Breast Cancer Histopathological Image Classification Based on Deep Second-order Pooling Network

Jiasen Li*, Jianxin Zhang† *, Qiule Sun‡, Hengbo Zhang†, Jing Dong*, Chao Che*, Qiang Zhang* §
*Key Lab of Advanced Design and Intelligent Computing (Ministry of Education), Dalian University, Dalian, China
†School of Computer Science and Engineering, Dalian Minzu University, Dalian, China
‡School of Information and Communication Engineering, Dalian University of Technology, Dalian, China
§School of Computer Science and Technology, Dalian University of Technology, Dalian, China
Email addresses:jxzhang0411@163.com, zhangq@dlut.edu.cn

***Abstract*—With the breakthrough performance in a variety of computer vision and medical image analysis problems, convolutional neural networks (CNNs) have been successfully introduced for the classification task of breast cancer histopathological images in recent years. Nevertheless, existing breast cancer histopathological image classification networks mainly utilize the first-order statistic information of deep features to represent histopathological images, failing to characterize the complex global feature distribution of breast cancer histopathological images. To address the problem, this work makes a first attempt to explore global second-order statistics of deep features for the above task. More specifically, we propose a novel deep second-order pooling network (DSoPN) for breast cancer histopathological image classification, in which a robust global covariance pooling module based on matrix power normalization (MPN) is embedded into a simple yet effective CNN architecture. The given DSoPN model can capture richer second-order statistical information of deep convolutional features and produce more informative global representations for breast cancer histopathological images. Experimental results on the public BreakHis dataset illuminate the promising performance of the second-order pooling for breast cancer histopathological image classification. Besides, our DSoPN achieves very competitive performance compared to the state-of-the-art methods.**

***Keywords*—Breast cancer histopathological image classification, second-order pooling, covariance estimation, matrix power normalization, convolutional neural network**

I. Introduction

Breast cancer is one of the women cancers with the highest mortality rate in the world [1]. Investigating histopathological images has been regarded as the most important means for the diagnosis of cancers. However, with the number of cancer patients increasing continuously, the manual classification of histopathological images of breast cancer tissues becomes fatigued, tedious, and sensitive to the subjective influence of observers. Therefore, it is of great importance to explore computer-aided diagnostic means, providing more effective diagnosis for the breast cancer.

Recently, convolutional neural networks (CNNs) [2]–[6] have played an important role in various computer vision and medical image analysis tasks. Meanwhile, CNN models related to the classification of pathological images have been attracting wide attentions [7]–[14], which also achieve the breakthrough performance compared to traditional machine learning models. Generally speaking, CNN methods for breast cancer pathological images classification can be divided into three categories as follows. Firstly, some researchers utilize representative or newly constructed CNNs as feature extractors followed by traditional classifiers to distinguish the extracted CNN features, which belong to non-end-to-end models. Deniz et al. [7] adopt pre-trained AlexNet and VGG16 models to capture deep features of breast cancer histopathology images, and then use a support vector machine to distinguish the extracted deep features. Spanhol et al. [12] extensively evaluate the feature representation capability of the pre-trained AlexNet model for this task, whose results illuminate the superior performance to traditional hand-crafted textural descriptors and task-specific CNNs in certain conditions. Besides, Gupta and Bhavsar [10] adopt multi-layered deep features extracted from the fine-tuned DenseNet as the final representations of histopathology images, which are further distinguished through the combined XGBoost classifiers. Next, several previous works directly employ classical networks for end-to-end breast tumor histopathology images classification. To boost the classification performance, these methods usually employ pre-trained networks with fine-tuning. Thus, they are also called as task-specific CNN methods [8], [9]. Shallu and Mehra [8] explore the transfer learning ability in histopathological images compared to fully trained networks on VGG16, VGG19, and ResNet50. Zhang et al. [9] systematically study a family of VGGNet and ResNet models for breast cancer histopathology images classification, and experimental results show that ResNet50 achieves the best performance among the six networks. Lastly, some works attempt to construct novel end-to-end networks using representative networks or modules for the classification of pathological images. Bayramoglu et al. [13] recommend two CNN models with independent and different magnifications, inputting mixed images with various magnifications for training. Spanhol et al. [11] construct a simple CNN with five trainable layers for breast cancer classification, and this model utilizes small batches of pathological images as

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

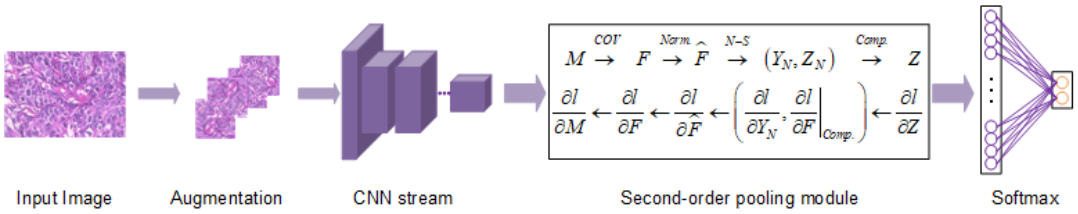

Fig. 1. The overall architecture of our deep second-order pooling network (DSoPN) for breast cancer histopathological image classification. The core of DSoPN consists of two blocks, i.e., a simple CNN as the backbone and a second-order pooling (SoP) module to boost its performance. COV, Norm., N-S and Comp. given in SoP module represent covariance estimate, pre-normalization, Newton-Schultz iteration and post compensation operation, respectively.

network inputs to prevent the overfitting problem, achieving a 6% performance improvement over traditional methods. Jiang et al. [14] design a novel CNN model consisting of small SE-ResNet modules, achieving favorable performance on this task. To further promote its performance, a Gauss error scheduler is also applied for the learning of the network.

All of these studies supply a broad prospect for the application of CNNs in large-scale histopathological image classification of breast cancer, which has significantly promoted the research of this field. However, these methods primarily employ simple first-order statistics of deep features to represent histopathological images, ignoring characterize complex global feature distribution (high-order statistics) of histopathological images, leading to the limitation of classification performance. Actually, high-order representations have shown advantages in obtaining global deep feature statistics of visual images, which have recently attracted more and more attentions [15]–[20]. Compared with first-order CNNs, high-order models effectively capture more complex relationships of features through second-order or higher-order functions inserted into convolutional layers. Relevant researches on high-order statistical modeling mainly include DeepO2P, B-CNN, MPN-COV, etc. [15]–[18] DeepO2P [17] embeds a learnable O2P layer into existing convolutional networks, and it further proposes a matrix backpropagation methodology for network training in an end-to-end manner. Different from DeepO2P, B-CNN [18] captures second-order statistic features by performing an outer product of convolutional activations extracted from two different convolutional networks. In addition, Li et al. [15] utilize covariance square root normalization to construct deep second-order representations of input images, achieving the brilliant performance on large-scale visual recognition tasks. To obtain more discriminant second-order statistic features, GSoP-Net [16] plugs global covariance pooling in-between network, further boosting the performance of convolutional networks.

Inspired by the brilliant performance of high-order networks achieved in computer visual tasks, this work makes an attempt to explore global second-order pooling for breast cancer histopathological image classification problem. Specifically, we propose a novel deep second-order pooling network (DSoPN) for this medical image task, in which a robust covariance pooling module [15], [20] based on matrix power normalization (MPN) is embedded into a simple yet efficient CNN architecture. The proposed DSoPN can capture richer second-order statistic information of deep convolutional features and produce more informative global representations of breast cancer histopathological images. Experiments conducted on public BreakHis dataset illuminate the effectiveness of second-order pooling for this problem. The overall architecture of our DSoPN can be illustrated as in Fig. 1. The main contributions of our paper can be summarized into three aspects. (1) A novel deep second-order pooling network (DSoPN) is proposed for the breast cancer histopathological image classification problem in an end-to-end learning manner. To the best of our knowledge, it is the first attempt to introduce second-order statistic features for this medical image task. (2) Our DSoPN consists of a simple yet effective CNN architecture as the backbone and a second-order pooling module to boost the performance of the network. To obtain more discriminant representations, DSoPN utilizes a covariance estimation module based on matrix power normalization (MPN) to compute second-order statistics of deep activations for histopathological images. (3) Experiment results on the public BreakHis dataset demonstrate the effectiveness of the second-order pooling for this problem. Moreover, our DSoPN achieves very competitive performance compared to the state-of-the-art methods.

II. Method

In this section, we give a detailed description of the deep second-order pooling network (DSoPN) proposed for the breast cancer histopathological image classification problem, which mainly consists of two core blocks, as shown in Figure 1. The leftmost block is a simple yet efficient CNN architecture named SimpleNet as the backbone, and the rightmost block is the second-order pooling module based on the faster training of global covariance pooling module (fast-MPN-COV) [20].

A. SimpleNet

The topology of SimpleNet given in this work can be shown in Figure 2, which mainly consists of five convolutional layers and three fully connected layers. We apply the ReLU [21] activation to provide nonlinearity for the SimpleNet. With the function $f(\boldsymbol{X}) = max(0, \boldsymbol{X})$, the ReLU activation has a faster gradient descent to back-propagate, compared to other activation functions. For the first two convolutional layers and the fifth convolutional layer, we add the Max pooling layer with $kernel\ size = 3 \times 3$ and $stride = 2$. The Max pooling

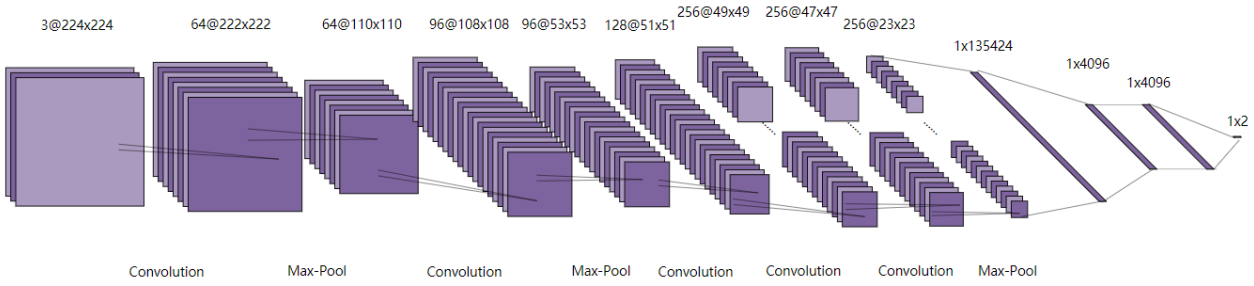

Fig. 2. Framework of the SimpleNet model, which mainly consists of five convolutional layers and three fully connected layers.

layer is behind the ReLU layer to reduce the spatial resolution of the feature maps, and achieves spatial invariance to input distortions and translations [22]. The weights of the network are initialized by using a Gaussian distribution with a low standard deviation of 0.01 for the convolutional layers and fully connected layers. The Dropout layer is applied after the first fully connected layer, with the probability $p = 0.5$, which effectively improves the generalization ability of the network.

Note that though our SimpleNet is similar to the VGGNet [3], we prefer to use a shallower, more efficient and simple network, which is also intuitive and natural. Simple and effective structure is our motivation for adopting the SimpleNet, with which we empirically show, a well-crafted yet simple and reasonably deep architecture can be comparable to or better than a complex network structure, such as VGG-F or VGG-M [3]. In the Section III, we also conduct extensive ablation studies and experiments on both image-level and patient-level classification tasks for evaluating the proposed SimpleNet's performance, compared to the counterparts of VGGNet models.

B. Covariance Pooling Module

Recent works have shown that second-order statistics in deep CNNs encourage obtaining more discriminative representations than first-order ones [15]–[20], [23]. These studies show that second-order statistics can capture the correlation between feature dimensions and significantly improve model classification performance. Specially, global second-order covariance representations of deep features based on matrix power normalization have illustrated brilliant performance on large-scale visual recognition tasks [15], [20], attracting more and more attentions in recent years. Inspired by the above observations, we also adopt covariance pooling to compute the second-order statistic features of breast cancer histopathological images.

1) Covariance normalization: Let $\boldsymbol{X} \in \mathbb{R}^{C\times H\times W}$ be the input feature of the network, we can get the following form:

$$\boldsymbol{M} = \varphi(\boldsymbol{X}; \boldsymbol{W}, \boldsymbol{b}), \tag{1}$$

where φ represents the feature function, which is a CNN consisting of a series of convolutional and pooling layers, such as SimpleNet. $\boldsymbol{W}$ and $\boldsymbol{b}$ represent the weight parameters and the bias of the feature function. We reshape the feature map $\boldsymbol{M} \in \mathbb{R}^{C\times H\times W}$ into a feature matrix $\boldsymbol{M} \in \mathbb{R}^{d\times N}$ with $N = H \times W$ features of d-dimension. Then, covariance differentiation is embedded into deep learning network, and covariance matrix $\boldsymbol{F}$ is established.

$$\boldsymbol{F} = \boldsymbol{M}\bar{\boldsymbol{I}}\boldsymbol{M}^T, \tag{2}$$

where $\bar{\boldsymbol{I}} = \frac{1}{N}\left(\boldsymbol{I} - \frac{1}{N}\boldsymbol{J}\boldsymbol{J}^T\right)$, $\boldsymbol{I}$ is the identity matrix of $N \times N$, $\boldsymbol{J}$ is the d-dimensional vector with all elements are one, and $\boldsymbol{T}$ is the transpose of the matrix.

Since the covariance matrix is a symmetric positive semi-positive, Singular Value Decomposition (SVD) or eigenvalue decomposition (EIG) can be performed as follows:

$$\boldsymbol{F} \to (\boldsymbol{U}, \Lambda), \boldsymbol{F} = \boldsymbol{U}\Lambda\boldsymbol{U}^T, \tag{3}$$

where $\boldsymbol{\Lambda} = diag(\lambda_1, ..., \lambda_d)$ is a diagonal matrix and $\lambda_i, i = 1, ..., d$ is an eigenvalue arranged in non-increasing order; $\boldsymbol{U} = \left[u_1, ..., u_d\right]$ is an orthogonal matrix, and its column u_i is the eigenvector corresponding to the eigenvalue λ_i. Therefore, the power operation of covariance matrix $\boldsymbol{F}$ can be calculated through EIG or SVD decomposition to the power operation on the eigenvalue:

$$(\boldsymbol{U}, \Lambda) \to \boldsymbol{Z}, \boldsymbol{Z} = F^\alpha = \boldsymbol{U}\Phi(\Lambda)\boldsymbol{U}^T, \tag{4}$$

here, $\Phi(\Lambda) = diag(\lambda_1^\alpha, ..., \lambda_d^\alpha)$, the exponent α is a positive real number, and empirically, the covariance matrix works best when α is 0.5.

2) Covariance normalization Acceleration: As described in Subsection II-B1, the power operation on the matrix depends heavily on the SVD or EIG operation, but the implementation of SVD or EIG on GPU is slower than the CPU counterparts [15], [20], [24]. For this reason, we utilize Newton-Schulz iteration to accelerate the computation of covariance normalization [20], [24]. In details, from Eq. (4), the $\boldsymbol{Z} = \boldsymbol{F}^{1/2} = \boldsymbol{Y} = \boldsymbol{U}diag(\lambda_i^{1/2})\boldsymbol{U}^T$ when $\alpha = 1/2$. Let $\boldsymbol{Y}_0 = \hat{\boldsymbol{F}}, \boldsymbol{Z}_0 = \boldsymbol{I}$ for $n = 1, ..., d$, as shown in [20], the coupled iteration can take the following form:

$$\begin{aligned} \boldsymbol{Y}_n &= \frac{1}{2}\boldsymbol{Y}_{n-1}(3\boldsymbol{I} - \boldsymbol{Z}_{n-1}\boldsymbol{Y}_{n-1}), \\ \boldsymbol{Z}_n &= \frac{1}{2}(3\boldsymbol{I} - \boldsymbol{Z}_{n-1}\boldsymbol{Y}_{n-1})\boldsymbol{Z}_{n-1}. \end{aligned} \tag{5}$$

The Eq. (5) can be computed with only matrix product, more suitable for implementation on GPU. Meanwhile, the iteration can achieve an approximate solution with a small number of iterations. In our method, the iteration number is 5.

In order to guarantee the covergence of the Newton-Schulz iteration, the $\boldsymbol{F}$ firstly passes through

$$\hat{\boldsymbol{F}} = \frac{1}{tr(\boldsymbol{F})}\boldsymbol{F}, \tag{6}$$

where $tr(\cdot)$ denotes the trace of matrix. And in order to compensate the data magnitudes caused by Eq. (6), we apply a post-compensation, i.e.,

$$\boldsymbol{Z} = \sqrt{tr(\boldsymbol{F})}Y_N. \tag{7}$$

C. Instantiation

As mentioned above, our proposed DSoPN model combines the merits of the SimpleNet and the covariance pooling module to capture second-order statistic information of deep features, and meanwhile it can be easily trained in an end-to-end learning manner, as shown in the Figure 1. In details, the breast cancer histopathological image passes through all convolutional layers, and first of all, the first five convolutional layers are used to extract feature maps of the input breast cancer histopathological image, and then the feature maps are passed through the covariance pooling module, capturing the second-order statistical information and providing more nonlinearity. Finally, the outputs as the final representation of the image are fed into the classifier.

III. Experiments

In this section, we first introduce the breast cancer histopathological image dataset adopted for evaluating deep second-order pooling network (DSoPN). Then, experimental setting and evaluation metrics are further described, followed by the detailed experimental results on patient level and image level, respectively.

A. Dataset

To effectively evaluate the performance of the proposed DSoPN, we adopt a commonly used breast cancer histopathological image dataset, i.e. BreaKHis dataset. BreaKHis dataset consists of 7909 pathological images from 82 patients, in which each pathological image has been annotated with a benign or malignant label. Besides, 2480 samples belong to benign images and the remained 5429 samples are malignant images. Each sample image has the size of 700x460 pixels in RGB channel pattern with an 8-bit color depth per channel. According to the different magnification factors, samples of each patient can be classified into four groups of 40x, 100x, 200x and 400x, respectively. A typical sample with four different magnification ratio images is shown in Figure 3. The rectangle in this figure is manually added for demonstration purpose, which indicates the region of interest selected to be detailed in the next higher magnification factor.

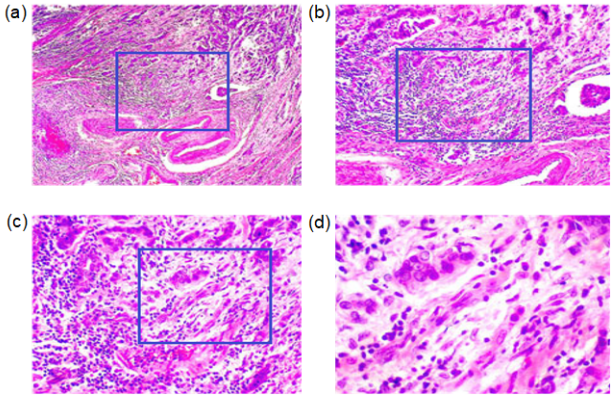

Fig. 3. A typical breast malignant sample in four magnification factors (a) 40×, (b) 100×, (c) 200×, (d) 400×.

B. Evaluation Metrics and Settings

In this work, two typical metrics of image-level recognition rate and patient-level recognition rate are computed for the model measure. The recognition rate of image level does not consider the factor of patient information, but only considers the classification accuracy of models for images. Given N_{all} represents the number of all pathological images in the test set, and N_{rec} represents the number of pathological images correctly classified by the test. The recognition rate of the image level IRR can be calculated as follows:

$$IRR = \frac{N_{rec}}{N_{all}}. \tag{8}$$

Meanwhile, the recognition rate on patient level is calculated by taking patient information into account. Let N_p be the number of pathological images of patient $\boldsymbol{P}$. For patient $\boldsymbol{P}$, if the classification of cancer images is correct, the score on the patient level can be defined as $\boldsymbol{P_s}$ and computed as

$$\boldsymbol{P}_s = \frac{\boldsymbol{N}_{rec}}{\boldsymbol{N}_p}. \tag{9}$$

Let N be the number of patients in the test set, the global patient recognition rate PRR can be further calculated by

$$PRR = \frac{\sum \boldsymbol{P}_s}{\boldsymbol{N}}. \tag{10}$$

For fair comparisons, we adopt the same experimental protocols following [25], [26]. More specifically, the original dataset is randomly divided into training set and testing set for each magnification, where training set consists of 70% images and other 30% of images make up testing set. Additionally, 25% images of the training set are retained for cross-validation to select model parameters. All of experiments adopt the same training dataset and testing dataset. To reduce the influence of the possible over-fitting problem, we perform simple crop and flip operators to augment the size of training set. For the training of networks, we set the initial learning rate to $lr = 0.001$, and attenuate it once during the training process through the anti-attenuation strategy if the loss value keep stable or increasing in 8 epochs. This parameter scale updates

the size or steps taken by the neural network. The stochastic gradient descent method with a batch size of 32 is used to optimize the loss function. The datasets are randomly scrambled to avoid any negative effects on learning from the use of orderly training data. We also set the value of the momentum factor to 0.9, which helps the loss function move to the global minimum instead of getting stuck in the local minimum. All of the experiments are implemented using the Pytorch deep learning framework on Server equipped with Tesla V100 GPU.

C. Experimental Results

1) Comparisons with Baselines: To illustrate the effectiveness of second-order statistic features for this task, we first compare DSoPN with its first-order baseline (i.e., SimpleNet) on BreakHis dataset. In addition, we also carry out experiments by using two milestone models (i.e., VGG-F and VGG-M) and corresponding second-order pooling networks (i.e., VGG-F-SoP and VGG-M-SoP) to further prove the superiority of second-order statistic features for this problem. The compared experimental results on image level and patient level are reported in Table I and Table II, respectively.

TABLE I
COMPARED RESULTS WITH CORRESPONDING FIRST-ORDER BASELINES ON IMAGE LEVEL

Methods	40x(%)	100x(%)	200x(%)	400x(%)
SimpleNet	91.49	92.12	91.39	91.44
DSoPN(SimpleNet-SoP)	96.00	**96.16**	**98.01**	95.97
VGG-F	90.65	88.16	88.25	87.55
VGG-F-SoP	95.49	94.40	96.85	**96.34**
VGG-M	89.15	88.16	88.91	86.26
VGG-M-SoP	**96.33**	94.56	96.36	95.42

Table I shows that DSoPN achieves image level recognition rates of 96.00%, 96.16%, 98.01% and 95.97% on 40x, 100x, 200x and 400x data sets, respectively. DSoPN obtains the optimal classification results on 100x and 200x data sets, while it is slightly worse than VGG-M-SoP on 40x data set and VGG-F-SoP on 400x date set, respectively. In addition, its first-order baseline, i.e., SimpleNet, is superior to both VGG-F and VGG-M on all of the four data sets, whose recognition rates are 91.49%, 92.12%, 91.39% and 91.44% on 40x, 100x, 200x and 400x data sets, respectively. Compared with its first-order counterpart of SimpleNet, DSoPN largely outperforms it by 4.51%, 4.04%, 6.62% and 4.53% gains on the four data sets, respectively. The illustrated promising performance results prove the effectiveness of DSoPN for this medical image problem. When it comes to VGG-based models, it is clear to see that the two second-order pooling models are also superior to corresponding first-order baselines by a large margin. By embedding the second-order pooling layer into VGG-F, the image level recognition rate on 40x, 100x, 200x and 400x data sets increases by 4.84%, 6.24%, 8.6% and 8.97%, respectively. Similarly, the second-order pooling model VGG-M-SoP gains 7.54% performance improvement on average over its first-order network. Specially, VGG-M-SoP outperforms VGG-M by 9.16% on 400x data set. These significant performance improvements establish the effectiveness of second-order statistic features for breast cancer histopathological image classification tasks.

TABLE II
COMPARED RESULTS WITH CORRESPONDING FIRST-ORDER BASELINES ON PATIENT LEVEL

Methods	40x(%)	100x(%)	200x(%)	400x(%)
SimpleNet	89.76	93.67	92.85	92.86
DSoPN(SimpleNet-SoP)	95.01	**96.84**	**97.92**	96.28
VGG-F	90.19	90.63	87.42	88.06
VGG-F-SoP	**95.22**	94.82	97.48	**96.72**
VGG-M	88.69	90.51	89.36	87.09
VGG-M-SoP	95.12	94.77	97.28	96.22

Table II illuminates the compared experiment results with corresponding first-order counterparts on patient-level classification problem. As shown in Table II, patient-level classification accuracy of 95.01%, 96.84%, 97.92% and 96.28% are achieved by DSoPN on 40x, 100x, 200x and 400x data sets, respectively. The optimal classification accuracy on 100x and 200x data sets are obtained by DSoPN, while the best results on 40x and 400x data sets belong to VGG-F-SoP. Though DSoPN is shallower and has fewer parameters, it has a similar classification performance with VGG-F-SoP on this medical image task, and both of them slightly outperform VGG-M-SoP model. When comparing second-order pooling networks with their first-order counterparts, similar to the comparison results on image level, by embedding second-order pooling layers, all of the three second-order pooling networks gain significant performance improvements on the four data sets. The compared performance results once again prove the effectiveness of deep second-order statistic features for breast cancer histopathological image classification application.

2) Comparisons with representative methods: To further evaluate the effectiveness of second-order statistic features for this task, we further compare DSoPN, VGG-F-SoP and VGG-F-SoP with other typical CNN-based methods that have achieved advanced level in the past four years, and the compared results on both image level and patient level are listed in Table III. As shown in this table, our second-order methods are superior to most of the previous methods, illustrating a competitive performance compared to several state-of-the-art methods.

More specially, for compared results on image level, our second-order models significantly outperform methods given in [11], [26], [28], [31]–[33]. Among the six methods, Benhammo et al. obtain optimal result of 90.20% on 40x data set by using Inception V3 model. Meanwhile, CNN+FV and ResHist-Aug have the best performance of 87.60% and 91.15% on 100x and 200x data sets, respectively. In addition, the highest accuracy on 400x data set of 86.60% is achieved by [32]. However, our proposed DSoPN, whose accu-

TABLE III
COMPARED RESULTS WITH REPRESENTATIVE CNN-BASED METHODS ON BOTH IMAGE LEVEL AND PATIENT LEVEL

References	Years	Methods	Image Level (%)				Patient Level (%)			
			40x	100x	200x	400x	40x	100x	200x	400x
Spanhol et al. [11]	2016	A variant of AlexNet	85.60	83.50	82.70	80.70	90.00	88.40	84.60	86.10
Bayramoglu et al. [13]	2016	Single Task CNN	–	–	–	–	83.08	83.17	84.63	82.10
Song et al. [28]	2017	CNN+FV	87.70	87.60	86.50	83.90	90.20	91.20	87.80	87.40
Han et al. [27]	2017	CSDCNN	95.80	**96.90**	96.70	94.90	**97.10**	95.70	96.50	95.70
Bardou et al. [25]	2018	CNN+Aug	–	–	–	–	96.82	**96.96**	96.36	95.97
Benhammo et al. [26]	2018	Inception V3	90.20	85.60	86.10	82.50	91.50	85.10	86.80	82.90
Gupta et al. [10]	2018	Sequential(Ave.)	–	–	–	–	94.71	95.90	96.76	89.11
Zhu et al. [31]	2019	Multiple CNNs	85.70	84.20	84.90	80.10	85.20	83.50	84.10	79.30
Lichtblau et al. [32]	2019	DE ensemble	83.90	86.00	89.10	86.60	85.60	87.40	89.80	87.00
Gupta and Bhavsar [29]	2019	Independent framework	–	–	–	–	96.81	95.26	93.78	90.76
Gour et al. [33]	2020	ResHist-Aug	87.40	87.26	91.15	86.27	87.47	88.15	92.52	87.78
Kumar et al. [30]	2020	VGGNET16-SVM	–	–	–	–	94.11	95.12	97.01	93.40
Ours	–	DSoPN	**96.00**	96.16	**98.01**	95.97	95.01	96.84	**97.92**	96.28
		VGG-F-SoP	95.49	94.40	96.85	**96.34**	95.22	94.82	97.48	**96.72**
		VGG-M-SoP	96.33	94.56	96.36	95.42	95.12	94.77	97.28	96.22

racy reaches 96.00%, 96.16%, 98.01% and 95.97%, achieves 5.80%, 8.56%, 6.86% and 9.37% gains over these optimal results on 40x, 100x, 200x and 400x data sets, respectively. In addition, DSoPN achieves competitive results with the state-of-the-art model of class structure-based deep CNN (CSDCNN) [27], which adopts an extra supervised distance constraint to supervise the learning process of the network, and outperforms CSDCNN on 40x, 200x, and 400x sets, respectively.

When it comes to results on patient level, we compare our second-order pooling models with 12 typical first-order models including CSDCNN, Multiple CNNs, Sequential model, etc. It is worth noting that our second-order methods still outperform most of the given methods. Among them, models proposed by [11], [13], [26], [28], [31]–[33] obtain the average accuracy of less than 90% on the four data sets, which are obviously inferior to other methods. CSDCNN and CNN+Aug gain brilliant performance with average accuracy over 96%, and are superior to the remained three methods [10], [29], [30] by appropriate 2.00%. Our second-order pooling models can achieve extremely competitive performance with CSDCNN and CNN+Aug. Moreover, they obtain the optimal accuracy of 97.92% and 96.72% on 200x and 400x data sets, respectively. The compared results on patient level once again show the effectiveness of second-order pooling models.

3) Visualization Results: Finally, we select and analyze some representative images that can be well classified by second-order network of DSoPN but failed to be correctly discriminated by its corresponding first-order baseline of SimpleNet model. Figure 4 shows some typical images in 100x data set under this condition, where images of the first two rows are breast benign tumor images that are misclassified as malignant samples, and images of the third and fourth rows are mispredicted as benign images by SimpleNet model. As shown in Figure 4, pathological images with large blank areas are often not able to be correctly classified by the baseline CNN network. When benign images contain irregular texture in most areas, SimpleNet easily fails to classify these images. Meanwhile, by analyzing misclassified malignant images, it

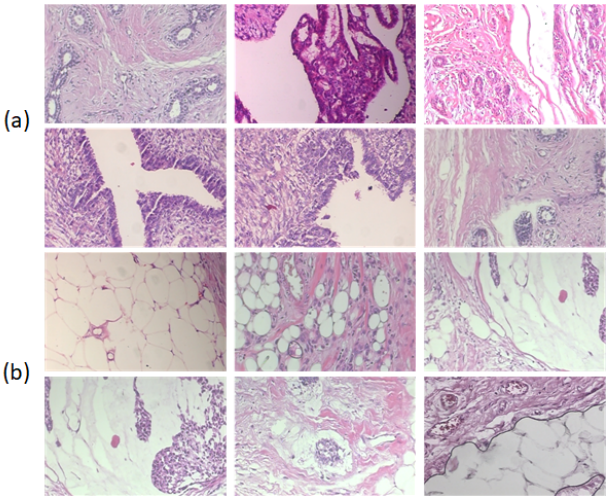

Fig. 4. Sample images from 100x data set that are correctly classified by DSoPN while misclassified by its first-order baseline, where images of the first two rows are benign images and the remaining belong to malignant.

can be seen that the inapparent texture feature could also be the possible reason leading to the misclassification of the baseline CNN. However, these images can be correctly classified by our second-order pooling network, indicating the effectiveness of DSoPN in capturing more discriminant features of breast cancer histopathological images.

IV. CONCLUSION

In this work, we try to explore global high-order statistics and propose a novel deep second-order pooling network (DSoPN) for breast cancer histopathological image classification. DSoPN employs a robust covariance estimation module to capture richer second-order statistic information of deep convolutional features, providing a more discriminant representation of histopathological images. Experimental results on public BreakHis dataset also illuminate the promising performance of DSoPN, proving the effectiveness of second-order statistic features for this medical image task. Our future

work will further explore more robust high-order statistic features for breast cancer histopathological image classification by performing more advanced statistic models and suitable backbone networks. Besides, the combination of first-order and high-order statistic features will also be addressed.

ACKNOWLEDGMENT

This work was partially supported by the National Key R&D Program of China (2018YFC0910506), the National Natural Science Foundation of China (61972062 and 61603066), the Natural Science Foundation of Liaoning Province (2019-MS-011), the Key R&D Program of Liaoning Province (2019 JH2/10100030) and the Liaoning BaiQianWan Talents Program. The authors are grateful to P&D Lab, Brazil, for providing the BreakHis dataset used in this work.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: a cancer journal for clinicians, vol. 68, no. 6, pp. 394-424, 2018.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[3] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proceedings of the International Conference on Learning Representations. 2015, pp. 1-14.

[4] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 770-778.

[5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp. 4700-4708.

[6] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 7132-7141.

[7] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, and Ü. Budak, "Transfer learning based histopathologic image classification for breast cancer detection," Health Information Science and Systems, vol. 6, no. 1, pp. 18, 2018.

[8] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," ICT Express, vol. 4, no. 4, pp. 247–254, 2018.

[9] J. X. Zhang, X. G. Wei, C. Che, X. P. Wei, and Q. Zhang, "Breast Cancer Histopathological Image Classification Based on Convolutional Neural Networks," Journal of Medical Imaging and Health Informatics, vol. 9, no. 4, pp. 735-743, 2019.

[10] V. Gupta, and A. Bhavsar, "Sequential modeling of deep features for breast cancer histopathological image classification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018, pp. 2254-2261.

[11] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in 2016 international joint conference on neural networks (IJCNN). 2016, pp. 2560-2567.

[12] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2017, pp. 1868-1873.

[13] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in 2016 23rd International conference on pattern recognition (ICPR). 2016, pp. 2440-2445.

[14] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," PloS one, vol. 14, no. 3, ID. e0214587, 2019.

[15] P. H. Li, J. T. Xie, Q. L. Wang, and W. M. Zuo, "Is second-order information helpful for large-scale visual recognition?," in Proceedings of the IEEE International Conference on Computer Vision. 2017, pp. 2070-2078.

[16] Z. L. Gao, J. T. Xie, Q. L. Wang, and P. H. Li, "Global Second-order Pooling Convolutional Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, pp. 3024-3033.

[17] C. Ionescu, O. Vantzos, and C. Sminchisescu, "Matrix backpropagation for deep networks with structured layers," in Proceedings of the IEEE International Conference on Computer Vision. 2015, pp. 2965-2973.

[18] T. Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear convolutional neural networks for fine-grained visual recognition," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 6, pp. 1309-1322, 2017.

[19] Q. L. Sun, Q. L. Wang, J. X. Zhang, and P. H. Li, "Hyperlayer bilinear pooling with application to fine-grained categorization and image retrieval," Neurocomputing, vol. 282, pp. 174-183, 2018.

[20] P. H. Li, J. T. Xie, Q. L. Wang, and Z. L. Gao, "Towards faster training of global covariance pooling networks by iterative matrix square root normalization," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018, pp. 947-955.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in Proceedings of the 27th international conference on machine learning (ICML). 2010, pp. 807-814.

[22] M. Ranzato, F. Huang, Y. Boureau, and Y. Lecun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in IEEE conference on computer vision and pattern recognition (CVPR). 2007, pp. 1-8.

[23] T. Dai, J. R. Cai, Y. B. Zhang, S. T. Xia, and L. Zhang, "Second-order Attention Network for Single Image Super-Resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019, pp. 11065-11074.

[24] T. Y. Lin, and S. Maji, "Improved bilinear pooling with cnns," in British Machine Vision Conference (BMVC). 2017, pp. 1-12.

[25] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," IEEE Access, vol. 6, pp. 24680-24693, 2018.

[26] Y. Benhammou, S. Tabik, B. Achchab, and F. Herrera, "A first study exploring the performance of the state-of-the art CNN model in the problem of breast cancer," in Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications. ACM, 2018, ID. 47.

[27] Z. Y. Han, B. Z. Wei, Y. J. Zheng, Y. L. Yin, and S. J. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," Scientific reports, vol. 7, no. 1, pp. 4172, 2017.

[28] Y. Song, H. Chang, H. Huang, and W. D. Cai, "Supervised intra-embedding of fisher vectors for histopathology image classification," in International Conference on Medical Image Computing and Computer-Assisted Intervention. 2017, pp. 99-106.

[29] V. Gupta, and A. Bhavsar, "Partially-Independent Framework for Breast Cancer Histopathological Image Classification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019, pp. 0-0.

[30] A. Kumar, S. K. Singh, S. Saxena, K.Lakshmanan, A. K. Sangaiah, and H. Chauhab, et al., "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," Information Sciences, vol. 508, pp. 405-421, 2020.

[31] C. Zhu, F. Z. Song, Y. Wang, H. H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," BMC medical informatics and decision making, vol. 19, no. 1, pp. 198, 2019.

[32] D. Lichtblau, C. Stoean, "Cancer diagnosis through a tandem of classifiers for digitized histopathological slides," PloS one, vol. 14, no. 1, ID. e0209274, 2019.

[33] G. Mahesh, J. Sweta, T. Sunil, "Residual learning based CNN for breast cancer histopathological image classification," International Journal of Imaging Systems and Teachnology. doi:10.1002/ima.22403, 2020.