

---

# Flood-DamageSense: 使用 SAR 遥感影像进行建筑物洪水损害评估的多模态 Mamba 与多任务学习

---

Yu-Hsuan Ho<sup>a,\*</sup>, Ali Mostafavi<sup>a</sup>

<sup>a</sup> Urban Resilience.AI Lab, Zachry Department of Civil and Environmental Engineering,  
Texas A & M University, College Station, TX

\* corresponding author, email: yuhsuanho@tamu.edu

## Abstract

大多数灾后损害分类器只有在破坏力留下明显的光谱或结构特征时才成功，而这些条件在淹水后很少出现。因此，现有模型在识别与洪水有关的建筑损害时表现不佳。本研究中提出的模型 Flood-DamageSense 填补了这一空白，作为首个专为建筑层级洪水损害评估量身打造的深度学习框架。其架构将事件前后 SAR/InSAR 场景与超高分辨率光学基图和内在洪水风险层融合，后者编码长期暴露概率，引导网络关注可能受影响的区域，即使构成变化最小。采用半 Siamese 编解码器与任务特定解码器的多模态 Mamba 骨干共同预测 (1) 分级建筑损害状态，(2) 洪水水域范围，和 (3) 建筑轮廓。在来自德克萨斯州哈里斯县的哈维飓风 (2017 年) 图像上进行训练和评估，由保险产生的财产损失范围支持，显示在“轻微”和“中度”损害类别中，F1 均值较最先进基准提升多达 19 个百分点。消融研究指出，内在风险特征是此性能提升的最重要贡献者。一个端到端后处理管道能在图像获取后的几分钟内将像素级输出转换为可操作的建筑尺度损害地图。通过结合风险感知模型和 SAR 的全天候能力，Flood-DamageSense 提供更快速、更细致和更可靠的洪水损害情报，以支持灾后决策和资源分配。

Keywords Rapid damage assessment · Mamba · Synthetic Aperture Radar (SAR) · Multimodal learning · Multitask learning · Remote sensing · Artificial intelligence · Disaster response

## 1 引言

快速进行建筑级别的洪水损害评估对于启动应急响应、调配资源和加速恢复至关重要，但传统的现场检查速度慢、成本高，并可能将人员置于危险境地 (Spencer et al., 2019)。受到这些限制的推动，研究人员越来越多地采用自动化技术进行洪水风险分析 (Ho et al., 2024, 2025; Huang et al., 2025)、预测 (Dong et al., 2021; Esparza et al., 2025) 和损失评估 (Dong et al., 2020; Nofal et al., 2024)。机器学习模型擅长从大型数据集中提取复杂的非线性模式 (Huang et al., 2025; Rafiei & Adeli, 2017)，而贝叶斯框架则提供了系统的不确定性量化 (Dong et al., 2020; Perez-Ramirez et al., 2019; Wang et al., 2024)。尽管有这些进步，但在个体结构尺度上提供近实时洪水损害情报的研究仍然处于起步阶段，主要原因是观测数据在洪水事件期间和紧接之后较为稀缺。

卫星影像具有快速获取和大陆尺度覆盖的双重优势。xBD 数据集 (Gupta et al., 2019) 的发布，其中包含来自各种灾难事件的建筑多边形和等级损害标记的灾前和灾后卫星影像，激发了使用光学传感器 (Braik & Koliou, 2024; Kaur et al., 2023) 开发有前景的多灾害损害映射管道。然而，洪水期间常见的云层在信息最需要的时候遮蔽了目标。其他领域中的类似可见度问题——例如通过红外热成像检测隐藏的混凝土缺陷 (Sirca Jr & Adeli, 2018) 或通过 MRI 揭示微妙的脑部异常 (Mirzaei & Adeli, 2018; Nogay & Adeli, 2020)——已经通过部署更适合该任务的传感方式得到解决。在地球观测中，合成孔径雷达 (SAR) 提供了一种全天候、全天时的替代方案，而用于洪水水域映射的 SAR 强度和 InSAR (干涉合成孔径雷达) 相干性已经得到充分确立 (Amitrano et al., 2024; Singh et al., 2021)。然而，粗糙的本地分辨率和斑点噪声限制了 SAR 在建筑级洪水损害评估中的应用 (Cao & Choe, 2020)。将 SAR 与光学数据融合 (Rudner et al., 2019) 或将

SAR 嵌入贝叶斯因果网络 (Wang et al., 2024) 的混合方法显示出前景，但要么继承了光学影像的天气限制，要么依赖于资源密集型推理。

更广泛地说，卫星变化检测研究已经从卷积的 Siamese U-Nets (Durnov, 2020) 发展到了 Transformer 架构 (Bandara & Patel, 2022; Chen et al., 2022; Dosovitskiy et al., 2020; Kaur et al., 2023; Vaswani et al., 2017)，再到最近的 Mamba 状态空间模型家族，其线性复杂度提供了高效的长程上下文建模 (A. Gu & Dao, 2023; Y. Liu et al., 2024; Zhu et al., 2024)。遥感变体，如 ChangeMamba (Chen et al., 2024)，已经开始利用这些特性进行一般变化检测和多灾害损害映射，然而没有一个针对典型洪水影响的微妙、低对比度特征。专门的、具有灾害特定性的模型强调了量身定制特征设计的价值：风和地震研究结合了航拍、街景或 SAR 数据以捕捉这些事件独有的现象 (Cheng et al., 2021; D. Gu et al., 2025; Khajwal et al., 2023; Wu et al., 2025; Xiao et al., 2025; Xiao & Mostafavi, 2025)。然而，对于洪水来说，现有的卫星处理管道很少超出二元损害标签，并且经常依赖于需大量人工且易出错的手工标注地面实况 (Cao & Choe, 2020; Rudner et al., 2019)。此外，大多数视觉类论文停留在像素级输出；只有很少的工作流尝试将模型预测转化为地理参考、建筑物级别的地图，而这些通常是从每一个足迹中提取孤立的图像片段，因而丢失了上下文提示并引入了额外的计算 (Braik & Koliou, 2024)。公共建筑足迹数据库 (Google, 2023; Microsoft, 2018, 2022) 仍然是将栅格预测直接关联到个别结构和简化图像到地图流程的未开发资产。

这些差距催生了 Flood-DamageSense，这是首个为多层次、建筑特定的洪水损害评估专门设计的深度学习框架。该系统采用了一种多模态 Mamba 骨干网，通过接收成对的事件前后 SAR-InSAR 场景、事件前超高分辨率 (VHR) 光学基础地图，以及从地形、水文和保险索赔记录中衍生出的历史洪水风险表面 (Mobley et al., 2021)。洪水水体分割被引入作为一个辅助任务，以规范学习并实施水文一致性，同时监督由来自全国洪水保险计划 (NFIP) 和联邦紧急事务管理署 (FEMA) 个人援助 (IA) 赔款 (C.-F. Liu et al., 2024) 编制的财产损失程度 (PDE) 标签提供。结合一个建筑物轮廓覆盖，该系统将像素概率转换为每座建筑的分数，该流程在 SAR 获取后的数分钟内生成带地理参考的损害地图，可在完全云覆盖下操作，保留细尺度的语义信息，并消除人工标注的需求。关于哈维飓风 (得克萨斯州哈里斯县, 2017 年) 的案例研究表明，Flood-DamageSense 可以准确区分轻度和中度损失——这是以前的光学为中心的模型最常失败的类别，从而填补了灾难响应工具箱中的关键空白。

## 2 方法论

本节详细介绍了提出的 Flood-DamageSense 框架，该框架专为通过主要利用 SAR 遥感影像克服光学传感器固有的天气相关限制来快速进行建筑级洪水损害评估而设计。该框架采用一种基于 Mamba 的多模态架构，利用多任务学习。多模态输入包括 SAR 影像以保持对天气条件的鲁棒性，历史洪水风险水平以增强对不涉及结构变化的损害的感知，以及事件前的 VHR 光学卫星影像以细化空间细节。为了进一步提升建筑损害评估的性能，在这个多任务学习设置中，洪水水体映射被纳入为一个辅助目标。综合方法，如图 1 所示，包括数据预处理、通过 VMamba (Y. Liu et al., 2024) 编码器的特征提取以及多任务解码以生成建筑定位掩模、建筑损害水平和洪水水图。在解码阶段，我们引入特征融合状态空间 (FFSS) 来共同解码由 VMamba 编码的多模态特征。最后，将建筑损害水平与地理参考的建筑足迹整合以生成详细的建筑级损害地图。

### 2.1 预备知识

状态空间模型 (SSMs) 起源于卡尔曼滤波器 (Kalman, 1960)，通过隐藏状态  $h(t) \in \mathbb{R}^N$  将一维序列  $x(t) \in \mathbb{R}$  转换为输出响应  $y(t) \in \mathbb{R}$ 。这种模型通常表示为以下形式的线性常微分方程 (ODEs)：

$$h'(t) = Ah(t) + Bx(t), \quad (1)$$

$$y(t) = Ch(t) + Dx(t), \quad (2)$$

其中  $h'(t)$  是当前状态  $h(t)$  的导数， $A \in \mathbb{R}^{N \times N}$ 、 $B \in \mathbb{R}^{N \times 1}$ 、 $C \in \mathbb{R}^{1 \times N}$ 、 $D \in \mathbb{R}$  是可学习参数。方程 1 定义了状态更新函数，其中矩阵  $A$  控制当前状态  $h(t)$  的时间演化，而矩阵  $B$  决定当前输入  $x(t)$  如何影响该状态。方程 2 表示输出函数，其中矩阵  $C$  将当前状态  $h(t)$  映射到输出，矩阵  $D$  决定当前输入  $x(t)$  对输出的直接影响。 $Dx(t)$  可被解释为从输入到输出的馈通或跳跃连接，在某些模型公式中常被省略。

为了使 SSM 适合离散标记序列并与深度学习环境兼容，必须对其进行离散化处理，将连续参数转换为离散参数。通常，离散化技术将一个连续时间线分为  $K$  个不同的时间间隔，理想情况下保持每个间隔的积分面积相等。一个显著的方法来完成这一转换是零阶保持 (ZOH) 方法，该方法已经在 SSM 中有效地应用。ZOH 方法的运作基础是假设函数值在每个时间间隔内保持不变，记为  $\Delta = [t_{k-1}, t_k]$ 。应用 ZOH 离散化后，方程 1 和方程 2 可以重新表述为

$$h_k = \bar{A}h_{k-1} + \bar{B}x_k, \quad (3)$$

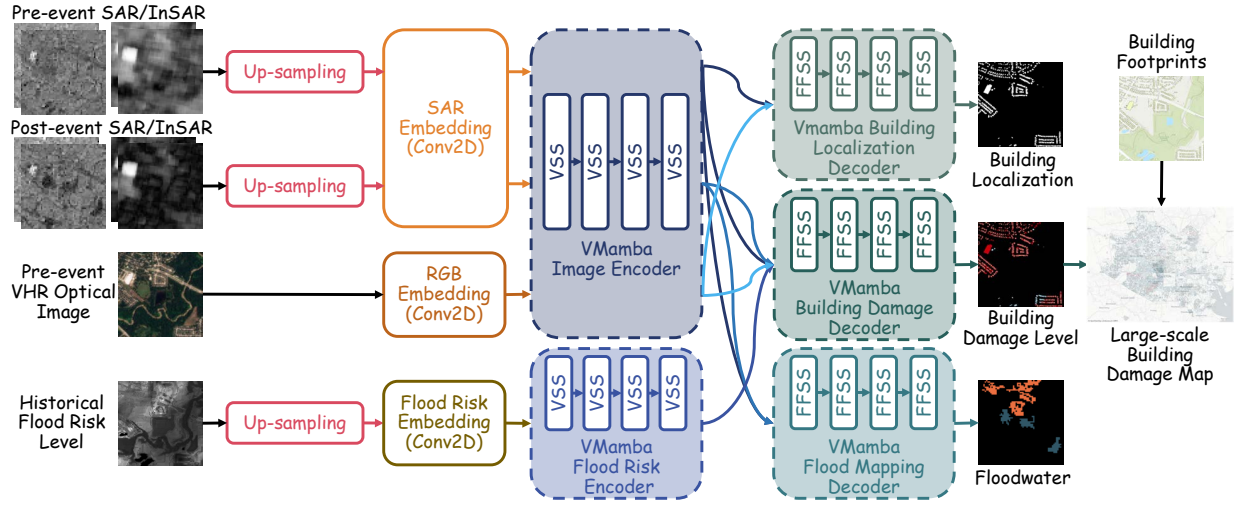


Figure 1: Flood-DamageSense 的架构。该框架处理多模态输入，包括事件前后 SAR 和 InSAR 影像、事件前的超高分辨率光学影像以及历史洪水风险水平，通过初始嵌入层后接 VMamba 编码器。多任务学习方法利用不同的解码器同时生成建筑物定位掩码、建筑物损坏等级和洪水地图。像素级的建筑物损坏输出随后与地理参考的建筑物轮廓相结合，生成最终的大规模建筑物损坏地图。

$$y_k = Ch_k, \quad (4)$$

其中

$$\bar{A} = \exp(\Delta A), \quad (5)$$

$$\bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I)\Delta B. \quad (6)$$

注意在此公式中省略了  $Dx_k$ 。

在经典的计算机视觉任务中，滤波器通常以卷积核的形式被用来从输入数据中提取聚合特征。在离散化后，SSM（状态空间模型）也可以被表示为卷积操作。鉴于 SSM 本质上处理的是序列数据，这种卷积表示使用的是一维卷积核。这一维卷积核的参数是通过离散化的 SSM 参数数学推导而来的。具体来说，

$$y_0 = Ch_0 = C\bar{B}x_0, \quad (7)$$

$$y_1 = Ch_1 = C(\bar{A}h_0 + \bar{B}x_1), \quad (8)$$

$$y_k = Ch_k = C(\bar{A}h_{k-1} + \bar{B}x_k) = C\bar{A}^k\bar{B}x_0 + C\bar{A}^{k-1}\bar{B}x_1 + \dots + C\bar{A}^1\bar{B}x_{k-1} + C\bar{B}x_k. \quad (9)$$

因此，卷积核  $\bar{K}$  可以被公式化为

$$\bar{K} = (C\bar{B}, C\bar{A}^1\bar{B}, \dots, C\bar{A}^k\bar{B}, \dots, C\bar{A}^{K-1}\bar{B}). \quad (10)$$

递归运算现在可以转换为卷积操作，如下所示

$$y = x * \bar{K}, \quad (11)$$

其中  $x = [x_0, x_1, \dots, x_{K-1}] \in \mathbb{R}^K$  和  $y = [y_0, y_1, \dots, y_{K-1}] \in \mathbb{R}^K$  分别表示输入和输出序列。

以前的 SSM 的一个主要限制是只有那些表现出线性时不变性 (LTI) 的才能有效地计算为卷积。在输入和输出之间具有可变间距的场景无法通过静态卷积核进行有效建模。然而，LTI 模型固定的动态限制了它们执行基于内容推理的能力。Mamba (A. Gu & Dao, 2023) 通过引入选择性扫描机制解决了 SSM 的这些限制。这种方法通过使 SSM 参数依赖于输入，从而克服了 LTI 模型的限制，实现了选择性信息处理。此外，Mamba 采用了一种硬件感知的并行算法以确保计算效率。具体来说，由于 Mamba 的时变设计排除了使用有效卷积的可能性，它利用并行扫描算法进行其递归计算。该算法通过内核融合进行了优化以便在图形处理单元 (GPU) 上运行，从而最大限度地减少了内存输入/输出操作，并在训练期间采用重新计算以避免存储中间状态以进行反向传播，从而管理其选择性状态的内存需求。

## 2.2 网络架构

所提出的 Flood-DamageSense 模型建立在 VMamba (Y. Liu et al., 2024) 架构和 ChangeMamba (Chen et al., 2024) 框架之上。图 1 展示了所提议的网络架构。输入数据包括事件前后在垂直-垂直 (VV) 和垂直-水平 (VH) 极化下的 SAR 强度和 InSAR 相干性、事件前的高分辨率光学卫星图像以及历史洪水风险水平地图。最初，将输入数据上采样以达到统一分辨率，与最高分辨率输入一致，从而确保后续操作的一致性。在编码阶段，采用基于 Visual State Space (VSS) 主干网的半孪生网络结构，具有两个不同的编码器：一个被指定为图像编码器，另一个作为洪水风险数据编码器。事件前后的 SAR 和 InSAR 图像，以及事件前的高分辨率光学图像，由一个共享权重的图像编码器处理，以促进空间信息的联合学习。另一方面，历史洪水风险数据由洪水风险数据编码器单独编码，考虑到与图像数据相比，洪水风险数据具有不同的特征模式和较低的信息密度。考虑到不同传感器类型的通道数不同，SAR/InSAR 和光学图像数据起初由独立的卷积嵌入层处理，以确保进入图像编码器之前的维度一致性。

在解码阶段，本研究引入了一个特征融合状态空间模块以联合解码多模态特征。这个 FFSS 模块扩展了 ChangeMamba 中的时空状态空间 (STSS) 概念，如图 2 所示，以适应多个特征输入。三个特定任务的解码器随后利用这些处理过的特征的不同组合。具体来说，

- 建筑损坏解码器处理来自事件前后 SAR 和 InSAR 影像、事件前 VHR 光学影像以及历史洪水风险水平的特征，作为输入通过时空关系以及与事件无关的信息感知建筑损坏的程度。
- 洪水水位映射解码器从事件前后 SAR 和 InSAR 图像中提取特征，作为一个纯粹的变化检测解码器。
- 建筑定位解码器使用来自事件前 SAR 和 InSAR 影像以及事件前 VHR 光学影像的特征，作为一个具有特征融合能力的语义解码器进行操作。

虽然在这个框架中，建筑物定位输出不是直接的最终产品，但它在细化建筑级别损害评估的关联方面起到了辅助作用，特别是在使用地理参照的建筑轮廓进行最终聚合之前。

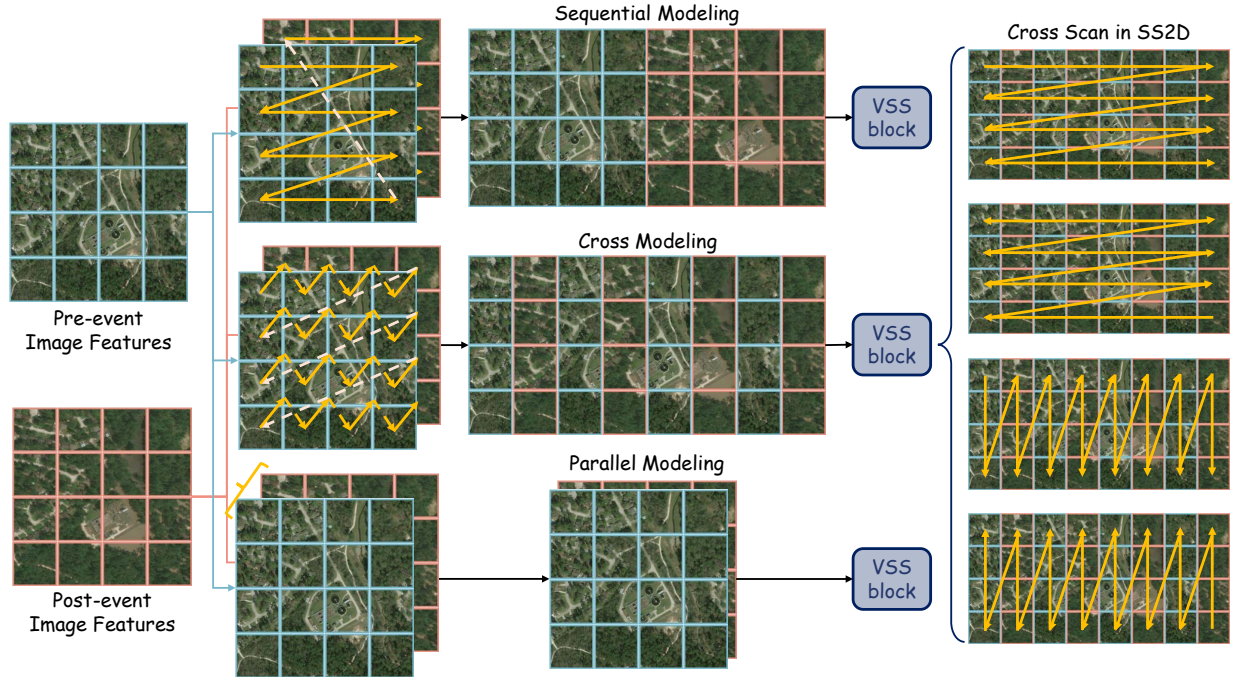


Figure 2: 时空状态空间的示意图。一个 STSS 模块包含三个视觉状态空间块。STSS 通过顺序、交叉和并行建模机制的组合来处理事件前后的图像特征。每种建模机制重新排列的特征然后被输入到对应的 VSS 块中。在 VSS 中，一个交叉扫描机制利用四个方向的扫描路径有效地从时空特征中建模全局上下文信息。需要注意的是，所描述的输入表示编码特征而非原始图像。所示的光学卫星图像源自于 xBD 数据集 (Gupta et al., 2019)，只在插图中用于增强概念理解。

## 2.2.1 半孪生编码

Mamba 架构最初是为处理具有因果关系的序列数据而开发的。因此，其原始的一维扫描机制不易转换为非序列数据模式，如图像，这些模式通常表现空间而非严格因果依赖关系。VMamba 通过引入 2D 选择性扫描 (SS2D) 将 Mamba 架构适应于视觉任务，该机制采用四向交叉扫描机制，通过沿四条不同路径重新排序标记，将二维图像块转换为多个扁平化的一维序列。这些标记序列随后由四个不同的选择性扫描结构状态空间序列 (S6) 块独立处理。然后合并每个扫描方向获得的特征，从而可以在最终特征图中有效捕获和重建全面的二维空间上下文。SS2D 机制详细说明于图 3 (a) 的 SS2D 面板中。

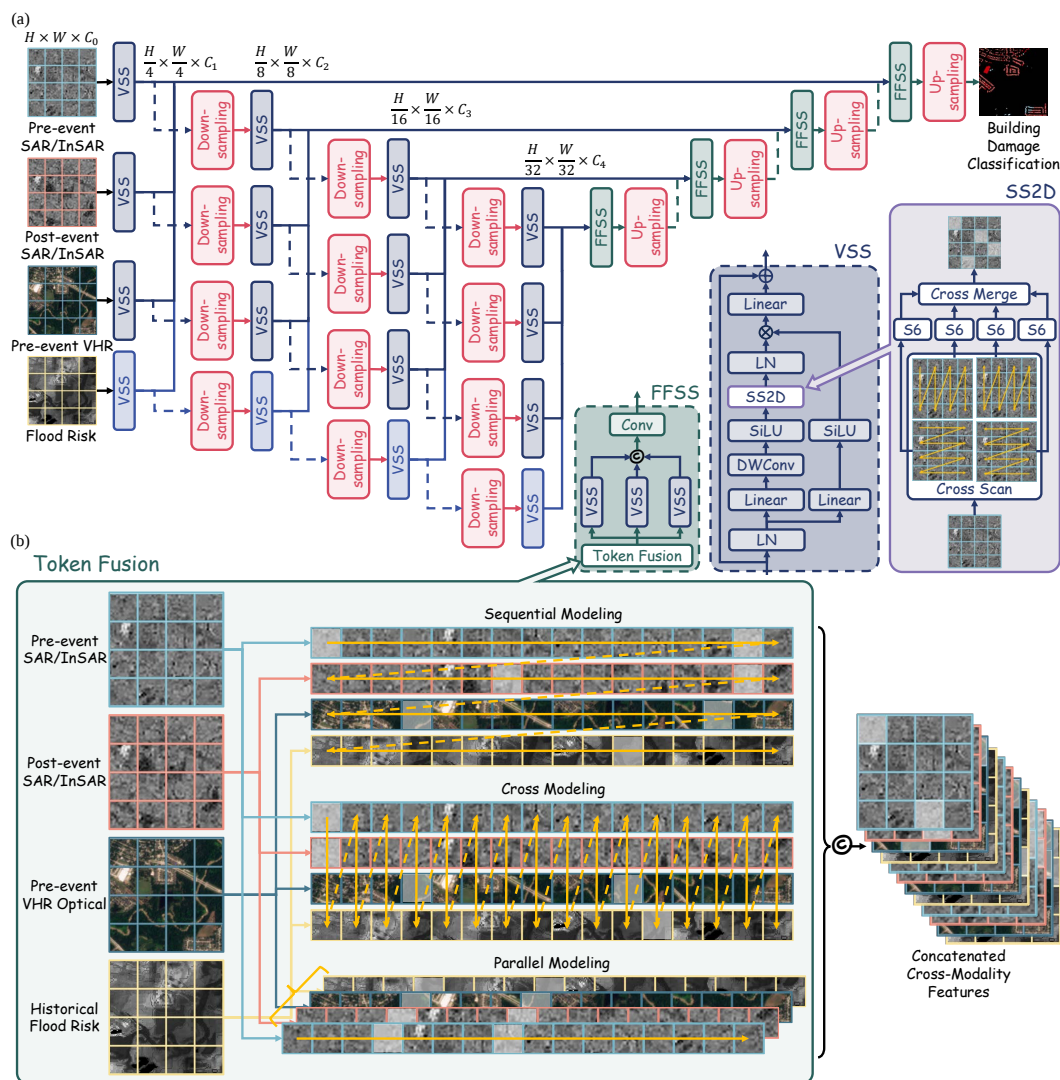


Figure 3: 在 Flood-DamageSense 模型中编码器-解码器网络和特征融合状态空间的详细信息。(a) 编码器-解码器网络的详细架构。编码器部分具有多个并行的视觉状态空间块流，通过逐步下采样来提取层次特征。随后，解码器通过与上采样阶段组合使用的模块重建空间细节，并生成建筑损坏分类输出。三个扩展视图详细展示了 FFSS 块、VSS 块和二维选择扫描机制的内部结构。尽管图中主要展示了建筑损坏解码器的路径，但三种任务特定解码器在完整模型中都采用了所展示的解码器架构。(b) FFSS 模块中标记重排机制的概念图。FFSS 被设计用于处理从多种输入模式得出的 VSS 编码特征。这些特征经过 ChangeMamba 中的时空建模技术启发的顺序、交叉和平行建模方式进行处理，作为重排和融合来自各种来源的标记的机制。由这三种建模机制产生的跨模式特征随后被拼接。特定 SS2D 操作没有在此图例中详细说明，因为它们的原理在图 2 中进行了详细说明。此外，与图 2 中的描述相似，此处展示的输入代表的是 VSS 编码特征，而非原始图像数据。

通过 SS2D 机制，利用 VMamba 架构有效地将 SSM 转用于视觉任务，提出的 Flood-DamageSense 模型的编码阶段被专门设计用于处理多模态输入并提取层次特征，以进行后续的损坏评估，如图 3 (a) 的左半部分所示。该编码阶段采用了一个半孪生网络结构。这种配置包含一个权重共享编码器和一个特定于特征的编

码器，二者均基于来自 VMamba 的视觉状态空间主干架构，如图 3 (a) 中的 VSS 面板所详述，其中 LN 表示层归一化，linear 指的是线性层，而 DWConv 则表示深度卷积操作。这个半孪生网络的第一个组件，被指定为权重共享图像编码器，处理各类图像输入，包括事件前和事件后的 SAR 和 InSAR 影像，以及事件前的 VHR 光学影像。这些不同的输入最初通过专门针对每个传感器模式定制的嵌入层进行处理，然后进入这个单一的、权重共享的图像编码器。这一结构选择被采用是因为权重共享提供了几种独特的优势：它确保在不同的时间状态和传感器模式下提取一致的特征表示，通过减少可学习参数的总数来提高参数效率，从而可能减轻过拟合，并引导模型专注于区分真正的时空变化和图像间关系，而不是由于不同处理路径导致的变化。结合旨在标准化维度并捕获传感器特定特征的不同嵌入层与旨在学习联合表示的共享编码器在多传感器地理空间模型 (Han et al., 2024) 中被证明是有效的。第二个组件，洪水风险数据编码器，专门用于处理历史洪水风险数据。这种结构性分离适应了洪水风险数据中固有的不同行为模式和通常较低的信息密度，与卫星图像的丰富内容相比不同。编码器网络包括四个分层阶段，旨在跨多个尺度捕获语义信息，从全局到局部。初始阶段将输入图像划分为小块。相反，后续的三个阶段中的每一个都以对其输入特征进行下采样开始。在这些初步操作（小块划分或下采样）之后，每个阶段都使用一系列 VSS 块对空间上下文信息进行全面建模，从而生成分层特征表示。

## 2.2.2 特征融合解码

解码阶段对于合成由 VSS 编码器生成的丰富多模态信息至关重要。这种复杂的整合主要由特征融合状态空间模块管理。FFSS 的设计灵感来源于 ChangeMamba 中的时空状态空间模块，这是因为融合时空信息和多模态信息在概念上具有相似性，这两个过程都涉及整合来自不同来源的特征。因此，FFSS 模块被设计为能够同时处理时空信息和多模态信息。

如图 3 (b) 所示，FFSS 模块接收从不同模态派生的 VSS 编码特征。在该模块中，这些多模态特征通过一个复杂的整合过程进行处理，该过程采用三种不同的建模策略来有效地重新排序和融合来自各种输入流的标记。这些策略重新排列的标记序列，然后由三个对应的 VSS 块逐个处理。FFSS 模块的内部结构，包括这些 VSS 组件，如图 3 (a) 中的 FFSS 面板所示。如图 2 (b) 所示，这三个重新排列的标记序列中的每一个随后通过由 SS2D 机制实现的四向交叉扫描。结合 FFSS 和 SS2D 的方法可以被概念化为一个双层叠加扫描过程，该过程能够探索 12 种跨模态和时空关系。因此，这种全面的扫描方法，通过多序列、多方向操作，显著增强了模型捕获多样化上下文信息的能力。随后，这些 VSS 块的输出被反向重新排列并连接，产生一个统一且丰富的跨模态特征表示。

解码器网络反映了编码器的四阶段层次结构，在对应的尺度上运行，如图 3 (a) 右侧所示。在每个尺度上，VSS 编码的特征被输入到相应的 FFSS 模块。令牌重排机制如图 3 ((b) 所示，用于每个解码阶段中的四个。在每个阶段的 FFSS 模块输出之后，又与前一、更深阶段的 FFSS 模块的上采样输出进行整合。来自最后一个 FFSS 模块的输出构成了特定任务的最终特征表示，有效地整合了来自所有解码阶段的层次信息。

上述机制涉及单个解码器路径的架构。所提出的 Flood-DamageSense 模型包含三个这样的特定任务解码器。虽然它们共享相同的基于 FFSS 的核心架构，但它们通过特定目标的编码特征的不同组合进行训练。建筑损坏解码器利用所有特征类型来共同学习时空关系和跨模态表示，其中，事件无关特征（例如历史洪水风险）的加入提高了模型对非结构性损坏的敏感性。洪水水位映射解码器处理事件前后 SAR 和 InSAR 图像的特征，以进行时空变化检测。建筑定位解码器则将事件前的 SAR/InSAR 和 VHR 光学卫星图像的特征作为输入，利用多模态信息执行语义分割。

## 2.3 损失函数和评估指标

Flood-DamageSense 模型使用复合损失函数  $L^{final}$  进行训练，该损失函数汇聚了三个主要任务的损失。这种多任务学习方法使模型能够在优化每个特定目标的同时学习共享表示。对于每个任务  $i \in \{BDA, FM, LOC\}$ ，表示建筑损坏分类、洪水映射和建筑定位，如果存在有效的真实标签，则损失  $L^i$  被计算为交叉熵 ( $ce$ ) 损失和 Lovasz-Softmax ( $lov$ ) 损失的加权组合：

$$L^i = \lambda_{ce}^i \cdot L_{ce}^i + \lambda_{lov}^i \cdot L_{lov}^i, \quad (12)$$

其中  $L_{ce}^i$  是任务  $i$  的交叉熵损失， $L_{lov}^i$  是任务  $i$  的 Lovasz-Softmax 损失， $\lambda_{ce}^i$  和  $\lambda_{lov}^i$  是任务特定的超参数，用于平衡每个损失组件的贡献。任务  $i$  的交叉熵损失  $L_{ce}^i$ ，衡量真实类别分布与模型为每个像素估计的概率分布之间的差异。对于任务  $i$  中的一组  $M_{valid}^i$  个有效像素和  $K_i$  个类别，其中  $y_{j,k}^i$  是像素  $j$  和类别  $k$  的真实标签（使用单热编码）并且  $p_{j,k}^i$  是像素  $j$  和类别  $k$  的输出 softmax 概率，交叉熵损失被公式化为

$$L_{ce}^i = -\frac{1}{M_{valid}^i} \sum_{j=1}^{M_{valid}^i} \sum_{k=1}^{K_i} (y_{j,k}^i \cdot \log(p_{j,k}^i)). \quad (13)$$

Lovasz-Softmax 损失 (Berman et al., 2018) 是一种专门为语义分割任务设计的损失函数。其开发灵感源于传统像素级损失函数（如交叉熵）无法直接优化常见分割评价指标，比如交并比 (IoU) 或称为 Jaccard 指数的观察结果。虽然交叉熵独立惩罚每个错误分类的像素，但 IoU 指标评估的是每个类别中预测分割掩码与真实掩码之间的重叠，从而在更结构化或基于区域的层次上捕捉分割的质量。一个仅用交叉熵训练的模型可能会达到很高的像素级准确度，但仍可能产生边界或形状不佳的分割，导致 IoU 评分不理想。为了弥合这一差距，在 multi-class 分类的背景下，Lovasz-Softmax 损失作为 IoU 指标的直接、可微分的替代。

为了解决类别不平衡问题，尤其是在建筑物损害分类和洪水映射任务中，将类别特定权重纳入交叉熵和 Lovasz-Softmax 损失计算中。这些权重是基于从训练数据集中导出的每个类别的逆频率确定的，以给予少数类别适当的重要性。任务  $i$  和类别  $k$  的权重  $w_k^i$  被制定为

$$w_k^i = \frac{N^i}{N_k^i}, \quad (14)$$

其中  $N^i$  是任务  $i$  的实例总数，而  $N_k^i$  是训练数据集中属于任务  $i$  的类别  $k$  的实例数。类别特定权重是在建筑物层面计算的，而不是更细粒度的像素层面，因此降低了相关的计算成本。给定方程 14，加权交叉熵损失可以表述为

$$L_{ce}^i = -\frac{1}{M_{valid}^i} \sum_{j=1}^{M_{valid}^i} \sum_{k=1}^{K_i} w_k^i (y_{j,k}^i \cdot \log(p_{j,k}^i)). \quad (15)$$

训练模型的最终复合损失  $L^{final}$  是来自三个任务的各自损失的总和：

$$L^{final} = L^{BDA} + L^{FM} + L^{LOC} = \lambda_{ce}^{BDA} \cdot L_{ce}^{BDA} + \lambda_{lov}^{BDA} \cdot L_{lov}^{BDA} + \lambda_{ce}^{FM} \cdot L_{ce}^{FM} + \lambda_{lov}^{FM} \cdot L_{lov}^{FM} + \lambda_{ce}^{LOC} \cdot L_{ce}^{LOC} + \lambda_{lov}^{LOC} \cdot L_{lov}^{LOC}. \quad (16)$$

如果给定批次中对于任务  $i$  没有有效的标签，则任务  $i$ ， $L^i$  对  $L^{final}$  的损失贡献为零。

为了评估所提出模型的性能，特别是在 multi-class 分类和分割任务的背景下，采用了一套稳健的评价指标。主要关注的指标是 F1 分数，对每个类别单独计算，并使用这些每类 F1 分数的调和平均得到一个聚合的度量。特定类别  $k$  的 F1 分数是精确度  $P_k$  和召回率  $R_k$  的调和平均，这在这两个常常竞争的度量之间提供了平衡。当处理不平衡的类别分布时，它特别有用。精确度衡量正预测的准确性，而召回率衡量模型识别所有真实正例的能力。类别  $k$  的精确度  $P_k$  和召回率  $R_k$  定义为

$$P_k = \frac{TP_k}{TP_k + FP_k}, \quad (17)$$

$$R_k = \frac{TP_k}{TP_k + FN_k}, \quad (18)$$

其中  $TP_k$ ，类别  $k$  的真正例数，是正确分类为类别  $k$  的实例数， $FP_k$ ，类别  $k$  的假正例数，是错误分类为类别  $k$  的实例数，而  $FN_k$ ，类别  $k$  的假负例数，是属于类别  $k$  但被错误分类为其他类别的实例数。类别  $k$  的 F1 分数  $F1_k$  随后计算为

$$F1_k = 2 \cdot \frac{P_k \cdot R_k}{P_k + R_k}. \quad (19)$$

计算每个类别的 F1 得分，可以详细了解模型在其被要求识别的不同类别上的表现。为了获得一个总结模型在所有类别上的整体性能的综合指标，使用单个 F1 得分的调和平均数。调和平均数优于简单的算术平均数，因为它对较低的 F1 得分更为敏感；即使模型在其他类别表现良好，它也会惩罚那些在一个或多个类别上表现非常差的模型，从而提供一个更加保守且往往更为现实的整体效能度量，特别是在每类性能不同的情况下。任务  $i$  的 F1 得分的调和平均数为

$$F1_{HMean}^i = \frac{K_i}{\sum_{k=1}^{K_i} \frac{1}{F1_k^i}}. \quad (20)$$

值得注意的是，为了使用 F1 得分的调和平均数计算综合性能指标，在建筑损坏评估的背景下排除了无损坏类别，类似地，在洪水映射任务中排除了非洪水类别。这一方法选择确保评估更加关注模型在检测和区分实际存在及损坏或洪水严重程度方面的能力；这些类别通常具有更高的操作重要性，并且往往代表更具挑战性的分类情况。通过排除这些通常占主导地位的负面类别，综合调和平均 F1 得分为模型在目标关注类别上的表现提供了一个更严格且具有代表性的衡量标准，减轻了由于在易于识别的无损伤或非洪水实例上高准确度导致的指标通胀的潜在影响。

除了标准的 F1 分数，还引入了两种修改的 F1 指标，即邻近 F1 分数和上邻近 F1 分数，用于评估建筑损坏评估中的性能，这是一项以有序类关系为特征的任务。这些指标特别适合于存在序关系的任务，因为它们通过将接近真实类别的预测视为部分正确，提供了更细致的理解。具体而言，对于给定类别  $k$  的邻近 F1 分数，

记作  $F1_{adj,k}$ ，通过赋予模型输出为真实类别  $k$  或其直接相邻类别 ( $k-1$  或  $k+1$ ) 的情况一定的信用，放宽了正确预测的标准。这种方法尤其适用于有序尺度，因为在这种情况下，分类在相邻类别之间的误分类通常被认为比涉及更远类别的错误不那么严重。要计算  $F1_{adj,k}$  的邻近 F1 分数，重新定义了精确度和召回率。邻近精确度  $P_{adj,k}$  衡量的是，对于所有被预测为类别  $k$  的实例，根据邻近标准正确识别的比例，其计算公式为

$$P_{adj,k} = \frac{N(GT \in \{k-1, k, k+1\} \wedge Pred = k)}{N(Pred = k)}. \quad (21)$$

；邻近召回率  $R_{adj,k}$  衡量的是，对于所有真实值为类别  $k$  的实例，根据邻近标准正确识别的比例，其计算公式为

$$R_{adj,k} = \frac{N(GT = k \wedge Pred \in \{k-1, k, k+1\})}{N(GT = k)}. \quad (22)$$

。然后使用标准的 F1 公式计算  $F1_{adj,k}$ 。上邻接 F1 分数  $F1_{upper\_adj,k}$  是一个更严格的变体。它认为一个预测是正确的，如果它是实际类别  $k$  或是紧邻更高的类别  $k+1$ ，这在损害的上下文中意味着更为严格。这种度量在避免低估严重性比轻微高估更为关键的情境中可能有用。对于此度量，上邻接精确度  $P_{upper\_adj,k}$  量化了对于所有被预测为类别  $k$  的实例，实际类别是  $k$  或  $k-1$  的比例，其可以被公式化为

$$P_{upper\_adj,k} = \frac{N(GT \in \{k-1, k\} \wedge Pred = k)}{N(Pred = k)}. \quad (23)$$

上邻接召回率  $R_{adj,k}$  测量了对于所有实际类别是  $k$  的实例，模型正确召回为类别  $k$  或紧邻更高类别  $k+1$  的比例，可以被公式化为

$$R_{upper\_adj,k} = \frac{N(GT = k \wedge Pred \in \{k, k+1\})}{N(GT = k)}. \quad (24)$$

这些序列感知的 F1 分数提供了比严格的每类别 F1 分数更深入的模型行为洞察，特别是对于诸如损坏评估这样的任务，类别邻近性具有重要的解释意义。

在 Flood-DamageSense 生成每个像素的损坏评分后，这些栅格数据通过将其叠加在一个已经重新投影到相同坐标参考系统的地理参考建筑物轮廓图层上转换为矢量产品。对于每一个轮廓多边形，该工作流程提取交叉的损坏等级像素，并为结构分配一个等于这些值中位数的单一标签——这是一种既对离散的错误分类有鲁棒性又忠实于损坏等级序数特性的统计量。当需要在低估和高估之间进行不同权衡时，可以通过简单的配置选项选择替代的汇总方式（例如，均值、众数或最大值）。在聚合期间，算法还记录像素覆盖计数和轮廓内标签的方差，并标记覆盖不足或内部分歧较大的建筑，以便分析人员能快速审查潜在的对齐错误或树冠遮挡。增强后的轮廓图层，现携带明确的损坏等级和置信属性，被导出为带有嵌入式样式规则的开放标准矢量文件 (GeoPackage 或 GeoJSON)，随时可以在常用 GIS 环境中使用或用于网络切片流式传输。通过将聚合推迟到最后一步，管道在整个处理中保留了卫星影像的全部语义丰富性，同时提供了适合快速分流、索赔优先化和恢复计划的建筑级损坏地图。

### 3 实验与结果

本节对所提出的 Flood-DamageSense 模型进行了全面评估。通过一个针对 2017 年德克萨斯州哈里斯县哈维飓风的案例研究，调查了其在建筑层级洪水损害评估中的效果。使用由事件前后的 SAR 和 InSAR 影像、事件前的 VHR 光学影像以及历史洪水风险数据组成的多模态数据集，对模型的性能进行了严格评估，并通过洪水保险索赔数据确定地面真实情况。评估框架涉及多个关键维度：通过每类 F1 分数、相邻 F1 分数、上级相邻 F1 分数及其在损害类别中的调和平均数来衡量的定量性能；与相关基线模型的比较分析；以及旨在确定 Flood-DamageSense 的架构组件和输入模态的具体贡献的消融研究。随后的子节将系统地介绍研究区域和数据集，基准与基线模型，对实验配置进行详细说明，报告定量和定性发现，并讨论消融研究的结果。

对 Flood-DamageSense 模型的验证是通过一个以 2017 年“哈维”飓风期间德克萨斯州哈里斯县为中心的案例研究来进行的。“哈维”飓风是一次灾难性且极其重大的天气事件，特点是前所未有的降雨导致广泛和毁灭性洪水，造成该地区建筑物的不同程度的广泛损害。选择这一重大灾难作为研究区域，为评估先进损害评估方法的能力提供了一个坚实且具有挑战性的真实场景。至关重要，此特定事件和地点的全面损害真实性数据和相关地理空间数据集的可获得性是选择此研究的决定性因素。后续用于模型训练和评估的多模态数据集将在下文中详细介绍。

#### 3.0.1 SAR 和 InSAR 图像

本研究的主要 SAR 和 InSAR 数据来源于 UrbanSARFloods 数据集 (Zhao et al., 2024)。UrbanSARFloods 是一个基准数据集，源自 Sentinel-1 单视复杂 (SLC) 影像，专门设计用于提高大规模洪水地图绘制能力，

特别强调解决城市洪水探测以及开放区域洪水映射的挑战。该数据集提供了空间分辨率为 20 米的预处理 Sentinel-1 数据产品,包括事件前后 SAR 强度和 InSAR 相干性,以垂直-垂直和垂直-水平极化方式可用,涵盖 18 个全球洪水事件。所有数据产品都分割成尺寸为  $512 \times 512$  像素的图像碎片。UrbanSARFloods 提供了用于洪水映射的相应标注,将区域分类为未被淹没的区域、被淹没的开放区域和被淹没的城市区域,分别由像素值 0、1 或 2 表示。这些标注是通过对所有图像片进行的半自动方法的组合生成的,涵盖了广泛的范围,同时使用高分辨率光学数据对选定子集进行详细的手动注释以确保准确性。洪水标注被采用为我们辅助洪水映射任务的真实标签。由于有真实的损害数据可用,本研究利用了与飓风哈维相对应的 UrbanSARFloods 数据集。该子集包含 1,296 张图像。鉴于 UrbanSARFloods 数据集最初是为更广泛的洪水映射设计的,每个原始  $512 \times 512$  像素图像覆盖了相对较大的地理区域。为了使这些图像适应精细化的建筑级别分析,它们被划分为  $8 \times 8$  的网格,每个子图像为  $64 \times 64$  像素。此预处理步骤生成了总计 82,944 个用于模型输入的图像块。选择 UrbanSARFloods 数据集的主要原因是其全面提供了 SAR 强度和 InSAR 相干性数据,后者对于在复杂的城市建筑环境中识别洪水特别重要,因为基于强度的检测可能会产生模糊。

我们从 Apollo Mapping (2017) 获得了事件前的高分辨率光学卫星图像,其空间分辨率为 0.5 米。然而,这些高分辨率图像仅适用于研究区域的一小部分。在某个位置没有高分辨率光学数据的情况下,输入模型的是一个值为 255 的空图像矩阵。在本研究中,值 255 在所有数据模式中一致地用于表示缺失数据或无数据像素。

为了结合一个独立于事件的基线洪水易发性测量,本研究利用了一张由 Mobley et al. (2021) 开发的历史洪水风险等级图,该图使用机器学习方法为德克萨斯东南部提供了洪水危险层。该方法采用了随机森林分类模型,该模型是在国家洪水保险计划洪水索赔的广泛长期记录的基础上训练的,并结合了全面的高分辨率地理空间数据集,涵盖了地形变量,如海拔、到海岸和河流的距离、到最近排水点的高度,水文参数,如饱和导水率和曼宁粗糙系数,以及地表覆盖特征,包括不透水率。所得的数据产品提供了一个分辨率为 30 米的洪水风险地图。包括这一历史风险评估的目的是提升 Flood-DamageSense 模型的敏感性,特别是用于检测可能不涉及显著或明显结构变化的损害。

### 3.0.2 真实数据

在本研究中,建筑损坏评估的真实标签基于来自 C.-F. Liu et al. (2024) 的财产损失程度 (PDE) 数据。该 PDE 指标的推导包括几个关键的数据处理步骤,其中包括最初关注建筑物结构损坏,而不包括内容损坏,并对 NFIP 和 IA 计划索赔值应用特征封顶和最小-最大归一化技术。这种归一化确保在合并之前这两个不同的数据集之间的一致性和可比性。处理过的 PDE 作为训练和评估 Flood-DamageSense 模型的主要目标变量。利用 PDE 作为真实标签解决了手动洪水损坏标注的困难和潜在的不准确性,尤其是对于大规模事件,通过提供更量化和系统得出的损坏测量。需要注意的是,PDE 数据集来自于实际索赔记录,这意味着没有索赔的财产不包含在这个已处理的损坏程度数据中。连续的 PDE 值通过 K-means 聚类被分为三级损坏程度,聚类的最佳数目通过肘部法确定。为了 Flood-DamageSense 模型的目的,采用了四类损坏分类方案,利用从 PDE 得出的三级损坏等级来表示轻微损坏、中度损坏和严重损坏,分别用像素值 1、2 或 3 表示。第四类,无损坏,则用像素值 0 表示,定义为研究区域内存在但未在基于索赔的 PDE 数据集中表示的财产。最后,使用从 Microsoft USBuildingFootprints 数据集 (Microsoft, 2018) 获取的哈里斯县的地理参考建筑轮廓,生成最终的大规模建筑洪水损坏地图,并为辅助建筑定位任务提供真实标签。

最初的 PDE 数据以点数据形式提供,经过几个预处理步骤后,关联到各个建筑物轮廓。首先,PDE 点被空间关联到相应的建筑物轮廓多边形上。对于那些没有直接包含任何 PDE 点的建筑物轮廓,与距离最近的 PDE 点建立关联,前提是距离不超过 100 米,该措施旨在尽量减少潜在的不匹配。为了进一步考虑那些初始没有洪水索赔但可能遭受损坏的建筑物,特别是在周围有索赔的房产中,一个加权平均  $k$  近邻 (kNN) 插补方法被采用。kNN 方法由三个关键参数控制: $k$ ,表示纳入距离加权平均的最近邻居数量; $d$ ,定义考虑邻居的最大距离限制;以及  $k_{min}$ ,规定需要用于推断无索赔财产损坏的最少邻居索赔财产数。这些参数  $d$  和  $k_{min}$  共同决定了识别有资格进行损坏插补的房产的标准。三个参数配置被用于实验,如图 4 中详细说明。在以高洪水索赔密度为特征的地区,测试的三个配置产生了相似的插补结果。然而,在低洪水索赔密度的区域,两个使用  $k_{min}$  值为 3 的配置导致超过的损坏插补。基于这些实验观察,选择了  $k = 5$ 、 $d = 100$  m 和  $k_{min} = 5$  的参数配置用于 PDE 值的插补。

## 3.1 实验配置

### 3.1.1 数据准备与增强

数据集被划分为训练集、验证集和测试集,比例分别为 60 %、20 % 和 20 %。对于所提出的 Flood-DamageSense 网络以及包含在消融研究中的模型,采用了一致的数据预处理流程来处理多模态输入。首先,

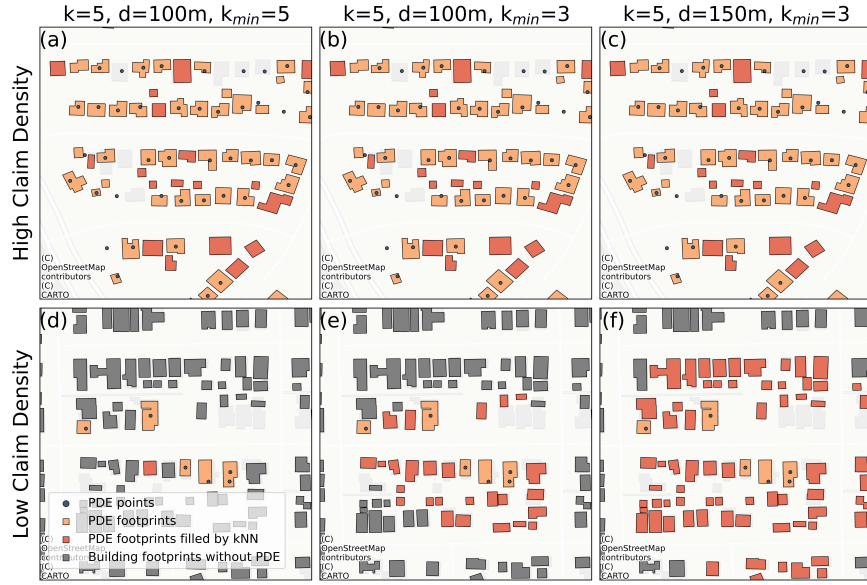


Figure 4: 在不同的洪水索赔密度和 kNN 参数设置下，对 PDE 的 kNN 插补结果进行视觉比较。上排 (a–c) 显示的是在高索赔密度区域的结果，而下排 (d–f) 对应于低索赔密度区域。探讨的 kNN 配置为：(a, d)  $k = 5$ 、 $d = 100$  米和  $k_{min} = 5$ ；(b, e)  $k = 5$ 、 $d = 100$  米和  $k_{min} = 3$ ；以及 (c, f)  $k = 5$ 、 $d = 150$  米和  $k_{min} = 3$ 。展示了原始 PDE 点数据、直接分配了 PDE 的建筑底图、通过 kNN 插补的 PDE 建筑底图以及没有 PDE 的建筑底图。

所有多模态输入数据及其对应的多任务标签都被上采样到统一的  $1280 \times 1280$  像素分辨率，以确保不同数据来源之间的分辨率一致性。在这些模型的训练阶段， $512 \times 512$  像素的图块会从这些输入数据和标签中随机裁剪。为了增强模型的鲁棒性和泛化能力，训练数据上应用了多种数据增强技术，包括随机旋转、水平翻转和垂直翻转。输入数据也进行了归一化。对于基线对比模型，由于这些模型使用单模态输入，因此不需要初始的上采样步骤。对这些模型的处理开始于输入数据原始的  $512 \times 512$  像素分辨率，并从中随机裁剪  $256 \times 256$  像素的图块用于训练。

使用 AdamW 优化器，Flood-DamageSense 网络以及用于基线比较和消融研究的其他网络进行了优化。初始学习率设为  $1 \times 10^{-4}$ ，权重衰减设为  $5 \times 10^{-3}$ 。在复合损失函数中，每个任务的交叉熵损失成分权重设为 1.0；建筑物定位和洪水水域映射任务的 Lovasz-Softmax 损失成分权重设为 0.5，类似于 ChangeMamba (Chen et al., 2024) 中的策略加权方法。将建筑物损伤分类作为 Flood-DamageSense 的主要目标，其 Lovasz-Softmax 损失成分的权重被赋予了更高的值 0.75。所有模型的训练有效批量大小为 16。由于某些实验设置中的 GPU 内存限制，对某些网络的实际批量大小使用 2，并补充 8 步梯度累积，或使用批量大小 8 并进行 2 步梯度累积。训练总共进行了 200 个 epoch。在测试和推理阶段，Flood-DamageSense 模型以其原始采样分辨率  $1280 \times 1280$  像素处理输入数据以生成预测。

### 3.2 基线比较

为了为提出的 Flood-DamageSense 模型建立一个稳健的基础，对 UrbanSARFloods 数据集中的哈佛飓风子集进行了全面的基准比较。该评估考虑了来自三个主要深度学习范式的代表性模型：基于卷积神经网络的方法，Siamese-UNet (Durnov, 2020)；两个基于 Transformer 的方法，BIT (Chen et al., 2022) 和 ChangeFormer (Bandara & Patel, 2022)；以及一个基于 Mamba 的方法，ChangeMamba (Chen et al., 2024)。输入数据为事件前后的 SAR 和 InSAR 影像。为了确保适合该研究的目标，这些基准模型的解码器架构被修改并适应于多任务学习框架。经过适应的这些基准模型的定量性能在表 1 中展示。

评估建筑损坏被证明是大部分基线模型的一项艰巨任务，特别是对于细微的损坏类别而言。大多数 Siamese-UNet、BIT 和 ChangeFormer 变体在轻微损坏和中度损坏类别上记录的 F1 分数为 0。尽管对于无损坏类别的 F1 分数普遍较高，但对于重大损坏类别的 F1 分数差异较大。ChangeMamba (0.3132)、BIT-34 (0.2396) 和大多数 ChangeFormer 版本（例如，V1: 0.2581, V3: 0.2525, V4: 0.2492）展示了在检测重大损坏方面的能力。重要的是，ChangeMamba 在轻微损坏 (0.0012) 和中度损坏 (0.0468) 类别中成功实现了非零虽然不高的 F1 分数。因此，其损坏评估的调和平均 F1 分数 (0.0034) 比其他在中间类别中经常完全失败而导致登记为零的模型更能说明其在多种损坏状态中的能力。在洪水水位测绘任务中，几个 ChangeFormer 变

Table 1: 对 UrbanSARFloods 数据集中哈维飓风子集的基线模型的定量结果进行了分析，该数据集包含额外的建筑物掩膜和损坏标签。在建筑物损坏分类中，对 F1 分数的调和平均值排除了无损坏类，而在洪水水域绘制中则排除了未被淹没的类。

| Model                                     | Batch Size | Building Localization F1 Score | Building Damage Assessment F1 Score |              |               |              |               | Floodwater Mapping F1 Score |            |               |
|-------------------------------------------|------------|--------------------------------|-------------------------------------|--------------|---------------|--------------|---------------|-----------------------------|------------|---------------|
|                                           |            |                                | No Damage                           | Minor Damage | Medium Damage | Major Damage | Harmonic Mean | Open Area                   | Urban Area | Harmonic Mean |
| Siamese-Unet-ResNet-34 (Durnov, 2020)     | 16         | 0.0031                         | 0.9746                              | 0            | 0             | 0.0035       | 0             | 0.3390                      | 0.0002     | 0.0003        |
| Siamese-Unet-SE-ResNext-50 (Durnov, 2020) | 16         | 0.0048                         | 0.9758                              | 0            | 0             | 0.0058       | 0             | 0.3674                      | 0          | 0             |
| Siamese-Unet-DPN-92 (Durnov, 2020)        | 16         | 0.0692                         | 0.9763                              | 0            | 0             | 0            | 0             | 0.3861                      | 0          | 0             |
| Siamese-Unet-SENet-154 (Durnov, 2020)     | 16         | 0.4257                         | 0.9711                              | 0            | 0             | 0.1868       | 0             | 0.4177                      | 0.0009     | 0.0017        |
| BIT-18 (Chen et al., 2022)                | 16         | \\                             | 0.9692                              | 0            | 0             | 0.1758       | 0             | 0.8145                      | 0.0058     | 0.0116        |
| BIT-34 (Chen et al., 2022)                | 16         | \\                             | 0.9711                              | 0            | 0             | 0.2396       | 0             | 0.8332                      | 0.2213     | 0.3498        |
| BIT-50 (Chen et al., 2022)                | 16         | \\                             | 0.9708                              | 0            | 0.0098        | 0.0978       | 0             | 0.5464                      | 0.0057     | 0.0113        |
| ChangeFormerV1 (Bandara & Patel, 2022)    | 16         | \\                             | 0.9713                              | 0            | 0.0003        | 0.2581       | 0             | 0.8656                      | 0.2161     | 0.3458        |
| ChangeFormerV2 (Bandara & Patel, 2022)    | 16         | \\                             | 0.9719                              | 0            | 0             | 0.2368       | 0             | 0.8606                      | 0.1526     | 0.2592        |
| ChangeFormerV3 (Bandara & Patel, 2022)    | 16         | \\                             | 0.9724                              | 0            | 0             | 0.2525       | 0             | 0.8630                      | 0.1326     | 0.2299        |
| ChangeFormerV4 (Bandara & Patel, 2022)    | 16*        | \\                             | 0.9735                              | 0            | 0             | 0.2492       | 0             | 0.8816                      | 0.3253     | 0.4753        |
| ChangeFormerV5 (Bandara & Patel, 2022)    | 16*        | \\                             | 0.9669                              | 0            | 0             | 0.1062       | 0             | 0.8790                      | 0.2974     | 0.4444        |
| ChangeFormerV6 (Bandara & Patel, 2022)    | 16*        | \\                             | 0.9722                              | 0            | 0             | 0.2317       | 0             | 0.8780                      | 0.3688     | 0.5194        |
| ChangeMamba (Chen et al., 2024)           | 16         | 0.6319                         | 0.9738                              | 0.0012       | 0.0468        | 0.3132       | 0.0034        | 0.7998                      | 0.2641     | 0.3971        |

\*Effective batch size is 16 and actual batch size is 8 with 2 accumulation steps due to limited available GPU memory.

体、ChangeMamba 和 BIT-34 表现出了竞争力。ChangeFormerV4 在开放区域洪水上取得了最高的 F1 分数 (0.8816)，而 ChangeFormerV6 在城市地区洪水测绘上表现卓越 (0.3688)，这导致该任务的最高调和平均数 (0.5194)。建筑定位的 F1 分数揭示了大多数基线模型的显著挑战。大多数 Siamese-UNet 变体表现出非常低的 F1 分数，范围从 0.003 到 0.4257。ChangeMamba 在此任务中表现最佳，取得了 0.6319 的建筑定位 F1 分数。建筑定位任务未在 BIT 和 ChangeFormer 中进行，因为这些模型是纯粹的变化检测模型，没有语义解码器。

比较分析表明，虽然某些模型，如特定的 ChangeFormer 变种，在洪水映射等辅助任务中表现出色，ChangeMamba 在建筑损坏分类中取得了最佳表现，这是我们的主要任务。此外，ChangeMamba 展示了最有前景的整体平衡，特别是在能够检测到所有级别的建筑损坏方面，而其他基线模型大多未能做到。然而，各个基线模型在轻度 and 中度损坏类别上普遍低的 F1 分数突显了现有架构在这些任务上的固有不适宜性。这凸显了一个重要的改进领域，并为开发 Flood-DamageSense 模型提供了关键动机，该模型旨在提高建筑洪水损害评估这一具有挑战性的方面。在建筑定位的结果方面，与现有文献中的发现一致，即 SAR 影像的粗略空间分辨率不足以进行详细的建筑层级评估，因此强调建筑底图数据在细化建筑层级的损害结果中的必要性。

### 3.3 实验结果

本节展示了所提出的 Flood-DamageSense 模型在哈维飓风数据集上的定量性能。评估重点在于其在不同粒度水平上的建筑损坏评估效果，以及其在建筑定位和洪水水域绘制的辅助任务中的表现。

在建筑损害评估中，Flood-DamageSense 模型的性能，采用结合所有多模态输入的配置，对应于表 3 中的配置 4，详见表 2。在像素级别，该模型在无损害类别上达到了 0.9702 的 F1 分数，表明在识别无损害属性方面表现强劲。对于实际损害类别，轻微损害、中等损害和重大损害的 F1 分数分别是 0.1109、0.2100 和 0.3818。在像素级别，小损害、中损害和大损害类别 F1 分数的调和平均值是 0.1829。当将输出聚合到建筑级别时，对于大多数损害类别，标准 F1 分数有所提高，分别是无损害 0.9473、轻微损害 0.1255、中等损害 0.2487，重大损害显著为 0.4610。在建筑级别，损害类别标准 F1 分数的调和平均值是 0.2629。这表明基于足迹内中值像素值的建筑级别聚合有助于改进分类。为了考虑损害评估的有序性，计算了建筑级别相邻 F1 分数，在真实标签的  $\pm 1$  类别内对预测进行评分。这一指标的 F1 分数分别为无损害 0.9840、轻微损害 0.9890、中等损害 0.4951 和重大损害 0.7545。使用此相邻标准的损害类别的调和平均值是 0.7446。尽管 F1 分数在考虑相邻因素后显著提高，但双侧相邻考虑可能会引起轻微损害类别的关注，因为在损害评估中，将轻微损害分类为无损害通常是不理想的。此外，在建筑层面考虑上邻近类别，将真实类别或下一个更高损害类别的预测给予信用，结果在 F1 分数中，无损为 0.9522，轻微损害为 0.2387，中等损害为 0.3923，显著的重大损害为 0.5999。包括上邻近类别的损害类别的谐调平均值提高至 0.4230。与标准 F1 分数相比，该指标的显著增加是一个积极信号，因为在真实世界的灾害响应和资源分配中，轻微高估损害往往比低估更为可取，以确保潜在的关键损害不会被忽视。

为了补充这些定量评估，图 5 展示了模型生成的建筑物损坏等级和洪水水域制图的定性示例。研究结果将在第 3.3.1 节中与洪水水域制图的结果一起讨论。图 6 进一步展示了所提框架在建筑损坏评估方面的能力。图 6 (a) 提供了像素级和汇总建筑级建筑损坏输出的详细比较，图像上证明了相应于表 2 的建筑级处理策略优势。图 6 (b) 展示了本分析的实际成果。由 Flood-DamageSense 为哈里斯县生成的综合建筑级洪水损坏

Table 2: 不同等级建筑损毁评估的 F1 分数。F1 分数的调和平均数不包括无损坏类。

| Metric                           | No Damage | Minor Damage | Medium Damage | Major Damage | Harmonic Mean |
|----------------------------------|-----------|--------------|---------------|--------------|---------------|
| Pixel-level F1                   | 0.9702    | 0.1109       | 0.2100        | 0.3818       | 0.1829        |
| Building-level F1                | 0.9473    | 0.1255       | 0.2487        | 0.4610       | 0.2629        |
| Building-level Adjacent F1       | 0.9840    | 0.9890       | 0.4951        | 0.7545       | 0.7446        |
| Building-level Upper Adjacent F1 | 0.9522    | 0.2387       | 0.3923        | 0.5999       | 0.4230        |

地图展示了该框架在大规模、可操作损坏评估中的潜力。为了更好地可视化地图生成，图 6（b）中展示了训练和测试数据。

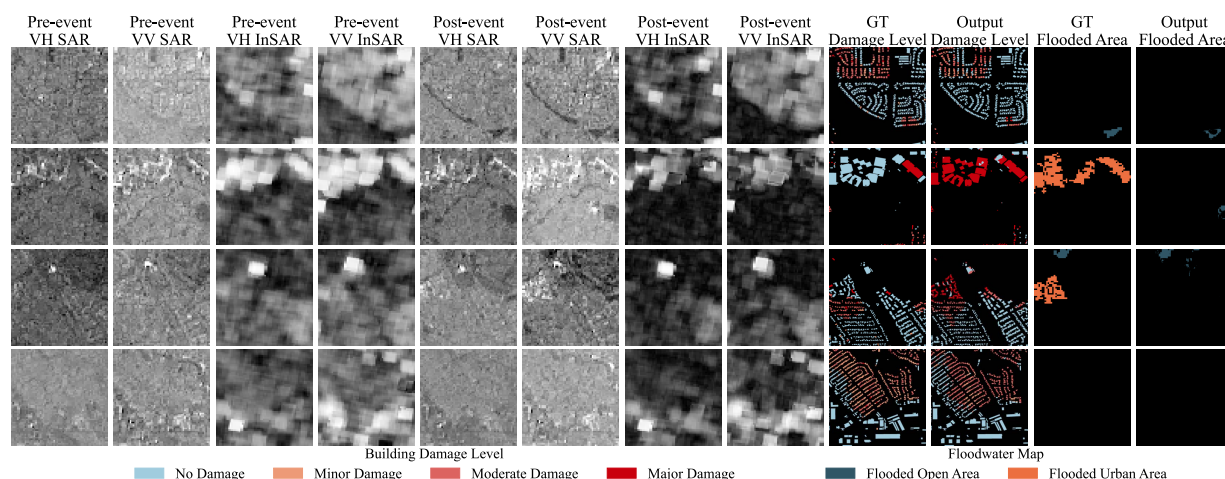


Figure 5: 用于建筑损害评估和洪水水位制图的 SAR 和 InSAR 输入示例及其对应的输出。列展示了事件前后的 SAR 和 InSAR 图像 (VH 和 VV 极化)，接着是建筑洪水损害等级的地面实况 (GT) 和模型输出，分类为无损害、轻微、中度和严重损害，以及洪水地图，区分被淹的开放区域和被淹的城市区域。需要注意的是，为了可视化建筑洪水损害输出，使用从建筑轮廓数据集中提取的地面实况建筑掩膜来描绘建筑位置，而不是使用模型的辅助定位任务生成的掩膜，以保持与建筑损害评估的 F1 分数计算方法的一致性，该方法在已知建筑范围内评估像素级分类，而且该框架的最终建筑损害地图产品依赖于将损害输出与现有的、具地理参考的建筑轮廓数据集成，而非同时产生的建筑掩膜。

### 3.3.1 辅助任务上的表现

Flood-DamageSense 模型实现了 0.3855 的建筑物定位 F1 得分。该性能低于关于 ChangeMamba 模型在以单模态输入运行时的 F1 得分，具体见表 1。造成这种差异的一个可能原因是用于评价的建筑物掩码标签在各自实验设置中分辨率的差异。具体来说，虽然这些实验中 SAR 和 InSAR 输入影像的分辨率保持一致，本研究用于评估 Flood-DamageSense 的建筑物掩码标签的分辨率比用于基线比较的 ChangeMamba 高出了 20 倍。本研究使用的建筑物掩码标签来自建筑物轮廓数据集，该数据集由矢量数据组成。矢量数据的一个主要特征是其可扩展性，可以被栅格化为任何所需分辨率。因此，在本研究中为了不同的实验目的，建筑物轮廓在各种分辨率下进行了栅格化，以适当与每个特定配置中输入数据的最高有效分辨率对齐。因此，在单模态基线比较中，ChangeMamba 模型的任务是勾勒出相对粗略的轮廓，可能包括成组的建筑物。相比之下，Flood-DamageSense 模型在此实验中被评估其实现更精确地划分单个建筑物边界的能力，这由于地面实况的更高细粒度而成为一项更具挑战性的定位任务。此外，尽管 Flood-DamageSense 中引入了事件前 VHR 光学影像用于建筑物定位，但由于研究区域内其可用性的限制，意味着它只能作为辅助数据，而非用于建筑物定位的一种始终可用的主要输入。

对于洪水淹没区域的制图，模型在淹没的开放区域获得了 0.8893 的 F1 分数。然而，淹没的城市区域的 F1 分数为 0，导致整体洪水淹没制图任务的调和平均值为 0。这表明虽然模型在检测淹没的开放区域上表现良好，但在城市洪水检测上存在困难。除了定量指标之外，图 5 提供了 Flood-DamageSense 模型表现的定性说明。图中显示了地面真实数据与模型输出的并列比较，涵盖了建筑物洪水损害等级和洪水边界的划分。这些输出的视觉检查揭示了一项值得注意的观察。尽管洪水淹没制图解码器未能检测到被淹没的城市区域，但这些城市区域的建筑物损害等级输出展示了反映淹没存在的分类。然而令人惊讶的是，这些区域内建筑物对应的地面真实 PDE 数据表明没有记录的经济损失。这一观察突显了一个重要的发现，即基于索赔的 PDE 数据所捕捉的经济损失或损害程度可能并不总是直接对应于可视的洪水存在。

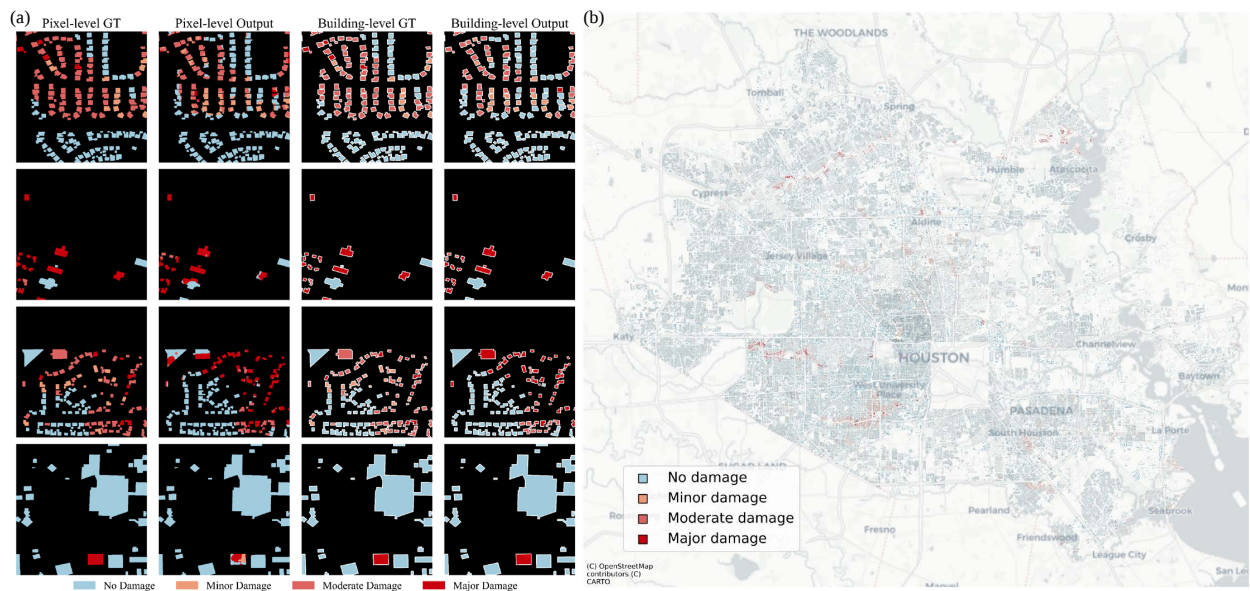


Figure 6: 建筑级洪水损害输出的展示。(a) 详细比较了像素和建筑级别的建筑洪水损害输出。(b) 显示了由 Flood-DamageSense 模型为哈里斯县生成的综合建筑级洪水损害地图。

为了系统地评估不同输入模式的贡献及其在 Flood-DamageSense 多任务学习框架中对并行任务的具体应用，进行了一系列消融研究。分析了在表 3 中详细列出的五种不同模型配置的性能。这些配置在如下三个并行任务中的应用中，利用了不同的前事件 SAR/InSAR、后事件 SAR/InSAR、前事件 VHR 光学影像和历史洪水风险数据：建筑物损害评估 (BDA)、洪水水域映射 (FM) 以及建筑物定位 (LOC)。

Table 3: 评估多任务输出和多模态输入对建筑定位、损坏评估和水淹映射性能的影响。左半部分列出了在每个模型的每个任务中使用了哪些输入。中间一栏报告了训练过程中输入数据的批量大小。右半部分显示了每个任务的模型性能。建筑损坏分类的 F1 分数的调和平均值排除了无损坏类别，水淹映射的 F1 分数的调和平均值排除了非淹没类别。

| No. | Pre-event SAR/InSAR |     |    | Post-event SAR/InSAR |     |    | Pre-event VHR |     |    | Flood Risk |     |    | Batch Size | Building Localization (LOC) F1 | Building Damage Assessment (BDA) F1 |              |               |              |               | Floodwater Mapping (FM) F1 |            |               |
|-----|---------------------|-----|----|----------------------|-----|----|---------------|-----|----|------------|-----|----|------------|--------------------------------|-------------------------------------|--------------|---------------|--------------|---------------|----------------------------|------------|---------------|
|     | LOC                 | BDA | FM | LOC                  | BDA | FM | LOC           | BDA | FM | LOC        | BDA | FM |            |                                | No Damage                           | Minor Damage | Medium Damage | Major Damage | Harmonic Mean | Open Area                  | Urban Area | Harmonic Mean |
| 0   | Y                   | Y   |    |                      | Y   |    |               |     |    |            |     |    | 16         | 0.5750                         | 0.9702                              | 0            | 0.0121        | 0.3061       | 0             | 0.7998                     | 0.2641     | 0.3971        |
| 1   | Y                   | Y   | Y  |                      | Y   | Y  |               |     |    |            |     |    | 16         | 0.6319                         | 0.9738                              | 0.0012       | 0.0468        | 0.2312       | 0.0034        |                            |            |               |
| 2   | (Y)*                | Y   | Y  |                      | Y   | Y  | Y             |     |    |            | Y   |    | 16*        | 0.0395                         | 0.9737                              | 0.1135       | 0.1198        | 0.3196       | 0.1479        | 0.7772                     | 0.1207     | 0.2090        |
| 3   | Y                   | Y   | Y  |                      | Y   | Y  |               |     |    |            | Y   |    | 16*        | 0.2759                         | 0.9764                              | 0.0909       | 0.1487        | 0.3302       | 0.1446        | 0.7969                     | 0.0652     | 0.1840        |
| 4   | Y                   | Y   | Y  |                      | Y   | Y  | Y             | Y   |    |            | Y   |    | 16*        | 0.3855                         | 0.9702                              | 0.1109       | 0.2100        | 0.3818       | 0.1829        | 0.8893                     | 0          | 0             |

\*Effective batch size is 16 and actual batch size is 2 with 8 accumulation steps due to limited available GPU memory.

\*(Y) indicates pre-event SAR/InSAR data were used only when pre-event VHR optical imagery was not available.

模型 0 是使用单模态输入的原始 ChangeMamba 模型，作为以 SAR 为中心的基准。模型 0 利用事件前后的 SAR/InSAR 数据进行建筑损坏评估，仅使用事件前的 SAR/InSAR 数据进行建筑定位。该配置不执行洪水映射任务。模型 1 在模型 0 的基础上扩展用于多任务学习，使用事件前后的 SAR/InSAR 数据进行建筑损坏评估和洪水映射，同时使用事件前的 SAR/InSAR 数据进行建筑定位。模型 1 在表格 1 中展示。模型 2 通过纳入历史洪水风险数据进行建筑损坏评估引入多模态输入。对于建筑定位，主要使用事件前的高分辨率 (VHR) 光学影像，当 VHR 图像不可用时，以事件前的 SAR/InSAR 数据为后备。模型 3 类似于模型 2，但仅依赖事件前的 SAR/InSAR 数据进行建筑定位，同时仍使用 SAR/InSAR 结合历史洪水风险数据进行建筑损坏评估。模型 4 是完全提出的 Flood-DamageSense 模型，利用所有输入模式进行建筑损坏评估。对于洪水映射，使用事件前后的 SAR/InSAR 数据。对于建筑定位，使用事件前的 SAR/InSAR 和事件前的高分辨率光学影像。建筑损坏评估这一主要任务从多重数据源的整合中显著受益。模型 0 在无损伤和重大损伤方面取得了较高的 F1 分数，但对于轻微损伤得分为 0，在中等损伤方面得分很低，导致损伤类别的调和平均 F1 为 0。模型 1 在无损伤和重大损伤方面表现相似，但在轻微和中等损伤方面稍高，使得调和平均 F1 达到 0.0034。引入洪水水位映射任务与建筑损伤评估相结合，使模型在所有损伤类别上取得非零 F1 分数。模型 2 引入历史洪水风险数据，显著改善了整体建筑损伤评估性能，调和平均 F1 提高到 0.1479。这主要得益于轻微损伤和中等损伤 F1 分数的显著提升。模型 4，即完整的提议模型，进一步结合了用于建筑损伤评估的事件前 VHR 影像，在损伤类别上实现了最高的调和平均 F1 分数 0.1829。它在所有损伤类别上表现最强，尽管无损伤的 F1 分数略低于模型 3。洪水水位映射的输入在所有模型中保持一致。然而，随着建筑损伤评估性能的提高，城市淹没区域映射的性能下降。模型 4 虽然在淹没的开放区域上拥有最高的 F1 分数，但完全未能检测到淹没的城市区域，导致调和平均 F1 为 0。随着建筑损伤评估性能提高，城市洪水映射性能下降，

表明解码器之间可能存在潜在的相互作用或资源争用。据推测,在更复杂的配置中,建筑损伤解码器可能学习到与洪水水位映射解码器检测城市洪水所需线索重叠的特征,这一发现与图 5 的定性观察结果一致。在建筑物定位任务上的表现强调了输入模态选择和多模态融合架构的重要性。模型 2 主要利用事件前的 VHR 影像进行定位,当 VHR 不可用时则使用 SAR/InSAR 作为备选,其 F1 分数大幅下降至 0.0395。这表明在该特定配置中,有效利用条件性可用的 VHR 数据或将其与 SAR 特征融合方面存在挑战。在此模型配置中,建筑物定位解码器接收到的特征交替来源于 VHR 光学影像或者在 VHR 数据不可用时来源于 SAR 影像。通过一个并未专门设计用于自适应多模态融合的通用解码器处理可能存在差异的特征类型可能会对达到最佳性能造成挑战。模型 3 使用事件前 SAR/InSAR 数据进行定位,获得了 0.2759 的 F1 分数。模型 4,即提出的完整模型,使用事件前的 SAR/InSAR 和 VHR 影像,并在解码器中采用专门的特征融合机制进行定位,获得了 0.3855 的 F1 分数。在 3.3.1 节中对模型 2-4 相对模型 0 和 1 的较低 F1 分数的解释进行了说明。

## 4 讨论

比较实验强调了多模态、多任务学习在快速洪水损害情报中的潜力和仍然存在的障碍。消融测试显示,单独使用 SAR/InSAR 很难区分轻微和中等损失;然而,结合内在风险优先信息可以使这些类别的性能提高一倍,引入事前 VHR 图像则通过增强中等和重度损害预测的精确性来提高整体准确性。收益证实了 SAR 的全天候覆盖必须与上下文信息——历史暴露、土地用途和精细建筑纹理——互为补充,以捕捉淹没的低对比度特征。将噪声像素分数转换为每个区域单一的中位损害标签再次提高了谐波平均 F1,并实现了 0.60 的上邻 F1,这确保了严重受影响的结构很少被误排等级——这一结果对于分类和资源分配至关重要。辅助目标揭示了更微妙的动态。当专门的特征融合状态空间模块结合 SAR 和 VHR 流进行建筑定位时,准确性最高;在缺少 VHR 的情况下有条件地回退到 SAR 会降低性能;而由于评价掩膜详细程度大幅增加,该任务比仅使用光学基线要困难得多。洪水水体映射则显示出相反的模式:尽管洪水水体映射解码器架构保持不变,当引入额外模态和更复杂的解码器以改善损害分级时,城市淹没检测恶化,尽管开放区域的性能仍然强劲。目视检查表明,损害解码器垄断了都市水体线索,指向共享编码器内部的表示竞争——这是多任务网络的已知风险,需要在架构或损失平衡方面进行改进。定性示例还揭示了可见水和基于索赔的财产损失程度标签之间的系统性差距:一些城市街区看起来被淹没,但却没有记录到经济损失,这可能是由于保险覆盖率或损失未达到申报门槛。虽然这种不匹配可能会压低相对于纯视觉地面实况的 F1 分数,但财产损失程度 (PDE) 仍然是评估事件后恢复成本最具可扩展性和最切实相关的代理;因此,其使用将模型与应急管理者的优先事项保持一致,尽管这会以较嘈杂的训练信号为代价。

三方面的限制需要在未来工作中加以解决。首先,索赔依赖性可能会低估未投保或投保不足建筑物的损失;通过补充 PDE 以群众外包或政府检查数据可以减少偏差。其次,稀疏的 VHR 覆盖限制了光学线索的能力;在全球档案上轻量级的自监督预训练或基于扩散的上采样可以增加有效覆盖范围。第三,不捕捉复杂的城市洪水特征表明需要特定于 SAR 的物理先验或更高分辨率的干涉堆栈。解决这些问题,并完善解码器之间的损失共享,应该能够使 Flood-DamageSense 在不同的水文环境中进行泛化,并在所有子任务中提供一致的高性能。总体而言,本研究表明,通过 Mamba 骨干融合互补模态,使用经济上有意义的标签进行监督,并在建筑物底图上汇总预测,可以在图像获取后的数小时内生成可操作的洪水损害地图——标志着向操作级、结构规模的灾害分析迈出了实质性一步。

## 5 总结论

Flood-DamageSense 在远程感知分析领域做出了三项基础性进展。方法上,它首次提供了一种特定于洪水的多层次建筑损坏模型,该模型 (i) 利用线性复杂度的 Mamba 序列骨架捕捉长距离空间依赖性,而无需注意力机制的计算负担,(ii) 结合了事件前后 SAR-InSAR,事件前的 VHR 光学图像,以及一种固有的洪水风险先验,整体嵌入于半对称视觉状态空间编码器和特征融合解码器中,(iii) 在单一多任务体制下学习三项紧密关联的任务——损害分级、洪水范围和建筑物定位。此设计克服了早期研究阻碍的双重障碍:云引起的数据缺口和水损伤微妙的光谱特征,这阻碍了通用变化检测管道。从实证上看,飓风哈维的案例研究确认了结合风险先验和 SAR 数据可在统计上获得显著的提升,特别是在“轻微”和“中等”损坏类别中,这些类别推动了多数恢复成本决策,但在 xBD 时代的网络中经常被错误分类。消融结果进一步揭示了在多任务框架中每种模式如何贡献——有时竞争——为未来架构调优提供了蓝图。

从实际的角度来看,该研究提供了一整套将像素级概率转化为可供检查的、地理参考的矢量层的图像到地图工作流程,使用中位数聚合和质量标志方案。这条流水线可以接收来自任何像素级模型的输出,使得各机构在保留地图生成主线的同时接入其他分类器。由于唯一的事后输入是免费的 SAR 场景,该系统可以在洪峰后数小时内运行,提供场地级别的损害情报,用于分诊调度、索赔优先级排序和恢复预算估算——这些能力远远超出了之前洪水研究中的二元图像块产品。

尽管有这些进步，本研究承认存在一些局限性，例如基于索赔的 PDE 基准数据的固有属性、VHR 光学数据的获取限制、在健全的城市洪水制图中的挑战，以及广泛测试以实现普适性的需要。未来的研究应旨在通过探索增强的多模态融合技术、结合替代或补充的基准数据源，以及在更广泛的洪水事件和多样的地理环境中评估模型性能来解决这些限制。进一步探索优化多任务解码器之间的相互作用也是必要的，以确保在所有并发目标下的稳健性能。Flood-DamageSense 代表了朝着更可靠和快速的建筑物洪水损害评估迈出的重要一步，为灾后响应和恢复工作提供了有价值的支持。

6

数据可用性 UrbanSARFloods 数据集可在 <https://github.com/jie666-6/UrbanSARFloods> 公开获取。其他使用的数据是保密的。

本研究中所提出模型的源代码可以在 <https://github.com/violayhho/Flood-DamageSense> 找到。

7

致谢 作者们要感谢德克萨斯 AM 大学（加尔维斯顿）的 Samuel D. Brody 博士提供了宝贵的历史洪水风险数据，以及我们城市恢复力实验室成员刘嘉福博士（德克萨斯 AM 大学）提供的用于模型开发的财产损失范围数据。作者们还要感谢国家自然科学基金会通过 CRISP 2.0 类型 2，资助编号 1832662，以及德克萨斯 AM 大学 X-Grant 699 的资助支持。本文中所表达的任何观点、发现、结论或建议均为作者个人观点，并不一定反映资助机构的意见。

## References

- Amitrano, D., Di Martino, G., Di Simone, A., & Imperatore, P. (2024). Flood detection with SAR: A review of techniques and datasets. *Remote Sensing*, 16(4). doi:10.3390/rs16040656
- Apollo Mapping. (2017). Apollo Mapping | The Image Hunters. <https://apollomapping.com/>.
- Bandara, W. G. C., & Patel, V. M. (2022). A transformer-based siamese network for change detection. In *Igarss 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium* (p. 207-210). doi:10.1109/IGARSS46834.2022.9883686
- Berman, M., Triki, A. R., & Blaschko, M. B. (2018). The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4413-4421).
- Braik, A. M., & Koliou, M. (2024). Automated building damage assessment and large-scale mapping by integrating satellite imagery, GIS, and deep learning. *Computer-Aided Civil and Infrastructure Engineering*, 39(15), 2389-2404.
- Cao, Q. D., & Choe, Y. (2020). Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks. *Natural Hazards*, 103(3), 3357-3376.
- Chen, H., Qi, Z., & Shi, Z. (2022). Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-14. doi:10.1109/TGRS.2021.3095166
- Chen, H., Song, J., Han, C., Xia, J., & Yokoya, N. (2024). ChangeMamba: Remote sensing change detection with spatiotemporal state space model. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-20. doi:10.1109/TGRS.2024.3417253
- Cheng, C.-S., Behzadan, A. H., & Noshadravan, A. (2021). Deep learning for post-hurricane aerial damage assessment of buildings. *Computer-Aided Civil and Infrastructure Engineering*, 36(6), 695-710.
- Dong, S., Yu, T., Farahmand, H., & Mostafavi, A. (2020). Bayesian modeling of flood control networks for failure cascade characterization and vulnerability assessment. *Computer-Aided Civil and Infrastructure Engineering*, 35(7), 668-684.
- Dong, S., Yu, T., Farahmand, H., & Mostafavi, A. (2021). A hybrid deep learning model for predictive flood warning and situation awareness using channel network sensors data. *Computer-Aided Civil and Infrastructure Engineering*, 36(4), 402-420.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929 [cs.CV]*.
- Durnov, V. (2020). xview2 1st place solution. [https://github.com/vdurnov/xview2\\_1st\\_place\\_solution](https://github.com/vdurnov/xview2_1st_place_solution).

- Esparza, M., Ho, Y.-H., Brody, S., & Mostafavi, A. (2025). Improving flood damage estimation by integrating property elevation data. *International Journal of Disaster Risk Reduction*, 105251.
- Google. (2023). Open Buildings. <https://sites.research.google/gr/open-buildings/>.
- Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.
- Gu, D., Shuai, Q., Zhang, N., Jin, N., Zheng, Z., Xu, Z., & Xu, Y. (2025). Multi-view street view image fusion for city-scale assessment of wind damage to building clusters. *Computer-Aided Civil and Infrastructure Engineering*, 40(2), 198–214.
- Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeed, S., Heim, E., Doshi, J., Lucas, K., Choset, H., & Gaston, M. (2019). Creating xBD: A dataset for assessing building damage from satellite imagery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops*.
- Han, B., Zhang, S., Shi, X., & Reichstein, M. (2024). Bridging remote sensors with multisensor geospatial foundation models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 27852–27862).
- Ho, Y.-H., Lee, C.-C., Diaz, N., Brody, S., & Mostafavi, A. (2024). Elev-vision: Automated lowest floor elevation estimation from segmenting street view images. *ACM Journal on Computing and Sustainable Societies*, 2(2), 1–18.
- Ho, Y.-H., Li, L., & Mostafavi, A. (2025). Integrated vision language and foundation model for automated estimation of building lowest floor elevation. *Computer-Aided Civil and Infrastructure Engineering*, 40(1), 75–90.
- Huang, L., Antolini, F., Mostafavi, A., Blessing, R., Garcia, M., & Brody, S. D. (2025). High-resolution flood probability mapping using generative machine learning with large-scale synthetic precipitation and inundation data. *Computer-Aided Civil and Infrastructure Engineering*.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45.
- Kaur, N., Lee, C.-C., Mostafavi, A., & Mahdavi-Amiri, A. (2023). Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. *Computer-Aided Civil and Infrastructure Engineering*, 38(15), 2072–2091.
- Khajwal, A. B., Cheng, C.-S., & Noshadran, A. (2023). Post-disaster damage classification based on deep multi-view image fusion. *Computer-Aided Civil and Infrastructure Engineering*, 38(4), 528–544.
- Liu, C.-F., Huang, L., Yin, K., Brody, S., & Mostafavi, A. (2024). FloodDamageCast: Building flood damage nowcasting with machine-learning and data augmentation. *International Journal of Disaster Risk Reduction*, 114, 104971. doi:<https://doi.org/10.1016/j.ijdr.2024.104971>
- Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., & Liu, Y. (2024). Vmamba: Visual state space model. In A. Globerson et al. (Eds.), *Advances in neural information processing systems* (Vol. 37, pp. 103031–103063). Curran Associates, Inc. Retrieved from [https://proceedings.neurips.cc/paper\\_files/paper/2024/file/baa2da9ae4bfed26520bb61d259a3653-Paper-C](https://proceedings.neurips.cc/paper_files/paper/2024/file/baa2da9ae4bfed26520bb61d259a3653-Paper-C)
- Microsoft. (2018). USBuildingFootprints. <https://github.com/microsoft/USBuildingFootprints>.
- Microsoft. (2022). GlobalMLBuildingFootprints. <https://github.com/microsoft/GlobalMLBuildingFootprints>.
- Mirzaei, G., & Adeli, H. (2018). Segmentation and clustering in brain mri imaging. *Reviews in the Neurosciences*, 30(1), 31–44.
- Mobley, W., Sebastian, A., Blessing, R., Highfield, W. E., Stearns, L., & Brody, S. D. (2021). Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: a pilot study in southeast texas. *Natural Hazards and Earth System Sciences*, 21(2), 807–822.
- Nofal, O., Rosenheim, N., Kameshwar, S., Patil, J., Zhou, X., van de Lindt, J. W., Duenas-Osorio, L., Cha, E. J., Endrami, A., Sutley, E., Cutler, H., Lu, T., Wang, C., & Jeon, H. (2024). Community-level post-hazard functionality methodology for buildings exposed to floods. *Computer-Aided Civil and Infrastructure Engineering*, 39(8), 1099–1122.
- Nogay, H. S., & Adeli, H. (2020). Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging. *Reviews in the Neurosciences*, 31(8), 825–841.
- Perez-Ramirez, C. A., Amezcua-Sanchez, J. P., Valtierra-Rodriguez, M., Adeli, H., Dominguez-Gonzalez, A., & Romero-Troncoso, R. J. (2019). Recurrent neural network model with bayesian training and mutual information for response prediction of large buildings. *Engineering Structures*, 178, 603–615.
- Rafiei, M. H., & Adeli, H. (2017). A novel machine learning-based algorithm to detect damage in high-rise building structures. *The Structural Design of Tall and Special Buildings*, 26(18), e1400.
- Rudner, T. G. J., Rußwurm, M., Fil, J., Pelich, R., Bischke, B., Kopačková, V., & Biličski, P. (2019). Multi3Net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In *Proceedings of the thirty-third aaai conference on artificial intelligence and thirty-first innovative applications of artificial intelligence conference and ninth aaai symposium on*

- educational advances in artificial intelligence. AAAI Press. doi:10.1609/aaai.v33i01.3301702
- Singh, P., Diwakar, M., Shankar, A., Shree, R., & Kumar, M. (2021). A review on SAR image and its despeckling. *Archives of Computational Methods in Engineering*, 28, 4633–4653.
- Sirca Jr, G. F., & Adeli, H. (2018). Infrared thermography for detecting defects in concrete structures. *Journal of Civil Engineering and Management*, 24(7), 508–515.
- Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2), 199–222. doi:<https://doi.org/10.1016/j.eng.2018.11.030>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Wang, C., Liu, Y., Zhang, X., Li, X., Paramygin, V., Sheng, P., Zhao, X., & Xu, S. (2024). Scalable and rapid building damage detection after hurricane ian using causal bayesian networks and InSAR imagery. *International Journal of Disaster Risk Reduction*, 104, 104371.
- Wu, K., Matsuoka, M., & Oshio, H. (2025). Earthquake damage detection and level classification method for wooden houses based on convolutional neural networks and onsite photos. *Computer-Aided Civil and Infrastructure Engineering*, 40(5), 674–694.
- Xiao, Y., Lu, X., & Zhang, H. (2025). Perspective correction and deep learning-based crack detection for concrete structures. *The Structural Design of Tall and Special Buildings*, 34(7), e70029.
- Xiao, Y., & Mostafavi, A. (2025). DamageCAT: A deep learning transformer framework for typology-based post-disaster building damage categorization. *arXiv preprint arXiv:2504.11637*.
- Zhao, J., Xiong, Z., & Zhu, X. X. (2024). UrbanSARFloods: Sentinel-1 SLC-based benchmark dataset for urban and open-area flood mapping. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (cvpr) workshops* (p. 419–429).
- Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., & Wang, X. (2024). Vision Mamba: Efficient visual representation learning with bidirectional state space model. In *Forty-first international conference on machine learning*.