Human Epithelial Type 2 Cell Classification with Convolutional Neural Networks

Neslihan Bayramoglu, Juho Kannala, and Janne Heikkilä

***Abstract*— Automated cell classification in Indirect Immunofluorescence (IIF) images has potential to be an important tool in clinical practice and research. This paper presents a framework for classification of Human Epithelial Type 2 cell IIF images using *convolutional neural networks* (CNNs). Previuos state-of-the-art methods show classification accuracy of 75.6% on a benchmark dataset. We conduct an exploration of different strategies for enhancing, augmenting and processing training data in a CNN framework for image classification. Our proposed strategy for training data and pre-training and fine-tuning the CNN network led to a significant increase in the performance over other approaches that have been used until now. Specifically, our method achieves a 80.25% classification accuracy. Source code and models to reproduce the experiments in the paper is made publicly available.**

I. Introduction

The analysis of Indirect Immunofluorescence (IIF) images are important in the diagnosis and monitoring of disease progression in patients with autoimmune diseases. The IIF images of Human Epithetial Type 2 (HEp-2) cells have been used to detect antinuclear antibodies (ANA) in human serum. The presence of these antibodies are strong markers for predicting and diagnosing autoimmune diseases. Type of the antibody and its binding to specific part of the cell results different nuclear patterns of fluorescence on the HEp-2 cells [1]. The identification of these IIF patterns has significant importance in making correct diagnosis.

Recently, automated analysis of indirect immunofluorescence images has received increasing attention because of the disadvantages of manual inspection such as its subjective nature and cost. The interest of image classification community in this problem is attracted with publicly available datasets and competitions and corresponding workshops [2]. The first edition of the HEp-2 cell dataset **(ICPR2012)**[1] was released in year 2012. This publicly released dataset contains a set of manually annotated training images associated with segmentation masks (Figure 2). A set of test images is also released, with the manual annotations and segmentation masks. The second edition of the dataset (used in contests in **ICIP2013** and **ICPR2014**)[2] is released with only training images. The test set is privately maintained by the organizers. ICPR2012 dataset allows comparison of HEp-2 image classification algorithms, and majority of existing approaches have evaluated their performance on this dataset. Therefore, in order to make comprehensive and fair comparisons, we performed experiments on the ICPR2012 dataset.

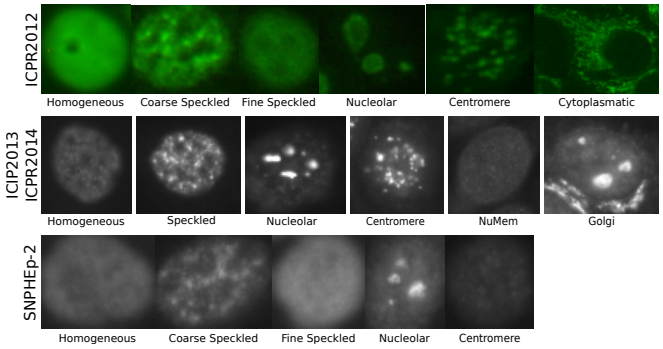


Fig. 1. Sample images from ICPR2012, ICPR2014, and SNPHEp-2 datasets.

In this paper, we study automated classification of HEp-2 cell images with Convolutional Neural Networks (CNNs). The main focus of the paper is to analyze data preprocessing, normalization and augmentation approaches for HEp-2 cell image classification.

II. Related Work

HEp-2 image classification problem has been studied in many recent papers [3], [4], [5], [6], [7], [8]. Most of the methods attempt to utilize texture information. The first prize winner [9] in the 2012 HEp-2 cells classification contest [2] employed *Local binary pattern* (LBP)-based method [10]. To date, the approach proposed by Qi et al.[6] achieved the best *average classification accuracy* on the ICPR2012 dataset by utilizing multi-resolution texture descriptor again based on LBP.

Recently, neural networks achieved a great success in object classification [11]. Specifically, *Convolutional Neural Network* (CNN) based approaches showed significant improvements over state-of-the-art recognition and classification approaches. Unlike hand-designed features like SIFT [12], CNNs learn image features from training data. It has been also shown that learnt image features with CNN architectures are often more discriminative than hand crafted image descriptors and perform better [13]. Today, there are numerous attempts to solve various computer vision problems using CNNs due to its remarkable achievements on several large scale image datasets [11]. It has also been applied to biomedical applications with success. For example, Ciresan et. al [14] detect mitosis in breast cancer

N. Bayramoglu (nyalcinb@ee.oulu.fi), J. Kannala (jkannala@ee.oulu.fi) and J. Heikkilä (jth@ee.oulu.fi) are with Center for Machine Vision Research, University of Oulu, Finland.

[1] http://mivia.unisa.it/hep2contest/index.shtml

[2] http://nerone.diem.unisa.it/contest-icip-2013-test/index.shtml

histology images and won the related contest. On the other hand, CNN is applied to HEp-2 cell image classification problem before only in a few studies [2], [15].

Despite CNN's success, there are several challenges in generalizing CNN frameworks to different applications such as biomedical images. First, CNNs typically require lots of training data due to large number of parameters to be learned. However, public databases in biomedical imaging applications are usually small due to the fact that annotation of such data requires expertise in that field. Second, the small interclass variation and high intraclass variation could be present in biomedical images as in HEp-2 datasets; therefore, optimizing CNN's hyper-parameters for small training data leads to overfitting and bias.

In this work, we study mentioned drawbacks of CNNs by experimentation. We apply CNNs to a small HEp-2 dataset published by [2] and compare our results with the state-of-the art. We explore the impact of data normalization, preprocessing and augmentation methods in HEp-2 cell classification. The most similar study to ours is described in [15], however, the focus in [15] is to optimize network hyper-parameters whereas we focus on the data itself. Our framework achieves a better classification accuracy than the previous state-of-the-art results on the ICPR2012 HEp-2 dataset [2].

III. Our Work

Color and Intensity: In fluorescent imaging, despite the existence of RGB image databases, pattern information is usually carried in a single (green) channel. We tested our framework both with color images and green channel images. Brightness and contrast properties vary greatly among database images. This could be due to the *"photobleaching effect"* and could be also due to the differences in labeling process and image acquisitions. Previous studies adopt different strategies to normalize image intensity. Scaling intensities between 0 and 1 and histogram equalization are the most common methods. In our study, we compare four different intensity schemes: i) original raw data, ii) histogram equalization, iii) scaling between 0 and 1, and iv) zero mean and unit variance normalization (MVN). Usually, normalization step is not an essential part of CNN based classifiers as the networks have capacity to handle variations as long as the training set has enough samples to represent these variations. Otherwise, intensity normalization may influence classification performance greatly.

Data Augmentation: Increasing the number of training examples is useful for small training sets in CNN frameworks. Most works perform data augmentation only artificially by employing affine transformations. We augment the training set in three ways: i) by affine transformations (rotations, flipping, scaling, etc.), ii) by importing real data from similar datasets, and iii) by adding intensity variations of the same sample.

Images are first flipped horizontally and then both original and flipped images are rotated. We rotate cell images around their centers with angles sampled uniformly between 0° and 360°. Borders in the rotated versions of input images require special intention due to discontinuities. Zero padding and inpainting techniques creates unrealistic images that could diminish the performance of data augmentation instead of enhancing. Therefore, we cropped rotated images in such a way that the cropped image corresponds to the largest rotated rectangle within the input image (Figure 2(f)). This transformation introduces scale and translation variations inherently. Augmenting training data with such transformations in CNN frameworks is useful in order to compensate for data variations between test and training sets.

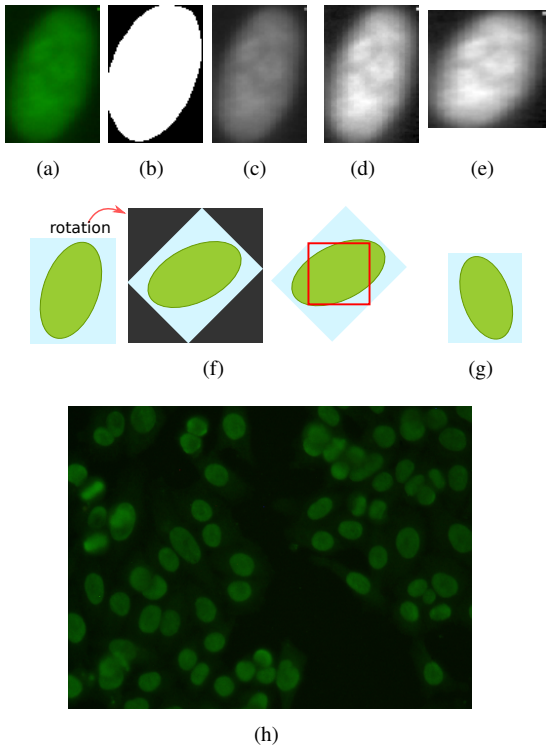


Fig. 2. a) A sample image from the ICPR2012 benchmark, b) segmentation mask, c) green channel, d) intensity normalized green channel, e) resampled image into a square size, f) to avoid border issues in the rotated images, largest rotated rectangle within the input image (*red*) is cropped, g) flipped horizontally, h) sample of a specimen image.

We expand the ICPR2012 training set with the ICPR2014 and the SNHEp-2 [8] datasets (Figure 1). Table I describes the compositions of test and training sets. ICPR2014 dataset is about ten times larger than the ICPR2012 but it has a different cell categorization. *Fine speckled* and *coarse speckled* classes in ICPR2012 are merged and represented as a new class called *speckled* in ICPR2014. In addition, two new categories *golgi* and *nuclear* membrane is introduced and *cytoplasmic* category is removed. Therefore, we add images only from the common classes of the ICPR2014 dataset, *homogeneous*, *nucleolar* and *centromere* (7833 samples), to the ICPR2012 training set. SNHEp-2 dataset, which has a similar size with the ICPR2012, comprised of five categories that exist in the ICPR2012. Therefore, SNHEp-2 dataset is entirely added to our training set. Addition of these images increased the training size considerably (721 vs. 9639).

Finally, we augment our training set by presenting exactly the same training examples with different intensity normalizations. In addition to the original raw images (Figure 2(c)), we add histogram equalized images (Figure 2(d)) to the training set. In the ICPR2012 dataset, individual cell images (Figure 2(a)) are actually cropped from a bigger specimen image (Figure 2(h)) which is comprised of multiple cells. Specimen images and individual cell locations are also provided by the benchmark. This enables us to perform another intensity correction. We compute histogram equalization on the specimen image and crop individual cell images. As a result, one sample is presented with three different intensities in the populated training set: i) raw pixel data; ii) histogram equalization before cropping, and iii) histogram equalization after cropping.

Pre-training and Fine Tuning: Fine tuning is the process of transferring learned parameters from pre-trained models to new ones. We pre-train our model on a larger dataset, then we fine tune the network parameters on the ICPR2012 training set. To avoid overfitting, we utilized all available HEp-2 image data also by including images from cell categories that do not exist in ICPR2012 from the ICPR2014 dataset. That is, we trained our network with 9 classes. Previously mentioned affine transformations are also applied to this larger pre-training set.

IV. Experiments and Results

The ICPR2012 dataset contains 28 HEp-2 specimen images (comprised of several cells) almost equally distributed among different classes were acquired by a fluorescence microscopy. Manual segmentation and annotation of individual cells were performed by specialists. After the segmentation, the ICPR2012 dataset contains $1,455$ cell images in total with six classes namely: homogeneous, coarse speckled, fine speckled, nucleolar, centromere, and cytoplasmic. Figure 1 shows sample images from the dataset. The dataset is divided into test (734 cell images) and training (721 cell images) sets. Since specimen images contain different number of cells, the distribution of cell patterns in the final dataset is imbalanced (Table I). Although it is a small set, there is a great variability among cell intensities and patterns within the same class.

In this work, we use Caffe [16] library to extract CNN-based features. We experiment with the architecture based on Alex Krizhevsky's *cuda-convnet* model [17]. The model composes of three convolutional layers, pooling layers, and rectified linear unit (ReLU) nonlinearities layers with a linear classifier on top of it all (Figure 3). Network layers and parameters are shown in Table II. These settings are deliberately kept fixed for all experiments to observe effects of changes in the training data. Despite having only three convolutional layers, the classification capacity of this network is sufficient for classifying our data since in our experiments we observed that the classification accuracy for the training set often reaches close to 100%. However, overfitting may still be a problem. Therefore, to avoid overfitting, the model is regularized with weight decay $w = 0.0004$. Additionally, we applied *dropout* regularization [18] which prevents complex node connections. While dropout has shown superior performance in [18], it does not have positive effect in our experiments. Therefore, we report our results without dropout regularization. Similar behavior with dropout regularization in HEp-2 image classification is also reported in [15].

The network is trained using the minibatch stochastic gradient descent with a momentum factor of 0.9. Each iteration operates on a minibatch of 256 images that are sampled randomly from the training set. The network is trained for 12,000 iterations with a learning rate of 0.001. For pre-training, fine tuning, and heavily populated data we use 120,000 iterations. All the images are resized to 64×64 both in training phase and testing phase (Figure 2(e)).

In our first experiment, we compared *color* images and *green* channel images. Figure 4(a) shows average classification accuracy of ICPR2012 test and training data in case of RGB images and green channel images. There is a clear difference between the results that shows the superiority of the green channel images over RGB. The reason may be due to the existence of intra-correlations between channel maps. Therefore, we used only green channel images in the rest of the experiments. Figure 4(b) shows performance of different intensity corrections. While histogram equalization and MVN improves the performance, the scaling approach has a negative effect. This observation is consistent with the common issue of having low-contrast images in biomedical imaging. In addition, scaling operation seems to introduce some numerical instability as the test set accuracy shows instant changes and training set reached its maximum accuracy much later than the others. On the other hand, both histogram equalization and MVN enhances contrast and brightness. Therefore, the intensity imbalance among samples are reduced and the network learns invariant properties.

We next investigated data augmentation. Results of the evaluations on the three datasets are shown in Figure 5(a). First we populate the ICPR2012 training set with the SNHEp-2 data (Exp2- blue curve). Then we expand the set with the ICPR2014 samples (Exp3-red curve). This time we add ICPR2014 samples only from the common cell categories (i.e. *homogeneous*, *nucleolar*, *centromere*). After augmenting the training set with the SNHEp-2 data, the accuracy is improved around 10% (Exp2) and an additional 10% improvement is gained after the introduction of the ICPR2014 samples (Exp3). Although augmenting with ICPR2014 samples makes the training set more imbalanced in numbers, the overall effect is positive. This indicates the significance of having more samples during learning. Note

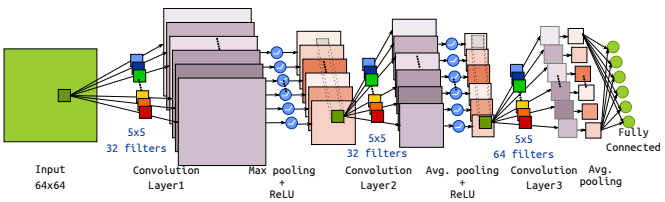


Fig. 3. Network architecture.

TABLE I

ICPR2012, ICPR2014, AND SNHEP-2 BENCHMARK PROPERTIES

	Number of Images in ICPR2012		
Pattern	**Test**	**Train**	**Total**
Homogeneous	180	150	330
Coarse speckled	101	109	210
Fine speckled	114	94	208
Nucleolar	139	102	241
Centromere	149	208	357
Cytoplasmic	51	58	109
Total	**734**	**721**	**1455**

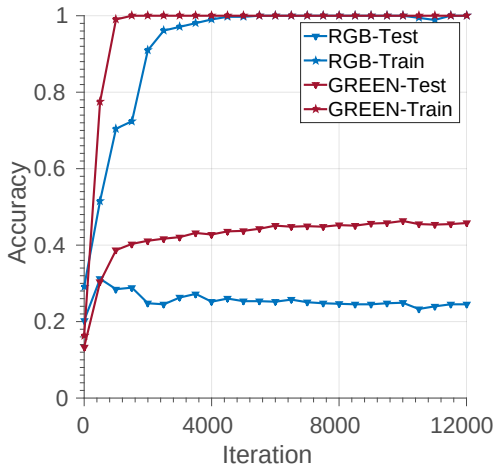
	No. of Images in ICPR2014
Pattern	**Train**
Homogeneous	2494
Speckled	2831
Nucleolar	2598
Centromere	2741
NuMem	2208
Golgi	724
Total	**13596**

	No. of Images in SNHEp-2		
Pattern	**Test**	**Train**	**Total**
Homogeneous	188	172	360
Coarse speckled	187	166	353
Fine speckled	191	188	379
Nucleolar	139	194	382
Centromere	149	149	332
Total	**937**	**869**	**1806**

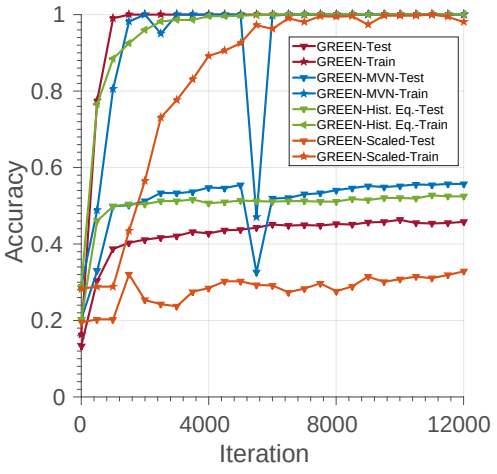
TABLE II

MODEL DESCRIPTION OF THE NETWORK.

Layer Type	Num. of Features	Filter Size	Pooling (Kernel/Stride)	Non-Linearity	Initial Weight
Convolutional	32	5×5	MAX., 3×3, 2	ReLU	$\mathcal{N}(0; 0.0001)$
Convolutional	32	5×5	AVG., 3×3, 2	ReLU	$\mathcal{N}(0; 0.01)$
Convolutional	64	5×5	AVG., 3×3, 2	ReLU	$\mathcal{N}(0; 0.01)$
Fully Connected	6	-	-	-	$\mathcal{N}(0; 0.01)$



(a)



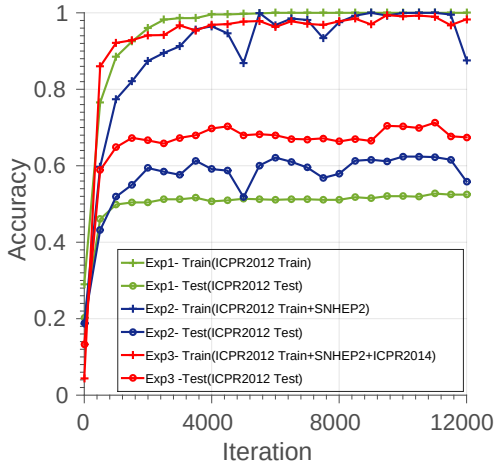
(b)

Fig. 4. Comparison of (a) color and (b) intensity normalization schemes.

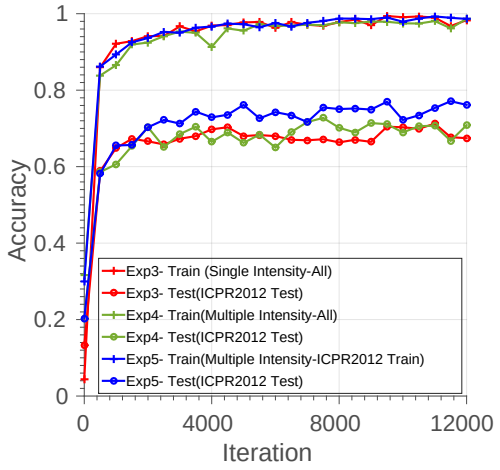
that in all experiments, histogram equalization is applied to both training and test sets.

After populating the training set with real samples, we apply artificial augmentation by presenting same training sample with different intensity normalizations (see Figure 5(b)). Specifically, our expanded training set (*Exp4-Train Multiple Intensity-All*) has the following composition: a) original ICPR2012 train set (*721 samples*), b) histogram equalized ICPR2012 train set before cropping (*721 samples*) and c) after cropping (*721 samples*), d) original SNHEp-2 (*1,806 samples*), e) histogram equalized SNHEp-2 (*1,806 samples*), f) original ICPR2014 (*common categories, 7,833 samples*), and g) histogram equalized ICPR2014 (*common categories, 7,833 samples*). Finally, there are $21,441$ samples in the training set. The classification performance is shown in Figure 5(b) with *green curve* (Exp4). Compared to our baseline (Exp3-*red curve*) which is the single intensity set, this arrangement does not bring clear improvements. One extra experiment is presented in Figure 5(b) with blue curve (*Exp5-Multiple Intensity-ICPR2012*). This time we have only ICPR2012 train set represented with several intensities. Interestingly, this composition boosts the performance of our classification model another 5%. Note that, in all experiments, the same single intensity test set *(ICPR2012-Test)* normalized with histogram equalization is used.

In Figure 6(a)-6(b), we compare training set performance augmented by all kind of transformations including rotations and flipping on top of multiple intensities against the baseline which does not include any artificial augmentation. We experiment here with 472,080 training samples obtained by horizontal flipping $11,802$ images of the previous augmentation and then rotating all with an angle step of $18°$. Maximum rectangular area within the rotated image is cropped as shown in Figure 2(f) and resized to 64×64. We report results for iterations up to $120,000$ as the training size is now much bigger than before. Figure 6(b) magnifies the early iterations for easy interpretation. As expected, the



(a)



(b)

Fig. 5. Average classification accuracy comparison for different data augmentation proposals. a) comparison of additional real samples, b) comparison of multiple intensity samples.

acceleration of training set accuracy is slower but rather high ($94,23\%$). On the other hand, the maximum of $78,20\%$ average classification accuracy is observed on the test set. This performance for the ICPR2012 is better than all other results reported in the literature (Table III).

Finally, in the pre-training process, we used a larger dataset by adding all categories from the ICPR2014. The same network used before is employed except the output of the last fully-connected layer. It is fed into a 9-way softmax layer with the multinomial logistic regression as the loss function, to produce a probability distribution over the 9 classes. Network parameters pre-trained on the larger dataset are directly transferred for fine-tuning, including the last fully-connected layer. In other words, for fine-tuning we also adopt 9-way softmax layer. The fine-tuning and related pre-training with all available data (including also additional classes) is applied in a similar setting as in the previous experiment which achieved the $78,20\%$ accuracy. The effects of pre-training and fine-tuning on CNN performance are shown in Figure 6(c). As we expected, pre-training with additional classes, followed by fine-tuning, brings another boost in performance. Finally, the average classification accuracy reaches up to $80,25\%$ at iteration $110,000$. The confusion matrix at this point is reported in Table IV. Note that, despite the high performance in five categories, the classification accuracy is rather low for the *fine speckled* category. Most samples in this category are classified as *homogeneous*. Previous works also report similar errors [7], [6]. The reason could be the similarities in the appearance of these classes (see Figure 1).

Table III compares our final result with the previous state-of-the-arts on the ICPR2012 benchmark. Our data augmentation and fine-tuning approach improves the previous state-of-the-art result by 5%. These results highlight four important points: (i) augmenting training data with real images is most beneficial, (ii) also augmentation with synthetic data using intensity transformations and geometric transformation is helpful, (iii) contrast normalization is substantially beneficial (e.g. histogram equalization), (iv) pre-training with all available data, including also additional classes, and followed by problem specific fine-tuning is useful.

V. Conclusion and Future Work

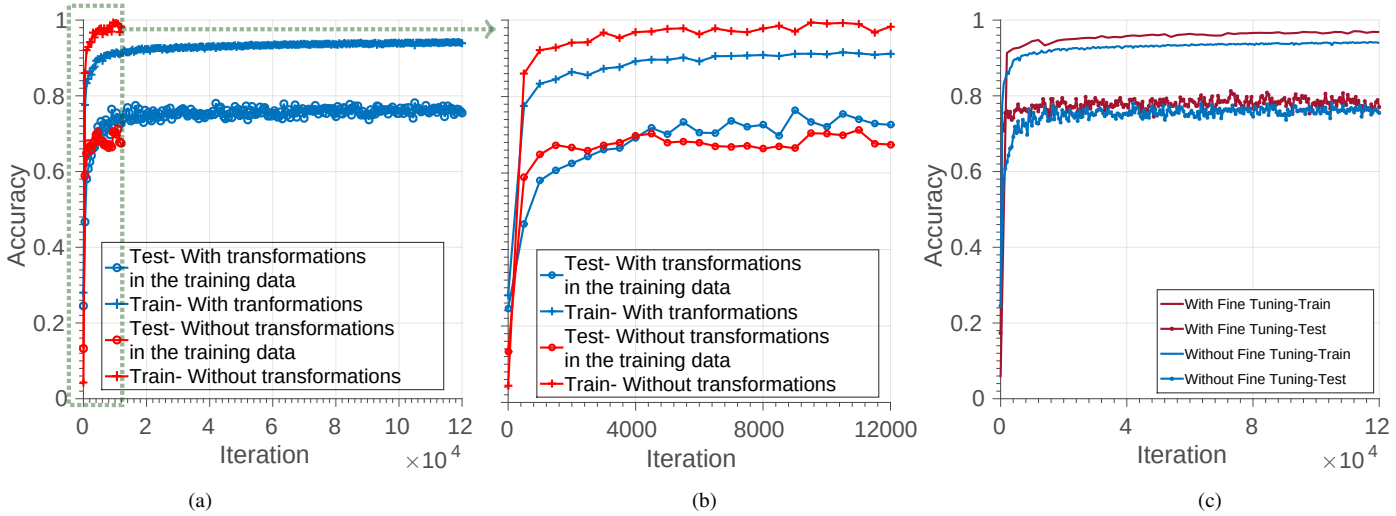
We conduct experiments on the ICPR2012 HEp-2 benchmark to analyze the effects of data pre-processing, augmenta-

TABLE III
COMPARISON OF OUR AVERAGE CLASSIFICATION ACCURACY WITH OTHER METHODS ON THE ICPR2012 DATASET

Method	Classification accuracy
Human expert [2]	**73.3**%
2012 contest winner [2]	**68.7**%
2012 contest CNN [2]	59.8%
Faraki et al. [19]	70.2%
Larsen et al. [5]	71.5%
Shen et al. [20]	74.4%
Gao et al.(CNN) [15]	74.8%
Theodorakopoulos et al. [7]	75.1%
Schaefer et al. [21]	75.3%
Qi et al. [6]	**75.6**%
Ours	**80.3**%

TABLE IV
CONFUSION MATRIX OF OUR CLASSIFICATION METHOD ON ICPR2012 HEP-2 DATASET WITH AVERAGE CLASS ACCURACY OF 80.25%

Pattern	Homoge–nous	Coarse Speckled	Fine Speckled	Nucle–olar	Centro–mere	Cytop–lasmic
Homoge–nous	97.78%	0%	1.11%	1.11%	0%	0%
Coarse Speckled	13.86%	73.27%	0.99%	1.98%	8.91%	0.99%
Fine Speckled	80.70%	1.75%	11.40%	2.63%	3.51%	0%
Nucle–olar	0.72%	0%	0.72%	94.96%	3.60%	0%
Centro–mere	0%	0%	0%	2.68%	97.32%	0%
Cytop–lasmic	0%	1.96%	0%	0%	1.96%	96.08%



(a) (b) (c)

Fig. 6. a) Evaluation showing the average classification accuracy over all types of transformations in the dataset compared against the dataset without transformations, 12×10^4 iterations, b) magnified version of a, c) pre-training and fine tuning.

tion, and pre-training within a CNN based framework. Since it is not easy to design an optimal CNN model in theory we focus on the data itself. We demonstrate how training data affects classification accuracy of cell classification. We find that additional real data-augmentation is incredibly helpful and domain specific pre-training still maintains an advantage and we provide the state-of-the-art performance. Most of our observations confirm recent findings about the benefits of data augmentation in CNN models. But unlike most previous studies, where affine transformations and contrast augmentation were randomly added to an image, we directly add each transformed copy to the training set. On the other hand, we believe that overfitting could still be a problem and additional gains can be achieved if there were more real HEp-2 cell image data available. Source code and models to reproduce the experiments in the paper is made publicly available on the project webpage (`http://www.ee.oulu.fi/~nyalcinb/papers/bibe2015`).

REFERENCES

[1] C. Buchner, C. Bryant, A. Eslami, and G. Lakos, "Anti-nuclear antibody screening using hep-2 cells," *JoVE (Journal of Visualized Experiments)*, no. 88, pp. e51211–e51211, 2014.

[2] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking hep-2 cells classification methods," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 10, pp. 1878–1889, 2013.

[3] P. Foggia, G. Percannella, A. Saggese, and M. Vento, "Pattern recognition in stained hep-2 cells: Where are we now?," *Pattern Recognition*, vol. 47, no. 7, pp. 2305–2314, 2014.

[4] Z. Gao, J. Zhang, L. Zhou, and L. Wang, "Hep-2 cell image classification with convolutional neural networks," in *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*, pp. 24–28, IEEE, 2014.

[5] A. Larsen, J. Vestergaard, and R. Larsen, "Hep-2 cell classification using shape index histograms with donut-shaped spatial pooling," *Medical Imaging, IEEE Transactions on*, vol. 33, pp. 1573–1580, July 2014.

[6] X. Qi, G. Zhao, C.-G. Li, J. Guo, and M. Pietikäinen, "Hep-2 cell classification via fusing texture and shape information," *arXiv preprint arXiv:1502.04658*, 2015.

[7] I. Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos, "Hep-2 cells classification via sparse representation of textural features fused into dissimilarity space," *Pattern Recognition*, vol. 47, no. 7, pp. 2367–2378, 2014.

[8] W. Arnold, W. Yongkang, S. Conrad, H. Peter, C. Shaokang, and L. B. C, "Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors," in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pp. 95–102, IEEE, 2013.

[9] R. Nosaka and K. Fukui, "Hep-2 cell classification using rotation invariant co-occurrence among local binary patterns," *Pattern Recognition*, vol. 47, no. 7, pp. 2428–2436, 2014.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[14] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pp. 411–418, Springer, 2013.

[15] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "Hep-2 cell image classification with deep convolutional neural networks," *arXiv preprint arXiv:1504.02531*, 2015.

[16] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[17] A. Krizhevsky, "cuda-convnet," 2012.

[18] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[19] M. Faraki, M. T. Harandi, A. Wiliem, and B. C. Lovell, "Fisher tensors for classifying human epithelial cells," *Pattern Recognition*, vol. 47, no. 7, pp. 2348–2359, 2014.

[20] L. Shen, J. Lin, S. Wu, and S. Yu, "Hep-2 image classification using intensity order pooling based features and bag of words," *Pattern Recognition*, vol. 47, no. 7, pp. 2419–2427, 2014.

[21] G. Schaefer, N. P. Doshi, and B. Krawczyk, "Hep-2 cell classification using multi-dimensional local binary patterns and ensemble classification," in *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on*, pp. 951–955, IEEE, 2013.