VAAS：用于数字取证中图像篡改检测的视觉注意力异常评分

Opeyemi Bamigbade[1],Mark Scanlon[2],*, John Sheppard[1]

[1]取证与安全研究组，东南理工大学，爱尔兰沃特福德[2]取证与安全研究组，计算机科学学院，都柏林大学，爱尔兰都柏林

arXiv:2512.15512v1 [cs.CV] 17 Dec 2025

摘要

人工智能驱动的图像生成技术近期取得进展，为取证调查中数字证据真实性的核验带来了新的挑战。现代生成模型能够生成视觉上连贯的伪造图像，从而逃避基于像素或压缩伪影的传统检测器。大多数现有方法也缺乏对异常强度的明确度量，因此难以量化篡改的严重程度。本文提出视觉注意力异常评分（VAAS），这是一种新颖的双模块框架，将基于视觉变换器（ViT）的全局注意力异常估计与源自SegFormer嵌入的图像块级自一致性评分相结合。该混合方法可提供连续且可解释的异常分数，同时反映篡改的位置和程度。在*DF2023* 和*CASIA v2.0*数据集上的评估表明，VAAS取得了具有竞争力的F1值和交并比性能，同时借助注意力引导的异常图提升了视觉可解释性。该框架将定量检测与人类可理解的推理相结合，有助于实现透明、可靠的图像完整性评估。所有实验的源代码及用于复现结果的相关材料均已开源。

关键词：数字取证 、图像篡改检测、篡改定位、可解释人工智能、视觉 Transformer 、SegFormer机制、注意异常评分。

1. 引言

数字图像作为可采证据的可靠性，已成为现代取证调查中的关键问题。多媒体取证技术在核实法律、新闻和情报场景中所呈现视觉内容的Spichiger 和 Adelstein, 2025真实性方面发挥着核心作用。在这些场景中，首要目标之一是检测和定位图像篡改，确保数字证据在法庭上仍然可信。在2024年数字取证从业者调查（Hargreaves 等人，2024）中， ~20% 的从业者表示，他们在调查中“偶尔”或“经常”遇到深度伪造。

生成模型（如生成对抗网络和扩散网络）的最新进展显著增加了检测篡改图像或伪造图像的难度，因为这些模型能够生成语义连贯且无伪影的编辑结果，从而规避传统的取证线索。传统的图像篡改技术大致分为三类：（1）拼接，即将一幅图像中的内容复制到另一幅图像中；（2）复制移动篡改，即在同一幅图像中复制某个区域；以及（3）图像修复，即使用合成或重建的内容填补缺失区域（Wu 等，2019；Verdoliva，2020；Ma 等，2023）。这些篡改的示例见图1。早期的取证方法依赖于识别像素级不一致、压缩伪影或元数据差异。然而，在采用扩散模型或对抗合成流程的人工智能增强篡改中，这些线索往往会被移除或隐藏。

随着生成式人工智能工具日益普及，图像如今可以经过修改，呈现出极其逼真的效果和语义连贯性（Ma 等人，2023）。这些模型会调整纹理、光照和物体边界等细粒度特征，留下的取证痕迹极少。因此，仅关注低级特征的传统取证算法难以捕捉定义视觉真实性的空间与上下文关系。应对这一挑战，需要能够学习高级空间依赖关系，并检测注意力与特征表示中细微不一致之处的方法。

在此背景下，注意力机制和视觉变换器（视觉变换器）为取证分析提供了有前景的基础。它们能够捕捉全局上下文并建模长距离依赖关系，因此适合识别真实区域与篡改区域之间的不一致。基于这一原理，本文提出了一种混合框架，将全局注意力分析与局部篡改分割相结合，以实现可解释异常检测。

1.1. 本文的贡献

所提出的VAAS框架引入了一种新颖的混合异常评分公式，将全局注意力偏

*通讯作者电子邮箱：opeyemi.bamigbade@postgrad.wit.ie (Opeyemi Bamigbade), mark.scanlon@ucd.ie (Mark Scanlon), john.sheppard@setu.ie (John Sheppard)

差与局部自一致性相结合，为图像篡改提供了一种统一且可解释的度量，而现有文献尚未对此进行探讨。主要贡献总结如下：

- **双模块取证架构：**统一设计融合了基于视觉Transformer的取证注意力提取器 *(Fx)*，用于全局异常表征；以及基于 SegFormer 的篡改分割器 *(Px)*，用于局部篡改分析。
- **混合异常评分机制：**一种有原则的融合策略，将全局注意力偏差与局部空间不一致性结合为单一混合分数，以量化图像完整性。
- **可解释的取证可解释性：**生成注意力驱动的异常图，提供直观且可解释的证据，从而提升取证决策的透明度与可信度。
- **实证验证与基准评估：**在两个数据集上进行全面评估，结果表明，该方法在图像篡改检测方面表现出竞争力，定位稳定，并且能够在传统篡改类型（*CASIAv2.0*）和生成式篡改类型（*DF2023*）上实现强泛化能力。

所有实验代码和可复现性资源均已开源，可通过以下地址获取：https://github.com/OBA-Research/VAAS-experiments。

2. 背景

数字图像的真实性始终是取证调查中的核心关注点，因为视觉证据必须可追溯、可复现且经得起质证。传统取证方法侧重于低层信号伪影，而现代篡改技术日益依赖生成模型，在改变全局语义的同时不留下明显的像素级痕迹。本节概述图像篡改检测的发展历程，并强调当代取证分析为何需要全局—局部推理。

2.1. 经典取证方法

早期取证技术依赖于分析编辑流程造成的像素级扰动。例如 DCT 块不一致性、CFA 伪影破坏和重采样痕迹（Monika 等人，2021；Mahdian 和 Saic，2008）。利用 SIFT/SURF 描述符进行复制移动检测（Pandey 等人，2014；Christlein 等人，2012）以及通过边界或纹理异常进行拼接检测（Bappy 等人，2019），能够提供适用于取证报告、且易于解释的结果。然而，这些手工设计的线索在图像经过重新压缩、滤波或由人工智能处理流程重新生成后会变得不可靠，从而限制了其鲁棒性。

2.2. 人工智能生成的篡改与现代挑战

生成对抗网络和扩散网络等生成模型能够在保持像素级连贯性的同时，改变光照、结构和场景语义（Ma 等人，2023）。此类篡改缺乏传统检测器所利用的伪影，因此往往能逃过基于信号的方法的检测，尤其是在使用大规模合成数据集进行训练时。这一转变催生了对新型取证方法的需求：这类方法应着眼于空间连贯性进行推理，而非仅仅依赖低层噪声模式。

2.3. 图像取证中的深度学习

基于CNN的取证模型通过直接从数据中学习判别性特征，提高了鲁棒性（Nguyen 等人，2022；Zhuang 等人，2021）。较新的基于Transformer的系统利用注意力机制捕捉长距离依赖关系和语义不一致，因此在检测生成式篡改方面表现出色（Hao 等人，2021；Wang 等人，2022）。尽管性能出色，大多数深度模型仍如同黑箱，无法提供可解释的连续篡改强度度量。这一不足促使人们探索混合方法，将全局上下文、局部一致性与可解释评分结合起来。

2.4. VAAS 的研究动机

本文提出的视觉注意力异常评分（VAAS）框架结合全局注意力分析与局部自一致性评分，建立在上述进展的基础之上。VAAS 引入了一种混合机制，用于量化图像遭到篡改的可能性，同时生成具有视觉可解释性的证据图。通过这种方式，该框架弥合了量化性能与定性透明度之间的差距，使计算检测与数字取证验证的实际需求相契合。

3. 相关工作

人工智能，尤其是计算机视觉的持续发展，推动了图像伪造检测取得长足进步。传统的数字图像完整性验证方法依赖像素级异常检测、压缩伪影或元数据分析；面对现代生成模型生成的复杂篡改，这些方法正日益显得力不从心。本节回顾图像篡改检测、真实性验证等主要研究方向，并探讨注意力机制和交叉注意力机制在取证图像分析中的新兴作用。

3.1. 图像篡改检测

人工智能驱动的图像篡改检测已成为新闻业、执法和国家安全领域的重要方向，因为视觉真实性至关重要（Singh，2022）。生成对抗网络和扩散模型的出现，使无缝且照片级逼真的伪造图像成为可能（Guarneraet al.，2024）。早期的取证方法依赖手工设计线索，例如像素级不一致、彩色滤光阵列伪影或 JPEG 量化痕迹。然而，这些方法无法应对不含伪影的

由高级合成流水线生成的内容（Du 等人，2020）。

深度学习，尤其是卷积神经网络的兴起，标志着取证检测转向自动化表征学习。基于卷积神经网络的架构，例如广泛应用于深度伪造检测挑战赛和 FaceForensics++ 基准数据集（Nguyen 等人、2022；Zhuang 等人，2021），提升了跨操纵领域的泛化能力。Matern 等人（2019）利用了语义不一致性，例如缺失的反射或不对称面部线索，而Li 等人（2018、2020）探究了由人脸变形和重采样产生的空间伪影。Feng 等人（2020）提出了两阶段三元组损失框架，用于学习真实与伪造人脸的判别性嵌入。同样，Kingra 等人（2024）提出了 SFormer，这是一种端到端的 Transformer 架构，融合空间与时间信息，以提升深度伪造检测的准确率。尽管基于卷积神经网络的系统表现有效，但其主要捕捉局部纹理信息，缺乏建模长程空间依赖关系的能力。这是检测复杂篡改时的一个关键因素。

3.2. 基于注意力与 *Transformer* 的取证 模型

注意力机制与 Transformer 架构的结合，近来重塑了图像取证，使模型能够对空间与语义上下文进行推理。注意力机制会有选择地突出信息丰富区域，有助于识别光照、纹理和结构中的细微异常。视觉变换器（ViTs）通过自注意力对图像块之间的全局关系进行建模，因此尤其适合分析篡改图像中的语义连贯性（Ma 等人，2023）。

Hao 等人 (2021) 提出了 TransForensics，这是一种密集型 Transformer 编码器模型，可捕捉多尺度图像块依赖关系，以实现伪造定位。 Wang 等人 (2022) 提出了 ObjectFormer，该模型采用注意力引导的特征融合来进行像素级篡改检测。 Xia 等人(2024) 进一步引入层次化注意力，以结合卷积特征和 Transformer 特征；而 Shi 等人 (2024) 则结合双重注意力与边缘监督，以提高边界精度。这些研究凸显了注意力机制在学习语义丰富且可解释表征方面的有效性。

3.3. 用于上下文一致性的交叉注意力 *y*

交叉注意力机制通过允许不同表征空间之间进行特征交互来扩展自注意力，例如全局与局部特征，或真实与篡改嵌入。这种设计使模型能够比较并对齐反映真实图像语义一致性的异质线索，从而增强上下文推理。 Chen 等人 (2021) 提出了 *CrossViT*，这是一种通过交叉注意力整合多尺度图像块标记的 Transformer 模型，能够有效结合细粒度表征与全局表征。最近，Munawar 和 Oussalah(2025) 提出了一种用于图像伪造检测与定位的可解释双流注意力网络，并采用对比学习来增强真实区域与篡改区域之间的特征判别能力。

这些进展表明，交叉注意力和双重注意力框架能够弥合不同特征领域之间的互补信息，从而推动了 VAAS 架构所采用的全局—局部交互策略。

3.4. 数据集与取证基准测试

基准数据集对图像取证的发展至关重要（Thomson 等人，2025）。CMFD 数据集（Christlein 等人，2012年）为几何失真条件下的复制-移动检测提供了早期基准。Dong 等人提出的 CA-SIA v2.0（Dong 等人（2013年）引入了拼接与合成篡改，推动了鲁棒的基于分割的取证网络的发展（Bappy 等人，2019；Zhou 等人，2018）。近期，*Fischinger* 和 *Boyer* 发布的*2023* 年数字取证图像伪造检测数据集（*DF2023*）Fischinger 和 Boyer（2023年b）通过纳入来自扩散和风格迁移流程的人工智能生成篡改与混合篡改，拓展了这一范式，从而支持带有细粒度标注的大规模基准测试。

图像篡改检测领域取证数据集的发展，体现了其朝着更高多样性、真实性和语义丰富度转变的趋势。DF2023 等现代基准数据集融合传统篡改操作与人工智能驱动的生成式篡改，正是这一转变的体现。通过在 CASIA 和 DF2023 上评估 VAAS，本研究顺应了推动可复现且具有代表性的取证测试这一趋势，确保性能指标能够反映经典和新兴的篡改场景。

3.5. 总结与研究空白

基于注意力的架构通过捕捉长程依赖关系并提升可解释性，显著推动了篡改检测的发展。然而，现有方法大多只关注二元真实性分类或定性定位，未能提供一致的篡改严重程度定量指标。本文提出的视觉注意力异常评分（VAAS）框架通过引入双模块架构，融合视觉变换器（ViT）的全局异常评分与 SegFormer 的局部篡改分割，从而提供定量且可解释的取证证据。

4. 方法论

本文提出的方法——用于图像完整性的视觉注意力异常评分（VAAS）——被构建为异常检测与定位框架，而非传统的分类或分割任务。该方法不单纯依赖类别标签或像素级掩码，而是着重学习自然图像内在的空间一致性，并识别由传统篡改或人工智能驱动的篡改所引入的偏差。

如图2所示，系统整合了两个互补模块：(1) 用于全局异常检测的全图注意力模块（*Fx*），以及 (2) 用于局部异常定位的图像块级自一致性模块（*Px*）。这两个模块共同量化空间不规则性，例如



图1：基于内容变化的图像篡改技术示例（复制-移动、拼接和移除）。摘自 Shi 等人（2024）

纹理不连续性、光照不一致和不自然的物体对齐。两个模块的输出通过混合加权评分机制（*HSM*）进行融合，将全局和局部异常线索整合为单一且易于解读的完整性评分。这一设计使系统能够平衡注意力驱动的全局一致性与图像块级空间异常，为数字取证调查人员提供篡改的定量证据和视觉证据。

4.1. 全图一致性模块（*Fx*）

Fx 模块采用视觉Transformer（ViT）从整幅图像中提取深层表征，并评估其全局空间连贯性。与基于图像块的方法不同，该模块以整体方式处理图像，并根据注意力图的强度和空间分布得出全局异常评分。这些注意力图突出显示模型在推理过程中最关注的区域，为识别结构异常提供可解释的线索。

为确保结果可复现，*Fx* 模块使用从 ViT-Base 最后四个编码器层提取的注意力图，并对所有注意力头的结果取平均。首先将图像调整为 224×224，并使用 ImageNet 统计数据进行归一化。评分前，使用双线性插值将注意力张量下采样至原始图像分辨率。所有实验均使用固定的随机种子运行，以确保训练和评估过程的一致性。

该过程包括：

1. **特征提取：** 输入图像经 ViT 处理后，会生成多层注意力图，反映图像块之间的上下文关系。

2. **注意力分析：** 在真实图像中，注意力分布通常平滑且在语义上保持一致。篡改区域会因特征交互不一致而引入不规则的注意力激活。

3. **统计评分：** 将注意力激活的均值（μ）和标准差（σ）与根据真实样本计算得到的参考分布进行比较。

全局异常评分 S_F 表示为：

$$S_F = \frac{|\mu - \mu_{\text{ref}}|}{\sigma_{\text{ref}}} \tag{1}$$

其中 μ_{ref} 和 σ_{ref} 表示从真实注意力分布中得出的参考统计量。较高的 S_F 表明其与自然空间连贯性之间的偏差更大，这对应着更高的篡改可能性。为评估 Transformer 表征深度对异常估计的影响，我们还考察了 `vit-large-patch16-224`、`swin-base-patch4-window7-224`和`dinov2-base`等其他骨干网络。本分析详见第5.1节，旨在评估全局注意力粒度是否会影响取证异常评分 S_F。

4.2. 基于图像块的自一致性模块 *(Px)*

Px 模块通过分析不重叠图像块之间的自一致性，执行局部异常检测。输入图像 I 被划分为大小为$k \times k$的图像块 $\mathbb{P}_i$。每个图像块都使用基于 SegFormer 的编码器进行嵌入，以获得其潜在表征 $F(\mathbb{P}_i)$。

通过测量每个图像块与其空间邻居之间的余弦相似度来量化局部不规则性：

$$Sim(P_i, P_j) = \frac{F(P_i) \cdot F(P_j)}{\|F(P_i)\| \|F(P_j)\|} \tag{2}$$

相似度较低表示存在上下文偏差，每个图像块的异常分数定义为：

$$S_P(i) = 1 - \frac{1}{N} \sum_{j \in N(i)} Sim(P_i, P_j) \tag{3}$$

其中，$\mathrm{N}(i)$ 是图像块 i 的邻域，N 是其大小。最终的图像块级异常分数汇总如下：

$$S_P = \frac{1}{M} \sum_{i=1}^{M} S_P(i) \tag{4}$$

其中，M 是图像块的数量。该表示编码了局部纹理和结构与其空间上下文的契合程度。

实际应用中，SegFormer 编码器输出 256 维的图像块嵌入。提取图像块时采用固定大小 32×32，并使用不重叠的步长。所有真实掩码均使用最近邻插值调整为 224×224，以保留边界清晰度。与全局分数融合前，先将图像块级异常分数上采样至全分辨率。

4.3. 混合异常评分机制（*HSM*）

最终异常分数整合了分别从*Fx* 和*Px* 模块中获取的篡改全局与局部证据。所提出的框架不依赖单一线索，而是采用加权融合策略，在全局上下文异常与局部空间不一致性之间实现平衡。统一分数记为 S_H，其定义如下：

$$S_H = \alpha\, S_F + (1 - \alpha)\, S_P \tag{5}$$

其中，S_F 是来自视觉Transformer的基于全局注意力的异常分数，S_P 是来自

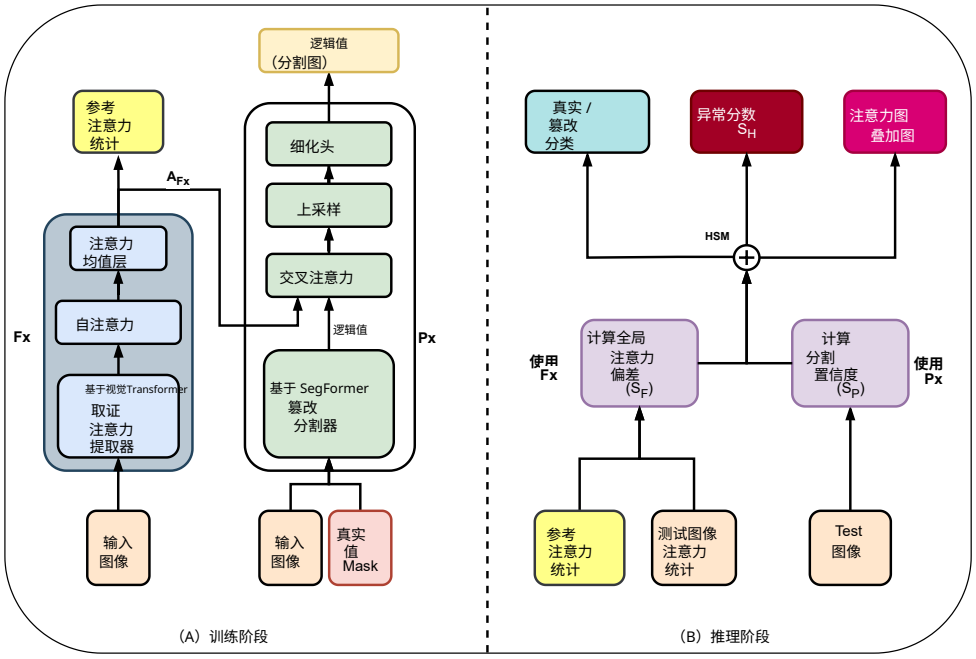


图 2：VAAS框架概览，展示了数据流经训练阶段（左）和推理阶段（右）的过程。该架构整合了全局注意力分析与局部一致性估计，以检测表明图像篡改的空间不一致性。

SegFormer 的局部图像块级一致性分数，而 $\alpha \in [0,1]$ 控制全局线索与局部线索的相对影响。

较高的 α 会更加重视全局注意力偏差，适用于会改变语义结构或光照一致性的篡改；较低的 α 则优先考虑由Px 模块捕捉的局部纹理和边界不规则性。这种自适应公式使得 α 可以根据数据集特征或模型部署的取证场景，通过经验确定或交叉验证得出。验证分析或阈值扫描还可进一步优化 α，确保该混合指标能够有效适应多种篡改类型。

这种加权融合提供了连续且可解释的图像完整性度量。所得分数不仅反映了图像遭到篡改的可能性，还保留了对细微全局异常和精细局部不一致性的敏感度，因此尤其适用于取证评估；在这类场景中，可解释性和均衡整合证据至关重要。

除加权形式外，我们还评估了融合函数的一种调和变体，以探究惩罚两个模块（Fx 和Px）之间的不一致是否能增强检测鲁棒性。这两种融合策略的性能对比见第5.2节。调和变体可表示为：

$$S_H^{harmonic} = \frac{2}{\frac{1}{S_F} + \frac{1}{S_P}} \tag{6}$$

4.4. 损失函数与训练策略

Px 模块采用监督方式，利用现有的篡改掩码进行训练，而Fx 模块则提供注意力引导，以增强空间连贯性。复合分割损失结合了二元交叉熵（BCE）、Dice 和焦点损失成分，以平衡像素级准确率与区域重叠：

$$\mathcal{L}_{seg} = \lambda_{bce}\mathcal{L}_{\text{BCE}} + \lambda_{dice}\mathcal{L}_{\text{Dice}} + \lambda_{focal}\mathcal{L}_{\text{Focal}} \tag{7}$$

为使Px的局部特征进一步与Fx学习到的全局上下文对齐，引入注意力对齐项：

$$\mathcal{L}_{fx} = 1 - \cos(F_{Px}, F_{Fx}) \tag{8}$$

该对齐项的影响由正则化系数 ω_{fx} 缩放，从而得到总损失：

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \omega_{fx}\mathcal{L}_{fx} \tag{9}$$

其中，L_{BCE} 表示二元交叉熵损失，L_{Dice} 表示 Dice 重叠损失，L_{Focal} 表示用于处理类别不平衡的焦点损失。λ_{bce}、λ_{dice}和 λ_{focal}是标量权重因子，用于控制各项在分割损失 L_{seg}中的贡献。ω_{fx}是正则化系数，用于平衡注意力对齐项相对于分割目标的影响，从而得到总优化损失 L_{total}。

$\omega_{fx} = 0$。1 的适中取值在各数据集上实现了特征稳定性与掩码清晰度之间的最佳平衡。Fx 模块本身在训练期间保持冻结，仅提供预训练注意力图，以指导Px特征细化。

VAAS 的性能通过多种检测和定位指标进行评估。所选指标如下：(1)

F1分数：平衡篡改检测中的精确率和召回率；(2) 交并比（IoU）：评估预测掩码与真实标注掩码之间的重叠程度；(3) 精确率和召回率：体现误报与漏检之间的权衡。

4.5. 数据集概述与实现

VAAS框架在两个成熟的取证数据集上进行了评估：*CASIA v2.0* 和*DF2023*，这两个数据集被选来代表传统篡改和人工智能驱动的篡改场景。

CASIAv2.0(Dong 等人，2013年) 是评估拼接和合成检测方法的基础基准数据集。该数据集包含约*12,614* 张图像，其中包括 *7,491* 张真实样本 和 *5,123* 张篡改样本，每个样本都有对应的真实掩码。该数据集涵盖多种真实场景和篡改风格，包括不同的后处理方式。所有图像均调整为 *224×224* ，并归一化至 [0,1]。 *Px* 模块使用篡改掩码进行训练，而*Fx* 模块则根据真实样本计算参考注意力统计量（μ_{ref}， σ_{ref}）。

DF2023(Fischinger 和 Boyer，2023年b) 旨在构建一个基准数据集，用于评估取证方法对人工智能生成和混合篡改的检测能力。完整数据集包含约一百万张伪造图像，分为四种篡改类型：*10万* 张移除篡改，*20万* 张增强篡改，*30万* 张复制-移动篡改，以及*40万* 张拼接篡改。关于*DF2023*的首项研究由Fischinger 和 Boyer(2023年a) 开展；由于该数据集计算开销巨大，他们只能使用其中一部分，且报告结果主要聚焦于其他相关篡改数据集，而非 *DF2023* 本身。本研究使用 DF2023 中按分层方式抽取的 10% 子集（约 10 万张图像），以确保训练在计算上可行。该子集按照各篡改类别的比例进行抽样（移除类 10 万张、增强类 20 万张、复制-移动类 30 万张、拼接类 40 万张），从而保留原始分布。所有实验均采用 80/20 的训练集—验证集划分；用于抽样的确切文件索引已包含在公开代码仓库中，以确保完全可复现。验证集作为所有报告指标的测试集，评估期间未使用该划分之外的任何 DF2023 样本。

该数据集的多样性和生成式内容，使其尤其适合评估 VAAS 在语义连贯但空间不一致的篡改下的混合异常评分表现。

实验使用 PyTorch 和 Hugging Face Transformers 库进行。 *Fx* 模块采用`google/vit-base-patch16-224-in21k`，而 *Px* 采用了`nvidia/segformer-b1-finetuned-ade-512-512`。两个模型均使用 Adam 优化器进行训练（学习率1×10^{-4}，以及 *8* 的批量大小）。复合损失由二元交叉熵、 Dice 和焦点损失组成，并由系数 $\lambda_{dice} = 0。7$和 $\lambda_{focal} = 1。0$进行平衡，详情见第 4节。模型在 NVIDIA GeForce RTX 5090 图形处理器（34 GB 显存）上训练了 80 个训练轮次，并根据验证集上的 F1 分数选择检查点。评估指标包括 F1分数、交并比、精确率和召回率。

尽管 *Hotels-50K* (Stylianou 等人，2019) 并非图像篡改数据集，但在涉及大规模图像验证的实际取证场景中仍具有应用价值。近期利用该数据集开展的研究，涉及面向室内场景识别的颜色敏感嵌入设计（Bamigbade 等人，2025）和室内多媒体地理定位（Aftab 等人，2026年），凸显了人们对开发可解释视觉表征、以支持数字取证调查任务的广泛关注。尽管这些先前工作与篡改检测没有直接关系，但它们促使人们进一步探索 VAAS 等异常评分方法，如何在类似的真实场景工作流程中支持完整性评估。

总而言之，VAAS 的方法设计提供了一个均衡的框架，可用于评估传统篡改和 AI 生成篡改的可解释性与取证鲁棒性。借助混合异常评分机制和*Fx*引导训练，下一节将介绍对 VAAS 的实证评估。我们将报告两个数据集上的结果，重点展示定量性能指标（F1值、交并比）以及通过注意力引导可视化呈现的定性可解释性。这些分析表明，所提出的架构如何在不同篡改类型和尺度下实现泛化，同时保持对数字取证可靠性至关重要的可解释决策路径。

5. VAAS 组件分析

为更好地理解 VAAS 背后的设计选择，并论证主要实验所采用的配置，我们开展了一系列组件分析研究。每项实验均单独考察一个特定的架构组件或超参数组件，并评估其对检测准确率、定位精确率和可解释性的影响。以下四项分析共同为主要组件的贡献及其相互作用提供了实证见解：

1. ***Fx* 骨干网络变体：** 评估 Transformer 的深度和架构类型对全局异常估计的影响。2. **融合机制（加权与调和）：** 考察混合评分函数的不同形式及其在不同数据集上的稳定性。3. **Fx引导正则化的影响：** 探究训练期间基于注意力的引导强度如何影响 *Px* 特征对齐和分割保真度。4.**数据集特异性敏感度 α：** 分析控制 *Fx*–*Px* 贡献的加权因子如何适应不同的操纵领域。

这些研究共同阐明了 VAAS 的内部动态，凸显了注意力耦合与融合加权如何协同作用，从而在取证图像分析中实现均衡的性能与可解释性。

5.1. Fx 骨干网络变体

为考察 Transformer 容量和注意力粒度对全局异常估计的影响，我们将默认的 ViT-Base 骨干网络与 ViT-Large、

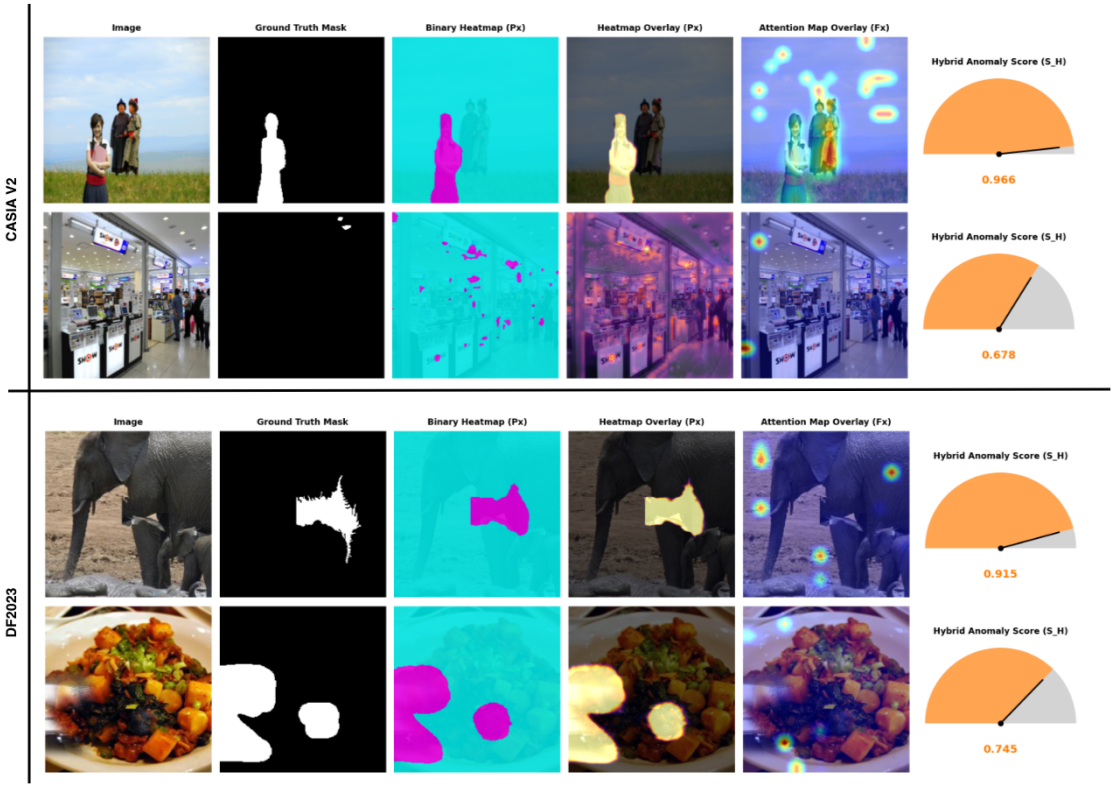


图3：VAAS 在*CASIA v2.0*（前两行）和*DF2023*（后两行）上的定性可视化。第（1）至（6）列依次展示：输入图像、真实掩码、二值*Px* 输出、*Px* 热图叠加图、*Fx* 注意力叠加图，以及最终混合异常分数。高异常样本（前几行）呈现出精准定位和显著的全局线索，而中等异常样本则显示出均衡但分散的注意力，体现了*Px* 与*Fx*之间的互补作用。

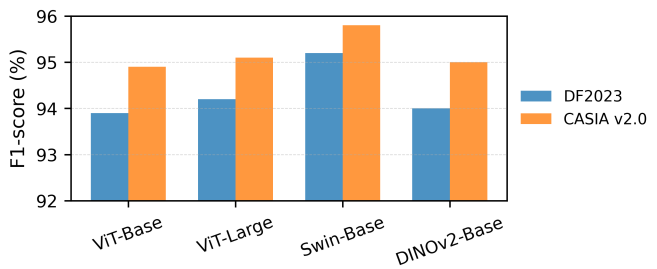


图4：*Fx* 骨干组件分析，展示了其在 DF2023 和 CASIA v2.0 上的 F1 分数。ViT-Base 在准确率和效率之间实现了最佳平衡。

Swin-Base和DINOv2-Base进行了比较。具有更深层或层次化注意力（如 Swin）的模型生成的异常热力图更平滑，但 F1值仅略有提升（+1。3% ，在*DF2023*上）。标准 ViT-Base 在可解释性与计算效率之间实现了最佳权衡，证实了其适用于 VAAS框架中的*Fx* 。图4展示了不同*Fx* 骨干网络在定量和定性方面的影响，说明 Transformer 深度和层次结构如何影响异常锐度与可解释性。

5.2. 融合机制：加权与调和

对线性加权融合（公式 5）与用于结合全局和局部异常分数的调和变体（公式 6）进行了比较。对加权和调和形式进行了考察，方法是将加权因子 α 在 [0.3,0.8]范围内进行扫描。如图 5所示，随着 α 增加，性能稳步提升，直至约0.6，此时全局注意力线索（*Fx*）与局部图像块一致性（*Px*）达到平衡。超过这一点后，过度依赖全局项会使性能提升趋于饱和。调和变体在不同数据集上表现得更加平滑、稳定，表明当*Fx* 和*Px* 提供冲突证据时，它具有更强的鲁棒性。这证实，调和融合不仅能稳定决策置信度，还能通过对全局和局部异常线索之间的一致性加权，提升可解释性。

5.3. Fx 引导正则化的影响

为评估*Fx*引导的注意力对齐项的影响，我们调整了正则化系数 ω_{fx}，取值为 $\omega_{fx} = 0、0, 0.05, 0.1, 0.2\}$。适中的取值（$\omega_{fx} = 0.1$）在局部边界精度和全局注意力一致性之间取得了最佳平衡。设为 $\omega_{fx} = 0$ 会导致分割掩码更加嘈杂，这表明*Fx* 线索有助于稳定*Px* 特征学习；而较高的权重（$\omega_{fx} > 0.2$）则会造成过度平滑并降低边缘保真度。

5.4. α 的数据集特异性敏感度

加权因子 α 用于控制*Fx* 和*Px* 分数的贡献，我们对其进行了评估，以识别特定数据集的趋势。对于*CASIAv2.0*，最佳 α 取值介于 0.4–0.6 之间，这反映出

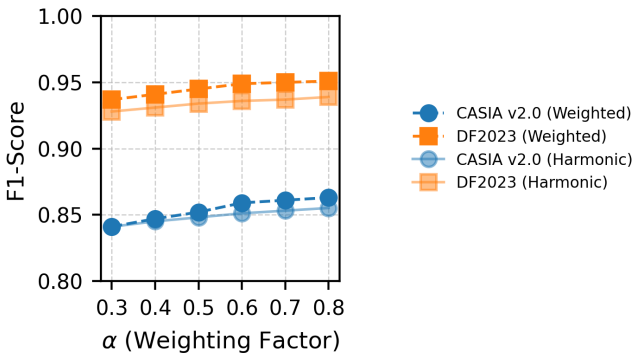


图5：异常评分融合组件分析，展示了不同 α 下的 F1 分数。调和融合的变化平稳，且在各数据集上保持稳定；加权融合则在 α=0.6 附近达到峰值。

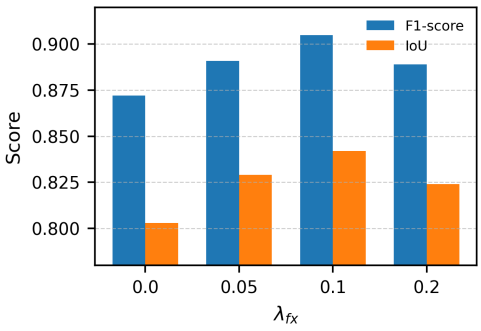


图6：Fx 引导的正则化权重 λ_{fx} 的影响。适中的取值（λ_{fx}=0.1）能在边界精度与注意力一致性之间取得最佳平衡。

对局部线索和全局线索的依赖较为均衡。对于*DF2023*，较高的取值（$\alpha \approx 0.7$）带来了更优结果，这是因为生成模型产生的篡改以语义连贯但全局不一致的情况为主。该分析凸显了 VAAS 在不同领域中的适应性，其中 α可作为一种可解释的控制参数，用于调节取证敏感度。

总之，组件级分析始终支持 VAAS 的设计理念。ViT-Base 骨干网络在全局异常估计的表征能力与可解释性之间实现了良好平衡，而加权融合策略则在各数据集上表现稳定。*Fx* 引导的正则化改善了特征对齐，并降低了 *Px* 输出中的噪声，凸显了训练期间交叉注意力耦合的价值。最后，α 对数据集的依赖行为表明，可根据不同的篡改特征调整异常敏感度，从而兼顾适应性与透明部署。

6. 结果

所提出的 VAAS 框架在两个互补的取证基准数据集上进行了评估：*DF2023* 数据集，代表大规模且多样化的图像篡改；以及*CASIA v2.0* 图像篡改检测评估数据库，代表经典的拼接和合成伪造。为确保基准比较具有上下文代表性，我们选取了变换器架构和卷积架构中具有代表性的先进方法，涵盖了现代取证检测策略的广泛范围。

对于*CASIA v2.0*，我们选取了四种已发表的方法进行比较：基于注意力的双流注意力网络（*DSCL-Net*），由 Munawar 和 Oussalah（2025）提出；视觉引导的 *VASLNet*（Yadav 和 Vishwakarma，2024）；混合卷积–Transformer 模型混合 *CNN–Transformer*（Sharma 等人，2025）；以及层次化细粒度定位框架*HiFi-IFDL*（Guo 等人，2023）。这些基线方法共同构成了一个均衡组合，涵盖经典策略和基于深度注意力的策略，用于篡改检测与定位。

对于*DF2023*，*DF-Net*（Fischinger 和 Boyer，2023年a）被用作该数据集的参考模型。尽管该数据集在 Fischinger 和 Boyer（2023年b）中首次介绍，但此前没有研究直接针对*DF2023*报告全面的定量指标。*DF-Net* 的研究仅评估了其在其他基准数据集上的跨数据集性能。本研究中，VAAS 首次在*DF2023*上建立了可复现的检测和定位指标，并使用完整语料库的*10%*（超过 *10* 万条）进行*80/20* 划分训练和验证，同时保留官方发布版本中所述的各类篡改比例。

表1 展示了定量比较结果。在*CASIA v2.0*上，VAAS 的性能与注意力模型和基于 CNN 的模型相当或更优，尤其是在定位（交并比）方面，这证实了其混合评分机制在不同篡改类型上的鲁棒性和泛化能力。在*DF2023*上，VAAS 首次为检测和定位提供了基准参考，为今后评估生成式和混合式篡改检测方法提供了支持。

图 3 展示了篡改图像、真实标注掩码与 VAAS 生成的注意力图之间的定性对比。*Fx* 模块（ViT）捕捉光照和场景连贯性方面的全局不一致性，而 *Px* 模块（SegFormer）则能分离融合边缘或图像修复接缝等细粒度伪影。二者加权融合后生成的异常图与篡改区域高度吻合，进一步增强了可解释性和取证可靠性。

6.1. 定量评估

如表1所示，VAAS 在*DF2023*上取得了 94.9% 的 F1分数和 91.1% 的交并比，优于数据集基线。这一提升体现了将全局注意力线索（S_F）与图像块级自一致性（S_P）相结合的有效性，使该框架能够捕捉生成式篡改中典型的语义和结构不一致。在*CASIA v2.0*上，VAAS 的 F1分数达到*94.1%*，交并比达到*89.0%*，优于大多数基于注意力的方法和混合 CNN–Transformer 方法。尽管*HiFi-IFDL*凭借层次化细化取得了略高的 F1 值，但其较低的交并比表明定位稳定性较差。相比之下，VAAS 在检测准确率和边界精度之间保持了更好的平衡，证实了其混合评分设计的优势。

表1：VAAS 与最先进的代表性方法在 DF2023 和 CASIA v2.0 上的定量比较。基线值直接取自所引用的出版物。“–”表示未报告的指标。VAAS 在检测（F1值）和定位（交并比）指标上均取得了相当或更优的性能。

数据集	方法	精确率（%）	召回率（%）	F1值（%）	交并比（%）
CASIA v2.0	DSCL-Net（双流注意力）（Munawar 和 Oussalah，2025）	–	–	92.7	–
	VASLNet（Yadav 和 Vishwakarma，2024）	–	–	91.9	85.1
	混合 CNN–Transformer（Sharma 等人，2025）	–	–	84.0	82.0
	HiFi-IFDL（Guo 等人，2023）	99.5	–	**97.4**	61.6
	VAAS	93.5	94.8	94.1	**89.0**
DF2023	DF-Net (Fischinger 和 Boyer，2023年a)	–	–	–	–
	VAAS	**95.9**	**94.2**	**94.9**	**91.1**

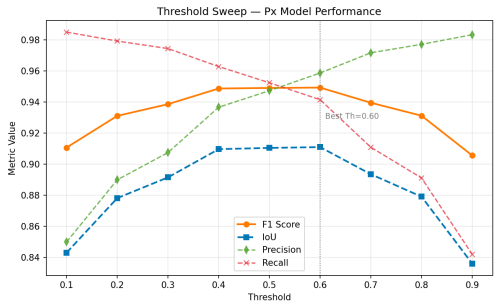


图 7 $\alpha S_F + (1-\alpha) S_P$ α S_H =：对（）中的混合评分权重进行阈值扫描分析，性能在 . 左右达到峰值。 $\alpha = 0.6$

为考察混合加权参数 α 对性能的影响，我们针对多个全局—局部比率进行了阈值扫描（图7）。F1 值和交并比均在α = 0.6左右达到峰值，表明适度强调全局注意力项（S_F）能够在检测与定位准确率之间实现最稳定的权衡。

6.2. 定性分析

图 3展示了所提出的 VAAS 框架在*CASIAv2.0* 和 *DF2023*上的定性表现。每幅图由四行六列组成：前两行对应 CASIA 样本，后两行对应*DF2023*。第 (1)–(6) 列依次显示输入图像、真实掩码、二值*Px* 输出、*Px* 热力图叠加图、*Fx*注意力叠加图，以及混合异常分数可视化。

每个数据集都展示了两个具有代表性的异常程度：高分和中等分数。在高异常案例中，*Px* 生成的分割掩码与真实值高度吻合，而*Fx* 则突出显示更广泛的上下文异常，例如光照变化或语义失衡。这种表现符合混合评分的设计意图：全局流（S_F）增强基于图像块的流（S_P）的精确率，从而生成空间一致的取证证据。

对于中等程度的异常，*Px* 仍能定位主要篡改区域；不过，边缘或纹理过渡处会出现轻微偏差。同时，*Fx* 继续检测细微的全局不一致性，其注意力偶尔会延伸至相邻区域或上下文相关区域；这一现象与全局语义推理相符。这些响应共同体现了 VAAS 如何整合互补线索：*Px* 提供空间保真度，而*Fx* 则增强上下文感知。

需要注意的是，*Fx* 注意力图并非旨在复现像素级篡改掩码。Vision Transformer 的注意力会突出显示因篡改而导致语义或结构连贯性受损的区域，这些区域可能超出确切的篡改边界。这种行为符合全局推理：模型会关注引入编辑后在上下文中变得不一致的对象或区域。这些全局线索与*Px*生成的精确空间定位相辅相成，共同构成对异常证据更完整的解释。

总体而言，图 3 中的定性模式与第 6节报告的定量趋势相呼应。较高的异常分数对应着置信度高、空间上连贯的篡改，而中等分数则反映出局部或模糊的编辑。异常强度与视觉可解释性之间的一致性，进一步彰显了 VAAS 作为透明、可靠取证工具的价值。

6.3. 可解释性与稳定性

除了定量准确性，VAAS 还注重透明且稳定的取证推理。*Fx* 注意力头生成具有语义意义的异常图，并与篡改区域相对应，使人工分析人员能够直观地验证模型推断。对于真实图像，注意力模式始终紧凑且连贯；而篡改样本则呈现碎片化或弥散激活，清晰地表明异常的存在。这种可解释性有助于实现可解释的决策，这也是确保取证与法律可靠性的基本要求。

该框架还展现出针对不同篡改尺度的自适应稳定性。在 *CASIA v2.0* 中，篡改通常局限于局部，*Px* 对异常响应的贡献占主导；而在包含全局语义修改的 *DF2023* 中，*Fx* 的贡献更为显著。自适应加权参数 α 会动态平衡这些贡献，使 VAAS 能够有效泛化到不同的数据集和篡改类型。

6.4. 研究结果总结

总之，VAAS 在两个基准测试中均展现出具有竞争力或更优的性能，同时保持了可解释性和稳定性。其混合评分机制兼顾了补丁级定位的细粒度敏感性与基于 Transformer 的注意力的上下文深度。通过整合这些互补线索，VAAS 提供了可解释、可复现且可扩展的取证推理；这些都是实现数字证据验证和跨领域部署的关键特性。

7. 讨论

在第6节的研究结果基础上，本节对所提出的 VAAS 框架的性能和架构行为进行解读。讨论重点在于全图注意力模块（*Fx*）、补丁级分割模块（*Px*）与混合评分机制（HSM）之间的交互如何共同提升检测准确率和可解释性。本节还阐述了这些结果对数字取证实践的影响；在该领域，可解释且可复现的证据生成是一项关键要求。

如第 6 节所示，定量结果表明，适度的全局—局部权重（$\alpha \approx 0.6$）能够在检测准确率和定位精度之间实现最佳平衡。这一发现与推理阶段的定性观察一致，其中 *Fx* 和 *Px* 模块在突出语义和结构不一致方面表现出互补性。

实验结果证实，VAAS 在保持篡改区域定位可解释性的同时，取得了出色的定量性能。混合评分机制有效融合了来自 *Fx* 的全局异常线索与来自*Px*的空间细节预测，使其能够检测出常被传统方法忽略的细微篡改或空间不一致篡改。生成的异常图提供了可视化解释，阐明了模型的决策边界；这对于需要推理过程可追溯的取证分析尤为重要。

基于 Transformer 的*Fx* 骨干网络提升了全局注意力表征的稳定性和表达能力。通过提供一致的上下文信号来引导分割，*Fx* 确保*Px* 的异常定位与实际篡改区域相吻合。与此同时，采用 SegFormer 实现的*Px* 模块通过图像块级监督细化定位结果。这些模块共同构成跨尺度交互，在检测灵敏度、定位精度和可解释性之间取得平衡；这三者是实现实际取证可靠性的关键支柱。

总体而言，这些发现表明，VAAS 不仅提升了定量性能，也为构建可解释且可扩展的取证人工智能系统提供了有原则的基础，促使我们进一步探究其局限性与未来扩展方向。

7.1. 局限性与未来工作

尽管 VAAS 展现出良好的可解释性和具有竞争力的准确率，但仍存在若干局限。首先，部分评估是在*DF2023* 的子集上进行的，可能无法充分体现实际取证图像中篡改的多样性与复杂性。另一个局限是，当前评估未纳入跨数据集测试，而这越来越被视为评估基于人工智能的取证系统对未见分布的鲁棒性所必需的措施。

尽管*Hotels-50K* 并非图像篡改数据集，也未用于本文的实验，但它仍然在实际应用中与数字取证相关，因为它用于现实世界中的酒店房间识别。纳入该数据集仅是为了说明 VAAS 未来可扩展至受控篡改基准之外。在这类领域中，目标从检测篡改转变为评估更广泛的视觉完整性和来源。

未来工作将把该框架扩展至更广泛的篡改类别，包括人工智能与人工共同参与的合成编辑，并将在未见数据集上评估跨领域泛化能力。将 VAAS 拓展至视频和多模态取证也是一个颇有前景的方向。在方法层面，将自适应阈值处理纳入调和评分机制，或可提升系统在不同内容领域中的鲁棒性。除检测指标外，未来研究还应量化可解释性本身，评估生成的异常图在帮助人工分析人员核实篡改证据方面的成效。

8. 结论

本文提出了 VAAS，这是一种用于数字取证中检测和定位图像篡改的视觉注意力异常评分框架。该方法通过混合调和评分机制，将全局注意力分析（*Fx*）与局部分割过程（*Px*）相结合，以量化图像完整性。在*CASIA v2.0*和 *DF2023* 上的评估表明，VAAS 能够实现较高的检测准确率和稳定的定位效果，同时生成具有可解释性的注意力热图，为取证决策提供支持。

研究结果证实，基于 Transformer 的注意力表征能够有效地对空间不一致性进行建模，且所提出的混合融合方法能够对真实性进行均衡且可解释的度量。通过将可解释性融入检测流程，VAAS 提升了取证准备度，并支持在调查和法律场景中对视觉证据进行透明验证。未来的工作将探索如何将 VAAS 应用于 Hotels-50K 等变化多样的真实世界数据集，届时研究重点将从篡改检测转向一般性的图像完整性与来源验证。

参考文献

Aftab, K., Adams, G., Scanlon, M., 2026年。《从插头到地点：基于电源插座的室内多媒体地理定位与数字调查》。Forensic Science International: Digital Inves- tigation 56S.

Bamigbade, O., Scanlon, M., Sheppard, J., 2025. 通过颜色特征改进室内场景地理定位中的图像嵌入。IEEE Access13, 79860–79870. doi:10.1109/ACCESS.2025.3564496.

Bappy, J.H., Simons, C., Nataraj, L., Manjunath, B.S., Roy-Chowdhury, A.K., 2019. 用于检测图像伪造的混合 LSTM 与编码器–解码器架构。IEEE图像处理汇刊 28, 3286–3300. doi:10.1109/TIP.2019.2895466。

Chen, C.F.R., Fan, Q., Panda, R., 2021. CrossViT：用于图像分类的跨注意力多尺度视觉Transformer，载于：IEEE/CVF国际计算机视觉会议（ICCV）论文集，第357–366 页。doi:10.1109/ICCV48922.2021.00042。

Christlein, V., Riess, C., Jordan, J., Riess, C., Angelopoulou, E., 2012年。《热门复制移动伪造检测方法评估》。IEEE信息取证与安全汇刊 7，1841–1854。doi：10.1109/TIFS.2012.2218597。

Dong, J., Wang, W., Tan, T., 2013年。CASIA图像篡改检测评估数据库，载于：2013年 IEEE 中国峰会暨信号与信息处理国际会议，页码 422–426。doi：10.1109/ChinaSIP.2013.6625374。

Du, X., Hargreaves, C., Sheppard, J., Anda, F., Sayakkara, A., Le-Khac, N.A., Scanlon, M., 2020。SoK：探讨人工智能在数字取证调查中的最新进展与未来潜力，载于：第15届国际可用性、可靠性与安全会议论文集，美国计算机协会，纽约，纽约州，美国。网址：https://doi.org/10.1145/3407023.3407068，doi：10.1145/3407023.3407068。

Feng, D., Lu, X., Lin, X., 2020. 用于人脸篡改的深度检测，载于：Yang, H., Pasupa, K., Leung, A.C.S., Kwok, J.T., Chan, J.H., King, I.（编），《神经信息处理》，Springer国际出版公司，Cham，第316–323页。doi：10.1007/978-3-030-63823-8_37。

Fischinger, D., Boyer, M., 2023年a. DF-Net：用于图像伪造检测的数字取证网络，载于：Corcoran, P.（编），《第25届爱尔兰机器视觉与图像处理会议论文集》（IMVIP 2023），第120–127页。网址：https://iprcs.github.io/pdf/IMVIP2023_Proceeding.pdf，doi：10.5281/zenodo.8142657。

Fischinger, D., Boyer, M., 2023年b. DF2023：用于图像伪造检测的数字取证 2023 数据集，载于：Corcoran, P.（编），第 25 届爱尔兰机器视觉与图像处理会议（IMVIP 2023）论文集，第 128–135 页。网址：https://iprcs.github.io/pdf/IMVIP2023_Proceeding.pdf，doi：https://doi.org/10.5281/zenodo.8214996。
Guarnera, L., Giudice, O., Battiato, S., 2024. 精通深度伪造检测：区分生成对抗网络图像与扩散模型图像的前沿方法。《ACM 多媒体计算、通信与应用汇刊》20。网址：https://doi.org/10.1145/3652027，doi：10.1145/3652027。Guo, X., Liu, X., Ren, Z., Grosz, S., Masi, I., Liu, X., 2023. 分层细粒度图像伪造检测与定位，载于：2023 IEEE/CVF计算机视觉与模式识别会议（CVPR），第3155–3165 页。doi：10.1109/CVPR52729.2023.00308。

Hao, J., Zhang, Z., Yang, S., Xie, D., Pu, S., 2021. Trans-Forensics：基于密集自注意力的图像伪造定位，载于：2021 IEEE/CVF国际计算机视觉会议（ICCV），第15035–15044 页。doi：10.1109/ICCV48922.2021.01478。

Hargreaves, C., Breitinger, F., Dowthwaite, L., Webb, H., Scanlon, M., 2024. DFPulse：2024 年数字取证从业者调查。《国际法庭科学：数字调查》51，301844。网址：https://www.sciencedirect.com/science/article/pii/S2666281724001719，DOI：https://doi.org/10.1016/j.fsidi.2024.301844。

Kingra, S., Aggarwal, N., Kaur, N., 2024. SFormer：一种用于深度伪造检测的端到端时空 Transformer 架构。《国际法庭科学：数字调查》51，301817。网址：https://www.sciencedirect.com/science/article/pii/S2666281724001410，DOI：https://doi.org/10.1016/j.fsidi.2024.301817。

Li, Y., Chang, M.C., Lyu, S., 2018. In Ictu Oculi：通过检测眨眼揭露人工智能生成的伪造视频，载于：2018年IEEE信息取证与安全国际研讨会（WIFS），第1–7页。doi：10.1109/WIFS.2018.8630787。

Li, Y., Yang, X., Sun, P., Qi, H., Lyu, S., 2020. Celeb-DF：用于深度伪造取证的大规模挑战性数据集，载于：2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），第3204–3213页。doi：10.1109/CVPR42600.2020.00327。

Ma, X., Du, B., Jiang, Z., Hammadi, A.Y.A., Zhou, J., 2023. IML-ViT：基于视觉Transformer的图像操纵定位基准测试。arXiv预印本 arXiv:2307.14863，网址：https://arxiv.org/abs/2307.14863，doi：10.48550/arXiv.2307.14863。

Mahdian, B., Saic, S., 2008. 通过噪声不一致性分析补充的重采样检测方法，用于图像取证，载于：2008 年国际计算科学及其应用大会论文集，第 546–556 页。doi:10.1109/ICCSA.2008.34。

Matern, F., Riess, C., Stamminger, M., 2019. 利用视觉伪影揭露深度伪造和人脸篡改，载于：2019 年 IEEE 冬季计算机视觉应用研讨会（WACVW）论文集，第 83–92 页。doi:10.1109/WACVW.2019.00020。

Monika, Bansal, D., Passi, A., 2021. 基于离散余弦变换的方法进行图像取证调查。Wirel. Pers. Commun. 119, 3241–3253. URL：https://doi.org/10.1007/s11277-021-08396-1，doi：10.1007/s11277-021-08396-1。

Munawar, M., Oussalah, M., 2025. 利用对比学习进行图像伪造检测与定位的可解释双流注意力网络。IET《雷达、声纳与导航》19, e70064. doi：10.1049/rsn2.70064。

Nguyen, H.H., Yamagishi, J., Echizen, I., 2022. 用于深度伪造检测的胶囊取证网络。Springer 国际出版，Cham，第 275–301 页。网址：https://doi.org/10.1007/978-3-030-87664-7_13，doi：10.1007/978-3-030-87664-7_13。

Pandey, R.C., Singh, S.K., Shukla, K.K., Agrawal, R., 2014. 使用 SURF 和 SIFT 图像特征的快速稳健的被动复制移动伪造检测，载于：2014年第九届工业与信息系统国际会议（ICIIS），第 1–6 页。doi：10.1109/ICIINFS.2014.7036519。

Sharma, S., Singh, B.K., Garg, H., 2025. 基于混合 CNN-Transformer 协同框架的稳健图像伪造定位。《计算机、材料与连续体》82，4691–4708。doi：10.32604/cmc.2025.061252。

Shi, C., Wang, C., Zhou, X., Qin, Z., 2024. DAE-Net：用于图像篡改检测与定位的双重注意力机制和边缘监督网络。《IEEE仪器与测量汇刊》73，1–17。doi：10.1109/TIM.2024.3451570。

Singh, J., 2022. 深度伪造：人工智能驱动的视听内容篡改时代对数据真实性和公众信任的威胁。《人工智能辅助科学发现期刊》2，428–467。

Spichiger, H., Adelstein, F., 2025. 保全演进系统中证据的含义。《法证科学国际：数字调查》52，301867。网址：https://www.sciencedirect.com/science/article/pii/S266628172500006X，doi：https://doi.org/10.1016/j.fsidi.2025.301867。DFRWS EU 2025——第十二届欧洲数字取证研究年会精选论文。

Stylianou, A., Xuan, H., Shende, M., Brandt, J., Souvenir, R., Pless, R., 2019。《Hotels-50K：全球酒店识别数据集》，载于：第三十三届AAAI人工智能会议、第三十一届创新人工智能应用会议及第九届AAAI人工智能教育进展研讨会论文集，AAAI出版社。网址：https://doi.org/10.1609/aaai.v33i01.3301726，doi：10.1609/aaai.v33i01.3301726。Thomson, M., McKeown, S., Macfarlane, R., Leimich, P., 2025。《探索数字取证中生成式人工智能图像修复定位的数据集多样性》，载于：数字取证博士生研讨会论文集，美国计算机协会，纽约，NY，美国。网址：https://doi.org/10.1145/3712716.3712724，doi：10.1145/3712716.3712724。

Verdoliva, L., 2020。《媒体取证与深度伪造：综述》。IEEE信号处理精选主题期刊 14，910–932。doi：10.1109/JSTSP.2020.3002101。Wang, J., Wu, Z., Chen, J., Han, X., Shrivastava, A., Lim, S.N., Jiang, Y.G., 2022。《用于图像篡改检测与定位的ObjectFormer》，载于：2022年IEEE/CVF计算机视觉与模式识别会议（CVPR），第2354–2363页。doi：10.1109/CVPR52688.2022.00240。Wu, Y., AbdAlmageed, W., Natarajan, P., 2019。《ManTra-Net：用于检测和定位具有异常特征的伪造图像的篡改追踪网络》，载于：2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），第9535–9544页。doi：10.1109/CVPR.2019.00977。Xia, X., Su, L.C., Wang, S.P., Li, X.Y., 2024。《DMFF-Net：用于图像伪造定位的双流多级特征融合网络》。人工智能工程应用 127，107200。网址：
h
t
t
p
s://
w
w
w.
sciencedirect.com/science/article/pii/S0952197623013842，doi：https://doi.org/10.1016/j.engappai.2023.107200。Yadav, A., Vishwakarma, D.K., 2024。《具备多域特征提取器和多感受野上采样器的视觉注意式拼接定位网络》。arXiv预印本 arXiv:2401.06995 doi：10.48550/arXiv.2401.06995。预印本。Zhou, P., Han, X., Morariu, V.I., Davis, L.S., 2018。《学习用于图像篡改检测的丰富特征》，载于：2018年IEEE/CVF计算机视觉与模式识别会议，第1053–1061页。doi：10.1109/CVPR.2018.00116。Zhuang, P., Li, H., Tan, S., Li, B., Huang, J., 2021。《使用密集全卷积网络进行图像篡改定位》。IEEE信息取证与安全汇刊 16，2986–2999。doi：10.1109/TIFS.2021.3070444。