Vol.2021-CVIM-225 No.13
2021/3/4

擬似カメラ画像の内部補完学習を用いた拡張内視鏡画像の生成

新田 潤平[1]　中尾 恵[1]　松田 哲也[1]

概要：内視鏡手術では視野の狭さから関心領域の三次元構造を捉えることは容易ではなく，手術支援のための画像生成技術が広く研究されている．本研究では胸腔鏡下肺がん切除術を対象に，機械学習に基づく画像補完によって内視鏡カメラ画像と 3D-CT モデルの位置合わせを達成し，臓器内部の腫瘍位置を可視化した拡張内視鏡画像を生成する手法を提案する．学習データが少ない問題に対して，臓器変形の統計的変位モデルから生成した擬似カメラ画像を用いて腫瘍位置の補完を学習した．さらに学習済みの画像補完モデルを内視鏡画像に適用する枠組みを提案し，臓器の姿勢推定を行うことなく位置合わせが達成された拡張現実画像を生成したので，結果を報告する．

キーワード：機械学習，画像補完，2D-3D 位置合わせ，拡張現実感

Augmented Endoscopic Image Generation by Image Completion Learning for Virtual Images

Jumpei Nitta[1]　Megumi Nakao[1]　Tetsuya Matsuda[1]

Abstract: In endoscopic surgery, it is not easy to capture the three-dimensional structure of the target region due to the narrow field of view, and image generation techniques for surgical guidance have been widely studied. In this study, we propose a method to generate an augmented endoscopic image which visualizes the tumor position inside the organ. We achieved 2D-3D registration between endoscopic images and 3D-CT models by machine learning-based image completion for thoracoscopic lung cancer resection. Since there was few training data, we trained image completion model using virtual images generated from a statistical model of organ deformation. In addition, we proposed a framework to apply the learned image completion model to endoscopic images and generated augmented reality images in which 2D-3D registration was achieved without organ pose estimation.

Keywords: Machine learning, Image completion, 2D-3D registration, Augmented reality

1. はじめに

近年医療技術および手術を支援する機器などの進歩により，内視鏡を用いた手術が広く行われるようになっている [1][2]．内視鏡手術ではカメラを体内に挿入し，モニタを通して手術を行うため安全性向上の観点から関心領域の三次元構造の把握が必要である．しかし，内視鏡の狭い術野から臓器の三次元構造を捉えることは容易ではなく，これを支援する研究が進められている [3][4]．例えば肝臓は術中の体積変化が小さい臓器であり，術前に撮影した Computed Tomography (CT) 画像から作成した臓器の腫瘍や血管構造を可視化した 3 次元 CT 像を手術のガイドとして腹腔鏡画像に重畳する研究が報告されている [5]．一方で術中の変形量が大きい臓器は，CT 画像から生成された臓器モデルと術中の臓器形状に差が大きく，肝臓のように単純に 3 次元 CT 像を重ね合わせることはできない．特に肺は術中に胸郭内部の圧変化に伴って肺実質が大きく虚脱し，その変形量も 50% 以上と大きいため [6]，腫瘍の位

[1] 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

置や血管構造の同定が課題となっている．

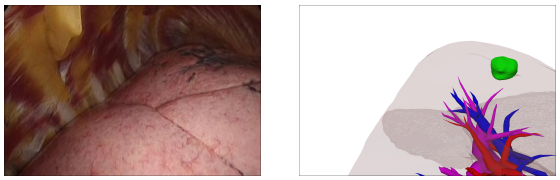
肺内部を可視化した 3 次元 CT 像を作成するためには，内視鏡画像内の肺に対して変形推定と姿勢推定を同時に行うことが必要であり，これは内視鏡画像と手術中に生じうる臓器の変形を表現した 3 次元 CT モデルとの 2D-3D 位置合わせに相当する．従来研究において，内視鏡画像における臓器形状に基づいた変形推定が可能であることが示されているが [5]，画像中に視認できる臓器が一部のみの場合，安定な位置合わせを達成することは難しい．肺のように一定範囲で変形する可能性がある物体を対象とした 2D-3D の変形・姿勢推定に関する研究も報告されているが，シミュレーションモデルでの検証に留まっている [7][8]．機械学習を用いて 2D-3D 位置合わせを実現する研究も報告されているが [9]，機械学習には大量のデータが必要であり医用画像においてはデータが十分に用意できないという問題がある．内視鏡画像中の肺と肺の 3 次元 CT モデルとの 2D-3D 位置合わせには，姿勢推定が課題であるが肺が一部しか写っていないことから難しく，機械学習を用いる場合においてもデータが十分でないという問題がある．特に機械学習による位置合わせにおいて腫瘍位置の可視化を実現する場合，ある肺に対応した腫瘍位置を学習する必要があるため，複数の肺データをまとめて学習に用いることが困難でありデータ数を増やせない大きな要因の一つとなっている．

本研究では，機械学習に基づく画像補完によって内視鏡画像と 3 次元 CT モデルの位置合わせを達成し，臓器内部の腫瘍や血管構造を可視化した拡張内視鏡画像を生成する手法を提案する．学習データとして，臓器変形のバリエーションが表現された統計的変位モデル [10][12] から生成した多数の擬似カメラ画像を用いることで，手術時に視認される臓器外形に対する臓器内部の腫瘍及び血管構造の補完を学習する．擬似カメラ画像と内視鏡画像は見た目が異なり，学習済みの画像補完モデルをそのまま適用することはできないため，本研究では内視鏡画像中の肺領域を抽出したラベル画像を用いることで入力画像の形式を統一して適用を可能にする枠組みを提案する．胸腔鏡下肺がん切除術を受けた 2 名の患者の 3 次元 CT モデルと内視鏡画像を用いて，提案方法による拡張内視鏡画像の生成を試みる．提案方法の有効性を検証するために，疑似カメラ画像の真値と生成画像間で評価値を算出する．また，画像補完モデルを内視鏡画像に適用して得られた生成画像に対しては，手動で位置合わせを行って得られた 3 次元 CT 像との間で評価値を算出する．

2. 提案手法

2.1 対象データ

本研究では京都大学医学部附属病院呼吸器外科から提供を受けた胸腔鏡下肺がん切除術の手術動画 2 例から，術具が含まれていないシーンをそれぞれ 6 シーン抽出した静止画像を用いる．これらの静止画像を内視鏡画像 I_r とよぶ．図 1(a) に I_r の例を示す．2 例の手術動画をそれぞれ case 1, 2 と呼称し，I_r は手術動画のうち肺の写り方がなるべく異なっているものを抽出した．



(a) 内視鏡画像 I_r　　(b) 3 次元像 I_v

図 1 内視鏡画像 I_r と 3 次元像 I_v の例

肺の内部構造情報としては，各症例の 3 次元 CT モデル M^I を用いる．これは手術前に撮像された 3 次元 CT 画像から三角形表面メッシュによって構成されたもので，肺葉，気管支，動脈，静脈及び腫瘍の情報を保持している．しかし，肺は術中に大きく虚脱し変形するため，M^I では直接的に術中の腫瘍位置や血管構造を表すことはできない．

また手術時の内視鏡カメラの位置と方向，焦点距離，画角を反映させて 3 次元 CT モデル M^I を透視投影によりレンダリングした画像を 3 次元像 I_v とよぶ．図 1(b) に I_v の例を示す．図 1(b) において，緑が腫瘍，マゼンタが気管支，青と赤がそれぞれ静脈，動脈を示している．I_v は肺の実質部分を半透明とし，内部構造を可視化した画像であり，内視鏡画像に対応した視点で肺を観察した際の内部構造を表す画像として用いる．なお，case 1 は腫瘍が 1 つの症例であり，case 2 は腫瘍が 2 つの症例である．

2.2 問題設定

拡張内視鏡画像は内視鏡画像中の肺の見た目はそのままに，その内部構造を重畳可視化した画像であり，直接腫瘍位置を確認できるようにしたものである．こうした画像の作成には内視鏡画像に位置合わせされた患者個人の 3 次元 CT モデルが必要であり，手術のガイドとして用いるためにはこれを実時間で算出する必要がある．

本研究では機械学習に基づく内視鏡画像の内部補完によって，肺の内部構造を重畳可視化した拡張内視鏡画像を生成することを目的とする．機械学習を用いることで内視鏡画像と 3 次元 CT モデルとの 2D-3D 位置合わせを行うことなく，内視鏡画像に位置合わせされた 3 次元 CT 像を直接的に得る．本研究の提案フレームワークを図 2 に示す．提案フレームワークは以下の 3 ステップに分けられる．

STEP 1 内視鏡画像 I_r に対する Semantic segmentation[11] により肺領域のラベル画像 L_r を得る

STEP 2 機械学習によって得られた画像補完モデル G を用いて，L_r に対応する患者個人の腫瘍や血管構造を内部補完し 3 次元像 I_v を得る

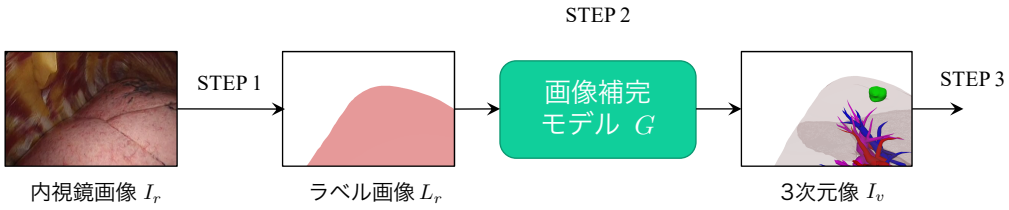


図 2 提案フレームワーク

STEP 3 得られた I_v と I_r をブレンディングし，拡張内視鏡画像 I_a を得る

提案フレームワークの要点は画像補完モデルによって位置合わせ済みの 3 次元像を生成することである．画像補完モデル G は機械学習を用いて構築するが，その際学習データが少ないという問題がある．ラベル画像の内部補完を行う G の学習にはラベル画像とそれに対応した 3 次元像が必要となる．しかし，内視鏡画像と 3 次元像のペアデータは得られておらず，3 次元 CT モデルを元にこれらを手動で作成することも可能ではあるが，学習に十分な数を用意するのは困難である．

そこで本研究では肺の虚脱変形のバリエーションを表現した統計的変位モデル [10][12] を用いて手術時に想定される肺形状を推定し，推定された 3 次元 CT モデルを多数の視点からキャプチャすることにより 3 次元像 I_v を生成する．同時に，生成した I_v に対するラベル画像 L_v も生成し，これらを学習データとすることでデータの少なさに対処する．これらの (L_v, I_v) をまとめて擬似カメラ画像とよび，(L_v, I_v) を用いて G の学習を行う．ここで，(L_v, I_v) を用いて学習を行うことで G の入力がラベル画像となり，I_r を直接入力して変換を行うことができないという問題が生じる．そこで入力画像形式を統一する目的で，提案フレームワークの STEP 1 において I_r から L_r を取得し，STEP 2 において L_r を G に入力して I_v を得る．このような流れで拡張内視鏡画像を生成することで，データが少ないという問題を解決しつつ学習済みモデルを内視鏡画像に適用することを可能とする．

2.3 統計的変位モデルを用いた擬似カメラ画像の生成

本節では，統計的変位モデルを用いた虚脱肺の 3 次元 CT モデルの生成方法および生成した 3 次元 CT モデルを用いた擬似カメラ画像の生成について述べる．2.1 節で説明した 3 次元 CT モデル M^I は含気時の肺形状を示すモデルであり，術中の肺の脱気変形のために虚脱時の形状とは大きく異なっている．そのため M^I から術中肺の内部構造を可視化した 3 次元像 I_v を作成することはできない．そこで統計的変位モデルを用いて虚脱肺形状を示す 3 次元 CT モデル M^D を生成する．

統計的変位モデルは，手術時の含気肺と虚脱肺を対象として撮像された Cone-beam CT 画像からメッシュ変形位置合わせ [10][13] によって頂点対応が取れた肺形状モデル群を得た後，カーネル法に基づいて頂点単位で変位を学習することによって得られる．患者個人の術前 CT モデルに適用することで，学習データが示す平均的な虚脱変形を表現することが可能である [12]．統計的変位モデルを用いた M^D の生成における頂点単位の変位は次式で表現される．

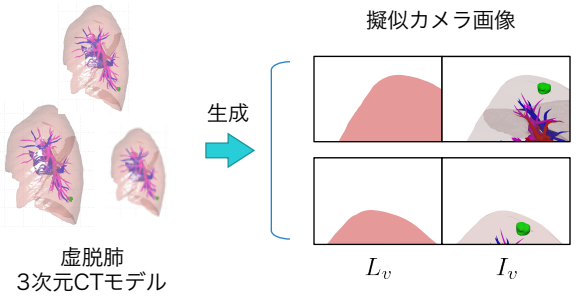


図 3 疑似カメラ画像 (L_v, I_v) 生成の様子

$$v_i' = v_i + w \cdot u_i. \tag{1}$$

ここで v_i' は M^D の頂点位置，v_i は M^I の頂点位置であり，u_i は統計的変位モデルによって求められた平均的な頂点の変位である．また w は平均的な変位からどれだけ虚脱させるかを調整する重みであり，$w < 1.0$ で平均変位よりもしぼませた形状が表現され，$w > 1.0$ で平均変位よりも膨らませた形状が表現される．本研究では w を変更することで手術時に想定される M^D を複数生成する．

次に統計的変位モデルから生成した M^D を用いた擬似カメラ画像 (L_v, I_v) の生成について説明する．概要を図 3 に示す．図 3 に示すように擬似カメラ画像は肺内部を可視化した I_v とその肺領域を示す L_v で構成されるもので，L_v は I_v の背景以外の部分をピンクとした画像である．(L_v, I_v) の生成では，あるカメラパラメータ θ で M^D をキャプチャして I_v を生成し，得られた I_v から L_v を作成する．I_v は (θ, w) を変化させることで多数のバリエーションが生成可能であり，図 3 では様々な (θ, w) による (L_v, I_v) 生成の様子を示している．

まず θ は，各 case に対して想定される内視鏡カメラの位置と注視点の範囲内でパラメータを変更し，多数の I_v の生成を行う．内視鏡手術ではすべての手技が体腔内で行われるため，手術時における内視鏡カメラの位置とその注視

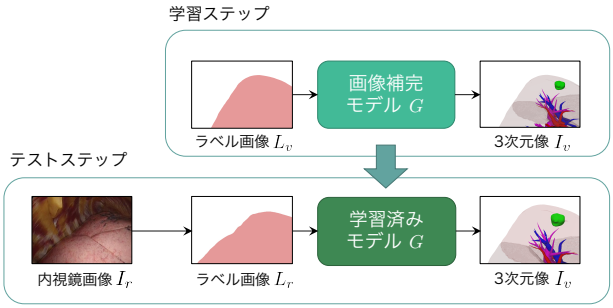


図 4　内部補完学習と内視鏡画像への適用の流れ

点は肺に対して一定の範囲内であると仮定でき，またその範囲は患者個人の 3 次元 CT モデルの座標系において事前に算出することが可能である．なおパラメータの変更間隔は，各 case 共通に内視鏡カメラの位置については 5mm 間隔で，注視点位置は 10mm 間隔で変更することとした．次に w は，変更することで異なる虚脱状態の M^D を生成し，これに対して θ を変更して I_v を生成する．手術時の肺の虚脱量には個人差があり，虚脱のバリエーションを学習するためにこれを行う．このとき w は 0.9 から 1.1 まで 0.1 刻みで変更することとした．

2.4　内部補完学習と内視鏡画像への適用

本研究では，内視鏡画像中の肺に対する変形推定や姿勢推定を行うことなく位置合わせが達成された 3 次元像の生成を目指しており，機械学習に基づく画像補完モデルを用いることでこれを達成する．ここで扱う画像補完は肺領域を示すラベル画像に対して肺の内部構造を補完するものであり，学習データとして用いるのはラベル画像 L_v とそれに対応する 3 次元像 I_v である．本研究では画像補完モデルとして GAN の拡張である pix2pix[14] を用いた．

pix2pix を用いてラベル画像の画像補完を学習することを内部補完学習とよぶ．疑似カメラ画像による内部補完学習および内視鏡画像への適用について概要を図 4 に示す．まず学習ステップでは疑似カメラ画像を学習データとして用いることで，ラベル画像から 3 次元像を生成する画像補完モデル G を学習する．次にテストステップでは学習済みの G を内視鏡画像 I_r に適用することで，内部構造を可視化した 3 次元像を生成する．このとき I_r から肺領域を抽出したラベル画像 L_r を取得し，L_r を G に入力することで学習済みモデルを内視鏡画像に適用することを可能とする．なお，本研究で目指す内部補完は患者個人の腫瘍位置や血管構造を正確に補完するものであるため，学習は症例ごとで別々に行う．このようにすることによって対象の患者の内部構造に特化した学習となり，より正確な補完が可能となると考えた．

次に内部補完学習について説明する．pix2pix による内部補完学習では，擬似カメラ画像中のラベル画像 L_v を条件画像，3 次元像 I_v を目標画像とする．学習を進めるために pix2pix で用いられている次の 2 つの損失関数を適用する．1 つ目の損失関数は GAN の基本的な損失関数である Adversarial loss を Conditional GAN[15] に拡張したもので，次の式 (2) で表される $\mathcal{L}_{\mathrm{cgan}}$ である．

$$\mathcal{L}_{\mathrm{cgan}}(G,D) = \mathbb{E}_{L_v,I_v}[\log D(L_v,I_v)] + \mathbb{E}_{L_v,z}[\log(1-D(L_v,G(L_v,z)))]. \quad (2)$$

ここで G は生成器，D は識別器，z はノイズベクトル，$G(L_v,z)$ は生成器に L_v を入力して得られた生成画像を示し，$D(L_v,I_v)$, $D(L_v,G(L_v,z))$ はそれぞれ (L_v,I_v), $(L_v,G(L_v,z))$ の正解ペアらしさを確率で表したものである．D は正解ペアと偽物ペアを正しく識別することを目的とするため，$D(L_v,G(L_v,z))$ が 0.0 に近づくことを目指すが，反対に G は 1.0 に近づくことを目指す．これによって G と D に敵対的関係が成り立つ．2 つ目は生成したい目標画像である 3 次元像 I_v と G によって生成した生成画像 $G(L_v,z)$ の L1 損失 $\mathcal{L}_{\mathrm{L1}}$ である．式を式 (3) に示す．

$$\mathcal{L}_{\mathrm{L1}}(G) = \mathbb{E}_{L_v,I_v,z}[\|I_v - G(L_v,z)\|_1]. \quad (3)$$

L1 損失は $G(L_v,z)$ が I_v に一致することを要請する損失関数である．これら 2 つの損失関数を合わせて pix2pix の最終的な目的関数は以下となる．

$$\min_G \max_D \mathcal{L}_{\mathrm{cgan}}(G,D) + \lambda \mathcal{L}_{\mathrm{L1}}(G). \quad (4)$$

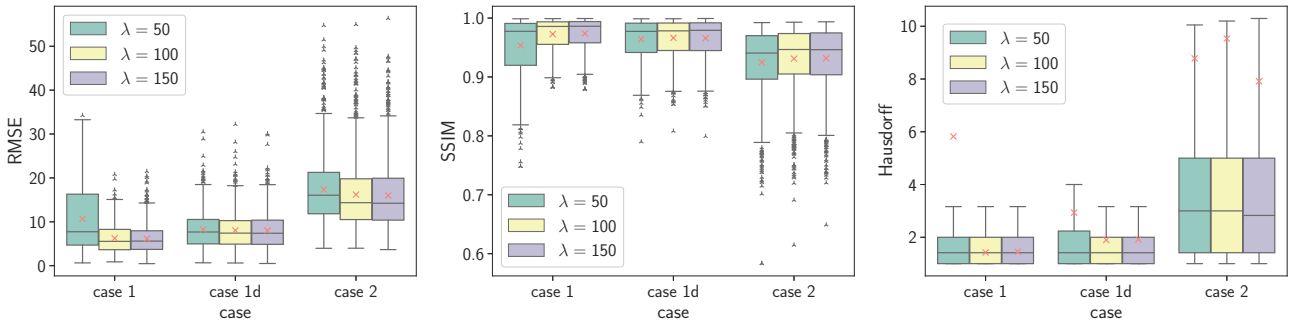
ここで λ は $\mathcal{L}_{\mathrm{L1}}$ の影響を制御するための重みである．提案する内部補完学習では式 (4) を解くことで L_v を I_v に変換する生成器 G を最適化し，学習済みの G を用いて肺の内部補完を行う．

3.　評価実験

3.1　擬似カメラ画像による学習

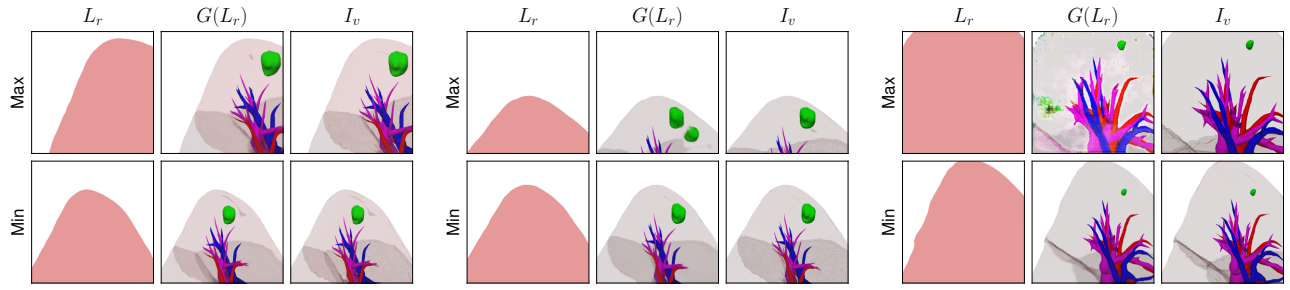
本研究では提案フレームワークの有効性を確認するために，胸腔鏡下肺がん切除術の手術動画から抽出した内視鏡画像およびそれに対応する患者個人の 3 次元 CT モデルを対象として，実験 1, 2 の 2 通りの評価実験を行った．本節では疑似カメラ画像の内部補完学習を複数の症例および条件下で行い，テストデータに対して肺内部の腫瘍及び血管構造の補完が可能であるかを確認する実験 1 について述べる．

実験 1 では 2.1 節で説明した case 1 と case 2 の症例および case 1 について虚脱バリエーションを表現した case 1d の 3 つのデータを用いた．これら 2 つの症例については，統計的変位モデルの重み w を 1.0 として生成した平均的な虚脱肺形状 M^D1 つを用いて疑似カメラ画像を生成し，学習に利用した．さらに本実験では，統計的変位モデルを用いた虚脱バリエーションの生成が画像補完に有効か否かを確認するため，case 1 に対して w を 0.9, 1.0, 1.1 として生



(a) RMSE　(b) SSIM　(c) ハウスドルフ距離（誤差 10 ピクセル以下）

図 5　テストデータに対する補完結果の評価値



(a) case 1　(b) case 1d　(c) case 2

図 6　ハウスドルフ距離最大及び最小の場合の生成画像例（実験 1）

表 1　学習およびテストで用いるデータ数

	case 1	case 1d	case 2
学習データ数	3000	3000	3000
テストデータ数	840	840	1032
総データ数	3840	11520	4032

成した 3 つの M^D を用いて疑似カメラ画像を生成し，内部補完学習を行った．これを case 1d とする．なお，疑似カメラ画像の生成において θ は 2.3 節に示した範囲で変化させた．

実験では疑似カメラ画像を学習データとテストデータに分割して学習およびテストを行った．表 1 にデータ数をまとめる．学習データの数による内部補完の精度に偏りが出ないようにするため，各 case における学習データの数は 3000 で統一した．なお case 1d については，3 通りの M^D それぞれから 3840 のデータを作成し，作成したデータから 1000 ずつのデータを抽出して合計で 3000 とし，テストデータについても case 1 と数を統一するため 280 ずつ抽出して合計で 840 とした．また実験で用いる画像は RGB 画像とし，解像度は学習時およびテスト時ともに 256×256 で統一した．

学習は各 case に対して，$\lambda = 50, 100, 150$ の 3 通りで行った．ここで λ は式 (4) における L1 損失の重みを決定する値であり，値を大きくすると生成画像が目標画像に近づく．λ 以外の学習に用いるパラメータは実験を通して共通とし，エポック数 1000，バッチサイズ 4，学習率 0.0002 を用いた．また学習の各反復において，pix2pix では学習データ読み込み時に画像の反転および切り取りを行うが，本研究ではこれを行わず学習した．

学習した画像補完モデル G の評価は，テストデータのラベル画像 L_v を G に入力し，得られた生成画像 $G(L_v)$ に対して目標画像である 3 次元像 I_v との間で RMSE，SSIM，および腫瘍位置に対するハウスドルフ距離の 3 つの評価値を算出して行った．生成画像において肺領域以外の背景部分は評価の対象としないため，評価値を算出する前に背景部分を白とする前処理を行った．評価値のうち RMSE と SSIM は $G(L_v)$ と I_v の画像全体の誤差を評価するために用いた．RMSE は 2 枚の画像間で同じ位置にあるピクセルの輝度値の差を 2 乗し，平均をとったものの平方根であり，値が小さいほど画像間の輝度値の差が少ないことを表している．SSIM は画像の輝度，コントラスト，構造の視覚的影響を評価する構造的類似性の指標であり，値が 1 に近いほど類似度が高く，0 に近いほど類似度が低いことを表す．

また $G(L_v)$ における内部構造の補完精度を評価するために，腫瘍位置に対するハウスドルフ距離を用いた．ハウスドルフ距離は 2 つの領域間のずれを定量する指標として用いられている．本研究で扱う I_v では腫瘍をすべて緑で描画しており，$G(L_v)$ と I_v それぞれから抽出した緑の領域間のハウスドルフ距離を算出することで，G によって補完された腫瘍位置と真値の腫瘍位置とのピクセル間距離誤差を評価した．

学習が完了したモデルを用いてテストデータの内部補完を行い，生成された画像に対して算出した評価値の箱

ひげ図を図 5 に示す．図 5 の各箱ひげ図では左から case 1，1d，2 の評価値を示しており，各 case の中では左から $\lambda = 50, 100, 150$ の評価値を示している．RMSE と SSIM については外れ値も含めた箱ひげ図を示しており，腫瘍位置のハウスドルフ距離については外れ値が極端に大きな値となるため，誤差 10 ピクセル以下のプロットに限定した箱ひげ図を示している．また図中の $\times$ は平均値を表している．図 6 に各 case の $\lambda = 100$ における生成画像のうち，腫瘍位置のハウスドルフ距離が最大のものと最小のものを示す．図では上段にハウスドルフ距離が最大のものを，下段に最小のものを示し，画像は左から入力した L_r，内部補完結果の $G(L_r)$，L_r に対する真値の I_v を示している．いずれの case でもハウスドルフ距離の最小値は 1.0 であり，正確な腫瘍位置を補完できていることが確認できた．

全体を通して $\lambda = 50$ の場合は，case 2 のハウスドルフ距離を除けば評価値はいずれも $\lambda = 100, 150$ に比べて悪かった．RMSE と SSIM に関しては多くの場合で $\lambda = 150$ の場合の方が $\lambda = 100$ の場合よりも良い値となったが，その差は僅かであり，$\lambda = 100$ の場合は RMSE，SSIM 外れ値の最大値が小さく，ばらつきが少ない．これは，$\lambda = 150$ の場合では $G(L_v)$ と I_v との輝度値の差を小さくする方向に過学習したことで，学習データに含まれない画像に対しての変換精度が低下した可能性がある．ハウスドルフ距離について，case 1 と case 1d の平均値は $\lambda = 100$ の場合がもっとも小さく，case 1 と case 1d を比較すると case 1d の方が補完精度が低い．これは疑似カメラ画像生成に用いた w が増えたことで学習データのバリエーションが増加したが，学習データの数は 3000 と一定のために十分な学習が行えなかったためと考えられる．case 2 では $\lambda = 150$ の場合が $\lambda = 100$ の場合よりもハウスドルフ距離が小さかったが，case 2 では case 1 と比較して画像中の大きな部分が肺領域であったものが多かったためであると考えられる．図 6(c) の Max が顕著な例であるが，画像中の大半の部分が肺領域である．こうしたラベル画像を元に学習した場合，テストデータに同様のラベル画像を入力しても区別がつきにくいため，正しい腫瘍位置の生成が困難である．しかしこのような場合は，肺を捉えている視点の区別がつかないため本提案手法の対象外であり，さらに評価値のばらつきは $\lambda = 100$ のほうがやや小さくなっている．また case 2 では $\lambda = 50$ の場合が $\lambda = 100$ の場合よりも平均値が小さかったが，RMSE および SSIM では $\lambda = 50$ における評価値は悪く，腫瘍以外の物体については正しく生成できない場合が多いと考えられる．以上の考察を踏まえて，疑似カメラ画像の内部補完学習においては $\lambda = 100$ を用いることにより，ある程度の汎化性能を維持したまま腫瘍位置を正しく補完できると考え，実験 2 ではこの値を用いることとした．

3.2 内視鏡画像への適用と手動位置合わせとの比較

本節では疑似カメラ画像の内部補完学習を行って得た画像補完モデルを内視鏡画像に適用することで 3 次元像を生成し，内視鏡画像に対する手動位置合わせによって作成した 3 次元像と比較する実験 2 を行った．実験 2 では 3.1 節の実験 1 と同様に case 1 と case 2 の症例および虚脱バリエーションを表現した case 1d の三つのデータを用いた．データ数は表 1 に示す通りである．内視鏡画像 I_r は case 1 と case 2 ともに 6 枚あり，case 1d は case 1 と同じ I_r を用いた．

3 つのデータに対する擬似カメラ画像の内部補完学習を実験 1 と同じ条件で行い，λ の値は実験 1 の考察にあるように $\lambda = 100$ とした．ただし学習データとテストデータの分割は行わず，すべてのデータを用いて学習を行った．次に，学習により得られた画像補完モデル G に I_r の肺領域を抽出したラベル画像 L_r を入力し 3 次元像 $G(L_r)$ を生成した．疑似カメラ画像中の輪郭線に類似した輪郭線を抽出するため，肺領域を示すラベル画像 L_r の作成は手動で行った．

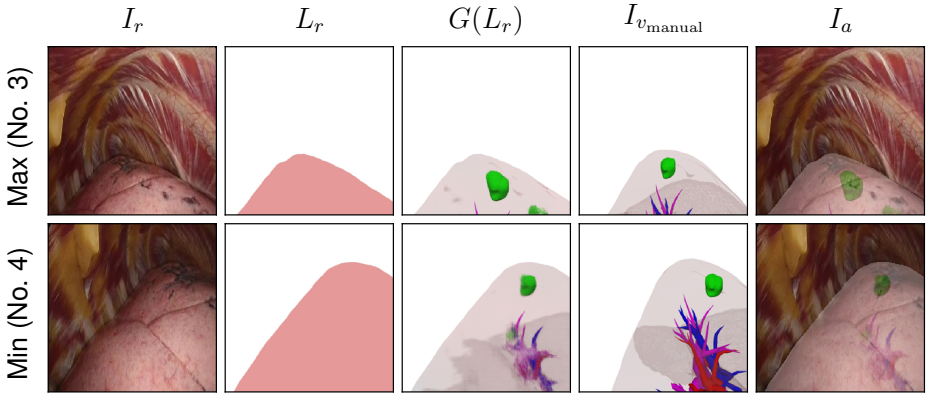
$G(L_r)$ の評価は，手動位置合わせ結果の 3 次元像 $I_{v_{\mathrm{manual}}}$ と比較することで行った．手動位置合わせとは虚脱肺の 3 次元 CT モデル M^D を手動で回転および平行移動させ，M^D をキャプチャする際のカメラパラメータを I_r と同じ視点から肺を捉える値に合わせることであり，これによって I_r に位置合わせされた 3 次元像を得る．評価は 3.1 節で説明した RMSE，SSIM および腫瘍位置のハウスドルフ距離を用いて行った．なお実験 1 と同様に，評価値算出の前に背景部分を白とする前処理を行った．

各 case ごとに学習した G を用いて L_r を変換し，生成した $G(L_r)$ と $I_{v_{\mathrm{manual}}}$ とで RMSE，SSIM，腫瘍位置のハウスドルフ距離を算出した結果を表 2 に示す．各 case の 6 枚の I_r から抽出した L_r に 1 から 6 の番号をつけて，表中の No. にその番号を示した．NaN と表記した画像については，生成した $G(L_r)$ 中に腫瘍が生成されなかったことで，ハウスドルフ距離を算出できなかったことを表している．なお，最終行には平均値を示した．表 2 の RMSE と SSIM より case 1 と case 2 を比較すると case 1 のほうがより $I_{v_{\mathrm{manual}}}$ に近い $G(L_r)$ を生成できていることがわかる．また case 1 と case 1d を比較するとわずかに case 1d のほうが評価値が良くなっていた．表 2 より，腫瘍位置のハウスドルフ距離が最小となったのは case 1 の No. 1 であり，case 2 の No. 1 で最大であった．

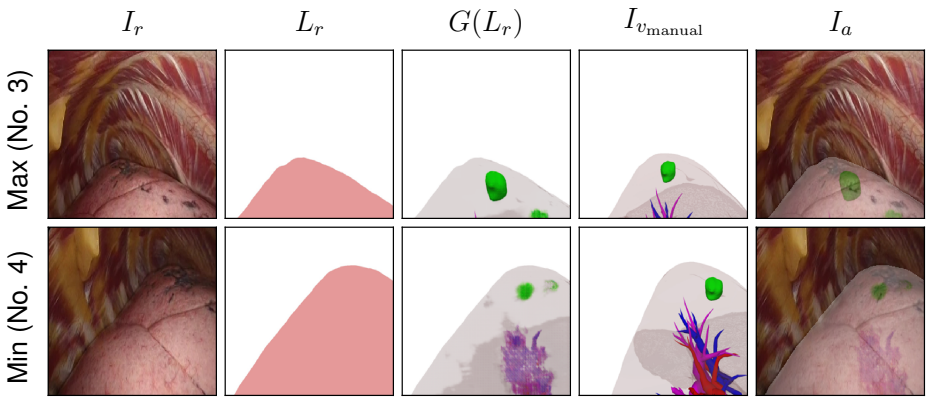
また図 7 に case 1，case 1d，case 2 の三つについてハウスドルフ距離が最大および最小になったものの生成画像例を示す．図中の画像は左から I_r, L_r, $G(L_r)$, $I_{v_{\mathrm{manual}}}$, I_a であり，上段がハウスドルフ距離最大，下段がハウスドルフ距離最小の例である．ここで I_a は I_r と $G(L_r)$ を次の式 (5) で $\alpha = 0.35$ としてアルファブレンディングを行い合成

表 2　3 次元像 $G(L_r)$ と手動位置合わせ結果 $I_{v_{\text{manual}}}$ との評価値

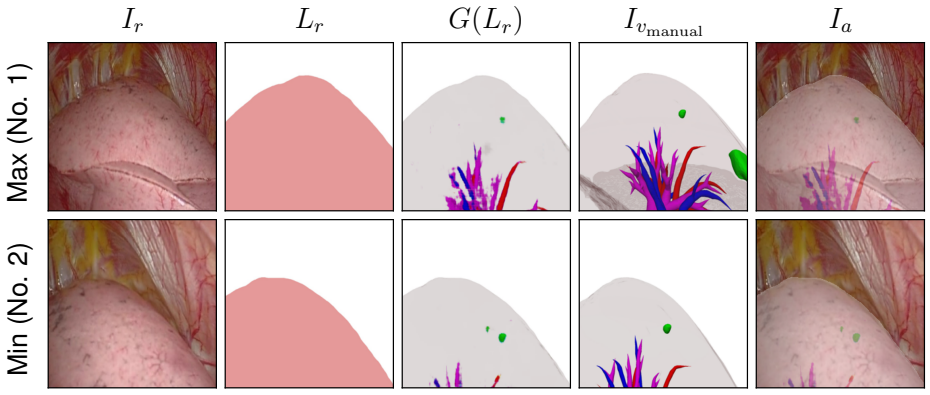
	RMSE			SSIM			ハウスドルフ距離		
No.	case 1	case 1d	case 2	case 1	case 1d	case 2	case 1	case 1d	case 2
1	36.13	35.31	53.36	0.807	0.801	0.777	15.23	19.42	134.62
2	25.13	22.28	28.38	0.933	0.935	0.916	63.82	52.35	22.20
3	28.59	28.22	49.77	0.910	0.912	0.788	98.86	98.86	85.00
4	24.24	24.65	39.66	0.943	0.943	0.878	27.86	28.18	91.59
5	38.86	34.58	57.44	0.824	0.834	0.722	38.95	36.06	NaN
6	39.47	38.39	61.55	0.788	0.789	0.711	47.01	48.08	NaN
Mean	32.07	30.57	48.36	0.867	0.869	0.799	48.62	47.16	83.35



(a) case 1



(b) case 1d



(c) case 2

図 7　ハウスドルフ距離最大及び最小の場合の生成画像例（実験 2）

した拡張内視鏡画像である．

$$I_a = (1-\alpha)\cdot I_r + \alpha \cdot G(L_r). \qquad (5)$$

実験結果より，疑似カメラ画像の内部補完学習によって得られた G を用いて，$I_{v_{\mathrm{manual}}}$ の腫瘍位置に近い位置に腫瘍を生成可能であることが確認できた．ハウスドルフ距離が最小であった case 1 の No. 1 では $I_{v_{\mathrm{manual}}}$ とおおよそ同じ位置に腫瘍を生成できている．3.1 節の実験結果より，G に入力するラベル画像が学習データ中のラベル画像に類似した画像であれば小さい誤差で 3 次元像を生成できると考えられる．したがって G によって生成された腫瘍位置が $I_{v_{\mathrm{manual}}}$ の腫瘍位置に近い場合は，入力したラベル画像の肺領域の輪郭が学習データに用いた疑似カメラ画像の肺領域の輪郭に類似しており，小さい誤差で $G(L_r)$ を生成できたと考えられる．一方で多くの場合では腫瘍を $I_{v_{\mathrm{manual}}}$ の腫瘍位置に近い位置に生成することはできなかった．図 7(a) 中の No. 3 および図 7(b) 中の No. 3 は腫瘍を 2 つ生成しているが，これはラベル画像の肺領域に類似した肺領域を示すラベル画像が学習データ中に 2 種類あり，それらの結果を混合して生成したと考えられる．また図 7(c) 中の No. 1 における $G(L_r)$ は小さい方の腫瘍しか生成できておらず，生成した腫瘍も不鮮明となっている．これは元の腫瘍が小さいことに起因すると考えられる．これらの問題を解決するためには学習データを充実させることがもっとも効果的であると考えられる．

4. おわりに

本研究では機械学習に基づく画像補完によって内視鏡画像と 3 次元 CT モデルの位置合わせを達成し，拡張内視鏡画像を生成する手法を提案した．2 症例の手術動画に対する内部補完学習を行った実験では，疑似カメラ画像を用いたテストにおいて高い精度での内部補完が可能であった．一方で学習したモデルを内視鏡画像に適用し，拡張内視鏡画像の生成を行った結果を手動位置合わせと比較した実験では，case 1d において腫瘍位置のハウスドルフ距離の平均値がもっとも小さく，47.16 ピクセルでの内部補完が可能であった．

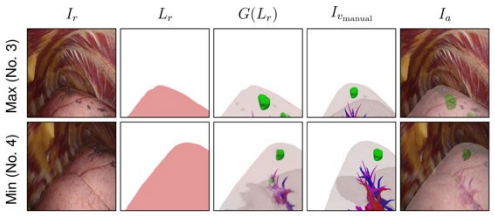
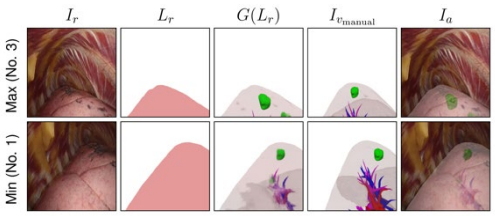
謝辞 本研究は科研費 挑戦的研究（萌芽）課題番号：18K19918 の助成による．患者個人の手術動画及び 3 次元 CT 画像を提供頂いた名古屋大学医学部附属病院呼吸器外科芳川豊史先生に感謝の意を表します．

参考文献

[1] Flores, R. M. and Alam, N.: Video-assisted thoracic surgery lobectomy (VATS), open thoracotomy, and the robot for lung cancer, *The Annals of thoracic surgery*, 85(2), S710-S715, (2008).
[2] Shaw, J. P., Dembitzer, F. R., Wisnivesky, J. P., Litle, V. R., Weiser, T. S., Yun, J., Chin, C. and Swanson, S. J.: Video-assisted thoracoscopic lobectomy: state of the art and future directions, *The Annals of thoracic surgery*, 85(2), S705-S709, (2008).
[3] Sato, M., Omasa, M., Chen, F., Sato, T., Sonobe, M., Bando, T. and Date, H.: Use of virtual assisted lung mapping (VAL-MAP), a bronchoscopic multispot dye-marking technique using virtual images, for precise navigation of thoracoscopic sublobar lung resection, *The Journal of thoracic and cardiovascular surgery*, Vol. 147, No. 6, pp. 1813–1819 (2014).
[4] Koo, B., Özgür, E., Le Roy, B., Buc, E. and Bartoli, A.: Deformable registration of a preoperative 3D liver volume to a laparoscopy image using contour and shading cues, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 326–334 (2017).
[5] Haouchine, N., Cotin, S., Peterlik, I., Dequidt, J., Lopez, M. S., Kerrien, E. and Berger, M. O.: Impact of soft tissue heterogeneity on augmented reality for liver surgery, *IEEE transactions on visualization and computer graphics*, Vol. 21, No. 5, pp. 584–597 (2014).
[6] Nakao, M., Tokuno, J., Chen-Yoshikawa, T. F., Date, H. and Matsuda, T.: Surface deformation analysis of collapsed lungs using model-based shape matching, *Int. J. Computer Assisted Radiology and Surgery*, Vol. 14, No. 10, pp. 1763–1774 (2019).
[7] Suwelack, S., Röhl, S., Bodenstedt, S., Reichard, D., Dillmann, R., dos Santos, T., Maier-Hein, L., Wagner, M., Wünscher, J., Kenngott, H. et al.: Physics-based shape matching for intraoperative image guidance, *Medical physics*, Vol. 41, No. 11, p. 111901 (2014).
[8] 齋藤陽，中尾恵，浦西友樹，松田哲也：カメラ画像の大域的な輝度情報に基づく弾性体の変形推定，電子情報通信学会技術研究報告 (MI)， Vol. 116, No. 393, pp. 13–18 (2017).
[9] Miao, S., Wang, Z. J. and Liao, R.: A CNN regression approach for real-time 2D/3D registration, *IEEE transactions on medical imaging*, Vol. 35, No. 5, pp. 1352–1363 (2016).
[10] Maekawa, H., Nakao, M., Mineura, K., Chen-Yoshikawa, T. F. and Matsuda, T.: Model-based registration for pneumothorax deformation analysis using intraoperative cone-beam CT images, *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5818–5821 (2020).
[11] Wu, S., Nakao, M. and Matsuda, T.: Continuous lung region segmentation from endoscopic images for intra-operative navigation, *Computers in Biology and Medicine*, Vol. 87, No. 1, pp. 200–210 (2017).
[12] Nakao, M., Maekawa, H., Mineura, K., Chen-Yoshikawa, T. F., Matsuda, T. et al.: Statistical modeling of pneumothorax deformation by mapping CT and cone-beam CT images, *arXiv preprint arXiv:2012.13237* (2020).
[13] Nakao, M., Nakamura, M., Mizowaki, T. and Matsuda, T.: Statistical deformation reconstruction using multi-organ shape features for pancreatic cancer localization, *Medical Image Analysis*, Vol. 67, p. 101829 (2021).
[14] Isola, P., Zhu, J. Y., Zhou, T. and Efros, A. A.: Image-to-image translation with conditional adversarial networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134 (2017).
[15] Mirza, M. and Osindero, S.: Conditional generative adversarial nets, *arXiv preprint arXiv:1411.1784* (2014).

正誤表

下記の箇所に誤りがございました．お詫びして訂正いたします．

訂正箇所	誤	正
7 ページ 図 7 (a)		
7 ページ 図 7 (b)