

基于自注意力机制U-net的微焦CT去射线源模糊方法

刘川江^{1,2}, 王奥², 张根源², 袁伟², 刘丰林^{1,2*}¹重庆大学机械与运载工程学院, 重庆 400044;²重庆大学工业CT无损检测教育部工程研究中心, 重庆 400044

摘要 在微焦CT成像中,通常利用增大X射线源管电压、管电流来提高扫描效率,但射线源功率增加会导致焦点尺寸增大,投影图像模糊,从而降低重建图像的空间分辨率。为了解决因非理想射线源焦点引起的图像模糊问题,本文提出利用深度学习在投影域映射非理想焦点与理想焦点投影之间的关系。推导了理想焦点投影与非理想焦点投影的正向关系,基于该关系构建配对数据集;提出一种基于自注意力机制的U-net网络(SU-net)学习非理想焦点投影到理想焦点投影的逆向关系。仿真实验和实际实验结果表明,提出的SU-net方法能准确地从非理想焦点投影中估计出理想焦点投影,可有效减少焦点导致的图像模糊。

关键词 计算机断层扫描; 微焦点CT; 空间分辨率; 深度学习; X射线源焦点

中图分类号 TP391 文献标志码 A

DOI: 10.3788/AOS231855

1 引言

在CT成像中,影响空间分辨率的因素有很多,包括X射线源焦点尺寸、探测器单元尺寸、几何放大比、机械精度、重建算法和物体材料等^[1-7]。其中,前三者决定了系统的物理分辨率。射线源焦点尺寸对分辨率产生影响的原因在于,在几何放大比 m 的作用下,X射线源有效焦点尺寸 a 会在探测器上产生半影模糊 $b = a(m - 1)$ 。当半影模糊大于一个探测器像素尺寸 d 时($b > d$),会使测量投影的边缘模糊。

减小半影模糊最直接的方法是减小焦点尺寸,焦斑尺寸越小,半影模糊越小,图像越清晰;焦斑尺寸越大,半影模糊越大,图像边缘越模糊。由于在X射线管中,焦点尺寸与电子束功率直接相关,电子束功率越小,焦点尺寸越小。因此,可以通过减小电子束功率减小焦点尺寸。但这会导致重建图像达到所需的信噪比需要更长的曝光时间,扫描效率降低;同时功率太小也会导致X射线的能量不足以穿透样品的风险。

另一方面,基于软件的方法可补偿非理想焦点模糊。该方法分为三类:基于投影域、基于图像域和基于模型迭代的方法。基于投影域的方法通过构建模糊模型直接对测量投影去卷积。例如,Chang等^[8]通过建立非理想焦点投影模型和理想点源投影模型之间的近似线性等价关系,从大焦点的测量模糊投影中解析出理想焦点的清晰投影,然后用经典迭代算法重建出高分

辨率CT图像。该类方法的优点在于,可以根据投影几何关系直观地建立数学模型,但对模型精度要求较高,因为投影域的微小误差会累积到图像域。基于图像域的方法通过求解模糊核在图像域的表达形式然后利用反卷积去模糊。例如,Hu等^[9]解析推导了滤波反投影(FBP)算法重建图像中的非理想焦点模糊核,Lorenz等^[10]通过在模糊图像高频段引入尺度不变稀疏正则项函数(NSM)估计出模糊核。这类方法不仅可以有效解决焦点模糊,还能解决探测器模糊、重建算法模糊、运动模糊等,但难点在于精确求解模糊核,因为Radon变换会将模糊核表达式复杂化。基于模型的重建方法(MBIR)将焦点模糊的各种物理和统计特性纳入重建模型中。例如,2015年,Tilley等^[11]建立了一个前向模型用于刻画有限焦点尺寸和探测器造成的模糊和噪声相关性,基于该模型对投影数据去卷积,然后进行广义最小二乘重建,利用非对角统计加权来解释采集和数据处理链中产生的相关性。随后,他们又提出了基于模型模糊和噪声一致性的高斯惩罚似然目标(GPL-BC)算法^[12],包含模糊模型和噪声相关性的高斯惩罚似然目标函数,可以比对投影去模糊后采样Feldkamp-Davis-Kress (FDK)重建的方法实现更高的分辨率。虽然MBIR具有模型可解释性强且不依赖于训练数据的优点,但其性能依赖于预先设计的先验知识,如能量函数中的正则化项、噪声统计模型,或有效优化算法,且计算量大。

收稿日期: 2023-11-29; 修回日期: 2024-01-19; 录用日期: 2024-01-25; 网络首发日期: 2024-02-20

基金项目: 国家重点研发计划(2022YFF0706400)、中央高校基本科研业务费(2023CDJXY-005)

通信作者: *liufl@cqu.edu.cn

随着深度学习的发展,一些基于神经网络的方法已成功应用于自然图像处理中的去模糊。例如,Wang等^[13]在图像超分辨率生成对抗网络(SRGAN)^[14]的基础上提出了增强型超分辨率生成对抗网络(ESRGAN)超分辨重建算法,能从低分辨率的模糊图像中恢复高分辨率的清晰图像。这些方法能够从大量训练数据集中提取特征,并生成符合训练数据分布的具有很强拟合能力的非线性映射函数。然而,当训练数据集和测试数据集差异大时,端到端神经网络可能并不总是能提供令人满意的结果。这个问题在CT成像中更为明显,因为模态对图像细节的保真度比自然图像要求更严格,并且获取大量配对的CT训练数据难度大。2021年,Zhang等^[15]结合迭代重建和深度学习,提出了一个模块化数据驱动重建框架,直接从模糊正弦图中重建高分辨率CT图像,同时可以根据控制模块化数量获得一系列渐进去模糊的图像。

为了解决非理想焦点引起的图像模糊,本文根据非理想焦点模糊投影与理想焦点清晰投影之间的关系,提出一种基于自注意力机制的U-net(SU-net)模型,通过大量数据集,拟合非理想焦点模糊投影与理想焦点清晰投影之间的关系。根据通用近似定理^[16],当数据量和网络容量足够时,能很好地逼近这种关系。本文利用训练好的网络,从实际测量的非理想焦点模糊投影中估计出理想焦点清晰投影,然后用经典的重建算法重建图像,实现去除由有限焦点尺寸引入的图像模糊。

2 基本原理

2.1 理论

在CT重建算法中,一般假设X射线源为理想点

源。因此,根据Beer-Lambert定律,对于入射强度为 $I_0(s, t)$ 的光子,经过物体衰减过后的射线强度为

$$I(s, t) = I_0(s, t) \exp[-p(s, t)],$$

$$\text{with } p(s, t) = \int_L f(l) dl, \quad (1)$$

式中: $p(s, t)$ 表示物体衰减系数的函数 $f(l)$ 沿着路径 L 的线积分(投影); s 和 t 分别是射线源坐标和探测器坐标。

在实际CT成像模型中,受X射线管散热能力和阳极靶材料的影响,焦点不可能无限小。对于微焦点X射线源,其焦点大小一般在几微米到几十微米,而对于医用X射线源,其焦点大小可达几毫米。当焦点满足 $a \leq a_0$ 时[其中, $a_0 = d/(m-1)$],理想点源模型能清晰重建CT图像,但当 $a > a_0$ 时,理想点源模型将不再有效。因此,考虑将焦点尺寸引入成像模型中,实际入射光子强度应为

$$I_a(t) = \int_{\Omega} I_0(s, t) ds, \quad (2)$$

经物体衰减后实测的射线强度应为

$$I_m(t) = \int_{\Omega} I(s, t) ds = \int_{\Omega} I_0(s, t) \exp[-p(s, t)] ds, \quad (3)$$

式中, Ω 为有限焦点支撑。因此基于非理想焦点的线积分(投影)可以表示为

$$q(t) = -\ln \left\{ \int_{\Omega} I_0(s, t) \exp[-p(s, t)] ds \right\} + \ln[I_a(t)], \quad (4)$$

式中, $I_a(t)$ 表示实际入射光子强度。为推导出非理想焦点投影 $q(t)$ 与理想点源投影 $p(s, t)$ 直接的关系,先引入一个辅助变量 $s_0 \in \Omega$,式(4)变为

$$q(t) = -\ln \left\{ \exp[-p(s_0, t)] \int_{\Omega} I_0(s, t) \exp\{-[p(s, t) - p(s_0, t)]\} ds \right\} + \ln[I_a(t)] =$$

$$p(s_0, t) - \ln \left\{ \int_{\Omega} I_0(s, t) \exp\{-[p(s, t) - p(s_0, t)]\} ds \right\} + \ln[I_a(t)], \quad (5)$$

由于焦点尺寸较小,可假设在给定焦点强度范围内的两个不同的点的投影差值较小。利用泰勒展开,式(5)可以展开为

$$q(t) = p(s_0, t) - \ln \left\{ \int_{\Omega} I_0(s, t) [1 - p(s, t) + p(s_0, t)] ds \right\} + \ln[I_a(t)] =$$

$$p(s_0, t) - \ln \left\{ 1 - \int_{\Omega} \frac{I_0(s, t)}{I_a(t)} [p(s, t) - p(s_0, t)] ds \right\} = \int_{\Omega} \frac{I_0(s, t)}{I_a(t)} p(s, t) ds. \quad (6)$$

如果进一步忽略X射线源的角强度分布,非理想焦点投影模型可简化为

$$q(t) = \int_{\Omega} w(s) p(s, t) ds, \text{ with } w(s) = \frac{I_0(s)}{\int_{\Omega} I_0(s) ds}. \quad (7)$$

式(7)表明,非理想焦点投影是给定点在有限焦点支撑

上的理想点源投影的加权积分。权重取决于给定点对整个有限焦点支撑的归一化强度贡献。式(7)可以离散为

$$q(t) = \sum_{i=1}^M w(s_i) p(s_i, t), \text{ with } w(s_i) = \frac{I_0(s_i)}{\sum_{i=1}^M I_0(s_i)}. \quad (8)$$

通常,非理想焦点 a 支撑内的射线强度服从均值为 0、标准差为 $a / (2\sqrt{2\ln 2})$ 的高斯分布。因此,可以很容易地通过理想点源投影近似得到非理想焦点投影。然而面临的主要问题是,如何通过实测的非理想焦点投影,估计出理想点源投影。

为了解决该问题,本文构建了一个基于自注意力机制的神经网络 \mathcal{G}_θ ,其中 θ 为神经网络参数。通过式(8)构建非理想焦点投影 $\mathbf{q} \in \mathbb{R}^{H \times W}$ -理想点源投影 $\mathbf{p} \in \mathbb{R}^{H \times W}$ 数据对,用于训练该神经网络。然后利用训练好的神经网络从非理想焦点投影中估计出理想点源投影,即

$$\hat{\mathbf{p}} = \mathcal{G}_\theta(\mathbf{q}), \quad (9)$$

式中, θ^* 为最优网络参数。

2.2 SU-net

U-net结构如图1(a)所示。该网络的主干由U-net^[17]构成,分为下采样部分和上采样部分。在下采样过程中,采用五个模块提取不同分辨率的特征。并在第二层至第五层中分别添加2、2、8、2个注意力机制模块,

用以更好地学习每层的特征信息。

与基于归一化点乘的注意力机制模块^[18]不同,本文采用卷积调制(CM)^[19]模块,如图1(b)所示。卷积调制模块与缩放点积注意力机制模块类似,主要包括用于实现空间编码的自注意力层和用于通道融合的全连接层,但它是通过用哈达玛乘积和卷积来实现。具体而言,对于一个给定的特征 $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$,通过如下方式计算输出:

$$\begin{cases} \mathbf{X}' = \mathbf{B} \odot \mathbf{V} \\ \mathbf{B} = \text{DConv}_{k \times k}(\mathbf{W}_1 \mathbf{X}), \\ \mathbf{V} = \mathbf{W}_2 \mathbf{X} \end{cases} \quad (10)$$

式中: $\text{DConv}_{k \times k}(\cdot)$ 表示卷积核大小为 $k \times k$ 的深度卷积; \odot 表示哈达玛乘积; \mathbf{W}_1 和 \mathbf{W}_2 是两个线性层的权重矩阵。采用卷积调制实现注意力机制的优点在于: 1) 可以从具有较大内核的卷积中受益更多,因此本文将卷积调制模块的卷积核大小设置为 11×11 ; 2) 与基于归一化点乘的注意力机制相比,该方式计算量线性增加,而不是二次方增加,因此可以减少计算量。

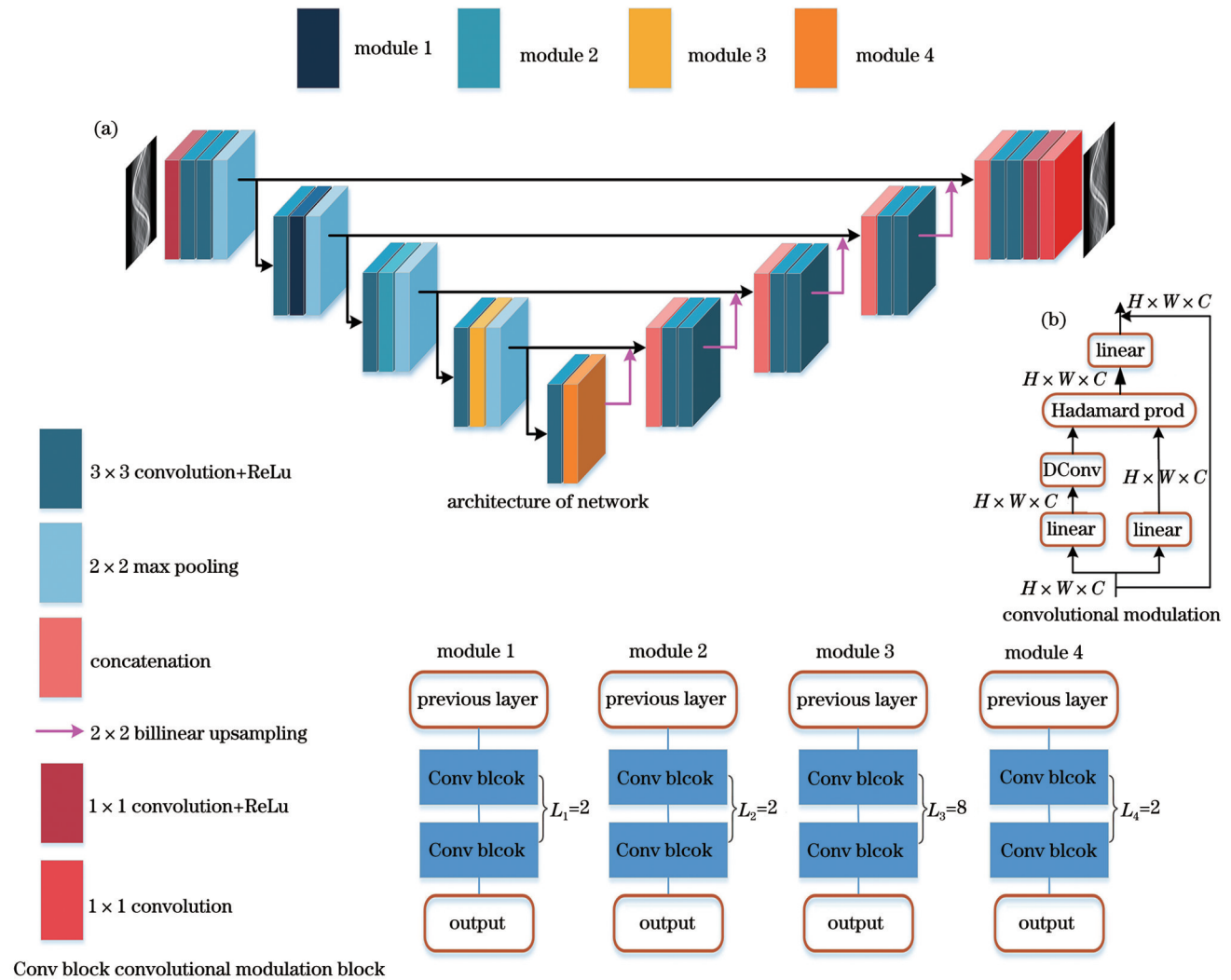


图1 网络结构图。(a) SU-net结构图;(b)卷积调制模块

Fig. 1 Network structure diagram. (a) SU-net structure; (b) convolutional modulation module

在上采样过程中,采用了四个模块去逐步上采样特征。为了避免棋盘状伪影^[20],采用双线性插值去扩大特征图。在每一层,来自下采样路径的高分辨率特征通过一个残差连接被连接到来自下一层的上采样输出中,以形成大量的特征通道。因此,网络可以将上下文信息传播到更高分辨率的层,以便于下一个卷积层能够学习更丰富的特征。

为了较好优化网络参数,提出一种联合损失函数,即

$$\theta^* = \arg \min_{\theta} \mathcal{L}[\mathcal{G}_{\theta}(\mathbf{q}), \mathbf{p}] = \mathcal{L}_2 + \alpha \mathcal{L}_{\text{GDL}}, \quad (11)$$

式中, α 表示权重因子,通常设置为0.01。

$$\mathcal{L}_2 = \frac{\sum_{i=1}^N [\mathcal{G}_{\theta}(\mathbf{q}_i) - \mathbf{p}_i]^2}{N}, \quad (12)$$

$$\mathcal{L}_{\text{GDL}} = \left| \left| \nabla_x \mathcal{G}_{\theta}(\mathbf{q}) \right| - \left| \nabla_x \mathbf{p} \right| \right|^2 + \left| \left| \nabla_y \mathcal{G}_{\theta}(\mathbf{q}) \right| - \left| \nabla_y \mathbf{p} \right| \right|^2. \quad (13)$$

\mathcal{L}_2 损失函数会导致图像边缘模糊、图像细节和解剖结构减少^[21]。因此,在 \mathcal{L}_2 损失函数的基础上引入了梯度偏差损失。文献[22-23]表明,梯度偏差损失能尽量保留具有强梯度的区域,能够对 \mathcal{L}_2 重建项进行有效补偿。

网络训练采用AdamW优化器优化所有参数,采用Xavier^[24]初始化所有参数。一共训练30个回合,批量大小设置为1,学习率采用余弦衰减策略从0.001衰减到0.0001。此外,为了提高训练的稳定性,采用了梯度裁剪^[25]策略,将梯度的范数限制为1。用于训练的计算机平台配置为Inter(R) Xeon(R) Silver 4210R CPU @2.40 GHz, 64 GB RAM 和 NVIDIA GeForce RTX 3090, 24 GB RAM。

3 实验与结果

3.1 仿真实验与结果

采用公开数据集2DeteCT^[26]构建用于训练的数据对,它是一个采用微焦CT设备扫描的2D扇束CT数据集,其扫描样品由各种各样的干果、坚果(如杏干、香蕉、枣子、无花果、芒果、葡萄干和咖啡豆,以及杏仁、腰果、榛子、副坚果、花生、山核桃、开心果和核桃),以及不同类型的岩石组成。该数据集包含5000组数据,每组数据都提供了高质量、低剂量、含射束伪影的重建图像和对应的投影数据。

将5000张高质量的重建切片用于生成理想点源投影-非理想焦点投影数据对。首先将整个数据集的值归一化到[0, 1],然后基于Astra^[27]工具包,采用扇束扫描几何计算理想点源投影。其几何参数为:射线源到旋转中心的距离为30 mm,射线源到探测器的距离为600 mm,探测器单元大小为0.1 mm,探测器单元个数为1024,在360°内均匀采集1024个投影,因此得到理想点源投影的大小为1024×1024。

将射线源焦点大小设置为50 μm ,根据式(8)计算出对应的非理想焦点投影。其加权函数均值为0,方差为21。由于在此参数下,避免半影模糊的最大焦点尺寸 a_0 为5.3 μm ,因此,将50 μm 焦点离散成10个理想点源(其中,离散点的个数通过 $N = \lceil a/a_0 \rceil$ 确定, $\lceil \cdot \rceil$ 为向下取整操作)。本文将构建的5000对理想点源投影-非理想焦点投影数据集,按8:1:1的比例分成训练集、验证集和测试集,用于训练和测试网络。在测试过程中,将网络估计的理想点源投影使用FBP^[28]算法重建。

为了评估该算法性能,将其与NSM^[10]算法和ESRGAN^[13]算法进行了对比。NSM算法是一种图像域去模糊算法,通过在模糊图像高频段引入尺度不变稀疏正则项 l_1/l_2 函数估计模糊核,然后采用迭代的方式从模糊图像中恢复清晰图像。该算法具有可适用于不同模糊模型、速度快、对参数稳健的特点。采用公开的MATLAB代码实现NSM算法,直接对有限焦点投影的FBP重建图像去模糊,并采用原文作者推荐的参数。ESRGAN算法是一种基于生成对抗性网络(GAN)^[29]的超分辨率重建算法。采用公开的PyTorch^[30]代码实现ESRGAN重建算法,将非理想焦点投影和理想点源投影的FBP重建图像作为数据对训练网络,迭代 10^6 次优化网络参数。采用均方根误差(RMSE)^[31]、峰值信噪比(PSNR)^[32]、特征相似度(FSIM)^[33]作为定量评价指标, RMSE越小、PSNR越大、FSIM越高,图像质量越好。

图2展示了两组随机挑选的测试结果,从图中可以看出,当焦点大小为50 μm 时,由于半影模糊,FBP重建图像边缘模糊。NSM方法能有效地去除这种模糊,但同时也放大了图像噪声。ESRGAN作为一种端到端的网络在测试集上也表现得较好,能有效去除模糊,但结果仍然不如提出的SU-net方法。相比于ESRGAN方法,SU-net能更真实地恢复图像细节,同时与NSM相比拥有更小的噪声。定量结果表明,SU-net拥有最小的RMSE、最大的PSNR和最高的FSIM,如图3所示。

为了验证SU-net的泛化性能,用数据库之外的图像做了测试。如图4所示,SU-net仍然能较好地恢复图像细节,而ESRGAN算法效果并不显著。NSM作为一种迭代重建算法不受数据库的影响,可以看到NSM几乎拥有和SU-net相同的效果。

3.2 消融实验与结果

为了验证所添加模块和梯度偏差损失的有效性,用与仿真实验相同的数据集及实验参数进行消融实验。实验结果如表1所示,其中WoU-net表示未添加卷积调制模块,SU-net-wo表示未包含梯度偏差损失,表中加粗表示最好结果。定量结果表明,SU-

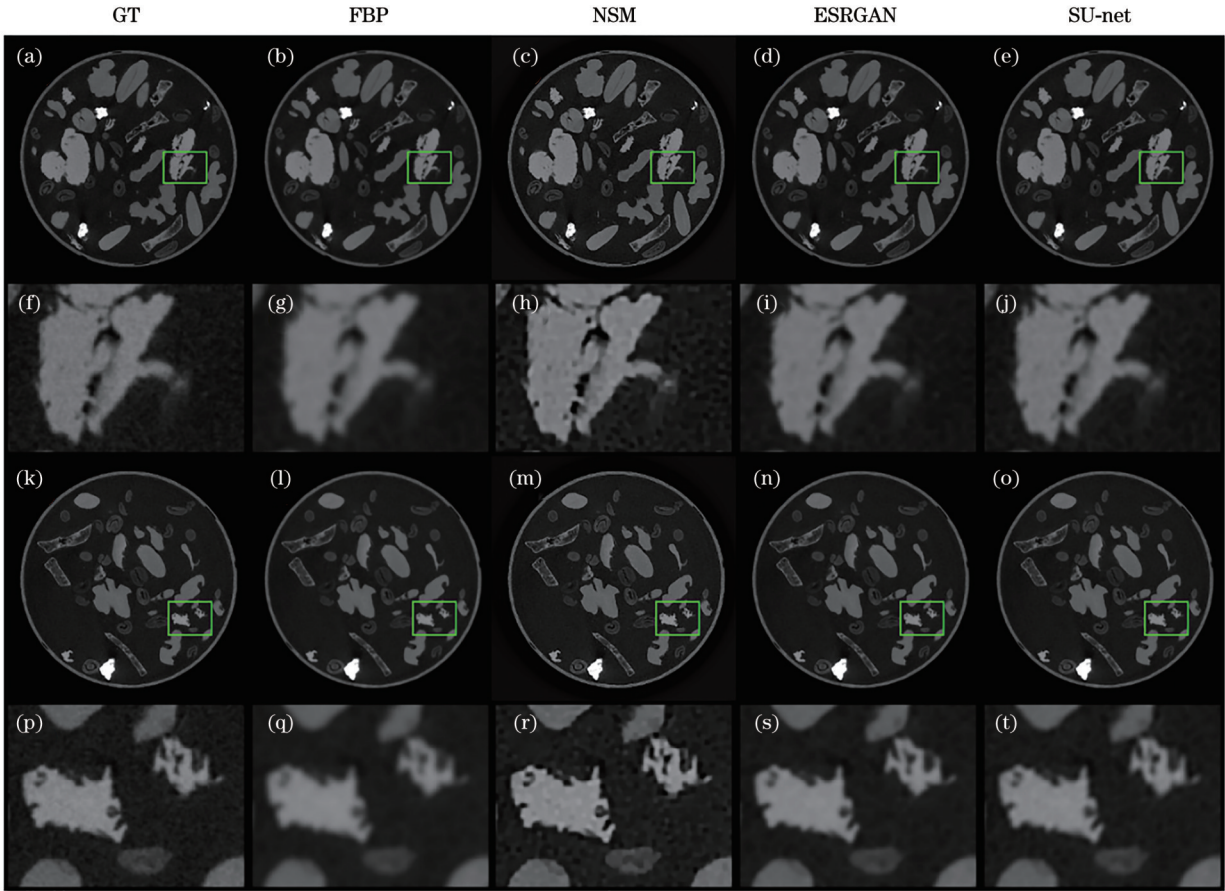


图2 测试结果图。(a)~(e)和(k)~(o)分别是真值图像、FBP、NSM、ESRGAN、SU-net重建图像；(f)~(j)分别是(a)~(e)的局部放大图；(p)~(t)分别是(k)~(o)的局部放大图。显示窗口为 $[0, 0.35] \text{ cm}^{-1}$

Fig. 2 Test result images. (a)~(e) and (k)~(o) are ground truth image, FBP, NSM, ESRGAN, and SU-net reconstruction images, respectively; (f)~(j) are partial enlarged images of (a)~(e), respectively; (p)~(t) are partial enlarged images of (k)~(o), respectively. Display window is $[0, 0.35] \text{ cm}^{-1}$

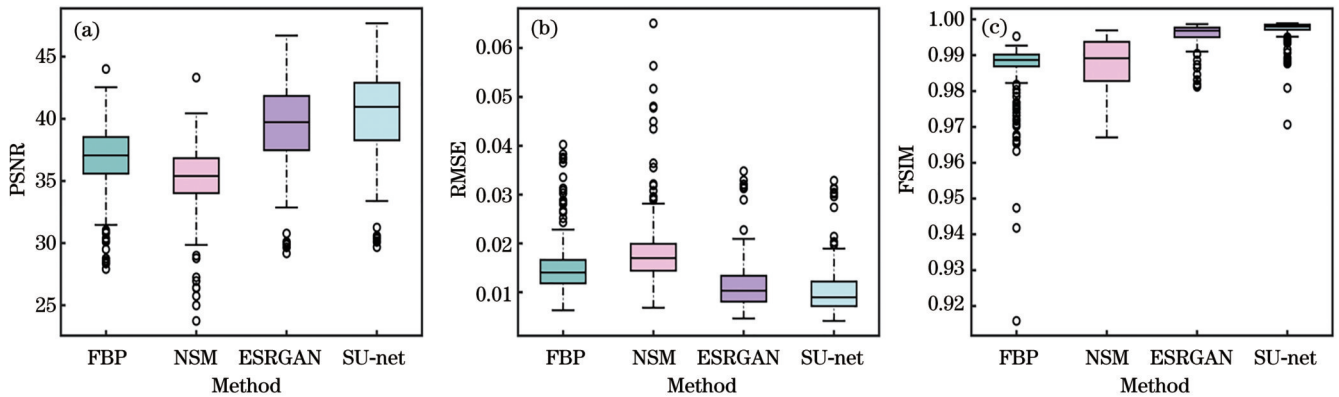


图3 定量指标。(a)峰值信噪比；(b)均方根误差；(c)特征相似度
Fig. 3 Quantitative index. (a) PSNR; (b) RMSE; (c) FSIM

net 相比 WoU-net, RMSE 减少了 0.001, PSNR 和 FSIM 分别提高了 0.8356 和 0.0006; 相比 SU-net-wo, RMSE 减少了 0.0004, PSNR 和 FSIM 分别提高了 0.3053 和 0.0001。

3.3 实际实验与结果

为了评价本文算法在实际数据上的效果,采用与

仿真实验相同的参数采集了实际数据。唯一不同的是,探测器大小为 0.2 mm。射线源电流、电压分别为 100 mA、120 kV。实际实验中,为了避免不同值的范围对结果的影响,将测得的实际投影用 FBP 重建后,归一化到 $[0, 1]$, 然后重投影获得模糊的投影图像。如图 5 所示,分别展示了第 600 行和第 650 行的结果,

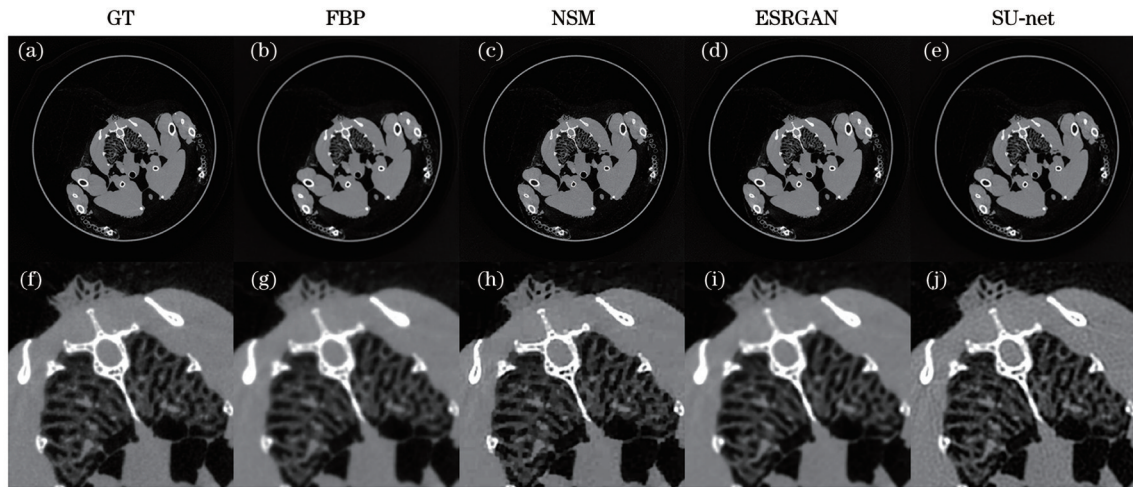


图4 数据集之外的实验结果。(a)~(e)分别为真值图像、FBP、NSM、ESRGAN、SU-net重建结果；(f)~(j)分别为(a)~(e)的局部放大图。显示窗口为 $[0, 0.35] \text{ cm}^{-1}$

Fig. 4 Experimental results outside dataset. (a)–(e) are ground truth image, FBP, NSM, ESRGAN, and SU-net reconstruction images, respectively; (f)–(j) are partial enlarged images of (a)–(e), respectively. Display window is $[0, 0.35] \text{ cm}^{-1}$

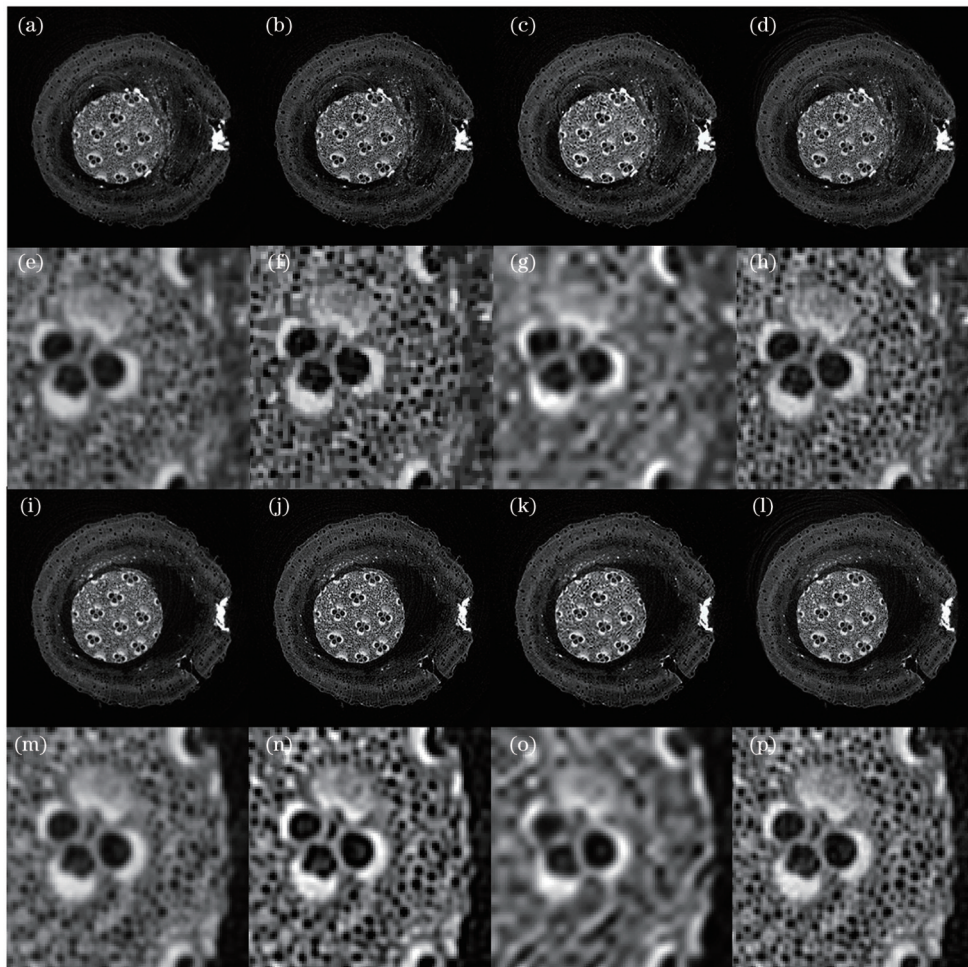


图5 实际实验结果。(a)~(d)分别为第600行的FBP、NSM、ESRGAN、SU-net重建结果；(e)~(h)分别为(a)~(d)的局部放大图；(i)~(l)分别为第650行FBP、NSM、ESRGAN、SU-net重建结果；(m)~(p)分别为(i)~(l)的局部放大图。显示窗口为 $[0, 0.35] \text{ cm}^{-1}$

Fig. 5 Results of actual experiment. (a)–(d) are reconstruction results of FBP, NSM, ESRGAN, and SU-net in line 600, respectively; (e)–(h) are partial enlarged images of (a)–(d), respectively; (i)–(l) are reconstruction results of FBP, NSM, ESRGAN, and SU-net in line 650, respectively; (m)–(p) are partial enlarged images of (i)–(l), respectively. Display window is $[0, 0.35] \text{ cm}^{-1}$

表 1 消融实验定量指标
Table 1 Quantitative index of ablation experiment

Method	CM	\mathcal{L}_{GDL}	RMSE	PSNR	FSIM
WoU-net	—	✓	0.0109	39.7491	0.9969
SU-net-wo	✓	—	0.0103	40.2794	0.9974
SU-net	✓	✓	0.0099	40.5847	0.9975

受半影模糊影响,直接FBP重建图像边缘模糊,部分微小细节丢失。NSM算法可以较好地去除模糊。对于ESRGAN算法,可以看到图像存在部分失真。提出的SU-net算法可以有效地让图像更清晰。实际情况下,很难获得参考图像,采用非参考图像质量评估模型:基于感知的图像质量评估器(PIQE)^[34]、自然图像

质量评估器(NIQE)^[35]及图像清晰度评价函数:离散余弦变换(DCT)函数^[36]作为定量评价指标。PIQE和NIQE值越小,图像的感知质量越好;DCT值越大,图像越清晰。定量结果如表2所示,其中加粗表示最好的结果。结果表明,提出的SU-net算法拥有最小的PIQE和NIQE值及最大的DCT值。

表 2 实际实验定量指标
Table 2 Quantitative index of actual experiment

Method	600			650		
	PIQE	NIQE	DCT	PIQE	NIQE	DCT
FBP	42.4410	3.7894	2.3910	40.6988	3.9677	2.3643
NSM	37.3960	5.0011	2.6883	36.3874	4.8068	3.2259
ESRGAN	76.8350	4.7775	2.0623	80.9565	4.7411	1.9283
SU-net	34.9580	3.3270	4.0553	35.8899	3.2659	4.2822

4 结 论

在CT成像中,实际X射线源焦点为有限尺寸的非理想焦点,当焦点尺寸较大时,用基于理想点源模型的CT算法直接对测量的投影进行重建会导致图像模糊,空间分辨率降低。为了解决此问题,本文从理论上推导了非理想焦点投影和理想焦点投影模型之间的关系,并在此基础上提出了一个SU-net用于从实际测量的模糊投影中估计出理想点源投影。仿真实验和实际实验表明,该方法能有效地从模糊投影中估计出清晰投影。

本文方法的优点是可以通过非理想焦点投影模型与理想点源投影模型之间的关系构建数据集,而不需要实际采集大焦点投影与小焦点投影组成的数据对,降低了构建数据集的难度;其次,直接根据非理想焦点投影模型与理想点源投影模型的关系构建网络具有较好的可解释性,即通过网络学习从非理想焦点模型到理想点源模型的逆向关系。与端到端的ESRGAN相比,具有更好的泛化性,尤其是对于图像细节的保真度较高的CT图像而言。

本文工作的局限性在于,只针对特定焦点大小和特定扫描几何进行训练,且未考虑噪声的影响,后续研究将训练适用于不同焦点大小和几何参数的网络和考虑含噪声的情况。

参 考 文 献

- [1] Hiller J, Maisl M, Reindl L M. Physical characterization and performance evaluation of an X-ray micro-computed tomography system for dimensional metrology applications[J]. Measurement Science and Technology, 2012, 23(8): 085404.
- [2] Welkenhuyzen F, Kiekens K, Pierlet M, et al. Industrial computer tomography for dimensional metrology: overview of influence factors and improvement strategies[C]//4th International Conference on Optical Measurement Techniques for Structures and Systems (OPTIMESS2009), May 25-26, 2009, Antwerp, Belgium. [S.l.: s.n.], 2009: 401-410.
- [3] Hiller J, Reindl L M. A computer simulation platform for the estimation of measurement uncertainties in dimensional X-ray computed tomography[J]. Measurement, 2012, 45(8): 2166-2182.
- [4] Yu H J, Li L, Tan C D, et al. X-ray source translation based computed tomography (STCT)[J]. Optics Express, 2021, 29(13): 19743-19758.
- [5] Chen J, Yu H J, Ni S, et al. Weighted filtered back-projection for source translation computed tomography reconstruction[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 4506409.
- [6] Ni S, Yu H J, Chen J, et al. Hybrid source translation scanning mode for interior tomography[J]. Optics Express, 2023, 31(8): 13342-13356.
- [7] 戈文杰, 余海军, 陈杰, 等. 基于求导-希尔伯特变换-反投影的源直线扫描计算机断层成像解析重建[J]. 光学学报, 2022, 42(11): 1134025.
Ge W J, Yu H J, Chen J, et al. Analytical reconstruction for source translation scanning computed tomography based on derivative-hilbert transform-back projection[J]. Acta Optica

- Sinica, 2022, 42(11): 1134025.
- [8] Chang M, Xiao Y S, Chen Z Q. Improve spatial resolution by Modeling Finite Focal Spot (MFFS) for industrial CT reconstruction[J]. *Optics Express*, 2014, 22(25): 30641-30656.
- [9] Hu X Y, Zhong Y C, Huang Y Q, et al. Improving small animal cone beam CT resolution by mitigating X-ray focal spot induced blurring via deconvolution[J]. *Physics in Medicine and Biology*, 2022, 67(12): 125005.
- [10] Hehn L, Tilley S, Pfeiffer F, et al. Blind deconvolution in model-based iterative reconstruction for CT using a normalized sparsity measure[J]. *Physics in Medicine and Biology*, 2019, 64(21): 215010.
- [11] Tilley S, Siewerdsen J H, Stayman J W. Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise[J]. *Physics in Medicine and Biology*, 2016, 61(1): 296-319.
- [12] Tilley S, Jacobson M, Cao Q, et al. Penalized-likelihood reconstruction with high-fidelity measurement models for high-resolution cone-beam imaging[J]. *IEEE Transactions on Medical Imaging*, 2018, 37(4): 988-999.
- [13] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[M]//Leal-Taixé L, Roth S. *Computer vision-ECCV 2018 workshops. Lecture notes in computer science*. Cham: Springer, 2019, 11133: 63-79.
- [14] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21–26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 105-114.
- [15] Zhang Z C, Yu L Q, Zhao W, et al. Modularized data-driven reconstruction framework for nonideal focal spot effect elimination in computed tomography[J]. *Medical Physics*, 2021, 48(5): 2245-2257.
- [16] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. *Neural Networks*, 1989, 2(5): 359-366.
- [17] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. *Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science*. Cham: Springer, 2015, 9351: 234-241.
- [18] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4–9, 2017, Long Beach, California, USA. New York: ACM Press, 2017: 6000-6010.
- [19] Hou Q B, Lu C Z, Cheng M M, et al. Conv2Former: a simple transformer-style ConvNet for visual recognition[EB/OL]. (2022-11-22)[2023-11-09]. <https://arxiv.org/abs/2211.11943>.
- [20] Odena A, Dumoulin V, Olah C. Deconvolution and checkerboard artifacts[J]. *Distill*, 2016, 1(10): e3.
- [21] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [22] Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-aware generative adversarial networks[M]//Descoteaux M, Maier-Hein L, Franz A, et al. *Medical image computing and computer-assisted intervention-MICCAI 2017. Lecture notes in computer science*. Cham: Springer, 2017, 10435: 417-425.
- [23] Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error[EB/OL]. (2015-11-17)[2023-11-09]. <https://arxiv.org/abs/1511.05440>.
- [24] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[J]. *Journal of Machine Learning Research*, 2010, 9: 249-256.
- [25] Pascanu R, Mikolov T, Bengio Y. Understanding the exploding gradient problem[EB/OL]. (2012-11-21)[2023-11-09]. <https://arxiv.org/abs/1211.5063v1>.
- [26] Kiss M B, Coban S B, Batenburg K J, et al. 2DeteCT-A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning[J]. *Scientific Data*, 2023, 10(1): 576.
- [27] van Aarle W, Palenstijn W J, Cant J, et al. Fast and flexible X-ray tomography using the ASTRA toolbox[J]. *Optics Express*, 2016, 24(22): 25129-25147.
- [28] Feldkamp L A, Davis L C, Kress J W. Practical cone-beam algorithm[J]. *Journal of the Optical Society of America A*, 1984, 1(6): 612-619.
- [29] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [30] Imambi S, Prakash K B, Kanagachidambaresan G R. PyTorch[M]//Prakash K B, Kanagachidambaresan G R. *Programming with TensorFlow. EAI/Springer innovations in communication and computing*. Cham: Springer, 2021: 87-104.
- [31] Chai T, Draxler R R. Root mean square error (RMSE) or mean absolute error (MAE)? -arguments against avoiding RMSE in the literature[J]. *Geoscientific Model Development*, 2014, 7(3): 1247-1250.
- [32] Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//2010 20th International Conference on Pattern Recognition, August 23–26, 2010, Istanbul, Turkey. New York: IEEE Press, 2010: 2366-2369.
- [33] Zhang L, Zhang L, Mou X Q, et al. FSIM: a feature similarity index for image quality assessment[J]. *IEEE Transactions on Image Processing*, 2011, 20(8): 2378-2386.
- [34] Venkatanath N, Praneeth D, Chandrasekhar B M, et al. Blind image quality evaluation using perception based features[C]//2015 Twenty First National Conference on Communications (NCC), February 27-March 1, 2015, Mumbai, India. New York: IEEE Press, 2015.
- [35] Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer[J]. *IEEE Signal Processing Letters*, 2013, 20(3): 209-212.
- [36] Marichal X, Ma W Y, Zhang H J. Blur determination in the compressed domain using DCT information[C]//Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348), October 24–28, 1999, Kobe, Japan. New York: IEEE Press, 2002: 386-390.

Source Blur Elimination in Micro-CT Using Self-Attention-Based U-Net

Liu Chuanjiang^{1,2}, Wang Ao², Zhang Genyuan², Yuan Wei², Liu Fenglin^{1,2*}

¹College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China;

²Engineering Research Center of Industrial Computed Tomography Non-Destructive Testing, Ministry of Education, Chongqing University, Chongqing 400044, China

Abstract

Objective Spatial resolution of X-ray imaging systems is crucial for microstructural object studies due to the small size of the subjects. Specifically, the focal spot size of the X-ray source is a main factor affecting the spatial resolution of micro-computed tomography (micro-CT), which will produce penumbra blur on detectors and thus blur the reconstructed images and reduce spatial resolution. Meanwhile, reducing the focal spot size by decreasing the X-ray tube power is a straightforward solution, but will prolong the scan duration. Therefore, we aim to develop a deep learning-based strategy by learning the inverse finite focal spot model to mitigate the penumbra blur for obtaining CT images with high spatial resolution even in the case of a non-ideal X-ray source.

Methods First, we derive the finite focal spot model that builds a relationship from the ideal point source projection to the finite focal spot projection. Based on the derived model, we numerically compute a paired projection dataset. Second, we utilize the neural network U-net and an attention mechanism module of convolution modulation block to build a self-attention mechanism-based U-net (SU-net) and thus learn the inverse finite focal spot model. The goal is to estimate the ideal point source projection from the actual non-ideal focal spot projection. SU-net (Fig. 1) which introduces convolution modulation blocks into the contracting path of the U-net is proposed to boost the U-net property. Finally, the standard filtered back-projection (FBP) is employed for reconstruction using the estimated ideal point projection.

Results and Discussions Simulation experiments are performed by the public dataset 2DeteCT to verify the effectiveness of the SU-net, which consists of a wide variety of dried fruits, nuts, and different types of rocks. Two groups of results are randomly selected in the test dataset for visualization (Fig. 2) and quantitative indicators are tested on the whole test dataset (Fig. 3). The results show that our proposed SU-net can estimate the ideal point source projection from the non-ideal focal spot projection. To verify the robustness of the SU-net, we test it with data outside of the simulation experimental dataset (Fig. 4), and the results show that it has better generalization than the end-to-end enhanced super resolution generative adversarial network (ESRGAN). Meanwhile, the ablation experiment is conducted with the same dataset and experimental parameters as the simulation experiment to confirm the validity of the added convolutional modulation module (CM) and gradient deviation loss, with quantitative indicators measured (Table 1). The results show that both the CM module and gradient deviation loss added by us can improve the network performance. Practical experiments are carried out to evaluate the effectiveness of the SU-net algorithm on real data (Fig. 5). Since it is difficult to obtain label data in the actual experiment, we select three evaluation indicators that do not require label data (Table 2), including PIQE (perception-based image quality evaluator), NIQE (natural image quality), and image sharpness evaluation function DCT (discrete cosine transform). The results show that our proposed SU-net algorithm achieves the optimal results compared with the comparison methods.

Conclusions In micro-CT imaging, the focal spot size of the actual X-ray source is limited, and under the relatively large focal spot size, the projected image will be blurred, and the reconstruction of the measured projection directly using the CT algorithm based on the point source model will cause the image to be blurred. We propose a U-net based on the self-attention mechanism to estimate the ideal point source projection from the actual measured non-ideal focal spot projection. Meanwhile, we establish a training dataset according to the relationship between the non-ideal focal spot projection and the ideal point source projection to optimize the network. Simulation and practical experiments show that this method can effectively estimate clear projection from blurred projection. The advantage of the proposed method is that we can construct a dataset by the relationship between the finite focal spot projection model and the ideal point source projection model, without collecting data pairs composed of non-ideal focal spot projection and ideal point source projection, which greatly reduces the difficulty of constructing datasets. Secondly, the proposed network directly based on the relationship between the finite focal spot projection model and the ideal point source projection model has strong interpretability, which means the inverse relationship from the finite focal spot model to the ideal point source model is learned through the network. Therefore, this method has better generalization than end-to-end ESRGAN, especially for CT images with high fidelity of image details. Our limitation is that the training is conducted for a specific focal spot size and a specific scanning geometry without considering the influence of noise. Subsequent studies will train networks with different focal spot sizes and geometric parameters and consider situations with noise.

Key words computed tomography; micro-computed tomography; spatial resolution; deep learning; X-ray source focus