

DDPTransformer: Dual-Domain With Parallel Transformer Network for Sparse View CT Image Reconstruction

author¹, author², author³

(Dated: April 6, 2022)

Abstract: 对于临床诊断来说，CT 技术是必不可少的。但由于 CT 使用有害的电离辐射，所以每做一次 CT 会对人体造成不可逆的损伤。为了减少辐射对人体的影响，研究人员通过减少采样点的个数得到稀疏采样图像，但稀疏采样下的图像会带来图像不清晰，条纹伪影严重等问题，从而为诊断带来不好的影响。我们的研究内容是从稀疏采样下的 sinogram 中恢复高质量的 CT 图像。近年来基于深度学习的方法被广泛应用于 CT 重建，然而长期以来，基于 CNN 的模型由于感受野通常很小，所以在处理大尺寸的图像上往往丢失全局信息。为了解决这个问题，我们提出了一种基于 Transformer 的模型 (called:DDPTransformer) 来获取全局信息并在”Low Dose CT Image and Projection Data (LDCT-and-Projection-data)”数据集上进行训练和验证模型性能，更具体的说，我们使用 Transformer 来代替传统的卷积操作，并通过不同的切 Patch 方式使得两个 Transformer 可以互相拟补 Patch 块的边缘信息。结果显示在不同的稀疏采样下模型均表现出优异的性能，和其他先进的算法对比有了很大提高。并最后通过在不同的数据集上验证了模型的鲁棒性。代码和模型可以在此公开获得：xxxxx

Keywords: Deep learning;sparse view CT reconstruction;Transformer;dual domains

1 Introduction

计算机断层扫描 (COMPUTED tomography, CT) 由于能够在不破坏物体的情况下实现物体的内部视觉，已广泛应用于临床、工业和其他领域 [3]。但 CT 的电离辐射会对人体造成危害 [18]，这严重限制了它的实际应用。在临床诊断中，为了降低辐射剂量和减少扫描时间而采用 sparse view CT。然而，投影视图的缺陷给重建的图像带来 ill-posed inverse problems [29]。许多重建算法被提出用来解决这些问题，他们一般被分为 3 类：(a) sinogram domain pre-processing,(b) iterative algorithm, and (c)image domain post-processing。

sinogram domain pre-processing 首先对 sinograms 进行上采样和去噪，然后再将它们重建出 CT 图像。在过去的几十年中，在 singram domain 中提出了非线性平滑 [27]、结构自适应滤波 [4] 和基于字典学习的图像修补方法 [26] 方法用于上采样和去噪。然后再用经典的解析算法 (en:classic analytical method) 如 filtered back-projection (FBP) [22] 重建出 CT 图像。然而，由于重建对 sinogram domain 中产生的误差很敏感，这些方法的性能往往会受到影响。

除了简单的 back projection 和改进的 FBP 算法, 在过去的几十年中, 图像重建更多使用的是 iterative algorithm。尤其是将压缩感知 (CS) [7] [12] 引入到迭代重建之后, CT 图像质量大幅度提升。其中最著名的是总变异 (TV) [37]。除此之外, iterative algorithm 还包括 nonlocal means (NLM) [45], tight wavelet frames [15], dictionary learning [5, 44], low rank [6] 以及 TV 之后的改进版 [34, 48, 49]。然而, 上述迭代重建方法由于计算量巨大以及难以调优的参数, 导致其需要较长的计算时间以及很难去泛化不同的扫描方案或人体部位产生的不同图像。

最后一类稀疏视图 CT 重建算法是 image domain post-processing。稀疏视图 CT 投影数据经过解析算法 (en:classic analytical method, 如 FBP) 重构后, CT 图像中会出现噪声和条纹伪影等问题。image domain post-processing 方法通过去除解析算法重建之后的这些问题来提高 CT 图像的质量。受稀疏表示理论的启发, Aharon 等人将字典学习 [1] 应用于 LDCT 去噪, 显著提高了腹部图像的质量。Han [19] 等人考虑到光谱中的其他信息, 提出 low-rank Hankel matrix(ALOHA) 从 sparse-view 中重建出 CT 图像。与其他两种方法相比, 噪声在图像域中的分布不能准确确定, 这使得 post-processing 方法无法在保持结构和噪声替代之间实现最优的权衡。随着计算机硬件性能的提升以及深度学习的快速发展促成了许多在医学图像重建领域的成功应用。例如, Chen 等人 [10] 提出了一种残差的编码器-解码器卷积神经网络 (Res-CNN) 来从 FBP 重建中去除伪像。Jin 等人 [21] 提出了一种结合了 FBP、u-net 和残差学习的深度卷积网络 (deep convolutional network, FBPconvNet), 在保留图像结构的同时去除 CT 图像的伪影。Eunhee 等人 [23] 结合小波变换和深度残差学习解决低剂量 CT 降噪, 这篇文章的方法赢得了 2016 Low-Dose CT Grand Challenge 第二名。为了使预测图像服从与 NDCT 相同的统计分布, Wolterink 等人 [43] 中引入生成对抗网络 (generative adversarial network, GAN), 并利用判别器网络实现该约束。Yang 等人 [46] 提出使用带有 Wasserstein 距离和感知损失的生成式对抗网络 (WGAN-GP) 从 FBP 重建的 CT 图像中恢复微妙的结构。Chen 等人 [9] 展开了最陡梯度下降算法, 提出了稀疏视图 CT 的基于学习专家评估的重构网络 (LEARN)。尽管这些方法已经显示出良好的效果, 然而基于 CNN 的模型通常感受野很小, 不利于捕获全局信息。

近年来, 为了解决 CNN 感受野有限导致很难捕获全局信息, 一种基于注意力的编码器-解码器结构-Transformer [39] 被提出来, 并在许多计算机视觉任务中取得了巨大成功。Dosovitskiy 等人 [14] 提出了 Vision Transformer(ViT) 用于图像分类, 这是 Transformer 首次用于计算机视觉, 并在著名的 ImageNet 数据集 [13] 上突破了 SOTA。Liu 等人 [28] 提出了通过滑动窗口来弥补 patch 边缘信息的 Swin Transformer, 这也是本文主要的灵感来源。Chen 等人提出了 IPT [8], 这是 transformer 第一次应用于 low-level 计算机视觉任务。Wang 等人提出了一种基于 U-net 架构的 Transformer 模型 [42] 用于图像重建。Luthra 等人 [30] 提出基于边缘增强的 transformer 模型, 用于医学图像去噪。尽管以上 Transformer 模型在其他计算机视觉的任务上表现优异, 然而据我们所知, 目前还没有将 Transformer 用于在 sparse-view 下的双域 CT 重建任务中。

In this paper, 我们提出了一个全新的模型-DDPTransformer 用于 sparse-view 双域 CT 重建, 利用 transformer 相比与 CNN 获得的全局信息的优势, 我们超越了现有的最先进的方法并且证明了 transformer 对于双域重建的有效性。本文的其余部分组织如下。In Section II, 我们详细阐述了 DDPTransformer 的实现细节。大量的实验结论 in Section III。In Section IV, 我们将讨论模型的鲁棒性。the remarks on conclusions and future works are presented in Section V。

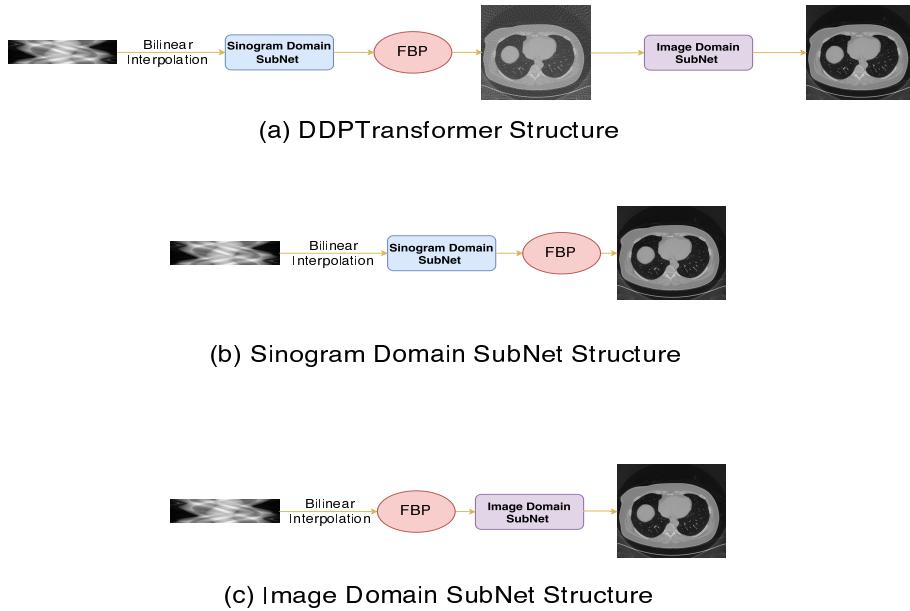


Figure 1: The Overall structures of the DDPTTransformer, Sinogram Domain SubNet and Image Domain SubNet.

2 Method

在本节中，我们详细阐述了 DDPTTransformer 的实现过程。首先介绍 DDPTTransformer 的整体架构，之后描述搭建网络过程中的细节以及 loss function 的选择。

2.1 Overall Network Architecture

从 sinogram 中重建出 CT 图像可表示为以下线性逆问题：

$$g = Af + \eta \quad (1)$$

其中 g 是已经被测量出的 singram(投影数据), f 是其对应的高质量的 CT 图像, A 是对成像系统进行建模的矩阵。 η 是噪声。事实上在 2DCT 重建任务中, A 等价于平行光束成像几何的 Radon 变换，并且它是不可逆的。 η 大部分是 Poisson noise。因此, this inverse problem is highly ill-posed。

我们的任务就是通过神经网络去拟合 this inverse problem¹。网络的整体架构如图 1(a) 所示¹, 它由四个阶段组成, 第一步对输入的 sinogram 经过双线性插值法模拟出正常采样下得到的 sinogram, 当然它的误差和噪声很大; 第二步将其输入 Sinogram Domain SubNet 进行去噪; 第三步通过 Filter BackProjection(FBP) 算法将 sinogram 转为 CT 图, 但 FBP 算法会带来条纹伪影和噪声; 所以最后一步通过 Image Domain SubNet 去修正 FBP 算法带来的误差, 得到高质量的 CT 图像。在 DDPTTransformer 的四个阶段中, 阶段 2 的 Sinogram Domain SubNet 和阶段 4 的 Image Domain SubNet 是两个深度神经网络, 所以我们要分别对其进行训练。图 1(b) 和图 1(c)¹ 表示这两个子网络训练时的总体架构。

在 DDPTTransformer 中, Sinogram Domain SubNet 和 Image Domain SubNet 均由 DDPTTransformer Block 组成。所以在介绍两个 SubNet 之前, 我们将详细说明 DDPTTransformer Block 的结构和细节。

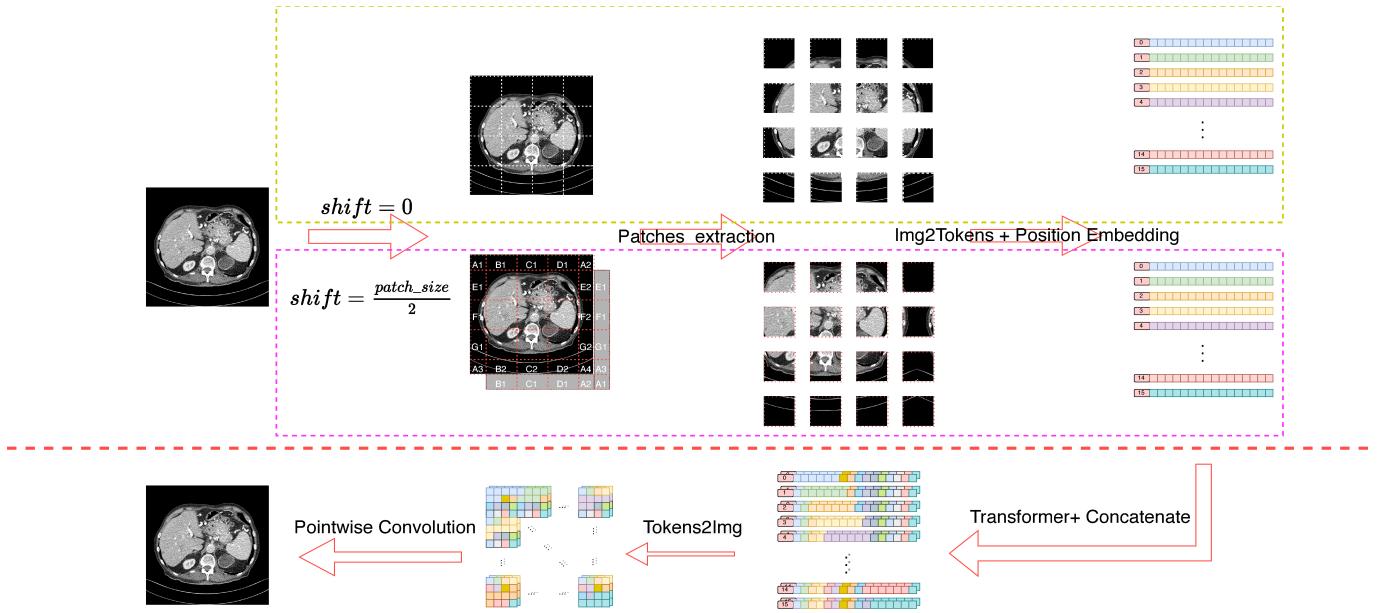


Figure 2: DDPTransformer Block

2.2 DDPTransformer Block Design

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，这会导致浅层的卷积层不会得到太多的全局信息。但 transformer 并不假定从局部信息开始，而是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。我们提出的 DDPTransformer Block 流程如图 2 所示。

首先我们将输入图片 $x \in \mathbb{R}^{H \times W \times C}$ 划分成相同大小的 Patch 块（例如在图 2 中我们划分了 16 个 Patch 块），并根据两种不同的划分方案分别进行操作，where (H, W) 是输入图片 x 的尺寸， C 是通道。一种划分方案为 $shift = 0$ （即图 2 黄色矩形框部分），即将输入图片划分成 16 宫格的形状 $x_p \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 16C}$ ，之后因为标准 Transformer 接收的输入是一维序列，所以我们将每个 x_p 拉平为 $x_d \in \mathbb{R}^{N \times 16C}$ ，where $N = \frac{H \times W}{16}$ 是每个 x_d 的长度。并参考 Vision Transformer 为每个序列加上 Position embeddings 以保留位置信息。另一种划分方案为 $shift = \frac{patch_size}{2}$ （即图 2 紫色矩形框部分），首先将图片 x 沿右沿下平移 $\frac{patch_size}{2}$ 得到 x' ，即图 2 阴影部分。之后在 x' 上按照 16 宫格的形状在 x 上划分出尺寸不一的 patch 块 x_p ，并按照图中所示将部分 x_p 进行 shift，如将 x 中的 $A1$ 块移动到 x' 中的 $A1$ 块， x 中的 $B1$ 块移动到 x' 中的 $B1$ 块等。所有移动完成之后就可以得到尺寸一样的 $x_p \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 16C}$ ，之后与第一种方案一样，将 x_p 拉平为 $x_d \in \mathbb{R}^{N \times 16C}$ 并加上 Position embeddings。

接着我们将两种方案得到的是一维序列分别输入到 Transformer 中。如图 3(a)所示，Transformer 由标准的多头自注意力 (multi-head self attention (MSA)) 模块和 Layer-Conv-Layer(LCL) 模块组成（如图 3(b) 所示）。A layer normalization(LN) [2] 被用在每个模块之前，and a Skip Connection 被用在

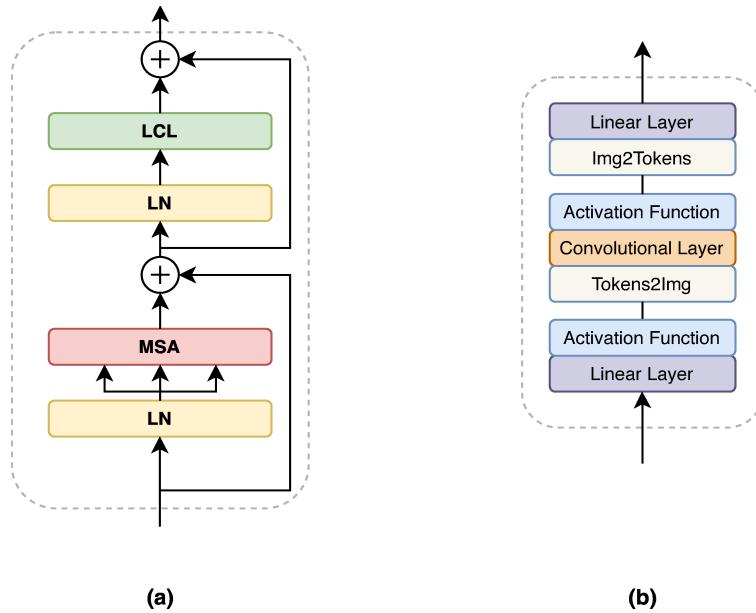


Figure 3: Transformer and LCL structure.

每个模块之后。Transformer 块的计算表示为：

$$\begin{aligned}\hat{x}_d &= \text{MSA}(\text{LN}(x_{d-1})) + x_{d-1} \\ x_d &= \text{LCL}(\text{LN}(\hat{x}_d)) + \hat{x}_d\end{aligned}\quad (2)$$

其中 \hat{x}_{d-1} 和 x_d 分别为 MSA 模块和 LCL 模块的输出。下面，我们分别对 MSA 和 LCL 进行详细说明

Multi-head Self-Attention (MSA)。我们将 $x_{d-1} \in \mathbb{R}^{N \times 16C}$ 按 Patches 的个数将其分开并计算他们的自注意力，公式如下：

$$\begin{aligned}x_{d-1} &= \{x_{d-1}^1, x_{d-1}^2, \dots, x_{d-1}^n\}, n = \text{number of Patch} \\ y_{d-1}^i &= \text{Attention}(x_{d-1}^i w^q, x_{d-1}^i w^k, x_{d-1}^i w^v), i = 1, 2, \dots, n \\ \hat{x}_d &= \{y_{d-1}^1, y_{d-1}^2, \dots, y_{d-1}^n\}, n = \text{number of Patch}\end{aligned}\quad (3)$$

其中 w^q, w^k, w^v 表示 query, key and value 的投影矩阵, Attention 公式为：

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{D}}\right)V \quad (4)$$

其中 D 表示 query/key 的维数。上述过程为计算一个头的自注意力，假设头数为 k，即上述公式重复 k 次，得到 $\{\hat{x}_d^1, \hat{x}_d^2, \dots, \hat{x}_d^k\}$ ，并将其 Concat 得到最终 MSA 的输出 \hat{x}_d 。而为了保证参数的计算和数量不变，需要将 D 缩小 k 倍。

Layer-Conv-Layer(LCL)。MSA 模块注重于融合更多的全局信息，但是对于如何学习这些信息的能力却不足，所以我们通过 LCL 模块来弥补学习能力。由于 MSA 的输出是 1-D 序列，所以大多数都是通过 MLP 来进行学习。虽然我们加了位置信息来拟补 2-D 拉平为 1-D 导致降维带来的部分特征丢失，但我们认为更好的解决方案是在 2-D 上进行学习。所以在 MLP 中加入一个卷积操作。整个流程如图 3(b)所示，我们首先对 \hat{x}_d 用 Linear Layer 并使用激活函数，之后 reshape 为 2-D 并使用卷积和激活函数进行学习。之后在拉平回 1-D 并再次用 Linear Layer 得到输出 x_d 。

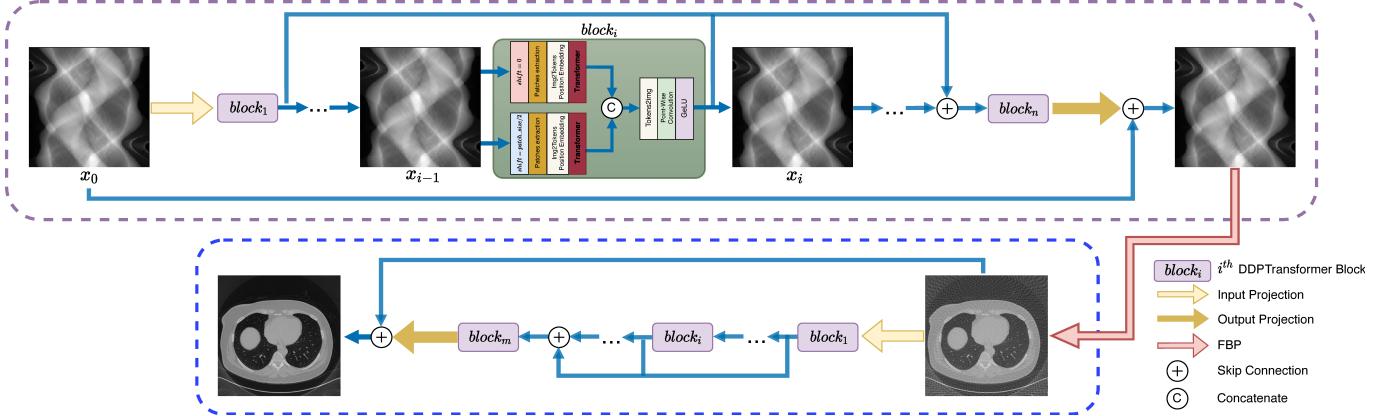


Figure 4: DDPTransformer, 紫色虚线框为 Sinogram Domain SubNet，蓝色虚线框为 Image Domain SubNet。

两种方案 ($shift = 0$ 和 $shift = \frac{patch_size}{2}$) 分别使用 Transformer 之后，我们将其输出 Concat。使用两种方案是为了互相弥补 patch 块的边缘信息，所以我们选择将其再次 reshape 为 2-D 而不是直接在 1-D 上进行互补。最后通过 Point-Wise Convolution [11] 得到输出。

2.3 Sinogram Domain SubNet

如图 4 紫色虚线框所示，Sinogram Domain SubNet 采用流式架构，他由 3 部分组成。首先将经过插值之后的 sinogram image 通过 Input Projection 扩大通道数并将每个通道进行 Flatten 操作，变成 1-D 的 token。其次我们使用 n 个 DDPTransformer Block 组成 Sinogram Domain SubNet 的 backbone, n 的值我们会通过实验选出最优值，并且使用 skip connection 使模型更容易拟合。最后通过 Output Projection 将 1-D 的 token 输出得到重建之后的 Sinogram Domain。在训练 Sinogram Domain SubNet 时，我们使用 the mean square error (MSE) 作为我们的损失函数：

$$MSE_{Loss}(X', Y) = |X' - Y|^2 \quad (5)$$

其中 X' 表示经过模型计算并通过 FBP 算法得到的 CT 图， Y 是对应的高质量的 CT 图 (label)。

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 4 蓝色虚线框所示，Image Domain SubNet 整体架构和 Sinogram Domain SubNet 类似。不同的是我们选用了 m 个 DDPTransformer Block 模块，同样 m 的值会通过实验得出。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 [25] 作为损失函数：

$$Charbonnier_{Loss}(X', Y) = \sqrt{(X' - Y)^2 + \epsilon^2} \quad (6)$$

其中 X' 表示我们最终重建得到的 CT 图， Y 是对应的高质量的 CT 图 (label)。 ϵ 的值为常量 10^{-3} 。

3 Experiments

3.1 数据集和实验设置

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)*” [33] 公共数据集。其图像数据集总共包含 25908 张 1mm 厚度高质量 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 5 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。最终得到每个视图下训练集的总数为 20800，验证集的总数为 2586，测试集的总数为 2522。

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 [21], DD-Net [50], DP-ResNet [47], Adaptive-Net [16], EEDeepNet [40]。FBPCConvNet 是一种后处理方法，采用 U-Net [36] 来减少 FBP 重构中的伪影。DD-Net 结合了 DenseNet [20] 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”简写为 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。DDPTransformer 是由 Adam 算法 [24] 训练的，学习率从初值 3×10^{-4} 缓慢下降到 1×10^{-6} , mini-batch 的 size 设为 4, 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)。由于 Transformer 参数量和计算量巨大导致训练时间更长，我们使用 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。另外我们通过 torchRadon [35] 在 PyTorch 上实现 FBP 算法。

Implementation details: 在 Input Projection 中，我们选择尺寸为 4×4 ，通道数为 32 的卷积核，并且为了节省空间，将 stride 设为 4。同样在 Output Projection，选择尺寸为 4×4 ，通道数为 32，stride 为 4 的转置卷积。并且在 Input Projection 和 Output Projection 的卷积之后使用 LeakyReLU [31] 去稳定我们的训练。对于每个 DDPTransformer Block，我们选择 MSA 的头数为 16，并在 LCL 中将隐藏层的大小设为输入的四倍进行更好的学习。我们使用了参数为 0.2 的 Dropout [38] 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 ($\mu = 0.0, \sigma = 0.02$)。

Quantitative evaluation metrics: 通过 root mean square error(RMSE), peak signal-to-noise ratio (PSNR) [17], the structural similarity index metric (SSIM) [41] 等量化评价指标，比较了不同 CT 图像重建方法的性能。RMSE 的定义如下：

$$\text{RMSE}(X, Y) = \frac{\sqrt{\sum_i^N (X_i - Y_i)^2}}{N} \quad (7)$$

其中 X 表示重建结果，Y 表示对应的参考图像, n is the number of pixels in a single image。

$$\text{PSNR}(X, Y) = 20 \times \log_{10}\left(\frac{\text{MAX}(X, Y)}{\text{RMSE}}\right) \quad (8)$$

其中 $\text{MAX}(X, Y)$ 表示 X 和 Y 中的最大值。

$$\text{SSIM}(X, Y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{x,y} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (9)$$

其中 μ 表示图像的平均值, σ^2 表示图像的方差, $\sigma_{x,y}$ 表示两张图像的协方差。 $C_1 = (0.03 \times R)^2$ 和 $C_2 = (0.01 \times R)^2$ 是两个用来稳定具有弱分母除法的常数, 其中 R 表示图像 X 的取值范围。用于计算 ssim 的图像大小为 512×512 。

3.2 性能评价结果

表一列出了在不同采样下 (views = 64 and 128) 的六种模型重构出测试集的 CT 图像的 PSNR 和 SSIM 的均值和方差, 另外还给出了两种不用模型计算的参考值 (FBP 和 bilinear+FBP)。可以看出, 在不同的扫描设置下我们的网络都获得了最高的 PSNR 和 SSIM 的值以及最低的 RMSE 值, 说明我们的网络可以重建出更高质量的 CT 图。和第二高的相比, PSNR 在不同采样下分别高出了 1.85dB(64view) 和 0.7dB(128view)。

method	PSNR	SSIM	RMSE*100	PSNR			SSIM			RMSE*100		
				64 views			128 views					
FBP	20.3683±0.3074	0.3141±0.0113	1.2390±0.0831	24.8409±0.2854	0.5076±0.0115	0.4033±0.0261						
bilinear+FBP	21.4815±0.1842	0.5326±0.0182	0.7301±0.0307	25.0491±0.2243	0.6682±0.0215	0.3195±0.0164						
FBPConvNet	27.7944±0.1953	0.6584±0.0074	0.1729±0.0076	34.1109±0.3703	0.8635±0.0104	0.0398±0.0033						
DD-Net	26.8456±0.2214	0.6415±0.0089	0.2132±0.0110	33.0648±0.3304	0.8320±0.0102	0.0512±0.0041						
DP-ResNet	23.5609±0.2156	0.6210±0.0194	0.4469±0.0221	29.8236±0.3308	0.7437±0.0195	0.1076±0.0085						
Adaptive-Net	24.2013±0.3093	0.6462±0.0290	0.3856±0.0275	31.7390±0.6146	0.7853±0.0271	0.0692±0.0095						
EEDeepNet	23.4033±0.2300	0.5975±0.0247	0.4636±0.0246	34.2305±0.6116	0.8707±0.0148	0.0388±0.0052						
DDPTransformer	29.6453±0.9910	0.7590±0.0495	0.1129±0.0236	34.8335±1.8713	0.8731±0.0385	0.0362±0.0126						

Table 1: 不同方法的性能评价结果 (均值 ± 方差), 最好的值用红色标出。

3.3 视觉效果图

图 45 显示了 64views 下不同器官 (头部, 胸部和腹部) 的重建结果。可以看到通过 FBP 重建出的 CT 图像有很严重的条纹伪影。并且不同的模型对条纹伪影的抑制程度不同, 有的甚至还会降低图像分辨率 (如 bilinear+FBP 和 AdaptiveNet)。尽管由于 views 过于稀疏, 导致每个方法重建出的 CT 图像都与原图有明显差异, 但 DDPTransformer 相比与其他方法相比仍然有最好的视觉效果。图 56 显示了来自 128views 下的重建结果。可以看到由于 views 的增加, 每个模型重建出的图像更加清晰。但由于 DDPTransformer 可以获得更多的全局信息, 可以看出在清晰度和图像的轮廓上效果依然是最好的, 和其他相比效果更接近于原图。图 67 我们列出了图 4 和图 5 中选取的测试集样本的 psnr, ssim 和 rmse 值, 第一行表示 64view 下的值, 第二行表示 128view 下的值。图 78 和图 89 分别表示图 4 和图 5 中的 roi 区域 (红色虚线框), 我们可以看到, 我们的方法重建的 CT 图像可以保留更多的细节。

3.4 Ablation Study

在本节中, 我们将评估所提出方法中不同组件的有效性。如验证 Sinogram Domain SubNet(SD-Net) 和 Image Domain SubNet(ID-Net) 的有效性、确定 DDPTransformer block 的个数 n 和 m 、每个

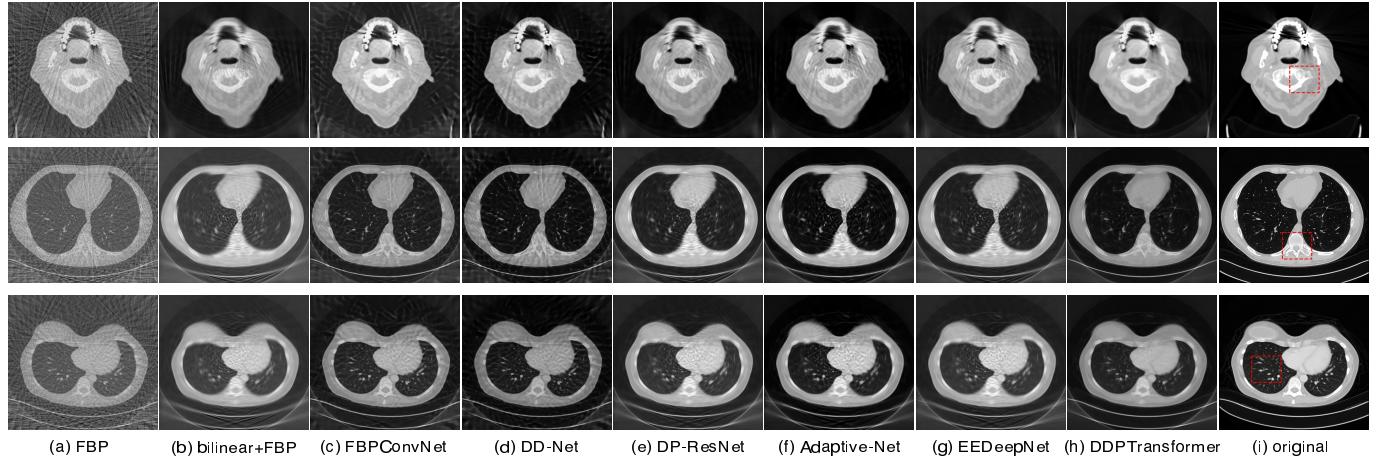


Figure 5: 64 views 不同器官的 CT 重建结果图。第一行是头部，第二行是胸部，第三行是腹部

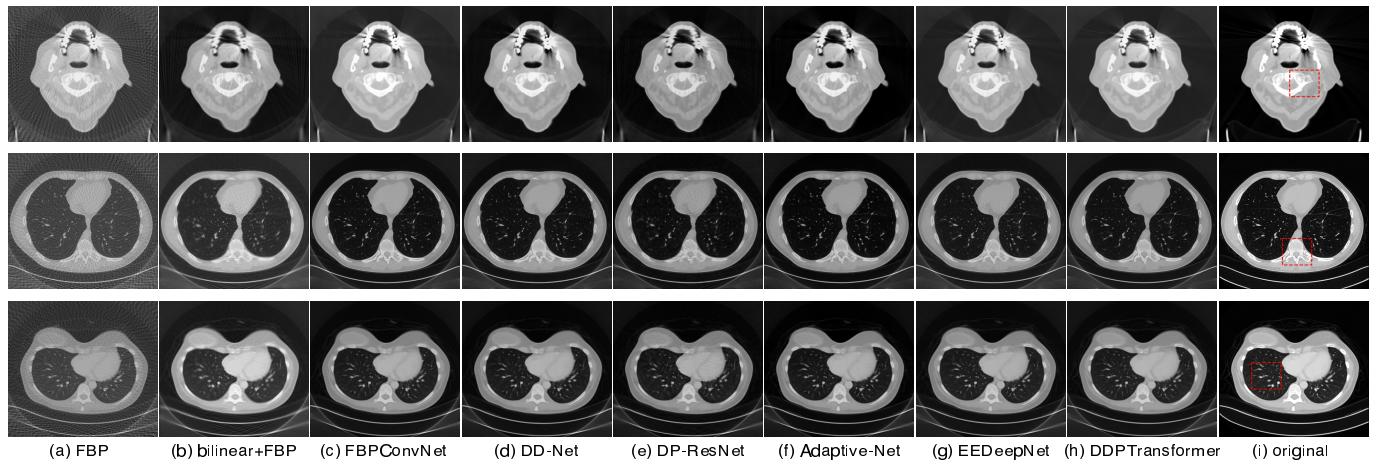


Figure 6: 128 views 不同器官的 CT 重建结果图。第一行是头部，第二行是胸部，第三行是腹部

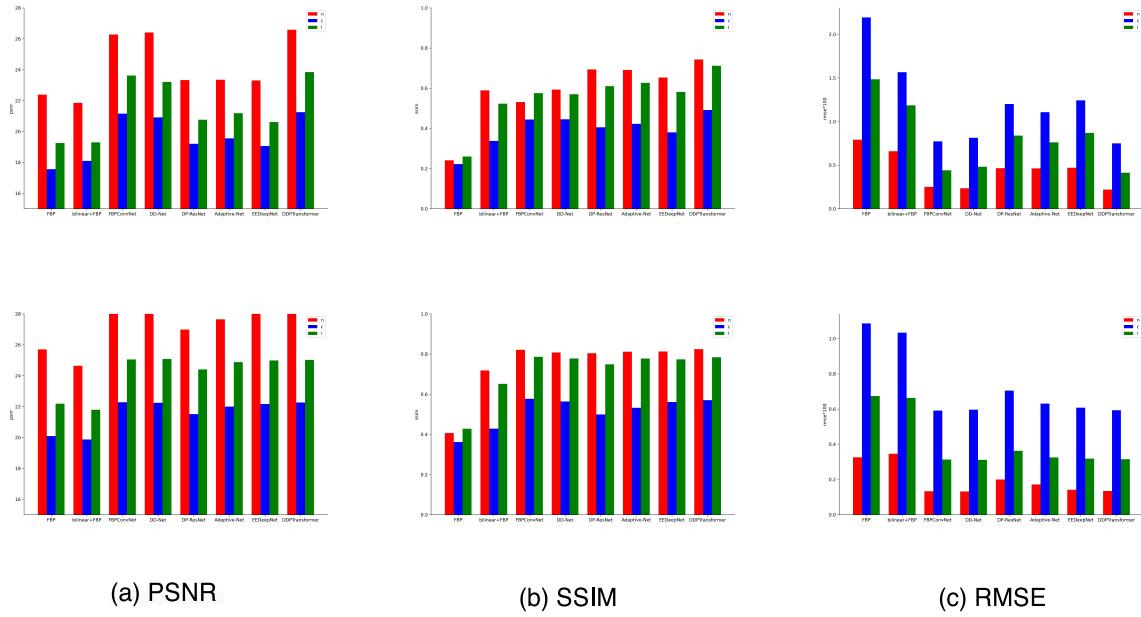


Figure 7: 图 4 和图 5 的评价指标, 第一、二行分别为 64、128views 下不同的评价指标。n 表示头部, c 表示胸部, l 表示腹部。

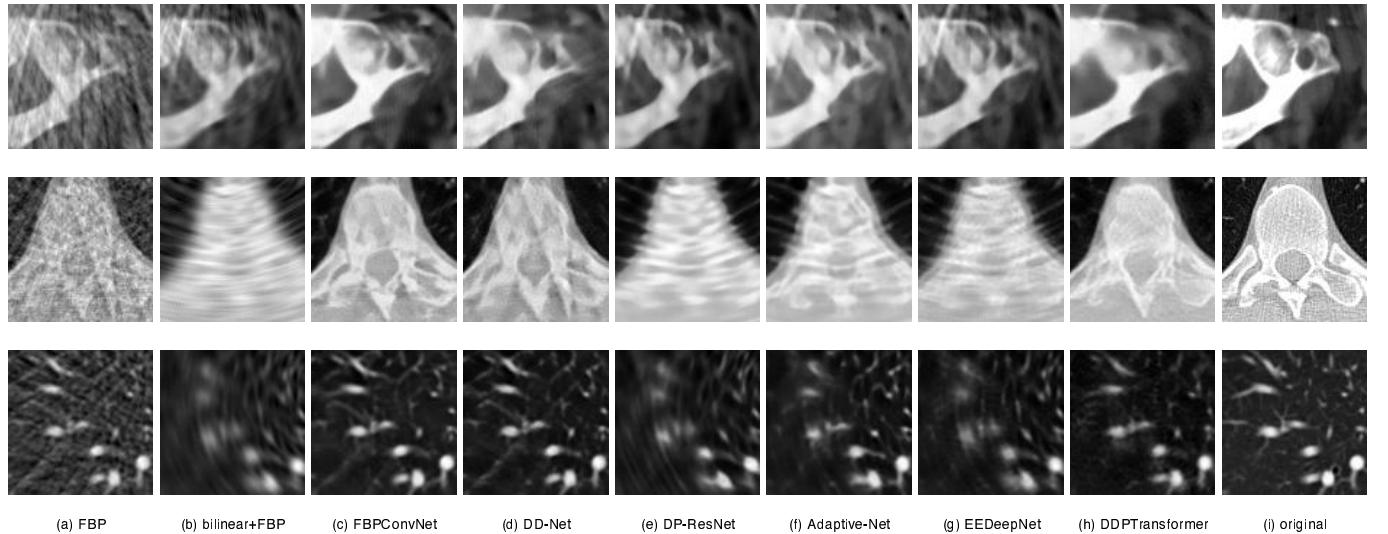


Figure 8: 图 4(i) 中红色框标记的缩放区域

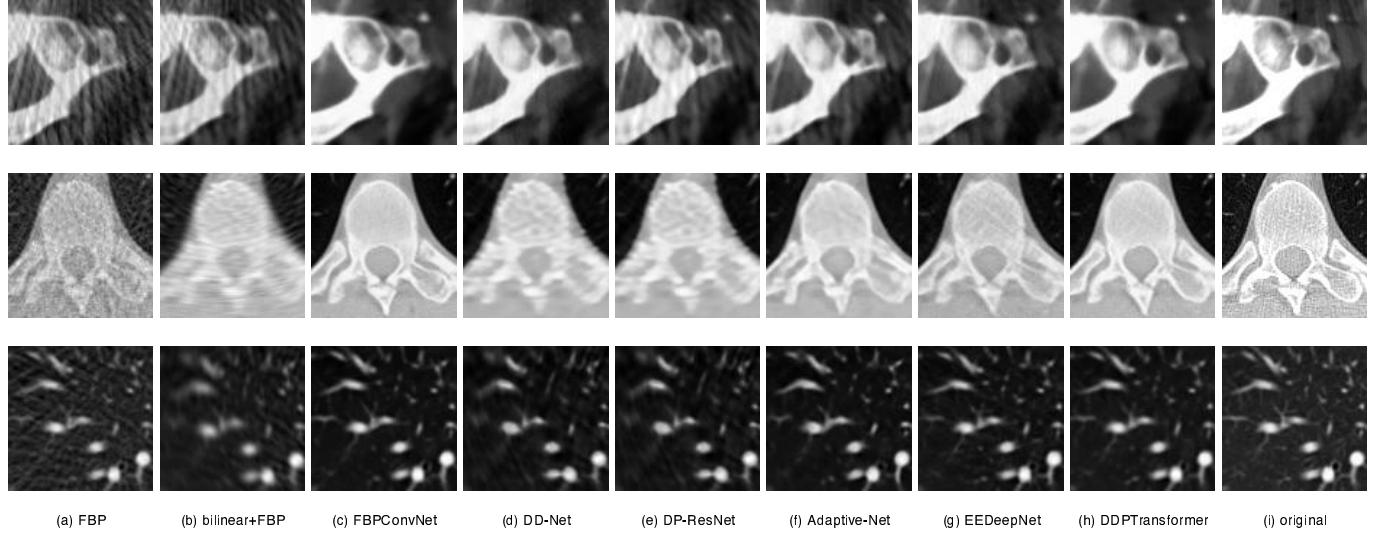


Figure 9: 图 5(i) 中红色框标记的缩放区域

DDPTransformer block 只有单个 Transformer(SiT) 或者 serial Transformer(SeT)。通过对一个参数进行改动，使其他参数保持不变，并对验证数据集的重构 CT 图像进行定量分析，确定各参数的最佳值。计算验证集中所有 CT 图像的 PSNR、SSIM 值和 RMSE 值。

表 2²列出了 SD-Net 和 ID-Net 的性能评价结果，可以看出单个子网络的性能与 DDPTransformer 的差距还是比较大的。图 10¹⁰列出了不同 DDPTransformer block 个数的折线图，在 SD-Net 中我们设置 n 为从 1 到 7，可以看出 n 从 1 到 5 时结果越来越好，5 之后效果趋于平稳，因此我们确定了 n 为 5。而 ID-Net 中我们选择了 m 为从 1 到 9 进行实验，可以看出当 m 为 7 时效果最好。表 3³列出了与 SiT 和 SeT 进行对比的性能结果，更好的结果说明我们用并行的方式来互补 Patch 块边缘信息的效果是显著的。

method	PSNR	SSIM	RMSE*100			
				64 views		
Sinogram Domain SubNet	28.9623 \pm 0.9575	0.7297 \pm 0.0466	0.1323 \pm 0.0273	33.7763 \pm 1.5785	0.8571 \pm 0.0383	0.0453 \pm 0.0143
Image Domain SubNet	21.5160 \pm 0.6347	0.5450 \pm 0.0615	0.7241 \pm 0.1018	25.1324 \pm 0.6829	0.6944 \pm 0.0617	0.3142 \pm 0.0478
DDPTransformer	29.6453 \pm 0.9910	0.7590 \pm 0.0495	0.1129 \pm 0.0236	34.8335 \pm 1.8713	0.8731 \pm 0.0385	0.0362 \pm 0.0126

Table 2: 子网络的性能评价结果 (均值 \pm 方差)，最好的值用红色标出。

method	PSNR	SSIM	RMSE*100			
				64 views		
Single Transformer	27.4160 \pm 0.3170	0.6693 \pm 0.0194	0.1848 \pm 0.0135	30.9338 \pm 0.4125	0.7783 \pm 0.0164	0.0828 \pm 0.0077
Serial Transformer	27.8340 \pm 0.2849	0.7104 \pm 0.0170	0.1677 \pm 0.0107	32.7085 \pm 0.4841	0.8350 \pm 0.0144	0.0551 \pm 0.0060
DDPTransformer	29.6453 \pm 0.9910	0.7590 \pm 0.0495	0.1129 \pm 0.0236	34.8335 \pm 1.8713	0.8731 \pm 0.0385	0.0362 \pm 0.0126

Table 3: 不同 Transformer 的性能结果 (均值 \pm 方差)，最好的值用红色标出。

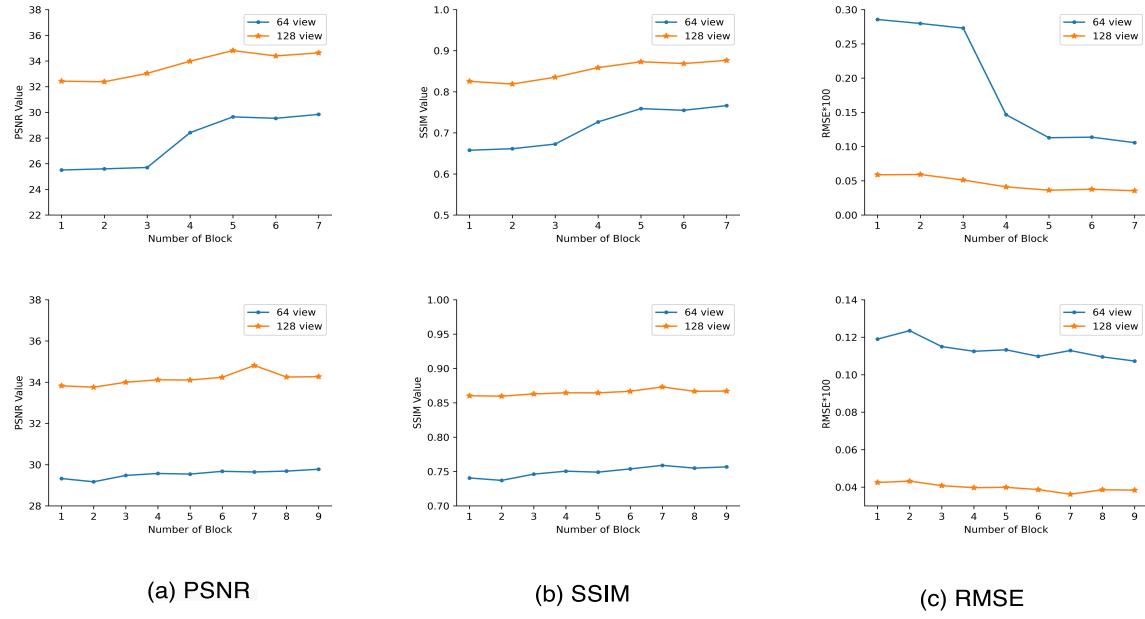


Figure 10: Sinogram Domain SubNet(第一行) 和 Image Domain SubNet(第二行) 中的 block 个数。

3.5 Running Time Comparisons

在 NVIDIA RTX A5000 GPU 上我们对运行时间进行测试。如表 4所示, 尽管与卷积相比 Transformer 在参数量、训练时间和难度上都有巨大的优势, 但在单张切片上 DDPTransformer 的测试时间(即不计算梯度的反向传播算法)却与其他卷积网络相差不大。在几乎相同的运行时间中, DDPTransformer 实现了卓越的性能。

FBP	bilinear+FBP	FBPConvNet	DD-Net	DP-ResNet	Adaptive-Net	EEDeepNet	DDPTransformer
222	210	180	121	211	148	78	204

Table 4: 不同方法的运行时间 (单位:ms)

4 Discussion and Conclusions

In this paper, 我们提出了 DDPTransformer 模型来解决 sparse-view 下双域 CT 重建。并在实验部分证明它在 LDCT-and-Projection-data 数据集上的优异表现。但这并不能证明我们模型的鲁棒性。所以, 我们从著名的”2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge”(2016aapm) [32] 数据集中随机抽取了两位病人共 1086 个切片组成测试集进行测试。表 5显示了对全部测试集进行测试的性能评价结果的均值和方差, 可以看出在完全没有训练过的数据集上 DDPTransformer 的效果依然是最好的, 和第二高的相比, PSNR 在不同采样下分别高出了 0.037dB(64view) 和 1.155dB(128view)。如图 11所示, 我们随机挑选了一个切片展示其视觉效果, 并在图 12 中列出了他们的性能评价指标, 通过对比我们发现 DDPTransformer 尽管 RMSE 值上并不是最好的, 但在 PSNR 和 SSIM 上效果最好, 并且在不同 view 下重建出的 CT 图整体质量更高。图 13对于在放大的 ROI 下进行观察,

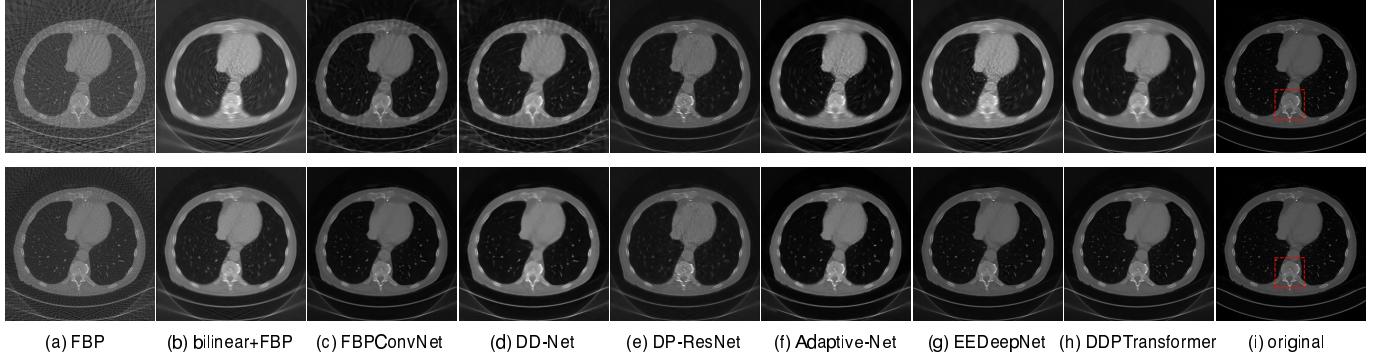


Figure 11: 2016AAPM 数据集上的可视化结果, 第一行 64views, 第二行为 128views

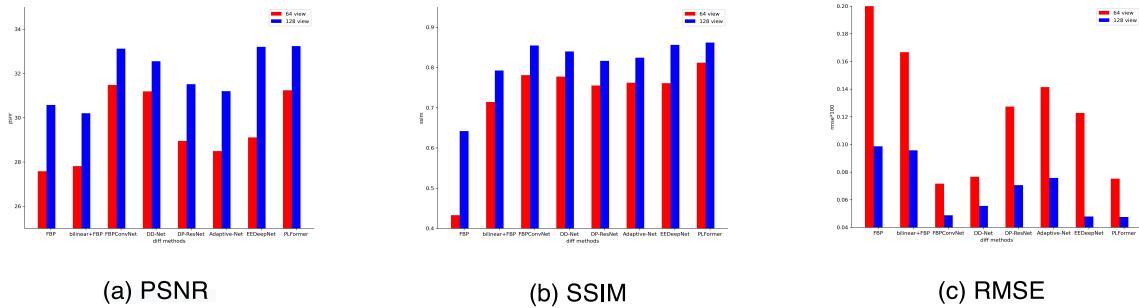


Figure 12: 图 11 的性能评价指标。

DDPTransformer 重建出 CT 图像在条纹伪影上处理的更好。因此，通过在 2016AAPM 数据集上的优异表现，我们验证了 DDPTransformer 的鲁棒性。

method	PSNR	SSIM	RMSE*100	PSNR	SSIM	RMSE*100
64 views				128 views		
FBP	29.9765±0.1209	0.5416±0.0063	0.1350±0.0035	34.4040±0.0949	0.7454±0.0061	0.0446±0.0010
bilinear+FBP	30.4448±0.1365	0.8206±0.0034	0.0922±0.0029	34.3314±0.1436	0.9075±0.0021	0.0374±0.0012
FBPConvNet	36.9814±0.1666	0.8939±0.0032	0.0206±0.0008	42.7887±0.0948	0.9682±0.0004	0.0054±0.0001
DD-Net	30.4719±0.0546	0.8525±0.0015	0.0900±0.0011	33.7510±0.0444	0.9039±0.0010	0.0424±0.0004
DP-ResNet	31.2974±0.1674	0.8426±0.0036	0.0746±0.0029	35.3386±0.1026	0.9120±0.0018	0.0295±0.0007
Adaptive-Net	31.3441±0.2889	0.8659±0.0041	0.0738±0.0050	35.3676±0.2762	0.9284±0.0016	0.0293±0.0018
EEDeepNet	32.4082±0.1377	0.8714±0.0024	0.0578±0.0018	43.9455±0.1820	0.9750±0.0006	0.0041±0.0002
DDPTransformer	37.0181±0.3035	0.9270±0.0041	0.0202±0.0014	45.1001±0.3832	0.9787±0.0013	0.0032±0.0003

Table 5: 2016AAPM 数据集上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。

Acknowledgments

These are acknowledgments. These are acknowledgments.

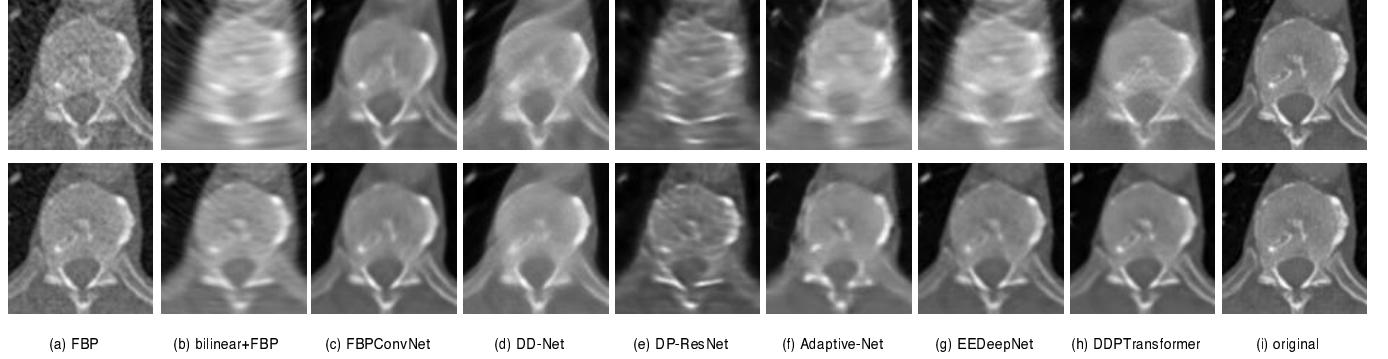


Figure 13: 图 11(i) 中红色框标记的缩放区域

References

- [1] M. Aharon, M. Elad, and A. Bruckstein. *rmk-svd*: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54:4311–4322, 2006.
- [2] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [3] Mihalj Bakator and Dragica Radosav. Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies and Interaction*, 2(3), 2018.
- [4] M. Balda, J. Hornegger, and B. Heismann. Ray contribution masks for structure adaptive sinogram filtering. *IEEE Transactions on Medical Imaging*, 31(6):1228–1239, 2012.
- [5] P. Bao, W. Xia, K. Yang, W. Chen, M. Chen, Y. Xi, S. Niu, Zhou. J, H. Zhang, and H. Sun. Convolutional sparse coding for compressed sensing ct reconstruction. *IEEE Transactions on Medical Imaging*, pages 1–1, 2019.
- [6] J. F. Cai, X. Jia, H. Gao, S. B. Jiang, Z. Shen, and H. Zhao. Cine cone beam ct reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study. *IEEE Trans Med Imaging*, 33(8):1581–1591, 2014.
- [7] E. J Candes, J Romberg, and T Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.
- [8] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021.

- [9] Hu Chen, Yi Zhang, Yunjin Chen, Junfeng Zhang, Weihua Zhang, Huaiqiang Sun, Yang Lv, Peixi Liao, Jiliu Zhou, and Ge Wang. Learn: Learned experts' assessment-based reconstruction network for sparse-data ct. *IEEE Transactions on Medical Imaging*, PP(99):1–1, 2018.
- [10] 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. 36:2524–2535, 06 2017.
- [11] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [12] L. David. Donoho. compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [13] Jia Deng, Wei Dong, Richard Socher, Li Jia Li, Kai Li, and Fei Fei Dept. Imagenet : A large-scale hierarchical image database. *Proc. CVPR, 2009*, 2009.
- [14] 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. *ICLR*, 2021.
- [15] H. Gao, H. Yu, S. Osher, and G. Wang. Multi-energy ct based on a prior rank, intensity and sparsity model (prism). *Inverse Problems*, 27(11):115012–115033(22), 2011.
- [16] 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.
- [17] S. E. Ghrare, Mam Ali, and M. Ismail. The effect of image data compression on the clinical information quality of compressed computed tomography images for teleradiology applications. *european journal of scientific research*, 2008.
- [18] E J Hall and D J Brenner. Hall ej, brenner djcancer risks from diagnostic radiology: the impact of new epidemiological data. br j radiol 85(1020): e1316-e1317. *The British journal of radiology*, 85(1020):e1316–7, 2012.
- [19] Y. S. Han, K. H. Jin, K. Kim, and J. C. Ye. Sparse-view x-ray spectral ct reconstruction using annihilating filter-based low rank hankel matrix approach. In *IEEE International Symposium on Biomedical Imaging*, 2016.
- [20] G. Huang, Z. Liu, Vdm Laurens, and K. Q. Weinberger. Densely connected convolutional networks. *IEEE Computer Society*, 2016.
- [21] 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.
- [22] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

- [23] Eunhee Kang, Won Chang, Jaejun Yoo, and Jong Chul Ye. Deep convolutional framelet denoising for low-dose ct via wavelet residual network. *IEEE Transactions on Medical Imaging*, 37(6):1358–1369, 2018.
- [24] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014.
- [25] 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.
- [26] S. Li, Q. Cao, Y. Chen, Y. Hu, L. Luo, and C. Toumoulin. Dictionary learning based sinogram inpainting for ct sparse reconstruction. *Optik - International Journal for Light and Electron Optics*, 125(12):2862–2867, 2014.
- [27] T. Li, L. Xiang, W. Jing, J. Wen, H. Lu, H. Jiang, and Z. Liang. Nonlinear sinogram smoothing for low-dose x-ray ct. *Nuclear Science IEEE Transactions on*, 51(5):2505–2513, 2004.
- [28] 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.
- [29] Alfred K. Louis and Andreas Rieder. Incomplete data problems in x-ray computerized tomography. *Numerische Mathematik*, 56(4):371–383, 1989.
- [30] A. Luthra, H. Sulakhe, T. Mittal, A. Iyer, and S. Yadav. Eformer: Edge enhancement based transformer for medical image denoising. *arXiv e-prints*, 2021.
- [31] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. 2013.
- [32] Cynthia McCollough and Samuel Armato. Tu-fg-207a-00: Grand challenges in medical imaging and radiomics. *Medical Physics*, 43(6Part35):3759–3760, 2016.
- [33] 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.
- [34] S. Niu, Y. Gao, Z. Bian, J. Huang, W. Chen, G. Yu, Z. Liang, and J. Ma. Sparse-view x-ray ct reconstruction via total generalized variation regularization. *Physics in Medicine and Biology*, 59(12):2997, 2014.
- [35] Matteo Ronchetti. Torchradon: Fast differentiable routines for computed tomography. *arXiv preprint arXiv:2009.14788*, 2020.
- [36] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Springer International Publishing*, 2015.

- [37] E. Y. Sidky and X. Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777, 2008.
- [38] 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.
- [39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [40] 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.
- [41] Z. Wang. Image quality assessment : From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004.
- [42] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. *arXiv preprint 2106.03106*, 2021.
- [43] Jelmer M. Wolterink, Tim Leiner, Max A. Viergever, and Ivana Išgum. Generative adversarial networks for noise reduction in low-dose ct. *IEEE Transactions on Medical Imaging*, 36(12):2536–2545, 2017.
- [44] Q. Xu, H. Y. Yu, X. Q. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-dose x-ray ct reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging*, 2012.
- [45] C. Yang, D. Gao, N. Cong, L. Luo, W. Chen, X. Yin, and Y. Lin. Bayesian statistical reconstruction for low-dose x-ray computed tomography using an adaptive-weighting nonlocal prior. *Computerized Medical Imaging & Graphics*, 33(7):495–500, 2009.
- [46] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, pages 1348–1357, 2018.
- [47] 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.
- [48] Y. Zhang, Y. Wang, W. Zhang, F. Lin, Y. Pu, and J. Zhou. Statistical iterative reconstruction using adaptive fractional order regularization. *Biomedical Optics Express*, 7(3):1015–1029, 2016.
- [49] Yi Zhang, Wei Hua Zhang, Hu Chen, Meng Long Yang, Tai Yong Li, and Ji Liu Zhou. Few-view image reconstruction combining total variation and a high-order norm. *International Journal of Imaging Systems and Technology*, 23(3):249–255, 2013.

- [50] 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.