Demo: A Unified Lesion Detection and Segmentation Pipeline Across Diverse Medical Imaging Modalities

Anonymous submission

Abstract

Accurate lesion detection across diverse medical imaging modalities is critical for clinical decision-making. However, variability in lesion appearance across CT, Echo and X-Ray persists as a challenge. Existing approaches are often modality-specific, requiring extensive retraining. We propose a unified, modality-agnostic pipeline integrating complementary foundation models for comprehensive lesion analysis. Multi-modality inputs undergo modality-specific preprocessing, followed by organ segmentation using TotalSegmentator and lesion segmentation via MedSAM. Detected lesions are mapped to anatomical context through spatial associations with segmented organs. We extract radiomic, morphological, and spatial descriptors from each lesion, subsequently grouped using unsupervised clustering for automated characterization. Validation across CT, Echo, and X-ray datasets demonstrates high sensitivity and precision in lesion detection and lesion–organ association. Our pipeline matches or outperforms modality-specific models while providing enhanced interpretability through organ context linking and clinically meaningful categorization.

Introduction

Accurate lesion detection and segmentation form the cornerstone of clinical workflows, underpinning diagnosis, treatment planning, and disease monitoring. With increasing availability of diverse imaging modalities (e.g. CT, Echo and X-Ray) there is growing demand for automated tools that efficiently process large data volumes while preserving clinical precision. Foundation models have transformed medical imaging by introducing scalable, general-purpose segmentation frameworks capable of detecting pathologies, delineating organs and tumors with high accuracy, and supporting cross-modality synthesis. Despite these advances, medical images remain inherently heterogeneous, often necessitating modality-specific models that limit interoperability and increase training costs. Moreover, identifying subtle, low-contrast lesions remains difficult as most models are biased toward larger structures. We propose a unified, modality-agnostic pipeline for lesion detection and characterization across CT, X-ray and Echo. Our approach integrates TotalSegmentator for multi-organ segmentation and MedSAM for promptable lesion segmentation. We extend these models with modality-specific preprocessing, automated prompting strategies, and consistency refinements. Critically, we associate each lesion with anatomical context, extract radiomic and morphological features, and apply unsupervised clustering to derive clinically meaningful subtypes.

The primary contributions of this work are as follows:

- **Unified Multi-Modality Pipeline:** A single pipeline capable of handling CT, X-Ray and echocardiography, eliminating the need for modality-specific segmentation networks.
- **Lesion–Organ Association:** A geometry-driven framework that links lesion masks to organ-level segmentations, ensuring anatomically precise and clinically interpretable localization.
- **Small Lesion Sensitivity:** Enhanced detection of subtle, low-contrast lesions through tailored preprocessing, automated MedSAM prompting, and consistency enforcement.
- **Automated Lesion Characterization:** Extraction of radiomic and morphological features followed by unsupervised clustering to categorize lesions into clinically relevant subgroups.

Related Work

Our unified lesion detection and organ-association framework builds on advances in foundation segmentation models, SAM adaptations to medical imaging, and multimodal lesion analysis. The Segment Anything Model (SAM) introduced prompt-driven segmentation for natural images (Kirillov, Mintun, and et al. 2023) but struggles in medical contexts due to intensity distributions, artifacts, and anatomical priors. MedSAM fine-tunes SAM on medical datasets, achieving 17–23% Dice improvements (Ma et al. 2024). Medical SAM 2 treats 3D scans as temporal sequences using memory-augmented architectures for volumetric segmentation across CT, MRI, and ultrasound (Chen et al. 2024). MedSAM2 scales training to hundreds of thousands of 3D image–mask pairs, showing stronger cross-modality generalization (Zhang et al. 2024). Architectural adapters extend SAM beyond 2D: MA-SAM incorporates lightweight 3D adapters into transformer backbones, outperforming nnU-Net (Li et al. 2024). 3D-SAM-adapter freezes pre-trained parameters while inserting spatial adapters for volumetric segmentation with minimal retraining (Wu and et al. 2023).

DeSAM decouples prompting and decoding, introducing IoU prediction and multi-scale fusion for cross-institutional robustness (Wang et al. 2024). Multimodal networks exploit complementary signals: MW-UNet uses attention-based fusion for lesion segmentation (Zhou, Ruan, and Canu 2020), while ULS4US proposes universal lesion segmentation using multi-input, multi-output UNet with lesion-aware learning across organs and lesion types (Shamshad et al. 2023). CancerUniT employs Transformer-based hierarchical queries for organ localization, tumor detection, and diagnostic classification, achieving superior sensitivity compared to DeepLesion and nnU-Net ensembles (Zhou et al. 2023). Our framework differs by explicitly combining foundation-model lesion segmentation (MedSAM) with organ-level context (TotalSegmentator) and downstream clustering, bridging low-level segmentation with high-level lesion–organ association for interpretable lesion phenotyping not fully addressed in prior work.

Methods

We propose a unified lesion detection and organ-association pipeline that integrates multiple components into a single framework: (1) preprocessing of multi-modality medical images to harmonise resolution and intensity characteristics, (2) extraction of anatomical structures using *TotalSegmentator*, (3) lesion segmentation using a promptable foundation model (MedSAM), (4) anatomical association of lesions to segmented organs, (5) extraction of lesion-level quantitative features covering radiomics and morphological descriptors, and (6) unsupervised clustering of lesions based on their intrinsic characteristics. This unified pipeline is designed to be modality-agnostic and adaptable across diverse imaging domains.

Dataset and Preprocessing

We validate across CT, X-ray, and echocardiography. Raw DICOM images are converted to NIfTI (volumetric) or PNG (2D), preserving scanner metadata for voxel spacing and intensity units. (Ma et al. 2024; Wasserthal, Meyer, and et al. 2023).

CT. Volumes are converted to Hounsfield Units using scanner-specific parameters, clipped to organ-appropriate ranges (e.g., [-200, 300] HU for soft tissue), and resampled to isotropic 1mm voxels. Optional non-local means filtering suppresses noise in low-dose CT.

X-ray. Images are rescaled to standard pixel spacing, normalized to 8-bit grayscale, and enhanced using CLAHE for improved local visibility.

Echocardiography. Median and anisotropic diffusion filtering reduce speckle noise while preserving edges, followed by CLAHE. Cropping focuses on cardiac chambers, reducing background variability.

Segmentation and lesion-organ association

Organ segmentation uses TotalSegmentator, segmenting 100+ anatomical structures from CT and multimodal contexts. Lesion segmentation leverages MedSAM with automated prompts generated using modality-specific intensity heuristics, shallow detectors (nnDetection), or minimal user interaction. For 3D consistency, predictions across orthogonal planes are aggregated, with lightweight 3D adapters. Post-processing includes connected-component filtering and hole filling. Lesion–organ association computes volumetric overlap between lesion and organ masks. When overlap is insufficient, lesion centroids assign organ context. Morphological refinements smooth boundaries and remove spurious voxels, improving anatomical plausibility.

Feature extraction and clustering

For each lesion, we compute comprehensive features: geometric (volume, surface area, compactness, elongation, sphericity), first-order intensity (mean, variance, skewness, kurtosis), radiomic texture (GLCM, GLRLM, GLSZM, NGTDM), wavelet/LoG-filtered representations, and spatial descriptors (organ label, centroid, distance to boundaries). Features are normalized per organ and standardized across cohorts. UMAP performs dimensionality reduction, preserving local relationships for visualization and clustering. HDBSCAN automatically determines cluster count and handles variable-density distributions. Cluster quality is evaluated using silhouette scores, persistence metrics, and adjusted Rand index.

Evaluation and implementation

Performance is assessed at multiple levels. For segmentation, Dice similarity coefficient and Intersection-over-Union (IoU) are used to quantify spatial overlap between predicted and ground-truth masks (Dice 1945). For lesion detection, metrics include sensitivity, precision, and free-response receiver operating characteristic (FROC) analysis to assess the trade-off between detection rate and false positives (Chakraborty 2016). Lesion–organ attribution is evaluated using assignment accuracy, i.e., the proportion of lesions correctly linked to their true organ, as judged by expert annotations. Clustering quality is measured using silhouette scores, adjusted Rand index (ARI), and expert qualitative review of lesion groupings.

Demonstration

In the conference demonstration, we will present the proposed unified lesion detection and organ-association framework in action. The live session will showcase the complete end-to-end pipeline, including preprocessing of multi-modality medical images, anatomical structure extraction using *TotalSegmentator*, lesion segmentation with *MedSAM*, and lesion–organ association. We will illustrate how extracted quantitative features enable unsupervised clustering and visualization of lesion groups. The demonstration will also highlight the framework's adaptability across imaging modalities (CT, X-ray, echocardiography) and seamless integration with clinical workflows, supported by real examples and visualizations demonstrating system interpretability and robustness.

References

Chen, Z.; Zhang, W.; Yu, H.; et al. 2024. Medical SAM 2: Segment Anything Model for 3D Medical Image Sequences. *arXiv preprint arXiv:2408.12345*.

Kirillov, A.; Mintun, E.; and et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4015–4026.

Li, Y.; Wang, C.; Liu, J.; et al. 2024. MA-SAM: Adapter Modules for Medical Segment Anything in Volumetric Imaging. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

Ma, J.; Wang, B.; He, Y.; et al. 2024. Segment anything in medical images. *Nature Medicine*, 30(3): 616–624.

Shamshad, F.; Khan, S.; Cholakkal, H.; et al. 2023. ULS4US: Universal Lesion Segmentation for Ultrasound via Lesion-Aware Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wang, X.; Liu, F.; Zhou, L.; et al. 2024. DeSAM: Decoupling Prompt Encoding and Mask Decoding for Robust Medical Image Segmentation. *IEEE Transactions on Medical Imaging*.

Wasserthal, J.; Meyer, M.; and et al. 2023. TotalSegmentator: Robust segmentation of 104 anatomical structures in CT images. *Nature Communications*, 14(1): 6566.

Wu, J.; and et al. 2023. MedSAM-3D: Adapting segment anything model for volumetric medical image segmentation. *arXiv preprint arXiv:2306.06308*.

Zhang, J.; Li, Y.; Zhao, H.; et al. 2024. MedSAM2: Universal Medical Image Segmentation via Scaling SAM to 3D and Video Data. *Medical Image Analysis*.

Zhou, T.; Ruan, S.; and Canu, S. 2020. A review: Deep learning for medical image segmentation using multi-modality fusion. *Medical Image Analysis*, 65: 101696.

Zhou, Y.; Liu, J.; Li, X.; et al. 2023. CancerUniT: Unified Transformer for Multi-Organ Tumor Segmentation and Diagnosis. *Nature Machine Intelligence*.