https://doi.org/10.2352/EI.2026.38.7.IMAGE-277

Local Attention and Detail-Enhanced Network for Mass Segmentation in Whole Mammograms

Qingkun Guo[a], Yixuan Wang[b], Wenzong Jiang[c], Bin Shen[d], Weifeng Liu[a], Lin Cong[a,*], Baodi Liu[a,*]
[a] **College of Control Science and Engineering, China University of Petroleum (East China), Qingdao, China**
[b] **Electronic and Computer Engineering, Shenzhen MSU-BIT University, Shenzhen, China**
[c] **College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China**
[d] **Celonis, New York, United States**
[*] **Corresponding author**

Abstract

Mammography is one of the most commonly used tools for early screening of breast cancer. Developing computer-aided diagnosis (CAD) based on mammographic images to assist doctors in making efficient and accurate diagnoses holds significant research value. Mass segmentation in mammograms is a core component of breast cancer CAD systems and an essential step in further qualitative analysis of breast cancer. However, significant challenges persist in the field of mass segmentation in whole mammograms, including model misalignment due to the small proportion of mass regions and difficulties in segmenting boundaries caused by blurred edges of mass areas. To solve these challenges, this paper proposes a local attention and detail-enhanced network (LADE-Net) for mass segmentation in whole mammograms. LADE-Net employs an asymmetric encoder-decoder architecture and introduces a lightweight local attention (LA) module aimed at early and precise localization of breast mass regions. Importantly, we design a new detail-enhanced fusion residual network (DEFRB) to refine and enhance the learning of edge features in breast masses. We evaluated the performance of LADE-Net on two publicly available datasets (INbreast, CBIS-DDSM). Compared to previous works, LADE-Net achieved superior performance.

Introduction

Breast cancer is a significant public health issue worldwide [1]. According to data from the World Health Organization, millions of women are affected by breast cancer every year [2]. Early diagnosis and swift intervention for breast cancer cases are particularly crucial [3]. Compared to other breast examination methods such as ultrasound, magnetic resonance imaging (MRI), and positron emission tomography (PET), mammography stands out as a globally recognized standard for breast cancer screening due to its affordability, simplicity, low radiation risk, ability to detect cancerous areas and suspicious masses in the early stage of lesions [4]. However, the interpretation of mammograms is typically assessed by radiologists, and accuracy of diagnosis largely relies on their level of expertise [5]. To solve these challenges, many computer-aided diagnostic (CAD) systems had been designed to help radiologists in making efficient and accurate diagnoses. Mass segmentation in mammograms is a core component of breast cancer CAD systems and an essential step in further qualitative analysis of breast cancer. Early-stage breast cancer often presents without clear clinical symptoms and manifests as painless masses [6]. Generally, benign masses typically exhibit uniform contours and regular edges, while malignant masses show irregular contours and edges [7].

In the field of breast mass segmentation, the vast majority of deep learning methods rely on supervised techniques. Mass segmentation methods can be categorized into segmentation of regions of interest (ROIs) or segmentation of the full fields of view (FOVs). Previous work had mainly focused on extracted ROIs rather than the original whole mammographic images. The methods for extracting ROI data primarily involve manual and automated detection algorithms [8]. However, extracting ROI is difficult, cumbersome, and costly.

In the field of FOVs of breast mass segmentation, early traditional methods relied on manually crafted features for large-scale segmentation, which were insufficient for handling complex scenarios and led to high misdiagnosis rates. As deep learning techniques have advanced, many CNN-based methods have been developed and implemented for breast mass segmentation. Currently, methods for mass segmentation in whole mammograms can primarily be categorized into two types: those based on UNet [9] and those based on GANs [10]. However, existing UNet-based methods mostly directly apply models designed for natural image segmentation to the task of breast mass segmentation without specific optimizations tailored to the challenges in this domain. The images generated by GAN-based methods are relatively similar, making it difficult to address the issue of large variations in mass shapes. Additionally, current methods still struggle to accurately segment masses in entire breast images, including issues such as inaccurate model localization due to small mass regions and difficulties in delineating boundaries caused by blurred mass edges.

To solve these, this paper designs a model, the local attention and detail-enhanced network (LADE-Net), for FOVs of mass segmentation. LADE-Net adopts an asymmetric structure with different encoder and decoder blocks. The main contributions of this work are as follows:

1. This paper proposes LADE-Net for mass segmentation in whole mammograms, which is capable of accurately locating mass regions and achieving precise segmentation of mass boundaries.
2. A lightweight local attention module (LA) is introduced to improve the early localization accuracy of breast masses. LA combines 1D convolutions and group normalization (GN) techniques, effectively encoding the location informa-

tion of the region of interest without sacrificing the channel dimension.

3. A detail-enhanced fusion residual block (DEFRB) is proposed to refine and enhance the edge features of breast masses. The designed detail-enhanced fusion convolution (DEFconv) consists of five parallel-deployed convolutions (one vanilla convolution and four different differential convolutions).
4. Experiments conducted on two public datasets, INbreast, CBIS-DDSM, demonstrated that LADE-Net consistently achieved the highest average Dice scores, outperforming other benchmark methods.

Methods

The overall architecture of LADE-Net

The whole architecture of LADE-Net is shown in Figure 1, consisting of the following components: encoder module, decoder module, upsampling module, attention module, and skip connection. The simple workflow is outlined as follows:

First, the mammographic images are inputted into the encoding pathway for feature extraction and resolution compression. Subsequently, through the decoding pathway, the image is gradually restored to the original spatial resolution. Additionally, skip connections are utilized to combine high-resolution feature maps from the encoder with feature maps from the decoder, enabling the network to integrate low-level and high-level features to further generate fine segmentation maps. Finally, a 1x1 convolution is applied to generate the predicted binary image of the mass. In this setup, the encoder pathway consists of five encoder modules, where max pooling is applied after the first four modules. The decoder pathway comprises four alternating decoder modules and upsampling modules.

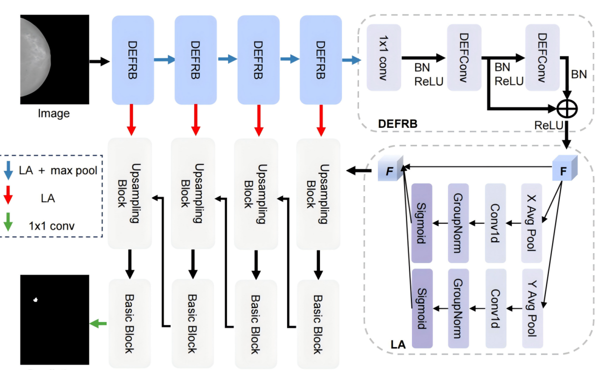


Figure 1. The overall architecture of the local attention and detail-enhanced network (LADE-Net). We deploy local attention (LA) modules in the encoder and skip connection parts to improve the early localization accuracy of breast masses and deploy detail-enhanced fusion residual blocks (DEFRB) in the encoder part to refine and enhance the edge features of breast masses.

Encoder Module: Detail-enhanced fusion residual block (DEFRB) serves as the model's encoder module to enhance the edge details of breast masses. The details of DEFRB will be provided in the following .

Dcoder Module: Basic block acts as the model's decoder module to restore feature maps, which consists of two consecutive 3x3 convolutions. It can be represented by the following formula:

$$Y_{basic}(x) = \delta(W_2 * \delta(W_1 * x + b_1) + b_2) \tag{1}$$

where σ represents the ReLU activation function. w and b denote the weight and bias of the convolution. $*$ corresponds to the convolution operation.

Upsampling Module: Attention-guided dense-upsampling (AU) block functions as the model's upsampling module to further enhance the fusion of extracted high and low-level features. AU block employs a dense upsampling convolution (F_{duc}) and a convolution layer after bilinear upsampling (F_{buc}) to upsample the high-level features (F_{high}). Subsequently, F_{duc} is merged with the lower-level features (F_{low}) through summation (F_{sum}). A convolution is then employed before concatenating F_{sum} with F_{buc} to create F_{concat}, facilitating a smoother process. Finally, channel-wise attention is applied to select important information from F_{concat}.

Attention Module: Local attention (LA) module is applied to the encoding and skip connection stages to achieve early and precise localization of breast mass regions. The details of LA will be provided in the following .

Detail-enhanced fusion residual block

Currently, the greatest challenge in FOVs of breast mass segmentation lies in the difficulty of accurately segmenting the edges of the masses. Some breast images have complex backgrounds that are similar to features in the mass region, such as the pectoral muscle. Additionally, there is minimal contrast between the inside and outside regions of the mass. Therefore, refining the model's capability to capture detailed features and deep-level characteristics of mass edges is crucial. Several deep learning methods aim to enhance model performance through depth or width augmentation in convolutions. The learning capacity of CNN structures has not been fully explored. Therefore, LADE-Net introduces a detail enhancement fusion residual block (DEFRB) to enhance the extraction of edge detail features. This modification is highly targeted, aiming to address the issue of blurred boundaries of masses in breast mass segmentation.

DEFRB consists of three convolutional layers. Specific details are illustrated in the upper right part of Figure 1. The first convolutional layer performs dimension expansion using a 1x1 kernel, without affecting the computational complexity and spatial dimensions of the feature map. The second and third convolutional layers are detail-enhancd fusion convolutions (DEFConv), utilizing a 3x3 kernel. A residual connection is employed after the second convolutional layer.

In the field of mass segmentation, previous methods [8] typically utilized vanilla convolution (VC) layers and dilated convolution layers for feature extraction. These convolutions explore the vast solution space without constraints, which limites their feature extraction capabilities. Additionally, we notice that high-frequency information (e.g., edges and contours) plays a crucial role in mass segmentation. Inspired by Chen et al. [18], we designed DEFConv to enhance the feature extraction capability of breast mass edges.

DEFConv consists of five parallel convolutional layers with 3x3 kernels, including VC, center difference convolution (CDC), angle difference convolution (ADC), horizontal difference convolution (HDC), and vertical difference convolution (VDC). Vanilla convolution is the most basic convolution operation commonly used in CNN. The latter four belong to difference convolutions, which improve the network's capability to capture image details by computing differences between pixels, thereby improving feature representation and model generalization. It proves to be effective for edge detection task[19]. Considering that using multiple parallel convolutional layers for feature extraction increases parameters and inference time. Therefore, we explore leveraging the additive properties of convolutional layers to simplify multiple convolutions into one standard convolution. Furthermore, to further enhance the fusion effect of the results from different convolutions, we designed a simple strategy that gives DEFRB the characteristic of dynamic weight adjustment. The architecture of DEFConv is shown in Figure 2, including the following steps:

Firstly, the input undergoes five parallel convolutions to obtain corresponding outputs, which are then summed to derive the weights (w_1) and biases (b_1). The formula is as follows (the biases are omitted for simplification):

$$w_1 = \sum_{i=1}^{5} w_i \tag{2}$$

where $w_i, i = 1:5$, represent the weights of five parallel convolutions corresponding to different types.

Subsequently, we concatenate this with the results of four differential convolutions along the channel dimension, and reduce the dimension by a factor of 5 using a 1x1 convolution to obtain weights (w_2) and biases (b_2) of the original dimension.

$$w_2 = Conv1\left(Concat\left(w_1, w_2, w_3, w_4, w_5\right)\right) \tag{3}$$

where $Conv1$ denotes the 1x1 convolution operation that reduces the channel dimension. $Concat$ denotes concatenation along the channel dimension.

Then, we add the results obtained after applying the re-parameterization technique separately and calculate the weight factor a through a sigmoid operation.

$$a = \delta\left(Con\left(x, w_1\right) + Con\left(x, w_2\right)\right) \tag{4}$$

where a is the gating weight calculated using the sigmoid function. Con denotes the re-parameterization technique.

Finally, we apply corresponding weights to the results of the reparameterization and sum these weighted outputs and perform a residual operation with the original input x to obtain the final output.

$$out = x + a \cdot Con\left(x, w_1\right) + (1-a) \cdot Con\left(x, w_2\right) \tag{5}$$

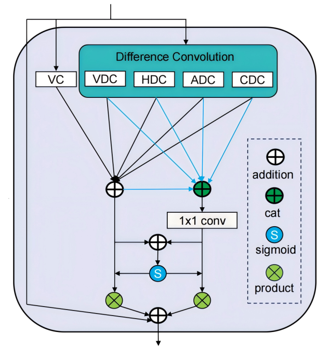


Figure 2. *Diagram of the detail-enhanced fusion convolution (DEFConv). It consists of five parallel deployed convolution layers (one vanilla convolution and four different differential convolutions).*

Local attention

The precise localization of mass area is a prerequisite and key step in FOVs of breast mass segmentation. Typically, the area containing the mass in mammographic images only represents a small proportion of the entire image, implying that most regions in the mammographic images do not significantly contribute to the segmentation task and may even have a negative impact. In traditional methods, it is challenging to locate the mass region in whole mammographic image without ROI data. Moreover, many approaches for locating mass regions require substantial additional computational complexity and complex models.

To address the challenge of accurately localizing small mass targets in FOVs of mammographic images while maintaining a low-complexity network, we introduced the local attention (LA), a lightweight structural design to enhance the accuracy of mass region localization. The structure of LA includes pooling layers, 1D convolutional layers with kernel sizes of 5 or 7, GN layers, and a sigmoid activation function. Specific details are illustrated in the lower right part of Figure 1. Initially, strip pooling [20] is employed to separately capture feature vectors along the horizontal and vertical spatial dimensions. Subsequently, these feature vectors undergo 1D convolution processing to enhance positional information. Following this, the GN layer normalizes the output of the convolution, and finally, precise position attention predictions are generated through a sigmoid function. The specific details are as follows:

Strip pooling was previously used in the coordinate information embedding part of Coordinate Attention (CA)[21], replacing spatial global pooling to capture distant spatial dependencies, leading to substantial enhancements in accuracy, particularly for deeper networks. Therefore, LA applies strip pooling to conduct average pooling across each channel within two spatial dimensions: horizontally (H,1) and vertically (1,W). This produces output representations for channel c with a height of h and channel c with a width of w. These steps can be expressed as:

$$z_c^h(h) = \frac{1}{H} \sum_{0 \le i < H} x_c(h, i) \tag{6}$$

$$z_c^w(w) = \frac{1}{W} \sum_{0 \le i < W} x_c(j, w) \tag{7}$$

where the values z_h and z_w capture not only the global receptive field but also encode precise positional information.

To make effective use of these features, we devised a simple processing method. Given that the positional information embeddings z_h and z_w are sequential signals within the channel, 1D convolutions are more suitable for handling these sequential signals compared to 2D convolutions and are also lighter. Therefore, LA employs 1D convolutions with kernel sizes of 5 or 7 to further enhance its feature extraction capability, effectively strengthening the interaction ability of embedded positional information as well. The adjustment ensures precise identification of the target area by LA. The conventional approach involves using a 7x7 kernel, although slightly heavier in terms of parameters, delivers outstanding performance. To achieve a balance between performance and

Table 1: Comparative Experiments

	INbreast					CBIS-DDSM				
	Dice	mIoU	Acc	Recall	Precision	Dice	mIoU	Acc	Recall	Precision
cGAN [17]	68.69	52.31	-	-	-	-	-	-	-	-
Attention Dense U-Net [12]	-	-	-	-	-	78.55	66.90	**99.81**	83.41	-
DS U-Net [13]	77.00	-	98.70	78.01	-	78.70	-	99.70	78.10	-
YOLO-LOGO [24]	69.37	61.09	-	92.50	-	74.52	64.04	-	**95.70**	-
SCSGNet [14]	61.00	49.94	-	65.50	65.40	48.78	39.04	-	58.48	46.57
Mammo-SAM [25]	75.10	64.88	-	83.35	74.71	61.28	50.68	-	71.56	58.56
AUNet-ASP [15]	74.59	-	98.65	78.16	-	51.48	-	99.43	52.00	-
UNet [9]	81.31	72.65	99.41	74.93	92.25	76.51	67.93	99.61	75.45	77.67
UNet++[11]	77.94	69.20	99.31	71.91	88.73	76.99	68.38	99.60	76.80	75.72
Deeplabv3+ [26]	78.14	69.41	99.36	70.91	93.36	71.70	63.62	99.20	85.52	65.70
AUNet [8]	87.55	79.98	99.55	83.89	92.11	77.96	69.31	99.63	76.70	**79.33**
LADE-Net	**91.69**	**85.70**	**99.70**	**97.47**	**96.99**	**79.13**	**70.46**	99.62	80.49	77.89

parameter count, the optimal group size for 1D convolutions is chosen as the plane's size or the plane's size divided by 8, to cater to diverse requirements.

Subsequently, considering that GN can enhance model generalization to some extent, LA utilizes GN to process the enhanced positional information, obtaining representations of positional attention in the horizontal and vertical directions.

$$y^h = \sigma\left(G_n\left(F_h\left(z_h\right)\right)\right) \tag{8}$$

$$y^w = \sigma\left(G_n\left(F_w\left(z_w\right)\right)\right) \tag{9}$$

where y^h and y^w represent the representations of positional attention in the horizontal and vertical directions, respectively. σ denotes the operation of the sigmoid activation function.

Finally, by utilizing Eq. 10, we can derive the output of the LA module, denoted as Y.

$$Y = x^c \times y^h \times y^w \tag{10}$$

After preliminary experimental validation, we selected the LA-S version and applied it to the encoding and skip connection stages of the model to achieve early and precise localization of breast mass regions.

Experiments

Dataset

This paper validates the proposed LADE-Net using two publicly available datasets (INbreast, CBIS-DDSM). Similar to previous works [15], we randomly split them into training, validation, and test sets in an 8:1:1 ratio.

INbreast [22]: The INbreast dataset contains 107 full-field digital mammograms digital mammograms derived from 115 multi-view cases.

CBIS-DDSM [23]: The CBIS-DDSM dataset contains 1592 images containing masses.

Training details and evaluation metrics

Training details

The experimental data input consists of original RGB images and grayscale labels, adjusted to a standard size of 512×512, without cropping non-breast regions as done in prior works. The resized images were adjusted to the range of [0,1] by dividing by 255 to ensure uniform pixel scaling. To mitigate overfitting, random transformations such as scaling, distortion, flipping, and color changes were applied to both images and labels to enhance dataset diversity.

The code in this study was developed within the PyTorch framework and executed on an RTX 3090 GPU. No pretrained models were used during training. The model was trained using the Adam optimizer, starting with a learning rate of 1 x 10-4, a minimum decay rate of 1 × 10-6, a momentum parameter of 0.9, and a cosine annealing scheduler for learning rate decay. The training was conducted with a batch size of 4 and halted after 150 epochs. Binary cross-entropy (BCE), dice loss, and focal loss were combined to supervise the training process.

$$L = L_{BCE} + L_{Dice} + L_{Focal} \tag{11}$$

where L is the composite loss function used during the training process.

Evaluation metrics

Five common metrics were employed to assess mass segmentation performance, including dice coefficient (Dice), mean intersection over union (mIoU), accuracy (Acc), recall, and precision. In the field of FOVs of breast mass segmentation, Dice is the most commonly used and authoritative metric, followed by mIoU.

Comparative experiments

Table 1 presents a performance comparison between LADE-Net and previous methods as well as related models across two datasets. LADE-Net achieved Dice scores of 91.69%, 79.13%, respectively, outperforming state-of-the-art methods. We also present four examples on INbreast dataset, covering major challenges in FOVs of breast mass segmentation such as small masses, masses with complex boundaries, and rare-shaped masses, as shown in rows 1 - 4 of Figure 3. AUNet achieved relatively high scores in terms of Dice and IoU. However, it still has shortcomings in mass region localization and boundary delineation. For these challenging scenarios, LADE-Net demonstrated good performance.

Ablation experiments

The study conducted a series of ablation experiments across two datasets, primarily aimed at validating the effectiveness of

Table 2: Ablation experiments

	INbreast					CBIS-DDSM				
	Dice	mIoU	Acc	Recall	Precision	Dice	mIoU	Acc	Recall	Precision
Default	87.55	79.98	99.55	83.89	92.11	77.96	69.31	99.63	76.70	79.33
LA-Net	91.20	84.98	99.68	88.09	94.87	77.76	69.11	99.62	77.62	77.89
DE-Net	88.32	80.99	99.58	84.86	92.57	79.02	70.35	**99.65**	78.17	**79.92**
LADE-Net	**91.69**	**85.70**	**99.70**	**97.47**	**96.99**	**79.13**	**70.46**	99.62	**80.49**	77.89

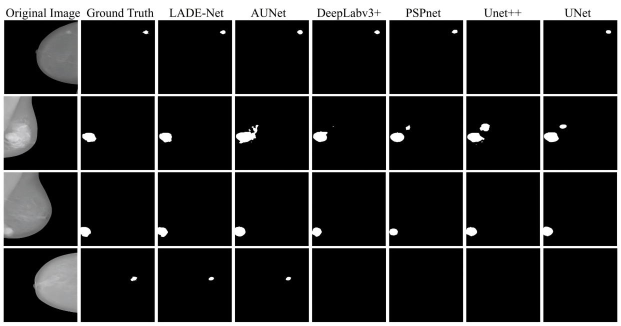


Figure 3. *Comparative results of four samples in INbreast dataset.*

LA and DEFRB in breast mass segmentation. The settings of each experiment are introduced. Here, removing DEFRB means replacing it with a basic block. Detailed experimental results are recorded in Table 2, providing a basis for subsequent analysis and summary.

- Default: Removal of LA and DEFRB from LADE-Net.
- DE-Net: Removal of LA from LADE-Net.
- LA-Net: Removal of DEFRB from LADE-Net.

Firstly, we can observe that the introduction of LA improved the performance of Default, with increases in Dice of 3.65% on the INbreast. The proposed DEFRB led to improvements in Dice scores by 0.77% on the INbreast dataset and 1.06% on the CBIS-DDSM dataset. When both LA and the DEFRB were used simultaneously, the network performance reached its peak, with improvements in Dice over the baseline of 4.14% and 1.17% on the INbreast, CBIS-DDSM. This indicates that LADE-Net can, to some extent, tackle the challenges in current breast mass segmentation.

Conclusion

To address the challenges of small mass region proportions and blurred mass edges in mass segmentation in in whole mammograms, this paper proposes LADE-Net. LADE-Net combines local attention module and detail-enhanced module. LA integrates 1D convolution and group normalization techniques to effectively encode positional information of the regions of interest without sacrificing channel dimensions, thereby improving early localization accuracy of breast masses. DEFRB utilizes detail-enhanced convolutions with dynamic weight adjustment and multi-path fusion mechanisms to refine and enhance edge features of breast masses. A series of experiments on two publicly available datasets (INbreast, CBIS-DDSM) demonstrate that LADE-Net outperforms other baseline methods.

Acknowledgment

This work was supported by the Shandong Natural Science Foundation under Grants ZR2024MF102, ZR2023MF008, the National Natural Science Foundation of China under Grant 62372468, the Major Basic Research Projects in Shandong Province under Grant ZR2023ZD32, the Qingdao Natural Science Foundation under Grant 23-2-1-161-zyyd-jch, the research and development of key technologies for clinical medicine of "heart brain treatment", Yunnan Province, under Grant No.202203AC100007, the State Key Laboratory of Shale Oil and Gas Enrichment Mechanisms and Effective Development under Grant No.33550000-22-ZC0613-0243, and the Jinan Clinical Medical Science and Technology Innovation Plan under Grant 202225018.

References

[1] C. Xia, X. Dong, H. Li, M. Cao, D. Sun, S. He, F. Yang, X. Yan, S. Zhang, N. Li *et al.*, "Cancer statistics in china and united states, 2022: profiles, trends, and determinants," *Chinese medical journal*, vol. 135, no. 05, pp. 584–590, 2022.

[2] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 74, no. 3, pp. 229–263, 2024.

[3] R. Selvi, *Breast diseases: imaging and clinical management*. Springer, 2014.

[4] A. Bleyer and H. G. Welch, "Effect of three decades of screening mammography on breast-cancer incidence," *New England Journal of Medicine*, vol. 367, no. 21, pp. 1998–2005, 2012.

[5] M. S. Bae, W. K. Moon, J. M. Chang, H. R. Koo, W. H. Kim, N. Cho, A. Yi, B. La Yun, S. H. Lee, M. Y. Kim *et al.*, "Breast cancer detected with screening us: reasons for nondetection at mammography," *Radiology*, vol. 270, no. 2, pp. 369–377, 2014.

[6] M. L. Giger, N. Karssemeijer, and J. A. Schnabel, "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer," *Annual review of biomedical engineering*, vol. 15, no. 1, pp. 327–357, 2013.

[7] L. Liberman and J. H. Menell, "Breast imaging reporting and data system (bi-rads)," *Radiologic Clinics*, vol. 40, no. 3, pp. 409–430, 2002.

[8] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D. D. Feng, and S. Wang, "Aunet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Physics in Medicine & Biology*, vol. 65, no. 5, p. 055005, 2020.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
[11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.
[12] S. Li, M. Dong, G. Du, and X. Mu, "Attention dense-u-net for automatic breast mass segmentation in digital mammogram," *Ieee Access*, vol. 7, pp. 59 037–59 047, 2019.
[13] N. R. Rajalakshmi, R. Vidhyapriya, N. Elango, and N. Ramesh, "Deeply supervised u-net for mass segmentation in digital mammograms," *International Journal of Imaging Systems and Technology*, vol. 31, no. 1, pp. 59–71, 2021.
[14] Q. Li, J. Xu, R. Yuan, Y. Zhang, and R. Feng, "Scsgnet: Spatial-correlated and shape-guided network for breast mass segmentation," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.
[15] P. Aliniya, M. Nicolescu, M. Nicolescu, and G. Bebis, "Improved loss function for mass segmentation in mammography images using density and mass size," *Journal of Imaging*, vol. 10, no. 1, p. 20, 2024.
[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
[17] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. Kamal Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig *et al.*, "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network," *arXiv e-prints*, pp. arXiv–1809, 2018.
[18] Z. Chen, Z. He, and Z.-M. Lu, "Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention," *IEEE Transactions on Image Processing*, 2024.
[19] Z. Su, W. Liu, Z. Yu, D. Hu, Q. Liao, Q. Tian, M. Pietikäinen, and L. Liu, "Pixel difference networks for efficient edge detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 5117–5127.
[20] Q. Hou, L. Zhang, M.-M. Cheng, and J. Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4003–4012.
[21] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 713–13 722.
[22] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "Inbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.
[23] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, no. 1, pp. 1–9, 2017.
[24] Y. Su, Q. Liu, W. Xie, and P. Hu, "Yolo-logo: A transformer-based yolo segmentation model for breast mass detection and segmentation in digital mammograms," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106903, 2022.
[25] X. Xiong, C. Wang, W. Li, and G. Li, "Mammo-sam: Adapting foundation segment anything model for automatic breast mass segmentation in whole mammograms," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2023, pp. 176–185.
[26] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

Author Biography

Qingkun Guo is currently pursuing his master's degree at the School of Control Science and Engineering, China University of Petroleum (East China). His research interests include computer vision and medical image processing.

Yixuan Wang is currently pursuing his bachelor's degree at the School of Electronic and Computer Engineering, Shenzhen MSU-BIT University. His research interests is computer vision.

Wenzong Jiang received his Master's degree in Information and Communication Engineering from China University of Petroleum (East China). His research interests include computer vision and medical image processing.

Bin Shen is Engineering Director (AI/ML) at Celonis. He was with the Advanced Technology Group (ATG) of Pinterest from 2019 to 2022. Before that, he was with Google New York from 2015 to 2019. Bin earned B.S. and M.S. degrees from Department of Electronic Engineering, Tsinghua University, Beijing, in 2007 and 2009, respectively, and M.S. and Ph.D. degrees from Department of Computer Science, Purdue University, West Lafayette, Indiana, in 2011 and 2014.

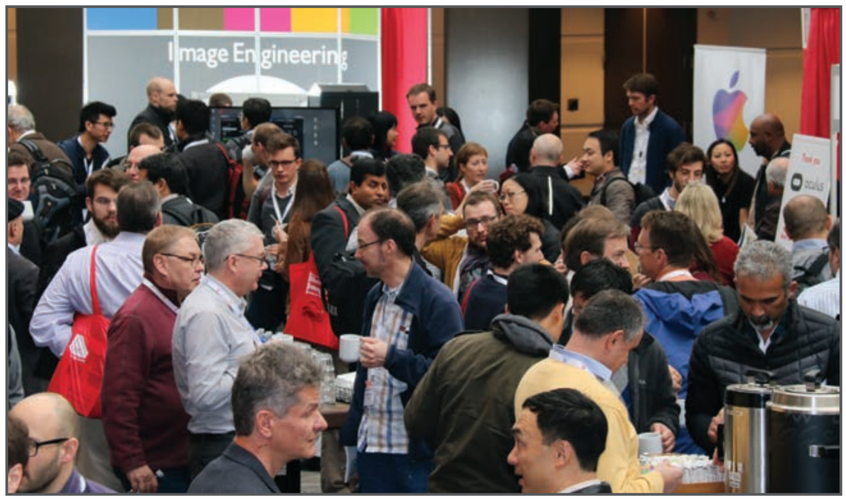
Weifeng Liu is a Professor at the College of Control Science and Engineering, China University of Petroleum (East China). He holds a Ph.D. in pattern recognition and intelligent systems from the University of Science and Technology of China. His research interests include pattern recognition and machine learning, and he has published extensively in top journals. He serves as an associate editor for Neural Processing Letters and co-chair for the IEEE SMC technical committee.

Lin Cong received his Ph.D. in Control Science and Engineering from Zhejiang University, China. He is currently an associate professor at the School of Control Science and Engineering, China University of Petroleum, Qingdao, China. His research interests include intelligent modeling of complex systems.

Baodi Liu received the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China. He is currently an Associate Professor with the College of Control Science and Engineering, China University of Petroleum, Qingdao, China. His research interests include computer vision and machine learning.

JOIN US AT THE NEXT EI!

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org