Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8862353, 9 pages
https://doi.org/10.1155/2020/8862353

Research Article

URDNet: A Unified Regression Network for GGO Detection in Lung CT Images

Weihua Liu, Yuchen Ren, and Huiyu Li

Beijing Lab of Intelligent Information, School of Computer Science, Beijing Institute of Technology, Beijing, China

Correspondence should be addressed to Weihua Liu; liuweihua@bit.edu.cn

Received 30 July 2020; Revised 19 August 2020; Accepted 3 September 2020; Published 17 October 2020

Academic Editor: Chao-Yang Lee

We present a 3D deep neural network known as URDNet for detecting ground-glass opacity (GGO) nodules in 3D CT images. Prior work on GGO detection repurposes classifiers on a large number of windows to perform detection or fine-tuning by box regression based on a previous window classification step. Instead, we consider GGO detection as a multitarget regression problem to focus on the location of GGO. Furthermore, to capture multiscale information, we introduce a backbone network which is a contracting-expanding structure similar to 2D U-net, but we inject the source CT inputs into each layer in the contracting pathway to prevent source information loss at different scales. At last, we propose a two-stage training method for URDNet. In the first stage, the backbone of the network for feature extraction is trained, and in the second, the overall URDNet is fine-tuned based on the previous pretrained weights. By using this training method in conjunction with data augmentation and hard negative mining techniques, our URDNet can be effectively trained even on a small amount of annotated CT images. We evaluate the proposed method on the LIDC-IDRI dataset. It achieves the sensitivity of 90.8% with only 1 false positive per scan. Experimental results show that our detection method achieves the superior detection performance over the state-of-the-art methods. Due to its simplicity and effective, URDNet can be easier to apply to medical IoT systems for improving the efficiency of overall health systems.

1. Introduction

Lung cancer is currently a leading cause of cancer death worldwide and is responsible for more than 1.3 million deaths annually [1]. Detection and treatment of lung cancer at an early stage can improve the survival rate. GGO is a highly important CT imaging sign for detection of lung cancer at an early stage [2], which is defined as increased attenuation of the lung parenchyma without obscuration of the pulmonary vascular markings on the CT images [3]. Recently, the new coronavirus COVID-19 pandemic is prevalent, and its main symptoms are also related to GGO. However, due to their indistinct boundaries and no clear rules for brightness and shape, GGO nodules are easily overlooked, even by experienced radiologists. A promising solution to this problem is the use of computer-aided detection techniques.

The traditional architecture for computer-aided GGO detection typically consists of two stages: GGO candidate detection and false-positive reduction [4]. A small number of papers have been published on this topic. Bastawrous et al. [5] applied a Gabor filter to choose candidates and used an ANN to reduce false positives. Kim et al. [6] extracted tentative regions using binarization and classified the GGO nodules with a linear discriminant function. Jacobs et al. [7] first used intensity, shape, and context features to describe the appearance of candidates and subsequently applied a linear discriminant classifier and a gentle boost classifier to classify candidate regions. Although the conventional methods have yielded promising results, they still suffer from the low sensitivity and poor generalization, especially for notably small GGO nodules.

In recent years, nodule detection based on deep neural networks has achieved state-of-the-art detection performance. For example, Ginneken et al. [8] presented promising results for the extraction of nodule features using an off-the-shelf convolutional neural network (CNN) that was pretrained for a natural image classification task. Setio et al. [9] used multiple CNNs to extract discriminative features from the candidates,

and these features were used to classify candidates as nodules or background. Superior performance was achieved in the false-positive reduction track. Roth et al. [10] proposed an effective 2.5D representation for lymph node detection to exploit the 3D information of nodules when training a deep network by taking slices of the CT images from a point of interest in 3 orthogonal views. The slices were subsequently combined into a 3-channel image as the network input. Han et al. [11] proposed hybrid resampling in multi-CNN models for 3D GGO nodules to cover a large range scale, which reduced the risk of missing small or large GGO nodules. In general, these methods rely on classification or a combination of classification and regression for detection. These types of methods usually do not pay enough attention to the location problem of the detection and often produce missed detection and inaccurate locations. In addition, various types of neural works have been applied in various applications, e.g., graph neural network for creative works [12], LVQ neural network for traffic prediction [13], generative adversarial networks (GANs) for style transfer [14], and 3D GANs for the simulation of creative stage scene [15]. The most important is that GGO detection requires a huge amount of computation in 3D CT and generally requires a more efficient detection method to meet actual needs in medical IoT.

In order to overcome the above limitations, we propose a unified regression deep neural network for GGO detection. We consider GGO detection as a multitarget regression problem, straight from 3D CT to bounding box coordinates. To acquire the discrimination information between the object and the background, the same pseudotarget (zero) is set for all the negative samples, so the pseudotarget also is denoted as zero target in our paper. Compared with the classification-based method, the learning goal of our whole detection is just a unified object location regression, which can guide the network to learn better object location information. Therefore, more attention can be paid to the localization problem in the detection by using a unified regression objective function in our approach. And a multi-input and multioutput backbone convolutional network is also applied in our approach to make it more representative. Furthermore, we design a two-stage transfer training method to train our URDNet on small annotated GGO data. To evaluate the effectiveness of our proposed URDNet for GGO detection, we conduct GGO detection experiments on the LIDC-IDRI [16], the currently largest publicly available and mostly often used database of lung nodules. The experiment results show that the network is effective and accurate.

Our main contributions are summarized as follows:

(1) We present an end-to-end deep convolutional neural network which is unified and only regression for GGO detection in 3D CT scans which leads to outstanding performance of GGO detection on LIDC-IDRI. The resultant detection sensitivity is 90.8% at 1 false positive per scan

(2) We introduce a multi-input and multioutput structure for our network's backbone. The backbone not only reserves the subtle locations but also represents the discriminate information of GGO nodules

(3) We propose a unified regression objective function for all samples. For positive samples, the position prediction is regarded as only a conventional regression task. For negative samples, zero target is set. The location of negative samples (boxes) will be regressed to a pseudotarget (zero).

(4) We adopt a two-stage training method to train the complicated 3D detection network given a small amount of annotated samples

The remainder of this paper is organized as follows. Sections 2 and 3 presents our URDNet architecture and its training method, respectively. The implementation details and experimental results are discussed in Section 4. We conclude in Section 5.

2. URDNet Architecture

The network architecture is illustrated in Figure 1 and is composed of a backbone network for feature extraction, which is a multiscale input-output structure, and a detection head which is a single prediction module by location regression, which directly generates a fixed set of 3D bounding boxes. Due to the memory limitation of GPU, the input of the method is only a CT cube with a fixed size ($128 \times 128 \times 128$ in our experiments). The final detection result is the combination of all the detected GGOs in each cube. Figure 1 illustrates the architecture of our network, which takes as input a $128 \times 128 \times 128$ CT cube. The details of CT preprocessing are introduced here. For a 3D CT scan, a voxel size normalization is firstly performed due to various voxel sizes of subjects. The voxel size is set to $1 \times 1 \times 1\,\mathrm{mm}^3$ by using bilinear interpolation (after voxel normalization, even the number of slices in axial dimension is greater than 128 for all the data in our experiments). Then, we divide a 3D CT volume into several $128 \times 128 \times 128$ cubes in a slider-patch way. For each cube, it will be inputted into our URDNet and be processed. More details of URDNet are given in the following subsections.

2.1. Backbone Network. The backbone network includes two main symmetric pathways: a contracting pathway and an expanding pathway. The contracting pathway follows the typical architecture of a convolutional network. To aid the network in capturing spatial information at different scales, we construct a multiscale input structure by downsampling the source input and feeding it into each layer in the contracting pathway, not only the first layer. The feed operation is indicated by the green lines in Figure 1. We referred to these lines connecting the source input CT to the layers in the contracting pathway as source connections. The contracting pathway can be divided into five blocks, in which the output feature maps are downsampled by 2, 4, 8, 16, and 32 w.r.t. the input cube size. This block is a composite function of four consecutive operations: 3D convolution (Conv), batch normalization (BN) [17], rectified linear units (ReLU), and pooling (Avg: average or Max: maximum).

The expanding pathway is an information-expanding process that elucidates higher resolution features via an upsampling strategy. The upsampling procedure is

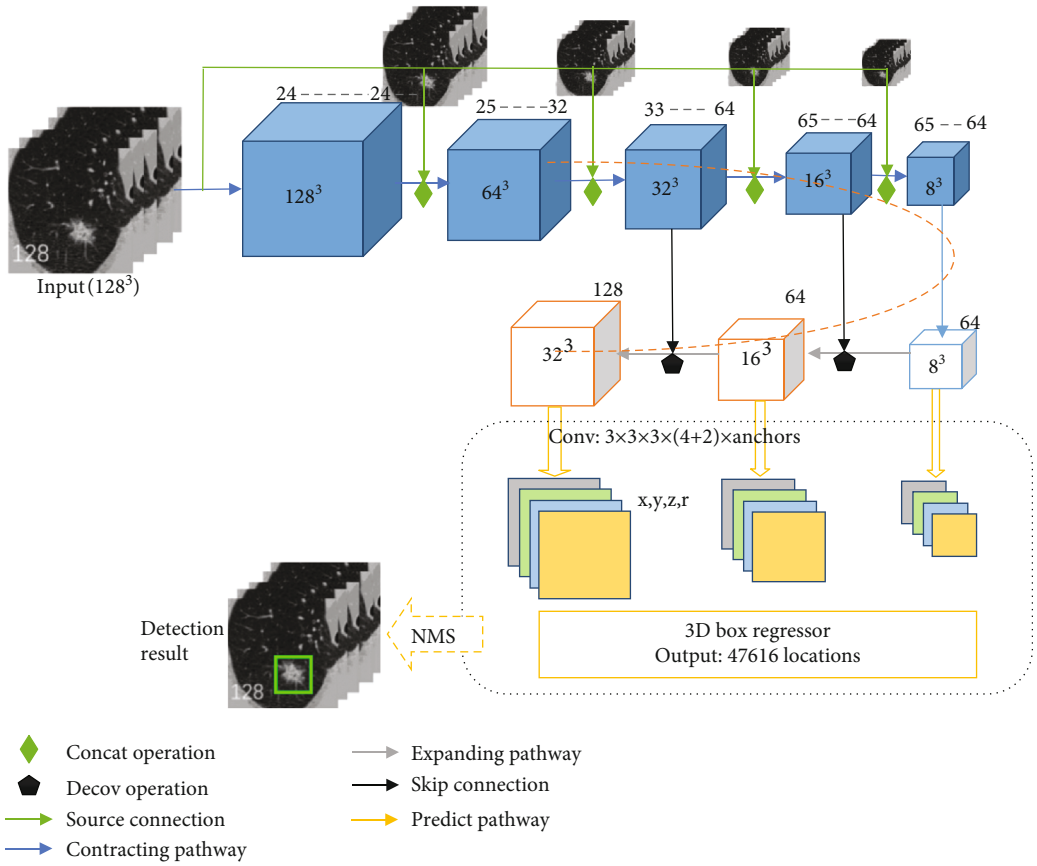


FIGURE 1: The URDNet architecture, which mainly comprises two components, the backbone network for feature extraction and the detection head. The backbone network is a multi-input and multioutput structure. The detection head merely contains a $3 \times 3 \times 3$ convolutional layer, followed by only one prediction layer, for regression the center location (x, y, z) and the corresponding diameter r of the nodule. (the digits on the cube denote the corresponding numbers of channels.)

implemented by a series of layers including unpooling, deconvolution, BN, and ReLU operations to perform a complicated deconvolution, as described in a previous paper [18]. The expanding pathway is semantically stronger because of the feature map from higher levels (the top of the contracting pathway). In contrast, the contracting feature map consists of lower-level semantics, but its activations are more accurately localized because it was subsampled fewer times. To preserve the localized information, the feature map of the expanding pathway is enhanced with the feature map from the contracting pathway via skip connections. Skip connections associate low-level feature maps across resolutions and semantic levels. Moreover, to create a multiscale feature map that has strong semantics and precise spatial information at all scales, we combine low-resolution and semantically strong features with high-resolution and semantically weak features via a top-down pathway and skip connections. Skip connections [19] are connections that can skip one or more layers. Similar architectures adopting a top-down pathway and skip connections are popular in recent research [20–22]. However, only a single high-level feature map of fine resolution was applied for prediction in previous networks. In contrast, our backbone leverages multiscale feature maps in which predictions are independently generated on each level for GGO detection.

2.2. Detection Head. In the traditional sliding-window detection methods, the entire detection space is eventually discretized into a series of windows. Our network also discretizes the output space of bounding boxes into a set of anchor boxes with different scales over multiple feature maps. Each anchor box is a predefined box centered at a location of the feature map and is associated with a special initial scale, similar to the anchor box used in Faster R-CNN [23].

Our regression prediction head predicts bounding boxes based on a fixed set of anchor boxes and is implemented by regressing 3D box relative offsets from anchor boxes to satisfiable boxes (to better match the GGO shape) using small convolutional filters applied to multiple feature maps. These processes are indicated on the yellow lines and a yellow rectangular box in the bottom area of Figure 1. According to the previous section, the feature map of the expanding pathway progressively increases in size. The multiscale feature maps are composed of multiple feature maps of different resolutions, and each feature map can produce a set of detection predictions. To detect GGO of various sizes, the prediction pathway of our network can naturally combine predictions from the multiscale feature maps. Additionally, at each location of the feature map, we simultaneously predict multiple boxes with different scales but the same center. The multiple boxes are parameterized relative to the corresponding anchor

boxes. For example, the left block shown in Figure 2 is used to regress to the resultant boxes. Since we consider 3 sizes of anchors and each resultant result is a 4D vector, the outcome of this block is $12 = 3 \times 4$ vectors for each anchor. The number of anchors for each feature map location must be carefully set to cover a wider and finer range of scale, and more details are provided in Section 4.1.

After the resultant boxes are obtained from the prediction pathway, we perform nonmaximum suppression (NMS) to rule out the overlapping boxes. For each anchor, the prediction finally produces the four-position component map. If the area is background, the corresponding location of the component map is very close to zero. Only the box with non-background will be retained and others will be deleted. Then, the retained boxes will be decided as GGOs if their position is larger than a threshold or otherwise as non-GGO (background). Such threshold will be set up in the applications.

3. Training Method

Training of URDNet is a multitarget regression procedure because it simultaneously regresses the GGO center and the diameter of GGO. The details of the objective of our network is given in the following subsection. Besides, only a small amount of CT data with annotated GGO nodules is given, and the overall network is difficult to converge. We adopt a two-stage transfer training strategy to solve the problem. More useful strategies also are given in the below subsections.

3.1. Loss Function. In our method, a true object can be expressed by a box. Each box corresponds to a 3D square and thus is represented by a 4D vector (x, y, z, r), where (x, y, z) is the 3D center point and r is the side length of the square. We also use a box to express the position of any background, and the components of this box (x, y, z, r) are set to zero. We define the position of this background as zero target (pseudo-target). The target position of true objects is defined as real target. In this way, both positive samples (target window) and negative samples (background window) can be used as position targets. We can express the detection problem as the same learning target and only need position regression.

The conventional regression loss for positive samples and a new design regression loss for negative samples are included in the overall localization loss.

The regression loss for positives is a modified smooth L_1 loss [24] between the predicted 3D bounding box (denoted as l) and the ground-truth 3D bounding box (denoted as g). Similar to Faster R-CNN [12], it can be regressed to the offset terms for the center (xc, yc, and zc) of the anchor 3D bounding box (denoted as d) and its radius (denoted as r).

$$L_{\text{pos}}(l, g) = \sum \left(\beta \sum_{k \in \{xc,yc,zc\}} \text{smooth}_{L_1} \left(l_i^k - \hat{g}_j^k \right) + (1 - \beta) \text{smooth}_{L_1} \left(l_i^r - \hat{g}_j^r \right) \right), \tag{1}$$

where $\hat{g}_j^{xc} = (g_j^{xc} - d_i^{xc})/d_i^r$, $\hat{g}_j^{yc} = (g_j^{yc} - d_i^{yc})/d_i^r$, $\hat{g}_j^{zc} = (g_j^{zc} - d_i^{zc})/d_i^r$, $\hat{g}_j^r = \log(g_j^r/d_i^r)$, and the weight term β is set to 0.6 through careful experiments in this paper, which means that we focus additional attention on the center point.

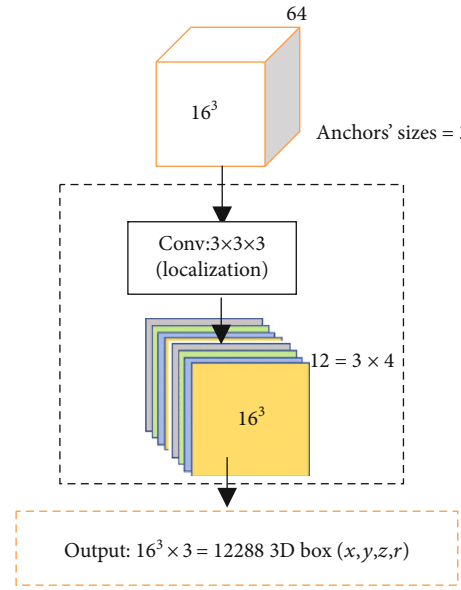


Figure 2: The illustration of the regression in a prediction pathway for one scale of feature map, which is the detail of "3D box regressor" shown in Figure 1.

At the same time, because of the characteristics of zero target in our negative samples, we design a loss function which regresses to a zero target (denoted as o),

$$L_{\text{neg}}(l, o) = \sum \log \left(\sum_{k \in \{xc,yc,zc,r\}} \|l_k - o\| \right), \tag{2}$$

which means that several position components can be cohered to zero (zero target).

To sum up, the full optimization objective is

$$L_{\text{all}} = L_{\text{pos}} + \lambda_{\text{n}} L_{\text{neg}}, \tag{3}$$

where λ_{n} is the weight for the regression loss for the negative samples and is set to 0.5 in this paper.

3.2. Two-Stage Training. We add a classifier module with a 3D Avg-pooling $(4 \times 4 \times 4)$ layer and a two-class softmax layer behind the backbone of our network to construct a solo GGO classifier. The $64 \times 64 \times 64$ positive cubes are cropped from the lung scans such that they contain only one GGO nodule. More positive cubes are generated by data augmentation. The $64 \times 64 \times 64$ negative cubes without nodules are randomly cropped. It is easier to construct a relatively large-scale dataset for training the nodule classifier than the detector network. Moreover, the solo classifier can be trained much faster than our network because the input size is 1/8 of the URDNet input size, and it is only a binary classification problem. To avoid slow convergence, the weights are initialized with Glorot and Bengio [25] initialization. The classifier

network is trained by minimizing the softmax-loss function with the stochastic gradient descent (SGD).

After training the classifier network, we truncate the weights of the classifier network to initialize our backbone and train the URDNet by fine-tuning with a smaller learning rate. We find that the URDNet loss converges quicker than training from scratch. Training our network in this way can effectively alleviate the overfitting problem.

3.3. Data Augmentation. Considering that only a few training samples are available, we tested certain data augmentation operations. We exploit both lossless and lossy augmentation and find the following two simple but useful augmentations for GGO nodule patches:

(1) Flipping. The patches are randomly flipped with respect to coronal, sagittal, and axial dimensions

(2) Resizing. The patches are randomly resized with a ratio between 0.9 and 1.1

We also tested other augmentation methods, such as axis swapping and rotation. However, it has no significant effects. Furthermore, to boost nodule classification performance, we adopt an online hard negative mining technique, similar to the online hard example mining (OHEM) method proposed in a previous paper [26]. We use online hard negative mining in the training procedure. First, we process the input patches using the network to conduct a forward pass and obtain tens of thousands of detections, each of which is associated with a box with a classification confidence score. Second, the acceptable N negative samples are randomly selected from these boxes as a candidate pool. Third, the negative samples in this pool are sorted in a descending order based on their overall loss, and the top K samples are selected as the hard negatives.

4. Experiments

In this section, we first describe the experimental setup. We introduce our dataset's details. The use of anchor boxes improves the speed and efficiency for the detection portion of a deep learning neural network. Due to its importance, we then give more details about anchors. At last, we give the results of our two groups of experiments. One group is used to evaluate our URDNet, and the other group is used to compare the performance with that of other systems.

4.1. Experimental Setup

4.1.1. Dataset. The dataset is collected from the largest publicly available reference database for lung nodules, the LIDC-IDRI [16]. The LIDC-IDRI database contains a total of 1018 CT scans. Each LIDC-IDRI scan was annotated by experienced thoracic radiologists using a two-phase reading process. Four radiologists annotated scans and marked all suspicious lesions as nodule ≥ 3 mm, nodule < 3 mm, or non-nodule. We only considered the GGO nodules. Simultaneously, we only considered the nodules annotated by the majority of the radiologists (at least 3 out of 4 radiologists) in our standard reference. 302 CT scans from 299 patients were collected from LIDC-IDRI. In these CT scans, 635 GGO nodules (271 nodules annotated by 3 radiologists and 364 nodules annotated by 4 radiologists) were found. The diameter of the nodules varied from 3 to 34 mm with a median of 10.3 mm. We use 250 CT scans as a training set and the remaining 52 as a test set. A slice from one patient with a single nodule location is shown in Figure 3.

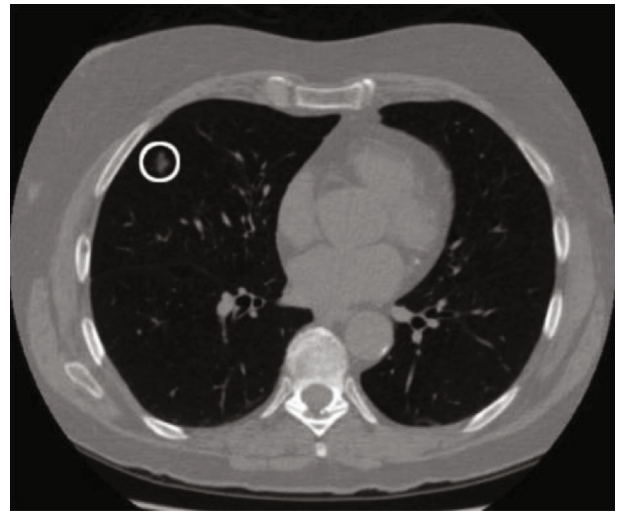


Figure 3: A slice from LIDC-IDRI with a nodule annotated.

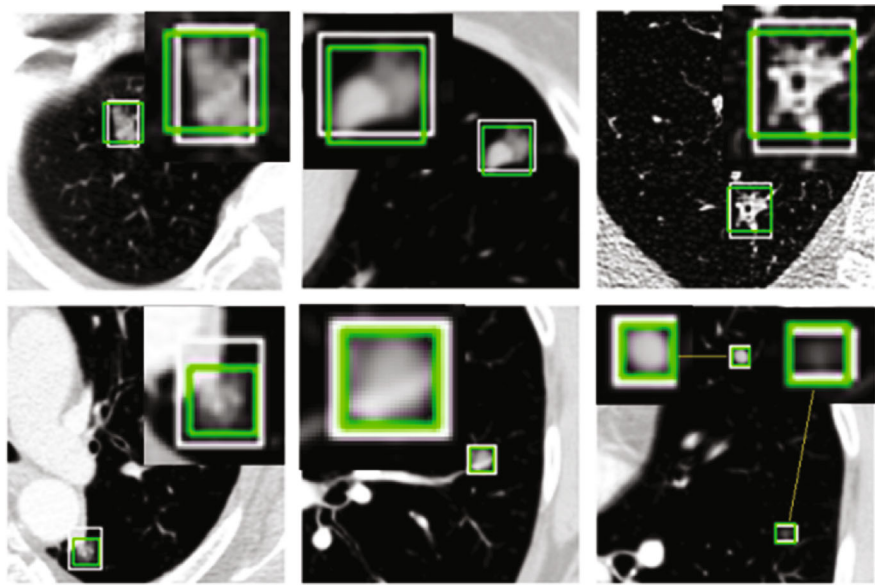
Table 1: Configuration of anchors' sizes for each scale of decoded feature maps.

Feature map	No. of anchor size	No. of anchor box	Size (mm)				
$32 \times 32 \times 32$	1	32768	4	\	\	\	\
$16 \times 16 \times 16$	3	12288	6	8	10	\	\
$8 \times 8 \times 8$	5	2560	12	16	20	26	32
Sum		47616					

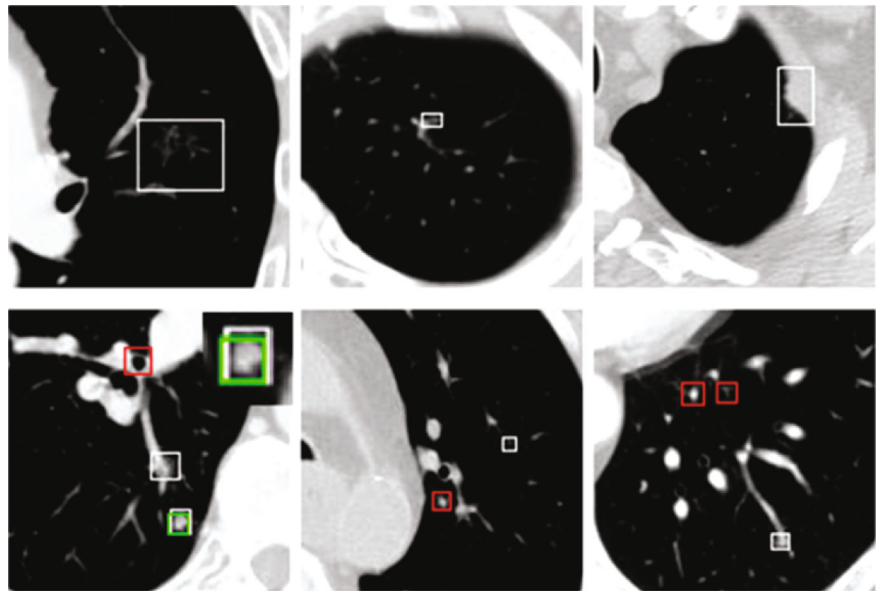
Table 2: GGO detection results using SSD, U-net, and our URDNet.

Methods	Sensitivity	FPs/scan
SSD	78.2%	9.1
U-net	83.7%	8.9
Our URDNet	93.5%	6.8

4.1.2. Anchor Configuration. All the scales of information decoded in the expanding pathway of our URDNet, including $32 \times 32 \times 32$, $16 \times 16 \times 16$, and $8 \times 8 \times 8$, are used and inputted into the detection head. The numbers of anchors' sizes are set to be 1, 3, and 5 for these three scales, respectively, which lead to 9 anchors' sizes in total: 4, 6, 8, 10, 12, 16, 20, 26, and 32 mm. The details are listed in Table 1. To sum up, we will generate 47616 ($= 32^3 \times 1 + 16^3 \times 3 + 8^3 \times 5$) anchors for each CT cube, based on which 47616 resultant boxes will be computed. We find that the detection scale range is between 3 mm and 33 mm, which is larger than the scale range of the anchor box, in our test set. Benefitting from the



(a) Correct detection samples



(b) False detection samples

Figure 4: Examples of the detection by URDNet (the white rectangles denote the ground-truth boxes, the green rectangles denote the detection results, and they are zoomed at the top-right area or the left-bottom area, and the red rectangles denote the wrong results).

multiscale processing, our network can adapt to the changing of nodules' sizes.

4.2. Experimental Results. In our experiments, we use a free-response receiver operating characteristic (FROC) analysis [27] on the filtered GGO dataset from LIDC-IDRI [16] for comparison. In the FROC curve, the sensitivity is plotted as a function of the average number of false positives per scan (FPs/scan). In this work, the sensitivity is defined as the fraction of detected true positives divided by the number of ground-truth GGO nodules. The FROC overall score is defined as the average of the sensitivity at seven predefined false-positive rates: 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs per scan. This performance metric was introduced into the ANODE09 challenge and referred to as the competition performance metric (CPM) in a previous paper [28].

We first conduct three experiments using a different backbone or head detection to evaluate the effect of our network. We use the same training dataset and data augmentation strategy. Other hyperparameters of the training network are also shared, except for specified changes to components. The SSD method predicts bounding boxes and confidence scores based on a fixed set of anchor boxes, which directly related to our URDNet's head detection. In contrast, U-net has an elegant decoder-encoder structure, but it only uses the last feature map for biomedical applications and ignores the multi-input and multioutput structure. Table 2 lists the results in the GGO detection task.

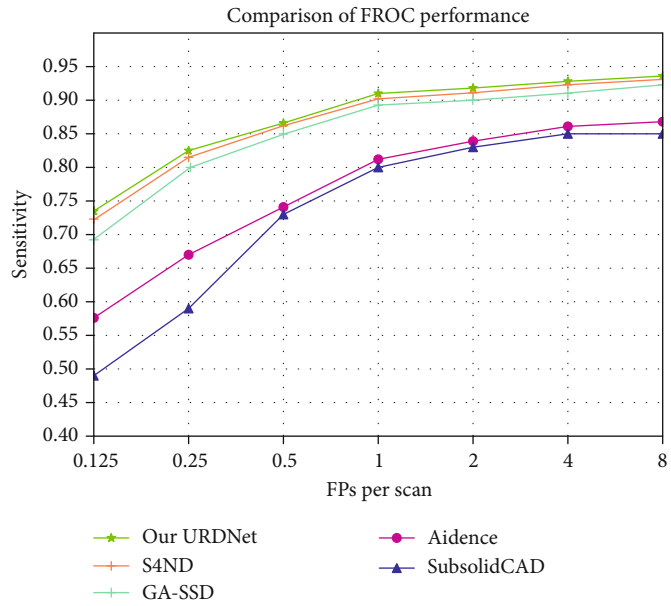


Figure 5: Comparison of performance (our URDNet with four counterparts).

It is obvious that our URDNet achieves the highest sensitivity (93.5%) at the lowest FPs/scan (6.8) among these detection experiments, which demonstrates the superiority of our detection network. Certain examples that are correctly detected are illustrated in Figure 4(a). These results indicate that URDNet can accurately locate the centers of GGO nodules and regress the size of GGO nodules. Examples of GGO nodules false detections are also shown in Figure 4(b). Typically, these nodules are notably low in contrast (pure or diffuse) or located close to the other tissues (blood or chest wall) and can be considered notably low-quality GGO nodules. To further improve the detection performance, additional discriminative features must be learned by a new learning method.

Although a few methods have been developed for GGO nodule detection, it is trivial to compare all other methods. In this paper, we choose three GGO detection systems or methods from different categories for comparison. We first select the SubsolidCAD [29] system, which is a state-of-the-art conventional method that uses 4 categories of hand-crafted features to describe the appearance or the internal characteristic of the GGO. The system can reach a sensitivity of 80% at an average of 1.0 false positives per scan with a CPM [28] of 0.734. We also compare our performance with the Aidence [30] system based on convolutional networks, which is the strongest competitor and the top performer in the LUNA16 Challenge. Referring to the report [30], the best score was achieved by Aidence with a CPM of 0.764 in the GGO nodule candidate detection task. The Aidence system uses end-to-end convolutional networks that are trained on a subset of studies from the National Lung Screening Trial [31]. Last, we chose the S4ND [32] method and GA-SSD [33] method for lung nodule detection to compare which are current state-of-the-art methods. The S4ND method is a single-shot and single-scale method, while GA-SSD is an improved method based on SSD by implementing the attention mechanism. Additionally, we list a comparison of performance among our URDNet, SubsolidCAD, Aidence, S4ND, and GA-SSD in Figure 5. We observe that our URDNet attained superior performance. The CPM can reach 0.874, surpassing the SubsolidCAD system (CPM: 0.734) with relative performance gains of 19.07%, the Aidence system (CPM: 0.767) with 13.95%, S4ND (CPM: 0.866) with 0.92%, and GA-SSD (CPM: 0.855) with 2.26%, respectively.

5. Conclusions

In this paper, we present a 3D convolutional detector network known as URDNet that was constructed of a multi-scale input-output U-shaped network for GGO detection in CT images. A unified regression objective function is proposed in URDNet in which the location of an object is focused on during learning that can directly regress a fixed set of 3D boxes for all samples. Furthermore, a two-stage training method is designed to help our complicated 3D detector network converge and prevent overfitting, even if given a small amount of annotated GGO nodule CT images. By this training method incorporated with data augmentations and hard negative mining, our network can be efficiently and effectively trained in an end-to-end manner for GGO detection.

We believe that URDNet offers a useful tool for GGO nodule detection in the clinical diagnosis of lung cancer. Moreover, our independent GGO detection algorithm can be easily integrated into the existing lung nodule CAD systems to boost the overall system performance. Immediate future work will extend our network for the detection of other nodules. Because this method does not require large amounts of labeled data and has a simple, unified regression objective, it is easier to be applied to other nodule processing tasks in medical IoT systems.

Data Availability

The data used to support the findings of this study are from previously reported LIDC-IDRI dataset, which is the currently largest publicly available database. The data are available at relevant places with text as reference.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 60973059, 81171407, and 61901533) and Program for New Century Excellent Talents in University of China (grant number NCET-10-0044).

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA: A Cancer Journal for Clinicians*, vol. 65, no. 1, pp. 5–29, 2015.

[2] C. I. Henschke, D. F. Yankelevitz, R. Mirtcheva, G. McGuinness, D. McCauley, and O. S. Miettinen, "CT screening for lung Cancer," *AJR American Journal of Roentgenology*, vol. 178, no. 5, pp. 1053–1057, 2002.

[3] W. T. Miller Jr. and R. M. Shah, "Isolated diffuse ground-glass opacity in thoracic CT: causes and clinical presentations," *AJR American Journal of Roentgenology*, vol. 184, no. 2, pp. 613–622, 2005.

[4] L. Linying, L. Xiabi, Z. Chunwu, Z. Xinming, and Z. Yanfeng, "A review of ground glass opacity detection methods in lung CT images," *Current Medical Imaging Reviews*, vol. 13, no. 1, pp. 20–31, 2017.

[5] H. A. Bastawrous, T. Fukumoto, M. Tsudagawa, and N. Nitta, "Detection of ground glass opacities in lung CT images using Gabor filters and neural networks," in *2005 IEEE Instrumentationand Measurement Technology Conference Proceedings*, pp. 251–256, Ottawa, Ont., Canada, May 2005.

[6] H. Kim, T. Nakashima, Y. Itai, S. Maeda, J. K. Tan, and S. Ishikawa, "Automatic detection of ground glass opacity from the thoracic MDCT images by using density features," in *2007 International Conference on Control, Automation and Systems*, pp. 1274–1277, Seoul, South Korea, October 2007.

[7] C. Jacobs, C. I. Sánchez, S. C. Saur, T. Twellmann, P. A. de Jong, and B. van Ginneken, "Computer-aided detection of ground glass nodules in thoracic CT images using shape, intensity and context features," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, pp. 207–214, Springer, 2011.

[8] B. Van Ginneken, A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *2015 IEEE 12th International symposium on biomedical imaging (ISBI)*, pp. 286–289, New York, NY, USA, April 2015.

[9] A. A. A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[10] H. R. Roth, L. Lu, A. Seff et al., "A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, pp. 520–527, Springer, 2014.

[11] G. Han, X. Liu, G. Zheng, M. Wang, and S. Huang, "Automatic recognition of 3D GGO CT imaging signs through the fusion of hybrid resampling and layer-wise fine-tuning CNNs," *Medical & Biological Engineering & Computing*, vol. 56, no. 12, pp. 2201–2212, 2018.

[12] F. Zhang, Y. Wang, and C. Wu, "An automatic generation method of cross-modal fuzzy creativity," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 5685–5696, 2020.

[13] F. Zhang, T. Y. Wu, Y. Wang et al., "Application of quantum genetic optimization of LVQ neural network in smart city traffic network prediction," *IEEE Access*, vol. 8, pp. 104555–104564, 2020.

[14] F. Zhang and C. Wang, "MSGAN: generative adversarial networks for image seasonal style transfer," *IEEE Access*, vol. 8, pp. 104830–104840, 2020.

[15] F. Zhang, G. Ding, Q. Lin, L. Xu, Z. Li, and L. Li, "Research of simulation of creative stage scene based on the 3DGans technology," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 6, pp. 1430–1443, 2018.

[16] S. G. Armato III, G. McLennan, L. Bidaut et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

[17] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, https://arxiv.org/abs/1502.03167.

[18] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," 2015, https://arxiv.org/abs/1511.00561.

[19] C. Bishop and C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Munich, Germany, 2015.

[21] S. Honari, J. Yosinski, P. Vincent, and C. Pal, "Recombinator networks: learning coarse-to-fine feature aggregation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5743–5752, Las Vegas, 2016.

[22] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Computer Vision – ECCV 2016*, pp. 483–499, 2016.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[24] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, 2015.

[25] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Journal of Machine Learning Research*, vol. 9, pp. 249–256, 2010.

[26] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in

Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 761–769, Las Vegas, 2016.

[27] H. L. Kundel, K. S. Berbaum, D. D. Dorfman, D. Gur, C. E. Metz, and R. G. Swensson, "Receiver operating characteristic analysis in medical imaging," *Journal of the Icru*, vol. 79, no. 8, p. 1, 2008.

[28] M. Niemeijer, M. Loog, M. D. Abràmoff, M. A. Viergever, M. Prokop, and B. van Ginneken, "On combining computer-aided detection systems," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 215–223, 2011.

[29] C. Jacobs, E. M. van Rikxoort, T. Twellmann et al., "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical Image Analysis*, vol. 18, no. 2, pp. 374–384, 2014.

[30] A. A. A. Setio, A. Traverso, T. de Bel et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.

[31] C. S. White, "National lung screening trial," *Journal of Thoracic Imaging*, vol. 26, no. 2, pp. 86-87, 2011.

[32] N. Khosravan and U. Bagci, "S4ND: single-shot single-scale lung nodule detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 794–802, Granada, Spain, 2018.

[33] J. Ma, X. Li, H. Li et al., "Group-attention single-shot detector (GA-SSD): finding pulmonary nodules in large-scale CT images," 2018, https://arxiv.org/abs/1812.07166.