Deep Feature Learning for Medical Image Analysis with Convolutional Autoencoder Neural Network

Min Chen, *Senior Member, IEEE*, Xiaobo Shi, Yin Zhang, *Senior Member, IEEE*, Di Wu, *Senior Member, IEEE*, and Mohsen Guizani, *Fellow, IEEE*

Abstract—At present, computed tomography (CT) is widely used to assist disease diagnosis. Especially, computer aided diagnosis (CAD) based on artificial intelligence (AI) recently exhibits its importance in intelligent healthcare. However, it is a great challenge to establish an adequate labeled dataset for CT analysis assistance, due to the privacy and security issues. Therefore, this paper proposes a convolutional autoencoder deep learning framework to support unsupervised image features learning for lung nodule through unlabeled data, which only needs a small amount of labeled data for efficient feature learning. Through comprehensive experiments, it shows that the proposed scheme is superior to other approaches, which effectively solves the intrinsic labor-intensive problem during artificial image labeling. Moreover, it verifies that the proposed convolutional autoencoder approach can be extended for similarity measurement of lung nodules images. Especially, the features extracted through unsupervised learning are also applicable in other related scenarios.

Index Terms—Convolutional autoencoder neural network, lung nodule, feature learning, hand-craft feature, unsupervised learning

1 Introduction

Computed tomography (CT) is an effective approach to diagnose disease, by which the doctor can intuitively examine a patient's body structure and efficiently analyze the possibility of illness. However, each patient often includes hundreds of medical images, so it is a great challenge to process and analyze the massive amount of medical image data. Therefore, intelligent healthcare is an important research direction to assist doctors in harnessing medical big data [1].

Especially, it is difficult to identify the images containing nodules, which should be analyzed for assisting early lung cancer diagnosis, from a large number of pulmonary CT images. At present, the image analysis methods for assisting radiologists to identify pulmonary nodules consist of four steps: i) region of interest (ROI) definition, ii) segmentation [2], iii) hand-crafted feature [3] and iv) categorization. In particular, radiologist has to spend a lot time on checking each image for accurately marking the nodule, which is critical for diagnosis and is a research hotspot in intelligent healthcare [4].

For example, it is proposed to extract texture features for nodules analysis, but it is hard to find effective texture feature parameters [5]. In [6], nodules were analyzed by morphological method through shape, size and boundary, etc. However, this analytical approach is difficult to provide accurate descriptive information. It is because even an experienced radiologist usually give a vague description based on personal experience and understanding. Therefore, it is a challenging issue to effectively extract features for representing the nodules. In [7], [8], convolutional neural network (CNN) is proposed to extract nodule features for avoiding the problems caused by hand-crafted feature extraction, but this approach requires a large number of labeled data for effectively training features.

To address these challenges, we propose a deep learning architecture based on convolutional autoencoder neural network (CANN) for the classification of pulmonary nodules. As shown in Fig. 1, the proposed method first utilizes the original image patch for unsupervised feature learning, with the use of a small amount of labeled data for supervised fine-tuning parameters. Then, the feature representation can be extracted from the input image. For the recognition and classification of lung nodules, the CT images are imported and the patch images are extracted according to the proposed CANN method. Each patch obtains a corresponding verification

- *M. Chen is with Wuhan National Laboratory for Optoelectronics (WNLO) and with School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China. E-mail: minchen@ieee.org.*
- *X. Shi is with the College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China. E-mail: xiaoboshi.cs@qq.com.*
- *Y. Zhang is with the School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China, and the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China. E-mail: yin.zhang.cn@ieee.org.*
- *D. Wu is with the Department of Computer Science, School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China. E-mail: wudi27@sysu.edu.cn.*
- *M. Guizani is with the Electrical and Computer Engineering Department, University of Idaho, MS 1023, Moscow, ID 83844. E-mail: mguizani@ieee.org.*

Manuscript received 1 Jan. 2017; revised 24 Mar. 2017; accepted 10 June 2017. Date of publication 0 . 0000; date of current version 0 . 0000.
(Corresponding author: Xiaobo Shi.)
Recommended for acceptance by S. Yu, P. Mueller, J. Pei, and K. Liu.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TBDATA.2017.2717439

2332-7790 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

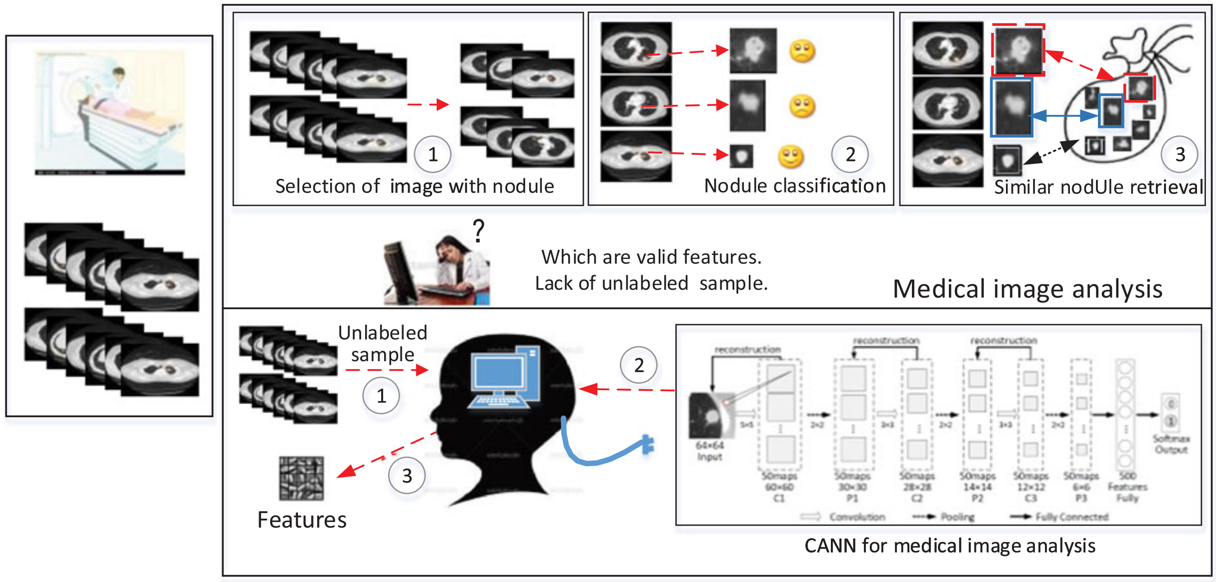


Fig. 1. Illustration of medical image analysis with CANN.

result set for classification after extracting feature through the network structure. The experimental results shows that the proposed method is effective to extract the image features via data-driven approach, and achieves faster labeling for medical data. Specifically, the main contributions of this paper are as follows.

- From the original CT images, the patches are automatically selected for analyzing the existence of nodules, which efficiently reduces the doctor's workload for image viewing and ROI labeling. Due to the small proportion of the pulmonary nodules in the original image, sub-regional learning approach is implemented to accurately extract the pulmonary nodule features.
- CANN is proposed for features learning from large amounts of data, avoiding the uncertainty of hand-crafted features. By the use of the advantages of both unsupervised learning and unlabeled data learning, CANN efficiently addresses the issue of the insufficiency of training data caused by difficulty of obtaining labeled medical images.
- Image features are available to be directly extracted from the raw image. Such an end-to-end approach doesn't use image segmentation method to find the nodules, avoiding the loss of important information which may affect the classification results.
- The unsupervised data driven approach is able to extend to implement in other data sets and related applications.

The remainder of this article is organized as follows. Section 2 briefly introduces the related work. In Section 3, the proposed approach and relational algorithm are presented. Section 4 describes dataset, experimental environments and the produced results. Finally, Section 5 concludes this paper and future work.

2 Relation Work

Feature selection is an essential procedure to obtain extracted features for raw data representation. In recent year, it is a hot research topic in the field of machine learning. Compared with the conventional methods by heuristic approach or manual approach with human-intervention, data-driven feature learning through deep learning exhibits its much higher performance. In [9], Bengio et al. introduce the advantages of deep learning for feature learning, which is a layered architecture like human brain. Through deep learning, the simple features are extracted from the raw data, and then more complex features are learned through multiple layers [10]. Finally, considerable features are generated through multi-iteration learning, in which the parameters, i.e., forward propagation and backward propagation are continuously optimized. Specifically, feature learning is often classified into two categories, i.e., supervised learning and unsupervised learning.

Through supervised learning, the sample data is forwarded from input to the top layer for prediction. By minimizing the value of the cost function between the target value and the predicted value, backward propagation is used to optimizes the connection parameters between each pair of layers. In particular, CNN [11] is a transformation based on neural network, which is used to represent features via supervised learning. CNN is often implemented in image and video analysis [12], [13], speech recognition [14], [15] and text analysis, etc.. Especially in the field of image analysis, CNN has been a great success, such as face recognition [16], scene parsing [17], cell segmentation [18], neural circuit segmentation [19], analysis of images the breast [20], [21] and brain lesion segmentation [22], [23]. For example, a novel 3D-CNN is proposed to categorize in polyp candidates on circulating tumor cell (CTC) [24]. In [7], [25], [26], and evolved convolution networks are proposed to classify the lung nodules through supervised feature learning from medical images. Gao et al. [27] and Schlegl et al. [28] CNN-based methods for classifying the lung tissue according based on lung CT images.

In unsupervised learning approaches, unlabeled data are used to learn features, while a small amount of labeled data are used to fine-tuning the parameters, such as restricted boltzmann machine (RBM) [29], deep belief network [30],

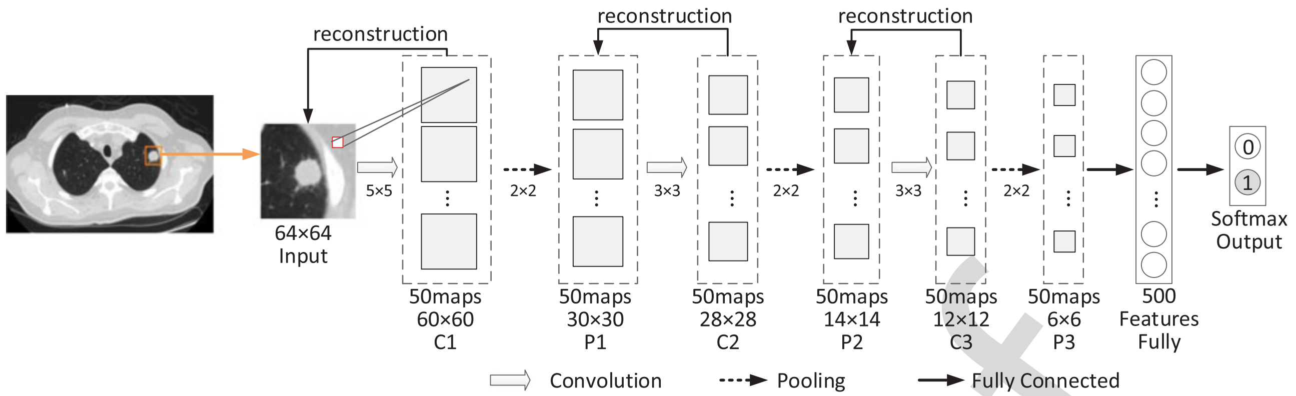


Fig. 2. Convolutional autoencoder neural network for medical image analysis.

autoencoders [31] and stacked autoencoders [32]. Devinder Kumar et al. propose an autoencoder approach for unsupervised feature learning and classification of pulmonary nodules [33]. Kalleberg et al. propose a convolutional autoencoder approach to analyze breast images [34], and Li et al. design a RBM-based approach for lung tissue classification in [35], Tulder et al. analyze lung CT with convolutional restricted boltzmann machines in [36].

In this paper, we propose a convolution autoencoder unsupervised learning algorithm for lung CT features learning and pulmonary nodules classification. Compared with the conventional CNN [7], [25], the proposed scheme is significantly improved that the unsupervised autoencoder and CNN are collaborative to extract the features from the image. Due to the scarcity of medical image labeling, we use a large amount of unlabeled data for training the feature learning network, while only a small amount of labeled data are used to fine-tuning the network. Moreover, because the workload for labeling ROI is high and the pulmonary nodules are difficult to be recognized, the raw CT images are divided into small patch areas for training the network.

3 The Proposed Convolutional Autoencoder Neural Network

The patch divided from the raw CT image is input to CANN for the purpose of learning the feature representation, which is used for classification. The parameters of convolution layers in CNN are determined by autoencoder unsupervised learning, and a small amount of labeled data are used for fine tuning the parameters of CANN and training the classifier. This section describes the proposed CANN structure, parameter settings and training methods, etc.

Specifically, the patch divided from the original CT image can be represented as $x{\in}X$, $X{\subset}\mathbb{R}^{m\times d\times d}$, where m represents the number of input channel, and $d{\times}d$ represents the input image size. The labeled data are represented as $y{\in}Y$, $Y{\subset}\mathbb{R}^{n}$, where n represents the number of output classification. Through the proposed model, it is expected to deduce the hypothesis function from the training, i.e., $f: X{\mapsto}Y$ and the set of parameters θ .

In the proposed model, the hypothesis function f based on deep learning architecture consists of multiple layers, which is not a direct mapping from X to Y. Specifically, the first layer L_1 receives the input image x, and the last layer L_N is the output layer. Middle layers include three convolution layers, three pooling layers and one fully connected layer. The structure of the proposed CANN is shown in Fig. 2.

In this paper, the training data include two datasets, i.e., the unlabeled dataset $UD = \{x|x{\in}X\}$ and the labeled dataset $D = \{x, y|x \in X, y \in Y\}$. In particular, UD is used for unsupervised training, while D is used for supervised fine tuning and classifier training.

3.1 Standard Autoencoder

Supervised approach is available for data-driven features learning, in which the connection weights are updated through forward and backward propagation algorithms. Compared with supervised approach, unsupervised approach can directly receive unlabeled input data, which effectively reduce the workload for labeling data.

In this paper, we propose an autoencoder method for unsupervised learning. Autoencoder extract output data to reconstruct input data and compare it with original input data. After numerous times of iterations, the value of cost function reaches its optimality, which means that the reconstructed input data is able to approximate the original input data to a maximum extent.

The input data I represents m-dimension vector $I \in R^m$. The output data $code$ is a n-dimension vector $code \in R^n$. Standard autoencoder includes three main steps:

1) Encode: Convert input data I into $code$ of the hidden layer by $code = f(I) = \sigma(w \cdot I + b)$, where $w \in R^{m\times n}$ and $b \in R^n$. σ is an activate function, the sigmod or hyperbolic tangent function can be used.
2) Decode: Based on the above code, reconstruct input value O' by equation $O' = f'(code) = \phi(\hat{w}{\cdot}code + \hat{b})$, where $\hat{w} \in R^{n\times m}$ and $\hat{b} \in R^m$. The activate function ϕ is the same as σ.
3) Calculate square error $L_{recon}(I, O') = \|I - O'\|^2$, which is the error cost function. Error minimization is achieved by optimizing the cost function:

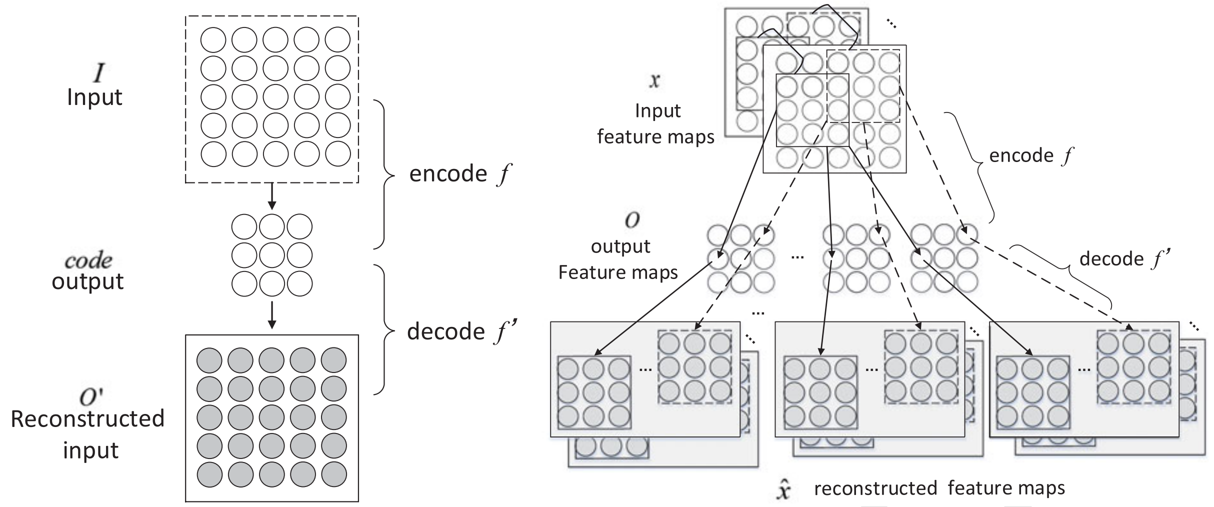
$$J(\theta) = \sum_{I\in D} L(I, f'(f(I)) \qquad \theta = \{w, \hat{w}, b, \hat{b}\}. \tag{1}$$

Fig. 3a shows the unsupervised feature learning with autoencoder.

3.2 Convolution Autoencoder

Convolution autoencoder combines the local convolution connection with the autoencoder, which is a simple

IEEE Proof



(a) Unsupervised feature learning with autoencoder.

(b) Learning multi tunnel feature of convolutional layer with autoencoder

Fig. 3. Two architecture of unsupervised feature learning.

operation to add a reconstruction input for the convolution operation. The procedure of the convolutional conversion from feature maps input to output is called convolutional decoder. Then, the output values are reconstructed through the inverse convolutional operation, which is called convolutional encoder. Moreover, through the standard autoencoder unsupervised greedy training, the parameters of the encode and decode operation can be calculated.

The operation in the convolutional autoencoder lay is illustrated in Fig. 3b, where $f(.)$ represents the convolutional encode operation and $f'(.)$ represents the convolutional decode operation. Input feature maps $x \in R^{n\times l\times l}$, which are obtained from the input layer or the previous layer. It contains n feature maps, and the size of each feature map is $l \times l$ pixels. The convolutional autoencoder operation includes m convolutional kernels, and the output layer output m feature maps. When the input feature maps produced from the input layer, n represents the number of input channels. When the input feature maps from the previous layer, n represents the number of output feature maps from the previous layer. The size of convolutional kernel is $d \times d$, where $d \leqslant l$.

$\theta = \{W, \hat{W}, b, \hat{b}\}$ represents the parameters of convolutional autoencoder layer need to be learned, while $b \in R^m$ and $W = \{w_j, j = 1, 2, \ldots, m\}$ represent the parameters of convolutional encoder, where $w_j \in R^{n\times l\times l}$ is defined as a vector $w_j \in R^{nl^2}$. And $\hat{W} = \{\hat{w}_j, j = 1, 2, \ldots, m\}$ and $\hat{b}$ represent the parameters of convolutional decode, where $\hat{b} \in R^{nl^2}$, $\hat{w}_j \in R^{1\times nl^2}$.

First, the input image is encoded that each time a $d \times d$ pixels patch $x_i, i{=}1, 2, \ldots, p$, is selected from the input image, and then the weight w_j of the convolution kernel j is used for convolutional calculation. Finally, the neuron value $o_{ij}, j = 1, 2, \ldots, m$ is calculated from the output layer

$$o_{ij} = f(x_i) = \sigma\big(w_j \cdot x_i + b\big). \tag{2}$$

In Eq. (2), σ is a non-linear activation function, often including three fuctions, i.e., the sigmod function, the hyperbolic tangent function, and the rectified linear function (Relu). And Relu is implemented in this paper.

$$\mathrm{Relu}(x) = \begin{cases} x & x \geqslant 0 \\ 0 & x < 0. \end{cases} \tag{3}$$

Then o_{ij} output from the convolutional decode is encoded that x_i is reconstructed via o_{ij} for generated $\hat{x}_i$

$$\hat{x}_i = f'(o_{ij}) = \phi(\hat{w}_i \cdot o_{ij} + \hat{b}). \tag{4}$$

$\hat{x}_i$ is generated after each convolutional encode and decode. We get P patch obtained from the reconstruction operation with $d \times d$. We use the mean square error between the original patch of input image $x_i, (i{=}1, 2, \ldots p)$ and the reconstructed patch of image $\hat{x}_i, (i{=}1, 2, \ldots p)$ as the cost function. Furthermore, the cost function is described in Eq. (5), and the reconstruction error is described in Eq. (6)

$$J_{CAE}(\theta) = \frac{1}{p}\sum_{i=1}^{p} L[x_i, \hat{x}_i] \tag{5}$$

$$L_{CAE}[x_i, \hat{x}_i] = \|x_i - \hat{x}_i\|^2 = \|x_i - \phi(\sigma(x_i))\|^2. \tag{6}$$

Through stochastic gradient descent (SGD), the weight and error are minimized, and the convolutional autoencoder layer is optimized. Finally, the trained parameters are used to output the feature maps which are transmitted to the next layer.

3.3 Pooling

The proposed CANN is similar to the common CNN, where the convolutional layer is connected to the pooling layer. Especially, in CANN after the convolutional autoencoder layer is the max pooling layer, as shown in

$$o_j^i = \max\Big(x_j^i\Big). \tag{7}$$

Each input feature map is divided into n no-overlapping regions according to the size of the pooling region, where x_j^i represents the ith region of the jth feature map, and o_j^i

represents the ith neuron of the jth output feature map. The number of input feature maps is equal to the number of output feature maps in the pooling layer. Neurons in the feature map can be reduced after the pooling operation, thus the computational complexity is also reduced.

3.4 Cost Function

As shown in Fig. 2, softmax classification layer, which is used for classification according to the features, is after multiple convolutional autoencoder layers, max pooling layer and full connected layer. In this paper, the lung CT images are divided into two categories. Specifically, $\hat{y}_i$ from the classifier represents the probability of nodules and no nodules

$$\hat{y}_i = \frac{e^{(o_i)}}{\sum_{k=1}^{2} e^{(o_k)}}, i = 0, 1. \tag{8}$$

$o_i = \sigma(\sum_{t=1}^{T} x^f \cdot w^f + b^f)$ represents the T output features x^f generated through the full connected layer, where w^f and b^f represent the weight and error respectively, σ represents the nonlinear function sigmoid.

Furthermore, in the supervised training network, the cost function is cross entropy L, as shown in Eq. (9), and SGD is implemented for minimizing L. Where y is the label of sample data. Specifically, 0 and 1 represents no nodules and nodules respectively

$$L = -(y \log\ \hat{y}_0 + (1-y) \log\ \hat{y}_1). \tag{9}$$

3.5 Training Parameters

3.5.1 Convolution Autoencoder

$N = 50{,}000$ unlabeled samples are used to train the autoencoder through unsupervised learning at the convolutional layer, the gradient is calculated through the cost function Eq. (5), and the parameters are optimized through SGD. Specifically, every 100 samples are included in a mini batch, and the number of iterations for each batch is 50, so the number of iterations per layer is $50 \times N/100$. Moreover, the number of channels must be set in the convolutional encoder Eq. (2) and convolutional decode Eq. (4) respectively.

3.5.2 Full Connected Layer and Classifier

The input of the full connected layer is from the last pooling layer. In particular, the features are represented through 500 neurons, which are connected to the softmax classifier. The parameters are supervised trained at full connected layer and softmax classifier. There are 1,800 labeled data for classification training, and each mini batch including 50 samples are used for parameter optimization via 500 SGD-based iteration.

3.5.3 Algorithm of Training CANN

The training in CANN is based on the work in [38] and [31], including unsupervised training and supervised fine-tunning, which are described in Algorithms 1 and 2.

4 Experiment and Results

4.1 Dataset

The experimental data are collected from a second-class hospital in China, including about 4,500 patients' lung CT images from 2012 to 2015.

Algorithm 1. Unsupervised Training CANN

1: UD: given unlabeled dataset;
2: desired number of convolution layer and pooling layer;
3: Initialize all weight matrices and bias vectors randomly convolution layer and pooling layer;
4: $i \leftarrow 1$;
5: **if** $i == 1$ **then**
6:     The input of C_i is UD;
7: **else**
8:     The input of C_i is the output of P_i;
9: **end if**
10: Greedy layer-wise training C_i;
11: Find parameters for C_i by cost function;
12: Use the output of C_i as the input of the P_i;
13: Max pooling operator;
14: **if** $i < N$ **then**
15:     goto line 5;
16: **end if**

Algorithm 2. Supervised Fine-Tunning CANN

1: Initialize all weight matrices and bias vectors randomly of fully connect layer;
2: Given labeled dataset D as the input of network;
3: Use BP algorithm with SGD parameter optimization method tune the network's parameters in top-down direction;

Doctors identify the ROI for each nodule (the total number of nodules is around 1,800) by some mark, based on which a centering the marked area, 64×64 region is segmented as the patch of nodule. Specifically, the data are divided into three datasets:

- D1: unlabeled data for unsupervised training, which contains 50,000 64×64 patches. These small patches are randomly captured from all the patients' lung CT slides in this hospital.
- D2: labeled data for classification, which include 3,662 64×64 patches. They are labeled by two professional radiologists. In the labeled data, 1,754 patches contain nodules, while the other 1,908 patches are normal ones.
- D3: labeled data for similarity judgement contain 500 pairs of labeled patches. The images are marked by two doctors, and the similarity is generated according to the intersection of the labeled results. The range of similarity is from 1 to 3, where 3 represents the highest level of similarity and 1 means the lowest similarity. We deleted 61 samples with similarity of 2 , i.e., those with the middle level of similarity are deleted. Finally, 214 samples with similarity of 1 are labeled as "0" (i.e., they are not similar), while 225 samples with similarity of 3 are labeled as "1" (i.e., they are similar).

4.2 Convolutional Architecture

In this paper, we propose two kinds of CANN, i.e., C-CANN for classifying as shown in Fig. 2, S-CANN for similarity check in Fig. 4. In particular, S-CANNs can be regarded as two parallel CAANS with the same structures and parameters.

IEEE Proof

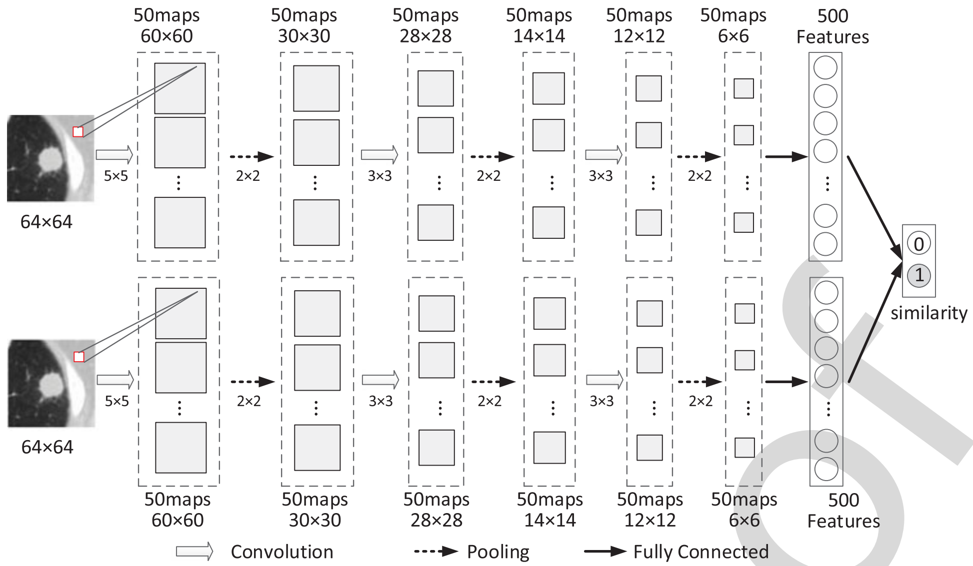


Fig. 4. S-CANN for estimating image similarity.

The C-CANN consists of 3 groups of connections between convolutional layer and pooling layer, followed by a full connected layer and a classifier, i.e., eight layers in the structure. The network parameters are list as follows:

- Input: 64×64 patch captured from CT image.
- C1: convolution kernel is 5×5, the step 1, the number of convolution kernel is 50, non-linear function is Relu.
- P1: max pooling is used, the size of pooling area is 2×2.
- C2: convolution kernel is 3×3, the step 1, the number of convolution kernel 50, non-linear function is Relu.
- P2: max pooling is used, the size of pooling area 2×2.
- C3: convolution kernel is 3×3, the step 1, the number of convolution kernel 50, non-linear function: Relu.
- P3: max pooling with 2×2 size of pooling area .
- Full: fully connected layer, 500 neurons.
- Output: softmax classifier, 2 classes.

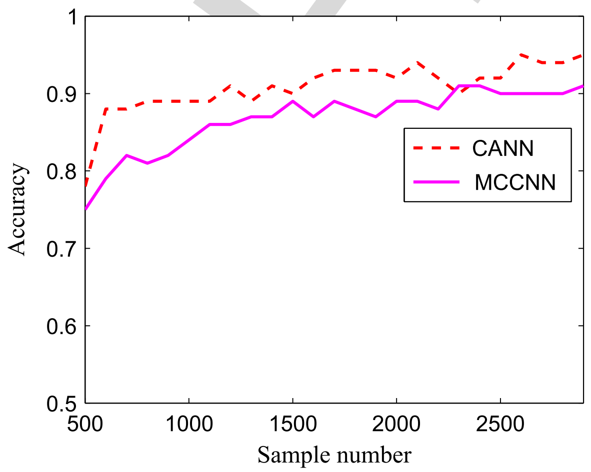


Fig. 5. The impact of training sample number on classification accuracy of CANN and MCNN.

S-CANN also includes eight layers which are the same as those in C-CANN. Through S-CANN, the features are extracted from a pair of images to be compared for the calculation of similarity by two identical C-CANNs respectively.

4.3 Classification

4.3.1 *Impact of Sample Number*

Fig. 5 illustrates the impact of the number of training sample on the classification accuracy of CANN and MCNN. The results shows that the performance is optimal when the number reaches 2,900 for both CANN and MCCNN methods. When the number is around 700 or 800, CANN starts to ourperform MCCNN. With the increase of the number to 1,500 or 1,600, CANN exhibits a tendency.

4.3.2 *Performance Comparison of Classification*

Convolutional neural network for learning the lung nodule image feature is similar to common image feature learning. Both CNN and conventional learning use labeled dataset, and learn the network parameters between each layer from the input layer to the output layer by the use of forward and backward propagation methods. MCNN is a variant of CNN. Its difference from CNN is that the pooling operation adopts multiple methods with different pooling area and fuses multi-scale pooling results as the output of pooling layer.

We compare the classification performance of (CANN), autoencoder (AE) [33], convolutional neural network (CNN) and MCCNN [25] with dataset D2, the results are shown in Table 1 and the Rate of Change (ROC) is shown in Fig. 6.

TABLE 1
Comparison of Different Method's Classification Performance on D2

Method	accuracy	precision	recall	F1	AUC
CANN	**95.00%**	**95.00%**	**95.00%**	**95.00%**	**0.98**
AE [33]	77.00%	76.00%	77.00%	77.00%	0.83
CNN	89.00%	88.00%	90.00%	89.00%	0.95
MCCNN [25]	91.00%	91.00%	90.00%	91.00%	0.97

IEEE Proof

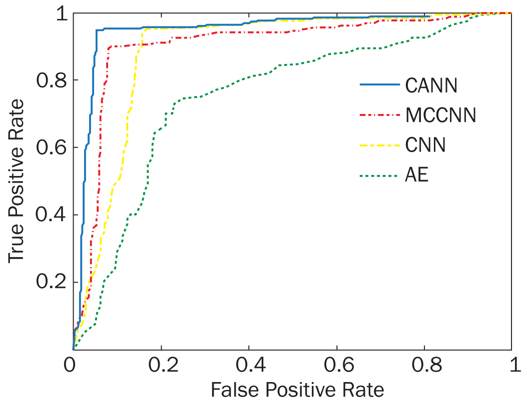


Fig. 6. ROC of classification on D2.

The CNN and MCCNN methods use the same convolutional architecture as CANN. The accuracy, precision, recall, F1 and AUC of proposed method are 92, 91, 91, 91 percent and 0.97 respectively. For AE method, we use the same unlabeled training database and test it on the same database, and full connected layer has 1,024 neurous. The accuracy, precision, recall, F1 and AUC of AE are 77, 76, 77, 77 and 83 percent respectively. Because unsupervised method can not learning optimal feature, its performance is lower than CANN. The five evaluation index of CNN method are 89, 88, 90, 89, 95 percent respectively. The performance index of MCNN method are 91, 91, 90, 91 and 97 percent respectively. The classification performance of both CNN and MCCNN method are lower than the proposed method. The evaluation verifies that the combination of unsupervised feature learning and supervised fine-tunning can significantly improve performance.

4.4 Similarity Check

Image similarity judgment is used to retrieve the similar nodules for providing reference to doctors. Similarity judgment and nodules classification have to consider several features, such as nodule's morphology, density, size, edge, etc. The full connected layer network and similarity judgment layer are trained through unsupervised approach, while five-fold cross validation are trained by using dataset D3 with supervised approach. CANN performance for image similarity and classification, such as accuracy, precision, recall, F1 and etc., are shown in Fig. 7. The evaluation verifies that unsupervised feature learning and supervised fine-tunning with a small training set can obtain better performance.

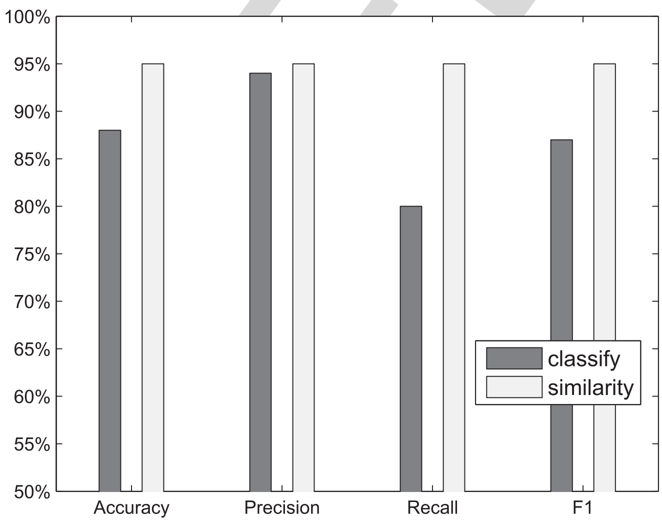


Fig. 7. The image similarity performance with D4.

5 Conclusion

In this paper, we investigate two representative approaches to assist CT image analysis. The approach based on segmentation and hand-craft-features is time consuming and labor-intensive, while the data-driven approach is available to avoid the loss of important information in nodule segmentation. However, due to the scarcity of labeled medical data, these two approaches are not practicable. Hence, this paper proposes a CANN-based approach for data-driven feature learning, in which the network is unsupervised trained with a large amount of unlabeled patch and a small amount of labeled data is used for fine-tuning the network structure. The proposed approach is applied for lung nodule recognition, classification and similarity check, which significantly solves the issues of time consuming for ROI labeling and inadequate labeled data. Compared with other data-driven approaches, it verifies that the proposed method is superior through comprehensive experiments. Moreover, it proves that the system performance and feasiblity may be affected by the quality of data, because the role of expert is ignored. Therefore, we will combine domain knowledge and data-driven feature learning in our future work.

Acknowledgments

Prof. Min Chen's work was supported by the National Natural Science Foundation of China (Grant No. 61572220). Di Wu's work was supported by the National Science Foundation of China under Grant 61572538, the Fundamental Research Funds for the Central Universities under Grant 17LGJC23.

References

[1] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, "Wearable 2.0: Enable human-cloud integration in next generation healthcare system," *IEEE Commun.*, vol. 55, no. 1, pp. 54–61, Jan. 2017.

[2] T. Messay, R. C. Hardie, and T. R. Tuinstra, "Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset," *Med. Image Anal.*, vol. 22, no. 1, pp. 48–62, 2015.

[3] Y. Balagurunathan, et al., "Reproducibility and prognosis of quantitative features extracted from CT images," *Translational Oncology*, vol. 7, no. 1, pp. 72–87, 2014.

[4] M. Chen, J. Yang, Y. Hao, S. Mao, and K. Hwang, "A 5G cognitive system for healthcare," *Big Data Cognitive Comput.*, vol. 1, no. 1, 2017, Art. no. 2, doi: 10.3390/bdcc1010002.

[5] F. Han, et al., "A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng.*, 2015, pp. 14–18.

[6] T. W. Way, et al., "Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features," *Med. Physics*, vol. 36, no. 7, pp. 3086–3098, 2009.

[7] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2015, pp. 588–599.

[8] S. Hamidiana, B. Sahinerb, N. Petrickb, A. Pezeshkb, and M. Spring, "3D convolutional neural network for automatic detection of lung nodules in chest," *SPIE Med. Imag.*, vol. 10134, pp. 101340901–101340906, 2017.

IEEE Proof

[9] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[10] K. Hwang and M. Chen, *Big Data Analytics for Cloud/IoT and Cognitive Computing*. Chichester, U.K.: Wiley, 2017.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[12] L. Zhou, "Mobile device-to-device video distribution: Theory and application," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 12, no. 3, pp. 1253–1271, 2015.

[13] L. Zhou, "QoE-driven delay announcement for cloud mobile media," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 1, pp. 84–94, Jan. 2017.

[14] M. Chen, P. Zhou, and G. Fortino, "Emotion communication system," *IEEE Access*, vol. 5, pp. 326–337, 2017, doi: 10.1109/ACCESS.2016.2641480.

[15] L. Zhou, "On data-driven delay estimation for media cloud," *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 905–915, May 2016.

[16] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1891–1898.

[17] C. Farabet, C. Couprie, L. Najman, and Y. LeCun , "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

[18] H. Su, Z. Yin, S. Huh, T. Kanade, and J. Zhu, "Interactive cell segmentation based on active and semi-supervised learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 762–777, Mar. 2016.

[19] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Advances Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.

[20] P. Fonseca, et al., "Automatic breast density classification using a convolutional neural network architecture search procedure," *SPIE Med. Imag.*, vol. 9414, pp. 2801–2808, 2015.

[21] A. R. Jamieson, K. Drukker, and M. L. Giger, "Breast image feature learning with adaptive deconvolutional networks," *SPIE Med. Imag.*, vol. 8315, pp. 601–613, 2012.

[22] K. Kamnitsas, et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 37, pp. 61–78, 2017.

[23] A. Patel, S. van de Leemput, M. Prokop, B. Ginneken, and R. Manniesing, "Automatic cerebrospinal fluid segmentation in non-contrast CT images using a 3D convolutional network," *SPIE Med. Imag.*, vol. 10134, pp. 201–206, 2017.

[24] T. Uemura, et al., "Classification of polyp candidates on CTC based on 3D-CNN," in *Proc. Int. Forum Med. Imag. Asia*, 2017, pp. 103–105.

[25] W. Shen, et al., "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognit.*, vol. 61, pp. 663–673, 2016.

[26] X. Li, Y. Kao, W. Shen, X. Li, and G. Xie, "Lung nodule malignancy prediction using multi-task convolutional neural network," *SPIE Med. Imag.*, vol. 10134, pp. 241–247, 2017.

[27] M. Gao, et al., "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Comput. Methods Biomechanics Biomed. Eng.: Imag. Vis.*, pp. 1–6, 2016, doi: 10.1080/21681163.2015.1124249.

[28] T. Schlegl, S. M. Waldstein, W.-D. Vogl, U. Schmidt-Erfurth, and G. Langs, "Predicting semantic descriptions from medical images with convolutional neural networks," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2017, pp. 437–448.

[29] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[30] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[31] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. 19th Int. Conf. Neural Inf. Process. Syst.*, 2007, vol. 19, pp. 153–160.

[32] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.

[33] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Proc. 12th IEEE Conf. Comput. Robot Vis.*, 2015, pp. 133–138.

[34] M. Kallenberg, et al., "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1322–1331, May 2016.

[35] Q. Li, W. Cai, and D. D. Feng, "Lung image patch classification with automatic feature learning," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 6079–6082.

[36] G. Tulder and M. Bruijne, "Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1262–1272, May 2016.

[37] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Int. Conf. Artif. Neural Netw.*, 2011, pp. 52–59.

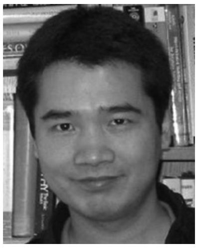
[38] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

Min Chen is a full professor in the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST) since Feb. 2012. He is chair of IEEE Computer Society (CS) Special Technical Communities (STC) on Big Data. He was an assistant professor in the School of Computer Science and Engineering, Seoul National University (SNU). He worked as a post-doctoral fellow in the Department of Electrical and Computer Engineering, University of British Columbia (UBC) for three years. Before joining UBC, he was a post-doctoral fellow at SNU for one and half years. He received Best Paper Award from QShine 2008, IEEE ICC 2012, ICST IndustrialIoT 2016, and IEEE IWCMC 2016. He serves as editor or associate editor of the *Information Sciences*, the *Information Fusion*, and the *IEEE Access*, etc. He is a guest editor of the *IEEE Network*, the *IEEE Wireless Communications*, and the *IEEE Transactions on Services Computing*, etc. He is co-chair of IEEE ICC 2012-Communications Theory Symposium, and co-chair of IEEE ICC 2013-Wireless Networks Symposium. He is general co-chair for IEEE CIT-2012, Tridentcom 2014, Mobimedia 2015, and Tridentcom 2017. He is Keynote Speaker for CyberC 2012, Mobiquitous 2012, Cloudcomp 2015, IndustrialIoT 2016, and The 7th Brainstorming Workshop on 5G Wireless. He has more than 280 paper publications, including more than 150 SCI papers, more than 60 IEEE Trans./Journal papers, 13 ISI highly cited papers, and seven hot papers. He has published two books: *OPNET IoT Simulation* (2015) and *Big Data Inspiration* (2015) with HUST Press, a book on big data: *Big Data Related Technologies* (2014) and a book on 5G: *Cloud Based 5G Wireless Networks* (2016) with Springer Series in Computer Science. His latest book (co-authored with Prof. Kai Hwang), entitled *Big Data Analytics for Cloud/IoT and Cognitive Computing* (Wiley, U.K.) is in press to appear in 2017. His Google scholars citations reached more than 9,200 with an h-index of 47. His top paper was cited more than 960 times. He got IEEE Communications Society Fred W. Ellersick Prize in 2017. His research focuses on cyber physical systems, IoT sensing, 5G networks, mobile cloud computing, SDN, healthcare big data, medica cloud privacy and security, body area networks, emotion communications and robotics, etc. He is a senior member of the IEEE since 2009

Xiaobo Shi received the PhD degree from the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), China. She is an associate professor of the College of Computer and Information Engineering, Henan Normal University, China. Her research interests include healthcare big data, machine learning, data analysis, congnitive computing, and Internet of Things.

IEEE Proof

Yin Zhang is an assistant professor of the School of Information and Safety Engineering, Zhongnan University of Economics and Law (ZUEL), China. He is an Excellent Young Scholar at ZUEL. He is vice-chair of IEEE Computer Society Big Data STC. He was a poster-doctoral fellow in the School of Computer Science and Technology, Huazhong University of Science and Technology, China. He has published more than 50 prestigious conference and journal papers. His research interests include intelligent service computing, big data and social network, etc. He is a senior member of the IEEE since 2016.



Di Wu is a professor and assistant dean in the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include broadly in the areas of networking and distributed systems, cloud computing, big data analytics, network security, etc. He is a senior member of the IEEE, a member of the ACM and the Sigma Xi.

Mohsen Guizani (S'85-M'89-SM'99-F'09) received the BS (with distinction) and MS degrees in electrical engineering, the MS and PhD degrees in computer engineering from Syracuse University, Syracuse, New York, in 1984, 1986, 1987, and 1990, respectively. He is currently a professor and the ECE Department chair with the University of Idaho. Previously, he served as the associate vice president of Graduate Studies, Qatar University, chair of the Computer Science Department, Western Michigan University, and chair of the Computer Science Department, University of West Florida. He also served in academic positions with the University of Missouri-Kansas City, University of Colorado-Boulder, Syracuse University, and Kuwait University. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He currently serves on the editorial boards of several international technical journals and the founder and the editor-in-chief of *Wireless Communications and Mobile Computing Journal* (Wiley). He is the author of nine books and more than 450 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE journals and magazines. He also served as a member, chair, and general chair of a number of international conferences. He received the teaching award multiple times from different institutions as well as the best Research Award from three institutions. He was the chair of the IEEE Communications Society Wireless Technical Committee and the chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker from 2003 to 2005. He is a fellow of the IEEE and a senior member of the ACM.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.

IEEE Proof