

3D MEDICAL IMAGE SEGMENTATION WITH ANATOMY-GUIDED CONDITIONING FROM SURROUNDING STRUCTURES

Anonymous authors

Paper under double-blind review

ABSTRACT

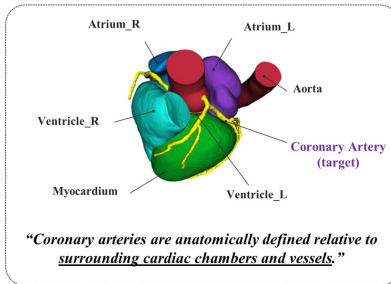
Accurate segmentation of complex anatomical structures in 3D medical images is challenged by low contrast, unclear features, and complex topology. We propose Anatomy-Guided Conditioning (AGC), which integrates signed distance maps of surrounding anatomy into segmentation networks via feature modulation in the decoder. Anatomical priors are obtained from automatic tools such as TotalSegmentator, requiring no additional training and enabling use in both multi-class and single-target tasks. We evaluate AGC on CTA coronary arteries, PET/CT visceral fat, CT head-and-neck organs, and CBCT dental canals. Across CNN, Transformer, and hybrid backbones, AGC improves Dice, HD95, and topology-aware metrics (cDice, Betti error), reducing boundary errors and fragmentation. These results demonstrate that conditioning on surrounding anatomy provides a simple and broadly applicable inductive bias for anatomically constrained 3D segmentation.

1 INTRODUCTION

Deep learning has achieved remarkable success in 3D medical image segmentation, with models reaching expert-level accuracy on benchmarks such as MSD (Antonelli et al., 2022) and BraTS (Menze et al., 2015; Bakas et al., 2018). Yet most methods still rely heavily on local image appearance and voxel-wise objectives, often neglecting global anatomical context. This limitation leads to anatomically implausible predictions, including fragmented vessels, leakage beyond organ boundaries, and topological inconsistencies (Wyburd et al., 2024; Bassi et al., 2024). Such problems are particularly severe in coronary artery segmentation, where continuity must be preserved through low-contrast and artifact-prone regions (Duan et al., 2019).

Radiologists routinely delineate structures by referencing surrounding anatomy rather than considering each organ in isolation. This observation motivates the need for segmentation frameworks that explicitly incorporate anatomical context from anchor anatomical regions.

(i) Importance of Surrounding Structures



(ii) Anatomy-Guided Conditioning (Proposed)

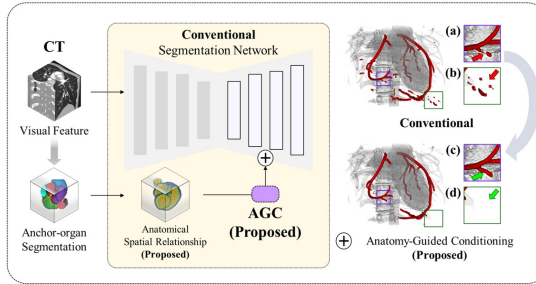


Figure 1: Illustration of coronary artery segmentation. Left: baseline prediction with fragmented branches and boundary leakage. Right: our proposed AGC reduces errors by leveraging surrounding anatomical context.

We introduce **Anatomy-Guided Conditioning (AGC)**, a simple conditioning module that integrates signed distance maps of neighboring structures into segmentation networks. Unlike binary mask or coordinate priors Guo et al. (2020); Goncharov et al. (2024), distance maps encode continuous proximity information, enabling more precise spatial reasoning. By injecting these priors at the feature level via FiLM-based modulation Perez et al. (2018), AGC enhances structural coherence without altering inputs or encoders.

We evaluate AGC on diverse datasets, including thin tubular vessels (ImageCAS), diffuse abdominal regions (VAT PET/CT), small head-and-neck organs (HAN-Seg), and narrow dental canals (Tooth-Fairy2). These benchmarks present complementary challenges such as continuity, boundary precision, and variability across both CT and CBCT modalities. AGC consistently improves overlap, boundary, and topology-aware metrics, and provides complementary benefits when combined with recent topology-preserving losses such as cIDice Menten et al. (2023) and Skeleton Recall Kirchhoff et al. (2024).

Our contributions are threefold:

- We propose **AGC**, a conditioning mechanism based on signed distance maps of surrounding anatomy, enabling proximity-aware reasoning beyond categorical masks or coordinates.
- We demonstrate that AGC consistently improves structure-aware metrics and acts synergistically with topology-preserving objectives, enhancing continuity and reducing topological errors.
- We show that AGC is **simple, backbone- and dataset-agnostic**, requiring no manual annotation by leveraging off-the-shelf tools such as TotalSegmentator Wasserthal et al. (2023).

2 RELATED WORKS

Network Design for Medical Image Segmentation. Volumetric segmentation has been explored with a broad range of architectures, from convolutional encoder-decoders (e.g., 3D U-Net (Çiçek et al., 2016), V-Net (Milletari et al., 2016), SegResNet (Myronenko, 2019)) to transformer-based and hybrid designs (e.g., TransUNet (Chen et al., 2024), UNETR (Hatamizadeh et al., 2022), Swin-UNETR (Hatamizadeh et al., 2021)). Large-scale benchmarks such as nnU-Net (Isensee et al., 2021) and Touchstone (Bassi et al., 2024) show that performance is often task-dependent, and that carefully tuned CNNs can match more complex backbones Azad et al. (2024). However, architectural refinements alone are insufficient for thin or low-contrast structures, where local features are unreliable. These cases motivate the use of anatomical priors that are model-agnostic and can provide organ-relative context beyond raw image intensities.

TotalSegmentator as a Source of Anatomical Priors. *TotalSegmentator* (Wasserthal et al., 2023) is a foundation model trained on over 1,200 CT scans that outputs robust pseudo-labels for 104 organs, bones, and vessels. Its coverage and generalization make it a scalable source of anatomical context, readily applicable to CT, CBCT, and even MRI. Importantly, these pseudo-labels require no manual annotation, enabling downstream tasks to access surrounding anatomy without extra supervision. This capability allows priors to be introduced even in single-target problems such as coronary artery segmentation, where auxiliary multi-class labels are typically unavailable. In our framework, this factor is a key enabler of broad applicability.

Anatomical Prior-based Conditioning. Several strategies have been proposed to incorporate anatomical knowledge. Topology-aware losses such as cIDice (Shit et al., 2021) and Skeleton Recall (Kirchhoff et al., 2024) encourage connectivity via differentiable skeletonization, but remain confined to the optimization stage and require task-specific tuning. Shape-based approaches, e.g., SPGNet (Song et al., 2024), depend on explicit shape models learned from annotated datasets, limiting scalability to diverse anatomies or 3D volumes. Coordinate embeddings (Das et al., 2024) add absolute voxel positions but lack anatomical semantics, reducing generalizability across architectures. Anchor-based methods (Guo et al., 2020) sequentially segment organs using binary masks as anchors; while effective in multi-target head-and-neck settings, they are unsuitable for single-target tasks and cannot encode continuous distance information.

3 METHODS

3.1 PROPOSED FRAMEWORK

AGC enhances segmentation by embedding spatial context from surrounding anatomy into the decoding process. Surrounding structures are obtained from off-the-shelf TotalSegmentator, which produces pseudo-labels for over 100 organs across CT, CBCT, and MRI without requiring manual annotation. These pseudo-labels are converted into signed distance maps to provide continuous proximity information, which is then injected via FiLM-based feature modulation in the decoder (Figure 2). This design distinguishes AGC from prior mask- or shape-based conditioning, offering a simple, backbone-agnostic mechanism that supplies proximity-aware guidance even in single-target tasks such as coronary artery segmentation.

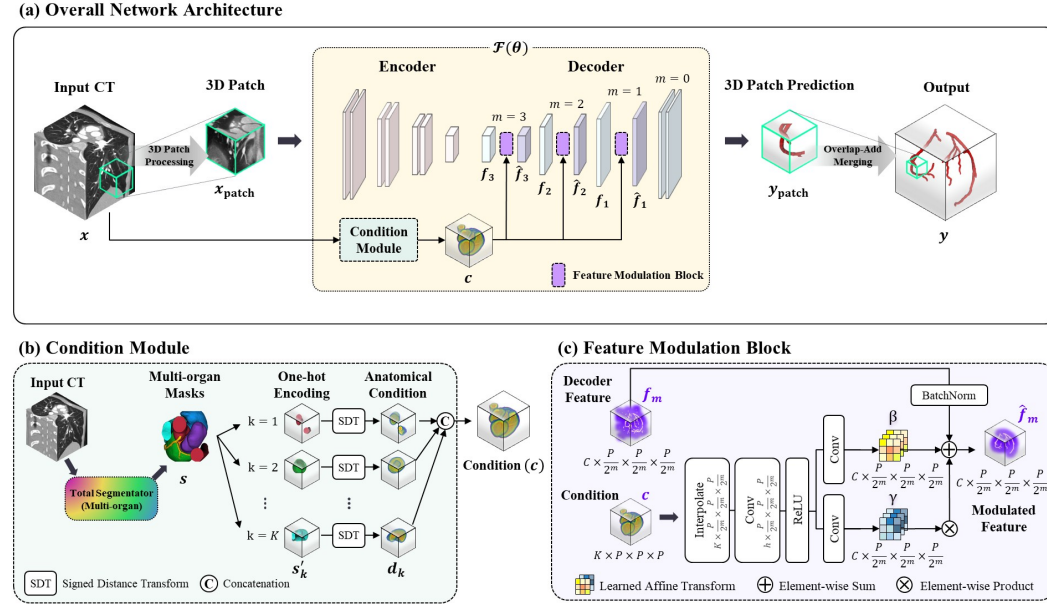


Figure 2: Architecture of the proposed Anatomy-Guided Conditioning (AGC) framework, including the overall network (a), condition module (b), and feature modulation block (c).

Network Architecture The proposed AGC method is integrated into a standard encoder-decoder architecture and follows a patch-based framework commonly adopted in 3D medical image segmentation Isensee et al. (2021). To support clarity, we briefly outline the principle of general patch-based segmentation framework.

As illustrated in Figure 2 (a), a full CT volume $x \in \mathbb{R}^{1 \times H \times W \times D}$ is divided into overlapping 3D patches $x_{\text{patch}} \in \mathbb{R}^{1 \times P \times P \times P}$, where $P = 96$ in our experiments. Each patch is processed independently, and the resulting predictions are aggregated into a full-resolution output $y \in \mathbb{R}^{2 \times H \times W \times D}$ using a weighted averaging scheme with a 3D Gaussian kernel and 25% overlap, mitigating boundary artifacts at patch edges. As seen in the notation above, binary segmentation tasks are formulated as two-class problems (foreground and background) following standard practice to maintain stability and compatibility with conventional loss functions with multi-class based cross-entropy. The following sections detail how anatomical priors are extracted and encoded, and how they modulate decoding features via our feature modulation mechanism.

Anatomical Condition Module We extract anatomical priors by first applying a multi-organ segmentation tool, such as **TotalSegmentator** (Wasserthal et al., 2023), to an input CT volume $x \in \mathbb{R}^{H \times W \times D}$. This model produces a semantic pseudo label map s :

$$s \in \{1, 2, \dots, K\}^{H \times W \times D},$$

where each voxel is assigned an integer label corresponding to one of K anatomical anchor structures. We convert this label map into a one-hot encoded tensor s' :

$$s' \in \{0, 1\}^{K \times H \times W \times D},$$

where the k -th slice s'_k is a binary mask indicating the presence of organ k . The region of organ k is defined as:

$$\mathcal{O}_k = \{\mathbf{p} \in \Omega \mid s'_k(\mathbf{p}) = 1\},$$

where $\Omega = \{1, \dots, H\} \times \{1, \dots, W\} \times \{1, \dots, D\}$ is the full spatial domain.

To represent soft spatial context beyond hard anatomical boundaries, we compute a **signed distance transform** for each organ k , resulting in a signed distance map $d_k \in \mathbb{R}^{H \times W \times D}$ defined as:

$$d_k(\mathbf{p}) = \begin{cases} -\min_{\mathbf{q} \in \Omega \setminus \mathcal{O}_k} \|\mathbf{p} - \mathbf{q}\|_2, & \text{if } \mathbf{p} \in \mathcal{O}_k, \\ \min_{\mathbf{q} \in \mathcal{O}_k} \|\mathbf{p} - \mathbf{q}\|_2, & \text{if } \mathbf{p} \notin \mathcal{O}_k, \end{cases}$$

where $\mathbf{p}, \mathbf{q} \in \Omega$ denote voxel coordinates and $\|\cdot\|_2$ is the Euclidean norm. This results in negative values inside the organ ($\mathbf{p} \in \mathcal{O}_k$), positive values outside the organ ($\mathbf{p} \notin \mathcal{O}_k$), and a zero level set at the organ boundary. Finally, we stack all organ-wise distance maps into an anatomical prior tensor c :

$$c = \{d_k\}_{k=1}^K \in \mathbb{R}^{K \times H \times W \times D},$$

The anatomical prior c serves as a rich spatial prior that captures both the absolute and relative anatomical positioning. It is extracted as $K \times P \times P \times P$ 3D patches and used as conditioning input for downstream components (see Figures 2(a) and (b)).

Anatomy-Guided Feature Modulation To integrate anatomical context with minimal interference, we inject the generated priors into the decoder using a Feature Modulation Block (see Figure 2(c)). In particular, many existing methods leverage pre-trained encoders on large-scale datasets, which produce features where general-purpose priors are well-preserved. Modifying the encoder can disrupt these valuable features and degrade performance. Therefore, we focus on integrating the generated anatomical priors into the decoder. This design allows anatomical priors to guide feature refinement at later stages, where semantic information is richer, without altering early feature extraction.

At each decoder stage m , let $f_m \in \mathbb{R}^{C \times \frac{P}{2^m} \times \frac{P}{2^m} \times \frac{P}{2^m}}$ be the decoder feature map, and let c be the anatomical prior tensor. We first apply trilinear interpolation to c to match the spatial dimensions of f_m , resulting in a down-sampled prior of shape $K \times \frac{P}{2^m} \times \frac{P}{2^m} \times \frac{P}{2^m}$. The interpolated prior is then processed by a lightweight convolutional block to extract modulation parameters. Specifically, a $3 \times 3 \times 3$ convolution first projects the K -channel prior into a higher-dimensional hidden representation (128 channels), followed by a ReLU activation. This intermediate representation is then passed through two parallel $3 \times 3 \times 3$ convolutions to produce the spatially adaptive scaling and shifting parameters, γ and β , respectively, both in $\mathbb{R}^{C \times \frac{P}{2^m} \times \frac{P}{2^m} \times \frac{P}{2^m}}$. These modulation parameters are applied to the normalized decoder features via FiLM-style conditioning Perez et al. (2018):

$$\hat{f}_m = \gamma \odot \text{BN}(f_m) + \beta, \quad (1)$$

where $\text{BN}(\cdot)$ denotes batch normalization and $\odot$ is element-wise multiplication. This modulation allows the network to adjust decoder features in a spatially adaptive yet minimally invasive manner, correcting anatomical inconsistencies while preserving the underlying feature representation.

3.2 DATASETS

We benchmark AGC on four CT/Cone-Beam CT-based datasets that span diverse anatomical targets and challenges: i) ImageCAS for thin, highly branched coronary vessels, ii) VAT PET/CT for spatially diffuse abdominal fat, iii) HAN-Seg for small, heterogeneous head & neck structures under data scarcity, and iv) Tooth-Fairy2 for narrow dental canals. Anatomical masks used for conditioning were either provided with the datasets or generated using off-the-shelf **TotalSegmentator v2.6.0**. Here we summarize only the keypoints and illustrative details of conditioning for each dataset and preprocessing schemes are further demonstrated in **Supplementary Figure 1**.

ImageCAS—Coronary Arteries (Zeng et al., 2023). 1,000 contrast-enhanced chest CTA volumes (700/50/250 train/val/test). Target: coronary arteries, prone to centerline fragmentation due to their thin and densely branched morphology. Conditioning: aorta, myocardium, atria, and ventricles from TotalSegmentator.

Low-dose PET/CT VAT—Visceral Adipose Tissue. 147 low-dose cone-beam CT scans (108/10/29 train/val/test). Target: visceral adipose tissue spanning the abdominal region, difficult to delineate under low-contrast CBCT. Conditioning: body trunk, torso fat, liver, heart, spine, and ribs.

HAN-Seg—Head & Neck Organs-at-Risk (Podobnik et al., 2023). 42 CT scans (30/12 train/test). Targets: carotids, spinal cord, mandible, and thyroid, with challenges of discontinuity, high shape variability, and low tissue contrast. Conditioning: dataset-provided labels of targets and neighboring structures.

Tooth-Fairy2—Inferior Alveolar Canal (Bolelli et al., 2024). 417 dental CBCT scans (292/25/100 train/val/test). Target: inferior alveolar canal, a narrow, low-contrast tubular structure critical in dental surgery. Conditioning: dataset-provided label of mandible and surrounding tooth masks.

Implementation Details All scans were resampled to isotropic spacing and clipped to task-specific Hounsfield windows (e.g., -150 – 550 HU for ImageCAS). Intensity values were min–max normalized to $[0, 1]$. We followed MONAI’s reference pipeline Cardoso et al. (2022) using 96^3 voxel patches, with augmentations including random axis flips (10% probability) and intensity jittering of $\pm 5\%$ (50%).

Models were trained with the Adam optimizer for 200 epochs using a starting learning rate of 0.001 decayed via cosine annealing. The Dice–Focal loss was used to balance overlap accuracy and class imbalance. We report four metrics: (1) **Dice coefficient** for region overlap; (2) **95th-percentile Hausdorff distance** (HD95) for boundary error; (3) **Betti number error** (Hu et al., 2019), which quantifies topological mismatches via β_0 (components) and β_1 (loops); and (4) **clDice** (Shit et al., 2021), a connectivity-aware variant suited for thin structures. Higher Dice and clDice are better; lower HD₉₅ and Betti error indicate better geometric and topological fidelity.

4 EXPERIMENTS AND RESULTS

4.1 EXPERIMENTAL RATIONALE

The experiments are designed to evaluate AGC as a simple and general mechanism for embedding anatomical context across models, datasets, and training objectives. We organize the evaluation along four complementary axes.

Backbone generality. We first test whether AGC improves performance consistently across diverse segmentation architectures. Since AGC operates at the feature level, it integrates seamlessly with both convolutional and transformer-based models without modifying inputs or encoders.

Dataset generalization. Second, we evaluate AGC across four benchmarks spanning distinct anatomies and modalities: thin coronary vessels (ImageCAS, CTA), diffuse abdominal fat (VAT, PET/CT), small head-and-neck organs (HAN-Seg, CT), and narrow dental canals (Tooth-Fairy2, CBCT). These tasks capture complementary challenges such as continuity, boundary precision, and variability under heterogeneous imaging conditions.

Relation to topology-aware objectives. Third, we compare AGC with topology-preserving losses such as clDice Menten et al. (2023) and Skeleton Recall Kirchhoff et al. (2024). We hypothesize complementarity: topology-aware losses regularize connectivity during optimization, while AGC guides feature representations with anatomical priors.

Alternative conditioning. Finally, we compare AGC with positional encodings (PPE, APE) and mask priors. These provide only coarse spatial cues, whereas AGC encodes continuous proximity information via distance maps, offering a more principled conditioning mechanism.

4.2 MODEL-AGNOSTIC EVALUATION

We first assess whether AGC improves segmentation across convolutional, transformer-based, hybrid, and vascular-specialized backbones. Table 1 shows consistent improvements not only in overlap but also in boundary and topology-aware metrics. Convolutional models benefit most, with AGC reducing discontinuities and false positives, while transformers such as UNETR show alleviated fragmentation and restored connectivity. Qualitative results in Figure 3 confirm enhanced structural coherence across architectures. Supplementary Table 1 reports model-agnostic performance on other datasets.

Table 1: Model-agnostic evaluation of AGC across **seven segmentation backbones on the ImageCAS dataset**. AGC denotes whether our anatomy-guided conditioning was used or not

Backbone	Architecture	AGC	DSC $\uparrow$	HD95 $\downarrow$	clDice $\uparrow$	β_0 error $\downarrow$	β_1 error $\downarrow$
SegResNet	Convolutional		0.77 ± 0.03	81.8 ± 29.1	0.82 ± 0.08	25.2 ± 8.71	22.7 ± 8.30
	Convolutional	✓	0.79 ± 0.02	36.4 ± 12.0	0.86 ± 0.05	10.6 ± 3.94	10.0 ± 3.99
VNet	Convolutional		0.78 ± 0.02	66.0 ± 18.9	0.84 ± 0.06	19.0 ± 6.75	17.3 ± 6.34
	Convolutional	✓	0.79 ± 0.05	38.2 ± 25.8	0.86 ± 0.06	12.4 ± 5.21	11.5 ± 4.97
Attention-UNet	Convolutional		0.79 ± 0.02	57.1 ± 17.4	0.84 ± 0.06	18.4 ± 6.69	16.9 ± 6.22
	Convolutional	✓	0.78 ± 0.02	31.1 ± 10.9	0.86 ± 0.05	10.2 ± 4.46	9.56 ± 4.26
UNETR	Transformer		0.69 ± 0.03	162.7 ± 17.5	0.60 ± 0.09	221.4 ± 50.6	208.9 ± 48.2
	Transformer	✓	0.76 ± 0.02	47.9 ± 11.7	0.74 ± 0.07	86.3 ± 24.0	81.0 ± 22.4
Swin-UNETR	Hybrid		0.79 ± 0.03	55.3 ± 16.2	0.84 ± 0.06	20.6 ± 7.27	19.1 ± 6.90
	Hybrid	✓	0.79 ± 0.02	43.9 ± 13.0	0.84 ± 0.06	19.6 ± 6.25	18.6 ± 5.76
nnFormer	Hybrid		0.72 ± 0.03	122.3 ± 24.3	0.74 ± 0.08	74.8 ± 27.5	72.3 ± 26.5
	Hybrid	✓	0.77 ± 0.02	33.9 ± 10.8	0.82 ± 0.06	31.8 ± 15.6	29.8 ± 14.7
CS ² Net	Vascular-spec.		0.72 ± 0.04	92.1 ± 20.5	0.76 ± 0.08	55.1 ± 18.8	51.4 ± 18.1
	Vascular-spec.	✓	0.76 ± 0.02	40.2 ± 11.2	0.83 ± 0.06	28.2 ± 10.6	26.0 ± 9.85

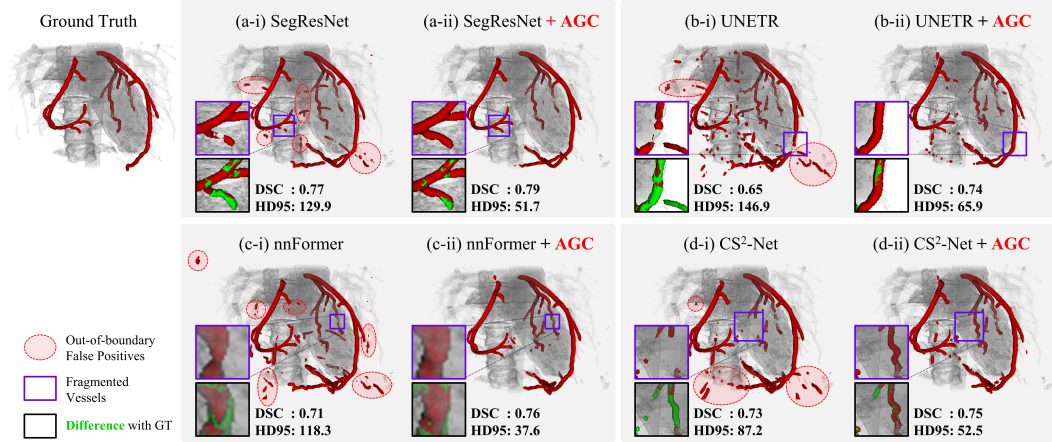


Figure 3: Qualitative 3D visualization of segmentation results on the ImageCAS dataset across multiple segmentation backbone models, with and without Anatomy-Guided Conditioning (AGC).

4.3 EVALUATION ON MULTIPLE DATASET

We next evaluate AGC on VAT, HAN-Seg, and Tooth-Fairy2 (Table 2). In VAT, conditioning sharpens diffuse abdominal boundaries. In HAN-Seg, AGC improves challenging small structures such as carotids and mandible under limited data. For Tooth-Fairy2, conditioning on mandible and teeth provides strong anchors for the alveolar canal, yielding highly stable segmentations with minimal topological errors. Qualitative examples are shown in Figure 4, with additional cases in the Supplementary.

Table 2: Comparison between conventional and proposed models on PET/CT VAT, HaN-Seg, and Tooth-Fairy2 datasets. **AGC** column indicates whether our conditioning was used.

Dataset	AGC	DSC $\uparrow$	mIoU $\uparrow$	HD95 $\downarrow$	β_0 error $\downarrow$	β_1 error $\downarrow$
Visceral Adipose Fat (PET-CT VAT)	✓	0.94 ± 0.02 0.96 ± 0.03	0.90 ± 0.04 0.92 ± 0.04	1.82 ± 1.55 1.56 ± 1.05	137.4 ± 73.5 128.6 ± 48.1	129.1 ± 70.0 119.8 ± 44.6
Carotid Artery (HAN-Seg)	✓	0.65 ± 0.07 0.73 ± 0.04	0.48 ± 0.07 0.57 ± 0.05	269.6 ± 74.1 19.8 ± 5.15	36.6 ± 20.2 6.25 ± 2.13	33.3 ± 19.5 5.50 ± 1.94
Spinal Cord (HAN-Seg)	✓	0.73 ± 0.07 0.74 ± 0.05	0.58 ± 0.08 0.59 ± 0.07	235.0 ± 132.5 162.9 ± 27.1	9.58 ± 6.82 5.17 ± 9.35	8.33 ± 6.17 5.08 ± 9.31
Mandible (HAN-Seg)	✓	0.84 ± 0.04 0.92 ± 0.02	0.73 ± 0.06 0.84 ± 0.03	261.9 ± 31.1 3.06 ± 4.11	51.3 ± 23.7 3.25 ± 6.11	47.0 ± 20.1 2.75 ± 5.07
Thyroid (HAN-Seg)	✓	0.78 ± 0.09 0.81 ± 0.08	0.64 ± 0.01 0.69 ± 0.11	215.6 ± 130.0 6.57 ± 4.62	12.6 ± 5.24 1.25 ± 0.83	11.6 ± 4.27 1.17 ± 0.80
Alveolar Canal (Tooth-Fairy2)	✓	0.89 ± 0.02 0.98 ± 0.01	0.79 ± 0.02 0.96 ± 0.02	2.56 ± 6.89 0.68 ± 0.77	1.82 ± 1.41 0.45 ± 0.84	1.64 ± 1.32 0.36 ± 0.68

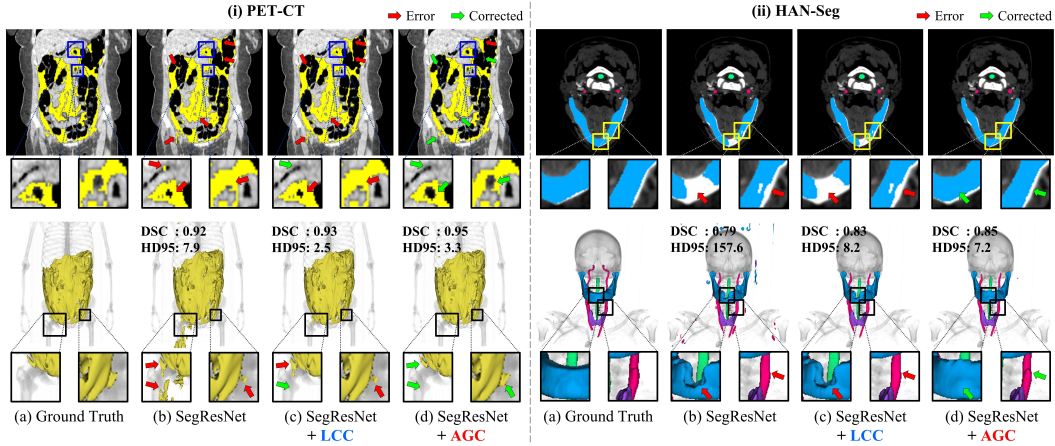


Figure 4: Qualitative comparison of segmentation results on the PET-CT and HAN-Seg datasets. Each column presents the ground truth, SegResNet baseline, and outputs refined by either the largest connected component (LCC) post-processing or our proposed AGC.

4.4 LOSS-AGNOSTIC EVALUATION

We next compare AGC with recent topology-aware losses, namely cIDice (Menten et al., 2023) (ICCV 2023) and Skeleton Recall (Kirchhoff et al., 2024) (ECCV 2024), both of which enforce 3D centerline and topology preservation during optimization. The comparison addresses two questions: (i) whether AGC performs competitively against these state-of-the-art losses, and (ii) whether they act synergistically when combined. Table 3 shows consistent improvements across losses and datasets, indicating that AGC and topology-aware objectives are complementary: the former injects anatomical context at the feature level, while the latter enforces connectivity during training. The full table, including standard deviations and Betti errors, is provided in Supplementary Table 2.

Metrics such as Dice can sometimes favor DiceFocal despite inferior continuity, reflecting their voxel-wise nature. Qualitative figures (Figure 5) clarify such cases, showing that AGC with topology-aware losses reduces fragmentation and improves global coherence even when Dice values are similar.

4.5 SPATIAL GUIDANCE ALTERNATIVES

Finally, we compare AGC against other conditioning mechanisms, including physical positional embeddings (PPE), anatomical positional embeddings (APE), and binary mask pri-

Table 3: Loss-agnostic evaluation across datasets. Each subtable reports three different losses (DiceFocal, SkelRec, cIDice) with/without AGC.

(a) ImageCAS (Coronary)					(b) HAN-Seg (Carotid)					(c) Tooth-Fairy2 (Alveolar Canal)				
Loss	AGC	DSC	HD95	cIDice	Loss	AGC	DSC	HD95	cIDice	Loss	AGC	DSC	HD95	cIDice
DiceFocal	✓	0.766	81.8	0.815	DiceFocal	✓	0.646	269.6	0.646	DiceFocal	✓	0.887	2.56	0.987
		0.788	36.4	0.862			0.727	19.8	0.789			0.981	0.68	0.994
SkelRec	✓	0.777	40.6	0.856	SkelRec	✓	0.671	108.1	0.741	SkelRec	✓	0.879	1.54	0.994
		0.778	39.2	0.844			0.698	111.5	0.796			0.988	0.36	0.997
cIDice	✓	0.786	27.9	0.855	cIDice	✓	0.671	39.3	0.801	cIDice	✓	0.891	1.47	0.994
		0.794	27.2	0.871			0.673	9.28	0.841			0.989	0.31	0.997

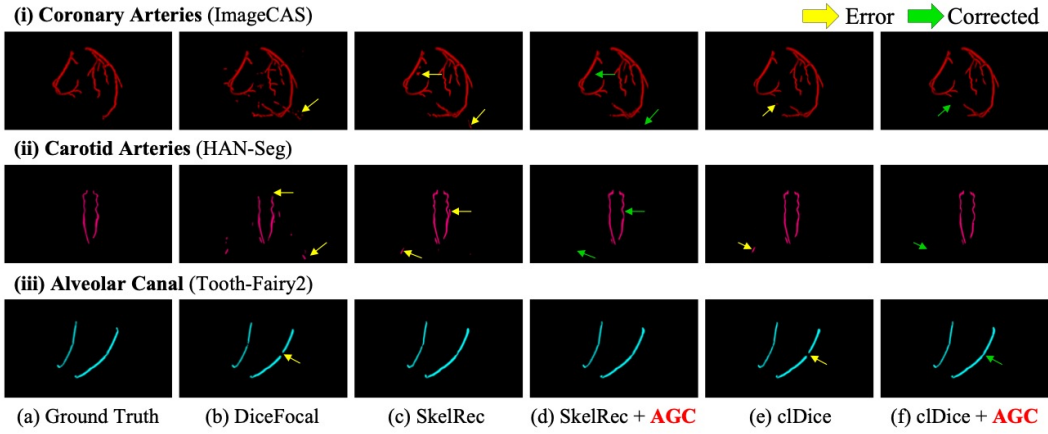


Figure 5: Evaluation across loss functions. Yellow and Green arrow indicate errors in wrong segmentation and where AGC has corrected them

ors. As reported in Table 4, distance map conditioning provides the most consistent improvements. PPE introduces raw coordinates but fails to encode anatomical relevance, while APE shows moderate gains but lacks robustness across tasks.

Mask priors provide categorical boundaries without proximity information, limiting their utility. In contrast, AGC encodes continuous spatial context, resulting in lower topological errors and more coherent predictions. Figure 6 provides qualitative comparisons, confirming that distance maps restore connectivity where alternative priors remain fragmented.

Table 4: Comparison of SegResNet variants using PPE, APE, Mask, and DistMap.

Method	DSC $\uparrow$	HD95 $\downarrow$	$\beta_0 \downarrow$	$\beta_1 \downarrow$
SegResNet (Base)	0.77 ± 0.03	81.8 ± 29.1	25.2 ± 8.71	22.7 ± 8.30
+ PPE	0.59 ± 0.03	140.2 ± 11.8	88.4 ± 23.0	78.5 ± 20.4
+ APE	0.77 ± 0.03	72.8 ± 22.7	19.9 ± 8.10	18.6 ± 7.62
+ Mask	0.78 ± 0.02	60.3 ± 18.2	24.5 ± 7.45	22.4 ± 6.99
+ DistMap (Ours)	0.79 ± 0.02	36.4 ± 12.0	10.6 ± 3.94	10.0 ± 3.99

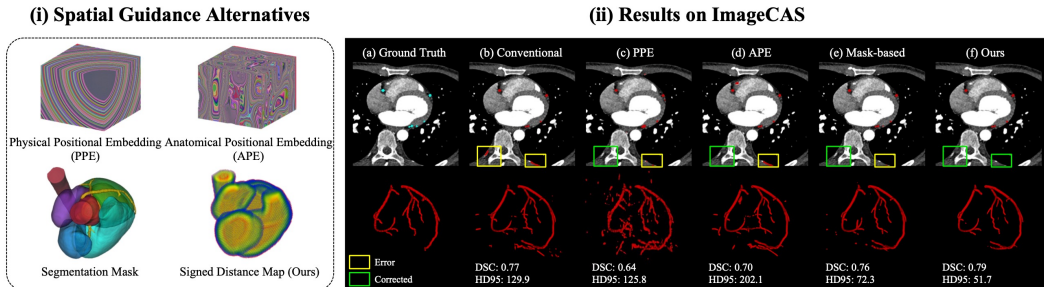


Figure 6: Comparison of SegResNet variants using PPE, APE, Mask, and DistMap.

4.6 ABLATION STUDY

We conduct two ablation studies to analyze the design choices of AGC. For efficiency and reproducibility, all experiments are performed on a subset of the ImageCAS dataset (210/20/100 train/val/test).

Encoding Strategy of Distance Map Prior: A natural concern is whether distance maps actually provide any benefit over simple mask-based priors, since both originate from the same anatomical labels. To address this, we progressively clipped the signed distance maps, starting from a degenerate binary mask ($[0, 0]$) and extending the range to include both interior (negative) and exterior (positive) values. As summarized in Table 5a, the binary case yields limited improvements. The full-range variant achieves the best performance overall, demonstrating that distance maps are effective because they encode global spatial positioning across the entire volume.

Injection Location of the Modulation Block: The second study examines whether priors should be injected into the encoder, the decoder, or both. This question is particularly relevant when using pretrained encoders such as Swin-UNETR, where modifying early feature representations can be detrimental. Table 5b shows that decoder-only integration achieves the most consistent results, while encoder-only or dual injection provides less stable improvements.

Table 5: Ablation studies: (a) different distance map clipping for anatomical prior injection; (b) injection at different network components.

(a) Distance map clipping strategies						(b) Injection depth in the network					
Threshold	Purpose	DSC $\uparrow$	HD95 $\downarrow$	$\beta_0 \downarrow$	$\beta_1 \downarrow$	Encoder	Decoder	DSC $\uparrow$	HD95 $\downarrow$	$\beta_0 \downarrow$	$\beta_1 \downarrow$
$[0, 0]$	Boundary only	0.76 ± 0.04	54.1 ± 25.0	27.5 ± 6.85	25.3 ± 6.23						
$[-20, 20]$	Narrow symmetric	0.76 ± 0.04	48.9 ± 26.0	13.0 ± 4.43	12.2 ± 4.22						
$[-50, 50]$	Mid-range symmetric	0.77 ± 0.04	47.2 ± 25.7	20.0 ± 6.29	18.5 ± 5.96						
$[-150, 150]$	Wide symmetric	0.76 ± 0.04	52.1 ± 25.7	18.7 ± 5.75	17.4 ± 5.50						
$[\text{min}, \text{max}]$	Full range	0.77 ± 0.04	36.5 ± 23.0	14.3 ± 5.29	13.3 ± 4.94						
$[\text{min}, 0]$	Inside only	0.75 ± 0.04	58.9 ± 23.1	40.8 ± 10.9	37.0 ± 9.42						
$[0, \text{max}]$	Outside only	0.76 ± 0.04	49.7 ± 23.8	20.9 ± 6.78	19.2 ± 6.39						

Encoder	Decoder	DSC $\uparrow$	HD95 $\downarrow$	$\beta_0 \downarrow$	$\beta_1 \downarrow$
	✓	0.77 ± 0.04	36.5 ± 23.0	14.3 ± 5.29	13.3 ± 4.94
✓		0.76 ± 0.04	46.5 ± 23.1	26.2 ± 7.94	24.1 ± 6.95
✓	✓	0.77 ± 0.04	41.4 ± 24.2	25.7 ± 7.90	23.4 ± 7.32

5 DISCUSSIONS AND CONCLUSIONS

The purpose of this study was not to establish state-of-the-art results on every benchmark, but to examine whether the proposed anatomical guidance can serve as a simple and effective conditioning mechanism in 3D medical image segmentation. Our experiments provide strong evidence that incorporating surrounding anatomy improves both robustness and plausibility. The results on Tooth-Fairy2 are particularly interesting: by conditioning on coarse tooth segmentations, the network was able to delineate the alveolar canal with much improved precision. This illustrates how relative spatial anchors can play a decisive role, and suggests that models can exploit neighbor-aware information in a way reminiscent of human reasoning. More broadly, we observed that AGC consistently reduces false positives appearing in anatomically implausible regions, showing that conditioning provides a direct mechanism for enforcing spatial plausibility that conventional voxel-wise losses cannot guarantee.

While our implementation relies on FiLM-based feature modulation, which acts as a conservative form of error correction, the results indicate that even such a simple strategy is highly effective. This raises the possibility that more expressive conditioning mechanisms could further enrich feature representations while preserving stability. Similarly, our current formulation relies on distance maps from neighboring anchors, capturing relative proximity but not global position. Combining such relative cues with global anatomical encodings, in the spirit of positional embeddings, could provide complementary benefits. Finally, our experiments suggest that conditioning does not require perfect labels: pseudo-labels from TotalSegmentator were already sufficient to drive substantial gains, reinforcing the view that coarse anchors can be repurposed as powerful priors for downstream tasks.

In conclusion, AGC demonstrates that anatomical context offers a principled inductive bias for volumetric segmentation. By leveraging pseudo-label priors, encoding them as distance maps, and integrating them through feature-level modulation, AGC enhances boundary precision, reduces implausible predictions, and promotes anatomical coherence across diverse datasets and architectures.

REPRODUCIBILITY STATEMENT

All datasets used in this study are publicly available: ImageCAS Zeng et al. (2023), PET/CT VAT, HAN-Seg Podobnik et al. (2023), and Tooth-Fairy2 Bolelli et al. (2024). Preprocessing details, model configurations, and training settings are described in Section 3 and Supplementary Notes 1–2. Ablation and comparison studies in Section 4 and Supplementary Note 3 further clarify design choices. To ensure reproducibility, we include the basic implementation code and the corresponding segmentation results (in `.nii.gz` format) within the supplementary zip file.

REFERENCES

- M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, et al. The medical segmentation decathlon. *Nat. Commun.*, 13(1):4128, 2022.
- R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, et al. Medical image segmentation review: The success of U-Net. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):10076–10095, 2024.
- S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BraTS challenge. *arXiv:1811.02629*, 2018.
- P. R. A. S. Bassi, W. Li, Y. Tang, F. Isensee, Z. Wang, J. Chen, et al. Touchstone benchmark: Are we on the right way for evaluating AI algorithms for medical segmentation? *Adv. Neural Inf. Process. Syst.*, 37:15184–15201, 2024.
- Federico Bolelli, Luca Lumetti, Shankeeth Vinayahalingam, Mattia Di Bartolomeo, Arrigo Pellacani, Kevin Marchesini, Niels Van Nistelrooij, Pieter Van Lierop, Tong Xi, Yusheng Liu, et al. Segmenting the inferior alveolar canal in cbct volumes: the toothfairy challenge. *IEEE Transactions on Medical Imaging*, 2024.
- M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*, 2022.
- Ö. Cıçık, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Proc. MICCAI*, pp. 424–432, 2016.
- J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. *Med. Image Anal.*, 97:103280, 2024.
- B. K. Das, G. Zhao, S. Islam, T. J. Re, D. Comaniciu, E. Gibson, and A. Maier. Co-ordinate-Based positional embedding that captures resolution to enhance transformer’s performance in medical image analysis. *Sci. Rep.*, 14(1):9380, 2024.
- Y. Duan, J. Feng, J. Lu, and J. Zhou. Context aware 3D fully convolutional networks for coronary artery segmentation. In *Stat. Atlases Comput. Models Heart (STACOM), MICCAI Workshop*, pp. 85–93, 2019.
- M. Goncharov, V. Samokhin, E. Soboleva, R. Sokolov, B. Shirokikh, et al. Anatomical positional embeddings. In *Proc. MICCAI*, pp. 68–77, 2024.
- Dazhou Guo, Dakai Jin, Zhuotun Zhu, Tsung-Ying Ho, Adam P Harrison, Chun-Hung Chao, Jing Xiao, and Le Lu. Organ at risk segmentation for head and neck cancer using stratified learning and neural architecture search. In *Proc. CVPR*, pp. 4223–4232, 2020.
- A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In *BrainLes 2021 (MICCAI Workshop, Part I)*, volume 12962 of *LNCS*, pp. 272–284, 2021.
- A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, et al. UNETR: Transformers for 3D medical image segmentation. In *Proc. WACV*, pp. 574–584, 2022.

- X. Hu, F. Li, D. Samaras, and C. Chen. Topology-Preserving deep image segmentation. *Adv. Neural Inf. Process. Syst.*, 32, 2019.
- F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnU-Net: A Self-Configuring method for deep Learning-Based biomedical image segmentation. *Nat. Methods*, 18(2):203–211, 2021.
- Yannick Kirchoff, Maximilian R Rokuss, Saikat Roy, Balint Kovacs, Constantin Ulrich, Tassilo Wald, Maximilian Zenk, Philipp Vollmuth, Jens Kleesiek, Fabian Isensee, et al. Skeleton recall loss for connectivity conserving and resource efficient segmentation of thin tubular structures. In *Proc. ECCV*, pp. 218–234. Springer, 2024.
- Martin J. Menten, Johannes C. Paetzold, Veronika A. Zimmer, Suprosanna Shit, Ivan Ezhov, Robbie Holland, Monika Probst, Julia A. Schnabel, and Daniel Rueckert. A skeletonization algorithm for gradient-based optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 21394–21403, October 2023.
- B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging*, 34(10):1993–2024, 2015.
- F. Milletari, N. Navab, and S. A. Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proc. 3DV*, pp. 565–571, 2016.
- A. Myronenko. 3D MRI brain tumor segmentation using autoencoder regularization. In *BrainLes (MICCAI Workshop)*, volume 11384 of *LNC3*, pp. 234–244, 2019.
- Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proc. AAAI*, volume 32, 2018.
- G. Podobnik, P. Strojani, P. Peterlin, B. Ibragimov, and T. Vrtovec. HaN-Seg: The head and neck Organ-at-Risk CT and MR segmentation dataset. *Med. Phys.*, 50(3):1917–1927, 2023.
- S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, et al. cDice: A Topology-Preserving loss function for tubular structure segmentation. In *Proc. CVPR*, pp. 16560–16569, 2021.
- Zhengxuan Song, Xun Liu, Wenhao Zhang, Yongyi Gong, Tianyong Hao, and Kun Zeng. Spgnet: A shape-prior guided network for medical image segmentation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1263–1271, 2024. doi: 10.24963/ijcai.2024/140.
- J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, et al. TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiol.: Artif. Intell.*, 5(5):e230024, 2023.
- M. K. Wyburd, N. K. Dinsdale, M. Jenkinson, and A. I. L. Namburete. Anatomically plausible segmentations: Explicitly preserving topology through prior deformations. *Med. Image Anal.*, 97:103222, 2024.
- A. Zeng, C. Wu, G. Lin, W. Xie, J. Hong, et al. ImageCAS: A Large-Scale dataset and benchmark for coronary artery segmentation based on CTA. *Comput. Med. Imaging Graph.*, 109:102287, 2023.

A APPENDIX

USE OF LARGE LANGUAGE MODELS

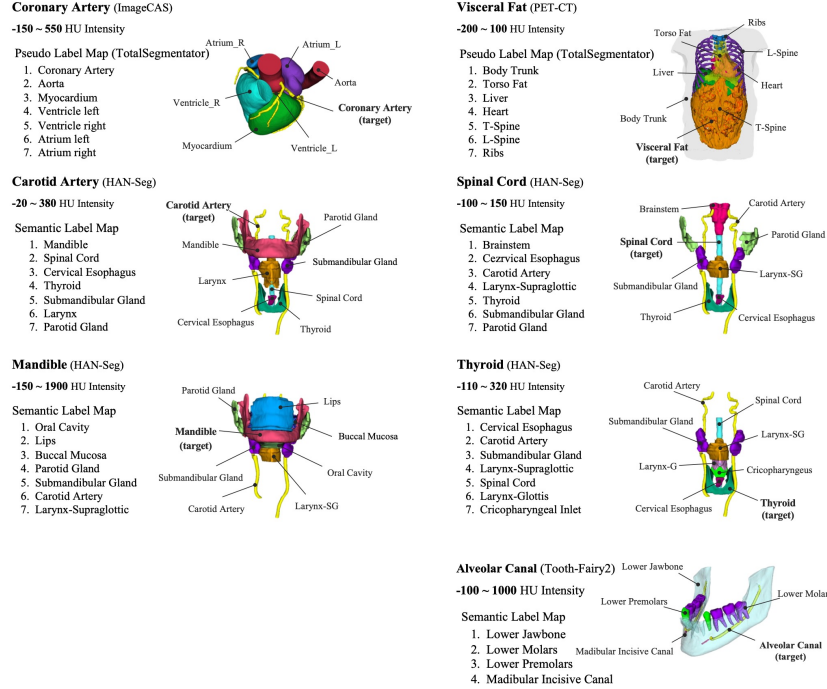
During manuscript preparation, we used a large language model (ChatGPT) to assist in polishing English and improving readability. The model was employed exclusively for grammar correction, style refinement, and phrasing suggestions. All scientific content, experimental design, implementation, analysis, and conclusions are the work of the authors.

Supplementary Materials

Title: 3D Medical Image Segmentation with Anatomy-Guided Conditioning from Surrounding Structures

B SUPPLEMENTARY NOTE 1: DATASET-SPECIFIC ANATOMICAL PRIORS

We visualize the anatomical priors used to guide segmentation across seven tasks: **Coronary Artery**, **Visceral Fat**, **Carotid Artery**, **Mandible**, **Spinal Cord**, **Thyroid**, and **Alveolar Canal**. For each task, the semantic label map, HU intensity range, and 3D rendering of anatomical structures are provided. For ImageCAS and PET-CT datasets, priors were generated by TotalSegmentator, while HAN-Seg and Tooth-Fairy2 used original dataset labels. These priors are injected into the network through Anatomy-Guided Conditioning (AGC).



Supplementary Figure 1: Anatomical segmentation masks used for AGC across datasets. Target and contextual organs are shown in task-specific pastel colors, highlighting their spatial relation.

C SUPPLEMENTARY NOTE 2: MODEL- AND LOSS-AGNOSTIC EVALUATION

To evaluate generality, we provide two complementary analyses. **(i) Model-agnostic evaluation:** Supplementary Table 1 reports results on three anatomically distinct targets (VAT, Carotid, Tooth-Fairy2) using SegResNet, UNETR, and Swin-UNETR. **(ii) Loss-agnostic evaluation:** Supplementary Table 2 reports results with three losses (DiceFocal, SkelRec, cIDice) on three datasets (ImageCAS, HAN-Seg, Tooth-Fairy2). In both cases, AGC improves overlap, boundary, and topology-aware metrics, with boldface marking improvements over baselines.

Supplementary Table 1: Model-agnostic evaluation of AGC on three datasets (VAT, Carotid, Tooth-Fairy2) with SegResNet, UNETR, and Swin-UNETR. Bold = improvement with AGC.

Dataset	Backbone	AGC	DSC $\uparrow$	HD95 $\downarrow$	$\beta_0 \downarrow$	$\beta_1 \downarrow$
VAT (PET-CT)	SegResNet (Conv)		0.94 $\pm$ 0.02	1.82 $\pm$ 1.55	137.4 $\pm$ 73.5	129.1 $\pm$ 70.0
		✓	0.96 $\pm$ 0.03	1.56 $\pm$ 1.05	128.6 $\pm$ 48.1	119.8 $\pm$ 44.6
	UNETR (Transf.)		0.89 $\pm$ 0.04	38.3 $\pm$ 21.8	961.3 $\pm$ 509.5	927.9 $\pm$ 494.3
		✓	0.95 $\pm$ 0.02	1.62 $\pm$ 1.19	159.1 $\pm$ 79.1	150.3 $\pm$ 74.1
Carotid (HAN-Seg)	Swin-UNETR (Hybrid)		0.88 $\pm$ 0.08	194.4 $\pm$ 27.6	977.5 $\pm$ 262.0	928.0 $\pm$ 246.3
		✓	0.96 $\pm$ 0.02	1.75 $\pm$ 2.08	176.7 $\pm$ 79.1	167.7 $\pm$ 74.2
	SegResNet (Conv)		0.65 $\pm$ 0.07	269.6 $\pm$ 74.1	36.6 $\pm$ 20.2	33.3 $\pm$ 19.5
		✓	0.73 $\pm$ 0.04	19.8 $\pm$ 5.15	6.25 $\pm$ 2.13	5.50 $\pm$ 1.94
Tooth-Fairy2 (Alveolar Canal)	UNETR (Transf.)		0.38 $\pm$ 0.06	437.1 $\pm$ 37.7	1618.3 $\pm$ 770.0	1459.8 $\pm$ 663.2
		✓	0.55 $\pm$ 0.07	134.0 $\pm$ 164.7	78.9 $\pm$ 44.8	73.2 $\pm$ 42.2
	Swin-UNETR (Hybrid)		0.27 $\pm$ 0.06	546.3 $\pm$ 38.9	561.5 $\pm$ 138.8	546.8 $\pm$ 137.8
		✓	0.75 $\pm$ 0.05	27.0 $\pm$ 28.1	12.4 $\pm$ 5.53	11.2 $\pm$ 5.24
Tooth-Fairy2 (Alveolar Canal)	SegResNet (Conv)		0.888 $\pm$ 0.016	2.56 $\pm$ 6.89	1.82 $\pm$ 1.41	1.64 $\pm$ 1.32
		✓	0.982 $\pm$ 0.012	0.69 $\pm$ 0.77	0.45 $\pm$ 0.84	0.36 $\pm$ 0.69
	UNETR (Transf.)		0.737 $\pm$ 0.076	22.0 $\pm$ 34.9	85.8 $\pm$ 22.6	82.6 $\pm$ 21.9
		✓	0.966 $\pm$ 0.006	1.26 $\pm$ 1.07	3.88 $\pm$ 8.27	3.84 $\pm$ 8.22
Tooth-Fairy2 (Alveolar Canal)	Swin-UNETR (Hybrid)		0.889 $\pm$ 0.029	1.85 $\pm$ 1.63	1.43 $\pm$ 1.58	1.28 $\pm$ 1.47
		✓	0.966 $\pm$ 0.005	1.01 $\pm$ 0.05	0.44 $\pm$ 0.75	0.42 $\pm$ 0.72

Supplementary Table 2: Loss-agnostic evaluation of AGC across three datasets (ImageCAS, HAN-Seg, Tooth-Fairy2). Bold = improvement with AGC.

Dataset	Loss	AGC	DSC $\uparrow$	HD95 $\downarrow$	clDice $\uparrow$	$\beta_0 \downarrow$	$\beta_1 \downarrow$
ImageCAS (Coronary)	DiceFocal		0.766 $\pm$ 0.032	81.8 $\pm$ 29.1	0.816 $\pm$ 0.077	25.2 $\pm$ 8.71	22.7 $\pm$ 8.30
		✓	0.788 $\pm$ 0.021	36.4 $\pm$ 12.0	0.862 $\pm$ 0.054	10.6 $\pm$ 3.94	10.0 $\pm$ 3.99
	SkelRec		0.778 $\pm$ 0.030	40.6 $\pm$ 17.8	0.856 $\pm$ 0.056	10.3 $\pm$ 5.13	9.77 $\pm$ 4.90
		✓	0.778 $\pm$ 0.030	39.2 $\pm$ 14.7	0.845 $\pm$ 0.071	8.90 $\pm$ 3.89	8.32 $\pm$ 3.70
HAN-Seg (Carotid)	clDice		0.787 $\pm$ 0.033	27.9 $\pm$ 16.2	0.855 $\pm$ 0.057	18.1 $\pm$ 8.10	17.1 $\pm$ 7.77
		✓	0.794 $\pm$ 0.023	27.2 $\pm$ 10.9	0.872 $\pm$ 0.051	7.19 $\pm$ 3.83	6.65 $\pm$ 3.58
	DiceFocal		0.646 $\pm$ 0.072	269.6 $\pm$ 74.1	0.646 $\pm$ 0.086	36.6 $\pm$ 20.2	33.3 $\pm$ 19.5
		✓	0.727 $\pm$ 0.044	19.8 $\pm$ 5.15	0.789 $\pm$ 0.044	6.25 $\pm$ 2.13	5.50 $\pm$ 1.94
Tooth-Fairy2 (Alveolar Canal)	SkelRec		0.671 $\pm$ 0.060	108.2 $\pm$ 73.2	0.741 $\pm$ 0.105	36.5 $\pm$ 31.8	32.8 $\pm$ 26.4
		✓	0.698 $\pm$ 0.062	111.5 $\pm$ 230.3	0.797 $\pm$ 0.192	14.9 $\pm$ 16.5	13.9 $\pm$ 15.8
	clDice		0.671 $\pm$ 0.067	39.3 $\pm$ 31.6	0.801 $\pm$ 0.123	21.0 $\pm$ 20.2	17.8 $\pm$ 16.6
		✓	0.673 $\pm$ 0.048	9.28 $\pm$ 4.87	0.842 $\pm$ 0.073	7.00 $\pm$ 4.43	6.00 $\pm$ 3.65
Tooth-Fairy2 (Alveolar Canal)	DiceFocal		0.888 $\pm$ 0.016	2.56 $\pm$ 6.89	0.987 $\pm$ 0.018	1.82 $\pm$ 1.41	1.64 $\pm$ 1.32
		✓	0.982 $\pm$ 0.012	0.69 $\pm$ 0.77	0.994 $\pm$ 0.009	0.45 $\pm$ 0.84	0.36 $\pm$ 0.69
	SkelRec		0.880 $\pm$ 0.015	1.54 $\pm$ 0.32	0.994 $\pm$ 0.009	0.26 $\pm$ 0.48	0.24 $\pm$ 0.43
		✓	0.988 $\pm$ 0.006	0.36 $\pm$ 0.60	0.998 $\pm$ 0.004	0.14 $\pm$ 0.35	0.12 $\pm$ 0.33
Tooth-Fairy2 (Alveolar Canal)	clDice		0.891 $\pm$ 0.012	1.48 $\pm$ 0.39	0.994 $\pm$ 0.011	0.24 $\pm$ 0.49	0.18 $\pm$ 0.45
		✓	0.989 $\pm$ 0.006	0.31 $\pm$ 0.34	0.997 $\pm$ 0.005	0.24 $\pm$ 0.51	0.22 $\pm$ 0.50

D SUPPLEMENTARY NOTE 3: ABLATION AND COMPARISON STUDY

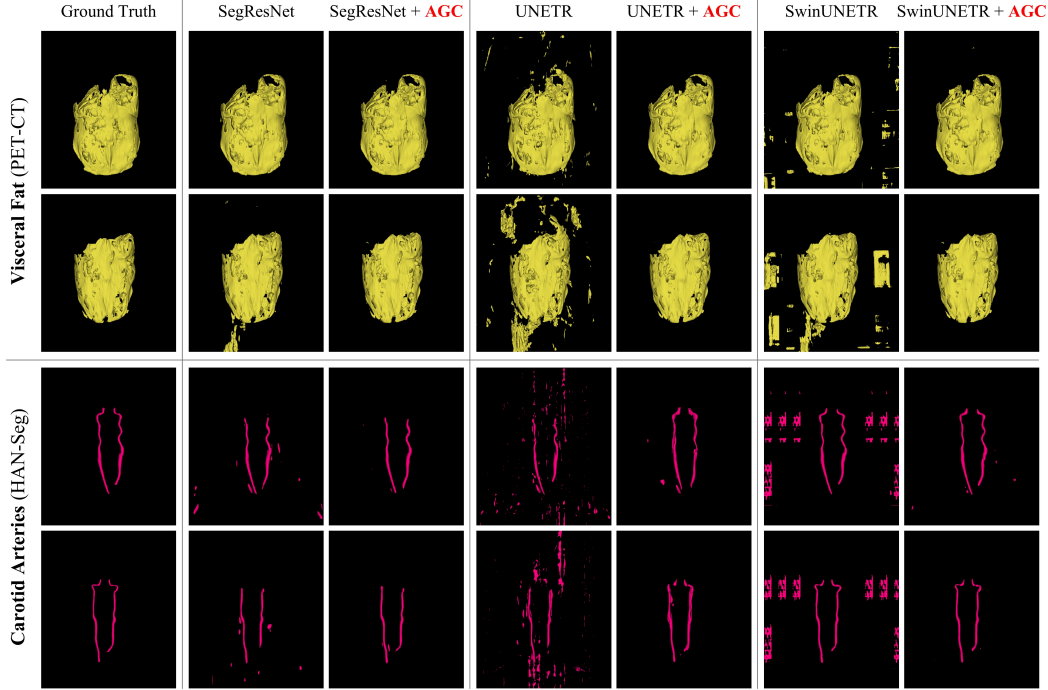
Supplementary Table 3 quantitatively ablates AGC and largest-connected-component pruning (LCC), showing that AGC raises structural accuracy, LCC removes disconnected false positives, and the two combined deliver the best Dice and HD95.

Supplementary Table 3: Quantitative comparison of segmentation performance across four strategies: (1) no conditioning and no post-processing, (2) LCC post-processing only, (3) AGC only, and (4) AGC combined with LCC. All experiments were conducted using the SegResNet for clarity.

Dataset	AGC	LCC	DSC $\uparrow$	HD95 $\downarrow$
Coronary Artery (ImageCAS)			0.7658 $\pm$ 0.0321	81.8044 $\pm$ 29.0627
			0.7822 $\pm$ 0.0601	34.5192 $\pm$ 28.9655
	✓		0.7883 $\pm$ 0.0208	36.4332 $\pm$ 12.0385
	✓	✓	0.7945 $\pm$ 0.0570	29.6120 $\pm$ 27.3706
Visceral Adipose Fat (PET-CT VAT)			0.9444 $\pm$ 0.0202	1.8175 $\pm$ 1.5481
		✓	0.9400 $\pm$ 0.0392	3.4425 $\pm$ 9.8962
	✓		0.9564 $\pm$ 0.0255	1.5603 $\pm$ 1.0483
	✓	✓	0.9565 $\pm$ 0.0267	1.7725 $\pm$ 2.2565
Carotid Artery (HAN-Seg)			0.6459 $\pm$ 0.0722	269.6247 $\pm$ 74.0853
		✓	0.7276 $\pm$ 0.0616	26.8810 $\pm$ 9.1371
	✓		0.7273 $\pm$ 0.0435	19.7951 $\pm$ 5.1479
	✓	✓	0.7035 $\pm$ 0.0745	34.1748 $\pm$ 16.6633
Spinal Cord (HAN-Seg)			0.7282 $\pm$ 0.0708	235.0251 $\pm$ 132.4872
		✓	0.7617 $\pm$ 0.0362	5.8269 $\pm$ 3.7512
	✓		0.7385 $\pm$ 0.0547	162.9246 $\pm$ 271.1081
	✓	✓	0.7503 $\pm$ 0.0403	6.4769 $\pm$ 3.9617
Mandible (HAN-Seg)			0.8446 $\pm$ 0.0393	261.8987 $\pm$ 31.0779
		✓	0.9176 $\pm$ 0.0189	2.1617 $\pm$ 1.0773
	✓		0.9150 $\pm$ 0.0198	3.0604 $\pm$ 4.1141
	✓	✓	0.9177 $\pm$ 0.0171	1.7538 $\pm$ 0.4173
Thyroid (HAN-Seg)			0.7782 $\pm$ 0.0925	215.6387 $\pm$ 129.9535
		✓	0.8008 $\pm$ 0.0747	64.2508 $\pm$ 93.7234
	✓		0.8131 $\pm$ 0.0811	6.5673 $\pm$ 4.6229
	✓	✓	0.8134 $\pm$ 0.0809	6.5946 $\pm$ 4.6689

E SUPPLEMENTARY NOTE 4: ADDITIONAL RESULTS AND VISUALIZATION

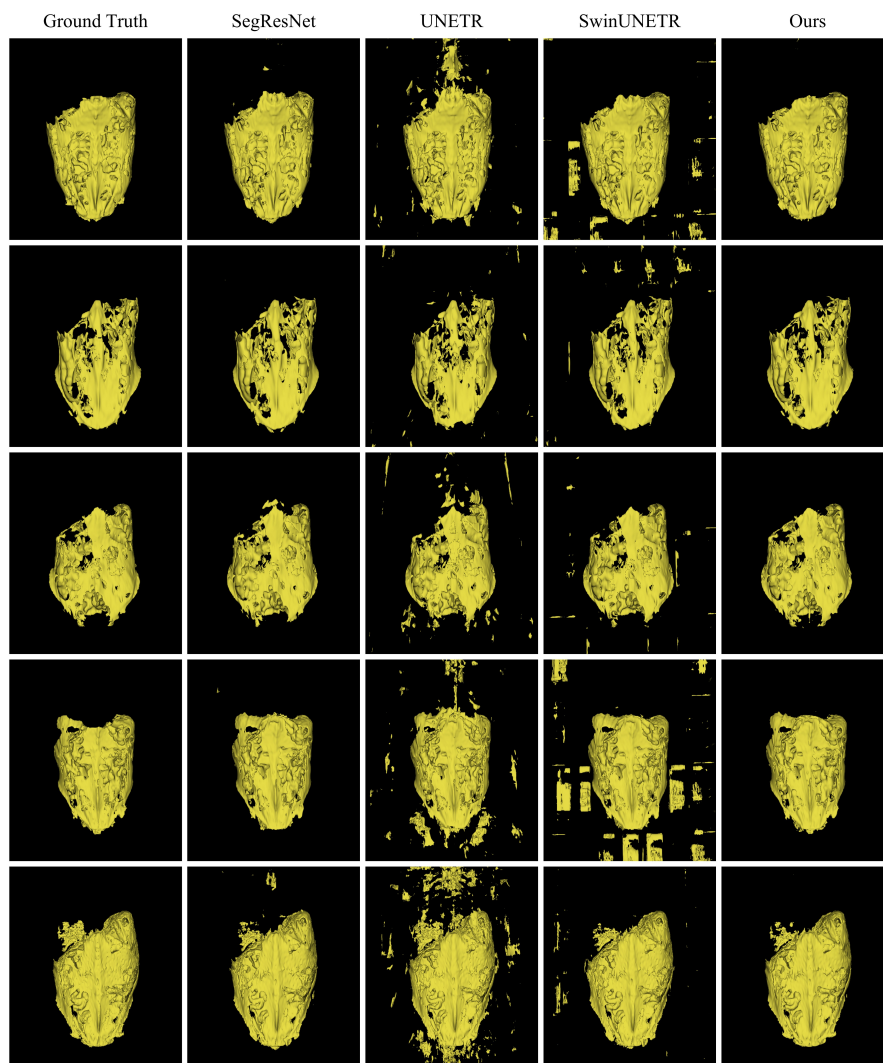
We show additional qualitative examples that could not be accommodated in the main paper. Supplementary Figure 3 demonstrates that AGC consistently improves segmentation across two external datasets and several backbones. Supplementary Figures 4–6 then provide multi-subject 3D renderings for each benchmark—ImageCAS (coronary artery), PET-CT (visceral fat), and HAN-Seg (carotid artery, spinal cord, mandible, thyroid)—allowing visual inspection of common failure modes (fragmentation, boundary leakage, false positives) and their correction by AGC.



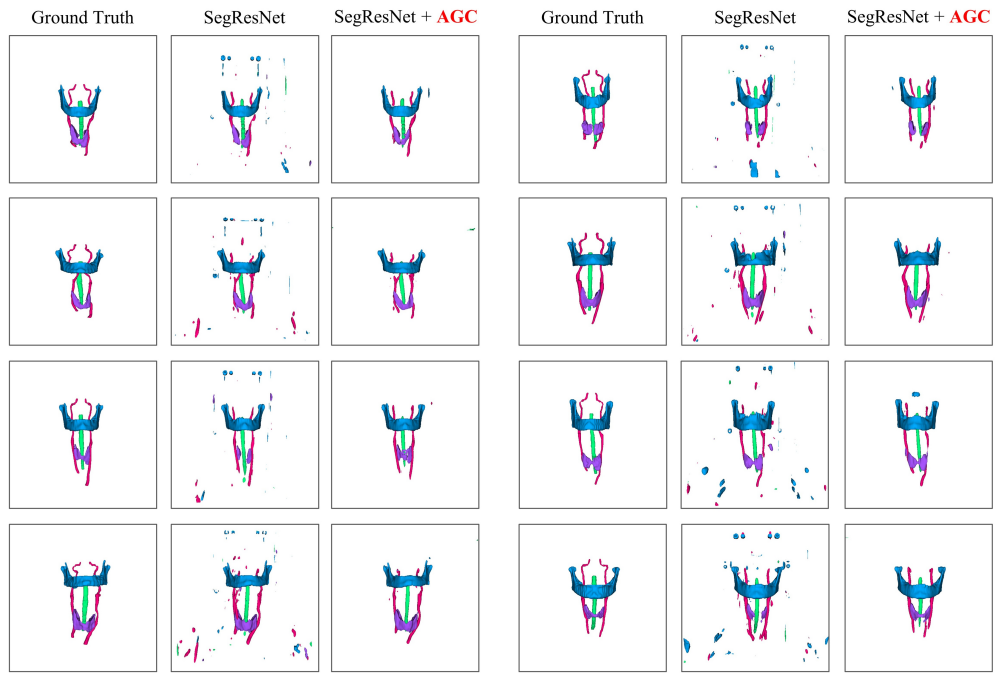
Supplementary Figure 3: Qualitative 3D visualization of segmentation results on two additional datasets, PET-CT (Visceral Fat) and HAN-Seg (Carotid Artery), using multiple segmentation backbones with and without AGC.



Supplementary Figure 4: Additional qualitative 3D visualization of coronary artery segmentation on multiple subjects from the ImageCAS dataset. Each row corresponds to a different subject, and each column shows the segmentation output from different segmentation backbones: Ground Truth, SegResNet, UNETR, SwinUNETR, CS²-Net, and our proposed method (SegResNet + AGC).



Supplementary Figure 5: Additional qualitative 3D visualization of visceral fat segmentation on multiple subjects from the PET-CT dataset. Each row corresponds to a different subject, and each column shows the segmentation output from different segmentation backbones: Ground Truth, SegResNet, UNETR, SwinUNETR, and our proposed method (SegResNet + AGC).



Supplementary Figure 6: Additional 3D visualization results on the HAN-Seg dataset for multiple subjects. This figure presents the segmentation outputs of four anatomical structures—carotid artery, spinal cord, mandible, and thyroid. For each subject, we show the ground truth, SegResNet baseline, and the prediction refined by our proposed AGC method.