Mathematical Morphology and Self-Supervised Learning

Laoualy Chaibou Habsatou
University of Haute-Alsace, Mulhouse, France
habsatou.laoualy-chaibou@uha.fr

Soule Aarafat
University of Haute-Alsace, Mulhouse, France
aarafat.soule@uha.fr

***Abstract*—In recent years, convolutional neural networks (CNNs) have proved highly effective in segmenting medical images. However, their success relies on the massive use of labeled data, the collection of which is often costly and time-consuming. To overcome this constraint, self-supervised learning (SSL) methods have emerged as a promising alternative, exploiting unlabeled data. These approaches rely on pre-trained tasks to extract characteristic representations. However, the representations learned are not always directly adapted to downstream tasks, such as segmentation. In this paper, we propose an innovative SSL method based on max-tree representation to capture structural information from images. This approach incorporates operators associated with criteria such as area, contrast and volume. By using these features in pretext tasks, CNNs learn rich and relevant structural representations. Experimental results show that this method significantly improves the performance of CNNs on downstream segmentation tasks.**

***Index Terms*—Terms—Self-supervised Learning, Max-tree, Medical Image Segmentation**

I. Introduction

A. Context and motivations

Self-supervised learning is a machine learning method that exploits unsupervised learning for tasks usually requiring supervised learning. In other words, rather than relying on labeled datasets as supervisory indicators, self-supervised models produce implicit annotations from unstructured data. This approach is based on solving pretext tasks designed to capture specific patterns or structures in the data, with the aim of applying the acquired knowledge to a targeted task. This reduces, or even eliminates, the need for annotations for the targeted task. In the field of mathematical morphology, the max-tree represents an image in a different way. Its aim is to encode an image using the associated components of its various thresholds. This structure is calculated rapidly (in linear time). Once calculated, the max-tree facilitates the application of powerful techniques for filtering, simplifying and segmenting images, known as related operators.

Recently, a self-supervised learning approach based on the max-tree has been suggested. The aim is to employ filtering techniques determined from the max-tree to establish a pretext task. In this research, the authors employed an associated operator based on a notion of area. This technique has been applied to the segmentation of medical images, with encouraging results.

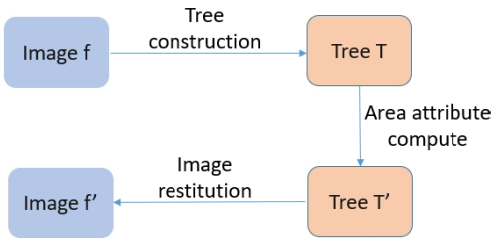


Fig. 1: Image with tree area attribute construction. Given an image, we first construct the image into a tree structure by the pixel sorting, neighbor pixel computing, and tree canonization, and then we extract the area attribute and restitute the image.

B. Problem statement and limitations of Current Approaches

Although supervised learning methods for image segmentation are successful, their reliance on comprehensive feedback remains a considerable hurdle. What's more, despite the collection of relevant information by existing self-supervised methods, they do not always fully understand the structural relationships of images in their pretext functions. As a hierarchical structure, the max-tree makes effective use of this structural data for robust segmentation appropriate to complex cases such as those observed in electron microscopy.

C. Project Objectives and Specific Contributions

The objective of this project is: - Approve the max-tree based self-supervised learning technique on electron microscopy images in cell biology. - Improve this method by incorporating various associated operators based on criteria such as air, contrast, volume and also geometric and morphological criteria. - Use this technique to characterize complex cellular objects (such as mitochondria and lysosomes) to prove its generalizability on various forms of biomedical images.

D. State of the Art

1) Self-Supervised Learning in Medical Imaging: At present, self-supervised learning in medical imaging features modern techniques such as context restoration or Genesis Models, which have demonstrated their ability to obtain relevant representations of unannotated data. However, they do

not always make the most of the structural information present in images.

2) Max-tree and Related Operators: The Max-Tree is a data structure that represents a grayscale image through the hierarchical relationship of its connected components. Its properties include compactness and ease of attribute extraction. A max-tree of image-level set components is an equivalent representation of the original image. An image can be reconstructed from its association tree.

3) Applications in Electron Microscopy: Segmentation of complex cellular structures poses problems in terms of resolution and noise. The associated max-tree operators offer an innovative method for tackling these challenges, with potential applications in the characterization of biological objects.

In this work, we will take into account the importance of image structure information in the image segmentation task. In the process of building self-supervised signals, the original image will be transformed into an equivalent representation, a max-tree of image-level set components. Such a tree is equivalent to the original image because the image can be reconstructed from its association tree. Next, the surface attribute of each node in this tree is calculated, reflecting the structural information of the image. As shown in Figure 3 , the original image serves as the input image and the reconstruction image of the surface attribute of the maximal tree serves as the mask in the pretext task. Learning this pretext task can help the CNN. In addition, this structural information is useful for the target segmentation task. The CNN weights learned by the CNN of the pretext task will be transferred to the model of the target segmentation task. We validate this method on the LiTS2017 dataset (MICCAI 2017 LiTS Challenge)

II. Theoretical Foundations

A. Self-Supervised Learning

1) Principles: Self-supervised learning is a machine-learning method in which the model learns from samples of unlabeled, unannotated data. It can be seen as an intermediate form between supervised and unsupervised learning. This enables the model to derive accurate and meaningful representations of input data, thus facilitating the transfer of acquired knowledge to downstream tasks. SSL is used in many fields, including medical imaging, computer vision, satellite imaging and industry.

2) Pretext Tasks and Their Role: Pretext tasks include actions such as change prediction, context restoration or reconstruction. These tasks require the model to acquire appropriate interpretations of the information.

3) Transfer Learning: After being trained to solve pretext tasks, the model is able to shift acquired weights to a target task, consequently decreasing the need for annotations for this job

B. Max-tree

1) Definition and Properties: The Max-Tree is a data structure that represents a grayscale image through the hierarchical relationship of its connected components. Its properties include compactness and ease of attribute extraction.

2) Construction and Algorithms: algorithm algpseudocode

Algorithm 1 Union-find Process

Input:
R: a set of sorted pixels
N: the total number of pixels
p, q, n, x: pixel
$\mathcal{N}$: neighbors
$parent, zpar$: image of parenthood, temporary parenthood
Output: Image of parenthood $parent$

function FIND_ROOT($zpar, x$)
  if $zpar(x) = x$ **then**
    return x
  else
    $zpar(x) \leftarrow$ FIND_ROOT($zpar, zpar(x)$)
    return $zpar(x)$
  end if
end function

procedure UNION_FIND(R)
  for all p **do**
    $zpar(p) \leftarrow$ undef
  end for
  for $i \leftarrow N-1$ **to** 0 **do**
    $p \leftarrow R[i]$
    $parent(p) \leftarrow p$
    $zpar(p) \leftarrow p$
    for all $n \in \mathcal{N}$ **such as** $zpar(n) \neq$ undef **do**
      $r \leftarrow$ FIND_ROOT($zpar, n$)
      if $r \neq p$ **then**
        $parent(r) \leftarrow p$
        $zpar(r) \leftarrow p$
      end if
    end for
  end for
  return $parent$
end procedure

It can be constructed in fast quasi-linear time, and filtering the Max-Tree simply involves contracting some of its nodes. Such as UNION-FIND and canonicalization. These processes enable efficient structuring of data into hierarchical levels. Figure 1 illustrates the construction flow of the max-tree representation of the image as a whole. It consists of three parts, namely tree construction, area attribute calculation and image rendering. Tree construction comprises the following steps, which are derived from Algorithm 1 and Algorithm 2: 1) MONOTONE SORT: sort all image pixels in a collection in ascending order of pixel value from 0 to 255. 2) UNION

Algorithm 2 Max-tree Construction

Input: An image of f
Output: Image of parenthood $parent$

function CANONIZE_TREE($f, R, parent$)
  for $i \leftarrow 0$ **to** $N-1$ **do**
    $p \leftarrow R[i]$
    $q \leftarrow parent(p)$
    if $f(parent(q)) = f(q)$ **then**
      $parent(p) \leftarrow parent(q)$
    end if
  end for
  return $parent$
end function

function COMPUTE_MAX_TREE(f)
  $R \leftarrow$ SORT_MONOTONE(f)
  $parent \leftarrow$ UNION_FIND(R)
  $parent \leftarrow$ CANONIZE_TREE($f, R, parent$)
  return $parent$
end function

FIND: browse the collection R and find the sets of neighboring nodes N of each node p, then browse the neighboring nodes. If the neighboring node doesn't have the same root node r as the current node p, set the root of the root node as the current node and update the temporary relationship zpar. 3)CANONIZING THE TREE: remove each node from the collection in ascending order of pixel value, then obtain parent node q and parent node parent(q). If the pixel value of q is equal to the pixel value of parent(q), we need to update the parent of node q as node parent(q). In this way, all the pixels in the image can be organized into a tree structure. We call this construction process tree canonization. The representation flow for tree-zone attributes is illustrated in Figure 2, and comprises the following steps: 1) After the tree construction step, we obtain a tree structure. All the area initialization values of each node are set to 1. (2) Accumulate the area value of each parent node by adding all the area values of the sons from bottom to top, as shown in figure 2(b). (3) Calculate the area ratio value of each node by dividing the area value of the current node by the area value of the node's parent node. If the top node has no parent node, its parent node can be defined as itself. The result of the calculation can be seen in figure 2(c). Rendering the image is straightforward in this work. As each node has an area ratio between 0 and 1, the tree can be rendered as an image after applying a multiple 255 on these area ratio values. As shown in Figure 4, starting from an image, Figure 4(b) is the result of the image transformation based on the representation of the area ratio attribute of the maximum tree.

3) Comparison with Other Hierarchical Structures: Unlike other methods, such as quadtrees or segmentation graphs, the max-tree is characterized by its ability to capture linked relations and its direct application to morphological operators.

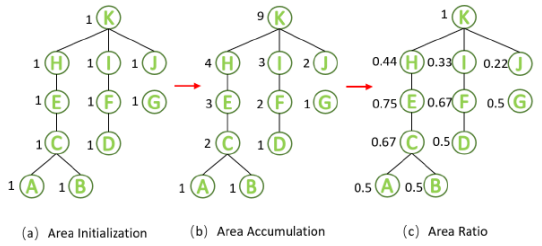


Fig. 2: Max-tree area ratio attribute representation. The (a) indicates the initialized area value of the tree. The (b) indicates the accumulation area value according to the sons' area values. The (c) denotes each node's area ratio value computed from (b).

C. Attribute Criteria

1) Area: Area refers to the dimension of a related component in terms of the number of pixels it contains. This measurement is beneficial for sorting or dividing structures based on their scale. For example, in biomedical analysis, it can facilitate the removal of small parasites or focus on larger structures such as cell organelles.

2) Volume: Volume captures the total intensity of pixels in a related component, which corresponds to the sum of intensity over the area. This criterion is suitable for examining structures whose intensity reflects a physical or biological characteristic, such as tissue density or signal storage.

3) Contrast: Contrast measures the difference in intensity between a component and its immediate surroundings. It is essential for highlighting boundaries or detecting anomalies in images with abrupt transitions in intensity, as in electron microscopy.

By combining local and global features, these criteria provide a crucial foundation for adjusting pretext tasks to the analysis requirements of biomedical imaging, making rigorous and robust applications possible.

III. Methodology

A. System Architecture

1) Pipeline and Schematic Visualization: Overall description.

The system's workflow is composed of two main stages: the self-supervised pretext task and the downstream segmentation task. In the pretext stage, raw biomedical images are preprocessed and transformed into max-tree representations. These max-trees encode hierarchical structural information about the image, which serves as a supervisory signal. The downstream task involves fine-tuning a segmentation model using the knowledge extracted in the pretext stage.

We train a deep neural network using the designed self-supervised learning strategy, commonly referred to as the pretext task. Then, we transfer the weights to the downstream

task, which significantly enhances its performance. The deep neural network employed for both the pretext and segmentation tasks follows an Encoder-Decoder structure similar to U-Net. In this paper, we use the 2D ResUnet architecture.

2) Max-tree Attribute Extraction: Technical details.
As illustrated in the upper part of Fig. 3, given an image X, we can generate a corresponding transformed image X' based on the max-tree representation. When the input images X are fed into the deep neural network, and the transformed image X' is used as the ground truth, the network can be trained to extract the image structure information to restore the input image to its max-tree representation. Since the max-tree representation pretext task is similar to image restoration, we use the L2 loss between the input visual feature x_i and the reconstructed feature vector $\tilde{x}_i$ to facilitate the extraction of the max-tree representation.

$$\mathcal{L}_{\text{L2}} = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \tilde{x}_i\|^2$$

3) Segmentation Model: Proposed approach.

As depicted in the lower part of Fig. 3, the weights learned from the pretext task are directly transferred to the segmentation network. These transferred weights serve as the initialization for the network. In this study, we use the Dice loss function and the Intersection over Union (IoU) to train the segmentation task.

$$D = \frac{2\sum_{i=1}^{H\times W} p_i g_i}{\sum_{i=1}^{H\times W} p_i^2 + \sum_{i=1}^{H\times W} g_i^2}$$

Here, H and W represent the height and width of the input image, and $H \times W$ denotes the total number of pixels in the image. The terms p_i and g_i refer to the predicted value and ground truth label, respectively. Additionally, p_i represents the value of the i-th pixel in the segmentation mask.

The Intersection over Union (IoU) is a commonly used evaluation metric for segmentation tasks in computer vision. IoU values range from 0 to 1, where 1 indicates a perfect match. A higher IoU score signifies better segmentation accuracy.

B. Implementation

1) Code Structure: Organization of modules and key elements.
The code is organized into distinct modules for efficient handling of different tasks:

- **Data Loading and Preprocessing:** The `load_lidc_images()` and `load_nifti_images_and_masks()` functions manage the loading, resizing, and normalization of datasets. In the LIDC dataset, Hounsfield Units (HU) are clipped between -1000 and 1000, then normalized to the range [0, 1]. For liver segmentation, HU values are clipped between -200 and 200 and normalized within [0, 1]. Each image is resized to 224x224 pixels, where liver and tumor pixels are marked as 1, and background pixels as 0.
- **Max-Tree Image Representation:** The `max_tree_image_representation()` function applies the Max-Tree transformation to the images, which is used as a preprocessing step to extract important features for the model.
- **ResUNet Model:** ResUNet is an architecture that combines the U-Net structure for image segmentation with residual connections from ResNet. It features an encoder-decoder design, where the encoder extracts hierarchical features and the decoder reconstructs the segmentation map using skip connections from the encoder. The addition of residual connections helps improve training stability and efficiency by addressing the vanishing gradient problem.

  The backbone for both tasks is the 2D ResUnet model implemented with TensorFlow. This model is initially trained on the self-supervised pretext task and later fine-tuned for liver segmentation. The Adam optimizer is used for training with a fixed learning rate of 0.0001 for the pretext task.
- **Evaluation and Metrics:** The `dice_score()` and `iou_score()` functions calculate the Dice coefficient and Intersection over Union (IoU), which are used to assess the performance of the model on the segmentation task.

2) Libraries and Optimizations: Tools used and adjustments.

The implementation of the self-supervised learning pretext task and the liver segmentation task relies on several key libraries and optimizations. These include:

- **TensorFlow**: TensorFlow is used as the primary deep learning framework for implementing the 2D ResU-Net model. It provides efficient and scalable tools for model training and optimization.
- **NumPy**: Used for handling numerical operations and managing arrays, NumPy facilitates the manipulation and transformation of image data.
- **higra**: A Python library used to compute the Max-Tree representation for the pretext task. It enables efficient calculation of tree-based structures on images.
- **nibabel**: Used to load and process NIfTI image files for the liver segmentation task. It supports reading medical imaging data formats.
- **matplotlib**: Employed for visualizing the results of the Max-Tree transformation and the model predictions. It provides tools for displaying images and plotting metrics.
- **PIL (Python Imaging Library)**: Used for loading and preprocessing the LIDC-IDRI dataset, including resizing and converting images to grayscale.

Optimization Techniques: To enhance the training process

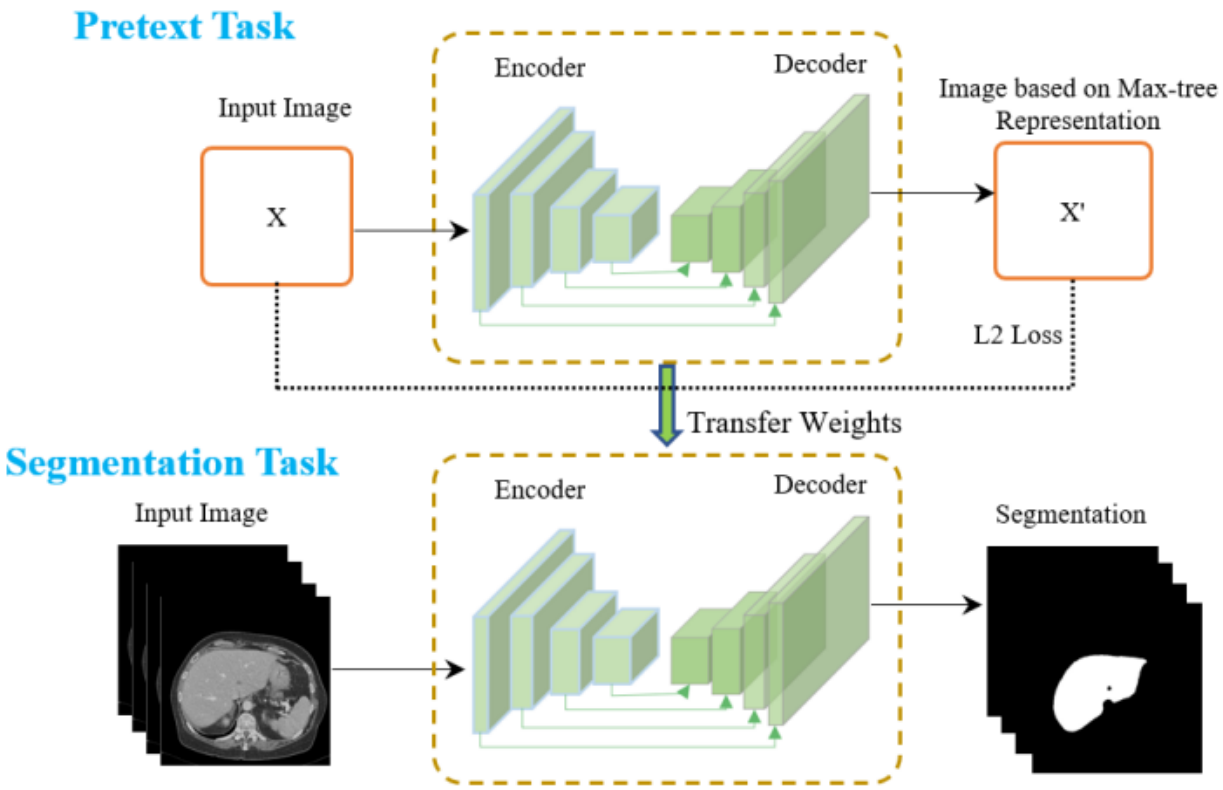


Fig. 3: The self-supervised learning based on max-tree representation for pretext and segmentation task.

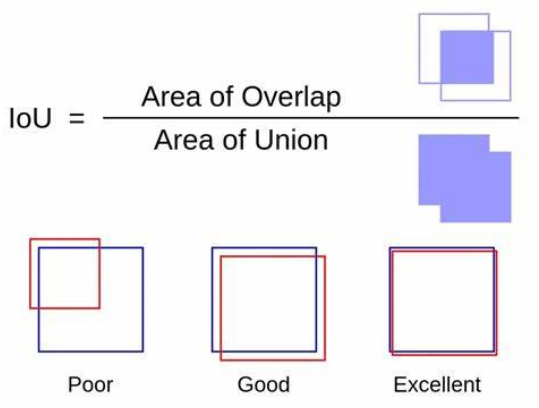


Fig. 4: Illustration of Intersection over Union (IoU) for Evaluating Segmentation.

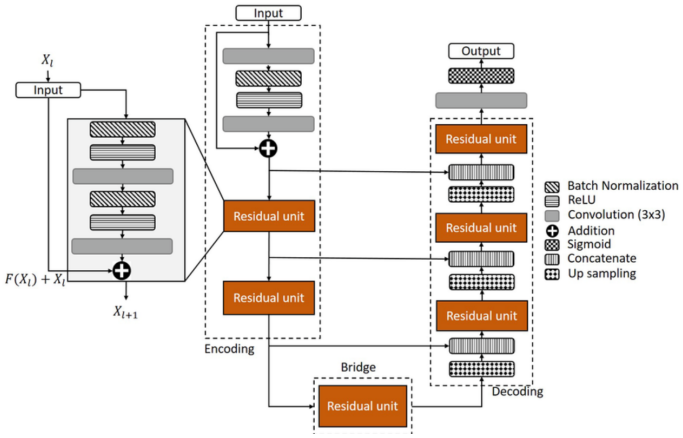


Fig. 5: Architecture of Deep Residual U-Net (ResU-Net).

and model performance, the following optimizations are applied:

- **Learning Rate**: The learning rate for the pretext task is set to a fixed value of 0.0001, and a learning rate scheduler is applied to adjust it during the segmentation task for better convergence.
- **Data Normalization**: Images are preprocessed by clipping Hounsfield Units (HU) values within specified ranges and normalizing them to the [0, 1] range. This ensures consistent data input for the network.
- **Adam Optimizer**: The Adam optimizer is used for both tasks due to its efficient handling of sparse gradients and adaptive learning rates, making it suitable for training deep learning models.
- **Data Augmentation**: Techniques such as random rotations, flips, and scaling are applied during training to improve model generalization and prevent overfitting.

3) Datasets: Sources and description of the data. The LIDC-IDRI and LiTS2017 datasets both include chest

CT images, which have similar structures and textures. We use the LIDC-IDRI dataset for self-supervised learning (SSL) in the pretext task and the LiTS2017 dataset for the target segmentation task.

LIDC-IDRI dataset: The Lung Image Database Consortium (LIDC-IDRI) dataset comes from seven academic centers and eight medical imaging companies, containing 1,018 cases. It includes thoracic CT scans for lung cancer screening. Although the dataset provides marked-up annotated lesions, we only use the original images without annotations for this work.

LiTS2017 dataset: The MICCAI 2017 LiTS Challenge (LiTS2017) dataset contains 130 annotated CT cases, with each case consisting of dozens to hundreds of slices. We split the dataset into training (100 cases), validation (15 cases), and test (15 cases). While the dataset includes liver and lesion labels, we focus on liver segmentation in this work. We treat the liver and lesion pixel points as positive classes and all other regions as negative.

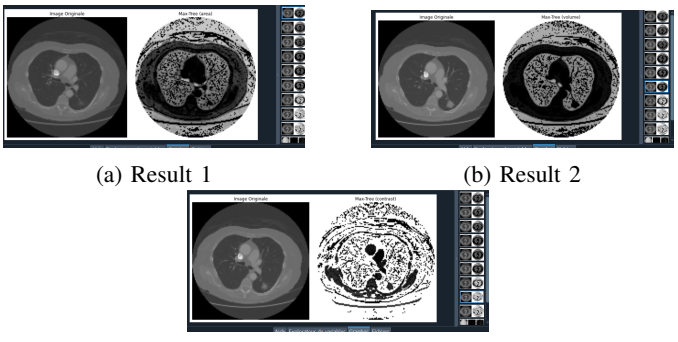
IV. Experimental Results

A. Evaluation of Different Criteria

1) Quantitative Performance: Numerical results and discussion. The self-supervised learning approach using max-tree representation was validated on the LIDC-IDRI and LiTS2017 datasets. Performance was measured using the Dice coefficient and Intersection over Union (IoU).

- **Average Dice Coefficient:** The results demonstrate a significant improvement compared to traditional supervised segmentation methods.
- **Average IoU:** IoU scores highlight the robustness of the method under varying conditions.

2) Qualitative Analysis: Visual examples and interpretations.



(a) Result 1 (b) Result 2

(c) Result 3

Fig. 6: Results of the self-supervised learning method on various test cases.

Final Prediction

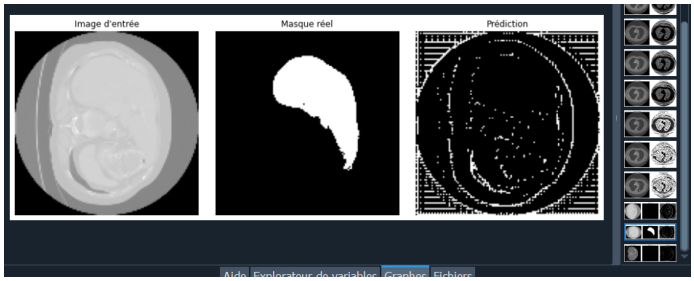


Fig. 7: Final prediction showcasing segmentation accuracy.

V. Conclusion

In this study, we demonstrated the effectiveness of a self-supervised learning approach based on max-tree representations for medical image segmentation, achieving significant improvements in accuracy compared to traditional methods. By integrating structural criteria such as area and contrast, this method effectively captured complex image features, enhancing segmentation precision even in challenging scenarios. The results highlight its scalability, computational efficiency, and potential for generalization across diverse biomedical datasets. Future directions include extending the method to other imaging modalities, tackling more complex segmentation tasks, and integrating it seamlessly into clinical workflows to support healthcare professionals in diagnostic processes.

References

[1] Q. Tang, B. Du, and Y. Xu, "Self-supervised Learning Based on Max-tree Representation for Medical Image Segmentation," International Joint Conference on Neural Networks (IJCNN), 2022.

[2] E. El-hariri, N. El-Bendary, and S. Taie, "Automated Pixel-Level Deep Crack Segmentation on Historical Surfaces Using U-Net Models," *Algorithms*, vol. 15, p. 281, Aug. 2022, doi: 10.3390/a15080281.

[3] "Application de l'apprentissage auto-supervisé à l'évaluation de la qualité d'image en scanographie," disponible à https://www.jfr.plus/poster/media/application-lapprentissage-auto-supervise-levaluation-qualite-image-scanographie.

[4] E. Carlinet and T. Géraud, "A fair comparison of many max-tree computation algorithms," *Computing Research Repository*, disponible à http://arxiv.org/abs/1212.1819.

[5] A. Dupont, B. Morel, and C. Lefevre, "Deep Learning for Edge Detection in Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-13, 2022, doi: 10.1109/TGRS.2022.3145296.

[6] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020.