A Second-Order Subregion Pooling Network for Breast Lesion Segmentation in Ultrasound

Lei Zhu[1,5], Rongzhen Chen[2], Huazhu Fu[3], Cong Xie[2], Liansheng Wang[2(✉)], Liang Wan[1,4], and Pheng-Ann Heng[5]

[1] College of Intelligence and Computing, Tianjin University, Tianjin, China
[2] Department of Computer Science at School of Informatics, Xiamen University, Xiamen, China
lswang@xmu.edu.cn
[3] Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, UAE
[4] Medical College of Tianjin University, Tianjin , China
[5] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong

Abstract. Breast lesion segmentation in ultrasound images is a fundamental task for clinical diagnosis of the disease. Unfortunately, existing methods mainly rely on the entire image to learn the global context information, which neglects the spatial relation and results in ambiguity in the segmentation results. In this paper, we propose a novel second-order subregion pooling network (S^2P-Net) for boosting the breast lesion segmentation in ultrasound images. In our S^2P-Net, an attention-weighted subregion pooling (ASP) module is introduced in each encoder block of segmentation network to refine features by aggregating global features from the whole image and local information of subregions. Moreover, in each subregion, a guided multi-dimension second-order pooling (GMP) block is designed to leverage additional guidance information and multiple feature dimensions to learn powerful second-order covariance representations. Experimental results on two datasets demonstrate that our proposed S^2P-Net outperforms state-of-the-art methods.

Keywords: Ultrasound image · Breast lesion segmentation · Second-order subregion pooling

1 Introduction

Breast cancer is a leading cause of women death in the world [17]. Ultrasound imaging is an useful tool for breast cancer detection in clinical due to its versatility, safety and high sensitivity [18]. Segmenting breast lesions from ultrasound imaging is an important step of computer-aided diagnosis systems, which

L. Zhu and R. Chen—Joint first authors of this work.

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12266, pp. 160–170, 2020.
https://doi.org/10.1007/978-3-030-59725-2_16

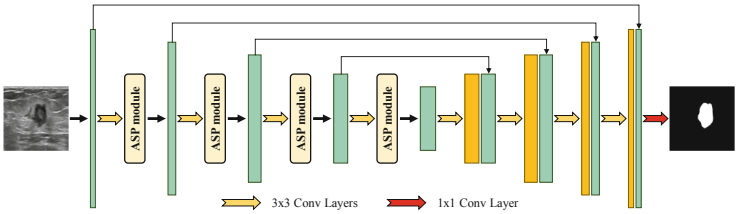

Fig. 1. Schematic illustration of the proposed breast ultrasound lesion segmentation network. Please see Fig. 2 for the ASP module. Best viewed in color. (Color figure online)

assists radiologists in the ultrasound-based breast cancer diagnosis [22]. However, accurate segmentation in ultrasound images is challenging due to the missing/ambiguous boundary, the inhomogeneous intensity distribution of breast ultrasound image, the similar visual appearance between lesions and non-lesion backgrounds, as well as the irregular shapes and complex variants of breast lesions [10,22].

Early methods [2,16,20] mainly examine hand-crafted features for inferring breast lesion boundaries in ultrasound images. Madabhushi [14] incorporated empirical domain knowledge from radiologists and low-level image features (e.g., texture, intensity and directional gradient) into a deformable shape-based model for the segmentation. Gómez-Flores et al. [6] segmented breast lesions by analyzing textures of ultrasound images. Later, several methods based on convolution neural networks (CNNs) [10,21] have achieved superior breast lesion performance than early methods by learning discriminative features from annotated images. Yap et al. [22] investigated the patch-based LeNet, U-Net, and transfer learning with a pre-trained FCN-AlecNet for segmenting breast ultrasound lesions. Lei et al. [10] designed a boundary regularized encoder-decoder network for predicting segmentation maps. Unfortunately, due to diverse ambiguous boundaries and complex shape variances, existing CNN-based methods mainly rely on the entire ultrasound image to learn the global context information, which neglects the spatial relation and results in ambiguity in the segmentation results.

In this work, we develop a second-order subregion pooling network (S^2P-Net) for boosting breast lesion segmentation performance by aggregating the multi-context information from both the whole image and multiple subregions. The contributions of this work could be summarized as: **1)** A new second-order subregion pooling network is proposed for boosting breast lesion segmentation in ultrasound images. **2)** The ASP module is utilized to attentively aggregate global and multiple subregion representations for inferring breast lesion regions. **3)** The GMP block is designed to leverage additional guidance information and all three feature dimensions for modeling higher-order statistics for more discriminative image representations. **4)** Moreover, experimental results on two datasets

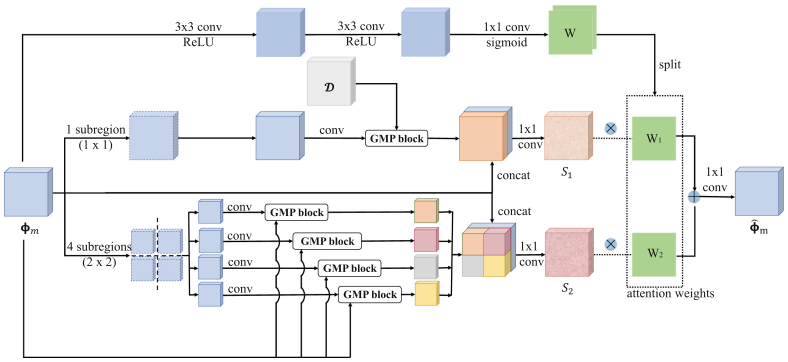

Fig. 2. Schematic illustration of the ASP module at the m-th CNN layer. $\mathbf{\Phi}_m$ is the feature map at the m-th CNN layer. Please see Fig. 3 for the details of the GMP block. Note that W has 2 channels. The first channel is W_1 while the second channel is W_2. $\mathcal{D}$ is an unit matrix (all elements are 1). Best viewed in color. (Color figure online)

demonstrate that our network sets a new state-of-the-art performance on breast lesion segmentation in ultrasound images.

2 Proposed Method

Figure 1 illustrates the architecture of the proposed breast lesion segmentation network, which takes a breast lesion ultrasound image as the input and produces a segmentation resultant image of breast lesions. In our method, the U-Net architecture [15] is utilized as the backbone, which consists of the encoder and decoder paths. Each CNN block of the encoder contains two 3×3 convolutional layers and a max-pooling layer (stride $= 2$), and decoder blocks have two 3×3 convolutional layers and a upsampling operation. After each CNN block of the encoder, we introduce an attention-weighted subregion pooling (ASP) module (see Fig. 2) to refine the CNN features by learning second-order statistics from multiple subregions. In each subregion, a guided multi-dimension second-order pooling (GMP) block (see Fig. 3) is designed to leverage guidance information for learning second-order covariance matrices. Then, we iteratively merge two adjacent layers by first upsampling the low-resolution feature map and then applying two 3×3 convolutional layers. Finally, the feature map with the largest spatial resolution is used to predict the final segmentation result.

2.1 Attention-Weighted Subregion Pooling (ASP) Module

Existing methods mainly rely on the entire ultrasound image to learn a global information for inferring breast ultrasound lesions, which loses the spatial relations and tends to contain non-lesion regions or loss parts of breast lesions in the segmentation result. We develop an ASP module (see Fig. 2) to fuse the

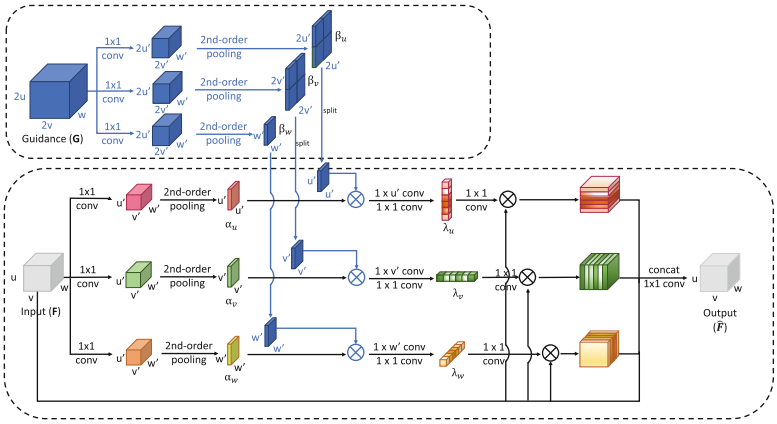

Fig. 3. Schematic illustration of the GMP block. Best viewed in color. (Color figure online)

global second-order features from the whole image and the local second-order features from multiple subregions together. Intuitively, image subregions have less non-breast-lesion details than the whole input image, and thus reduce the interference from non-lesion regions, resulting in a superior segmentation performance via our ASP modules.

Figure 2 shows the detailed architecture of the ASP module at m-th CNN layer. It refines the feature map (denoted as $\mathbf{\Phi}_m$) at m-th CNN layer by aggregating features learned from different subregion separation branches. The first branch of the ASP module uses a 1×1 separation on $\mathbf{\Phi}_m$. It passes the whole $\mathbf{\Phi}_m$ to a GMP block to learn a second-order feature map, concatenates the resultant features with Φ_m, and uses a 1×1 convolutional layer to produce a new feature map (denoted as S_1). The second branch separates the input $\mathbf{\Phi}_m$ into 4 (2×2) subregions and then extracts second-order features from each subregion by feeding its feature map into a GMP block. The resultant second-order features from the four subregions are then combined together to form a feature map, which is merged with $\mathbf{\Phi}_m$ by using a concatenation and a 1×1 convolution to produce a feature map (denoted as S_2). To fuse the two feature maps, we generate an attention map (denoted as W; 2 channels) by employing two successive convolutional layers (with 3×3 kernels) followed by a ReLU non-linear operation [9,26,27], and the third layer (with 1×1 kernels and a sigmoid activation layer). Finally, we split W into two maps (W_1 and W_2), multiply them with S_1 and S_2, add two multiplication results, and apply a 1×1 convolutional layer on the addition result to produce the output feature map (denoted as $\hat{\mathbf{\Phi}}_m$) of the ASP module. Mathematically, we can compute $\hat{\mathbf{\Phi}}_m$ as:

$$\hat{\mathbf{\Phi}}_m = f_{conv}(W_1 \times S_1 + W_2 \times S_2) \ , \tag{1}$$

where f_{conv} is the 1×1 convolutional parameters.

2.2 Guided Multi-dimension Second-Order Pooling (GMP) Block

Breast lesion segmentation in ultrasound images is challenging due to the ambiguous boundary, the inhomogeneous intensity distribution, and the complex variants of breast lesions. CNN features in existing methods almost are first-order, and tend to produce unsatisfactory results of the challenging breast lesion segmentation in ultrasound. To enhance the segmentation accuracy, our work introduces the second-order information into an end-to-end breast lesion segmentation network. In this regard, we devise a novel guided multi-dimension second-order pooling (GMP) block to leverage an entire feature map as a guidance to provide more spatial information from other subregions for helping to learn second-order covariances in each subregion of the entire features. Moreover, instead of considering only channel dimension, GMP further augments the resulting second-order covariances by exploring element statistical dependencies from all three feature dimensions. Hence, GMP enables our method to better identify breast lesions than original second-order pooling [5]; see the ablation study results in Table 3.

Given a 3D CNN feature map $\mathbf{F}$ (size: $u \times v \times w$), original second-order pooling block [5] learns a $w \times w$ region covariance matrix from $\mathbf{F}$ by reformulating $\mathbf{F}$ as a set of points $\{P_{i,j},\ 1 \leq i \leq u; 1 \leq j \leq v\}$, and each point $P_{i,j}$ is a vector with w elements. The resulting covariance matrix represents the pair-wise channel correlations among all w elements, and has a clear physical meaning, i.e., its k-row ($\{1 \leq k \leq w\}$) indicates the statistical dependencies of the k-th element (channel) with all w elements (channels). We argue that only exploring element dependencies along the channel dimension suffers from a limited capability to capture the second-order statistics, due to ignoring other two feature dimensions. On the other hand, as presented in classical guided filters [8,23], incorporating additional information from a guidance image generates a better filtering behavior.

In this regard, our GMP block leverages all three dimensions to learn the second-order covariance matrix by considering additional information from another feature map (we call it "guidance feature map"). Before going into details of GMP block, we first present which information is employed as the guidance feature map. As shown in Fig. 2, our ASP module uses two branches to divide the input features ($\mathbf{\Phi}_m$). In the 1×1 subregion branch, we set the guidance feature map (denoted as $\mathcal{D}$ in Fig. 2) as an all-1 3D feature map (all elements are 1) to bypass the filtering, while the guidance feature map for all the four subregions in the 2×2 subregion branch is set as $\mathbf{\Phi}_m$. The reason behind is that $\mathbf{\Phi}_m$ has information of all the four subregions, and thus can provide additional spatial relations in other three subregions for each GMP block to better learn second-order statistics.

Figure 3 shows the architecture of our GMP block, which takes the feature map $\mathbf{F}$ (size: $u \times v \times w$) and guidance feature map $\mathbf{G}$ (size: $2u \times 2v \times w$) as two inputs and produces a refined feature map $\hat{\mathbf{F}}$. Specifically, we first apply a 1×1 convolution layer on $\mathbf{F}$ to reduce its feature dimension size into $u^{'} \times v^{'} \times w^{'}$ ($u^{'} < u; v^{'} < v; w^{'} < w$) for saving the computational cost, and take three copies of the

resultant features for learning three region covariance representations along u, v and w dimensions: a $u^{'} \times u^{'}$ covariance matrix (α_u), a $v^{'} \times v^{'}$ covariance matrix (α_v), and a $w^{'} \times w^{'}$ covariance matrix (α_w). To do so, apart from considering the w dimension in original second-order pooling [5], we reformulate $\mathbf{F}$ into $v^{'} \times w^{'}$ points $\{Q_{i,j},\ where\ 1 \leq i \leq v^{'}; 1 \leq j \leq w^{'}\}$, and the vector $Q_{i,j}$ has $u^{'}$ elements. After that, we learn a $u^{'} \times u^{'}$ covariance matrix to capture the statistical dependencies among all $u^{'}$ elements in the vector $Q_{i,j}$. Moreover, we decompose $\mathbf{F}$ into $u^{'} \times w^{'}$ vectors $\{R_{i,j},\ where\ 1 \leq i \leq u^{'}; 1 \leq j \leq w^{'}\}$, and each $R_{i,j}$ has $v^{'}$ elements. Then, we learn a $v^{'} \times v^{'}$ covariance matrix to compute the second-order statistics of all $v^{'}$ elements of $R_{i,j}$.

Similarly, we can learn three region covariance matrices from $\mathbf{G}$ along its three dimensions. To this end, we first reduce the dimension of $\mathbf{G}$ to $2u^{'} \times 2v^{'} \times w'$ by applying a 1×1 convolutional layer on $\mathbf{G}$ and produce three covariance matrices (i.e., a $2u^{'} \times 2u^{'}$ covariance matrix (β_u), a $2v^{'} \times 2v^{'}$ covariance matrix (β_v), and a $w^{'} \times w^{'}$ covariance matrix (β_w)) from three copies of the resized feature map. Once obtaining three covariance matrices from $\mathbf{F}$ and $\mathbf{G}$ respectively, we multiply them (e.g., $\alpha_w \times \beta_w$) together to integrate the guidance features for learning guided second-order covariance matrices. Note that β_u ($2u^{'} \times 2u^{'}$) and β_v ($2v^{'} \times 2v^{'}$) have not the same size of α_u ($u^{'} \times u^{'}$) and α_v ($v^{'} \times v^{'}$). Hence, according to the 2×2 subregion partition manner in Fig. 2, we similarly split β_u into four $u^{'} \times u^{'}$ subregion matrices, select the corresponding one (subregion) with the size of $u^{'} \times u^{'}$, and multiply it with α_u. The same operations are applied to β_v and α_v. After that, following [5], we use a row-wise convolution, a 1×1 convolution, and a sigmoid activation function on each resultant covariance matrix to produce three statistic weight vector (λ_u, λ_v, and λ_w), which are then multiplied with the input $\mathbf{F}$ for scaling different channels, in order to emphasize useful channel and suppress bad channel information for detecting breast lesion boundaries. Finally, we concatenate the scaled feature maps from three dimensions and the input $\mathbf{F}$, followed by a 1×1 convolution, to generate the output feature map ($\hat{\mathbf{F}}$) of the developed GMP block.

Implementation Details. We train our network from scratch and all the network parameters are initialized by a normal distribution. All the training images are randomly rotated, cropped, and horizontally flipped for data augmentation. The focal loss [12] is employed to compute the total loss of our network, and we utilize the Adam optimizer to minimize our total loss for training the whole framework with $15,000$ iterations. The learning rate is initialized as 0.0001 and then reduced to 0.00001 at $3,000$ iterations. We implement our network on Keras and train it on two GPUs with a mini-batch size of 4. Our method In our experiments, we empirically set $u^{'}$=24, $v^{'}$=24, and $w^{'}$=128,

3 Experiments

Datasets. We used two breast ultrasound image datasets to evaluate the effectiveness of the proposed network. The first one is a public dataset, BUSI [1],

Table 1. The results (mean ± variance) of different methods on our dataset.

Method	#paras (M)	Dice ↑	ADB ↓	Jaccard ↑	Precision ↑	Recall ↑
FCN [13]	513	0.8289±0.0007	7.1340±1.4860	0.7461±0.0008	0.8600±0.0005	0.8465±0.0002
U-Net [15]	30	0.7907±0.0003	14.428± 3.2128	0.7020±0.0004	0.8212± 0.0004	0.8168±0.0003
U-Net++ [25]	35	0.7933±0.0003	13.9142±3.0879	0.7036±0.0005	0.8300±0.0004	0.8101±0.0002
FPN [11]	52	0.8336±0.0007	7.6704±4.8599	0.7548±0.0009	0.8742±0.0009	0.8438±0.0001
DeeplabV3+ [4]	159	0.8180±0.0007	8.5677±3.4262	0.7318±0.0009	0.8533±0.0008	0.8271±0.0002
GSoP [5]	48	0.8361±0.0005	8.4699±5.6814	0.7540±0.0006	0.8640±0.0012	0.8473±0.0004
ConvEDNet [10]	310	0.8428±0.0003	6.2834±1.3013	0.7652±0.0004	0.8847±0.0005	0.8485±0.0003
Our method	96	**0.8967±0.0001**	**5.5967±0.7930**	**0.8311±0.0001**	**0.9072±0.0003**	**0.9047±0.0002**

Table 2. The results (mean ± variance) of different methods on the BUSI dataset.

Method	#paras (M)	Dice ↑	ADB ↓	Jaccard ↑	Precision ↑	Recall ↑
FCN [13]	513	0.8259±0.0120	16.4322±0.7647	0.7412±0.0089	0.8551± 0.0093	0.8435±0.0173
U-Net [15]	30	0.7660±0.0041	28.8543±4.0009	0.6710±0.0058	0.8326±0.0148	0.7799±0.0060
U-Net++ [25]	35	0.7621±0.0053	33.0030±2.1888	0.6664±0.0042	0.8267±0.0060	0.7802±0.0059
FPN [11]	52	0.7993±0.0033	22.5294±1.4860	0.7012±0.0083	0.8352±0.0171	0.8254±0.0154
DeeplabV3+ [4]	159	0.8210±0.0056	14.6643±1.3872	0.7321±0.0024	0.8648±0.0113	0.8266±0.0100
GSoP [5]	48	0.8309±0.0062	13.0799±1.0415	0.7455±0.0050	0.8745±0.0109	0.8311±0.0044
ConvEDNet [10]	310	0.8254±0.0015	14.6643±1.8361	0.7386±0.0056	0.8408±0.0136	0.8516±0.0109
Our method	96	**0.8470±0.0094**	**11.1760±0.9436**	**0.7639±0.0107**	**0.8762±0.0081**	**0.8551±0.0100**

from the Baheya Hospital for Early Detection & Treatment of Womenś Cancer (Cairo, Egypt). It has a total of 780 tumor images from 600 female patients (25–75 years old). Second, we collected 632 breast ultrasound images from Shenzhen People's Hospital to build the second dataset for evaluation, and informed consent forms were obtained from all patients. We invited experienced clinicians to manually annotate the breast lesion regions of each image. Moreover, we further adopt the five-folder cross-validation to statistically test different segmentation methods on the two datasets.

Evaluation Metrics. We employ widely-used segmentation metrics for quantitatively comparing different methods. They are Dice Similarity Coefficient (Dice), Average Distance of Boundaries (ADB, in pixel), Jaccard, Precision, and Recall; see [3,7,19,24] for details of these five metrics. A better segmentation result shall have smaller ADB and larger values for all other four metrics.

3.1 Segmentation Performance

We validate our segmentation network by comparing with seven state-of-the-art methods, including the fully convolutional network (FCN) [13], U-Net [15], U-Net++ [25], feature pyramid network (FPN) [11], DeeplabV3+ [4], a very recent second-order method (i.e., GSoP [5]), and a recent ultrasound breast lesion segmentation method (i.e., ConvEDNet [10]). For a fair comparison, we obtain the segmentation results of all the competitors by exploiting its public implementations or implementing them by ourselves, and fine-tuning the network training parameters for best segmentation results.

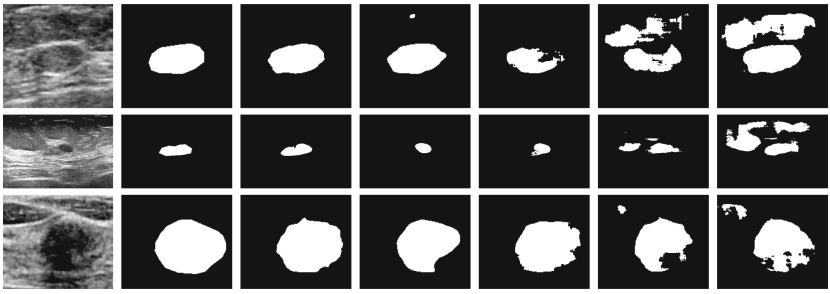

Fig. 4. Comparing segmentation maps produced by different methods. (a) Breast ultrasound lesion images. (b) Ground truths (denoted as GT). (c)–(g): Segmentation results produced by our method, ConvEDNet [10] (denoted as CEDNet), GSoP [5], U-Net++ [25], and U-Net [15], respectively.

Table 3. Metric results of different components on the BUSI dataset.

Method	Dice ↑	ADB ↓	Jaccard ↑	Precision ↑	Recall ↑
w/o-GMP	$0.801 + 0.0024$	$21.1114 + 3.5251$	$0.7113 + 0.0034$	$0.8436 + 0.0143$	$0.8199 + 0.0116$
w/o-subregions	$0.8315 + 0.0023$	$13.7236 + 1.8757$	$0.7463 + 0.0036$	$0.8635 + 0.0055$	$0.8397 + 0.0080$
Ours-channelRC	$0.8341 + 0.0151$	$12.9171 + 2.4227$	$0.7488 + 0.0170$	$0.8672 + 0.0145$	$0.8422 + 0.0171$
w/o-guidance	$0.8410 + 0.0098$	$14.2395 + 3.4063$	$0.7547 + 0.0148$	$0.8709 + 0.0193$	$0.8485 + 0.0186$
Our method	$\mathbf{0.8470 \pm 0.0094}$	$\mathbf{11.1760 \pm 0.9436}$	$\mathbf{0.7639 \pm 0.0107}$	$\mathbf{0.8762 \pm 0.0081}$	$\mathbf{0.8551 \pm 0.0100}$

Quantitative Comparisons. Tables 1 and 2 report the metric results of different segmentation methods on our collected dataset and BUSI dataset, respectively. Apparently, our method consistently and stably has the superior performances of the mean and variance values of all the five metrics over all the competitors. It indicates that our method has more accurate segmentation results than all the competitors.

Visual Comparisons. Figure 4 visualizes the segmentation results produced by different methods. Apparently, the compared methods tend to neglect some details of the breast lesion regions or include other non-lesion regions into their predicted segmentation results, while our method can more accurately detect the blurry breast lesion boundaries for the input images (1st row), and can better detect the whole breast lesion regions for the images (last two rows) with multiple intensity distributions. The superior segmentation results of our method show that our subregion based second-order features have more discriminative capabilities in inferring breast lesion regions from ultrasound images. Apparently, our network is more complex than standard UNet, but the GMP blocks first downsample the resolutions of the input feature map and the guidance features a lot for computing second-order statistics. Thus our network does not increase the inference time too much. As shown in Table 1, our method only has about a half model size of the Deeplabv3+, but achieves a better breast lesion segmentation performance.

3.2 Ablation Study

We conduct the ablation study experiments to verify the major components in our network design. Here, we consider four baseline networks, and report their quantitative results on the BUSI dataset in Table 3. (i) The first baseline (denoted as "w/o-GMP") removes all the GMP modules (see Fig. 3) from our network, while (ii) the second baseline (denoted as 'w/o-subregions') removes the 2×2 subregions (the second branch of Fig. 2) from our network, meaning that we do not model the second-order representation from any subregion. (iii) The third baseline (denoted as "Ours-1channelRC") is to replace our GMP block (see Fig. 3) with the original second-order block [5], which uses only the feature channel dimension. (iv) The last baseline (denoted as "w/o-guidance") is to remove the guidance features from our GMP block (see Fig. 3).

Table 3 shows the comparison results. Apparently, our method has superior metric results than "w/o-GMP" and "w/o-subregions", which demonstrates that both ASP module and GMP block have the contributions to the superior segmentation results of our method. Our method can also more accurately segment breast lesions than "Ours-channelRC" and "w/o-guidance", showing that both guidance feature and the multiple feature dimension help to learn more powerful second-order features.

4 Conclusion

This paper presents a second-order subregion network for the breast lesion segmentation from an ultrasound image by harnessing second-order statistics from multiple feature subregions. Our key idea is to develop an ASP module at each CNN layer to aggregate global features from the whole image and local high-order features from multiple subregions, and a GMP block in each subregion to leverage additional guidance information and all the three feature dimensions for learning powerful second-order covariance features. Experiments on two datasets demonstrate that our method clearly outperforms state-of-the-art methods. In addition, the proposed segmentation network has the potential for other similar medical image segmentation tasks, e.g., the prostate segmentation. Although our method obtained best performance in two datasets (i.e., the Dice score is 0.847 for BUSI, and 0.8967 for our dataset), there is a large room to further improve the segmentation accuracy.

Acknowledgements. This work was supported by National Natural Science Foundation of China (Project No. 61902275, 61671399), the Fundamental Research Funds for the Central Universities (Grant No. 20720190012), and Hong Kong Innovation and Technology Fund (GHP/002/13SZ and GHP/003/11SZ). We thank Yunzhu Wu for her efforts of data collection and annotations.

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data Brief **28**, 104863 (2020)

2. Boukerroui, D., Basset, O., Guerin, N., Baskurt, A.: Multiresolution texture based adaptive clustering algorithm for breast lesion segmentation. Eur. J. Ultrasound **8**(2), 135–144 (1998)
3. Chang, H.H., Zhuang, A.H., Valentino, D.J., Chu, W.C.: Performance measure characterization for evaluating neuroimage segmentation algorithms. Neuroimage **47**(1), 122–135 (2009)
4. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV, pp. 801–818 (2018)
5. Gao, Z., Xie, J., Wang, Q., Li, P.: Global second-order pooling convolutional networks. In: CVPR, pp. 3024–3033 (2019)
6. Gómez-Flores, W., Ruiz-Ortega, B.A.: New fully automated method for segmentation of breast lesions on ultrasound based on texture analysis. Ultrasound Med. Biol. **42**(7), 1637–1650 (2016)
7. Gu, Z., et al.: CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging **38**(10), 2281–2292 (2019)
8. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1397–1409 (2012)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (NIPS), pp. 1097–1105 (2012)
10. Lei, B., et al.: Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder-decoder network. Neurocomputing **321**, 178–186 (2018)
11. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: CVPR, pp. 2117–2125 (2017)
12. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV, pp. 2980–2988 (2017)
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
14. Madabhushi, A., Metaxas, D.N.: Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. IEEE Trans. Med. Imaging **22**(2), 155–169 (2003)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Shan, J., Cheng, H., Wang, Y.: Completely automated segmentation approach for breast ultrasound images using multiple-domain features. Ultrasound Med. Biol. **38**(2), 262–275 (2012)
17. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics. CA: Cancer J. Clin. **67**(1), 7–30 (2017)
18. Stavros, A.T., Thickman, D., Rapp, C.L., Dennis, M.A., Parker, S.H., Sisney, G.A.: Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology **196**(1), 123–134 (1995)
19. Wang, Y., et al.: Deep attentional features for prostate segmentation in ultrasound. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 523–530. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_60

20. Xian, M., Zhang, Y., Cheng, H.D.: Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains. Pattern Recogn. **48**(2), 485–497 (2015)
21. Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., Carson, P.L.: Medical breast ultrasound image segmentation by machine learning. Ultrasonics **91**, 1–9 (2019)
22. Yap, M.H., et al.: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Health Inform. **22**(4), 1218–1226 (2018)
23. Zhang, Q., Shen, X., Xu, L., Jia, J.: Rolling guidance filter. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 815–830. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10578-9_53
24. Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., Shao, L.: ET-Net: A generic edge-aTtention guidance network for medical image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 442–450. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_49
25. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1
26. Zhu, L., et al.: Aggregating attentional dilated features for salient object detection. IEEE Trans. Circ. Syst. Video Technol. **PP**(99), 1 (2019)
27. Zhu, L., et al.: Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 121–136 (2018)