Learning Image Representations using Deep Siamese CNNs for Content-Based Medical Image Retrieval

Yu-An Chung* **Wei-Hung Weng†**
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge, MA 02139
{andyyuan,ckbjimmy}@mit.edu

1 Introduction and Related Works

Effective feature extraction and data representation are key factors of successful medical imaging predictive modeling tasks [Litjens et al., 2017]. Researchers usually adopt domain knowledge and labeling from clinical experts to design image features for image learning tasks. However, using predefined features for representation limits the chance to discover novel features. It is also very expensive to have clinicians and experts to label the data manually, and such labor-intensive approach is hard to be scaled and generalized. Recently, deep neural networks have been adopted in medical image analysis and yielded the state-of-the-art performance in different tasks, such as the medical image classification [Esteva et al., 2017], segmentation [Havaei et al., 2017], image generation [Nie et al., 2017], captioning [Shin et al., 2015], and content-based medical image retrieval (CBMIR) due to its capability of learning representations [Litjens et al., 2017, Bengio et al., 2013].

CBMIR helps clinicians make decisions by retrieving similar cases and images from the electronic medical image database. CBMIR for knowledge discovery and similar image identification in massive medical image database have been explored. However, deep learning is not widely adopted in the CBMIR task except for few studies on lung CT [Sun et al., 2017], prostate MRI [Shah et al., 2016] and X-ray [Anavi et al., 2016, Liu et al., 2016]. Nevertheless, the previous works focused more on combining single pre-trained CNN structure with other techniques and heavily depended on exact manually annotated label information.

To address the issues, we proposed CNN-based end-to-end deep Siamese convolutional neural networks (SCNN) [Bromley et al., 1994] (Figure 1 left) that can learn fixed-length image representation from only image pair information and performed the experiment using CBMIR of diabetic retinopathy (DR) fundus images as an application to validate our approach. We hypothesized that the proposed deep SCNN can reduce the dependency of expert labeling but still learn image representations well.

2 Methods and Materials

Deep Siamese Convolutional Neural Networks SCNN architecture is a variant of neural network that can find the relationship and similarity between the input objects. It has multiple symmetric subnetworks tying the same parameters and weights and updating mirrorly, and cojoining at the top by an energy function. Two identical CNNs with the same weights were constructed. Each identical CNN was constructed using ResNet-50 [He et al., 2016] architecture with the ImageNet pre-trained weight. We used 25% dropout for regularization to reduce overfitting and adopted batch normalization [Srivastava et al., 2014, Ioffe and Szegedy, 2015]. The rectified linear units (ReLU) nonlinearilty is applied as the activation function for all layers, and we used Adam optimizer [Kingma and Ba, 2014] to control learning rate. The similarity between paired images was calculated by

*Co-first author
†Co-first author, corresponding author

Submitted to 31st Conference on Neural Information Processing Systems (NIPS 2017). Do not distribute.

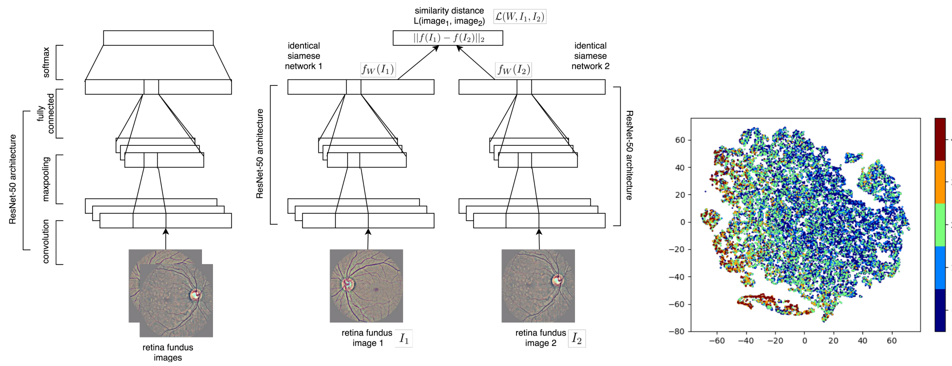

Figure 1: (Left) Structure of proposed deep SCNN and (right) the t-SNE visualizations for the distribution of learned retina fundus image representation embedding from the last layer of SCNN. Colors represent the real expert-labeled severity.

Euclidian distance, and we defined loss function by computing the contrastive loss [Hadsell et al., 2006]. In this study, we compared the deep SCNN to the single supervised ResNet-50 architecture.

We used mean reciprocal rank (MRR, $\frac{1}{Q}\Sigma_{i=1}^{Q}\frac{1}{rank_i}$, where Q is the query size and $rank_i$ means that the rank of the real first-ranked item in the i-th query) and mean average precision (MAP, $\frac{1}{Q}\Sigma_{i=1}^{Q}AveP$, where $AveP$ is the area under precision-recall curve) for evaluation.

Data and Preprocessing We used the full training set of Kaggle Diabetic Retinopathy Detection challenge with 35,125 fundus images. Five clinical severity labels from normal to severe were labeled by experts and used for single supervised CNN. Further preprocessing and data augmentation were done to handle the variation between image conditions and class imbalance. The original and augmented images were pooled together and split into 70% train and 30% test data based on stratification of class labels.

3 Results

For both single supervised CNN and deep SCNN architecture, we extracted the last bottleneck layer as our latent representations of retina fundus images. We visualized the data distribution of the deep SCNN's image representations using principal component analysis and t-Distributed Stochastic Neighbor Embedding (t-SNE) [Maaten and Hinton, 2008] (Figure 1 right). A clear clinically interpretable transition from healthy cases (label 0) to severe disease (label 3 and 4) is shown in the t-SNE embedding. For CBMIR, Table 1 shows that the proposed deep SCNN architecture yielded a comparable performance even with minimal expert labeling information compared to the single supervised CNN architecture, which relied on the exact expert labeling.

Table 1: Performance measurement of CBMIR using latent representations from single pre-trained CNN or deep SCNN

Layer	CNN third-last	CNN second-last	CNN softmax	SCNNs last
MAP	0.6209	0.6369	0.6673	0.6492
MRR	0.7608	0.7691	0.7745	0.7737

4 Conclusions

In this paper, we have presented a new strategy to learn latent representation of medical images by learning an end-to-end deep SCNN with only image pair information. We performed the experiment on the CBMIR task using publicly DR image dataset and demonstrated that our proposed deep SCNN approach is comparable to the commonly used single pre-trained CNN architecture, which requires actual expert labeling that is expensive in the machine learning tasks.

References

Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan. Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval. In *SPIE Medical Imaging*, 2016.

Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah. Signature verification using a siamese time delay neural network. In *NIPS*, 1994.

A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *CVPR*, 2006.

M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. C. Courville, Y. Bengio, C. Pal, P. Jodoin, and H. Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35: 18–31, 2017.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014.

G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

X. Liu, H. R. Tizhoosh, and J. Kofman. Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In *IJCNN*, 2016.

L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen. Medical image synthesis with context-aware generative adversarial networks. In *MICCAI*, 2017.

A. Shah, S. Conjeti, N. Navab, and A. Katouzian. Deeply learnt hashing forests for content based image retrieval in prostate mr images. In *SPIE Medical Imaging*, 2016.

H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers. Interleaved text/image deep mining on a very large-scale radiology database. In *CVPR*, 2015.

N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1): 1929–1958, 2014.

Q. Sun, Y. Yang, J. Sun, Z. Yang, and J. Zhang. Using deep learning for content-based medical image retrieval. In *SPIE Medical Imaging*, 2017.