PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

From patch-level to ROI-level deep feature representations for breast histopathology classification

Mercan, Caner, Aksoy, Selim, Mercan, Ezgi, Shapiro, Linda, Weaver, Donald, et al.

Caner Mercan, Selim Aksoy, Ezgi Mercan, Linda G. Shapiro, Donald L. Weaver, Joann G. Elmore, "From patch-level to ROI-level deep feature representations for breast histopathology classification," Proc. SPIE 10956, Medical Imaging 2019: Digital Pathology, 109560H (18 March 2019); doi: 10.1117/12.2510665

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

From Patch-level to ROI-level Deep Feature Representations for Breast Histopathology Classification

Caner Mercan[a], Selim Aksoy[a], Ezgi Mercan[b], Linda G. Shapiro[b], Donald L. Weaver[c], and Joann G. Elmore[d]

[a]Department of Computer Engineering, Bilkent University, Ankara, 06800, Turkey
[b]Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA 98195, USA
[c]Department of Pathology, University of Vermont, Burlington, VT 05405, USA
[d]David Geffen School of Medicine, University of California, Los Angeles, USA

ABSTRACT

We propose a framework for learning feature representations for variable-sized regions of interest (ROIs) in breast histopathology images from the convolutional network properties at patch-level. The proposed method involves fine-tuning a pre-trained convolutional neural network (CNN) by using small fixed-sized patches sampled from the ROIs. The CNN is then used to extract a convolutional feature vector for each patch. The softmax probabilities of a patch, also obtained from the CNN, are used as weights that are separately applied to the feature vector of the patch. The final feature representation of a patch is the concatenation of the class-probability weighted convolutional feature vectors. Finally, the feature representation of the ROI is computed by average pooling of the feature representations of its associated patches. The feature representation of the ROI contains local information from the feature representations of its patches while encoding cues from the class distribution of the patch classification outputs. The experiments show the discriminative power of this representation in a 4-class ROI-level classification task on breast histopathology slides where our method achieved an accuracy of 66.8% on a data set containing 437 ROIs with different sizes.

Keywords: Digital pathology, computer aided diagnosis, breast histopathology, region of interest classification

1. INTRODUCTION

Breast cancer is the most prevalent type of cancer among women. The treatments of the patients depend on the types of the decisions that the pathologists make based on their interpretation of the biopsy slides. Histopathological image analysis systems aim to aid the pathologists in their interpretation of these slides by filtering out benign regions or by pointing out malignant areas so that they are investigated in more detail.

Whole slide imaging (WSI) involves digitization of biopsy slides at high resolution. The slides may contain several different types of malignant areas, and associating different parts with a diagnosis poses a challenge. Multi-class classification of whole slide images by learning from slide-level annotations using multi-instance and multi-label learning based approaches have shown promising results at predicting malignant areas in whole slides.[1–4] However, the slide-level diagnosis may only correspond to a relatively small portion of the whole slide image. Training of classifiers can be performed more effectively when the tissue structures used in learning are in isolation and have no ambiguity in their diagnostic labels. This setting requires the pathologists to annotate the regions of interest (ROI) in whole slides and associate each ROI with one of the diagnostic labels. Performance improvements on the multi-class classification of malignant regions can be achieved when isolated ROIs with associated labels are used in the training of the state-of-the-art deep learning based models.

Convolutional neural networks (CNNs) have had great success in several different domains including histopathological image analysis.[5–8] CNNs typically require the input image to be of specific size, often very small, due to computational limitations. This poses a problem for the classification of the ROIs in histopathology images,

Send correspondence to S.A.: E-mail: saksoy@cs.bilkent.edu.tr, Telephone: +90 (312) 2903405

Medical Imaging 2019: Digital Pathology, edited by John E. Tomaszewski, Aaron D. Ward,
Proc. of SPIE Vol. 10956, 109560H · © 2019 SPIE · CCC code:
1605-7422/19/$18 · doi: 10.1117/12.2510665

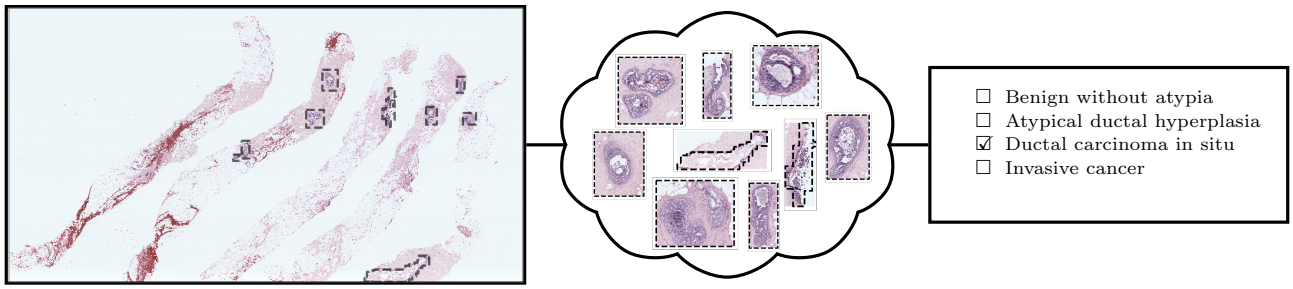


Figure 1. An example slide, manually annotated ROIs, and the diagnosis of the slide.

due to them having arbitrarily different sizes in high-resolution. CNNs trained on cropped patches from an ROI may not be able to preserve the contextual information within the ROI, and using a resized version of the ROI as input to CNNs may result in structural information loss. In this paper, we propose a framework that involves learning a feature representation for large ROIs from the feature representations of their patches. The proposed method aims to simultaneously preserve the local information contained in the patches and encode the class distributions of these patches within the ROI. Then, the multi-class classification of the ROIs using their feature representations can be performed in a separate experiment involving a classifier of choice.

2. DATA SET

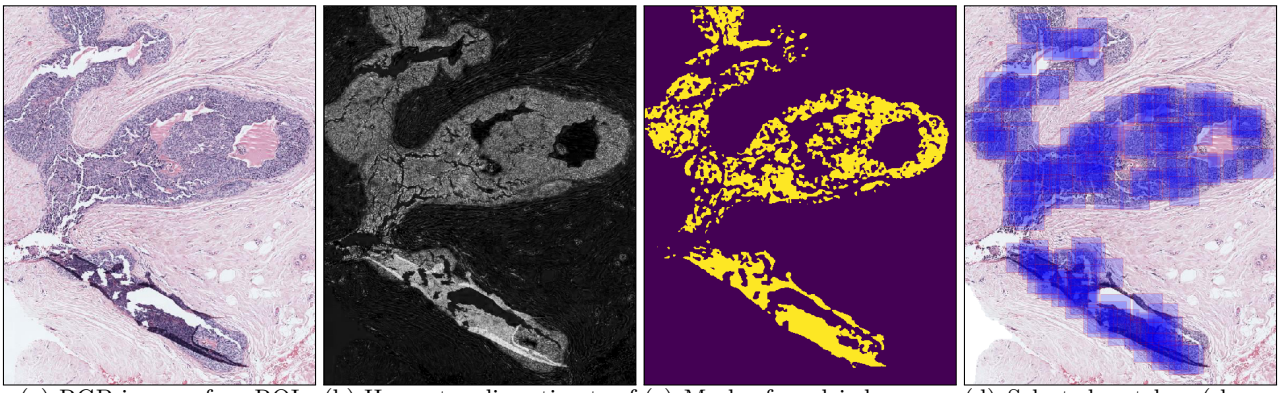
We use a data set of 240 breast histopathology images that were collected as part of an NIH-sponsored project titled "Digital Pathology, Accuracy, Viewing Behavior and Image Characterization (digiPATH)". The data set contains haematoxylin and eosin (H&E) stained breast biopsy slides with an average image size of 100,000×64,000 pixels at 40× magnification. All slides were scanned by the same iScan Coreo Au digital slide scanner. Each slide was examined by three experienced pathologists in their consensus meetings, and was associated with a single consensus diagnosis. Each slide has one or more ROIs that were also marked to correspond to the most severe diagnosis within the associated slide. In total, there are 437 ROIs, each associated with a single diagnostic label, available within the 240 whole slide images that we investigate at 10× magnification. An example slide with its ROIs and the associated diagnostic label is presented in Figure 1.

3. METHODOLOGY

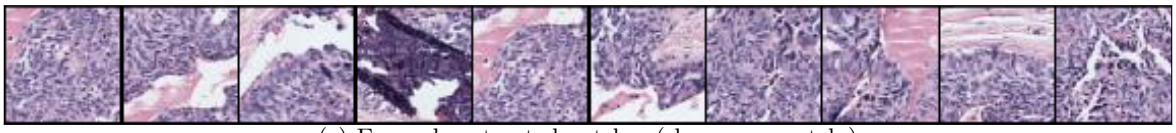
CNNs have been used in various forms for classification problems. One of the most prevalent uses of the CNNs is to exploit them as feature extractors and use the resulting feature representations in a separate classification formulation. The input to such networks is required to be of specific size and resolution, which limits their use for domains with variable-sized images. Breast histopathology analysis is one of the domains that faces this limitation. Whole slide breast histopathology images contain variable number of variable-sized ROIs with complex structural information. In this work, we propose a framework that learns the feature representation of an ROI from weighted average pooling of the feature representations of its patches.

3.1 Patch-level Deep Network Training

The patch-level convolutional neural network training is done in two stages. The first step involves the extraction of the fixed-sized patches from the ROIs, and the second step involves the CNN training on the extracted patches. The identification of good patches from an ROI is the first building block to obtain effective ROI-level feature representations. Thus, the patches that are going to be sampled from an ROI need to be informative and diverse enough to represent the ROI. For this, we use the haematoxylin channel estimated from the RGB image using a color deconvolution procedure[9] to locate the nuclei dense regions. A non-parametric Parzen density estimate was built in the image domain by applying a Gaussian window on the haematoxylin values of the pixels, and a threshold was applied to this estimate to eliminate the regions with little to no nuclei. The points inside the resulting nuclei region correspond to the center points of candidate patches. We extract fixed-sized patches inside the ROI from these nuclei-dense regions, achieving diversity in the sampled patches by imposing a maximum overlap constraint on pairs of patches. The number of patches to extract within an ROI is automatically



(a) RGB image of an ROI (b) Haematoxylin estimate of the ROI (c) Mask of nuclei dense regions (d) Selected patches (shown on the RGB image)



(e) Example extracted patches (shown separately)

Figure 2. Patch selection process shown on an example ROI.

determined by the size of the region covered by the nuclei. The patch selection process is presented in Figure 2 on an example ROI. Once the patches are selected, one can train a convolutional neural network on these small fixed-sized patches directly. However, instead of training a network from scratch, we opt to use a pre-trained network and fine-tune its classification parameters in the fully-connected layers using the ROI label as the label of each extracted patch. Due to the limited number of ROIs for patch extraction, only the fully-connected layer parameters are updated during backpropagation, and the kernel coefficients in the convolutional layers are kept frozen.

3.2 ROI-level Deep Feature Representations

Our framework aims to generate feature representations for variable-sized ROIs. The detailed process of learning the feature representation of an ROI is as follows. First, the patches sampled from the ROI are fed to the fine-tuned network, and the deep feature vector of each patch is obtained from the penultimate layer of the network. Similarly, the class probabilities are obtained from the output softmax layer of the network. Then, each class-specific probability is separately used as a weight applied to the components of the feature vector, and the final feature representation of the patch is obtained by the concatenation of the weighted deep feature vectors. This patch-level feature generation procedure is repeated for each patch extracted from the ROI. Finally, the feature representation of the ROI is computed by aggregating the feature representations of its patches by average pooling. A simplified visualization of the procedure of patch-level deep feature generation can be seen in Figure 3.

3.3 Classification

As a pre-processing step before the ROI-level classification, we apply principle component analysis to reduce the redundancy and the dimensionality of the feature representations of the ROIs. Finally, we use the feature representations of the ROIs in the training set to train a multi-layer perceptron (MLP) classifier to perform multi-class classification on unseen ROIs whose feature representations also follow the same procedure.

4. EXPERIMENTS

The data set of 437 ROIs belonging to one of the 4 classes (benign without atypia (UDH), atypical ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS), and invasive cancer (INV)) is split into two equal sized

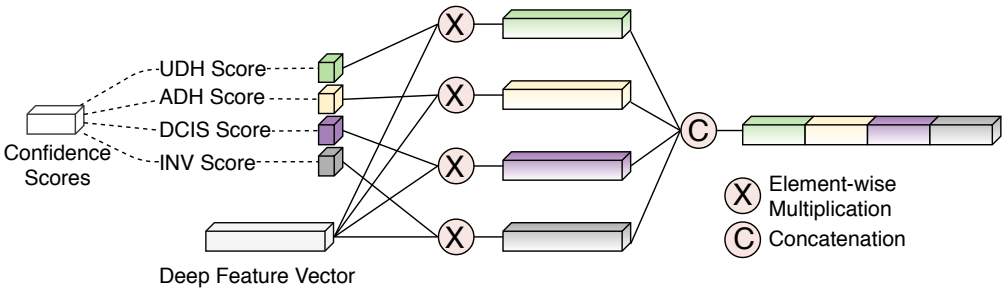


Figure 3. The patch-level deep feature representation is generated from the aggregation of deep feature vectors with class-specific deep network output.

Table 1. The class distribution of the slides and the ROIs in the training and test sets.

		UDH	ADH	DCIS	INV	Total
Slide	Training Set	34	35	41	10	120
	Test Set	22	48	38	12	120
ROI	Training Set	60	58	85	17	220
	Test Set	37	81	80	19	217

sets containing slides from different patients for training and test. The class distribution between the two sets are kept as close as possible. The split was performed based on the slide-level class distribution. The training and test set distributions of the slides and the ROIs for each diagnostic class label is presented in Table 1. The ROI sizes showed a very large variance where the largest ROI in the data set was 2000 times greater than the smallest ROI. The class-specific ROI size statistics in Table 2 show that the ROIs labeled as invasive cancer, on average, spread over the largest area, and the ROIs labeled as atypical ductal hyperplasia, on average, were the smallest in size.

In our experiments, we used the VGG16 network[10] as the base CNN architecture due to its good depth and representational capabilities. However, the proposed method is applicable to any choice of convolutional neural network. To improve the generalization performance of the network, we applied random rotation and random horizontal and vertical flipping as well as random perturbations on the Hue channel in the HSV color space as part of the data augmentation routine. The perturbations on the Hue channel aimed to provide a workaround to the problem of variations in staining. However, the data set was collected from a single whole slide scanner and the variations in staining were expected to be negligible. We fine-tuned the network on the augmented training set, leaving a small portion of the training set for validation. We used batches of 12 patches, employed the Adam optimizer with a learning rate set to $1e{-}4$, and fixed the dropout to 0.75 on the fully-connected layers.

The parameters of the VGG16 network were trained on the patches extracted from the ROIs in the training data according to the procedure described in Section 3.1. The network is then used to extract the feature representations of the ROIs in the training and test sets. We aggregated the penultimate layer activations weighted by each class specific softmax output to represent a patch inside an ROI. Then, we employed average pooling on the patch-level feature representations to obtain the deep feature representation of the associated ROI. We refer to this representation as Agg-Penultimate, as detailed in Section 3.2. Finally, the deep feature representations in the training data were used to train a 4-class multi-layer perceptron (MLP) classifier on a 3-fold cross-validation setting. The feature representation of an unseen ROI follows the same feature extraction routine, and the MLP classifier is used to predict the label of the ROI. We compared the performance of the

Table 2. The average and standard deviation of the sizes of the ROIs, as well as the ratio of the largest ROI to the smallest one (max-min ratio) for each diagnostic label. The values for mean and standard deviation correspond to the number of pixels at $10\times$ magnification.

	UDH	ADH	DCIS	INV
Mean	$6460K$	$3437K$	$7857K$	$36785K$
Standard deviation	$9093K$	$6364K$	$14543K$	$47711K$
Max-min ratio	780.0	930.3	1170.5	414.8

Table 3. The ROI-level classification performance comparison.

Method	Accuracy
Pathologists[11]	0.700
Max-Pooling[2]	0.548
Decision-Fusion[4]	0.649
Y-Net[7]	0.625
Base-Penultimate	0.622
Agg-Penultimate	**0.668**

proposed algorithm with the following methods:

- Base-Penultimate: A patch-level feature representation is extracted directly from the penultimate layer activations of the fine-tuned network. The feature representations of the patches inside an ROI were aggregated by average pooling to obtain the feature representation of the ROI without involving class-specific scores. An MLP classifier was trained on the ROI-level feature representations and the labels in the training set.
- Max-Pooling: The ROI-level feature representation is obtained through pooling the patch-level class probabilities of the fine-tuned network.[2] From the set of patch classification scores, the ones that were below a confidence threshold were discarded. Then, the predictions were filtered through a frequency threshold on the pixel-level histogram of the labels so that the predictions with insufficient coverage in the ROI were eliminated. Finally, the most severe diagnostic label remaining determined the label of the ROI. In our experiments, we used 0.50 as the classification threshold and set the frequency threshold to 0.25.
- Decision-Fusion: The class probabilities of the patches from the final layer of the network were summed over each patch for each class label to create a class frequency histogram for the ROI.[4] Then, the four dimensional frequency vector was used to train an MLP classifier.
- Y-Net: A deep network that combined semantic tissue segmentation with discriminative patch selection[7] was trained on the same data set used in this paper, and was directly used for comparison.

The performance of the proposed approach and the comparisons are shown in Table 3. The proposed method, Agg-Penultimate, achieved the best accuracy score of 0.668 on the four-class classification problem. Our method improved the base accuracy of 0.622 obtained from the aggregation of the patch-level features without class-specific scores by 4.8%. Between the pooling based approaches, Max-Pooling and Decision-Fusion, the latter performed better due to training a classifier on the pooled probabilities instead of only rule-based pooling.

The classification performance of the Agg-Penultimate features in the form of a confusion matrix are presented in Table 4. Out of the four classes, benign without atypia (UDH), atypical ductal hyperplasia (ADH), ductal carcinoma in situ (DCIS), and invasive cancer (INV), the classifier could predict ADH and DCIS better than UDH and INV. The classifier was able to predict the ROIs with ADH 56 out of 81 times, misclassifying only 15 of them as UDH and 10 of them as DCIS. Additional class-specific statistics for the classifier were given in Table 5. The classifier achieved the best precision value on ADH, only misclassifying 19 out of the 75 ROIs as ADH. The ROIs with DCIS were best captured by the classifier compared to other classes, misclassifying only 13 out of the 80 ROIs. The precision performance on DCIS was below ADH, with the classifier favoring DCIS in more cases, whereas majority of the incorrectly labeled ROIs with UDH were predicted as ADH, followed by DCIS. In addition, the ROIs with INV were correctly predicted in seven ROIs compared to the 10 ROIs which were predicted as DCIS which also makes sense given that a large number of cases with INV also involved DCIS in their pathology reports.[12]

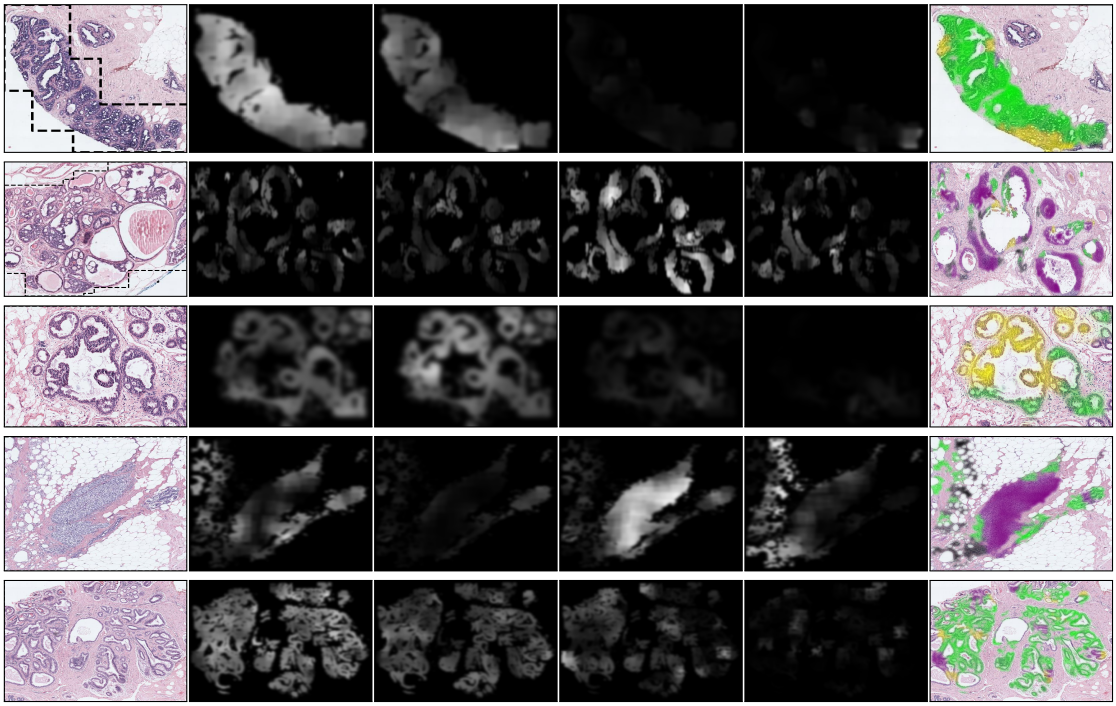
We present the patch-level CNN predictions and class-specific scores from the fine-tuned VGG16 network on example ROI images in Figure 4. The ROIs in the first three rows with consensus labels, UDH, DCIS, and ADH, had very good patch level CNN predictions as most of the individual patches were predicted as the correct class by the network outputs only. The methods involving the proposed feature representations, Agg-Penultimate

Table 4. Confusion matrix of the proposed method with Agg-Penultimate features for ROI-level classification.

		Predicted			
		UDH	ADH	DCIS	INV
True	UDH	15	12	8	2
	ADH	15	56	10	0
	DCIS	5	6	67	2
	INV	1	1	10	7

Table 5. Class-specific statistics on the performance of the proposed method with Agg-Penultimate features for ROI-level classification. The number of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) are given. Precision, recall (also known as true positive rate and sensitivity), false positive rate (FPR), specificity (also known as true negative rate) and F-measure are also shown.

Class	TP	FP	FN	TN	Precision	Recall/ Sensitivity	FPR	Specificity	F-measure
UDH	15	21	22	159	0.417	0.405	0.117	0.883	0.411
ADH	56	19	25	117	0.747	0.691	0.140	0.860	0.718
DCIS	67	28	13	109	0.705	0.837	0.204	0.796	0.766
INV	7	4	12	194	0.636	0.368	0.020	0.980	0.467



(a) RGB (b) UDH (c) ADH (d) DCIS (e) INV (f) Predictions

Figure 4. Patch-level classification outputs from CNN on an example ROI. From left to right: RGB image of an ROI; scores for individual classes, UDH, ADH, DCIS, INV; predicted classes from the fine-tuned VGG16 network. UDH is shown as green, ADH is shown as yellow, DCIS is shown as purple, and INV is shown as gray.

and Base-Penultimate, as well as the methods used for comparison, Max-Pooling and Decision-Fusion, correctly classified the ROIs as UDH, DCIS, and ADH, respectively. However, when the ROIs contained high-confidence patch-level predictions different from the consensus diagnoses, the compared methods performed poorly as seen on the ROIs in the fourth and the fifth rows. The proposed method, Agg-Penultimate, was able to classify the ROI in the fourth row as DCIS whereas only Decision-Fusion out of the rest of the methods was able to correctly predict DCIS. The fifth row demonstrated one of the most interesting cases where the most dominant patch-level predictions did not agree with the ROI-level consensus label, ADH. When the individual class scores were investigated, the network was not certain in some areas and the majority of the patches were predicted as UDH, instead of ADH. The proposed method that considered patch-level network outputs for each class as well as the features extracted from the penultimate layer of the network was able to correctly classify the ROI as ADH while all of the compared methods failed. The proposed approach was able to handle such cases even when the patch-level predictions were completely different from the target diagnosis of the ROI. The quantitative as well as the visual evaluation showed that the patch-level CNN predictions were not individually representative of the ROI structure and more information from the network and the class-specific network outputs improved the ROI-level classification performance.

5. CONCLUSIONS

Convolutional networks are often trained on fixed-sized small patches and require the input patches to be exactly the same size during testing. Whole slide breast histopathology images containing ROIs with drastically different sizes and shapes make the process of training the networks directly on these ROIs difficult. We proposed a framework to generate deep feature representations for variable-sized ROIs in breast histopathology images. The proposed method extracted fixed-sized patches from potentially relevant areas in the ROI automatically. The structural information preserved in a patch and the class probability distribution of the patch were considered in the generation of its deep feature representation by concatenating the weighted penultimate layer activations with the class probability scores from the softmax layer. The final ROI-level representation was obtained by average pooling of the patch-level representations. We demonstrated the representation power of the proposed approach in comparative experiments on a breast pathology data set. Our future work involves extensions of the deep representation in an end-to-end framework.

ACKNOWLEDGMENTS

C. Mercan and S. Aksoy were supported in part by the Scientific and Technological Research Council of Turkey (grant 117E172) and in part by the GEBIP Award from the Turkish Academy of Sciences. E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore were supported in part by the National Cancer Institute of the National Institutes of Health (awards R01-CA172343, R01-140560, and KO5-CA104699). The content is solely the responsibility of the authors and does not necessarily represent the views of the National Cancer Institute or the National Institutes of Health. C. Mercan and S. Aksoy also gratefully acknowledge the support of NVIDIA Corporation with the donation of a Quadro P6000 GPU used for this research.

REFERENCES

[1] Mercan, C., Mercan, E., Aksoy, S., Shapiro, L. G., Weaver, D. L., and Elmore, J. G., "Multi-instance multi-label learning for whole slide breast histopathology," in [*Proceedings of SPIE Medical Imaging Symposium, Digital Pathology Conference*], (February 27–March 3, 2016).

[2] Mercan, C., Aksoy, S., Mercan, E., Shapiro, L. G., Weaver, D. L., and Elmore, J. G., "Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images," *IEEE Transactions on Medical Imaging* **37**, 316–325 (January 2018).

[3] Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., and Chang, E. I.-C., "Deep learning of feature representation with multiple instance learning for medical image analysis," in [*IEEE International Conference on Acoustics, Speech and Signal Processing*], 1626–1630 (2014).

[4] Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., and Saltz, J. H., "Patch-based convolutional neural network for whole slide tissue image classification," in [*IEEE Conference on Computer Vision and Pattern Recognition*], 2424–2433 (2016).

[5] Bejnordi, B. E., Zuidhof, G., Balkenhol, M., Hermsen, M., Bult, P., van Ginneken, B., Karssemeijer, N., Litjens, G., and van der Laak, J., "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images," *Journal of Medical Imaging* **4**(4), 044504 (2017).

[6] Mehta, S., Mercan, E., Bartlett, J., Weaver, D., Elmore, J., and Shapiro, L., "Learning to segment breast biopsy whole slide images," in [*IEEE Winter Conference on Applications of Computer Vision*], (2018).

[7] Mehta, S., Mercan, E., Bartlett, J., Weaver, D., Elmore, J. G., and Shapiro, L., "Y-net: Joint segmentation and classification for diagnosis of breast biopsy images," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], (2018).

[8] Gecer, B., Aksoy, S., Mercan, E., Shapiro, L. G., Weaver, D. L., and Elmore, J. G., "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks," *Pattern Recognition* **84**, 345–356 (December 2018).

[9] Ruifrok, A. and Johnston, D., "Quantification of histochemical staining by color deconvolution," *Analytical and Quantitative Cytology and Histology* **23**(4), 291–299 (2001).

[10] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," in [*International Conference on Learning Representations*], (2015). arXiv:1409.1556.

[11] Elmore, J. G., Longton, G. M., Carney, P. A., Geller, B. M., Onega, T., Tosteson, A. N. A., Nelson, H. D., Pepe, M. S., Allison, K. H., Schnitt, S. J., O'Malley, F. P., and Weaver, D. L., "Diagnostic concordance among pathologists interpreting breast biopsy specimens," *Journal of American Medical Association* **313**(11), 1122–1132 (2015).

[12] Mercan, E., *Digital Pathology: Diagnostic Errors, Viewing Behavior and Image Characteristics*, PhD thesis, University of Washington, Seattle, Washington (2017).