Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation

Long D. Nguyen, Dongyun Lin, Zhiping Lin
School of Electrical and Electronic Engineering,
Nanyang Technological University, 639798, Singapore

Jiuwen Cao
Key Lab for IOT and Information Fusion Technology of Zhejiang,
Hangzhou Dianzi University, 310018, China

***Abstract*—Deep convolutional neural networks (CNNs) have become one of the state-of-the-art methods for image classification in various domains. For biomedical image classification where the number of training images is generally limited, transfer learning using CNNs is often applied. Such technique extracts generic image features from nature image datasets and these features can be directly adopted for feature extraction in smaller datasets. In this paper, we propose a novel deep neural network architecture based on transfer learning for microscopic image classification. In our proposed network, we concatenate the features extracted from three pretrained deep CNNs. The concatenated features are then used to train two fully-connected layers to perform classification. In the experiments on both the 2D-Hela and the PAP-smear datasets, our proposed network architecture produces significant performance gains comparing to the neural network structure that uses only features extracted from each single CNN and several traditional classification methods.**

Keywords—feature concatenation; deep Convolutional Neural Networks; transfer learning; microscopic image

I. INTRODUCTION

Microscopic image analysis has become one of the most important fields in biomedical research as it helps scientists to understand organisms of several biological phenomena at a cellular or subcellular level. Applications of microscopic images range from diagnosing patient conditions to studying complex processes of cells. Among microscopic image analysis tasks, classification of images is of great significance. Various applications related to microscopic image classification have been developed. For example, classification of PAP-smear images [1] is performed for diagnosing cancers and classification of 2D-Hela images [2] is carried out for investigating subcellular structures of endogenous proteins. Since manual microscopic image classification is time-consuming and high-cost, several automatic methods have been proposed. The traditional automatic classification methods share the same pipeline using hand-crafted features: a features extractor (such as Local Binary Pattern (LBP) [3], Zernike moment [4], Gabor [5] or combinations of them [6]) is used jointly with a classifier (support vector machines [7] or artificial neural network [4][8]). These methods have achieved reasonably good classification results but the accuracy could still be further improved.

With the advance of computational capability of hardware and the availability of large labelled image datasets, deep convolutional neural networks (CNNs) have become one of the state-of-the-art methods for image classification in various domains. Such deep CNN structures [9][10][11][12] perform well on large datasets such as ImageNet. However, deep CNNs may suffer from serious overfitting on biomedical image datasets, which generally have only hundreds or thousands of images [13].

One promising approach to exploiting deep neural networks on small datasets is transfer learning. In transfer learning, the deep network structure is trained on a large nature image dataset before being used as a feature extractor on a small dataset. The extracted features from pretrained deep CNNs are generic and applicable to other datasets [14][15]. Recently, several approaches based on transfer learning for biomedical image classification have been proposed, such as combining extracted features from deep CNNs with hand-crafted features [16][17], fine-tuning deep CNNs without further network structure adjustment [18], ensembling the outputs generated by various CNNs [19]. These approaches are either computationally complicated for biomedical applications [16][17][19] or do not consider jointly applying features extracted from various CNNs [18].

In this paper, we use three different deep CNNs, namely Inception-v3 [9], Resnet152 [11], Inception-Resnet-v2 [12]. These CNNs are pretrained on ImageNet [20]. In [21], the features from various CNN layers are combined to incorporate both low and high-level information. Also, in [22], CNN features from images of multiple resolutions are used. Motivated by the success of exploiting multiple features for classification in [21][22], we propose to concatenate the features from these pretrained networks. The classification part of our network is inspired by [18][23], where the softmax layer is used as the classifier. We modify the classifier by adding one hidden layer for better learning capability.

The contributions in this research are summarized as follows: i) The use of transfer learning on small microscopic datasets; ii) The concatenation of features extracted from different networks to improve classification accuracy; iii) The proposal of the last two fully-connected layers to adapt the

generic features extracted from the pretrained CNNs to biomedical data. Our experiments on two microscopic images datasets, namely PAP-smear and 2D-Hela show significant accuracy improvements comparing to several traditional classification methods. Specifically, the proposed method achieves classification accuracy of 92.57% on 2D-Hela dataset and 92.63% on PAP-smear dataset. Compared with the best method available to us, the proposed method achieves the accuracy gains of 3.20% on 2D-Hela and 2.67% on PAP-smear, respectively.

II. PROPOSED METHOD

In this paper, we introduce a novel deep neural network architecture for microscopic image classification using transfer learning. In the feature extraction layers of the proposed architecture, three state-of-the-art CNNs are used and their extracted features are concatenated. The concatenated feature is fed into two fully-connected layers to generate classification outputs.

A. Pretrained CNNs for feature extraction

In this section, we adopt three deep CNN architectures, namely Inception-v3, Resnet152 and Inception-Resnet-v2 as the feature extractors of the proposed method for microscopic image classification tasks. These CNNs are pretrained on a nature image dataset (ImageNet) for distinct generic image descriptors and they can be applied to extract discriminative features from biomedical images based on transfer learning theory [14]. The structure of each adopted CNN is briefly described as follows:

a) Inception-v3

Inception-v3 [9] is an extended network of the popular GoogLeNet [10] which has achieved good classification performance in several biomedical applications using transfer learning [18][19]. Following GoogLeNet, Inception-v3 proposed an inception model which concatenates multiple different sized convolutional filters into a new filter. Such design decreases the number of parameters to be trained and thereby reduces the computational complexity. The basic architecture of Inception-v3 is illustrated in Fig. 1.

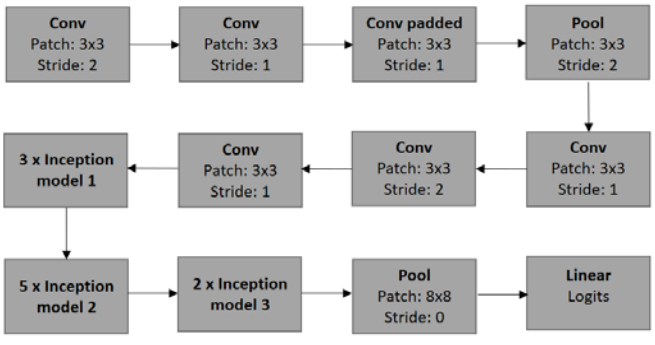

Fig. 1. The basic architecture of Inception-v3.

b) Resnet152

Residual networks (Resnet) [11] were proposed as a family of multiple deep neural networks with similar structures but different depths. Resnet introduces a structure called residual learning unit to alleviate the degradation of deep neural networks. This unit's structure is a feedforward network with a shortcut connection which adds new inputs into the network and generates new outputs. The main merit of this unit is that it produces better classification accuracy without increasing the complexity of the model. We select Resnet152 as it achieves the best accuracy among Resnet family members [11]. Fig. 2 illustrates the basic architecture of Resnet152.

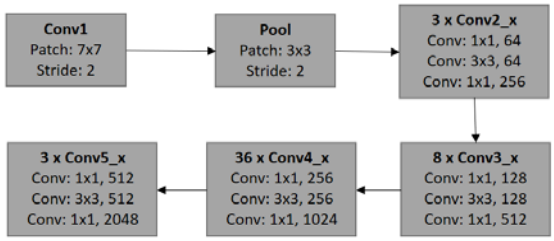

Fig. 2. The basic architecture of Resnet152 .

c) Inception-Resnet-v2

Inception-Resnet-v2 [12] is formulated based on a combination of the Inception structure and the Residual connection. In the Inception-Resnet block [12], multiple sized convolutional filters are combined by residual connections. The usage of residual connections not only avoids the degradation problem caused by deep structures but also reduces the training time. Fig. 3 shows the basic network architecture of Inception-Resnet-v2.

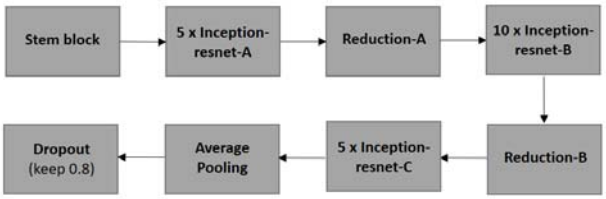

Fig. 3. The basic architecture of Inception-Resnet-v2.

B. The proposed network structure

First, the three CNN models are trained over more than 1 million natural images from 1000 categories in ImageNet [20]. As discussed in [19], after being trained on a very large labelled dataset (e.g., ImageNet), transfer learning technique can be adopted, i.e., these deep CNNs are capable to learn generic image features that are applicable to other image datasets without training from scratch. Fig. 4 illustrates the transfer learning structure for a single CNN. In this figure, the pretrained networks perform as feature extractors for generic image features and the two last layers are fully-connected layers for classification. We refer to this structure as single transfer learning network.

The details of the features generated by the pretrained deep CNNs are summarized as follows:

- Inception-v3: For one image, we extract a 2048-dimensional feature from the last fully-connected logits layer as shown in Fig. 1.
- Resnet152: For one image, we extract a 2048-dimensional feature from the last fully-pooling layer (Conv5x layer) as shown in Fig. 2.
- Inception-Resnet-v2: For one image, we extract a 1536-dimensional feature from the last fully connected layer after dropout as shown in Fig. 3.

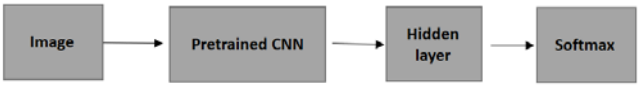

Fig. 4. The transfer learning structure for a single CNN (single transfer learning network).

After pretraining, we concatenate the extracted features from three CNNs to form a 5632-dimensional feature vector. Fig. 5 shows the structure of the proposed feature concatenation scheme. As discussed in [19], since different CNN architectures can capture diverse information in microscopic images, such concatenation of multiple CNN features integrates the information from different CNNs together to create a more discriminative feature representation compared with a single CNN structure.

Last, we feed the concatenated feature vector into two fully-connected layers for classification. Compared with the network structures in [18][19][23], we adjust the classification architecture by adding one more hidden layer. Such modification extends the learning capability of our network and helps to adopt the generic features extracted by the pertained CNNs to the specific microscopic image data. It is suggested in [19] to apply fine-tuning to the pretrained network. However, fine-tuning may cause the overfitting problem as the number of microscopic images for training might not be sufficient. Therefore, we formulate a simple structure for better adoption without overfitting.

To summarize, we propose a transfer learning network structure based on feature concatenation and design a two-layer fully-connected structure for generic feature adoption to microscopic image data. Fig. 5 illustrates the entire architecture of the proposed network.

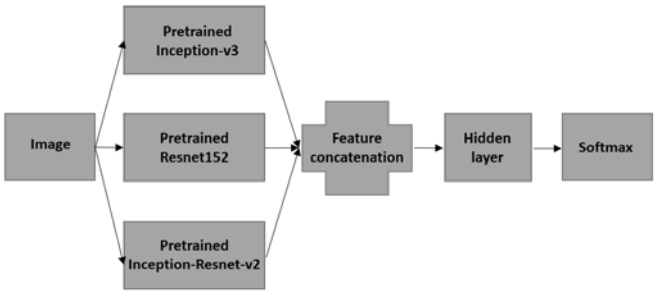

Fig. 5. Proposed feature concatenation network structure

III. EXPERIMENTS

A. Dataset Description

We evaluated our proposed network structures on two benchmark microscopic datasets: PAP-smear [1] and 2D-Hela [2]. The PAP-smear dataset consists of 917 images of various resolutions belonging to two big categories: normal and abnormal. The 2D-Hela dataset consists of 862 fluorescence microscopic images in 10 categories. The typical images of PAP-smears and 2D-Hela datasets are shown in Fig. 6 and Fig. 7, respectively.

Fig. 6. Typical images in PAP-smear dataset

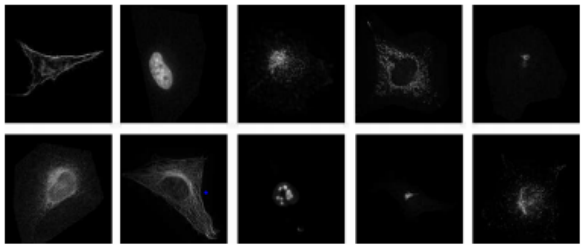

Fig. 7. Typical images in 2D-Hela dataset

B. Experimental Settings

We randomly divide our dataset into training plus validation set (80% of the images) and testing set (20% of the images). Within the training plus validation set, 75% images are used for network training while the rest 25% are for validation. For hyper parameter optimization, a grid search is performed using four-folds cross validation and early stopping to avoid overfitting. The early stopping criterion is based on the validation performance, i.e., the training will be stopped if no further validation performance improvement is made after 500 iterations.

The weights of the network are randomly initialized with a zero-mean Gaussian distribution with standard deviation 0.001. The learning rate is updated with an exponential decay factor as:

$$adaptive\_learning\_rate = learning\_rate \times decay\_rate^{(step / decay\_step)} \quad (1)$$

where the decay step is set to be 1000. The experiments are designed with 30 independent trials and the average testing results are recorded for comparison.

C. Experimental results and Analysis

We first compare the classification accuracy of our proposed feature concatenation network structure with the three single transfer learning network structures. Among

them, the single transfer learning network structures with Inception-v3 and Resnet152 are similar to the GoogLeNet and Resnet version of the network proposed in [23], respectively, except that we replace GoogLeNet (which is known as Inception-v1) with a newer version Inception-v3, Resnet50 with a deeper version Resnet152. Table I shows the average classification accuracy and its standard deviation obtained by the proposed network and three compared transfer learning network structures. It is noted from Table I that the proposed feature concatenation network consistently performs better than the three compared single transfer learning networks on the two datasets. Particularly, it is noted that the performance of the three single transfer learning networks varies greatly on the two datasets. For example, the single transfer learning network with Inception-Resnet-v2 produces the best performance on the 2D-Hela dataset among the three single transfer learning networks (with average accuracy of 92.00%), but it performs the worst on the PAP-smear dataset (with average accuracy of 89.25%). From the comparison made among the proposed feature concatenation structure and the single CNN networks, we conclude that:

- For a particular dataset, each of the three pretrained deep CNNs extracts distinct features from input images. This results in different capabilities of capturing the subtle differences between categories. Thus, the performance of each single transfer learning network varies greatly with different datasets. Choosing one pretrained network that suits all datasets at hand with a single transfer learning network structure is difficult.
- The concatenation of features from various pretrained networks helps to overcome the limitations of single network and produces robust and superior performance.

TABLE I. CLASSIFICATION ACCURACY OF TRANSFER LEARNING METHODS

Methods	***2D-Hela***	***PAP smear***
Single transfer learning network with Inception-v3 [23]	90.72 ± 1.85	89.66 ± 1.89
Single transfer learning network with Resnet152 [23]	89.72 ± 2.18	90.87 ± 1.48
Single transfer learning network with Inception – Resnet v2	92.00 ± 1.97	89.25 ± 2.23
Proposed features concatenation network	**92.57 ± 2.46**	**92.63 ± 1.68**

The format of the table is accuracy ± std (%)

We also compare the proposed method with five traditional classification methods. In these methods, several hand-crafted features, such as SIFT, LBP, SAHLBP are used, combining with classifiers such as SVM and Softmax. Table II shows the average classification accuracies of all the compared methods. Among all compared algorithms, the one proposed in [6] achieves the highest classification accuracy (89.37 % for 2D-Hela and 89.96 % for PAP-smear). Comparing with the method in [6], our proposed method shows significant accuracy gains of 3.20% on 2D-Hela and 2.67% on PAP-smear datasets, respectively.

TABLE II. CLASSIFICATION ACCURACY COMPARISON WITH EXISTING METHODS

Methods	***2D-Hela***	***PAP smear***
SIFT(BoW(VQ)+SPM+SVM) [24]	83.79 ± 2.5	84.03 ± 2.3
LBP(BoW(VQ)+SPM+SVM) [25]	81.47 ± 2.1	81.43 ± 2.1
SAHLBP(BoW(VQ)+SPM+SVM) [3]	84.49 ± 2.2	86.21 ± 2.0
SIFT+SAHLBP(BoW(VQ)+SPM+SVM) [3]	86.20 ± 2.5	87.63 ± 2.1
SIFT(BoW(LLC)+SPM+Softmax) [6]	89.37 ± 1.5	89.96 ± 1.4
Proposed features concatenation network	**92.57 ± 2.46**	**92.63 ± 1.68**

The format of the table is accuracy ± std (%)

IV. CONCLUSION

In this paper, we have proposed a transfer learning network by exploiting feature concatenation from three deep CNNs. In the experiments, the feature concatenation network shows superior performance to single CNN networks without concatenation. We also compare the proposed network with several traditional classification methods. Compared with the best competing method, the proposed method achieves significant accuracy gains (3.20% for 2D-Hela and 2.67% for PAP smear, respectively). Since transfer learning is adopted, the proposed method produces good classification performance without training deep neural networks from scratch. One future consideration regarding this method is extending the current concatenation scheme to combine both CNN and hand-crafted features or combine features from both low and high layers of the current deep CNNs to further improve classification performance.

ACKNOWLEDGMENT

We wish to acknowledge the funding support for this project from Nanyang Technological University under the Undergraduate Research Experience on Campus (URECA) program.

REFERENCES

[1] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," Nature inspired Smart Information Systems (NiSIS 2005), pp. 1–9, 2005.

[2] M. V. Boland and R. F. Murphy, "A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells," *Bioinformatics*, vol. 17, no.12, pp. 1213–1223, 2001.

[3] D. Liu, S. Wang, D. Huang, G. Deng, F. Zeng, and H. Chen, "Medical image classification using spatial adjacent histogram based on adaptive local binary patterns," *Computers in biology and medicine*, vol. 72, pp.185–200, 2016.

[4] N. A. Hamilton, R. S. Pantelic, K. Hanson, and R. D. Teasdale, "Fast automated cell phenotype image classification," *BMC bioinformatics*, vol. 8, no. 1, p. 110, 2007.

[5] S.-C. Chen, T. Zhao, G. J. Gordon, and R. F. Murphy, "Automated image analysis of protein localization in budding yeast," *Bioinformatics*, vol. 23, no. 13, pp. i66–i71, 2007.

[6] D. Lin, Z. Lin, L. Sun, K. A. Toh, and J. Cao, "LLC encoded BoW features and softmax regression for microscopic image classification," 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, 2017, pp. 1-4.

[7] D. Lin, Z. Lin, S. Sothiharan, L. Lei, and J. Zhang, "An SVM based scoring evaluation system for fluorescence microscopic image classification," in 2015 IEEE International Conference on Digital Signal Processing (DSP). IEEE, 2015, pp. 543–547.

[8] M. V. Boland, M. K. Markey, and R. F. Murphy, "Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images," Cytometry, vol. 33, no. 3, pp. 366–375, 1998.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2818-2826.

[10] C. Szegedy, W. Liu, Y. Jia and P. Sermanet et al., "Going deeper with convolutions," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1–9.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning". In AAAI, 2017, pp. 4278-4284.

[13] D. Lyndon, A. Kumar, J. Kim, P. H. W. Leong, and D. Feng, "Convolutional neural networks for medical clustering," in Proc. Workshop CLEF 2015 Working Notes, vol. 1391, 2015.

[14] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. "How transferable are features in deep neural networks?", in Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14), Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.), Vol. 2. MIT Press, Cambridge, MA, USA, 3320-3328.

[15] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition," 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, 2014, pp. 512-519.

[16] F. Ciompi, K. Chung, B. V. Ginneken, and B. D. Hoop et al, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box", Medical image analysis, vol .26, no. 1, pp. 195-202, 2015.

[17] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification", Proc. SPIE Med. Imag. Computer-Aided Diagnosis, vol. 9414, 2015.

[18] H. C. Shin, H. R. Roth, M. Gao, and L. Lu, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," in IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, May 2016.

[19] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 31-40, Jan. 2017.

[20] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 248–255.

[21] L. Zheng, Y. Zhao, S. Wang, J. Wang, and Q. Tian, "Good practice in CNN feature transfer" in, Apr. 2016, [online] Available: https://arxiv.org/abs/1604.00133.

[22] J. Kawahara and G. Hamarneh, "Multi- resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers", in Wang L., Adeli E., Wang Q., Shi Y., Suk HI. (eds) Machine Learning in Medical Imaging. MLMI 2016. Lecture Notes in Computer Science, vol 10019, pp. 164-171. Springer, Cham (2016).

[23] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum and C. H. Davis, "Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery," in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 4, pp. 549-553, April 2017.

[24] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2. IEEE, 2006, pp. 2169–2178.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971–987, 2002.