WILEY | Hindawi

Research Article

Feature Representation Using Deep Autoencoder for Lung Nodule Image Classification

Keming Mao [ID],[1] Renjie Tang,[2] Xinqi Wang,[1] Weiyi Zhang,[1] and Haoxiang Wu[1]

[1]*College of Software, Northeastern University, Shenyang, Liaoning Province 110004, China*
[2]*China Mobile Group Zhejiang Co., Ltd., Hanzhou, Zhejiang Province 310016, China*

Correspondence should be addressed to Keming Mao; maokm@mail.neu.edu.cn

This paper focuses on the problem of lung nodule image classification, which plays a key role in lung cancer early diagnosis. In this work, we propose a novel model for lung nodule image feature representation that incorporates both local and global characters. First, lung nodule images are divided into local patches with Superpixel. Then these patches are transformed into fixed-length local feature vectors using unsupervised deep autoencoder (DAE). The visual vocabulary is constructed based on the local features and bag of visual words (BOVW) is used to describe the global feature representation of lung nodule image. Finally, softmax algorithm is employed for lung nodule type classification, which can assemble the whole training process as an end-to-end mode. Comprehensive evaluations are conducted on the widely used public available ELCAP lung image database. Experimental results with regard to different parameter setting, data augmentation, model sparsity, classifier algorithms, and model ensemble validate the effectiveness of our proposed approach.

1. Introduction

Lung cancer is one of the most deadly diseases around the world, with about 20% among all cancers in 2016. The 5-year cure rate is only 18.2% in spite of great progress in recent diagnosis and treatment. It is noted that if the patient can be accurately diagnosed in the early stage and suitable treatment can be implemented, there will be a greater chance for their survival [1]. Therefore, it is of great significance to do research about early diagnosis of lung cancer. Computed Tomography (CT) is currently the most popular method among lung cancer screening technologies [2]. CT can generate high resolution data, which enable small/low-contrast lung nodules effectively detected compared with conventional radiography methods. According to the report of National Lung Screening, low-dose CT scan reduces lung cancer mortality by a rate of 20% [3]. Due to the fact that traditional lung cancer diagnosis only relies on professional experts, two main drawbacks will be caused: (1) subjectivity, different doctors have different diagnostic results for the same CT scan image; (2) huge workload, reading CT images consumes much time

and effort. This makes the efficiency inevitably weakened. With the development of computer vision technology, some benefits are brought for medical image process and analysis. Its efficiency and stability provide auxiliary help for doctors with automatically or semiautomatically pattern.

During the last two decades, a number of researchers have been devoted to the development of medical image process and analysis with computer vision and machine learning technologies especially for lung disease diagnosis [4]. Among these studies, lung nodule image classification has attracted much attentions for it is a key step for lung cancer analysis. The lung nodule is characterized by its appearance and relation between surrounding regions. Usually, the lung nodule can be classified into 4 types [5], as shown in Figure 1. To be specific, Figures 1(a)–1(d) demonstrate nodule types W, V, J, and P, respectively, where

> W is well-circumscribed nodule located centrally in the lung without any connection to other structures;
>
> V is vascularized nodule that is also central in the lung but closely attached to the neighbouring vessels;
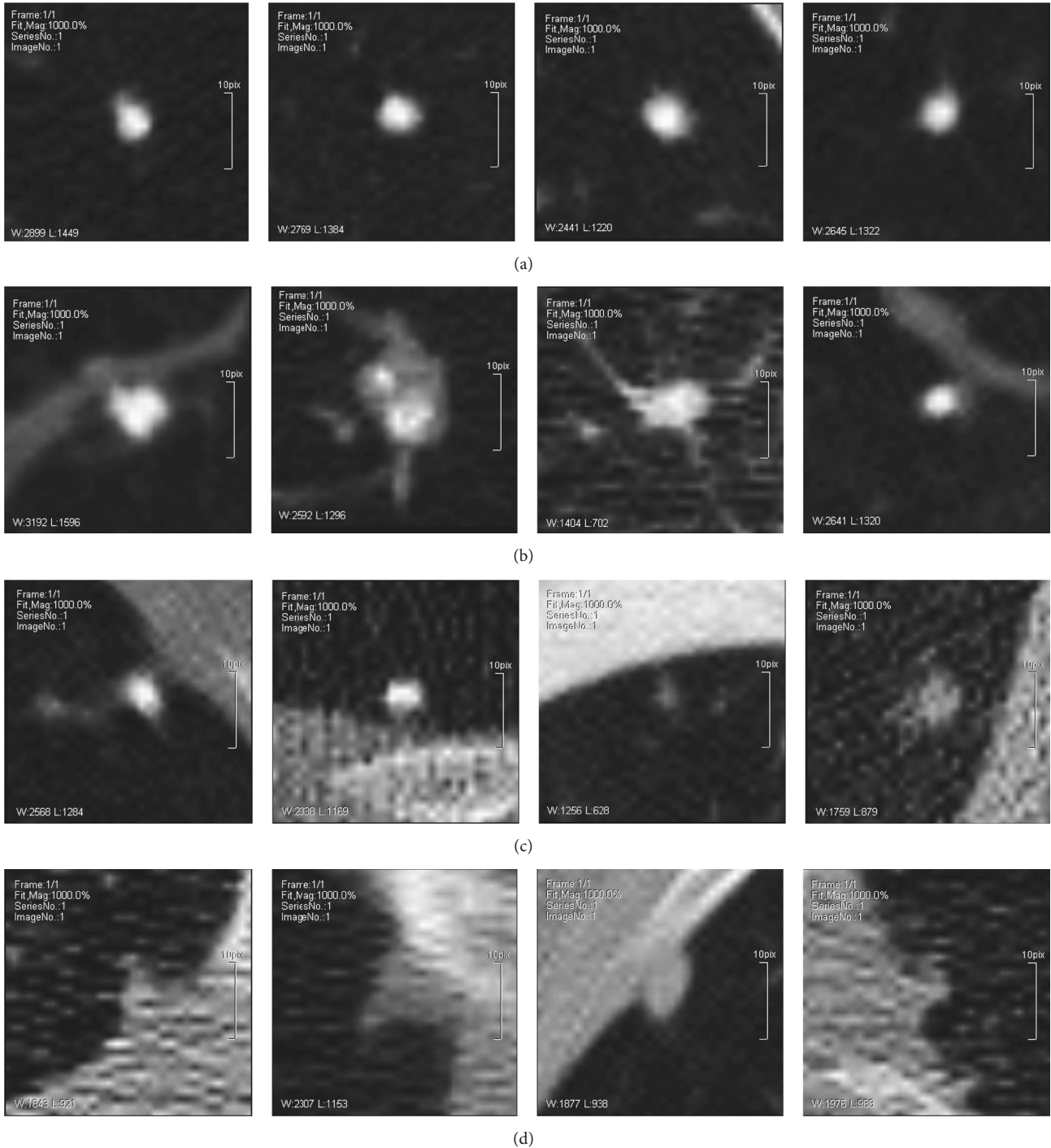
(a)

(b)

(c)

(d)

Figure 1: Demonstration of four types lung nodule image samples (cropped from images in [6]).

J is juxtapleural nodule that has a large portion connected to the pleura;

P is pleural-tail nodule that is near the pleural surface connected by a thin tail.

Lung nodule CT image classification includes two main steps. First, feature extraction and representation use segmentation, filter, and statistical method to describe feature of lung nodule based on shape and texture. Second, classifier design constructs classifier based on supervised or unsupervised machine learning method. However, these methods belong to the fields of traditional image processing and machine learning, which can only characterize the abstraction of lung nodule image in a shallow layer and make the research at low level. As a result, the complex structure of lung nodule makes the classification still a challenging problem. This paper proposes a novel model for lung nodule feature representation and classification. The model considers both

local feature and global feature. Lung nodule CT images are first divided into local patches with Superpixel, and each patch is associated with a relatively intact tissue. Then local feature is extracted from each patch with deep autoencoder. Visual vocabulary is constructed with local features. Global representation is constructed by bag of visual word (BOVW) model and classifier is trained using softmax algorithm. The main contributions of our work are as follows: (i) a novel feature representation model for lung nodule image classification is proposed. Local and global features are constructed by unsupervised deep autoencoder and BOVW model, and (ii) comprehensive evaluations are conducted, and performance analyses are reported from multiple aspects.

The structure of this paper is organized as follows. Related works are introduced in Section 2. Section 3 gives the framework. Local feature representation and global feature representation are given in Sections 4 and 5. Section 6 presents the classifier model. Experimental evaluations are shown in Section 7. Section 8 concludes this paper.

2. Related Works

Many studies have reported the classification of lung nodule in CT image. Some representative works are introduced in this section. Many researches designed feature based on texture, shape, and intensity of lung nodule image. A feature extraction method based on morphological and shape of lung nodule was designed in [7]. A subclass local constraint based method is proposed in [8]. Spectral clustering and approximate affine matrix were used to construct data subclass and each subclass was used as reference dictionary. The testing image was represented by sparse dictionary. Finally, two metrics based on approximation and distribution degree were merged. In [9], spectrum was sampled around center of lung nodule and feature was constructed by FFT. All features were used to construct the dictionary, and then BOVW mode was used to represent the feature of lung nodule. The Haralick texture feature based on spatial direction distribution was proposed in [10], and SVM was used as classifier finally. Ridge direction information was adopted in [11]. Local random comparison method was used to construct the feature vector, and then random forest was used as classifier. Reference [12] first labeled nodule as solid, part-solid, and nonsolid. Then shape based feature was extracted and kNN was used train the classifier. Reference [13] adopted smoothness and irregularity of lung nodule as feature representation. Texture, shape, statistics, and intensity were extracted as feature representation and ANN was used as classifier in [14]. An eigenvalue of Hessian matrix based feature extraction method is adopted in [15], and AdaBoost was used as classifier. Reference [16] used rotation-invariant second-order Markov-Gibbs random field to model the intensity distribution of lung nodule, and Gibbs energy was used to describe the feature vector. Finally, Bayes classifier was constructed. LDA and 3D lattice were used to construct the mapping between lung nodule image and feature representation in [17]. Reference [18] used topology histogram to represent feature vector of lung nodule, and discriminant and K-means were used as classifier. These methods

represent the lung nodule image feature in relatively low level, and they lack sophisticated extraction. On the other hand, these methods need heavy participation of professional expert and they have less generality.

Some well-engineered feature extraction and representation methods widely used in computer vision domain were adopted in lung nodule image classification. Reference [22] proposed a method based on texture and context of lung nodule. Lung images are divided into nodule level and context level; then SIFT, LBP, and HOG features were extracted. Reference [19, 23] divided lung nodule as foreground and background with graph model and conditional random field. Then SIFT was used to extract feature and SVM was used as classifier. In [24], SIFT feature was first extracted. Then PCA and LDA were used for dimension reduction. Finally, complex Gabor response was used for representation. In [25], a supervised method was used for initial classification with 128-length SIFT descriptor and weighted Clique was constructed using 4-length probability vector against the 4 nodule types. The overlap that lung nodule belongs to different types was used for optimizing the final classification result. These methods adopt general designed features. They obtain higher performance compared with traditional low-level features, while such methods are considered as mid-level abstraction of lung nodule and with less flexibility.

Several methods were concerned with other aspects. An ensemble based method was applied in [26] for lung nodule classification. Lung nodule image patch was used as input, and six large scale artificial neural networks were trained for classification. Data imbalance problem was discussed in [27]. It used downsampling and SMOTE algorithms to train lung nodule classifier.

Due to its breakthrough in the field of image processing and speech recognition, deep learning has become one of the most hottest topics in machine learning research and application [20, 28–30]. High-level abstraction of image object can be described using deep learning model. Meanwhile, feature extraction and representation are more efficient and effective. In [28], curvature, hu-moment, morphology, and shape features were used to detect nodule candidate region. Then convolutional neural network (CNN) was used to extract feature for candidate region and multiple classifiers were merged for final result. Some changes were made in [29, 30]. OverFeat was used for CNN parameter initialization. In [20], a deep feature extraction with one hidden layer autoencoder was adopted, and a binary decision tree was used as classifier for lung cancer detection. This paper proposes a lung nodule image classification method combining both local and global feature representation. Our proposed work is close but has essential difference from the work of [20]. Method in [20] just applied one hidden layer autoencoder to lung nodule image. Our proposed method uses Superpixel to generate intact patches and deep autoencoder to extract local feature. Moreover, method BOVW is incorporated for lung nodule global feature representation and method in [20] has no consideration.

3. Framework

The procedure of proposed lung nodule classification method is shown in Figure 2. It contains training and testing stages.
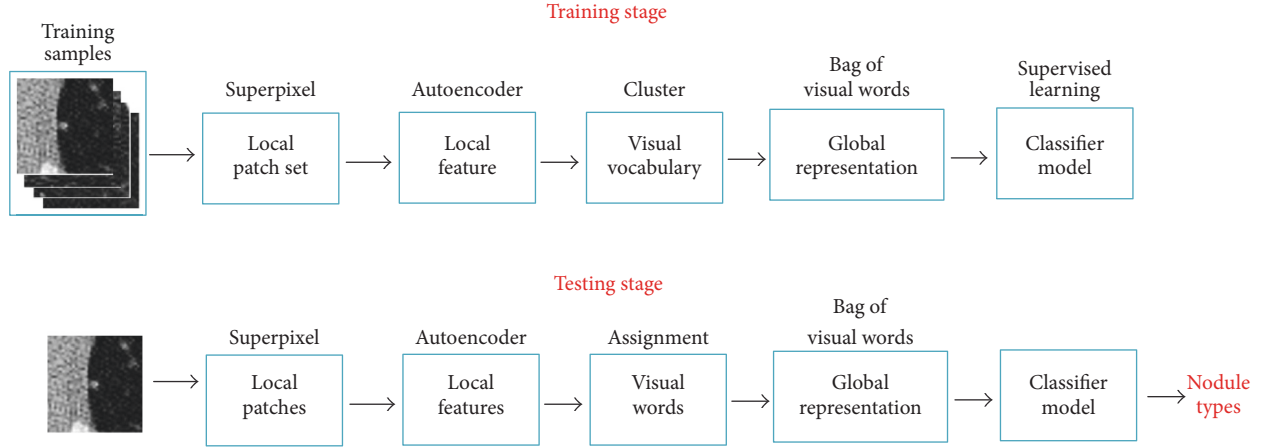
FIGURE 2: Framework of the proposed method.

In training stage, lung nodule image samples are used as input and the output is a trained classifier model. Collected training image samples are first divided into local patches with Superpixel. Local patches are assigned with no class label and constitute local patch set. With the local patch set, local features are extracted by unsupervised learning model, deep autoencoder. Next, visual vocabulary is constructed based on clustering all local feature vectors. A lung nodule image can therefore be described by a global feature representation with bag-of-visual-words model. Finally, classifier is trained by supervised learning with nodule type labels. In testing stage, the input is a lung nodule image with unknown type, and the output is its predicted type label. Similar to training stage, a test image is divided into multiple patches. Each local patch is transformed into local feature and assigned with a visual word. Finally, global feature representation of test image is used for classification by the trained model. Details of the proposed method will be introduced in the following sections.

4. Local Feature Representation

Local feature representation is proposed in this section. The process consists of two steps: (1) local patch generation and (2) local feature extraction and representation.

4.1. Local Patch Generation. Decomposing a lung image into small patches is useful and practical and for important tissues can be picked up and unrelated ones can be get rid of. As shown in (1), a lung nodule image x can be composed of a group of image patches p_i, where n denotes the number of local patches:

$$x = \{p_1, p_2, \ldots, p_n\}. \tag{1}$$

The location and scale of local patches are determined through generation [22, 24]. Useless part will be contained for large size patch, while small part may not cover enough intact tissue. Superpixel is a popular method that can partition the

image into small similar regions with better representativeness and integrity [31]. So it is adopted in this work.

Figure 3 illustrates the process of the proposed local patch generation method. For a lung nodule image (Figure 3(a)), it is first segmented into local patches using Superpixel and a Superpixel map is obtained (Figure 3(b)). Local patches essentially indicate the uniform regions. Figure 3(c) is an individual patch sample. However, the region that Figure 3(c) gives is an irregular shape, and it is inconvenient for local feature extraction and representation. So we expand local patch with its minimum enclosing rectangle, as shown in Figure 3(d). Finally, a lung nodule image is decomposed into a set of local patches, as shown in Figure 3(e). Besides, there are some additional criterions to determine whether an image patch is qualified for local feature extraction:

(i) Let p_i be a local patch; it is removed when the area of p_i is larger than $A_{\max}$ or smaller than $A_{\min}$.

(ii) Let p_i and p_j be two local patches; if the ratio between their intersection and their union is larger than O_t, then the smaller one is removed.

$A_{\max}$, $A_{\min}$, and O_t are predefined thresholds.

4.2. Local Feature Extraction and Representation. With the rapid development of unsupervised learning in recent years, using unlabeled data to extract feature with autoencoder has become an appropriate way. Autoencoder model is essentially a multilayered neural networks. Its original version is a forward network with one hidden layer. Let x_i be the input data, $a_i^{\,j}$ be the activation of unit i in layer j, and w_i be the matrix of weights controlling function mapping from layer i to layer $i + 1$. If layer i has s_i units and layer $i + 1$ has s_{i+1} units, then w_i will be a matrix with size of $s_i * s_{i+1}$. The activation can be formulated as (2), where a_1^2 is the 1st unit in the 2nd layer and x_0–x_3 are 4 input features:

$$a_1^2 = g\left(w_{10}^1 x_0 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3\right). \tag{2}$$

The main difference between ordinary forward neural network and autoencoder is that an autoencoder's output is
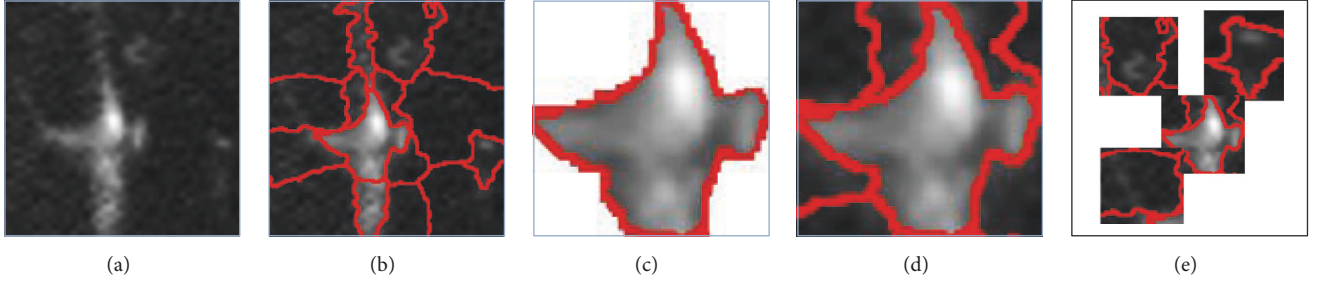
(a)   (b)   (c)   (d)   (e)

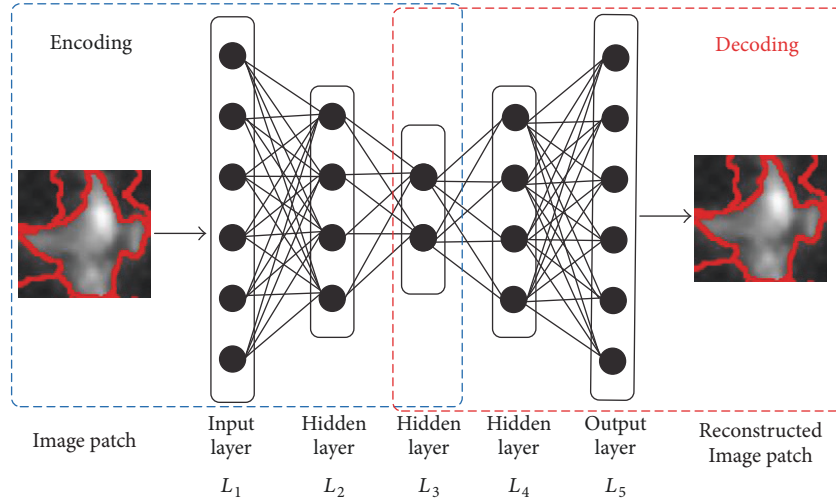FIGURE 3: The process of local patch generation by Superpixel.

FIGURE 4: The process of stacked deep autoencoder model.

always the same as or similar to its input. The basic formula can be expressed as follows:

$$a = h(x) = f\left(W^E x + b\right),$$
$$x' = h'(x) = g\left(W^D a + b'\right) = g\left(W^D h(x) + b'\right). \tag{3}$$

An autoencoder can be seen as a combination of encoder and decoder. The encoder includes an input layer and a hidden layer, which converts an input image x into feature vector a. The decoder includes a hidden layer and an output layer that transform feature a to output feature x'. W^E and W^D are weight matrices of encoder and decoder, respectively. Functions $f(\cdot)$ and $g(\cdot)$ can be either sigmoid or tanh activation functions, which is used to activate the unit in each layer. When x' approximates x, it is considered that the input feature can be reconstructed from an abstract and compressed output feature vector a. The cost function can be generally defined as follows:

$$J(W, b) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left\| x_i' - x_i \right\|^2 + \lambda \sum_{l=1}^{N_l - 1} \sum_{i=1}^{M_i} \sum_{j=1}^{M_{i+1}} \left(W_{ij}^l\right)^2. \tag{4}$$

A deep autoencoder can be constructed by stacking more hidden layers. As shown in Figure 4, there are 5 layers in the model (including 3 hidden layers). L_1 to L_3 are encoding layers, and L_3 to L_5 are decoding layers. L_i is used as the input of the layer L_{i+1}, and the weights can be gained based on (3). There are 2 stacked autoencoders. The activation of 1st hidden layer is the input of the 2nd stacked autoencoder. The network can be trained in a fine-tuning stage by minimizing the equation (4). W_1 and W_4 are trained through the encoding and decoding weights of the 1st stacked autoencoder, and W_2 and W_3 are trained through the encoding and decoding weights of the 2nd stacked autoencoder. Finally, the whole network can be constructed layer by layer in a stacked way. Moreover, Figure 4 just shows an example of symmetric encoding and decoding structures, and other variational structures can also be adopted.

Therefore, each local patch of a lung nodule image p_i can be represented by a fixed-length feature vector pf_i with deep autoencoder model. Then (1) is transformed as follows:

$$x = \{p_1, p_2, \ldots, p_n\} = \{pf_1, pf_2, \ldots, pf_n\}. \tag{5}$$

5. Global Feature Representation

For BOVW model, visual vocabulary is first constructed based on clustering all local patch descriptors (local feature representation) generated by a set of training images. Then each lung nodule image can be represented globally by a histogram of visual words. Distance between histograms of
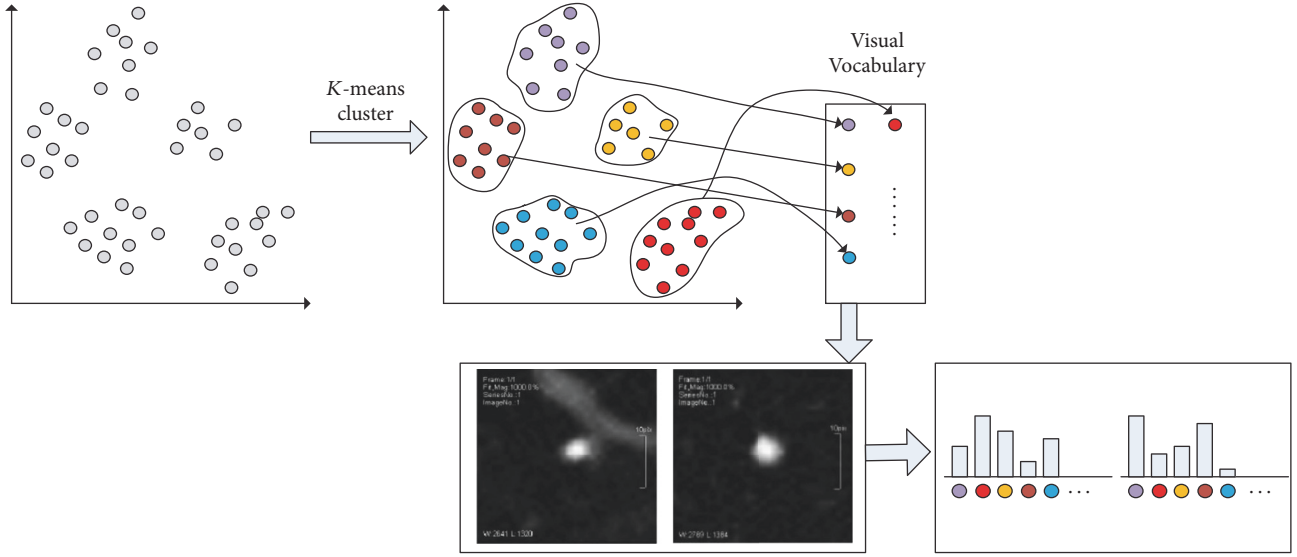
FIGURE 5: Procedure of BOVW representation of lung nodule image.

visual words can be treated as similarity between lung nodule image samples.

Recall that a lung nodule image is decomposed into a group of local patches and each patch is represented with a feature vector based on deep autoencoder. Assume there are D local patches generated from all lung nodule training images and each local patch is represented with d-dimensional feature vector; then all local feature vectors can be assembled into a feature space with size of $d * D$. Clustering is performed with $d * D$ features, and k-means clustering method is adopted since it has relatively low time and storage complexity, irrelevant to data process ordering. Each cluster center c_i represents a visual word i, and k cluster centers constitute the visual vocabulary. A lung nodule image sample x can be represented by the encoded local patches as a bag, which is the occurrence frequency of visual word in vocabulary. To get the histogram representation $h(x)$ of an image x, all local patch feature vectors of x are mapped onto the cluster center of the visual vocabulary, and each local feature is assigned with the label of its closest cluster center using Euclidean distance in feature space. Then a k-bins histogram $h(x)$ is obtained by counting all the label of local patches generated by image x, as shown in (6). Figure 5 exhibits the procedure of global representation of lung nodule image.

$$h(x) = [h(x)_1, h(x)_2, \ldots, h(x)_k]. \quad (6)$$

6. Classifier Model

With global representation of lung nodule image, softmax algorithm is used to train nodule type classifier. Let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$ denote training data set. x_i denotes the lung nodule image sample and $y_i \in \{0, 1, 2, 3\}$ denotes nodule type label.

For an input image sample x_i, we want to compute $p(y = j \mid x_i)$ ($j \in \{0, 1, 2, 3\}$). The output, a 4-dimensional vector, is estimated to represent the probability of each type that x_i belongs to. The hypothesis function can be expressed as follows:

$$h_\theta(x_i) = \begin{bmatrix} p(y_i = 0 \mid x_i; \theta) \\ p(y_i = 1 \mid x_i; \theta) \\ p(y_i = 2 \mid x_i; \theta) \\ p(y_i = 3 \mid x_i; \theta) \end{bmatrix} = \frac{1}{\sum_{j=0}^{3} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_0^T x_i} \\ e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ e^{\theta_3^T x_i} \end{bmatrix}, \quad (7)$$

where $\theta = \{\theta_0, \theta_1, \theta_2, \theta_3\}$ is model parameter set. This equation normalizes the result and makes the sum to 1. For training procedure, the loss function is given as follows:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^{m} \sum_{j=0}^{3} 1\{y_i = j\} \log \frac{e^{\theta_j^T x_i}}{\sum_{l=0}^{3} e^{\theta_l^T x_i}} \right] + \frac{\lambda}{2} \sum_{i=1}^{k} \sum_{j=0}^{n} \theta_{ij}^2, \quad (8)$$

where $1\{\cdot\}$ is an indicative function, and stochastic gradient descent (SGD) is used for function optimization and the corresponding derivative functions are given as follows:

$$\nabla_{\theta_j} J(\theta)$$

$$= -\frac{1}{m} \sum_{i=1}^{m} \left[x_i \left(1\{y_i = j\} - p(y_i = j \mid x_i; \theta) \right) \right] + \lambda \theta_j, \quad (9)$$

$$p(y_i = j \mid x_i; \theta) = \frac{e^{\theta_j^T x_i}}{\sum_{l=0}^{3} e^{\theta_l^T x_i}}.$$

FIGURE 6: Demonstration of lung CT images (downloaded from [6]).

7. Experimental Evaluations

7.1. Dataset and Program Implementation. In order to evaluate the performance of the proposed lung nodule image representation and classification method, a widely used public available lung nodule image dataset, ELCAP, is used for testing [6]. The dataset contains 379 lung CT images, which are collected from 50 distinct low-dose CT lung scans. The center position of lung nodule is marked in an extra ∗.csv file.

Figure 6 demonstrates the lung nodule CT scan images, which are sampled from different slices. Table 1 shows the format of a ∗.cvs file. Each row denotes a lung nodule. The 4th column indicates the slice number where the lung nodule exists. The 2nd and 3rd columns give the positions that the lung nodule is located in. In this section, lung nodule images are cropped from the raw CT images based on the x- and y-coordinates of nodule center given in Table 1. The raw lung CT scan image is fixed with $512 * 512$ pixels, and the cropped nodule images are too small to implement the algorithm. Therefore we further resize the cropped lung nodule image into $180 * 180$ pixels with bicubic method. The lung nodule images are labeled with one of four types according to the guidance by an expert. Programs are implemented with Matlab 2016a programming language and tested on a Pentium i7 CPU, 8 G RAM, NVIDIA GTX 960 GPU, Windows OS PC.

The experiments include the following aspects: (1) parameter setting; (2) classification rate with different parameters; (3) classification rate with data augmentation; (4) classification rate with model sparsity; (5) classification rate with different classifier algorithms; (6) comparing with other methods; (7) classification rate with model ensemble. The performance of lung nodule image classification is computed with overall classification rate, as shown in the following:

$$\text{Classification rate} = \frac{N_{\text{correct}}}{N_{\text{all}}}, \tag{10}$$

where N_{correct} is the number of correctly labeled images and N_{all} is the number of all testing images. Cross validation mode is adopted. The dataset is divided into 8 groups: 7 randomly chosen groups are used for training and the left group is used

TABLE 1: Format of lung nodule position.

Type	x	y	Slice
Nodule	98	218	54
Nodule	355	153	84
Nodule	139	366	130
Nodule	436	213	169
Nodule	372	163	239
Nodule	328	175	229
Nodule	54	224	169

for testing. This process is repeated 7 times and the result is computed by averaging 7 independent tests.

7.2. Parameter Setting. The parameters are needed to be set in local patch generation, local feature representation, and global feature representation. For local patch generation, we need to set the number of superpixels that each lung nodule image generates. For local feature representation, the number of hidden layers and nodes that each layer contains should be set. For global feature representation, the size of visual vocabulary should be set.

As shown in Table 2, the number of patches that each lung nodule image generates is set with 15, 20, 25, and 30. The number of hidden layers in deep autoencoder is set with 1, 2, and 3. The number of nodes in deep autoencoder is set with 50, 75, 100, 125, and 150. The size of visual vocabulary is set with 200, 300, 400, and 500. The classification rate is evaluated on the combination of these parameters. For convenience, parameters are expressed with p_1, p_2, p_3, and p_4, respectively.

7.3. Classification Rate with Different Parameters. The size of local patch is set with $30 * 30$ pixels in our experiment. Table 3 gives the average performance of lung nodule image classification based on combination of parameters p_1, p_2, p_3, and p_4. It can be seen that classification model with $p_1 = 25$, $p_2 = 2$, $p_3 = (100, 50)$, and $p_4 = 400$ gets the optimal result,

Table 2: Parameter setting.

Parameter	Parameter explanation	Setting
p_1	Number of Superpixel generation	15, 20, 25, 30
p_2	Hidden layers of deep autoencoder	1, 2, 3
p_3	Nodes of deep autoencoder	50, 75, 100, 125, 150
p_4	Size of visual vocabulary	200, 300, 400, 500

Table 3: Performance with different parameters.

Parameters setting			Performance
p_1	p_2 (p_3)	p_4	
15	1 (50)	200	0.8199
15	1 (50)	400	0.832
15	1 (50)	600	0.829
15	2 (100, 50)	200	0.858
15	2 (100, 50)	400	0.86
15	2 (100, 50)	600	0.845
15	3 (150, 100, 50)	200	0.836
15	3 (150, 100, 50)	400	0.845
15	3 (150, 100, 50)	600	0.85
25	1 (50)	200	0.83
25	1 (50)	400	0.842
25	1 (50)	600	0.849
25	2 (100, 50)	200	0.887
25	2 (100, 50)	400	0.895
25	2 (100, 50)	600	0.891
25	3 (150, 100, 50)	200	0.835
25	3 (150, 100, 50)	400	0.832
25	3 (150, 100, 50)	600	0.86
40	1 (50)	200	0.824
40	1 (50)	400	0.82
40	1 (50)	600	0.813
40	2 (100, 50)	200	0.824
40	2 (100, 50)	400	0.815
40	2 (100, 50)	600	0.833
40	3 (150, 100, 50)	200	0.81
40	3 (150, 100, 50)	400	0.80
40	3 (150, 100, 50)	600	0.806

reaching 89.5%. We can also see that different parameter settings have great impact on the classification results.

7.4. Classification Rate with Data Augmentation. Overfitting is common in machine learning, and it is influenced by both model complexity and the size of training data. Data augmentation scheme is usually adopted to lessen this problem [32]. In this section, data augmentation is used to enlarge the size of training data. Random rotation, random cropping, and random perturbation (brightness, saturation, hue, and contrast) are used as basic augment techniques.

For original lung nodule image, it is sampled with possibility of 0.5 for data augmentation. The new created examples are set with same labels as original. As shown in Table 4, data augmentation can increase classification rate

Table 4: Performance with data augmentation.

Data augmentation method			Average performance
Random rotation	Random cropping	Random perturbation	
No	No	No	0.895
Yes	No	No	0.899
No	Yes	No	0.887
Yes	Yes	No	0.908
No	No	Yes	0.903
Yes	No	Yes	0.911
No	Yes	Yes	0.912
Yes	Yes	Yes	0.924

with 3%. This shows that adding more augmented data for training can improve the compatibility and generalization of the classification model.

7.5. Classification Rate with Model Sparsity. In this subsection, a sparsity constraint is imposed on the hidden layer. Sparsity is a recently proposed technique to improve the generalization of the model [33]. A sparsity regularization term is added to (4), and the new objective functions are given as follows:

$$
\begin{aligned}
J(W, b) = {} & \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left\| x_i' - x_i \right\|^2 \\
& + \lambda \sum_{l=1}^{N_l-1} \sum_{i=1}^{M_i} \sum_{j=1}^{M_{i+1}} \left(W_{ij}^l \right)^2 \\
& + \beta \sum_{i=1}^{N_l-1} \mathrm{KL}\left(\rho \parallel \rho_i \right),
\end{aligned}
\tag{11}
$$

$$
\mathrm{KL}\left(\rho \parallel \rho_i \right) = \rho \log \frac{\rho}{\rho_i} + (1 - \rho) \log \frac{(1 - \rho)}{(1 - \rho_i)},
$$

$$
\rho_i = \frac{1}{N} \sum_{i=1}^{N} h_i(x).
$$

The sparsity regularization term is regulated by Kullback-Leibler divergence $\mathrm{KL}(\rho \parallel \rho_i)$. ρ_i is the average activation of ith layer of deep autoencoder and ρ is the target activation. ρ with small value can reduce the mean activation of the model. β is a trade-off parameter. Table 5 gives the result of classification performance with different ρ (values from 0.1–0.9). It can be seen that ρ set around 0.3–0.4 leads to the superior performance.

7.6. Classification Rate with Different Classifier Algorithms. In this subsection, we evaluate the performances of 4 commonly used classifier algorithms. Softmax (which is used in this paper), SVM, kNN, and decision tree are used. The same feature representation is adopted. Table 6 shows that softmax slightly outperforms SVM, kNN, and decision tree. The

TABLE 5: The effect of model sparsity.

Sparsity (ρ)	Performance
0.1	0.92
0.2	0.929
0.3	0.938
0.4	0.939
0.5	0.93
0.6	0.905

TABLE 6: The effect of different classifier algorithms.

Classifier model	Performance
Softmax	0.939
SVM	0.931
kNN	0.927
Decision tree	0.919

TABLE 7: Performance comparing with other methods.

Classification method	Performance
Ref. [19]	0.877
Ref. [7]	0.88
Ref. [20]	0.82
Ref. [21]	0.895
Ref. [9]	0.891
Our proposed method	0.939

TABLE 8: Performance of model ensemble.

Number of models	Performance
1	0.939
5	0.952
6	0.954
7	0.955

results demonstrate that, compared with classifier algorithm, the feature representation is the key problem. Meanwhile, it is easy to combine the softmax algorithm and the proposed feature representation method into an end-to-end structure, which can make model training more convenient.

7.7. Comparing with Other Methods. In order to evaluate the classification rate of different methods, 5 related algorithms are used for testing. Reference [19] studies the same problem as ours. Reference [20] adopts the primitive autoencoder method. References [7, 21] use non-deep-learning methods for classification. Reference [9] employs the BOVW model. The compared methods are reimplemented and are tested with diverse parameters. Table 7 gives the testing result. Among all testing methods, the proposed one demonstrates the best performance. Comparing with non-deep-learning method, our method can construct better feature representation, while, comparing with primitive autoencoder method, the Superpixel and DAE used in our method can catch more detailed information.

7.8. Classification Rate with Model Ensemble. Model ensemble can improve the classification performance by aggregating multiple individual classifiers [34]. We evaluate model ensemble based on Majority Rule in this subsection. In Majority Rule, the class label is assigned with the one that most classifier votes. The function to evaluate the class label e for image I is given as follows:

$$p(e) = \sum_{i=1}^{N} S(C_i(I) = e),$$
$$e = \arg\max_i p(e_i),$$
(12)

where I is a testing image, e is a class label, N denotes number of selected models, and C_i means ith classifier. $S(C_i(I) = e) = 1$, if C_i classifies I as e. The label with maximal value of $p(\cdot)$ is determined as the final result. If multiple labels have the same

votes, the arithmetic average of class probabilities predicted by individual model is used as classification result.

With different parameters combination, models with top performances are retained for ensemble. Table 8 gives the testing result. The 1st row denotes the single model. The 2nd to 4th rows denote model ensemble with 5, 6, and 7 individual models, respectively. The result demonstrates that model ensemble can complement individual ones and the performance is improved with about 1.5%.

8. Conclusion and Future Works

In this paper, a novel feature representation method is proposed for lung nodule image classification. Superpixel is first used to divide lung nodule image into local patches. Then local feature is extracted and represented from local patches with deep autoencoder. Bag-of-visual-words model is used as global feature representation with visual vocabulary constructed by local feature representation. Finally, an end-to-end training is implemented with a softmax classifier. The proposed method is evaluated from many aspects, including parameter setting, data augmentation, model sparsity, comparison among different algorithms, and model ensemble. We draw a conclusion that the proposed method achieves superior performance. The merits of our method are the combination of local and global feature representation, and better model generalization can be gained by incorporating unsupervised deep learning model.

Our future works will focus on two aspects: (1) study new classification framework and method according to up-to-date convolutional neural network and (2) analysis of our method in large data set for making further improvement and optimization.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] National Cancer Institute, "SEER Stat Fact Sheets: Lung and Bronchus Cancer. NCI online 2016," http://seer.cancer.gov/statfacts/html/lungb.html.

[2] American Cancer Society, https://www.cancer.org.

[3] Y. Shieh and M. Bohnenkamp, "Low-Dose CT Scan for Lung Cancer Screening: Clinical and Coding Considerations," *CHEST*, vol. 152, no. 1, pp. 204–209, 2017.

[4] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "Content-Based Image Retrieval System for Pulmonary Nodules: Assisting Radiologists in Self-Learning and Diagnosis of Lung Cancer," *Journal of Digital Imaging*, vol. 30, no. 1, pp. 63–77, 2017.

[5] S. Diciotti, G. Picozzi, M. Falchini, M. Mascalchi, N. Villari, and G. Valli, "3-D segmentation algorithm of small lung nodules in spiral CT images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 7–19, 2008.

[6] ELCAP public lung image database, http://www.via.cornell.edu/databases/lungdb.html.

[7] J. N. Stember, "The Normal Mode Analysis Shape Detection Method for Automated Shape Determination of Lung Nodules," *Journal of Digital Imaging*, vol. 28, no. 2, pp. 224–230, 2015.

[8] Y. Song, W. Cai, H. Huang, Y. Zhou, Y. Wang, and D. D. Feng, "Locality-constrained subcluster representation ensemble for lung image classification," *Medical Image Analysis*, vol. 22, no. 1, pp. 102–113, 2015.

[9] F. Ciompi, C. Jacobs, E. T. Scholten et al., "Bag-of-frequencies: a descriptor of pulmonary nodules in computed tomography images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 4, pp. 962–973, 2015.

[10] F. Han, G. Zhang, H. Wang et al., "A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database," in *Proceedings of the 2013 IEEE International Conference on Medical Imaging Physics and Engineering, ICMIPE 2013*, pp. 14–18, IEEE, Shenyang, China, October 2013.

[11] J. Bai, X. Huang, S. Liu, Q. Song, and R. Bhagalia, "Learning orientation invariant contextual features for nodule detection in lung CT scans," in *Proceedings of the 12th IEEE International Symposium on Biomedical Imaging, ISBI 2015*, pp. 1135–1138, IEEE, New York, NY, USA, April 2015.

[12] C. Jacobs, E. M. van Rikxoort, T. Twellmann et al., "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical Image Analysis*, vol. 18, no. 2, pp. 374–384, 2013.

[13] T. W. Way, B. Sahiner, H.-P. Chan et al., "Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features," *Medical Physics*, vol. 36, no. 7, pp. 3086–3098, 2009.

[14] S. Akram, M. Y. Javed, and A. Hussain, "Automated thresholding of lung CT scan for Artificial Neural Network based classification of nodules," in *Proceedings of the 14th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2015*, pp. 335–340, IEEE, Las Vegas, NV, USA, July 2015.

[15] A. Robert, G. Jonathan, A. Fereidoun et al., "Automated classification of lung bronchovascular anatomy in CT using AdaBoost," *Medical Image Analysis*, vol. 11, no. 3, pp. 315–324, 2007.

[16] A. El-Baz, G. Gimel'farb, R. Falk, and M. El-Ghar, "Appearance analysis for diagnosing malignant lung nodules," in *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2010*, pp. 193–196, IEEE, Rotterdam, Netherlands, April 2010.

[17] H. Takizawa, S. Yamamoto, and T. Shiina, "Accuracy improvement of pulmonary nodule detection based on spatial statistical analysis of thoracic CT scans," *IEICE Transaction on Information and Systems*, vol. 90, no. 8, pp. 1168–1174, 2007.

[18] K. Yoshiki, N. Noboru, O. Hironobu et al., "Hybrid Classification Approach of Malignant and Benign Pulmonary Nodules Based on Topological and Histogram Features," *Medical Image Computing & Computer Assisted Intervention*, pp. 297–306, 2000.

[19] F. Zhang, Y. Song, W. Cai et al., "A ranking-based lung nodule image classification method using unlabeled image knowledge," in *Proceedings of the IEEE 11th International Symposium on Biomedical Imaging (ISBI '14)*, pp. 1356–1359, Beijing, China, May 2014.

[20] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Proceedings of the 12th Conference on Computer and Robot Vision (CRV '15)*, pp. 133–138, IEEE, Halifax, Canada, June 2015.

[21] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "A Combination of Shape and Texture Features for Classification of Pulmonary Nodules in Lung CT Images," *Journal of Digital Imaging*, vol. 29, no. 4, pp. 466–475, 2016.

[22] F. Zhang, Y. Song, W. Cai et al., "Lung nodule classification with multilevel patch-based context analysis," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1155–1166, 2014.

[23] Y. Song, W. Cai, Y. Wang, and D. D. Feng, "Location classification of lung nodules with optimized graph construction," in *Proceedings of the 9th IEEE International Symposium on Biomedical Imaging (ISBI '12)*, pp. 1439–1442, IEEE, Barcelona, Spain, May 2012.

[24] A. Farag, S. Elhabian, J. Graham, A. Farag, and R. Falk, "Toward precise pulmonary nodule descriptors for nodule type classification," *Medical Image Computing and Computer-Assisted Intervention*, vol. 13, no. 3, pp. 626–633, 2010.

[25] F. Zhang, W. Cai, Y. Song, M.-Z. Lee, S. Shan, and D. Dagan, "Overlapping node discovery for improving classification of lung nodules," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '13)*, pp. 5461–5464, Osaka, Japan, July 2013.

[26] K. Suzuki, F. Li, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1138–1150, 2005.

[27] Y. Sui, Y. Wei, and D. Zhao, "Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE," *Computational and Mathematical Methods in Medicine*, Article ID 368674, pp. 1–13, 2015.

[28] A. A. A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[29] B. Van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Proceedings of the 12th IEEE International Symposium on Biomedical Imaging (ISBI '15)*, pp. 286–289, Brooklyn, NY, USA, April 2015.

[30] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 9123, pp. 588–599, 2015.

[31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, 2012.

[32] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," https://arxiv.org/abs/1207.0580.

[33] X. Zhang and R. Wu, "Fast depth image denoising and enhancement using a deep convolutional network," in *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2016*, pp. 2499–2503, Shanghai, China, March 2016.

[34] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1–39, 2010.