Accurate and Efficient LIF-Nets for 3D Detection and Recognition

YUETING SHI, HAI LI, HEHUI ZHANG, ZHENZHI WU, AND SHIWEI REN

1School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
2School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
3Beijing Lynxi Technology Company Ltd., Beijing 100080, China

Corresponding author: Shiwei Ren (renshiwei@bit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801024, in part by the National Key Research and Development Program under Grant 2018AAA0102600, in part by the Key Scientific Technological Innovation Research Project by the Ministry of Education, and in part by the Institute for Guo Qiang, Tsinghua University.

ABSTRACT

3D object detection and recognition are crucial tasks for many spatiotemporal processing applications, such as computer-aided diagnosis and autonomous driving. Although prevalent 3D Convolution Nets (ConvNets) have continued to improve the accuracy and sensitivity, excessive computing resources are required. In this paper, we propose Leaky Integrate and Fire Networks (LIF-Nets) for 3D detection and recognition tasks. LIF-Nets have rich inter-frame sensing capability brought from membrane potentials, and low power event-driven mechanism, which make them excel in 3D processing and save computational cost at the same time. We also develop ResLIF Blocks to solve the degradation problem of deep LIF-Nets, and employ U-LIF structure to improve the feature representation capability. As a result, we carry out experiments on the LUng Nodule Analysis 2016 (LUNA16) public dataset for chest CT automated analysis and conclude that the LIF-Nets achieve 94.6% detection sensitivity at 8 False Positives per scan and 94.14% classification accuracy while the LIF-detection net reduces 65.45% multiplication operations, 65.12% addition operations, and 65.32% network parameters. The results show that LIF-Nets have extraordinary time-efficient and energy-saving performance while achieving comparable accuracy.

INDEX TERMS

Spiking neural network, leaky integrate and fire model, 3D detection, 3D recognition, pulmonary nodule screening.

I. INTRODUCTION

Advanced computer-aided diagnosis systems (CADs) using deep learning analysis to detect and recognize lung nodules have been conducted in recent years, which helps free radiologists from the time-consuming work and reduce inter-observer variability [1]. However, solving the automated CT analysis problems in the real world would certainly require a more sophisticated model with a vast number of parameters and result in a substantial amount of computation overhead and power consumption. This technical difficulty is related to the fact that nodules have variable sizes and shapes and have similar appearances to normal tissues. Thus, it is paramount to develop a methodology to capture the spatiotemporal features of CT slices both accurately and efficiently.

The current body of research on automated chest CT analysis mainly includes nodule detection [2]–[5] and nodule classification [6]–[9]. Nodule detection typically consists of two stages, namely region proposal generation and false positive reduction. Benefiting from the feature learning ability, convolutional neural networks (CNNs) have shown great performance over other traditional feature extraction methods and thus been widely used in the first stage for generating candidate bounding boxes [2], [3]. In the second stage, complex features identities are used to remove false positive nodules [4]. Furthermore, to extract better nodule-sensitive features from 3D CT images, state-of-the-art frameworks often utilize the 3D region proposal network (RPN) [10] for nodule screening [11]–[13], followed by a 3D classifier for false positive reduction [2], [14] or malignancy evaluation [15], [16]. Although the use of deep ConvNets has continued to improve the accuracy and sensitivity, some limitations exist.
Firstly, the computational cost of 3D convolution is one order of magnitude higher than that of the 2D both for training and inference purposes; the existing 3D CNN methods require much more sophisticated computers due to the memory constriction [15]. Secondly, the excessively large number of weights of the 3D ConvNets makes it difficult to train on public chest CT datasets of relatively small sizes [13] and prone to overfitting. Thirdly, with the popularity of low-dose CT screening and the development of high-resolution CT scanning technology, an increasing number of nodules are being identified, which poses many challenges for quick lesion screening. During the influenza epidemic period or coronavirus disease epidemic period, improving clinical diagnosis efficiency is one of the core needs.

To overcome these difficulties, the development of an efficient and rich-temporal-sensing-capability spatiotemporal processing method is much needed. Spiking Neural Networks (SNNs) are brain-inspired computing models that use spatiotemporal dynamics to mimic neural behaviors and communicate between units [17], [18]. Leaky Integrate and Fire (LIF) neuron model is one of the popular neuron models in SNNs [19], [20].

Lately, SNNs have shown a competitive performance compared with Artificial Neural Networks (ANNs) with the recent developments of improved training algorithms [21]–[23], and have been applied in Electroencephalogram (EEG) brain data [24], [25] and functional Magnetic Resonance Imaging (fMRI) for spatial-temporal cognitive processes [26], [27]. In computer vision tasks, SNNs have achieved outstanding performance in N-MNIST dataset and Cifar10-DVS dataset for image classification and dynamic visual recognition [28]–[30]. However, seldom work is related to object detection based on SNN because it involves both identifying multi-scale objects and calculating precise coordinates of bounding boxes. Solving the regression problem in object detection requires much higher accuracy than selecting the highest probability with the argmax function in image classification [31], [32]. Furthermore, increasing the depth of a network leads to decrease in performance on both test and training data (i.e. degradation problem) [33]. As a result, even comparatively less attention has been given to SNN-based 3D processing.

In this work, we address the above mentioned problems and propose accurate and efficient LIF-Nets for 3D object detection and recognition. ConvLIF Layer, ResLIF Block, and U-LIF structures are developed for improving the performance of our deep model. Furthermore, we evaluate the performance and efficiency of LIF-Nets and compare to that of 3D convolution by calculating the number of computations and the number of parameters.

Our contributions in this work are summarized as follows:

1) We propose the LIF-Nets for 3D object detection and recognition, which uses membrane potential to obtain the inter-frame information. LIF-Nets exploit the deep SNN successfully to the 3D volumetric detection and recognition tasks.

2) We develop ResLIF Blocks to solve the degradation problem of deep LIF-Nets, and develop the U-LIF structure to improve the feature representation capability, which creates opportunities for further LIFNet-based models to be applied in various and more completed applications.

3) We realize the dendritic integration by addition operations instead of multiplications (i.e. spikes can be simplified as 0 or 1 value), which significantly reduces the latency and saves the energy while achieving high accuracy.

The rest of the paper is organized as follows. Section II presents some closely related work. Section III describes the LIF theory and details LIF-Nets structures. Efficiency evaluation and experiment results are presented in Section IV. Section V is the discussion, and Section VI concludes the paper.

II. RELATED WORK

A. SNN

In recent years, SNN has received extensive attention and become a popular research topic as the third-generation artificial neural network due to the event-driven manner and low energy consumption nature.

SNN-based object detection is considered as a challenging task. Kheradpisheh et al. [32] used a temporal coding scheme and proposed a swallow 2D detection net (a temporal-coding layer followed by a cascade of consecutive convolutional and pooling layers) trained by spike timing dependent plasticity (STDP). They demonstrated the capability of SNN to recognize several natural objects even under severe variations. Kim et al. [31] used Deep Neural Network-to-SNN conversion methods and proposed a Spiking-YOLO model. They proposed the first deep SNN for 2D object detection and achieved 51.61% mAP on PASCAL VOC dataset. Doborjeh et al. [34] and Kasabov et al. [35] proposed 3D SNN models to cluster and classify the EEG signals. These researches are different from the digital image classification task in computer vision, but they all make good use of the spatiotemporal processing capability of SNN. Despite these efforts, the most prevalent methods have been limited to shallow SNNs and 2D cases.

B. NODULE DETECTION

Since CT sequences present human organs in 3D, 3D contexts play an important role in recognizing nodules. Liao et al. [15] proposed a 3D Faster R-CNN for nodule detection with 3D convolutional operations and a UNet-like encoder-decoder structure to learn latent information. Qin et al. [36] introduced the dense connection structure, which aimed to reuse nodules’ features and boost feature propagation. Zhu et al. [13] reported that the 3D ConvNet had too many parameters, which made it challenging to train on public chest CT datasets.
FIGURE 1. The neuron model in (a) ANNs and (b) SNNs, and (c) Diagram of LIF-based 3D CT detection. The red curve shows one of the increasing membrane potentials of the neurons screening the suspected nodule. The black curve shows one of the membrane potentials of the neurons screening the background regions. As the features of CT slices are gradually extracted, the membrane potential representing the lesion site increases and exceeds the fire threshold (i.e., activates the neuron). However, the membrane potential of the background site changes randomly and cannot be accumulated to the threshold. Overall, LIF-Net extracts temporal (i.e., depth direction) information of CT slices by the accumulation of membrane potential; thus, 3D nodule detection is realized in both the temporal domain and spatial domain.

of small sizes; they also proposed a 3D dual-path network which reduced some parameters and achieved comparable performance. Tang et al. [14] integrated nodule candidate screening subnet, false positive reduction subnet, and nodule segmentation subnet to reduce the number of parameters. However, they only compared the number of parameters, whereas it is the number of multiply-add operations that dominate the complexity of a network.

III. METHODS
A. SPATIOTEMPORAL SUPERIORITY OF LIF-NETS
The typical artificial neuron model is shown in Fig.1(a). In ANNs, multiply and accumulate of inputs and weights are the major operations. Neurons propagate information only in the spatial domain. LIF neuron model shown in Fig.1(b) has more dynamic behaviours. In LIF neuron model, the dendrites integrate the input information and update membrane potential, and the soma performs leaky and fire [19]. The information is propagated in spatiotemporal domain, and the current state is tightly affected by the history in the temporal domain.

We take the spatiotemporal superiories brought by membrane potential to realize the 3D object detection and recognition. When it comes to CT automated analysis, the information from each CT slice is first modulated by inter-connecting synaptic weights and then integrated into LIF neurons, which leads to changes in membrane potentials. Once membrane potential reaches the threshold, the soma will be activated and fire a spike, which indicates the nodule-sensitive features are being extracted in both the spatial domain and temporal domain.

As shown in Fig.1(c), the membrane potential of the suspect nodule region (i.e. red curve) increases and integrates from a few previous slices. When it exceeds the threshold, the soma will send a spike down its axon and will deliver to other neurons, and the membrane potential will be reset. New presynaptic inputs will then affect the membrane potential. In contrast, the membrane potentials of the background area (i.e. black curve) less likely exceed above the threshold and have no output ideally. The weights on synapses in the LIF-Nets are trained by Back Propagation Through Time (BPTT) [37].

B. LIF LAYERS AND ResLIF BLOCK
1) LIF MODEL, LIF LAYER, AND ConvLIF LAYER
The basic Leaky Integrate and Fire (LIF) model [19] is used to describe evolution of neuron’s potential and spike activity, which is defined as:

\[
\begin{align*}
\tau_m \frac{dV(t)}{dt} &= -V(t) + \sum_{j \in S^*} W_j I_j(t) \\
\text{fire a spike, } V &\leftarrow V_{\text{reset}}, \text{ if } V \geq V_{th}
\end{align*}
\]

where \(\tau_m\) is a constant representing leaky decay, \(V(t)\) is the membrane potential of the neuron, \(V_{th}\) is the firing threshold, \(V_{\text{reset}}\) is a resting voltage, \(W_j\) is the synaptic weight connecting the \(j\)th presynaptic neuron, \(S^*\) is set of all presynaptic neurons, and \(I_j(t)\) represents the input spike train of the \(j\)th pre-synapses at the current time step.
whether the membrane potential needs to be reset to $V_{\text{reset}} = 1$, otherwise it will be set to 0. Then, we determine $I^t$ is updated based on (2) and (3), where $V_{\text{th}}$ is the threshold of the LIF neuron membrane potential; (c) The approximation at the non-differentiable point; (d) Convolutional LIF.

At time step $t$, the neuron receives the spike train emitted by presynaptic neurons and updates its membrane potential shown in Fig.2(a), according to (1). During the integration stage, the evolution of the neuron membrane potential can be represented by the blue curve in Fig.2(b). When the membrane potential exceeds its threshold $V_{\text{th}}$, the neuron will fire a spike and $V$ will be reset to $V_{\text{reset}}$, this mutation can be indicated by the red curve in Fig.2(b).

However, for mainstream deep learning platforms, codes are executed in patterns of sequential and discrete [38], which makes it difficult to implement the complex evolution of SNN in such continuous time domain. Moreover, the discrete spike train makes it scarcely possible for training SNN with BP like ANN. Therefore, we convert (1) into the Euler format as follows to solve these difficulties.

$$u^t = \begin{cases} I^t W, & \text{FC} \\ \text{Conv}(I^t, W), & \text{Conv} \end{cases}$$ (2)

$$u^{t-1}_m = u^t + \tau I^{t-1}_m$$ (3)

For each time step, the membrane potential of LIF neuron is updated based on (2) and (3), where $I^t$ is the input of the pre-synapses at time step $t$, $W$ is the synaptic weight matrix, $V^{t-1}_m$ is the membrane potential of neuron at time step $t-1$, and $u^t_m$ is the membrane potential of time step $t$.

$$F^t = u^{t}_m \geq V_{\text{th}}$$ (4)

$$V^t_{\text{m}} = \tau_{\alpha} \odot (F^t \odot V_{\text{reset}} + (1 - F^t) \odot u^{t}_m + \tau_{\beta})$$ (5)

According to (4), if $u^t_m$ exceeds the threshold $V_{\text{th}}$, $F^t$ will be set to 1, otherwise it will be set to 0. Then, we determine whether the membrane potential needs to be reset to $V_{\text{reset}}$ by $F^t$ according to (5). $\tau_{\alpha}$ and $\tau_{\beta}$ are introduced to simulate the role of $\tau_m$ (defined in (1)).

$$O^t = \begin{cases} F^t, & \text{for 3D Recognition Net} \\ u^{t-1}_m - V_{\text{th}}, & \text{for 3D Detection Net} \end{cases}$$ (6)

For enhancing the differentiability of the output, $u^{t-1}_m - V_{\text{th}}$ is applied as an alternative to the spiking on detection part in (6). It is noteworthy that the Euler format of LIF can be expanded into convolution patterns (short for Conv), as depicted in Fig.2(d).

Since $F^t$ is retrieved from a step function, the gradient is infinite at $V = V_{\text{th}}$, which leads to the inability to update learnable weights. To this end, we introduce a derivative approximation [39] (see green typical curve in Fig.2(c)) to solve this problem with the hyperparameters $b$ and $g$. The hyperparameters $b$ and $g$ identified in Fig.2(c) are set to 0.2 and 1 respectively in subsequent tasks. We implement the overall training of proposed convolutional LIF in Pytorch with learnable hyperparameters $V_{\text{th}}, V_{\text{reset}}, \tau_{\alpha},$ and $\tau_{\beta}$.

2) ResLIF BLOCK

ResLIF Block is short for Residual Block based on ConvLIF, and the structure is shown by the dotted box with block title in Fig.3. Inspired by ANN-based ResNet [33], ResLIF Block is developed to address the degradation problem of deep LIF-Net when dealing with 3D object detection. ResLIF block has the same skip connection as the residual block in ANN, and consists of two ConvLIF layers, each followed by a 3D BatchNormalization layer and a ReLU layer. In Winterer’s research [40], we can find a similar skip connection structure in the human brain (cortical layer VI neurons get input from layer I, skipping intermediary layers), which gives us a better understanding of ResLIF block from the perspective of bionics.

C. LIF-NET FOR DETECTION

U-Net-like [41] structure is adopted as the backbone of LIF-Net for 3D detection. We conduct two-stage detection, which is analogous to Faster R-CNN [10] in 3D. LIF-Net receives the 3D patches cropped from the 3D CT images as input, and outputs the 4D tensor format of RPN with a size of $15 \times 32 \times 32 \times 32$, as illustrated in Fig.3. The neuron model in LIF faster R-CNN has been defined in the previous section.

The 3D CT patches are first sent to two LIF convolutional layers which have the same $3 \times 3 \times 3$ kernel size and 24 channels, and then to a 3D maxpooling layer which is used to downsample the feature map by setting stride to 2 and kernel size to $2 \times 2 \times 2$. After that, four ResLIF Blocks are used to extract features, with each ResLIF Block followed by a 3D max-pooling layer. The parameters set in LIF convolutional layers and 3D max-pooling layers are consistent with the previous counterparts. Then the two 3D Transposed Convolution Blocks with a stride of 2 and a kernel size of $2 \times 2 \times 2$ cooperate with concatenation units, thus allowing the LIF Faster R-CNN to capture multi-scale nodules...
information. In particular, a $3 \times 32 \times 32 \times 32$ location information is concatenated to the feature map to enhance the robustness of detection model [13]. Finally, the LIF Faster R-CNN squeezes the size of the channels to 15 through two 3D convolution layers and outputs a tensor with a size of $15 \times 32 \times 32 \times 32$.

Each location in the final output tensor contains three scales of nodule-detecting information, obtained from anchor boxes of different scales: 5, 10, and 20. The anchor boxes of each scale contain five pieces of information: the confidence score indicating whether the box is a nodule or not $p_i^*$, the normalized coordinates of the nodule position $x_i^*, y_i^*, z_i^*$, and the nodule size $d_i^*$. If the degree of intersection between the current box and the ground truth bounding box is less than 0.02, it is treated as a negative sample ($p_i' = 0$). Furthermore, we used the hard negative mining [15] to solve the problem of excessive negative samples by selecting the negative samples of the top 800 confidence scores as the hard negatives and discarding the others.

The label of bounding box contains confidence score $p_i'$ (0 for negative samples and 1 for positive samples) and regression labels $l_i'$, which is defined by (7).

$$l_i' = \left( \frac{x_j - \tilde{x}_i}{\tilde{d}_i}, \frac{y_j - \tilde{y}_i}{\tilde{d}_i}, \frac{z_j - \tilde{z}_i}{\tilde{d}_i}, \frac{d_j}{\tilde{d}_i} \right)$$  (7)

The ground truth bounding box of an anchor is denoted by $(x_j, y_j, z_j, d_j)$, and the current bounding box of an anchor by $(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i, \tilde{d}_i)$. The loss function for each anchor is defined as:

$$L(p_i^*, p_i', l_i', l_i') = \alpha \times L_{cls}(p_i, p_i') + p_i'L_{res}(l_i, l_i')$$  (8)
where \( l^n_i \) is the corresponding predicted nodule coordinates and the diameter \((x^*_i, y^*_i, z^*_i, d^*_i)\). The hyper-parameter \( \alpha \) is set to 0.5. The total regression loss \( L_{\text{reg}} \) is designed as the smooth \( l_1 \) regression loss function [42] and the classification loss \( L_{\text{cls}} \) which is designed as the binary cross entropy loss function.

**D. LIF-NET FOR CLASSIFICATION**

A light weight model for 3D classification is built for recognizing the malignancy of lung nodules. As the spike mode has been conducted in the LIF-Net, there is only 0 or 1 value in the axons, thus the multiplication-free method is realized (i.e. replace multiplication with addition). As a result, the LIF-Net for 3D recognition has an extraordinary low computational cost and low latency.

As shown in Fig.3, we first center-crop CT slices at predicted nodule locations with a size of 32 \( \times \) 32 \( \times \) 32. Then two 3 \( \times \) 3 ConvLIF layers (channels being 128 and 256, and stride being 2 and 1, respectively) are used to extract features. After that, a SumLayer is designed to get information from all time steps. These features are flattened into a 128 \( \times \) 1 vector and then are followed by three dense layers with ReLU activation. Finally, softmax is set up for a benign or malignant diagnosis.

The design of SumLayer in the LIF-Net is to integrate temporal information contained in each timesteps. Aiming to reserve more information and reduce the number of fully connected nodes, we put the Sumlayer in front of the flatten layer. This procedure also contributes to reducing overfitting.

**IV. EXPERIMENTS AND RESULTS**

**A. DATASET AND PREPROCESSING**

The Lung Nodule Analysis 2016 dataset (LUNA16) is used in this work, which includes 1186 nodule labels in 888 patients annotated by radiologists. LUNA16 is a subset of LIDC-IDRI, which is the largest publicly available dataset for pulmonary nodules. LUNA16 dataset removes CTs with slice thickness greater than 3mm, slice spacing inconsistent or missing slices from LIDC-IDRI dataset, and explicitly gives the patient-level 10-fold cross-validation split of the dataset [11].

Due to the difference between CT instruments, the spacing between different raw CT images varies. To keep the input images of LIF-Nets unified, we interpolate the original pixels in the x, y, and z directions with a fixed sampling interval and convert original CT slices into new 3D images. The pixels are real-valued numbers, and the pixel intervals in the three axis are equal to the sampling intervals. For the second step, we clip and adjust the raw data to the Hounsfield Unit (HU) value of [-1200, 600]. Then, we normalize the range linearly into \([0, 1]\). Finally, we use LUNA16’s annotated segmentation ground truth and remove the background.

**B. NODULE DETECTION**

We train and validate LIF-Net for 3D nodule detecting on the LUNA16 dataset following 10-fold cross-validation. The evaluation metric is the Free Response Operating Characteristic (FROC) curve [11], which is the average sensitivity (recall rate) at the average number of false positives (FPs) at 1/8, 1/4, 1/2, 1, 2, 4, 8. In training, we augment samples by 3D randomly flipping, 3D rotation, and random rescale between 0.75 and 1.25. The augmentations are applied at random to the training samples during training. We use 150 epochs in total with stochastic gradient descent optimization with the initial learning rate being 0.01, and 0.001 after half the total number of epoch, and 0.0001 after epoch 120. In the testing, the detection probability threshold is set to \(-3\) before sigmoid, and NMS with IoU threshold is set to 0.1. The detection is considered as a true positive if the location falls within the radius of a nodule centroid.

We implement the 3D CNN method [13] for comparison. The FROC curve is shown in Fig.4. The detection result demonstrates that LIF-detection net achieves comparable accuracy with 3D CNN model, which yields a performance of 94.6% at 8 FPs/scan compared to 94.4% at 8 FPs/scan.

**C. NODULE CLASSIFICATION**

In nodule classification, we label 450 positive nodules and 554 negative nodules determined by the annotations matched from LIDC-IDRI dataset, which are identified by four experienced thoracic radiologists. The benign or malignant score ranges from 1 to 5; we label each nodule as the positive sample if the score is greater than 3, and vice versa.
**TABLE 2.** Formulas for calculating the computational complexity and the weights of different network layers. FMUL, FADD, and FWeights refer to formulas for calculating the number of multiplication, addition and weights, respectively.

| Parameter | Description |
|-----------|-------------|
| \((T, H, W, C_{in})\) | Shape of the feature map, as Input |
| \((T, M, N, C_{out})\) | Shape of the feature map, as Output |
| \((U, I, J)\) | Kernel size of 3D convolution |
| \((I, J)\) | Kernel size of LIF convolution |
| \(B\) | Batch size |

| Layer/Name | FMUL | FADD | FWeights |
|------------|------|------|----------|
| 3D Conv \(U \times I \times J \times T \times M \times N \times C_{in} \times C_{out}\) | \(U \times I \times J \times T \times M \times N \times C_{in} \times C_{out}\) | \(I \times J \times C_{in} \times C_{out} \times U + C_{out}\) |
| LIF Conv \((I \times J \times C_{in} + 1) \times M \times N \times C_{out} \times T\) | \((I \times J \times K + 2) \times M \times N \times L \times T\) | \(I \times J \times C_{in} \times C_{out} + L\) |
| 3D Batch Normalization \(2 \times B \times T \times H \times W \times C_{in}\) | \(5 \times B \times W \times H \times C - 2\) | \(2 \times C_{in}\) |

**TABLE 3.** Efficiency comparison of 3D CNN residual blocks and LIF (Conv) residual blocks with the same parameter set. Both blocks have input size \((128, 128, 128, 24)\), kernel size of \(3 \times 3 \times 3\), batch size \(= 1\), stride \(= 1\), padding \(= 1\) and output channel \(= 24\). #MUL refers to the number of multiplication, #ADD refers to the number of addition, Weights refer to the number of weights.

| Block Name | #MUL | #ADD | Weights |
|------------|------|------|---------|
| 3D Residual | 65,431,142,400 | 65,733,123,288 | 31,248 |
| LIF Residual | 22,045,261,824 | 22,447,915,008 | 10,512 |
| Reduced | 66.31% | 65.85% | 66.36% |

**TABLE 4.** Efficiency comparison of 3D-U-Net and LIF-U-Net with the same network structure.

| Model Name | #MUL | #ADD | Weights |
|------------|------|------|---------|
| 3D-U-Net | 170,214,686,720 | 170,773,643,264 | 5,381,319 |
| LIF-U-Net | 58,815,479,808 | 59,560,493,056 | 1,866,135 |
| Reduced | 65.45% | 65.12% | 65.32% |

In training, 3D data augmentation is applied, including random flip, random angle rotation, and random rescale. The total number of training epochs is 150. The initial learning rate is 0.001, and the learning rate reduces to 0.0001 after 50 epochs. We use each fold for the test, and the final performance is the average performance on ten test folds. The nodule classification performance is concluded in Table 1.

From Table 1, our LIF-classification net achieves the best performance among Multi-scale CNN [8], Vanilla 3DCNN [6], Multi-crop CNN [7], and Dual-Path Nets [13], and at the same time more than 60% amount of computation is saved. Our testing results also demonstrate that even when small sample sizes are used for training, the LIF-classification net can achieve reasonably outstanding separation.

Note that, our LIF-classification is light in network structure, and is successful in providing doctors with reliable and clear diagnostic recommendations. There are methods achieved over 95% accuracy in literatures [44], [45]; however, they use different evaluation criteria (composite rank of malignancy 1 and 2 as benign, 4 and 5 as malignant, and 3 is neglected) or use transfer learning methods pre-training on other datasets.

### D. EFFICIENCY EVALUATION

The most significant advantage of SNNs over ANNs is the low power consumption due to several aspects, including conduct multiplication-free operations, saving storage of activation values, and using low complexity model [18]. The main reason for the first two advantages is that the activations are binary. In addition, SNNs benefit from low power consumption due to the small amount of weights used compared with 3D CNN, measured by the equivalent number of neurons.

Saving weights leads to the reduction of computational complexity and memory capacity [18]. We evaluate the computational complexity and the number of network weights through the layer block level and network level. For layer level, we compare the residual blocks based on 3D CNN and LIF (Conv) with the same parameter set, and the results are shown in Table 3.

For the network level, Table 4 provides a comparison between LIF faster R-CNN proposed in the detection task and 3D faster R-CNN with the same network structure. It is noticeable that we summarize the formulas needed [46], [47] for evaluation in Table 2, which includes all the layers required to evaluate the residual block and network.
FIGURE 5. Visualization of central slices for nodule ground truth (shown in blue boxes) and LIF-Net nodule detection results (shown in red boxes).

As the evaluation results shown in Table 3 and Table 4, LIF-Net has about 65% less computation and network parameters than the network based on 3D CNN, under the same conditions. The results confirm the conclusion that the proposed LIF-Net has lower power consumption without reducing network performance. Compared with the traditional 3D network, it can be more easily deployed to hardware with less computing resources and has more flexible application prospects.

Different from traditional ANN-SNN conversion schemes [48], we conduct convolutional LIF and take the advantage of membrane potential to obtain inter-frame information of CT slices, in other words, each CT slice can be considered as a time step. Parameter description, formulas for computation complexity, and weights of different network layers are shown in Table 2. FMUL, FADD, and FWEIGHT refer to formulas for calculating the number of multiplications, additions, and weights, respectively.

E. VISUALIZATION OF DIAGNOSIS RESULTS

In this section, we present the visualization of nodule detection and recognition results of LIF-Nets. As shown in Fig.5(a)(b), our LIF-Nets can screen the suspect nodule from other nodule-like organs, and perform well when the nodule is located near the thoracic cavity. Fig.5(c)(d) shows that various nodules can be accurately detected by the anchors of different sizes produced by LIF-Nets. The benign and malignant nodules recognized by LIF-Nets are shown in Fig.5(e)(f). The results demonstrate the LIF-Nets perform well.

V. DISCUSSION

LIF-Nets can be exploited in other object detection applications such as Lidar data processing in autonomous driving. Real-time object perception is essential for self-driving cars at high speed to take evasive actions in time. Moreover, building a portable capture platform necessitates operating under limited computational resources. Given these conditions, LIF-Nets could outperform other methods owing to its low computational complexity, low storage cost, low energy consumption, yet rich spatiotemporal capability.

In the medical field, advanced artificial intelligent human-computer interaction systems, real-time operation assistant systems, and wearable SoC devices can be further explored by developing power-efficient LIF-Nets.

VI. CONCLUSION

In this paper, we propose light weight LIF-Nets for 3D detection and 3D recognition. To our best knowledge, this is the first SNN that is exploited to 3D object detection and 3D computer-aided diagnosis, which achieves comparable accuracy while significantly saves computational costs up to 65%. The LIF classification net outperforms prevalent 3D models. U-LIF structure with ResLIF Blocks is developed to
address degradation problem and improves feature representation capability, which makes easier to develop deeper LIF-Nets exploited in more complicated 3D processing applications. This work highlights the spatiotemporal superiorities of LIF-Nets, which contributes to decreasing complexity, saving the energy, and reducing the latency. We believe that this work is the first step to demonstrate the strong competitiveness of SNNs in 3D processing and look forward that SNNs considered as the third generation of artificial intelligent networks to increase the quality of life ubiquitously.

ACKNOWLEDGMENT

The authors gratefully acknowledge Beijing Lynxi Technology Company Ltd. for the support of Brain-Inspired Network Training Platform used in this research. The authors would like to thank all the reviewers and editors for their insightful comments and helpful suggestions.

REFERENCES

[1] A. El-Baz, G. M. Beach, G. Gimel'farb, K. Suzuki, K. Okada, A. Elnakib, A. Soliman, and B. Abdollahi, “Computer-aided diagnosis systems for lung cancer: Challenges and methodologies,” Int. J. Biomed. Imag., vol. 2013, pp. 1–46, Jan. 2013.
[2] 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. Interventions. Cham, Switzerland: Springer, 2017, pp. 559–567.
[3] Q. Dou, H. Chen, Y. Jin, H. Lin, J. Qin, and P.-A. Heng, “Automated pulmonary nodule detection via 3D ConvNets with online sample filtering and hybrid-loss residual learning,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interventions. Cham, Switzerland: Springer, 2017, pp. 630–638.
[4] E. L. Torres, A. Traverso, S. Bagnasco, C. Bracco, D. Campanella, M. Fantacci, S. Lusso, D. Regge, M. Saletta, and M. Stasi, “Clinical validation of the MSL lung computer-assisted detection system,” Radiotherapy Oncol., vol. 118, pp. S22–S23, Apr. 2016.
[5] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. S. Tavares, “Automatic 3D pulmonary nodule detection in CT images: A survey,” Comput. Methods Programs Biomed., vol. 124, pp. 91–107, Feb. 2016.
[6] X. Yan, J. Pang, H. Qi, Y. Zhu, C. Bai, X. Geng, M. Liu, D. Terzopoulos, and X. Ding, “Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2D and 3D strategies,” in Proc. Asian Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 91–101.
[7] 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.
[8] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, “Multi-scale convolutional neural networks for lung nodule classification,” in Proc. Int. Conf. Inf. Process. Med. Imag., vol. 24, Jun. 2015, pp. 588–599.
[9] S. Hussein, K. Cao, Q. Song, and U. Bagci, “Risk stratification of lung nodules using 3D CNN-based multi-task learning,” in Information Processing in Medical Imaging. Cham, Switzerland: Springer, 2017, pp. 249–260.
[10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[11] 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. 2016.
[12] H. Tang, D. R. Kim, and X. Xie, “Automated pulmonary nodule detection using 3D deep convolutional neural networks,” in Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI), Apr. 2018, pp. 523–526.
[13] 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.
**HEHUI ZHANG** received the B.Eng. degree in computer science and technology from the Tianjin University of Technology, Tianjin, China, in 2018. He is currently pursuing the M.E. degree with the Beijing University of Posts and Telecommunications, Beijing, China. His research interests include spiking neural networks and object detection.

**ZHENZHI WU** received the B.S. degree in electronic science and technology and the Ph.D. degree in information and communication engineering from the Beijing Institute of Technology (BIT), in 2008 and 2015, respectively. He was a Visiting Researcher with the Division of Computer Engineering, Linköping University, Linköping, Sweden, from 2012 to 2014. From 2016 to 2018, he was an Assistant Researcher with the Center for Brain Inspired Computing Research (CBICR), Tsinghua University, Beijing, China. He is currently the Technical Director of Beijing Lynxi Technology Company Ltd. His research interests include brain-inspired algorithms and chip design.

**SHIWEI REN** received the B.Eng. degree in information engineering from the Beijing Institute of Technology, in 2008, and the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, in 2013. In 2013, she joined the School of Information and Electronics, Beijing Institute of Technology, as a Faculty Member. Her research interests include direction-of-arrival estimation, sparse array, and signal processing.

**YUETING SHI** received the B.Eng. degree in telecommunication engineering from the Beijing Institute of Technology, Beijing, China, in 2013, where she is currently pursuing the Ph.D. degree. From 2016 to 2017, she was a Visiting Scholar with the Department of Bioengineering, Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA. Her research interests include medical imaging, image processing, and deep learning.

**HAI LI** (Member, IEEE) received the B.Sc. and Ph.D. degrees from the Beijing Institute of Technology, in 1997 and 2002, respectively. He is currently an Associate Professor with the School of Information and Electronics, Beijing Institute of Technology. His research interests include signal processing, protocol engineering, and wireless communications.