MVPNets: Multi-Viewing Path Deep Learning Neural Networks for Magnification Invariant Diagnosis in Breast Cancer

Padmaja Jonnalagedda[1], Daniel Schmolze[3], Bir Bhanu[1,2]
[1]Department of Electrical Engineering, University of California at Riverside, Riverside, CA 92521, USA
[2]Department of Bioengineering, University of California at Riverside, Riverside, CA 92521, USA
[3]Department of Pathology, City of Hope, Duarte, CA 91010, USA
E-mail: sjonn002@ucr.edu, dschmolze@coh.org, bhanu@cris.ucr.edu

Abstract— **Breast cancer diagnosis requires a pathologist to analyze the histology slides under various magnifications. An automated diagnosis method to aid pathologists that is magnification independent will significantly save time, reduce cost and mitigate subjectivity and errors in current histopathological diagnosis procedures. This paper presents a deep learning network, called MVPNet and a customized data augmentation technique, called NuView, for magnification independent diagnosis. MVPNet is tailored to tackle the most common issues (diversity, relatively small size of datasets and manifestation of diagnostic biomarkers at various magnification levels) with breast cancer histology data to perform the classification. The network simultaneously analyzes local and global features of a given tissue image. It does so by viewing the tissue at varying levels of relative nuclei sizes. MVPNet has significantly less parameters than standard transfer learning deep models with comparable performance and it combines and processes local and global features simulatenously for effective diagnosis. Additionally, NuView extracts tumor nuclei location and points the attention of MVPNet to the informative region specifically. The method gives an average magnification independent classification accuracy of 92.2% as compared to 83% reported in literature on the BreaKHis database.**

Keywords - Deep convolutional networks, multi-view network, tumor location, Gaussian mixture models

I. INTRODUCTION

Breast cancer persists as the most common form of cancer in women, with reports projecting high incidence rates and high mortality and morbidity [1]. While researchers point positively towards a reduction in mortality rates, the steeply increasing numbers call for a higher diagnostic power with a feasibility in daily clinical routine.

Histopathological image analysis has been the basis of breast cancer diagnosis for decades. A common practice is to stain the tissues with Hematoxylin and Eosin (H&E) in order to observe them more clearly. Despite advances in digital imaging techniques, majority of breast cancer diagnosis is performed by pathologists manually examining the tissue slides under a microscope. This task is time consuming as pathologists have to scan the tissue slide at multiple magnifications to identify biomarkers. With breast cancer incidence reports consistently projecting increasing numbers, the individual cases manifest in complex forms, making the task of diagnosis a complex one. Protocol, therefore, suggests the pathologists to look at more number of quantitative biomarkers, making the diagnosis subjective. These biomarkers are magnification dependent, with features like infiltrative patterns, size and invasiveness of tumors revealing in lower magnifications and individual mitotic activity showing in higher magnifications. Assessment of each individual tissue for all features at multiple magnifications also makes this a time consuming process. Breast cancer tissues have very subtle differences amongst different classes in texture, color and morphology which makes breast cancer histology data an extremely diverse one. Pathologists need to possess rich experience and knowledge to assess the cancer tissues with accuracy and finesse. This also causes inter-pathologist variability. There is a large window for errors and variations in the diagnosis owing to human errors and pathologist experience, adding to the inconsistency and variability. Research suggests that pathologists prefer to analyze tissue slides under a microscope as compared to analyzing digital images; with no appreciable change in accuracy. Storage of these glass slides raises the cost of diagnostic process.

To be a practical aid for pathologists, an automated method should diagnose cancer at any magnification, without any additional immunohistochemical markers or human intervention. This would make the diagnosis fast, low-cost, objective and consistent and can be introduced in daily routine to address rising breast cancer incidence.

In this paper, the authors propose a method that aims to tackle these problems using a deep learning system. The network combines high and low level features to be processed together and integrates learning principles of modular residual networks to avoid overfitting and cater to smaller datasets, point convolutions for dimensionality reduction and widening the network to handle the highly diverse nature of breast cancer histology.

II. RELATED WORK AND CONTRIBUTIONS

A. *Related work*

There has been extensive work in automating tumor detection. The research community has identified automated diagnosis as a vital need. Early work attempted binary classification of a tumor into benign and malignant by extracting biomarkers such as nuclei shape, size, mitosis etc. and applying machine learning classifiers. Ciresan et al., detected mitotic activity [3] which is one of the features used to detect cell types in tumor [4, 5, 6]. These hand-crafted features have also been used in past research [7, 8, 9, 10] for classification of tumors. Doyle et. al., [11] have attempted to achieve cancer detection using structural and textural features such as Haralick features, Gabor filters and graph

2471-7819/18/$31.00 ©2018 IEEE
DOI 10.1109/BIBE.2018.00044

methods such as Voronoi diagrams. ROI extraction using top-bottom hat transform for segmentation and SVM for subsequent classification has been done by Wang et. al. [12].

In recent years, CNNs have played a major role in breast cancer detection tasks [13, 14, 15] instead of hand crafted features. Xu et al., [16] have successfully pointed out the role played by the regions surrounding nuclei like stroma and tubules and have also used deep networks to achieve tumor classification. Havaei et. al. [17] used CNNs to locate tumors by passing two different sized patches and multiple paths of different kernel sizes to achive tumor segmentation in greyscale MRI images. Despite increasingly sophisticated methods being introduced to automate the tumor classification and segmentation processes, very little has been done to address magnification invariant classification. All the above mentioned methods achieve detection with a prior knowledge of the incoming tissue slide's magnification. Bayramoglu et. al., have identified the need for a standalone model for detection that does not need a-priori information of the tissue resolution and proposed a method using CNNs [22]. Since breast cancer histology is considered to be extremely diverse and complex [2] as illustrated in Fig. 1, with each magnification posing a new set of distinguishable features as can be seen in Fig. 2, a lot of preprocessing is involved in most of these methods. If not, a different model is used at each of the resolutions [18]. Making the classification magnification dependent has been a way to control one of the many variabilities that come with the data. Removing this dependency gives another degree of robustness to automated diagnostics.MVPNet with NuView data augmentation, proposed below, tackles these issues.

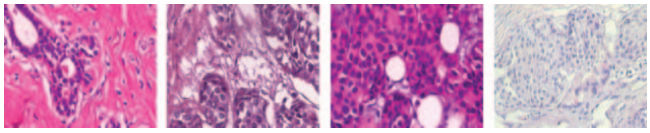

Figure 1: Diversity in a class of Breast Cancer (malignant).

B. *Contributions*

The important contributions of the proposed method are: (1) Use of multiple viewing paths in MVPNet that can view the tissue in different relative nuclei sizes, so as to handle multi-level feature extraction simultaneously (2) Use of NuView data augmentation method to introduce random magnifications using contextual information provided by location of tumor regions which makes learning robust to changes in resolution (3) Striking a balance between depth and width of a network for optimal classification performance. The depth is decided by the number of modular residual network style blocks. The width is decided via the use of 1x1 convolutions and multiple densely connected layers. A description of each of these techniques and their impact has been provided in detail in Sections III and IV.

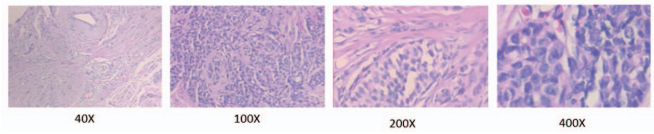

Figure 2: Magnifications of same tissue slide in the BreaKHis database.

III. TECHNICAL APPROACH

A. *Overview of algorithm*

To successfully detect tumor types, MVPNet, with NuView emulate the manual dexterity of viewing tissue slides and pathologists' ability to quantitatively analyze biomarkers at multiple magnifications in histopathology images. A tumor is detected by inspecting and quantitatively analyzing various biomarkers like nuclei shape, size, uniformity, tumor location, shape, infiltrative growth patterns etc. All of these indicators manifest at different resolutions. To address this issue, a framework should be able to view and process both high and low level features of a tissue slide. MVPNet achieves this by passing the same tissue slide parallelly through two paths. The multi-path approach mimics the manual perceptive abilities by incorporating local and globalview of the features. Mathematically, this is achieved by having different kernel sizes for filters in parallel paths. Smaller kernel size looks at low level features such a nuclei size, shape, mitotic figures etc., while the larger kernel analyzes the infiltrative growth patterns, tumor size etc. The kernel sizes in current network are 7x7 and 11x11. These kernel sizes account for the wide range of magnifications (40X to 400X) in the training set. To make the network robust to any new resolutions that might be available in the future, a customized data augmentation (NuView) is employed to efficiently locate tumor regions and extract patches with randomized magnifications from that location; thereby ensuring high information flow into the network. To account for the high diversity of breast cancer data (Fig. 1), the network is widened (more number of neurons) by adding multiple fully connected layers. Typically, deeper networks reuse features and wider networks have higher capacity to learn newer features. The network is composed of multiple modular residual blocks that automatically bypass unused modules in the face of smaller datasets, and can cater to larger datasets should new data be available in the future. An overview of the algorithm with custom data augmentation (NuView) is shown in Fig. 3. In addition to standard data augmentation, the patches of various magnifications on tumor location are obtained. They are further subjected to standard data augmentation, constituting an enriched and balanced training set. Then classification is done using MVPNet.

B. *NuView augmentation*

In histopathology images, nuclei are by far the most significant indicators of tumor. The data augmentation technique, termed NuView, in this proposed system is customized to supplement the benefits of standard operations for data augmentation such as random zooming, cropping, resizing of training data. NuView leverages the location of the areas where most nuclei are clustered, indicating where the tumor is localized. Bringing the attention of the network to these areas can boost the performance of automated diagnosis. The images in the training set are subjected to clustering using a mixture of Gaussian models (GMM).

GMM assumes that an overall population is a mixture of a finite number of components. In this segmentation, the tissue is segmented into three components, one of them being nuclei. This clustering is done using the Expectation-Maximization (EM) algorithm. EM assumes random initial cluster centers and variances and each point is assigned a (posterior) probability that the point belongs to that cluster. Thus, each point belongs to all clusters with some probability. Using this and the priors of each cluster (set to uniform for the first iteration), new cluster centers and variances are found. Initial cluster centers and variances are initialized to preset values. This process is repeated until convergence is obtained, i.e., the likelihood becomes constant. This locates the nuclei as shown in Fig. 4.

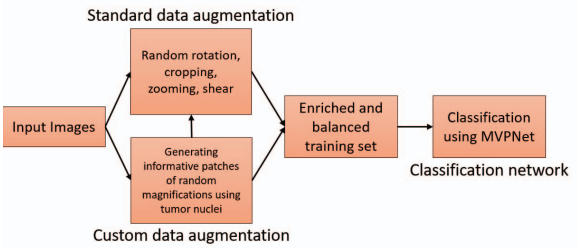

Figure 3: Overview of proposed algorithm as described in Section IIIA.

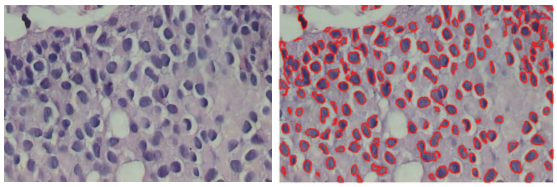

Figure 4: EM based segmentation for nuclei extraction at 200X.

The mask that extracts nuclei is used to find the region with most information in the tissue slide. This is done by nuclei counting, or effectively, the area with the highest concentration of nuclei. An image displaying one of the clusters contains only nuclei, called the *intermediate image*. Across this intermediate image, a patch of desired size slides to find the area with the maximum number of nuclei. That area is then extracted and resized to the original image (see Fig. 5). The size of the patch determines the magnification of the resultant informative patch. In Fig. 5, the patch used is half the size of an original image at 100X magnification, thus the resultant patch is generated at 200X magnification.

This informative patch is added to the training set the bolster it. Depending on the size of patches, randomized magnifications can be added to the new training set as needed. An added advantage of this method is to tackle the imbalanced class problem. If proportionally higher number of patches are extracted from the less dominant class images, it balances both classes to avoid classification bias towards more dominant classes. The importance of using the contextual information of nuclei concentration as random resolution patches in data augmentation step can be seen in Section V where results are compared to no patches (just standard data augmentation) and using randomly magnified patches without using contextual information in extracting them (random patches). The network architecture used in the proposed model has been described in detail in Section IV.

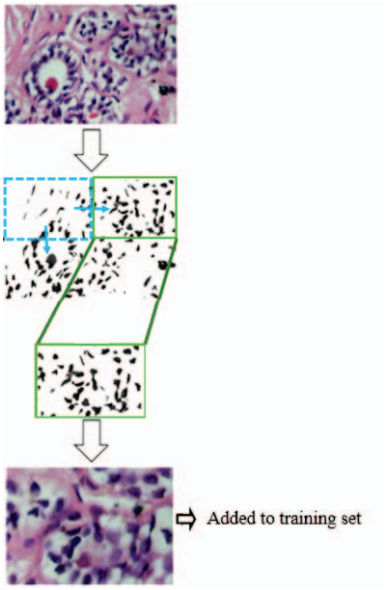

Figure 5: Localizing significant tumor regions using GMM. The blue dotted box shows the sliding window and the green solid box shows the subsequent informative patch selected.

IV. MODEL ARCHITECTURE

The augmented training set is given as an input to MVPNet for magnification invariant classification. This paper proposes a multi-view architecture that has the capacity to view regional and global features of the tissue at the same time. The following sub-sections describe the use of various functionalities of the network.

A. Multi-view module

To assess features at varying level of relative nuclei size at the same time, the multi-view block is designed. While a single stack of multiple convolutional layers can analyze a set of features end to end at a given time, this block sends the same input into two parallel blocks of varied convolution kernel sizes (11x11 and 7x7 kernel sizes). The path with higher kernel size views a relatively more holistic view such as infiltrative growth pattern as compared to the smaller kernel size which views more regional features, native to mitotic activity. A typical multi-view block looks as shown in Fig. 6. The 1x1 convolutions help in dimensionality reduction and are explained in detail in Section IV C. Each of the convolutional block is followed by a maxpooling block with stride 2 and size of kernel $(k+1)/2$, where k is the kernel size of convolutional filter. This is to keep the size of activation map consistent in the two paths.

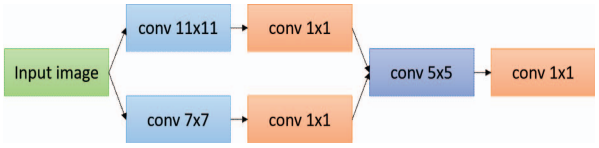

Figure 6: Multi-view module.

B. *Identity and projection blocks*

Identity blocks are a key feature introduced in residual networks and have many advantages in deep networks [19].

An identity block contains an identity path that can pass information through the convolutional path or the identity path based on the amount of information learnt in that block, avoiding over learning of features. Each identity block acts as its own mini-network and can be added or removed according to need – which provides modularity to the network. This modular structure with identity path also helps in avoiding vanishing and exploding gradient problems.

If the identity path in the above block is replaced with a projection (1x1 convolution), some extra features may be learnt at the cost of high dimensionality and time complexity. To avoid possible degradation problem, it is practice to commonly introduce a few projection blocks after every few identity blocks. Every few identity blocks are followed by a projection block to facilitate and regulate the learning process. In MVPNet, a projection block followed by a predefined number of identity blocks is collectively called a *combinatorial module*. Typical 64 filter identity and projection blocks are shown in Fig. 7(a) and (b) respectively.

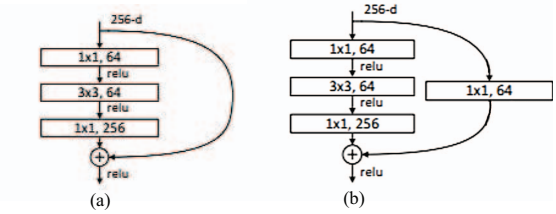

Figure 7: (a) Identity block and (b) Projection block, both with 64 filters.

C. *Use of 1x1 convolutions*

1x1 convolutions were initially proposed in the "Network in Network" paper by Lin et. al. [23]. These blocks with 1x1 kernel size achieve dimensionality reduction as well as makes the network orders of magnitude faster and less memory intensive. Typically, a 1x1 layer has 9X fewer parameters than a 3x3 block. This block helps reduce overall depth of a network, since adding one of these is mathematically equivalent to adding a multi-layer perceptron [23]. This can also be effectively viewed as making the network denser since each block acts like a fully connected network. Each of 1x1 convolution is followed by ReLU activation, therefore, adding to non-linearity in feature mapping. Summarizing, the 1x1 convolutions help build the balance between depth and width, make it less memory intensive and faster with a locally denser and non-linear feature mapping and lower information loss.

D. *Specifications of model architecture*

The full model architecture is shown in Fig. 9. The input of the network is given to the first section of MVPNet - *the multi-view module*. To accommodate all resolutions in the range of 40 to 400X, the filter sizes chosen are 11x11 and 7x7 in each of the two paths. A stride of 2x2 is used to prevent large activation maps. The convolutions in each path are followed by 1x1 convolutions, immediately followed by ReLU activation. Once the features from both paths are combined, a convolutional block with a 5x5 kernel size and a 1x1 convolution is used. This entire block is collectively called the multi-view module. In each of these convolutions, 64 filters have been used. The second section of MVPNet follows the multi-view module, and is a series of 3 combinatorial modules. The first combinatorial module consists of three blocks - one projection and two identity blocks. The number of filters in each of these blocks is 64, 64 and 256 (Fig. 8). The next combinatorial module has 128, 128, 512 filters with five blocks – one projection and four identity blocks. The third and final combinatorial module has 256, 256, 1024 filters with 3 blocks – one projection and two identity. Maxpooling is done after every convolution in the multi-path module and batch normalization is employed after convolution in all combinatorial modules. The kernel size for all the filters in the combinatorial modules is 3x3. The combinatorial modules make the network deeper. After these modules, an average pooling layer with 7x7 kernel size is added. The final section of MVPNet has three fully connected layers, which in combination with the multi-view block effectively make it wider. The first two fully connected (FC) layers have 1024 and 512 neurons. All the layers up to this point have ReLU activation. The third FC layer has # of neurons equal to # of classes. This has softmax activation.

Figure 8: Order of combinatorial modules used in MVPNet.

The input images in Fig. 9 come from the balanced training set shown in Fig. 3. The multi-view module is shown in Fig. 6. The (64) in the figure refers to number of filters. This is followed by the combinatorial module as shown in Fig. 8. The features from these modules go into 2 fully connected layers of 1024 and 512 neurons each. This output is given to the softmax classifier.

E. *Model training specifications*

Stochastic Gradient Descent (SGD) optimizer is used to train the MVPNet from scratch with an adaptive learning rate function. Adam optimizer was also tested which converges faster but doesn't always converge to the best solution when trained from scratch. The weights are initialized according to He initialization [20] and random normal distribution (with no major change in performance). He initialization draws samples from a truncated normal distribution with mean 0 and standard deviation given by:

$$\sqrt{2/f_{in}}$$

where f_{in} is number of input units in the weight of that layer.

Among many batch sizes tested, the batch size used is 24. Nesterov criteria is applied to the optimizer and the training runs for a maximum of 1000 epochs. All input images are resized to 224x224. The loss used is the categorical cross-entropy loss and the model with the best performance is saved among all epochs.

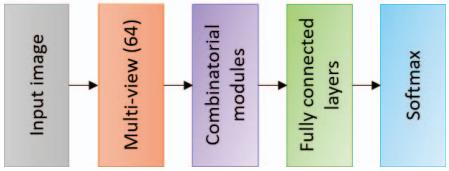

Figure 9: Complete MVPNet architecture.

V. RESULTS

A. Description of dataset

The BreaKHis database has data from 82 anonymous patients from the Pathological Anatomy and Cytopathology (P&D) Lab, Brazil [18]. BreaKHis has data from various tumors divided into 4 magnifications: 40X, 100X, 200X and 400X, as shown in Fig. 2, and is comprised of 7909 H&E stained surgical biopsies, (460x700 pixels, RGB images) divided into 2480 benign and 5429 malignant tumor images. The training/validation and testing sets are divided as 70% and 30% such that data from one patient is not simultaneously included in both training and testing sets.

B. Extraction of contextual information

The extraction of contextual information is the most crucial step in the customized data augmentation. This requires segmenting the tissue slide so as to obtain the location of nuclei clusters. The training set is augmented using these segmented results (see Section III). The results of EM segmentation at 200X and 400X magnifications are shown in Fig. 10. It can be seen that nuclei clusters are well detected and are subjected to the process in Section III.

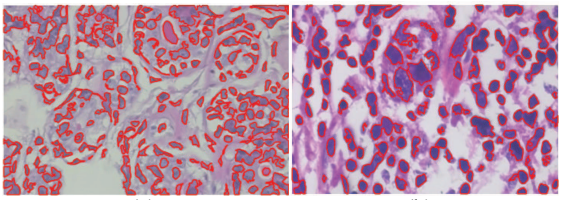
(a) (b)

Figure 10: Segmentation results (a) benign at 200X, (b) malignant at 400X.

C. Magnification invariant classification results

Both MVPNet and ResNet50 are trained and tested using 7-fold cross-validation with and without employing NuView. HSV and YCbCr spaces mitigate the illumination and redundant color effects and thus, they have been used here to show comparison with the proposed method. The model performs just as well with any color space implying it uses the parallel views to obtain a holistic view of features and is not dependent on color for its classification. A comparison of results for magnification dependent classification is offered in research [21]. AlexNet results provided [21] are magnification dependent, hence, comparison in Table 1 is done by considering average AlexNet performance across all magnification levels. This gives an upper bound on its performance considering the AlexNet were to give optimum performance at individual magnification level when all magnifications are combined in the training set. Thus, the error percentage in AlexNet does not come from 7-fold validation but the performance average over all magnifications. Bayramoglu et. al. [22] has reported results on 5 fold cross validation. The results are for magnification invariant detection, but separated by magnification levels. Table 1 shows the average of all results.

In Table 1, SDA stands for Standard Data Augmentation without addition of any informative patches. *Random patches* in Table 1 is SDA augmented by randomly selected patches unlike NuView which extracts informative patches. Accuracy in Table 1 is defined as the ratio of correctly classified instances over all instances, i.e., the percentage of correct classifications. Sensitivity is defined as the true positive rate, or the percentage of actual malignant tumors that are correctly classified as such. The specificity in the table is defined as the true negative rate, measuring the actual benign tumors being correctly identified as such.

TABLE 1: RESULTS OF MAGNIFICATION INVARIANT CLASSIFICATION

Model	Accuracy	Sensitivity	Specificity
Bayramoglu et al [22]	83.2 ± 3.1%	NA	NA
AlexNet	87.2 ± 5.5%	NA	NA
ResNet50+NuView+ RGB images	92.5 ± 1.4%	94.6 ± 1.4%	91.2 ± 2.2%
ResNet50+random patches+RGB	89.9 ± 1.8%	94.3 ± 1.7%	88.3 ± 2.1%
MVPNet+NuView+ RGB images	92.2 ± 1.6%	94.2 ± 2.2%	92.3 ± 2.4%
MVPNet+NuView+ HSV images	90.3 ± 1.7%	92.6 ± 2.1%	91 ± 1.8%
MVPNet+NuView+ YCbCr images	91.1 ± 1.6%	91.4 ± 2.3%	92.6 ± 1.7%
MVPNet+random patches+RGB	88.3 ± 2.1%	84.1 ± 1.9%	93.3 ± 2.3%
MVPNet+SDA+ RGB	84.1 ± 2.2%	80.5 ± 1.8%	94.2 ± 2.4%

The average values of ROC plots and standard deviation over 7 folds are given in Fig. 11. The threshold to obtain the ROC plot in this case has been varied from 0 to 1 over steps of 0.1. It can be seen from the Area Under the Curve (AUC) that the performance for MVPNet is comparable to ResNet50. The ROC plots are shown for MVPNet with NuView data augmentation and ResNet50 with NuView. Both results are obtained on RGB input images.

D. Model parameters and memory requirements

MVPNet has approximately 10.4M parameters, as compared to 60M parameters in AlexNet and about 26M parameters in ResNet50. Number of parameters to be learnt during the training of data depend on number of layers and number of neurons per layer. It can be seen that for almost comparable performance, MVPNet uses 2.6 times less number of parameters as ResNet50 and further less than AlexNet. This reduces the memory requirements of the model when storing only weights of the model, with MVPNet requiring less half as much memory to store its weights as compared to ResNet50. MVPNet needs about 4-5 mins to train per epoch and testing takes about 0.05 sec/image using NVIDIA GeForce GTX1080 graphics card.

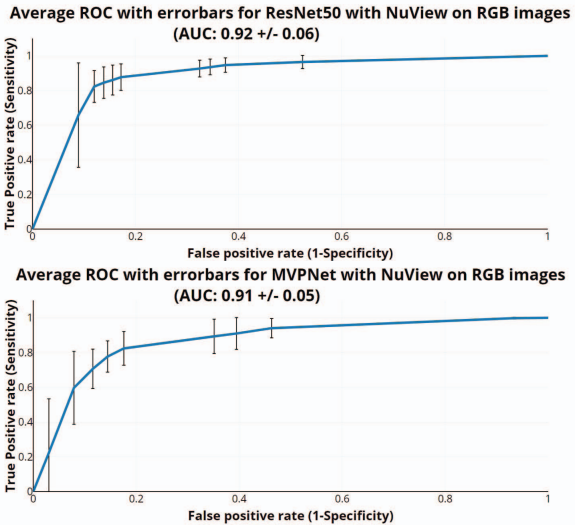

Figure 11: ROC plot for 7-fold validation of ResNet50 and MVPNet under similar conditions.

TABLE 2: MODEL PARAMETERS

Model	# parameters	Model size in memory
AlexNet	60M	240MB
ResNet50	~27M	103MB
MVPNet	~10M	41.3MB

VI. CONCLUSIONS

MVPNet, using the NuView data augmentation scheme, surpasses the current standard in automated, magnification invariant, breast cancer histology classification task. The use of contextual information of tumor location and nuclei density in a tissue image improves learning in a cancer detection task as compared to standard data augmentation. Widening the network with multiple viewing paths instead of only depth in a network (like ResNets) significantly reduces the training and memory requirements without appreciable loss in performance, increasing the learning capability of the network. This is achieved in a fast, consistent and objective way; without any human intervention, thus, improving the reliability and robustness of automated methods for use in clinical pathology.

ACKOWLEDGEMENT

This work is supported in part by Bourns Endowment funds.

REFERENCES

[1] Stewart, B. W. & Wild, C. World cancer report 2014. International agency for research on cancer, WHO, 505, 2014.

[2] Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S. Breast cancer multi-classification from histopathological images with structured deep learning model. Scientific reports. June 2017.

[3] Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. In MICCAI 2013.

[4] Bayramoglu N, Kannala J, Heikkilä J. Human epithelial type 2 cell classification with convolutional neural networks. In IEEE BIBE, 2015.

[5] Bayramoglu N, Heikkilä J. Transfer learning for cell nuclei classification in histopathology images. In ECCV, 2016.

[6] Mao Y, Yin Z, Schober JM. Iteratively training classifiers for circulating tumor cell detection. In IEEE ISBI, 2015

[7] Lo Loukas C, Kostopoulos S, Tanoglidi A, Glotsos D, Sfikas C, Cavouras D. Breast cancer characterization based on image classification of tissue sections visualized under low magnification. Comp. & Math Methods in Medicine. 2013.

[8] Peikari M, Gangeh MJ, Zubovits J, Clarke G, Martel AL. Triaging diagnostically relevant regions from pathology whole slides of breast cancer: A texture based approach. IEEE TMI. 2016 Jan;35(1):307-15.

[9] Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological image analysis: A review. IEEE reviews in biomedical engineering. 2009;2:147.

[10] Wan S, Huang X, Lee HC, Fujimoto JG, Zhou C. Spoke-LBP and ring-LBP: New texture features for tissue classification. In IEEE ISBI, 2015.

[11] Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszewski J. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In IEEE ISBI, 2008.

[12] Wang P, Hu X, Li Y, Liu Q, Zhu X. Automatic cell nuclei segmentation and classification of breast cancer histopathology images. Signal Processing. May 2016.

[13] Cruz-Roa A, Basavanhally A, González F, Gilmore H, Feldman M, Ganesan S, Shih N, Tomaszewski J, Madabhushi A. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In SPIE Medical Imaging 2014: Digital Pathology, Vol. 9041, 2014.

[14] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of pathology informatics. 2016;7.

[15] Spanhol FA, Oliveira LS, Petitjean C, Heutte L. Breast cancer histopathological image classification using convolutional neural networks. In IEEE IJCNN, 2016 (pp. 2560-2567). I

[16] Xu J, Luo X, Wang G, Gilmore H, Madabhushi A. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing. May 2016, 26;191:214-23.

[17] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with deep neural networks. Medical image analysis. 2017 Jan 1;35:18-31.

[18] Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. IEEE TBME 63(7) 2016.

[19] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE CVPR 2016.

[20] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In IEEE ICCV 2015.

[21] Wei B, Han Z, He X, Yin Y. Deep learning model based breast cancer histopathological image classification. In IEEE Cloud Computing and Big Data Analysis (ICCCBDA), 2017 .

[22] Bayramoglu N, Kannala J, Heikkilä J. Deep learning for magnification independent breast cancer histopathology image classification. In ICPR 2016, pp. 2440-2445.

[23] Lin M, Chen Q, Yan S. Network in network. ArXiv preprint arXiv:1312.4400. 2013 Dec 16.