2024年度 卒業論文

論文題目

生成モデルによるデータ拡張および 選別によるリンパ節転移の分類精度向上

指導教員 舟橋 健司 准教授

名古屋工業大学 工学部 情報工学科 2021年度入学 33114099番

服部 駿希

目 次

第1章	はじめに	1
第2章 2.1 2.2 2.3	背景知識 造影 CT 画像 DICOM 画像 リンパ節のがん転移有無による特徴の変化	5 5 6
第3単 3.1 3.2 3.3	データの概要	8 8 9
第4章 4.1 4.2	先行研究とその課題 DDIM の作成と評価	11 11 12
第5章 5.1 5.2	二つの生成モデルによるデータ拡張の比較 DDPM の作成と評価	16 16 18
第6章 6.1 6.2 6.3 6.4	二つの選別手法の提案と比較 U-Net によるリンパ節検出を用いた選別 DreamSim による類似度比較を用いた選別 選別枚数による比較 リンパ節分類実験による選別効果検証 6.4.1 選別の有無による比較 6.4.2 選別の DDIM と DDPM に対する効果の比較	21 21 24 24 25 26 26 27
第7章	むすび	29
謝辞		30
参考文南	Ŕ	31

第1章 はじめに

がん診断において,X線CTをはじめとする画像診断技術が広く活用されている. X線CTとは,X線管が人体の周囲を回転しながらX線を照射し,人体を通過する ことで減衰したX線を検出することで,人体の断層を撮影する医用画像技術である. 現在多くの施設で使われるCT装置はマルチスライスCTと呼ばれ,短時間で全身 の撮影ができ,立体的な画像を容易に作成できる.そのため病気と腫瘍の発見や進 行具合の確認など多岐にわたって活用されている.また,最小で0.5mmの間隔で断 層画像を作成できるため,微小な病変や初期段階の疾患を発見することが可能であ る.一方で,画像診断には高い専門性が求められるため,医師の経験に依存した診 断精度のばらつきが生じるといった問題点もある.例えば短径5mmから10mm未 満のリンパ節に関するがん転移有無の診断では,医師の経験量の違いによって正診 率が51-86%と大きくばらつきがある[1].これらの問題に対処するために,医用画 像に対して機械学習を用いて診断支援を行う研究が盛んに行われており,CT画像 からリンパ節を検出する研究[2-5]や,CT画像をMRI画像へ変換する研究[6]など がある.

続いて、本研究で対象としている直腸がんとそのリンパ節転移について述べる.直 腸がんとは、大腸の末端部である直腸の組織内に悪性がん細胞が発生する疾患であ る.日本では1975年以降、罹患数と死亡数が増加傾向にあり[7]、2023年には直腸 がんにより15,737人もの死亡数が報告されている[8].そのためこの疾患に対する 機械学習を活用した診断支援の確立が強く求められている.がんはリンパの流れに 沿ってリンパ節へ転移することが可能性があるため、直腸がんと診断された場合に は、他の部位への転移の有無を確認する病期診断が重要となる.特に側方リンパ節 への転移が認められた場合、直腸がんのステージは進行しているものと考えられ、 治療方針や予後に大きく影響を与える[9].例えば、側方リンパ節転移が陽性の場合、 5年生存率は36-42%と低く,転移が陰性の場合の約80%と比較すると予後が著しく 悪い [10, 11]. さらに,直腸がんは治癒切除後でも局所再発の頻度が約10%と言わ れており,同じく大腸の一部である結腸に悪性がん細胞が発生する結腸がんの再発 頻度が1%以下であることと比較するとかなり再発リスクが高いといえる[12].直腸 がんの再発リスクを減少させ、5年生存率を改善する治療法として,側方リンパ節 郭清がある.この治療はがん周辺のリンパ節に加え側方リンパ節を切除することで 骨盤内再発のリスクを約50%減少させ、5年生存率を8~9%改善できるとされてい る[13].一方で,側方リンパ節郭清には排尿機能障害や性機能障害といった後遺症 が伴う可能性があるため、すべての症例に対して実施することは推奨されない.転 移が陰性の場合には、側方リンパ節郭清を省略することで、後遺症のリスクを回避 する判断も重要となる[14].したがって、手術前に側方リンパ節への転移の有無を 正確に診断し、側方リンパ節郭清が必要かどうかを適切に判断することが求められ ている.しかし、側方リンパ節に対してのがん転移を正確に診断する手法は確立さ れておらず、新たな診断技術の開発が望まれている.

近年,CT 画像を用いた機械学習によるリンパ節分類の研究が活発に進められて いる.例えば,非小細胞肺がんの縦隔リンパ節転移分類の性能について CNN と従 来の機械学習手法および医師の診断を比較する Wang らの研究 [15] や,食道がんの リンパ節転移を CNN で予測する磯崎らの研究 [16] がある.一般に十分なデータが 確保できない場合,データの特徴を捉えきれずモデルの能力が十分に発揮されなく なることが多い.そのためリンパ節分類の精度をよくするためには,多くのデータ とそれに対するアノテーションが求められる.しかし,医用画像のアノテーション ができるのは専門医師に限られるため,大規模なデータセットの構築が困難である. その解決策として,近年急速に発展している Diffusion モデルを用いたデータ拡張 手法が提案されている.Diffusion モデルとは,画像生成やデータ生成の分野で使用 される機械学習モデルの一種であり,データにノイズを少しずつ加えて,そこから ノイズを取り除くことで新たなデータを生成することができる.近年 Diffusion モ デルは生成品質や効率の向上を目指して大きく進化しており,Stable Diffusion [17] や Imagen [18] のように,テキストや画像を条件にした画像生成が可能なモデルも 登場している.Denoising Diffusion Implicit Models (DDIM) は,Diffusion モデル の基盤となる Denoising Diffusion Probabilistic Models (DDPM)を改良したもの で、より短時間での画像生成を実現するモデルである.しかし、DDIM は高速化の 利点がある一方で、生成画像の多様性が制限されている可能性があることが課題で ある [19].

また,生成画像の品質をさらに向上させるためには,生成画像から選別を行い, 意図しない画像を除外することが重要である.例えば,岩崎らは,ナラティブ(物 語,語り)に合致した画像の自動生成手法を提案しており,複数枚の生成画像を Vision Language Modelにより評価することでより合致度の高い画像の選別を行っ ている [20].また,海老沢らの研究では生成画像に対して二段階の評価に基づいた 選別手法を用いることで,機械学習を用いたメロン等級判定が 81.3%から 86.3%へ 精度向上することが確認されている [21].これらの研究のように生成画像に対する 選別が広く行われているが,医用画像に対してその影響に言及したものは少ない.

リンパ節転移の分類に関する当研究室の先行研究では先述の DDIM を用いて CT 画像に写るリンパ節を生成することでデータ拡張する手法を検討している [22]. が ん転移ありリンパ節画像とがん転移なしリンパ節画像をそれぞれ 2000 枚生成し,そ れを Vision Transformer (ViT)の学習に用いた結果,分類精度が 88.89%まで向上し, 専門医師の正診率 86%を超える精度が得られた.この成果は,Diffusion モデルを用 いたデータ拡張が医用画像分野における診断精度のさらなる向上に寄与する可能性 を示している.しかし,生成画像の品質に関する課題として,意図しない画像が生 成される問題が指摘されている.

そこで、本研究では先行研究で用いた Diffusion モデル(DDIM)に加えて異なる Diffusion モデル(DDPM)も採用し、生成画像の多様性を改善することで分類精度 の向上を図る.その上で DDIM と DDPM で生成した画像を用いたリンパ節分類の 精度を比較し、生成画像の多様性が分類精度に与える影響について考察する.また、 意図しない画像を除外することによる分類精度の向上を期待して、機械学習を用い た生成画像の二つの選別手法を提案する.一つ目は、生成画像に対して U-Net [23] によりリンパ節検出を行うことでリンパ節として適切な画像を選別する手法であり、 二つ目は DreamSim [24] により生成したリンパ節画像を、がん転移ありリンパ節と がん転移なしリンパ節のどちらに類似しているか比較することで、選別を行う手法 である.それぞれの手法による選別後にリンパ節分類実験を行うことで,それぞれ の効果を比較検証する.

第2章 背景知識

本章では,診断に用いるリンパ節の画像に関する背景知識として,画像形式とリ ンパ節のがん転移有無による特徴の変化を説明する.

2.1 造影 CT 画像

CT 画像を撮影する際に,診断精度を高めるために,医師が造影剤と呼ばれる薬 剤を使用する場合がある.造影剤はX線透過性が低い性質があり,造影剤を撮影前 に静脈に注射することで各臓器に行き渡り,血液が豊富な組織ほどX線透過性が低 くなる.そのため血管や臓器,リンパ節などの組織に対してコントラストが生じる ため,造影剤を使用しない場合ではわかりにくい小さな病変をより明瞭に視認でき るようになる.造影剤を使わずに撮影する方法を単純CTと呼び,造影剤を使用し て撮影する方法を造影CTと呼ぶ.造影剤を使うことによるCT 画像の違いを図2.1 に示す.リンパ節について取り扱う本研究では,リンパ節の特定が容易となる造影 CT で撮影されたデータセットを使用する.

2.2 DICOM 画像

DICOM (Digital Imaging and Communications in Medicine) は医用画像の保存, 転送,管理などを標準化するために制定された国際規格であり,患者の情報や撮影 日時といったメタデータを画像データと同時に管理できる.一般的なデジタル画像 の多くが8ビットの階調で表現されるのに対して、DICOMは10ビット以上の階調 で画像を表現することが可能である.医師がCT 画像による診断を行う際には,体内 組織や病変を識別しやすくするためにウィンドウレベル変換という処理を行う.ウィ ンドウレベル変換とは,画像の輝度の中央値であるウィンドウレベルと,表示する 輝度の範囲であるウィンドウ幅を調整することで,特定の組織の強調を行う処理で



(a) 単純 CT

(b) 造影 CT

図 2.1: 造影剤による CT 画像の違い

ある.実際のリンパ節診断でもこの処理が用いられており,リンパ節の腫大や内部 構造の異常をより正確に把握することが可能になる.

2.3 リンパ節のがん転移有無による特徴の変化

リンパ節にがん細胞が転移している場合と転移していない場合ではさまざまな異 なる特徴が現れる.最も大きく表れる特徴は大きさである.正常なリンパ節は基本 的に10mm以下であるのに対し,転移したリンパ節は10mm以上に大きくなるとい われている[25].また,正常なリンパ節が一様な模様であるのに対し,転移したリ ンパ節は腫瘍内部が壊死することもあり,血流が届かなくなるため,まばらな模様 が見えることもある[26].その他にも,がん転移することで,形態が楕円体から球 体に近づき,周囲との境界が不明瞭になると考えられている[27].図2.2,2.3に転 移別のリンパ節を示す.



図 2.2: 転移なしリンパ節の例(画像中央)



図 2.3: 転移ありリンパ節の例(画像中央)

第3章 データセット

本研究では愛知県がんセンターから提供された 54 人分の造影 CT 画像を使用する. そのデータの概要と,前処理について説明する.

3.1 データの概要

提供されたデータは、スライス間隔が 1mm で、患者一人当たり 500 から 1400 ス ライスの 512 × 512 ピクセルの DICOM 形式 CT 画像である.また、そのデータに 対し 1 人の医師によってリンパ節の位置とがん転移の有無がアノテーションされた PowerPoint プレゼンテーションファイル(以下、pptx ファイルと呼ぶ)が提供さ れている.なお、提供されたデータ内に写るすべてのリンパ節にアノテーションさ れている.リンパ節は転移のないリンパ節でも平均短径が 1.9mm ほどの大きさ [28] であるため、複数のスライスに描写されることが多いが、pptx ファイルにアノテー ションされているのは、リンパ節の中心が写る 1 スライスのみである.pptx ファイ ルの例を転移別に図 3.1 に示す.627 個のリンパ節がアノテーションされており、そ の内訳はがん転移なしリンパ節(以下、G1 と呼ぶ)が 603 個、がん転移ありリンパ 節(以下、G3 と呼ぶ)が 24 個である.

3.2 マスク画像の作成

リンパ節領域を学習に用いるときの処理を容易にするため, pptx ファイルを元に マスク画像の作成を行う.マスク画像作成の手順は以下の通りである.CT 画像と 対応するマスク画像の例を図 3.2 に示す.

1. CT 画像を DICOM 形式から TIFF 形式に変換する.



(a) 転移なしリンパ節(G1)



(b) 転移ありリンパ節(G3)



(a) CT 画像



(b) マスク画像

図 3.2: CT 画像とマスク画像の例

図 3.1: アノテーションされた pptx ファイルの例

- 画像編集ソフトを用いて pptx ファイルでアノテーションされているリンパ節
 を, (R,G,B)=(255,255,255) で塗りつぶす.
- 3. それ以外の部分を (R,G,B)=(0,0,0) で塗りつぶす.

3.3 データセットの前処理

分類用モデルにリンパ節画像を学習させるため、リンパ節領域のクロップを行う. 先行研究において、リンパ節のみではなく、周辺領域もモデルに入力すると分類精 度が向上することが示されている [29]. そのため、DICOM 画像からリンパ節の座標 を取得し、リンパ節を中心に 64 × 64 ピクセルにクロップする. さらに、実際の画 像診断と同様にリンパ節領域を明瞭にするため、ウィンドウレベル変換を行う.本 研究では実際のリンパ節診断に用いられる値を参考にして、ウィンドウレベルを63、 ウィンドウ幅を256としてウィンドウレベル変換を行った.クロップ後の、ウィン ドウレベル変換前後のリンパ節画像を図3.2に示す.



(a) 変換前



(b) 変換後

図 3.3: ウィンドウレベル変換による違い

第4章 先行研究とその課題

本章では,先行研究で用いた生成モデル DDIM (Denoising Diffusion Implicit Models)の詳細と,指摘された品質に関する課題を説明する.

4.1 DDIM の作成と評価

DDIM はデータに対して段階的にノイズを加える Forward process と,完全なノ イズから徐々にノイズを除去することでデータを再構成する Reverse process によっ て画像を生成するモデルである.この手順を図で表したものが図 4.1 である.DDIM では Reverse process が決定論的に進行するため,画像生成を高速に行うことができ る.しかし,決定論的な進行は生成画像の品質や多様性に制約を与える可能性があ る [19, 30].このモデルで G1 を生成する際は,G1 のデータのみを訓練データとし て 120 エポック学習する.しかし,G3 はデータセットに含まれる数が少ないため, 学習を 2 段階に分ける手法を提案する.手順を以下に示す.

1. G1とG3両方のデータを訓練データとして100エポック学習する.

2. G3 のデータのみを訓練データとして 50 エポック学習する.

なお,手順2においてはG3のデータ数が少ないため,リンパ節の中心の前後のス ライスも訓練データに追加する.手順1が完了した時点の生成画像の例を図4.2に 示し,手順2が完了した時点での生成画像の例を図4.3に示す.図4.2からリンパ節 とその周辺領域の特徴に基づいた画像が生成できていることが確認できる.しかし, この時点では.G1とG3のデータ数の差によってG1の特徴に沿った画像が生成さ れている.図4.3から,追加でG3のデータのみで学習することでG3の特徴に沿っ た画像が生成されることが確認できる.実際に機械学習によるリンパ節診断が実用 化されるためには、高い分類精度だけでなく、優れた汎用性も必要となる.しかし,

Forward process



Reverse process

図 4.1: Diffusion モデルの学習プロセス

データ数が少ないことから過学習によって訓練データと酷似した画像が生成される 可能性がある.そのため、生成画像と既存のデータとの類似度を PSNR と SSIM を 用いて評価する.PSNR はピクセル単位での誤差を計算し類似度を評価する指標で あり、SSIM は画像をパッチ単位で構造の類似性を計算する指標である.生成画像 と既存データの全ての組み合わせの中で最も PSNR が高かったのは 15.67 であった. その画像の組を図 4.4 に示す.また、最も SSIM が高かったのは 0.2977 であった.そ の画像の組を図 4.5 に示す.どちらの数値も低く、生成画像は訓練データに酷似し たものは存在しないことを示している.そのため、この学習手順による生成画像を データ拡張として用いるのは妥当であると考え、リンパ節分類に使用する.

4.2 先行研究での課題

先述の方法で作成した DDIM では,G3 として生成されることを期待しているに もかかわらず,G1 として分類されるべき特徴を持った画像が生成されてしまうとい う課題があった.その例を図 4.6 に示す.図 4.7 の既存データの例と図 4.6 を見比べ ると,生成画像は画像中央に写るリンパ節のようなものがかなり小さく描写されて いることが確認でき,これは G1 の特徴に近いことがわかる.先行研究では,この



図 4.2: 手順1 完了時点での生成例 (DDIM)



図 4.3: 手順2 完了時点での生成例 (DDIM)

期待とは異なる生成画像がリンパ節分類に悪影響を与えた可能性が考えられており, この課題を解決できればさらなる分類精度の向上が期待できる.この課題の原因と して考えられるのは,G3の生成にG1のデータが使われていることである.しかし データ数が少ないためG3のみで学習すると過学習により多様なリンパ節画像を生 成することができなくなる.



(a) 訓練データ



(b) 生成画像

図 4.4: PSNR が最大(15.67)であった組 (DDIM)



(a) 訓練データ



(b) 生成画像

図 4.5: SSIM が最大(0.2977)であった組 (DDIM)



図 4.6: 意図しない生成画像の例



(a) G1 の例



(b) G3 の例

図 4.7: 生成画像との比較のための既存データ例

第5章 二つの生成モデルによるデータ拡張の 比較

本章では,DDIMよりも多様な画像生成が可能であるDDPM (Denoising Diffusion Probabilistic Models) について説明する.またDDIMとDDPM によるデータ拡張 後の Vision Transformer によるリンパ節分類における両者の分類性能の比較につい て述べる.

5.1 DDPM の作成と評価

DDPMはDDIMの基盤となるモデルであり,DDIMと同様にForward processと Reverse processから構成される [31].DDIMと異なる特徴は,Reverse process において条件付き確率分布に基づいてノイズを取り除くことである.DDIMは決定論的に画像生成を行うため,同じ条件では常に同じ結果が得られる.それに対し,DDPMは確率的であるため,同じ条件でも毎回異なる結果が得られる.そのため,DDPMはDDIMより多様な画像が生成できると考えられる.このモデルでG1を生成する際は,G1のデータのみを訓練データとして,700エポック学習する.G3を生成する際は,DDIMと同様に学習を2段階に分けて行う.手順を以下に示す.

1. G1とG3両方のデータを訓練データとして600エポック学習する。

2. G3 のデータのみを訓練データとして 300 エポック学習する。

手順2においてはG3のデータ数が少ないため、リンパ節の中心の前後のスライス も訓練データに追加する. 手順1が完了した時点と, 手順2が完了した時点での生 成結果の例を図5.1, 5.2に示す. 図5.1から手順1によりG1の特徴に沿った画像 が生成されることが確認できる. また図5.2から, DDPMでもG3の特徴を満たす 画像が生成できることが確認できる. 酷似した画像がないことを確認するため, 生 成画像と既存データの類似度を PSNR と SSIM によって評価する. 生成画像と既存 データで PSNR または SSIM が最大であった組を図 5.3, 5.4 に示す. 最大の PSNR は 16.42,最大の SSIM は 0.3274 であった. どちらの評価においても DDIM よりも 数値が高く,既存データにより近い画像が生成されているといえる. しかし,これ らの数値は十分低く,酷似しているわけではない. また,目視での評価では構造に 似ている部分があると感じるが,中央のリンパ節部分には形状の違いがあることも わかる. このことから DDPM による生成画像もデータ拡張として用いるのは妥当 であると考えた.



図 5.1: 手順1完了時点での生成例 (DDPM)



図 5.2: 手順 2 完了時点での生成例 (DDPM)



(a) 訓練データ

(b) 生成画像





(a) 訓練データ



(b) 生成画像

図 5.4: SSIM が最大(0.3274) であった組 (DDPM)

5.2 リンパ節分類実験における比較

DDPMとDDIMいずれのモデルの生成画像がよりリンパ節分類に効果があるか確認するため、Vision Transformer (ViT) [32] による学習を行った。ViT は自然言語処理で高い性能を発揮する Transformer アーキテクチャ[33] を、視覚的特徴の抽出と理解に適応することで画像処理に応用したモデルである。ViT は画像分類において高い精度を発揮するが、学習には大量のデータセットが必要である。例えば、文献[32] では ViT の事前学習には 1400 万枚以上の画像が使用されている。本実験では、生成データの枚数は各クラス 500 から 5000 枚で変化させて比較した。G1 と G3 のそ

れぞれ15枚ずつをランダムに抽出しテストデータとする.結果を比較するため,全 ての実験を通してテストデータは固定する.テストデータを除いた元のデータセッ トに生成画像を加えた上で,元のデータと生成画像が均等に含まれるように4分割 したうちの1つを評価データとし,残りの3つを訓練データとして層化4分割交差 検証を行った.

拡張枚数	Accuracy	Precision	Recall	F值
各クラス 500 枚	0.8500	0.9613	0.7333	0.8269
各クラス 1,000 枚	0.8583	1.000	0.7167	0.8299
各クラス 2,000 枚	0.8917	0.9833	0.8000	0.8786
各クラス 4,000 枚	0.9000	0.9833	0.8167	0.8844
各クラス 5,000 枚	0.9333	1.000	0.8667	0.9279

表 5.1: DDIM でのデータ拡張による分類結果

拡張枚数	Accuracy	Precision	Recall	F値
各クラス 500 枚	0.8333	1.000	0.7667	0.8636
各クラス 1,000 枚	0.8667	0.8709	0.8667	0.8675
各クラス 2,000 枚	0.9167	0.9651	0.8667	0.9091
各クラス 4,000 枚	0.9417	0.9208	0.9667	0.9432
各クラス 5,000 枚	0.9833	0.9844	0.9833	0.9833

表 5.2: DDPM でのデータ拡張による分類結果

DDIMでのデータ拡張による分類結果を表 5.1 に示す. DDPMでのデータ拡張によ る分類結果を表 5.2 に示す. どちらのモデルも 5000 枚まで生成データ拡張の枚数を 増やすほど性能が向上している.また,どちらのモデルも 5000 枚データ拡張した際 は、専門医師の 86%を超える結果が得られている.DDIM と DDPM を比較すると、 生成枚数によらず DDPM の方が F 値が高く、分類精度が高いことがわかる.Recall も生成枚数によらず DDPM の方が高いため、G3 の分類精度が高いといえる.その ため、DDIM より G3 に近い画像を生成することができていると考えられる.しか し、Precision に関しては DDIM の方が高いことが多い.これは DDPM が今回の学 習法により多様な G3 を生成し、その中に G1 と G3 どちらの特徴もある画像が生成 されたためと考えられる.Recall が高く Precision が低いことは G3 と予測した数が 多くなっていることを示す.これは G3 の特徴をより捉えることができていると考え ることができ,DDPMによって生成したデータの質が高いことが予想される.さら に,がん診断においては,G3の見落としを防ぐことが最も重要であるためPrecision より Recall が重要視される.そのため DDPM は生成に時間がかかるという側面も あるが,時間よりも精度を優先すべきリンパ節分類においては DDIM より DDPM によるデータ拡張の方が有効であるといえる.

第6章 二つの選別手法の提案と比較

生成画像の中に期待とは異なる画像が生成されてしまう課題を解決するために, リンパ節の特徴を踏まえて適切な画像のみ選別する二つの手法を提案する.また, DDIM と DDPM による生成画像に対し,それぞれの選別手法を用いることによる リンパ節分類への効果とそれらの比較について述べる.

6.1 U-Net によるリンパ節検出を用いた選別

U-Net で生成画像からリンパ節検出を行うことで G1, G3 として適切であること を判定する手法を提案する.U-Net とはセマンティックセグメンテーションに多く 用いられる畳みこみニューラルネットワークである.全結合層が存在せず,ピクセ ル単位でのセグメンテーションが可能である.U-Net は概形が図 6.1 のように U 字 になっており,エンコーダ部 (図 6.1 の左側)とデコーダ部 (図 6.1 の右側)から なるという特徴がある.この U-Net を用いて G1 と G3 をそれぞれ学習し,生成画 像からリンパ節領域を検出する.正しく検出できた場合,その生成画像はリンパ節 画像として適切であるとし,検出できない場合は不適切であると判定することで選 別を行う.一般的に物体検出を行う際は,画像全体を対象に処理を行う.しかし本 研究では,リンパ節として適切であるか否か検出によって判定することが目的であ るため,腹部全体ではなくリンパ節が中心に位置する 64 × 64 ピクセルの画像を使 用する.G1 として適切である画像を選別するには G1 だけを検出し,G3 を検出し ないように学習を行う.入力データを以下に示し,その例を図 6.2 に示す.なお,検 出対象がないマスク画像は空マスク画像とする.

- G1 画像と G1 マスク画像: 603 個
- G3 画像と空マスク画像:24 個



図 6.1: U-Net の概形

• リンパ節でない部位の画像と空マスク画像:579個

G1とG3のデータのみで学習した場合,必ず中央にリンパ節があるため位置情報を 学習し,生成画像にリンパ節のような形状がない場合でも中央部分を検出する可能 性がある.そのため,リンパ節でない部位の画像を加えることで,リンパ節がない 場合は検出しないようにする.そこで次に,G3として適切である画像を選別するに はG3を検出しG1を検出しないように学習を行う.入力データを以下に示す.

- G1 画像と空マスク画像:603 個
- G3 画像と G3 マスク画像: 24 個

上記の入力データで検出を行っただけでは,図 6.3 のような小さくまばらな検出が 行われることがある.そのため,検出した連続領域のピクセル数が基準値以上の画 像のみを選別する.ピクセル数の基準値を以下に示す.

- G1 選別での基準値:20
- G3 選別での基準値:90

これらの基準値はそれぞれ最も小さいリンパ節のピクセル数をもとに決定する.



(a) G1 画像



(c) G3 画像



(b) G1 マスク画像



(d) 空マスク画像



(e) リンパ節でない部位の画像





(a) 生成画像



(b) U-Net 出力画像

図 6.3: まばらな検出の例



図 6.4: DreamSim による類似度比較の例

6.2 DreamSim による類似度比較を用いた選別

DreamSim を用いて生成画像のリンパ節部分が G1 と G3 どちらに類似している かを判別することで,適切な画像であることを判定する手法を提案する. PSNR や SSIM などの従来の知覚類似性メトリクスがピクセルとパッチレベルで画像間の視 覚的な類似性を評価するのに対し,DreamSim は画像全体の視覚的特徴と内容を総 合的に評価する.そのため,従来の知覚類似性メトリクスより人間の知覚に近い類 似性判断を行うことが可能である.DreamSim により選別を行うには,生成画像1 枚につき既存のデータ全てとの類似度を計算する.そして最も類似した画像と同じ 分類であると判断する.例えば G1 であることを期待して生成した画像に最も類似 した画像が G1 であれば,期待通り適していると判断し、そうでなければ適してい ないと判断する.G3 であることを期待して生成した画像に対しても同様である.リ ンパ節部分の類似度を評価するため,画像の中央 32 × 32 ピクセルをクロップした 画像を使用する.G3 として生成されることを期待した画像に対し,類似度を計算し た例を図 6.4 に示す.類似度を距離で表しており,生成画像はDreamSim によりG3 と同じ分類であると判断されている.

6.3 選別枚数による比較

まず, DDIM と DDPM がそれぞれ期待通りの画像を生成できているか否か,提案 の選別手法により比較する. DDIM と DDPM により G1 と G3 を 10000 枚ずつ生成 し,両者を二通りの手法で選別する. 選別後の枚数を表 6.1 に示す.

表 6.1: 選別後の枚数

生成モデル	U-Net による選別	DreamSim による選別
DDIM	G1: 8211, G3: 569	G1: 9862, G3: 761
DDPM	G1: 7394, G3: 1338	G1: 9296, G3: 1979

どちらの選別においても、DDIM によって G1 を期待して生成した画像の選別枚 数は DDPM によって G1 を期待して生成した画像の選別枚数より多い. このことか ら DDIM の方が質の高い G1 生成ができていると考えられる.一方、両選別とも、 DDIM によって G3 を期待して生成した画像の選別枚数は DDPM によって G3 を期 待して生成した画像の選別枚数より少なく、DDPM の方が質の高い G3 生成ができ ていると考えられる. 続いて U-Net による選別と DreamSim による選別を比較する と、U-Net による選別の方が選別される数が少ない. U-Net による選別の方がより 基準の高い選別を行っている可能性がある一方で、U-Net の学習が十分でなかった ため選別枚数が少なくなった可能性もある.

6.4 リンパ節分類実験による選別効果検証

選別後の生成画像によりデータ拡張しリンパ節分類実験を行う.そしてその結果 から選別の有無による比較や,選別手法の比較,選別のDDIMとDDPMに対する 効果の比較を行う.DDIMの生成画像に対して,U-Netによる選別後にデータ拡張 を行った分類結果を表 6.2 に示し,DreamSimによる選別を行った後にデータ拡張を 行った分類結果を表 6.3 に示す.同様にDDPMの生成画像に対して,U-Netによる 選別を行った後にデータ拡張を行った分類結果を表 6.4 に示し,DreamSimによる選 別を行った後にデータ拡張を行った分類結果を表 6.5 に示す.

拡張枚数	Accuracy	Precision	Recall	F値
各クラス 500 枚	0.8750	0.9599	0.7833	0.8624

表 6.2: 生成:DDIM, 選別:U-Net でのデータ拡張による分類結果

拡張枚数	Accuracy	Precision	Recall	F値
各クラス 500 枚	0.9083	1.000	0.8167	0.8988

表 6.3: 生成: DDIM, 選別: DreamSim でのデータ拡張による分類結果

拡張枚数 Accuracy Precision Recall F 值

抵抵权致	Accuracy	Precision	Recall	F川旦
各クラス 500 枚	0.8917	1.000	0.7833	0.8782
各クラス 1000 枚	0.9000	0.9143	0.8834	0.8983
各クラス 1338 枚	0.9083	0.9154	0.9000	0.9074

表 6.4: 生成: DDPM, 選別: U-Net でのデータ拡張による分類結果

拡張枚数	Accuracy	Precision	Recall	F値
各クラス 500 枚	0.9083	0.9821	0.8333	0.9007
各クラス 1000 枚	0.9250	0.9458	0.9000	0.9204
各クラス 1979 枚	0.9500	0.9667	0.9312	0.9490

表 6.5: 生成: DDPM, 選別: DreamSim でのデータ拡張による分類結果

6.4.1 選別の有無による比較

選別の有無による分類結果を比較することで選別が効果的であるかを確認する.表 5.1,表6.2,表6.3の各クラス500枚のデータ拡張数で比較すると,DDIMの生成画 像に対して選別をすることでAccuracy,Recall,F値が向上している.このことか ら,DDIMによる生成画像に対して選別が有効であることがわかる.また、表5.2, 表6.4,表6.5を比較することで,DDPMに対しても同じデータ拡張数では選別後の 方がAccuracy,Recall,F値が高いことが確認できる.そのため,DDIMとDDPM どちらに対しても選別が効果的であり,不適切な画像が削減できたと考えられる.

6.4.2 選別手法の比較

それぞれの選別手法のどちらがより効果的であるかを検証するため,選別手法 ごとの分類結果を比較する.表 6.2,表 6.3 を比較することで,DDIM においては DreamSim による選別の方が全ての評価項目において高いことが確認できる.次に, DDPM において選別手法による比較を行う.表 6.4,表 6.5 を比較すると,DreamSim

モデル/転移の有無	U-Net による選別後の枚数	DreamSim 選別後の枚数	共通して選別された画像数
DDIM / G1	8211	9862	8105
DDIM / $G3$	569	761	97
DDPM / $G1$	7394	9296	6987
DDPM / G3	1338	1979	317

表 6.6: 共通して選別された画像数

による選別の方が500枚の拡張と1000枚の拡張どちらにおいてもAccuracy, Recall, F値が高い. そのため, DDIMとDDPMどちらにおいてもDreamSimによる選別の方 が効果的であることがわかる. 前節のとおり,選別枚数のみを参照すると, DreamSim による選別よりもU-Netによる選別の方が選別枚数が少ないため選別基準が高い可 能性も考えた. しかし,分類結果ではより適切な画像を選別しているのはDreamSim による選別であった. このことから,単にそれぞれの手法の選別基準の高低ではな さそうだと考えた. 二つの選別手法で共通して選別された画像数を表 6.6 に示す. ど ちらの手法においても G1 は 10000枚の生成画像の大半が選別されており,共通し て選別された画像数もかなり多い. 一方で,G3で共通して選別された画像数は少な い. このことから,それぞれの手法による選別の違いは単純な基準の高低ではない と確認できる. すなわち U-Net による選別とDreamSim による選別では,選別され た画像の数ではなく質に違いがあったため,分類結果にも差が生じたのだろう.

6.4.3 選別の DDIM と DDPM に対する効果の比較

選別を行うことで分類精度が向上することが確認できたが、その効果はDDIMと DDPM で異なる可能性がある.それを確認するため、表 6.2 と表 6.4 の組み合わせ と、表 6.3 と表 6.5 の組み合わせで比較する.いずれの組み合わせでも、選別後の分 類精度はDDIM よりDDPM の方が高い.DDPM の方が分類精度が高いことは選別 なしで分類実験を行ったときと同じであるが、選別を行うことでDDIM とDDPM の 分類精度の差が縮まっている.このことから選別はDDIM に対してより効果的であ るといえる.この原因として、DDIM による生成画像の方が質の低い画像が多かっ たため、選別の効果が大きく現れたと考えられる.一方でDDPM は選別によって精 度は向上するがその効果は少ない.これはDDPM による生成画像の質が全体的に 高いためだと思われる.

第7章 むすび

本論文ではリンパ節分類の精度向上を目的に、学習用のデータ拡張のための画像 生成において期待とは異なる画像が生成されてしまう課題に対して、これまでとは 別の画像生成手法の導入と二種類の選別手法を提案した。新たに導入した DDPM に よるデータ拡張後にリンパ節分類を行うことで、先行研究で使用した DDIM による データ拡張よりも Recall と F 値が向上することを示した。特に Recall が大きく向上 していることから、がん転移ありリンパ節 G3 の分類精度が高くなっており、DDPM によって質の高い G3 が生成できたと考えられる。このことから、G3 の見落としを 防ぐことが重要なリンパ節分類において、DDIM よりも DDPM の方が適切であると いえる。次に、生成画像から適切な画像を選別する手法を二つ提案し、どちらの選別 も同じ拡張枚数では精度が向上することを示した。U-Net による選別は DreamSim による選別に対して選別される枚数が少なかった。また分類結果から DreamSim に よる選別の方が効果的であった。それぞれの手法は選別の基準の方向性が異なるこ とも確認した。また、DDIM によって生成した画像の方が選別による影響は大きいこ とがわかり、この原因として DDIM と DDPM の生成画像の質の違いが考えられる。

本研究では選別後の枚数から, DDIMの方が質の高いG1生成ができており, DDPM の方が質の高いG3生成ができていると考えた.そのため,今後の課題としてDDIM によって生成したG1とDDPMにより生成したG3に対し選別を行い,ViTの学習 をすることが挙げられる.また,ViTに入力する訓練データは数が多く,質が高い ことが理想的である.しかし,質を高くするために選別を行うことでデータが減少 してしまうため,データの質と数のバランスが重要となる.そのため今後の展望と して,選別の基準を変更することで,分類精度に最も寄与するデータの質と数の最 適なバランスを明らかにすることが挙げられる.

謝辞

本研究を行うにあたり,終始熱心にご指導をいただいた名古屋工業大学 舟橋健司 准教授,伊藤宏隆 助教に深く感謝いたします.また,研究の最終方針までご指導い ただいた中部大学 岩堀祐之教授 (2024年12月24日逝去)に感謝するとともに結果 をお見せできなかったことを残念に思います.この論文を捧げます.なお,本研究 において,CT 画像の提供をはじめ,ご協力を頂きました愛知県がんセンター大内 晶先生,清水泰博先生に感謝いたします.さらに,本研究を進めるにあたり多くの 助言を頂き,御協力いただいた名古屋工業大学舟橋研究室の皆様に深く感謝し,厚 く御礼申し上げます.

参考文献

- [1] "直腸癌側方骨盤リンパ節転移の術前診断の妥当性に関する観察研究 厚生労 働科学研究成果データベース," https://mhlw-grants.niph.go.jp/project/ 24780 (Accessed on 01/22/2025).
- [2] Y. Manjunatha, V. Sharma, Y. Iwahori, M. Bhuyan, A. Wang, A. Ouchi, and Y. Shimizu, "Lymph node detection in CT scans using modified U-Net with residual learning and 3D deep network," International Journal of Computer Assisted Radiology and Surgery, vol.18, no.4, pp.723–732, 2023.
- [3] A.-A. Nayan, B. Kijsirikul, and Y. Iwahori, "Mediastinal lymph node detection and segmentation using deep learning," IEEE Access, vol.10, pp.89289–89307, 2022.
- [4] T. Ono, Y. Iwahori, H. Usami, B. Kijsirikul, M.K. Bhuyan, T. Oshiro, and Y. Shimizu, "Detection of Lymph Nodes using CNN from Contrast-Enhanced CTImages," 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI)IEEE, pp.471–474 2020.
- [5] J.P. Singh, Y. Iwahori, M.K. Bhuyan, H. Usami, T. Oshiro, and Y. Shimizu, "Mediastinal Lymph Node Detection using Deep Learning.," ICPRAM, pp.159-166, 2020.
- [6] S. Kundu, Y. Iwahori, M.K. Bhuyan, M. Bhatt, A. Ouchi, and Y. Shimizu, "Improved Residual Vision Transformer for CT to MRI Translation," 2023 Fifth International Conference on Transdisciplinary AI (TransAI)IEEE, pp.58–65 2023.
- [7] "直腸:[国立がん研究センター がん統計],"https://ganjoho.jp/reg_stat/ statistics/stat/cancer/7_rectal.html.(Accessed on 01/22/2025).

- [8] "国立がん研究センターがん情報サービス「がん統計」(厚生労働省人口動態統計),"https://ganjoho.jp/reg_stat/statistics/data/dl/index.html#
 a7 (Accessed on 01/22/2025).
- [9] "JSCCR 大腸癌研究会," https://www.jsccr.jp/guideline/2019/index_guide.html (Accessed on 01/22/2025).
- [10] T. Akiyoshi, T. Watanabe, S. Miyata, K. Kotake, T. Muto, and K. Sugihara, "Results of a Japanese Nationwide Multi-Institutional Study on Lateral Pelvic Lymph Node Metastasis in Low Rectal Cancer: Is it regional or distant disease?," Annals of surgery, vol.255, no.6, pp.1129–1134, 2012.
- [11] S. Fujita, S. Yamamoto, T. Akasu, and Y. Moriya, "Prognostic Factors of Rectal Cancer Patients with Lateral Pelvic Lymph Node Eetastasis.," Hepatogastroenterology, vol.59, no.120, pp.2494–2497, 2012.
- [12] "直腸がんの再発率は?再発しやすい時期、再発パターンを完全解説," https://www.senju-ge.jp/media/rectal-cancer-recurrence-rate (Accessed on 01/22/2025).
- [13] Sugihara K, Kobayashi H, Kato T, Mori T, Mochizuki H, Kameoka S, Shirouzu K, Muto T. Indication and benefit of pelvic sidewall dissection for rectal cancer. Dis Colon Rectum. 2006 Nov;49(11):1663-72. doi: 10.1007/s10350-006-0714-z. PMID: 17041749.
- [14] "JSCCR-大腸癌研究会," http://www.jsccr.jp/guideline/2019/cq.html. (Accessed on 01/22/2025).
- [15] H. Wang, Z. Zhou, Y. Li, Z. Chen, P. Lu, W. Wang, W. Liu, and L. Yu, "Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images," EJN-MMI research, vol.7, pp.1–11, 2017.

- [16] 磯崎哲郎,吉村裕一郎,平田篤史,村上健太郎,津村徳道,林秀樹,松原久裕, "機械学習を用いた食道癌リンパ節転移予測手法の検討,"日本コンピュータ外 科学会誌,vol.22, no.4, pp.279-280, 2020.
- [17] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," https://arxiv.org/pdf/ 2112.10752, 2022, (Accessed on 01/22/2025).
- [18] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. Seyed Ghasemipour, B. Karagol Ayan, S. Mahdavi, R. Gontijo Lopes, T. Salimans, J. Ho, D. Fleet, M. Norouzi, "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding," https://arxiv.org/pdf/2205.11487, 2022, (Accessed on 01/22/2025).
- [19] R. Yan, K. Wu, K. Ma, "Flow Score Distillation for Diverse Text-to-3D Generation," https://arxiv.org/pdf/2405.10988, 2024, (Accessed on 01/22/2025).
- [20] 岩崎翔,粟井修司,青木俊彦,石塚昌平,紺野剛史,"Narrative-to-image:ナ ラティブに合致した画像の自動生成,"人工知能学会全国大会論文集,第38回, 2024.
- [21] 海老沢源,島田拓人,平原健太郎,小池誠,小川晋,野村祐一郎,峰野博史,"画像生成モデルを用いたドメイン適応型データ拡張手法によるメロン等級判定モデルの評価,"マルチメディア,分散,協調とモバイルシンポジウム 2024 論文集, pp.603-610, 2024.
- [22] K. Suzuki, Y. Iwahori, K. Funahashi, M. K. Bhuyan, A. Ouchi and Y. Shimizu, "Data Augmentation Method for Classification of Lymph Node Metastasis using Diffusion Model," Proceedings of the International Conference on Image Processing and Robotics, 2024.
- [23] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv preprint arXiv:1505.04597, 2015

- [24] S. Fu, N. Y. Tamir, S. Sundaram, L. Chai, R. Zhang, T. Dekel, P. Isola, "Dream-Sim: Learning New Dimensions of Human Visual Similarity using Synthetic Data," arXiv preprint arXiv:2306.09344, 2023
- [25] T. Akasu, G. Iinuma, M. Takawa, S. Yamamoto, Y. Muramatsu, and N. Moriyama, "Accuracy of High-Resolution Magnetic Resonance Imaging in Preoperative Staging of Rectal Cancer," Annals of surgical oncology, vol.16, no.10, pp.2787–2794, 2009.
- [26] G. Brown, C.J. Richards, M.W. Bourne, R.G. Newcombe, A.G. Radcliffe, N.S. Dallimore, and G.T. Williams, "Morphologic Predictors of Lymph Node Status in Rectal Cancer with Use of High-Spatial-Resolution MR Imaging with Histopathologic Comparison," Radiology, vol.227, no.2, pp.371–377, 2003.
- [27] 林孝文,新國農,西山秀昌,"頸部リンパ節腫脹の画像診断:超音波診断を中心 として,"日本口腔外科学会雑誌,70巻,2号,pp.42-48,2024.
- [28] 瀧山博年,"進行下部直腸癌における術前放射線療法と術前化学放射線療法の 臨床病理学的効果の比較,"東京大学大学院医学系研究科博士論文, 2019.
- [29] 岡田佳朗, "CNN を用いた造影 CT 画像からのリンパ節転移の有無分類,"中部 大学修士論文, 2021.
- [30] J. Song, C. Meng, S. Ermon, "DENOISING DIFFUSION IMPLICIT MOD-ELS," https://arxiv.org/pdf/2010.02502, 2020, (Accessed on 01/22/2025).
- [31] J. Ho, A. Jain, P. Abbeel, "Denoising Diffusion Probabilistic Models," https:// arxiv.org/pdf/2006.11239, 2020, (Accessed on 01/24/2025).
- [32] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in neural information processing systems, pp.5998–6008, 2017.