Balance Accuracy and Efficiency: Segment 3D medical images with inter-slice context information guidance

Risheng Xie
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
chino1204@mail.ustc.edu.cn

Shouhong Wan✉
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
wansh@ustc.edu.cn

Peiquan Jin
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
jpq@ustc.edu.cn

Weiyi Deng
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
jnfkdsn@mail.ustc.edu.cn

***Abstract*—Medical image segmentation plays a crucial role in disease detection, diagnosis and treatment. However, with the rapid increase in three-dimensional (3D) image data size, existing medical segmentation methods struggle in balancing computational cost and model performance. This paper proposes a slice-based segmentation method on 3D images and several plug-and-play Slice Context Modules (SCMs) to enable 2D networks to perform 3D segmentation. With an autoregressive method to extract inter-slice context information, SCMs use image morphological operations to expand this information and predict the inter-slice morphological changes with previous inter-slice context information, and use a channel-wise fusion module to guide the segmentation. SCMs improve the model's ability to sense and predict target morphological changes, while balancing the accuracy and efficiency in model training process. This paper also designed a specific training method called random partitioning to further accelerate the training process. Experiments conducted on open datasets have shown that the proposed methods increased the performance of baseline networks by over 4.6% in Dice score, surpassing multiple state-of-the-art 3D segmentation networks. Furthermore, SCMs achieve 30% faster training speed and 40% less GPU memory usage during training, demonstrating their effectiveness in 3D medical segmentation tasks.**

***Index Terms*—3D Medical Image Segmentation, Plug-and-play module, Autoregressive method, Morphology Prediction.**

I. Introduction

Recently, automatic medical image segmentation methods have been widely researched due to their key importance in the diagnosis of numerous diseases and the difficulties encountered in real-world clinical practice [1]–[3], which include complex noise interference, low foreground-background contrast, and a lack of annotated data. To address these problems, numerous medical segmentation methods [4]–[6] have been proposed.

However, as the volume of data accumulated in practical applications continues to grow, the high computational cost of these methods limits their practical deployment. Since the size of 3D images increases significantly, transformer-based methods [7]–[10] face severe problems due to their quadratic computation complexity.

In this paper, we propose SCMs, an autoregressive method that resolves the efficiency-accuracy dilemma in 3D segmentation by leveraging inter-slice context information as guidance while maintaining linear computation complexity.

The primary contributions of this paper are as follows:

1) This paper proposes several plug-and-play inter-slice modules, termed SCMs, which enable 2D segmentation frameworks to segment 3D medical images with a negligible additional computational cost (less than 1% parameters and at most 4 GFLOPs).

2) To verify the adaptability of SCMs across different network architectures, we integrate SCMs into multiple 2D segmentation networks [11]–[13], resulting in SCNet, SCAU, and SCUNext.

3) We compare our method with multiple state-of-the-art 3D segmentation methods on open medical image segmentation datasets. The results demonstrate that the proposed method achieves better segmentation performance while using fewer computational resources.

All training and evaluation code are now available at this repository.

II. Related Works

A. 3D Medical Image Segmentation

3D segmentation methods have been developed to automatically analyze medical images and facilitate the delineation of tumors and organs from volumetric data, such as CT and MRI scans. For example, Liu et al. [14] employ cascaded densely connected encoder blocks to improve segmentation

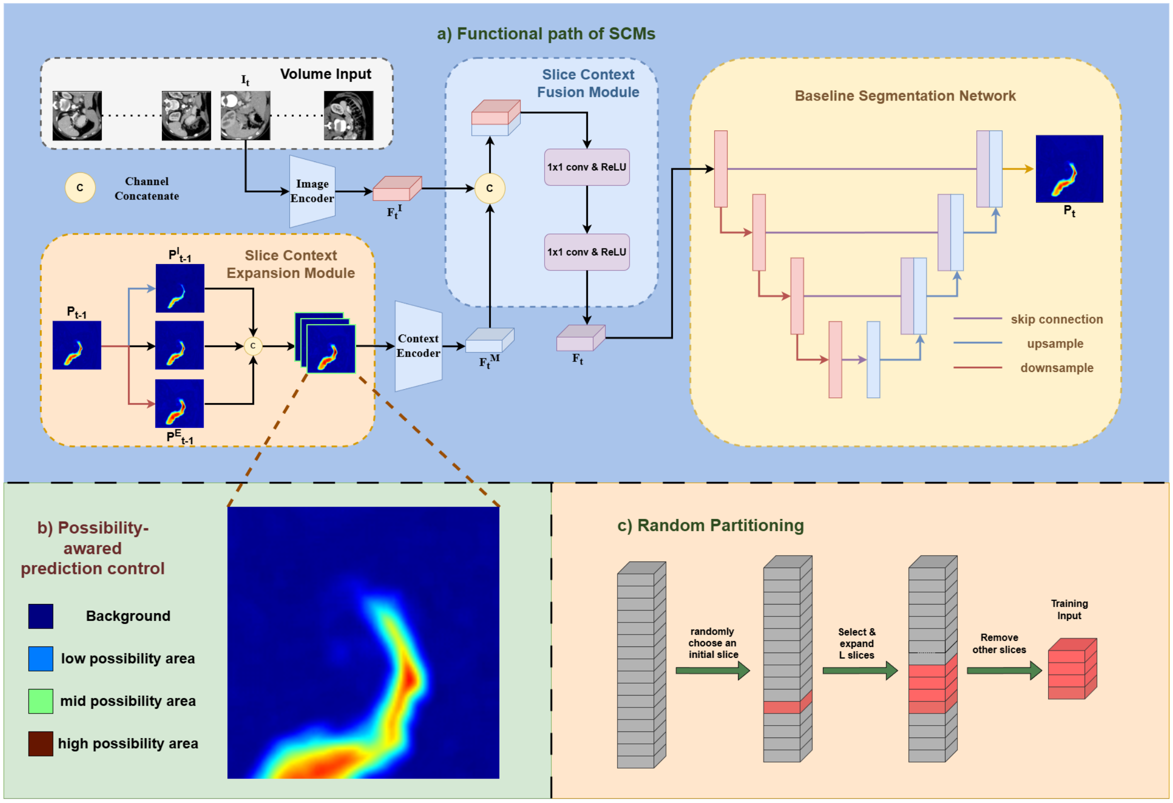

Fig. 1. a) Structure of SCMs, which receive image I_t and autoregressive inter-slice prediction P_{t-1} as input, and P_{t-1} is the unbinarized mask prediction of the last slice. When $t = 0$, P_{-1} is defined to $\mathbf{0}$. b) Possibility-awared prediction control, aiming to decrease the inter-slice error accumulation. c) The random partitioning strategy employed during training process, making model more concentrate on learning short-range structural features while accelerating the training process.

stability, while Xu et al. [15] introduce an adaptive aggregation module to capture structural characteristics across different lesion shapes and sizes. Additionally, Chowdary et al. [16] incorporate a generative network for data augmentation and train a transformer-based model for segmentation. Although these methods produce promising results, the increasing clinical demand for precise 3D image segmentationparticularly for CT and MRI scansreveals certain limitations. In particular, the efficient handling of 3D volumetric data remains underexplored in existing work, which motivates further investigation into dedicated, computationally efficient 3D segmentation frameworks.

For the 3D medical image segmentation task, due to the high dimensionality and heterogeneous structures, the primary challenge lies in effectively extracting features from the images. Some works attempt to expand convolutional kernels from 2D to 3D, thereby enabling networks to acquire 3D extraction capabilities. For instance, Jia et al. [17] introduce 3D convolutional kernels and Transformer blocks into the network architecture to directly segment 3D images, while Li et al. [18] employ both 2D and 3D convolutional kernels for improved feature extraction. Other approaches seek to incorporate Transformers into network structures by flattening input images and/or features. For example, Yu et al. [19] utilize specifically designed Transformer blocks to reprocess segmentation results. However, these efforts to directly expand kernel dimensions lead to a significantly increased computational cost.

Recently, Wu et al. [20] proposed a slice-awared segmentation method, with low computation cost. However, its segmentation performance is limited due to the lack of inter-slice interaction, and the 3D morphological continuity is not fully utilized in the reconstruction module, So SCMs try to increase the 3D morphological continuity by enhancing the inter-slice context information guidance with autoregressive method, i.e. segmenting images based on the previous segmentation result. In this paper, we will use it to predict the inter-slice morphological changes of organs, including shapes and sizes, and balance the segmentation accuracy and efficiency.

B. Autoregressive Methods

Autoregressive (AR) methods have been widely used in multiple tasks, particularly in recent generative AI and diffusion models [21]–[23]. Their compelling results indicate the efficiency of autoregression in short-term data prediction. Song et al. [24] employ an autoregressive method in their work to predict pixel values between slices, achieving lossless 3D medical image compression. Fabio et al. [25] demonstrate that autoregressive methods are useful for predicting pixel-wise covariances. These studies have explored the use of autoregressive methods as powerful predictors in medical images but have also revealed a critical issue: when the model produces an incorrect prediction early in training, that error can propagate throughout the entire segmentation, leading to inter-slice error accumulation. SAM2 [26] introduced an attention-based memory mechanism that incorporates autoregressive methods into image segmentation tasks. By adding a memory encoder, a storage module, and an integration module, SAM2 successfully tracks objects between video frames and reduces manual intervention in the video segmentation process. However, its level of automation still requires improvement, as it still depends on manual guidance. SCMs address these limitations by using image morphological operations to expand the predicted feature map and applying a possibility-aware prediction control mechanism. This approach increases segmentation flexibility, ensures that inter-slice errors decrease during forward propagation, and enables the entire network to complete 3D segmentation tasks automatically without manual intervention.

III. METHODS

A. Overview

As shown in Fig.1, The proposed SCMs are consist of a slice context expansion module, a context encoder, and a slice context fusion module, which can be plugged into the network as an independent autoregressive path. To describe possible inter-layer spatial changes, we use image morphological operations to expand the slice context features. The context encoder is used to better uniform image features and slice context features, ensuring the similarities between them to reducing the burden of fusion module, and let inter-slice morphological information efficiently mentor main segmentation path. Random partitioning strategy is also deployed in the training process for better short-range structural features extraction and further acceleration.

Noticing that 3D medical segmentation has similarity with video segmentation since they all have spatial continuous in small intervals, this paper introduced an autoregression method to predict inter-slice morphological changes. But different from traditional autoregression models and SAM2 [26], Our method do not use exponential moving average(EMA) algorithm and transformer-based long range memory mechanism, but enhanced SCMs' short-range inter-slice analysis ability by adding slice context expansion module with possibility-awared control into the autoregressive path. It forced SCMs to more concentrate on short-range dependencies, meanwhile reduces probably inter-slice error accumulation.

B. Data Preprocessing

The high cost of GPU memory in 3D model training process is mainly due to the high backward propagation cost, which is directly caused by the large size of 3D images. Reducing batch size in training process can partially alleviate this high computation cost but will affect model training since it will reduce the stability of parameters especially those in batch normalization layers. What's more, the reduction of batch size leads to the decrease of training efficiency, which is unacceptable. So slice-based 3D segmentation is considered as a more efficient and visionary solution. So the training datasets are divided into 2D slices, while the testing dataset remains to judge the performance.

C. Dual Path Segmentation

As discussed in the previous subsection, slice-based segmentation method is more efficient. However, the direct application of 2D segmentation methods on slices result in sacrifice of inter-slice context, which is totally ignored. To handle this problem, SCMs add an autoregressive path besides the image forward propagation path to enhance slices interaction and improves 3D segmentation accuracy. The detailed methodology of expansion, encoding, and fusion will be discussed in the following subsections.

D. Slice Context Expansion Module

In 3D medical segmentation task, predicting morphological changes in forward propagation process is the key to increase the performance of networks. Precision is crucial because prediction errors can amplify and lead to inter-slice error accumulation.

To better predict morphological changes we use multiple image morphological operations to prefabricate predictions as an enhancement. Learnable expansion modules are not chosen out of the consideration for its extra computation costs and unstable predictions.

The slice context expansion method can be described as follows, where $\sigma(.)$ presents for the concatenation, $\psi(.)$ and $\phi(.)$ presents for dilation and corrosion operation, with a kernel size of 3:

$$P_t^{dil} = \psi(P_{t-1}), P_t^{cor} = \phi(P_{t-1}) \tag{1}$$

$$P_t^{E} = \sigma(P_{t-1}, P_t^{dil}, P_t^{cor}) \tag{2}$$

These formulas presents how SCMs derive expanded prediction P_t^E from the input P_{t-1}. As shown in Fig.1, the concatenation forms a confidential map, which borders possibly inter-slice error accumulation.

E. Context Encoder

Transformer context encoders used in other works have shown furnish refined and comprehensive features. But in SCMs, to better align the feature of images and memories and keep the low computation complexity, we choose to use a double convolutional encoder as context encoder. This simple but effective design method can help decrease the training difficulty of slice context fusion module by ensuring the feature alignment.

F. Slice Context Fusion Module

A channel-wised convolutional fusion module is used in SCMs to fuse inter-slice context features and image features. As shown in Fig.1, the fusion process can be divided into two stages: In first stage, we concatenate the image feature $\mathbf{F_t^I}$ with the inter-slice context feature $\mathbf{F_t^M}$. Then, in the second stage, SCMs use double 1x1 convolutional layers to achieve channel fusion target. This method has extra benefits as it ensures the spatial correspondence of features, and since the features are extracted with similar encoder, the following modules can directly use the output of the second stage as the fused feature, $\mathbf{F_t}$, without extra operation. The entire process can be described as follows:

$$F_t^{fus} = \sigma(F_t^I, F_t^M) \tag{3}$$

$$F_t = \kappa(F_t^{fus}) \tag{4}$$

This formula presents how the features is fused, in which the $\kappa(.)$ presents the double convolutional fusion layer, and $\sigma(.)$ presents the concatenation operation.

G. Random Partitioning

Random partitioning is a strategy which randomly crops images with a certain length L, which can force the network to be more concentrate on short-range spatial dependencies, meanwhile accelerating the training process. For each image $\mathbf{I_o} \in R^{H*W*D}$, the random partitioning algorithm will randomly select a number $n \in N$ in [0, D-L) with same possibility, and choose L consecutive slices as the final input I. The L is set to 4 in practice since a smaller L will harm the morphological continuity and higher L will slow down the training process. SCMs only use this strategy in the training process.

IV. Experiments and Results

In this section, we conduct multiple experiments including comparison experiment and ablation experiment to evaluate the resulting 3D segmentation network on two open datasets. All the experiments are conducted with a single nvidia 3090 GPU. The learning rate is set to 1e-4 with cosine annealing, and we using Adam optimizer with a weight decay of 5e-4.

A. Dataset

In the comparison experiment, two public 3D medical datasets are used to train networks and evaluate the segmentation results. The NIH segmentation dataset [27] contains 80 3D CT images, and the MSD segmentation dataset [28] contains 420 3D CT images.

B. Metrics

To evaluate the segment accuracy of networks, we choose four metrics: Dice Score(DSC), Intersection Over Union (IOU), Recall(Rec), Precision(Pr), and Hausdorff Distance(HD). To calculate them, we need True Positive(TP), False Positive(FP), False Negative(FN). Basing on these basic metrics, we can conduct four metrics as follows:

$$Dice = \frac{2|y \cap y'|}{|y| + |y'|} = \frac{2TP}{2TP + FP + FN} \tag{5}$$

$$IoU = \frac{|y' \cap y|}{|y' \cup y|} = \frac{TP}{TP + FP + FN} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

And the HD metric is defined and calculated as follow:

$$HD(A, B) = \max(\max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(a, b)) \tag{9}$$

For evaluating efficiency, we measure the amount of model parameters and computation cost e.g. Flops, training latency(TL) and GPU memory footprint.

C. Comparison Experiments

We applied our methodologies upon several 2D SOTA baselines, including UNet [11], AttUNet [13], and UNeXt [12], and proposed SCNet, SCAU, and SCNeXt. These methods are compared with other 3D segmentation methods and their baseline. Experimental results about segmentation accuracy are conducted in Table I and Table II.

TABLE I
Comparison of Segmentation Result on NIH Dataset

	DSC	IOU	Recall	Precision	HD
UNet [11]	76.20	64.10	77.25	80.25	18.12
AttUNet [13]	75.91	63.99	75.10	82.09	17.85
UNeXt [12]	76.67	64.15	78.99	79.09	15.91
TrNet [29]	82.42	71.43	82.22	85.44	13.25
UNetr [30]	74.37	60.05	40.27	75.73	17.09
MenUNet [31]	85.90	76.27	**90.72**	81.90	12.35
Our methods					
SCUNeXt	81.20	69.51	87.03	76.72	5.641
SCNet	**88.23**	**79.45**	87.75	89.26	**4.032**
SCAU	86.02	76.51	83.61	**89.83**	5.687

In NIH comparison experiment, the SCMs methods are better than other works in all aspect, only with a single exception which is a around 3% inferior in recall metric when compared to MenUNet [31]. But on the other hand, SCMs performs 7% better than it on precision metrics, and on more crucial dice score metrics, the SCMs also perform better. Taking all these factors into account, the exception only shows SCMs tend to give a more conservative segmentation result, which is actually a signal of the reliability of the segmentation results. As different from NIH dataset, images in MSD dataset contains more cancer affected organs, the morphological changes in MSD are more frequent and intense, and the result of comparison experiment on MSD dataset also shows the effectiveness of SCMs in more complex situation. In total, SCMs methods get good performance, which indicates the effectiveness of our approach.

We also compared the parameter amount and computation cost of all methods. Comparison results are shown in the table III.

Combining results from previous tables, it shows that only with about 0.2% extra parameters and around 15% extra Flops,

TABLE II
COMPARISON OF SEGMENTATION RESULT ON MSD DATASET

	DSC	**IOU**	**Recall**	**Precision**	**HD**
UNeXt [12]	72.00	59.34	76.90	73.18	18.46
UNet [11]	69.99	57.66	69.55	77.27	23.49
AttUNet [13]	73.67	61.66	73.88	79.11	17.59
TrNet [29]	66.71	54.30	64.31	75.36	24.59
SwinUNetr [7]	79.06	65.41	82.43	76.40	19.91
MenUNet [31]	78.40	64.88	80.58	78.79	33.84
Our methods					
SCUNeXt	76.63	63.45	77.02	78.99	32.42
SCNet	79.15	66.99	76.05	**85.53**	28.82
SCAU	**81.64**	**70.11**	**84.50**	80.81	**16.09**

TABLE III
COMPARISON OF COMPUTATION COST

	MParams	**GFlops**	**Memory(M)**	**TL(s/iter)**
UNeXt [12]	1.471	0.4242	968	23.63
U-Net [11]	31.03	23.49	6220	75.33
AttUNet [13]	34.87	50.93	8184	114.48
TrNet [29]	49.80	69.84	18476	298.38
UNetr [30]	101.6	960.6	19280	38.36
SwinUNetr [7]	15.50	303.32	19938	46.83
MenUNet [31]	20.25	401.4	36480	54.03
Our methods				
SCUNeXt	1.473	0.5230	1450	9.76
SCNet	31.09	45.04	8716	13.41
SCAU	34.93	54.12	10798	15.84

SCMs-based networks get much better performance on 3D segmentation tasks. What's more, compared to existing 3D segmentation methods, although the SCMs models gets 40% fewer parameters and 15% less Flops, they still get better dice score on 3D segmentation tasks.

In summary, SCMs networks gets a significant increase on segmentation accuracy and efficiency in 3D segmentation task when compared to other SOTA 3D segmentation methods, which shows the effectiveness of this method on one aspect.

D. Ablation Study

The ablation experiment is conducted on NIH dataset, which results are shown in Table IV, in which 'Backbone' refers to the baseline model, 'Encoder' refers to the proposed context encoder, 'Fusion' represent to the slice context fusion module, and 'Expansion' denotes the proposed slice context expansion operation.

TABLE IV
THE RESULTS OF ABLATION EXPERIMENT

Backbone	Encoder	Fusion	Expansion	DSC(%)	IOU(%)
✓				76.20	64.10
✓	✓			83.94	73.60
✓	✓	✓		85.16	75.15
✓	✓		✓	85.03	75.04
✓	✓	✓	✓	**88.23**	**79.45**

By adding autoregressive path and context encoder into the network structure, the preliminary network get a peak performance increase, including an advantage of 7.74% DSC and 9.50% IOU. The slice context fusion module and slice context expansion module both gives around 1% DSC advantage and around 1.4% IOU advantage when applied to the network separately, And there is a cooperative effect of the expansion and fusion module: when they are applied to the network together gives 4.29% advantage in DSC and 6.85% advantage in IOU in total. Finally, by adopting all the proposed methods, SCMs get 12.03% more dice score and 15.35% more IOU than the baseline in total, which firmly supports the effectiveness of these proposed methods in this paper.

E. Visible Results

This subsection shows the segmentation results of the proposed method and other sota methods to provide a directly view to the segmentation results of SCMs methods. These segmentation results are shown in the Fig. 2 and Fig. 3.

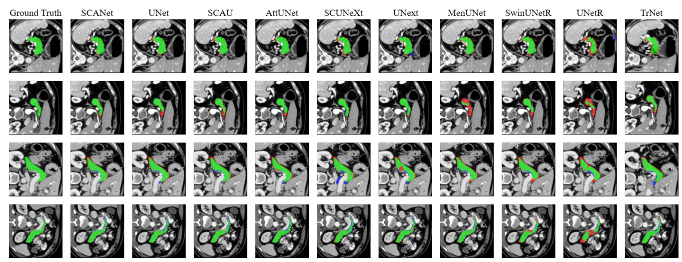

Fig. 2. Sliced segmentation results, in which blue refers to the false prediction, red refers to the false negative and green refers to the true prediction

Fig. 2 shows the slice segmentation results of all methods concluded in this paper. The upper three lines are the results from MSD dataset, and the result of the bottom line is from NIH dataset. Compared to other methods, the SCMs network has fewer missed inspections and better morphology similarity with ground truth. Some methods detected extra objects, which is a signal of lacking spatial information guidance. It shows the spatial guidance effect of autoregressive path in SCMs. The result of SCMs-based networks also has fewer fracture and get smoother edges, which shows the morphological effect of expansion.

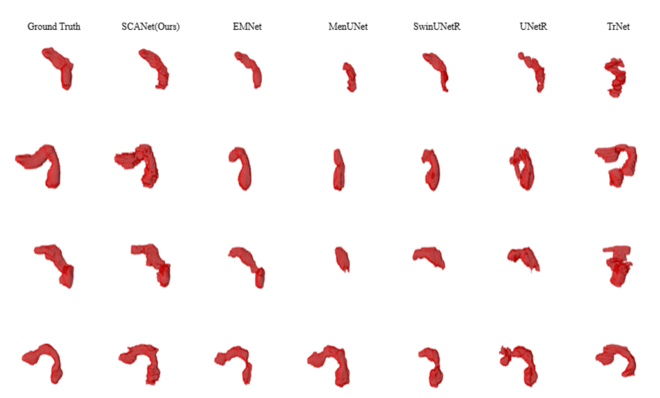

Fig. 3. 3D Segmentation results, showing the segmentation mophological accuracy of SCMs method

Fig. 3 shows the 3D segmentation results comparison of works, providing another aspect of evaluating the segmentation results. All results are selected from the experiment on MSD dataset, since tumors brings more interruption to the segmentation task. It is clear that compares to other 3D and 2.5D works, SCMs maintains better morphological character similarities with ground truth, which also supported the effectiveness of all the methods proposed in this paper.

Generally speaking, all visible results shows that SCMs works have better morphological accuracy, lower false prediction ratio and false negative ratio in 3D medical segmentation tasks, which shows the effectiveness of SCMs.

V. Conclusion

This paper focuses on the high computational cost problem caused by the rapid increase in data size for 3D medical image segmentation tasks and proposes the plug-and-play modules, termed SCMs, to help balance accuracy and efficiency. By employing an autoregressive slice-based segmentation method that introduces inter-slice context information and predicts morphological changes of foreground objects along the segmentation forward propagation path, SCMs significantly improve the performance of 2D baseline networks on 3D segmentation tasks. Experiments conducted on multiple 3D medical segmentation datasets show that SCMs increase the performance of baseline networks by over 4.6% in Dice score, surpassing many state-of-the-art 3D segmentation networks, thereby demonstrating the accuracy of SCMs-based networks. Furthermore, SCMs-based networks exhibit superior efficiency. Compared to existing 3D segmentation methods, SCMs-based networks train over 30% faster with a GPU memory footprint that is over 40% smaller during the training process.

References

[1] H. E. Sarolu, I. Shayea, B. Saoud, M. H. Azmi, A. A. El-Saleh, S. A. Saad, and M. Alnakhli, "Machine learning, iot and 5g technologies for breast cancer studies: A review," *Alex. Eng. J.*, vol. 89, pp. 210–223, 2024.

[2] P. Gurunathan, P. S. Srinivasan, and R. S, "Deep transfer learning based feature fusion model with bonobo optimization algorithm for enhanced brain tumor segmentation and classification through biomedical imaging," *Sci Rep*, vol. 15, no. 1, Sept. 2025.

[3] Z. Wang, Y. Wang, and Y. Wang, "The application of deep learning-based pulmonary image segmentation in disease screening and prevention," *Am. J. Prev. Med.*, vol. 69, p. 107871, Aug. 2025.

[4] J. Bae, X. Guo, H. Yerebakan, Y. Shinagawa, and S. Farhand, "Samu: An efficient and promptable foundation model for medical image segmentation," in *MedAGI*, 2025, pp. 134–142.

[5] A. Chang, J. Zeng, R. Huang, and D. Ni, "Em-net: Efficient channel and frequency learning with mamba for 3d medical image segmentation," in *MICCAI*, 2024, pp. 266–275.

[6] F. Song, Z. Cheng, L. Li, and H. Zhang, "Mtpnet: Learning multiple twin-support prototypes for few-shot medical image segmentation," in *BIBM*, 2024, pp. 2414–2421.

[7] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *BRAINLES*, 2022, pp. 272–284.

[8] B. Yu, Q. Zhou, L. Yuan, H. Liang, P. Shcherbakov, and X. Zhang, "3d medical image segmentation using the serialparallel convolutional neural network and transformer based on cross-window self-attention," *CAAI T. Intell. Technol.*, vol. 10, no. 2, pp. 337–348, Apr. 2025.

[9] M. Li, J. Ma, and J. Zhao, "Lkda-net: Hierarchical transformer with large kernel depthwise convolution attention for 3d medical image segmentation," *PLoS One*, vol. 20, no. 8, Aug. 2025.

[10] Y. Xiang, Q. He, T. Xu, R. Hao, J. Hu, and H. Zhang, "Adaptive transformer attention and multi-scale fusion for spine 3d segmentation," arXiv:2503.12853[cs.CV], 2025.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[12] J. M. J. Valanarasu and V. M. Patel, "Unext: Mlp-based rapid medical image segmentation network," in *MICCAI*, 2022, pp. 23–33.

[13] O. Oktay *et al.*, "Attention u-net: Learning where to look for the pancreas," arXiv:1804.03999[cs.CV], 2018.

[14] L. Liu, J. Qiu, K. Wang, Q. Zhan, J. Jiang, Z. Han, J. Qiu, T. Wu, J. Xu, and Z. Zeng, "Eu-net: Efficient dense skip-connected autoencoder for medical image segmentation," *IEEE Access*, vol. 11, pp. 135 959–135 967, 2023.

[15] Q. Xu, X. Luo, C. Huang, C. Liu, J. Wen, J. Wang, and Y. Xu, "Hacdr-net: Heterogeneous-aware convolutional network for diabetic retinopathy multi-lesion segmentation," in *AAAI*, vol. 38(6), 2024, pp. 6342–6350.

[16] G. J. Chowdary and Z. Yin, "Diffusion transformer u-net for medical image segmentation," in *MICCAI*, 2023, pp. 622–631.

[17] Y. Jia, G. Duan, and Y. Sheng, "Sctransnet: 3d medical image segmentation model based on the fusion of cnn and transformer," in *BIBM*, 2023, pp. 1158–1165.

[18] J. Li, S. Chen, S. Ma, F. Guo, and J. Tang, "Mixunet: Mix the 2d and 3d models for robust medical image segmentation," in *BIBM*, 2023, pp. 1242–1247.

[19] B. Yu, Q. Zhou, and X. Zhang, "Spctnet: A series-parallel cnn and transformer network for 3d medical image segmentation," in *CICAI*, 2024, pp. 376–387.

[20] Y. Wu, B. Zheng, J. Chen, D. Z. Chen, and J. Wu, "Self-learning and one-shot learning based single-slice annotation for 3d medical image segmentation," in *MICCAI*, 2022, pp. 244–254.

[21] R. Kumar, M. Prakash, and B. Shakila, "A comparative analysis of time series and machine learning models for wind speed prediction," in *MysuruCon*, 2023, pp. 1–6.

[22] A. Amer, S. Ahmed, and F. Kopsaftopoulos, "Probabilistic state estimation under varying loading states via the integration of time-varying autoregressive and gaussian process models," *Struct. Health Monit.*, vol. 23, no. 6, pp. 3545–3580, Nov. 2024.

[23] J. Gao, Q. Cao, and Y. Chen, "Auto-regressive moving diffusion models for time series forecasting," in *AAAI*, vol. 39, no. 16, 2025, pp. 16 727–16 735.

[24] C. Song *et al.*, "Lvpnet: A latent-variable-based prediction-driven end-to-end framework for lossless compression of medical images," in *MICCAI*, 2026, pp. 288–298.

[25] F. D. S. Ribeiro, O. Todd, C. Jones, A. Kori, R. Mehta, and B. Glocker, "Flow stochastic segmentation networks," arXiv:2507.18838[cs.CV], 2025.

[26] N. Ravi *et al.*, "Sam 2: Segment anything in images and videos," arXiv:2408.00714[cs.CV], 2024.

[27] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," arXiv:1506.06448[cs.CV], 2015.

[28] M. Antonelli *et al.*, "The medical segmentation decathlon," *Nat. Commun.*, vol. 13, no. 1, p. 4128, Jul. 2022.

[29] F. Xia, Y. Peng, J. Wang, and X. Chen, "Mtr-net:a multipath fusion network based on 2.5d for medical image segmentation," in *BIBM*, 2023, pp. 2896–2903.

[30] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," in *WACV*, 2022, pp. 1748–1758.

[31] H. Lin, X. Cheng, and Z. Chen, "Menunet: An end-to-end 3d neural network for meningioma segmentation from multiparametric mri," in *BRATS*, 2024, pp. 291–299.