arXiv:2603.24296v1 [cs.CV] 25 Mar 2026

AMIF：内置身份验证的可授权医学图像融合模型

Jie Song[1], Jun Jia[2†], Wei Sun[3],Wangqiu Zhou[4],Tao Tan[1†], Guangtao Zhai[2†]

[1] 澳门理工大学[2] 上海交通大学[3] 华东师范大学[4] 合肥工业大学

† 通讯作者。

摘要。 多模态图像融合能够精确定位和表征病灶，从而实现准确诊断，强化临床决策，并日益成为医学影像研究的关注重点。强大的多模态图像融合模型依赖高质量且具有临床代表性的多模态训练数据，以及经过严格设计的模型架构。因此，此类专业放射组学模型的开发是基于标准化采集、临床专业知识和算法设计能力的协作成果，相关知识产权亟需得到保护。然而，当前的多模态图像融合模型生成融合结果时，并未内置保护知识产权的机制，因而可能通过推理泄漏无意间暴露专有模型知识和敏感训练数据。例如，恶意用户可以利用融合结果、模型蒸馏或其他基于推理的逆向工程技术，近似复现专有模型的融合性能。为解决这一问题，我们提出**AMIF**，首个内置身份验证的**可授权医**学图像融合模型，将授权访问控制融入图像融合目标。对于未经授权的使用，AMIF 会在融合结果中嵌入明确且可见的版权标识符。相反，成功通过基于密钥的认证后，用户即可获得高质量的融合结果。为保证授权用户获得的融合质量，并增强对水印去除的鲁棒性，AMIF采用内容条件化水印记忆（CCWM）和通道—空间注意力调制可逆耦合（C-SAMIC）机制，将水印嵌入从外部输入依赖机制转变为模型内部能力。大量实验表明，AMIF能够灵活地保护版权，同时对水印移除攻击表现出很强的鲁棒性。

关键词： 多模态医学图像融合 · 可授权 · 版权保护

1 引言

多模态图像融合（MMIF）旨在将不同模态的互补信息整合为统一的表征。它已得到广泛研究和应用

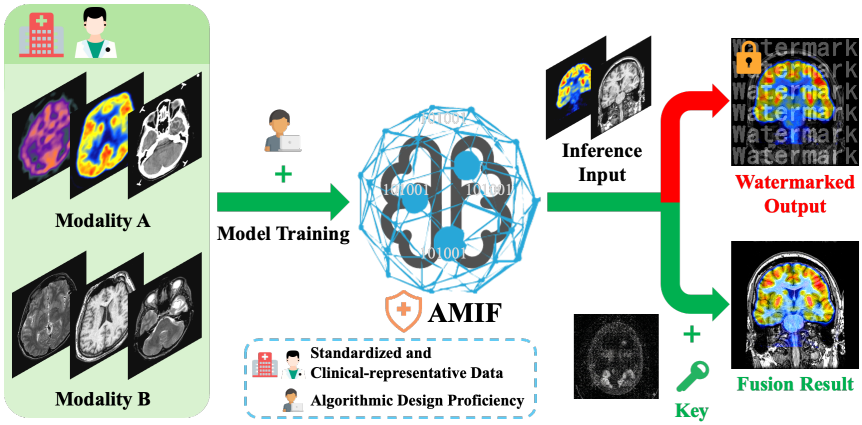

图1. AMIF 的设计动机。医学图像涉及隐私，不应随意分发。然而，现有融合模型通常忽视数据泄露风险和模型知识产权保护。AMIF 使用标准化且具有临床代表性的多模态数据进行训练，在学习融合的同时内置版权保护机制。默认情况下，它会输出带水印的融合图像，只有使用有效密钥才能还原无水印结果。

于医学影像、遥感和红外–可见光（IR-VIS）图像融合领域[50,1,49,38]。在医学影像中，不同模态侧重呈现解剖结构和生理状况的不同方面。例如，计算机断层成像（CT）擅长呈现骨骼及其他高密度结构。磁共振成像（MRI）能够提供更出色的软组织对比度和病灶勾画。正电子发射断层扫描/单光子发射计算机断层成像（PET/SPECT）等功能成像模态则能提供有关代谢活动和灌注的线索。融合这些互补信息可获得信息更加丰富的视觉表征，有助于提升肿瘤分割等下游任务的表现[43]，也能更好地支持临床诊断和治疗方案规划[15,47]。为提升融合质量和下游任务表现，研究人员在网络架构设计[48,41,36]，、特征交互[24]，、融合策略[22]，和损失函数优化[42]方面付出了大量努力。因此，此类专业放射组学模型的开发是一项协作成果，植根于标准化的医院影像采集、放射科医生和临床医生的领域专业知识，以及医学影像人工智能研究人员的算法设计能力。此类由医院、放射科医生和算法研究人员协作开发的专业医学图像融合模型，其相关知识产权需要得到保护。

然而，当前医学图像融合模型在开发和临床传播过程中忽视了知识产权（IP）保护。医学图像是敏感且高价值的资产。随着远程医疗和开放研究仓库的快速发展，融合图像面临更高的未授权使用、恶意篡改和知识产权盗用风险。现有融合模型通常输出“裸露”的融合结果，缺乏内置的所有权验证和版权保护。事后添加水印是一种可行的

补救方法。然而，这通常依赖外部水印模型或第三方服务，可能引入数据泄露风险。此外，它还将“融合”与“保护”分为两个独立阶段。因此，水印与融合内容的耦合较弱，在攻击下更容易被移除。在实际应用中，真实场景部署还需要授权。授权用户应能够在下游分析或临床诊断前无损地移除水印，并通过授权密钥实现这一操作。因此，需要一种统一模型，实现高质量融合、可追溯的版权保护，并通过可控的水印去除支持可撤销授权。这对于医学图像在现实流通中的安全合规使用至关重要。因此，我们提出了**AMIF**，首个具备内置身份验证的**可授权医学 图像 融合**模型。AMIF 首次在统一框架中联合建模版权保护约束与融合目标。AMIF 进一步引入基于密钥的授权机制，以满足临床完整性和可用性的严格要求。图1展示了 AMIF 的端到端应用工作流程。在未经授权的场景中，模型输出带有可见版权水印的融合图像，适用于公开共享和存储。提供正确密钥后，模型可按需去除水印，并恢复干净、无水印的融合结果。这支持授权临床使用和下游分析。为保持授权使用时的融合质量，我们设计了内容条件化水印记忆（CCWM）模块。该模块使模型无需显式水印输入，即可根据内部参数生成水印表征。CCWM 引入一组可学习水印记忆向量，并利用双向交叉注意力将水印记忆与输入特征耦合，从而在水印表征与输入特征之间建立内容绑定关系。这有助于将水印信号更紧密地写入受保护表示。此外，为增强对水印移除攻击的鲁棒性，我们提出通道—空间注意力调制可逆耦合（C-SAMIC）机制。该机制在特征层面增强图像表示与水印信号的融合，提高水印与图像特征之间的兼容性，同时增强对未经授权的移除操作的鲁棒性。

总之，我们的贡献如下：

– 我们提出了**AMIF**，这是首个**可授权的医用图像融合**模型，并内置身份认证功能。AMIF 将版权保护明确纳入融合目标，同时支持获得授权的用户使用高质量融合结果。

– 我们设计了内容条件化水印记忆（CCWM）模块和通道-空间注意力调制可逆耦合（C-SAMIC）机制，分别提升授权融合质量和去除水印的鲁棒性。

– 在公开医学数据集上的实验表明，AMIF 能够在有效密钥授权下，按需移除水印并恢复无水印结果。与现有最先进水平（SOTA）方法相比，该方法也取得了具有竞争力的融合性能。

2 相关工作

2.1 基于深度学习的多模态图像融合

基于深度学习的多模态图像融合已成为主流研究方向。早期方法大多基于卷积神经网络（CNNs），通过学习模态特定表示，并在特征空间中进行融合，以提升融合质量[48,41,36,37,42,6,9]。除了卷积神经网络框架外，一些方法还利用生成对抗网络（GANs）学习分布映射，以实现融合[17,46]。近年来，基于Transformer的方法利用注意力机制交换跨模态信息并建模长程依赖，从而提升对结构和纹理细节的表示能力[24]。此外，一些研究还探索了用于融合的基于Mamba的选择性状态空间建模[35,29]。例如，Xie 等人[35]提出了动态特征融合模块，以增强动态纹理、差异感知和跨模态特征强化，同时抑制冗余信息。另一方面，Wang 等人[22] 从多尺度或频域视角设计融合策略，以更好地保留显著结构和细节。Zhao 等人[49]提出了相关性驱动的特征分解融合网络，并设计了相关性驱动损失，以促进低频相关性和高频互补性。Liang 等人[19] 提出了 DeFusion，通过基于分解的表示学习实现融合。为解决跨模态错位问题，Huang 等人[14]提出了 ReCoNet，该方法采用轻量级配准和循环校正来减轻伪影并提升鲁棒性。Nam 等人[27] 利用隐式坐标神经表示进行多图像对齐与融合。此外，图像融合通常还会与分割等下游任务相结合，更具信息量的融合表示有助于提高分割准确率[20,25,43]。

2.2 版权保护与图像水印

图像水印是一种广泛使用的版权保护技术。它将所有权信息嵌入图像，以支持归属认定和可追溯性。根据呈现方式，水印可分为可见水印[28,44] 或不可见水印[45]。传统水印方法通常依赖手工设计的信号处理，以及在变换域中的嵌入规则[2,12]。近年来，基于深度学习的水印方法[30,4] 进一步提升了水印嵌入的灵活性。然而，现有的大多数水印处理流程仅将水印作为后处理步骤。它们与图像融合等上游生成任务相互分离。这往往导致水印信号与图像内容之间的耦合较弱。在本研究中，我们将基于水印的版权保护与多模态医学图像融合统一到一个建模框架中。

2.3 可逆神经网络

可逆神经网络的概念可追溯至 Dinh等人提出的非线性独立成分估计（NICE），该方法采用可逆加性耦合变换来实现双向映射。基于这一思路，Dinh等人进一步将卷积结构引入耦合模型，并采用

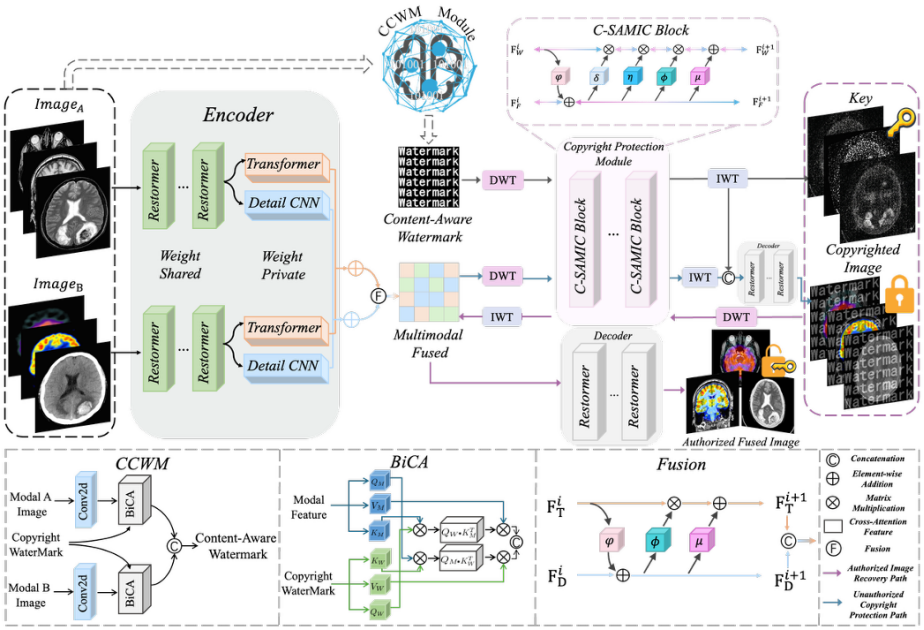

图 2. AMIF 通过内置身份验证实现可授权的多模态医学图像融合。共享的 Restormer 编码器提取共有特征，专用编码器则捕获各模态的特有线索，并将其融合为多模态表示。CCWM 首先生成一种内部内容感知水印，其中携带隐式内容线索。随后，C-SAMIC 通过紧密耦合的小波域可逆耦合，将其嵌入受保护特征中，生成受版权保护的输出及相应密钥。使用有效密钥时，逆映射会移除水印，并恢复高质量的无水印融合结果，供授权使用。整个流程以统一且协同的方式运行。

多尺度设计以降低计算成本。Kingma等人引入了可逆卷积等关键组件，提升了模型的表达能力和可扩展性，使 INN 更适用于图像建模。后续研究将 INN 拓展到多种视觉任务，包括深度图像隐藏[39], 对抗样本生成[5], 超分辨率[34],以及提升用于图像分类的主干网络特征表示[3,10]。为了将水印信号有效嵌入图像特征，并在获得授权后恢复高质量的无水印融合结果，我们量身设计了一种可逆框架，以加强水印标识符与图像表示之间的绑定关系。

3 方法

3.1 总体架构

为实现高质量融合，同时为融合结果提供版权保护并为融合模型提供知识产权保护，我们提出了一种可授权

医学图像融合框架 AMIF，如图 2 所示。AMIF 在单一模型中统一了三种紧密耦合的机制：跨模态互补融合、耦合水印注入和密钥驱动的可逆恢复。这一设计兼顾了融合质量与授权版权保护要求。

给定来自不同模态的两幅输入图像， $Image_A$ 和 $Image_B$，我们使用 Restormer[40] 作为共享编码器和解码器，提取跨模态基础表征。随后，我们为每种模态引入一个专用编码器，以捕获模态特定特征。每个专用编码器都包含一个基于 Lite Transformer[33]的全局分支，以及一个基于 INN[8]的局部细节 CNN 分支。这种设计能够同时建模长程依赖和精细纹理。通过融合模块整合两种模态的特征，该模块由 φ、 ϕ和 μ 函数组成，以获得融合后的多模态特征。在未授权模式下，CCWM 模块根据内部可学习记忆参数生成水印表征，并以输入图像为条件进行调节，详见第 3.2 节。由于水印在生成过程中与输入特征相互作用，因此会隐式编码当前多模态图像对的内容线索。这使水印表征与融合内容更加兼容，也有助于后续将其写入受保护特征，并稳定后续基于密钥的恢复。接着，我们将融合内容表征和 CCWM 水印表征映射至小波域。我们对它们应用离散小波变换 (DWT)，将其分解为多频分量。这种频率分解使我们能够以更可控的方式在不同频带中注入水印信号。随后，我们拼接小波域特征，并将其输入版权保护模块。该模块由 N 个 C-SAMIC 模块组成（详见第 3.3 节）。每个 C-SAMIC 模块都会执行可逆特征耦合，并在特征层面逐步将水印线索写入内容流。

重要的是，版权保护模块会产生两个输出：受保护的融合特征，以及剔除了冗余信息、用作逆变换密钥的特征流。Restormer 解码器对受保护特征进行解码，生成受版权保护的融合图像。在授权模式下，给定有效密钥后，模型会对 C-SAMIC 块应用逆映射，在特征空间中移除水印并重建授权融合图像。

3.2 CCWM 模块

现有的图像水印方法通常依赖显式提供的水印模板或载荷，并将其嵌入受保护图像中。这种外部输入范式暴露了攻击面。攻击者可以篡改、替换或绕过水印输入，从而削弱保护效果。为解决这一问题并提升无水印恢复质量，我们提出了 CCWM，将水印转变为模型的一项内在能力。版权水印表征由可学习的内部参数生成，并与源图像内容耦合，作为条件。因此，水印能够从输入图像中提取并编码内容相关信息。它还保留了支持后续图像恢复的隐式跨模态线索，并在嵌入过程中提升了与融合特征的兼容性。CCWM 将一组可学习的水印记忆向量初始化为版权标识符的载体。轻量级 2D 卷积从两种输入模态中提取特征。接下来，

双向交叉注意力（BiCA）在水印向量与输入特征之间建立双向关联。一方面，水印向量关注输入特征，以形成内容条件化水印特征。另一方面，输入特征反向关注水印向量，从而增强二者的相互感知。最后，我们沿通道维度拼接两个水印特征，得到最终的内容感知水印表示。

3.3 C-SAMIC机制

为了在特征层面更有效地将版权水印嵌入受保护特征，并使嵌入能够抵御未经授权的水印去除，同时仍能在授权下实现可控的无水印恢复，我们设计了一个C-SAMIC模块。该模块通过可逆映射交替耦合两个输入特征流，以交换信息。具体而言，我们将输入拆分为两部分： F_F^i 表示待保护的内容特征， F^i 表示水印特征。$_W$每个耦合模块由一组可学习函数组成，记作 φ、 δ、 η、 ϕ 和 μ。函数 φ、 ϕ 和 μ 通过密集块 [32]实现。 δ 采用 [13]中的通道注意力机制，以增强沿通道维度的信息流交换。 η 引入空间注意力机制，以引导空间位置间的特征交互与对齐。在未授权模式下，版权保护的计算方式如下：

$$F_{\mathrm{F}}^{i+1} = F_{\mathrm{F}}^{i} + \varphi\left(F_W^i\right), \tag{1}$$

$$F_W^{i+1} = F_W^i \odot \delta\left(F_{\mathrm{F}}^{i+1}\right) \odot \exp\left(\alpha\left(\eta\left(F_{\mathrm{F}}^{i+1}\right) \odot \phi\left(F_{\mathrm{F}}^{i+1}\right)\right)\right) + \mu\left(F_{\mathrm{F}}^{i+1}\right). \tag{2}$$

其中， $\alpha(\cdot)$表示由常数因子缩放的Sigmoid函数， $\odot$ 表示逐元素乘法。

在授权模式下，水印恢复对应于版权保护过程的逆向计算，信息流方向相反。其公式如下：

$$F_K^i = F_K^{i+1} - \mu\left(F_{\mathrm{C}}^{i+1}\right) \odot \exp\left(-\alpha\left(\phi(F_{\mathrm{C}}^{i+1}) \oslash \eta\left(F_{\mathrm{C}}^{i+1}\right)\right)\right) \oslash \delta\left(F_{\mathrm{C}}^{i+1}\right), \tag{3}$$

$$F_{\mathrm{C}}^{i} = F_{\mathrm{C}}^{i+1} - \varphi\left(F_K^i\right). \tag{4}$$

其中， F_K^i 表示密钥， F_C^i 表示受版权保护图像， $\oslash$ 表示逐元素除法。

3.4 损失函数

AMIF 经过训练，可适配两种使用模式。在未授权模式下，前向传播会生成含水印融合输出。在授权模式下，逆向传播会在给定密钥的作用下移除水印，并恢复出无水印结果。因此，我们将训练目标分为两部分。第一部分在未授权模式下约束融合质量和水印嵌入。第二部分在授权模式下约束可逆恢复质量。下面介绍这两个损失项中的

详情如下。

授权模式损失。 在授权模式下，AMIF 需要生成无水印结果，以支持临床应用和下游分析。受 [31,49], 等研究的启发，我们通过考虑强度保真度、梯度一致性、特征分解约束以及密钥条件下的特征恢复来设计损失函数，定义如下。

$$\mathcal{L}_{\text{fusion}} = \mathcal{L}_{\text{int}} + \alpha_1 \mathcal{L}_{\text{grad}} + \alpha_2 \mathcal{L}_{\text{decomp}} + \alpha_3 \mathcal{L}_{\text{krecov}} \tag{5}$$

$$\mathcal{L}_{\text{int}} = \frac{1}{HW} \left\| I_f - \max(I_{ModalA}, I_{ModalB}) \right\|_1 \tag{6}$$

$$\mathcal{L}_{\text{grad}} = \frac{1}{HW} \left\| |\nabla I_f| - \max(|\nabla I_{ModalA}|, |\nabla I_{ModalB}|) \right\|_1 \tag{7}$$

$$\mathcal{L}_{decomp} = \frac{\left(\mathcal{L}_{CC}^{D}\right)^2}{\mathcal{L}_{CC}^{B}} = \frac{\left(CC(\Phi_{ModalA}^{D}, \Phi_{ModalB}^{D})\right)^2}{CC\left(\Phi_{ModalB}^{B}, \Phi_{ModalB}^{B}\right) + \epsilon} \tag{8}$$

$$\mathcal{L}_{\text{krecov}} = \frac{1}{N} \sum_{n=1}^{N} \left(F_{\text{recov}}^{(n)} - F_{\text{ori}}^{(n)} \right)^2 \tag{9}$$

其中， I_f表示无水印融合图像， I_{ModalA} 和 I_{ModalB} 分别表示来自两种不同模态的输入图像。 ∇ 表示Sobel梯度算子。 Φ^D 和 Φ^B 分别表示细节特征和分解后的基础特征。 $CC(\cdot,\cdot)$表示相关系数算子， ϵ设为1。01，以确保分母为正。 α_1、α_2和 α_3 为调节参数。

未授权模式损失。 在未授权模式下，AMIF会从可学习的内置参数中生成水印表征，并将其写入融合结果。因此，输出携带可验证的版权标识符。为了监督这些内部水印参数的学习，我们将预测的水印与预定义水印标签对齐，并通过结合二元交叉熵（BCE）和Dice损失进行优化。具体而言，BCE提供稳定的逐像素监督，而Dice损失则缓解稀疏水印区域造成的类别不平衡，并更好地约束水印形状与结构。损失定义如下。

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{n=1}^{N} \left(I_{\text{wmt}}^{(n)} \log\left(\sigma(I_{\text{pwm}}^{(n)})\right) + \left(1 - I_{\text{wmt}}^{(n)}\right) \log\left(1 - \sigma(I_{\text{pwm}}^{(n)})\right) \right) \tag{10}$$

$$\mathcal{L}_{\text{Dice}} = \frac{1}{B} \sum_{b=1}^{B} \left(1 - \frac{2 \sum_{n=1}^{N_b} \sigma(I_{\text{pwm}}^{(b,n)})\, I_{\text{wmt}}^{(b,n)}}{\sum_{n=1}^{N_b} \left(\sigma(I_{\text{pwm}}^{(b,n)}) + I_{\text{wmt}}^{(b,n)}\right) + \epsilon} \right) \tag{11}$$

此外，为了以更自然、稳定的方式将水印信号写入受保护特征，我们引入了基于均方误差的内容一致性约束。具体而言，我们在像素域中对含水印融合输出 I_{wf}与参考图像 I_{wft}施加均方误差损失。参考图像定义为两个输入

模态与水印标签的平均值，即 1 3 ($I_{\text{modalA}} + I_{\text{modalB}} + I_{\text{wt}}$)。此外，为了进一步规范结构一致性，我们将特征映射到小波域，并对离散小波变换的低频分量计算均方误差损失。这有助于在低频结构带中稳定地注入水印。由此得到的损失定义如下：

$$\mathcal{L}_{\text{wm}} = \frac{1}{N}\sum_{n=1}^{N}\left(I_{\text{wf}}^{(n)} - I_{wft}^{(n)}\right)^2 \tag{12}$$

$$\mathcal{L}_{\text{wmlow}} = \frac{1}{N}\sum_{n=1}^{N}\left(\text{DWT}_{\text{low}}(F_{\text{wf}})^{(n)} - \text{DWT}_{\text{low}}(F)_{wft}^{(n)}\right)^2 \tag{13}$$

其中， F_{wf} 和 F_{wft} 分别表示带水印的低频特征和目标低频水印特征。 F_{wft} 定义为 1 2 ($F_f + F_{\text{wmt}}$)。

综上，AMIF 的总体优化损失函数可概括为：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{fusion}} + \alpha_4((\mathcal{L}_{\text{wm}} + \mathcal{L}_{\text{wmlow}}) + \alpha_5(\mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}})) \tag{14}$$

4 实验

4.1 实验设置

数据集和指标。 为评估所提出的 AMIF 模型，我们使用来自哈佛医学图像网站[5]的公开医学图像开展实验。该数据集涵盖三种典型的多模态图像对，包括 MRI-CT、MRI-PET 和 MRI-SPECT。由于 MRI-CT 子集相对较小，我们采用随机旋转进行数据增强，以扩大其规模。我们将所有图像对合并为一个统一数据集，其中 983 对用于训练，104 对用于验证，81 对用于测试。测试集包括 16 对 MRI-CT、25 对 MRI-PET 和 40 对 MRI-SPECT。对于定量评估，我们遵循 [23]并报告五种常用的融合指标。空间频率（SF）用于衡量空间活性，反映细节和纹理的丰富程度。互信息（MI）量化源图像中的信息在融合结果中的保留程度。视觉信息保真度（VIF）通过衡量视觉信息的保留程度来评估感知保真度。$Q^{AB/F}$ 评估从源图像到融合图像的边缘信息传递。结构相似性（SSIM）衡量图像之间的结构一致性。通常，数值越高表示融合质量越好。

实现细节。所有实验均在配备 NVIDIA GeForce RTX A6000 图形处理器（48GB 显存）的服务器上进行。所有图像均调整为 256×256。我们训练 200 个周期，批量大小为 2。我们使用 Adam 优化器，初始学习率为 1×10^{-4}，每 100 个周期衰减至原来的 0.5 倍。在编码器中，Restormer 和 Transformer 均使用 4 个模块、8 个注意力头和 64 个特征维度。解码器采用相同配置。对于公式（5）和（14），我们将 α_1 设为 α_5，依次取值 10、2、100、0.1 和 0.1，以匹配预期的优化优先级。

[5] http://www.med.harvard.edu/AANLIB/home.html

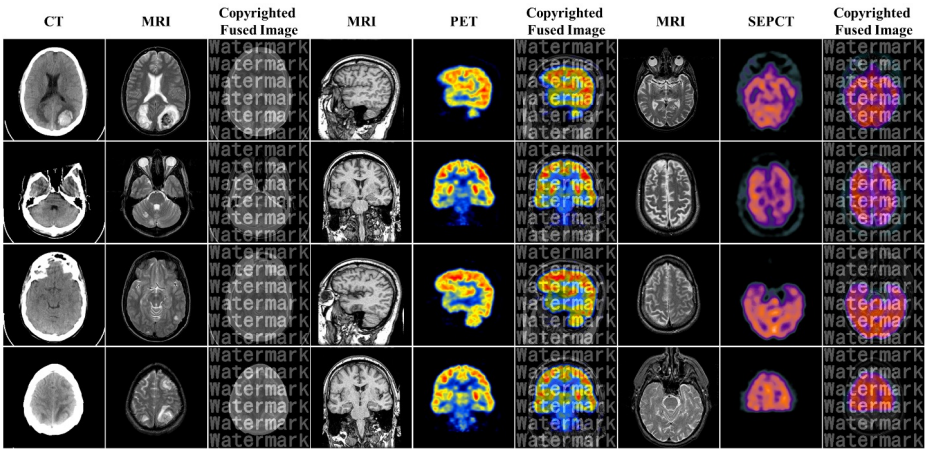

图3. 未授权模式下融合结果的可视化。AMIF 输出带有清晰可见水印的融合图像，以实现版权保护。

4.2 未授权融合

在本小节中，我们通过定性实验考察 AMIF 在未授权使用情况下的表现。在未经授权的场景中，输入图像无法直接生成用于临床分析的无水印融合结果。相反，AMIF 会生成带有嵌入式版权标识符的受保护融合输出。如图3所示，对于不同的模态对和样本，AMIF 始终会在融合图像上叠加清晰易读的可见水印，从而直观地声明视觉所有权并提供版权保护。同时，水印叠加不会造成明显的结构断裂或大面积伪影。融合图像仍保留了关键的跨模态信息，包括解剖轮廓以及显著的功能或代谢区域。这些结果表明，AMIF 在保持融合内容可用的同时，主动生成可验证的版权标记，展现了其在未授权传播场景中的内置自我保护能力和版权意识。

4.3 授权融合的最先进方法比较

在本节中，我们在授权模式下的测试集上评估 AMIF，此时水印会被有效移除。我们使用定量指标和定性结果，将 AMIF 与最先进的方法进行比较，包括 IFCNN [48], U2FUSION [36], FUSIONMAMBA [35], PSLPT [22],SWINFUSION [24],和 MMIF-INET [11]。

定量分析。 表1显示，AMIF 在 MRI–CT 上的空间频率、互信息、视觉信息保真度和$Q^{AB/F}$均达到最佳。其结构相似性指数也最高，与最佳方法并列。这表明，AMIF 能够恢复出空间细节活跃度高、边缘传递效果好的无水印结果。它还能保留更多跨模态信息，提升感知保真度，同时保持稳定的结构一致性。在 MRI–PET 和 MRI–单光子发射计算机断层扫描上，AMIF 在互信息、视觉信息保真度和 $Q^{AB/F}$方面优势更为明显，这些指标更能反映

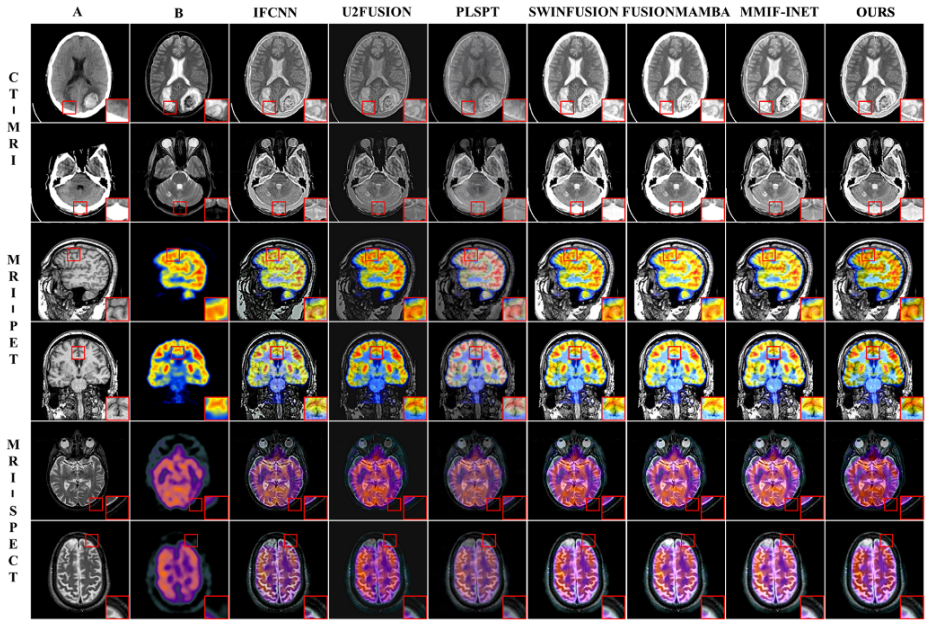

图 4. AMIF在三个多模态数据集上与不同融合模型生成的授权（无水印）融合图像定性结果。A和B表示两种输入模态。

信息保留、感知质量和边缘传递效果。其结构相似性指数与最佳方法相当。这表明，AMIF 在保留可靠解剖结构的同时，能够注入功能或代谢线索。尽管部分方法的空间频率更高，AMIF 在互信息、视觉信息保真度和 $Q^{AB/F}$方面仍始终表现更好。总体而言，定量结果与定性可视化结果一致，进一步验证了 AMIF 在经授权去除水印后仍能生成高质量且具有互补性的融合结果。

定性分析。 如图4所示，在授权模式下，AMIF使用有效密钥去除可见水印，且未引入明显的水印去除伪影。放大区域表明，AMIF在恢复后保留了更清晰的纹理和边缘。对于MRI–PET融合，尽管SWINFUSION、PLSPT和MMIF-INET保留了合理的细节，AMIF对细微皮层脑沟模式的保留效果更好。对于MRI–SPECT融合，IFCNN、U2FUSION和PLSPT在局部区域容易造成边缘平滑和精细纹理减弱。相比之下，AMIF保留了更清晰的边界和更丰富的细粒度纹理。对于MRI–CT融合，一些方法会削弱高密度CT结构或改变对比度。SWINFUSION和FUSIONMAMBA可能会过度增强高密度区域，并损害MRI软组织细节。AMIF则能更好地保留CT结构轮廓和MRI软组织细节。总体而言，AMIF在授权去除水印后能够恢复高质量的融合结果，为下游分析和临床参考提供支持。

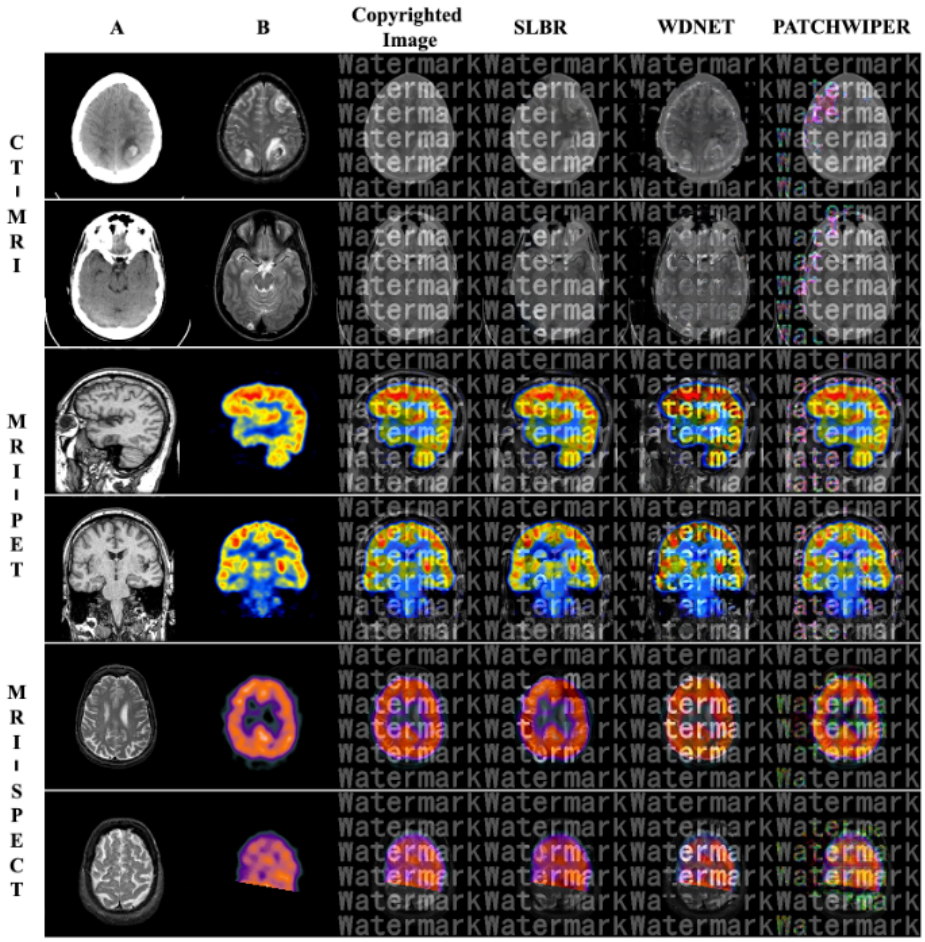

图5. 水印移除攻击下的视觉结果。A列和B列展示两种输入模态，随后依次为受版权保护的融合图像，以及 SLBR、WDNET 和 PATCHWIPER 的攻击结果。在未经授权的场景中强行移除水印会引入失真和伪影，从而降低融合内容的质量。

表1. 授权模式下与最先进方法的定量比较。使用有效密钥后，AMIF 会移除水印，并恢复无水印融合结果以供评估（红色 表示最佳结果，蓝色 表示次佳结果）。

方法	MRI-CT					MRI-PET					MRI-SPECT				
	SF	互信息指数	视觉信息保真度空间频率	$Q^{AB/F}$	结构相似性	SF	互信息指数	视觉信息保真度空间频率	$Q^{AB/F}$	结构相似性	SF	互信息指数	视觉信息保真度结构相似性指数	$Q^{AB/F}$	结
IFCNN[48]	31.57	1.81	0.4	0.55	0.65	30.09	1.63	0.44	0.55	0.43	18.52	1.56	0.46	0.57	0.56
U2FUSION[36]	17.97	1.66	0.37	0.40	0.44	17.19	1.76	0.43	0.38	0.34	11.02	1.61	0.43	0.37	0.39
SWINFUSION[24]	21.56	2.01	0.56	0.59	0.69	30.07	2.63	0.79	0.74	0.57	18.6	2.23	0.84	0.76	0.64
PSLPT[22]	18.98	1.62	0.37	0.35	0.5	16.03	1.74	0.51	0.39	0.38	10.36	1.63	0.53	0.41	0.46
FUSIONMAMBA[35]	34.35	2.01	0.57	0.52	0.70	35.58	2.01	0.59	0.62	0.47	20.97	1.93	0.69	0.7	0.62
MMIF-INET[11]	29.7	1.74	0.48	0.58	0.62	30.01	1.9	0.57	0.64	0.44	18.96	1.92	0.74	0.74	0.58
OURS	36.15	2.04	0.59	0.64	0.70	30.24	2.97	0.85	0.76	0.57	19.12	2.68	0.93	0.78	0.64

表 2. AMIF 在三个多模态测试集上有无 CCWM 的定量消融对比（**粗体** 表示最佳，W/ 和 W/O 分别表示有和没有）。

方法	MRI-CT					MRI-PET					MRI-SPECT				
	SF	互信息	VIF	$Q^{AB/F}$	SSIM	SF	互信息	VIF	$Q^{AB/F}$	SSIM	SF	互信息	视觉信息保真度结构相似性指数	$Q^{AB/F}$	结
我们的方法（不含 CCWM）	36.05	1.99	0.57	0.64	0.70	30.11	2.91	0.84	0.76	0.57	19.04	2.62	0.91	0.78	0.63
我们的方法（含 CCWM）	**36.15**	**2.04**	**0.59**	0.64	0.70	**30.24**	**2.97**	**0.85**	0.76	0.57	**19.12**	**2.68**	**0.93**	0.78	**0.64**

4.4 未经授权的水印移除攻击。

为评估 AMIF 抵御未经授权的水印移除的鲁棒性，我们采用了三种具有代表性的最先进可见水印去除方法，包括 PATCH- WIPER [26], SLBR [18], 和 WDNET [21]。我们将这些方法应用于授权模式下生成的带水印融合图像。如图 5 所示，在这些攻击下，我们的水印仍清晰可辨，仅出现局部损坏或平滑处理。同时，攻击后的结果，尤其是融合输出，表现出明显的信息损失。在水印减弱或部分移除的区域，医学内容显著退化。典型伪影包括边缘损坏、组织纹理过度平滑，以及显著功能或代谢区域的形状或对比度失真。具体而言，SLBR 通过渐进式复原来优化背景，但在 AMIF 输出中通常只能抑制部分水印图案。WD-NET 尝试通过分解分离水印成分，但在这种情况下往往会造成更严重的全局去纹理化。这会降低对比度并抹平细节，而水印仍未能被彻底移除。对于 AMIF 而言，水印信号在小波域中与内容深度耦合，因此 PATC HWIPER 的作用明显受限，还可能引入额外的局部颜色异常。这些结果表明，水印与特征空间中的融合内容高度耦合。没有密钥，攻击者几乎无法在不损害关键医学信息的情况下移除水印，从而有效遏制未经授权的删除。

消融研究。 本小节对 CCWM 和 C-SAMIC 进行了消融研究。所有变体均采用相同的训练方案、超参数和测试集。对于 CCWM，我们将其替换为简化设计，仅保留可学习的水印

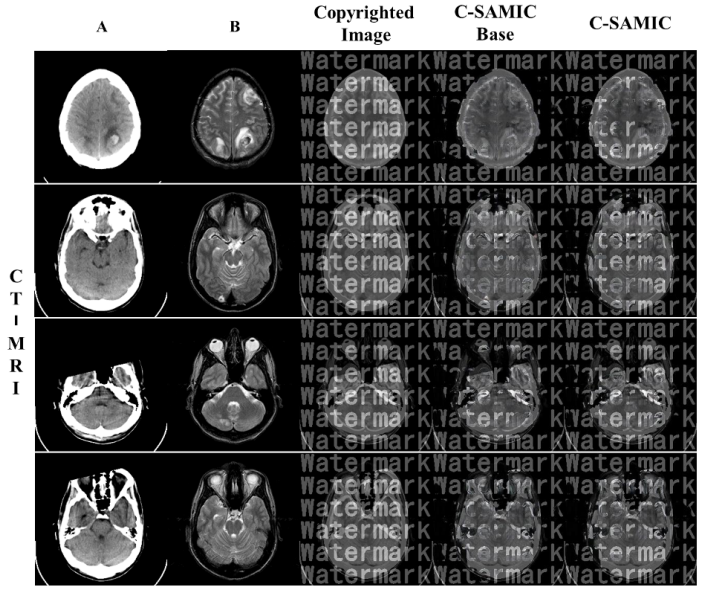

图 6. C-SAMIC 和 C-SAMIC Base 在 MRI-CT 测试集上的消融可视化结果。我们展示了 WDNet 攻击下的几组代表性样本对。C-SAMIC 对未经授权的水印移除具有更强的鲁棒性。

参数，其他模块保持不变。表 2 显示，移除 CCWM 后，三个数据集上的指标均一致下降。互信息和视觉信息保真度的降幅更为明显，它们分别反映信息保留和感知保真度。这表明 CCWM 会根据输入内容调整水印表征。由此产生的水印特征隐式编码了多模态线索。这为水印嵌入和密钥驱动恢复提供了更多信息丰富的内容信号，从而提升无水印重建质量和跨模态信息保留。我们进一步将完整的 C-SAMIC 与一个基础变体进行比较，该变体从耦合函数中移除了 δ 和 η。我们使用 WDNET [21] 作为水印去除攻击，并在 MRI–CT 测试集上进行评估。如图 6 所示，遭受攻击后，C-SAMIC 仍能保持水印清晰且连续，受保护内容仅出现轻微退化。相比之下，基础变体的水印图案有所减弱，内容损坏也更严重，例如细节过度平滑、解剖边界退化。这些结果表明，函数 δ 和函数 η 增强了通道和空间维度上的特征交互，使水印与结构内容耦合得更加紧密，从而提高了嵌入强度以及抵御未经授权的水印移除的鲁棒性。

5 结论

我们提出 AMIF，这是首个内置身份验证的可授权医学图像融合模型。AMIF 将多模态融合与版权保护统一于一个

框架中，实现了对融合结果的数据级所有权保护以及对模型内嵌知识产权的保护。在未授权模式下，它会输出带有明确可见水印的融合图像，以便进行所有权识别。在授权模式下，它通过基于密钥的认证去除水印，并恢复出高质量、无水印的融合结果，供临床应用和下游分析使用。AMIF 引入了 CCWM，它从内部可学习参数生成内容感知水印表示，并提供隐式模态线索，进一步提升融合质量。它还采用 C-SAMIC 模块，通过在小波域中将水印信号与受保护特征紧密耦合，增强水印嵌入鲁棒性。在 MRI–CT、MRI–PET 和 MRI–SPECT 数据集上的实验表明，AMIF 在授权恢复后实现了有竞争力的融合质量，并且能够抵御具有代表性的可见水印移除攻击。

参考文献

1. Bandara, W.G.C., Patel, V.M.：Hypertransformer：一种用于全色锐化的纹理与光谱特征融合 Transformer。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第1767–1777页（2022年） 2. Barni, M., Bartolini, F., Piva, A.：通过逐像素掩蔽改进基于小波的水印。IEEE图像处理汇刊**10**(5)，783–791（2001年） 3. Behrmann, J., Grathwohl, W., Chen, R.T., Duvenaud, D., Jacobsen, J.H.：可逆残差网络。载于：国际机器学习会议。第573–582页。PMLR（2019年） 4. Bui, T., Agarwal, S., Yu, N., Collomosse, J.：Rosteals：利用自动编码器潜在空间实现鲁棒隐写。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第933–942页（2023） 5. Chen, Z., Wang, Z., Huang, J.J., Zhao, W., Liu, X., Guan, D.：通过可逆神经网络实现不可感知对抗攻击。载于：AAAI人工智能会议论文集。第37卷，第414–424页（2023） 6. Deng, X., Dragotti, P.L.：用于多模态图像复原与融合的深度卷积神经网络。IEEE模式分析与机器智能汇刊**43**(10)，3333–3348（2020年） 7. Dinh, L., Krueger, D., Bengio, Y.：NICE：非线性独立成分估计。arXiv预印本 arXiv:1410.8516（2014年） 8. Dinh, L., Sohl-Dickstein, J., Bengio, S.：使用 Real NVP 进行密度估计。arXiv预印本 arXiv:1605.08803（2016年） 9. Gao, F., Deng, X., Xu, M., Xu, J., Dragotti, P.L.：多模态卷积字典学习。IEEE图像处理汇刊**31**，1325–1339（2022年） 10. Gomez, A.N., Ren, M., Urtasun, R., Grosse, R.B.：可逆残差网络：无需存储激活值的反向传播。神经信息处理系统进展**30** （2017年） 11. He, D., Li, W., Wang, G., Huang, Y., Liu, S.：MMIF-INET：基于可逆网络的多模态医学图像融合。信息融合**114**，102666（2025年） 12. Hernandez, J.R., Amado, M., Perez-Gonzalez, F.：静态图像的 DCT 域水印技术：检测器性能分析与一种新结构。IEEE图像处理汇刊**9**(1)，55–68（2000年） 13. Hu, J., Shen, L., Sun, G.：挤压-激励网络。载于：IEEE计算机视觉与模式识别会议论文集。第7132–7141页（2018年）

14. Huang, Z., Liu, J., Fan, X., Liu, R., Zhong, W., Luo, Z.：ReCoNet：用于快速高效多模态图像融合的循环校正网络。载于：欧洲计算机视觉会议。第539–555页。施普林格（2022年） 15. James, A.P., Dasarathy, B.V.：医学图像融合：最先进水平综述。信息融合**19**，4–19（2014年） 16. Kingma, D.P., Dhariwal, P.：Glow：采用可逆1×1卷积的生成流。神经信息处理系统进展**31**（2018年） 17. Li, J., Huo, H., Li, C., Wang, R., Feng, Q.：Attentionfgan：使用基于注意力的生成对抗网络进行红外与可见光图像融合。IEEE多媒体汇刊**23**，1383–1396（2020年） 18. Liang, J., Niu, L., Guo, F., Long, T., Zhang, L.：通过自校准定位和背景细化去除可见水印。载于：第29届ACM国际多媒体会议论文集。第4426–4434页（2021年） 19. Liang, P., Jiang, J., Liu, X., Ma, J.：从分解到融合：一种用于图像融合的自监督分解方法。载于：欧洲计算机视觉会议。第719–735页。施普林格（2022年） 20. Liu, J., Liu, Z., Wu, G., Ma, L., Liu, R., Zhong, W., Luo, Z., Fan, X.：用于图像融合与分割的多交互特征学习及全时多模态基准。载于：IEEE/CVF国际计算机视觉会议论文集。第8115–8124页（2023） 21. Liu, Y., Zhu, Z., Bai, X.：WDNET：用于可见水印去除的水印分解网络。载于：IEEE/CVF冬季计算机视觉应用会议论文集。第3685–3693页（2021年） 22. Liu, Y., Liu, S., Wang, Z.：一种基于多尺度变换和稀疏表示的图像融合通用框架。信息融合**24**，147–164（2015年） 23. Ma, J., Ma, Y., Li, C.：红外与可见光图像融合方法及应用：综述。信息融合**45**，153–178（2019年） 24. Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., Ma, Y.：SWINFUSION：通过Swin Transformer实现通用图像融合的跨域长程学习。IEEE/CAA自动化学报**9(7)**，1200–1217（2022年） 25. Maiti, A., Elberink, S.O., Vosselman, G.：Transfusion：用于语义分割的多模态融合网络。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第6537–6547页（2023） 26. Mo, Z., Chen, J., Fang, C., Li, G.：PATCHWIPER：利用动态分块参数去除真实场景中的可见水印。载于：第33届ACM国际多媒体会议论文集。第3537–3545页（2025年） 27. Nam, S., Brubaker, M.A., Brown, M.S.：用于多图像融合与层分离的神经图像表示。载于：欧洲计算机视觉会议。第216–232页。施普林格（2022年） 28. Niu, L., Zhao, X., Zhang, B., Zhang, L.：细粒度可见水印去除。载于：IEEE/CVF国际计算机视觉会议论文集。第12770–12779页（2023） 29. Peng, S., Zhu, X., Deng, H., Deng, L.J., Lei, Z.：FUSIONMAMBA：利用状态空间模型实现高效遥感图像融合。IEEE地球科学与遥感汇刊**62**，1–16（2024年） 30. Tancik, M., Mildenhall, B., Ng, R.：Stegastamp：实体照片中的隐形超链接。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第2117–2126页（2020年） 31. Tang, L., Yuan, J., Ma, J.：高层视觉任务闭环中的图像融合：一种具备语义感知能力的实时红外与可见光图像融合网络。信息融合**82**，28–42（2022年）

32. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: ESRGAN：增强型超分辨率生成对抗网络。载于：欧洲计算机视觉会议（ECCV）研讨会论文集。第 0–0 页（2018年）33. Wu, Z., Liu, Z., Lin, J., Lin, Y., Han, S.: 具有长短程注意力机制的 Lite Transformer。arXiv 预印本 arXiv:2004.11886（2020年）34. Xiao, M., Zheng, S., Liu, C., Lin, Z., Liu, T.Y.: 可逆重缩放网络及其扩展。国际计算机视觉杂志 **131**(1), 134–159 （ 2023 ） 35. Xie, X., Cui, Y., Tan, T., Zheng, X., Yu, Z.: FUSIONMAMBA：利用 Mamba 进行多模态图像融合的动态特征增强。视觉智能 **2**(1), 37（2024）36. Xu, H., Ma, J., Jiang, J., Guo, X., Ling, H.: U2FUSION：统一的无监督图像融合网络。IEEE模式分析与机器智能汇刊 **44**(1), 502–518（2020年）37. Xu, H., Ma, J., Le, Z., Jiang, J., Guo, X.: FusionDN：用于图像融合的统一密集连接网络。载于：AAAI人工智能会议论文集。第 34 卷，第 12484–12491 页（2020年）38. Xu, S., Zhang, J., Zhao, Z., Sun, K., Liu, J., Zhang, C.: 用于全色锐化的深度梯度投影网络。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第 1366–1375 页（2021年）39. Xu, Y., Mou, C., Hu, Y., Xie, J., Zhang, J.: 鲁棒可逆图像隐写。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第 7875–7884 页（2022年）40. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer：用于高分辨率图像复原的高效 Transformer。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第 5728–5739 页（2022年）41. Zhang, H., Ma, J.: SDNet：用于实时图像融合的通用挤压与分解网络。国际计算机视觉杂志 **129**(10), 2761–2785（2021）42. Zhang, H., Xu, H., Xiao, Y., Guo, X., Ma, J.: 重新思考图像融合：一种基于梯度与强度比例保持的快速统一图像融合网络。载于：AAAI人工智能会议论文集。第 34 卷，第 12797–12804 页（2020年）43. Zhang, H., Zuo, X., Zhou, H., Lu, T., Ma, J.: 一种基于视觉语义一致性的三维多模态医学图像融合鲁棒互增强框架。载于：AAAI人工智能会议论文集。第 38 卷，第 7087–7095 页（2024）44. Zhang, J., Jiang, X., Wang, H., Yuan, L., Luo, X., Ma, B., Wang, J.: DVW：扩散可见水印。载于：第 33 届 ACM国际多媒体会议论文集。第 8769–8777 页（2025年）45. Zhang, J., Chen, D., Liao, J., Zhang, W., Feng, H., Hua, G., Yu, N.: 通过深度水印实现深度模型知识产权保护。IEEE模式分析与机器智能汇刊**44**(8), 4005–4020（2021）46. Zhang, J., Jiao, L., Ma, W., Liu, F., Liu, X., Li, L., Chen, P., Yang, S.: 基于 Transformer 的多模态图像融合条件 GAN。IEEE多媒体汇刊 **25**, 8988–9001（2023）47. Zhang, T., Tan, T., Han, L., Wang, X., Gao, Y., van Dijk, J., Portaluri, A., Gonzalez-Huete, A., D'Angelo, A., Lu, C., 等：Important-Net：用于缺失数据合成的集成式 MRI 多参数增量融合生成器与注意力网络。信息融合 **108**, 102381（2024）48. Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., Zhang, L.: IFCNN：一种基于卷积神经网络的通用图像融合框架。信息融合 **54**, 99–118（2020年）49. Zhao, Z., Bai, H., Zhang, J., Zhang, Y., Xu, S., Lin, Z., Timofte, R., Van Gool, L.: CDDFuse：面向多模态图像融合的相关性驱动双分支特征分解。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第 5906–5916 页（2023）

18 Jie Song[1], Jun Jia[2†], Wei Sun[3], WangqiuZhou[4], Tao Tan[1†], GuangtaoZhai[2†]

50. Zhou, T., Cheng, Q., Lu, H., Li, Q., Zhang, X., Qiu, S.: 医学图像融合的深度学习方法：综述。计算机生物学与医学**160**, 106959（2023）

补充材料

6 补充实验

表 3。具有版权水印保护的融合图像的定量结果。

方法	MRI-CT			MRI-PET			MRI-SPECT		
	视觉信息保真度 $Q^{AB/F}$		结构相似性指数	视觉信息保真度 $Q^{AB/F}$		结构相似性指数	视觉信息保真度 $Q^{AB/F}$		结构相似性指数
OURS(CP)	0.09	0.12	0.19	0.12	0.13	0.15	0.09	0.11	0.16

表4。 三种不同水印去除攻击模型下的定量比较结果。所有指标的值越高，表示水印去除程度越低，意味着保护性能越好（**粗体**表示最佳结果）。

方法	MRI-CT			MRI-PET			MRI-SPECT		
	视觉信息保真度 $Q^{AB/F}$		结构相似性指数	视觉信息保真度 $Q^{AB/F}$		结构相似性指数	视觉信息保真度 $Q^{AB/F}$		结构相似性指数
PATCHWIPER	0.58	0.82	0.83	0.54	0.79	0.83	0.50	0.75	0.77
WDNET	0.41	0.63	0.69	0.47	0.71	0.75	0.57	0.79	0.80
SLBR	**0.82**	**0.89**	**0.94**	**0.74**	**0.83**	**0.90**	**0.86**	**0.90**	**0.94**

6.1 受版权保护的融合图像分析

表3展示了采用版权水印保护的融合图像的定量结果，评估指标为**视觉信息保真度、$Q^{AB/F}$**和**结构相似性指数**。由于版权水印嵌入在未经授权的输出中，这些输出与正常融合图像的相似度会降低。因此，指标值越低，表明版权水印对未经授权的输出影响越明显。此外，图7展示了三种不同模态组合下受版权保护结果与授权结果的定性比较。可以清楚地看到，嵌入水印后，受版权保护的融合图像得到了有效保护，实现了有效的版权追溯，并且无法直接使用，因为大量有用的视觉信息被遮挡。相比之下，使用密钥恢复的授权融合图像呈现出更清晰的纹理细节和更锐利的结构边界，证明了所提出的授权机制是有效的。

6.2 水印去除的定量分析

我们采用**视觉信息保真度、$Q^{AB/F}$和结构相似性指数** 作为定量评估指标。为了衡量不同水印移除攻击对版权保护结果造成的损害程度，我们将受攻击图像与未受攻击的版权保护融合图像进行比较，从而评估受攻击图像的退化程度以及水印信息的保留情况。指标值越高，表明受攻击结果越接近原始版权保护融合图像，对应的水印去除程度越低。因此，较高的数值意味着嵌入的版权水印对相应攻击具有更强的鲁棒性。如表4所示，在三种模态组合中，SLBR的总体指标值最高，表明这种攻击对嵌入的版权水印造成的损害最小，且我们的方法能够在SLBR攻击下更完整地保留水印信息。相比之下，在MRI-CT和MRI-PET模态组合中，WDNET的指标值最低，表明其水印去除效果更明显、攻击能力更强。对于MRI-SPECT模态组合，PATCHWIPER的指标值最低，说明在这种情况下，它对版权水印造成的损害更为严重。总体而言，AMIF的版权水印保护展现出一定程度的鲁棒性，但不同攻击模型在不同模态组合下表现出的破坏效果存在明显差异。

6.3 C-SAMIC的定量消融分析

我们在WDNET水印攻击模型下对C-SAMIC进行了消融研究，并采用**视觉信息保真度、$Q^{AB/F}$和结构相似性指数**作为三项定量评估指标。如表5所示，在所有三种模态组合中，配备C-SAMIC的AMIF均优于配备C-SAMIC Base的AMIF。其中，MRI-CT上的提升最为显著，**视觉信息保真度、$Q^{AB/F}$和结构相似性指数**分别从0.31、0.52和0.60提高至0.41、0.63和0.69。在MRI-PET上，这三项指标分别从0.46、0.70和0.74提高至0.47、0.71和0.75。在MRI-SPECT上，这些指标进一步分别从0.56、0.78和0.79提高至0.57、0.79和0.80。结果表明，完整的C-SAMIC设计有效提升了版权水印的保留能力。

6.4 C-SAMIC 的补充定性消融分析

对于 C-SAMIC 的定性消融结果，我们进一步在 MRI-PET 和 MRI-SPECT 测试集上提供了具有代表性的示例，如图8和图9所示。与 C-SAMIC Base 相比，C-SAMIC 在 WDNET 攻击后保留了更完整的水印信息，表明 C-SAMIC 能够更有效地将水印嵌入受保护特征中。

表5。 在 WDNET水印攻击模型下，分别配备 C-SAMIC Base 和 C-SAMIC 的 AMIF 消融结果（**粗体** 表示最佳，W/ 表示配备）。

方法	MRI-CT			MRI-PET			MRI-SPECT		
	VIF	$Q^{AB/F}$	结构相似性指数	VIF	$Q^{AB/F}$	结构相似性指数	VIF	$Q^{AB/F}$	结构相似性指数
WDNET(AMIF 搭配 C-SAMICBase)	0.31	0.52	0.6	0.46	0.7	0.74	0.56	0.78	0.79
WDNET(AMIF 搭配 C-SAMIC)	**0.41**	**0.63**	**0.69**	**0.47**	**0.71**	**0.75**	**0.57**	**0.79**	**0.80**

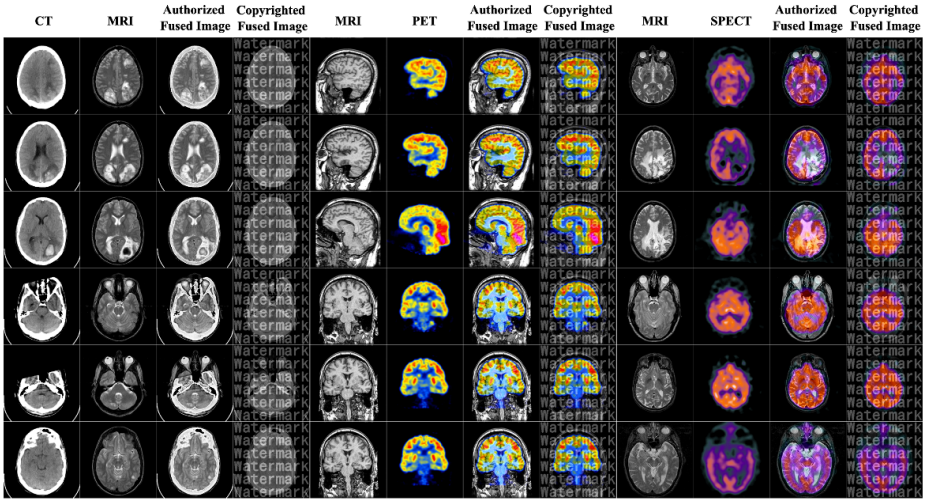

图7。 未授权模式和授权模式下三种模态组合（CT-MRI、MRI-PET 和 MRI-SPECT）的定性融合结果。未授权模式生成带有版权水印的融合图像，而授权模式生成使用密钥恢复的融合图像。

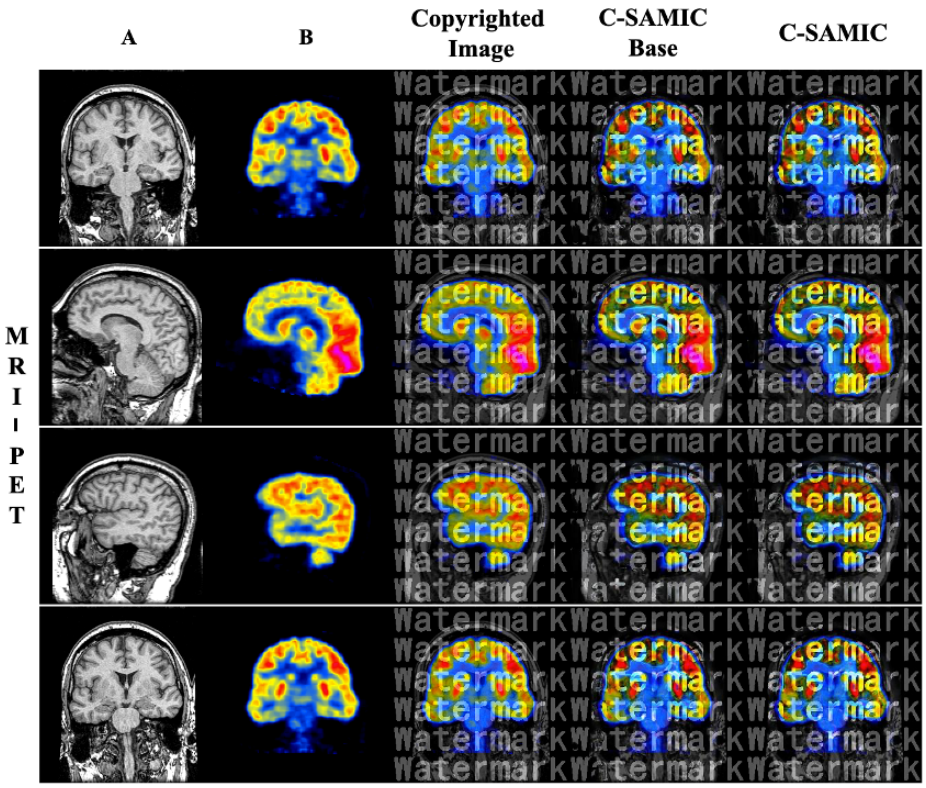

图 8。 C-SAMIC 和 C-SAMIC Base 在 MRI-PET 测试集上的消融研究可视化结果。展示了 WDNET 攻击下的若干组代表性图像对。

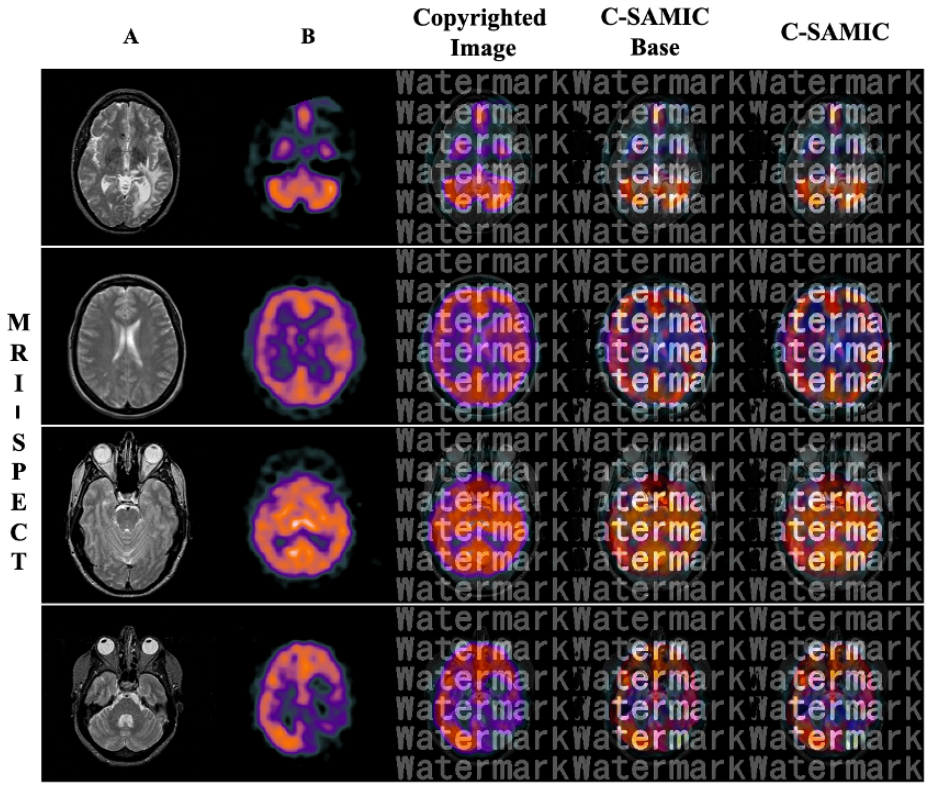

图 9. MRI-SPECT 测试集上 C-SAMIC 和 C-SAMIC Base 的消融研究可视化结果。图中展示了 WDNET 攻击下的几组代表性图像对。