Uncertainty-Aware Spatio-Semantic Contextual Prompts for Multimodal Medical Segmentation

Soumitri Chattopadhyay[1] Başar Demir[1] Marc Niethammer[1,2]

[1]Dept. of Computer Science and Engineering, University of California, San Diego
[2]Dept. of Neurological Surgery, University of California, San Diego
{sochattopadhyay, bdemir, mniethammer}@ucsd.edu

Abstract. The vast heterogeneity of medical imaging demands developing universal and modality-transferable segmentation models that can ideally work in low-data regimes. Although few-shot cross-domain, in-context learning, and promptable foundational models have emerged as promising data-efficient domain-agnostic solutions, they are all limited either in dimensionality (2D only), scalability (interactive prompting being too slow and iterative), or require re-training for each new task, limiting their general applicability. In this work, we address these limitations and propose a novel framework that harnesses the representational capabilities of foundational models to generate spatial and semantic contextual priors that holistically describe the target structure to be segmented. We also propose a confidence-weighted dynamic gating scheme to fuse these context maps into a single dense prompt, and re-purpose a frozen foundational segmentation model, SAM-Med3D, to predict segmentations using this fused representation instead of sparse points. Our framework is modality-agnostic, training-free, scalable, and enables rapid and robust universal segmentation. We validate our approach on two abdominal CT and MRI datasets under cross-modal and intra-modal settings, and show it outperforms existing state-of-the-art methods by significant margins. Source codes are available at: https://github.com/ucsdbiag/spatio-semantic-in-context-segmentation.

Keywords: Medical Image Segmentation · Cross-Modal · Promptable

1 Introduction

Deep learning models often struggle when applied to data that differs significantly from their training distribution [4,5,17,33]. This challenge is particularly pronounced in medical imaging, a domain studded with distribution heterogeneity, where datasets, for example, vary with respect to modaility (e.g., MRI vs. CT), sequence (T1w vs. T2w vs. FLAIR), scanner protocols, and resolutions [5,17] even when imaging the same anatomical region. Expert annotations of medical images is time-consuming and expensive, resulting in a scarcity of high-quality labeled datasets. Although anatomical structures are shared across different imaging modalities, models often fail to generalize effectively. Hence,

developing methods that enable domain transfer of segmentation knowledge is critical for medical imaging.

Recent work has explored three directions to enable segmentation knowledge transfer across domains. The first direction includes few-shot [23, 26] and cross-domain segmentation [1, 2, 26, 32, 35], which relies on training with labeled source and target domain datasets. Unsurprisingly, these approaches do not easily scale in the face of medical data heterogeneity, and mostly require re-training for each new domain pair. The second direction involves in-context learning [3, 28, 34] where a support (or source) image provides contextual guidance for a target region to be segmented. While these works have shown promise, they are currently limited to 2D and within-modality transfer only, hindering their adoption for 3D multi-modal segmentation. The third direction encompasses interactively promptable foundational models [6, 9, 12, 15, 22, 31] that are trained over large and diverse data distributions. However, although these models yield powerful representations for various tasks, interactive prompting requires manual intervention and therefore does not scale to large sample sizes; this is only feasible if they can be faithfully prompted automatically.

Effective cross-domain segmentation relies on *preserving structural and semantic correspondence between the source and the target domains*. Anatomical structures maintain consistent geometric and (to a certain extent) appearance properties across imaging modalities, enabling feature-based transfer. To validate this assumption, we conducted a pilot study measuring cosine similarity between SAM-Med3D [31] feature embeddings of abdominal volumes across three conditions: (1) uni-modal same-organ pairs: 0.72 ± 0.11, (2) cross-modal same-organ pairs: 0.65 ± 0.14, and (3) different-organ pairs: 0.53 ± 0.17 (where organ-specific features are extracted via masking the features). These numbers reveal that same-organ pairs exhibit significantly higher similarity than different-organ pairs, even across modalities. This confirms that (i) organs retain semantic consistency across imaging domains, and (ii) SAM-Med3D features may be suitable to provide semantic correspondence for cross-domain matching.

Hence, we propose a unified training-free framework that leverages both spatial and semantic contextual features to enable universal multi-modal segmentation of medical volumes. Our method leverages ***(i)*** support-to-query image registration [8, 29] to obtain a *geometry-preserving contextual region-of-interest* (ROI) on the query image (i.e., the image to be segmented); as well as ***(ii)*** feature similarity with respect to an ROI to estimate a *semantic-preserving context* of the target organ. To effectively fuse these ROIs into a unified dense estimate, we formulate an ***(iii)*** uncertainty-aware gating mechanism, which adaptively weights different regions of the estimated ROI. Finally, ***(iv)*** we re-purpose sparse-prompted foundational models (SAM-Med3D [31] specifically) and use our dense spatio-semantic fused map as prompt input, to segment the query image. We comprehensively evaluate our approach on *one/few-shot cross-modal* and *uni-modal* abdominal segmentation tasks [16, 18], significantly surpassing prior state-of-the-art works in segmentation accuracy.

The key contributions of our work are:

1. We propose a training-free, multimodal 3D medical segmentation framework that leverages rich **spatial** and **semantic** contextual priors, and an **uncertainty-aware** fusion scheme to obtain *high-quality domain transferable dense visual prompts* for segmentation.
2. Our method re-purposes manually prompted foundational segmentation models for *auto-promptable one/few-shot cross-domain segmentation*.
3. Our method outperforms prior universal segmentation methods by significant margins for both cross-modal and uni-modal segmentation, *establishing a new state-of-the-art* in generalizable 3D medical segmentation.

2 Methodology

Our framework synthesizes dense prompts for SAM-Med3D [31] by fusing spatial and semantic priors from a single annotated support volume. We extract spatial priors via deformable registration (Sec. 2.1) and semantic priors via cross-modality feature matching (Sec. 2.2), then combine them through entropy-based gating (Sec. 2.3) for segmentation prediction (Sec. 2.4). The overall framework is depicted in Figure 1.

2.1 Spatial Context Generation

We leverage image registration [7, 8, 24, 29] to capture cross-modal spatial interactions between the labeled support volume and the unlabeled query volume. Let $\mathbf{X}_s \in \mathbb{R}^{D\times H\times W}$ denote the support image with its corresponding ground-truth binary mask $\mathbf{Y}_s$, and let $\mathbf{X}_q$ denote the query image. In our training-free framework, we employ MultiGradICON [8], a multimodal foundational registration model, to estimate a deformation map $\Phi_{s\to q}$ between $\mathbf{X}_s$ and $\mathbf{X}_q$. Applying this transformation, we warp the support mask into the query coordinate space to obtain an initial pseudo-label: $\tilde{\mathbf{Y}}_q = \mathbf{Y}_s \circ \Phi_{s\to q}$. Since $\tilde{\mathbf{Y}}_q$ is a binary map and does not capture boundary uncertainty, we convolve it with a 3D Gaussian kernel G_σ (σ=2.0) and normalize the result to produce a soft spatial prior:

$$\mathbf{P}_{reg} = \frac{G_\sigma \circledast \tilde{\mathbf{Y}}_q}{\max(G_\sigma \circledast \tilde{\mathbf{Y}}_q)} \quad \text{where} \quad G_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}\right). \tag{1}$$

This smoothing operation *injects boundary uncertainty* into the spatial region-of-interest context. To enable compatibility with the continuous dense prompt representations expected by SAM-Med3D [31], we transform this soft spatial prior into logit space, obtaining the spatial context prompt

$$\mathbf{L}_{reg}(v_q) = \log\left(\frac{\mathbf{P}_{reg}(v_q)}{1-\mathbf{P}_{reg}(v_q)}\right). \tag{2}$$

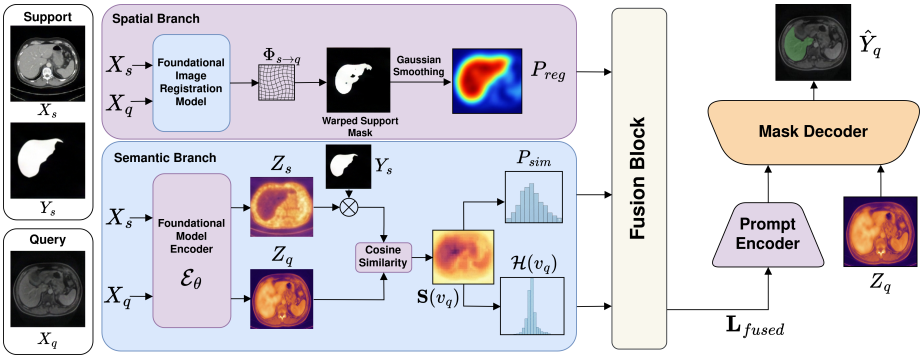

Fig. 1: Proposed uncertainty-aware spatio-semantic prompt fusion framework. We extract spatial and semantic priors from the support set and fuse them via an entropy-based dynamic gate, yielding a robust dense prompt for SAM-Med3D.

2.2 Semantic Context Generation

Complementary to the spatial prior, we extract semantic correspondences by matching query voxel features against target-class prototypes derived from the support volume. We use the frozen 3D ViT encoder $\mathcal{E}_\theta(\cdot)$ of SAM-Med3D [31] to extract l_2-normalized feature embeddings $\mathbf{Z}_s = \mathcal{E}_\theta(\mathbf{X}_s)$ and $\mathbf{Z}_q = \mathcal{E}_\theta(\mathbf{X}_q)$.

Next, we extract a set of foreground prototypes $\mathcal{F}$ via mask-pooling of support features: $\mathcal{F} = \{\mathbf{Z}_s(v) \mid \mathbf{Y}_s(v) > 0\}$. Let $N_p = |\mathcal{F}|$ denote the number of prototypes. For every query voxel feature $v_q \in \mathbf{Z}_q$, we compute the cosine similarity $\texttt{sim}_j(v_q)$ with each prototype $\mathbf{f}_j \in \mathcal{F}$, and aggregate them into a unified semantic similarity map $\mathbf{S}$ via temperature-scaled softmax weighting

$$\alpha_j(v_q) = \frac{\exp(\texttt{sim}_j(v_q)/\tau_s)}{\sum_{k=1}^{N_p} \exp(\texttt{sim}_k(v_q)/\tau_s)}, \quad \mathbf{S}(v_q) = \sum_{j=1}^{N_p} \alpha_j(v_q) \cdot \texttt{sim}_j(v_q), \tag{3}$$

where $\tau_s = 0.1$ and the weights α_j act as soft attention over prototypes. The resulting map $\mathbf{S} \in [-1, 1]$ is rescaled to $[0, 1]$ via $\mathbf{P}_{sim}(v_q) = \texttt{clamp}\left[\frac{\mathbf{S}(v_q)+1}{2}, 0, 1\right]$ and converted to logit space using Equation 2 to form the semantic context $\mathbf{L}_{sim}$.

2.3 Uncertainty-Aware Spatio-Semantic Contextual Fusion

Averaging of $\mathbf{L}_{reg}$ and $\mathbf{L}_{sim}$ may be suboptimal for anatomically homogeneous regions where texturally similar structures yield ambiguous semantic matches. We address this via an uncertainty-aware gating mechanism that dynamically weights each prior according to its voxel-wise distributional confidence.

Voxel-wise Entropy: We compute per-voxel uncertainty using normalized Shannon Entropy [21] of the cosine similarity distribution. Specifically, we compute a softmax over the semantic similarity values $\{\texttt{sim}_j(v_q)\}_{j=1}^{N_p}$, using a lower

temperature $\tau_e = 0.07$ ($\tau_e < \tau_s$), followed by entropy computation ($\mathcal{H}$) as

$$\omega_j(v_q) = \frac{\exp(\mathtt{sim}_j(v_q)/\tau_e)}{\sum_{k=1}^{N_p} \exp(\mathtt{sim}_k(v_q)/\tau_e)}, \quad \mathcal{H}(v_q) = \frac{-\sum_{j=1}^{N_p} \omega_j(v_q) \cdot \log \omega_j(v_q)}{\log N_p}. \quad (4)$$

A lower temperature τ_e *sharpens the predicted similarity distribution*, suppressing spurious uncertainty to ensure $\mathcal{H}(v_q)$ selectively isolates genuine semantic ambiguity [10,13,20]. Consequently, while $\mathbf{P}_{sim}$ captures the magnitude of the match, $\mathcal{H}$ provides a complementary measure of its distributional spread.

Uncertainty-Aware Gated Fusion: We synthesize the final dense context by fusing the spatial ($\mathbf{L}_{reg}$) and semantic ($\mathbf{L}_{sim}$) logits via the following adaptive gating mechanism

$$\mathcal{W}_{sim}(v_q) = 1 - \mathcal{H}(v_q), \quad \mathcal{W}_{reg}(v_q) = \mathbf{P}_{reg}(v_q) \cdot (\mathbf{P}_{sim}(v_q) + \epsilon), \quad (5)$$

where ϵ is a small stabilization constant. Functionally, $\mathcal{W}_{sim}$ actively suppresses the semantic context in regions characterized by high predictive uncertainty (i.e. where $\mathcal{H}(v_q) \to 1$). Conversely, $\mathcal{W}_{reg}$ propagates the structural confidence of the spatial context, but ensures that this propagation is strictly bounded by localized semantic correspondence. This product form enforces a conservative policy: the spatial logit receives full weight only when both priors concur, preventing the registration prior from dominating in regions of semantic disagreement.

The final fused logit map is the spatially weighted linear combination

$$\mathbf{L}_{fused}(v_q) = \mathcal{W}_{reg}(v_q) \cdot \mathbf{L}_{reg}(v_q) + \mathcal{W}_{sim}(v_q) \cdot \mathbf{L}_{sim}(v_q). \quad (6)$$

2.4 Segmentation Prediction via SAM-Med3D

The fused dense map $\mathbf{L}_{fused}$ (Eq. (6)) is encoded via SAM-Med3D's prompt encoder and decoded with query features $\mathbf{Z}_q$ to produce $\hat{\mathbf{Y}}_q \in \mathbb{R}^{D \times H \times W}$. We leverage SAM-Med3D's ***native support for dense mask inputs*** (used in its interactive refinement mode [31]) to enable fully automatic prompting.

3 Experiments and Results

Datasets. Following prior works [2,23,37] we evaluate our proposed method on two publicly available 3D medical imaging datasets spanning different modalities and organs, namely **Abdomen-MRI** comprising 20 T2-SPIR MR volumes from the ISBI 2019 Combined Healthy Organ Segmentation challenge [16], and **Abdomen-CT** comprising 30 3D CT images from the MICCAI 2015 Beyond The Cranial Vault challenge [18].

Implementation. We implement our method in PyTorch [27] on a single NVIDIA RTX A6000 GPU. We use the official SAM-Med3D [31] codebase and leverage UniGradICON/MultiGradICON [8, 29] for unimodal/multimodal registration. Segmentation is evaluated using the Dice Similarity Coefficient (DSC %).

Table 1: Quantitative Comparison (DSC % ↑) of **cross-modal** abdominal segmentation methods for both **CT→ MRI** and **MRI → CT** settings. The best value is shown in **bold** font, and the second-best value is underlined. nnUNet is trained on the target modality (CT or MR). It is not cross-modal and provides an empirical in-domain upper bound. "†" and "‡" markers refer to 2D and 3D methods, respectively.

Methods	Reference	Abdomen CT → MRI					Abdomen MRI → CT				
		Liver	LK	RK	Spleen	Mean	Liver	LK	RK	Spleen	Mean
‡nnUNet [14]	Nature'21	92.02	92.09	92.74	89.38	91.56	95.57	86.75	89.39	90.83	90.64
†PANet [32]	ICCV'19	39.24	26.47	37.35	26.79	32.46	40.29	30.61	26.66	30.21	31.94
†SSL-ALP [26]	TMI'22	70.74	55.49	67.43	58.39	63.01	71.38	34.48	32.32	51.67	47.46
†RPT [36]	MICCAI'23	49.22	42.45	47.14	48.84	46.91	65.87	40.07	35.97	51.22	48.28
†PATNet [19]	ECCV'22	57.01	50.23	53.01	51.63	52.97	75.94	46.62	42.68	63.94	57.29
†IFA [25]	CVPR'24	50.22	35.99	34.00	42.21	40.61	46.62	25.13	26.56	24.85	30.79
†APM-M [30]	NeurIPS'24	70.85	55.41	58.68	53.11	59.51	74.48	56.01	49.83	64.12	61.11
†RobustEMD [35]	AIIM'25	60.16	66.34	70.26	53.71	62.61	69.82	63.79	50.34	59.88	60.95
†FAMNet [2]	AAAI'25	73.01	57.28	74.68	58.21	65.79	73.57	57.79	61.89	65.78	64.75
†C-Graph [1]	TMI'25	70.92	73.69	82.51	64.18	72.83	69.60	70.00	63.95	65.23	67.20
‡**Ours** (DSC_{2D})	–	73.47	82.59	87.49	83.43	81.74	82.14	81.91	84.76	80.55	82.34
‡**Ours** (DSC_{3D})	–	**79.82**	**90.77**	**89.81**	**85.75**	**86.54**	**90.42**	**86.13**	**86.87**	**85.88**	**87.33**

Data processing. We followed the exact input processing as in SAM-Med3D [31]; all input volumes were resampled to $1.5mm \times 1.5mm \times 1.5mm$ space, with ROI cropping dimensions of $128 \times 128 \times 128$. Both CT and MR voxel intensities undergo Z-score normalization before being fed into the model.

Evaluations. We evaluate cross-domain segmentation transferability through: ***(i)*** cross-modal transfer (Sec. 3.1); ***(ii)*** uni-modal transfer (Sec. 3.2) and ***(iii)*** ablation studies (Sec. 3.3). Following prior work [1, 2, 37], we use **1-way-1-shot** evaluation (i.e., all organs are evaluated independently with only 1 support image per query), except for support set ablations. For each query, we sample 5 random support images ($= |\mathrm{D}| \times 5$ evaluations per organ per dataset D) and report mean DSC per organ, both slice-based (DSC_{2D}) and volumetric (DSC_{3D}).

SOTA Competitors. We compare against methods spanning cross-domain segmentation (IFA [25], RobustEMD [35], FAMNet [2], C-Graph [1], MAUP [37]) and in-context learning (UniverSeg [3], Tyche [28], AtlasSegFM [34]). Our work addresses the intersection of these paradigms – cross-domain in-context learning, which remains largely unexplored.

3.1 Cross-Modal Segmentation

We evaluate cross-modal transfer on abdominal organs (liver, left kidney, right kidney, spleen) in both the CT→MRI and MRI→CT directions. Table 1 shows that our method substantially outperforms all prior work, exceeding the second-best method (C-Graph [1]) by +14% (CT→MRI) and +20% (MRI→CT) mean DSC. These gains establish a new state-of-the-art for one-shot cross-modal segmentation, demonstrating the effectiveness of our spatio-semantic prompt fusion for cross-domain transfer. Note that our method shows average Dice scores not

Table 2: Quantitative Comparison (DSC % ↑) of **uni-modal** segmentation methods on **Abdomen-MRI** and **Abdomen-CT**. The best value is shown in **bold** font, and the second-best value is underlined. nnUNet is trained on the target modality (CT or MR) and provides an empirical in-domain upper bound for all models. "†" and "‡" markers refer to 2D and 3D methods, respectively.

Methods	Reference	Abdomen-MRI					Abdomen-CT				
		Liver	LK	RK	Spleen	Mean	Liver	LK	RK	Spleen	Mean
‡nnUNet [14]	Nature'21	92.02	92.09	92.74	89.38	91.56	95.57	86.75	89.39	90.83	90.64
†SSL-ALP [26]	TMI'22	76.10	81.92	85.18	72.18	78.84	78.29	72.36	71.81	70.96	73.35
†ADNet [11]	MedIA'22	82.11	73.86	85.80	72.29	78.51	77.24	72.13	79.06	63.48	72.97
†UniverSeg [3]	ICCV'23	67.27	58.80	47.83	39.16	53.27	77.36	32.45	46.28	36.61	48.17
†Tyche [28]	CVPR'24	61.20	51.37	77.92	40.56	57.76	68.78	18.16	38.53	15.59	35.26
†CoWPro [23]	ICPR'24	75.77	75.30	80.45	71.51	75.56	73.11	62.66	58.99	67.97	65.83
†MAUP [37]	MICCAI'25	78.16	58.23	72.34	59.65	67.09	78.25	59.41	71.80	60.38	67.46
†FAMNet [2]	AAAI'25	80.77	71.20	87.21	67.14	76.58	74.29	71.14	66.13	70.08	70.41
†C-Graph [1]	TMI'25	74.95	83.48	88.34	73.44	80.05	75.89	77.51	67.64	71.35	73.10
‡AtlasSegFM [34]	arXiv'25	83.77	81.11	83.54	77.83	81.22	90.06	63.27	62.50	75.79	72.91
‡**Ours** (DSC_{2D})	–	77.78	82.14	86.96	82.42	82.33	82.47	84.31	87.10	80.80	83.67
‡**Ours** (DSC_{3D})	–	**85.13**	**88.37**	**91.24**	**86.60**	**87.84**	**90.63**	**89.54**	**88.42**	**86.99**	**88.90**

too far from the empirical upper bound provided by an nnUNet model trained specifically on the target modality demonstrating that our auto-prompting approach works effectively across modalities.

3.2 Uni-Modal Segmentation

Following recent benchmarks [1,34,37], we evaluate within-dataset one-shot segmentation. Table 2 shows our method achieves the best performance on both CT and MRI, outperforming AtlasSegFM [34] and C-Graph [1] by +6% and +15% mean DSC, respectively. As expected, all methods improve under uni-modal transfer compared to the cross-modal setting (Table 1). Notably, while our CT liver gains are modest relative to AtlasSegFM, performance improvements on other organs are substantial, confirming robust generalization across diverse anatomical structures with minimal supervision.

Comparison with in-domain specialist *(upper bound).* We compare against a fully-supervised nnUNet [14] trained on the target domains (highlighted in pink in Tables 1 and 2) to quantify the domain gap. Our method approaches this upper bound within ≈ 4% mean DSC (without being trained on these target domains), substantially closer than all prior work. Notably, we even surpass nnUNet on CT left kidney segmentation (89.54% vs. 86.75%).

3.3 Ablation Study

Contribution of each component. We ablate individual components to validate our design choices. Table 3 compares: (a) spatial prompt alone, (b) semantic

Table 3: Ablation study on each component. We report mean 3D DSC % across all organs.

Spatial Branch	Semantic Branch	Uncertainty-gated Fusion	CT→MRI	MRI→MRI	MRI→CT	CT→CT
✓	✗	✗	79.27±5.87	80.30±5.40	64.43±9.82	84.51±6.38
✗	✓	✗	80.78±7.64	84.67±3.19	83.60±7.62	84.45±5.92
✓	✓	✗	84.48±5.53	86.40±3.08	86.08±2.36	87.68±1.65
✓	✓	✓	**86.54**±4.98	**87.84**±2.63	**87.33**±2.11	**88.90**±1.56

prompt alone, (c) naive mean fusion instead of our proposed dynamic gating, and (d) our full method with uncertainty-aware gating. Spatial prompting alone exhibits high variance in cross-modal settings due to registration challenges [8,29], while semantic matching provides a stronger baseline leveraging SAM-Med3D's multi-modal encoder [31]. Fusing both priors improves mean performance and reduces variance. Uncertainty-aware gating further enhances robustness, confirming that entropy-based modulation effectively resolves contextual ambiguity.

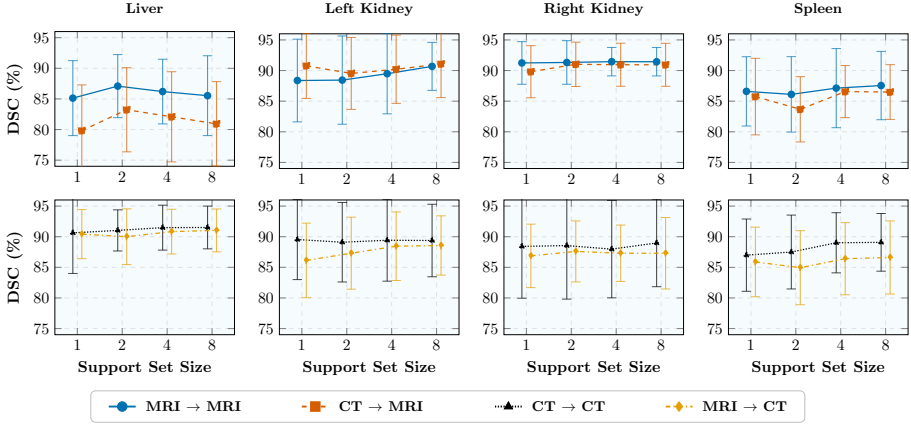

Fig. 2: Performance variation with increasing support set size for each organ.

Increasing support set sizes. We evaluate performance across $N \in \{1, 2, 4, 8\}$ support images (Figure 2). While additional support images do not yield monotonic improvement, performance remains stable with minor fluctuations. This plateau may be attributable to SAM-Med3D's fixed decoder capacity, which cannot fully exploit richer prompts beyond its representational bottleneck. Critically, our method achieves strong performance in the one and two shot settings, enabling rapid annotation scaling with minimal expert supervision.

4 Conclusion

We proposed a training-free segmentation approach that generates spatial and semantic correspondence-driven contextual prompts, and uses an uncertainty-aware gating to fuse these prompts. This fusion approach yields a high quality dense prompt which is then used to auto-prompt an interactive image segmentation foundation model. We validated our approach across cross-modal and uni-modal one-shot abdominal segmentation tasks, and showed that it outperforms all prior SOTA methods by large margins. Our approach reduces the gap with respect to an in-domain specialist trained nnUnet model, and is robust in ultra low-data regimes. Our work has immense potential to scale up data annotation. While we do not yet achieve the same performance of the specialist model our approach is generic: to segment a new structure we simply need to provide example segmentations, even if they are not provided for the same modality.

Acknowledgements. This research was, in part, funded by the National Institutes of Health (NIH) under other transactions 1OT2OD038045-01 and NIAMS 1R01AR082684. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the NIH.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Bo, Y., Zhou, T., Li, Z., Zhang, H., Shao, L.: Contrastive graph modeling for cross-domain few-shot medical image segmentation. TMI (2025)
2. Bo, Y., Zhu, Y., Li, L., Zhang, H.: FAMNet: Frequency-aware matching network for cross-domain few-shot medical image segmentation. In: AAAI Conference on Artificial Intelligence (2025)
3. Butoi, V.I., Ortiz, J.J.G., Ma, T., Sabuncu, M.R., Guttag, J., Dalca, A.V.: UniverSeg: Universal medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21438–21451 (2023)
4. Cekmeceli, K., Himmetoglu, M., Tombak, G.I., Susmelj, A., et al.: Do vision foundation models enhance domain generalization in medical image segmentation? arXiv preprint arXiv:2409.07960 (2024)
5. Chattopadhyay, S., Demir, B., Niethammer, M.: Zero-shot domain generalization of foundational models for 3D medical image segmentation: An experimental study. arXiv preprint arXiv:2503.22862 (2025)
6. Cheng, J., Ye, J., Deng, Z., Chen, J., et al.: Sam-med2d. arXiv preprint arXiv:2308.16184 (2023)
7. Demir, B., Niethammer, M.: Multimodal image registration guided by few segmentations from one modality. In: Medical Imaging with Deep Learning (2024)
8. Demir, B., Tian, L., Greer, H., Kwitt, R., et al.: multigradICON: A foundation model for multimodal medical image registration. In: MICCAI WBIR (2024)

9. Du, Y., Bai, F., Huang, T., Zhao, B.: SegVol: Universal and interactive volumetric medical image segmentation. In: NeurIPS (2024)
10. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: ICML. pp. 1321–1330. PMLR (2017)
11. Hansen, S., Gautam, S., Jenssen, R., Kampffmeyer, M.: Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Medical Image Analysis **78**, 102385 (2022)
12. He, Y., Guo, P., Tang, Y., Myronenko, A., et al.: Vista3d: Versatile imaging segmentation and annotation model for 3D computed tomography. arXiv preprint arXiv:2406.05285 (2024)
13. Hendrycks, D., Gimpel, K.: A baseline for detecting misclassified and out-of-distribution examples in neural networks. In: ICLR (2017)
14. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods (2021)
15. Isensee, F., Rokuss, M., Krämer, L., Dinkelacker, S., et al.: nnInteractive: Redefining 3D promptable segmentation. arXiv preprint arXiv:2503.08373 (2025)
16. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., et al.: CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis (2021)
17. Kondrateva, E., Pominova, M., Popova, E., Sharaev, M., Bernstein, A., Burnaev, E.: Domain shift in computer vision models for MRI data analysis: an overview. In: SPIE ICMV (2021)
18. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: MICCAI multi-atlas labeling beyond the cranial vault–workshop and challenge. In: MICCAI multi-atlas labeling beyond cranial vault—workshop challenge (2015)
19. Lei, S., Zhang, X., He, J., Chen, F., Du, B., Lu, C.T.: Cross-domain few-shot semantic segmentation. In: ECCV. pp. 73–90 (2022)
20. Liang, S., Li, Y., Srikant, R.: Enhancing the reliability of out-of-distribution image detection in neural networks. In: ICLR (2018)
21. Lin, J.: Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory **37**(1), 145–151 (2002)
22. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications (2024)
23. Manna, S., Bhattacharya, S., Pal, U.: Correlation weighted prototype-based self-supervised one-shot segmentation of medical images. In: International Conference on Pattern Recognition. pp. 16–33. Springer (2024)
24. Modersitzki, J.: Numerical methods for image registration. Oxford University Press on Demand (2004)
25. Nie, J., Xing, Y., Zhang, G., Yan, P., Xiao, A., Tan, Y.P., Kot, A.C., Lu, S.: Cross-domain few-shot segmentation via iterative support-query correspondence mining. CVPR pp. 3380–3390 (2024)
26. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervised learning for few-shot medical image segmentation. IEEE Transactions on Medical Imaging **41**(7), 1837–1848 (2022)
27. Paszke, A., Gross, S., Massa, F., Lerer, A., et al.: PyTorch: An imperative style, high-performance deep learning library. In: NeurIPS (2019)
28. Rakic, M., Wong, H.E., Ortiz, J.J.G., Cimini, B.A., Guttag, J.V., Dalca, A.V.: Tyche: Stochastic in-context learning for medical image segmentation. In: CVPR. pp. 11159–11173 (2024)
29. Tian, L., Greer, H., Kwitt, R., Vialard, F.X., et al.: unigradICON: A foundation model for medical image registration. In: MICCAI (2024)

30. Tong, J., Zou, Y., Li, Y., Li, R.: Lightweight frequency masker for cross-domain few-shot semantic segmentation. Advances in Neural Information Processing Systems **37**, 96728–96749 (2024)
31. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., et al.: SAM-Med3D: a vision foundation model for general-purpose segmentation on volumetric medical images. IEEE Transactions on Neural Networks and Learning Systems (2025)
32. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: Few-shot image semantic segmentation with prototype alignment. In: CVPR. pp. 9197–9206 (2019)
33. Yoon, J.S., Oh, K., Shin, Y., Mazurowski, M.A., Suk, H.I.: Domain generalization for medical image analysis: A survey. arXiv preprint arXiv:2310.08598 (2023)
34. Zhang, Z., Yu, Y., Zhu, S., Aly, A., Gao, Y., Gu, N., Xue, Y.: Atlas is your perfect context: One-shot customization for generalizable foundational medical image segmentation. arXiv preprint arXiv:2512.18176 (2025)
35. Zhu, Y., Li, M., Ye, Q., Wang, S., Xin, T., Zhang, H.: RobustEMD: Domain robust matching for cross-domain few-shot medical image segmentation. Artificial Intelligence in Medicine **167**, 103197 (2025)
36. Zhu, Y., Wang, S., Xin, T., Zhang, H.: Few-shot medical image segmentation via a region-enhanced prototypical transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 271–280. Springer (2023)
37. Zhu, Y., Zhang, H.: MAUP: Training-free multi-center adaptive uncertainty-aware prompting for cross-domain few-shot medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 326–336. Springer (2025)