IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 16, 2020, accepted September 30, 2020, date of publication October 8, 2020, date of current version October 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029684

Contextual Level-Set Method for Breast Tumor Segmentation

SUMAIRA HUSSAIN[1,7], XIAOMING XI[2], INAM ULLAH[1], YONGJIAN WU[1], CHUNXIAO REN[3], ZHAO LIANZHENG[4], CUIHUAN TIAN[5,6], AND YILONG YIN[1], (Member, IEEE)

[1]School of Software Engineering, Shandong University, Jinan 250101, China
[2]School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China
[3]Shandong Province Science and Technology Exchange Center, Jinan 250101, China
[4]Center for Big Data of Shandong Province, Jinan 250101, China
[5]School of Medicine, Shandong University, Jinan 250100, China
[6]Health Management Center, Qilu Hospital of Shandong University, Jinan 250000, China
[7]Department of Computer Science, Sindh Madressatul Islam University, Karachi 74000, Pakistan

Corresponding authors: Xiaoming Xi (fyzq10@126.com) and Cuihuan Tian (shandamla@163.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61701280, Grant 61801263, Grant 61703235, and Grant 61701281; in part by the National Key Research and Development Program of China under Grant 2018YFC0830100; and Grant 2018YFC0830102; in part by the Natural Science Foundation of Shandong Province under Grant ZR2018BF012; and in part by the Foundation of Distinguished Associate Professor in Shandong Jianzhu University.

ABSTRACT Breast ultrasound image segmentation is the foundation of the diagnosis and treatment of breast cancer. The level set method is widely used for medical image segmentation. However, it remained a challenge for traditional level set methods because they cannot fully understand the tumor regions with complex characteristics by only low-level features. Considering that contextual features can provide complementary discriminative information to low-level features, this paper proposed a contextual level set method for breast tumor segmentation. Firstly, an encoder-decoder architecture network such as UNet is developed to learn high-level contextual features with semantic information. After that, the contextual level set method has been proposed to incorporate the novel contextual energy term. The proposed term has the ability to embed the high-level contextual knowledge into the level set framework. The learned contextual features with semantic information can provide more discriminative information, which has been directly associated with category labels, instead of the original intensity. Therefore, it is robust to serious intensity inhomogeneity, which is helpful to improve segmentation performance. The experiments had taken place with the help of three databases, which indicates that the proposed method outperformed traditional methods.

INDEX TERMS Breast ultrasound images, contextual feature, level-set method, tumor segmentation.

I. INTRODUCTION

As one of the second-highest causes of death among women, breast cancer has been marked as one of the fatal diseases that exist.

IARC statistics quote that worldwide nearly 25% of patients are of breast cancer, among all other types of cancer patients. This tumor is quite severe in terms of its diagnosis, and women majorly experience this medical failure globally [1]. It has been proven in multiple campaigns that early detection of breast cancer can increase the probability of survival among patients. For early diagnosis, medical imaging plays an important role. The images from ultrasound are commonly used to identify and classify mass abnormalities rather than other methods like mammography [2]. This methodology has proved to be accurate, fast, and economical. Hence, it has been recommended to use the ultrasound technique to monitor any early breast cancer signs.

Automatic tumor segmentation is the foundation of the diagnosis and treatment of breast cancer. Many methods have been proposed for this task.Compared with other methods, the level set framework has some advantages, such as the ability to represent contours/surfaces with complex topology and naturally changing their topology. Therefore, it has been widely used for medical image segmentation.

However, the challenge arises for traditional level set methods to segment complex ultrasound images. The traditional level-set methods detect edges according to the

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

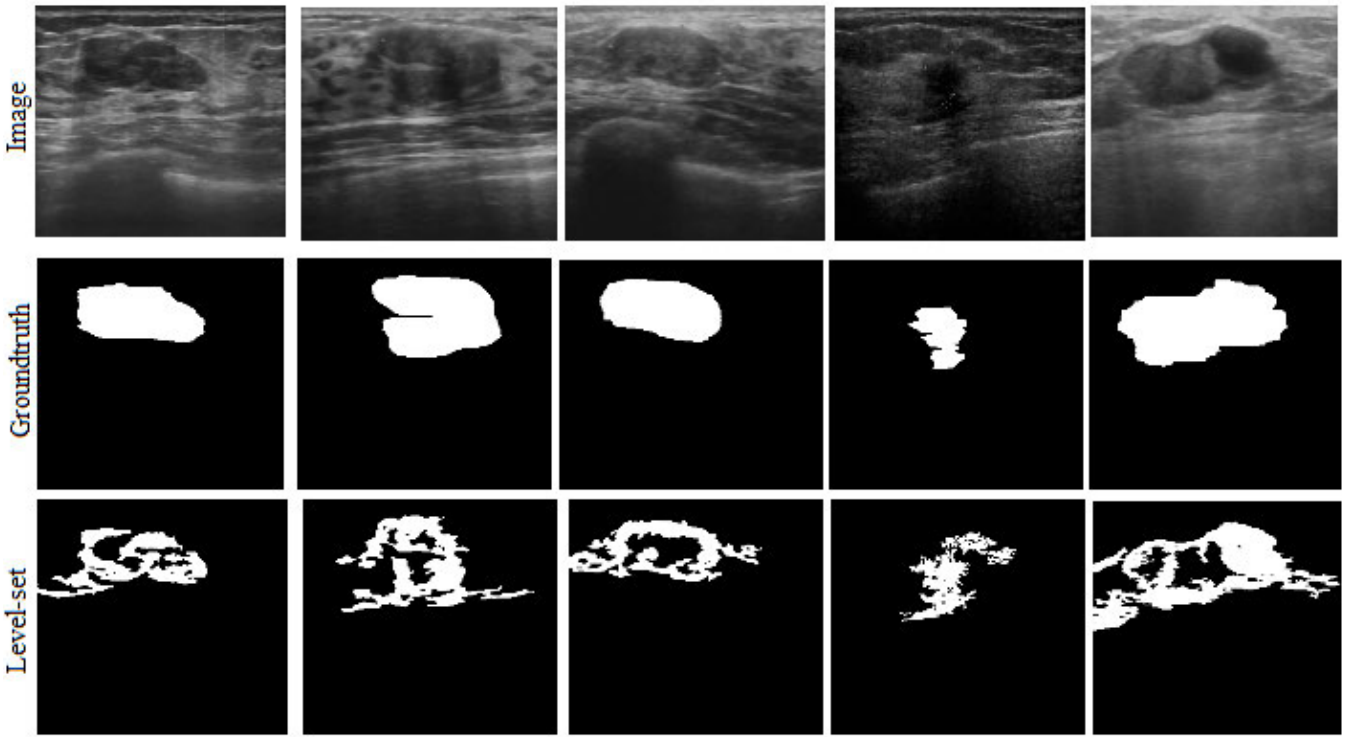


FIGURE 1. Segmentation samples of Level-set.

low-level features. These features include texture, intensity, and many other relevant properties. However, a problem does occur with such traditional methods, i.e., a lack of distinction in understanding tumor regions because of intensity inhomogeneity. This usually arises because low-level features cannot distinguish between tumors and their relative non-homogenous surroundings due to the unavailability of sufficient discriminative information.

Fig.1 gives several segmentation examples of level-set [3]; it can be seen that level-set fails to generate smooth segmentation because some tumor pixels are segmented as background incorrectly in the presence of severe intensity inhomogeneity.

Considering that contextual features provide complementary discriminative information to low-level features [4], [5], a novel contextual level-set method has been proposed for breast tumor segmentation.

This proposition involves two phases: (1) High-level contextual feature learning. An encoder-decoder network is developed so that contextual features could be understood clearly and easily. We have chosen this because, on the one hand, the encoder extracts multi-level features smoothly. On the other hand, un-sampling is performed in gradual steps by the decoder to restore resolutions. In addition to this, the decoder also integrates multi-level features to form contextual features. An upgraded level of data could be collected from contextual features, which has been associated with category labels, instead of the original intensity [6]. Therefore, the contextual feature is insensitive to severe intensity inhomogeneity.

(2) Segmentation with the incorporation of high-level contextual information. A contextual energy term has also been introduced in the level-set formulation, which is a different approach from conventional level-set methods. This approach contains a distinct ability to integrate high-level contextual knowledge to our level-set formulations. The high-level contextual knowledge that is generated responds strongly to intensity inhomogeneity. Hence, noisy labels have smoothened that cater to improved performance of the overall segmentation. We conducted the experiments on three datasets consisting of one private and two public, containing 186, 163, and 780 images, respectively, of breast ultrasound, compiled from multiple ultrasonic devices.

II. RELATED WORK

This section aims to discuss the previous work that has already been done in this particular area. Popularly, the level-set segmentation method qualifies to be robust in the presence of noise and can simplify complex topological changes. A common problem that arises in applications of medical image segmentation is that of intensity inhomogeneity. To handle such an issue, Wang *et al.* [7] presented a level set method that analyzed intensities in the local region. The level-set method had its foundations dependent on multi-scale segmentation.

Balla-Arabé *et al.* [8] used FCM and LBM solver to resolve the issue of intensity inhomogeneity. Dong *et al.* [9] and Wang *et al.* [10] had also put forward a methodology to sync intensities from local and global sources to cope with the

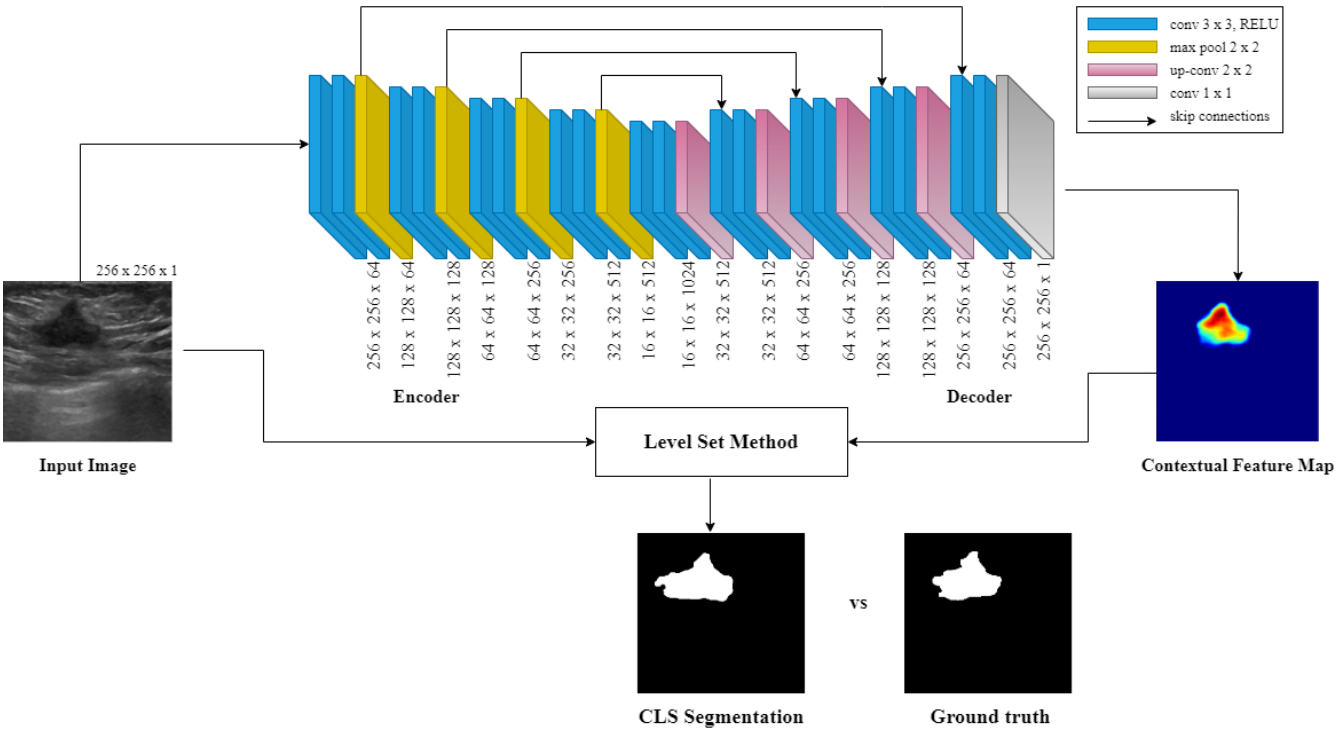


FIGURE 2. **Contextual Level-set Method framework.**

issues arising from non-homogenous behavior of intensities. The local source was responsible for directing intensities to the non-uniform area while the global source performs a contrary function for uniform intensities. Bui *et al.* [11] have used Random decision forests to obtain a pre-divided classification of images to start processing the level-set method. This will refine the segmentation results for lymph node ultrasound images to tackle the problem of intensity inhomogeneity of images and poor-quality contrasts.

Li. *et al.* had also put forward that with the integration of spatial fuzzy clustering, the level-set method could achieve more vigorous segmentation results [12].Zhang. *et al.* also points towards the fact that each inhomogeneous object needs to be treated by Gaussian distribution having variant mean and variances. Furthermore, for each local region, these would form statistical energies, respectively. The energy consists of level-set function, bias field, and true signal of the application object, represented by a constant [13], [14]. Huang *et al.* [15] has forwarded a level-set method, based on edges and regions using information from a local source of images. Li. *et al.* introduced a local clustering assumption and bias field estimation in the level set method. This was [3] to improvise the overall performance of the done segmentation. Using the local clustering assumption [3], Ivanovska *et al.* [16] classified the level-set function with the convexification method's help. The energy levels were lowered down using the Split Bregman method to achieve the global minimum.

These methods' specialty revolves around the fact that they can deal with intensity inhomogeneity problems via manually designed prior. In contrast, cases of complex imaging cause irregular distribution. This becomes a rather difficult situation to grasp information effectively, even with manual prior learning. Let us consider an example: In the real world, complex images might not demonstrate the local clustering assumption because the intensity variation could occur in the local region for complex images. Xi *et al.* [17] proposed a prior learning model and embedded the learned model into the level set. However, prior learning is based on the low-level features that result in limited performance improvement.

III. PROPOSED METHOD

As observed in Fig. 2, the general architecture for the proposed contextual level set method has been shown. By seeing the figure, it could be seen that our proposed method mainly contains two stages. This method's initial step is to train neural networks based on the encoder-decoder model [18]. The model utilizes UNet as a building block to gain knowledge about intrinsic features of breast tissues to generate a contextual feature map.The encoder-decoder model has been used because of its better performance than other methods to restore the contextual feature map. Tracking pixels that respond to tumors ultimately leads to an improvement in the process of segmentation.

The second phase proceeds based on the proposed level-set method, used for segmentation. Combining our segmentation method and the encoder-decoder model for feature map generation removes user involvement in tumor identification. On the other hand, it provides an automated approach for

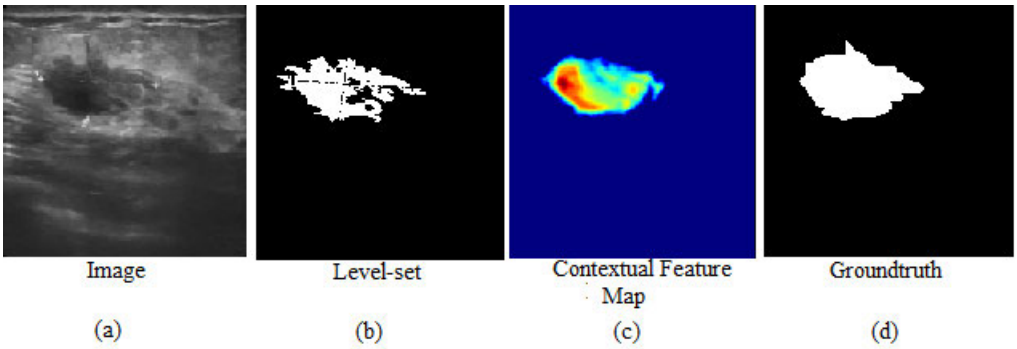


FIGURE 3. Visualization of contextual feature map.

breast tumor segmentation. The function of energy for this contextual level-set model has been established to incorporate the learned contextual features that are robust to intensity inhomogeneity. The high-level contextual features from the encoder-decoder network reserve a significant role in evolving level-set function, resulting in more accurate segmentation results.

A. HIGH-LEVEL CONTEXTUAL FEATURES LEARNING

During the current times, it has been observed that deep learning methodologies are forming its roots in core areas that integrate computer visuals. Effective contextual features' visual results have better performance than those obtained by conventional methods [4], [19]–[22]. These modern methods represent the encoder-decoder as an integral network architecture to learn pixel-wise contextual features [4], [19]–[22]. Therefore, in this paper, we have developed an encoder-decoder architecture for contexture feature learning.

The encoder-decoder network is comprised of two main components; encoder and decoder (as seen in Fig. 2). The encoder aims to learn significant features using the first 13 convolutional layers of the VGG-16 [23]. The kernel size for convolution layers and deconvolution layers is set to 3 x 3 following VGG-16 architecture because of image recognition. Kernel size for both convolution layers and deconvolution layers must be kept same for assuring the network input and network output to bear similar results.

Each convolutional layer contains a 3×3 kernel and generates feature maps via convolution operation. Then these feature maps are further processed by an element-wise rectified-linear Unit (ReLU), which serves as an activation (non-linear) function to extract features that exhibit non-linear changes for the given input. After that, the max-pooling layer reduces the feature map sizes and retain robust rotation information. There are five max-pool layers that have a 2×2 window-size along with a stride-size of 2. In the encoder, convolutional layers extract the local details, while the max-pooling layers extract global features. The spatial resolution of the input image reduces to 32 times after the extraction of features.

The decoder is developed using un-pooling, deconvolution operations, and skip-layers to obtain the same resolution level from the input image. The un-pooling is done to expand the view of the resolution of feature maps. The deconvolution layers perform a related operation as the encoder. The skip layers are responsible for forwarding each pooling layer's feature-maps at the encoder side to link it with relative feature maps at the decoder side. This is done to restore the spatial details that have been lost during max pooling.

At the end of this process, a 1×1 convolution layer generates the contextual feature map following U-Net [20] and Google Inception [24] in the network's decoder side, compared to the traditional fully connected layer. The 1 x 1 convolution layer contains the channels of an input image (features) into a single channel. The use of the convolution layer at the end reduces the prediction time during testing. Another reason to choose a 1 x 1 convolution layer is its efficient feature pooling and high-performance computational operations, with the ability to deplete in dimensions.

A visualization of the generated contextual feature map can be observed in Fig. 3, following serious intensity inhomogeneity. Nevertheless, the learned contextual feature is independent of intensity. It can distinguish them from the background, which contributes to the improvement of performance ultimately.

B. CONTEXTUAL LEVEL-SET METHOD (CLS)

In level-set framework, the segmentation problem could be categorized as an advancement of the contour, represented by zero level-set function (x,y,t). Zero level contour is responsible for the division of domains of an image into two broad regions: $\phi(x, y, t) > 0$ for pixel (x, y) inside the contour and $\phi(x, y, t) < 0$ for pixel (x, y) outside the contour. With the determination of level-set function catering minimal energy function, image segmentation's objective could be achieved [25], [26].

To embed the high-level contextual information into the model of level-set, a novel energy function F is given as follows:

$$F = \alpha + \chi + C \tag{1}$$

It is composed of three terms: α represents an intensity fitting term that calculates contour for tumor boundary with the evaluation of surrounding intensities. χ defines a contextual energy term that ensures the final segmentation results are

closer to learned high-level context feature maps. It further improves segmentation performance by using high-level contextual features, which will make it more robust for inhomogeneous intensity. C represents the regularization term.

Li *et al.* [3] proposed that the intensity clustering and bias field property can improve the ultrasound image's segmentation results. With this inspiration; the introduction of bias field (b) and the original image (I) can be written as:

$$I = bJ + n \tag{2}$$

Here I represents the observed image, b is the bias field, J means the true image, and n is additive noise. Regarding Eq. (2), our intensity fitting term for pixel (x, y) can be written as:

$$\alpha = \int_{\Omega} (I(x, y) - b(x, y)J(x, y))^2 \partial x \partial y \tag{3}$$

In the above equation, Ω refers to the image domain, and I: $\Omega \longrightarrow$ R denotes image in the gray level. Here, we follow the assumptions utilized in the work of Li *et al.* [3], to integrate the intensity clustering property of local intensity due to the bias field into our segmentation model. (a)The bias field b gradually changes. Now we understand that at any point d in a circularly surrounded region, radius r is centered at point e.

Hence, e ϵ Ω can be defined as: $O_e \triangleq \{d : |d-e| \leqslant r\}$ The bias field b values for all points d in a circular surrounding region O_e are almost similar at any point of e. Hence b(d) $\approx$ b(e). (b) A true image J for N disjoint regions $\Omega_1, \ldots, \Omega_N$ takes N distinct constant values $c_1, \ldots, c_N$, respectively. In light of the above statements, we can rewrite Eq. (3) as:

$$\alpha = \int (\sum_{i=1}^{N} \int K(e-d)(I(d) - b(e)c_i)^2 \partial d)\partial e \tag{4}$$

where Kernel function K(e-d)has represented the Gaussian truncated function defined as K(e-d) = 1 if point d is not a neighbor of point e else K(e-d) = 0.

Thus, the level set function ϕ, the vector c, and the bias field b show intensity variables denoted by α, which can, therefore, be represented as $\alpha(\phi, c, b)$. From (4), the intensity fitting terms could be written as:

$$\alpha(\phi, c, b) = \int \sum_{i=1}^{N} P_i(d) M_i(\phi(d))\partial d \tag{5}$$

where P is the function defined as,

$$P_i(d) = \int (K(e-d)(I(d) - b(e)c_i)^2)\partial e \tag{6}$$

The energy χ for contextual energy term can be given as:

$$\chi = \sum_{i=1}^{N} \int_{\Omega_i} (P(d) - c_i)^2 \partial d \tag{7}$$

Adding contextual feature map to $P_i(d)$ results into:

$$P_i(d) = I^2 1_K - 2c_i I(b * K) + c_i^2 (b^2 * K) + c_i^2 - 2c_i I_c + I_c^2 \tag{8}$$

$$P_i(d) = I^2 1_K - 2c_i [I(b * K) + I_c] + (b^2 * K + 1)c_i^2 \tag{9}$$

In the above equation, P(d) denotes the probability of pixel d associated with the tumor, according to contextual feature learner. The term χ makes sure that the achieved segmentation results are closer to the obtained contextual feature with semantic information. This is not dependent on intensity. Hence, the proposed contextual level-set model acts strongly to the inhomogeneous behavior of intensities. The energy in the level-set formulation can, therefore, be rewritten as:

$$F(\phi, c, b) = \alpha(\phi, c, b) + \chi(\phi, c) + \Upsilon(\phi) + \vartheta(\phi) \tag{10}$$

where,

$$\alpha(\phi, c, b) = \int \sum_{i=1}^{N} P_i(d) M_i(\phi(d))\partial d \tag{11}$$

$$\chi(\phi, c) = \sum_{i=1}^{N} \int (P(d) - c_i)^2 M_i(\phi(d))\partial d \tag{12}$$

$$\Upsilon(\phi) = \int |\bigtriangledown H(\phi(d))| \partial d \tag{13}$$

$$\vartheta(\phi) = \int \rho(|\bigtriangledown(\phi(d))|)\partial d \tag{14}$$

ϕ denotes the level-set function. The level-set function splits the image domain into two disjoint regions i.e. $\Omega_1 = \{\phi(d) > 0\}$ and $\Omega_2 = \{\phi(d) < 0\}$ that represents the tumor itself and the relative background in an image. These regions are characterized by their membership functions respectively i.e. $M_1(\phi) = H(\phi)$ and $M_2(\phi) = 1 - H(\phi)$ where H is Heaviside function [27]. In Eq. (13), $\Upsilon(\phi)$ calculates the arc length for level-set (zero level contour), that is responsible to smoothen the curve [3]. In Eq.(14), $\vartheta(\phi)$ is the distance regularization term proposed in DRLSE [28].

This term eliminates the necessity for costly re-initialization process. Segmentation results could be obtained by estimating the level-set function values through minimizing the energy, shown in Eq. (6). ϕ,c,b values are determined by using the same optimization process used by Li *et al.* [3] given in equation 15 and 16. The optimization process used to determine c is given in equation 17.

$$\frac{\partial \phi}{\partial t} = -\delta(\phi)(p_1 - p_2) + \nu\delta(\phi) div(\frac{\nabla\phi}{\nabla\phi}) + \mu div(d_p(\nabla\phi)\nabla\phi) \tag{15}$$

where ∇ is the gradient operator, div(.) is the divergence operator, and the function d_p is defined as $d_p(s) \triangleq \frac{p'(s)}{s}$

$$b = \frac{(IJ^{(1)}) * K}{J^{(2)} * K} \tag{16}$$

where $J^{(1)} = \sum_{i=1}^{N} c_i u_i, J^2 = \sum_{i=1}^{N} c_i^2 u_i$ and $u_i(d) = M_i(\phi(d))$ Note that the convolutions with a kernel function K in (12) confirms the slowly varying property of the derived optimal estimator b of the bias field.

$$c_i = \frac{\int [(b * K)I + I_c] u_i \partial e}{\int (b^2 * K + 1) u_i \partial e} \tag{17}$$

where i = 1,...,N with $u_i(e) = M_i(\phi(e))$ The energy minimization can be achieved by the iterative optimization. In each iteration, we minimize the energy with respect to each of its variables ϕ,c,b provided the other two variables have been updated in the previous iteration

IV. EXPERIMENT

A. DATASET DESCRIPTION

To evaluate the reliability of our model, experiments have taken place on three datasets. Dataset I (Private) – QHSP: The dataset is constructed with the help of the Qianfoshan Hospital of Shandong Province. It consists of 186 breast ultrasound grayscale images with the tumor. Among them, 135 cases are from benign tumors, and the remaining 51 cases are from a malignant tumor. The images are taken from four different devices (ALOKA α 10, AplioXG, GE LOGIQ E7, and SIEMENS Sequoia 512) in the presence of technically skilled staff. The radiologists were responsible for delineating the tumors manually.

Dataset II (Public) – UDIAT: The dataset is constructed in the UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell, Spain [29]. The images are taken with a Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz). This sample size had 163 images with a mean image size of 760 × 570 pixels (with a nominal pixel size of 0.084mm). These images are further classified into 110 images with benign lesions and 53 with malignant lesions. The credible radiologists delineated the lesions.

Dataset III (Public) – BHEDTWC: This data was collected from Baheya Hospital for Early Detection and Treatment of Women's Cancer, Cairo (Egypt) in 2018 [30]. For the collection of this dataset, they took 600 female patients. It consisted of 780 images with a standardized image size of 500*500 pixels. The images have been classified into three classes; normal, benign, and malignant, with 133, 437, and 210 images respectively for each category. The scanning process took place in instruments such as the LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound system.

B. EXPERIMENTAL SETTING

In the experiment that has been done by us, all the images have been normalized and transformed into grayscale images in the pre-processing image phase. It had to be made sure that this process was performed before forwarding it to the neural network. We adopt the encoding-decoding structure (UNet) in our proposed model as a foundation network for contextual feature learning. The image dataset has been divided into two broad types of training data and testing data. The size of the dataset is small in the first two datasets. Therefore, data augmentation (in-place-data-augmentation) is performed by doing rotation, horizontal flips, shearing, zooming, and cropping. This is done to train our network on different perspectives and also able to handle noise. Hence, data augmentation creates batches with real-time data augmentation that train models on more than 91,000 transformed images for the given dataset to avoid overfitting. Here, Adam optimizer has been used. The learning rate has also been set to $1e^{-5}$. The computation platforms used include NVIDIA TITAN XP 12G and Intel Xeon Gold 5115 2.4G, respectively.

C. EVALUATION

The contextual level-set method results have been analyzed under six metrics to estimate our derived method's performance successfully. These metrics are Accuracy (Acc), Sensitivity or True Positive rate (TP), Specificity or True Negative rate (TN), Jaccard coefficient (Jaccard), F1-score, and Peak signal-to-noise ratio (PSNR):

$$Acc = \frac{|(T_m \cap T_l) \cup (B_m \cap B_l)|}{|T_m \cup B_m|} \tag{18}$$

$$Sensitivity(TP) = \frac{|T_m \cap T_l|}{|T_m|} \tag{19}$$

$$Specificity(TN) = \frac{|B_m \cup B_l|}{|B_m|} \tag{20}$$

$$Jacc = \frac{|T_m \cap T_l|}{|T_m \cup T_l|} \tag{21}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{22}$$

where

$$Precision = \frac{T_m \cup T_l}{|(T_m \cup T_l) \cap (T_m \cup B_l)} \tag{23}$$

$$Recall = \frac{T_m \cup T_l}{|(T_m \cup T_l) \cap (B_m \cup T_l)} \tag{24}$$

$$PSNR(O, S) = 10log_{10}(\frac{MAX^2}{MSE(O, S)}) \tag{25}$$

where

$$MSE(O, S) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (O_{ij} - S_{ij})^2 \tag{26}$$

Here T_m is the Tumor region observed by manual segmentation and T_l is the Tumor region observed by our method. The same can be represented for B_m and B_l where B represents the background of the image.

In the PSNR equation, O represents the image's ground truth, and S denotes the segmented image with both size m x n. MAX value represents the maximum value for the size of a given image.

D. OPTIMIZATION ANALYSIS

To ensure the effectiveness of the encoder-decoder network as an optimization method for CLS. Here we present the test that evaluates the Particle Swam optimization method's performance as a CLS optimization method. Particle swarm optimization (PSO) is a well-known bioinspired mechanism for its global optimization ability based on swarm intelligence. PSO, compared to the genetic algorithm (GA), can cluster more quickly and computationally cheap. PSO works on the mechanism of finding global optimum from the image pixels.

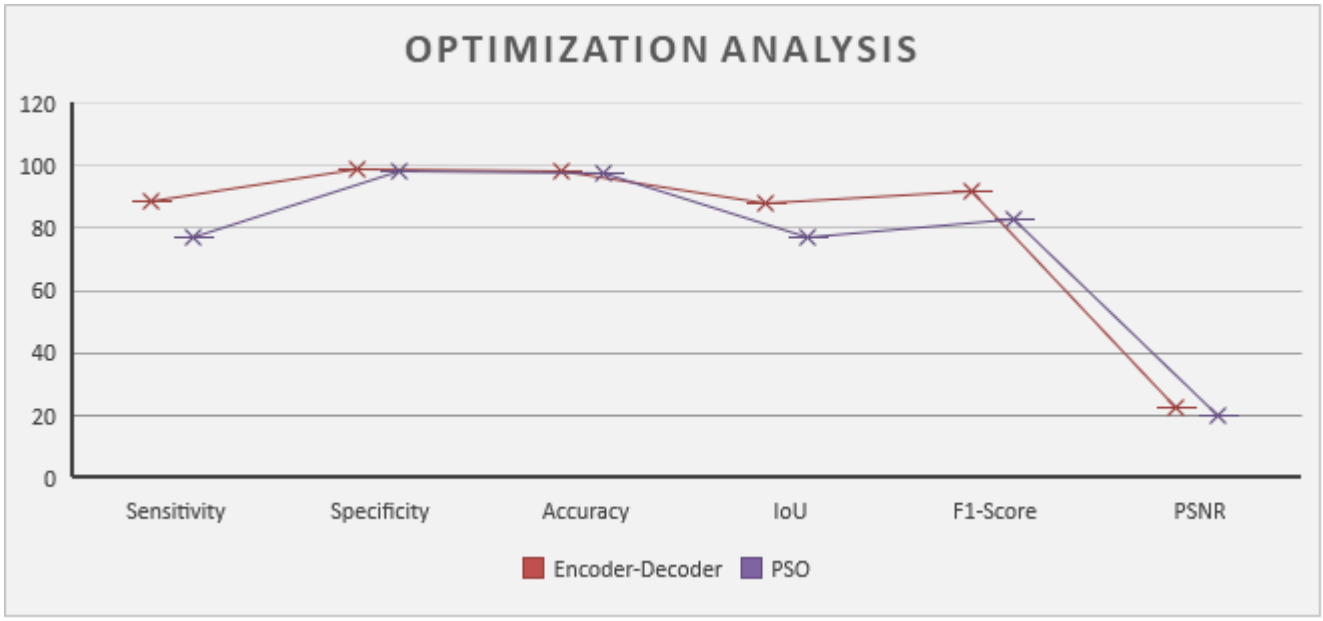


FIGURE 4. PSO vs Encoder-Decoder optimization comparison for CLS.

TABLE 1. Quantitative evaluation of CLS on multiple datasets.

	Dataset	Accuracy	IoU	F1-score	Sensitivity	Specificity	PSNR
Benign	QHSP	0.986	0.919	0.951	0.920	0.990	24.462
	UDIAT	0.999	0.968	0.983	0.985	0.981	32.010
	BHEDTWC	0.988	0.845	0.914	0.923	0.903	20.979
Malignant	QHSP	0.963	0.770	0.837	0.968	0.886	17.126
	UDIAT	0.929	0.659	0.725	0.749	0.833	20.844
	BHEDTWC	0.959	0.746	0.852	0.787	0.944	17.496

According to Fig. 4, encoder-decoder-based optimized CLS achieved better segmentation performance on every evaluation metric. It can be seen from the results that PSO optimization results are closer to accuracy and specificity; however, it shows lower performance on other quantitative measures. This signifies the proposed optimization strategy's ability to prevent incorrect segmentation.

E. EFFECTIVENESS OF CONTEXTUAL FEATURES

1) QUANTITATIVE EVALUATION

In this experiment, we try to validate the performance of CLS. For this purpose, we perform experiments on three different datasets: QHSP (Qianfoshan Hospital of Shandong Province), UDIAT (UDIAT Diagnostic Centre of the Parc Taulí Corporation), and BHEDTWC (Baheya Hospital for Early Detection and Treatment of Women's Cancer) on breast ultrasound tumor. We have used six different metrics to evaluate our proposed model's quantitative virtues to attain optimal results. The results depicted in Table 1. have been divided into two sections. One section has been specified for a benign tumor while the other has a malignant tumor.

With the comparison of results for all datasets, our proposed method has demonstrated adequate performance, in terms of both F1-score and IoU metrics on both private and public datasets. In the case of a private dataset, all the metrics have shown a significant increase for a malignant tumor. Moreover, UDIAT has demonstrated the highest performance on all metrics for a benign tumor. The overall performance differences in both benign and malignant tumors are minimal. Lastly, we notice from the results that our proposed method achieves a robust performance on both types of datasets; small and large.

2) QUALITATIVE EVALUATION

Fig.5 highlights some examples for comparing our proposed method with the level-set method, as observed visually. These examples show different intensity inhomogeneity images from benign and malignant tumors, where tumor and background regions overlap. These closely related regions can majorly affect the overall segmentation accuracy of the level-set.

The contextual feature map was generated with high-level semantic information, which is not dependent on intensity. In our general observation, neighbor pixels have a similar semantic label [6]. Hence, the contextual feature map produced could be regarded as the semantic label map, which

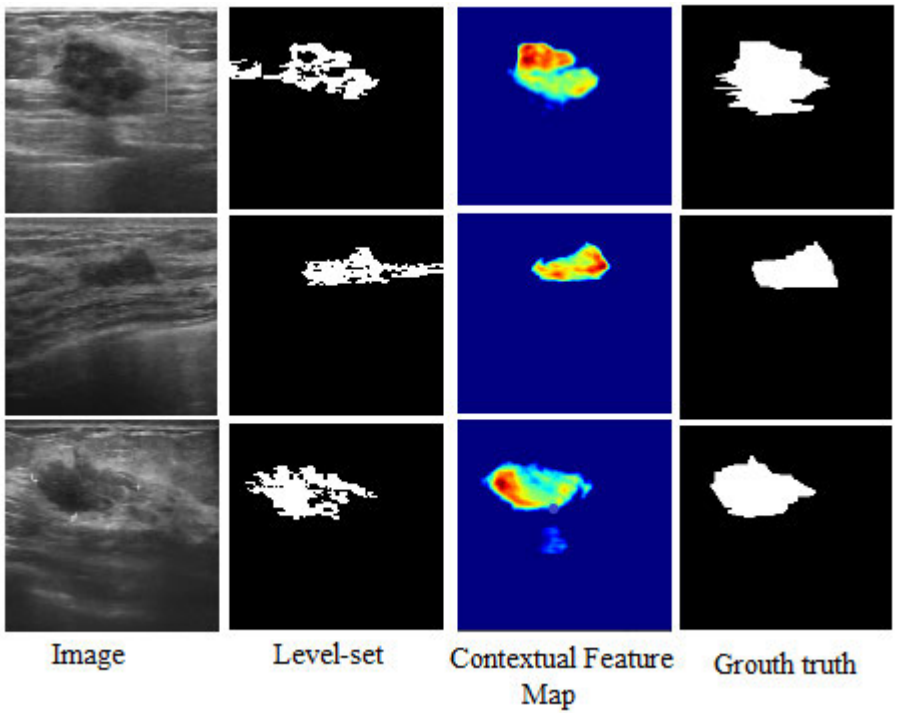


FIGURE 5. Several segmentation examples of CLS.

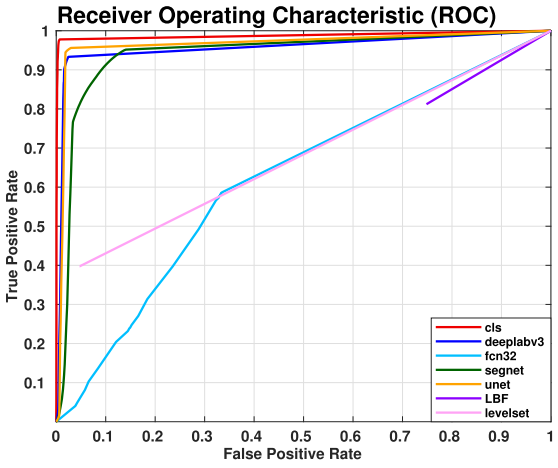


FIGURE 6. ROC curve for CLS and other benchmark methods.

can smoothen noisy labels caused by severe intensity inhomogeneity. Therefore, embedding the contextual information has been helpful to improve segmentation performance. This shows that our proposed model has consistently produced precise segmentation in non-homogenous regions, having sharp edges and precise details.

F. COMPARISON WITH DIFFERENT METHODS

In this experiment performed by us, the proposed method CLS against 12 highly commendable methods including: dynamic active contour namely Local Binary Fitting Model (LBF) [31], Level-set method [3], Prior Knowledge Learning-Level-set (PKL-LS) [17], learning-based methods(MLF1 [32], MLF2 [33]), region grow-based method(RG [34]) and deep learning-based methods, FCN-16 [35], FCN-32 [35], UNet [20], PSPNET [36], SegNet [18], DeepLab v3 [37]. A quantitative analysis is performed on malignant and benign tumors, as presented in Table. 2. The accuracy of our proposed method comes out to be 98% and 96% for benign and malignant tumors, respectively. Table. 2 has shown that CLS has achieved significant performance improvement compared to other methods.

To formulate the segmentation task into a binary classification task, learning-based methods are used commonly. However, it is challenging to train an accurate classifier for pixel classification because serious intensity inhomogeneity can manipulate data distribution. There is difficulty developing an effective growth rule for the region growing-based method to manage the intensity inhomogeneity problem. Although with PKL-LS, the prior learned knowledge proceeds to the level set, it is difficult for a shallow model to learn effectively about images with serious intensity inhomogeneity. The dynamic snake active contour model outperforms by breaking down the images with complex targets into several smaller targets [31]. The local binary fitting model [31] utilizes local image information to identify the targeted object. However, CLS has achieved more consistent results by taking advantage of contextual information, shown in Fig. 6. We can notice from the ROC curve in Figure 6, CLS gets a significant altitude, achieving a very robust performance and significantly overtaking other deep learning methods.

TABLE 2. Segmentation Results of different methods on Benign and Malignant Tumor.

	Methods	Accuracy	IoU	F1-score	Sensitivity	Specificity	PSNR
Benign	RG	0.682	0.157	0.362	0.449	0.392	7.712
	MLF1	0.461	0.076	0.316	0.447	0.009	3.004
	MLF2	0.671	0.120	0.425	0.564	0.052	5.432
	LBF	0.290	0.081	0.571	0.805	0.000	1.526
	PKL-LS	0.928	0.467	0.550	0.547	0.633	14.397
	Levelset	0.901	0.307	0.393	0.389	0.565	12.307
	FCN16	0.593	0.108	0.461	0.626	0.045	4.636
	FCN32	0.928	0.240	0.379	0.438	0.542	12.322
	U-Net	0.971	0.595	0.693	0.684	0.808	17.062
	PSPNET	0.947	0.376	0.517	0.537	0.669	13.992
	SegNet	0.909	0.550	0.553	0.709	0.175	10.859
	DeepLab v3	0.960	0.542	0.702	0.812	0.488	14.819
	CLS	**0.986**	**0.919**	**0.951**	**0.920**	**0.990**	**24.460**
Malignant	RG	0.743	0.340	0.519	0.572	0.554	8.633
	MLF1	0.574	0.182	0.440	0.608	0.013	3.989
	MLF2	0.623	0.280	0.540	0.691	0.192	5.391
	LBF	0.341	0.157	0.566	0.798	0.000	1.851
	PKL-LS	0.921	0.606	0.695	0.667	0.810	13.835
	Levelset	0.855	0.350	0.448	0.412	0.702	10.315
	FCN16	0.885	0.370	0.526	0.507	0.650	10.070
	FCN32	0.889	0.550	0.694	0.732	0.619	11.233
	U-Net	0.934	0.753	0.825	0.962	0.739	15.020
	PSPNET	0.931	0.550	0.859	0.926	0.773	14.943
	SegNet	0.877	0.550	0.709	0.904	0.344	10.205
	DeepLab v3	0.936	0.764	0.826	0.925	0.780	15.228
	CLS	**0.963**	**0.770**	**0.837**	**0.968**	**0.886**	**17.126**

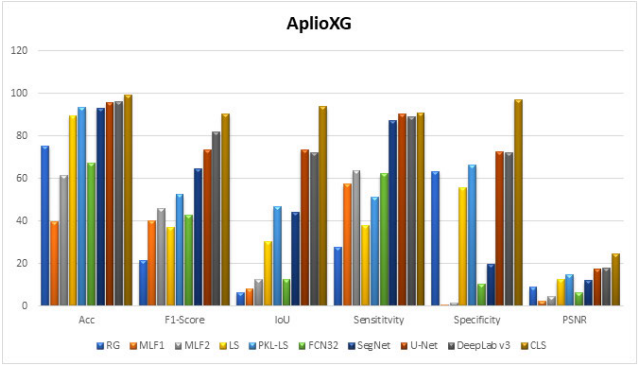


FIGURE 7. Segmentation results of different methods on device AplioXG.

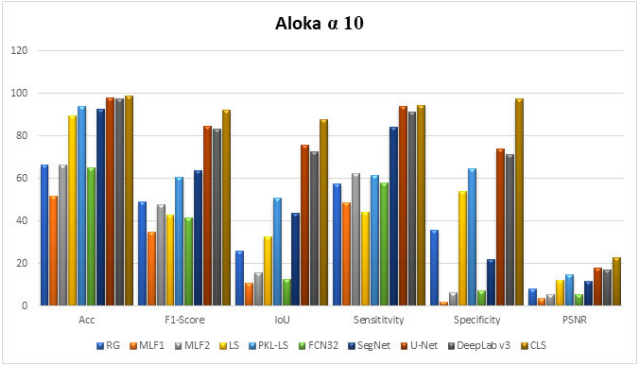


FIGURE 8. Segmentation results of different methods on device ALOKA α 10.

G. ROBUSTNESS OF CLS

To understand our proposed model's robustness, we further test our method on four types of ultrasound machines. Fig. 7 to Fig. 10 show segmentation results of different

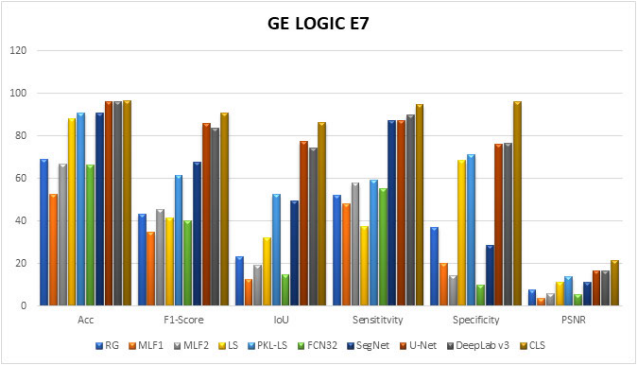


FIGURE 9. Segmentation results of different methods on device GE LOGIC E7.

methods for a dataset, collected from sources of multiple ultrasound machines. The quantitative results indicate clearly that CLS results perform better than other conventional methods. The characteristics of captured images from various devices are different due to imaging factors such as individual devices' parameters. As shown in these figures, the conventional methods are sensitive to the device.

Consider an example: MLF1 can achieve better IoU and F1-score than RG on device AplioXG. However, RG outperforms MLF1 on the remaining devices.

Apart from conventional and deep learning techniques, methods experience some performance variation on different devices. Our proposed approach has excelled in performance in all the devices that have been used. Different devices with low-level variant feature such as intensity, texture, etc. affect the strength of obtained images in traditional methods.

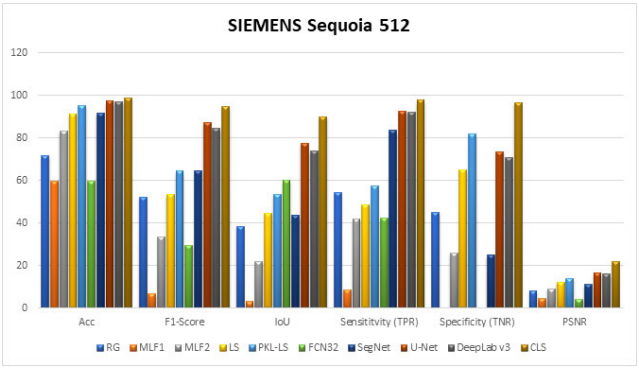


FIGURE 10. Segmentation results of different methods on device SIEMENS Sequoia 512.

However, the learned high-level contextual feature is independent of these low-level features, achieving robust segmentation results.

V. CONCLUSION

Apart from conventional level set methods that use low-level features, this paper proposed a contextual level set method that uses high level discriminative contextual features for breast tumor segmentation. Firstly, the encoder-decoder has been developed to learn contextual features with semantic information. Then, a novel level set has been designed with the introduction of contextual energy term. This term aims to embed the high-level contextual knowledge into the level set framework. The contextual features provide high-level semantic information, which has been related to category labels, rather than original intensity. Therefore, it is robust to intensity inhomogeneity, which is helpful to improve segmentation performance. We also conducted experiments on both public and private databases, and the experiment results prove that the proposed method outperformed traditional methodologies.

We also focus on extending the proposed framework on other medical segmentation applications such as MRI, CT, etc.

ACKNOWLEDGMENT

The authors would like to thank all the editors and reviewers for their valuable time, comments, and suggestions.

REFERENCES

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

[2] A. Jalalian, S. B. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: A review," *Clin. Imag.*, vol. 37, no. 3, pp. 420–426, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0899707112002938

[3] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2007–2016, Jul. 2011.

[4] Z. Wang, L. Xie, and J. Qi, "Dynamic pixel-wise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI," *Magn. Reson. Imag.*, vol. 66, pp. 131–140, Feb. 2020.

[5] D. Mahapatra, "Graph cut based automatic prostate segmentation using learned semantic information," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 1316–1319.

[6] Q. Wang, D. Wu, L. Lu, M. Liu, K. L. Boyer, and S. K. Zhou, "Semantic context forests for learning-based knee cartilage segmentation in 3D MR images," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.* Cham, Switzerland: Springer, Sep. 2013, pp. 105–115.

[7] X.-F. Wang, H. Min, and Y.-G. Zhang, "Multi-scale local region based level set method for image segmentation in the presence of intensity inhomogeneity," *Neurocomputing*, vol. 151, pp. 1086–1098, Mar. 2015.

[8] S. Balla-Arabe, X. Gao, and B. Wang, "A fast and robust level set method for image segmentation using fuzzy clustering and lattice Boltzmann method," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 910–920, Jun. 2013.

[9] F. Dong, Z. Chen, and J. Wang, "A new level set method for inhomogeneous image segmentation," *Image Vis. Comput.*, vol. 31, no. 10, pp. 809–822, Oct. 2013.

[10] X.-F. Wang, H. Min, L. Zou, and Y.-G. Zhang, "A novel level set method for image segmentation by incorporating local statistical analysis and global similarity measurement," *Pattern Recognit.*, vol. 48, no. 1, pp. 189–204, Jan. 2015.

[11] T. M. Bui, A. Coron, L. Bridal, J. Mamou, E. J. Feleppa, E. Saegusa-Beecroft, and J. Machi, "Random forest classification and local region-based, level-set segmentation for quantitative ultrasound of human lymph nodes," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2015, pp. 1–4.

[12] B. N. Li, C. K. Chui, S. Chang, and S. H. Ong, "Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation," *Comput. Biol. Med.*, vol. 41, no. 1, pp. 1–10, Jan. 2011.

[13] K. Zhang, L. Zhang, K.-M. Lam, and D. Zhang, "A local active contour model for image segmentation with intensity inhomogeneity," 2013, *arXiv:1305.7053*. [Online]. Available: http://arxiv.org/abs/1305.7053

[14] K. Zhang, L. Zhang, K.-M. Lam, and D. Zhang, "A level set approach to image segmentation with intensity inhomogeneity," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 546–557, Feb. 2016.

[15] J. Huang, F. Jian, H. Wu, and H. Li, "An improved level set method for vertebra CT image segmentation," *Biomed. Eng. OnLine*, vol. 12, no. 1, p. 48, 2013.

[16] T. Ivanovska, R. Laqua, L. Wang, A. Schenk, J. H. Yoon, K. Hegenscheid, H. Völzke, and V. Liebscher, "An efficient level set method for simultaneous intensity inhomogeneity correction and segmentation of MR images," *Comput. Med. Imag. Graph.*, vol. 48, pp. 9–20, Mar. 2016.

[17] X. Xi, H. Shi, L. Han, T. Wang, H. Y. Ding, G. Zhang, Y. Tang, and Y. Yin, "Breast tumor segmentation with prior knowledge learning," *Neurocomputing*, vol. 237, pp. 145–157, May 2017.

[18] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[19] J. Liu, H. Liu, Z. Tang, W. Gui, T. Ma, S. Gong, Q. Gao, Y. Xie, and J. P. Niyoyita, "IOUC-3DSFCNN: Segmentation of brain tumors via IOU constraint 3D symmetric full convolution network with multimodal auto-context," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, Dec. 2020.

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

[21] W. Yu, K. Yang, H. Yao, X. Sun, and P. Xu, "Exploiting the complementary strengths of multi-layer CNN features for image retrieval," *Neurocomputing*, vol. 237, pp. 235–241, May 2017.

[22] S. K. Devalla, P. K. Renukanand, B. K. Sreedhar, G. Subramanian, L. Zhang, S. Perera, J.-M. Mari, K. S. Chin, T. A. Tun, and N. G. Strouthidis, "Drunet: A dilated-residual u-net deep learning network to segment optic nerve head tissues in optical coherence tomography images," *Biomed. Opt. Express*, vol. 9, no. 7, pp. 3244–3265, 2018.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

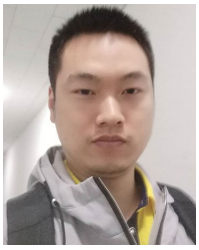
[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[25] N. Paragios and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *Int. J. Comput. Vis.*, vol. 46, no. 3, pp. 223–247, Feb. 2002.
[26] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia, "A variational model for image classification and restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 5, pp. 460–472, May 2000.
[27] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 271–293, Dec. 2002.
[28] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3243–3254, Dec. 2010.
[29] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018.
[30] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, Feb. 2020, Art. no. 104863. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352340919312181
[31] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding, "Implicit active contours driven by local binary fitting energy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–7.
[32] R. Rodrigues, A. Pinheiro, R. Braz, M. Pereira, and J. Moutinho, "Towards breast ultrasound image segmentation using multi-resolution pixel descriptors," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 2833–2836.
[33] B. Liu, H. D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu, "Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images," *Pattern Recognit.*, vol. 43, no. 1, pp. 280–298, Jan. 2010.
[34] J. Shan, H. D. Cheng, and Y. Wang, "A completely automatic segmentation method for breast ultrasound images using region growing," in *Proc. 11th Joint Conf. Inf. Sci. (JCIS)*, 2008, pp. 332–337.
[35] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[36] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.
[37] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

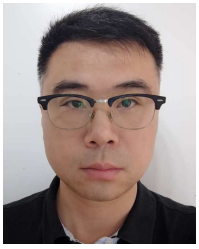
SUMAIRA HUSSAIN received the M.S. degree in computer science from Virtual University, Pakistan, in 2014. She is currently pursuing the Ph.D. degree in computer science with Shandong University, Jinan, Shandong, China. Her research interests include machine learning, medical image analysis, and computer vision.

XIAOMING XI is currently an Associate Professor with Shandong Jianzhu University. His research interests include machine learning, data mining, and medical image analysis.

INAM ULLAH received the bachelor's degree from the University of Peshawar, Pakistan, and the master's degree from International Islamic University Islamabad, Pakistan. He is currently pursuing the Ph.D. degree with the School of Software Engineering, Shandong University, Jinan, China. His current research interests include image processing, computer vision, machine learning, deep learning, and biometric recognition.

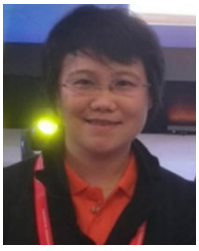


YONGJIAN WU received the master's degree from Shandong University. His research interests include machine learning and medical image analysis.



CHUNXIAO REN is currently an Associate Research Fellow with the Shandong Province Science and Technology Exchange Center. His research interests include machine learning, data mining, and medical image analysis.

ZHAO LIANZHENG is currently a Researcher with the Center for Big Data of Shandong Province. His research interest includes image processing.



CUIHUAN TIAN is currently a Professor with Shandong University. Her research interests include image analysis, data mining, and so on.

YILONG YIN (Member, IEEE) is currently the Director of the MLA Laboratory and a Professor with Shandong University. His research interests include machine learning, data mining, medical image analysis, and biometrics.