

令和7年度 卒業論文

# 深層学習を利用した三次元医用画像超解像 における学習時の計算削減

— 知覚損失計算・モデル規模・学習入力設計の検討 —

指導教員 椋木 雅之 教授

宮崎大学 工学部 工学科 情報通信工学プログラム

学籍番号 60222509

西岡 駿

2026年2月6日 提出

# 目次

第 1 章	はじめに	1
第 2 章	超解像技術の変遷と三次元医用画像への拡張における課題	3
2.1	補間に基づく従来の超解像	3
2.2	深層学習に基づく超解像技術	3
2.2.1	画素精度 (MSE) 重視のアプローチ	3
2.2.2	知覚品質 (Perceptual Quality) 重視のアプローチ	4
2.3	三次元医用画像への拡張	4
2.3.1	2D スライス処理の限界と 3D モデルの必要性	4
2.3.2	研究室内先行研究: 3D-SRCGAN の三次元医療データ超解像への応用	4
2.3.3	3D RRDB-GAN と計算コストの課題	5
2.4	小結	5
第 3 章	知覚損失の近似とモデル・入力設計による計算削減手法	6
3.1	提案手法の位置付けと概要	6
3.1.1	本研究の位置づけ	6
3.1.2	提案手法の概要	6
3.2	問題設定と前処理	7
3.2.1	問題設定	7
3.2.2	データ形式と正規化	7
3.2.3	学習用パッチ	7
3.3	ネットワーク構造	8
3.3.1	Generator: 3D RRDB-Net	8
3.3.2	Discriminator: 3D CNN スカラー判別器	9
3.4	損失関数の設計	9
3.4.1	再構成損失 (MSE, SSIM)	10
3.4.2	敵対的損失	10
3.4.3	2.5D VGG 知覚損失	10
3.5	学習手順と実装設定	11
3.5.1	学習アルゴリズム	11
3.5.2	最適化とハイパラメータ	11
3.6	推論	11

<b>第 4 章</b>	<b>実験</b>	12
4.1	実験条件 . . . . .	12
4.1.1	データセット . . . . .	12
4.1.2	実験環境 . . . . .	12
4.1.3	実験設定と学習設定 . . . . .	12
4.1.4	評価 . . . . .	13
4.2	提案手法の定量評価結果 . . . . .	13
4.3	実験 1：VGG ランダムサンプリング枚数の影響 . . . . .	14
4.4	実験 2：RRDB 規模（ブロック数）の影響 . . . . .	15
4.5	実験 3：パッチサイズ（入力サイズ）の影響 . . . . .	15
4.6	定性評価 . . . . .	16
4.6.1	視覚的品質の比較 . . . . .	16
4.6.2	ランダム抽出スライスによる出力例 . . . . .	17
4.7	本章の総合考察 . . . . .	17
<b>第 5 章</b>	<b>おわりに</b>	21
5.1	結論 . . . . .	21
5.2	今後の課題 . . . . .	22
	<b>参考文献</b>	24

# 第1章

## はじめに

三次元医用画像（MRI・CT など）は、本質的に三次元ボリュームデータとして扱われ、診断や治療計画において体積内の連続性や三次元的な解剖学的構造の把握が極めて重要となる。高解像度な三次元医用画像は詳細な診断を可能にする一方で、その取得には大きなコストが伴う。例えば、高分解能での撮像は、十分な信号対雑音比（SNR）を確保するために高磁場 MRI など高性能機器が必要となる場合があるほか、撮像時間の増加につながり得る。その結果、患者の身体的負担や病院の運用コストの観点から、常に高解像度データが利用できるとは限らない。このような背景から、低解像度の三次元医用画像から高解像度画像を推定する技術である超解像（Super-Resolution; SR）を、深層学習（Deep Learning）により実現するソフトウェア的アプローチが近年盛んに研究されている。

超解像に関しては二次元画像を対象とした研究が先行しており、ESRGAN [1] などに代表される手法により、写真のような自然な画像の復元が可能となっている。しかし、三次元医用画像に対して、各スライスを二次元画像として独立に切り出して超解像を行う方法では、スライス間（奥行き方向）の相関が明示的に扱われず、再構成された三次元構造に不連続が生じたり、血管や臓器の形状が破綻したりする可能性がある。このため、三次元ボリュームを入力とし、三次元的な相関を考慮して三次元ボリュームを出力する「三次元超解像」が求められている。

三次元超解像の代表的な成功例として、Ha ら [2] は 4 倍超解像を対象に、2D 超解像で高性能が報告されている ESRGAN[1] で導入された RRDB 系の設計を三次元医用画像へ拡張した 3D RRDB-GAN を提案している。3D RRDB-GAN は、RRDB（Residual-in-Residual Dense Block）と呼ばれるネットワーク構造を三次元へ拡張して生成器（Generator：画像を生成するネットワーク）に組み込み、三次元ボリュームの表現能力を高めることを狙う。さらに、事前学習済みネットワークの特徴量差に基づく知覚損失（Perceptual Loss）[3] を用い、2D VGG [4] を特徴抽出器として各軸方向に展開・集約する 2.5D 方式を導入することで、画素誤差だけでは捉えにくい構造・テクスチャ情報を保った復元、すなわち視覚的リアリズムの向上を目指している。ここで GAN とは、生成器と識別器を競合的に学習させる枠組みであり、より自然な画像分布へ近づけることを狙う。実験では、LPIPS[5] などの知覚指標および可視化において有効性が報告されており、三次元医用画像超解像は画質面で進展を見せている。

一方で、三次元超解像は二次元超解像に比べて計算コストが著しく増大するという課題がある。三次元畳み込みを用いるネットワークは計算量とメモリ消費が増大し、学習および推論に要する時間が研究開発のサイクルや実用化を制約する要因となっている。さらに、3D RRDB-GAN のような高品質化手法で導入される知覚損失（VGG 特徴量を用いた損失計算）は、追加的な計算リソースを必要とし、計算コスト増大の一因となる。したがって今後の三次元医用画像超解像においては、先行手法（3D RRDB-GAN）が実現する高品質を大きく損

なわない範囲で、どの程度の計算削減が可能かを系統的に整理し、効率的な設計指針として示すことが重要である。本研究では、計算コストに影響する要因を (i) 知覚損失計算, (ii) モデル規模, (iii) 学習入力設計の観点に分解して検討する。

本研究は、3D RRDB-GAN [2] を基盤とした三次元医用画像の 4 倍超解像を対象とし、**同手法の持つ高品質を維持したまま、学習の計算コストをどこまで削減できるか**を実験的に検証することを目的とする。ここで品質とは、画素（ボクセル）値の誤差を測る MSE（Mean Squared Error）および、MSE から算出される PSNR（Peak Signal-to-Noise Ratio）、構造的な類似度を評価する SSIM（構造的類似度）[6] に加え、深層学習モデルの特微量を用いて人間の知覚に近い画質評価を行う LPIPS（Learned Perceptual Image Patch Similarity）[5] を用いて多面的に評価する。

計算効率化の方策は多岐にわたるが、本研究では以下の三点を中心に比較検討する。

第一に、**知覚損失計算の効率化**としてスライスのランダムサンプリングを検討する。知覚損失計算とは、生成画像と正解画像を特徴抽出ネットワーク（VGG）に通し、その特微量の差を計算する手法である。VGG は画像認識で事前学習された畳み込みニューラルネットワークであり、本研究では特徴抽出器として用いる。三次元医用画像に対してこれを行う場合、各軸方向（xy/xz/yz）からすべてのスライスを抽出して計算すると膨大な時間を要する。そこで本研究では、各平面から抽出するスライス枚数を削減し、少数枚をランダムにサンプリングして損失を近似計算することで、サンプリング枚数が画質と計算時間に与える影響を評価する。

第二に、**モデル規模の見直し**として RRDB ブロック数を検討する。RRDB ブロックとは生成器を構成する基本単位であり、この直列数はネットワークの表現能力に寄与する一方、計算量に直結する。そこで本研究では RRDB ブロック数を変化させ、計算削減効果と品質変化の関係を実験的に検証する。

第三に、**学習入力設計の効率化**として小パッチ学習を検討する。学習入力設計とは、学習時にネットワークへ入力するデータのサイズや形式を定めることである。三次元ボリューム全体を一度に扱うのではなく、小さな三次元領域（パッチ）に切り出して入力とすることで、計算を局所領域に限定し、学習時間を短縮できる可能性がある。本研究ではこのパッチサイズ等の条件を変更し、計算時間短縮と品質維持の両立可否を検証する。

以上より、本研究の貢献は、三次元医用画像超解像における計算コストの支配要因を整理し、品質と計算効率のトレードオフを明確化する点にある。特に、知覚損失計算におけるスライス・ランダムサンプリング、RRDB ブロック数、および小パッチ学習という三つの設計要素について、品質指標（MSE/PSNR/SSIM/LPIPS）と計算量の関係を実験的に示す。本論文の構成として、第 2 章で関連研究を述べ、第 3 章で提案手法を説明する。第 4 章で実験条件と評価結果および考察を示し、第 5 章で結論と今後の課題を述べる。

## 第2章

# 超解像技術の変遷と三次元医用画像への拡張における課題

本章では、三次元医用画像超解像に関する関連研究を整理し、本研究の位置付けを明確化する。まず補間に基づく従来法を概観し、次に深層学習に基づく超解像技術の変遷（画素精度重視から知覚品質重視へ）を整理する。続いて、三次元医用画像への適用における課題（スライス間の整合性）と、それを解決する三次元モデル（3D CNN/GAN）について述べる。最後に、三次元超解像の実用化を阻む計算コストの問題に焦点を当て、本研究の目的である「知覚品質と計算効率の両立」について位置付ける。ただし、計算量の削減を過度に追求すると、生成器の表現能力や損失設計の制約により微細構造の復元性能が大きく低下する可能性がある。したがって本研究では、品質劣化を許容する単純な軽量化ではなく、知覚品質と計算効率の両立点を探索する設計方針を取る。

### 2.1 補間に基づく従来の超解像

医用画像を高解像度グリッドへ再標本化（リサンプリング）する際の基本操作として、最近傍補間、線形補間、三次（cubic convolution / bicubic）補間などが広く用いられてきた [7, 8]。Keys [7] は三次畳み込み（cubic convolution）補間を定式化し、最近傍補間・線形補間などの代表的手法と比較して計算効率や誤差特性（エイリアシング等）を議論している。また Meijering ら [8] は、医用画像に対する幾何変換（回転・サブピクセル平行移動）における補間カーネルを体系的に評価し、精度と計算コストの観点からスプライン補間が良好なトレードオフを示すことを報告している。

しかし補間は観測済みサンプルから未観測点を内挿する再構成であり、実画像ではナイキスト条件を満たさない場合も多く、厳密な復元には限界がある [8]。

### 2.2 深層学習に基づく超解像技術

#### 2.2.1 画素精度（MSE）重視のアプローチ

CNN（Convolutional Neural Networks）を用いた超解像は SRCNN[9] の提案以降、急速に発展した。初期の研究では、ネットワークの深層化や残差学習（Residual Learning）の導入により、入力画像と正解画像の間の画素ごとの誤差（Mean Squared Error: MSE）を最小化する手法が主流であった [10, 11]。これらの

手法は、定量的な歪み指標である PSNR (Peak Signal-to-Noise Ratio) を向上させる点では有効である。

一方で、MSE を最小化する学習は、不確定な高周波成分に対して統計的な平均値を出力する傾向があるため、結果として生成画像が平滑化され、人間が見た際に「ぼやけた (blurry)」印象を与える問題が指摘されている [3]。医用画像において、組織の微細な境界や質感 (テクスチャ) が消失することは診断能の低下に直結する恐れがある。したがって、PSNR/SSIM の改善のみを目的とした最適化に加え、知覚的観点での品質改善を考慮した枠組みが重要となる。

## 2.2.2 知覚品質 (Perceptual Quality) 重視のアプローチ

上述の「ぼけ」問題に対し、GAN (Generative Adversarial Networks) [12] を用いた超解像が提案された。SRGAN[3] やその改良である ESRGAN[1] は、画素ごとの誤差だけでなく、識別器 (Discriminator) による敵対的損失と、特徴抽出器 (VGG Network 等 [4]) の中間層特徴に基づく知覚損失 (Perceptual Loss) を導入した。知覚損失は、画像のピクセル値の差ではなく「構造や特徴の差」を評価するため、平均化による平滑化を抑え、より自然に見えるエッジやテクスチャの復元に寄与する [3]。

ただし、歪み最小化 (PSNR/SSIM) と知覚品質の向上は一般に同時に最適化できるとは限らず、両者の間にトレードオフが存在することが理論的にも整理されている [13]。そのため、医用画像への応用では、単に「見た目が良い」だけでなく、忠実性 (歪み指標) と知覚品質のバランスを明確な目的に基づいて設計・評価する必要がある。

また、超解像は観測されていない高周波成分を推定する問題であるため、手法によらずアーチファクトや非実在構造が生じ得る。特に GAN に基づく手法は知覚品質を高める反面、視覚的にもっともらしいテクスチャを生成しやすく、診断用途では忠実性とのバランス設計と慎重な評価が重要となる。本研究では、従来の MSE/PSNR/SSIM に加え、人間の知覚に近い評価として提案されている LPIPS[5] を併用し、品質を多面的に評価する。

## 2.3 三次元医用画像への拡張

### 2.3.1 2D スライス処理の限界と 3D モデルの必要性

CT や MRI などの三次元ボリュームデータに対し、ボリュームを二次元スライスへ分解して 2D 超解像モデルを適用する方法が考えられる。しかしこの方法は、スライス方向 (奥行き方向) の相関を明示的に扱にくく、再構成後のボリュームにおいて血管や臓器の形状がスライス間で不連続になる可能性がある [2]。特にスライス厚が大きいデータでは、三次元的な連続性の欠如が診断上の解釈を難しくする。したがって、三次元畳み込み (3D Convolution) を用いてボリューム全体の文脈を考慮しつつ超解像を行う 3D CNN / 3D GAN が重要となる [2]。

### 2.3.2 研究室内先行研究：3D-SRCGAN の三次元医療データ超解像への応用

研究室内の先行研究として、倉永 [14] は 3D-SRCGAN を三次元医療データへ適用し、ボリュームデータ特有の構造保持と学習・推論の実用性の観点から、損失設計およびパッチ処理の影響を検証している。従来の 3D-SRCGAN では画素単位の MSE に基づく最適化が中心であったのに対し、医療ボリュームでは局所構造や濃淡境界の保持が重要であるという問題意識から、MSE Loss, SSIM Loss, およびその併用 (MSE+SSIM)

を比較している。

さらに、パッチング（ボリュームを部分に分割して学習・推論する処理）についても検討し、視覚的評価ではパッチングにより濃淡や細部の再現が改善する例が報告されている。一方で、定量指標（MSE/PSNR/SSIM）と視覚的評価が必ずしも一致しないことも指摘されている。この事実は、歪み指標のみの最適化では知覚的な品質向上に限界がある可能性を示唆しており、高次特徴に基づく損失設計（例：VGG 特徴に基づく知覚損失）の導入が有効であると考えられる。本研究はこの流れを踏まえ、パッチベース処理を前提としつつ、知覚損失の導入および RRDB 系構造の採用により、微細構造の復元と知覚品質の向上を狙う（詳細は第 3 章で述べる）。

### 2.3.3 3D RRDB-GAN と計算コストの課題

三次元医用画像の高品質化手法として、Ha ら [2] は ESRGAN の構造を三次元化した 3D RRDB-GAN を提案している。この手法は、Generator に 3D RRDB ブロックを採用して三次元的な連続性を担保しつつ、損失関数の計算には 2D VGG を各断面（Axial, Sagittal, Coronal）に適用する「2.5D 知覚損失」を導入している [2, 4]。これにより、画素誤差だけでは捉えにくい構造・テクスチャ情報の復元を促し、高い知覚品質を実現している点で有効である。

一方で一般に、三次元畳み込みは二次元畳み込みに比べて計算量・メモリ消費が増大しやすく、学習・推論に要する時間が研究開発サイクルや実用化を制約する要因となり得る。さらに、2.5D 知覚損失はボリュームから多数の二次元スライスを切り出し、別途 VGG で特徴量を計算する必要があるため、学習時の追加計算負荷の要因となる。したがって、知覚品質を大きく損なわない範囲で計算コストを削減する設計指針を整理することが重要である。

そこで本研究では、既存の三次元医用画像超解像手法の枠組みを踏まえつつ、(i) 知覚損失計算の効率化（スライスのランダムサンプリング）、(ii) 生成器規模（RRDB 数）、(iii) 入力パッチ設計（パッチサイズ）を調整し、知覚品質と計算効率の両立点を系統的に検証する。

## 2.4 小結

本章では、三次元医用画像超解像に関する関連研究を整理し、補間法の限界、深層学習ベース超解像の発展（歪み最小化から知覚品質重視へ）、および三次元化に伴うスライス間整合性と計算コストの課題を概観した。次章では、本研究が採用するモデル構成と損失設計、学習・推論手順を述べる。

## 第3章

# 知覚損失の近似とモデル・入力設計による 計算削減手法

### 3.1 提案手法の位置付けと概要

#### 3.1.1 本研究の位置づけ

第2章で整理したように、三次元医用画像超解像では、(1) 画素忠実性 (PSNR/SSIM) と知覚品質 (LPIPS 等) のトレードオフ、(2) 3D 畳み込みと知覚損失 (VGG 特徴) 計算に起因する計算コスト、が実用化上の主要な課題となる。特に 3D RRDB-GAN のような高品質手法は、知覚損失を計算するために多数のスライス VGG に入力する必要がある、学習・評価サイクルの長期化を招きやすい。

そこで本研究は 3D RRDB-GAN を基盤とし、三次元医用画像超解像における「品質 (画素忠実性・知覚品質)」と「計算コスト」の両立を主題として位置づける。本研究では計算コストの支配要因を分解し、以下の3つの観点に基づいてトレードオフを実験的に整理する。

1. 知覚損失計算の近似： $xy/xz/yz$  各平面からスライスをランダムサンプリングし、全スライス評価を近似して計算を削減する。
2. ネットワーク規模の調整：RRDB の直列数をやや小規模にしてモデル容量と計算量の関係性を評価する。
3. 入力設計の調整：学習時の入力でフルボリュームを入力するのではなくパッチングで切り出したパッチを用いることで、メモリの節約と計算量の削減を図る。

これらを統一条件下で比較し、MSE/PSNR/SSIM と LPIPS の両観点から品質を評価しつつ、学習時間を含む計算コストの削減可能性を明確化する。

#### 3.1.2 提案手法の概要

本研究では、医用 3 次元画像をボクセル値からなる体積データ (ボリューム) として扱い、空間解像度の低い LR (Low-Resolution) ボリュームから参照となる HR (High-Resolution) ボリュームを推定する 4 倍超解像を対象とする。ここで LR ボリュームは、同一の視野を保ったままボクセル間隔を粗くした低解像度観測であり、HR ボリュームは対応する高解像度の基準ボリュームである。

GAN に基づく学習フレームワークにより LR ボリュームから HR ボリュームを推定する。Generator には ESRGAN の設計思想 (RRDB, BN 除去, 残差スケール) を三次元へ拡張した 3D RRDB-Net を採

用し、損失関数として MSE/SSIM に加え、敵対的損失および 2.5D VGG 知覚損失を組み合わせる。また 2.5D VGG では、各平面からのスライスをランダムサンプリングすることで、知覚損失計算の計算量を抑えつつ、微細構造の復元を狙う。

## 3.2 問題設定と前処理

### 3.2.1 問題設定

単一ボリューム超解像は、観測された低解像度 (LR) ボリューム  $\mathbf{y}$  から、その元となった高解像度 (HR) ボリューム  $\mathbf{x}$  を推定する逆問題として定式化される。一般に、LR ボリューム  $\mathbf{y}$  は理想的な HR ボリューム  $\mathbf{x}$  に対し、ぼかし (Blurring)、ダウンサンプリング (Downsampling)、および加法性ノイズ (Noise) を含む退化過程を経て得られるとモデル化できる。この退化過程は次式で表される：

$$\mathbf{y} = (\mathbf{x} \otimes \mathbf{k}) \downarrow_s + \mathbf{n}. \quad (3.1)$$

ここで、 $\otimes$  は畳み込み演算、 $\mathbf{k}$  はぼけカーネル (Point Spread Function)、 $\downarrow_s$  はスケールファクタ  $s$  によるダウンサンプリング操作 (本研究では  $s = 4$ )、 $\mathbf{n}$  は加法性ノイズを表す。なお、本研究における学習データの作成では、理想的な縮小処理 (平均プーリング) のみを行い、明示的なノイズ付加は行わないものとした (すなわち  $\mathbf{n} = \mathbf{0}$  と仮定する)。

本研究の目的は、多数の学習ペア  $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^N$  を用いて、 $\mathbf{y}$  から  $\mathbf{x}$  を復元する写像関数 (Generator)  $G_\theta$  を学習することである：

$$\hat{\mathbf{x}} = G_\theta(\mathbf{y}). \quad (3.2)$$

ここで  $\hat{\mathbf{x}}$  は推定された高解像度ボリュームを表し、本論文ではこれを **SR (Super-Resolved) ボリューム** と呼ぶ

### 3.2.2 データ形式と正規化

学習データは、LR/HR ボリュームの対応する 3D ボリューム (NIFTI 形式) から構成し、LR/HR ボリュームのファイルパスとクラスラベルを列挙したリストを用いて読み込む。各サンプルについて HR ボリュームの最大値  $m = \max(\mathbf{x})$  を計算し、 $m > 0$  のとき LR/HR ボリュームを同一の  $m$  で正規化する...

$$\mathbf{y} \leftarrow \mathbf{y}/m, \quad \mathbf{x} \leftarrow \mathbf{x}/m. \quad (3.3)$$

これにより画素値レンジが概ね  $[0, 1]$  となる。なお、本研究で用いたデータでは HR ボリュームの最大値  $m$  が 0 となる例は確認されなかったため、 $m > 0$  を前提として式 (3.3) を適用した。また、評価時の前処理として SR ボリュームは  $[0, 1]$  にクリップし、HR ボリュームは  $\max(\text{HR})$  で正規化する (指標計算と整合させるため、詳細は第 4 章に示す)。

### 3.2.3 学習用パッチ

医用 3D ボリュームは空間サイズが大きく、フルボリューム ( $60 \times 60 \times 39 \rightarrow 240 \times 240 \times 156$ ) をそのまま学習に入力すると計算資源を要する。そこで、LR/HR ボリュームの対応関係を保ったまま 3D パッチを事前に切り出し、学習データを構成する。具体的には 4 倍超解像を前提として、LR パッチ  $30 \times 30 \times 20$  と対応す

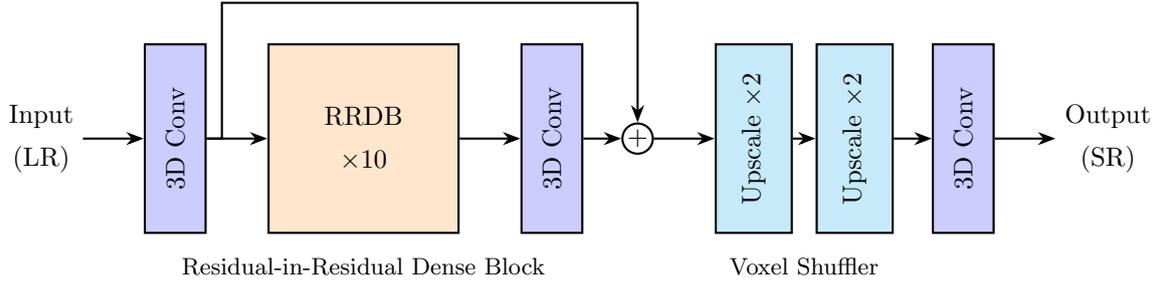


図 3.1: 提案手法の Generator 構造. RRDB を 10 個直列に配置し, グローバルな残差結合 (Global Residual) を採用している.

る HR パッチ  $120 \times 120 \times 80$  を用いる. このパッチ学習により, ミニバッチ学習とデータ多様性を両立する. (推論・評価では, フルボリューム  $60 \times 60 \times 39 \rightarrow 240 \times 240 \times 156$  を直接入力として扱う.)

### 3.3 ネットワーク構造

#### 3.3.1 Generator : 3D RRDB-Net

Generator の全体構造を図 3.1 に示す. 本モデルは ESRGAN[1] の設計思想を三次元畳み込み (3D Convolution) へ拡張した構造であり, 具体的には以下の 6 つのステージから構成される.

- (i) 初段の特徴抽出を行う 3D 畳み込み
- (ii) 深層特徴抽出を行う 10 個の RRDB (Residual-in-Residual Dense Block)
- (iii) RRDB の出力を整える 3D 畳み込み
- (iv) グローバルな残差結合 (Global Residual) による浅層・深層特徴の統合
- (v) ボクセルシャッフル (Voxel Shuffler) による 2 倍アップサンプリングの 2 段適用 (計 4 倍拡大)
- (vi) 最終的な SR ボリュームを再構成する 3D 畳み込み

次に, 本モデルの中核となる RRDB ブロックの詳細について述べる. 図 3.2 に ESRGAN における RRDB の概念図を示す. SRGAN で用いられた Residual Block (RB) から Batch Normalization (BN) を除去した RB w/o BN を基盤とし, さらに DenseBlock を 3 個直列に配置して内部に dense 接続を導入することで特徴再利用を促進している. また, 複数段の残差接続 (Residual-in-Residual) により学習を安定化させる設計である. 本研究では, この接続様式 (dense 接続および残差経路) を維持したまま 3D へ拡張し, ボリュームの空間連続性を考慮した特徴抽出を行う.

RRDB は残差接続を持つブロックであり, 入力特徴  $\mathbf{h}$  に対してブロック内部の変換結果をスケールリングして加算することで出力を得る. 本実装では, DenseBlock を 3 個直列に適用した後, RRDB 全体でも同じ係数  $\alpha = 0.2$  による residual scaling を行う.

まず, DenseBlock (5 層 3D 畳み込みと concat による dense 接続から構成) の写像を  $\text{DB}(\cdot)$  とすると, 実装は

$$\text{DB}(\mathbf{u}) = \alpha \mathcal{T}(\mathbf{u}) + \mathbf{u}, \quad (\alpha = 0.2) \quad (3.4)$$

に対応する. ここで  $\mathcal{T}(\cdot)$  は DenseBlock 内部の畳み込み計算を表す.

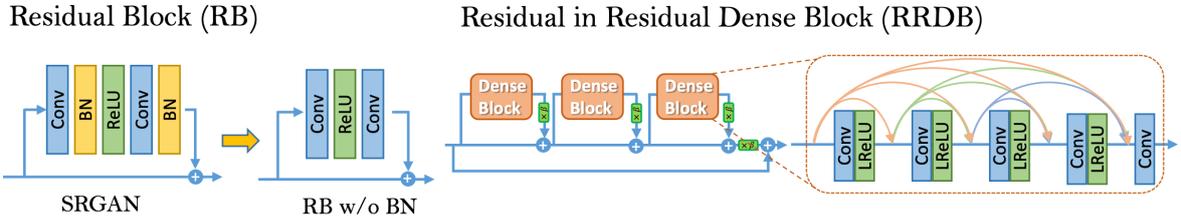


図 3.2: ESRGAN における RB (Residual Block) から RRDB (Residual-in-Residual Dense Block) への発展 (文献 [1] Fig. 4 より引用). 本研究では図の接続様式を保持したまま, 各畳み込みを 3D 畳み込みへ拡張して用いる.

次に, RRDB は DenseBlock を 3 回直列に適用した出力

$$\mathbf{u}_3 = \text{DB}(\text{DB}(\text{DB}(\mathbf{h}))) \quad (3.5)$$

を用いて,

$$\text{RRDB}(\mathbf{h}) = \alpha \mathbf{u}_3 + \mathbf{h}, \quad (\alpha = 0.2) \quad (3.6)$$

と表される. residual scaling は, 深い残差ブロックを多数直列化した際に活性化値や勾配が過度に増大することを抑え, 学習を安定化させる目的で用いる.

アップサンプリングにはボクセルシャッフル (Voxel Shuffler) を用いる. これはチャンネル方向に保持した  $r^3$  倍の情報を空間方向へ再配置し,  $(D, H, W) \rightarrow (rD, rH, rW)$  の拡大を実現する (本研究では  $r = 2$  を 2 回適用し 4 倍). 実装では, アップサンプル直前に 3D 畳み込みでチャンネル数を  $r^3$  倍 (= 8 倍) し, ボクセルシャッフルを適用する.

以上をまとめると, 入力 LR ボリュームを  $\mathbf{x}$  として初段特徴を  $\mathbf{f}_0 = \text{Conv}_{\text{in}}(\mathbf{x})$  により抽出し, RRDB trunk により  $\mathbf{f}_1 = \text{Trunk}(\mathbf{f}_0)$  を得る. その後,  $\mathbf{f}_2 = \text{Conv}_{\text{trunk}}(\mathbf{f}_1)$  を計算し, グローバル残差として  $\mathbf{f} = \mathbf{f}_0 + \mathbf{f}_2$  を形成する. 最終的な SR 出力は,  $\mathbf{f}$  に対してボクセルシャッフルによる  $\times 2$  アップサンプリングを 2 段適用して  $\times 4$  に拡大した後,  $\text{Conv}_{\text{out}}$  により再構成する.

### 3.3.2 Discriminator : 3D CNN スカラー判別器

Discriminator の全体構造を図 3.3 に示す. Discriminator は, 入力ボリュームが Real (HR) か Fake (SR) かを判別する 3D CNN である. スライド付き 3D 畳み込みと LeakyReLU を積層して特徴を抽出し, Global Average Pooling により空間次元を集約した後,  $1 \times 1 \times 1$  畳み込みでスカラーの logit を出力する. 出力は確率ではなく logit とし, 損失には二値交差エントロピー (logits 形式) を用いる. また最終層は混合精度環境での数値安定性を考慮し, float32 で計算する.

## 3.4 損失関数の設計

以下の複合損失を最小化して Generator を学習する :

$$\mathcal{L}_G = \mathcal{L}_{MSE} + (1 - \text{SSIM}(\hat{\mathbf{x}}, \mathbf{x})) + \lambda_{adv} \mathcal{L}_{adv} + \lambda_{VGG} \mathcal{L}_{VGG}. \quad (3.7)$$

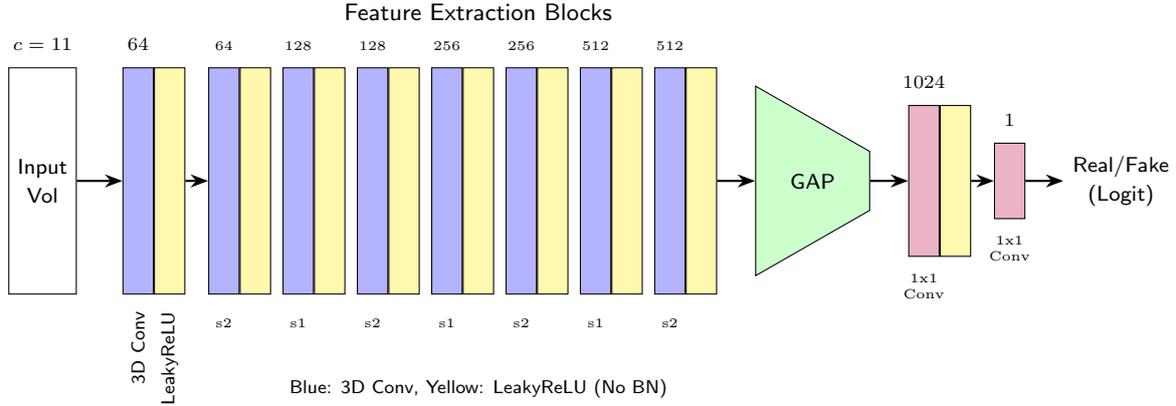


図 3.3: 提案手法の Discriminator 構造. 入力はボリュームデータとクラスラベルを結合した 11 チャンネルを受け取る. 正規化層 (Batch Norm) は使用せず, Global Average Pooling (GAP) により空間情報を集約した後,  $1 \times 1$  畳み込みにより最終的な Logit を出力する.

### 3.4.1 再構成損失 (MSE, SSIM)

画素単位の整合性として平均二乗誤差 (MSE) を用いる:

$$\mathcal{L}_{MSE} = \frac{1}{N} \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2. \quad (3.8)$$

また構造的整合性として SSIM[6] を使い, 損失は  $1 - \text{SSIM}$  として加算する. 画像  $\mathbf{x}$  と  $\hat{\mathbf{x}}$  の間の SSIM は次式で定義される:

$$\text{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}. \quad (3.9)$$

ここで  $\mu_x, \mu_{\hat{x}}$  はそれぞれの平均輝度,  $\sigma_x^2, \sigma_{\hat{x}}^2$  は分散,  $\sigma_{x\hat{x}}$  は共分散を表す. また  $C_1 = (k_1L)^2, C_2 = (k_2L)^2$  は分母がゼロになるのを防ぐための定数であり,  $k_1 = 0.01, k_2 = 0.03$  が一般に用いられる. 本研究では学習データを最大値 1.0 に正規化しているため, 画素値のダイナミックレンジ  $L = 1.0$  として計算した.

### 3.4.2 敵対的損失

Discriminator の出力 logit を  $D(\cdot)$  とすると, Generator の敵対的損失は

$$\mathcal{L}_{adv} = \text{BCE}(D(\hat{\mathbf{x}}), 1) \quad (3.10)$$

で与える. ここで BCE は logits 形式の二値交差エントロピーを表す. 一方 Discriminator は

$$\mathcal{L}_D = \text{BCE}(D(\mathbf{x}), 1) + \text{BCE}(D(\hat{\mathbf{x}}), 0) \quad (3.11)$$

を最小化する. 実装上は Real ラベルに 0.99 を用いるラベルスムージングを適用する.

### 3.4.3 2.5D VGG 知覚損失

3D ボリュームに対して 2D 画像認識用に事前学習された VGG モデルを適用するため, 2.5D 入力を構成する. 各平面 ( $xy/xz/yz$ ) において, ランダムに選択されたスライスインデックスの集合  $\mathcal{I}$  を用いて  $n$  枚の

スライスを抽出する（本実装では  $n = 10$ ）。各インデックス  $i \in \mathcal{I}$  について、中心スライスとその前後を含む 3 枚をチャンネル方向にスタックし、疑似 RGB 画像  $\mathbf{s}_i$  を構成する：

$$\mathbf{s}_i = \text{Concat}(S_{i-1}, S_i, S_{i+1}) \in \mathbb{R}^{3 \times h \times w}. \quad (3.12)$$

得られた 2D 入力  $\mathbf{s}_i$  は、ImageNet 事前学習済みモデルの入力仕様に合わせるため  $224 \times 224$  にリサイズ (bilinear) し、ImageNet 学習時と同様の平均・分散を用いた正規化処理を行ってから VGG16 に入力し、第 3 プーリング層 (pool3) 直前の特徴マップ (ReLU3.3 の出力) を特徴表現として抽出する（実装上は `vgg16.features[:16]`）。特徴空間での MSE を知覚損失として定義し、サンプリングされた全スライス・全平面について平均する：

$$\mathcal{L}_{VGG} = \mathbb{E}_{plane} \left[ \frac{1}{n} \sum_{i \in \mathcal{I}} \|\phi(\mathbf{s}_i^{SR}) - \phi(\mathbf{s}_i^{HR})\|_2^2 \right]. \quad (3.13)$$

ここで  $\phi(\cdot)$  は VGG 特徴抽出器であり、学習中は勾配を停止する。

## 3.5 学習手順と実装設定

### 3.5.1 学習アルゴリズム

各ミニバッチに対して以下を繰り返す：(1) HR を Real, SR を Fake として Discriminator を更新, (2) Discriminator を固定して Generator を式 (3.7) で更新する。Fake 側の Discriminator 更新では  $\hat{\mathbf{x}}$  から Generator への勾配伝播を遮断し、Discriminator のみを更新する ( $\hat{\mathbf{x}}$  を detach)。また、学習ログでは  $\sigma(D(\cdot))$  を用いて確率としての平均値を出力し、学習の進行 (判別器の強弱) を可視化する。

### 3.5.2 最適化とハイパラメータ

最適化には Adam[15] を用い、学習率は  $lr = 2.0 \times 10^{-4}, \beta = (0.9, 0.99)$  とした。Generator の学習率は StepLR により 70 エポックごとに 0.5 倍に減衰させ、Discriminator の学習率は一定とした。バッチサイズは 16、総エポック数は 200 とした。損失重みは  $\lambda_{adv} = 0.001, \lambda_{VGG} = 0.2$  とした。

また、本研究では複数 GPU 環境での効率的な学習を行うため、Hugging Face Accelerate<sup>\*1</sup> ライブラリを用いた分散学習環境を構築した。混合精度学習 (Mixed Precision, bf16) を有効化し、計算資源の効率化を図った。

## 3.6 推論

推論では、学習済み Generator に LR ボリュームを入力し SR ボリュームを得る。推論スクリプトでは、リストから LR/HR ボリュームとクラスラベルを読み込み、式 (3.3) で正規化した LR ボリュームを入力として  $\hat{\mathbf{x}} = G_\theta(\mathbf{y})$  を計算する。得られた  $\hat{\mathbf{x}}$  は NIFTI 形式で保存し、HR ボリュームと比較して各種指標を算出する。

<sup>\*1</sup> <https://github.com/huggingface/accelerate>

## 第 4 章

# 実験

本章では, 提案手法の有効性を検証するために行った評価実験について述べる. 4.1 節で実験の前提となるデータセットおよび評価指標について述べ, 4.2 節で提案手法の定量的な評価結果を示す. 続く 4.3 節以降では, 各要素 (VGG 損失, ネットワーク規模, パッチサイズ) が性能に与える影響について詳細な比較検証を行う.

### 4.1 実験条件

#### 4.1.1 データセット

本研究では, Medical Segmentation Decathlon (MSD) の Task01\_BrainTumour [16] を用いた. 前処理済みデータは HR ボリュームを  $240 \times 240 \times 156$ , LR ボリュームを  $60 \times 60 \times 39$  に整形して用いた. 評価はテストデータ 266 ボリュームの平均値で報告する.

#### 4.1.2 実験環境

学習および推論は, NVIDIA RTX 6000 Ada 世代 GPU を 2 基用い, Hugging Face Accelerate によるデータ並列学習で実行した.

#### 4.1.3 実験設定と学習設定

本研究では, 4 倍超解像を対象とし, 学習は 3 次元ボリュームからのランダムパッチ抽出に基づいて行う. 各症例のボリュームサイズは, LR ボリュームが  $60 \times 60 \times 39$ , 対応する HR ボリュームが  $240 \times 240 \times 156$  である. 学習では, 各 LR ボリュームからサイズ  $p_x \times p_y \times p_z$  の LR パッチをランダムに抽出し, 同一位置に対応する HR パッチ (サイズ  $4p_x \times 4p_y \times 4p_z$ ) を対として生成する. このとき, 各ボリュームから 8 個のパッチ対を抽出する設定は全実験で共通である.

MSD Task01 BrainTumour の学習用データは 484 ボリュームであり, 各ボリュームから 8 個のパッチ対を生成するため, 学習サンプル数は  $484 \times 8 = 3872$  とした. なお, パッチサイズ  $(p_x, p_y, p_z)$  はアブレーション実験において変更しており, 各実験で用いた具体的なパッチサイズは該当節にて示す.

パッチサイズはモデル構造により GPU メモリ消費量が異なるため, メモリ不足を回避する目的でモデルごとに調整した. 一方で学習効率を考慮し, 許容される範囲で可能な限り大きいパッチサイズを採用した.

学習は 200 epoch まで実行した。各 epoch 終了時に、学習データに対する平均損失 (MSE/SSIM に基づく損失) を記録し、損失が十分に収束したことを確認した上で、終盤 (概ね 200 epoch 近傍) のチェックポイントを推論・評価に用いた。なお、テストデータはモデル選択には用いず、最終評価のみに使用した

#### 4.1.4 評価

本研究では、再構成品質と知覚品質の両面から性能を評価する。再構成品質は MSE, PSNR, SSIM により評価し、MSE は小さいほど、PSNR/SSIM は大きいほど良好である。知覚品質は LPIPS を用い、LPIPS は小さいほど良好である。

■SR/HR の対応付けと入出力 評価用スクリプトでは、SR/HR を同一のファイル命名規則に基づくものとみなし、SR ディレクトリおよび HR ディレクトリ内のボリュームデータを辞書順に整列して 1 対 1 に対応付けた。

■正規化と MSE/PSNR/SSIM SR ボリューム  $V_{SR}$  および HR ボリューム  $V_{HR}$  に対して次の正規化を適用する：

$$V'_{SR} = \text{clip}(V_{SR}, 0, 1), \quad V'_{HR} = \begin{cases} \frac{V_{HR}}{\max(V_{HR})} & \max(V_{HR}) \neq 0, \\ V_{HR} & \max(V_{HR}) = 0. \end{cases} \quad (4.1)$$

ここで  $\max(V_{HR})$  はボリューム内の最大画素値である。MSE, PSNR, SSIM は正規化後の 3 次元ボリューム  $V'_{SR}, V'_{HR}$  全体に対して算出する。PSNR は最大値を 1 として計算した。SSIM は `skimage.metrics.structural_similarity` により算出し、浮動小数点入力であるため `data_range` を  $\max(V'_{HR}) - \min(V'_{HR})$  として指定した。

■LPIPS (全スライス, 3 平面) 3 次元的な知覚品質を網羅的に評価するため、 $xy$  (Axial),  $xz$  (Coronal),  $yz$  (Sagittal) の 3 軸すべての断面に対して全スライスを評価対象とした。各軸方向のスライス画像を  $S$  としたとき、LPIPS は入力レンジを  $[-1, 1]$  とするため、次の変換を適用した：

$$S_{\text{input}} = 2 \cdot \text{clip}(S, 0, 1) - 1. \quad (4.2)$$

なおこの変換は SR/HR ボリュームの双方に適用し、LPIPS 計算に限り HR ボリュームも  $[0, 1]$  へのクリップを行った (MSE/PSNR/SSIM の算出手順には影響しない)。最終的な LPIPS スコアは、3 平面すべてのスライスにおける算出値をスライス枚数で重み付け平均した `LPIPS[mean]` として報告する。バックボーンネットワークには AlexNet を使用した。

■計算コスト 計算コストは、ログに出力される 1 epoch あたりのおおよその所要時間と、200 epoch の総学習時間を併記する。ただし時間は計算機環境の負荷や状態 (熱スロットリング等) により変動し得るため、相対比較のための参考値として扱う。

## 4.2 提案手法の定量評価結果

提案手法 (基準モデル) の評価結果を表 4.1 に示す。本結果はテストデータ 266 例に対する平均値である。表 4.1 より、提案手法は、テストデータ 266 例の平均で PSNR=35.47, SSIM=0.9608, LPIPS[mean]=0.0324 を達成した。本研究では知覚損失として 2.5D VGG 特徴を用いるが、全スライスを問わず各平面 10 枚をラ

表 4.1: 提案手法の定量評価結果

Model	MSE ↓	PSNR ↑	SSIM ↑	LPIPS[mean] ↓
RRDB10, VGG=10/平面 入力 30×30×20	0.000307	35.47	0.9608	0.0324

表 4.2: VGG サンプル枚数の比較 (RRDB10, 入力 30 × 30 × 20 を固定)

VGG sampling	MSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Time/epoch [min]	200epochs [h]
0 枚/平面 (VGG なし)	0.000284	35.79	0.9662	0.0584	5.16	17.23
10 枚/平面 (提案: 計 30 枚)	0.000307	35.47	0.9608	0.0324	8.75	29.24
20 枚/平面 (計 60 枚)	0.000305	35.51	0.9609	0.0316	9.78	32.63
30 枚/平面 (計 90 枚)	0.000309	35.48	0.9606	0.0305	10.43	35.16
フル (計 320 枚)	0.000313	35.45	0.9608	0.0289	16.64	55.53

ンダムにサンプリングして損失を計算したにもかかわらず、学習の不安定化（発散や生成崩壊）は観測されず、安定に超解像が実行できた。このことは、VGG 特徴に基づく監督信号が、限定枚数のスライスからでも一定の有効性を持つ可能性を示唆する。以降では、サンプリング枚数や RRDB 規模の変更が画質指標および計算コストに与える影響を検証する。

### 4.3 実験 1：VGG ランダムサンプリング枚数の影響

本節では、VGG 特徴に基づく知覚損失の計算に用いるスライス枚数を変化させ、知覚品質 (LPIPS) と計算コスト (時間) のトレードオフを検証する。比較条件は、0 枚、10 枚/平面 (提案)、20 枚/平面、30 枚/平面、フルスライス (合計 320 枚) とする。結果を表 4.2 に示す。表の見方としては VGG sampling の列は各平面あたりのサンプリング枚数を示している。MSE は低いほうが良好であり、PSNR、SSIM は高いほうが良好である。LPIPS は低いほうが良好である。Time/epoch は 1 epoch あたりの学習時間を示し、200epochs は 200 epoch までの学習にかかった総学習時間を示している。

表 4.2 の結果を見ると、サンプリング枚数を 20 枚、30 枚、あるいはフルスライスへと増やすことで、LPIPS は 0.0324 (10 枚) から 0.0289 (フル) へと確実に改善している。したがって、最良のモデル性能を追求し、かつ計算時間やメモリ量に制約がない環境下であれば、VGG 損失計算におけるスライス枚数は可能な限り多く確保することが望ましいといえる。

しかしながら、フルスライス計算では 1 epoch あたり 16.64 min を要し、提案設定 (10 枚/平面) の 8.75 min/epoch に対して約 1.9 倍 (+7.89 min/epoch) である。また、200 epoch 換算の総学習時間も 55.53 h であり、提案設定の 29.24 h に対して約 1.9 倍 (+26.29 h) となる。一方で提案手法 (10 枚) は、VGG なし (0 枚) と比較して LPIPS を劇的に改善 (0.0584 → 0.0324) させており、「わずかなサンプリング数であっても、知覚品質の向上に大きく寄与する」ことが確認できる。また、10 枚から 30 枚へ増やしても LPIPS の改善幅は 0.0019 に留まっており、学習時間の増加コストに対して性能向上のリターンは逡減している。

表 4.3: RRDB 直列数の比較 (VGG=10/平面, 入力  $30 \times 30 \times 20$  を固定)

RRDB	MSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Time/epoch [min]	200epochs [h]
5	0.000306	35.47	0.9616	0.0344	8.35	27.85
10(提案)	0.000307	35.47	0.9608	0.0324	8.75	29.24
15	0.000302	35.55	0.9611	0.0311	9.20	30.77

以上のことから, 本研究の提案設定 (10 枚/平面) は, 膨大な 3 次元データセットに対し計算資源の制約がある状況下において, 実用的な学習時間で十分な知覚品質を持つモデルを構築する本条件下では実用的な折衷点であると結論付ける.

#### 4.4 実験 2: RRDB 規模 (ブロック数) の影響

本節では, Generator に含める RRDB の直列数を 5,10,15 と変化させ, 性能と計算コストの関係を比較する. 比較に際しては RRDB 数以外の条件を統一し, VGG は各平面 10 枚, 入力 LR パッチは  $30 \times 30 \times 20$  に固定した. 結果を表 4.3 に示す.

表 4.3 に RRDB 直列数の比較結果を示す. RRDB 数を増やすことで, LPIPS は  $0.0344 \rightarrow 0.0324 \rightarrow 0.0311$  と段階的に改善しており, 知覚品質の観点ではモデルの表現力増加が有効であることが分かる. 一方, MSE/PSNR は RRDB5 と RRDB10 でほぼ同等であり, RRDB15 で僅かな改善が見られる程度である.

RRDB5 は, RRDB10 (提案) と比較して MSE/PSNR が同等である一方, LPIPS は  $0.0344$  と高く (差  $0.0020$ ), 知覚品質では不利である. ただし学習時間は Total200 が  $27.85$  h となり, RRDB10 の  $29.24$  h に対して  $1.39$  h 短い ( $4.8\%$ ). 計算資源が厳しい条件では, RRDB5 は候補となり得る.

RRDB15 は RRDB10 に対して LPIPS を  $0.0324$  から  $0.0311$  へ改善し (差  $0.0013$ ), 同時に MSE も  $0.000307$  から  $0.000302$  へ低下した. 一方で Total200 は  $29.24$  h から  $30.77$  h へ  $1.53$  h 増加 ( $5.2\%$ ) しており, 性能向上は得られるものの計算コストの増加を伴う. また SSIM は RRDB5 が  $0.9616$  と僅差で最大であり, 指標間で最良構成が必ずしも一致しない点にも留意が必要である.

以上より, 知覚品質 (LPIPS) を優先し計算資源に余裕がある場合は RRDB15 が有利である. 一方, 本研究の主眼である「計算コストを抑えつつ一定の品質を維持する」観点では, RRDB10 は RRDB15 に近い定量性能を保ちながら計算増を抑えたバランスの良い設定であると結論付ける.

#### 4.5 実験 3: パッチサイズ (入力サイズ) の影響

本節では, パッチベース学習における入力サイズを変化させた場合の影響を比較する. 提案手法の  $30 \times 30 \times 20$  に対し,  $20 \times 20 \times 15$ ,  $15 \times 15 \times 10$  を評価する. パッチングはすべて 1 つのボリュームから 8 個のパッチ対を抽出する設定は共通である. VGG は各平面 10 枚, RRDB は 10 個直列に固定する. 結果を表 4.4 に示す.

表 4.4 より, 入力パッチを小さくすると学習時間は大幅に短縮される一方で, MSE/PSNR/SSIM および LPIPS が悪化する傾向が見られる. これは小パッチ化により, Generator が利用できる 3 次元コンテキストが制限され, 微細構造や連続性の復元が難しくなるためと考えられる. 以上より, 本研究の目的 (知覚品質を保

表 4.4: 入力パッチサイズの比較 (RRDB10,VGG=10/平面 を固定)

入力パッチ	MSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Time/epoch [min]	200epochs [h]
30 × 30 × 20 (提案)	0.000307	35.47	0.9608	0.0324	8.75	29.24
20 × 20 × 15	0.000319	35.32	0.9597	0.0414	2.41	10.51
15 × 15 × 10	0.000365	34.68	0.9544	0.0462	1.67	6.03

ちつつ計算量を抑える) に対しては, 30 × 30 × 20 を基準設定として採用する.

## 4.6 定性評価

本節では, 提案手法および比較手法によって生成された超解像画像の視覚的な品質について比較・検討を行う. 評価対象として, テストデータセット MSD Task01 BrainTumour[16] に含まれる症例のうち, 全スライスを通して目視比較を行った際に, 脳回 (Gyri) や脳溝 (Sulci) などの皮質微細構造の差異が最も判別しやすいと判断した症例 (ID: 530) を例として提示する.

スライス位置は, Axial ( $xy$ ) 断面としてボリュームの  $z$  軸方向における中央付近 (HR および SR では  $z = 76$ , LR では  $z = 19$ ) を採用した. 加えて, 3 次元的な再構成品質 (特に  $z$  軸方向の連続性) を確認するため, Sagittal ( $yz$ ) 断面として  $x$  軸方向の中央付近 (HR および SR では  $x = 120$ , LR では  $x = 30$ ) も併せて可視化した.

なお, 各画像の輝度スケールは公平な比較を行うため, 基準となる正解画像 (HR) の画素値分布における 99.5% パーセンタイル値を上限として, 全画像に対し統一的な正規化 (クリッピング) を行っている.

図 4.1 に, 入力画像 (LR), 正解画像 (HR), および各手法 (提案手法, Small Patch モデル, Full VGG モデル) による出力結果を示す. 各行は手法ごとの結果を示しており, 左列は Axial 断面, 右列は Sagittal 断面である. また, 各断面図においては左側にスライス全体像, 右側に赤枠内の拡大図を配置した.

### 4.6.1 視覚的品質の比較

図 4.1 に, 入力 LR, 参照 HR, および各手法による超解像結果の比較を示す. まず図 4.1(a) は入力 LR であり, 解像度不足に起因して全体に平滑化が強く, 脳溝の切れ込みや組織境界といった微細構造が十分に分離して見えにくい. 図 4.1(b) は教師データである HR であり, 本節ではこれを視覚的な基準として扱う.

図 4.1(c) は提案手法の結果である. 本手法は VGG 損失の計算に用いるスライス数を各平面 10 枚のランダムサンプリングに制限し, 計算コストを抑える設計としている. 図 4.1(d) は Small Patch モデルであり, 入力パッチサイズを 15 × 15 × 10 に縮小することで学習時間が最短となる構成である一方, 定量評価では指標が概ね低下した設定である. 図 4.1(e) は Full VGG モデルであり, 超解像後の解像度が 120 × 120 × 80 であることから, 3 平面の全スライス (合計 320 枚) に対して VGG 中間特徴量に基づく損失を計算する. 本研究で扱った範囲では学習時間が最長である一方, LPIPS は最も良好であった.

各結果を拡大図で比較すると, 図 4.1(d) の Small Patch は構造の復元自体は確認できるものの, 全体としてやや平滑化が残り, しわ状の高周波成分や細い境界の立ち上がり十分に強調されない箇所が見られる. これは表 4.4 における LPIPS の傾向とも整合的である.

図 4.1(e) の Full VGG は,3 構成の中では輪郭や溝の切れ込みがより明瞭に見える傾向があり, 図 4.1(b) の HR が持つ局所コントラストやテクスチャに近い印象を与える. ただし, 図 4.1(c) の提案手法との差は限定的であり, 少なくとも本図の範囲では, 注意深く観察して初めて判別できる程度の差に留まる.

最後に 図 4.1(c) の提案手法は, ランダムサンプリングにより VGG 損失計算を大幅に削減しているにもかかわらず, LR や Small Patch と比べて境界の見え方や微細構造の保持が良好である. 以上より, 提案手法は知覚品質を一定程度確保しつつ, Full VGG に比べて計算コストを抑える方向性を, 定性的にも支持する結果が得られた.

## 4.6.2 ランダム抽出スライスによる出力例

前節では, 皮質微細構造の差異が比較的判別しやすい症例 (ID:530) を例として詳細比較を行った. 一方で, 単一症例・特定断面の提示は選定バイアスを含む可能性がある. そこで補助的検証として, 提案手法により生成したテストデータの超解像結果 (SR) から, 症例 ID を所定範囲より無作為抽出し, 各症例について断面 (xy/xz/yz) およびスライス位置 (軸方向インデックス) をランダム (端部の影響を避けるためスライス位置は一定範囲に制限した) に選択して, 5 件の可視化例を作成した.

なお, 可視化における輝度スケールは前節と同様に, 各症例の正解画像 (HR) の画素値分布における 99.5% パーセンタイル値を上限 (下限 0) として設定し, 同一症例内で LR/SR/HR を同一スケールで表示した. 図 4.2 に, ランダム抽出したスライスに対する提案手法の出力例を示す.

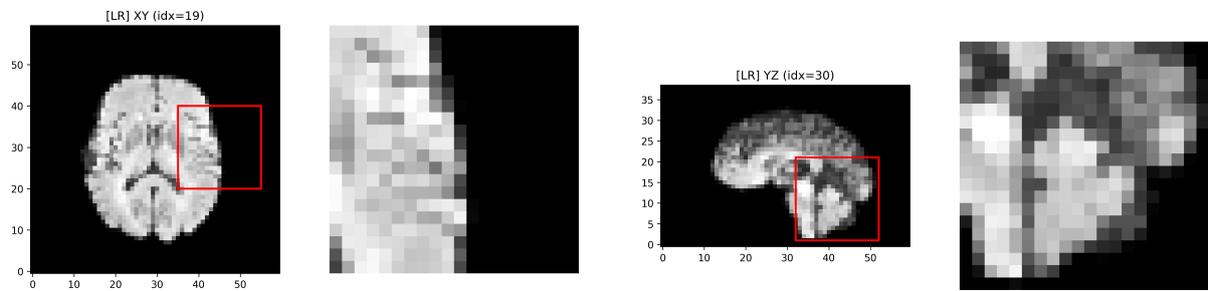
## 4.7 本章の総合考察

本章では, 提案手法の有効性を, 定量評価 (表 4.1) と, 構成要素ごとのアブレーション (表 4.2, 表 4.3, 表 4.4), ならびに定性評価 (図 4.1) により検証した. ここでは各実験結果を横断的に整理し, 「計算コスト削減」と「品質維持」の観点から得られた知見をまとめる.

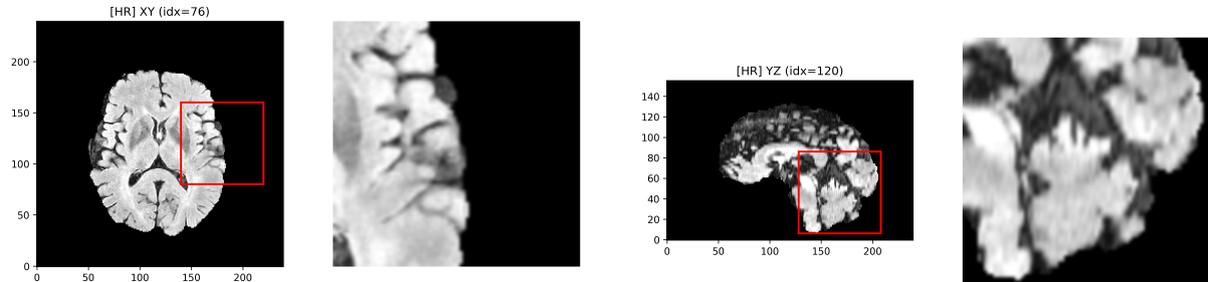
■**画素指標と知覚指標のトレードオフ** まず VGG 損失の有無に着目すると, 表 4.2 の通り, VGG なし (0 枚/平面) は MSE/PSNR/SSIM が最良である一方, LPIPS が大きく悪化した. これは, 画素単位損失を主に最適化した場合には平均化による平滑化が生じやすく, 見た目の自然さ (知覚品質) と画素指標が一致しないことを示している. したがって本研究の目的 (知覚品質の改善) に対しては, VGG 特徴に基づく知覚損失の導入が本質的である.

■**VGG サンプリング近似の有効性と逓減** VGG 損失の計算枚数を増やすほど LPIPS が改善する傾向は一貫して確認できるが, その改善幅は枚数の増加に対して逓減する (10 枚/平面以降で改善が緩やかになる). 一方で学習時間はサンプリング枚数に応じて増加し, 特にフルスライス (計 320 枚) は総学習時間 (200 epoch 換算) が 55.53 h となり, 提案設定 (10 枚/平面) の 29.24 h に対して約 1.9 倍 (+26.29 h) であった (表 4.2). このことから, 提案する「各平面 10 枚のランダムサンプリング」は, 知覚品質を大きく損なわずに計算コストを現実的な水準へ抑える近似として合理的である. 定性評価 (図 4.1) でも, Full VGG と提案手法の差は限定的であり, 定量結果と整合する.

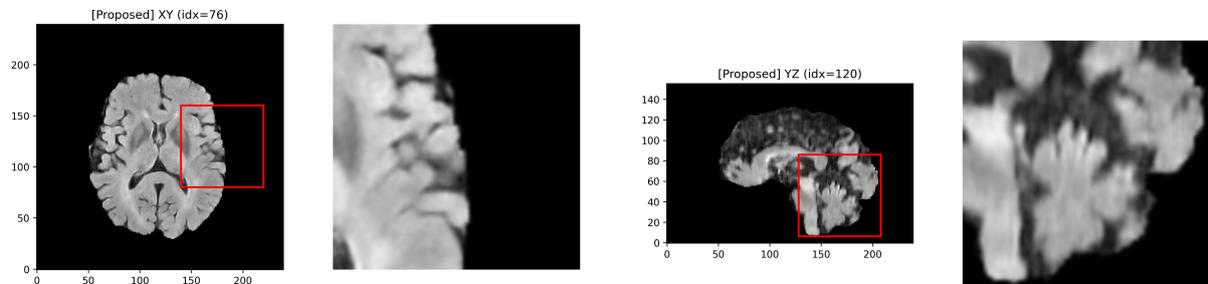
■**ネットワーク規模 (RRDB 数) の効果** RRDB 数の増加は, LPIPS や MSE/PSNR の改善に寄与する一方, 学習時間の増加を伴う (表 4.3). ただし 5-15 の範囲では性能差が僅差であり, 計算増に対する改善は大きくない. 本研究が志向する「計算資源制約下での実用性」を踏まえると, RRDB10 は RRDB15 に近い性能を保



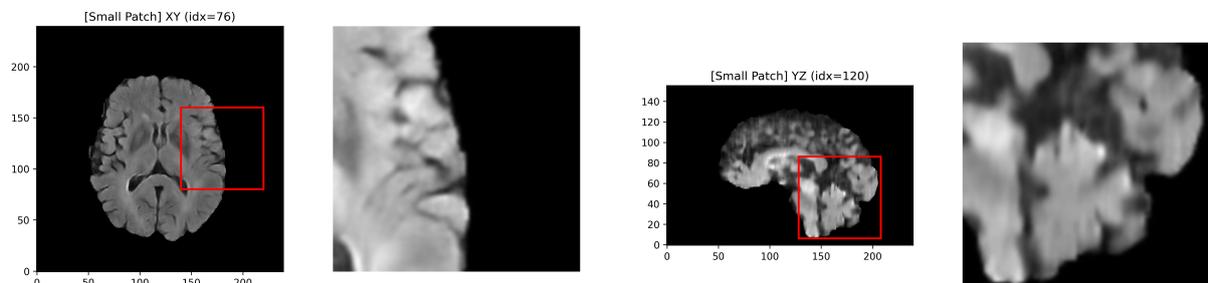
(a) Low-Resolution (LR)



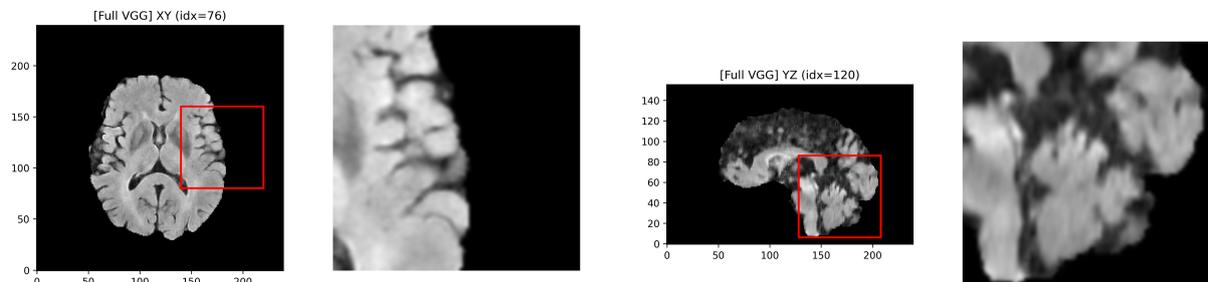
(b) High-Resolution (HR)



(c) Proposed Method



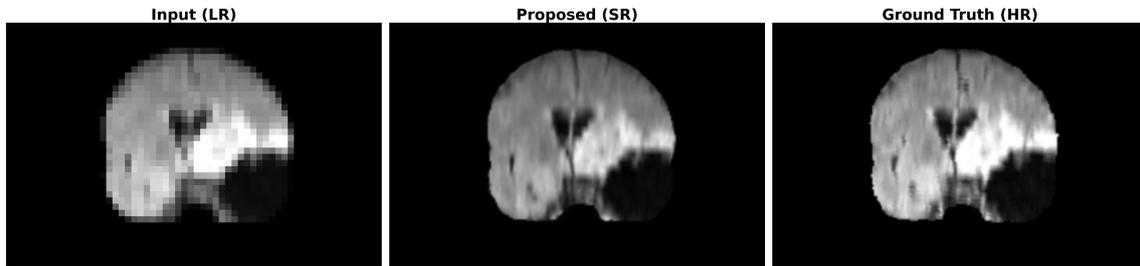
(d) Small Patch



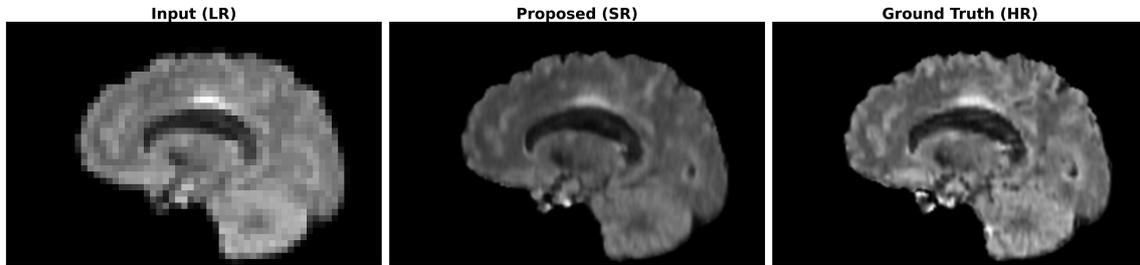
(e) Full VGG (Reference)

図 4.1: 代表的なテスト症例 (ID: 530) における定性評価の比較. 各行は手法ごとの結果を示し, 左列は Axial ( $xy$ ) 断面, 右列は Sagittal ( $yz$ ) 断面である. (a) 入力 LR 画像. (b) 正解 HR 画像. (c) 提案手法の結果. (d) Small Patch モデルの結果. (e) Full VGG モデルの結果.

Case ID: 584 | Plane: Coronal (XZ) | Slice: 125



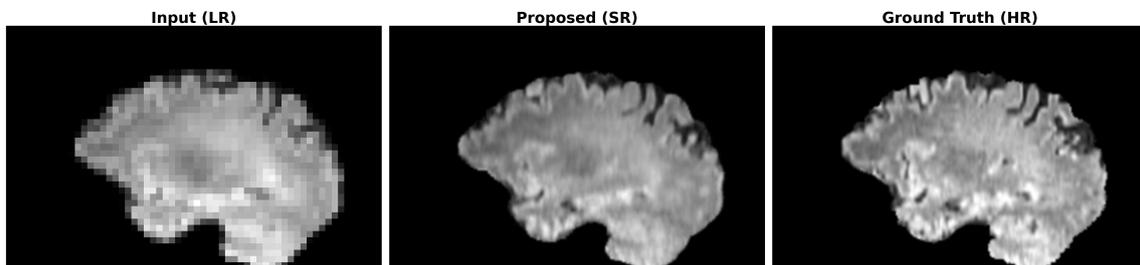
Case ID: 591 | Plane: Sagittal (YZ) | Slice: 134



Case ID: 598 | Plane: Coronal (XZ) | Slice: 106



Case ID: 613 | Plane: Sagittal (YZ) | Slice: 151



Case ID: 711 | Plane: Axial (XY) | Slice: 40

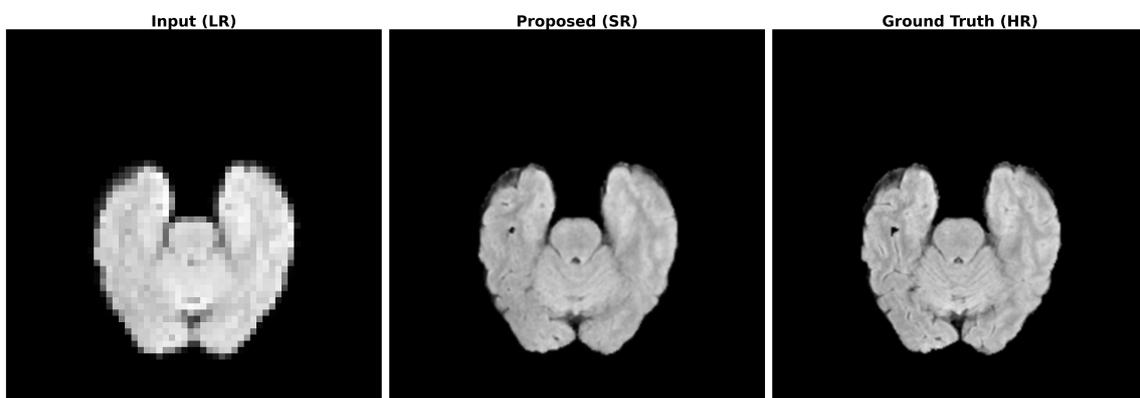


図 4.2: テストデータから無作為に抽出した 5 枚のスライスに対する提案手法の出力例. 左から, LR, SR, HR の並びになっている. 各画像の断面 (xy/xz/yz) とスライス位置は画像内に注記した.

ちながら計算増を抑えた, バランスの良い構成であると解釈できる.

■**パッチサイズと 3 次元コンテキスト** 入力パッチを小さくすると学習時間は大幅に短縮されるが, MSE/PSNR/SSIM に加えて LPIPS も悪化した (表 4.4). これは小パッチ化により, Generator が参照できる 3 次元コンテキストが制限され, 連続構造や高周波成分の復元が難しくなるためと考えられる. 定性評価でも Small Patch は平滑化が残りやすい傾向が確認でき, 定量結果を支持する (図 4.1).

■**総括** 以上を踏まえると, 提案設定 (RRDB10, 入力  $30 \times 30 \times 20$ , VGG サンプリング 10 枚/平面) は, (1) VGG 損失導入により知覚品質 (LPIPS) を大きく改善しつつ, (2) フルスライス計算に比べ学習時間を大幅に抑え, (3) ネットワーク規模およびパッチサイズの観点でも過度な計算増を避けた, 「計算コスト削減と品質維持」の両立を狙った実用的な折衷点として位置付けられる. なお, 本章で報告した学習時間は計算機の負荷や状態により変動し得るため, 絶対値ではなく相対比較の参考値として解釈する必要がある.

## 第5章

# おわりに

### 5.1 結論

本研究では、三次元医用画像（MRI）に対する4倍超解像を対象に、RRDB構造と2.5D VGG知覚損失、およびGAN学習を基盤とした枠組みの下で、学習時の計算コストを削減しながら実用上の画質を維持できる設計条件を整理した。特に、計算コストへ影響が大きい要因を(i)知覚損失計算、(ii)モデル規模(RRDB数)、(iii)学習入力(バッチサイズ)に分解し、統一条件下でアブレーション実験を行うことで、品質(MSE/PSNR/SSIM/LPIPS)と学習時間のトレードオフを定量的に示した。

第一に、VGG特徴に基づく知覚損失はLPIPSの改善に有効であり、画素指標(MSE/PSNR/SSIM)だけでは捉えにくい知覚品質の向上に寄与することを確認した(表4.2)。一方で、VGG損失の計算枚数を増やすほどLPIPSが改善する傾向はあるものの、改善幅は逓減し、計算時間はほぼ比例的に増加する。この点に対し、各平面10枚(計30枚)のランダムサンプリングは、フルスライス計算(計320枚)に比べて学習時間を大幅に短縮しつつ、定性的品質を一定程度維持できる近似として有効であることを示した(図4.1)。

第二に、RRDB直列数を増やすほどLPIPSやMSE/PSNRが改善する一方で、学習時間も単調に増加し、5-15の範囲では性能差が僅差に留まることを確認した(表4.3)。

第三に、入力パッチの小型化は学習時間短縮に直結するが、三次元的コンテキストが不足し、微細構造や高周波成分の復元に不利となり得ることを示した(表4.4, 図4.1)。

以上より、提案設定(RRDB10, 入力 $30 \times 30 \times 20$ , VGGサンプリング10枚/平面)は、知覚品質(LPIPS)を改善しつつ学習コストを抑えるという本研究の目的に対して、実用的な折衷点として位置付けられる。ただし、本研究の設計は「計算資源に制約がある状況で、学習・検証サイクルを現実的な時間で回す」ことを主眼としており、**品質を最優先し、学習リソースを度外視できる条件では最適解とは限らない**。そのような条件では、RRDB直列数を増やし、VGG損失もフルスライスで計算する構成が、より高い知覚品質を得るうえで有利になる可能性が高い(表4.2, 表4.3)。すなわち本研究の貢献は、「最高品質モデルの更新」そのものではなく、計算制約下での品質確保に向けた設計指針とトレードオフを整理し、近似(VGGサンプリング)を含む現実的な選択肢を提示した点にある。

なお、本研究で報告した学習時間は計算機の負荷や状態により変動し得るため、絶対値ではなく相対比較の参考値として解釈する必要がある。

## 5.2 今後の課題

本研究は、学習時の計算削減に焦点を当てて設計要因を整理した一方、以下の課題が残る。

■より広いデータ・条件での一般性検証 本研究は MSD Task01 BrainTumour[16] を用いて評価したが、撮像条件、臓器部位、モダリティ (CT 等) が異なる場合に同様の傾向が成立するかは追加検証が必要である。特に医用領域ではデータ分布の差が大きく、サンプリング枚数やモデル規模の最適点が変化する可能性がある。

■推論時コストおよび実運用指標の評価 本研究では学習時間を主な計算コスト指標として扱ったが、推論時間、GPU メモリ使用量、スループットなど、運用上重要な指標も併せて評価し、「学習が短いだけでなく推論も軽い」設計へ拡張する必要がある。また、品質指標についても、臨床的に重要な構造の保持に関する評価 (専門家評価や下流タスク性能など) を導入することで、医用応用としての妥当性をより強固にできる。

■知覚損失近似の高度化 本研究のランダムサンプリングは単純で実装容易である一方、病変周辺や高周波成分が多い領域を優先してサンプリングする重要度サンプリング、学習の進行に応じて枚数を動的に変更するスケジューリングなど、同等の計算量でより高い品質を得る余地がある。さらに、2D VGG に依存しない軽量特徴抽出器や、医用画像に適した自己教師あり特徴を用いた知覚損失の検討も今後の課題である。

■過生成 (ハルシネーション) リスクへの配慮 GAN を用いた高品質化は知覚品質の改善に有効である一方、医用画像では「存在しない構造」を生成するリスクが問題となり得る。歪み指標 (PSNR/SSIM) と知覚指標 (LPIPS) のバランスに加え、不確実性推定や保守的な損失設計など、安全側に倒す設計の検討が求められる。

以上の課題に取り組むことで、本研究で得られた設計指針をより一般化し、限られた計算資源下でも高品質な三次元医用画像超解像を現実的に運用可能な形へ発展させられると考える。

# 謝辞

本研究の遂行にあたり、終始懇切丁寧なご指導とご鞭撻を賜りました、指導教員である椋木雅之教授に深く感謝の意を表します。先生には、研究の方向性から論文の細部に至るまで、多角的な視点から貴重なご助言をいただきました。

また、本研究は倉永先輩が作成された実装を引き継いで進めたものであり、研究遂行の基盤として大きく寄与しました。ここに記して感謝申し上げます。

さらに、本研究で用いたデータセット (MSD Task01 BrainTumour) は Medical Segmentation Decathlon により公開されたものであり、貴重なデータを提供いただいた関係者各位に感謝いたします。

さらに、日々の研究生活において有益な助言と励ましをいただいた、椋木研究室の諸氏に感謝いたします。

## 参考文献

- [1] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.
- [2] Juhjung Ha, Nian Wang, Surendra Maharjan, and Xuhong Zhang. 3d volumetric super-resolution in radiology using 3d rrdg-gan. In *Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2024.
- [3] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
- [5] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [6] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [7] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160, 1981.
- [8] Erik H.W. Meijering, Wiro J. Niessen, and Max A. Viergever. Quantitative evaluation of convolution-based methods for medical image interpolation. *Medical Image Analysis*, 5(2):111–126, 2001.
- [9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 184–199, 2014.
- [10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [11] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

- [12] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, volume 27, pages 2672–2680, 2014.
- [13] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [14] 倉永 将宏. 三次元医療データ超解像への 3D-SRCGAN の応用. 修士論文, 宮崎大学大学院 工学研究科 工学専攻, 2024. 指導教員: 椋木 雅之.
- [15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
- [16] Amber L. Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M. Summers, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc Gollub, Jennifer Golia-Pernicka, Stephan H. Heckers, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Eugene Vorontsov, Lena Maier-Hein, and M. Jorge Cardoso. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.