ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/323945807

Improve the Performance of Transfer Learning Without Fine-Tuning Using Dissimilarity-Based Multi-view Learning for Breast Cancer Histology Images

Conference Paper · June 2018

DOI: 10.1007/978-3-319-93000-8_88

CITATIONS
6

READS
170

4 authors, including:

Hongliu Cao
École de Technologie Supérieure
7 PUBLICATIONS **13** CITATIONS

SEE PROFILE

Simon Bernard
Université de Rouen
25 PUBLICATIONS **322** CITATIONS

SEE PROFILE

Robert Sabourin
École de Technologie Supérieure
455 PUBLICATIONS **7,673** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Deep Feature learning, dynamic selection of classifiers, meta-learning, PR in changing environments, evolutionary computation, offline signature verification & Face recognition in video View project

Vérification identité View project

All content following this page was uploaded by Robert Sabourin on 12 October 2018.

The user has requested enhancement of the downloaded file.

Improve the performance of transfer learning without fine-tuning using dissimilarity-based multi-view learning for breast cancer histology images

Hongliu CAO[1,2*], Simon Bernard[2], Laurent Heutte[2], and Robert Sabourin[1]

[1] Laboratoire d'Imagerie, de Vision et d'Intelligence Artificielle, École de Technologie Supérieure,
Université du Québec, Montreal, Canada,
[2] Université de Rouen Normandie, LITIS (EA 4108), BP 12 - 76801 Saint-Étienne du Rouvray, France

Abstract. Breast cancer is one of the most common types of cancer and leading cancer-related death causes for women. In the context of ICIAR 2018 Grand Challenge on Breast Cancer Histology Images, we compare one handcrafted feature extractor and five transfer learning feature extractors based on deep learning. We find out that the deep learning networks pretrained on ImageNet have better performance than the popular handcrafted features used for breast cancer histology images. The best feature extractor achieves an average accuracy of 79.30%. To improve the classification performance, a random forest dissimilarity based integration method is used to combine different feature groups together. When the five deep learning feature groups are combined, the average accuracy is improved to 82.90% (best accuracy 85.00%). When handcrafted features are combined with the five deep learning feature groups, the average accuracy is improved to 87.10% (best accuracy 93.00%).

Keywords: Breast Cancer, Dissimilarity, Random forest, Deep Learning, Multi-View, Transfer learning, high dimensional low sample size

1 Introduction

The detection and treatment of cancer are still very challenging. The normal process of cancer detection is from certain signs and symptoms to the further investigation by medical imaging and at last confirmed by biopsy [1, 2]. The diagnosis of breast cancer usually uses the biopsy tissue. The pathologists can histologically assess the microscopic structure and elements of the tissue from breast tissue biopsies [3].

One of the most important method for tumor histological examination in pathology is Hematoxylin and eosin (H&E) staining [4]. However, manual analysis is experience based, qualitative and always causes intra- or inter-observers

*Corresponding author, email: caohongliu@gmail.com

variation even for experienced pathologists [5]. Hence developing a more efficient, accurate, quantitative and automated system is necessary and urgent. Due to the high performance of deep learning networks, more and more studies used deep learning for the classification of breast cancer images [6]. However, the number of images available has always been an obstacle for the use of deep learning. Many studies divide images into patches for data augmentation, but the new problem is that there are no label information for patches.

In this paper, transfer learning without fine-tuning is proposed to solve the above problems. Six different feature extractors are compared, including five deep learning architectures and a traditional feature extractor combining PFTAS (Parameter-Free Threshold Adjacency Statistics) and GLCM (Gray Level Co-Occurrence Matrices) features. When all features are combined, there are mainly three challenges from the machine learning point of view: (i) small sample size: size: like most other medical applications, the number of breast cancer histology images is very small (400 images); (ii) high dimensional feature space: as six groups of features may be combined, the size of the feature space may be up to 31855, which is over 80 times bigger than the sample size; (iii) multiple feature groups: it may be hard to improve the learning performance by exploiting the complementary information that different groups contain [7]. To deal with these three challenges, we propose to treat breast cancer histology image classification as a multi-view learning problem. A multi-view RFSVM method proposed in our previous work [7] is then used as a solution.

The remainder of this paper is organized as follows: the six feature extractors are detailed in Section II; in Section III, the dissimilarity based multi-view learning solution is introduced; we describe in Section IV the data sets chosen in this study and provide the protocol of our experimental method; we analyze in Section V the results of our experiments; the final conclusion and future works are drawn in Section VI.

2 Feature extractors

In total six different feature extractors are used in this work: handcrafted features, ResNet-18, ResNeXt, NASNet-A, ResNet-152 and VGG16. In this section, a brief introduction of each feature extractor is given. The handcrafted features include PFTAS and GLCM and have been chosen due to their good performance on breast cancer histology image classification [8]. The five deep learning networks have been chosen for their performance and because they are built on different structures with different depths, and the pre-trained models are available online[3,4].

Handcrafted features: Two kinds of feature extractors are combined together to form the handcrafted feature group: PFTAS and GLCM. TAS (Threshold Adjacency Statistics) is a simple and fast morphological measure for cell phenotype

[3]https://github.com/Cadene/pretrained-models.pytorch
[4]https://github.com/pytorch/vision/tree/master/torchvision

image classification presented by Hamilton et al. in [9]. Similar to the work of [8], we use the Parameter-Free Threshold Adjacency Statistics (PFTAS) from the python library Mahotas [10] to build a 162-dimensional PFTAS-feature vector. GLCM features are widely used to describe the texture of tumor in cancer applications. Same as PFTAS, the library Mahotas is used to calculate the GLCM features leading to a 175-dimensional GLCM-feature vector.

ResNet-18 and ResNet-152: ResNet is one of the deepest deep learning architectures proposed by Microsoft researchers. The deep residual nets based methods have won the first places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation as well as the first place on the ILSVRC 2015 classification task [11]. We use two ResNet in this work: ResNet-18 and ResNet-152. Both networks take as input a {3, 224, 224} RGB image and are pretrained on ImageNet with 1000 classes[5,6]. Features are extracted from the average pool layer (i.e. before the last classification layer), which results in 512 features for ResNet-18 and 2048 features for ResNet-152.

ResNeXt: ResNeXt is one of the state-of-the-art techniques for object recognition. It builds upon the concepts of repeating layers while exploiting the split-transform-merge strategy to bring about a new and improved architecture [12]. The input space of ResNeXt is a {3, 224, 224} RGB image and we use the network pretrained on ImageNet with 1000 classes[7]. 2048 features are extracted from the average pool layer (i.e. before the last classification layer).

NASNet-A: In the work of [13], the authors proposed to search for an architectural building block on a small dataset and then transfer the block to a larger dataset to reduce the computation cost and improve the efficiency. They used NAS (Neural Architecture Search) framework from [14] as the main search method for their NASNets. The three networks constructed from the best three searches are named NASNet-A, NASNet-B and NASNet-C respectively. In this work, a NASNet-A pretrained on ImageNet is used. The input space of NASNet-A is a {3, 331, 331} RGB image and we use the network pretrained on ImageNet with 1001 classes (ImageNet+background)[8]. 4032 features are extracted from the last layer before the classification layer.

VGG16: The VGG Network was introduced by the researchers at Visual Graphics Group at Oxford [15]. This network is specially characterized by its pyramidal shape. VGG16 takes as input a {3, 224, 224} RGB image and we use the network pretrained on ImageNet with 1000 classes[9]. Features are extracted from the last max pooling layer, which results in 512x7x7 features.

[5] https://download.pytorch.org/models/resnet18-5c106cde.pth
[6] http://data.lip6.fr/cadene/pretrainedmodels/fbresnet152-2e20f6b4.pth
[7] http://data.lip6.fr/cadene/pretrainedmodels/resnext101-64x4d-e77a0586.pth
[8] https://data.lip6.fr/cadene/pretrainedmodels/nasnetalarge-a1897284.pth
[9] https://download.pytorch.org/models/vgg16-397923af.pth

3 DISSIMILARITY-BASED LEARNING

In our previous work [7], we proposed to use RFSVM to integrate information from different views together (each feature group is a view in multi-view learning framework). We have shown that RFSVM offers a good performance on Radiomics data. The RFSVM method can deal well with high dimensional low sample size multi-view data because: (i) RFSVM uses random forest dissimilarity measure to transfer each view of the data to a dissimilarity matrix so that the data dimension is reduced without feature selection, and at the same time the data in each view become directly comparable; (ii) RFSVM can take advantage of the complementary information contained in each view by combining the dissimilarity matrices together. We now recall the RFSVM method.

Random forest: Given a training set $\mathbf{T}$, a Random Forest classifier $\mathbf{H}$ is a classifier made up of M trees denoted as in Equation (1):

$$\mathbf{H}(\mathbf{X}) = \{h_k(\mathbf{X}), k = 1, \ldots, M\} \tag{1}$$

where $h_k(\mathbf{X})$ is a random tree grown using the Bagging and the Random Feature Selection techniques as in [16]. For predicting the class of a given query point $\mathbf{X}$ with such a tree, $\mathbf{X}$ goes down the tree structure, from its root till its terminal node. The prediction is given by the terminal node (or leaf node) in which $\mathbf{X}$ has landed. We refer the reader to [16] for more information about this process. Hence if two query points land in the same terminal node, they are likely to belong to the same class and they are also likely to share similarities in their feature vectors, since they have followed the same descending path.

Random Forest Dissimilarity (RFD): the RFD measure is inferred from a RF classifier $\mathbf{H}$, learned from $\mathbf{T}$. Let us firstly define a dissimilarity measure inferred by a decision tree $d^{(k)}$: let L_k denote the set of leaves of the kth tree, and let $l_k(\mathbf{X})$ denote a function from $\mathbb{X}$ to L_k that returns the leaf node of the kth tree where a given instance $\mathbf{X}$ lands when one wants to predict its class. The dissimilarity measure $d^{(k)}$, inferred by the kth tree in the forest is defined as in Equation (2): if two training instances $\mathbf{X}_i$ and $\mathbf{X}_j$ land in the same leaf of the kth tree, then the dissimilarity between both instances is set to 0, else set to 1.

$$d^{(k)}(\mathbf{X}_i, \mathbf{X}_j) = \begin{cases} 0, & \text{if } l_k(\mathbf{X}_i) = l_k(\mathbf{X}_j) \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

The RFD measure $d^{(\mathbf{H})}$ consists in calculating the $d^{(k)}$ value for each tree in the forest, and to average the resulting dissimilarity values over the M trees, as in Equation (3):

$$d^{(\mathbf{H})}(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{M} \sum_{k=1}^{M} d^{(k)}(\mathbf{X}_i, \mathbf{X}_j) \tag{3}$$

Multi-view learning dissimilarities: For multi-view learning tasks, the training set $\mathbf{T}$ is composed of K views: $\mathbf{T}^{(k)} = \{(\mathbf{X}_1^{(k)}, y_1), \ldots, (\mathbf{X}_N^{(k)}, y_N)\}$, k=1..K. Firstly, for each view $\mathbf{T}^{(k)}$, the RFD matrix is computed and noted

as $\{\mathbf{D}_\mathbf{H}^k, k = 1..K\}$. In multi-view learning, the joint dissimilarity matrix can typically be computed by averaging over the K matrices as in Equation (4):

$$\mathbf{D_H} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{D}_\mathbf{H}^i \tag{4}$$

Multi Random Forest kernel SVM (RFSVM): From the joint RFD matrix $\mathbf{D_H}$ of Equation (4), one can calculate the joint similarity matrix $\mathbf{S_H}$ as in Equation (5):

$$\mathbf{S_H} = \mathbf{1} - \mathbf{D_H} \tag{5}$$

where $\mathbf{1}$ is a matrix of ones. SVM is one of the most successful classifier. Apart from the most used gaussian kernel, a lot of custom kernels can also be used: we use the joint similarity matrix $\mathbf{S_H}$ inferred from the RF classifier $\mathbf{H}$ as a kernel in a SVM classifier.

4 EXPERIMENTS

The dataset used in this work is from ICIAR 2018 Grande Challenge on BreAst Cancer Histology images[10]. It is composed of Hematoxylin and eosin stained breast histology microscopy images. Microscopy images are labeled as normal, benign, in situ carcinoma or invasive carcinoma according to the predominant cancer type in each image. It is a balanced dataset with in total 400 images.

The protocol of the experiments is as follows:

- First, the 6 feature extractors described in Section 2 are used to extract features from histology image data. As there is no patch label provided, to simplify the feature extracting process, all images are rescaled to the network input size.
- Second, for each group of features, a random forest with 500 trees is built. The performance of each feature group is measured by the classification accuracy of the random forest. The random forest dissimilarity matrix is calculated for each group too.
- Finally, the RFSVM method described in Section 3 is used to combine all the groups together. Two RFSVMs are used: *RFSVM (DL only)* combines the five deep learning based feature groups; *RFSVM-All* combines all the six feature groups. For RFSVM, the search range of parameter C for SVM is $\{0.01, 0.1, 1, 10, 100, 1000\}$.

Note that in [17], the authors found that when dealing with high dimensional low sample size data, stratification of the sampling is central for obtaining minimal misclassification. In this work, the stratified random splitting procedure is repeated 10 times, with 75% as training data and 25% as testing data. In order to compare the methods, the mean and standard deviations of accuracy were evaluated over the 10 runs. However, for the contest, only one model can be submitted. Hence the best performance among the 10 runs is also presented and chosen as the model for the contest.

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

5 Results

The results of the experiments are shown in Table 1. We can tell that the best feature extractor is ResNet-152 with an average accuracy of 79.30% and best accuracy of 83.00%. Followed by ResNeXt with an average accuracy of 78.60% and the best accuracy of 81.00%. Surprisingly, the worst feature extractor is handcrafted features with PFTAS and GLCM with an average accuracy of 67.00%. In the work of [8], PFTAS and GLCM are the best features for breast cancer histology image classification. By comparing the performance of the six feature extractors, we can see that even though the deep learning networks are pretrained on ImageNet dataset, which is very different from histology images, they still have a better performance as a feature extractor for breast cancer data than the best handcrafted feature extractor used in the field of breast cancer histology image classification.

Table 1. The image wise classification results with 75% training data and 25% test data. *Average* is the average accuracy over 10 runs, *Best* is the best accuracy among the 10 runs.

	Average	Best
Handcrafted	$67.00\% \pm 5.46$	76.0%
ResNet-18	$75.10\% \pm 5.46$	78.0%
ResNeXt	$78.60\% \pm 1.74$	81.0%
NASNet-A	$74.70\% \pm 2.33$	78.0%
ResNet-152	$79.30\% \pm 3.20$	83.0%
VGG16	$68.00\% \pm 5.04$	78.0%
RFSVM(DL only)	$82.90\% \pm 1.37$	85.0%
RFSVM-All	$\mathbf{87.10}\% \pm 2.17$	**93.0**%

With *RFSVM (DL only)* integrating all the five deep learning based feature groups together, the average accuracy is improved to 82.90% and the best performance is improved to 85.00%. However, when all feature groups are combined with *RFSVM-All*, the average accuracy is improved to 87.10% and the best performance is improved to 93.00%. It shows that even though the handcrafted features do not have a very good performance individually, they can still provide useful complementary information for breast cancer classification when combined with deep learning based feature groups.

The confusion matrix, sensitivity and specificity of our best model are shown in Table 1. From the results we can see that our model has very high sensitivity on all four classes, and very high specificity too for two classes, i.e. *InSitu* and *Invasive*.

Table 2. The confusion matrix, sensitivity and specificity of our best model.

	Benign	InSitu	Invasive	Normal
Benign	23	1	0	1
InSitu	0	23	0	2
Invasive	1	0	24	0
Normal	2	0	0	23
Sensitivity	92%	92%	96%	92%
Specificity	85%	96%	100%	85%

Note that the state of the art performance on this dataset is considered to be from [3]. In this work, the authors used CNN patch-wise training on a previous version of the dataset with 249 images for training and 20 images for testing (7.4% of the whole dataset as test data). They obtained as best performance an accuracy of 85.00%. In our work, 300 images are used for training and 100 images are used as test data. Hence, even if the results are not directly comparable with [3], the accuracy of our best model is 8% higher than the accuracy reported in [3] while using 25% of the whole dataset as test data, which is much more than 7.4% in [3].

6 Conclusion

In this work, we firstly compared the popular handcrafted features used in breast cancer histology image classification with five deep learning based feature extractors pretrained on ImageNet. Not surprisingly, the experimental results show that the deep learning based features are better than the handcrafted. To improve the performance of transfer learning, we tackled the problem of breast cancer histology image classification as an HDLSS multi-view learning task and applied an RFSVM method previously proposed for the classification of Radiomics data. The results obtained with *RFSVM (DL only)* show that the performance of transfer learning can be improved by combining multiple feature extractors together. The results obtained with *RFSVM-All* show that even though deep learning based features have better performance than handcrafted features for breast cancer histology image classification, the accuracy can be improved significantly when they are combined together and surpass the state of the art performance on the dataset used.

Acknowledgment

This work is part of the DAISI project, co-financed by the European Union with the European Regional Development Fund (ERDF) and by the Normandy Region.

References

1. T. P. Coroller, P. Grossmann, Y. Hou, E. R. Velazquez, R. T. Leijenaar, G. Hermann, P. Lambin, B. Haibe-Kains, R. H. Mak, and H. J. Aerts, "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma," *Radiotherapy and Oncology*, vol. 114, no. 3, pp. 345–350, 2015.
2. H. Aerts, E. R. Velazquez, R. Leijenaar, C. Parmar, P. Grossmann, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat Commun*, vol. 5, pp. 1–8, 2014.
3. T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," *PloS one*, vol. 12, no. 6, p. e0177544, 2017.
4. J. K. Chan, "The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology," *International journal of surgical pathology*, vol. 22, no. 1, pp. 12–32, 2014.
5. J. S. Meyer, C. Alvarez, C. Milikowski, N. Olson, I. Russo, J. Russo, A. Glass, B. A. Zehnbauer, K. Lister, and R. Parwaresch, "Breast carcinoma malignancy grading by bloom–richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index," *Modern pathology*, vol. 18, no. 8, p. 1067, 2005.
6. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *Neural Networks (IJCNN), 2016 International Joint Conference on*, pp. 2560–2567, IEEE, 2016.
7. H. Cao, S. Bernard, L. Heutte, and R. Sabourin, "Dissimilarity-based representation for radiomics applications," *arXiv preprint arXiv:1803.04460*, 2018.
8. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016.
9. N. A. Hamilton, R. S. Pantelic, K. Hanson, and R. D. Teasdale, "Fast automated cell phenotype image classification," *BMC bioinformatics*, vol. 8, no. 1, p. 110, 2007.
10. L. P. Coelho, "Mahotas: Open source software for scriptable computer vision," *Journal of Open Research Software*, vol. 1, 2013.
11. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
12. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pp. 5987–5995, IEEE, 2017.
13. B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," *arXiv preprint arXiv:1707.07012*, 2017.
14. B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578*, 2016.
15. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
16. G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.
17. J. Bill and E. Fokoué, "A comparative analysis of predictive learning algorithms on high-dimensional microarray cancer data," *Serdica Journal of Computing*, vol. 8, no. 2, pp. 137–168, 2014.

View publication stats