SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

FewSAMNet - A Hybrid SAM-CNN Framework for Semi-Supervised Few-Shot Segmentation and Multi-Institutional Generalization

Chetana Krishnan, PhD Student, University of Alabama at Birmingham; Ezinwanne Onuoha; Alex Hung; Kyung H. Sung; Harrison Kim

Introduction/Background

Prostate segmentation is vital for cancer diagnosis and treatment planning. However, current models struggle to generalize across institutions due to scanner variability and domain shifts. To overcome this, we introduce FewSAMNet, a hybrid SAM-CNN framework tailored for semi-supervised and few-shot segmentation. By fusing global attention with local feature encoding, FewSAMNet ensures robust performance across diverse, unseen clinical datasets.

Methods/Intervention

From Fig. 1, FewSAMNet employs a hybrid encoder-decoder architecture that integrates the global contextual capabilities of SAM's vision transformer with the local feature sensitivity of CNNs. The encoder branches extract parallel features: a CNN pipeline captures fine-grained spatial cues, while a SAM-based ViT branch models long-range dependencies. These representations are fused via a multi-scale attention and cross-attention mechanism. The fused features are passed to a hybrid decoder integrating SAM's prompt-guided mask decoder and a multi-scale CNN decoder path, enabling precise segmentation even in limited-label or cross-domain settings. A knowledge retention block further enhances generalization by preserving critical spatial-semantic context during the upsampling process. We used 45 UAB and 300 ProstateX 3D prostate MRI volumes, sliced axially for 2D training (UAB: 464/146; ProstateX: 2679/279 train/test). Images were intensity-normalized, resized to 128×128, and enhanced using histogram equalization and Gaussian smoothing to improve contrast and reduce noise. FewSAMNet was trained for 20 epochs using the Adam optimizer (Learning Rate = 1e-4) with weighted Dice loss. SAM-based pre-trained weights were used for initialization. A batch size of 2 and custom collate functions supported 2D slice-level training with mixed prompt supervision.

Results/Outcome

From Table 1, FewSAMNet achieves a 3.7–4.7% DSC gain and a 0.43–0.51 lower MSE in general settings ($p < 0.01$), and a 2.8–3.3% higher DSC with a 0.22–0.27 lower MSE in few-shot transfer ($p = 0.0105–0.0684$), confirming its robust cross-domain performance. From Fig. 2, the FewSAMNet model delineates critical boundaries across different datasets, highlighting its superior localization capability.

Conclusion

FewSAMNet combines prompt-guided attention with CNN precision, providing a scalable and real-world solution for cross-institutional prostate segmentation with minimal supervision.

Statement of Impact

FewSAMNet sets a new benchmark in prostate segmentation by combining substantial quantitative gains with visibly superior boundary precision, enabling reliable deployment across diverse clinical settings with limited annotations.

Uploaded File(s)

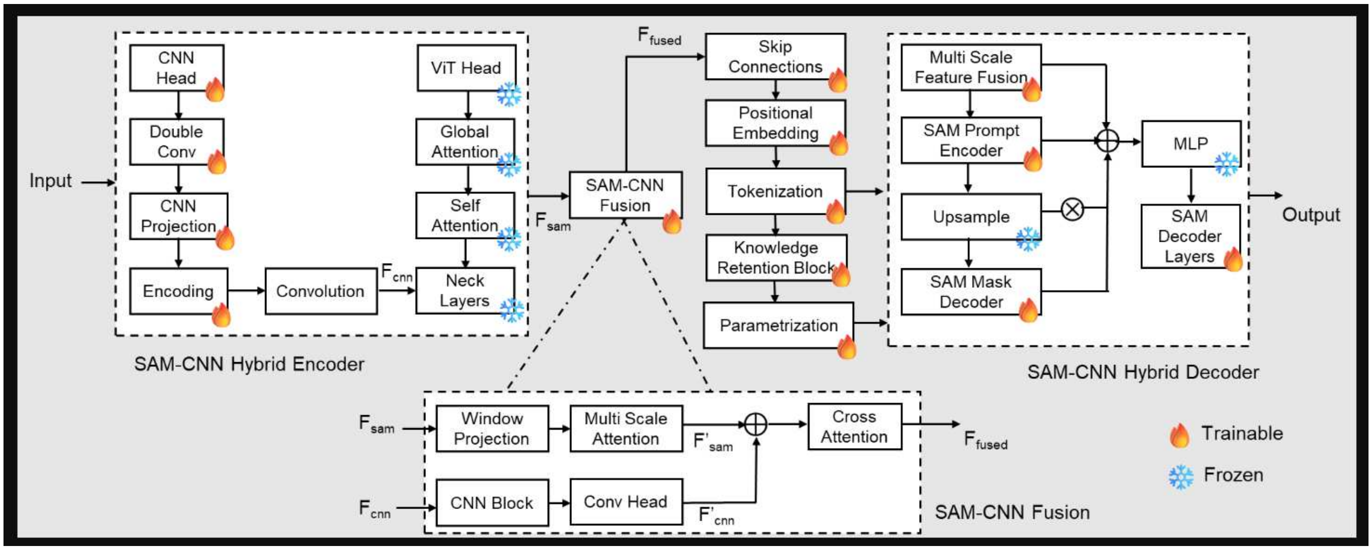

Figure 1. Architecture of the FewSAMNet Model. The proposed hybrid model integrates convolutional and attention-based mechanisms to enhance medical image segmentation. The architecture comprises three main components: the SAM-CNN Hybrid Encoder, which extracts both local (CNN-based) and global (ViT-based) features from the input image; the SAM-CNN Fusion module, which combines the CNN and SAM feature streams using multi-scale and cross-attention mechanisms; and the SAM-CNN Hybrid Decoder, which performs multi-scale feature fusion and mask decoding to generate the final segmentation output. The decoder also leverages skip connections and positional embeddings to retain spatial context.

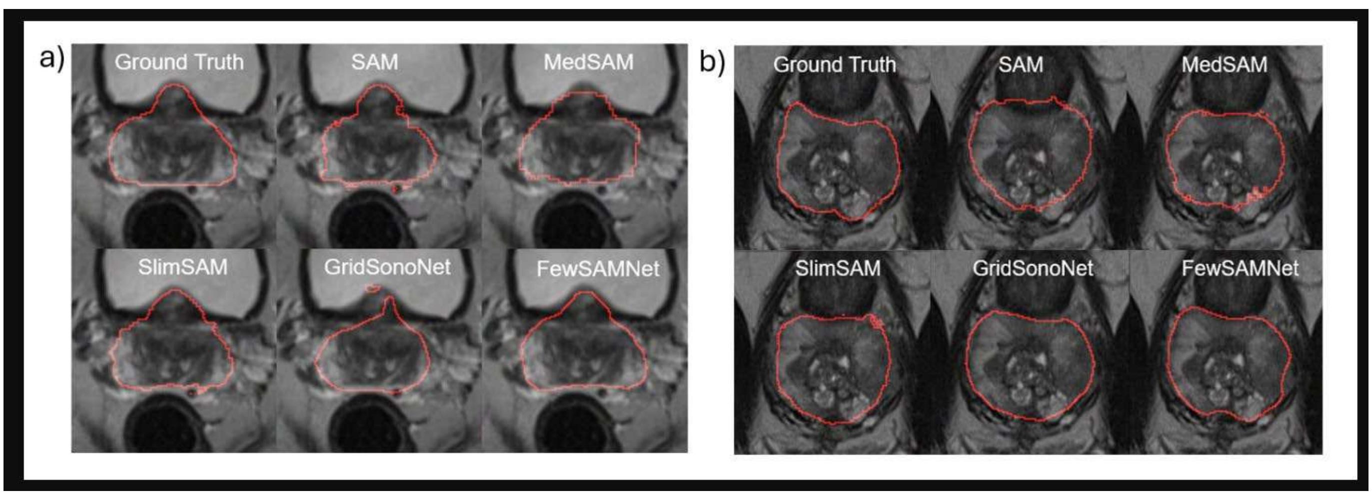

Figure 2. Segmentation Results of Few-Shot Learning of Various Foundational Models. Representative few-shot segmentation results by the proposed FewSAMNet model and four other models, along with the ground truth, when a) UAB data is finetuned and tested with ProstateX weights, and b) ProstateX data is finetuned and tested with UAB weights.

Model	General Performance				Few Shot Performance			
	UAB		ProstateX		UAB on X Weights		X on UAB Weights	
	DSC	MSE	DSC	MSE	DSC	MSE	DSC	MSE
SAM	0.92 ± 0.01	1.06 ± 0.23	0.92 ± 0.02	0.66 ± 0.14	0.94 ± 0.01	0.62 ± 0.06	0.90 ± 0.02	0.82 ± 0.20
MedSAM	0.93 ± 0.01	0.80 ± 0.12	0.91 ± 0.02	0.84 ± 0.22	0.92 ± 0.00	0.86 ± 0.14	0.92 ± 0.02	0.69 ± 0.16
SlimSAM	0.93 ± 0.01	0.83 ± 0.12	0.91 ± 0.02	0.76 ± 0.17	0.94 ± 0.01	0.73 ± 0.09	0.92 ± 0.02	0.65 ± 0.15
GridSonoNet	0.93 ± 0.01	0.80 ± 0.19	0.90 ± 0.03	0.96 ± 0.25	0.93 ± 0.01	0.84 ± 0.17	0.89 ± 0.04	1.00 ± 0.27
SonoNet	0.93 ± 0.01	0.81 ± 0.15	0.90 ± 0.03	1.00 ± 0.27	-	-	-	-
SwinUNETR	0.92 ± 0.02	0.97 ± 0.24	0.90 ± 0.03	0.95 ± 0.26	-	-	-	-
UNETR	0.88 ± 0.02	1.57 ± 0.34	0.77 ± 0.05	1.98 ± 0.46	-	-	-	-
ViT	0.87 ± 0.05	1.62 ± 0.50	0.86 ± 0.04	1.35 ± 0.43	-	-	-	-
nUNet	0.93 ± 0.01	0.87 ± 0.21	0.90 ± 0.04	0.92 ± 0.25	-	-	-	-
UNet	0.89 ± 0.02	1.42 ± 0.27	0.86 ± 0.05	1.21 ± 0.36	-	-	-	-
FewSAMNet	**0.95 ± 0.01**	**0.57 ± 0.09**	**0.93 ± 0.02**	**0.63 ± 0.19**	**0.96 ± 0.00**	**0.49 ± 0.09**	**0.94 ± 0.02**	**0.57 ± 0.17**

Table 1. Performance Comparison of Proposed FewSAMNet Hybrid Model with Baselines. Dice Similarity Coefficient (DSC) and Mean Surface Distance (MSD) on UAB and X datasets under general and few-shot settings. A few-shot evaluation on a different dataset than the one trained on (e.g., UAB data tested on a model trained with ProstateX) was conducted only for models with results comparable to those of the proposed method. The proposed model achieves the best performance across all metrics, highlighted in bold.

Keywords

Few-shot segmentation; Prostate MRI; SAM-CNN hybrid; Semi-supervised learning; Multi-institutional generalization