

# Stacked U-Net による CT 再構成

水澤 悟

電気通信大学 大学院 情報理工学研究科  
博士（情報学）の学位論文

2022 年 3 月



電気通信大学 大学院 情報理工学研究科  
博士（情報学）の学位論文

博士論文審査委員会

- 主査 清雄一准教授  
委員 折原良平客員教授  
委員 庄野逸教授  
委員 柳井啓司教授  
委員 高橋裕樹准教授

著作権所有者

水澤 悟

2021 年

# Abstract

In X-ray CT image reconstruction, faster reconstruction speed enables quicker diagnosis, more accurate reconstruction and less noise makes it more difficult to miss diseases, and fewer X-rays reduce the burden on the patient. Therefore, research is being conducted to speed up the reconstruction, improve the accuracy, and reduce the amount of X-ray irradiation, respectively. The FBP method, which is a conventional direct method, is fast, but has the problem that it cannot be reconstructed when the number of X-rays is reduced and the noise becomes large. The iterative approximation method, which is another conventional algebraic method, can reduce the number of X-rays, but has the problem of slow reconstruction speed due to the iterative operation. In order to overcome these problems, CT reconstruction methods combining existing methods with deep learning methods have been proposed in recent years. For example, a method that removes noise from an image reconstructed by the FBP method using a deep learning model, and a method that replaces each step of the iterative approximation method with a deep learning model have been proposed as a combination of deep learning and existing methods. However, these methods are slower than the existing methods in terms of reconstruction speed because they combine deep learning with the existing methods. In this study, we proposed a model that can recover reconstructed images directly using only deep learning models (Stacked U-Net) without combining with existing methods. In addition, one of the challenges in medical applications of deep learning is that there are few images for training due to patient privacy issues. If the

number of training images is small, the accuracy of the deep learning model cannot be improved, and it is inferior to existing methods. To solve this problem, we proposed to use natural images for training. We showed that the model trained on natural images can recover medical images, and outperformed existing methods when the amount of X-ray irradiation is small. We also tried to solve the problem of the small number of images for training by using a different approach. In order to increase the amount of training data, there is a conventional method called data augmentation, which adds more modified data to the original training data to increase the amount of training data. In the conventional method, if the amount of modification is too large, correct learning cannot be achieved, so it is not possible to create more data than the original data. In contrast, we have proposed a method called interlayer augmentation. This is a method of augmenting data in a batch and in the middle of a model in batch learning, and the data can be augmented by a power depending on the number of layers inserted into the model. We proposed batch generalization (BG) and random BG (RBG) as specific algorithms, and applied them to a classification task, and showed that they improved the accuracy in learning from scratch. In addition, Stacked U-Net has the advantage that the accuracy is improved by simply stacking U-Net and is easy to create. On the other hand, if the number of stacks is increased too much, the accuracy deteriorates. In order to solve this problem, we focused on the skip-connection of the stacked U-Net and proposed a structure in which high frequency components can propagate easily. As a result of experiments with the proposed structure and increasing the number of stacks, we were able to construct a model that is resistant to the deterioration of accuracy due to the increase in the number of stacks. The method presented in this study is expected to be used for the development of CT equipment that enables rapid diagnosis with a small amount of X-ray exposure. In the future, it is necessary to improve the accuracy of the model by integrating the above methods and to evaluate

the optimal image reconstruction method combined with AI diagnosis.

## 概要

X 線 CT 画像再構成では再構成速度が早ければ素早く診断ができる, 再構成精度が良くノイズが少なければ, 疾病の見逃しがしづらくなる, X 線照射数が減れば患者の負荷が減る. といった点があるためそれぞれ再構成速度を早くし, 精度を良くし, X 線照射量を減らす研究が行われている. 従来手法の直接的な手法である FBP 法では, 計算速度は速いものの, X 線照射数を減らすと復元できない, ノイズが大きくなるという課題がある. 別の従来手法の代数的な手法である逐次近似法では X 線照射数を減らすことができるが, 繰り返し演算を行うため, 再構成速度が遅くなるという課題がある. それらの課題を克服するために, 近年既存手法とディープラーニング手法と組み合わせた CT の再構成手法が提案されている. ディープラーニング手法と既存手法を組み合わせたものとして, FBP 法により再構成された画像からディープラーニングのモデルによりノイズを除去する手法や, 逐次近似法の各ステップをディープラーニングモデルに置き換える手法が提案されている. しかしこれらは既存手法に加えてディープラーニングをさらに組み合わせているという点においてどうしても既存手法より再構成速度が遅くなってしまいう課題がある. そのため本研究では既存手法との組み合わせなしにディープラーニングのモデルだけで直接再構成画像を復元するモデル (Stacked U-Net) を提案した. またディープラーニングの医療適用上の課題として患者のプライバシーの問題から学習用画像が少ないという課題がある. 学習用画像がすくないとディープラーニングモデルの精度を上げることができず, 既存手法に劣ってしまう. 我々はこの課題に対して自然画像を用いて学習することを提案した. 自然画像で学習を行ったモデルにおいても, 医療用画像が復元できることを示し, X 線照射量が少ない場合において, 既存手法より優れた結果を示した. また, 学習用画像が少ないという課題に対して, 別のアプローチでも解決を狙った. 少ないデータを増やすためには従来手法として data augmentation と

いう、元の学習データに対して、修正を行ったデータをさらに加えることにより学習データを増やす手法がある。従来手法はあまりに修正量が大きいと正しい学習ができなくなるため、元のデータ数を大きく超えた数のデータを作ることはできない。これらに対し Interlayer augmentation という手法を提案した。これはバッチ学習においてバッチ内かつモデルの途中で data を augment する方法で、モデルに挿入するレイヤの数に応じて data をべき乗で増やすことができる。具体的なアルゴリズムとして batch generalization (BG) と random BG (RBG) を提案し、クラス分類タスクに適用し、1 からの学習において精度が向上することを示した。また、Stacked U-Net は U-Net を stack するだけで精度が良くなり、作成が簡単といった利点がある。一方 stack 数を増やしすぎると、精度が悪化していく。この課題を解決するために、Stacked U-Net の skip-connection に着目し、高周波成分が伝搬しやすい構造を提案した。提案した構造に対して stack 数を増やした実験を行った結果、stack 数の増加による精度悪化に耐性のあるモデルが構築できた。本研究で示した手法により、少ない X 線被ばく量で迅速に診断ができる CT 機器の開発が期待できる。今後、上記の手法を統合しよりモデルの精度を高める、AI 診断と組み合わせた最適な画像再構成法の評価を行うといったことが課題である。

# 目次

第 1 章	序論	14
1.1	研究の背景	14
1.2	直接法の課題	15
1.3	逐次近似法の課題	15
1.4	従来手法とディープラーニングの組み合わせ	16
1.4.1	ディープラーニングを組み合わせることによる課題	16
1.5	研究の目的と貢献	17
1.6	論文の構成	18
第 2 章	背景	21
2.1	X 線 CT 再構成	21
2.1.1	X 線 CT 装置による投影	21
2.1.2	CT 画像の再構成	23
2.2	従来手法	23
2.2.1	直接法	23
2.2.2	FBP 法	25
2.3	代数的手法	25
2.3.1	CT 再構成問題の代数的定式化	26
2.3.2	SART 法	27
2.3.3	ML-EM 法	27
2.3.4	OS-EM 法	29
第 3 章	Stacked U-Net による CT 再構成	31

---

3.1	はじめに . . . . .	31
3.2	関連研究 . . . . .	33
3.2.1	MRI 再構成へのディープラーニングの適用 . . . . .	33
3.2.2	CT 再構成へのディープラーニングの適用 . . . . .	34
3.2.3	U-Net とその改良である Stacked U-Net . . . . .	34
3.2.4	医用におけるトレーニングデータセット . . . . .	34
3.3	Stacked U-Net の構造 . . . . .	35
3.4	予備実験 . . . . .	37
3.4.1	32 × 32 画像の復元でのレイヤ関数の選択 . . . . .	37
3.4.2	32 × 32 画像での復元の考察 . . . . .	40
3.5	実験結果 . . . . .	41
3.5.1	512 × 512 の画像の復元 . . . . .	41
3.5.2	医用画像での結果 . . . . .	42
3.5.3	FBP U-Net を ImageNet で学習させたときとの比較 . . . . .	45
3.5.4	セグメンテーションタスクを用いた再構成結果画像の客観 評価 . . . . .	46
3.6	考察 . . . . .	47
3.7	まとめ . . . . .	52
第 4 章	Interlayer Augmentation による学習データ増強	57
4.1	はじめに . . . . .	57
4.2	関連研究 . . . . .	59
4.2.1	正則化 (Regularization) . . . . .	59
4.2.2	データ増強 (Data augmentation) . . . . .	59
4.2.3	特徴空間データ増強 (Feature space augmentation) . . . . .	59
4.3	データの分布距離の分析 . . . . .	60
4.4	レイヤ間データ増強 (Interlayer augmentation method) . . . . .	61
4.4.1	RBG (Random Batch Generalization) . . . . .	61
4.4.2	BG (Batch Generalization) . . . . .	63
4.5	実験結果 . . . . .	63

4.5.1	クラス分類タスクへの適用 . . . . .	63
4.5.2	CT 再構成タスクへの適用 . . . . .	65
4.6	考察 . . . . .	68
4.6.1	クラス分類タスク考察 . . . . .	68
	256 × 256 画像での評価 . . . . .	70
	CutMix との併用評価 . . . . .	71
4.6.2	CT 再構成タスク考察 . . . . .	72
	補間関数の内挿化 . . . . .	73
4.7	まとめ . . . . .	74
第 5 章	Stacked U-Net の改良	76
5.1	はじめに . . . . .	76
5.2	関連研究 . . . . .	77
5.2.1	U-Net の改良 . . . . .	77
5.2.2	Stacked U-net . . . . .	77
5.3	Skip Connection の改良 . . . . .	78
5.4	stack 数の増加に耐える skip connection 構造 . . . . .	78
5.5	実験結果 . . . . .	83
5.5.1	セグメンテーションタスク . . . . .	84
5.5.2	セグメンテーションタスクの結果考察 . . . . .	84
5.5.3	CT 再構成問題 . . . . .	85
5.5.4	CT 再構成の考察 . . . . .	87
5.5.5	CT 再構成問題での skip connection の周波数解析 . . . . .	88
5.6	考察 . . . . .	90
5.7	まとめ . . . . .	90
第 6 章	結論	92
6.1	今後の展望 . . . . .	94
	発表実績	96
	謝辞	98

参考文献	99
著者略歴	112

## 目次

1.1	Overview of CT Reconstruction. . . . .	14
1.2	Issues CT reconstructions and algorithm correspondence. . .	17
1.3	Issues deeplearning for medical and correspondence. . . . .	17
2.1	Conceptual diagram of X-ray CT scanner. . . . .	22
2.2	X-ray, generator and detector detail. . . . .	22
2.3	Sinogram image. . . . .	23
2.4	Conceptual diagram of sinogram image. . . . .	26
2.5	Algebraic image. . . . .	26
3.1	List of open access datasets for medical study and research [1].	35
3.2	Structure of the proposed U-Net. . . . .	36
3.3	Stacked U-Net . . . . .	36
3.4	Connect previous U-Net input to next U-Net input. . . . .	39
3.5	Connect stacked U-Net input to all U-Net input. . . . .	39
3.6	Comparison of $32 \times 32$ images. . . . .	40
3.7	Results for validation sets . . . . .	42
3.8	TV-EM projection 64 result. . . . .	43
3.9	Reconstructed image with 64 projections. . . . .	44
3.10	Comparison of details with 64 projections. . . . .	45
3.11	Results of medical images of FBP U-Net trained by ImageNet at projection number 64. . . . .	46
3.12	Results of reconstructing medical images with 512 projections.	47

---

3.13	Intermediate output corresponding to the validation set with the number of projections equal to 512. . . . .	48
3.14	Reconstruction of a phantom by the SART method. . . . .	49
4.1	Distribution of feature values. . . . .	60
4.2	vgg's RGB and BG insert positions. . . . .	65
4.3	resnet's RGB and BG insert positions. . . . .	65
4.4	ViT's RGB and BG insert positions. . . . .	66
4.5	4 stack Stacked U-Net and insert positions. . . . .	67
4.6	resnet110 cifar100 train curve. . . . .	68
4.7	resnet110 cifar100 test curve. . . . .	68
4.8	resnet110 cifar100 accuracy. . . . .	69
4.9	Insertion number 3 learning curve. . . . .	72
4.10	Example of image intrpolation by RGB. . . . .	73
4.11	Insertion points. . . . .	73
5.1	Structure U-Net. . . . .	79
5.2	Encoding layer's in U-Net. . . . .	79
5.3	Encoding layer. . . . .	80
5.4	Decoding layer's in U-Net. . . . .	80
5.5	Decoding layer. . . . .	80
5.6	Skip connection between encoder and decoder layers. . . . .	80
5.7	Skip Normal. . . . .	81
5.8	Skip Down. . . . .	82
5.9	Skip Feature. . . . .	83
5.10	Input image. . . . .	85
5.11	Output image. . . . .	85
5.12	Ground truth. . . . .	85
5.13	Segmentation results. . . . .	85
5.14	Difference in gradient part. . . . .	85
5.15	Disappearing objects. . . . .	85

5.16	Structure of U-Net for CT reconstruction. . . . .	86
5.17	Skip Normal best result stack:4. . . . .	87
5.18	Skip Down best result stack:2. . . . .	87
5.19	Skip Feature best result stack:4. . . . .	87
5.20	Skip Down result stack:4. . . . .	88
5.21	Layer 1. . . . .	88
5.22	Layer 2. . . . .	88
5.23	Layer 3. . . . .	88
5.24	Frequency of skip connections at stack 1. . . . .	88
5.25	Layer 1. . . . .	88
5.26	Layer 2. . . . .	88
5.27	Layer 3. . . . .	88
5.28	Frequency of skip connections at stack 2. . . . .	88
5.29	Layer 1. . . . .	89
5.30	Layer 2. . . . .	89
5.31	Layer 3. . . . .	89
5.32	Frequency of skip connections at stack 3. . . . .	89
5.33	Layer 1. . . . .	89
5.34	Layer 2. . . . .	89
5.35	Layer 3. . . . .	89
5.36	Frequency of skip connections at stack 4. . . . .	89
6.1	Issues CT reconstructions and algorithm correspondence. . .	93
6.2	Issues deeplearning for medical and correspondence. . . . .	93

# 表目次

3.1	Applied methods . . . . .	38
3.2	Relationship between the number of stacks and evaluation values for $32 \times 32$ . . . . .	40
3.3	Comparison of the proposed method and existing ones. . . . .	54
3.4	Comparison of details with 64 projections. . . . .	55
3.5	Comparison of segmentation task. . . . .	56
4.1	Statistic of feature space. . . . .	60
4.2	cifar10 results. . . . .	64
4.3	cifar100 results. . . . .	66
4.4	Per epoch execution time for each augment rate. . . . .	67
4.5	Results of RBG on CT Reconstruction Problem . . . . .	67
4.6	Imangenette 256 results. . . . .	71
4.7	CutMix combination results. . . . .	71
4.8	Impoved RBG's mean and variance. . . . .	74
5.1	Segmentation task stack results. . . . .	84
5.2	CT reconstruction task stack results. . . . .	86

# 第1章

## 序論

本章では、まず背景として CT 再構成の説明とその課題を示す。次にディープラーニングを従来手法に組み合わせて適用した時の課題を合わせて説明する。最後に研究の目的と貢献、論文の構成を説明する。

### 1.1 研究の背景

X 線 CT 画像再構成では再構成速度が早ければ素早く診断ができる。再構成精度が良くノイズが少なければ、疾病の見逃しがしづらくなる。X 線照射数が減れば患者の負荷が減る。といった点があるためそれぞれ再構成速度を早くし、精度を良くし、X 線照射量を減らす研究が行われている。

X 線 CT 画像再構成は、X 線検出器を様々な角度で対象物に照射して撮影したサイノグラム画像から、断面画像を再構成する処理のことである。投影とは、発生器で

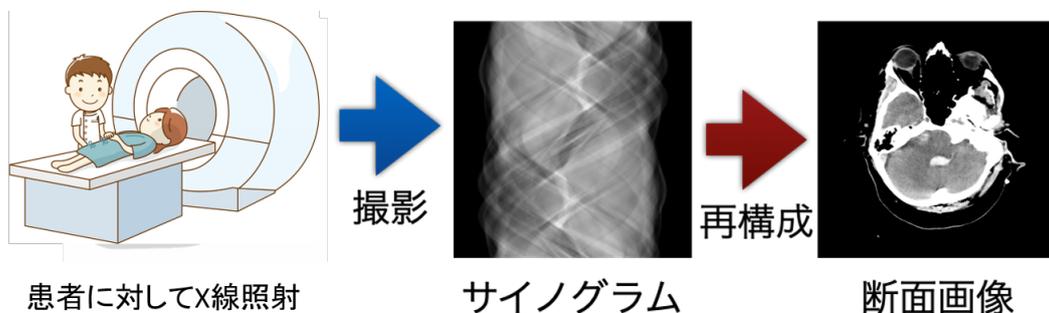


図 1.1: Overview of CT Reconstruction.

作られた X 線を対象物に照射し、反対側の検出器で撮影することである。対象物を

さまざまな角度で投影し、各角度で得られた画像を組み合わせることでサイノグラムを構成する。検出器の数と投影角度によって、サイノグラム画像の座標が決まる。検出器に観測された X 線の強度は、対象物の性質に応じて特定の範囲で減衰する。減衰の程度は、X 線の経路上にある物体の特性に依存する。例えば、X 線の通り道に骨や他の障害物があると、検出器で検出される X 線の強度はそれに応じて減衰する。つまり検出器は、X 線の通り道にある物体の X 線吸収の結果を積算して観測する。このようにして得られたサイノグラム画像をもとに、X 線 CT 画像再構成により、対象物の分布を表す断面画像を得ることができる。

X 線 CT 画像再構成には、大きく分けて「直接法」と「逐次近似法」の 2 つの方法がある。直接法は、FBP (Filtered Back-Projection) 法が代表的なアルゴリズムとして存在する。逐次近似法としては SART (Simultaneous Algebraic Reconstruction Technique) がある。直接法と逐次近似法は、計算コスト、ノイズ、アーティファクトの点で違いがある。

## 1.2 直接法の課題

直接法は解析的手法でありは計算コストが低いが、再構成時の投影回数が不足すると、得られる画像のノイズやアーチファクトが増えてしまう。

代表的な手法である FBP 法は、フィルタ計算とフーリエ逆変換を組み合わせた解析的な手法であり、様々な CT 装置に採用されている。しかし、ナイキスト周波数以下の投影結果に FBP 法を適用すると、再構成された断面画像に顕著なアーチファクトとノイズが発生する。例えば、肥満の患者の CT 画像には大きなノイズが乗り、石灰化やステントはブルーミングアーチファクトを、金属インプラントや骨構造はストリークアーチファクトを引き起こす。そのため、ノイズを減らす研究が行われている。

## 1.3 逐次近似法の課題

逐次近似法は低ノイズ、低アーチファクトの結果を得ることができるが、計算コストが直接法に比べ高い。

逐次近似法は代数的手法であり、初期値を反復的に修正しながら問題の解に近づい

ていくことで解を求める。この方法では、断面画像のスパース性を考慮しているため、ナイキスト周波数以下の場合でもアーチファクトやノイズを低減することができる。しかし、この方法では繰り返し演算を行う必要があるため、計算コストが高くなってしまう。逐次近似法には代表的なものとして SART 法や ML-EM (Maximum Likelihood- Expectation Maximization) 法がある。

医療分野では、X 線が患者に与える影響を抑えるために、投影回数をできるだけ少なくする必要がある。そのため、近年では逐次近似法の開発が活発に行われている。臨床の現場に適応するために、計算コストと再構成に要する時間を短縮する研究が行われている [2,3]。逐次近似法による再構成に要する時間は、10~90 分程度である。これでは緊急性の高い患者には使えない。しかし、肥満の患者など診断が難しいとされる患者からも、ノイズが少ない画像を獲得できる。さらに放射線量を減らすことで、肺がんや大腸がん、小児の画像診断などのスクリーニング検査などへの応用が期待できる [4]。もちろん COVID-19 のスクリーニング検査においても患者の負担軽減につながる。

## 1.4 従来手法とディープラーニングの組み合わせ

上記 FBP 法や SART 法の課題を克服するために、既存手法とディープラーニング手法と組み合わせた CT の再構成手法が提案されている。ディープラーニング手法と既存手法を組み合わせたものとして、FBP 法により再構成された画像からディープラーニングのモデルによりノイズを除去する手法や、逐次近似法の各ステップをディープラーニングモデルに置き換える手法が提案されている。

### 1.4.1 ディープラーニングを組み合わせることによる課題

上記手法は既存手法の弱点をディープラーニングによりカバーする手法だが、既存手法に加えてディープラーニングの計算を実施するという点において、既存手法よりも再構成の速度が遅くなってしまうのが課題である。またディープラーニングの医療適用上の課題として患者のプライバシーの問題から学習用画像が少ないという課題がある。ディープラーニングは単に画像を多くするだけで精度が上がっていくという利点がある一方、学習用画像がすくなくと精度を上げることができず、既存手法に劣っ

てしまう。

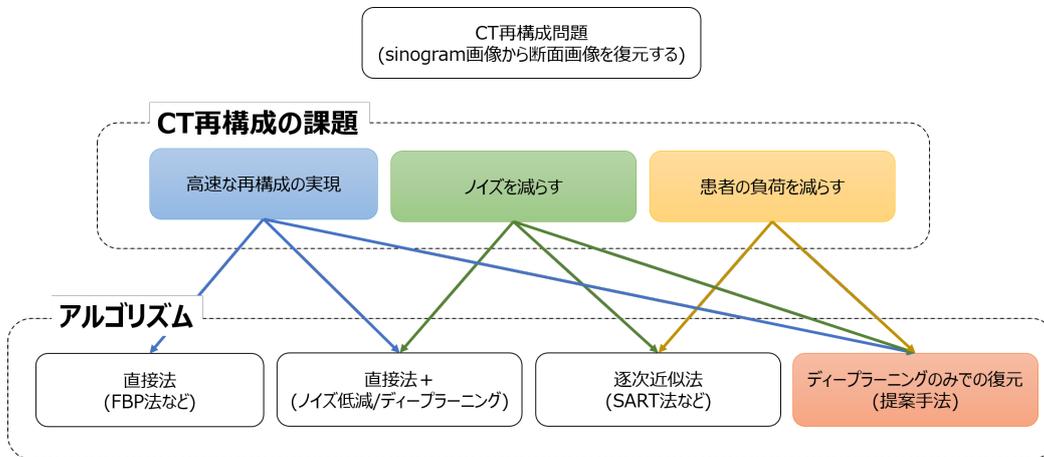


図 1.2: Issues CT reconstructions and algorithm correspondence.

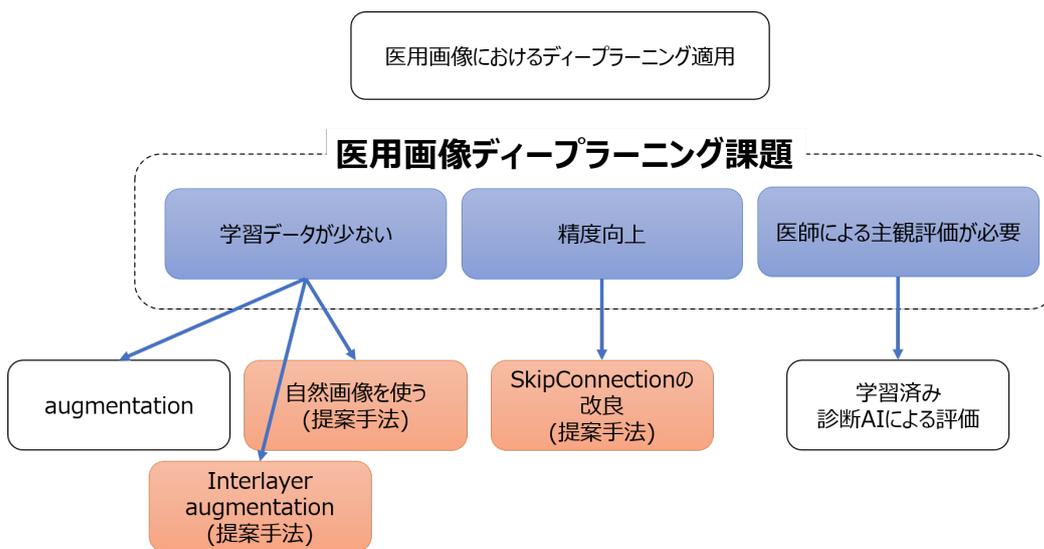


図 1.3: Issues deeplearning for medical and correspondence.

## 1.5 研究の目的と貢献

本研究の目的は CT の再構成において X 線投影数が少なく、短時間で再構成できる手法を作成し、診断を患者の負荷を少なく実行できるようにすることである。特に今後発展する可能性の高い AI 診断に有用な手法をめざした。

既存手法にはそれぞれ課題があるが、以下に示すように課題に取り組んだ。

既存手法とディープラーニング手法の組み合わせでは、再構成の速度が既存手法より遅くなるという課題に対しては、ディープラーニングのモデル (Stacked U-Net) だけで直接再構成画像を復元できることを示した。またディープラーニングの医療適用上の課題である患者のプライバシーの問題から学習用画像が少ないという課題に対しては自然画像を用いて学習することを提案した。自然画像で学習を行ったモデルにおいても、医療用画像が復元できることを示し、X 線照射量が少ない場合において、既存手法より優れた結果を示した。

また、学習用画像が少ないという課題に対して、別のアプローチでも解決を狙った。少ないデータを増やすためには従来手法として data augmentation という、元の学習データに対して、修正を行ったデータをさらに加えることにより学習データを増やす手法がある。従来手法はあまりに修正量が大きいと正しい学習ができなくなるため、元のデータ数を大きく超えた数のデータを作ることはできない。これらに対し Interlayer augmentation という手法を提案した。これはバッチ学習においてバッチ内かつモデルの途中でデータを augment する方法で、モデルに挿入するレイヤの数に応じてデータをべき乗で増やすことができる。具体的なアルゴリズムとして batch generalization (BG) と random BG (RBG) を提案し、クラス分類タスクに適用し、1 からの学習において精度が向上することを示した。

また、Stacked U-Net は U-Net を stack するだけで精度が良くなり、作成が簡単といった利点がある。一方 stack 数を増やしすぎると、精度が悪化していく。この課題を解決するために、Stacked U-Net の skip-connection に着目し、高周波成分が伝搬しやすい構造を提案した。提案した構造に対して stack 数を増やした実験を行った結果、stack 数の増加による精度悪化に耐性のあるモデルが構築できた。

本研究で示した手法により、少ない X 線被ばく量で迅速に診断ができる CT 機器の開発が期待できる。

## 1.6 論文の構成

本論分は全 6 章から構成される。以下に各章の概要を述べる。

第 1 章「序論」では、本研究の背景として CT 再構成およびその課題を示し、本論分の目的、構成を述べた。

第 2 章「研究背景」では本研究の前提となる CT 再構成の原理と典型的な再構成手法である FBP 法, SART 法, ML-EM 法について整理した。

第 3 章「**Stacked U-Net** による CT 再構成」では Stacked U-Net により従来手法との組み合わせなしに, ディープラーニングのモデルだけでダイレクトに再構成ができることを示した。医用領域にディープラーニングを適用する際の課題として, 学習用データが集めにくいという課題があるが, 自然画のデータセットにより再構成を学習させることによりこれを解決した。医用画像での評価の結果 FBP, SART, TV-norm, FBP U-Net, DD-Net と比べて低投影数の時に優れた結果を示した。また画像の評価として, セグメンテーションモデルを用いた定量評価を行い, FBP, SART, TV-norm, FBP U-Net, DD-Net と比べて低投影数の時に優れた結果を示した。

上記提案モデルは, 精度向上の余地がある。そのため, augmentation 手法およびモデルの改良それぞれについて検討を行った。

第 4 章「**Interlayer Augmentation** による学習データ増強」ではディープラーニング手法と学習データ量の近年の傾向について述べ, 医用画像では大きな学習データを用意できないため, 学習データを増やすための augmentation 手法について述べた。従来の augmentation 手法は再利用性や汎用性に制限があるといわれており, また事前に学習データを増やす必要がある。そのため, 既存モデルに挿入するだけで適用できる Interlayer augmentation を提案した。Interlayer augmentation をクラス分類タスク (CIFAR10, CIFAR100) に適用し, VGG, ResNet, ViT それぞれに適用し, Interlayer augmentation を適用しない場合と比較し適用しない場合と比べて, 優れた結果が得られることを示した。

第 5 章「**Stacked U-Net** の改良」では精度向上の検討のため, U-Net 自体の改良手法, Stacked U-Net の改良手法について述べた。Stacked U-Net は stack 数を増やしすぎると精度が悪化するという課題があるが, この課題に対して, U-Net の skip connection に着目し高周波が伝わりやすい構造を提案した。それぞれの構造についてセグメンテーションタスクや, CT 再構成問題に適用し評価した。結果 CT 再構成問題では, skip connection の改良により優れた結果が得られることを示した。

第 6 章「まとめ」では, 本論分を通して得られた結果をまとめ, 本論分で提案した

Stacked U-Net による CT 再構成や Interlayer augmentation の今後の展望について述べた。

## 第 2 章

# 背景

本章では X 線 CT 再構成の構成や、問題設定を示し、代表的な再構成手法である、直接法の FBP 法、逐次近似法の SART 法、ML-EM 法について述べる。

### 2.1 X 線 CT 再構成

CT による患者の輪切り画像 (断面画像) は 1. CT 機器による投影, 2. 投影された画像を元に画像を再構成するというステップにより獲得される。

#### 2.1.1 X 線 CT 装置による投影

まず X 線 CT 装置の構成の説明, および X 線 CT 装置による撮影 (投影) について説明を行う。

図 2.1 は X 線 CT 装置の模式図である。X 線 CT 装置は円筒に配置された X 線照射装置 (X-ray generator) および X 線検出器 (detector) で構成されている。患者 (patient) は円筒の中に入り X 線 CT 装置で撮影される。

X 線照射機, X 線および検出器の対応関係を図 2.2 に示す。X 線は照射機から、平行に照射される。X 線照射機から照射された X 線は検出器に当たる。検出器は位置によって分割されており、それぞれ X 線が当たった位置を把握できるようになっている。

X 線 CT 装置による撮影は投影と呼ばれる。

患者は X 線照射装置および X 線検出器の間に入ることで、X 線照射装置から送出された X 線が患者の体を通過し減衰し、検出器で X 線の強度が観測される。X 線

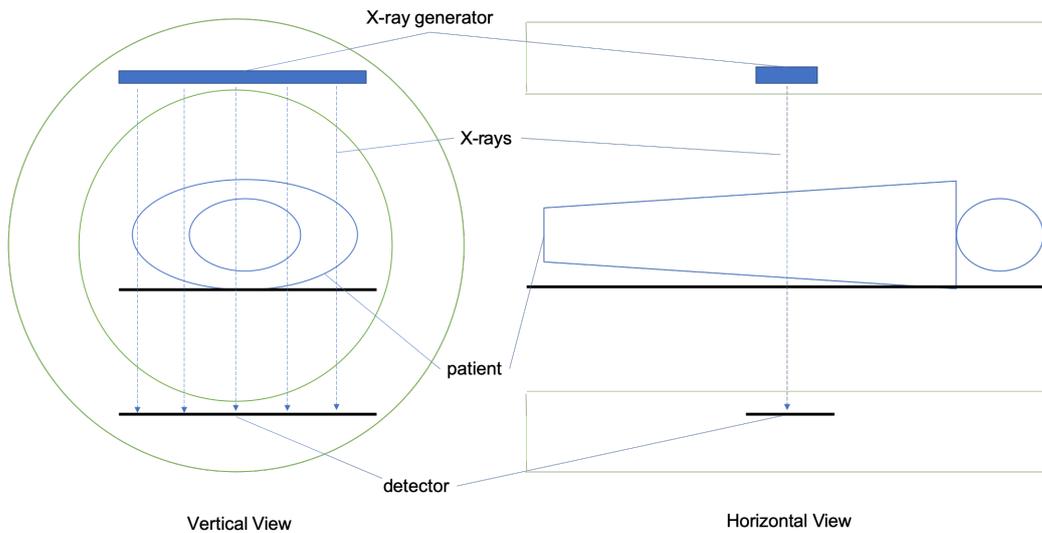


図 2.1: Conceptual diagram of X-ray CT scanner.

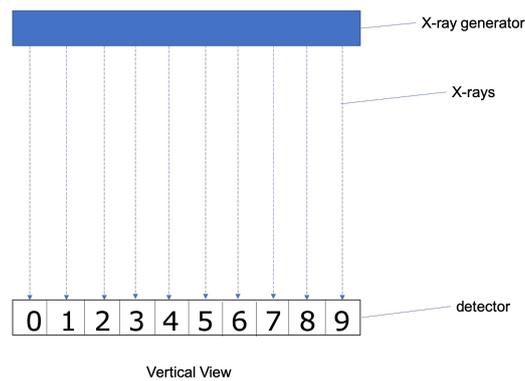


図 2.2: X-ray, generator and detector detail.

照射装置および X 線検出器が縦方向を軸に患者の周りを回することで、様々な角度の患者の体の情報を観察することができる。縦方向を軸にした回転および X 線照射を連続して行うことで、患者の周りを 360 度回った情報を得ることができる。このように角度ごとの X 線強度を観測することが投影である。また X 線照射装置および X 線検出器が患者の横方向に移動することで、様々な箇所の患者の体の情報を観察することができる。

X 線検出器で検出された画像は図 2.3 に示すようなサイノグラム (sinogram) 画像という画像になる。サイノグラム画像は横軸が検出器の位置、縦軸が X 線照射時の角度となっているため、人の目には理解できない画像となっている。

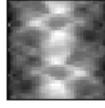


図 2.3: Sinogram image.

### 2.1.2 CT 画像の再構成

上記の投影により獲得されたサイノグラム画像を元に横軸を体の横位置、縦軸を体の縦位置とし人間の断面を計算するのが CT 再構成問題である。

CT 再構成問題は、Radon 定理”全ての角度から透視した時、得られる無限の投影データから 2次元、3次元のデータを再構成できる” [5,6] を元にしており、この原理を人体に応用したのが CT 機器である。人間の体は骨や水分などの分布の違いにより、X 線の通り方に違いがありこれが X 線強度の違いとして現れる。そのため、観測された X 線強度には人体の構成の情報が乗っている。そのため、直接法や逐次近似法を用いて断面画像を復元できる。

## 2.2 従来手法

### 2.2.1 直接法

CT の投影は対象の周りを回転し積分するため以下の radon 変換として示すことができ、投影面切断定理によりフーリエ変換により逆 radon 変換を解くことができる [7]。以下にそれを示す [8]。

$s$  を X 線照射機と検出機とを結ぶ線分とすると、 $s$  は下記になる。

$$x \cos \theta + y \sin \theta = s \quad (2.1)$$

ここで  $(x, y)$  は物体の位置、 $\theta$  は回転角度である。

$s$  および  $\theta$  を用いて、radon 変換は以下に示す式で表現できる。

$$p(s, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy \quad (2.2)$$

ここで  $f(x, y)$  は物体の位置  $(x, y)$  での X 線減衰係数を表す。 $\delta$  はディラックのデルタ関数である、

$p(s, \theta)$  は  $s$  の線積分つまり、X 線照射機と検出機とを結ぶ線分にある物体の X 線減衰係数の積分となる。  $p(s, \theta)$  は検出器での観測値であるので、  $p(s, \theta)$  から  $f(x, y)$  が解ければ物体の情報が得られる。これを逆 radon 変換という。

逆 radon 変換は投影面切断定理 (Fourier Slice Theorem) により実現できる。投影面切断定理は CT の投影結果 (サイノグラム画像) の 1 次元フーリエ変換が、断面の 2 次元フーリエ変換と一致するという定理である。具体的には以下から導くことができる。

$p(s, \theta)$  のフーリエ変換を  $P(\omega, \theta)$  とすると

$$P(\omega, \theta) = \int_{-\infty}^{\infty} p(s, \theta) e^{-2\pi i \omega s} ds \quad (2.3)$$

となる。

上述した  $p(s, \theta)$  の定義にもとづくと

$$P(\omega, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy e^{-2\pi i \omega s} ds \quad (2.4)$$

となり、積分の順序を整理すると

$$P(\omega, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \int_{-\infty}^{\infty} \delta(x \cos \theta + y \sin \theta - s) e^{-2\pi i \omega s} ds dx dy \quad (2.5)$$

となる。デルタ関数の積分は 1 なので、

$$P(\omega, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i (x \cos \theta + y \sin \theta) \omega} dx dy \quad (2.6)$$

となる。

最後に 2 次元フーリエ変換の定義を用いると

$$P(\omega, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i (xu + yu)} dx dy \quad (2.7)$$

となる。ここで  $u$  は  $\omega \cos \theta$ ,  $y$  は  $\omega \sin \theta$  である。

つまり、左辺はサイノグラム画像の 1 次元フーリエ変換であり、右辺は  $f(x, y)$  の 2 次元フーリエ変換、つまり断面画像のフーリエ変換である。

これは下記のようにあわせ

$$P(\omega, \theta) = F(\omega \cos \theta, \omega \sin \theta) = F_{polar}(\omega, \theta) \quad (2.8)$$

$\theta$  を変化させることで、極座標  $(\omega, \theta)$  における未知の関数  $f(x, y)$  の完全なフーリエ変換  $F_{polar}$  を得ることができるといえる。

投影面切断定理に基づけばサイノグラム画像に対して 1 次元フーリエ変換を行い、2 次元フーリエ逆変換を行うと断面画像が得られる。

### 2.2.2 FBP 法

FBP (Filtered Back Projection 法) は投影面切断定理に基づき、フィルタ計算とフーリエ逆変換により断面画像を復元する方法である。以下にその導出方法を示す。

フーリエ逆変換を  $F(u, v)$  とすると  $f(x, y)$  は

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{-2\pi i(xu + yv)} du dv \quad (2.9)$$

と表せる。

これを極座標表現  $F_{polar}(\omega, \theta)$  に直すと

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} F_{polar}(\omega, \theta) |\omega| e^{2\pi i\omega(x \cos \theta + y \sin \theta)} d\omega d\theta \quad (2.10)$$

となる。

投影面切断定理に基づくと上記は

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} P(\omega, \theta) |\omega| e^{2\pi i\omega(x \cos \theta + y \sin \theta)} d\omega d\theta \quad (2.11)$$

となる。  $P(\omega, \theta) |\omega|$  は一次元フーリエ変換に対するフィルタ計算 (Ramp フィルタ) と考えることができる。つまり、サイノグラム画像に対して 1 次元フーリエ変換を行い、フィルタ計算を行い、2 次元フーリエ逆変換を行うと断面画像が得られるということである。これが FBP 法となる。

## 2.3 代数的手法

FBP 法は解析的手法に基づいていたが、一方投影角度や検出器の数は有限である。その点からみると、サイノグラム画像は図 2.4 に示すように獲得されることができると考えることができる。図 2.4 の真ん中の四角部分を人体の断面 (断面画像) とすると、X 線検出器では X の通り道にある、人体の構造を加算した値が検出される。

上述したようにサイノグラム画像は極座標変換された積算画像として表される。これは真ん中の人体部分の各ピクセルを変数とした連立方程式と考えることができる。

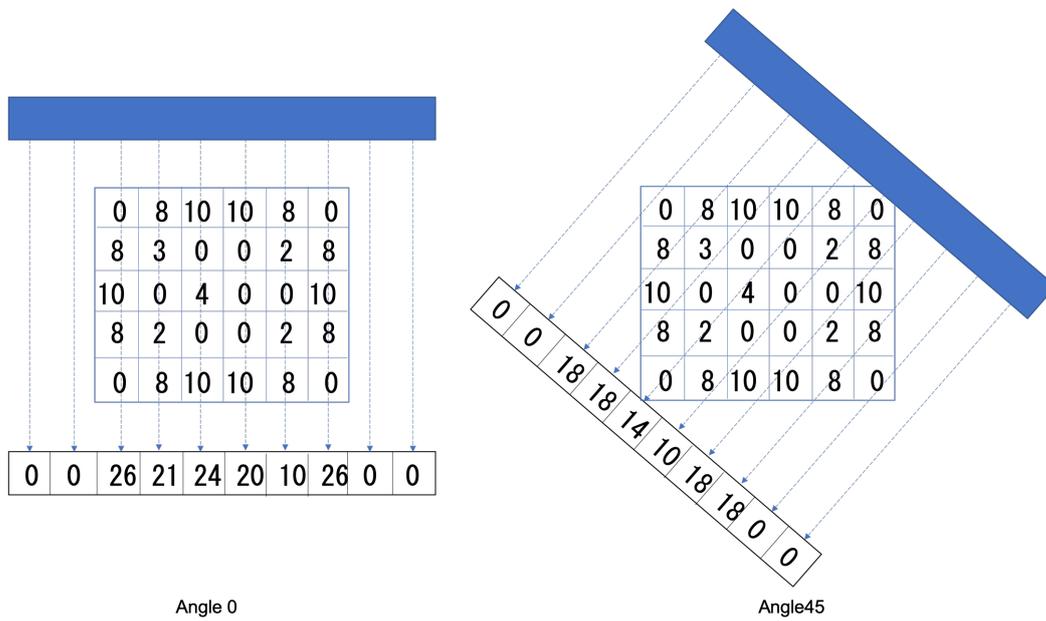


図 2.4: Conceptual diagram of sinogram image.

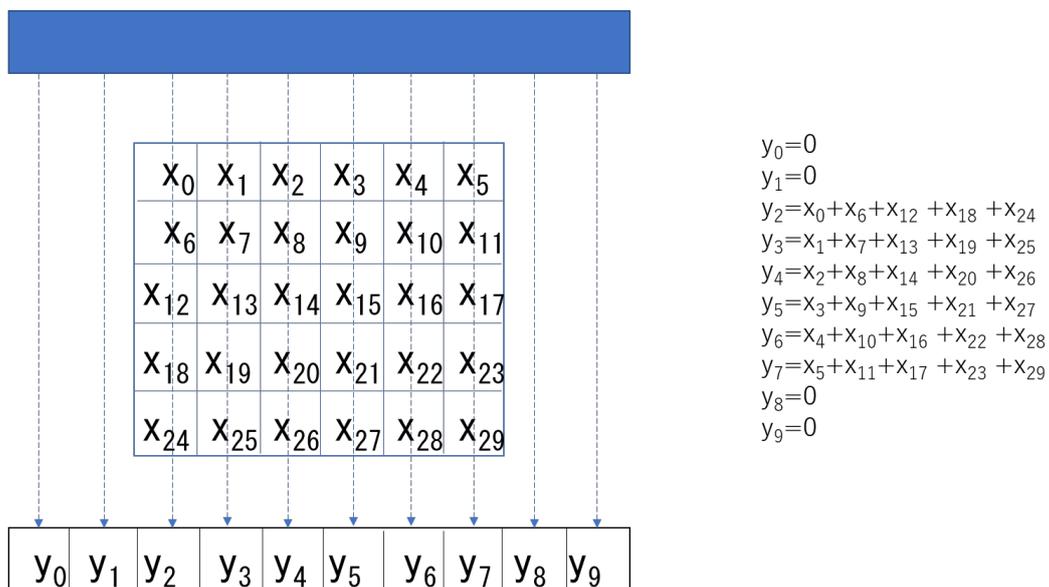


図 2.5: Algebraic image.

### 2.3.1 CT 再構成問題の代数的定式化

上記を鑑みるとデジタルな CT 画像の投影は下記のように定義できる。

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \quad (2.12)$$

ここで、 $\mathbf{y} \in \mathbb{R}^{N_y}$  は、サイノグラム画像であり、 $A = a_{ij} \in \mathbb{R}^{N_y \times N_x}$  は、投影行列と呼ばれる、物体の周りに放射された X 線の投影を表した行列。 $\mathbf{x} \in \mathbb{R}^{N_x}$  は、復元する断面画像である。また、 $\mathbf{b} \in \mathbb{R}^{N_y}$  はノイズである。

$\mathbf{x}$  の復元は  $A^{-1}$  を求めてノイズ  $\mathbf{b}$  の影響をのぞけば復元できるが、

$$\mathbf{x} = A^{-1}\mathbf{y} - \mathbf{b} \quad (2.13)$$

$A$  は特異行列であるため、陽に求めることが難しく、 $\mathbf{x}$  の復元は、下記の式の最適化問題としてあらわされる。

$$\arg \min \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda \text{norm}(A) \quad (2.14)$$

ここで  $\lambda > 0$  である。

### 2.3.2 SART 法

式 (2.13) をノイズの影響を無視し、逐次的に変形すると以下になる。

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \frac{\mathbf{y}_i - A_i \mathbf{x}_i}{A_i A_i^T} A_i^T \quad (2.15)$$

ここで  $A_i$  はシステムマトリックス  $A$  の  $i$  番目の行のベクトルであり、 $\mathbf{y}_i$  は  $\mathbf{y}$  の  $i$  番目の要素である。上記は  $\mathbf{y}_i - A_i \mathbf{x}_i$  が十分少なくなれば、収束しており、解が確からしいと言える。

上記より SART 法はアルゴリズム 1 に示すアルゴリズムとして表せる。

ここで  $a$  は定数である。典型的には想定画素範囲の中央値を設定すると収束が良いと言われている。また、 $\epsilon$  は収束条件であり、大きいと画質が悪くなり、小さいと終了しなくなるため、実際にアルゴリズムを動かし適切な値を設定する必要がある。

### 2.3.3 ML-EM 法

ML-EM 法は統計的手法であり尤度に基づいて解を求める。

ある体内でのボクセルでの吸収量が平均が  $x_j$  のポワソン分布に従うランダムな変数  $X_j$  であると仮定する。

$$P(X_j = k; x_j) = \frac{1}{k!} e^{-x_j} x_j^k \quad (2.16)$$

---

**Algorithm 1** SART method.
 

---

**Require:** system matrix  $A$ **Require:** sinogram image  $\mathbf{y}$ **Ensure:** reconstructed image  $\mathbf{x}$ 

```

 $\mathbf{x} \leftarrow a$ 
while  $y_i - A_i x_i > \epsilon$  do
  for  $i = 0$  to  $n$  do
     $x_i = x_{i-1} + \frac{y_i - A_i x_i}{A_i A_i^T} A_i^T$ 
  end for
end while

```

---

ここで  $k$  はボクセル内で発生した実際の吸収量である。複数のボクセルを通過する光線はベルヌーイ過程に従うとされている。その過程が別のポワソン分布ランダム変数  $Y_i$  を導く [9].

$$P(Y_i = y_i; \mathbf{x}) = \frac{1}{y_i!} e^{-y_i(\mathbf{x})} y_i(\mathbf{x})^{y_i} \quad (2.17)$$

ここで  $y_i$  は検出器で観測された強度を表し、 $y_i(\mathbf{x})$  はそれぞれの光線に沿ったボクセル値  $\mathbf{x}$  による累積吸収量を表す。検出器での観測値が  $y_i$  ごとに独立だとすると、あるボクセル値  $\mathbf{x}$  に対する全確率は、尤度関数  $L(\mathbf{x})$  を用いて表すことができる。

$$L(\mathbf{x}) = \prod_{y_i} P(Y_i = y_i; \mathbf{x}) \quad (2.18)$$

単調増加関数では、 $\max$  の位置は変化しなく、すべての確率が非負であるため、両辺の対数をとると対数尤度関数を求めることができる。

$$\log L(\mathbf{x}) = \log \prod_{y_i} P(Y_i = y_i; \mathbf{x}) = \sum_{y_i} (y_i \log y_i(\mathbf{x}) - (y_i(\mathbf{x}) + r_i) + \log \frac{1}{y_i!}) \quad (2.19)$$

ここで  $r_i$  は測定誤差である。

この関数を最適化することで、maximum likelihood (最尤) で、観測結果に最もふさわしいボクセル値  $\mathbf{x}$  を決定することができる。なお、 $\log \frac{1}{y_i!}$  は  $\mathbf{x}$  に依存しないため削除することができる。

上述した式はあるボクセル (画素) についての尤度推定の式であるが、画像全体について考えるときは、EM 法に基づく解法が有効である。上述した式を EM 法に基

づいて逐次化したのが，ML-EM 法であり下記式であらわされる [10].

$$x_j^k = \frac{x_j^{k-1}}{\sum_{i=0}^{N-1} C_{ij}} \sum_{i=0}^{N-1} \frac{y_i C_{ij}}{\sum_{j=0}^{M-1} C_{ij} x_j^k} \quad (2.20)$$

ここで  $k$  は繰り返し回数を表す． $j$  は画像の画素番号を表し，すべての画素数は  $M$  である． $i$  は投影データの検出器番号を示し，すべての検出器の数は  $N$  である． $x_j^k$  と  $x_j^{k-1}$  はそれぞれ  $k$  番目， $k-1$  番目の画素値である． $C_{ij}$  は検出確率である．

上記より ML-EM 法はアルゴリズム 2 に示すアルゴリズムとして表せる．

---

**Algorithm 2** ML-EM method.
 

---

**Require:** detection probability matrix  $C$

**Require:** sinogram image  $\mathbf{y}$

**Ensure:** reconstructed image  $\mathbf{x}$

```

 $\mathbf{x} \leftarrow a$ 
while  $k < b$  do
  for  $j = 0$  to  $M$  do
     $x_j^k = \frac{x_j^{k-1}}{\sum_{i=0}^{N-1} C_{ij}} \sum_{i=0}^{N-1} \frac{y_i C_{ij}}{\sum_{j=0}^{M-1} C_{ij} x_j^k}$ 
  end for
   $k = k + 1$ 
end while

```

---

ここで， $a$ ， $b$  は定数である．繰り返し替えしの最大回数である  $b$  は経験的に決定される．

### 2.3.4 OS-EM 法

ML-EM 法は再構成時間が長いため，改良を行い収束が早くなるようにした OS-EM (ordered subset expectation maximization) 法が提案されている．OS-EM 法では式 (2.20) において画像  $\mathbf{x}$  の更新部分をすべての投影を使うのではなく，サブセットに分割してサブセットごとに画像の更新を行う方法である．アルゴリズム 3 にアルゴリズムを示す．

ここで  $n$  は分割するサブセット数である．サブセットはサイノグラムデータを投影角度ごとに分割したもので，特に制約はないものの，なるべく投影角度が均等になる

---

**Algorithm 3** OS-EM method.
 

---

**Require:** detection probability matrix  $C$ 
**Require:** sinogram image  $\mathbf{y}$ 
**Require:** subset  $S = y_0, y_n, \dots, y_1, y_{1+n}, \dots, y_{n-1}, y_{2n-1}, \dots$ 
**Ensure:** reconstructed image  $\mathbf{x}$ 

```

 $\mathbf{x} \leftarrow a$ 
while  $k < b$  do
  for s in S do
    for j = 0 to M do
      for s in S do

$$x_j^k = \frac{x_j^{k-1}}{\sum_{i=0}^s C_{ij}} \sum_{i=0}^s \frac{y_i C_{ij}}{\sum_{j=0}^{M-1} C_{ij} x_j^k}$$

      end for
    end for
  end for
   $k = k + 1$ 
end while

```

---

ように選ぶのが良いとされている。OS-EM 法では  $\mathbf{x}$  の更新をすべてのサイノグラムデータを使用せずにサブセットごとに更新を行うので、解から遠い場合に解に近づくやすく、ML-EM 法に比べて収束が早くなる。サブセットの数  $n$  はあまり大きくするとノイズが増えてしまうので、適切なサブセット数に抑えることが必要である。

## 第 3 章

# Stacked U-Net による CT 再構成

前述したように，従来手法ではノイズがのる，再構成に時間がかかるという課題があった．近年，ディープラーニング技術の進展もあり，従来手法に組み合わせてディープラーニングによる再構成を実現しノイズの低減が行われている．一方，既存手法との組み合わせているという点において，従来手法よりは時間がかかっている．

本章では Stacked U-Net により既存手法との組み合わせなしにディープラーニングのモデルのみで再構成ができることを示す．

### 3.1 はじめに

近年，ディープラーニング技術の進歩に伴い，畳み込みニューラルネットワーク (CNN) [11, 12] を用いたノイズ除去 [13]，超解像 [14]，セグメンテーション [15] などの研究が盛んに行われている．

再構成においてもディープラーニング技術の適用が進んでおり，Yang ら [16] らは magnetic resonance imaging (MRI) の再構成において，逐次近似法の各ステップをディープラーニングに基づいた alternating direction method of multipliers [17] に置き換えることで，既存手法より優れた結果を示した．

また，Yang ら [18] は generic adversarial network (GAN) [19] を MRI の再構成に適用し，16,095 枚の画像で訓練を行い，5ms という短い時間で再構成ができることを示した．

また, Zhu ら [20] は病変部を抽出するディープラーニングのモデルと GAN を組み合わせて MRI の再構成に適用し, ROI (Region Of Interest) 抽出した領域に対して超解像で画質を改善するモデルを提案した. ROI を用いることで, 画像のサイズが小さくなるため, 超解像を行う GAN モデルの学習の収束や生成が早くなる. GAN の評価関数としては VGG loss [21] を使用し, 既存手法を上回る結果を得た [22–24]. 結果学習済みモデルからの高解像度画像の出力は 1 枚あたり 0.22~0.33ms と高速だった. また彼らは, 医療関係者の主観的な評価基準である mean opinion score を提案した.

CT 画像の再構成としては, Jin ら [25] は FBP 法と U-Net [15] と residual learning [26, 27] を組み合わせた物を提案している. また, Zhang ら [28] は DenseNet [29] を FBP 法と組み合わせたモデルを提案している. これらの手法は FBP 法で発生するノイズやアーティファクトの量を減らすことができ, 既存手法より良い結果を得ることができる.

U-Net [15] は, 画像の入力と出力が同じサイズの CNN モデルであり, 高解像度の特徴量をスキップ接続によって後段に渡す所に特徴がある. U-Net は, 高解像度画像に対して畳み込み演算を繰り返し, 畳み込み画像を縮小することで, 低解像度特徴を抽出する. 抽出された低解像度の特徴は反復的に拡大および統合され, 最終的に高解像度の画像が生成される.

U-Net それ自体も改良されており, [30] らは Resblock を適用した U-Net を 15 層重ねた Stacked U-Net においてセグメンテーションタスクにおいて U-Net より優れた結果を示した. Shah ら [31] は Dense block を適用した Stacked U-Net をクラス分類タスクに適用し, モデルに転移学習を適用し, セグメンテーションタスクにおいて既存の手法より優れた結果を示した.

本章では, 少ない投影数でも, 既存の逐次近似法より再構成速度が早く安定した画質を得ることを目的とした. 提案手法ではディープラーニングの技術を使用し Stacked U-Net を構造として採用した.

一般的に, 患者のプライバシー保護のため, 学習データセットとして医療画像を収集することは容易ではない. CT の結果画像は患者の断面の画像であるため, 取得することは不可能である. 可能な方法は, 既存手法により取得された断面画像を利用することである.

FBP U-Net では学習データとして、既存手法により取得された断面画像を radon 変換によりサイノグラム画像に変換し、サイノグラム画像から FBP 法でさらに変換して得た断面画像を学習に使用している。しかし、この方法では量子化誤差に起因するノイズが学習データに入りこむ、そのため、FBP U-Net の学習では実際とは異なる断面画像を生成していると考えられる。

我々は、学習には自然画像つまり医療機器で取得されていない画像を使用することにした。具体的には ImageNet [12] を使用した。この中から、28,463 枚の画像を学習データセットとして使用し、300 イテレーションを 90 時間かけて学習した。

結果として、512 個の検出器の 360°回転 64 投影を想定したデータセットで、Peak signal-to-noise ratio (PSNR) で 27.93db, structural similarity (SSIM) [32] で 0.886 の断面画像を、CPU で 2.35s, GPU で 0.11s で再構成することができた。

## 3.2 関連研究

### 3.2.1 MRI 再構成へのディープラーニングの適用

MRI は磁場の変化により、体内の水素原子が発する電磁波を観測することにより、体内の情報を得る手法である。そのため、外部から X 線を照射する CT とは再構成の方法が違う。しかし、使用するのはフーリエ変換や逐次近似法であるため、MRI 再構成の知見は CT 再構成にも応用可能なことが多い。

Yang ら [16] らは magnetic resonance imaging (MRI) の再構成において、逐次近似法の各ステップをディープラーニングに基づいた alternating direction method of multipliers [17] に置き換えることで、既存手法より優れた結果を示した。

また、Yang ら [18] は generic adversarial network (GAN) [19] を MRI の再構成に適用し、16,095 枚の画像で訓練を行い、5ms という短い時間で再構成ができることを示した。

また、Zhu ら [20] は病変部を抽出するディープラーニングのモデルと GAN を組み合わせて MRI の再構成に適用し、ROI 抽出した領域に対して超解像で画質を改善するモデルを提案した。ROI を用いることで、画像のサイズが小さくなるため、超解像を行う GAN モデルの学習の収束や生成が早くなる。

### 3.2.2 CT 再構成へのディープラーニングの適用

CT の再構成へのディープラーニングも行われているが、既存手法との組み合わせで実現されている先行研究が多い。

Jin ら [25] は FBP 法と U-Net [15] と residual learning [26, 27] を組み合わせた FBP U-Net を提案している。また、Zhang ら [28] は DenseNet [29] を FBP 法と組み合わせた DD-Net を提案している。

### 3.2.3 U-Net とその改良である Stacked U-Net

U-Net [15] は、画像の入力と出力が同じサイズの CNN モデルであり、高解像度の特徴量をスキップ接続によって後段に渡す所に特徴がある。U-Net は、高解像度画像に対して畳み込み演算を繰り返し、畳み込み画像を縮小することで、低解像度特徴を抽出する。抽出された低解像度の特徴は反復的に拡大および統合され、最終的に高解像度の画像が生成される。

U-Net それ自体も改良されており、[30] らは Resblock を適用した U-Net を 15 層重ねた Stacked U-Net においてセグメンテーションタスクにおいて U-Net より優れた結果を示した。Shah ら [31] は Dense block を適用した Stacked U-Net をクラス分類タスクに適用し、モデルに転移学習を適用し、セグメンテーションタスクにおいて既存の手法より優れた結果を示した。

### 3.2.4 医用におけるトレーニングデータセット

医用画像は学習用データの数が患者のプライバシーなどの理由から自然画のデータに比べて少ないのが課題である。

Bhatt ら [1] らは、医療におけるディープラーニングの研究のほとんどが自ら収集したデータで実施されていると指摘している。それは研究所や病院でデータセットを集めるのは時間がかかるため、オープンアクセスなデータセットの使用を推奨している。ただ現状オープンなデータセットで最大のものでも、数万枚であり CT スキャンのデータに限ると 1000 枚程度が限界である。

Tajbakhsh ら [33] は特にセグメンテーションタスクにおいて、アノテーションが

図 3.1: List of open access datasets for medical study and research [1].

Organization/dataset	Description	Source link
EPILEPSIAE database	Data sets from more than 200 patients with epilepsy	<a href="https://www.epilepsy.uni-freiburg.de/database">https://www.epilepsy.uni-freiburg.de/database</a>
Cancer data set	7909 breast microscopic images of 82 patients	<a href="https://www.web.inf.ufpr.br/vri/breast-cancer-database/">https://www.web.inf.ufpr.br/vri/breast-cancer-database/</a>
Image CLEF 2017	CT images and biomedical images	<a href="https://www.imageclef.org/2017">https://www.imageclef.org/2017</a>
ISIC2017	Skin disease dataset contain more than 24k images	<a href="https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery">https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery</a>
Mini-MIAS database	3369 mammograms of 967 patients. It is classified on the basis of lesion type, morphology, breast tissue and pathology type	<a href="https://www.mammoimage.org/databases/">https://www.mammoimage.org/databases/</a> <a href="https://www.peipa.essex.ac.uk/info/mias.html">https://www.peipa.essex.ac.uk/info/mias.html</a>
DDSM-BCRP	2620 scanned film contain normal, benign and malignant cases with verified pathology information	<a href="https://www.eng.usf.edu/cvprg/Mammography/DDSM/BCRP/bcrp_mass_01.html">https://www.eng.usf.edu/cvprg/Mammography/DDSM/BCRP/bcrp_mass_01.html</a>
INbreast	115 cases with 410 images, 90 cases are from women with both breast and 25 cases from mastectomy patients	<a href="https://www.medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database">https://www.medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database</a>
ADNI	ADNI dataset contain data of Alzheimer's disease patient	<a href="https://www.adni.loni.usc.edu/data-samples/access-data/">https://www.adni.loni.usc.edu/data-samples/access-data/</a>
LIDC	Lung image database consortium. CT scan images of lung for detection of lung cancer	<a href="https://www.wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI#">https://www.wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI#</a>
DRIVE	Digital Retinal images for extraction. It contains 40 images of retina	<a href="https://www.isi.uu.nl/Research/Databases/DRIVE/download.php">https://www.isi.uu.nl/Research/Databases/DRIVE/download.php</a>
ALL-IDB	It is a new and pubic dataset of microscopic images of blood samples	<a href="https://www.homes.di.unimi.it/scotti/all/">https://www.homes.di.unimi.it/scotti/all/</a>
LUNA	888 CT images are in this dataset	<a href="https://www.luna16grand-challenge.org/data/">https://www.luna16grand-challenge.org/data/</a>
LIDC	Lung images database Consortium contains lung screening thoracic CT scan	<a href="https://www.wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI">https://www.wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI</a>
NLST	It has SCT screening images and lung cancer progression dataset	<a href="https://www.biometry.nci.nih.gov/cdas/learn/nlst/images/">https://www.biometry.nci.nih.gov/cdas/learn/nlst/images/</a>
TCGA	The cancer genome atlas, it has genomic, transcriptomic, epigenomic and proteomic data	<a href="https://www.portal.gdc.cancer.gov/">https://www.portal.gdc.cancer.gov/</a>

問題となることを指摘している。アノテーションとは画像へのラベル付けであるが、セグメンテーションタスクでは疾病領域をマーキングすることである。一枚一枚画像にアノテーションするのは専門家である医師でないと難しく、とても時間がかかる作業となる。

少ない学習データを増やすためには augmentation 手法を適用することが考えられる。augmentation においては自然画と同じ手法がつかえる。Christ ら [34] は CT 画像のセグメンテーションタスクに対してガウスぼかしを採用している。

しかし、augmentation はもともとなるデータセットを増強する方法であり、増強したデータの数が元のデータの何倍にもなると、元のデータの分布から外れていくために自然画の学習ほど多くのデータを作ることはできない。

### 3.3 Stacked U-Net の構造

図 3.2 と図 3.3 は今回提案した stacked U-Net の構造である。基本構造は U-Net と同じだが、これを図 3.3 に示すように U-Net 自体を 6 層重ねている。

もともとの U-Net は画像のセマンティックセグメンテーションのために作成されており、再構成とはタスクが違うが、サイノグラムから断面画像の変換はサイノグラ

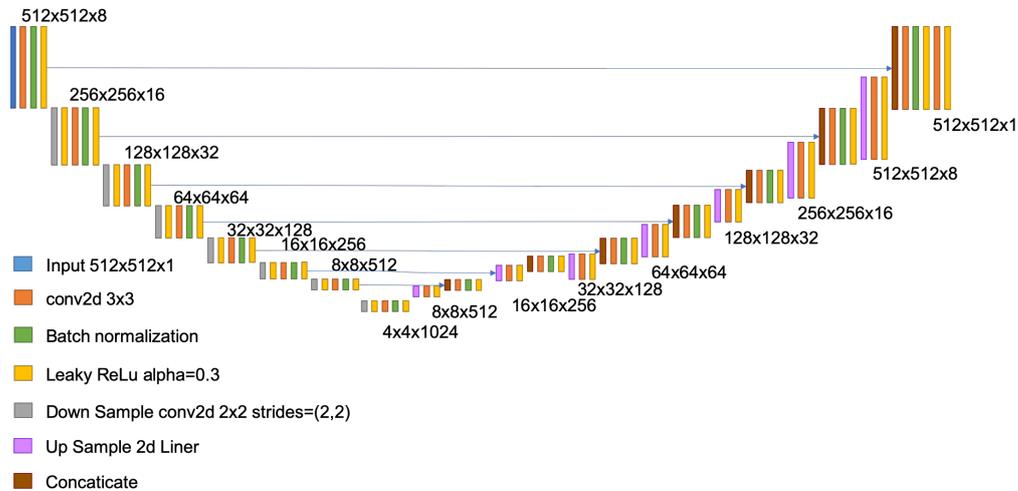


図 3.2: Structure of the proposed U-Net.

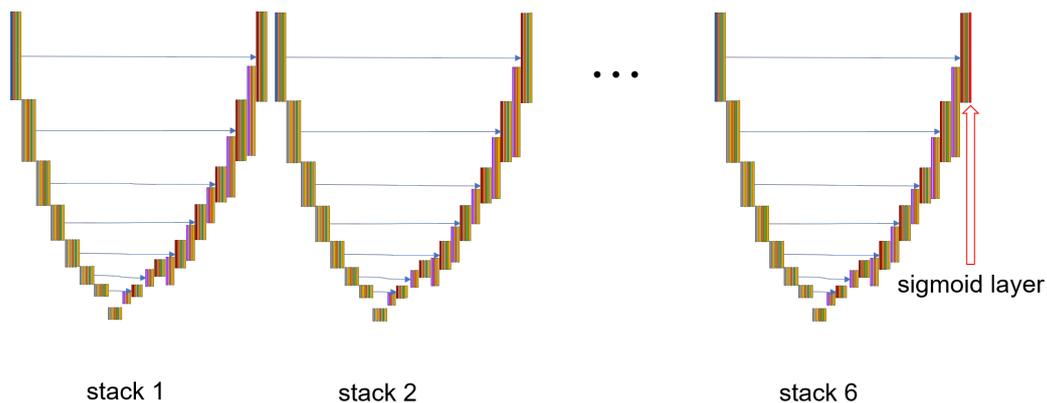


図 3.3: Stacked U-Net

ムの検出器方向の位置情報が必要であり、また、stack しているため、stack の後半の U-Net で位置を意識した超解像やノイズ除去が行われることを期待し U-Net を採用した。

stacked U-Net は [30] や [31] が提案しているが、通常の U-Net と同様セマンティックセグメンテーションを対象のタスクとしている。クラス分類の目的のためモデルの最終出力は n-class となっている。

本論では、最終出力はグレイスケールの画像となるように 1 チャンネルとなっている。途中の出力結果は活性化関数である LeakyReLU を使用し、最終出力は 0 – 1 に正規化された画像となるようにシグモイド関数を使用している。

Shah ら [31] は U-Net はそもそもパラメタ数を多く必要としており、それを多数重ねた stacked U-Net はさらに多くのパラメタ必要とするため、メモリサイズが大きくなり、現実的な GPU のメモリサイズで訓練するのは難しいと述べている。そのため、彼らは  $3 \times 3$  の convolution をもともとの U-Net の 2 つから 1 つに減らし、さらに縮小を  $3 \times 3$  のストライド 3 の convolution によって行い、縮小レイヤの数も 3 に減らしている。本論では、我々もメモリサイズを減らすため、convolution を 2 から 1 に減らしている。

また Shah ら [31] は前処理により U-Net に入力するメモリのサイズを減らしている。そのため、縮小レイヤの数は 3 で十分である。しかし、我々は 5 レイヤにしている。これは縮小レイヤ数を多くし、詳細な特徴を十分に抽出するためである。

また、精度向上のため、レイヤごとの関数も見直しをおこなった。縮小処理は max pooling から  $2 \times 2$  のストライド 2 の convolution に、拡大処理は deconvolution から線形補間に、活性化関数は rectified linear unit (ReLU) [35] から LeakyReLU [36] にした。また batch normalization も適用した。

## 3.4 予備実験

### 3.4.1 $32 \times 32$ 画像の復元でのレイヤ関数の選択

$512 \times 512$  の復元の学習は時間がかかるため、 $32 \times 32$  の画像の復元で、U-Net の詳細な構造の検討をおこなった。それぞれのレイヤ関数について 50epoch 学習させ評価をおこなった。評価は評価関数の値を使用した。今回評価関数としたのは、Mean Squared Error (MSE) で復元画像とモデルの出力結果を比較している。表 3.1 は評価の結果を示している。縮小レイヤ (down scaling) での評価結果は max pooling で 0.0107, 線形縮小 (linear scaling) で 0.0099,  $2 \times 2$  の stride 2 の convolution で 0.0086 となった。そのため  $2 \times 2$  convolution を採用した。

拡大レイヤ (up scaling) では縮小レイヤで convolution を採用しているため、対称性を考えると deconvolution [37] が適していると考えられる。しかし、評価値は、線形拡大 (linear scaling) で 0.0086, deconvolution で 0.0107 となった。そのため、線形拡大を採用した。

汎化方法としては dropout と batch normalization [38] を検討したが、評価

表 3.1: Applied methods

class	methods	adoption	validation loss
down scaling	$2 \times 2$ stride 2 convolution	✓	0.0086
	max pooling		0.0107
	linear scaling		0.0099
up scaling	$2 \times 2$ stride 2 deconvolution		0.0107
	linear scaling	✓	0.0086
generarization	batch normalization	✓	0.0083
	drop out		0.0141
activate function	sigmoid		0.0567
	ReLU		0.0132
	LeakyReLU	✓	0.0086
skip-connection	previous U-Net input image to the next U-Net input image		-
	first input image to all the U-Net input		-
Optimizer	Adam	✓	0.0071
	AdaBound		0.0142

値は dropout では 0.0141 と悪くまた出力結果も悪化したため、採用しなかった。batch nomalization は効果があり評価値も 0.0083 と改善した。そのため、汎化方法としては batch normalization を採用した。

活性化関数はシグモイド関数で 0.0567, ReLU で 0.0132, LeakyReLU で 0.086 だった。シグモイド関数はスタック数が少ない場合は評価結果がよかったが、スタック数が多くなると、収束しなかった。ReLU と LeakyReRU は収束したが、LeakyReRU のほうが結果がよかったため、活性化関数としては LeakyReRU を採用した。

U-Net の入力と出力は画像である。U-Net 内部の skip connection は情報損失をなくすといわれており、stack 数の増加による情報損失に対応するため、図 3.4, 3.5 に赤線で示す二種類の大域な skip connection を試した、一つは前の U-Net の入力を U-Net の入力に入れるもの (図 3.4), もう一つは入力をすべての U-Net の入力に入れるもの (図 3.5) である。それぞれ、収束しなくなったため今回は採用しなかった。

最適化アルゴリズムとしては Adam [39] と Adabound [40] を試して、それぞれ

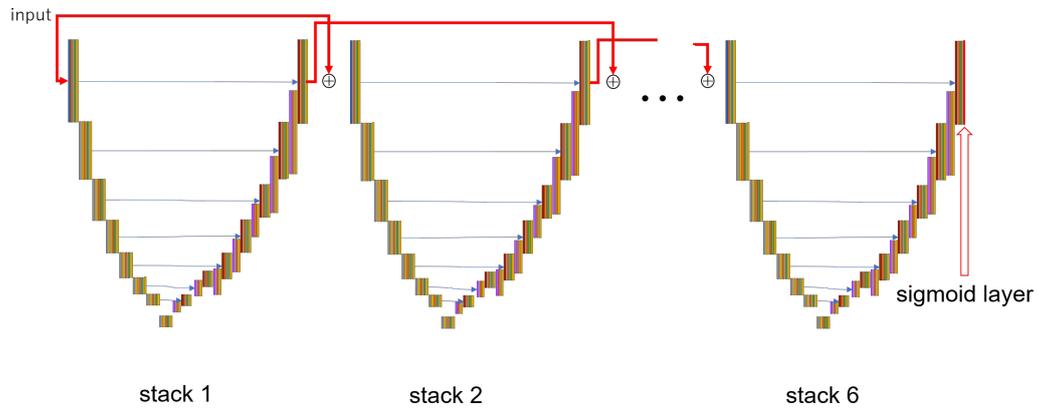


図 3.4: Connect previous U-Net input to next U-Net input.

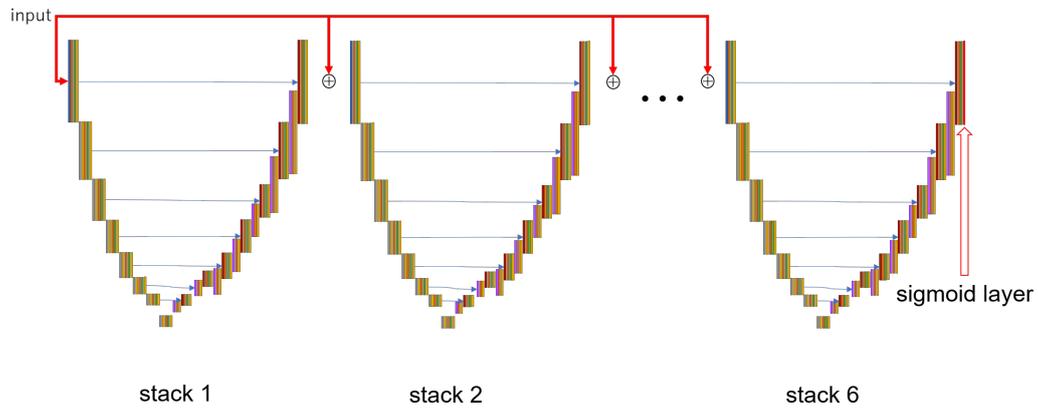


図 3.5: Connect stacked U-Net input to all U-Net input.

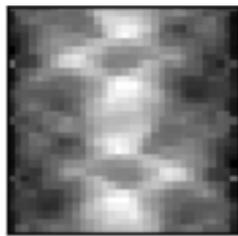
の評価値は 0.0071, 0.0142 だった. そのため, Adam を採用した.

最後にスタック数を調整した. 我々はスタック数を 1 から 8 まで変更して 50epoch 学習を行わせた. 結果を図 3.2 に示す. 4 スタックまでは評価値はよくなっていったが, 5 スタック以上では 30epoch 以上になると発散の挙動を示し, 訓練損失 (training loss) と評価損失 (validation loss), の差が大きくなり過学習の挙動を示したので学習を打ち切った. 結果として 4 スタックを採用した.

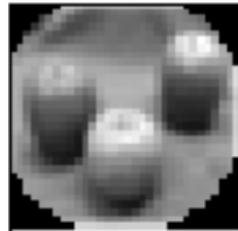
学習用のデータは CIFAR-100 [41] を使用した. CIFAR-100 は  $32 \times 32$  の自然画の RGB のデータセットであるが, これを白黒画像に変換し, CT の条件と同じように円で切り取り, その画像をもとにサイノグラム画像を生成した. 生成したサイノグラム画像を入力, 生成元の画像を出力とし 100epoch 学習させた. 評価値とし

表 3.2: Relationship between the number of stacks and evaluation values for  $32 \times 32$ .

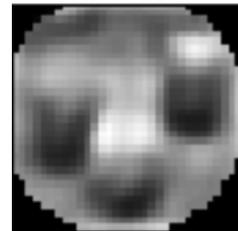
num stacks	training loss	validation loss
1	0.0070	0.0073
2	0.0070	0.0072
3	0.0051	0.0053
4	0.0053	0.0054
5	0.0053	0.0301*
6	0.0050	0.0173*
7	0.0052	0.0232*
8	0.0054	0.0479*



(a) sinogram



(b) ground truth



(c) reconstructed image

図 3.6: Comparison of  $32 \times 32$  images.

ては Mean squared error (MSE) を使用した。図 3.6a-3.6c に結果を示す。Peak signal-to-noise ratio (PSNR) と structural similarity index measure (SSIM) はそれぞれ 26.83db と 0.951 であった。

### 3.4.2 $32 \times 32$ 画像での復元の考察

図 3.6c ではリンゴの輪郭と影が復元できているのがわかる。しかしながら、リンゴのへたは復元できていない。PSNR も低く復元が不十分である。

表 3.2 に示すように、スタック数を大きくするとスタック数 4 までは評価値は徐々に改善していく。しかし、スタック数 5 以上になると悪化する。

これは、スタック数を大きくすれば大きくするほど後段の U-Net に伝わる情報の欠落が欠落していつているためと考えられる。そのため、図 3.4, 3.5 に示すような、

大域的な skip-connection も試したが、結果は悪化した。この部分はさらに改良の余地があると考え。

この程度の復元では、あまり実用に耐えうる復元ではない。しかし、低解像度での再構成では元の画像を復元するための情報量も少なくなるため、詳細な構造の復元に限界があると考え  $512 \times 512$  の学習に移行した。

## 3.5 実験結果

### 3.5.1 $512 \times 512$ の画像の復元

$32 \times 32$  での復元の検討をもとに  $512 \times 512$  での学習を行った。

$512 \times 512$  の学習ではスタック数 3 から 7 で 30 イテレーションを予備的に学習させ、収束度合いが一番急だったスタック数 6 を採用し、学習を行った。

また、今回投影数が少ない画像での比較実験を行ったが、今回作成したニューラルネットは  $512 \times 512$  の入力しか受け付けられないため、投影数が少ない場合入力となるサイノグラム画像は  $64 \times 512$  (eg. 投影数 64) などの画像となり、入力できないが、BICUBIC 処理を行い  $512 \times 512$  に拡大し入力した。

訓練データは構造決定の時とはちがいが、ある程度高解像度の画像が必要であるため、自然画のデータセットである Image Net [12] から 28463 枚をトレーニングセットに 1537 枚をバリデーションセットに使用した。テストセットとしては The Cancer Imaging Dataset [42] [43] を使用し、467 枚の画像を評価した。

ImageNet の画像の解像度はばらばらであるため、グレースケール化の後、アスペクト比を変えないように正方形に中心を切り抜き、 $512 \times 512$  に線形補間で scale し、円形に cropping し出力画像とし (図 3.7b)、出力画像を元にサイノグラム画像 (図 3.7a) を生成し入力画像とした。

学習環境としては CPU は Intel Xeon E5 2680 v4 2.4GHz 14 コアを 2CPU、Memory は 256GByte、GPU は NVIDIA TESLA P100 for NVlink-Optimized Servers を 4GPU を使用した。学習のためのイテレーション回数は 300 回であり、学習時間は 90 時間だった。

図 3.7b-図 3.7c にバリデーションセットで復元した投影数 512 の自然画の例を示す。図 3.7a がモデルに入力されるサイノグラム画像で、モデルは図 3.7c を出力す

る。この時の図 3.7b と図 3.7c の PSNR は 32.24db, SSIM は 0.963 であった。

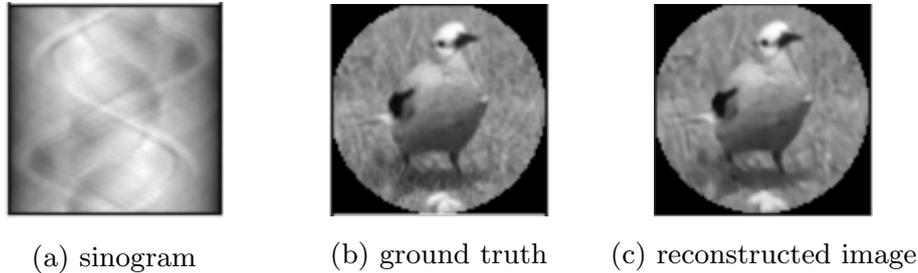


図 3.7: Results for validation sets

### 3.5.2 医用画像での結果

学習は自然画像で行っており、本来の目的の医用画像の復元ができるかどうかはわからない。そのため、学習に使用していない医用画像をテストセットとし、復元を行った。学習には一切医用画像を使用していないが、復元ができた。

表 3.3 にテストセットである医用画像を提案モデルで復元した場合と既存手法との比較を示す。今回比較したのは FPB, SART, TV-norm, FBP U-Net, DD-Net である。

逐次近似法には MLEM 法に TV-norm を使用しノイズを低減した TV-EM 法 [44] がある。本実験でも比較対象にするか予備実験を行い、iteration 数および反映係数  $\alpha$  の調整を行なったが、投影数 64 の場合で、iteration45,  $\alpha$  0.001 で PSNR24.75db, SSIM 0.80, 実行時間 243 秒, 投影数 128 の場合で、iteration45,  $\alpha$  0.001 で PSNR25.59db, SSIM 0.84, 実行時間 421 秒と似た特徴を持つ SART 法と比べても劣った結果となっており、実行時間も大きいため比較は行わなかった。

FBP および SART は skimage [45] の iradon 関数, iradon\_sart 関数を使用した。TV-norm は proxTV [46] の tv1\_2d 関数を使用した。FBP U-Net は [25] に示されているように投影数 512 の場合 (FBP U-Net512) と投影数 64 の場合 (FBP U-Net64) の場合それぞれに対して 475 枚の医用画像を 101epoch 学習させた。DD-Net は [47] の学習済みモデルを使用した。

計測環境は Intel Core i3 4130T CPU 2.90GHz, Memory 24GByte を使用し、各々の PSNR および SSIM および処理時間を計測した。

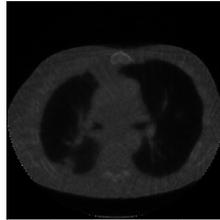


図 3.8: TV-EM projection 64 result.

ここで FBP, SART, TV-norm は CPU 上で処理を行った結果の処理速度を計測した. FBP U-Net と DD-Net および提案手法は BICUBIC による拡大処理を含めた処理時間を計測し, 特に GPU:GeForce GTX 1080 を使用して計算を行った時の処理速度も計測した.

提案手法の処理時間はすべての投影数において CPU で 2.35s, GPU で 0.11s と一定である. 投影数 64 の再構成の CPU での処理速度を比べると FBP は 0.21s, SART は 2.07s, TV-norm は 0.21s, FBP U-Net は 6.54s, DD-Net は 4.99s となっている. 投影数 64 の再構成の GPU での処理速度は FBP U-Net で 1.02s, DD-Net で 1.16s となっている.

また投影数 64 において平均の PSNR は, 提案手法は 27.93db, FBP は 21.41db, SART は 25.03db, TV-norm は 25.73db, FBP U-Net512 は 19.79db, FBP U-Net64 は 22.27db, DD-Net は 22.75db となった. 平均の SSIM は, 提案手法は 0.886, FBP は 0.547, SART は 0.829, TV-norm は 0.750, FBP U-Net512 は 0.581, FBP U-Net64 は 0.666, DD-Net は 0.717 となった. 既存手法は投影数 64 の場合従来手法より優れている.

図 3.9a-図 3.9h に投影数 64 の時の再構成結果を示す. 提案手法には中心から放射上に広がるアーティファクトがないことが見て取れる.

断面画像の右上を拡大したものを図 3.10a-図 3.10h に示す.

拡大してみると, 原画像にある微細な血管はどの手法においても消えてしまっていることが分かる.

提案手法と FBP U-Net64 以外ではアーティファクトが顕著である. 提案手法は大きな構造物のエッジは良く出ており, また中心に位置する太い血管もはっきり出ている.

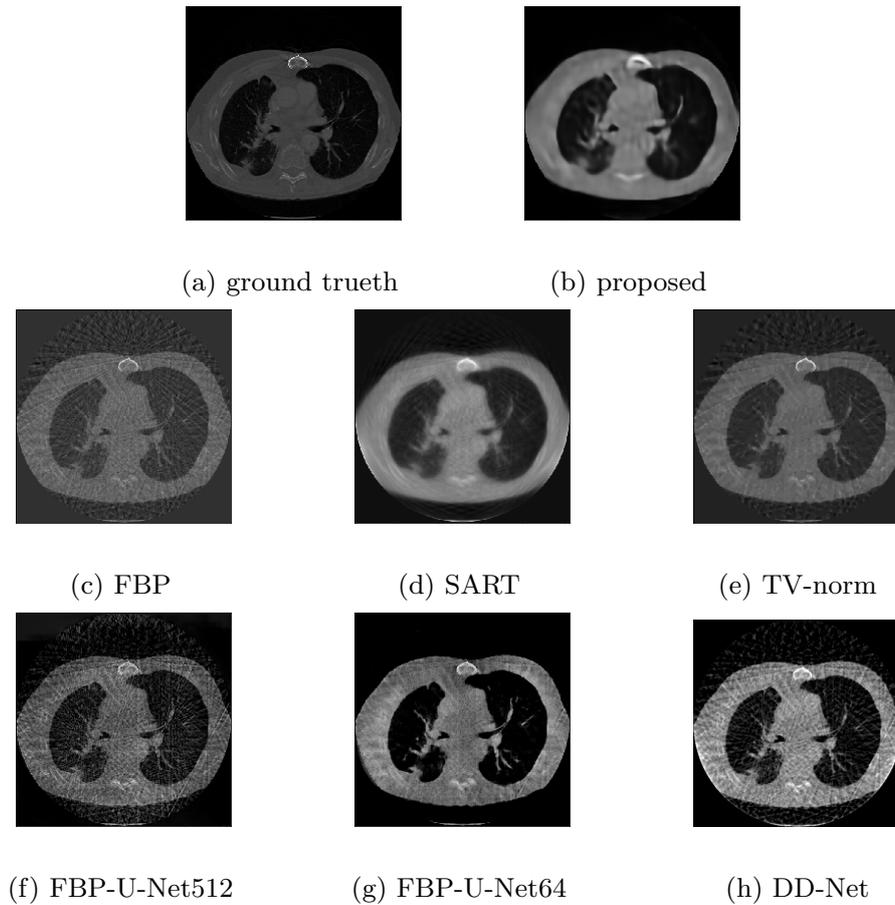


図 3.9: Reconstructed image with 64 projections.

FBP U-Net64 では細い血管がもっともよく出ている。一方、中央左下付近の大きな血管の構造は提案手法では先端の詳細構造が失われており、FBP 法や TV-norm では詳細が出ている。

表 3.4 に詳細部分の PSNR と SSIM を示す。提案手法の PSNR は 30.37db, SSIM は 0.896, FBP 法の PSNR は 25.17db, SSIM は 0.570 SART 法の PSNR は 29.34db, SSIM は 0.863 TV-norm の PSNR は 29.61db, SSIM は 0.816 FBP U-Net512 は 18.75db, SSIM は 0.494 FBP U-Net64 は 22.92db, SSIM は 0.447 DD-Net は 18.03db, SSIM は 0.510 となっており、詳細部分においては提案手法がすべての手法より優れた結果となっている。

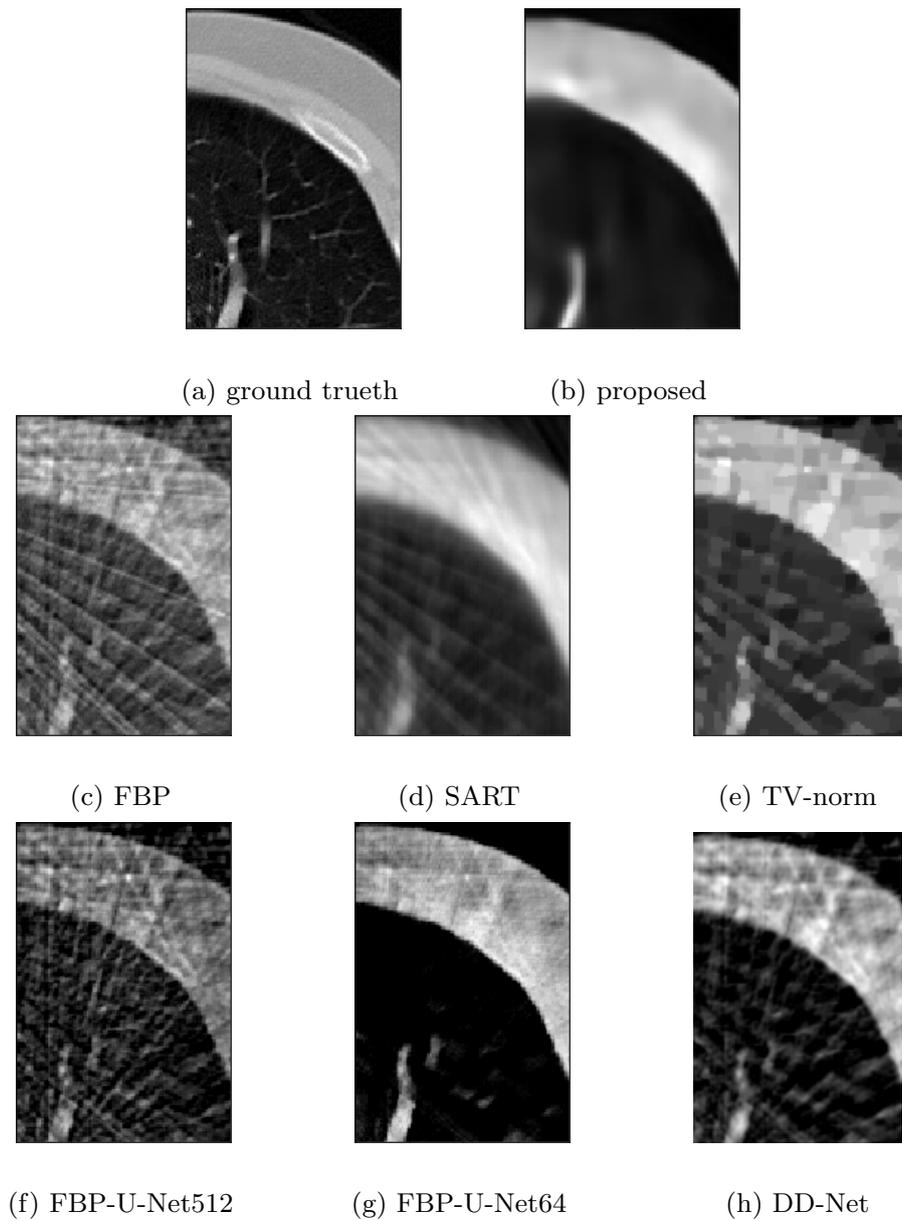


図 3.10: Comparison of details with 64 projections.

### 3.5.3 FBP U-Net を ImageNet で学習させたときとの比較

FBP U-Net は学習データが少ない医用画像で学習させているため、データ量という面で学習条件が公平でない提案手法とおなじように ImageNet で学習させて比較を行った。学習させたのは投影数 64 の時のデータで、図 3.11 に結果を示す。平均の PSNR は 18.99db で SSIM は 0.580 であった。これは医用画像で学習させた

場合より低い。提案手法は図 3.9b に示すようにアーティファクトは発生していないが、FBP U-Net の場合図 3.11 に示すように、画像全体がぼんやりとし、アーティファクトでおおわれている。

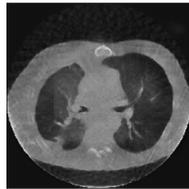


図 3.11: Results of medical images of FBP U-Net trained by ImageNet at projection number 64.

#### 3.5.4 セグメンテーションタスクを用いた再構成結果画像の客観評価

上記結果は、実際に画像を用いた診断に使えるかが重要であるが、医療関係者に大量の画像を評価してもらうのは困難である。そこで、我々はセグメンテーションタスクを学習したモデルに再構成した画像を入力した。その結果を比較することで、生成された画像が医療目的に使用できるかの評価をおこなった [48]。

セグメンテーション用のモデルとしては UNet++ [49] を使用し、MedSeg Covid Dataset 1 [50] の  $512 \times 512$  の画像を 80 枚学習用として使用し、20 枚を検証用として使用し、250epoch の学習を行った。学習したモデルの検証セットにおける平均 Dice スコアは 0.555 だった。

表 3.5 にセグメンテーションタスクにおける平均 Dice スコアを示す。

提案手法は投影数 64 とき既存手法と比べもっともよい結果となっており、ディープラーニングを用いた手法である FBP U-Net や DD-Net に対してすべての投影数で良い結果になっている。また、TV-norm は投影数 512 から 128 ですべての手法の中でもっともよい結果となっている。対して SART 法はすべての手法の中で一番悪い結果となっている。

## 3.6 考察

図 3.7c に示す検証セットでの再構成結果では、鳥の画像の背景の草も正確に復元されており、再構成がうまくできていることを示している。

しかし、テストセットの医用画像では詳細な構造が失われている。これは医用画像でなく自然画をもちいて学習を行ったせいだと考えられる。そのため、学習セットに医用画像である断面画像を加えることで、提案モデルの性能を向上させる余地がある。また、医用画像に似た特徴を持つ工業製品の画像を加えても精度向上させる余地があると考えられる。

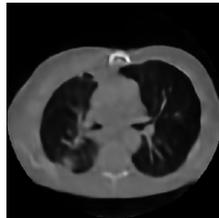


図 3.12: Results of reconstructing medical images with 512 projections.

また、提案手法による再構成結果と FBP U-Net による再構成結果を比較した。

図 3.9b に示すように、投影数 64 において提案手法は 512 で学習したにも関わらず再構成ができており、ノイズとアーティファクトが軽減されている。図 3.9f に示すように、FBP U-Net512 と比べると、FBP U-Net512 ではノイズは除去されていない。図 3.9g に示すように、FBP U-Net64 では FBP U-Net512 に比べるとノイズが除去されている。

この違いは FBP U-Net512 が 512 投影の画像を FBP した結果を用いて学習を行っていることで現れていると考えられる。これは、FBP U-Net は FBP 法が発生させる投影数に応じたノイズを除去する方法を学んでいると言い換えることができる。つまり、FBP U-Net512 は投影数 64 のような学習していないノイズを除去することはできない。また FBP U-Net は再構成の方法自体は FBP 法に依存しており、単にノイズを除去しているだけである。

提案手法は投影数 512 で学習したにも関わらず、投影数 64 の座像でもノイズが生じておらず、再構成にも成功している。これは提案手法が再構成の手法自体を学習し

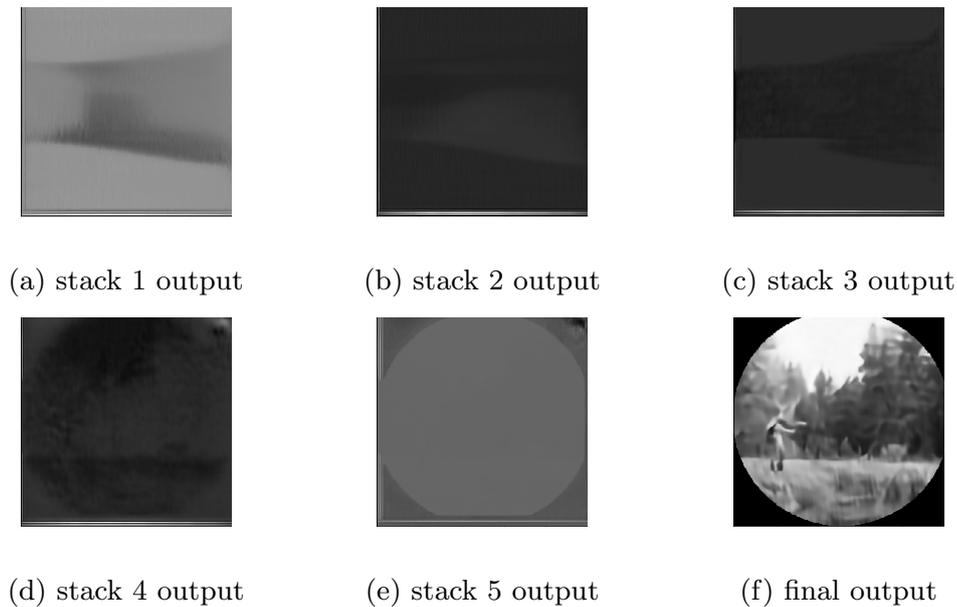


図 3.13: Intermediate output corresponding to the validation set with the number of projections equal to 512.

たためであると考えられる。そのため、提案手法は投影数の違いに対して、高い汎化性能を示している。

これに対して、FBP U-Net は投影数ごとに再学習をする必要がある。提案手法は投影数が変わっても同じモデルを使用することができる。

詳細な構造を見ると (図 3.10b および 3.10g), 画像上部の脂肪を表す領域ではアーティファクトやノイズのない滑らかな画像が得られている。一方 FBP U-Net64 では線状のアーティファクトが除去できず残っていることが確認できる。

画像の左下にある大きな血管の部分では、提案手法と FBP U-Net64 の両方で血管がはっきり見えていることがわかる。大きな血管にかんしては、FBP U-Net64 は提案手法に比べて詳細な構造が再構成できていることがわかる。

画像の右下にある細い血管を比較すると、提案手法では血管が消えているのに対し FBP U-Net では約 4 つの大きなノイズが見えていることがわかる。上記の違いは、提案手法と FBP U-Net の PSNR および SSIM の値の違いとして表れている。

図 3.13a から図 3.13f は投影数を 512 にした時の提案手法の、検証セットでの各 U-Net の中間出力を示している。

図 3.13a から図 3.13c には、水平方向に円錐状の帯が見て取れる。図 3.13d か

ら円の形状が現れ始め, 図 3.13e で円の形状が鮮明になる. 図 3.13f で出力画像となる.

SART による再構成では, 図 3.14 に示すように断面の形状が徐々に現れていく.

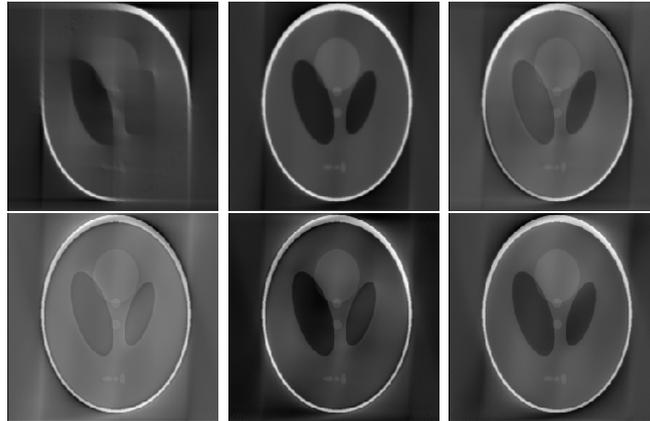


図 3.14: Reconstruction of a phantom by the SART method.

上述した違いから見て取れるように, 提案手法では中間画像の再構成方法が, 逐次近似法と異なっており, 従来とは異なる再構成の方法を獲得していることがわかる.

実行時間の観点では, CPU を用いた場合, 提案手法は 512 から 128 までの投影数において SART より優れている. また投影数 64 の場合でも提案手法が 2.35s, SART 法が 2.07s となった. GPU での処理の場合提案手法は, 既存手法より全ての場合で上回っている.

GPU のハードウェアアーキテクチャは一般的に CPU とは異なるため, 適切な処理を実行するためにはアルゴリズムや実装を適宜変更する必要がある [51]. ディープラーニングは線形計算の組み合わせであるため, 提案手法は実装やアルゴリズムを CPU の時から変更することなく, GPU でモデルを実行することができる.

画質の面では, 提案手法は中心から放射状に広がるアーティファクトや破綻がなく, 投影数が 64 の時に高い汎化性能を示している. また, 本研究では学習に投影数を減らした画像を利用していないが, 投影数の条件を変えた未知の入力を失敗なく復元できている.

投影数の減少と画質の関係という面で見ると, 提案手法では, 投影数を半分に減らしても PSNR の低下は約 0.30db で済んでいる. これは未知のデータに対する高い汎化性能と解釈できる. 一方 SART 法では投影数を半分に減らした場合の PSNR

の低下は約 2.0db となっている。これは提案手法が投影数が低いほど良い結果を得られるということを示している。

再構成された画像の詳細部分についてみると、提案手法は比較した全ての手法よりも優れた結果であった。これは提案手法がアーティファクトを発生させないためであると考えられる。この結果は、血管のより詳細な構造を見る必要がある石灰化の判定を行う診断において、提案手法が有効であることを示している。

しかし、提案手法では画像の詳細部分を見ると、全体的に平坦で、高周波成分が復元できていないことがわかる。一方、FPB 法ではノイズの中に詳細な構造を見ることができる。提案手法では、自然画では画像の詳細構造を復元できていることを考えると、医用画像における詳細構造（高周波成分）の欠落は学習に医用画像を利用していないことに起因すると考えられる。学習に医用画像を利用していないため、医用画像に特徴的な詳細の構造の特徴が復元できていない。学習データセットに医用画像を混ぜることで、性能向上を図ることができると考える。

再構成結果画像のセグメンテーションタスクによる、画質評価では、提案手法は 64 投影の時に既存手法に比べて最も良い結果を示した。また、提案手法は、投影数が 512, 256, 128 および 64 すべての場合で、ディープラーニングを利用した手法である FBP U-Net64, FBP U-Net512 および DD-Net を上回る結果となった。投影数が 512 から 128 の場合に最も良い結果となったのは TV-norm だった。これは、TV-norm のスパース性によると考えられる。TV-norm はノイズを除去するために使用されるが、ノイズ除去では必要な情報以外も除去してしまうことがある。今回の場合は TV-norm は診断に必要な本質的な情報は保持し不要なノイズ部分だけ除去でき、診断に必要な画像を復元できたと考えられる。一方、SART 法は一般にノイズやアーティファクトに強いと言われているが、セグメンテーションタスクの観点からは、不利な画像になってしまうことがわかった。その点において、提案手法は、必要な領域情報を保持したままノイズやアーティファクトを除去でき、優れているということが出来る。今回は、セグメンテーションタスクモデル自体の Dice スコアが 0.555 と低かった。これは学習不足のためと考えられる。今後は、より精度の高いセグメンテーションタスクモデルを作成し、評価する必要があると考える。

ディープラーニングでは、学習データセットが大きければ大きいほど精度の高いモデルを作成できることが知られている [52, 53]。そのため、提案手法を投影数の少な

い画像も加えて学習データを増やすことで、様々な投影数で再構成後の画像の精度が向上することが期待できる。また、投影数が少ない画像の再構成の場合においても、画質の低下がさらに少なくなることが期待される。

また、モデル自体も精度改良の余地があると考え。Guo ら [54] は Attention Block [55] を MRI の GAN による再構成に適用した CT 再構成に応用することを考えた時、Attention Block の特定の領域に着目するという特性は、サイノグラム画像を断面画像に変換するというタスクの性質とは相性が悪く単純には適用できない。なぜならサイノグラム画像は断面画像の全体の情報を各画素に積算して保持しており、それらを全体的に見ないとサイノグラム画像を断面画像に変換する問題を解くことができないためである。一方 Attention Block は Recurrent Model と組み合わせることで、全体的な文脈を保持することができると言われている。そのため、復元する画素に必要な情報を画像全体から選択することができ、精度が向上する可能性があると考え。

また、Schlemper ら [56] は MRI の再構成に DC-CNN [56] と Stochastic Depth [57] を適用し、良い結果を得たと述べている。Stochastic Depth は Res-Block の適用を前提とし、学習時にランダムに残渣パスを無視することにより、深いネットワークの学習の汎化性能を高め、学習時間を短くすると言われている。Stochastic Depth は活性化関数として正の値を前提とする ReLU を必要としているため、LeakyReLU を採用している提案手法には単純に適用することはできない。一方、今回の実験では、stack num を増やしすぎると学習時間が長くなり、結果も悪化するという課題があった。Stochastic Depth のような接続の制御を適用することで学習時間の短縮とスタック数の増加による精度の向上が期待できる。

Zhu ら [20] は病変部の ROI 画像と GAN を組み合わせて、学習収束が早い、高精度なモデルを構築した。提案手法は、サイノグラム画像から断面画像を直接復元するために、病変部を特定することはできない。しかし本手法を適用した後に、病変部を特定し、ROI 画像を作成し、さらに GAN を適用することで、精度が向上する可能性があると考え。また、病変部の ROI をサイノグラム画像から直接復元できれば、1. 復元、2. 病変部の ROI の作成、3. GAN による精度向上した病変部画像の作成といったステップを行うことなくダイレクトに病変部の ROI 画像を生成できるため計算コストを削減することができると考え。

また VGGLoss など loss 関数を工夫し精度を向上する方法も考えられる。VGGLoss は CT 画像の雑音除去などで成果を上げている [58] が本研究では詳しく述べていないが、予備実験時は単純に VGGLoss を適用した場合は精度が向上しなかった。TV-norm など既存手法で実績がある評価関数を組み込むことで、精度が向上できる可能性があると考えられる。

### 3.7 まとめ

提案手法では、Stacked U-Net に対して、レイヤ関数と stack 数を調整することで、サイノグラム画像から断面画像を再構成することができた。

提案モデルは、少ない投影回数において断面画像を再構成でき、逐次近似法と同等の画像を得ることができる。また、提案手法は既存手法に比べて処理速度が速く再構成ができるといった点で優れている。

提案されたモデルは、医用画像なしに断面画像を再構成できるように学習ができ、また、サイノグラム画像から直接断面画像を復元できる。この特徴により、FBP 法と U-Net を組み合わせた既存モデルの問題点である、投影数に応じたノイズ除去しできないという問題を回避することができる。

提案手法では、再構成のための CPU による処理時間は 2.35 秒であり、既存手法と同等である。しかし、GPU での実行も容易であり、既存手法より早い処理時間 (0.11s) を実現できる。また、少ない投影数でも再構成ができ、既存の手法よりアーティファクトの少ない断面画像を出力することができる。投影数を減らすことで、X 線の照射量を減らすことができ、患者の負荷を減らすことができる。

しかし、現在のモデルでは、復元された断面画像において詳細な構造が失われてしまう。そのため、精度の向上が今後の課題である。この問題は、学習データセットに投影数が少ない画像が含まれていないことや、学習量データセットが小さいこと、モデルが学習した医用画像の知識が不足していることが影響していると考えられる。

今後は、精度を向上させるために、投影数の少ない画像を加えて学習データセットを拡張する、トレーニングに用いる画像を増やし、再構成の知識を得やすくするために、医用画像や医用画像に似た特徴をもつ工業用製品の画像を自然画像に対して混ぜる、などを行い精度を向上していく。

また、ディープラーニングのモデルは日々新しい構造が提案されているため、最近された構造に対しても評価を行い、モデル自体の精度向上につなげる必要があると考える。Attention Block や Res Block, Stochastic Depth, などをモデルに適用し、精度向上につなげる。またモデルアーキテクチャ自体も U-Net が最適なのか考えていく必要がある。近年だと Position Embedding を使用し位置情報を埋め込み、3D view の精度を向上した NeRF [59] などがあり、レイトレーシングの復元を行っている。レイトレーシングは投影と同じ原理であり、再構成に相性がいいと思われる。様々なアーキテクチャの中から CT 再構成に最適なアーキテクチャを考える必要がある。

今回の研究では、画像の評価を Ground truth との定量的な比較と、学習済みのセグメンテーションタスクによって行なった。AI 診断用のセグメンテーションタスクによる評価は、多くの画像を評価できるため、モデルの改善につながりやすく、今後 AI 診断に向けて提案手法を改善していく。一方、実際の医療現場においては医師が診断しやすい画像かどうかという点も重要である。今後は、医師に評価してもらい、医療現場で使える画像に改善していく必要もあると考える。

CT 画像は臓器によって見やすいコンフィグレーションが違う、今回提案のモデルは上記コンフィグレーションの違いを反映していないため、それぞれのコンフィグレーションごとに応じた、復元精度が変化する可能性がある。モデルにコンフィグレーションパラメタを追加し、より医用現場で使用されている場面に近い再構成ができるようにしていく必要がある。

今回の研究では、stacked U-Net を利用したが、スタック数を深くすると結果が悪化することが確認された。これは、スタック数が深くなると、入力画像の特報が深いレイヤに対して伝わらなくなるためと考えられる。このような結果の悪化がおこならない仕組みを提案することで、stacked U-Net 自体の改善が行え、また、精度の向上が行えると考え。また、提案手法は SART 法とは異なる方法で再構成を実現していることを観察した。さらに低投影数の復元を行わせていき挙動や投影数の下限を観察することで、提案手法がどのような方法で再構成を実現しているかを明らかにし、提案手法の精度向上と既存手法の精度の向上にも寄与できると考える。

表 3.3: Comparison of the proposed method and existing ones.

methods	num projections	PSNR (db)	SSIM	processing time (s) (GPU)
Proposed	512	29.67	0.907	2.35(0.11)
	256	28.43	0.894	2.35(0.11)
	128	28.20	0.891	2.35(0.11)
	64	<b>27.93</b>	<b>0.886</b>	2.35(0.11)
FBP	512	<b>35.17</b>	<b>0.961</b>	1.71
	256	<b>31.93</b>	0.901	0.82
	128	26.56	0.734	0.41
	64	21.41	0.547	0.21
SART	512	31.07	0.949	16.79
	256	29.35	<b>0.926</b>	8.30
	128	27.24	0.884	4.11
	64	25.03	0.829	2.07
TV-norm	512	30.56	0.910	2.31
	256	30.50	0.909	1.49
	128	<b>29.86</b>	<b>0.893</b>	1.01
	64	25.73	0.750	0.21
FBP U-Net512	512	26.79	0.858	8.16(2.70)
	256	25.38	0.798	7.24(1.72)
	128	22.93	0.687	6.75(1.26)
	64	19.79	0.581	6.54(1.04)
FBP U-Net64	512	23.13	0.733	8.06(2.70)
	256	23.02	0.726	7.14(1.72)
	128	23.01	0.709	6.76(1.26)
	64	22.27	0.666	6.54(1.02)
DD-Net	512	26.75	0.911	6.49(2.66)
	256	26.58	0.898	5.60(1.77)
	128	25.51	0.832	5.19(1.36)
	64	22.75	0.717	4.99(1.16)

表 3.4: Comparison of details with 64 projections.

method	PSNR (db)	SSIM
proposed	<b>30.37</b>	<b>0.896</b>
FBP	25.17	0.570
SART	29.34	0.863
TV-norm	29.61	0.816
FBP U-Net512	18.75	0.494
FBP U-Net64	22.92	0.447
DD-Net	18.03	0.510

表 3.5: Comparison of segmentation task.

methods	num projections	Dice
Proposed	512	0.556
	256	0.545
	128	0.540
	64	<b>0.546</b>
FBP	512	0.556
	256	0.552
	128	0.538
	64	0.503
SART	512	0.458
	256	0.462
	128	0.478
	64	0.490
TV-norm	512	<b>0.562</b>
	256	<b>0.562</b>
	128	<b>0.561</b>
	64	0.539
FBP U-Net512	512	0.550
	256	0.541
	128	0.522
	64	0.476
FBP U-Net64	512	0.543
	256	0.544
	128	0.543
	64	0.533
DD-Net	512	0.514
	256	0.502
	128	0.478
	64	0.437

## 第 4 章

# Interlayer Augmentation による 学習データ増強

前述した通り，Stacked U-Net で CT の再構成を実現したものの，精度はいまだに改良の余地がある．学習用データを増やして，精度を改良することも考えられるが，医用画像は学習用データの数が多い患者のプライバシーなどの理由から自然画のデータに比べて少なく課題である．本節では，新しい Data argumentation の方法を提案し，少ないデータ数でもより精度が上がる方法を検討する．

### 4.1 はじめに

近年ディープラーニング技術は広く使用されており [60, 61]，精度の良いモデルを作成するための，訓練データの量も増えている．Chen ら [62] は学習データ量に応じて視覚タスクの性能が対数的に増加することを示した．例えば，ViT [63] は 3.75 億以上のノイズ付きラベルの 300 億枚の画像の非公開データセットである JFT-300M [62] で訓練されている．しかし，大きなデータセットは作成するのが難しい．特に，医用画像のような特定のドメインでは大きなデータを収集するのは不可能である [64, 65]．そのため，少量のデータで正確なモデルを作成する方法が研究されてきた [66]．

正確なモデルを作成する典型的な方法は regularization [67] と data augmentation である．

Regularization はモデルの汎化性能を向上させる方法である．Dropout [68] は

regularization の代表的な手法であり、モデルの学習時の途中データをランダムに0埋めすることで、汎化性能を向上させる。これはノイズへの耐性を向上させ、モデルがアンサンブル学習として機能するようになっているとされている [69]。しかし、Zhao ら [70] や Santra ら [71] は dropout が0埋めによる情報損失により、精度低下を引き起こすと報告している。

データ増強 (Data augmentation) は既存のデータに少しの修正を加えたコピーや、既存のデータから新たに作成した合成データを追加することで、学習データ量を増やす手法である。

たとえば、CutMix [72] は学習画像の領域を別の画像に置き換え、組み合わせた画像のピクセル数に比例してラベルを混合するように変更する。

Data augmentation はタスクごとに慎重に設計、実装、テストする必要がある再利用や汎用性に制限があると言われている [73]。

Feature space augmentation はモデルの特徴出力に基づいて、generative adversarial network (GAN) などを用いて、データを生成し、増強する方法である。ただし、学習するモデルのエンコード部分を丸ごとコピーする必要があり、深いモデルの場合、大規模なオートエンコーダになってしまい、学習が困難かつ、時間がかかるようになってしまう [64]。

本節では、レイヤ間でデータを増強する interlayer augmentation を提案する。そして、具体的な方法として batch generalization (BG) と random BG (RBG) を提案する。提案手法は、モデルにレイヤを挿入するだけなので、既存のモデルに容易に適用でき、さらに学習中に多くの種類のデータを生成することができる。

我々は BG と RBG を VGG [74], ResNet [75] および ViT に適用し、CIFAR10 と CIFAR100 の分類タスクで、スクラッチ学習した場合と比較し評価した。結果として、CIFAR10 では RBG で平均 0.39%, %BG で平均 0.27% の改善が見られ、CIFAR100 では RBG で平均 1.07%, BG で平均 0.30% の改善が見られた。特に、全てのケースにおいて、RBG はスクラッチ学習より優れた結果を示した。

## 4.2 関連研究

### 4.2.1 正則化 (Regularization)

正則化は、モデルの汎化性能を向上させるための手法である。代表的な手法としては、dropout [68] や、batch normalization [76] や、group normalization [77] がある。これらの手法は過学習 (overfitting) を防ぐことで、同じデータ量での汎化性能や性能を向上させることができる。

### 4.2.2 データ増強 (Data augmentation)

データ増強 (Data augmentation) は既存のデータに少しの修正を加えたコピーや、既存のデータから新たに作成した合成データを追加することで、学習データ量を増やす手法である。代表的なデータ増強の方法としては入力画像をシフトさせたり、ランダムに反転させたりする方法がある [67]。また、二つの入力画像を混ぜ合わせる Mixup [78] や、画像の一部を黒く塗りつぶす Cutout [79]、入力画像の一部と別の入力画像の一部を合成する CutMix [72] などの手法もある。

### 4.2.3 特徴空間データ増強 (Feature space augmentation)

特徴空間データ増強とは、モデルの特徴出力に基づいて、敵対的生成モデル (generative adversarial network:GAN) などを用いてデータを生成し、データ増強を行う方法である。De Vries らは [73]、分類タスクでの特徴空間でのデータ増強を示した。彼らは、まずオートエンコーダに学習を行わせた。オートエンコーダはエンコーダとデコーダで構成されるモデルであり、画像を入力し、同じ画像を出力するモデルである。エンコーダの出力は特徴ベクトルと解釈され、デコーダは特徴ベクトルに基づいて、エンコーダに入力された画像を復元する。彼らは学習させたエンコーダの特徴ベクトルに対し、補間を行うことで、新たな特徴ベクトルを生成し、それをデコーダにより復号化することによってデータ増強を行った。最後に、データ増強を行ったモデルに対して分類モデルを作成し、学習させた。

### 4.3 データの分布距離の分析

なぜ小さいデータでは過学習 (overfitting) が起こり、大きなデータでは過学習せずに精度が向上するのか。その理由は、小さいデータの分布は疎であるのに対し、大きなデータでは分布が密であるためである。

図 4.1 は小さいデータと大きいデータの場合のデータの分布の様式図である。

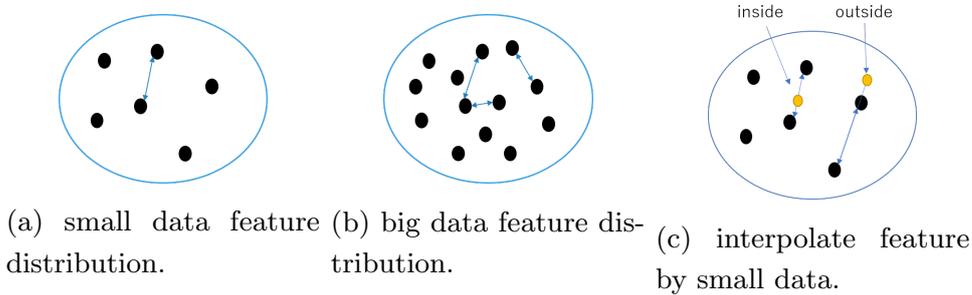


図 4.1: Distribution of feature values.

小さいデータの分布は大きいデータの分布より疎である、小さいデータの場合、モデルはデータの間を間違ってフィットする、そのため、モデルがデータの間を誤って補間し、過学習が起きると考えられる。

表 4.1 は様々なデータサイズに対する VGG11 の特徴空間の出力の分布の統計値を示している。

表 4.1: Statistic of feature space.

data size	min	max	mean	std
10	0.0143	0.166	0.0393	0.0242
100	0.00866	0.252	0.0377	0.0254
1000	0.00477	0.422	0.0382	0.0272
5000	0.00477	0.8433	0.0451	0.0338

特徴空間の分布はビンサイズ (bin size)255 のヒストグラムで作成し、各分布の距離は Kullback-Leibler divergence で測定した。表 4.1 に示すようにデータサイズが増加するに従って、平均は変化しないが、最小値は減少し、最大値は増加してい

る。これは図 4.1 の仮説と一致している。

上記から、データサイズの増加を模擬するためには、平均値が変わらないように分布の最小値と最大値を増やすことが重要である。

そこで我々は、図 4.1c のように二つのデータの間を補間することで新たなデータを増やすことができると考えた。もし新たなデータが 2 つの特徴の間に作成されれば、最小値は少なくなり、2 つの特徴点の外に作成されれば、最大値は大きくなる。

## 4.4 レイヤ間データ増強 (Interlayer augmentation method)

以上のことから、学習データは入力データだけでなくレイヤ間でも増強できることがわかる。ディープラーニングの各レイヤは特徴を抽出していると言われており、分類レイヤの出力の前のレイヤの出力は特徴量になると言われている。特徴量を出すレイヤより前のレイヤでは、データと特徴量が混在していると考えられる。また、Yun ら [72] や De Vries ら [73]、が示したように、データ空間でも特徴空間でもデータ増強は可能である。

もしレイヤ間でデータ増強を行うことができれば、各々のレイヤの間でデータ増強ができるということで、各層で増えたデータの最終的なパターンは、各層のデータを組み合わせることになり、データ量をべき乗で増やすことができる。

我々はこの概念をレイヤ間データ増強と名付けた。特に、レイヤ間データ増強の具体的な方法として RBG と BG を提案する。

### 4.4.1 RBG (Random Batch Generalization)

RBG はパッチのデータからランダムに二つのデータを選択し、データ増強する方法である。アルゴリズム 4 に順伝播 (forward propagation) アルゴリズムを示す。

ここで、 $\mathbf{x}$  はバッチ入力、 $\mathbf{l}$  はバッチ入力に対応するラベル、 $n$  はバッチサイズ、 $a$  および  $b$  はランダムに選択されたバッチ入力のインデックス値、 $l_a$  および  $l_b$  は  $x_a$  および  $x_b$  に対応するラベル、 $r(0 \leq r \leq 1)$  はデータ増強率、 $m$  (平均:0, 分散: $\epsilon$  のガウス分布) は修正率である。アルゴリズムはランダム  $batchsize * r$  個バッチ入力からペア  $P(x_a, x_b)$  を選び、 $x_a$  を選択されたペアに基づき変化させる。 $m$  はガウ

---

**Algorithm 4** RBG (forward in train).
 

---

**Require:** batch input  $\mathbf{x} = \{x_i : i = 0, 1, \dots, n-1\}$ ,  
 batch label  $\mathbf{l} = \{l_i : i = 0, 1, \dots, n-1\}$

**Ensure:** batch output  $\mathbf{x}' = \{x'_i : i = 0, 1, \dots, n-1\}$ ,  
 batch label  $\mathbf{l}' = \{l'_i : i = 0, 1, \dots, n-1\}$

$\mathbf{x}' \leftarrow \mathbf{x}$   
 from set  $\{(x_0, x_0), (x_0, x_1), \dots, (x_{n-1}, x_{n-1})\}$   
 choose  $rn$  pairs  $P(x_a, x_b)$   
**for** all P **do**  
    $m \leftarrow$  random number (Gauss distribution)  
    $x'_a \leftarrow (1 - m)x_a + mx_b$   
    $l'_a \leftarrow (1 - m)l_a + ml_b$   
**end for**

---

ス分布からランダムに選択された値であり、正の値にも負の値にもなりうる。もし  $m$  が正の値であるなら、 $x_a$  は  $x_a$  と  $x_b$  の間に变化する。これは最小距離が变化することを意味する。もし  $m$  が負の値であるなら、 $x_a$  は  $x_a$  と  $x_b$  の外に变化する。これは最大距離が变化することを意味する。 $x_a$  はペア  $P(x_a, x_b)$  から变化させるため、 $x_a$  のラベルも  $m$  に応じて变化させる。この振る舞いは、訓練時のみ実施され、RBG は評価時には何も行わない。

アルゴリズム 5 は RBG の逆伝播 (backward propagation) アルゴリズムを示している。

---

**Algorithm 5** backward
 

---

**Require:** gradient  $dL/dx_a, dL/dx_b$ , modification rate  $m$

**Ensure:** gradient  $dL/dx_a, dL/dx_b$

$dL/dx_a \leftarrow dL/dx_a(1 - m)$   
 $dL/dx_b \leftarrow dL/dx_b + dL/dx_a m$

---

順伝播時に  $x_a$  をペアに応じて变化させたため、修正率  $m$  を保存し  $m$  を傾きに反映させている。

### 4.4.2 BG (Batch Generalization)

BG はバッチのデータから同じラベルのペアを選んで、データ増強する方法である。アルゴリズム 6 に BG の順伝播アルゴリズムを示す。

---

**Algorithm 6** BG (forward in train).

---

**Require:** batch input  $\mathbf{x} = \{x_i : i = 0, 1, \dots, n - 1\}$ ,

batch label  $\mathbf{l} = \{l_i : i = 0, 1, \dots, n - 1\}$

**Ensure:** batch output  $\mathbf{x}' = \{x'_i : i = 0, 1, \dots, n - 1\}$

$\mathbf{x}' \leftarrow \mathbf{x}$

from set  $\{(x_0, x_0), (x_0, x_1), \dots, (x_{n-1}, x_{n-1})\}$

choose  $rn$  same label pairs  $P(x_a, x_b)$

**for** all  $P$  **do**

$m \leftarrow$  random number (Gauss distribution)

$x'_a \leftarrow (1 - m)x_a + mx_b$

**end for**

---

基本的な処理は RBG の場合と同じだが、BG は同じラベルを選んでデータを混合しているため、ラベルを変化させる必要がない、そのため、ラベルは出力されていない。BG も RBG と同じように評価時には何も行わない。逆伝播アルゴリズムは RBG のものと同じである (アルゴリズム 5)。

## 4.5 実験結果

### 4.5.1 クラス分類タスクへの適用

我々は、CIFAR10 と CIFAR100 [41] に RBG と BG を適用し、学習結果を比較し、提案手法の有効性を評価した。

表 4.2 と表 4.3 に CIFAR10 と CIFAR100 に元々のモデル (scrach learning), RBG を適用した結果および BG を適用し学習させた結果の peak signal-to-noise ratio (PSNR) の平均と標準偏差を示す

実験環境としてはフレームワークとして Pytorch を使用した。CIFAR10 と

CIFAR100 のトレーニングセットの数は 50,000 で、テストセットの数は 10,000 である。出力であるラベル数は CIFAR10 で 10, CIFAR100 で 100 である。そのため、各モデルの回収出力の全結合レイヤ (fully connected layer) の出力を 10 と 100 に修正した。ViT では [63] にならい、レイヤ数, ヘッド数, パッチサイズをそれぞれ 12, 12, (16, 16) とし、埋め込みレイヤ (embedding layer) はストライド 16, カーネルサイズ (16, 16) の畳み込み (convolution) としたが、CIFAR10 で、正解率が約 60% と論文で示されている値とだいぶ違っていた。そのため、埋め込みレイヤをストライド 2, カーネルサイズ (2, 2) の四層の畳み込みにした。結果 CIFAR10 で、正解率が約 84% に改善したので、実験ではこの構造を使用した。

各学習は 5 回実行し、200 エポック実行した。最適化には確率的勾配降下法 (stochastic gradient descent) を用いた。学習率スケジューラには CosineAnnealing を使用し T\_max は 200 に設定した。

RBG と BG はそれぞれ図 4.2 から図 4.4 に示す場所に挿入した。VGG では Conv2d の前に挿入した (図 4.2)。そのため、RBG と BG は VGG11 で 7 回、VGG19 で 15 回挿入された。ResNet では ResBlock の前に挿入した (図 4.3)。そのため、RBG と BG は ResNet20 で 9 回、ResNet110 で 54 回挿入された。ViT では埋め込み (embedding) の後ろと feedforard の前に挿入した (図 4.4)。そのため、RBG と BG は 13 回挿入された。 $r$  は VGG11 と ResNet20 で 0.1 を使用し、VGG19, ResNet110, ViT では 0.01 を使用した。 $\epsilon$  は 0.4 を使用した。

表 4.2: cifar10 results.

model	scratch	RBG	BG
VGG11	91.10 $\pm$ 0.09	<b>91.41</b> $\pm$ 0.10	91.24 $\pm$ 0.11
VGG19	92.43 $\pm$ 0.06	92.97 $\pm$ 0.16	<b>93.06</b> $\pm$ 0.06
ResNet20	92.82 $\pm$ 0.10	93.19 $\pm$ 0.20	<b>93.22</b> $\pm$ 0.12
ResNet110	94.69 $\pm$ 0.06	<b>94.97</b> $\pm$ 0.09	94.92 $\pm$ 0.17
ViT	84.88 $\pm$ 0.13	<b>85.34</b> $\pm$ 0.06	85.23 $\pm$ 0.14

CIFAR10 での平均の改善は RBG で 0.39%, BG で 0.27% である。

CIFAR100 での平均の改善は RBG で 1.07%, BG で 0.30% である。

表 4.4 に各モデルの augment rate ( $r$ )0.1, 0.01 それぞれでのエポックごとの実

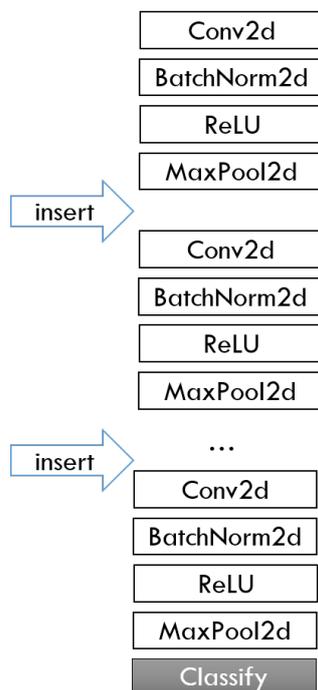


図 4.2: vgg's RBG and BG insert positions.

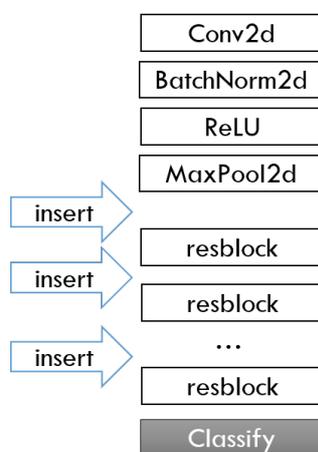


図 4.3: resnet's RBG and BG insert positions.

行時間をしめす.

#### 4.5.2 CT 再構成タスクへの適用

上記の結果を受けて CT 再構成タスクへ適用し評価を行った. 実験では CIFAR100 [80] を使用し, 50,000 枚をトレーニングセットに, 10,000 枚をバリデー

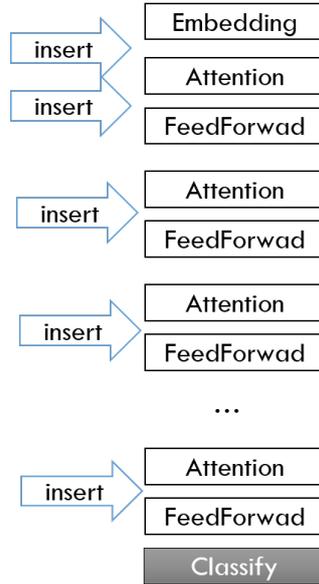


図 4.4: ViT's RBG and BG insert positions.

表 4.3: cifar100 results.

model	scratch	RBG	BG
VGG11	67.10 $\pm$ 0.29	<b>69.65</b> $\pm$ 0.52	66.80 $\pm$ 0.29
VGG19	70.21 $\pm$ 0.16	<b>71.10</b> $\pm$ 0.18	69.80 $\pm$ 0.21
ResNet20	64.09 $\pm$ 0.30	64.67 $\pm$ 0.11	<b>65.59</b> $\pm$ 0.16
ResNet110	66.71 $\pm$ 0.21	<b>67.92</b> $\pm$ 0.15	67.79 $\pm$ 0.27
ViT	60.84 $\pm$ 0.21	<b>60.96</b> $\pm$ 0.18	60.47 $\pm$ 0.22

ションセットとして使用した。出力画像である断面画像は CIFAR100 の画像をグレースケール化ののち、画像の真ん中を円状に切り取り生成した。入力画像であるサイノグラム画像は、出力画像をもとに scikit-image [45] の radon 関数により生成した。CT 再構成タスクは画像を入力し、画像を出力とするタスクであるため、ラベル出力を前提とする BG には適用できない。そのため RBG のみを評価し、 $r$  は 0.05、 $\epsilon$  は 0.1 を使用した。また Optimizer は Adam を使用し 100epoch 学習を行った。使用したモデルの構造を図 4.5 に示す。

スタック数 4、レイヤ数 3 の Stacked U-Net であり。RBG の挿入箇所は数字で示してある、Stacked U-Net の入力と出力を除いた、各 U-Net の最初の入力、特徴量変換後、特徴点から画像変換の各々の部分で 11 箇所ある。結果を表 4.5 に示す。

表 4.4: Per epoch execution time for each augment rate.

model	scratch	RBG		BG	
		0.1	0.01	0.1	0.01
vgg11	24.8	30.5	25.0	31.0	24.6
vgg19	28.9	41.0	29.2	42.2	29.0
resnet20	18.9	27.9	19.0	28.1	18.9
resnet110	35.1	90.0	35.9	91.6	35.6
ViT	30.0	40.3	30.6	42.1	30.1

0.1 and 0.01 are augment rate  $r$ 

Execution times unit is sec

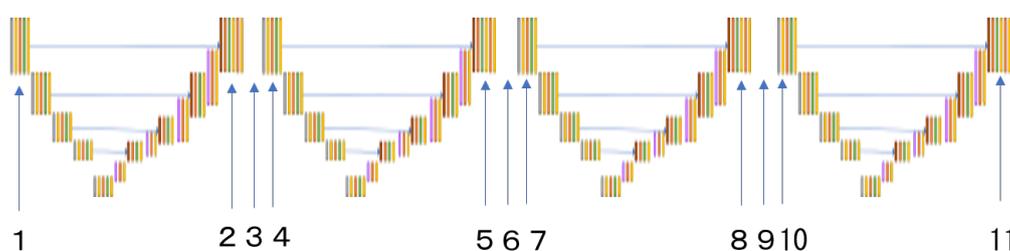


図 4.5: 4 stack Stacked U-Net and insert positions.

表 4.5: Results of RBG on CT Reconstruction Problem

insert num	PSNR(db)
2	26.06
3	26.25
4	25.39
6	25.34
8	26.05
11	25.66
without RBG	<b>27.09</b>

挿入箇所それぞれ, 2 の場合 1 と 2, 3 の場合 3 箇所で 1~3, 4 の場合 1~4, 6 の場合 1~6, 8 の場合 1~8, 11 の場合 1~11 である. 比較のため RBG を適用しなかった場合も without RBG として載せている. CT 再構成問題の場合一番良い結果は RBG を適用しなかった場合であり, 改善しなかった.

## 4.6 考察

### 4.6.1 クラス分類タスク考察

表 4.2 に示すように CIFAR10 では RBG と BG 両方ともスクラッチ学習を上回る結果となった。VGG と ResNet では RBG と BG の間に差がなく、ViT は BG よりも RBG のほうが良い結果を示した。

表 4.3 に示すように、CIFAR100 では VGG と ViT が RBG がスクラッチ学習を上回る結果となった。一方、ResNet では、RBG と BG の両方がスクラッチ学習を上回り、BG が RBG を上回った。

図 4.6, 図 4.7, および図 4.8 は CIFAR100 での ResNet110 の学習曲線を示している。

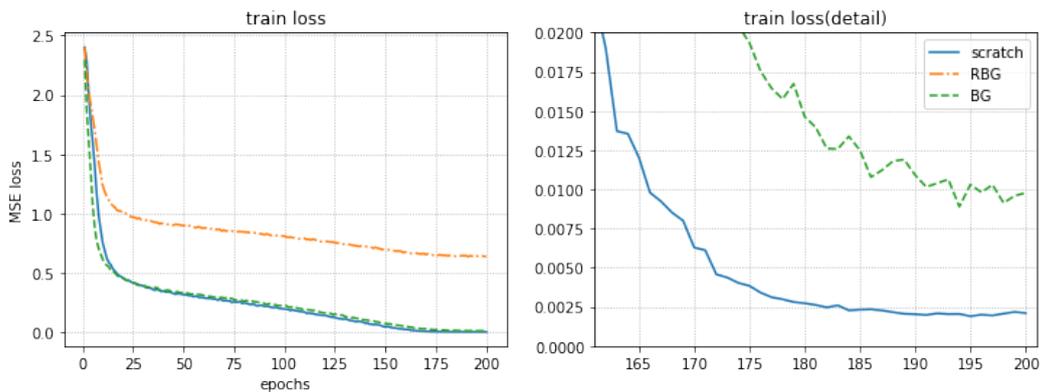


図 4.6: resnet110 cifar100 train curve.

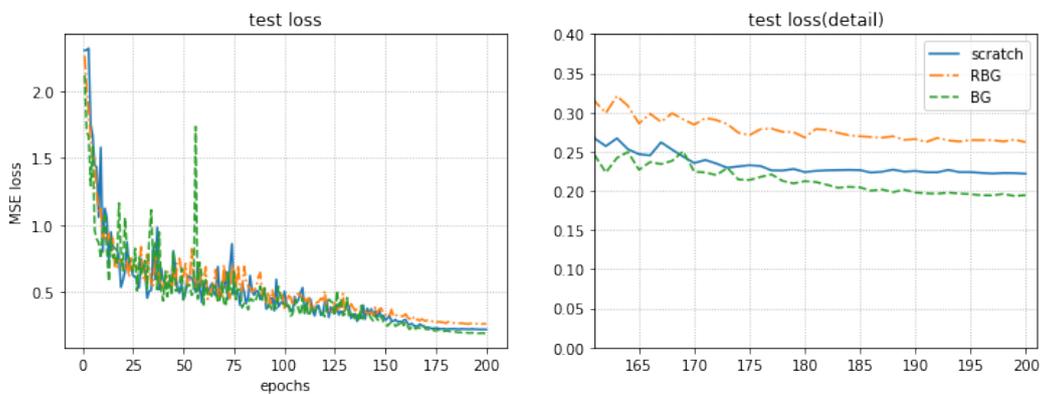


図 4.7: resnet110 cifar100 test curve.

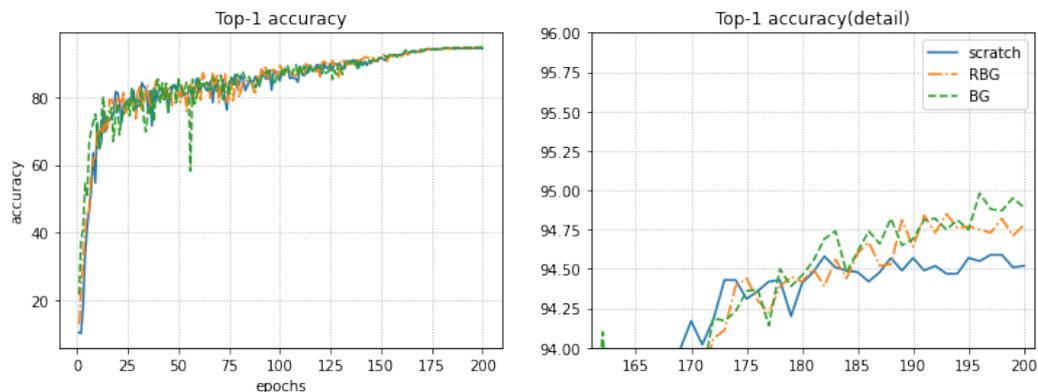


図 4.8: resnet110 cifar100 accuracy.

scratch と BG では、訓練損失 (training loss) が テスト損失 (test loss) よりも小さく、過学習 (overfitting) が起きている。しかし、RBG では訓練損失がテスト損失よりも大きく、過学習は発生していない。BG の場合、訓練損失は 0.19、テスト損失は 0.10 に収束した。scratch では、訓練損失が 0.22、テスト損失が 0.02 に損失した。そのため、BG のほうが過学習に強いといえる。

訓練損失とテスト損失の差は、scratch と BG では小さかったが、RBG では訓練損失は 0.64 に収束し、テスト損失の 0.26 とは大きく異なっていた。これは、訓練データとテストデータの間乖離があったことを示しており、補間方法として採用した 2 点間の線形補間とラベルの変更では、データをうまく補間できていないことを示唆している。本研究では、単純にベクトルの間の点を補間した。しかし、特徴空間の局面に沿って補間を行えば、訓練データとテストデータの乖離が少なくなり、結果が改善されることが考えられる。

また、BG の場合、訓練損失とテスト損失の乖離は RBG より小さく、同じラベルの間で補間したデータではテストデータと近い空間にデータが分布していることがわかる。これは、同じラベルの間では、近い空間に分布があるため、補間がうまくいき、データ増強がうまくいっていることを示している。

scratch, BG, RBG のテスト損失はそれぞれ 0.02, 0.10, 0.64 だったのに対し、正答率 (accuracy) はそれぞれ 94.54%, 94.76%, 94.90%, だった。正答率の面では、RBG と BG の差は有意ではなく、訓練損失とテスト損失の差ほどは、データ拡張の良さに差はない。

VGG11 と ResNet では  $r$  として 0.1 を、その他のモデルでは  $r$  として 0.01 を

使用した。これは VGG19, ResNet110 や ViT のような深いモデルではレートが大きくなると、性能が悪くなるからである。これは、今回は各モデルに機械的に RBG と BG を挿入したため、深いモデルでは挿入数が多くなり、データの変化率が大きくなりすぎたためであると考えられる。2 点間をより正確に補間できる手法や、挿入箇所を減らすなどの工夫により、レートが大きくても良い結果が得られると考える。提案手法では、レートが大きければ大きいほど、挿入された各レイヤで多くのデータが生成され、汎化性能が向上することになる。また、モデルごとに適切な挿入位置とレートを探索すれば、さらに精度を高めることができると考える。

ViT では表 4.2 および 4.3 に示すように、正解率は CIFAR10 で 85.34%, 60, 96% となっており、VGG や ResNet より大幅に低い。RGB により正解率は向上しているものの、大規模なデータセットで学習した、ViT は 99.50% を記録したといわれている。提案手法は大規模なデータセットを模擬するにはまだ不十分であるといえる。

また表 4.4 にしめすように、RBG および BG の実行時間はレート  $r$  が 0.01 の時 scratch の実行時間とほぼ同じだった。このように、提案手法の利点として実行時間がほぼ変わらないことが挙げられる。 $r$  が 0.1 の時は提案手法の実行時間は scratch より大きくなっている、これは深いモデルでは挿入回数が多くなるためであり、挿入数を調整することで、実行時間を短くすることができる。

#### 256 × 256 画像での評価

本手法が小さい画像のみで有効である可能性があるため、256×256 の画像に対し RBG および BG の適用を行い評価を行った。学習に使用したデータセットは Imagenette [81] を使用した。Imagenette は ImageNet のサブセットであり、クラス数は 10 である。学習データは 12900 枚を Train set として、100 枚を Test set として使用した。Optimizer は Adam を使用し、学習率は 0.1 とし、resnet18 および resnet20 に対し 200epoch 学習させた。結果を表 4.6 に示す。

resnet18 においては accuracy は scratch で 80.20%, RBG では 84.20%, BG では 78.60% である。resnet20 においては accuracy は scratch で 73.00%, RBG では 80.20%, BG では 75.20% である。resnet18 および resnet20 では、RBG の結果が scratch よりも結果が良く、BG は resnet20 では、scratch よりも良い結

表 4.6: Imagenette 256 results.

model	method	r	$\epsilon$	accuracy
resnet18	scratch	-	-	80.20
	RBG	0.1	0.4	84.20
	BG	0.1	0.4	78.60
resnet20	scratch	-	-	73.00
	RBG	0.01	0.4	80.20
	BG	0.01	0.4	75.20

果が得られた。このことから提案手法は大きい画像でも有効であることがわかった。かつ CIFAR10 や CIFAR100 での学習の結果よりも、改善効果大きい。そのため、大きい画像に対して適用したほうがより改善効果大きい可能性がある。

#### CutMix との併用評価

また、既存の Augment 手法とどの程度親和性があるかを確認するために、CutMix と本手法を併用し、実験を行った。学習対象は CIFAR10 である。Optimizer は SGD で、resnet20 に対して 200epoch 学習を行った。CutMix はラベルを変更するために、BG は併用できない、そのため RBG のみで評価を行った。結果を表 4.7 にしめす。

表 4.7: CutMix combination results.

model	method	parameter	accuracy
resnet20	CutMix	$\beta = 1.0$ $p = 1.0$	92.45 $\pm 0.10$
	CutMix	$\beta = 1.0$ $p = 1.0$	91.89 $\pm 0.07$
	+RBG	$r = 0.1$ $\epsilon = 0.4$	

CutMix 単体の場合は精度は 92.45%、RBG を併用した場合は 91.89% である。CutMix と RBG を併用した場合は、精度は低下している。そのため提案手法は CutMix とは併用できない。しかし、表 4.2 にあるように、RBG のみの適用の場合の精度は 93.19% であり、提案手法のほうが CutMix より優れている。

### 4.6.2 CT 再構成タスク考察

今回 RGB を CT 再構成タスクに適用したが、RGB 適用なしのモデルと比べて結果が悪くなった。これは、クラス分類タスクではラベルという大まかな正解に対して、補間を行うことで大体あったラベルが作成されていたが、CT 再構成タスクでは正解は画像であり、画像は細かな変化が精度に効くため上手くいかなかったためであると考えられる。

実際の学習値の精度を epoch ごとに観察すると図 4.9 に示すように、発散するような挙動を示している。これは RGB による補間により、本来の学習データ上の特

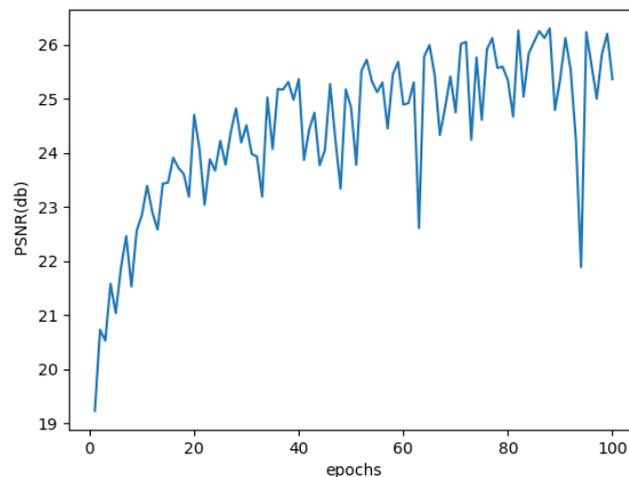


図 4.9: Insertion number 3 learning curve.

微量とかけ離れたデータが生成され、学習が安定しなくなっているとみることができる。

また図 4.10 に示すのは、左 2 枚の出力教師画像を RGB により補間した画像を生成した例である。右 2 枚の画像が補間された画像であるが、1 つめは画像の周囲の円周部分が灰色になっている。CT 再構成において生成されるべき画像は周囲の円周部分は黒であるべきである。また 2 つめは、画像の中心付近のきのこのテクスチャが白飛びしている。このように現状の RGB では補間がうまくいかず、本来ないはずの値域の画像が正解画像として生成されている。このことから、正解画像のデータとかけ離れた教師画像が作成されることがうまくいかない原因になっていると予測でき

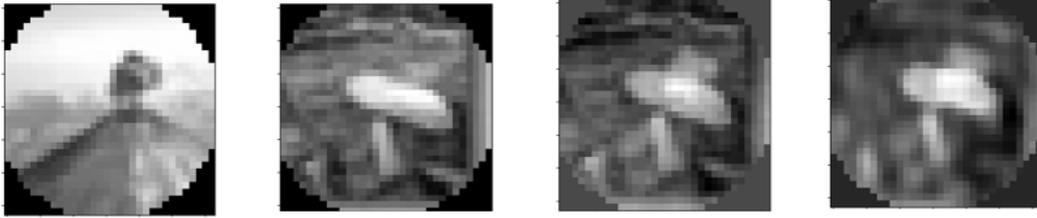


図 4.10: Example of image intrpolation by RBG.

る。これらの改善のためには、正解画像の補間関数を工夫し、本来の値域内に補間の範囲を収めることができるようにする必要がある。

#### 補間関数の内挿化

上記観察から、白飛びや周囲部の灰色化がおこらないように補間の方法を式 4.1, 式 4.2 に示す方法 (内挿) に変更した。

$$x'_a \leftarrow (1 - |m|)x_a + |m|x_b \quad (4.1)$$

$$l'_a \leftarrow (1 - |m|)l_a + |m|l_b \quad (4.2)$$

上記変更の元 RBG を再度 CIFAR100 画像の CT 再構成タスクに実施した結果を表 4.8 に示す。使用したのは 4 段の Stacked U-Net で図 4.11 に示す箇所それぞれの箇所に対して、RBG を挿入して  $r = 0.1$ ,  $\epsilon = 0.1$  とし実験を行った。

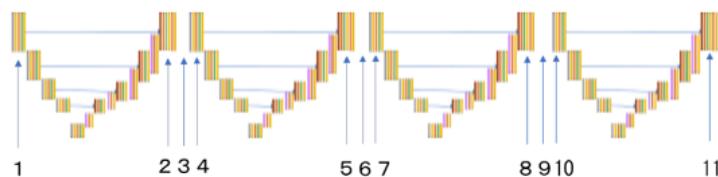


図 4.11: Insertion points.

結果は、表 4.8 に示す通り、RBG を適用しない場合 (scratch) の 26.68db と比較し、挿入箇所 1, 6, 7, 8, 10, 11 の時に優れた結果を得た。特に 11 の時は 27.13db と一番優れた結果を得た。表 4.8 では前半に挿入するより後半に挿入するほうが PSNR が改善しやすい傾向が見える。これはサイノグラム画像と断面画像に相関がないためと考えられ、断面画像が復元されてきている、後半の補間が効果的に働いていると考えられる。上記では単純に一か所挿入しただけだったが、多数挿入し、 $r$  や  $\epsilon$  を前半は RBG のでの変化が少ないように調整し、後半では効果が大きい

表 4.8: Impoved RBG's mean and variance.

insert point	PSNR
scratch	26.68 $\pm$ 0.58
1	26.94 $\pm$ 0.40
2	26.17 $\pm$ 0.31
3	26.67 $\pm$ 0.14
4	26.18 $\pm$ 0.34
5	26.30 $\pm$ 0.52
6	26.70 $\pm$ 0.22
7	26.94 $\pm$ 0.25
8	26.77 $\pm$ 0.23
9	26.60 $\pm$ 0.24
10	27.05 $\pm$ 0.16
11	27.13 $\pm$ 0.45

ように調整することで augment される data の量が増えさらに精度改善ができると考える。

## 4.7 まとめ

本章では, interlayer augmentation を提案し, RBG と BG を具体的な手法として提案した. これらの手法は単純であり, そのため学習時の計算コストは少なくて済む. これらの手法は, CIFAR10 および CIFAR100 タスクにおいて, VGG11, VGG19, ResNet20, ResNet110, ViT のスクラッチ学習を上回る結果を得た. 特に, すべてのケースにおいて, RBG はスクラッチ学習より良い結果を示した. 提案手法は, 増強レート ( $r$ ), ガウス分布の標準偏差 ( $m$ ), 挿入点などのハイパーパラメタがある. これらのハイパーパラメタの最適値の設定は困難である. また, 提案した手法は 2 つのデータベクトルの線形補間で, 新たなデータを作成しており, 真の空間を十分に再現できていない.

今後の課題としては, 2 つのデータベクトル間の補間を特徴空間の曲面に沿うようにすることで, 精度を向上させるとともに, モデルに応じて最適なパラメータや挿入

点を見つけるためのハイパーパラメータ探索を行うことが挙げられる。また、アルゴリズムの性質上増強レート ( $r$ ) と挿入数を多くすると学習速度が悪化する。これらのハイパーパラメータの設定と学習の改善率にはトレードオフが存在すると思われ、最適な増強レート ( $r$ ) と挿入数と学習改善率の関係を明らかにすることも必要である。

また、RBG を CT 再構成に適用した場合は、学習精度が悪化した。元の RBG は元の特徴の値域をうまく表現できておらず、学習も発散の挙動を示している。クラス分類ではおおまかな範囲での分類のため、RBG により精度が向上したが、画像の再構成タスクではクラス分類と比べてより細かい画素の精度が要求される。そのため、CT 画像の再構成にはそのままでは適用できなかった。しかし、学習タスクの特性を考慮し、補間式を少し変更するだけで、精度が向上した結果を得ることができた。これにより、CT 再構成タスクにおいても Interlayer augmentation は精度の改善に寄与できることが示された。

## 第 5 章

# Stacked U-Net の改良

前述した通り，Stacked U-Net で CT の再構成を実現したものの，精度はいまだに改良の余地がある．そのためモデル自体を改良することで精度を良くする方法を提案する．

Stacked U-Net は U-Net を多層重ねたモデルであり，画像から画像の変換に使用され，重ねる数（スタック数）を多くするだけで精度が向上する．Stacked U-Net はスタック数を多くするだけで，精度が向上するために，使用しやすい．一方でスタック数が大きくなりすぎると，精度が悪くなる．

本節では U-Net の skip connection を改良し，高周波成分を伝えやすくするモデルを提案し，スタック数を増やしても精度が低下しづらくすることを示す．

### 5.1 はじめに

Ronneberger ら [15] は U-Net をセグメンテーションタスクに対して提案した．U-Net は Convolutional Neural Network (CNN) モデルで，画像を入力し，同じサイズの画像を出力する fully convolutional network である．U-Net の構造はエンコーディングと，デコーディングに大きく分かれ，エンコーディングでは画像は段階的に縮小され，縮小の過程で空間情報は特徴情報として抽出される．デコーディングでは特徴情報をもとに空間情報を段階的に拡大する．また skip connection により同じ階層のエンコーディング結果の特徴情報が，デコーディングに渡される．これらの構造が U 字構造となっているため U-Net と呼ばれている．

U-Net はセグメンテーションタスクにおいて成功したため，様々な改良が施さ

れおり, U-Net に residual block [27] を組み合わせた Res-U-Net [28], Dense Net [82] と組み合わせた Dense-U-Net [83], U-Net++ [49] が提案されている.

Stacked U-Net においても Res-U-Net を適用し [30], 重ねたもの (stack したものの) が提案されているが, 同時に U-Net を重ねすぎると精度が悪化することが指摘されている.

本章では, stack 数による精度劣化に対する構造として, Skip Down および Skip Feature を提案し, 実験を行った. 結果として, CT 再構成タスクでは Skip Feature においては既存手法である Skip Normal より良い評価値を得ることができた. また Skip Down も浅いスタック数では良い結果を示した.

## 5.2 関連研究

### 5.2.1 U-Net の改良

U-Net は高周波成分が skip connection により後段に伝わりやすいと言われている. 一方で, 精度の面ではさらなる改良の余地があるため, U-Net はいろいろな方法で改良されている. Zhang ら [28] は residual block [27] と U-Net を組み合わせた Res-U-Net を提案している. residual block は残渣を伝えるブロックであり, これの適用により, 勾配損失問題に対処し, 層が深いモデルを作成しやすくしている. また, Zeng ら [83] Dense Net [82] と U-Net を組み合わせた Dense-U-Net を提案している. Dense Net はすべてのレイヤーを直接接続させることで勾配消失に対処している.

### 5.2.2 Stacked U-net

Shah ら [31] は U-Net を 7 層重ねたモデルを提案している. 彼らはセグメンテーションタスクである土地被覆問題 (land cover classification task) に stacked U-Net を適用し, state-of-the-art の結果を得た. Sevastopolsky ら [30] も stacked U-Net を optic cap segmentation に適用した. 彼らは U-Net を 1 層から 20 層まで重ねたモデルで, 実験を行い, 1 から 15 層までは徐々に結果が改善し 15 層で, 一番良い結果を得た. しかし 15 層以降は徐々に精度が悪化していった. そのため, stacked U-Net はスタック数が大きくなりすぎると, 精度が悪くなることを指摘し

ている [30]. 彼らはこの問題に対処するために, resblock を採用した, resblock は残渣のみを後段に伝えるため, 高周波成分が後段に伝わりやすくなり, 深いネットワークでも学習が可能になると言われている.

### 5.3 Skip Connection の改良

Sevastopolsky ら [30] は stack 数を深くする方法として, residual block の適用を行ったが, 依然として stack 数を深くすると精度は悪化している.

そこで我々は, U-Net に特有の skip connection に着目した. U-Net はこの skip connection により高周波成分が保存され, 後段に伝わりやすくなると言われている. しかし Han ら [84] が示すように, 高周波成分は十分に伝えられていない.

また我々の実験では, residual block は簡単に既存のレイヤと置き換えが可能であるが, 予備的な実験では stacked U-net に適用しても収束しなかった. Dense U-Net は残渣を Dense に skip connection でつないでいくが, モデルパラメタが多くなり学習が遅くなる, またこちらも予備的な実験で収束しなかった.

そのため元々の skip connection 自体を改良する方が, モデルパラメタ面, 収束面で有利と考え, 改良した skip connection を提案する.

### 5.4 stack 数の増加に耐える skip connection 構造

スタック数の増加による精度の低下の理由は高周波成分がうまく伝搬されていないからである. そのため, 我々は U-Net の skip connection を改良し, 高周波のみを伝える構造である Skip Down (図 5.8) と Skip Feature (図 5.9) を提案した.

U-Net の構造は図 5.1 に示す通りである.

エンコーディングレイヤは図 5.2 の赤丸で示す部分である. エンコーディングレイヤはすべて同じ構造をしており (図 5.3), 最初の  $3 \times 3$  Convolution, BatchNorm, Leaky ReLU により, 複数のチャンネルに特徴量を抽出し, Down Sample レイヤ (図では stride2 の  $2 \times 2$  Convolution) により画像のサイズを  $1/2$  に縮小する. エンコーディングレイヤは何層にも重ねられて, 特徴量を出力する.

デコーディングレイヤは図 5.4 の赤丸で示す部分である. デコーディングレイヤはすべて同じ構造をしており (図 5.5),





図 5.3: Encoding layer.

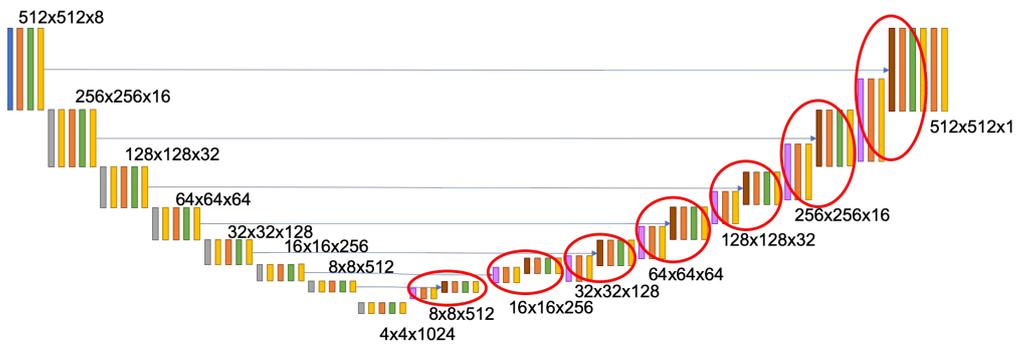


図 5.4: Decoding layer's in U-Net.

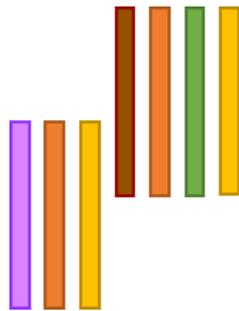


図 5.5: Decoding layer.



図 5.6: Skip connection between encoder and decoder layers.

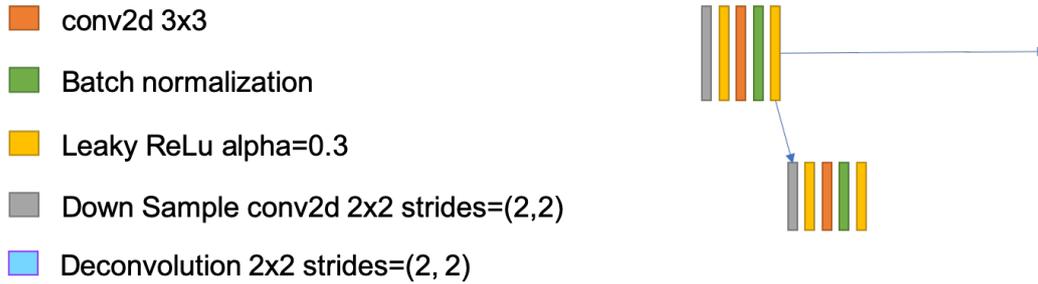


図 5.7: Skip Normal.

**Algorithm 7** SkipNormal.

**Require:** Leaky ReLU output  $\mathbf{x} = \{x_i : i = 0, 1, \dots, n - 1\}$

**Ensure:** batch output  $\mathbf{y} = \{y_i : i = 0, 1, \dots, n - 1\}$

skip connection output  $\mathbf{z} = \{z_i : i = 0, 1, \dots, n - 1\}$

$\mathbf{y} \leftarrow \text{DownSample}(\mathbf{x})$

$\mathbf{y} \leftarrow \text{LeakyReLU}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{Conv2d}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{BatchNorm}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{LeakyReLU}(\mathbf{y})$

$\mathbf{z} \leftarrow \mathbf{x}$

から引くことで、高周波成分が残るようにした。

通常の U-Net の skip connection を Skip Normal (図 5.7) とする。アルゴリズム 7 に疑似コードを示す。ここで  $\mathbf{x}$  はエンコーディングレイヤの特徴出力である縮小する前の LeakyReLU の結果、 $\mathbf{y}$  はエンコーディングレイヤの出力、 $\mathbf{z}$  は skip connection の出力である。DownSample, LeakyReLU, Conv2d, BatchNorm はそれぞれ関数であり。DownSampe は  $2 \times 2$  の stride2 の Convolution, LeakyReLU の  $\alpha$  は 0.3, Conv2d は  $3 \times 3$  の stride1 の Convolution, BatchNorm の  $\epsilon$  は 0.00001 momentum は 0.1 である。アルゴリズム 7 では、単に  $\mathbf{z}$  に  $\mathbf{x}$  を代入している。これは、縮小する前の情報を単に decoding layer に伝えているということである。

我々が提案する高周波を伝えやすくする構造の一つ目は、縮小された画像を単に拡

大する方法である，これを Skip Down と名付けた．(図 5.8)．アルゴリズム 8 に疑似コードを示す．ここで *Deconvolution* は関数であり，kernel サイズは  $2 \times 2$  で stride2 である．SkipNormal とは違い *DownSample* および *LeakyReLU* を適用した結果に *Deconvolution* を適用しその結果と  $\mathbf{x}$  の差分を  $\mathbf{z}$  としている．*DownSample* によっていったん縮小された画像を *Deconvolution* により拡大することにより，縮小により低周波成分だけになった情報を  $\mathbf{x}$  から引くことにより，高周波成分だけが  $\mathbf{z}$  に残ることを意図している．

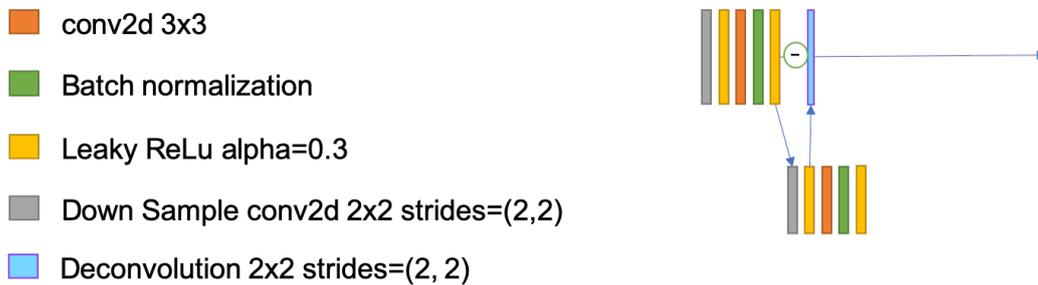


図 5.8: Skip Down.

---

**Algorithm 8** SkipDown.

---

**Require:** Leaky ReLU output  $\mathbf{x} = \{x_i : i = 0, 1, \dots, n - 1\}$

**Ensure:** batch output  $\mathbf{y} = \{y_i : i = 0, 1, \dots, n - 1\}$

skip connection output  $\mathbf{z} = \{z_i : i = 0, 1, \dots, n - 1\}$

$\mathbf{y} \leftarrow \text{DownSample}(\mathbf{x})$

$\mathbf{y} \leftarrow \text{LeakyReLU}(\mathbf{y})$

$\mathbf{z} \leftarrow \mathbf{x} - \text{Deconvolution}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{Conv2d}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{BatchNorm}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{LeakyReLU}(\mathbf{y})$

---

もう一つは，convolution がかった特徴量に対して，拡大を行う方法で，これを Skip Feature と名付けた (図 5.9)．アルゴリズム 9 に疑似コードを示す．Skip-Down とは違い，*DownSample* により縮小されたうえで *LeakyReLU*，*Conv2d*，*BatchNorm* および *LeakyReLU* を適用した結果に対し，*Deconvolution* して

いる。画像のサイズを変えない *Conv2d*, *BatchNorm* および *LeakyReLU* の適用は、局所的な特徴量を抽出しているといわれ、そのため縮小され高周波が失われたらうえ、特徴的な情報のみになった画像を拡大することになり、その過程で失われた、高周波および特徴的でないと判断された情報を skip connection で伝えることになる。

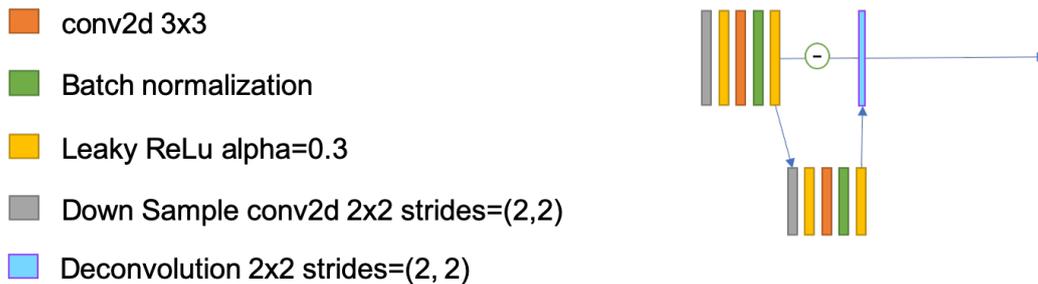


図 5.9: Skip Feature.

---

**Algorithm 9** SkipFeature.

---

**Require:** Leaky ReLU output  $\mathbf{x} = \{x_i : i = 0, 1, \dots, n - 1\}$

**Ensure:** batch output  $\mathbf{y} = \{y_i : i = 0, 1, \dots, n - 1\}$

skip connection output  $\mathbf{z} = \{z_i : i = 0, 1, \dots, n - 1\}$

$\mathbf{y} \leftarrow \text{DownSample}(\mathbf{x})$

$\mathbf{y} \leftarrow \text{LeakyReLU}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{Conv2d}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{BatchNorm}(\mathbf{y})$

$\mathbf{y} \leftarrow \text{LeakyReLU}(\mathbf{y})$

$\mathbf{z} \leftarrow \mathbf{x} - \text{Deconvolution}(\mathbf{y})$

---

## 5.5 実験結果

上述した構造に対して効果を確認するため、二つのタスクで評価を行った。下記の実験では、Intel (R) Pentium (R) CPU G3260 3.30GHz Memory 24Gbyte GPU: RTX2080Ti を使用した。

### 5.5.1 セグメンテーションタスク

U-Net はもともとセグメンテーションタスク用に開発された，そのため，まずセグメンテーションタスクで評価を行った．セグメンテーションタスクでは，図 5.10 に示すような，画像から図 5.12 に示す細胞の境界を示す画像を出すことを目的としている．

実験では，2D EM segmentation [85] データセットから，168 枚の画像をトレーニングに使用し，12 枚の画像をバリデーションに使用した．モデルの構造は図 5.1 に示す通りである．

スタック数は 1 から 8 に変化させトレーニングを 30epoch 実施した．バリデーションセットにおける平均の MSE は表 5.1 に示す結果となった．

表 5.1: Segmentation task stack results.

stack num	skip-normal	skip-down	skip-feature
1	0.380410	0.38422	0.380422
2	0.380421	0.38422	0.380422
3	0.380422	0.38422	0.380421
4	0.380422	0.38422	0.380422
5	0.380422	0.38422	0.380422
6	0.380422	0.38422	0.380422
7	0.380422	0.38422	0.380422
8	0.380422	0.38422	0.380422

結果は，Skip Normal はスタック数 1 から 3 で改善している．しかし，Skip Normal, Skip Down, Skip Feature のほぼ全ての結果が同じだった．特に Skip Down と Skip Feature は 0.34822 と Skip Normal のスタック数 3 以降と同じだった．出力された結果は 0.34822 の場合，図 5.11 に示すようになった．

### 5.5.2 セグメンテーションタスクの結果考察

実験では上述したように，Skip Normal ではスタック数が 1 から 3 の間は結果が改善したものの，Skip Down や Skip Feature では全ての場合で同じ評価値となっ

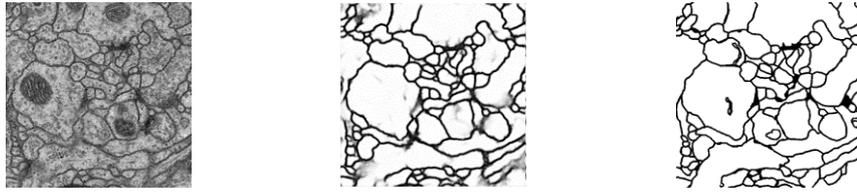


図 5.10: Input image. 図 5.11: Output image. 図 5.12: Ground truth.

図 5.13: Segmentation results.

た. Skip Down や Skip Feature では Skip Normal においてのスタック数 3 以降の評価値が, スタック数 1 において実現されているため, 効率的に欠けている情報を伝えられているものと思われる. ほぼ全てのスタック数において同一の評価値となっているのは, タスクが簡単すぎるためであると考え. 一方出力画像 (図 5.11) を ground truth (図 5.12) と比べると, 出力画像では境界がグラデーションになっており曖昧な部分 (図 5.14) や, 完全になくなっている部分 (図 5.15) がある.

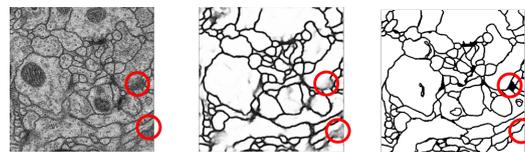


図 5.14: Difference in gradient part.

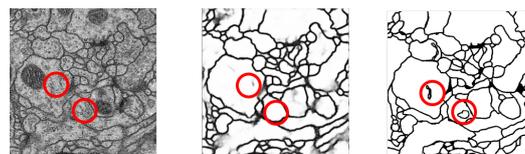


図 5.15: Disappearing objects.

これらは Skip Normal でも同じように消えているため, U-Net と skip connection の改良では, 対応できない部分であると考え.

### 5.5.3 CT 再構成問題

次に CT 再構成問題で効果を確認した.

モデルの構造を図 5.16 に示す.

スタック数は 1 から 10 まで変化させ 100 エポック学習させた. バリデーション

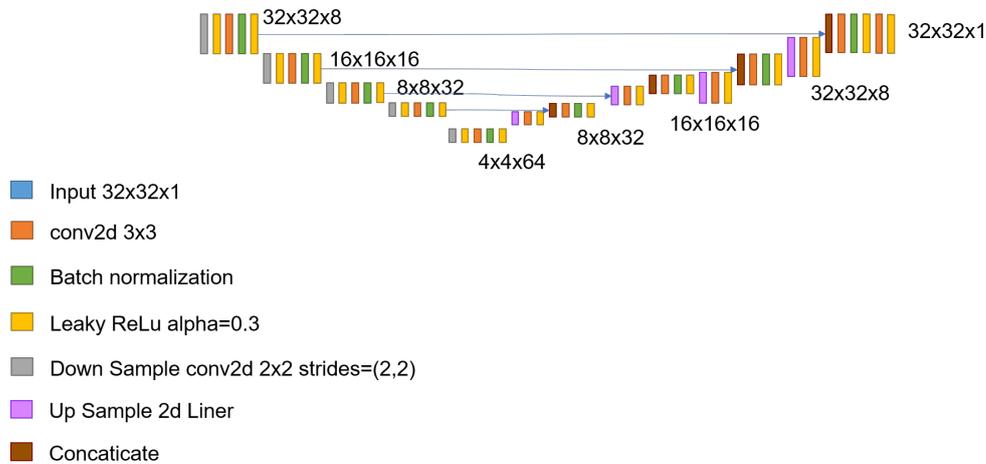


図 5.16: Structure of U-Net for CT reconstruction.

セットでの平均の PSNR を表 5.2 に示す.

表 5.2: CT reconstruction task stack results.

stack num	Skip Normal	Skip Down	Skip Feature
1	25.58	<b>25.95</b>	25.73
2	25.41	<b>27.33</b>	27.32
3	<b>26.65</b>	24.76	26.13
4	27.35	24.6	<b>27.86</b>
5	26.38	26.32	<b>27.18</b>
6	25.27	25.37	<b>25.71</b>
7	26.83	26.29	<b>27.00</b>
8	24.49	26.09	<b>27.65</b>
9	25.84	24.88	<b>26.68</b>
10	24.11	<b>24.20</b>	23.79

Skip Normal はスタック数が 3 の場合をのぞいて, Skip Down や Skip Feature より悪い結果となった. Skip Down はスタック数が 1 や 2 の時に良い結果となった. 特に, スタック数が 2 の時に最も良い結果になっており, Skip Down はスタック数が少ないときに良い結果になることがわかる. Skip Feature ではスタック数が 4 から 9 の場合に他の手法より良い結果となった. スタック数が多い場合は Skip Feature が有効であることがわかる. また, スタック数が 4 の場合 27.86db と Skip

Feature がすべての条件で最も良い評価結果となった。

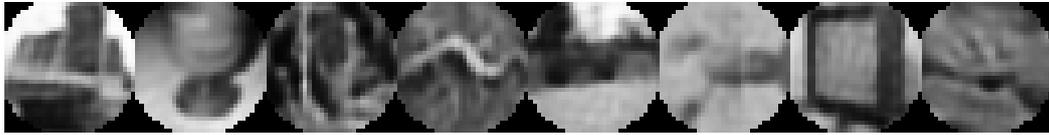


図 5.17: Skip Normal best result stack:4.

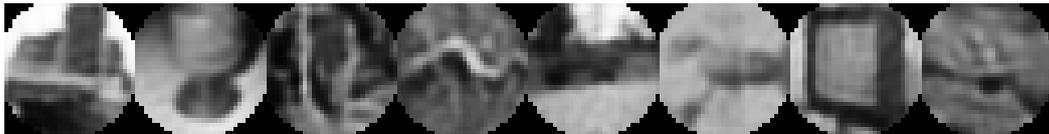


図 5.18: Skip Down best result stack:2.

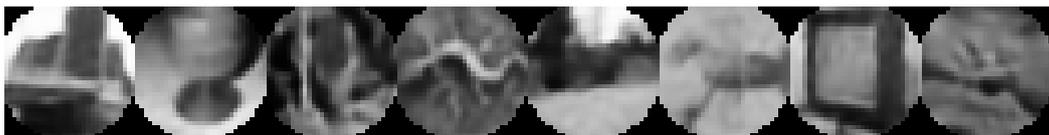


図 5.19: Skip Feature best result stack:4.

最後に図 5.17～図 5.19 にそれぞれのベストの結果の画像を示す。stack 数はそれぞれ Skip Normal で 4, Skip Down で 2, Skip Feature で 4 である。

#### 5.5.4 CT 再構成の考察

Skip Feature では Skip Normal や Skip Down と比べて、スタック数の増加による評価値の減少は少なく、ある程度スタック数の増加による評価値の減少に対応できていると思われる。一方で評価値のピークは Skip Normal では 4, Skip Down では 2, Skip Feature では 4 とほぼかわっていない。これは CT 再構成においては提案手法である Skip Feature は評価値は改善するものの、Stack 数の増加時の評価値の減少に対する効果は既存手法である Skip Normal と変わらないということである。

なお図 5.17～図 5.19 に示すように、視覚的には手法によって差はみられなかった。これはそれぞれのベストな時を比べたからで図 5.20 のように 24.6db の時と比べると、左から 5 番目の公園の画像で中央の塔が薄くなっているなど明らかな違いがわかる。



図 5.20: Skip Down result stack:4.

### 5.5.5 CT 再構成問題での skip connection の周波数解析

提案した Skip Down および Skip Feature が意図通り高周波を伝えているかを見るために、Stack 数 4 での skip connection での skip connection で伝わっている画像の周波数をフーリエ変換により解析した。ここで各 Skip connection はチャンネルがありそれぞれ違ったデータを伝播するが、今回は全体の傾向を見るために各 skip connection ごとの周波数成分のチャンネル全体の平均をとった。結果を図 5.21 から図 5.35 に示す。

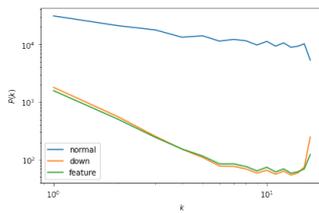


図 5.21: Layer 1.

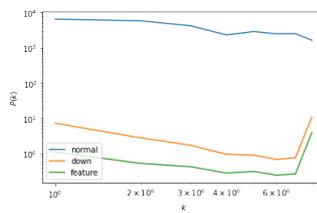


図 5.22: Layer 2.

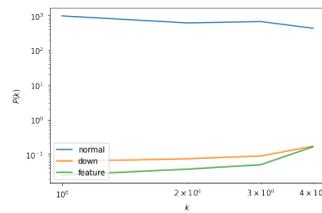


図 5.23: Layer 3.

図 5.24: Frequency of skip connections at stack 1.

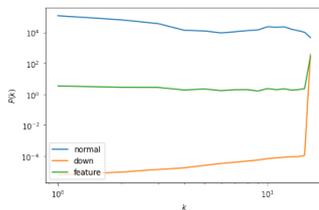


図 5.25: Layer 1.

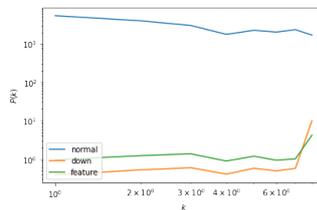


図 5.26: Layer 2.

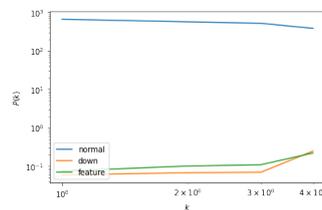


図 5.27: Layer 3.

図 5.28: Frequency of skip connections at stack 2.

グラフは両対数表時である。横軸は周波数であり縦軸はパワースペクトラムである。グラフ中の normal, down, feature はそれぞれ Skip Normal, Skip Down,

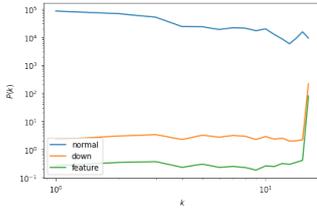


図 5.29: Layer 1.

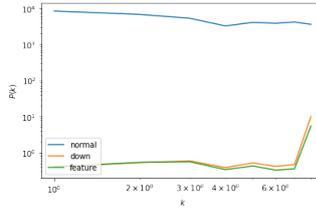


図 5.30: Layer 2.

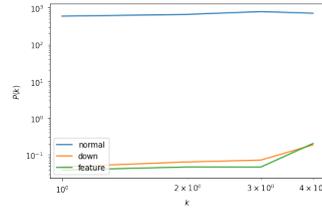


図 5.31: Layer 3.

図 5.32: Frequency of skip connections at stack 3.

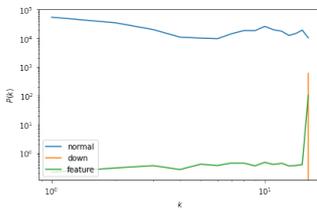


図 5.33: Layer 1.

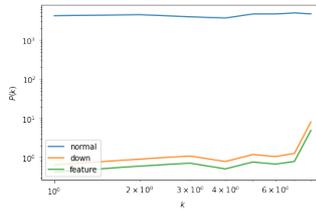


図 5.34: Layer 2.

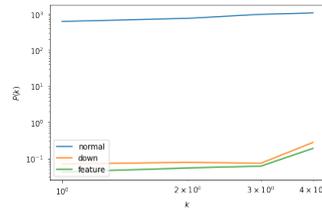


図 5.35: Layer 3.

図 5.36: Frequency of skip connections at stack 4.

Skip Feature の結果を示している。これらの結果を見ると、グラフの右側が跳ね上がっており、Skip Down および Skip Feature では高周波の成分が、他の成分に比べて強調されていることがわかり、概ね意図通りになっていることがわかる。Skip Normal は Skip Down や Skip Feature と比べてパワースペクトラムの値が大きくなっているが、これは Skip Down や Skip Feature が画像と拡大した画像というほぼ同じ情報差分を伝えており、Skip Normal が画像そのものの値を伝えているのに比べて、値が小さくなっているためである。

stack 数が 4 の場合 Skip Feature が 27.86db と一番良い結果になっていた。一方 Skip Down では 24.6db と Skip Normal の 27.35db とくらべ顕著に悪い状況になっている。Skip Down と Skip Feature の周波数解析の結果を見比べると、全体の傾向は似ている。このことから今回提案の手法は高周波を伝えやすくしており、Skip Feature の精度改善には貢献しているが、Skip Down の精度改善には貢献しておらずむしろ悪化させているといえる。細かくグラフを見ると Skip Feature は 2 スタック目の 1 レイヤ (図 5.25) において、Skip Down より低周波数成分が大きい。また、Skip Down は 4 スタック目の 1 レイヤ (図 5.33) において、低周波が

まったくなくなっている。これらの違いが Skip Feature と Skip Down の精度の違いを生み出しており、一概に高周波を skip connection で伝えることが精度改善に貢献するとは言えないことがわかる。Skip Down はスタック数が 2 の時一番良い結果を出しており、この結果と見比べることで、さらなる改良ポイントが見つかるものとする。

## 5.6 考察

今回実験した、2D EM segmentation [85] データセットにおけるセグメンテーションでは、既存手法 (Skip Normal) と提案手法では評価値に差がなかった。これはタスクが簡単すぎるためと考えられる。一方 Skip Normal がスタック数を 1 から 3 に増やすと、微小に結果が改善したが、Skip Down や Skip Feature で stack 数 1 から Skip Normal が stack 数 3 の時と同じ評価値となっていた。ここから、提案手法は効率的に情報を伝えられていることがうかがえ、ちょうど良い難しさのタスクでは既存の U-Net より良くなる可能性があると考えられる。

また、CT 再構成タスクでは stack 数に対する耐性はある程度できるものの大規模に stack すると評価値が悪化するのとは変わらなかった。これは stack 数に対する劣化の対処が今回の改良では不十分であるということである。伝搬する成分を分析し、何が stack 耐性につながったのか詳細に分析し改良していく必要がある。

さらに、既存研究では optic cap segmentation におけるベストの stack 数は 15 となっており、今回の 2D EM segmentation のベストの stack 数 3、CT 再構成のベストの stack 数 4 と大幅に違う、これはタスクに応じてベストな stack 数も変わってくるということを示唆しており、タスクと stack 数との関係を明らかにしていくことで、stack 数が増加しても精度が落ちない手法の提案につなげることができると考える。

## 5.7 まとめ

Stacked U-Net の Stack 耐性を向上するための手法として、Skip Down と Skip Feature を提案し、実験を行った。CT 再構成タスクでは Skip Feature においては既存手法である Skip Normal より良い評価値を得ることができた。また Skip

Down も浅いスタック数では良い結果を示した。

Skip Feature はスタック数を増加させても Skip Normal や Skip Down と比べて評価値の減少は少なく，スタック数が多い場合は Skip Feature が有効であることがわかる。一方でスタック数と評価値のピークの観点では提案手法は既存手法と変わりがなく，Stack 数が多すぎると評価値が悪くなるという Stacked U-Net 改良点を改善するまでには至らなかった。

既存研究ではスタック数が 15 の場合ベストの評価値をだしており，今回の結果と違う結果になっている。これはタスクやモデル構造により最良のスタック数が違うということを示唆している。ほかのさまざまなタスクに適用し，評価することが必要である。また，提案手法では高周波成分が伝わりやすくなることを狙っているが，実際に skip connection で伝わっているデータの周波数解析をすることで，実際に高周波成分が伝わりやすくなっていることが確認できた。これらの解析により，さらに手法を改善する方法を発見する必要がある。

## 第6章

# 結論

本研究では、CT再構成の課題である「高速な再構成の実現」「ノイズを減らす」「患者の負荷を減らす」(6.1:再掲)を実現するために既存手法なしにディープラーニングのみで、再構成を実現するモデルを提案した。医用画像を用いたディープラーニングの課題としては「学習データがすくない」という課題があるが(6.2:再掲)自然画のデータセットであるImageNetを学習に使用することで、この課題を解決した。提案したStacked U-Netのモデルは512個の検出器の360°回転64投影を想定したデータセットで、CPUで2.35s、GPUで0.11sで再構成することができた。提案手法は投影数64の場合画質はPSNRで27.93db、SSIMで0.886である。これらは既存手法より優れた結果であった。また、セグメンテーションタスクの学習済みモデルを用いた客観評価においても、投影数64の場合においてすべての既存手法に比べて優れた結果を示し、投影数512、256、128および64すべての場合においてディープラーニングを利用した既存手法より優れた結果を示した。

また、「学習データが少ない」という課題に対して新たなaugmentation手法であるInterlayer augmentationを提案した。Interlayer augmentationはモデルのレイヤ間に挿入しバッチ学習時にデータを増強するaugmentation手法である。従来のaugmentation手法はあらかじめ入力データを増強しておく方法であり、増強できるデータ数に限りがあるが、提案手法では挿入数を増やせば各々のレイヤでデータ増強を行うことになり、ベキ乗で学習データ量を増やすことが可能である。具体的なアルゴリズムとしてRBG, BGを提案し、CIFAR10およびCIFART100の分類タスクでVGG, ResNetおよびViTでスクラッチ学習した場合と比較し評価を

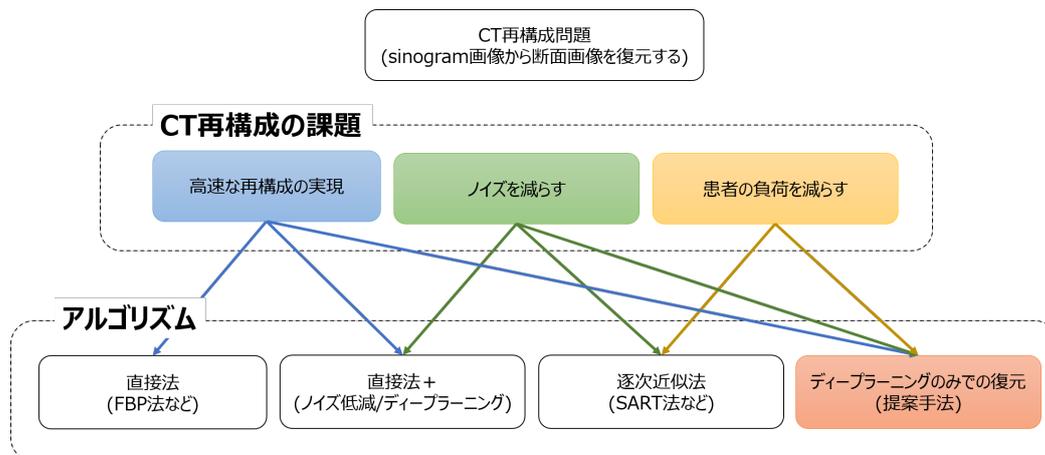


図 6.1: Issues CT reconstructions and algorithm correspondence.

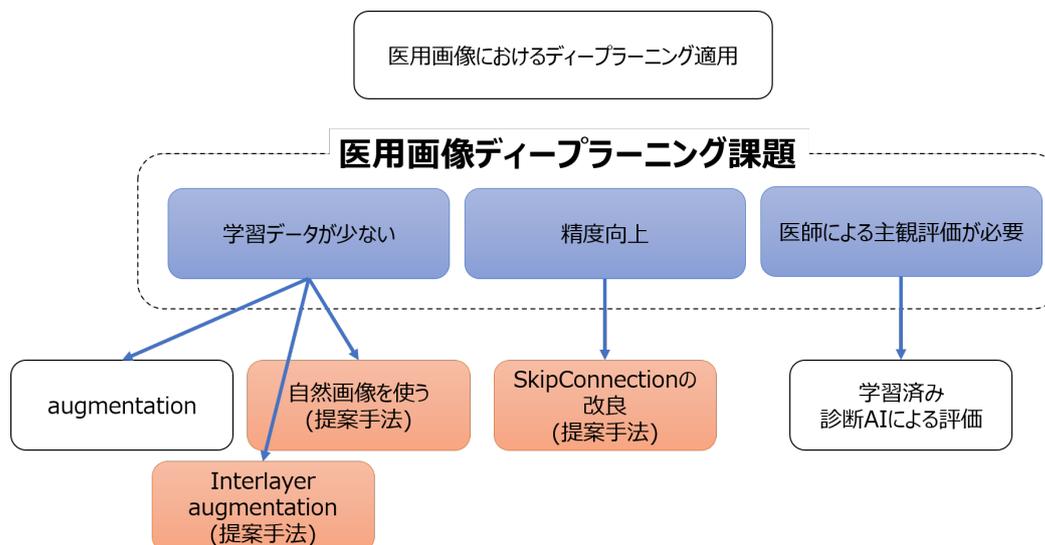


図 6.2: Issues deep learning for medical and correspondence.

行った. 結果として, CIFAR10 では RGB で平均 0.39%, BG で平均 0.27% の改善が見られ, CIFAR100 では RGB で平均 1.07%, BG で平均 0.30% の改善が見られた. 特に, 全てのケースにおいて, RGB はスクラッチ学習より優れた結果を示した.

提案した Stacked U-Net は画像の見た目としては, 詳細構造が失われているためさらに画質を良くする必要がある. 「精度向上」という課題に対して Stacked U-Net が Stack 数が多くなりすぎると精度が悪化することに着目し, U-Net の SkipConnection に対して高周波成分が伝わりやすい Skip Down および Skip

Feature を提案した。結果として、CT 再構成タスクにおいて既存手法である Skip Normal と比べて SkipFeature は PSNR で 0.51db 改善した。また、SkipDown は stack 数 2 の場合既存手法より、1.98db 改善した。SkipFeature は Stack 数が 4 の時に最大性能を示したものの、スタック数が大きくなっても性能の低下度合いは Skip Normal と比べて少なかった。

## 6.1 今後の展望

上述したように、CT 再構成が行えるモデルを提案し、その弱点を補う方法を提案したが、現状それらの手法を統合できていない。

512 × 512 の CT 再構成問題に対して精度改善のための方法を以下の方法を適用し、血管などの詳細な構造が消えないかつ高速な再構成方法を確立するのが今後の課題である。

医用画像の学習データの少なさに対しては、

- 既存の data augmentation であるボケの付加、MixUpなどを適用する、
- Interlayer augmentation を適用する。
- data augmentation と Interlayer augmentation を同時に適用する。

などが考えられる。

また精度向上に対しては

- Stacked U-Net に対して Skip Feature を適用する。
- 自然画像を用いた学習に対して、医用画像を少し混ぜる。
- 低投影の画像も混ぜる。
- Attention や GAN など最新のモデルに対応させる。

などが考えられる。

CT 再構成画像は最終的には診断に使われる。そのため医者による主観評価や学習が十分にされた AI 診断用モデルによる客観評価を行うことが重要である。AI 診断という面では Interlayer augmentation はクラス分類タスクに効果がある。CT

---

再構成から診断まで一貫したモデルを作成し、Interlayer augmentation により、データの不足を補うことができれば、システムとして CT 画像が AI 診断に有効に機能するモデルを提案できると考える。

## 発表実績

### 1. 査読付き学術雑誌論文 (1 件)

- [1] Satoru Mizusawa, Yuichi Sei, Ryohei Orihara and Akihiko Ohsuga: Computed Tomography Image Reconstruction using Stacked U-Net, Computerized Medical Imaging and Graphics, Vol.90, No.101920, pp.1-10, Elsevier, 2021

### 2. 査読付き国際会議発表論文 (2 件)

- [1] Satoru Mizusawa and Yuichi Sei: Interlayer augmentation in a classification task, 4th IEEE International Conference on Computing, Electronics & Communications Engineering (IEEE iCCECE), pp.59-64, 2021
- [2] Satoru Mizusawa, Yuichi Sei, and Akihiko Ohsuga: Stack performance improvement of stacked U-Net, 9th IEEE Joint International Information Technology and Artificial Intelligence Conference (IEEE ITAIC), pp.2055-2060, 2020

### 3. 査読付き国内会議発表論文 (1 件)

- [1] 水澤悟, 清雄一, 折原良平, 大須賀昭彦: Stacked U-Net による CT 画像再構成, 合同エージェントワークショップ&シンポジウム (JAWS), pp.256-262, 2019

### 4. 国内会議発表論文 (0 件)

### 5. 受賞 (1 件)

- [1] Satoru Mizusawa and Yuichi Sei: Interlayer augmentation in a classification task, 4th IEEE International Conference on Computing, Electronics & Communications Engineering (IEEE iC-

CECE), 2021, **Best Paper Award**

## 謝辞

本論分は、執筆者がトップエッセイの修了製作をもとに電気通信大学博士後期課程入学以来発展させたテーマをまとめたものです。

主任指導教員として、研究の方向性、論文構成など研究活動全般にわたりご指導をいただいた電気通信大学情報理工学系研究科 大須賀昭彦教授，清雄一准教授，折原良平客員教授に深く感謝申し上げます

本論分をまとめるにあたり、論文審査をしていただいた電気通信大学大学院情報理工学研究科 庄野逸教授，柳井啓司教授，高橋裕樹准教授に深く感謝いたします。

最後に、博士課程での研究活動や社会人学生生活を支えてくれた妻や子供たちに心から感謝します。

上記以外にも、執筆者の研究活動を支えていただいたすべての方々に感謝いたします。ありがとうございました。

## 参考文献

- [1] Chandradeep Bhatt, Indrajeet Kumar, V. Vijayakumar, Kamred Udham Singh, and Abhishek Kumar. The state of the art of deep learning models in medical science and their challenges. *Multimedia Systems 2020 27:4*, Vol. 27, No. 4, pp. 599–613, sep 2020.
- [2] H. Malcolm Hudson and Richard S. Larkin. Accelerated Image Reconstruction Using Ordered Subsets of Projection Data. *IEEE Transactions on Medical Imaging*, Vol. 13, No. 4, pp. 601–609, 1994.
- [3] Donghwan Kim, Sathish Ramani, and Jeffrey A. Fessler. Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction. *IEEE Transactions on Medical Imaging*, Vol. 34, No. 1, pp. 167–178, jan 2015.
- [4] Lucas L. Geyer, U. Joseph Schoepf, Felix G. Meinel, John W. Nance, Gorka Bastarrika, Jonathon A. Leipsic, Narinder S. Paul, Marco Rengo, Andrea Laghi, and Carlo N. De Cecco. State of the Art: Iterative CT reconstruction techniques1. *Radiology*, Vol. 276, No. 2, pp. 339–357, aug 2015.
- [5] J Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Akad. Wiss.*, Vol. 69, pp. 262–277, 1917.
- [6] Ohnishi Hideo. Introduction to Medical Technology : Analytical Image Reconstruction (lecture in journal). *Nuclear medicine subspecialty journal*, Vol. 57, pp. 37–43, oct 2008.
- [7] Chiou Shann Fuh and Shih Schön Lin. The Fourier slice theorem for

- range data reconstruction. *Image and Vision Computing*, Vol. 16, No. 9-10, pp. 689–701, jul 1998.
- [8] Joachim Hornegger, Andreas Maier, and Markus Kowarschik. *CT Image Reconstruction Basics*. Wolters Kluwer, 2016.
- [9] Jeffrey Fessler. Statistical Image Reconstruction Methods for Transmission Tomography. In *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, Vol. PM80, pp. 1–71. SPIE, jun 2000.
- [10] 篠原広行, 桑山潤, 小川瓦, 橋本雄幸. 断層映像法の基礎第 32 回 : ML-EM 法と OS-EM 法. 断層映像研究会雑誌, Vol. 37, No. 3, pp. 1–8, 2011.
- [11] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, Vol. 1, No. 4, pp. 541–551, dec 1989.
- [12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, Vol. 60, No. 6, pp. 84–90, jun 2017.
- [13] Stamatios Lefkimmiatis. Non-local color image denoising with convolutional neural networks. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Vol. 2017-Janua, pp. 5882–5891, 2017.
- [14] Wei Sheng Lai, Jia Bin Huang, Narendra Ahuja, and Ming Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Vol. 2017-Janua, pp. 5835–5843. Institute of Electrical and Electronics Engineers Inc., nov 2017.
- [15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net : Convolutional Networks for Biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015.
- [16] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Deep ADMM-Net

- 
- for compressive sensing MRI. In *Advances in Neural Information Processing Systems 29*, pp. 10–18. Curran Associates, Inc., 2016.
- [17] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, Vol. 3, No. 1, pp. 1–122, 2010.
- [18] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiang Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, and David Firmin. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging*, Vol. 37, No. 6, pp. 1310–1321, jun 2018.
- [19] Mehdi Mirza and Simon Osindero. Conditional Generative Adversarial Nets. nov 2014.
- [20] Jin Zhu, Guang Yang, and Pietro Lio. Lesion focused super-resolution. In Elsa D. Angelini and Bennett A. Landman, editors, *Medical Imaging 2019: Image Processing*, Vol. 10949, p. 56. SPIE, mar 2019.
- [21] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9906 LNCS, pp. 694–711, 2016.
- [22] Pennell; Zhu, Jin; Guang, Yang; Pedro, Ferreira; Andrew, Scott, Sonia, Nielles-Vallespin; Jennifer, Keegan; Dudley and Firmin; Pietro, Lio; David. A ROI Focused Multi-Scale Super-Resolution Method for the Diffusion Tensor Cardiac Magnetic Resonance. In *In the International Society for Magnetic Resonance in Medicine 27th Annual Meeting*, p. 1, 2019.
- [23] Jin Zhu, Guang Yang, and Pietro Lio. How Can We Make GAN Perform Better in Single Medical Image Super-Resolution? A Lesion Fo-

- cused Multi-Scale Approach. In *Proceedings - International Symposium on Biomedical Imaging*, Vol. 2019-April, pp. 1669–1673. IEEE Computer Society, jan 2019.
- [24] Simiao Yu, Hao Dong, Guang Yang, Greg Slabaugh, Pier Luigi Dragotti, Xujiang Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, David Firmin, and Yike Guo. Deep De-Aliasing for Fast Compressive Sensing MRI. *arXiv*, may 2017.
- [25] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael Unser. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, Vol. 26, No. 9, pp. 4509–4522, sep 2017.
- [26] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2016-Decem, pp. 1646–1654. IEEE Computer Society, dec 2016.
- [27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2016-Decem, pp. 770–778. Microsoft, 2016.
- [28] Foivos I. Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 162, pp. 94–114, apr 2019.
- [29] Simon Jegou, Michal Drozdal, David Vazquez, Adriana Romero, and Yoshua Bengio. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Vol. 2017-July, pp. 1175–1183. IEEE Computer Society, aug 2017.

- 
- [30] Artem Sevastopolsky, Stepan Drapak, Konstantin Kiselev, Blake M. Snyder, Jeremy D. Keenan, and Anastasia Georgievskaya. Stack-U-Net: refinement network for improved optic disc and cup image segmentation. *Medical Imaging 2019: Image Processing*, Vol. 10949, p. 78, mar 2019.
- [31] Sohil Shah, Pallabi Ghosh, Larry S Davis, and Tom Goldstein. Stacked U-Nets: A No-Frills Approach to Natural Image Segmentation. apr 2018.
- [32] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612, apr 2004.
- [33] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N. Chiang, Zhihao Wu, and Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, Vol. 63, p. 101693, jul 2020.
- [34] Patrick Ferdinand Christ, Mohamed Ezzeldin A. Elshaer, Florian Ettl, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D’Anastasi, Wieland H. Sommer, Seyed-Ahmad Ahmadi, and Bjoern H. Menze. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9901 LNCS, pp. 415–423, oct 2016.
- [35] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve Restricted Boltzmann machines. In *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 807–814, 2010.
- [36] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *ICML Workshop*

- on Deep Learning for Audio, Speech and Language Processing*, 2013.
- [37] Matthew D. Zeiler, Graham W. Taylor, and Rob Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2018–2025, 2011.
- [38] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *32nd International Conference on Machine Learning, ICML 2015*, Vol. 1, pp. 448–456, 2015.
- [39] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
- [40] Liangchen Luo, Yuanhao Xiong, Yan Liu, and Xu Sun. Adaptive gradient methods with dynamic bound of learning rate. In *7th International Conference on Learning Representations, ICLR 2019*, sep 2019.
- [41] Alex Krizhevsky and G Hinton. Learning multiple layers of features from tiny images. 2009.
- [42] Shanah Kirk, Y. Lee, C. Roche, E. Bonaccio, J. Filippini, and R. Jarosz. Radiology Data from The Cancer Genome Atlas Thyroid Cancer [TCGA-THCA] collection, 2016.
- [43] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, Vol. 26, No. 6, pp. 1045–1057, dec 2013.
- [44] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. *IEEE Nuclear Science Symposium and Medical Imaging Conference*, Vol. 3, pp. 1562–1566, 1999.
- [45] Stefan der Walt, Johannes L Schönberger, Juan Nunez-Iglesias,

- 
- François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillard, and Tony Yu. scikit-image: image processing in Python. *PeerJ*, Vol. 2, p. e453, 2014.
- [46] Barbero Alvaro, Suvrit, and Sra. Modular Proximal Optimization for Multidimensional Total-Variation Regularization. *Journal of Machine Learning Research*, Vol. 19, No. 56, pp. 1–82, 2018.
- [47] Zhicheng Zhang. zzc623/DD\_Net: code for the paper entitled "A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution".
- [48] Maximilian Seitzer, Guang Yang, Jo Schlemper, Ozan Oktay, Tobias Würfl, Vincent Christlein, Tom Wong, Raad Mohiaddin, David Firmin, Jennifer Keegan, Daniel Rueckert, and Andreas Maier. Adversarial and Perceptual Refinement for Compressed Sensing MRI Reconstruction. In Alejandro F Frangi, Julia A Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 232–240, Cham, 2018. Springer International Publishing.
- [49] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. *UNet++: A Nested U-Net Architecture for Medical Image Segmentation: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*, Vol. 11045, pp. 3–11. 2018.
- [50] MedSeg; and Tomas; Jenssen, Håvard Bjørke; Sakinis. MedSeg Covid Dataset 1, 2021.
- [51] Vítězsav Vlček. Computation of filtered back projection on graphics cards. Vol. 4, pp. 34–39. WSEAS Korfu, 2005.
- [52] Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, and Synho Do. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? In *undefined*. ICLR 2016, 2015.

- [53] Rosa L. Figueroa, Qing Zeng-Treitler, Sasikiran Kandula, and Long H. Ngo. Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making*, Vol. 12, No. 1, p. 8, dec 2012.
- [54] Yifeng Guo, Chengjia Wang, Heye Zhang, and Guang Yang. Deep Attentive Wasserstein Generative Adversarial Networks for MRI Reconstruction with Recurrent Context-Awareness. *arXiv*, p. 2006.12915, jun 2020.
- [55] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, sep 2015.
- [56] Jo Schlemper, Guang Yang, Pedro Ferreira, Andrew Scott, Laura Ann McGill, Zohya Khaliq, Margarita Gorodezky, Malte Roehl, Jennifer Keegan, Dudley Pennell, David Firmin, and Daniel Rueckert. Stochastic deep compressive sensing for the reconstruction of diffusion tensor cardiac MRI. In *The 21st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2018)*, Vol. 11070 LNCS, pp. 295–303. Springer Verlag, sep 2018.
- [57] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Weinberger. Deep Networks with Stochastic Depth. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9908 LNCS, pp. 646–661, mar 2016.
- [58] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K. Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low Dose CT Image Denoising Using a Generative Adversarial Network with Wasserstein Distance and Perceptual Loss. *IEEE transactions on medical imaging*, Vol. 37, No. 6, p. 1348, jun 2018.

- 
- [59] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*, 2020.
- [60] Nazia Hameed, Antesar Shabut, Fozia Hameed, Silvia Cirstea, Alamgir Hossain, and Sorrel Harriet. "Mobile-based Skin Lesions Classification Using Convolution Neural Network", *Annals of Emerging Technologies in Computing (AETiC)*, Print Mobile-based Skin Lesions Classification Using Convolution Neural Network. Vol. 4, No. 2, pp. 26–37, 2020.
- [61] Khai Dinh Lai, Thuy Thanh Nguyen, and Thai Hoang Le. "Detection of lung nodules on CT images based on the Convolutional Neural Network with Attention Mechanism", *Annals of Emerging Technologies in Computing (AETiC)*, Print Detection of Lung Nodules on CT Images based on the Convolutional Neural Network with . Vol. 5, No. 2, pp. 77–89, 2021.
- [62] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In *Proceedings of the IEEE International Conference on Computer Vision*, Vol. 2017-Octob, pp. 843–852. Institute of Electrical and Electronics Engineers Inc., dec 2017.
- [63] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. oct 2020.
- [64] Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, Vol. 6, No. 1, pp. 1–48, dec 2019.
- [65] Satoru Mizusawa, Yuichi Sei, Ryohei Orihara, and Akihiko Ohsuga. Computed tomography image reconstruction using stacked U-Net. *Computerized Medical Imaging and Graphics*, Vol. 90, p. 101920, jun

- 2021.
- [66] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. *arXiv*, dec 2020.
- [67] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pp. 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.
- [68] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, Vol. 15, No. 56, pp. 1929–1958, 2014.
- [69] Kazuyuki Hara, Daisuke Saitoh, and Hayaru Shouno. Analysis of Dropout Learning Regarded as Ensemble Learning. In Alessandro E P Villa, Paolo Masulli, and Antonio Javier Pons Rivero, editors, *Artificial Neural Networks and Machine Learning – ICANN 2016*, pp. 72–79, Cham, 2016. Springer International Publishing.
- [70] Cairong Zhao, Kang Chen, Zhihua Wei, Yipeng Chen, Duoqian Miao, and Wei Wang. Multilevel triplet deep learning model for person re-identification. *Pattern Recognition Letters*, Vol. 117, pp. 161–168, 2019.
- [71] Bikash Santra, Angshuman Paul, and Dipti Prasad Mukherjee. Deterministic dropout for deep neural networks using composite random forest. *Pattern Recognition Letters*, Vol. 131, pp. 205–212, 2020.
- [72] Sangdoon Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. *Proceedings of the IEEE International Conference on Computer Vision*, Vol. 2019-Octob, pp. 6022–6031, may 2019.
- [73] Terrance De Vries and Graham W. Taylor. Dataset augmentation in

- 
- feature space. In *Proceedings of the international conference on machine learning, workshop track*. arXiv, feb 2017.
- [74] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, sep 2015.
- [75] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2016-Decem, pp. 770–778. IEEE Computer Society, dec 2016.
- [76] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, pp. 448–456. JMLR.org, 2015.
- [77] Yuxin Wu and Kaiming He. Group Normalization. *International Journal of Computer Vision*, Vol. 128, No. 3, pp. 742–755, mar 2020.
- [78] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond Empirical Risk Minimization. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, oct 2017.
- [79] Terrance DeVries and Graham W. Taylor. Improved Regularization of Convolutional Neural Networks with Cutout. aug 2017.
- [80] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. CIFAR-100 (Canadian Institute for Advanced Research).
- [81] Jeremy Howard. Imagenette(fast.ai).
- [82] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition*,

- CVPR 2017*, Vol. 2017-Janua, pp. 2261–2269. Institute of Electrical and Electronics Engineers Inc., nov 2017.
- [83] Yuanyi Zeng, Xiaoyu Chen, Yi Zhang, Lianfa Bai, and Jing Han. Dense-U-Net: densely connected convolutional network for semantic segmentation with a small number of samples. In Hui Yu, Yifei Pu, Chunming Li, and Zhigeng Pan, editors, *Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*, Vol. 11069, p. 159. SPIE, may 2019.
- [84] Yoseob Han and Jong Chul Ye. Framing U-Net via Deep Convolutional Framelets: Application to Sparse-view CT. *IEEE Transactions on Medical Imaging*, Vol. 37, No. 6, pp. 1418–1429, aug 2017.
- [85] Albert Cardona, Volker Hartenstein, Stephan Saalfeld, Stephan Preibisch, Benjamin Schmid, Anchi Cheng, Jim Pulokas, and Pavel Tomancak. An integrated micro- and macroarchitectural analysis of the *Drosophila* brain by computer-assisted serial section electron microscopy. *PLoS Biology*, Vol. 8, No. 10, 2010.
- [86] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H Wallach, H Larochelle, A Beygelzimer, F d\textquotesingle Alché-Buc, E Fox, and R Garnett, editors, *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.
- [87] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9351, pp. 234–

241. Springer Verlag, 2015.

## 著者略歴



**Satoru Mizusawa** The University of Tokyo, Graduate School of Information Science and Technology 2009. Admitted to the Ph.D. program in Informatics at the University of Electro-Communications in 2018. Engaged in research on image reconstruction using deep learningss