Unregistered Bosniak Classification with Multi-phase Convolutional Neural Networks

Myunggi Lee[1], Hyeogjin Lee[1], Jiyong Oh[1], Hak Jong Lee[2], Seung Hyup Kim[2], Nojun Kwak[1*]

[1]Graduate School of Convergence Science and Technology
[2]Department of Radiology, College of Medicine
Seoul National University, Korea
{myunggi89,hjinlee,yong97,hakjlee,kimshrad,nojunk}@snu.ac.kr

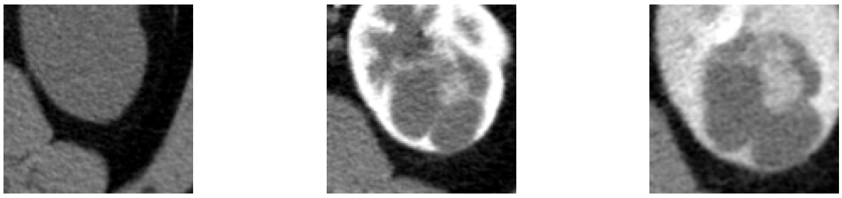
Abstract. Deep learning has been a growing trend in various fields of natural image classification as it performs state-of-the-art result on several challenging tasks. Despite its success, deep learning applied to medical image analysis has not been wholly explored. In this paper, we study on convolutional neural network (CNN) architectures applied to a Bosniak classification problem to classify Computed Tomography images into five Bosniak classes. We use a new medical image dataset called as the Bosniak classification dataset which will be fully introduced in this paper. For this data set, we employ a multi-phase CNN approach to predict classification accuracy. We also discuss the representation power of CNN compared to previously developed features (Garbor features) in medical image. In our experiment, we use data combination method to enlarge the data set to avoid overfitting problem in multi-phase medical imaging system. Using multi-phase CNN and data combination method we proposed, we have achieved 48.9% accuracy on our test set, which improves the hand-crafted features by 11.9%.

Keywords: Medical image, Bosniak Classification, Deep Convolutional Neural Network, Unregistered Medical image

1 Introduction

Deep Learning has made a significant breakthrough in natural image classification systems and has become a powerful tool in various fields of artificial intelligence. In particular, the use of deep Convolutional Neural Network (CNN) in computer vision problems has been an essential. In the recent progress in deep learning, many deep convolutional neural network architectures have been proposed, such as AlexNet [1], VGGNet [2] and GoogLeNet [3] which generate state-of-the-art results on a variety of challenging tasks. One of the advantages brought by deep learning model is the high-level features produced by the top layers of the model compared to the previous hand-crafted features.

* Corresponding author. This work was supported by Brain Fusion Program of Soul National University (800-20140265).



(a) (b) (c)

Fig. 1. Example of three phases of CT images. (a) Before contrast injection, (b) 50 seconds after contrast injection and (c) 5 minutes after contrast injection of ROI cropped from CT images.

Though deep learning has made great advances in natural image classification problems, its potential in the field of medical image has not been completely explored because it is difficult to obtain large enough data in certain medical image problems and the DICOM[1] files are hard to annotate manually. For the training of CNN, it usually requires hundreds of thousands of data to achieve reasonable results avoiding overfitting. For this reason, the amount of publicly available medical images is increasing rapidly. Nevertheless, it is difficult to gather a large quantity of such medical images because of patient privacy and security policies. Moreover, regardless of policies, the hardness of annotating such a unique data format leads to an exhaustive work. In spite of the issues above, we have seen many approaches using deep learning applied to medical image problems [4] [5].

In this paper, we propose a multi-phase CNN trained in a supervised fashion with labeled data to automatically diagnose the Bosniak categories. The Bosniak classification system of renal cystic masses divides renal cystic masses into five categories based on image characteristics on contrast-enhanced Computed Tomography (CT). We manually annotate regions of interest (ROI) for CT images of three phases which will be explained in Section 2. Using our data augmentation schemes, each sample of unregistered ROI triplets[2] is fed to the proposed multi-phase CNN structure. The proposed multi-phase CNN has achieved 48.9% classification accuracy which enhances both a single CNN and a hand-crafted feature like Gabor Features plus SVM by more than 10% respectively.

2 Bosniak Classification Problem

The Bosniak classification is a practical and accurate method to evaluate renal cystic lesions which divides renal cystic masses into five categories according to image characteristics on contrast-enhanced CT. It is helpful for doctors in predicting a risk of malignancy and suggesting either follow up or treatment.

[1] Digital image and Communications in Medicine (DICOM) is a standard for handling, storing, printing, and transmitting information in medical image.

[2] The term ‘unregistered’ is used to indicate that the three images from different phases have different shapes and sizes of lesions.

CT image criteria makes it possible to analyse renal cysts' contour and contents, presence of septations or calcifications, and enhancement after contrast agent injection. The categories are determined by characteristics of kidney region CT images from the three phases which are 'before contrast injection', '50 seconds after contrast injection' and '5 minutes after contrast injection' as shown in Fig. 1. Combining all the detailed characteristics of each phase, we are able to classify which Bosniak category the patient belongs to. One of the challenges of Bosniak classification with renal cystic lesions is that diverse characteristics of such lesions per Bosniak class as well as overall analysis of three phases are needed for accurate classification [6]

There has been some approaches to automatically classify Bosniak class in machine learning fashion [7] [8]. Most previous works in medical image focus on hand-crafted features like Gabor features, which are unable to comprehensively represent low-variant medical images for classification [9]. Moreover, as far as we know, it has not been reported the use of CNNs in Bosniak classification problem as well as the analysis of unregistered input images with the multi-phase CNN approach.

3 Algorithm

3.1 Data Acquisition

The database we use in our experiment is CT images of patients from Seoul National University BunDang Hospital (SNUH) with an Institutional Review Board (IRB) approval. The dataset maintains overall 371 patients' CT images. Each patient has three DICOM files which is taken from 'before contrast-enhanced injection', '50 seconds after injection' and '5 minutes after injection'.

To classify Bosniak class, we manually annotate several ROIs using our annotation software. In slice-by-slice images from DICOM files, we cropped ROIs that have characteristics of complex renal cysts as shown Fig. 1. The number of cropped images per patient varies from 2 to 28. Overall 8,758 image patches were extracted from DICOM files. However, as it can be seen from Table 1, the number of image patches per class is highly unbalanced and not enough for training a CNN. Also, these three-phase image patches are unregistered, which makes it more difficult to classify accurately.

Table 1. Data Distribution of Bosniak Classification Database

Category	# Patient	Pre	50 sec	5 min	Total
Bosniak I	112	798	799	807	2404
Bosniak II	145	1102	1163	1139	3404
Bosniak IIF	26	302	296	295	893
Bosniak III	45	274	298	307	879
Bosniak IV	43	390	394	394	1178
Total	371	2866	2950	2942	8758

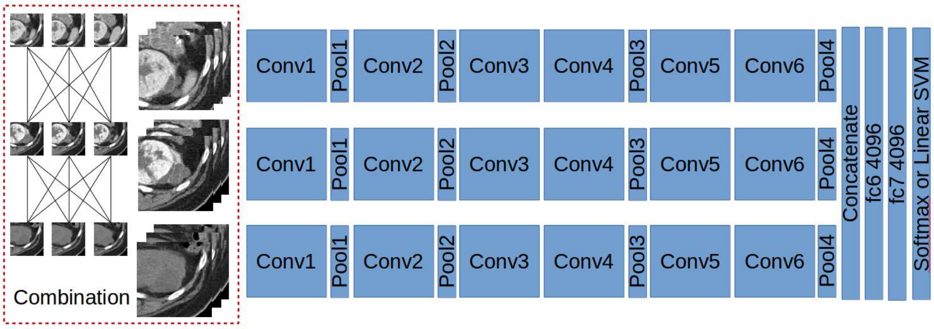


Fig. 2. Overall architecture and data combination used in our experiment. Our proposed architecture has three independent CNNs (not sharing weights), each CNN has 9 layers. The first 6 layers has convolution operation following by pooling operation. Due to the padded convolutions for the third and fifth layers, the output feature map sizes are equal to the input feature map size. All the feature maps extracted from layer 6 (pool4) are concatenated followed by a fully-connected layer. The concatenated feature map is encoded into fully-connected layer with 28,800 dimension.

3.2 Data Augmentation

In order to train our proposed approach, we enrich our data set by applying data augmentation techniques to each image patches. We also applied a new input strategy, namely data combination for each patient. The image cropped from DICOM files has different sizes that varies from 40 by 40 to 160 by 160. First we prepare two sets of image sizes by resizing each image to 128 by 128 and 256 by 256. Then, we generate 10 times more image patches by random translations and horizontal reflections. The resulting input image sizes to the CNN are 112×112 and 227×227, respectively.

Another approach to enrich our dataset is the data combination method specifically designed for our application. For each patient we have more than 2 images per phase from which the image patches are cropped in slice-by-slice manner. With multi-phase CNN architecture taken into consideration, we proceed the inter-phase data combination to enlarge our data set, that is, choosing one image from each phase for a combination. For example, suppose a patient has two pre-injection, three 50 seconds, four 5 minutes after injection patches, then we get 24 ($2 \times 3 \times 4$) different input combinations fed to three-stream CNN architecture we propose. Figure 2 illustrates overall architecture and data combination method used in our experiment.

3.3 Multi-Phase Convolutional Neural Network

In this section, we describe the algorithm used in our experiment and methods we used to evaluate the Bosniak Classification system. Our goal is to predict Bosniak class according to CT images of three phases. We should consider overall

Table 2. Single Neural Network Architecture (for input image size of 128×128)

Layer	1	2	3	4	5	6	7	8	9
Stage	conv+max	conv+max	conv	conv+max	conv	conv+max	full	full	full
# Channel	32	64	96	96	128	128	4096	4096	5
Filter Size	5x5	3x3	3x3	3x3	3x3	3x3	-	-	-
Conv.Stride	2	2	1	1	1	1	-	-	-
Pooling Size	2	2	-	2	-	2	-	-	-
Pooling Stride	2	1	-	2	-	2	-	-	-
Padding Size	0	0	1	0	1	0	-	-	-
Spatial Input	112x112	54x54	26x26	26x26	12x12	12x12	5x5		

analysis of the characteristics of ROIs of three phases which are cropped from CT images.

A single CNN model consists of multiple convolution and sub-sampling (pooling) layers. Each of the layers is performed by using different convolutional operation. The result of convolution defines the features of input data which can be interpreted as corners, curves, lines and so on. After each convolution layer, a pooling layer follows to reduce the dimension of each feature map. After repeating these two operations, the feature map from the last pooling layer is followed by several fully-connected MLPs (multi-layer perceptron) whose output is a vector fed to a classifier. Given our own dataset size and the characteristics of low-variance in medical image, we choose to implement the modified AlexNet [1] for our specific application. In our application, we applied consecutive convolution layers in two places, 1) layer 3 and 4 and 2) layer 5 and 6 in order to smooth the noise in the feature map. To avoid overfitting, we apply 70% dropout to fully-connected layer [10]. The architecture of the single network for input image size of 128×128 is depicted in Table 2.

We propose a multi-phase CNN which is a slight variant of the multi-scale CNN architecture used in [11]. Originally, the multi-scale convolutional net contains multiple copies of a single network (not sharing the weights) that are applied to a Laplacian pyramid version of the images. Being slightly different from this, our method applies ROIs of three phases as the input data. More precisely, the three images of the different phases are fed into three independent single networks whose architectures are as described in Table 2. The last pooling layer (layer 5) produces the feature maps of each single network. Then the three feature maps of each phase are concatenated. After the concatenated feature maps are encoded into fully-connected layer (layer 7), the extracted features from the last fully-connected layer are trained using softmax with stochastic gradient descent method.

4 Experiment

In this part, we will fully introduce the dataset used in our application and set-ups for our experiment. Using the algorithm described in Section 3, we have

Table 3. Accuracy of Single Network with Each Phase Input

Method	Pre	50 sec	5 min
Alexnet-256 (from scratch)	27.09	30.26	28.50
Alexnet-256(Fine Tuning)	22.34	26.72	25.83
NIN-256(Fine Tuning)	27.04	30.45	29.15
Ours-256	32.15	34.98	35.63
Ours-128	32.26	31.02	37.21

a comprehensive analysis on the Bosniak classification problem using different strategies of the input data fed to our proposed CNN architecture. We will also show how we deal with unregistered low-variance CT images and discuss how to avoid the overfitting problem caused by relatively small dataset. Furthermore, we compare the proposed method with the traditional Gabor feature extraction plus SVM classification. In this paper, the classification accuracy is measured by counting the number of correctly classified unregistered ROI triplets.

4.1 Single Convolutional Networks

Before implementing our multi-phase CNN, we experiment on single-phase-only data with different single CNN architectures and different input conditions.

First of all, we compare our model with AlexNet and NIN [12] fine-tuned with pre-training on ImageNet dataset. In many cases, fine-tuning on ImageNet pre-trained model performs slightly better than training from scratch. However, we have figured out that the trained features with ImageNet pre-trained model are not suitable in the case of medical image problem as it can be seen in Table 3. Using our method depicted in Table 2, the single phase architecture performed around 5% better than the conventional network architectures. It is interesting to see that single network with 5-min-only input data has slightly better performance than the other phases. It can be seen that 5-min data has more significant features compared to the others.

4.2 Multi-phase Convolutional Networks

In the multi-phase CNN experiment, we conducted three different experiments using two different sizes (128×128 and 256×256) of training data sets. We also compare our method against the features extracted from Gabor filter trained with linear and nonlinear kernel SVM classifiers. Gabor filters have rich applications in image processing [13] [14], especially in feature extraction for the texture analysis. In our experiment, Gabor filters with different frequencies and different orientation directions have been used to extract complex features from cropped CT images. We extract 1,000 dimensional Gabor features per phase and concatenate features from the three phases together. In this way, we have feature vectors with 3,000 dimension for each input data. Training with SVM with linear

Table 4. Accuracy of multiple network with each phase input. The 256 and 128 means the size of input data. Comb means using our data augmentation method of data combination.

Method	Accuracy
Gabor & Linear SVM	28.1
Gabor & Kernel SVM	37.0
Ours-256 & Comb.	47.58
Ours-128 & Comb.	48.9

and kernel method, we obtain 28.1% and 37.0% classification accuracies, respectively. Compared to Gabor features trained with SVM, our proposed method shows 48.9% accuracy, which is 11.9% better.

4.3 Implementation details

The training is completed on Titan-X GPU with caffe. The weights in the networks are initialized randomly with Gaussian distribution. They are then updated by stochastic gradient descent, accompanied by a momentum term of 0.9 and an L_2 weight decay of 1×10^{-5}. The learning rate is 1×10^{-4} and is decreased by a factor of 0.5 in each 10,000 steps. Drop-out with a rate of 0.7 is employed on the fully connected layers (6th and 7th) in the classifier.

5 Conclusion

In this paper, we implemented multi-phase convolutional neural networks to solve the Bosniak classification problem which classifies unregistered CT images into five different Bosniak classes. The performance of the proposed multi-phase CNN shows an improvement over the single-phase CNNs, demonstrating that the high-level features of the multi-phase CNN provide a robust representation of the three-phase input images since the three input images are all useful for the classification.

We also proposed data combination method specially designed for our application. The method efficiently enlarges our training data set and avoids overfitting problem in our application. In this way, the registration of different phase images is not needed. This is particularly important in CT image analysis, where registration is a big challenge due to non-rigid deformations. Although the performance of the multi-phase CNN is better than the single CNNs, it is still not satisfactory and there is more room to be improved. We can infer the reason for the relatively low performance as below: 1) The lack of sufficient training data in Bosniak IIF, III and IV leads to an overfitting in our architecture. 2) the class-imbalance causes a bias on the classification accuracy. 3) because different people annotated the images, the quality of ROI patches are irregular. Despite the issues mentioned above, we want to emphasize that our results can be used as a baseline in the Bosniak classification problem in the future works.

References

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
2. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
3. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
4. Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, E. I. Chang *et al.*, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on.* IEEE, 2014, pp. 1626–1630.
5. D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013.* Springer, 2013, pp. 411–418.
6. C. M. N. R. d. Miranda, C. P. d. M. Maranhão, C. J. J. d. Santos, I. G. Padilha, L. d. P. G. d. Farias, and M. S. d. Rocha, "Bosniak classification of renal cystic lesions according to multidetector computed tomography findings," *Radiologia brasileira*, vol. 47, no. 2, pp. 115–121, 2014.
7. Y. Lee, N. Kim, K.-S. Cho, S.-H. Kang, D. Y. Kim, Y. Y. Jung, and J. K. Kim, "Bayesian classifier for predicting malignant renal cysts on mdct: early clinical experience," *American Journal of Roentgenology*, vol. 193, no. 2, pp. W106–W111, 2009.
8. N. S. Curry, S. T. Cochran, and N. K. Bissada, "Cystic renal masses: accurate bosniak classification requires adequate renal ct," *American Journal of Roentgenology*, vol. 175, no. 2, pp. 339–342, 2000.
9. J. V. Soares, J. J. Leandro, R. M. Cesar Jr, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 9, pp. 1214–1222, 2006.
10. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
11. C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1915–1929, 2013.
12. M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.
13. V. Kyrki, J.-K. Kamarainen, and H. Kälviäinen, "Simple gabor feature space for invariant object recognition," *Pattern recognition letters*, vol. 25, no. 3, pp. 311–318, 2004.
14. M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *Computer Vision–ECCV 2010.* Springer, 2010, pp. 448–461.