

オープンソース AI ツールの手術動画解析への適用

Application of open source AI tools to surgical video analysis

山下 一郎^{1*} 山田 敏哉¹ 宮崎 淳^{1,2}
西田 健次³ 村上 貴志⁴

¹ (株) オレンジテクラボ

¹ Orange Tech Lab Inc.

² 東京国際工科専門職大学

² International Professional University of Technology in Tokyo

³ 東京工業大学 工学院 システム制御系

³ Dept. of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology

⁴ 東京歯科大学市川総合病院 心臓血管外科

⁴ Tokyo Dental College Ichikawa General Hospital

Abstract:

模擬手術環境における看護師の技量評価を自動化するための AI ツールの適用可能性の検討、および、評価を行った。具体的には、器械出し看護師から医師への手術器材の手渡しの自動検出を試みた。器材の手渡しは、手術中の看護師と医師の協調作業の重要な側面を捉えるものである。本研究では、看護師と医師の手首間距離の分析や発声認識を通じて、手渡しタイミングの適切さを判断する新しい手法を開発した。現行の技術では誤検出や録音の質の問題があったが、これらは将来の改善のための重要な指針を提供する。また、視線検出に代えて顔の向き推定を行うことで、手術室における看護師の行動パターンを把握するための基礎的なデータを得ることができた。本研究の成果は、医療現場における AI の活用を進め、手術室での看護師の作業を効率化し、全体的な医療の質を向上させる重要な一歩となる。

1 はじめに

器械出し看護師は手術において重要な役割を果たしており、その技量を向上させるために種々の訓練手法が提案されてきた。

Skov らは器械出し看護師に対してシミュレーションに基づく教育を行うことで、ストレスを軽減し手術パフォーマンスが向上することを示しており [1]、Siah らはバーチャルリアリティシミュレーションを用いて、看護学生の有効性、態度、信頼度を評価した [2]。しかし、訓練の結果、看護師の技量が向上したか否かの評価に関しては、人間の判断に頼る部分が多いと言える。原らは初心者看護師に対するシミュレーションによる教育の効果を検証したが、技量の評価に関してはアンケートなどに基づいて行っており、客観的な評価指標を得てはいない [3]。また、Nasiri らは初心者看護師の器械出し技量を評価するためのチェックリストを開発し、その信頼性を評価した [4]。評価項目としての有効性は示し

ているが、アンケートに基づく評価指標となっており、人間の判断が基盤となっていると考えられる。Bracq らは熟練看護師と初心者看護師のバーチャル手術室におけるエラー認識の差を評価しているが、これは必ずしも看護師の技量を評価したものではなかった [5]。

一方、近年では深層学習などの AI 技術を用いて、手術室での動作解析を行う試みがなされてきており、それを活用することにより、看護師の技量評価を行える可能性もある。Funke らは三次元 CNN を用いて時空間特徴を学習することにより、手術における動作認識を行うことを目指した [6]。また、岸らは CNN を用いて時間的姿勢特徴を学習することにより、手術手技認識を行うことを目指し [7]。模擬手術に対する動作認識において、一部の動作に対しては 90% 程度の認識精度が得られたが、平均的な精度は 75% 程度であり、動作ごとの認識精度の差が大きいことが示されていた。Goldbraikh らは時間方向の畳み込みネットワーク (temporal Convolutional network) の一種である MS-TCN++ による手術動作認識を目指した [8]。Khalid らは深層学習による手術動作

*ichiro.yamashita@orange-tech-lab.com

の認識とパフォーマンス測定手法を評価した [9]. 北口らは深層学習を用いた実時間での手術の段階認識手法を提案し [10], 腹腔鏡下 S 字結腸切除術において, 高い認識率を示した. Menegozzo らは時間遅れニューラルネットによる運動データに基づく手術動作認識を提案した [11]. しかし, これらの研究は, 主に, 術者の動作を認識することを目的としており, 看護師の技量を評価するためのものとはなっていない.

また, 器材の検出, 追跡を行うことで, 器材の手渡し時のタイミングを検出することができるが, Bajraktari らの CNN を用いた機器検出による手術補助システムでは, ベンチマークセットに対しては良好な識別率を示しているが, 実際に用いられる手術器具の種類をカバーしきれてはいない [12].

看護師, および, 術者の厳密な動作認識, 器材の追跡などを行えば, 器材の手渡しタイミングなどの検出が可能になり, 看護師動作の適否の評価が可能になる. しかし, 現実には, 動作認識のための学習サンプルが不足しており, 汎用性のある認識評価システムを構築することは困難であると考えられる. そこで, 看護師の技量評価を行うための学習サンプルが不足していることを前提に OpenPose[13], YoloV5[14], Whisper[15] などの AI ツールを活用して, 簡便に得られる特徴量を元に, 看護師の技量評価が可能か否かを検証することとした [16].

2 器械出し看護師の技量評価自動化への特徴量抽出実験

器械出し看護師の技量は, 看護師から術者に対して適切なタイミングで適切な器材が渡されているか否かで評価できると考えられる. 器材の手渡しの検出, 追跡を行えば, 手渡しのタイミングを検出することは可能となるが, 現状では高い精度での手渡しタイミングを検出することは難しく, また, タイミングの適切さの評価基準は定まっていない. そこで, 現状の AI ツールによって, 看護師の技量評価に資する特徴量が抽出できるか否かを検証する実験を行った. 模擬手術の動画に対して OpenPose による看護師, および術者の姿勢解析, YoloV5 による器材の検出, Matlab の Speech detection[17] による音声区間切り出し, Whisper による音声認識を適用し, どの程度の認識が可能で, 実際に器材の手渡しが行われたタイミング検出への有効性の検証を行った. 本節では, 模擬手術動画の撮影環境, および, 動画に対する AI ツールの適用結果について述べる.

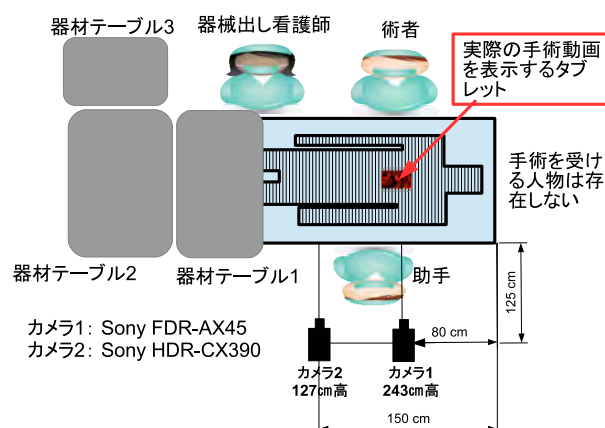


図 1: 実験環境



図 2: 取得データ例

2.1 実験環境

図 1 に, 模擬手術動画撮影の環境を示す. 模擬手術の内容は, 心臓手術における開胸から人工心肺の開始, 上行大動脈遮断までである. 実際の手術で撮影された術野の動画を手術台上に置かれたタブレットで再生しながら, 手術手順を追った. カメラ 1 は器材テーブルおよび看護師, 医師の手元を上から写すために, 高さ 243 cm に設置し, カメラ 2 は看護師の顔を写すために高さ 127 cm に設置した. 音声の録音は, ビデオカメラ付属のマイクを利用した. また, 看護師, 術者, 助手の三者とも, 今回は熟練者である. 取得された動画に OpenPose を適用した例を図 2 に示す.

2.2 器材の検出

YoloV5 による器材の検出の状況を図 3 に示す. 器材の学習には Kaggle の Labeled Surgical Tools and Images[18] を用いたが, 今回使用した器材に対応したデータセットではなかったため, 良好な性能を得ることはできなかった.

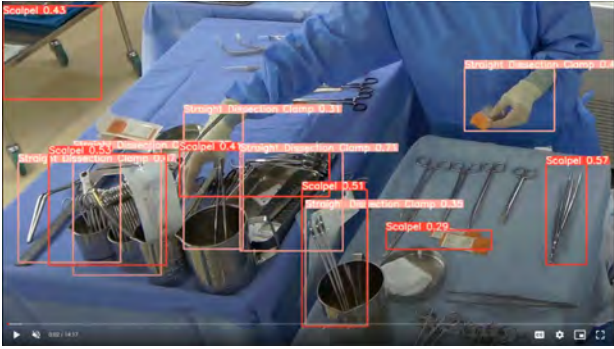


図 3: YoLoV5 による器材の検出

表 1: 手首間の距離による手渡し検出精度

看護師	術者	手渡し回数	検出回数	精度
右手	右手	22	35	62.9%
右手	左手	1	1	100%
左手	右手	7	64	11.0%
左手	左手	1	3	33.3%
トータル		31	103	30.1%

2.3 手渡しタイミングの検出

OpenPose による姿勢検出を元に、器材の手渡しタイミングの検出を試みた。看護師と術者の手首の位置を推定し、その距離が近づいた時を「機材手渡し」の瞬間とした場合の検出精度を検証した。実際に機材が手渡された瞬間を目視で確認し、その時刻をゼロとして、看護師と術者の手首の位置を図 4 に示す。術者が手を伸ばした後に看護師の手が近づいていくことが確認できる。

看護師と術者の手首間の距離が 500 画素よりも小さくなったと検出された回数は 103 であり、そのうち実際に機材が手渡しされていたのは 31 回であった。その内訳は、看護師の右手から術者の右手への手渡しは距離による検出 35 回に対して実際の手渡しは 22 回、看護師の右手から術者の左手は検出 1 回に対して手渡し 1 回、看護師の左手から術者の右手は検出 64 回に対して手渡し 7 回、看護師の左手から術者の左手は検出 3 回に対して手渡し 1 回となっている。看護師の左手と術者の右手の距離による誤検出が多いのは、両者の位置関係により、手渡しを行わない場合でも距離が近づいてしまうためであろうと考えられる (表 1)。

2.4 手首間距離最小と手渡しの時間差

手首間距離最小による手渡しの検出精度は高くないものの、手首間距離最小の瞬間と実際に手渡しが行われた瞬間の時間差によって、看護師の技量を推定でき

る可能性がある。そこで、実際に手渡しが確認された事例において、手首間距離最小と手渡しの時間差を検証した (図 4)。図 5 に手首間距離最小と手渡しの瞬間の時間差のヒストグラムを示した。時間差が負の値をとるのは、手渡しが行われるよりも前に手首間の距離が最小になっていることを示す。本実験では、手首間距離最小のほとんどが手渡しの前後 0.2 秒の間に入っていることが示された。

2.5 術者の発声と手渡しの時間差

術者の発声から手渡しまでの時間についても評価した。音声認識ツール Whisper では発生時刻が特定できないため、MATLAB の speech detection を適用し音声活動検出 (Voice Activity Detection: VAD) を行い、その結果を Whisper で解析してみたが満足のいく認識結果は得られなかった。しかし、器材の種類にかかわらず手渡しを検出するのであれば、VAD によって検出された発声時刻を手掛かりとすることはできる。図 6 に術者の発声と手渡しの時間差を示した。多くの手渡しは、術者の発声から 2 秒以内に行われており、4 秒以上遅れているものは少ないことが示された。

2.6 看護師の顔の向き

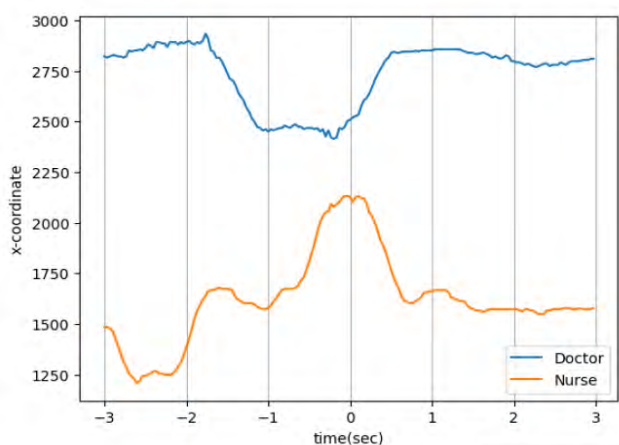
手術中に看護師が何処を (何を) 見ているのかは、看護師の技量を推定するのに重要な手掛かりとなると考えられている。しかし、手術中は保護メガネを着用しているため、視線の検出は困難である。そこで、顔の向きを検出することにより、看護師の視点推定に代えることとした。顔の向き検出に関しても、マスクにより口などは隠されており、手掛かりとなるのは目 (メガネ)、鼻、首であった。図 7 に、顔の向きの検出結果を示し、図 8 に、顔が向きの傾向を示す。図中、-150 から 0 は画面に向かって左のテーブルに向いており、0 から 100 は前の機材テーブル、100 から 200 は術野、あるいは、術者に向いていると考えられる。この結果、看護師は器材テーブル全体と術野 (あるいは術者) を、ほぼ半々に見ていることを示している。

2.7 手渡し時刻の判読が困難な例

今回の実験の録画からは手渡し時刻の判読が困難な例を幾つか見つけられた (図 9)。その原因は、以下に示す 5 つであった。

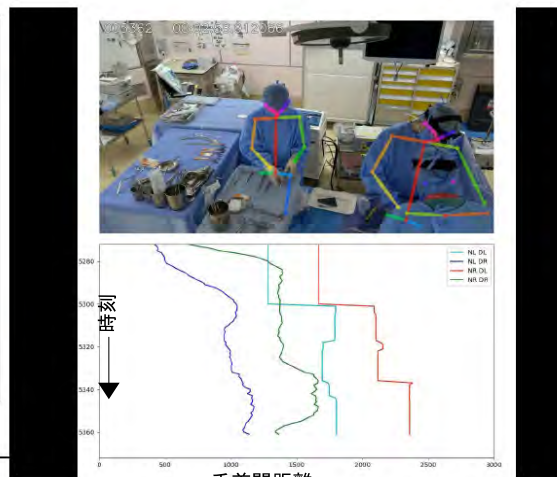
1. 術者が直前に別の動作を行う
2. 長い器材を渡す

縦軸
画面上の
手首のX
座標



横軸: 時間、手渡された瞬間を0

青: 術者の手首
橙: 看護師の手首



青: 看護師左手と術者左手間の距離
緑: 看護師左手と術者右手間の距離
赤: 看護師右手と術者左手間の距離
青緑: 看護師右手と術者右手間の距離

図 4: 術者が手を出してから渡されるまでの時間

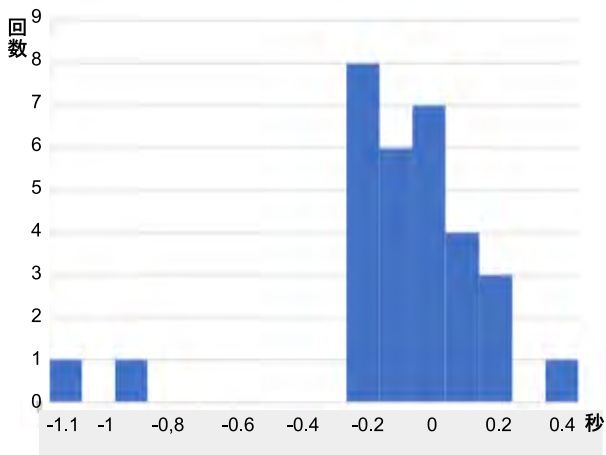


図 5: 手首間の距離最小と手渡しの時間差

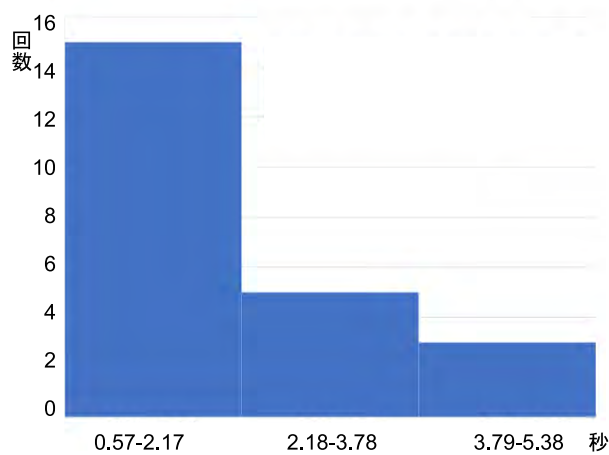


図 6: 術者の発声と手渡しの時間差

3. 看護師が先に手を伸ばす
4. 術者が手を伸ばさない
5. 術者が左手で受け取る

3 考察

看護師の技量評価を自動化するために AI ツールを用いて手術器具の手渡しを検出する試みを行った。現行の AI ツールの適用により、一部の課題は確認されたが、これらの課題は将来の研究の有望な方向性を示している。具体的には、看護師と医師の手首間の距離に

基づく検出法は、単一カメラの限界を露呈したが、これは複数カメラの統合によって克服できる可能性がある。また、誤検出はあったものの、検出漏れはなく、潜在的な手渡しの候補を絞り込むための有効なツールとして機能した。一方で、現状では、手術器材の学習サンプルが不足しているため、器材の検出追跡に関して十分な性能が得られないことが確認された。

術者の発声に基づく検出では、録音状態の問題から、Whisper による器材名の識別は出来なかった。これは、ビデオカメラ付属のマイクによる録音のため、マスクによる声のこもり、周囲の雑音の影響が大きかったと考えられる。術者にヘッドセットを装備してもらうことなどにより改善が期待される。また、器材名の識別

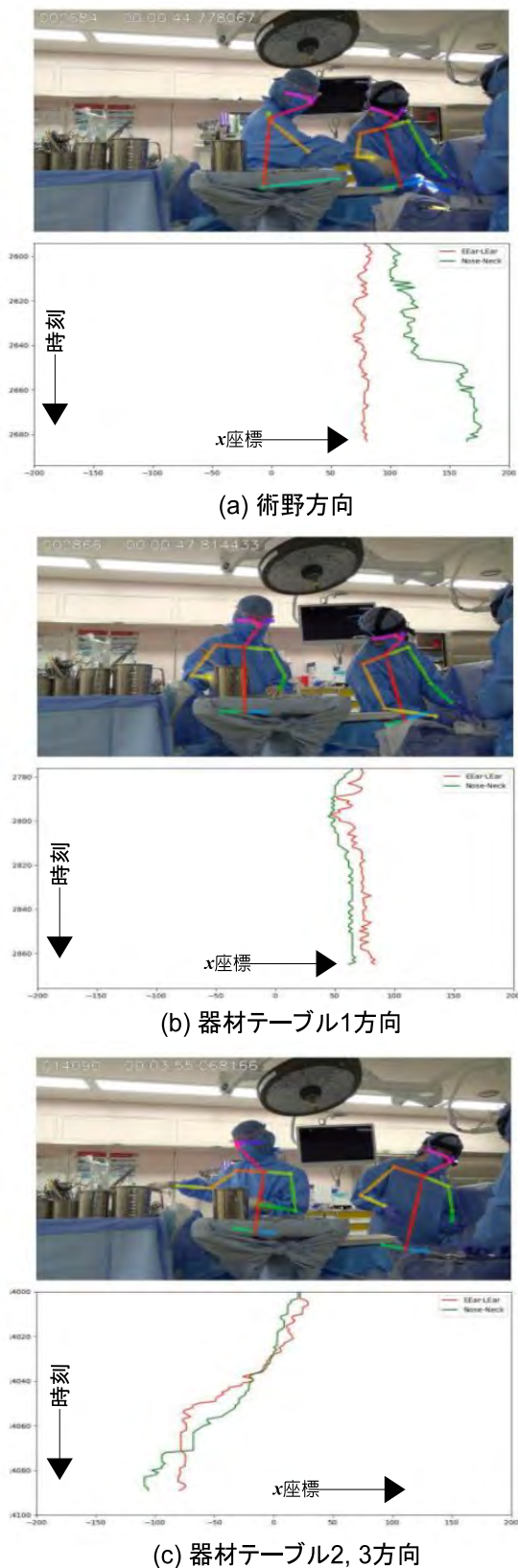


図 7: 看護師の顔の向き検出

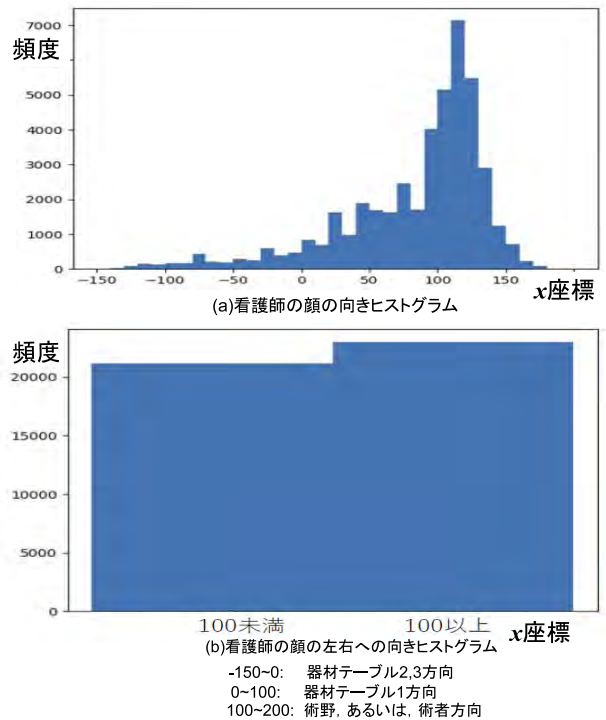


図 8: 看護師の顔の向きの傾向

は出来なくとも、術者の発声を捉えることで、手渡し検出、および、手渡しまでの時間推定が可能である。これを利用することで、手渡し検出精度の向上が期待されるとともに、看護師の技量評価の一助となると考えられる。

手術中の看護師が何処を見ているかも技量評価の大きなポイントとなるが、術中は保護メガネを着用しているため視線検出は困難である。そこで、視線検出に代え顔の向き推定を行った。目（保護メガネ）と鼻が検出できたため、大まかな顔の向き推定が可能であることが示された。これにより、看護師が何処を見ているかの推定が可能であることが示された。

4 結論

本研究では、模擬手術動画に AI ツールを適用し、看護師の技量評価の自動化の可能性を検証した。現在の技術では誤検出の問題が存在するが、これは今後の改善の余地を残している。特に、術者の要求と手渡しのタイミングの差を推定することで、看護師の技量評価の基準を設定する新しいアプローチが見出された。さらに、術者の発声を検出トリガとして利用することで、精度の向上が期待できる。今後の研究では、器材検出のための学習モデルの構築、複数カメラの視点統合、録音状態の改善など、さらなる精度向上に向けた取り組みが予定されている。これらの進展は、手術室での看

看護師の役割をより効果的にサポートし、全体的な医療の質を向上させる大きな一歩となると考えられる。

今後は、器材検出のための学習モデルの構築、複数カメラ視点の統合、録音状態の改善など、精度向上のための検討を行っていきたい。

参考文献

- [1] Rebecca Andrea Conradsen Skov, Jonathan Lawaetz, Lars Konge, Lise Westerlin, Eske Kvaner Aasvang, Christian Sylvest Meyhoff, Katja Vogt, Tomas Ohrlander, Timothy Andrew Resch, and Jonas Peter Eiberg. Simulation-based education of endovascular scrub nurses reduces stress and improves team performance. *Journal of Surgical Research*, Vol. 280, pp. 209–217, 2022.
- [2] Rosalind CJ Siah, Ping Xu, Cheang L Teh, and Alfred WC Kow. Evaluation of nursing students' efficacy, attitude, and confidence level in a peri-operative setting using virtual-reality simulation. In *Nursing Forum*, Vol. 57, pp. 1249–1257. Wiley Online Library, 2022.
- [3] Kentaro Hara, Tamotsu Kuroki, Masashi Fukuda, Toru Onita, Hiromi Kuroda, Emi Matsuura, and Terumitsu Sawai. Effects of simulation-based scrub nurse education for novice nurses in the operating room: A longitudinal study. *Clinical Simulation in Nursing*, Vol. 62, pp. 12–19, 2022.
- [4] Morteza Nasiri, Shahrzad Yektatalab, Marzieh Momennasab, and Fatemeh Vizeshtar. Development and assessment of validity and reliability of a checklist to evaluate the circulating and scrub skills of operating room novices (cssorn checklist). *Journal of Education and Health Promotion*, Vol. 12, , 2023.
- [5] Marie-Stéphanie Bracq, Estelle Michinov, Marie Le Duff, Bruno Arnaldi, Valérie Gouranton, and Pierre Jannin. Training situational awareness for scrub nurses: Error recognition in a virtual operating room. *Nurse education in practice*, Vol. 53, p. 103056, 2021.
- [6] Isabel Funke, Sebastian Bodenstedt, Florian Oehme, Felix von Bechtolsheim, Jürgen Weitz, and Stefanie Speidel. Using 3d convolutional neural networks to learn spatiotemporal features for

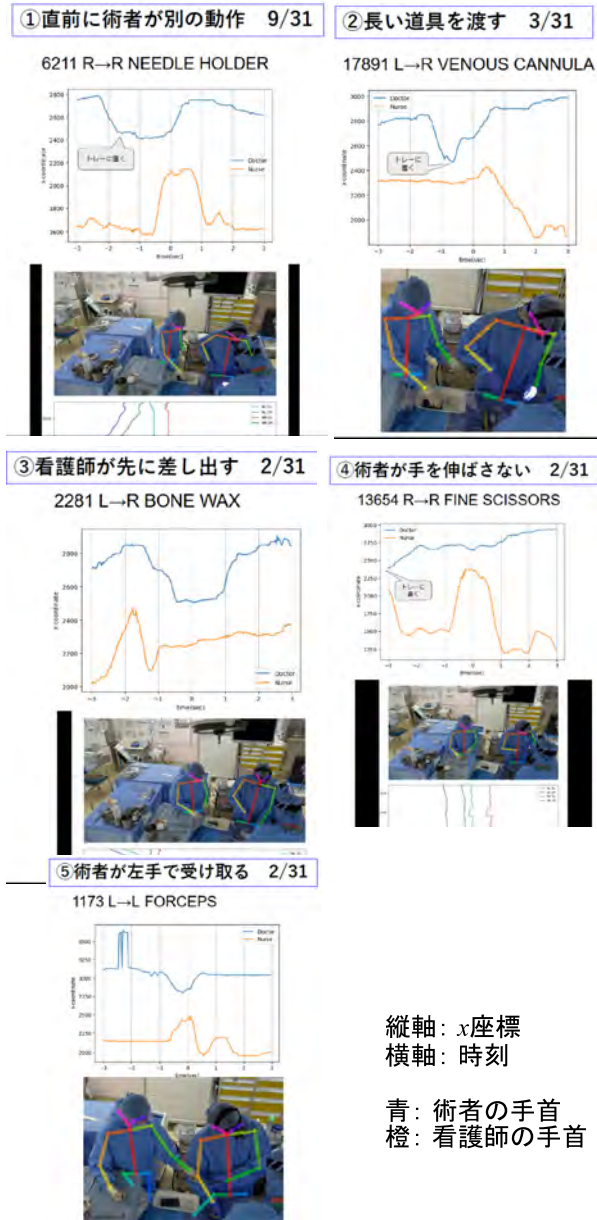


図 9: 手渡し時刻の判読が困難な例

- automatic surgical gesture recognition in video. In *International conference on medical image computing and computer-assisted intervention*, pp. 467–475. Springer, 2019.
- [7] Shota Kishi, Nozomu Suzuki, Shota Tsuyuki, Takio Kurita, Fujio Miyawaki, and Akinori Hidaka. Convolutional neural network based on temporal pose features for surgical procedure recognition. In *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications*, Vol. 2021, pp. 60–64. The ISCIE Symposium on Stochastic Systems Theory and Its Applications, 2021.
- [8] Adam Goldbraikh, Netanell Avisdris, Carla M Pugh, and Shlomi Laufer. Bounded future ms-ctn++ for surgical gesture recognition. In *European Conference on Computer Vision*, pp. 406–421. Springer, 2022.
- [9] Shuja Khalid, Mitchell Goldenberg, Teodor Grantcharov, Babak Taati, and Frank Rudzicz. Evaluation of deep learning models for identifying surgical actions and measuring performance. *JAMA network open*, Vol. 3, No. 3, pp. e201664–e201664, 2020.
- [10] Daichi Kitaguchi, Nobuyoshi Takeshita, Hiroki Matsuzaki, Hiroaki Takano, Yohei Owada, Tsuyoshi Enomoto, Tatsuya Oda, Hirohisa Miura, Takahiro Yamanashi, Masahiko Watanabe, et al. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surgical endoscopy*, Vol. 34, pp. 4924–4931, 2020.
- [11] Giovanni Menegozzo, Diego Dall’Alba, Chiara Zandona, and Paolo Fiorini. Surgical gesture recognition with time delay neural network based on kinematic data. In *2019 International symposium on medical robotics (ISMR)*, pp. 1–7. IEEE, 2019.
- [12] Flakë Bajraktari, Kathrin Fleissner, and Peter P Pott. A comparison of two cnn-based instrument detection approaches for automated surgical assistance systems. In *Current Directions in Biomedical Engineering*, Vol. 9, pp. 599–602. De Gruyter, 2023.
- [13] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [14] Glenn Jocher et al. ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements, Oct. 2020.
- [15] Vadisetti G.P. Niranjana A. Saranu K. Sarma R. Shaik M.A.B. Gudepu, P.R. and P Paramasivam. Whisper Augmented End-to-End/hybrid Speech Recognition System - CycleGAN approach. In *InterSpeech 2020*, pp. 2302–2302, 2020.
- [16] 村上貴志. 音声トリガーも利用した AI による手術自動分析システムの提案と初期評価報告. https://www.surgicaleducation.jp/images/ses2023/10th_program.pdf, 2023.
- [17] 深層学習を使用したノイズに含まれる音声区間の検出. <https://jp.mathworks.com/help/deeplearning/ug/voice-activity-detection-in-noise-using-deep-learning.html>.
- [18] Labeled Surgical Tools and Images. <https://www.kaggle.com/datasets/dilavado/labeled-surgical-tools>.