

# Localization-Guided Semantic Segmentation Framework for Breast Ultrasound Lesions Using BUS\_UCLM

Faycal Arioua<sup>1</sup>, Faycal Touazi<sup>1</sup>, Djamel Gaceb<sup>1</sup>, Tayeb Benzenati<sup>1</sup>, Ibtissem Telkhoukh<sup>1</sup> and Nour Elhouda Magaz<sup>1</sup>

<sup>1</sup>LIMOSE Laboratory, Computer Science Department, University M'hamed Bougara, Independence Avenue, 35000 Boumerdes, Algeria

## Abstract

Breast cancer tumor detection and segmentation in ultrasound imaging is crucial for improving early diagnosis and guiding treatment decisions. This paper presents a two-stage deep learning framework that integrates lesion localization-guided semantic segmentation to enhance tumor detection in breast ultrasound images. In the first stage, YOLOv11-n-nano, a lightweight detection model with approximately 2 million parameters, is employed to localize suspicious regions of interest. In the second stage, advanced segmentation architectures—including U-Net++, U-Net3+, Attention U-Net, and TransUNet—are applied to refine the segmentation of the localized areas. The proposed framework is evaluated on the publicly available BUS\_UCLM dataset. Our approach achieves state-of-the-art performance, with DeepLabV3 obtaining a Dice score of 94.31% and an Intersection over Union (IoU) of 92.10%. These results highlight the effectiveness of localization-guided segmentation in reducing false positives and improving segmentation accuracy in breast ultrasound imaging.

## Keywords

Breast tumor segmentation, Ultrasound Images, Deep Learning, YOLOv11-n, DeepLabV3

## 1. Introduction

According to the World Health Organization, breast cancer is the most commonly diagnosed cancer among women worldwide, with over 2.3 million new cases in 2020 and accounting for approximately 685,000 deaths annually [1]. The diagnosis and treatment of breast cancer rely on multiple steps, including clinical examinations and imaging techniques, to assess tumor presence, stage, and biological characteristics. Among imaging modalities, ultrasound plays a central role in breast cancer diagnosis, especially in young women and in patients with dense breast tissue, where mammography may be less effective [2].

Breast ultrasound imaging offers several advantages: it is non-invasive, cost-effective, portable, and free of ionizing radiation. However, ultrasound suffers from inherent limitations such as speckle noise, low contrast, operator dependency, and variability in acquisition protocols [3]. These factors make accurate interpretation a challenging task, often requiring considerable radiologist expertise. Consequently, the integration of Artificial Intelligence (AI), particularly deep learning, into ultrasound analysis has emerged as a promising direction for automating lesion detection, classification, and segmentation.

In recent years, deep learning techniques, especially convolutional neural networks (CNNs), have demonstrated remarkable success in medical image analysis tasks including classification, detection, and segmentation [4, 5, 6, 7]. Several studies have focused on the diagnosis of breast lesion in mammography and MRI; however, fewer studies have systematically addressed the breast tumor segmentation in

---

*ProfIT AI'25: 5th International Workshop of IT-professionals on Artificial Intelligence, October 15–17, 2025, Liverpool, UK*

\*Corresponding author.

†These authors contributed equally.

✉ f.arioua@univ-boumerdes.dz (F. Arioua); f.touazi@univ-boumerdes.dz (F. Touazi); d.gaceb@univ-boumerdes.dz (D. Gaceb); t.benzenati@univ-boumerdes.dz (T. Benzenati); i.telkhoukh@univ-boumerdes.dz (I. Telkhoukh); nh.magaz@univ-boumerdes.dz (N.E. Magaz)

ORCID 0009-0003-2043-4412 (F. Arioua); 0000-0001-5949-5421 (F. Touazi); 0000-0002-6178-0608 (D. Gaceb); 0000-0002-0123-0022 (T. Benzenati); 0009-0009-2077-9510 (I. Telkhoukh); 0009-0008-0166-1544 (N.E. Magaz)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ultrasound images. This task remains difficult due to factors such as heterogeneous tissue appearance, irregular tumor boundaries, small lesion sizes, and class imbalance between benign, malignant, and normal samples [8, 9].

To address these challenges, incorporating a lesion localization step prior to segmentation has been proposed as an effective strategy. By guiding the segmentation model toward regions of interest, this two-stage approach reduces false positives and improves delineation of lesion boundaries [4]. This design also reflects the clinical workflow, where radiologists typically identify suspicious regions before performing detailed analysis. Furthermore, focusing segmentation on localized regions can alleviate the effects of class imbalance and improve computational efficiency during both training and inference [10]. Ensemble and hybrid deep learning strategies can further enhance robustness and generalization across diverse imaging conditions [11, 12].

In this study, we present a two-stage deep learning pipeline for breast lesion analysis in ultrasound images. The first stage localizes suspicious regions using a YOLO-based model, while the second stage performs precise semantic segmentation with advanced architectures. We evaluate our framework on the BUS\_UCLM dataset, a publicly available breast ultrasound lesion segmentation dataset containing 683 images across benign, malignant, and normal categories. We integrate four state-of-the-art CNN-based segmentation models: U-Net, U-Net++, TransUNet, Attention U-Net and DeepLab. We provide a comprehensive comparison of their performance.

The structure of this paper is as follows: Section 2 reviews related literature on breast tumor analysis in ultrasound imaging, with emphasis on deep learning-based methodologies. Section 3 introduces the proposed methodology, including the two-stage localization-segmentation framework, dataset preprocessing, and network architectures. Section 4 introduces the evaluation metrics and loss functions employed to assess the performance of the different models. Section 5 presents the experimental results and comparisons across different models. Finally, Section 6 concludes the paper with key findings, limitations, and future research directions.

## 2. Related Work

Several studies have explored deep learning approaches for breast cancer detection and segmentation in ultrasound imaging.

Cho et al. [13] proposed a multi-stage segmentation framework combining classification (BTEC-Net, based on DenseNet and ResNet) and segmentation (RFS-UNet with residual and spatial attention). Evaluated on the BUSI and UDIAT datasets, their approach significantly reduced false positives and improved IoU (77%) and Dice (85%), outperforming conventional models such as UNet and PSPNet.

Raza et al. [14] introduced *DeepBreastCancerNet*, a 24-layer CNN with inception modules, enhanced through transfer learning from nine pre-trained networks. Using the BUSI dataset (780 images) and an additional set of 250 images, the model achieved remarkable accuracy (99.35% and 99.63%), surpassing ResNet-50 (98.06%) and GoogLeNet (98.71%).

Mukasheva et al. [15] compared five UNet variants (UNet, Attention UNet, UNet++, DenseInception UNet, Residual UNet) on 780 ultrasound images. DenseInception UNet achieved the highest Dice (0.976) after augmentation, while Attention UNet underperformed, highlighting the influence of noise and preprocessing strategies.

Khaledyan et al. [16] studied UNet-based architectures with preprocessing (CLAHE, augmentation) on the BUSI dataset. Their proposed Sharp Attention UNet achieved Dice 0.93 and accuracy 97.9%, outperforming UNet, Sharp UNet, and Attention UNet.

MohammadiNasab et al. [17] presented a self-supervised multi-task approach (DATTR2U-Net) for Automated Breast Ultrasound (ABUS). Using inpainting and denoising as pretext tasks, their model improved segmentation robustness. On TDSC-ABUS, it outperformed Faster R-CNN and YOLO, though challenges remain for detecting very small lesions.

Madhu et al. [18] proposed UCapsNet, a two-stage model combining U-Net for segmentation and Capsule Networks for classification. On BUSI, it achieved outstanding results (segmentation Dice 99.07%,

classification accuracy 99.22%), surpassing pre-trained CNNs like VGG-19 and ResNet-50.

Almajalid et al. [19] used an enhanced U-Net with preprocessing (denoising, contrast adjustment) and heavy augmentation on 221 images, achieving a Dice score of 82.5%, outperforming other automatic methods.

Vallez et al. [20] introduced the BUS-UCLM dataset (683 annotated ultrasound images). In baseline tests, Mask R-CNN performed best with Dice 77.09% and IoU 65.46%.

Zhang et al. [21] proposed a dual-branch DenseNet-UNet combining classification and segmentation on 1600 images, reaching AUC 0.991 and Dice 89.8%, and showing good generalization on external data.

Vakanski et al. [22] integrated visual attention maps into a modified UNet for 510 images, achieving Dice 90.5%, demonstrating the benefit of medical prior knowledge in guiding segmentation. Pan et al [23] introduces a diffusion-based framework for synthesizing breast ultrasound images, integrating text prompts and a shape-aware mask generator to enhance realism and diversity. It effectively addresses data scarcity, with downstream evaluations showing strong performance boosts. On the BUS\_UCLM dataset, segmentation achieves a peak DSC of 83.5% (Attention UNet at 25% synthetic ratio). Vallez et al. [24] evaluates deep learning approaches for breast ultrasound lesion detection and classification using the BUS\_UCLM dataset. Among the models tested, Sk-UNet achieved the highest performance for multi-class segmentation with a IoU score of 79.2% and Dice of 87.3%.

Table 1 summarizes these related works.

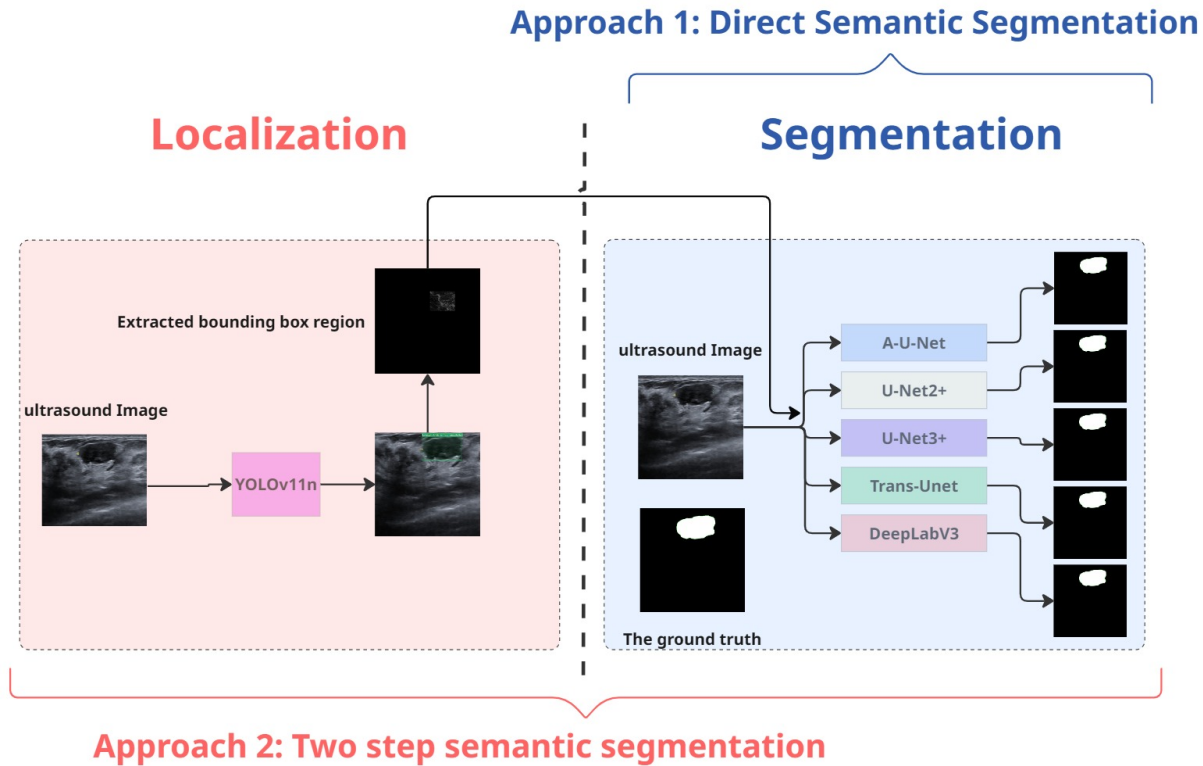
**Table 1**

Comparison of deep learning-based approaches for breast cancer diagnosis from ultrasound images.

Reference	Dataset	Task / Method	Model	Results
Cho et al. [13]	BUSI	Segmentation	RFS-UNet	IoU = 0.77 ; Dice = 0.85
	UDIAT	Classification	MSSE-ResNet101	Acc = 0.97
Asaf Raza et al. [14]	BUSI	Classification	ResNet101	Precision = 0.99
Mukasheva et al. [15]	BUSI	Segmentation	Dense Inception UNet	Dice = 0.97
Donya et al. [16]	BUSI	Segmentation	Sharp Attention U-Net	Dice = 0.93
Poorya et al. [17]	TDSC-ABUS	Segmentation	DATTR2U-Net	Recall = 0.79 ; Precision = 0.56 ; SF = 0.65
Golla Madhu et al. [18]	BUSI	Segmentation	U-Net	IoU = 0.94 ; Dice = 0.95 ; Precision = 0.92
	BUSI	Classification	Capsule Network	Acc = 0.99
Almajalid et al. [19]	221 images	Segmentation	U-Net	Dice = 0.82
Vallez et al. [20]	BUS-UCLM	Segmentation	Mask R-CNN	Dice = 0.77 ; IoU = 0.65
Pan et al. [23]	BUS-UCLM	Segmentation	Diffusion-based framework	Dice = 0.835
Vallze et al. [24]	BUS-UCLM and other datasets	Segmentation	SK-Unet	Dice = 0.873 , IoU=0.792
Zhang et al. [21]	1600 images	Segmentation + Classification	U-Net + DenseNet	AUC = 0.991 ; Dice = 89.8%
Vakanski et al. [22]	BUSI	Segmentation with attention	U-Net-SA ; U-Net-SA-C	Dice = 90.5% ; Jaccard = 83.8%

### 3. Proposed Approach

This section presents two deep learning strategies developed for the automatic segmentation of breast lesions in ultrasound images. The goal is to enhance diagnostic accuracy and robustness by leveraging both direct and localization-guided segmentation approaches.



**Figure 1:** Overview of the proposed segmentation strategies for breast ultrasound. **Approach 1** (right) performs direct semantic segmentation on full ultrasound images using state-of-the-art architectures. **Approach 2** (left) introduces a two-stage pipeline where lesion localization is first achieved via YOLOv11-n, followed by refined segmentation of the extracted region. This hybrid design aims to improve robustness and precision by focusing on regions of interest.

- **Approach 1: Direct semantic segmentation** using deep neural networks such as U-Net++, U-Net3+, Attention U-Net, TransUNet, and DeepLabv3.
- **Approach 2: A localization-guided two-stage segmentation** where YOLOv11-n is used to localize regions of interest, which are subsequently segmented using the same set of models.

Both approaches were implemented and evaluated on the **BUS\_UCLM dataset** (Breast Ultrasound Lesion Segmentation Dataset) [20].

#### 3.1. Approach 1: Direct Semantic Segmentation

This approach applies semantic segmentation directly to full breast ultrasound images, aiming to produce pixel-wise boundaries of tumoral regions. The objective is to generate precise binary masks that distinguish between lesion and non-lesion areas without relying on any prior localization step.

To evaluate this strategy, we implemented and compared five state-of-the-art segmentation models widely used in medical imaging: U-Net++ [25], U-Net3+ [26], Attention U-Net [27], TransUNet [28], and DeepLabv3 [29]. Each model was trained on the BUS\_UCLM dataset and evaluated using standard performance metrics. The results obtained serve as a baseline for comparison with localization-guided segmentation strategies (see Figure 1).

### 3.2. Approach 2: Localization-Guided Semantic Segmentation

In the second approach, we introduce a two-step pipeline to refine segmentation by concentrating on the region of interest (ROI) of the lesion area identified in the original image. This strategy aims to reduce false positives and improve precision.

- **Lesion localization:** The YOLOv11-n-nano (YOLOv11-n) model, with around 2 million parameters, is a lightweight detector that delivers accurate results with low computational cost and fast inference. YOLOv11-n is applied to the full ultrasound image to localize bounding boxes around suspicious lesions.
- **Targeted Segmentation:** Detected regions are cropped, resized, and passed to a segmentation model (same as those used in Approach 1) for semantic mask prediction. The resulting mask is then reintegrated into the original image coordinate space.

This guided strategy helps the model concentrate on lesion-relevant pixels, potentially improving performance in noisy or low-contrast ultrasound contexts (see Figure 1).

The localization-guided inference process is summarized in Algorithm 1.

---

#### Algorithm 1: Localization-Guided Segmentation Pipeline using YOLOv11-n

---

**Input:** Preprocessed dataset  $\mathcal{D}$  (ultrasound images  $I$ , ground truth masks  $M_{GT}$ )  
YOLOv11-n detector  
segmentation model  $\mathcal{S}$

**Output:** Segmentation metrics (Dice, IoU, etc.)

```

1 foreach  $I \in \mathcal{D}$  do
2    $ROI = YOLOv11 - n(I)$  // detect region(s) of interest;
3   if no localization then
4     Predict empty mask;
5   else
6     1.  $I_c =$  Crop image using predicted  $ROI$  bounding box;
7     2.  $M_p = \mathcal{S}(I_r)$  //Apply segmentation model to obtain binary lesion mask;
8     3. Reproject segmented mask into original image dimensions;
9     4. Compute [Dice, IoU, Precision, Sensitivity, Recall]( $M_p, M_{GT}$ );
6 Return: [Dice, IoU, Precision, Sensitivity, Recall];

```

---

### 3.3. BUS\_UCLM Dataset

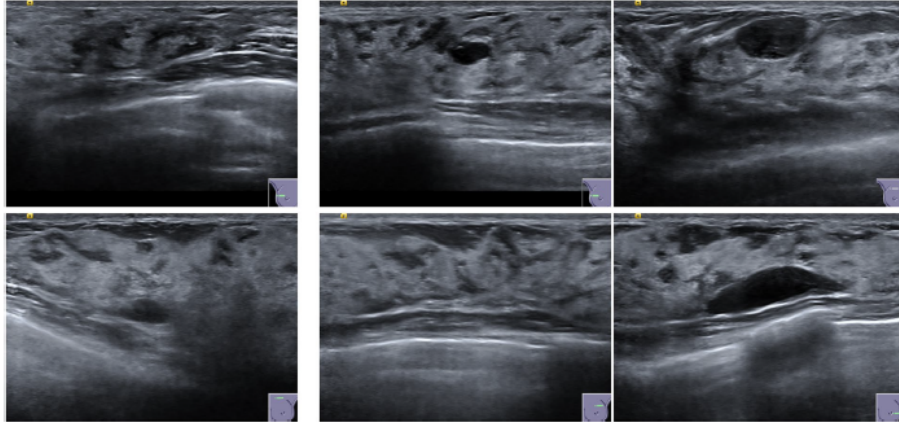
The dataset we selected for ultrasound images in our approaches, BUS\_UCLM (*Breast ultrasound lesion segmentation dataset*) [20], contains a total of 683 images acquired from 38 patients, divided into three categories: benign (174), malignant (90), and normal (419).

We split the dataset into 80% for the training set and 20% for the test set.

**Table 2**

Technical details of the BUS\_UCLM dataset

<b>Dataset name</b>	BUS_UCLM (Breast ultrasound lesion segmentation dataset)
<b>Medical center</b>	University of Castilla-La Mancha
<b>Total number of images</b>	683 ultrasound images
<b>Original format</b>	2D (.png)
<b>Total number of patients</b>	38
<b>Annotations</b>	Segmentation masks + bounding box coordinates
<b>Main tasks</b>	Semantic segmentation & Object detection (YOLO)
<b>Usage</b>	Training, validation, testing



**Figure 2:** Example of ultrasound images collected from the same patient, showing different transverse breast slices extracted from the BUS-UCLM dataset

### 3.4. Preprocessing of the BUS\_UCLM Dataset

Data preprocessing is an essential step to ensure the quality and consistency of the dataset, while also improving model performance and reducing the risk of overfitting. For the BUS\_UCLM dataset, which consists of breast ultrasound images with corresponding lesion annotations, we applied the following preprocessing steps:

- **Resizing:** All images were resized to a fixed resolution of  $224 \times 224$  pixels, which is a common requirement for convolutional neural network (CNN) architectures.
- **Normalization:** Pixel intensity values were normalized to the range  $[0, 1]$  in order to stabilize training and ensure consistency across images.
- **Tensor Conversion:** Images were converted into tensors, enabling efficient processing by deep learning models during training and inference.
- **Cropping Regions of Interest (ROI):** In the second approach, we applied cropping to focus on tumor-specific regions of interest. Using the provided ground-truth masks, the regions containing lesions were extracted and cropped from the original ultrasound images.
- **Data Augmentation:** To improve robustness and reduce overfitting, data augmentation techniques were applied. Specifically:
  - **Random Horizontal Flip:** Applied to both the image and its corresponding mask with a probability of 50%.
  - **Random Rotation:** Images and masks were rotated by a random angle in the range  $[-10^\circ, +10^\circ]$ . This augmentation was mainly used in the second approach to enhance model generalization on the cropped ROIs.

## 4. Evaluation Metrics and Loss Functions

To assess the effectiveness of our proposed approach for breast cancer diagnosis, we employed a set of standard evaluation metrics, including accuracy, F1-score, sensitivity, and precision, which are defined below. These metrics were used to quantify the performance of both the detection and segmentation components of the pipeline. In addition, appropriate loss functions were selected to guide model optimization during training.

### 4.1. Classification Metrics

- **Accuracy:**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:**

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall** (also called Sensitivity or True Positive Rate):

$$\text{Recall} = \frac{TP}{TP + FN}$$

Represents the counts of correct and incorrect classifications for binary classification problems.

## 4.2. Segmentation Metrics

- **Intersection over Union (IoU):**

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN}$$

- **Dice Coefficient (F1 Score):**

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN}$$

- **Relation between Dice and IoU:**

$$\text{Dice} = \frac{2 \cdot \text{IoU}}{1 + \text{IoU}} \quad \text{and} \quad \text{IoU} = \frac{\text{Dice}}{2 - \text{Dice}}$$

## 4.3. Segmentation Loss Functions

- **Binary Cross Entropy (BCE):**

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [t_i \log(y_i) + (1 - t_i) \log(1 - y_i)]$$

where  $t_i$  is the ground truth label,  $y_i$  the predicted probability.

- **Dice Loss** (derived from Dice coefficient):

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_i y_i t_i}{\sum_i y_i + \sum_i t_i}$$

- **Focal Loss** (to address class imbalance):

$$\mathcal{L}_{\text{Focal}} = -\sum_{i=1}^N (1 - y_i)^\gamma \cdot t_i \log(y_i)$$

where  $\gamma$  is the focusing parameter, and  $y_i$  is the predicted probability.

## 4.4. YOLO Evaluation Metrics

- **Average Precision (AP):**

$$\text{AP} = \int_0^1 \text{Precision}(r) dr$$

- **Mean Average Precision (mAP):**

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i$$

- **mAP@50:**

$$\text{mAP@50} = \text{AP at IoU threshold of 0.50}$$

- **mAP@50:95:**

$$\text{mAP@50:95} = \frac{1}{10} \sum_{\text{IoU}=0.50}^{0.95} \text{AP}_{\text{IoU}}$$

**Table 3**

Performance of segmentation models with the direct approach on BUS\_UCLM ultrasound images.

Model	Loss	Dice (%)	IoU (%)	Recall (%)	Precision (%)
UNet++	0.5044	79.19	66.26	77.44	84.43
UNet3+	0.2905	77.67	64.09	75.02	83.37
Attention U-Net	0.4288	77.41	63.84	80.50	77.76
TransUNet	0.3722	70.18	54.86	69.50	73.65
<b>DeepLabV3</b>	<b>0.2687</b>	<b>84.49</b>	<b>73.57</b>	<b>82.62</b>	<b>88.60</b>

## 5. Experimental Results

### 5.1. Approach 1: Direct Semantic Segmentation

In the first strategy, segmentation models were directly applied to all ultrasound images of the BUS\_UCLM dataset. We evaluated five architectures: UNet++, UNet3+, Attention U-Net, TransUNet, and DeepLabV3.

Table 3 summarizes the obtained results across different metrics (Loss, Dice, IoU, Recall, and Precision). As illustrated, DeepLabV3 significantly outperformed other models, achieving the highest Dice (84.49%) and IoU (73.57%), confirming its strong capability for tumor delineation.

The results demonstrate that DeepLabV3 clearly outperforms the other segmentation models in terms of Dice and IoU, which confirms its ability to capture fine tumor boundaries more effectively. UNet++ also shows competitive performance with a balanced precision and recall, making it a reliable alternative when computational efficiency is considered. In contrast, UNet3+ and Attention U-Net perform moderately, with Attention U-Net providing slightly higher recall but lower precision, indicating a tendency to over-segment. TransUNet achieved the weakest results, suggesting that transformer-based representations may require larger training datasets or additional fine-tuning to handle ultrasound-specific noise and variability. Overall, the direct segmentation approach proves effective, but the variability in performance across models highlights the importance of architecture choice for breast ultrasound analysis.

Figure 3 presents visual examples of segmentation masks generated by the different models, highlighting the superior precision of DeepLabV3 compared to the other architectures.

### 5.2. Approach 2: Localization-Guided Semantic Segmentation

The second approach was designed as a two-step process: (1) lesion detection using YOLOv11, followed by (2) focused segmentation on detected regions with the same five architectures.

YOLOv11 achieved competitive detection performance with an **mAP@50 of 81.8%** and **mAP@50-95 of 65%**, ensuring robust region proposals for the subsequent segmentation stage.

**Table 4**

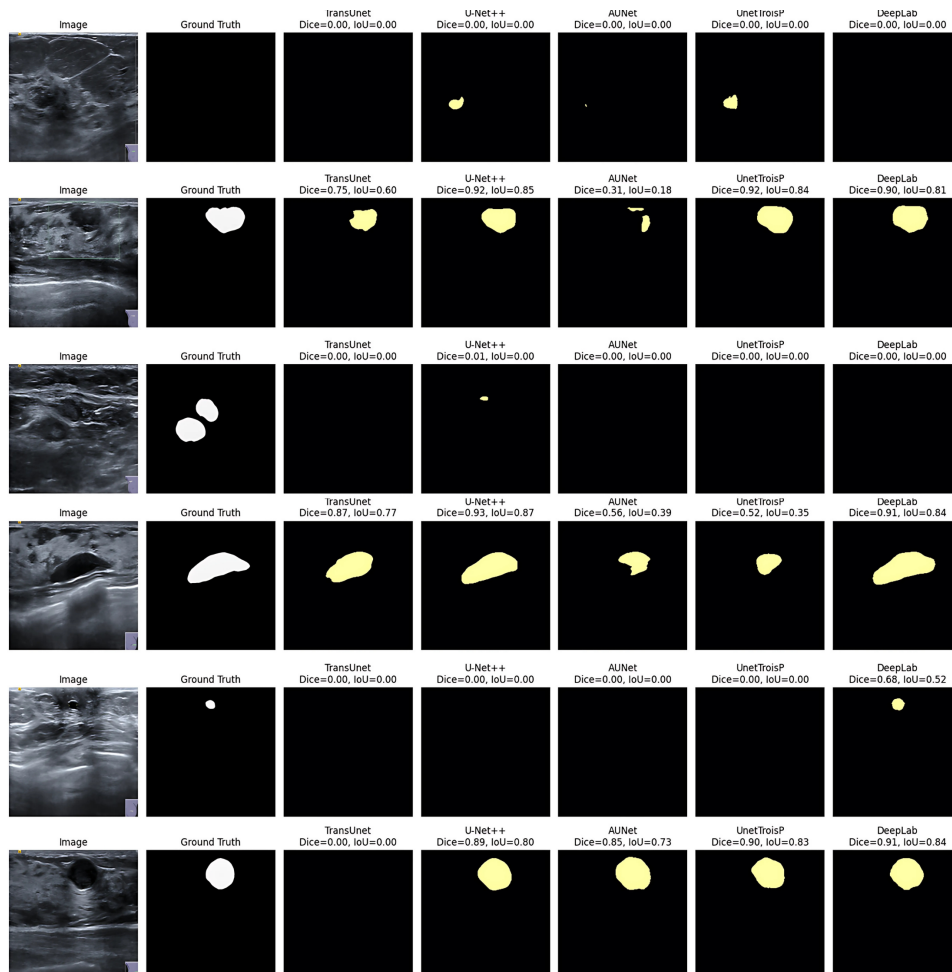
Detection results obtained by the YOLOv11 model on the breast BUS\_UCLM dataset

Metric	Value
mAP@50	81.8%
mAP@50:95	65%

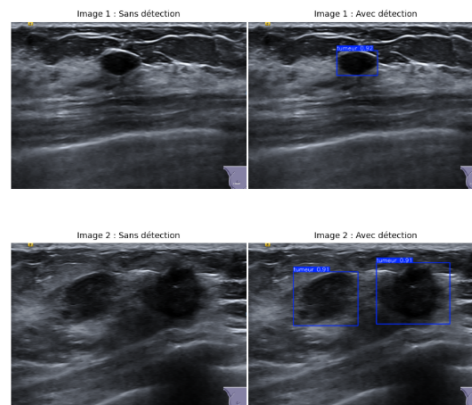
Figure 4 illustrates several examples showcasing the performance of YOLOv11n on BUS\_UCLM dataset images.

Table 5 reports the segmentation performances obtained after YOLOv11 detection. Results show a clear improvement compared to the direct approach. All models achieved Dice and IoU above 91%. DeepLabV3 achieved the best overall performance with a Dice of 94.31% and an IoU of 92.10%.

The results presented in Table 5 indicate that all segmentation models reached very high performances, with Dice scores consistently above 94% and IoU values exceeding 91%. Among the tested architectures,



**Figure 3:** Comparison of segmentation masks obtained with Approach 1: Direct Semantic Segmentation for the different five models.



**Figure 4:** Masses detected with YOLOV11-n on ultrasound images.

DeepLabV3 obtained the best overall results, achieving the highest Dice (94.31%) and IoU (92.10%), confirming its robustness for accurate lesion delineation.

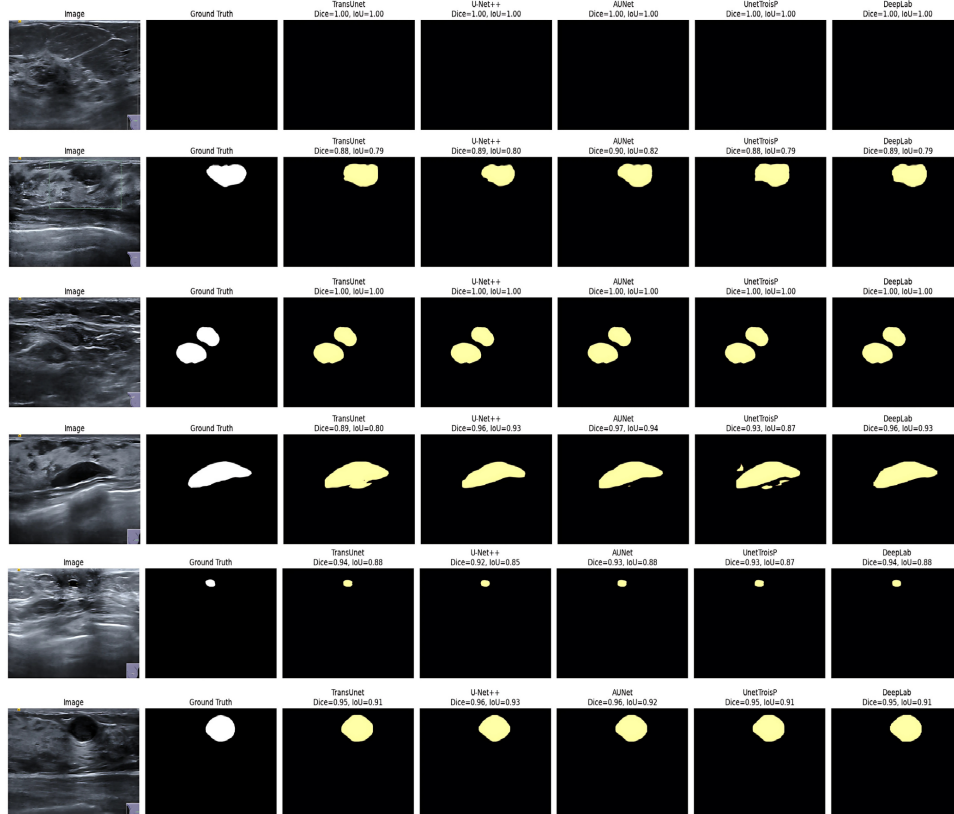
TransUNet also performed competitively, recording the lowest loss (0.3061) and the highest recall (94.93%), which suggests a strong sensitivity in detecting lesion areas. In contrast, Attention U-Net achieved the best precision (94.67%), reflecting its ability to reduce false positives while maintaining high segmentation accuracy. Both UNet++ and UNet3+ yielded stable and reliable performances, with Dice scores of 94.27% and 94.08% respectively, further confirming their effectiveness in this task.

**Table 5**

Performance of segmentation models after YOLOv11 detection (two-stage approach).

Model	Loss	Dice (%)	IoU (%)	Recall (%)	Precision (%)
UNet++	0.3071	94.27	92.03	94.66	94.52
UNet3+	0.3072	94.08	91.74	94.83	94.00
Attention U-Net	0.3063	94.20	91.95	94.38	<b>94.67</b>
TransUNet	<b>0.3061</b>	94.20	91.93	<b>94.93</b>	94.14
<b>DeepLabV3</b>	0.3071	<b>94.31</b>	<b>92.10</b>	94.64	94.58

Figure 5 illustrates segmentation masks generated by the two-stage approach. Compared to the direct segmentation, the contours are more accurate, and false positives are reduced, confirming the benefit of integrating a detection step before segmentation.



**Figure 5:** Comparison of segmentation masks obtained with Approach 2: localization-guided semantic segmentation using different models.

### 5.3. Discussion

Table 6 summarizes the performance of the two segmentation strategies. The direct segmentation approach (Approach 1) provides reasonable results, with DeepLabV3 achieving the highest Dice (84.49%) and IoU (73.57%), demonstrating its capability to capture tumor boundaries more accurately than the other architectures. UNet++ also shows competitive performance, while TransUNet and Attention U-Net are slightly weaker, either in precision or recall.

In contrast, the two-stage, YOLOv11-guided segmentation approach (Approach 2) substantially improves performance across all models. By first localizing lesion regions, the models can focus their learning on relevant areas, resulting in consistently higher Dice and IoU scores (all above 91%). DeepLabV3 again achieves the best overall performance (Dice 94.31%, IoU 92.10%), confirming its robustness for precise tumor delineation. Notably, TransUNet attains the highest recall (94.93%),

indicating strong sensitivity to lesion regions, while Attention U-Net reaches the highest precision (94.67%), reflecting effective reduction of false positives.

Overall, the results highlight the benefits of integrating a detection step prior to segmentation. The two-stage framework not only enhances mask accuracy but also reduces over-segmentation and false positives, making it a reliable strategy for breast ultrasound analysis.

**Table 6**

Comparison of segmentation model performances under both approaches (Direct vs. YOLOv11-Guided) on the BUS\_UCLM dataset.

Approach	Model	Loss	Dice (%)	IoU (%)	Recall (%)	Precision (%)
<b>Approach 1</b>	UNet++	0.5044	79.19	66.26	77.44	84.43
	UNet3+	0.2905	77.67	64.09	75.02	83.37
	Attention U-Net	0.4288	77.41	63.84	80.50	77.76
	TransUNet	0.3722	70.18	54.86	69.50	73.65
	<b>DeepLabV3</b>	<b>0.2687</b>	<b>84.49</b>	<b>73.57</b>	<b>82.62</b>	<b>88.60</b>
<b>Approach 2</b>	UNet++	0.3071	94.27	92.03	94.66	94.52
	UNet3+	0.3072	94.08	91.74	94.83	94.00
	Attention U-Net	0.3063	94.20	91.95	94.38	<b>94.67</b>
	TransUNet	<b>0.3061</b>	94.20	91.93	<b>94.93</b>	94.14
	<b>DeepLabV3</b>	0.3071	<b>94.31</b>	<b>92.10</b>	94.64	94.58

## 5.4. Comparison with Related Work

The results clearly show that our approach substantially outperforms previous methods. While the models from Vallez et al. [20] reach only 68–77% in Dice and around 56–65% in IoU, all of our models exceed 94% in Dice and 91% in IoU, reflecting far more precise and reliable segmentation. Pan et al [23] 83.5% in Dice with Attention UNet at 25% synthetic ratio.

Among our models, DeepLabV3 slightly stands out with the best overall scores, but all others (UNet++, UNet3+, Attention UNet, and TransUNet) perform very closely, confirming the robustness of the proposed approach. In summary, our method brings a significant advancement in breast ultrasound

**Table 7**

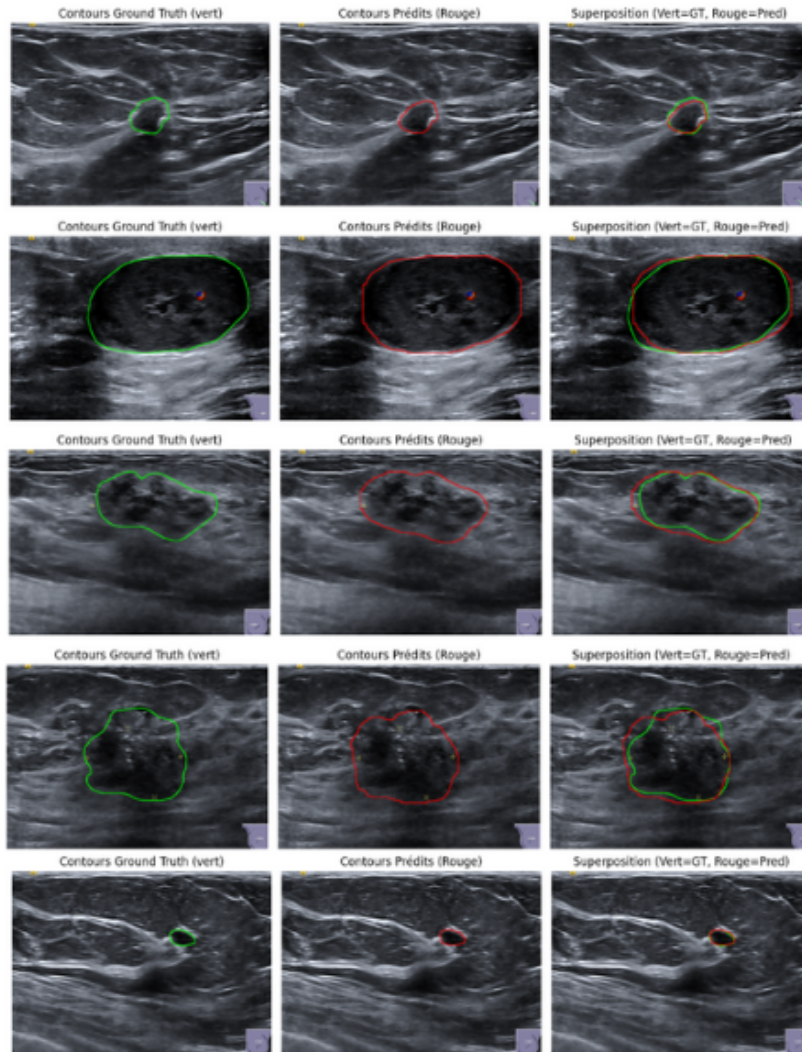
Comparison with related work.

Article	Models	Dice	IoU
Vallez et al. [20]	DeepLabV3	68.46%	56.56%
Vallez et al. [20]	UNet	68.87%	56.94%
Vallez et al. [20]	AttUNet	69.45%	57.53%
Vallez et al. [20]	sk-UNet	69.43%	57.50%
Vallez et al. [20]	Mask R-CNN	77.09%	65.46%
Pan et al[23]	Diffusion-based framework	83.50%	–
Vallze et al. [24]	SK-Unet	87.3%	79.2%
Our approach	UNet3+	94.08%	91.74%
Our approach	Attention UNet	94.20%	91.95%
Our approach	TransUNet	94.20%	91.93%
Our approach	UNet++	94.27%	92.03%
<b>Our approach</b>	<b>DeepLabV3</b>	<b>94.31%</b>	<b>92.10%</b>

image segmentation by effectively combining detection and segmentation.

## 6. Conclusion and Future Work

Breast tumor segmentation in ultrasound imaging is essential for enhancing early diagnosis and improving patient management. In this study, we introduced a two-stage deep learning framework that



**Figure 6:** Comparison of segmentation contours: Ground truth (Green) vs. Predictions (Red) using DeepLabV3 on breast ultrasound images.

integrates lesion localization with semantic segmentation to better delineate tumors in breast ultrasound images. The pipeline uses YOLOv11-n-nano, a lightweight detector of around 2 million parameters, for efficient and robust lesion localization, followed by leading segmentation architectures—U-Net3+, TransUNet, U-Net++, DeepLabV3, and Attention U-Net—to refine tumor boundaries. Evaluations on the BUS\_UCLM dataset demonstrate that our approach significantly outperforms conventional methods.

The localization-guided segmentation strategy boosts precision by focusing the segmentation models on tumor-relevant areas, reducing false positives and emphasizing critical structures. Among the models tested, DeepLabv3 achieved the highest performance, with a Dice score of 94.31% and an IoU of 92.10%, while TransUNet’s strong results underscore the potential of hybrid convolutional-transformer designs. The proposed two-stage framework significantly enhances breast ultrasound lesion segmentation, achieving a Dice score of 94.31% and reducing diagnostic variability by minimizing manual delineations.

Looking ahead, future work will aim to enhance the clinical adaptability and generalization of this framework through several directions:

- Extending the pipeline to process full ultrasound sequences or volumes, allowing the model to capture contextual continuity between consecutive frames.
- Integrating multimodal ultrasound data—such as Doppler or elastography—to enrich lesion characterization and improve segmentation accuracy.

- Validating the model across additional ultrasound datasets and acquisition protocols to ensure robustness and broad applicability.
- Collaborating with clinicians for user-centered evaluations, assessing the tool's utility in routine diagnostic workflows.
- Optimizing the pipeline for real-time inference and deploying it on edge devices to support point-of-care ultrasound applications.

## Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

## References

- [1] World Health Organization, Breast cancer, WHO Fact Sheets (2021). URL: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [2] A.-R. Wuni, B. Botwe, T. Akudjedu, Impact of artificial intelligence on clinical radiography practice: futuristic prospects in a low resource setting, *Radiography* 27 (2021) S69–S73.
- [3] A. N. Tarekegn, M. Giacobini, K. Michalak, A review of methods for imbalanced multi-label classification, *Pattern Recognition* 118 (2021) 107965.
- [4] F. Touazi, D. Gaceb, M. Chirane, S. Hrzallah, Two-stage approach for semantic image segmentation of breast cancer: Deep learning and mass detection in mammographic images., in: *IDDM, 2023*, pp. 62–76.
- [5] F. Touazi, D. Gaceb, N. Boudissa, S. Assas, Enhancing breast mass cancer detection through hybrid vit-based image segmentation model, in: *International Conference on Computing Systems and Applications*, Springer, 2024, pp. 126–135.
- [6] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, J. Wang, Medical image segmentation: A comprehensive review of deep learning-based methods, *Tomography* 11 (2025) 52.
- [7] F. Touazi, D. Gaceb, C. Belkadi, B. Loubar, A self-supervised learning approach for detecting brca mutations in breast cancer histopathological images, in: N. Shakhovska, J. Jiao, I. Izonin, S. Chrétien (Eds.), *Proceedings of the 7th International Conference on Informatics & Data-Driven Medicine*, Birmingham, United Kingdom, November 14-16, 2024, volume 3892 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 183–195.
- [8] M. Rahimpour, M.-J. Saint Martin, F. Frouin, P. Akl, F. Orhac, M. Koole, C. Malhaire, Visual ensemble selection of deep convolutional neural networks for 3d segmentation of breast tumors on dynamic contrast enhanced mri, *European Radiology* 33 (2023) 959–969.
- [9] S. Saikia, T. Si, D. Deb, K. Bora, S. Mallik, U. Maulik, Z. Zhao, Lesion detection in women breast's dynamic contrast-enhanced magnetic resonance imaging using deep learning, *Scientific reports* 13 (2023) 22555.
- [10] N. e. a. Tajbakhsh, Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation, *Medical Image Analysis* 63 (2020) 101693.
- [11] M. Khaled, F. Touazi, D. Gaceb, Improving breast cancer diagnosis in mammograms with progressive transfer learning and ensemble deep learning, *Arabian Journal for Science and Engineering* 50 (2025) 7697–7720.
- [12] N. Laribi, D. Gaceb, F. Touazi, A. Rezoug, A. Sahad, M. O. Reggai, Ensemble deep learning of cnn vs vision transformers for brain lesion classification on mri images, in: *Proceedings of the International Conference on Intelligent Data and Digital Media (IDDM)*, CEUR-WS.org, 2024, pp. 203–219. URL: <https://ceur-ws.org/Vol-3892/paper15.pdf>, open Access.
- [13] S. W. Cho, N. R. Baek, K. R. Park, Deep learning-based multi-stage segmentation method using

- ultrasound images for breast cancer diagnosis, *Journal of King Saud University-Computer and Information Sciences* 34 (2022) 10273–10292.
- [14] A. Raza, N. Ullah, J. A. Khan, M. Assam, A. Guzzo, H. Aljuaid, DeepBreastCancerNet: A novel deep learning model for breast cancer detection using ultrasound images, *Applied Sciences* 13 (2023) 2082.
- [15] A. Mukasheva, D. Koishiyeva, Z. Suimenbayeva, S. Rakhmetulayeva, A. Bolshibayeva, G. Sadikova, Comparison evaluation of unet-based models with noise augmentation for breast cancer segmentation on ultrasound images., *Eastern-European Journal of Enterprise Technologies* 125 (2023).
- [16] D. Khaledyan, T. J. Marini, T. M. Baran, A. O’Connell, K. Parker, Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques: Sharp attention unet, *Plos one* 18 (2023) e0289195.
- [17] P. MohammadiNasab, A. Khakbaz, H. Behnam, E. Kozegar, M. Soryani, A multi-task self-supervised approach for mass detection in automated breast ultrasound using double attention recurrent residual u-net, *Computers in Biology and Medicine* 188 (2025) 109829.
- [18] G. Madhu, A. M. Bonasi, S. Kautish, A. S. Almazyad, A. W. Mohamed, F. Werner, M. Hosseinzadeh, M. Shokouhifar, Ucapsnet: A two-stage deep learning model using u-net and capsule network for breast cancer segmentation and classification in ultrasound imaging, *Cancers* 16 (2024).
- [19] R. Almajalid, J. Shan, Y. Du, M. Zhang, Development of a deep-learning-based method for breast ultrasound image segmentation, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018, pp. 1103–1108.
- [20] N. Vallez, G. Bueno, O. Deniz, M. A. Rienda, C. Pastor, Bus-uclm: Breast ultrasound lesion segmentation dataset, *Scientific Data* 12 (2025) 242.
- [21] S. Zhang, M. Liao, J. Wang, Y. Zhu, Y. Zhang, J. Zhang, R. Zheng, L. Lv, D. Zhu, H. Chen, W. Wang, Fully automatic tumor segmentation of breast ultrasound images with deep learning, *Journal of Applied Clinical Medical Physics* 24 (2023) e13863. URL: <https://doi.org/10.1002/acm2.13863>. doi:10.1002/acm2.13863.
- [22] A. Vakanski, S. G. Lee, N. Baker, Attention-based segmentation of breast tumors in ultrasound images using saliency maps, *Sensors* 20 (2020) 2612. doi:10.3390/s20092612.
- [23] H. Pan, H. Lin, Z. Feng, C. Lin, J. Mo, C. Zhang, Z. Wu, Y. Wang, Q. Zheng, Breast ultrasound tumor generation via mask generator and text-guided network: A clinically controllable framework with downstream evaluation, *arXiv preprint arXiv:2507.07721* (2025).
- [24] N. Vallez, I. Mateos-Aparicio-Ruiz, M. A. Rienda, O. Deniz, G. Bueno, Comparative analysis of deep learning methods for breast ultrasound lesion detection and classification, *Physica Medica* 134 (2025) 104993.
- [25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 3–11.
- [26] H. Huang, et al., Unet 3+: A full-scale connected u-net for medical image segmentation, in: *ICASSP 2020 – IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2020, pp. 1055–1059.
- [27] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, *arXiv preprint arXiv:1804.03999* (2018).
- [28] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, *arXiv preprint arXiv:2102.04306* (2021).
- [29] Y. Tang, D. Tan, H. Li, M. Zhu, X. Li, X. Wang, J. Wang, Z. Wang, C. Gao, J. Wang, et al., Rtc\_tonguenet: An improved tongue image segmentation model based on deeplabv3, *Digital Health* 10 (2024) 20552076241242773.