

厚生労働科学研究費補助金（臨床研究等 ICT 基盤構築・人工知能実装研究事業）
『AI の眼』による医療安全確保に関する研究

分担研究報告書

「人工知能技術を用いた手術訓練用模型シミュレーター動画解析に関する研究」

研究分担者 上浦尚武 兵庫県立大学教授

【研究要旨】 眼内レンズ挿入を伴う白内障手術は外科手術全体においてもその件数は突出しており日本国内だけでも年間 150 万件を超える。眼内レンズが登場してからすでに 30 年以上が経過しており、眼内で眼内レンズが固定位置から偏位し、眼内に落下するという事例が急激に増加している。光学系が著しく損なわれるために視力は失明領域まで一挙に低下し、外科的な回復手術が必須となる。現況で回復手術の主流となっているのが強膜内固定術であるが、この手術は眼内レンズのループ部分操作が初心者には難しく、事前の模型を用いたドライシミュレーション訓練が非常に有効だとされている。今回我々はこの模型訓練動画を AI で学習させることで、実手術可能なレベルに到達したかどうかについて AI 判断を行い、手術トレーニングの自学自習システムを構築することを目的とした研究を開始した。今回は大量のデータを処理しなければならない動画 AI について、最適なフレームレートおよびモデル構築について複数の条件設定において比較研究した。その結果、従来の CNN に加えて **Optical Flow** を併用することでよりリアルタイム性が高い検討が可能であることが示された。

A. 研究目的

眼内レンズ強膜内固定術動画を対象とし、手術技術評価システム構築の一環として、ニューラルネットワークと **Optical Flow** に基づく手術器具追跡を提案する。提案法ではニューラルネットワークの 1 種である **scSE-FC-DenseNet** と **Optical Flow** を併用する。**scSE-FC-DenseNet** のみを用いた手術動画像における術器具追跡法は、高精度で手術器具の検出を可能とするが、比較的長い処理時間を要する。一方、**Optical Flow** のみを用いた場合は高速処理が可能となる反面、一度あるフレームで誤った部位を手術器具として認識すると、それが後に続くフレームでの誤検出の一因となる。本研究では、**scSE-FC-DenseNet** と **Optical Flow** の併

用により、リアルタイムでの正確な手術器具検出を目指す。

B. 研究方法

本研究では、社会医療法人三栄会ツカザキ病院で撮影された模擬眼に対する眼内レンズ強膜内固定術 13 件の動画像に対し、手術器具検出を行う。これらには眼内レンズ強膜内固定術の重要過程である後方眼内レンズ支持部の受け渡しと引き抜きまでの処理が記録されている。動画の解像度は 1920×1080 でフレームレート 30FPS、動画時間の平均は約 26 秒である。手術器具検出では学習データを 12 件、テストデータを 1 件とする。提案法では動画フレームレートを 10FPS に、解像度を 256×168 にする。

これにより 13 件の動画から、学習データ 3480 枚、テストデータ 300 枚の画像が得られる。主要な手術器具についてアノテーションを行い、ニューラルネットワークの学習に使用した。手術過程のフレーム例を図 1 に示す。同図において右下の主要な器具が提案法における主な検出対象となる。

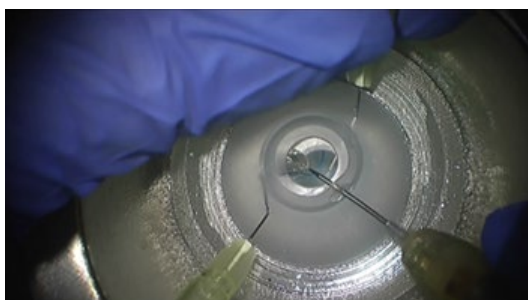


図 1) 眼内レンズ強膜内固定術の主要器具

ニューラルネットワークの学習については、必要に応じてデータ拡張の採用により学習データの枚数を増やす。本研究では、データ拡張採用時、提案法は原画像、左右反転、上下反転、上下左右反転の 4 枚およびそれらに対して最大 20% の上下左右にランダムに平行移動した画像を 2 枚ずつ作成する。これらによって 1 枚を 12 枚に拡張し、学習データ 41760 枚、テストデータ 300 枚の画像を得る。入力画像と主要な器具のマスク画像を図 2 に示す。

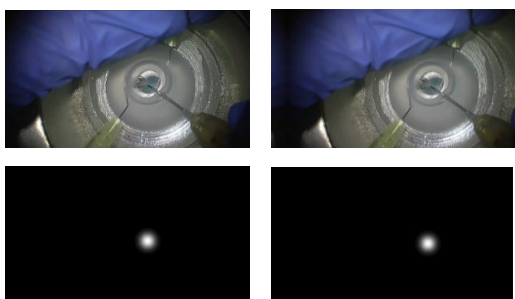


図 2 入力画像とマスク画像の例

(左:原画像, 右:平行移動画像)

本研究を担当する田淵および上浦は、

白内障手術動画に記録された前囊切開過程における手術器具の検出のために、ニューラルネットワークの 1 種で、40 層の Dense ブロックからなる scSE-FC-DenseNet 40 を使用している。scSE-FC-DenseNet40 の入力は $256 \times 128 \times 3$ のカラー画像である。scSE-FC-DenseNet40 の構造を表 1 に示す。本研究では、scSE-FC-DenseNet40 を使用する器具追跡方法に加え、Optical Flow を以下に示すように併用する。まず、M フレームに対して scSE-FC-DenseNet40 で対象器具の先端を検出し、それを特徴点として確定する。そして、次に (M+1) フレームから (M+N-1) フレームまで Optical Flow で特徴点の追跡を続ける。そして、(M+N) フレームで再び scSE-FC-DenseNet40 で特徴点を確定させる。以上を繰り返し、特徴点を見失うことなく、高速で追跡を実行する。以下では scSE-FC-DenseNet40 で特徴点を確定し、その後 (N-1) 枚のフレームについて Optical Flow が適用される場合を optN と記す。また、scSE-FC-DenseNet40 のみの連続使用により器具先端を追跡し続ける場合、それを cnn とよぶことにする。cnn は田淵、上浦が提案した先行研究である scSE-FC-DenseNet40 のみを用いる方法と等価である。本研究で行う実験では、工業的生産品である模擬眼に対する手術動画を対象とする。よって、倫理面での留意点は発生しない。

表 1. 使用した scSE-FC-DenseNet40 の構造

Type	Patch size/stride	Input shape	Skip Connection
Convolution	3 × 3/1	256 × 128 × 3	
2 × (Dense Block & scSE module)		256 × 128 × 48	
Convolution	1 × 1/2	256 × 128 × 80	
4 × (Dense Block & scSE module)		128 × 64 × 80	Output (1)
Convolution	1 × 1/2	128 × 64 × 144	
8 × (Dense Block & scSE module)		64 × 32 × 144	Output (2)
Convolution	1 × 1/2	64 × 32 × 272	
6 × (Dense Block & scSE module)		32 × 16 × 272	Output (3)
Transposed Convolution	3 × 3/2	32 × 16 × 96	
8 × (Dense Block & scSE module)		64 × 32 × 368	Concat (3)
Transposed Convolution	3 × 3/2	64 × 32 × 128	
4 × (Dense Block & scSE module)		128 × 64 × 272	Concat (2)
Transposed Convolution	3 × 3/2	256 × 128 × 64	
2 × (Dense Block & scSE module)		256 × 128 × 144	Concat (1)
Convolution	1 × 1/1	256 × 128 × 176	
Sigmoid		256 × 128 × 3	

C. 研究結果

提案法と **cnn** によって、テストデータの膜内固定術の手術画像 300 枚における手術器具の先端を検出する。本研究では、GPU 搭載マシン（GPU：GeForce GTX 1080Ti，メモリ：8GB，HDD：1TB）上での検出で、学習データ生成時に拡張処理無し，データ拡張ありの 2 ケースについて実験する。また，CPU 搭載マシン（CPU：Intel Core i7-10700，メモリ：8GB，HDD：1TB）上で学習データ生成時に拡張処理ありの場合の検出結果も示す。さらに，Optical Flow の使用間隔，すなわち optN における N をパラメータとした場合の結果の違いについても論じる。N の値は 3，5，7，10，20 と変化させる。よって，ここでは **cnn** に加えて opt3，opt5，opt7，opt10，opt20 について検証することとなる。実験のパラメータの一つとして IoU を用いる。これは 2 つの領域がどれくらい重なっているのかを表す。テストデータに対して手術器具の先端を検出した結果の一例を図 3 に示す。ただし，赤が提案法による検出結果で，緑が正解部位であり，この場合の IoU は 0.542 である。一般に IoU が 0.4 程度であれば目視によっても実用上問題はない。したがって，本研究では IoU が 0.4 以上のと

き，手術器具の先端は検出できたとみなす。

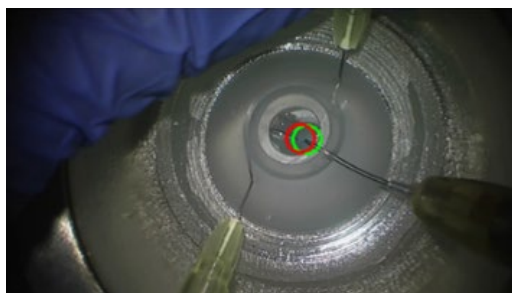


図 3 手術器具の検出結果

（opt5，データ拡張有り）

IoU 0.4 以上を成功としてテストデータ 300 枚について計算した検出率（すなわち，提案法により正しく先端を検出できているフレーム数をデータ総数 300 で割った値）と，フレーム 100 枚目～300 枚目までの計 200 枚の検出についての処理速度を計算し，1 秒間に検出できる画像枚数を計算した。結果を図 4 に示す。なお，opt の後の数字から 1 減じた値は，Optical Flow が連続して適用され続けるフレームの枚数を表している。同図において，横軸を 1 秒間に検出できる画像枚数，縦軸を検出率としている。また，同図は，データ生成時に拡張ステップを採用し，かつ提案法を GPU 搭載マシン上で実装した場合の結果である。

次に，上記と同様の条件で提案法を CPU 搭載マシン上で実装した場合の結果を図 4 に示す。この実験は，手術室に GPU 搭載マシンの設置が困難であり，CPU 搭載の一般的なコンピュータのみ使用できる場合を想定して実施したものである。

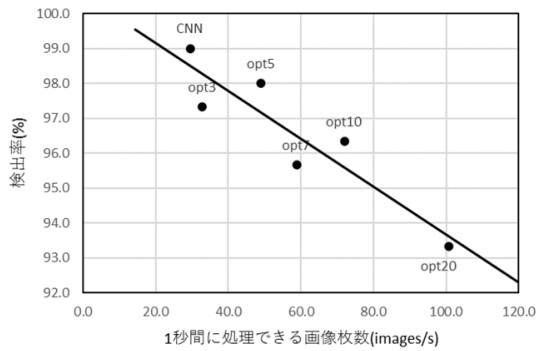


図4 GPU搭載マシンで1秒間に検出できる画像枚数と検出率の関係 (データ拡張有り)

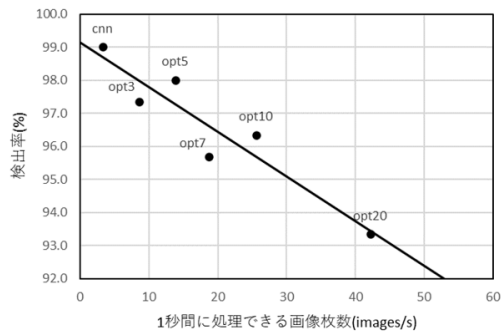


図5 CPU搭載マシンで1秒間に検出できる画像枚数と検出率の関係 (データ拡張有り)

次に、上記と同様の条件で提案法をCPU搭載マシン上で実装した場合の結果を図4に示す。この実験は、手術室にGPU搭載マシンの設置が困難であり、CPU搭載の一般的なコンピュータのみ使用できる場合を想定して実施したものである。

D. 考按

実験結果において、1秒間に検出できる画像枚数と検出率の関係はトレードオフの関係にある。すなわち、提案しているoptNはcnnと比べると検出率が少し下がってしまうが、検出速度を上げることができる。特に、一般的なCPUのみ搭載のコンピュータで器具追跡を行った場合、ニューラルネットワークのみで検出すると1秒間に約3.3枚しか画像を処理できないため、リアルタイムで手術器具の追跡ができていない。一方、opt5では1秒間に約13.9枚処理できるため、リアルタイム性は

保持できていると考えられる。また、cnnでは検出率として1%落ちてしまっているが、誤検出が増える傾向にある。すなわち、手術器具がフレームに明確に現れているときの追跡についてはcnnとopt5はほとんど変わらないと示される。

E. 結論

実験の結果、1秒間に処理できるフレーム枚数と検出率の関係はトレードオフ関係にあることが示された。CPU搭載マシン上で従来法(cnn)を用いる場合と、提案法であるopt5を実行した場合、後者は前者と比べて検出率は1%下がっているが、1秒間に検出できる枚数が10.9枚増えている。よって、提案法は従来法に比べて、リアルタイム性の観点から優れていることが明らかとなった。

GPU搭載マシンでopt5を実行した場合、1秒間にフレームを49.0枚処理できる。これは高速ではあるが、眼内レンズ強膜内固定術の手術動画が30FPSであるので、過剰な処理性能と断じることでもできる。ただし、システムをより現実的にするのであれば、追跡点を増やし手術の評価を行う処理が必須となる。すなわち、リアルタイム性を維持しつつ実用性も高めるためには、手術器具の追跡に費やせるリソースは確実に減じる。よって、過剰性能といえる提案法は、本格的に運用可能な手術教育システムの開発に余地を残しているとみなすことができる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

なし

2. 学会発表

平野颯大, 森田翔治, 礪川悌次郎, 上浦尚武, 田淵仁志, “眼内レンズ強膜内固定術動画に対するニューラルネットワークとOptical Flowに基づくリアルタイム器具追跡,” 電子情報通信学会技術研究報告 FIIS-22-559, 2022年3月18日

H. 知的所有権の取得状況

1. 特許取得

申請検討・準備中

2. 実用新案特許

なし

3. その他

なし