IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 22, 2019, accepted November 8, 2019, date of publication November 14, 2019,
date of current version December 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953517

Liver Detection Algorithm Based on an Improved Deep Network Combined With Edge Perception

KAIJIAN XIA[1,2] **AND HONGSHENG YIN**[1]

[1]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2]The affiliated Changshu Hospital of Soochow University (Changshu No.1 people's Hospital), Changshu 215500, China

Corresponding authors: Kaijian Xia (xiakaijian@163.com) and Hongsheng Yin (xuzhouyhs@sina.com)

This work was supported in part by the Jiangsu Committee of Health on the Subject under Grant H2018071, and in part by the Open Fund Project of Jiangsu Key Laboratory of Media Design and Software Technology (Jiangnan University) under Grant 19ST0205.

ABSTRACT Due to the large number of organs and the similar grayscale in abdominal medical images, accurately locating and identifying the liver in an abdominal image is a challenging problem. To improve the accuracies of liver detection and localization, this paper proposes an improved deep network that is combined with edge perception. The network improves the contour-detection accuracy of the liver via an edge-perception fusion module and captures the high-level semantic features of the abdominal image using a multiscale pyramid pooling layer. The complementary characteristics of the edge-related features can effectively preserve the clear boundaries of the liver, while rich global context information can be extracted from the combination of the auxiliary channel output and the pyramid pooling layer output. Many qualitative and quantitative experimental results demonstrate that the proposed model can effectively improve the performance of detection and localization networks, which can narrow the range of regions of interest, and can enhance the accuracy of subsequent segmentation and recognition.

INDEX TERMS Liver detection, edge perception, contour prior, pyramid pooling, deep learning, semantic features.

I. INTRODUCTION

The locations of anatomical structures in medical images are highly important for various clinical image interpretation applications. This information can improve the accuracies of target segmentation, detection and recognition [1]. In the diagnosis of liver lesions, to quantify the regional characteristics of lesions and to evaluate the progress of or therapeutic efficacy against tumors, using liver localization as a pretreatment step will effectively improve the accuracy of detection/segmentation of tumors [2].

Although CT images provide valuable information for organ detection and segmentation, it is difficult to locate target organs directly via traditional detection algorithms due to the low contrast and similar gray levels between organs. In recent years, experts and scholars at home and abroad have proposed many target detection and location algorithms [4]. Available localization methods for abdominal imaging of single or multiple organs can be roughly divided into two groups: (1) localization methods that are based on maps, which rely on image registration technology that is based on massive voxel information [5]; and (2) localization methods that are based on machine learning (ML), which rely on training algorithms to learn the discriminating features [6] of organs. Machine learning methods can be classified into classical learning methods and deep learning methods. The former extract artificially designed features and use them as input for training subsequent classifications or regression models. The latter do not manually design feature operators but typically select or design deep structures and provide relatively large training data sets.

After years of development, there remain urgent problems with liver detection technology to be solved, which are mainly due to the non-rigid characteristics of abdominal organs and the influence of the complex background [7]. Most traditional organ detection algorithms adopt semi-supervised learning methods, which match by designing reference maps and locate via feature classification. Since traditional methods must design artificial features for each task, such as gray-level features, contour features, and HOG features, these features are easily affected by the imaging angle, and their generalization performance is poor [8]. Tran *et al.* [9] proposed a multifeature target detection model that is based on saliency segmentation. This method combines multiple feature space

The associate editor coordinating the review of this manuscript and approving it for publication was Yongtao Hao.

models and uses them to segment candidate regions from a complex background, and it uses improved HU invariant moments to describe the contours to obtain multifeature cascade vectors of the segmentation region. Finally, SVM is used to classify feature data. This method has high positioning accuracy for high-contrast test data; however, its detection accuracy for abdomen images that are captured under clinical conditions is poor.

Object detection and location models that are based on machine learning are the current research direction, which improve the accuracy of recognition. Varol *et al.* [10] proposed a liver detection method that is based on intelligent learning, which intelligently analyzed complex abdominal images via a machine learning algorithm, acquired the temporal and spatial characteristics of abdominal images, and conducted the corresponding high-precision classification. Although the accuracy of model classification can be improved by directly extracting features from the detection area and using an intelligent algorithm to classify and judge, the feature representation performance directly affects the recognition accuracy of the system. Traditional contour recognition models, such as models that are based on image processing and machine learning, are not susceptible to noise or other disturbances but cannot effectively recognize large-scale medical image data. Compared with SVM, decision tree, neural network and other shallow learning model detection algorithms [11], CNN, which is the representative deep learning approach, emphasizes the depth of the model structure, highlights the importance of learning behavior characteristics, and more accurately characterizes data-rich feature information by using big data to learn features. Ng *et al.* [12] applied a convolutional neural network to the field of abdominal organ recognition for the first time. By designing a deeper network structure and using dropout technology to enhance the generalization performance of the model, its detection accuracy is higher than that of traditional shallow learning models. A deep model is essentially a deep nonlinear network structure, which possesses learning ability with powerful essential features and can implement complex function approximation by combining low-level features into high-level features that are more abstract using multiple hidden nodes in the network.

With the rapid development and extensive application of deep learning, the available deep learning algorithms, such as AlexNet [13], GoogleNet [14], and residual network (ResNet) [15], have yielded satisfactory results in image classification and have high application prospects. However, there are also shortcomings, such as large network computations, complex models, and poor real-time performance. At the same time, the available deep learning network structure only uses high-level features for image classification and detection, which renders it difficult to distinguish targets that require the detection of fine features, such as thyroid nodules and small tumors. Jiang *et al.* [16] designed a hippocampal localization method for brain CT images that is based on the LeNet-5 network. This method does not extract features manually. Therefore, the available deep learning model cannot be used directly for medical image detection and localization.

Due to the complex backgrounds of abdominal images, based on the multiscale deep model, an improved multiscale deep learning network is proposed, which improves the accuracy of liver contour detection through the edge perception fusion module and uses a multiscale pyramid pooling layer to capture the high-level sub-semantic features of abdominal images. The complementary features of edge-related features can effectively preserve the clear boundary of the liver, while the combination of the auxiliary side output and the pyramid pooling layer output can be used to extract abundant global context information.

II. DENSENET DEEP MODEL

The gradient on the front layer in a multilayer deep network comes from the product of the gradient on the back layer. With the deepening of the network depth, the learning rate of the front hidden layer may be lower than that of the back hidden layer, which causes the classification accuracy of the back layer to decline, thereby resulting in the disappearance of the gradient or gradient explosion [3]–[5]. Most available networks adopt cross-connections, which are similar to identical mappings, such as ResNet [6] and FractalNets [7], whose unit structure output is expressed in equation (1).

$$x_l = H_l(x_{l-1}) + x_{l-1} \tag{1}$$

where x_l is the output data of layer l and H_l is the nonlinear transformation of the network. The output of layer l is the output of layer $l-1$ plus the nonlinear transformation of the output of the opposite layer, namely, layer $l-1$. The features that are extracted from each layer of a traditional deep network are equivalent to a nonlinear transformation of the input data. With the increase of the depth, the complexity of nonlinear functions increases gradually [8], [9]. The DenseNet network also connects all layers, but the input of each layer comes from the outputs of all the previous layers, which is equivalent to each layer directly connecting to the input layer and the loss layer, to alleviate the gradient disappearance phenomenon, and the network structure becomes increasingly compact. Its output is expressed in Equation (2).

$$x_l = H_l([x_0, x_1, \cdots, x_{l-1}]) \tag{2}$$

where $[x_0, x_1, \cdots, x_{l-1}]$ is a cascade of feature graphs from the outputs of layers 0 to $l-1$, which is similar to the Inception operation [10]. To avoid the feature dimension of the latter layer growing too fast with the increase of the network layers, DenseNet pools and balances after each cell module, as illustrated in Figure 1.

III. PROPOSED LIVER LOCATION ALGORITHM

The key strategy of liver localization is to obtain the contour information of liver and accurately localize the liver region via feature modeling. However, due to the lack of constraints

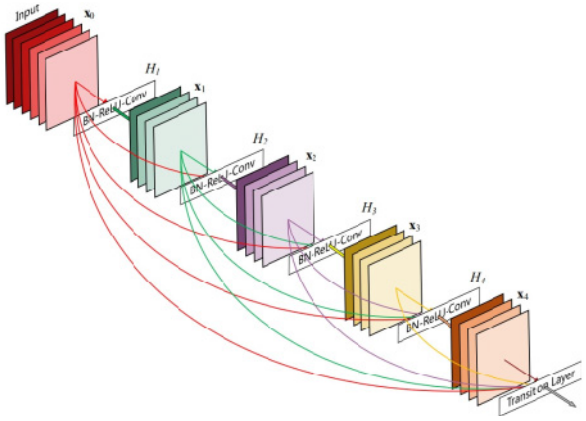

FIGURE 1. 5 Layer DenseNet network.

on the boundary area of the existing liver contour segmentation algorithm, it cannot effectively improve the accuracy of the target boundary contour [17]. In addition, although the method that is based on multichannel fusion integrates a variety of feature information, it lacks a representation of global structure information, and the network is still unable to separate the liver from the complex background. When using the semantic features of the liver for supervised recognition, the available methods mainly roughly estimate the location of the liver from the reference position, which cannot blur the edge of the liver, thereby resulting in low detection efficiency. Therefore, this paper improves the prediction accuracy by using a liver contour detection network that is based on edge perception. Moreover, the available liver location detection methods only utilize bypass output features (feature addition or feature stitching) that are based on channel operation, while ignoring the importance of global structural features. Therefore, this paper combines ASPP [18] and auxiliary supervision [19] to overcome the deficiencies of the previous models. The network structure of the proposed method is illustrated in Figure 2, which consists of two modules: the edge perception fusion module is based on the VGG-16 network [20], and the multi-scale ASPP is based on the DenseNet network.

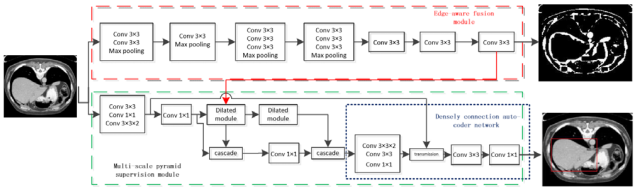

FIGURE 2. Model frameworks.

For preserving the edges, the proposed edge perception fusion module can learn additional fine liver contour information. Inspired by the HED model, this model trains an edge detection stream (EDS) that is based on the VGG network on the IRCAD liver data set, and the parameters of EDS are fixed when training the loss function. Then, for each input

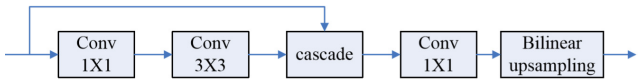

FIGURE 3. Decoder block.

image, we extract edge maps from the trained EDS and fuse them with DenseBlocks bypass output features to help train the detection stream (SDS) to locate the liver. In conclusion, EFM aims at providing additional edge-related information for liver detection tasks.

From a global perspective, this paper adopts a multiscale pyramid monitoring module, which is based on the pyramid pooling model (PPM). PPM implements a series of pooling operations to collect global structural information, and multiple bypass output monitoring of the monitoring module can generate liver location predictions hierarchically. Combining these two parts, the module can effectively improve the final prediction results.

A. DENSELY CONNECTED AUTO-CODER NETWORK

Due to the excellent performance of the densely connected model in classification tasks and the efficient implementation of memory, this paper chooses densely connected network 121 (Dense Net121) as the deep coding model [21]. Compared with VGG16 and ResNet-50, the DenseNet-based training process can converge faster through dense residual connections. Our encoder structure consists of the first convolution layer and the pooling layer (DenseBlock 0), together with the first three DenseBlocks (DenseBlock 1, DenseBlock 2, and DenseBlock 3), from the pretrained DenseNet 121. The resolution of the output feature map is 1/32 the size of the input image [22]. As illustrated in Figure 3, each decoder module is designed as a residual conversion module, in which 1*1 convolution is responsible for increasing and reducing the dimensions, and the last layer is designed to double the resolution of the feature map via bilinear interpolation. The input of each decoder module is composed of two parts with the same spatial resolution. The first part is the feature map from each corresponding DenseBlock. For simplicity, the batch normalization layer and the Relu layer after each convolution layer in the decoder module are omitted.

B. EDGE-AWARE FUSION MODULE

The coding and decoding module feedforward network that is described above can roughly localize the liver; however, because the upsampling operation cannot recover spatial information and finer details, it cannot retain the edge structure well. Therefore, we propose a novel edge perception fusion module for preserving more edge structure information and integrate the edge perception feature map into the liver detection task. The edge detection task is to detect the liver edge and the object edge in an abdominal image. Edge detection is a basic computer vision task, and it is an important step in segmentation and target detection

tasks [23]–[25]. Here, we use complementary information with edge-related information to facilitate liver contour detection. First, the monitoring flow module, which is based on the VGG network, is trained separately on the benchmark data set for liver detection. The module is constructed from the HED module, which splices multiple side outputs together. Then, the spliced outputs are fed into the fusion layer to obtain a unified output, in which the fusion layer is a convolution with a core size of 51. The structure of the liver contour monitoring module, which is based on the VGG network, is illustrated in the red box of Figure 1.

We connect each input image to the monitoring module and the detection module. The model parameters of the monitoring module are fixed, and the multilevel edge perception feature map is fused into the detection module to help train the liver detection task on the abdominal liver data set. The four levels of the monitoring module, namely, conv2-2 (128 channels), conv3-3 (256 channels), conv4-3 (512 channels) and conv5-3 (512 channels), are used for feature mapping. Then, they are fused with the outputs of Decoder4, Decoder3, Decoder2 and Decoder1 via splicing operations. The fusion mechanism of the edge perception fusion module can be expressed as follows:

$$\mathrm{F}_i = Cat(X_i, E_\mathrm{j}) \tag{3}$$

where F_i denotes the output characteristic diagram of the module, X_i is the intermediate feature of the corresponding block in the SDS module, and $\mathrm{E_j}$ is the intermediate feature of the corresponding block in the detection module. The values of (i, j) are {(5, 1), (4, 2), (3, 3), (2, 4)}. Cat is a channel-by-channel splicing operation. In addition, for Decoder 5, which has the same spatial resolution as the input image, we fuse the final edge detection result (1 channel) with the original gray image to form the most complementary detail [26].

A denotes the output characteristic diagram of the module, B is the intermediate feature of the corresponding block in the SDS module, and B is the intermediate feature of the corresponding block in the detection module. The values of i and j are (i, j) {(5, 1), (4, 2), (3, 3), (2, 4)}. **Cat** is a channel-by-channel splicing operation. In addition, for Decoder 5 with the same spatial resolution as the input image, we fuse the final edge detection result (1 channel) with the original gray image to form the most complementary detail [26].

Therefore, this paper also uses the VGG-16 network structure to extract contour features. To detect edges, modifications are made to the contour detection network: First, remove all fully connected layers. Then, remove the last maximum pooling layer. Finally, connect the output to the refinement module. The largest pooled conv1, conv2, conv3, conv4 and conv5 in the VGG-16 network are used as the front-end to extract target features. Then, the back-end is modified to enable the network to extract contour information. This strategy originates from the five multiscale aggregation layers in holistically nested edge detection (HED): the lower layer captures more spatial details but lacks sufficient semantic information, whereas the deeper layer encodes more semantic information but lacks spatial details. In this task, useless background information and abstract target contour interference must be reduced. Therefore, this study uses deeper features to build templates. However, especially at a deeper level, edge output in HED networks is more susceptible to severe edge problems. Therefore, it is necessary to refine the output boundary to generate a clear and accurate target contour. In addition, the final output must be adjusted to the original size via upsampling maximum pooling and deconvolution. To realize the optimal performance in the convolution process, this paper chooses the smallest convolution filter (3 *3), whose stride is one pixel, which can capture left/right, up/down and central motions; the largest pooling operates with a step of 2 on a 2 *2-pixel window. The parameters of each convolution layer and the maximum pooling layer in the contour detection network that are used in this paper are listed in Table 1, where RF, C and P denote receptive field, convolution and pooling, respectively.

TABLE 1. Parameter settings for each layer in the contour detection network.

Layer	C1	P1	C2	P2	C3	P3	C4	P4	C5
RF size	5	6	14	16	40	44	92	100	196
Stride	1	2	2	4	4	8	8	16	16

C. MULTISCALE PYRAMID SUPERVISION MODULE

Although the coding and decoding network can combine feature maps from various levels through splicing operations, due to the lack of a global context structure, it cannot capture multiscale liver objects in input images. To overcome this problem, this paper uses the pyramid pooling module (PPM), which has realized high performance in classification and segmentation tasks. First, four-level average pooling operations are performed in cascaded networks. Then, the pooling characteristics are sampled to the same size as the original input. Finally, the multistage pooling feature is combined with the original feature. Based on the basic pyramid pooling operation, a new multiscale pyramid monitoring module is proposed for more effectively utilizing the multiscale global context structure. Instead of only applying one PPM network to the final prediction part, it combines multiple auxiliary side outputs with the PPM network to realize multiscale monitoring. Therefore, we add a PPM network module after each edge perception fusion module to accelerate the learning process. The equation is presented in (4):

$$S_i = \sigma(W_i^* Cat(W_{i,1}^* PP_1(X_i), \cdots, W_{i,n}^* PP_n(X_i)) \tag{4}$$

where $X_i(i \in \{2, 3, 4, 5\})$ is the output of the ith decoding block, PP_n is the pyramidal pooling operation $n \in \{1, 2, 3, 4\}$ at level n, $W_{i,n}$ is the weight of the 1*1 convolution, σ is the sigmoid function that is used to scale parameters, and $S_i()$ is the last ith bypass predictive output.

IV. EXPERIMENT AND ANALYSIS

A. EXPERIMENTAL DATA AND PARAMETER SETTINGS

To evaluate the performance of the improved deep network liver detection algorithm, in this paper, a liver database that is published by IRCAD International Medical Center in France [16] and an abdominal image database are used as training image data sets. Among them, 8,500 CT images of the liver have been accurately labeled by medical experts with clinical experience. Except for the target area, the remaining area is marked as background. Hence, the labeled data sets can be used to train and test the liver localization model. These data contain many abdominal images with complex backgrounds. There are distorted images in the data set, but they do not affect the medical experts' labeling and positioning.

This paper proposes a network module that is completed under the framework of TensorFlow in-depth learning [27], [28]. To increase the efficiency of optimization, the ADAM optimization algorithm is used. The parameters are as follows: alpha = 0.001, beta 1 = 0.9, beta 2 = 0.999 and epsilon = 10-8. The batch size of each stream is 1. For the liver contour detection module, the learning rate is set to 10^{-6}, and the training iteration is set to 200 k/time; and for the monitoring flow module that is based on the VGG network, the learning rate is set to 300 k/time for iteration. The training sample size was adjusted to 384×384 and served as the input of the network. Graphic processing unit GTX 1080TI is used for training and prediction. It takes approximately 8 hours to train the edge detection network and 7 hours to train the liver location detection network. At the detection end, the proposed model can process 5 frames per second.

B. EVALUATION INDEX AND ITS CONTRAST ALGORITHMS

The medical image intelligent analysis system has been able to identify various organs, tumors and other salient areas. However, there are substantial errors in the available algorithms, which occur mainly because the grayscale of the liver is similar among sections and the liver shape differs substantially among sections. In this paper, the detection rate (DR) and the number of false-positive detections per image (FPPI) are used as evaluation criteria. The relationship between DR and FPPI is as follows:

$$\mathrm{DR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{5}$$

$$\mathrm{FPPI} = FP/(FP + TN) \tag{6}$$

where TP represents the number of positive examples that are correctly detected; $TP + FN$ represents the total number of positive examples in the image; and FP represents the number of false-positive cases. To facilitate the analysis, we divide the test data into 5 categories; the identifications and characteristics are listed in Table 2.

In addition, to evaluate the accuracy of liver detection and location results, the pixel accuracy (PA) and the mean intersection over union (MIoU) are used as evaluation criteria [30]. The calculation formulas are as follows:

TABLE 2. Characteristics of the test subsets.

Identification	Characteristic
A	Fuzzy, low contrast
B	The liver is small and there are many highlighted areas in the background.
C	The liver region is large and fills whole CT images.
D	The boundary of the liver is not readily identifiable.
E	Noise is high, and there are tumors and internal heterogeneity.

$$\mathrm{PA} = \sum_i n_{ii} / \sum_i t_i \tag{7}$$

$$\mathrm{IoU} = (1/n_{cl}) \sum_i n_{ii} / (t_i + \sum_j n_{ji} - n_{ii}) \tag{8}$$

where n_{ij} is the number of pixels in class i that have been correctly classified as class j, and t_i is the number of samples in class i. According to the definition of IOU, this is equivalent to dividing the overlap of two regions by the set of two regions. A score of greater than 0.5 corresponds to an accurate detection and segmentation result.

Experts at home and abroad have proposed many target detection algorithms for abdominal images. However, few models are available for liver detection in abdominal CT images, and most of the algorithms are for tumor detection. The objectives of this paper are to conduct target detection as the pretreatment step of the subsequent cancer detection, segmentation and recognition algorithm, to narrow the scope of subsequent processing and to reduce the impact of background interference. The approach that was proposed by Dr. Debin Lei, Institute of Automation, Chinese Academy of Sciences, is to optimize the energy function through edge constraints to identify the suspected liver region in CT images. However, since open-source code is not provided for this method, it is difficult to conduct a quantitative analysis. Although open-source code is available for the model that was proposed by Azizi *et al.* [26], this method is mainly aimed at PET images and is not suitable for the data that are processed in this paper. In addition, most evaluation indicators of available liver detection algorithms are based on the positioning accuracy, which directly depends on the size of the data voxels and cannot be used directly. Although there are few algorithms that can be directly compared qualitatively and quantitatively, comparative experiments are conducted in this paper.

Detection algorithms that are based on deep learning have yielded highly satisfactory results in the field of natural image processing. To qualitatively and quantitatively analyze the accuracy of liver detection in the field of medical imaging, in this paper, we consider ConvNet [31], YOLO-v2 [32],

SSD [33], DenseNet [34], and ResNet [35]. For all comparison algorithms, the source codes or executable files that were provided by the authors are used. Because the compared algorithm detects natural images, to facilitate fair qualitative and quantitative comparison, all deep algorithm models are trained with the same training set.

C. QUANTITATIVE COMPARISON AND ANALYSIS

1) COMPARISON IN TERMS OF DETECTION ACCURACY

Table 3 presents the test results of the evaluation algorithms for several test subsets, where the best and the second-best results are shown in bold and italics, respectively.

TABLE 3. Detection rates under various data sets.

Data set	Algorithms					
	SDD	DenseNet	ResNet	ConvNet	YOLO-v2	The proposed algorithm
A	74.11%	85.47%	*87.22%*	86.28%	*90.15%*	**91.26%**
B	61.11%	66.93%	*71.09%*	77.00%	*77.59%*	**78.35%**
C	69.12%	78.22%	*81.25%*	*88.24%*	87.09%	**88.28%**
D	68.28%	85.28%	*88.20%*	85.69%	*88.36%*	**89.37%**
E	51.91%	52.90%	*59.75%*	59.11%	58.69%	**59.78%**

According to the test results in Table 3, the improved DenseNet deep network that is proposed in this paper outperforms the original DenseNet. The main reason is that the edge perception fusion module improves the accuracy of liver contour detection and captures the high-level semantic features of abdominal images using a multiscale pyramid pooling layer. Complementary features of edge-related features can effectively preserve the clear boundary of the liver, and the combination of auxiliary side output and the pyramidal pooling layer output can extract rich global context information to adapt to the binary classification problem of liver detection. Data set E contains the most complex images in the whole test data set. There are many interference targets in the image, the noise level is high, and there are tumors in the liver, which cause an uneven internal gray level. This directly affects the detection performance of the network. According to the accuracy comparison results in Table 3, the detection rate on data set E is the lowest, but it is higher compared to the other deep networks.

2) COMPARISON OF FPPI

FPPI represents the statistical results of various detection rates of liver detection algorithms, as shown in Figure 4. To facilitate the comparison of the liver detection performances in abdominal CT images among various deep networks, the detection results of each algorithm with FPPI = 1 are considered for visual analysis. In the same image database, when FPPI = 1, the detection rate of the proposed algorithm is 97.21%, while the best results of the comparison algorithms are 95.44% for the YOLO-v2 algorithm, 89.1% for the ConvNet algorithm, 91.88% for the ResNet algorithm, 92.15% for the SSD algorithm, and 87.1% for the DenseNet algorithm. The main reason is that most of the deep detection methods only use bypass output features (feature addition or feature splicing) that are based on the channel operation for the liver but ignore the importance of global structural features [36]. The improved network model that is proposed in this paper is based on edge perception by the liver contour detection network to improve the prediction accuracy. Therefore, the deep network that is proposed in this paper can more accurately extract the liver features of abdominal images, and the training is strengthened via the improved predictive rectangular box method, thereby further reducing the false detection rate on each image.

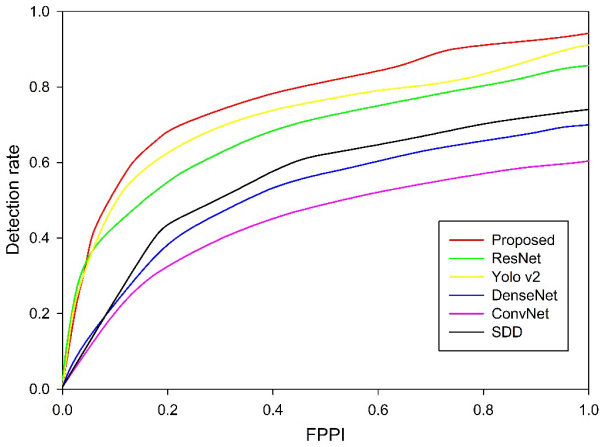

FIGURE 4. Relationship curves between the detection rate and FPPI.

3) COMPARISON OF PERFORMANCE INDICATORS OF TEST BOXES

Figure 5 plots the accuracy and success curves of target detection.

According to Fig. 5, the proposed detection method realizes the best detection performance among the six compared detection algorithms. The distance accuracy and the overlap accuracy are plotted in Fig. 5 (a) and (b), respectively. In the medical image analysis system, the accuracy of the center position determines whether the test results satisfy the requirements. To facilitate quantitative analysis, this paper compares the percentage center position error within a 10-pixel threshold. The center error accuracy of the proposed detection algorithm within a 10-pixel threshold is 0.91, and the optimal detection performance is realized. The detection success rate of the proposed algorithm is 0.872 at the overlap rate threshold of 0.5, which is 4.7% higher than that of suboptimal model Yolo v2. The quantitative analysis results demonstrate that the proposed contour perception deep network has high performance and improves the accuracy of detection

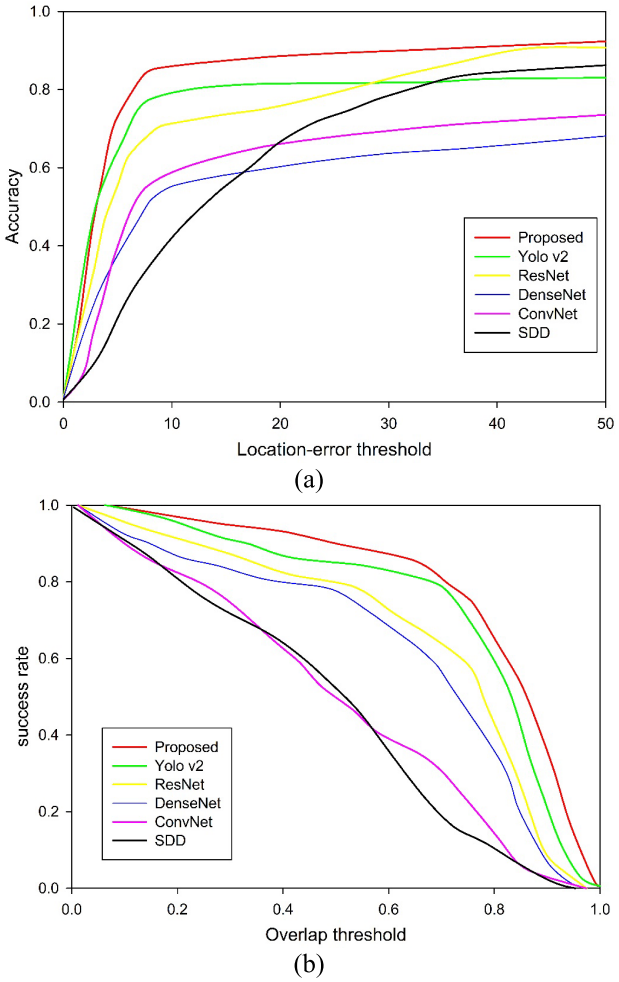

FIGURE 5. Performance comparison of six detection algorithms in terms of (a) accuracy and (b) success rate.

through contour detection and recognition. Especially for liver images from various perspectives, this algorithm can accurately mark the location of the liver, reduce the consumption of artificial markers, and improve the efficiency of marking.

D. QUALITATIVE COMPARISON AND ANALYSIS

Fig. 6 presents the results of liver detection for three abdominal images by various algorithms, in which the red border corresponds to the results of the proposed model. Due to space limitations, only three representative images are presented in this paper. Between Fig. 6 (a) and Fig. 6 (b), the morphology of the liver differs substantially, especially when the liver is adjacent to the adjacent organs, and the grayscale is similar. According to the maximum analysis of the response graph, due to these appearance changes, it is not possible to identify the exact boundary; hence, the loss function cannot converge and the detection difference is large. Therefore, the ConvNet, SSD, DenseNet and ResNet networks inaccurately locate the targets. The results of the proposed model demonstrate that deformation, low contrast and other disturbances can be overcome by contour

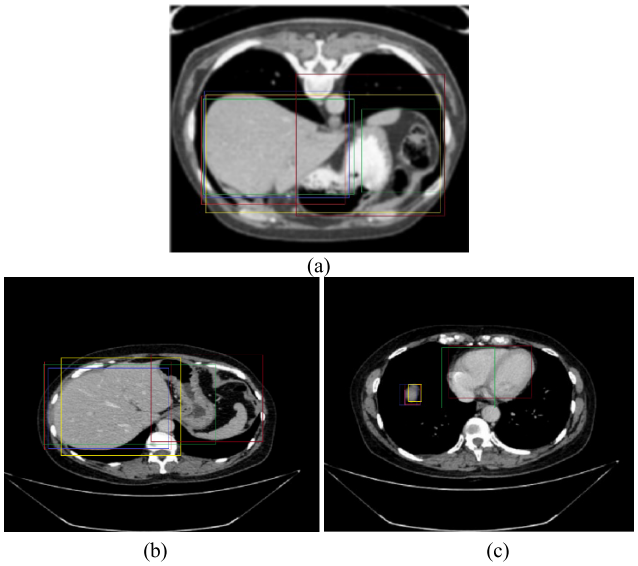

FIGURE 6. Results of comparison algorithms on three abdominal liver images. The red bounding boxes correspond to the results of our proposed model; the yellow bounding boxes correspond to the results of SDD; the orange bounding boxes correspond to the results of DenseNet; the green bounding boxes correspond to the results of ResNet; the purple bounding boxes correspond to the results of ConvNet; and the blue bounding boxes correspond to the results of YOLO-v2.

sensing networks. Yolo V2 is a lightweight network structure that is based on an improved VGG network. Although it can balance the robustness and speed, it remains vulnerable to background drying, thereby resulting in deviation of detection center. In Figure 6 (c), the boundary of the target is not readily identifiable. According to the test results, the ConvNet detection box has deviated from the target center. The model that is proposed in this paper combines the advantages of a contour detection network and a target detection network so that the algorithm can better adapt to shape changes of the target in abdominal image detection.

V. CONCLUSION

This paper proposes an improved deep network that is based on edge perception, which improves the accuracy of liver contour detection via an edge perception fusion module and captures high-level semantic features of abdominal images using a multiscale pyramid pooling layer. Complementary features of edge-related features can effectively preserve the clear boundary of the liver, and from the combination of the auxiliary side output and the pyramidal pooling layer output, rich global context information can be extracted. Extensive qualitative and quantitative experimental results demonstrate that the proposed model can effectively improve the performance of the available liver detection and localization network.

REFERENCES

[1] J. Chen, X. Song, L. Nie, X. Wang, H. Zhang, and T.-S. Chua, "Micro tells macro: Predicting the popularity of micro-videos via a transductive model," in *Proc. 24th ACM Multimedia Conf.*, 2016, pp. 898–907.

[2] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," Nov. 2016, *arXiv:1611.08408*. [Online]. Available: https://arxiv.org/abs/1611.08408

[3] J. Wang, Z. Wang, D. Tao, S. See, G. Wang, "Learning common and specific features for RGB-D semantic segmentation with deconvolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 664–679.
[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[5] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.
[6] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," vol. 11, no. 6, pp. 947–956, May 2015, *arXiv:1505.07293*. [Online]. Available: https://arxiv.org/abs/1505.07293
[7] P. Qian, J. Zhou, Y. Jiang, F. Liang, K. Zhao, S. Wang, K. Su, and R. Muzic, "Multi-view maximum entropy clustering by jointly leveraging inter-view collaborations and intra-view-weighted attributes," *IEEE Access*, vol. 6, pp. 28594–28610, 2018.
[8] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Washington, DC, USA: IEEE Computer Society, Jun. 2014, pp. 1725–1732.
[9] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.* Washington, DC, USA: IEEE Computer Society, Dec. 2015, pp. 4489–4497.
[10] G. Varol, I. Laptev, and C. Schmid, "Long-term temporal convolutions for action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1510–1517, Jun. 2018.
[11] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Conf. Neural Inf. Process. Syst.* New York, NJ, USA: Curran Associates, 2014, pp. 568–576.
[12] J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderic, "Beyond short snippets: Deep networks for video classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Washington, DC, USA: IEEE Computer Society, Jun. 2015, pp. 4694–4702.
[13] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. Van Gool, "Temporal segment networks: Towards good practices for deep action recognition," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2016, pp. 22–36.
[14] l. wang, Y. Qiao, and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Washington, DC, USA: IEEE Computer Society, Jun. 2015, pp. 4305–4314.
[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Washington, DC, USA: IEEE Computer Society, Jun. 2016, pp. 2818–2826.
[16] Y. Jiang, D. Wu, Z. Deng, P. Qian, J. Wang, G. Wang, F.-L. Chung, K.-S. Choi, and S. Wang, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, Dec. 2017.
[17] J. Bano, S. A. Nicolau, and A. Hostettler, "Multiphase liver registration from geodesic distance maps and biomechanical modelling," in *Abdominal Imaging. Computation and Clinical Applications* (Lecture Notes in Computer Science), vol. 12, no. 12. 2013, pp. 761–775.
[18] Y. Jiang, Z. Deng, F.-L. Chung, G. Wang, P. Qian, K.-S. Choi, and S. Wang, "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 3–20, Feb. 2017.
[19] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, "Collaborative fuzzy clustering from multiple weighted views," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 688–701, Apr. 2015.
[20] W. Li, F. Jia, and Q. Hu, "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks," *J. Comput. Commun.*, vol. 3, no. 11, pp. 146–151, 2015.
[21] C. S. Reiner, M. A. Fischer, T. Hany, P. Stolzmann, D. Nanz, O. F. Donati, D. Weishaupt, G. K. von Schulthess, and H. Scheffel, "Molecular imaging of malignant tumor metabolism: Whole-body image fusion of DWI/CT vs. PET/CT," *Academic Radiol.*, vol. 18, no. 8, pp. 940–946, 2011.
[22] H. Spitzer, K. Kiwitz, and K. Amunts, "Improving cytoarchitectonic segmentation of human brain areas with self-supervised siamese networks," in *Proc. 21st Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Granada, Spain, 2018.
[23] F. J. Estrada and A. D. Jepson, "Quantitative evaluation of a novel image segmentation algorithm," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*. San Diego, CA, USA: IEEE, Jun. 2005, pp. 1132–1139.
[24] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," Jun. 2016, *arXiv:1606.04797*. [Online]. Available: https://arxiv.org/abs/1606.04797
[25] L. Yu, X. Yang, H. Chen, J. Qin, and P. A. Heng, "Volumetric convnets with mixed residual connections for automated prostate segmentation from 3D MR images," in *Proc. 21st AAAI Conf. Artif. Intell.*, 2017, pp. 2512–2531.
[26] S. Azizi, F. Imani, A. Tahmasebi, B. Wood, P. Mousavi, and P. Abolmaesumi, "Detection of prostate cancer using temporal sequences of ultrasound data: A large clinical feasibility study," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 6, pp. 947–956, 2016.
[27] Y. Qi, Y.-Z. Song, H. Zhang, and J. Liu, "Sketch-based image retrieval via Siamese convolutional neural network," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 68–79.
[28] D. Chung, K. Tahboub, and E. J. Delp, "A two stream siamese convolutional neural network for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sep. 2017, pp. 659–668.
[29] K.-J. Xia, H.-S. Yin, and Y.-D. Zhang, "Deep semantic segmentation of kidney and space-occupying lesion area based on SCNN and ResNet models combined with SIFT-flow algorithm," *J. Med. Syst.*, vol. 43, no. 1, p. 2, 2019.
[30] J. Hosang, R. Benenson, and B. Schiele, "A convnet for non-maximum suppression," in *Proc. German Conf. Pattern Recognit.*, 2016, pp. 192–204.
[31] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR).*, Jul. 2017, pp. 6517–6525.
[32] W. Liu, D. Anguelov, and D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2015, pp. 21–37.
[33] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–23.
[35] Y. Jiang, F.-L. Chung, H. Ishibuchi, Z. Deng, and S. Wang, "Multitask TSK fuzzy system modeling by mining intertask common hidden structure," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 534–547, Mar. 2015.
[36] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.

KAIJIAN XIA was born in Jiangsu, China, in 1983. He is currently a Senior Engineer with The affiliated Changshu Hospital of Soochow University and also an Associate Professor with Xuzhou Medical University, China. He has authored or coauthored more than 50 research articles in international/national journals. His current research interests include medical information, medical image processing, deep learning, transfer learning, and computational intelligence and their applications in smart medicine. He is currently a member of the Information Professional Committee of the Chinese Hospital Association. He is the Program Vice-Chair of the CyberLife 2019 Procedure Committee. He is an Associate Editor of *Journal of Medical Imaging and Health Informatics*. He is also a Lead Guest Editor for *IEEE/ACM Transactions on Computational Biology and Bioinformatics* and *Journal of Medical Systems*.

HONGSHENG YIN was born in March 1967. He received the Ph.D. degree. He is currently a Professor. He was with the China University of Mining and Technology. He has been engaged in the teaching and research work of mine monitoring and monitoring, information processing, and computer network communication.