# Multi-view Multi-scale CNNs for Lung Nodule Type Classification from CT Images

Xinglong Liu[a], Fei Hou[b,*], Hong Qin[c], Aimin Hao[a]

[a] *State Key Laboratory of Virtual Reality Technology and Systems, Beihang University.*
[b] *State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences.*
[c] *Department of Computer Science, Stony Brook University.*

**Abstract**

In this paper, we propose a novel convolution neural networks (CNNs) based method for nodule type classification. Compared with classical approaches that are handling four solid nodule types, i.e., *well-circumscribed*, *vascularized*, *juxta-pleural* and *pleural-tail*, our method could also achieve competitive classification rates on *ground glass optical (GGO)* nodules and *non-nodules* in computed tomography (CT) scans. The proposed method is based on multi-view multi-scale CNNs and comprises four main stages. First, we approximate the spherical surface centered at nodules using icosahedra and capture normalized sampling for CT values on each circular plane at a given maximum radius. Second, intensity analysis is applied based on the sampled values to achieve estimated radius for each nodule. Third, the re-sampling (which is the same as the first step but with estimated radius) is conducted, followed by a high frequency content measure analysis to decide which planes (views) are more abundant in information. Finally, with approximated radius and sorted circular planes, we build nodule captures at sorted scales and views to first pre-train a view independent CNNs model and then train a multi-view CNNs model with maximum pooling. The experimental results on both Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [1] and Early Lung Cancer Action

*Corresponding author
*Email address:* houfei0801@gmail.com (Fei Hou)

Program(ELCAP) [2] have shown the promising classification performance even with complex *GGO* and *non-nodule* types.

*Keywords:* Computed Tomography, Lung Nodule, CNNs

## 1. Introduction

Lung cancer has been the leading cause of cancer-related deaths in United States in 2016 [3]. There are over 150,000 deaths caused by lung and bronchus cancer, accounting for over 25% of all cancer-related deaths in the past year. A total of over 220,000 new lung and bronchus cancer cases are projected to occur in 2017. Fortunately, early detection and localization of nodules could significantly improve the survival rate to 52% [4]. Among all the cases with nodules, approximately 20% represent lung cancers [5]. Therefore, identification of nodules from malignant to innocent is an essential part for both screening and diagnosis of lung cancer [6].

Shapes of lung nodules are usually spherical, but can be complicated when surrounded by anatomical structures such as vessels, and pulmonary walls. Intra-parenchymal and GGO type nodules have higher chance to be lung cancer than those only connected with surrounding structures. Based on the fact that malignancy has strong correlation with intensity distribution and relative positions, lung nodules are categorized into different types: well-circumscribed (**W**), located in the center of the lung without any connection to other tissues; vascularized (**V**), the same as **W** except for connecting to vascular structures; juxta-pleural (**J**), fully connected to pleural surface; pleural-tail (**P**), close to pleural but only connected with a thin tail; GGO (**G**), the same with **W** except for its non-solid CT value distribution. Besides these, we also conduct investigation on the so-called non-nodules (**N**), which have partial nodule features but actually are not nodules. Typical cases for each type are shown in Fig. 1.

CT could capture luxuriant details of both lung nodules and surrounding structures, leading to a powerful tool towards diagnosis. However, lung nodules are so complex that even experienced radiologists can not correctly tell

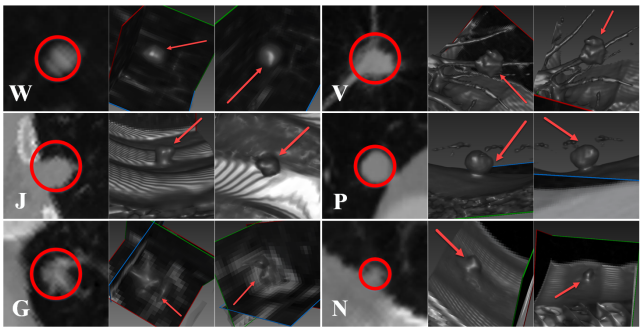


Figure 1: Typical cases for each nodule type. Both images on axial plane and in 3-D spaces are presented. Approximated nodule radii are shown in red circles. Red arrows are used to point out corresponding nodules in 3-D space.

malignancy from benign in some cases. It is reasonable if we can label lung nodules on CT images to their corresponding types, which will be of great help for radiologists for early lung cancer diagnosis.

### *1.1. Previous Work*

To detect lung cancer at early stage, various researches have been carried out. Image intensity-based thresholding or edge preserving smooth is most commonly used because of their ability of overcoming large image contrast between lung regions and surrounding body tissues [7]. Gao et al. [8] propose another threshold-based approach consisting of four steps which take anatomical information into consideration. Although simple and fast, it is extremely hard to determine the threshold since it is greatly related to the image acquisition types, protocols, intensities, especially when there are severe lung pathologies [9].

Compared with processing original CT images, the other straightforward way is to design features that can best describe the characteristics of nodules. The grey-level distribution represented as a histogram [12] has been effectively used to characterize intensity variations.

Also, filter-based feature extraction techniques are widely applied to highlight the edge and shape information for images [13]. Besides these commonly-used computer vision features, new methods that are more incorporated with

medical image analysis are also proposed. Scale-invariant feature transform (SIFT) [15] provides a robust way invariant to image transformation, scaling and rotation. Histogram of oriented gradients (HOG) [16] is interpreting objects by occurrences of gradient orientation in localized portions of an image. Local binary patterns (LBP) [17] provides a powerful tool for local texture classification by applying multi-scale and rotation-variant property. Ciompi et al. [18] have encoded the nodule intensity distribution patterns into frequency domain and classified nodules in a bag-of-words fashion. Although these descriptors are effective individually, it is becoming more complex to achieve outstanding human-design features towards huge and various data nowadays. Gu et al. [**?** ] propose an automatic lung lesion segmentation method based on existing "Click & Grow" method, which only require one seed point. The average similarity indexes on 129 CT lung tumor images have shown the accuracy and stability of the proposed method.

Besides intimate connection with local image analysis, Computer Aided Detection (CADe) systems have been developed and widely used to assist radiologists in diagnosis and thereby making screening more effective [19]. CADe usually comprise of two stages: nodule candidate detection stage which usually uses double thresholding and morphological operations to detect a large number of candidates, and false positive reduction stage, which uses features and supervised classification. Although CADe systems are effective, there do exist a considerable number of nodules undetected, prohibiting their widespread use in clinical practice [20]. Abdel-Nasser et al. [**?** ] propose a strain-tensor-based method for quantifying and visualizing changes of breast tumors for patients with undergoing medical treatments. They evaluate the performance of eight robust and recent optical flow methods and use the aggregated approaches to estimate the displacement fields.

Recently, inspired by a large amount of available data and more powerful computational resources, especially parallelization ability empowered by Graphic Processing Units (GPUs), convolutional neural networks (CNNs) [21] have shown their abilities of outperforming the state-of-the-art in classical computer

vision applications [23], as well as in the field of medical image analysis [25]. Parveen et al. [26] have conducted reviews on computer aided detection and diagnosis methods for lung cancer nodules including both classical and modern methods for preprocessing, segmentation and classification, showing the high efficiency of artificial neural network based methods. In [27], they investigate more details on deep learning techniques including Deep Belief Network (DBN) and CNNs for classification of lung nodules. Their experiments and results have shown that deep learning framework can outperform the conventional feature computing CADe methods. Shin et al. [28] have deployed researches on different CNNs architectures, different data sets on two specific CADe problems including thoraco-abdominal lymph node detection and interstitial lung disease classification, indicating the high potentials of CNNs in CADe field. Meanwhile, since CNNs can be trained end-to-end with the layers automatically learning discriminative features without handcrafting design, they are very suitable for lung nodule type classification considering the complex intensity and surrounding anatomical structure distributions. Lo et al. [29] proposed the CNNs based method for pulmonary nodule detection in chest radiology images. Arnaud et.al [30] have used the multi-view CNNs for pulmonary nodule false positive reduction in CT images, with a complete performance discussion over different fusion methods [24]. Cao et al. [**?** ] have proposed a multi-kernel based framework for feature selection and imbalanced data learning, using multiple kernel learning, multi-kernel feature selection and multi-kernel over-sampling. A few other studies have also extended the use of 2D CNNs to 3D volumetric analysis on 3D images, i.e., CT images. In all of these methods, volumetric images are projected to fixed views (planes), followed by that each view is processed under 2D CNNs and finally integrated under a multi-view fashion with the best fusion methods. Li et al. [33] have applied deep CNNs for nodule classification, which achieve better feature representations for solid, semisolid and GGO nodules. Lin et al. [**?** ] have proposed a set of fractal features based on fractional Brownian motion model to distinguish nodule malignancy instead of classical CT attenuation values. Experiments on 107 CT data have shown that their method achieve

promising results in accuracy, sensitivity, etc. Liu et al. [34] have adopted multi-view CNNs with two convolutional, two pooling and one fully connected layers for binary and ternary nodule classification. Shen et al. [**?** ] have proposed an end-to-end Multi-crop Convolutional Neural Network to automatically extract salient nodule information from raw CT images with multi-cropped regions and max-pooling operations to unify outputs of all regions, which can predict not only nodule malignancy but also nodule semantic attributes. Jiang et al. [35] cut multi-group patches from original images, followed by Frangi filter enhancement to generate feedings for a four-channel CNNs. They acquire a relative low false positive rate at high sensitivity. Christodoulidis et al. [36] have proposed an improved method for lung pattern analysis by transferring knowledge from the similar domain of general texture classification. Anthimopoulos et al. [37] have applied deep CNNs to classify healthy, ground glass opacity (GGO), micro-nodules, consolidation, reticulation, honeycombing and a combination of GGO/reticulation nodules for interstitial lung diseases diagnosis. Their method has achieved 85.5% accuracy and demonstrating the potentials of CNNS in analyzing lung patterns. Chen et al. [38] exploit three multi-task learning schemes including stacked denoising autoencoder, CNNs and hand-crafted Haar-like and HoG features, to obtain semantic feature descriptions for lung nodules in CT images. Their method may provide quantitative assessments of nodule for better diagnosis.

While many studies have reported detection and segmentation of pulmonary nodules, there are limited researches in nodule type classification. Farag et al. [39] have reported some basic studies in classification problem. Zhang et al. [14] designed an overlapping nodule identification procedure for nodules located at intersections among different types. After that, they have proposed a multilevel patch-based context analysis for nodule classification [40] in light of prior work proposed by Song et al. [41] and validated their method on the public ELCAP database.

All the aforementioned methods are based on designed features that are calculated from candidate regions perhaps already segmented from images. How-

ever, segmentation from blurry medical images has always been a non-trival task and at the same time it is doubtful how well human-crafted features could characterize candidate regions. Therefore, we apply CNN-based approaches because of the fact that they are taking raw images as input without the need of segmentation and its self-adaptive weight tuning between adjacent layers. We believe that CNN-based approaches would achieve comparable performance to tackle the complexity of the pulmonary nodule classification problem.

### *1.2. Contributions*

Inspired by the aforementioned works, this paper presents a CNN-based nodule type classification method in a multi-view multi-scale fashion. The pipeline of the proposed method is described in Fig. 2. Different from other classical methods, we also conduct investigation on GGO and non-nodules. The major contributions of our work include:

- A comprehensive method for classifying not only solid nodule types such as well-circumscribed and vascularized ones, but also GGO and non-nodule types.
- A normalized spherical sampling pattern based on icosahedron and a nodule radius approximation method based on thresholding.
- A view selection method for nodules on CT images based on high frequency content analysis.
- A multi-view multi-scale re-sampling and color projection method for nodules, based on which the CNNs with maximum pooling is trained.
- A comprehensive validation on the publicly accessible datasets of LIDC-IDRI and ELCAP.

The rest of this paper is organized as follows. Section 2 provides details for data source we have been using to train and validate our method. Section 3 documents how to sample nodule volumes, estimate nodule radii, select views

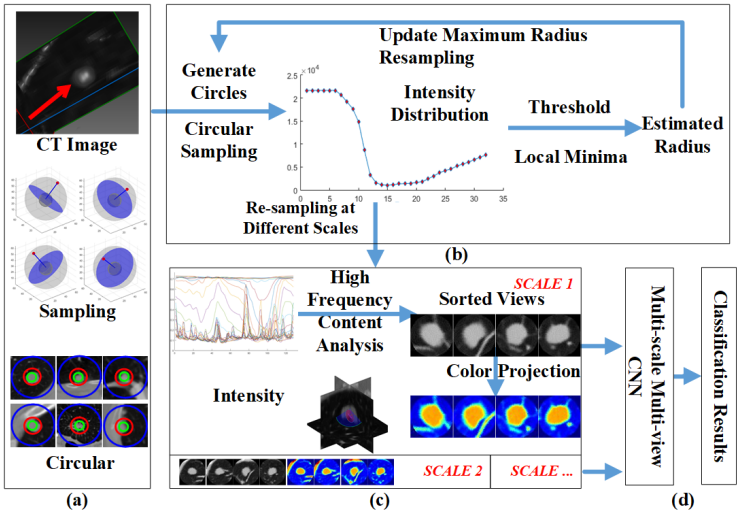


Figure 2: Pipeline of the proposed method. (a) Sampling patterns. (b) Intensity threshold based nodule radius estimation. (c) High frequency content based view selection. (d) multi-view multi-scale CNNs with soft-max layer.

and ultimately train CNNs in a multi-view multi-scale fashion. A comprehensive comparison between the proposed method and PB [40] method, as well as comparison among our own methods is presented in Section 4. Finally, we draw a brief conclusion and pinpoint possible future work in Section 5.

## 2. Materials

In our approach, we train and validate the CNNs both on LIDC-IDRI [1] and ELCAP [2] separately. Original data from both sets are processed and annotated by an experienced radiologist to obtain corresponding nodule types and labels.

### *2.1. LIDC-IDRI*

LIDC-IDRI [1] consists of totally 1018 chest CT image cases. Each case in the database includes images from a clinical thoracic CT scan and an associated eXtensible Markup Language (XML) file that records the locations of the image pixels on the nodule boundary in each image and nine characteristics of the

nodule malignancy given by up to four experienced thoracic radiologists. LIDC-IDRI contains a large quantity of non-blind expert- labeled nodule cases with multiple types from different institutions [42] and provides a feasible platform to quantitatively compare between different methods.

From the original LIDC-IDRI, only nodules with at least agreement level 2 (count of radiologists labeling the area as nodules or non-nodules) and sizes larger than 3 mm are extracted. Totally, we obtain 1738 nodules (**W**: 905, 52.0%; **P**: 329, 19%; **V**: 219, 12.5%; **G**: 82, 6%; **J**: 203, 11.5%) and 1000 non-nodules from 744 chest CTs with various imaging qualities of different patients. The CT screening parameters include reconstruction intervals of 0.6-3.0 mm, and tube currents of 120 mA with 200-400 kVp. The original CT slice is sampled with 512 $\times$ 512 pixels with in-plane spacing as 0.4-1.0 mm, while the extracted nodules are organized at 64 $\times$ 64 pixels with normalized in-plane spacing 1 mm. 80% of nodules for each type are used in training and validating, while the other 20% are used to test trained CNNs. To enrich varieties, both training and testing cases are augmented at different scales and views, separately.

### *2.2. ELCAP*

ELCAP consists of an image set of 50 low-dose documented whole lung CT scans with labeled nodules annotated by two experienced radiologists. Scans are achieved in a single breath hold with 1.25 mm slice thickness. The in-plane spacing is 0.5 mm while most nodule sizes range from 2 mm to 5 mm.

In our approach, we select 46 cases with 421 nodules (**W**: 92, 21.8%; **V**: 49, 11.6%; **P**: 155, 36.8%; **J**: 106, 25.2%, **G**: 19, 4.6%). It should be noted that, ELCAP is a bit different towards LIDC-IDRI in two fields: its nodule sizes are very small; it contains no annotated non-nodules. In fact, tiny nodules mostly can only be observed on one single slice even for the ELCAP data, which has already sustained perfect slice thickness as thin as 1.25 mm. That means annotations for these nodules are not reliable and they can be recognized as non-nodules even for radiologists during clinical practice. Meanwhile, our CNNs model is trained, validated and tested with nodules larger than 3 mm on

LIDC-IDRI. Therefore, to achieve fair comparisons, we have changed labels of nodules whose sizes ¡= 2 mm as non-nodules in our experiments.

### *2.3. Data Augmentation*

Extracted labeled nodules from LIDC-IDRI data set are unbalanced for each type, which may mislead the optimization of CNNs to local optima and cause predictions biased towards more frequent samples. Data augmentation is an efficient method to prevent overfitting and misleading by adding in-variances to existing data.

Nodule numbers for each type are much smaller than non-nodules, therefore augmentation is only done to nodule types (type **W**, **V**, **P**, **J**, **G**). Trainning data augmentation is done by two ways: (1) nodules are scaled [2.5, 3.0, 3.5] times of their estimated radii (maximum 32 mm) and enlarged to 64 mm $\times$ 64 mm by linear sampling; (2) nodules are re-sampled for each scale with random selection in the sorted views. After augmentation, we achieve almost 640 training cases for each type (nodules and non-nodules) in LIDC-IDRI.

It is reported that test data augmentation can also help to improve the performance and robustness of CNNs [43]. Test data augmentation is performed on each type by the same multi-view multi-scale way as training data. Finally, we obtain almost 160 cases for each test type and resulting in around 1000 cases in total in LIDC. Meanwhile, there is no need for consideration of balancing for training in ELCAP since it is only used for validation. Therefore, each type in ELCAP is augmented according to the original count, but finally resulting in almost 690 in total.

## 3. Methods

### *3.1. Overview*

In our approach, we are adopting a multi-view multi-scale CNNs based approach to overcome the shortcomings of hand-crafted features. Instead of calculating nodule features from segmented CT images, we are using an icosahedron

based sampling pattern to encode CT values for both nodules and their surrounding areas. Meanwhile, we have also added a multi-scale way to ensure nodules and their surroundings are both captured, as well as a high frequency content based view sorting method to output more informatics to CNNs compared with classical fixed sampling patterns. In addition, we have applied a linear color projection method to re-sliced CT values to preserve original image characteristics as much as possible.

The pipeline of the proposed method comprises five steps. First, we preprocess input CT data using a linear interpolation method (Sec. 3.2). Second, we construct a normalized sphere partitioned by icosahedra at the center of nodules and sample the volume using concentric circle planes whose normal vectors are from the nodule center to icosahedron inner centers (Sec. 3.3). Third, we use a threshold approach to estimate the nodule radii, with which we re-sample the nodule volume (Sec. 3.4). Fourth, with estimated radii, we re-sample the data again in the same ways as that in the second step but at different scales and compute the high frequency content to sort views depending on their importance for each scale (Sec. 3.5). Finally, selected views at all scales are used to first pre-train an independent CNN and then fine-tune it in a multi-view fashion using maximum pooling, resulting in a multi-view multi-scale network (Sec. 3.6).

Compared with current methods, the proposed method has several obvious advantages. In processing steps, the icosahedron-partitioned sphere ensures normal sampling across nodule regions. Meanwhile, we are building multi-scale views based on approximated radius, making built views capture not only nodules but also nodule-surrounding anatomical structures, which would be of great importance for classification of pleural-type nodules. Furthermore, instead of fixed views, the view-sorting procedure based on high frequency content analysis is forwarding dynamic views to CNNs according to their importance, which could afford more informative details for correct classification.

### *3.2. Preprocessing*

The original CT images are with various imaging qualities and different inner plane spacings, making these data anisotropic. Therefore, before all training and testing processes, we first re-sample the original images to a unified inner plane spacing as 1 mm linearly. Then, all cases are clamped to [-1000,1000] Hounsfield Unit (HU) to remove noises whose CT values definitely do not belong to nodules. These preprocessing steps are essential especially when encountering terrible imaging qualities and help to reduce artifacts and noises on CT images.

### *3.3. Normalized Spherical Sampling*

Sampling is the very first and important step in our method. It is commonly agreed that nodule types are related to not only nodule intensity distributions, but also nodule surrounding anatomical structures. Therefore, we are applying a sampling method based on the icosahedron divided sphere which is described in Fig. 3. The schematic representation of the normalized spherical sampling is described in Fig. 4. The algorithm for sampling is presented in Algorithm. 1.

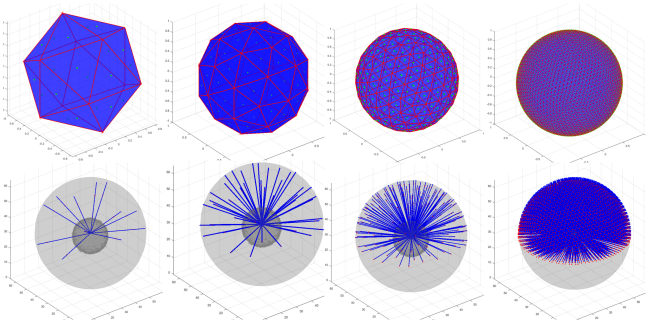

Figure 3: Icosahedron based sampling sphere division. Top row, icosahedra with different levels; from left to right, original , level 1, level 2, level 4 icosahedra; Icosahedron edges are in red and inner centers in green dots. Bottom row, generated sampling planes (views) corresponding to top row, normal vector for each sample plane is from the center to the icosahedron inner center, described using lines colored in blue. Due to symmetry, only half inner centers are used in bottom row.

Denote the candidate nodule region as $V$ and the center of nodule as $v_{xyz}$. First, we divide the sphere containing $V$ using icosahedra and compute inner

**Algorithm 1** Normalized Sphere Sampling

**Input:**

$V$, nodule volume. $r$, maximum sampling radius.

$N$ / $M$ / $K$, specified parameters.

**Output:**

$S$, sampled matrix with size $N \times M \times K$ for $V$.

**function** SPHERE_SAMPLING

    $C \leftarrow$ generate $N$ inner centers

    $\triangle r \leftarrow r/M$

    **for** $n = 1 \rightarrow N$, $m = 1 \rightarrow M$, $k = 1 \rightarrow K$ **do**

        $r_{cur} \leftarrow \triangle r \times m$

        $S_{nmk} \leftarrow linear\_interp(V, n, m, k)$

    **end for**

    **return** $S$

**end function**

centers $C = \{c_1, c_2, ..., c_N\}$ for each triangles on the divided mesh surface as described in Fig. 3. These $N$ inner centers spread uniformly on the surface, making a homogeneously sampling in nodule space. Then, $N$ circle planes centered at $v_{xyz}$ with a given maximum sample radius $r_{max}$ and normal vector $normal_n = (c_n - v_{xyz}), n \in \{1, 2, 3, ..., N\}$ are calculated. These circles are used to approximate the sphere in a normalized sampling fashion. After that, we slice the volume $V$ onto these $N$ circle planes (Fig. 4 (a)), and generate $M$ concentric circles for each plane with $r_m = \triangle r \times m$, $m \in \{1, 2, 3, ..., M\}$ and $\triangle r = r_{max}/M$, as described in Fig. 4 (b) and (c). Finally, $K$ points are sampled along counter clockwise direction on each circle, resulting in a vector describing CT values along the corresponding circle, described as $S_{nm} = \{S_{nm1}, S_{nm2}, ..., S_{nmK}\}$ in Fig 4 (d) and (e).

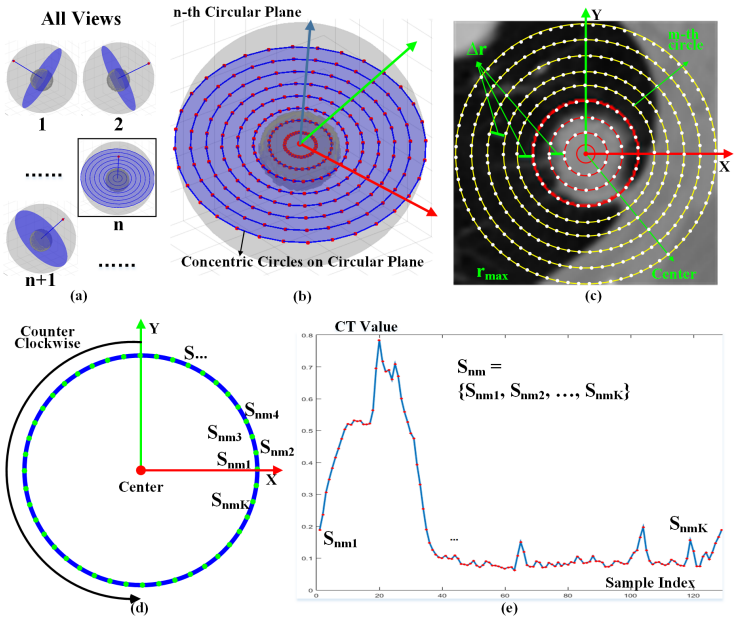


Figure 4: Schematic view of the normalized spherical sampling. (a) All sampling views. (b) The $n$-th sampling circular plane (view) with concentric circles. (c) Sampling pattern for $n$-th plane. Inner and outer nodule sampling are described using red and blue circles separately. (d) For the $m$-th circle, $K$ points (white dots) are sampled in a counter clockwise direction. (e) Corresponding sampled values.

### *3.4. Nodule Radii Estimation*

It is essential to confirm the volume of interest (VOI) to extract features more accurately. During sampling, it can be observed that the intensity value distribution of nodules on different sampled circles have distinguished characteristics. Intensity values for an inner circle are relatively high and stable; values for an outer circle are somehow low but also stable; values for circles at the boundary are very unstable and can be from very high to very low. In our approach, we apply a threshold operation on sampled data to estimate nodule radii. The algorithm is presented in Algorithm 2.

With sampled CT values $S$, $N$ circular planes generated by icosahedra and $M$ concentric circles on each plane, we first calculate count of the $K$ sampled values greater than the given *threshold* (Algorithm 2 Ln.6 - Ln.15). This leads

to a counter matrix $Counter$ with size $N \times M$. Then, we accumulate the counter through all the $N$ circular planes for each concentric circle, leading to a vector $R\_Counter$ with size $M$ (Algorithm 2 Ln.16 - Ln.22). Finally, the indices of the first zero value and local minima in $R\_Counter$ are extracted as $r_1$ and $r_2$ separately, and the estimated radius is calculated as the minimum between $r_1$ and $r_2$ (Algorithm 2 Ln.23 - Ln.30). Typical estimated nodule radii are described in Fig 1 using red circles.

### *3.5. View Selection*

After we obtain the VOI, the next problem is how we can project the CT data to 2-D images as input for CNNs. Although original CT images are 3-D, they are usually projected to three 2-D fixed planes naming *Axial, Sagittal* and *Coronal* for a better view for radiologists. It is reasonable to have a 2-D projection based on both the complexness of CT images with many anatomical structures and human's better understanding. However, most classical approaches are adopting the same fixed views for every nodule candidate, which are not efficient for all CT images since many pathological areas are irregular and fixed views can not present the best characteristics.

In our approach, we regard the quality of projection as how many irregular distributions, in other words, the variational changes, are presented on the projected image. More irregular distributions captured by projected views should lead to better visualization for the characteristics and morphological features of candidate nodules.

The procedure of view selection is described in Fig. 5, after achieving estimated nodule radii, we re-sample $V$ again with $r_{est}$ as maximum radius. The problem of selecting views is transformed into how to sort the $N$ circular planes according to their importance. Actually, since each circular plane comprises of $M$ concentric sampling circles, planes with more information should be those inside which the sampling circles show more amount of variations. This can be straight-forward solved in frequency domain since variations of signals in a spatial domain imply presence of high frequency components in frequency

**Algorithm 2** Nodule Radius Estimation

**Input:**

$S$, sampled 3-D matrix with size $N \times M \times K$ for $V$.

$N$ / $M$ / $K$, specified parameters.

**Output:**

$r_{est}$, estimated radius for $V$.

**function** RADIUS_ESTIMATION

    $Counter \leftarrow 0$ // Counter for $M$ circles $\times$ $N$ planes

    **for** $n = 1 \rightarrow N$, $m = 1 \rightarrow M$, $k = 1 \rightarrow K$, $S_{nmk} > threshold$ **do**

        $Counter_{nm} = Counter_{nm} + 1$

    **end for**

    $R\_Counter \leftarrow 0$ // Counter for $M$ concentric circles

    **for** $m = 1 \rightarrow M$, $n = 1 \rightarrow N$, $R\_Counter_m < counter\_threshold$ **do**

        $r_1 \leftarrow m$

    **end for**

    $r_2 \leftarrow local\_min(R\_Counter)$

    $r_{est} \leftarrow min(r_1, r_2)$

    **return** $r_{est}$

**end function**

domain.

Therefore, we are applying the high frequency content analysis [44] to be an indicator of importance as,

$$freq_n = \sum_{m=1}^{M} D_C(S_{nm}), \tag{1}$$

where $D_C(S_{nm})$ is the complex domain difference between target and observed Short Time Fourier Transform (STFT) and defined as,

$$D_c(S_{nm}) = \frac{1}{K} \sum_{k=1}^{K} |\widehat{X}_k - X_k|, \tag{2}$$

where $X_k$ is the STFT value of $S_{nmk}$ and $\widehat{X}_k$ is the polar form of $X_k$, $K$ is sampling count for each concentric circle.

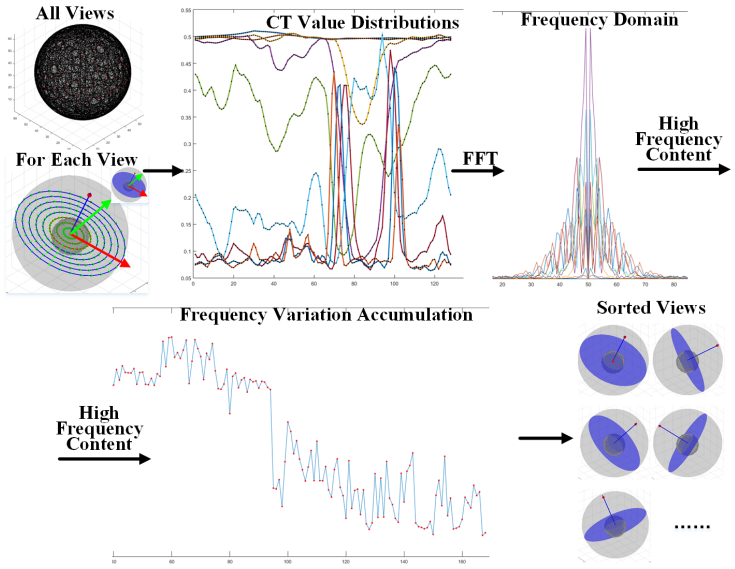


Figure 5: Schematic view of the view sorting procedure. Analysis for one view is shown. Fast Fourier Transform (FFT) is applied to sampled CT values to transform them into frequency domain. Then, computed high frequency contents are accumulated towards each view and used as sign of variations. Finally, all views are sorted based on their corresponding frequency variations.

The procedure is described in Algorithm 3. With re-sampled intensity distribution $S$, we compute the STFT and frequency for data grouped by $K$ sampling points for each $m$-th ($m \in 1, 2, ..., M$) concentric circle on the $n$-th ($n \in 1, 2, ..., N$) circular plane (Algorithm 3 Ln.7 - Ln.13). Then we sum up the frequency for all $M$ circles on the $n$-th plane and use this accumulated frequency as the indicator for this view. Finally, all $N$ views are sorted by their frequencies from high to low, indicating from most important to less.

### *3.6. CNNs Input Generation and Training*

Multi-view CNNs has been proven efficient in 3-D object classification [45]. However, different from object classification, it is a common sense that nodule type classification is related to not only nodule shapes and intensity distributions, but also positions and surrounding anatomical structures. Thus, following the

multi-view approach, with estimated radius $r_{est}$ achieved in Sec. 3.4 and sorted $Views$ in Sec. 3.5, we design the CNNs for original volume $V$ in a multi-view multi-scale fashion to capture both nodule and its surrounding characteristics. The process is presented in Algorithm 4.

Although all views are acquired according to their significance, another big difference between common computer vision and medical image analysis is from the characteristics of imaging protocol. CT images are gray-scale while each pixel is 12-bit, usually holding a value ranging from $-1024$ HU to $+3071$ HU. Specific to lung cancer diagnosis, we are using values within [-1000,1000] HU. However, this is also out of the ability of normal gray-scale images holding values from 0 to 255 ($2000 \rightarrow 255, \sim 10x$) . Therefore, we use sampled concentric circles from $S$ at specified $Views$ to re-build the images; then, we adopt a color bar with 2000 color values smoothly changing from blue to red and projecting values [-1000,1000] HU to these colors to generate RGB colored 2-D images. Each intensity value on CT image is correspond to one specific RGB color. This projection will encode original data in CT images as much as possible and projected color images will definitely hold more information compared with normal gray-scale images. The processed 2-D images are shown in Fig 6.

In our approach, we are first pre-training independent CNNs and then fine tuning the pre-trained model with a maximum pooling layer to take all views and scales into consideration. Totally, for each nodule candidate, we are adopting 3 scales, while 4 views for each scale, leading to 12 views finally. The architecture of the proposed network is presented in Fig. 7. At the end of the proposed CNNs, we are using a soft-max layer to process prior $6 \times 1$ fully connected layer and decide the final classification results.

The characteristics of our proposed network lies in two aspects. In pre-training stage, the CNNs will adjust weights according to the input twelves images separately, which summarizes the common nodule type information in each 2-D slice image at one single scale and view. After that, all 2-D slice images corresponding to the same nodule will be feed in together, processed by several convolutional layers to extract feature maps and finally pooled by the

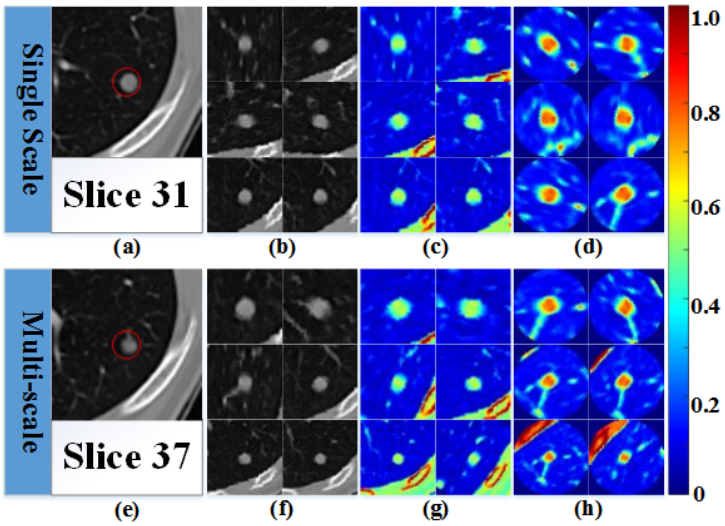


Figure 6: Generated multi-scale inputs for proposed CNNs. (a) and (e): axial view of original CT images on slice 31 and 37. (b) and (f): 2-D images sliced from original volume and CT values compressed to gray-scale images. (c) and (g): same with (b) and (f), but images are re-built using color projections. (d) and (h): re-built images using concentric sampling data and color projection. The color map used is also shown on right.

**Algorithm 3** High Frequency Content Analysis

**Input:**

$S$, $C$, $N$ / $M$ / $K$

**Output:**

$Views$, sorted views corresponding index in $C$.

**function** HIGH_FREQ_ANALYSIS

    $Freq\_all \leftarrow 0$

    **for** $n = 1 \rightarrow N$, $m = 1 \rightarrow M$ **do**

        $this\_freq \leftarrow freq\_analysis(S_{nm})$

        $Freq\_all_n \leftarrow Freq\_all_n + this\_freq$

    **end for**

    $Views \leftarrow sort(Freq\_all)$

    **return** $Views$

**end function**

**Algorithm 4** Generate CNNs Images

**Input:**

$V$, candidate volume. $Scales$, given multi-scales.

$r$, maximum sampling radius.

**Output:** $CNNImages$, input images for CNNs.

**function** GENERATE_CNN_VIEWS
    $S \leftarrow$ SPHERE_SAMPLING($V$, $r$)
    $r_{est} \leftarrow$ RADIUS_ESTIMATION($S$)
    **for** *scale* in $Scales$ **do**
        $S \leftarrow$ SPHERE_SAMPLING($V$, $r_{est} \times scale$)
        $Views \leftarrow$ HIGH_FREQ_ANALYSIS($S$)
        $Images \leftarrow$ BUILD_CIRCLES_IMAGES($S, Views$)
        $CNNImages \leftarrow \{CNNImages, Images\}$
    **end for**
    $CNNImages \leftarrow color\_project(CNNImages)$
    **return** $CNNImages$
**end function**

max-pooling operation to collect maximum responses from all slice images. The max-pooling will unite corresponding feature maps of different slice images into single one, which next will be flatten by the fully-connected layer and processed by the soft-max layer for final prediction.

## 4. Experiments and Validations

As described in Sec. 2, we totally select 1738 nodules and 1000 non-nodules from 744 chest CTs in LIDC-IDRI. These nodules and non-nodules are extracted with a $64 \times 64 \times 64$ $mm^3$ volume. 80% of all these candidates are used for training, while 20% are used for testing. Meanwhile, 421 nodules from 46 ELCAP cases are extracted for testing. ELCAP nodules smaller than 3 mm are classified as non-nodules. Data augmentation is done on both LIDC-IDRI and

Table 1: Parameter Selection for the proposed method

| **Parameter** | **Value** | **Description** |
|---|---|---|
| Views | 12 | number of views |
| Scales | 2.5, 3.0, 4.0 | sampling scales |
| $N$ | 168 | number of circular planes |
| $M$ | 32 | number of concentric circles |
| $K$ | 128 | number of sampling points |
| $r_{max}$ | 32 mm | maximum sampling radius |
| $threshold$ | 0.15 (-700 HU) | maximum accepted CT value |
| $counter\_threshold$ | 3200 | minimum count of threshold |

ELCAP and finally we obtain almost 3900 training (all from LIDC-IDRI) and 1600 testing (LIDC-IDRI: $\sim$ 960, ELCAP: $\sim$ 640) cases separately.

### *4.1. Parameter Selection*

The parameter *number of views* has been detailedly discussed in [45]. Following that, we are using three scales [2.5, 3.0, 4.0] in 12 views. The selection of these three scales is under consideration: since estimated radii are not accurate, scale 2.5 is used to capture the whole nodule inside; scale 4.0 is used to capture more surrounding details. In preprocessing, all images are re-sliced to an inner plane space 1 mm; CT values within [-1000,1000] HU are linearly normalized into [0,1]. In sampling stage, we are using a level-2 icosahedron mesh, but only 168 ($N$=168) inner centers are used due to symmetry, leading to a same amount of circular planes; 32 ($M$=32) concentric circles are built for each circular plane; 128 ($K$=128) sampling points are evenly distributed on each concentric circle; maximum sampling radius ($r_{max}$) is set to 32 mm. All default parameters are presented in Table 1.

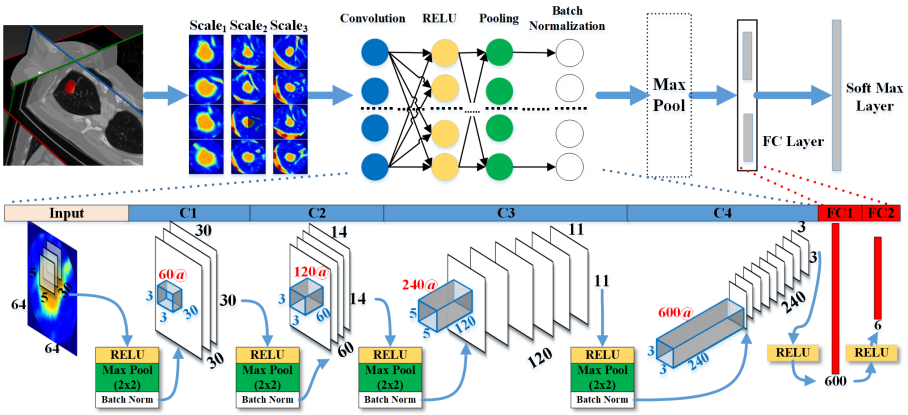


Figure 7: Network architecture of the proposed method. The pre-trained model is without *max pooling* layer to train an independent CNN without affection of views. This model is then fine-tuned by combining all views together through the *max pooling*.

### *4.2. Results*

Classification results for typical cases in each type are shown in Fig. 8. Labeled classification types are labeled out on top. Different boxes are used to indicate different data set LIDC and ELCAP. Both re-built images for CNNs input and original CT image on axial plane are shown. It should be noted that these results are achieved using *msnodulecircles* described in Sec. 4.4. Due to page limit, more results can be found in the supplementary material.

### *4.3. Performance*

It is hard to completely assess a classification model, especially in medical image analysis field without so many public open-access benchmarks like other computer vision fields to compete. However, LIDC-IDRI and ELCAP provide some standards for comparison between different methods.

As described in Sec. 2, we are training the proposed CNNs on LIDC-IDRI and testing the model on both LIDC-IDRI and another independent data set ELCAP. We are testing our model on almost 1000 cases (∼160 cases for each

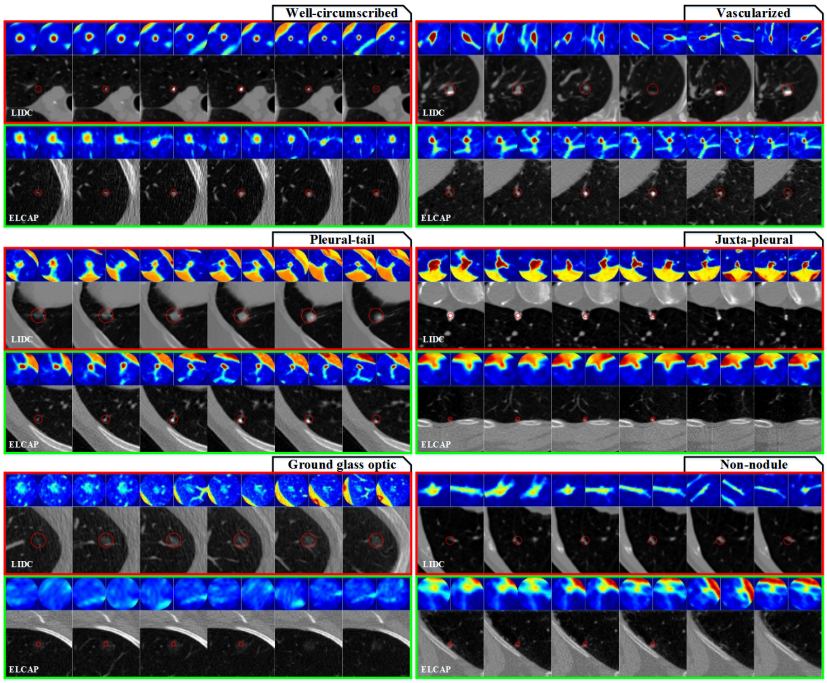


Figure 8: Classification results for typical cases of each type. Classified types are labeled. Red rectangles: data from LIDC-IDRI. Green rectangles: data from ELCAP. In each rectangle, top row: re-built images with color projection; bottom row: original CT images on axial plane with nodule radius labeled in red circles.

type) of LIDC augmented data. The confusion matrix on LIDC-IDRI is presented in Table 2. Classification rate for each type is presented in main diagonal of the confusion matrix. Totally, we achieve a classification rate as 92.1% (932 out of 1012) through all types of candidates. Through the matrix, we can figure out that even though the classification rates for solid nodules (type **W**, **V**, **J**) are high (>90%), more false classifications appear in type GGO. We think this is because of the limitation of count for GGO type nodules.

We have also tested the proposed model on the public accessible data set ELCAP. The confusion matrix is shown in Table. 3. We achieve an overall classification rate at 90.3% (624 out of 691). The testing results show a little difference compared with LIDC-IDRI, which is related to the characteristics of these two sets. Most nodule sizes in ELCAP are under 5 mm, making it even harder to classify them into correct types.

Table 2: Confusion Matrix for Proposed Model on LIDC-IDRI

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **0.83** | 0.02 | 0.13 | 0.02 | 0.00 | 0.00 |
| **W** | 0.01 | **0.97** | 0.01 | 0.00 | 0.02 | 0.00 |
| **N** | 0.00 | 0.01 | **0.99** | 0.00 | 0.00 | 0.00 |
| **P** | 0.00 | 0.00 | 0.10 | **0.90** | 0.00 | 0.00 |
| **V** | 0.01 | 0.06 | 0.01 | 0.00 | **0.92** | 0.00 |
| **J** | 0.00 | 0.00 | 0.09 | 0.01 | 0.00 | **0.90** |

G = Ground glass optic W = Well-circumscribed N = Non-nodule

P = Pleural-tail V = Vascularized J = Juxta-pleural

From Table 3, we can figure out that the proposed model will give a higher response towards small nodules not touching too much with surrounding anatomical structures into non-juxta-pleural types.

It can also observed that classification for juxta-pleural nodules is not as efficient as other types. This has great correlation with nodule sizes based on the fact that nodule radius estimation is not so efficient as other types towards nodules attached on lung plurals when nodules are pretty small.

### *4.4. Comparison*

As described in Sec. 1.1, although much work has been done on lung nodule detection and segmentation, only a few researches are focusing on nodule type classifications. However, comparison between methods is an essential and reasonable step for validating the efficiency of our method. Thus, we are doing comparisons in two ways.

On one way, we have adopted the most similar work in [40], in which they are focusing on classifying **W**, **V**, **P** and **J** type nodules using a patch-based model.

On the other way, we compare different configurations to have an investigation in the effect of the multi-view multi-scale fashion:

Table 3: Confusion Matrix for Proposed Model on ELCAP

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **1.00** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **W** | 0.00 | **0.95** | 0.05 | 0.00 | 0.00 | 0.00 |
| **N** | 0.04 | 0.04 | **0.92** | 0.01 | 0.00 | 0.00 |
| **P** | 0.01 | 0.00 | 0.12 | **0.92** | 0.00 | 0.00 |
| **V** | 0.05 | 0.00 | 0.02 | 0.00 | **0.93** | 0.00 |
| **J** | 0.06 | 0.00 | 0.19 | 0.01 | 0.00 | **0.75** |

- Configuration 1, called as *nodules*. The input 12 views for CNNs comprises of 9 random views from the sorted $Views$ and 3 specific views at center of $v_{xyz}$ on sagi ttal, axial and coronal planes. All views are sliced from original data with linear interpolation and projected to [0,255] gray-scale images.

- Configuration 2, called as *colornodules*. All 12 views are selected from the sorted $Views$, randomly. Meanwhile, sliced data from original CT values are re-projected to color space to generate color images (see Sec. 3.6) as input for CNNs.

- Configuration 3, called as *nodulecircles*. All 12 views are selected from the sorted $Views$ from top to bottom, meaning from more important to less. Instead of using original sliced CT values, the sampled circle data $S$ are projected to color space to build color concentric circle images.

Each configuration is coming with a normal form and a multi-scale (*ms-*) form, resulting in 6 configurations *nodules, ms-nodules, colornodules, ms-colornodules, nodulecircles, ms-nodulecircles*. It should be noted that, all performance results presented in Sec. 4.3 are achieved using *msnodulecircles*. We validate all configurations on both LIDC-IDRI and ELCAP.

Table 4: Confusion Matrix for Comparison with PB [40]

| | **W** Prop / PB | **P** Prop / PB | **V** Prop / PB | **J** Prop / PB |
|---|---|---|---|---|
| **W** | **0.95** / 0.89 | 0.00 / 0.06 | 0.00 / 0.04 | 0.00 / 0.01 |
| **P** | 0.00 / 0.03 | **0.92** / 0.91 | 0.00 / 0.03 | 0.00 / 0.03 |
| **V** | 0.00 / 0.05 | 0.00 / 0.06 | **0.93** / 0.86 | 0.00 / 0.03 |
| **J** | 0.00 / 0.03 | 0.01 / 0.06 | 0.00 / 0.04 | 0.75 / **0.87** |

Prop = proposed method PB = patch based [40] P = Pleural-tail
W = Well-circumscribed V = Vascularized J = Juxta-pleural

*4.4.1. Comparison with other methods*

Zhang et al. [40] proposed a multilevel patch based (PB) nodule classification method and evaluated the classification performance for **W**, **V**, **J** and **P** type noduls on ELCAP data set. Therefore, only performance of ELCAP of our method for these four types is compared. The confusion matrix is presented in Table. 4. It should be noted that, classification results for GGO and non-nodule are not shown, therefore, sum up of each row of the proposed method is not equal to 1.

The overall claimed classification rate of PB [40] is 89% on ELCAP. Meanwhile, even though we are processing two more complex types of nodules (GGO and non-nodule), we also achieve a higher rate at 90.3%. When diving into results for each individual type, we can observe two facts: the proposed method outperforms PB at nodule types **W**, **P** and **V** but falls behind at **J**; most classification errors of the proposed methods are lying in GGO and non-nodule types. If we wipe out these two types, the proposed method is outperforming PB at all four types.

*4.4.2. Comparison among different configurations*

To have a comprehensive understanding of the proposed method and the effect of each step we use, we are validating the method using 3 different con-

figurations with both normal and multi- scale form on LIDC and ELCAP data set, resulting in 6 confusion matrices totally. Each confusion matrix consists of the normal and multi-scale evaluation of one configuration on one data set.

The confusion matrix for *nodules* and *ms-nodules*, *colornodules* and *ms-colornodules*, *nodulecircles* and *ms-nodulecircles* on LIDC-IDRI are presented in Table 5, Table 6 and Table 7 with overall classification rate as 83.1%, 84.1%, 81.1%, 85.9%, 88.2%, 92.1%, separately. Meanwhile, the confusion matrix for *nodules* and *ms-nodules*, *colornodules* and *ms-colornodules*, *nodulecircles* and *ms-nodulecircles* on ELCAP are presented in Table 8, Table 9 and Table 10 with overall classification rate as 79.6%, 86.5%, 84.1%, 84.3%, 84.9%, 90.3%, separately.

Comparing from left to right in the same table, it can be figured out that, multi-scale approach helps raise accuracy in classifying most types. Comparing from Table. 5 to Table. 6 and Table. 8 to Table. 9, it is obvious that color projection from CT values to color images does not indeed improve performance. Meanwhile, comparing from Table. 6 to Table. 7 and Table. 9 to Table. 10, it is obvious that when used combined with circular sampling, the generated color images definitely help to raise accuracy. With all techniques in the multi-view multi-scale fashion, we are achieving an overall classification rate as high as 92.1% (Table. 7 *msnodulecircles*) on LIDC-IDRI and 90.3% (Table. 10 *msnodulecircles*) on ELCAP, denoting the promising performance of our method.

### *4.5. Discussions and Limitations*

Although the proposed method achieves promising results on both LIDC-IDRI and ELCAP data set, the overall classification rate on ELCAP is slightly lower than LIDC-IDRI. The most obvious reason is LIDC-IDRI contains more various testing cases and all cases distribute more normally than ELCAP. However, the other important reason should be nodule sizes on ELCAP are too small to be estimated using threshold based method, leading to wrong classification results. Therefore, a better investigation based on the proposed method against

Table 5: Confusion Matrix for *nodules/ms-nodules* on LIDC-IDRI

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **0.80 / 0.47** | 0.00 / 0.09 | 0.20 / 0.35 | 0.00 / 0.01 | 0.01 / 0.08 | 0.00 / 0.00 |
| **W** | 0.01 / 0.00 | **0.83 / 1.00** | 0.05 / 0.00 | 0.00 / 0.00 | 0.13 / 0.00 | 0.01 / 0.00 |
| **N** | 0.05 / 0.02 | 0.07 / 0.05 | **0.92 / 0.87** | 0.06 / 0.02 | 0.04 / 0.02 | 0.07 / 0.02 |
| **P** | 0.00 / 0.01 | 0.00 / 0.01 | 0.12 / 0.08 | **0.91 / 0.89** | 0.05 / 0.00 | 0.00 / 0.00 |
| **V** | 0.07 / 0.01 | 0.04 / 0.03 | 0.02 / 0.05 | 0.00 / 0.00 | **0.88 / 0.89** | 0.01 / 0.01 |
| **J** | 0.01 / 0.01 | 0.00 / 0.00 | 0.19 / 0.10 | 0.00 / 0.00 | 0.00 / 0.00 | **0.86 / 0.89** |

Table 6: Confusion Matrix for *colornodule/ms-colornodule* on LIDC-IDRI

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **0.62 / 0.65** | 0.00 / 0.00 | 0.38 / 0.35 | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 |
| **W** | 0.00 / 0.02 | **0.86 / 0.89** | 0.03 / 0.04 | 0.00 / 0.01 | 0.11 / 0.04 | 0.00 / 0.00 |
| **N** | 0.00 / 0.01 | 0.02 / 0.03 | **0.94 / 0.95** | 0.02 / 0.00 | 0.00 / 0.01 | 0.03 / 0.00 |
| **P** | 0.02 / 0.01 | 0.12 / 0.04 | 0.07 / 0.03 | **0.79 / 0.93** | 0.00 / 0.01 | 0.00 / 0.00 |
| **V** | 0.02 / 0.03 | 0.07 / 0.06 | 0.05 / 0.02 | 0.03 / 0.00 | **0.83 / 0.89** | 0.00 / 0.00 |
| **J** | 0.00 / 0.00 | 0.00 / 0.00 | 0.23 / 0.20 | 0.00 / 0.02 | 0.00 / 0.00 | **0.76 / 0.78** |

Table 7: Confusion Matrix for *nodulecircles/ms-nodulecircles* on LIDC-IDRI

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **0.80 / 0.83** | 0.01 / 0.02 | 0.20 / 0.13 | 0.00 / 0.02 | 0.00 / 0.00 | 0.00 / 0.00 |
| **W** | 0.01 / 0.01 | **0.97 / 0.97** | 0.03 / 0.01 | 0.00 / 0.00 | 0.00 / 0.02 | 0.00 / 0.00 |
| **N** | 0.00 / 0.00 | 0.01 / 0.01 | **0.98 / 0.99** | 0.01 / 0.00 | 0.00 / 0.00 | 0.01 / 0.00 |
| **P** | 0.01 / 0.00 | 0.00 / 0.00 | 0.14 / 0.10 | **0.86 / 0.90** | 0.00 / 0.00 | 0.00 / 0.00 |
| **V** | 0.01 / 0.01 | 0.01 / 0.06 | 0.05 / 0.01 | 0.00 / 0.00 | **0.94 / 0.92** | 0.00 / 0.00 |
| **J** | 0.00 / 0.00 | 0.00 / 0.00 | 0.21 / 0.09 | 0.00 / 0.01 | 0.01 / 0.00 | **0.79 / 0.90** |

tiny nodules ($\leq$ 3mm) will draw our more attention in the future.

Table 8: Confusion Matrix for *nodules/ms-nodules* on ELCAP

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **0.74 / 0.81** | 0.21 / 0.02 | 0.05 / 0.17 | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 |
| **W** | 0.00 / 0.02 | **0.61 / 0.86** | 0.37 / 0.03 | 0.00 / 0.01 | 0.01 / 0.08 | 0.01 / 0.01 |
| **N** | 0.00 / 0.00 | 0.00 / 0.00 | **1.00 / 1.00** | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 |
| **P** | 0.00 / 0.08 | 0.00 / 0.00 | 0.23 / 0.17 | **0.77 / 0.75** | 0.00 / 0.00 | 0.00 / 0.00 |
| **V** | 0.03 / 0.01 | 0.05 / 0.01 | 0.14 / 0.09 | 0.00 / 0.00 | **0.78 / 0.82** | 0.00 / 0.06 |
| **J** | 0.01 / 0.02 | 0.00 / 0.00 | 0.40 / 0.16 | 0.01 / 0.00 | 0.00 / 0.00 | **0.58 / 0.81** |

Table 9: Confusion Matrix for *colornodules/ms-colornodules* on ELCAP

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **0.57 / 0.79** | 0.00 / 0.00 | 0.43 / 0.21 | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 |
| **W** | 0.03 / 0.01 | **0.84 / 0.85** | 0.06 / 0.09 | 0.00 / 0.00 | 0.07 / 0.04 | 0.00 / 0.01 |
| **N** | 0.01 / 0.00 | 0.01 / 0.00 | **0.97 / 1.00** | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 |
| **P** | 0.00 / 0.09 | 0.02 / 0.00 | 0.18 / 0.11 | **0.79 / 0.79** | 0.02 / 0.02 | 0.00 / 0.00 |
| **V** | 0.04 / 0.02 | 0.07 / 0.06 | 0.09 / 0.06 | 0.00 / 0.00 | **0.79 / 0.83** | 0.01 / 0.03 |
| **J** | 0.00 / 0.01 | 0.00 / 0.00 | 0.28 / 0.36 | 0.00 / 0.00 | 0.02 / 0.01 | **0.70 / 0.62** |

Table 10: Confusion Matrix for *nodulecircles/ms-nodulecircles* on ELCAP

| | **G** | **W** | **N** | **P** | **V** | **J** |
|---|---|---|---|---|---|---|
| **G** | **1.00 / 1.00** | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 | 0.00 / 0.00 |
| **W** | 0.04 / 0.00 | **0.92 / 0.95** | 0.03 / 0.05 | 0.00 / 0.00 | 0.01 / 0.00 | 0.00 / 0.00 |
| **N** | 0.10 / 0.04 | 0.00 / 0.04 | **0.90 / 0.92** | 0.00 / 0.01 | 0.00 / 0.00 | 0.00 / 0.00 |
| **P** | 0.04 / 0.02 | 0.01 / 0.00 | 0.11 / 0.06 | **0.83 / 0.92** | 0.01 / 0.00 | 0.00 / 0.00 |
| **V** | 0.10 / 0.05 | 0.00 / 0.00 | 0.09 / 0.02 | 0.00 / 0.00 | **0.82 / 0.93** | 0.00 / 0.00 |
| **J** | 0.02 / 0.05 | 0.01 / 0.00 | 0.32 / 0.20 | 0.02 / 0.00 | 0.00 / 0.00 | **0.63 / 0.76** |

## 5. Conclusion

In this paper, we have devised a multi-view multi-scale CNNs model for lung nodule type classification. We employed concentric circles and icosahedra to conduct sampling over nodule volume normally. A threshold-based method is used to approximate nodule radii. We made use of high frequency content analysis to select the views. Finally, the built multi-view, multi-scale, color-projected images are used to first pre-train an independent CNNs and then fine-tune the pre-trained model with maximum pooling. The overall classification rates have reached 92.3% for almost 1000 cases on LIDC-IDRI and 90.3% for 690 cases on ELCAP, and such impressive results have shown the promising performance of the proposed method. It is our hope that our newly-proposed method could be of great assistance for radiologists to diagnosis of lung cancer in clinical practice. Future work will lie in two aspects. One is a scale in-variant and more robust radius estimation method towards very tiny nodules (radius $\leq$ 3 mm) and juxta-pleural nodules. The other is automatically labeling out types, positions and sizes for nodules not centered in images with less human interactions.

## Acknowledgments

This research is supported in part by National Natural Science Foundation of China (Grant No. 61532002, 61672149, 61672077), National Science Foundation of USA (Grant No. IIS-1715985).

## References

[1] G. Armato III, Samueland McLennan, et al., The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans, Med. Phys. 38 (2) (2011) 915–931.

[2] Vison, I. A. Group, I-elcap: International early lung cancer action program (2013).

[3] R. Siegel, K. Miller, A. Jemal, Cancer statistics, 2017, CA-CANCER J. Clin.

[4] D. Henschke, D. Yankelevitz, D. Naidich, G. McGuinness, O. Miettinen, D. Libby, M. Pasmantier, J. Koizumi, N. Altorki, J. Smith, Early Lung Cancer Action Project: Overall design and findings from baseline screening, Lancet 354 (9173) (1999) 99–105.

[5] J. Erasmus, J. Connolly, H. McAdams, V. Roggli, Solitary pulmonary nodules: Part i. morphologic evaluation for differentiation of benign and malignant lesions, Radiographics 20 (1) (2000) 43–58.

[6] D. Wu, L. Lu, J. Bi, Y. Shinagawa, K. Boyer, A. Krishnan, M. Salganicoff, Stratified learning of local anatomical context for lung nodules in ct images, in: Proc. CVPR, 2010, pp. 2791–2798.

[7] Y. Song, W. Cai, Y. Zhou, D. Feng, Feature-based image patch approximation for lung tissue classification, IEEE Trans. Med. Imag. 32 (4) (2013) 797–808.

[8] Q. Gao, S. Wang, D. Zhao, J. Liu, Accurate lung segmentation for X-ray CT images, in: Proc. ICNC, Vol. 2, 2007, pp. 275–279.

[9] E. van Rikxoort, B. de Hoop, M. Viergever, M. Prokop, B. van Ginneken, Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection, Med. Phys. 36 (7) (2009) 2934–2947.

[10] A. Kumar, J. Kim, W. Cai, M. Fulham, D. Feng, Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data, J. Digit. Imaging 26 (6) (2013) 1025–1039.

[11] J. Galaro, A. Judkins, D. Ellison, J. Baccon, A. Madabhushi, An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas, in: Proc. EMBC, 2011, pp. 3443–3446.

[12] F. Zhang, Y. Song, W. Cai, Y. Zhou, M. Fulham, S. Eberl, S. Shan, D. Feng, A ranking-based lung nodule image classification method using unlabeled image knowledge, in: Proc. ISBI, 2014, pp. 1356–1359.

[13] Y. Song, W. Cai, J. Kim, D. D. Feng, A multistage discriminative model for tumor and lymph node detection in thoracic images, IEEE Trans. Med. Imag. 31 (5) (2012) 1061–1075.

[14] L. Sorensen, S. Shaker, M. de Bruijne, Quantitative analysis of pulmonary emphysema using local binary patterns, IEEE Trans. Med. Imag. 29 (2) (2010) 559–569.

[15] F. Ciompi, C. Jacobs, E. Scholten, M. Wille, P. de Jong, M. Prokop, B. van Ginneken, Bag-of-frequencies: A descriptor of pulmonary nodules in computed tomography images, IEEE Trans. Med. Imag. 34 (4) (2015) 962–973.

[16] Y. Gu, V. Kumar, L. O. Hall, D. B. Goldgof, C.-Y. Li, R. Korn, C. Bendtsen, E. R. Velazquez, A. Dekker, H. Aerts, P. Lambin, X. Li, J. Tian, R. A. Gatenby, R. J. Gillies, Automated delineation of lung tumors from ct images using a single click ensemble segmentation approach, Pattern Recognition 46 (3) (2013) 692 – 702.

[17] C. Jacobs, E. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J. M. Kuhnigk, et al., Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images, Med. Imag. Anal. 18 (2) (2014) 374–384.

[18] M. Firmino, A. Morais, R. Mendoça, M. Dantas, H. Hekis, R. Valentim, Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects, BioMedical Engineering OnLine 13 (2014) 41.

[19] M. Abdel-Nasser, A. Moreno, H. A. Rashwan, D. Puig, Analyzing the evolution of breast tumors through flow fields and strain tensors, Pattern Recognition Letters 93 (Supplement C) (2017) 162 – 171, pattern Recognition Techniques in Data Mining.

[20] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[21] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR abs/1409.1556.

[22] T. Brosch, Y. Yoo, D. K. Li, A. Traboulsee, R. Tam, Modeling the variability in brain morphology and lesion distribution in multiple sclerosis by deep learning, in: Proc. MICCAI, Part II, 2014, pp. 462–469.

[23] S. S. Parveen, C. Kavitha, A review on computer aided detection and diagnosis of lung cancer nodules, International Journal of Computers & Technology 3 (3).

[24] K. L. Hua, C. H. Hsu, S. C. Hidayati, W. H. Cheng, Y. J. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, Oncotargets & Therapy 8 (2015) 2015–2022.

[25] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imag. 35 (5) (2016) 1285–1298.

[26] S. Lo, H. Chan, J. Lin, H. Li, M. Freedman, S. Mun, Artificial convolution neural network for medical image pattern recognition, Neural Netw. 8 (7) (1995) 1201–1214.

[27] A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. van Riel, M. Wille, M. Naqibullah, C. Snchez, B. van Ginneken, Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks, IEEE Trans. Med. Imag. 35 (5) (2016) 1160–1169.

[28] H. R. Roth, L. Lu, A. Seff, K. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R. Summers, A new 2.5d representation for lymph node detection using random sets of deep convolutional neural network observations, in: Proc. MICCAI, Part I, 2014, pp. 520–527.

[29] P. Cao, X. Liu, J. Yang, D. Zhao, W. Li, M. Huang, O. Zaiane, A multi-kernel based framework for heterogeneous feature selection and oversampling for computer-aided detection of pulmonary nodules, Pattern Recognition 64 (Supplement C) (2017) 327 – 346.

[30] L. Wei, C. Peng, D. Zhao, J. Wang, Pulmonary nodule classification with deep convolutional neural networks on computed tomography images, Computational and Mathematical Methods in Medicine 2016 (2016) 1–7.

[31] P.-L. Lin, P.-W. Huang, C.-H. Lee, M.-T. Wu, Automatic classification for solitary pulmonary nodule in ct image by fractal analysis based on fractional brownian motion model, Pattern Recognition 46 (12) (2013) 3279 – 3287.

[32] K. Liu, G. Kang, Multiview convolutional neural networks for lung nodule classification, International Journal of Imaging Systems & Technology 27 (1) (2017) 12–22.

[33] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, J. Tian, Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification, Pattern Recognition 61 (Supplement C) (2017) 663 – 673.

[34] H. Jiang, H. Ma, W. Qian, M. Gao, Y. Li, An automatic detection system of lung nodule based on multi-group patch-based deep learning network, IEEE J. Biomed. Health PP (99) (2017) 1–1.

[35] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, S. Mougiakakou, Multisource transfer learning with convolutional neural networks for lung pattern analysis, IEEE J. Biomed. Health 21 (1) (2017) 76–84.

[36] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, S. Mougiakakou, Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, IEEE Trans. on Med. Imag. 35 (5) (2016) 1207–1216.

[37] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, J. Z. Cheng, Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in ct images, IEEE Trans. on Med. Imag. 36 (3) (2017) 802–814.

[38] A. Farag, A. Ali, J. Graham, A. Farag, S. Elshazly, R. Falk, Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose ct scans of the chest, in: Proc. ISBI, 2011, pp. 169–172.

[39] F. Zhang, W. Cai, Y. Song, M. Z. Lee, S. Shan, D. Dagan, Overlapping node discovery for improving classification of lung nodules, in: Proc. EMBC, 2013, pp. 5461–5464.

[40] F. Zhang, Y. Song, W. Cai, M. Lee, Y. Zhou, H. Huang, S. Shan, M. Fulham, D. Feng, Lung nodule classification with multilevel patch-based context analysis, IEEE Trans. Biomed. Eng. 61 (4) (2014) 1155–1166.

[41] Y. Song, W. Cai, Y. Wang, D. D. Feng, Location classification of lung nodules with optimized graph construction, in: Proc. ISBI, 2012, pp. 1439–1442.

[42] R. Wiemker, M. Bergtholdt, E. Dharaiya, S. Kabus, M. Lee, Agreement of cad features with expert observer ratings for characterization of pulmonary nodules in ct using the lidc-idri database, in: Proc. SPIE, Vol. 7260, 2009, pp. 72600H–72600H–12.

[43] C. Ken, S. Karen, V. Andrea, Z. Andrew, Return of the devil in the details: Delving deep into convolutional nets, CoRR abs/1405.3531.

[44] P. Brossier, J. Bello, M. Plumbley, Real-time temporal segmentation of note objects in music signals, Proc. ICMC.

[45] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller, Multi-view convolutional neural networks for 3d shape recognition, in: Proc. ICCV, 2015, pp. 945–953.