

## 深層学習による 8K 内視鏡映像の高精細化

和田 歩<sup>†</sup> (正会員) 大和 淳司<sup>†</sup> 合志 清一<sup>†</sup> (正会員)<sup>†</sup>工学院大学情報学部

## Image Enhancement for 8K Endoscope Videos Based on Deep Learning

Ayumu WADA<sup>†</sup> (Member), Junji YAMATO<sup>†</sup>, Seiichi GOHSHI<sup>†</sup> (Member)<sup>†</sup>Department of Information Science, Kogakuin University

〈あらまし〉 内視鏡手術は切開創が小さく術後の回復が早いいため手術件数は年々増加傾向にある。開腹手術よりも患者への負担が少なく高齢者に対しても有用な内視鏡手術であるが、手術難易度が高く執刀可能な医師の数が不足している。近年、内視鏡手術の技術水準を向上させるため 8K 解像度の内視鏡カメラが開発された。8K 内視鏡カメラは患部の質感や奥行き感など多くの情報を取得するため、患部をより詳細に観察可能であり手術難易度を下げることには貢献し得ると期待されている。しかし、8K 内視鏡カメラは従来よりも被写界深度が浅く、血管や縫合糸など複数の対象にフォーカスを同時に合わせることに困難である。そこで本研究では 8K 内視鏡映像に対し信号処理を施すことで、フォーカス合わせが不十分であっても精細感を得られるよう映像を改善する手法を提案する。また、提案手法の性能を定量化するため主観評価実験により手術で使用される縫合糸の視認性とノイズによる画質劣化について評価した。実験の結果、ノイズの増幅による画質劣化が避けられ、縫合糸をより短時間で視認できることを確認した。

キーワード：8K 内視鏡カメラ、深層学習、高精細化、ノイズ低減

<Summary> Number of endoscopic surgery trends upward because the postoperative recovery is fast and the patient can be discharged immediately. Although endoscopic surgery is less burden on patients than open surgery, its difficulty is high for unskilled doctors. Recently, an 8K resolution endoscope camera has been developed to improve the technical level of endoscopic surgery. An 8K endoscope camera is expected to reduce the difficulty of surgery because it can obtain a lot of information about texture and depth of the organ. However, an 8K endoscope camera has a shallower depth of field than traditional cameras, and it is difficult to simultaneously focus on multiple objects such as blood vessels and sutures. In this paper, the method which makes 8K endoscopic videos with insufficient focusing sharp is proposed. And then, in order to quantify the performance of the proposed method, the visibility of sutures used in surgery and the image quality degradation due to noise were evaluated by subjective assessments. As a result of the experiments, it is shown that image quality degradation was avoided, and that sutures could be found in a shorter time.

**Keywords:** 8K endoscope cameras, deep learning, image enhancement, noise reduction

## 1. はじめに

患者への負担が少ない外科治療として内視鏡手術が普及している。内視鏡手術では腹部に開けた 1cm の穴から内視鏡や鉗子を挿入し、モニターに映し出された患部を観察しながら治療を行う。開腹手術よりも切開創が小さく術後の回復が早いいため、患者は早期に退院が可能である。高齢化が進む日本では体力が少ない高齢者に対して内視鏡手術の有用性が期待されている<sup>1)</sup>。しかし、内視鏡手術ではモニターに映し出された映像を基に治療を進めるため、手術難易度が高く執刀可

能な医師の数が不足している<sup>2)</sup>。

近年、内視鏡手術の技術水準を向上させるため 8K 内視鏡カメラが開発された<sup>3)</sup>。8K 内視鏡カメラは血管や臓器を従来よりも高い解像度で撮影することで、患部の質感や奥行き感などの情報を執刀医に提供する。臨場感が高い映像は手術の根治性を向上させるため、より安全な手術として 8K 内視鏡手術が期待されている。また、従来の HD/4K 内視鏡カメラは患部を観察するため対象に近づく必要があった。これは内視鏡とその他の医療器具とが接触する恐れがあり、臓器を損傷させるリスクがある。しかし、8K へと解像度が向上したこ

とで電子ズームによって患部を精細に映し出せるため、内視鏡カメラを動かす必要が無く、臓器損傷のリスクが減少すると期待されている。従来よりも優位性がある 8K 内視鏡カメラであるが、実用化とともにフォーカス合わせが新たな課題として表面化した。8K 内視鏡カメラの撮像素子は従来の HD/4K 解像度よりも画素あたりの面積が小さく、各素子に入射する光量が少ない。より多くの光を撮像素子に集めるため 8K 内視鏡カメラでは絞りを開くが、その結果として被写界深度が浅くなる。このためフォーカスを臓器に合わせた場合、内視鏡手術で用いる直径 0.02mm 以下の縫合糸は 8K 本来の解像感が得られず視認性が悪い。そこで本研究では 8K 内視鏡映像の微弱な高周波成分を強調することで、フォーカス合わせが不十分であっても縫合糸を高精細化する手法を提案する。

## 2. 従来手法

画像や映像を高精細化する手法として超解像技術が存在する。超解像は原画像のナイキスト周波数を超える高周波成分を生成する技術であり画像の輪郭を鮮明にする。いくつかの超解像技術が提案されている<sup>4,5)</sup>が、いずれもリアルタイムで処理することが困難であった。一方、テレビでのアップコンバートを目的とした再構成型超解像<sup>6)</sup>や非線形超解像<sup>7)</sup>はリアルタイムで計算可能である。再構成型超解像は複数のフレームから被写体が映る画素をサンプリングすることで高精細化を行う。被写体を鮮明に撮影したフレームが存在すれば高精細な映像を得られるが、縫合糸にフォーカスが合っていない 8K 内視鏡映像への適用は困難である。一方で、非線形超解像は 1 枚のフレームに対して処理可能であり、8K 内視鏡映像の高精細化手法としても検討された<sup>8)</sup>。しかし、非線形超解像には高周波成分であるノイズの増幅が伴うため、縫合糸を鮮明にしたが同時に画質が劣化してしまった。そこで窪田らはノイズ低減手法の一つである Total variation 法を 8K 内視鏡映像に適用した<sup>9)</sup>。Total variation 法は入力画像のエッジ成分を保持しつつ画像の劣化過程を推定するため、鮮明なノイズ低減画像が期待できる。しかし、先行研究において処理後の画像を評価した結果、臓器表面上ではノイズ低減効果が見られたが、処理に伴い発生したアーティファクトが縫合糸を一部欠落させていた。

## 3. 提案手法

縫合糸や血管は執刀医が手術中に注視する重要な要素であり、欠落させずに映像をノイズ低減及び高精細化する必要がある。そこで提案手法ではノイズ低減や超解像技術で高い精度を持つニューラルネットワーク<sup>10)</sup>に、ノイズの特徴を学習させることを試みた。これによりノイズと他の高周波成分が区別され、従来手法<sup>9)</sup>よりもアーティファクトが少ない状態で、8K 内視鏡映像の高周波成分を強調することができた。

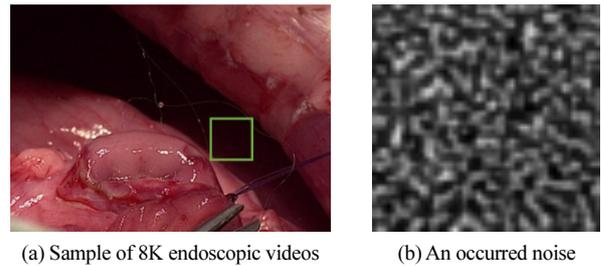


図 1 8K 内視鏡映像で発生したノイズの一例

Fig.1 The noise occurred in 8K endoscopic videos

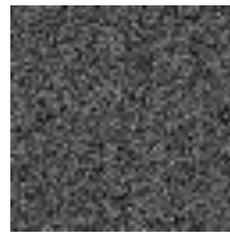


図 2 ガウシアンノイズ

Fig.2 Gaussian noise

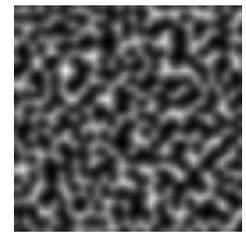


図 3 パーリンノイズ

Fig.3 Perlin noise

### 3.1 8K 内視鏡映像のノイズ

映像に発生するノイズは一般にガウシアンノイズとしてモデル化されることが多い<sup>11)</sup>。しかし、8K 内視鏡映像にはガウシアンノイズとは特徴が異なるノイズが発生している。図 1(a)に示す 8K 内視鏡映像から緑枠の範囲を切り取り、輝度成分の抽出した結果を図 1(b)に示す。抽出した画像に注目すると、8K 内視鏡映像には隣り合う画素と連続するようなノイズが発生していることがわかる。一方で、図 2 に示すガウシアンノイズは画像内で粒状となって現れており、図 1(b)のノイズと特徴が異なることがわかる。本研究では 8K 内視鏡映像に発生するノイズと同じ特徴を学習させるため、図 3 に示すパーリンノイズ<sup>12)</sup>を使用する。パーリンノイズはコンピュータグラフィックスにおいてテクスチャ生成などに使用される手法であり、図 1(b)のノイズと特徴が似て不規則に連続している。予備実験においてパーリンノイズを学習に用いた結果、8K 内視鏡映像に対するノイズ低減効果が確認できた。画像に対するパーリンノイズの付加は式(1)で表される。

$$\mathbf{F}(x, y) = \mathbf{I}(x, y) + \mathbf{Z}(x, y) \quad (1)$$

$$\mathbf{Z}(x, y) = \sum_{i=1}^N \alpha_i \times \mathbf{G}(\sigma_i) * \tilde{\mathbf{N}}_i(x, y) \quad s.t. \sum_{i=1}^N \alpha_i = 1 \quad (2)$$

式中の $\mathbf{F}(x, y)$ はノイズ混入画像、 $\mathbf{I}(x, y)$ は原画像を表す。 $\mathbf{Z}(x, y)$ は式(2)に従い生成されたパーリンノイズである。パーリンノイズの生成では複数のノイズを合成する。式中の $\tilde{\mathbf{N}}_i(x, y)$ は標準正規分布に従ったノイズであり、原画像とサイズが同じとなるように生成されている。パーリンノイズは

N 個のノイズを合計 1 である  $\alpha_i$  によって加重平均することで生成される。平均化される各ノイズにはガウシアンフィルタ  $G(\sigma_i)$  が畳み込まれ、ノイズ模様のスケールを変化させている。提案手法では以下のパラメータ  $\alpha_1 = 0.9, \alpha_2 = 0.1, \sigma_1 = 0.7, \sigma_2 = 0.2$  により、パーリンノイズを生成する。

### 3.2 学習方法

フォーカス合わせが不十分な 8K 内視鏡映像には不鮮明な臓器や縫合糸が映し出されている。提案手法では 8K 内視鏡映像の各フレーム画像 (8K 内視鏡画像) に対し、ノイズ低減処理及び高精細化処理を施すことで、ノイズを増幅せずに縫合糸や血管などを鮮明にする。提案手法の概要を図 4 に示す。ここで入力にはノイズを含んだ画像  $I_{in}$ 、中間出力はノイズ低減後の画像  $I_{middle}$ 、出力は縫合糸が強調された画像  $I_{out}$  である。提案手法は 2 つのニューラルネットワークによって構成され、入力画像に対しノイズ低減処理を行った後、高精細化処理を施す。これらネットワークは除去するノイズの特徴や強調する画像の高周波成分を事前に画像データセットから学習する必要がある。しかしながら、8K 内視鏡画像の高精細化を目的とした画像データセットは存在しない。そこで、新たに画像データセットを作成し、2 つのニューラルネットワークを学習させる方法について説明する。

提案手法では 2 つのニューラルネットワークを学習させるため、図 4 の入力画像  $I_{in}$ 、中間画像  $I_{middle}$ 、出力画像  $I_{out}$  に対応する 3 つの画像を用意する必要がある。これら学習用データセットを新たに作成するため、本研究では画像に対しノイズ付加と高周波成分の増幅を行う。学習用データセットの作成には DIV2K データセット<sup>13)</sup>を用いる。DIV2K データセットは 800 枚の自然画像から構成されている。初めにデータセット内の画像からサイズ  $144 \times 144$  のパッチ画像をランダムに抽出する。次に、パッチ画像から輝度成分を抽出するため、国際的な規格である ITU-R BT709<sup>14)</sup> が定める式(3)に従い赤 (R) 緑 (G) 青 (B) 成分から輝度成分 (Y) へと変換する。ここで抽出した輝度成分をノイズの無い画像と見なし中間画像に対応する学習データとする。

$$Y = 0.2126R + 0.7152G + 0.0722B \quad (3)$$

生成した中間画像の一例を図 5 に示す。図 5(a)は 8K 内視鏡映像から抽出したパッチ画像の輝度成分であり、(b)~(d)は学習データの中間画像である。それぞれの中間画像は(b)像の皮膚、(c)山肌、(d)石畳の道路から抽出されており、(a)の臓器表面上の凹凸と特徴が類似している。このような自然画像を用いることで、8K 内視鏡画像内の血管や臓器を強調させる高周波成分を学習可能にする。

次に、入力画像に対応する学習データを生成するため、上述した中間画像に対しノイズを付加する。この処理ではニューラルネットワークに 8K 内視鏡画像に発生するノイズの特徴を学習させるため、前節で説明したパーリンノイズを使用

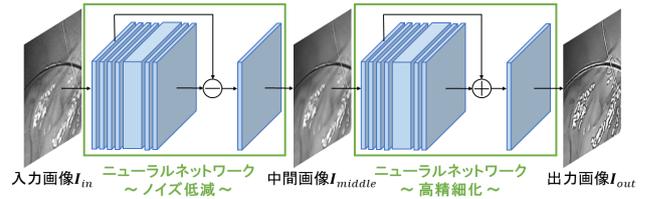
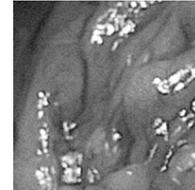


図 4 提案手法の概要

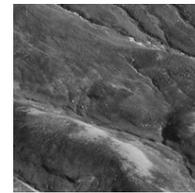
Fig.4 Outline of the proposed method



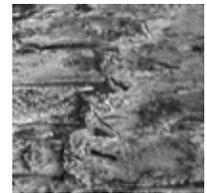
(a) An extracted patch image from 8K endoscopic videos



(b) Training data 1



(c) Training data 2



(d) Training data 3

図 5 中間画像の一例

Fig.5 Sample of  $I_{middle}$  images

する。最後に、出力画像に対応する学習データを生成するため、画像の高周波成分を増幅する。この増幅処理では、まずノイズ付加前の中間画像に対しガウシアンフィルタ ( $\sigma = 5$ ) を畳み込むことで低周波成分のみを抽出する。その後、抽出した低周波成分を元の中間画像から差し引くことで画像の高周波成分を抽出する。この高周波成分を中間画像に加算することで出力画像に対応した学習データを得る。

2 つのニューラルネットワークにノイズ低減と高精細化処理を学習させるため、提案手法では中間画像及び出力画像において推論結果の誤差を計算する。それぞれの誤差は次の式に従って求める。

$$loss_1 = \|\Delta I_{middle}\|_2 \quad (4)$$

$$loss_2 = \|\Delta I_{out}\|_1 \quad (5)$$

ここで式中のデルタ記号は推論結果と生成した学習用画像データとの差を意味する。中間画像における誤差は式(4)に従い、差分値に対して  $L_2$  ノルムを求める。これは 8K 内視鏡画像のノイズ低減に関する先行研究<sup>9)</sup>において、 $L_1$  ノルムよりも  $L_2$  ノルムの方が効果的であることが示されたためである。一方、出力画像における誤差は式(5)に従い、差分値に対して  $L_1$  ノルムを求める。ここでは高周波成分が失われにくい  $L_1$  ノルムを用いることで、 $L_2$  ノルムよりも解像感を向上させる<sup>15)</sup>。提案手法では式(6)に従いこれら誤差を足し合わせた損失値が小さくなるようニューラルネットワークを繰り返し学習させる。

$$\text{minimize}\{loss_1 + loss_2\} \quad (6)$$

### 3.3 ネットワーク構造

8K 内視鏡画像の高周波成分にはノイズや血管などの輪郭成分が含まれる。ノイズは不規則に発生するため一定の形状を持たない。一方、血管や神経は網目状に広がりノイズとは異なる形状を持つ。ノイズ低減や高精細化処理は画像の高周波成分を対象とするため、ノイズと輪郭成分を区別することが重要である。提案手法では畳み込みニューラルネットワークを構築することで、高周波成分から大局的な特徴を抽出し、その形状を識別していく。ノイズ低減及び高精細化を行う2つのニューラルネットワークの構造を図6、図7に示す。両者とも入力は左上、出力は左下であり、下段の加減算記号以外は同じ構成である。図中の画像は各層によって抽出された8K内視鏡画像の特徴マップを表す。

畳み込み処理を繰り返すと物体の形状を表す大局的な特徴マップが得られる。しかし、ノイズ低減や高精細化を行うには、ノイズや縫合糸の位置を表す局所的な特徴マップも必要である。2種類の特徴マップを並行して抽出するため、提案手法ではマルチスケールで特徴マップを抽出可能なResidual dense block (RDB)<sup>16)</sup>を用いる。最初に、8K内視鏡画像の高周波成分を抽出するため、“Conv<sub>1,1</sub>”、“Conv<sub>2,1</sub>”ではサイズ3×3の2次元形状のカーネルが64個畳み込まれる。この畳み込みは学習の結果ハイパスフィルタとして機能し、血管や臓器表面の凹凸から高周波成分を多角的に抽出する。次に、8K内視鏡画像の高周波成分をRDBに入力することでマルチスケールな特徴マップを抽出する。これは“RDB<sub>1,2~4</sub>”、“RDB<sub>2,2~4</sub>”で処理され、3回処理を重ねることで縫合糸や臓器表面の凹凸の形状などの大局的な特徴マップを段階的に抽出していく。各RDBで抽出された計192枚の特徴マップは取りうる値の範囲にバラつきがある。この不均一性はニューラルネットワークの学習を困難にするため、値を均一化する必要がある<sup>16)</sup>。そこで“Conv<sub>1,5</sub>”、“Conv<sub>2,5</sub>”ではサイズ1×1×192の3次元形状のカーネルを64個畳み込むことで、192枚の特徴マップを画素単位で均一化する。次に、64枚の特徴マップを基にノイズの位置や強調する縫合糸の形状を識別する。これは“Conv<sub>1,6</sub>”、“Conv<sub>2,6</sub>”で処理され、サイズ3×3×64の3次元形状のカーネルを64個畳み込むことで、特徴マップから多様な形状を識別していく。以上の処理により、ノイズの位置や縫合糸の輪郭を表す特徴マップが抽出される。これらを入力画像の高周波成分に対し、図6では加算、図7では減算することで、ノイズ低減及び高精細化した特徴マップを得る。最後の畳み込み処理“Conv<sub>1,7</sub>”、“Conv<sub>2,7</sub>”ではサイズ3×3×64の3次元形状のカーネルを畳み込むことで、高周波成分から物体の形状やテクスチャを再構成している。

提案手法で使用したRDBのネットワーク構造を図8に示す。RDBは畳み込み処理を階層的に構築したニューラルネットワークである。入力は左上、出力は左下である。RDBでは

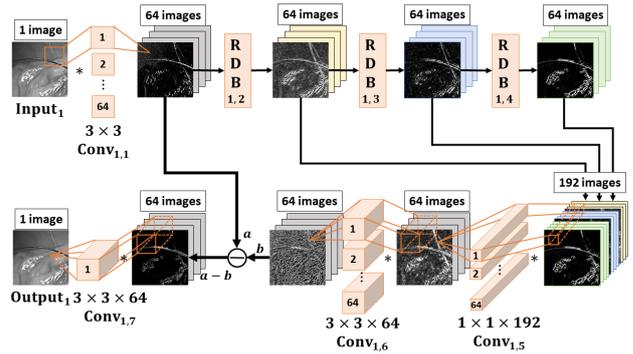


図6 提案するニューラルネットワークの構造 (ノイズ低減)  
Fig.6 The architecture of proposed method for noise reduction

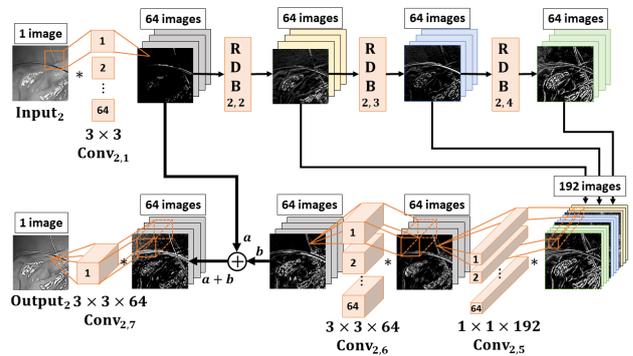


図7 提案するニューラルネットワークの構造 (高精細化)  
Fig.7 The architecture of proposed method for image enhancement

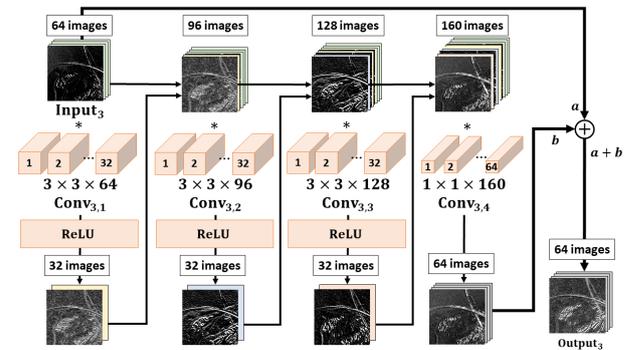


図8 提案手法で使用した Residual dense block  
Fig.8 Residual dense block used in proposed method

マルチスケールな抽出を行うため、Skip connection<sup>17)</sup>を用いる。Skip connectionとは入力された特徴マップを出力前に畳み込みの処理結果に付加する処理手順のことである。これにより畳み込み処理によって失われるノイズの位置情報を保持しつつ、縫合糸の形状などの大局的な特徴マップを同時に抽出することが可能になる。RDBではサイズ3×3×64の3次元形状のカーネルによる畳み込み処理“Conv<sub>3,1~3,3</sub>”が行われる。この時、畳み込み前後で特徴マップの数を減らし、非線形処理であるRectified Linear Units (ReLU)<sup>18)</sup>を用いること

で、複数の特徴マップに現れる入力画像の主成分を抽出する。これにより臓器表面の凹凸などの複雑な形状を抽出されるが、各畳み込み処理によって得られる計 160 枚の特徴マップは取りうる値の範囲にバラつきがある。そこで“Conv<sub>3,4</sub>”ではサイズ 1×1×160 のカーネルを 64 個畳み込むことで、160 枚の特徴マップを画素単位で均一化する。

### 3.4 高精細化の性能

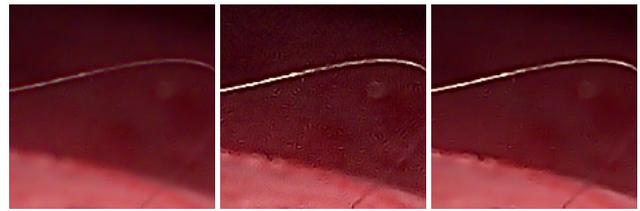
提案手法はノイズ低減及び高精細化処理の 2 段階に分かれている。このためノイズ低減後の中間画像に対し従来の鮮鋭化フィルタを施すことも可能である。そこで提案手法と従来手法による高精細化の違いについて説明する。まず、両手法により 8K 内視鏡映像の中間画像を高精細化した結果を図 9 に示す。(a)はノイズ低減後の中間画像、(b)は従来手法、(c)は提案手法による処理結果である。ここでは従来手法として学習用データセット作成時と同じ鮮鋭化フィルタを適用している。両者を比較すると縫合糸は同程度に強調されていることが分かる。しかし、臓器の表面に注目すると、従来手法の方が提案手法よりもアーティファクトが多く見られる。この違いは提案手法がノイズ低減と高精細化を同時に学習しているためである。2 つのタスクを独立させず同時に学習させることで、提案手法は従来手法よりもノイズ増幅が少ない高精細化処理を実現している。

## 4. 評価実験

提案手法の性能を定量化するため、ノイズ低減および高精細化の効果を評価する必要がある。しかし、評価基準となる縫合糸にフォーカスが合ったノイズの無い 8K 内視鏡映像が存在しないため、画質を PSNR などの数値で単純に評価することは困難である。そこで本研究では主観評価実験により縫合糸の視認性とノイズによる画質劣化について評価を行った。

### 4.1 設定

提案したニューラルネットワークを学習させるため、本実験では最適化手法 Adam<sup>19)</sup>を用い、学習率 $1 \times 10^{-4}$ 、バッチサイズ 16 で式(6)の損失値が小さくなるよう 1000 回繰り返し計算した。また、少量データによる過学習を防ぐため、学習データに対し Data augmentation を行った。まず、1 回の学習ごとにデータセット内の画像からサイズ 144×144 のパッチ画像をランダムに抽出する。次に、抽出したパッチ画像を水平/垂直方向にそれぞれ 50%の確率で反転させる。これにより学習を繰り返す度に 1 枚のパッチ画像から 4 パターンの画像が確率的に生成される。水増ししたパッチ画像を用いて学習したニューラルネットワークに 8K 内視鏡映像を入力し、高精細化した出力映像を主観評価実験に用いた。この 8K 内視鏡映像は焦点距離 35mm、対角視野 52° の 8K カメラ<sup>20)</sup>を使用し、約 1m の距離から手術中のブタの内臓を撮影した映像である。使用した 5 つの実験映像を図 10 に示す。ただし、



(a)  $I_{middle}$  image (b) Conventional method (c) Proposed method

図 9 高精細化処理の結果

Fig.9 The results of image enhancement

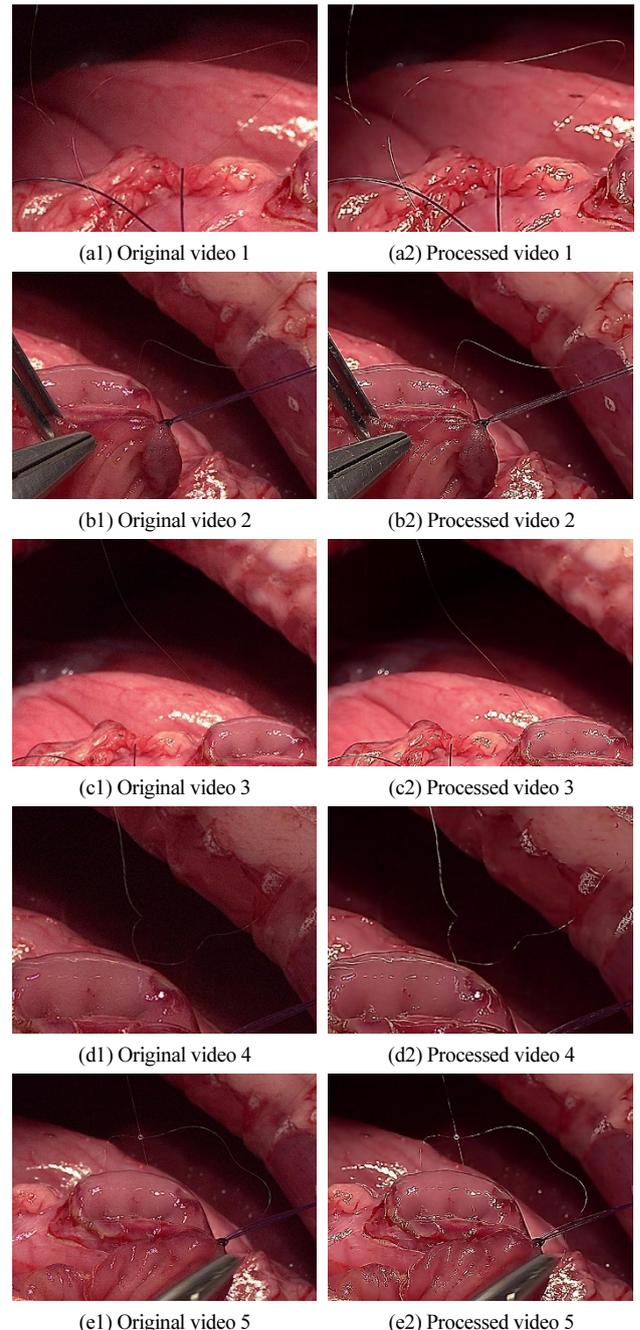


図 10 8K 内視鏡映像および処理結果

Fig.10 8K endoscopic videos and processed videos

図 10(a1)~(e1)は 8K 内視鏡映像から縫合糸が映る領域を 5 秒間切り抜いた結果である。図 10(a2)~(e2)は処理映像から同領域を切りぬいた結果である。

## 4.2 視認時間計測実験

### 4.2.1 実験条件

手術中は出血を避けるため血管が存在する臓器表面に 8K 内視鏡カメラのフォーカスを合わせることが多い。このためフォーカスがずれた縫合糸は視認性が悪化する。提案手法による縫合糸の視認性向上を検証するため、処理前後の映像に対して縫合糸を発見するまでの時間を計測した。

実験は非専門家である 10 人の被験者に対して行われ、55 インチの 4K ディスプレイに表示された映像から縫合糸を見つける作業を行う。被験者とディスプレイとの距離はディスプレイの高さの 1.5 倍離れた位置 (約 1m) となるように調整した<sup>21)</sup>。使用した実験映像の冒頭には 5 秒のカウントダウン映像が挿入され、その後 5 秒の手術映像が 2 回繰り返される。実験では被験者ごとに映像をランダムに提示することで処理前後での偏りを無くした。また、非専門家である被験者間の個人差が小さくなるよう、事前にサンプル映像を使用して縫合糸に関する説明を行った。加えて、被験者には縫合糸全体の形状が把握できるまで作業を行うよう指示することで、計測結果から縫合糸の動きによる反射的な視認を除いた。

### 4.2.2 実験結果

5 つの実験映像について縫合糸を視認するまでの時間を計測した結果を表 1~表 5 に示す。表の各行は被験者ごとの計測結果を表し、2,3 列目は原映像及び処理映像において縫合糸を視認するまでに要した秒数を表す。4 列目は、2,3 列目の差分値を表し、値が負であると処理映像の方が縫合糸を視認するまでの時間が短いことを意味する。実験の結果、Video 2 では 1 人の被験者が原映像の方が処理映像よりも早く縫合糸を発見したが、10 人中 9 人の被験者は処理映像の方が視認するまでに要した時間が短くなった。上記以外の実験映像では全被験者 10 人において処理映像の方が原映像よりも早く縫合糸を発見していた。全実験映像について平均した結果、処理映像の方が原映像よりも約 1.51 秒早く縫合糸を視認していた。

## 4.3 画質評価実験

### 4.3.1 実験条件

高精細化処理にはノイズの増幅が伴う。提案手法ではノイズ低減を高精細化処理の直前で行うことで画質の劣化を抑制する。提案手法が画質にどの程度影響するか検証するため、映像の画質を 5 段階で評価した。実験は非専門家である 10 人の被験者に対して行われ、55 インチの 4K ディスプレイに交互に表示される処理前後の映像のノイズを比較する。このとき被験者は前節の実験と同様に、非専門家である被験者間

表 1 Video 1 の視認時間

Table 1 Recognition time of video 1

被験者	時間[s] ~原映像~	時間[s] ~処理映像~	時間[s] ~差分~
A	3.27	1.90	-1.37
B	2.17	1.47	-0.70
C	4.12	2.30	-1.82
D	3.53	2.63	-0.90
E	6.55	3.33	-3.22
F	6.83	3.30	-3.53
G	2.00	1.25	-0.75
H	1.88	1.78	-0.10
I	3.72	2.68	-1.04
J	6.25	3.07	-3.18

表 2 Video 2 の視認時間

Table 2 Recognition time of video 2

被験者	時間[s] ~原映像~	時間[s] ~処理映像~	時間[s] ~差分~
A	4.72	2.07	-2.65
B	3.35	1.63	-1.72
C	3.88	3.53	-0.35
D	4.87	4.60	-0.27
E	10.00	4.78	-5.22
F	9.27	4.55	-4.72
G	6.12	1.95	-4.17
H	5.60	1.58	-4.02
I	4.55	4.73	0.18
J	9.77	5.25	-4.52

表 3 Video 3 の視認時間

Table 3 Recognition time of video 3

被験者	時間[s] ~原映像~	時間[s] ~処理映像~	時間[s] ~差分~
A	3.78	2.78	-1.00
B	3.38	3.00	-0.38
C	4.07	3.22	-0.85
D	3.98	2.22	-1.76
E	4.42	3.45	-0.97
F	4.12	3.75	-0.37
G	2.62	1.78	-0.84
H	2.42	1.30	-1.12
I	4.20	2.35	-1.85
J	4.65	3.17	-1.48

の個人差が小さくなるよう、事前にサンプル映像を使用して映像のノイズに関する説明を行った。

### 4.3.2 実験結果

5 つの実験映像におけるノイズについて比較した結果を表 6 に示す。表の各行は実験映像ごとの評価結果を表し、2~6 列目の数値はノイズによる画質劣化を 5 段階で評価した時の被験者数である。これは右の列ほど処理映像の方が原映像よりもノイズが少ないと評価したことを意味する。実験の結果、

Video 3, 5 では2人の被験者が原映像の画質の方が良いと評価したが、8名以上の被験者は同等かそれ以上にノイズが少ないと処理映像を評価した。上記以外の実験映像では全被験者10人が処理映像に画質劣化を感じないと評価した。また、50試行中47回と94%の割合で処理映像に画質劣化を感じないと評価した。

## 5. 考察

評価実験では5秒の手術映像を2回繰り返し表示している。これは被験者によって縫合糸を視認するまでの時間に個人差があり、1回の提示では縫合糸を発見出来なかったためである。本研究では縫合糸の視認性に個人差があることを考慮し、計測結果に対してノンパラメトリックな手法であるウィルコクソンの符号順位検定<sup>2)</sup>を行った。その結果、1%の有意水準で処理映像の方が原映像よりも縫合糸を視認しやすいことが示された。

一方で、提案手法の処理前後の8K内視鏡映像について画質を評価した結果、76%の被験者はノイズが少なくなったと評価した。さらに、画質が変わらなかったと評価した被験者を含めた場合、比率は94%となる。この比率はノイズによる画質劣化を感じていなかった被験者の割合を意味し、提案手法による高精細化にはノイズを増幅しないといつて良い。以上から、提案手法はノイズを低減することで画質劣化を抑制しつつ、フォーカス合わせが不十分な8K内視鏡映像を高精細化可能であると考えられる。

提案手法は執刀医が手術中に注視する4K解像度の拡大映像を処理対象としている。GPUにNvidia Titan Xpを使用した場合、提案手法は1フレームあたり約18ミリ秒の処理時間を要した。これは56fpsで処理可能なことを意味する。一方、カイロス株式会社によって開発された8K内視鏡カメラは60fpsで映像を出力する性能を持つ<sup>23)</sup>。提案手法の処理能力では60fpsでのリアルタイム処理を行うには僅かに不足だが、ハードウェアレベルでの最適化を行うことでリアルタイム処理が十分に期待できる。今後は処理速度の向上を目指し、レイテンシの低いField Programmable Gate Array (FPGA)により提案手法を実装する必要があると考える。

8K内視鏡映像にはガウシアンノイズとは特徴が異なるノイズが発生していた。これには撮像素子そのものが発生させる熱が原因の一つとして考えられる。8K内視鏡カメラの撮像素子は従来のHD/4Kカメラに比べサイズが小さい。このため熱による影響を強く受け、暗電流が多く発生してしまった可能性がある。暗電流は斑点模様の固定パターンノイズを発生させる要因であり、8K内視鏡映像に発生していたノイズの特徴と一致する。また他の要因として、解像度が向上したことで従来では撮影不可能であった臓器表面上の微細な凹凸を捉えた可能性がある。今後はノイズの発生原因を明らかにし、実用化に向けて8K内視鏡の特性を調査する必要がある。

表4 Video 4の視認時間

Table 4 Recognition time of video 4

被験者	時間[s] ～原映像～	時間[s] ～処理映像～	時間[s] ～差分～
A	3.12	2.27	-0.85
B	2.95	2.48	-0.47
C	3.90	2.70	-1.20
D	3.82	3.43	-0.39
E	4.42	2.92	-1.50
F	4.53	3.15	-1.38
G	2.62	1.90	-0.72
H	2.50	1.33	-1.17
I	3.68	2.83	-0.85
J	4.28	3.20	-1.08

表5 Video 5の視認時間

Table 5 Recognition time of video 5

被験者	時間[s] ～原映像～	時間[s] ～処理映像～	時間[s] ～差分～
A	3.40	2.18	-1.22
B	2.22	1.40	-0.82
C	4.37	2.37	-2.00
D	4.65	3.10	-1.55
E	5.38	4.88	-0.50
F	5.63	5.18	-0.45
G	2.17	1.62	-0.55
H	3.23	1.08	-2.15
I	4.77	3.35	-1.42
J	5.88	5.22	-0.66

表6 画質評価の結果

Table 6 The result of image quality evaluation

映像	Bad	Poor	Fair	Good	Excellent
Video 1	0	0	1	5	4
Video 2	0	0	2	3	5
Video 3	0	1	2	4	3
Video 4	0	0	3	3	4
Video 5	0	2	1	2	5

## 6. 結論

本研究は深層学習を基にフォーカスが不十分な8K内視鏡映像を高精細化する手法を提案した。高精細化に伴い増幅されるノイズを抑制するため、提案手法にはノイズ低減処理と高精細化処理を組み合わせた。実験の結果、処理前の8K内視鏡映像では視認困難であった直径0.02mm以下の縫合糸が鮮明となり、1%の有意水準で発見するまでの時間が33%程度短縮したことが示された。さらに、処理後の映像はノイズによる画質劣化がほとんど感じられないことが示された。今後はリアルタイム処理を目指し、提案手法をハードウェア実装することが課題である。

## 謝辞

本研究で使用した8K内視鏡映像をご提供いただいたカイ

ロス株式会社の山下紘正様, 映像出力機器をご提供いただいた計測技研株式会社の田端宏至様に感謝申し上げます。

### 参考文献

- 1) 須藤 剛, 佐藤 敏彦: “高齢者大腸疾患における腹腔鏡補助下手術の検討”, 老年医学会雑誌, Vol. 48, No. 6, pp. 665-671 (2011).
- 2) 近藤 大祐, 飯塚 博幸, 安藤 英由樹, 小濱 和貴, 坂井 義治, 前田 太郎: “腹腔鏡手術トレーニングにおける視野共有手法による学習効果とその実証”, 日本バーチャルリアリティ学会論文誌, Vol. 18, No. 4, pp. 487-496 (2013).
- 3) H. Yamashita, H. Aoki, K. Tanioka, T. Mori, T. Chiba: “Ultra-High Definition (8K UHD) Endoscope: Our First Clinical Success”, SpringerPlus, Vol. 5 (2016).
- 4) J. Gu, H. Lu, W. Zuo, C. Dong: “Blind Super-Resolution With Iterative Kernel Correction”, Proc. of Conference on Computer Vision and Pattern Recognition (2019).
- 5) V. Bannore: “Iterative-Interpolation Super-Resolution Image Reconstruction: A Computationally Efficient Technique”, Studies in Computational Intelligence Springer (2009).
- 6) T. Ida, N. Matsumoto, K. Isogawa: “画像の自己合同性を利用した再構成型超解像”, 情報処理学会研究報告, Vol. 107, No. 380, pp. 135-140 (2007).
- 7) S. Gohshi, S. Nakamura, H. Tabata: “Development of Real-time HDTV-to-8K TV Upconverter”, Proc. of Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (2017).
- 8) S. Gohshi, C. Mori, K. Tanioka, H. Yamashita: “Real-time Nonlinear Signal Processing Super Resolution of 8K Endoscope Cameras”, Proc. of Conference on Signal Processing and Multimedia Applications (2019).
- 9) A. Kubota, S. Gohshi: “Noise Reduction for 8K Endoscope Images”, Proc. of International Workshop on Advanced Image Technology (2019).
- 10) Y. Lecun, L. Bottou, Y. Bengio, P. Haffner: “Gradient-Based Learning Applied to Document Recognition”, IEEE, Vol. 86, pp. 2278-2324 (1998).
- 11) 高松 淳: “画像化プロセスと画像ノイズ”, 情報処理学会研究報告, Vol. 2010CVIM174, No. 17, pp. 1-8 (2010).
- 12) K. Perlin: “An Image Synthesizer”, Proc. of ACM SIGGRAPH Computer Graphics, Vol. 19, No. 3, pp. 287-296 (1985).
- 13) E. Agustsson R. Timofte: “NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study”, Proc. of Conference on Computer Vision and Pattern Recognition (2017).
- 14) ITU-R: “BT.709: Parameter Values for the HDTV Standards for Production and International Programme Exchange”, <https://www.itu.int/rec/R-REC-BT.709-6-201506-1/en> (2015).
- 15) B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee: “Enhanced Deep Residual Networks for Single Image Super-Resolution”, Proc. of Conference on Computer Vision and Pattern Recognition (2017).
- 16) Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu: “Residual Dense Network for Image Super-Resolution”, Proc. of Conference on Computer Vision and Pattern Recognition (2018).
- 17) J. Long, E. Shelhamer, T. Darrell: “Fully Convolutional Networks for Semantic Segmentation”, Proc. of Conference on Computer Vision and Pattern Recognition, pp. 3431-3440 (2015).
- 18) V. Nair, G.E. Hinton: “Rectified Linear Units Improve Restricted Boltzmann Machines”, Proc. of 27th International Conference on Machine Learning (2010).
- 19) Diederik P. Kingma, Jimmy L. Ba: “Adam: A Method for Stochastic Optimization”, Proc. of International Conference on Learning Representations (2015).
- 20) 船津 良平, 山下 誉行, 添野 拓司, 柳 忠明, 高橋 雄大, 吉田 哲男: “スーパーハイビジョン用小型カメラヘッドの開発”, 映像情報メディア学会誌, Vol. 68, No. 3, pp. J117-J124 (2014).
- 21) 川嶋 喜美子, 岡本 淳: “4K 映像の品質評価技術”, NTT 技術ジャーナル, Vol. 26, No. 1, pp. 68-71 (2014).
- 22) Wilcoxon Signed-Ranks Test Real Statistics Using Excel, <http://www.real-statistics.com/non-parametric-tests/wilcoxon-signed-ranks-test/> (2020).
- 23) カイロス株式会社, <http://kairos-8k.co.jp/product/scopeeight.html> (2020).

(2020年2月27日 受付)  
(2020年7月3日 再受付)



和田 歩 (正会員)

2018年工学院大学情報学部情報デザイン学科卒業。2018年より、同大学院工学研究科情報学専攻 修士課程に在籍。主として超解像技術の研究に従事。現在、(株)日立製作所所属。



大和 淳司

1988年東京大学工学部精密機械工学科卒業。1990年同大学院工学系研究科精密機械工学専攻修士課程修了。同年日本電信電話(株)入社。1998年MIT電気工学及びコンピュータ科学科修士課程修了。NTTヒューマンインタフェース研究所, NTTコミュニケーション科学基礎研究所等を経て、2016年工学院大学情報学部教授。博士(工学)。



合志 清一 (正会員)

1981年早稲田大学博士前期課程(修士)修了。同年日本放送協会(NHK)入局。主としてNHK放送技術研究所において、映像の伝送, 符号化, 映像処理, 番組応用, セキュリティの研究に従事。2008年シャープ(株)ディスプレイシステム研究所。2011年より、工学院大学情報学部教授。工学博士。主として映像処理及びセキュリティの研究に従事。2019年9月逝去。

画像電子学会誌

第 49 卷 第 4 号 (通巻 254 号)

2020(令和2)年 10 月 30 日発行 (年 4 回発行) 〒116-0002 東京都荒川区荒川 3-35-4 ライオンズマンション三河島第二 101 号

©2020 画像電子学会

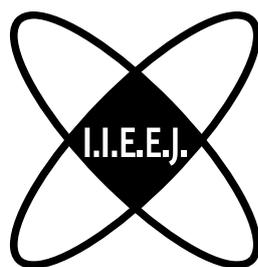
E-mail : hensyu@iieej.org

<http://www.iieej.org/>

発行所 一般社団法人 画像電子学会

TEL (03) 5615-2893 FAX (03) 5615-2894 (振替 00180-3-166232)

編集・発行者 児玉 明



ビジュアル  
コンピューティング論文特集号

第49巻  
第4号