Classification of histopathology images with random depthwise convolutional neural networks

Anonymous[1]

Anonymous

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 histopathology 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. We examine top similar images to a randomly selected one in the ISIC and Gleason datasets and see that it indeed most of the similar images belong to the same category as the query in the RDCNN feature space compared to ResNet50 and VGG16. For such histopathology datasets where similarity also implies same class membership we can expect RDCNN to be highly accurate and useful.

Keywords: 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 [9] have previously been proposed for this problem [12, 2, 6].

Here we investigate a depthwise convolutional neural network with random weights (RDCNN) [15]. Previously this has been shown to classify images with similar background, color, and texture accurately as evaluated on existing benchmarks [1]. 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

We describe our datasets followed by methods.

2.1 Datasets

We obtain four publicly available datasets spanning thee different cancers. These are available upon request to reproduce our results in the paper.

IDC - breast cancer The Invasive Ductal Carcinoma (IDC) dataset is provided by ICPR 2012 contest [3]. 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].

ISIC - skin cancer This dataset is provided by the ISIC 2019 Challenge [4, 13]. This is for classifying skin cancer images among nine different diagnostic categories: Actinic Keratosis, Squamous Cell Carcinoma, Basal Cell Carcinoma, Seborrheic 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.

Gleason - prostate cancer Gleason 2019 dataset contains prostate cancer from H&E-stained histopathology images, which is provided by Gleason 2019 challenge (https://bmiai.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.

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 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) [15]. Each convolutional block in this 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 Figure 1 we show an example of the RDCNN 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 [5] on the final feature space. The source code of RDCNN is freely available from https://github.com/xyzacademic/RandomDepthwiseCNN.

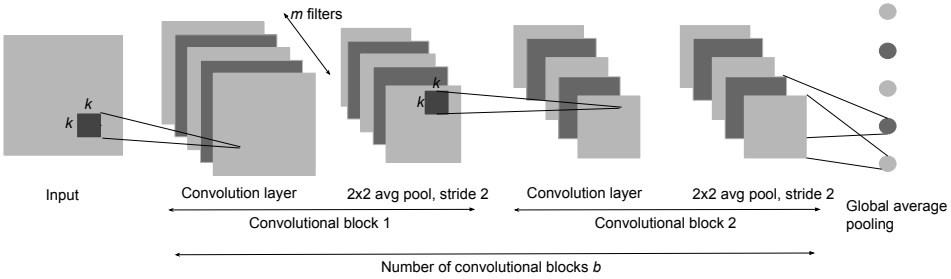


Fig. 1. A random depthwise convolutional neural network with two convolutional blocks, kernel size of k, and $m = 5$ kernels in each layer

2.4 Deep networks compared in our study

We compare the RDCNN network 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 [7]: Residual convolutional networks contain connections from previous layers and not just the last one.
- VGG16 [10]: Deep convolutional neural network with several layers of convolution and pooling.

3 Results

We conducted preliminary experiments on RDCNN to study the effect of increasing features. As observed before [14, 15] we see that increasing features increases the test accuracy. We trained both ResNet50 and VGG16 with a batch size of 32 and center cropped images for the Gleason dataset (whose images are very large in dimensions). We also performed center cropping for ISIC since it improved its test accuracy. Images in IDC are already in small patches.

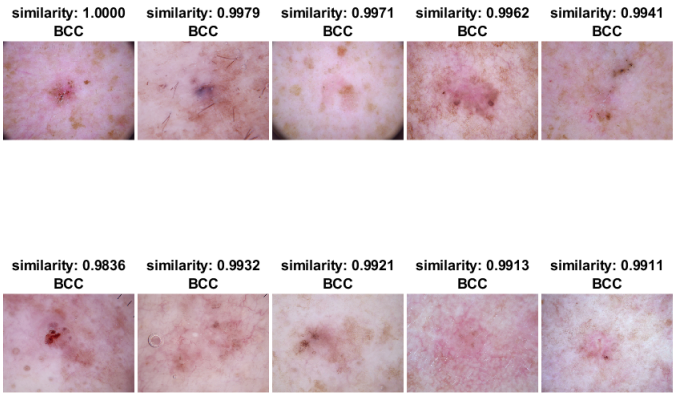
In Table 1 we see the train and test accuracies of VGG16, ResNet50, and RDCNN on our four datasets. On ISIC and Gleason we see that RDCNN achieves a much higher accuracy and comparable on the remaining datasets. The kernel size of RDCNN has little effect on the datasets shown here. Overall across all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

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

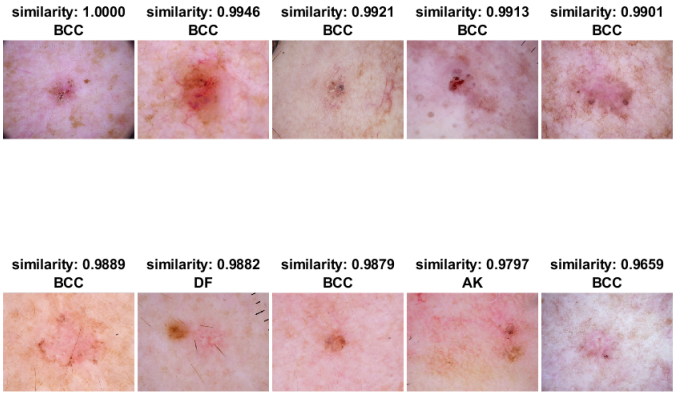
IDC		
Method	Train	Test
VGG16	92.2	83.3
ResNet50	100	88.2
RDCNN (30K features, k=3, 4 layers)	87.8	87.6
RDCNN (50K features, k=5, 4 layers)	86.3	87.6
ISIC		
Method	Train	Test
VGG16	89.8	85.9
ResNet50	90.3	87.5
RDCNN (65K features, k=3, 4 layers)	100	100
RDCNN (65K features, k=5, 4 layers)	100	100
Gleason		
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

To explain why RDCNN performs better in our study we take a random image from the ISIC dataset and show its top ten similar images in the final layer feature space of each network (see Figure 2). In the RDCNN space all ten similar images are in the same category as the query. But in ResNet50 and VGG16 there are two and three respectively from different classes. In datasets such as ISIC where similarity also implies same category, the unsupervised RDCNN network will outperform trained models.

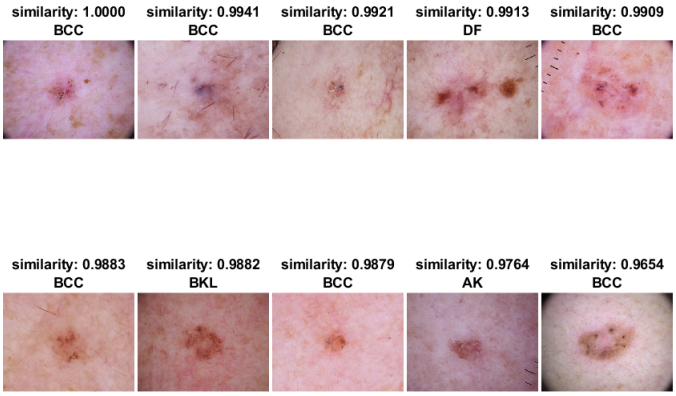
In Figure 3 we show the top ten similar images to a randomly selected one from the Gleason dataset. We see that in the RDCNN final layer space seven of the ten images are in the same category as the query whereas in the ResNet50 and VGG16 final layer space only six and five respectively are in the same category as the query.



(a) Top similar images to the query in RDCNN final layer feature space

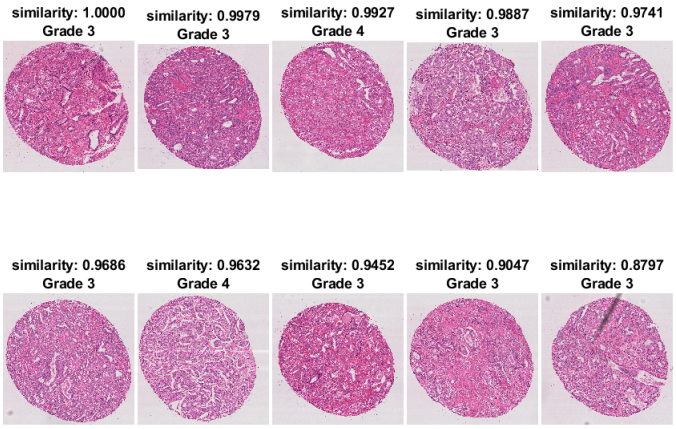


(b) Top similar images to the query in the ResNet50 final layer feature space

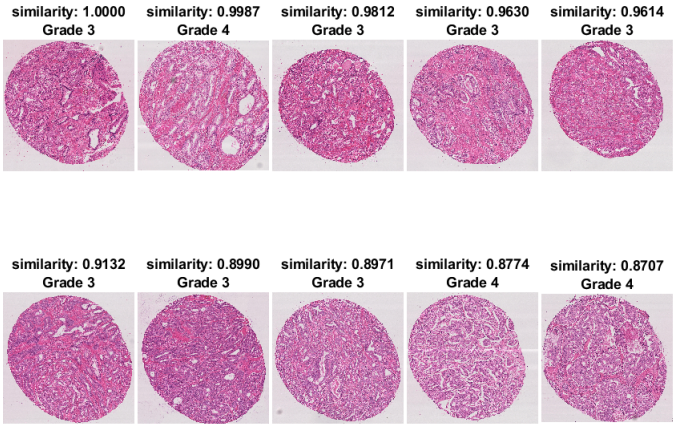


(c) Top similar images to the query in the VGG16 final layer feature space

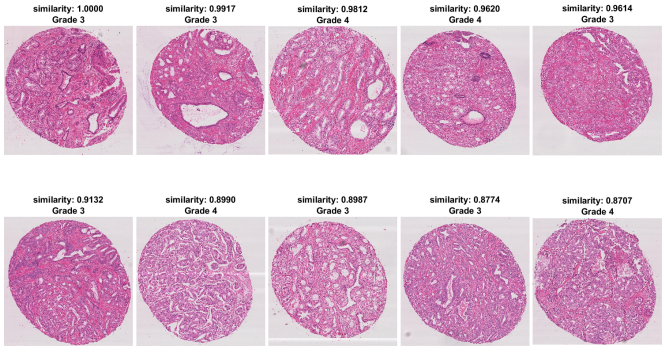
Fig. 2. Top 10 similar images in the ISIC dataset to a randomly selected one (also from ISIC) are in the same class as the query in the RDCNN feature space but not in ResNet50 and VGG16 space.



(a) Top similar images to the query in RDCNN final layer feature space



(b) Top similar images to the query in the ResNet50 final layer feature space



(c) Top similar images to the query in the VGG16 final layer feature space

Fig. 3. Most of the top 10 similar images in the Gleason dataset to a randomly selected one (also from Gleason) are also in the same class as the query in the RDCNN feature space but less so in the ResNet50 and VGG16 space.

4 Conclusion

Our preliminary results suggest that the unsupervised RDCNN can be highly useful in the classification of histopathology images where similarity also implies same class membership. We see that a kernel size of 3 and a total of 4 layers works well in most cases.

References

1. Corel-princeton image similarity benchmark. http://www.cs.princeton.edu/cass/benchmark/
2. Bejnordi, B.E., Mullooly, M., Pfeiffer, R.M., Fan, S., Vacek, P.M., Weaver, D.L., Herschorn, S., Brinton, L.A., van Ginneken, B., Karssemeijer, N., et al.: Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. Modern Pathology **31**(10), 1502 (2018)
3. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 411–418. Springer (2013)
4. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., 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). pp. 168–172. IEEE (2018)
5. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297 (1995)
6. Gertych, A., Swiderska-Chadaj, Z., Ma, Z., Ing, N., Markiewicz, T., Cierniak, S., Salemi, H., Guzman, S., Walts, A.E., Knudsen, B.S.: Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. Scientific reports **9**(1), 1483 (2019)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. Janowczyk, A., Basavanhally, A., Madabhushi, A.: Stain normalization using sparse autoencoders (stanosa): Application to digital pathology. Computerized Medical Imaging and Graphics **57**, 50–61 (2017)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
10. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
11. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering **63**(7), 1455–1462 (2015)
12. Talo, M.: Convolutional neural networks for multi-class histopathology image classification. arXiv preprint arXiv:1903.10035 (2019)
13. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**, 180161 (2018)

14. Xie, M., Roshan, U.: Exploring classification, clustering, and its limits in a compressed hidden space of a single layer neural network with random weights. In: International Work-Conference on Artificial Neural Networks. pp. 507–516. Springer (2019)
15. Xue, Y., Roshan, U.: Image classification and retrieval with random depthwise signed convolutional neural networks. In: International Work-Conference on Artificial Neural Networks. pp. 492–506. Springer (2019)