基于深度学习的无参考CT图像质量评估模型

沈菁莹,章浩伟,刘 颖*

上海理工大学健康科学与工程学院,上海

收稿日期: 2025年2月13日; 录用日期: 2025年3月6日; 发布日期: 2025年3月14日

摘要

无参考CT图像质量评估(NR-IQA)旨在建立与放射科医生主观评估高度一致的客观图像质量评估体系。目前诸多临床CT图像数据集没有实际IQA评分,基于此,本研究提出一种基于深度学习的NR-IQA模型,并对其进行验证。NR-IQA模型将卷积模块(CNN)与视觉Transformer模块(ViT)结合,同时训练4个CNN-ViT网络集成作为教师模型,以模拟放射科医生多次主观IQA过程;接着结合知识蒸馏框架,将教师模型的信息蒸馏到1个学生模型(单一CNN-ViT网络)中。本研究结合峰值信噪比(PSNR)和结构相似性(SSIM)两个客观指标来评估CT图像质量,并使用其标注临床胸部CT图像数据集以验证NR-IQA模型。提出的NR-IQA模型综合性能达到2.8070,PLCC为0.9916,SROCC为0.9683,KRCC为0.8471,MAE低至0.0259,MSE仅为0.0010,验证了其预测CT图像IQA精度的优越性。

关键词

无参考图像质量评估,临床胸部CT图像,视觉Transformer,知识蒸馏

A No-Reference CT Image Quality Assessment Model Based on Deep Learning

Jingying Shen, Haowei Zhang, Ying Liu*

School of Health Sciences and Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 13th, 2025; accepted: Mar. 6th, 2025; published: Mar. 14th, 2025

Abstract

No-reference CT image quality assessment (NR-IQA) aims to establish an objective image quality evaluation system that achieves high consistency with radiologists' subjective assessments. Given the current lack of actual IQA scores in numerous clinical CT image datasets, this paper proposed and validated a NR-IQA model based on deep learning. The proposed model integrated convolutional

*通讯作者。

文章引用:沈菁莹,章浩伟,刘颖.基于深度学习的无参考 CT 图像质量评估模型[J].建模与仿真,2025,14(3):186-198. DOI:10.12677/mos.2025.143214 neural network (CNN) modules with visual Transformer (ViT) modules, and trained an ensemble of four CNN-ViT networks as teacher models to simulate radiologists' repeated subjective IQA processes. Subsequently, a knowledge distillation framework was employed to transfer the information from teacher models to a student model (a single CNN-ViT network). This paper combined two objective metrics, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), to evaluate the quality of CT images, and used them to annotate the CT image dataset to validate the proposed NR-IQA model. The comprehensive performance of the proposed NR-IQA model reached 2.8070, achieving a Pearson linear correlation coefficient (PLCC) of 0.9916, Spearman rank-order correlation coefficient (SROCC) of 0.9683, Kendall rank correlation coefficient (KRCC) of 0.8471, with mean absolute error (MAE) reduced to 0.0259 and mean squared error (MSE) as low as 0.0010, validating its superior accuracy in predicting CT image IQA scores.

Keywords

NR-IQA, Clinical Chest CT Images, Vision Transformer, Knowledge Distillation

Copyright © 2025 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <u>http://creativecommons.org/licenses/by/4.0/</u> CO Open Access

1. 引言

计算机断层扫描(Computed Tomography, CT)是一种广泛应用于临床疾病诊断的成像技术[1] [2]。CT 图像质量通常受到成像设备、辐射剂量和患者因素等不同因素的影响,对患者使用高辐射剂量可能会诱 发恶性肿瘤,但低辐射剂量却又会导致 CT 图像质量下降,从而导致诊断偏差[3] [4]。因此,使用 CT 最 重要的原则是将辐射剂量保持在"合理可行范围内尽可能低(As Low As Reasonably Achievable, ALARA)" [5]。为了实现这一原则,我们需要在保证 CT 图像质量的同时,在 CT 图像质量和辐射剂量之间找到最佳 平衡。CT 图像质量评估(Image Quality Assessment, IQA)是优化辐射剂量与权衡图像质量的重要环节。

CT-IQA 分为主观评估和客观评估两大类。主观评估依赖于放射科医生的经验和视觉判断,然而,在 面对大量 CT 图像时,评估过程难以实时进行。此外,不同放射科医生的评估结果也会存在不一致性[4] [6]。因此,需要一种高效稳定的客观 IQA 方法。在客观评估方法中,依据是否需要参考高质量图像,IQA 方法可分为全参考(Full-Reference, FR) [7]-[9]、半参考(Reduced-Reference, RR) [10]-[12]和无参考(No-Reference, NR) [4] [5], [13]-[16]三种。在患者进行 CT 扫描的过程中,出于患者安全考虑,不会使用高辐射 剂量,而较高的辐射剂量才可能获得更好的图像质量,因此,在临床环境中获取高质量 CT 图像较为困 难。相较于 FR-IQA 和 RR-IQA 方法, NR-IQA 方法无需参考高质量图像作为评估图像质量的标准,适用 于临床应用,因此, NR-IQA 成为当前研究热点。

目前深度学习技术在诸多图像处理任务中取得了优异性能,研究者将深度学习技术引入 NR-IQA 问题中。Gao 等人[4]提出了一个新的结合全局和局部信息的 CT 图像 NR-IQA 框架 NR-GL-IQA。NR-GL-IQA 方法采用了自动标记策略来获取大量训练数据,以预训练模型,然后使用少量主观评估数据对预训练模型进行了微调,以预测全局图像质量。接着使用基于感知的图像质量评估器来预测局部区域的质量。 最后,通过结合全局和局部信息来估计整体图像质量。NR-GL-IQA 方法可以准确预测 CT 图像质量,并 且全局和局部 IQA 的组合比单一评估更接近放射科医生的评估。Lee 等人[5]提出了一种基于自监督的 NR-IQA 方法。在该研究中,一个 CascadeR-CNN 目标检测模型尝试定位插入 CT 图像中的虚拟低对比度 目标,检测模型的平均精度(mAP)被用作 CT 图像的质量分数。Lu 等人[13]提出了一种用于冠状动脉 CT 血管级 IQA 的强化 Transformer 网络。研究者利用中心线跟踪算法检测和裁剪冠状动脉的立方区域,通过逐步强化学习剔除模块将与图像质量无关的立方区域剔除,而质量相关的立方区域被聚合并传送到 Transformer 中以预测质量分数。除此之外,很少有研究工作关注基于深度学习的 CT 图像的 NR-IQA, 一个主要原因是需要大量手动标记的 CT 图像数据来训练模型,对放射科医生来说,为大量 CT 图像注释 质量分数既费力又耗时。

因此,本研究提出一种基于深度学习的 NR-IQA 模型,该模型将卷积(Convolutional Neural Network, CNN)模块与视觉 Transformer (Vision Transformer, ViT)模块结合,实验结果表明,通过添加 CNN 模块,能够提升模型性能。同时训练 4 个 CNN-ViT 网络集成作为教师模型,以模拟放射科医生多次主观 IQA 过程,接着结合知识蒸馏框架,将教师模型的信息蒸馏到 1 个学生模型(单一 CNN-ViT 网络)中。本研究 针对深度学习模型训练缺乏大量手动标记的 CT 图像数据,结合 PSNR 和 SSIM 两个客观指标来评估 CT 图像质量,并使用其标注临床胸部 CT 图像数据集(Clinical Chest CT Image Dataset, CCCIDB)来训练提出 的模型。本研究提出的 NR-IQA 模型在将网络尺寸缩小 4 倍的同时,实现了更高的 IQA 准确性,为 CT 图像的质量评估任务提供了新的思路和方法。

2. 数据集与模型构建

2.1. 数据集

本研究使用的 CCCIDB 数据集由上海某医院提供,使用扫描设备 SIEMENS SOMATOM Definition Edge (西门子,德国)采集,采集对象为志愿者。CCCID 数据集包括以常规剂量采集的 10 名志愿者的 CT 图像数据,共计 471 张,所有 CT 图像数据均以 dicom 格式储存。

由于每位患者的数据只包含以常规剂量扫描获得的 CT 图像,为了增加数据集的多样性,本研究通 过向常规剂量 CT 图像中插入泊松噪声以模拟生成低剂量 CT 图像,同时通过向常规剂量 CT 图像添加零 均值独立噪声,以分别在常规剂量与低剂量之间 50%和 75%的剂量水平上合成 CT 图像[17]。原数据集包 含 471 张 CT 图像,在合成三个新剂量的 CT 图像后,CCCID 数据集中 CT 图像数量达到 1884 张。

由于 CCCID 数据集未提供 IQA 评分。因此,本研究通过计算常规剂量和其余三个新剂量 CT 图像之间的 SSIM 和 PSNR 指标,并结合两个指标值作为 CT 图像的实际 IQA 评分[16]。SSIM 和 PSNR 分别从解剖结构保真度和噪声抑制水平两个维度提供了对 CT 图像的量化评价依据,二者的协同分析可为 IQA 提供关键技术支撑[16] [18] [19]。具体来说,本研究分别计算每张 CT 图像的 SSIM 和 PSNR,并通过公式(1),来获取每张 CT 图像的 IQA 评分 Q,其中,PSNR 指标经过归一化处理,*a*₁和 *a*₂是调整每个指标 贡献的权重,本研究中,*a*₁和 *a*₂均为 0.5。

$$Q = a_1 Q_{PSNR} + a_2 Q_{SSIM} \tag{1}$$

最后,数据集被划分为训练集、验证集和测试集,其中8位患者的CT图像用于训练和验证,剩余2 位患者的CT图像用于测试,用于训练和验证的8位患者的CT图像样本80%随机分配到训练集,剩余的 20%分配到验证集。

2.2. 模型构建

本研究所提出的 NR-IQA 模型结合了集成 CNN-ViT 网络和知识蒸馏策略,用于 CT 图像的质量评估。请参阅图 1,模型由 CNN 模块、ViT 模块以及能够将信息从 4 个 CNN-ViT 网络集成得到的教师模型中蒸馏到 1 个学生模型(单一 CNN-ViT 网络)中的知识蒸馏框架组成。



Figure 1. NR-IQA model framework diagram 图 1. NR-IQA 模型框架图

2.2.1. CNN 模块

请参阅图 2, CNN 模块用于对输入的尺寸大小为(512, 512)的 CT 图像进行特征提取。该模块由 4 个 卷积块、1 个下采样层及 1 个裁剪层组成。每个卷积块包括 1 个卷积层,其卷积核大小为 3 × 3,步幅为 1,随后的是批量归一化层及 ReLU 激活函数,CNN 模块中 4 个卷积块的卷积层的滤波器依次分别为 8、 16、32、64。接着通过 1 个卷积核大小为 2 × 2,步幅为 2 的下采样层将卷积块产生的特征图的大小缩小 一半,并生成 3 个输出通道。最后,通过裁剪层将特征图的大小调整为(3, 224, 224),使其尺寸与构建的 ViT 架构的输入相匹配。





2.2.2. ViT 模块

在临床评估过程中,放射科医生主要考虑与特定病理状况相关的图像区域质量,特别是对准确诊断

至关重要的器官或结构的可见性[20]。ViT 是一种基于 Transformer 架构的视觉模型,它将 CT 图像分成 多个小块(Patch),通过自注意力机制来捕捉 CT 图像中的全局和局部关系[21]。这种机制使得 ViT 能够模 拟放射科医生评估 CT 图像的过程,使其能够观察整个图像并专注于对图像质量影响更大的区域。

请参阅图 3,本研究中,ViT 模块接受 CNN 模块输出的尺寸大小为(3,224,224)的特征图,并将其划 分为 *N* = 196 个小块,每个小块的大小为 16×16 像素。这些小块被展平,并通过可学习的线性投影层映 射到 *D* = 768 维度的向量。为了保留定位信息,向每个小块中嵌入 1 个可学习的位置编码。此外,在小块的开头添加了 1 个"类别标记",以聚合所有其他标记的信息,并生成 IQA 评分。本研究使用的 ViT 模块包含 12 个 Transformer 编码器块,每个 Transformer 编码器块由 1 个层归一化、1 个具有 12 个头的多头自注意力机制层、1 个层归一化和 1 个多层感知机组成。在多头自注意力机制层中,每个头将输入转换为三个矩阵:键 *K*、查询 *Q* 和值 *V*。自注意力机制的计算过程如公式(2),其中 *d*_k表示 *K* 和 *Q* 的维度,多头自注意力机制层的输出是每个注意力头输出的拼接。

Attention
$$(Q, K, V) = \operatorname{Softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V$$
 (2)

多层感知机由两个全连接层组成,并在两个全连接层之间使用 GeLU 激活函数。第一个全连接层有 3072 个神经元,第二个全连接层有 768 个神经元。最后,类别标记被送入一个全连接层以预测 IQA 评 分。





2.2.3. 知识蒸馏框架

知识蒸馏旨在将来自大型复杂模型(即教师模型)的信息迁移到更简单、更小的模型架构中,即学生模型[22]-[24]。请参阅图 1,本研究第一阶段在 CCCID 数据集上训练了 4 个 CNN-ViT 网络,并将其集成为教师模型。在第二阶段设计了一个蒸馏损失,教师模型通过蒸馏损失将 4 个 CNN-ViT 网络的信息蒸馏到 1 个学生模型(单一 CNN-ViT 网络)中,最终通过学生模型预测得到 IQA 评分。

教师模型的预测通过对 4 个 CNN-ViT 网络的预测进行平均得到。每个 CNN-ViT 网络在训练过程中

使用 AdamW 优化器,同时添加权重衰减以防止过拟合。此外,本研究在训练过程中引入 p-Huber 损失函数,其定义如公式(3):

$$\operatorname{Loss}_{\operatorname{teacher}} = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} \left[\frac{|\hat{y}_{i} - y_{i}|^{p}}{p} \right], & \text{if } |\hat{y}_{i} - y_{i}| \leq \lambda \\ \frac{1}{N} \sum_{i=1}^{N} \left[\lambda |\hat{y}_{i} - y_{i}| - \frac{\lambda^{p}}{p} \right], & \text{otherwise} \end{cases}$$
(3)

其中, y_i 表示图像 *i* 的实际 IQA 评分, \hat{y}_i 表示图像 *i* 的 CNN-ViT 网络预测 IQA 评分, N 为 CT 图像数量,本研究中 p=1.6, $\lambda=1.0$ 。P-Huber 损失函数在 y_i 与 \hat{y}_i 之间的误差小于 λ 时使用 Lp 范数损失,而在误差大于 λ 时使用修改的 L1 损失,在保留 Huber 损失鲁棒性的同时,通过可调的 p 参数增强了对 CT 图像中不同误差模式的适应能力,是 IQA 任务的理想选择。

使用集成 CNN-ViT 网络的目的是模仿不同放射科医师进行 IQA 评分。通常, CT 图像的质量分数是 通过对所有放射科医师提供的 IQA 评分进行平均得出的。本研究的教师模型中,每个 CNN-ViT 网络类 似于一位放射科医师,学习来自输入 CT 图像的不同特征,以确保多样性,而后通过蒸馏,教师模型学习 到的信息可以有效地转移到学生模型中。本研究中教师模型使用的 4 个 CNN-ViT 网络与学生模型中的单 一 CNN-ViT 网络架构相似,唯一的不同是在最后一层,学生模型预测的是一个包含 4 个值的向量,而不 是单一值。

在学生模型中使用两个目标损失函数进行训练,分别为预测损失和蒸馏损失。在预测损失中,学生 模型的预测 IQA 评分与给定 CT 图像的实际 IQA 评分进行比较,使用 P-huber 损失函数,如公式(4),其 中 \overline{y}_i 表示图像 *i* 的学生模型预测损失阶段得到的 IQA 评分。由于学生模型预测的是一个包含 4 个分量的 向量,学生模型的最终预测是通过对该向量中的值进行平均得出的。

$$\operatorname{Loss}_{\operatorname{student}}\left(\overline{y}_{i}, y_{i}\right) = \begin{cases} \frac{\left|\overline{y}_{i} - y_{i}\right|^{p}}{p}, & \text{if } \left|\overline{y}_{i} - y_{i}\right| \leq \lambda \\ \lambda \left|\overline{y}_{i} - y_{i}\right| - \frac{\lambda^{p}}{p}, & \text{otherwise} \end{cases}$$
(4)

第二个目标函数是蒸馏损失。在该损失中,标签由一个四维向量组成,该向量通过聚合教师模型的四个集成成员的预测生成。本研究引入了一个最大误差阈值,记作 ϵ 。该误差阈值 ϵ 代表教师模型预测的最大可接受误差,用来指导学生模型。具体来说,如果教师模型的平均预测超过了 ϵ ,它将不被视为适合的标签用于学生训练。然而,当教师模型的平均预测误差低于 ϵ 时,将其作为新的实际 IQA 标签向量使用,而不会使学生的预测与真实分布发生显著偏差[25][26]。本研究中, ϵ 值经过实验,取 0.05。蒸馏损失定义如公式(5),其中, \tilde{y} 是教师模型集成成员的平均预测 IQA 评分。

$$\operatorname{Loss}_{\operatorname{distillation}}\left(\overline{y}_{i}, \tilde{y}_{i}, y_{i}\right) \begin{cases} \operatorname{Loss}_{\operatorname{student}}\left(\overline{y}_{i}, \tilde{y}_{i}\right), & \text{if } |\tilde{y}_{i} - y_{i}| \leq \varepsilon \\ 0, & \text{otherwise} \end{cases}$$
(5)

最后,学生模型的总损失公式如下:

$$\operatorname{Loss}_{\operatorname{total}} = \frac{1}{N} \sum_{i=1}^{N} \left[\operatorname{Loss}_{\operatorname{student}} \left(\overline{y}_{i}, y_{i} \right) + \beta \operatorname{Loss}_{\operatorname{distillation}} \left(\overline{y}_{i}, \widetilde{y}_{i}, y_{i} \right) \right]$$
(6)

其中, β 是蒸馏损失的权重。学生模型的训练方式与教师模型集成成员的训练方式相似, 使用 AdamW 优化器,同时添加权重衰减以防止过拟合。

2.2.4. 评估指标

本研究使用皮尔逊线性相关系数(Pearson Linear Correlation Coefficient, PLCC)、斯皮尔曼秩相关系数 (Spearman Rank-Order Correlation Coefficient, SROCC)和肯德尔秩相关系数(Kendall Rank-Order Correlation Coefficient, KROCC)来量化实际 IQA 评分 y_i 与 NR-IQA 模型预测 IQA 评分 \hat{y}_i 之间的相关性。这些指标在 IQA 领域已被广泛验证并用于衡量模型的预测性能[27] [28]。

PLCC 用于评估 $y_i = \hat{y}_i$ 之间线性相关性的强度及其方向性,其值越接近 1 表明预测值与实际值之间的线性一致性越高。PLCC 的定义如公式(7),其中 μ_{y_i} 和 $\mu_{\hat{y}_i}$ 分别表示 $y_i = \hat{y}_i$ 的均值。

$$PLCC = \frac{\sum_{i=1}^{N} (y_i - \mu_{y_i}) (\hat{y}_i - \mu_{\bar{y}_i})}{\sqrt{\sum_{i=1}^{N} (y_i - \mu_{y_i})^2 \sum_{i=1}^{N} (\hat{y}_i - \mu_{\bar{y}_i})^2}}$$
(7)

SROCC 的定义如公式(8),其中 d_i 表示 y_i 与 \hat{y}_i 之间的秩差。

SROCC =
$$1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2 - 1)}$$
 (8)

KROCC 的定义如公式(9), 用于衡量 $y_i 与 \hat{y}_i$ 之间的秩序相关性, 其中 N_c 表示一致数据对的数量, N_d 表示不一致数据对的数量。

$$KROCC = \frac{2(N_c - N_d)}{N(N - 1)}$$
(9)

$$OPM = |PLCC| + |SROCC| + |KROCC|$$
(10)

上述评估指标的取值范围均为[-1,1],在本研究中,上述评估指标的值越接近 1 表明模型的预测性能越优。此外,本研究进一步计算了一个综合性能度量(Overall Performance Metric, OPM) [29],其定义如公式(10)。OPM 通过对 PLCC、SROCC 和 KROCC 加权求和,提供了一个整体性能评估指标,其值越接近 3 表明模型的综合性能越优。

此外,本研究还使用均方误差(Mean Squared Error, MSE)和平均绝对误差(Mean Absolute Error, MAE) 来量化 $y_i = \hat{y}_i$ 之间的差异。MSE 是 $y_i = \hat{y}_i$ 之间差异的平方的平均值,其定义如公式(11); MAE 是 $y_i = \hat{y}_i$ 之间绝对差异的平均值,其定义公式(12)。MSE 和 MAE 的值越小,表明模型的预测性能越好。

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$
(11)

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$
(12)

2.2.5. 实验环境

所有实验均在一台配备 NVIDIA GeForce RTX 3060Ti 显卡和 Intel I5-12400F (Intel Corporation,美国) 的计算机上运行。该计算机基于 Windows 11 操作系统,实验代码基于 Python 3.10 编程语言开发,并采用 Pytorch 2.5.0 深度学习框架构建深度学习模型。

3. 结果

3.1. 实验结果

本研究使用了 4 个 CNN 模块和 1 个 ViT 模块实现 NR-IQA 模型,并在 CCCID 数据集上进行训练、

验证以及测试。教师模型中的每个集成成员(CNN-ViT 网络)训练了 100 个 epoch, 批处理大小为 10 张 CT 图像, 学习率设置为 0.0002。使用在验证集上具有最佳整体表现的权重来形成教师模型。学生模型训练 了 20 个 epoch, 蒸馏损失权重 β 设置为 0.75,误差阈值 ε 设置为 0.05,学生模型的训练参数设置与教师 模型相同。在 IQA 预测过程中,本研究使用了在整个训练集上表现最好的权重。教师模型的 4 个集成成 员和学生模型的损失变化如图 4 所示,本研究提出的 NR-IQA 模型在性能评估中展现出显著的优越性, 其综合性能指标达到 2.8070, PLCC 达到 0.9916,表明模型预测 IQA 评分与实际 IQA 评分之间具有极强 的线性相关性; SROCC 为 0.9683, KRCC 为 0.8471 则共同验证了模型在排序一致性上的鲁棒性。在误 差控制方面,MAE 低至 0.0259, MSE 仅为 0.0010,体现了模型预测 IQA 精度的优越性。





3.2. 消融实验

3.2.1. CNN 模块消融实验

本研究同时分析了 CNN 模块数量、教师模型集成成员数量以及学生模型蒸馏损失权重对模型性能的影响,所有实验的训练参数设置均与本研究提出的 NR-IQA 模型相同。在第一个实验中,本研究改变 了 CNN 模块数量,数量从 0 到 5 依次进行实验,所有消融模型中 ViT 架构保持不变,在完成教师模型 的训练后再进行学生模型的训练,实验结果详见表 1。实验结果表明,增加 CNN 模块的数量有助于提升 模型预测 IQA 评分的性能。此外,当 CNN 模块数量为 4 时,模型综合性能最优,其次是 5 个 CNN 模块。

Table 1. The impact of different numbers of CNN modules on the performance of NR-IQA model 表 1. 不同 CNN 模块数量对 NR-IQA 模型性能的影响

CNN 模块数量	OPM	PLCC	SROCC	KRCC	MAE	MSE
CNN-0	2.6325	0.9311	0.9292	0.7723	0.0655	0.0102
CNN-1	2.5841	0.8863	0.9295	0.7683	0.0977	0.0141
CNN-2	2.6838	0.9363	0.9446	0.8028	0.0735	0.0113
CNN-3	2.6943	0.9579	0.9456	0.7908	0.0486	0.0050
CNN-4	2.8070	0.9916	0.9683	0.8471	0.0259	0.0010
CNN-5	2.7154	0.9613	0.9525	0.8016	0.0504	0.0049

3.2.2. 集成模型消融实验

在第二个实验中,本研究评估了不同教师模型集成成员数量对模型综合性能的影响。本研究从1到5逐步增加集成成员的数量,每个集成成员的网络架构保持不变,均为 CNN-ViT 网络,在完成教师模型的训练后再进行学生模型的训练,实验结果详见表2。实验表明,增加集成成员数量能提高所有评估指标的性能,这归因于网络泛化能力的增强。此外,表现最好的教师模型是由4个集成成员组成的。

集成成员数量 OPM PLCC SROCC KRCC MAE MSE 1 0.9519 0.9376 0.7789 0.0499 0.0054 2.6684 2 2.7536 0.9755 0.9587 0.8195 0.0322 0.0025 3 2.7922 0.9761 0.9679 0.8481 0.0320 0.0025 4 2.8070 0.9916 0.9683 0.0259 0.0010 0.8471

Table 2. The impact of the number of integrated members on the performance of NR-IQA model 表 2. 不同集成成员数量对 NR-IQA 模型性能的影响

0.9505

3.2.3. 蒸馏损失权重实验

2.7439

5

在第三个实验中,本研究将学生模型的蒸馏损失权重 β 从 0 到 1 逐步进行调整,每个学生网络训练 20 个 epoch,实验结果详见表 3。实验结果表明,蒸馏损失权重在 β =0.75 时,模型所有评估指标表现最 优。蒸馏损失在模型中起到了正则化的作用,使得模型能够学习 IQA 的整体分布,蒸馏损失权重使用 0.2 到 0.8 之间的数值是有效的,而给蒸馏损失和预测损失(β =1)相同的权重则会导致模型性能下降。

0.9609

0.0074

0.0701

0.8325

蒸馏损失权重	OPM	PLCC	SROCC	KRCC	MAE	MSE
$\beta = 0$	2.3926	0.8994	0.8511	0.6422	0.0798	0.0116
$\beta = 0.2$	2.6823	0.9609	0.9386	0.7828	0.0849	0.0104
$\beta = 0.25$	2.6546	0.9429	0.9343	0.7774	0.0650	0.0090
$\beta = 0.4$	2.6406	0.8872	0.9498	0.8037	0.0761	0.0150
$\beta = 0.5$	2.6373	0.9328	0.9373	0.7673	0.0577	0.0070
$\beta = 0.6$	2.7482	0.9603	0.9584	0.8295	0.0333	0.0041
$\beta = 0.75$	2.8070	0.9916	0.9683	0.8471	0.0259	0.0010
eta=0.8	2.7237	0.9788	0.9477	0.7972	0.0423	0.0034
eta=1	2.6699	0.9588	0.9379	0.7732	0.0471	0.0047

Table 3. The impact of different student model distillation loss weights on the performance of NR-IQA model 表 3. 不同学生模型蒸馏损失权重对 NR-IQA 模型性能的影响

3.3. 对比实验

本研究对比分析了提出的 NR-IQA 模型与六个已经应用于 IQA 任务的网络: ViT、CNN-ViT、ResNet [30]、DenseNet [31]、VGGNet [32]、GoogleNet [33]。为了确保训练的一致性,保持训练超参数不变,epoch 均设置为 100,实验结果请参阅表 4。实验结果表明,本研究提出的 NR-IQA 模型相较于对比模型,在所 有评估指标上均存在显著的优越性,在综合性能上,本研究提出的 NR-IQA 模型以 2.8070 领先,较次优 模型 VGGNeT 提升约 0.8%,表明其在综合性能上具有明显优势;同时本研究提出的 NR-IQA 模型在 PLCC 指标上表现突出,较第二名的 DenseNet 提升 1.17%。实验结果验证了本研究提出的 NR-IQA 模型在 IQA 任务中的先进性和实用性,为 NR-IQA 技术在临床场景的应用提供了新的技术基准。

表 4. 不同对比模型	型的实验结果 1	sient comparative	models	
模型	OPM	PLCC	SROCC	KRC

Table 4 Experimental regults of different comparative models

模型	OPM	PLCC	SROCC	KRCC	MAE	MSE
ViT	2.7224	0.9562	0.9533	0.8128	0.0552	0.0073
CNN-ViT	2.7836	0.9746	0.9669	0.8420	0.0378	0.0031
ResNet	2.7126	0.9577	0.9508	0.8041	0.0583	0.0063
DenseNet	2.7731	0.9799	0.9621	0.8311	0.0591	0.0049
VGGNeT	2.7839	0.9762	0.9654	0.8423	0.0333	0.0024
GoogleNet	2.7423	0.9633	0.9597	0.8194	0.0665	0.0090
本研究	2.8070	0.9916	0.9683	0.8471	0.0259	0.0010

4. 讨论

NR-IQA 旨在建立与放射科医生主观评估高度一致的客观图像质量评估体系。本研究提出一种基于 深度学习的 NR-IQA 模型,该模型将 CNN 模块与 ViT 模块结合,同时训练 4 个 CNN-ViT 网络集成作为 教师模型,以模拟放射科医生多次主观 IQA 过程,接着结合知识蒸馏框架,将教师模型的信息蒸馏到 1 个学生模型(单一 CNN-ViT 网络)中。本研究利用 CCCID 数据集对提出的 NR-IQA 模型进行验证,针对 CT 图像数据集没有实际 IQA 评分的问题,结合 PSNR 和 SSIM 两个客观指标来评估 CT 图像质量,并使

用其标注 CCCID 数据集来训练提出的模型,最终实现了更高的 IQA 准确性,为 CT 图像的质量评估任务 提供了新的思路和方法。

结合表 1 和表 4 可知,相较于单独使用 ViT 模块,结合 CNN 模块和 ViT 模块中的自注意力机制更能够提高模型性能。且表 1 中的实验结果表明当 CNN 模块数量增至 4 个时,OPM 达到 2.8070, PLCC 突破 0.99 且 MAE 降至 0.0259,揭示多层级特征提取可有效捕获 CT 图像的多尺度质量特征。然而,当 CNN 模块数量增至 5 个时,OPM 下降 3.26%,MAE 反弹 94.6%,表明过深的网络结构会引发梯度退化。

请参阅表2,集成成员数量对模型性能的影响实验表明,当集成成员增至4个时,OPM达峰值2.8070, PLCC 突破 0.99 且 MAE 降至 0.0259,证明适度集成通过特征互补显著增强评估鲁棒性。然而,当集成成员增至5个时,OPM 下降2.25%、MAE 上升,表明过度集成引发的决策冲突,对模型性能产生一定的影响。

本研究还提出了使用最大误差阈值 来选择指导学生模型训练的教师预测。如果教师模型的平均预测 超过了 ε ,它将不被视为适合的标签用于学生模型训练;当教师模型的平均预测误差低于 ε 时,将其作为 新的实际 IQA 标签向量用于学生模型训练,有效地减轻了教师模型的极端预测对学生模型可能产生的不 利影响,从而导致更稳健、更准确的学生模型训练过程。表 3 的实验结果表明,蒸馏损失权重的选择在 训练中的重要性,当 β =0.75 时,模型达最优性能(OPM = 2.8070, PLCC = 0.9916),MAE 较无蒸馏(β =0) 降低 67.5%,验证适度软化教师监督可协调特征空间对齐与预测精度。

此外,未来的研究工作应该着重于测试提出的 NR-IQA 模型的泛化能力,通过在其他医学图像数据 集上进行评估。

5. 结论

本研究提出了一种基于 ViT 模块和知识蒸馏策略的 NR-IQA 模型,该模型将 CNN 模块与 ViT 模块 结合,同时训练 4 个 CNN-ViT 网络集成作为教师模型,以模拟放射科医生多次主观 IQA 过程,接着结 合知识蒸馏框架,将教师模型的信息蒸馏到 1 个学生模型中。针对 CT 图像数据集没有实际 IQA 评分的 问题,本研究结合 PSNR 和 SSIM 两个客观指标来评估 CT 图像质量,并使用其标注 CCCID 数据集以训 练提出的 NR-IQA 模型,最终实现了更高的 IQA 准确性,为 CT 图像的质量评估任务提供了新的思路和 方法。

基金项目

上海介入医疗器械工程技术研究中心(18DZ2250900)。

参考文献

- [1] Xun, S.Y., Li, Q.Y., Liu, X.H., *et al.* (2025) Charting the Path Forward: CT Image Quality Assessment—An In-Depth Review. arXiv: 2405.00075.
- [2] Yi, X. and Babyn, P. (2018) Sharpness-Aware Low-Dose CT Denoising Using Conditional Generative Adversarial Network. *Journal of Digital Imaging*, **31**, 655-669. <u>https://doi.org/10.1007/s10278-018-0056-0</u>
- [3] Kasban, H., El-Bendary, M. and Salama, D. (2015) A Comparative Study of Medical Imaging Techniques. *International Journal of Information Science and Intelligent System*, **4**, 37-58.
- [4] Gao, Q., Li, S., Zhu, M., Li, D., Bian, Z., Lv, Q., et al. (2020) Combined Global and Local Information for Blind CT Image Quality Assessment via Deep Learning. *Medical Imaging* 2020: *Image Perception, Observer Performance, and Technology Assessment*, Houston, 15-20 February 2020. <u>https://doi.org/10.1117/12.2548953</u>
- [5] Lee, W., Cho, E., Kim, W., Choi, H., Beck, K.S., Yoon, H.J., *et al.* (2022) No-Reference Perceptual CT Image Quality Assessment Based on a Self-Supervised Learning Framework. *Machine Learning: Science and Technology*, 3, Article ID: 045033. <u>https://doi.org/10.1088/2632-2153/aca87d</u>

- [6] Zarb, F., Rainford, L. and McEntee, M.F. (2010) Image Quality Assessment Tools for Optimization of CT Images. *Radiography*, 16, 147-153. <u>https://doi.org/10.1016/j.radi.2009.10.002</u>
- Bevabcmlal, A. (2016) Knowledge-Based Taxonomic Scheme for Full-Reference Objective Image Quality Measurement Models. *Journal of Imaging Science and Technology*, **60**, 60406-1-60406-15.
- [8] Sara, U., Akter, M. and Uddin, M.S. (2019) Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *Journal of Computer and Communications*, 7, 8-18. <u>https://doi.org/10.4236/jcc.2019.73002</u>
- [9] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13, 600-612. <u>https://doi.org/10.1109/tip.2003.819861</u>
- [10] Rehman, A. and Wang, Z. (2012) Reduced-Reference Image Quality Assessment by Structural Similarity Estimation. IEEE Transactions on Image Processing, 21, 3378-3389. <u>https://doi.org/10.1109/tip.2012.2197011</u>
- [11] Bampis, C.G., Gupta, P., Soundararajan, R. and Bovik, A.C. (2017) Speed-QA: Spatial Efficient Entropic Differencing for Image and Video Quality. *IEEE Signal Processing Letters*, 24, 1333-1337. <u>https://doi.org/10.1109/lsp.2017.2726542</u>
- [12] Zhang, Y., Phan, T.D. and Chandler, D.M. (2017) Reduced-Reference Image Quality Assessment Based on Distortion Families of Local Perceived Sharpness. *Signal Processing: Image Communication*, 55, 130-145. https://doi.org/10.1016/j.image.2017.03.020
- [13] Lu, Y., Fu, J., Li, X., Zhou, W., Liu, S., Zhang, X., et al. (2022) RTN: Reinforced Transformer Network for Coronary CT Angiography Vessel-Level Image Quality Assessment. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S. and Li, S., Eds., Medical Image Computing and Computer Assisted Intervention—MICCAI 2022, Springer, 644-653. https://doi.org/10.1007/978-3-031-16431-6_61
- [14] Baldeon Calisto, M.G., Rivera-Velastegui, F., Lai-Yuen, S.K., Riofrío, D., Pérez, N., Benítez, D., et al. (2024) Distilling Vision Transformers for No-Reference Perceptual CT Image Quality Assessment. Medical Imaging 2024: Image Processing, San Diego, 19-22 February 2024. <u>https://doi.org/10.1117/12.3004838</u>
- [15] Xun, S., Jiang, M., Huang, P., Sun, Y., Li, D., Luo, Y., et al. (2024) Chest CT-IQA: A Multi-Task Model for Chest CT Image Quality Assessment and Classification. *Displays*, 84, Article ID: 102785. <u>https://doi.org/10.1016/j.displa.2024.102785</u>
- [16] Gao, Q., Li, S., Zhu, M., Li, D., Bian, Z., Lyu, Q., et al. (2019) Blind CT Image Quality Assessment via Deep Learning Framework. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Manchester, 26 October-2 November 2019, 1-4. <u>https://doi.org/10.1109/nss/mic42101.2019.9059777</u>
- [17] Ayaan, H., Adam, W. and Abdullan-al-zubaer, I. (2022) Noise2Quality: Non-Reference, Pixel-Wise Assessment of Low Dose CT Image Quality. *Image Perception, Observer Performance, and Technology Assessment: Medical Imaging* 2022, San Francisco, 20-24 February 2022, 120351C-1-120351C-6.
- [18] Mudeng, V., Kim, M. and Choe, S. (2022) Prospects of Structural Similarity Index for Medical Image Analysis. Applied Sciences, 12, Article 3754. <u>https://doi.org/10.3390/app12083754</u>
- [19] Hore, A. and Ziou, D. (2010) Image Quality Metrics: PSNR vs. SSIM. 2010 20th International Conference on Pattern Recognition, Istanbul, 23-26 August 2010, 2366-2369. <u>https://doi.org/10.1109/icpr.2010.579</u>
- [20] Cavaro-Menard, C., Zhang, L. and Le Callet, P. (2010) Diagnostic Quality Assessment of Medical Images: Challenges and Trends. 2010 2nd European Workshop on Visual Information Processing (EUVIP), Paris, 5-6 July 2010, 277-284. <u>https://doi.org/10.1109/euvip.2010.5699147</u>
- [21] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2021) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [22] Hinton, G., Vinyals, O. and Dean, J. (2015) Distilling the Knowledge in a Neural Network. arXiv: 1503.02531.
- [23] Zhao, Q., Zhong, L., Xiao, J., Zhang, J., Chen, Y., Liao, W., *et al.* (2023) Efficient Multi-Organ Segmentation from 3D Abdominal CT Images with Lightweight Network and Knowledge Distillation. *IEEE Transactions on Medical Imaging*, 42, 2513-2523. <u>https://doi.org/10.1109/tmi.2023.3262680</u>
- [24] 刘泽奇, 王宁, 张冲, 魏国辉. 基于轻量化网络与知识蒸馏策略的心脏核磁共振图像分割[J]. 生物医学工程学杂志, 2024, 41(6): 1204-1212.
- [25] Chen, G., Choi, W., Yu, X., et al. (2018) Learning Efficient Object Detection Models with Knowledge Distillation. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, 4-9 December 2017, 742-751. https://dl.acm.org/doi/10.5555/3294771.3294842
- [26] Saputra, M.R.U., Gusmao, P., Almalioglu, Y., Markham, A. and Trigoni, N. (2019) Distilling Knowledge from a Deep Pose Regressor Network. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 263-272. <u>https://doi.org/10.1109/iccv.2019.00035</u>
- [27] Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., et al. (2020) Blindly Assess Image Quality in the Wild Guided by a

Self-Adaptive Hyper Network. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 3664-3673. <u>https://doi.org/10.1109/cvpr42600.2020.00372</u>

- [28] Wu, J., Ma, J., Liang, F., Dong, W., Shi, G. and Lin, W. (2020) End-to-End Blind Image Quality Prediction with Cascaded Deep Neural Network. *IEEE Transactions on Image Processing*, 29, 7414-7426. https://doi.org/10.1109/tip.2020.3002478
- [29] Lee, W., Wagner, F., Galdran, A., Shi, Y., Xia, W., Wang, G., et al. (2025) Low-Dose Computed Tomography Perceptual Image Quality Assessment. *Medical Image Analysis*, 99, Article ID: 103343. https://doi.org/10.1016/j.media.2024.103343
- [30] Xu, L. and Chen, Q. (2019) Remote-sensing Image Usability Assessment Based on Resnet by Combining Edge and Texture Maps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12, 1825-1834. https://doi.org/10.1109/jstars.2019.2914715
- [31] Jiang, T., Hu, X., Yao, X., Tu, L., Huang, J., Ma, X., et al. (2021) Tongue Image Quality Assessment Based on a Deep Convolutional Neural Network. BMC Medical Informatics and Decision Making, 21, Article No. 147. https://doi.org/10.1186/s12911-021-01508-8
- [32] Gao, F., Yu, J., Zhu, S., Huang, Q. and Tian, Q. (2018) Blind Image Quality Prediction by Exploiting Multi-Level Deep Representations. *Pattern Recognition*, 81, 432-442. <u>https://doi.org/10.1016/j.patcog.2018.04.016</u>
- [33] Sun, J., Wan, C., Cheng, J., Yu, F. and Liu, J. (2017) Retinal Image Quality Classification Using Fine-Tuned CNN. In: Cardoso, M., et al., Eds., Fetal, Infant and Ophthalmic Medical Image Analysis, Springer, 126-133. <u>https://doi.org/10.1007/978-3-319-67561-9_14</u>