Pulmonary Nodule Detection Using V-Net and High-Level Descriptor Based SVM Classifier

YUYUN YE, MIAO TIAN (Member, IEEE), QIYU LIU, AND HENG-MING TAI (Senior Member, IEEE)

1Department of Electrical and Computer Engineering, The University of Tulsa, Tulsa, OK 74104, USA
2School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
3Department of Radiology, Mianyang Central Hospital, Mianyang 621000, China

Corresponding author: Miao Tian (ziguiren@uestc.edu.cn)

ABSTRACT Early detection of the pulmonary nodule is critical to increase the five-year survival rate of lung cancer. Many computer-aided diagnosis (CAD) systems have been proposed for nodule detection to assist radiologists in diagnosis. Along this direction, this paper proposes a novel automated pulmonary nodule detection model using the modified V-Nets and a high-level descriptor based support vector machine (SVM) classifier. The former is for nodule candidate detection and the latter is for false positive (FP) reduction. A hard mining scheme for retraining is devised to improve the FP reduction performance. The proposed SVM classifier, which employs more critical features of CT images, performs superior in FP reduction than other SVM based classifiers and CNN classifiers. Experimental results using the LIDC-IDRI dataset are presented to demonstrate the effectiveness of the proposed CAD model.

INDEX TERMS Convolutional neural networks (CNNs), pulmonary nodule detection, lung CT image, classification, support vector machine (SVM).

I. INTRODUCTION Pulmonary cancer is a leading cause of cancer death in the world. Global cancer statistics reported that nearly 1.83 million new cases of lung cancer occurred and the estimated annual deaths are over 1.5 million [1]. Pulmonary cancers have five-year survival rate between 10% and 16% [2]. The survival rate is low because most lung cancers are found mostly at advanced stages, at which cancer cells have already spread widely. If the lung nodule can be detected in the early stage, then the overall five-year survival rate would increase to 54%. Thus, early detection of the nodule can be crucial in improving the five-year survival rate of pulmonary cancer.

Nowadays, many imaging modalities such as the computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI), have been used by radiologists for pulmonary nodules detection. Of these imaging modalities, CT has been adopted as the standard screening tool for pulmonary cancer diagnosis because of being cost effective and widely available [3]. In pulmonary nodule detection from CT scan, radiologists read the images, extract suspicious pulmonary nodules, and label the possible malignancy based on the nodule information which includes morphology, shape, size, and textural features. The patient’s treatment plan is determined directly by the diagnosis result. But many factors, such as distraction, fatigue, and experience, may bring about the misinterpretation of available data. Thus, it is crucial to have a computer-aided diagnosis (CAD) tool providing radiologists of assistance and helping increase the detection accuracy.

CAD systems provide the detection result which can be used as the second diagnosis opinion before radiologists make the final decision. In the pulmonary cancer diagnosis, accurate automated pulmonary nodule detection in CT scan plays an important role in the CAD field to serve radiologists. But it is very challenging to design a reputable pulmonary nodule detection system because pulmonary nodules have diverse sizes, shapes, locations, and types, as well as the abundance of tissues, e.g., blood vessels, chest wall, that resemble the appearance of nodules. Pulmonary nodule usually has a size about 5 mm to 30 mm in diameter. It can be discriminated based on its size (large or medium or small), location (well circumscribed, juxta-pleural, and juxta-vascular), shape (ball-like and irregular), and internal texture (solid, part-solid and nonsolid). Fig. 1 shows various marked lung nodule examples.
Most pulmonary nodule detection systems perform two tasks: nodule candidate detection and false positive (FP) reduction. The former aims to identify suspicious nodule candidates in the CT scan images. Any nodules not detected at this stage will no longer be retrieved later. Thus, the first stage usually allows to have very high sensitivity which comes with the cost of many false positive cases. Thus, the latter process is designed to reduce the number of FPs.

Conventional pulmonary nodule detection methods usually adopt complicated image processing algorithms, such as image transformation, enhancement, and segmentation, to detect the lung parenchyma, select nodule candidates, extract distinctive features, and perform the false positive reduction. Owing to their heavy dependence on hand-crafted features, these conventional methods still have room for improvement [4]. Recently, deep learning techniques have been introduced to medical image analysis for a variety of applications with excellent results [5], [6]. Convolutional neural networks (CNNs), in particular, have attracted the attention of many researchers [5]–[7]. CNNs can be trained and learn distinct features that may not be interpreted by radiologists, without the handcrafting descriptors.

Most CNNs used in the nodule candidate detection stage are 2D CNNs [5], [7]. But 2D CNNs may not exploit well the 3D data. Some variants of 2D CNNs attempt to introduce multi-view planes [5], [8] or hand-crafted 3D features to aggregate the 3-D spatial information. However, due to the nature of 2D network architecture, these 2D CNNs variants still not sufficiently exploit 3D spatial information within volumetric data. A variety of 3D CNNs and their variants have been proposed to overcome the limitations of 2D CNNs [9], [10].

In FP reduction, few 2D CNNs has been adopted because of the poor discrimination ability. 3D CNNs are most common in reducing the FP cases [11]–[13]. However, they may fail to correctly capture all distinct features of a nodule in the CT scan, due to the high variability of shape, size, and texture of the nodule. Thus, several high-level descriptor based machine learning techniques have been developed for effective FP reduction [14]–[20].

This paper proposes a pulmonary nodule detection model using the modified V-Net and a high-level descriptor based SVM classifier. To improve the detection performance, the former is combined with the multilevel receptive fields [8] and the latter is paired with the hard mining algorithm. To the best of our knowledge, this is the first time the V-Net is used in pulmonary nodule detection. In addition, the descriptor SVM classifier considers the wavelet feature of CT image, which has not been used in the existing classifiers.

The rest of this paper is organized as follows. Section II describes related works on pulmonary nodule detection. Section III elaborates detailed operation of the proposed nodule detection model, including the description of modified V-Net and a retraining scheme for improvement of false positive reduction. Simulation results on LIDC-IDRI dataset are given in Section IV to demonstrate the effectiveness of the proposed model. Conclusions and future research are given in Section V.

II. RELATED WORKS

Many CAD systems for pulmonary nodule detection have been proposed in the last decade [5], [6], [11], [15], [16], [21]–[24]. These conventional approaches generally adopt the voxel clustering and pixel threshold techniques using the handcrafted features such as morphological features. For instance, the shape and curve index were used for nodule candidate detection and the k-nearest neighbor algorithm for FP reduction [5]. Image enhancement, segmentation, and feature extraction were also used for nodule candidate detection and the Weber local descriptor (WLD) based SVM for FP reduction [6].

Recently, deep learning models have shown promising results in the extraction of image features for successful detection and classification of CT scan images [9], [10]. A variety of CNN models, including R-CNN [7], R-FCN [10], hierarchical semantic CNN [25], and Mask R-CNN [26], enable the generation of nodule candidate regions with higher precision and less time than traditional approaches. CNNs also can be used as classifiers to reduce the false positives [23].

3D CNNs are increasingly used for nodule candidate detection due to that chest CT scans are inherently 3D images. 3D CNNs may have better ability to model spatial information due to its intrinsic 3D convolution and 3D pooling operations. For example, a de-convolutional layer was incorporated into a 3D Faster R-CNN for lung nodule candidate detection [9].

3D CNN based deep learning models have also been used for FP reduction. For example, a residual multiple view 3D CNN was proposed to extract the 3D patch [13]; a model comprising three 3D CNNs and multi-level contextual 3D CNNs were used to cope with the nodule variations [12]. False positives also can be reduced by a 2D CNN extension using 3D kernels [11], by a 3D CNN network with dual pooling structure [25], a nodule-size-adaptive 3D CNN [26] or a maximum intensity projection based CNN [28].

Image features of pulmonary nodules, such as its shape, texture, and intensity are critical to the classification and reduction of FP cases. 3D CNNs may not capture the whole discriminative features because of high variability on these features. Several image processing techniques have been proposed to capture as many features as possible [15], [16].

FIGURE 1. Examples of nodules with various size (large or medium or small), location (well circumscribed, juxta-pleural and juxta-vascular), shape (ball-like and irregular), and internal texture (solid, part-solid and nonsolid).
These object detection techniques include linear interpolation, K-means, median filtering, morphological operation, and so on. SVM based classifiers also have been proposed to exploit a variety of features of CT images for FP reduction [17], [20]. These SVM classifiers employed the WLD features [14], the HOG and LBP features [17], intensity, shape, texture features [20], and GLCM [19].

CT images contain critical features, which affect greatly the classification accuracy and performance of FP reduction. These features are shape, texture, and wavelet coefficient. Most of the existing methods consider only part of these critical features. In this paper, a SVM classifier incorporating high-level descriptor is proposed to extract the complete critical features. Moreover, a hard mining scheme is used to improve the performance by focusing on samples that are difficult to classify [5]. Hard mining puts more weights on the erroneous samples.

**III. PROPOSED LUNG NODULE DETECTION MODEL**

Pulmonary nodule detection model generally performs two main functions: nodule candidate detection and false positive reduction. Fig. 2 depicts the proposed module detection structure. At the nodule candidate detection stage, three modified V-Nets with multilevel receptive fields are used. Each receptive field encodes a specific scale of contextual information. The outputs are obtained by combining the results of these nets. At the FP reduction stage, a high-level descriptor based SVM classifier trained by a hard mining algorithm is developed to identify whether the candidate images actually contain the nodules.

V-Net is a 3D CNN and has been used in the prostate MRI image segmentation [29]. It extends the functionality of U-Net [30] to 3D images. V-Net performs better in terms of nodule candidate selection than other CNNs because it solves the imbalance problem between background and foreground pixels in CT images. V-Net calculates the loss using a dice coefficient based loss function, which ignores large part of background pixels so as to achieve the balance between amounts of background and foreground pixels [29].

Moreover, different receptive fields strengthen V-Net’s capability in handling drastic variations of the nodules. The size of receptive field plays a crucial role in the discrimination performance of CNN [8]. If too small, only limited contextual information will be extracted to train the network and the recognition capability would be deficient to handle large variations of nodules. If too large, more redundant messages and strong unbalance between foreground and background information would be involved in the training. This degraded the performance, especially when the number of training samples is limited [29].

Due to the variety of pulmonary nodules, CNNs can reach high sensitivity, but at the expense of redundant false positive cases. Many works focus on exploiting new CNN models to reduce the number of FPs [8], [11]. CNNs, however, may not capture all the crucial nodule features. This paper proposes a high-level descriptor based SVM classifier to fill the gap.

Sample results of the proposed nodule detection model are shown in Fig. 3. The top panel displays the patches extracted from six CT scan images. The first four have the nodule marked by radiologists and the last two have nodule like lesion not marked by radiologists. The middle panel depicts the corresponding patches after the nodule candidate detection stage. The bottom panel shows the patches with marked nodules after false positive reduction. It can be seen from the bottom panel of Fig. 3 that two patches are classified as FP cases after FP reduction. This result agrees with that by the radiologists (top panel). This shows that FP reduction is critical to the accuracy of lung nodule detection.

**A. MODIFIED V-Net**

Fig. 4 depicts a schematic diagram of the modified V-Net. The input to V-Net is a CT image of size $16 \times 512 \times 512$. The output is also an image of size $16 \times 512 \times 512$. The pulmonary nodules are marked if there is any. The upper part of V-Net is comprised of a compression path, while the lower part is the decompression path. The orange box shows the residual connections and the purple box the convolutional block.

The convolutions performed at each stage use the volumetric kernels having $5 \times 5 \times 5$ voxels. As the data proceeds along the compression path, its resolution is reduced via the convolution with $2 \times 2 \times 2$ voxel wide kernels with stride 2, such as from $16 \times 512 \times 512$ to $8 \times 256 \times 256$. To speed up the computation, each stage is formulated such that it learns a residual function processed via ReLU and added to the output of the last convolutional layer.
FIGURE 3. Results of the proposed nodule detection model. Patches extracted from six CT scan images. (a) Top panel: nodule marked by radiologists. (b) Middle panel: nodule marked after nodule candidate detection. (c) Bottom panel: nodule marked after false positive reduction.

FIGURE 4. Modified V-Net model for nodule candidate detection.

The lower part of V-Net extracts features and expands the spatial support of lower resolution feature maps so as to gather and assemble the necessary information and output the volumetric segmentation. The low and high resolution feature maps are converted to probabilistic segmentations of the foreground and the background regions by applying the soft-max function. A de-convolution operation is employed to increase the size of the inputs. Features extracted from early stages in the compression path are forwarded to the lower part to improve the quality of the contour prediction.

Fig. 4 shows that the proposed V-Net model employs three receptive fields of size $96 \times 96 \times 16$, $48 \times 48 \times 8$, and $24 \times 24 \times 4$. The $96 \times 96 \times 16$ receptive field provides rich contextual information for large and middle-sized lesions, with the risk of bringing in noisy surrounding signals to some small-sized cases. The $48 \times 48 \times 8$ field provides rich context for small nodules and appropriate amount of contextual information for middle-sized lesions. The $24 \times 24 \times 4$ receptive field encompasses small-sized nodules with proper amount of context.

In the fusion stage, contour retrieval and an approximation method [27] are combined to extract and refine the predicted candidate information, such as coordinates and size of the bounding box. For candidates with centers are too close to each other, a distance ratio of 1.1 is used to distinguish whether two candidates are actually one finding or two individual findings. The distance ratio is defined as the distance between the centers of two detected candidates divided by the predicted side of the bounding box from larger candidates. The results from each V-Net are merged to obtain the final prediction.

Fig. 5 shows the improvement of task accommodation. The left column shows the original operation and the right column the modified accommodation. CONV 3D/2 denotes the convolution of $2 \times 2 \times 2$ voxel kernels applied with stride 2. It serves similar goal as the pooling layer which halves the resolution of input data. In the convolutional block, the batch normalization (BN) is used prior to ReLU for the internal covariate shift. The modified V-Net adopts two CONV 3Ds of size $3 \times 3 \times 3$ instead of one CONV 3D of size $5 \times 5 \times 5$. Small kernel speeds up the computation and increases the depth of convolutional block for performance improvement. Although the convolution of large kernel can catch more global features than that of small kernel, two small kernels could achieve the same receptive field function as one large kernel. In the residual connections, modifications improve
Once the candidate images are identified, it is needed to each iteration to alleviate the excessive storage requirements. This augmentation is performed before through a $2 \times 2$ grid of control-points and B-spline interpolation [29]. The inputs to the network are randomly deformed versions of the training images by a dense deformation field obtained on the other hand, employs a loss function based on dice coefficients to optimize the training process and a residual modification. (a) Convolutional block. (b) Residual connections.

FIGURE 5. Comparison between the original and the modified V-Net modules. The left column is the original and the right column is the modification. (a) Convolutional block. (b) Residual connections.

the gradient flow across the network with full pre-activation connection.

Since the nodule occupies only a very small region of the CT image, the learning process is often caught trapped in the local minima of the conventional logistic loss function. This makes the network prediction strongly biased towards the background. Consequently, the foreground (nodule) region is either missing or partially detected.

To tackle this problem, most CNNs utilize the logistic loss function with sample re-weighting. The foreground regions are assigned more weights than the background during learning [29]. Since sample re-weighting is carried out manually, it is difficult to find the optimal loss function. The V-Net, on the other hand, employs a loss function based on dice coefficients to optimize the training process and a residual learning scheme is incorporated into the model to improve the performance [29]. This loss function does not need to assign different weights according to samples of the foreground voxel (nodule) and the background voxel. Thus, quantity imbalance between the foreground and background voxels would not affect the convergence performance.

The modified V-Net is trained by the stochastic gradient descent algorithm [8]. The training step follows the block diagram of Fig. 4. All the content volumes processed by the RSFS algorithm repetitively chooses a random subset of features from the feature vector. The RSFS selected features are fed into the SVM classifier for nodule validation.

Support vector machine (SVM) has been used for pulmonary nodule discrimination [14], [17]–[20]. SVM classifier builds a hyperplane to maximize the margin between positive and negative samples. The hyperplane is determined by support vectors closest to the decision surface, which is determined by the inner product of feature vectors obtained from the high-level descriptor. This maps the input vectors to a higher dimensional feature space. SVM classifier finds the hyperplane in the higher dimensional space in the training process to separate the nodules from non-nodules.

The hard mining scheme selects all FPs and randomly selected true positives (TPs) to train the classifier and merges FP samples of SVM with TP samples from the training set. Then check if FP samples are either the same as that of the previous round, or the number of iterations reaches its maximum. If yes, the training is complete; otherwise, repeat the training.

Feature selection are often used to reduce the feature dimensionality to improve performance. RSFS enables the selection of the most critical features from the feature vector. The RSFS algorithm repetitively chooses a random subset of features from the set of all possible features. Relevance is ranked with score and updated every iteration according to the classification performance of the subset.

A common practice is to perform the false positive reduction. The idea is to capture the features of CT images that may be neglected or missed by the V-Net in the detection stage.

Fig. 6 depicts the proposed high-level descriptor based SVM classifier. Its purpose is to validate the CT images detected from the nodule detection stage. The output of the classifier indicates whether or not the detected CT image indeed contains any nodule. The input to the classifier is a CT image of size $16 \times 512 \times 512$ from the nodule candidate detection stage. It is categorized as CT image containing the nodule. The high-level descriptor forms the feature vector for feature extraction. The vector is comprised of the image features of shape, texture, and wavelet information. The random subset feature selection (RSFS) algorithm is applied to extract as much relevant features as possible from the feature vector [31]. The RSFS selected features are fed into the SVM classifier for nodule validation.

B. HIGH-LEVEL DESCRIPTOR BASED SVM CLASSIFIER
Once the candidate images are identified, it is needed to ensure that they contain the actual lung cancer nodules.

FIGURE 6. Proposed high-level descriptor based SVM classifier.

The RSFS algorithm repetitively chooses a random subset of features from the set of all possible features. Relevance is ranked with score and updated every iteration according to the classification performance of the subset.

Fig. 7 shows the scores of features in CT images from the Luna16 dataset by RSFS. Features with scores higher than 13 are regarded as the critical features and marked in red color; others are the non-critical features. For instance, the HOG (histogram of oriented gradient) method [32] describes the shape feature regarding 3D geometric properties such as gradient orientations of the nodule. The textural
features quantify the intra-nodule heterogeneity and specify the spatial relationships in a CT image. The grey-level co-occurrence matrix (GLCM) is used for texture analysis [33]. The Hu moments are a set of seven numbers calculated by central moments that are invariant to image transformations [34]. Although Fig. 7 shows seven critical features, five of them (HOG, Hu2, Hu3, Hu4, and Hu7) belongs to the shape feature. Thus, the critical features shown in Fig. 7 are the shape, texture (GLCM correlation), and wavelet features.

The high-level descriptor selects all the critical features for SVM classification. In contrast, current methods mainly consider some critical features and some non-critical features in the CNN or SVM classifiers. For example, features HOG and LBP were considered in [17]. The shape, texture, and intensity features were considered in [18], [20] and the GLCM in [19]. None of the existing classifiers consider the wavelet feature, which is effective in subtle feature detection of the nodules.

IV. EXPERIMENTAL RESULTS
A. TEST DATASET
CT scan images used for test are from Luna16 dataset [35]. This dataset is a subset of the publicly accessible LIDC/IDRI dataset [36]. It contains a total of 888 thoracic CT scan images. In each case, the lesions have been consensually marked by at least three of the four radiologists. Moreover, only the nodules with diameters larger than 3 mm are considered as positive samples, while the others are referred as the negative samples. Each sample is comprised of 16 CT images (slices) of size 512 × 512. These 1471 samples were partitioned into two groups, one as the training set while another as the test set. The ratio of the training set to the test set is 8:2. This division provides the best performance after extensive trials of different ratios.

Performance evaluation metrics used in this study include the Recall, Precision, and $F_1$ score. These metrics are widely used for performance analysis of image classification. They can be calculated by

$$\text{Recall} = \frac{TP}{TP + FN}$$  \hspace{1cm} (1)

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$  \hspace{1cm} (3)

TP, FP, and FN denote the true positive, false positive, and false negative, respectively. TP means that the model correctly detects the nodules that are marked by at least three of the four radiologists. FP denotes the opposite; the nodules detected in the nodule detection stage are not marked by three or more of the four radiologists. FN denotes the number of CT images having nodules marked by at least three of the four radiologists, but were not detected.

Precision and Recall measure the performance with respect to false positives and false negatives, respectively. Recall is an indicator that tells how many of the actual positives are correctly predicted. Precision is an indicator that shows how many of predicted positives are actually positive. So, if we want to minimize FN, we would want the Recall as large as possible. If we want to minimize FP, the focus should be to make the Precision as large as possible. The $F_1$ score is the harmonic average of the precision and recall. It reaches the maximum value at 1 (perfect precision and recall) and the minimum at 0.

B. NODULE CANDIDATE DETECTION
The goal of nodule candidate detection is to identify as many as possible the CT images that contain nodules. Table 1 shows the performance of the proposed nodule candidate detection model. The metrics were evaluated considering all the slices of a CT scan. In the training process, the proposed V-Net model achieves 89.48% of Precision, 96.72% of Recall, and 92.37% of the $F_1$ score. When inputs are from the test set, the model obtains the Precision, Recall, and $F_1$ score of 70.81%, 83.50%, and 74.72%, respectively. It is observed from Table 1 that Recall is always larger than Precision, because the model intends to identify as many nodule candidates as possible, which would come at the cost of many FP cases. It is inevitable because any actual nodules not detected at this stage would no longer be retrieved later. Thus, the false positive reduction process is required.

Classification accuracy was also evaluated. It is the ratio of the number of correct predictions over the total number of samples and is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$  \hspace{1cm} (4)

Accuracy is a good measure when the classes in the data are nearly balanced. For each 16-slice sample in the training set, 9.2 slices contain the nodule on average. In the test set, 9.1 slices contain the nodule.

|                | Precision (%) | Recall (%) | $F_1$ score (%) |
|----------------|--------------|------------|-----------------|
| Training Set   | 89.5         | 96.7       | 92.4            |
| Test Set       | 70.8         | 83.5       | 74.7            |

TABLE 1. Performance of nodule candidate detection by the proposed detection model.
For nodules larger than 3 mm, 928 lesions are classified by all four radiologists, while 2,669 lesions are classified by at least one radiologist [35]. This means that, for nodules \( \geq 3 \) mm, the accuracy evaluated by one radiologist is about 34.8%. Fig. 8 presents the accuracy performance of various methods. It can be seen from Fig. 8 that the proposed system is able to produce an average accuracy of 66.7%, which is better than other CNN models, such as the CNN via deep reinforcement learning, which shows 55.3% accuracy using the same LIDC database [37]. The proposed system achieves superior accuracy performance because it resolves the imbalance issue regarding the numbers of background and foreground pixels in CT images [29].

The precision performance can be improved because V-Net may falsely discriminate some non-nodules as positive cases. It can be seen from Fig. 3(b) and Fig. 3(c) that the nodules in the last two slices are misclassified as positives. Thus, FP reduction is required for performance improvement.

### C. FALSE POSITIVE REDUCTION

FP reduction aims to reduce the number of FP cases obtained from nodule candidate detection. The high-level descriptor based SVM classifier performs this task. Performance evaluation was conducted by the following means. The test set used in the nodule detection stage was partitioned into two new groups (again, a ratio of 8:2). The 80% group is added to the training set as the new training set to train the proposed SVM classifier by the hard mining scheme. Meanwhile, new group 2 was used as the new test set.

Performances of the proposed high-level descriptor based SVM classifier on the reduction of FP cases are shown in Table 2. The results of the nodule detection in Table 2 are obtained by the proposed V-Net model. It can be seen from Table 2 that, after applying the proposed SVM classifier, both the precision and F1 score improve significantly, while the recall is slightly lowered. This verifies the effectiveness of the proposed high-level based SVM classifier. Slightly lower recall may be related to the fact that large reduction in FP cases also causes some true positive cases being removed.

Performance comparison of the proposed SVM classifier with other SVM classifiers and two deep CNN classifiers [6], [38] was conducted. The same datasets were employed for the classifiers. Results are shown in Table 3. It can be seen from Table 3 that the proposed classifier exhibits the best FP reduction in terms of the precision performance. The reason may be due to the fact that the proposed SVM classifier considers more critical features (especially the wavelet feature) than other SVM classifiers. It is also observed from F1 score in Table 4 that the use of 3D deep CNN or deep residual network only improves the FP reduction slightly. This may be due to the fact that CNNs used in the FP reduction stage extract the same CT scan features as that in the candidate detection.

Receiver operating characteristic (ROC) plot is an excellent tool for performance evaluation in the classifier system. This study adopts the FROC curve, which is an alternative to the ROC curve [39]. On the x-axis stands the average number of false positives (FP) per scan. FROC is particularly useful for imbalance detection problem. In actual data set, the negative samples are much more than the positive samples or vice versa.

Fig. 9 shows the FROC curves of five state-of-the-art CNN models and the proposed model on the LIDC/IDRI data. Detailed FROC values are given in Table 4. The models are Faster R-CNN [5], ResNet [6], Mask R-CNN [26], DeepSeed [28], and 3D CNN [38]. Performance was evaluated at different false positive rates. It can be seen from Fig. 9 that

### TABLE 2. Performance comparison for false positive reduction by various SVM and CNN classifier.

| Model              | Precision (%) | Recall (%) | F1 score (%) |
|--------------------|---------------|------------|--------------|
| ResNet [6]         | 3.9           | -3.7       | 0.3          |
| 3D CNN [38]        | +4.8          | -4.5       | +0.4         |
| [14]               | +3.5          | -0.8       | -1.7         |
| [17]               | +3.8          | -0.7       | -1.9         |
| [18]               | +3.1          | -1.4       | +1.1         |
| [19]               | +3.8          | -0.9       | -1.6         |
| Proposed           | +6.4          | -0.8       | +3.4         |

Table 2 that, after applying the proposed SVM classifier, both the precision and F1 score improve significantly, while the recall is slightly lowered. This verifies the effectiveness of the proposed high-level based SVM classifier. Slightly lower recall may be related to the fact that large reduction in FP cases also causes some true positive cases being removed.

Performance comparison of the proposed SVM classifier with other SVM classifiers and two deep CNN classifiers [6], [38] was conducted. The same datasets were employed for the classifiers. Results are shown in Table 3. It can be seen from Table 3 that the proposed classifier exhibits the best FP reduction in terms of the precision performance. The reason may be due to the fact that the proposed SVM classifier considers more critical features (especially the wavelet feature) than other SVM classifiers. It is also observed from F1 score in Table 4 that the use of 3D deep CNN or deep residual network only improves the FP reduction slightly. This may be due to the fact that CNNs used in the FP reduction stage extract the same CT scan features as that in the candidate detection.

Receiver operating characteristic (ROC) plot is an excellent tool for performance evaluation in the classifier system. This study adopts the FROC curve, which is an alternative to the ROC curve [39]. On the x-axis stands the average number of false positives (FP) per scan. FROC is particularly useful for imbalance detection problem. In actual data set, the negative samples are much more than the positive samples or vice versa.

Fig. 9 shows the FROC curves of five state-of-the-art CNN models and the proposed model on the LIDC/IDRI data. Detailed FROC values are given in Table 4. The models are Faster R-CNN [5], ResNet [6], Mask R-CNN [26], DeepSeed [28], and 3D CNN [38]. Performance was evaluated at different false positive rates. It can be seen from Fig. 9 that

### TABLE 3. Performance comparison for false positive reduction by various SVM and CNN classifier.

| Model              | Precision (%) | Recall (%) | F1 score (%) |
|--------------------|---------------|------------|--------------|
| ResNet [6]         | 0.734         | 0.744      | 0.763        | 0.796 | 0.824 | 0.832 | 0.834 |
| 3D CNN [38]        | 0.636         | 0.727      | 0.792        | 0.844 | 0.876 | 0.905 | 0.906 |
| Mask R-CNN [26]    | 0.472         | 0.706      | 0.849        | 0.884 | 0.887 | 0.887 | 0.887 |
| DeepSeed [28]      | 0.739         | 0.803      | 0.858        | 0.888 | 0.907 | 0.916 | 0.920 |
| 3D CNN [38]        | 0.640         | 0.698      | 0.750        | 0.804 | 0.847 | 0.874 | 0.897 |
| Proposed           | 0.621         | 0.733      | 0.849        | 0.899 | 0.918 | 0.924 | 0.934 |

Table 4. FROC by different models on LIDC/IDRI dataset.

| CNN Models | # of False Positives per Scan |
|------------|-------------------------------|
| Faster R-CNN [5] | 0.125 | 0.25 | 0.5 | 1.0 | 2.0 | 4.0 | 8.0 |
| ResNet [6] | 0.734 | 0.744 | 0.763 | 0.796 | 0.824 | 0.832 | 0.834 |
| Mask R-CNN [26] | 0.472 | 0.706 | 0.849 | 0.884 | 0.887 | 0.887 | 0.887 |
| DeepSeed [28] | 0.739 | 0.803 | 0.858 | 0.888 | 0.907 | 0.916 | 0.920 |
| 3D CNN [38] | 0.640 | 0.698 | 0.750 | 0.804 | 0.847 | 0.874 | 0.897 |
| Proposed | 0.621 | 0.733 | 0.849 | 0.899 | 0.918 | 0.924 | 0.934 |
the proposed model exhibits the best sensitivity at the test rates higher than 1.0 false positive per scan. Table 4 shows that the proposed model provides the best nodule detection sensitivity of 0.899, 0.918, 0.924 and 0.934 at 1, 2, 4, and 8 FPs/scan, respectively, over other CNN models. This implies that the incorporation of features extracted by the proposed high-level descriptor based SVM classifier improves the nodule detection performance.

It is observed from Table 4 that the proposed model excels at medium and high FPs/scan (rate higher than 1.0 FPs/scan) and DeepSeed model [28] provides better FROC values at low FPs/scan. The proposed hard mining retraining scheme puts more weights on the erroneous samples (FP cases), which effectively improve the FP reduction performance. Thus, the proposed model may be of help to radiologists and patients by providing accurate diagnosis results when multiple FP samples exist in the CT scans [26].

D. DISCUSSION

Results on FROC from Fig. 9 show that the proposed model performs well at medium and high FPs/scan (rate higher than 1.0 FPs/scan), but not at low FPs/scan. Although the proposed SVM classifier performs better than other SVM based classifier, the improvement is not significant. Feature selection may need to be revisited to find the optimal set of features to improve the FP reduction performance of SVM classifier. Another consideration is to use the collaborative learning to train the multiple classifiers simultaneously on the same training data to examine the generalization capability of the proposed model.

V. CONCLUSION

This paper has presented an automated pulmonary nodule detection system, which performs nodule candidate detection using a modified 3D V-Net model and the false positive reduction with a high-level descriptor based SVM classifier. The modified V-Net with multilevel receptive fields enables more nodule candidate CT images to be detected than 2D CNNs. The proposed SVM classifier trained by the hard mining algorithm enables improved feature selection of CT image for the increase of detection accuracy. Test results on LIDC/IDRI dataset demonstrate that the proposed framework provides more reliable information to the doctor in lung nodule detection from CT scan than several existing deep learning CAD models.

Future research includes the performance evaluation of other classifiers (e.g., KNN and random forest classifiers) for false positives reduction, after the detection of nodules carried out by the modified V-Net. The proposed model will also be used in other medical applications, such as the breast cancer prediction.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2015,” CA, Cancer J. Clinician, vol. 65, no. 1, pp. 5–29, Jan. 2015.
[2] Y. Xie, J. Zhang, Y. Xie, M. Fullham, and Y. Zhang, “Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT,” Inf. Fusion, vol. 42, pp. 102–110, Jul. 2018.
[3] G. Zhang, S. Jiang, Z. Yang, L. Gong, X. Ma, Z. Zhou, C. Bao, and Q. Liu, “Automatic nodule detection for lung cancer in CT images: A review,” Comput. Biol. Med., vol. 103, pp. 287–300, Dec. 2018.
[4] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, “Texture feature analysis for computer-aided diagnosis on pulmonary nodules,” J. Digit. Imag., vol. 28, no. 1, pp. 99–115, Feb. 2015.
[5] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, “Automated pulmonary nodule detection in CT images using deep convolutional neural networks,” Pattern Recognit., vol. 85, pp. 109–119, Jan. 2019.
[6] A. Nibali, Z. He, and D. Wollersheim, “Pulmonary nodule classification with deep residual networks,” Int. J. Comput. Assist. Radiol. Surg., vol. 12, no. 10, pp. 1799–1808, 2017.
[7] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sanchez, and B. van Ginneken, “Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks,” IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1160–1169, May 2016.
[8] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, “Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection,” IEEE Trans. Biomed. Eng., vol. 64, no. 7, pp. 1558–1567, Jul. 2017.
[9] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 91–99.
[10] J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object detection via region-based fully convolutional networks,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 379–387.
[11] A. Pezeshk, S. Hamidian, N. Petrick, and B. Sahiner, “3-D convolutional neural networks for automatic detection of pulmonary nodules in chest CT,” IEEE J. Biomed. Health Informat., vol. 23, no. 5, pp. 2080–2090, Sep. 2019.
[12] H. Cao, H. Liu, E. Song, G. Ma, X. Xu, R. Jin, T. Liu, and C.-C. Hung, “Multi-branch ensemble learning architecture based on 3D CNN for false positive reduction in lung nodule detection,” IEEE Access, vol. 7, pp. 67380–67391, 2019.
[13] H. Cao, H. Liu, E. Song, G. Ma, X. Xu, R. Jin, T. Liu, and C.-C. Hung, “A two-stage convolutional neural networks for lung nodule detection,” IEEE J. Biomed. Health Informat., vol. 24, no. 7, pp. 2006–2015, Jul. 2020.
[14] S. A. Khan, S. Hussain, S. Yang, and K. Iqbal, “Effective and reliable framework for lung nodules detection from CT scan images,” Sci. Rep., vol. 9, no. 1, Dec. 2019, Art. no. 4989.
[15] P. Monjum, S. Qi, H. Ma, W. Gao, Y. Yao, and W. Qian, “Detection and classification of pulmonary nodules using convolutional neural networks: A survey,” IEEE Access, vol. 7, pp. 78075–78091, 2019.
[16] T. Liu, F. Hou, H. Qin, and A. Hao, “Multi-view multi-scale CNNs for lung nodule type classification from CT images,” Pattern Recognit., vol. 77, pp. 262–275, May 2018.
[17] S. P. Kailasam and M. M. Sathik, “A novel hybrid feature extraction model for classification on pulmonary nodules,” Asian Pacific J. Cancer Prevention, vol. 20, no. 2, pp. 457–468, Feb. 2019.
[18] Y. Gu, X. Lu, B. Zhang, Y. Zhao, D. Yu, L. Gao, G. Cui, L. Wu, and T. Zhou, “Automatic lung nodule detection using multi-scale dot nodule-enhancement filter and weighted support vector machines in chest computed tomography,” PLoS ONE, vol. 14, no. 1, Jan. 2019, Art. no. e0210551.

[19] S. G. and J. P., “A fully-automated system for identification and classification of subsolid nodules in lung computed tomographic scans,” Biomed. Signal Process. Control, vol. 53, Aug. 2019, Art. no. 101586.

[20] L. Boroczky, L. Zhao, and K. P. Lee, “Feature subset selection for improving the performance of false positive reduction in lung nodule CAD,” IEEE Trans. Inf. Technol. Biomed., vol. 10, no. 3, pp. 504–511, Jul. 2006.

[21] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, “An automatic detection system of lung nodule based on multigroup patch-based deep learning network,” IEEE J. Biomed. Health Inform., vol. 22, no. 4, pp. 1227–1237, Jul. 2018.

[22] G. L. F. da Silva, T. L. A. Valente, A. C. Silva, A. C. de Paiva, and M. Gattass, “Convolutional neural network-based PSO for lung nodule false positive reduction on CT images,” Comput. Methods Programs Biomed., vol. 162, pp. 109–118, Aug. 2018.

[23] G. Kasinathan, S. Jayakumar, A. H. Gandomi, M. Ramachandran, S. J. Song, and R. Patan, “Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier,” Expert Syst. Appl., vol. 134, pp. 112–119, Nov. 2019.

[24] Ö. Çiçek, “3D U-Net: Learning dense, volumetric segmentation from sparse annotation,” in Proc. Int. Conf. Med. Image Comput. Assist. Interv., 2016, pp. 424–432.

[25] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, “An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification,” Expert Syst. Appl., vol. 128, pp. 84–95, Aug. 2019.

[26] L. Cai, T. Long, Y. Dai, and Y. Huang, “Mask R-CNN-Based detection and segmentation for pulmonary nodule 3D visualization diagnosis,” IEEE Access, vol. 8, pp. 44400–44409, 2020.

[27] S. Zheng, J. Guo, X. Cui, R. N. J. Veldhuis, M. Oudkerk, and P. M. A. van Ooijen, “Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection,” IEEE Trans. Med. Imag., vol. 39, no. 3, pp. 797–805, Mar. 2020.

[28] Y. Li and Y. Fan, “DeepSEED: 3D Squeeze-and-Excitation encoder-decoder convolutional neural networks for pulmonary nodule detection,” in Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI), Apr. 2020, pp. 1866–1869.

[29] F. Miletati, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in Proc. 4th Int. Conf. 3D Vis. (3DV), Oct. 2016, pp. 565–571.

[30] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Assist. Intervent., Munich, Germany, 2015, pp. 231–234.

[31] J. Pohjalainen, O. Räsänen, and S. Kadioglu, “Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits,” Comput. Speech Lang., vol. 29, no. 1, pp. 145–171, Jan. 2015.

[32] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2005, pp. 886–893.

[33] P. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” Pattern Recognit., vol. 29, no. 1, pp. 51–59, Jan. 1996.

[34] M.-K. Hu, “Visual pattern recognition by moment invariants,” IEEE Trans. Inf. Theory, vol. IT-8, no. 2, pp. 179–187, Feb. 1962.

[35] A. A. Assitio, “Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge,” Med. Image Anal., vol. 42, pp. 1–13, Dec. 2017.

[36] S. G. Armato, “The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans,” Med. Phys., vol. 38, no. 2, pp. 915–931, Jan. 2011.

[37] I. Ali, G. R. Hart, G. Gunabushanam, Y. Liang, W. Muhammad, B. Nartowt, M. Kane, X. Ma, and J. Deng, “Lung nodule detection via deep reinforcement learning,” Frontiers Oncol., vol. 8, p. 108, Apr. 2018.

[38] H. Tang, D. Kim, and X. Xie, “Automated pulmonary nodule detection using 3D deep convolutional neural networks,” in Proc. IEEE 15th Int. Symp. Biomed. Imag., Birmingham, AL, USA, 2018, pp. 523–526.

[39] A. I. Bandos, H. E. Rockette, T. Song, and D. Gur, “Area under the free-response ROC curve (FROC) and a related summary index,” Biometrics, vol. 65, no. 1, pp. 247–256, Mar. 2009.

YUYUN YE received the B.E. degree in electrical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2003, the M.S. degree in electrical engineering from The University of Tulsa, Tulsa, OK, USA, in 2005, and the Ph.D. degree in electrical engineering from the University of Colorado, Boulder, CO, USA, in 2012. From 2013 to 2014, he worked as a Research Associate with the Earth System Science Interdisciplinary Center (ESSIC), University of Maryland at College Park, College Park, MD, USA. From December 2014 to 2016, he was a Senior Engineer with Earth Resources Technology (ERT), Inc. He is currently an Associate Researcher with UESTC. His research interests include deep learning, image processing and pattern recognition, passive remote sensing, instrument calibration, and antenna design.

MIAO TIAN (Member, IEEE) received the B.E. degree in electrical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2003, the M.S. degree in electrical engineering from The University of Tulsa, Tulsa, OK, USA, in 2005, and the Ph.D. degree in electrical engineering from the University of Colorado, Boulder, CO, USA, in 2012. He also serves as the Deputy Chairman of the Interventional Medicine Committee of Sichuan Medical Association.

QIYU LIU received the bachelor’s degree from Southwest Medical University, Luzhou, Sichuan, China, in 1988, and the master’s degree in medicine from Sichuan University, Chengdu, Sichuan, in 2002. He is a Radiologist with the Mianyang Central Hospital, Mianyang, Sichuan, in 2012. Since 2009, he has been the Director of the Radiology Department, since 2007. He has published tens of papers in referred journals or conferences. His research interests include neurovascular intervention, image processing, radiomics, and deep learning.

HENG-MING TAI (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Texas Tech University, Lubbock, TX, USA, in 1987. He is currently a Professor with the Department of Electrical and Computer Engineering, The University of Tulsa, Tulsa, OK, USA. His research interests include signal processing, power system reliability, and industrial electronics.