Sensors and Materials, Vol. 36, No. 6 (2024) 2481–2493
MYU Tokyo

S & M 3681

Delineation of Clinical Target Volume of Esophageal Cancer Based on 3D Dense Network with Embedded Capsule Modules

Yong Huang,[1] Feixiang Zhang,[2] Kai Xu,[2*] and Chengcheng Fan[3,4,5**]

[1]Department of Medical Oncology, The Second People's Hospital of Hefei, Hefei 230011, China
[2]School of Internet, Anhui University, Hefei 230039, China
[3]Innovation Academy for Microsatellites of CAS, Shanghai 201210, China
[4]Shanghai Engineering Center for Microsatellites, Shanghai 201210, China
[5]Key Laboratory of Satellite Digital Technology, Shanghai 201210, China

(Received July 7, 2023; accepted January 16, 2024)

Keywords: deep learning, esophageal cancer, medical image processing, radiation therapy, target delineation

In this study, we propose a 3D dense network with embedded capsule modules (3D-DUCaps) for automatically delineating the clinical target volume of esophageal cancer, addressing the spatial dependence issue between parts and the whole that cannot be effectively captured by 2D networks. The network integrates capsule modules into the encoding layers of the U-Net to enhance feature learning capabilities and preserve more information, enabling the inference of poses and learning the relationship between parts and the whole. Additionally, dense connections are introduced to further promote the fusion of high-level semantic information and low-level feature information, enhancing the network's information propagation capabilities. Compared with traditional 2D deep learning networks, the proposed 3D deep learning network demonstrates stronger spatial awareness and superior boundary delineation capabilities, resulting in the more precise delineation of the clinical target volume of esophageal cancer. Experimental results indicate that the 3D-DUCaps network achieves a 2.4% improvement in the Dice Similarity Coefficient metric compared with the classical 3D-UNet network.

1. Introduction

Esophageal cancer, which originates from esophageal epithelial cells, encompasses various types, including squamous cell carcinoma, adenocarcinoma, and adenosquamous carcinoma. Globally, it exhibits high incidence and mortality rates, particularly in East Asia.[1,2] Treatment options include surgery, radiation therapy, and chemotherapy, with radiation therapy being a highly effective method that utilizes high-energy radiation to eradicate cancer cells, alleviate symptoms, and extend patient lifespans. Accurate localization in radiation therapy significantly affects treatment outcomes.[3] Hence, the precise delineation of tumor boundaries, ensuring proper coverage while minimizing harm to normal tissues, is essential. Traditionally, radiation oncologists have manually delineated tumors on medical images such as computed tomography

*Corresponding author: e-mail: kaixu@ahu.edu.cn
**Corresponding author: e-mail: fancc@microsate.com
https://doi.org/10.18494/SAM4573

ISSN 0914-4935
https://myukk.org/

(CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) images.[4–6] However, this method is subjective and prone to inter-observer variability.[7,8] To address this issue, the automated delineation of tumor targets has become a hot topic in the field of medical image processing.

Recent advancements have spotlighted deep learning, which offers automated feature extraction and learning from medical images, particularly in classification and segmentation.[9,10] The utilization of deep learning for esophageal cancer target volume delineation increases accuracy and efficiency while reducing subjectivity and variability,[11] providing a robust foundation for radiation treatment planning. Deep-learning-based convolutional neural network (CNN) models are widely applied in medical image segmentation. For instance, Hasan *et al.*[12] employed depthwise separable convolution for the pixel-level segmentation of skin lesions. Yan *et al.*[13] introduced a two-stage model with a multilevel detection strategy for fine-scale segmentation. Tulsani *et al.*[14] proposed cup and disc-based segmentation for glaucoma identification. Skouta *et al.*[15] improved the CNN U-Net for retinal hemorrhage identification, incorporating a novel loss function. Zhou *et al.*[16] presented ERV-Net, a 3D CNN for efficient brain tumor segmentation with lower GPU memory usage and computational complexity.

Esophageal cancer image segmentation using deep learning represents a breakthrough by automatically classifying and segmenting regions of interest and clinical target volumes (CTVs). This approach enhances information extraction, which may be challenging for human detection, and accelerates cancer diagnosis, thereby saving doctors' time. Huang *et al.*[17] introduced the Channel-Attention U-Net for esophagus and CTV segmentation, combining channel attention and cross-level feature fusion modules. Cai and Wang[18] proposed a multiscale, attention-based model to reduce semantic ambiguity. Han *et al.*[19] proposed the ConvUNeXt model, which was constructed using standard ConvNet modules and demonstrated competitive performance in various image applications.

In this study, we propose a 3D dense network embedded with capsule modules for the automatic delineation of CTVs in esophageal cancer. The model is specifically designed and optimized to improve the accuracy, efficiency, and practicality of delineation. In comparison with traditional 2D deep networks, we offer the following contributions:

1. Enhanced spatial perception of the model: Unlike 2D models that can only handle planar images, the 3D model can simultaneously process multiple slices of images from different directions, more precisely capturing the spatial morphology of tumors.
2. Improved characterization of boundary information: With the ability to process multiple slice images simultaneously, the 3D model can more precisely capture the information regarding tumor boundaries, facilitating the accurate localization and segmentation of the tumor target region.
3. Enhanced robustness and generalization of the model: The 3D model can learn features from a larger data space, enabling the more precise handling of diverse imaging qualities and tumor morphologies, thereby improving the model's robustness and generalization capability.
4. Reduced dependence on manual preprocessing: As the 3D model can directly process raw 3D image data, it reduces reliance on manual preprocessing, thereby reducing potential errors and uncertainties introduced during the preprocessing stage.

Overall, this study's approach addresses the challenges of automatic delineation in esophageal cancer CTVs by leveraging a 3D dense network with embedded capsule modules. It enhances the model's spatial perception, boundary characterization, robustness, and generalization, while reducing dependence on manual preprocessing, contributing to a more accurate and efficient delineation in clinical practice.

2. Materials and Methods

2.1 Datasets

Stage I and stage II esophageal cancers are usually diagnosed at an earlier stage and commonly treated with radical esophageal cancer surgery. In contrast, stage III esophageal cancer requires different surgical approaches, often involving extensive resection. As a result, the relative positions among anatomical structures after surgery become complex and variable. This variation adds to the complexity of managing stage III esophageal cancer and its surgical outcomes. This complexity can potentially introduce interference during model training. Therefore, we collected data from 91 patients diagnosed with stage I and stage II esophageal cancers at the first author's institution between January 2015 and December 2019. We used preoperative CT because contouring included the primary tumor.

During the treatment of esophageal cancer, it is necessary for doctors to localize the radiation therapy within a certain range around the esophageal tumor. In this study, the delineation of the CTV was defined as follows: The upper boundary of the CTV extended to the inferior margin of the cricoid cartilage, while the lower boundary encompassed an area extending 3 cm below the tracheal carina. This CTV included the tumor bed within the esophagus, the anastomotic stoma, and the lymph nodes located in regions 2, 4, 5, and 7 of the chest. This ensures that the radiation therapy covers the entire lesion area of the esophageal cancer, minimizing the risk of missed diagnosis or mistreatment. However, owing to variations in tumor location among patients, the lower boundary of the CTV may need to be further adjusted downward. Patients within this CTV range were positioned in the supine position using a vacuum cushion and a thermoplastic mask to immobilize the neck and shoulders. CT data was acquired using the Siemens Healthineers Somatom Definition AS 40-slice CT system or the Philips Healthcare Brilliance CT Big Bore system, and the acquisition sequence employed was the plain CT scan protocol. The CT images were reconstructed with a matrix size of 512 × 512 and a thickness of 2.5 mm. Radiation oncologists used the Pinnacle treatment planning system from Philips Radiation Oncology Systems to delineate the contours of the CTV on the planning CT scans, which served as the ground truth (GT). The contours of each CTV were drawn by experienced oncologists and further reviewed and verified by senior radiation oncologists. The CT data exported from the treatment planning system, along with the radiation structures of the patients, were used to extract and label all voxels belonging to the GT segmentation of the CTV. Only CT images containing the CTV were included as training and testing data. In this study, the CTV portion of the patients' CT images and radiation structures were labeled as class 1, while the remaining background was labeled as class 0.

In this section, we will describe the process of handling the raw images of esophageal cancer, which are typically stored in DICOM format. The DICOM format has been widely used in radiology and medical devices. However, this format is not suitable for deep learning tasks, so it must be converted into the .NII format that can be used by 3D deep learning networks and then organized into a dataset. Specifically, in this study, the DICOM files' metadata is first read, and then the images are preprocessed to obtain standard Hounsfield unit values. To remove irrelevant details, all CT image intensity values are truncated to the range of [−150, 200] Hounsfield units. For each CT slice and annotation image, an adaptive cropping method is applied to crop them to a matrix size of 256 × 256. Next, the CTV region in the annotation image is read, and the upper and lower boundaries of the CTV are determined and appropriately extended. Finally, the entire dataset is saved separately as image data and label data in .NII format, completing the dataset creation process.

After the aforementioned data preprocessing steps, in this section, we focus on data analysis. The label information is utilized to perform statistical analyses on tumor size, location, and volume, as shown in Fig. 1. From the figure, it can be observed that the CTVs of esophageal cancer are mostly located in the central region of the images and exhibit irregular shapes. The slices are closely connected to each other. On the basis of this information, the dataset can undergo center-adaptive cropping, as illustrated in Fig. 2.

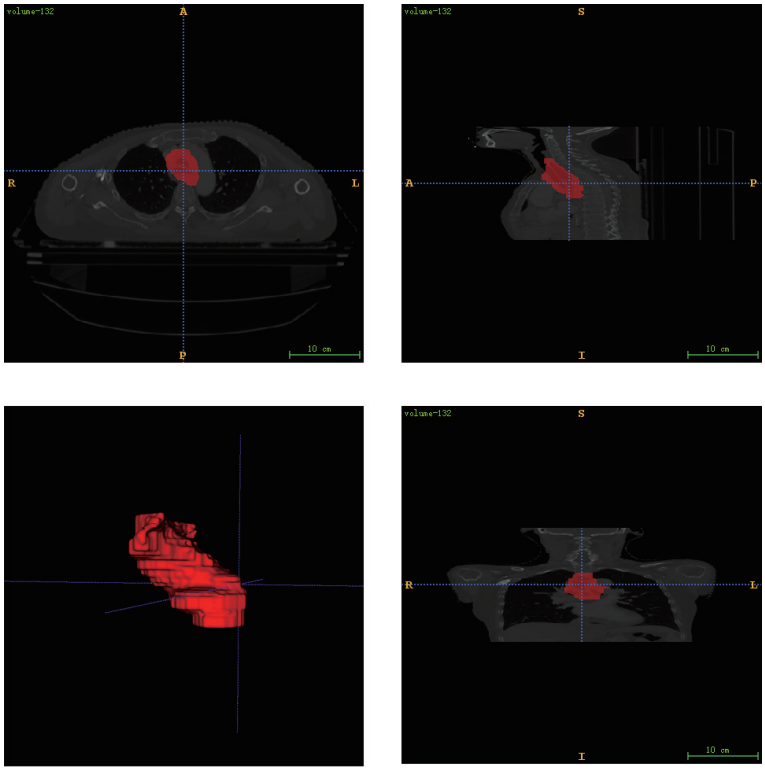


Fig. 1. (Color online) Presentation of 3D esophageal tumor CT image data.

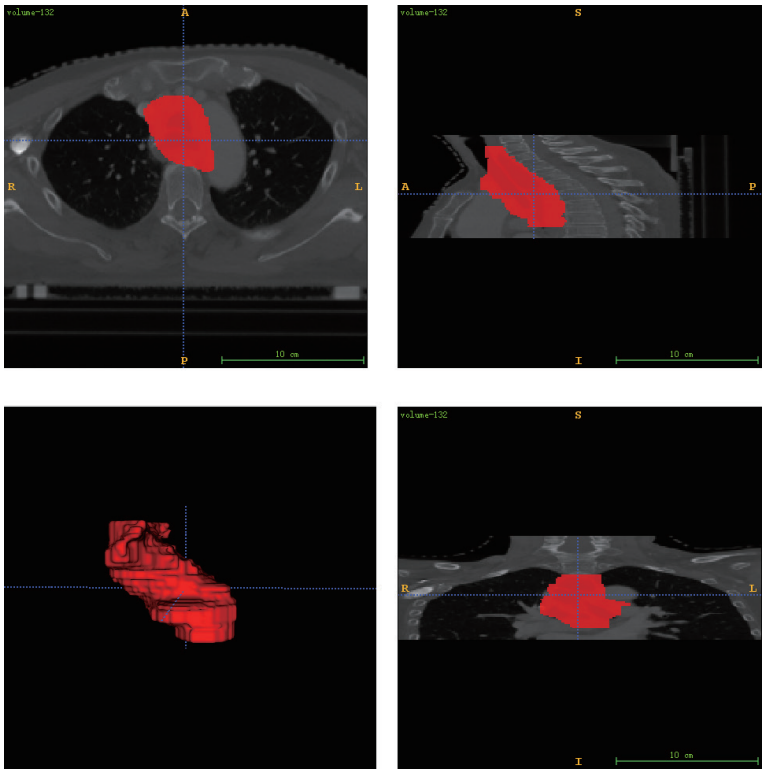


Fig. 2. (Color online) Visualization of adaptive cropping results of 3D esophageal tumor CT images.

2.2 Proposed 3D-DUCaps network architecture

In this paper, we propose a 3D dense network with embedded capsule modules. The capsule modules consider the relationship between position and orientation, while the dense connections focus on extracting contextual visual representations. By incorporating the capsule network and dense connections into the convolutional neural network, the encoding layers consist of a combination of deep dense networks and 3D capsule networks, while the decoding blocks consist of traditional convolutional neural networks. The network is named 3D-DUCaps, as shown in Fig. 3. Therefore, 3D-DUCaps inherits the advantages of preserving spatial relationships from the capsule network and learning visual representations from the convolutional neural network. In a convolutional neural network, each filter in a convolutional layer works as a feature detector in a small region of the input feature. As the network goes deeper, the detected low-level features are aggregated and combined into high-level features that can be used for discrimination. However, this approach results in each feature map containing information about the presence of a specific feature only, and the network relies on fixed weight matrices to connect features between layers. This leads to the model's inability to generalize well to variations in input images and often performs poorly in such cases. To address this issue,

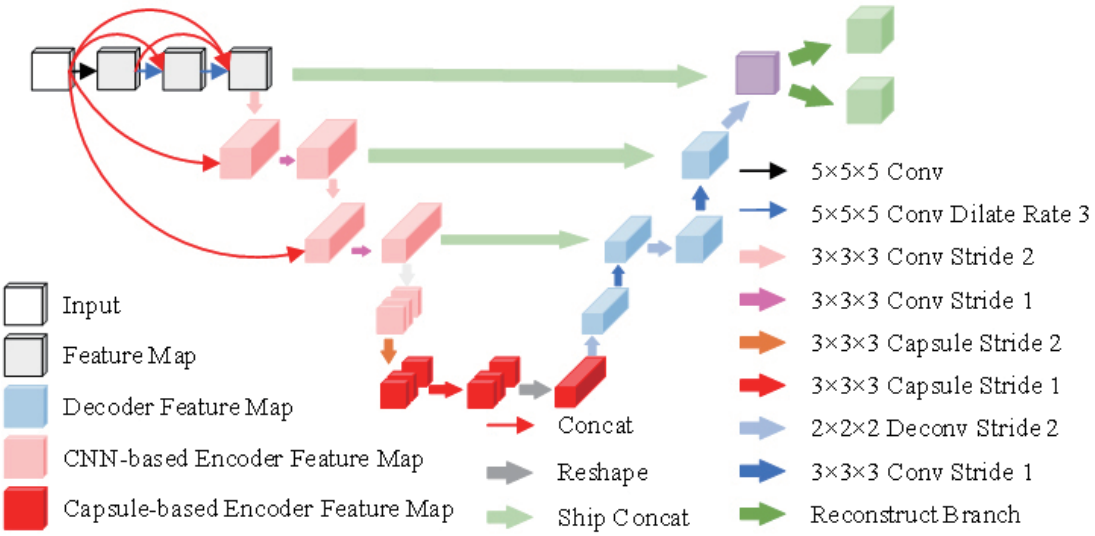


Fig. 3. (Color online) 3D-DUCaps network architecture.

shortcut connections are created between the preceding and succeeding layers to pass information, effectively mitigating the problem of vanishing gradients during the back-propagation process of deep learning networks. By doing so, the weights closer to the input layers can receive updated gradients, which improves the training effectiveness of the entire network.

The 3D-DUCaps network in this study inherits the advantages of both the capsule network and the convolutional neural network. In the encoding layers, dense connections are utilized to effectively alleviate the risk of vanishing gradients. This means that each layer can directly utilize the inputs and loss gradients from all previous layers through the back-propagation algorithm, allowing for effective weight updates. This approach enhances the training effectiveness of the neural network, enabling it to more efficiently learn feature information from the data and train deeper networks, serving as an implicit form of deep supervision. Simultaneously, in the encoding layers, the capsule network is employed to enhance feature learning and preserve more information for inferring poses and capturing relationships between parts and the whole. In the decoding layers, we still employ deep neural networks combined with skip connections to recover feature maps. By combining these advantages, the 3D-DUCaps network improves the feature learning and information propagation capabilities of the network, leading to excellent performance.

2.3 Loss function

We utilized cross entropy (*CE*) loss to train the network, which is represented in Eq. (1). In the equation, c represents the number of classes, p_i represents the true values, and q_i represents the predicted values.

$$CE(p:q) = -\sum_{i=1}^{c} p_i \log(q_i) \tag{1}$$

To achieve clear boundary detection, we employed *Dice loss* to train the network. The *Dice* coefficient is used to assess the similarity between two samples. For binary classification problems, the *Dice* coefficient is calculated by comparing the predicted binary mask with the true binary mask to measure their similarity. The *Dice* coefficient ranges from 0 to 1, where 1 indicates a perfect match and 0 indicates no match. It is calculated as

$$Dice = \frac{2|A \cap B|}{|A| + |B|}, \tag{2}$$

where $A \cap B$ represents the intersection of A and B, and $|A|$ and $|B|$ denote the numbers of elements in A and B, respectively. The numerator coefficient of 2 is used to account for the repeated calculation of common elements between A and B in the denominator. The *Dice loss* is derived from the *Dice* coefficient and can be expressed as

$$DiceLoss = 1 - \frac{2|A \cap B|}{|A| + |B|}. \tag{3}$$

By using the *Dice loss*, the differences between positive and negative samples can be effectively captured. Additionally, the *Dice loss* can lead to sharper boundaries, resulting in more precise image segmentation. When applying the *Dice loss* to image segmentation, it is common practice to convert the *Dice* coefficient into a loss function. This is achieved by taking the negative of the *Dice* coefficient and subtracting it from 1, yielding the value of the loss function. The purpose of this transformation is to maximize the *Dice* coefficient, which is equivalent to minimizing the loss function during model training. In summary, utilizing the *Dice loss* can help improve the clarity of boundary detection and increase the accuracy and precision of image segmentation in this study.

Although the *Dice loss* can help with rapid boundary localization, it is unstable and difficult to fit. To address the challenge of fitting the *Dice loss*, we adopt a composite loss function that combines the advantages of both the *Dice loss* and the *CE loss*. Specifically, the *CE loss* aids in more accurately fitting the target shape, whereas the *Dice loss* aids in more precisely localizing the target boundaries. By weighting and combining these two losses, improved segmentation results can be achieved, as shown in Eq. (4). In the experiments, the weights between the two loss functions are adjusted to balance their contributions to the overall loss. Through this approach, more accurate and stable segmentation results can be obtained, enhancing the performance and robustness of the network, thereby further improving the training outcomes.

$$DiceCE\ Loss = \alpha Dice\ Loss + \beta CE\ Loss \tag{4}$$

2.4 Experiments

The experiments were conducted using the Python programming language on a GeForce RTX 3090 graphics card, utilizing the PyTorch and PyTorch Lightning deep learning frameworks. To enhance the stability and convergence of the model, the image input size was set to 256 × 256, and the data was normalized to have pixel values within the range of [0, 1]. During the training process, a batch size of 8 was used, with 2000 epochs. The Adam optimizer was employed with an initial learning rate of $1e^{-4}$ and an L2 weight decay of $2e^{-6}$. Additionally, an early stopping strategy was implemented, where training would halt if there was no significant improvement in performance over 20 consecutive epochs.

2.4.1 Evaluation metrics

In this study, the Dice similarity coefficient (*DSC*) and Hausdorff distance (*HD*) were used to evaluate the similarity between automatic and manual segmentation results.

DSC is a commonly used evaluation metric that is sensitive to the filling of the region. It is calculated as

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|}, \tag{5}$$

where A represents the automatic segmentation, B represents the manual segmentation, $A \cap B$ is the intersection of the manual and automatic segmentations, and $|A|$ and $|B|$ are the sizes of the manual and automatic segmentation regions, respectively. The *DSC* value ranges between 0 and 1, where a higher value indicates higher overlap and segmentation performance.

HD is another widely used evaluation metric that is sensitive to the boundary regions of the segmentation. It is calculated as

$$HD(A,B) = \max\left(h(A,B), h(B,A)\right), \tag{6}$$

$$h(A,B) = \max\left(\min\|a-b\|\right), a \in A, b \in B. \tag{7}$$

Here, $HD(A, B)$ represents the maximum value among the minimum distances of each point in set A to set B and each point in set B to set A. A smaller *HD* value indicates higher overlap and segmentation performance between A and B.

2.4.2 Data splitting

To assess the overall performance of the model, a random selection of 23 patients was designated as the test set, while the remaining 68 patients underwent a fourfold cross-validation.

Each of the four models partitioned the 68 patients into a training set (80%) and a validation set (20%). Among these four trained models, the one exhibiting the highest validation *DSC* was selected for evaluation on the test set. This approach serves to mitigate the risk of overfitting on the validation set and ensures the credibility of the final experimental results.

3. Experimental Results and Discussion

In this experiment, the 3D-UNet was used as the baseline network, and the encoding layers of 3D-UCaps, 3D-EUCaps, and 3D-DUCaps were compared with different modules against the baseline network.

3.1 Network performance comparison

A comparison of the decoding layers of 3D-SegNet with those of the other networks was conducted in this section. Here, "Encoder" represents the encoding layers, "Decoder" represents the decoding layers, "Network Efficiency" represents the model performance, "Conv" represents the embedded convolution module, "Cap" represents the embedded capsule module, "Dense" represents the embedded dense connections, and "Model Size" represents the model size, where smaller values are preferable. To ensure fair experiments, all experiments were conducted using the same initial training parameters. After training completion, the optimal trained network was selected and applied to the test set. The automatically delineated CTVs of esophageal cancer are shown i a comparative analysis of the results was performed.

In this experiment, we employed different combinations of encoders and decoders to verify the performance of medical image segmentation for esophageal cancer CTVs. Here is a preliminary analysis of the experimental results:

1. Network performance comparison: From the table, it can be observed that the 3D-DUCaps model in this study achieved the highest score in terms of network efficiency, reaching 0.812. The scores of the other models were slightly lower but were generally good, ranging from 0.75 to 0.81.
2. Analysis of encoder and decoder combinations: The table reveals that different combinations of encoders and decoders can affect the model's performance. Specifically, the 3D-SegCaps network, where the Cap module was placed in the decoder, exhibited the lowest performance.

Table 1
Model performance comparison.

Method	Encoder			Decoder		*DSC*		Network Efficiency	
	Conv	Cap	Dense	Conv	Cap	Min	Max	Avg	Model Size (MB)
3D-UNet	✓	×	×	✓	×	0.751	0.808	0.788	69.1
3D-SegCaps	×	✓	×	×	✓	0.727	0.775	0.757	83.8
3D-UCaps	×	✓	×	✓	×	0.737	0.780	0.765	36
3D-EUCaps	✓	✓	×	✓	×	0.730	0.818	0.796	42
Our 3D-DUCaps	✓	✓	✓	✓	×	0.779	0.833	0.812	93.3

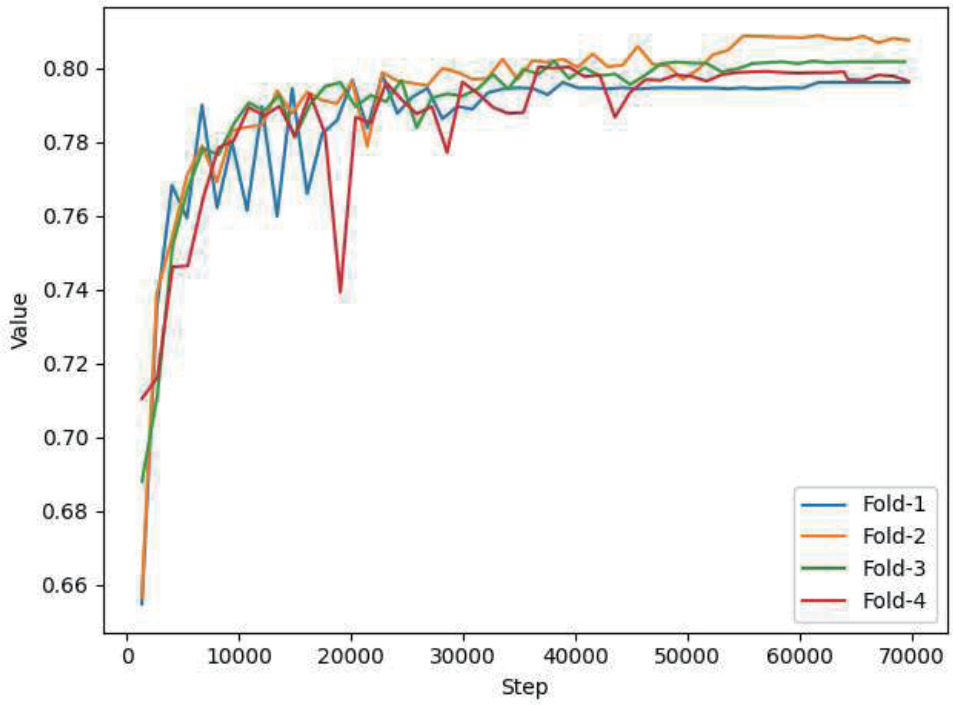


Fig. 4. (Color online) Dice curves for 3D-DUCaps.

Hence, the Cap module is not suitable for the decoding layer. In comparison with the baseline 3D-UNet network, the performance of the 3D-UCaps network, which only utilized the Cap module in the encoding layer, slightly decreased. This suggests that the Cap module is not suitable for standalone usage. This observation was further confirmed in the 3D-EUCaps network, where both the Cap and Conv modules were used in the encoding layer. Finally, the 3D-DUCaps network, which incorporated the Cap, Conv, and Dense modules in the encoding layer, achieved the best results, validating the feasibility of embedding capsule layers and dense connections in the encoding layer.

3. Model size: The size of the 3D-DUCaps model studied in this research was 93.3 MB, which was the largest among all the models. The sizes of the other models ranged from 30 MB to 90 MB. This indicates that selecting different network structures and components can impact the model size, necessitating a trade-off between model performance and computational resources.
4. Visualization of results: Figure 5 presents the automatic delineation results of the tumor target area using the 3D-DUCaps model in two esophageal cancer patients. The “Ground Truth” in red represents the target area manually annotated by radiation oncologists, serving as the “gold standard.” The green annotations indicate the predicted results of this experiment.

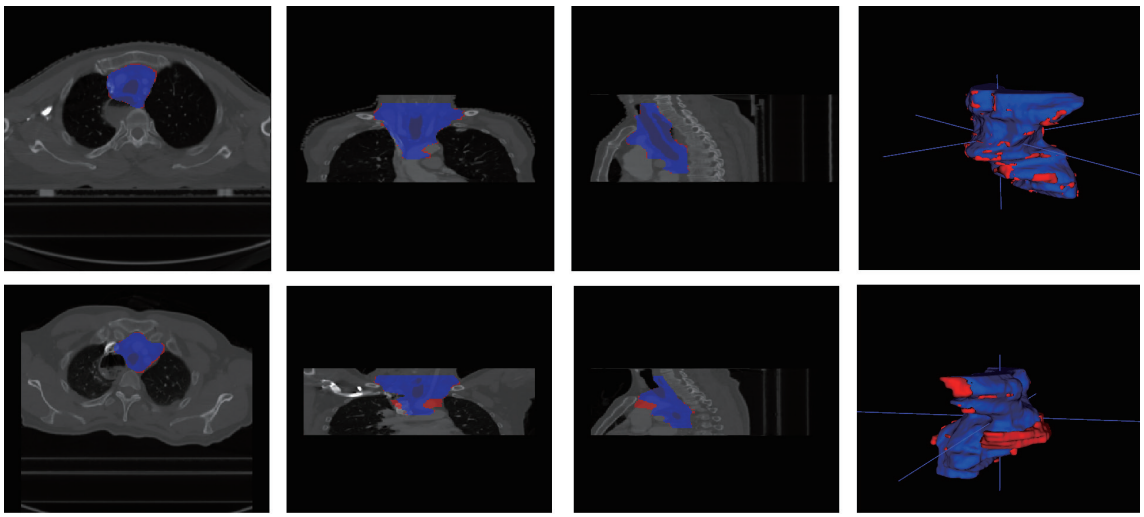
It can be observed that the automatic delineation results obtained from the network show good consistency with the manual annotations by physicians.

In summary, the 3D-DUCaps model in this research exhibited the highest performance in medical image segmentation tasks, but it had a relatively larger model size. Therefore, comprehensive considerations must be made on the basis of task requirements and computational resources.

3.2 Loss function comparison

The performance characteristics of *CE* loss, *Dice loss*, and *DiceCE loss* were tested on the basis of the 3D-EUCaps network in this experiment. The experimental results are shown in Table 2.

The performance of *DiceCE loss* is superior to those of *CE* loss and *Dice loss*. The *DSC* metric for *DiceCE loss* is 0.796, which is slightly higher than the 0.785 of *CE loss* and significantly higher than the 0.729 of *Dice loss*. Therefore, using *DiceCE loss* can improve the performance of the model for this task. In practical applications, it is recommended to train the 3D-DUCaps network using *DiceCE loss*.



(a) (b) (c) (d)

Fig. 5. (Color online) Automatic visualization of clinical target volume of esophageal cancer: descriptions of (a) transverse section, (b) sagittal plane, (c) coronal plane, and (d) 3D representation.

Table 2
Loss function comparison.

Loss	*CE*	*Dice*	*DiceCE*
DSC	0.785	0.729	0.796

4. Conclusions

In recent years, medical image analysis has gained increasing popularity in clinical medicine owing to advancements in image acquisition and storage technologies. Medical image segmentation, which accurately identifies lesion areas, is a crucial task for diagnosis and treatment planning. However, it faces challenges such as noise, low contrast, and motion artifacts. To address these issues, we proposed a novel convolutional neural network that combines traditional convolutional and capsule networks. This proposed neural network captures spatial relationships between features and prevents spatial information loss through dynamic routing. Experimental results demonstrate significant improvements in segmentation performance while maintaining reasonable complexity and computational costs. We provided a new solution and valuable insights for medical image analysis. Future research will focus on robustness analysis and exploring applications in other domains. The 3D-DUCaps network outperforms classical 3D-UNet by improving the *DSC* by 2.4% for esophageal cancer target area delineation. This architecture shows promise in enhancing medical image segmentation and can benefit tumor detection, localization, and treatment planning.

Acknowledgments

This work was supported by the Anhui Provincial Natural Science Foundation under Grant no. 2108085QD154, China Postdoctoral Science Foundation under Grant no. 2020M681993, National Natural Science Foundation of China under Grant no. 42001408, and Natural Science Foundation of Anhui Provincial Education Department under Grant no. KJ2021A0022.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal: Cancer J. Clinicians **70** (2020) 313. https://doi.org/10.3322/caac.21609
2 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray: Cancer J. Clinicians **71** (2021) 209. https://doi.org/10.3322/caac.21660
3 R. Baskar, K. A. Lee, R. Yeo, and K.-W. Yeoh: Int. J. Med. Sci. **9** (2012) 193.
4 M. F. Kircher, U. Mahmood, R. S. King, R. Weissleder, and L. Josephson: Cancer Res. **63** (2003) 8122.
5 M. AustinSeymour, G. T. Y. Chen, J. Rosenman, J. Michalski, K. Lindsley, and M. Goitein: Int. J. Radiat. Oncol. Biol. Phys. **33** (1995) 1041. https://doi.org/10.1016/0360-3016(95)00215-4
6 C. F. Njeh: J. Med. Phys. **33** (2008) 136. https://doi.org/10.4103/0971-6203.44472
7 L. V. van Dijk, L. Van den Bosch, P. Aljabar, D. Peressutti, S. Both, R. J. H. M. Steenbakkers, J. A. Langendijk, M. J. Gooding, and C. L. Brouwer: Radiother. Oncol. **142** (2020) 115. https://doi.org/10.1016/j.radonc.2019.09.022
8 M. Wilke, B. de Haan, H. Juenger, and H.-O. Karnath: Neuroimage **56** (2011) 2038. https://doi.org/10.1016/j.neuroimage.2011.04.014
9 Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I. C. Chang: IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) (2014) 1626.
10 G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan: J. Imaging Sci. Technol. **64** (2020) jist0710. https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508
11 J. Van der Veen, S. Willems, D. Robben, W. Crijns, F. Maes, and S. Nuyts: Radiother. Oncol. **133** (2019) S371. https://doi.org/10.1016/s0167-8140(19)31143-0
12 M. K. Hasan, L. Dahal, P. N. Samarakoon, F. I. Tushar, and R. Martí: Comput. Biol. Med. **120** (2020) 103738.
13 Y. Yan, P.-H. Conze, G. Quellec, M. Lamard, B. Cochener, and G. Coatrieux: Biocybern. Biomed. Eng. **41** (2021) 746. https://doi.org/10.1016/j.bbe.2021.03.005

14 A. Tulsani, P. Kumar, and S. Pathan: Biocybern. Biomed. Eng. **41** (2021) 819.
15 A. Skouta, A. Elmoufidi, S. Jai-Andaloussi, and O. Ouchetto: J. Big Data **9** (2022) 1.
16 X. Zhou, X. Li, K. Hu, Y. Zhang, Z. Chen, and X. Gao: Expert Syst. Appl. **170** (2021) 114566.
17 G. Huang, J. Zhu, J. Li, Z. Wang, L. Cheng, L. Liu, H. Li, and J. Zhou: IEEE Access **8** (2020) 122798.
18 Y. Cai and Y. Wang: Third Int. Conf. Electronics and Communication; Network and Computer Technology (ECNCT 2021) (2022) 205.
19 Z. Han, M. Jian, and G.-G. Wang: Knowledge-Based Syst. **253** (2022) 109512.

About the Authors

Yong Huang is currently the director of the Department of Medical Oncology, the Second People's Hospital of Hefei, Hefei, China. His main research interests focus on radiotherapy for solid tumors and head and neck tumors, and cancer treatment using deep learning.

Feixiang Zhang has been pursuing a bachelor's degree at Anhui University. His research interests include remote sensing geodesy, machine learning, and image target detection.

Kai Xu received his Ph.D. degree from Wuhan University, Wuhan, China in 2017. He is currently an associate professor in Anhui University. His research interests include geometric calibration and image processing using deep learning. (kaixu@ahu.edu.cn)

Chengcheng Fan received his Ph.D. degree from Wuhan University, Wuhan, China in 2017. He is currently an associate researcher in the Innovation Academy for Microsatellites of the Chinese Academy of Sciences. His research interests include remote sensing satellite design and on-orbit intelligent processing.