

統計的変位場ガイドを用いた拡散モデルによる胸部4D-CT画像生成

金室 美紀[†] 平島 英明^{††} 中村 光宏[†] 中尾 恵[†]

[†] 京都大学大学院医学研究科 人間健康科学系専攻 〒606-8507 京都市左京区聖護院川原町 53

^{††} 京都大学医学部附属病院 放射線治療科 〒606-8507 京都府京都市左京区聖護院川原町 54

E-mail: [†]kanamuro.miki.56m@st.kyoto-u.ac.jp

あらまし 患者個人の呼吸動態を得るために用いられる 4D-CT 撮影は画像内に臓器の移動によるモーションアーチファクトが生じ、高い被ばく線量をもたらす。本研究では、条件付き拡散モデルによって変位場を生成し、1 時相のみの画像から異なる時相の画像を生成する 4D-CT 画像予測の枠組みの構築を目的とする。本モデルでは、変位量によって制御可能な統計的変位場をガイドとした拡散モデルによって患者個人の変位場を予測し、空間変換によって単一時相の画像からターゲット位相の変形 CT 画像を生成する。提案する枠組みには予測対象とする時刻の画像計測を必要とせず、従来モデルより高品質な画像生成が可能であることを確認したので報告する。

キーワード 4D-CT 画像予測, 変位場, 拡散モデル, 呼吸変形

Chest 4D-CT Image Generation Using Statistical Displacement Field-Guided Diffusion Model

Miki KANAMURO[†], Hideki HIRASHIMA^{††}, Mitsuhiro NAKAMURA[†], and Megumi NAKAO[†]

[†] Human Health Sciences, Graduate School of Medicine, Kyoto University, Sakyo, Kyoto, 606-8507, Japan

^{††} Radiation Oncology and Image-Applied Therapy, Kyoto University Hospital, Sakyo, Kyoto, 606-8507, Japan

E-mail: [†]kanamuro.miki.56m@st.kyoto-u.ac.jp

Abstract To obtain individual respiratory dynamics, time-series 3D-CT (i.e., 4D-CT) imaging is employed. However, 4D-CT acquisition suffers from motion artifacts caused by organ motion within the images and increases radiation exposure. This study aimed to establish a 4D-CT image prediction framework using a diffusion model that incorporates both the initial-phase CT image and a controllable statistical displacement vector field (sDVF) parameterized by the displacement magnitude. We report that the framework enables high-quality image generation compared to conventional models without the acquisition of target-phase images.

Key words 4D-CT image prediction, displacement vector field, diffusion model, respiratory deformation

1. はじめに

コンピュータ断層撮影 (Computed Tomography, CT) や磁気共鳴画像 (Magnetic Resonance Imaging, MRI) などの三次元医用画像は、臓器や腫瘍の位置や形状など患者固有の形態的情報を画像化できるため、診断および手術計画、治療計画などに広く用いられている。特に、肺がんのような呼吸による臓器の移動や変形を伴う腫瘍に対する放射線治療では、腫瘍領域のみに効率的に放射線を照射する必要がある [1]。治療計画時に患者個人の呼吸動態を得るために、体表に設置されたマーカーをトラッキングし、呼吸同期下で CT 画像を時系列的に計測する 4D-CT 撮像が用いられている。しかし、4D-CT 撮影は画像内に臓器の移動によるモーションアーチファクト

が生じ、より高い被ばく線量をもたらす。

これまでに治療中に想定される生体内部の動きを反映した画像を得るために、治療前に撮像された 3D-CT 画像を用い、治療中にリアルタイムで撮像可能な二次元投影像から、臓器の三次元的な変動を得る試みがなされている。近年、機械学習によってターゲット時刻の画像や変位場 (Displacement Vector Field, DVF) を生成するアプローチ [2], [3] や、事前に取得された患者個人の形状データの 2D/3D 位置合わせの方法 [4], [5] が報告されている。これらの方法は呼吸によって変動する腫瘍を追跡するようなイメージガイド下の治療の用途に有効である。しかし、予測ターゲットとなる時刻の画像計測を前提としており、治療前の線量計画や事前予測等への利用は想定されていない。

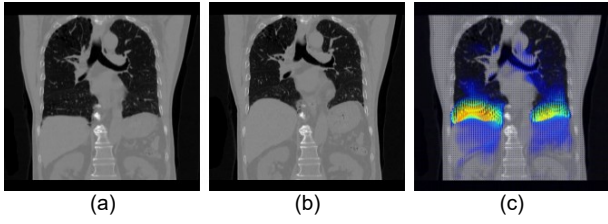


図 1 4D-CT 画像から得られる変位場, (a) 吸気終端相の CT 画像 ($T = 00$), (b) 対応する呼気終端相の CT 画像 ($T = 50$), (c) 3 次元変位ベクトル場

本研究では、条件付き拡散モデルによって変位場を生成し、1 時相のみの画像から異なる時相の画像を生成する 4D-CT 画像予測の枠組みの構築を目指している。従来研究とは異なり、提案する枠組みには予測対象とする時刻の画像計測を全く必要としない。具体的には、4D-CT 画像から得られる吸気終端相と呼気終端相の 3D-CT 画像間の位置合わせによって、事前に図 1 に示すような 3 次元変位場を取得しておく。また、患者個人の変位場の復元に適したガイドとして、初期位相の CT 画像に加えて、変位量によって制御可能な胸部の呼吸変位の空間分布を統計的に表現したベクトル場（統計的変位場, statistical displacement vector field, sDVF）を与える条件付き拡散モデルを提案する。これらの概念により、1 時相の CT 画像のみから、患者個人の形態的特徴と統計的な呼吸変形を反映した画像生成の達成が期待される。

本研究の学術的な貢献は以下の通りである。

- 1 時相の画像のみから予測される呼吸変形を反映した 3D-CT 画像を時系列的に生成する新しい拡散生成モデル
- 直接的に 4D-CT 画像を生成せず、変位場の生成と空間的な変形によって臓器特徴を維持した画像予測を達成する枠組み
- 生体が共通に有する変位場分布と変位量を事前知識として拡散モデルに与えるガイドの設計

2. 提案手法

2.1 統計的変位場をガイドとした変位場生成モデル

提案モデルの全体的な枠組みとデータ処理の流れを図 2 に示す。本研究では、前処理として 4D-CT 画像に対する VoxelMorph [6] による可変形位置合わせにより、教師変位場を算出する。具体的には、吸気終端相（初期位相, ソース）と呼気終端相（ターゲット）の 3D-CT 画像間の位置合わせによって二つの位相間の変位場 ϕ_0 を算出する。

変位場生成モデルの学習では、拡散モデルの考え方に基づいて、ノイズ ϵ を付与した変位場 ϕ_t から元の変位場 ϕ_0 を復元するように学習を行う。条件付き拡散モデル [7] では画像を拡散モデルのガイド x (Condition) として与え、画像生成タスクの方向付けが可能である。

本研究では、変位場の復元性能を高めるために、初期位相の 3D-CT 画像に加えて、統計的変位場 $\phi(u)$ を新たにガイドとして導入する。 $\phi(u)$ は変位量 u によって制御可能とした変位場の事前分布であり、呼吸変形において生体が共通に有する胸部

変位の平均的な空間分布 ϕ_m (Mean displacement vector field, mDVF) を拡張することによって得られる。なお、mDVF をガイドに用いた変位場生成は先行研究 [8] において報告したモデルであり、Diffusion-based DVF generation model (DDVF) と呼称する。本論文において新たに提案する統計的変位場 $\phi(u)$ をガイドに用いるモデルは、その拡張として DDVF+ と表記する。加えて、本研究では、高速な変位場生成のために拡散モデルの一種である一貫性モデル [9] を採用する。

推論時には学習済みモデルにソースの 3D-CT 画像とターゲットとなる変位量に対応する統計的変位場 $\phi(u)$ を結合してガイドに与え、変位場の予測結果 $\hat{\phi}_t$ を得る。得られた変位場 $\hat{\phi}_t$ を空間的なマッピング関数とし、ソースの 3D-CT 画像を変形することによって、呼吸変形を再現した変形 3D-CT 画像 (synthetic CT image, sCT) を得る。

2.2 統計的変位場

本節では、拡散モデルのガイドとして用いる統計的変位場の算出方法について述べる。

最初に、ソース 3D-CT 画像に対応する患者 i の変位場 ϕ_i を得る。全患者 $i = 1, \dots, m$ について、 ϕ_i の患者間平均を算出することによって平均変位場 ϕ_m を得る。次に、得られた mDVF を変位量 u によって制御可能とするため、 ϕ_m の空間平均のユークリッドノルム $\bar{u}$ を用いて mDVF を正規化する。すなわち、正規化変位場を $\phi_N = \bar{\phi}/\bar{u}$ によって得る。 ϕ_N は単位変位量に相当する変位の空間分布を表すため、求める統計的変位場は $\phi(u) = \phi_N \times u$ によって定義される。

統計的変位場の算出に用いる変位場 ϕ_i は、深層学習に基づく位置合わせモデルとして広く利用されている VoxelMorph [6] によって得る。本研究では、ソース画像 I_s とターゲット画像 I_t 間の可変形位置合わせタスクを学習した学習済みモデルを用い、患者 i に対するモデルの出力として、各ボクセルに変位ベクトルが格納された三次元変位場 ϕ_i を取得する。

3. 実験

提案モデルの有効性を確認するために、1) 拡散モデルに与えるガイドの違いが変位場生成に与える影響の検証と 2) 既存手法を比較による sCT 画像の評価の二つの実験を行った。提案モデルは Python 3.12.10, PyTorch 2.10 を用いて実装し、実験には CPU: Intel Core i9 3.70GHz, メモリ: 64.0GB, GPU: NVIDIA RTX A6000 48GB を搭載した PC を用いた。

3.1 対象データ

モデルの学習及び推論の対象とするデータセットには、京都大学医学部附属病院において肺がんに対する放射線治療を受けた患者 62 名の胸部 4D-CT ボリュームデータを用いた。吸気終端相と呼気終端相の 2 時相の 3 次元 CT 画像 ($T = 00, 50$) から構成され、図 1(a), (b) に示すように ($T = 00$) は吸気終端相 (end-inhalation)、($T = 50$) は呼気終端相 (end-exhalation) にそれぞれ対応している。なお、本研究における患者データの利用は京都大学医学研究科 医の倫理委員会によって承認された (承認番号: R1446-3)。

最初に症例ごとの撮像範囲や解像度の違いを軽減するため

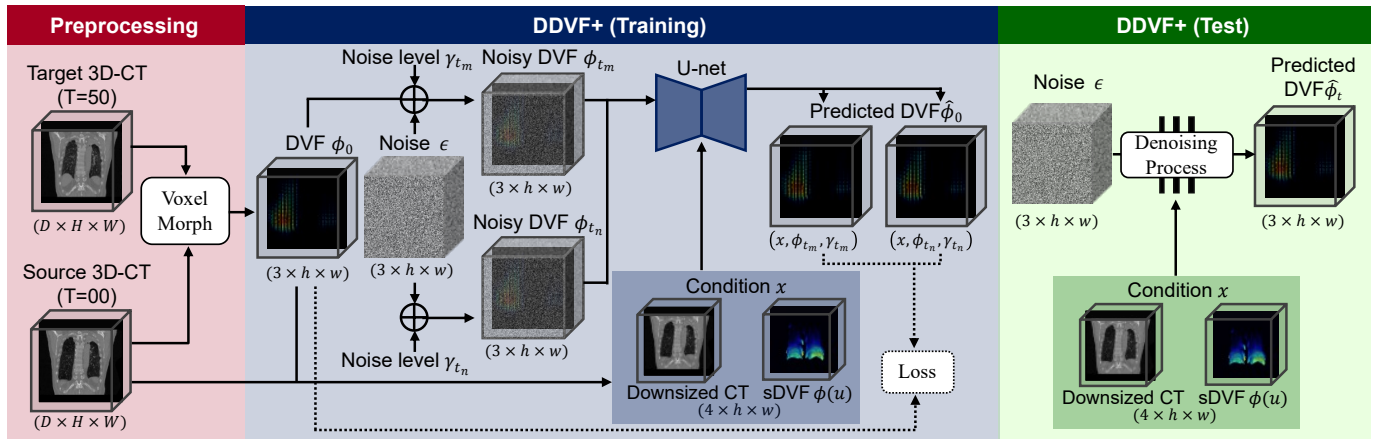


図2 提案する DDVF+ による統計的変位場をガイドとした変位場生成の枠組みと処理の流れ

に、症例間で平行移動のみによる三次元剛体位置合わせを実施した。次に、各ボリュームが第1肋骨から第12肋骨の範囲を示すように体軸方向にクロップし、各ボクセルが等方的になるようにリサイズした。CT値の $[-1000\text{HU}, 1000\text{HU}]$ にクリップ後、 $[-1, 1]$ に正規化し、ボリュームのサイズが 512^3voxel となるようゼロパディングを行った。最後に VoxelMorph による可変形位置合わせにより、学習時のメモリ容量の上限から $128^3\text{voxel} \times 3$ チャンネルの教師変位場を算出した。

3.2 統計的変位場が拡散モデルに与える影響

最初の実験では、変位場予測モデル DDVF+ のガイドとして統計的変位場を用いることの効果を検証した。以下の3通りのガイドを用いた学習済みモデルによって生成される変位場を比較評価する。

ガイド A (I_d, u): 初期位相 ($T=00$) の冠状面 CT スライス画像 I_d に変位量 u を結合したガイド

ガイド B (I_d, ϕ_m): 条件 A の CT スライス画像に平均変位場 ϕ_m を結合したガイド

ガイド C ($I_d, \phi(u)$): 条件 A の CT スライス画像に統計的変位場 $\phi(u)$ を結合したガイド

モデルの学習には 62 例の 3D-CT 画像をトレーニングデータ 54 例 (6912slices)、検証データ 2 例 (6slices)、テストデータ 6 例 (768slices) に分割して用いた。学習は 200epoch とし、推論には検証データに対して MAE が最良となった epoch のモデルを用いた。予測変位場の評価指標には、中央の 64^3voxel を関心領域 (ROI) とし、3.1 節において得たターゲット変位場 DVF (Reference) との各軸方向の MAE とユークリッドノルムの MAE を用いた。

表1に各ガイドによって得られた変位場の左右 (LR)、背腹 (AP)、頭尾 (SI) の各軸方向の MAE とユークリッドノルムの MAE を示す。いずれのガイドも各軸において誤差 0.5mm 以内の変位ベクトルの予測が達成されており、ガイド C ($I_d, \phi(u)$) の誤差が最も小さくなる結果が得られた。

図3は2症例に対して、異なるガイドによって予測された変位場をターゲット時相 ($T=50$) の冠状面の CT スライス画像に重畳して可視化した結果であり、左から (I_d, u)、(I_d, ϕ_m)、($I_d, \phi(u)$) およびターゲットの変位場 (Reference) を示し

表1 拡散モデルに与える条件の違いが予測変位場に与える影響

MAE [mm]↓	(I_d, u)	(I_d, ϕ_m)	($I_d, \phi(u)$)
LR	0.15 ± 0.01	0.13 ± 0.01	0.11 ± 0.01
AP	0.24 ± 0.04	0.19 ± 0.04	0.17 ± 0.04
SI	0.36 ± 0.05	0.31 ± 0.05	0.28 ± 0.05
Euclidean	0.41 ± 0.06	0.34 ± 0.06	0.32 ± 0.06

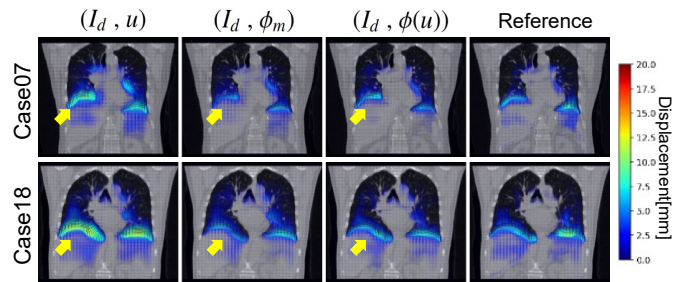


図3 異なる3つのガイドによって予測された変位場

ている。ガイド A (I_d, u) を用いたモデルでは、Reference と比較して全体的に大きな変位が生成されており、ガイド B (I_d, ϕ_m) では全体的に小さな変位として予測されている。これらと比較して、ガイド C ($I_d, \phi(u)$) を用いた提案モデルは、より Reference に近い変位分布を生成している。

3.3 変形 3D-CT 画像の比較

二つ目の実験では、初期位相の CT 画像からターゲット位相の 3D-CT 画像を生成する問題において、既存モデルとの比較を通して提案モデルの性能を検証した。本実験では、比較手法として初期位相の CT 画像からターゲット位相の sCT 画像を直接生成可能な深層学習ベースの条件付き画像生成モデルを比較対象とし、得られる sCT 画像の精度を DDVF/DDVF+ と比較した。

比較モデルは、初期位相の 3D-CT 画像からターゲット位相の 3D-CT 画像への変換タスクを学習する Generative Adversarial Networks (GAN) ベースの pix2pix [10]、初期位相からターゲット位相の 3D-CT 画像への変換を、ガイド付きのノイズ除去タスクとして学習する Conditional Denoising Diffusion Probabilistic Model (DDPM [7]), DDPM の学習を異なる 2 時

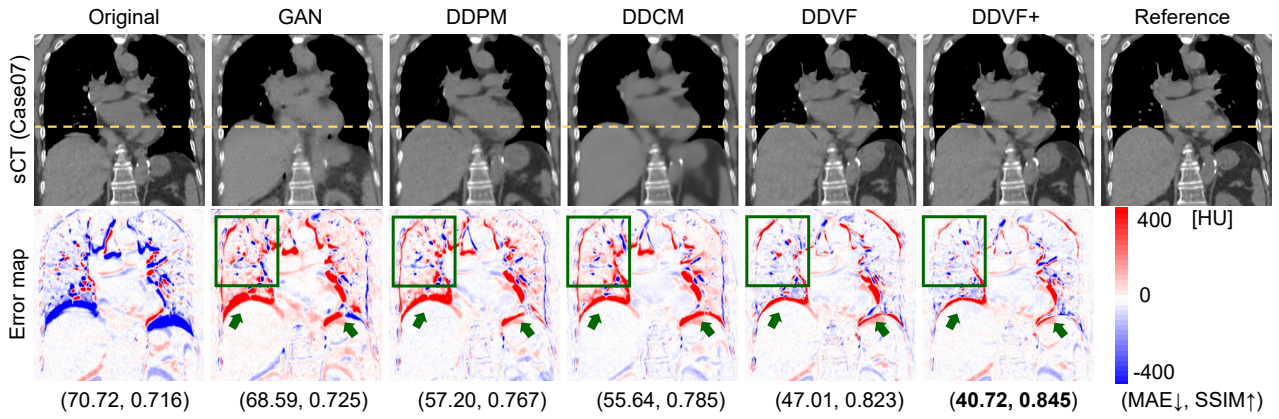


図 4 各手法によって得られたターゲット位相の sCT 画像と予測誤差

表 2 ターゲット位相の sCT 画像の手法間比較

	MAE↓	SSIM↑	PSNR↑
GAN	35.0±5.5	0.857±0.025	27.1±1.3
DDPM	36.0±4.7	0.816±0.019	27.6±1.1
DDCM	32.0±4.8	0.869±0.019	27.7±1.0
DDVF	30.0±4.2	0.868±0.018	27.3±0.9
DDVF+	28.0±4.5	0.876±0.020	28.1±0.9

刻から目標となる画像を直接予測する Conditional Denoising Diffusion Consistency Model (DDCM [9]) を採用した。

表 2 に得られた sCT 画像の中央 256^3 voxel に対する評価値の平均と標準偏差を示す。すべての指標において DDVF+ が最良値を示した。また、図 4 にコロナ断面における sCT 画像の事例を示す。各症例について、上段には初期位相の CT 画像 (Original, $T = 00$)、ターゲット位相の CT 画像 (Reference, $T = 50$) および各モデルにおいて最終的に得られた sCT 画像を、下段はターゲット位相の CT 画像との誤差を可視化したエラーマップ示している。赤は予測 CT 画像がターゲット CT 画像より高い CT 値を示す領域、青は低い CT 値を示す領域を示す。また、一番下に各画像の全体領域の MAE および SSIM を示している。

GAN, DDPM, DDCM は、図中の点線で示す横隔膜位置に着目すると、ターゲット画像に対して横隔膜が上方へ過度に変形している。また、胸郭内の枠で示す領域では CT 値の変化が大きく、画素値が過度に修正される傾向が認められる。これに対し、DDVF および DDVF+ では、横隔膜位置ならびに胸郭内 CT 値の双方において、ターゲット画像との差が小さい。さらに、DDVF+ では、胸郭内領域における誤差が DDVF よりも低い傾向が確認された。

4. おわりに

本研究では、拡散モデルによって変位場を予測し、予測された変位場を用いて 3D-CT 画像を変形することによって、呼吸変形を再現する方法を提案した。比較実験において、吸気終端相の 3D-CT 画像に加えて、呼吸変位の一次元指標である変位量を反映した変位分布である統計的変位場を条件として加えることによって、より精度の高い変位場が生成できること

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 (A) 24H00795, 基盤研究 (B) 23K24282 の支援による。

文 献

- [1] H. Teske, P. Mercea, M. Schwarz, N. H. Nicolay, F. Sterzing, and R. Bendl. Real-time markerless lung tumor tracking in fluoroscopic video: Handling overlapping of projected structures. *Medical Physics*, Vol. 42, No. 5, pp. 2459–2540, 2015.
- [2] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng. X2CT-GAN: Reconstructing CT from biplanar X-rays with generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10611–10620. IEEE, 2019.
- [3] R. Miura, M. Nakamura, and M. Nakao. Image-to-volume deformable registration by learning displacement vector fields. *IEEE Transactions on Radiation and Plasma Medical Sciences*, Vol. 9, No. 1, pp. 69–82, 2025.
- [4] M. Nakao, M. Nakamura, and T. Matsuda. Image-to-graph convolutional network for 2D/3D deformable model registration of low-contrast organs. *IEEE Transactions on Medical Imaging*, Vol. 41, No. 12, pp. 3747–3761, 2022.
- [5] T. Oya, Y. Kadomatsu, T. F. Chen-Yoshikawa, and M. Nakao. 2D/3D deformable registration for endoscopic camera images using self-supervised offline learning of intra-operative pneumothorax deformation. *Computerized Medical Imaging and Graphics*, p. 102418, 2024.
- [6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, Vol. 38, No. 8, pp. 1788–1800, 2019.
- [7] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, No. 15, pp. 1–10. Association for Computing Machinery, 2022.
- [8] M. Kanamuro, H. Hirashima, M. Nakamura, and M. Nakao. Diffusion model-based displacement field generation for 4D-CT chest image generation. *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2025.
- [9] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever. Consistency models. *arXiv preprint, Art no.:2303.01469*, 2023.
- [10] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint, Art no.:1611.07004*, 2016.