A System for One-Shot Learning of Cervical Cancer Cell Classification in Histopathology Images

Dig Vijay Kumar Yarlagadda[a], Praveen Rao[a], Deepthi Rao[b], and Ossama Tawfik[c]

[a]University of Missouri-Kansas City
[b]University of Missouri-Columbia
[c]Saint Luke's Hospital of Kansas City

ABSTRACT

Convolutional neural networks (CNNs) have been popularly used to solve the problem of cell/nuclei classification and segmentation in histopathology images. Despite their pervasiveness, CNNs are fine-tuned on specific, large and labeled datasets as these datasets are hard to collect and annotate. However, this is not a scalable approach. In this work, we aim to gain deeper insights into the nature of the problem. We used a cervical cancer dataset with cells labeled into four classes by an expert pathologist. By employing pre-training on this dataset, we propose a one-shot learning model for cervical cell classification in histopathology tissue images. We extract regional maximum activation of convolutions (R-MAC) global descriptors and train a one-shot learning memory module with the goal of using it for various cancer types and eliminate the need for expensive, difficult to collect, large, labeled whole slide image (WSI) datasets. Our model achieved 94.6% accuracy in detecting the four cell classes on the test dataset. Further, we present our analysis of the dataset and features to better understand and visualize the problem in general.

Keywords: One-shot learning, cervical cancer, whole slide images

1. INTRODUCTION

Cervical cancer causes upwards of 500,000 new cases and more than 250,000 patients die each year worldwide.[1] In the United States alone, it is expected that more than 13,170 cervical cancer cases will be diagnosed in 2019.[2] Pap smear is a screening test for cervical cancer and can lead to early diagnosis and treatment. Cancer diagnosis from tissue biomarker scoring is a vital technique in determining the type and grade of cancer. The process demands a lot of effort and time from pathologists. It is also prone to errors and lacks inter-pathologist agreement. Whole slide imaging is touted as a disruptive technology in digital pathology for enabling the automatic detection and analysis of cellular and morphological features. Despite the promise of CNNs for deep learning on histopathology images,[3,4] they are fine-tuned on specific, large and labeled datasets. Also, creating large fully supervised datasets like ImageNet[5] is not feasible for WSIs; therefore, semi-supervised or unsupervised learning approaches need to be explored for this problem. Traditional deep neural networks for classification can be combined them other learning representations such as transfer learning, reinforcement learning, and one-shot learning to improve their usability for the task of WSI analysis. In one-shot learning,[6] we observe a single example and use the learned knowledge to predict on an unseen task. In classification tasks where dataset is small and contains very few examples in each class, one-shot learning is a viable choice.

Motivated by these reasons, we present a one-shot learning model for automatically classifying cervical cells in histopathology images (pap smears) into four classes: ASCUS (atypical squamous cells of undetermined significance), HGSIL (high grade squamous intraepithelial lesion), LGSIL (low grade squamous intraepithelial lesion), and normal. Our model is based on Inception-ResNet-v2.[7] Using a dataset curated and labeled by an expert pathologist, we demonstrate that our model achieves good accuracy for classifying cervical cancer cells in histopathology images. We performed experiments with standard deep learning networks such as Inception-v3[8] and ResNet[9] and visualized the hidden layers of our network to assess how well it understood the features in

Author email information: dy5kc@umkc.edu, raopr@umkc.edu, raods@health.missouri.edu, otawfik@saint-lukes.org

histopathology images. Our system can be trained end-to-end and can be extended to other cancer types in a lifelong manner.

The rest of the paper is organized as follows: In Section 2, we provide some background and discuss related work. In Section 3, we present our approach of one-shot learning for classifying cervical cells. In Section 4, we describe our dataset, experimental setup, and report the cervical cell classification results. Finally, we conclude and present some directions for future work in Section 5.

2. BACKGROUND AND RELATED WORK

In recent years, a lot of advances have been made in the area of deep learning due to the availability of modern hardware such as graphics processing units (GPUs) and large datasets. This has resulted in many state of art results for many tasks and development of scalable frameworks for deep learning like TensorFlow.[10] CNNs have been widely used for object recognition on large-scale image datasets. A typical CNN architecture has convolutional, rectified linear unit (ReLU), and pooling layers, followed by one or more fully connected layers. Networks with more than hundred layers such as GoogLeNet[11] and ResNet[9] usually outperform others on several classification tasks. Our work in this paper is inspired by the success of these very deep networks. GoogLeNet is based on the idea of inception module, it applies multiple convolutions (e.g., $1 \times 1, 3 \times 3, 5 \times 5$) at different scales and the results are then concatenated to form multi-level features. ResNet, on the other hand, introduced skip connections to allow for training deeper networks without degradation in performance.

Previous approaches for cervical cell classification[12,13] use pixel-based or segmentation techniques[14] to extract nuclei features and use shallow CNNs to classify the features. Most of these techniques use preprocessing techniques like blue ratio transformation, non-local means filters, or patch extraction techniques[15] before feeding data into the network. Because some of these classifiers utilized specific handcrafted features chosen specially for the dataset at hand, they suffer from the skewness of dataset. To the best of our knowledge, one-shot learning approaches have never been used for the task of cell classification in histopathology images.

3. PROPOSED APPROACH

Our system architecture contains a region proposal network followed by a pre-trained CNN feature extractor for cell detection and classification and a memory module for one-shot learning containing R-MAC global descriptors[16] representing CNN layer features. The system architecture is shown in Figure 1. Our classification network is based on Inception-ResNet-v2.[7] Inception-ResNet combines residual connections and cheaper inception blocks to improve the accuracy while retaining computational efficiency. We have observed that 1×1 convolutions and residual connections played a critical role for improving the accuracy for our task along with the improved training speed due to residual connections. In our experiments, Inception-ResNet-v2 outperforms various forms of ResNet and DenseNet[17] architectures. We used the region proposal network as originally introduced in Faster R-CNN architecture.[18]

Our one-shot learning memory component is based on the idea of using a memory module in deep neural networks for classifying learned features.[19] Each slot in the module contains key-value pairs with key being the memory keys of dimension 128 and value being the target label. We obtain R-MAC global descriptors from CNN layers and store them along with the label in the memory module. Given a memory query and target value, we compute nearest neighbors in memory sorted in decreasing order by cosine similarity and train the module using memory loss:[19] we maximize the similarity to positive key and minimize similarity to negative key till they are apart by a certain margin. As shown in Figure 1, our system can be trained end-to-end in a lifelong manner. In contrast to other one shot learning approaches like Siamese neural networks[20] and memory augmented neural networks,[21] we believe that our model provides us the opportunity to train on different cancer types and add to the capabilities of our model in lifelong manner.

4. EVALUATION

4.1 Dataset

Our goal was to classify the cervical cells in histopathology images into four classes: ASCUS, HGSIL, LGSIL and normal. An example of each class is shown in Figure 2.

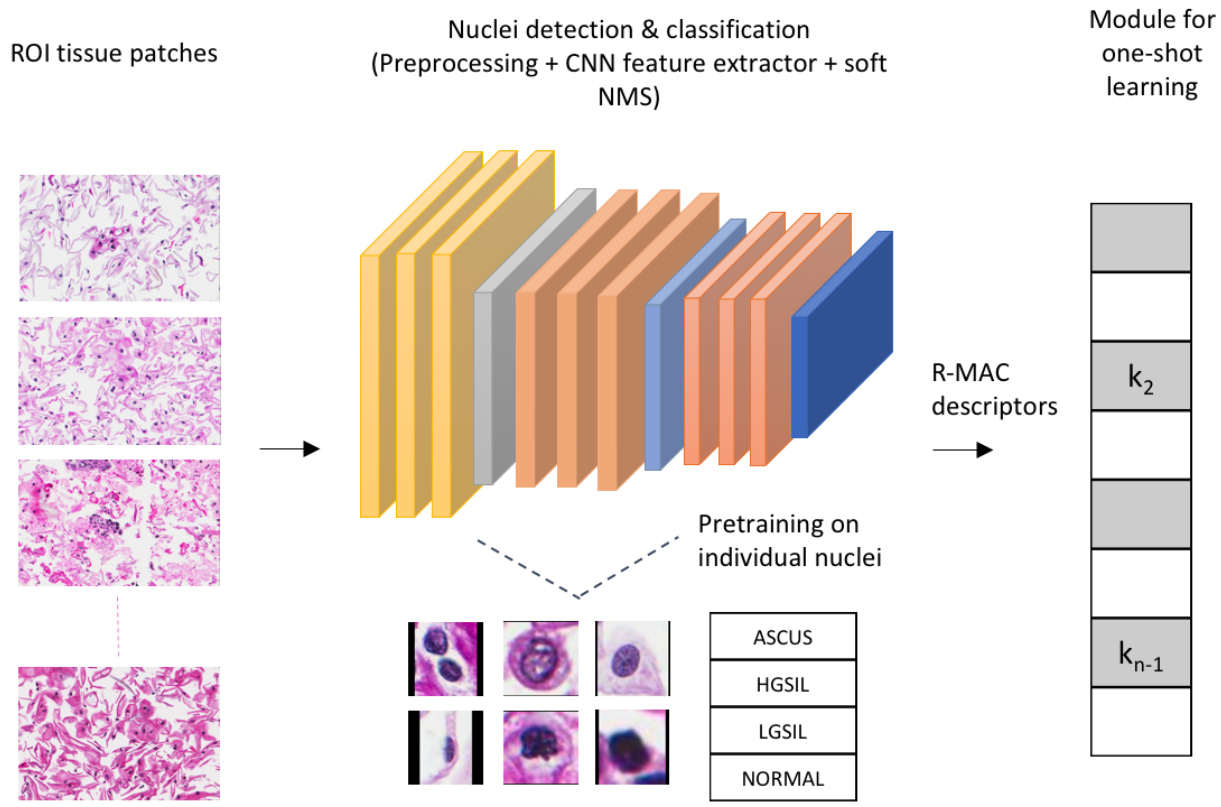

Figure 1. System architecture

ASCUS	HGSIL	LGSIL	normal
46	34	51	100

Table 1. Dataset class distribution

The dataset curated for this work consists of 90 cervical cancer WSI tissue samples, 15 of which were labeled by an expert pathologist into the aforementioned four classes as shown in Figure 3. The WSIs represents 90 patients and are digitized at 20X resolution into SVS format by Aperio scanners at the University of Kansas Medical Center. The labeled data was extracted by using OpenCV,[22] NumPy and Python Imaging Library. The extracted dataset for classification contained 231 patches. The dataset was split into 80%, 10%, and 10% for training, validation and test sets, respectively. The distribution of each cell class is shown in Table 1. The main challenges with our dataset were limited number of training samples and data imbalance. This dataset captured complex morphology, texture, scale, variability due to staining process and scanner and high intra-class variations

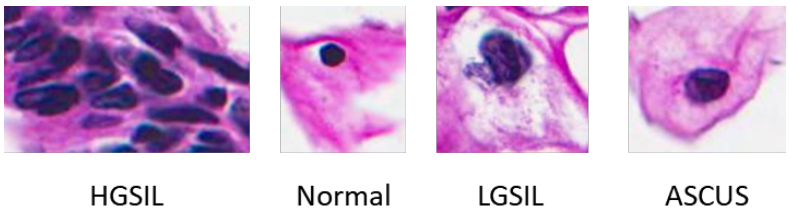

Figure 2. Cervical cancer classes labeled by an expert pathologist

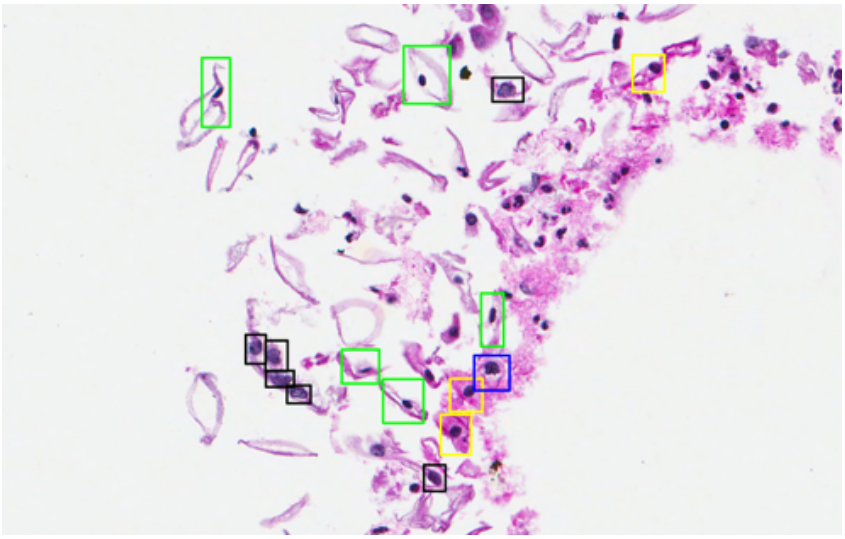

Figure 3. Tissue sample labeled by an expert pathologist. Note that the green boxes correspond to "normal", yellow boxes correspond to ASCUS, black boxes correspond to HGSIL, and blue boxes correspond to LGSIL.

of a typical histopathology dataset. (For example, Aperio and Hamamatsu slide scanners have variations in their procedure for digitizing glass slides). A sample region of interest is shown in Figure 4. We observed that curating a dataset in this manner takes a pathologist less time and effort when compared to exhaustively annotating WSIs.

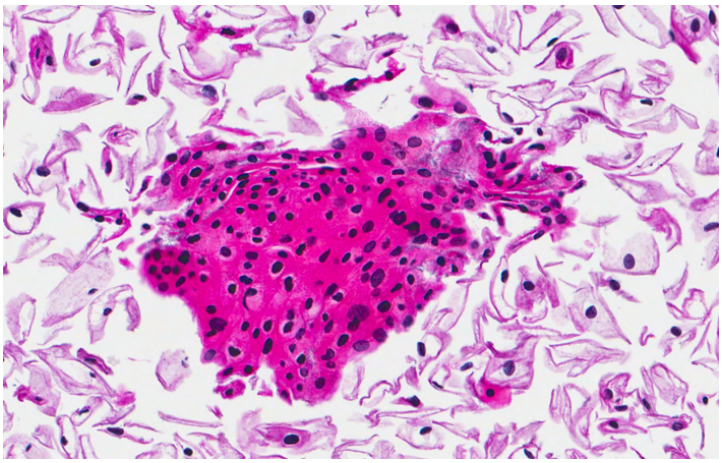

Figure 4. Sample WSI region of interest

4.2 Experiments

The cell detection and classification network was trained on our dataset for 940 epochs with the following hyperparameters: fixed learning rate of 1e-2, batch size of 32, RMSProp optimizer with a weight decay of 4e-5, and input image zero-padded and resized to 224 x 224 pixels. We implemented our system in Python using TensorFlow, CUDA, NumPy and OpenCV libraries. Recall that the training methodology is provided in Figure 1.

Our one-shot learning module was trained for 100,000 training episodes, each episode with 4 distinct labels, a batch size of 4 and episode length of 31. Our memory module had 40,960 slots in memory with dimension of each key was set to 128. We kept aside 10 episodes for validating our model. It took 12 hours to train the system from end to end.

4.3 Results

We evaluated our model on a test dataset of 23 images. The experiments were run on a test bench running Intel i7 6700 HQ CPU, 16 GB RAM and NVIDIA GTX 960m GPU with 640 CUDA cores, memory clock of 2500 MHz, and 4 GB GDDR5 memory. Our model achieved an accuracy of 94.6% in classifying cells into one of four classes: ASCUS, HGSIL, LGSIL and normal. For comparison, we performed experiments with Inception-v3[8] and ResNet,[9] they have achieved 88% and 91% accuracy, respectively on the same dataset. When compared to Omniglot[23] - a standard dataset for evaluating one-shot learning models, our dataset is more challenging, our results demonstrate that augmenting models using a memory module could be a good approach for small histopathology datasets.

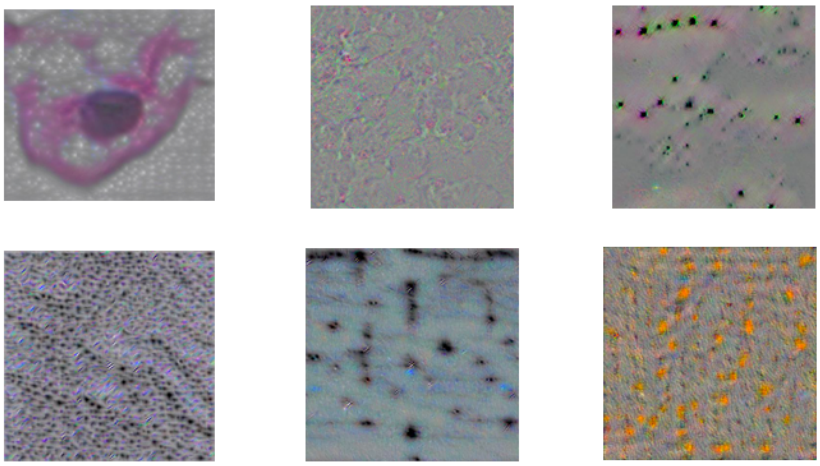

Figure 5. Features learned by our network

Visualizing the hidden layers of our network reveals that our network did understand important features in histopathology images. As shown in Figure 5, the purpose of these features is to illustrate which part of the image is considered important for the classifier and to prove the expressiveness of chosen CNN features.

5. CONCLUSION AND FUTURE WORK

In this work, we proposed an end-to-end architecture to facilitate detection and classification of cervical cancer cell classes on histopathology image datasets. Using CNNs, R-MAC global descriptors and one-shot learning module, we achieved 94.6% classification accuracy on our dataset. We also visualize the learned features to better understand the classification task in general.The code and data are available at https://goo.gl/mNwxMf.

In the future, we plan to expand our dataset and evaluate our model on several other cancer types. We also plan to use attention mechanisms to improve detection and classification, scale number of classes, optimize size of memory module and the model. In addition, we plan to integrate this work with our storage platform[24] to create an end-to-end pipeline for Whole Slide Images.

Acknowledgments

We would like to thank the University of Missouri Research Board for partial support.

REFERENCES

[1] "TCGA study identifies genomic features of cervical cancer," (Jan 2017).

[2] "Key statistics for cervical cancer." American Cancer Society (2019). https://www.cancer.org/cancer/cervical-cancer/about/key-statistics.html.

[3] Liu, Y., Gadepalli, K. K., Norouzi, M., Dahl, G., Kohlberger, T., Venugopalan, S., Boyko, A. S., Timofeev, A., Nelson, P. Q., Corrado, G., Hipp, J., Peng, L., and Stumpe, M., "Detecting cancer metastases on gigapixel pathology images," tech. rep., arXiv (2017). https://arxiv.org/abs/1703.02442.

[4] Arajo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polnia, A., and Campilho, A., "Classification of breast cancer histology images using convolutional neural networks," *PLOS ONE* **12**, 1–14 (06 2017).

[5] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., "ImageNet: A Large-Scale Hierarchical Image Database," in [*2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], (2009).

[6] Fei-Fei, L., Fergus, R., and Perona, P., "One-shot learning of object categories," *IEEE transactions on pattern analysis and machine intelligence* **28**(4), 594–611 (2006).

[7] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A., "Inception-v4, inception-resnet and the impact of residual connections on learning.," in [*Proceedings of 2017 AAAI Conference*], (2017).

[8] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., "Rethinking the inception architecture for computer vision," *CoRR* **abs/1512.00567** (2015).

[9] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[10] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., "Tensorflow: a system for large-scale machine learning.,"

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going deeper with convolutions," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1–9 (2015).

[12] Zhang, L., Lu, L., Nogues, I., Summers, R. M., Liu, S., and Yao, J., "Deeppap: Deep convolutional networks for cervical cell classification," *arXiv preprint arXiv:1801.08616* (2018).

[13] Forslid, G., Wieslander, H., Bengtsson, E., Whlby, C., Hirsch, J., Stark, C. R., and Sadanandan, S. K., "Deep convolutional neural networks for detecting cellular changes due to malignancy," in [*2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*], 82–89 (Oct 2017).

[14] Zhao, L., Li, K., Wang, M., Yin, J., Zhu, E., Wu, C., Wang, S., and Zhu, C., "Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search mrf," *Computers in biology and medicine* **71**, 46–56 (2016).

[15] Xu, Y., Jia, Z., Wang, L.-B., Ai, Y., Zhang, F., Lai, M., Eric, I., and Chang, C., "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC Bioinformatics* **18**(1), 281 (2017).

[16] Tolias, G., Sicre, R., and Jégou, H., "Particular object retrieval with integral max-pooling of cnn activations," *arXiv preprint arXiv:1511.05879* (2015).

[17] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q., "Densely connected convolutional networks," in [*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 2261–2269, IEEE (2017).

[18] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: Towards real-time object detection with region proposal networks," in [*Advances in neural information processing systems*], 91–99 (2015).

[19] Kaiser, Ł., Nachum, O., Roy, A., and Bengio, S., "Learning to remember rare events," *arXiv preprint arXiv:1703.03129* (2017).

[20] Koch, G., Zemel, R., and Salakhutdinov, R., "Siamese neural networks for one-shot image recognition," in [*2015 ICML Workshop on Deep Learning*], (2015).

[21] Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T. P., "One-shot learning with memory-augmented neural networks," *CoRR* **abs/1605.06065** (2016).

[22] Bradski, G., "The OpenCV Library," *Dr. Dobb's Journal of Software Tools* (2000).

[23] Lake, B., Salakhutdinov, R., Gross, J., and Tenenbaum, J., "One shot learning of simple visual concepts," in [*Proceedings of the Annual Meeting of the Cognitive Science Society*], **33**(33) (2011).

[24] Barron, D. E. L., Yarlagadda, D. V. K., Rao, P., Tawfik, O., and Rao, D., "Scalable storage of whole slide images and fast retrieval of tiles using apache spark," in [*Medical Imaging 2018: Digital Pathology*], **10581**, 1058113, International Society for Optics and Photonics (2018).