Learning Deep Representations of Medical Images using Siamese CNNs with Application to Content-Based Image Retrieval

Yu-An Chung, Wei-Hung Weng

Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Abstract

Deep neural networks have been investigated in learning latent representations of medical images, yet most of the studies limit their approach in a single supervised convolutional neural network (CNN), which usually rely heavily on a large scale annotated dataset for training. To learn image representations with less supervision involved, we propose a deep Siamese CNN~(SCNN) that can be trained with only binary image pair information. We evaluated the learned image representations on a task of content-based medical image retrieval using a publicly available multiclass diabetic retinopathy fundus image dataset. The experimental results show that our proposed deep SCNN is comparable to the state-of-the-art single supervised CNN, and requires much less supervision for training.

Background

Motivation

- Effective feature extraction and data representation are key factors to successful medical imaging tasks. Researchers usually adopt domain knowledge and ask for annotations from clinical experts.
- However, using predefined features for data 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 annotation task limits the scalability of learning medical imaging representations.

Deep Representations of Medical Images

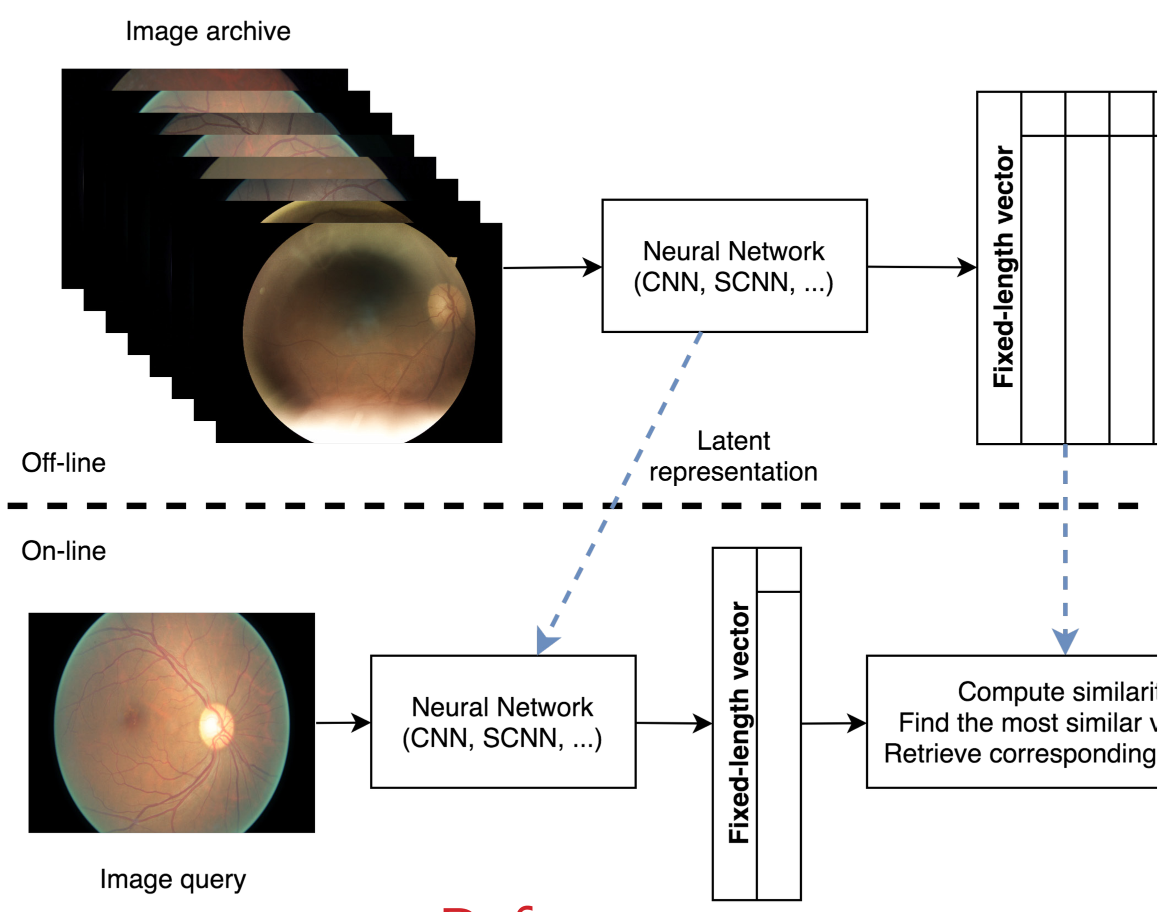
- Recently, deep neural networks have been adopted in medical image and yield the state-of-the-art performance [Litjens 2017].
- For medical image, researchers mainly use CNN for different tasks such as image classification [Esteva 2017], segmentation, generation, captioning, and content-based image retrieval.

Content-based Image Retrieval (CBMIR)

- CBMIR is a task that helps clinicians make decisions by retrieving similar cases and images from the electronic medical image database [Muller 2004].
- CBMIR for knowledge discovery and similar image identification in massive medical image database have been explored [Kumar 2013].
- Deep learning is not yet widely adopted in CBMIR except for few studies on lung CT, prostate MRI, and X-ray image.
- However, the previous works focused more on combining single pre-trained CNN structure with other techniques. Furthermore, they also relied heavily on high quality manual ground truth labeling.

Proposed Approach

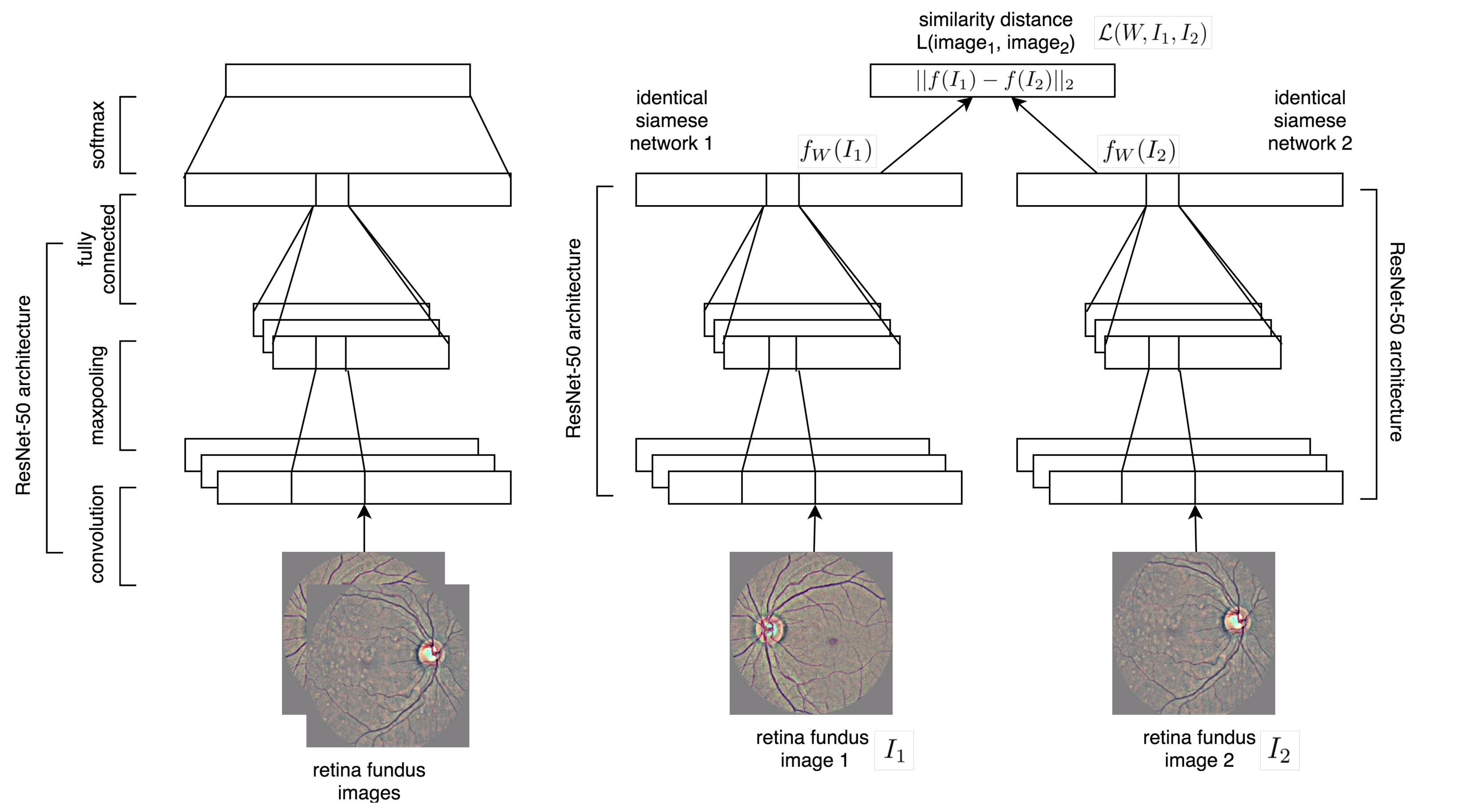
- Deep Siamese CNN (SCNN) that can learn fixed-length latent image representation from only image pair information in order to reduce the dependency of using actual class labels annotated by human experts [Bromley 1994, Koch 2015].
- Evaluate the learned image representations on the task of CBMIR using a publicly available diabetic retinopathy(DR) fundus image dataset.



Reference

- Bromley et al. Signature verification using a siamese time delay neural network. NIPS, 1994.
- Esteva et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017.
- Hadsell et al. Dimensionality reduction by learning an invariant mapping. CVPR, 2006.
- Koch et al. Siamese neural networks for one-shot image recognition. ICML EL Workshop, 2015.
- Litjens et al. A survey on deep learning in medical image analysis. Medical Image Analysis 2017.
- Müller et al. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. International Journal of Medical Informatics 2004.

Materials and Methods



Deep Siamese CNN

- Deep SCNN can find the similarity between input objects.
- It has multiple symmetric subnetworks tying the same parameters and weights and updating mirrorly, and cojoining by an energy function.
- Two identical CNNs were built using ResNet-50 with the ImageNet pre-trained weight, with ReLU, 25% dropout, batch normalization, and Adam.
- The similarity between images was calculated by Euclidean distance, and we defined loss function by computing the contrastive loss [Hadsell 2006].

$$\mathcal{L}(W, I_1, I_2) = \mathbf{1}(L=0)\frac{1}{2}D^2 + \mathbf{1}(L=1)\frac{1}{2}[\max(0, margin - D)]^2 \qquad MRR = \frac{1}{Q}\Sigma_{i=1}^{Q}\frac{1}{rank_i}$$

- Use mean reciprocal rank (MRR) and mean average precision (MAP) for evaluation.

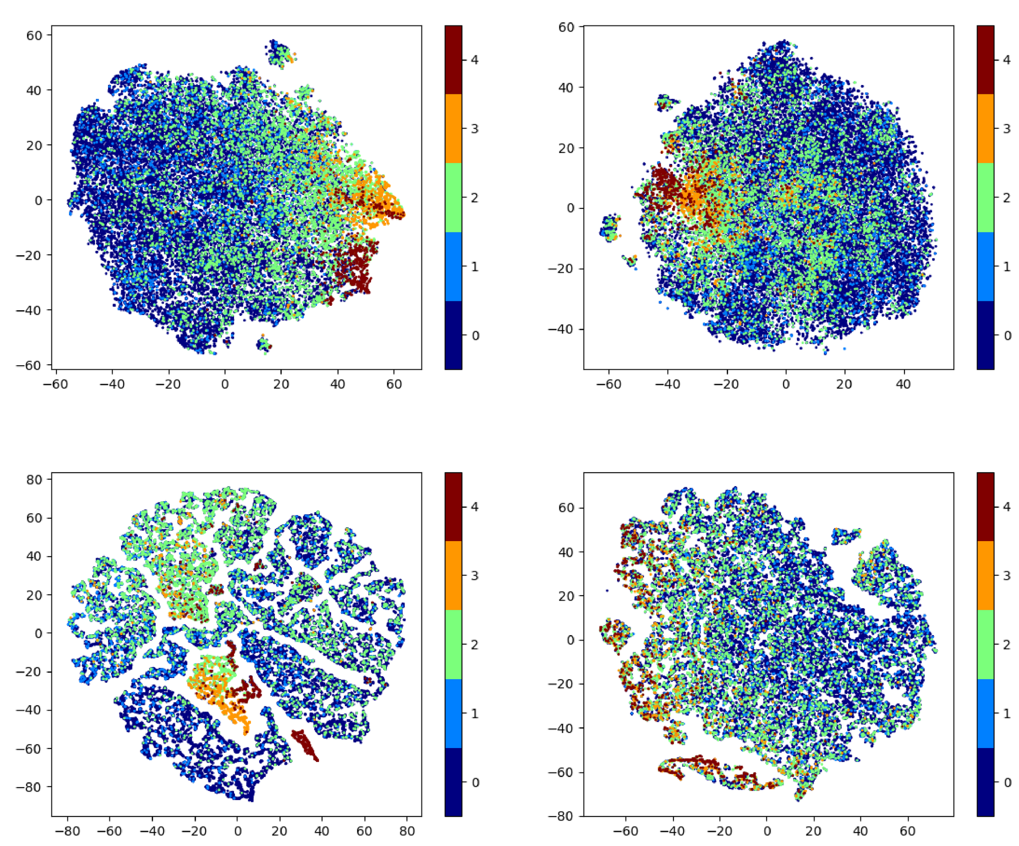
$$MAP = \frac{1}{Q}\Sigma_{i=1}^{Q} AveP$$

Experiment

Data Source

- Full training set of Kaggle Diabetic Retinopathy Detection challenge with 35,125 fundus images and five clinical severity labels from normal to severe were given by experts and used for the single CNN approach.
- Data preprocessing includes rescaling, quantitative and qualitative data augmentation.

Learning Representations



- Extract the last bottleneck layer.
- PCA was first adopted to reduce the feature dimension to 50, then t-SNE was applied to further reduce the dimension to two.
- Ground truth multiclass labels given by experts are arbitrary.
- The sliding scale representations are therefore more desirable to express the real DR pathology.

CBMIR

Layer	CNN (third-last)	CNN (second-last)	CNN (softmax)	SCNN (last layer)
MAP	0.6209	0.6369	0.6673	0.6492
MRR	0.7608	0.7691	0.7745	0.7737

- The representation learned by deep SCNN with binary labeling outperformed those learned by single CNN.

Conclusion

We propose an end-to-end deep SCNN model for learning latent representations of medical images with minimal expert labeling efforts, and apply them in the task of CBMIR using retina fundus images. Experimental results show that SCNN's performance is comparable to that of the state-of-the-art CBMIR method using single supervised pre-trained CNN, but requires much less supervision for training. Future investigation will focus on different network architectures, other ranking metrics and applying to different medical image datasets.