

CLASSIFICATION OF HISTOPATHOLOGY IMAGES WITH RANDOM DEPTHWISE CONVOLUTIONAL NEURAL NETWORKS

Yanan Yang, Fadi G. Farhat, Yunzhe Xue, Frank Y. Shih, and Usman Roshan

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102

ABSTRACT

The classification of whole slide images plays an important role in understanding and diagnosing cancer. Pathologists typically have to work through numerous pathology images that can be in the order of hundreds or thousands which takes time and is prone to manual error. Here we investigate an automated method based on a random depthwise convolutional neural network (RDCNN). In previous work this network has shown to achieve high accuracies for image similarity. We conjecture that for pathology images similarity may play an important role in accurate classification of the images. We evaluate RDCNN against trained deep convolutional neural networks VGG16 and ResNet50 on four pathology image datasets. We find RDCNN to give the average highest accuracy across the four datasets. On two datasets RDCNN is significantly higher in accuracy and comparable in the others. This suggests that for whole slide image data a network with random weights can better capture similarity and thus classification.

Index Terms— histopathology, whole slide images, random depthwise convolutional network

1. INTRODUCTION

The classification of histopathology images play a key role in diagnosing and understanding cancer. Pathologists typically have to browse numerous images to determine the tumor type which requires considerable training, is time intensive, and is prone to manual errors. The automated classification of tumor type can greatly speed up physician diagnosis and lead to better care and treatment. Convolutional neural networks that attain the state of the art in image recognition [1] have previously been proposed for this problem [2, 3, 4].

Here we investigate a depthwise convolutional neural network with random weights (RDCNN) [5]. Previously this has been shown to classify images with similar background, color, and texture accurately as evaluated on existing benchmarks [6]. We conjecture this may be useful in the problem of histopathology images where image similarity may play a role in classification. To test this hypothesis we performed an experimental performance study comparing the accuracy

of trained convolutional networks to RDCNN. Below we describe our datasets and methods followed by our results.

2. METHODS

2.1. Datasets

We obtain four publicly available datasets spanning three different cancers.

2.1.1. IDC - breast cancer

The Invasive Ductal Carcinoma (IDC) dataset is provided by ICPR 2012 contest [7]. The original dataset consisted of 162 whole mount slide images of Breast Cancer histology specimens scanned at $40x$. From that patches of size 50×50 were extracted of which 198,738 were IDC negative and 78,786 IDC positive. We used the same train and test split by Janowczyk et. al. [8].

2.1.2. ISIC - skin cancer

This dataset is provided by the ISIC 2019 Challenge [9, 10]. This is for classifying skin cancer images among nine different diagnostic categories: Actinic Keratosis, Squamous Cell Carcinoma, Basal Cell Carcinoma, Seborrhic Keratosis, Solar Lentigo, Dermatofibroma, Nevi, Melanoma, and Vascular Lesions. This dataset includes a total of 25331 images each of size 600×400 . We split them into train and test with a ratio of 80:20.

2.1.3. Gleason - prostate cancer

Gleason 2019 dataset contains prostate cancer from H&E-stained histopathology images, which is provided by Gleason 2019 challenge (<https://bmi.ubc.ca/research/miccai-automatic-prostate-gleason-grading-challenge-2019>). This challenge is part of MICCAI 2019 Conference, and will be one of the three challenges under the MICCAI 2019 Grand Challenge for Pathology. Data used in this challenge consists of 267 tissue micro-array (TMA) images, the size of each image is 5120×5120 . Each TMA image is annotated in detail by several expert pathologists. We select Map1 (the

first expert pathologist labels) as true labels, and split these images into train and test with a ratio of 80:20.

2.1.4. BreakHis - breast cancer

The microscopic biopsy images in the BreakHis dataset were collected from 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). The images are provided in their raw PNG (Portable Network Graphic) format, without normalization or color standardization and are all the same size (700×460 pixels, 3-channel RGB, 8-bit depth per channel). The samples were collected by Surgical Open Biopsy method, also called partial mastectomy or excisional biopsy [11]. This type of procedure removes a large tissue sample and is done in a hospital with general anesthesia.

The benign and malignant image groups are further divided into sub-groups describing the specific kind of anomaly. For benign lesions, the anomalies present are fibroadenoma, Phyllodes tumor and tubular adenoma. For the malignant lesions, the anomalies present are ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma.

We only consider images at the 400X magnification level, where we count a total of 1,606 samples (1285 for training, 321 for testing). Out of that total, 374 samples are benign and 1,232 are malignant. By using augmentation (adding more samples by rotating and flipping the original images), we will have 8120 samples (6496 for training, 1624 for testing).

2.2. Convolutional neural networks

Convolutional neural networks are typically composed of alternating convolution and pooling layers followed by a final flattened layer. A convolution layer is specified by a kernel size and the number of kernels in the layer. Briefly, the convolution layer performs a moving non-linearized dot product against pixels given by a fixed kernel size $k \times k$ (usually 3×3 or 5×5). The dot product is usually non-linearized with the sigmoid or hinge (relu) function since both are differentiable and fit into the gradient descent framework. The output of applying a $k \times k$ convolution against a $p \times p$ image is an image of size $(p - k + 1) \times (p - k + 1)$.

2.3. Random depthwise convolutional neural networks (RDCNN)

Consider applying random convolutional blocks repeatedly and then averaging all the values in the final representation of the image. If we repeat this k times it gives us k new features. This can be described as a random depthwise convolutional neural network (RDCNN) [5]. Each convolutional block in our network is a convolutional kernel followed by 2×2 average pooling with stride 2.

Our network is parameterized by the number of convolutional blocks b , the size of each kernel $k \times k$ and the number of kernels m in each layer (this is the same in each layer). In

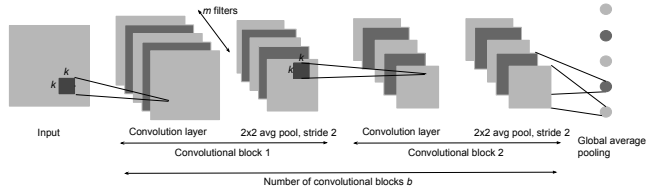


Fig. 1. A random depthwise convolutional neural network with two convolutional blocks, kernel size of k , and $m = 5$ kernels in each layer

Figure 1 we show an example of our network with two layers ($l = 2$) and five 3×3 convolution blocks in each layer ($m = 5, k = 3$). We set the values in each convolutional kernel randomly from the Normal distribution with mean 0 and variance 1.

We non-linearize the output of each convolution with the sign function and our convolution is *depthwise*. This means the i^{th} convolution is applied on the i^{th} kernel only of the previous layer. In the input layer, however, the convolution is applied in the conventional way to account for RGB images that have three layers. After we are done with convolutions we globally average pool the final layer which gives us a flattened feature space. We then apply a linear support vector machine or stochastic gradient descent on the final feature space.

2.4. Deep networks compared in our study

We compare our method to modern networks used in image recognition today. These are convolutional neural networks designed to enable deeper architectures and are trained with stochastic gradient descent.

- ResNet50 [12]: Residual convolutional networks contain connections from previous layers and not just the last one.
- VGG16 [13]: Deep convolutional neural network with layers of convolution and pooling.

3. RESULTS

In Table 1 we see the train and test accuracies of VGG16, ResNet50, and RDCNN on our four datasets. On ISIC and Gleason we see that RDCNN achieves a much higher accuracy and comparable on the remaining datasets. Overall across all the datasets RDCNN has the highest accuracy of 95% whereas VGG16 and ResNet50 have 79% and 89.8% respectively.

4. DISCUSSION AND CONCLUSION

Our preliminary results suggest that the unsupervised RDCNN can be highly useful in the classification of histopathol-

IDC			
Method	Train	Test	
VGG16	92.2	83.3	
ResNet50	100	88.2	
RDCNN (30K features, k=3, 4 layers)	87.8	87.6	
RDCNN (50K features, k=5, 4 layers)	86.3	87.6	
ISIC			
Method	Train	Test	
VGG16	89.8	85.9	
ResNet50	90.3	87.5	
RDCNN (65K features, k=3, 4 layers)	100	100	
RDCNN (65K features, k=5, 4 layers)	100	100	
Gleason			
Method	Train	Test	
VGG16	83.3	73.4	
ResNet50	87.5	75	
RDCNN (68K features, k=3, 4 layers)	100	93.5	
RDCNN (70K features, k=5, 2 layers)	100	93.5	
BreakHis 2 class			
Method	Train	Test	
VGG16	82.8	81.8	
ResNet50	100	99.8	
RDCNN (10K features, k=3, 7 layers)	100	98.8	
BreakHis 7 class			
Method	Train	Test	
VGG16	99.14	70.61	
ResNet50	94.1	98.6	
RDCNN (10K features, k=3, 7 layers)	99.5	95.4	

Table 1. Train and test accuracies (shown as percentages) of fully trained VGG16 and ResNet50 and the unsupervised RDCNN on our datasets.

ogy images. We see that a kernel size of 3 and a total of 4 layers works well in most cases. We plan to follow up with a detailed look at images where RDCNN performs better than the trained models.

5. REFERENCES

- [1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [2] Muhammed Talo, “Convolutional neural networks for multi-class histopathology image classification,” *arXiv preprint arXiv:1903.10035*, 2019.
- [3] Babak Ehteshami Bejnordi, Maeve Mullooly, Ruth M Pfeiffer, Shaoqi Fan, Pamela M Vacek, Donald L Weaver, Sally Herschorn, Louise A Brinton, Bram van Ginneken, Nico Karssemeijer, et al., “Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies,” *Modern Pathology*, vol. 31, no. 10, pp. 1502, 2018.
- [4] Arkadiusz Gertych, Zaneta Swiderska-Chadaj, Zhaoxuan Ma, Nathan Ing, Tomasz Markiewicz, Szczepan Cierniak, Hootan Salemi, Samuel Guzman, Ann E Walts, and Beatrice S Knudsen, “Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides,” *Scientific reports*, vol. 9, no. 1, pp. 1483, 2019.
- [5] Yunzhe Xue and Usman Roshan, “Image classification and retrieval with random depthwise signed convolutional neural networks,” in *International Work-Conference on Artificial Neural Networks*. Springer, 2019, pp. 492–506.
- [6] “Corel-princeton image similarity benchmark,” <http://www.cs.princeton.edu/cass/benchmark/>.
- [7] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber, “Mitosis detection in breast cancer histology images with deep neural networks,” in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2013, pp. 411–418.
- [8] Andrew Janowczyk, Ajay Basavanahally, and Anant Madabhushi, “Stain normalization using sparse autoencoders (stanosa): Application to digital pathology,” *Computerized Medical Imaging and Graphics*, vol. 57, pp. 50–61, 2017.
- [9] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al., “Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic),” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 168–172.
- [10] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler, “The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific data*, vol. 5, pp. 180161, 2018.
- [11] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte, “A dataset for breast cancer histopathological image classification,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2015.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in

Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

- [13] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.