MedOpenSeg: Open-World Medical Segmentation with Memory-Augmented Transformers

MedOpenSeg: Open-World Medical Segmentation with Memory-Augmented Transformers

Luisa Vargas[1,*]
luisa.vargas@eurecom.fr
Eleonora Poeta[2,*,†]
eleonora.poeta@polito.it
Tania Cerquitelli[2]
tania.cerquitelli@polito.it
Elena Baralis[2]
elena.baralis@polito.it
Maria A. Zuluaga[1]
maria.zuluaga@eurecom.fr

[1] Eurecom
Sophia Antipolis, Biot, France
[2] Politecnico di Torino
Turin, Italy

Abstract

Open-world segmentation in medical imaging presents unique challenges, as models must generalize to seen and unseen classes while retaining knowledge of previously seen structures. We propose MedOpenSeg, a Memory-Augmented transformer framework that dynamically stores and updates class prototypes to enhance segmentation accuracy, improve adaptability to new anatomical structures, and detect novel regions during inference. MedOpenSeg integrates a Swin-Transformer 3D backbone with a memory bank module that retrieves class-specific feature embeddings and facilitates prototype-based novelty detection using cosine similarity and Euclidean Distance Sum (EDS). We benchmark MedOpenSeg on multiple datasets against state-of-the-art closed-set segmentation and foundation models, demonstrating its effectiveness in handling open-set medical segmentation. Code is publicly available at https://github.com/robustml-eurecom/MedOpenSeg.git.

1 Introduction

Semantic segmentation of medical images plays a crucial role in clinical decision-making, enabling precise delineation of anatomical structures and pathological regions. Recent advances in deep learning-based segmentation models, particularly Convolutional Neural Networks (CNNs) and transformer-based architectures, have led to state-of-the-art performance in various medical imaging tasks [6, 10, 22, 26]. However, the vast majority of these models operate under a closed-set assumption, where all semantic categories are known and fixed during training. This assumption breaks down in real-world clinical scenarios, which

* denotes equal contribution, †Work performed as a visiting student at Eurecom.

frequently present unseen anatomical variations, rare pathologies, or device artifacts not represented in the training set [22]. As a result, standard segmentation models underperform or misclassify in open-set scenarios, where new categories emerge at inference. When the segmentation task involves unfamiliar anatomies or labels, conventional deep-learning models often require retraining or fine-tuning, which is impractical in clinical environments due to computational cost. This highlights the need for flexible segmentation frameworks to detect and handle unseen classes without extensive retraining. In this work, we adopt the open-world medical segmentation setup: unknown categories may appear at test time, and the system should flag them while maintaining accuracy on known classes, without post-deployment learning. A detailed review of related paradigms open-set/OOD detection, prototype-memory mechanisms, universal, zero-shot segmentation, and promptable foundation models is provided in Sec. 2.

To address these challenges, we propose MedOpenSeg, a Memory-Augmented transformer framework designed for open-world medical image segmentation. Our method dynamically stores and updates class prototypes in a memory bank, enabling both robust segmentation of known anatomical structures and effective identification of novel regions at inference. Crucially, MedOpenSeg does not require model retraining or class prompts for unseen structures, instead relying on voxel-wise comparisons in a learned embedding space to score novelty and guide segmentation decisions. In summary, our key contributions are:

(1) **Memory-Augmented Segmentation:** via a memory bank module that dynamically stores and updates class prototypes, allowing robust segmentation of known categories while detecting novel ones.

(2) **Prototype-Based Learning:** with a loss function that optimizes feature embeddings by enforcing class consistency and prototype alignment, improving representation learning for known and unknown structures.

(3) **Prototype-Based Novelty Detection:** using voxel-wise novelty maps from cosine similarity and Euclidean Distance Sum (EDS), providing a continuous measure of feature divergence, and

(4) **Open-Set Segmentation Benchmarking:** where we evaluate MedOpenSeg on multiple datasets against state-of-the-art closed-set segmentation and foundation models, highlighting the key challenges in open-set medical segmentation.

We provide our source code and dataset splits to facilitate reproducibility and future research in open-set medical segmentation.

2 Related Work

Open-set and out-of-distribution (OOD) detection for medical segmentation. Open-set recognition and OOD detection aim to identify inputs that deviate from the training distribution while preserving performance on known classes; in medical imaging, recent surveys have examined this specifically for segmentation and highlighted persistent challenges under distribution shift [9, 28]. A common strategy is to attach post-hoc scores to features of a trained segmenter, e.g., energy-based scores [19], Mahalanobis distances [16], and several works adapt such detectors to medical segmentation [11, 13, 17]. While effective for detecting abnormalities, these approaches typically output image- or region-level flags and lack a prompt-free mechanism for producing dense, voxel-wise novelty maps integrated into a unified segmentation pipeline [27, 28].

Prototype and memory-augmented representations. Prototype learning and memory banks have shown strong benefits in classification, and few-shot and semi-supervised segmentation: class prototypes compactify intra-class variation, stabilize training, and support transfer by comparing embeddings to stored representatives [5, 30, 32]. In medical imaging, prototype-centric objectives and memory-enhanced schemes have improved data efficiency and generalization in few-shot and semi-supervised settings [1, 23]. However, most prior works assume a closed label space and do not explicitly convert prototype distances into an open-set decision mechanism [8]. MedOpenSeg leverages prototypes during training to shape the embedding space and at test time to produce training-free voxel-wise novelty scores, bridging representation learning and open-set inference.

Universal, zero-shot, and few-shot segmentation. Universal segmentation frameworks seek task- or label-agnostic models that generalize across organs, modalities, or institutions. CLIP-driven models integrate text embeddings to encode label semantics, enabling zero-shot extension by providing a new class name [18, 33]. Few-shot frameworks exploit limited support masks to adapt to new targets with minimal tuning. While compelling, these approaches generally assume either textual descriptors that align with anatomy or a small number of support annotations—assumptions that may fail for rare or ambiguous findings. Moreover, they typically lack an explicit unknown-class detector, focusing instead on transferring to named classes. Our approach targets the complementary setting where unknowns are unspecified and must be surfaced without text or support.

Promptable foundation models for medical segmentation. Segment Anything (SAM) [14] and its medical adaptations (e.g., MedSAM, SAM-Med2D) [2, 20, 21, 29] demonstrate strong generalization through prompt-conditioned inference (points/boxes/masks). Recent work has also reduced manual effort by automating prompt generation or propagation within SAM-based pipelines, including microscopy-focused systems for segmentation and tracking [3]. Despite these advances, SAM-family methods remain gated by the presence and policy of prompts at inference; absent a suitable prompt, they do not natively declare "unknown". In contrast, MedOpenSeg is prompt-free and produces an intrinsic novelty map, making it suitable as an autonomous unknown detector.

Compared to (i) OOD detectors that score anomalies but do not yield dense novelty maps integrated into the segmentation pipeline, (ii) prototype-memory methods that predominantly operate in closed-set regimes, (iii) universal and few-shot approaches that require class names or supports, and (iv) SAM-based pipelines, including those with automatic prompting that remain prompt-dependent, MedOpenSeg contributes a 3D prototype-memory segmentation framework with a training-free, voxel-wise novelty scorer for open-set medical segmentation.

3 Method

We propose MedOpenSeg, a Memory-Augmented transformer framework designed for open-set medical segmentation. Our approach integrates a Swin-Transformer backbone for volumetric segmentation [26], a memory module that dynamically stores and updates class prototypes, enabling novelty detection by comparing voxel-wise embeddings against learned prototypes. Figure 1 shows an overview of the proposed method.

Unlike conventional closed-set segmentation models operating within a fixed anatomical structure set, MedOpenSeg explicitly accounts for previously unseen structures by incorpo-

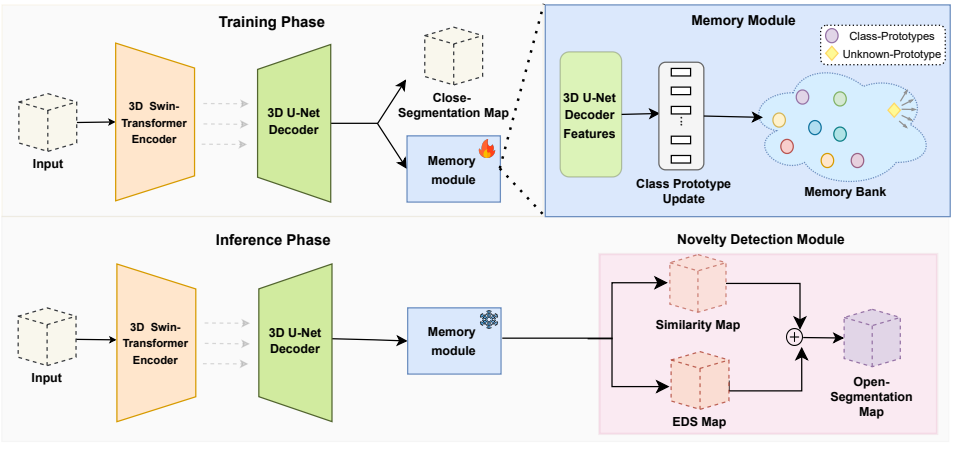

Figure 1: Overview of **MedOpenSeg**, a memory-augmented transformer framework for open-world medical segmentation. It integrates a Swin-Transformer encoder, a U-Net-based decoder, and a memory bank module. During training, the memory bank stores class prototypes from decoder embeddings. At inference, voxel-wise embeddings are compared against these prototypes to generate segmentation and novelty detection outputs, identifying novel structures based on feature divergence.

rating an adaptive memory mechanism. This design allows the model to perform segmentation as usual while simultaneously detecting novel anatomical regions based on their feature divergence from known prototypes.

3.1 Multi-Scale Feature Representation

MedOpenSeg adopts a hybrid architecture that integrates a Swin-Transformer based encoder with a U-Net-style decoder, enabling multi-scale feature representation and fine-grained semantic segmentation. Our encoder is based on the SwinUNETR architecture [26], which applies a hierarchical Swin-Transformer for multi-scale feature extraction. The input 3D volume is first partitioned into non-overlapping patches, which are processed by a patch embedding layer before passing through a sequence of shifted window self-attention mechanisms. This architecture allows the model to capture both local fine-grained and global contextual dependencies. As the encoder deepens, the spatial resolution is progressively reduced while the feature dimensionality increases, producing a hierarchical latent representation.

To restore spatial resolution and refine segmentation predictions, MedOpenSeg employs a U-Net-like decoder with skip connections that progressively integrate contextual information from the encoder while refining feature representations at multiple scales. Each decoder block consists of upsampling operations, transposed convolutions, and residual connections, ensuring a balance between coarse-to-fine reconstruction and detailed segmentation refinement. In addition to segmentation, MedOpenSeg incorporates an embedding projection layer at the final decoding stage. This projection is implemented via a $1 \times 1 \times 1$ convolutional layer, which maps the final segmentation feature maps into a compact embedding space. The resulting feature embeddings serve as input to the memory bank, where they are com-

pared against stored class prototypes. If a voxel's embedding significantly deviates from all known class prototypes, it is flagged as belonging to an unseen category, triggering novelty detection.

3.2 Memory Bank Module

The memory bank serves as a structured repository of learned class representations, storing a prototype $p_c \in \mathbb{R}^F$ for each class $c \in \mathcal{C}$, where $\mathcal{C}$ is the set of known classes, and F represents the feature dimensionality of the learned embedding space. Each prototype p_c acts as a centroid that captures the characteristic distribution of embeddings associated with class c, allowing MedOpenSeg to guide segmentation and flag novel anatomical structures.

During training, the memory bank is iteratively updated as embeddings for each class are observed. Specifically, a moving average update rule is applied:

$$p_c \leftarrow \alpha p_c + (1-\alpha) \cdot \bar{x}_c \tag{1}$$

where $\bar{x}_c$ is the mean embedding of class c in the current batch, and $\alpha \in [0,1]$ is a momentum term ensuring smooth adaptation. All prototypes are L2-normalized to stabilize training and enable reliable distance-based comparisons.

To model unseen anatomical regions, the memory bank reserves a dedicated prototype $\mathbf{p}_{\text{unk}}$ that is allocated for unknown structures. This prototype is optionally initialized using a single annotated instance of unseen structures, serving as an embedding anchor rather than as a supervised signal. Crucially, $\mathbf{p}_{\text{unk}}$ is not updated via backpropagation, but is adapted solely through unsupervised embedding statistics. This preserves the open-set assumption while providing a structured reference for novelty detection.

This strategy enables MedOpenSeg to benefit from minimal supervision while maintaining test-time autonomy. Notably, unlike few-shot segmentation methods [24] that require support-query supervision [25, 31], our approach uses no query-time labels and operates fully autonomously during inference.

It shares similarities with recent one-shot prototype-based methods such as ProtoSAM [4], yet distinguishes itself by explicitly avoiding any retraining or supervised adaptation on the unseen class. These design choices make MedOpenSeg suitable for practical clinical deployment, where manual annotations for novel conditions are scarce or unavailable.

3.3 Prototype-Guided Representation Loss

We introduce the *Prototype-Guided Representation (PGR) Loss* to enhance the discriminative power and consistency of learned voxel embeddings with respect to their corresponding class prototypes. The core idea is to explicitly encourage embeddings to tightly cluster around their respective prototypes, thereby ensuring strong representational robustness for seen and unseen anatomical structures. The PGR Loss comprises two complementary terms: the *Class Assignment (CA) Loss* and the *Prototype Consistency (PC) Loss*.

Class Assignment (CA) Loss: The CA Loss ensures accurate assignment of each voxel embedding to its correct class prototype by minimizing a softmax-normalized negative squared Euclidean distance. Formally, given a voxel embedding x with ground-truth class label c, we define the probability of voxel x belonging to class c as:

$$P(y = c \mid x) = \frac{\exp(-\|x - p_c\|^2)}{\sum_{c' \in \mathcal{C}} \exp(-\|x - p_{c'}\|^2)} \tag{2}$$

where $\mathcal{C}$ represents the set of known classes, and p_c denotes the learned prototype embedding of class c. The corresponding classification loss for voxel x is computed using the negative log-likelihood:

$$\mathcal{L}_{CA}(x, c) = -\log P(y = c \mid x) \tag{3}$$

Minimizing this loss encourages the embedding vectors to be closely aligned with their correct prototypes and promotes clear decision boundaries between different classes.

Prototype Consistency (PC) Loss: To further enhance representational coherence within each class, we introduce the PC Loss. This loss penalizes deviations of voxel embeddings from their corresponding class prototypes, explicitly enforcing tight clustering around each class prototype. Formally, it is defined as:

$$\mathcal{L}_{PC} = \frac{1}{|\mathcal{C}|} \sum c \in \mathcal{C} \frac{1}{|\mathcal{V}_c|} \sum_{x \in \mathcal{V}_c} \|x - p_c\|^2 \tag{4}$$

where $\mathcal{C}$ denotes the set of known classes, $\mathcal{V}_c$ represents the set of voxel embeddings belonging to class c, x is the embedding of a voxel belonging to class c, and p_c is the corresponding learned prototype embedding for class c. By enforcing low variance around each class prototype, this loss ensures compactness within classes, significantly improving intra-class consistency.

Finally, the overall PGR Loss function combines the CA Loss and PC Loss through a weighted summation, formally expressed as:

$$\mathcal{L}_{PGR} = \mathcal{L}_{CA} + \lambda_{PC} \mathcal{L}_{PC} \tag{5}$$

where λ_{PC} is a hyperparameter balancing the importance of accurate class assignment against intra-class representational compactness. We set $\lambda_{PC} = 0.01$ based on validation performance in our current implementation. For training, MedOpenSeg optimizes a combined objective that integrates PGR loss with Dice-Cross Entropy loss, ensuring accurate segmentation while refining the prototype-based representation space.

3.4 Novelty Detection Module

During inference, MedOpenSeg detects voxel-wise novelty by assessing how each voxel's embedding aligns with known class prototypes stored in the memory bank. Rather than relying on hard class predictions, the model computes two complementary voxel-wise novelty scores: one based on cosine similarity and the other on cumulative squared Euclidean distance. These maps allow the model to highlight potentially novel anatomical regions that deviate from known anatomical structures. The novelty head is parameter-free, introducing no learnable weights; at inference it reduces to batched matrix multiplications between the $(N \times F)$ embedding map and the $(F \times |\mathcal{C}|)$ prototype matrix, with memory $O(F|\mathcal{C}|)$ for the bank.

To compute the novelty maps, the feature embeddings extracted from the decoder are first projected into a lower-dimensional space using the embedding projection layer.

Let $x_v \in \mathbb{R}^F$ denote the L2-normalized embedding of voxel v, and let $p_c \in \mathbb{R}^F$ be the normalized prototype of class $c \in \mathcal{C}$. The cosine similarity between x and prototype p_c is computed as:

$$s_{v,c} = \frac{x_v \cdot p_c}{|x_v||p_c|}. \tag{6}$$

Since all embeddings and prototypes are L2-normalized, this reduces to a dot product. The maximum similarity to any class is then:

$$s_v = \max_{c \in \mathcal{C}} s_{v,c}. \tag{7}$$

This similarity score quantifies how closely voxel v aligns with the closest known class. A voxel is flagged as novel if its similarity score falls below a predefined threshold τ, i.e., if $s_v < \tau$. τ is set on a validation split to maximize performance; we keep it fixed across test cases.

To complement this local similarity score, we also compute the Euclidean Distance Sum (EDS), which captures the total deviation of the voxel embedding from all class prototypes:

$$\mathcal{S}_{EDS} = \sum_{c \in \mathcal{C}} \|x_v - p_c\|^2. \tag{8}$$

While cosine similarity focuses on alignment with the most likely prototype, EDS provides a holistic distance measure to the entire prototype manifold. Although both measures relate to L2-normalized vectors, they offer complementary insights: cosine similarity reflects local class membership confidence, while EDS captures global distributional shift. To generate a robust open-set segmentation map, we normalize both novelty maps and apply an adaptive threshold. Voxels with low cosine similarity and high EDS are flagged as potentially novel. This dual-criterion fusion improves sensitivity to subtle outliers and mitigates overconfidence in ambiguous regions. The cosine-based map emphasizes class alignment, while the EDS map captures deviations from learned distributions.

4 Experimental Setup

We evaluate MedOpenSeg on multiple 3D medical imaging datasets, covering MRI and CT modalities. Our experiments focus on open-world segmentation, where models are trained only on a subset of known anatomical categories and evaluated on their ability to detect both known and previously unseen structures at inference time.

Dataset Protocol: To assess the robustness and effectiveness of MedOpenSeg, we conduct experiments on three widely used 3D medical imaging datasets: AMOS 2022 [12], BTCV [15], and MSD-Pancreas [2]. Each dataset provides voxel-wise annotations for multiple abdominal organs, enabling a controlled evaluation of open-world segmentation. AMOS 2022 consists of 500 CT and 100 MRI scans from multi-center, multi-modality sources, with annotations for 15 abdominal organs. We train MedOpenSeg on 10 common organs (e.g., liver, kidneys, spleen) and evaluate its generalization to duodenum, prostate, bladder, and adrenal glands as unseen categories. BTCV contains 50 abdominal CT scans collected from patients with liver cancer or post-operative conditions, with 13 annotated organs. During training, we exclude the pancreas and adrenal glands and, at inference, evaluate MedOpenSeg's ability to recognize them as novel structures. Because these selected

categories are relatively small or thin and have shown lower segmentation performance in prior work, we anticipate lower AUROC on these held-out classes. To assess generalization to pathological regions, we include MSD-Pancreas, which provides 281 CT scans with annotations for the pancreas and pancreatic tumors. This dataset allows us to evaluate MedOpenSeg's ability to identify unseen tumor regions. Crucially, unseen classes are never presented to the model during training except in one controlled ablation ("*w/o Unknown Prototype*"), where a single instance is used to initialize an adaptive prototype. This prototype is updated only through latent embedding statistics and is not trained using segmentation supervision.

Evaluation Metrics: Open-set semantic segmentation combines closed-set segmentation and anomaly detection elements. We assess performance using the Dice Score (DSC) to evaluate segmentation accuracy on known structures. Area Under the ROC Curve (AUROC) measures the model's ability to distinguish novel from known anatomical regions [5].

Implementation Details: We implement MedOpenSeg in PyTorch and train the models on an NVIDIA A100 GPU for 30k iterations. We use an Adam optimizer with a learning rate of 1×10^{-4} to minimize the Dice-Cross entropy loss and the Prototype-Guided Representation Loss with a batch size of 6.

5 Results

5.1 Comparison with the State-of-the-Art

We compare MedOpenSeg with SwinUNETR [26], MedSAM [20], SAM-Med2D [7] and CLIP-Universal [18]. SwinUNETR is a strong transformer-based baseline trained in a closed-set setting, while MedSAM and SAM-Med2D leverage SAM [14] for medical image segmentation, incorporating foundation model pretraining. We evaluate two training regimes: *CW (closed-world)* trained with access to all categories (seen + unseen) to establish a supervised performance upper bound. *OW (open-world)* trained only on seen classes, following the open-world setting. This variant does not observe any information about unseen classes. MedOpenSeg is trained strictly on known classes. The unknown class prototype is not trained using labeled examples of unseen structures. Instead, it is either: (1) uninitialized (random), or (2) optionally initialized using a single one-shot example, and then adapted solely through statistics over the embedding space without backpropagation on unseen annotations. This preserves the open-set assumption and avoids leakage of segmentation supervision from unseen classes.

Table 1 presents the performance in both known class segmentation and novel class detection. MedOpenSeg achieves superior Dice scores on known classes and significantly higher AUROC on unseen structures. For instance, on AMOS, MedOpenSeg improves AUROC by 13% over the strongest baseline. This performance stems from the synergy between hierarchical encoding, prototype-guided representation learning, and training-free novelty detection using embedding-prototype distances.

MedOpenSeg consistently outperforms MedSAM and SAM-Med2D on known structure segmentation despite these baselines leveraging large-scale foundation model pretraining. We attribute this to our prototype-driven training strategy, which explicitly optimizes intra-class compactness and inter-class separation through the PGR loss.

In Figure 2, we present qualitative results illustrating MedOpenSeg's performance in closed-set segmentation and open-set novelty detection on a representative CT slice from

Table 1: **Performance Comparison.** MedOpenSeg outperforms other methods in known segmentation accuracy (Dice) and unseen class detection (AUROC). Best results in **bold** and percentage improvements compared to the best baseline.

Method	AMOS		BTCV		MSD-Pancreas	
	Known ↑	Unseen ↑	Known ↑	Unseen ↑	Known ↑	Unseen ↑
MedOpenSeg (Ours)	**90.11** ▲ 2%	**79.20** ▲ 13%	**88.18** ▲ 2%	**68.13** ▲ 3%	**88.96** ▲ 8%	**75.80** ▲ 8%
SwinUNETR (CW)	87.00	62.70	86.68	63.50	81.67	58.36
SwinUNETR (OW)	81.45	39.82	80.39	34.53	77.40	12.36
CLIP-Universal (OW)	88.51	52.70	83.17	42.70	82.43	49.60
MedSAM	67.81	70.24	64.40	66.34	58.86	70.45
SAM-Med2D	72.06	52.83	70.00	46.59	56.50	64.60

Figure 2: **Qualitative Results:** The left panel shows the ground truth segmentation overlaid on the CT scan. The middle panel displays the closed-set segmentation output of MedOpenSeg. The right panel illustrates the novelty map, where the model highlights previously unseen structures via the fused cosine/EDS novelty map.

the AMOS dataset. The left panel shows the ground truth segmentation overlaid on the original image, highlighting both seen and unseen anatomical structures. The middle panel displays the closed-set segmentation output from MedOpenSeg, which is trained exclusively on seen classes. As expected, the model correctly segments structures it was trained on (e.g., liver, kidney, spleen), while completely ignoring or misclassifying regions corresponding to unseen organs such as the bladder and prostate.

The right panel visualizes the novelty detection output using the fused prototype-based scoring mechanism. Warm colors (e.g., red and orange) indicate regions where voxel embeddings exhibit low similarity to all known class prototypes and high cumulative distance in embedding space, suggesting a distributional shift from the seen class manifold. Notably, these highlighted regions spatially correspond to the unseen anatomical structures present in the ground truth, validating the effectiveness of MedOpenSeg's novelty scoring mechanism.

5.2 Ablation Experiments

We conduct ablation studies to assess the impact of the PGR loss, the unknown prototype mechanism, and different novelty detection scoring strategies (Table 2). We observe that removing the PGR loss degrades both known-structure Dice and AUROC on unseen classes, confirming its role in learning well-clustered, discriminative prototypes. Similarly, removing

Table 2: **Ablation Study on MedOpenSeg.** Evaluation of key components affecting segmentation accuracy (Dice) and unseen class detection (AUROC). We analyze the impact of the Prototype-Guided Representation Loss (PGR), the unknown class prototype, and alternative novelty detection strategies. The best results for each setting are highlighted in **bold**.

	AMOS		BTCV		MSD-Pancreas	
Method	**Known ↑**	**Unseen ↑**	**Known ↑**	**Unseen ↑**	**Known ↑**	**Unseen ↑**
MedOpenSeg (Ours)	**90.11**	**79.20**	**88.18**	**68.13**	**88.96**	**75.80**
w/o PGR	85.6	76.45	87.09	65.34	83.24	74.90
w/o Unknown Prototype	88.01	45.23	87.97	39.47	86.75	28.90
Cosine Similarity Only	-	68.30	-	60.18	-	70.67
EDS Only	-	77.82	-	65.45	-	75.60

the unknown class prototype drastically reduces novelty detection performance (e.g., 45.23 AUROC on AMOS, 39.47 on BTCV, and 28.90 on MSD-Pancreas), demonstrating its importance for effective novelty detection. The unknown prototype can be optionally initialized from a single annotated instance. We acknowledge that this can bias the prototype toward that seed. To mitigate this, the seed is selected from a validation set disjoint from test data, no gradients are propagated through it, and the prototype is subsequently updated only via unsupervised embedding statistics. We also report a seed-free variant at the cost of lower AUROC. Among novelty detection strategies, Euclidean Distance Sum (EDS) outperforms Cosine Similarity alone, as it captures global divergence rather than alignment to a single prototype. Their combination yields the best overall AUROC.

6 Conclusions

This work presents MedOpenSeg, a Memory-Augmented transformer framework for open-world medical image segmentation. By integrating a hierarchical Swin-Transformer encoder with prototype-driven representation learning, MedOpenSeg achieves strong performance in both closed-set segmentation and open-set novelty detection. The key innovation lies in its prototype-guided representation loss (PGR) and training-free novelty scoring mechanism, which together enforce compact, class-aligned prototypes and enable the model to flag anatomically coherent novel regions without requiring retraining. Our experiments demonstrate that MedOpenSeg outperforms state-of-the-art methods, including vision-language pre-trained models, on standard benchmarks like AMOS, BTCV, and MSD-Pancreas.

Despite these strengths, MedOpenSeg has some limitations. In particular, its reliance on a dedicated unknown class prototype introduces a trade-off between performance and supervision. While the prototype is not trained with segmentation masks, it is initialized using a one-shot annotated example from an unseen class. This improves detection precision but introduces minimal supervision that may not be viable in all settings. In deployments where such a seed is unavailable, MedOpenSeg remains fully unsupervised at test time but with reduced sensitivity. To mitigate these limitations, future work will investigate fully unsupervised novelty detection, along with extensions to diverse imaging modalities and clinically relevant tasks, including rare pathologies and cross-domain generalization.

References

[1] Inigo Alonso, Alberto Sabater, David Ferstl, Luis Montesano, and Ana C Murillo. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8219–8228, 2021.

[2] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022.

[3] Anwai Archit, Luca Freckmann, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Carolin Teuber, Melanie Spitzner, Constanza Tapia Contreras, et al. Segment anything for microscopy. *Nature Methods*, 22(3):579–591, 2025.

[4] Lev Ayzenberg, Raja Giryes, and Hayit Greenspan. Protosam: One-shot medical image segmentation with foundational models. *arXiv preprint arXiv:2407.07042*, 2024.

[5] Jun Cen, Peng Yun, Junhao Cai, Michael Yu Wang, and Ming Liu. Deep metric learning for open world semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15333–15342, 2021.

[6] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[7] Junlong Cheng, Jin Ye, Zhongying Deng, Jianpin Chen, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyan Huang, Jilong Chen, Lei Jiang, et al. Sam-med2d. *arXiv preprint arXiv:2308.16184*, 2023.

[8] Theekshana Dissanayake, Yasmeen George, Dwarikanath Mahapatra, Shridha Sridharan, Clinton Fookes, and Zongyuan Ge. Few-shot learning for medical image segmentation: A review and comparative study. *ACM Computing Surveys*, 2025.

[9] Zesheng Hong, Yubiao Yue, Yubin Chen, Lele Cong, Huanjie Lin, Yuanmei Luo, Mini Han Wang, Weidong Wang, Jialong Xu, Xiaoqi Yang, et al. Out-of-distribution detection in medical image analysis: A survey. *arXiv preprint arXiv:2404.18279*, 2024.

[10] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[11] Masoumeh Javanbakhat, Md Tasnimul Hasan, and Cristoph Lippert. Assessing uncertainty estimation methods for 3d image segmentation under distribution shifts. *arXiv preprint arXiv:2402.06937*, 2024.

[12] Yuanfeng Ji, Haotian Bai, Jie Yang, Chongjian Ge, Ye Zhu, Ruimao Zhang, Zhen Li, Lingyan Zhang, Wanling Ma, Xiang Wan, et al. Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *arXiv preprint arXiv:2206.08023*, 2022.

[13] Davood Karimi and Ali Gholipour. Improving calibration and out-of-distribution detection in deep models for medical image segmentation. *IEEE transactions on artificial intelligence*, 4(2):383–397, 2022.

[14] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.

[15] Bennett Landman, Zhoubing Xu, Juan Igelsias, Martin Styner, Thomas Langerak, and Arno Klein. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *Proc. MICCAI multi-atlas labeling beyond cranial vault—workshop challenge*, volume 5, page 12. Munich, Germany, 2015.

[16] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31, 2018.

[17] Shiman Li, Mingzhi Yuan, Xiaokun Dai, and Chenxi Zhang. Evaluation of uncertainty estimation methods in medical image segmentation: Exploring the usage of uncertainty in clinical deployment. *Computerized Medical Imaging and Graphics*, page 102574, 2025.

[18] Jie Liu, Yixiao Zhang, Jie-Neng Chen, Junfei Xiao, Yongyi Lu, Bennett A Landman, Yixuan Yuan, Alan Yuille, Yucheng Tang, and Zongwei Zhou. Clip-driven universal model for organ segmentation and tumor detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 21152–21164, 2023.

[19] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020.

[20] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15:654, 2024.

[21] Jun Ma, Sumin Kim, Feifei Li, Mohammed Baharoon, Reza Asakereh, Hongwei Lyu, and Bo Wang. Segment anything in medical images and videos: Benchmark and deployment. *arXiv preprint arXiv:2408.03322*, 2024.

[22] Md Eshmam Rayed, SM Sajibul Islam, Sadia Islam Niha, Jamin Rahman Jim, Md Mohsin Kabir, and MF Mridha. Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked*, page 101504, 2024.

[23] Yuhui Song, Xiuquan Du, Yanping Zhang, and Chenchu Xu. Multi-shot prototype contrastive learning and semantic reasoning for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 578–588. Springer, 2023.

[24] Liyan Sun, Chenxin Li, Xinghao Ding, Yue Huang, Zhong Chen, Guisheng Wang, Yizhou Yu, and John Paisley. Few-shot medical image segmentation using a global correlation network with discriminative embedding. *Computers in biology and medicine*, 140:105067, 2022.

[25] Song Tang, Shaxu Yan, Xiaozhi Qi, Jianxin Gao, Mao Ye, Jianwei Zhang, and Xiatian Zhu. Few-shot medical image segmentation with high-fidelity prototypes. *Medical Image Analysis*, 100:103412, 2025.

[26] Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20730–20740, 2022.

[27] Anton Vasiliuk, Daria Frolova, Mikhail Belyaev, and Boris Shirokikh. Exploring structure-wise uncertainty for 3d medical image segmentation. In *International Conference on Medical Imaging and Computer-Aided Diagnosis*, pages 15–26. Springer, 2022.

[28] Anton Vasiliuk, Daria Frolova, Mikhail Belyaev, and Boris Shirokikh. Limitations of out-of-distribution detection in 3d medical image segmentation. *Journal of Imaging*, 9 (9):191, 2023.

[29] Haoyu Wang, Sizheng Guo, Jin Ye, Zhongying Deng, Junlong Cheng, Tianbin Li, Jianpin Chen, Yanzhou Su, Ziyan Huang, Yiqing Shen, et al. Sam-med3d: towards general-purpose segmentation models for volumetric medical images. In *European Conference on Computer Vision*, pages 51–67. Springer, 2024.

[30] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *proceedings of the IEEE/CVF international conference on computer vision*, pages 9197–9206, 2019.

[31] Xiaoxiao Wu, Zhenguo Gao, Xiaowei Chen, Yakai Wang, Shulei Qu, and Na Li. Support-query prototype fusion network for few-shot medical image segmentation. *arXiv preprint arXiv:2405.07516*, 2024.

[32] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3474–3482, 2018.

[33] Ziqin Zhou, Yinjie Lei, Bowen Zhang, Lingqiao Liu, and Yifan Liu. Zegclip: Towards adapting clip for zero-shot semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11175–11185, 2023.