

电子科技大学  
UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

# 硕士学位论文

MASTER THESIS



论文题目 基于机器学习的低剂量 CT 图像降噪  
方法研究

学科专业	<u>计算机科学与技术</u>
学 号	<u>201921080512</u>
作者姓名	<u>陈柯成</u>
指导教师	<u>蒲晓蓉 教授</u>
学 院	<u>计算机科学与工程学院</u>

分类号

TN828.6

密级

公开

UDC<sup>注1</sup>

621.39

# 学位论文

## 基于机器学习的低剂量CT图像降噪方法研究

(题名和副题名)

陈柯成

(作者姓名)

指导教师

蒲晓蓉

教授

电子科技大学

成都

(姓名、职称、单位名称)

申请学位级别

硕士

学科专业

计算机科学与技术

提交论文日期

2022年3月21日

论文答辩日期

2022年5月20日

学位授予单位和日期

电子科技大学

2022年6月

答辩委员会主席

高辉 教授

评阅人

注1: 注明《国际十进分类法UDC》的类号。

# **Low-dose CT Image Denoising with Machine Learning**

A Master Thesis Submitted to  
University of Electronic Science and Technology of China

Discipline **Computer Science and Technology**

Student ID **201921080512**

Author **Chen Kecheng**

Supervisor **Prof. Pu Xiaorong**

School **School of Computer Science and Engineering**

## 独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

作者签名：\_\_\_\_\_

日期：2022年5月21日

## 论文使用授权

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权电子科技大学可以将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、汇编学位论文。

(保密的学位论文在解密后应遵守此规定)

作者签名：\_\_\_\_\_

导师签名：\_\_\_\_\_

日期：2022年5月21日



## 摘要

CT (Computed Tomography) 技术已经成为使用频率最高的医学检测手段之一。CT 成像过程产生的辐射会对人体产生不可避免的危害。因此,低辐射剂量的 CT 成像技术被逐渐发展和推广。然而,辐射剂量的降低会产生噪声的干扰,导致 CT 图像的质量显著下降,进而影响医生的诊断。为此,面向低剂量 CT 图像的降噪方法被提出,旨在实现低剂量 CT 图像的噪声移除,得到较高质量的 CT 图像。目前,基于深度学习的低剂量 CT 图像降噪方法由于其卓越地降噪效果成为了主流方法。这些方法通常利用仿真的低/正常剂量的 CT 图像对,搭建基于深度卷积神经网络的模型进行监督学习。本文针对现有基于深度学习的低剂量 CT 图像存在的严重依赖仿真数据集、下游任务(如病灶检测)脱节和模型可解释性差等问题展开深入研究,主要创新点如下:

1. 创新性提出基于噪声学习的无监督低剂量 CT 图像降噪新框架。现有方法大多严重依赖仿真的成对低/正常剂量 CT 数据集,提出一种无监督低剂量 CT 图像降噪方法;提出低剂量 CT 图像噪声学习技术,旨提取 CT 图像的真实噪声并扩充样本;利用丰富易得的正常剂量 CT 数据集和学习到的噪声样本,构建配对数据集。该方法能学习 CT 图像的真实噪声,可较好地克服基于仿真数据集噪声模式单一的缺点。

2. 创新性提出将低剂量 CT 图像降噪与下游高阶任务关联的新架构。现有方法大多忽略下游的高阶任务,本文提出利用一个联合的损失函数将上下游任务进行有效地连接。利用病灶检测任务关注的感兴趣区域,提出了一种基于感兴趣区域的感知损失,用于降噪效果的评价和损失计算,使降噪过程充分考虑下游的病灶检测效果。针对降噪模型与病灶检测模型训练的不适配问题,提出一种协同训练的策略。

3. 首次提出基于医生阅片行为启发的低剂量 CT 降噪模型。针对现有的基于深度卷积神经网络模型偏重局部信息忽略非局部信息的问题,通过总结放射科医生阅片的流程,将医生对非局部信息、局部信息、切片间信息的综合考虑行为抽象为模型表征。创新性的引入了图卷积神经网络用于非局部信息和切片间信息的提取;同时将传统卷积操作提取的局部信息与上述两者信息进行有机融合,实现综合考虑非局部和局部信息的降噪模式,提升降噪过程的可解释性。

**关键词:** 低剂量 CT, 降噪, 图卷积神经网络, 噪声学习, 感兴趣区域

**ABSTRACT**

Computed Tomography (CT) has become one of the most frequently used medical tests. Radiation generated during CT imaging will inevitably cause harm to human body. Therefore, CT imaging technology with low radiation dose is gradually developed and promoted. However, the reduction of radiation dose will produce noise interference, leading to a significant decline in the quality of CT images, and then affect the diagnosis of doctors. Therefore, a noise reduction method for low-dose CT images is proposed to achieve noise removal from low-dose CT images and obtain higher quality CT images. At present, the low dose CT image denoising method based on deep learning has become the mainstream method due to its excellent denoising effect. These methods usually use simulated low-dose/normal-dose CT image pairs to build a model based on deep convolutional neural network for supervised learning. Existing deep learning-based low-dose CT images denoising methods rely heavily on simulation data sets, disconnection of downstream tasks (such as focus detection) and poor interpretability of models. To address these limitations, the main innovations are as follows:

1. Innovatively proposed a new framework for unsupervised low-dose CT image denoising based on noise learning. Most existing methods rely heavily on simulated pairwise low/normal dose CT data sets. An unsupervised low dose CT image denoising method is proposed. Low dose CT image noise learning technology is proposed to extract the noise and expand the sample of CT image. Paired data sets were constructed using the widely available normal-dose CT data sets and the learned noise samples. This method can be used for specific noise reduction tasks and can overcome the shortcoming of single noise mode based on simulation data set.

2. Innovatively proposed a new architecture for connecting low-dose CT image denoising with downstream high-level tasks. Most of the existing methods ignore the downstream high-level tasks. In this paper, a joint loss function is proposed to connect the two tasks effectively. Based on the region of interest (ROI) of focus detection task, this paper proposed a novel perception loss based on ROI, which is used for the evaluation and loss calculation of noise reduction effect, so that the noise reduction process can effectively focus on the effect of focus region, and fully consider the feedback of downstream task. Aiming at the problem of mismatch between the training of noise reduction model and the

training of lesion detection model, a cooperative training strategy was proposed.

3. A low-dose CT image denoising model inspired by radiologist was proposed for the first time. This model mainly solves the problem that existing models based on deep convolutional neural network emphasize local information and ignore non-local information. By summarizing the process of radiologists' film reading, the comprehensive consideration behavior of doctors for non-local information, local information and intersection information is abstracted as model representation. The graph convolutional neural network is innovatively introduced to extract non-local information and inter-slice information. At the same time, the local information extracted by traditional convolution operation is organically fused with the above two information to realize the noise reduction mode that comprehensively considers non-local and local information and improves the interpretability of the noise reduction process.

**Keywords:** Low-dose CT Image, Denoising, Graph Convolutional Network, Noise Learning, Region of Interest

目 录

第一章 绪 论 .....	1
1.1 研究工作的背景与意义 .....	1
1.2 国内外研究历史与现状 .....	1
1.3 主要贡献与创新 .....	4
1.4 本论文的结构安排 .....	4
第二章 相关理论基础介绍 .....	6
2.1 低剂量 CT 图像的产生机理 .....	6
2.2 相关模型介绍 .....	7
2.2.1 基于卷积神经网络的低剂量 CT 图像降噪模型 .....	7
2.2.2 基于生成对抗网络的低剂量 CT 图像降噪模型 .....	11
第三章 基于噪声学习的无监督低剂量 CT 图像降噪 .....	15
3.1 问题定义 .....	15
3.2 算法动机及其概述 .....	15
3.3 基于噪声学习的无监督低剂量 CT 图像降噪算法 .....	16
3.3.1 数学变量声明 .....	16
3.3.2 基于变分自编码器的噪声提取与学习 .....	16
3.3.3 基于生成对抗网络的噪声提取与学习 .....	19
3.4 实验结果及其分析 .....	20
3.4.1 基线方法和评估指标 .....	20
3.4.2 数据集 .....	21
3.4.3 实验细节 .....	21
3.4.4 基于变分自编码的噪声提取与学习的实验结果与分析 .....	22
3.4.5 基于生成对抗网络的噪声提取与学习的实验结果与分析 .....	23
3.4.6 消融实验 .....	26
3.5 本章小节 .....	27
第四章 基于上下游任务关联的低剂量 CT 图像降噪 .....	28
4.1 问题定义 .....	28
4.2 算法动机及其概述 .....	28
4.3 基于上下游任务关联的低剂量 CT 图像降噪架构 .....	29
4.3.1 数学变量声明 .....	29

4.3.2	基于感兴趣区域的感知损失.....	29
4.3.3	上下游任务的关联机制.....	30
4.3.4	基于协同训练的连接策略.....	32
4.4	实验结果及其分析.....	34
4.4.1	基线方法和评估指标.....	34
4.4.2	网络结构.....	35
4.4.3	检测任务的结果及其分析.....	37
4.4.4	降噪任务的结果及其分析.....	38
4.4.5	消融实验.....	39
4.5	本章小节.....	40
<b>第五章</b>	<b>医生阅片行为启发的低剂量 CT 图像降噪</b> .....	<b>42</b>
5.1	问题定义.....	42
5.2	算法动机及其概述.....	42
5.3	医生阅片行为启发的低剂量 CT 图像降噪算法.....	44
5.3.1	数学变量声明.....	44
5.3.2	医生行为启发的低剂量 CT 图像降噪模型.....	44
5.3.3	整体架构和损失函数.....	47
5.4	实验结果及其分析.....	47
5.4.1	数据集.....	47
5.4.2	双盲评分实验的测试集.....	47
5.4.3	双盲阅读研究的细节.....	48
5.4.4	训练细节.....	49
5.4.5	实验结果和分析.....	49
5.4.6	消融实验.....	54
5.4.7	讨论.....	55
5.5	本章小节.....	56
<b>第六章</b>	<b>全文总结与展望</b> .....	<b>57</b>
6.1	全文总结.....	57
6.2	后续工作展望.....	58
致 谢	.....	59
参考文献	.....	60
攻读硕士学位期间取得的成果	.....	65

## 第一章 绪论

### 1.1 研究工作的背景与意义

CT (Computed Tomography) 图像是第一种不需要打开人体内部结构就能生成不重叠的轴向片的方法。与传统的 X 射线成像相比, CT 具有优越地对比分辨率和极为详细的解剖表征能力。同时, CT 成像具有可以从图像中选择性增强或移除结构的能力。日常生活中, CT 扫描已经成为必不可少的检查手段。然而, CT 的成像是基于电磁辐射穿透人体, 并由于人体的密度差异 (如人体中骨骼的密度大于脂肪的密度), 导致电磁波的衰减不同, 产生不同的成像能量。然而, 有研究表明 CT 扫描的辐射暴露可能会带来潜在的癌症风险, 尤其是对儿童会产生较大影响。因此, 低剂量 CT 成像技术已成为降低辐射风险一种临床选择。低剂量 CT 成像技术是相对于正常剂量的 CT 检查来说, 其辐射剂量显著降低, 常常对于肺癌的早期筛查来说, 低剂量 CT 的辐射剂量仅为正常剂量的 10%。尽管由于辐射剂量的降低, 成像的质量会有所下降, 但基于低剂量 CT 的肺癌早期筛查已经被证实可以大幅度减小肺癌的致死率。因此, 低剂量 CT 成像技术在最近 20 年受到极大地欢迎。

然而, 更精细化地诊断需要更为高质量的 CT 图像。相矛盾的是, CT 扫描剂量的降低会不可避免的会增加低剂量 CT 图像的噪声。图像的质量的降低会影响放射科医生判断和计算机辅助影像诊断高阶视觉任务的实现。因此, 发展切实有效的低剂量 CT 图像的降噪方法成为了一件十分有意义的研究工作。当前, 低剂量 CT 图像的降噪方法大致包括三类, 即基于正弦图滤波的、基于迭代重构的和基于深度学习的后处理方法 (将在下一小节详细阐述)。其中, 基于深度学习的方法取得了非常令人印象深刻地降噪效果, 这些方法常常采用基于深度卷积神经网络的模型和仿真的成对 (即低剂量和对应的组织结构完全匹配的正常剂量的 CT 图像对) 的训练集进行有监督的训练, 将降噪问题抽象为低剂量 CT 图像向正常剂量 CT 图像的一个映射问题。

### 1.2 国内外研究历史与现状

目前, 低剂量 CT 图像降噪技术大致包含四类。

#### (a) 基于正弦图滤波的方法

基于正弦图滤波的方法是经典的低剂量 CT 图像降噪方法, 这些方法直接对 CT 图像重建前的投影域正弦图数据进行降噪。各种变换的滤波方法可以被广泛采用<sup>[1]</sup>。显然, 由于基于正弦图滤波的方法可以直接感知到最原始的数据, 因此这

种方法常常具有十分不错地降噪效果。然而，由于 CT 重建前的投影域的正弦图数据是集成在设备中的，因此研究人员想要获取这部分数据进行处理是十分困难的<sup>[2]</sup>。此外，该方法据报道会使得 CT 图像的结构信息和分辨率产生不同程度的降低<sup>[3]</sup>。

### (b) 基于迭代重构的方法

在过去二十年中，基于迭代重构的方法在低剂量 CT 降噪领域中得到广泛的关注<sup>[4]</sup>。该方法将正弦图域数据的各种先验特性（如统计属性、图像域的先验知识和图像重建过程的参数）纳入一个逆问题求解的目标函数<sup>[5]</sup>，通过迭代求解的方式，将数据从正弦域不断地在图像域进行转换。基于迭代重构的方法常常利用自然图像具有的稀疏先验来处理低剂量 CT 降噪问题。一些自然图像降噪问题常用到的方法被引入低剂量 CT 图像降噪方法中，包括基于非局部均值的方法 (non-local means)<sup>[6]</sup>、基于总变差 (TV) 的方法<sup>[7]</sup>、基于字典学习的方法 (dictionary learning)<sup>[8]</sup> 和基于低秩 (low-rank<sup>[9]</sup> 先验的方法。尽管基于迭代重构的方法取得了令人印象深刻地降噪结果，其本身依然存在着一些局限性：首先，在大多数现代 CT 扫描仪上，基于反向滤波投影的图像重建方法已经被红外技术所取代，因此辐射剂量被大幅度的降低。同时，和基于正弦图滤波的方法相似，投影域的正弦图数据对于研究人员来说是不易获取的，该方法的实用性被大大降低<sup>[10]</sup>。最后，基于迭代重构的方法一般被认为是十分耗时的。以上所列的局限性极大限制了迭代重构技术在临床上的使用<sup>[2]</sup>。

### (c) 基于深度学习的后处理方法

受益于深度学习技术的发展，机器学习技术在诸如计算机视觉、自然语言处理等领域都得到了比较好的应用。对于医学图像领域，基于深度学习的方法已取得了具有影响力的成果<sup>[11]</sup>。基于深度学习的低剂量 CT 降噪，已成为目前的主流方法。该方法也是一种后处理方法，常直接作用于图像域。得益于易于获取的数据集<sup>[11,12]</sup> 和有价值的临床可用性，低剂量 CT 图像降噪领域是医学图像后处理中最活跃的研究方向。与传统方法相比，基于深度学习的低剂量 CT 图像降噪方法取得了较好的效果。一般来说，Chen 等人<sup>[13]</sup> 被认为首先采用简单的卷积神经网络 (Convolutional Neural Network, CNN) 来抑制低剂量 CT 图像的噪声。之后，研究人员提出了多种基于 CNN 低剂量 CT 图像降噪方法，包括基于 CNN 单点单框架<sup>[14]</sup> 和生成基于对抗网络 (Generative Adversarial Network, GAN) 的框架<sup>[15,16]</sup>。Red-CNN<sup>[17]</sup> (Residual Denoising Convolutional Neural Network) 是典型的基于 CNN 的低剂量 CT 图像降噪的基线方法，它利用均方误差损失来抑制噪声。Yang 等人 citingRN276 利用 W-GAN<sup>[18]</sup> (Wasserstein Generative Adversarial Network) 和感知损

失相结合<sup>[19]</sup>，以实现从低剂量 CT 图像到正常剂量的 CT 图像的风格学习。基于此，许多研究都旨在提高生成对抗网络的生成器性能。Wolterinkd 等人<sup>[18]</sup>提出了一种基于 GAN 的低剂量 CT 图像降噪方法。该方法结合卷积神经网络和对抗性卷积神经网络，分别训练一个使用了体素损失的降噪生成器网络和判别降噪结果的判别器网络。除了基于经典神经网络的降噪方法以外，有人提出了改进神经网络基础部分的降噪方法，例如，Q-AE<sup>[20]</sup>(Quarter Autoencoder) 使用了二次神经元代替传统神经网络中的内积，从而提高单个神经元的能力。CPCE<sup>[15]</sup>(Conveying Path-based Convolutional Encoder-decoder) 采用了一种基于传递路径的 CNN 模型。MAP-NN<sup>[16]</sup>(Modularization Neural Network) 是一种模块化的低剂量 CT 图像降噪深度神经网络。为在基于 GAN 的框架下，实现基于预训练的 VGG 模型<sup>[19]</sup>的感知损失计算，最近的一些研究也提出了用其他损失(如基于结构相似度损失<sup>[21]</sup>和基于自动编码器的损失<sup>[22]</sup>)取代感知损失)。然而，由于自然图像和医学图像存在分布特点的差异，预训练的 VGG 模型(通过 ImageNet 训练)在直觉上可能会遭受域的不匹配问题。

然而，现有的基于深度学习的低剂量 CT 图像降噪方法存在的潜在问题包括：

- 现有的基于深度学习的低剂量 CT 降噪方法常常简单的利用深度卷积神经网络进行降噪，但这种降噪过程对于放射科医生来说过于“黑盒”。因为卷积操作一般来说可以看作是一个局部算子来提取局部信息。对于放射科医生来说，他们常常利用多种信息去帮助他们从带有噪声的低剂量 CT 图像中得到一个高质量的观察。因此，医生对于现有的基于深度卷积神经网络的模型往往缺乏信任。探索基于放射科医生阅片行为模仿的低剂量 CT 降噪新机制值得研究。
- 现有基于深度学习的低剂量 CT 降噪方法往往基于仿真噪声的低/正常剂量成对训练集进行训练，并希望训练的模型能够在临床的特点降噪任务中发挥作用。由于成像设备、参数和辐射剂量的变化，不同数据集可能具有不同的噪声特点。现有的架构可能会在临床中由于噪声分布或类型的不同，造成潜在的降噪效果的退化。然而，在临床中，往往不能获得组织结构完全匹配的低/正常剂量的数据集，为一个无监督的问题。存在因此，探索基于不同临床环境的噪声，实现低剂量 CT 图像的无监督降噪具有很大潜力。
- 现有基于深度学习的低剂量 CT 降噪任务大多属于独立任务，未与后续病灶检测等任务相互连接，往往导致将一些不显著的病灶信息当成噪声被消除或削弱，严重影响临床诊断。同时，现有的降噪方法没有显式的对感兴趣的病灶区域进行针对性的降噪。因此，如何将降噪任务和病灶检测等高阶相互关



联，并使得在降噪过程尽可能保留小目标病灶信息，使医学影像降噪与临床诊断任务贯通构成一个闭环路径值得研究

### 1.3 主要贡献与创新

为了解决以上以上问题，提出 3 个主要创新点：

- 创新性地提出基于噪声学习的无监督低剂量 CT 图像降噪新框架，用于解决临床中缺少配对数据集的问题。通过噪声的提取，提出利用基于变分自编码器和生成对抗网络的进行噪声样本的扩充和学习，充分使用丰富易得的正常剂量 CT 数据集和学习的噪声样本，构建伪数据对，实现无监督低剂量 CT 图像降噪问题的有效解决。
- 创新性提出将低剂量 CT 图像降噪与下游高阶任务关联的新架构，通过联合学习一个复合的损失函数，将医学图像降噪任务与下游检测任务连接起来，实现对于病灶区域的降噪效果和检测精度的协同提高；同时，为了更好地连接两个任务，提出了一种新颖的感兴趣区域驱动的兴趣损失。最终，提出一种病灶启发的低剂量 CT 图像降噪算法。
- 首次提出基于医生阅片行为启发的低剂量 CT 降噪模型，增强模型的可解释性和可用性，有机地融合现有的深度卷积神经网络和图卷积神经网络，实现综合考虑非局部、局部和上下文信息的降噪模式，提升降噪过程的可解释性，实现一种医生阅片行为启发的低剂量 CT 图像降噪算法。

### 1.4 本论文的结构安排

本文的章节结构安排如下：

第一章，绪论。介绍的研究背景和意义，然后逐步分析现有国内外工作的现状和对应的局限性，最后提炼针对这种局限性的主要创新点和贡献进行详细的阐述。

第二章，相关理论基础介绍。本章对低剂量 CT 图像的噪声的产生机理进行介绍，概述具体的噪声类型。同时，对采用的深度神经网络的基本组件，如卷积操作、非线性函数、转置卷积等进行介绍，进而对采用的基于深度卷积神经网络低剂量 CT 图像降噪的代表模型进行概述。也扩展的介绍了将要高频采用的基于生成对抗网络的低剂量 CT 图像降噪模型的基本原理、降噪机理和模型基本架构。

第三章，基于噪声学习的无监督低剂量 CT 图像降噪算法。本章旨在解决现有的方法往往只基于仿真噪声的低/正常剂量成对训练集进行训练的缺点，旨在提出能够适应于特点降噪场景的算法。本章详细介绍了如何通过噪声学习的方式将临床场景的无监督学习问题转化为监督学习问题，充分利用可以广泛获取的正常剂

量的 CT 图像进行伪的低剂量 CT 图像的构建。采用了基于变分自编码器和生成对抗网络的两种噪声学习的方式对噪声进行扩充。最后，进行了噪声学习的实验验证，证明两种噪声学习的有效性。同时，通过量化结果证明了对特定降噪场景进行针对性的降噪的优良的效果。

第四章，基于上下游任务关联的低剂量 CT 图像降噪算法。本章旨在解决现有的医学图像降噪算法与下游的高阶任务（如病灶检测）脱节的问题。阐明现有的方式可能造成最优的降噪结果可能并不是最优的下游任务的输入的问题，详细介绍了如何通过构建一个统一的损失函数进行两个任务的有效连接。同时，本章对现有基于深度学习的降噪方法对感兴趣区域的较少关注，提出了一种基于感兴趣的感知损失方法，用于指导降噪的进行。最后，本章通过详细的降噪任务和检测任务的量化指标、降噪任务与检测任务可视化的对比分析说明了算法的有效性。

第五章，医生阅片行为启发的低剂量 CT 图像降噪算法。本章针对现有的基于深度卷积神经网络缺乏模型很好地可解释为动机，提炼放射科医生阅片过程中的克服噪声干扰的行为抽象，用于指导模型的设计。详细介绍了提出的基于图卷积神经网络的非局部信息的获取和基于图神经网络的上下文信息的获取。介绍了构建的统一的医生阅片行为启发的低剂量 CT 图像降噪模块。最后，通过严谨的双盲评分实验、消融实验和训练细节的可视化说明了本章介绍算法的合理性。并进一步对算法的优缺点通过讨论环节进行了详细的阐述。

第六章，全文总结和展望，本章进行全文主要内容的总结，然后针对以后的工作进行详细的阐述。

## 第二章 相关理论基础介绍

### 2.1 低剂量 CT 图像的产生机理

CT 扫描是在计算机处理的帮助下使用 x 射线的精细化的检测方法。相比 x 射线，它能够获取多个横切面的图像信息。CT 扫描比常规 x 线更加详细，可以揭示各种解剖结构的形状、尺寸、密度和内部缺陷。图 2-1 为 CT 图像的产生机理示意图。因此，CT 扫描仪使用机动的 x 射线源，在围绕患者旋转时发射狭窄的 x 射线束。在 x 射线源的正对面有特殊的数字 x 射线探测器。当 x 射线穿过病人的身体时，它们被探测器捕获并传送到计算机上。通过氩和反氩变换对透射投影数据进行进一步处理。在此过程中，利用反投影算法将图像重构为 CT 图像。最后，重建的图像切片可以以二维的形式单独显示，也可以堆叠在一起生成三维图像。对重建后的 CT 图像质量进行分析和校正是一项强制性的后处理任务。这主要是由于为了防止患者受到不良地辐射照射而降低了 x 光管电流，影响了重建质量。

在低剂量 CT 图像采集过程中，CT 图像通常会受到量子噪声和各种伪影的影响。其中，在图像采集<sup>[15]</sup>过程中，由于 x 射线光子饥饿，量子噪声嵌入到低剂量 CT 图像中。量子噪声在视觉上的退化主要表现在边缘的分离、目标精细结构的平滑以及由于缺乏 x 射线光子而产生的低对比度视觉。物理上，量子噪声图像空间呈非均匀分布。因此，由于很难区分 CT 图像 [45] 中实际的噪声内容，低剂量 CT 图像恢复算法的验证和学习变得具有挑战性。在实验<sup>[23]</sup>过程中，通常采用泊松分布近似计算量子噪声。此外，还有一些应用是利用混合 Poisson 高斯分布 (MPGD)<sup>[24]</sup>估计 CT 图像的噪声分布。在 MPGD 中，电子噪声和量子噪声分量将分别使用高斯分布和泊松分布建模。

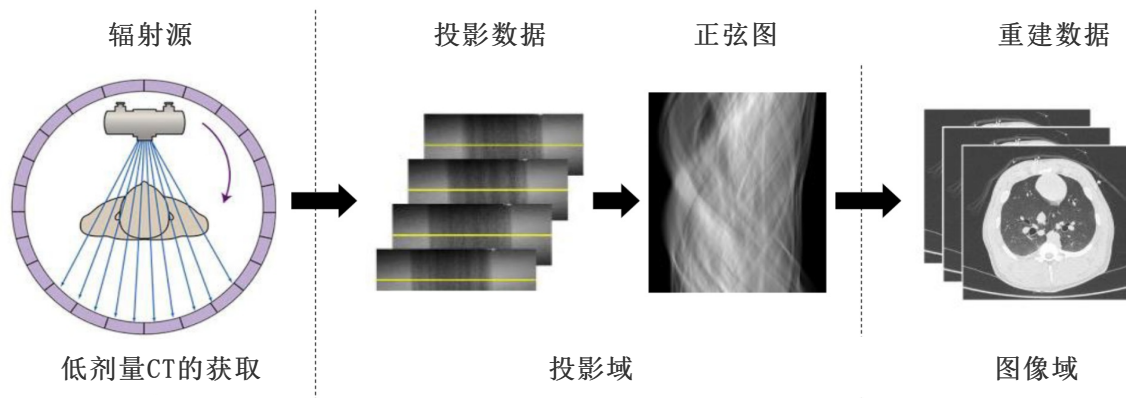


图 2-1 低剂量 CT 图像的产生机理

除了噪声之外，低剂量 CT 图像还因模糊和条纹伪影<sup>[25,26]</sup>而退化。CT 扫描和患者运动期间缺少 x 射线光子也会导致模糊。此外，当扫描一些细微结构时，CT 扫描设备可能会产生一些障碍（如肝脏的病灶）也会导致条纹伪影的产生，出现深色的条纹带。总之，低剂量 CT 图像的噪声产生是由于辐射剂量的降低产生的，噪声分布又是非平稳、分布非均一的复杂噪声，这导致了低剂量 CT 图像降噪任务十分困难，极具研究价值。

## 2.2 相关模型介绍

在第三章“基于噪声学习的无监督低剂量 CT 图像降噪算法”和第五章“基于上下游任务关联的低剂量 CT 图像降噪算法”使用了两类主要的降噪模型，即基于卷积神经网络和生成对抗网络的低剂量 CT 图像降噪模型。因此，将在此介绍相关理论基础。

### 2.2.1 基于卷积神经网络的低剂量 CT 图像降噪模型

#### 1. 卷积运算

卷积运算是基于卷积神经网络的低剂量 CT 图像降噪模型的基本组件，它能够通过窗口滑动的形式，以固定的卷积核（即需要学习的参数）大小提取图像的局部特征，并且具有旋转不变性等特点。同时，单层的卷积神经网络可以通过不断的堆叠，扩大自身的感受野的大小，以实现更大范围的图像特征的感知。数学上，基本的卷积操作，可以写为

$$y_k = \sum_{n=0}^{N-1} a_{k,n} x_n + b_k = \mathbf{a}_k^T \mathbf{x} + b_k, \quad k = 0, 1, \dots, K-1 \quad (2-1)$$

其中  $\mathbf{a}_k$  代表第  $k$ -th 卷积核，是一个需要学习的参数。需要注意的是，卷积核具有“权值共享”的特性。具体来说，对于某一层的卷积神经网络来说，整个特征图  $\mathbf{x}$  上共享相同的卷积核参数。

此外，卷积核也是“局域连接”的，具体来说，卷积核只与特征图上的局部区域进行卷积元素，这个局部区域的大小由卷积核的大小决定。卷积核在特征图上以预设定的步长进行不断的滑动，以覆盖整个特征图。通常来说，常常也需要对特征图进行填充操作，以实现更好的滑动，避免在边缘出现无法以整数进行局部区域定义的问题。通常  $\mathbf{a}_k$  被定义为一个  $m \times m$  的形式， $m$  可根据网络结果自行决定。 $b_k$  成为偏置项，它通过添加一个可学习的常量（即给定的偏差）到输入，使得结果能够获得更自由的移动量。图 2-2 为卷积操作的计算流程。

#### 2. 激活函数

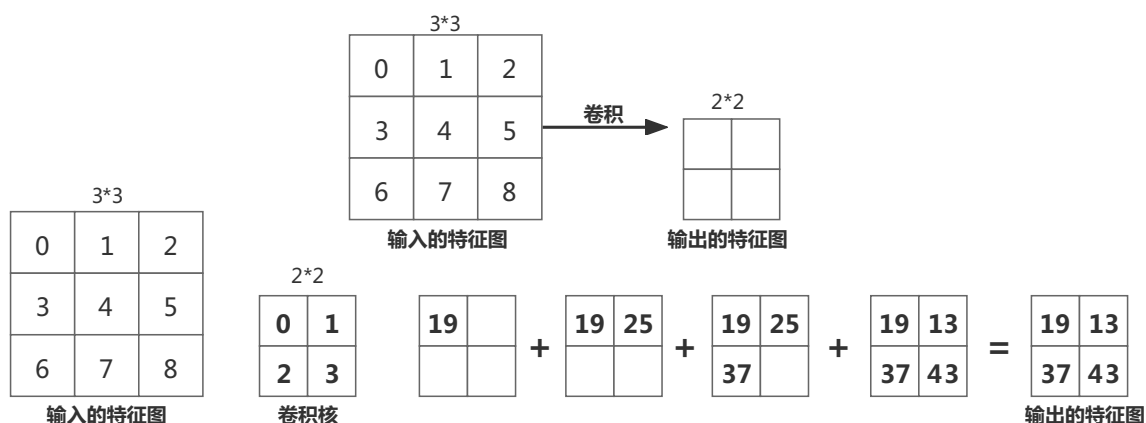


图 2-2 卷积操作的计算流程

激活函数对于整个卷积神经网络来说扮演者至关重要的地位。从公式 2-1 可以看出，整个卷积运算都是一个线性的操作。显然，在实际应用中，非线性的特征和数据分布时常存在，如果不引入非线性的操作，对于整个卷积神经网络的学习来说是十分受限的。为此，激活函数常常被引入，以提供非线性的能力。则整个卷积操作可以写为

$$y_k = \sum_{n=0}^{N-1} a_{k,n}x_n + b_k = f(a_k^T \mathbf{x} + b_k), \quad k = 0, 1, \dots, K-1 \quad (2-2)$$

其中  $f(\cdot)$  代表激活函数，对于激活函数的选择各不相同。同时，不同激活函数也具有不同的功能。如图 2-2 所示，为线性整流函数（Rectified Linear Unit, ReLU）

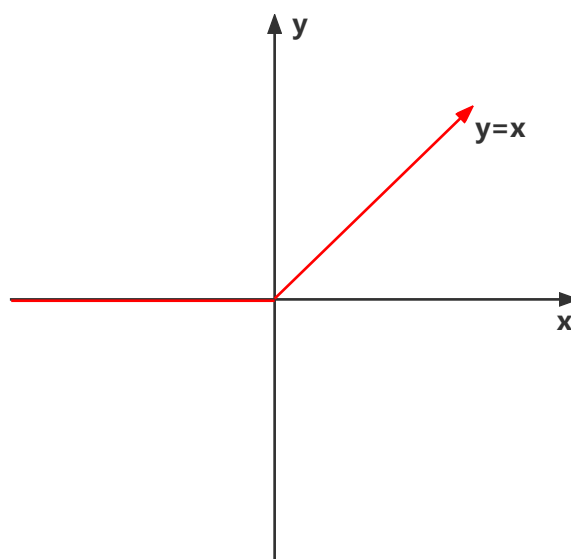


图 2-3 线性整流函数

[27] 的函数图, 可以发现 ReLU 是在负向区域抑制的。同时，对于正向激活，ReLU

进行了能量的全部保留。由于 ReLU 可以有效地缓解梯度消失现象和计算高效性，ReLU 被广泛的应用于图像相关的各种模型（如目标检测 [28,29]、低阶任务 [30,31]）。对于基于卷积神经网络的低剂量 CT 图像降噪模型来说，由于的图像常常都是归一化到 0 ~ 1 之间的。同时，考虑到计算效率来说，现有的大多数模型也采用 ReLU 函数作为激活函数，并取得非常优异的成绩。因此，整个卷积层可以被写为

$$y_k = \max(\mathbf{a}_k^T \mathbf{x} + b_k, 0), \quad k = 0, 1, \dots, K - 1 \quad (2-3)$$

图 2-3 为线性整流函数的基本形状。可以发现这条曲线是负向抑制的。

### 3. 转置卷积操作

卷积操作可以进行特征提取，如果不进行边缘的填充，特征图将会变小。然而，转置卷积 [32] 操作可以看作是卷积操作的逆操作，通常可以扩大特征图。这种转置卷积操作对于编解-解码的网络结构是十分有必要的。因为通过连续的编码操作，输入图像的特征图将会不断减小。为了不断的恢复特征图的大小，反卷积操作可以实现这个目的。

具体来说，转置卷积用于使用一些可学习参数将输入特征映射上采样到期望的输出特征映射。通过一个简单的转置卷积的运算来描述。考虑一个 2x2 编码的特征图，它需要上采样到一个 3x3 的特征图，如图 2-4 所示。

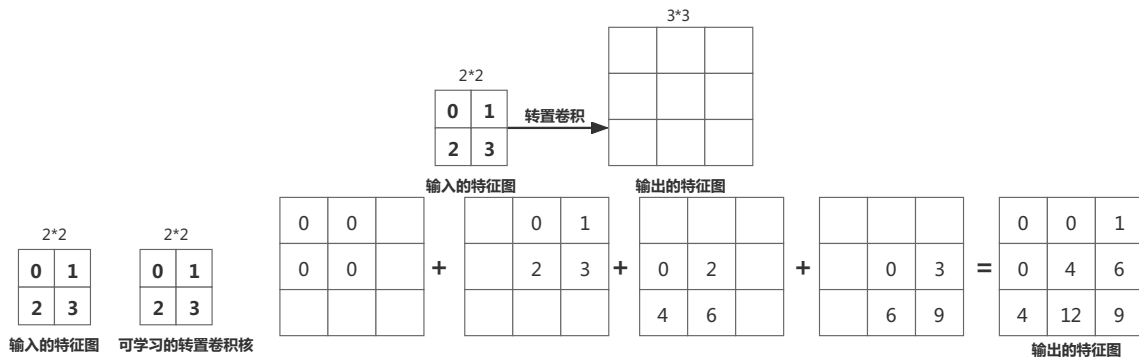


图 2-4 转置卷积的计算流程

如图 2-4 所示，对于一个  $2 \times 2$  的特征图来说，想要通过转置卷积得到  $3 \times 3$  的上采样的特征图。定义一个  $2 \times 2$  的转置卷积核，进行如同卷积一样的滑动点积的操作。特征图上的 0 和转置卷积核做点积得到的 4 个值分别为 0, 0, 0, 0。这个得到的 4 个值就为求得的输出特征图的左上角的 4 个位置的值。通过进行相同的操作，把这个转置卷积扩展到整个输入特征图（如图所示）。最后，将四个得到的子特征图进行加法操作，就可以得到最后的输出的特征图。需要注意点是由于在输出上的步幅等于在输入上的步幅，所以转置卷积也被称为分数步幅卷积。例如，输出上的步幅为 2，则输入上的步幅为 1/2。

同时，需要注意的是，转置卷积也并不是没有缺点的。可以看到第 2 个得到的输出的特征图和第三个输出的特征图存在值的重叠部分。如果这种重叠扩展到比较大范围的特征图，会产生一种叫做“棋盘伪影”的问题，这种产生的“棋盘伪影”现象会产生一些网格状的输出，导致不佳的图像可视化效果。

#### 4. 编码-解码的神经网络结构

编码-解码的神经网络结构<sup>[23,33]</sup>被广泛应用于图像的恢复<sup>[33,34]</sup>、超分辨率<sup>[35,36]</sup>、分割<sup>[4,37]</sup>等任务。编码-解码的神经网络结构的编码部分是不断的进行下采样。具体来说，可以通过不断的卷积操作进行特征图大小的减小，每一层的信息会更加的抽象和高阶，所以称为输入信息的“编码”，意为对输入信息通过神经网络的推理进行信息的浓缩。这样可以保留信息最重要的部分，诸如图像恢复任务都需要保留大范围的语义内容区域，同时抛弃掉不重要的噪声或伪影的干扰。这些大范围的语义区域毫无疑问，是更容易在编码过程中进行保留的。所以编码过程是十分有效地保留关注的语义内容区域，同时不编码一些不重要地干扰的。相反，解码部分，可以实现一个上采样的过程。具体来说，使用在之前介绍的转置卷积的操作，不断地提高特征图的大小，相当于从编码的信息中逐步地恢复信息，所以这一部分叫做解码的过程。

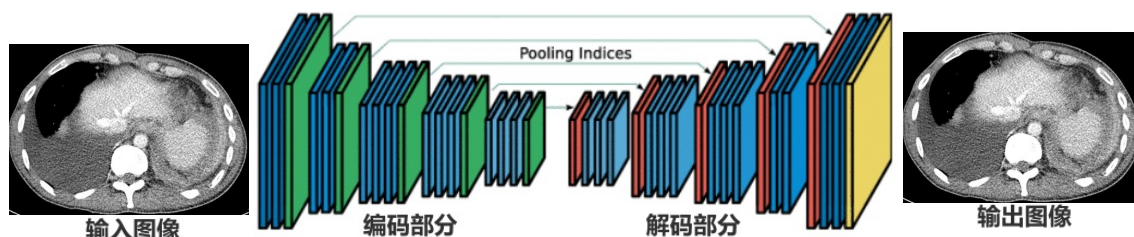


图 2-5 编码-解码神经网络结构的计算流程

编码-解码的神经网络结构如图 2-5 所示，可以发现，在左边部分，特征图不断地进行卷积编码，特征图的大小不断缩小。当达到最小的瓶颈层的时候，信息被完全的编码，可以认为这一部分的信息被完全的编码表示。当进行解码的时候，如图 2-5 的右边部分，特征图使用转置卷积进行特征图的上采样，这一部分被称为解码的部分，信息不断的从编码的信息解码出来。根据已有的文献可以发现，现在的低剂量 CT 降噪方法大多都采用了编码-解码的神经网络结构，这种结构十分的高效。在整个编码-解码的神经网络结构的过程中，可以在编码和解码层还可以插入一些跳跃连接和信息的交互，这样可以在解码过程中不断使用编码过程中的低阶的表示，这样十分有益于最后的学习。

#### 5. 基于卷积神经网络的编码-解码结构的低剂量 CT 降噪模型



在此，已经介绍完了基于卷积神经网络的编码-解码结构的低剂量 CT 降噪模型所有的组件。现在开始介绍主流的基于卷积神经网络的编码-解码结构的低剂量 CT 降噪模型的架构。在此，以 RED-CNN<sup>[17]</sup> 做为一个例子，如图 2-6 所示，编码分支采用卷积操作和 ReLU 进行不断交替，这里的编码部分，卷积层是不断地做下采样的，所以他们的特征图在不断地减小。同时，对于编码来说，RED-CNN 采用的是转置卷积和 ReLU 的结合，不断地进行上采样进行特征图的大小的提升。同时，在编码和解码过程之间，也进行了拼接的操作，使在编码阶段的低阶信息表示在后端的解码过程也进行利用，实验证明这样有利用噪声的有效学习。最后，RED-CNN 采用了一个残差学习的思路，即在最后的阶段将输入图像跳跃连接到最后的输出层之前，这样可以有助于进行噪声的学习。

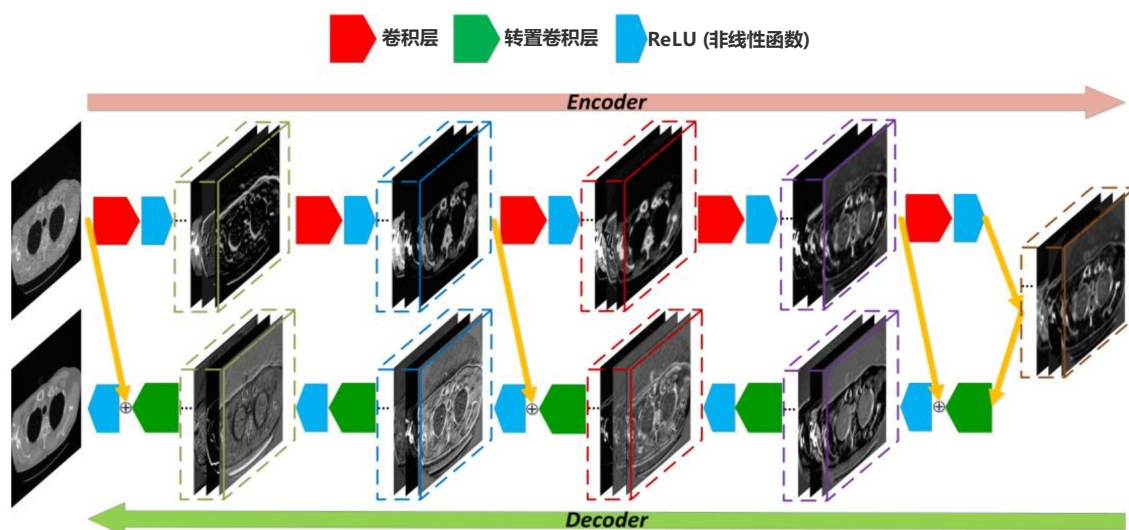


图 2-6 RED-CNN 编码-解码神经网络结构的计算流程<sup>[17]</sup>

可以发现，RED-CNN 的结构是十分简单和高效的。因此，本文在第四章和第五章都使用了 RED-CNN 作为降噪器进行噪声的学习。

## 2.2.2 基于生成对抗网络的低剂量 CT 图像降噪模型

### 1. 生成对抗网络的基本架构

生成对抗网络<sup>[38,39]</sup> 包括两部分，即生成器和判别器。生成器的输入可以是任意分布的噪声（一般设置为  $0 \sim 1$  均匀分布的噪声），通过生成器这个噪声能够产生一些伪样本。随后，判别器尝试去判断真实样本和生成器产生的伪样本的真假，并产生损失来提高自己的判别能力。生成器和判别器可以通过多层全联接网络或者深度卷积神经网络进行搭建。整个生成对抗网络可以看作是一个最大最小游戏，



所以损失函数可以定义为：

$$\begin{aligned} \min_{G_\phi} \max_{D_\theta} L(G_\phi, D_\theta) &= E_{s \sim p_r} [\log(D_\theta(\mathbf{s}))] \\ &+ E_{x \sim p_n} [\log(1 - D_\theta(\mathbf{m}))], \end{aligned} \quad (2-4)$$

其中  $E[\cdot]$  表示给定输入的期望值， $D_\theta$  是一个参数化的网络作为鉴别器。更具体地说，希望所学到的样本分布  $p_\phi(z|q)$  与真实的样本分布  $p(\varepsilon)$  相似。生成器寻求去映射潜在的噪声  $\mathbf{z} \sim p_z(\mathbf{z})$  到真实的样本分布（即  $G(\mathbf{z}; \theta_g)$ ， $\theta_g$  代表生成器的参数）。

同时，判别器被训练去判别是真实样本还是产生的伪样本（可以表示为  $D(\mathbf{x}; \theta_d$ ，其中， $\theta_d$  代表判别器的参数）。一般通过判别器的输出二进制的表示来代表样本的真和伪。具体来说，通过判别器输出 0 代表输入是一个伪样本，通过判别器输出 1 代表输入是一个真实样本。生成器和判别器不断地交替更新，提升自己的能力。因此，判别器的损失可以写为：

$$\begin{aligned} L(\theta_d) &= E_{s \sim p_r} [\log(D(\mathbf{x}; \theta_d))] \\ &+ E_{x \sim p_n} [\log(1 - D(\mathbf{z}; \theta_g))], \end{aligned} \quad (2-5)$$

需要注意的是，判别器进行更新的时候，生成器并不会进行参数更新。相反，在判别器进行判别的时候，生成器的参数是相对固定的。同样地，生成器的损失函数可以写为

$$L(\theta_g) = E_{x \sim p_n} [\log(1 - D(\mathbf{z}; \theta_g))], \quad (2-6)$$

有理论可以证明，当  $D(\mathbf{z}; \theta_g) = 0.5$  时候模型趋于收敛。这意味着判别器不能再区分真实样本和伪样本的真假。一旦整个模型收敛，生成器能够产生一个符合真实样本分布的合成的样本，以便于判别器不能够再区分这个产生的样本是真还是假。生成对抗网络的算法流程如图所示。

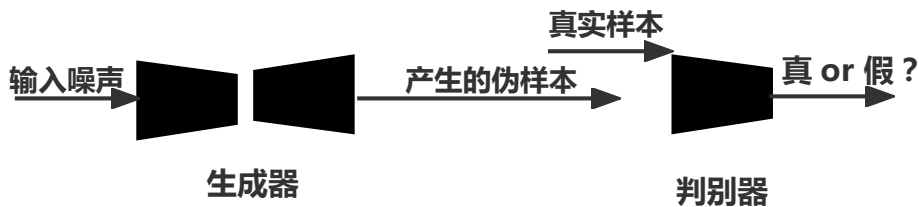


图 2-7 生产对抗网络模型的计算流程

然而，在现有的基于生成对抗网络的低剂量 CT 图像降噪模型并没有采用上面描述的最基本的生成对抗网络的工作模型。相反，这些方法通常对于生成器的输入并不是一个随机的噪声，而是直接是低剂量的 CT 图像。所以这可以看作是

一种变体的生成对抗网络<sup>[40]</sup>：生成器学习在低剂量 CT 图像的输入条件下到一个正常剂量 CT 图像的映射。因此，损失函数可以写为：

$$\begin{aligned} \min_G \max_D L(D, G) &= E_{x \sim p_{data}(x)} [\log D(x | y)] \\ &+ E_{z \sim p_z(z)} [\log(1 - D(G(z) | y))] \end{aligned} \quad (2-7)$$

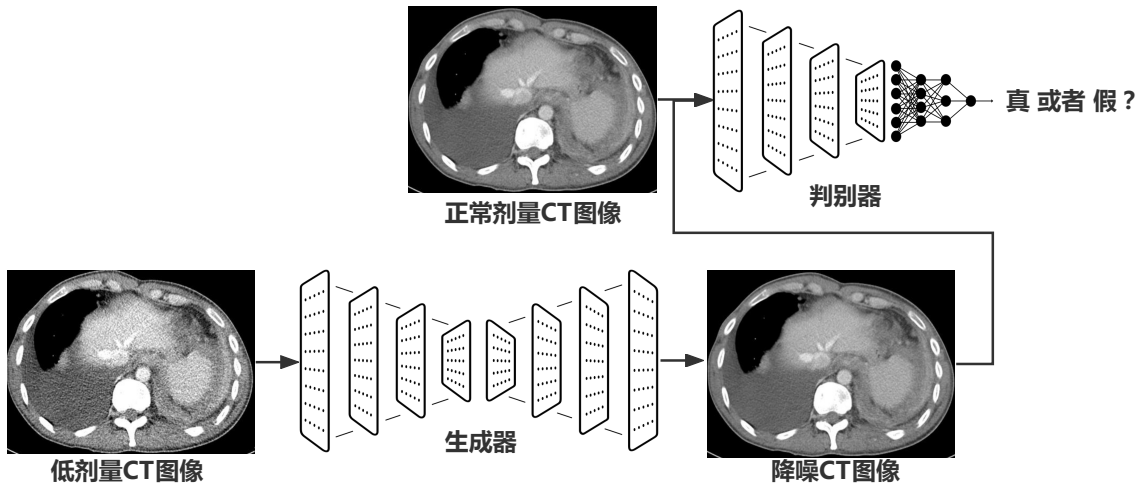


图 2-8 基于生成对抗网络的低剂量 CT 图像降噪方法架构

由式 2-7 可以发现，生成器的输入是低剂量的 CT 图像。同时，基础的生成对抗网络经常遭受梯度消失和训练不稳定的现象。因此，现有基于生成对抗网络的方法通常通常引入 Wasserstein 距离<sup>[41]</sup> 到损失函数中来提升训练的稳定性。更为重要的是，梯度如果容易消失很难获得想要的降噪效果。所以，现有的基于生成对抗网络的低剂量 CT 图像降噪方法通常会在 Wasserstein 距离的基础上，引入梯度惩罚来更进一步来改善整个网络的梯度流。因此，整个生成对抗网络被叫做基于 Wasserstein 距离和梯度惩罚的生成对抗网络<sup>[42]</sup>。整个损失函数可以写为：

$$\begin{aligned} L(F, G) &= \sup_{\|F\|_L \leq 1} E_{x \sim p_{data}(x)} [F(x)] - E_{z \sim p_z(z)} [F(G(z))] \\ &+ \lambda E_{\hat{x} \sim p_{\hat{x}}(\hat{x})} [(\|\nabla_{\hat{x}} F(\hat{x})\|_2 - 1)^2] \\ \hat{x} &= \rho \tilde{x} + (1 - \rho)x \quad \text{和} \quad 0 \leq \rho \leq 1. \end{aligned} \quad (2-8)$$

最终，如图 2-8 所示，可以得到现有广泛采用的基于生成对抗网络的低剂量 CT 图像降噪模型的框架图。具体来说，生成器的输入为低剂量 CT 图像，经过生成器产生一个伪样本（即为降噪后的 CT 图像），这个样本经过判别器和真实样本

(即正常剂量的 CT 图像) 进行比较。生成器和判别器相互对抗, 生成器在损失函数的指导下提高自己的降噪能力 (产生更好地降噪的图像)。同时, 判别器也不断提高自己的判别能力, 希望自己能够较好地分辨是正常剂量的 CT 图像还是产生的伪样本 (降噪后的图像)。

需要指出的是, 在第三章和第四章都使用了基于生成对抗网络的低剂量 CT 图像的降噪方法, 因此该部分是的核心基础部分。

## 第三章 基于噪声学习的无监督低剂量 CT 图像降噪

### 3.1 问题定义

现有的基于深度学习的低剂量 CT 图像降噪方法通常会在基于仿真噪声的低/正常剂量配对 CT 图像数据集上进行训练（例如：NIH-AAPM-Mayo 数据集），并希望该训练模型能够在特定的临床低剂量 CT 图像上进行降噪。本质上来说，这种工作方式假设不同的低剂量 CT 图像的噪声特点（或分布）是十分相似的。为了检验这个假设，本文通过噪声提取方法对不同低剂量数据集的噪声进行提取。如图 3-1 中所示，是从不同数据集中提取的噪声子块的结果，可以观察到估计噪声的特征之间存在有明显的差异。这种现象可能是由于不同的成像协议、设备的供应商或 CT 图像的采集环境造成的。因此，现有的基于深度学习的低剂量 CT 图像降噪方法由于不同使用场景的噪声特性的差异，低剂量 CT 图像降噪模块可能在临床上会遭受效果不佳地困扰。

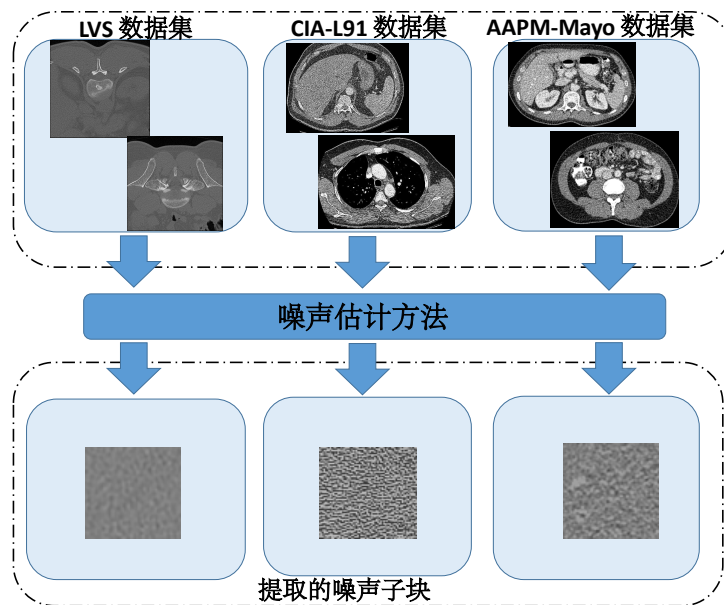


图 3-1 提取的不同噪声子块的示例，分别来自 LVS、CIA-L91 和 AAPM-Mayo 数据集。对于噪声子块的估计，采用最新的噪声估计方法<sup>[43]</sup>进行噪声提取

### 3.2 算法动机及其概述

如果想解决上述问题，一个直观的想法是收集大量多样的配对的低/正常剂量 CT 数据集，这些配对数据必须是结构信息完全匹配的。基于这些严苛的条件，降

噪模型是可以进行比较好的训练的。同时，得到的模型具有很强的泛化性能力。不幸的是，由于医学伦理（不能对同一病人施加多次 CT 检查），配对的低/正常剂量的数据集在临床中是十分难以获取的。因此，目前大多数的基于深度学习的低剂量 CT 图像降噪方法只能在仿真的成对数据集上进行训练。更糟糕的是，在大多数情况下，临床上常常只有低剂量 CT 图像仅可获取，因此需要面对无监督的低剂量 CT 图像降噪问题。

为解决上述问题，创新性的提出一种基于噪声学习的无监督低剂量 CT 降噪算法。总的来说，通过构建一个适用于特定任务的配对的从临床场景低剂量 CT 图像诱导的低/正常剂量 CT 数据集。因此，无监督问题可以转换为监督问题。降噪模型可以在这个构建的配对数据集上训练。由于该数据集中的低剂量 CT 图像是从需要降噪的低剂量 CT 图像中诱导的，因此该模型自然的在临床场景的低剂量 CT 图像数据集上会有一个较好地表现。具体来说，首先估计低剂量 CT 图像中出现的噪声以提取噪声子块。然后，进一步利用变分自动编码器 (Variational Autoencoder, VAE) 和生成对抗网络 (Generative Adversarial Network, GAN) 两种模型来增加噪声样本并提高多样性。受益于医学图像高阶任务（例如，CT 图像实例分割、感兴趣区域的 CT 图像检测），大量的开源正常剂量 CT 图像数据集可以很容易获得。因此，提出将学习到的噪声样本添加到这个获取到的正常剂量 CT 图像中。通过这样，结构信息完全匹配的低/正常剂量 CT 数据集可以构建。提出的方法不仅享有数据驱动的深度神经网络方法的强大能力，同时也旨在解决由不同数据集之间噪声特征差异。

### 3.3 基于噪声学习的无监督低剂量 CT 图像降噪算法

#### 3.3.1 数学变量声明

给定一批没有正常剂量 CT 图像参考的低剂量 CT 图像图像，表示为  $X = \{x_1, x_2, \dots, x_N\}$ 。假设有丰富的正常剂量 CT 图像可以从开源网站中轻松地获取到（例如，Cancer Imaging Archivel<sup>①</sup>），表示为  $Y = \{y_1, y_2, \dots, y_M\}$ 。的目标是从低剂量 CT 图像  $x$  中移除噪声以获得潜在的相应的正常剂量 CT 图像  $h$ 。

#### 3.3.2 基于变分自编码器的噪声提取与学习

首先做出以下假设：

**假设 1.** 低剂量 CT 图像  $x$  可以用潜在的相对应的正常剂量 CT 图像  $h$  和加性

<sup>①</sup> <https://www.cancerimagingarchive.net/collections-list>

无关且均值为零的噪声  $\varepsilon$  表示, 即

$$x = h + \varepsilon. \quad (3-1)$$

请注意, 假设 1 是计算上易于处理的假设。在 CT 图像降噪领域, 同样的假设广泛存在且也是比较合理的。同时, 该假设广泛存在于自然图像降噪中<sup>[7]</sup>。

基于这个假设, 可以采用一种零均值独立噪声的估计方法<sup>[44]</sup>来提取低剂量 CT 图像中估计的噪声  $\hat{\varepsilon}$ 。该方法旨在寻找一些低剂量 CT 图像中的平滑的区域, 该区域必须满足两个约束条件 (图像块的均值和方差变化小), 记为  $S = \{s_1, s_2, \dots, s_t\}$ 。通过使用零均值独立噪声的估计方法进行平滑区域的搜索, 估计的噪声块可以表示为

$$\hat{\varepsilon}_i = s_i - \text{Mean}(s_i), i = 1, 2, \dots, t, \quad (3-2)$$

其中  $\hat{\varepsilon}_i$  和  $\text{Mean}(\cdot)$  表示估计噪声的子块形式和均值算子。估计的噪声子块集合可以表示为  $W = \{w_1, w_2, \dots, w_t\}$ 。然后, 根据**假设 1**, 可以构建成对的伪低剂量 CT 图像/正常剂量 CT 图像, 即通过收集的正常剂量 CT 图像  $Y = \{y_1, y_2, \dots, y_M\}$  和估计噪声子块  $W = \{w_1, w_2, \dots, w_t\}$ 。需要注意的是正常剂量 CT 图像也需要划分为同等大小的子块形式。

通过的初步实验, 可以观察到估计的噪声块的数量不是很大, 这会导致构造的成对数据集的数量是十分有限的。为了增加噪声样本量, 创新性的提出使用生成式深度学习模型进行噪声样本的扩充和学习。

具体来说, 本文提出使用变分自动编码器<sup>[4]</sup> (Variational Autoencoder, VAE) 来学习估计噪声的潜在分布, 并进一步用于生成一系列随机地输出样本。对于学习模型来说, 学习的潜在分布可以表示为  $p_\varphi(z|w)$ , 其中的学习通过参数化变分编码网络  $F_\varphi$  实现 (即通常使用两个独立的网络实现)。然后, 潜在变量  $z$  通过重新参数化的网络从  $p(z|w)$  中使用采样技巧得到<sup>[2]</sup>。随机输出样本  $W' = \{w'_1, w'_2, \dots, w'_t\}$  可以通过似然分布  $q_\varphi(w'|z, w)$  获得, 即由解码网络  $T_\varphi$  进行参数化。

VAE 的优化目标可以表示为:

$$L(\varphi, \Phi; w) = L_{KL} + L_{rec}, \quad (3-3)$$

第一项是 Kullback-Leibler (KL) 损失<sup>[37]</sup>, 它将推动用于对齐先验分布的潜在分布, 即  $N(0, 1)$ 。这一项对于 VAE 来说非常重要, 以确保输出具有多样性。这一项又可

以具体写为

$$L_{KL} = D_{KL}(p_{\Phi}(z|w) \| N(0, 1)). \quad (3-4)$$

第二项是重建损失保证输出  $w'$  更接近输入  $w$ :

$$\mathbb{E}_{z \sim p_{\Phi}(z|w)}(\log q_{\Phi}(w'|z, w)). \quad (3-5)$$

这个方案不仅增加了噪声样本，而且提高了样本的多样性。VAE 模型是无监督的训练模型，这对于整个架构是十分方便的。最后，构造的配对数据集可以正式写为  $(X', Y)$ ，其中  $X' = \{x'_1, x'_2, \dots, x'_l\}$ 。

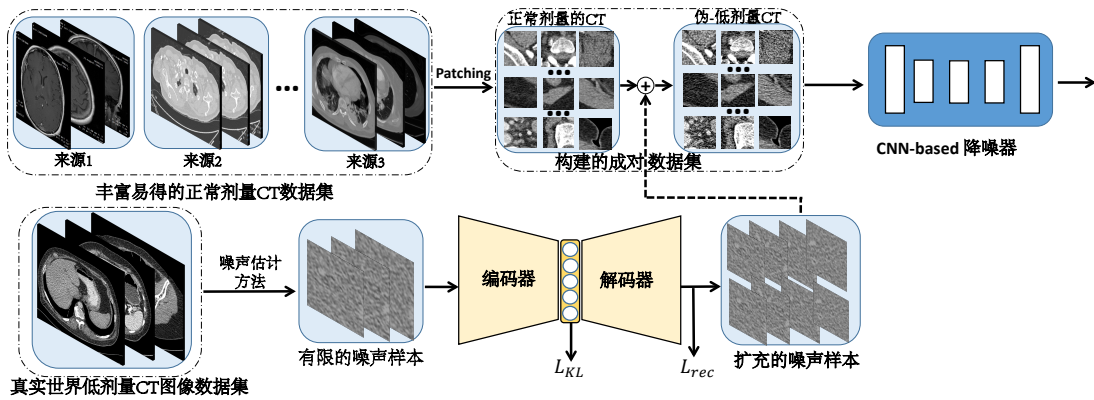


图 3-2 基于变分自编码器的噪声学习无监督低剂量 CT 图像降噪算法整体架构。在整个算法中，变分自编码器用于噪声样本的提取和学习。本算法利用正常剂量的 CT 图像和提取的噪声进行配对样本的构建用于后端的降噪学习，其中 Patching 操作代表将图像随机划分为子块表示

基于获得的低剂量 CT 图像诱导的配对数据集，一种广泛采用的低剂量 CT 图像降噪思路是学习一个降噪网络  $D: x \rightarrow y$ ，进行两个域  $X'$  和  $Y$  之间的映射，其中配对的训练样本表示为  $(x', y) \in (X', Y)$ <sup>[21]</sup>。最近，许多自然图像降噪模型被提出，它们通过噪声建模的方式实现了具有竞争力的降噪表现，即替换映射

$$\{D: x' \rightarrow y\} \rightarrow \{D: x' \rightarrow (x' - y)\} \quad (3-6)$$

从式 3-6 可以直观地发现，降噪器网络可以更专注于捕捉出现的在低质量的图像中噪声。相似的噪声建模策略在本文中也采用，其中潜在的正常剂量 CT 图像表示为  $\hat{y} = x' - D(x')$ ，其中  $D(x')$  可以看作是估计的噪声。

对于降噪器的选择，如先前的研究所采用的<sup>[28,31]</sup>，深度卷积神经网络被作为的降噪器，记为作为  $D_{\theta}$ 。需要注意的是，本节的工作主要集中在构建成对的诱导

低剂量 CT 图像/正常剂量 CT 图像，而不是设计的降噪器。因此，考虑到模型的性能和便利性，已有的基于深度卷积神经网络的降噪器 CPCE-2D<sup>[30]</sup> 主干网络被采用。类似于一般的低剂量 CT 图像降噪问题，使用均方误差（MSE）作为损失函数来指导模型的优化过程，定义为：

$$\text{loss}(\hat{y}, y) = L(\Theta) = \frac{1}{2M} \sum_{i=1}^M \|(x' - D(x'; \theta) - y)\|_2^2 \quad (3-7)$$

其中  $(x'_i, y_i)_{i=1}^M$  表示  $M$  个低/正常剂量 CT 图像对。

提出的基于变分自编码器的噪声学习无监督低剂量 CT 图像降噪算法整体架构如图 3-2 所示。总的来说，提出的低剂量 CT 图像降噪框架的优势有两个方面：

- 与现有的深度学习方法相比，提出算法中的伪低剂量 CT 图像/正常剂量 CT 训练样本直接从需要降噪的低剂量 CT 图像中诱导，而不是使用特定的已知噪声模拟配对数据集。
- 丰富的正常剂量 CT 图像数据集被利用去构造成对的数据集，这是通常被先前的研究所忽略的。

### 3.3.3 基于生成对抗网络的噪声提取与学习

虽然已知低剂量 CT 图像理论上存在泊松噪声，研究发现高斯电子噪声也存在<sup>[45]</sup>于低剂量 CT 图像的噪声中。在临床场景中，低剂量 CT 图像的噪声分布会更加复杂。以前的工作经常将这个问题简化为同一种未知的零均值噪声<sup>[46]</sup>以达到包含范围广泛的复杂噪声。因此，为了构建用于训练目的的低剂量/正常剂量 CT 图像数据对，首先采用通过遵噪声特性来近似表示低剂量 CT 图像中的噪声信息。

和 3.3.2 小节相似，为了提取可靠的噪声信息，需要相对平滑的图像块（即内部结构非常相似的区域），即其中背景的结构化信息较弱，噪声信息占主导地位的区域。对于自然图像，可以通过测量候选子块与其亚子块之间的均值和方差的相似性来获得平滑子块，这种方法也是 3.2.2 小节所采用的约束条件。与之不同的是，本文提出融入医学图像的自身特性进行噪声的估计。具体来说，相比自然图像，低剂量 CT 图像的大多数区域自然地具有相似的平均值（除了高组织密度的区域<sup>[47]</sup>）。

因此，本文创新性的提出利用低剂量 CT 图像的这一固有特性和基本的噪声提取方法从相对平滑的低剂量 CT 图像块中提取噪声。在这种情况下，给定一个低剂量 CT 图像子块  $\mathbf{H}_k$ ，以及它的第  $i$  个亚子块  $\mathbf{h}_{k,i}$ ，其中  $\mathbf{h}_i$  是  $\mathbf{H}_k$  上的一个子区域。在提出的算法中，规定低剂量 CT 图像的平滑子块必须满足给定的下述约束条件

$$|\text{Variance}(\mathbf{h}_{k,i}) - \text{Variance}(\mathbf{H}_k)| \leq \beta \cdot \text{Variance}(\mathbf{H}_k), \quad (3-8)$$



其中  $Variance(\cdot)$  表示“方差”运算符。 $\beta$  是预定义的超参数，范围从零到一。如果对于每个图像子块都满足约束，则  $\mathbf{H}_k$  将被视为平滑子块。

将满足约束的平滑块分组为  $\mathbb{H}$ ，并将其表示为  $\mathbb{H} = \{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_K\}$ 。在噪声分布的期望为零的假设下，平滑块  $\mathbf{H}_j$  的噪声信息可以近似为

$$\mathbf{N}_j = \mathbf{H}_j - \text{Mean}(\mathbf{H}_j), \quad j = 1, 2, \dots, K, \quad (3-9)$$

其中  $\text{Mean}(\cdot)$  表示“均值”运算符。噪声块可以进一步分组为  $\mathbb{N} = \{\mathbf{N}_1, \mathbf{N}_2, \dots, \mathbf{N}_K\}$ 。此外，临床场景中的 CT 图像的成像区域具有两个固有特性，即大面积的中低组织密度和较高的分辨率。为了获得大量符合条件的平滑区域，相对于上一章使用的变分自编码器，提出通过生成对抗网络 (Generative Adversarial Network, GAN) 来生成足够的噪声块用于后端的降噪。

尽管低剂量 CT 图像可以生成相对大量的噪声块，但在临床场景中可用的低剂量 CT 图像的数量是有限的。此外，噪声样本的多样性也会影响数据驱动方法的性能。为此，本文创新性的提出利用 GAN 对低剂量 CT 图像噪声模式的分布进行建模，因为它可以学习目标样本上的复杂分布并输出接近目标样本<sup>[48]</sup>的分布，其中目标样本是  $\mathbb{N}$  的集合。对于所提出的算法，采用 WGAN-GP<sup>[49]</sup> 模型进行噪声建模，其目标可以写为

$$\begin{aligned} \mathcal{L}_{GAN} = & \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [D(\tilde{x})] - \mathbb{E}_{n \sim \mathbb{P}_n} [D(n)] \\ & + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|^2 - 1)] \end{aligned} \quad (3-10)$$

其中  $\mathbb{P}_g$  是生成器的分布，它将一个高斯分布  $\mathcal{N} \sim (0, 0.5)$  作为 GAN 的输入， $\mathbb{P}_n$  是提取的噪声子块集合之间的分布，它满足规定的约束条件。 $\mathbb{P}_{\hat{x}}$  被定义为在从  $\mathbb{P}_g$  和  $\mathbb{P}_g$  采样的点对之间沿直线均匀采样的分布。

在继续后续的降噪任务之前，可以获得足够的噪声块，并利用额外的正常剂量 CT 图像来进一步使用它们来构建训练数据对。更具体地说，通过将干净的正常剂量 CT 图像表示为  $\mathbf{I}$ ，将通过 GAN 生成的随机噪声模式表示为  $\mathbf{N}$ ，可以获得训练样本对  $\{(\mathbf{I} + \mathbf{N}, \mathbf{N})\}$ 。

## 3.4 实验结果及其分析

### 3.4.1 基线方法和评估指标

考虑采用数据驱动的降噪模型，包括 DnCNN<sup>[50]</sup>、FFDNet<sup>[51]</sup>、MAP-NN 和 Red-CNN 进行比较。特别指出的是 DnCNN、FFDNet 是用于自然图像降噪的盲降噪方法。BM3D<sup>[49]</sup>、KSVD<sup>[52]</sup> 和 TV-L1<sup>[53]</sup> 是用于低剂量 CT 图像降噪的方法。

考虑使用基于图像质量评估的各种指标进行评估。由于有可用于合成低剂量 CT 图像的正常剂量 CT 图像，因此通过采用峰值信噪比 (Peak Signal to Noise Ratio, PSNR)、结构化相似性指数测量 (Structural Similarity Index Measure, SSIM) 和均方根误差 (Root-Mean-Square Error, RMSE) 来作为评估的基准方法。对于临床场景中的低剂量 CT 图像，由于没有可用的对应的干净的 CT 图像，采用一种基于无参考的图像质量评估指标对比噪声比 (Contrast Noise Ratio, CNR)<sup>[29]</sup> 进行评估。

### 3.4.2 数据集

**基于生成对抗网络的噪声学习方式:** 由于没有可用于临床场景的低剂量 CT 图像对应的干净图像，因此使用 LVS CT 图像数据集<sup>[48]</sup>，其中包括 7850 个临床场景低剂量腰椎 CT 图像，大小为  $1024 \times 1024$ 。需要注意的是，他们都没有对应的正常剂量的 CT 图像。一些 LVS 数据集的样本如图 3-3 所示。使用来自 FUMPE<sup>①</sup>数据集作为正常剂量 CT 图像样本的来源，该数据集由 35 位不同受试者的肺栓塞的正常剂量 CT 图像组成，其中包括 5000 张正常剂量图像大小为  $512 \times 512$ 。

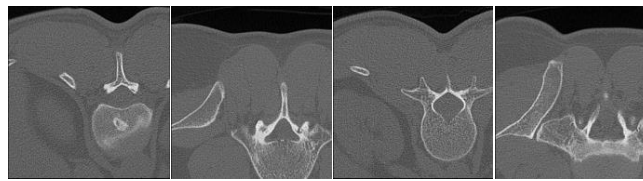


图 3-3 数据集示例

**基于变分自编码器的噪声学习方式:** 低剂量 CT 图像数据集<sup>[23]</sup> 用于评估降噪性能。该数据集包括低/正常剂量的 CT 图像对，但相比 NIH-AAPM-Mayo 数据集具有不同的成像协议。对于本文的实验设计，只利用低剂量 CT 图像来构建配对数据集，其中的正常剂量的 CT 图像仅用于评估定量和可视化结果的对比。为了构建诱导的低剂量 CT 图像/正常剂量 CT 图像对，额外的正常剂量 CT 数据集是需要。因此，采用 LIDC-IDRI 数据集<sup>[3]</sup> 用于诊断和肺癌筛查胸部 CT 扫描标记的注释病灶。最后，32k 个  $64 \times 64$  结构信息完全匹配的低/正常剂量 CT 图像数据集可以得到。该数据集随机分为 90% 的训练集，10% 的验证集。

### 3.4.3 实验细节

关于低剂量 CT 图像噪声提取的参数，图像子块的大小、亚子块的大小和  $\beta$  分别设置为 64、32 和 0.1。对于降噪部分，遵循已有的低剂量 CT 图像降噪方法中提出的网络结构。采用现有的噪声提取和学习方法将搜索窗口大小设置为  $56 \times 56$ ,

① <https://figshare.com/collections/FUMPE/4107803>

$K = 16$ 。采用 Adam 算法进行优化，学习率从  $10^{-4}$  开始，然后在训练误差变化较小时固定为  $10^{-5}$ 。

### 3.4.4 基于变分自编码的噪声提取与学习的实验结果与分析

低剂量 CT 图像数据集中的 L209 病例被用于比较不同方法的降噪性能。如图 3-4 所示，可以观察到所提出方法和 MAP-NN、FFDNet 的可视化比较结果。总的来说，所提出的方法实现了相对更好的噪声抑制。通过放大 MAP-NN 的结果，一些明显的噪声点和白色伪影（在正常剂量 CT 中不存在）可以被观察到。为了进

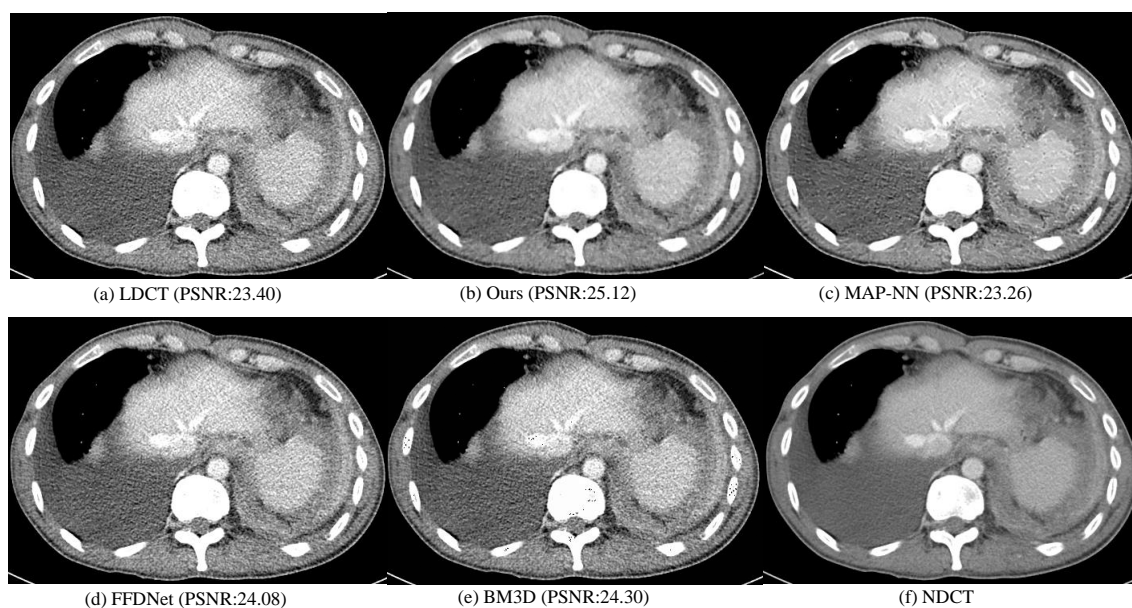


图 3-4 可视化对比结果

一步评估细节的表现，放大图 3-4 中的感兴趣区域 (Region of Interest, ROI) 的结果被显示在图 3-5 中。如绿色箭头所示，MAP-NN 的明显噪声点可以被明显地观察到。与 MAP-NN 相比，提出的方法具有更自然的纹理和更强的噪声压制表现。这些差异可能源于两个方面：一是噪声特性的区别存在于低剂量 CT 图像数据集和 NIH-AAPM-Mayo 数据集（用于训练 MAP-NN）中，这可能会导致 MAP-NN 处理出现的噪声的泛化能力较弱于低剂量 CT 图像数据集。此外，MAP-NN 的性能退化也可能导致白色伪影。第二，提出的方法对低剂量 CT 图像的噪声更敏感，因为基于深度卷积神经网络的降噪器直接作用在低剂量 CT 图像诱导的配对数据集上。从细节和整体的角度来看，提出的方法达到最佳的降噪效果。这也获得了如表 3-1 所示的最具竞争力的 PSNR 和 SSIM 表现。

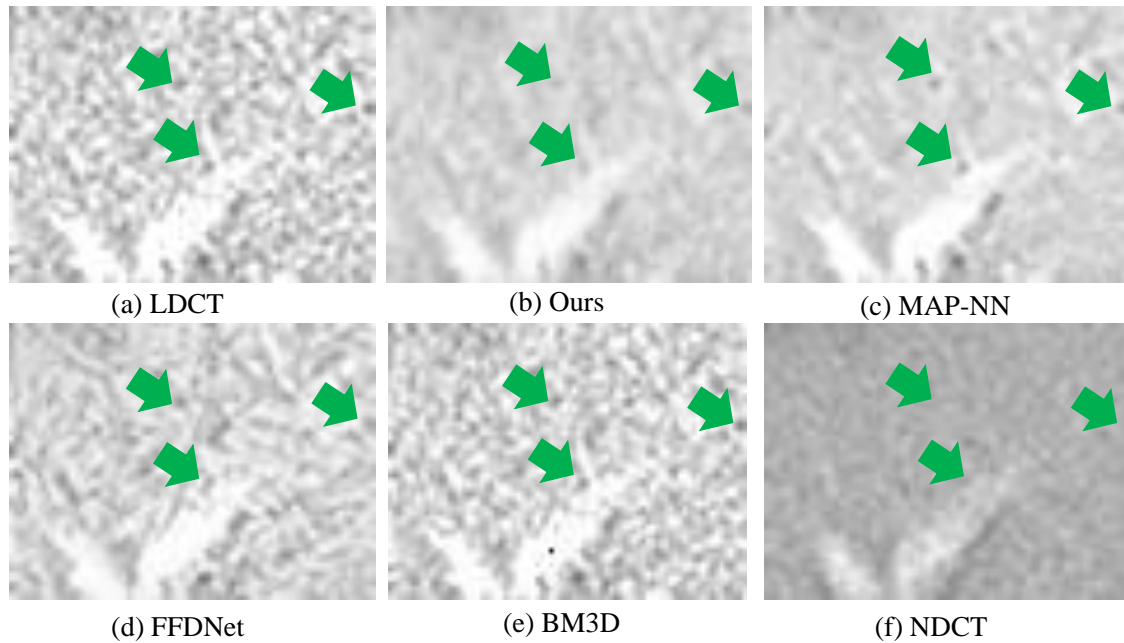


图 3-5 可视化对比结果

表 3-1 不同对比方法的峰值信噪比 (PSNR)、结构相似度 (SSIM) 和均方根误差 (RMSE) 指标对比的结果, 其中 PSNR 得分越高越好, SSIM 和 RMSE 得分越小越好

方法	PSNR	SSIM	RMSE
MAP-NN	25.01	0.72	6.31
FFDnet	25.44	0.76	<b>5.33</b>
BM3D	25.49	0.73	5.66
CPCE	24.89	0.68	6.38
<b>Ours</b>	<b>25.80</b>	<b>0.77</b>	5.52

### 3.4.5 基于生成对抗网络的噪声提取与学习的实验结果与分析

首先在合成的低剂量 CT 图像上通过定量的和可视化的结果来验证提出算法的有效性。对比结果如表 3-2 所示。可以有如下的观察结果:

- 首先, 与基于非深度学习的方法 (例如 WNNM) 相比, 基于深度学习的方法 (例如 FFDNet、Red-CNN) 通常可以获得更好的性能。这是合理的, 因为与基于非深度学习的相比, 深度学习的方法可以从数据中学习到有用的特征。
- 其次, 提出的基于基于噪声学习 (生成对抗网络) 的无监督低剂量 CT 图像降噪方法优于其他基于仿真数据集的低剂量 CT 图像降噪方法, 如 MAP-NN 和 Red-CNN, 这表明通过引入噪声学习的机制很好的解决了低剂量 CT 图像降噪中训练噪声和目标噪声不匹配的问题。尽管 TV-L1 在 RMSE 方面可以

表 3-2 不同模型的峰值信噪比 (PSNR)、结构相似度 (SSIM) 和均方根误差 (RMSE) 指标对比的结果, 其中 PSNR 得分越高越好, SSIM 和 RMSE 得分越小越好  
PSNR、SSIM 和 RMSE 结果, B/N-B 分别表示基于盲降噪的方法和非盲降噪的方法

方法	B/N-B	PSNR	SSIM	RMSE
DnCNN	B	21.62	0.43	7.59
FFDNet	N-B	22.76	0.57	7.58
BM3D	N-B	24.77	0.77	6.04
WNNM	B	20.39	0.64	6.62
TV-L1	B	23.12	0.75	<b>5.40</b>
MAP-NN	N-B	23.54	0.71	6.48
Red-CNN	N-B	24.79	0.74	6.84
CPCE	N-B	21.65	0.65	6.48
KSVD	B	23.61	0.72	6.89
<b>Ours</b>	<b>B</b>	<b>25.16</b>	<b>0.85</b>	6.21

取得更好的结果, 但它可能会导致潜在的模糊伪影问题。

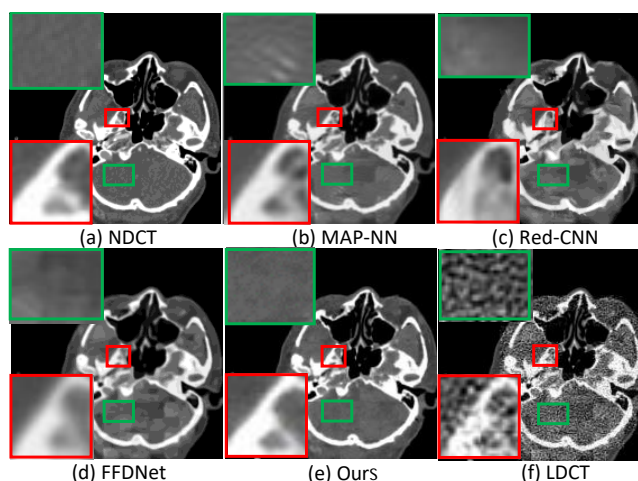


图 3-6 FUMPE 在合成低剂量 CT 图像评价中的比较结果, 请放大以获得更好的视角

为进一步验证算法的有效性, 根据可视化结果进行方法之间表现的比较。一些可视化的结果如图 3-6 所示, 可以发现虽然所有的方法都可以在一定程度上降低噪声, 但使用提出的方法重建的图像更为清晰, 额外的伪影也会更少 (如图 3-6 所示的红框和绿框)。

最后, 使用 CNR 进行无参的对比评估。该对比对临床场景的低剂量 CT 图像进行实验, 结果如表 3-3 所示。可以观察到, 提出的方法可以有效地处理临床场景的低剂量 CT 图像降噪问题。进一步地可视化的对比结果如图 3-7 所示, 可以发现使用 BM3D 方法的结果过于平滑, KSVD 和 Red-CNN 的噪声抑制效果较差, 尤其

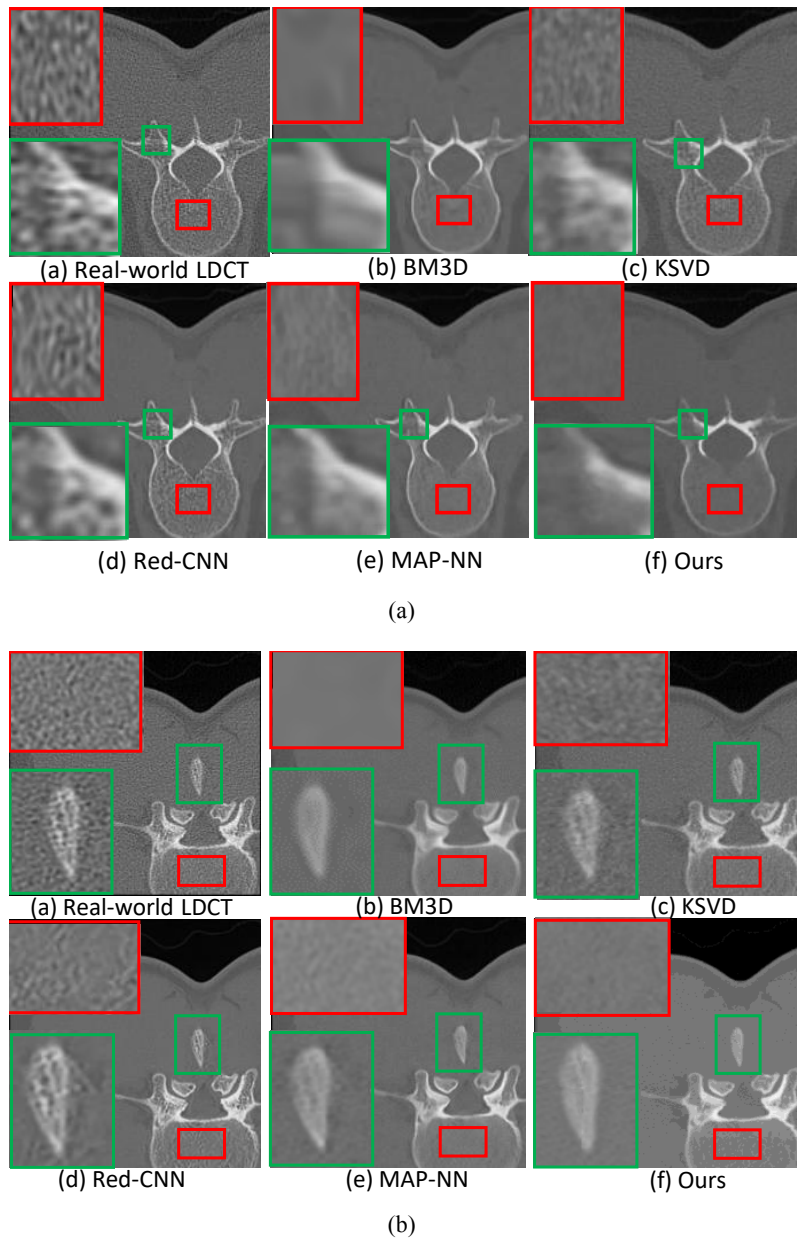


图 3-7 在 LVS 临床低剂量 CT 图像数据集进行降噪的可视化比较

表 3-3 使用 LVS 临床数据集进行测试，并对不同方法的降噪结果进行比较，对于相对噪声比 (CNR) 指标来说，得分越低越好

Methods	MAP-NN	Red-CNN	KSVD	BM3D	Our Method
Results	2.55	2.53	1.45	1.31	<b>1.23</b>

是在 ROI 的结构内部。相反，所提出的方法可以产生更逼真的纹理（见红色框）与临床场景噪声条件下的其他方法进行比较。另一方面，与其他方法（见绿色框）相比，所提出的方法可以产生细节更自然地过渡，这可能受到放射科医生的青睐。



### 3.4.6 消融实验

**噪声建模分析:** 合成的低剂量 CT 图像的质量决定着后续降噪任务的效果, 因此本文对此部分进行可视化验证。图 3-8 通过添加来自 GAN 的生成噪声来合成低剂量 CT 图像样本, 其中干净的正常剂量 CT 图像来自 FUMPE 数据集。可以注意到图中可视化的合成的低剂量 CT 图像的纹理与图 3-6 所示的临床场景的低剂量 CT 图像十分相似。为此, 提出的方法可以受益于该属性来处理未知的临床场景低剂量 CT 图像。进一步可视化了提出的框架产生的噪声。可视化结果如图 3-9 所示, 可以发现生成的噪声与低剂量 CT 图像中出现的真实噪声也非常相似。

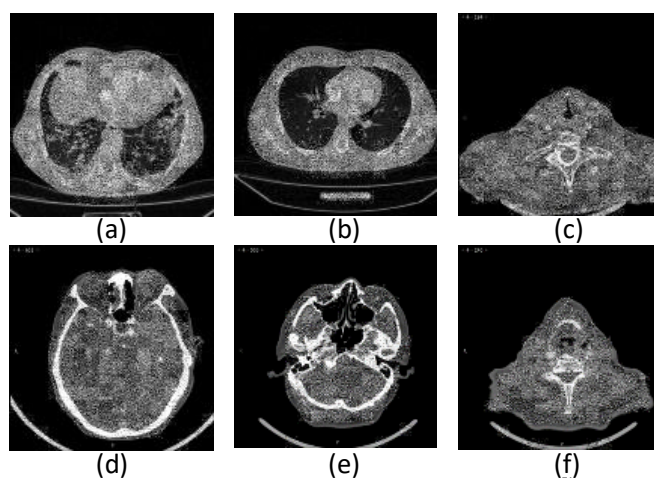


图 3-8 合成低剂量 CT 图像样本

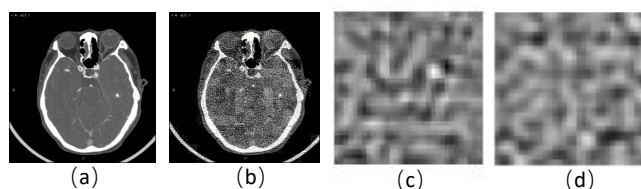


图 3-9 (a): FUMPE 数据集上的正常剂量 CT 图像样本, (b): 合成的低剂量 CT 图像, (c): 从真实低剂量 CT 图像中提取的临床场景噪声, (d) GAN 产生的噪声

**KL 损失与重建损失的对比:** 对于变分自编码器, 重构损失和 KL 损失决定了模型学习的有效性 (它们分别决定了重构质量和样本多样性)。因此, 本文通过可视化它们随训练的损失变化来观察整体变化。如图 3-10 和图 3-11 所示, 重构损耗逐渐减小并最终收敛, 这说明生成样本与真实样本非常相似。此外, KL-loss 在一个值附近快速收敛并稳定, 这对于保持多样性是必不可少的。

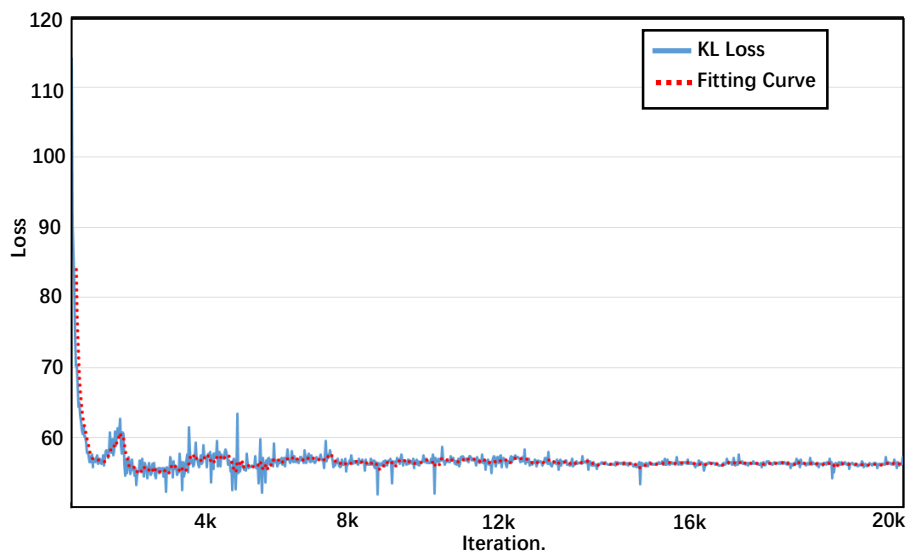


图 3-10 KL 损失随训练的变化曲线

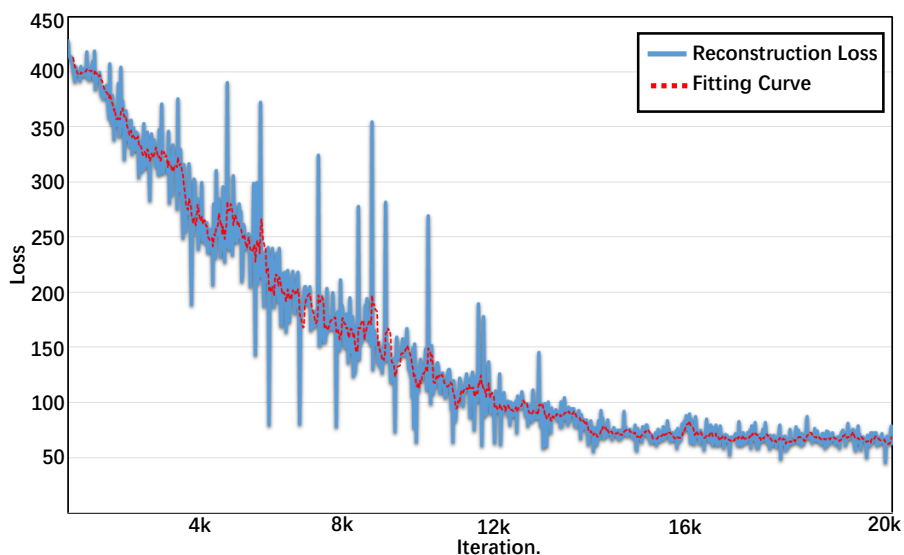


图 3-11 重建损失随训练的变化曲线

### 3.5 本章小节

主要研究低剂量 CT 图像的无监督降噪问题，旨在设计一种有效的方法解决训练数据集与目标数据集在噪声分布不匹配的问题。具体来说，的贡献包括两部分，利用噪声提取的方法获得平滑区域的噪声子块，并提出利用生成对抗网络和变分自编码器进行噪声分布的学习，以此构造数据对进行训练。通过对丰富的正常剂量 CT 图像の利用，构造了诱导的低剂量/正常剂量 CT 图像数据集，并使用基准的降噪器进行实验验证了算法的有效性。



## 第四章 基于上下游任务关联的低剂量 CT 图像降噪

### 4.1 问题定义

医学图像的质量对于医生的准确诊断至关重要<sup>[2,20]</sup>。特别是对于医学图像相关的人工智能算法来说,相当多的高阶任务(如病灶的自动检测和分割、多模态图像配准)都严重依赖于高质量的输入图像,因为:1)低质量图像中轻微的噪声扰动可能会导致模型出现意想不到的退化;2)一些小的病灶(如肺部的低剂量 CT 图像中微小的肺结节)会遭受严重的噪声,导致后处理和诊断十分困难。更重要的是,低质量的医学图像会受到噪声和伪影的干扰。因此,研究人员将重点放在低质量的医学图像恢复上,以便改进后的图像可以很好地用于潜在的下游任务。然而,现有的低剂量 CT 图像降噪算法最大的问题是现有的医学图像降噪任务与下游任务脱节。具体来说,现有的基于深度学习的方法通常通过多个评估指标展示其更好地降噪性能,例如定量结果(例如,峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)<sup>[15]</sup>,结构相似性指数测量 (Structural Similarity Index Measure, SSIM)<sup>[13]</sup>、纹理匹配损失 (Texture Match Loss, TML)<sup>[23]</sup>)和双盲评分实验<sup>[51]</sup>(通过噪声抑制、对比度和诊断可接受性的得分)。然而,对于基于深度神经网络的下游医学图像任务(如相关病灶的自动检测和分割、多模态图像配准)的实际可用性尚未得到探索。这些下游任务对于与医学相关的 AI 社区极为重要。

### 4.2 算法动机及其概述

正如上小节阐述的,现有的方法使得低剂量 CT 图像降噪算法与下游的高阶任务脱节,这可能导致最优的降噪结果对于下游任务来说可能并不是最优的。例如,现有的基于深度学习的低剂量 CT 图像降噪方法通常会在基于整个特征图上进行降噪,但是病灶自动检测任务更喜欢关注局部感兴趣区域 (Region of Interest, ROI),这导致这两个任务之间的目标不匹配。

为了解决这些问题,一个直觉的想法是将医学图像降噪任务与下游的高阶任务直接联系起来。通过这样做,可以通过下游任务的输出明确反映降噪图像的实际可用性。同时,降噪任务也可以利用下游任务的反馈,这样医学图像降噪任务就可以学习到一个综合性能最优的模式。

在本章中,考虑到临床使用频率、数据集的可用性和噪声特性,专注于基于深度学习的低剂量 CT 图像降噪任务和基于深度学习的下游病灶检测任务。根据上述思想,提出了一种即插即用的医学图像降噪框架,即病灶启发的降噪网络

(Lesion-Inspired Denoising Network, LIDnet), 以协同提高医学图像的降噪性能和检测结果的准确性。更具体地说, 本算法通过联合学习一个多损失目标函数, 将下游检测任务的反馈传送到现有的降噪框架中, 从而将医学图像降噪任务与下游检测任务联系起来。同时, 本算法通过由病灶检测任务产生的感兴趣区域的感知损失来进一步连接这两个任务。最后, 为了更好地优化整体框架, 本算法针对 LIDnet 提出了一种定制化的协同训练策略。

### 4.3 基于上下游任务关联的低剂量 CT 图像降噪架构

#### 4.3.1 数学变量声明

针对现有低剂量 CT 图像降噪模型和病灶检测模型的训练过程通常是脱节的问题, 本章算法旨在将这两个任务进行连接。将联合空间  $X \times Y \times Z$  上的训练样本表示为  $D = \{(x_i, y_i, (z_{y_r}^i, z_{y_c}^i, z_{y_w}^i, z_{y_h}^i, L_y^i))\}_{i=1}^N$ , 其中  $x_i$  和  $y_i$  表示第  $i$  个低剂量 CT 图像和相应的正常剂量 CT 图像。 $(z_{y_r}^i, z_{y_c}^i, z_{y_w}^i, z_{y_h}^i)$  表示第  $i$  个正常剂量 CT 图像的检测目标的边界框的 4 个坐标图像,  $L_y^i$  是检测目标的分类标签。

对于降噪任务来说, 的目标是学习一个映射网络  $F$ 。这个映射网络可以利用配对训练样本  $(x, y) \in (X, Y)$  进行学习, 从而学习到域  $\mathcal{X}$  到域  $\mathcal{Y}$  之间映射  $x \rightarrow y$  关系。对于病灶检测任务来说, 它大致可以分为两个分支, 即区域建议网络和病灶识别。

#### 4.3.2 基于感兴趣区域的感知损失

在介绍提出的区域和兴趣 (ROI) 感知损失之前, 回顾一下现有的低剂量 CT 图像降噪算法是如何计算感知损失的。正如 [23] 所指出的, 感知损失对于降噪是非常有效, 因为感知损失可以度量降噪后的 CT 图与正常剂量 CT 图像在特征空间中图像风格的相似性, 进而迫使网络朝着更接近正常剂量 CT 的方向进行学习。数学上, 感知损失可表示为

$$L_{perceptual\_loss} = L_{pl} = \mathbb{E}_{(x,y)} \left[ \frac{\|\varphi_{VGG}(F(x)) - \varphi_{VGG}(y)\|_F^2}{whd} \right], \quad (4-1)$$

其中  $\varphi_{VGG}$  表示一个特征提取器, 它可以通过 VGG 网络 [20] 的第 16 层卷积层生成输出。 $\|\cdot\|_F$  表示 Frobenius 范数。 $w, h,$  和  $d$  分别表示特征图的宽度、高度和数量。由公式 4-1 所示, 可以观察到感知损失是在整个特征图上计算的, 称之为**全局感知损失**。直观地说, 这种全局感知损失存在一个潜在的局限性: 在医学图像领域, 通常更关注 ROI 中的局部特征而不是全局特征, 因为相关病灶 (例如肺结节和肺栓塞) 通常位于器官的局部区域 [26]。然而, 现有的低剂量 CT 图像降噪方法倾向

于实现全局特征上的最佳的感知损失，而不是 ROI 中的局部特征。因此，通过全局感知损失优化的降噪模型可能无法保证 ROI 中的最优降噪结果。

为了解决上述问题，提出了一种新颖的基于感兴趣的感知损失。需要指出的是，提出的 ROI 感知损失是一种直接的方式，它不仅有利于局部特征的改善，而且有助于下游检测任务（因为这些任务主要作用于感兴趣区域）。有趣的是，如何获得 ROI 来计算局部特征的感知损失呢？受益于降噪和检测任务的连接不能问，本文创新性的提出利用目标检测模型的区域建议网络 (Region Proposals Networks, RPN) [54] 获得的感兴趣区域来计算感知损失（如图 4-1 所示）。RPN 是目标检测模型的一个重要模块，它以图像作为输入并输出一组建议的潜在目标区域。该过程可表述如下

$$\{(t_1, p_1), (t_2, p_2), \dots, (t_{M-1}, p_{M-1}), (t_M, p_M)\} = RPN(H(F(x))) \quad (4-2)$$

其中  $H(\cdot)$  表示检测网络中的特征提取网络（如图 5-1 中的 ResNet50）。 $RPN(\cdot)$  可以产生一组区域建议，由 4 个坐标点组成的矩形边界框  $t_i$  和对象得分  $p_i$  组成。 $M$  表示建议的目标区域的数量，通常非常大。考虑到效率，采用对象得分来选择前  $K$  个建议的目标区域来计算感知损失，

$$\{(t_1, p_1), (t_2, p_2), \dots, (t_K, p_K)\} = \text{select}(\{RPN(H(F(x)))_{i=1}^M\}), \quad s.t. \quad p_i > p_K. \quad (4-3)$$

这个过程可以确保获得的感兴趣区域相对有意义，因为图像的前景（通常是感兴趣的对象）与图像的背景（例如 CT 图像中黑色的空气区域）相比具有更高的目标得分。相比之下，现有的感知损失只是融合了全局特征，无法显式地关注有意义的感兴趣区域。最后，特征图上的基于感兴趣区域的感知损失可以表示为

$$L_{ROI\_pl} = \mathbb{E}_{(x,y)} \frac{1}{K} \sum_{i=1}^K \left[ \frac{\|T(F(x))_{t_i} - T(y)_{t_i}\|_F^2}{whd} \right], \quad (4-4)$$

其中  $T(\cdot)$  表示特征提取器。对于  $T(\cdot)$  的选择，现有方法通常采用 VGG 网络。本文提出采用检测网络的主干网络作为特征提取器，用于计算降噪后的图像与其参考图像在特征空间中的差异。

### 4.3.3 上下游任务的关联机制

在本小节，降噪网络和病灶检测网络的关联机制将被详细介绍。这种连接的动机有两个方面：首先，利用一种多损失的优化函数（包括检测网络的损失、降噪网络的损失和感知损失）来共同优化降噪网络。这是一种简单有效的方式，降

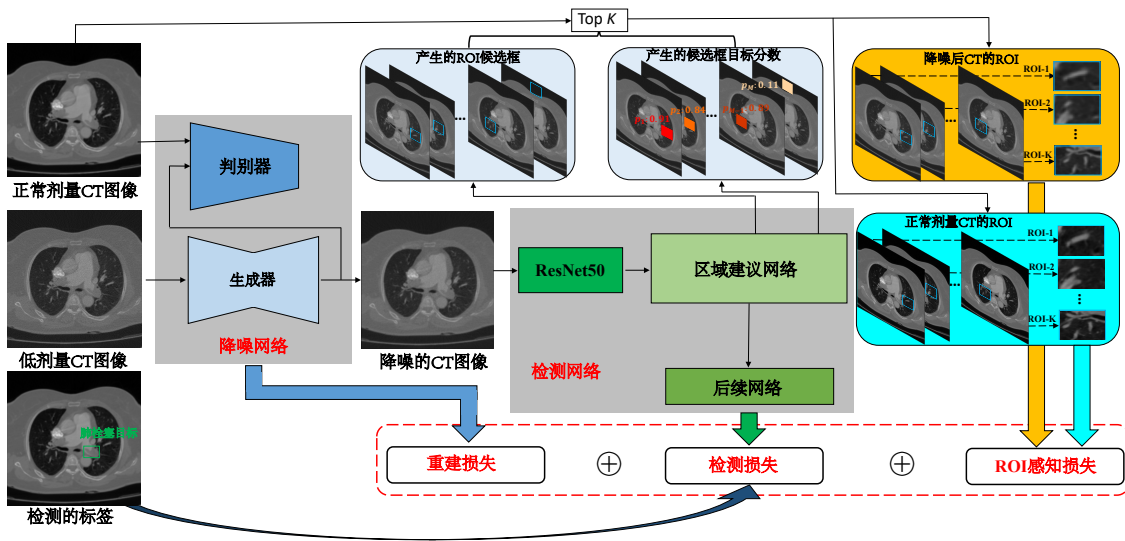


图 4-1 提出的基于上下游任务关联的低剂量 CT 图像降噪算法的整体架构。降噪网络包括生成器和判别器。降噪后的 CT 图像由生成器生成。检测网络包括主干网络（ResNet50）、区域建议网络和后续网络。候选的 ROI 区域和相应的对象分数由区域建议网络获得。根据目标得分，降噪的 CT 图像中有意义的前景 ROI 和相应的正常剂量 CT 的 ROI 被选择用于计算感知损失

噪网络可以感知下游检测任务的反馈，从而协同提高降噪效果和病灶检测的准确性。其次，提取的 ROI 是从检测网络中直接获得来计算感知损失的，这可以促进基于感兴趣区域的感知损失用于优化降噪网络，这也是连接的一个方面。

本文所提出的 LIDnet 的整体架构如图 4-1 所示。对于降噪网络来说，现有的低剂量 CT 图像降噪方法通常采用生成对抗网络（GAN）<sup>[52]</sup> 的方案。GAN 是一个具有生成器和判别器的极小极大游戏。对于提出的 LIDnet，降噪网络中生成器的参数  $\Theta_G$  通过一个联合的损失函数进行优化，如下所示：

$$L_{total} = \mathbb{E}_x[-D(G(x))] + \lambda_1 \mathbb{E}_{(\hat{x}, y)} \frac{1}{K} \sum_{i=1}^K \left[ \frac{\|T(\hat{x})_{t_i} - T(y)_{t_i}\|_F^2}{whd} \right] + \lambda_2 L_{det}, \quad (4-5)$$

其中第一项是生成器损失，它保证生成器的数据分布可以从较高噪声的 CT 图像转换为相应的无噪声版本。这一项也称为重建损失，它考虑了不太详细的信息。第二项和第三项分别是提出的基于感兴趣区域的感知损失函数和检测网络损失函数。

$\lambda_1$  和  $\lambda_2$  用于平衡这两项。第三项可以表述为

$$L_{det} = \mathbb{E}_{(t_i^*, p_i^*, \hat{x})} [Cross\_entropy(p_i^*, S_1(H(\hat{x}))) + L1_s(t_i^*, S_2(H(\hat{x})))] \quad (4-6)$$

对于公式 4-5 和公式 4-6,  $\hat{x}$  表示降噪网络的输出,  $t_i^*$  和  $p_i^*$  表示正常剂量的 CT 图像中感兴趣区域的 4 个参数化的坐标及其标签。预测的目标得分  $S_1(H(\hat{x}))$  和预测的目标坐标  $S_2(H(\hat{x}))$  由检测模型的后续网络生成。  $Cross\_entropy(\cdot)$  表示预测的目标分数和相应的真实标签之间的交叉熵损失的计算。  $L1_s$  表示平滑的 L1 损失。检测损失的更多细节可以在 [48] 中找到。

对于判别器的损失函数, 遵循带梯度惩罚的推土机距离 (Wasserstein Distance) 来优化判别器  $D$  [42], 可以表示为

$$L_D = \mathbb{E}_{\hat{x} \sim \mathbb{P}_G} [D(\hat{x})] - \mathbb{E}_{y \sim \mathbb{P}_y} [D(y)] + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D_{\theta}(\hat{x})\|^2 - 1)], \quad (4-7)$$

$\mathbb{P}_y$  是正常剂量 CT 图像之间的分布。  $\mathbb{P}_G$  是生成器输出之间的分布。  $\mathbb{P}_{\hat{x}}$  被定义为在从  $\mathbb{P}_y$  和  $\mathbb{P}_{\hat{x}}$  采样的点对之间沿直线均匀采样的分布  $G, \lambda$  控制梯度惩罚的重要性。

除了 GAN 的方案, 一些主流的低剂量 CT 图像降噪方法 (如 Red-CNN [17]) 也采用基于卷积神经网络 (CNN) 的降噪架构。通过基于 CNN 的网络替换基于 GAN 的降噪网络, 它们也可以被插入到提出的 LIDnet 框架中。通过这样做, 降噪网络的目标可以表述为:

$$L_{total} = \mathbb{E}_{x,y} \left[ \frac{1}{N} \|F(x) - y\|_F^2 \right] + \lambda_1 L_{ROI\_pl} + \lambda_2 L_{det}, \quad (4-8)$$

其中第一项也是重建损失。  $N$  表示批次中的样本数。对于检测网络, 采用 faster R-CNN 中提出的损失函数来优化参数。

从公式 (4-5) 和公式 (4-8) 可以发现, 检测网络的输入是上游降噪网络产生的降噪结果。这意味着如果检测网络 (最后一项) 的损失较高, 则降噪网络的降噪质量可能会更差。为此, 较高检测损失将推动降噪网络学习更好的模式, 这可以协同提高降噪性能和检测结果的准确性。

#### 4.3.4 基于协同训练的连接策略

所提出的 LIDnet 框架涉及多个网络, 可以发现这些网络是相互约束的。具体来说, 降噪网络需要检测网络提供感兴趣区域来计算感知损失。同时, 检测网络

**Algorithm 1** 基于协同训练的连接策略**Input:**

- 1: 低剂量 CT 图像  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ , 对应的正常剂量 CT 图像  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ , 正常剂量 CT 图像对应的目标坐标  $\mathbf{T}^* = \{\mathbf{t}_1^*, \dots, \mathbf{t}_N^*\}$ , 正常剂量 CT 图像对应的目标类别  $\mathbf{P}^* = \{\mathbf{p}_1^*, \dots, \mathbf{p}_N^*\}$ . 初始的降噪网络的参数  $\Theta$ , 初始的检测网络的参数  $\psi$
- 2: 检测网络预训练的迭代次数  $T_1$
- 3: 降噪网络每一轮的训练步数  $T_2$
- 4: 检测网络每一轮的训练步数  $T_3$

**Output:** 学习到的参数:  $\Theta^*$  和  $\psi^*$ 

- 5: 冻结  $\Theta$
- 6: **for**  $i=0$  到  $T_1$  **do**
- 7: 采样一些样本  $\mathbf{X}_d, \mathbf{Y}_d, \mathbf{T}_d^*, \mathbf{P}_d^*$ , 他们分别来自  $\mathbf{X}, \mathbf{Y}, \mathbf{T}^*, \mathbf{P}^*$
- 8: 优化  $\psi$  通过 (4-6) 即  $S_1, S_2$  和  $RPN$  on  $\mathbf{Y}_d, \mathbf{T}_d^*$ , 和  $\mathbf{P}_d^*$
- 9: **end**
- 10: **While** 当不满足条件停止训练 **do**
- 11: 解冻  $\Theta$ ; 冻结  $\psi$
- 12: **for**  $i=0$  到  $T_2$  **do**
- 13: 采样一些样本  $\mathbf{X}_d, \mathbf{Y}_d, \mathbf{T}_d^*$ , 和  $\mathbf{P}_d^*$  分别来自  $\mathbf{X}, \mathbf{Y}, \mathbf{T}^*$ , 和  $\mathbf{P}^*$
- 14: 优化  $\Theta$  通过 (4-5), (4-7) 即 D 和 G (or 优化  $\Theta$  通过 (4-8) 即 G 在  $\mathbf{X}_d, \mathbf{Y}_d, \mathbf{T}_d^*$ , 和  $\mathbf{P}_d^*$
- 15: **end**
- 16: 冻结  $\Theta$ ; 解冻  $\psi$
- 17: **for**  $i=0$  to **do**
- 18: 采样一些样本  $\mathbf{X}_d, \mathbf{Y}_d, \mathbf{T}_d^*$ , 和  $\mathbf{P}_d^*$  分别来自  $\mathbf{X}, \mathbf{Y}, \mathbf{T}^*$ , 和  $\mathbf{P}^*$ .
- 19: 通过降噪网络计算降噪的输出:  $X_{denoise} \leftarrow G(X_d)$
- 20: 优化  $\psi$  通过 (4-6) 即  $S_1, S_2$  和  $RPN$  在  $X_{denoise}, \mathbf{T}_d^*$ , 和  $\mathbf{P}_d^*$ .
- 21: **end**
- 22: **end**

需要降噪网络的输出作为其输入。如果两个网络中的一个不够强大，另一个网络也将遭受糟糕地优化方向。

为了解决这个问题，本算法为 LIDnet 框架提出了一种定制化的协作训练策略 (Collaborative Training Strategy, CTS)，它类似于 GAN (迭代优化) 的训练过程。更具体地说，首先使用高质量的正常剂量 CT 图像以预定义的  $T_1$  步骤训练检测网络，使检测网络朝着正确的方向优化，并逐渐提供可接受的精确的 ROI。然后，多轮交替进行降噪网络和检测网络的训练。一轮过后，降噪网络首先根据其损失函数进行训练，即 (4-5) 和 (4-8)，执行预定义的  $T_2$  步迭代。在这里，感知损失计算所需要的 ROI 是通过检测网络的区域建议网络提供的。在训练了一定数量的步数后，检测网络会根据预定义的  $T_3$  步后的降噪网络的输出再次训练。整个算法在算法 1 中描述。

表 4-1 数据集的细节

数据集	PE-CT <sup>①</sup>	L-CT-A <sup>②</sup>	L-CT-G <sup>③</sup>
扫描部位	胸部	肺部	肺部
病灶	肺结节	恶性腺瘤	肾细胞瘤
病灶大小	小	大	大
切片厚度	$\leq 1\text{mm}$	$2\text{mm}$	$2\text{mm}$
切片间隔	$\leq 1.5\text{mm}$	$0.625\text{mm} \sim 5\text{mm}$	$0.625\text{mm} \sim 5\text{mm}$
分辨率	$512 \times 512$	$512 \times 512$	$512 \times 512$
总共数量	2304	3000	3000
用于训练	2000	2500	2500
用于测试	304	500	500

#### 4.4 实验结果及其分析

本节设计大量实验以评估所提出的 LIDnet 框架的有效性。根据已有文献，目前还没有开源医学图像数据集同时考虑 CT 降噪和检测。为此，使用正常剂量 CT 图像检测数据集来模拟相应的低剂量 CT 图像。最终可以获得联合模拟的低/正常剂量和检测数据集。使用低剂量和正常剂量的 CT 对患者进行两次扫描是不可接受的，类似的模拟方案因此在低剂量 CT 图像降噪<sup>[19]</sup>领域被广泛采用。所用数据集的详细信息如表 5-1 所示，采用了一种简单有效的低剂量 CT 图像模拟方法，如<sup>[40]</sup>所示。对于 PE-CT、L-CT-A 和 L-CT-G 数据集，噪声水平  $N_0$  分别设置为 3000、1000 和 1000。

##### 4.4.1 基线方法和评估指标

对于降噪任务，使用降噪图像的定量指标、可视化图像和降噪图像的检测精度等三方面将提出的 LIDnet 框架与以下基线方法进行了比较。基线方法的细节描述如下：

- **WGAN-VGG** (2018)<sup>[52]</sup>：考虑提出的 WGAN-VGG 模型在<sup>[52]</sup>作为基线。WGAN-VGG 采用基于推土机距离的生成对抗网络和感知损失（由 VGG 网络计算）用于低剂量 CT 图像降噪。
- **CPCE**<sup>[15]</sup> (2018)：训练了一个基于传送路径的卷积编码器-解码器 (CPCE) 网络。该网络实现了令人印象深刻的低剂量 CT 图像降噪性能，并且还使用基于 VGG 的感知损失进行训练。
- **MAP-NN**<sup>[16]</sup> (2019)：应用模块化低剂量 CT 图像降噪深度神经网络 (MAP-NN)。该网络包括多个相等的降噪模块。选择具有 3 个模块的版本，以平衡降噪水平和细节保留。MAP-NN 也采用了基于 VGG 的感知损失。

- **SSMS** <sup>[21]</sup> (2019): 该网络采用结构敏感多尺度 (SSMS) 深度神经网络进行低剂量 CT 降噪。不使用基于 VGG 的感知损失, 它使用结构化相似性指数测量 (SSIM) 损失来衡量两个图像之间的结构和感知相似性。
- **Res-HLF** <sup>[24]</sup> (2020): 将具有混合损失函数的生成对抗网络应用于低剂量 CT 图像降噪。与 SSMS 类似, Res-HLF 也将基于 VGG 的感知损失替换为 SSIM 损失, 以测量感知相似度。此外, 它还引入了残差学习。

为了评估不同方法的降噪性能, 分别计算了降噪结果在感兴趣区域 (ROI) 和整体图像上的量化性能。选择的 ROI 来自检测标签 (病灶区域), 它可以提供更多有用的病理信息。对于 ROI, 使用降噪图像和相应正常剂量 CT 图像之间的影响组学特征的平均绝对差异来评估潜在疾病特征的结构保留能力。影像组学函数具有强大的能力, 可以揭示 ROI 中病灶的微观特征。根据经验选择了三个典型的影像组学特征 (包括相关性、同质性和能量)。Pyradiomics 平台<sup>①</sup> 用于计算感兴趣区域的每个特征的单个值。关于计算放射组学特征的更多细节可以在<sup>[33]</sup>中找到。对于整体性能, 采用峰值信噪比 (Peak Signal to Noise Ratio, PSNR)、结构化相似性指数测量 (Structural Similarity Index Measure, SSIM) 和均方根误差 (Root-Mean-Square Error, RMSE)。

表 4-2 检测任务在三个数据集上的定量结果, 分别为 AP-50 和 AP-75。基线方法和基于本文提出框架的基线方法之间的较高的分数用蓝色粗体表示, 对于 AP-50 和 AP-75, 越高越好

数据集		WGAN-VGG <sup>[52]</sup>	Ours-WGAN-VGG	CPCE <sup>[15]</sup>	Ours-CPCE	MAPNN <sup>[16]</sup>	Ours-MAPNN	SSMS <sup>[21]</sup>	Ours-SSMS
PE-CT	AP-50	74.56	<b>74.59</b>	74.60	<b>75.69</b>	75.03	<b>76.79</b>	73.16	<b>74.07</b>
	AP-75	25.46	<b>26.70</b>	25.74	<b>26.26</b>	25.75	<b>27.01</b>	23.02	<b>24.62</b>
L-CT-G	AP-50	93.90	<b>95.31</b>	91.62	<b>93.74</b>	92.84	<b>95.14</b>	91.15	<b>93.89</b>
	AP-75	46.15	<b>48.14</b>	35.89	<b>46.67</b>	45.04	<b>45.53</b>	38.76	<b>44.00</b>
L-CT-A	AP-50	<b>91.51</b>	91.31	88.01	<b>89.76</b>	86.98	<b>91.44</b>	84.02	<b>88.50</b>
	AP-75	41.34	<b>41.48</b>	41.30	41.30	40.00	<b>43.21</b>	37.49	<b>40.39</b>

为了评估不同方法的检测性能, 使平均精度 (Average Precision ,AP)-50 和 AP-75。需要注意的是, 每个数据集的定量检测结果都是通过预训练的检测网络计算出来的。具体来说, 基于正常剂量 CT 图像训练的检测网络。直观地说, 如果降噪图像比相应的正常剂量 CT 图像更接近, AP 的得分会更高。

#### 4.4.2 网络结构

所提出的 LIDnet 是一个即插即用框架, 它旨在协同提高现有方法的降噪和检测性能。因此, 分别采用 5 种基线方法作为降噪网络的主干, 以及改进的 Faster

① <https://pyradiomics.readthedocs.io/en/latest/>



表 4-3 L-CT-A 数据集的感兴趣区域和整体区域降噪任务的量化结果。基线方法和基于本文提出框架的基线方法之间的较高的分数用蓝色粗体表示。对于相关性 (Correlation)、同质性 (Homogeneity)、均方根误差 (RMSE) 和能量 (Energy) 指标, 分数越低越好。对于峰值信噪比 (PSNR) 和结构相似度 (SSIM) 指标, 分数越高越好

L-CT-A 数据集						
	感兴趣区域			整体区域		
	Correlation	Homogeneity	Energy	PSNR	SSIM	RMSE
WGAN-VGG	<b>0.042</b>	<b>0.085</b>	<b>0.025</b>	27.18	99.94	0.049
Ours-WGAN-VGG	0.048	0.104	0.029	<b>29.13</b>	<b>99.96</b>	<b>0.042</b>
CPCE	0.049	0.109	0.030	26.30	99.93	0.057
Ours-CPCE	<b>0.043</b>	<b>0.105</b>	0.030	<b>28.01</b>	<b>99.95</b>	<b>0.047</b>
MAPNN	0.047	0.102	0.027	28.20	99.95	0.046
Ours-MAPNN	<b>0.045</b>	<b>0.101</b>	0.027	<b>28.45</b>	<b>99.96</b>	<b>0.045</b>
SSMS	0.047	0.125	0.032	<b>27.56</b>	99.94	<b>0.049</b>
Ours-SSMS	<b>0.046</b>	<b>0.105</b>	<b>0.030</b>	25.85	99.94	0.057
Res-HLF	0.027	<b>0.036</b>	0.105	0.029	29.70	99.96
Ours-Res-HLF	<b>0.026</b>	0.054	<b>0.097</b>	<b>0.026</b>	<b>30.78</b>	<b>99.98</b>

表 4-4 L-CT-G 数据集的感兴趣区域和整体区域降噪任务的量化结果。基线方法和基于本文提出框架的基线方法之间的较高的分数用蓝色粗体表示。对于相关性 (Correlation)、同质性 (Homogeneity)、均方根误差 (RMSE) 和能量 (Energy) 指标, 分数越低越好。对于峰值信噪比 (PSNR) 和结构相似度 (SSIM) 指标, 分数越高越好

L-CT-G 数据集						
	感兴趣区域			整体区域		
	Correlation	Homogeneity	Energy	PSNR	SSIM	RMSE
WGAN-VGG	<b>0.048</b>	<b>0.068</b>	0.019	26.74	99.97	0.049
Ours-WGAN-VGG	0.049	0.076	<b>0.018</b>	<b>30.94</b>	<b>99.99</b>	<b>0.032</b>
CPCE	0.055	0.090	0.022	28.45	99.98	0.042
Ours-CPCE	<b>0.054</b>	<b>0.078</b>	<b>0.018</b>	<b>29.90</b>	<b>99.99</b>	<b>0.035</b>
MAPNN	0.052	0.078	0.018	23.37	99.91	0.077
Ours-MAPNN	<b>0.046</b>	<b>0.077</b>	<b>0.017</b>	<b>30.70</b>	<b>99.99</b>	<b>0.032</b>
SSMS	0.071	0.109	0.022	24.55	99.95	0.061
Ours-SSMS	<b>0.070</b>	<b>0.010</b>	0.022	<b>29.84</b>	<b>99.99</b>	<b>0.035</b>
Res-HLF	<b>0.025</b>	0.080	0.022	<b>32.30</b>	99.99	0.027
Ours-Res-HLF	0.033	<b>0.063</b>	<b>0.016</b>	31.67	<b>99.99</b>	<b>0.026</b>

表 4-5 PE-CT 数据集的感兴趣区域和整体区域降噪任务的量化结果。基线方法和基于本文提出框架的基线方法之间的较高的分数用蓝色粗体表示。对于相关性 (Correlation)、同质性 (Homogeneity)、均方根误差 (RMSE) 和能量 (Energy) 指标, 分数越低越好。对于峰值信噪比 (PSNR) 和结构相似度 (SSIM) 指标, 分数越高越好

PE-CT 数据集						
	感兴趣区域			整体区域		
	Correlation	Homogeneity	Energy	PSNR	SSIM	RMSE
WGAN-VGG	0.023	<b>0.038</b>	<b>0.009</b>	27.49	99.97	0.046
Ours-WGAN-VGG	0.023	0.058	0.013	<b>35.51</b>	<b>1.00</b>	<b>0.017</b>
CPCE	0.021	0.045	0.012	26.34	99.97	0.049
Ours-CPCE	<b>0.019</b>	<b>0.034</b>	<b>0.009</b>	<b>31.76</b>	<b>99.99</b>	<b>0.028</b>
MAPNN	0.024	0.045	0.011	27.33	99.97	0.043
Ours-MAPNN	<b>0.018</b>	<b>0.038</b>	<b>0.009</b>	<b>31.85</b>	<b>99.99</b>	<b>0.027</b>
SSMS	0.040	0.138	0.043	<b>33.89</b>	99.99	<b>0.021</b>
Ours-SSMS	<b>0.028</b>	<b>0.087</b>	<b>0.021</b>	31.86	99.99	0.027
Res-HLF	<b>0.023</b>	0.107	0.031	31.35	99.99	0.030
Ours-Res-HLF	0.122	<b>0.106</b>	<b>0.021</b>	<b>32.11</b>	99.99	<b>0.027</b>

RCNN 作为检测网络。请注意, ResNet50 在经验上被用作特征提取器的主干。最终, 可以获得 5 个基于 LIDnet 的模型, 即 **Ours-model\_name** (例如 **Ours-CPCE**)。比较了基线方法与其基于 LIDnet 的版本之间的降噪和检测性能, 以表现提出的 LIDnet 的有效性。对于所有模型, Adam 优化器以小批量方式进行参数优化。批量大小为 8。

#### 4.4.3 检测任务的结果及其分析

在上述三个数据集上训练了 4 种基线方法及其基于 LIDnet 的版本。请注意, 4 种基线方法遵循其默认训练协议。对基于 LIDnet 的版本采用提出的协作训练策略。训练好地降噪模型用于对每个数据集的测试集进行降噪任务。然后, 使用预训练的检测网络来评估降噪 CT 图像的检测性能。

检测任务的定量结果如表 4-2 所示。正如所见, 与所有基线方法相比, 基于 LIDnet 的基线方法的版本可以实现更好的性能。这是合理的, 因为提出的 LIDnet 框架直接插入通过联合学习多损失目标 (参见公式 4-5 和公式 4-8) 将检测网络反馈到上游降噪网络。

同时可以观察到, 在 PE-CT 数据集 (小目标) 中, 小物体的检测对于现有的检测模型<sup>[26]</sup>来说通常是一项非常困难的任务, 这反过来又需要高质量的输入图像 (这里的降噪图像)。在图 4-2 中展示了 PE-CT 数据集的可视化肺栓塞检测示例。

该 CT 图像中肺栓塞的病灶位置如图 4-2 的正常剂量 CT 图像所示。

可以注意到，与低剂量 CT 图像（甚至无法检测到物体）相比，大多数方法（SSMS 除外）都可以提高检测性能，这是合理的，因为降噪后的图像具有较少的噪声扰动。与所有基线相比，提出的模型（基于 LIDnet 的 MAP-NN 模型）在 IoU（Intersection over Union，可以评估对象检测的准确性）上取得了更好的分数，这是降噪网络和检测网络连接的结果。有趣的是，虽然可以在 SSMS 的降噪图像中发现非常强地降噪效果，但仍然无法检测到物体。

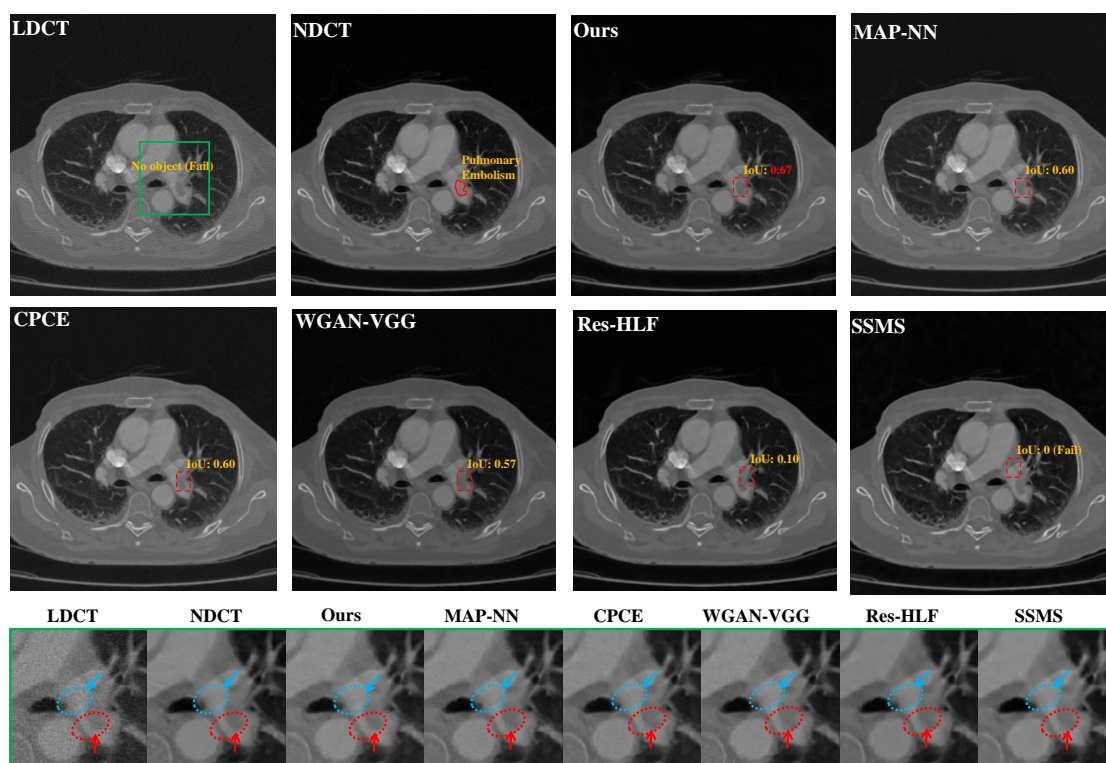


图 4-2 PE-CT 数据集上的可视化肺栓塞检测示例。肺栓塞病灶在正常剂量 CT 图像中用红色标记。对于每个降噪的 CT 图像，红色虚线框是检测对象的边界框。请注意，在低剂量 CT 图像中无法检测到物体。对于 IoU（Intersection over Union）的得分，越高越好。绿色框是为了分析详细的降噪性能而选择的观察区域，可以生成每个 CT 图像的放大子图像。在放大的子图中，蓝色和红色虚线圆圈是两个选定区域，以便更好地比较

#### 4.4.4 降噪任务的结果及其分析

降噪表现的定量结果见表 4-3、表 4-4、表 4-5 所示。有一些观察结果如下：首先，可以观察到，基于 LIDnet 的模型在大多数情况下具有更好的 ROI 定量影像组学特征，这意味着潜在的病灶特征是与他们的基线方法相比，保存得更好。一方

面，ROI 中更好地量化性能有利于所提出的 ROI 感知损失，这可以使所提出的模型更加关注 ROI 的效果。另一方面，检测性能也受到 ROI 图像质量的影响。

为了最小化检测损失，所提出的模型必须在 ROI 中产生更好的性能。其次，与基线方法相比，基于 LIDnet 的模型在大多数情况下也提高了整体性能（包括 PSNR、SSIM 和 RMSE），这意味着降噪图像在整体视图中更接近正常剂量 CT 图像。第三，可以观察到，在 PE-CT 数据集中，SSMS 的整体定量结果优于其基于 LIDnet 的版本，尤其是对于 PSNR。然而，如 4.4.3 节所述，SSMS 的降噪图像的检测性能极差。可以观察到 SSMS 的放大子图（见图 4-2 的最后一行）是过度平滑的，这通常会导致更高的 PSNR 分数。总之，推测过度平滑的降噪结果对检测任务没有贡献。有趣的是，得益于降噪任务和检测任务的联系，提出的模型不仅提高了所有数据集的检测性能，而且在大多数情况下提高了整体量化结果。

图 4-2 的最后一行显示了可视化的放大的结果。在这个选定的区域可以很好地观察到病灶。有以下观察：首先，如红色箭头所示，与其他方法相比，提出的方法实现了最好的伪影抑制，导致病灶边缘非常清晰。这可能受益于 ROI 感知损失的影响。其次，如蓝色和红色圆圈所示，可以发现与 MAP-NN、CPCE 和 WGAN-VGG 相比，提出的方法具有更好的噪声抑制性能。Res-HLF 和 SSMS 显示出较少的噪声，但它们存在过平滑问题。

额外的可视化肺栓塞检测示例如图 4-3 所示。与其他方法相比，所提出的模型（基于 LIDnet 的 MAP-NN 模型）实现了最佳检测性能。进一步分析图 4-3 最后一行的放大结果。如红色圆圈所示，可以发现 Res-HLF 和 SSMS 与其他模型相比是过于平滑的结果。对于病灶区域（由蓝色虚线圈出），文提出的模型和 Res-HLF 实现了更好的灰度水平的保留（如蓝色箭头所示）。反而，其他模型相对削弱了灰度水平。总体而言，与其他模型相比，提出的模型实现了全面更好地可视化性能。

#### 4.4.5 消融实验

首先验证所提出的协作训练策略（CTS）是否有效。图 4-4 展示了 CTS 的训练过程与基线训练过程的比较（这表示降噪网络和检测网络同时进行了优化）。可以观察到，在初始步骤中，与提出的 CTS 相比，基线训练策略具有更高的 AP 分数。这是合理的，因为提出的 CTS 的降噪网络在此期间没有得到优化（降噪后的图像质量很差）。受益于协同训练策略，可以充分优化降噪和检测网络，从而为训练指明正确的方向，从而在后续步骤中获得更好的性能。

本文还对生成的 ROI 候选框是否对计算 ROI 感知损失有意义感兴趣。两个输入的低剂量 CT 图像的可视化生成 ROI 候选如图 4-5 所示。如图 4-5(a) 所示，生成

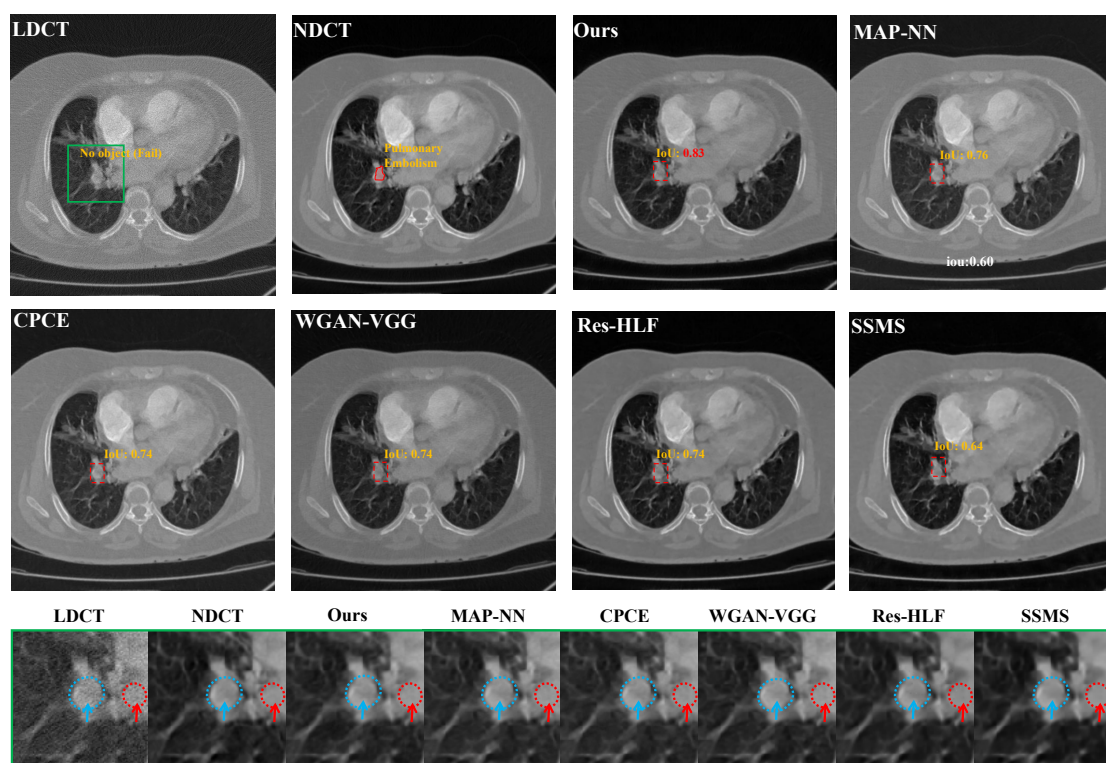


图 4-3 PE-CT 数据集上的可视化肺栓塞检测示例。肺栓塞病灶在正常剂量 CT 图像中用红色标记。对于每个降噪的 CT 图像，红色虚线框是检测对象的边界框。请注意，在低剂量 CT 图像中无法检测到物体。对于 IoU 的得分，越高越好。绿色框是为了分析详细的降噪性能而选择的观察区域，可以生成每个 CT 图像的放大子图像。在放大的子图中，蓝色和红色虚线圆圈是两个选定区域，以便更好地比较

的 ROI 候选 #2 确实位于有意义的位置（结节）。如图 4-5(b) 所示，这些生成的 ROI 候选框与图 4-2 中标记的病灶非常相似，这是合理的，因为检测网络将通过训练产生越来越准确的 ROI 候选框。受益于准确的 ROI 候选框的产生，ROI 感知损失将评估降噪 CT 图像和相应的正常剂量 CT 图像在特征空间中的图像质量，这可以提高 ROI 中的降噪性能（见表 4-2）。

#### 4.5 本章小节

在本章，创新性的提出了一种即插即用的医学图像降噪方法，以协同提高降噪性能和检测精度。通过联合学习多损失目标将下游检测任务的反馈插入到现有的降噪框架中。还提出了一种新的 ROI 感知损失来进一步连接这两个任务。所提出的协同训练策略有助于更好地优化这两个任务。考虑到临床使用频率、数据集



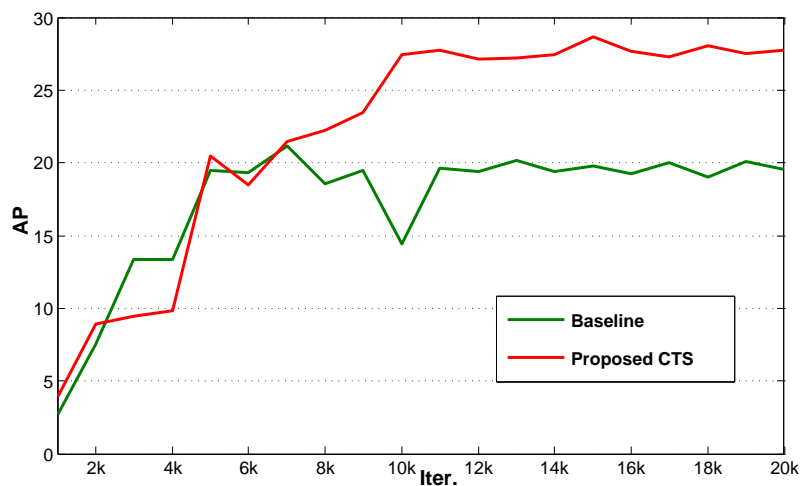


图 4-4 消融研究：提出的 LIDnet 协作训练策略 (CTS) 的有效性

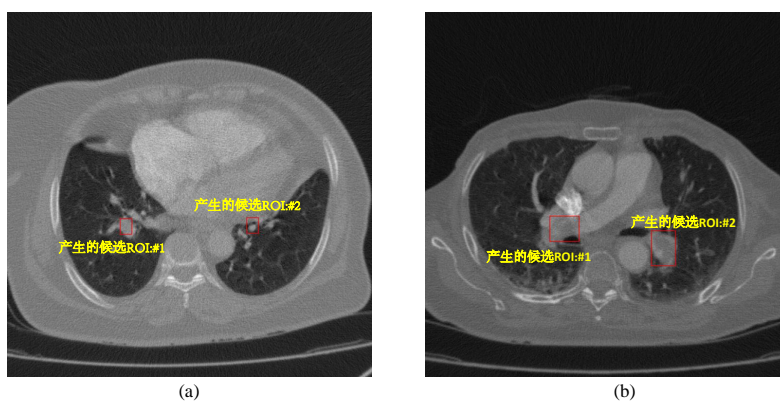


图 4-5 消融研究：生成的 ROI 候选框的有效性

的可用性和噪声特性，使用三个低剂量 CT 数据集来评估所提出框架的有效性。

## 第五章 医生阅片行为启发的低剂量 CT 图像降噪

### 5.1 问题定义

在现有的基于深度学习的低剂量 CT 图像降噪方法中，降噪任务通常被认为是将低剂量 CT 图像映射到对应的正常剂量图像的过程，其中深度卷积神经网络常被作为映射函数。然而，对于所采用的模型的内部机制却很少被关注。比如，该模型较少关注对低剂量 CT 图像降噪过程中学到了什么信息？现有研究表明，基于卷积神经网络的模型更专注于局部信息的提取<sup>[55]</sup>。这种降噪机制与放射科医生使用的降噪机制常常很不同，在临床应用中导致了潜在信任问题。在阅读低剂量 CT 图像时，放射科医生更倾向于相信一个受自身行为或工作流程启发的模型，而不是一个与他们的工作机制明显不同的模型。有趣的是，什么是放射科医生的阅读行为或工作流程呢？

### 5.2 算法动机及其概述

通过与放射科医生的深入沟通，放射科医生的低剂量 CT 图像降噪机制通常包含三步，大致可归纳如图 5-1 所示。首先，放射科医生将重点关注表示局部信息的感兴趣区域 (如图 5-1 的步骤 1 中的红框所示)，例如肺的微小支气管和肺的微小结节。然而，这种细微地组织结构和血管很容易被噪声淹没，使得放射科医生很难观察到。其次，为了进一步解读这些不易观察到的组织，放射科医生会利用这些非局部但容易观察到的类似组织进行辅助观察 (如图 5-1 的步骤 2 中的绿框所示)。这一步显示了非局部信息的重要性。第三，滑动鼠标是放射科医生的自然行为，这样类似的感兴趣区域可以在前后切片观察到 (如图 5-1 的步骤 3 的绿框所示)。由于噪声的随机性，大量受噪声干扰的感兴趣区域可能在前或后片切片中存在低噪声的相似区域。这隐式地利用了上下文信息。根据上述讨论，本文可以得出结论，即放射科医生的阅读工作流程采用了综合的信息表示，即局部信息、非局部信息和上下文信息的综合利用，以从充满噪声的感兴趣区域中获得最佳的观察。

相反，现有的基于深度学习的低剂量 CT 图像降噪方法严重依赖于卷积神经网络。卷积本质上是一种局部运算符，在它的滤波器大小内提取一些有用的局部信息。尽管这些基于卷积操作的方法取得了令人印象深刻的性能，但显式的非局部和上下文信息 (正如前面讨论中提到的，这也是放射科医生所利用的) 通常被大多数现有方法所忽略。应该注意的是，放射科医生可能不信任一个工作流程或者操作方式与他们有很大不同的深度神经网络模型。

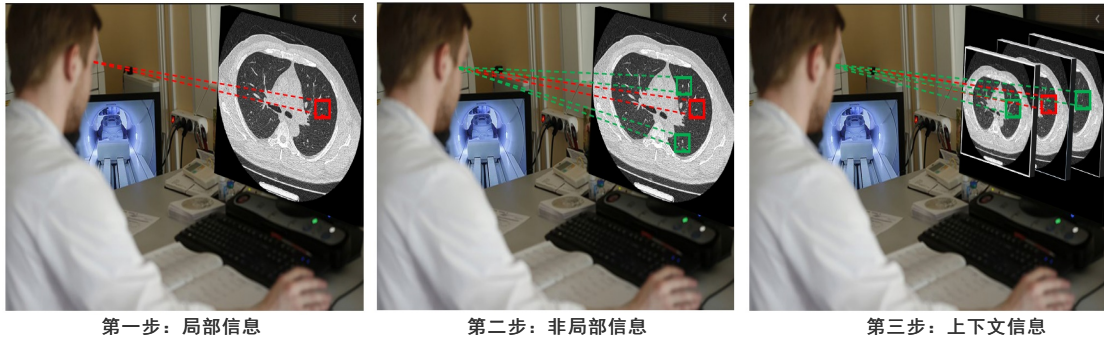


图 5-1 放射科医生阅读低剂量 CT 图像时的工作流程。在步骤 1 中，放射科医生利用局部信息聚焦于感兴趣区域 (Region of Interest, ROI)。在第二步中，放射科医生利用那些非局部但容易观察到的类似组织进行辅助观察。在步骤 3 中，放射科医生将鼠标滑动，观察前后切片相似的 ROI

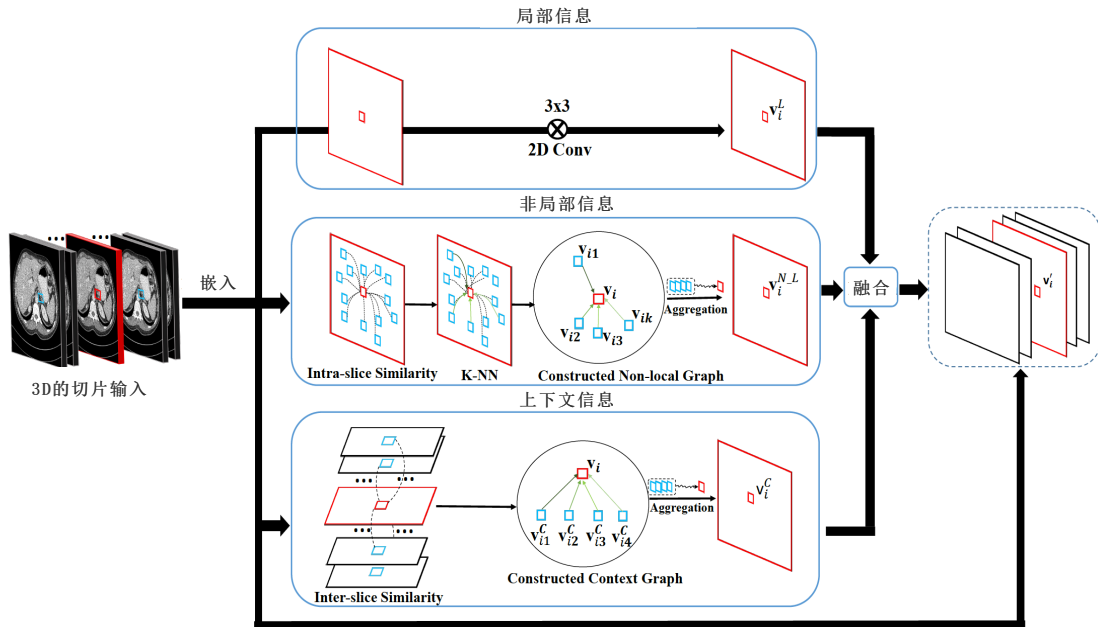


图 5-2 受医生行为启发的低剂量 CT 图像深度降噪网络。在特征图提取之后，步骤一采用经典的二维卷积运算；步骤二进行切片内相似度计算、K 近邻 (K- nn) 计算、非局部图的构造以及特征聚合；步骤三的输入是整个特征图，然后计算片间、上下文图的构造和特征聚合；最后，用特征融合的结果替换原始输入位置的特征图，以保持原始的三维形状

为此，本文首次提出了一种新的深度学习模型，即受放射科医生行为启发的模型 (称为 RIDnet, Radiologist-Inspired Deep Neural Network)，以模拟放射科医生阅读低剂量 CT 图像的工作流程。如图 5-2 所示，遵循放射科医生启发的综合化的



信息聚合方案, 首先引入图卷积网络 (Graph Convolutional Network, GCN), 通过构造非局部图来模仿放射科医生的步骤 2 (如图 5-1), 即显式提取非局部信息, 这是现有的基于卷积神经网络的方法无法直接利用的。此外, 通过建立切片间的上下文图来获取上下文信息, 提出了模拟放射科医生步骤 3 的基于上下文的图卷积神经网络。与卷积神经网络<sup>[47]</sup>隐式提取的上下文信息相比, 基于上下文的图卷积神经网络隐式提取的信息更为显式, 在广泛的应用中得到了更好的效果。在这里, 步骤 2 和步骤 3 的模仿一起组成了一个新的深度学习模块, 即三维图卷积模块。所提出的三维图卷积模块旨在模拟放射科医生的行为, 即考虑切片内的非局部信息和切片间的上下文信息。局部信息也很重要, 因此利用经典的卷积运算来获得有用的局部模式。最后, 提出的 RIDnet 将上述三个部分结合起来, 通过特征融合的方式来学习最佳的自适应组合。需要注意的是, 本文提出的 RIDnet 模型是可以堆叠的, 以进一步提高降噪能力。考虑到综合性能, 本文采用三个 RIDnet 模型构建整个网络 (将在消融实验中讨论了这个问题), 如图 5-3 所示为提出的整个模型框架图。需要指出的是, 提出的 RIDnet 模型遵循放射科医生对低剂量 CT 图像降噪的工作流程, 与广泛采用的基于卷积神经网络的模型不同。

### 5.3 医生阅片行为启发的低剂量 CT 图像降噪算法

#### 5.3.1 数学变量声明

首先, 进行一些数学变量的声明以实现更好地解释。假设有一个低剂量/正常剂量配对的 CT 图像数据集,  $T = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \cdots, (\mathbf{x}_N, \mathbf{y}_N)\}$ , 其中  $N$  表示成对训练样本的个数。 $(\mathbf{x}, \mathbf{y}) \in (X, Y)$  中, 其中  $X$  和  $Y$  分别是两个图像域,  $\mathbf{x} \in \mathbb{R}^{n \times n}$  和  $\mathbf{y} \in \mathbb{R}^{n \times n}$  分别表示带噪声的低剂量 CT 图像和高质量的正常剂量 CT 图像的矩阵表达式。假设  $i$  表示第  $i$  个切片  $\mathbf{x}_i$  是患者整体切片中需要降噪的切片。 $\mathbf{x}_i$  的三维输入可以表示为  $\mathbf{x}_i = \text{concat}(\mathbf{x}_{-s}, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}, \cdots, \mathbf{x}_s)$ , 其中  $2s + 1$  是预定义的切片数, 并按照<sup>[22]</sup>中的建议设置为 3。 $\text{concat}(\cdot)$  操作符表示维度的连接操作符。

#### 5.3.2 医生行为启发的低剂量 CT 图像降噪模型

提出的 RIDnet 模型由嵌入层、局部信息提取层、三维图卷积网络 (3-Dimensional Graph Convolutional Network, 3D-GCN) 和特征融合层四部分组成。

**嵌入层:** 对于模型学习来说, 通常将输入转换成它在特征空间中的嵌入会更有效。每个像素都可以看作是特征向量沿通道方向的形式。为此, 提出的 RIDnet 模型采用了经典的 3D 卷积操作。其中, 第一个 RIDnet 模型有两个 3D 卷积层, 通道数分别为 64 和 32。由于在模型开始时需要更强的特征提取能力, 所以在最开始设

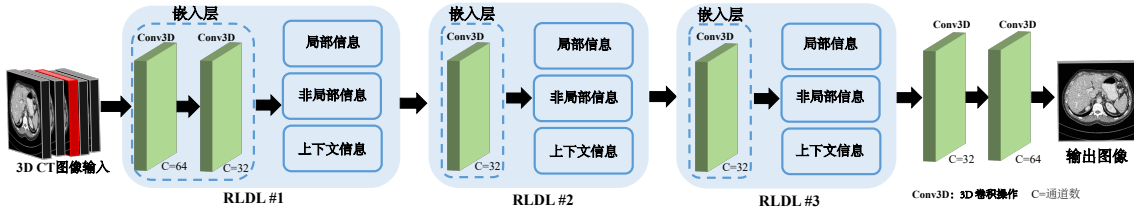


图 5-3 受医生行为启发的堆叠式深度降噪网络，该模型堆叠了三个 RIDnet 模块。最后两个卷积层的目的是生成一个通道数上的编码器-解码器结构

置更多的通道数。需要注意的是，其他的 RIDnet 模型只有一个 3D 卷积层，通道数为 32。为了保持特征图的大小，采用了“反射”填充的操作，这有助于避免在边缘出现不希望的人工伪影。这样，嵌入层的输出可以记为

$$\mathbf{X}'_i = (\mathbf{x}'_{-s}, \mathbf{x}'_{i-1}, \mathbf{x}'_i, \mathbf{x}'_{i+1}, \dots, \mathbf{x}'_s) = \text{embedding}(\mathbf{X}_i). \quad (5-1)$$

**三维图卷积网络:** 三维图卷积网络是所提出的 RIDnet 模型中最重要的部分，由两个模块组成，分别是平面图卷积模块和深度图卷积模块。

第一步是提出顶点  $j$  对顶点  $i$  的重要性的定义：在图像域中，一个像素相对于另一个像素的重要性通常用相似度来度量。不依靠权重矩阵来获得重要性（例如自注意力机制），本文提出在特征空间中使用像素间的欧几里得距离来衡量像素间的重要性。这是受经典的非局部平均降噪算法的启发。作为一种非深度学习方法，基于非局部平均的低剂量 CT 图像降噪方法对真实环境中的复杂噪声水平具有令人印象深刻的自适应降噪能力，这也是现有的基于深度学习的模型所需要的。在形式上，相似度可以计算为

$$e_{ij} = \frac{\|v_i - v_j\|_2^2}{h_{ij}}, \quad (5-2)$$

其中  $h_{ij}$  表示  $v_i$  或  $v_j$  的平方根。为了减少计算复杂度，计算预定义的邻居范围  $N_i$  中的像素  $j$  与中间像素  $i$  的  $e_{ij}$ ，其中  $N_i$  是  $d \times d - 8$  (顶点  $i$  直接相邻的像素总数为 8) 的顶点  $i$  的非局部邻域。与现有文献不同的是，本文并没有根据权重矩阵得来的相似性来利用 CT 切片中的所有像素，而是只在特征空间中选择  $(K-1)$  最近邻 (即 K-NN) 进行信息聚合。通过这样的方式，所提出的算法既避免了无效计算，又可以达到预期的表现。在本文中， $K$  设为 8。最后，提出将  $j$  的所有候选像素的相似性的表示进行标准化，这样顶点之间的重要性就可以很容易地表示出来，即

$$a_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp\left(\frac{-\|v_i - v_j\|_2^2}{h_{ij}}\right)}{\sum_{k=1}^{K-1} \exp(-e_{ik})}. \quad (5-3)$$

从式 5-3 可以观察到,  $v_i$  的邻居顶点  $v_j$  的重要性可以用由特征空间距离导出的概率结果  $a_{ij}$  表示。将每条边  $(v_i, v_j)$  赋值为  $a_{ij}$ 。之后是在这个加权图上聚合那些非局部信息。为了实现特定的任务方式, 受边缘条件卷积<sup>[56]</sup>的启发, 将基于概率的边缘表示引入边缘条件卷积中, 即

$$\begin{aligned} \mathbf{s}_i^{N-L} &= \frac{1}{K-1} \sum_{j=1}^{K-1} (F(a_{ij}, w) \mathbf{v}_{ij} + b) \\ &= \sum_{j=1}^{K-1} \frac{\Theta_{ij} \mathbf{v}_{ij}}{(K-1)} + b, \end{aligned} \quad (5-4)$$

其中  $F^l$  表示以  $w^l$  为参数的网络的输出, 用于对不同的边  $z^{l,j \rightarrow i}$  动态生成权值矩阵  $\Theta_{j,i}^l$ , 而  $b^l$  是一个可学习的偏差。本文采用一个多层感知网络作为  $F^l$ 。

**深度图卷积模块:** 深度图卷积模块旨在显式地获得有用的片间的相似性信息, 这也放射科医生的行为的一种表现。也就是说, 放射科医生可以滑动鼠标来利用切片间的信息, 由此可以帮助对较大噪声的感兴趣区域进行降噪。对于 CT 切片间的特征空间中的每个像素, 构造深度图为  $G_d = (V, E)$ 。假设深度图  $G_d$  中有  $M$  顶点, 对于某个顶点, 用  $V$  表示为  $v_i$ 。每条边是一对顶点, 用  $(v_i, v_j) \in E, \forall i < j$ 。的目标包括表示切片间像素之间的上下文关系, 并对它们进行聚合。本文采用与平面图卷积中相同的聚合方法。将  $M$  设为 3。

**局部信息提取层:** 在实际操作中, 放射科医生首先会将注意力集中在病灶的局部区域。因此, 局部信息对于噪声区域的降噪也至关重要。正如之前的研究, 利用  $3 \times 3$  卷积滤波器进行图像域的卷积运算, 提取需要降噪的特征图中的局部信息。

**特征融合层:** 到目前为止已经获取了非局部信息、局部信息和上下文信息。现在棘手的问题是如何把它们结合起来呢? 对于放射科医生来说, 处理这种综合信息的融合可能是非常容易的。为了解决这个问题, 必须首先回顾放射科医生的工作流程。从直观上看, 与上下文信息相比, 放射科医生更关注需要降噪的切片的信息 (包括非局部信息和局部信息)。此外, 由于不同身体区域或成像供应商的切片厚度可能不同, 切片间的信息的可用性可能不是很稳定 (因为如果一组切片的厚度很大, 切片间的关系对学习来说就不是很强)。因此, 将切片间的上下文信息视为辅助信息。根据前面的讨论, 特征融合的过程可以表示为

$$\mathbf{x}''_i = \alpha \cdot \text{Mean}(\mathbf{p}'_{i,NL} + \mathbf{p}'_{i,L}) + (1 - \alpha) \cdot \mathbf{p}'_{i,C}, \quad (5-5)$$

其中  $\mathbf{p}'_{i,NL}, \mathbf{p}'_{i,L}$  和  $\mathbf{p}'_{i,C}$  表示  $\mathbf{x}'_i$  的非局部信息,  $\mathbf{x}'_i$  的局部信息,  $\mathbf{x}'_i$ , 分别。  $\alpha \in [0, 1]$  是一个可学习的参数。  $\alpha$  被初始化为 0。  $\text{Mean}$  运算符表示逐像素平均运算。

最后提出融合结果  $\mathbf{x}''_i$  来替换原始输入位置的特征图 (如图 5-2 所示), 即

$\mathbf{X}'_i = \text{concat}(\mathbf{x}_{-s}, \mathbf{x}_{i-1}, \mathbf{x}''_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_s)$ , 由于考虑到保持原始 3D 形状和类似的快捷方式 [56]。如图 5-3 所示, 整个模型堆叠了三个 RIDnet 模块。最后两个卷积层旨在生成通道编码器-解码器结构。

### 5.3.3 整体架构和损失函数

提出的医生行为启发的低剂量 CT 图像降噪模型的整体框架是基于的生成对抗网络。具体来说, 所提出的 RIDnet 模型作为生成器  $G$  的角色。鉴别器  $D$  遵循 [15] 中的结构。生成器的损失函数由对抗损失和感知损失组成:

$$\min_{\theta_G} = -\mathbb{E}_{\mathbf{X}_i} D(G(\mathbf{X}_i)) + \lambda \cdot \mathbb{E}_{\mathbf{X}_i, \mathbf{y}_i} \|\varphi(G(\mathbf{X}_i)) - \varphi(\mathbf{y}_i)\|_2^2. \quad (5-6)$$

与之前的研究一样, 采用 VGG19 [57] 作为特征提取器  $\varphi$ ,  $\lambda$  是一个平衡项, 如 [15] 中所建议的那样设置为 0.1。 $\theta_G$  表示生成器的参数。鉴别器的损失函数遵循基于梯度惩罚 [49] 的生成对抗网络结构。

## 5.4 实验结果及其分析

### 5.4.1 数据集

正如之前的研究 [49], 采用 2016 年 NIH-AAPM-Mayo Clinic 低剂量 CT 图像质量恢复挑战赛中发布的公共的低/正常剂量 CT 图像数据集用于训练。该数据集包括在类似扫描仪的模型 (Somatom Definition AS+, 或以单源模式运行的 Somatom Definition Flash, Siemens Healthcare, Forchheim, Germany) 上获得的 10 名患者的腹部检查结果作为基础。正常剂量 CT 图像是在 120kV、50mAs 的设置下获取的。需要注意的是, 该数据集是通过插入泊松噪声模拟低剂量的 CT 图像获得的。最后的结果可以达到正常剂量 CT 图像的 25% 的噪声水平<sup>①</sup>。具体来说, 使用 6 个患者的低/正常剂量的 CT 图像对作为训练集, 4 个患者的 CT 图像作为验证集。为了平衡学习效率和内存消耗, 将每张大小为 512×512 的 CT 图像随机划分为不重叠的 64×64 子图。子图的总数约为 27K。

### 5.4.2 双盲评分实验的测试集

为了充分比较提出的模型与其他最先进模型的降噪性能。采用具有不同检查区域的低/正常剂量配对的临床 CT 数据集<sup>②</sup>, 用于的测试研究。该测试数据集由 Mayo Clinic 发布, 包括为急性认知或运动缺陷患者采集的 99 次非对比头部 CT 扫

① [https://www.aapm.org/Gr and Challenge/LowDoseCT/#](https://www.aapm.org/Gr_and_Challenge/LowDoseCT/#)

② <https://wik i.cancerimagingarchive.net/pages/viewpage.action?pageId=52758026>

描，为筛查高危肺病患者而获得的 100 次低剂量非对比胸部扫描而来。最后包含 100 次腹部增强 CT 扫描，以寻找转移性肝病灶的患者。与<sup>[16]</sup>类似，选择胸部和腹部 CT 图像进行双盲实验。每个部分包括 67 位患者的扫描。正常剂量胸部和腹部 CT 图像分别在 120kV、250mAs 和 100kV、300mAs 设置下进行扫描。对应的低剂量图像是正常剂量的 10% 和 25%。

需要注意的是，采用的测试数据集提供了一份完备的临床诊断报告，包括病灶的标记位置和病灶的诊断类型。基于此，因此选择一个标记有病灶的 CT 切片及其相邻切片（一个病灶切片及其前 4 张切片和后 4 张切片，共 9 个 CT 切片）来表示该患者。因此，可以研究不同降噪结果对病灶的影响。

有两点需要注意：首先，训练集和测试集的成像参数、设备供应商和获取位置存在差异，这很好地反映了临床复杂的环境。为了获得有竞争力的降噪性能，基于深度学习的模型必须具有良好的自适应和泛化能力。这是双盲测试的动机之一，即基于深度学习的降噪模型是否在临床上具有有效性，以及哪些模型在复杂条件下的降噪能力最好。其次，双盲实验中的测试图像参考了大量带有标记病灶的 CT 图像，这对于放射科医师来说非常有意义，可以让他们评估降噪结果是否会影响对病灶类型或图像特征的判断。

### 5.4.3 双盲阅读研究的细节

每个选定患者的低剂量 CT 图像由三个基于深度学习的模型（MAP-NN<sup>[16]</sup>、RedCNN<sup>[15]</sup> 和提出的 RIDnet）进行降噪。对于每个患者，可以获得 5 个子文件夹，即 3 个不同降噪结果的子文件夹、低剂量 CT 图像子文件夹和正常剂量 CT 图像子文件夹。降噪结果的子文件夹是随机命名的（例如 measure1、measure2 和 measure3）。最后，共获得 134 个文件夹进行双盲实验。三位经验丰富的放射科医师进行整个实验。这三个医生分别来自四川大学华西第一人民医院和华西第四人民医院。

对于评价标准，在图像噪声、结构保真度和总分方面采用了与之前研究<sup>[14]</sup>相同的 4 分制方案。具体而言，1 分表示 CT 图像的质量不能用于临床诊断。2 分表示 CT 图像只能提供有限地诊断信息。3 分表示 CT 图像可以接受，可以提供平均地诊断信息。4 分表示 CT 图像在准确诊断和解释方面具有良好的质量。与<sup>[16]</sup>中的双盲实验不同的是将总分作为评估的一部分。这是由于某些结果在感兴趣区域方面可能具有非常好地降噪性能，但在非感兴趣区域方面可能是次优的，因此放射科医师可以使用总分来反映真实的图像质量。

表 5-1 本文采用的基于深度学习的对比方法细节

方法	RED-CNN [TMI,2017]	MAP-NN [Nature MI,2019]	RIDnet (本文提出的方法)
主干网络	卷积神经网络	卷积神经网络	图卷积神经网络 + 卷积神经网络
模拟医生行为	否	否	是
训练集	AAPM-Mayo	AAPM-Mayo	AAPM-Mayo

#### 5.4.4 训练细节

对于训练集的预处理细节，根据不同的身体区域，使用不同的窗口宽度和窗位。具体来说，将 CT 域中的数据（提供的数据集为 DICOM 格式）归一化为 0 到 1。对于腹部扫描，窗口宽度和窗位分别为 400、40。对于胸部扫描，窗口宽度和窗位分别为 1500、-600。模型的输出被重新归一化到相应的 CT 域范围以生成降噪后的 DICOM 文件。对于模型训练的细节，使用 tensorflow<sup>①</sup> 来构建提出的 RIDnet 模型。

对于模型结构来说，采用 Adam 优化器<sup>[58]</sup> 用于以指数衰减的学习率优化模型的参数。生成对抗网络采用已有论文<sup>[48]</sup> 中的框架，对于生成器的初始学习率设置为  $1 \times 10^{-4}$  和鉴别器的初始学习率设置为  $4 \times 10^{-4}$ 。为了平衡内存消耗和学习效率，批量大小设置为 32。该模型使用 4 个 Geforce 1080Ti GPU 训练了 40 个 epoch 并最终收敛。选择了在验证集上性能最好的模型进行测试实验。

#### 5.4.5 实验结果和分析

为了比较具有代表性的基于深度学习的方法，本文进行了一项双盲研究。比较的代表性方法包括 RED-CNN<sup>[15]</sup>、MAP-NN<sup>[16]</sup> 和提出的 RIDnet 模型。相应的细节可以在表 5-1 中找到。RED-CNN 使用均方误差作为损失函数，因其良好的噪声抑制而被广泛采用。通过系统的双盲实验，MAP-NN 与商业迭代重建方法相比显示出具有竞争力的降噪性能。应该注意的是，提出的模型是唯一尝试考虑放射科医生行为的模型。为了公平，所有的方法都是在 AAPM-Moyo 数据集上训练，其中包括 10 位患者的腹部扫描（成对的正常剂量和低剂量（正常剂量的 25%）CT 图像）。

本文利用一个单独的数据集，即低剂量 CT 图像-PD 数据集，来评估不同方法在临床场景复杂环境中的降噪能力。在低剂量 CT 图像-PD 数据集中，选择的测试数据共包括 67 对正常/低剂量（正常剂量的 25%）腹部扫描和 67 对正常/低剂量（正常剂量的 10%）胸部扫描。可以很容易地注意到，测试集中低剂量胸部扫描的

① [www.tensorflow.org](http://www.tensorflow.org)

剂量水平远低于训练集中的剂量水平。因此，这些测试模型必须具有适应和泛化的能力来处理这种不同的环境，将在实验分析中进一步分析这一效应。与以往的双盲实验不同，采用不同方法的降噪性能完全在病灶区域进行评估。

本文相信这种设置可能对临床环境更有价值。具体来说，腹部和胸部 CT 图像分别有转移性肝病灶和肺结节病灶。来自华西第一人民医院和华西第四人民医院的三位经验丰富的放射科医师参与了本次研究。他们分别被命名为“radiologist #1”、“radiologist #2”和“radiologist #3”。本文采用 4 分制度量表，即从噪声、保真度和总分三个方面评估降噪性能。简而言之，分数越高越好。

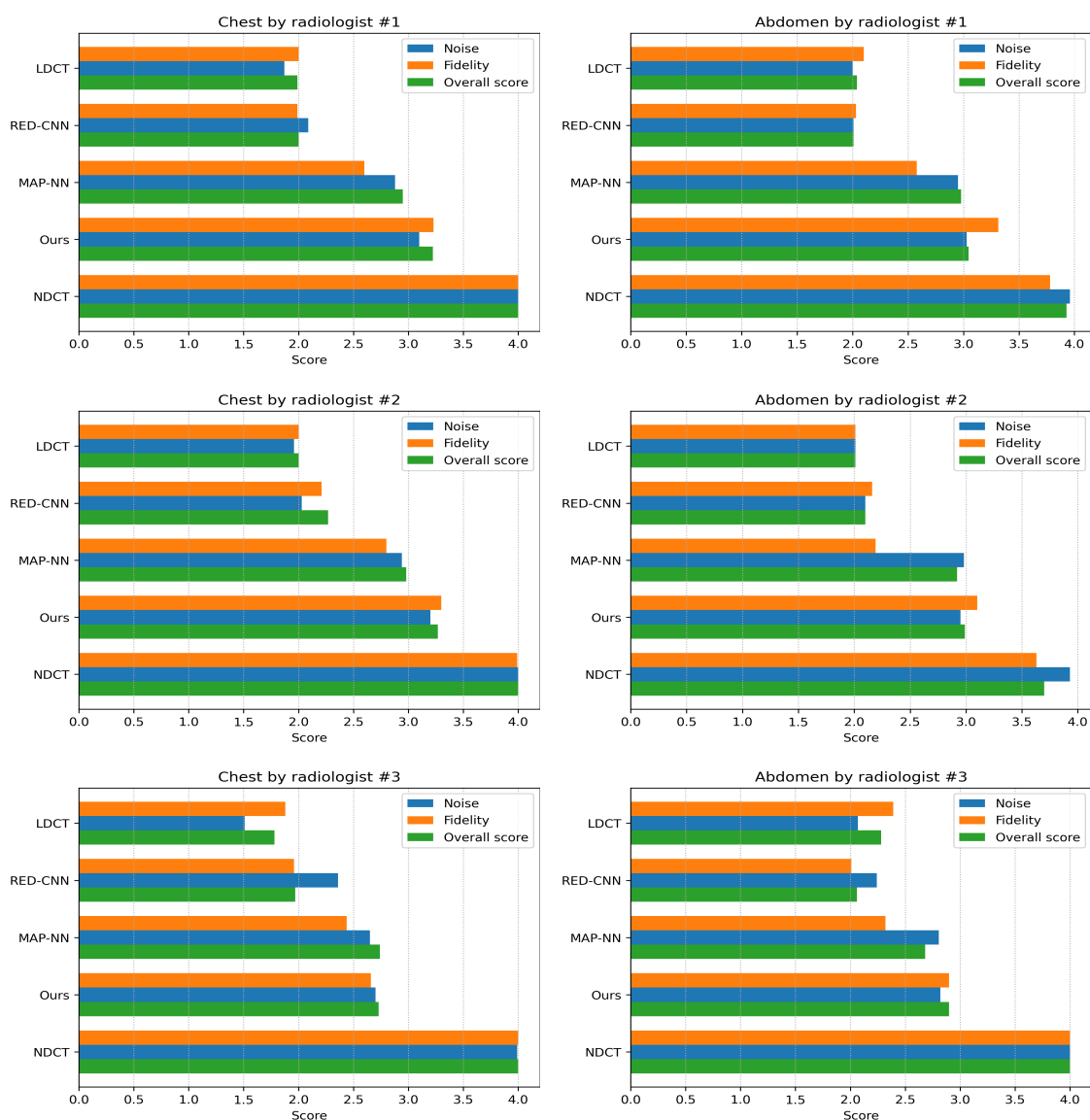


图 5-4 由三位放射科医师对胸部和腹部进行双盲研究的结果。对于噪声、保真度和整体评分来说，得分越高越好，满分为 4 分。Chest: 胸部；Abdomen: 腹部

### 胸部扫描结果的分析

如图 5-4 的第一列所示，可以总结出一些实验发现。首先，提出的模型在噪声、保真度和总体得分方面明显优于其他方法，除了“radiologist #3”的两个结果（噪声和保真度）略有改善。这显示了整合放射科医生启发的降噪机制的优越性。其次，可以注意到 RED-CNN 明显落后于 MAP-NN 和提出的模型。尽管 MSE 引导的架构（即 RED-CNN 采用的）会直观地生成非常平滑且噪声很小的结果，但放射科医生可能觉得这种方式对于诊断的改进非常有限。相反，与低剂量 CT 图像相比，生成对抗网络引导的架构的结果有了明显地改进。第三，与 MAP-NN 相比，尽管使用相同的框架，提出的模型仍然取得了更好的性能。得益于提出的模型中引入了模拟放射科医生工作流程的结构（例如非局部和上下文信息）使得整体模型更为有效。第四，因为测试集的剂量水平与训练集的剂量水平不匹配，本文提出的模型更好的结果反映了提出的模型具有更好的自适应和泛化能力。这一特性将在复杂的临床环境中发挥巨大的潜力，特别是对于各种成像参数、设备供应商。

### 腹部扫描结果分析

图 5-4 的第二列显示了腹部的评估结果。可以做出以下的实验结果总结：首先，将三个指标的结果放在一起对比，提出的模型实现了综合上最佳的性能，尤其是在保真度评估这一方面。这也极大地受益于模拟放射科医生行为的模型。同时，所提出的模型利用综合的信息表示，与单一信息（如 RED-CNN 和 MAP-NN）相比，这对结构的保真度在直觉上是非常地有用的。其次，由于转移性肝的病灶通常以黑色区域为特征，因此灰度和形状的保真度对诊断非常重要。有趣的是，提出的模型的保真度评估在所有放射科医师的结果中明显优于其他模型。这将有助于诊断一些无法观察到的病灶。

**得分高于低剂量 CT 图像的结果分析**除了不同方法对不同身体部位的平均得分，得分高于低剂量 CT 图像图像的病例数对于指示在整个测试环境下的降噪结果的整体性能也很重要。如图 5-5 的第一列所示，可以做出如下的结论：首先，就“radiologist #1”和“radiologist#2”的结果来说，可以发现无论噪声、保真度和总体得分如何，提出的模型基本上在所有情况下都实现了性能的提升。虽然 MAP-NN 在噪声和整体得分上也取得了不错的成绩（提出的模型仍然有一点优势），但保真度的提升并不能完全反映整体的性能。直观地说，平衡降噪和结构保真度是一件十分困难的事情。MAP-NN 的实验结果清楚地说明了这一挑战的存在。

相反，提出的模型享有额外的非局部和上下文信息，这表明通过集成放射科医生启发的工作流程可能更适合应对这一挑战。其次，由于放射科医师之间可能存在偏好，因此改进并未涵盖放“radiologist#3”结果中的所有病例。然而，提出



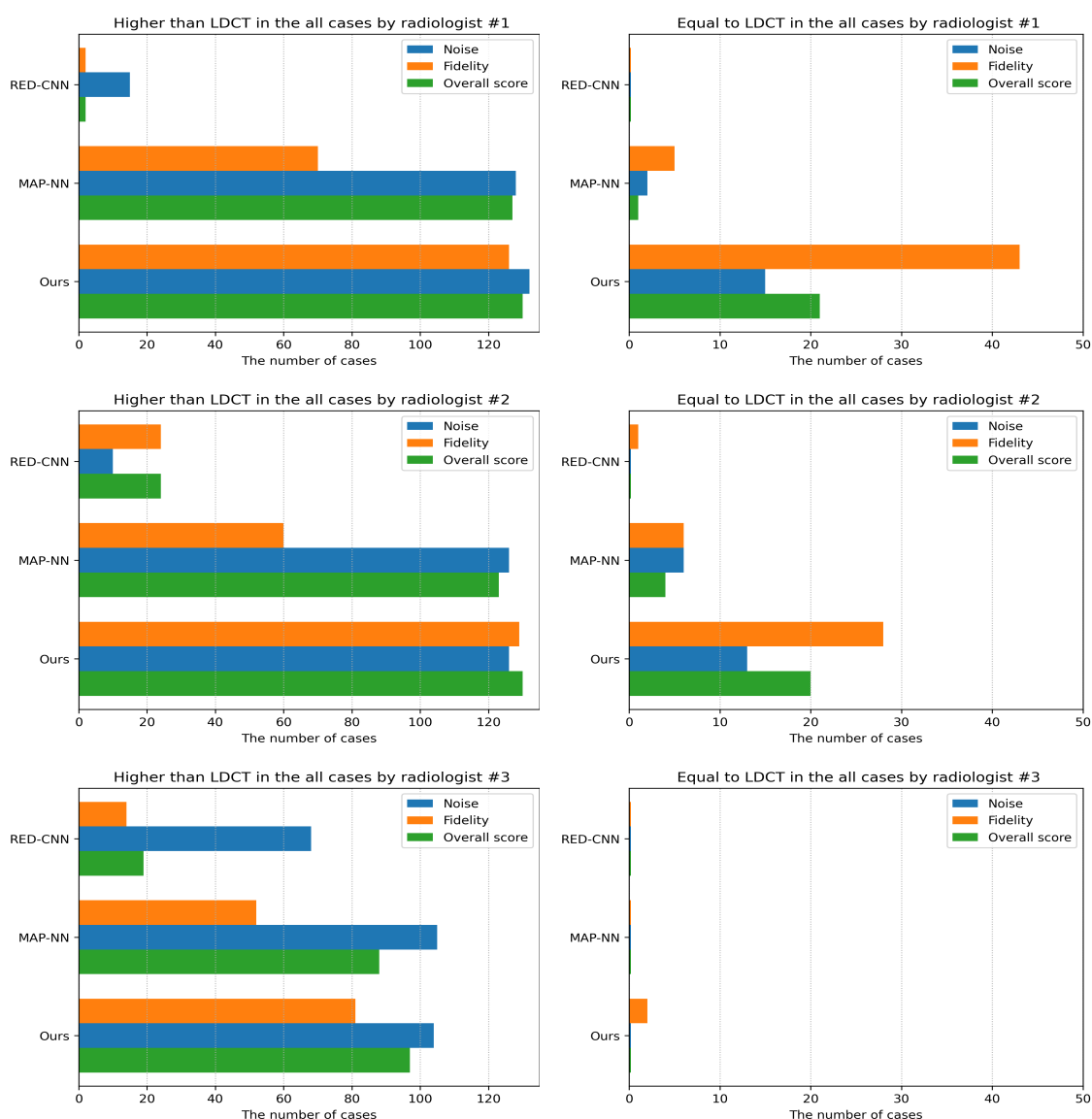


图 5-5 由三位放射科医师对胸部和腹部进行双盲研究的结果。对于噪声、保真度和整体评分来说，得分越高越好，满分为 4 分。Chest: 胸部；Abdomen: 腹部

的模型在保真度方面仍然明显优于其他模型。

### 得分等于正常剂量 CT 图像的结果分析

本文同时感兴趣有多少降噪的案例可以达到正常剂量 CT 图像的水平，这可以进一步反应降噪能力。通过分析所有放射科医生的结果，可以发现本文所提出模型实现了最佳性能。如图 5-5 的第二列所示，发现在“radiologist#3”的结果中，提出的模型是唯一可以达到正常剂量 CT 图像水平的模型，且图像的保真度也十分好。总之，受益于放射科医师启发行为的整合，提出的模型具有实现了正常剂量 CT 图像质量的潜力。

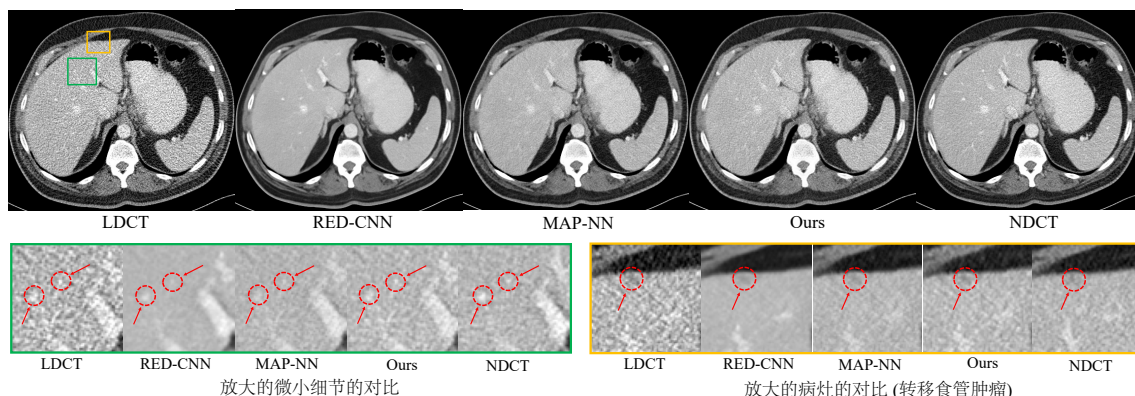


图 5-6 以腹部为例，放大细微结构和病灶的视觉比较。绿框和黄框分别为细微结构放大区域和病灶放大区域。红色的圆圈和箭头是用于比较的建议区域

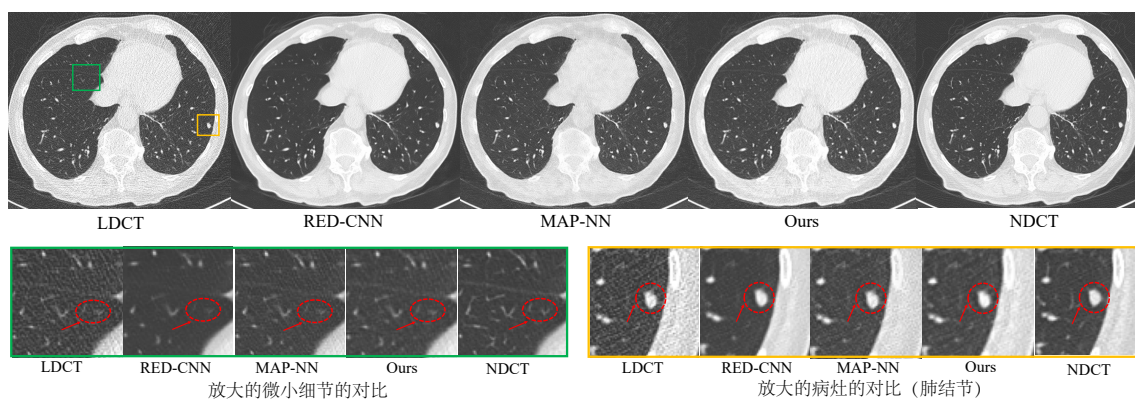


图 5-7 放大的细微结构和病灶的视觉比较，以胸部为例。绿色框和黄色框分别是放大的细微结构区域和放大的病灶区域。红色圆圈和箭头是建议的比较区域

### 放大的细微结构和病灶的可视化分析结果

胸部: 如图 5-6 所示，可以得到一些重要的结果。首先，绿色框显示了放大的细微结构的比较图，可以发现这些结构（在红色圆圈内）在 RED-CNN 的结果中几乎消失了。这可能是因为降噪的结果过度平滑。类似地，MAP-NN 也存在同样的问题，并且也有轻微的过度平滑。相反，提出的模型极大地保留了细微结构并产生了最接近正常剂量 CT 的纹理。深入分析可以发现，RED-CNN 和 MAP-NN 都只利用基于卷积神经网络的局部信息，因此很难在有限的信息中平衡局部细节和降噪效果。相反，本文提出的模型采用了受放射科医生行为启发的综合的信息集成框架，这些额外的信息可以作为补充来产生最佳结果，而不管结构和纹理如何。

其次，如图 5-6 中黄色框的红色圆圈所示，提出的模型对病灶（诊断为转移-食管瘤）的降噪结果最为明显，尤其是它的灰度水平。然而，可以发现 RED-CNN 和 MAP-NN 中的病灶显得非常模糊。病灶区域降噪结果的优越性进一步证明了放

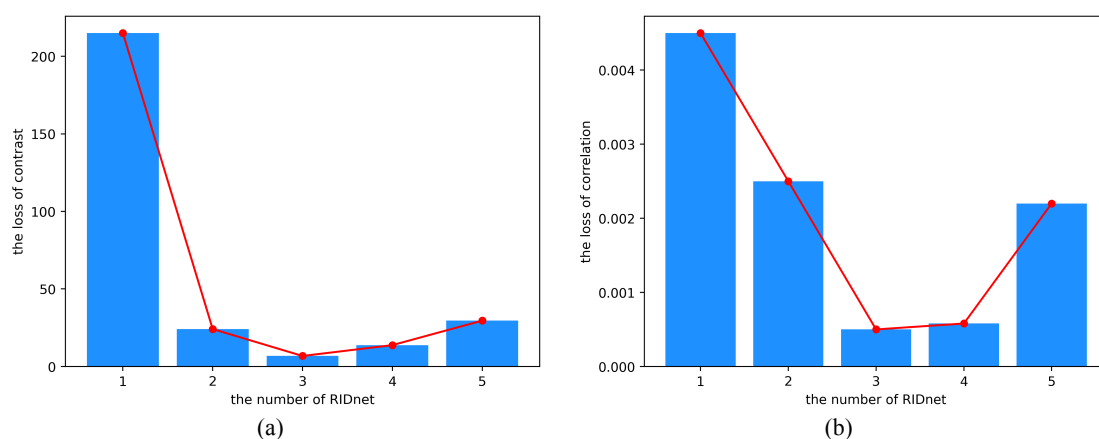


图 5-8 最佳 RIDnet 数量的结果：(a) 对比度损失与 RIDnet 数量之间的关系；(b) 相关性损失与 RIDnet 数量之间的关系

射科医师启发框架的效果。肺部: 如图 5-7 黄色方框所示, 可以发现所有的模型都实现了病灶的较好保留 (如红圈内)。然而, 与 RED-CNN 和提出的模型相比, MAP-NN 更容易观察到噪声点。RED-CNN 基本损失了微妙的结构 (如绿框红圈所示)。总的来说, 提出的模型具有最令人印象深刻地可视化的降噪效果, 获得了在双盲实验中比较好的综合性能。

#### 5.4.6 消融实验

**RIDnet 的最优数量** 确定多少 RIDnets 能够取得最优的降噪效果是非常重要的。为此, 本文构建了 5 个模型, 它们分别有 1、2、3、4、5 个 RIDnet (s) 模型。对于评估指标, 本文使用影像组学指标 (即 Radiomics 指标) 来衡量。这些指标包括对比度损失和相关性损失。对比度损失反映了图像的清晰度和纹理的局部特征。相关性损失度量局部灰度值的相似性。

利用正常剂量 CT 图像对比度/相关性值与降噪结果之间的绝对损失来评估降噪的性能。对于对比度损失和相关性损失, 值越低越好。如图 5-8 所示, 可以发现, 与其他模型相比, 采用 3 个 RIDnets 的模型性能最好。因此, 本文采用的模型是基于三个 RIDnets 模型的架构, 如图 5-3 所示。

**K 的最优值** 在本文中,  $K$  表示一个像素的非局部信息的数量。从图 5-9(a) 和 (b) 可以看出, 随着  $K$  的增加, 对比度损失逐渐减小。这显示了非局部信息对模型的有效性。与此同时, 过多的非局部信息可能会产生负面的效果。当  $K$  大于 8 时, 对比度损失和相关性损失会逐步增加。为此,  $K$  的数目被设置为 8。

**平均绝对损失和感知损失随训练的变化过程** 如图 5-10(a) 所示, 可以观察到平均绝对损失逐渐的收敛。同时, 感知损失也逐步的振荡收敛, 如图 5-10(b) 所示。

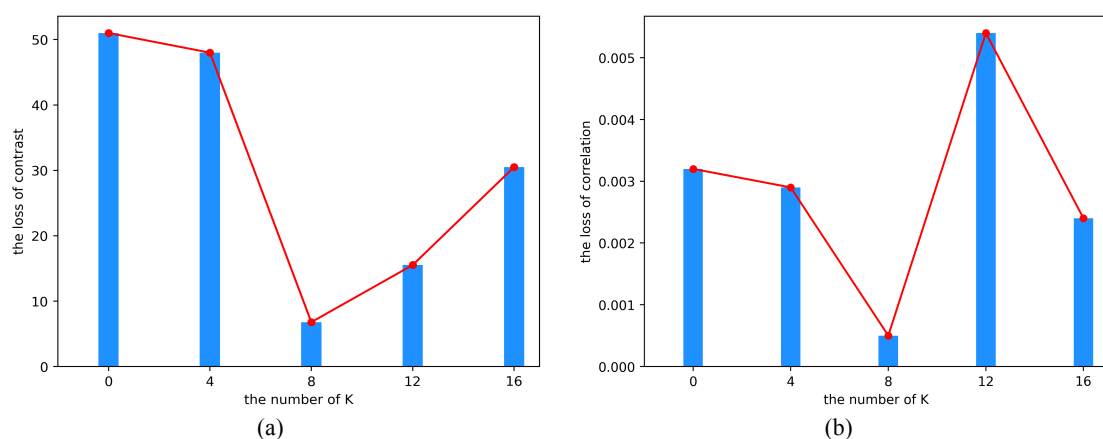


图 5-9 最优的  $K$  的结果：(a) 对比度损失与  $K$  数量之间的关系；(b) 相关性损失与  $K$  个数的关系

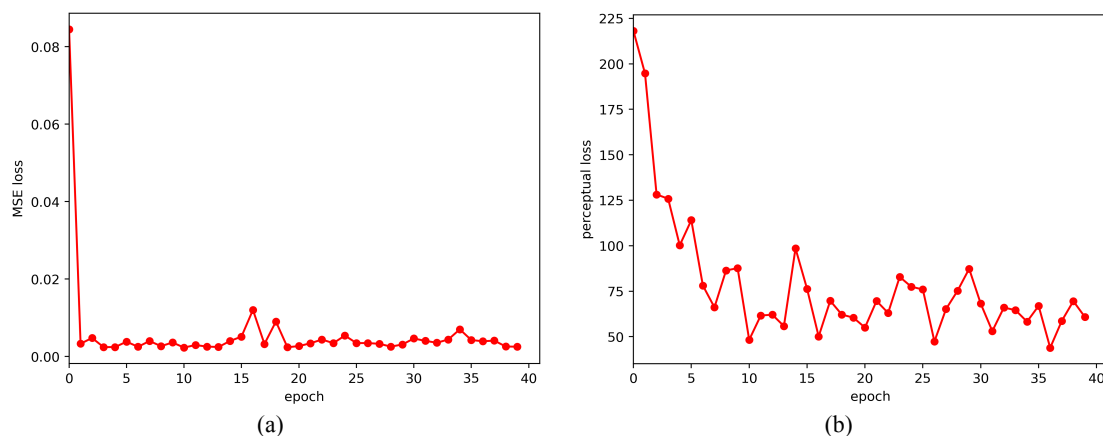


图 5-10 模型训练分析：(a) 平均平方误差损失收敛的过程；(b) 感知损失收敛的过程

### 5.4.7 讨论

与现有的基于深度学习的低剂量 CT 图像降噪方法相比，提出的方法的优越性可以总结为四个部分。

- 首先，提出的模型旨在通过全面整合低剂量 CT 图像降噪的局部、非局部和上下文信息，深入模仿放射科医生的行为及其工作流程。相比之下，现有模型缺乏放射科医生的行为或工作流程的介绍。
- 其次，通过放射科医师启发的工作流程，尽管临床环境复杂，所提出的模型仍显示出最佳的自适应和泛化能力（如图 3-4 所示，提出的模型在训练集和测试集之间剂量不匹配的挑战下具有最具竞争力的性能）。
- 第三，如何平衡细节保留和降噪水平一直是现有基于深度学习的模型的难

题。然而，提出的模型不仅保留了细微的结构和病灶，而且达到了最接近正常剂量 CT 的纹理（纹理可以看作是降噪水平。如果降噪水平很高，降噪结果的纹理将是非常光滑）。这种自适应能力可能得益于图卷积的引入，它也可以被视为可学习的非局部平均算法的扩展。<sup>[3]</sup>（作为非深度学习方法具有完美的自适应能力）。

- 第四，的双盲研究完全基于对病灶区域的评估。认为这对临床目的更有价值。

根据已有文献，本文所提出的模型是第一个尝试将放射科医生启发的工作流程集成到深度神经网络中的模型。在未来，认为应该考虑更多潜在的放射科医师的行为，以进一步改进，例如泛化能力和鲁棒性。此外，在真实的临床环境中，各种扫描条件（如供应商、重建类型、成像参数的差异）不可避免地对低剂量 CT 图像图像降噪提出了巨大的挑战。

在实践中，强烈建议使用双盲实验来评估不同的降噪方法的可用性，因为现有的客观指标，例如峰值信噪比 (PSNR) 和结构相似性指数测量 (SSIM) 不能完全反映降噪结果的质量（例如，过度平滑的结果通常具有较高的 PSNR 分数）。然而，双盲实验对于放射科医生来说将花费大量时间。因此，需要一些更好的客观指标，尤其是对于病灶区域的评价指标。

## 5.5 本章小节

在本章中，一种基于医生行为启发的低剂量 CT 图像降噪模型被首次提出，该模型可以模拟放射科医生阅读低剂量 CT 图像的工作流程。一项基于公共临床数据集的双盲实验显示了该模型的有效性。作为一个受医生启发的模型，RIDnet 给出了一个新的研究路线图，在设计辅助临床诊断的决策支持工具时，考虑了医生的行为。

## 第六章 全文总结与展望

### 6.1 全文总结

关注低剂量 CT 图像降噪问题，从临床的真实场景可能存在的问题出发，详细分析了现有的基于深度学习的低剂量 CT 图像降噪任务的一些局限性，如模型缺乏解释性、难以实现盲降噪、降噪任务与下游的高阶任务脱节等问题进行解决。主要的提出了 3 个新颖的算法，如下：

1. 创新性提出基于噪声学习的无监督低剂量 CT 图像降噪新框架。现有方法大多严重依赖仿真的成对低/正常剂量 CT 数据集，提出一种无监督低剂量 CT 图像降噪方法；提出低剂量 CT 图像噪声学习技术，旨提取 CT 图像的真实噪声并扩充样本；利用丰富易得的正常剂量 CT 数据集和学习到的噪声样本，构建配对数据集。该方法能学习 CT 图像的真实噪声，可较好地克服基于仿真数据集噪声模式单一的缺点。

2. 创新性提出将低剂量 CT 图像降噪与下游高阶任务关联的新架构。现有方法大多忽略下游的高阶任务，本文提出利用一个联合的损失函数将上下游任务进行有效地连接。利用病灶检测任务关注的感兴趣区域，提出了一种基于感兴趣区域的感知损失，用于降噪效果的评价和损失计算，使降噪过程充分考虑下游的病灶检测效果。针对降噪模型与病灶检测模型训练的不适配问题，提出一种协同训练的策略。

3. 首次提出基于医生阅片行为启发的低剂量 CT 降噪模型。针对现有的基于深度卷积神经网络模型偏重局部信息忽略非局部信息的问题，通过总结放射科医生阅片的流程，将医生对非局部信息、局部信息、切片间信息的综合考虑行为抽象为模型表征。创新性的引入了图卷积神经网络用于非局部信息和切片间信息的提取；同时将传统卷积操作提取的局部信息与上述两者信息进行有机融合，实现综合考虑非局部和局部信息的降噪模式，提升降噪过程的可解释性。提出的模型从医生的阅片行为出发，相比现有模型，可能更受到医生的信任，同时具有较好地降噪效果。

同时，进行详细的主观评价和客观量化对比，说明所提算法的合理性和有效性。

## 6.2 后续工作展望

对于医生阅片行为启发的低剂量 CT 图像降噪算法的未来工作，可以尝试引入一些更具透明的模型来提升模型的整体可解释性，如引入矩阵分解的思想，将低剂量 CT 图像的字典获得，并构建编码矩阵来恢复降噪后的 CT 图像。尝试对模型的每一层进行可视化，体现出每一层学习到信息的本质。

对于基于噪声学习的无监督低剂量 CT 图像降噪算法的未来工作，可以尝试引入更好的噪声提取方式，实现非平稳和非均一的噪声提取方式，以实现更好地无监督问题到监督问题的转换；可以尝试直接引入无监督的低剂量 CT 图像降噪架构，诸如 cycleGAN 等无监督低剂量 CT 图像降噪方法都取得了不错的效果。

对于基于上下游任务关联的低剂量 CT 图像降噪算法的未来工作，可以尝试连接更多的下游任务，诸如病灶的分割和识别等，以实现更多的下游的任务信息引入到上游的降噪任务中。



## 致 谢

时间过得很快，一晃就要在成电呆三年了。这三年我从成都理工大学大学来到电子科技大学，感受到了身边人的优秀和无时无刻需要的谦逊的态度。求学生涯总是充满快乐的，希望我的这种快乐能够不断延续。

感谢我的恩师，蒲晓蓉教授对我的学术和生活的关心和帮助，使我受益匪浅。还记得我在上本科的时候，蒲老师细致的给我改论文的经历使我明白了学术的严谨和态度。蒲老师提供给我各种有形和无形的帮助，包括写第一篇论文时细致的语法、用词的讨论，给予我和校外高水平学者的合作机会 (在此也一并感谢香港城市大学 Haoliang Li 博士, 也希望我在香港城市大学的求学之路顺利), 为我协调计算资源等。如果没有蒲老师, 我也没有机会完成本篇毕业论文和诸多支撑的学术论文。如果没有蒲老师, 我也没有继续进一步深造的条件和视野。最后, 感谢蒲老师这三年来给我提供的优越的条件!

同样, 感谢任亚洲副教授对我的学术上的巨大帮助。基本每一篇论文的创新点提出、内容打磨、完善和投稿前的校对等, 都少不了任老师的建议和指正, 我也深深地学习到很多。希望以后能够有更多的合作的机会。

同时, 科研不是研究生生活的全部, 也伴随着和健康大数据研究所所有同门的快乐时光, 他们是: 龙坤、张辰翰、沈望舒、唐磊师兄, 罗纪翔、平明、赵越、李聪硕士, 邵剑剑、黄宗莫、刘军池、张鼎文、魏笑雨师弟, 黄佳欣、王梦洁师妹。每次和大家在“川工第一漂”、“大懒人冒菜”、“豆花冒菜”、“万州烤鱼”、“龙湖猪脚饭”、“经典螺蛳粉”都充满了欢声笑语, 希望以后大家有时间再聚一堂。

感谢我的朋友们, 特别是远在西安电子科技大学的郭志成、在成都的刘哲谛、李攀, 和你们一起分享生活中、工作中、学习中的苦涩和快乐的时光尤为难忘。

最后, 感谢一直在背后默默支持我的父母。从小到大, 父母总是教导我要不断充实自己, 让我明白知识就是财富的真谛, 同时感谢父母让我可以无所顾忌的继续求学之路。衷心感谢一直陪伴我整个硕士生涯挚爱的女友王明越女士, 山水一程, 冷暖相知, 我也不断的从你那里学会了很多比科研更宝贵的道理, 愿与你携手余生!

感谢评阅老师和专家在百忙之中对论文的评阅!



## 参考文献

- [1] A. Manduca, L. Yu, J. D. Trzasko, et al. Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct[J]. *Medical physics*, 2009, 36(11): 4911-4919
- [2] K. Chen, X. Pu, Y. Ren, et al. Low-dose ct image blind denoising with graph convolutional networks[C]. *International Conference on Neural Information Processing*, 2020, 423-435
- [3] J. Wang, H. Lu, T. Li, et al. Sinogram noise reduction for low-dose ct by statistics-based nonlinear filters[C]. *MIIP*, 2005, 2058-2066
- [4] K. Chen, K. Long, Y. Ren, et al. Lesion-inspired denoising network: Connecting medical image denoising and lesion detection[C]. *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, 3283-3292
- [5] M. Balda, J. Hornegger, B. Heismann. Ray contribution masks for structure adaptive sinogram filtering[J]. *IEEE TMI*, 2012, 31(6): 1228-1239
- [6] J. Ma, J. Huang, Q. Feng, et al. Low-dose computed tomography image restoration using previous normal-dose scan[J]. *Medical physics*, 2011, 38(10): 5713-5731
- [7] Y. Liu, J. Ma, Y. Fan, et al. Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction[J]. *Physics in Medicine & Biology*, 2012, 57(23): 7923
- [8] Y. Zhang, X. Mou, G. Wang, et al. Tensor-based dictionary learning for spectral ct reconstruction[J]. *IEEE TMI*, 2016, 36(1): 142-154
- [9] Y. Chen, X. Yin, L. Shi, et al. Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing[J]. *Physics in Medicine & Biology*, 2013, 58(16): 5803
- [10] M. H. Al-Mallah, A. Aljizeeri, M. Alharthi, et al. Routine low-radiation-dose coronary computed tomography angiography[J]. *European Heart Journal Supplements*, 2014, 16(suppl\_B): B12-B16
- [11] T. Flohr, K. Stierstorfer, S. Ulzheimer, et al. Image reconstruction and image quality evaluation for a 64-slice ct scanner with-flying focal spot[J]. *Medical physics*, 2005, 32(8): 2536-2547
- [12] T. R. Moen, B. Chen, D. R. Holmes III, et al. Low-dose ct image and projection dataset[J]. *Medical physics*, 2021, 48(2): 902-911
- [13] H. Chen, Y. Zhang, W. Zhang, et al. Low-dose ct via convolutional neural network[J]. *Biomedical optics express*, 2017, 8(2): 679-694

- 
- [14] Y. Liu, Y. Zhang. Low-dose ct restoration via stacked sparse denoising autoencoders[J]. *Neurocomputing*, 2018, 284: 80-89
- [15] H. Shan, Y. Zhang, Q. Yang, et al. 3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network[J]. *IEEE Trans Med Imaging*, 2018, 37(6): 1522-1534
- [16] H. Shan, A. Padole, F. Homayounieh, et al. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose ct image reconstruction[J]. *Nature Machine Intelligence*, 2019, 1(6): 269-276
- [17] H. Chen, Y. Zhang, M. K. Kalra, et al. Low-dose ct with a residual encoder-decoder convolutional neural network[J]. *IEEE transactions on medical imaging*, 2017, 36(12): 2524-2535
- [18] M. Arjovsky, S. Chintala, L. Bottou. Wasserstein generative adversarial networks[C]. *International conference on machine learning*, 2017, 214-223
- [19] J. Johnson, A. Alahi, L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution[C]. *European conference on computer vision*, 2016, 694-711
- [20] A. Sengupta, Y. Ye, R. Wang, et al. Going deeper in spiking neural networks: Vgg and residual architectures[J]. *Frontiers in neuroscience*, 2019, 13: 95
- [21] C. You, Q. Yang, H. Shan, et al. Structurally-sensitive multi-scale deep neural network for low-dose ct denoising[J]. *IEEE Access*, 2018, 6: 41839-41855
- [22] M. Li, W. Hsu, X. Xie, et al. Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network[J]. *IEEE Trans Med Imaging*, 2020, 39(7): 2289-2301
- [23] C. You, Q. Yang, L. Gjestebj, et al. Structurally-sensitive multi-scale deep neural network for low-dose ct denoising[J]. *IEEE Access*, 2018, 6: 41839-41855
- [24] Y. Ma, B. Wei, P. Feng, et al. Low-dose ct image denoising using a generative adversarial network with a hybrid loss function for noise learning[J]. *IEEE Access*, 2020, 8: 67519-67529
- [25] K. Chen, J. Huang, J. Sun, et al. Task-driven deep learning for ldct image denoising[C]. *The Fourth International Symposium on Image Computing and Digital Medicine*, 2020, 35-39
- [26] K. Long, L. Tang, X. Pu, et al. Probability-based mask r-cnn for pulmonary embolism detection[J]. *Neurocomputing*, 2021, 422: 345-353
- [27] P. Priyanka, G. Nisarga, S. Raghuram. Cmos implementations of rectified linear activation function[C]. *International Symposium on VLSI Design and Test*, 2018, 121-129

- [28] Z.-Q. Zhao, P. Zheng, S.-t. Xu, et al. Object detection with deep learning: A review[J]. IEEE transactions on neural networks and learning systems, 2019, 30(11): 3212-3232
- [29] A. R. Pathak, M. Pandey, S. Rautaray. Application of deep learning for object detection[J]. Procedia computer science, 2018, 132: 1706-1717
- [30] C. Gsaxner, P. M. Roth, J. Wallner, et al. Exploit fully automatic low-level segmented pet data for training high-level deep learning algorithms for the corresponding ct data[J]. PloS one, 2019, 14(3): e0212550
- [31] A. Du Plessis, J. Strauss, A. Rix. Short-term solar power forecasting: Investigating the ability of deep learning models to capture low-level utility-scale photovoltaic system behaviour[J]. Applied Energy, 2021, 285: 116395
- [32] S. Hickel, N. A. Adams, J. A. Domaradzki. An adaptive local deconvolution method for implicit les[J]. Journal of Computational Physics, 2006, 213(1): 413-436
- [33] F. Tatsugami, T. Higaki, Y. Nakamura, et al. Deep learning-based image restoration algorithm for coronary ct angiography[J]. European radiology, 2019, 29(10): 5322-5329
- [34] K. Zhang, W. Zuo, S. Gu, et al. Learning deep cnn denoiser prior for image restoration[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 3929-3938
- [35] W. Yang, X. Zhang, Y. Tian, et al. Deep learning for single image super-resolution: A brief review[J]. IEEE Transactions on Multimedia, 2019, 21(12): 3106-3121
- [36] C. Dong, C. C. Loy, K. He, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(2): 295-307
- [37] D. Johnson, S. Sinanovic. Symmetrizing the kullback-leibler distance[J]. IEEE Transactions on Information Theory, 2001
- [38] A. Creswell, T. White, V. Dumoulin, et al. Generative adversarial networks: An overview[J]. IEEE Signal Processing Magazine, 2018, 35(1): 53-65
- [39] K. Wang, C. Gou, Y. Duan, et al. Generative adversarial networks: introduction and outlook[J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4): 588-598
- [40] Y. Wang, B. Yu, L. Wang, et al. 3d conditional generative adversarial networks for high-quality pet image estimation at low dose[J]. Neuroimage, 2018, 174: 550-562
- [41] S. Vallender. Calculation of the wasserstein distance between probability distributions on the line[J]. Theory of Probability & Its Applications, 1974, 18(4): 784-786
- [42] I. Gulrajani, F. Ahmed, M. Arjovsky, et al. Improved training of wasserstein gans[J]. Advances in neural information processing systems, 2017, 30

- [43] D. Kang, P. Slomka, R. Nakazato, et al. Image denoising of low-radiation dose coronary ct angiography by an adaptive block-matching 3d algorithm[C]. MIIP, 2013, 86692G
- [44] F. Lin, K. Chen, X. Wang, et al. Denoising stacked autoencoders for transient electromagnetic signal denoising[J]. Nonlinear Processes in Geophysics, 2019, 26(1): 13-23
- [45] K. Perlin. Improving noise[C]. Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002, 681-682
- [46] Y. Yao, T. Liu, B. Han, et al. Dual t: Reducing estimation error for transition matrix in label-noise learning[J]. Advances in neural information processing systems, 2020, 33: 7260-7271
- [47] P. Kim. Convolutional neural network[M]. Springer, 2017, 121-147
- [48] J. C. Venter, M. D. Adams, E. W. Myers, et al. The sequence of the human genome[J]. science, 2001, 291(5507): 1304-1351
- [49] Q. Jin, R. Lin, F. Yang. E-wacgan: enhanced generative model of signaling data based on wgan-gp and acgan[J]. IEEE Systems Journal, 2019, 14(3): 3289-3300
- [50] Y. Zhao, Y. Li, X. Dong, et al. Low-frequency noise suppression method based on improved dncnn in desert seismic data[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 16(5): 811-815
- [51] K. Zhang, W. Zuo, L. Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4608-4622
- [52] Q. Yang, P. Yan, Y. Zhang, et al. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss[J]. IEEE Trans Med Imaging, 2018, 37(6): 1348-1357
- [53] J. Feng, X. Liu, X. Li, et al. Low-rank tensor minimization method for seismic denoising based on variational mode decomposition[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5
- [54] K. Chen, J. Sun, J. Shen, et al. Ridnet: Radiologist-inspired deep neural network for low-dose ct denoising[J]. arXiv preprint arXiv:2105.07146, 2021
- [55] V. Badrinarayanan, A. Kendall, R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495
- [56] M. Simonovsky, N. Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 3693-3702

- [57] T. Carvalho, E. R. De Rezende, M. T. Alves, et al. Exposing computer generated images by eye' s region classification via transfer learning of vgg19 cnn[C]. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 2017, 866-870
- [58] L.-C. Chen, Y. Zhu, G. Papandreou, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European conference on computer vision (ECCV), 2018, 801-818

## 攻读硕士学位期间取得的成果

### 学术论文:

[1] "RIDnet: Radiologist-Inspired Deep Neural Network for Low-dose CT Denoising." IEEE Transactions on Medical Imaging, 2021. (under review, 支撑第五章内容, 第一作者)

[2] "Low-dose ct image blind denoising with graph convolutional networks." International Conference on Neural Information Processing. Springer, Cham, 2020. (CCF-C类, 支撑第三章内容, 第一作者)

[3] "Lesion-Inspired Denoising Network: Connecting Medical Image Denoising and Lesion Detection." Proceedings of the 29th ACM International Conference on Multimedia. 2021. (CCF-A类, 支撑第四章内容, 第一作者)

[4] "Task-Driven Deep Learning for low-dose CT Image Denoising." The Fourth International Symposium on Image Computing and Digital Medicine. 2020. (EI, 支撑第四章内容, 第一作者)

[5] "TEMDnet: A Novel Deep Denoising Network for Transient Electromagnetic Signal With Signal-to-Image Transformation." IEEE Transactions on Geoscience and Remote Sensing 60 (2020): 1-18. (CCF-B类, SCI, JCR-Q1, 第一作者)

[6] "TEM-NLnet: A Deep Denoising Network for Transient Electromagnetic Signal with Noise Learning." IEEE Transactions on Geoscience and Remote Sensing (2022). (\*共同一作, (CCF-B类, SCI, JCR-Q1, 共同第一作者)

[7] "EEGDnet: Fusing Non-Local and Local Self-Similarity for 1-D EEG Signal Denoising with 2-D Transformer." IEEE Transactions on Neural Systems & Rehabilitation Engineering, 2022. (under review, 通讯作者)

[8] "Machine learning-based method for interpreting the guidelines of the diagnosis and treatment of COVID-19." Sheng Wu Yi Xue Gong Cheng Xue Za Zhi= Journal of Biomedical Engineering= Shengwu Yixue Gongchengxue Zazhi 37.3 (2020): 365-372. (EI, 第二作者(导师一作))

[9] “A semanticbased shorttext fast clustering method on hotline records in chengdu.” 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2019, 516521.(EI, 第三作者)

**国际比赛:**

1. The 5th International Symposium on Image Computing and Digital Medicine (ISICDM 2021) 低剂量 CT 图像质量优化挑战赛, **冠军**. (团队负责人, 排名第一, 支撑第三章和第四章)

**个人称号:**

1. 2021 年, 电子科技大学”**学术青苗**” 称号

**其他奖励:**

1. 2019 年, 电子科技大学新生二等奖学金
2. 2020 年, 电子科技大研究生一等奖学金
3. 2020 年, 电子科技大学优秀研究生
4. 2021 年, 电子科技大学研究生一等奖学金
5. 2021 年, 电子科技大学优秀研究生
6. 2021 年, 国家奖学金