



TITLE:

# 生成型バーチャル学習を用いた高遮蔽カメラ画像からの形状再構成

AUTHOR(S):

大矢, 友貴; 中尾, 恵; 松田, 哲也

CITATION:

大矢, 友貴 ...[et al]. 生成型バーチャル学習を用いた高遮蔽カメラ画像からの形状再構成. 情報処理学会研究報告 コンピュータビジョンとイメージメディア (CVIM) 2022, 2022-CVIM-229(10): 1-8

ISSUE DATE:

2022-03

URL:

<http://hdl.handle.net/2433/275927>

RIGHT:

ここに掲載した著作物の利用に関する注意 本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うをお願いいたします。; The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright © 2022 Information Processing Society of Japan.

# 生成型バーチャル学習を用いた 高遮蔽カメラ画像からの形状再構成

大矢 友貴<sup>1</sup> 中尾 恵<sup>1</sup> 松田 哲也<sup>1</sup>

**概要：**近年、単視点画像を用いた三次元形状の再構成問題に内在する不確実性に対し、様々な機械学習の応用が考えられている。十分なデータ数を用意できない医用画像等の分野では学習時にシミュレーション画像を活用したバーチャル学習が試みられているが、人間が同等とみなす画像であっても実画像との微小な差が推定に大きな影響を与える点が課題となっている。本研究ではシミュレーション画像と実画像に共通の潜在変数を仮定した画像変換に基づく生成型バーチャル学習の枠組みを提案する。胸腔鏡下肺がん切除術における肺の内視鏡カメラ画像に対し、患者個人の三次元 CT 画像から生成したシミュレーション画像と実画像間の類似度を改善した学習が可能であることを確認したので報告する。

**キーワード：**三次元形状推定, バーチャル学習, Variational Autoencoder, 内視鏡画像

## Shape reconstruction from occluded camera images using generative virtual learning

TOMOKI OYA<sup>1</sup> MEGUMI NAKAO<sup>1</sup> TETSUYA MATSUDA<sup>1</sup>

**Abstract:** In recent years, various applications of machine learning have been considered for uncertainty included in reconstructing 3D shapes from single-viewpoint images. In the field of medical imaging, where a sufficient amount of data cannot be prepared, virtual learning using simulation images has been attempted. However, small differences between virtual and real images can degrade the estimation performance even if the images are regarded as equivalent by humans. In this study, we propose a framework for generative virtual learning based on image transformation that assumes common latent variables between simulation and real images. We confirmed that our methods could improve the similarity between simulation images generated from 3D CT images of individual patients and real images for endoscopic camera images of the lung in thoroscopic lung cancer resection.

**Keywords:** 3D shape reconstruction, Virtual learning, Variational Autoencoder, Endoscopic image

### 1. はじめに

コンピュータビジョン分野ではカメラ画像から撮像対象の三次元形状を再構成する手法が広く研究されてきた。複数のカメラで撮影された多視点画像から撮像対象の形状を高精度に復元可能であるが、撮像時の物理的な制約や撮像条件から利用機会は限定的である。このため、単視点画像からの形状再構成が広く研究されている。例えば、異なる

向きや姿勢の人体を対象に撮像された単視点画像から人体の三次元形状を生成し、コンピュータグラフィクスや動作解析等に利用することが考えられている。しかし、人体は関節の自由度が高く複雑な姿勢を取りうるため、高頻度で手や足など人体の一定領域が遮蔽された画像が発生する。近年、そのような高遮蔽カメラ画像からの形状再構成問題に内在する不確実性に対して、機械学習の応用が多数試みられている。特に最近では、画像内に生じる遮蔽に対応するため、学習時に入力画像の一部をマスクし、遮蔽状態を生成しつつ学習する枠組み等が研究されている [1]。

<sup>1</sup> 京都大学大学院情報学研究科  
Graduate School of Informatics, Kyoto University

医療分野においても生体内部の臓器の三次元構造の把握が重要であり、手術時に取得可能な内視鏡画像から臓器の三次元形状を再構成する研究が進められている。治療前に撮像されたコンピュータ断層撮影 (Computed Tomography, CT) や磁気共鳴画像 (Magnetic Resonance Imaging, MRI) 等により、関心領域の形状再構成が可能である。しかし、外科手術や放射線治療の開始後にはこれら三次元イメージングは利用できない。また、骨盤や下肢のような剛体を対象とした形状再構成は広く研究されているが、手術時の変形が大きい臓器 [2] においては形状再構成はさらに困難な課題である。特に、内視鏡画像を対象とした生体臓器の形状推定は学習データの集積が難しいことに加え、視認範囲の狭さに起因する不確実性の高さから研究例は限定的である。これまでに胃の内視鏡映像からテクスチャを含めた胃全体の三次元情報の復元を行った研究 [3] や肝臓のモデルベース位置合わせ [4] などが報告されている。前者は動画から形状情報を再構成しているが、視認範囲内の臓器表面の形状復元に留まる、後者はパラメータ最適化によって変形を算出しているが、最適化計算における計算コストの高さと安定性に課題が残っている。

これらの問題に対して、生体臓器の形態的特徴を低次元のパラメータによって記述した統計形状アトラスやグラフ畳み込みネットワークを活用して三次元形状や変形を予測する試み [5] [6][7][8] が報告されている。また、Nitta らは、教師なし画像変換によって内視鏡画像に位置合わせされた臓器の三次元像を直接的に生成する枠組みを提案した [9]。また、データの少なさに対応するため、統計形状アトラスから生成したシミュレーション画像を用いたバーチャル学習が採用されている。患者の 3D-CT から生成したシルエット画像に基づいて形状推定を行っており、専門家によって実画像に類似した外観を持つと判断されたシミュレーション画像から三次元像が生成できることが確認された。しかし、学習済みモデルを実画像に適用した際、不正確な像が生成されることが報告されており、シミュレーション画像と実画像間の微小な差異が推定に大きな影響を与える点が課題となっている。

本研究では、内視鏡画像に代表される単一高遮蔽カメラ画像からの変形物体の形状再構成問題に対して、シミュレーション画像と実画像に共通の潜在変数を仮定した画像変換に基づく生成型バーチャル学習の枠組みを提案する。学習データの少なさに対して、統計的形状アトラスから生成されたシミュレーション画像を活用したバーチャル学習を用いる。Variational Autoencoder (VAE) [10] によってシミュレーション画像と実画像に共通の潜在変数を抽出し、人間が同等とみなす特徴の異なる二つの画像が入力された際に同一の画像に変換される枠組みを構築する。VAE は病態や形態的、解剖学的なバリエーションを表現した医用画像の生成を行った研究 [11] が報告されているが、本研

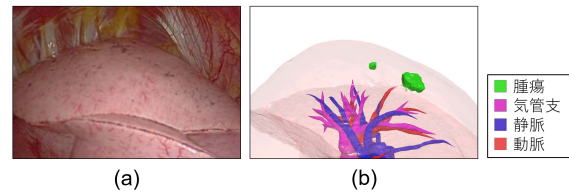


図 1 内視鏡画像における実画像とシミュレーション画像の例, (a) 手術動画から抽出した実画像  $I_R$ , (b) 実画像に対応するシミュレーション画像  $I_S$

究が対象とするシミュレーションベースの学習へ応用された事例は報告されていない。

胸腔鏡下肺がん切除術における肺の内視鏡カメラ画像からの形状推定の実験を実施し、患者個人の三次元 CT 画像から生成したシミュレーション画像と実画像間の類似度を改善することで、推定精度の向上が可能であることを確認したので報告する。

## 2. 提案手法

内視鏡下手術では、手術時に撮影される内視鏡画像および対象臓器に幾つかの条件が想定される。本研究では、多くの手術において共通に見られる以下の状況を想定する。

- 手術対象となる臓器の初期形状  $M$  が手術前に計測された 3D-CT 等から得られている
- 内視鏡画像  $I_R$  では臓器の一部表面が観察できるが、臓器全体の 50%以上が遮蔽された状態である
- 内視鏡カメラの位置や視線方向は得られていない
- 観測時に臓器は変形しているが、変形の自由度は一定範囲内である

本研究では、以上の条件下で統計形状アトラスから生成可能なシミュレーション画像  $I_S$  を活用したバーチャル学習と、内視鏡画像  $I_R$  とシミュレーション画像  $I_S$  に共通の潜在変数に基づく画像変換によって、内視鏡画像から直接的に手術中の臓器の三次元形状を求めることを目指す。

図 1 は内視鏡画像と統計的形状アトラスから生成されたシミュレーション画像の例を示している。図 1(a) には肺上葉の表面の一部と胸腔の内壁が確認される。図 1(b) は外科医の監修の下、その手術シーンに対応するよう手動で統計形状アトラスの主成分に対する重みとカメラパラメータが調整された結果である。肺表面が透過され、腫瘍位置、気管支、血管構造が把握できるため、手術支援に利用可能である。内視鏡画像に合わせて統計形状アトラスのパラメータを自動的に算出し、両者を位置合わせできれば手術時の臓器形状と解剖学的構造の同定が可能である。

### 2.1 提案フレームワーク

本研究で提案するフレームワークを図 2 に示す。提案フレームワークは実画像とシミュレーション画像間に共通の潜在変数に基づく画像変換を達成する画像変換モデルと、

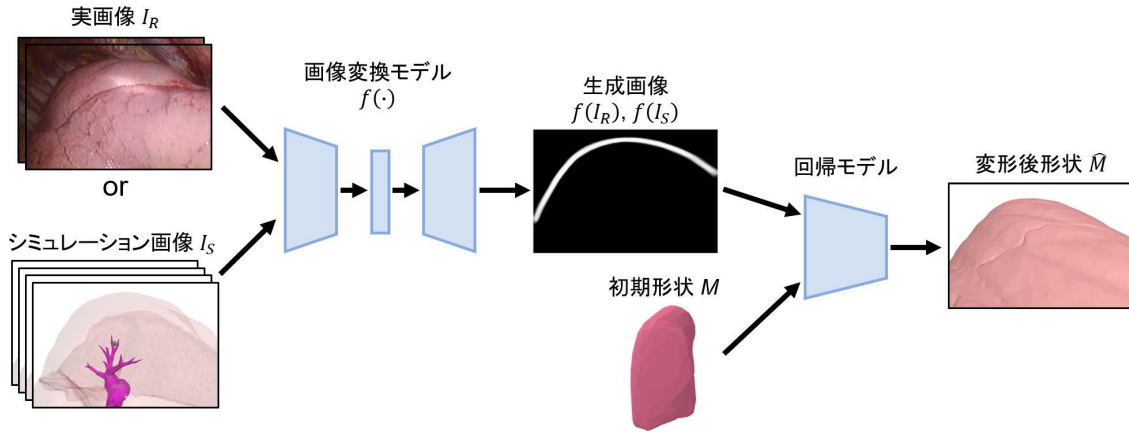


図2 提案フレームワーク. 画像変換モデルは実画像  $I_R$  とシミュレーション画像  $I_S$  に共通の潜在変数を学習し, 回帰モデルでは生成画像とモデルパラメータ  $\theta$  間の関係を学習する.

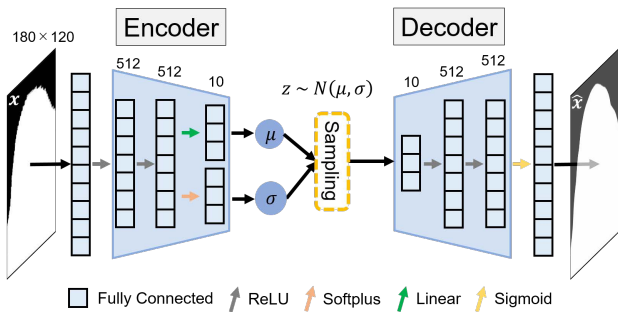


図3 VAEの層設計. 全結合層で構成

生成画像とモデルパラメータ間の関係を学習する回帰モデルにより構成される.

画像変換モデルでは細部の特徴は異なるが人間が同等とみなす実画像  $I_R$  とシミュレーション画像  $I_S$  を同一の生成画像に変換する. 本研究では, 実画像とシミュレーション画像に共通の潜在変数を仮定したことから, 画像変換モデルには同様の概念を持つVAEを採用した. 一方, 回帰モデルは生成画像と臓器の初期形状  $M$  を入力とし, 変形及びカメラパラメータを出力する. この回帰モデルにはConvolutional Neural Network(CNN)を採用し, 自己教師ありバーチャル学習を行う.

## 2.2 VAEによる画像変換

本節では, 細部の特徴は異なるが人間が同等と認識する実画像とシミュレーション画像を同一の生成画像に変換する変換関数  $f$  を提案する.

本研究において, 画像変換モデルへの入力画像は実画像  $I_R$  と  $I_R$  を人工的に再現したシミュレーション画像  $I_S$  の2種類である. 変換関数  $f$  は  $I_R$  から  $I_S$  への変換と,  $I_S$  から  $I_S$  を再構成する役割を担う関数であり, 実画像とシミュレーション画像を同一画像に変換することを期待する. そのため, 変換関数  $f$  の学習では,  $I_R$  を入力した場合,  $I_S$  に近い画像を出力し,  $I_S$  を入力した場合,  $I_S$  を出力する

ように学習を進める. その結果,  $I_R$  入力時の出力  $f(I_R)$  と  $I_S$  入力時の出力  $f(I_S)$  間の類似度が高まることが期待される. 変換関数  $f$  により特徴の異なる  $I_R, I_S$  が同等であることを学習し, 同一の画像に変換することが可能となる.

続いて, 本研究の画像変換モデルに採用するVAEについて説明する. VAEは入力データ  $x$  を潜在変数  $z$  に変換するEncoderと, 潜在変数  $z$  から元の入力データ  $x$  を再構成するDecoderから構成される. 本研究で用いたVAEの構造を図3に示す. 各層は全結合層で構成されている. Encoder, Decoderそれぞれのニューラルネットワークにおけるパラメータを  $\phi, \theta$  とし, Encoder, Decoderの確率分布  $q_\phi(z|x), p_\theta(x|z)$  とすると,  $z$  が従う真の確率分布は  $p(x)$  と表記できる. Encoderでは入力データ  $x$  を直接潜在変数  $z$  に変換するのではなく, 潜在変数  $z$  がガウス分布に従うと仮定し, 潜在変数  $z$  が従う分布  $p(x)$  の平均  $\mu$  と分散  $\sigma$  を推定する. これにより潜在変数  $z$  が広がりを持った状態で推定できるため, 潜在変数  $z$  を調整することで様々なバリエーションを表現したデータが生成可能となる. VAEの学習では各ネットワークの確率分布に関するパラメータ  $\phi, \theta$  を学習する.

$I_R, I_S$  を共通の潜在変数へと圧縮し, 共通の潜在変数から  $f(I_R), f(I_S)$  が出力される点から, 共通の潜在変数から同一の画像を再構成可能になると考えたため, 本研究ではVAEを採用した. ここで, 通常の学習手法におけるVAEの変換関数を  $g$  とすると, 変換関数  $g$  へ実画像  $I_R$  を入力した場合, 入力  $I_R$  と出力  $g(I_R)$  は  $g(I_R) \approx I_R$  の関係性が成り立つ. シミュレーション画像  $I_S$  を入力とした場合, 入力  $I_S$  と出力  $g(I_S)$  は  $g(I_S) \approx I_S$  の関係性が成り立つ.

一方, 本研究ではVAEの変換関数  $f$  へ実画像  $I_R$  を入力した場合, 入力  $I_R$  と出力  $f(I_R)$  は  $f(I_R) \approx I_S$  の関係性が成り立つ. シミュレーション画像  $I_S$  を入力とした場合, 入力  $I_S$  と出力  $f(I_S)$  は  $f(I_S) \approx I_S$  の関係性が成り立つ. 特徴の異なる二画像を同一の画像に変換する変換関

数  $f$  により,  $I_R$  と  $I_S$  が同等であることを学習することが可能であると考えた. また,  $I_R$ ,  $I_S$  を変換した Synthetic Image  $f(I_R)$ ,  $f(I_S)$  は, 互いに類似度が高い画像となる.

VAE による画像変換では潜在変数  $z$  が従う分布  $p(x)$  の尤度が最大となるように学習を行う.  $p(x)$  の対数尤度の変分下限は式 (1) で表すことができる.

$$L(x|z) = \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x) \| p(z)) \quad (1)$$

ここで, 第 1 項は Decoder における再構成誤差を, 第 2 項は Encoder における潜在変数  $z$  を推定する確率分布  $q(z|x)$  と, 潜在変数  $z$  が従う真の確率分布  $p(x)$  の Kullback-Leibler Divergence (KLD) である. VAE の学習では, 再構成誤差の最小化と KLD の最小化を目的とする. 本研究では VAE による画像変換モデルの損失関数を式 (2) により算出した.

$$\mathcal{L} = \|x - \hat{x}\|_2^2 + \lambda_{KLD} D_{KL}(q(z|x) \| p(z)) \quad (2)$$

### 2.3 バーチャル学習

本研究では医用画像分野の課題となる学習データの少なさを統計的に生成可能なシミュレーション画像を活用したバーチャル学習により解決する. 本節ではシミュレーション画像の生成および学習方法について述べる.

術中の肺の虚脱変形により, 手術前に取得可能な含気時肺形状と術中の虚脱時肺形状は大きく異なる. そこで, 本研究では肺形状アトラスを用いて臓器変形のバリエーション生成を行う. 手術前に撮像された CT データから構成された含気時の肺形状  $M$  と 術中に撮像された Cone-beam CT により構成された虚脱時肺形状  $\hat{M}$  間でメッシュ変形位置合わせ [12][13] を行い, 頂点对応が取れた肺形状メッシュ群が取得可能となる. 肺形状アトラスは対応する肺形状メッシュ群から得られる頂点単位の変位ベクトルを低次元のパラメータで記述した統計モデルである. 肺の虚脱量に関わる重みを変更することで手術前の患者個人の術前 CT モデルから, 学習による平均的な虚脱変形を基準として, 手術時に想定される虚脱変形を表現することが可能である. 虚脱変形後のメッシュ  $\hat{M}$  の頂点座標  $v'_i$  は次式で求められる.

$$v'_i = v_i + \bar{u} + \sum_{j=1}^m \omega_j \cdot u_i \quad (3)$$

ここで,  $v_i$  は含気時肺形状の頂点位置,  $\bar{u}$  は平均的な虚脱変形による変位,  $u_i$  は肺形状アトラスによって得られる変位ベクトルを表す. また,  $\omega_j$  は各方向に対する変位量の重みであり,  $\omega_j < 1.0$  で虚脱変形による平均変位に比べてしぼませた形状を表現でき,  $\omega_j > 1.0$  で虚脱変形による平均変位に比べて膨らませた形状が表現できる. 本研究では重み  $\omega$  を変形させ, 手術時に想定される虚脱変形後のメッシュを多数生成する.

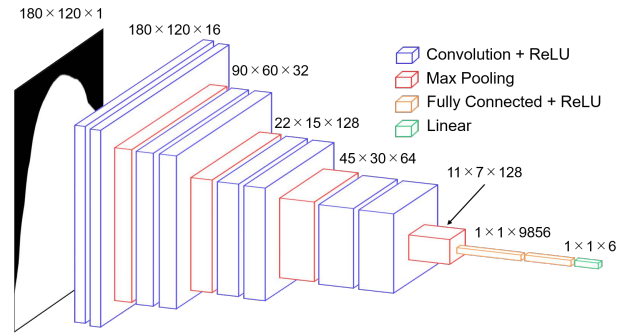


図 4 CNN の層設計. 最終層では Linear 関数によりパラメータ値を出力

次に, 肺形状アトラスにより生成した  $\hat{M}$  を設定したカメラパラメータでレンダリングを行い, シミュレーション画像  $I_S$  を生成する. カメラパラメータはカメラ位置 (position), 注視点 (focus), 上方向 (up vector) 各 3 次元の計 9 次元である. 本研究では手術時に想定されるカメラパラメータを中心にパラメータをランダムに変更し, 肺形状アトラスにより虚脱変形における変位量  $\omega$  を変更して生成した  $\hat{M}$  をレンダリングすることで, シミュレーション画像を多数生成する. 2.2 節の画像変換を厳密に行うため, バーチャル学習も共通して生成したシミュレーション画像を肺領域のみを抽出したシルエット画像へ変換し, モルフォロジー変換 [14] により肺輪郭画像に変換する.

続けて, 生成した肺輪郭画像を 2.2 節の VAE により Synthetic Image  $f(I_S)$  に変換する.  $f(I_S)$  から  $I_S$  生成時に定めたモデルパラメータ  $\theta$  (変位量の重み  $\omega$  とカメラパラメータ) への回帰を CNN により学習する. 本研究で用いた CNN の層設計を図 4 に示す. 本ネットワークでは全 4 層の畳み込み層と全 3 層の全結合層から構成される. 畳み込み層では ReLU 関数を各層の活性化関数としており, 各層 3 × 3 の Convolution と 2 × 2 の Maxpooling を行う. 第 1 層目には Batch Normalization [15] と Dropout [16] を適用している. 全結合層では最終層の活性化関数に Linear 関数を適用しており, 各モデルパラメータ値が出力される. 学習において推定されたモデルパラメータを  $\hat{\theta}$  とし, 次に示す 2 つの損失関数をもとに学習を進める.

### Reconstruction Error

本研究の目的は内視鏡画像に位置合わせされた臓器の三次元形状を求めることであるため, 推定形状と目的形状間の差に加え, 得られた形状の位置, 姿勢のずれも小さくなることが望ましい. 提案手法では式 (4) で定義される Reconstruction Error を投影後のメッシュ頂点間の誤差を算出する損失関数として導入する.

$$\mathcal{L}_R = \frac{1}{n} \sum_{i=1}^n |M_{\theta} v_i - M_{\hat{\theta}} \hat{v}_i| \quad (4)$$

ここで  $M_\theta$ ,  $M_{\hat{\theta}}$  はそれぞれ推定対象及び推定された射影行列であり,  $v_i$ ,  $\hat{v}_i$  はそれぞれ推定対象及び推定されたメッシュの頂点位置である.  $M$  はカメラパラメータに基づいて臓器メッシュのレンダリング像を生成する際の透視投影変換に用いられる  $4 \times 4$  の行列である.

### Parameter Loss

胸腔内の空間は限られるため, 内視鏡カメラは一定範囲内に動きが制約される. カメラパラメータの変動範囲に制約を持たせるために, 提案手法では式 (5) で定義される Parameter Loss をカメラパラメータおよび肺の虚脱変形における変位量  $\omega$  の推定値  $\hat{\theta}$  と真値  $\theta$  との誤差を算出する損失関数として導入する.

$$\mathcal{L}_P = \frac{1}{n} \sum_{i=1}^n |\theta_i - \hat{\theta}_i|^2 \quad (5)$$

バーチャル学習全体の損失関数  $\mathcal{L}$  は次のように 2 つの損失関数の重み付き線形和として定義する.

$$\mathcal{L} = \mathcal{L}_R + \lambda_P \mathcal{L}_P \quad (6)$$

## 3. 評価実験

1 枚の内視鏡画像を用いた臓器形状推定における提案フレームワークの有効性を確認するために, VAE による画像変換の確認と変換後画像を用いた形状推定における手法間比較の二つの実験を実施した. VAE の学習においては, バッチサイズ 20, エポック数 300, バーチャル学習においては, バッチサイズ 60, エポック数 500 で, 両学習に共通して学習率  $1.0 \times 10^{-3}$  とした最適化手法 Adam を用いた. 損失関数の組み合わせは複数組の実験結果から,  $\lambda_{KL} = 5$ ,  $\lambda_P = 0.5$  を採用した.

### 3.1 データセット

本実験では, 胸腔鏡下肺がん切除術右肺 2 症例, 左肺 1 症例の手術動画から抽出した静止画像における肺輪郭画像に対して, 肺形状の推定を行った. 各症例において構成したデータセット数を表 1 に示す. 提案手法における画像変換モデルのデータセットには, 全 12 症例のうちテストケース 1 症例を除いた交差検証を実施した. すなわち, 11 症例分の胸腔鏡下肺がん切除術の手術動画から抽出した静止画像における肺輪郭画像を学習データに, それに対応した同数のシミュレーション画像を VAE の再構成時の標的画像とした. 実画像の少なさを補うため, 全実画像に対して 2 パターンの回転処理を行った肺輪郭画像を学習データに, 回転前の肺輪郭画像に対応したシミュレーション画像を VAE の標的画像へ追加した. さらに, シミュレーション画像を入力とした際に共通の潜在変数からシミュレーション画像を再構成するための学習データと VAE の再構成標的

表 1 各症例のデータ数

Case	VAE 学習		計	バーチャル学習	テスト
	Real	Sim			
01	363	10,363	10,726	10,000	7
02	345	10,345	10,690	10,000	13
03	345	10,345	10,690	10,000	13

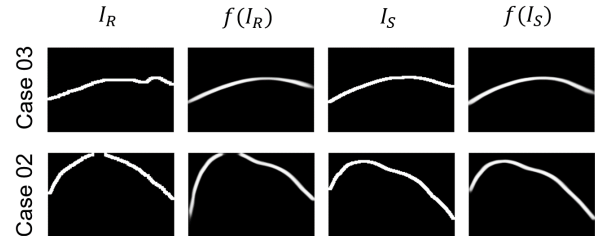


図 5 VAE による画像変換例. Case03: 平均類似度が最大の例, Case02: 平均類似度が最小の例

表 2 VAE による画像変換の評価結果 [%]

	$DSC(I_R, I_S)$	$DSC(f(I_R), f(I_S))$
Case01	35.2±13.7	62.3±9.2
Case02	25.6±15.0	44.6±21.6
Case03	37.8±13.7	57.5±14.9

画像には, 肺輪郭画像, 回転処理を行った肺輪郭画像に対応したシミュレーション画像, および各症例において外科医が事前に想定したカメラパラメータを中心として, 右肺症例ではカメラ位置の奥行方向に 15mm, 横軸方向, 縦軸方向に 25mm, 注視点の横軸方向, 縦軸方向に 15mm, 左肺症例では前述のカメラ位置の 3 方向, および注視点の 2 方向に 10mm, 肺のしぼみ量を平均的なしぼみを中心として 0.1 の幅で各パラメータを変化させ生成したシミュレーション画像 10,000 枚をデータセットに追加した. 同様のシミュレーション画像 10,000 枚をバーチャル学習のデータセットとする. なお, 本実験で用いた画像は  $180 \times 120$  pixel に統一している.

### 3.2 画像変換

本節では形状推定への応用を想定した, 胸腔鏡下肺がん切除術右肺 2 症例, 左肺 1 症例におけるシミュレーション画像と実画像に共通の潜在変数を仮定した画像変換の前実験を行った. 学習データ, およびテストデータの構成は表 1 の通りである.

図 5 は提案方法における画像変換結果である.  $I_R$  は胸腔鏡下肺がん切除術の手術動画から抽出した静止画像における肺輪郭画像,  $I_S$  は  $I_R$  に対応したシミュレーション画像,  $f(I_R)$  と  $f(I_S)$  は  $I_R$ ,  $I_S$  の変換後画像である. 提案方法において人間が同等とみなす特徴の異なる二枚の画像が同一の画像に変換された場合, シミュレーション画像と実画像間の類似度が高まることが考えられる. そのため, 提

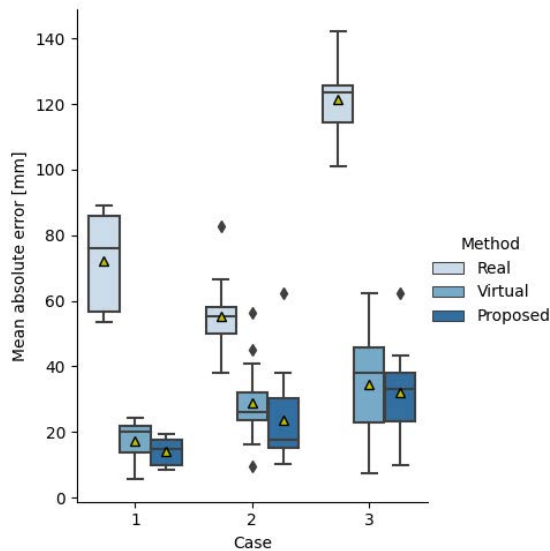


図 6 手法比較結果. Real はリアル学習, Virtual はバーチャル学習, Proposed は提案手法を表している. 平均値 (黄色△) は Proposed が最小

案方法における画像変換が成功しているか  $I_S$  と  $I_R$ ,  $f(I_S)$  と  $f(I_R)$  の類似度を参考に確認した. 画像間の類似度として Dice Similarity Coefficient(DSC) を用いる. DSC は 0 から 1 の値をとり, 値が大きいほど類似度が高いことを示す. 図 5 の例では, Case03 で最大 DSC76.0%, Case02 で最小 DSC16.1% であり, 変換前の  $I_R$ ,  $I_S$  間の DSC はそれぞれ 45.1%, 15.3% であった. Case03 では  $I_R$  特有の輪郭構造がシミュレーションで再現できる特徴へ変換されたことが目視で確認できる. また, Case02 においても  $f(I_R)$  は  $I_S$  に類似した特徴を有しているが, 輪郭の差から低い類似度となった.

表 2 に各症例における  $I_S$  と  $I_R$ ,  $f(I_S)$  と  $f(I_R)$  の類似度の平均値を示す. 今回テストを行った 3 症例全てで, 本手法により実画像とシミュレーション画像間の類似度が改善された. 3 症例において  $f(I_S)$  と  $f(I_R)$  間の DSC が最も高かったシーン (Case03 scene01), 最も低かったシーン (Case02 scene05) を図 5 に示す. それぞれ  $f(I_S)$  と  $f(I_R)$  間の DSC は 76.0%, 16.1% であった.

図 5 から, 本手法によりシミュレーションでは再現が困難な実画像特有の輪郭がシミュレーションで再現できる特徴へ近づいたことが確認された.

### 3.3 形状推定

本節では, 胸腔鏡下肺がん切除術右肺 2 症例, 左肺 1 症例の手術動画から抽出した静止画像における肺輪郭画像に対して, 肺形状推定の実験を行った. 提案手法の有効性を確認するため, 全 12 症例のうち, テストケース 1 症例を除いた 11 症例分の胸腔鏡下肺がん切除術の手術動画から抽出した静止画像における肺輪郭画像を学習データとした学

表 3 形状推定誤差 [mm]

	リアル学習	バーチャル学習	提案手法
Case01	72.0±15.8	17.3±6.7	13.9±4.5
Case02	55.1±11.4	28.8±12.6	23.7.2±14.1
Case03	121.5±12.4	34.7±18.3	32.0±9.7

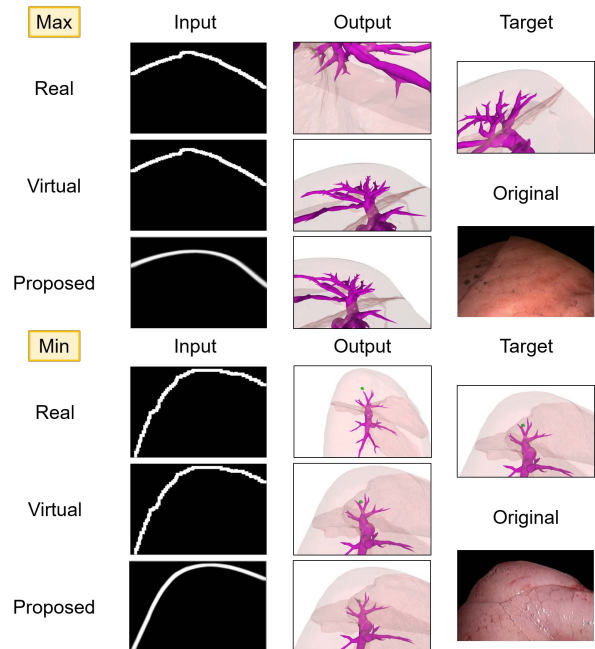


図 7 形状推定結果の手法比較. 各手法の Output と Target のレンダリング像を比較

習手法 (以下, リアル学習と呼称), 各症例において外科医が事前に想定した手術時のカメラパラメータに基づき, 3.2 節と同様の範囲で各パラメータを変化させ生成したシミュレーション画像 10,000 枚を学習データとしたバーチャル学習, 表 1 に示すデータを学習データとした提案手法で形状推定精度を比較した.

リアル学習, バーチャル学習, 提案手法の 3 手法において, 推定した臓器形状と真値とのずれを評価するため, 式 (7) で定義される Mean Absolute Error(MAE) により, カメラ座標系での三次元形状間のずれを評価する.

$$MAE = \frac{1}{n} \sum_{i=1}^n |M_{\theta} v_i - M_{\hat{\theta}} \hat{v}_i| \quad (7)$$

ここで  $M_{\theta}$ ,  $M_{\hat{\theta}}$  はそれぞれ推定対象及び推定された射影行列であり,  $v_i$ ,  $\hat{v}_i$  はそれぞれ推定対象及び推定されたメッシュの頂点位置である.  $M$  はカメラパラメータに基づいて臓器メッシュのレンダリング像を生成する際の透視投影変換に用いられる  $4 \times 4$  の行列である.

図 6 に各症例における従来手法との比較結果を示す. リアル学習 (Real), バーチャル学習 (Virtual), 提案手法 (Proposed) の形状再構成誤差範囲, および平均値 (△) から, 全症例において提案手法の平均値が減少し, 誤差範囲も縮小されていることが確認でき, Case01 では誤差範囲の縮小が顕著である. 各症例のリアル学習, バーチャル学習,

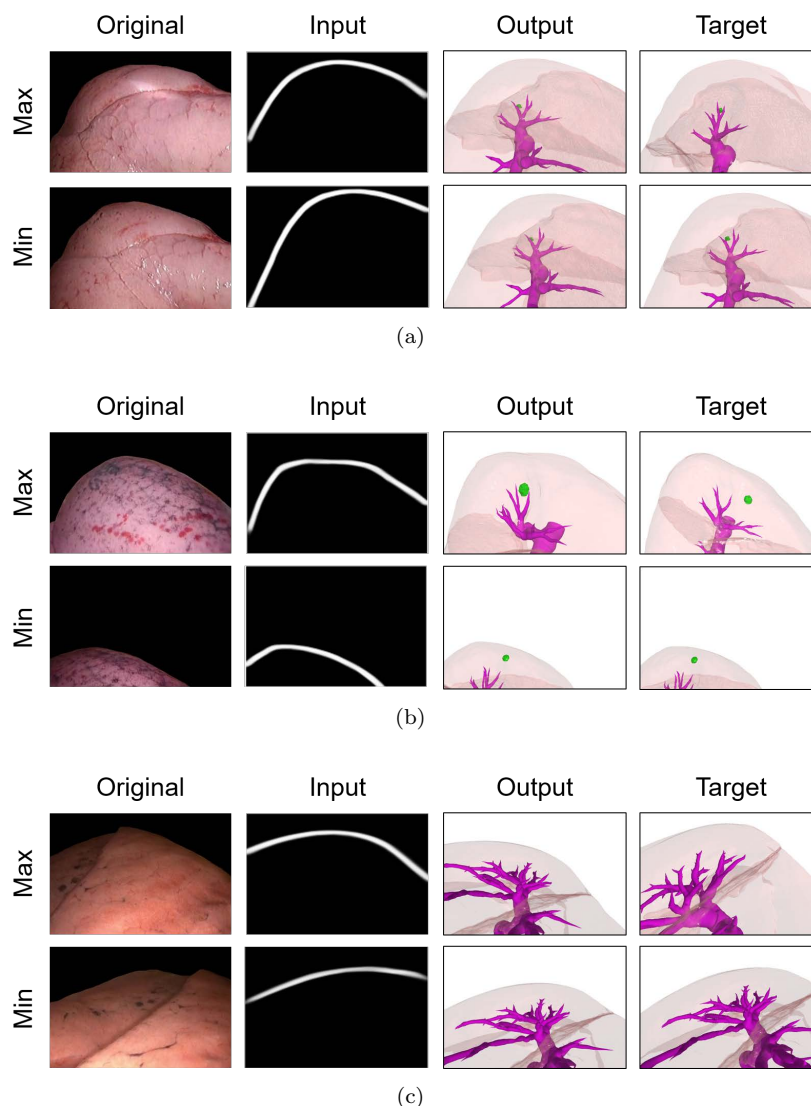


図 8 提案手法による形状推定結果. (a) Case01, (b) Case02, (c) Case03 における形状再構成誤差最大・最小の場合.

提案手法における誤差は表 3 の通りである. 症例間で精度にばらつきはあるが, リアル学習と他手法における精度の差は明らかであり, バーチャル学習と提案手法においても形状再構成精度の向上が確認できる. また, 図 7 に 3 症例における提案手法の形状再構成結果から形状誤差が最大の場合と最小の場合を, 図 8 に各症例における形状誤差が最大の場合と最小の場合を示す. 図 7 は肺輪郭画像を入力とし, リアル学習, バーチャル学習, 提案手法における形状推定結果, および形状の真値を並べたものである. 提案手法における形状誤差最大の例では MAE 62.4mm(同シーンのリアル学習では MAE 138.7mm, バーチャル学習では MAE 54.2mm), 最小の例では MAE 8.3mm(同シーンのリアル学習では MAE 85.4mm, バーチャル学習では MAE 5.5mm)となっている. 形状再構成誤差最大の例では画像変換の精度のため, 形状再構成誤差最小の例では画像変換前の画像がシミュレーションで再現できる輪郭構造を有し

ていたため, 図 7 ではバーチャル学習の方が提案手法より精度が良い結果となった. 図 8 では, 葉間や輪郭の一致度合いから, 各症例で形状再構成誤差が最小となった場合では高精度に形状再構成が可能であることが確認できる. 形状再構成誤差の値は, 図 8(a): Case01 最大 19.4 mm 最小 8.3 mm, 図 8(b): Case02 最大 62.1 mm 最小 10.2 mm, 図 8(c): Case03 最大 62.4 mm 最小 9.7 mm であった. 形状再構成誤差最大の場合には画像変換により輪郭構造が変化することで精度が低下した可能性がある. 図 8(b) では幅の広い輪郭へと変換され, 図 8(c) では輪郭の傾き度合いが変化した.

そのため, テストケースとした 3 症例において, 本提案手法によって形状推定精度が向上することが確認されたものの臨床利用可能なレベルとはならなかった. 特に, カメラ位置と注視点の距離が平均 87mm と比較的近い Case01 では高精度に形状推定が可能であったが, カメラ位置と注視



点の距離が平均 118mm, 102mm となった Case02, Case03 では形状誤差が大きくなる結果となった。

#### 4. おわりに

本研究では内視鏡画像に代表される単一高遮蔽カメラ画像からの臓器の形状再構成問題に対して、生成型バーチャル学習の枠組みを提案した。統計的形状アトラスから生成されたシミュレーション画像に基づくバーチャル学習によって学習データの少なさへ対応し、VAE によって実画像とシミュレーション画像間に共通の潜在変数に基づく共通の生成画像への変換を実現した。

右肺 2 症例, 左肺 1 症例の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に対して形状推定を行い、提案手法は従来手法と比較してより高精度に形状推定が可能であることを確認した。

**謝辞** 本研究は日本学術振興会 科研費補助金 基盤研究 (B) 19H04484 及び 21H03020 の支援による。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] Lin, K., Wang, L. and Liu, Z.: End-to-End Human Pose and Mesh Reconstruction with Transformers, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1954–1963 (2021).
- [2] Nakao, M., Tokuno, J., Chen-Yoshikawa, T., Date, H. and Matsuda, T.: Surface deformation analysis of collapsed lungs using model-based shape matching, *International Journal of Computer Assisted Radiology and Surgery*, Vol. 14, No. 10, pp. 1763–1774 (2019).
- [3] Widya, A. R., Monno, Y., Okutomi, M., Suzuki, S., Gotoda, T. and Miki, K.: Stomach 3D Reconstruction Using Virtual Chromoendoscopic Images, *IEEE Journal of Translational Engineering in Health and Medicine*, Vol. 9, pp. 1–11 (2021).
- [4] Koo, B., Ozgur, E., Roy, B. L., 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 (MICCAI)*, pp. 326–334 (2017).
- [5] Wu, S., Nakao, M., Tokuno, J., Chen-Yoshikawa, T. and Matsuda, T.: Reconstructing 3D Lung Shape from a Single 2D Image during the Deaeration Deformation Process using Model-based Data Augmentation, *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pp. 1–4 (2019).
- [6] Nakao, M., Nakamura, M. and Matsuda, T.: Image-to-Graph Convolutional Network for Deformable Shape Reconstruction from a Single Projection Image, *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 259–268 (2021).
- [7] Modrzejewski, R., Collins, T., Bartoli, A., Hostettler,

- A. and Marescaux, J.: Soft-Body Registration of Pre-operative 3D Models to Intra-operative RGBD Partial Body Scans, *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 39–46 (2018).
- [8] Ketcha, M. D., De Silva, T., Uneri, A., Jacobson, M. W., Goerres, J., Kleinszig, G., Vogt, S., Wolinsky, J.-P. and Siewerdsen, J. H.: Multi-stage 3D-2D registration for correction of anatomical deformation in image-guided spine surgery, *Phys Med Biol*, Vol. 62, No. 11, pp. 4604–4622 (2017).
- [9] 新田潤平, 中尾 恵, 松田哲也: 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, 情報処理学会コンピュータビジョンとイメージメディア研究会 (CVIM), No. 13, pp. 1–8 (2021).
- [10] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, *Proc. 2nd International Conference on Learning Representations (ICLR)* (2014).
- [11] Sengupta, S., Athwale, A., Gulati, T., Zelek, J. and Lakshminarayanan, V.: FunSyn-Net: enhanced residual variational auto-encoder and image-to-image translation network for fundus image synthesis, *Proc. SPIE Medical Imaging*. 113132M (2020).
- [12] 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, *Proc. 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5818–5821 (2020).
- [13] Nakao, M., Maekawa, H., Mineura, K., Chen-Yoshikawa, T. F., Date, H. and Matsuda, T.: Kernel-based modeling of pneumothorax deformation using intraoperative cone-beam CT images, *Proc. SPIE Medical Imaging*. 115980P (2021).
- [14] Serra, J.: Introduction to mathematical morphology, *Proc. Computer Vision, Graphics, and Image Processing*, Vol. 35, No. 3, pp. 283–305 (1986).
- [15] Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *Proc. 32nd International Conference on International Conference on Machine Learning*, pp. 448–456 (2015).
- [16] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, Vol. 15, No. 56, pp. 1929–1958 (2014).