A Deep Residual Inception Network for HEp-2 Cell Classification

Yuexiang Li and Linlin Shen(✉)

Computer Vision Institute, Shenzhen University, Shenzhen, China
llshen@szu.edu.cn

Abstract. Indirect-immunofluorescence (IIF) of Human Epithelial-2 (HEp-2) cells is a commonly-used method for the diagnosis of autoimmune diseases. Traditional approach relies on specialists to observe HEp-2 slides via the fluorescence microscope, which suffers from a number of shortcomings like being subjective and labor intensive. In this paper, we proposed a hybrid deep learning network combining the latest high-performance network architectures, i.e. ResNet and Inception, to automatically classify HEp-2 cell images. The proposed Deep Residual Inception (DRI) net replaces the plain convolutional layers in Inception with residual modules for better network optimization and fuses the features extracted from shallow, medium and deep layers for performance improvement. The proposed model is evaluated on publicly available I3A (Indirect Immunofluorescence Image Analysis) dataset. The experiment results demonstrate that our proposed DRI remarkably outperforms the benchmarking approaches.

Keywords: HEp-2 cells · Image classification · Deep Learning Network

1 Introduction

Detecting antinuclear antibodies (ANAs) by Human Epithelial-2 (HEp-2) cell patterns is a recommended approach for autoimmune disease diagnosis. However, in current clinical practice, inspection of the indirect-immunofluorescence (IIF) slides requires highly-skilled pathologists, resulting in a time-consuming analysis subject to inter-observer variations. To address the problem, computer-aided diagnostic (CAD) systems are proposed to assist pathologists. As one of the most challenging tasks, HEp-2 staining pattern recognition has received increasing attention from the research community. Recent contests [1–3] of HEp-2 cell image processing further promoted the developments of CAD systems for automatic classification of HEp-2 cells.

The performances of previous works on HEp-2 cell pattern classification highly rely on the choices of hand-crafted features. Nosaka et al. [4] proposed a rotation invariant Local Binary Pattern descriptor (CoALBP) for HEp-2 cell classification, which won the first prize of ICPR 2012 contest. In ICIP 2013 contest, the framework combining LBP and bag of words proposed by Shen et al. [5] achieved the highest accuracy, i.e. 83.65%. Manivannan et al. [6], the winner of ICPR 2014 contest, extracted Root-SIFT features and multi-resolution local patterns from HEp-2 cell

M.J. Cardoso et al. (Eds.): DLMIA/ML-CDS 2017, LNCS 10553, pp. 12–20, 2017.
DOI: 10.1007/978-3-319-67558-9_2

images and employed ensembles of SVMs for classification. In more recent studies, an increasing number of methods using hand-crafted features were proposed to automatically classify HEp-2 cell images. For examples, Xu et al. [7] trained a linear SVM with a Co-occurrence Differential Texton (CoDT) feature to identify the staining patterns of HEp-2 cells. Taalimi et al. [8] employed joint sparse representation of HEp-2 cell images for classification. Kastaniotis et al. [9] developed a method, named VHAR, hierarchically aggregating the residual of the sparse representation of feature vectors derived from SIFT descriptor for HEp-2 images classification.

Different from traditional approaches, convolutional neural networks (CNN) do not require the design of hand-crafted features. They automatically construct feature representations from input images through multi-layer processing. The feature learned from CNN is verified to provide better classification performance than hand-crafted features [10]. Researchers have made their efforts to implement CNN for HEp-2 classification. Gao et al. [11] presented the first work using CNN to classify HEp-2 cell images. Their model is a VGG-based network [12], which consists of 8 layers. Bayramoglu et al. reported their new progress on using CNN for HEp-2 classification in [13]. Phan et al. [14] finetuned a model pre-trained on ImageNet to HEp-2 dataset. The published CNN frameworks for HEp-2 classification share the same characteristics: the network architectures are straightforward with a single softmax classifier. As the deep learning model developed, multi-branches networks, e.g. ResNet [15], Inception [16, 17], have gradually surpassed straightforward networks and become the main-stream model for researchers.

In this paper, we proposed a hybrid multi-branches model, Deep Residual Inception (DRI), instead of straightforward CNN to improve the accuracy of classifying HEp-2 cell images. The proposed model combines the architectures of two advanced high-performance deep learning networks, i.e. ResNet and Inception, by replacing the plain convolutional layers in Inception with residual modules. Therefore, the proposed DRI can exploit the strengths of both ResNet and Inception for network training and image classification, i.e. easy network optimization from ResNet and multi-scale feature extractors from Inception. Furthermore, to better utilize the features learned by DRI, we proposed a novel scheme to fuse the features extracted by the shallow, medium and deep layers. Performance evaluation has been conducted on I3A dataset. Experiment results demonstrate that our DRI can provide outstanding HEp-2 cell classification performance.

2 Deep Residual Inception

2.1 Network Architecture

The architecture of proposed Deep Residual Inception network is shown in Fig. 1. The input cell images are reshaped to a size of 76 × 76. The parameters of 1/2 and 1/3 represent the stride size of 2 and 3, respectively. The green layers are convolutional layers. The numbers represent the size and amount of convolutional kernels. Our DRI uses two kinds of pooling, i.e. max pooling and average pooling. The numbers in pooling layers represent the kernel size. Layers named with ‘FC’ are fully-connected

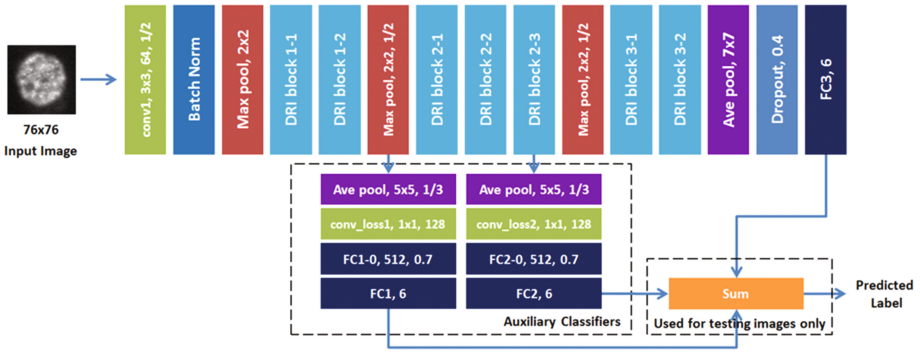

Fig. 1. Network architecture of Deep Residual Inception. (Color figure online)

layers. The second parameter of fully-connected layers is the number of containing neurons. The third parameter of FC1-0 and FC2-0 indicates that the layers are followed by a dropout layer [18] of 0.7. The non-linear activation function used in DRI is PReLU [19]. Two auxiliary classifiers, i.e. FC1 and FC2, and one primary classifier, i.e. FC3, are involved in the proposed DRI. The three classifier branches are trained with separate softmax loss (L), as defined in (1), respectively.

$$L = \frac{1}{N}\sum\nolimits_i L_i = \frac{1}{N}\sum\nolimits_i -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \tag{1}$$

where f_j denotes the j-th element ($j \in [1, K]$, K is the number of classes) of vector of class scores $\boldsymbol{f}$, y_i is the label of i-th input feature and N is the number of training data.

Different weights (w) are assigned to the softmax loss of different classifiers. Assume L_n represents the softmax loss of classifier FCn, the total joint loss (L_{total}) for DRI can be defined as:

$$L_{total} = \sum\nolimits_{n=1}^{3} w_n \times L_n \tag{2}$$

where w = 0.3, 0.3 and 1 for L_1, L_2 and L_3, respectively.

According to the calculated total loss, the optimization of DRI is performed using the stochastic gradient descent (SGD) with back-propagation [20].

The original Inception net uses auxiliary classifiers to improve the convergence of very deep network during training. In our experiments, we found the uses of auxiliary classifiers not only assist network convergence, but also have positive effect to the testing classification performance. Thus, the proposed DRI integrates the features from shallow (FC1), medium (FC2) and deep (FC3) classifier branches with a sum layer and accordingly predicts the staining cell pattern for testing HEp-2 image.

2.2 DRI Module

The DRI has 7 multi-branches modules, named DRI block, whose architecture is presented in Fig. 2. Branches with different sizes of convolutional kernels, i.e. 1×1, 3×3, 5×5, enable the network to extract multi-scale features. Inspired by [21], the Batch

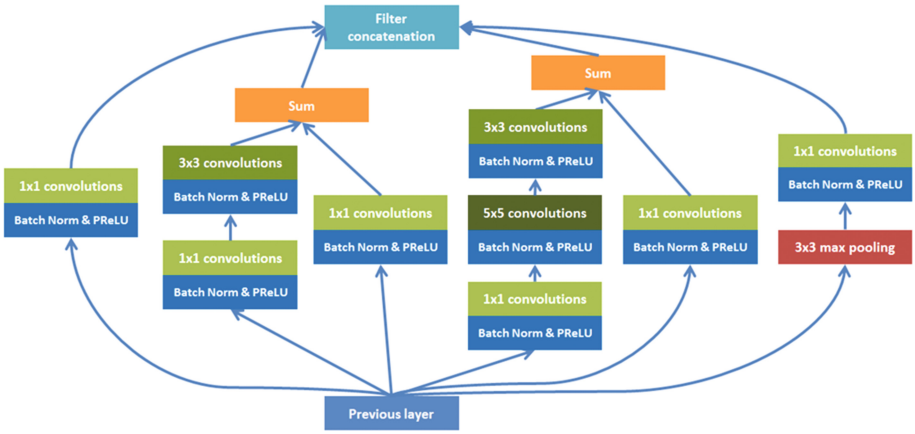

Fig. 2. Architecture of DRI module. Two primary changes of module structure: (1) Two identity shortcuts are added to 3×3 and 5×5 branches. (2) The depth of 5×5 branch is increased.

Normalization [22] and PReLU are placed in front of convolutional layers. Compared to the original Inception module [16], two identity shortcut connections [15] are added to 3×3 and 5×5 branches to alleviate the gradient vanishing problem as network goes deeper. Another difference between the proposed DRI module and the original inception module is that we increased the depth of 5×5 branch by adding a 3×3 convolutional layer to the end of 5×5 convolutions. This change of network structure is inspired by the recent research [23], which illustrates the different-depth branches can benefit the overall performance of multi-branches model.

A detailed configuration of DRI is shown in Table 1. The first number of each entry in Table 1 is the kernel size and the second number is the amount of kernels. Similar to original Inception, the amount of convolutional kernels in our DRI smoothly increases for extracting features with better representation capacity.

2.3 Network Training

The proposed DRI is implemented utilizing Caffe toolbox [24]. The network is initialized with the 'xavier' scheme and trained with a mini-batch size of 24 on one GPU (GeForce GTX TITANX, 12 GB RAM). A momentum of 0.9 is used to assist network optimization. The initial learning rate is set to 0.001 and decayed with a gamma of 0.79. The network converges after 290,000 iterations.

Table 1. The configuration of the proposed DRI architecture

	DRI 1-1	*DRI 1-2*	*DRI 2-1*	*DRI 2-2*	*DRI 2-3*	*DRI 3-1*	*DRI 3-2*
1 × 1 branch	1 × 1, 64	1 × 1, 128	1 × 1, 192	1 × 1, 160	1 × 1, 244	1 × 1, 256	1 × 1, 384
3 × 3 branch	1 × 1, 96	1 × 1, 128	1 × 1, 112	1 × 1, 128	1 × 1, 144	1 × 1, 160	1 × 1, 192
	3 × 3, 128	3 × 3, 192	3 × 3, 224	3 × 3, 256	3 × 3, 288	3 × 3, 320	3 × 3, 384
5 × 5 branch	1 × 1, 16	1 × 1, 32	1 × 1, 16	1 × 1, 24	1 × 1, 24	1 × 1, 32	1 × 1, 48
	5 × 5, 32	5 × 5, 96	5 × 5, 48	5 × 5, 64	5 × 5, 64	5 × 5, 128	5 × 5, 128
	3 × 3, 64	3 × 3, 128	3 × 3, 96	3 × 3, 128	3 × 3, 128	3 × 3, 256	3 × 3, 256
3 × 3 max	1 × 1, 32	1 × 1, 64	1 × 1, 64	1 × 1, 64	1 × 1, 64	1 × 1, 128	1 × 1, 128

3 Results

3.1 Dataset

We use I3A (Indirect Immunofluorescence Image Analysis) dataset for testing in this paper. The **I3A dataset**[1] was first released on the HEp-2 cell classification competition hosted by ICIP 2013, and used again in the contest organized by ICPR 2014. Participants trained their algorithms on the publicly available I3A training set and submit them to the contest organizers for testing. As the I3A test set is not publicly available, we partitioned the I3A training set based on the protocol stated in [11] to evaluate the performance of proposed DRI. The dataset contains 13,596 cell images which can be separated to six categories: Homogeneous (**H**), Speckled (**S**), Nucleolar (**N**), Centromere (**C**), Nuclear membrane (**Nm**) and Golgi (**G**) (Fig. 3).

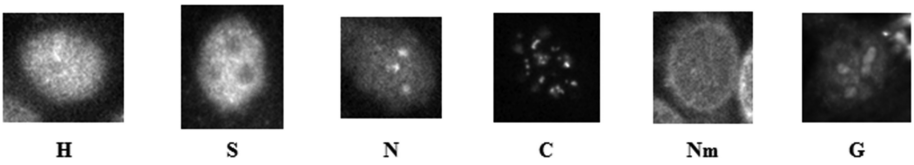

Fig. 3. Example images of I3A dataset

3.2 Data Augmentation

The I3A dataset contains 13,596 cell images. For comparison convenience, we partition the dataset according to the same percentages, i.e. 64% for training, 16% for validation and 20% for testing, as reported in [11]. A class-balanced augmentation scheme [25] is employed to the training set. Each Golgi (**G**) image was rotated for 59°, i.e. 6° for each rotation, which is three times more than that for the images of other cell patterns, i.e. 18° for each rotation. The class-balanced augmentation approach addressed the problem of unbalanced volumes of cell images contained in I3A dataset. Detailed information of

[1] http://nerone.diem.unisa.it/hep2-benchmarking/dbtools/.

Table 2. Details of augmented training datasets

	H	S	N	C	Nm	G	Total
I3A	2,494	2,831	2,598	2,741	2,208	724	13,596
Training set	1,596	1,812	1,663	1,754	1,413	463	8,701
Augmented set	31,920	36,240	33,260	35,080	28,260	27,780	192,540

Table 3. Performance evaluations of the proposed DRI. (a) MCA of using DRI Module (%) (b) MCA of using auxiliary classifiers (%)

	MCA
Shallow Inception	96.54
SRI	97.85
Inception [16]	97.93
DRI	**98.37**

(a)

	MCA
$\mathbf{DRI_3}$	96.23
$\mathbf{DRI_{1+3}}$	97.41
$\mathbf{DRI_{2+3}}$	96.23
$\mathbf{DRI_{1+2+3}}$	**98.37**

(b)

augmented training set is listed in Table 2. The original training set is augmented from 8,701 to 192,540 images.

3.3 Performance Analysis

We use the mean class accuracy (MCA) adopted by ICPR 2014 competition [3] as the criterion for performance evaluation. It measures the average of per-class accuracies, as defined in (3).

$$MCA = \frac{1}{K}\sum_{k=1}^{K} CCR_k \tag{3}$$

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

DRI Module. To investigate the effect of DRI model and network depths, we developed a Shallow Residual Inception (SRI) from the DRI by removing the ***DRI 1-2***, ***DRI 2-3***, and ***DRI 3-2*** modules. Shallow Inception was also developed by replacing the DRI modules in SRI to original Inception modules. Table 3(a) listed the MCA of Shallow Inception, SRI, and the deep frameworks using original Inception/DRI modules evaluated on testing set. It can be observed that with similar network depth, the frameworks using the proposed DRI modules, i.e. SRI and DRI, outperform the ones with original Inception modules, i.e. Shallow and Deep Inceptions, illustrating the DRI is a more advanced network architecture compared to the original Inception. Furthermore, the deeper networks, i.e. Inception and DRI, yield better results, i.e. 97.93% and 98.37%, than the shallow ones, i.e. Shallow Inception (96.54%) and SRI (97.85%), which demonstrates that the network depth enables the models to learn better feature representations for HEp-2 cell classification.

Auxiliary Classifiers. Auxiliary classifiers were only used as regularizers to improve network convergence during training in original Inception. In our experiments, combinations of the features extracted by auxiliary classifier branches, i.e. FC-1, FC-2, are found to produce improvements of classification performance. Table 3(b) presents the accuracies of combinations of different classifier branches. The MCA increases from 96.23% to 97.41% by adding FC1 branch. As the FC2 branch extracts similar features compared to FC3, there is no improvement of classification accuracy by only using FC2 and FC3. However, FC2 branch can be used as a corrector to amend errors occurred in the model using shallow classifier branch, i.e. DRI_{1+3}. The model using all branches, i.e. DRI_{1+2+3}, achieves the highest MCA of 98.37%.

3.4 Comparisons

Table 4 compares our DRI with the benchmarking algorithms, i.e. the VHAR approach [9], the Bag-of-features (BoF) model [11], the Fisher Vector (FV) model [11] and the CNN model proposed by Gao et al. [11]. The average classification accuracy (ACA) [11] is also measured for evaluation. All deep-learning approaches outperform the models using hand-crafted features, i.e. *VHAR*, *BoF* and *FV*. Our DRI achieves the highest MCA, i.e. 98.37%, and ACA, i.e. 98.49%, which are 1.61% and 1.25% higher than that of the straight-forward CNN model proposed by Gao et al.

Table 4. Comparsion with benchmarking algorithms (%)

	VHAR [9]	BoF	FV	Gao et al. [11]	SRI (ours)	DRI (ours)
MCA	93.60	94.23	95.73	96.76	97.85	**98.37**
ACA	-	94.38	96.07	97.24	97.98	**98.49**

4 Conclusion

In this paper, we presented a hybrid deep-learning network, named Deep Residual Inception, combining the architectures of two most advanced frameworks, i.e. ResNet and Inception, for HEp-2 cell image classification. The DRI model adds short-cut connections to the original Inception model for better network convergence and fuses the features extracted from shallow, medium and deep classifier branches to improve classification performance. The experimental results show that our DRI provides a significant improvement compared to existing HEp-2 classification methods.

References

1. Foggia, P., Percannella, G., Soda, P., Vento, M.: Benchmarking HEp-2 cells classification methods. IEEE Trans. Med. Imag. **32**, 1878–1889 (2013)
2. Hobson, P., Lovell, B.C., Percannella, G., Vento, M., Wiliem, A.: Benchmarking human epithelial type 2 interphase cells classification methods on a very large dataset. Artif. Intell. Med. **65**, 239–250 (2015)

3. Hobson, P., Lovell, B.C., Percannella, G., Saggese, A., Vento, M., Wiliem, A.: HEp-2 staining pattern recognition at cell and specimen levels: datasets, algorithms and results. Pattern Recognit. Lett. **82**, 12–22 (2016)
4. Nosaka, R., Fukui, K.: HEp-2 cell classification using rotation invariant co-occurrence among local binary patterns. Pattern Recognit. **47**, 2428–2436 (2014)
5. Shen, L., Lin, J., Wu, S., Yu, S.: HEp-2 image classification using intensity order pooling based features and bag of words. Pattern Recognit. **47**, 2419–2427 (2014)
6. Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., Mckenna, S.J.: An automated pattern recognition system for classifying indirect immunofluorescence images of HEp-2 cells and specimens. Pattern Recognit. **51**, 12–26 (2016)
7. Xu, X., Lin, F., Ng, C., Leong, K.P.: Adaptive co-occurrence differential texton space for HEp-2 cells classification. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 260–267. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_31
8. Taalimi, A., Ensafi, S., Qi, H., Lu, S., Kassim, A.A., Tan, C.L.: Multimodal dictionary learning and joint sparse representation for HEp-2 cell classification. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 308–315. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_37
9. Kastaniotis, D., Fotopoulou, F., Theodorakopoulos, I., Economou, G., Fotopoulos, S.: HEp-2 cell classification with vector of hierarchically aggregated residuals. Pattern Recognit. **65**, 47–57 (2017)
10. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
11. Gao, Z., Wang, L., Zhou, L., Zhang, J.: HEp-2 cell image classification with deep convolutional neural networks. IEEE J. Biomed. Health Inform. **21**(2), 416–428 (2016)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv e-print arXiv:1409.1556 (2015)
13. Bayramoglu, N., Kannala, J., Heikkila, J.: Human epithelial type 2 cell classification with convolutional neural networks. In: BIBE, pp. 1–6 (2015)
14. Phan, H.T.H., Kumar, A., Kim, J., Feng, D.: Transfer learning of a convolutional neural network for HEp-2 cell image classification. In: ISBI, pp. 1208–1211 (2016)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
16. Szegedy, C., Liu, W., Jia, Y., Sermanet, P.: Going deeper with convolutions. In: CVPR, pp. 1–9 (2015)
17. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. arXiv e-print arXiv:1602.07261 (2016)
18. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv e-print arXiv:1207.0580 (2012)
19. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: ICCV, pp. 1026–1034 (2015)
20. Lecun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**, 541–551 (1989)
21. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: ECCV, pp. 630–645 (2016)
22. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML, pp. 448–456 (2015)
23. Zhao, L., Wang, J., Li, X., Tu, Z., Zeng, W.: On the connection of deep fusion to ensembling. arXiv e-print arXiv:1611.07718 (2016)

24. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. arXiv e-print arXiv:1408.5093 (2014)
25. Jia, X., Shen, L., Zhou, X., Yu, S.: Deep convolutional neural network based HEp-2 cell classification. In: ICPR Contest and Workshop: Pattern Recognition Techniques for Indirect Immunofluorescence Images Analysis (2016)

Springer

http://www.springer.com/978-3-319-67557-2

Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings

Cardoso, J.; Arbel, T.; Carneiro, G.; Syeda-Mahmood, T.; Tavares, J.M.R.S.; Moradi, M.; Bradley, A.; Greenspan, H.; Papa, J.P.; Madabushi, A.; Nascimento, J.C.; Cardoso, J.S.; Belagiannis, V.; Lu, Z. (Eds.)

2017, XIX, 385 p. 169 illus., Softcover
ISBN: 978-3-319-67557-2