SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

February 2023, Vol. 66 122104:1–122104:15
https://doi.org/10.1007/s11432-021-3340-y

Difficulty-aware prior-guided hierarchical network for adaptive segmentation of breast tumors

Sumaira HUSSAIN[1,2], Xiaoming XI[3*], Inam ULLAH[1], Syeda Wajiha NAIM[2], Kashif SHAHEED[4], Cuihuan TIAN[5,6*] & Yilong YIN[1*]

[1]*School of Software Engineering, Shandong University, Jinan* 250101*, China;*
[2]*Department of Computer Science, Sindh Madressatul Islam University, Karachi* 74000*, Pakistan;*
[3]*School of Computer Science and Technology, Shandong Jianzhu University, Jinan* 250101*, China;*
[4]*School of Computer Science and Engineering, South China University of Technology, Guangzhou* 510006*, China;*
[5]*School of Medicine, Shandong University, Jinan* 250100*, China;*
[6]*Health Management Center, Qilu Hospital of Shandong University, Jinan* 250000*, China*

Received 26 March 2021/Revised 1 July 2021/Accepted 1 September 2021/Published online 5 January 2023

Abstract Breast tumor segmentation is vital to tumor detection at the early stages. Deep learning methods are typically used in automatic tumor segmentation tasks. However, in existing methods, the difference between pixels is disregarded, and the union network architecture is used to segment all pixels; these methods involve a tradeoff between accuracy and efficiency. A novel, difficulty-aware, prior-guided hierarchical network for the adaptive segmentation of breast tumors is presented herein. A difficulty prior learning module is proposed to learn the pixel's difficulty prior to guild adaptive segmentation in the proposed network. To achieve a more accurate segmentation of hard pixels, a hard pixel processing unit is presented to learn more discriminative features for hard pixels. Experiments are conducted based on three datasets. The experimental results show that the proposed methods outperform traditional deep learning methods and achieve a balance between accuracy and efficiency.

Keywords breast tumor, ultrasound image segmentation, deep neural network, difficulty-awareness

Citation Hussain S, Xi X M, Ullah I, et al. Difficulty-aware prior-guided hierarchical network for adaptive segmentation of breast tumors. Sci China Inf Sci, 2023, 66(2): 122104, https://doi.org/10.1007/s11432-021-3340-y

1 Introduction

Breast cancer is the second-highest cause of death among women worldwide [1]. It is reported that in 2020, approximately one in eight women (approximately 12%) will likely be diagnosed with breast cancer. Early detection and classification of tumors can save lives [2,3].

Ultrasonography is a well-known tool to diagnose breast cancer. Compared with mammography, it offers advantages such as usability, inexpensiveness, and non-ionizing nature [4,5].

Breast tumor segmentation is vital to the diagnosis of breast cancer. Recently, deep learning has been widely used in medical image segmentation [6–8]. Typical deep networks [9–13] are constructed based on the encoder-decoder architecture. An encoder is used for feature learning, which is significant for performance improvement.

However, the pixel characteristics in an image are different, and existing models process all pixels based on the union network architecture, where the tradeoff between accuracy and efficiency is disregarded. In reality, hard pixels (complex regions) may be caused by variations in tumor intensity, size, density, texture, and contrast. Figure 1 shows an example of an image with hard pixels. The image in Figure 1(c) highlights hard pixel regions that are difficult to identify owing to their significant ambiguity. These regions result in the incorrect segmentation of tumors, as shown in Figure 1(d). The network is trained on the entire image and typically fails to accurately segment tumor regions with ambiguous surfaces surrounded by thick tissues; these pixels can be regarded as hard pixels. A more complex network architecture needs to be

* Corresponding author (email: fyzq10@126.com, shandamla@163.com, ylyin@sdu.edu.cn)

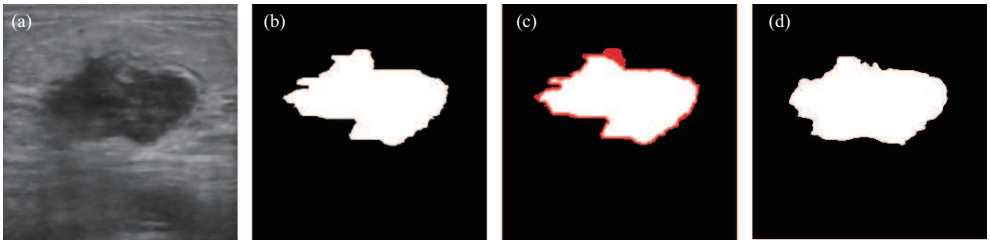

Figure 1 (Color online) (a) Original ultrasound image; (b) ground truth; (c) highlighted hard pixel regions; (d) segmentation result.

developed to process hard pixels. However, the computational complexity increases with the complexity of the network architecture.

A novel, difficulty-aware, prior-guided hierarchical network (DHN) for breast ultrasound image segmentation is proposed herein. Generally, the incorporation of a prior is observed to acquire more precise results [14]. Therefore, we first developed a difficulty prior learning module (DPLM) to learn the difficulty prior, which can be used to classify easy and hard pixels. In this study, pixels correctly segmented with probability scores exceeding 95% were regarded as easy, whereas the remaining were regarded as hard.

Furthermore, a novel hierarchical network is proposed for the adaptive segmentation of pixels using the learned prior guidance. It detects easy pixels in shallow layers and integrates a complex hard pixel processing unit (HPPU) at deep layers to recognize a small group of hard pixels. The HPPU is developed to learn high-level discriminative features of hard pixels. It can learn more effective features of hard pixels and improve the segmentation accuracy of hard pixels. Traditional networks learn the features of all pixels using the same feature extractor. Therefore, the balance between the segmentation accuracy and efficiency is difficult to guarantee. The proposed network can adaptively learn easy and hard pixels with different feature extractors, unlike conventional methods. Therefore, it improves the segmentation accuracy of hard pixels and the efficiency of easy pixels. In this study, the proposed method was evaluated using three breast ultrasound datasets. The experimental results show that the proposed method outperformed state-of-the-art methods.

The main contributions of this study are as follows:

(1) A novel, difficulty-aware, prior-guided hierarchical network is proposed for the adaptive segmentation of breast tumors. It can improve the segmentation accuracy of hard pixels and the efficiency of easy pixels.

(2) A DPLM is proposed to learn the difficulty prior of pixels. It can guide the network to segment images with different feature extractors.

(3) An HPPU is introduced to learn the discriminative features of hard pixels. A more complex feature extractor is developed for the HPPU to learn discriminative features. It overcomes the inadequate prediction of hard pixels for breast tumor segmentation and achieves high-quality image segmentation.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of related studies. Section 3 presents the proposed network architecture. Section 4 explains the experiments and provides the findings, Section 5 gives a detailed discussion of the findings, and Section 6 summarizes the findings of this study.

2 Related work

Breast tumor segmentation methods can be categorized into conventional and deep learning methods [15]. Conventional methods such as threshold [16], watershed [17], region growing [18, 19], active contour model [20, 21], and clustering [22] techniques are typically used for image segmentation. Ilhan et al. [16] recommended a morphological approach based on a median filter and a threshold to segment tumors. Nayak et al. [17] proposed breast tumor segmentation and breast density extraction using a watershed algorithm. However, the accuracy deteriorated in fatty and dense regions. Fang et al. [20] used a local bias correction function and a probability score to segment complex ultrasound images with a minimized energy function. Chowdhary et al. [22] presented an intuitionistic possibilistic fuzzy c-mean clustering method based on the hybridization of fuzzy c-mean and possibilistic fuzzy c-mean algorithms to segment and classify breast tumors. The algorithm conserves the positive points of the possibilistic fuzzy c-mean to address the issue of coincidence clusters. However, feature extraction from breast ultrasound images

is difficult because of artifacts such as noise, intensity inhomogeneity, and indistinct boundaries. The sensitivity of the traditional method to noise and inferior scalability affects the robustness of the methods.

Deep learning methods [7, 10, 23–26] have demonstrated significant performance improvements. DeepLab [24] outperformed the state-of-the-art methods for semantic segmentation. DeepLab uses atrous convolution to expand the receptive field and embed contextual information at multiple scales without increasing the computation time. Ronneberger et al. [7] presented an encoder-decoder structure with skip connections for medical image segmentation. The encoder in the network exploits contextual features, and the decoder obtains spatial resolution, whereas the skip connections conserve high-frequency details to ensure better object detection. Neural network performance and design have improved considerably in recent years, owing to the use of AlexNet [27], VGG [28], ResNet [11], ResNetXt [29], and Xception [30]. Researchers are developing new methods by adopting such networks as the backbone or utilizing the encoder-decoder network to capture contextual information while maintaining spatial learning. Yap et al. [31] performed breast ultrasound tumor segmentation on two different datasets using the patch-based LeNet [32], UNet [7], and a transfer learning strategy based on a pretrained FCN-AlexNet. Huang et al. [33] proposed information extension by training a fully convolutional network (FCN) with the addition of wavelength features to the original image. Conditional random fields with prior breast structure knowledge were utilized to improve the segmentation performance. Zhang et al. [34] developed a hierarchical mask-guided learning framework using a two-stage FCN model for coarse-to-fine breast tumor segmentation. Lei et al. [2] proposed a mask scoring region-based convolutional neural network (R-CNN) to perform automatic breast tumor segmentation. The network was composed of five subnetworks. A network block was used to establish a direct relation between mask quality and region class; subsequently, it was incorporated into the mask scoring R-CNN to segment images containing unclear regions of interest. The network yielded outstanding results, although it was limited by the image quality and tumor size. Chiang et al. [35] proposed an automated tumor detection technique based on a three-dimensional (3D) convolutional neural network and a prioritized candidate aggregation that extracts the volume of interest for tumor regions to prioritize tumors based on the computed probability. The model achieved high sensitivities, thereby demonstrating the system's potential; however, false positives with 100% sensitivity generated should be further decreased. Al-antari et al. [36] designed a computer aided diagnosis (CAD) system based on a full-resolution convolutional network and a deep convolutional neural network to segment and classify breast tumors. However, the model indicated limited accuracy and precision in localizing small objects. Zhou et al. [37] used a multitasking learning mechanism for joint classification and segmentation tasks. The framework utilizes an encoder-decoder structure to segment images and a lightweight multiscale network to classify images. Ho et al. [38] devised a multi-magnification network with multi-encoder, multi-decoder, and multi-concatenation operations to perform patch-wise image segmentation for breast tumors. In certain scenarios, the application of the multi-magnification network was beneficial. However, when the number of training examples is low and the context of the global tissue microenvironment cannot be integrated, its performance is deteriorated. Ahmed et al. [39] designed a hybrid methodology for breast tumor classification and segmentation based on two well-known deep learning methods, i.e., the mask-R-CNN [40] and DeepLab v3 [41]. However, despite their success, these techniques are deficient in segmenting minor tumor regions. Deep convolutional networks that utilize regularization methods and large network depths [1, 42] have demonstrated outstanding performances. Although these networks improved classification performance, significant computational time and memory requirements were incurred [43].

By contrast, prior knowledge learning has achieved considerably better performance in terms of tumor segmentation [37]. However, it is difficult to learn an effective prior using the shallow model.

Although the abovementioned methods can achieve a certain degree of performance, the difficulty of pixels was disregarded without considering the segmentation accuracy and efficiency. Recently, different importance-aware or difficulty-aware methods have been proposed to perform semantic segmentation. For example, Shareef et al. [44] presented a small tumor-aware network with two encoders and applied multiple kernels to segment small tumors. These encoders use receptive fields of different sizes to obtain contextual information at multiple scales and then fuse them. The architecture improves the overall performance for small tumor detection but yields high false positives for images with few small tumors. Nie et al. [45] used a fully convolutional adversarial network to perform the confidence learning of voxels and regions for a segmentation network and a difficult-aware attention mechanism to understand the structural information of hard regions. The proposed method yielded satisfactory dice similarity for prostrate, bladder, and rectum segmentation. However, the efficiency of the method was not analyzed

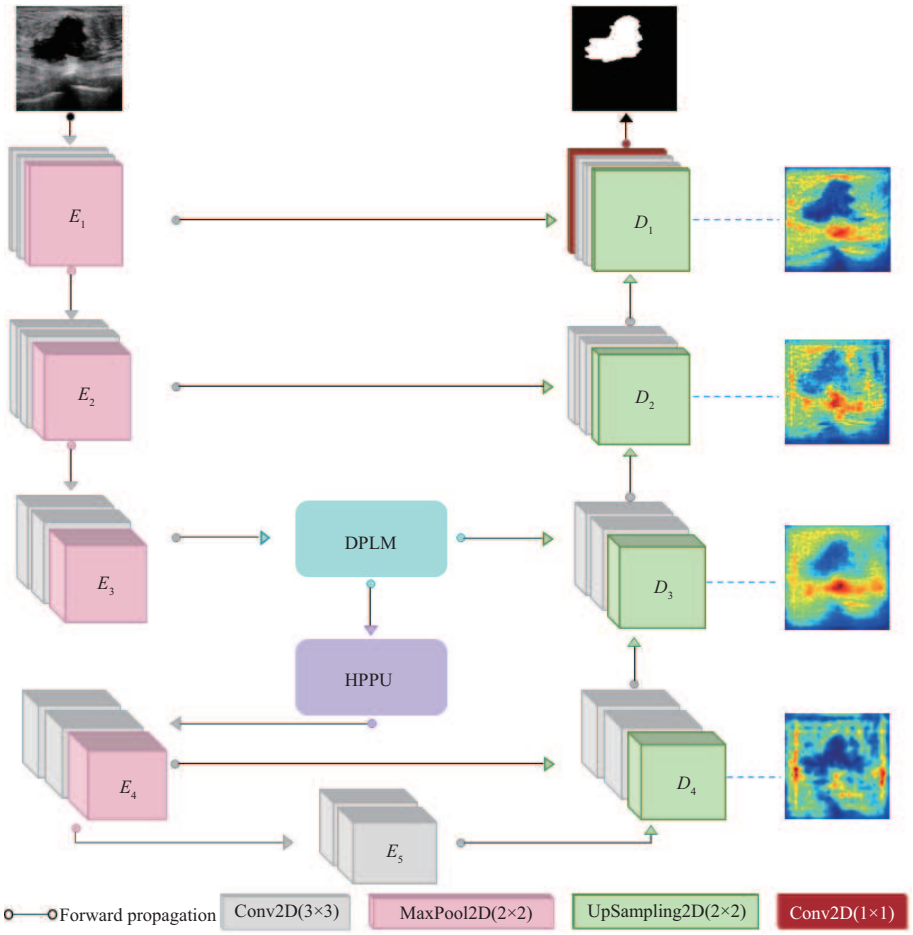

Figure 2 (Color online) Overall architecture of difficulty-aware prior-guided hierarchical network (best viewed in color with 150% zoom).

based on false positives. Xie et al. [46] designed a segmentation network with two branches, i.e., a common segmentation and a semantic difficulty branch. The former branch segments common pixels, whereas the latter uses the probability attention module to estimate the error mask and uses it to learn the semantic difficulty. The proposed method is used to segment objects in the environment and effectively improve hard-area segmentation. However, the current studies do not use a specific difficult-aware network to manage hard pixels in breast ultrasound (BUS) images. Therefore, we propose a difficult framework to segment easy/hard pixels in BUS images based on the estimated difficulty.

3 Methodology

3.1 Framework overview

The DHN was constructed based on the encoder-decoder architecture. It comprises fully convolutional encoder and decoder networks linked via skip connections. The encoder comprises four consecutive blocks E (as shown in Figure 2), which are composed of two convolutional layers with a kernel size 3×3, a rectified linear unit (ReLU) as an activation function, followed by a 2×2 max-pooling layer. Unlike traditional networks, the DPLM is introduced after the third layer to identify easy or hard pixels in our encoder. Based on the results of the DPLM, easy pixels are processed via shallow layers, whereas more discriminative high-level features of hard pixels are learned via the HPPU. To restore the spatial resolution of the extracted feature map to the same size as the input image, a decoder is constructed. It comprises four consecutive blocks (D1–D4 as shown in Figure 2) composed of a 2×2 upsampling layer, followed by two convolution layers with a kernel size 3×3 and a ReLU as an activation function. The encoder and decoder are connected via skip connections to transmit feature maps from the encoder to

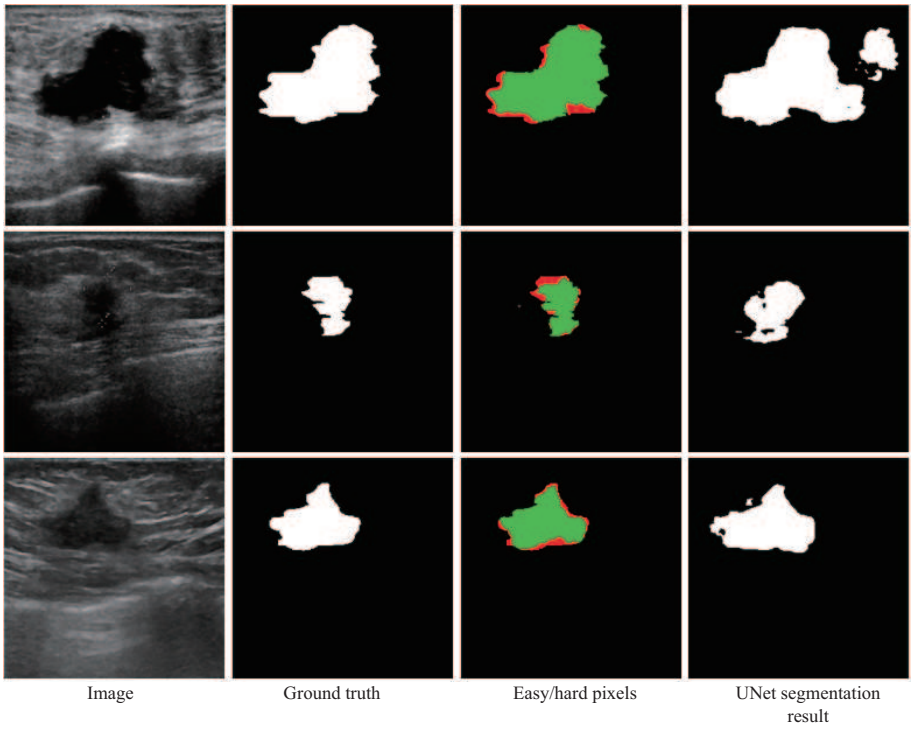

Figure 3 (Color online) Easy pixels vs. hard pixels.

the corresponding decoder to refine the segmentation [37, 47]. The final segmentation result is generated through a convolutional layer with a kernel size 1×1 and a sigmoid function based on the feature maps output via the encoder-decoder.

3.2 DPLM

In our model, difficulty learning is vital for learning features and segmenting images. Figure 3 shows some images whose easy pixels (green regions) are predicted accurately, and areas with hard pixels (red regions) that exhibit erroneous segmentation results. Therefore, we aim to identify the difficulty level of the pixels in the image.

In this study, a DPLM was constructed using two convolutional layers with a kernel size of 3×3 to realize more effective feature learning. Subsequently, the softmax layer was used to generate a $2 \times 32 \times 32 \times 1$ pixel-wise probability map L^0. The vector column $L_i^0 \in \mathbb{R}^{(N \times W \times H \times C)}$ for each $N \times C$ column denotes the i-th pixel's probability of belonging to the tumor/object in an image, where N represents the number of classes, W and H are the width and height of an image, respectively, and C represents the channels. If the maximum value of the i-th pixel, i.e., $L\_\max_i^0 = \mathrm{MAX}(L_i^0)$ and $L\_\max_i^0 \in \{L_{ij}^0 | j = 0, 1$ for tumor/background$\}$ is greater than a threshold value p ($L\_\max_i^0 \geqslant p$), then we assume that it is correctly predicted and hence forwarded to the symmetric layer of the decoder network.

The DPLM (as shown in Figure 4) classifies pixels with a probability higher than the threshold value as easy pixels. By contrast, pixels that cannot satisfy the condition are classified as hard pixels. The threshold value p generally exceeded 0.95. Pixels with probability ($L\_\max_i^0 \geqslant 0.95$) constituted approximately 40% of an image region that contained plenty of easy pixels and relatively few challenging hard pixels that were at high risk of being misclassified. In particular, p denotes the number of accepted easy pixels and exceptionally hard pixels rejected at that level. The increase or decrease in the value of p determines the number of easy pixels to be processed in the early layers, and hard pixels are managed in the later layers. Figure 5 presents a graph showing the percentage of easy pixels accepted at different threshold values. The result shows that the pixels were evaluated more strictly as the value of p increased, where challenging hard pixels were prioritized. The performance of the DPLM was affected by the small values of p and resulted in early decisions that did not provide a satisfactory awareness of the hard pixels. To select a threshold value more precisely, we measured the IoU for selected pixels over different threshold values. The best performance was achieved when a threshold value of 95% was used. However, the value

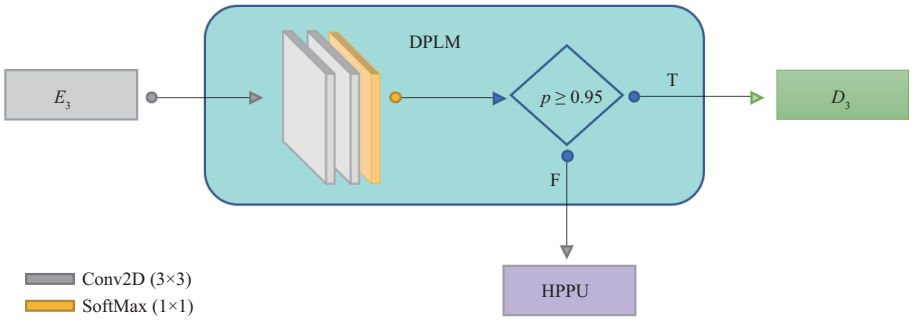

Figure 4 (Color online) Difficulty prior learning module (DPLM).

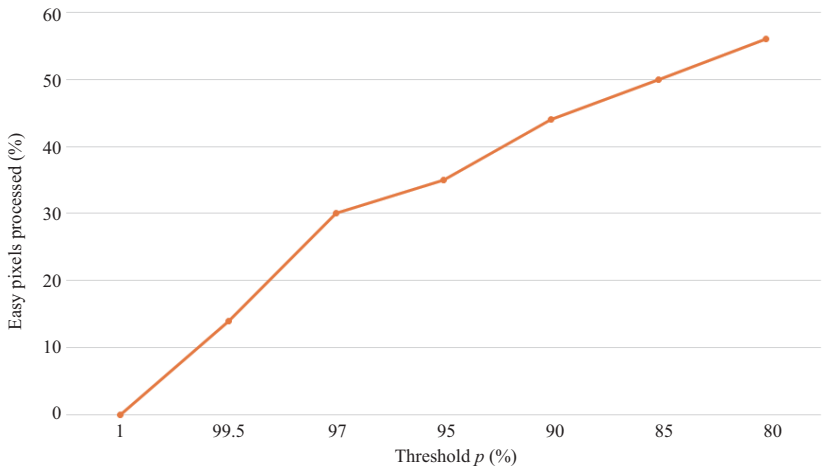

Figure 5 (Color online) Easy pixel classification at different threshold values.

of p can be tentatively selected using the validation results.

3.3 HPPU

To manage hard pixels by learning more discriminative features, an HPPU (shown in Figure 6) was developed. The HPPU was composed of five parallel convolutional layers and a global average pooling layer. The traditional convolution layer uses several filters to extract visual elements, thereby increasing the computational cost. Dilated convolutions permit the receptive field of the filters to be expanded at any convolution layer without additional parameters or computational cost [24]. Because dilated convolutions can accumulate more spatial information specifically to indistinctly recognizable boundaries [48], we applied dilation rates to the HPPU convolution layers. The variation in the dilation rates resulted in additional useful receptive fields that enable the image context to be obtained at various scales. Dilated convolutions with dilation rate d insert $d-1$ zeros between consecutive filter values. The kernel filter with a size $k \times k$ with dilation rate d is altered to $k_i = k + \langle k{-}1\rangle\langle d{-}1\rangle$ with a zero increase in the number of parameters and computations. This technique allows the module to control the receptive field, thereby channelizing the surrounding distinct information to indistinct object regions, as well as to establish the finest balance between precise localization and context inculcation.

Unlike traditional convolutions, the HPPU performs convolution only on hard pixels. In the proposed HPPU, multiscale features are learned using multiple parallel convolutional layers with different dilation rates. The final feature map confining intensely highlighted tumor regions is generated by concatenating multiscale features from four different dilated convolutions via a 1×1 convolution to reduce and restore the dimensions; additionally, a global average pool layer is used to obtain an image's global context. The HPPU improves hard pixel learning by expanding the receptive field, thereby allowing hard pixels to achieve refinement based on knowledge acquired regarding the surrounding well-defined regions. Hence, high-level learning at multiple scales produces more distinct contextual knowledge and localization for hard pixels. Consequently, the hard pixels are discoverable and easily captured by the network. Eq. (1)

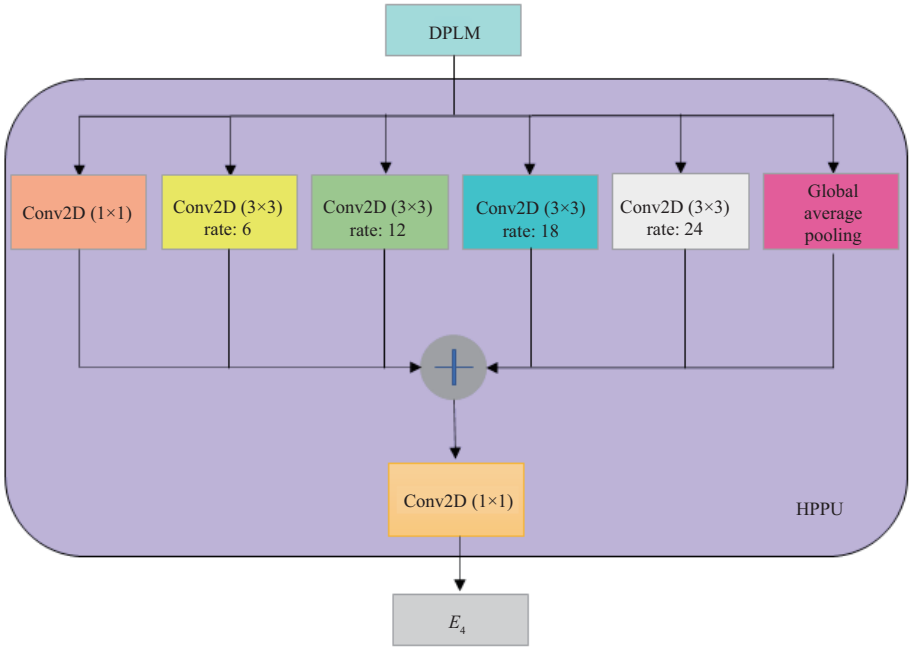

Figure 6 (Color online) Hard pixel processing unit (HPPU).

shows a mathematical representation of the HPPU.

$$\begin{aligned}\text{HPPU} = &\text{conv}_k(I) \oplus \text{conv}_{k,d}(I) \oplus \text{conv}_{k,d}(I) \\ &\oplus \text{conv}_{k,d}(I) \oplus \text{conv}_{k,d}(I) \oplus \text{GAP}(I,1),\end{aligned} \tag{1}$$

where $\text{conv}_{k,d}$ shows the convolution operation over image I with kernel size $k \times k$ where $k = 3$ for a dilated convolution; otherwise, the value k will be 1 and dilation rate $d = 6, 12, 18, 24$, respectively. GAP denotes the global average pooling layer over I with dimension 1, and $\oplus$ represents the concatenation operation.

4 Experimental results

To validate our model, we conducted a series of experiments on three datasets: QHSP (private), UDIAT (public), and BHAYE (public).

4.1 Datasets

QHSP. This dataset was obtained from the Qianfoshan Hospital of Shandong Province. It contains 186 breast ultrasound grayscale images with a tumor that belongs to one of two categories: benign (135) and malignant (51). The images were captured from four devices, i.e., ALOKA α 10, AplioXG, GE LOGIQ E7, and SIEMENS Sequoia 512.

UDIAT. This dataset is obtained from the UDIAT Diagnostic Center of the Parc Tauli Corporation, Sabadell, Spain [31]. It contains 163 images classified into benign (110) and malignant (53) images. The image resolution is 760 $\times$ 570 pixels (with a nominal pixel size of 0.084 mm). A Siemens ACUSON Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz) was used to construct the ultrasound images.

BHAYE. The dataset contains 780 images from the Bhaye Hospital for Early Detection and Treatment of Women's Cancer, Cairo (Egypt) [49], captured using the LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound system. The dataset is classified into normal (133), benign (437), and malignant (210) images with a standardized image size of 500 $\times$ 500 pixels.

4.2 Implementation details

We used PyTorch and NVIDIA TITAN XP 12 G, and an Intel Xeon Gold 5115 2.4 G GPU for training and validation, respectively. We performed a five-fold cross-validation on the QHSP dataset, where three

folds (60% of the images) were used for training, one fold (20% of images) for validation, and one fold (20% of images) for testing. In each fold, images were selected randomly for training, validation, and testing. The UDIAT and BHAYE datasets were fully utilized for testing. All images in the dataset were first resized to a 256 × 256 pixel resolution; subsequently, they were normalized and then transformed into grayscale images. Data augmentation was performed to facilitate small-sized datasets, which were randomly rotated, sheared, zoomed, cropped, and flipped horizontally with real-time data augmentation for each batch, thereby generating more than 90000 images to improve the model's performance and robustness. The model was trained on 1000 epochs using the Adam optimizer with a learning rate of 5×10^{-7}, a batch size of 8, and a weight decay rate of 0.08.

4.3 Evaluation metrics

We used six metrics to evaluate the performance of the proposed method. The metrics were accuracy (Acc), sensitivity or true-positive rate (TP), specificity or true-negative rate (TN), false-positive rate (FP), intersection over union (IoU), and the Dice coefficient. These metrics are expressed as follows:

$$\mathrm{Acc} = \frac{|(T_g \cap T_p) \cup (B_g \cap B_p)|}{|T_g \cup B_g|}, \tag{2}$$

$$\mathrm{TP} = \frac{|T_g \cap T_p|}{|T_g|}, \tag{3}$$

$$\mathrm{TN} = \frac{|B_g \cap B_p|}{|B_g|}, \tag{4}$$

$$\mathrm{FP} = \frac{|T_g \cup T_p - T_g|}{|T_g|}, \tag{5}$$

$$\mathrm{IoU} = \frac{|T_g \cap T_p|}{|T_g \cup T_p|}, \tag{6}$$

$$\mathrm{Dice} = \frac{2|T_g \cap T_p|}{|T_g| + |T_p|}. \tag{7}$$

In the equations above, T_g and B_g denote the pixels of the tumor and background regions in the ground-truth image, respectively. Similarly, T_p and B_p indicate the pixels of the tumor and background regions in the predicted image, respectively. Additionally, a receiver-operating curve (ROC) was used for the performance analysis.

4.4 Effectiveness of difficulty prior

The effectiveness of the difficulty of prior guidance on breast tumor segmentation was investigated. The DPLM estimates the difficulty prior, whereas the HPPU segments the hard pixels. In our experiment, we compared the performances of a convolutional encoder-decoder network with and without difficulty prior guidance. Table 1 presents a detailed analysis of the effect of our network architecture on multiple datasets. The DPLM separates the easy and hard pixels by evaluating the probability of the pixel. The pixels that cannot satisfy the threshold value p are forwarded to the HPPU for further processing. Table 1 shows the effects of the DPLM on the final segmentation results and the accuracy of the model.

We observed that the convolutional encoder-decoder network without a DPLM exhibited performance discrepancies on the datasets with varying pixel difficulties. A network without DPLM considers all pixels equally, which affects the performance of the network. By contrast, a convolutional encoder-decoder network with a DPLM, which can estimate pixel difficulty and train accordingly, achieves outstanding performance. Our approach yielded high values across all metrics. The encoder network with a DPLM managed hard pixels better than an encoder network without a DPLM. The findings indicate that the proposed method effectively addressed images with different difficulties.

As mentioned in Section 3, we proposed a mechanism to differentiate easy/hard pixels and introduced an HPPU to learn hard pixels. The network can focus on hard pixel segmentation without complicating the network and wasting resources on easy pixels. We used UNet as the backbone for our model; the HPPU visualization results and their effect on the final prediction surpassed the UNet results. Figure 7 shows a comparison between the results of the proposed model with those of the UNet. The HPPU refines the

Table 1 Experimental results of DPLM and HPPU in a convolutional encoder-decoder network on three datasets

		QHSP	UDIAT	BHAYE
UNet	Accuracy	0.960	0.982	0.979
	IoU	0.706	0.601	0.836
	Dice	0.814	0.752	0.877
UNet+ DPLM (HPPU)	Accuracy	**0.963**	**0.983**	**0.980**
	IoU	**0.728**	**0.652**	**0.847**
	Dice	**0.828**	**0.762**	**0.887**

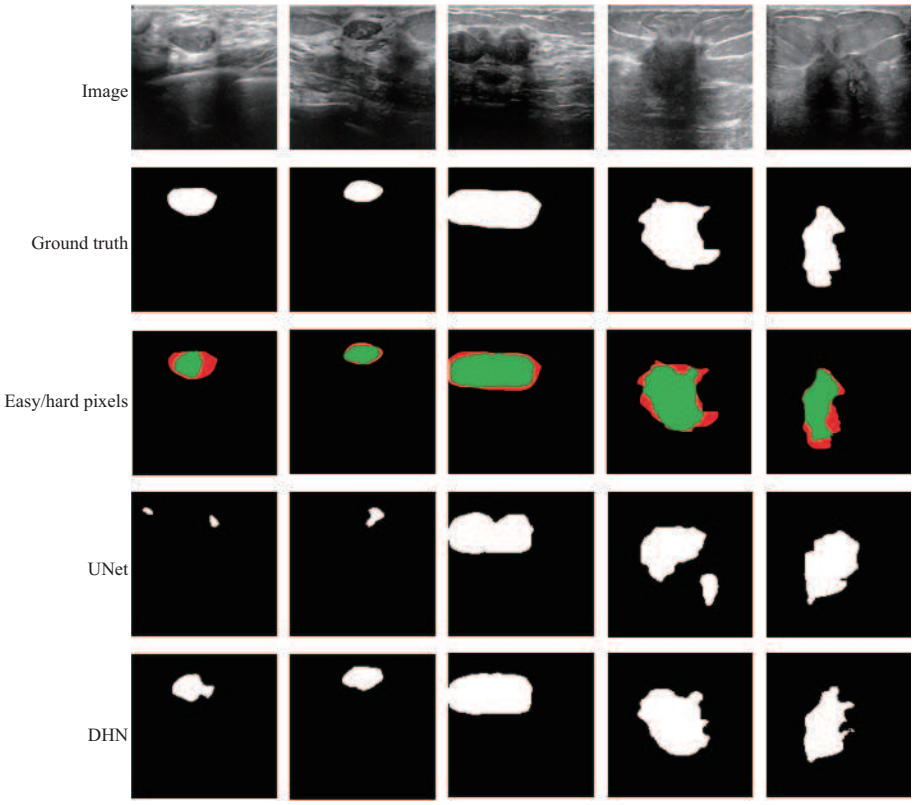

Figure 7 (Color online) UNet vs. DHN.

features of the hard pixels by performing multiple parallel convolution operations. Dilated convolutions capture more contextual information than normal convolutions. The fusion of these convolutions provides better access to hard pixel regions because convolutions with different dilation rates generate feature maps with semantic knowledge at multiple levels. The concatenation of explicit location and the contextual information of hard pixels attained at multiple scales progressively improve the network's learning and result in accurate segmentation. Figure 7 shows images highlighting easy (green) and hard (red) pixels. The UNet segments the hard pixels incorrectly, whereas the DHN segments the easy and hard pixels accurately. The incorporation of the HPPU in the DHN resulted in a more accurate display of the edges of a tumor, along with a significant increase in segmentation performance, thereby validating the effectiveness of our model.

4.5 Comparison with state-of-the-art methods

To demonstrate the effectiveness of our model, we compared our proposed method, i.e., the DHN, with the following five state-of-the-art methods: FCN32 [9], UNet [7], SegNet [10], DeepLab v3 [41], and PSPNET [25] on three different datasets. Table 2 shows the quantitative results comprehensively. The DHN surpasses the other five methods with a clear margin in terms of different metrics for the three datasets. Our method achieved a better IoU and Dice coefficient than the other techniques, particularly for hard pixels that are more difficult to segment. The DHN increased the accuracy significantly as

Table 2 Performance evaluation of our model compared with state-of-the-art methods on different datasets

	Methods	QHSP						UDIAT						BHAYE					
		TP	TN	FP	Acc	Dice	IoU	TP	TN	FP	Acc	Dice	IoU	TP	TN	FP	Acc	Dice	IoU
Benign	FCN32	0.403	0.506	0.395	0.912	0.378	0.262	0.245	0.707	0.193	0.968	0.230	0.164	0.161	0.890	0.110	0.942	0.166	0.117
	UNet	0.733	0.506	0.486	0.928	0.621	0.497	0.746	0.736	0.255	0.974	0.680	0.583	0.894	0.931	0.069	0.985	0.885	0.848
	PSPNET	0.709	0.576	0.416	0.929	0.627	0.520	0.674	0.827	0.273	0.982	0.664	0.571	0.863	0.936	0.064	0.985	0.883	0.829
	SegNet	0.743	0.217	0.774	0.895	0.569	0.359	0.773	0.045	0.954	0.822	0.589	0.142	0.906	0.927	0.073	0.985	0.896	0.854
	DeepLab v3	0.735	0.608	0.383	0.936	0.646	0.526	0.790	0.756	0.244	0.984	0.713	0.620	0.880	0.939	0.057	0.986	0.891	0.846
	DHN	**0.746**	**0.561**	**0.389**	**0.936**	**0.666**	**0.562**	**0.798**	**0.751**	**0.248**	**0.984**	**0.772**	**0.657**	**0.907**	**0.943**	**0.057**	**0.987**	**0.901**	**0.862**
Malignant	FCN32	0.587	0.560	0.399	0.872	0.549	0.388	0.350	0.845	0.155	0.955	0.372	0.279	0.289	0.853	0.142	0.886	0.332	0.239
	UNet	0.785	0.558	0.441	0.882	0.682	0.520	0.810	0.824	0.176	0.979	0.783	0.692	0.856	0.912	0.082	0.965	0.861	0.813
	PSPNET	0.779	0.580	0.420	0.881	0.683	0.522	0.690	0.922	0.178	0.977	0.713	0.634	0.821	0.908	0.088	0.960	0.842	0.783
	SegNet	0.766	0.353	0.646	0.850	0.615	0.407	0.800	0.042	0.957	0.830	0.628	0.244	0.836	0.912	0.080	0.963	0.848	0.804
	DeepLab v3	0.767	0.636	0.364	0.882	0.682	0.522	0.764	0.907	0.193	0.981	0.781	0.702	0.820	0.898	0.093	0.956	0.840	0.786
	DHN	**0.782**	**0.569**	**0.365**	**0.883**	**0.697**	**0.554**	**0.814**	**0.862**	**0.137**	**0.984**	**0.788**	**0.708**	**0.862**	**0.913**	**0.080**	**0.968**	**0.862**	**0.818**

Table 3 Overall segmentation performance of DHN

Methods	UDIAT				BHAYE			
	TP	FP	IoU	Dice	TP	FP	IoU	Dice
STAN	0.80	0.27	0.70	0.78	0.76	0.42	0.66	0.75
ESTAN	0.84	0.22	0.74	0.82	0.80	0.36	0.70	0.78
DSAG	0.77	0.19	0.62	0.73	0.86	0.07	0.82	0.87
DHN	**0.83**	**0.21**	**0.76**	**0.82**	**0.87**	**0.04**	**0.85**	**0.88**

compared with the UNet, demonstrating the effectiveness of multiscale cascade layers in learning the hard pixels' information, thereby validating the efficiency of our model.

We conducted another quantitative comparison with two of the latest available tumor-aware networks for breast ultrasound image segmentation, i.e., the STAN [50] and ESTAN [44], for publicly available datasets. In addition, we used another difficulty-aware method to segment tumors in breast ultrasound images, i.e., DSAG [51]. As shown in Table 3, the IoU and Dice rates of the DHN were higher than those of the STAN, ESTAN, and DSAG. In addition, the FP rate of our approach was lower than that of the other methods. Our approach achieved the highest positive rate for the BHAYE dataset. However, for the UDIAT dataset, its TP was higher than those of the STAN and DSAG, whereas it differed by one point as compared with that of the ESTAN. Both the STAN and ESTAN used two encoders with different receptive fields to capture more contextual information for tumor regions and fuse them subsequently. By contrast, the DSAG used a difficulty grading module to identify the difficulty of images and then used a bi-network to segment images adaptively on different branches. Meanwhile, the DHN uses difficulty measures to classify pixels and capture more discriminative information through an HPPU based on convolution layers with receptive fields, thereby yielding the best results for tumor segmentation on both datasets.

PSPNET and DeepLab v3 demonstrated high specificity for the QHSP and UDIAT datasets, which are smaller than the BHAYE dataset. The specificity of the BHAYE dataset was relatively high in our model. FCN32 indicated a higher FP rate in benign and malignant tumors, although the DHN indicated a significantly higher FP rate than the other methods. In benign tumors, the accuracy of the DHN was similar to the highest accuracy achieved among the other models; however, in malignant tumors, the DHN demonstrated the highest accuracy among all the methods. Figure 8 presents a visual example of a qualitative comparison between the DHN and other benchmark methods. As shown, DHN surpassed the other methodologies with consistent results. Figures 9(a)–(c) show a comparison of the ROC between our method and state-of-the-art methods on different datasets to verify the performance of our method. The results show that our model is superior to the other state-of-the-art methods, thereby proving the effectiveness of the DHN model. Figure 10 shows the outputs of different layers and provides sufficient evidence that the DPLM-HPPU integration improves breast ultrasound image segmentation results. The segmentation results of the DHN and ground truth for the input images were 99% similar, confirming that the difficulty-aware mechanism aided the network in emphasizing hard pixels.

4.6 Performance analysis

To demonstrate the tradeoff between speed and accuracy, we compared the speed (frames per second, fps) and accuracy (IoU) of the proposed model with those of other state-of-the-art models. We evaluated the performances of all the methods on the validation and test sets of the UDIAT dataset. The fps for each

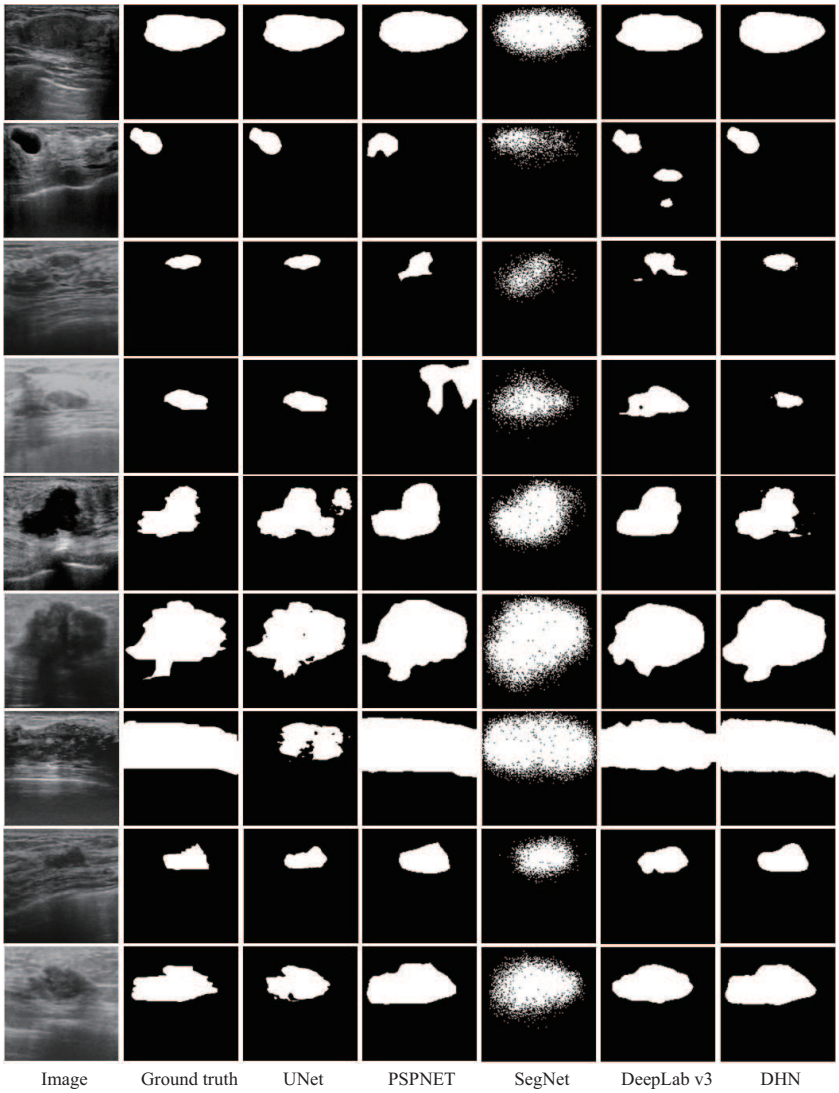

Figure 8 Qualitative evaluation of DHN compared with other benchmark methods.

model was measured on a TITAN XP GPU. All the models were evaluated without any pre-processing or post-processing to ensure a fair assessment. In general, the DHN achieved the best performance among the methods evaluated, as shown in Figure 11.

PSPNET and FCN32 achieved higher fps rates owing to their small sizes but a lower mIoU than the DHN. Meanwhile, DeepLab v3 uses ResNet as the backbone network, which increased its accuracy. However, its inference speed was comparatively slow, i.e., 10.81 fps. The DPLM allows the network to focus only on hard pixels in the HPPU and manage easy pixels in the early layers. The proposed model achieves the best tradeoff between speed and accuracy, with 86% mIoU and 12.94 fps.

5 Discussion

Herein, a mechanism was proposed that distinguished between easy and hard pixels in a breast ultrasound image by estimating the probability of each pixel in a probability map generated at a certain layer. Subsequently, each pixel probability was compared against a threshold value and categorized as correctly predicted or rejected. The best threshold value was finalized by monitoring the increase in IoU observed for correctly predicted pixels at different threshold values. A further integration of the hard pixel processing unit resulted in the better localization and contextual information of hard pixels captured through the different receptive fields across different convolutional layers. The gain of the network's overall performance confirmed the effectiveness of the DPLM and HPPU in better understanding the easy and hard

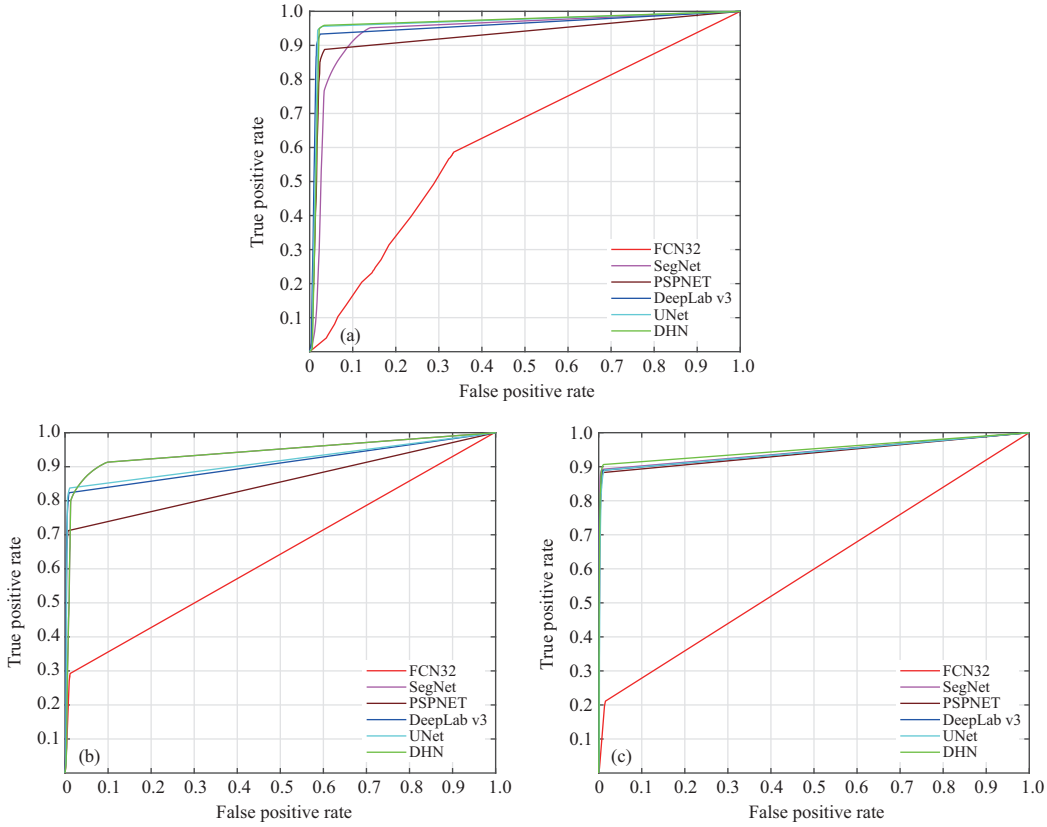

Figure 9 (Color online) ROC curves of (a) QHSP dataset, (b) UDIAT dataset, and (c) BHAYE dataset.

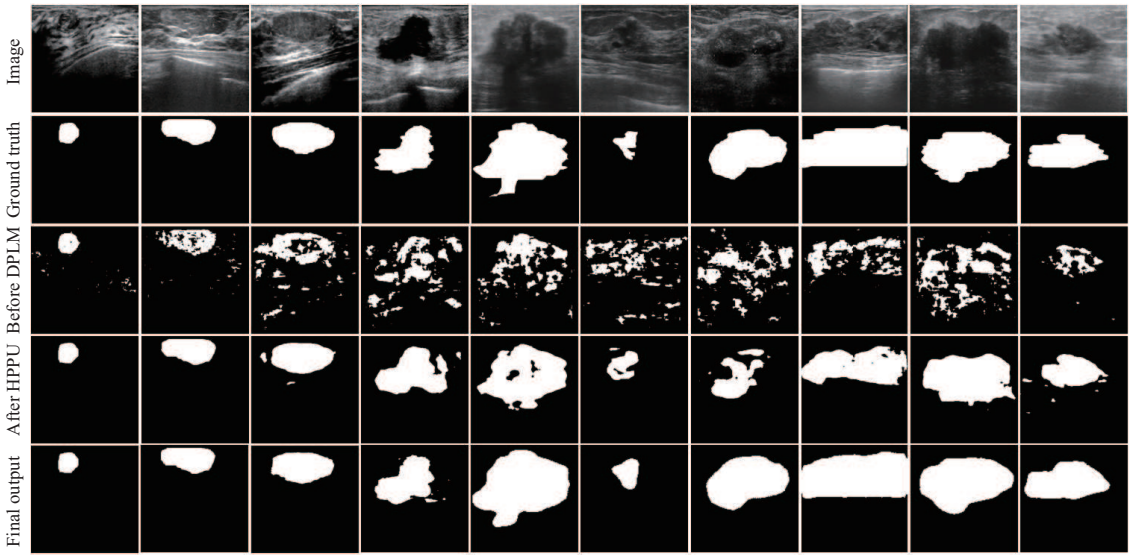

Figure 10 Visualization of outputs of different layers.

pixels. Furthermore, the network's ability to focus on difficult and complex regions in an image instead of using a unified model for all pixels resulted in the accurate and swift segmentation of tumors without any additional cost.

We compared the proposed method with several state-of-the-art methods, specifically tumor-aware breast tumor segmentation methods, in terms of image segmentation. As shown by the performance evaluation presented in Tables 2 and 3, and Figure 8, the UNet, PSPNET, and DeepLab v3 demonstrated better feature extraction and segmentation ability for breast ultrasound images than the other models. The UNet uses a contractive and expansive path with multiple convolution layers to extract the semantic

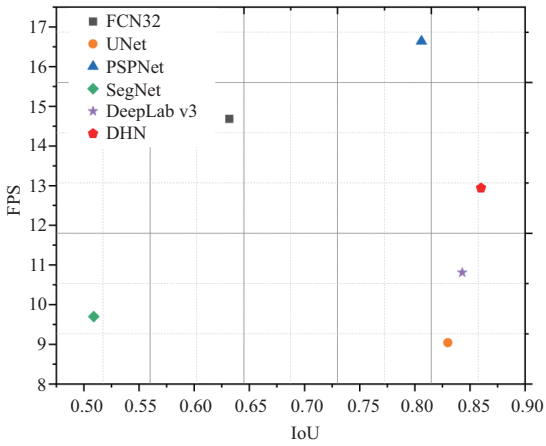

Figure 11 (Color online) Accuracy (IoU) and speed (fps) of various networks.

features of an image. In addition, it utilizes upsampling layers to restore the original image resolution. The most significant benefit of the network is that it employs skip connections to link the contractive and expansive convolution layers at each level; consequently, the grained details from the compression path are merged into the expansive path that is lost owing to pooling, and the segmentation quality is enhanced. However, the fusion of semantic and spatial features results in objects captured at different scales; as such, it does not benefit the relationship between objects globally and yields limited performance when the object changes in size and shape. The PSPNET utilizes spatial pooling at multiple scales to obtain a contextual prior that demonstrates effective performance in semantic segmentation. Although the feature map contains abundant semantic information, it lacks detailed boundary information regarding the object, thereby resulting in performance deterioration when complex objects with ambiguous boundaries are encountered. DeepLab v3 establishes an atrous spatial pyramid pooling (ASPP) module with global pooling to improve contextual awareness. DeepLab v3 demonstrates improved performance by recovering dense encoder features from the ASPP and detailed object boundaries through the decoder. However, high spatial resolution is crucial in dense image prediction, and an inconsistent increase in the dilation rate can result in an ineffective model. The deficiencies of these methods result in improper segmentation results for ultrasound images with complex regions.

DSAG presents a difficulty-aware mechanism in a network with two separate branches to manage hard and easy images. A difficulty grading module is adapted to determine the difficulty of images; subsequently, a bi-network is used to segment images adaptively on different branches. In DSAG, the complex branch employs a spatial attention module and graph-based energy incorporating a spatial attention constraint after an SE-UNET to learn the complex features of hard images. The STAN and ESTAN involve a tumor-aware strategy for segmenting tumor regions of different sizes. Both models use a two-encoder architecture with multiple kernels to learn multiscale features to improve performance.

For our model, we used the encoder-decoder structure of the UNet to retrieve semantic information for tumors in BUS images. The DHN employs a DPLM to estimate the difficulty of a pixel at a certain level of the encoder, where most of the easy features are already captured. Pixel discrimination allows the network to separate all correctly predicted pixels and forward them to the symmetric decoder for further use. Hence, the unnecessary wastage of resources is reduced, and the network will only focus on the feature retrieval of hard pixels. The HPPU is designed to function based on the principle of dense image segmentation with dilated convolutions. It enlarges the receptive field, thereby enabling a wide-range context utilization with better pixel localization, and global average pooling allows us to obtain the global context of an image. The fusion of these features provides a better understanding of hard pixels in an image. The decoder with skip connections restores the resolution size of the image and utilizes the contextual and spatial knowledge of the features to predict the nature of a tumor in an ultrasound image.

Among these comparison methods, DSAG is an image-level, difficulty-aware network that cannot accurately obtain the prior of the local hard pixels easily. Unlike DSAG, the proposed method can learn the prior of the pixel-level difficulty, which facilitates the identification of prior hard pixels. In addition, the proposed method can segment an image using a network. However, a graph cut is used for additional post-processing in DSAG. Therefore, the proposed method outperforms DSAG in terms of segmentation

accuracy and efficiency. The STAN and ESTAN use two encoders with different receptive fields to capture more contextual information for tumor regions. However, every pixel is considered equally. Therefore, it is difficult to acquire sufficient useful information regarding the hard pixels. By contrast, the proposed method can focus on hard pixels, and an HPPU is constructed specifically for the feature extraction of hard pixels. Consequently, it can learn more discriminative information regarding hard pixels, improve the segmentation accuracy of hard pixels, and eventually improve the performance.

In future studies, the method can be generalized to multiple tumor segmentation in an image. In addition, the proposed method can be evaluated on two-dimensional ultrasound images. Furthermore, it can be used on 3D ultrasound images as well as extended to applications involving the segmentation of tumors in other body regions.

6 Conclusion

A difficulty-aware prior-guided adaptive segmentation method was proposed herein. In contrast to traditional methods that use union network architectures to segment all pixels, the proposed method can segment breast ultrasound images adaptively. A DPLM is used to learn the prior difficulty. Based on the difficulty prior, the features of easy pixels are learned using a simple feature extractor, whereas those of hard pixels are learned using an HPPU. The HPPU is a complex feature extractor composed of several parallel dilated convolutional layers and a global average pooling layer. It can learn more discriminative features of the hard pixels. Compared with other methods, the proposed method uses a different architecture that can achieve a balance between accuracy and efficiency when managing easy and hard pixels. Experiments on three datasets demonstrated the effectiveness and efficiency of the proposed method.

In the future, we will apply our model to segment tumors in other regions of the human body, such as the brain, liver, lymph nodes, and uterus.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61701280, 61801263, 61703235, 61701281), National Key R&D Program of China (Grant Nos. 2018YFC0830100, 2018YFC0830102), Natural Science Foundation of Shandong Province (Grant No. ZR2018BF012), Foundation of Distinguished Associate Professor in Shandong Jianzhu University. The authors would like to thank all the anonymous reviewers for their valuable time, comments, and suggestions.

References

1 Kuen J, Kong X, Wang G, et al. DelugeNets: deep networks with efficient and flexible cross-layer information inflows. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017. 958–966
2 Lei Y, He X, Yao J, et al. Breast tumor segmentation in 3D automatic breast ultrasound using Mask scoring R-CNN. Med Phys, 2021, 48: 204–214
3 Xi X, Xu H, Shi H, et al. Robust texture analysis of multi-modal images using local structure preserving ranklet and multi-task learning for breast tumor diagnosis. Neurocomputing, 2017, 259: 210–218
4 Rani V M K, Dhenakaran S S. Classification of ultrasound breast cancer tumor images using neural learning and predicting the tumor growth rate. Multimed Tools Appl, 2020, 79: 16967–16985
5 Guo R, Lu G, Qin B, et al. Ultrasound imaging technologies for breast cancer detection and management: a review. Ultrasound Med Biol, 2018, 44: 37–70
6 Yang X, Yu L, Wu L, et al. Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2017
7 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention, 2015. 234–241
8 Xue C, Zhu L, Fu H, et al. Global guidance network for breast lesion segmentation in ultrasound images. Med Image Anal, 2021, 70: 101989
9 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. 3431–3440
10 Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell, 2017, 39: 2481–2495
11 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 770–778
12 Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the impact of residual connections on learning. 2016. ArXiv:1602.07261
13 Zhang Z, Wu C, Coleman S, et al. DENSE-INception U-net for medical image segmentation. Comput Methods Programs Biomed, 2020, 192: 105395
14 Ghosh P, Mitchell M, Tanyi J A, et al. Incorporating priors for medical image segmentation using a genetic algorithm. Neurocomputing, 2016, 195: 181–194
15 Ilesanmi A E, Chaumrattanakul U, Makhanov S S. Methods for the segmentation and classification of breast ultrasound images: a review. J Ultrasound, 2021, 24: 367–382
16 Ilhan U, Ilhan A. Brain tumor segmentation based on a new threshold approach. Procedia Comput Sci, 2017, 120: 580–587
17 Nayak T, Bhat N, Bhat V, et al. Automatic segmentation and breast density estimation for cancer detection using an efficient watershed algorithm. In: Data Analytics and Learning. Berlin: Springer, 2019. 347–358
18 Raja N S M, Fernandes S L, Dey N, et al. Contrast enhanced medical MRI evaluation using Tsallis entropy and region growing segmentation. J Ambient Intell Human Comput, 2018. doi: 10.1007/s12652-018-0854-8

19 Punitha S, Amuthan A, Joseph K S. Benign and malignant breast cancer segmentation using optimized region growing technique. Future Computing Inf J, 2018, 3: 348–358
20 Fang L, Pan X, Yao Y, et al. A hybrid active contour model for ultrasound image segmentation. Soft Comput, 2020, 24: 18611–18625
21 Niaz A, Memon A A, Rana K, et al. Inhomogeneous image segmentation using hybrid active contours model with application to breast tumor detection. IEEE Access, 2020, 8: 186851
22 Chowdhary C L, Mittal M, Kumaresan P, et al. An efficient segmentation and classification system in medical images using intuitionist possibilistic fuzzy c-mean clustering and fuzzy SVM algorithm. Sensors, 2020, 20: 3903
23 Li Y, Qi H, Dai J, et al. Fully convolutional instance-aware semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 2359–2367
24 Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell, 2018, 40: 834–848
25 Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 2881–2890
26 Ullah I, Jian M, Hussain S, et al. Global context-aware multi-scale features aggregative network for salient object detection. Neurocomputing, 2021, 455: 139–153
27 Ragab D A, Sharkas M, Marshall S, et al. Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ, 2019, 7: e6201
28 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2015. ArXiv:1409.1556
29 Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 1492–1500
30 Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. 1251–1258
31 Yap M H, Pons G, Marti J, et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform, 2018, 22: 1218–1226
32 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEEE, 1998, 86: 2278–2324
33 Huang K, Cheng H D, Zhang Y, et al. Medical knowledge constrained semantic breast ultrasound image segmentation. In: Proceedings of the 24th International Conference on Pattern Recognition (ICPR), 2018. 1193–1198
34 Zhang J, Saha A, Zhu Z, et al. Hierarchical convolutional neural networks for segmentation of breast tumors in MRI with application to radiogenomics. IEEE Trans Med Imag, 2019, 38: 435–447
35 Chiang T C, Huang Y S, Chen R T, et al. Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation. IEEE Trans Med Imag, 2019, 38: 240–249
36 Al-antari M A, Al-masni M A, Choi M T, et al. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int J Med Inf, 2018, 117: 44–54
37 Zhou Y, Chen H, Li Y, et al. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. Med Image Anal, 2021, 70: 101918
38 Ho D J, Yarlagadda D V K, D'Alfonso T M, et al. Deep multi-magnification networks for multi-class breast cancer image segmentation. Computized Med Imag Graph, 2021, 88: 101866
39 Ahmed L, Iqbal M M, Aldabbas H, et al. Images data practices for semantic segmentation of breast cancer using deep neural network. J Ambient Intell Human Comput, 2020. doi: 10.1007/s12652-020-01680-1
40 He K, Gkioxari G, Dollár P, et al. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 2961–2969
41 Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. 2017. ArXiv:1706.05587
42 Lv E, Wang X, Cheng Y, et al. Deep ensemble network based on multi-path fusion. Artif Intell Rev, 2019, 52: 151–168
43 Ullah I, Jian M, Hussain S, et al. A brief survey of visual saliency detection. Multimed Tools Appl, 2020, 79: 34605–34645
44 Shareef B, Vakanski A, Xian M, et al. Estan: enhanced small tumor-aware network for breast ultrasound image segmentation. 2020. ArXiv:2009.12894
45 Nie D, Wang L, Xiang L, et al. Difficulty-aware attention network with confidence learning for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2019. 1085–1092
46 Xie S, Feng Z, Chen Y, et al. DEAL: difficulty-aware active learning for semantic segmentation. In: Proceedings of the Asian Conference on Computer Vision, 2020
47 Ullah I, Jian M, Hussain S, et al. DSFMA: deeply supervised fully convolutional neural networks based on multi-level aggregation for saliency detection. Multimed Tools Appl, 2021, 80: 7145–7165
48 Mathai T S, Lathrop K L, Galeotti J. Learning to segment corneal tissue interfaces in OCT images. In: Proceedings of IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019. 1432–1436
49 Al-Dhabyani W, Gomaa M, Khaled H, et al. Dataset of breast ultrasound images. Data Brief, 2020, 28: 104863
50 Shareef B, Xian M, Vakanski A. Stan: small tumor-aware network for breast ultrasound image segmentation. In: Proceedings of IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020. 1–5
51 Xu Q, Xi X M, Meng X J, et al. Difficulty-aware bi-network with spatial attention constrained graph for axillary lymph node segmentation. Sci China Inf Sci, 2022, 65: 192102