

医学图像伪造检测技术研究综述

汤婧婧¹, 徐立中²

¹南京信息职业技术学院数码艺术学院, 江苏 南京

²河海大学计算机与信息学院, 江苏 南京

收稿日期: 2025年8月7日; 录用日期: 2025年10月30日; 发布日期: 2025年11月7日

摘要

医学图像伪造检测技术是保障医疗数据安全、确保临床诊断准确性的重要研究方向。随着数字图像处理技术和生成式人工智能的快速发展, 医学图像伪造手段日益复杂, 其检测技术也在不断完善。近年来, 医学图像伪造检测技术的突破性研究日新月异, 为了全面了解该领域的研究进展, 本文对医学图像伪造检测的核心问题进行系统综述: 首先, 分类总结伪造检测方法及其演进历程; 其次, 详细介绍了常用数据集的特点与适用范围以及评估指标; 重点梳理了医学CT、MRI图像小区域伪造和抗压缩干扰的优化策略; 最后, 提出医学图像伪造检测技术研究的未来方向, 是深度学习的多模态融合技术和主动防御策略的医学信息伪造检测研究, 以应对生成式AI时代的医学图像伪造挑战。

关键词

医学图像伪造检测, 深度学习, 生成对抗网络, 性能评估

A Review of Research on Medical Image Forgery Detection Technology

Jingjing Tang¹, Lizhong Xu²

¹Digital Art College, Nanjing Vocational College of Information Technology, Nanjing Jiangsu

²College of Computer and Information Science, Hohai University, Nanjing Jiangsu

Received: August 7, 2025; accepted: October 30, 2025; published: November 7, 2025

Abstract

Medical image forgery detection technology is an important research direction to ensure the safety of medical data and the accuracy of clinical diagnosis. With the rapid development of digital image processing technology and generative Artificial intelligence, the means of medical image forgery are increasingly complex, and its detection technology is also constantly improving. In recent years, the

文章引用: 汤婧婧, 徐立中. 医学图像伪造检测技术研究综述[J]. 生物医学, 2025, 15(6): 1102-1113.

DOI: 10.12677/hjbm.2025.156119

breakthrough research of medical image forgery detection technology is changing rapidly. In order to fully understand the research progress in this field, this paper systematically reviews the core issues of medical image forgery detection: first, classify and summarize forgery detection methods and their evolution process; Secondly, the characteristics, application scope and evaluation indicators of common data sets are introduced in detail; Focused on the optimization strategy of medical CT and MRI image small area forgery and anti-compression interference; Finally, the future direction of research on medical image forgery detection technology is proposed, which is the multimodal fusion technology of deep learning and the research on medical information forgery detection of active defense strategy, to meet the challenge of medical image forgery in the era of generative AI. This electronic document defines the standard format of the Chinese academic journals published by the Hans Publishing. The elements such as the paper title, author, affiliation, abstract, section title, main text, figure, table and references are defined, and this document is formatted according to the Hans standard, which illustrates all the formats.

Keywords

Medical Image Forgery Detection, Deep Learning, Generative Adversarial Networks, Performance Evaluation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

医学图像采集与识别是医学研究和教育的重要领域, 主要内容包括普通 X 光片、计算机断层扫描(CT)、磁共振成像(MRI)、超声成像、核医学技术和通过显微镜或数字化技术对组织样本进行分析的病理切片检查等, 目前已成为病人临床诊断、治疗监测不可或缺的技术手段。随着数字设备的普及和图像处理技术的进步, 一般人都能通过拼接、复制 - 移动等技术轻而易举的修改图像, 图 1 是图像复制移动伪造的例子[1]。如果在医学图像中伪造病灶区域或掩盖异常组织[2] [3]。这种篡改可能导致医生误诊, 进而引发错误的治疗方案, 严重威胁患者生命[4]。在司法鉴定、医疗保险等领域, 伪造的医学图像可能被用于欺诈索赔或掩盖医疗事故[3] [5]。此外, 医学图像的隐私泄露, 患者信息绑定不完整, 也加剧了数据滥用的风险[6] [7], 这对医疗信息的核心安全构成挑战。因此, 医学图像的伪造性检测, 即真实性认证(Authentication)和篡改定位(Tampering Localization)对病人正确诊断与治疗具有重要意义, 不仅是技术问题, 更关乎医疗伦理与公共安全[8] [9]。



Figure 1. (a) is the original image, (b) is the copy-move image, (c) shows the copied and pasted regions which are marked green and red, respectively, (d) is the binary mask of copy-move image

图 1. (a) 是原始图像, (b) 是复制 - 移动图像, (c) 显示标记为绿色和红色的复制和粘贴区域, (d) 是复制移动图像的二值掩码

2. 医学图像伪造

由于临床影像检查技术的不断成熟和费用下降, 医学 CT、MRI、超声等检查得到了广泛应用, 这使得患者早期诊断和及时处理成为现实, 在提高医疗水平的同时, 一些医学图像伪造也随之发生发展, 按时间顺序分为: ① 2000 年以前(早期阶段), 医学图像伪造主要依赖基础的图像编辑软件, 手段相对简单, 使用 Photoshop 等工具进行亮度调整、对比度修改或简单的复制粘贴操作[10]。这一时期的伪造技术较为原始, 留下的痕迹明显, 检测难度相对较低。② 2000~2015 年(中期发展阶段), 此期见证了数字图像处理技术的飞速进步, 伪造技术也随之变得更加复杂和隐蔽。随着视觉传感器的广泛应用和数字图像处理技术的发展, 图像复制移动伪造变得越来越普遍[1]。这一时期出现了区域复制、对象移除、图像拼接等更高级的伪造技术[11]。在医学影像领域, 图像伪造开始被滥用, 主要发生在有先进医疗设备的地区[12]。这一阶段的伪造技术虽然比早期复杂, 但由于缺乏智能化的伪造手段, 仍可通过传统的取证分析方法进行检测。③ 2015 年至今(深度学习时代), 这一时期的伪造技术发生了质的飞跃。生成对抗网络等深度学习模型的引入, 使得伪造技术能够生成高度逼真的医学影像[13]。这一时期出现了几种典型的伪造技术: 一是基于 GAN 的医学影像生成, 如 CT-GAN 可以生成逼真的 CT 扫描图像[14]; 二是深度伪造(Deep fake)技术在医学影像中的应用, 可以生成高度逼真的伪造面部医学图像[15]; 三是多级伪造技术, 即在一个伪造上叠加另一个伪造, 极大增加了检测难度[16]。特别值得关注的是, 随着人工智能生成(AIGC)技术的快速发展, 区分真实与伪造的医学面部图像变得越来越困难[17]。生成模型的进步使得伪造医学影像的成本大幅降低[18], 而检测这些高度逼真的伪造则面临前所未有的挑战。医学影像伪造技术已经从简单的图像处理发展为能够生成完整医学影像的智能系统, 这对伪造检测技术提出了更高的要求[13][14]。

当前, 医学影像伪造技术仍在不断进化, 呈现出几个明显趋势: 一是伪造区域越来越小, 有些甚至只占原始图像的不到 1% [14]; 二是伪造手段越来越多样化, 从简单的复制粘贴到复杂的生成模型应有尽有[1][13]; 三是抗取证伪造技术日益成熟, 能够有针对性地规避检测系统的识别[19]。这些发展趋势使得医学图像伪造检测面临持续的技术挑战。

3. 医学图像伪造检测

3.1. 医学图像伪造检测发展历程

医学图像伪造检测技术经历了从传统方法到深度学习方法的演进过程, 检测精度和效率得到了显著提升。早期的医学图像伪造检测主要依赖手工特征提取和传统的机器学习方法。这些方法通常基于图像的底层特征, 如颜色、纹理、噪声分布等, 通过分析这些特征的异常来识别潜在的伪造区域[1]。然而, 随着伪造技术的进步, 这些传统方法在面对复杂的医学图像伪造时逐渐显示出局限性, 对复制-移动伪造这类异常较小的伪造类型检测效果不佳[1]。医学图像伪造检测技术的发展按时间顺序分为以下几个阶段: ① 早期阶段(2000 年代初期): 早期的医学图像伪造检测主要依赖于传统的数字图像分析方法, 如基于像素的比对和简单的特征提取技术。由于医学图像的特殊性, 伪造手段相对简单, 主要是通过复制-粘贴伪造等手段进行局部修改[12][20]。这一阶段的检测技术主要关注于识别图像中的重复区域或异常像素分布, 但受限于计算能力和算法复杂度, 检测精度和效率较低。② 中期阶段(2010 年代): 随着数字图像处理技术的进步, 医学图像伪造检测开始引入更复杂的算法, 如基于块匹配和尺度不变特征变换(SIFT)的方法[1][10]。这些技术能够更好地应对图像缩放、旋转等变换带来的伪造痕迹。同时, 机器学习方法, 如支持向量机(SVM)开始被应用于特征分类, 提高了检测的准确性[21]。然而, 这一阶段的检测技术仍面临泛化能力不足的问题, 难以应对新型伪造手段。③ 深度学习兴起阶段(2010 年代末至 2020 年代初): 卷积神经网络(CNN)和深度学习的引入标志着医学图像伪造检测技术的重大突破。研究者提出了基于

CNN 的检测框架, 如 Faster R-CNN 和 U-Net, 能够更高效地定位伪造区域[11] [13]。此外, 生成对抗网络(GAN)的出现使得伪造技术更加复杂, 同时也催生了针对 GAN 生成图像的检测方法, 如针对 CT-GAN 的小区域伪造检测框架[13] [22]。这一阶段的检测技术显著提升了精度, 但仍受限于训练数据的多样性和泛化能力。④ 近期发展(2020 年代至今): 当前的研究重点转向多模态融合和元学习(Meta-Learning)技术, 以应对不断演变的伪造手段, 如结合超边界像素方向(Super-BPD)分割和深度 CNN 的 SD-Net 框架, 能够更准确地定位复制 - 移动伪造区域[10] [23]; 针对 JPEG 压缩图像的伪造检测技术, 如 MSCSCC-Net 通过多尺度机制提升了鲁棒性[24]。元学习方法也被用于快速适应新型伪造技术, 仅需少量样本即可调整模型参数[25]; 同时, 研究者开始关注检测技术的可解释性和公平性, 如基于块通道对应(Patch-Channel Correspondence)的检测模型[18] [26]。

3.2. 医学图像伪造检测研究现状

医学图像伪造检测是当前数字医疗安全领域的重要研究方向, 由于医疗图像的特殊性, 伪造直接影响临床决策, 尤其在缺乏先进医疗设备的地区, 篡改可能导致滥用, 在复制 - 移动伪造中, 篡改区域可能仅占原图的 1% 以下, 传统方法难以检测[14]。GAN 生成的伪造区域(如 CT-GAN)因局部细节逼真, 对检测算法提出更高要求[27]。针对复制 - 移动伪造、GAN 生成的小区域篡改等常见伪造手段展开研究, 通过匹配相似图像块识别篡改区域, 但传统方法对复杂变换(如旋转、缩放)效果有限[28]。用于特征提取和篡改定位的 CNN 检测复制 - 移动和拼接伪造, 要依赖人工特征提取[29], 单一 CNN 模型准确率较低, SD-Net 是结合超边界 - 像素方向(super-BPD)分割技术, 是一种改进方法, 提升相似区块的关联性检测[10]; DFST-UNet (Deep Feature Supervision and Training U-Net)是一种改进的 U-Net 架构, 它在传统 U-Net 的基础上引入了深度特征监督(Deep Feature Supervision, DFS)和多阶段训练策略, 融合频域和空域特征, 增强局部伪造痕迹的定位能力[13], 提高了精度和泛化能力; 多尺度上下文网络 MSCSCC-Net: 针对 JPEG 压缩图像, 利用多尺度机制保留伪造痕迹, 显著提升检测鲁棒性[24]; 自监督与轻量化方法, SFTA-Net: 基于自监督和三重损失, 无需标注数据即可检测篡改区域, 适用于资源受限场景[30]。快速区域 CNN(FR-CNN): 实时定位篡改区域, 平衡检测速度与精度[11]。

当前, 国内外研究聚焦于深度学习模型的优化与泛化能力提升, 但医疗图像的特殊性, 如小区域篡改、高精度要求, 仍带来显著挑战。未来方向包括: 开发轻量化实时检测系统、增强跨域泛化性, 以及结合可解释性技术满足临床需求。

4. 医学图像伪造检测方法分类、数据集和评价指标

医学图像伪造检测是随着伪造技术发展而发展的, 近 20 年来, 新理论, 新技术不断涌现, 创造出许多医学图像伪造检测方法, 以及对这些方法训练、测试和验证的数据集, 还有对这些方法性能比较的评价指标, 简介如下。

4.1. 医学图像伪造检测方法分类

医学图像伪造检测方法从方法学角度分为三种类型: ① 传统检测方法: 传统医学图像伪造检测方法主要包括以下几类: 基于块匹配的方法、基于关键点的方法和基于频域分析的方法。基于块匹配的方法将图像分割为重叠或非重叠的块, 然后比较这些块之间的相似性来检测复制区域[1]。基于关键点的方法则提取图像中的显著关键点(如 SIFT、SURF 等), 通过匹配这些关键点来定位伪造区域[31]。基于频域分析的方法则利用离散余弦变换(DCT)、离散小波变换(DWT)等工具分析图像的频域特征异常[24]。这些方法在特定条件下表现尚可, 但存在计算复杂度高、对几何变换敏感、难以检测复杂伪造等缺点[10]。② 深

深度学习的方法：近年来，深度学习技术在医学图像伪造检测领域取得了显著进展，CNN 成为检测医学图像伪造的主流架构之一[10]。与传统方法相比，CNN 能够自动学习更高级的图像特征，对复杂的伪造模式具有更强的识别能力。有研究提出将误差水平分析(ELA)与 CNN 相结合来检测图像伪造，该系统主要专注于检测图像中的复制移动和拼接伪造，ELA 算法首先识别输入图像中的潜在伪造区域，然后由 CNN 进行进一步分析和验证[29]。为提升复制移动伪造检测的准确性，有研究者提出了基于超边界到像素方向(super-BPD)分割和深度 CNN(DCNN)的图像复制移动伪造检测与定位方法 SD-Net，该方法首先利用分割技术增强相同或相似图像块之间的连接，提高检测精度；其次使用 DCNN 提取图像特征，取代传统手工特征提取方法[10]。这种方法在医学影像伪造检测中表现出优于传统方法的性能。更先进的网络架构也不断被引入医学影像伪造检测领域。基于区域的快速卷积神经网络(Faster Region-Based Convolutional Neural Networks FR-CNN)通过快速识别和定位被伪造区域，改善了图像伪造检测性能，增强了特征提取的准确性和实时取证分析能力[11]。针对 JPEG 压缩产生的伪影可能掩盖伪造痕迹的问题，有研究提出了多尺度上下文空间-通道相关网络(MSCSCC-Net)，该网络利用多尺度机制，通过更好地处理伪造区域的尺度变化来提高伪造检测和定位性能，并有助于在不同尺度上去除 JPEG 伪影[28]。两阶段级联框架被提出用于检测基于 GAN 的医学影像小区域伪造，如 CT-GAN [14]。在局部检测阶段，使用小子图像训练检测器网络，以减少真实区域中的干扰信息；在全局验证阶段，则利用全局信息进一步验证局部检测结果，这种方法针对医学影像中只占原始图像不到 1% 的小伪造区域特别有效[14]。③ 多模态融合方法：随着伪造技术的复杂化，单一检测方法往往难以应对所有类型的伪造，因此多模态融合方法逐渐成为研究热点。有研究尝试结合多种互补的检测方法，通过分析这些方法的互补性，客观测量互补程度，并计算目标性能值来增强伪造检测和定位能力，虽然这种融合方法在泛化能力方面仍存在局限，但这在监督学习中是一个普遍问题，也为未来工作提供了方向[32]。针对医学影像的特殊性，一些研究开始关注如何提取固有像素不一致性的伪造指纹。有人整合了 12 个数据集上的 16 个代表性检测模型，建立一个综合基准，专注于捕捉固有像素级伪影而非挖掘语义伪造痕迹的方法，实验表明，能够取得先进的泛化性和鲁棒性表现[33]。这种方法特别适合医学图像伪造检测，因为医学图像通常有严格的像素级一致性要求。与自然图像不同，医学图像通常具有特定的模态特征和结构约束，这要求伪造检测方法必须适应医学图像的特殊性，有研究将图像预处理、分割、特征提取、深度学习和迁移学习方法专门应用于医学成像数据[34]，这些针对医学图像的专门化方法正成为该领域的重要研究方向。

4.2. 医学图像伪造检测数据集

医学图像伪造检测数据集是医疗 AI 安全领域的重要研究内容，现在 CT/MR 图像伪造检测数据集较多，超声、病理、内镜等相对稀缺，存在模态不均衡，多数数据集都是针对特定篡改类型，跨模态检测效果差，泛化性不足。目前医学图像伪造检测数据集数据规模有限，远小于 Face Forensics++ 等自然图像伪造数据集，这是因为医学图像标注篡改区域需要专业医学知识，耗时费力，成本高。现在常用的医学图像伪造检测数据集的主要来源，是对现有医学影像数据集进行改造/标注的数据集，还有将通用图像篡改检测数据集中的部分图像替换为医学图像，如 CASIA、GRIP、CoMoFoD、CMH、MICC-F、Coverage 等，或者在医学图像上应用通用篡改技术来创建测试集，如拼接、复制-移动、移除，但这类数据集的医学特异性、篡改痕迹的自然性以及真实医学场景的相关性往往不足。图 2 为部分医学图像伪造检测数据集[35]。常用的医学图像伪造检测数据集简介如下：

1) CASIA-CMFD 数据集(CASIA Copy-Move Forgery Dataset) [36]，是由中国科学院自动化研究所(CASIA)发布的大规模复制-移动篡改检测数据集。该数据集规模大、篡改类型多样、标注精细等特点，成为 CMFD 领域的重要基准之一。原始图像：5063 张，有自然场景、不同光照和背景；篡改图像：5063

张, 每张原始图像对应 1 个篡改版本, 总计 10,126 张图像, 是当前复制 - 移动篡改检测领域最全面、最适用于深度学习的数据集。

2) CoMoFoD 数据集(Copy-Move Forgery Detection Dataset) [37]: 是一个专门用于图像复制 - 移动篡改检测研究的公开数据集。原始图像 40 张, 篡改图像 260 张, 共 300 张图像。是 CMFD 领域的经典数据集。

3) CMH (Copy-Move Hard Dataset) [38]: 它由四个 CMH 子数据集组成, 是一个面向高难度复制 - 移动篡改检测的专用数据集。原始图像: 30 张, 高分辨率, 涵盖自然和人工场景。篡改图像: 每张原始图像生成多个篡改版本, 总共有 108 个复制移动篡改图像,

4) MICC-F220 [39]: 是由意大利 Media Integration and Communication Center (MICC)开发的用于复制 - 移动篡改检测的数据集, 广泛应用于数字图像取证领域, 评估算法在复杂篡改场景下的性能。是复制 - 移动篡改检测领域的经典基准数据集。有自然场景、不同分辨率的原始图像 110 张, 篡改图像 110 张, 每张原始图像对应 1 个篡改版本, 总计 220 张图像, 因此命名为“F220”。

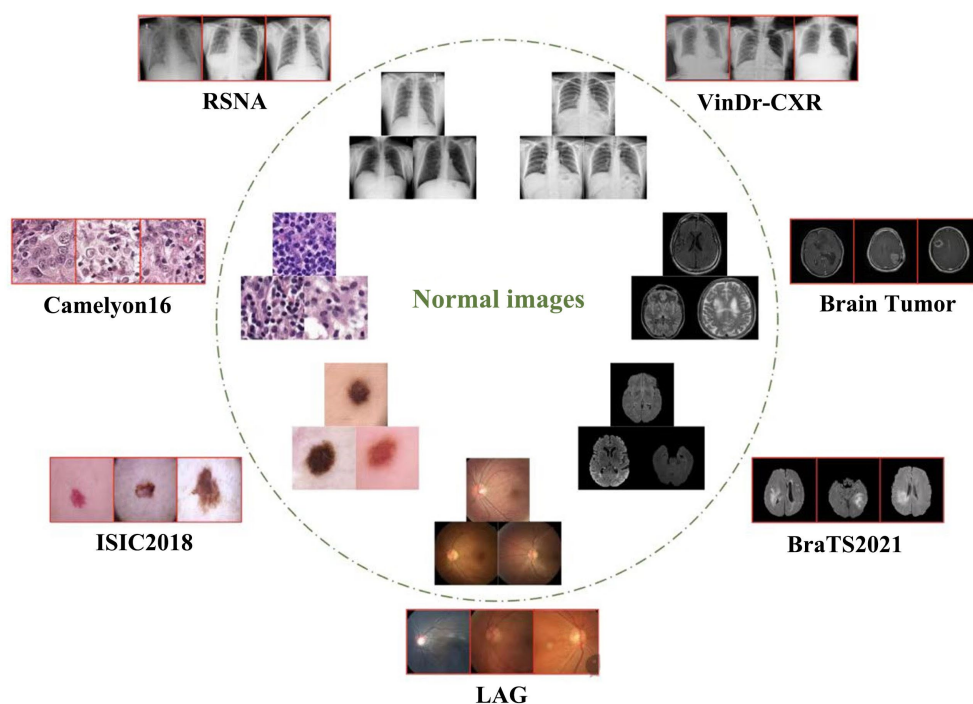


Figure 2. Some medical image forgery detection datasets. Samples enclosed by the green dashed circle are normal, while others are abnormal [35]

图 2. 部分医学图像伪造检测数据集, 绿色虚线圆圈包围的样本是正常的, 而其他样本是异常的[35]

5) MICC-F600 [40]: 是 MICC 在 MICC-F220 基础上扩展的更大规模复制 - 移动篡改检测数据集, 旨在提供更丰富的篡改类型和更具挑战性的测试案例。是复制 - 移动篡改检测领域的重要基准数据集。原始图像 440 张; 篡改图像 160 张, 总计 600 张图像。

6) GRIP [41]: 是一个专注于复杂图像篡改检测与鲁棒性分析的公开数据集。该数据集由国际图像取证研究团队构建, 原始图像 3500+张; 篡改图像 7000+张, 每张原始图像生成 2~3 种篡改版本, 总计 10,500+。

7) Coverage 数据集[42]: 包含 100 幅基本图像和相应的篡改图像, 是一个专门用于 CMFD 的基准数据集。原始图像 100 张; 篡改图像 100 张; 总计 200 张图像,

- 8) CFDMI-SEC 数据集[12]: 主要来自公开的医学影像数据库 DICOM 样本库, 包含 CT 扫描、MRI 和 X 射线等多种医学影像模态, 有正常医学图像 100 张, 伪造图像 200 张, 伪造图像使用不同参数和区域进行篡改以模拟真实伪造场景, 有不同身体部位的医学影像, 涵盖简单和复杂的伪造案例。
- 9) RSNA 数据集[43]: 是由北美放射学会(Radiological Society of North America, RSNA)提供的公开医学影像数据集, 包含各类医学影像(如 X 光、CT、)及相应的标注信息, 涵盖多种疾病(如肺炎、骨折等)。有正常图像 8851 张, 肺部异常阴影图像 6012 张, 有助与医学影像领域的 AI 研究、算法开发和临床诊断技术进步。
- 10) VinDr-CXR 数据集[44]: 是一个公开的胸部 X 光医学影像数据集, 包含大量标注的胸部 X 光片, 正常 10,606 张, 异常 4394 张, 有 14 类胸部疾病, 专注于胸部疾病的检测和分类。主要是针对常见胸部异常的自动检测, 推动人工智能在放射学诊断中的应用。
- 11) Camelyon16 数据集[45]: 是首个专注于全幻灯片图像(WSI)分析的公开数据集, 包含患者 160 人, WSI400 张。由荷兰拉德堡德大学医学中心发布, 其目的是推动乳腺癌淋巴结转移的自动检测算法研究, 辅助病理诊断, 促进计算病理学的发展。
- 12) BraTS2021 数据集[46]: 是全球最大规模的脑肿瘤多模态 MRI 数据集之一, 涵盖胶质母细胞瘤(GBM)和低级别胶质瘤(LGG), 包含 1251 例多模态 MRI 扫描: T1、T1c、T2、FLAIR。有像素级的分割标签, 图像分辨率 $155 \times 240 \times 240$ 。常用医学图像伪造检测数据集的主要特征总结见表 1。

Table 1. Common medical image forgery detection dataset and characteristics
表 1. 常用医学图像伪造检测数据集及特点

数据集	图像数量	主要特点	适用场景	文献
CASIA-CMFD	10,126	大规模 + 高分辨率 + 深度学习友好	训练深度网络 + 精细评估	[36]
CoMoFoD	300	纹理篡改 + JPEG 压缩	传统方法对比	[37]
CMH	150+	低纹理 + 复杂变换	平滑区域篡改 + 鲁棒性测试	[38]
MICC-F220	220	基础几何变换 + 遮挡	算法初步验证	[39]
MICC-F600	600	多篡改区域 + 复杂变换 + 后处理	鲁棒性 + 泛化性测试	[40]
GRIP	10,500+	多类型 + 多干扰 + 多模态	鲁棒性 + 跨模态测试	[41]
Coverage	200	低纹理 + 复杂几何变换 + 投票优化	边界定位精度评测	[42]
CFDMI-SEC	300	高分辨率 + 不同身体部位 + 简单和复杂	检测算法的性能	[12]
RSNA	14,863	图像级异常分类	检测算法开发	[43]
VinDr-CXR	15,000	图像级异常分类	胸部疾病的检测和分类	[44]
Camelyon16	WSI400	图像级异常分类	自动检测算法研究, 辅助病理诊断	[45]
BraTS2021	1251	图像级异常分类像素级异常分割	脑肿瘤多模态检测	[46]

4.3. 常用的医学图像伪造检测评估指标

1) 图像级评价指标: Precision(精准度)表示正确检测到的伪造图像/所有检测的图像或像素的像素比例; Recall (召回率)表示正确检测到的伪造图像或像素占有所有伪造图像或像素的比例; F1 分数是精准度和召回率的调和平均数; AUC 是 ROC 曲线下的面积, 用于评估不同阈值下的分类, 显示模型在不同类别之间的区分程度, ROC 越光滑拟合的概率就越小, AUC 越高, 越能覆盖整个区域, 即越接近 1, 模型在正确预测每个类别方面就越好, 更容易分离, 也更好分辨, 他们的计算公式:

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (1)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (2)$$

$$F1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (3)$$

2) 像素级评价指标: Dice 系数为预测区域与标签区域的重叠面积的两倍除以两幅图像的总像素数, 是评价分割结果和标签之间差距的重要指标; IoU 交并比(分割篡改区域)是模型对某一类别预测结果和真实值的交集与并集的比值, 他们的计算公式为:

$$\text{DiceCoefficient} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (4)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (5)$$

3) 视频级评价指标: 在医学图像伪造检测, 尤其是动态影像如超声、血管造影或功能 MRI 中, 时序一致性误差(Temporal Consistency Error TCE)用于量化伪造视频在时间维度上的运动连贯性与生理合理性缺陷, 是通过真实视频与伪造视频之间的运动模式差异计算的, 计算公式如下:

$$\text{TCE} = \frac{1}{N} \sum_{i=1}^N \|P_{t+1}(i) - \text{Warp}(P_t(i), \text{Flow}_{t \rightarrow t+1})\|_p \quad (6)$$

5. 医学伪造检测主要挑战和解决方法

医学图像伪造检测主要挑战: ① 伪造技术复杂性: 医学图像中常见的复制 - 移动伪造技术由于异常区域较小, 可能小于原图的 1%, 使得检测变得极具挑战性[14] [24]。特别是当伪造区域经过模糊处理、对比度调整或压缩等后处理操作后, 传统检测方法更难识别[10]; ② 检测准确性不足: 现有的基于 CNN 的方法, 在检测精度上仍存在明显不足[47], 对 GAN 生成的医学图像伪造时, 现有方法难以有效捕捉微小区域的伪造痕迹[24]; ③ 泛化能力局限: 现有深度学习算法在遇到训练集中未出现过的背景区域或新型伪造技术时表现不佳[30], 且多数方法仅支持单一伪造类型的检测[16]; ④ 数据稀缺与规模小: 相比自然图像, 高质量的医学图像获取本身就困难且受隐私保护限制, 构建大规模、多样化的伪造数据集成本极高; 标注困难: 精确标注篡改区域需要专业医学知识, 耗时费力; 直接使用真实患者数据构建的伪造检测数据集, 存在严重的隐私和伦理问题, 通常需要严格的匿名化和伦理审批。许多数据集要在受控的研究环境下访问。多样性不足: 现有数据集在疾病种类、成像设备、伪造方法、篡改复杂度等方面覆盖不足。应对医学图像伪造检测研究存在的挑战, 解决方案有: ① 研究特定伪造技术的检测方法: 针对 CT-GAN 等生成对抗网络生成的医学图像小区域伪造, 提出两阶段级联框架, 通过局部检测阶段训练检测器网络, 有效减少真实区域的干扰信息[14]。针对 JPEG 压缩伪影掩盖伪造痕迹的问题, 提出多尺度上下文空间 - 通道相关网络, 通过融合不同尺度的特征区分 JPEG 伪影与伪造痕迹[24]。② 融合深度学习与特征增强技术: 结合超边界到像素方向分割与深度卷积神经网络方法, 通过分割技术增强相似图像块的联系, 提升检测精度; 使用空间感知模块与约束卷积结合的方法, 专注于检测篡改痕迹, 同时抑制语义内容干扰; 自监督三元组辅助损失网络, 通过自引导机制和交换损失实现轻量级篡改区域检测, 无需预训练或领域特定损失; 医学图像水印技术不仅检测篡改, 还能恢复被篡改区域, 保障远程医疗中的图像真实性[48]。融合错误级别分析与卷积神经网络的算法, 专注于检测复制 - 移动和拼接伪造[29]。从单一技

术检测向多模态、自适应和可解释性方向演进,同时加强对抗新型生成式 AI 伪造检测的研究[11]。③ 鲁棒性与泛化能力提升:现有方法在低质量压缩图像等复杂场景中性能显著下降,需增强对多级伪造的检测能力[16];通过数据增强(如局部伪影放大)、多源学习和指纹检测等策略,提高模型对未知伪造类型和受损图像的泛化能力[49]。④ 标准化数据集与评估基准:医学图像伪造检测技术的发展,需要建立新一代医学图像伪造检测专用数据集[49],涵盖多样化医学图像篡改类型的大规模数据集,如复制-移动、AI生成图像等,并制定统一评估指标以推动研究可比性。

6. 未来研究方向与展望

面对日益复杂的医疗数据安全挑战,医学图像伪造检测正向多技术融合、特征精细化等方向发展[30],以应对医学图像伪造技术的快速演进,构建兼具高泛化性、实时性和可解释性的检测系统是未来的方向:

① 无监督学习与跨域应用:医学图像伪造检测具有特殊性,数据标注专业性很强,存在标注数据不足问题[11],解决这个问题,需整合无监督学习技术,减少对标注数据的依赖,并探索跨领域(如保险、司法、科研)的检测应用。通过生成合成伪造数据,如 GAN 生成的小区域篡改 CT 图像,来增强模型对未知伪造方法的适应性[24];② 实时检测与计算效率优化:针对医疗场景的实时性需求,需优化算法计算效率,开发更轻量化的实时检测模型[11],FR-CNN 等快速区域检测网络和轻量化 CNN 架构通过改进特征提取速度,已展现出实时分析的潜力[11]。未来需进一步平衡检测精度与计算资源消耗[28];③ 局部伪造与小区域检测:针对医学图像中仅占 1%面积的小区域篡改,如 CT-GAN 伪造,需开发级联检测框架。文献提出通过局部子图像训练和像素级不一致性指纹捕捉,用 patch-channel correspondence 技术来提升微小篡改的定位能力[32];④ 多模态融合与互补性分析:结合互补性检测方法,如 ELA 与 CNN 融合可提升对复制移动和拼接等多类型伪造的识别率[29]。要研究如何融合传统取证方法,如 JPEG 重压缩分析与深度学习模型,探索大模型在医学图像伪造检测中的应用潜力[49];⑤ 安全性与公平性研究:研究检测模型自身的安全性,在对抗扰动防御中,主要针对 Deep fake 伪造和身份保护机制,如 FOR 框架防止特定身份被伪造[27],同时需解决数据偏见导致的公平性问题[49]。医学图像伪造检测技术研究展望:① 轻量化实时检测与跨领域应用:在医疗场景中,伪造图像可能直接影响临床决策,这就需要模型轻量化,提升实时检测能力,才能避免伤害病人的诊断与治疗,同时,在保险、司法鉴定、Deep fake 与 AI 生成图像等跨领域应用时,需开发通用性强的检测系统[50],以应对不同场景下的伪造类型。② 数据集与泛化能力优化:现有方法依赖已知伪造数据训练,对未知方法表现不佳。未来需构建更全面的医学标准数据集,通过数据增强或测试时自适应技术提升泛化性;针对 JPEG 压缩医学图像伪造的检测,需保留伪造痕迹同时消除压缩干扰[24]。③ 多模态与可解释性研究:结合视觉-语言检测或空间感知模块,如 SRM 与约束卷积结合[48],可提升复杂场景下的检测精度,开展多模态模型研究,才能适应医学图像不断进步的要求。此外,通过 patch-channel 对应关系可视化伪造特征,增强模型可解释性。④ 安全与伦理考量:研究需关注检测系统的安全性和公平性,同时开发主动防护技术,如 FOR 机制保护特定身份图像,以应对生成式 AI 的威胁[27]。

参考文献

- [1] Zhao, K., Yuan, X., Xie, Z., Xiang, Y., Huang, G. and Feng, L. (2023) SPA-Net: A Deep Learning Approach Enhanced Using a Span-Partial Structure and Attention Mechanism for Image Copy-Move Forgery Detection. *Sensors*, **23**, Article No. 6430. <https://doi.org/10.3390/s23146430>
- [2] Hussain, S., Mubeen, I., Ullah, N., Shah, S.S.U.D., Khan, B.A., Zahoor, M., et al. (2022) Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review. *BioMed Research International*, **2022**, Article ID: 5164970. <https://doi.org/10.1155/2022/5164970>
- [3] Albakri, A.A., Alzahrani, M.M. and Alghamdi, S.H. (2024) Medical Imaging in Pregnancy: Safety, Appropriate Utilization,

- and Alternative Modalities for Imaging Pregnant Patients. *Cureus*, **16**, e54346. <https://doi.org/10.7759/cureus.54346>
- [4] Rong, C., Li, Z., Li, R. and Wang, Y. (2024) Spatial-Aware Contrastive Learning for Cross-Domain Medical Image Registration. *Medical Physics*, **51**, 8141-8150. <https://doi.org/10.1002/mp.17311>
- [5] Supriyadi, M.R., Samah, A.B.A., Muliadi, J., Awang, R.A.R., Ismail, N.H., Majid, H.A., *et al.* (2025) A Systematic Literature Review: Exploring the Challenges of Ensemble Model for Medical Imaging. *BMC Medical Imaging*, **25**, Article No. 128. <https://doi.org/10.1186/s12880-025-01667-4>
- [6] Jia, H., Jiao, Q. and Liu, M. (2024) Special Issue: Artificial Intelligence in Advanced Medical Imaging. *Bioengineering*, **11**, Article No. 1229. <https://doi.org/10.3390/bioengineering11121229>
- [7] Estaji, M., Hosseini, B., Bozorg-Qomi, S. and Ebrahimi, B. (2023) Pathophysiology and Diagnosis of Diabetic Retinopathy: A Narrative Review. *Journal of Investigative Medicine*, **71**, 265-278. <https://doi.org/10.1177/10815589221145040>
- [8] Gao, W., Wang, C., Li, Q., Zhang, X., Yuan, J., Li, D., *et al.* (2022) Application of Medical Imaging Methods and Artificial Intelligence in Tissue Engineering and Organ-on-a-Chip. *Frontiers in Bioengineering and Biotechnology*, **10**, Article ID: 985692. <https://doi.org/10.3389/fbioe.2022.985692>
- [9] Ghaderi, H. and April, A. (2023) Assessing Available Open-Source PACS Options. *Journal of Digital Imaging*, **36**, 2323-2328. <https://doi.org/10.1007/s10278-021-00435-4>
- [10] Li, Q., Wang, C., Zhou, X. and Qin, Z. (2022) Image Copy-Move Forgery Detection and Localization Based on Super-BPD Segmentation and DCNN. *Scientific Reports*, **12**, Article No. 14987. <https://doi.org/10.1038/s41598-022-19325-y>
- [11] Radhakrishnan, A., Sawant, T.N., Raghuram, C., Railis, D.J. and Singh, H. (2025) Exploring Machine Learning Approaches for Efficient Image Forgery Detection. *Journal of Forensic Sciences*, **70**, 1375-1391. <https://doi.org/10.1111/1556-4029.70069>
- [12] Amiri, E., Mosallanejad, A. and Sheikhamadi, A. (2024) CFDMI-SEC: An Optimal Model for Copy-Move Forgery Detection of Medical Image Using SIFT, EOM and CHM. *PLOS ONE*, **19**, e0303332. <https://doi.org/10.1371/journal.pone.0303332>
- [13] Xue, Z., Jiang, X., Liu, Q. and Wei, Z. (2023) Global-Local Facial Fusion Based GAN Generated Fake Face Detection. *Sensors*, **23**, Article No. 616. <https://doi.org/10.3390/s23020616>
- [14] Zhang, J., Huang, X., Liu, Y., Han, Y. and Xiang, Z. (2024) Gan-Based Medical Image Small Region Forgery Detection via a Two-Stage Cascade Framework. *PLOS ONE*, **19**, e0290303. <https://doi.org/10.1371/journal.pone.0290303>
- [15] Lu, W., Liu, L., Zhang, B., Luo, J., Zhao, X., Zhou, Y., *et al.* (2024) Detection of Deepfake Videos Using Long-Distance Attention. *IEEE Transactions on Neural Networks and Learning Systems*, **35**, 9366-9379. <https://doi.org/10.1109/tnnls.2022.3233063>
- [16] Singh, U., Rathor, S. and Kumar, M. (2024) Advanced Framework for Multilevel Detection of Digital Video Forgeries. *Annals of the New York Academy of Sciences*, **1543**, 180-193. <https://doi.org/10.1111/nyas.15257>
- [17] Peng, C., Sun, F., Liu, D., Wang, N. and Gao, X. (2024) Local Artifacts Amplification for Deepfakes Augmentation. *Neural Networks*, **180**, Article ID: 106692. <https://doi.org/10.1016/j.neunet.2024.106692>
- [18] Zhao, H., Li, X., Xu, B. and Liu, H. (2024) Combined Spatial and Frequency Dual Stream Network for Face Forgery Detection. *PeerJ Computer Science*, **10**, e1959. <https://doi.org/10.7717/peerj-cs.1959>
- [19] Bao, Q., Wang, Y., Hua, H., Dong, K. and Lee, F. (2024) An Anti-Forensics Video Forgery Detection Method Based on Noise Transfer Matrix Analysis. *Sensors*, **24**, Article No. 5341. <https://doi.org/10.3390/s24165341>
- [20] Amiri, E., Mosallanejad, A. and Sheikhamadi, A. (2024) The Optimal Model for Copy-Move Forgery Detection in Medical Images. *Journal of Medical Signals & Sensors*, **14**, Article No. 5. https://doi.org/10.4103/jmss.jmss_35_22
- [21] Sharma, P., Kumar, M. and Sharma, H. (2022) Comprehensive Analyses of Image Forgery Detection Methods from Traditional to Deep Learning Approaches: An Evaluation. *Multimedia Tools and Applications*, **82**, 18117-18150. <https://doi.org/10.1007/s11042-022-13808-w>
- [22] Yang, J., Xie, A., Mai, T. and Chen, Y. (2025) Dfst-unet: Dual-Domain Fusion Swin Transformer U-Net for Image Forgery Localization. *Entropy*, **27**, Article No. 535. <https://doi.org/10.3390/e27050535>
- [23] Akram, A., Jaffar, M.A., Rashid, J., Mahmood, K. and Ghani, A. (2025) Advanced Digital Image Forensics: A Hybrid Framework for Copy-Move Forgery Detection in Multimedia Security. *Journal of Forensic Sciences*, **70**, 1801-1823. <https://doi.org/10.1111/1556-4029.70076>
- [24] Shan, W., Yue, J., Ding, S.X. and Qiu, J. (2025) MSCSCC-Net: Multi-Scale Contextual Spatial-Channel Correlation Network for Forgery Detection and Localization of JPEG-Compressed Image. *Scientific Reports*, **15**, Article No. 12509. <https://doi.org/10.1038/s41598-025-97555-6>
- [25] Lin, Y.K. and Yen, T.Y. (2023) A Meta-Learning Approach for Few-Shot Face Forgery Segmentation and Classification. *Sensors*, **23**, Article No. 3647. <https://doi.org/10.3390/s23073647>

- [26] Hua, Y., Shi, R., Wang, P. and Ge, S. (2023) Learning Patch-Channel Correspondence for Interpretable Face Forgery Detection. *IEEE Transactions on Image Processing*, **32**, 1668-1680. <https://doi.org/10.1109/tip.2023.3246793>
- [27] Zhang, H., Hu, C., Min, S., Sui, H. and Zhou, G. (2024) TSFF-Net: A Deep Fake Video Detection Model Based on Two-Stream Feature Domain Fusion. *PLOS ONE*, **19**, e0311366. <https://doi.org/10.1371/journal.pone.0311366>
- [28] Joshi, D., Kashyap, A. and Arora, P. (2025) Optimized Detection and Localization of Copy-Rotate-Move Forgeries Using Biogeography-Based Optimization Algorithm. *Journal of Forensic Sciences*, **70**, 1392-1413. <https://doi.org/10.1111/1556-4029.70068>
- [29] Nagm, A.M., Moussa, M.M., Shoitan, R., Ali, A., Mashhour, M., Salama, A.S., *et al.* (2024) Detecting Image Manipulation with ELA-CNN Integration: A Powerful Framework for Authenticity Verification. *PeerJ Computer Science*, **10**, e2205. <https://doi.org/10.7717/peerj-cs.2205>
- [30] Alabrah, A. (2025) SFTA-Net: A Self-Supervised Approach to Detect Copy-Move and Splicing Forgery to Leverage Triplet Loss, Auxiliary Loss, and Spatial Attention. *PeerJ Computer Science*, **11**, e2803. <https://doi.org/10.7717/peerj-cs.2803>
- [31] Atallah, A.M., Mahmoud, I.I. and Ali, H.S. (2023) Robust Dense-Field Based Copy-Move Forgery Localization Using Generic Radial Harmonic Fourier Moment Invariants. *Journal of Forensic Sciences*, **69**, 139-152. <https://doi.org/10.1111/1556-4029.15420>
- [32] Mareen, H., De Neve, L., Lambert, P. and Van Wallendael, G. (2023) Harmonizing Image Forgery Detection & Localization: Fusion of Complementary Approaches. *Journal of Imaging*, **10**, Article No. 4. <https://doi.org/10.3390/jimaging10010004>
- [33] Kong, C., Luo, A., Wang, S., Li, H., Rocha, A. and Kot, A.C. (2025) Pixel-Inconsistency Modeling for Image Manipulation Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **47**, 4455-4472. <https://doi.org/10.1109/tpami.2025.3541028>
- [34] Ahmad, I. and Alqurashi, F. (2024) Early Cancer Detection Using Deep Learning and Medical Imaging: A Survey. *Critical Reviews in Oncology/Hematology*, **204**, Article ID: 104528. <https://doi.org/10.1016/j.critrevonc.2024.104528>
- [35] Cai, Y., Zhang, W., Chen, H. and Cheng, K. (2025) Medianomaly: A Comparative Study of Anomaly Detection in Medical Images. *Medical Image Analysis*, **102**, Article ID: 103500. <https://doi.org/10.1016/j.media.2025.103500>
- [36] Dong, J., Wang, W. and Tan, T. (2013). CASIA Image Tampering Detection Evaluation Database. 2013 *IEEE China Summit and International Conference on Signal and Information Processing*, Beijing, 6-10 July 2013, 422-426. <https://doi.org/10.1109/chinasip.2013.6625374>
- [37] Tralic, D., Zupancic, I., Grgic, S. and Grgic, M. (2013) CoMoFoD-New Database for Copy-Move Forgery Detection. *Proceedings of the Proceedings ELMAR-2013*, Zadar, 25-27 September 2013, 49-54.
- [38] Silva, E., Carvalho, T., Ferreira, A. and Rocha, A. (2015) Going Deeper into Copy-Move Forgery Detection: Exploring Image Telltales via Multi-Scale Analysis and Voting Processes. *Journal of Visual Communication and Image Representation*, **29**, 16-32. <https://doi.org/10.1016/j.jvcir.2015.01.016>
- [39] Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A. and Serra, G. (2011) A Sift-Based Forensic Method for Copy-Move Attack Detection and Transformation Recovery. *IEEE Transactions on Information Forensics and Security*, **6**, 1099-1110. <https://doi.org/10.1109/tifs.2011.2129512>
- [40] Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., Del Tongo, L. and Serra, G. (2013) Copy-Move Forgery Detection and Localization by Means of Robust Clustering with J-Linkage. *Signal Processing: Image Communication*, **28**, 659-669. <https://doi.org/10.1016/j.image.2013.03.006>
- [41] Cozzolino, D., Poggi, G. and Verdoliva, L. (2015) Efficient Dense-Field Copy-Move Forgery Detection. *IEEE Transactions on Information Forensics and Security*, **10**, 2284-2297. <https://doi.org/10.1109/tifs.2015.2455334>
- [42] Wen, B., Zhu, Y., Subramanian, R., Ng, T., Shen, X. and Winkler, S. (2016). COVERAGE—A Novel Database for Copy-Move Forgery Detection. 2016 *IEEE International Conference on Image Processing (ICIP)*, Phoenix, 25-28 September 2016, 161-165. <https://doi.org/10.1109/icip.2016.7532339>
- [43] RSNA Pneumonia Detection Challenge. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
- [44] Nguyen, H.Q., Lam, K., Le, L.T., Pham, H.H., Tran, D.Q., Nguyen, D.B., *et al.* (2022) Vindr-cxr: An Open Dataset of Chest X-Rays with Radiologist's Annotations. *Scientific Data*, **9**, 1-7. <https://doi.org/10.1038/s41597-022-01498-w>
- [45] Bejnordi, B.E., Veta, M., Van Diest, *et al.* (2017) Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer. *JAMA*, **318**, 2199-2210.
- [46] Baid, U., Ghodasara, S., Mohan, S., *et al.* (2021) The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification.
- [47] Li, L., Zhang, K., Lu, J. and Zhang, S. (2025) Multi-Label Classification for Image Tamper Detection Based on Swin-T Segmentation Network in the Spatial Domain. *PeerJ Computer Science*, **11**, e2775. <https://doi.org/10.7717/peerj-cs.2775>

-
- [48] Taj, R., Tao, F., Kanwal, S., Almogren, A., Altameem, A. and Ur Rehman, A. (2024) A Reversible-Zero Watermarking Scheme for Medical Images. *Scientific Reports*, **14**, Article No. 17320. <https://doi.org/10.1038/s41598-024-67672-9>
 - [49] Peng, C., Chen, T., Liu, D., Guo, H., Wang, N. and Gao, X. (2025) Revisiting Face Forgery Detection Towards Generalization. *Neural Networks*, **187**, Article ID: 107310. <https://doi.org/10.1016/j.neunet.2025.107310>
 - [50] Martin-Rodriguez, F., Garcia-Mojon, R. and Fernandez-Barciela, M. (2023) Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks. *Sensors*, **23**, Article No. 9037. <https://doi.org/10.3390/s23229037>