Micro and Macro Breast Histology Image Analysis by Partial Network Re-use

Quoc Dang Vu, Minh Nguyen Nhat To, Eal Kim, and Jin Tae Kwak(✉)

Department of Computer Science and Engineering, Sejong University, Seoul 05006, Korea
dangvuquocl993@gmail.com, tnnhatminh@gmail.com, laoavae@naver.com, jkwak@sejong.ac.kr

Abstract. Convolutional neural networks (CNN) have shown to be effective in medical image processing and analysis. Herein, we propose a CNN approach to perform patch- and pixel-wise histology labeling on breast microscopy and whole slide images (WSI), respectively. We devise a processing block that is capable of extracting compact features in an efficient manner. Based upon the processing block, classification and segmentation networks are built. Two networks share an encoder via partial transformation and transfer learning to maximally utilize the trained network and available dataset. 400 microscopy images and 10 WSI were employed to evaluate the proposed approach. For patch classification, an accuracy of 71% and 65% were obtained on the training and testing dataset, respectively. As for segmentation, we achieved an overall score of 0.7343 and 0.4945 on the training and testing dataset, respectively.

Keywords: Breast cancer · Tissue segmentation · Histology analysis CNN

1 Introduction

Breast cancer is one of the most prevalent causes of cancer-related death in women [1] and early diagnosis significantly increases the patient survival rate. In breast pathology, manual histologic assessment of tissue specimens forms the gold standard for cancer diagnosis, prognosis, and treatment. The manual process, however, not only limits speed and throughput but also places a huge demand on clinical services. It also suffers from substantial intra- and inter-observer variability [2]. Automated, robust, and precise tools for analyzing tissues will, therefore, aid in improving breast pathology.

With the advent of high-resolution and cost-effective digital scanners, numerous machine learning approaches have been applied to improve the analysis of digitized tissue specimen images [3]. Recently, deep learning, in particular convolutional neural networks (CNN), has been increasingly applied for pathology images [4, 5]. Its great learning capability has been recognized and confirmed by multiple applications such as detection of mitosis [6], invasive ductal carcinoma [7], and metastases [8] in breast and gland segmentation [5]. CNN approaches often fall into two categories: (1) region or patch classification and (2) segmentation (or pixel-wise classification). Although the

A. Campilho et al. (Eds.): ICIAR 2018, LNCS 10882, pp. 895–902, 2018.
https://doi.org/10.1007/978-3-319-93000-8_102

ultimate goal and overall structure (slightly) differ, both networks try to extract and utilize high-level feature representation of tissues. The networks, therefore, could be shared or transferred from one to the other. This may aid in overcoming the limited availability of dataset and ground truth in image analysis of breast pathology.

In earlier CNN architectures, convolution and pooling layers are, in general, (repeatedly) stacked on top of each other, forming a processing block of the network such as in VGG [12] and Unet [7], and shown to be effective in various tasks [13]. Later, a more efficient layout like residual blocks [14] further improves upon the performance of CNN [15]. A processing block, however, often generates high-dimensional features, i.e., the number of convolution kernels becomes much larger than the spatial dimension of the input and convolution kernels. It is likely that the resultant feature maps are redundant [9]. This redundancy not only decreases the learning capability of the network but also wastes memory storage. Therefore, a processing block that is able to learn a set of compact and efficient features will lead to a more efficient and powerful network.

In this manuscript, we present a deep learning approach for analyzing breast histology images at both micro-level (patch-based image classification) and macro-level (whole slide image (WSI) segmentation). Both breast image classification and segmentation networks are built based upon identical processing blocks, which attempt at producing high-level, compact feature representation of the breast images. Moreover, the architecture and weights of the two networks are shared to maximally utilize the networks and available dataset. The proposed networks have been evaluated using tissue specimen imaging dataset provided by the ICIAR 2018 BACH Challenge.

2 Methodology

The proposed approach contains two networks: (1) a classification network to conduct patch-based image classification and (2) a segmentation network to provide segmentation maps for WSI (Fig. 1).

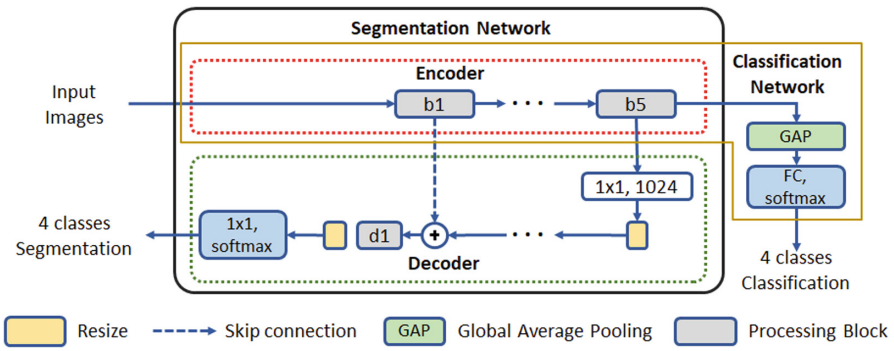

Fig. 1. Overview of the proposed approach. A classification network consists of an encoder and two processing layers. A segmentation network contains an encoder and decoder.

2.1 Processing Block

A processing block (Fig. 2) is composed of three units: the top and the bottom units are standard convolutions for spatial/feature dimensionality reduction; the unit in the middle follows the design principle of DenseNet [10] where a series of convolutions (designated as a sub-unit) are performed, each of which takes all the preceding inputs via concatenation (called as feature re-use). Feature re-use is known to reduce the number of parameters without losing the learning capability of the network. Each sub-unit is designed to find a more compact and structured feature representation by conducting two convolutions with a dilation rate of 1 (standard) and 2 (dilated) [11]; the first convolution is followed by a squeeze excitation (or self-gating) mechanism [12] to facilitate dynamic feature selection; the second convolution adopts a split-transform-merge strategy (or grouped convolution) [13] to improve feature correlation, leading to a more structured feature representation as well as a dilation rate of 2 to incorporate contextual information. In addition, the pre-activation [14] layout of Batch Normalization – ReLU – Convolution is adopted for all the convolutions in the proposed networks to ease the gradient flow during the optimization procedure. The top unit of each processing block lacks ReLU [15].

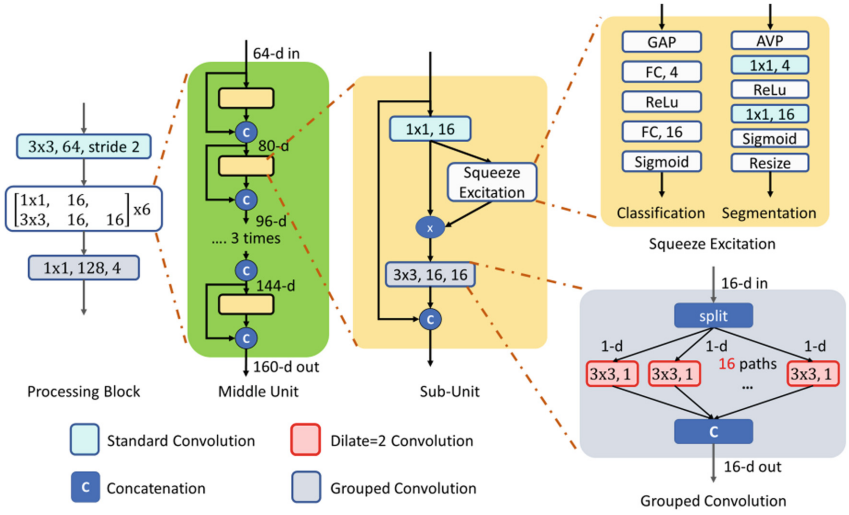

Fig. 2. Layout of the processing block *b2* in the encoder. GAP and AVP denote global average pooling and average pooling, respectively.

2.2 Classification Network

A classification network is comprised of five processing blocks (called as "encoder"), global average pooling (GAP) layer and fully-connected (FC) softmax layer (Table 1). The middle unit in the first processing block does not use the squeeze excitation.

Table 1. Architecture of the proposed approach. [AVP, 17 x 17, 8] denotes the region-wise squeeze excitation using average pooling (AVP) with kernel size of 17 x 17 and stride 8.

	Encoder					
	b1	*b2*	*b3*	*b4*	*b5*	
Classification	3x3, 64	3x3, 64, stride 2	3x3, 128, stride 2	3x3, 256, stride 2	3x3, 512, stride 2	
	3x3, 64, 16	[1x1, 16; 3x3, 16, 16] x6	[1x1, 16; 3x3, 16, 16] x12	[1x1, 16; 3x3, 16, 16] x24	[1x1, 16; 3x3, 16, 16] x32	Global AVP, FC, softmax
	3x3, 64	1x1, 128, 4	1x1, 256, 8	1x1, 512, 16	1x1, 512, 32	
Segmentation	3x3, 64	3x3, 64, stride 2	3x3, 128, stride 2	3x3, 256, stride 2	3x3, 512, stride 2	
	3x3, 64, 16	[1x1, 16; AVP, 17x17, 8; 3x3, 16, 16] x6	[1x1, 16; AVP, 9x9, 4; 3x3, 16, 16] x12	[1x1, 16; AVP, 5x5, 2; 3x3, 16, 16] x24	[1x1, 16; AVP, 3x3, 1; 3x3, 16, 16] x32	
	3x3, 64	1x1, 128, 4	1x1, 256, 8	1x1, 512, 16	1x1, 512, 32	
	Decoder					
	d4	*d3*	*d2*	*d1*		
Segmentation	5x5, 256	5x5, 128	5x5, 64	5x5, 32		
	[5x5, 128, 5x5, 128, 32] x2	[5x5, 64, 5x5, 64, 16] x2	[5x5, 32, 5x5, 32, 8] x2	[5x5, 16, 5x5, 16, 4] x2	1x1, Softmax	
	1x1, 256	1x1, 128	1x1, 64	1x1, 32		

2.3 Segmentation Network

A segmentation network consists of an encoder and decoder. The encoder utilizes the identical layout of the encoder from the classification. Weights are also transferred from the classification network. The processing blocks, however, utilize a modified squeeze excitation. In the classification network, it excites the whole feature map in response to a pertinent class, which contradicts with the segmentation task where a class label is assigned per pixel. Pixel-wise excitation is not applicable since the intermediate output does not preserve the exact location information of an object of interest. Alternatively, a local, region-wise excitation, where GAP is replaced by average pooling (AVP) with stride, is applied per feature map as illustrated in Fig. 2.

The decoder (Table 1) includes four processing blocks, each of which is followed by a nearest neighbor resizing layer (up-sampling), instead of other popular techniques such as deconvolution [4] or un-pooling [16]. Each processing block obtains two inputs, one from the preceding layer in the decoder and the other from the encoder via a skip connection [4, 16–18]. No squeeze excitation is employed. All the convolutions use no padding and a stride 1.

3 Dataset and Training Details

3.1 ICIAR2018 BACH Challenge Dataset

Hematoxylin and eosin (H&E) stained breast histology images are provided by ICIAR2018 Grand Challenge on Breast Cancer Histology images (BACH).

Training Dataset. For classification, 400 microscopy image patches that are labeled as one of four classes (normal, benign, in situ carcinoma or invasive carcinoma; 100 images per class) are employed. Each patch is of size 2048 x 1536 pixels with the pixel scale of 0.42 um x 0.42 um. For segmentation, 10 WSI images are used that was acquired through Leica SCN400 with a pixel resolution of 0.467 um/pixel. Three distinct regions are marked as benign (red), in situ carcinoma (green) and invasive carcinoma (blue). The rest of tissues are considered to be normal (black). For both tasks, the ground truth annotation was performed by two medical experts and those image patches or WSI regions on which they disagree were not included.

Testing Dataset. A total of 100 image patches are used for classification task. Each path is of size 2048 x 1536. For segmentation, a total of 10 WSI images are utilized. The ground truth annotation is blinded to the challenge participants.

3.2 Training Details

Classification. The classification network (encoder) is trained and evaluated via a 5-fold cross validation. The original image patches are down-sampled by a factor of 4 (scale x0.25) and then directly used to train the network without any additional pre-processing steps. During training, a random cropping, resizing [19], horizontal and vertical flipping are applied for data augmentation. The network is initialized with He method [20] and trained via Stochastic Gradient Descent for 120 epochs. The learning rate is initially set to 3.5e−3 and decreases by a factor of 10 at every 60 epochs.

Segmentation. WSI dataset is massive and highly imbalanced. By resizing WSI at x0.25 scale and manually sampling sub-regions, we sought to efficiently obtain the segmentation network that is less biased. 6,129 sub-regions of size 1000 x 1000, maintaining roughly the same number of regions per class and accounting for 17% the total volume of the WSI dataset, are utilized for training. For each sub-region, the center region of 630 x 630 is fed into the network to avoid the zero-padding during augmentation, and the segmentation map is generated for the inner most central region of size 204 x 204. Moreover, we re-use the weight of the encoder from the classification network. He method [20] is used to initialize the decoder and the extended layers. During training, random affine transformations (shifting, scaling, rotation and minor shearing) are applied to ensure the robustness of the network. Alongside the (main) loss for the targeted four classes, an auxiliary loss is jointly computed, at the same level as the main loss, to emphasize the closeness among the four classes; it is computed via categorical cross-entropy for normal and benign group versus in situ and invasive carcinoma group. The network is trained for 30 epochs using Adam optimizer

with default parameter values. The learning rate is initially set to 3.5e−5 and reduced by a factor of 10 at each subsequent 15th epoch.

4 Experimental Results and Discussion

4.1 Evaluation Metric

For classification task, the overall prediction accuracy, which is the ratio of correctly predicted patches to the total number of patches, is measured. As for segmentation task, the performance is assessed using the following formula $s = 1 - \frac{\sum_{i=1}^{N} |pred_i - gt_i|}{\sum_{i=1}^{N} \max(|gt_i - 0|, |gt_i - 3|) \times [1 - (1 - pred_{i,bin})(1 - gt_{i,bin})}$ where "*pred*" is the predicted class (0, 1, 2 or 3), and "*gt*" is the ground truth class, *i* is the linear index of a pixel in an image, *N* is the total number of pixels in an image and bin is the binarized value, i.e., is 0 if the label is 0 and 1 if the label is 1, 2 or 3. This score aims at penalizing more the predictions that are farther from the ground truth label (or value).

4.2 Classification and Segmentation Performance

To evaluate the performance of the proposed networks, we conducted two separate experiments per network. Each network was evaluated based upon the training dataset and then tested on the testing dataset. The performance on the testing dataset was evaluated by the challenge organizer. The testing results are available online at https://iciar2018-challenge.grand-challenge.org/results/.

Classification. In a 5-fold cross-validation on the training dataset, an average accuracy of 71% with 6.4% standard deviation was obtained. By using the best model from the 5-fold cross validation, an accuracy of 65% was obtained on the testing dataset.

Segmentation. Trained on the manually selected sub-regions and tested on the remaining regions, the network achieved an overall score of 0.7343 at x0.25 scale. Tested on the testing dataset, we obtained a score of 0.4945 at x0.5 scale.

5 Conclusion

In this manuscript, we propose a CNN approach for micro-level and macro-level analysis of breast tissue images. The two networks share the architecture and weights, in particular the encoder in the network, to improve the utility of the trained network and available dataset. Both networks are built based upon a processing block that is designed for compact and efficient feature learning. The future work will entail the investigation of an optimal design of a processing block and CNN that can maximize its memory capacity and learning capability and the extended study on other types of tissues and disease.

Acknowledgments. This study is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2016R1C1B2012433).

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2016. CA Cancer J. Clin. **66**, 7–30 (2016)
2. Elmore, J.G., Longton, G.M., Carney, P.A., et al.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA **313**, 1122–1132 (2015)
3. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: challenges and opportunities. Med. Image Anal. **33**, 170–175 (2016)
4. 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
5. Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.A.: DCAN: deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. **36**, 135–146 (2017)
6. Ciresan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. Med. Image Comput. Comput. Assist. Interv. **16**, 411–418 (2013)
7. Cruz-Roa, A.A., Arevalo Ovalle, J.E., Madabhushi, A., Gonzalez Osorio, F.A.: A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Med. Image Comput. Comput. Assist. Interv. **16**, 403–410 (2013)
8. Litjens, G., Sanchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-van de Kaa, C., Bult, P., van Ginneken, B., van der Laak, J.: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci. Rep. **6**, 26286 (2016)
9. Chen, W., Wilson, J.T., Tyree, S., Weinberger, K.Q., Chen, Y.: Compressing neural networks with the hashing trick. ArXiv e-prints, vol. 1504 (2015)
10. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. ArXiv e-prints, vol. 1608 (2016)
11. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. ArXiv e-prints, vol. 1511 (2015)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. ArXiv e-prints, vol. 1709 (2017)
13. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. ArXiv e-prints, vol. 1611 (2016)
14. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. ArXiv e-prints, vol. 1603 (2016)
15. Dong, X., Kang, G., Zhan, K., Yang, Y.: EraseReLU: a simple way to ease the training of deep convolution neural networks. ArXiv e-prints, vol. 1709 (2017)
16. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. ArXiv e-prints, vol. 1511 (2015)
17. Lin, G., Milan, A., Shen, C., Reid, I.: RefineNet: multi-path refinement networks for high-resolution semantic segmentation. ArXiv e-prints, vol. 1611 (2016)
18. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large Kernel matters – improve semantic segmentation by global convolutional network. ArXiv e-prints, vol. 1703 (2017)

19. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. ArXiv e-prints, vol. 1409 (2014)
20. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. ArXiv e-prints, vol. 1502 (2015)