Cell Images Classification using Deep Convolutional Autoencoder

Caleb Vununu and Ki-Ryong Kwon
Dept. of IT Convergence and Application Engineering
Pukyong National University
Busan, Republic of Korea
exen.xmen@gmail.com

Suk-Hwan Lee
Dept. of Information Security
Tongmyong University
Busan, Republic of Korea
skylee@tu.ac.kr,
kiryongkwon@gmail.com

Eung-Joo Lee
Dept. of Information & Communication Engineering
Tongmyong University
Busan, Republic of Korea
ejlee@tu.ac.kr

ABSTRACT

The present work proposes a classification method for the Human Epithelial of type 2 (HEp-2) cell images using an unsupervised deep feature learning method. Unlike most of the state-of-the-art methods in the literature that utilize deep learning in a strictly supervised way, we propose here the use of the deep convolutional autoencoder (DCAE) as the principal feature extractor for classifying the different types of the HEp-2 cellular images. The network takes the original cellular images as the inputs and learns how to reconstruct them through an encoding-decoding process in order to capture the features related to the global shape of the cells. A final feature vector is constructed by using the latent representations extracted from the DCAE, giving a highly discriminative feature representation. The created features will then be fed to a nonlinear classifier whose output will represent the final type of the cell image. We have tested the discriminability of the proposed features on one of the most popular HEp-2 cell classification datasets, the SNPHEp-2 dataset and the results show that the proposed features manage to capture the distinctive characteristics of the different cell types while performing at least as well as some of the actual deep learning based state-of-the-art methods.

KEYWORDS

Bioimage Analysis and Processing, Computer-aided Diagnosis, HEp-2 Cell Images Classification, Convolutional Autoencoders, Deep Learning

1 Introduction

Computer-aided diagnostic (CAD) systems have gained tremendous interests since the unfolding of various machine learning techniques in the past decades. They comprise all the systems that aim to consolidate the automation of the disease diagnostic procedures. One of the most challenging tasks regarding those CAD systems is the complete analysis and understanding of the images representing the biological organisms. In case of the autoimmune diseases, the automatic classification of the different types of the Human Epithelial type 2 (HEp-2) cell patterns is one of the most important steps of the diagnosis procedure.

Automatic feature learning methods have been widely adopted since the unfolding of deep learning [1]. They have shown outstanding results in the object recognition problems [2] and many researchers have adopted them as principal tool for the HEp-2 cell classification problem. Unlike conventional methods whose accuracy depends on the subjective choice of the features, deep learning methods, such as deep convolutional neural networks (CNNs), have the advantage of offering an automatic feature learning process. In fact, many works have demonstrated the superiority of the deep learning based features over the hand-crafted ones for the HEp-2 cell classification task.

The first work to apply CNN to the HEp-2 cell classification problem was presented during the 2012 edition of the ICPR HEp-2 cell classification contest. Although the results were outstanding, the datasets available on that time were not heterogeneous enough and needed a lot of improvements. Since then, many available datasets have been significantly diversified and the different proposed CNN models continue to push the limits in terms of classification accuracy. Gao et al. [3] have presented a simple CNN architecture that was tested over different datasets. They were the first to test the data augmentation techniques, such as rotation in different angles, for the HEp-2 cell images.

Li et al. [4] have adopted the deep residual inception model, the DRI, which combines two of the most popular CNN models, the ResNet [5] architecture and the “Inception” modules from the GoogleNet [6]. Phan et al. [7] have performed transfer learning, which consists of using an already trained network in a new dataset, by using a model that was trained on the ImageNet dataset. Note that all these methods prefer to address the HEp-2 cell classification problem in a strictly supervised way, where the feature extraction and classification processes are forced to belong to the same module.

Although the performance obtained with the supervised learning methodology continues to reach impressive levels, the exigency of always having labelled datasets in hand, knowing that deep-learning methods necessitate huge amount of images, can represent a relative drawback for these methods.

We propose an unsupervised deep feature learning process that uses the deep convolutional autoencoder (DCAE) as the principal feature extractor. The DCAE, which learns to reproduce the original cellular images *via* a deep encoding-decoding scheme, is used for extracting the features. The DCAE takes the original cell

image as an input and will learn to reproduce it by extracting the meaningful features needed for the discrimination part of the method. The latent representations trapped between the encoder and the decoder of the DCAE will be extracted and used as the final high-level features of the system.

The DCAE will help to encode the geometrical details of the cells contained in the original pictures. The discrimination potentiality carried by the extracted features allows us to feed them as the inputs of a shallow nonlinear classifier, which will certainly find a way to discriminate them. The proposed method was tested on the SNP HEp-2 Cell dataset [8] and the results show that the proposed features outperform by far the conventional and popular handcrafted features and perform at least as well as the state-of-the-art supervised deep learning based methods.

2 Proposed Method

Auto-encoders [9] are unsupervised learning methods that are used for the purpose of feature extraction and dimensionality reduction of data. Neural network based auto-encoder consists of an encoder and a decoder. The encoder takes an input x of dimension d, and maps it to a hidden representation y, of dimension r, using a deterministic mapping function f such that:

$$y = f(\boldsymbol{W}x + \boldsymbol{b}), \quad (1)$$

where the parameters $\boldsymbol{W}$ and $\boldsymbol{b}$ are the weights and bias associated with the layer that takes the input x. They must be learned by the encoder. The decoder then takes the output y of the encoder and uses the same mapping function f in order to provide a reconstruction z that must be of the same shape or in the same form (which means almost equal to) as the original input signal x. Using equation (1), the output of the decoder is also given by:

$$z = f(\boldsymbol{W}'y + \boldsymbol{b}'), \quad (2)$$

where the parameters **W'** and **b'** are the weights and bias associated with the decoder layer. In final, the network must learn the parameters **W**, **W'**, **b** and **b'** so that z must be close or, if possible, equal to x. In final, the network leans to minimize the differences between the encoder's input x and the decoder's output z.

This encoding-decoding process can be done with the use of convolutional neural networks by using what we call the deep convolutional autoencoder (DCAE). Unlike conventional neural networks, where you can set the size of the output that you want to get, the convolutional neural networks are characterized by the process of down-sampling, accomplished by the pooling layers, which are incorporated in their architecture. And this sub-sampling process has as consequence the loss of the input's spatial information while we go deeper inside the network.

To tackle this problem, we can use DCAE instead of conventional convolutional neural networks. In the DCAE, after the down-sampling process accomplished by the encoder, the decoder tries to up-sample the representation until we reconstruct the original size. This can be made by backwards convolution often called "deconvolution" operations. The final solution of the network can be written in the form:

$$(\mathbf{W}, \mathbf{W}', \mathbf{b}, \mathbf{b}') = \underset{W,W',b,b'}{\operatorname{argmin}} L(xz), \quad (3)$$

where z denotes the decoder's output and x is the original image. The function L in equation (3) estimates the difference between the x and z. So, the solution of equation (3) represents the values that minimize the most the difference between x and z.

The architecture of our DCAE is shown in Table 1 and we can remark how the network applies the encoding and decoding processes.

TABLE I. THE PROPOSED DCAE ARCHITECTURE

Layer	Filter size	#Feature maps	Stride	Padding	Output
Input	-	-	-	-	**224×224**
Conv 1	3×3	64	1	1	224×224
Pool 1	2×2	64	2	0	112×112
Conv 2	3×3	128	1	1	112×112
Pool 2	2×2	128	2	0	56×56
Conv 3	3×3	256	1	1	56×56
Pool 3	2×2	256	2	0	28×28
Conv 4	3×3	512	1	1	28×28
Pool 4	2×2	512	2	2	14×14
Conv 5	14×14	4096	1	1	**1×1**
Deconv5	14×14	512	1	1	14×14
Unpool 4	2×2	512	2	2	28×28
Deconv 4	3×3	512	1	1	28×28
Unpool 3	2×2	256	2	0	56×56
Deconv 3	3×3	256	1	1	56×56
Unpool 2	2×2	128	2	0	112×112
Deconv 2	3×3	128	1	1	112×112
Unpool 1	2×2	64	2	0	224×224
Deconv 1	3×3	1	1	1	**224×224**

In Table 1, we can see a process of down-sampling (encoding) with the stacking of many convolutional and pooling layers. Just after we reach a 1x1x4096 feature volume, we start the up-sampling process (decoding) with the stacking of many deconvolutional and unpooling layers. The convolutional layers are denoted by "Conv" in the table, the pooling layers are mentioned by "Pool", the deconvolutional layers are represented by "Deconv" and the unpooling layers are depicted as "Unpool". After the encoding-decoding process, the activations contained in the output of the "Conv 5" layer, which are in form of a one-dimensional vector containing 4096 values, will be the latent representations that are we are seeking for. These values will be extracted and will represent the final feature representation of our original images. We will use these values as the inputs of a shallow artificial neural network in order to test their discrimination potential.

Artificial neural networks are strong linear classifiers capable of discovering the hidden characteristics of the data. Knowing that they can simulate any nonlinear function, we have adopted them as our nonlinear discriminator. The feature vectors extracted from the DCAE, which contain 4096 elements, will be given to the network and a supervised learning process will be conducted in order to predict the cell type.

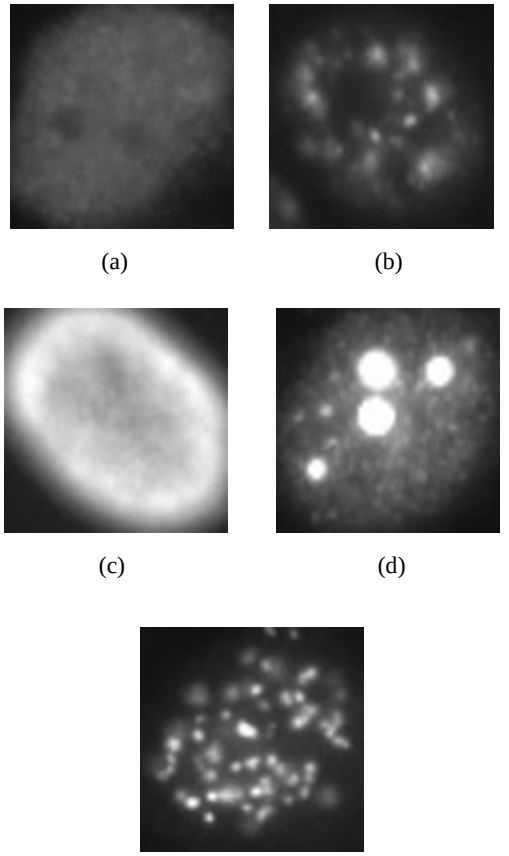

Figure 1: Examples of the five different HEp-2 cell images from the dataset used during the experiments: (a) homogeneous, (b) coarse speckled, (c) fine speckled, (d) nucleolar, and (e) centromere.

3 Results and Discussion

There are 1,884 cellular images in the dataset, all of them extracted from the 40 different specimen images. Different specimens were used for constructing the training and testing image sets, and both sets were created in such a way that they cannot contain images from the same specimen. From the 40 specimen, 20 were used for the training sets and the remaining 20 were used for the testing sets. In total there are 905 and 979 cell images for the training and testing sets, respectively. Each set (training and testing) contains five-fold validation splits of randomly selected images. In each set, the different splits are used for cross validating the different models, each split containing 450 images approximatively. The SNPHEp-2 dataset was presented by Wiliem et al. [8]. Figure 1 shows the example images of the five different cell types randomly selected from the dataset.

As presented before, the created feature vectors extracted from the DCAE contain 4096 elements. So, our network will have 4096 neurons in the input layer. The best results were obtained using a 4096-250-50-5 architecture, meaning that we have 4096 neurons in the input layer, 250 neurons in the first hidden layer, 50 neurons in the second hidden layer and a final layer containing 5 neurons corresponding to the 5 cell types of our dataset. The total accuracy reached by the network was 88.08 %. The details of the results are shown in the confusion matrix depicted in Figure 2. In the figure, 'Homo", "Coarse", "Fine", "Nucl" and "Centro" denote the homogeneous, the coarse speckled, the fine speckled, the nucleolar and the centromere cell types, respectively.

Output Class \ Target Class	Homo	Coarse	Fine	Nucl	Centro
Homo	**91.07**	0.53	9.97	0	0
Coarse	0.46	**88.27**	0	5.77	6.54
Fine	6.11	2.24	**86.42**	0	0.10
Nucl	1.38	3.79	0.94	**85.09**	0.17
Centro	0.98	5.17	2.67	9.14	**93.19**

Figure 2: Confusion matrix of the results obtained with a 4096-250-50-5 neural network using the extracted features from the DCAE as the inputs.

In the confusion matrix in Figure 2, we can see that the most distinguishable cells for the classifier are the centromere cells, for which the classification accuracy reaches 93.19 %. But, in the same time, we can notice that there is a significant confusion between the centromere and the coarse speckled cells: 6.54 % of the coarse speckled cells were misclassified as centromere. This is mainly explained by the similarity in shape and intensity level between the two types of cells. We can remark that in Figure 1, the coarse speckled and centromere cells, respectively shown in Figure 1 (b) and (e), have a lot of similarity in terms of shape and intensity. The confusion is confirmed by also taking a look at the classification rate of the coarse speckled: 5.17 % of them were misclassified as centromere, as we can see in the second column (fifth row) of the confusion matrix.

The homogeneous cells also are well classified in general, over 91 % of them were correctly recognized by the classifier. Another important confusion comes between the homogeneous and fine speckled cells. As we can notice in the confusion matrix, 6.11 % of the homogeneous cells were misclassified as fine speckled. Here also we can remark a strong similarity between the two types of cells in their shape. The two cells are the ones that exhibit a clear circular form (ellipsoidal, precisely). They have less similarity in their intensity level, explaining why the confusion in this case is slightly less significant than the confusion mentioned in the previous paragraph. And in the case of the fine speckled cells, the confusion is even more noticeable. We can see in the matrix that almost 10 % (9.97) of the fine speckled were misclassified as homogeneous. Trying to decrease the confusion between the cells

that show strong similarities in terms of shape and intensity level can be the direction of any consideration about the future works.

As mentioned before, the overall classification rate of the proposed method is 88.08 %.

TABLE II. COMPARATIVE RESULTS

Method	Accuracy
Texture features + SVM	80.90%
LPB descriptors + SVM	85.71%
5 layers CNN	86.20%
Present work (DCAE features + ANN)	**88.08%**

We have conducted a comparative study with the handcrafted features and one deep learning method using the CNN in a strictly supervised manner for the classification of the cellular images. The results of the comparative study are shown in Table 2. We can clearly see that the proposed method outperforms the handcrafted features. The proposed features from the DCAE perform also slightly better than the supervised deep-learning method proposed in [3] using a 5 layers' network.

4 Conclusion

We have presented a cell classification method for the images portraying the microscopy data, the HEp-2 cells, a method that has adopted the DCAE as the principal feature extractor. Unlike most of the methods in the literature that are based on the supervised learning, we have used the DCAE in order to construct the feature vectors in an unsupervised way. These obtained vectors were then given to a nonlinear classifier whose outputs determine the cell type of the image. The results show that the proposed feature extraction method really captures the characteristics of each cell type. The comparative study demonstrates that our proposed features perform far better than the handcrafted ones and slightly better than the supervised deep learning method.

But, as we have discussed in the results, many cell types exhibit strong similarities between them in terms of shape and intensity level. These similarities encourage a significant confusion during the discrimination step of the proposed features. We consider that the next step of our work is to try to find a way of minimizing the confusion between these cells that show strong similarities.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. 2016R1D1A3B03931003, No. 2017R1A2B2012456), and the MSIT (Ministry of Science and ICT), Korea, under the ICT Consilience Creative program (IITP-2019-2016-0-00318) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and the Ministry of Trade, Industry and Energy for its financial support of the project titled "the establishment of advanced marine industry open laboratory and development of realistic convergence content."

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hintton, "Deep Learning," Nature, Vol. 521, pp. 436-444.

[2] Y. LeCun, F.J. Huang, and L. Bottou, "Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting," In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), Washington, DC, USA, 27 June-2 July 2004.

[3] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "Hep-2 Cell Image Classification with Deep Convolutional Neural Networks," IEEE J. Biomed. Health Inf., Vol. 21, pp. 416-428.

[4] Y. Li, and L. Shen, "A Deep Residual Inception Network for HEp-2 Cell Classification," In Proceedings of Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Québec City, QC, Canada, 14 September 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 27-30 June 2016; pp. 770–778.

[6] C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, "Going Deeper with Convolutions," In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, USA, 7-12 June 2015; pp. 1–9.

[7] H.T.H. Phan, A. Kumar, J. Kim, and D. Feng, "Transfer Learning of a Convolutional Neural Network for HEp-2 Cell Image Classification," In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 16 June 2016; pp. 1208–1211.

[8] A. Wiliem, Y. Wong, C. Sanderson, P. Hobson, S. Chen, and B.C. Lovell, "Classification of Human Epithelial Type 2 Cell Indirect Immunofluorescence Images via Codebook based Descriptors," In 2013 IEEE Workshop on Applications of Computer Vision (WACV), Tampa, Florida, USA, 15-17 January 2013; pp. 95–102.

[9] G.E. Hinton and R.R. Salakhutdinov, "Reducing the Dimensionality of the Data with Neural Networks," *Science*, Vol. 313, Issue 5786, pp. 504-507, 2006.