kCBAC-Net: Deeply Supervised Complete Bipartite Networks with Asymmetric Convolutions for Medical Image Segmentation

Pengfei Gu, Hao Zheng, Yizhe Zhang, Chaoli Wang, and Danny Z. Chen

Department of Computer Science and Engineering, University of Notre Dame,
Notre Dame, IN 46556, USA
{pgu, hzheng3, yzhang29, chaoli.wang, dchen}@nd.edu

Abstract. Accurate and automatic medical image segmentation is challenging due to significant size and shape variations of objects (e.g., in multi-scales) and missing/blurring object borders. In this paper, we propose a new *deeply supervised k-complete-bipartite network with asymmetric convolutions* (kCBAC-Net) to exploit multi-scale features and improve the capability of standard convolutions for segmentation. (1) We leverage a generalized complete bipartite network to reuse multi-scale features, consolidate feature hierarchies at different scales, and preserve maximum information flow between encoder and decoder layers. (2) To further capture multi-scale information, we sequentially connect k complete bipartite network modules together to facilitate their processing in different image scales. (3) We replace the standard convolution by asymmetric convolution block to strengthen the central skeleton parts of standard convolution, enhancing the model's robustness on exploiting more discriminative features. (4) We employ auxiliary deep supervisions to boost information flow in the network and extract highly discriminative features. We evaluate our kCBAC-Net on three datasets (ultrasound lymph node segmentation (2D), 2017 ISIC Skin Lesion segmentation (2D), and MM-WHS CT (3D)), achieving state-of-the-art performance.

1 Introduction

Accurate image segmentation is critical to medical image analysis, disease diagnosis, and clinical applications (e.g., quantitative analysis of lymph node sizes and shapes [28], melanoma diagnosis [3], and cardiovascular surgical planning [18]). However, automatic medical image segmentation with high accuracy is a very non-trivial task due to multiple challenging factors, including (I) multi-scale objects: the variations of object sizes and appearances can be very large, and in some extreme cases, the ratio between the largest and smallest objects could be hundreds; (II) missing/blurring object borders: due to low contrast or noisy background, the borders between objects or background may be missing or very ambiguous. Fig. 1 gives some examples to illustrate such challenges.

Many deep learning based image segmentation methods were proposed to address these challenges. There are three main types of methods for dealing with

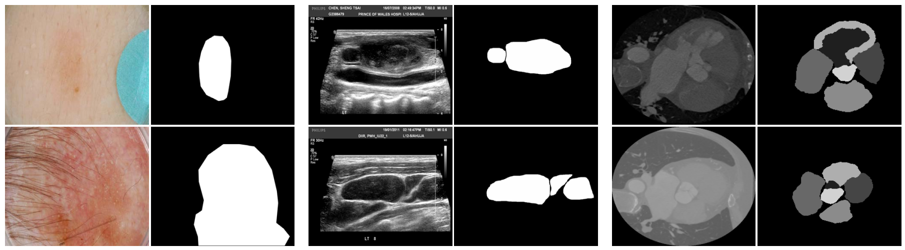

Fig. 1. Input image examples (left) and their ground truth (right) (best viewed in color). (a) 2017 ISIC Skin Lesion segmentation dataset [10]: Lesions vary a lot in sizes and shapes and their borders could be very ambiguous due to low contrast between lesion regions and surrounding skin. (b) Ultrasound lymph node images: Multi-scale objects and unclear borders increase the difficulty of distinguishing real lymph nodes and regions highly similar to lymph nodes. (c) 2017 MM-WHS CT dataset [31]: Different detailed structures are in various scales and boundaries are unclear.

the issue of multi-scale objects. (1) Encoder-decoder structures: U-Net [20] and its variants [30,16] utilized skip connections to fuse multi-scale features extracted from the encoder to the counterparts in the decoder. (2) Multi-path diagram: These frameworks [21,9] fed multi-scale images to the same network (with shared weights) and captured features from various scale inputs. (3) Spatial pyramid design: DeepLabv3 [8] utilized an atrous spatial pyramid pooling module containing multiple parallel atrous convolution layers to extract multi-scale contextual information. An array of studies sought to tackle the missing/blurring border issue. In [5], two branches were used to segment the main regions-of-interest (ROIs) and the corresponding contours separately. In [27,28], multiple sub-modules were built to encourage deep neural networks to learn richer and more comprehensive features. Despite yielding promising performance on various segmentation tasks, these known methods tended to emphasize on addressing one challenge while neglecting the other. It is highly desirable to tackle these two major issues simultaneously for accurate and robust segmentation.

In this paper, we propose a new *deeply supervised k-complete-bipartite network with asymmetric convolutions* (kCBAC-Net), aiming to extensively exploit multi-scale features and enhance the capability of standard convolutions on extracting discriminative features. Specifically, we develop kCBAC-Net with the following ideas. (1) A generalized complete bipartite network adopted from CB-Net [6] is employed to reuse and aggregate multi-scale features and boost information flow between encoder and decoder layers. (2) Following the structure of kU-Net [7], k complete bipartite network modules for processing different image scales are systematically connected to enhance and assimilate multi-scale information. (3) The standard convolution is incorporated with asymmetric convolutions [11] to strengthen the central skeleton portions of the standard convolution, reinforcing the network's robustness on exploiting discriminative features.

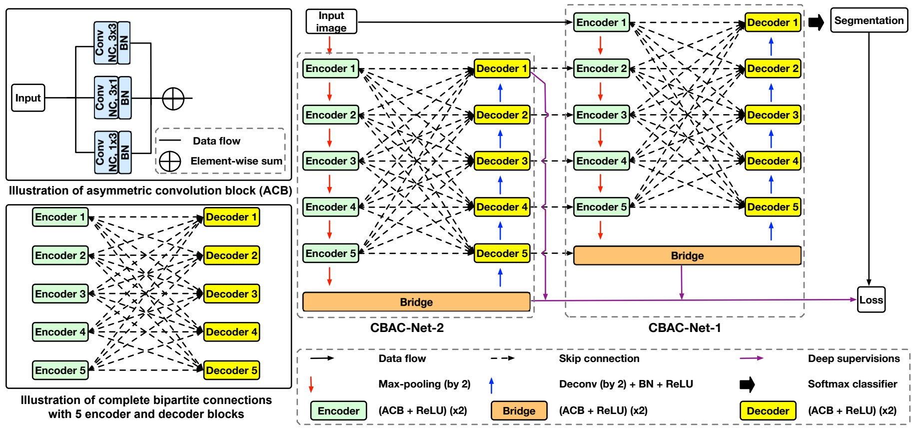

Fig. 2. An overview of kCBAC-Net with $k = 2$. The top-left box illustrates an asymmetric convolution block. The bottom-left box presents complete bipartite connections with 5 encoder and decoder blocks. The architecture is shown in 2D manner here; however, we also implement its 3D version. Best viewed in color.

(4) Auxiliary deep supervisions are applied to bolster the network on extracting discriminative features for tackling the missing/blurring border issue.

Our experiments on kCBAC-Net with two public datasets (2017 ISIC Skin Lesion segmentation (2D) [10] and MM-WHS CT (3D) [31]) and one in-house dataset (for lymph node segmentation in ultrasound images (2D)) show that kCBAC-Net outperforms state-of-the-art methods on these datasets.

2 Method

Fig. 2 shows the architecture of our kCBAC-Net. It contains three main components: (1) a k-complete-bipartite network (kCB-Net) that extracts and reuses rich multi-scale features, using complete bipartite connections (CBC) and k sequentially connected network modules; (2) kCBAC-Net that exploits more discriminative features via asymmetric convolutions; (3) auxiliary deep supervisions employed at layers far from the layer for computing loss functions that benefit effective network training and highly discriminative feature capturing.

2.1 k-Complete-Bipartite Network (kCB-Net)

A common way to reuse multi-scale features is to leverage skip connections. Using skip connections, multi-scale features extracted by the encoder can be utilized by the counterparts of the decoder. But, such skip connections may not ensure full exploitation and reuse of multi-scale features captured by the encoder. To further exploit multi-scale features, we design kCB-Net with complete bipartite connections and the structure of sequentially connected k network modules.

In an encoder-decoder structure (e.g., U-Net) with l encoder and decoder blocks, let x_i and y_i denote the outputs of the ith encoder and decoder blocks, respectively. Then x_i (or y_i) can be computed by a transformation function from the output x_{i-1} (or y_{i-1}) of the previous encoder (or decoder) block as $x_i = E_i(x_{i-1})$ and $y_i = D_i(y_{i-1})$, where E_i (or $D_i(x)$) can be a composite of operations such as Max Pooling (Deconvolution (Deconv)), Convolution (Conv), Batch Normalization (BN), Rectified Linear Unit (ReLU), etc.

Complete Bipartite Connections (CBC). To better reuse multi-scale features and enhance information flow among encoder and decoder blocks, the complete bipartite network (CB-Net) [6] used the idea of CBC, introducing one skip connection between every pair of encoder and decoder blocks, while HR-Net [23] gradually expands in deeper stages of encoder. Specifically,

$$y_i = D_i(E_1(x_0) \odot \cdots \odot E_i(x_{i-1}) \odot \cdots \odot E_l(x_{l-1}) \odot y_{i-1}),$$

where $\odot$ denotes concatenation operation. An example of CBC is given in the bottom-left box of Fig. 2. With a skip connection between each pair of encoder/decoder blocks, multi-scale features can be consolidated in the network hierarchy, and can be effectively reused at the decoder blocks. In addition, information flow between encoder and decoder blocks can be enlarged by the introduced skip connections. These characteristics are essential for the network to exploit and reuse multi-scale information.

***k*CB-Net Organization.** Another common deep learning based solution for handling multi-scale objects is a multi-path diagram. However, the design using the same network modules with shared weights may not be suitable for dealing with objects with significant size and appearance variations. In [7], kU-Net was proposed to handle multi-scale objects. Specifically, k U-Nets (without sharing parameters) working on different image scales are sequentially connected to extract multi-scale information. This structure has two compelling advantages. (1) The networks can view different-size regions of the same image with different scales, which facilitates processing multi-scale objects. (2) The multi-scale information extracted by one network can be propagated to the subsequent networks, which may assist further exploiting multi-scale information.

The organization of our kCB-Net is adopted from kU-Net [7]. First, an image of scale s_t ($t = 1, \ldots, k$) generated after $t-1$ max pooling layers is processed by a network module CB-Net-t. Second, CB-Net-($t-1$) takes every piece of information from CB-Net-t in the commensurate layers to well assimilate multi-scale information. We use this way to connect our CB-Net modules because such connections yield the best performance in our experiments. Note that the memory and time costs become larger when k increases. For example, on the lymph node dataset, for $k = 1, 2, 3, 4$, the F1 scores are 0.871, 0.897, 0.903, and 0.902 (see the F1 scores of CB-Net, kCB-Net, 3CB-Net, and 4CB-Net in Table 1), and the memory costs are 5.5GB, 11.8GB, 18GB, and 24.4GB, respectively. The performance saturates with larger k, and the costs increase largely. Considering the trade-off between performance and computation, we set $k = 2$ in our experiments.

Note that it is highly desirable to address the two major issues discussed in Section 1 simultaneously by the same model. Below we elaborate how to tackle the other issue, missing/blurring object borders, in our kCB-Net.

2.2 kCBAC-Net: Leveraging Asymmetric Convolutions

Recent studies attempting to tackle the missing/blurring border issue tended to focus on leveraging multiple modules to deal with the objects and boundaries separately, or encouraging the networks to learn richer and more distinguishable features. Such methods did not explore the possibility of identifying missing/blurring object borders by enhancing the network's capability via exploiting more discriminative features from the perspective of convolution operations.

An asymmetric convolution block (ACB) [11] adds additional $1 \times d$ and $d \times 1$ Conv layers on the basis of a standard $d \times d$ Conv layer, to strengthen the central skeleton parts of standard Conv, as:

$$ACB(x) = BN\left(Conv_{d\times d}(x)\right) + BN\left(Conv_{1\times d}(x)\right) + BN\left(Conv_{d\times 1}(x)\right)$$

where d is an integer and $+$ denotes element-wise sum. It is more sophisticated than inception-v2 [22] that decomposes only standard Conv. The added asymmetric convolutions can enhance the network's robustness with respect to rotational distortions, thus improving its capability to learn more distinguishable and discriminative features. This extension is important for us to address missing/blurring borders, as it can reduce ambiguities of object borders more effectively than standard convolution. Hence, we replace the standard convolution layers in our kCB-Net by ACBs, and call the resulted network kCBAC-Net.

The architecture of kCBAC-Net is shown in Fig. 2. The structure of CBAC-Net is adopted from CB-Net [6], with five encoder and decoder blocks. Each encoder or decoder block contains two ACBs, with each ACB followed by a ReLU. Starting from the first encoder block, we double the number of feature channels, and half the number of feature channels in the decoder blocks. The structure of kCBAC-Net uses the organization of kCB-Net (see Section 2.1).

2.3 Deep Supervision (DS)

Effectively training deep learning networks (especially deeper networks with a large number of parameters) is non-trivial, as the gradient can be gradually vanishing when it is propagated back to early layers. Hence, the layers that are far away from the layer computing loss functions may not be trained well [2]. Deep supervision [13] is a technique to address this issue. However, simply adding deep supervisions to the network may collapse task-relevant information at shallow layers, and thus hurt the performance [24].

An important question is: *At which layers should we add deep supervisions?* It was shown experimentally [1] that (1) adding deep supervisions at layers far away from the layer computing loss functions can effectively address performance deteriorate, and (2) the last layer of the encoder in U-Net is its farthest layer

Table 1. Comparison of segmentation results on the ultrasound lymph node dataset. CB-Net w/o CBC is a deeper U-Net with one more scale of encoding and decoding block. The best results are marked in bold (the same for the other tables in this paper).

Method	IoU	Precision	Recall	F1 Score
U-Net [20]	0.661	0.834	0.761	0.796
Zhang et al. [27]	0.810	0.901	0.889	0.895
kCB-Net	0.814	0.907	0.888	0.897
kCB-Net + ACB (kCBAC-Net)	0.819	0.900	0.900	0.900
kCBAC-Net + DS	**0.829**	**0.909**	**0.904**	**0.906**
Ablation Study				
CB-Net w/o CBC (deeper U-Net)	0.757	0.869	0.854	0.861
CB-Net	0.771	0.893	0.849	0.871
3CB-Net	0.824	0.908	0.898	0.903
4CB-Net	0.822	0.911	0.894	0.902

Table 2. Comparison of segmentation results on the 2017 ISIC Skin Lesion dataset.

Method	Jaccard Index	Dice	Sensitivity	Specificity
Yuan et al. [26]	0.765	0.849	0.825	0.975
Li et al. [15]	0.765	0.866	0.825	0.984
Lei et al. [14]	0.771	0.859	0.835	0.976
Mirikharaji et al. [17]	0.773	0.857	0.855	0.973
Xie et al. [25]	0.788	0.868	**0.884**	0.957
kCB-Net	0.784	0.881	0.827	0.984
kCBAC-Net	0.788	0.884	0.848	0.980
kCBAC-Net + DS	**0.794**	**0.887**	0.847	**0.984**

to the layer computing loss functions. To boost information flow in our network and extract more discriminative features to further handle the missing/blurring border issue, we employ three auxiliary deep supervisions in our kCBAC-Net (see Fig. 2). Specifically, we first add two auxiliary deep supervisions at the last layer of the Bridge block, which is the farthest layer of each CBAC-Net; then we add an auxiliary deep supervision to the last layer of Decoder 1 of CBAC-Net-2 to further boost information flow. We denote the auxiliary losses at the last layer of the Bridge and Decoder 1 blocks of CBACNet-2 as $Loss_{aux1}$ and $Loss_{aux2}$, respectively, and the auxiliary loss at the last layer of the Bridge of CBACNet-1 as $Loss_{aux3}$. Adding these three auxiliary losses to the main loss ($Loss_{main}$), our total loss ($Loss_{total}$) becomes:

$$Loss_{total} = \sum_{i=1}^{3} \lambda_i Loss_{auxi} + Loss_{main}$$

where λ_i is a balancing weight, and $Loss_{auxi}$ and $Loss_{main}$ are computed using the auxiliary output and main output of kCBAC-Net. Our kCBAC-Net employs DS in multiple modules rather than only in a certain one (e.g., U-Net++ [30]).

3 Experiments and Results

Two 2D Datasets and One 3D Dataset. (1) **The ultrasound lymph node segmentation dataset:** This in-house ultrasound lymph node dataset contains

Table 3. Comparison of segmentation results on the 2017 MM-WHS CT dataset. Note: "—" means that the corresponding results were not reported by that method.

Method	Metric	LV	RV	LA	RA	LV-myo	AO	PA	Mean
Payer et al. [19]	Dice	0.918	0.909	0.929	0.888	0.881	0.933	0.840	0.900
Dou et al. [12]	Dice	0.888	—	0.891	—	0.733	0.813	—	—
Chen et al. [4]	Dice	0.919	—	0.911	—	0.877	0.927	—	0.909
HFA-Net [29]	Dice	0.946	0.893	0.925	0.897	0.910	0.964	0.830	0.909
	Jaccard	0.898	0.810	0.861	0.816	0.836	0.930	0.722	0.839
	Hausdorff [voxel]	7.148	33.128	42.173	22.903	36.954	12.075	37.845	27.461
	ADB [voxel]	0.076	0.562	0.210	0.334	0.225	0.103	1.685	0.456
*k*CB-Net	Dice	0.950	0.894	0.929	**0.913**	0.916	0.968	0.837	0.915
	Jaccard	0.905	0.811	0.869	**0.841**	0.845	0.938	0.730	0.849
	Hausdorff [voxel]	8.887	**12.639**	14.630	24.229	10.642	20.713	33.663	17.915
	ADB [voxel]	0.074	0.338	0.205	**0.244**	0.122	0.074	1.467	0.361
*k*CBAC-Net	Dice	0.951	0.894	**0.939**	0.909	0.920	0.966	**0.843**	0.917
	Jaccard	0.907	0.812	**0.886**	0.834	0.851	0.934	**0.742**	0.852
	Hausdorff [voxel]	5.534	13.315	15.296	**14.670**	7.649	17.451	**30.758**	14.953
	ADB [voxel]	0.075	0.339	**0.139**	0.268	**0.115**	0.068	**1.391**	0.342
***k*CBAC-Net + DS**	Dice	**0.951**	**0.902**	0.938	0.911	**0.922**	**0.974**	0.837	**0.919**
	Jaccard	**0.907**	**0.825**	0.883	0.838	**0.855**	**0.949**	0.734	**0.856**
	Hausdorff [voxel]	**5.500**	14.940	**12.403**	15.081	**7.337**	**6.848**	32.499	**13.516**
	ADB [voxel]	**0.074**	**0.285**	0.163	0.248	0.119	**0.059**	1.403	**0.336**

137 training and 100 test images. The task is to segment lymph nodes in 2D ultrasound images. (2) **The 2017 ISIC Skin Lesion segmentation dataset:** This public dataset [10] contains 2000 training, 150 validation, and 600 test images. The task is to segment lesion boundaries in 2D dermoscopic images. Following [25,26], we resize the images to 224×224, and apply a dual-threshold method to generate the final results. (3) **The 2017 MM-WHS CT dataset:** This public dataset [31] contains 20 unpaired 3D CT images. Similar to [29], the dataset is randomly split into 16 images and 4 images for training and test. The task is to segment seven cardiac structures: the left/right ventricle blood cavity (LV/RV), left/right atrium blood cavity (LA/RA), myocardium of the left ventricle (LV-myo), ascending aorta (AO), and pulmonary artery (PA).

Implementation Details. Our new network is implemented with TensorFlow, trained on an Nvidia Tesla V100 Graphics Card with 32GB GPU memory using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e{-}10$). The "poly" learning rate policy, $L_r \times \left(1 - \frac{iter}{\#iter}\right)$, is applied, with the initial learning rate $= 5e - 4$, and the maximum number of iterations is 140k for the ultrasound lymph node and 2017 ISIC Skin Lesion datasets, and 180k for the 2017 MM-WHS CT dataset (using about 26, 21, and 92 hours, respectively, for training). Standard data augmentation (e.g., image crop, flip, and rotation in 90, 180, and 270 degrees) is utilized to reduce overfitting. The balancing weights λ_1, λ_2, and λ_3 are set as 0.25, 0.50, and 0.75 for the ultrasound lymph node and 2017 MM-WHS CT datasets and 1.0, 1.0, and 1.0 for the 2017 ISIC Skin Lesion dataset.

Quantitative Results. Table 1 gives quantitative comparison of our method with U-Net [20] and the best-known method [27] on the ultrasound lymph node dataset. First, our *k*CB-Net already obtains a new state-of-the-art performance. We attribute this to the compelling advantages of CBC and the structure of k connected network modules (without sharing parameters) on exploiting multi-scale features. Second, by leveraging ACB and DS, our method further improves the state-of-the-art performance (1.9% IoU and 1.1% F1 score over Zhang et al. [27]), showing the capability of ACB and the employed DS for capturing more distinguishable and discriminative features to handle the missing/blurring border issue. Results on the 2017 ISIC Skin Lesion dataset are given in Table 2. We compare our results with five recent skin lesion segmentation methods, including the 2017 ISIC Challenge winner [26], dense deconvolution network [15], dense convolution U-Net [14], FCN with a star shape prior [17], and a state-of-the-art method [25]. First, kCB-Net attains comparable Jaccard index score (the only evaluation metric used by the challenge organizers to rank competitors) as the state-of-the-art method [25]. This indicates the effectiveness of CBC and the structure of k connected network modules on exploiting multi-scale information. Second, by utilizing ACB and DS, our method achieves the highest Jaccard index, Dice, and specificity, though its sensitivity is slightly lower than that of some other methods. We attribute the improvement to the effects of ACB and DS on handling missing/blurring borders. To further show the effectiveness of our method on 3D images, we experiment with the 2017 MM-WHS CT dataset, and compare with four representative methods. As shown in Table 3, our method yields better results than the other methods on all the evaluation measures and achieves new state-of-the-art performance. Qualitative results showing our strong capability of delineating missing/blurring borders and handling multi-scale objects are given in Fig. 3.

Ablation Study. We conduct ablation study using the ultrasound lymph node dataset to examine the importance of CBC and the structure of k connected network modules on exploiting multi-scale features. As shown in Table 1, (1) when removing CBC from CB-Net, the F1 score drops by 1%; (2) when experimenting with CB-Net (i.e., setting $k = 1$), the F1 score drops by 2.6%. These observations demonstrate that CBC and the structure of k connected network modules indeed play a meaningful role in exploiting multi-scale features.

4 Conclusions

In this paper, we presented a new *deeply supervised k-complete-bipartite network with asymmetric convolutions* (kCBAC-Net) for medical image segmentation. Our kCBAC-Net leverages a generalized complete bipartite network and the structure of k connected network modules to exploit and reuse multi-scale features for dealing with the multi-scale object issue. To further enhance the capability of our model on handling the missing/blurring border issue, ACB and auxiliary deep supervisions are employed. Experiments on two public datasets and one in-house dataset demonstrated the effectiveness of our kCBAC-Net.

Acknowledgement. This research was supported in part by NSF grants IIS-1455886, CCF-1617735, CNS-1629914, and IIS-1955395.

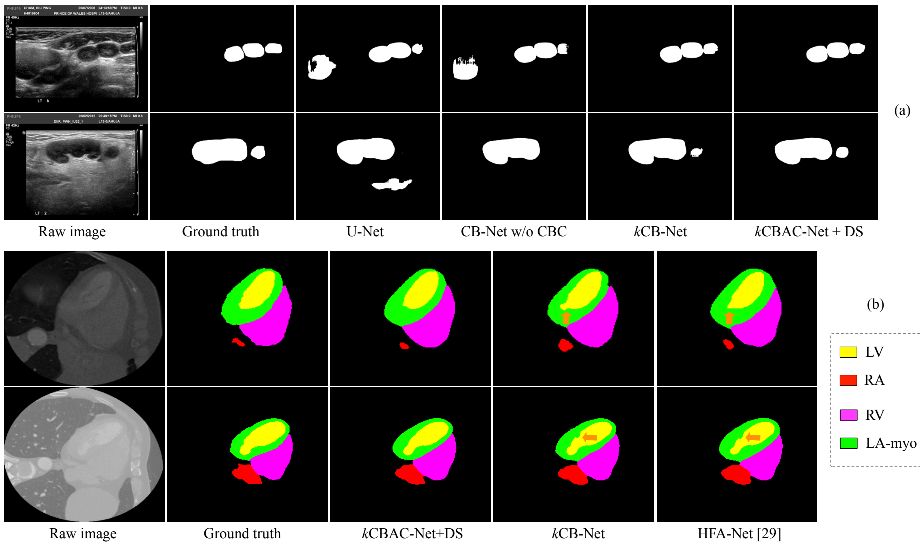

Fig. 3. Some visual qualitative results on the ultrasound lymph node dataset (a) and 2017 MM-WHS CT dataset (b), showing the capability of our new method (kCBAC-Net with deep supervisions) on handling the issues of multi-scale objects and missing/blurring borders. Magenta arrows mark some errors.

References

1. Ando, T., Hotta, K.: Cell image segmentation by feature random enhancement module. arXiv preprint arXiv:2101.07983 (2021)
2. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks **5**(2), 157–166 (1994)
3. Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H.: A methodological approach to the classification of dermoscopy images. Computerized Medical Imaging and Graphics **31**(6), 362–373 (2007)
4. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Trans. on Medical Imaging **39**(7), 2494–2505 (2020)
5. Chen, H., Qi, X., Yu, L., Heng, P.A.: DCAN: Deep contour-aware networks for accurate gland segmentation. In: CVPR. pp. 2487–2496 (2016)
6. Chen, J., Banerjee, S., Grama, A., Scheirer, W.J., Chen, D.Z.: Neuron segmentation using deep complete bipartite networks. In: MICCAI. pp. 21–29 (2017)
7. Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z.: Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: NIPS. pp. 3036–3044 (2016)
8. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)
9. Chen, L.C., Yang, Y., Wang, J., Xu, W., Yuille, A.L.: Attention to scale: Scale-aware semantic image segmentation. In: CVPR. pp. 3640–3649 (2016)
10. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward

melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: IEEE, ISBI. pp. 168–172 (2018)
11. Ding, X., Guo, Y., Ding, G., Han, J.: ACNet: Strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks. In: ICCV. pp. 1911–1920 (2019)
12. Dou, Q., Ouyang, C., Chen, C., Chen, H., Heng, P.A.: Unsupervised cross-modality domain adaptation of ConvNets for biomedical image segmentations with adversarial loss. In: IJCAI. pp. 691–697 (2018)
13. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: AISTATS. pp. 562–570 (2015)
14. Lei, B., Xia, Z., Jiang, F., Jiang, X., Ge, Z., Xu, Y., Qin, J., Chen, S., Wang, T., Wang, S.: Skin lesion segmentation via generative adversarial networks with dual discriminators. Medical Image Analysis **64**, 101716 (2020)
15. Li, H., He, X., Zhou, F., Yu, Z., Ni, D., Chen, S., Wang, T., Lei, B.: Dense deconvolutional network for skin lesion segmentation. IEEE Journal of Biomedical and Health Informatics **23**(2), 527–537 (2019)
16. Liang, P., Chen, J., Zheng, H., Yang, L., Zhang, Y., Chen, D.Z.: Cascade decoder: A universal decoding method for biomedical image segmentation. In: IEEE, ISBI. pp. 339–342 (2019)
17. Mirikharaji, Z., Hamarneh, G.: Star shape prior in fully convolutional networks for skin lesion segmentation. In: MICCAI. pp. 737–745 (2018)
18. Pace, D.F., Dalca, A.V., Geva, T., Powell, A.J., Moghari, M.H., Golland, P.: Interactive whole-heart segmentation in congenital heart disease. In: MICCAI. pp. 80–88 (2015)
19. Payer, C., Štern, D., Bischof, H., Urschler, M.: Multi-label whole heart segmentation using CNNs and anatomical label configurations. In: International Workshop, STACOM. pp. 190–198 (2017)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
21. Sun, C., Paluri, M., Collobert, R., Nevatia, R., Bourdev, L.: ProNet: Learning to propose object-specific boxes for cascaded neural networks. In: CVPR. pp. 3485–3493 (2016)
22. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR. pp. 2818–2826 (2016)
23. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al.: Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
24. Wang, Y., Ni, Z., Song, S., Yang, L., Huang, G.: Revisiting locally supervised learning: An alternative to end-to-end training. In: ICLR (2021)
25. Xie, Y., Zhang, J., Lu, H., Shen, C., Xia, Y.: SESV: Accurate medical image segmentation by predicting and correcting errors. IEEE Transactions on Medical Imaging **40**(1), 286–296 (2021)
26. Yuan, Y., Lo, Y.C.: Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks. IEEE Journal of Biomedical and Health Informatics **23**(2), 519–526 (2019)
27. Zhang, Y., Ying, M.T., Chen, D.Z.: Decompose-and-integrate learning for multi-class segmentation in medical images. In: MICCAI. pp. 641–650 (2019)
28. Zhang, Y., Ying, M.T., Yang, L., Ahuja, A.T., Chen, D.Z.: Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images. In: IEEE, BIBM. pp. 443–448 (2016)

29. Zheng, H., Yang, L., Han, J., Zhang, Y., Liang, P., Zhao, Z., Wang, C., Chen, D.Z.: HFA-Net: 3D cardiovascular image segmentation with asymmetrical pooling and content-aware fusion. In: MICCAI. pp. 759–767 (2019)
30. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: DLMIA/ML-CDS. pp. 3–11 (2018)
31. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Medical Image Analysis **31**, 77–87 (2016)