Tied Prototype Model for Few-Shot Medical Image Segmentation

Hyeongji Kim[1](✉), Stine Hansen[2], and Michael Kampffmeyer[1,3]

[1] Department of Physics and Technology, UiT - The Arctic University of Norway, Tromsø, Norway
hyeongji.kim@uit.no
[2] Norwegian Centre for Clinical Artificial Intelligence (SPKI), University Hospital of North-Norway, Tromsø, Norway
[3] Norwegian Computing Center, Oslo, Norway

Abstract. Common prototype-based medical image few-shot segmentation (FSS) methods model foreground and background classes using class-specific prototypes. However, given the high variability of the background, a more promising direction is to focus solely on foreground modeling, treating the background as an anomaly—an approach introduced by ADNet. Yet, ADNet faces three key limitations: dependence on a single prototype per class, a focus on binary classification, and fixed thresholds that fail to adapt to patient and organ variability. To address these shortcomings, we propose the Tied Prototype Model (TPM), a principled reformulation of ADNet with tied prototype locations for foreground and background distributions. Building on its probabilistic foundation, TPM naturally extends to multiple prototypes and multi-class segmentation while effectively separating non-typical background features. Notably, both extensions lead to improved segmentation accuracy. Finally, we leverage naturally occurring class priors to define an ideal target for adaptive thresholds, boosting segmentation performance. Taken together, TPM provides a fresh perspective on prototype-based FSS for medical image segmentation. The code can be found at https://github.com/hjk92g/TPM-FSS.

Keywords: Few-Shot Segmentation · Medical Image Segmentation · Multi-Class Segmentation · Prototype Model

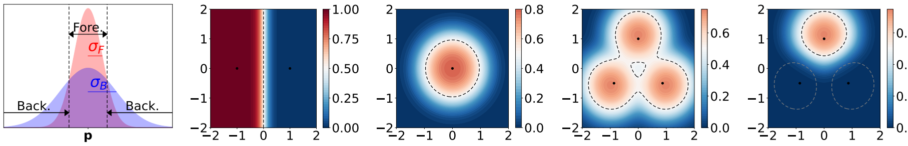
1 Introduction

Medical image segmentation is a critical component of numerous clinical applications such as diagnostics [19] and treatment planning [3]. While supervised deep learning approaches can achieve good performance, their applications are constrained by the limited availability of annotated medical images. To address this, few-shot segmentation (FSS) approaches have been proposed to effectively adapt models trained with labeled datasets to new, previously unseen, classes.

The dominant FSS approach employs prototype networks [16], first introduced by PANet [21], where each class is represented by a single prototype obtained through mean average pooling (MAP). Recognizing the limitation of a single prototype in modeling rich feature variants, PPNet [11] proposed using multiple prototypes for each class to improve expressiveness. More recently, there has been growing interest in leveraging self-supervised learning (SSL) to circumvent the need for labeled medical data altogether. ALPNet [13] pioneered the use of SSL by using pseudo-labels generated from superpixels, with additional local prototypes aiding intra-class local information. However, the background class typically exhibits significant heterogeneity, making it difficult to model using a fixed number of prototypes. Recognizing this, ADNet [6] employs an anomaly-detection-inspired approach, focusing on modeling the foreground prototype while treating the background as anomalous and separating regions via inside-outside (IO) classification. It assigns anomaly scores to each spatial location, using a fixed threshold at inference for segmentation. This simple method has demonstrated greater robustness to background heterogeneity relative to ALPNet and has been successfully used for the detection of brain tumors [1], ischemic stroke lesions [18], and lung lesions [17].

However, ADNet model [6] has three main drawbacks. First, its reliance on a single foreground prototype limits the expressiveness of the foreground class in medical images with significant intra-class variation. While non-anomaly-detection multi-prototype methods exist [13,5], they struggle to capture the full diversity of the background class. Meanwhile, existing anomaly-detection-based multi-prototype variants [14,25,23,26] of ADNet involve cumbersome modifications such as the introduction of new layers or hyperparameters. Hence, it is desirable for a more grounded yet simple approach that efficiently leverages multiple foreground prototypes at the feature level. Second, ADNet is limited to binary classification settings. While ADNet++ [7] extends it to multi-class classification, it only adapts the inference phase and thus does not effectively model the class relationships in training. Finally, ADNet's fixed threshold is insufficient to accommodate the inherent variability across patients and organs in diverse applications. While adaptive thresholds have been proposed to address this by utilizing support features [4], query information [15], or both [24], their threshold values are learned through cross-entropy-based losses, which may not necessarily optimize segmentation accuracy as further discussed in Sec. 3.4.

To address these issues, we take a step back and propose the Tied Prototype Model (TPM), a principled reformulation of ADNet [6] that leverages tied (shared) prototype locations for foreground and background distributions. As visualized in Fig. 1, the distributional foundation of TPM allows for its natural extension to both multiple prototypes and multi-class classification settings while retaining its simplicity and ability to separate non-typical background features. In particular, we propose a Gaussian mixture model (GMM)-based approach for multi-prototype segmentation, demonstrating its effectiveness in improving segmentation accuracy. Moreover, we introduce multi-foreground training for multi-class segmentation and show that it consistently offers superior representation learning capability. Lastly, by employing naturally occurring class priors



(a) 1D-distrib. (b) Std. model (c) TPM-SP (d) TPM-MP (e) TPM-MC

Fig. 1: (a) TPM class distributions. The (first) foreground class probability $p(F|x)$, based on (b) the standard prototype model, (c) TPM with a single prototype (SP), similar to the ADNet [6]—which is limited to spherical space, (d) TPM with multiple prototypes (MP), and (e) TPM for multi-class (MC) classification. For (b)–(e), the xy-axis shows the 2D Euclidean feature space positions, and the points denote prototypes. Dashed lines indicate decision boundaries for the shown (black) and the other (gray) foreground classes. Compared to standard prototype models, all TPM models achieve IO separation.

and feature distance distributions, we provide an ideal threshold estimation approach, aiming to enhance segmentation accuracy.

2 Preliminaries

2.1 Self-Supervised Few-Shot Learning Problem Setting

We adopt the self-supervised few-shot learning from [6], where pseudo-labels for unlabeled volumes are generated by performing supervoxel segmentation and randomly sampling a supervoxel as the foreground. Two 2D slices with the supervoxel are then selected to serve as support and query images, with one being augmented. The support image $\mathbf{X}^s$ and its pseudo-label $\mathbf{Y}^s$, where $\mathbf{Y}^s_c$ denotes the mask of class c, guide the segmentation of the query image $\mathbf{X}^q$, all of size (H, W). A feature extractor network f_θ encodes images into feature maps $\mathbf{F}^s = f_\theta(\mathbf{X}^s)$ and $\mathbf{F}^q = f_\theta(\mathbf{X}^q)$, with spatial dimensions (H', W') and feature dimension d. The upsampled predicted query label $\hat{\mathbf{Y}}^q$ is then compared to label $\mathbf{Y}^q$, training the network f_θ with a cross-entropy-based segmentation loss.

2.2 Anomaly Detection-Inspired Few-Shot Segmentation

In ADNet [6], the foreground prototype $\mathbf{p}$ is computed by upsampling the feature maps to the original size and applying masked average pooling (MAP) as:

$$\mathbf{p} = \frac{\sum_{\mathbf{r}} \mathbf{F}^s(\mathbf{r}) \odot \mathbf{Y}^s_F(\mathbf{r})}{\sum_{\mathbf{r}} \mathbf{Y}^s_F(\mathbf{r})}, \tag{1}$$

with $\odot$ as the Hadamard product, F the foreground class, and $\mathbf{r}$ the pixel location. It defines the anomaly score $S(\mathbf{r}) = -\alpha \cos(\mathbf{F}^q(\mathbf{r}), \mathbf{p})$ for each query feature

$\mathbf{F}^q(\mathbf{r})$, where $\alpha = 20$ is a scaling factor following [12], and $\cos(\cdot, \cdot)$ is the cosine similarity. Then, ADNet estimates foreground class probability as:

$$p(F|\mathbf{F}^q(\mathbf{r})) = 1 - \mathrm{sig}(S(\mathbf{r}) - T_S), \tag{2}$$

where $\mathrm{sig}(\cdot)$ is the sigmoid function with a steepness parameter $\kappa = 0.5$, and T_S is a learnable anomaly score threshold. This estimate is then upsampled and, given the binary task focus, is combined with its complement to generate the predicted label $\hat{\mathbf{Y}}^q$.

3 Method

3.1 Tied Prototype Model

The tied prototype model (TPM), with our established equivalence to ADNet [6] under a certain condition, adopts a probabilistic view of discriminative classification. This provides valuable insights for advancing few-shot medical image segmentation while addressing ADNet's limitations without modifying its architecture.

As illustrated in Fig. 1a, the key idea of our model is to use class distributions with a tied (shared) center position $\mathbf{p}$ for foreground and background classes while differing in their dispersion parameters. This contrasts with standard classification, which uses distinct centers for different classes. Though counterintuitive, enforcing the same center for both foreground and background classes is essential in separating the background class, as shown in Fig. 1c and Thm. 1.

TPM assumes that the foreground and background class distributions follow multivariate normal distributions, $\mathcal{N}(\mathbf{p}, \sigma_F^2 I)$ and $\mathcal{N}(\mathbf{p}, \sigma_B^2 I)$, respectively. Here, the tied prototype $\mathbf{p}$ serves as the location parameter, while σ_F and σ_B are class standard deviations, with $\sigma_F < \sigma_B$. The symbol I denotes the identity matrix. By applying Bayes' theorem, the foreground class probability is given by:

$$p(F|\mathbf{F}^q(\mathbf{r})) = \frac{p_F \phi(\mathbf{p}, \sigma_F; \mathbf{r})}{p_F \phi(\mathbf{p}, \sigma_F; \mathbf{r}) + p_B \phi(\mathbf{p}, \sigma_B; \mathbf{r})} = \frac{1}{1 + \frac{p_B}{p_F} \frac{\phi(\mathbf{p}, \sigma_B; \mathbf{r})}{\phi(\mathbf{p}, \sigma_F; \mathbf{r})}} \tag{3}$$

where p_F and p_B denote class priors and $\phi(\mathbf{p}, \sigma; \mathbf{r})$ represents the normal distribution density function with a mean $\mathbf{p}$ and covariance matrix $\sigma^2 I$. The choices of priors are discussed in Sec. 3.4.

Fig. 1c visualizes the estimated class probability of TPM with a single prototype, assigning higher probabilities to features closer to the prototype $\mathbf{p}$ while giving lower values to those further from the prototype. This inside-outside (IO) classification behavior highlights the model's effectiveness in distinguishing typical features from diverse non-typical features, which is particularly useful for foreground and background classification. As illustrated in Fig. 1b, this beneficial IO separation property is absent in the standard prototype model, which relies on distinct (non-tied) prototype positions and assigns class probabilities along the direction of the prototypes' difference.

Theorem 1. *In one foreground classification with a single prototype with a unit spherical embedding* $\mathbb{S}^{d-1}$*, the tied prototype model is equivalent to ADNet [6].*

Proof. To establish equivalence, we must show that Eqs. (2) and (3) are identical under a proper parameter correspondence. We first define $\mathrm{TP}(\mathbf{r})$ to simplify Eq. (3) and replace the multiplication with the exponential function, yielding:

$$\mathrm{TP}(\mathbf{r}) = \frac{1}{1 + \frac{p_B}{p_F} \frac{\phi(\mathbf{p},\sigma_B;\mathbf{r})}{\phi(\mathbf{p},\sigma_F;\mathbf{r})}} = \frac{1}{1 + \exp\left(\ln\left(\frac{\phi(\mathbf{p},\sigma_B;\mathbf{r})}{\phi(\mathbf{p},\sigma_F;\mathbf{r})}\right) + \ln\left(\frac{p_B}{p_F}\right)\right)}. \tag{4}$$

By substituting the density function $\phi(\cdot,\cdot)$ with its definition, we obtain:

$$\mathrm{TP}(\mathbf{r}) = \frac{1}{1 + \exp\left(\frac{1}{2}\left\|\mathbf{F}^q(\mathbf{r}) - \mathbf{p}\right\|^2 \left(\frac{1}{\sigma_F^2} - \frac{1}{\sigma_B^2}\right) + d\ln\left(\frac{\sigma_F}{\sigma_B}\right) + \ln\left(\frac{p_B}{p_F}\right)\right)}. \tag{5}$$

We then set $\alpha = 2\left(\frac{1}{\sigma_F^2} - \frac{1}{\sigma_B^2}\right)$ and $T_S = 2\ln\left(\frac{p_F}{p_B}\right) - 2d\ln\left(\frac{\sigma_F}{\sigma_B}\right) - \alpha$ to obtain:

$$\mathrm{TP}(\mathbf{r}) = \frac{1}{1 + \exp\left(\frac{1}{2}\left(\left(\frac{1}{2}\left\|\mathbf{F}^q(\mathbf{r}) - \mathbf{p}\right\|^2 - 1\right)\alpha - T_S\right)\right)}. \tag{6}$$

Since we use the unit spherical embedding $\mathbb{S}^{d-1}$, embedding vectors $\mathbf{F}^q(\mathbf{r})$ and $\mathbf{p}$ satisfy the equation $\frac{2-\|\mathbf{F}^q(\mathbf{r})-\mathbf{p}\|^2}{2} = \cos(\mathbf{F}^q(\mathbf{r}), \mathbf{p})$, leading to:

$$\mathrm{TP}(\mathbf{r}) = \frac{1}{1 + \exp\left(\frac{1}{2}\left(-\alpha\cos(\mathbf{F}^q(\mathbf{r}), \mathbf{p}) - T_S\right)\right)}. \tag{7}$$

Using the definitions of the anomaly score $S(\mathbf{r})$ and the sigmoid function $\mathrm{sig}(\cdot)$, we can observe that Eq. (7) is equivalent to Eq. (2). ☐

Thm. 1 reveals that our TPM, under specific conditions, is a reparameterized ADNet [6]. Unlike ADNet which inherently uses spherical embedding due to its reliance on cosine similarity, our model can utilize other geometries. Furthermore, it enables extensions beyond a single prototype as discussed below.

3.2 Binary Classification with Multiple Prototypes

Rooted in its distributional formulation, TPM naturally extends to multiple prototypes modeled by Gaussian mixture models (GMMs). As illustrated in Fig. 1d, this extension allows it to capture diversity in foreground features while retaining the ability to distinguish non-typical background features.

To learn the GMM parameters from the foreground features, we apply the EM algorithm from [22], but without enforcing equal mixture weights. This algorithm alternates between the E-step and M-step, to extract multiple prototypes $\mathbf{p}_\mathbf{m}$ along with their weights $w_\mathbf{m}$, satisfying $\sum w_\mathbf{m} = 1$. When multiple prototypes are obtained for the foreground class, we can simply use the same prototypes for the background class as well. Specifically, GMM $\sum_\mathbf{m} w_\mathbf{m}\phi(\mathbf{p}_\mathbf{m}, \sigma_F; \mathbf{r})$

is used for the foreground class distribution, while the corresponding GMM $\sum_{\mathbf{m}} w_{\mathbf{m}} \phi(\mathbf{p}_{\mathbf{m}}, \sigma_B; \mathbf{r})$ represents the background class distribution. Using Bayes' theorem, the foreground class probability $p_{\mathrm{MP}}(F|\mathbf{F}^q(\mathbf{r}))$, estimated from multiple prototypes and associated class priors, is given by:

$$p_{\mathrm{MP}}(F|\mathbf{F}^q(\mathbf{r})) = \frac{p_{F;\mathrm{MP}} \sum_{\mathbf{m}} w_{\mathbf{m}} \phi(\mathbf{p}_{\mathbf{m}}, \sigma_F; \mathbf{r})}{p_{F;\mathrm{MP}} \sum_{\mathbf{m}} w_{\mathbf{m}} \phi(\mathbf{p}_{\mathbf{m}}, \sigma_F; \mathbf{r}) + p_{B;\mathrm{MP}} \sum_{\mathbf{m}} w_{\mathbf{m}} \phi(\mathbf{p}_{\mathbf{m}}, \sigma_B; \mathbf{r})}. \tag{8}$$

3.3 Multi-Class Classification

The distributional foundation of TPM allows for another expansion: multi-class classification. While multiple prototypes can be used for each foreground class, this work focuses on the simplified case of using a single prototype $\mathbf{p}_{F_i}$, obtained by MAP in Eq. (1), for foreground class F_i. Fig. 1e visualizes the class probability $p(F_1|x)$ of the first foreground class with prototype $\mathbf{p}_{F_1} = (0, 1)$. The model effectively distinguishes features from other foreground classes while ensuring the separation of non-typical background features.

With a similar assumption on class distributions as in Secs. 3.1 and 3.2, the class probability $p_{\mathrm{MC}}(F_i|\mathbf{F}^q(\mathbf{r}))$ of foreground class F_i can be obtained as:

$$p_{\mathrm{MC}}(F_i|\mathbf{F}^q(\mathbf{r})) = \frac{p_{F_i} \phi(\mathbf{p}_{F_i}, \sigma_F; \mathbf{r})}{\sum_{i'} p_{F_{i'}} \phi(\mathbf{p}_{F_{i'}}, \sigma_F; \mathbf{r}) + \sum_{i'} p_B \phi(\mathbf{p}_{F_{i'}}, \sigma_B; \mathbf{r})}, \tag{9}$$

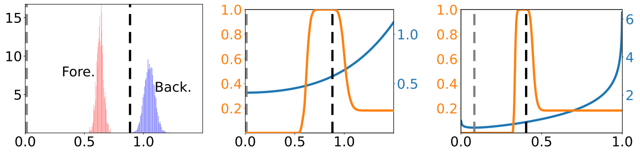
where p_{F_i} is the class prior of F_i. The background class probability can then be computed as $p_{\mathrm{MC}}(B|\mathbf{F}^q(\mathbf{r})) = 1 - \sum_{i'} p_{\mathrm{MC}}(F_i'|\mathbf{F}^q(\mathbf{r}))$.

Ignoring the background class, i.e., when $p_B = 0$, Eq. (9) corresponds to the normalized softmax [20]. This suggests that the model enables training to separate different foreground features into their respective prototype positions while pushing background features away from the prototypes. This is not possible in multi-class classification with ADNet++ [7], as its use of the max operation over foreground classes inherently makes it a piecewise binary classification model.

3.4 Targeting the Ideal Threshold

The foreground-background imbalance is a critical challenge in few-shot medical segmentation. While adaptive thresholds [15,4,24] attempt to tackle this with cross-entropy (CE), as illustrated in Fig. 2, minimizing CE does not necessarily yield the highest Dice score, showing that CE minimization does not directly optimize segmentation accuracy. Furthermore, using the *ideal* proportion—where the predicted foreground pixel counts match the true counts—achieves a near-optimal Dice score. Inspired by this and the significant performance gains observed in [2], we propose ideal thresholds of training query images as targets for learning adaptive thresholds. For brevity, we consider the single-prototype binary classification setting, which readily extends to the other settings.

First, we compute the upsampled Euclidean distance between features and prototypes. We then calculate the ideal distance threshold (IDT) T_D^* for each



(a) Dist. distrib. (b) Varying T_D (c) Varying p_F

Fig. 2: (a) Distributions of distances to a prototype for Gaussian samples projected onto a sphere. Dice score (orange) vs. CE (blue) for (b) varying distance threshold T_D and (c) varying foreground class prior p_F. Dashed lines indicate values for the ideal threshold (black) and the CE-minimizing threshold (gray).

training query image. Mathematically, $T_D^* = \frac{D;\text{sort}_{|F|} + D;\text{sort}_{|F|+1}}{2}$, where D;sort is a vector of distances sorted in ascending order and $|F|$ is the foreground pixel count. By leveraging sorted distances, IDT ensures the number of predicted foreground pixels matches the true count $|F|$. To employ Eq. (3), we convert IDT into the corresponding ideal class prior (ICP) as $p_F^* = 1 - \text{sig}(-T_D^{*2}\left(\frac{1}{\sigma_F^2} - \frac{1}{\sigma_B^2}\right) - 2d\ln(\frac{\sigma_F}{\sigma_B}))$. Replacing p_F and the background prior with ICP p_F^*, in Eq. (3), shifts its decision boundary distance to IDT T_D^*.

4 Experiments

4.1 Setup

Dataset. Two public abdominal datasets are used for evaluation: ABD-MRI and ABD-CT. ABD-MRI [9] consists of 20 abdominal MRI scans with annotations for the liver, left kidney (L.K.), right kidney (R.K.), and spleen (Spl.). ABD-CT [10] contains 30 abdominal CT scans with the same organ annotations as ABD-MRI. For supervoxel generation, based on [6,7], we set the size parameter to 5000 for ABD-MRI and 2000 for ABD-CT.

Implementation Details. Following [6] for the binary-class and [7] for the multi-class setting, we conduct experiments using 5-fold cross-validation, performing three training runs per split, and leverage a single support image slice to segment the entire query volume and measure the mean Dice score. We initialize the feature extractor f_θ with a pre-trained ResNet-101 [8], using randomly transformed and reshaped 256×256 images, and optimize for 50k iterations using SGD with a learning rate of 10^{-3}. Features are 256-dimensional and normalized onto a spherical embedding, with $d = 1$ for simplicity. We run experiments in PyTorch (v1.9.0) on an NVIDIA RTX 3090 GPU.

Training and Threshold Options. To showcase the effectiveness of our approaches, we compare results across various options. For binary segmentation, we assess single-prototype-trained models using both single- and five-prototype evaluations. The baseline is standard ADNet [6], trained with the additional T loss, which regularizes T_S. In contrast, our TPM training excludes the T loss

Table 1: Binary segmentation. Best result in **bold**, second-best underlined. * denotes our contributions and OCP is the oracle result using test labels.

Proto.	Inference	ABD-MRI [9]					ABD-CT [10]				
		Liver	L.K.	R.K.	Spl.	Mean	Liver	L.K.	R.K.	Spl.	Mean
Single	ADNet[6]	74.86	59.41	84.27	59.54	69.52	56.18	34.82	37.56	49.14	44.42
	CE-T*	**75.31**	56.91	78.87	56.93	67.00	71.76	29.27	28.99	46.90	44.23
	AvgEst*	69.94	70.38	**85.99**	69.40	73.93	33.36	39.42	**41.49**	56.47	42.68
	LinEst*	73.98	**71.23**	85.64	69.11	74.99	65.34	38.92	40.88	58.49	50.91
	OCP	85.71	87.19	89.47	81.05	85.85	81.01	65.65	67.34	76.15	72.54
Multi (5)	CE-T*	74.85	54.92	77.60	56.75	66.03	**75.50**	28.48	28.79	46.06	44.71
	AvgEst*	73.60	68.08	85.82	69.65	74.29	46.26	39.97	40.49	58.71	46.36
	LinEst*	74.83	70.18	85.61	**70.18**	**75.20**	70.40	**40.08**	41.40	**58.96**	**52.71**
	OCP	85.39	86.00	89.00	81.04	85.36	82.96	64.14	66.00	76.14	72.31

and only optimizes the segmentation loss, as our proposed estimation schemes provide more principled thresholds. To assess the impact of aligning predicted and true pixel counts, we report TPM results using the CE-trained threshold (CE-T), estimated ideal class prior (ICP) values from training data—specifically, the average estimation (AvgEst) and a linear estimation (LinEst), with the latter based on the support foreground size $|F|$ and query slice location—as well as the oracle ideal class prior (OCP) for test queries. For multi-class segmentation, we compare results from single- and five-foreground training, using ADNet++ [7] as the baseline. While it is possible to consider each class count separately, we report LinEst estimated by matching the sum of all foreground class counts.

4.2 Results

Table 1 presents the binary segmentation results, confirming the merits of the ICP and multi-prototype (MP) approaches. In the single-prototype (SP) setting, while ADNet [6] tends to outperform CE-T (ADNet equivalent without T loss), AvgEst generally yields higher Dice scores than ADNet, except for the liver class, which has a relatively large foreground size. Moreover, LinEst further improves performance by incorporating size information, revealing ICP's benefit in this setting. The notably higher Dice scores of the OCP, which leverages oracle information, suggest significant room for improvement with advanced ICP estimation. Compared to SP predictions, the proposed MP extension further enhances the performance of the estimated ICPs. Table 2 presents the results for TPM's multi-class segmentation extension, demonstrating the efficacy of ICP and multi-foreground (MF) training. Focusing on single foreground (SF) training, TPM with estimated ICPs significantly outperforms the ADNet++ baseline [7]. Furthermore, MF consistently enhances TPM's performance over SF. This highlights the superior representation learning capability of the proposed extension of our TPM.

Table 2: Multi-class segmentation. * denotes our contributions.

Train	Inference	ABD-MRI [9]					ABD-CT [10]				
		Liver	L.K.	R.K.	Spl.	Mean	Liver	L.K.	R.K.	Spl.	Mean
SF	ADNet++[7]	64.13	54.62	59.86	41.36	54.99	65.58	20.67	16.66	19.59	30.63
	LinEst*	65.50	69.43	70.72	47.15	63.20	61.97	31.77	24.46	**29.15**	36.84
MF (5)	ADNet++[7]	57.22	45.85	54.57	29.93	46.89	61.14	15.43	12.00	12.88	25.36
	LinEst*	**71.91**	**71.00**	**73.85**	**49.94**	**66.67**	**66.84**	**37.90**	**27.55**	28.95	**40.31**

5 Conclusion

In this work, we introduce the tied prototype model (TPM), a probabilistic reformulation of ADNet [6] that advances its capabilities. Notably, TPM enables seamless extensions to multi-prototype and multi-class segmentation. Furthermore, we highlight the significance of ideal thresholds as target thresholds. Our experimental results demonstrate the performance improvements of each component, paving new research directions in prototype-based few-shot segmentation.

Acknowledgments. This work was supported by The Research Council of Norway (Visual Intelligence, grant no. 309439 as well as FRIPRO grant no. 315029 and IKT-PLUSS grant no. 303514).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Balasundaram, A., Kavitha, M.S., Pratheepan, Y., Akshat, D., Kaushik, M.V.: A foreground prototype-based one-shot segmentation of brain tumors. Diagnostics **13**(7), 1282 (2023)
2. Boudiaf, M., Kervadec, H., Masud, Z.I., Piantanida, P., Ben Ayed, I., Dolz, J.: Few-shot segmentation without meta-learning: A good transductive inference is all you need? In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13979–13988 (2021)
3. Chen, X., Sun, S., Bai, N., Han, K., Liu, Q., Yao, S., Tang, H., Zhang, C., Lu, Z., Huang, Q., et al.: A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. Radiotherapy and Oncology **160**, 175–184 (2021)
4. Cheng, Z., Wang, S., Xin, T., Zhang, H.: Frequency-aware adaptive filtering network for few-shot medical image segmentation. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1890–1897. IEEE (2024)
5. Cheng, Z., Wang, S., Xin, T., Zhou, T., Zhang, H., Shao, L.: Few-shot medical image segmentation via generating multiple representative descriptors. IEEE Transactions on Medical Imaging (2024)
6. 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)

7. Hansen, S., Gautam, S., Salahuddin, S.A., Kampffmeyer, M., Jenssen, R.: Adnet++: A few-shot learning framework for multi-class medical image volume segmentation with uncertainty-guided feature refinement. Medical Image Analysis **89**, 102870 (2023)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021)
10. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge. vol. 5, p. 12 (2015)
11. Liu, Y., Zhang, X., Zhang, S., He, X.: Part-aware prototype network for few-shot semantic segmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. pp. 142–158. Springer (2020)
12. Oreshkin, B., Rodríguez López, P., Lacoste, A.: Tadam: Task dependent adaptive metric for improved few-shot learning. Advances in neural information processing systems **31** (2018)
13. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. pp. 762–780. Springer (2020)
14. Salahuddin, S.A., Hansen, S., Gautam, S., Kampffmeyer, M., Jenssen, R.: A self-guided anomaly detection-inspired few-shot segmentation network. CEUR Workshop Proceedings (2022)
15. Shen, Q., Li, Y., Jin, J., Liu, B.: Q-net: Query-informed few-shot medical image segmentation. In: Proceedings of SAI Intelligent Systems Conference. pp. 610–628. Springer (2023)
16. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in Neural Information Processing Systems **30**, 4077–4087 (2017)
17. Tian, Y., Liang, Y., Chen, Y., Zhang, J., Bian, H.: Multilevel support-assisted prototype optimization network for few-shot medical segmentation of lung lesions. Scientific Reports **15**(1), 3290 (2025)
18. Tomasetti, L., Hansen, S., Khanmohammadi, M., Engan, K., Høllesli, L.J., Kurz, K.D., Kampffmeyer, M.: Self-supervised few-shot learning for ischemic stroke lesion segmentation. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)
19. Tsochatzidis, L., Koutla, P., Costaridou, L., Pratikakis, I.: Integrating segmentation information into cnn for breast cancer diagnosis of mammographic masses. Computer Methods and Programs in Biomedicine **200**, 105913 (2021)
20. Wang, F., Xiang, X., Cheng, J., Yuille, A.L.: Normface: L2 hypersphere embedding for face verification. In: Proceedings of the 25th ACM international conference on Multimedia. pp. 1041–1049 (2017)
21. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: Few-shot image semantic segmentation with prototype alignment. In: proceedings of the IEEE/CVF international conference on computer vision. pp. 9197–9206 (2019)

22. Yang, B., Liu, C., Li, B., Jiao, J., Ye, Q.: Prototype mixture models for few-shot semantic segmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16. pp. 763–778. Springer (2020)
23. Zhao, Z., Gao, J., Luo, Z., Li, S.: Cpnet: Cross prototype network for few-shot medical image segmentation. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 135–149. Springer (2024)
24. Zhu, Y., Cheng, Z., Wang, S., Zhang, H.: Learning de-biased prototypes for few-shot medical image segmentation. Pattern Recognition Letters **183**, 71–77 (2024)
25. 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)
26. Zhu, Y., Wang, S., Xin, T., Zhang, Z., Zhang, H.: Partition-a-medical-image: Extracting multiple representative sub-regions for few-shot medical image segmentation. IEEE Transactions on Instrumentation and Measurement (2024)