International Journal of Advanced Science and Technology
Vol. 29, No. 08, (2020), pp. 2229- 2234

Low Computational Model for Classifying Medical Histology Images.

Aakash Garg[1]
Software Engineering
Delhi Technological University
New Delhi, India
aakashgarg bt2k16@dtu.ac.in

Karan Aggarwal[1]
Software Engineering
Delhi Technological Univeristy
New Delhi, India
karanaggarwal bt2k16@dtu.ac.in

Manan Saxena[1]
Software Engineering
Delhi Technological University
New Delhi, India
manan bt2k16@dtu.ac.in

Aruna Bhat[2]
Computer Science and Engineering
Delhi Technological University
New Delhi, India
aruna.bhat@dtu.ac.in

Abstract—Breast Cancer is one of the most common types of cancer amongst women, and early detection of carcinogenic tissue from histology images can go a long way in effective treatment as well. Recent approaches to solving this problem of classification of tissues utilize heavy deep learning-based architectures which takes higher inference time and storage space for model parameters. In this paper, we propose a CNN architecture with a lesser number of parameters that can be effectively deployed on a resource-constrained device, with the utilization of the Knowledge Distillation technique. We evaluate our approach on High-Resolution Breast Cancer histology slides of the BACH 2018 dataset.

Index Terms—Deep Learning, Residual Networks, Patchwise extraction, Pattern Recognition, Stain Normalization, Knowledge Distillation

I. INTRODUCTION

In 2019, according to the American Cancer Society [1] approximately 268,600 new cases of breast cancer were reported out of which 41, 760 cases resulted in the death of the patient in question. Moreover, according to a report published by the American Cancer Society, breast cancer is one of the most prevalent cancers in women. For the diagnosis of breast cancer in a patient, Histopathology plays a crucial role. It is the foremost technique for distinguishing between malignant and benign breast cancer tissue and thereby also differentiating patients suffering from in-situ or invasive carcinoma in the process.

Normally, the procedure for determining the nature of the biopsy tissue involves the analysis of the tissue under a microscope wherein the pathologist can look for specific features such as the spatial arrangement of the cells, morphometric characteristics of the nuclei, how many of the cancer cells are in the process of dividing(mitotic count), etc [2]. However, the manual examination of cancerous tissue is prone to errors, and two highly-trained pathologists may have different opinions on the same biopsy tissue. To overcome that, Computer-aided diagnosis(CAD) has proven to be quite useful for giving accurate information, which is quite close to what a trained pathologist would give.

To aid the Computer-aided diagnosis of biopsy slides, breast tissue slides stained with Hematoxylin and Eosin (H&E) is used. These slides typically determine a crucial rule in determining the nature of the breast tissue, namely if it is non-cancer(i.e., normal or benign) or a malignant(in situ or invasive carcinoma) tissue. When the breast tissue is stained with Hematoxylin, the nuclei of the tissue in question is stained purple, whereas Eosin turns the cytoplasm surrounding the nuclei to a pinkish color. The nature of this staining is used by pathologists and consequently, in CAD to identify the grade of carcinoma present in the tissue.

To develop a solution to this problem, we have chosen to move forward with The Breast Cancer Histology Challenge (BACH) 2018 dataset, which consists of high-resolution H&E stained breast histology microscopy

ISSN: 2005-4238 IJAST

images [3]. The images contained in this dataset are RBG images of the size 2048 × 1536 pixels. Each pixel covers $0.42\mu m \times 0.42\mu m$ of tissue area. The labeling of the images in this dataset is done by two medical experts, and in cases of disagreement among the experts, the image in question was not added to the final dataset. The images are supposed to be classified into four categories: i) normal tissue, ii) benign lesion, iii) in situ carcinoma, and iv) invasive carcinoma as per the diagnosis of the two medical experts. Each category has 100 images, thereby giving a dataset containing 400 images.

Many researchers have worked on the issue of early diagnosis of breast cancer using the tissue biopsy slides and have achieved significant success in this area [4]–[6]. However, since the images present in the BACH 2018 dataset are very high-resolution in comparison to other datasets, like MNIST [7], the model architectures proposed in these papers generally have very high computational resource requirements in terms of inference and training time and requires good storage space for model parameters. In this paper, we propose a lower parameter model, trained using the knowledge distillation [8] technique that performs reasonably well as compared to a high parameter model.

The motivation behind using a deep learning architecture with lower computational requirements instead of one with a higher number of parameters stems from the fact that deployment of such a model would be troublesome,

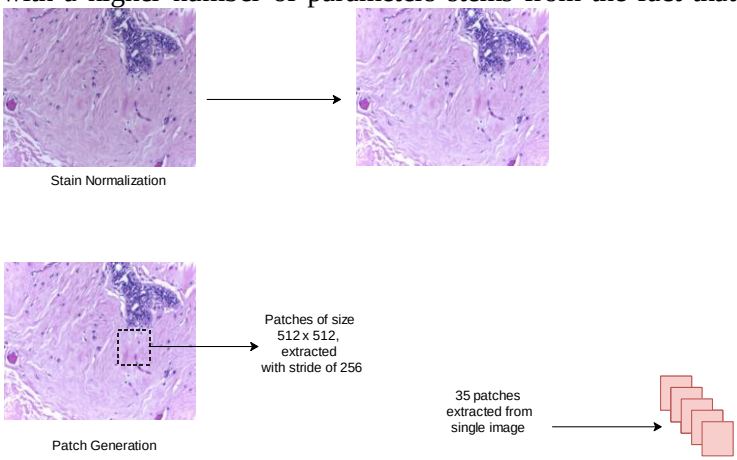

Fig. 1: Data Preprocessing

especially if the model is designed to cater to a large number of people. Running the model would be too computationally expensive in comparison to a "smaller" model as well. Moreover, the advantages of having a smaller model would also include the opportunity to deploy the model in question to resourceconstrained devices, which can not support larger models with higher resource requirements. The fact that a smaller model could be made available on a resource-constrained device makes it possible to detect the presence of carcinogenic breast tissue in a faster and cheaper manner, thereby making this method commercially viable.

To summarize, our main contributions are:

1) We utilized Knowledge Distillation [8], for training a low resource requirement model for Medical Histology Image Classification.
2) We demonstrate the effectiveness of Knowledge Distillation for the task consisting of very high-resolution images and a low number of classes, compared to standard datasets such as CIFAR-10, MNIST.

The rest of the paper is organized as follows. Background and Related Work are given in Section 2. Methodology in Section 3, Experiment and Results in Section 4 and Conclusion and Future Work in Section 5.

II. BACKGROUND AND RELATED WORK

Deep Learning has become the principal tool to achieve high-level performance for many vision and non-vision problems. But such high accuracy is usually acquired by exploiting a very deep neural network or their ensemble. Also, utilizing these deep neural networks require significant memory space and computational resources. Hence, over the years, many techniques for model compression have been devised, such as Knowledge Distillation [8].

ISSN: 2005-4238 IJAST

Hinton et al. introduced knowledge distillation with neural networks. [8] where they drew attention towards training a less powerful student network by ”distilling” the knowledge gained from training a more powerful teacher network. Distillation can be done by using ”soft targets” as class probabilities for the student model implying usage of the last layers of the teacher model as ”feature representation” to train the student model [9].

Detection and classification of breast cancer in microscopic tissue images at early stages can significantly help in

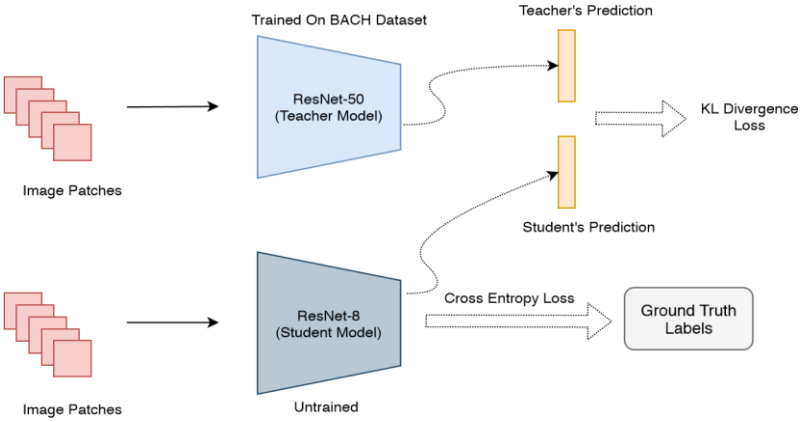

Fig. 2: Model Architecture

developing the course of action used to treat the patient; hence it has been a critical area of research for a significant amount of time. Recently, several works aimed at usage of deep learning models for breast cancer histology image classification have been published [4]–[6]. Differences in these works lie in the model and the patch extraction technique used. Patches are extracted either by convolutional based split [4], [5] or nuclei based split [6]. The deep learning models used vary from being a standard CNN architecture [4] to using huge models like InceptionV3 for transfer learning [5], [6]. The major drawback of using these models is the the need for very high computational resources required to train and later on deploy these networks.

III. METHODOLOGY

In this paper, we propose to solve this problem using a ResNet-50 network as a teacher and a ResNet-8 network as a student. Since for this paper, we are making the use of highresolution breast cancer histology images, we preprocess the images using stain normalization and a patchwise image split strategy, as mentioned in [4]. Stain normalization is a process wherein the variations present in the histology slides due to factors such as Staining duration, Stain concentration, etc. are overcome by a standardisation process described by Reinhard et al. [10]. This is necessary because the variations present in the histology slide don’t cope well with the deep learning networks used in this project, thereby giving inaccurate results.

After the images are preprocessed, they are used in the training phase of the ResNet-50 network wherein the classes assigned to the slides are used as the ground truth label. Post training, the accuracy of the network is measured with respect to the assumed ground truth. Once ResNet-50 has been trained, we use knowledge distillation to train the ResNet-8 model.

A. Stain Normalization

A well-known problem in histological image analysis is of variations present in the stained images, which are mainly caused by unquantified staining procedures for preparing tissues before microscopic imaging for cancer diagnosis. The variety is primarily in color and intensity. Therefore, stain normalization is used to make the images usable for computerbased analysis. In this paper, we have utilized the approach proposed by Reinhard et al. [10], which maps the color histogram of the image to that of the target image, following the transformation of RGB colorspace to the de-correlated LAB colorspace. It’s done by using the linear transforms to match the standard deviation and mean of each color channel of the source and target images in LAB colorspace.

ISSN: 2005-4238 IJAST

B. *Patchwise Image Split*

Training deep learning models on high-resolution images of the dataset used requires ample memory space or downsampling of the images. However, neither a large memory is readily available, nor downsampling can be done as it leads to loss of discriminative features in the image. Also, if we train on the whole high-resolution images, then the models might learn the most distinctive features and disregard the intricate details of the image. Hence, feeding images in the form of patches to the network is the simplest solution. In this paper, we have used the patchwise image generation mentioned by [4]. For an image, we extract fixed-sized patches by sliding a window(patch) of size $k \times k$ over the image with stride s. With this, we would get a total number of $[1+\frac{I_H-k}{s}] \times [1+\frac{I_W-k}{s}]$ patches where I_H and I_W are image

height and width respectively. In our experiments, we take a patch size of 512 × 512 with a stride of 256, which gives us 35 patches of each image. The patches generated here are overlapping, which is beneficial as it helps in learning of features shared between patches.

C. *Training Residual Networks*

In this paper, we have used the ResNet [11] as our base deep learning model. ResNet is based on residual learning framework, which enables the network to retain the learning of previous residual blocks by doing an identity mapping weight function where the output is equal to the input, preserving by adding on to what the network already learned. Such a network is easier to optimize and, therefore, enable the learning of deeper networks. For our teacher model, we have used ResNet50, which has 23.5 million parameters and 151 layers, whereas for student model, we have used ResNet-8 which had only 94.7k parameters and merely 37 layers. The low number of parameters makes the model less computationally expensive. To train the model, we use the cross-entropy loss function which is described as:

$$H_y 0(y) = -\sum_i y_i^0 \log(y_i) \quad (1)$$

D. *Knowledge Distillation*

Knowledge Distillation [9] is a technique that aims to "distill" or transfer the knowledge contained in a cumbersome model to a smaller model that is more suitable to be deployed in resource-constrained devices. As compared to traditional Knowledge Transfer techniques, Knowledge Distillation aims to make the smaller models generalize better to the training data and reflect the real objective of the user rather than to optimize the performance on the training data. Training a model using this approach would, however, requires information about the generalizations, patterns, and the relations inherent in the data to make the generalizations possible. In deep learning applications, the class probabilities for a particular case depict a lot of hidden information and not just the label to be chosen. Analyzing the class probabilities sometimes reveals hidden patterns amongst the data that may prove to be very useful. However, these relationships have very little influence on the cross-entropy cost function used during the learning phase due to the small probabilities associated with them. Knowledge Distillation addresses this problem by raising the temperature of the softmax until the cumbersome model produces a sufficiently soft set of targets. This temperature is later used in the training phase of the small model. The loss function used is the KL divergence function given by the following equation:

$$\begin{aligned} D_{\mathrm{KL}}(P\|Q) &= \int_{x_a}^{x_b} P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx \\ &= \int_{y_a}^{y_b} P(y) \log\left(\frac{P(y)\frac{dy}{dx}}{Q(y)\frac{dy}{dx}}\right) dy \\ &= \int_{y_a}^{y_b} P(y) \log\left(\frac{P(y)}{Q(y)}\right) dy \end{aligned} \quad (2)$$

ISSN: 2005-4238 IJAST

IV. EXPERIMENTS AND RESULTS

We used the BACH 2018 dataset [1]which consists of 400 high-resolution Hematoxylin and Eosin stained breast cancer histology images having labels as normal, benign, invasive and in-situ with 100 images of each category. We took 70 samples, 15 samples, and 15 samples from each class for training, validation, and testing, as to avoid class imbalance. Since the images are of high-resolution, it is impractical to train a CNN using whole images. Thus, the patchwise generation is beneficial in this case. From each image, 35 patches are generated covering the whole surface of the input image. Each patch was assigned a class label of the image which it is extracted from. We train the networks on a single NVIDIA Tesla P-100 GPU with 12GB memory using Adam optimizer [12], with a mini-batch size of 16 and initital learning rate of 0.001, with a decay of 0.1 every 20 epochs. The results reported in I, II, III are reported on patchwise classification of data. From I and II, we can see that the ResNet-50 (Teacher Model) has the best precision on majority of classes. This is due to the fact that a higher parameter model has a better generalization performance [13]. The precision on ResNet-8 (without distillation training) showed poor results, but after using distillation, the model showed good generalization over the data. Similarly, in III, we report the F1 score of all the classes and Accuracy of the prediction to demonstrate the effectiveness of using the Knowledge Distillation technique for Medical Histology Image Classification.

V. CONCLUSION AND FUTURE WORK

In this paper, we have demonstrated the fact that an architecture with lower parameters can perform well with the help of Knowledge Distillation. This has a wide array of applications in the industrial sector, which we intend to research upon. For further research, we can use several novel methods proposed in Medical Image Classification by replacing the base models with a low parameter models. Also, we can utilize semisupervised methods on these to improve the generalizability of our models.

TABLE I: Precision

Model *Architecture*	Precision			
	Benign	*In-situ*	*Invasive*	*Normal*
ResNet-50	0.85	0.83	0.92	0.72
Resnet-8	0.68	0.79	0.82	0.74
ResNet-8 with KD	0.81	0.81	0.80	0.79

TABLE II: Recall

Model *Architecture*	Recall			
	Benign	*In-situ*	*Invasive*	*Normal*
ResNet-50	0.62	0.84	0.91	0.90
Resnet-8	0.74	0.77	0.76	0.75
ResNet-8 with KD	0.76	0.78	0.82	0.85

TABLE III: F1 Score and Accuracy

[1] https://iciar2018-challenge.grand-challenge.org/dataset/

ISSN: 2005-4238 IJAST

Model *Architecture*	F1 Score				Accuracy
	Benign	*In-situ*	*Invasive*	*Normal*	
ResNet-50	0.72	0.84	0.92	0.80	82.00 %
ResNet-8	0.71	0.78	0.79	0.75	75.71 %
ResNet-8 with KD	0.79	0.79	0.81	0.82	80.24 %

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal. (2019) Cancer statistics, 2019. [Online]. Available: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21551

[2] D. I. A. Group, "Breast histology," 2019, online; accessed 15 March 2020. [Online]. Available: http://diagnijmegen.nl/index.php/Breast Histology

[3] G. Aresta, T. Araujo,´ S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, G. Fernandez, J. Zeineh, M. Kohl, C. Walz, F. Ludwig, S. Braunewell, M. Baust, Q. D. Vu, M. N. N. To, E. Kim, J. T. Kwak, S. Galal, V. Sanchez-Freire, N. Brancati, M. Frucci, D. Riccio, Y. Wang, L. Sun, K. Ma, J. Fang, I. Kone, L. Boulmane, A. Campilho, C. Eloy, A. Polonia,´ and P. Aguiar, "Bach: Grand challenge on breast cancer histology images," *Medical Image Analysis*, vol. 56, pp. 122 – 139, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841518307941

[4] K. Nazeri, A. Aminpour, and M. Ebrahimi, "Two-Stage Convolutional Neural Network for Breast Cancer Histology Image Classification," *arXiv e-prints*, p. arXiv:1803.04054, Mar. 2018.

[5] S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, and A. Maier, "Classification of breast cancer histology images using transfer learning," *arXiv e-prints*, p. arXiv:1802.09424, Feb. 2018.

[6] A. Golatkar, D. Anand, and A. Sethi, "Classification of Breast Cancer Histology using Deep Learning," *arXiv e-prints*, p. arXiv:1802.08080, Feb. 2018.

[7] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[8] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," *arXiv e-prints*, p. arXiv:1503.02531, Mar. 2015.

[9] J. Karlekar, J. Feng, Z. Sian Wong, and S. Pranata, "Deep Face Recognition Model Compression via Knowledge Transfer and Distillation," *arXiv e-prints*, p. arXiv:1906.00619, Jun. 2019.

[10] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21 , no. 5, pp. 34–41, July 2001.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv e-prints*, p. arXiv:1512.03385, Dec. 2015.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.

[13] A. Brutzkus and A. Globerson, "Why do Larger Models Generalize Better? A Theoretical Perspective via the XOR Problem," *arXiv e-prints*, p. arXiv:1810.03037, Oct. 2018.

ISSN: 2005-4238 IJAST