

## 平成 29 年度 情報工学コース卒業研究報告要旨

森 研究室	氏 名	小 澤 卓 也
卒業研究題目	深層学習を用いた内視鏡動画像に対する 大規模アノテーションデータの構築手法に関する研究	

本稿では、内視鏡動画像に対して、術具や解剖領域の詳細なアノテーションを深層学習を用いて効率的かつ半自動的に構築する手法について述べる。内視鏡動画像に対して詳細に術具や解剖領域がアノテーションされたデータは、手術支援機器における画像からの解剖構造の自動識別技術の実現や内視鏡手術手技の評価技術の実現など、内視鏡手術の安全性を高める研究において利用価値が高い。しかし現状、手術映像中における術具や解剖構造を詳細にセグメンテーションしたデータベースは存在しない。先述のように手術映像アノテーションデータベースは手術機器開発等に欠かせなく、これらの動画に対して効率的にアノテーション作業を行う手法が求められている。

そこで本研究では 1) 手動アノテーションのためのツール, 2) 深層学習を用いた半自動アノテーション手法をそれぞれ開発する。1) では、内視鏡映像の各フレームに対してアノテーションを行うことのできるツール「NuVAT.exe」を開発する。2) では手動アノテーションを、深層学習による自動認識により半自動化する手法を実現する。NuVAT を用いて内視鏡手術ビデオシーケンスの一部のフレームに対してアノテーションデータを作成する (図 1)。このデータを用いて深層学習の 1 つである U-Net を用いて、ラベル付けされていないフレームに対して自動アノテーションを行う (図 2)。誤検出が存在する自動抽出結果をアノテーションツールで修正し、修正データを再び深層学習で学習させる。この処理を反復し、全てのフレームに対して正確なアノテーションを行う。

実験では、4 例の内視鏡映像全 104 万フレームを用いた。1326 枚のフレーム画像を用いて U-Net を訓練し、8 種類の鉗子に対して自動アノテーションを行った際の検出率を調査した。鉗子領域は柄と先端で異なるクラスとして扱った。1) のツールでラベル付けされた学習用画像と評価用画像 80 枚を用いて学習と評価を行った。評価は各鉗子ごとの F 値を用いて行った。実験の結果、それぞれの鉗子の柄と先端のラベル付け精度 (F 値) は表 1 に示す通りになった。学習データに高頻度で出現する鉗子に対する自動アノテーションの精度が高かった一方で、出現頻度が低い鉗子の検出率は低く、評価用データに出現するが自動アノテーションで 1 ピクセルもラベル付けされない鉗子領域が 4 種類存在した。ラベルの修正のみに要する時間は、修正元となるラベル画像が存在しない状態から手動で生成する場合と比べて 8%短縮された。半自動アノテーション手法は、内視鏡動画の全フレームにラベル付けする上で有効だと知られた。

表 1: 各鉗子の柄・先端に対する検出率 (F 値)

	Grasper	Point dissector	Linear dissector	Maryland	Suction
柄	0.779	0.824	0.832	0	0
先端	0.873	0	0.763	0.686	0

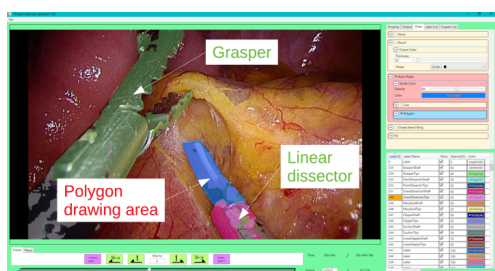


図 1: アノテーションツールでラベル付けを行う様子。多角形描画ツールの描画領域が青色で示されている。

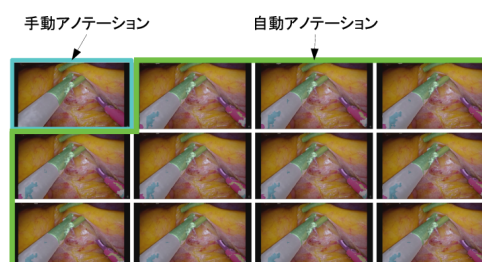


図 2: 連続のフレームにラベル画像を統合した画像。左上は手動生成したデータ。それ以外は自動生成したデータ。