Automated Pseudo-Label Generation and Parallel Computing for Enhanced Few-Shot Medical Image Segmentation

Trong-Duc Nguyen* and Tien-Dung Do † and Thanh-Ha Do‡
* VNU University of Science
E-mail: nguyentrongduc_t66@hus.edu.vn
† Post and Telecommunication Institute of Technology (PTIT), Vietnam
E-mail: dungdt@ptit.edu.vn
‡ VNU University of Science
E-mail: dothanhha@hus.edu.vn

***Abstract*—Few-shot semantic segmentation is a technique with significant potential for medical image segmentation tasks. Most existing few-shot semantic segmentation methods require fully annotated labels for the training process. However, these methods may not be suitable for medical images, where data collection and labeling are challenging. To address this issue, this paper proposed an enhanced, few-shot semantic segmentation model with a new pre-processing step to generate pseudo-labels automatically. In this paper, parallel computing is also developed to accelerate image pre-processing. Experiments done on MRI image datasets present the effectiveness of the new approach since it outperforms conventional few-shot semantic segmentation methods.**

I. Introduction

Medical image segmentation classifies and locates areas in medical images, aiding in the diagnosis and treatment of diseases. Its applications are diverse, from detecting cancer and classifying organs over images to analyzing images of the brain and heart. Medical image segmentation presents a significant challenge for medical experts and image processing engineers because of the complexity, noise, and large size of medical images. Therefore, many types of research have been done to segment medical images. They are divided into two directions: traditional image segmentation methods and deep-learning-based segmentation methods.

Traditional image segmentation methods use images' characteristics, such as color and edge, to analyze the color values of pixels in the image over different color spaces, such as RGB, HUE, and HSV. Then, similar color values are grouped into regions with similar colors. Color-based image segmentation can be used for simple applications, but outliers, such as lighting, shadows, and transformations, can reduce accuracy.

Edge-based approaches detect edges and contours to segment the image into regions. These methods use edge detection algorithms such as Sobel , Canny , or Laplacian , followed by an algorithm such as RANSAC to detect points on the edge. After detecting points on the edge, clustering methods such as K-means and DBSCAN are applied to group the points on the edge into objects. In general, edge-based segmentation methods are often affected by noise and lighting variations, and segmenting objects of different shapes and sizes is challenging.

Threshold-based image segmentation methods set thresholds to divide pixels into regions of similar contrast. However, these methods also have limitations, such as setting thresholds, which can be challenging and require much experimentation to find the suitable threshold. Additionally, they could be more effective with low-contrast images or are affected by noise or uneven lighting.

Traditional image segmentation methods generally allow quick and easy implementation. Their fast processing speed is also advantageous in applications that require fast processing speed. However, they have some notable disadvantages, such as producing regions segmented incorrectly or confused with others in complex images or objects with complex shapes, being unstable to outliers such as noise or uneven lighting, and requiring adjusting many parameters to achieve the best segmentation results, which requires extensive experience and expertise.

Deep learning methods have recently developed efficient and accurate medical image segmentation solutions. Notable among these are the U-net [1], a convolutional neural network designed for biomedical image segmentation; the V-net, an extension of the U-net that incorporates volumetric data; and DeepLab, a state-of-the-art deep learning model for semantic image segmentation. In general, the performance of deep learning approaches is excellent. However, these networks all require labels for training. This is a big drawback because labeling medical images takes a lot of time and effort, and it requires well-qualified doctors to be able to label them correctly.

Recently, deep-learning-based approaches have combined the traditional model and deep learning architecture to form a self-learning method that takes advantage of the deep learning model and does not need to label the data (see Section II). Among approaches, the ALPNet represents a significant advancement in image segmentation by combining active learn-

ing with probabilistic deep learning models. It reduces the labeled data required for training models by iteratively selecting the most informative samples for labeling. By integrating with self-supervised learning (SSL) techniques, SSL-ALPNet proposed further efficiency and effectiveness than ordinary ALPNet since pre-training the network using self-supervised tasks on a large amount of unlabeled data can provide a strong initialization, reducing the number of learning cycles needed to achieve high performance.

Although SSL-ALPNet presents a promising approach, it still has room for improving performance, especially the computational complexity (iterative training, uncertainty estimation). Moreover, the risk of model over-fitting to the noise in the pseudo-labels during SSL phase can also be enhanced. Thus, this paper presents a new approach that improves the above limitations of SSL-ALPNet.

The paper's organization is as follows: Section II presents the state-of-the-art deep learning-based approach for medical segmentation. Section III presents the ALPNet and its enhanced version by investigating more on the clustering stage and time computing using parallel ideas. Section IV evaluates the proposed idea on the publication dataset, and the conclusion and future work are included in Section V.

II. Automated Pseudo-Label Generation for Medical Image Segmentation

There are medical image segmentation deep-learning-based methods for medical image segmentation, such as U-net, V-net, and DeepLab. While these networks require labels for training, which can be time-consuming for medical images, there is hope for the future. The self-learning method, pixel grouping, and deep-learning-based module are potential directions for improving the efficiency of image segmentation.

The authors in [2] presents an uncertainty-aware semi-supervised framework for 3D MR image segmentation with an uncertainty-aware scheme through the student and teacher models. The teacher model guide and target the student one to minimize segmentation loss and consistency loss. This uncertainty-aware scheme enables the student model to gradually learn from meaningful and reliable targets by exploiting the uncertainty information.

In [3], the authors proposed a multi-task deep network using a shareware semi-supervised segmentation strategy to leverage abundant unlabeled data and enforce a geometric shape constraint on the segmentation output. This network predicts semantic segmentation and signed distance maps of object surfaces. This paper also developed an adversarial loss to enhance the network's performance further. This loss function is designed to create a relationship between the predicted signed distance maps of labeled and unlabeled data. Doing so encourages the network to capture shape-aware features more effectively, thereby improving the quality of the segmentation output.

The authors in [4] present the MC-Net+ network, a potential approach in semi-supervised medical image segmentation. The MC-Net+ model introduces two designs. The first design features a shared encoder and multiple slightly different decoders. In contrast, the second design implements a mutual consistency constraint between one decoder's probability output and other decoders' soft pseudo labels. These two designs enable the MC-Net+ model to minimize the discrepancy of multiple outputs and generate consistent results in challenging regions, thereby promising to revolutionize model training.

In [5], Huimin Wu *et al.* propose a novel Compete-to-Win method (ComWin) to enhance the pseudo-label quality and reduce the critical limitation of the existing state-of-the-art semi-supervised approach cross with pseudo supervision. Their essential idea is to generate high-quality pseudo labels by comparing multiple confidence maps produced by different networks to select the most confident one. In addition, a boundary-aware enhancement module is integrated to refine pseudo labels further in near-boundary areas.

The SemiSL model [6] is a semi-supervised patch-based contrastive learning framework for medical image segmentation. In SemiSL, the pseudo-labels generated provide additional guidance, whereas discriminative class information learned in contrastive learning accurate multi-class segmentation. The novel loss in the SemiSL model also synergistically encourages inter-class separability and intra-class compactness among the learned representations.

Compared to the successful methods mentioned above, the model SSL-ALPNet [7] also significantly advanced medical image segmentation. In this paper, we research and then develop SSL-ALPNet [7] to improve performance, both in computational complexity and in reducing the over-fitting of the noise in the pseudo-labels during the SSL phase. On the other hand, our proposed method can be enhanced by incorporating unsupervised models in the SSL phase and adapted with other deep-learning models instead of ALPNet to boost semi-supervised medical image segmentation further.

III. Proposed Approach for Automated Pseudo-Label Generation

Our model is developed from the SSL-ALPNet model [7], which consists of two phases: offline pseudo-label generation and online training. The first phase directly affects the model's feature learning capability (see Figure 1). Therefore, we focus on this phase by proposing a novel pseudo-label generation algorithm instead of using Felsenszwalb's efficient graph-based [8] image segmentation algorithm as in the original SSL-ALPNet model. We also focus on optimizing prediction time to address more complex real-world challenges.

We selected the SLIC [9] algorithm to generate pseudo-segmentation labels to ensure segmentation accuracy and computational efficiency. Compared to the method used in [7], which employs a graph-based approach for segmentation, SLIC offers several advantages. Firstly, It is computationally more efficient due to its linear iterative clustering, with a computational complexity of $O(N)$, making it suitable for real-time applications. Secondly, SLIC produces more uniform superpixels that adhere closely to image boundaries, enhancing

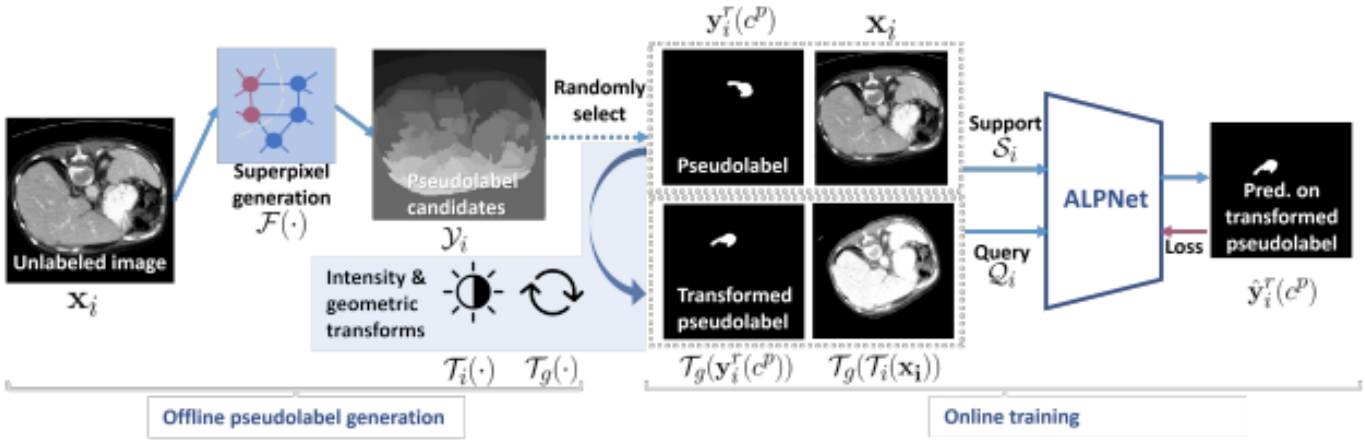


Fig. 1: Workflow of the superpixel-based self-supervised learning (SSL-ALPNet) [7]

the quality of the segmentation. In contrast, the method used in [7], with a computational complexity of $O(N \log N)$, while effective in capturing irregular and varied segment shapes, can be more computationally intensive and may produce segments that do not align as precisely with image edges. These characteristics make SLIC a preferred choice for applications requiring both speed and accuracy in superpixel generation.

A. The SLIC Model

The Simple Linear Iterative Clustering (SLIC) algorithm is an efficient method for generating superpixels, which are compact and nearly uniform regions that adhere to image boundaries. Due to its simplicity and speed, It is particularly effective for image segmentation tasks, making it a popular choice in image processing applications.

SLIC operates by clustering pixels in the combined five-dimensional color and image plane space, each cluster representing a superpixel. The algorithm can be summarized in the following steps:

1) **Initialization:** Initialize cluster centers $C_k = [l_k, a_k, b_k, x_k, y_k]$ by sampling pixels at regular grid steps S. Perturb the centers in a 3×3 neighborhood to the lowest gradient position to avoid placing centers at edges.
2) **Assignment:** For each pixel p, assign it to the nearest cluster center whose search region overlaps the pixel. The distance measure D combines color similarity and spatial proximity:

$$D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \tag{1}$$

where d_c is the color distance in CIELAB space, d_s is the spatial distance and m is a compactness factor that balances the two distances.
3) **Update:** Update the cluster centers by computing the mean l, a, b, x, y values of all pixels assigned to each cluster.
4) **Iteration:** Repeat the assignment and update steps until convergence, typically defined by a minimal change in cluster centers.

B. Improving the Time Computing

Although SLIC has lower computational complexity than Felsenszwalb's graph-based image segmentation algorithm used in [7], we aimed to accelerate the pre-processing step further to reduce the time training model. By utilizing parallel computing on multiple CPU threads, we allow for concurrent image processing and significantly reduce the processing time by distributing the workload across several threads.

1) Parallel Computing Approach: Our approach to parallel computing is a testament to efficiency. By dividing the task into smaller sub-tasks and executing them simultaneously on multiple processors, we've significantly enhanced the efficiency of the pre-processing step. This allows us to process various images concurrently, assigning each to a different CPU thread for simultaneous processing.

- **Thread Allocation:** We allocate a specific number of CPU threads based on the available cores. Each thread processes a distinct image, ensuring balanced workload distribution and maximizing CPU utilization.
- **Synchronization:** While each thread processes its assigned image independently, synchronization mechanisms ensure that all threads start and complete their tasks coordinately. This avoids conflicts and ensures efficient resource management.
- **Load Balancing:** Dynamic load balancing techniques ensure that all threads complete their tasks approximately simultaneously. This avoids scenarios where some threads are idle while others are still processing.
- **Memory Management:** Efficient memory management is crucial in parallel processing to prevent bottlenecks. Shared resources are carefully managed to ensure that threads can access necessary data without significant contention.

2) Implementation Details: To implement parallel processing of multiple images, we used threading modules in Python, which provide high-level abstractions for parallel programming. The steps involved in our parallel implementation are as follows:

- **Image Allocation:** The set of images to be processed is divided among the available threads. Each thread is assigned a subset of the images.
- **Parallel Processing:** Each thread independently processes its assigned images using the SLIC algorithm. This includes all steps of the SLIC process, from initialization to convergence.
- **Global Synchronization:** After processing all assigned images, threads synchronize to ensure that the entire batch of images has been processed before proceeding to the next batch.
- **Result Compilation:** The results from all threads are compiled into a unified output, maintaining the segmentation results for each image.

3) Performance Evaluation: We conducted extensive experiments to evaluate the performance of our parallel implementation. The results (see Section IV) show a significant reduction in processing time compared to the sequential version, especially when processing large batches of images. The algorithm's scalability was also tested, demonstrating that the parallel implementation scales well with the number of CPU cores. This parallelization approach accelerates the preprocessing step and enhances the overall efficiency of the image processing pipeline, making it highly suitable for real-time and high-resolution image processing tasks.

In conclusion, by utilizing parallel computing for the preprocessing step, we achieved significant improvements in processing time without altering the core SLIC [9] algorithm. This approach ensures that the algorithm can meet the demands of modern image processing applications, providing both speed and accuracy in superpixel generation.

IV. Experimental And Result

A. Dataset

We utilized the Abd-MRI dataset sourced from the ISBI 2019 Combined Healthy Abdominal Organ Segmentation Challenge (Task 5) [10], comprising 20 3D T2-SPIR MRI scans. Each scan in the dataset was converted into 2D axial slices and resized to dimensions of 256 $\times$ 256 pixels. Standard pre-processing procedures were applied, including normalization and augmentation. Each 2D slice was replicated three times along the channel dimension to accommodate the network architecture.

B. Experimental Settings

The parameters are established in a similar experiment to [7]. We use the Dice score (0-100, with 0 indicating no overlap and 100 indicating perfect overlap), a standard medical image segmentation research metric, to assess the overlap between predicted and ground truth segmentation. For evaluating 2D segmentation on 3D volumetric images, we adhere to the protocol established by [11]. Within a 3D image, for each class $\hat{c}_j$, images between the top and bottom slices containing $\hat{c}_j$ are divided into C equally spaced chunks. The middle slice in each chunk from the support scan is the reference for segmenting all slices in the corresponding chunk in the query scan. In our experiments, C is set to 3. Notably, the support and query scans originate from different patients.

To evaluate generalization to unseen test classes, we ensure test classes are entirely unseen by removing any image containing a test class from the training dataset. We evaluate the model's ability to generalize to unseen semantic classes, testing its robustness in scenarios where new classes arise that were not present during training.

1) **Dataset Preparation:** Exclude any images containing the testing classes from the training dataset to ensure the model is trained without exposure to these specific classes.
2) **Training:** Train the model using only the remaining annotated data. The model learns to segment based on standard features and structures in the training data without explicit knowledge of the testing classes.
3) **Testing:** Evaluate the model on images containing the unseen testing classes. These classes are entirely novel to the model, having been excluded from the training phase. Measure how effectively the model can generalize its learned features and segmentation strategies to new, previously unseen classes.

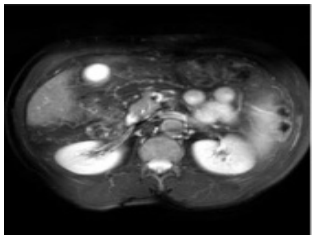
The experiments in this section are conducted under a 1-way 1-shot setting to simulate the limited availability of labeled data in clinical practice.

The network is implemented in PyTorch based on the official PANet implementation [12]. To achieve high spatial resolutions in feature maps, $f_\theta(\cdot)$ is configured as an off-the-shelf fully convolutional ResNet101, pre-trained on a subset of MS-COCO to enhance segmentation performance [12], [13] (similar to the vanilla PANet in our experiments). It processes a $3 \times 256 \times 256$ image and generates a $256 \times 32 \times 32$ feature map. The local pooling window (L_H, L_W) for prototypes is set to 4×4 during training and 2×2 during inference on feature maps.

C. Comparative Experiments

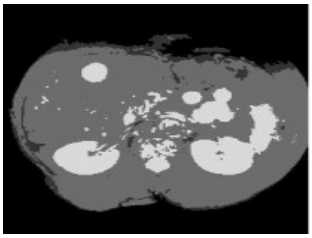
Table II shows that the proposed SLIC [9] method is generally more stable and accurate than existing methods.

Tables II compare our method with vanilla PANet, one of the most advanced methods for natural images, as well as SE-Net [11], the latest approach for few-shot segmentation in medical images. Our proposed method, utilizing automatic label generation combined with ALPNet [7], achieves superior performance in these comparisons without manual annotation. Importantly, Table II highlights the strong generalization capability of our approach to unseen classes. Our proposed self-supervised learning method based on automatic label generation has successfully trained the network to learn diverse and generalizable image representations from unlabeled data.

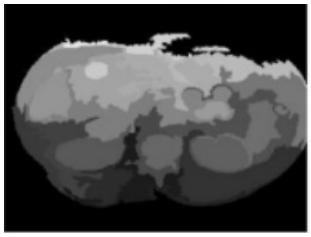


Original image

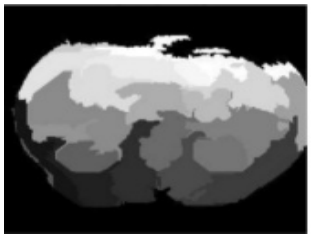
Truth ground



MeanShift



SLIC



Felzenszwalb

TABLE I: Experiment results (in Dice score) on abdominal images.

Method	Manual Annotation	Left kidney	Right kidney	Spleen	Liver	Mean
SE-Net [11]	✓	62.11	61.32	51.80	27.43	50.66
Vanilla PANet [12]	✓	53.45	38.64	50.90	42.26	46.33
MeanShift-ALPNet	×	50.03	46.25	56.44	42.78	48.88
SSL-ALPNet [7]	×	73.63	78.39	67.02	73.05	73.02
Proposed method	×	**82.53**	**82.04**	**80.61**	**80.35**	**81.38**

The ALPNet model has demonstrated strong generalization capabilities for object features compared to existing methods. As seen in Figure II, the newly proposed pseudo-label generation method improves over the technique used in [11], [12]. SLIC generates segments closer to the ground truth labels compared to [8], [14].

In addition, SLIC can speed up the pre-processing process compared to the method used in [7]; the results can be tracked in Table II. The execution time is significantly reduced, mainly when parallel computing is applied.

TABLE II: Time computing in pre-processing process done on Intel Core i7-10750H CPU, 6 Core(s)

Method used in offline pseudo-label generation	Frames/s
Felzenszwalb [8]	13
SLIC [9]	28
SLIC-Parallel	91

To further elucidate the impact of the size and shape of pseudo-labels on the model, we experimented with another pseudo-label generation method, MeanShift [14]. MeanShift can generate pseudo-labels with more accurate contours for smaller objects than SLIC. However, it tends to produce more significant segments for larger organs than usual. As shown in Table I, the model's accuracy significantly decreased.

The results show that the size of the pseudo label significantly impacts the model's accuracy, not only if the labels are too coarse. Alternatively, pseudo labels that are too fine-grained might divert the granularity of clusters in the learned representation space from that of actual semantic classes.

V. Conclusion

The paper introduces a method for pseudo labels automatically combined with existing deep learning models, forming a self-supervised few-shot segmentation framework. The proposed method has outperformed the modern method; the state-of-the-art method requires no manual training. It also represents a significant improvement over existing supervised self-learning models. The ability to generalize to unseen classes outperforms its predecessor. Furthermore, we optimized the pre-processing process through parallel computing, enabling the model to solve more complex problems. Importantly, this optimization also allows the model to operate in real time, demonstrating the practical applicability of our method and its potential for the future of supervised self-learning models.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[2] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, *Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation*, 2019. arXiv: 1907.07034 [cs.CV].

[3] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3d semantic segmentation for medical images," in *Lecture Notes in Computer Science*. Springer International Publishing, 2020, pp. 552–561, ISBN: 9783030597108.

[4] Y. Wu, Z. Ge, D. Zhang, *et al.*, *Mutual consistency learning for semi-supervised medical image segmentation*, 2022. eprint: 2109.09960.

[5] H. Wu, X. Li, Y. Lin, and K.-T. Cheng, "Compete to win: Enhancing pseudo labels for barely-supervised medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 42, no. 11, pp. 3244–3255, Nov. 2023, ISSN: 1558-254X.

[6] H. Basak and Z. Yin, "Pseudo-label guided contrastive learning for semi-supervised medical image segmentation," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 19 786–19 797.

[7] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, and D. Rueckert, "Self-supervision with superpixels: Training few-shot medical image segmentation without annotation," *arXiv preprint arXiv:2007.09886*, 2020.

[8] P. F. Felsenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," in *International Journal of Computer Vision*, Springer, vol. 59, 2004, pp. 167–181.

[9] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels," *Computer Vision–ECCV 2010*, pp. 849–862, 2010.

[10] A. Kavur, N. Gezer, M. Barı¸s, *et al.*, "Chaos challenge–combined (ct-mr) healthy abdominal organ segmentation," *arXiv preprint arXiv:2001.06535*, 2020.

[11] A. Roy, S. Siddiqui, S. Pölsterl, N. Navab, and C. Wachinger, "'squeeze excite' guided few-shot segmentation of volumetric images," *Medical Image Analysis*, vol. 59, p. 101 587, 2020.

[12] K. Wang, J. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," pp. 9197–9206, 2019.

[13] H. Shin, H. Roth, M. Gao, *et al.*, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[14] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, IEEE, 2002, pp. 603–619.