

マルチボリュームガイド拡散モデルを用いた狭角 CT 画像再構成と 未知投影角度幅への汎化性能

磯貝日向子[†] 村橋 成紀[†] 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

E-mail: jisogai.hinako.35z@st.kyoto-u.ac.jp

あらまし 本研究では、臨床実問題における 狭角 CT (Limited-angle computed tomography: LACT) 再構成と異なる投影角度幅による投影データへの汎化性能の調査を目的として、複数の有効視野から得られる三次元的な潜在表現をガイドとして用いるマルチボリューム潜在拡散モデルを提案する。評価実験の結果、複数の投影角度幅から得られる狭角 CT 画像により自己訓練されたモデルは特定の角度に特化して訓練されたモデルより高い性能を示した。学習時に含まれない未知角度幅に対しても安定した再構成性能を有することが確認されたので報告する。

キーワード 潜在一貫性モデル, 狭角 CT 画像再構成, 画像生成

Multi-Volume-Guided Diffusion Model for Limited-Angle CT Reconstruction and Generalization Ability to Unseen Angle Ranges

Hinako ISOGAI[†], Naruki MURAHASHI[†], Mitsuhiro NAKAMURA[†], and Megumi NAKAO[†]

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

E-mail: jisogai.hinako.35z@st.kyoto-u.ac.jp

Abstract This study investigates limited-angle computed tomography (LACT) reconstruction in clinical settings and its generalization to different projection angle ranges. We propose a multi-volume latent diffusion model guided by three-dimensional latent representations derived from multiple effective fields of view. Experimental results show that a self-supervised model trained with narrow-angle CT images from multiple angle ranges outperforms models specialized for a specific angle range. Furthermore, the proposed method maintains stable reconstruction performance even for unseen projection angle ranges not included during training.

Key words Latent Consistency model, Limited angle CT reconstruction, Image synthesis

1. はじめに

X-ray Computed Tomography (X 線 CT) は定量性と分解能の高さから臨床において広く用いられてきたが、放射線治療時や手術中など即時性や可搬性が求められる状況では利用が制限される。この制約に対し、撮像の高速化や省スペース化を目的として照射角度を制限した Cone-beam CT の利用や画像再構成への機械学習の導入が取り組まれてきた。しかし、狭角 X 線画像撮影は再構成に必要な情報が不足しているため、臨床においては 2 次元画像としての参照に留まっている。

狭角 CT (Limited-angle computed tomography: LACT) 再構成は、投影角度の欠損に起因する強い不良設定性を有する逆問題であり、十分な事前知識なしに高精度な画像を復元することは困難である。近年、拡散モデル (Diffusion Models) [1] に代表される機械学習手法が高い画像生成能力を示し、LACT

再構成への応用が進められている。拡散モデルは高い画像生成能力を有することが示されており、拡散モデルを用いた LACT 再構成の手法が提案され、視覚的に高品質な画像生成が達成されている [2], [3]。しかし、画像生成における臓器の 3 次元構造の正確な復元やコントラストの達成は依然として課題であり、有効視野 (Field of view, FOV)、撮像方向や範囲など、臨床において様々に想定される撮像条件の違いが生成画像に与える影響も十分に明らかになっていない。

本研究では、臨床における実問題に対する狭角 CT 画像再構成を目的として、複数の有効視野から得られる潜在変数をガイドに用いるマルチボリューム潜在一貫性モデルを提案する。提案モデルでは一貫性モデル [4] を用いて高速な画像生成を可能にする。また、3D-CT 画像から潜在表現へ変換し、逆拡散プロセスの算出を潜在空間において実施することによって、効率のかつ高解像度の画像生成を達成できる。FOV が広

い画像では、再構成後に臓器の細部構造が不鮮明になりやすいという問題に対し全体と局所の大きく異なる画像領域から潜在変数を獲得する Multi-volume encoder を導入し、変換対象のスライス外の領域からも 3 次元的に潜在変数を得ることで、臓器の 3 次元構造のより正確な復元を目指す。また、臨床での応用を想定し、様々な照射条件に対応可能なモデルであるか調査する。

2. 提案手法

2.1 問題設定

本研究では、挟角 CT イメージングにおいて、拡散モデルを用いた画像変換によって、全投影条件で得られる CT 画像に近い品質を有する画像を生成する問題を対象とする。全投影画像によって得られる CT 画像を参照画像 y_0 、LA 設定で得られる X 線投影画像群から再構成された LACT 画像を x と定義する。提案モデルの目的は LACT 画像 x を条件として、 y_0 に近い品質を有する synthetic CT (sCT) 画像 $\hat{y}_0$ を生成すること、未学習角度データでの推論による提案手法の汎化性能の確認である。

本研究では、Cone-beam CT イメージングと同様の X 線撮像を LA 撮像設定で実施する場合を想定している。撮像時の設定として、仰臥位における左手方向を 0° と定義し、時計回りに以下の角度条件によるイメージングが行われるケースを対象とする。

- θ_s : 投影開始時点の角度
- θ_r : 投影開始から終了までの角度範囲
- $\Delta\theta$: 各投影画像間の角度間隔

ここで、 θ_r は可変とし、[3] で調査された最小角である $\theta_r = 60^\circ$ を代表的な設定として、 $\theta_r = 120^\circ, 90^\circ, 60^\circ$ の LA 設定と、 $\theta_r = 45^\circ, 30^\circ, 15^\circ$ の extreme LA 設定を対象に sCT 画像の復元タスクの学習と推論性能の検証を目指す。さらに、LA, extreme LA 設定でトレーニングされた学習済みモデルの汎化性能を調査するために、テストのみに用いる $\theta_r = 105^\circ, 75^\circ, 37.5^\circ, 22.5^\circ$ の 4 通りの未知投影角度幅についても LACT 画像データベースを構築する。なお、 θ_s 及び $\Delta\theta$ については、臨床における撮像条件を考慮して、 $\theta_s = -\theta_r/2 + 90^\circ, \Delta\theta = 0.5^\circ$ と設定した。

上述の計 10 通りの LA setting によって CT volume からデジタル再構成 X 線画像 (Digitally Reconstructed Radiograph, DRR) を生成し、得られた DRR 画像群を Feldkamp-Davis-Kress (FDK) 法 [5] による逆投影計算によって LACT volume を得た。LA 条件では欠損角度が大きいほど逆投影に必要な情報が不足するため、構造が大きく欠損した不正確な CT 値を有する再構成画像が得られる。事前知識を用いた不完全な LACT 画像から CT 画像への変換問題において、両者に含まれる画像特徴や 3 次元構造に関する相関を効果的に捉える学習を促すガイド設計が本研究が対象とする中心的な課題である。

2.2 マルチボリューム潜在一貫性モデルの概要

本研究では挟角 CT 再構成に条件付き潜在一貫性モデル (Conditional Latent Consistency Model, CLCM) を用いた。CLCM は、LACT 画像から得られる潜在表現をガイドとする

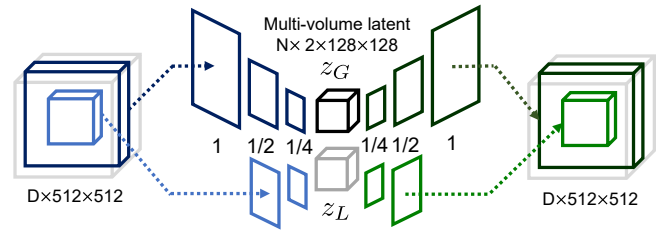


図 1 提案するマルチボリュームエンコーダの概念図

条件付き潜在拡散モデルであり、一貫性モデル [4] に基づき逆拡散過程を 1 ステップで実行する。さらに、次元圧縮された潜在空間で学習を行うことで計算効率を向上させつつ、臓器構造の保持と生成の安定性を両立する。ここで、従来のエンコーダ [6] では画像全体を一様に圧縮するため、LACT 再構成において重要となる臓器内部の微細構造や高周波成分が潜在表現への変換過程で失われやすく、コントラスト低下を招く。本研究では画像の全体領域と局所領域を異なる解像度で同時に扱うマルチボリューム潜在表現を導入する。

図 1 に提案する Multi-volume encoding の概念図を示す。本エンコーダは大域構造抽出を目的とした Global path と、局所構造抽出を目的とした Local path から構成される。3D-CT 画像から生成対象スライスを中心とする N 枚のサブボリュームを入力とし、その中央領域を切り出した局所ボリュームを同時に処理する。Global path ではダウンサンプリングを伴う畳み込みによって大域的形状情報を抽出し、Local path では最小限のダウンサンプリングにより高周波構造を保持したまま特徴変換を行う。デコーダでは、大域成分と局所成分を別経路で復号し、最終的に統合することで、画像全体の整合性を保ちつつ臓器の局所的高周波構造を強調した再構成を実現する。

拡散モデルの学習時には、 z_0 に対してランダムな 2 つの時刻 $t = t_n, t_m$ ($1 \leq t_n, t_m \leq T-1$), ($t_n \neq t_m$) に対応するノイズを付与する。ここで、 T は画像が完全にノイズになる時刻のことであり、本研究では先行研究 [7] を参考に $T = 1000$ を用いた。ノイズが付与された 2 時刻の潜在表現に対し、条件 x と一貫性制約の下で z_0 を直接予測するようにモデルをトレーニングする。推論時には狭角 CT 画像から潜在表現を取得し、内部で生成したガウシアンノイズに対して、取得した狭角 CT 画像の潜在表現をガイドとして 1 ステップでノイズ除去を行う。得られた潜在表現をデコードし、最終的な sCT 画像を得る。以上の推論処理はスライス単位でなされるが、3 次元的に連続的な画像補完となるようすべてのスライスにおいて同一の初期ノイズを用いる。

3. 実験と結果

3.1 対象データと実験条件

実験には京都大学医学部附属病院において膀胱がんの放射線治療を受けた 135 例の 3D-CT ボリュームを用いた。本研究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)。3D-CT ボリュームは 512×512 pixel からなる 93-174 枚のスライス画像から構成され、各スライスの解

表 1 Multi-volume encoding が sCT 画像の画質に及ぼす影響

Encoding	N	512 × 512		256 × 256 (ROI)	
		MAE [HU] ↓	SSIM ↑	MAE [HU] ↓	SSIM ↑
Single	1	12.26	0.9586	44.07	0.8385
Multi	1	11.11	0.9637	39.97	0.8591
Multi	11	10.34	0.9672	36.85	0.8723
Multi	31	10.29	0.9672	36.76	0.8731
Multi	61	10.36	0.9671	37.05	0.8724

像度は $1.0\text{mm} \times 1.0\text{mm}$ である。すべての実験において、CT 画像内の CT 値の範囲を $[-1000\text{HU}, 1000\text{HU}]$ にクリップ後、 $[-1, 1]$ に正規化して用いることとした。CT 画像と潜在変数間の変換に用いるマルチボリュームエンコーダには、先行研究 [8] において腹部領域を対象に撮像された CT スライス計 27,438 枚のデータベースを用いてトレーニングされた学習済みモデルを用いた。

本データベースをトレーニングデータ 118 例、検証データ 4 例、テストデータ 13 例に分割し、検証データに対する平均絶対誤差 (Mean Absolute Error, MAE) が最良となったモデルを用いてテストを行った。テスト時の評価指標には MAE に加えて、構造的類似指標 (Structural Similarity, SSIM) を用い、ボリューム単位で 1 つの MAE と SSIM が算出される。なお、 512×512 pixel の画像全体を評価対象とする場合に加え、画像中央の 256×256 pixel を対象に評価指標を算出した。この関心領域 (Region of interest, ROI) 設定は撮像時の FOV が広く、構造が画像中央のみに画像化されている場合をより厳密に評価することを目的としている。

3.2 マルチボリュームガイドの有効性評価

最初に、Multi-view encoding が生成精度に与える影響を調査するため、再構成対象のスライスに対応する 1 枚 ($N = 1$) の LACT 画像から得られる潜在変数のみにガイドを固定して sCT 画像を生成した。本検証には $\theta_r = 60^\circ$ の代表的設定によって LACT 画像を得ることとし、学習は 400epoch とした。

表 1 に、各エンコーダ構成における生成画像と参照画像間の誤差の中央値を示す。また、図 2 にテストデータから無作為に抽出した CaseA について、提案モデルによる水平断面の sCT 画像、参照画像 (Reference) を示す。骨盤領域において、Multi-view を用いた場合に骨格構造の輪郭がより明確に表現されている。また、画像上部の軟組織のテクスチャについても、Single-view では不鮮明である一方、Multi-view では細部構造の生成が可能であることが確認された。次に Multi-view encoding を固定し、ガイドとして与えるスライス枚数 N を 1, 11, 31, 61 と変化させた。 $N = 31$ において各評価指標で最良の画質 (MAE: 36.76 HU, SSIM: 0.8731) を示すことを確認した。以上から、Multi-volume encoding によって、MAE と SSIM が改善することを確認した。

3.3 提案モデルの汎化性能

本研究では、提案モデルの臨床利用を想定して、投影角度幅の異なる LACT 画像を同時に用いた CLCM の評価実験を行った。さらに学習時に用いない投影角度幅の LACT 画像をテストに用い、臨床において想定される撮像条件への汎化性能

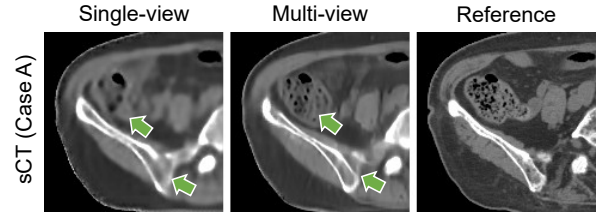


図 2 Multi-view encoding が生成画像に与える影響

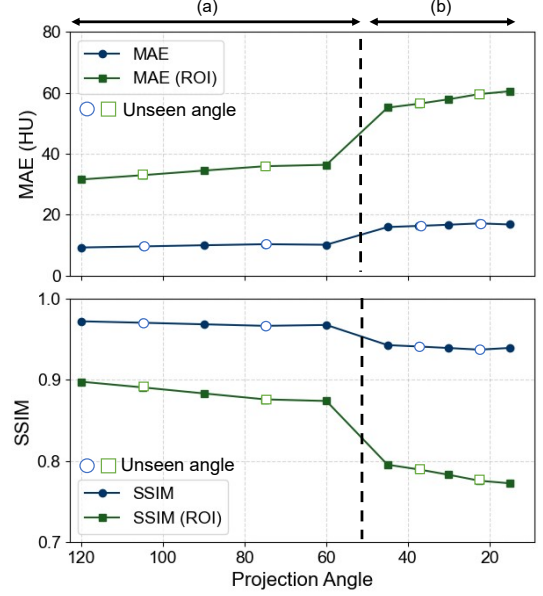


図 3 投影角度幅の違いによる画像再構成精度の変化 (a) 狭角学習データ, (b) 超狭角学習データ

を調査する。以下の LA 設定, Extreme LA 設定による LACT 画像をガイドする設定を用意した。

a) LA 設定: $\theta_r = 120^\circ, 90^\circ, 60^\circ$

b) Extreme LA 設定: $\theta_r = 45^\circ, 30^\circ, 15^\circ$

それぞれ異なる三つの投影角度幅によって撮像される LACT 画像をすべて混合した学習データベースを構成し、投影角度を特定せずにモデルを 200epoch でトレーニングする。学習済みモデルに各投影角度幅の LACT テスト画像を適用し、得られる sCT 画像の誤差をそれぞれ評価した。さらに、学習時に含まれていない未知の投影角度幅に対する再構成性能を評価するため、a) の学習済みモデルには $\theta_r = 105^\circ, 75^\circ$ の LACT 画像を、b) の学習済みモデルには $\theta_r = 37.5^\circ, 22.5^\circ$ の LACT 画像から sCT 画像を得る推論を行った。この実験設定により、対象とする投影角度幅を拡張したより汎用性の高いモデルの構築や、学習条件と異なる LA 設定に対する再構成性能の検証が可能である。

表 2 に、10 通りの LACT 画像から得られる sCT 画像と参照画像間の評価指標の中央値を示す。図 3 に各投影角度幅における sCT 画像の評価値をプロットした結果を示す。図 4 には、テストデータから無作為に選択した 2 例について、 θ_r の順に得られた sCT 画像を示している。特に 60° の LA 画像単独で学習・推論を行い得られた sCT 画像の MAE が 36.76[HU]

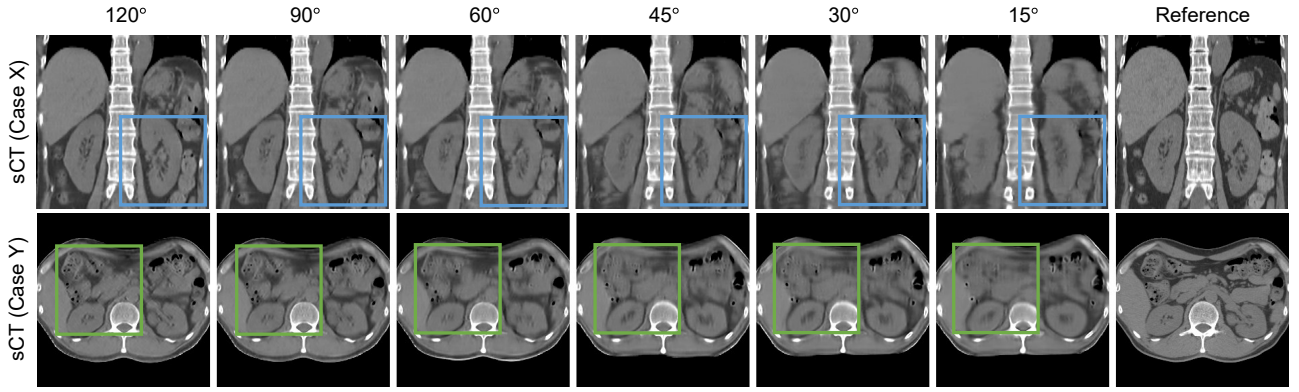


図 4 異なる投影角度幅から得られる LACT 画像をガイドとして生成された sCT 画像

表 2 投影角度幅が sCT 画像に与える影響と未知の投影角度幅に対する再構成性能

θ_r	LA					Extreme LA				
	120°	105°	90°	75°	60°	45°	37.5°	30°	22.5°	15°
MAE ↓	9.215	9.576	9.953	10.30	10.12	15.92	16.29	16.69	17.12	16.77
MAE (ROI) ↓	31.55	32.95	34.46	35.89	36.35	55.11	56.36	57.86	59.56	60.49
SSIM ↑	0.9722	0.9704	0.9685	0.9666	0.9677	0.9429	0.9412	0.9393	0.9373	0.9394
SSIM (ROI) ↑	0.8978	0.8908	0.8834	0.8761	0.8741	0.7957	0.7900	0.7833	0.7760	0.7728

であったのに対して、 $\theta_r = 120^\circ, 90^\circ, 60^\circ$ の LA 設定 で学習されたモデルでは 36.35[HU] と再構成誤差の低下が確認された。このような評価値悪化の抑制は (b) の領域においても確認され、特定の角度幅に特化した学習よりも、複数の投影角度幅の LACT 画像からの復元を一つのモデルで自己トレーニングすることの有効性が示されている。また、図 3 において (a) の領域は LA 設定、(b) の領域は Extreme LA 設定 で学習したモデルによる sCT 画像の評価値を示しており、その境界において品質に顕著な差が生じている。(a) の領域において θ_r の減少に対して MAE は徐々に増加しているが、ガイドに含まれる情報の欠損と比較して増加は抑制されている。 $\theta_r = 105^\circ, 75^\circ, 37.5^\circ, 22.5^\circ$ の未知投影角度幅に対しても MAE の増加は概ね線形を維持していることが確認される。提案モデルが学習時に含まれていない投影角度幅に対しても安定した画像生成を達成する汎化性能を有することが確認できる。図 4 において、LA 設定においても、主要な解剖構造の再現性が保たれており、投影角度幅の変化に対して頑健な再構成が可能であることが視覚的にも確認できる。45° 以下など超狭角条件では、臓器輪郭や構造に不鮮明な箇所が現れており、今後の課題である。

4. おわりに

本研究では、LACT 条件下における高精度な画像再構成を目的として、Multi-Volume encoding と潜在一貫性モデルを組み合わせた画像再構成モデルとその自己教師あり学習方法を提案した。画像全体と中央領域のサブボリュームでエンコードする枠組みによって、臓器内部の高周波構造や 3 次元の境界情報を効率的に学習可能とした。提案モデルは学習時に含まれていない未知の LA 設定に対しても安定した再構成性能を示し、実臨床における多様な撮像条件への応用可能性を有

することを確認した。

謝 辞

本研究は日本学術振興会科研費補助金基盤研究 基盤研究 (A) 24H00795, 基盤研究 (B) 23K24282 の支援による。

文 献

- [1] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. *Proc. International Conference on Machine Learning (ICML)*, pp. 2256–2265, 2015.
- [2] Y. Sun, H. Baroudi, T. Netherton, L. Court, O. Mawlawi, A. Veeraraghavan, and G. Balakrishnan. DIFR3CT: Latent diffusion for probabilistic 3D CT reconstruction from few planar X-rays. *arXiv preprint, Art no.:2408.15118*, 2024.
- [3] J. Liu, R. Anirudh, J. J. Thiagarajan, S. He, K. A. Mohan, U. S. Kamilov, and H. Kim. Dolce: A model-based probabilistic diffusion framework for limited-angle ct reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10498–10508, 2023.
- [4] T. Luo, A. Blattmann, R. Rombach, K. Zhang, T. Tang, K. Karra, A. Schwing, Y. Bengio, and B. Ommer. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378*, 2023.
- [5] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am. A*, Vol. 1, p. 612–619, 1984.
- [6] A. van den Oord, O. Vinyals, and K. Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, Vol. 30, pp. 6306–6315, 2017.
- [7] N. Murahashi, M. Nakamura, and M. Nakao. A trial for improvement in image quality of cone-beam ct images using conditional diffusion model. In *IEICE Technical Report*, Vol. 123, pp. 109–112. IEICE, 2024. in English.
- [8] H. Isogai, N. Murahashi, M. Nakamura, and M. Nakao. Limited-angle ct reconstruction using a latent diffusion model. *Transactions of Japanese Society for Medical and Biological Engineering*, Vol. Annual63, No. Proc, pp. 368–370, 2025.