Onco-Seg: Adapting Promptable Concept Segmentation for Multi-Modal Medical Imaging

Ashish Makani[1], Anjali Agrawal[2], and Anurag Agrawal[1]

[1]Koita Centre for Digital Health – Ashoka (KCDH-A), Ashoka University, Sonipat, India
[2]Teleradsol Pvt Ltd, India

Abstract

Medical image segmentation remains a critical bottleneck in clinical workflows, from diagnostic radiology to radiation oncology treatment planning. We present **Onco-Seg**, a medical imaging adaptation of Meta's Segment Anything Model 3 (SAM3) that leverages promptable concept segmentation for automated tumor and organ delineation across multiple imaging modalities. Unlike previous SAM adaptations limited to single modalities, Onco-Seg introduces a unified framework supporting CT, MRI, ultrasound, dermoscopy, and endoscopy through modality-specific preprocessing and parameter-efficient fine-tuning with Low-Rank Adaptation (LoRA). We train on 35 datasets comprising over 98,000 cases across 8 imaging modalities using sequential checkpoint chaining on a 4-GPU distributed training infrastructure. Evaluation on 12 benchmark datasets spanning breast, liver, prostate, lung, skin, and gastrointestinal pathologies yields Dice coefficients of 0.752 (breast ultrasound), 0.714 (polyp segmentation), and 0.641 (liver CT). We further present two clinical deployment patterns and release an open-source napari plugin enabling interactive segmentation with text-based prompting and DICOM-RT export for radiation oncology workflows. Code and models are available at https://github.com/inventcures/onco-segment.

Keywords: Medical image segmentation, foundation models, SAM3, deep learning, radiation oncology, multi-modal imaging, LoRA, napari plugin, DICOM-RT

1. Introduction

Medical image segmentation is fundamental to modern clinical practice. Radiologists rely on precise tumor delineation for diagnosis and treatment response assessment. Radiation oncologists require accurate organ-at-risk (OAR) contouring for treatment planning, where segmentation errors can result in inadequate tumor coverage or excessive normal tissue irradiation [4]. Despite decades of algorithmic development, manual segmentation remains the clinical standard, consuming significant physician time and introducing inter-observer variability [5].

The Segment Anything Model (SAM) series has transformed computer vision by demonstrating that large-scale pretraining enables zero-shot segmentation across diverse domains. SAM1 [1] introduced promptable segmentation with points and boxes. SAM2 [2] extended this to video with memory-based tracking. SAM3 [3], released in November 2025, introduced Promptable Concept Segmentation (PCS)—the ability to segment all instances of a concept specified by text or image exemplars.

SAM3's innovations present opportunities for medical imaging:

1. **Text-Based Prompting:** Clinicians specify targets using natural language ("liver tumor," "left parotid gland") rather than manual annotations.
2. **Multi-Instance Segmentation:** SAM3 identifies all lesions matching a concept, which is relevant for metastatic disease assessment.
3. **Presence Token Architecture:** Decouples recognition from localization, enabling discrimination between similar anatomical concepts.
4. **Unified Detector-Tracker Design:** Extends naturally to 3D volumetric data by treating slices as frames.

In this work, we present Onco-Seg, an adaptation of SAM3 for medical imaging. Our contributions include: (1) a unified preprocessing pipeline supporting ten imaging modalities; (2) parameter-efficient fine-tuning using LoRA on SAM3's 848M parameter architecture; (3) training on 35 datasets with over 98,000 cases; (4) evaluation across 12 benchmark datasets; (5) two clinical deployment patterns; and (6) an open-source napari plugin for interactive clinical use.

2. Methods

2.1 Problem Definition

Given a medical image volume $V \in \mathbb{R}^{H \times W \times D}$ and a clinical prompt P (text, bounding box, or point), we seek a segmentation mask $M \in \{0,1\}^{H \times W \times D}$ that accurately delineates target structures. Challenges include: domain gap between medical and natural images, volumetric consistency across slices, extreme class imbalance ($<1\%$ lesion pixels), and multi-modal heterogeneity.

2.2 Onco-Seg Architecture

Onco-Seg builds upon SAM3's three-component architecture sharing a unified 848M-parameter vision backbone:

Vision Encoder: A hierarchical vision transformer (Hiera) processing images at 1008×1008 resolution (72 patches of 14 pixels each).

Detector (DETR-based): For single-image segmentation with presence token prediction.

Tracker (Memory-based): Inherited from SAM2 for propagation across frames/slices.

2.3 Modality-Specific Preprocessing

We implement ten dedicated normalization transforms, each outputting to SAM3's expected range via final normalization $(x-0.5)/0.5 \rightarrow [-1,1]$.

- **CT:** Hounsfield unit windowing [-1000, 1000], linear scaling to [0,1]
- **MRI:** Percentile clipping (1st–99th), channel-wise for multi-modal inputs
- **Ultrasound:** Speckle-robust 2nd–98th percentile clipping
- **Dermoscopy:** Per-channel RGB percentile (1st–99th)
- **Endoscopy:** Per-channel RGB, 5th–95th percentile bounds
- **X-Ray:** Wide dynamic range compression, 1st–99th percentile
- **PET:** NaN handling for SUV data, percentile clipping
- **Mammography:** Percentile-based for variable exposure settings
- **CXR:** Chest X-ray specific normalization
- **SRH:** Log transform for Stimulated Raman Histology

2.3.1 Multi-Channel MRI Handling

For multi-channel MRI data such as BraTS (FLAIR, T1w, T1gd, T2w), we implement channel selection to reduce 4 channels to SAM3's expected 3:

Listing 1: Channel selection for BraTS

```
# BraTS: FLAIR=0, T1w=1, T1gd=2, T2w=3
# Selected: [T1w, T1gd, FLAIR] for tumor contrast
indices = [1, 2, 0]  # t1_t1gd_flair mode
output = input[indices[:3]]  # -> 3 channels
```

2.4 Parameter-Efficient Fine-Tuning

We employ Low-Rank Adaptation (LoRA) [6] on attention layers:

$$W' = W + BA \tag{1}$$

where $B \in \mathbb{R}^{d\times r}$, $A \in \mathbb{R}^{r\times k}$ with $r \ll \min(d,k)$. Configuration: $r = 16$, $\alpha = 32$, dropout 0.1, targeting query-key-value and projection layers.

The perception encoder (800M parameters) remains frozen while LoRA adapts attention layers, resulting in ∼42M trainable parameters (5% of total).

2.4.1 DDP Device Placement Fix

SAM3's decoder caches coordinate tensors on the first forward pass. In DDP training, this causes device mismatch errors when ranks 1–3 attempt to use coordinates cached on rank 0's device. We implement a monkey-patch for the relative position bias matrix function to move cached tensors to the correct device.

2.5 Loss Function

We use combined Dice-Focal loss with modality-specific weighting:

$$\mathscr{L} = \lambda_{\text{dice}}\mathscr{L}_{\text{dice}} + \lambda_{\text{focal}}\mathscr{L}_{\text{focal}} \tag{2}$$

Standard modalities (CT, MRI) use equal weighting ($\lambda_{\text{dice}} = \lambda_{\text{focal}} = 1.0$, $\gamma = 2.0$). Extreme imbalance modalities (Mammography, Dermoscopy, Endoscopy, Ultrasound) use increased focal loss weighting ($\lambda_{\text{dice}} = 0.5$, $\lambda_{\text{focal}} = 1.5$, $\gamma = 3.0$).

3. Training Infrastructure

3.1 Hardware Configuration

Training was conducted on RunPod cloud infrastructure using 4× NVIDIA RTX 4090 GPUs (96GB total VRAM), AMD EPYC 7543 CPU (32 cores), 256 GB DDR4 RAM, and 2 TB NVMe SSD storage.

3.2 Distributed Data Parallel Strategy

We employ PyTorch Lightning's DDP strategy with the following configuration:

Table 1: DDP Training Configuration

Parameter	Value
Strategy	DDP (find_unused_params=True)
Backend	NCCL
Batch Size	1 per GPU
Gradient Accumulation	4 steps
Effective Batch Size	16
Precision	16-mixed (AMP)
Image Size	1008×1008

3.3 Optimizer and Scheduler

- **Optimizer:** AdamW with $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 0.05
- **Learning Rate:** 1×10^{-4}
- **Scheduler:** CosineAnnealingWarmRestarts with $T_0 = 5$, $T_{mult} = 2$
- **Early Stopping:** Patience of 10 epochs on validation Dice

4. Training Pipeline

4.1 Dataset Inventory

Onco-Seg was trained on 35 datasets across 8 phases, totaling over 98,000 training cases spanning 8 imaging modalities.

4.2 Sequential Checkpoint Chaining

Training proceeds sequentially with checkpoint chaining:

1. **Phase B (MSD):** Establishes CT/MRI foundation on Medical Segmentation Decathlon [8]
2. **Phase C (BraTS):** Adds brain tumor segmentation with multi-region targets

Table 2: Training Datasets by Phase

Phase	Datasets	Cases	Modality
B (MSD)	10 tasks	1,741	CT/MRI
C (BraTS)	3 tasks	1,410	MRI
D (OpenSRH)	1 task	250	SRH
E1 (Breast)	2 tasks	3,714	MRI/Mammo
E2 (Multi-Organ)	3 tasks	38,053	CT
E3 (CXR)	5 tasks	32,001	X-ray
F1 (High-Impact)	4 tasks	3,483	CT/MRI/PET
F2 (Modality)	4 tasks	12,697	Multi
F3 (Specialized)	3 tasks	4,805	US/XR/CT
Total	**35**	**98,154**	**8 modalities**

3. **Phase E1 (Breast):** Introduces breast MRI and mammography
4. **Phase E3 (CXR):** Adds chest X-ray modality
5. **Phase F1–F3:** Expands to oncology-critical sites and additional modalities

5. Evaluation Results

5.1 Benchmark Performance

Table 3 presents evaluation results across 12 benchmark datasets.

Table 3: Onco-Seg Evaluation Results

Dataset	Modality	Dice	IoU
BUSI	US	0.752±0.24	0.653±0.26
Hyper-Kvasir	Endo	0.714±0.32	0.637±0.33
Kvasir-SEG	Endo	0.714±0.32	0.637±0.33
ISIC 2018	Derm	0.680±0.28	0.572±0.28
LiTS	CT	0.641±0.12	0.554±0.13
3D-IRCADb	CT	0.621±0.10	0.535±0.11
PROMISE12	MRI	0.495±0.15	0.393±0.16
MSD Colon	CT	0.330±0.32	0.249±0.27
BTCV	CT	0.303±0.10	0.204±0.08
LNDb	CT	0.264±0.15	0.219±0.13
MSD Pancreas	CT	0.239±0.32	0.186±0.28
AutoPET-III	PET-CT	0.215±0.15	0.152±0.13

US=Ultrasound, Endo=Endoscopy, Derm=Dermoscopy

5.2 Performance Analysis

5.2.1 Strong Performance (Dice > 0.65)

BUSI Breast Ultrasound (0.752): Analysis of 647 images reveals category-dependent performance: benign tumors achieve Dice 0.807 while malignant tumors achieve 0.638. The performance gap reflects the irregular, spiculated boundaries characteristic of malignant lesions, which are inherently more difficult to delineate with high inter-observer agreement even among expert radiologists [5].

Kvasir-SEG/Hyper-Kvasir (0.714): Performance on 2,000 endoscopy images is consistent with specialized polyp segmentation models.

ISIC 2018 (0.680): Dermoscopy performance across 2,594 images.

LiTS/3D-IRCADb (0.64/0.62): Consistent liver CT performance across two independent datasets indicates cross-dataset generalization.

5.2.2 Moderate Performance (Dice 0.30–0.50)

PROMISE12 (0.495): Prostate MRI segmentation reflects challenges from variable MRI sequences and field strengths across institutions.

MSD Colon (0.330): Colon cancer primaries exhibit high variability in size, shape, and contrast enhancement patterns.

5.2.3 Challenging Cases (Dice < 0.30)

BTCV (0.303), LNDb (0.264), MSD Pancreas (0.239), AutoPET-III (0.215): These datasets present known challenges: BTCV requires multi-class segmentation of 13 organs; LNDb targets lung nodules (median 6mm diameter); MSD Pancreas targets a low-contrast organ with high anatomical variability; AutoPET-III requires PET-CT fusion for metabolically active lesion detection. These represent areas requiring further model development.

5.3 Limitations

Several factors affect performance on challenging datasets:

1. **Small lesion detection:** Targets below 10mm diameter require high spatial resolution that may be degraded during resizing to 1008×1008.
2. **Multi-class segmentation:** BTCV requires distinguishing 13 abdominal organs simultaneously, which the single-class prompt paradigm does not directly support.
3. **PET-CT fusion:** AutoPET-III requires integrating metabolic (PET) and anatomical (CT) information, which our preprocessing handles as separate channels rather than learned fusion.

6. Clinical Deployment

6.1 Deployment Pattern A: Interactive Sidecar

For diagnostic radiology, Onco-Seg integrates as an interactive assistant within existing PACS/viewer infrastructure:

Architecture: PACS → OHIF Viewer → FastAPI + Triton → Mask Overlay

Workflow: Radiologist clicks on suspected lesion; Onco-Seg returns segmentation mask within 500ms; auto-computed measurements (volume, diameter) populate structured report fields.

Clinical utility: Reduces manual measurement time for RECIST (Response Evaluation Criteria in Solid Tumors) assessments in oncology follow-up imaging.

6.2 Deployment Pattern B: Silent Assistant

For radiation oncology, Onco-Seg operates as an automated contouring pipeline:

Architecture: CT Scanner → AI Node → Onco-Seg → DICOM-RT Structure Set → Treatment Planning System

Workflow: CT simulation scan triggers automatic segmentation of organs-at-risk; DICOM-RT Structure Set generated for treatment planning system import.

Clinical utility: Initial auto-contours require physician review and editing; efficiency gains depend on segmentation accuracy and clinical workflow integration.

7. napari Plugin for Interactive Segmentation

To facilitate clinical deployment and research use, we developed **napari-oncoseg**, an open-source napari plugin available at https://github.com/inventcures/onco-segment/tree/main/napari_plugin.

7.1 Plugin Architecture

The plugin implements a Qt-based widget integrating with napari's layer system:

- **OncoSegWidget:** Main interface providing model loading, modality selection, and segmentation controls
- **OncoSegModel:** Wrapper around the trained checkpoint enabling point and box prompt inference
- **Preprocessing module:** Modality-specific normalization matching training transforms

7.2 Features

7.2.1 Interactive Prompting

Users provide prompts through napari's layer system:

- **Point prompts:** Click on target structure; coordinates passed to SAM3's point encoder
- **Box prompts:** Draw rectangle around target; bounding box coordinates passed to SAM3's box encoder
- **Text prompts:** Future functionality; current implementation supports point and box prompts

7.2.2 Multi-Modality Support

The plugin includes modality selection with automatic preprocessing:

Table 4: Supported Imaging Modalities

Modality	Preprocessing
CT	HU windowing [-1000, 1000]
MRI	Percentile normalization (1st–99th)
Ultrasound	Speckle-robust (2nd–98th)
Dermoscopy	Per-channel RGB
Endoscopy	Per-channel RGB (5th–95th)
PET	SUV normalization with NaN handling
X-Ray/CXR	Wide dynamic range compression
Mammography	Percentile-based

An auto-detect mode analyzes image statistics (intensity range, histogram shape) to suggest appropriate modality preprocessing.

7.2.3 3D Volume Propagation

For volumetric data, the plugin provides slice-by-slice propagation:

1. User segments one slice interactively
2. Plugin extracts mask centroid as seed point
3. Automatic propagation runs inference on remaining slices using centroid-derived prompts
4. Progress bar indicates completion status

This enables full-volume segmentation with a single click on a representative slice.

7.2.4 Export Formats

The plugin supports two export formats aligned with clinical workflows:

NIfTI (.nii.gz): Standard neuroimaging format preserving 3D structure with affine transformation matrix. Compatible with ITK-SNAP, 3D Slicer, and FSL.

DICOM-RT Structure Set: For radiation oncology integration. Requires source DICOM series for geometric reference. Uses rt-utils library to generate compliant RTSTRUCT files importable by treatment planning systems (Eclipse, RayStation, Monaco).

7.3 Installation

Listing 2: napari plugin installation

```
# Clone repository
git clone https://github.com/inventcures/onco-
    segment.git
cd onco-segment/napari_plugin

# Install with dependencies
pip install -e ".[dev]"

# Launch napari
napari
# Plugins > OncoSeg
```

7.4 Checkpoint Management

The plugin downloads pre-trained checkpoints from HuggingFace Hub:

Table 5: Available Checkpoints

Name	Training Data	Best For
latest	All phases (F3)	General purpose
breast	Duke-Breast-Cancer-MRI	Breast lesions
liver	MSD Task03	Liver/hepatic tumors
brain	BraTS-GLI	Brain tumors

Users can also load custom checkpoints via file browser.

8. Discussion

8.1 Summary

Onco-Seg demonstrates that SAM3's promptable concept segmentation transfers to medical imaging across multiple modalities. The unified preprocessing framework enables a single model to handle CT, MRI, ultrasound, dermoscopy, and endoscopy.

Performance on breast ultrasound (Dice 0.752), polyp segmentation (0.714), and liver CT (0.641) indicates clinical utility for these anatomical targets. Performance on small targets

(lung nodules, pancreatic lesions) and multi-class segmentation remains limited.

8.2 Clinical Considerations

The napari plugin provides an accessible interface for clinical evaluation. Several considerations apply to clinical use:

1. **Validation requirement:** All auto-generated contours require physician review and editing before clinical use. Automated segmentation should augment, not replace, clinical judgment.
2. **Device registration:** DICOM-RT export preserves source image geometry but requires verification against treatment planning system import to ensure spatial accuracy.
3. **Regulatory status:** Onco-Seg is a research tool not cleared for clinical use. Clinical deployment would require institutional review board approval, prospective validation studies, and potentially regulatory clearance (FDA 510(k) or CE marking).
4. **Population generalization:** Training data demographics may not represent all patient populations. Performance should be validated on local institutional data before clinical integration.
5. **Edge cases:** Atypical presentations, rare pathologies, and cases with imaging artifacts may produce unreliable segmentations requiring manual correction.

8.3 Future Directions

Short-term: Uncertainty quantification for identifying low-confidence segmentations; 3D Slicer extension.

Medium-term: Multi-center validation studies; federated learning for privacy-preserving training on institutional data.

Long-term: Integration with clinical trial imaging endpoints; regulatory pathway exploration.

9. Code and Data Availability

Source code, trained model checkpoints, and the napari plugin are available at:

- **Repository:** https://github.com/inventcures/onco-segment
- **napari plugin:** `onco-segment/napari_plugin/`
- **Checkpoints:** HuggingFace Hub (`tp53/oncoseg`)
- **Project page:** https://inventcures.github.io/onco-seg/

Training datasets are publicly available from their respective sources (Medical Segmentation Decathlon, BraTS, LiTS, ISIC, Kvasir-SEG, TCIA).

10. Acknowledgments

This work was supported by the Koita Centre for Digital Health at Ashoka University (KCDH-A). We thank RunPod for GPU infrastructure and Weights & Biases for experiment tracking.

We acknowledge Meta AI and the SAM team, led by Nikhila Ravi, for their open release of SAM, SAM2, and SAM3. Their decision to publish technical reports and model weights has enabled research extensions like Onco-Seg.

We thank Bo Wang (Vector Institute, University of Toronto) and colleagues for MedSAM [9], which demonstrated SAM adaptation for medical imaging.

We are grateful to the National Cancer Institute (NCI), CBIIT, and The Cancer Imaging Archive (TCIA) for providing open-access datasets. Thanks to Justin Kirby at TCIA for technical assistance.

We thank the creators of benchmark datasets: Medical Segmentation Decathlon, BraTS, LiTS, ISIC, Kvasir-SEG, PROMISE12, BUSI, IRCADb, BTCV, LNDb, and AutoPET.

References

[1] Kirillov, A., et al. (2023). Segment Anything. *ICCV 2023*.

[2] Ravi, N., et al. (2024). SAM 2: Segment Anything in Images and Videos. *arXiv:2408.00714*.

[3] Carion, N., et al. (2025). SAM 3: Segment Anything with Concepts. *arXiv:2511.16719*.

[4] Brouwer, C.L., et al. (2012). 3D Variation in delineation of head and neck organs at risk. *Radiotherapy and Oncology*, 106(3), 392–396.

[5] Vinod, S.K., et al. (2016). Uncertainties in volume delineation in radiation oncology. *Radiotherapy and Oncology*, 121(2), 169–179.

[6] Hu, E.J., et al. (2022). LoRA: Low-Rank Adaptation of Large Language Models. *ICLR 2022*.

[7] Isensee, F., et al. (2021). nnU-Net. *Nature Methods*, 18(2), 203–211.

[8] Antonelli, M., et al. (2022). The Medical Segmentation Decathlon. *Nature Communications*, 13, 4128.

[9] Ma, J., He, Y., et al. (2024). Segment Anything in Medical Images. *Nature Communications*, 15, 654.

[10] napari contributors (2019). napari: a multi-dimensional image viewer for Python. https://napari.org

[11] Wacker, A. (2021). rt-utils: A Python library for RT structure set creation. https://github.com/qurit/rt-utils

[12] Biewald, L. (2020). Experiment Tracking with Weights and Biases. https://www.wandb.com/

Corresponding Author: Ashish Makani (ashish.makani@ashoka.edu.in)

Project Page: https://inventcures.github.io/onco-seg/