H-028

手技習得を目的とした腹腔鏡手術映像に対する映像補完用マスク画像の自動生成
Automatic Generation of Mask Images for Video Inpainting in Laparoscopic Surgery Videos for Skill Acquisition

大西 涼介† 大野 将樹‡ 獅々堀 正幹‡
Ryosuke Onishi Masaki Oono Masami Shishibori

1. はじめに

近年，映像教材を用いた手術手技の教育支援が進んでおり，鉗子の操作や手技の流れを視覚的に学習することが可能となっている．しかし，多くは手技を視聴するだけの受動的な学習にとどまっている．この課題に対し，鉗子を除去し，補完した映像を活用することで，学習者自身の判断を促す能動的な学習が可能となる．映像補完技術では，補完したい領域をマスクした画像を事前に準備し，Lama[1]などを用いることで，マスク領域が補完される．映像の全フレームのマスク画像を手作業で作成するには多大な労力が必要であるが，XMem[2]を用いると，先頭フレームに対するマスク画像のみを作成しておけば，後続フレームの鉗子の動きを追跡したマスク画像を自動的に生成できる．しかし，鉗子が遮蔽された場合や他の鉗子と重なった場合にはマスク画像の精度が低下することが多い．本稿では，この課題を解決するために，YOLOv5[3]によって検出された鉗子領域とマスク領域の面積比に基づき，不備マスクを自動的に検出する手法を提案する．なお，本稿では腹腔鏡手術映像から鉗子部分を補完する目的のために本手法を適用する．鉗子が補完された映像は，術者の手技習得のために用いられる．

2. 提案手法

本手法は，XMem を実行して鉗子のマスク画像を自動生成し，手術画像を用いて鉗子を学習した YOLOv5[3]を活用し，精度の低いマスク画像を自動的に検出する．提案手法の概要を図 1 に示す．

2.1 XMem を用いたマスク画像の自動生成

XMem の実行前の事前準備として，YOLOv5 を用いて，入力映像とは異なるシーンの鉗子画像を用いた学習を行う（図 1 左上）．この学習モデルはマスク不備フレームの検出に用いられる．まず，マスク生成対象となる映像を 1 フレームごとに画像に変換し，先頭フレームのマスク画像を手作業で作成する．次に，全フレームの画像と先頭フレームのマスク画像を XMem に入力することで，図 1 中央に示すように，全フレームに対応するマスク画像が自動生成される．

2.2 マスク不備フレームの検出手順

マスク不備フレームの自動検出手順は，主に 3 つの処理に分類される．以下にその概要を示す．

2.2.1 鉗子画像の推論

XMem の実行で入力した鉗子画像を事前に学習した YOLOv5 の学習モデルに入力し，鉗子検出の推論処理を行う．推論によって出力されたバウンディングボックスの座標情報をマスク不備フレームの検出に利用する．

†徳島大学大学院創成科学研究科, Graduate School of Sciences and Technology for innovation, Tokushima University
‡徳島大学大学院社会産業理工学研究部, Graduate School of Technology, Industrial and Social Science, Tokushima University

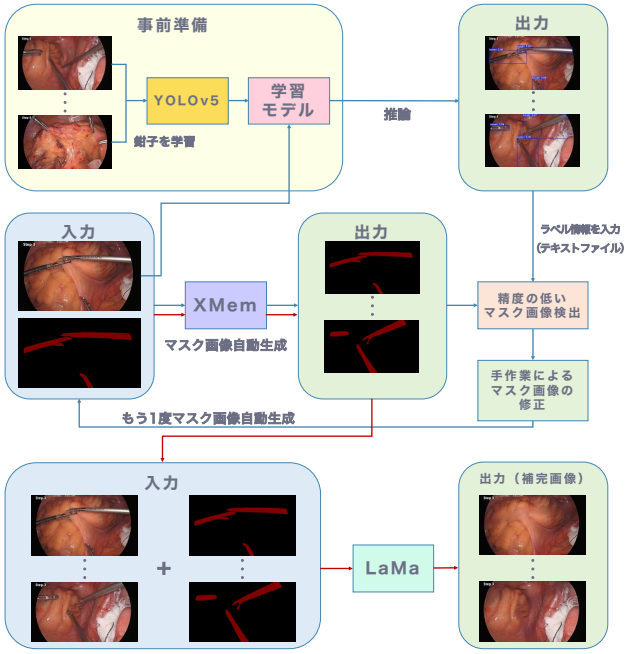
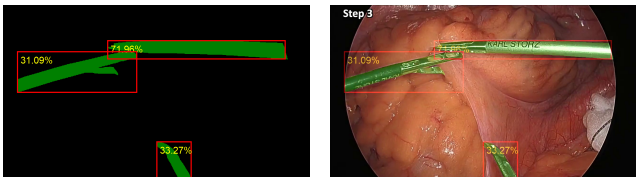


図 1　提案手法の概要

2.2.2 バウンディングボックス内の鉗子面積割合計算

YOLO の推論で出力されたラベル情報を用いて，バウンディングボックスの面積とバウンディングボックス内でのマスク領域の面積との比率を算出する（図 2 参照）．面積と面積比率はテキストファイルとして出力し，マスク不備フレームの検出プログラムに入力される．



(a) マスク画像　(b) 鉗子画像と重ね合わせた画像

図 2　鉗子の割合計算結果の例

2.2.3 マスク不備フレームの検出

マスク不備フレームの検出は，バウンディングボックスの面積と，その内部にマスクされた鉗子の割合との関係を回帰分析により解析することで行う．正しくマスクされた画像を用い，バウンディングボックスの面積（ピクセル値）をx，バウンディングボックス内のマスクされた鉗子の割合をyとして回帰分析を行い，6 次の回帰式を導出した．検出時には，この回帰式から得られる鉗子の割合の予測値と実測値の残差が標準偏差の 1.5 倍を超えた場合に，不備フレームと判断する．なお，判断基準は下限のみに設定し，予測より割合が大きい場合はマスク精度に問題がないとみなす．

3. 評価

徳島大学大学院医歯薬学研究部消化器・移植外科教室より腹腔鏡下 S 状結腸切除術の手術映像[4]を提供いただき，2 つのシーンを切り取って評価に使用する．この手術は，腹腔鏡を使用した手術であり，鉗子などの器具の操作に高度な技術を要する．使用するデータを表 1 に示す．

表 1　評価使用データ

評価データ	フレーム画像総数（枚）
データ 1	436
データ 2	359

これらのデータに提案手法を適用し，マスク不備フレームを検出する．マスク不備フレームの正解データは，XMem により生成されたマスク画像と元画像を比較し，目視により特定したものである．これらのデータと検出結果を比較し，評価を行う．評価指標として検出率を用いる．

$$\text{検出率（\%）} = \frac{\text{提案手法で検出したマスク不備フレームの数}}{\text{正解データのマスク不備フレームの数}}$$

3.1 評価結果

2 つのデータのそれぞれの正解データと提案手法を用いた検出の比較結果を表 2，表 3 に示し，図 3，図 4 に検出結果を示す．

表 2　データ 1 の評価結果

フレーム番号	正解データ	提案手法	比較結果
0～312	正常	正常	一致
313～315	マスク不備	正常	提案手法で不備未検出
316～435	マスク不備	マスク不備	一致

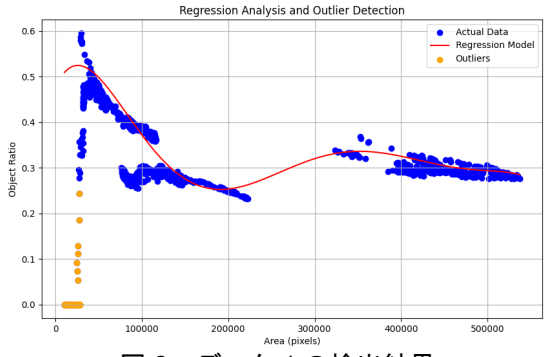


図 3　データ 1 の検出結果

図 3，図 4 のグラフにおける赤線は回帰曲線であり，青色の点は実際のバウンディングボックスの面積と鉗子の割合，黄色の点はマスク不備フレームと識別されたデータを示している．

3.2 考察

データ 1 において，提案手法で未検出であったのは 3 フレームに限定された．これらのフレームは，鉗子部分の大部分が適切にマスクされており，マスク領域の膨張処理を実施した上で画像補完を適用することにより，鉗子部分は十分に補完可能であると考えられる．全体として，マスク不備フレームの大部分が検出されており，検出率は 97.6%と高く，提案手法の有効性が確認された．

一方，データ 2 では YOLOv5 による鉗子認識精度が十分ではなく，特に映像の端に位置する鉗子が検出されないケースが多く確認された．その結果，検出率 79.7%となり，データ 1 と比べて低下した．

以上の結果から，提案手法の平均検出率は 88.7%であり，いずれのデータにおいても概ねマスク不備フレームを検出できていることが確認された．一方で，YOLOv5 による鉗子検出精度が低下した場合には，不備フレームの検出精度にも影響を及ぼすことが明らかとなった．

表 3　データ 2 の評価結果

フレーム番号	正解データ	提案手法	比較結果
0～204	正常	正常	一致
205～209	マスク不備	正常	提案手法で不備未検出
210～219	正常	正常	一致
220～227	マスク不備	正常	提案手法で不備未検出
228～238	正常	正常	一致
239	正常	マスク不備	正解データで不備未検出
240～292	正常	正常	一致
293～296	マスク不備	正常	提案手法で不備未検出
297～358	マスク不備	マスク不備	一致

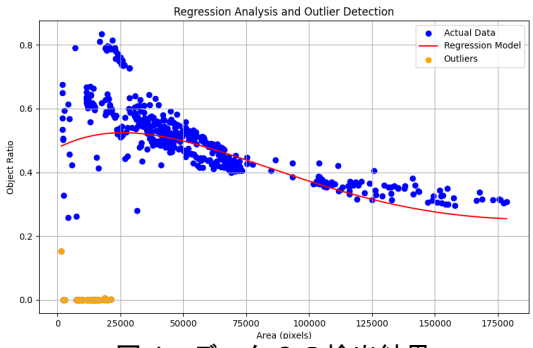


図 4　データ 2 の検出結果

4. まとめ

本稿では，映像補完用マスク画像の自動生成におけるマスク不備の問題に着目し，マスク不備フレームを自動的に検出する手法を提案した．具体的には，YOLOv5 による鉗子領域の検出と回帰分析を組み合わせることで，従来手作業で行っていたマスク不備の検出を自動化するものである．評価の結果，2 つのデータにおいて平均検出率は 88.7%となり，マスク不備フレームを概ね正確に検出できることが確認された．また，軽微な不備に関しては検出が困難であったが，それらはマスク領域の膨張処理により補完可能であり，実用上の影響は小さいと考えられる．しかし，提案手法は YOLOv5 の検出精度に依存しており，鉗子の検出が不完全な場合にはマスク不備の検出精度も低下するという課題がある．今後は学習データの拡充や検出モデルの改善が重要な課題である．

参考文献

[1] Suvorov, Roman, et al. "Resolution-robust large mask inpainting with fourier convolutions." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022.
[2] Cheng, Ho Kei, and Alexander G. Schwing. "Xmem: Long-term video object segmentation with an atkinson-shiffrin memory model." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.
[3] Ultralytics. YOLOv5. https://github.com/ultralytics/yolov5
[4] 竹政伊知朗 浜部敦史．“Procedure Book Vol.18 腹腔鏡下 S 状結腸切除術を極める！-超音波凝固切開装置-”．Medtronic．徳島大学大学院医歯薬学研究部消化器・移植外科教室より提供．