Towards an Affordable Deep Learning System: Automated Intervertebral Disc Detection in X-ray Images

Ruhan Sa[a], William Owens[b], Raymond Wiegand[c], Vipin Chaudhary[a]
[a]Department of Computer Science and Engineering
State University of New York (SUNY) at Buffalo, Buffalo, NY 14260–2500
[b]School of Medicine and Biomedical Sciences
State University of New York (SUNY) at Buffalo, Buffalo, NY 14214-3013
[c]Spine Metrics, Inc. 520 Huber Park Court, Weldon Spring. MO 63304

ABSTRACT

Adult Spinal Deformity is a prominent medical issue with about 68% of the healthy, elderly population suffering from the disease [1]. Detailed biomechanical assessment is needed both in the presurgical planning of structural spinal deformity as well as in early functional biomechanical compensation in ambulatory spinal pain patients. When considering automation of this process, we have to look at photographic intervertebral disc detection technique as a way to produce a detailed model of the spine with appropriate measurements required to make efficient and accurate decisions on patient care. Deep convolutional neural network (CNN) has given remarkable results in object detection tasks in recent years. However, massive training data, computational resources and long training time is needed for both training a deep network from scratch or fine-tuning a network. Using pre-trained model as feature extractor has shown promising result for moderate sized medical data [2]. However, most work have extracted features from the last layer and little has been explored in terms of the number of layers needed for best performance. In this work we trained Support Vector Machine (SVM) classifiers on different layers of CaffeNet [3] features to show that "deeper the better" concept does not hold for intervertebral disc detection task. Furthermore, our experimental results show the potential of using very small training data to yield satisfactory classification performance with accuracy up to 97.2% using only 15 training data.

Keywords: Deep Learning, Convolutional Neural Network, intervertebral disc detection

1. INTRODUCTION

A large proportion of the population suffers from spinal deformity [1] and there is clear evidence showing the correlation between increasing deformity and increasing pain [4]. Kim et al [5] shows that the detailed knowledge of biomechanical parameters of human spine is important for accurate and efficient patient care. Spine measurement system (SMS) serves as a tool to provide interpretation of biomechanical data and increased efficiency in the collection. Automation of SMS has brought great challenges due to reduced image contrasts, variation of scales in X-ray images and different numbers of intervertebral discs in each image.

Automatic object detection has been extensively studied in computer vision field and among many algorithms, deep convolutional neural networks have outperformed many state of the art algorithms in visual perception challenges, which has brought much attention in medical image field. There are three major approaches in terms of deep learning based image analysis system: training a network from scratch [7][8], fine tuning a network[12][13] and using a deep learning network as a feature extractor [2][6][9][11]. Training a deeper network from scratch is harder [10] and usually requires at least thousands of training images or strong augmented training images [7][8]. In order to overcome this issue, fine tuning technique is explored for medium sized dataset in [13], where the network is initialized using pre-trained model parameters. Using pre-trained model as feature extractor has also shown promising experimental results for medium sized dataset. For example Razavian et al [11] have used pre-trained convolutional neural network called OverFeat as feature extractor and trained SVM directly on various database based on the feature. Ginneken et al [2] have used the same pre-trained model as feature extractor and tested on 3D CT images for pulmonary nodule detection. Both work have shown good results. However, in both of the above work, features are extracted from the last convolutional layer. Less work has been explored in the literature about how features extracted from different layers affect the detection performance. For example, Figure 1. shows the filter samples of

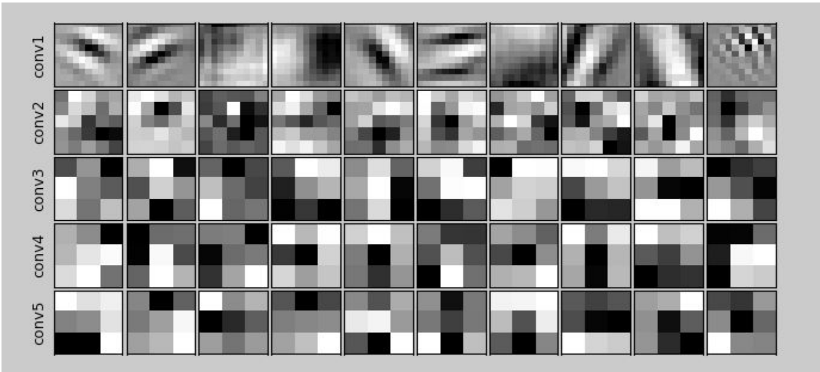

Figure 1. Samples of filters in pre-trained CaffeNet model.

five convolutional layers in CaffeNet. We can see that lower level convolutional filters describe low level features and higher level convolutional filters describe more abstract features. Does it always give better performance using higher level feature? In this work we show that for intervertebral disc detection problem, “deeper the better” concept does not hold when using pre-trained model as feature extractor. We utilized pre-trained CaffeNet, which is trained on ImageNet (natural images), extracted features from each convolutional layer and trained SVM classifiers on each of the layer separately. By comparing the performances between different layers, we observe that the classifiers trained on third and fourth layer perform better than the classifier trained on last layer. Detailed results shown in Section 3. Our experiment also shows that using only 15 training samples can yield sufficient good results with accuracy score up to 97.2%.

2. METHODS

2.1 Convolutional Neural Network Model

Krizhevsky et al proposed CaffeNet [14] that is a deep convolutional neural network trained on ImageNet as classifier of natural images. Here we use it as a feature extractor and train SVM classifiers on the output of each convolutional layer. Object detection will be conducted in a sliding window fashion. Each window size is 256x256 and is classified into either disc or non-disc class. Figure 2. shows the CaffeNet model. There are 5 convolutional layers in total and the output dimensions are: (55x55x96), (27x27x256), (13x13x384), (13x13x384) and (13x13x256) respectively. It is known in deep learning field that the depth of representation is important in many visual recognition tasks . He et al [10] trained a network as deep as 152 layers. However, for intervertebral disc detection task, we show that the classifier trained on feature extracted from deeper layer does not give better result.

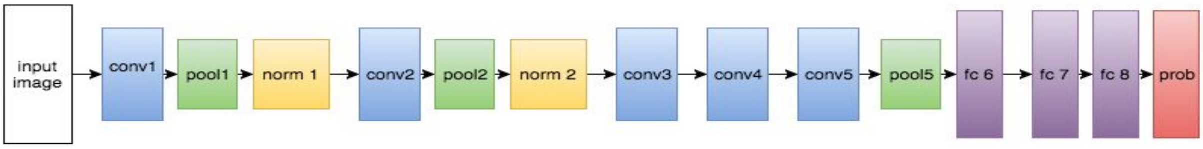

Figure 2. CaffeNet model structure

2.2 Data preparation and analysis

Our training data consists of 15 patient samples as shown in Figure 3. We preprocessed it into image patches of size 256x256 as shown in Figure 4. The total number of patches used for training is 25100.

Figure 3. Training samples: lateral lumbar X-ray images.

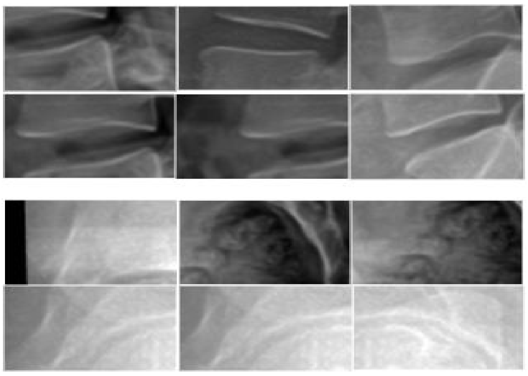

Figure 4. First two rows show the positive samples and last two rows show the negative samples

These training images are fed into CaffeNet and the output of different layers are extracted as features. In our experiment, we explore the features from layer conv3, conv4 and con5. conv1 and conv2 are not experimented, as we observe that conv1 and conv2 output feature dimensions are both a magnitude higher than the training sample number.

Table 1. Detection results compared between different convolution layers.

Configuration	True positive	True negative	False positive	False negative	Overall accuracy
conv3+SVM	0.785	0.986	0.214	0.014	0.970
conv4+SVM	0.790	0.987	0.210	0.013	0.972
conv5+SVM	0.741	0.984	0.258	0.015	0.966

3. RESULTS

In our experiment, we trained linear SVM classifiers based on the output of CaffeNet conv3, conv4 and conv5 layers. The performance of different classifiers on 4056 testing image patches are listed in Table 1. We can see that the classification accuracy is higher for classifier trained on lower level convolutional features. The best accuracy is as high as 97.2%, which shows that using very small training samples can yield satisfactory result when using deep convolutional network as feature extractor.

4. CONCLUSIONS

Deep learning is a state-of-the-art technique used in many visual recognition tasks. However the computational time and space it needs to train a network from scratch is very high and the demand for massive training data is not applicable for medical image field. In our work we show that using a pre-trained model as feature extractor with very small training dataset can give satisfactory result and we further conclude that in intervertebral disc detection task, deeper neural network features shows less performance compared to lower lever features.

REFERENCES

[1] Devon J. Ryan, Themistocles S. Protopsaltis, Christopher P. Ames, Richard Hostin, Eric Klineberg and *et al,* "T1 Pelvic Angle (TPA) Effectively Evaluates Sagittal Deformity and Assesses Radiographic Surgical Outcomes Longitudinally," Spine, Volume 39, Number 15, Papers 1203-1210 (2014).

[2] Bram van Ginneken, et al., "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans."2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE, 2015.

[3] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, "Caffe: Convolutional Architecture for Fast Feature Embedding," Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.

[4] Steven Glassman, et al., "The Impact of Positive Sagittal Balance in Adult Spinal Deformity," Spine, Volume 30(18), 15 September 2005, pp 2024-2029.

[5] Kim, Y. B., Kim, Y. J., Ahn, Y. J., Kang, G. B., Yang, J. H., Lim, H., & Lee, S. W.," A comparative analysis of sagittal spinopelvic alignment between young and old men without localized disc degeneration," European Spine Journal, 23(7), 1400-1406 (2014)

[6] Suzani, Amin, et al., "Fast Automatic Vertebrae Detection and Localization in Pathological CT Scans-A Deep Learning Approach," International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[7] Olaf Ronneberger, Philipp Fischer and Thomas Brox, " U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[8] Xie, Yuanpu, et al., "Beyond classification: structured regression for robust cell detection using convolutional neural network," International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[9] Carneiro, Gustavo, Jacinto Nascimento, and Andrew P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.

[10] Kaiming He, et al., "Deep residual learning for image recognition." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[11] Sharif Razavian, Ali, et al., "CNN features off-the-shelf: an astounding baseline for recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014.

[12] Yosinski, Jason, et al. ,"How transferable are features in deep neural networks?," Advances in neural information processing systems. 2014.

[13] Tajbakhsh, Nima, et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," IEEE transactions on medical imaging 35.5 (2016): 1299-1312.

[14] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems. 2012.