Received November 11, 2020, accepted November 24, 2020, date of publication December 15, 2020,
date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044941

A Joint Detection and Recognition Approach to Lung Cancer Diagnosis From CT Images With Label Uncertainty

LIU CHENYANG AND SHING-CHOW CHAN, (Member, IEEE)
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong
Corresponding author: Shing-Chow Chan (scchan@eee.hku.hk)

This work was supported by the Hong Kong Research Grant Council General Research Fund under Project GRF 17211320.

ABSTRACT Automatic lung cancer diagnosis from computer tomography (CT) images requires the detection of nodule location as well as nodule malignancy prediction. This article proposes a joint lung nodule detection and classification network for simultaneous lung nodule detection, segmentation and classification subject to possible label uncertainty in the training set. It operates in an end-to-end manner and provides detection and classification of nodules simultaneously together with a segmentation of the detected nodules. Both the nodule detection and classification subnetworks of the proposed joint network adopt a 3-D encoder-decoder architecture for better exploration of the 3-D data. Moreover, the classification subnetwork utilizes the features extracted from the detection subnetwork and multiscale nodule-specific features for boosting the classification performance. The former serves as valuable prior information for optimizing the more complicated 3D classification network directly to better distinguish suspicious nodules from other tissues compared with direct backpropagation from the decoder. Experimental results show that this co-training yields better performance on both tasks. The framework is validated on the LUNA16 and LIDC-IDRI datasets and a pseudo-label approach is proposed for addressing the label uncertainty problem due to inconsistent annotations/labels. Experimental results show that the proposed nodule detector outperforms the state-of-the-art algorithms and yields comparable performance as state-of-the-art nodule classification algorithms when classification alone is considered. Since our joint detection/recognition approach can directly detect nodules and classify its malignancy instead of performing the tasks separately, our approach is more practical for automatic cancer and nodules detection.

INDEX TERMS Deep learning, multi-task learning, nodule detection, nodule malignancy classification, label noise.

I. INTRODUCTION

Lung cancer is the primary cause of cancer deaths worldwide. The 2018 Global Cancer Statistics [1] shows that there are approximately 1.8 million deaths and 2.1 million new cancer cases caused by lung cancer, ranking first among other cancers. Early diagnosis of a small tumor can prevent metastasis of cancer and substantially improves the prognosis and survival rate [2]. Therefore, the development of an intelligent computer-aided diagnosis system (CADS) can be beneficial to the early treatment of lung cancer.

The volumetric thoracic computed tomography (CT) is the most commonly used imaging technique for lung scan [3], which can be used to detect lesions in the lung called pulmonary nodules. Such nodules can be benign or malignant, and the detection of the latter is of great importance. One difficulty in detecting the nodules from these CT scans is that the nodules absorb the same level of X-ray as normal body tissues. Thus, there is no apparent intensity discrepancy. The distinctive features of pulmonary nodules are primarily related to shape and location. Figure 1 shows an example 2D slice from such as volumetric or 3D- CT scan. It can be seen from Figure 1 (c) that the tiny pulmonary nodule has no distinctive feature compared with vessels in the 2-D image. However, the vessels have a continuous structure, while nodules are isolated. This motivates us to develop a network for detecting nodule and malignancy using 3-D volumetric data instead of fusing results from multiple 2D slices.

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

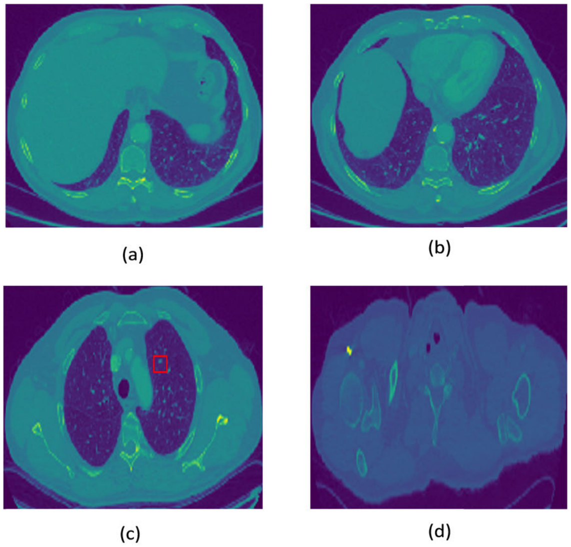


FIGURE 1. Four 2D slices from a CT scan of lung from the bottom to top: (a) is scanned on the bottom of lung and (d) is on the top. A red bounding box has been included in (c) to highlight the nodule. To better reveal the nodules and other tissue structure, the grey-scale image has been color mapped to improve visualization.

On the other hand, humans are more proficient in extracting information from 2-D images than 3-D volumetric images. Therefore, it is expected that a thorough analysis of CT scans by clinicians can take much time, increasing the cost of such check. Compared with checking by doctors, CADS has the potential advantage of taking the three-dimension image data into account and output potential nodule candidates for reference or confirmation quickly. More importantly, the CADS approach can even learn and accumulate the experience from radiologists via continuous training. Hence, they may provide very stable prediction comparable or even outperforming a single experienced radiologist [4]. Hence, it is helpful to develop an efficient CADS for the diagnosis of lung cancer from CT images.

In the literature, such automatic diagnosis usually consists of two steps: nodule detection and nodule classification [5]. With the success of deep learning in natural image processing, most recent studies on these two tasks are based on the convolution neural network (CNN) [6]–[9]. Methods for nodule detection usually rely on networks for object detection problems, including faster R-CNN [10] and YOLO [11], which outputs region proposals of the target objects. The nodule classification problem, on the other hand, is usually regarded as a 3-D image[1] recognition problem using the data at the detected regions as inputs. 3-D extensions of well-known image classification networks such as ResNet [12] are widely used.

Despite these advances, a fully automatic CADS for lung nodules detection and cancer classification still present several major challenges. First of all, separating the detection with classification tasks usually reduce the overall classification rate as considerable amounts of detected nodules are, in fact, false positives. By introducing a simple classification stage to refine the detected nodules after the detection task can considerably reduce the false-positive results [13], which, otherwise, will mislead the classification task later. Therefore, it is desirable to develop a methodology for joint nodule detection and malignancy classification.

Secondly, most pulmonary nodules are small and isolated in the raw CT scans. The shape of the nodule thus serves as an informative feature for distinguishing it from other body tissues. Therefore, it is desirable to exploit the 3D nature of the data for better classification. However, due to significantly increased parameters of 3D neural networks, most conventional approaches are still based on multiple 2D networks [14]–[16]. The primary obstacle of applying the 3-D model in nodule classification is the overfitting problem arising from the increased number of parameters and the limited number of training samples. For instance, while ImageNet [17] uses millions of images for training, there are only 1018 scans in the LIDC-IDRI [18]–[20] lung cancer CT dataset.

Finally, for some cases, the labels of the radiologists may not be consistent or missing (say the nodules may be labelled by 1 or 2, but not all the radiologists). This arises because labeling nodules as benign or malignant using CT images depends mostly on the experience of radiologists and the limitations in the data collection process. Unless a single consistent label can be agreed on (as in some dataset), such uncertain labels, which we shall also refer to as marginal labels, will arise for some nodules. In fact, it is commonly found in the LIDC-IDRI dataset. If the network is forced to fit these marginal samples, the performance usually deteriorates as reported in [15], [16]. This problem is usually referred to as the label uncertainty problem. Though a precise probabilistic model to describe such variations can be difficult to obtain, it is desirable that such adverse effect on the overall performance of the network can be mitigated. All these motivate us to develop a joint detection and recognition approach to lung cancer diagnosis and segmentation from CT images with possibly marginal or uncertain labels.

An important advantage of the proposed joint detection/recognition approach is that it can directly detect nodules and classify its malignancy instead of performing the two tasks separately. Therefore, our approach is more practical as it can be applied in an end-to-end manner to automatic cancer and nodules detection. Moreover, the proposed joint nodule segmentation/recognition (JNSC) network is capable of exploring the semantic segmentation information [21] to yield a more detailed segmentation of the nodules and their malignancy instead of conventional simple regional proposal. It is known that nodule malignancy is highly related to its morphology. The segmentation information offered by our proposed joint nodule segmentation/recognition (JNSC) network can provide valuable morphology description of

[1] The phrase 3D images, 3D volumetric images, and 3D CT scans will be used interchangeably in the paper with the same meaning, while a slice of such volume is referred to as a 2D slice image or 2D slice.

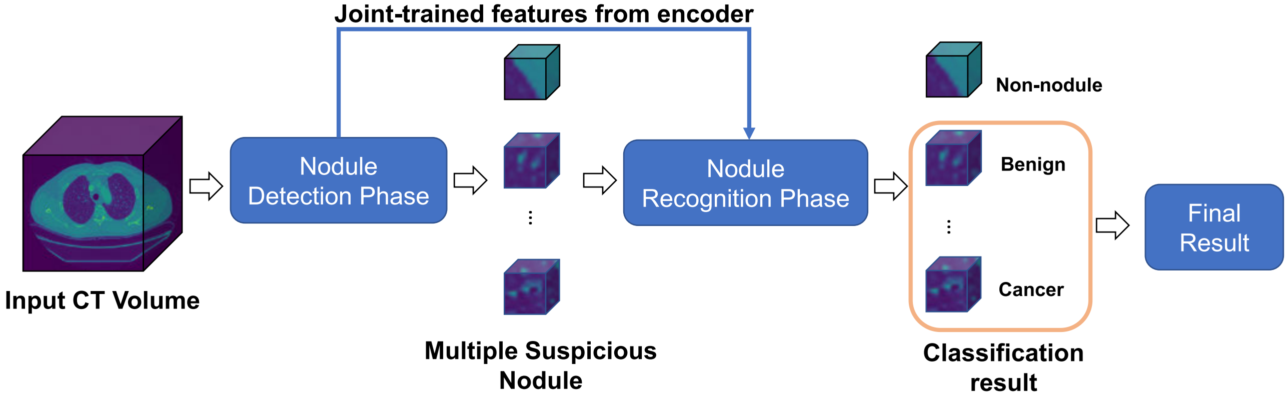


FIGURE 2. System overview of the proposed framework. The detection phase outputs multiple potential nodules. The recognition phase uses features of detection phase to build an additional classifier to discriminate them into three classes: benign, cancer and non-nodule. Only the benign and cancer nodules are then evaluated for nodule detection task and classification task. Importantly, in the classification task, the undetected nodules are directly labeled as benign to report the result. The architecture of the joint nodule detection and classification network is shown in Figure 3.

the detected nodules, which can be useful in differentiating malignant tumors from scars or other complications.

From the neural network training point of view, the encoded features and initial segmentation obtained in our nodule detection network serve as valuable prior information for the subsequent classification process. This not only helps to the classification network to extract more discriminative features but also makes possible the training of our 3D neural network for classification and further refinement of the segmentation map without suffering from excessive overfitting.

Figure 2 shows the system overview of the proposed network, where the input CT image is passed through the proposed joint nodule detection and recognition network to provide a segmentation map of the nodule as well as its malignancy prediction. Our JNSC network is a 3D network and it adopts the encoder-decoder architecture with multiscale features extraction, which has the advantages to encode the desired location information as well as shape information of the nodules. Moreover, instead of simply cascading the detection and classification networks, a path for extracting discriminative features from the output of the encoder of the nodule detection module to the classification network is proposed. These features are jointly trained from the two networks and provide valuable additional information for improving the classification performance.

Thanks to this additional information provided by the nodule detection network, the proposed 3D JNSC can be trained from scratch despite the limited number of training samples. Moreover, the encoder in our JNSC is trained on the whole CT image, which can also distinguish other body tissues for nodule detection. Experiment results to be presented later show that the joint detection and classification framework is superior to the sole classification approach with an improvement of 1.25% in terms of accuracy. This is in accordance with previous studies in scene geometry and semantics research [22], [23] where it has been demonstrated that multi-task learning can effectively boost the overall performance.

Finally, to address the label uncertainty problem, we treat the problem as a training problem with label noise[2] [24] where the noisy label will be corrected during the training phase. In the lung nodule diagnosis problem, samples with inconsistent or missing annotations are commonly encountered and they may be less reliably annotated. Here, we introduce the concept of pseudo-label to alleviate the adverse effect of these possible less reliable annotations. More precisely, the unreliable annotations are detected and their labels are re-estimated as "pseudo-labels" by minimizing a variant of the cross-entropy loss function, which is capable of seeking a better tradeoff between network prediction and fitting errors. While the true model of these less reliable labels is different to obtain in practice, the use of the more robust cross-entropy loss function effectively prevents the network from overfitting those less reliable marginal samples.[3] Experimental results show that training with the proposed pseudo-labels can improve the accuracy by 2.44% compared with the hard-label assignment and by 1.31% compared with the soft-label assignment.[4]

The proposed approach has been evaluated and compared with state-of-state algorithms on the publicly available LIDC-IDRI dataset. In particular, the nodule detection phase is validated on the LUNA 16 [13] competition, which is a subset of LIDC-IDRI. The result shows that our proposed nodule detection network outperforms state-of-the-art algorithms while achieving comparable results with state-of-art nodule classification algorithms. Since our joint detection/recognition approach can directly detect nodules and classify its malignancy in an end-to-end manner instead of performing the two tasks separately,[5] our approach is more

[2] Or simply the label noise problem.

[3] This is similar to the use of robust loss functions, instead of the true probability function of the measurement noise, in robust statistics.

[4] Note, we do not change the original labels of the samples in the evaluation of the accuracy. Instead, we adopt a more robust loss function to measure the error between the predicted and the given labels so as to reduce the sensitivity of the network trained. Thus, the overall accuracy can still be improved.

[5] This may reduce the overall classification rate.

TABLE 1. Summary of nodule detection approaches (CPM: Competition Performance Metric. AUC: Area Under Curve).

Approach	Year	Dataset	Result
DeepMed [8]	2019	LUNA16	CPM: 0.832
SDFPR [4]	2018	LUNA16	CPM: 0.831
DeepLung [9]	2018	LUNA16	CPM: 0.842
Attention 3D-CNN [25]	2018	LUNA16	CPM: 0.897
CAD-Multimodalities [26]	2019	JRST/LIDC	Accuracy at 3 FPs : 75.2
CAD-SVM [27]	2018	LUNA16	AUC: 0.811
ETROCAD [28]	2020	LUNA16	CPM:0.74
CAD-ST [29]	2018	LUNA16	AUC: 0.874

practical for automatic cancer and nodules detection. Moreover, the segmentation map of the nodules and its malignancy are available from the network output, which provides valuable information on the morphology of the tumor.[6]

The rest of the paper is organized as follows. Section II briefly reviews the literature of related works. The information of the dataset under study is given in Section III. The proposed network architecture, feature extraction, and joint optimization methods are presented in Section IV. The experimental results, analysis, and comparisons are presented in Section V. Section VI summarizes the major findings/contributions and possible limitations of the work. Finally, conclusions are drawn in Section VII.

II. RELATED WORKS

A. NODULE DETECTION

Nodule detection from CT images usually involves two steps: i) nodule candidate proposal and ii) false-positive reduction [30]. The goal of nodule detection is to identify potential nodule candidates from the remaining lung tissues, whereas the false positive reduction aims to suppress potential false positive due to interference from tissues such as blood vessels, etc. TABLE 1 summarizes some recent works on nodule detection and their performance.[7]

Traditional detection methods usually rely on hand-craft features and classic image segmentation methods [31]. Recently, a more extensive dataset LIDC-IDRI is made publicly available. Hence, more sophisticated deep learning-based methods can be applied and significantly better performance over traditional approaches in the larger dataset has been demonstrated [13], [32].

In Ding *et al.* [33], a 2-D region proposal network, which is transferred from the general image detection framework [10], was proposed and an impressive sensitivity of 94.6% under 15 candidates per scan is achieved. Though a 2-D network generally has fewer parameters than a 3-D network, it cannot fully utilize the 3-D shape information simultaneously. Therefore, more recent studies [9], [34], [35] tend to adopt 3-D CNN to solve the problem directly. For instance, Khosravan and Bagci [35] propose a 3-D densely connected region proposal network to acquire the region proposals. This densely connected network connects every two layers in the network, while the typical network only connects two successive layers. Therefore, it usually improves the overall performance over normal layer-by-layer connected network, while requiring much fewer parameters than many conventional 3-D networks. Besides the region proposal network, Pezeshk *et al.* [8] proposed to segment the nodules from the CT scans directly. Similar pixel-wise segmentation has been widely applied to biomedical-related applications, in which the 3-D U-net [36] and V-Net [37] are prevalent network architectures. While segmentation can provide more accurate information than detection only, it is also more involved as more detailed annotation will be required. Since LIDC-IDRI has released the pixel-wise segmentation label recently, training deep networks for nodule segmentation is now feasible, and it can potentially provide more information to the joint detection (segmentation) and classification of lung nodules.

The false-positive reduction is another essential step after nodule detection to eliminate false positive candidates, and 3-D CNN is usually preferred [4], [8], [32], [33] because of their excellent performance. The network usually undertakes a classical classification task, i.e., classifying nodule with non-nodule. Furthermore, there is no need to develop an independent network as features can be simply transferred from the detection stage for performing classification. In Qin *et al.* [4], the feature from the nodule detection network is directly cropped. As the LUNA 16 competition provides an additional false-positive reduction (FPR) task which labels many possible false-positive nodules, better performance is achieved if a FPR network is trained to refine the detection result. Moreover, it is observed that even if the false positive samples in the detection task are collected without additional labels from FPR task, training their own FPR networks can also improve the result [25], [35].

B. NODULE CLASSIFICATION

Currently, nodule classification is performed either on the patient-level or nodule level. On the patient-level, only the binary label for each patient is available regardless of the number of nodules of the patient. Liao *et al.* [34] proposed an end-to-end CADS and won the competition for patient-level lung cancer classification. The nodule-level evaluation is popular because it has an accurate label for each nodule and avoids the variance raising from the multiple instance problem. Indeed, the framework of both levels is quite similar, except for the training strategy.

Some classical image processing descriptors, including Local Binary Pattern (LBP) [38], Histogram of Oriented Gradients (HOG) [39], and Fourier shape descriptor [40], are firstly exploited in nodule classification. Nevertheless, deep learning-based approaches usually outperform these hand-craft features [15]. Zhao *et al.* [41] propose a hybrid approach using well-known AlexNet and LeNet to classify the nodule slice, the performance is superior to single model

[6] This information can be used in say studying the relationship between the CT image features with genomic features in radiomics.

[7] It should be noted that some of these works may not be directly compared due to differences in modalities, etc. They are listed to give the reader an impressive of the state-of-the-art performance in nodule detection.

methods. Moreover, in order to alleviate the overfitting problem, the 3-D nodules can be decomposed into multi-views [32] therefore the 3-D network is simplified to multiple 2-D networks. Recently, Xie *et al.* [16] adopt, in total, 27 ResNet for classifying the 3-D nodules from 9 viewpoints. Similarly, Hussein *et al.* [42] adopt a slice-by-slice approach by fusing the results from all the slices. Although many studies [9], [15] have focused on 3-D architecture, the performance is usually inferior to these 2-D ensemble methods Liao *et al.* [34] firstly incorporate the nodule classification into the nodule detection network and train the detection and classification network alternatively. Zhang *et al.* [43] fine tune the classification network from the detection network and shows that classification performance can be benefited from information of the detection stage. Moreover, Xie *et al.* [44] show that joint training can boost segmentation and classification in skin lesion. While the choice of 2-D or 3-D networks in nodule classification remains controversial, we shall focus on 3-D network as it is more promising in exploring the morphology information of pulmonary nodules. Notably, we extended the co-training method in [34] for training our 3-D network to be described in Section IV.

C. LABEL NOISE

Estimating the malignancy level of nodules from morphology depends mainly on the experience of the clinicians and there are inevitably variations and perhaps errors for difficult cases. Therefore, labels may not always be consistent, especially when only a few annotations are available. Although up to 4 radiologists will label the data in the LIDC-IDRI database [18]–[20], many samples are only labeled by only one radiologist. Such uncertainty in the labels are usually referred to as label noise. Frenay and Verleysen [45] give a comprehensive review on tackling label noise. Manwani and Sastry [46] studied the noise tolerance performance of various loss function and found that the 0-1 loss has the best noise toleration ability. Zhang *et al.* [47] developed a probabilistic model to deal with potential misclassification where the noise label is used as prior information for updating the posterior probability. These algorithms mainly focus on loss function and label correction. Other improvements proposed include data cleansing [48], [49] and model-based methods [50], [51].

Since training a neural network is time-consuming, it is hard to train a neural network several times until the noise correction converges. Patrini *et al.* [52] recently proposed a two-stage training method which adapts the loss function at the first stage and re-trains the network at the second stage. Adjusting the loss function is preferred on the neural network-based model because it can be easily integrated into the current framework if the loss function is differentiable.

III. DATASET

In this study, the LIDC-IDRI [18]–[20] dataset from The Cancer Imaging Archive (TCIA) is used to evaluate the performance of our proposed network. There are 1018 scans obtained from seven institutions in the dataset, and four experienced thoracic radiologists annotate each scan with detailed nodule location as well as malignancy level. However, the radiologists sometimes cannot reach a consensus for some lesions, and therefore, some nodules are annotated by one to three radiologists.

The diameter of nodules ranges from 3 mm to 30 mm, and the malignancy level is evaluated in a 5-point scale where 1 represents 'Highly unlikely' nodule, 3 represents 'Indeterminate' and 5 represents 'Highly suspicious'. Following the settings in the previous studies [15], [16], [53], we calculate the malignancy score (MS) by taking the median of the malignancy levels from different annotations and label the nodules whose MS$<$3 as benign, MS=3 as uncertain, and MS$>$3 as malignant. Note that uncertain nodules are excluded in the testing phase. Moreover, we observe that a considerable number of nodules are marginally classified as benign or malignant, and some nodules are only annotated by one radiologist, which may introduce label uncertainty. Thus, we further categorize the benign and malignant nodules as certain and marginal nodules. Marginal nodules are defined as the nodules which are labelled by only one or two radiologists, and the median malignancy levels are between 2 and 4, including 2 and 4. We list the precise number of nodules in each class in TABLE 2.

TABLE 2. Number of nodules in LIDC-IDRI dataset.

Benign (MS<3)		Uncertain (MS=3)	Malignant (MS>3)	
Certain	Marginal	-	Certain	Marginal
648	495	947	446	89

For nodule detection, we adopt the Lung Nodule Analysis 2016 (LUNA16) [13] to evaluate the performance of the nodule detection algorithms. The LUNA16 dataset is a subset of the previous LIDC-IDRI dataset. To better evaluate the nodule detection algorithms, the scans with a slice thickness greater than 2.5 mm are excluded from the LIDC-IDRI dataset. LUNA16 only consists of nodules whose diameters are larger than 3 mm and annotated by at least three radiologists. Therefore, there are in total of 888 scans with 1186 nodules in the challenge. Due to the large image size, most works tend to train the detection and classification network on a small size voxel[8] like 64 × 64 × 64, randomly sampled from the entire image. Afterward, to obtain the final detection/classification for a particular subject, one needs to apply the network to the many sub-voxels of the entire image and aggregate the respective outputs. For instance, in the LUNA challenge, the results obtained by applying the detection to 64 × 64 × 64 voxels with a shift of multiples of 32 voxels in any of the three directions are averaged to form the performance metric.

[8] The choice of a size of "64×64×64" is to avoid possible memory limitation in GPUs, which is a commonly used method. Moreover, the performance degradation is found to be minimal.

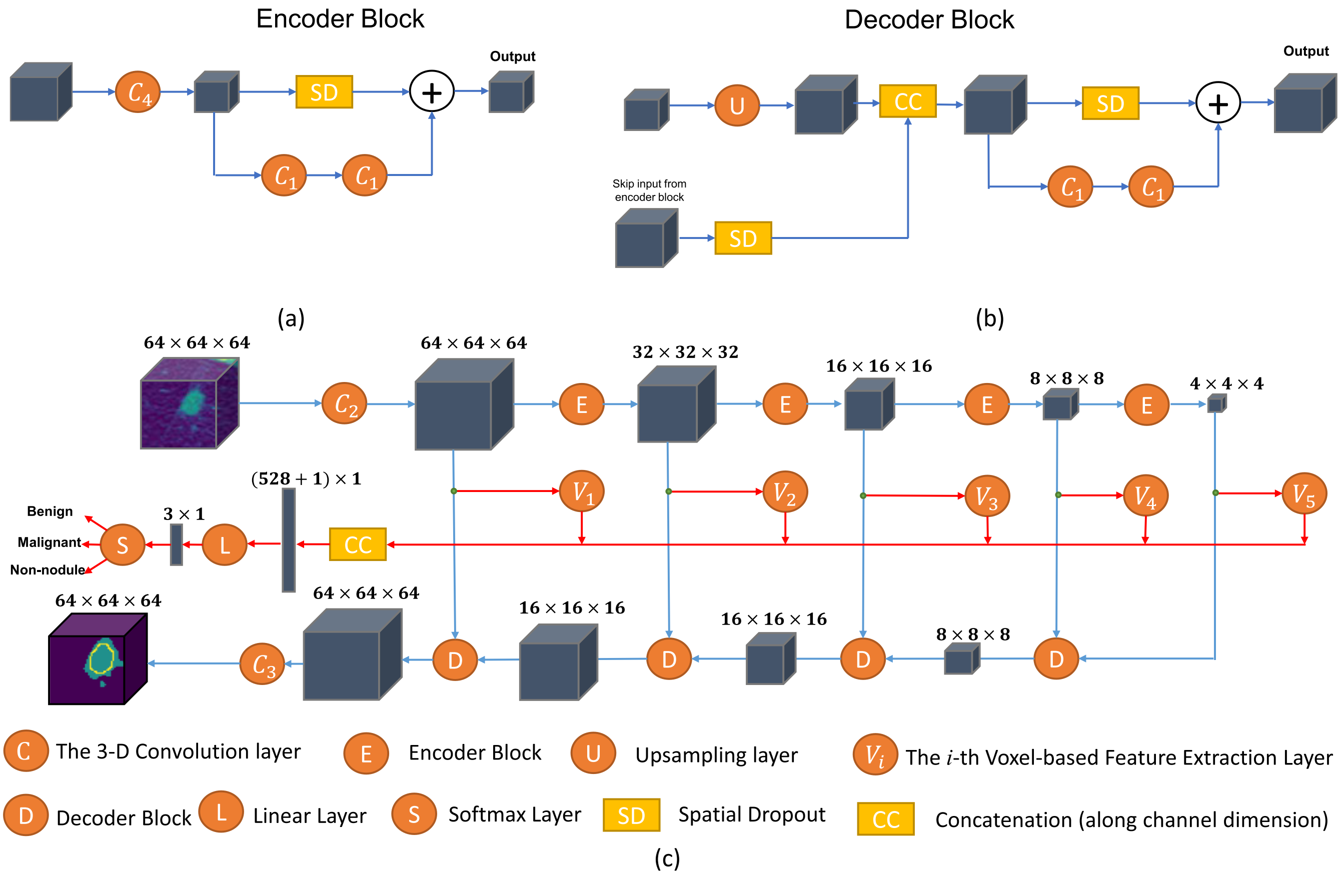


FIGURE 3. The basic structures for (a) encoder block "E" and (b) decoder block "D" used in the proposed joint network. (c) Block diagram of the proposed joint nodule segmentation and classification network. The red line represents the data flow of nodule classification network and the blue line represents the nodule detection network. The cubes in the figure represent 4-D tensor and the number on the top of the cube is the dimension for width, height and length while the channel dimension is not plotted. The voxel-based feature extraction layer V_i is shown in Figure 4. The parameters of the convolution layers are shown in TABLE 3.

In this study, we use the official 10-fold split in LUNA16 to report the detection performance by randomly splitting the scans in LIDC-IDRI to 10-fold for five times to report the nodule classification performance.

IV. PROPOSED METHOD

We now present our joint nodule segmentation and recognition network (JNSC) and its construction, which consists of the following step: 1) data pre-processing and data augmentation (DPA), 2) multiscale voxel-based feature-extraction and nodule size estimation (MVFNSE), 3) pseudo-label assignment for marginal samples (PSA), and 4) jointly-optimized nodule segmentation and classification (JNSC). In the DPA step, the training samples are generated from the CT scans data after standard processing procedure. Moreover, additional training samples are generated using data augmentation technique to improve the robustness of the neural networks against various variations such as rotation of the input, etc. The input voxel is assumed to be a voxel cube with size 64 × 64 × 64. Next, we shall introduce the network architecture and the details of the above four steps will be presented.

A. NETWORK ARCHITECTURE

The proposed joint nodule segmentation and recognition network (JNSC) is shown in Figure 3. It adopts the V-Net [37] as the backbone as the V-Net adopts a multiscale encoder-decoder architecture, and it can perform pixel-wise segmentation. The upper and lower branches form the encoder and decoder in a V-Net architecture where the input voxels are segmented to yield the segmented output at the left lower corner. The encoder and decoder are arranged in a multiscale manner where features are extracted at each scale via the voxel-based feature extraction layer (see also Figure 4). The multiscale features and the nodule size are also estimated in the MVFNSE step which are then concatenated (denoted by the block CC in Figure 3) for predicting whether the current block is a nodule, and whether they are benign or malignant (the middle path in Figure 3).

In the MVFNSE step of the nodule detection subnetwork, the possible locations of the nodule at each scale are estimated from the initial segmentation outputs to form the nodule location map (NLM), which consists of bounding boxes containing potential nodules (as is shown in Figure 4).

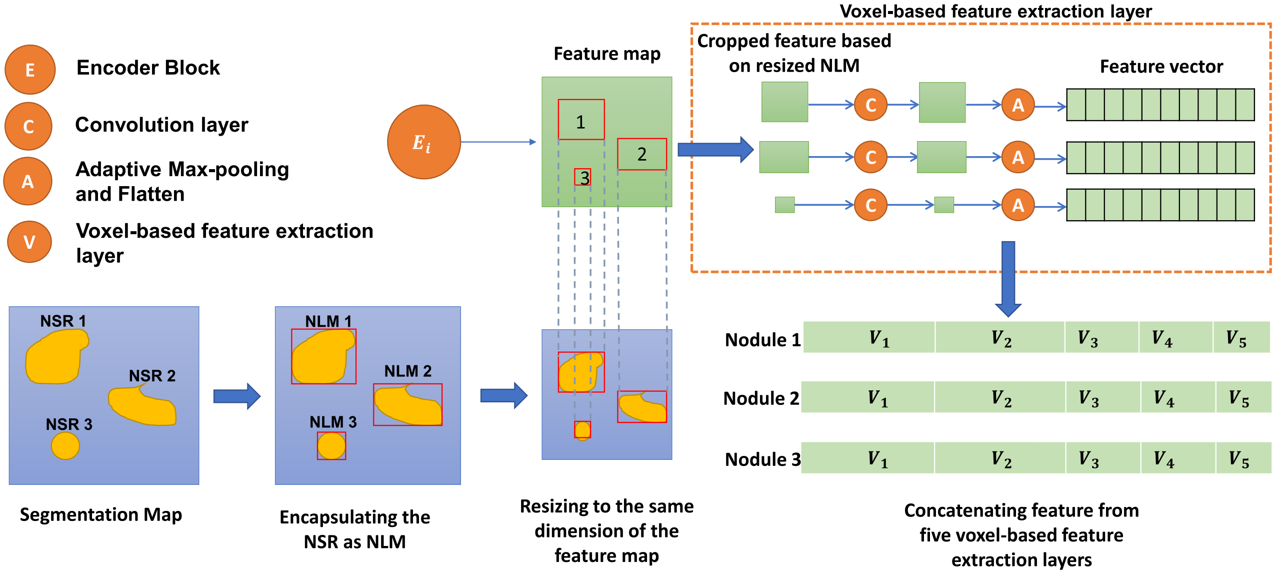

Encoder Block
Convolution layer
Adaptive Max-pooling and Flatten
Voxel-based feature extraction layer
E
C
A
V
E_i
Feature map
Voxel-based feature extraction layer
Cropped feature based on resized NLM
Feature vector
1
2
3
NSR 1
NSR 2
NSR 3
NLM 1
NLM 2
NLM 3
Nodule 1
Nodule 2
Nodule 3
V_1
V_2
V_3
V_4
V_5
Segmentation Map
Encapsulating the NSR as NLM
Resizing to the same dimension of the feature map
Concatenating feature from five voxel-based feature extraction layers

FIGURE 4. Illustration of the proposed multi-scale voxel-based feature extraction. Our implementation is in 3-D while a 2-D slice of the 3D volumetric image is shown here for sake of presentation. The nodule specific region (NSR) is obtained by applying a threshold to the nodule segmentation map. The nodule location map (NLM) is generated as 3D bounding boxes encapsulating the NSR, which is introduced to tolerate the irregular shape of the potential nodules. Nodule specific features are extracted at the location of NLM and are fed into the voxel-based feature extraction layer. Finally, the flatted feature vectors from multiple scales as defined in Figure 3 are concatenated (block CC in Figure 3) for classification using the soft-max criterion. Note, the in the above example, it is assumed that three possible nodules are detected, each with a multiscale feature vector. Each of these candidate nodules will pass through the linear layer and the softmax unit as shown in the middle of Figure 3 to yield the classification output for all these nodules candidates. The number of nodules detected (i.e. the number of NLM) can be variable from each input of voxels.

For classification, the multiscale features of each nodule candidate in each NLM and the nodule size will be fed to the linear layer and softmax layer for classification as shown in the middle path of Figure 3.[9] The PSA step will adjust the label for the marginal nodules to avoid possible overfitting of the marginal samples. The feature vector, together with the segmentation outputs, enables us to jointly optimize the segmentation and classification in a single network at the JNSC step. Training and other details of the above operations will now be discussed.

B. DATA PRE-PROCESSING AND AUGMENTATION (DPA)

The LIDC-IDRI dataset consists of CT scans from seven institutions. Therefore, the pixel spacing and slice thickness may vary on different scans. To reduce the variation from inconsistent resolution, we simply normalize all scans into a resolution of 1.0 mm × 1.0 mm× 1.0 mm by spline interpolation. Besides, the raw CT images are clipped to between -1000 and 400 Hounsfield unit (HU), which can reduce the effect of air and bone in the images. The last step is normalizing the CT images to zero mean and unit variance as commonly used in training neural networks. n each epoch, we extract two voxels from each scan. One of the voxels consists of a nodule, and if a scan has multiple nodules, we randomly pick one of the nodules every time. The other voxel is extracted from the normal region, which does not include any nodule. The motivation for sampling voxels from nodules is to increase the occurrence of the nodule in the training data while sampling other position is to encourage the network to distinguish other body tissues better.

Different from many studies [15], [16], which mainly consider nodule classification, we do not require the nodules to be located in the center of the voxels. To reduce overfitting and improve the generalization ability of the network, we further adopt data augmentation by random rotating the extracted voxels. The rotation is done in one of the x-y plane, x-z plane, and y-z plane with equal probability at each time. To avoid the blank region caused by rotation, we only rotate the image with one of the following angles $[0°, 90°, 180°, 270°]$ with equal probability.

C. MULTISCALE VOXEL-BASED FEATURE EXTRACTION AND NODULE SIZE ESTIMATION (MVFNSE)

As mentioned, we choose the V-Net [37] as the backbone of our JNSC as the V-Net adopts a multiscale encoder-decoder architecture as it can perform pixel-wise segmentation. The multiscale voxel-based feature extraction has three steps: i) generation of the nodule location map (NLM), ii) extraction of the multiscale features, and iii) concatenation of the nodule size information to the feature vector. We summarize these procedures in Figure 4.

[9] It should be noted that there may be more than one nodule candidate (or none) detected inside each voxel volume, each with its own multiscale feature vector and each of these feature vectors will pass through the linear layer and the softmax unit to yield the classification output for all these nodules candidates (please refer to Figure 4 for more details) inside the voxel volume.

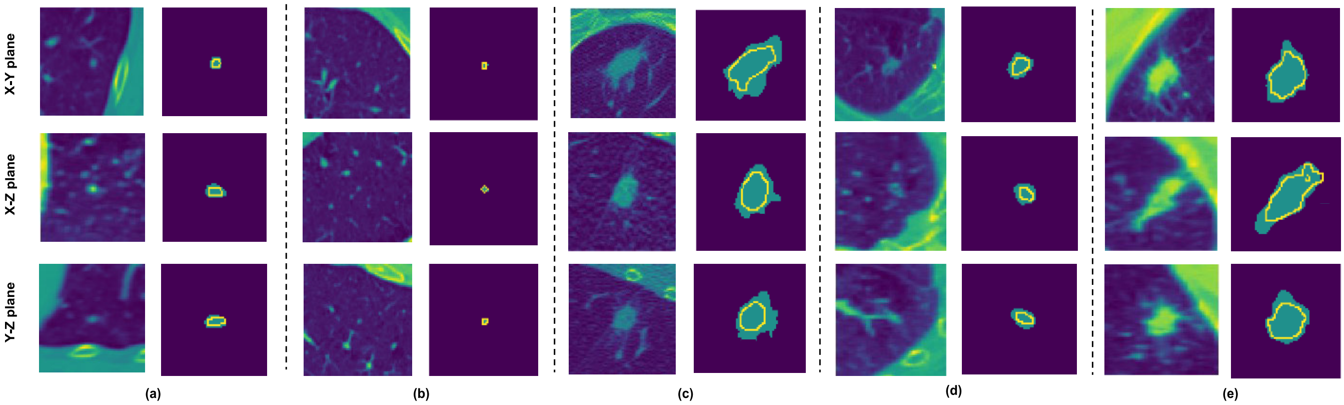


FIGURE 5. Examples of nodule segmentation results. (a)-(e) show five examples of nodule segmentation results along the x-y, y-z and x-z planes, of which (a) and (b) are benign nodules and (c), (d) and (e) are malignant nodules. The yellow contours denote the ground truth nodule boundary annotated by at least three radiologists. The final segmentation result is a binary map obtained by threshold the network output having a value from 0 to 1. Thus, the segmentation map will depend on the applied threshold. In the above illustration, a conservative threshold of 0.4 is used. For the best performance, it can be further optimized via cross validation. It should be noted that the CT images (nodules) are 3D volumetric images and the 2D images (nodules) shown above are their x-y, y-z and x-z cross sections.

To generate the nodule location map, the network is trained on the pixel-wise segmentation from radiologists. Therefore, we can acquire the corresponding nodule probability map from the output of the detection network. The nodule probability map contains the probability of each pixel being classified as nodules. Note that the dimension of the map is identical to the input voxel, which is 64 × 64 × 64. Afterward, we empirically use a detection threshold of 0.4 (40%) to include more suspicious regions for detection. The probability map is then transformed into a binary segmentation map, where 1 represents nodules, and 0 represents non-nodules. Because the shape of the detected nodule is irregular at this stage, as shown in Figure 5, we propose to draw a bounding box[10] encapsulating each nodule to tolerate the irregular shape and reduce the variance in extracting nodule specific features. Then the region inside the box is called a nodule-specific region (NSR). The NSR is found based on its voxel connectivity in the binary map [54]. It should be noted that the segmentation results at this stage may contain errors, say a single or small patch of voxels may be detected, which are likely to be false positives. Therefore, the NSR extracted may be false positives. Fortunately, these false positives are not that many, and their labels are available. Therefore, they are also extracted and will be labelled as non-nodule against benign and malignant nodules, and this preliminary decision information can then be corrected at the classification stage. To this end, we pre-train the detection network at initialization so as to simplify its joint training with the classification network.

Compared with pixel-wise NSR, using the NSR for feature extraction the following benefits. Firstly, accurate morphology information is prone to segmentation errors. Secondly, it allows information/features surrounding the nodules to be extracted for performing the classification at the final stage. Finally, even if the segmentation is extremely accurate, it may be smeared by the subsequent convolution layers. Therefore, more emphasis should be paid on the features of the nodule voxels as well as its neighborhood. Hence, the final nodule location map (NLM) is then generated based on NSR to tolerate the mentioned effect.

TABLE 3. Convolution layer parameter.

Name	Type	Stride	Kernel Size	Input Channel	Output Channel
C_1	Conv	1	(5,5,5)	N	N
C_2	Conv	1	(3,3,3)	1	16
C_3	Conv	1	(5,5,5)	32	2
C_4	Conv	2	(2,2,2)	N	$2N$
U	De-conv	2	(2,2,2)	N	$N/2$

For the extraction of the multiscale feature, the size of the input voxel is 64 × 64 × 64, which will be down-sampled 4 times in the encoder network. Therefore, we have feature maps of size 64, 32,16,8,4 as shown in Figure 3. The NLM is also down-sampled to the same size of each feature maps, as shown in Figure 4. For each feature map, we crop the feature from the corresponding location in NLM. Following the feature cropping, we further add 1 × 1 convolution layers to aggregate inter-channel information. An adaptive max-pooling operation on the features is then performed where the features from the first two voxel-based feature extraction layers V_1, V_2 are pooled into a uniform spatial size of 2 while those at the third to fifth layers V_3, V_4, V_5 are pooled into a spatial size of 1. Because of the adaptive max-pooling layer, the length of the final feature vector is invariant to the size of the NSR, and it can be flattened and concatenated among different scales.

The last step of the MVFNSE step is to concatenate the nodule size information on the feature vector. It is widely recognized that nodule size is highly related to the malignancy

[10] Note, the bounding box is used for feature extraction. The final segmentation output will be derived from these features as shown in the lower branch of the joint network in Figure 3

level, and larger size usually increases with the probability of being malignant. The pooling operation in step 2 is invariant to nodule size, and therefore, we can directly add the information to the concatenated features. The nodule size is estimated as:

$$V = \frac{1}{10}\sqrt[3]{P} \tag{1}$$

where V is the estimated nodule size and P is the number of pixels for the given nodule in the NSR. The nodule diameters vary from 3 mm to 30 mm and the resolution of segmentation result is 1.0 mm × 1.0 mm × 1.0 mm. Since large values in the features may dominate the classification performance, the estimated size is scaled by a factor of 0.1, which is determined empirically. It was found that the performance is relatively insensitive to the choice. The final feature used for classification consists of concatenated multiscale features from step 2 and a dimension of estimated nodule size. Each vector will pass through the linear layer and the softmax unit to yield the classification output for all the nodules candidates detected inside the voxel volume (please refer to Figure 4 for more details).

D. PSEUDO-LABEL ASSIGNMENT FOR MARGINAL SAMPLES (PLA)

In nodule classification, some nodules are labelled by 1 or 2 radiologists. However, radiologists are likely to be inconsistent on the malignancy level, especially all with a marginal level of malignancy. To address this issue in training our network, we propose a pseudo-label approach for those marginal nodules to alleviate the effect caused by label uncertainty. More precisely, the cross-entropy loss we based for training is given by:

$$L_{ce} = -\frac{1}{N}\sum_{i}^{N} T_i \log(p_i) + (1 - T_i)\log(1 - p_i) \tag{2}$$

where T_i and p_i are the malignancy score and the predicted probability by the network respectively. Here, the labels "0" and "1" represent the benign and malignant nodules respectively. However, due to label uncertainty, T_i is usually not chosen as either 0 or 1 and the following soft-label is preferred:

$$T_i = 0.25(M_i - 1) \tag{3}$$

where M_i is the MS for the i-th nodule.

Here, we re-estimate the underlying label called the pseudo-label $\hat{p}_i$ for addressing those marginal nodule samples and continuously adapting them based on the network prediction obtained as well as the MS. Specifically, by initializing the initial value of the pseudo-label with the soft-label in (3), the resultant loss function using the pseudo-label is given by

$$\tilde{L}_{ce} = -\sum_{i} \hat{p}_i \log(p_i) + (1 - \hat{p}_i)\log(1 - p_i) - \alpha\left(\hat{p}_i - T_i\right)^2 \tag{4}$$

where α is a regularization parameter that balances the influence of MS and network prediction on the pseudo-label. If α is large, the pseudo-label will mainly depend on MS and $\tilde{L}_{ce}$ will approach the cross-entropy loss. On the contrary, if α is small, the pseudo-label is dominated by the network output, which is not desirable because the training information T_i cannot guide the learning process. The influence of alpha on the classification result will be further studied in the experiment section. By introducing the regularization in $\tilde{L}_{ce}$, the pseudo-label becomes adjustable. The gradient of $\tilde{L}_{ce}$, which is required for performing the optimization, is given by:

$$\frac{\partial \tilde{L}_{ce}}{\partial p_i} = \frac{1 - \hat{p}_i}{1 - p_i} - \frac{\hat{p}_i}{p_i} \tag{5}$$

$$\frac{\partial \tilde{L}_{ce}}{\partial \hat{p}_i} = \log\left(\frac{1 - p_i}{p_i}\right) + 2\alpha\left(\hat{p}_i - T_i\right). \tag{6}$$

We now briefly explain the advantage of the proposed pseudo-label approach. Firstly, if the network prediction result is consistent with the MS, the first term in (6) will increase the certainty of the pseudo-label, which will implicitly increase the weight on this sample. For example, if the network prediction value p_i is 0.7, the first term in (6) is negative and the corresponding $\hat{p}_i$ will become larger during optimization. This larger $\hat{p}_i$ will increase the absolute value of the gradient in (5), which in turn will encourage learning from the sample. On the other hand, if the network prediction is contradicting the MS, forcing the network to fit the sample may lose the generalization ability of the network due to the MS noise. Thus, for such samples, the first term in (6) will drive the $\hat{p}_i$ towards p_i, which will implicitly lower the weights of learning from such samples. Besides, the second term in (6) is used to penalize the pseudo-label for large deviation from T_i, which avoids large fluctuation in the pseudo-variable. Thus, the pseudo-label can be regarded a weight reflecting our confidence on the marginal label given the original annotation as well as the current network knowledge.

The pseudo-label can be updated using gradient descent:

$$\hat{p}_i^{t+1} = \hat{p}_i^{t} - r_2 \frac{\partial \tilde{L}_{ce}}{\partial \hat{p}_i} \tag{7}$$

where r_2 is the learning rate for the pseudo-labels. Since the pseudo-label represents the probability of malignancy, it should be bounded between 0 and 1. Therefore, the update in (7) is further projected on these bound constraints as:

$$\hat{p}_i^{t+1} = \begin{cases} 0, & \mathit{if}\ \hat{p}_i^{t+1} < 0 \\ \hat{p}_i^{t+1}, & \mathit{otherwise} \\ 1, & \mathit{if}\ \hat{p}_i^{t+1} > 1. \end{cases} \tag{8}$$

E. JOINTLY-OPTIMIZED OF NODULE SEGMENTATION AND CLASSIFICATION (JNSC)

The proposed JNSC network comprises of a nodule detection module and a nodule classification module with a shared structure for information exchange. The features for nodule classification can be extracted from the encoder of the nodule detection module, which provides additional information for feature extraction. For training this joint network, we first train the nodule classification network for 100 epochs using the pixel-wise cross-entropy loss:

$$L_{seg} = -\sum_{i} S_i \log(p_i) + (1 - S_i) \log(1 - p_i), \tag{9}$$

where S_i denotes the probability of the pixel belonging to the nodule. After the initialization of the nodule segmentation network, the output segmentation may still generate many false-positive nodules. To overcome this problem, we extract not only features for true positive nodules, but also those false positive nodules for classification. Moreover, the false-positive nodules are labelled as non-nodule with probability 1.

The network is then trained jointly. For the following 100 epochs, we do not update the pseudo-label because the network prediction is unstable at these early stages. Finally, the segmentation and classification modules are properly initialized, and the network can be optimized using the following cost function:

$$L = L_{seg} + \left(L_{ce} + \tilde{L}_{ce}\right). \tag{10}$$

Different from [34] where the segmentation and classification networks are trained iteratively, the parameters in both the detection and classification modules of the proposed JNSC can be updated simultaneously.

Additionally, because the parameters in our network are differentiable, the parameters can be optimized by efficient optimizer like Adam. In each epoch, which consists of a number of iterations, the network parameters are updated at each iteration. Since the parameters are likely to sufficient training after each epoch, each pseudo-label will be updated after each epoch. To reduce the effect of previous gradient, the pseudo-labels are directly updated by gradient descent without momentum.

F. IMPLEMENTATION DETAILS

Our proposed network mainly consists of three convolution layers, and the parameters of the convolution layers are listed in TABLE 3. Each convolution layer is followed by an instance normalization [55] layer and a ReLU layer. The Adam optimizer optimizes the parameters in our network with default settings in PyTorch. The initial learning rate is 0.001, and it is decreased every 250 epochs with a factor of 0.2. The maximum training epoch is set to 1000 and the batch size is 12. The spatial dropout strategy is applied to the 3-D convolutions with a dropout rate of 0.1. We also employ gradient clipping during the optimization by clipping the gradient to 1 if the L_2 norm of the gradient is larger than 1 for the sake of stability.

Since the number of benign nodules is almost 2 times that of the cancer nodules, class-imbalance problem will occur. Specifically, the non-nodule pixels in L_{seg} and benign nodules in L_{ce} will dominate the training phase if no balancing mechanism is used. To leverage this problem, we, therefore, adopt different weights in the cross-entropy loss. Specifically, in the nodule detection module, they are chosen as 0.01 and 0.99 for nodule and non-nodule pixels, respectively. On the nodule classification module, the weights for the malignant, benign, and non-nodule classes are set to 0.35, 055 and 0.1, respectively. In principle, the weights are chosen as the ratio of samples in the two classes. Of course, one can increase the weight to allow the network to focus more on the cancer samples. The weights in the nodule segmentation also adopt a similar criterion, where the weight of non-nodule pixels is about 100 times that of the nodule pixels. The performance does not depend critically on these weights as long as they can reflect the difference in the sample number between classes.

V. EXPERIMENT RESULTS

A. NODULE DETECTION

We first evaluate the performance of the nodule detection performance of our JNSC and other state-of-the-art algorithms on the LUNA16 dataset. The standard ten-fold cross-validation of LUNA16 competition is adopted and the standard evaluation script is used to compute the Free-response Receiver Operating Characteristic FROC curve.

To extract the nodule candidate from the 3-D nodule detection probability maps, we first set the detection threshold to 0.4 and label the connected regions in the segmentation map based on their voxel connectivity [54]. Then, the region proposals can be extracted from the labelled map, and the center is calculated by the centre of mass of the proposed regions. Lastly, we use non-maximum suppression [56] on the proposed regions and exclude those with diameter less than 3 mm. Figure 5 shows five examples of our nodule detection results with a wide range of nodule diameters. To visualize the 3-D segmentation result in a 2D figure, we present the cross sections of the nodules as well as the corresponding segmentation maps along the *x*-*y*, *y*-*z* and *x*-*z* planes. It can be seen from Figure 5 that the detected regions are relatively larger than the ground truth.

Moreover, as shown in Figure 5 (b), our network can detect tiny nodules while distinguishing the small nodule from other body tissues like vessels. The resolution of CT scans in the *z*-axis is much lower than the resolution in the *x*- and *y*-axis. For instance, the resolution in the *x*- and *y*-axis is usually 0.7 mm per pixel, but the resolution in the z-axis can vary from 1.25 to 3 mm per pixel. To ensure similar accuracy in the three dimensions, we employ interpolation to convert the resolution along the three dimensions to 1 mm per pixel. The result shows that our network can tolerate the problem of different resolutions and achieve similar performance on the three dimensions.

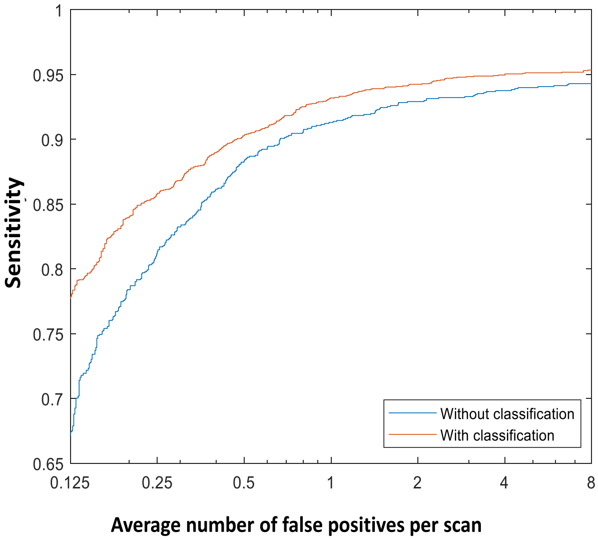


FIGURE 6. FROC of nodule detection of our JNSC. The orange line is the result by joint optimization of detection and classification. The blue line shows the result of JNSC without classification where the detection outputs are regarded as final result.

1) PERFORMANCE OF JOINTLY OPTIMIZED NODULE DETECTION

To verify the effectiveness of the structure, we compare the performance of our proposed approach on nodule detection under standard settings with and without the classification phase.[11] The FROC under the two settings is

shown in Figure 6. As shown in Figure 6, the jointly-optimized approach significantly outperforms the detection only case. More specifically, the sensitivity of JNSC with classification at 0.125 false positives per scan is 0.776, while that of the classification only case is 0.630. Because the undetected nodules at low false-positive levels are primarily small ones, the joint optimization approach is capable of significantly improving the detection performance on such tiny nodules, which is essential to the early detection of the disease.

The detection module of the JNSC is trained using the pixel-wise cross-entropy cost function. Since large nodules have more pixels, they will dominate the performance at the training phase as the gradients are mainly backpropagated from these large nodules. Consequently, the small but important nodules can easily be neglected. Moreover, despite the shortcut path, the gradient backpropagated from the decoder may be less sensitive to the small nodules. On the other hand, the direct path of our JNSC to every encoder helps to propagate the gradient from the classification network to train the encoder so that undetected small nodules can be distinguished from the non-nodule region during the training phase. It can facilitate the detection of those small nodule regions which are not detected by the detection network alone. It is also observed that the information backpropagated from the direct path is much more direct and effective than those back-propagating from the gradient of the classification network, due to the large separation between the classification output and the detection encoder. Moreover, the classification phase performs the simultaneously false-positive reduction, which further improves the detection rate.

Additionally, from Figure 6, the proposed JNSC with and without classification achieves respectively an impressive sensitivity of 0.953 and 0.942 at 8 false positives per scan, which further demonstrates the effectiveness of joint optimization.

2) COMPARISON WITH STATE-OF-THE-ART ALGORITHMS ON NODULE-DETECTION

We adopt the standard train-test split in the LUNA16 competition for a fair comparison. The competition performance metric (CPM) is defined as the average sensitivity at 0.125, 0.25, 0.5, 1, 2, 4, 8 false positives per scan. The result is shown in TABLE 4.

TABLE 4. Comparison of Competition Performance Metric (CPM) with the State-of-the-Art Algorithms on LUNA16 Dataset.

Methods	CPM
ZNET [13]	0.811
Aidence [13]	0.807
DeepMed [8]	0.832
SDFPR [4]	0.831
DeepLung [9]	0.842
Attention 3D-CNN [25]*	0.878
S4ND [35]	0.897
JNSC (without classification)	0.866
JNSC*	**0.902**

*The additional false positive refinement is applied to the detection result where the false positives in the training phase is collected and trained for another time. The methods using additional false positive reduction data is not reported for fair comparison. The 3D-CNN and S4ND only produce bounding box instead of detailed segmentation.

ZNET [13] and Aidence [13] are the participants of the competition and win the first and second places. ZNET uses a 2-D U-Net [36] architecture and computes the nodule probability map slice by slice. Though the 2-D network cannot fully utilize the 3-D structure of the nodules, the parameters to be trained are much less than the 3-D network. The ZNET achieves a CPM of 0.811 and a sensitivity of 0.915 at 8 false positives per scan. The detailed method of Aidence is unavailable because of commercial confidentiality. The Aidence also achieves a CPM of 0.807 on the competition.

Despite the advantage of having fewer parameters in 2-D networks, 3-D neural networks are preferred recently due to its ability to detect 3-D patterns and the increased availability of computational power. DeepMed [8] was extended to a 3-D architecture, but the network is relatively shallow. Also, an independent false-positive network is trained

[11] In the detection without classification case, the outputs from the detection phase are directly evaluated using standard evaluation script. The joint training case will further classify the outputs as non-nodule, benign and malignant. Afterwards, benign and malignant nodules in the final result are evaluated. The result shows the classification stage can significantly reduce false positive nodules.

to distinguish the detected candidates. Our JNSC is deeper than [8], which can capture more complicated structures and the false-positive reduction stage is implicitly incorporated into the JNSC. SDFPR [4] and DeepLung [9] adopt faster R-CNN structure which performs the regression of nodule location as well as probability but not pixel-wise segmentation as in our JNSC. Their encoder-decoder architecture is similar to our network, but our network has an additional shortcut path to the encoder. Hence our network can be more sensitive at a low false-positive level. For example, our JNSC obtains 0.776 sensitivity at 0.125 false positives per scan, while SDFPR [4] is approximately 0.62.

The 3D-CNN in [25] uses a combination of 2-D and 3-D networks where the 2-D network is used for candidate detection while the 3-D network is used to classify false positives. The candidate detection network can benefit from the pre-trained VGG network while the 3-D network can only be trained from scratch. The conditional non-maximum suppression in [25] is superior to normal NMS. However, the two networks are still independent of each other while our network adopts a joint optimization approach. The result shows that the CPM of our JNSC outperforms [25] by 2.4%.

The S4ND [35] employs a single end-to-end network and replace convolution blocks with densely connected convolution blocks. The results from [35] show that densely connected block outperforms regular residual connection.

However, S4ND does not perform false positive reduction after detection, while a considerable number of tiny nodules are, in fact, body tissues. Our JNSC jointly achieves false positive reduction with the help of the classification network and outperforms state-of-the-art algorithms.

B. NODULE CLASSIFICATION

We now evaluate the nodule classification performance using the LIDC-IDRI dataset. As described in section VI, the uncertain nodules are excluded from evaluation.[12] We randomly split the 1018 scans into ten subsets and adopt 10- fold cross-validation to report the result. Additionally, each fold is trained five times to reduce the effect of network initialization. Note that the uncertain nodules in the testing set are excluded from calculating the accuracy.

The classification network in our proposed JNSC requires the segmentation result from the nodule detection network to perform multiscale voxel-based feature extraction. In order to compare with other classification only algorithms, those undetected nodules are directly labelled as benign. We have also neglected the false positives in the nodule detection process.

1) COMPARISON WITH THE STATE-OF-THE-ART ALGORITHMS ON NODULE CLASSIFICATION

To our knowledge, few studies report the end-to-end result, and therefore, the comparisons can hardly be absolutely fair. Therefore, we report algorithms using the same MS and CT scans as ours. It should be noted that our system is end-to-end, which is more challenging than just classification of the nodules as nodules detection process may itself be error-phone. On the other hand, our framework is closer to a realistic operating environment.

TABLE 5. Lung Nodule Classification Results of the State-of-the-Art and The Proposed Algorithms on LIDC-IDRI Dataset.

Methods	Results (%)			
	Accuracy	Accuracy (Balanced)	Sensitivity	Specificity
MC-CNN [15]*	87.14	85.00	77.00	93.00
Fuse-TSD [40]*	88.73	87.64	84.40	90.88
TMME [14]*	91.01	89.15	83.83	**94.56**
2D-MV-KBC* [16]	**91.60**	90.26	86.52	94.00
3D-MV-KBC* [16]	90.07	88.24	81.50	94.98
JNSC (without detection)	89.57	88.76	86.54	90.99
JNSC	90.82	**90.29**	**88.79**	91.78

*The algorithms listed here for comparison solely perform classification task with pre-cropped nodules as inputs based on ground truth location. Meanwhile, our JNSC algorithm operates in an end-to-end manner, which does not require the information of nodules location. It is shown here just to illustrate the classification performance of our network. In fact, it is far more challenging to simultaneously detect and predict the nodule location and its malignancy, which is however the situation encountered in practical applications.

The accuracy, sensitivity, and specificity of the proposed approach on nodule classification is reported and compared with state-of-the-art algorithms. Moreover, as there are more negative samples than positive samples in the dataset, the network is likely to perform better on the negative samples (thus, the specificity is usually higher than the sensitivity). Hence, the negative samples will have more influence on the accuracy. To illustrate the overall performance of the algorithms despite these effects, we also report the balanced accuracy to better reveal and compare the performance. More precisely, the definition of the balanced accuracy is

$$BalancedACC = \frac{Sensitivity + Specificity}{2} \tag{11}$$

Furthermore, to verify the effectiveness of the segmentation information in the proposed joint-optimization approach, the NSR is replaced by the ground truth region and the JNSC is trained without segmentation, i.e. it is operated in classification only mode. Particularly, we do not backpropagate the gradient from the segmentation module so that the encoder is trained only by the classification network. The results show that the joint training performs better than the classification only mode.

As shown in TABLE 5, our proposed JNSC achieves the highest balanced accuracy and sensitivity among the algorithms. Although the 2D-MV-KBC [16] has the best accuracy, the higher accuracy results from the imbalanced classes where specificity can contribute more to the overall accuracy. Moreover, 2D-MV-KBC only considers the classification on

[12] We follow the common practice that nodules with MS = 3 are excluded, as such these nodules are uncertain as to benign or cancer.

the extracted nodule patches while our algorithm does not require the nodule location to be known in the training phase. Although the 2-D U-Net is adopted for labelling the nodule from the patch, training the network on the extracted patches is still much easier than for the entire CT scans because the extracted regions will be free from the interference of many other body tissues. Moreover, it is required to train 27 independent networks in 2D-MV-KBC so that their results can be aggregated. Its complexity will be significantly increased. In [16], a three-dimension network with 3 independent networks based on ResNet-50 is also proposed. Experimental results show the 2-D network outperforms the 3-D network, which is likely due to the fact that the 2-D network can benefit from the pre-trained ResNet-50 network.

On the other hand, the proposed 3D JNSC can be trained from scratch since the nodule detection network can provide additional information in the form of regularization to alleviate the overfitting problem caused by insufficient training samples. Moreover, the encoder in our JNSC is trained on the whole CT image which can also distinguish other body tissues for nodule detection. The experiment results show that the joint detection and classification framework is superior to the classification only approach with an improvement of 1.25% accuracy. Overall, our approach is more practical for automatic cancer and nodules detection.

The MC-CNN [15] is the first to introduce the approach of cropping nodule-specific feature, which is similar to our multiscale feature extraction method. However, our algorithm differs from [15] in that: i) our extraction is based on the nodule detection while MC-CNN uniformly extracts multiscale feature by using successive max-pooling on each feature, ii) MC-CNN requires nodule-centric inputs (i.e. the first identification of the location of the nodules to be classified by the network) while our JNSC is more flexible in that the nodule can occur anywhere in the voxels and our feature extraction is invariant to the nodule location. Moreover, MC-CNN employs 2-D convolution given the 3-D inputs (i.e. as multiple 2D channels), and hence the information among slices may not be efficiently exploited.

In conclusion, our JNSC is at least comparable to the state-of-the-art nodule classification algorithms with respect to accuracy, sensitivity, and specificity for classification alone task. On the other, the JNSC is fully automatic and does not require pre-selected inputs of the detected nodules. Actually, it can be operated in an end-to-end manner.

2) ANALYSIS OF THE EFFECT OF PSEUDO-LABEL

To examine the effect of labels on the classification performance of our approach, experiments are performed on the following three cases: 1) assigning hard label to nodules, by which each nodule is labelled either "0" or "1"; 2) substituting the hard label by soft label, by which nodule is labelled based on MS in (3); and 3) replacing the soft label by our pseudo-label for the marginal nodules. The results are shown in TABLE 6.

TABLE 6. Evaluation of the Nodule Classification Performance over Regularization Parameter on L_{ce}^2.

Label	Results (%)			
	Accuracy	Accuracy (Balanced)	Sensitivity	Specificity
Hard Label	88.38	87.39	84.67	90.11
Soft Label	89.51	88.94	87.29	90.58
Mixture ($\alpha = 1$)	89.86	88.46	84.61	**92.30**
Mixture ($\alpha = 5$)	90.05	89.27	87.10	91.44
Mixture ($\alpha = 10$)	**90.82**	**90.29**	**88.79**	91.78
Mixture ($\alpha = 20$)	90.30	89.82	88.41	91.24

Apparently, the performance of using the hard label is the worst among the three methods. This phenomenon reveals that classification in the biomedical area is different from natural image recognition because ground truth is not absolutely correct. Inconsistent labels may arise in the biomedical area due to human errors. It is noted that we are not proposing a physical model to accurately model the probability that the label is uncertain. Instead, we empirically estimate the reliability of the marginal samples and its associated labels via the cross-entropy loss function so as to prevent the network from overfitting these less reliable samples, which affect the overall performance. Consequently, assigning soft-label in classification can significantly improve classification accuracy. However, soft label requires the estimation of probability, which may also introduce additional noise when only a few annotations are available. In this study, we assume that the nodules annotated by at least three radiologists are reliable, while nodules annotated by less than three radiologists and not highly confident are marginal. We then estimate and update a soft label in the form of pseudo-labels for the marginal nodules based on the annotation and network prediction to reduce the noise mentioned above.

To visualize and validate the effectiveness of the proposed pseudo-label during the training phase, the histograms of pseudo-labels before and after training under different regularization parameter α are shown in Figure 7. Figure 7 (a) plots the initial distributions of pseudo-labels. Note that the data is acquired on a randomly selected fold. We then examine the effect of α on pseudo-labels. As shown in Figure 7 (b), lower α pushes the pseudo-label towards the boundary, where pseudo-labels are similar to hard labels on the marginal samples. This can be explained by the fact that the network prediction results dominate the pseudo-label update. However, this is undesired because little information can be learned from the marginal nodules. The result shows that the network tends to fit the benign nodules, and the highest specificity is achieved and the overall performance is inferior to the soft label. When α grows larger, we observe from Figure 7 (d) that α still forces the pseudo-label towards the boundary, but the changes are less severe than before.

Moreover, the network increases the malignancy probability of some benign nodules, revealing that the network treats such nodules as malignant. The network is not trained to mine the marginal samples. Instead, it relies more on the certain data for classification as the marginal samples may not

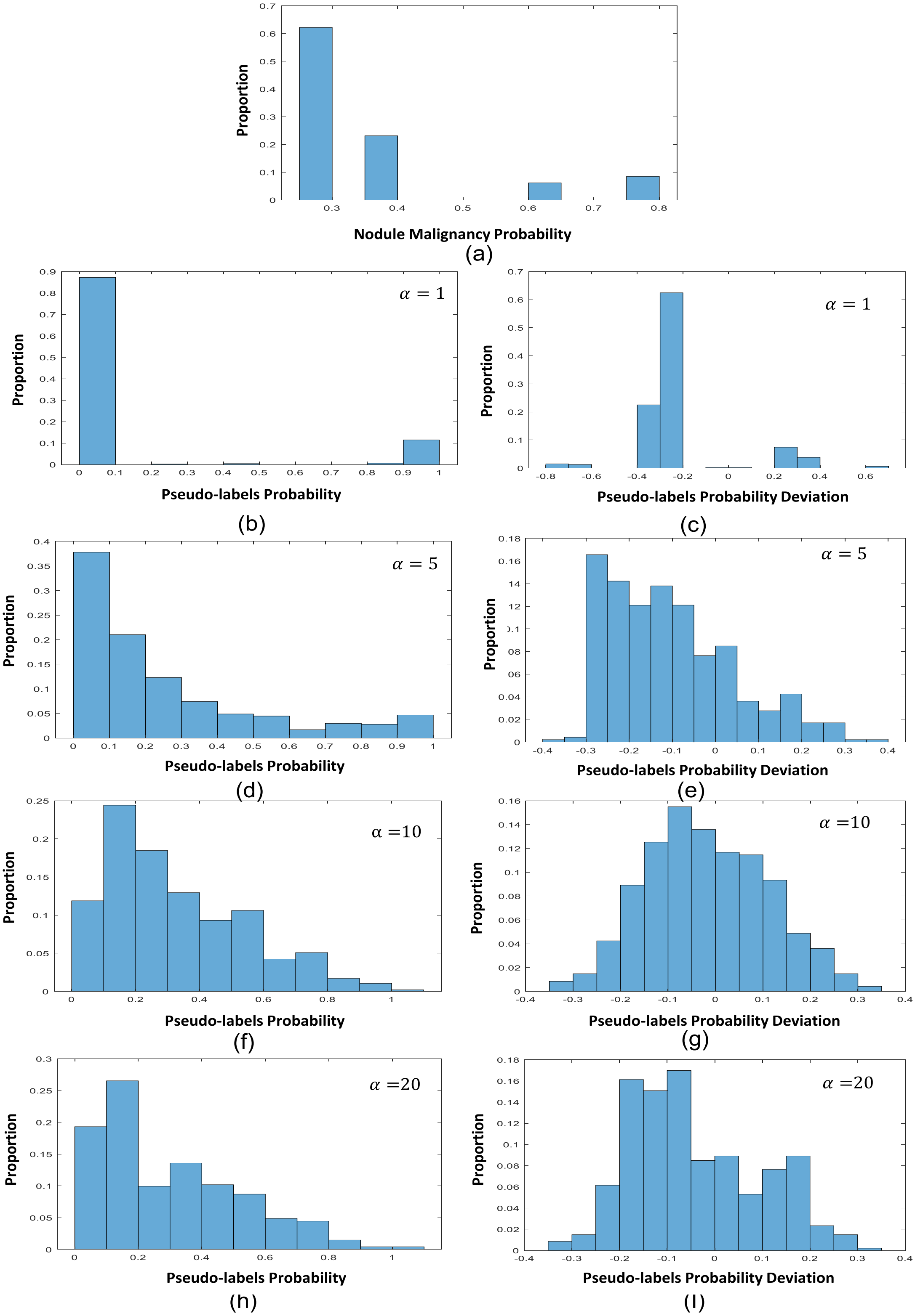


FIGURE 7. Visualization of pseudo-label before and after optimization on a random selected fold. The nodules with malignancy probability less than 0.5 are regarded as benign nodule. The probability deviation is calculated by subtracting the pseudo-label after training to its initial value. (a) is the histogram of initial pseudo-labels. The four figures on the left column are the histograms of pseudo-labels after the optimization under different regularization power. The four figures on the right are the deviation of pseudo labels after the optimization.

be absolutely correct due to label uncertainty. The problem is commonly encountered in biomedical applications where ground truth may not be precisely gauged from limited human labels. This is in great contrast to natural image classification and language understanding where such labels are usually correct, except for occasion human errors. In summary, the pseudo-label approach addresses the label uncertainty by incorporating the network prediction results or knowledge in addition to the label provided.

Next, we observe that the regularization power does not grow linearly with increasing α. Figure 7 (h) shows that $\alpha = 20$ performs similarly as 10. When α is set to 10, the majority of pseudo-labels only vary in a small range as is shown in Figure 7 (g). It is reasonable that the network prediction and ground truth annotation are balanced under $\alpha = 10$, thus achieving the best overall performance. Theoretically, when α grows to infinity, the annotation should govern the pseudo-label, which is identical to soft label. We do not explore larger α and 10 is selected as the default value in this study.

3) ANALYSIS ON MULTISCALE FEATURE EXTRACTION

Our proposed JNSC relies on the features from several encoders to perform nodule classification. Hence, it is important to evaluate the effect of the number of the multiscale features on the classification performance. The experiment is designed to observe the classification performance over concatenating features from first encoder V_1 to the deepest level V_5. Note that the nodule size is still concatenated to the feature.

As shown in TABLE 7, the classification performance generally improves as deeper features are added. Although discarding the feature up to V_4 yields higher accuracy and specificity, the performance is comparable to that of concatenating all features after considering the balanced accuracy and sensitivity. Therefore, to maintain the consistency of the structure, we do not discard the feature from V_5.

TABLE 7. **Performance of nodule classification with multi-scale features.**

Feature	Results (%)			
	Accuracy	Accuracy (Balanced)	Sensitivity	Specificity
V_1	75.70	76.63	79.20	74.06
V_1-V_2	86.86	86.53	85.61	87.45
V_1-V_3	89.50	88.76	86.73	90.79
V_1-V_4	**90.88**	89.77	88.23	**92.13**
V_1-V_5	90.82	**90.29**	**88.79**	91.78

The reason for such a behaviour can be explained as followed. As features of different scales are extracted from the corresponding location in multiscale feature map, the convolution operation can expand the reception field, which means that the extracted features usually represent a larger region in the input CT images. For the features from V_1 and V_2, the effect is negligible. For the feature from V_5, such effect can somewhat affect the classification, especially on small nodules because it may encode the information of other body tissues. Meanwhile, the small nodules are likely benign nodules and thus the specificity decreases after adding V_5 features.

VI. DISCUSSION AND FUTURE WORK

A deep-learning based approach for joint detection, segmentation and classification of nodules from 3-D CT scans has been proposed. Moreover, the concept of pseudo-label has been proposed to tackle the problem of label uncertainty, which is commonly encountered in biomedical data. While most algorithms proposed focus on either detection or classification, the proposed algorithm operates in an end-to-end manner, which provides detection and classification of nodules simultaneously together with a segmentation of the detected nodules. Experimental results show that it outperforms the state-of-the-art nodule detection algorithm, and yields comparable performance as state-of-the-art nodule classification algorithm while classification alone is considered.

While natural images are often in two-dimension, biomedical images, such as CT and MRI, are often in three-dimension. Since it is usually difficult for human to efficiently visualize these three-dimension data for detection, detail segmentation and classification of region of interest, the proposed algorithm offers a promising approach in developing similar computer-aided diagnosis systems.

In this work, we have employed a multi-task framework, which combines the detection and classification in a single network. Such an integrated approach allows essential information to be exchanged between individual subnetworks and lead to higher performance in both tasks. Moreover, in many practical applications, it is required to be able to provide users with the detailed location or morphology of the objects of interest, in addition to the final decision. In this work, we further extend the nodule detection to pixel-wise nodule segmentation, where a more accurate shape or morphology description of nodules can be obtained. Therefore, the present framework may also be useful in related applications.

Some limitations do exist in our study. Firstly, the patient-level prediction is not studied in this work. Secondly, the slice thickness of various CT scans can vary dramatically. The nodule detection competition (LUNA16) manually excludes the scans whose slice thicknesses are larger than 2.5 mm. The diameter of the small nodules is around 3 mm, which is very close to the slice thickness. Therefore, the low and variant resolution on the z-axis is another difficulty in nodule detection, especially for the small nodules. Many studies [57-60] have adopted the deep-learning-based super-resolution approaches to address the problem in CT and MRI images. It is interesting to incorporate the super-resolution into the proposed nodule detection and classification framework.

VII. CONCLUSION

A joint lung nodule detection and classification network for end-to-end lung nodule detection, segmentation and classification subject to possible label uncertainty in the training set has been presented. It operates in an end-to-end manner, which provides detection and classification of nodules simultaneously together with a segmentation of the detected nodules. A 3D encoder-decoder architecture is adopted for better exploration of the 3D nature of the data. The nodule classification subnetwork of the joint network utilizes the features from the encoder output of the detection subnetwork and the multiscale nodule-specific features for boosting the classification performance. This valuable prior information also allows the more complicated 3D nodule classification encoder network to be optimized directly with improved performance on both tasks. Evaluation using the LUNA16 and LIDC-IDRI datasets shows that the proposed nodule detector outperforms the state-of-the-art algorithms and yields comparable performance as state-of-the-art nodule classification algorithms when classification alone is considered. Finally, since our joint detection/recognition approach can directly detect nodules and classify its malignancy instead of performing the tasks separately, our approach is more practical for automatic cancer and nodules detection.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, ''Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,'' *CA, Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[2] M. Burch, S. Kapur, and S. Starnes, ''Lung cancer screening: Insights from a thriving clinical practice,'' *Current Pulmonol. Rep.*, vol. 8, no. 3, pp. 96–103, Sep. 2019, doi: 10.1007/s13665-019-00231-0.

[3] M. Infante, S. Cavuto, F. R. Lutman, G. Brambilla, G. Chiesa, G. Ceresoli, E. Passera, E. Angeli, M. Chiarenza, G. Aranzulla, and U. Cariboni, ''A randomized study of lung cancer screening with spiral computed tomography: Three-year results from the DANTE trial,'' *Amer. J. Respir. Crit. Care Med.*, vol. 180, no. 5, pp. 445–453, 2009, doi: 10.1164/rccm.200901-0076OC.

[4] Y. Qin, H. Zheng, Y.-M. Zhu, and J. Yang, ''Simultaneous accurate detection of pulmonary nodules and false positive reduction using 3D CNNs,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1005–1009, doi: 10.1109/ICASSP.2018.8462546.

[5] P. Monkam, 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, doi: 10.1109/ACCESS.2019.2920980.

[6] F. Ciompi, K. Chung, S. J. van Riel, A. A. A. Setio, P. K. Gerke, C. Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. M. W. Wille, A. Marchiano, U. Pastorino, M. Prokop, and B. van Ginneken, ''Towards automatic pulmonary nodule management in lung cancer screening with deep learning,'' *Sci. Rep.*, vol. 7, no. 1, p. 46479, Jun. 2017, doi: 10.1038/srep46479.

[7] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. P. Naidich, and S. Shetty, ''End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,'' *Nature Med.*, vol. 25, no. 6, pp. 954–961, Jun. 2019, doi: 10.1038/s41591-019-0447-x.

[8] 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 Inform.*, vol. 23, no. 5, pp. 2080–2090, Sep. 2019, doi: 10.1109/JBHI.2018.2879449.

[9] W. Zhu, C. Liu, W. Fan, and X. Xie, ''DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification,'' in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 673–681, doi: 10.1109/WACV.2018.00079.

[10] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards real-time object detection with region proposal networks,'' presented at the Adv. Neural Inf. Process. Syst., 2015.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ''You only look once: Unified, real-time object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[12] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[13] A. A. A. Setio *et al.*, ''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, doi: 10.1016/j.media.2017.06.015.

[14] Y. Xie, Y. Xia, J. Zhang, D. D. Feng, M. Fulham, and W. Cai, ''Transferable multi-model ensemble for benign-malignant lung nodule classification on chest CT,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 656–664.

[15] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, and J. Tian, ''Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification,'' *Pattern Recognit.*, vol. 61, pp. 663–673, Jan. 2017, doi: 10.1016/j.patcog.2016.05.029.

[16] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, and W. Cai, ''Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT,'' *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 991–1004, Apr. 2019, doi: 10.1109/TMI.2018.2876510.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ''ImageNet large scale visual recognition challenge,'' *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[18] S. G. Armato, III, *et al.*, ''Data from LIDC-IDRI,'' *Cancer Imag. Arch.*, 2015, doi: 10.7937/K9/TCIA.2015.LO9QL9SX.

[19] S. G. Armato *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.*, vol. 38, no. 2, pp. 915–931, Jan. 2011, doi: 10.1118/1.3528204.

[20] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, ''The cancer imaging archive (TCIA): Maintaining and operating a public information repository,'' *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013, doi: 10.1007/s10278-013-9622-7.

[21] J. Long, E. Shelhamer, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[22] R. Cipolla, Y. Gal, and A. Kendall, ''Multi-task learning using uncertainty to weigh losses for scene geometry and semantics,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7482–7491, doi: 10.1109/CVPR.2018.00781.

[23] O. Sener and V. Koltun, ''Multi-task learning as multi-objective optimization,'' in *Proc. Adv. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2018, pp. 527–538.

[24] E. Arazo, P. Albert, N. E. O'Connor, K. McGuinness, and D. Ortego, ''Unsupervised label noise modeling and loss correction,'' in *Proc. Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, USA, ,2019, pp. 1–17.

[25] B. Wang, G. Qi, S. Tang, L. Zhang, L. Deng, and Y. Zhang, ''Automated pulmonary nodule detection: High sensitivity with few candidates,'' in *Medical Image Computing and Computer Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 11071, A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-030-00934-2_84.

[26] B. N. Narayanan, R. C. Hardie, T. M. Kebede, and M. J. Sprague, ''Optimized feature selection-based clustering approach for computer-aided detection of lung nodules in different modalities,'' *Pattern Anal. Appl.*, vol. 22, no. 2, pp. 559–571, May 2019, doi: 10.1007/s10044-017-0653-4.

[27] B. N. Narayanan, R. C. Hardie, and T. M. Kebede, ''Performance analysis of feature selection techniques for support vector machine and its application for lung nodule detection,'' in *Proc. NAECON-IEEE Nat. Aerosp. Electron. Conf.*, Jul. 2018, pp. 262–266, doi: 10.1109/NAECON.2018.8556669.

[28] A. Sóñora-Mengan, P. Gonidakis, B. Jansen, J. Garcia-Naranjo, and J. Vandemeulebroucke, ''Evaluating several ways to combine handcrafted features-based system with a deep learning system using the LUNA16 Challenge framework,'' *Proc. SPIE*, vol. 11314, Mar. 2020, Art. no. 113143T, doi: 10.1117/12.2549778.

[29] B. N. Narayanan, R. C. Hardie, and T. M. Kebede, ''Performance analysis of a computer-aided detection system for lung nodules in CT at different slice thicknesses,'' *Proc. SPIE*, vol. 5, Feb. 2018, Art. no. 014504, doi: 10.1117/1.JMI.5.1.014504.

[30] B. van Ginneken *et al.*, ''Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study,'' *Med. Image Anal.*, vol. 14, no. 6, pp. 707–722, Dec. 2010, doi: 10.1016/j.media.2010.05.005.

[31] E. Lopez Torres, E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. E. Fantacci, and P. Cerello, "Large scale validation of the M5L lung CAD on heterogeneous CT datasets," *Med. Phys.*, vol. 42, no. 4, pp. 1477–1489, Mar. 2015, doi: 10.1118/1.4907970.

[32] 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, doi: 10.1109/TMI.2016.2536809.

[33] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 559–567.

[34] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3484–3495, Nov. 2019, doi: 10.1109/TNNLS.2019.2892409.

[35] N. Khosravan and U. Bagci, "S4ND: Single-shot single-scale lung nodule detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer 2018, pp. 794–802.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland: Springer, 2015, pp. 234–241.

[37] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," 2016, *arXiv:1606.04797*. [Online]. Available: http://arxiv.org/abs/1606.04797

[38] L. Srensen, S. B. Shaker, and M. D. Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, Feb. 2010, doi: 10.1109/TMI.2009.2038575.

[39] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, and J.-Z. Cheng, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images," *IEEE Trans. Med. Imag.*, vol. 36, no. 3, pp. 802–814, Mar. 2017, doi: 10.1109/TMI.2016.2629462.

[40] Y. Xie, J. Zhang, Y. Xia, M. Fulham, 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, doi: 10.1016/j.inffus.2017.10.005.

[41] X. Zhao, L. Liu, S. Qi, Y. Teng, J. Li, and W. Qian, "Agile convolutional neural network for pulmonary nodule classification using CT images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 4, pp. 585–595, Apr. 2018, doi: 10.1007/s11548-017-1696-0.

[42] S. Hussein, R. Gillies, K. Cao, Q. Song, and U. Bagci, "TumorNet: Lung nodule characterization using multi-view convolutional neural network with Gaussian process," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 1007–1010, doi: 10.1109/ISBI.2017.7950686.

[43] C. Zhang *et al.*, "Toward an expert level of lung cancer detection and classification using a deep convolutional neural network," *Oncologist*, vol. 24, no. 9, pp. 1159–1165, Sep. 2019, doi: 10.1634/theoncologist.2018-0908.

[44] Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2482–2493, Jul. 2020, doi: 10.1109/TMI.2020.2972964.

[45] B. Frenay and M. Verleysen, "Classification in the presence of label noise: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 845–869, May 2014, doi: 10.1109/TNNLS.2013.2292894.

[46] N. Manwani and P. S. Sastry, "Noise tolerance under risk minimization," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 1146–1151, Jun. 2013, doi: 10.1109/TSMCB.2012.2223460.

[47] W. Zhang, R. Rekaya, and K. Bertrand, "A method for predicting disease subtypes in presence of misclassification among training samples using gene expression: Application to human breast cancer," *Bioinformatics*, vol. 22, no. 3, pp. 317–325, Feb. 2006, doi: 10.1093/bioinformatics/bti738.

[48] A. Malossini, E. Blanzieri, and R. T. Ng, "Detecting potential labeling errors in microarrays by data perturbation," *Bioinformatics*, vol. 22, no. 17, pp. 2114–2121, Sep. 2006, doi: 10.1093/bioinformatics/btl346.

[49] F. Muhlenbach, S. Lallich, and D. A. Zighed, "Identifying and handling mislabelled instances," *J. Intell. Inf. Syst.*, vol. 22, no. 1, pp. 89–109, 2004, doi: 10.1023/A:1025832930864.

[50] W. An and M. Liang, "Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises," *Neurocomputing*, vol. 110, pp. 101–110, Jun. 2013, doi: 10.1016/j.neucom.2012.11.023.

[51] S. Yu, B. Krishnapuram, R. Rosales, and R. B. Rao, "Bayesian co-training," *J. Mach. Learn. Res.*, vol. 12, no. 1, pp. 2649–2680, Sep. 2011.

[52] G. Patrini, A. Rozza, A. K. Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: A loss correction approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2233–2241, doi: 10.1109/CVPR.2017.240.

[53] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "A combination of shape and texture features for classification of pulmonary nodules in lung CT images," *J. Digit. Imag.*, vol. 29, no. 4, pp. 466–475, Aug. 2016, doi: 10.1007/s10278-015-9857-6.

[54] K. Wu, E. Otoo, and A. Shoshani, "Optimizing connected component labeling algorithms," *Proc. SPIE*, vol. 5747, pp. 1965–1976, Apr. 2005, doi: 10.1117/12.596105.

[55] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*. [Online]. Available: http://arxiv.org/abs/1607.08022

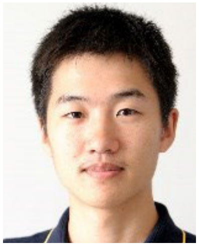
[56] A. Neubeck and L. V. Gool, "Efficient non-maximum suppression," presented at the Int. Conf. Pattern Recognit., 2006.

[57] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, "Brain MRI super resolution using 3D deep densely connected neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 739–742, doi: 10.1109/ISBI.2018.8363679.

[58] H. Yu, D. Liu, H. Shi, H. Yu, Z. Wang, X. Wang, B. Cross, M. Bramler, and T. S. Huang, "Computed tomography super-resolution using convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3944–3948, doi: 10.1109/ICIP.2017.8297022.

[59] C. You, W. Cong, M. W. Vannier, P. K. Saha, E. A. Hoffman, G. Wang, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, and Z. Zhang, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 188–203, Jan. 2020, doi: 10.1109/TMI.2019.2922960.

[60] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 91–99.



LIU CHENYANG received the B.Eng. degree from the Harbin Institute of Technology, in 2016. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong. His research interests include computer vision, machine learning, and image retrieval.

SHING-CHOW CHAN (Member, IEEE) received the B.Sc. (Eng.) and Ph.D. degrees in electrical engineering from The University of Hong Kong, in 1986 and 1992, respectively. Since 1994, he has been with the Department of Electrical and Electronic Engineering, The University of Hong Kong, where he is currently a Professor and an Associate Head of the Department. His research interests include fast transform algorithms, filter design and realization, multirate and biomedical signal processing, communications and array signal processing, high-speed A/D converter architecture, bioinformatics, smart grid, and image-based rendering. He is also a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society and an Associate Editor of the Journal of Signal Processing Systems. He was the Chair of the IEEE Hong Kong Chapter of Signal Processing from 2000 to 2002, and an Organizing Committee Member of the 2003 IEEE ICASSP and 2010 IEEE ICIP. He has served as an Associate Editor for *Digital Signal Processing*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II from 2012 to 2016, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I from 2008 to 2009.