Cross-graph Interaction and Diffusion Probability Models for Lung Nodule Segmentation

Huaqiang Su[1], Haijun Lei[1], Chen Guoliang[1], and Baiying Lei*[2]

[1] Key Laboratory of Service Computing and Applications, Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China.

[2] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Shenzhen University Medical school, Shenzhen 518060, China. (*Email: leiby@szu.edu.cn)

Abstract. Accurate segmentation of lung nodules in computed tomography (CT) images is crucial to advance the treatment of lung cancer. Methods based on diffusion probabilistic models (DPMs) are widely used in medical image segmentation tasks. Nevertheless, conventional DPM encounters challenges when addressing medical image segmentation issues, primarily attributed to the irregular structure of lung nodules and the inherent resemblance between lung nodules and their surrounding environments. Consequently, this study introduces an innovative architecture known as the dual-branch Diff-UNet to address the challenges associated with lung nodule segmentation effectively. Specifically, the denoising UNet in this architecture interactively processes the semantic information captured by the branches of the Transformer and the convolutional neural network (CNN) through bidirectional connection units. Furthermore, the feature fusion module (FFM) helps integrate the semantic features extracted by DPM with the locally detailed features captured by the segmentation network. Simultaneously, a lightweight cross-graph interaction (CGI) module is introduced in the decoder, which uses region and edge features as graph nodes to update and propagate cross-domain features and capture the characteristics of object boundaries. Finally, the multi-scale cross module (MCM) synergizes the deep features from the DPM with the edge features from the segmentation network, augmenting the network's capability to comprehend images. The Diff-UNet has been proven effective through experiments on challenging datasets, including self-collected datasets and LUNA16.

Keywords: Diffusion probabilistic models · Lung nodule · Cross graph interaction · Feature fusion module · Multi-scale cross module.

1 Introduction

Lung cancer, one of the most common and severe cancers, can have devastating effects on human lives [1]. It is expected to be the leading cause of death in

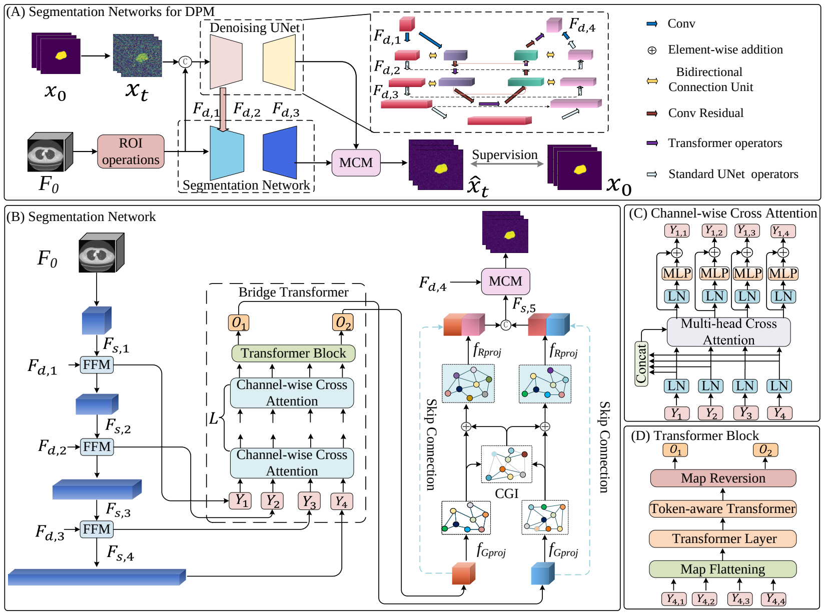

Fig. 1. Overview of the Diff-UNet proposed in this study. $F_{s,i}$ is the feature of i stage of the encoder in the segmentation network. The bidirectional connection units use convolutional layers to refine deep features and interactively connect the two branches.

Europe in 2019 [2]. This disease can be treated effectively with radiotherapy and chemotherapy. A computed tomography (CT) examination is one of the screening tools for lung nodules in clinical practice. Nevertheless, manual lung nodule detection from CT images is time-consuming and labour-intensive, radiologists meticulously examine each layer of chest CT scans to identify nodules [3, 4].

Recently, deep convolutional neural networks (CNNs) have made significant advances in natural image processing, particularly in the fields of medical image segmentation [5–8]. Researchers leverage CNNs to automatically learn high-level semantic features from images to eliminate the need for hand-crafted descriptors [9–11]. Among various CNN network architectures, UNet [2] has become a mainstream framework for segmentation tasks because it utilizes low-level and high-semantic features in the encoder into the decoder. Subsequently, the performance of the network was improved by using variant UNet (e.g., DAF3D [12], ErNet [9], SurrigateNet [10], GenerativeNet [11], CsNet [13], WingsNet [14]). However, most existing CNNs encounter several challenges. Firstly, they may not effectively capture global features by employing shared weights in convolutional layers across various spatial locations. Secondly, the networks often utilize many feature channels, potentially leading to feature redundancy [15–17, 14, 18].

The attention mechanism is anticipated to enhance the segmentation performance of CNNs by concentrating on the most pertinent information in the feature map while mitigating irrelevant components [7, 19, 20, 12]. SwinTrans [18] and TransUNet [15] use Transformer blocks to extend model depth from input to output to capture multiple levels of global feature representation. Segmentation models based on deep learning primarily concentrate on acquiring the intensity features of the input image while neglecting the intrinsic relationships between regional boundaries [21]. The structure based on graph reasoning can skillfully promote long-range information propagation [21–23]. However, the performance of the segmentation networks may be affected by noise in medical images.

The combination of diffusion probabilistic model (DPM) and CNN, as a denoising method, has been used in image segmentation tasks [24–26, 13]. This combination enables the CNN to effectively eliminate noise in the segmentation process. Wolleb et al. [5] proposed to combine the DPM with CNN to address the challenge of medical image segmentation. However, the backbone network of UNet [27] is the basis of all these methods, and the features captured by Transformer are incompatible with those of the backbone network, impeding the network's ability to capture global features.

To address these issues, we propose a DPM-based medical image segmentation network for lung nodules, called Diff-UNet. This architecture interactively processes semantic information captured by the Transformer and CNN branches through bidirectional connection units, forming a dual-branch denoising UNet. Subsequently, incorporating a feature fusion module (FFM) aims to merge the global context information obtained from the DPM with the local features acquired by the segmentation network. Using the decoder's cross-graph interaction (CGI) module facilitates multi-level node updating and feature fusion, enabling Diff-UNet to effectively capture detailed features within boundary areas. Finally, the multi-scale cross module (MCM) fully uses the diffusion model to extract shallow image details and edge information. The contributions of this paper are:

- The decoder incorporates a CGI module to capture the complex correlation between pixels and their boundary details by performing feature transfer and aggregation operations on the graph structure to achieve accurate segmentation along the boundary.
- FFM effectively extracts and integrates scale features from the segmentation network and DPM, improving Diff-UNet's ability to identify structures of various sizes, thereby improving segmentation accuracy.

2 Method

2.1 Dual-branch Denoising UNet of DPM

To enhance the ability to capture both local details and global structural information in images, we introduce a novel UNet architecture that utilizes a dual-branch UNet. As shown in Fig. 1 (A), the central part of Diff-UNet consists of a segmentation network and denoising UNet (DU).

The DU comprises a feature-interactive dual-branch network based on ResNet [28] and Transformer branches. First, given 3D volume data F_0 for the Diff-UNet, F_0 and the noisy label x_t are channel concatenated into DU's encoder to yield the multi-scale feature. The ResNet and Transformer branches interact with features at each scale through bidirectional connection units. The global context information obtained from the diffusion model is fused with the local features obtained from the segmentation network through FFM. The four input feature maps obtained from the encoder are represented as $Y_i, i \in \{1, 2, 3, 4\}$, which are input into channel-wise cross attention. After the feature is reshaped, it is converted into a tensor with uniform depth in four channels. The tensors are fed into the multi-head channel cross-attention module. This process utilizes multi-scale features to refine features at each UNet encoder level. Subsequently, we input features into the attention module, To strengthen global dependencies through multi-stage concatenation. The final output feature vector of the Transformer block is $O_i, i \in \{1, 2\}$, and then the CGI module performs multi-level node update and feature fusion on the feature vector O_i so that Diff-UNet can effectively capture the detailed features in the boundary area. Finally, the MCM is used to fuse the semantic information of the DPM and the segmentation network.

2.2 Graph Reasoning

A lightweight cross-graph interaction (CGI) module is introduced in the decoder, which uses region and edge features as graph nodes to update and propagate cross-domain features and capture features of object boundaries, thereby adapting to complex image structures. As shown in Fig.1, the decoder obtains regional depth feature extraction (O_1) and object-aware edge extraction (O_2). Specifically, it enforces the following three operations:

(1) Graph Projection: f_{Gproj} is used to transform feature vectors, O_1^l or O_2^l, into graph node embeddings, we parameterize f_{Gproj} by W. Each column w_k of W specifies a learnable anchor centre for the k-th node and forms the k-th column of the node feature matrix V.

(2) Cross-Graph Interaction (CGI):f_{CGI} The process of CGI emulates the interaction among graphs, transferring inter-graph messages from V_{O_1} to V_{O_2}, and computes inter-graph dependencies through an attention mechanism. Initially, Multi-layer perceptrons (MLPs) are utilized to transform V_{O_1} into the key graph $V^{\theta}_{O_1}$ and the value graph $V^{\gamma}_{O_1}$, while transforming V_{O_2} into the query graph $V^{k}_{O_2}$. Then, a similarity matrix is calculated using matrix multiplication $A^{inter}_{O_1 \longrightarrow O_2} \in R^{K \times K}$ as follows:

$$\begin{cases} A^{inter}_{O_1 \longrightarrow O_2} = f_{norm}(\mathcal{V}^{K^T}_{O_2} \times \mathcal{V}^{\theta}_{O_1}) \\ \mathcal{V}'_{O_2} = f_{CGI}(\mathcal{V}_{O_1}, \mathcal{V}_{O_2}) = \chi(A^{inter}_{O_1 \longrightarrow O_2} \times \mathcal{V}^{\gamma}_{O_1}) + \mathcal{V}_{O_2} \end{cases} \tag{1}$$

where the semantic information can be transmitted from $\mathcal{V}_{O_1}$ to $\mathcal{V}_{O_2}$. $\chi(\cdot)$ is used as a weighting parameter to adjust $\mathcal{V}'_{O_2}$ in CGI.

(3) Graph Reasoning:f_{GR} **and Graph Reprojection:**f_{Rproj} We use $\mathcal{V}_{O_1}$ and $\mathcal{V}'_{O_2}$ as inputs and perform intra-graph reasoning to create enhanced graph representations. The function f_{GR} can be applied through graph convolution.

$$\begin{cases} V_{O_1} = f_{GR}(\mathcal{V}_{O_1}) = g(A^{intra}_{O_1}\mathcal{V}_{O_1}W_{O_1}) \in R^{C\times K} \\ V^{'}_{O_2} = f_{GR}(\mathcal{V}^{'}_{O_2}) = g(A^{intra}_{O_2}\mathcal{V}^{'}_{O_2}W_{O_2}) \in R^{C\times K} \end{cases} \tag{2}$$

where $g(\cdot)$ denotes a non-linear activation function, W_{O_1} and W_{O_2} are parameters that can be learned from the graph convolution layer, and $A^{intra}_{O_1}$ and $A^{intra}_{O_2}$ represent the graph adjacency matrices corresponding to $\mathcal{V}_{O_1}$ and $\mathcal{V}^{'}_{O_2}$. By revisiting the assignments from the f_{Rproj} step, graph representations can be mapped back to the original coordinates using the enhanced representations.

2.3 Feature Fusion Module and Multi-scale Cross Module

Medical images are complex in structure and rich in detail, considering both global information and local features. Feature fusion enables the segmentation network to fully use the global context information generated by the diffusion model, mitigating mis-segmentation due to insufficient global information. Therefore, we introduce FFM to leverage the connection between the characteristics of the segmentation network and those of DPM, as shown in Fig.2 (A).

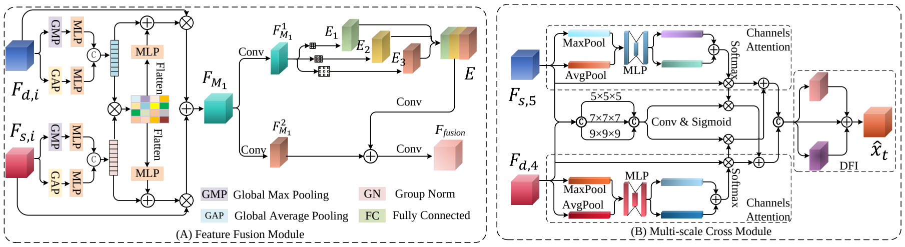

Fig. 2. (A) FFM. $F_{d,i}$ is the features of i stage in the encoder of the DU; (B) MCM. This module is processed with a set of convolutional blocks with kernels of different sizes to capture multi-scale features. $F_{s,5}$ is the feature obtained through graph reasoning.

The attention vectors $att_{s,i}$ and $att_{d,i} \in R^C$ are produced for $F_{s,i}$ and $F_{d,i}$, respectively. Subsequently, $att_{s,i}$ and $att_{d,i}$ are sequentially partitioned into k groups with a length of r, denoted as $G_{s,i}$ and $G_{d,i} \in R^{k\times r}$. Next, we establish the relationship matrix $\hat{R} \in R^{k\times r}$ by computing the inner product between $G_{s,i}$ and $G_{d,i}$ for each group pair, and the modulation matrix M as:

$$\begin{cases} \hat{R} = G_{s,i}G^T_{d,i}, M = \delta\left(att + \alpha f_c\left(Flatten\left(\hat{R}\right)\right)\right) \\ F_{M_1} = M_{s,i} \otimes F_{s,i} + M_{d,i} \otimes F_{d,i} \end{cases} \tag{3}$$

Then, to learn scale-aware feature representations, we feed $F^1_{M_1}$ into three convolutions by using $3\times3\times3$ dilated convolution denoted as $C^{r_i}_{3\times3\times3}$ with a dilated rate of $r_i \in \{2, 4, 8\}$, respectively, thus we can obtain $E_l = C^{r_i}_{3\times3\times3}\left(F^1_{M_1}\right), l \in$

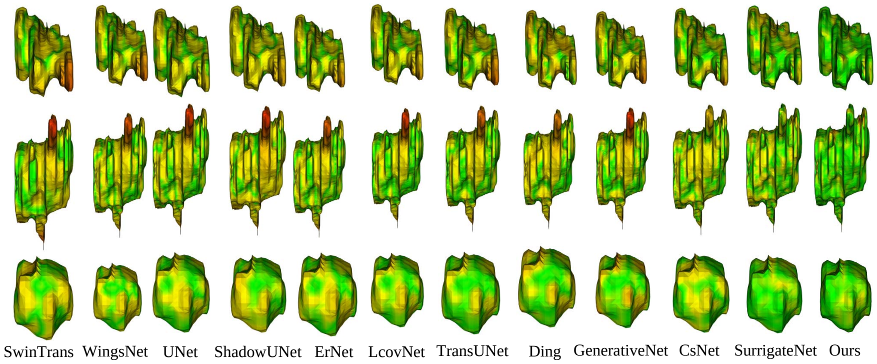

Fig. 3. Example of segmentation results for the self-collected CT image dataset.

$\{1, 2, 3\}$. Finally, the concatenated feature is fused with the original feature $F_{M_1}^2$. Thus, this entire procedure can be expressed as:

$$F_{fusion} = C_{3\times3\times3}\left(C_{3\times3\times3}\left(E\right) \oplus C_{3\times3\times3}\left(F_{M_1}^2\right)\right) \tag{4}$$

where $E = Cat\left(E_1, E_2, E_3\right)$ is the feature obtained by connecting the scale-aware features $E_l, l \in \{1, 2, 3\}$. F_{fusion} is the output fusion feature of FFM.

The MCM can fuse the cross-feature maps between the DPM and the segmentation network. Such feature fusion helps to fully use the global context and local features in the segmentation process, thus effectively solving the size inconsistency problem of lung nodules, as shown in Fig.2 (B). Specifically, channel attention [29] handles the input features for the segmentation network and the DPM. A matrix of adaptive coefficients is derived by concatenating input features and applying convolution operations to compress and separate channels. Different convolution kernels are set to capture multi-scale features in applying the cross structure. This procedure is represented as follows:

$$\begin{cases} A, B = Split(Conv_{\sum_{k\times k\times k}}\left(Cat\left(F_{d,i}, F_{s,i}\right)\right)) \\ \tilde{F_i} = Cat(A \otimes Cha\left(F_{d,i}\right) \oplus B, B \otimes Cha\left(F_{s,i}\right) \oplus A) \end{cases} \tag{5}$$

where A and B represent the adaptive coefficient matrices; $Split\left(\cdot\right)$ represents the operation of separating features from channels; $Cha\left(\cdot\right)$ indicates channel attention, respectively; $Conv_{\sum_{k\times k\times k}}, k \in \{5, 7, 9\}$ corresponds to convolution with a kernel size of $k \times k \times k$. The resulting features are concatenated and subsequently refined using the dual-branch feature integration (DFI) module. The DFI involves depth-wise convolutions and a residual connection. The final output of the MCM is $\hat{x}_t$.

3 Experiments

3.1 Datasets and Implementation Details

Our experiments used a self-collected dataset of CT images of patients with lung nodules collected by a local hospital, and a publicly accessible dataset called LUNA16[30]. The self-collected dataset contains 1299 samples with an image resolution of 1 *mm*. The self-collected data were ethically reviewed and informed consent was obtained from the patients. Experienced radiologists annotated the images based on surgical pathology results using ITK-SNAP software. In the preprocessing stage of the CT image, the grayscale image is utilized to determine the approximate location and size of the nodules. Subsequently, a region of interest (ROI) is selected to encompass the surrounding area of the nodules. The ROI size of the self-collected CT image and the LUNA16 datasets are $160 \times 160 \times 48$ and $32 \times 64 \times 64$, respectively.

Table 1. Evaluation of segmentation performance on the self-collected dataset.

Model	**Dice** (%)	**Pr** (%)	**Re** (%)	**JI** (%)
UNet [8]	73.4 ± 4.7	82.7 ± 7.1	71.7 ± 5.4	60.5 ± 5.2
TransUNet [15]	76.1 ± 3.3	75.3 ± 6.1	83.0 ± 3.5	64.5 ± 4.3
WingsNet [14]	71.6 ± 2.5	79.1 ± 3.9	72.9 ± 6.0	59.2 ± 2.6
SwinTrans [18]	71.2 ± 4.3	72.7 ± 7.9	78.3 ± 6.6	58.7 ± 4.6
ErNet [9]	75.3 ± 3.1	81.3 ± 5.7	76.2 ± 6.5	63.9 ± 3.5
SurrigateNet [10]	78.9 ± 1.7	86.1 ± 2.1	77.9 ± 3.9	68.4 ± 1.9
GenerativeNet [11]	77.7 ± 2.7	85.4 ± 1.8	76.8 ± 5.1	67.1 ± 3.0
CsNet [13]	78.5 ± 1.9	83.9 ± 3.4	79.3 ± 3.9	67.6 ± 2.2
LcovNet [7]	75.9 ± 2.2	82.1 ± 4.3	76.3 ± 6.2	64.0 ± 2.4
ShadowUNet [19]	73.5 ± 3.3	76.8 ± 6.0	77.7 ± 1.0	61.3 ± 4.0
Ding [6]	77.1 ± 1.3	77.2 ± 2.0	81.9 ± 3.0	65.3 ± 1.6
Ours	**83.2 ± 1.5**	**86.7 ± 2.6**	**81.8 ± 4.6**	**71.8 ± 2.3**

3.2 Experiments and Results Analysis

The quantitative comparison results of the Diff-UNet and other competing models on the self-collected and LUNA16 data sets are Table 1 and Table 2, respectively. It can be observed that since Diff-UNet benefits from the superior image generation ability of DPM and the global context capture ability of Transformer, it can generate segmentation maps with precise and accurate details, even in areas characterized by low contrast or blurriness. The Diff-UNet was also applied to the LUNA16 challenges as part of the validation process for the lung nodule segmentation task. We analyze the Dice coefficient (Dice), Precision (Pr), Jaccard Index (JI), and Recall (Re) of Diff-UNet and other models.

Table 2. Evaluation of segmentation performance on the LUNA16 dataset.

Model	**Dice** (%)	**Pr** (%)	**Re** (%)	**JI** (%)
UNet [8]	75.1 ± 2.1	80.9 ± 5.1	63.0 ± 3.2	76.0 ± 4.7
DAF3D [12]	77.5 ± 2.9	80.4 ± 4.6	65.6 ± 3.5	80.0 ± 6.7
LcovNet [7]	75.2 ± 1.9	83.1 ± 4.0	75.1 ± 4.2	64.1 ± 2.1
ShadowUNet [19]	73.5 ± 1.5	67.5 ± 3.0	89.7 ± 1.4	60.1 ± 1.8
Ding [6]	75.1 ± 1.8	81.8 ± 2.9	76.0 ± 5.2	63.9 ± 1.9
ErNet [9]	71.4 ± 1.5	84.9 ± 3.2	68.1 ± 1.6	60.4 ± 1.5
SurrigateNet [10]	72.4 ± 2.2	82.6 ± 5.0	71.6 ± 5.7	61.0 ± 2.2
GenerativeNet [11]	70.7 ± 1.9	84.4 ± 1.5	69.6 ± 2.7	59.7 ± 2.0
CsNet [13]	73.1 ± 2.5	84.8 ± 5.0	70.7 ± 6.0	62.2 ± 2.6
Ours	**81.6 ± 1.0**	**82.9 ± 2.4**	**84.9 ± 2.3**	**71.1 ± 1.1**

Fig. 3 is the 3D visual surface distance between the predicted result surface and the ground truth on the self-collected dataset. The segmentation results approach the ground truth more closely as the green area increases. It can be observed that Diff-UNet performs more accurate segmentation on parts that are difficult for human eyes to recognize. Since it can benefit from the diffusion model's superior generative ability and the Transformer's semantic representation ability, it can produce segmentation maps with precise and accurate details, even in regions with low contrast or blurriness. These comparisons demonstrate the effectiveness of our model in capturing lesion boundaries. This is due to the introduction of the CGI module in the decoder, which enables graph nodes to propagate cross-domain features and capture features of object boundaries.

3.3 Ablation Study

The ablation experiments were conducted on the self-collected CT image dataset, and the outcomes are presented in Table 3. We can see that FFM makes the segmentation network more focused on the lung nodule area in the CT image by fusing the particular noise information learned by DPM-based architecture with

Table 3. The ablation experiments of the Diff-UNet in the self-collected dataset. Baseline: Consists of DPM, Encoder, Bridge Transformer, and Decoder

Model				**Segmentation**			
Baseline	**CGI**	**FFM**	**MCM**	**Dice** (%)	**Pr** (%)	**Re** (%)	**JI** (%)
✓				76.9 ± 2.0	75.5 ± 4.4	84.0 ± 4.4	64.8 ± 2.4
✓	✓			79.6 ± 1.5	82.5 ± 5.9	81.1 ± 4.5	67.8 ± 1.5
✓	✓	✓		81.7 ± 1.4	85.5 ± 4.7	80.6 ± 4.2	70.1 ± 1.7
✓	✓	✓	✓	**83.2 ± 1.5**	**86.7 ± 2.6**	**81.8 ± 4.6**	**71.8 ± 2.3**

the semantic features of the segmentation network. The MCM combines the low-level features and texture information captured by the diffusion model with the high-level semantic features extracted by the segmentation network to obtain a comprehensive feature representation, thereby enhancing the ability to identify inconsistent lesion structures in images. The CGI module uses graph nodes to update and propagate cross-domain features that capture object boundaries.

4 Conclusion

We propose a new, comprehensive lung nodule segmentation framework that exploits the intuitive correlation between region and boundary features in CT images to produce more accurate segmentations. Experiments show that this model uses image area and edge features as graph nodes to update and propagate cross-domain features and uses DPM to weaken the impact of noise in CT images on segmentation effects and improve robust segmentation.

Acknowledgement. This work was supported partly by the National Natural Science Foundation of China (Grant Nos. 62171312,62276172, U22A2024, and 62271328), Shenzhen Science and Technology Program (Grant No. JCYJ2022081 8095809021), Shenzhen Medical Research Funds(No.C2301005), National Natural Science Foundation of Guangdong Province (Nos. 2024A1515011950).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., Jemal, A.: Global cancer statistics, 2012. CA, Cancer J. Clin **65**(2), 87–108 (2015)
2. Malvezzi, M., Carioli, G., Bertuccio, P., Boffetta, P., Levi, F., La Vecchia, C., Negri, E.: European cancer mortality predictions for the year 2019 with focus on breast cancer. Ann. Oncol. **30**(5), 781–787 (2019)
3. Wang, Q., Zhang, X., Zhang, W., Gao, M., Huang, S., Wang, J., Zhang, J., Yang, D., Liu, C.: Realistic lung nodule synthesis with multi-target co-guided adversarial mechanism. IEEE Tran. Med. Imaging **40**(9), 2343–2353
4. Hadjiiski, L., Napel, S., Goldgof, D., Perez, G., Arbelaez, P., Mehrtash, A., Kapur, T., Yang, E., Moon, J., Perez, G., et al.: Lung nodule malignancy prediction in sequential ct scans: Summary of isbi 2018 challenge. IEEE Tran. Med. Imaging **40**(12), 3748–3761
5. Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., Cattin, P.C.: Diffusion models for implicit image segmentation ensembles. In: International Conference on Medical Imaging with Deep Learnin. pp. 1336–1348. PMLR (2022)
6. Tripathi, P.C., Bag, S.: An attention-guided cnn framework for segmentation and grading of glioma using 3d mri scans. IEEE/ACM Trans. Comput. Biol. Bioinf. **20**(3), 1890–1904 ((2022))
7. Zhao, Q., Zhong, L., Xiao, J., Zhang, J., Chen, Y., Liao, W., Zhang, S., Wang, G.: Efficient multi-organ segmentation from 3d abdominal ct images with lightweight network and knowledge distillation. IEEE Tran. Med. Imaging **42**(9), 2513–2523 (2023)

8. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432. Springer, Cham (2016)
9. Xia, L., Zhang, H., Wu, Y., Song, R., Ma, Y., Mou, L., Liu, J., Xie, Y., Ma, M., Zhao, Y.: 3d vessel-like structure segmentation in medical images by an edge-reinforced network. Med. Image Anal. **82**, 102581 (2022)
10. Ezhov, I., Mot, T., Shit, S., Lipkova, J., Paetzold, J.C., Kofler, F., Pellegrini, C., Kollovieh, M., Navarro, F., Li, H., et al.: Geometry-aware neural solver for fast bayesian calibration of brain tumor models. IEEE Tran. Med. Imaging **41**(5), 1269–1278 (2021)
11. Chen, C., Zhou, K., Wang, Z., Xiao, R.: Generative consistency for semi-supervised cerebrovascular segmentation from tof-mra. IEEE Tran. Med. Imaging **42**(2), 346–353 (2022)
12. Wang, Y., Dou, H., Hu, X., Zhu, L., Yang, X., Xu, M., Qin, J., Heng, P.A., Wang, T., Ni, D.: Deep attentive features for prostate segmentation in 3d transrectal ultrasound. IEEE Tran. Med. Imaging **38**(12), 2768–2778 (2019)
13. Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., Liu, J.: Cs-net: Channel and spatial attention network for curvilinear structure segmentation. In: MICCAI. pp. 721–730. Springer (2019)
14. Zheng, H., Qin, Y., Gu, Y., Xie, F., Yang, J., Sun, J., Yang, G.Z.: Alleviating class-wise gradient imbalance for pulmonary airway segmentation. IEEE Tran. Med. Imaging **40**(9), 2452–2462 (2021)
15. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
16. Myronenko, A.: 3d mri brain tumor segmentation using autoencoder regularization. In: MICCAI. pp. 311–320. Springer (2019)
17. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: MICCAI. pp. 272–284. Springer (2021)
18. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218 (2022)
19. Xu, X., Sanford, T., Turkbey, B., Xu, S., Wood, B.J., Yan, P.: Shadow-consistent semi-supervised learning for prostate ultrasound segmentation. IEEE Tran. Med. Imaging **41**(6), 1331–1345 (2021)
20. Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H.: Transformers in medical imaging: A survey. Med. Image Anal. p. 102802 (2023)
21. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: MICCAI. pp. 263–273. Springer (2020)
22. Kumar, A., Tripathi, A.R., Satapathy, S.C., Zhang, Y.D.: Sars-net: Covid-19 detection from chest x-rays by combining graph convolutional network and convolutional neural network. Pattern Recognition **122**, 108255 (2022)
23. Wu, Y., Zhang, G., Gao, Y., Deng, X., Gong, K., Liang, X., Lin, L.: Bidirectional graph reasoning network for panoptic segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 9080–9089
24. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)

25. Brempong, E.A., Kornblith, S., Chen, T., Parmar, N., Minderer, M., Norouzi, M.: Denoising pretraining for semantic segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 4175–4186 (2022)
26. Rahman, A., Valanarasu, J.M.J., Hacihaliloglu, I., Patel, V.M.: Ambiguous medical image segmentation using diffusion models. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 11536–11546
27. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: MICCAI. pp. 528–538. Springer (2022)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 770–778 (2016)
29. Guo, M.H., Xu, T.X., Liu, J.J., Liu, Z.N., Jiang, P.T., Mu, T.J., Zhang, S.H., Martin, R.R., Cheng, M.M., Hu, S.M.: Attention mechanisms in computer vision: A survey. Comput. Vis. Media **8**(3), 331–368 (2022)
30. Setio, A.A.A., Traverso, A., De Bel, T., Berens, M.S., Van Den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M.E., Geurts, B., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. Med. Image Anal. **42**, 1–13 (2017)