

改良型 U-Net を用いた低線量 CT 画像の画質改善法 A Method for Denoising from Low-dose CT Images Based on Improved U-Net

澤田 修志* 村上 誠一† 李 光旭‡ 神谷 亨*

1. はじめに

CT(Computed Tomography)検査は、放射線を用い体内の断面写真を撮影する医用画像診断法の一つで、がんや腫瘍の検出、治療効果の判定に広く用いられている。日本は、人口 100 万人当たりの CT 装置設置台数が先進国の中で最も多く、他の主要国と比較しても放射線検査が容易に受けられる環境が整備されている[1]。こうした放射線を用いた検査が多大な恩恵をもたらしている一方で、検査時に放射線を浴びることによる医療被ばくが問題視されている。日本人一人当たりの年間医療被ばく線量は約 2.6 ミリシーベルトで、これは世界平均の四倍以上にあたるとされている。医療被ばくについては、直接的な健康への影響は報告されていないものの、確率的影響と呼ばれる年月が経過してがんや白血病を引き起こすリスクが増大する可能性があること示唆されている[2]。この問題を解決するため、CT 画像撮影時の放射線量を下げることが解決策の一つとして提案されている。しかし、放射線量を減少させると、図 1 に示すように、画像のノイズが増加し、重要な画像特徴が失われる問題が生じる。このため、病変部の見落としにつながり、結果的に診断の信頼性が損なわれる可能性がある。

本稿では、上記の問題解決のため、画像の鮮鋭度を保ったままノイズ除去が可能な画像処理手法を開発する。畳み込みニューラルネットワーク(CNN:Convolutional Neural Network)による画像処理は、現在多くの手法が提案され、従来の性能を超えることが報告されている[3]。そこで、CNNを活用し、低線量に起因するノイズの除去を行う手法を提案する。具体的には、学習の際にクリーンな画像を使用しない自己教師あり学習の Noise2Noise[4]のメソッドを用いた独自の CNN モデルを構築し、子豚の低線量 CT 画像データセット[5]を用いた実験結果について述べる。

2. 手法

本章では、Noise2Noise の学習法を用い、低線量 CT 画像から低線量に起因するノイズを除去し、画質改善を行う手法について述べる。ネットワークアーキテクチャについては、エンコーダデコーダ構造を持つ完全畳み込みネットワークの U-Net[6]をベースモデルとし、EfficientNetV2[7]で提案された Fused-MBConv と、画像特徴を強調する CBAM[8]と Feature Pyramid Attention[9]の 2 種類の Attention 機構を加味した新しい CNN モデルを構築する。

2.1 Noise2Noise

本稿では、ベースモデルとして Noise2Noise の自己教師あり学習によって学習を行う。一般に、ノイズ除去の学習はノイズなしの画像とノイズありの画像をペアとして学習を行う。その一方で、Noise2Noise は別のノイズ分布を持つ画像を複数枚用いて学習を行う点が特徴として挙げられる。

* 九州工業大学 Kyushu Institute of Technology

† 純真学園大学 Junshin Gakuen University

‡ 天津工業大学 Tiangong University

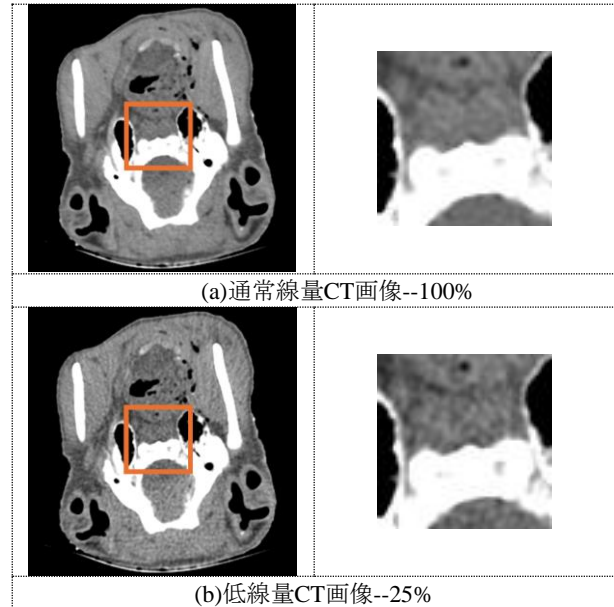


図 1 低線量 CT 画像の例：右側は拡大画像

Noise2Noise では、想定されるランダムノイズが元の画像との相関を持たず、平均が 0 であるという条件下であれば、ターゲットにノイズが含まれていても学習が可能であるとされている。特に、CT や MRI といった元々ノイズのない画像やデータを入手することが難しい条件下では、有効性や実用性が高いとして注目されている。一般に、ノイズ除去タスクにおける学習は式(1)に示すような、 \hat{x}_i で表すノイズ画像と、 y_i で表す正解画像をペアとして学習を行い、損失 L を最小化させるフィルタ関数 f_θ を求めることで学習を進める。一方で、Noise2Noise は以下の式(2)に示すような、ノイズ画像 \hat{x}_i と別のノイズ画像 \hat{y}_i のペアで学習を行う。

$$\operatorname{argmin}_{\theta} \sum_i L(f_{\theta}(\hat{x}_i), y_i) \quad (1)$$

$$\operatorname{argmin}_{\theta} \sum_i L(f_{\theta}(\hat{x}_i), \hat{y}_i) \quad (2)$$

一般に、低解像度画像 x に対応する高解像度画像 y は一意に定められず、高解像度画像 y は複数存在する。そこで、学習されたフィルタ関数 f_θ は、高解像度画像 y の平均値を出力するような傾向が見られ、その結果、ぼかし処理を生成する。ここで、 y の平均値を出力するフィルタ関数 f_θ が獲得できるとわかっている場合、そのフィルタ関数 f_θ が得られる条件下で、 y を \hat{y} に置き換えることが可能となる。この方法を用いることにより、クリーンなデータの分布やノイズ分布等を必要とせず、ノイズデータのみを収集することによりノイズ低減処理が可能となる。

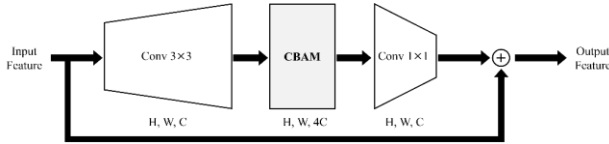


図 2 構築した Fused MB-Conv の構造

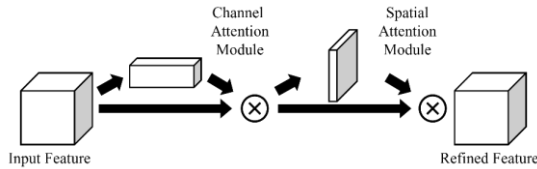


図 3 CBAM の構造

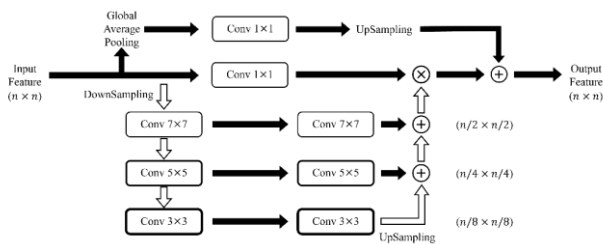


図 4 FPA の構造

2.2 U-Net

ノイズ除去を学習するネットワークアーキテクチャの具体的なベースモデルとしては、医用画像処理分野におけるセマンティックセグメンテーションで有効性を示した U-Net[6]を用いる。U-Net は全結合層を使用せず、線形層が畳み込み層のみで構成される完全畳み込みネットワークの一つで、解像度を縮小しつつ画像特徴を抽出するエンコーダと、獲得した特徴を受け取り、画像サイズを復元するデコーダ、エンコーダの特徴マップをデコーダの特徴マップに連結させるスキップ接続の 3 つから構成される。これにより、特徴抽出の際に失われる位置情報や細かい特徴を直接伝播させることにより、勾配を保ちつつ最適化を行うことが可能となる。また、ノイズ成分も特徴として獲得することを防ぎ、少ない特徴量でノイズ除去効果を最大限にすべく、セグメンテーションタスクに使用される U-Net の層数と特徴量を変更したモデルを用いる。

2.3 Fused-MBConv

Fused-MBConv[7]は、EfficientNetV2 アーキテクチャに導入された畳み込みであり、depthwise convolution と pointwise convolution からなる MBConv[10]の改良版として提案されている。解像度が大きい画像や浅い層で学習が遅いといった MBConv の問題点に対し、depthwise convolution を単一の通常畳み込みに変更することにより、層の深さに対してロバストな畳み込みを実現している。また、Fused-MBConv に内蔵される SEBlock(Squeeze-and-Excitation Block)[11]と呼ばれるチャンネル方向の特徴を強調するモジュールを、後述する CBAM(Convolutional Block Attention

Module) [8]に変更することで、チャンネルと空間の別次元に沿った特徴の強調を図る。本稿で構築した Fused-MBConv の構造を図 2 に示す。

2.4 Attention 機構

Attention とは、入力された特徴マップに対し、どこに注目するかを動的に特定する仕組みで、自然言語処理や画像処理分野において広く用いられる手法の一つである。本稿では、先述した CBAM(Convolutional Block Attention Module) と、複数スケールの特徴を考慮したピラミッドアテンションを行う Feature Pyramid Attention の導入により、特徴マップの補完を図る。

2.4.1 CBAM

CBAM の構造を図 3 に示す。CBAM は、チャンネル方向の Attention である Channel Attention と、空間方向の Attention である Spatial Attention の二つの Attention 機構を持ち、別次元に沿って重みづけされた特徴マップを推定することにより、特徴マップの洗練を行う。以下の式(3)に CBAM により出力される特徴マップ F_A を示す。

$$F_A = M_S(M_C(F) \otimes F) C(M_C(F) \otimes F) \quad (3)$$

ここで、 F は入力される特徴マップの情報、 M_C, M_S はそれぞれチャンネル、空間方向のアテンションマップを示す。ここで、Channel Attention は特徴チャンネル間の依存関係を用いることとなるため、周波数成分フィルタリングのような作用を発揮し、周波数の違いが重要であるデノイズタスクにおいて有用であるとされている。また、Spatial Attention については、画像中の何がどこにあるのかを強調するようなマスクを作成するため、重要なエッジを含むノイズ以外の要素を強調することを目的として導入している。

2.4.2 Feature Pyramid Attention

Feature Pyramid Attention(FPA) [9]の構造を図 4 に示す。FPA は Pyramid Attention Network で提案されたアテンションモジュールであり、ピクセルレベルで正確なアテンションを行うことが可能となる。FPA はそれぞれ $7 \times 7, 5 \times 5, 3 \times 3$ の畳み込みをピラミッド構造で段階的に用いることで異なるスケールの特徴マップを獲得する。その後、得られた特徴マップを段階的に足し合わせ統合することにより、ピラミッドアテンションを作成する。そして、元の特徴マップにピクセル単位でピラミッドアテンションを乗算し、大局的な特徴を考慮した特徴マップと足し合わせることで、複数のスケール情報を加味したアテンションを行う。本稿では、受容野を拡大することによるエッジ保持性能向上のため、U-Net のボトルネック部に FPA を導入する。

2.5 提案モデル

本稿で提案するモデルを図 5 に示す。デノイズタスク用に層数と最大チャンネル数を改良した U-Net に対し、エンコーダデコーダ共に畳み込み層を Fused-MBConv に変更した。加えて、Fused-MBConv の SEBlock を CBAM に変更し、U-Net のボトルネックに Feature Pyramid Attention を導入した改良型 U-Net を提案モデルとし、エッジ情報を保持しつつ平滑化を最大限に発揮することを図る。また、学習に際する損失関数は MSE、最適化アルゴリズムには Adam を使用し、学習率 0.0003、 $\beta_1 = 0.9, \beta_2 = 0.999$ とし、エポック数 1000 で学習を行った。

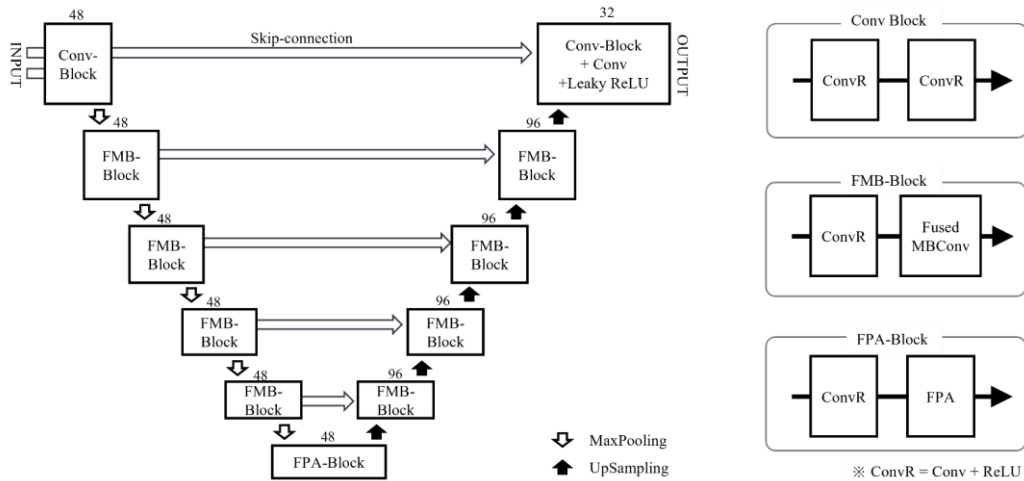


図 5 提案するネットワーク

表 1 CT 撮影条件

	通常線量 100%	低線量 25%
ガントリ	1	
回転時間(s)	0.625	
スライス厚(mm)	100	
管電位(kV)	300	75
管電流(mA)	30.83	7.71
CTDI _{vol} (mGy)	943.24	235.81
DLP(mGy-cm)	14.14	3.54
有効線量(mSv)	512×512	512×512
画像サイズ(px)		

表 2 実験結果

	平均 PSNR
	低線量 25%
元画像	30.2503
Model(1)	30.3381
Model(2)	30.2955
提案手法	30.4294

3. 実験と結果

本稿では、子豚の低線量 CT 画像(通常線量に対して線量が 25%の CT 画像)906 枚に対し、Noise2Noise にならった自己教師あり学習を行った。使用した CT 画像データセットにおける撮影条件を表 1 に示す。学習データについては、通常線量画像に対し、標準偏差 $\sigma = 40$ の正規分布に従ったガウシアンノイズを付与したものを使用した。

3.1 評価指標

画質改善の評価指標として、本稿では Peak Signal to Noise Ratio(PSNR)を使用した。PSNR の定義式を以下の式(4)に示す。PSNR は 2 枚の画像の類似度を測る指標であり、値が大きいほど高画質と評価される。ここで MSE(Mean Squared Error)は平均二乗誤差を表し、2 枚の画像の画素ごとの差分の二乗平均を表す。また、MAX は画像がとりうる最大の画素値を表す。本稿では、元の通常線量画像と、低線量画像や低線量画像に画質改善処理を適用した画像等と比較することで、提案手法の有用性を検討した。

$$PSNR = 10 \log_{10} \frac{MAX^2}{MSE} \quad (4)$$

3.2 実験結果

実験では、低線量 25%の CT 画像 906 枚をテストデータとして用いた。比較モデルとしては、(1)ベースモデルであるデノイズ用に改良した U-Net、(2)U-Net に SwinTransformer を搭載し、局所性・非局所性を強化した SCUNet[12]、本稿の提案手法である(3)改良型 U-Net である。

通常線量と低線量 CT 画像とを比較した平均 PSNR を表 2、ならびに低線量画像、各モデルを適用した出力画像を図 6 に示す。同図において、画像全体を比較した PSNR と注視した矩形領域についての PSNR を示す。

本実験により、量み込み層の変更と 2 種類の Attention 機構の導入により、他モデルと比較して高い平均 PSNR を獲得した。ノイズ除去タスクで高い精度を記録する SCUNet と比較しても、特徴量を抑えつつ、骨領域や臓器領域等の注視領域に特化した学習を行うことにより、視認性の向上につながり、重要なエッジ部位を保持したノイズ除去が可能となった。

4. おわりに

本稿では、Noise2Noise の学習を用いた改良型 U-Net による低線量 CT 画像のノイズ除去を試みた。子豚の低線量 25%の CT 画像データセットに対して実験を行った結果、平均 PSNR は 30.4294 となり、提案手法による画質改善が実現できた。さらなる精度向上を図るべく、モデルの改良や、異常組織に特化したノイズ除去の実現、さまざまな低線量に対する検証が今後の課題として挙げられる。また、通常線量画像では確認できる細かな脂肪組織部位が、画像の平滑化とともに消失しているため、学習の際に加えるノイズ分布の再検討を実施し、注視領域に対してノイズ除去とエッジ鮮鋭化のバランスを加味した学習が必要であると考察する。加えて、臨床現場での実用性を考慮し、より軽量なメソッドでの学習や、未知データに対する頑健性や即応性に対する検証についても課題である。

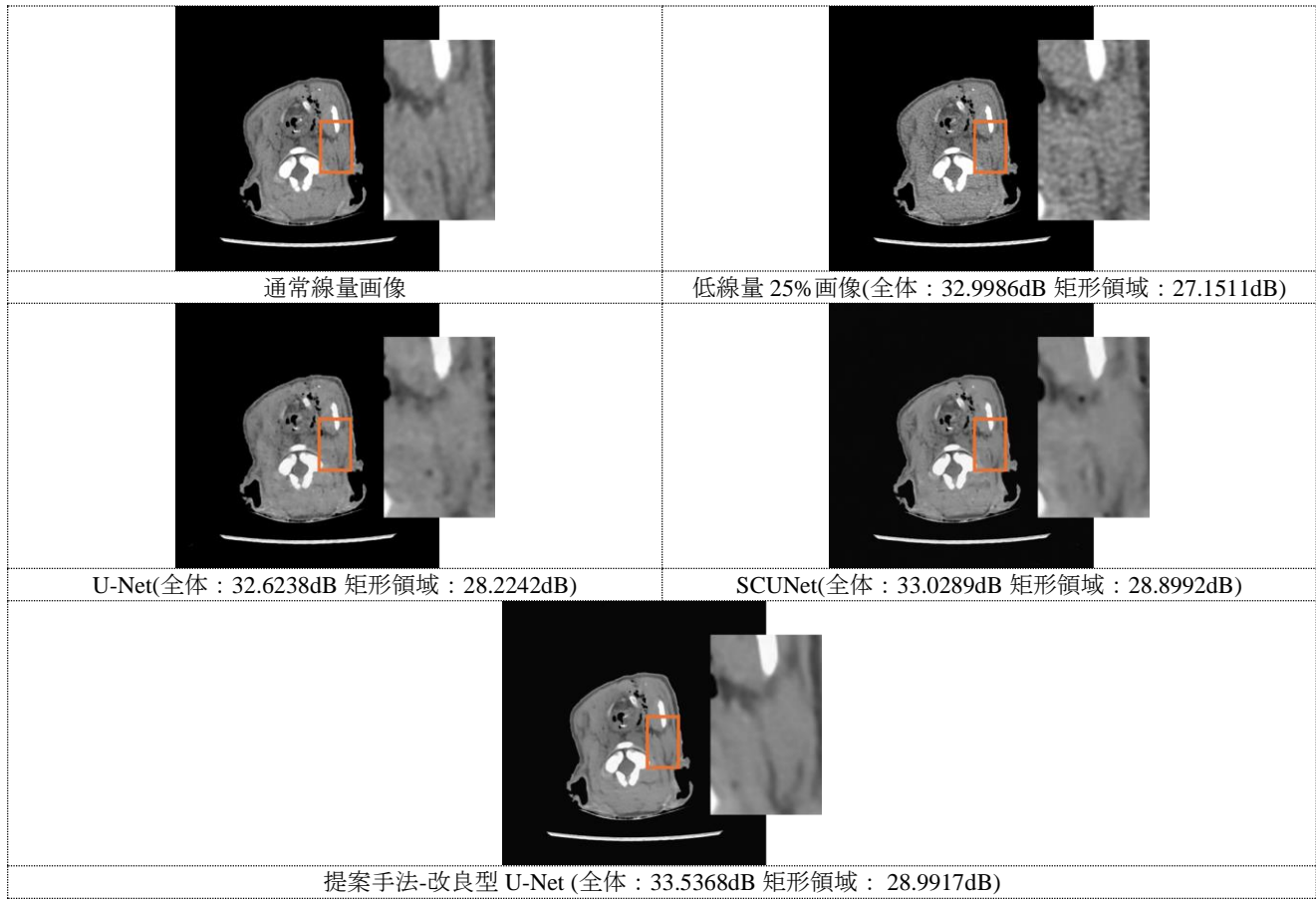


図 6 ノイズ除去画像(低線量 25%)

参考文献

- [1] OECD, Computed tomography (CT) scanners, <https://www.oecd.org/tokyo/statistics/ct-scanners-japanese-version.htm>, (accessed:2024/05/29).
- [2] 環境省, 放射線による健康影響等に関する統一した基礎資料, <https://www.env.go.jp/chemi/rhm/r4kisoshiryo/r4kiso-02-05-03.html>, (accessed:2024/05/29).
- [3] M. Gholizadeh-Ansari et al., "Deep learning for low-dose CT denoising using perceptual loss and edge detection layer", Journal of digital imaging, Vol.33, No.2, pp.504-515(2020).
- [4] L. Jaakko et al., "Noise2Noise: Learning image restoration without clean data", arXiv:1803.04189v3(2019).
- [5] X. Yi, Recent publication., <https://onedrive.live.com/?authkey=%21APVzpTOocrEDidk&cid=F4D4F832E5E42083E&id=FD4F832E5E42083E%21582&parId=root&action=locate>, (accessed:2024/05/29).
- [6] O. Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation", Medical Image Computing and Computer-Assisted Intervention, pp.234- 241(2015).
- [7] M. Tan et al., "EfficientNetV2: Smaller Models and Faster Training", Proceedings of the 38th International Conference on Machine Learning, PMLR 139, pp.10096-10106(2021).
- [8] S. Woo et al., "CBAM: Convolutional block attention module", European Conference on Computer Vision, pp.3-19(2018).
- [9] H. Li et al., "Pyramid Attention Network for Semantic Segmentation", arXiv:1805.10180(2018).
- [10] M. Tan et al., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", Proceedings of the 36th International Conference on Machine Learning, 97:pp.6105-6114(2019).
- [11] J. Hu et al., "Squeeze-and-Excitation Networks", IEEE Conference on Computer Vision and Pattern Recognition, pp.7132-7141(2018).
- [12] K. Zhang et al., "Practical Blind Image Denoising via Swin-ConvUNet and Data Synthesis", arXiv:2203.13278v4(2023).