LEARNING LOCAL AND DEEP FEATURES FOR EFFICIENT CELL IMAGE CLASSIFICATION USING RANDOM FORESTS

Zakariya A. Oraibi,† Hayder Yousif,† Adel Hafiane,‡ Guna Seetharaman,∗ Kannappan Palaniappan†

†Department of Electrical Engineering and Computer Science University of Missouri-Columbia, MO 65211 USA
‡INSA Centre Val de Loire, Université d’Orléans PRISME EA 4229, F-18022 Bourges, France
∗Advanced Computing Concepts/ITD, Naval Research Laboratory, USA

ABSTRACT

Automatic image classification systems for indirect immunofluorescence (IIF) labeling of human epithelial (HEp-2) cell specimens are needed to improve the efficient management of autoimmune diseases. In this paper, we propose to classify HEp-2 cell specimen imagery using a combination of local features and deep learning features extracted from the IIF images. Two local descriptors are used to capture texture information, namely: Rotation Invariant Co-occurrence among Local Binary Patterns (RIC-LBP) extending the LBP descriptor and Joint Motif Labels (JML) based on the Peano scan motif concept. Deep learning features are then extracted using the VGG-19 image classification network. Finally, all descriptors are combined using a late fusion approach with a Random Forests (RF) classifier with seven output classes. Experimental results show that our proposed framework achieves a mean class accuracy of 92.11% with five-fold cross validation using the RF classifier with 1000 trees on the HEp-2 specimen benchmark dataset, which outperforms the state-of-the-art accuracy on this dataset.

Index Terms— Feature Extraction, Random Forests, Local Features, Deep Learning, Image Classification

1. INTRODUCTION

HEp-2 cell staining pattern analysis from indirect immunofluorescence (IIF) imaging is important for autoimmune diseases. In recent years, computer aided pattern recognition techniques were introduced to reduce the burden of manual annotation and classification. Due to the international contests conducted in the last few years, great strides have been made towards obtaining efficient automatic pattern recognition and image analysis system for HEp-2 cell and specimen images [1].

The first HEp-2 cells classification contest was held in 2012 in conjunction with International Conference on Pattern Recognition (ICPR). The contest received 22 papers with significant focus on feature extraction techniques. Many image local descriptors were used in this competition, including Local Binary Pattern (LBP), Gray Level Co-occurrence Matrix (GLCM), Adaptive Robust Structure Tensors-Histogram of Oriented Gradients (ARST-HOG), and shape index histograms [2, 3, 4, 5]. Most methods used k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) classifiers. While others, like Malon et. al [1] used neural network classifier.

Two contests were held after that in 2014 and 2016 as part of the ICPR which included two main tasks: cell classification and specimen classification. Methods submitted for cell classification task used local feature extraction, feature encoding, and deep learning techniques. Local descriptors adopted in this task included LBP, motif features, and dense scale-invariant features [6, 7]. Methods with feature encoding adopted various local features like: SURF, SIFT, LBP, and co-occurrence of adjacent LBPs and used bag of visual words and Vectors of Locally Aggregated Descriptors (VLAD) to encode them [8, 9]. For deep learning methods, Jia et al. [10] extracted CNN features from a deeper network architecture and used these features to classify ICPR 2016 HEp-2 cell images with an accuracy of 98.26%. For the specimen level task, both 2014 and 2016 competitions used the seven class specimen cell images of I3A 2014 Task 2 competition. Voting methods and morphological features were used by [8, 11]. Prasath et. al [7] also used RIC-LBP descriptor and achieved an accuracy of 73.43% using RF classifier with 500 trees. Li et al. [12] used a fully convolutional network (FCN) adapted from VGG-16 and achieved a classification accuracy of 90.89%.

In this paper, we introduce a framework to classify specimen cell images of I3A 2014 Task 2 competition dataset using both local and deep learning features. Two local descriptors based on LBP [13] and motif Peano scan concept [14, 15] were combined together along with convolutional neural network features extracted from VGG-19 architecture. The proposed framework achieves high accuracy using Random Forests (RF) classifier.

The rest of this paper is organized as follows: Section 2 details the background approach of RIC-LBP and JML descriptors. Section 3 details the CNN model used in our work.
Network training and feature extraction is introduced in Section 4. Section 5 demonstrates the experiments conducted on specimen cell dataset. Finally, section 6 concludes the paper.

2. LOCAL FEATURES

In this section, we illustrate the local descriptors used in our work. Two local descriptors are used to extract texture features from the specimen images. These descriptors are:

RIC-LBP descriptor: In [4], a new texture descriptor called Rotation Invariant Co-occurrence among LBP (RIC-LBP) was proposed and used successfully to classify HEp-2 cell images. RIC-LBP makes use of the relationships among the binary patterns by finding the co-occurrences patterns among the histogram features.

As a result, RIC-LBP histogram will be represented in the form of many LBP pairs and each pair will be attached with a specific label to account for rotation invariance. The reason to choose this descriptor is because it achieved higher accuracy than other LBP based descriptors like Joint Adaptive Median Binary Patterns (JAMBP) [16]. The size of RIC-LBP is 408 bins.

JML descriptor: Joint Motif Labels (JML) descriptor was introduced by us in 2016 [7]. JML is based on the motif concept where 12 motif patterns are extracted from each 2 × 2 grid and labeled from 1 - 12. After that, three moments are found from these 12 patterns: minimum, median, and maximum which are the difference of two adjacent pixels in the motif pattern of the grid. Two bins mean and variance matrices are also found as extra information extracted from the original image. Finally, JML descriptor histogram is computed using 3D joint histogram of the three matrices: motif labels, mean, and variance. Translational invariance was achieved by computing the descriptor on four images: original image, image rotated horizontally, vertically, and diagonally by one pixel. The size of the descriptor is 576 bins [17]. RIC-LBP and JML descriptors are illustrated in Figure 1.

3. NETWORK ARCHITECTURE

In 2014, Simonyan et al. [18] proposed very deep convolutional neural networks called VGG(16-19) that extend AlexNet [19] network by adding more convolutional layers. The VGG model is applied to RGB image with a small
In this paper, we depicted the deeper VGG-19 network and extracted the features of the Fully Connected ("fc7") layer to perform classification using RF classifier.

4. PROPOSED APPROACH

4.1. Data Transformation and Augmentation

Since CNN operates on fixed size square images, the 1388 × 1040 pixels original images are extended by zero-padding to have 1388 × 1388 size before rescaling into the required CNN size (224 × 224 for VGG-19). This enabling step overcomes losing shape features which differentiate a class from another. The training HEP2 specimen images provided by I3A contest contains only 1008 images distributed over 7 classes. Images vary in each class widely from 40 to 212. We enrich the training set with augmented images to equalize the number of images in each class which helps in reducing training overfitting. The training set with augmented images to equalize the number of images in each class which helps in reducing training overfitting. Figure 3 shows the effect of the proposed data augmentation. The results are 1298 samples were used in the training phase. After applying this augmentation process, classes will contain images that vary between 172 to 195 as shown in Table 1 which provides enough samples to train the CNN model.

4.2. Fusing CNN and Local Features

We trained our VGG-19 network using the Caffe toolbox [20] on a single GPU. The minimum batch size used in training was 40 samples. We fixed the learning rate to 0.001 and the total number of epochs was also fixed to 10. After obtaining our model, we send our test images to the network for classification. The result of network classification for all test images is 78.85%. In order to improve this result, we propose two fusion methods:

**Early Fusion.** The Early Fusion combines information across the pixel level and the 12 motif patterns. This 13-layer image is then projected into Principle Component Analysis (PCA) and the first three principal components are used as CNN input. PCA eliminates the spatial redundancy of the multi-layer feature maps. Figure 4 shows the basic steps of the proposed early fusion.

**Late Fusion.** The Late Fusion combines the handcrafted and CNN features for better image representation. Specifically, the trained CNN model is used to extract the automated features and then we combine these features with the handcrafted features. Finally, we train a classifier on the combined features. We use the 4096 Fully Connected (FC) layer features of the pretrained VGG-19 model with different combination of local descriptors. The combined features could have

![Fig. 2. Sample specimen image from each class with a zoomed view showing details.](image)

![Fig. 3. Augmentation using image transformation; (a) original image; (b) resized image; (c) 90° rotated; (d) 180° rotated.](image)

![Classes](image)

<table>
<thead>
<tr>
<th>Classes</th>
<th>Hom</th>
<th>Spe</th>
<th>Nuc</th>
<th>Cen</th>
<th>Num</th>
<th>Gol</th>
<th>Mit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>170</td>
<td>167</td>
<td>160</td>
<td>163</td>
<td>67</td>
<td>32</td>
<td>48</td>
</tr>
<tr>
<td>Augmented</td>
<td>195</td>
<td>194</td>
<td>193</td>
<td>193</td>
<td>177</td>
<td>172</td>
<td>174</td>
</tr>
</tbody>
</table>

![Fig. 4. Early fusion approach for fusing information over spatial and feature domains. The raw input image and 12 Motif Pattern (MP) images are projected into PCA. The first three principal components are then fed into CNN.](image)
Table 2. Confusion matrix for specimen cell classification using RIC-LBP, JML, and CNN features using RF classifier. The Mean Class Accuracy (MCA) is 92.11%.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Homogeneous</th>
<th>Speckled</th>
<th>Nucleolar</th>
<th>Centromere</th>
<th>Numem.</th>
<th>Golgi</th>
<th>Mitsp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogeneous</td>
<td>91.08</td>
<td>6.12</td>
<td>0</td>
<td>0.47</td>
<td>2.33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Speckled</td>
<td>7.17</td>
<td>89.5</td>
<td>0.47</td>
<td>2.86</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nucleolar</td>
<td>0</td>
<td>0.50</td>
<td>99.0</td>
<td>0.50</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Centromere</td>
<td>0</td>
<td>4.40</td>
<td>1.47</td>
<td>94.13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Numem.</td>
<td>7.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>89.41</td>
<td>0</td>
<td>3.53</td>
</tr>
<tr>
<td>Golgi</td>
<td>5.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95.0</td>
<td>0</td>
</tr>
<tr>
<td>Mitsp.</td>
<td>10.0</td>
<td>1.66</td>
<td>0</td>
<td>0</td>
<td>1.67</td>
<td>0</td>
<td>86.67</td>
</tr>
</tbody>
</table>

Table 4. Comparison with state-of-the-art methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>MCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu et al. cited in [21]</td>
<td>86.10</td>
</tr>
<tr>
<td>Gragnanenelio et al. [6]</td>
<td>86.77</td>
</tr>
<tr>
<td>Mannivannan et al. [21]</td>
<td>89.93</td>
</tr>
<tr>
<td>Li et al. [12]</td>
<td>90.89</td>
</tr>
<tr>
<td>Ours (early fusion)</td>
<td>81.90</td>
</tr>
<tr>
<td>Ours (late fusion)</td>
<td>92.11</td>
</tr>
</tbody>
</table>

Table 3. Results comparing late fusion of deep and local features with our other approaches using RF classifier.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Size</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG-19</td>
<td>4096</td>
<td>90.81</td>
</tr>
<tr>
<td>RIC-LBP</td>
<td>408</td>
<td>70.14</td>
</tr>
<tr>
<td>JML</td>
<td>576</td>
<td>66.35</td>
</tr>
<tr>
<td>VGG-19+RIC-LBP</td>
<td>4504</td>
<td>91.32</td>
</tr>
<tr>
<td>VGG-19+RIC-LBP+JML</td>
<td>5080</td>
<td>92.11</td>
</tr>
</tbody>
</table>

redundancy or noise. We propose to utilize RF as a classifier to sample the best 1000 bins that can efficiently represent the input image.

5. EXPERIMENTAL RESULTS

In order to test our framework, we used the I3A Task 2 specimen dataset. This dataset consists of 1008 images from 252 specimens which were captured at four different locations. The size of each image is $1388 \times 1040$ pixels along with a cell mask for every image. Figure 2 contains image samples from each of the seven classes. The evaluation metric used is the Mean Class Accuracy (MCA) which is defined as:

$$MCA = \frac{1}{K} \sum_{k=1}^{K} CCR_k$$ (1)

where $K$ is the number of classes and $CCR_k$ is the correct classification rate for each class.

Table 3 shows the results of applying our framework with 1000 Random Forests (RF) classifier after carrying out five-fold cross validation experiments. First, deep learning features (4096 bins) generated a very good result of 90.81% MCA. This result was further improved by 1.3% after combining both local descriptors: RIC-LBP and JML with deep learning features. Table 2 contains the confusion matrix of the best method.

Table 4 compares our approach and the previously state-of-the-art techniques. Mannivannan et al. [21] extracted a combination of Root-SIFT features and multi-resolution LBPs from HEp-2 image cells with ensembles of SVMs for the classification phase. They achieved high accuracy of (89.93%) and were the winners of I3A 2014 competition. Li et al. [12] employed a fully convolutional network and used the VGG-16 softmax layer to achieve an accuracy of 90.89%. Our framework of combining local and deep learning features slightly outperforms these state-of-the-art methods.

6. CONCLUSION

In this paper, we address the problem of automatic classification of HEp-2 specimen cells. We showed that the features extracted from a VGG-19 CNN model achieve superior accuracy using RF classifier. Moreover, we improved the results of deep learning by combining CNN features with robust texture features called RIC-LBP and JML. The framework of CNN, RIC-LBP, and JML features achieved a mean class accuracy of 92.11% using five-fold cross validation and outperformed state-of-the-art methods on I3A Task 2 dataset.

Acknowledgments

High performance computing infrastructure was partially supported by the U.S. National Science Foundation (award CNS-1429294). ZAO and HY were supported by HCED Government of Iraq doctoral scholarships. This work was partially supported by a U.S National Institute of Health NIBIB award R33 EB00573 (KP).
7. REFERENCES


