

# Title

author1<sup>1</sup>, author2<sup>2</sup>, author3<sup>3</sup>

(Dated: January 7, 2022)

**Abstract:** This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.This is abstract.

**Keywords:** Keyword1; Keyword2; Keyword3;...

## 1 Introduction

This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.

### 1.1 title

This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.[\[1\]](#)

#### 1.1.1 title

This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.This is introduction.

## 2 Method

在本节中，我们给出了我们提出的网络结构和参数，以及 loss 的选择。

### 2.1 整体的网络结构

假设  $x \in R^{MN \times 1}$  是已被测量得出的在稀疏下的正弦图， $y \in R^{HWS \times 1}$  是其对应的高质量 CT 图像。我们的任务是找到一个函数  $f$ ，其满足

$$y = f(x) \tag{1}$$

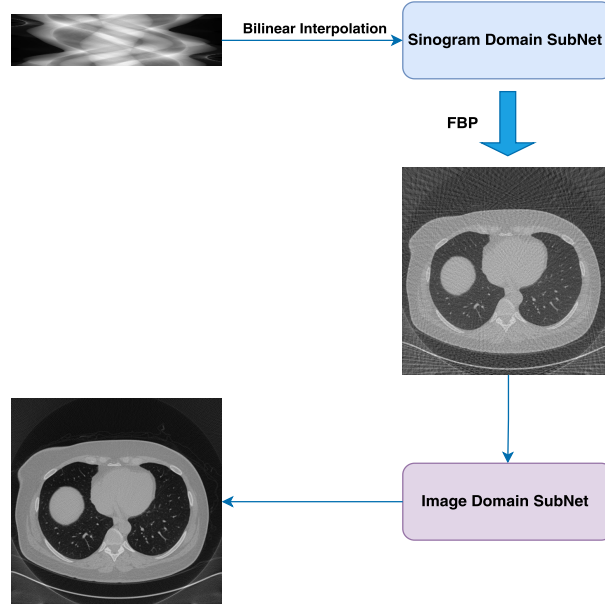


Figure 1: PLformer 的整体架构，该网络有三个阶段：Sinogram Domain SubNet, Filter Back Projection (FBP), 和 Image Domain SubNet。

其中  $M$  和  $N$  分别表示探测器的数量和投影角度， $H \times W \times S$  表示 CT 图像  $Y$  的分辨率， $S$  是 CT 图片切片的数量，对于 fan beam geometry 来说， $S$  的值为 1。我们将使用深度学习网络去拟合函数  $f$ ，网络的整体架构和流程如图 1 所示。

网络整体由三个阶段组成，首先通过 Sinogram Domain SubNet 对双线性插值后的稀疏正弦图进行重建，然后通过 FBP[7] 算法完成投影域到图像域的转换，最后通过 Image Domain SubNet 重建出高质量的 CT 图片。

## 2.2 Parallel LeWinTransformer Block (PLB)

过去基于卷积的神经网络的特征提取是非常成功的，但卷积操作需要不断堆积卷积层来完成对图像从局部信息到全局信息的提取，不断堆积的卷积层慢慢地扩大了感受野直至覆盖整个图像。但是 transformer 并不假定从局部信息开始，而且一开始就可以拿到全局信息，尽管训练难度更大一些（如所需要的数据集要更大些，以及所需要学习的参数量更多），但是一旦完成好训练，更早得到全局信息的优势会使得效果更好。受到 Swin Transformer[10] 以及 Uformer[16] 的启发，我们使用由 non-overlapping window-based Multi-head Self-Attention (W-MSA) 和 Locally-enhanced Feed-Forward Network (LeFF) 所组成的 locally-enhanced window (LeWin) Transformer 来获取图像信息，其中 W-MSA 用来捕获全局信息，而 LeFF 用来捕获局部信息。LeWin Transformer 的计算表示为：

$$\begin{aligned} X'_n &= \text{W-MSA}(\text{LN}(X_{n-1})) + X_n \\ X_n &= \text{LeFF}(\text{LN}(X'_n)) + X'_n \end{aligned} \quad (2)$$

其中 LN 表示 the layer normalization,  $X'_n$  和  $X_n$  分别为 W-MSA 和 LeFF 的输出。

由于切 Patch 会导致边界信息的不易捕获，如图 2 所示，我们提出 Parallel LeWinTransformer Block (PLB)，将 transformer block 并行 concatenate，并且 shift 互不相等。之后，使用 Point-Wise

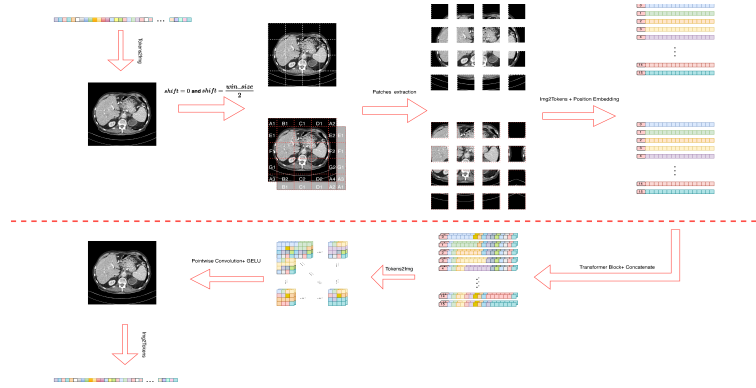


Figure 2: Parallel LeWinTransformer Block(PLB) 整体架构。

Convolution[2] 将二者互补，达到弥补边界信息的作用。整个过程可表示为：

$$X_l = \text{PW}(\text{Concat}(\text{LeWin}_1(X_{l-1}), \text{LeWin}_2(X_{l-1})) + X_{l-1} \quad (3)$$

其中 Concat 表示拼接操作，LeWin<sub>*i*</sub> (*i* = 1, 2) 分别表示 shift=0 和 shift=n 的 LeWin Transformer Block。PW 表示 Point-Wise Convolution。

GELU[4] 作为激活函数。

### 2.3 Sinogram Domain SubNet

如图 3 Part1 所示，Sinogram Domain SubNet 采用编码-解码的架构，他由 5 部分组成。首先通过 Input Projection 对输入的 Sinogram 2-D 图片通过一个 3x3 的卷积提取特征并通过 Flatten 将每个特征变成 1-D 的 token。然后在这些 token 上应用 PFLB 来获取长范围依赖，并且根据 LeWinTransformer 可以通过层次化的表示来减少 token 数量的特点，我们在 PFLB 之后对采用 token 长度变为 1/4, token 维数加倍的下采样，这样 token 的总数会减半。在两层 encoder 之后的使用 PFLB 作为 bottleneck layer。我们同样使用 2 层 decoder 来进行 token 重建。和 encoder 相似，decoder 由 token 长度 ×4, 维数减半的上采样和 PFLB 组成。但不同的是我们将上采样和对应 encoder 阶段的 token 合并起来，并用 point-wise convolution 进行融合之后在输入到 PFLB 中。最后通过 Output Projection 将 1-D 的 token 转为 2-D 的特征，再用一个 3x3 的卷积获取 residual image。假设输入的正旋图为  $X$ ，得到的 residual image 为  $R$ 。则我们最终的结果定义为  $X' = X + R$ ，而正常采样下的正弦图为  $Y$ ，我们使用 the mean square error (MSE) 作为我们的损失函数：

$$l(X', Y) = |X' - Y|^2 \quad (4)$$

**参数选择：** 我们选择特征的通道数为 16，并且在 Input Projection 和 Output Projection 的卷积之后使用 LeakyReLU leaky rectified linear activation functions (LReLU)[11] 去稳定我们的训练。下采样我们使用参数为 kernel size 为 4, stride 为 2, padding 为 1 的卷积，而上采样使用 kernel size 为 2, stride 为 2 的转置卷积。对于每个 W-MSA，我们使用了参数为 0.5 的 Dropout[14] 来防止过拟合。最后，我们将卷积层中的权重初始化为正态分布 ( $\mu = 0.0, \sigma = 0.02$ )。

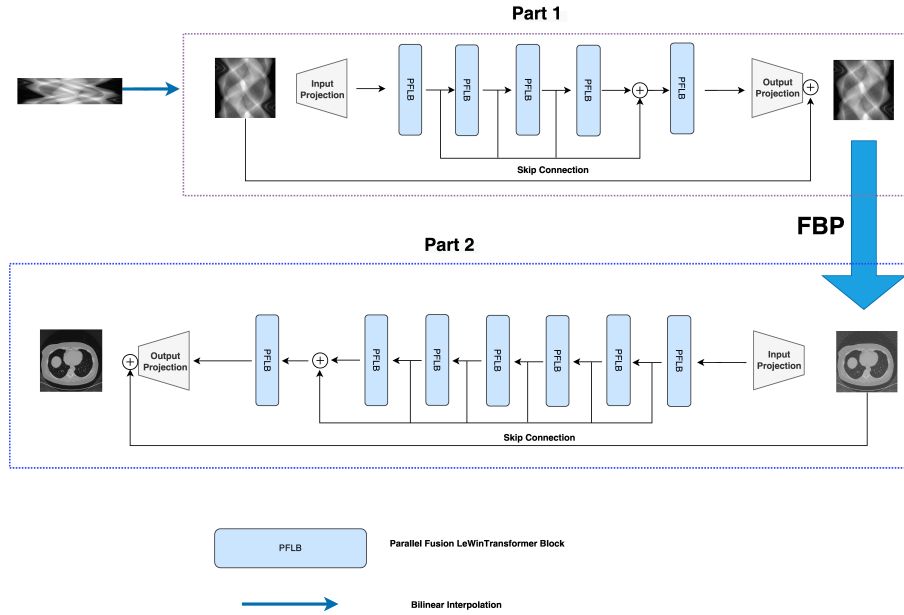


Figure 3: PLFormer, Part1 为 Sinogram Domain SubNet, Part2 为 Image Domain SubNet。

## 2.4 Image Domain SubNet

由于 FBP 重构的 CT 图像受到条纹伪影和噪声的影响而退化，所以我们通过 Image Domain SubNet 来重建出高质量的 CT 图。如图 3 Part2 所示, Image Domain SubNet 整体架构和 Sinogram Domain SubNet 类似。不同的是为了得到更多的 token, 我们分别选用了 5 层的 encoder layer 和 decoder layer, 以及将特征的通道数设为 18。为了加快模型的收敛速度以及使得模型的 performance 更好, 我们使用了 Charbonnier Loss[9] 作为损失函数:

$$\text{Charbonnier\_Loss}(X', Y) = \sqrt{(X' - Y)^2 + \epsilon^2} \quad (5)$$

其中  $X'$  表示我们最终重建得到的 CT 图,  $Y$  是高质量的 CT 图,  $\epsilon$  的值为常量  $10^{-3}$ 。

# 3 Experiments

## 3.1 数据集和实验环境

我们使用的是由梅奥诊所 (Mayo clinic) 在 2020 年提供的“Low Dose CT Image and Projection Data (LDCT-and-Projection-data)”[12] 公共数据集。其图像数据集总共包含 25908 张 1mm 厚度全剂量 CT 图片来自总共 150 个病例。参考图像是使用 FBP 方法从 512 个投影视图生成的, 我们简单地将投影数据下采样到 128 和 64 个视图, 以模拟采样率分别为 1/4 和 1/8 的稀疏视图情况。150 个病例分别为 50 个头部病例, 50 个胸部病例和 50 个腹部病例。我们将随机选取 40 个头部病例, 40 个胸部病例以及 40 个腹部病例作为训练集以及选取个头部病例, 5 个胸部病例以及 5 个腹部病例作为验证集, 再将剩下的 5 个头部病例, 5 个胸部病例以及 5 个腹部病例作为测试集。

(自己网络的参数。。)

我们对比了近几年几种基于深度学习的方法的其他性能, 包括 FBPCConvNet[6], DD-Net[18], DP-ResNet[17], Adaptive-Net[3], EEDeepNet[15] ((后续可能补充其他网络以及自己的消融实验)。FBPCConvNet 是一种后处理方法, 采用 U-Net[13] 来减少 FBP 重构中的伪影。DD-Net 结合了 DenseNet[5] 和反卷积的优点, 采用快捷连接将 DenseNet 和反卷积连接起来, 提高了网络的训练速度。DP-ResNet 是一种用于 CT 图像重建的双域网络。该算法在投影域和图像域分别对输入的测量数据进行处理, 并使用 FBP 连接两个子网络。EEDeepNet 是一种用于 CT 图像重建的端到端深度网络, 该网络直接将稀疏的正弦图映射到 CT 图像上, 因为原论文中并没有对其提出的网络起名, 所以我们将其论文的标题 “End-to-End Deep Network” 简称为 EEDeepNet 来表示该论文所提出的网络。所有对比实验的训练参数都充分参考原论文或代码中的设置。PLFormer 是由 Adam 算法 [8] 训练的, 学习率从初值  $3 \times 10^{-4}$  缓慢下降到  $10^{-6}$ 。mini-batch 的 size 设为 4。采用峰值信噪比 (PSNR) 和结构相似性指数 (SSIM) 来评价所有方法的性能。实验环境为 Python3.8+PyTorch1.7.1 在 PC 上 (Ubuntu20.04+Intel Xeon Silver 4210R CPU + 64G RAM 以及两张 NVIDIA RTX A5000), 并且使用 PyTorch 提供的 DistributedDataParallel(DDP) 去尽可能的缩短训练时间。所有工作的代码我们放在 (github) 上。

### 3.2 性能评价结果

采用不同的基于深度学习的方法对整个测试集的统计定量结果如表 1 所示, 给出了 PSNR 和 SSIM 的均值和方差。<sup>1</sup>

method	PSNR	SSIM
FBP	24.8415 $\pm$ 0.2952	0.5076 $\pm$ 0.0114
bilinear+FBP	25.0514 $\pm$ 0.2593	0.6683 $\pm$ 0.0221
FBPCovNet	34.1223 $\pm$ 0.4181	0.8637 $\pm$ 0.0111
DD-Net	33.0753 $\pm$ 0.2515	0.8324 $\pm$ 0.0081
DP-ResNet	29.8253 $\pm$ 0.2874	0.7441 $\pm$ 0.0159
Adaptive-Net	31.7135 $\pm$ 0.3483	0.7854 $\pm$ 0.0151
EEDeepNet	34.2187 $\pm$ 0.5583	0.8706 $\pm$ 0.0137

Table 1: 不同方法的性能评价结果 (均值  $\pm$  方差), 最好的值用红色标出。(双域的话应该不止 LDCT, 但是先放在这做一个提醒作用, 别忘了还要直接拿 input 和 target 去做对比)

## 4 Conclusions

This is conclusion. This is conclusion. This is conclusion. This is conclusion. This is conclusion.  
This is conclusion. This is conclusion. This is conclusion. This is conclusion. This is conclusion.

## Acknowledgments

These are acknowledgments. These are acknowledgments. These are acknowledgments. These are acknowledgments. These are acknowledgments. These are acknowledgments.

## References

- [1] Hu Chen, Yi Zhang, Mannudeep Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36:2524–2535, 06 2017.
- [2] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [3] Y. Ge, T. Su, J. Zhu, X. Deng, Q. Zhang, J. Chen, Z. Hu, H. Zheng, and D. Liang. Adaptive-net: deep computed tomography reconstruction network with analytical domain transformation knowledge. *Quantitative Imaging in Medicine and Surgery*, 10(2), 2020.
- [4] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.
- [5] G. Huang, Z. Liu, Vdm Laurens, and K. Q. Weinberger. Densely connected convolutional networks. *IEEE Computer Society*, 2016.
- [6] K. H. Jin, M. T. Mccann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, PP(99):4509–4522, 2016.
- [7] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.
- [8] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014.
- [9] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [10] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *International Conference on Computer Vision (ICCV)*, 2021.
- [11] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. 2013.
- [12] Taylor R Moen, Baiyu Chen, David R Holmes III, Xinhui Duan, Zhicong Yu, Lifeng Yu, Shuai Leng, Joel G Fletcher, and Cynthia H McCollough. Low-dose ct image and projection dataset. *Medical physics*, 48(2):902–911, 2021.
- [13] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Springer International Publishing*, 2015.
- [14] 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*, 15(1):1929–1958, 2014.

- [15] W. Wang, X. G. Xia, C. He, Z. Ren, and B. Lei. An end-to-end deep network for reconstructing ct images directly from sparse sinograms. *IEEE Transactions on Computational Imaging*, 6:1548–1560, 2020.
- [16] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. *arXiv preprint 2106.03106*, 2021.
- [17] Xiangrui Yin, Qianlong Zhao, Jin Liu, Wei Yang, Jian Yang, Guotao Quan, Yang Chen, Huazhong Shu, Limin Luo, and Jean-Louis Coatrieux. Domain progressive 3d residual convolution network to improve low-dose ct imaging. *IEEE Transactions on Medical Imaging*, 38(12):2903–2913, 2019.
- [18] Zhicheng Zhang, Xiaokun Liang, Xu Dong, Yaoqin Xie, and Guohua Cao. A sparse-view ct reconstruction method based on combination of densenet and deconvolution. *IEEE Transactions on Medical Imaging*, 37(6):1407–1417, 2018.