Multi-scale Region-aware Implicit Neural Network for Medical Images Matting

Yanyu Xu[1], Yingzhi Xia[2], Huazhu Fu[2], Rick Siow Mong Goh[2], Yong Liu[2], Xinxing Xu[2*]

[1] The Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan, 250100, P. R. China. xu_yanyu@sdu.edu.cn

[2] The Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore. xuxinx@ihpc.a-star.edu.sg

Abstract. Medical image segmentation is a critical task in computer-assisted diagnosis and disease monitoring, where labeling complex and ambiguous targets poses a significant challenge. Recently, the alpha matte has been investigated as a soft mask in medical scenes, using continuous values to quantify and distinguish uncertain lesions with high diagnostic values. In this work, we propose a multi-scale regions-aware implicit function network for the medical matting problem. Firstly, we design a regions-aware implicit neural function to interpolate over larger and more flexible regions, preserving important input details. Further, the method employs multi-scale feature fusion to efficiently and precisely aggregate features from different levels. Experimental results on public medical matting datasets demonstrate the effectiveness of our proposed approach, and we release the codes and models in GitHub.

Keywords: Implicit Neural Network · Medical Images Matting.

1 Introduction

Medical image segmentation, essential for diagnosis and monitoring [7] [10] [26], faces challenges with ambiguous targets and consensus among clinicians using binary masks. The alpha matte, using continuous values for soft masking, offers a solution for quantifying and distinguishing uncertain lesions effectively.

Matting assumes the image I is a mixture of foreground F and background B, with the alpha matte α representing the mixing coefficients. This is expressed as $I = \alpha F + (1 - \alpha)B$, where α ranges from 0 to 1. In medical matting [25] [24], the diseased lesion is the foreground F, and the normal tissue is the background B. As shown in Fig. 1 (a), it's more difficult for multiple clinicians to reach a consensus annotation using binary masks. The alpha matte in Fig. 1 (b) could pay attention to the uncertainties related to the characteristics of lesions and has better representation capability than simple binary masks. Recent methods have

* : Corresponding author

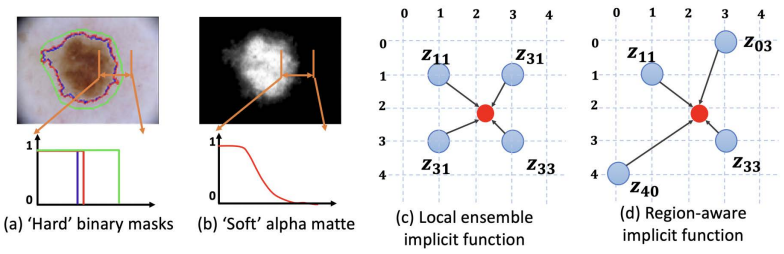

Fig. 1. Illustrations of motivations. (a) 'Hard' binary masks meet challenges with ambiguous targets and consensus among clinicians; (b) 'Soft' alpha matte uses continuous values for quantifying and distinguishing uncertain lesions; (c) Local ensemble implicit function decodes feature maps within a fixed region with limited ability to interpolate over larger regions; (d) The region-aware implicit function interpolates over larger and more flexible regions, preserving important details.

leveraged image matting to refine mask boundaries for improved segmentation performance [27] [7] or to construct trimaps as an aid for more precise manipulation of uncertain regions [10] [28] [15] as well as focus on medical matting [25] [24] in the views of uncertainty. They usually apply U-Net [22] as the backbone and ignore feature collaboration at different scales. However, the commonly used bilinear up-sampling and convolutions on feature maps of different scales might blur the precise information learned in these feature maps.

To efficiently and precisely aggregate features from different levels, we introduce an implicit neural function to define continuous feature maps and align multi-scale features. The implicit neural representations are recently designed and use multi-layer perceptron (MLP) to map coordinates to signals, including representing objects and scenes in 3D reconstruction [13], image super-resolution [6], decoding RGB values in image super-resolution [6] or feature alignment [12]. Local implicit image function [12] decodes from original feature maps within a fixed near region (e.g., 2×2) around the query coordinate, as shown in Fig. 1 (c). These methods might have limited ability to interpolate over larger regions.

To address this limitation, we propose a new region-aware implicit function to interpolate over larger and more flexible regions. It could be learned to preserve important details in the inputs, as depicted in Fig. 1 (d). The features can be viewed as latent codes distributed in spatial dimensions and each latent code will represent a field of information. The information from the fixed local near regions, e.g. 2×2 in local implicit image function, might be limited, especially for the boundary regions. In particular, the region-aware implicit function first learns from the features to determine the dynamic regions and might find similar information from far away. Inspired by the deform convolution, we employ its PyTorch implementation to realize the region-aware learning process. Moreover, we employ multi-scale feature fusion to enhance the quality of predictions by utilizing different levels of features. The Deformable convolution expands the network's receptive field, but coarser scale latent codes (as in IFA and IOS-

Net) have limitations in capturing fine-grained details, especially in small or arbitrary regions. Region-Aware mechanism addresses this by introducing offset mechanisms on feature maps at each scale, enhancing the capture of detailed information across scales and preserving tiny and intricate features that might otherwise be lost. Besides, there exist the following clinical values of new modules: 1) Detailed information from ambiguous and small regions enhances the accuracy of identifying and characterizing lesions, improving the overall diagnostic process. 2) Clear delineation of ambiguous and small regions reduces the risk of misdiagnosis, ensuring that both benign and malignant areas are correctly identified and treated appropriately.

The main contributions of this paper can be summarized as follows: Clinically, our method enhances the accuracy of lesion segmentation by preserving intricate details in ambiguous areas, which is crucial for precise diagnosis and treatment planning. Additionally, the multi-scale feature fusion mechanism ensures that even tiny and arbitrary regions are accurately captured, preserving critical details that might be overlooked by traditional methods. Technically, we introduce a region-aware implicit neural representation that interpolates over larger and more flexible regions, preserving important details missed by conventional locality-based approaches. Our multi-scale feature fusion mechanism integrates features across different scales, enhancing prediction quality and ensuring detailed information capture. The experiments have shown the effectiveness of the proposed method on three public medical matting datasets. The codes and models are released on `https://github.com/xuyanyu-shh/MedicalMattingMLP`.

2 Method

2.1 Overview

The network directly takes the medical images as inputs to simulate the alpha matte predictions, without trimap as inputs like most existing matting work [1] [17], considering the practice and difficulty of obtaining trimaps in medical diagnosis scenes. Similar to [24], the network employs a Probabilistic UNet to output the uncertainty maps as trimaps, generated from multiple binary segmentation maps. The binary masks predicted by the Probabilistic UNet can be regarded as the simulation of clinicians' labeling. Then we feed the uncertainty maps as the trimaps to the Matting UNet to predict alpha matte.

Probabilistic UNet. Following the medical matting work [24], we also have a Probabilistic UNet [8] to generate uncertainty maps as auxiliary information. It produces a bunch of binary masks and uses the intermediate score maps to build an uncertainty map. To note that, the uncertainty map indicates the challenging areas in continuous values, which could be regarded as trimap, in a similar even same role in the matting problem. In particular, the uncertainty map U_{map} is defined as the entropy $U_{map}(x_i) = -\sum_{c=1}^{C} \bar{p}^c(x_i) \log \bar{p}^c(x_i)$, where C is the number of classes, and $\bar{p}^c(x_i)$ is the probability of the pixel x_i in class c of the average score map of the Prob. UNet predictions. Suppose we generate N score maps $\hat{p}_1^c, \hat{p}_2^c, \hat{p}_3^c, ..., \hat{p}_N^c$ per image class, then $\bar{p}^c(x_i) = \frac{1}{N} \sum_{n=1}^{N} \hat{p}_n^c(x_i)$.

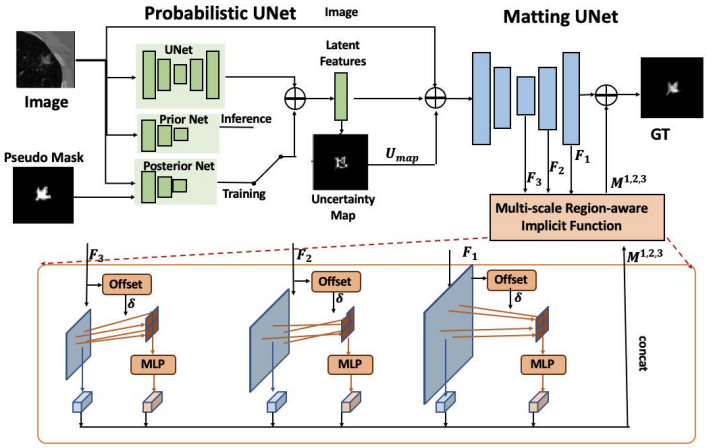

Fig. 2. Overview of the proposed method. The Probabilistic UNet outputs the uncertainty maps as trimaps and Matting UNet receives it to predict alpha matte.

Matting UNet. We employ a U-Net [22] as a backbone to generate the final alpha matte prediction, as shown in Fig. 2. The output block consists of two convolution layers at the end of the pipeline. It takes the concatenation of input image and latent features from the Prob. UNet as inputs. Furthermore, the uncertainty map is also applied to the last two propagation units, which provide constraint information [4] [24].

2.2 Region-aware Implicit Function

We propose a new region-aware implicit neural function by interpolating over larger and more flexible regions to preserve important details in the inputs. Further, to efficiently and precisely aggregate features from different levels, we involve a multi-scale feature fusion to align multi-scale features.

Vanilla and Local Ensemble Implicit Functions. An implicit feature function is utilized to obtain a continuous feature map from a discrete feature map, which can be decoded at any coordinate. The decoding function f_θ, typically an MLP, is defined over the discrete feature map, where feature vectors are considered as latent codes distributed evenly in the 2D space and assigned with 2D coordinates. The feature value at the coordinate (u_q, v_q) in the feature map M is $M(u_q, v_q) = f_\theta(z^*, u_q - u^*, v_q - v^*)$, where z^* is the interpolated latent code from (u_q^*, v_q^*) location and (u_q^*, v_q^*) is nearest one to (u_q, v_q). The local ensemble implicit feature function [6] directly decodes from the original feature maps in a fixed region, such as 2×2, around the query coordinate:

$$M(u_q, v_q) = \sum_{i \in \{00,01,10,11\}} f_\theta(z_i^*, u_q - u_i^*, v_q - v_i^*), \tag{1}$$

where $z_i^*(i \in \{00, 01, 10, 11\})$ are the nearest latent code in top-left, top-right, bottom-left, bottom-right sub-spaces.

Region-aware Implicit Function. We introduce a novel region-aware implicit function for larger and more flexible regions, as shown in Fig. 2. The proposed region-aware implicit function also decodes from the original feature maps in a $2 \times 2 = 4$ feature code around the query coordinate, while the feature codes might be near around the query coordinate and far away from it. In particular, the feature value at x_q in the feature map M is defined as

$$M(u_q, v_q) = \sum_i^4 f_\theta(z_i^*, u_q - u_i^*, v_q - v_i^*), (u_i^* \in [0, H^W], v_i^* \in [0, W^M]), \quad (2)$$

where H^W and W^M are the width and height of the feature map M. f_θ is a two-layer MLP. To learn flexible regions, we employ an offset convolution layer to learn the offset, taking the feature value at x_q in the continuous feature map M as input. Then, the decoding function f_θ decodes directly from the original feature maps from the learned positions. The decoding function f_θ is jointly learned with the whole matting U-Net, enabling the learned features to precisely represent continuous fields of information.

Both the vanilla implicit function and local ensemble implicit function use distance measurement to decode the nearest one or the top four original features around the query. Unlike them, our region-aware implicit function pays more attention to the feature itself and decodes the similar or related original features around the query. The regions determined by the learned offsets have similar patterns, as shown in the Region-aware Implicit Function Visualization of learned offset in the Experiment Section.

Multi-scale Feature Fusion. To take advantage of different levels of features, we employ a multi-scale feature fusion to further improve the quality of the predictions. Taking aligning the feature maps $\{F\}_{i=1}^3$ as an example, we extend the region-aware implicit function to fuse multi-scale features. The resolutions of feature maps F_1, F_2, F_3 are $H \times W$, $H/2 \times W/2$ and $H/4 \times W/4$, respectively. It defines a continuous feature map $M^{1,2,3}$ over multi-level discrete feature maps in different resolutions. Specifically, we define the the value of $M^{1,2,3}$ as

$$M^{1,2,3}(u_q, v_q) = [F^1; M^{2,3}(u_q, v_q)], M^{2,3}(u_q, v_q) = [F^2; M^3(u_q, v_q)], \quad (3)$$

where $[;]$ is the concatenation operation and $M^3(u_q, v_q)$ is obtained from equation (2). Intuitively, each latent code still represents a field of feature that can be decoded by relative coordinate, and f_θ can decode the field for each level and model the interaction across different levels at the same time. We use the augmented and aligned feature map $M^{1,2,3}(u_q, v_q)$ to predict the alpha matte.

2.3 Implementation of Region-aware Implicit Function

The proposed region-aware implicit function needs to operate on features located in arbitrary places, rather than a regular rectangular region. To reduce

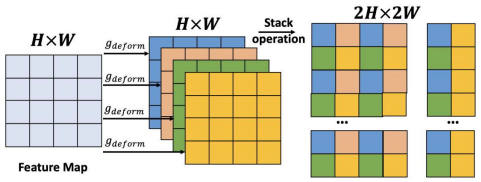

Fig. 3. Implementation of Region-aware Implicit Function. To learn flexible regions, an offset convolution layer (g_{deform}) is used to learn the offset, taking the feature value at x_q in the continuous feature map M as input.

the additional operation cost on decoding features at arbitrary places, we use the deformable convolution layer *deform_conv2d* on the PyTorch platform to efficiently implement the region-aware implicit function. As the output size of the deformable convolution is the same as the input, we repeat running it and stack the outputs to arbitrary resolution, as shown in Fig. 3. The final continuous feature map M in equation (2) is then defined as follows: $M = f_{\text{stack}}(\sum f_\theta(g_{\text{deform}}(x, x_\delta), x_\delta))$, where x_δ is the learned offsets from the offset convolution layer g_δ. f_{stack} is the stack operation.

2.4 Loss Functions

The total loss comprises the binary mask prediction loss L_{seg} and alpha matte prediction loss L_{mat}. For the binary mask prediction, we use a combination of cross-entropy loss L_{ce} and Kullback-Leibler loss L_{kl}. The former aims to match the predicted mask and the pseudo ground truth mask, while the latter is used to minimize the divergence between the prior distribution P and the posterior distribution Q [16]. For the alpha matte prediction, we use the absolute difference L_{mae} and the gradient difference between the predicted alpha matte and the ground truth alpha matte L_{grad}. Furthermore, we use an uncertainty map to generate a mask, which is applied to concentrate the gradient loss in the uncertain regions. To further improve performance, we adopt the uncertainty weighting strategy [14], as proposed in [24].

3 Experiment

3.1 Experimental Setting

We implemented our proposed model using the PyTorch framework [20]. The Adam optimizer is adopted with a base learning rate of 5×10^{-5}. All experiments are trained 100 epochs with a batch size of 4. We employed a cosine annealing schedule [3] [18] after a 1-epoch long steady increasing warm-up from 0 to base learning rate. To augment the data, we applied several techniques during the data pre-processing stage, including flipping, rotation, and elastic transformation [23].

Datasets	LIDC-IDRI			ISIC			Brain-growth		
Model	SAD	MSE	Grad	SAD	MSE	Grad	SAD	MSE	Grad
Bayesian	0.0778	0.0819	0.1535	7.7535	0.1624	9.2632	0.8435	0.1662	1.5921
Closed-Form	0.3040	0.4736	0.7584	21.7274	0.9062	2.7009	1.5419	0.4410	2.6960
KNN	0.0737	0.0451	0.1381	7.6282	0.1861	4.1263	0.6534	0.1073	1.1548
Information-Flow	0.0663	0.0351	0.1001	5.3062	0.1061	2.8643	0.6819	0.1056	1.5007
Learning Based	0.0554	0.0286	0.0826	8.4567	0.2113	4.8210	0.6061	0.0898	1.0559
FBA	0.0598	0.0395	0.1143	8.8235	0.2590	4.9446	0.7711	0.1390	1.2350
MatteFormer	0.0831	0.0224	0.1972	7.6443	0.2429	6.9984	1.0651	0.2443	1.9827
Medical matting	**0.0396**	0.0214	0.0587	1.1889	**0.0178**	0.2551	0.4150	0.0467	0.6015
Ours	0.0398	**0.0188**	**0.0540**	**0.3611**	0.0184	**0.0887**	**0.4032**	**0.0405**	**0.5411**

Table 1. Qualitative comparison of our proposed model and other state-of-the-art methods on three datasets using three evaluation metrics. (**Best**, Second)

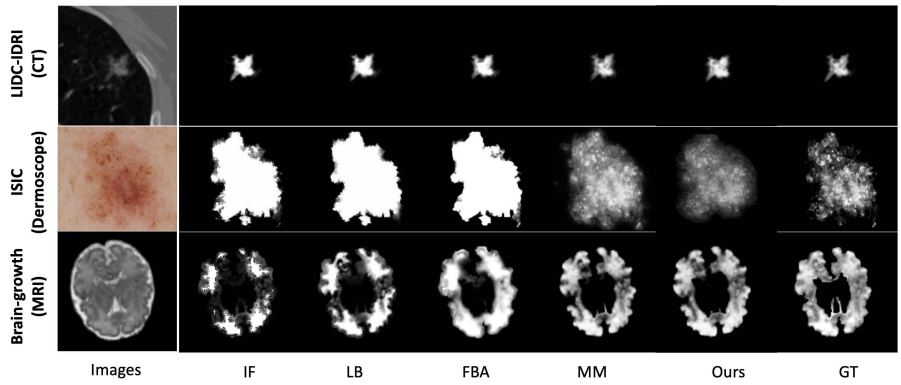

Fig. 4. Visual comparison of alpha matte generated by different methods.

Datasets. We use 3 public medical matting datasets: a subset of LIDC-IDRI [2], Brain-growth of QUBIQ [9], and a part of ISIC 2018 dataset [9]. To reduce the interference caused by random errors, we follow the settings in [24] to use part of two datasets for a fair comparison and perform a four-fold cross-validation.

Metrics. We use three commonly used evaluation metrics [21]: absolute differences (SAD), mean squared error (MSE), and gradient (Grad.).

3.2 Performance Comparison

We perform a comprehensive evaluation of our proposed model on the three medical matting datasets, employing three commonly used evaluation metrics. To benchmark the performance of our method, we compare it against several state-of-the-art matting methods, including a Bayesian-based method (Bayesian [8]), four Laplacian-based methods (ClosedForm [17], KNN [5], Information-Flow [1], Learning Based [29]), a deep learning-based method (FBA [11]), MatteFormer [19] and a medical matting method with uncertainty [25].

Upsampling	Scales	Kernel size	SAD	MSE	Grad
Bilinear	3	-	0.0401	0.0194	0.0550
Region-aware	3	2x2	0.0398	0.0188	0.0540
Region-aware	1	2x2	0.0397	0.0198	0.0559
Region-aware	2	2x2	0.0396	0.0194	0.0550
Region-aware	3	2x2	0.0398	0.0188	0.0540
Region-aware	3	1x1	0.0406	0.0198	0.0554
Region-aware	3	2x2	0.0398	0.0188	0.0540
Region-aware	3	3x3	0.0401	0.0196	0.0551

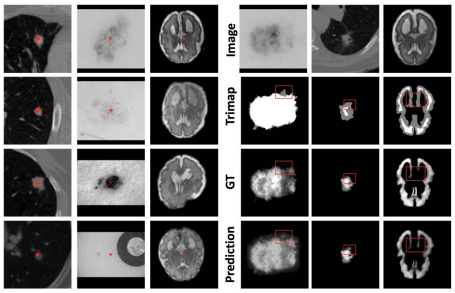

Table 2. The Left: The ablation studies on the LIDC-IDRI dataset. **The Middle**: The visualization of learned offsets in red points. **The Right**: More qualitative analysis.

Qualitative Comparison Table 1 shows the qualitative comparison of the results. We can see that our model outperforms other methods in terms of MSE and Grad metrics and achieves comparable performance in terms of SAD metric on the LIDC-IDRI dataset. On the ISIC dataset, ours performs much better than others in terms of SAD and Grad metrics, and achieves comparable performance in terms of MSE metric. Our model achieves better performance on the Brain-growth datasets. It indicates our method better expresses the edge of the fuzzy transition zone and subtle structural features in the matting results.

Quantitative Comparison We also show some example predictions in Fig. 4. The differences between the foreground and background of medical images are sometimes less prominent than that in natural scenes, and even sometimes the foreground area is hard to give a precise range, and the non-foreground component in the foreground leads to the failure of the trimap mechanism. The IF, LB, FBA methods tend to produce over-segmented or binary mask-like results, missing finer details. Both MM and ours provide a more detailed alpha matte closer to the ground truth (GT). Our method captures the complex texture of the lesion more effectively, maintaining details in ambiguous areas. The additional visual examples on the Right in Table 2 highlight finer details in ambiguous and tiny areas to demonstrate the new clinical values more comprehensively.

3.3 Ablation Studies

We conducted an extensive ablation study on the Brain-growth dataset to investigate the properties of our proposed method and its components.

Effect of the implicit function: To assess the impact of the implicit function module, we designed a baseline model that replaced it with simple upsample operations. The results in the Left on Table 2 revealed a significant performance gap, indicating the importance of adaptive and dynamic scales in interpolation.

Effect of different scales: We evaluated the impact of using different levels of features, ranging from 1, 2 to 3. The results in Table 2 Left, show that incorporating more information from multiple scales could improve model performance.

Effect of different kernel sizes: We also designed baselines using different kernel sizes, such as 1×1, 2×2, and 3×3. The results on the Left in Table 2

showed a performance drop for smaller and larger kernel sizes, highlighting the importance of using a 2×2 kernel size.
Visualization of learned offset: We visualize the learned offsets (in red points) in the Middle in Table 2. We can see that the regions determined by the learned offsets are more flexible and larger than the regular grids, such as 2×2.

4 Conclusion

We discussed the challenges of medical image segmentation and the recent exploration of using alpha matte as a soft mask to represent uncertain regions with high diagnostic value. We proposed a multi-scale regions-aware implicit function network for the medical matting problem to generate high-quality and resolution-free alpha matte. Experimental results on public medical matting datasets demonstrate the effectiveness of our proposed approach. Our work contributes to the development of accurate and efficient medical image segmentation, which assists clinicians in computer-assisted diagnosis and monitoring.

5 Acknowledgements

This work was supported by the Agency for Science, Technology, and Research (A*STAR) through its AME Programmatic Funding Scheme Under Project A20H4b0141, the National Research Foundation (NRF) Singapore under its AI Singapore Programme (AISG Award No: AISG2-TC-2021-003), the Agency for Science, Technology, and Research (A*STAR) through its RIE2020 Health and Biomedical Sciences (HBMS) Industry Alignment Fund Pre-Positioning (IAF-PP) (grant no. H20C6a0032), the 2022 Horizontal Technology Coordinating Office Seed Fund (Biomedical Engineering Programme – BEP RUN 3, grant no. C221318005),and the Career Development Fund (CDF) C233312010, and Taishan Scholars Program (Grant No. tsqn202312067). The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aksoy, Y., Ozan Aydin, T., Pollefeys, M.: Designing effective inter-pixel information flow for natural image matting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 29–37 (2017)
2. Armato III, S.G., McLennan, G., McNitt-Gray, M.F., Meyer, C.R., Yankelevitz, D., Aberle, D.R., Henschke, C.I., Hoffman, E.A., Kazerooni, E.A., MacMahon, H., et al.: Lung image database consortium: developing a resource for the medical imaging research community. Radiology **232**(3), 739–748 (2004)
3. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
4. Cai, S., Zhang, X., Fan, H., Huang, H., Liu, J., Liu, J., Liu, J., Wang, J., Sun, J.: Disentangled image matting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8819–8828 (2019)

5. Chen, Q., Li, D., Tang, C.K.: Knn matting. IEEE transactions on pattern analysis and machine intelligence **35**(9), 2175–2188 (2013)
6. Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8628–8638 (2021)
7. Cheng, J., Zhao, M., Lin, M., Chiu, B.: Awm: Adaptive weight matting for medical image segmentation. In: Medical Imaging 2017: Image Processing. vol. 10133, pp. 769–774. SPIE (2017)
8. Chuang, Y.Y., Curless, B., Salesin, D.H., Szeliski, R.: A bayesian approach to digital matting. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 2, pp. II–II. IEEE (2001)
9. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). pp. 168–172. IEEE (2018)
10. Fan, Z., Lu, J., Wei, C., Huang, H., Cai, X., Chen, X.: A hierarchical image matting model for blood vessel segmentation in fundus images. IEEE Transactions on Image Processing **28**(5), 2367–2377 (2018)
11. Forte, M., Pitié, F.: f, b, alpha matting. arXiv preprint arXiv:2003.07711 (2020)
12. Hu, H., Chen, Y., Xu, J., Borse, S., Cai, H., Porikli, F., Wang, X.: Learning implicit feature alignment function for semantic segmentation. In: European Conference on Computer Vision. pp. 487–505. Springer (2022)
13. Jiang, C., Sud, A., Makadia, A., Huang, J., Nießner, M., Funkhouser, T., et al.: Local implicit grid representations for 3d scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6001–6010 (2020)
14. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7482–7491 (2018)
15. Kim, T., Lee, H., Kim, D.: Uacanet: Uncertainty augmented context attention for polyp segmentation. In: Proceedings of the 29th ACM International Conference on Multimedia. pp. 2167–2175 (2021)
16. Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S., Jimenez Rezende, D., Ronneberger, O.: A probabilistic u-net for segmentation of ambiguous images. Advances in neural information processing systems **31** (2018)
17. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. IEEE transactions on pattern analysis and machine intelligence **30**(2), 228–242 (2007)
18. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
19. Park, G., Son, S., Yoo, J., Kim, S., Kwak, N.: Matteformer: Transformer-based image matting via prior-tokens. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11696–11706 (2022)
20. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach,

H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019)
21. Rhemann, C., Rother, C., Wang, J., Gelautz, M., Kohli, P., Rott, P.: A perceptually motivated online benchmark for image matting. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 1826–1833. IEEE (2009)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
23. Simard, P.Y., Steinkraus, D., Platt, J.C., et al.: Best practices for convolutional neural networks applied to visual document analysis. In: Icdar. vol. 3. Edinburgh (2003)
24. Wang, L., Ju, L., Zhang, D., Wang, X., He, W., Huang, Y., Yang, Z., Yao, X., Zhao, X., Ye, X., et al.: Medical matting: a new perspective on medical segmentation with uncertainty. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 573–583. Springer (2021)
25. Wang, L., Ye, X., Ju, L., He, W., Zhang, D., Wang, X., Huang, Y., Feng, W., Song, K., Ge, Z.: Medical matting: Medical image segmentation with uncertainty from the matting perspective. Computers in Biology and Medicine p. 106714 (2023)
26. Xu, Y., Zhou, M., Feng, Y., Xu, X., Fu, H., Goh, R.S.M., Liu, Y.: Minimal-supervised medical image segmentation via vector quantization memory. In: MICCAI. pp. 625–636. Springer (2023)
27. Zeng, Z., Wang, J., Shepherd, T., Zwiggelaar, R.: Region-based active surface modelling and alpha matting for unsupervised tumour segmentation in pet. In: 2012 19th IEEE International Conference on Image Processing. pp. 1997–2000. IEEE (2012)
28. Zhao, H., Li, H., Cheng, L.: Improving retinal vessel segmentation with joint local loss by matting. Pattern Recognition **98**, 107068 (2020)
29. Zheng, Y., Kambhamettu, C.: Learning based digital matting. In: 2009 IEEE 12th international conference on computer vision. pp. 889–896. IEEE (2009)