分类号：TP391
密　级：公开

学校代码：10363
学　　号：2210920119

安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目　基于Transformer网络的稀疏角度CT成像方法研究与应用

论 文 作 者　张庭宇
指 导 教 师　刘进 副教授 刘冬 工程师

学科（专业）　计算机技术
研 究 方 向　人工智能及应用

论文提交日期：2024 年 5 月 23 日

分类号：TP391　　单位代码：10363
密　级：公开　　学　　号：2210920119

题　　目　基于 Transformer 网络的稀疏角度 CT 成像方法研究与应用

英文并列

题　　目　**Research and application of sparse view CT imaging based on Transformer network**

学生姓名：张庭宇

指导教师：刘进（副教授）/刘冬（工程师）

专　　业：计算机技术

研究方向：人工智能及应用

论文答辩日期：2024.05.24

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风，所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：

日期：2024 年 5 月 23 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ___ 年解密后适用本版权书。

本学位论文属于

不保密 □。

学位论文作者签名：张[illegible]　　指导教师签名：[illegible]

日期：2024 年 5 月 23 日　　日期：2024 年 5 月 23 日

基于 Transformer 网络的稀疏角度 CT 成像方法研究与应用

摘 要

CT 是利用人体组织在 X 射线下的衰减差异进行成像的，可获得人体组织衰减分布。主要包括获取扫描断层的投影信息、利用投影数据进行重建得到衰减分布，最后获得对应的 CT 值图像等步骤。为减少 CT 扫描中 X 射线辐射剂量，通常可采用降低管电流和稀疏角度扫描这两种途径。与降低管电流方法相比稀疏角度扫描具有数据量少，计算速度快，适应场景广等优势，但稀疏角度扫描导致投影角度缺失，进而造成重建图像中充满星条状伪影，影响医师的诊断。为此提高稀疏角度扫描下的成像效果，一直是 CT 关键技术研发领域的重要问题。随着海量数据集和广泛样本库的构建，深度学习方法已深入应用于稀疏视角 CT 重建图像的后期处理领域，以其简便的算法实现、对原始数据的独立性以及出色的可扩展性而备受瞩目，成为当前研究的热门趋势。为了解决因投影数据不足造成 CT 图像噪声和伪影加剧的挑战，本文提出了一种创新策略，即利用 Transformer 神经网络架构，开发一系列专为稀疏角度 CT 成像定制的算法，旨在提升图像的质量。主要的研究工作和创新点体现在：

（1）稀疏角度 CT 图像图像退化复杂，单一图像域难以取得满意的效果。为此，本课题研究了联合投影域及图像域处理策略，建立基于 Transformer 块的混合域网络的重建流，以促进稀疏角度 CT 成像质量。该方法采用图像域—投影域—图像域三阶段混合处理流程，通过多阶段信息的联合互补提高成像质量。此外，针对不同阶段数据噪声伪影特点设计不同的 Transformer 模块，以实现差异化的处理，更进一步，算法采用可微分的解析重建和投影运算，建立投影域与图像域数据的转换，最后，在此基础上，实现高质量的端对端 CT 图像。结果显示，所建立的网络模型能够有效地压制各种层次的伪影与噪音，并最大限度地保持原始影像的细节。量化结果表明重建图像质量更高。

（2）为了更好的解决稀疏角度 CT 图像成像过程中边缘细节难以修复的问题，论文借助于生成对抗学习的思想，提出了一种面向稀疏角度 CT 图像伪影抑制的 Transformer 生成对抗网络。在生成器网络中，设计基于 Transformer 的编码器-解码器结构，取代 U 形网络中的卷积滤波器，提取高质量的全局信息。为了保留图像细节信息、抑制伪影特征，生成器使用跳跃连接结构，使得底层和高层特征可以进行交流和融合。通过增加判别器，构建生成对抗网络，以优化处理网络的性能。实验表明，面向稀疏角度 CT 图像处理的 Transformer 生成对抗网络在在抑制噪声伪影的同时，可有效的保留了 CT 图像的结构和纹理信息。

（3）结合（1）和（2）的工作，进一步，将混合域处理与生成对抗学习策

略相结合，构建基于双域 Transformer 生成对抗网路的稀疏角度 CT 成像算法。在投影域子网络，设计专用于投影数据处理的 Transformer 网络，对投影数据缺失部分进行修补与复原。在图像域子网络，由于在投影域子网络已去除部分伪影，可进一步简化第四章的 Transformer 模块，从而减少网络计算量提高推断速度，同时图像域的生成对抗网络，可对处理后结果和目标图像进行判断，能不断优化投影数据处理网络及图像数据处理网络的性能。实验表明，融合后的 Transformer 网络可获得高质量的 CT 图像。

最后，对本文中使用的三种不同 Transformer 网络进行了总结，并给出现阶段稀疏角度 CT 图像成像算法仍需面临的挑战，同时针对这些问题讨论了稀疏角度 CT 图像处理的未来研究趋势。

关键词：稀疏角度，Transformer，图像复原，混合域，生成对抗网络

Research and application of sparse view CT imaging based on Transformer network

ABSTRACT

CT imaging is conducted by utilizing the differences in X-ray attenuation among various human tissues, thereby enabling the acquisition of the attenuation distribution within the body. After obtaining the projection data from the scan, the process primarily involves reconstructing the attenuation distribution utilizing this projection data, ultimately yielding the corresponding CT value image. In order to reduce the X-ray radiation dose in CT scan, two approaches are usually used: reducing tube current and sparse view scanning. Compared with the method of reducing the tube current, sparse view scanning has the advantages of less data, faster calculation speed, and wide adaptation to the clinic scene. However, sparse view scanning leads to the loss of projection data, which results in the reconstruction image full of steak artifacts, affecting the doctor’ s diagnosis. Therefore, improving the imaging quality under sparse view scanning has always been an important issue in the research and development of key CT technologies. With the establishment of extensive data sets and comprehensive sample libraries, deep learning technology has been extensively applied in image processing following sparse view CT imaging. This approach boasts the advantages of algorithm ease of implementation, independence from original data, and excellent scalability, thus representing the primary research direction in recent years. To solve the problem of serious noise artifacts in CT images due to the reduction of projection data, this paper proposed a series of sparse view CT imaging methods with the advantage of Transformer neural network to improve imaging quality. The main work and research innovation are as follows:

(1) The image degradation of sparse view CT images is more complex, and it is difficult to obtain satisfactory results in a single image domain. Therefore, this dissertation was presented the processing strategy of combined projection domain and image domain, and establishes the reconstruction flow of hybrid domain network based on Transformer block to promote the imaging quality of sparse view CT. The method adopts a three-stage hybrid processing process of image domain, projection domain and image domain, and improves the image quality by combining and complementing multi-stage information. In addition, different Transformer modules

are designed according to the featrues of noise artifacts at different stages to achieve differentiated processing. Furthermore, the method adopts differentiable analytic reconstruction and projection operation to establish the conversion of projection domain and image domain data, and finally achieve end-to-end sparse view CT high-quality imaging flow. The experimental results show that the network model can initially suppress artifacts and noise at different sparse view levels CT images, and retain the organizational details of the original image as much as possible while removing artifacts and noise. The quantitative results show that the reconstructed image has higher quality.

(2) In order to better solve the problem of tissue edge details restoration in the imaging process of sparse view CT images, this dissertation introduce Transformer network generator into the generative adversarial network framework from the idea of generative adversarial learning, and proposes a Transformer based generative adversarial network for sparse view CT image artifact suppression. The network consists of two main parts, the generator and the discriminator. In the generator network, an encoder-decoder structure based on Transformer is designed to replace the convolutional filter in the U-shaped network and extract high quality global feature information. In order to recovery image details and suppress artifact features, the generator use skip connections that allows the low-level and high-level features to communicate and merge. By adding a discriminator, a generative adversarial network is constructed to optimize the performance of the artifact suppression. Experiment results show that Transformer for sparse view CT image processing can effectively retain the detial and texture information of CT images while suppressing noise artifacts.

(3) Combined with the work in (1) and (2), we further combine hybrid domain processing with generative adversarial learning strategies to build a sparse view CT imaging algorithm based on dual-domain Transformer to generate adversarial networks. In the projection domain sub-network, Transformer network is designed for projection data processing to repair and restore the missing data of projection. In the image domain sub-network, since some artifacts have been removed in the projection domain sub-network, the Transformer module in Chapter 4 can be further simplified to reduce the network complexity and improve the inference speed. Meanwhile, the generative adversarial framework in the image domain can discriminate the processed results and target images. It can continuously optimize the performance of projection domain sub-network and image domain sub-network. Experimental results

demonstrate that the hybrid Transformer network is capable of achieving high-quality CT images, exhibiting superior performance in artifact removal, structural preservation, and visual observation. This advancement can significantly enhance the quality of sparse view CT imaging, generating diagnostic images that are deemed acceptable by clinicians.

In conclusion, this dissertation summarizes three distinct Transformer-based networks utilized in the study. Furthermore, it presents the ongoing challenges encountered by sparse view CT image imaging methods at the current stage and discusses potential future research trends in sparse view CT imaging in light of these issues.

KEY WORDS: Sparse view, Transformer, Image restoration, Mixed domain, Generative adversarial network

目 录

第 1 章 绪论

1.1 研究背景及意义

1895 年，德国科学家伦琴首次发现了能够穿透物体的 X 射线，从此开创了使用 X 射线进行医学诊断的放射学研究[1]。1972 年首台计算机断层扫描（Computed Tomography, CT）设备被研制成功，并辅助医生完成临床的诊断。如今，CT 成像技术越来与成熟，使用也越来越广泛，已成为医学诊疗中不可或缺的一部分[2]-[3]。然而 X 射线被人体部分吸收，会对组织器官造成一定伤害，并可能导致遗传或癌症疾病的产生[4]。为此，国际辐射防护委员会曾经建议，在不影响图像诊断的前提下，要尽可能降低 X 射线剂量。在 2023 年发布的最新的《中国肺癌低剂量 CT 筛查指南》中提到，低剂量的 CT 扫描（Low Dose CT，LDCT）是目前肺癌筛查的重要方法，能够极大的提升早期肺癌的检出率，Meta 机构的分析显示 LDCT 筛查可使肺癌死亡率显著降低 16%-21%。因此，我国专家通过研究近年来国内外 LDCT 筛查进展，在 2018 年版中国肺癌 LDCT 筛查指南的基础上，修订并完成了 2023 版的中国肺癌筛查指南，同时发布了《中国肺癌低剂量 CT 筛查指南（2023 年版）》文件。

CT 成像原理是利用人体组织对 X 射线束的不同吸收程度，进行人体组织的衰减成像。目前，CT 成像系统大都采用 X 线数字化成像技术，基本原理包括获取扫描层面的数字化信号、获取扫描层面各个扫描角度的 X 线吸收系数、获取组织断层的 CT 图像等方面。通常，减少 CT 扫描中 X 射线辐射剂量有两种主要途径，降低管电流和减少扫描角度。降低管电流不可避免地导致较低有效光子和高投影数据（或正弦图）噪声，从而导致降低重建图像的质量。目前已有大量针对低管电流扫描模式的 CT 成像方法研究，并取得优异成绩。而减少扫描角度作为另一种有效降低辐射剂量措施，与降低管电流方法相比具有一定优势，如，稀疏角度扫描更简单容易执行，适应场景多，数据量少，计算速度快。但稀疏角度扫描会导致投影数据的采样缺失，进而造成重建图像中充满星条状伪影，影响医师的诊断。

改善稀疏角度扫描成像效果有以下三种方法：一、从 CT 投影信息视角考虑：由研发人员设置专门的投影信息恢复和处理方法，对原始数据以及经过对数变换后的投影信息进行恢复和处理，以恢复缺失投影信息，并补充数据的缺失，以增强原始数据的稳定性。由于投影数据灵敏度较高，对信息的还原很容易产生误差，从而造成重建图像的错误信息。二、从 CT 图像数据角度出发：通过采用图像处理技术，以抑制高稀疏角度 CT 图象上的条状伪影，可得到高信噪比和高对比图象。在各种扫描系统、扫描模式、扫描部位和重建方式下，由于稀疏角度与 CT

图象的伪影特征差别较大，因此不易实现整体建模。在过程中，也易发生过平滑、细节信息丢失及对比度下降等情况。三、从图像重建算法角度出发：近年来，众多迭代重建技术应运而生，并在实践中展现出卓越的效能，特别是那些融入了先验知识约束的统计迭代方法。然而，这些算法普遍存在的挑战在于，由于涉及复杂的超参数设置，难以实现自动化调参与优化；此外，它们的复杂性较大，需要进行大量的反复迭代；而且，由于先验信息的不确定性和泛化性较低[5]，使得迭代重建在临床应用场景下难以充分发挥其价值。尽管目前 CT 成像技术仍存在许多挑战，但它将是未来发展的重要方向，特别是在稀疏角度 CT 成像方面，将会带来更多的突破。

随着科技的进步，基于数据驱动的学习型算法已经取得了巨大的进步，它们可以更好地处理海量数据，从而为医学图像重建提供更有效的解决方案。尤其是在影像大数据环境下，深度学习技术已经成为稀疏角度 CT 成像的主流，如，Xu 等人利用 U-Net 构建的残差学习网络，可以有效地去除条纹伪影，从而极大地提升 CT 成像的精准度和准确性，但该网络在局部区域的伪影去除中效果不佳[6]；为解决深度学习（Convolutional Neural Network，CNN）感受野有限的问题，Google 机器翻译团队提出了一种基于注意力的编码器-解码器结构的 Transformer 网络，并在许多计算机视觉任务中取得了巨大成功[7]。诸多研究表明，Transfomer 网络模型具有较少的循环结构，可大幅度提高并行计算效果，提高计算速率。为此，论文将借助 Transformer 网络的优势，建立多种特征学习模型，实现对稀疏角度 CT 重建图像中不同特征信息的约束和处理，提高最终成像效果。通过设计新型成像算法，可不增加额外硬件，提高诊断收益，具有开发成本小，经济价值高的优势。通过对 Transfomer 网络模型的研究，设计供临床应用的新型稀疏角度 CT 重建算法，以促进稀疏角度 CT 成像技术的临床应用，缓解 CT 诊疗设备的损耗降低成本，为检查者减少辐射伤害，对疾病的筛查和诊断具有重要意义。

1.2 国内外研究现状

1.2.1 传统的 CT 成像方法研究现状分析

（1）投影域复原算法

投影域复原类是对投影数据进行滤波复原及处理，以去除其中的噪声伪影，及修复数据中的缺失，令其接近正常角度的投影数据，再用处理后的投影数据重建出 CT 图像的一种方法。稀疏角度 CT 投影数据复原主要是对缺失投影数据进行插值或修补，此外还结合一致性处理，滤波处理等操作，以降低重建后的 CT 图像的伪影。代表性的研究如，Li 等为减少稀疏角度扫描成像中的条状伪影，提出了一种字典学习的投影数据修补方法[8]。Shen 等为修补投影数据而提出了

一种改进型稀疏学习及表示模型，可在提高重建效果减少图像分辨率的损失[9]。Zhang 等提出了一种噪声特性相关性的投影域字典学习方法，以降低噪声提高成像质量[10]。Karimi 等利用三维投影数据的冗余性，构建联合稀疏学习模型，对投影数据进行编码解码处理，得到了较好的成像效果[11]。

（2）图像域后处理算法

直接解析重建后的稀疏角度 CT 图像可能会出现严重的条纹伪影，因此这里的图像域后处理方法，是通过相关图像处理技术以消除减缓这些伪影的干扰，提高 CT 图像的质量。对重建后的图像进一步处理，不仅拥有丰富的算法基础，而且具有较强的应用前景。具有代表性的工作有：Chen 等将钝化滤波融入到字典学习中并对 CT 图像进行处理，可改善图像质量提高肿瘤识别率[12]。Chen 等在小波域中使用区分性的伪影抑制字典来减少伪影，在图像域中使用传统字典降低噪声和伪影，最终提高了腹部 CT 图像质量[13]。Ha 等利用海量图像块构建 CT 图像块数据库，并将图像块相似性作为先验知识，对待处理图像块进行匹配和特征约束，从而恢复 CT 图像中的结构信息，提高组织的辨识度[14]。Green 等提出一种基于局部窗口约束的非局部均值降噪方法，在降低噪声和抑制伪影方面拥有较好的效果[15]。Cui 等借助形态分量分析的思想，提出了一种可自动生成自适应的区别字典学习方法，实现无监督有区分性的稀疏表示，能很好的抑制稀疏角度 CT 图像中的结构伪影，而且避免对偶字典设计时人工选取原子的不足[16]。Karimi 等利用稀疏学习方式，建立双字典，并通过伪影字典编码系数与干净字典编码系数之间的关系，可改善的稀疏角度 CT 图像质量[17]。

（3）迭代重建算法

由于计算技术的进步，迭代方法受到了广泛的关注。与传统解析重建方式相比，迭代方式在处理稀疏角度 CT 图像中表现出了更高的精度，并且可以轻松地添加限制因素。迭代重建主要分为代数重建算法和统计迭代算法。在数据迭代重构领域，使用最高后验概率密度（MaximumA Posteriori， MAP）作为参数是一种非常流行的方法，它通过对目标函数的正则化，从而大大增加了算法的稳健度，同时也能够有效地抵消外部干扰，并且能够维护原始的数据。Sidky 和其他研究人员发明的 TV 约束迭代重建技术，能够显著降低 CT 图像中的伪影，从而极大地改善了 CT 扫描的精确度和准确率[18]。随着时间的推移，越来越多的改良型 TV 重建算法不断涌现。与传统的 TV 类约束相比，特征字典在这方面发挥了更大作用，通过学习，可以获得更多的结构特征，从而有效地捕捉图像中的细微差别。例如， Xu 等通过将约束条件纳入迭代重建框架，构建了两种不同的重建目标函数：全局字典学习和自适应字典学习，并取得了显著的改进[19]。在这之后系列新型的字典学习方法被提出，如 Lu 等提出的对偶字典[20]，Liu 等提出的三维特征

字典[21]，Zheng 等提出的快速聚类字典[22]等，都在一定程度上提高了稀疏角度 CT 图像重建效果。Bao 等首次将卷积稀疏框架用于 CT 重建中，实现对稀疏角度 CT 的噪声和伪影抑制[23]。然而，由于存在大量的超参数，以及算法复杂度高，从而限制了它的临床应用潜力。

1.2.2 基于深度学习的 CT 成像方法研究现状分析

深度重建算法中，根据其应用方式，可以将这类成像技术划分为两大类：投影及图像域处理类算法和图像重建类算法。

（1）投影及图像域处理类算法

深度学习的投影数据处理方法，代表性工作如：Isola 等通过条件生成对抗网络，补全稀疏角度投影数据，可有效减少稀疏采样导致的条状伪影[24]。Li 等提出了一种面向弦图修复的生成对抗网络，以补偿投影数据中的损失，最终通过联合 U-net 的生成器与图像块操作的判别器共同提高成像质量[25]。Meng 等联合有监督与无监督的双子网络来捕捉关键特征，实现在抑制投影数据中噪声的同时保持衰减特征，重建后 CT 图像质量得到明显的改善[26]。Zainulina 等提出一种自监督学习方法，并汲取相邻投影数据冗余信息，实现无参考低剂量投影数据的处理[27]。Choi 以自监督方法直接处理三维的低剂量 CBCT 投影数据，可在降低噪声同时提高结构信息的复原效果[28]。Wu 等提出一种自监督的坐标投影重建网络，该网络可通过原始解的重投影策略，生成全角度投影数据，能大幅度提高稀疏角度 CT 重建质量[29]。Wang 等人对不完成投影数据的成像进行了详细的综述，其中涉及大量投影数据补全及复原的网络模型[30]。

随着大数据集和大样本库的不断发展，深度学习已经成为一种广泛的、有效的方法来实现对低剂量 CT 图像的处理。代表性的工作如：Yang 等提出将三维残差网络用于低剂量 CT 图像后处理，以降低重建图像中的噪声[31]。Kang 等在利用方向波变换对低剂量 CT 图像进行预处理后，再使用 U-net 处理，可进一步去除低剂量 CT 图像中的噪声[32]。Chen 等将残差编码解码 RED 网络应用于低剂量 CT 图像降噪，可在保持图像细节与抑制噪声直接取得较好平衡[33]。Yang 等提出使用感知损失替代均方误差损失，通过多个网络对比实验发现该损失函数能保留更多的关键细节并增强对比度[34]。Shan 等以生成对抗网络框架为基础，通过渐进式去噪来实现低剂量 CT 图像的模块化自适应处理，并允许放射科医师个性化设置网络映射的深度[35]。Liu 等通过实验表明，采用高质量迭代重建图像的样本数据训练网络，可进一步提升网络性能[36]。Li 等提出一种自注意力机制的 3D 卷积神经网络，并采用自编码监督的感知损失来提升网络性能，以实现低剂量 CT 图像去噪[37]。Choi 等采用增加感知域的方式来统计 CT 图像的空间变化，并融合层间相关信息，构建了一种 StatNet 网络，以实现低剂量 CT 图像的快速

复原[38]。Liu 等将残差模型引入到 U-net 架构中，并采用噪声功率谱作为训练效果的评判标准之一，使得处理后的低剂量 CT 图像纹理更自然[39]。Feng 等利用一种双残差网络实现投影及图像域特征互补，建立噪声抑制和图像误差减小的回归模型，取得了较 RED 模型更好的处理效果[40]。Huang 等考虑到不同剂量下图像的表征差异，并结合辐射剂量预估计与通道特征变换策略，设计了一种自适应先验选取的剂量意识网络，并取得较好的复原效果[41]。Liu 等融合空洞卷积、注意力机制及权重卷积稀疏编码，设计了一种多尺度卷积编码网络，可提升复原后低剂量 CT 图像质量[42]。

（2）图像重建类方法

近年来，深度学习已经被广泛应用于逆问题求解和 CT 成像，例如 Kelly 等人利用深度神经网络进行不完全投影数据重建，从而大大降低了重建的复杂性，提高了重建的效率[43]。Wu 等以预训练网络为先验函数，融入迭代重建框架中，实现低剂量 CT 的高质量重建[44]。Adler 等利用逆问题的对偶关系，以神经网络来替代重建目标函数并进行求解，成像效果及效率上具有较大优势[45]。Ge 等提出了一种解析域变换的端到端网络 ADAPTIVE-NET，可在较少的硬件资源下实现投影数据快速重建[46]。Mehmet 等以深度生成网络模型为先验，实现少噪声少伪影的迭代重建[47]。Ding 等在近似向前向后分离算法框架下，以数据驱动的深度神经网络为约束，建立迭代重建模型，结果表明在低剂量 CT 成像中要优于图像后处理网络[48]。He 等以生成密度梯度为先验，建立评分式生成网络并引入迭代重建中，在采用模拟退火方法优化后，重建具有较强的降噪与细节保持能力[49]。Xia 等通过联合空间卷积与图卷积来分析流行空间的非局部拓扑特征，并将其融入迭代重建中，可实现少标签下重建 CT 图像的噪声抑制[50]。为提高有限角度 CT 图像重建质量，Hu 等从投影数据补偿角度出发，设计了一种单步投影数据误差校正的对抗网络，实现了不完全投影数据的高质量重建[51]。Liu 等将卷积稀疏表示的先验知识模型应用于统计迭代重建框架中，以投影残差成分为约束对象，形成一种深度残差约束重建算法，具有较好的鲁棒性和可解释性[52]；Kang 等进一步通过引入快速滑动窗口的非局部相似性质，构建基于组块稀疏先验约束的重建目标函数并以网络形式展开，能更好的抑制条状伪影并恢复组织细节[53]；深度学习类迭代重建算法为当前研究的热点，但大多未形成真正意义的深度重建，此外网络可解释性差，难以保证其重建稳定性，还处在临床落地应用的初期阶段。

近年来，为了解决 CNN 感受野有限的问题，人们提出了一种基于注意力的编码器-解码器结构 Transformer 网络[7]，并在许多计算机视觉任务中取得了巨大成功。在 Transformer 网络思想基础上，已有一些新的 CT 图像处理及重建策略被提出，如，Wang 等提出 DuDoTrans 的稀疏角度 CT 重建方法，该方法使用 Transformer 网络同时处理投影域及图像域数据，同时增强投影域数据的注意力，

以减缓解析重建后图像中的伪影[64]；Li 等提出一种 DDPTransformer 网络的稀疏角度 CT 图像重建方法，该方法对投影域和图像域数据使用平行化的 Transformer 网络处理，可联合促进重建质量的提升[55]；Sun 等通过设计一种轻量化的 Transformer 网络，并通过提高双域数据的注意力来抑制重建图像中的伪影，提高组织细节的恢复能力[56]；Pan 等借助 Swin Transformer 网络优势，提出一种多域合成式的 Swin Transformer 网络架构，实现了稀疏角度 CT 图像的优质重建[57]。Li 等使用 Swin-Transformer 块作为投影（残差）域和图像（残差）领域子网络中的主要构建块，对投影和重建图像的全局和局部特征进行建模[58]。Du 等将 Transformer 与生成对抗网络网络结合，成功将 Transformer 插入生成对抗性网络进行图像重建[59]。通过调研发现，Transformer 网络在 CT 图像处理及重建中具有明显的优势，越来越多的基于 Transformer 方法被提出，借助其卓越的注意力及感知能力有望进一步提高 CT 图像重建性能[60-69]。为此，课题通过 Transformer 网络的研究，设计针对稀疏角度 CT 投影或图像处理网络，建立多种先验特征学习模型，以实现对稀疏角度重建图像中噪声的抑制，提高稀疏角度 CT 的成像效果。

1.3 论文主要工作

稀疏角度 CT 图像成像过程复杂，呈现的伪影水平具有较大差异，因此，本文在参考我国有关稀疏角度 CT 图像处理方法的基础上，以 Transformer 网络作为突破口，提出了三种基于稀疏角度 CT 的成像方法：一、联合投影域及图像域处理策略，建立基于 Transformer 块的混合域网络，并实现多阶段的稀疏角度 CT 图像重建；二、在生成对抗学习框架下，设计了面向稀疏角度 CT 图像处理的 Transformer 生成器，并借助于判别器进行生成对抗学习，以提升网络性能，实现稀疏角度 CT 图像高质量处理；三、结合多阶段处理的及生成对抗学习策略，设计多阶段数据处理的 Transformer 生成器网络，实现稀疏角度 CT 图像的优质重建。本文的主要工作如下所述：

（1）稀疏角度 CT 图像中呈现的伪影水平具有较大差异，单一图像域难以取得满意的效果。为此，课题研究了联合投影域及图像域处理策略，建立基于 Transformer 块的混合域网络的重建流，以促进稀疏角度 CT 成像质量。此部分研究将包括三部分：A. 投影域 Transformer 网络设计：设计专用于投影数据处理的 Transformer 网络，对投影数据缺失部分进行修补与复原，以增加原始数据的信息量及可靠度，使后续重建图像质量提升。B. 图像域 Transformer 网络设计：设计专用于图像数据处理的 Transformer 网络，对稀疏角度 CT 图像中的伪影去除，进一步提高重建图像视觉效果，减缓诊断干扰。C. 借助解析重建算法，实现基于 Transformer 块的混合域网络的稀疏角度 CT 重建：按照解析重建的数据处理

流程，及投影域图像域处理的 Transformer 网络，完成从稀疏角度投影数据到高质量重建图像过程的数据映射。实验结果显示，此网络可以有效压制各种程度的伪影与噪声，并最大限度地保持原始图像的结构细节。

（2）为了高效的抑制稀疏角度 CT 图像中的伪影，课题在生成对抗网络框架中引入 Transformer 模块，设计了一种面向稀疏角度 CT 图像伪影抑制的 Transformer 生成器，形成稀疏角度伪影抑制的生成对抗网络。在生成器网络中，设计基于 Transformer 的编码器-解码器结构，取代 U 形网络中的卷积滤波器，以提取高质量的全局信息，提高图像中伪影去除效果。为了保留图像细节信息、提高算法鲁棒性，生成器中包括线性层、前馈网络、多头转换注意力模块和门控直流前馈网络等。在此基础上增加判别器，构建生成对抗网络，以优化处理网络的性能。判别器采用简洁的 9 层网络结构，使整个模型更容易优化，训练效果更好。实验表明，面向稀疏角度 CT 图像伪影抑制的 Transformer 生成对抗网络，具有更好的推断能力，在抑制噪声伪影的同时可有效的保留了 CT 图像的结构和纹理信息。

（3）结合（1）和（2）的工作中的优势，进一步，将多阶段混合处理的及生成对抗学习策略相结合，在投影域处理阶段，设计专用于投影数据处理的 Transformer 网络，对投影数据缺失部分进行修补与复原，增加生成对抗网络从而处理边缘细节伪影。在图像域处理阶段，对获得的 CT 图像进一步处理。利用前面所提的 Transformer 模块的优势，可提高各阶段的数据处理精度。同时，为提高网络训练性能及减少计算量，仅在图像域处理阶段增加判别器，以最终输出的结果为导向，优化双域的网络。实验表明，融合后的多阶段生成对抗 Transformer 网络可获得高质量的 CT 图像，在伪影去除方面优于其他两种方法。

1.4 论文章节安排

论文主要有 6 章内容，具体安排如下：

第 1 章，简要地阐述了 CT 成像技术的研究背景和意义，对目前国际上的研究状况做了简要的介绍，并对论文的主要工作做了简要的介绍，最后对每一章的重点进行简要总结。

第 2 章，对稀疏角度 CT 成像过程及产生的伪影噪声进行了概述，主要介绍了卷积神经网络及 Transformer 网络中不同模块的概念，同时介绍了现有基于 Transformer 网络模型的稀疏角度 CT 图像方法。同时还介绍了论文使用的数据集以及评估指标等。

第 3 章，基于 Transformer 块的混合域网络稀疏角度 CT 成像研究。详细介绍了混合域网络处理流程，包括投影域 Transformer 网络结构和图像域 Transformer 网络结构设计，以及借助于可微分的解析重建算法，建立投影域与

图像域数据的转换，最终实现完整的端到端的稀疏角度 CT 优质成像。通过一系列试验，对所提出的算法进行了验证，并对其超参数作了初步分析。

第 4 章，为了实现稀疏角度 CT 图像的伪影抑制，设计了专用与伪影抑制的生成对抗网络。网络中生成器采用专用的 Transformer 网络，通过提高感受野及信息的融合，增强生成器的推断能力，并且介绍了判别器以及网络结构。通过相关实验，分析网络模型与生成对抗网络结合的伪影去除效果。最后通过消融和参数分析实验，归纳出该方法的优点与不足。

第 5 章，基于前期的研究成果，本章整合了混合域处理技术和生成对抗学习方法，构建了一种新型的 Transformer 网络，专门应用于投影和图像数据的处理。该章详细阐述了投影域的网络架构以及图像域的网络设计，并通过在多种稀疏角度数据集上的实验，评估并对比了算法的效能，以此证明了所提网络模型的优越性。此外，通过实验验证和模型比较，深入探讨了不同 Transformer 网络的性能差异。

第 6 章，对论文的全部内容进行了全面的总结和未来展望。本章系统地回顾和整理了本文的研究工作，明确了提出的算法的优点和局限性，并对后续可能的研究方向提出了前瞻性的观点。

第 2 章　CT 成像及卷积神经网络理论

在扫描角度减少的情况下，如何获得高品质的影像，这是稀疏角度 CT 成像研究中的一种重要环节。其中，使用特定深度学习网络来提高稀疏角度 CT 图像重建质量，是当前研究者们采用的主要方法，具有成本少、效率高质量好等优势。针对 CT 图像重建过程中，因扫描角度不足和投影采样不完全而引发的严重条纹伪影和斑点噪声问题，本研究以 Transformer 架构为重点，探索适用于稀疏角度 CT 图像重建的创新算法。章节开始，将详细解析造成稀疏角度 CT 图像伪影的根本原因，并概述现有的主要稀疏角度 CT 图像处理技术的基本原理。接着将深入探讨几种运用深度学习网络的典型 CT 图像处理策略，最后指出实验数据集的选择以及所依据的评价标准。

2.1 CT 成像原理

CT 成像技术基于 X 射线对身体各组织的差异性穿透，它首先收集由 X 射线穿过人体后的投影数据。这一过程中，探测器捕获到的 X 射线信号被转化为光电信号，进而生成投影数据。接下来，运用特定的重构算法，对这些数据进行处理，从而生成 CT 图像。重建后的图像以断层的形式展现，每个像素的位置大致对应着人体内部不同区域的组织密度信息，为观察人体的解剖构造提供了直观的视觉展示。CT 成像的这一原理可参照图 2-1。

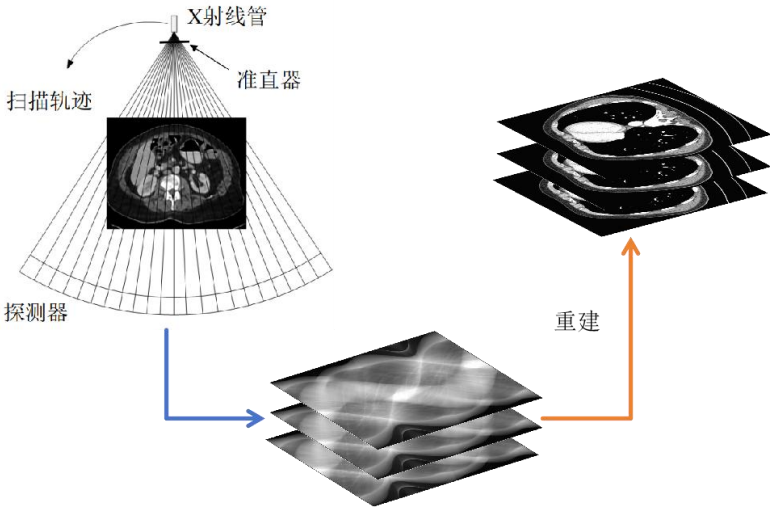


图 2-1 CT 成像过程原理示意图

Fig. 2-1　CT schematic diagram

重建后的 CT 图像能精细体现人体各种组织结构的区别。与常见的自然图像相异，CT 图像呈现为单一通道的灰阶图像，包含 2000 多级不同的 CT 值强度。由于人类视觉系统往往难以辨别微小的亮度变化，因此，在 CT 扫描中，医生通常会调整显示窗口的宽度和中心以区别那些 CT 值相差甚微的组织。窗口宽度定

义了图像上显示的 CT 值跨度，直接影响到图像对比度和细节可见性。若 CT 值差异不大，选用较窄的窗口宽度有助于提升图像的分辨率。窗口位置则是窗口宽度边界值的均值，即使窗口宽度固定，调整窗口位置也会改变 CT 图像的显示效果。为了观察特定的组织或构造，一般会依据这些组织的平均 CT 值来设置窗口位置。

CT 扫描获得的图像主要受这五种因素影响，如表 2-1 所示：

表 2-1　CT 图像中的噪声和伪影

Table 2-1　Noise and artifacts in CT images

伪影种类	伪影产生原因
设备固有特性的影响	设备自身的结构问题可能导致杯状或帽状的伪影，同时，部分容积效应会催生短条纹状的影像失真。
设备运行故障的影响	如， X 光管或检测器的位置不正确会导致环状伪像，而检测器的损伤也会造成这种问题。
患者体内因素	如果病人身体中存在着金属或者其它高密度的物体，会引起射线的不均匀分布，从而产生放射图像的畸形。
患者生理活动的影响	受检者在扫描期间的位置变动或是呼吸等自然生理动作，可能导致器官移动，从而产生运动伪影。
X 射线剂量的制约	如果射线剂量不足，可能导致探测信号弱化，进而形成条纹、斑点噪声或者带状的图像缺陷。

本论文主要针对由于扫描角度的缺失导致的条形伪影和斑点噪声问题。为了改善稀疏角度扫描带来的 CT 图像质量下降问题，论文设计不同算法解决方案。

2.2 卷积神经网络

多层神经网络架构的深度学习技术能揭示数据的多层次特性，生成富含语义的表示形式。其中，卷积神经网络作为一种关键的 DL 实现手段，其独特之处在于利用隐藏层内的神经元单元高效捕获训练样本的大量局部特征，进而对未知测试数据进行推理，展现出在机器视觉任务中对平移、缩放和倾斜等变换的强健性。CNN 逐级进行图像特征的提取和表述，通过卷积核在局部区域的连接，从图像中挖掘出视觉模式，随后将早期网络捕获的底层特征与高层网络结合，构建出更复杂的语义描述。在医学图像分析领域，一个典型的 CNN 模型通常包含卷积层、池化层以及反卷积层，这三者构成了 CNN 的基础结构。

2.2.1 卷积层

卷积神经网络是一个完整的、可学习的卷积神经网络（过滤器）。在对大规模、复杂的图像进行分析时，为了降低运算量，人们往往将卷积神经网络中的神经网络和上一层的神经网络联系起来。这种联系的面积被称为神经细胞的受体区

域。在卷积层中，该层本身的感知区域实际上是一个超级参数，即它所包含的卷积核的空间大小，它的宽度和高度都很小，而且是不规则的。但是，在深度（信道数目）的维数上，总是与输入的图象信息相吻合。也就是说，这种联系是本地的（宽度和高度），而在深度上是整体。通常，决定神经网络数目和使用模式的超参量有四种，从而决定了输入信号的维度：卷积核大小（Kernel Size）、深度（Depth）、步长（Stride）与零填充（Zero-Padding）。

(1)卷积核大小

卷积核心的尺寸就是前面所说的卷积核心的空间大小，这个大小将会对卷积网络的感知范围产生直接的影响。每个卷积层中的卷积核心尺寸大致相同，其取值从 3 到 11，个别情形下可以在同一卷积层中配置多种尺寸的卷积内核；这样就可以获得多个感知区域的取样。

(2)深度

输入数据的深度，又称为输入数据的信道数（Input Channels)，其自身也是一个超级参数，但是通常只需要选择上一层的深度就可以了。在建模过程中，用户可以自主选择要抽取的维度，也就是要抽取几个样本（每个样本中含有一个深度卷积核）的样本数量。

(3)步长

卷积核需要确定它每一次向右边和往下运动的长度，即为步长。卷积核（过滤器）在步长为 1 的情况下，卷积核向右侧移动 1 个像素，滑动到右侧移动 1 个单元。在步长为 2 的情况下，卷积核心在每一次移动两个象素，在向右侧移动两个象素，依次类推。

（4）零填充

零填充是一种允许保留原始输入大小的技术，保证输入和输出维度一致。可以在每卷积层的基础上指定。对于每个卷积层，就像定义卷积核和卷积核的大小一样，可以指定是否使用填充。

2.2.2 Batch Normalization

在神经网络的训练进程中，网络参数随着梯度下降策略的实施持续演变。每次穿越网络的卷积层，数据的分布特性都会呈现独特的转型。这种因参数更新所引发的输入数据分布的偏移现象，会倾向于将数据拉向特定的方向，进而对网络训练的效率产生负面影响。在 2015 年，由 Sergey 提出批归一化层（Batch Normalization, BN）的操作[70]，可对所有输入的卷积层特征图完成归一化，实现数据中均值与方差的统一，之后利用可训练参数进行修改，使得数据和特征拥有

类似的分布特性，这一操作还能有效缓解深度神经网络中梯度爆炸的问题。

2.2.3 池化层

池化层是深度学习中常用的一种层级结构，它可以对输入数据进行降采样，减少数据量，同时保留重要的特征信息。池化层通常紧跟在卷积层之后，可以有效地减少数据量和计算复杂度，提高模型的训练速度和泛化能力。池化层的结构与卷积层类似，它也由多个固定的滤波器组成，每个滤波器对输入数据进行池化操作，得到一个输出特征图。不同的是，池化层的池化操作通常不使用权重参数，而是使用一种固定的池化函数，例如最大池化、平均池化等。结构如图所示：

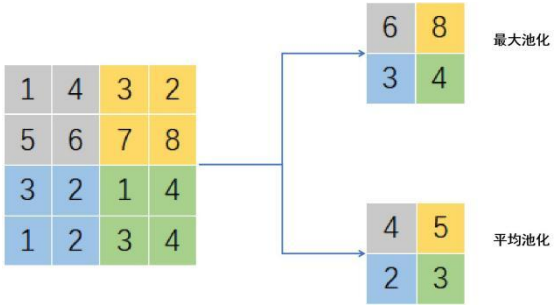


图 2-2 最大池化与平均池化过程示意图

Fig. 2-2 Schematic diagram of large pooling and average pooling

池化层主要用于减少数据量和计算复杂度，同时保留重要的特征信息。它可以提高模型的训练速度和泛化能力，避免过拟合等问题。具体应用包括：

2.2.4 Transformer 网络

Transformer 是在 2017 年由 Vaswani 等人所提出的一种新颖的神经网络体系结构。在以往的研究中，循环神经网络（RNN）与神经网络（CNN）在连续型数据以及自然语言处理方面都获得了很好的效果，但它们在处理长距离依赖关系和并行计算方面存在一些不足。Transformer 的出现正是为了解决这些问题[71-81]。Transformer 的核心思想是自注意力机制。传统的 RNN 和 CNN 在处理序列数据时，需要按顺序逐步处理每个元素，难以捕捉长距离数据的依赖关系。而自注意力机制允许模型在处理每个元素时，能够关注到序列中的其他元素，从而捕捉到更全局的上下文信息。在医学图像处理的 Transformer 网络中，通常包含层归一化（Layer Normalization, LN）、前馈网络（Feed Forward Network, FFN）和自注意力层（Multi-head Self-Attention, MSA）是其最基本的三个部分。

（1）层归一化

Transformer 网络的训练是一个高复杂度的问题，为降低训练时间、降低运算代价、加速收敛，可以使用层归一化的方式。层归一化操作会对一层的输入

$(x_1, x_2, ,, x_d)$进行归一化操作得到$(x_1', x_2', ,, x_d')$。计算层归一化的操作公式如下：

$$\mu = \frac{1}{d}\sum_{i=1}^{d} x_i, \sigma = \sqrt{\frac{1}{d}\sum_{i=1}^{d}(x_i - \mu)^2 + \varepsilon}$$
$$x_i' = f\left(g\frac{x_i - \mu}{\sigma} + b\right) \quad (2\text{-}1)$$

其中，g、b为参数，f为非线性函数。

（2）前馈网络

FFN 是第一个被提出来的一种简单的人工神经网络。在前向网络中，每个神经元都是一个独立的层级，每个层级的神经元都能从前面的神经元接受信息，然后将其输出给新的神经元。第 0 个层次叫做输入层，第三个层次叫做输出，第三个层次叫做隐藏层。在没有任何反馈的情况下，由输入层到输出层的信号可以用一个有向无环图来描述，图 2-1 为多层前馈神经网络。

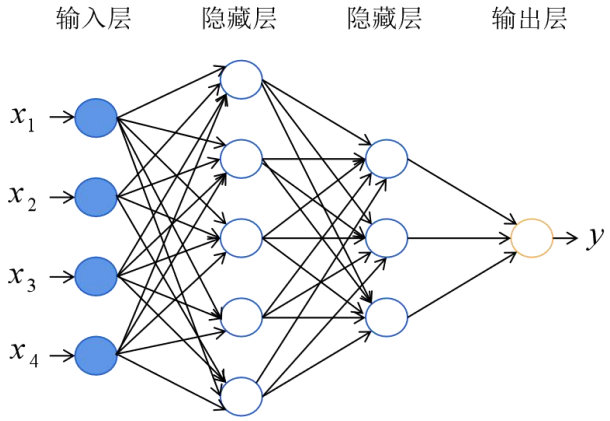


图 2-3 多层前馈神经网络

Fig. 2-3 Multilayer feedforward neural network

前馈神经网络中的符号通常描述如下：

表 2-2 前馈神经网络中的符号

Table 2-2 The notation of feedforward neural networks

记号	含义
L	神经网络的层数
M_l	第l层神经元的个数
$f_l(\bullet)$	第l层神经元的激活函数
$W^{(l)} \in R^{M_l \times M_{l-1}}$	第l-1 层到第l层的权重矩阵
$b^{(l)} \in R^{M_l}$	第l-1 层到第l层的偏置
$z^{(l)} \in R^{M_l}$	第l层神经元的净输入（净活性值）
$\alpha^{(l)} \in R^{M_l}$	第l层神经元的输出（活性值）

令$\alpha^0 = x$，前馈神经网络通过不断迭代下面的公式进行传播：

$$\begin{aligned} z^{(l)} &= W^{(l)}\alpha^{(l-1)} + b^{(l)}, \\ \alpha^{(l)} &= f_l\left(z^{(l)}\right) \end{aligned} \tag{2-2}$$

首先根据第$l-1$层神经元的活性值α^{l-1}计算出第l 层神经元的净活性值z^l，然后通过一个激活函数得到第l层神经元的活性值。式（2-2）若使用激活函数描述，则也可以写为：

$$\begin{aligned} z^{(l)} &= W^{(l)} f_{l-1}\left(z^{(l-1)}\right) + b^{(l)}, \\ \alpha^{(l)} &= f_l\left(W^{(l)}\alpha^{(l-1)} + b^{(l)}\right) \end{aligned} \tag{2-3}$$

（3）自注意力机制

自注意力机制（Self-Attention Mechanism），它也被称作“自我注意”或“注意机制”，是一个重要组成部分。自注意机制的原理是：当对连续的数据进行处理时，不仅要关注当前的输入，还是关注整个序列，并且为序列中的每一个元素赋予不同的重要性或权重。自注意力机制可以解决图像中的长程依赖，建立图像中的像素间的长程依赖性。例如，一幅图像中的某个部分可能会对图像的其他部分产生影响，Transformer 的自注意力机制可以捕捉到这些依赖关系。此外，自注意力机制可以实现端到端的全局优化，而传统的卷积神经网络通常会使用局部的卷积操作来提取图像特征，这些操作不容易捕获全局的图像信息。但 Transformer 的自注意力机制可以直接处理全局的信息，能实现端到端的全局优化。自注意力机制的运作流程如下：

（1）首先，每个输入元素都有三个向量表示：Query（查询），Key（键），Value（值）。这些向量是输入元素经过不同的线性转换（也就是乘以权重矩阵）得到的。

（2）然后，通过计算 Query 和所有 Key 的点积，得到每个元素的权重得分。这个得分反映了当前元素与其他元素的关联程度。更高的得分意味着这两个元素更相关。

（3）接着，将这些得分经过 Softmax 函数转换为概率值，这样所有的得分加起来就为 1。这个概率值就是注意力权重，反映了模型对每个元素的关注程度。

（4）最后，用这些注意力权重去加权求和所有的 Value，得到最终的输出。

这样，每个输出都包含了整个序列的信息，而且这个信息的整合是根据每个元素与当前元素的相关性进行的。这使得模型能够捕捉到长距离的依赖关系，并且可以并行处理整个序列，提高了计算效率。这就是自注意力机制的魅力所在，也是 Transformer 模型能够在很多任务上取得优异效果的主要原因。

2.2.5 损失函数

网络模型的训练目标是使模型预测的输出 $f(x)$ 与预设的标签 Y 尽可能一致，即减小两者间的 Loss 差距至极小。损失函数(Loss function)衡量这种预测偏差，表现为非负实数值的函数。它越小，象征着模型的适应性越优。在训练过程中，损失函数起到关键作用，它定义了一个标准，使模型能在任何时候优化其参数，有助于在最大化网络精度的情况下寻找到最佳参数，引导模型进入收敛状态。选择恰当的损失函数是网络训练过程中的核心决策。在 CT 图像的稀疏角度处理任务中，常见的损失函数包括均方误差和绝对值损失函数等。

（1）L2 范数损失，通常被称为最小平方误差，如公式（2-4）所表达，涉及第 i 个标签值与其对应的训练数据，致力于将预测输出 $f(x)$ 与实际值 y 之间的差值平方和降至最低：

$$L_2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{2-4}$$

在构建网络模型的训练过程中，L2 范数损失函数被广泛应用。

（2）绝对值损失函数，也称为 L1 范数损失，如公式(2-5)所示，其目标是最小化预测值 $f(x)$ 与标签 y 之间差值的绝对值，对数据中的离群值有较好的容忍性。

$$L_1 = \sum_{i=1}^{n} \left| y_i - f(x_i) \right| \tag{2-5}$$

L1 范数损失作为 L2 的替代选择，在深度网络模型的优化中展现出优势。大量的研究和实践证明，此方法能有效应对图像模糊问题，这主要归因于降低了的信噪比。然而，这种方法优化的深度网络模型也存在缺陷，可能导致预测结果出现一定程度的失真。

（3）对抗损失函数为 GAN 网络中的常规函数，如式（2-6）所示：

$$L_{adv} = E[\log D(x)] + E_{y,x}[\log(1 - D(G(x, y)))] \tag{2-6}$$

生成器 G 通过输入数据 x 和目标域标签 y 生成图像 $G(x, y)$，而判别器 D 来辨别真实数据与生成数据的真伪。

（4）感知损失是一种基于深度学习的图像风格迁移方法中常用的损失函数。与传统的均方误差损失函数相比，感知损失更注重图像的感知质量，更符合人眼对图像质量的感受。其损失函数如式（2-7）所示：

$$L_P = \frac{1}{K} \sum_{K-1}^{K} (F_K(x_i) - F_K(y_i))^2 \tag{2-7}$$

其中 y_i 和 x_i 分别代表着第 i 个标签值和其对应的预测数据，$F_K(\bullet)$ 表示它们在预训练的神经网络中的第 K 层的特征表示，K 是特征层数。感知损失可以用于各种图像处理任务中，如图像超分辨率、图像去噪、图像修复、图像风格迁移等。

在无监督学习中，虽然对抗丢失算法可以产生与物体分布类似的纹理，但是当样本数量较少时，也有可能引入错误的伪影。为此，提出了一种新的基于数据

的方法，即在图像内容信息重构中引入 L1 范数与 L2 范数。许多研究显示，对抗损耗网络具有较强的收敛性，因此，在文献中，为了解决这一问题，提出了一种新的基于梯度罚项的 Wasserstein 距离的方法[82]。另外，利用循环 GAN 与 LS-GA 相结合的循环一致性损耗[83]等方法，可以有效地解决人工免疫系统学习中出现的诸多问题。根据相关文献[84-91]，表 2-3 总结了深度神经网络模型中常用损失函数及其优缺点。

表 2-3　深度学习模型中常用损失函数

Table 2-3　Loss functions in deep learning

损失函数名称	优势	缺点
均方误差（MSE loss）	噪声抑制性能良好	过平滑、细节丢失
平均绝对误差（MAE loss）	噪声抑制性能良好	过平滑、细节丢失
全变差损失（TV loss）	噪声抑制性能良好	过平滑、细节丢失
对抗性损失（Adversarial loss）	图像视觉效果良好	容易出现错误结构
感知损失（Perceptual loss）	图像视觉效果良好	容易出现错误结构
结构相似性损失（SSIM loss）	更好的细节及边缘结构保护	对非结构性失真失效
边缘不相干损失（Edge incoherence）	更好的细节及边缘结构保护	噪声伪影抑制性能较差
一致性损失（Consistent loss）	使神经网络更具健壮性	容易出现错误结构

2.3 面向稀疏角度 CT 图像处理的相关模型

2.3.1 生成对抗网络

近年来，基于 Goodfellow 等人所构建的基于产生式对抗网络的学习思想得到了越来越多的关注。针对目前 CT 成像中存在的问题，提出了一种基于卷积神经网络的 CT 重建方法。GAN 网络可以很好地解决这个问题，它显示在图 2-4 中。在此基础上，提出了一种基于多个数据集的优化学习算法。在人工神经网络中，将具有稀疏角度 CT 图像作为数据的输入，产生经过预处理的影像。在此基础上，加入区分算子，判定后的结果与待识别的影像进行判定，并将其作为判定标记，从而持续地推动生成器的最优处理效果。在此基础上，引入感知损失、梯度约束等因素，构造出一种新型的混杂损失模型，以实现对复杂环境的有效控制。在此基础上，提出一种新的 GAN 模型，利用 GAN 模型与区分算子相结合的方法，实现对复杂图象的学习，并对其进行优化，以达到与目的相吻合的目的。

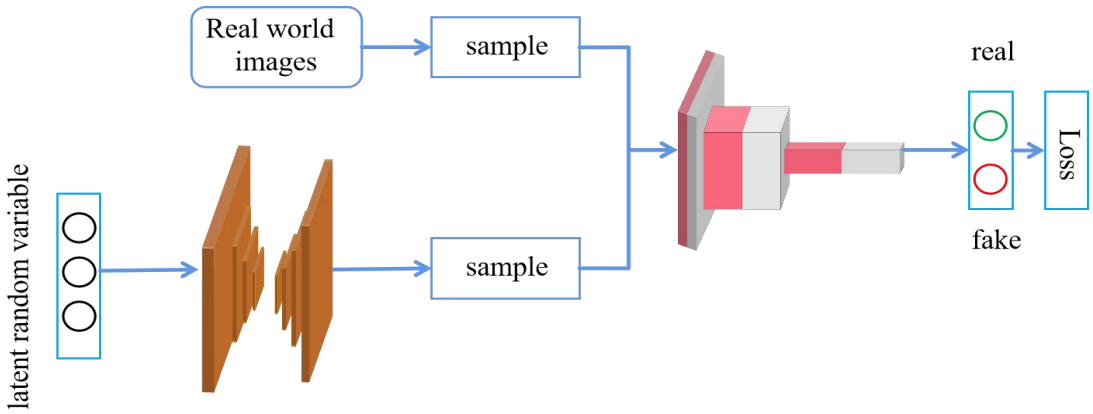


图 2-4　生成对抗网络流程图

Fig. 2-4 GAN network structure diagram

经过进一步的完善，比如 Wasserstein 最优距离和约束条件等，使得该方法具有较强的实用性。在这两个概率分布相互独立的情况下，利用 Wasserstein 距离构造出的 WGAN 模型具有较好的性能。另外，WGAN 通过对权重的限定来确保学习过程的稳定性，从而提高了整个网络的整体性能。在此基础上，通过对产生式对抗网络的研究，提出了一种基于仿真的 CT 影像生成方法。

2.3.2 StarGan 网络

由于生成对抗网络缺乏没有用户控制的能力和对低分辨率与低质量数据的处理不足等问题，2018 年， Choi 等人首次在多领域中建立了一种面向多领域的融合生成式对抗网络（StarGan）。StarGan 只需一种产生器和区分器即可完成跨域的影像自动生成与培训，并利用 mask vector 方式对有效区域内的影像进行标注，从而在人脸特征迁移与人脸表情合成等方面取得较好的效果。后来，很多研究人员扩大了它的适用范围，如应用于医学图像处理中，并取得了优异的性能。网络模型结构示意图如图 2-5 所示。

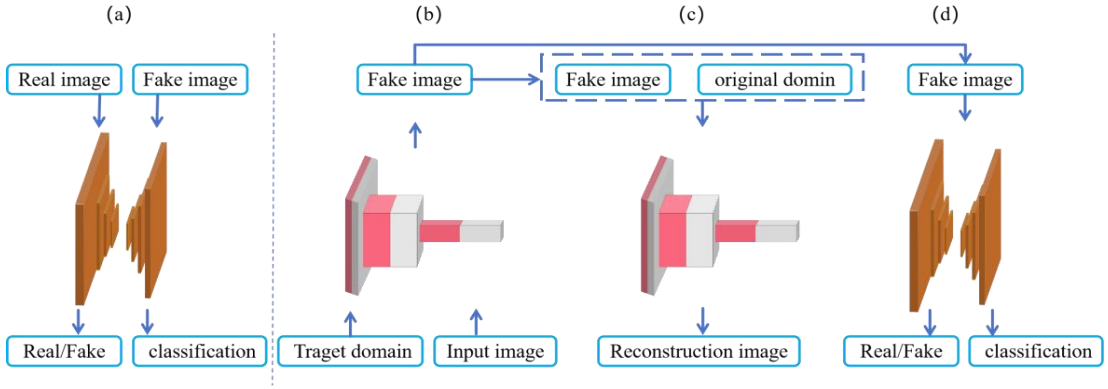


图 2-5 StarGan 概述

Fig. 2-5 Overview of StarGAN

从图 2-5 可以看出，StarGan 网络中的主要成分是判别器和生别器。其中：（a）：鉴别真伪图像，对真图像判定为真，对错误图像判定为错误，将实际图

像划分为对应的区域。（b）：接收一个真正的图象和一个对象字段标记，并且产生一个伪造的图象。（c）：通过给出初始领域标记，将伪图像重建成相似的原图像，从而得到重建损耗值。（d）：通过对相应的对象领域进行分类，尽量产生不能区别于真正的图像的图像。

2.3.3 CTTR 网络

传统的滤波反投影算法在稀疏角度 CT 图像重建中存在严重的伪影。采用迭代重建算法去除伪影，但由于重复投影和反投影，耗时长，容易产生块效应。为了克服稀疏角度 CT 图像重建的困难，2022 年，Shi 等提出了一种双域稀疏角度 CT 重建算法（CT Transformer, CTTR）。如图 2-6 所示，CTTR 包含四个主要组件：两个 CNN 残差块，用于从 FBP 图像和符号图中提取特征，一个编码器-解码器转换器，以及一个 CNN 块，用于将特征映射到恢复图像中。

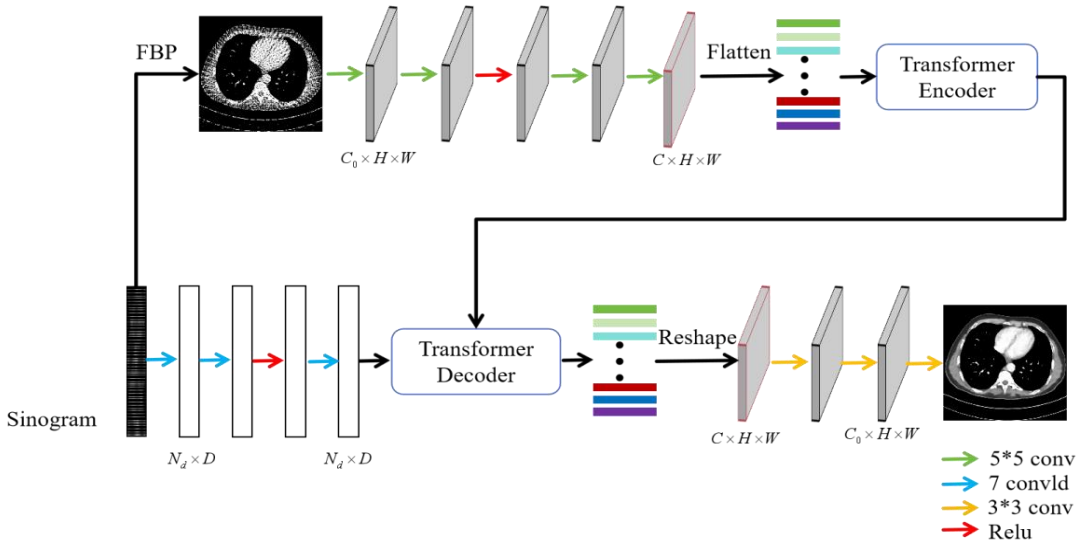


图 2-6 CTTR 网络结构

Fig. 2-6 CTTR Network Architecture

2.4 数据集

（1）为评估论文中所提的网络模型性能，实验将采用由 Mayo Clinic 授权的“2016 年 NIH-AAPM-Mayo LDCT 图像处理比赛”的数据集进行验证（简称 Mayo 数据集），该数据集包含来自 10 名患者的 4825 张 1mm 厚的高质量 CT 图像，可作为全角度的 CT 图像，也就是标签。通过对高质量 CT 图像进行 720 个投影，获得投影数据，并从这些投影数据中等角度间隔提取 90、120、180 三种情况的投影数据来重建，获得稀疏角度 CT 图像，用于网络训练。在投影实验中，CT 扫描几何为：扇形光束 X 射线；射线源到检测器和等中心的距离分别为 1085.6mm 和 595mm；该探测器有 736 个单元，每个单元的长度为 1.28mm。采用 FBP 算法进行重建，重建图像包含 512×512 个像素。实验中随机选择 8 个病例作为训

练集，1 个病例作为验证集，然后将剩余的 1 个病例用作测试集。此外对 CT 图像进行旋转，裁剪和调整大小等数据增广方法对数据集进行扩充，可增加更多的训练样本，以缓解模型过拟合问题。

（2）为了扩展数据范围并确保特性多样性，本研究引入了一个由 United Imaging Healthcare（UIH）公司提供的 CT 图像数据集，此数据集将被称为 UIH 数据集。该集合源于 UIH 的 uCT-760 系列扫描仪，包含了实际临床环境中的病例数据。具体来说，它包含了 14 位无名患者的常规 CT 扫描图像。所有扫描数据的采集与处理流程均事先获得了 UIH 公司内部伦理审查委员会的许可（批准代码：2015-07）。该数据集中标签为从 1200 个投影数据中重建出的图像，同时，从投影数据中等间隔提取 60、120 两种情况的投影数据，并重建出稀疏角度 CT 图像。数据采集的扫描几何为：锥束螺旋扫描；从源到检测器和等中心的距离分别为 950mm 和 570mm；该探测器有 936×80 个单元，每个单元大小为 1.548×1.405 mm2。重建的图像大小均为 512×512。采用与 Mayo 数据相同的预处理及增广策略。

2.5 实验验证及效果评估

2.5.1 实验验证

实验验证阶段，会充分考量网络架构和输入数据特性，实施精准的测试。（1）在模型优化环节，为了提升视觉呈现和细节纹理的精确性，不仅运用常规的均方误差损失函数，还将引入结构相似度损失和其他类型的损失函数，以此构建综合性的损失函数体系。（2）选取仿真数据作为优先训练集，利用监督学习在该数据集上评估所设计的三种深度学习网络的性能。同时，实验会对比其他 CT 图像处理技术，包括非深度学习的传统技术以及经典的深度学习算法，以凸显提出的网络优势。（3）对网络中的超参数进行精细调整，以确定最佳模型配置。（4）通过消融实验来检验每个网络部分的效果，对这些实验结果进行深入探讨，为后续研究提供指导。（5）在模拟数据上取得满意结果后，实验会引入部分真实数据，以增强模型的泛化能力和稳定性，同时提升预测 CT 图像的质量。最终，使用真实的患者数据进行验证和网络优化。

2.5.2 效果评估

在视觉效果评估方面，考虑到不同 CT 图像的表征差异，将按照预先设定好的观察点及尺度对 CT 图像进行定性分析，如图像的均匀度、组织纹理、细节对比度、伪影强度及分布等方面情况进行评价。

研究结果还采纳了严谨的量化评估标准进行深度分析。例如，均方误差（Mean Square Error, MSE）作为衡量处理后 CT 图像与参照图像间偏差的重要指

标；特征相似性指数（Feature Similarity Index Mersure, FSIM）则综合考量了亮度、对比度和纹理结构，通过对各要素的精确匹配来评判图像的优劣；峰值信噪比（Peak Signal to Noise Ratio, PSNR）则揭示了图像信号强度与噪声水平的比例关系。对于 MSE，其计算方法遵循公式（2-8）；FSIM 的计算依据是公式（2-9）；PSNR 的计算规则在公式（2-10）中明确；而 SSIM 则聚焦于对比图像在结构细节层面的契合度，其计算公式见公式（2-11）。：

$$M_{SE}(x,y)=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[x_{i,j}-y_{i,j}\right]^2 \tag{2-8}$$

$$F_{SIM}=\frac{\sum_{x,y}S_L(x,y)\bullet PC(x,y)}{\sum_{x,y}PC(x,y)} \tag{2-9}$$

$$P_{SNR}(x,y)=10\bullet\log_{10}\left(\frac{x_{MAX}^2}{M_{SE}(x,y)}\right) \tag{2-10}$$

$$S_{SIM}(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu x^2+\mu y^2+c_1)(\sigma x^2+\sigma y^2+c_2)} \tag{2-11}$$

其中，x表示为经过待评价的图像，y为参考图，$PC(\bullet)$表示为图像的相位一致性特征提取，$S_L(\bullet)$表示为图像相位一致性特征提取和梯度特征提取融合的相似度。$m\times n$为目标图像块的大小，x_{MAX}表示目标图像x中的最大像素值，MSE（•）为均方误差计算函数，μ_x和μ_y分别表示两个图像感兴趣区域，用来估计亮度；σ_x^2和σ_y^2分别表示两个图像的方差，用来估计对比度；σ_{xy}是两个图像的协方差，用来估计结构相似度；c_1和c_2为两个常数用来避免除数为 0 的情况。

2.6 本章小结

本章全面解析了 CT 图像的形成原理以及广泛应用的处理技术，同时详尽叙述了实验所依托的数据集和评价准则。稀疏角度扫描技术虽能有效减小 CT 投影数据和辐射影响，却也易引发重建图像的严重失真现象。洞悉这些失真的产生过程对构建高效的图像处理策略起着关键作用，为本研究奠定了坚实的理论根基。国内外学者在稀疏角度 CT 成像及相应处理技术方面已积累了丰富的研究成果，其实践意义深远，深入探究这些方法将为后续研究提供借鉴和比较分析的素材。此外，本章还详细介绍了用于训练神经网络模型的数据集具体情况。为了衡量不同方法在稀疏角度 CT 成像中的效能，本章确立了一系列核心评估指标，为后续算法开发提供了基本准则和方向指导。

第 3 章 基于 Transformer 的混合域网络稀疏角度 CT 成像

单一图像域网络在去除稀疏角度 CT 图像伪影上难以取得满意的效果，为此本章借助于新型的 Transformer 网络，构建适用于多阶段稀疏角度 CT 投影数据及图像数据的处理流，提出了一种基于 Transformer 的混合域网络稀疏角度 CT 成像算法（Hybrid Domain network for sparse view CT imaging based on Transformer，HDTransformer），以提高稀疏角度 CT 图像重建质量。本方法采用图像域—投影域—图像域三阶段混合处理流程，通过多阶段信息的联合互补提高成像质量。此外，针对不同阶段数据噪声伪影特点设计不同的 Transformer 模块，以实现差异化的处理，在第一阶段采用耦合 Transformer 模块对稀疏角度 CT 图像进行复原，在第二阶段采用浅层 Transformer 网络，可减缓投影数据过处理现象的发生，经过前两个阶段的修复，图像已去除大部分伪影，因此在第三阶段采用轻量化的 Transformer 网络设计方案，最终实现 CT 图像的高质量成像效果。更进一步，算法采用可微分的解析重建和投影运算，建立投影域与图像域数据的转换，最终实现端到端的稀疏角度 CT 优质成像流。通过 Mayo 数据集实验验证了所提网络的有效性，其视觉结果表明：处理后的不同部位 CT 图像噪声伪影均能够得到较好的抑制；量化结果表明：处理后的 CT 图像的峰值信噪比和特征相似性均优于对比方法，具有更高的重建质量。

3. 1 稀疏角度 CT 成像数学模型

人体组织（如骨骼和软组织）具有不同的 X 射线衰减系数 μ。当考虑二维 CT 图像时，衰减系数的分布 $X = \mu(a,b)$，其中(a,b)表示位置坐标，X表示断层解剖组织的衰减系数分布。CT 成像的原理是基于傅立叶切片定理，该定理保证了二维图像X可以从得到的密集投影中重构出来。在成像时，根据 Lambert-Beer 定律，通过发射和接收的 X 射线强度来推断解剖结构 X 的投影。进一步，在能量分布$\eta(E)$的多色 X 射线源下，CT 成像过程的数学表达式为：

$$P = -\log\int \eta(E)\exp\{-AX\} \tag{3-1}$$

式中 A 表示正弦图生成过程，即 Radon 变换（通常用扇形波束成像几何来定义）。通过上述正演过程，CT 成像的目的是利用估计/学习函数，从得到的投影 $P = AX$ 重建出图像 X。在实际的稀疏角度 CT 图像中，投影数据是不完整的，这种减少的正弦图信息严重限制了常规解析重建方法的性能，并且无法很好的复原 CT 图像，影响医师的诊断。为了更好的修复稀疏角度 CT 图像，本章提出了一种基于 Transformer 模块的混合域网络。

假设 P_f 为全角度扫描下的投影数据，P_s 为稀疏角度扫描下的投影数据，通过图像重建算法可以将 P_f 和 P_s 分别重建为全角度 CT 图像 I_f 和稀疏角度 CT 图像

I_s，如公式（3-2）所示：

$$\begin{cases} I_f = R\left(P_f\right) \\ I_s = R\left(P_s\right) \end{cases} \tag{3-2}$$

其中，$R(\bullet)$为表示图像重建算法。由于扫描角度的减少和缺失，稀疏角度 CT 图像 I_s 中会存在大量的条形伪影及噪声。为此，基于深度神经网络的稀疏角度 CT 成像方法，就是通过网络训练，以去除 I_s 中的大量伪影。由于单一图像域的修复难以完成所有的伪影去除，因此本章在投影域网络中对投影数据进行修复，即通过训练，使 P_s 接近全角度投影数据 P_f。

3.2 混合域网络稀疏角度 CT 成像

为了去除图像中的噪声伪影，本章采用了一种新型的成像网络流程，它由三个阶段组成。第一阶段是图像域处理子网络，将稀疏角度 CT 图像 I_s 输入图像域 Transformer 子网络进行复原，以去除图像中的高强度噪声和伪影。第一阶段的图像域处理可以表示为：

$$I_{s_1} = T_{I_1}\left(I_s\right) \tag{3-3}$$

其中，$T_{I_1}(\bullet)$是第一阶段图像域 Transformer 子网络，I_{s_1} 是经过 $T_{I_1}(\bullet)$网络修复过的 CT 图像。第二阶段的处理为：首先将经过第一阶段修复的 CT 图像 I_{s_1} 投影成全角度的投影数据 P_{s_1}，投影参数与全角度扫描参数相同；然后将投影数据进行分块并输入至投影域 Transformer 子网络进行投影数据修复；最后将修复好的投影数据借助重建算法生成二阶段的 CT 图像 I_{s_2}。第二阶段处理过程可以表示为：

$$I_{s_2} = R\left(T_s\left(P_{s_1}\right)\right) \tag{3-4}$$

其中，$T_s(\bullet)$是投影域 Transformer 子网络。第三阶段的处理是将 I_{s_2} 输入到第三阶段图像域 Transformer 子网络中实行进一步的图像细节微调和增强，以最终获得高质量的 CT 图像。第三阶段的处理过程可以表示为：

$$I_{s_3} = T_{I_2}\left(I_{s_2}\right) \tag{3-5}$$

其中，$T_{I_2}(\bullet)$是第二阶段图像域 Transformer 子网络，I_{s_3} 是经过 $T_{I_2}(\bullet)$修复过的高质量 CT 图像。通过三个阶段的处理，形成了用于稀疏角度 CT 成像的 HDTransformer 网络，整个成像算法流程图如图 3-1 所示。

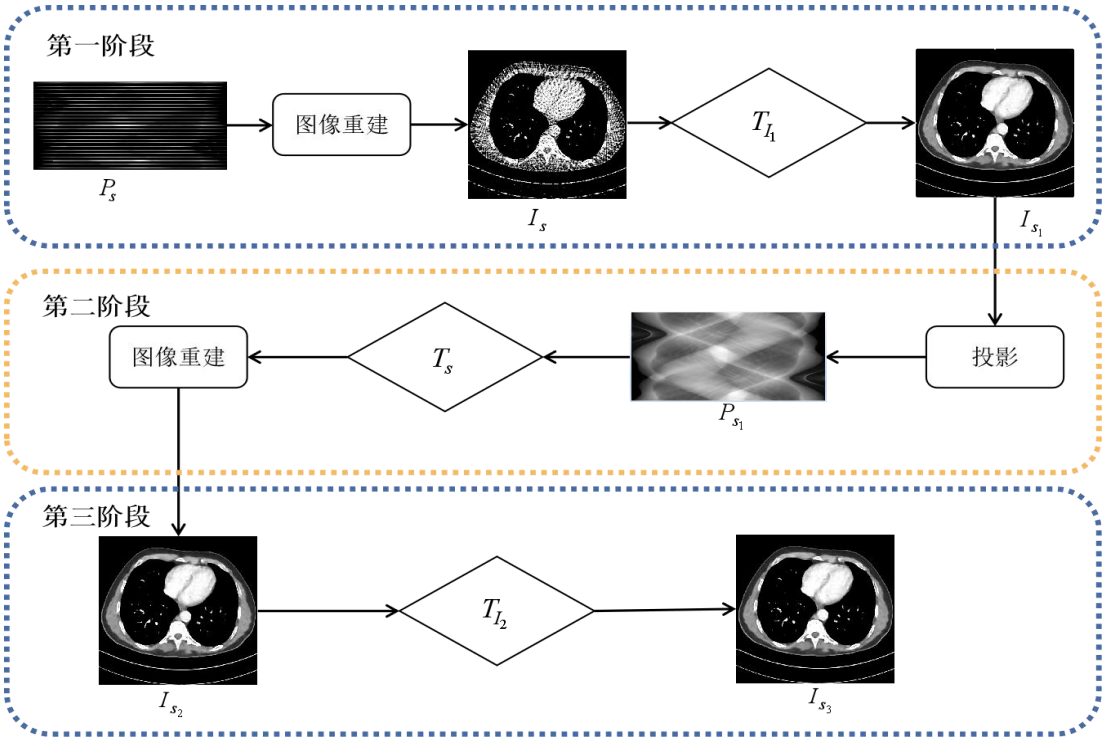


图 3-1 HDTransformer 流程图

Fig. 3-1 The workflow of HDTransformer

3.3 网络设计

HDTransformer 三个阶段的子网络中，首先使用 3×3 卷积来获得浅层特征。然后，通过不同的 Transformer 模块来提取不同阶段数据的主要特征并进行处理。与其他网络模块相比，Transformer 网络模块中不存在循环结构，可大幅的提高并行计算效率；同时 Transformer 中的自注意力机制允许模型从序列中任意位置获取信息，这使得 Transformer 能够更好地处理长距离依赖关系，而不需要通过增加卷积层数来处理这种长距离依赖，更适用于不同类型不同退化程度的数据处理。因此，这里 HDTransformer 在三个阶段的数据处理过程中，均选用 UNet 形式的架构，以 Transformer 模块为基本的编码解码单元，替代传统的卷积层，构建处理网络。如图 3-2 所示，为了简单起见，图 3-2 的 UNet 只显示了 2 层 Transformer 块。在 HDTransformer 中不同阶段的处理均设计了不同的 Transformer 模块来提取 CT 图像和投影图像的特征。其中第一阶段和第三阶段的 UNet 总共有 6 层 Transformer 模块，考虑到处理投影数据需要谨慎，容易出现过拟合的情况，第二阶段的 UNet 共有 4 层 Transformer 模块。最后，再采用 3×3 卷积获取最终处理后的数据。

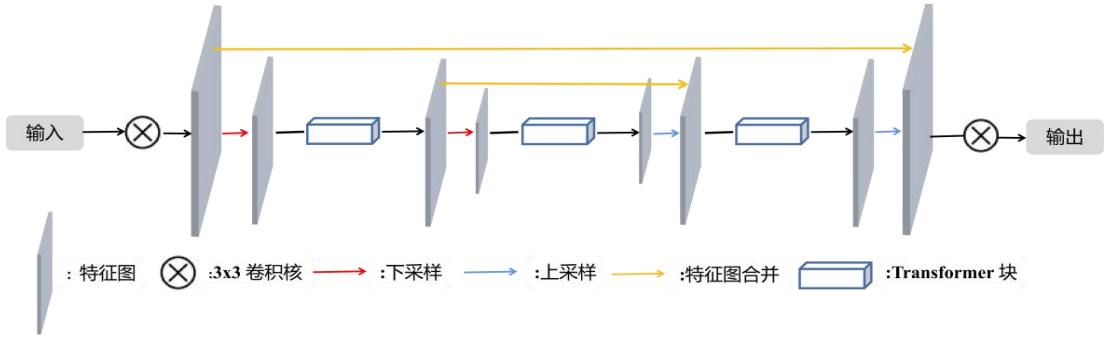


图 3-2 基于 Transformer 模块的 UNet 网络

Fig.3-2 UNet network based Transformer block

第一阶段图像域子网络所采用的是耦合 Transformer 块结构，主要由 3 部分组成，第 1 部分是 3×3 的卷积层；第 2 部分是归一化层（Layer Normalisation），窗口式多头自注意力（Window Multi-head Self-Attention，W-MSA）模块和 GELU（Gaussian Error Linear Unit，GELU）激活函数操作；第 3 部分是归一化层，具有 SWIN 滑动窗口的多头自注意力（Shifted Windows Multi-head Self-Attention，SW-MSA）和 GELU 激活函数操作。第 2 部分之前和第 3 部分后均包括 Patch Enbedding 层。第一阶段图像域子网络的 Transformer 模块如图 3-3（a）所示，该耦合形式的模块融合了传统卷积网络与 Transformer 优势，能在参数较少的情况下提取有效信息实现去伪影。在训练第一阶段图像域网络时，使用 MSE 作为网络的损失函数，其中 θ_1 为网络参数，损失函数表示为：

$$L_{T_1}\left(I_{s_1},I_f,\theta_1\right)=\left\|I_{s_1}-I_f\right\|_2^2 \tag{3-6}$$

在第二阶段投影域子网络中，Transformer 块采用线性叠加连接的设计思想，主要包含线性连接层（LN）、多头自注意力模块（MSA）和前馈网络（FFN）等部分，且该 Transformer 模块为浅层网络。浅层网络可减缓投影数据过处理现象的发生，同时可应对深层网络训练难不易收敛等问题，同时还可提高投影数据的复原效果。拟采用的 Transformer 结构如图 3-3（b）所示。在训练第二阶段投影域网络时，网络的损失函数如下，其中 θ_2 为网络参数：

$$L_{T_2}\left(P_{s_1},T_s\left(P_{s_1}\right),\theta_2\right)=\left\|P_{s_1}-T_s\left(P_{s_1}\right)\right\|_2^2 \tag{3-7}$$

第三阶段图像域子网络结构同样采用轻量化的设计方案，该 Transformer 块由三个部分组成：Layer Normalisation 层、W-MSA 模块和局部增强的前馈网络（Locally enhanced feedforward network，LeFF）。经过第二阶段输入到第三阶段的 CT 图像相较于输入第一阶段的 CT 图像伪影及噪声已大大减少，存在的主要问题是对比度不高，细节丢失。此阶段的轻量化 Transformer 模块不仅可以进一步的去噪去伪影，获得高质量的 CT 图像，同时尽可能的提高组织对比度，复原丢失的图像细节，减少训练计算量和内存。结构如图 3-3（c）所示，训练第三时

采用的损失函数如下，其中 θ_3 为网络参数：

$$L_{T_2}\left(I_{s_3}, I_f, \theta_3\right) = \left\| I_{s_3} - I_f \right\|_2^2 \tag{3-8}$$

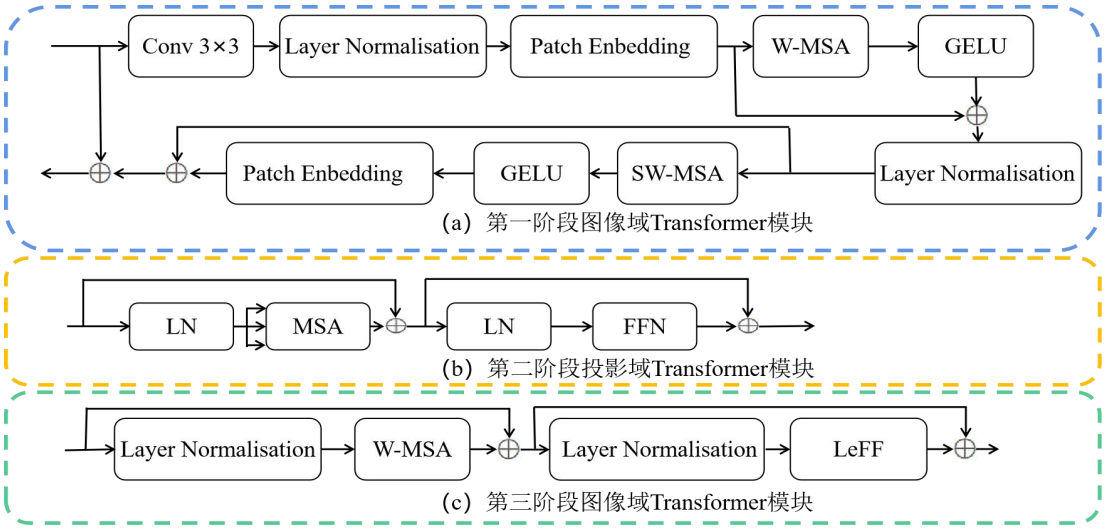


图 3-3 不同阶段子网络中 Transformer 模块结构

Fig. 3-3 Transformer block in different stages sub-net

3.4 实验结果与分析

3.4.1 实验设计

为了验证本章所设计 HDTranformer 网络在稀疏扫描条件下成像的有效性，实验选用 FBP、bilinear+FBP 和 CTTR 三种方法对比，其中 FBP 和 bilinear+FBP 是非深度学习的传统方法，CTTR 方法[18]为一种基于 Transformer 网络的双域稀疏角度 CT 重建算法，具有优异的稀疏角度图像处理性能，与本章方法类似。实验硬件配置为：Intel(R) Core(TM) i5-12600k 的 CPU，NVIDIA GTX 4090 显卡。为加速深度学习模型的收敛，所有的训练及测试均是在显卡上完成。

实验中采用的数据集是由 Mayo Clinic 授权的“2016 年 NIH-AAPM-Mayo LDCT 图像处理比赛”的数据集。该数据集中包括 10 名匿名患者匹配的 CT 图像与投影，选取了 9 名患者的 CT 数据作为训练集，1 名患者的 CT 数据作为测试集。在第一阶段将 10 名患者的 CT 稀疏角度采样投影数据进行重建，将重建后的稀疏角度 CT 图像作为样本，全角度下 CT 图像作为标签输入到网络中进行训练。在第二阶段中，将第一阶段修复的 CT 图像进行投影，将获得投影数据作为样本，将全角度 CT 图像的投影数据作为标签。在第三阶段，将经过第二阶段修复的投影数据进行重建，将重建后的 CT 图像作为样本，将对应的全角度 CT 图像作为标签。重建均采样 FBP 算法。重建 CT 图像大小为 512×512，全角度投影数据大小为 720×768，720 个投影角度。网络训练中，为减少单次数据量的

使用，BatchSize 大小为 3，网络通道数为 96，训练 epoch 设置为 50 次，学习率的初始值为2×10^{-4}，学习率递减到最小值为1×10^{-6}。

3.4.2 实验结果

图 3-4，图 3-5 和图 3-6 分别为 90 个角度，120 个角度和 180 个角度下的重建结果，其中第一行、第二行和第三行表示使用不同算法对所选胸部、腹部和胯部断层切片的成像结果，显示窗为[-150HU，250HU]。图中第一列为全角度下 FBP 重建图像，作为不同方法成像结果的对比。对比图中第二列成像结果，可以看到，由于投影数据的缺失，传统解析重建算法获得图像含有大量的伪影，图像细节被伪影淹没，严重影响诊断。同时也发现，随着扫描角度的减少，重建图中伪影越来越多。经过双线性插值后，缺失的投影数据被初略的补全，重建图像有了明显的改善，部分细节清晰，但依旧含有大量的条状伪影和部分噪声。通过对比第四列 CTTR 方法成像结果和第五列 HDTransformer 方法成像结果，可以发现图像质量有了大幅度的提升。进一步通过局部放大图可以看出，CTTR 算法结果会丢失一部分图像信息，细节对比度有所下降，出现一定的过平滑现象，而本章 HDTransformer 方法获得图像细节更多，对比度更高（如图 3-4-图 3-6（e）中红色箭头所示），更接近参考图，整体图像质量要优于其他三种对比方法。

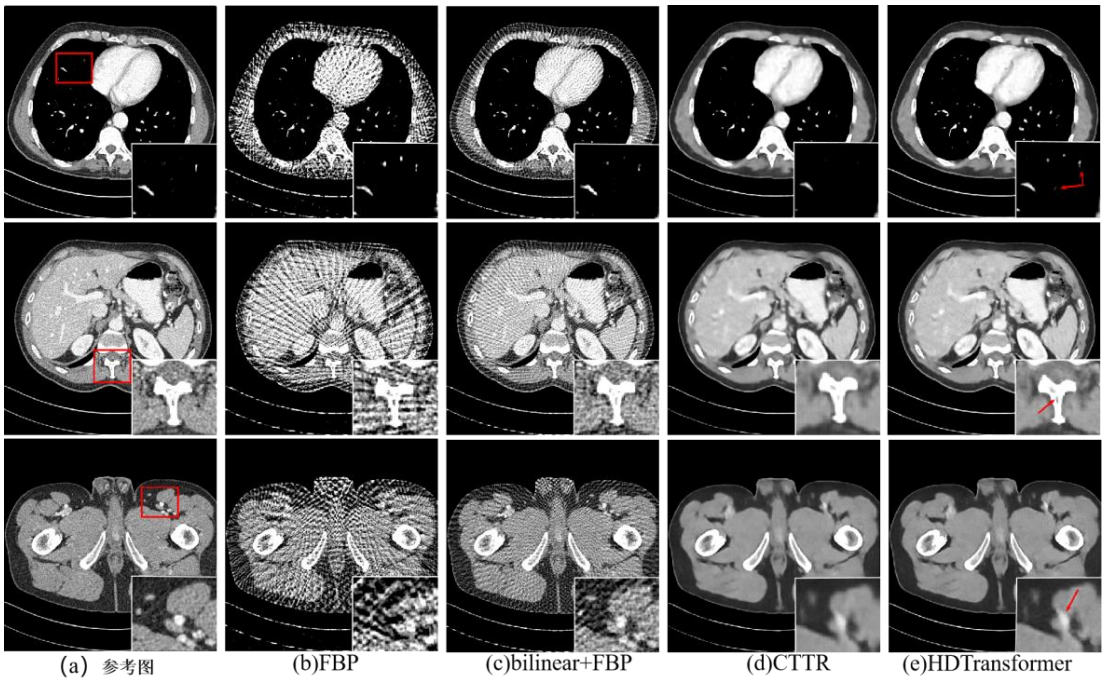


图 3-4 90 个扫描角度下不同方法的 CT 成像结果

Fig. 3-4 Axial view results of different methods under 90 views

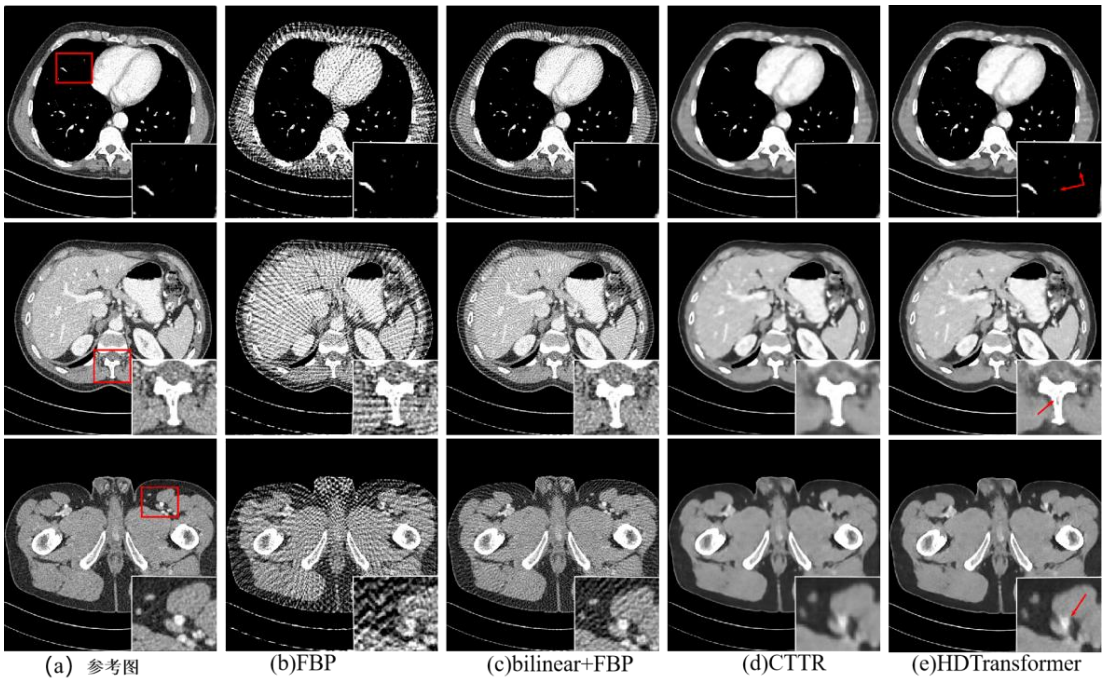


图 3-5 120 个扫描角度下不同方法的 CT 成像结果

Fig. 3-5 Axial view results of different methods under 120 views

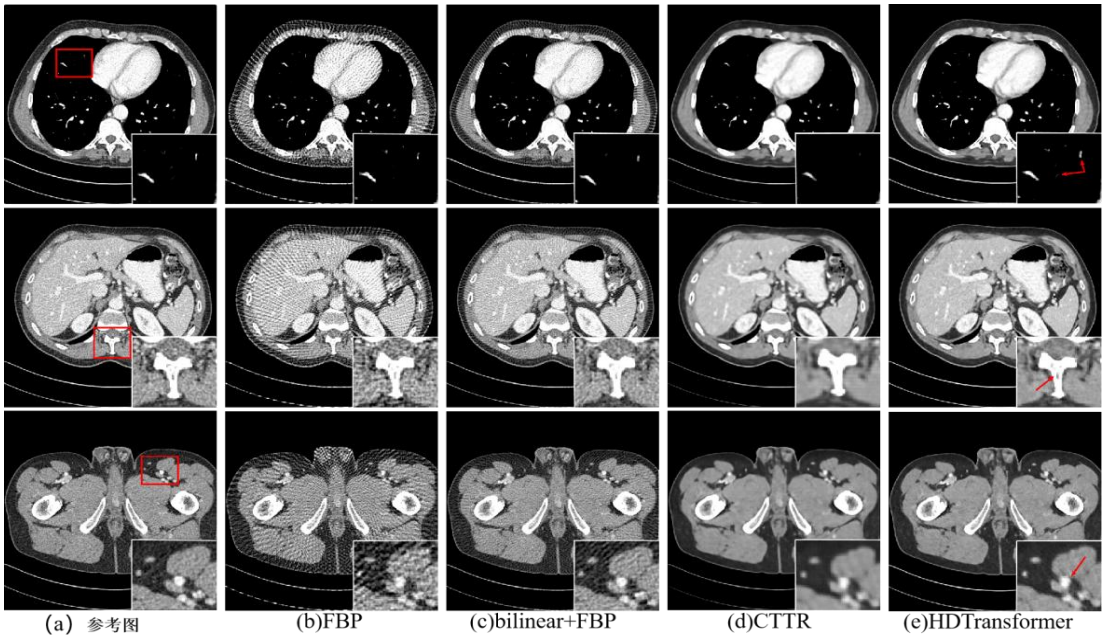


图 3-6 180 个扫描角度下不同方法的 CT 成像结果

Fig. 3-6 Axial view results of different methods under 180 views

为进一步分析不同方法在冠状面和矢状面中的成像效果，实验选取 90 个投影角度下冠状面和矢状面重建图进行比较，成像结果如图 3-7 所示，显示窗为[-150HU，250HU]。从图中可以看出，冠状面和矢状面中噪声伪影相对较少，扫描角度减少对这两个切面影响没有横断面大。对比方法 bilinear+FBP 虽然在一定程度上去除了图像伪影，但是整体复原度不高，部分细节丢失，同时在胯部位置

中存在伪影残留。CTTR 方法在噪声伪影上去除得非常干净，但图像组织纹理保持度不高，部分细节丢失，而本章 HDTransformer 方法显著降低了各部位噪声和伪影，同时也更好地保留了图像组织区域的边缘特征，处理后的图像具有较好的视觉效果。

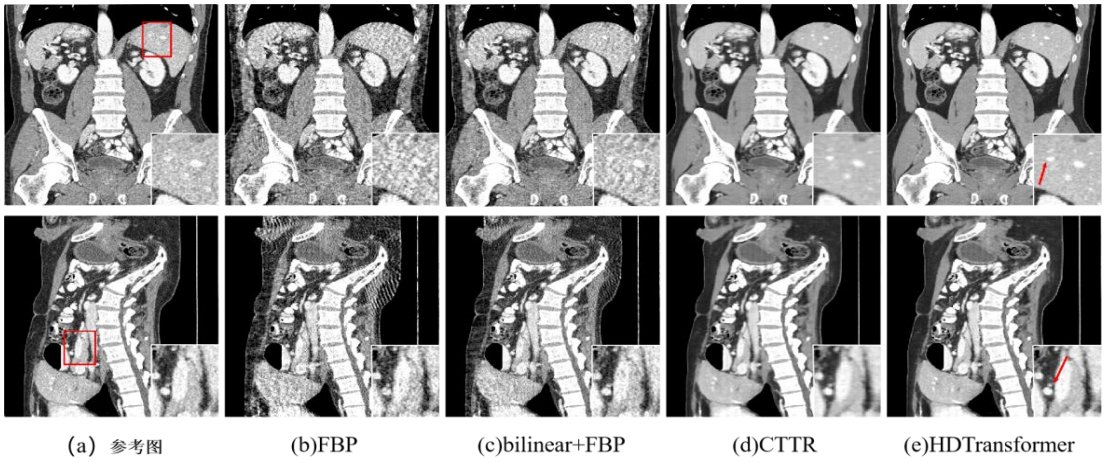


图 3-7 90 个扫描角度下矢状面和冠状面的成像结果

Fig. 3-7 Sagittal and Coronal results of different methods under 90 views

图 3-8 分别展示了 90 个角度，120 个角度和 180 个角度扫描下成像结果与对应参考图之间的差异，差异图由不同方法重建的图像与全角度扫描下参考 CT 图像相减得到，差异图反映处理后的图像与原始 CT 图像之间的接近程度，图 8 中显示窗口为[0HU，200HU]。通过差异图，可以发现前三种方法处理后的 CT 图像残留伪影较多，与参考图像之间差异较大，而本章 HDTransformer 方法的差值图强度较低，整体较均匀，这表明本章方法获得的图像更加接近于参考图，显著优于其他处理方法。综上，从视觉效果来看，本章方法的处理效果要优于对比方法。

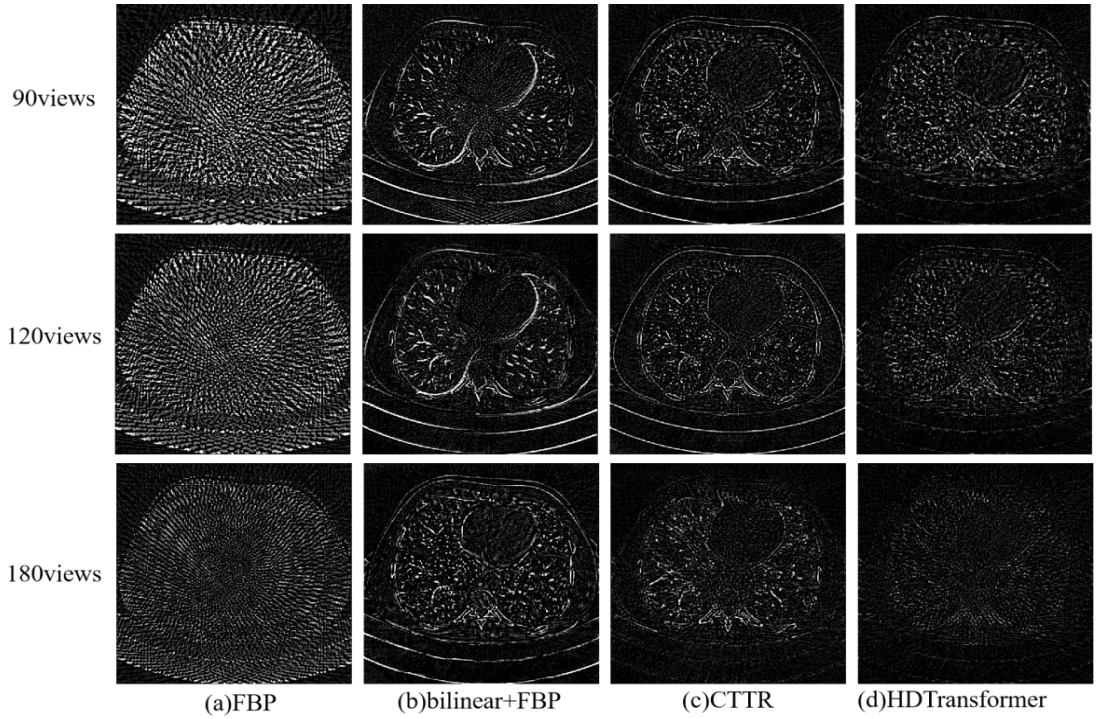


图 3-8 不同方法获得的 CT 图像的差异图

Fig. 3-8 Difference images relative to the reference image obtained by different methods

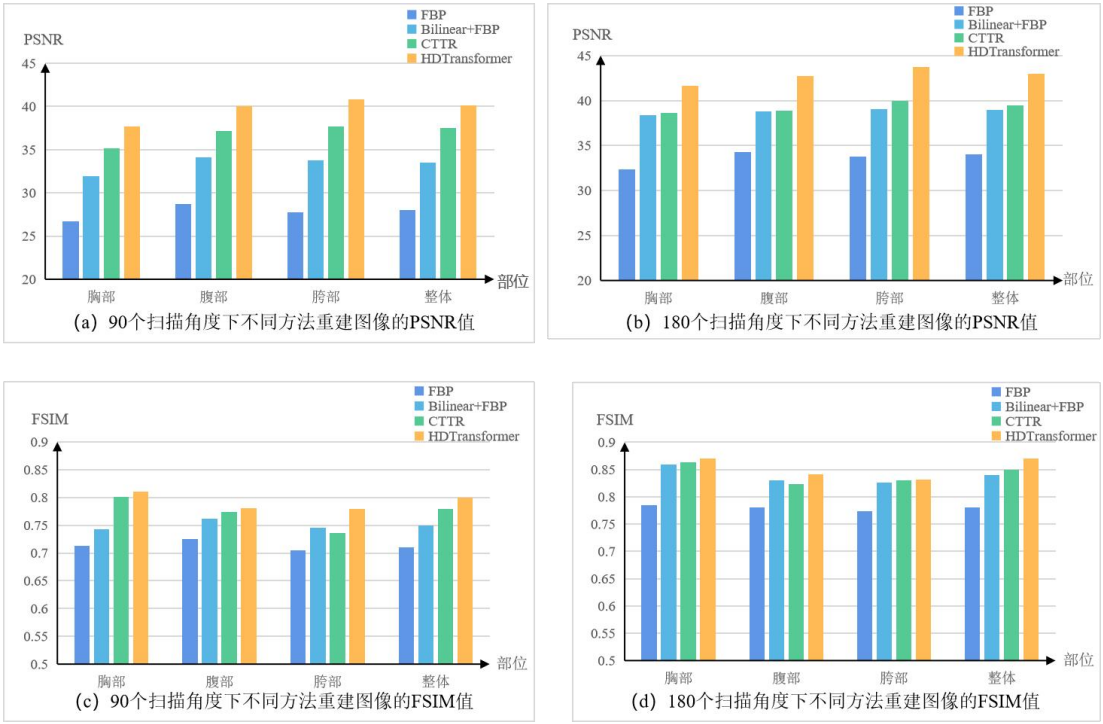


图 3-9 不同扫描角度下不同方法重建图像的 PSNR 值和 FSIM 值

Fig. 3-9 PSNR and FSIM values of reconstructed images using different views projection and methods

为更加直观及定量地比较不同算法的处理效果，本章采用 PSNR 和 FSIM 两个指标来评估实验结果。图 3-9 为 90 个角度和 180 个角度扫描下不同方法重建

图像的 PSNR 和 FSIM 直方图。通过对比直方图可以看出，本章 HDTransformer 方法在 PSNR 和 FSIM 两个指标下，均优于其他三种方法，具有明显的量化优势。为了评估像素级的 HU 拟合结果，在图 3-10 中绘制了紫线表示的轮廓。对于 90 角度和 120 角度数据，HDTransformer 的拟合性能优于 CTTR，表明 HDTransformer 具有更好的结构保存性能。

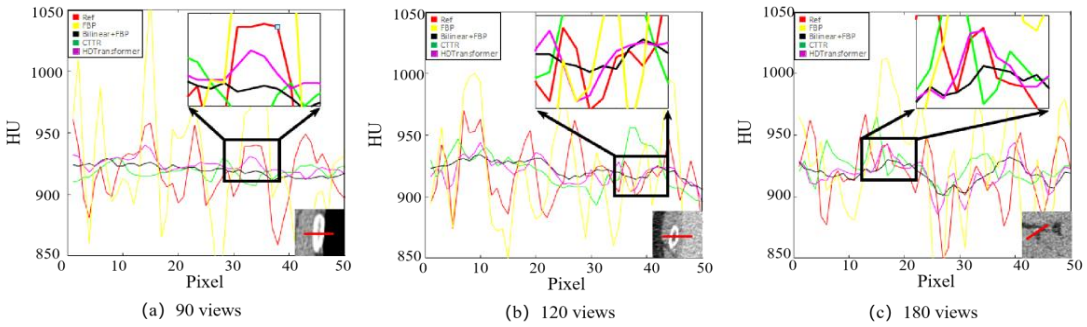


图 3-10 不同方法的三个轴向切片沿指定红线的强度分布图

Fig. 3-10 The intensity profiles along the specified red line in the three axial slices of different methods

3.4.3 网络模型分析

（a）消融实验

本章 HDTransformer 成像方法由三个阶段组成，分别是第一阶段图像域、第二阶段投影域和第三阶段图像域的处理，不同阶段对稀疏角度 CT 成像结果均有一定程度的促进作用，为了验证本章的混合域网络对图像提升作用，此部分分别对不同阶段成像效果进行了比较，包括第一阶段图像域（记为T_1）、第二阶段的投影域（记为S_1）、第一阶段图像域与第二阶段投影域（记为T_1S_2）、第二阶段投影域与第三阶段图像域（记为S_1T_2）以及本章的同时采用三个阶段处理（记为$T_1S_2T_3$）。实验结果选取胸部图像和腹部图像进行比较，如图 3-11，图 3-12 所示，其中图 3-11 为 90 个稀疏角度扫描后的 CT 图像，图 3-12 为 180 个稀疏角度扫描后的 CT 图像，显示窗口均为[-150HU，250HU]。通过对比图 3-11、图 3-12 中胸部的局部放大图可以发现，本章的三阶段处理策略$T_1S_2T_3$具有最好的成像效果，其余四种组合都存在不同程度的细节丢失或伪影残留情况。通过对比图 3-11 和图 3-12 中腹部局部放大图可以看出，在处理伪影和噪声方面，本章所采用的三阶段混合域网络同样处理效果最好，更加接近参考 CT 图像。

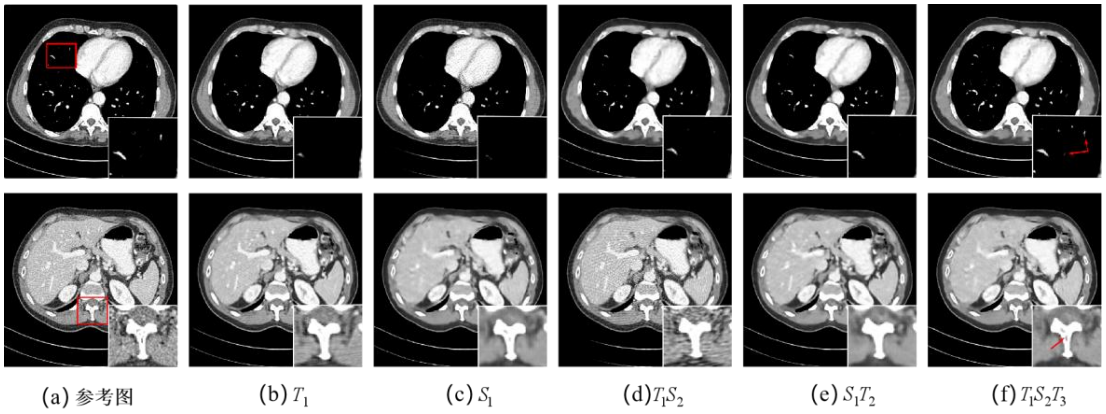


图 3-11 90 个扫描角度下胸部和腹部 CT 成像效果

Fig. 3-11 Chest and abdomen imaging results of different methods under 90 views

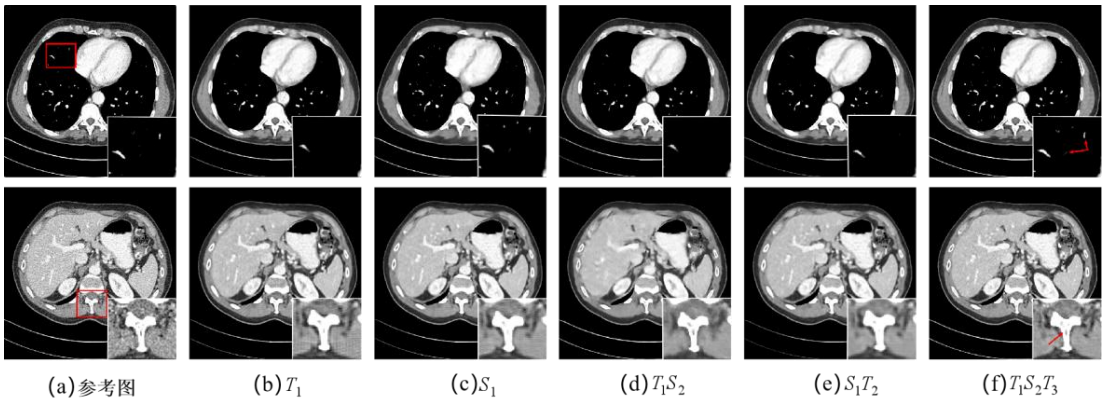
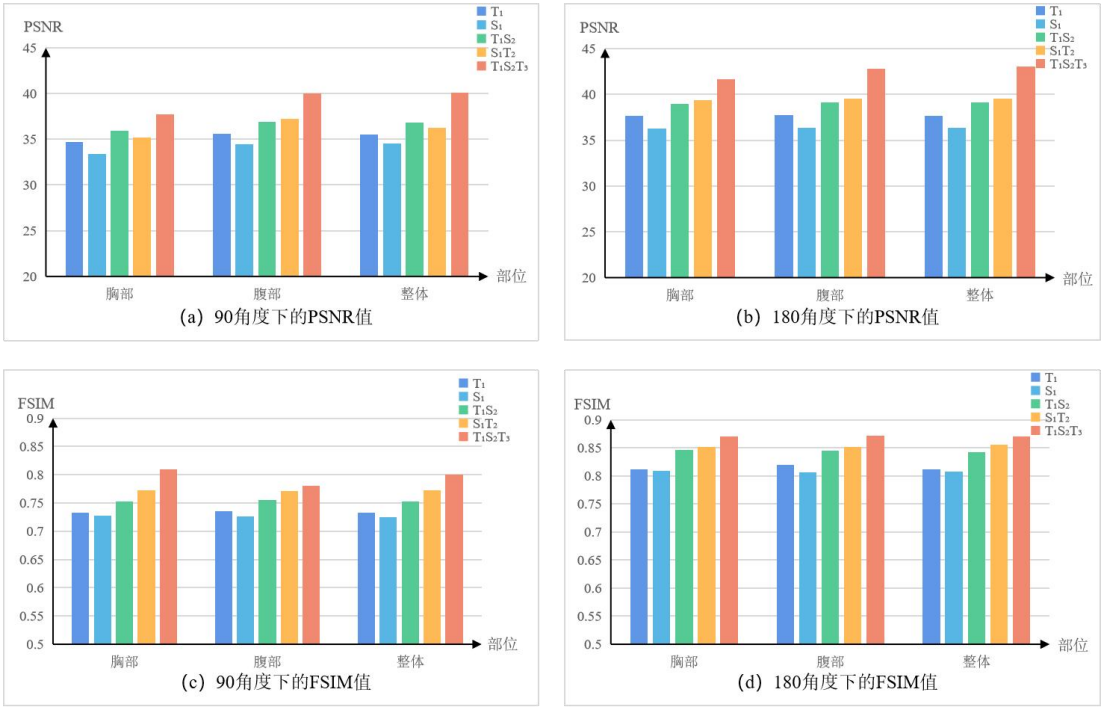


图 3-12 180 个扫描角度下胸部和腹部 CT 成像效果

Fig. 3-12 Chest and abdomen imaging results of different methods under 180 views

为更加直观及定量地比较不同阶段方法的对成像效果的促进情况，此部分将进一步采用 PSNR 和 FSIM 这两种客观评价指标进行量化。不同阶段重建后的 CT 图像的 PSNR 值和 FSIM 平均值如图 3-13 所示。通过 PSNR 和 FSIM 的直方图可以看出，单一的投影域处理网络其重建图像 PSNR 和 FSIM 值最低，两阶段的处理方法（T_1S_2 和 S_1T_2）要优于一阶段的单域处理方法（T_1 和 S_1），而本章提出的三阶段的混合域方法重建 CT 图像 PSNR 和 FSIM 最高，表明其重建图像更加接近于全角度下的参考图像。



(a) 90角度下的PSNR值　(b) 180角度下的PSNR值

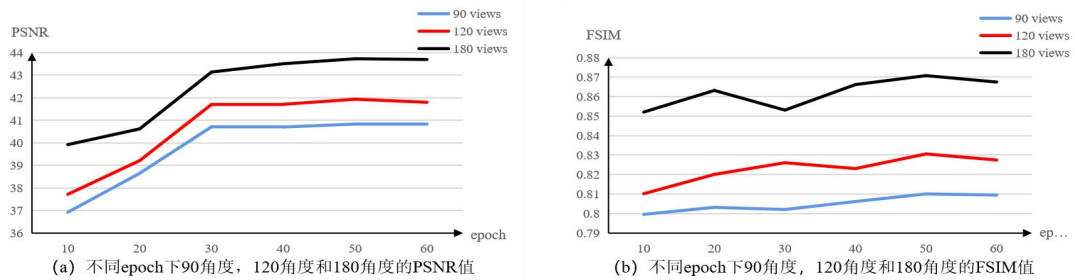
(c) 90角度下的FSIM值　(d) 180角度下的FSIM值

图 3-13 90 和 180 个角度下不同成像模型的 PSNR 值和 FSIM 值

Fig. 3-13 PSNR and FSIM values with different model under 90 and 180 views

（b）收敛性验证实验

此部分实验进一步分析了所提网络的收敛性。实验采用不同训练 epoch 下重建 CT 图像的 PSNR 均值和 FSIM 均值变化曲线，间接分析网络的收敛性和稳定性。实验选取 300 张 CT 图像为测试数据，在不同扫描角度下，网络训练后重建图像的 PSNR 和 FSIM 曲线如图 3-14 所示。从图 3-14（a）中可以看出，随着训练周期的增加，不同数据的 PSNR 值均逐步增高并在 30 个 epoch 后趋于平缓，并逐步收敛。从图 3-14（b）中可以看到，不同扫描数据训练后 FSIM 值变化差异较大。180 个扫描角度数据在 30 个 epoch 的训练下 FSIM 值反而有所下降，在 50 个 epoch 训练后均能达到最大值，随后有了略微的下降，表明网络开始出现过拟合现象。为此，综合视觉效果、PSNR 及 FSIM 的量化值，实验中选取训练周期为 50 个 epoch。



(a) 不同epoch下90角度，120角度和180角度的PSNR值　(b) 不同epoch下90角度，120角度和180角度的FSIM值

图 3-14 不同 epoch 下 90 角度，120 角度和 180 角度的 PSNR 值和 FSIM 值

Fig. 3-14 PSNR and FSIM values under 90, 120, and 180 views under different epochs

（c）模型复杂性

模型复杂度是指算法在编写成可执行程序后，运行时所需要的资源，包含时间复杂度和空间复杂度。论文采用模型训练时间，测试时间，参数量和计算量进行比较。由于 CTTR 算法和本章算法均采用 Transformer 网络结构，实验中将主要对比这两种方法的复杂性，结果如表 3-1 所示。从表中可以看出，CTTR 算法的参数和时间复杂度为：61M Parms 和 187G FLOPs，而本章提出的算法参数和时间复杂度为：30M Parms 和 99G FLOPs，具有更少的参数和计算量，并且在测试时间和训练时间上均少于 CTTR 算法。

表 3-1 CTTR 算法和 HDTransformer 算法复杂情况对比

Table3-1 Comparison of complexity between CTTR and HDTransformer

	训练时间			测试时间	参数	计算量
	Epoch =1	Epoch =10	Epoch = 50			
CTTR	16 min	152 min	635 min	3.4 min	61 M	187 G
HDTransformer	8 min	81 min	376 min	5 min	30 M	99 G

3.5 本章小结

由于扫描角度的缺失，导致重建后的 CT 图像中存在大量条形伪影和斑点噪声，单一图像域的复原难以达到满意的效果。为此，本章采用联合投影域及图像域处理策略，设计了基于 Transforme 的混合域成像网络 HDTransformer，以提高稀疏角度 CT 图像重建质量。HDTransforme 中，第一阶段图像域采用耦合 Transformer 网络，以融合传统 CNN 和 Transformer 的优势，能极大程度的提取有效信息实现伪影噪声的去除。第二阶段投影域采用线性叠加连接的网络设计思想，以尽可能的减缓投影数据过处理现象的发生，同时为避免处理不均匀问题，实施中采用数据分块策略进行网络训练。第三阶段的图像域设计的轻量化 Transformer 网络在减少计算量和内存空间的同时，实现对 CT 图像进一步的细节微调。实验结果表明，本章所提 HDTransformer 网络可适用于不同稀疏程度的 CT 数据，与其他 CT 图像算法相比，图像质量均有明显的提升，并优于对比算法结果。在 PSNR 和 FSIM 的量化指标上同样也取得了最好的结果。总的来说，本章所提模型在抑制噪声伪影的同时，可有效的保留了 CT 图像的结构和纹理信息，并获得了令人满意的定性指标。后续研究可进一步优化处理流程，提高成像效率。

第 4 章　基于 Transformer 生成对抗网络的稀疏角度 CT 图像伪影抑制

基于混合域的 HDTransformer 网络在去除图像伪影，提高 CT 成像质量上有着不错的效果，但是网络处理流程采用了三阶段，导致算法实施的复杂性较高。而单一网络在去除伪影方面难以达到理想效果，图像边缘细节的处理方面存在不足，为此本章提出一种基于 Transformer 生成对抗网络的稀疏角度 CT 伪影抑制算法（A Transformer based Generative Adversarial Network Algorithm for Sparse View CT Imaging，SVT-GAN），该方法的主要思想是使用 Transformer 模块构建了一个基于 Transformer 的编码器-解码器生成器网络，在 U 型网络中采用 Transformer 模块取代卷积滤波器可获得高质量的全局信息。同时，为实现 CT 图像的边缘细节的复原，将 Transformer 生成器网络和判别器模块组合在一起，构建了一个基于 Transformer 模块的生成对抗性网络 SVT-GAN，获得了优质的稀疏角度 CT 图像伪影抑制。生成对抗网络在生成器的架构上，同时整合了一个判别器，该判别器的任务是对生成器的输出与原始图像的真实度进行评估。这种设计促进了网络自我优化的过程，尤其是在捕获和再现复杂图像纹理特性方面。由此，处理后的图像能更逼真地逼近目标数据，显示出更强的细节表现力。此外，为进一步提升网络性能，还设计了混合损失函数。实验的定性和定量结果表明，所提算法在去除图像伪影噪声方面要优于其他算法，具有更好的伪影抑制效果。

4.1 基于 Transformer 块的生成对抗网络

依据 GAN 的对抗性学习理念，将稀疏角度 CT 图像的伪影消除任务拆分为两部分：一是生成去伪后的图像，二是判别生成图像与全角度 CT 图像的真伪。生成器的主要训练目的是减少稀疏角度 CT 图像中的伪影，以产生与全角度 CT 图像相似的复原图像。而判别器的任务是尽可能精确地区分真实的全角度 CT 图像和生成的复原图像，以此促进生成器生成更优的图像质量。因此，网络设计的核心挑战在于提升生成器消除伪影的效能和判别器辨别真实与虚假样本的能力。如图 4-1 所示，本章提出了一种用于稀疏角度 CT 图像伪影抑制的 SVT-GAN，其中包括一个基于 Transformer 模块的生成器和一个类似 VGG-128 网络的判别器。在生成器模块中，采用 4 级对称编码器将低级特征转换成深层特征，引入了一个多头部转置注意力模块和门控前馈网络。其中，多头部转置注意力模块能够聚合局部和非局部像素交互，并且足够有效地处理高分辨率图像，门控前馈网络可以执行受控特征转换，即抑制信息较少的特征，只允许有用的信息通过网络层次进一步传递。此外，训练中为生成器和判别器设计了不同的损失函数，并构通过构造混合损失函数来提高各网络的性能。

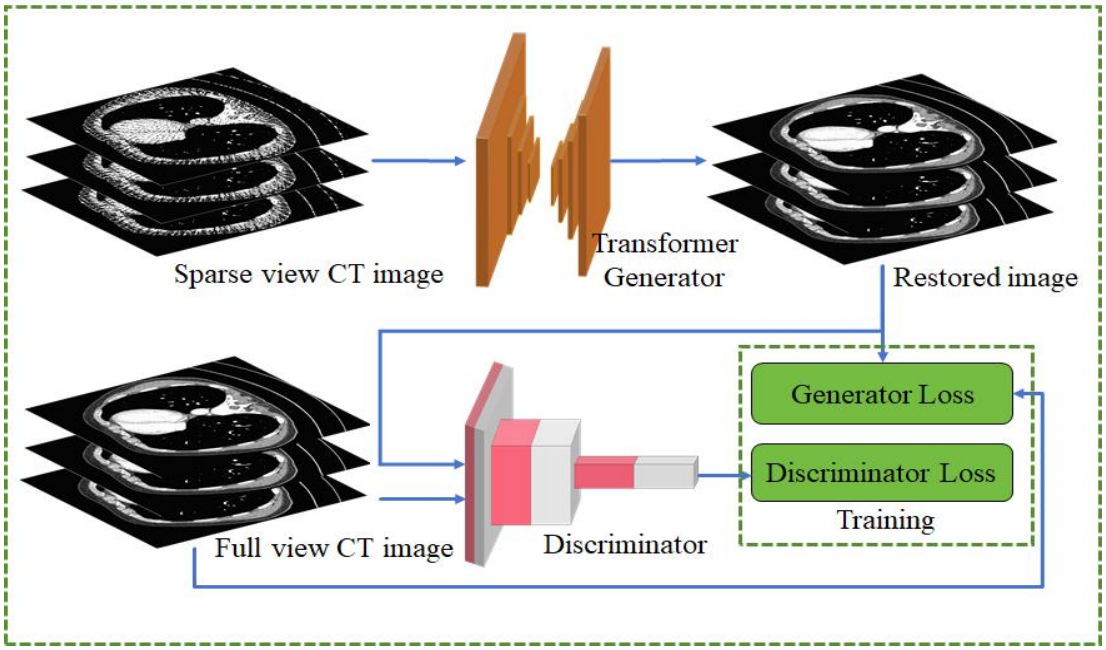


图 4-1 SVT-GAN 流程图

Fig. 4-1 The workflow of SVT-GAN

4.2 网络设计

4.2.1 生成器

在本章中，研究的目标是设计一个基于 Transformer 的稀疏角度 CT 图像伪影抑制模型，该模型以退化的稀疏角度 CT 图像作为输入，输出高质量的伪影抑制图像。一般设 X 为稀疏角度 CT 图像，Y 为其对应的全角度 CT 图像，两者的关系可表示为:

$$X = T^{-1}(Y) \tag{4-1}$$

式（4-1），$T: Y \to X$ 表示全角度 CT 图像通过投影下采样并重建出的稀疏角度 CT 图像的复杂退化过程。数学上，稀疏角度 CT 图像恢复问题可以建模为：

$$\arg\min \left\| T(X) - Y \right\|_2^2 \tag{4-2}$$

受第三章 Transformer 网络的启发，本章设计了一种高效的 Transformer 模型作为生成器，可以抑制稀疏角度 CT 图像的伪影。生成器 G 的结构如图 4-2 所示。给定稀疏角度 CT 图像 X，生成器 G 首先对其进行卷积，得到底层特征 F_0。然后，通过 4 级对称编码器将这些低级特征转换为深层特征 F_d。每个编码器-解码器级别包含多个 Transformer 块。从固定输入开始，编码器分层减少了空间大小，同时扩大了信道容量。解码器分层以低分辨率的潜在特征作为输入，逐步恢复高分辨率的表示。编码器功能通过跳跃连接，连接到解码器特征输入中。网络采用

像素对消和像素置乱操作分别作为特征下采样和上采样。为了减少通道和计算量，在拼接操作后采用1×1卷积。通过 Transformer 模块进一步丰富深度特征 F_d，进行高空间分辨率特征提取。最后，在残差图像 R 上加入退化图像，得到恢复图像：$X_G = X + R$。Transformer 块结构如图 4-2 所示，主要包括线性层、前馈网络、多头转换注意力模块（Multi-Dconv Head Transposed Attention，MDTA）和门控-直流前馈网络（Gated-Dconv Feed-Forward Network，GDFN）。MDTA 和 GDFN 的核心组成如图 4-3 所示。

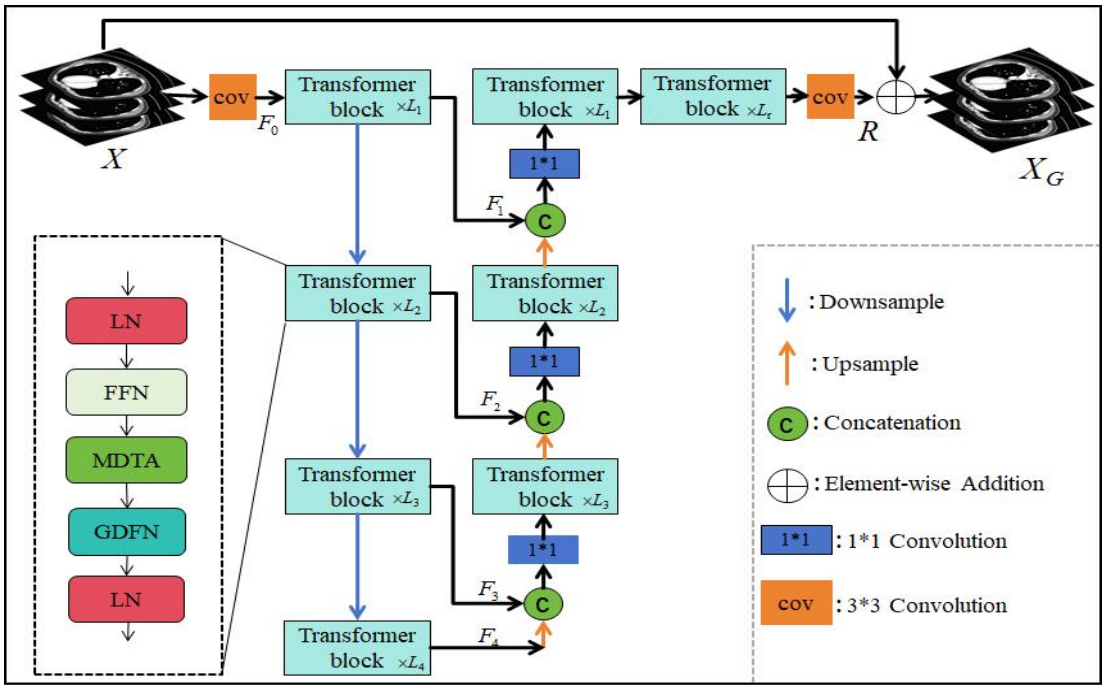


图 4-2 生成器 G 的网络结构

Fig. 4-2 The network architecture of generator G

为了实现特征变换，常规 FFN 分别对每个像素位置进行相同的操作。它使用两个1×1的卷积，一个用于扩展特征通道（通常是 4 倍），第二个用于将通道减少到原始输入维度。在隐藏层中加入非线性连接。MDTA 首先从层归一化张量出发，通过应用1×1和3×3卷积生成 Query，Key 和值 Value，丰富了局部上下文。接下来，通过修改查询和键投影，使它们的点积交互产生一个转置的注意力图。该模块的关键因素是 SA 的跨通道应用，而不是空间维度，它通过计算跨通道协方差生成全局上下文注意图的隐式编码。MDTA 进程定义如下：

$$\begin{aligned} Q &= R\left(W_1^Q W_3^Q LN(S)\right) \\ K &= R\left(W_1^K W_3^K LN(S)\right) \\ V &= R\left(W_1^V W_3^V LN(S)\right) \\ \hat{S} &= W_1 V \bullet soft\max\left(K \bullet Q \big/ \alpha\right) + S \end{aligned} \tag{4-3}$$

式中 S 、$\hat{S}$ 为输入、输出特征张量；$LN(\bullet)$ 为层归一化操作；$W_1^{(\bullet)}$ 为1×1逐点卷积，$W_3^{(\bullet)}$ 为3×3深度卷积；$G(\bullet)$ 表示维度变换操作；α 是一个可学习的缩放参数，用于控制大小。MDTA 模块可以聚合局部和非局部像素交互，有效地处理稀疏角度 CT 图像。在 MDTA 之后，Transformer 块进一步采用 GDFN 来改善表征学习。GDFN 被设计为两个线性投影层的逐元乘积，并制定了门控机制，其中一个是 GELU 激活函数。门控机制基于局部内容混合来强调空间语境。给定一个输入特征张量 S，GDFN 表示为:

$$\hat{S} = W_1^c G\left(W_1^a W_3^a LN(S) \Theta W_1^b W_3^a LN(S)\right) \tag{4-4}$$

式中 $G(\bullet)$ 表示 GELU 非线性，Θ 表示逐元素乘法。GDFN 执行控制信息流，通过通道中的各个层次，从而允许每个层次都专注于与其他层次互补的细节。与 FFN 相比，GDFN 实现了特征信息的互补选择。

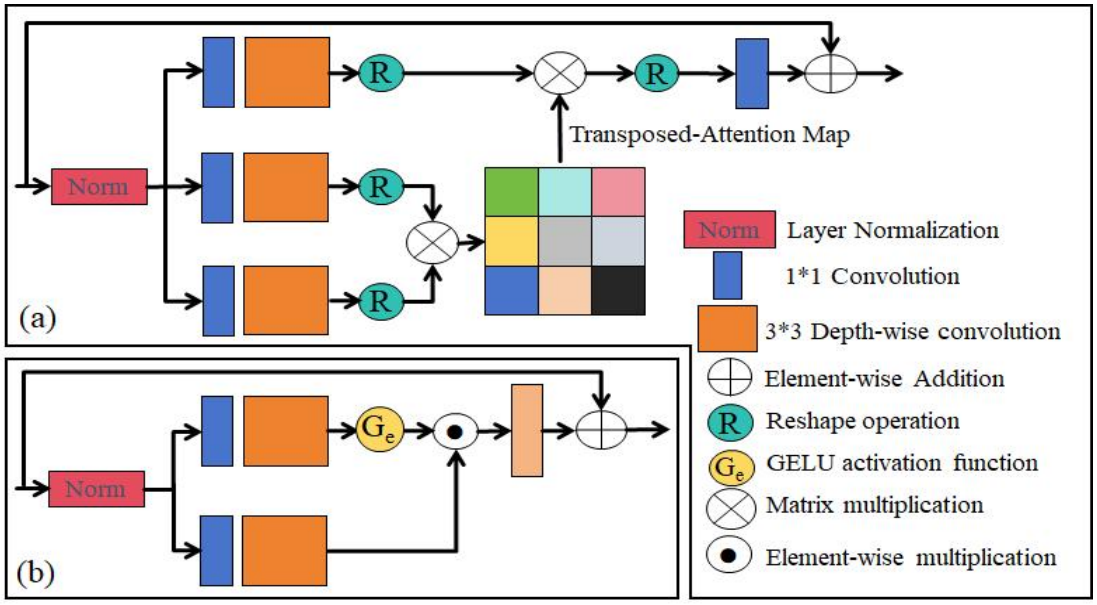


图 4-3（a）MDTA 模块，（b）GDFN 模块

Fig. 4-3 (a) MDTA module, (b) GDFN module

4.2.2 判别器

基于深度学习的图像处理方法，主要通过最小化参考图像与预测图像之间的

逐像素误差来进行拟合的，但像素误差最小化可能导致处理后的图像出现模糊效果，细微的结构也可能发生变形。GAN 在解决上述问题方面具有很大的潜力。作为一种强大的生成模型，GAN 通过交替训练两个具有不同目标的部分来工作。其中，判别器是区分标签样本和预测样本的分类网络，而生成器则产生预测样本，尽可能地混淆判别器。利用对抗学习策略，GAN 在各种任务上表现出了优异的性能。在这项工作中，本章采用平铺结构的判别器 D 来进一步提高生成器 G 的预测精度。本章设计的判别器 D 结构如图 4-4 所示，网络输入数据大小为 3×512×512。通过层间冗余性质，可以在不影响局部相关性的情况下增强判别器的表示能力。除了第一层，所有卷积层的卷积核大小都设置为 3×3。卷积处理可以从图像中提取不同层次的特征。最后，特征图通过全连接操作来区分输入图像的真实性。

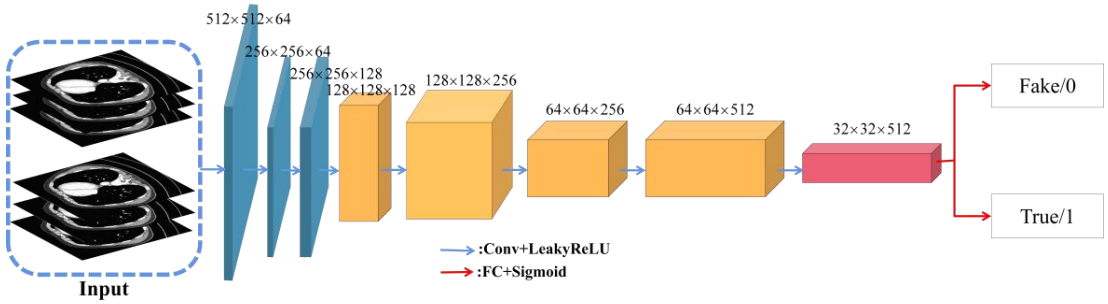


图 4-4 判别器 D 结构

Fig. 4-4 The architecture of the discriminator D

4.2.3 混合损失函数

在模型训练中，本章利用混合损失函数来达到期望的伪影抑制性能。最初使用像素级 L_1 损失函数作为混合生成器损失之一，主要用于计算预测图像与全角度标签 CT 图像之间的像素误差。定义如下：

$$L_{pixel}\left(X_n, Y_n, \Theta_G\right)=\frac{1}{N}\sum_{n=1}^{N}\left|G\left(X_n\right)-Y_n\right| \tag{4-5}$$

其中 X_n 和 Y_n 分别表示稀疏角度的 CT 图像和全角度 CT 图像。$G(\bullet)$ 表示生成器 G 的预测，Θ_G 为生成器 G 的参数，N 为样本个数。然而，作为一个平滑函数，L_1 注重于于临床相关特征空间中的像素距离，而不是测地线距离，容易导致过平滑。为了缓解这一问题，实验中进一步引入了多尺度感知损失。多尺度感知损失中使用预训练的 ResNet 提取特征映射。具体来说，计算中 ResNet 网络中最后一个线性层和平均池化层被移除，以形成特征提取网络。然后，将预测 CT 图像和全角

度标签 CT 图像数据输入到特征提取器网络，计算不同尺度下的 MSP 损失，公式如下：

$$L_{MSP}(X_n,Y_n,\Theta_G)=\frac{1}{NC}\sum_{n=1}^{N}\sum_{c=1}^{C}\left\|E_c(G(X_n))-E_c(Y_n)\right\|_2^2 \tag{4-6}$$

式中 $E_c(\bullet)$ 为尺度 c 下的提取器。因此，生成器 G 训练的目标是最小化以下混合损失函数 L_G：

$$\min_G L_G(X_n,Y_n,\Theta_G)=L_{pixel}(X_n,Y_n,\Theta_G)+\alpha\bullet L_{MSP}(X_n,Y_n,\Theta_G) \tag{4-7}$$

其中 α 是预定义的权重。

通过增加额外的对抗学习，SVT-GAN 将引导生成器 G 产生具有高细节保存和对比度的结果。SVT-GAN 中判别器 D 的损失函数定义如下：

$$\begin{aligned}&\min_G\min_D L_{GAN}(X,Y,\Theta_G,\Theta_D)=-E_Y[D(Y)]+E_{G(X)}[D(G(X))]\\&+\lambda E_{X'}\left[\left(\left\|\Delta_{X'}\left[\Delta_{X'}D(X)\right]\right\|_2-1\right)^2\right]\end{aligned} \tag{4-8}$$

其中 λ 为惩罚系数。X' 沿连接 $G(X)$ 和 Y 的直线均匀采样。Θ_D 表示判别器 D 的参数。为了获得稳定的训练，本章采用带梯度惩罚的 Wasserstein GAN 对 Θ_D 进行优化。因此，在 SVT-GAN 中，目标函数的最小值为公式（4-9）所示:

$$\min_G\min_D L_{GAN}(X,Y,\Theta_G,\Theta_D)+\beta\bullet L_G(X,Y,\Theta_G) \tag{4-9}$$

其中 β 是用来平衡不同损失值的权重参数。对于网络训练，本章采用默认参数的 Adam 方法对损失函数进行优化。

4.3 实验结果与分析

4.3.1 实验设计

A. Mayo 数据及实验参数

为了验证本章提出的模型的性能，实验中采用 Mayo Clinic 研究人员制作的 AAPM 数据集进行训练和验证。该数据集包含 10 例患者的 4825 张 1mm 厚的高质量 CT 图像。通过 FBP 算法从 720 个角度的投影数据中重建出高质量的图像，可以认为是全角度 CT 图像(标签)。然后从全角度投影数据中等间隔采样出 90、120、180 三种不同投影角度的数据，并重建稀疏角度 CT 图像。在网络训练过程中，将 4825 个单层的 CT 图像数据对通过滑动窗口重复采样为 4800 个连续的三层数据对，步长设置为 1，实验的输入数据大小为 3×512×512。所有数据均归

一化为零均值和单位方差。本实验将损失函数中α和β的最佳权重参数设为 1 和10^4。batchsize size 为 64，初始通道数为 96，训练周期 epoch 设置为 100，学习率的初始值为2×10^{-4}，学习率递减到最小值为1×10^{-6}。

B. UIH 数据准备

为了验证所提出的 SVT-GAN 在不同数据集上的效果，将从联合形象医疗（UIH）收集的真实患者数据作为另一个测试集，以丰富实验结果。该数据集包括 26 名匿名患者的全角度 CT 图像和相应的模拟稀疏角度 CT 图像本章从全角度投影数据中提取 60 和 120 种不同数量的投影数据来重建。所有的 CT 图像都用 512×512 像素的 FBP 重建。随机选择 24 名患者作为训练集，1 名患者作为验证集；其余 1 例患者作为试验集。本实验将损失函数中α和β的最佳权重参数设为 1 和10^4。batchsize size 为 64，初始通道数为 96，训练周期 epoch 设置为 100，学习率的初始值为2×10^{-4}，学习率递减到最小值为1×10^{-6}。

4.3.2 实验结果

A. Mayo 实验结果

图 4-5、图 4-6、图 4-7 分别显示了 90 个投影角度数据、120 个投影角度数据和 180 个投影角度数据下重建图的处理结果。第一、二、三行分别为所选择的胸、腹、胯部采用不同方法的成像结果，显示窗口为[-150HU，250HU]。第一列为全角度下的 FBP 重建图像，作为参考。第二列为稀疏角度 FBP 的重建结果，通过对比可以看出，由于缺乏投影数据，传统的解析重建算法得到的图像包含大量伪影，细节信息被伪影所覆盖。在图 5-5 中，随着扫描角度的减小，重建图像中的伪影也更加严重。通过对比第 3 列、第 4 列的结果，可以看到不同的方法对条纹伪影的抑制程度不同，有的甚至会降低图像分辨率（如 PRestor 和 StarGAN）。从图 4-5 的第五列和第六列可以看出，稀疏角度 CT 图像的质量有了明显的改善。通过观察图 4-5 中局部放大图，SVT-GAN 在精细结构保存方面明显优于对比方法（如红色箭头所示）。此外，通过局部放大图可以看出，CTTR 会导致一些微小的图像结构丢失，细节对比度降低，出现过平滑现象。从图 4-6 和图 4-7 可以看出，由于投影角度的增加，所有方法获得的图像质量逐渐提升了。相对而言，可以看到，由于 SVT-GAN 可以获得更多的全局信息，所以结果在图像的清晰度和细节保持上仍然是最好的，并且更接近参考图像。SVT-GAN 的整体图像质量优于其他四种对比方法。

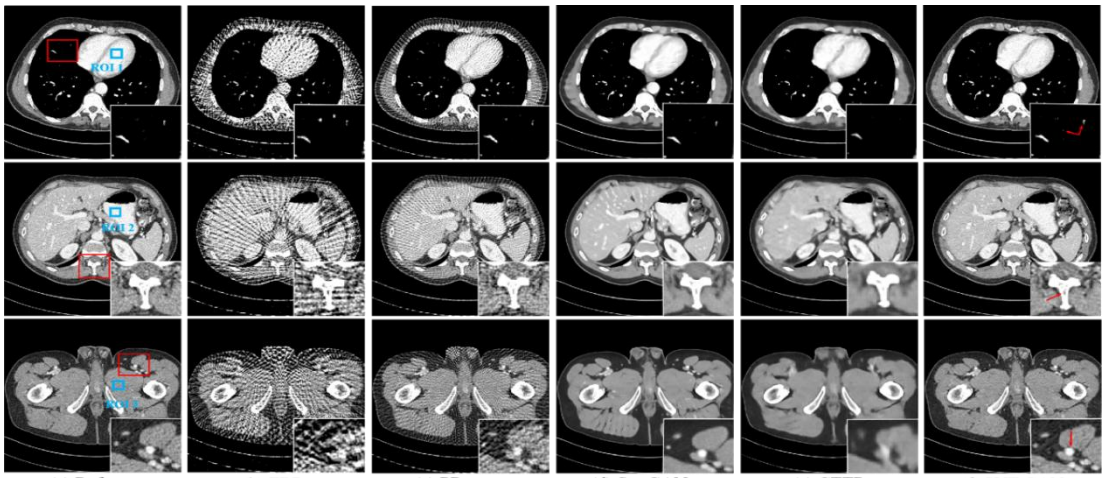


图 4-5 90 个扫描角度下不同方法的 CT 成像效果图

Fig. 4-5　Axial view results of different methods under 90 views

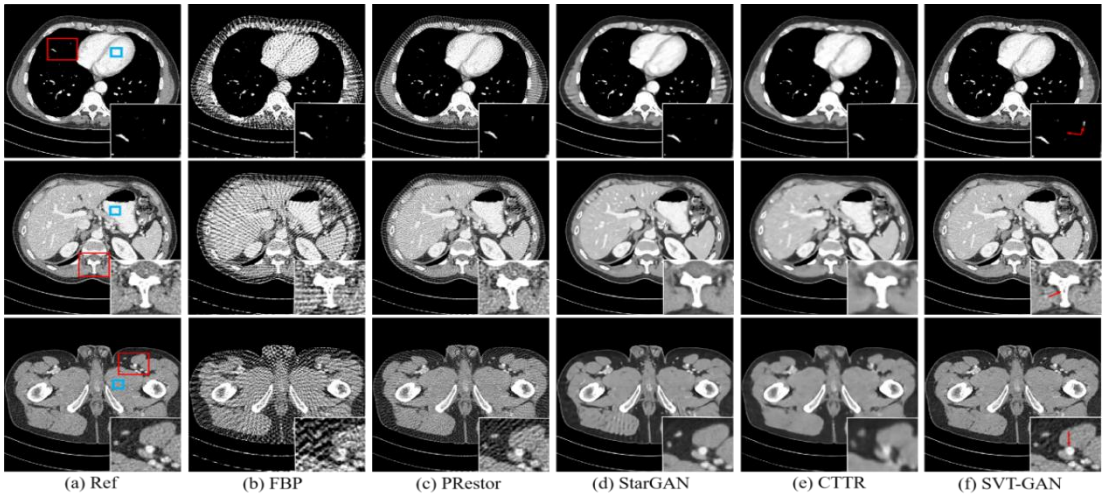


图 4-6 120 个扫描角度下不同方法的 CT 成像效果图

Fig. 4-6　Axial view results of different methods under 120 views

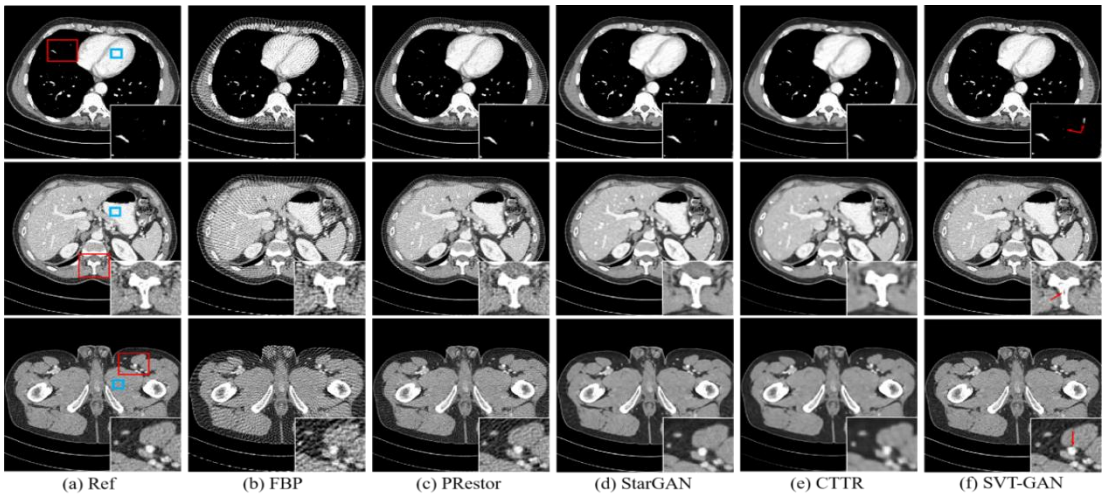


图 4-7 180 个扫描角度下不同方法的 CT 成像效果图

Fig. 4-7　Axial view results of different methods under 180 views

为了进一步验证不同方法在冠状位和矢状位图像处理中的性能，实验选取了90 个投影角度下的处理结果进行比较，如图 4-8 所示，显示窗口为[-150HU, 250HU]。从图 4-8 中可以看出，冠状面和矢状面图像的噪声伪影相对较少，相对于横断面图，扫描角度的减小对这两个视角上的图像的影响并不明显。PRestor 和 StarGAN 方法处理的结果，虽然图像伪影程度较低，但图像的整体质量不高，部分细节缺失。同时，可以发现在胯部位置仍有一些高强度的伪影。图 4-8（e）中 CTTR 方法在能够很好的除噪声伪影，图像清晰度也得到了大幅度的提高，但对组织的纹理保留度不高，与参考图像存在明显差异。从本章所提 SVT-GAN 方法的处理结果上看，能显著减少了整个区域的伪影，并且更好地保留了低对比度组织区域的边缘，组织纹理更加接近参考图。视觉结果表明，所提 SVT-GAN 方法能显著提高稀疏角度 CT 图像的质量。

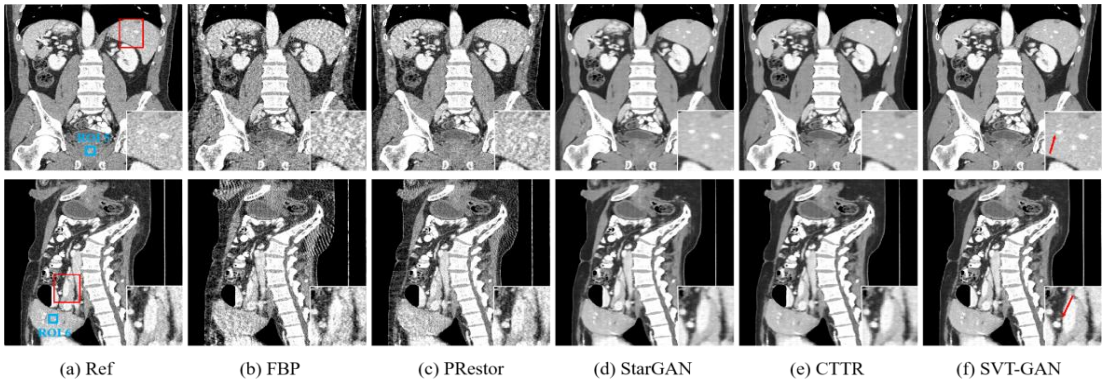


图 4-8 90 个扫描角度下矢状面和冠状面的成像效果图

Fig. 4-8 Coronal and sagittal images of the different methods under 90 views

图 4-9 为不同处理方法的结果与参考图像间的差值图。显示窗口为[0HU, 200HU]。差值图像反映了处理后的图像与全角度 CT 图像的接近程度。通过对比不同方法处理后图像与参考图的差异，可以看到，三种对比方法处理后的 CT 图像具有更多的伪影残留信息，这些残留信息与参考图像有明显的结构差异。然而，实验中，SVT-GAN 差值图像的强度更低，也更平坦，这表明 SVT-GAN 获得的图像更接近参考图像，要优于对比方法。综上所述，从视觉效果的角度来看，SVT-GAN 方法的处理性能优于对比方法。

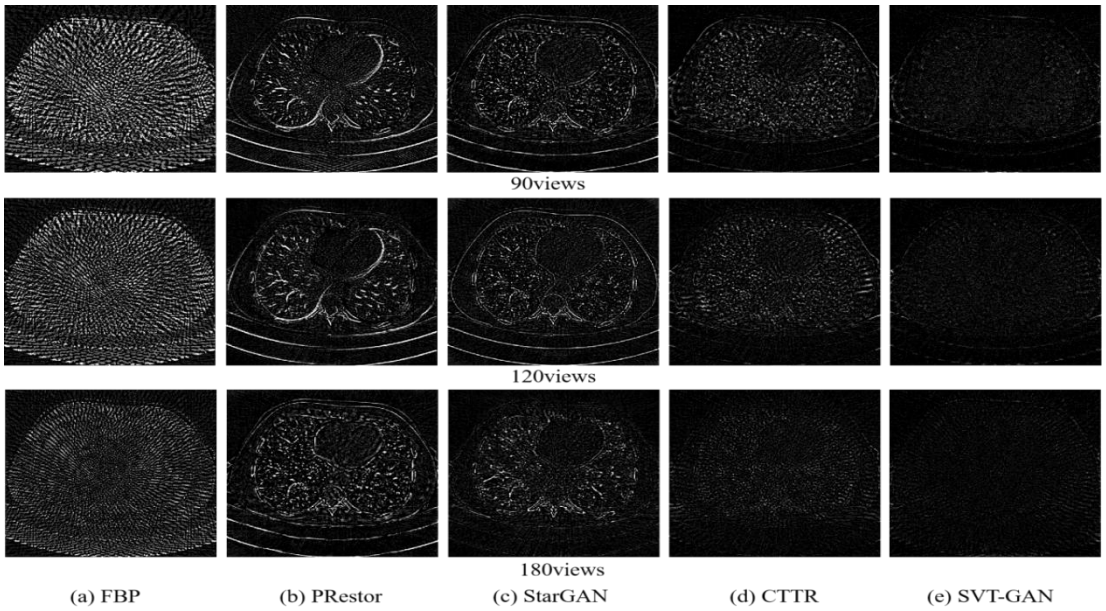


图 4-9 不同方法获得的 CT 图像的差异图

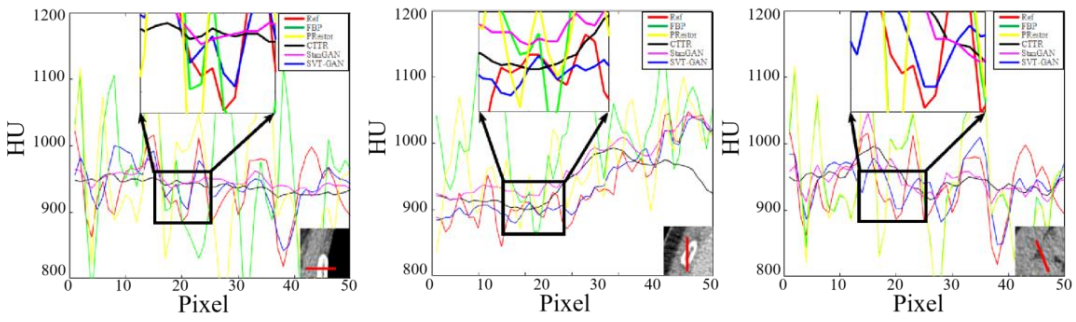
Fig. 4-9 Difference images relative to the reference image obtained by different methods

表 4-1 为 90 和 120 个角度下不同方法处理结果的量化值（PSNR，SSIM 和 FSIM 指标）。对比表 4-1 中的均值和方差可以发现，本章提出的 SVT-GAN 获得了最好的定量结果，其次是 CTTR。虽然 CTTR 在 90 角度数据中获得了更好的 PSNR 结果，但图 4-5 和图 4-8 所示的视觉结果证实 CTTR 产生了过平滑的结果。从表中可以看到，量化结果与可视化结果一致，SVT-GAN 获得了最佳的 PSNR，SSIM 和 FSIM 值，验证了所提方法的优势。为了评估像素级的 HU 拟合情况，图 4-10 中绘制了不同方法处理后图像感兴趣区域的 CT 值轮廓。从图中可以发现，在 90 个角度和 120 个角度数据情况下，SVT-GAN 的拟合性能优于 StarGAN 和 CTTR，表明 SVT-GAN 具有更好的结构保存性能。。

表 4-1 不同方法获得的 Mayo 数据的 PSNR、SSIM 和 FSIM 值

Table 4-1. PSNR, SSIM and FSIM values of Mayo data obtained by different methods

Views	90 views			120 views			180 views		
Metrics	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
FBP	28.33 ±0.77	0.73 ±0.031	0.72 ±0.032	32.12 ±0.68	0.76 ±0.029	0.75 ±0.029	34.24 ±0.65	0.78 ±0.026	0.77 ±0.027
PRestor	33.53 ±0.63	0.77 ±0.022	0.75 ±0.032	34.52 ±0.65	0.82 ±0.020	0.81 ±0.028	37.35 ±0.61	0.85 ±0.018	0.84 ±0.025
StarGAN	36.36 ±0.52	0.81 ±0.014	0.78 ±0.023	37.34 ±0.59	0.85 ±0.012	0.82 ±0.019	39.65 ±0.55	0.88 ±0.011	0.85 ±0.016
CTTR	38.40 ±0.67	0.83 ±0.018	0.79 ±0.023	40.65 ±0.62	0.86 ±0.016	0.83 ±0.018	41.87 ±0.59	0.91 ±0.013	0.88 ±0.016
SVT-GAN	42.47 ±0.49	0.94 ±0.013	0.91 ±0.018	43.32 ±0.32	0.93 ±0.011	0.92 ±0.015	46.34 ±0.29	0.96 ±0.010	0.95 ±0.013



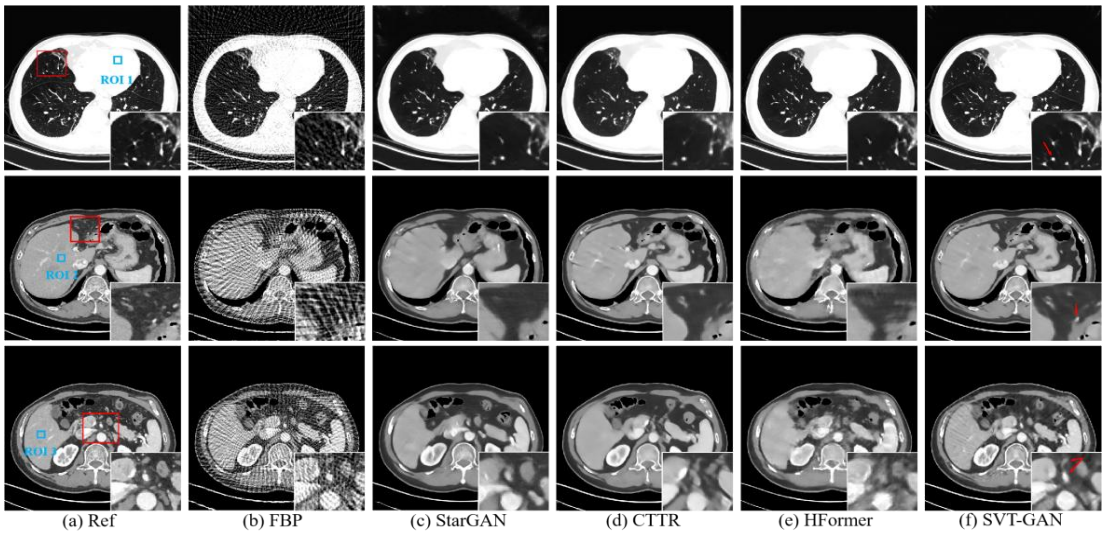
(a) 90 views (b) 120 views (c) 180 views

图 4-10 不同方法在三个横断面图中沿指定红线的 CT 强度图

Fig. 4-10 The intensity profiles along the specified red line in the three axial slices of different methods

B. UIH 实验结果

图 4-11，4-12 分别展示了 120 个投影数据和 200 个投影数据下的重建结果，腹部图像的显示窗口为 [-150HU, 250HU]，肺部图像的显示窗口为 [-900HU,100HU]。第一列为全角度下的 FBP 重建图像，作为参考。对比图 4-11 和图 4-12，可以看到随着扫描角度数量的减小，重建图像中的伪影也更加严重。而且，不同的方法对条纹伪影的抑制程度不同，有的甚至会降低图像分辨率（如 CTTR 和 StarGAN）。从图 4-11 的第五列和第六列可以看出，通过网络学习推断，稀疏角度 CT 图像有了明显的改善。通过观察图 4-11 中局部放大图，SVT-GAN 在精细结构保存方面明显优于其他方法（如红色箭头所示）。此外，通过局部放大图可以看出，HFormer 会导致一些微小的图像结构丢失，细节对比度降低，出现过平滑现象。从图 4-12 可以看出，由于投影角度的增加，所有模型的图像都更加清晰了。相对而言，可以看到，由于 SVT-GAN 可以获得更多的全局信息，所以结果在图像的清晰度和形状上仍然是最好的，并且更接近参考图像。SVT-GAN 的整体图像质量优于其他四种比较方法。



(a) Ref (b) FBP (c) StarGAN (d) CTTR (e) HFormer (f) SVT-GAN

图 4-11 120 个扫描角度下不同方法的 CT 成像效果图

Fig. 4-11 Imaging results of different methods under 120 views

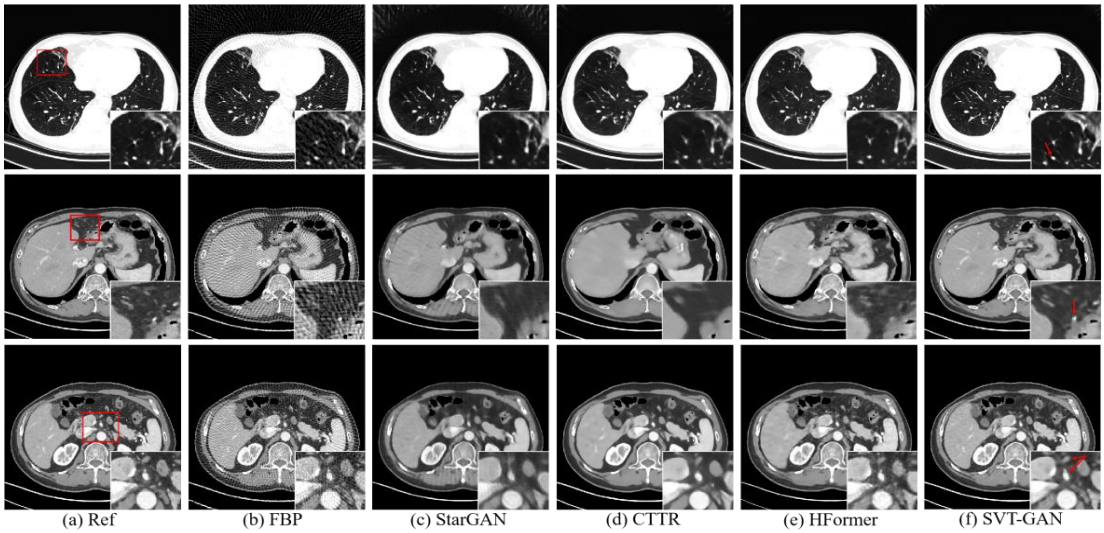


图 4-12 200 个扫描角度下不同方法的 CT 成像效果图

Fig. 4-12 Imaging results of different methods under 200 views

为了进一步验证不同方法在冠状面和矢状面图像处理中的性能，实验选取了 200 个投影角度下的处理结果进行比较，如图 4-13 所示，显示窗口为[-150HU, 250HU]。从图 4-13 中可以看出，冠状面和矢状面图像的噪声伪影相对较少，相对于横断面图，扫描角度的减小对这两个角度上的图像的影响并不明显。StarGAN 方法处理的结果，虽然图像伪影程度较低，但图像的整体质量不高，部分细节缺失。同时，可以发现在胯部位置仍有一些高强度的伪影。图 4-13（d）中 CTTR 方法和图（e）HFormer 能够很好的除噪声伪影，图像清晰度也得到了大幅度的提高，但对组织的纹理保留度不高，与参考图像存在明显差异。从本章所提 SVT-GAN 方法的处理结果上看，能显著减少了整个区域的伪影，并且更好地保留了低对比度组织区域的边缘，组织纹理更加接近参考图。视觉结果表明，所提 SVT-GAN 方法能显著提高稀疏角度 CT 图像的质量。

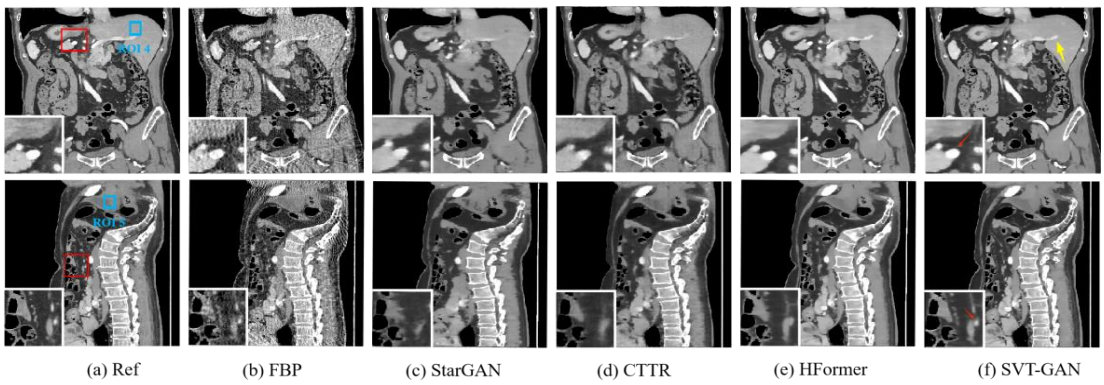


图 4-13 矢状面和冠状面下不同方法的 CT 成像效果图

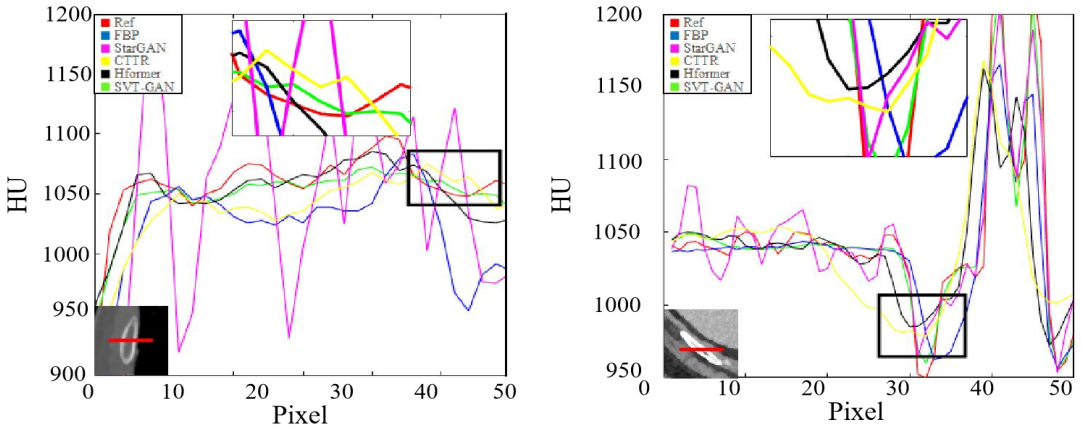
Fig. 4-13 Imaging results of different methods under sagittal plane and coronal plane

表 4-2 为 120 和 200 个角度下不同方法处理结果的量化值（PSNR，SSIM 和 FSIM 指标）。对比表 4-2 中的均值和方差可以发现，本章提出的 SVT-GAN 获得了最好的定量结果，其次是 HFormer。从表中可以看到，量化结果与可视化结果一致，SVT-GAN 获得了最佳的 PSNR，SSIM 和 FSIM 值，验证了所提方法的优势。为了评估像素级的 HU 拟合情况，图 4-14 中绘制了不同方法处理后图像的 CT 值轮廓。从图中可以发现，在 90 个角度和 120 个角度数据情况下，SVT-GAN 的拟合性能优于 StarGAN 和 CTTR，表明 SVT-GAN 具有更好的结构保存性能。

表 4-2 不同方法获得的 UIH 数据的 PSNR、SSIM 和 FSIM 值

Table 4-2. PSNR, SSIM and FSIM values of UIH data obtained by different methods

Views	120 views			200 views		
Metrics	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
FBP	30.58	0.60	0.54	34.70	0.74	0.64
	±0.78	±0.05	±0.023	±0.801	±0.047	±0.015
StarGAN	35.31	0.85	0.74	37.43	0.87	0.83
	±0.53	±0.017	±0.006	±0.48	±0.016	±0.005
CTTR	36.35	0.87	0.83	39.95	0.89	0.85
	±0.72	±0.012	±0.006	±0.54	±0.010	±0.005
HFormer	39.24	0.90	0.85	41.35	0.91	0.87
	±0.78	±0.012	±0.005	±0.56	±0.009	±0.004
SVT-GAN	43.13	0.92	0.90	43.77	0.97	0.93
	±0.69	±0.005	±0.003	±0.32	±0.002	±0.002



(a) 120 views　　(b) 200 views

图 4-14 不同方法在横断面图中沿指定红线的 CT 强度图

Fig. 4-14 The intensity profiles along the specified red line in the axial slices of different methods

4.3.3 网络模型分析

（a）消融实验

本章提出的 SVT-GAN 网络中，生成器 G 中的包含 FFN、MDTA、GDFN 等部分，同时训练中还包括判别器 D，为验证网络中各部分对稀疏角度 CT 图像伪

影的抑制效果，实验分别在 30、60、90 个角度投影数据下进行消融实验验证。该实验结果还可以探讨不同稀疏角度数据的鲁棒性。实验中，采用 MDTA + GDFN 表示不含 FFN 和 LN 的生成器 G，作为消融实验中的基线模型。接下来，MDTA + FFN 作为第一比较模型，生成器 G 作为第二比较模型。最后，将提出的 SVT-GAN 整体模型作为第三个比较模型。

图 4-15 为不同模型处理后图像，表 4-5 为定量结果。从图 4-15 可以看出，基线处理模型可以去除大部分伪影并能恢复一定的组织，但无法去除高强度伪影。与 MDTA + GDFN 和生成器 G 相比，可以发现生成器 G 中 FFN 能够带来一定的抑制噪声伪影效果。通过比较最后一列的结果，可以发现 SVT-GAN 中的对抗学习策略有助于恢复小细节，并减缓估计图像与参考图像之间的不一致问题。此外，PSNR 定量结果表明，本章使用的生成器 G 获得的图像比基线模型获得的高 0.51 dB。对抗学习要比单 Transformer 的生成器 G 网络具有更好的学习效果。重要的是，SVT-GAN 方法在 30 角度的情况也能产生几乎无伪影的图像。总体而言，SVT-GAN 获得了良好的视觉性能，图像结构清晰，在不同稀疏角度数据中具有一定的鲁棒性。

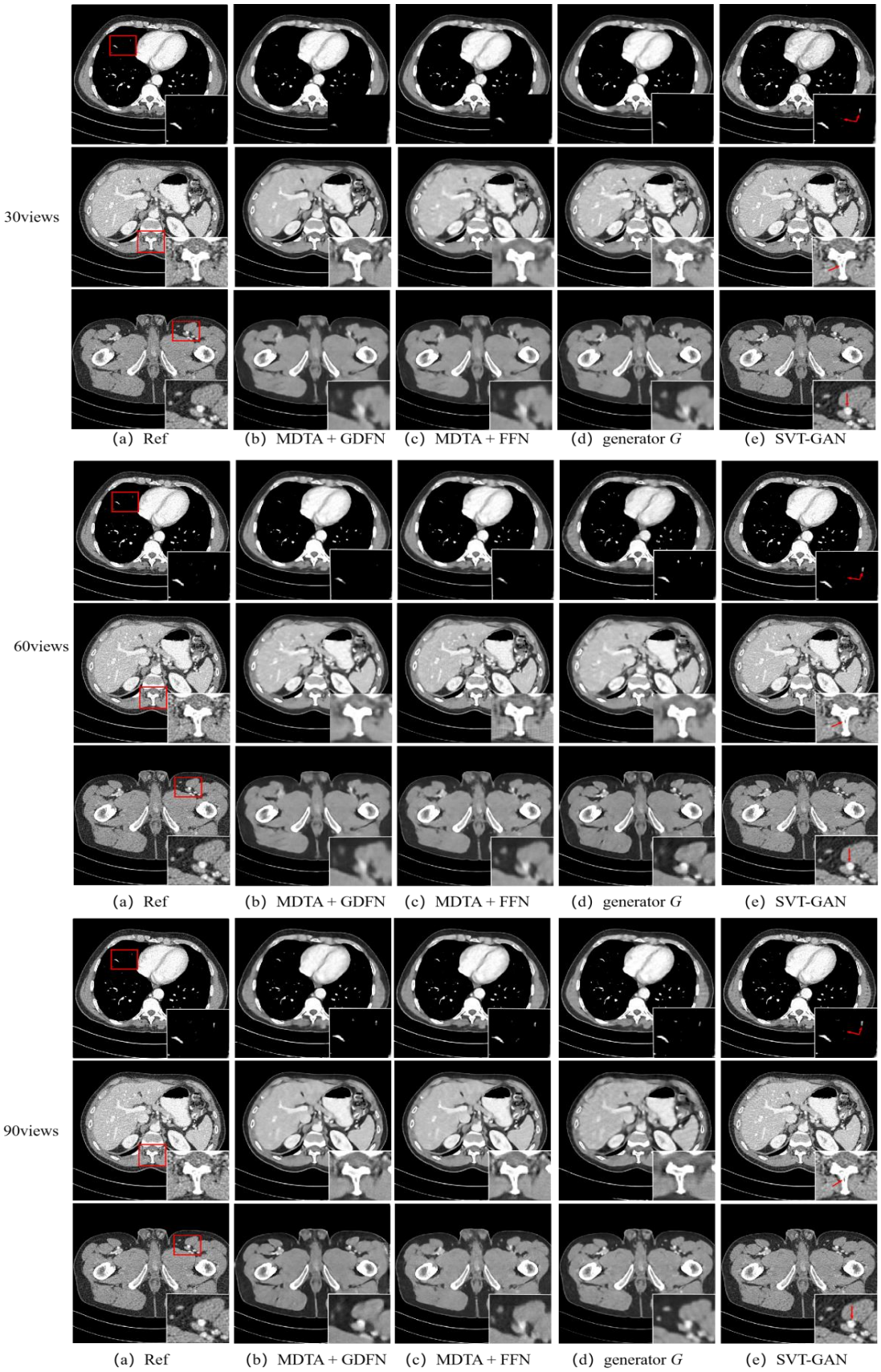


图 4-15 不同角度下消融实验的轴向结果

Fig. 4-15 The axial results for the ablation study under different views data

表 4-5 消融研究的定量结果

Table 4-5. Quantitative results of the ablation study.

Views	30 views			60 views			90 views		
Metrics	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
MDTA +	30.66	0.84	0.67	33.65	0.86	0.74	36.37	0.90	0.77
GDFN	±1.57	±0.03	±0.03	±1.24	±0.02	±0.03	±0.74	±0.01	±0.02
MDTA	31.41	0.87	0.69	35.92	0.89	0.75	38.26	0.91	0.77
+ FFN	±1.37	±0.03	±0.03	±1.11	±0.02	±0.03	±0.73	±0.01	±0.02
generator *G*	33.21	0.89	0.70	36.21	0.90	0.76	40.64	0.93	0.80
	±1.24	±0.03	±0.03	±1.04	±0.02	±0.02	±0.70	±0.01	±0.02
SVT-GAN	34.24	0.90	0.71	37.86	0.91	0.76	42.47	0.94	0.91
	±1.06	±0.02	±0.02	±0.88	±0.01	±0.02	±0.49	±0.013	±0.018

（b）参数分析实验

此外，在公式（4-7）和公式（4-9）中看出 SVT-GAN 具有与生成器损失相关的不同超参数的恢复性能。具体而言，本章采用了渐进改进策略来调整损失函数超参数。图 4-16 描绘了不同α在 90 角度下获得的视觉结果。从图 4-16 的局部放大图中可以看出，α的减小会导致一些微小的图像结构丢失，细节对比度降低，出现过平滑现象。当$\alpha=1$时，获得的 CT 图像视觉效果最好，图像更接近参考图。定量结果如表 4-6 所示。从表 4-6 可以看出，生成器只有像素损失没有产生优化的结果。从表 4-6 可以看出，恢复后的图像质量对 MSP 损失权值η比对抗性损失权值α更敏感。

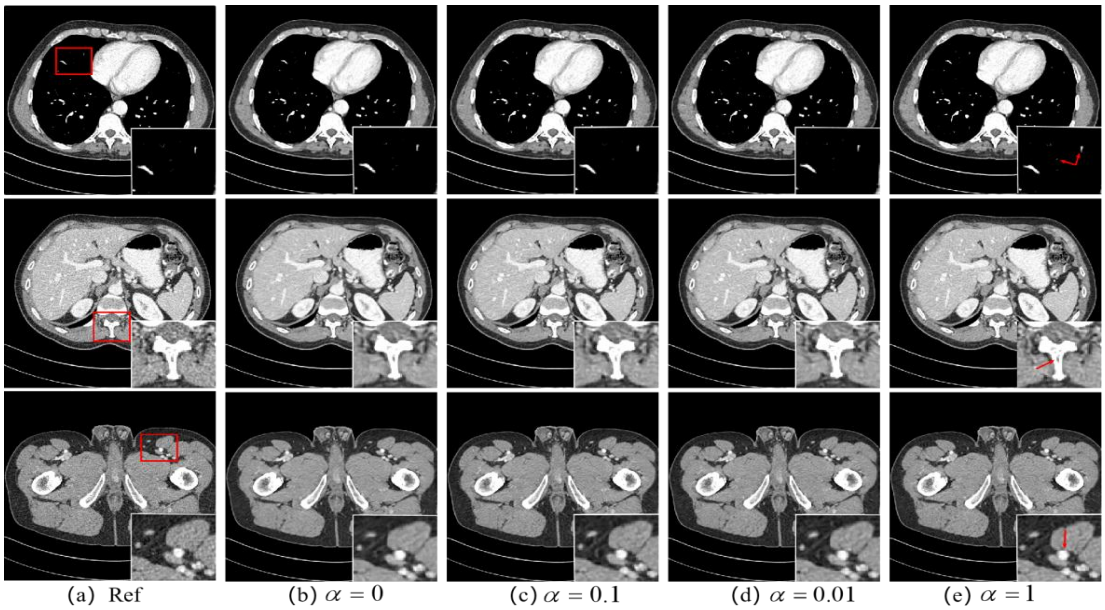


图 4-16 不同参数在 90 个角度下的结果

Fig. 4-16 The results of different parameters under 90 views

表 4-6 不同权重参数下的 PSNR 和 FSIM 值

Table 4-6: PSNR and FSIM values under different weight parameters

α	0	0	0	0	0.1	0.1	0.1	0.01	0.01	0.01	1	1	1
η	0	0.1	0.01	1	0	0.1	0.01	0	0.1	0.01	0	0.1	0.01
PSNR	42.3643	43.0975	42.7231	43.7563	42.8927	43.2020	42.8452	42.4021	42.6625	42.8731	43.0425	43.8216	43.7525
SSIM	0.8926	0.9021	0.9012	0.9138	0.9125	0.9102	0.9018	0.9017	0.9021	0.9075	0.9071	0.9183	0.9102
FSIM	0.9023	0.9181	0.9101	0.9235	0.9178	0.9217	0.9143	0.9076	0.9152	0.9209	0.9208	0.9252	0.9214

（c）收敛性验证实验

通过计算 PSNR 和 FSIM 分数，分析了 SVT-GAN 的收敛性。如图 4-17（a）所示，随着训练周期的增加，不同角度数据的 PSNR 值在 100 epoch 后逐渐增大，减小。从图 4-17（b）中可以看出，不同角度数据训练后的 FSIM 值可以达到最大值，然后略有下降。PSNR 和 FSIM 在 100 epoch 后下降，表明网络开始出现过拟合。为此，结合视觉效果、PSNR 和 FSIM，实验中训练 epoch 设置为 100。

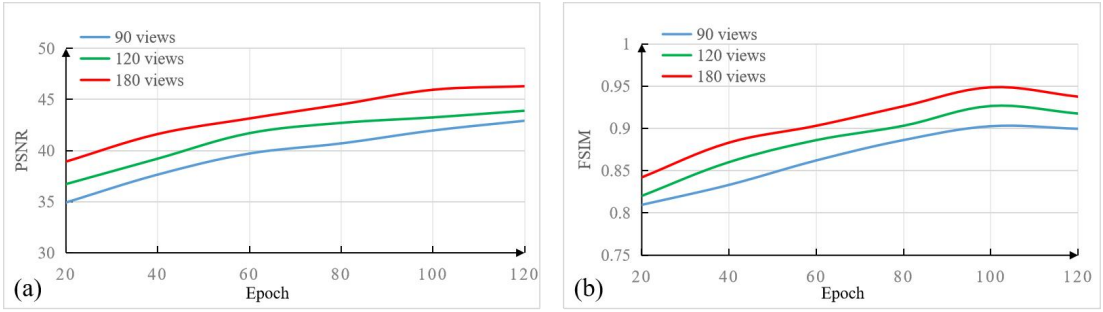


图 4-17 不同训练周期下 90、120、180 角度的 PSNR 和 FSIM 值

Fig. 4-17 PSNR and FSIM values for 90, 120, and 180 views under different epochs

（d）模型复杂性

实验以 120 个角度下的 300 张 CT 图像为测试数据，比较了不同方法的参数数量、FLOP 和运行时间，比较结果如表 4-7 所示。从表中可以看到，CTTR 方法和 HFormer 方法与本章提出的 SVT-GAN 均采用的 Transformer 网络结构。CTTR 的模型尺寸和计算复杂度分别为 61M 和 187G FLOPs，HFormer 的模型尺寸和计算复杂度分别为 1.6M 和 104G FLOPs，而 SVT-GAN 的参数和时间复杂度分别为 26M 和 98G FLOPs，计算复杂度更少。SVT-GAN 模型的测试时间和训练时间都比 CTTR 模型要短。

表 4-7 不同方法的模型大小和计算复杂度

Table 4-7. The model size and computational complexity of different methods

	Training time (min)			Test time (second / slice)	Params (M)	FLOPs (G)
	Iter =1000	Iter =5000	Iter = 10000			
StarGAN	18	123	401	3.8	53.2	192
CTTR	16	108	326	3.4	61	187
HFormer	14	92	251	2.8	1.65	104
SVT-GAN	5	30	64	1.2	26	98

4.4 本章小结

由于扫描角度的缺失，导致重建后的 CT 图像中存在大量条形伪影，干扰临床诊断。从图像处理的角度上来抑制伪影，具有较高的应用价值。为抑制噪声和伪影，提高稀疏角度 CT 图像的质量，促进稀疏角度 CT 成像的临床应用，本章改进了基于 Transformer 的网络，并以 GAN 为学习策略，来促进网络性能的提升。生成器网络采用了 MDTA 模块，提高了局部和非局部特征的提取能力，在结构和细节恢复方面具有竞争力；进一步，生成器网络中还采用 GDFN 操作来抑制特征转换；在训练过程中，专门设计了一个混合损失函数来提高网络性能。Mayo 数据和 UIH 数据的实验结果表明，本章所提网络模型在不同的稀疏角度图像的伪影抑制中均优于对比方法，具有较好的鲁棒性，并在 PSNR，SSIM 和 FSIM 等不同定量指标上取得了最佳效果。消融实验结果验证了网络各部分的设计在伪影抑制中的有效性，模型在不同数据中均表现出一定的收敛性，模型容量及计算效率也优于对比方法。

第 5 章 双域 Transformer 生成对抗网络的稀疏角度 CT 成像

稀疏角度 CT 图像成像过程复杂，由于不完全的扫描数据，导致图像充满斑点噪声和条状伪影。在生成对抗网络的架构中，扩展了基本的生成模型，引入了一个判别器，其任务是对生成样本与真实样本进行区分。通过这样的二元对立过程，不断微调生成器的性能。这种对抗性的学习策略形成了一个动态的博弈环境，有助于网络更有效地学习和表达复杂的图像纹理特性。因此，经由该模型预测出的图像能更逼真地逼近目标数据。但是单一图像域网络的复原难以达到满意的效果，而混合域网络的算法复杂度高，实用性不强。为此，本章将在前期工作的基础上，融合生成对抗网络和混合域网络的优势，构建基于双域 Transformer 生成对抗网路的稀疏角度 CT 成像算法（A Dual Domain Transformer Based Generative Adversarial Network for Sparse View CT Imaging，DuDoTrans-GAN），在投影域模块，设计专用于投影数据处理的 Transformer 网络，对投影数据缺失部分进行修补与复原。在图像域模块，由于在投影域模块已去除部分伪影，所以简化 SVT-GAN 的 Transformer 模块，从而减少网络计算量提高推断速度，同时图像域的生成对抗网络，可对处理后结果和目标图像进行判断，能不断优化投影数据处理网络及图像数据处理网络的性能。实验表明，融合后的 Transformer 网络可获得高质量的 CT 图像，在伪影去除、结构真实性和视觉观察方面均表现出较好的性能，可进一步提高稀疏角度 CT 图像复原效果，生成临床医师可接受的诊断图像。

5.1 双域 Transformer 生成对抗网络模型

稀疏角度成像是在常规全角度扫描下进行稀疏采样扫描，以获取更少的投影数据进行的 CT 图像重建。为此，本章希望通过双域处理的过程对投影域与图像域进行复原和伪影抑制，从而提高图像质量。本章所提 DuDoTrans-GAN 网络的主要思路为：首先将稀疏角度投影数据和全角度投影数据输入投影域 Transformer 子网络中，对稀疏角度投影数据进行复原，复原后的投影数据通过解析重建，获得重建图像，将重建图像与全角度 CT 图像输入图像域 Transformer 子网络中，得到最后处理后的图像，处理后的图像与全角度扫描重建的 CT 图像可送入判别器网络中，进行对抗学习。成像流程图如图 5-1 所示。DuDoTrans-GAN 方法中，投影域子网络借鉴了第三章的投影域处理 Transformer 网络结构，并对其进行改进，增加了跳跃连接以用于低级特征的保持。在损失函数上，使用局部内部结构损失函数，局部内部结构可借助正弦图的二阶导数稀疏性，以保持局部细节结构。在图像域子网络中，借鉴第四章的 Transformer 网络结构，将 4 级对称编码器简化成 3 级，以提高网络推断速度，同时还对 Transformer 结构进行优化改进，训

练时使用 MSE 函数作为图像域损失函数。

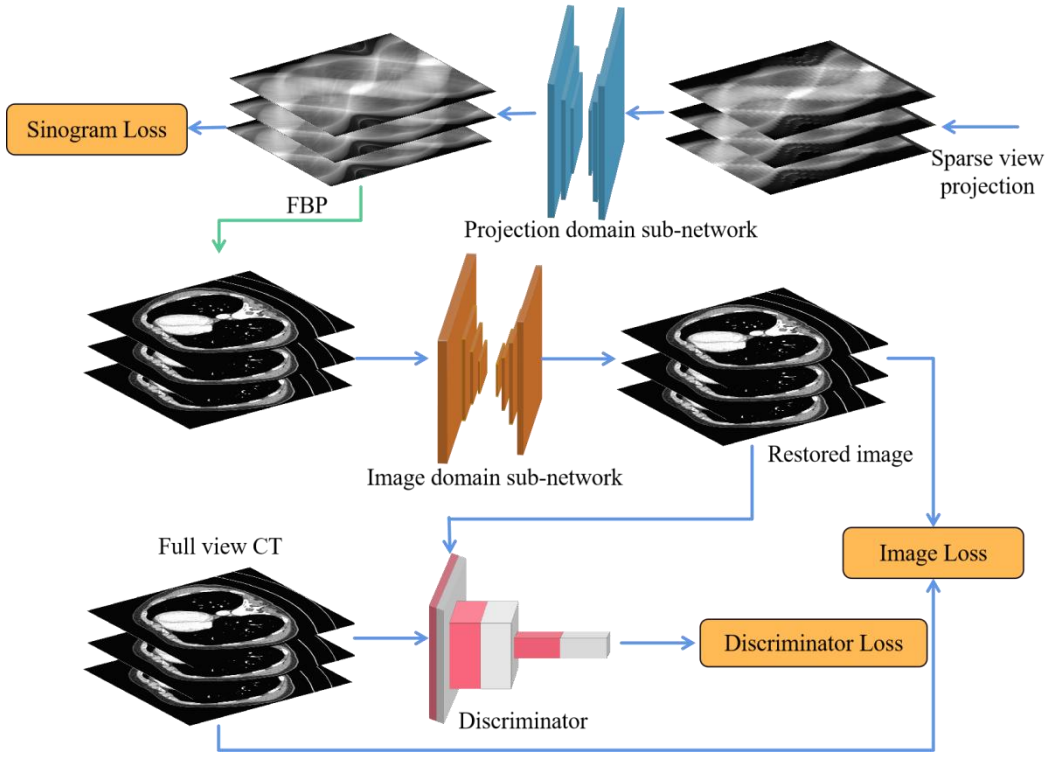


图 5-1 DuDoTrans-GAN 流程图

Fig. 5-1 The workflow of DuDoTrans-GAN

5.2 网络设计

5.2.1 投影域子网络

与常规的 CNN 网络相比，Transformer 网络中得益于其多头自注意力机制，能更充分的提取投影数据的特征。为更好的适应与投影数据特点，在投影域子网络采用改进型的 Transformer 模块。该 Transformer 模块由三个部分组成：头部，编码器和尾部。各部分的详细情况如下：

头部：由于标记化是性能的关键组成部分，因此将正弦图的每一行拆分为一组序列数据。然后使用 MLP 将正弦图嵌入到输入中：

$$H = \mathrm{Re}\,LU(LN(s_{ld}^{1},...,s_{ld}^{P}))_{\times n} \tag{5-1}$$

式中 LN 为线性层，将维数为 D 的正弦图转换为维数为 D_0 的嵌入。S_{ld} 为稀疏角度正弦图。$RELU(\bullet)$ 为激活函数。$H \in R^{H\times D}$ 是提取的特征作为 Transformer 的输入。MLP 由重复 n 次的块组成。

编码器：在这一部分中，沿用了第三章投影域 Transformer 的设计。在 Transformer 输入的每个投影 h_i 中加入可学习的位置编码 $e_i \in R^D$，以保持位置信

息。然后 $e_i + h_i$ 直接输入到 Transformer 编码器。Transformer 编码器由 MSA 和 MLP 组成。整体结构如下:

$$\begin{aligned} Z_0 &= \left[\mathrm{e}_1 + h_1, \ldots, e_p + h_p\right] \\ Q_i &= K_i = V_i = LN(Z_{i-1}) \\ Z_i^{'} &= MSA(Q, K, V) + Z_{i-1} \\ Z_{\mathrm{i}} &= MLP\left(Z_i^{'}\right) + Z_i^{'} \end{aligned} \tag{5-2}$$

其中 LN 是线性层。

尾部：在尾部，从头部添加了一个跳过连接，以维护低级功能。使用 MLP 将 Transformer 编码器的输出从 D_0 维恢复到 D 维：

$$S^{'} = \mathrm{Re}\, LU\left(LN\left(Z_l + H\right)\right)_{\times n} \tag{5-3}$$

经过这种细化后，得到了用于损失计算和下一阶段重建的投影图域输出。投影域网络设计如图 5-2 所示：

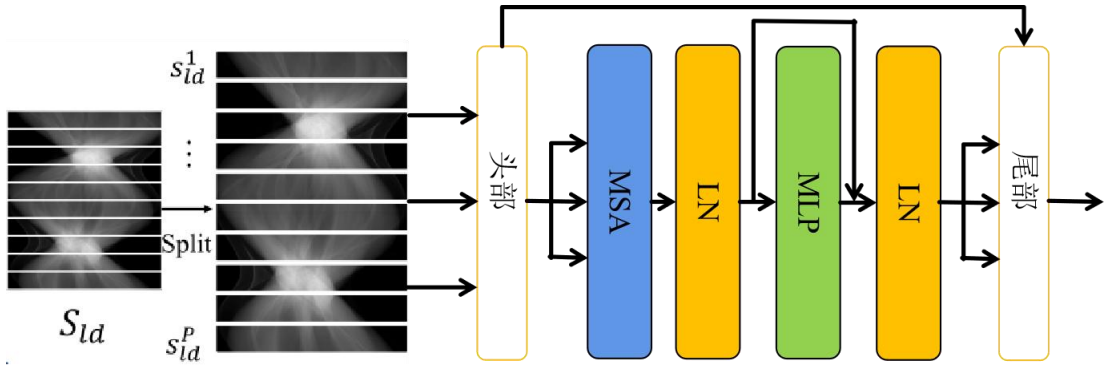


图 5-2 投影域子网络

Fig. 5-2 Projection domain sub-network

5.2.2 图像域子网络

在图像域模块中，沿用第四章的方法，给定稀疏角度 CT 图像 X，生成器 G 首先对其进行卷积，得到底层特征 F_0。然后，通过 3 级对称编码器将这些低级特征转换为深层特征 F_d。简化了第四章的生成器结构，为了不影响 CT 图像复原效果，对 Transformer 模块结构进行优化，减少线性连接层 LN 和前馈网络 FFN，MDTA 能够聚合局部和非局部像素交互，并且足够有效地处理稀疏角度 CT 图像，所以保留 MDTA 模块来代替常规 Transformer 中的多头注意力模块。GDFN 能够执行受控特征转换，即抑制信息量较小的特征，只允许有用的信息通过网络层次进一步传递，所以保留 GDFN 来代替常规 Transformer 中的线性连接层和前馈网络。在第四章消融实验中已经验证减少线性连接层和前馈网络会降低图像重建质量，但由于增加了投影域模块，所以输入图像质量提高，图像总体质量不会降低。在这项工作中，图像域模块结构图如图 5-3 所示：

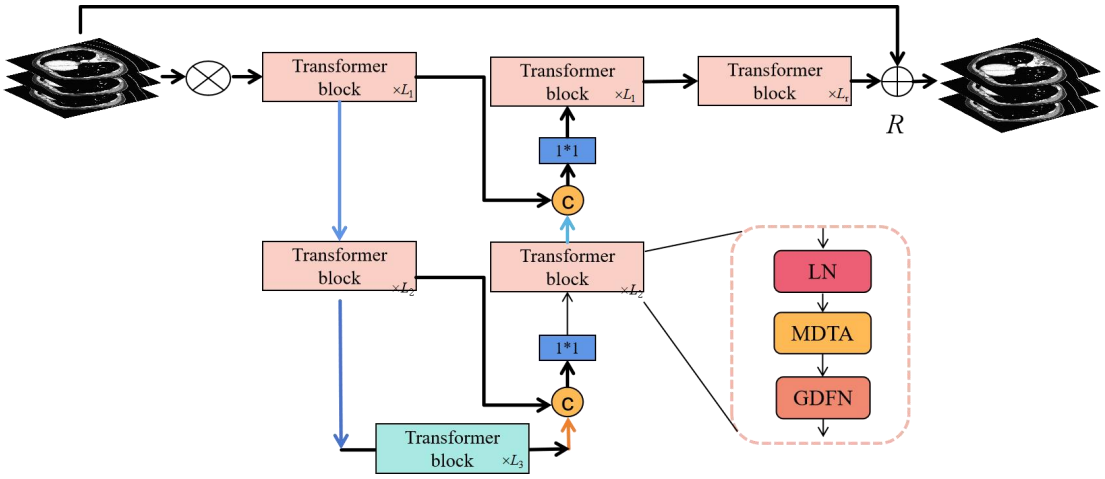


图 5-3 图像域子网络

Fig. 5-3 Image domain sub-network

判别器模块：本章沿用第四章的判别器模块，由于在生成器模块已经对于网络模型进行相对应的简化，为此在判别器模块决定不降低判别器功能，防止重建 CT 图像质量降低。在这一模块中采用平铺结构的判别器 D 来进一步提高生成器 G 的预测精度。本章设计的判别器 D 结构与第四章的相同，网络输入数据大小为 3×512×512。通过层间冗余性质，可以在不影响局部相关性的情况下增强判别器的表示能力。除了第一层，所有卷积层的卷积核大小都设置为 3×3。卷积处理可以从图像中提取不同层次的特征。最后，特征图通过全连接操作来区分输入图像的真实性。

5.2.3 损失函数

投影域子网络损失函数：投影图可以被描述为一个分段线性平滑的结构，其二阶导数具有一定的稀疏性。为此，在网络学习中可利用这一特点，构造损失函数。一种可行的方法是使用 Hessian 惩罚来引入二阶导数的稀疏性，但二阶导数稀疏化，容易导致图像过于平滑。为此本章使用稀疏角度投影数据与标签投影数据（即全角度扫描的投影数据）之间的误差作为主要损失，通过超参数控制二阶导数损失的引入。

$$
\begin{aligned}
L_S &= \sum_{i,j}\sqrt{D_{ii}^2 + D_{jj}^2 + D_{ij}^2 + D_{ji}^2} \\
D_{ii} &= \frac{\partial^2 \hat{S}(i,j)}{\partial i^2} - \frac{\partial^2 S_{nd}(i,j)}{\partial i^2} \\
D_{jj} &= \frac{\partial^2 \hat{S}(i,j)}{\partial j^2} - \frac{\partial^2 S_{nd}(i,j)}{\partial j^2} \\
D_{ij} &= \frac{\partial^2 \hat{S}(i,j)}{\partial i \partial j} - \frac{\partial^2 S_{nd}(i,j)}{\partial i \partial j} \\
D_{ji} &= \frac{\partial^2 \hat{S}(i,j)}{\partial j \partial i} - \frac{\partial^2 S_{nd}(i,j)}{\partial j \partial i}
\end{aligned}
\tag{5-4}
$$

其中，$D_{ii}, D_{jj}, D_{ij}, D_{ji}$ 是输出 $\hat{S}$ 和真值 S_{nd} 之间的二阶导数的差。

图像域子网络：使用 MSE 函数作为图像域模块损失函数，如式（5-5）所示：

$$
\min_{\theta} L = \left\| I_S - I_F \right\|_2^2 \tag{5-5}
$$

其中，I_S 为输出图像，I_F 为标签，θ 为网络参数。

判别器：判别器 D 的损失函数定义如下：

$$
\min_{G}\min_{D} L_{GAN}\left(X, Y, \Theta_G, \Theta_D\right) + \beta \bullet L_G\left(X, Y, \Theta_G\right) \tag{5-6}
$$

其中 X_n 和 Y_n 分别表示稀疏角度的 CT 图像和全角度 CT 图像。$G(\bullet)$表示生成器 G 的预测，Θ_G 为生成器 G 的参数，Θ_D 表示判别器 D 的参数，β 是用来平衡不同损失值的权重参数。对于网络训练，本章采用默认参数的 Adam 方法对损失函数进行优化。

5.3 实验结果与分析

5.3.1 实验设计

为了验证本章所设计 HDTranformer 网络在稀疏扫描条件下成像的有效性，实验选用 FBP、PRestor、StarGAN 和 CTTR 四种方法对比，其中 FBP 和 PRestor 是非深度学习的传统方法，StarGAN 方法是一种多域图像到图像转换的统一生成对抗网络，随后应用于医学图像处理中，并取得了优异的性能，CTTR 方法为一种基于 Transformer 网络的双域稀疏角度 CT 重建算法，具有优异的稀疏角度图像处理性能，与本章方法类似。实验中采用 Mayo Clinic 研究人员制作的 AAPM 数据集进行训练和验证。该数据集包含 10 例患者的 4825 张 1mm 厚的高质量 CT 图像。通过 FBP 算法从 720 个角度的投影数据中重建出高质量的图像，可以认

为是全角度 CT 图像(标签)。在第一阶段，将 CT 图像进行投影，将获得投影数据作为样本，将全角度 CT 图像的投影数据作为标签。在第二阶段，将经过第一阶段修复的投影数据进行重建，将重建后的 CT 图像作为样本，将对应的全角度 CT 图像作为标签。重建均采样 FBP 算法。重建 CT 图像大小为 512×512，全角度投影数据大小为 720×768，720 个投影角度。在投影域模块，为减少单次数据量的使用，BatchSize 大小为 3，网络通道数为 96，训练 epoch 设置为 80 次，学习率的初始值为 2×10^{-4}，学习率递减到最小值为 1×10^{-6}。在图像域模块，将 batchsize size 设为 64，初始通道数为 96，训练周期 epoch 设置为 80，学习率的初始值为 2×10^{-4}，学习率递减到最小值为 1×10^{-6}。

5.3.2 实验结果

图 5-4，图 5-5 和图 5-6 分别 90 个角度，120 个角度和 180 个角度下的重建结果，其中第一行、第二行和第三行表示使用不同算法对所选胸部、腹部和胯部切片的成像结果，显示窗为[-150HU，250HU]。图中第一列为全角度下 FBP 重建图像，作为不同方法成像结果的对比。通过对比图 5-4、图 5-5 和图 5-6 中第二列成像结果，可以看到，由于投影数据的缺失，传统解析重建算法获得图像含有大量的伪影，图像细节被伪影淹没，严重影响诊断。同时也发现，随着扫描角度的减少，重建图中伪影越来越多。通过对比第三列的成像结果，可以看到经过双线性插值后，缺失的投影数据被初略的补全，重建图像有了明显的改善，部分细节清晰，但依旧含有大量的条状伪影和部分噪声。通过对比 StarGAN、CTTR 和 DuDoTrans-GAN 方法成像结果，可以发现图像质量有了大幅度的提升。进一步通过局部放大图可以看出，StarGAN 方法获得图像仍存在伪影的残留，CTTR 算法结果会丢失一部分图像信息,细节对比度有所下降,出现一定的过平滑现象，而本章 DuDoTrans-GAN 方法获得图像细节更多，对比度更高（如图 5-4、图 5-5 和图 5-6 中第六列图中红色箭头所示），更接近参考图，整体图像质量要优于其他三种对比方法。

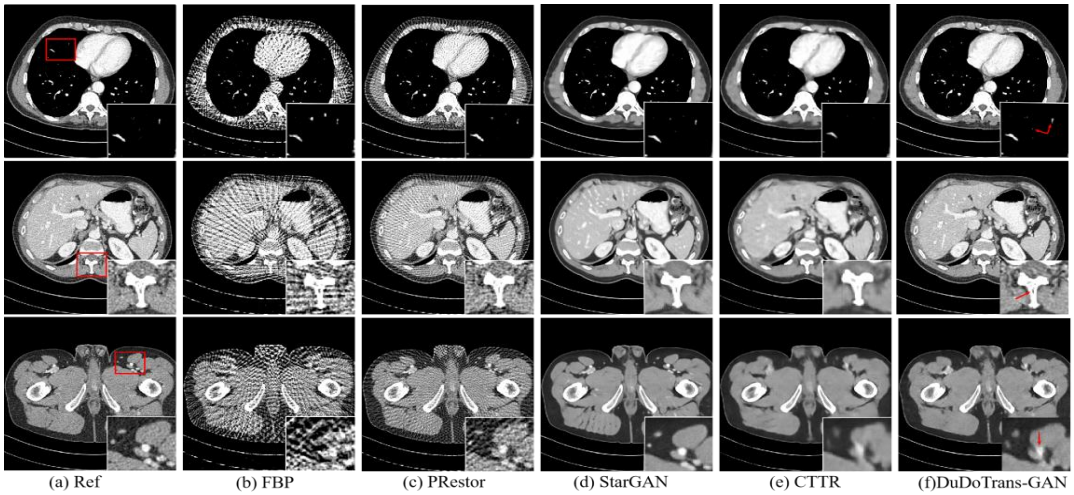


图 5-4 90 个扫描角度下不同方法的 CT 成像效果图

Fig. 5-4 Axial view results of different methods under 90 views

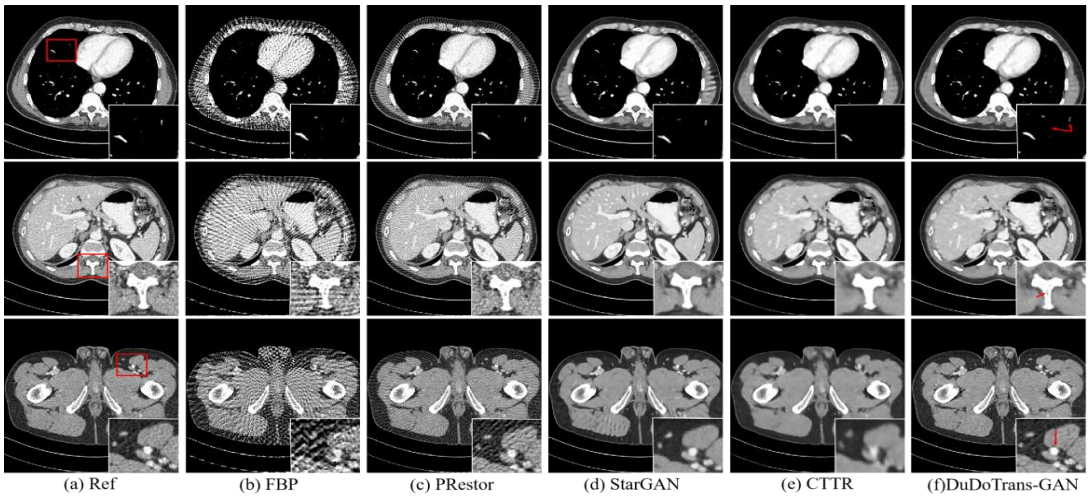


图 5-5 120 个扫描角度下不同方法的 CT 成像效果图

Fig. 5-5 Axial view results of different methods under 120 views

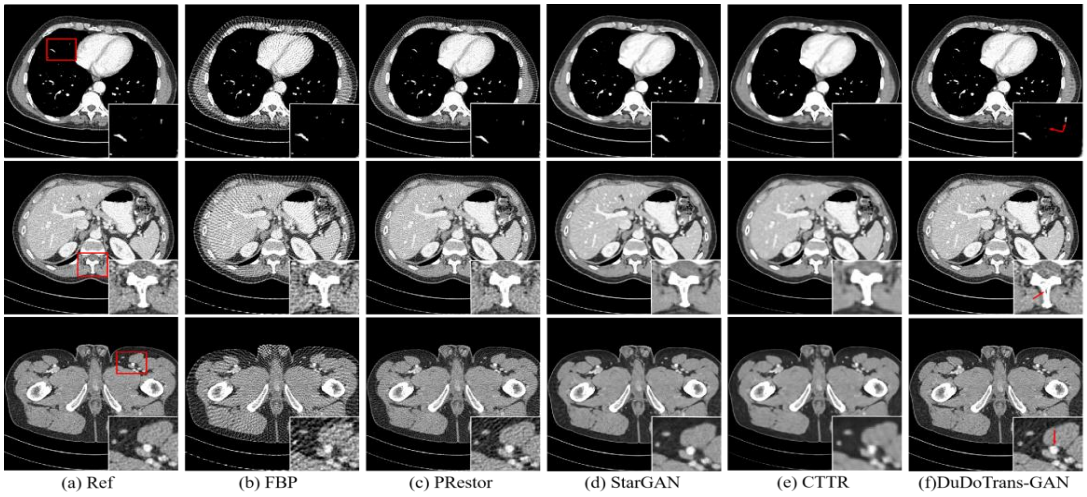


图 5-6 180 个扫描角度下不同方法的 CT 成像效果图

Fig. 5-6 Axial view results of different methods under 180 views

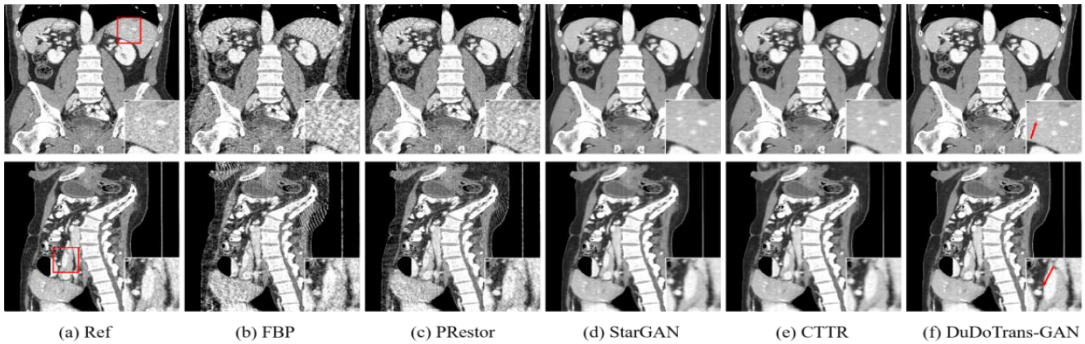


图 5-7 90 个扫描角度下矢状面和冠状面的成像效果图

Fig. 5-7 Coronal and sagittal images of the different methods under 90 views.

图 5-7 为不同方法在冠状面和矢状面中的成像效果，实验选取 90 个投影角度下冠状面和矢状面重建图进行比较，显示窗为[-150HU，250HU]。从图中可以看出，冠状面和矢状面中噪声伪影相对较少，扫描角度减少对这两个切面影响没有横断面大。CTTR 方法在噪声伪影上去除得非常干净，但图像组织纹理保持度不高，部分细节丢失，而本章 DuDoTrans-GAN 方法能显著降低各部位噪声和伪影，同时也更好地保留图像组织区域的边缘特征，处理后的图像具有较好的视觉效果。

图 5-8 分别展示了 90 个角度，120 个角度和 180 个角度扫描下成像结果与对应参考图之间的差异，显示窗口为[0HU，200HU]。通过差异图的对比，可以发现三种对比方法处理后的 CT 图像残留伪影较多，与参考图像之间差异较大，而本章所提方法的差值图强度较低，整体较均匀，这表明本章方法获得的图像更加接近于参考图，显著优于其他处理方法。综上，从视觉效果来看，本章方法的处理效果要优于对比方法。

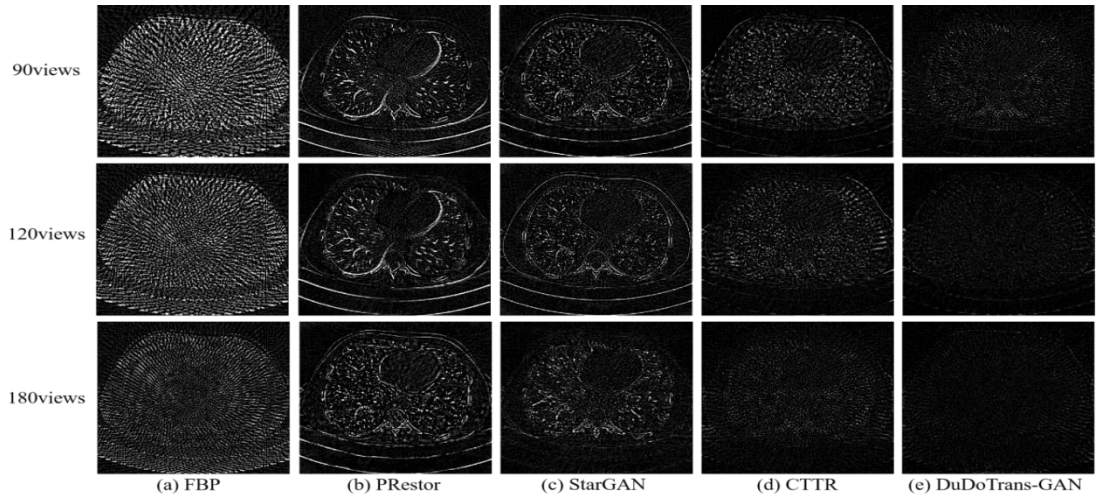


图 5-8 不同方法获得的 CT 图像的差异图

Fig. 5-8 Difference images relative to the reference image obtained by different methods

为更加直观及定量地比较不同算法的处理效果，本章采用 PSNR、SSIM 和

FSIM 三个指标来评估实验结果。表 5-1 为三种不同扫描角度下重建图像的 PSNR、SSIM 和 FSIM 值。通过表 5-1 可以看出，本章 DuDoTrans-GAN 方法在 PSNR 和 FSIM 两个指标下，要优于其他三种对比方法，具有一定量化优势。为了评估像素级的 HU 拟合效果，图 5-9 为三个不同感兴趣区域绘制的 CT 值轮廓曲线图。对于 90 角度和 120 角度数据，DuDoTrans-GAN 的拟合性能优于 CTTR，表明该方法具有更好的结构保存性能。

表 5-1 不同方法获得的 PSNR、SSIM 和 FSIM 值

Table 5-1. PSNR, SSIM, and FSIM values obtained by different methods

Views	90 views			120 views			180 views		
Metrics	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
FBP	28.33±	0.73±	0.72±	32.12±	0.76±	0.75±	34.24±	0.78±	0.77±
	0.77	0.031	0.032	0.68	0.029	0.029	0.65	0.026	0.027
PRestor	33.53±	0.77±	0.75±	34.52±	0.82±	0.81±	37.35±	0.85±	0.84±
	0.63	0.022	0.032	0.65	0.020	0.028	0.61	0.018	0.025
StarGAN	36.36±	0.81±	0.78±	37.34±	0.85±	0.82±	39.65±	0.88±	0.85±
	0.52	0.014	0.023	0.59	0.012	0.019	0.55	0.011	0.016
CTTR	38.40±	0.83±	0.79±	40.65±	0.86±	0.83±	41.87±	0.91±	0.88±
	0.67	0.018	0.023	0.62	0.016	0.018	0.59	0.013	0.016
DUDOTra	43.01±	0.94±	0.93±	44.21±	0.96±	0.94±	46.92±	0.96±	0.97±
ns-GAN	0.43	0.012	0.016	0.28	0.008	0.015	0.26	0.07	0.011

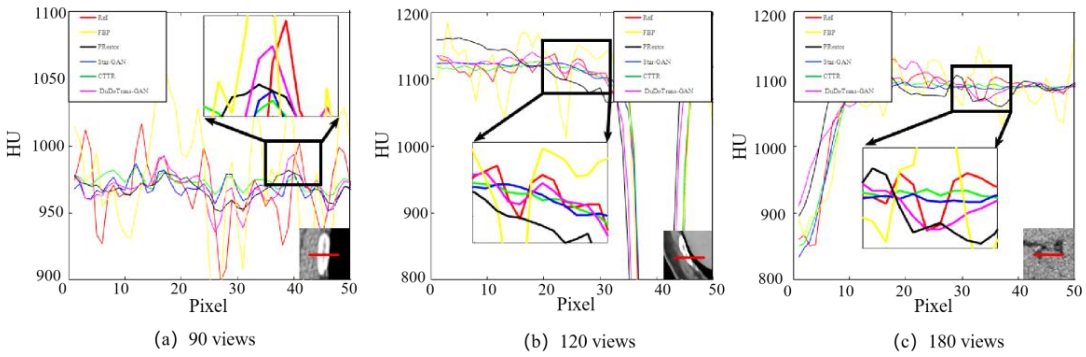


图 5-9 不同方法的三个轴向切片沿指定红线的强度分布图

Fig. 5-9 The intensity profiles along the specified red line in the three axial slices of different methods

本章融合了第三章和第四章的 Transforemr 算法，为此选取 120 个角度下的 CT 重建结果进行比较，其中第一行、第二行和第三行表示使用不同算法对所选胸部、腹部和胯部切片的成像结果，显示窗为[-150HU，250HU]。图中第一列为全角度下 FBP 重建图像，作为不同方法成像结果的对比，如图 5-10 所示。从图中可以看出，三种算法中 HDTransformer 重建效果最差，在跨部图像细节模糊。SVT-GAN 和 DuDoTrans-GAN 在胸部和跨部重建结果细节相似，但在腹部位置本章算法要优于 SVT-GAN。表 5-2 为 90 扫描角度下和 180 扫描角度下重建图像

的 PSNR、SSIM 和 FSIM 值。通过表 5-2 可以看出，本章 DuDoTrans-GAN 方法要优于其他三种对比方法，具有一定量化优势。

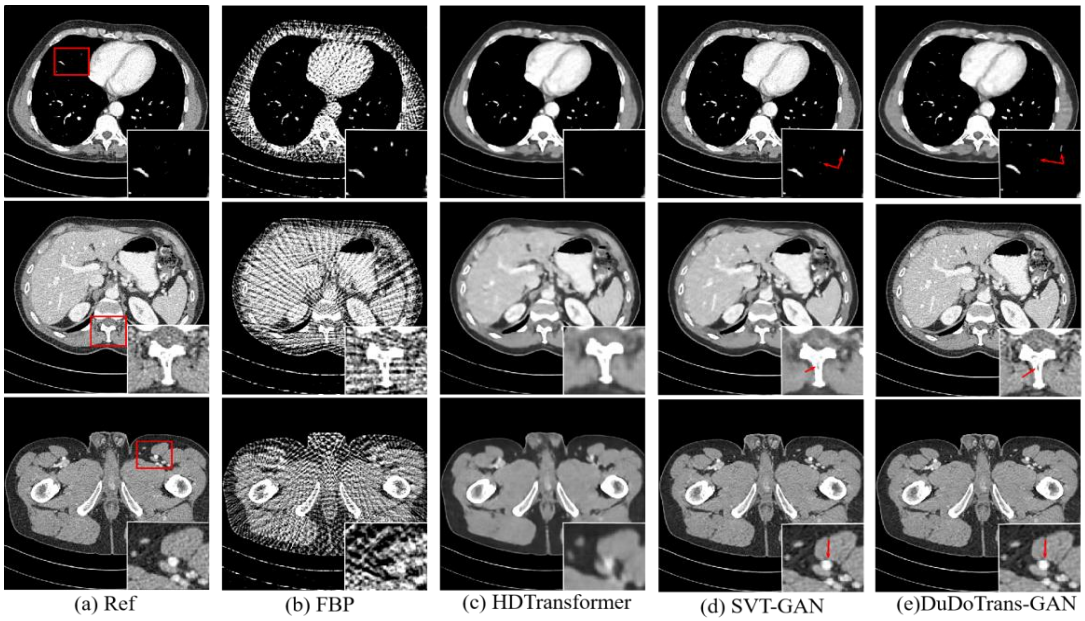


图 5-10 120 个扫描角度下不同方法的 CT 成像效果图

Fig. 5-10 Imaging results of different methods under 120 views

表 5-2 不同方法获得的 PSNR、SSIM 和 FSIM 值

Table 5-2. PSNR, SSIM, and FSIM values obtained by different methods

Views	90 views			180 views		
Metrics	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
FBP	28.33 ±0.77	0.73 ±0.031	0.72 ±0.032	34.24 ±0.65	0.78 ±0.026	0.77 ±0.027
HDTransformer	40.42 ±0.52	0.91 ±0.021	0.89 ±0.022	43.14 ±0.35	0.94 ±0.016	0.93 ±0.018
SVT-GAN	42.47 ±0.49	0.94 ±0.013	0.91 ±0.018	46.34 ±0.29	0.96 ±0.010	0.95 ±0.013
DuDoTrans-GAN	43.01 ±0.43	0.94 ±0.012	0.93 ±0.016	46.92 ±0.26	0.96 ±0.07	0.97 ±0.011

5.3.3 网络模型分析

（a）消融实验

本章 DuDoTrans-GAN 成像方法由三个部分组成，分别是投影域子网络、图像域子网络和对抗网络，子网络对稀疏角度 CT 成像结果均有一定程度的促进作用，其中投影域子网络由第三章投影域 Transformer 网络优化得到，图像域模块由第四章 SVT-GAN 简化而来，为了验证本章的融合网络对稀疏角度 CT 成像效果的提升作用，此部分分别将不同子网络及组成网络的成像效果进行了比较，将

投影域子网络设为S_1，FBP 重建图仅通过对抗网络处理，记为S_2。FBP 重建图仅通过图像域子网络处理，记为S_3。实验结果选取胸部图像和腹部图像进行比较，如图 5-11，图 5-12 所示，其中图 5-11 为 90 个稀疏角度扫描后的 CT 图像，图 5-12 为 120 个稀疏角度扫描后的 CT 图像，显示窗口均为[-150HU，250HU]。通过对比图 5-11、图 5-12 中胸部的局部放大图可以发现，S_2的图像细节要多于S_1，本章算法比S_2，S_3具有更好的成像效果。通过对比图 5-11 和图 5-12 中腹部局部放大图可以看出，在处理伪影和噪声方面，本章所采用的融合网络同样处理效果最好，更加接近参考 CT 图像。为更加直观及定量地比较不同阶段方法的对成像效果的促进情况，此部分将进一步采用 PSNR 和 FSIM 这两种客观评价指标进行量化，如图表 5-3 所示。通过表 5-3 可以看出，本章提出的融合算法 PSNR 和 FSIM 最高，表明其重建图像更加接近于全角度下的参考图像。

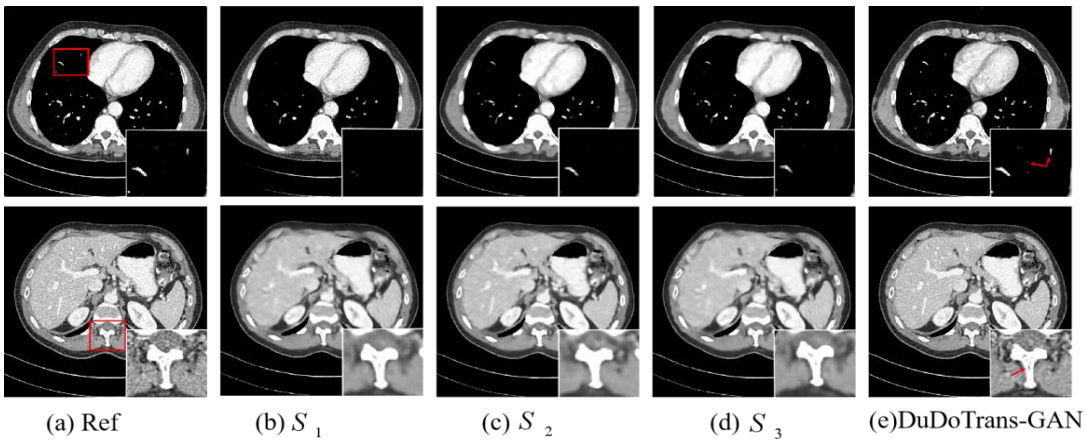


图 5-11 90 个扫描角度下胸部和腹部 CT 成像效果

Fig. 5-11 Chest and abdomen imaging results under 90 views

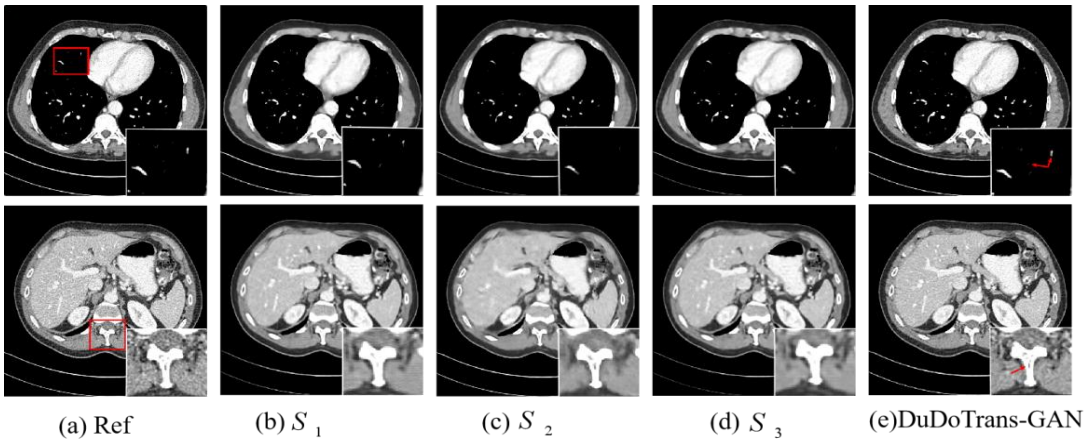


图 5-12 120 个扫描角度下胸部和腹部 CT 成像效果

Fig. 5-12 Chest and abdomen imaging results under 120 views

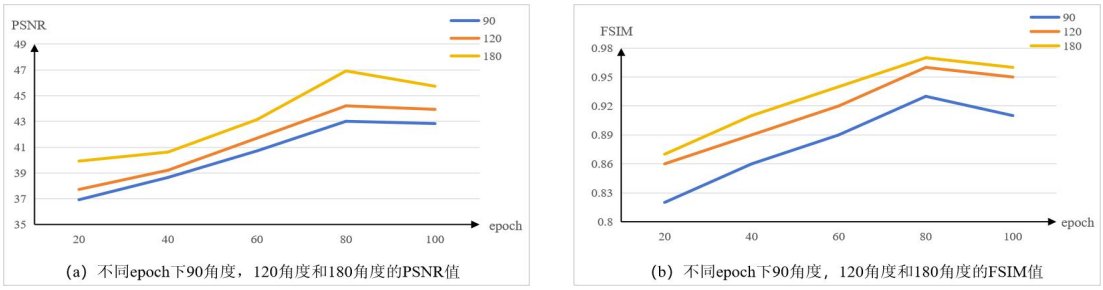
表 5-3 消融研究的定量结果

Table 5-3. Quantitative results of the ablation study.

Views	90 views			120 views		
Metrics	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
S_1	33.25 ±0.72	0.75 ±0.039	0.72 ±0.042	35.21 ±0.65	0.81 ±0.034	0.79 ±0.036
S_2	35.82 ±0.067	0.78 ±0.032	0.74 ±0.038	38.65 ±0.52	0.85 ±0.027	0.82 ±0.029
S_3	40.47 ±0.59	0.86 ±0.023	0.82 ±0.028	41.29 ±0.41	41.92 ±0.020	0.89 ±0.026
DuDoTrans-GAN	43.01 ±0.43	0.94 ±0.012	0.93 ±0.016	44.21 ±0.28	0.96 ±0.008	0.94 ±0.015

（b）收敛性验证实验

此部分实验进一步分析了 DuDoTrans-GAN 网络训练的收敛性。实验采用不同训练 epoch 下重建 CT 图像的 PSNR 均值和 FSIM 均值作为指标，绘制变化曲线，间接分析网络的收敛性和稳定性。PSNR 和 FSIM 变化曲线如图 5-13 所示。从图 5-13（a）中可以看出，随着训练周期的增加，不同扫描角度下重建图像的 PSNR 值均逐步增高至 80 个 epoch 后降低。从图 5-13（b）中可以看到，不同扫描角度测试数据重建图像的 FSIM 变化趋势与 PSNR 一致，在 80 个 epoch 后降低，表明网络开始出现过拟合。为此，综合视觉效果、PSNR 及 FSIM 的量化值，实验中选取训练周期为 80 个 epoch。



(a) 不同epoch下90角度，120角度和180角度的PSNR值　(b) 不同epoch下90角度，120角度和180角度的FSIM值

图 5-13 不同 epoch 下 90 角度，120 角度和 180 角度的 PSNR 值和 FSIM 值

Fig. 5-13 PSNR and FSIM values under 90, 120, and 180 views under different epochs

（c）模型复杂性

如表 5-4 所示，实验以 300 张 CT 图像为测试数据，在 120 个角度的基础上，比较了不同方法的参数数量、FLOP 和运行时间。CTTR 和本文提出的 DuDoTrans-GAN 均采用 Transformer 网络结构。CTTR 的模型尺寸和计算复杂度分别为 61M 和 187G FLOPs，HFormer 的模型尺寸和计算复杂度分别为 1.6M 和 104G FLOPs，而 DuDoTrans-GAN 的参数和时间复杂度分别为 28M 和 82G FLOPs，其计算复杂度更少。SVT-GAN 模型的测试时间和训练时间都比对比方法模型短。

表 5-4 不同方法的模型大小和计算复杂度

Table 5-4. The model size and computational complexity of different methods

	Training time (min)	Test time (min)	Params (M)	FLOPs (G)
StarGAN	401	3.8	53.2	192
CTTR	326	3.4	61	187
HFormer	251	2.8	1.65	104
HDTransformer	276	5	30	99
SVT-GAN	64	1.2	26	98
DuDoTrans-GAN	60	2.4	28	82

5.4 本章小结

CT 成像中，由于扫描角度的缺失，会导致重建图像中出现大量条形伪影和斑点噪声，干扰临床诊断。从图像处理的角度上来抑制伪影，具有较高的应用价值。但是单一图像域的复原难以达到满意的效果，而第三章所提的三阶段混合域网络的算法复杂度高，实用性不强。为此，本章在第三章和第四章的方法基础上融合，构建一种新的基于双域 Transformer 生成对抗网络模型，并且在投影域子网络和图像域子网络设计中，做了针对性的改进和优化。与此同时，还针对投影域子网络和图像域子网络中处理对象及特点的不同，设计不同应用于不同子网络损失函数。实验结果表明，本章所提网络模型可适用于不同稀疏程度的 CT 数据，与对比算法相比，图像质量均有明显的提升，并在 PSNR，SSIM 和 FSIM 等不同量化指标上取得了一定的提升。消融实验结果验证了网络各部分的设计在伪影抑制中的有效性，模型在不同扫描角度的成像任务重中均表现出一定的收敛性，模型容量及计算效率也优于对比方法。

第 6 章 总结与展望

6.1 论文总结

在 CT 成像中，由于扫描角度的缺失，导致重建后的 CT 图像中存在大量条形伪影和斑点噪声，现有的深度学习网络难以有效的去除图像中的伪影，尤其是图像中的高强度伪影，恢复图像组织边缘和细节。为此本章概括了研究的核心贡献，旨在优化 CT 成像在稀疏角度采样情况下的性能，以生成更适用于临床实践的图像，同时缩减辐射影响，减少潜在的遗传危害和肿瘤发生风险。这些成像模型的创新设计具有显著的理论价值和实际应用价值。以下详述了论文的主要研究内容：

（1）由于扫描角度的缺失，导致重建后的 CT 图像中存在大量条形伪影和斑点噪声。为解决单一图像域的复原难以达到满意的效果，论文第三章采用联合投影域及图像域处理策略，建立基于 Transformer 网络的混合域成像流，以提高稀疏角度 CT 图像重建质量。此外，针对不同阶段数据噪声伪影特点设计不同的 Transformer 块，以实现差异化的处理。更进一步，算法采用可微分的解析重建和投影运算，建立投影域与图像域数据的转换，实现了端到端的稀疏角度 CT 优质成像流。

（2）为降低算法复杂性，希望通过图像后处理的手段来抑制噪声和伪影，并提高稀疏角度 CT 图像的质量，论文第四章提出了一种基于 Transformer 的生成对抗网络 SVT-GAN。网络中生成器采用改进的 Transformer 模块，以充分提取局部和非局部特征，并促进结构和细节的恢复，同时借助于对抗学习进一步促进生成网络性能的提升。在训练过程中，专门设计了一个混合损失函数来提高网络性能。两种不同数据集的实验验证了该方法的有效性，并通过消融和参数分析实验，总结了其优缺点。

（3）考虑到单一图像域网络的复原难以达到满意的效果，而混合域网络的算法复杂度高，实用性不强。论文在第三章和第四章的工作基础上，进一步融合生成对抗网络和混合域网络，提出 DuDoTrans-GAN 网络学习框架，在投影域子网络中，设计专用于投影数据处理的 Transformer 网络，对投影数据缺失部分进行修补与复原。在图像域子网络块，简化第四章的 Transformer 模块，以减少网络计算量，提高计算速率。同时在训练中增加图像生成的判别网络，使其对处理后结果和目标图像进行判断，不断优化处理网络的性能。实验数据验证了网络各部分的有效性，具有较高的应用前景。

6.2 研究展望

通过对现有稀疏角度 CT 图像重建技术的广泛调查，本研究剖析了各种算法的强项与弱点，并提出了一种创新的基于 Transformer 架构的 CT 图像重建方法，

构建了三种独特的 Transformer 驱动的稀疏角度 CT 成像网络模型。实验模拟结果证实了提出的网络模型在处理稀疏角度 CT 图像时的有效性。尽管这些方法在处理稀疏 CT 图像时表现出色，但依然存在一些待解决的问题：

（1）Transformer 网络模型运算复杂度高，未来研究需探索在保证 CT 图像重建质量的同时，如何降低计算需求，提升实用性。同时，当前的 PSNR、SSIM 和 FSIM 等图像质量评估指标在诊断应用中的局限性也值得关注，需要寻找更全面的评估手段来评价处理后的 CT 图像。

（2）为了增强研究的实践意义，论文应更深入地结合临床实践，包括构建适应临床的模型、开发处理算法、设计评价和优化策略。由于临床数据的不足，部分处理效果的验证仍有欠缺。后续工作可侧重于以实际临床任务为导向的图像处理研究。

为了改善 Transformer 网络模型计算量大的问题，可以像文中针对 CT 图像不同模块设计出不同的 Transformer 网络模型来尽可能降低计算量，未来进一步希望通过对 Transformer 学习理论的深入研究，设计专门适用于 CT 投影数据和图像数据处理的 Transformer 模块，通过对不同域数据的处理及相互促进，来提升最终的成像效果。以此延申，Transformer 模块在具体使用和，实施过程中可借鉴效果较好的模型并改良使用，以获得整体更好的成像效果。其应用也可进一步扩展，如低剂量 CT，能谱 CT、锥束 CT、MRI 成像，PET 成像等。此外，针对成像效果的评价标准，论文中使用了常用的 PSNR，SSIM 和 FSIM 来综合评价图像质量好坏，后续可进一步建立更加完善的评价体系对处理后的图像进行评估，如邀请临床医生作为观察者，对运条纹伪影抑制、解剖结构保真等多方面进行盲评。临床应用评估也是重要的指标，量化比较稀疏角度 CT 图像质量改进所带来的临床应用精度的提升，同时还需进一步探究高质量稀疏角度 CT 图像在诊断中的临床价值。未来还需继续研究并设计更优异的深度学习网络模型，以进一步减少扫描角度，实现在极稀疏角度的情况下依旧可以有很好的复原效果，以进一步降低辐射剂量，减少扫描者辐射伤害。

参考文献

[1] 庄天戈. CT 原理与算法[M]. 上海: 上海交通大学出版社, 2-3, 1992.

[2] 林泽田,王单. 稀疏角度 CT 图像重建算法研究[J].计算机应用与软件, 35(10): 217-222+311, 2018.

[3] Zeeshan A., Aun I., Muazzam M.. An efficient U-Net framework for lung nodule detection using densely connected dilated convolutions[J]. The Journal of Supercomputing, 36(5): 151-164, 2021.

[4] Brennern D., Hall E.. Computed tomography: an increasing source of radiation exposure[J]. The New England Journal of Medicine, 357(22): 2277-2284, 2007.

[5] 陈海, 刘晓东, 董洁等. 基于缺陷投影的稀疏角度数据 CT 图像迭代重建仿真研究[J]. 医疗卫生装备, 38(04): 29-31+37, 2017.

[6] Xie J., Xu L., Chen E.. Image denoising and inpainting with deep neural networks [J]. Advances in Neural Information Processing Systems, 18(1): 107-114, 2012.

[7] Aswani A., Shazeer N., Parmar N., et al. Attention is all you need[J]. Advances in neural information processing systems 30, 15(3): 561-574, 2017.

[8] Li S., Cao Q., ChenY., et al. Dictionary learning based sinogram inpainting for CT sparse reconstruction[J]. Optik-International Journal for Light and Electron Optics, 125(12): 2862-2867, 2014.

[9] Shen L., Xu Y. , and Zhang N.. An approximate sparsity model for inpainting[J]. Applied and Computational Harmonic Analysis, 37(1): 171-184, 2014.

[10] Zhang H., Zhang L., Sun Y., et al. Projection domain denoising method based on dictionary learning for low-dose CT image reconstruction[J]. Journal of X-ray Science and Technology, 23(5): 567-578, 2015.

[11] Karimi D. and Ward R. K.. Sinogram denoising via simultaneous sparse representation in learned dictionaries[J]. Physics in Medicine and Biology, 61(9): 3536, 2016.

[12] Chen Y., Yin X., Shi L., et al. Coatrieux, and C. Toumoulin, improving abdomen tumor low-dose CT images using a fast dictionary learning based processing[J]. Physics in Medicine and Biology, 58(16): 5803, 2013.

[13] Chen Y., Shi L., Feng Q., et al. Artifact suppressed dictionary learning for low-dose CT image processing[J]. IEEE Transactions on Medical Imaging, 33(12): 2271-2292, 2014.

[14] Ha S. and Mueller K. . Low dose CT image restoration using a database of image patches[J]. Physics in Medicine and Biology, 60(2): 869, 2015.

[15] Green M., Marom E. M., Kiryati N., et al. Efficient low-dose CT denoising by locally-consistent non-local means (LC-NLM)[C]. In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 30(2): 769-781, 2016.
[16] Cui X., Gui Z., Zhang Q., et al. Learning-based artifact removal via image decomposition for low-dose CT image processing[J]. IEEE Transactions on Nuclear Science, 63(3): 1860-1873, 2016.
[17] Karimi D. and Ward R., Reducing streak artifacts in computed tomography via sparse representation in coupled dictionaries[J]. Medical Physics, 43(3): 1473-1486, 2016.
[18] Sidky E. Y. and Pan X., Image reconstruction in circular cone-beam computed tomography by constrained total-variation minimization[J]. Physics in Medicine and Biology, 53(17): 4762-4777, 2008.
[19] Xu Q., Yu H., Mou X., et al. Low-dose X-ray CT reconstruction via dictionary learning[J]. IEEE Transactions on Medical Imaging, 31(9): 1682-1697, 2012.
[20] Lu Y., Zhao J., and Wang G., Few-view image reconstruction with dual dictionaries[J]. Physics in Medicine and Biology, 57(1): 165-173, 2011.
[21] Liu J., Hu Y., Yang J., et al. 3D feature constrained reconstruction for low dose CT imaging[J]. IEEE Transactions on Circuits and Systems for Video Technology, 28(5): 1232-1247, 2018.
[22] Zheng X., Ravishankar S., LongY., et al. PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction[J]. IEEE Transactions on Medical Imaging, 37(6): 1498-1510, 2018.
[23] Bao P., Xia W., Yang K., et al. Convolutional sparse coding for compressed sensing CT reconstruction[J]. IEEE Transactions on Medical Imaging, 38(11): 2607-2619, 2019.
[24] Isola P., Zhu J., and Zhou T., Image-to-image translation with conditional adversarial networks[C], In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 33(6): 5967-5976, 2017.
[25] Li Z., Cai A., Wang L., et al, Promising generative adversarial network based sinogram inpainting method for ultra-limited-angle computed tomography imaging[J], Sensors, 19(18): 3941-3941, 2019.
[26] Meng M., Li S., Yao L., Li D., et al, Semi-supervised learned sinogram restoration network for low-dose CT image reconstruction[C], In Proceedings of SIPE Conference on Physics of Medical Imaging, 11312: 113120B, 2020.

[27]Zainulina E., Chernyavskiy A., and Dylov D. V., No-Reference Denoising Of Low-Dose CT Projections[C], In Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI), 77-81, 2021.
[28]Choi K., Self-supervised Projection Denoising for Low-Dose Cone-Beam CT[C], In Proceedings of IEEE Engineering in Medicine & Biology Society (EMBC), 3459-3462, 2021.
[29] Wu Q., Feng R., Wei H., et al. Self-supervised coordinate projection network for sparse-view computed tomography[J]. IEEE Transactions on Computational Imaging, 28(12): 321-338, 2023.
[30] Wang T., Xia W., Feng J., et al, A Review of Deep Learning CT Reconstruction From Incomplete Projection Data[J], IEEE Transactions on Radiation and Plasma Medical Sciences, 8(2): 138-152, 2024.
[31]Yang W., Zhang H., Yang J., et al, Improving low-dose CT image using residual convolutional network[J], IEEE Access, (5): 24698-24705, 2017.
[32]Kang E., Min J., and Ye J. C., A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction[J], Medical Physics, 44(10): 360-375, 2017.
[33]Chen H., Zhang Y., Kalra M. K., et al, Low-dose CT with a residual encoder-decoder convolutional neural network[J], IEEE Transactions on Medical Imaging, 36(12): 2524-2535, 2017.
[34]Yang Q., Yan P., Zhang Y., et al, Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss[J], IEEE Transactions on Medical Imaging, 37(6): 1348-1357, 2018.
[35]Shan H., Zhang Y., Yang Q., et al, 3D convolutional encoder-decoder network for low- dose CT via transfer learning from a 2D trained network[J], IEEE Transactions on Medical Imaging, 37(6): 1522-1534, 2018.
[36]Liu J., Zhang Y.,Zhao Q., et al, Deep Iterative Reconstruction Estimation (DIRE): Approximate iterative reconstruction estimation for low dose CT imaging[J], Physics in Medicine and Biology, 64(13): 135007, 2019.
[37]Li M., Hsu W., Xie X., et al, SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network[J], IEEE Transactions on Medical Imaging, 39(7): 2289-2301, 2020.
[38]Choi K., Lim J. S., and Kim S., StatNet: Statistical image restoration for low-dose CT using deep learning[J], IEEE Journal of Selected Topics in Signal Processing, 14(6): 1137-1150, 2020.

[39]Liu J.,Kang Y. , Qiang J., et al, Low-Dose CT Imaging via Cascaded ResUnet with Spectrum Loss[J], Methods, 202: 78-87, 2021.
[40]Feng Z., Cai A., Wang Y., et al, Dual residual convolutional neural network (DRCNN) for low-dose CT imaging[J], Journal of X-Ray Science and Technology, 32(12): 1202-1212, 2021.
[41] Huang Z., Chen Z., Chen J., et al, DaNet: dose-aware network embedded with dose-level estimation for low-dose CT imaging[J], Physics in Medicine and Biology, 66(1): 015055, 2021.
[42]Liu J., Xia Z., Kang Y., et al. Low Dose CT Noise Artifact Reduction Based on Multi-scale Weighted Convolutional Coding Network[C], In Proceedings of IEEE International Conference on Systems and Informatics (ICSAI), pp:1-6, 2021
[43]Kelly B., Matthews T. P., and Anastasio M. A., Deep learning-guided image reconstruction from incomplete data[J], arXiv: 1709. 00584, 2017.
[44]Wu D., Kim K., ElFakhri G., et al, Iterative low-dose CT reconstruction with priors trained by artificial neural network[J], IEEE Transactions on Medical Imaging, 36(12): 2479-2486, 2017.
[45]Adler J. and Oktem O., Learned primal-dual reconstruction[J], IEEE Transactions on Medical Imaging, 37(6): 1322-1332, 2018.
[46] Ge Y., Su T., Zhu J., et al, ADAPTIVE-NET: Deep computed tomography reconstruction network with analytical domain transformation knowledge[J], Quantitative Imaging in Medicine and Surgery, 10(2): 415-427, 2020.
[47]Mehmet O. U., Metin E., and Isa Y., Low-dose CT reconstruction using deep generative regularization prior[J], arXiv: 2012.06448, 2020.
[48]Ding Q., Chen G., Zhang X., et al, Low-dose CT with deep learning regularization via proximal forward-backward splitting[J], Physics in Medicine and Biology, 65(12): 125009, 2020.
[49] He Z., Zhang Y., Guan Y., et al, Iterative reconstruction for low-dose CT using deep gradient priors of generative model[J], arXiv: 2009.12760, 2020.
[50]Xia W., Lu Z., Huang Y., et al, MAGIC: Manifold and graph integrative convolutional network for low-dose CT reconstruction[J], arXiv: 2008.00406, 2020.
[51] Hu D., Zhang Y., Liu J., et al, SPECIAL: Single-Shot Projection Error Correction Integrated Adversarial Learning for Limited-Angle CT[J], IEEE Transactions on Computational Imaging, 7: 734-746, 2021.

[52] Liu J., Kang Y. Q., Qiang J., et al. Deep residual constrained reconstruction via learned convolutional sparse coding for low-dose CT imaging[J]. Biomedical Signal Processing and Control, 85 (2023): 104868, 2023.
[53] Kang Y. Q., Liu J., Wu F., et al. Deep Convolutional Dictionary Learning Network for Sparse View CT Reconstruction with a Group Sparse Prior[J]. Computer Methods and Programs in Biomedicine, 244 (2024): 108010, 2024.
[54] Wang C., Shang K., Zhang H., et al. Dudotrans: Dual-domain transformer provides more attention for sinogram restoration in sparse-view CT reconstruction[J]. arXiv preprint arXiv: 2111.10790, 2023.
[55] Li R., Li Q., Wang H., et al. DDPTransformer: Dual-Domain with Parallel Transformer Network for Sparse View CT Image Reconstruction[J]. IEEE Transactions on Computational Imaging, 69(1): 62-71, 2022.
[56] Sun C., Deng K., Liu Y., et al. A Lightweight Dual-Domain Attention Framework for Sparse-View CT Reconstruction[J]. arXiv preprint arXiv: 2202.09609, 2022.
[57] Pan J., Zhang H., Wu W., et al. Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction[J]. Patterns, 3(6): 100498, 2022.
[58] Li Y., Sun X., Wang S., et al. MDST: multi-domain sparse-view CT reconstruction based on convolution and swin transformer[J]. Physics in Medicine and Biology, 68(9): 095019, 2023.
[59] Du, W. Tian S.. Transformer and GAN-Based Super-Resolution Reconstruction Network for Medical Images[J]. Tsinghua Science and Technology, 29(1): 197-206, 2023.
[60] Wu Z., Liao W., Yan C., et al. Deep learning based MRI reconstruction with transformer[J]. Computer Methods and Programs in Biomedicine, 233: 107452, 2023.
[61] Wang Y., Sun L., Zeng Z., et al. TRCT-GAN: CT reconstruction from biplane X-rays using transformer and generative adversarial networks[J]. Digital Signal Processing, 104123, 37(1): 171-184, 2023.
[62] Zhao X., Yang T., Li B., et al. SwinGAN: A dual-domain Swin Transformer-based generative adversarial network for MRI reconstruction[J]. Computers in Biology and Medicine,153: 106513, 2023.
[63] Zhang Y., Hu D., Yan Z., et al. TIME-Net: Transformer-integrated multi-encoder network for limited-angle artifact removal in dual-energy CBCT[J]. Medical Image Analysis, 83: 102650, 2023.
[64] Zhu L., Han Y., Xia X., et al. STEDNet: Swin transformer-based encoder-decoder network for noise reduction in low-dose CT[J]. Medical Physics,

68(9): 095019, 2023.
[65] 樊雪林, 文昱齐, 乔志伟. 基于 Transformer 增强型 U-net 的 CT 图像稀疏重建与伪影抑制[J]. CT 理论与应用研究, 33(01):1-12, 2024.
[66] 陈蒙蒙. 基于 Transformer 的 CT 稀疏重建方法研究[D]. 山西大学, 2023.
[77] Wu M., Xu Y., Lu Y., et al. Adaptively re-weighting multi-loss untrained transformer for sparse-view cone-beam CT reconstruction[J]. arXiv preprint arXiv:2203.12476, 2022.
[68] Zhao B., Cheng T., Zhang X., et al. CT synthesis from MR in the pelvic area using Residual Transformer Conditional GAN[J]. Computerized Medical Imaging and Graphics, 103: 102150, 2023.
[69] Chen Z., Niu C., Gao Q., et al. LIT-Former: Linking in-plane and through-plane transformers for simultaneous CT image denoising and deblurring[J]. IEEE Transactions on Medical Imaging, 2024.
[70] SIoffe S., Szegedy C.. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International conference on machine learning. 37(15): 448-456, 2015.
[71] Hu D., Liu J., Lv T., et al. Hybrid-domain neural network processing for sparse-view CT reconstruction[J]. IEEE Transactions on Radiation and Plasma Medical Sciences, 5(1): 88-98, 2020.
[72] Yin X., Zhao Q., Liu J., et al. Domain progressive 3D residual convolution network to improve low-dose CT imaging[J]. IEEE Transactions on Medical Imaging, 38(12): 2903-2913, 2019.
[73] Unal M., Ertas M., Yildirim I. An unsupervised reconstruction method for low-dose CT using deep generative regularization prior[J]. Biomedical Signal Processing and Control, 75: 103598, 2022.
[74] Huang Z., Chen Z., Quan G., et al. Deep cascade residual networks (DCRNs): Optimizing an encoder-decoder convolutional neural network for low-dose CT imaging[J]. IEEE Transactions on Radiation and Plasma Medical Sciences, 38(2): 203-213, 2022.
[75] Gu Y., Tang H., Lv T., et al. Discriminative Feature Representation for Noisy Image Quality Assessment[J]. Multimedia Tools and Applications, 2020, 79(1): 7783-7809.
[76] 程小霞, 崔学英, 郭映亭, 等. 基于注意力机制 U-Net 的低剂量 CT 图像去噪方法[J]. 太原科技大学学报, 2022, 43(2): 147-153.
[77] Chan Y., Liu X., Wang T., et al. An attention-based deep convolutional neural

network for ultra-sparse-view CT reconstruction[J]. Computers in Biology and Medicine, 161: 106888, 2023.

[78] Guan B., Yang C., Zhang L., et al. Generative modeling in sinogram domain for sparse-view CT reconstruction[J]. IEEE Transactions on Radiation and Plasma Medical Sciences, 38(9): 903-913, 2023.

[79] Cheng W., He J., Liu Y., et al. CAIR: Combining integrated attention with iterative optimization learning for sparse-view CT reconstruction[J]. Computers in Biology and Medicine, 6(59):107161, 2023.

[80] Wang X., He J., Liu Y., et al. DIDR-Net: a sparse-view CT deep iterative reconstruction network with an independent detail recovery network[J]. Journal of Instrumentation, 18(05): P05038, 2023.

[81] Chen X., Xia W., Yang Z., et al. SOUL-net: A sparse and low-rank unrolling network for spectral CT image reconstruction[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024.

[82] 于淼, 许铮铧. 生成对抗网络医学图像去噪研究综述[J]. 中国生物医学工程学报, 41(06): 724-731, 2022.

[83] Yang H., Sun J., Carass Y., et al. Unpaired Brain MR-to-CT Synthesis Using a Structure-Constrained CycleGAN[J]. Lecture Notes in Computer Science, 20(5): 174–182, 2018.

[84] Salomonsson K., Oldgren E., Ström E., et al. Fast deep learning based reconstruction for limited angle tomography[J]. arXiv preprint arXiv:2402.12141, 2024.

[85] Wang D., Fan F., Wu Z., et al. CTformer: convolution-free Token2Token dilated vision transformer for low-dose CT denoising[J]. Physics in Medicine and Biology, 68(6): 065012, 2023.

[86] Zhang S., Wang Z., Yang H., et al. HFormer: highly efficient vision transformer for low-dose CT denoising[J]. Nuclear Science and Techniques, 34(4): 61-75, 2023.

[87] Yuan J., Zhou F., Guo Z., et al. HCformer: hybrid CNN-transformer for LDCT image denoising[J]. Journal of Digital Imaging, 36(5): 2290-2305, 2023.

[88] Adishesha A. S., VanselowD. J., La P. R., et al. Sinogram Domain Angular Upsampling of Sparse-View Micro-CT with Dense Residual Hierarchical Transformer and Noise-Aware Loss[J]. bioRxiv, 5(9):540072, 2023.

[89] Peng W., Xiao Y.. Sparse-view CT reconstruction method for in-situ non-destructive testing of reinforced concrete[J]. Nondestructive Testing and Evaluation, 1-18, 2023.

[90] Pan J., Yu H., Gao Z., et al. Iterative Residual Optimization Network for Limited-angle Tomographic Reconstruction[J]. IEEE Transactions on Image Processing, 39(1): 126-138, 2024.
[91] Lin J., Li J., Dou J., et al. Dual-Domain Reconstruction Network Incorporating Multi-Level Wavelet Transform and Recurrent Convolution for Sparse View Computed Tomography Imaging[J]. Tomography, 10(1): 133-158, 2024.
[92] Ni Y., Chen G., Feng Z., et al. DA-Tran: Multiphase liver tumor segmentation with a domain-adaptive transformer network[J]. Pattern Recognition, 149: 110233, 2024.

攻读学位期间发表的学术成果

已录用的论文：

[1] Zhang T. Y., Liu J., Wu F., et al. Artifact Suppression for Sparse View CT via Transformer-Based Generative Adversarial Network[J]，Biomedical Signal Processing and Control.

[2] 张庭宇，吴凡，金潼，等. 基于 Transformer 块的混合域网络稀疏角度 CT 成像[J]. 重庆工商大学学报(自然科学版)。

[3] Liu J., Zhang T. Y., Kang Y. Q., et al. SureUnet: sparse autorepresentation encoder U-Net for noise artifact suppression in low-dose CT [J]， Neural Computing and Applications，pp: 1-13, 2023.

[4] Liu J., Zhang T. Y., Kang Y. Q., et al. Deep residual constrained reconstruction via learned convolutional sparse coding for low-dose CT imaging [J], Biomedical Signal Processing and Control, 85:104868, 2023.

[5] Liu J., Kang Y. Q., Zhang T. Y., et al. Deep Iterative Reconstruction Network Based on Residual Constraint for Low-Dose CT Imaging [C], the 2022 8th International Conference on Systems and Informatics (ICSAI), pp: 1-6, 2022.

致 谢

在硕士毕业论文即将完成之际，我衷心感谢所有在我研究生生涯中给予我帮助、支持和鼓励的人。正是他们的陪伴和付出，让我能够顺利走过这段学术旅程，完成这篇论文。

首先，我要感谢我的导师。导师的严谨治学态度、深厚的学术造诣和敏锐的洞察力为我提供了宝贵的学术指导。在论文的选题、研究方法和撰写过程中，导师都给予了我悉心的指导和无私的帮助。在此，我向导师致以最崇高的敬意和衷心的感谢。

其次，我要感谢实验室的同学们。在研究生阶段，我们共同度过了许多难忘的时光。在学术上，我们互相学习、互相启发；在生活中，我们互相照顾、互相支持。

此外，我还要感谢我的家人和朋友。在我求学的道路上，他们始终是我最坚强的后盾。他们给予我无尽的关爱和鼓励，让我在面对困难和挑战时能够保持信心和勇气。

最后，我要感谢所有参与论文评审和答辩的老师们。他们的宝贵意见和建议使我的论文更加完善。在此，我向他们表示衷心的感谢，并期待在未来的学术道路上能够继续得到他们的指导和帮助。

在这段研究生生涯中，我收获了宝贵的学术成果和人生经历。我将铭记这段经历，继续前行，为学术研究和社会进步贡献自己的力量。再次感谢所有给予我帮助、支持和鼓励的人！