ORIGINAL ARTICLE

Dense gate network for biomedical image segmentation

Dongsheng Li[1] · Chunxiao Chen[1] · Jianfei Li[1] · Liang Wang[1]

Abstract

Purpose Deep learning has recently shown its outstanding performance in biomedical image semantic segmentation. Most biomedical semantic segmentation frameworks comprise the encoder–decoder architecture directly fusing features of the encoder and the decoder by the way of skip connections. However, the simple fusion operation may neglect the semantic gaps which lie between these features in the decoder and the encoder, hindering the effectiveness of the network.

Methods Dense gate network (DG-Net) is proposed for biomedical image segmentation. In this model, the Gate Aggregate structure is utilized to reduce the semantic gaps between features in the encoder and the corresponding features in the decoder, and the gate unit is used to reduce the categorical ambiguity as well as to guide the low-level high-resolution features to recover semantic information. Through this method, the features could reach a similar semantic level before fusion, which is helpful for reducing semantic gaps, thereby producing accurate results.

Results Four medical semantic segmentation experiments, based on CT and microscopy images datasets, were performed to evaluate our model. In the cross-validation experiments, the proposed method achieves IOU scores of 97.953%, 89.569%, 81.870% and 76.486% on these four datasets. Compared with U-Net and MultiResUNet methods, DG-Net yields a higher average score on IOU and Acc.

Conclusion The DG-Net is competitive with the baseline methods. The experiment results indicate that Gate Aggregate structure and gate unit could improve the performance of the network by aggregating features from different layers and reducing the semantic gaps of features in the encoder and the decoder. This has potential in biomedical image segmentation.

Keywords Convolution neural networks · Semantic segmentation · Biomedical images · Deep learning

Introduction

The rise of deep learning greatly promotes the development of machine vision during recent years. The success of convolution neural networks has highlighted its capabilities in image classification, object detection and semantic segmentation. Image segmentation is a significant and basic task in biomedical image processing and analysis. An efficient and automatic segmentation method is helpful for the preprocessing of 3D image reconstruction, computer-assisted diagnosis, etc.

Since the fully convolution network (FCN) [1] was proposed, semantic segmentation methods based on deep learning have led to great interests from researchers in various fields. Many convolution neural networks (CNNs) have been proposed to promote the performance on different datasets. Zhao [2] proposed pyramid pooling module to improve the performance of the network for segmentation. Chen et al. [3] combined the Xception model and atrous spatial pyramid pooling module for improvement. Yu et al. [4] designed a bilateral network that includes spatial path and context path to realize real-time segmentation. At present, CNN models for semantic segmentation have been applied in many fields, such as street scene segmentation [5–7].

The research on semantic segmentation is also extended into the field of biomedical image segmentation [8, 9]. With high representation power, high prediction accuracy and fast inference ability, CNNs are competent in this task. For biomedical image segmentation, U-Net is one of the most commonly used models. U-Net [10] has an elegant architecture and is able to work with very few training images. These characteristics of U-Net make it suitable for biomedical image processing. With the development of research in biomedical image segmentation, many variants of U-Net

✉ Chunxiao Chen
  ccxbme@nuaa.edu.cn

[1] Department of Biomedical Engineering, Nanjing University
  of Aeronautics and Astronautics, Nanjing 211106, China

have emerged. Fausto [11] replaced 2D convolution operation with 3D convolution operation to process CT and MRI images in the 3D domain. Oktay [12] introduced the attention mechanism into U-Net. Hasan [13] proposed U-Net Plus for surgical tool segmentation. Despite improvements in their task, these models are still based on encoder–decoder architecture and fuse features with simple skip connections.

In this paper, we propose a dense gate network (DG-Net) for semantic segmentation in which a novel feature fusion structure is used to aggregate features from different layers.

Related work

In this section, we describe the technology and background relevant to our work.

Encoder–decoder architecture

Most segmentation convolution neural networks are in the form of encoder–decoder architecture. The encoder network extracts features from images and enriches the semantic information. The decoder network processes the features from the encoder network and provides the final prediction of the segmentation. The encoder sub-network is typically alternately stacked by series of convolution, pooling and nonlinear activation, while the decoder sub-network consists of convolution, upsampling and nonlinear activation. The decoder sub-network gradually enlarges features provided by the encoder to recover final semantic and spatial information. Usually, the encoder sub-network and the decoder sub-network are connected by skip connections to fuse features with different semantic and spatial information.

Skip connections

Skip connections have been proved to be beneficial to segmentation [14] and are widely used in many CNNs [5, 10–12]. However, skip connections used in previous architectures usually merge features of the encoder with features of the decoder of the corresponding level by one operation, which could not make full use of the semantic and spatial information in features. Meanwhile, the semantic gaps between the corresponding level features of encoder and decoder may cause some discrepancies in the learning process and affect the prediction adversely. Moreover, even though the low-level high-resolution features could remedy the spatial information lost in downsample operation, they still lack semantic information [15]. The simple fusion operation cannot endow low-level high-resolution features with sufficient semantic information for pixel-wise classification. In [16], the MultiResUNet is proposed to reduce the semantic gaps between features by replacing the skip connection with

the *Res* path. The *Res* path is composed of a sequence of residual layers. This improvement has been tested to promote the performance of the network by reducing the semantic gaps between features in the encoder and the decoder.

Feature fusion

Because deep high-level layers in encoder extract more semantic features and shallow low-level layers provide more spatial information, it is necessary to explore how to connect and combine these layers. There are two main methods for layer fusion [17]. One performs on depths and channels by fusing different semantic features, which improves the CNNs' ability of classification, such as DenseNets [18]. The other performs on resolutions and scales by compounding features with different resolutions, which is helpful for localization, such as feature pyramid architectures [2]. Meanwhile, the skip connections, that allow networks to retrieve the spatial information lost in pooling, are usually used to bridge the gaps between encoder and decoder. Recently, several novel feature fusion methods have been proposed. Yu et al. introduced a deep layer aggregation structure to fuse features in different layers [17]. Zhou et al. presented a nested U-Net architecture [19] to bridge the gap between different level features. These architectures improved the performance of semantic segmentation models and gave us an important clue to design a light structure for multiple feature fusion.

Gate units

In most of the CNN models, the conventional way to fuse low-level high-resolution features and high-level low-resolution features is summing or concatenating them directly by skip connections. Even though the information from low-level high-resolution features may be helpful for localization, they are insufficient to recover the semantic information. Thus, it is sensible to constrain those features with discriminative semantic information. Islam et al. [20] presented the gate unit to filter finely localized layers with deeper layers. Zhang et al. [15] proposed semantic embedding branch to fuse features with different locations and semantic information. These methods are aimed to introduce discriminative semantic information of high-level low-resolution features to low-level high-resolution features and guiding the recovery of prediction in the semantic domain.

Methods

Our DG-Net combines dense nested feature connection structure with the gate unit. The main structure of DG-Net is shown in Fig. 1. Similar to many semantic segmentation CNN models, our DG-Net is in the form of encoder–decoder archi-
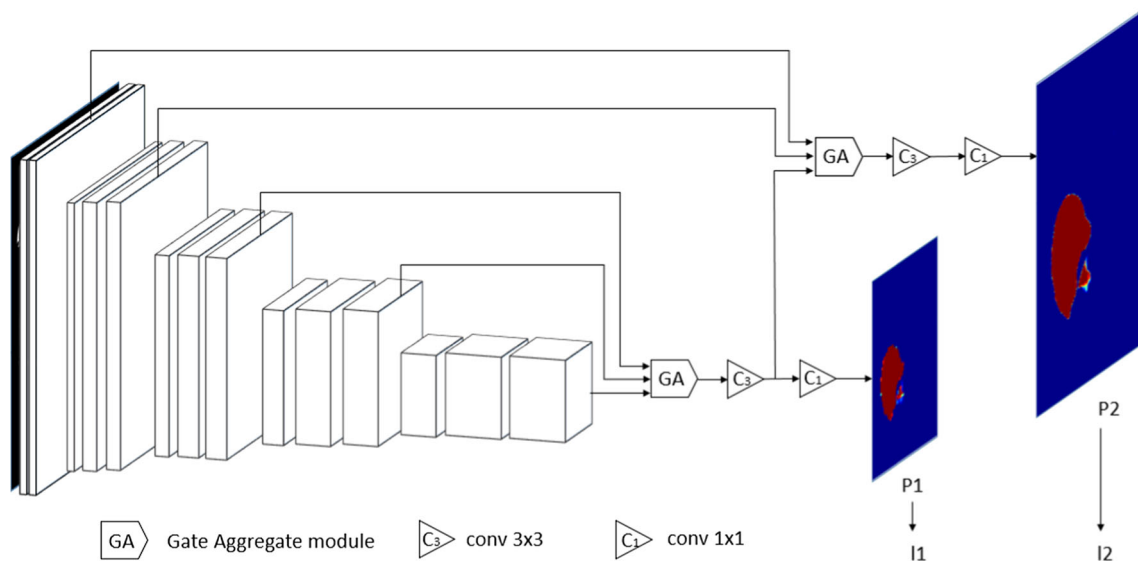
Fig. 1 The main architecture of DG-Net

tecture. The encode sub-network is the same as U-Net. Every encode stage consists of two 3×3 convolutions with batch normalization and leaky ReLU layer. And there are four 2×2 max-pooling operations in the encode pathway. In the decode pathway, DG-Net employs Gate Aggregate (GA) module to fuse features, and each of them fuses features of three adjacent levels. Features generated by the first GA module are processed by a 3×3 convolution with batch normalization and leaky ReLU layer before entering the second Gate Aggregate module. At the same time, a 1×1 convolution is used to provide a coarse score map P1 for deep supervision. Since P1 has smaller spatial resolution compared with the ground-truth label, we downsampled the ground-truth label to the same resolution of P1 to form loss function l1. The final score map P2 is obtained in the same way and is used to form loss function l2. The final loss function is the summation of l1 and l2.

The main novelty of our network is the designation of the Gate Aggregate module, which uses features in three levels to construct a small decoder network. This small decoder network makes full use of spatial and semantic information of features and bridges the gaps between different features. We describe the structure of the Gate Aggregate module in detail in the next section.

Gate Aggregate module

The conventional networks, such as U-Net, tend to directly combine features in the encoder with features in the decoder with the same spatial resolution. However, the features in the encoder undergo fewer operations than the features in the decoder do. So, there are semantic differences between these two types of features. The direct connections between these

two incompatible types of features may impede the learning procedure of networks. To solve this problem, the GA module is used to bridge the semantic gaps between the features in the encoder and the decoder sub-network before fusion.

The detail of the GA module is shown in Fig. 2. The GA module aggregates three adjacent level features (f1, f2, f3) and forms a dense connected fusion structure. The features with low resolution are upsampled by standard bilinear interpolation method to the same size of features with high resolution. And then, a 1×1 convolution is used to adjust the channels of upsampled low-resolution features to the same channels of high-resolution features. The gate units and convolutions build a grid-like structure, in which features are fused from low resolution to high resolution to make a progressive upsampling and decoding process. The dense gate units and convolutions reduce the semantic discrepancy between features in encoder and corresponding features in decoder gradually.

The backbone of the GA module can be regarded as a small U-Net++ L^2 network [19]. As implied in [19], the inference accuracy does not increase linearly with the complexity of U-Net++ models. The U-Net++ L^2 network is a nice compromise between accuracy and model complexity. Thus, a three-layer U-Net++ backbone may keep the expression of the network at a maximum degree with fewer parameters. Meanwhile, in contrast to U-Net++ which concatenates two features for fusion, the GA module uses the gate unit to implement features fusion. The motivation to introduce the gate unit into the network is that even though the dense convolutions reduce the gap between the features in the encoder and the decoder, the shallow low-level high-resolution features of encoder are still deficient on necessary semantic information. The detail of the gate unit is described in the next section.

Fig. 2 Structure of Gate Aggregate module
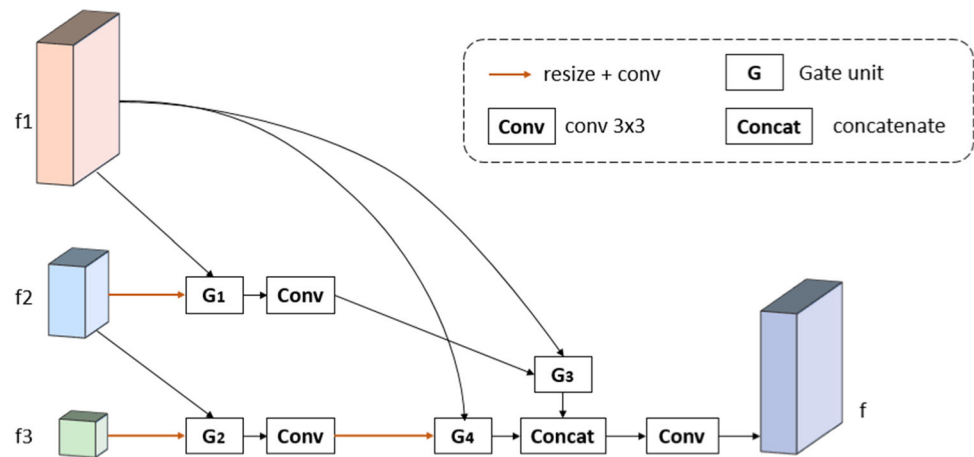
Gate unit

The structure of gate unit is shown in Fig. 3. The process in gate unit can be formulated as:

$$f = concat(f_l, f_l * f_h) \qquad (1)$$

where the *concat* operator denotes the concatenation in the dimension of channels. And f_l represents the upsampled low-resolution features. f_h is the high-resolution features, and f is the final features.

Before gate unit, features with high-semantic-level low resolution (features f_l) are upsampled and filtered with convolution to the same size of low-semantic-level high-resolution features (features f_h). In the gate unit, features f_h are filtered by features f_l with element-wise multiplication to enhance the semantic information and the location information. Finally, low-resolution features and the filtered high-resolution features are concatenated together as a new feature for the next convolution.

The gate unit is designed to control information passed on and reinforce the semantic information of low-level high-resolution features. The procedure of the gate unit and its preprocessing are similar to the method proposed in [20]. But unlike [20], the low-level high-resolution features are not abandoned after pixel-wise multiplication but are reused to provide inherent spatial information.
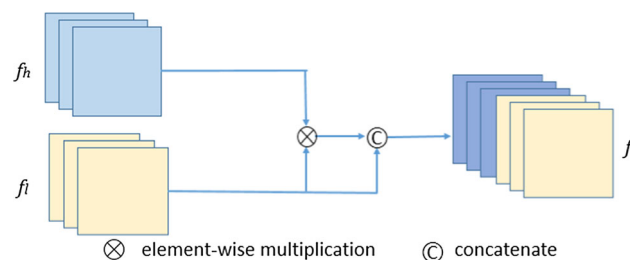

Fig. 3 The structure of gate unit

Experiments and results

Datasets

To evaluate the proposed model, four medical imaging datasets are used for comparison, including CT and microscopy images. The details of those datasets are listed in Table 1.

The liver segmentation dataset comes from liver tumor segmentation challenge (LiTS). There are 131 3D CT images in the training dataset of LiTS. The ground-truth labels of the test set of LiTS are not publicly available, so only the train set is used in this work. We sampled three sections that contain liver from each 3D CT image. The mouse segmentation dataset is provided by our team for 3D CT visualization. The main task of this dataset is to remove the CT scanning bed from the image to provide better data for 3D visualization and reconstruction. Both the PHC and N2DH datasets are obtained in cell tracking challenge. The PHC is the PhC-C2DH-U373 dataset, and the N2DH represents the Fluo-N2DH-SIM + dataset. The same to the LiTS dataset, only the training datasets of PHC and N2DH are available.

Data augmentation is essential for network training. It helps the network learn the invariance and perform more robustly. It is especially significant when the dataset is small. Thus, all four datasets were augmented before training by the Augmentor library in Python [23]. We augmented liver

Table 1 Datasets used in our experiments

Dataset	Image number	Modality
Liver[a]	393	CT
Mouse	320	CT
PHC [21, 22][b]	34	Microscopy
N2DH [21, 22][b]	215	Microscopy

[a]MICCAI 2018 LiTS challenge
[b]Cell tracking challenge

Table 2 Overview of augmented datasets

Dataset	Augment methods				Input size
	Flip	Rotation	Distortion	Shear	
Liver	√	√			256×256
Mouse	√	√			256×256
PHC	√	√	√	√	256×352
N2DH	√	√	√	√	288×256

and mouse dataset by flipping and rotation (integer multiple of 90°) operations. For PHC and N2DH, random distortions and shear were also used. We then downsampled every image and label to the input size due to computational constraints. The augment methods and input resolutions of datasets are listed in Table 2.

Baseline

The U-Net was used as a baseline in our work since it is a commonly used semantic segmentation network in biomedical image processing. The filter numbers of five decode procedures of encoder sub-network in U-Net are 16, 32, 64, 128 and 256. For a fair comparison, the filter numbers of decode sub-network of DG-Net are the same as U-Net.

To further evaluate the performance of DG-Net, the MultiResUNet [16] was used for comparison. The filter numbers of *MultiRes* blocks in MultiResUNet are 16, 32, 64, 128 and 256. To keep the number of parameters of MultiResUNet is on a similar scale as U-Net and DG-Net, we selected the scale coefficient $\alpha = 1.15$.

For a certain dataset, all models were trained under the same condition. The parameters and flops of U-Net, MultiResUNet and DG-Net are listed in Table 3.

Implementation details

We have implemented our method using Tensorflow 1.9.0 and Python 3.5.6. The parameters in convolution layers were initialized by Xavier initializer. We employed *softmax cross-entropy* as the loss function to be optimized. The Adam Optimizer was used for optimization. The batch size used in training process was 5.

Table 3 Parameters and flops of U-Net, MultiResUNet and DG-Net

Model	#Param ($\times 10^6$)	#Flops ($\times 10^6$)
U-Net	1.94	9.71
MultiResUNet	1.92	9.61
DG-Net	1.82	9.10

Metrics

Given a binary segmentation map *Seg* and the corresponding manually annotation *Ann*, the definition of *TP, FP, TN, FN* is as follows:

$$TP = \sum_{ij} (Seg_{ij} = 1 \text{ and } Ann_{ij} = 1),$$

$$FP = \sum_{ij} (Seg_{ij} = 1 \text{ and } Ann_{ij} = 0)$$

$$TN = \sum_{ij} (Seg_{ij} = 0 \text{ and } nn_{ij} = 0),$$

$$FN = \sum_{ij} (Seg_{ij} = 0 \text{ and } Ann_{ij} = 1)$$

To evaluate our method quantitatively, the performance is measured by pixel intersection-over-union (*IOU*) of objects and accuracy (*Acc*). The equations of these metrics are as follows:

$$IOU = \frac{TP}{FN + FP + TP} \tag{2}$$

$$Acc = \frac{TP + TN}{FN + FP + TP + TN} \tag{3}$$

Experimental results

In this paper, fivefold cross-validation is used to quantitatively evaluate the performance of U-Net, MultiResUNet and DG-Net. In each experiment, the training set consists of fourfold, and the other fold is used for testing. Among the training set, 20% of the data are held back for validation. We repeated the experiments three times to reduce the impact

Table 4 Quantitative results of U-Net, MultiResUNet and DG-Net

Dataset	Method	IOU (%)	Acc (%)
Mouse	U-Net	97.587 ± 0.997	99.680 ± 0.092
	MultiResUNet	97.380 ± 1.045	99.642 ± 0.117
	DG-Net	$\mathbf{97.953 \pm 1.010}$	$\mathbf{99.718 \pm 0.108}$
Liver	U-Net	89.323 ± 1.239	98.982 ± 0.160
	MultiResUNet	89.266 ± 1.134	98.938 ± 0.181
	DG-Net	$\mathbf{89.569 \pm 1.477}$	$\mathbf{99.024 \pm 0.169}$
PHC	U-Net	81.559 ± 3.297	98.622 ± 0.205
	MultiResUnet	80.966 ± 3.300	98.574 ± 0.211
	DG-Net	$\mathbf{81.870 \pm 3.261}$	$\mathbf{98.647 \pm 0.214}$
N2DH	U-Net	76.050 ± 0.921	97.832 ± 0.134
	MultiResUNet	76.149 ± 0.973	97.840 ± 0.141
	DG-Net	$\mathbf{76.486 \pm 0.765}$	$\mathbf{97.869 \pm 0.132}$

The best result of each experiment is marked in bold

Table 5 Statistical analysis of U-Net, MultiResUNet and DG-Net

Dataset	U-Net versus DG-Net		MultiResUNet versus DG-Net	
	IOU	Acc	IOU	Acc
Mouse	**0.002**	0.071	**< 0.001**	**< 0.001**
Liver	0.179	**0.039**	0.203	**0.032**
PHC	0.064	0.056	**< 0.001**	**< 0.001**
N2DH	**< 0.001**	**< 0.001**	**0.012**	**0.016**

The p-values which smaller than 0.05 are marked in bold

of randomness in the training process. The results of U-Net, MultiResUNet and DG-Net on mouse, liver, PHC and N2DH are listed in Table 4. It can be observed that though DG-Net performed differently on different datasets, it performed better than U-Net on all datasets. Compared with U-Net, DG-Net achieves at least 0.25% relative improvement on IOU. In terms of Acc, the difference between U-Net and DG-Net is not very obvious. Compared with MultiResUNet, though DG-Net has fewer parameters, it surpasses MultiResUNet on all four datasets.

Furthermore, the statistical analysis was performed to evaluate the performance of DG-Net. The paired t-tests were adopted to assess the significance of IOU and Acc between U-Net, MultiResUNet and DG-Net. As shown in Table 5, compared with U-Net and MultiResUNet, DG-Net could achieve better performance than baseline with significant difference in most cases.

As a popular used biomedical image segmentation model, U-Net can provide accurate segmentation results in most cases. The segmented results of U-Net, MultiResUNet and DG-Net are shown in Fig. 4. In these examples, these three methods could fit the ground truth nicely.

However, U-Net produces defective results on some other samples, as shown in Fig. 5. U-Net seems to be over-segmented or under-segmented on these samples. In Fig. 5a1, b1, c1, the foreground objects are not segmented completely, so there are some holes or gaps in the results. In Fig. 5a2, b2, c2, U-Net misdeems some small objects in other tissue or background as foreground objects. In Fig. 2d1, d2, some areas around the target are segmented as part of the foreground target. DG-Net performs more stable compared with U-Net.

For some images, these three methods may fail to provide accurate results, as shown in Fig. 6. In Fig. 6a1, these three methods all produce over-segmented results. In Fig. 6a2, all three methods make some wrong predictions and classify some background areas as foreground, but DG-Net makes relatively fewer mistakes.

In Fig. 6b1, b2, irregular results are produced by U-Net and MultiResUNet. Though DG-Net also cannot provide perfect results on these samples, it produces results with fewer irreg-

ularities along the boundaries and performs better than other methods, which is meaningful to provide better references for clinical diagnosis.

In Fig. 6c1, c2, all three methods under-segment or over-segment cells. Among them, DG-Net makes fewer mistakes. In Fig. 6d1, d2, the microscopy images contain more noise, and signals of cells are weaker than instances in Fig. 4d; thus, these images are more difficult for segmentation. The segmentation results of U-Net contain many outliers that are very similar to cells visually, which may mislead the subsequent analysis. MultiResUNet and DG-Net perform more stable compared with U-Net and can provide more valuable information.

The results of experiments show that DG-Net has the potential to produce competitive segmentation results of biomedical images. In some situations, DG-Net performs similarly to U-Net. However, in general, DG-Net is more stable and could provide more meaningful results.

Discussion

The results in the experiments show that DG-Net could yield remarkable accuracy. These achievements are partially attributed to the employment of the Gate Aggregate module and the gate unit. DG-Net employs the Gate Aggregate module, which is based on the backbone of UNet++ L^2, and gate unit to aggregate features in three layers. The light three layers U-Net++ structure can maintain the expression of the network as much as possible with less parameters. At the same time, the gate unit can enrich the semantic information of low-level high-resolution features which may improve the ability of the network further. In Fig. 7, features in G_4 (Fig. 2) of the first and second GA modules are visualized to show the effect of gate unit.

In the first stage, high-resolution feature f_h distinguishes the mouse and the CT scanning table from the background, but is unable to discriminate the mouse and the CT scanning table. However, the upsampled low-resolution feature f_l can effectively differentiate the mouse from the background and the CT scanning table. The result of gate unit f gains considerable semantic information compared with feature f_h. Features in the second stage have a higher resolution than features in the first stage. The f_h in the second stage contains ample spatial information, yet embodies less ability to distinguish targets. After filtered by the gate unit, the target area becomes clearer. Figure 7 demonstrates that the gate unit could efficiently enrich the semantic information of high-resolution features, which may be beneficial to the learning process.
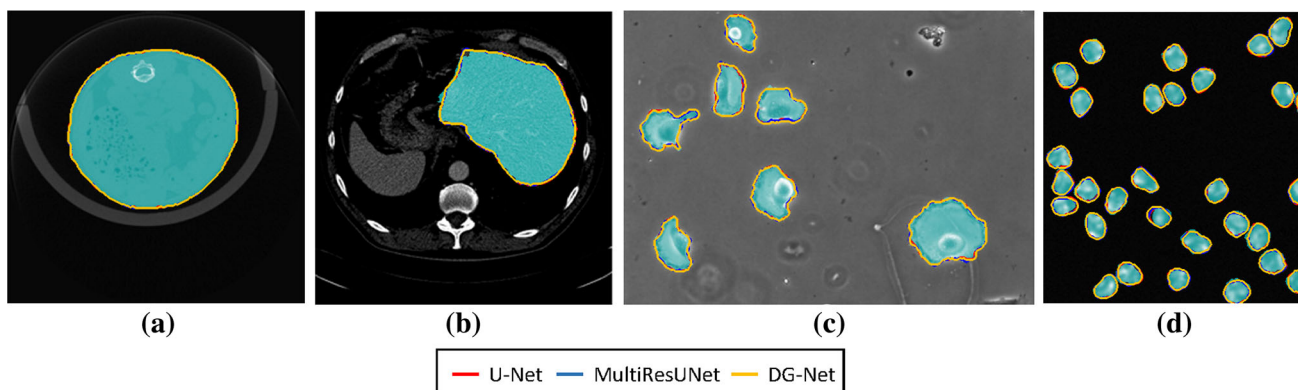
Fig. 4 The results of U-Net, MultiResUNet and DG-Net on different datasets. **a** Sample of mouse CT dataset. **b** Sample of liver dataset. **c** Sample of PHC dataset. **d** Result of N2DH cell dataset. The ground truth is marked in cyan. And the outline of U-Net result is marked in red, the outline of MultiResUNet result is marked in blue, and the outline of DG-Net result is marked in orange. The order of coloring is U-Net, MultiResUNet and DG-Net; therefore, in the overlap area the line drawn after will cover the previous line
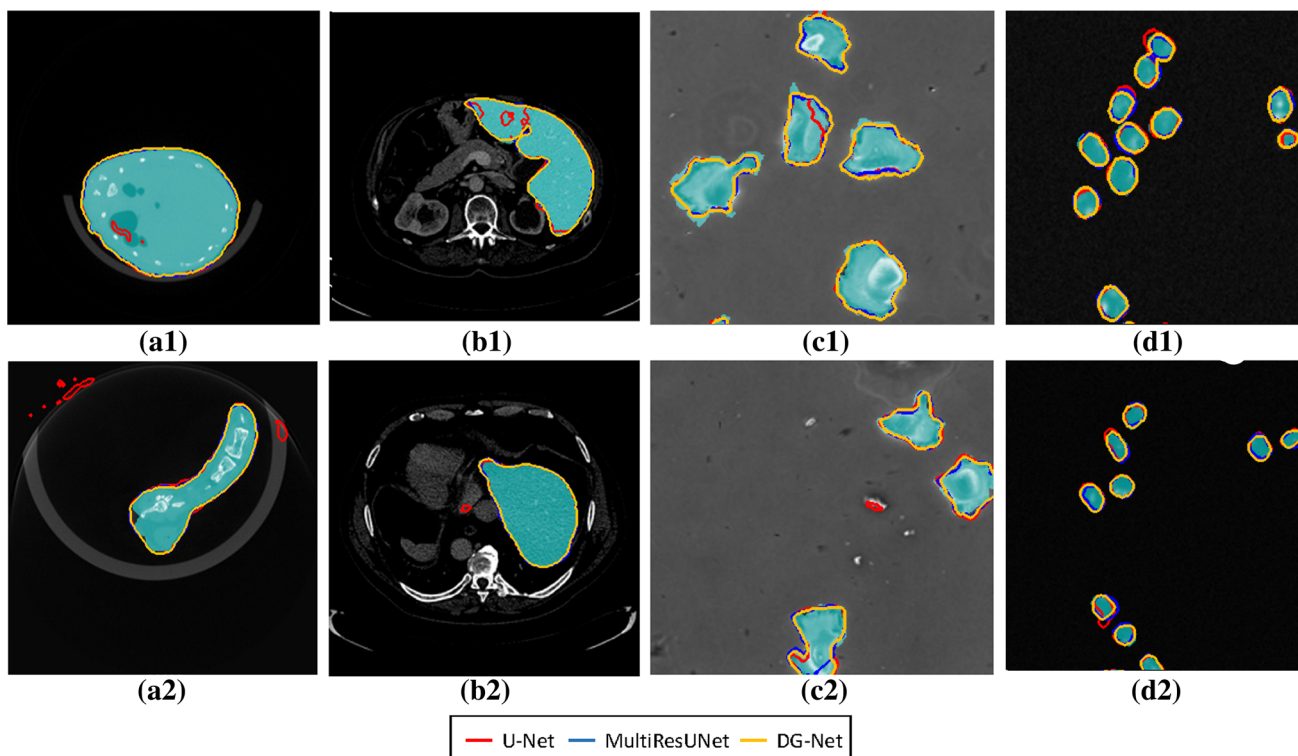

Fig. 5 Some defective segmentations of U-Net. **a1**, **a2** Samples of mouse CT dataset, **b1**, **b2** samples of liver dataset, **c1**, **c2** part of two samples of PHC dataset, **d1**, **d2** part of two samples of N2DH dataset. In order to show the detail of segmentation, the samples of PHC and N2DH are cropped to highlight the region of interest

Conclusion

In this paper, we focus on the feature fusion method in classical U-Net architecture and present a dense gate network for semantic segmentation. The Gate Aggregate module assembles features in the adjacent three levels as a small sub-network with a stratified structure. In this structure, semantic gaps between features are reduced gradually by the dense connected gate units and convolutions. Meanwhile, the gate unit is designed to enhance the semantic information of low-level high-resolution features to recover accurate results. The DG-Network outperforms U-Net in four datasets including CT images and microscopy images according to the experiments, which demonstrate that better fusion methods can
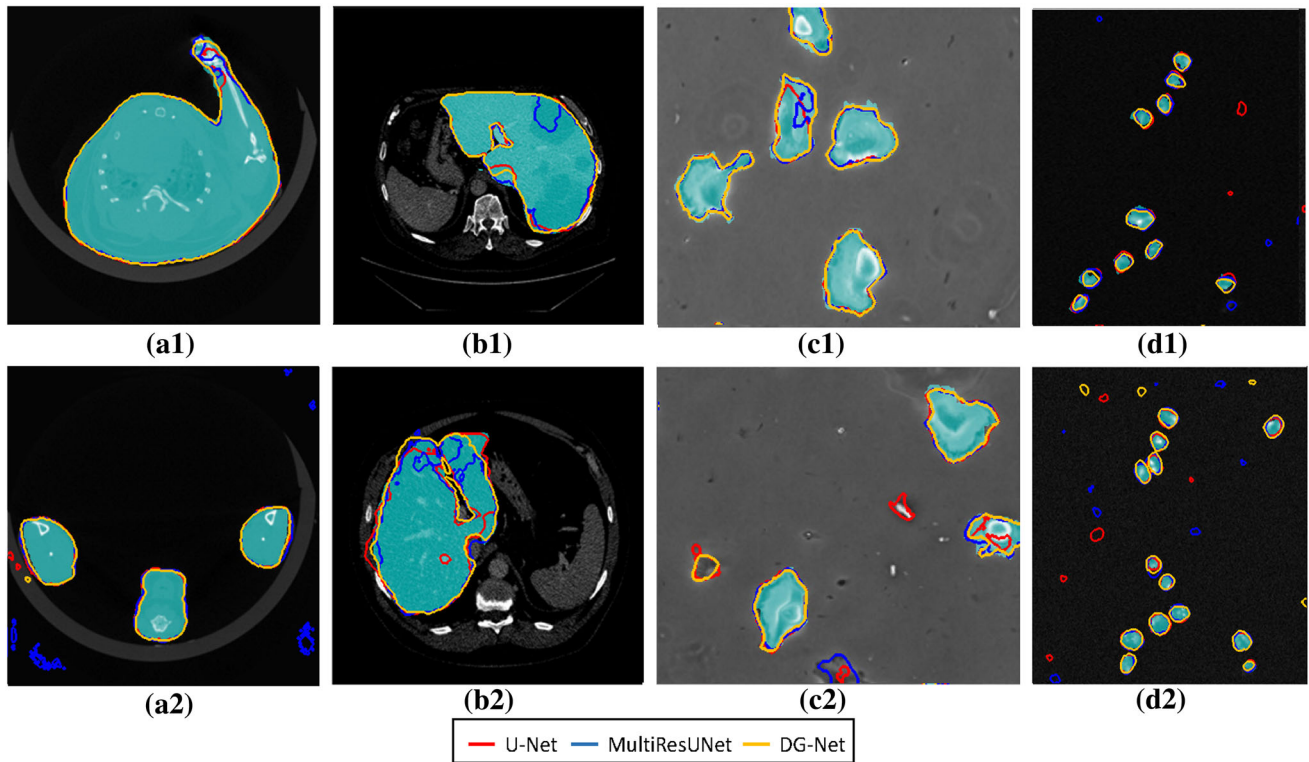
(a1)   **(b1)**   **(c1)**   **(d1)**

(a2)   **(b2)**   **(c2)**   **(d2)**

— U-Net   — MultiResUNet   — DG-Net

Fig. 6 Some results of U-Net, MultiResUNet and DG-Net. **a1**, **a2** Samples of mouse CT dataset, **b1**, **b2** samples of liver dataset, **c1**, **c2** results of PHC dataset, **d1**, **d2** samples of N2DH dataset

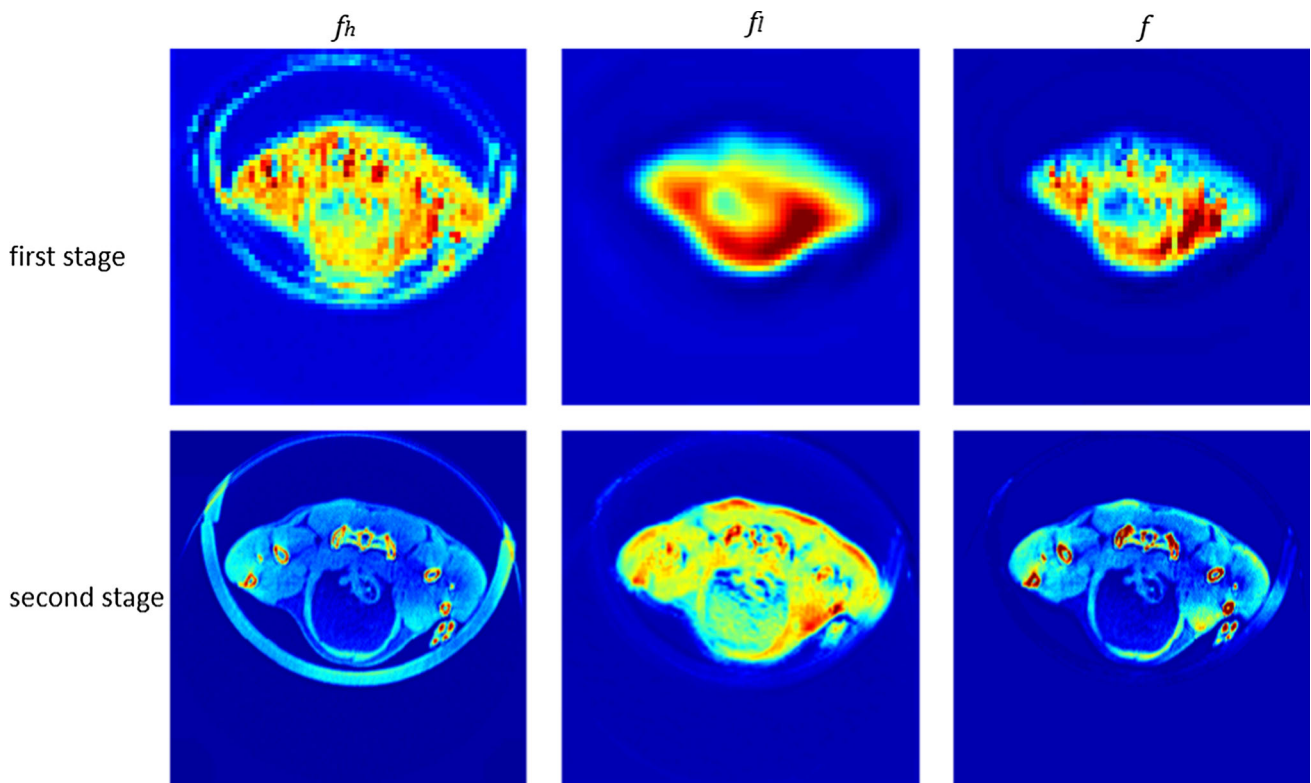f_h f_l f

first stage

second stage

Fig. 7 The effect of gate unit

improve the performance of networks. Therefore, DG-Net has the potential to become a reliable segment network.

Compliance of ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The mouse CT images were scanned in Southeast University, and the experiment was approved by the Animal Ethics Committee of Southeast University. All applicable Animal Ethics Committee of Southeast University guidelines for the care and use of animals were followed.

Informed consent This article does not contain patient data.

References

1. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: CVPR. arxiv:1411.4038
2. Zhao H, Shi J, Qi X, Wang X, Jia J (2017) Pyramid scene parsing network. In: CVPR. arxiv:1411.4038
3. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder–decoder with atrous separable convolution for semantic image segmentation. In: CVPR. arxiv:1802.0ds2611fds
4. Yu C, Wang J, Peng C, Gao C, Yu G, Sang N (2018) BiSeNet: bilateral segmentation network for real-time semantic segmentation. In: ECCV. arxiv:1808.00897
5. Chaurasia A, Culurciello E (2017) LinkNet: exploiting encoder representations for efficient semantic segmentation. In: VCIP 2017. arxiv:1707.0378
6. Pohlen T, Hermans A, Mathias M, Leibe B (2017) Full-resolution residual networks for semantic segmentation in street scenes. In: CVPR. arxiv:1611.08323
7. Meletis Panagiotis, Dubbelman Gijs (2018) Training of convolutional networks on multiple heterogeneous datasets for street scene semantic segmentation. IEEE Intell Veh Symp (IV) 2018:1045–1105
8. Fu H, Xu Y, Wing D, Wong K, Liu J (2016) Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). IEEE
9. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng P-A (2016) 3D Deeply supervised network for automatic liver segmentation from CT volumes. In: CVPR 2016. arxiv:1607.00582
10. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. arxiv:1505.04597
11. Beers A, Chang K, Brown J, Sartor E, Mammen CP, Gerstner E, Rosen B, Kalpathy-Cramer J (2017) Sequential 3D U-Nets for biologically-informed brain tumor segmentation. In: CVPR 2017. arxiv:1709.02967
12. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018). Attention U-Net: learning where to look for the pancreas. In: 1st conference on medical imaging with deep learning (MIDL 2018). arxiv:1804.03999
13. Kamrul Hasan SM, Linte CA (2019) U-NetPlus: a modified encoder–decoder U-Net architecture for semantic and instance segmentation of surgical instrument. arixv:1902.08994
14. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C (2016) The importance of skip connections in biomedical image segmentation. arxiv:1608.04117
15. Zhang Z, Zhang X, Peng C, Cheng D, Sun J (2018) ExFuse: enhancing feature fusion for semantic segmentation. In: European conference on computer vision. Springer, Cham. arxiv:1804.03821
16. Ibtehaz N, Sohel Rahman M (2019) MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. arxiv:1902.04049
17. Yu F, Wang D, Shelhamer E, Darrell T (2018) Deep layer aggregation. In: CVPR 2018. arxiv:1707.06484
18. Huang G, Liu Z, van der Maaten L (2016) Densely connected convolutional networks. 2016. In: CVPR. arxiv:1608.06993
19. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) UNet++: a nested U-Net architecture for medical image segmentation. In: *LMIA*. arxiv:1807.10165
20. Islam MA, Rochan M, Bruce NDB, Wang Y (2017) Gated feedback refinement network for dense image labeling. In: CVPR
21. Ulman V, Maška M, Magnusson KEG, Ronneberger O, Haubold C, Harder N, Matula P, Matula P, Svoboda D, Radojevic M, Smal I, Rohr K, Jaldén J, Blau HM, Dzyubachyk O, Lelieveldt B, Xiao P, Li Y, Cho S-Y, Dufour AC, Olivo-Marin J-C, Reyes-Aldasoro CC, Solis-Lemus JA et al (2017) An objective comparison of cell-tracking algorithms. Nat Methods 14:1141–1152
22. Maška M, Ulman V, Svoboda D, Matula P, Matula P, Ederra C, Urbiola A, Españ T, Venkatesan S, Balak DMW, Karas P, Bolcková T, Štreitová M, Carthel C, Coraluppi S, Harder N, Rohr K, Magnusson KEG, Jaldén J, Blau HM, Dzyubachyk O, Křížek P, Hagen GM, Escuredo DP, Jimenez-Carretero D, Ledesma-Carbayo MJ, Muñoz-Barrutia A, Meijering E, Kozubek M, Ortiz-de-Solorzano C (2014) A benchmark for comparison of cell tracking algorithms. Bioinformatics 30(11):1609–1617
23. Bloice Marcus D, Roth Peter M, Holzinger A (2019) Biomedical image augmentation using Augmentor. Bioinformatics 35(21):4522–4524. https://doi.org/10.1093/bioinformatics/btz259