TRANSFER LEARNING OF A CONVOLUTIONAL NEURAL NETWORK FOR HEP-2 CELL IMAGE CLASSIFICATION

Ha Tran Hong Phan[1], Ashnil Kumar[1], Jinman Kim[1], Dagan Feng[1]*

[1] Biomedical and Multimedia Information Technology (BMIT) Research Group
Institute of Biomedical Engineering & Technology, Faculty of Engineering & Information Technology
The University of Sydney, Australia
[*] tpha9954@uni.sydney.edu.au

ABSTRACT

The recognition of the staining patterns of Human Epithelial-2 (HEp-2) cells in indirect immunofluorescence (IIF) images is essential for the diagnosis of several autoimmune diseases. The main challenge is the extraction and selection of the optimal feature set that not only represents the cells' characteristics, but also distinguishes between the classes of cell images with similar appearances. In this paper, we propose a system to classify HEp-2 cell images by applying transfer learning from a pre-trained deep convolutional neural network (CNN) to extract the generic features and then using a feature selection method to get the most relevant features for classification. Although the CNN was trained with a dataset very different from cell images, our system is capable of extracting important semantic features that represent a HEp-2 cell image. When evaluated on the ICPR2012 cell dataset, our method outperforms all other methods on the dataset of the 2012 competition, and demonstrates stable performance under different test protocols.

Index terms – staining patterns, classification, indirect immunofluorescence, deep convolutional neural networks, transfer learning

1. INTRODUCTION

Indirect Immunofluorescence (IIF) is an important technique for the diagnosis of several autoimmune diseases. IIF images are obtained by staining a biological tissue with antibodies that are tied to a fluorescent chemical compound to detect the distribution of the target antibodies. The observed patterns are the essential information for the diagnosis. The analysis of a large volume of IIF images requires highly trained and specialized personnel, and is affected by very high inter laboratory variability, which is up to 24% [1]. The automation of IIF analysis will lead to cost and time savings and more importantly an improvement in the reproducibility of the results. Computer-Aided Diagnosis systems may overcome these limitations and provide easy, fast and reliable support to assist specialists in decision-making.

The growing interest for innovative solutions for the analysis of these images has led to the organization of the first HEp-2 Cell Classification Contest in 2012 (ICPR2012), which saw the participation of 28 research groups proposing different algorithms for the classification task of Hep-2 cell images. By measuring the algorithms' performance on a common database, researchers were given a comparative evaluation of their methods' performance and could improve their techniques based on the discoveries of other participants.

The accuracy rates achieved by published algorithms for the classification of these staining patterns varies depending on the test protocols. For example, Di et al.'s method [2] has 89.6% accuracy in the leave-one-out scheme, but scores only 49% accuracy when tested on the training and test datasets by the ICPR2012 contest.

1.1. Related works

A variety of methods have been proposed for pattern classification of HEp-2 cells. The majority of these methods utilize hand-crafted features, based on prior knowledge, such as local binary pattern (LBP) or scale-invariant feature transform (SIFT) and its variants, for feature extraction. These methods do not require extensive training and a massive amount of labelled samples, but are more problem-dependent than deep learning methods, which recently emerge in computer vision. For pattern classification, various classifiers are used, such as k-Nearest Neighbor, Multiclass SVM and Multiple Kernel SVM. Overall, the selection of the features is far more important than the choice of the classification algorithm [3].

Hand-crafted features may be application-specific and require devising new sets of features for new datasets. The ability to learn features automatically offers the freedom to apply the same method for various classification tasks on different datasets. The only current method that automatically extracts features is deep learning with convolutional neural networks (CNN). The architecture of CNNs, with convolutional layers interlaced with subsampling layers, allows feature extraction and pooling to obtain a feature representation of the original image. In [4], an eight-layer deep convolutional neural network system was designed for HEp-2 cells classification. This CNN was trained from scratch with extensive hyper-parameter optimization and data augmentation to achieve 69.9% accuracy on the ICPR2012 contest cell dataset.

978-1-4799-2349-6/16/$31.00 ©2016 IEEE

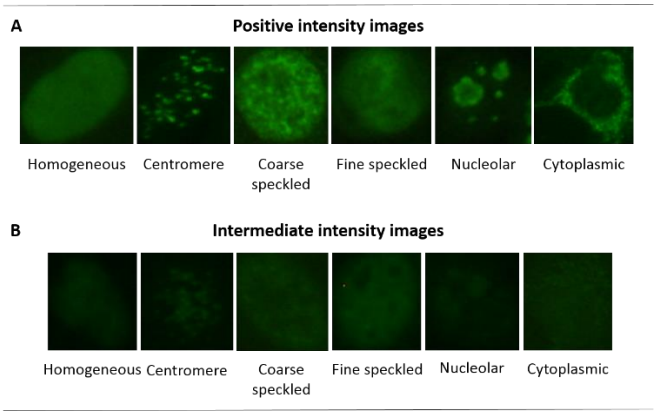

Fig. 1. HEp-2 cell images.

1.2. Motivation

A disadvantage of training a new CNN is that the training stage requires a large amount of labelled images. Furthermore, training or fine-tuning require extensive computation resources due to the estimation of millions of parameters. This is not feasible for medical studies. Fortunately, transfer learning from pre-trained models proves to ease the application of CNN and even boost the performance in some circumstances. Although many deep neural networks are trained on natural images, features are learned hierarchically from the first layers to deeper layers. Therefore, features from specific layers may be shared among images of different types. There are evidences that even features transferred from distant tasks are better than random weight initialization [5].

Besides the recently reported success among natural image data sets [6,7], transfer learning has been applied to study biological images. In [8], transfer learning proved to be beneficial for producing high-level representations of images of gene expression patterns in Drosophila melanogaster, a species of fly.

Because our aim is to overcome the shortage of training samples, and to provide a generic and problem-independent solution in biological applications, we chose transfer learning. With our newly devised framework, we will show that the last layers of our chosen CNN can also be used to extract high level semantic features that effectively represent different categories of biological images.

2. CONTRIBUTION

The main contributions of this paper are summarized below:

- We proposed a classification system that utilizes a pre-trained CNN model to extract features, a feature selection method to obtain the optimal feature set that distinguishes between the classes of cell images with similar appearances (E.g. coarse-speckled and fine-speckled classes), and a multiclass SVM classifier to detect the staining pattern of a HEp-2 cell image.
- We proved that a method based on transfer learning is effective in classifying cell images and performs stably regardless of test protocols.

3. METHOD

3.1. Dataset

The ICPR2012 dataset consists of fluorescence cell images in which each cell is manually segmented and annotated by specialists. Fluorescence intensity and staining pattern are also reported for each single cell. Mitotic cells and artifact are removed from the set of cell images, leaving only cells that can be categorized into six staining patterns: centromere (357 cells), nucleolar (241 cells), homogenous (330 cells), fine speckled (208 cells), coarse speckled (210 cells), and cytoplasmic (111 cells) (Fig. 1). In total, there are 1457 images of cells in the dataset, which are segmented from 28 large images, each belonging to one patient.

We utilized the ICPR2012 dataset because it has been publicly available for a long time since 2012, and there are many published methods tested on this baseline dataset that we can compare our method with.

3.2. The two layers classification system

Our system consists of two layers of classification with the first layer designed to differentiate between positive and intermediate intensity images, and the second layer performing the main classification task to assign each image to one of the six classes. As the low level reads primitive features that are responsible for the difference between the intensities of cell images, the output of the first layer of the CNN is used for the early classification stage. The second last layer of the CNN is used for the staining pattern classification because it generates a feature set of sufficient size that can represent the specific characteristics of HEp-2 cell images. Two separate SVMs are trained on two feature sets for high level representation that are selected and optimized specifically for the two categories of images, i.e. positive and intermediate intensity images (Fig. 2).

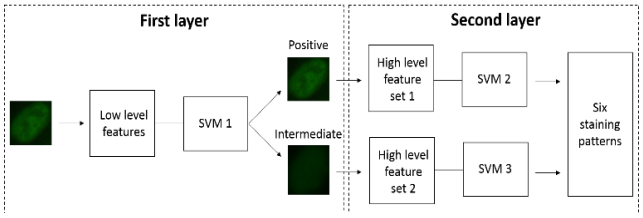

Fig. 2. Our system with two classification layers.

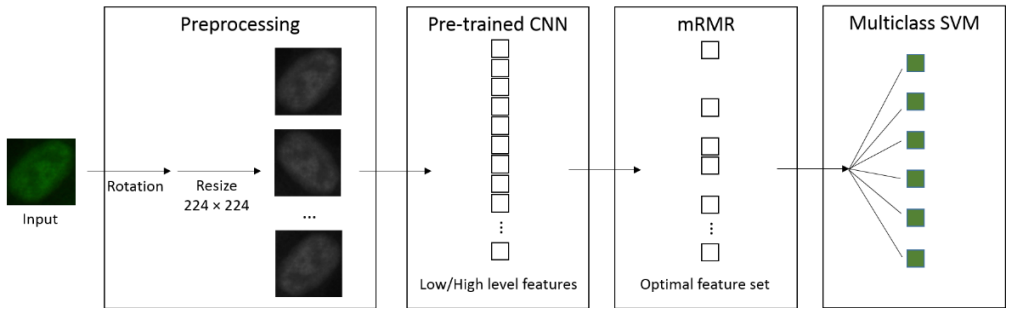

Fig. 3. The architecture for both classification layers in our system.

Conv 1	Conv 2	Conv 3	Conv 4	Conv 5	Full 6	**Full 7**	Full 8
64 filters 11×11 Stride 4 LRN Max pooling	256 filters 5×5 Stride 1 Padding 2 LRN Max pooling	256 filters 3×3 Stride 1 Padding 1	256 filters 3×3 Stride 1 Padding 1	256 filters 3×3 Stride 1 Padding 1 Max pooling	4096 Dropout	4096 Dropout	1000 Softmax

Fig. 4. CNN architecture. The CNN contains 5 convolutional layers (conv 1-5) and three fully-connected layers (full 6-8). LRN: Local Response Normalization. More details can be found in [10]. The output of the Conv 1 layer is the low level features that is used for the first classification layer. The output of the Full 7 layer is the 4096 dimensional high level representation of the input image that is used for the second classification layer.

3.3. Data augmentation

In order to increase the number of images when the training data is limited in the ICPR2012 dataset, data augmentation was performed by rotating each image with an angle step of 18 degree, which is suggested to achieve invariance in image rotation [4]. This also prevents overfitting by the classifier.

After data augmentation, if there is a skewed distribution of classes in the training dataset, the system will rebalance the distribution by automatically taking out random images in classes with excessive amount of training data.

3.4. Preprocessing

We converted the cell images into single-channel, grayscale images because they only exhibit shades of green. The images were also resized into a fixed size of 224×224 to guarantee uniform scale for all images, and subtracted by the mean value predefined for the CNN model.

3.5. Convolutional neural network

The pre-trained CNN model for feature extraction is the Fast architecture reported in [9]. This model is run using MatConvNet. There are more complex models that can achieve better performance, but with the trade-off of computation time. We chose this model to show that transfer learning is strongly effective in cell image analysis, even with a simple model. The CNN comprises 8 learnable layers, with the first 5 convolutional layers and the last 3 fully-connected layers (Fig. 4). It was trained on the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC-2012) dataset, which contains 1.2 million images and 1000 object categories. Considering the size and object categories of this dataset, this pre-trained model was intended to use for tasks that are very distant from our cell image classification task.

The CNN produces low and high level representations at the first and seventh hidden layers respectively. The outputs of these layers are taken as the sets of features that are later used in the two classification stages.

3.6. Feature selection and Classification

Feature selection is a data preprocessing step that can reduce the dimensionality of the input data and improve the chances of avoiding overfitting. The minimum-Redundancy-Maximum-Relevance (mRMR) algorithm, which sorts the features that are most relevant for the characterization of the classification variable by minimizing their mutual similarity and maximizing their correlation with the classification label, is chosen for its better performance over other conventional top-ranking methods [10].

This step takes as input the features extracted by CNN and selects the top ranked 5 features for the intensity classification task and 350 features for the staining pattern classification task. The later task is more complicated and requires a larger feature set than the intensity classification one. Linear multiclass SVM classifiers are trained with these feature sets.

4. DISCUSSION AND RESULTS

There are two complementing test protocols using the ICPR2012 dataset, presented in [3]. In the first protocol, the set of images is partitioned into training and test sets, with 723 and 734 cells respectively, in which the image pattern distribution over the two sets is maintained approximately the same as the original dataset. Despite the significant drop

in the amount of training samples, our proposed classification system (Fig. 2), with an accuracy rate of 77.1%, outperforms all the published algorithms [3] (Table I), including the CNN trained especially with this particular dataset [4], and a specialist in Immunology with 12 years of experience, who worked in exactly the same condition as the automatic methods [1].

In the second protocol, the leave-one-out scheme is applied when segmented cell images of one of the 28 large images are used as the test set and a classifier has to be trained using cell images from the remaining 27 large images [3]. Our method, with the accuracy rate of 91.5%, also obtained the first position among the published methods.

It is remarkable that no other algorithm can maintain high ranks under both test protocols. For example, the second most accurate algorithm by Di Cataldo [2] with the leave-one-out method, at 89.6%, only scores as low as 49% accuracy in the first protocol. By contrast, our method, always appearing on top of the lists, has shown its stable performance regardless of the test protocols.

TABLE I
CLASSIFICATION ACCURACY RATES (%).

Test protocol	Our algorithm (CNN)	Human expert [1]	Xu [12]	Nosaka [11]	Di Cataldo [2]
ICPR2012	**77.1**	73.3	75.2	69	49
Leave-one-out	**91.5**	-	-	68.5	89.6

5. CONCLUSION

In this paper, we proposed a classification system that extracts features using a pre-trained CNN model, selects an optimal feature set with mRMR, and utilizes multiclass SVM as the classifier. This system has achieved good and stable performance under rigorous testing schemes, and with limited training data, because of its ability to learn and select the best feature sets, depending on the given training data. It has outperformed methods using hand-crafted features, as well as other CNNs that learn features automatically. Our method also minimizes the involvement of humans, and proves to be applicable in many other classifications tasks for various types of cell images.

Moreover, our results indicate the applicability of transfer learning across distant tasks, when models trained with natural scene images can be used for analyzing biological images. This is a very promising idea and may become a future trend in biological and medical researches where there is a large variability in types of images and research intentions, but with only very limited available data.

Future work is in progress to further improve the accuracy of our algorithm. This will include using other larger and more complex pre-trained CNN models, fine-tuning, data augmentation, image preprocessing, and selection of optimal sizes for feature sets. Lastly, datasets other than the ICPR2012 one will be used for validating the generalization of our method.

6. REFERENCES

[1] Bizzaro, N., Tozzoli, R., Tonutti, E., Piazza, A., Manoni, F., Ghirardello, A., Rizzotti, P., "Variability between methods to determine ANA, anti-dsDNA and anti-ENA autoantibodies: A collaborative study with the biomedical industry," *J IMMUNOL METHODS*, 219, pp. 99–107, 1998.

[2] Di, S., Bottino, A., Ul, I., Figueiredo, T., & Ficarra, E., "Subclass Discriminant Analysis of morphological and textural features for HEp-2 staining pattern classification," *PATTERN RECOGN*, 47(7), pp. 2389–2399, 2014.

[3] Foggia, P., Percannella, G., Saggese, A., & Vento, M., "Pattern recognition in stained HEp-2 cells: Where are we now?," *PATTERN RECOGN,* 47, pp. 2305–2314, 2014.

[4] Gao, Z., Zhang, J., Zhou, L., & Wang, L., "HEp-2 Cell Image Classification with Convolutional Neural Networks," *I3A,* pp. 24–28, 2014.

[5] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H., "How transferable are features in deep neural networks ?," *ADV NEURAL INF PROCESS SYST,* 27, pp. 1–9, 2014.

[6] Sharif, A., Hossein, R., Josephine, A., Stefan, S., & Royal, K. T. H., "CNN Features off-the-shelf : an Astounding Baseline for Recognition," *IEEE CVPR*, pp. 512–519, 2014.

[7] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T., "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," *ICML*, 32, pp. 647–655, 2014.

[8] Wenlu Zhang, Rongjian Li, Tao Zeng, Qian Sun, Sudhir Kumar, Jieping Ye, and Shuiwang Ji., "Deep Model Based Transfer and Multi-Task Learning for Biological Image Analysis," *ACM SIGKDD*, pp. 1475-1484, 2015.

[9] Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A., "Return of the Devil in the Details: Delving Deep into Convolutional Nets," arXiv Preprint http://arxiv.org/abs/1405.3531, 2014

[10] Peng, H. C., Long, F. H., & Ding, C., "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE PAMI*, 27(8), pp. 1226–1238, 2015.

[11] Nosaka, R., & Fukui, K., "HEp-2 cell classification using rotation invariant co-occurrence among local binary patterns," *PATTERN RECOGN*, 47(7), pp. 2428–2436, 2014.

[12] Xu, X., Lin, F., Ng, C., & Leong, K., "Adaptive Co-occurrence Differential Texton Space for HEp-2 Cells Classification," In N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *MICCAI in LNCS* Vol. 9351, pp. 260–267, 2015.