S. Kiyofuji, T. Kin, T. Igarashi, and H. Koike. GAuZe-Microsuture-FICATION: Gamification in Microsuture training with real-time feedback. In *Proceedings of the Augmented Humans International Conference 2023*, AHs '23, p. 15–26, New York, NY, USA, 2023. Association for Computing Machinery.
- [7] Y. Tashiro, S. Miyafuji, Y. Kojima, S. Kiyofuji, T. Kin, T. Igarashi, and H. Koike. MR Microsurgical Suture Training System with Level-Appropriate Support. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [8] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri. A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 6450–6459, 2018.
- [9] 井上智弘, 國井尚人, 熊切敦, 大谷亮平, 田村晃, 齋藤勇, 堤一生. 脳卒中外科手術技量の継承における卓上型マイクロによる縫合練習の役割—8 万針の効果—. 脳卒中の外科, 37(4):247–252, 2009.