An Improved Pix2pix Generative Adversarial Network Model to Enhance Thyroid Nodule Segmentation

Huda F. AL-Shahad [1,2], Razali Yaakob [1,*], Nurfadhlina Mohd Sharef [1], Hazlina Hamdan [1], and Hasyma Abu Hassan [3]

[1] Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia
[2] College of Science, University of Kerbala, Karbala, Iraq
[3] Department of Radiology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia
Email: gs59967@student.upm.edu.my (H.F.A.); razaliy@upm.edu.my (R.Y.); nurfadhlina@upm.edu.my (N.M.S.); hazlina@upm.edu.my (H.H.); hasyma@upm.edu.my (H.A.H.)
*Corresponding Author

Abstract—**Thyroid nodules are a type of lesion, which doctors often need advanced diagnostic tools to detect and conduct follow-up diagnoses. Supervised deep learning techniques, particularly Generative Adversarial Networks (GANs), have been used to extract essential features, detect nodules and generate thyroid masks. However, these approaches suffer significant challenges in obtaining training data due to the high cost of identifying the cancer area and mode collapse during training. Therefore, this study proposed an improvement to one GAN model, namely, the pixel-to-pixel (pix2pix) model, for thyroid nodule segmentation, where the generator was incorporated with a supervised loss function to address instabilities during GAN training. The model used a generator with an encode-decoder structure inspired by U-Net architecture to produce the mask. The discriminator of the model consists of a multilayered Convolutional Neural Network (CNN) to compare the real and generated masks. In addition, three loss functions, namely, binary cross-entropy loss, soft dice loss and Jaccard loss, combined with loss GAN were used to stabilise the GAN model. Based on the results, the proposed model achieved 97% detection accuracy of the cancer area from the ultrasound thyroid nodule images and segmented it using the stabilised model with a generator loss function value of 0.5. In short, this study showed that the improved pix2pix model produced greater flexibility in nodule segmentation accuracy compared with semisupervised segmentation models.**

Keywords—**thyroid nodules segmentation, ultrasound image, deep learning, generative adversarial networks, pix2pix, loss function**

I. Introduction

A thyroid nodule is the most prevalent disease in the neck and is caused by many factors, such as iodine deficiency and radiation exposure [1]. Thyroid nodules are currently diagnosed using ultrasonography and fine needle aspiration and through the determination of thyroid-stimulating hormone levels in the blood [2]. Due to the small size of thyroid nodules, ultrasound imaging is the preferred tool to diagnose thyroid cancer [3]. Nevertheless, significant analytical expertise is needed to understand ultrasound images due to the poor image quality [4]. Boundary feature-based misdiagnoses may occur from inaccurate segmentation findings. Therefore, precise and accurate thyroid nodule segmentation in clinical applications is essential to differentiate between benign and malignant nodules [5]. The application of deep learning has progressed in image identification, classification and segmentation due to its capacity for self-learning and generalisation [6]. Machine-learning-based approaches, especially deep learning-based methods, may further enhance their segmentation performance, making them the preferred analytical tool for thyroid nodule segmentation if a substantial quantity of marked training data is gathered [7, 8].

The limited amount of labelled data is one of the challenges when diagnosing thyroids using ultrasound images. Therefore, researchers have used unsupervised learning techniques, such as Generative Adversarial Networks (GANs), to segment thyroid regions [9]. The popularity of various GAN-based methods and their variations have aided in the advanced use of deep learning algorithms in medical image processing.

However, GANs also suffer certain drawbacks, such as poor stability, repeatability and interpretability. Another problem with GANs is that all models in the cascade repeatedly extract identical low-level features. Achieving a proper balance between the generator and the discriminator is also a challenge [10].

Therefore, this study aimed to develop an improved pix2pix GAN model for segmenting the cancer region from ultrasound images. The convolution layer of the generator was replaced with a new deep network layer based on the U-Net design that focused on target

Manuscript received April 16, 2024; revised June 12, 2024; accepted September 10, 2024; published January 9, 2025.

doi: 10.12720/jait.16.1.37-48

structures. In this study, three loss functions, namely, binary cross-entropy loss (L_B) [11], soft dice loss (L_S) [12] and Jaccard loss [13], were combined and used for the generator to stabilise the model and assess segmentation quality.

II. Research Background

Thyroid cancer cases have significantly increased worldwide in recent decades [14]. Situated in the front neck area, the thyroid is a vital endocrine gland in the human body [8]. The thyroid nodules consist of two types: benign and malignant nodules. The former is often not treated until symptomatic, whereas the latter requires surgical excision [15]. The thyroid nodule's border and form are the key features for classifying thyroid nodules using ultrasonography. The edge of benign thyroid nodules is often smooth and clearly delineated, contrasting the atypical, poorly defined and vascularised appearance of malignant thyroid nodules [16, 17]. Accordingly, ultrasound image-based thyroid nodule evaluations primarily rely on radiologists' clinical expertise, making the diagnosis findings relatively subjective. Aside from the poor quality, resolution and contrast of the ultrasound images, speckle noises and echo perturbations also pose analysis challenges [18].

It is difficult for doctors to quantify these thyroid characteristics without advanced computer systems. Thyroid nodule and thyroid gland segmentation approaches are required to detect thyroid-related disorders and provide doctors with useful information to help them make the best diagnostic choices [8]. Moreover, automated segmentation may aid in the proper diagnosis of nodules by medical students or less experienced practitioners. Therefore, many researchers have gathered datasets and used artificial intelligence techniques to assist hospitals and medical research centres.

Traditional deep learning networks and hybrid models are two groups of deep learning-based thyroid and nodule segmentation techniques [19, 20]. Previous studies have suggested that various functional modules and networks can be added to the networks for precise segmentation. For instance, the multimodel method is useful to aid patch-based networks [21], whereas cascaded Convolution Neural Networks (CNNs) are recommended to enhance the localisation and segmentation of nodules [22]. Moreover, pretrained CNNs [7] could learn appropriate features and extract the region of interest from images, although this process requires marked training data.

Furthermore, missing or inaccurate automated detection might delay radiologists from making a timely diagnosis or even lead doctors to perform thyroid biopsies at the wrong location. Thus, a segmentation algorithm is more appropriate than an object detection algorithm, which can only describe a nodule's general shape and size. Segmenting regions also provide doctors with a clear visualization of the specific nodule details, enabling accurate determination of the biopsy position [10]. Fig. 1 shows an example of a segmentation nodule. The ultrasound image in Fig. 1(a) shows the thyroid cancer and the surrounding tissue, whereas Fig. 1(b) shows only the area of the thyroid cancer segmented from the images.

Many segmentation and detection algorithms have been investigated in thyroid ultrasound imaging studies. One study found that unsupervised learning techniques, such as GANs, can segment thyroid regions [9]. GANs are deep neural networks involving two simultaneous training networks [23], the generator and discriminator, similar to a semantic segmentation network. Some studies have also discovered that GANs can increase the image's accuracy when segmenting other organs [10]. Instead of identifying the mask as true or false, the discriminator determines the accuracy of the mask overlay using the original picture as a second input [10]. The two networks are also competitive because the discriminator targets the generator and forces it to advance to deceive it.

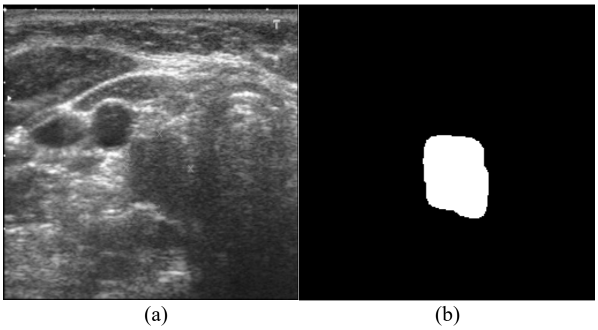


Fig. 1. Thyroid ultrasound images with masks showing (a) the thyroid cancer and the tissues surrounding it and (b) only the area of the thyroid cancer segmented from the images.

Various studies have employed GANs in medical image segmentation. The first reported study used simple adversarial networks for image segmentation [24]. Comparatively, the segmentation of the liver using Three-Dimensional (3D) Computed Tomography (CT) images is more computationally accurate than conventional deep learning segmentation networks. In addition, GANs can enhance the produced modality picture quality with more effective objective performance [25]. GANs also have better control over the morphological and structural details of the lesion in the resulting images at various levels, including brain tumour from Magnetic Resonance Imaging (MRI) images [26], retinal vessel images [27] and CT images for COVID-19 detection [28]. Furthermore, past studies [29–31] have successfully used GAN architecture to segment maligned brain tumours in the nervous system. Another study used U-Net as the generator for GANs used in image segmentation [32].

GANs offer a wide range of potential applications in medical image processing, where pixel-to-pixel (pix2pix) architecture has been used to diagnose low-dose CT [33] and reconstruct the MRI [34, 35]. Moreover, the popularity of various GAN-based techniques has facilitated the advanced use of deep learning algorithms in medical image processing. There are still many limitations to using GANs in image segmentation, such as the instability between training the Generator (G) and Discriminator (D) or the high performance of the discriminator compared with the generator, which struggles with gradient vanishing [23].

Hence, some studies used GANs with additional loss functions for the generator to enhance the model and segment regions accurately, such as in the segmentation of retinal blood vessels [36]. Another study combined the Lovasz hinge loss with the generator loss to improve GANs for segmenting thyroid nodules [10].

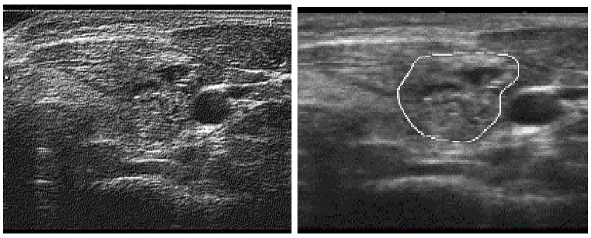
Therefore, this study introduced an improved pix2pix model to segment thyroid cancer from ultrasound images using a small number of marked datasets. The model used a generator with 14 convolutional Two-Dimensional (2D) layers and a combination of three loss functions to increase the stability of the model.

III. Materials and Methods

This section presents an overview of the proposed method in this study. A set of 2D ultrasound images of thyroid lesions and their surroundings were used and the images with their masks were generated by training the proposed pix2pix model.

A. Image Dataset

This work obtained approximately 747 thyroid nodule ultrasound images, which were collected from Hospital Sultan Abdul Aziz Shah in Malaysia (302 images) and an open-access dataset dedicated to thyroid nodule images (445 images) from [37]. Out of the total dataset, 545 images were allocated for training purposes, whereas the remaining 202 images were dedicated to evaluating the effectiveness and accuracy of the proposed pix2pix model. The training images consist of the original ultrasound images and samples of the masked thyroid cancer region, as shown in Fig. 2(a) and (b), respectively. All images were reshaped to 256×256 pixels.



(a) (b)

Fig. 2. Manual segmentation of the thyroid using the (a) original ultrasound image and (b) sample of the masked thyroid cancer region.

B. U-Net

Since 2015, U-Net has been recognised and commonly used as one of the fundamental deep learning architectures in medical image segmentation [38]. It comprises two networks: an encoder network that progressively decreases the input image's spatial resolution while capturing significant features and a decoder network that upscales the features to the original image resolution, ensuring accurate segmentation. Skip connections connect the encoder and decoder layers, enabling the model to preserve intricate spatial details [39]. Fig. 3 shows the standard U-Net architecture [40].

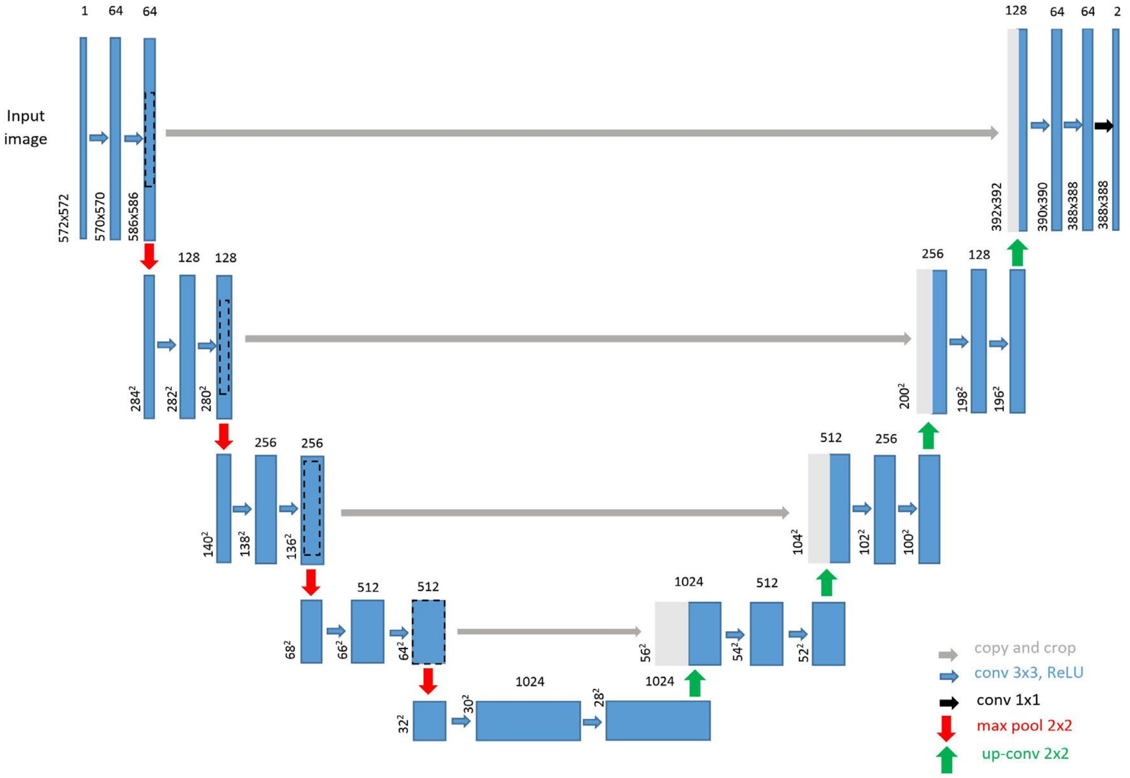


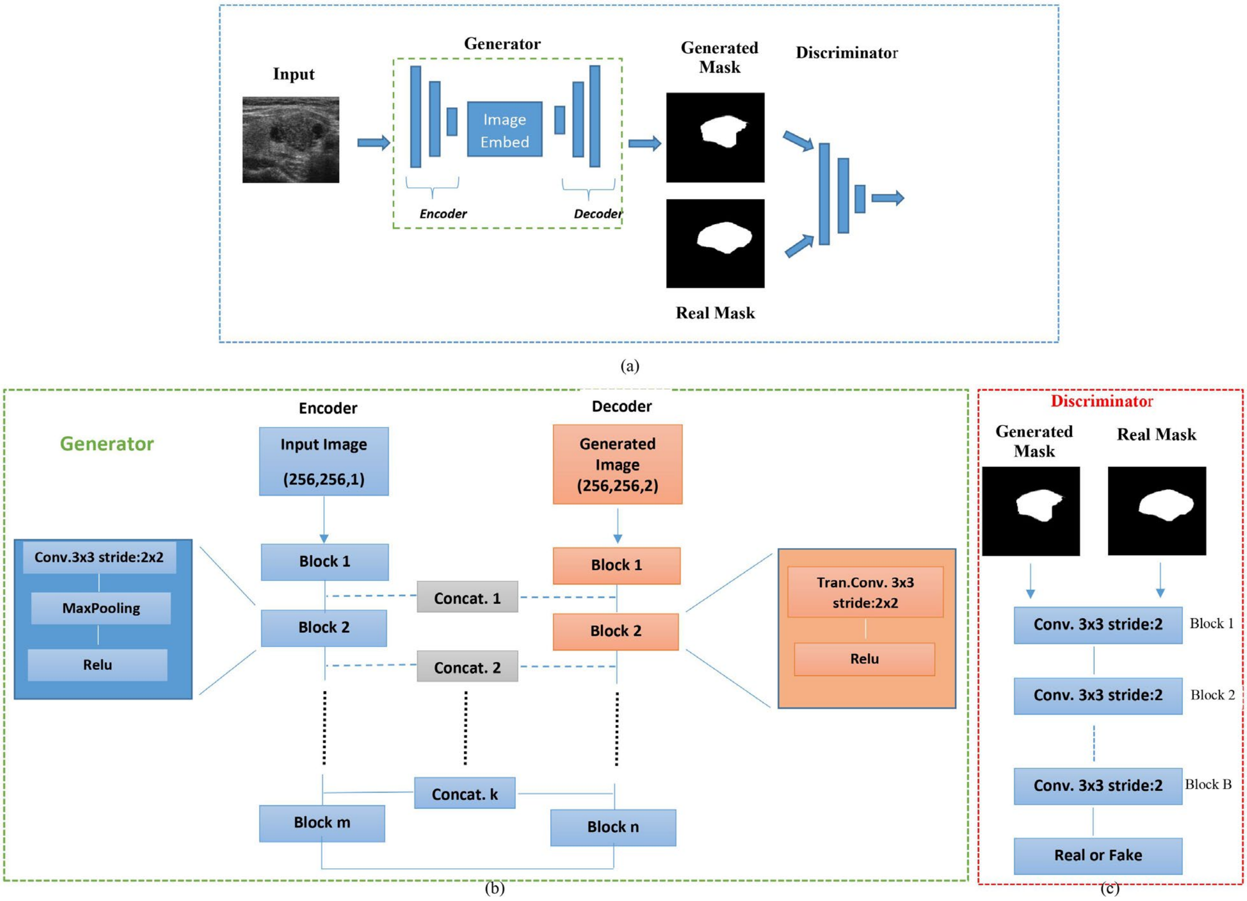
Fig. 3. U-Net standard architecture reprinted [39].

C. Pix2pix Model

The pix2pix model is a conditional GAN model developed in 2016 [41]. It is designed to learn the translation of an input image from one domain to another. The standard structure of pix2pix consists of a U-Net as a generator network and a PatchGAN as a discriminator network [42]. In simple terms, the generator strives to produce an output image consistent with it, whereas the discriminator distinguishes between the generated and real images from the target domain. Through adversarial training, the generator becomes skilled at generating realistic outputs that deceive the discriminator [43].

D. Proposed Pix2pix Model Structure

Fig. 4(a) shows the overall pix2pix GAN model, consisting of the generator and discriminator, as shown in Fig. 4(b) and (c), respectively. As medical image segmentation requires accurate pixel labelling, the output of classical GAN models may be ineffective in producing stable network feedback.



(a)

(b)

(c)

Fig. 4. Proposed pix2pix model: (a) pix2pix GAN model, (b) generator, and (c) discriminator.

Therefore, this study developed a deeper network inspired by a U-Net architecture, which is used as the generator to learn all features and generate masks. In addition, a discriminator with a new architecture was applied to critique the generator and force it to capture the features from the images. The first step of the process involves randomly selecting two images from the dataset: an input image and a reference image that serves as a manual segmentation. The images were transferred into the generator model, which attempts to generate a new mask for the original image as close as possible to the actual mask, which becomes the output of this step. Afterwards, the generated image was fed into the discriminator network to discern between the real and generated masks and determine the authenticity of the generated image (true or false). Meanwhile, Fig. 4(b) shows the setup generator comprising the fundamental blocks, each adopting the encoder–decoder architecture, producing the input image’s mask. There are seven downsampling blocks in the encoder, which were constructed as follows:

- Convolutional 2D layers with a 3×3 kernel size and a stride of 2×2; kernel weights were initialised using the ‘He normal initialisation’. The padding was set to “same”.
- A maxpooling layer with a 2×2 window plays a vital role in downsampling.
- A dropout layer with a dropout rate of 0.5 was applied to the last two blocks.
- The ‘Rectified Linear Unit’ (ReLU) activation function was used, which allows for a slight gradient when the unit is inactive.

Conversely, the decoder blocks were constructed as follows:

- Transposed convolutional 2D layers with a 3×3 kernel size and a stride of 2×2; kernel weights were initialised using the 'He normal initialisation'.
- Both the encoder and decoder blocks used connections for spatial information by concatenating layers that copy the features from the same network level.
- The ReLU activation function was used in the activation layer.

The initial number of kernels per block in the encoder–decoder was 18 before doubling in the subsequent blocks. The discriminator employed a straightforward CNN, as shown in Fig. 4(c). The discriminator was used to validate the generated mask. Each block within the network consisted of multiple layers and each block comprised the following layers:

- A 2D convolutional layer with a 3×3 kernel and a stride of 2×2; kernel weights were initialised using a "He normal initialisation".
- Padding was applied to ensure the output size remained the same as the input size.
- The ReLU activation function was used.

E. Loss Function

The loss function in GAN models is essential in the model's training and performance [12]. The GAN model's loss function comprises separate loss functions for the generator and the discriminator [44], as expressed in Eq. (1). The generator loss function stimulates the generator to produce more realistic samples, which in turn can fool the discriminator. In other words, the generator loss function measures the generator's ability to generate realistic data.

The loss function of the pix2pix model is defined by Eq. (2) and an additional regularisation of the real and generated images [45]. This regularisation component is crucial in image translation tasks to improve the quality of the generated images. This also helps enhance the fidelity of the generated images, making them more comparable to the actual images, as stated in previous work [46].

$$\underset{G}{min}\,\underset{D}{max}\;V_{cGAN}\,(D,G) \;=\; E_{X\sim P_{data}(X)}\,[\,log\,D(x|y)] \,+ \\ E_{z\sim P_z(z)}[\,log\,(1 \;-\; D\big(G(z|y)\big))] \qquad (1)$$

$$V_{Pix2Pix} = \underset{G}{min}\,\underset{D}{max}\;V_{cGAN}\cdot(D,G) + \\ \lambda E_{x,y,z}[\;\|x \;-\; G(z|y)\|_1] \qquad (2)$$

Here, VcGAN refers to the value function of the conditional GAN that pix2pix is inspired by. The generator (*G*) used a random number (*z*) to generate the fake image *G(z)* with a changing label (*y*). Concurrently, the discriminator (*D*) attempts to recognise the real image (x) from the fake image G(z), which uses the fake image space (P_z) to generate a high- quality image close to the real image (x) representing the real image space (P_{data}). The generator model uses a hyperparameter lambda (λ) to control the balance between loss functions and adversarial loss. Lambda represents the coefficient that measures the relative importance of the Generator's Loss (LG), adjusting the lambda value can influence the trade-off between the generator and discriminator objectives.

Goodfellow *et al.* [12] reported that a higher lambda value might prioritise more realistic sample generation (better at fooling the discriminator) at the expense of stability or diversity, potentially leading to mode collapse. In contrast, a lower lambda value may stabilise the training model but generate less realistic samples [12]. Note that lambda has no established or standard value; instead, it is typically found through experimentation, trial and error and consideration of the particular dataset, model architecture and training dynamics.

The gradual changes are essential for both the LG and the Discriminator's Loss (LD) to ensure efficient optimisation. A minimal gradient signal reaches the generator if the discriminator becomes over-confident too early. This phenomenon, known as the vanishing gradient, could lead to an imbalanced training process where the generator cannot learn effectively.

In such situations, the generator repeatedly produces only a single image, called mode collapse. Various methods can be used to overcome mode collapse in GAN training [47]. One approach involves incorporating a supervised loss function into the generator network, effectively transforming the technique into a semisupervised method [10]. Regarding the generator, the supervised loss function employed in this study is similar to the conventional pix2pix GAN. The model used two loss functions but failed to achieve perfect stabilisation. Alternatively, a combination of three supervised learning loss functions, namely, L_S, L_B and Jaccard loss, also called the intersection over union, was used to enhance the model training stabilisation, as expressed in Eq. (3). As GANs try to build the predicted mask through feature learning from the input images and the loss function of the pix2pix generator is L_B [10], it is necessary to measure the similarity between input and predicted images. For this reason, the Jaccard loss function is considered ideal [48]. In addition, the L_S function was used to directly measure the similarity between the predicted and truth masks without setting weights to imbalanced data [49].

The loss functions based on the region are used to solve the problems of unstable training and imbalanced data, especially in the segmentation of small tumours from medical images [50].

$$L_{total} \;=\; L_S + L_B + L_{IOU} \qquad (3)$$

where:

- **L_S** is widely used with segmentation tasks as it measures the overlap between the anticipated and ground truth segmentation masks, as described in Eq. (4). The basis for its calculation is the similarity between the two sets [11].

$$Ls \;=\; 1 \;-\; \frac{2y\hat{p}+1}{y+\hat{p}+1} \qquad (4)$$

- **L_B** measures the difference between the predicted probability and true labels during training, as expressed in Eq. (5); it minimises the divergence between them [12].

$$L_B = -(y\log(\hat{p}) + (1-y)(\log(1-\hat{p}))) \quad (5)$$

- L_{IOU} computes the ratio of the intersection to the union of two sets to determine how similar the two sets are, as explained in Eq. (6). It also works well for activities that need precise border delineation [13].

$$L_{IOU} = 1 - \frac{y.\hat{p} + \varepsilon}{(y + \hat{p} - y.\hat{p}) + \varepsilon} \quad (6)$$

where y represents the true mask and $\hat{p}$ is the prediction mask and stops zero division. By merging the unsupervised GAN loss functions with the supervised L_{total} loss, a comprehensive semisupervised loss for G [51] and stabilisation of the model were achieved, as formulated in Eq. (7).

$$G = arg\min_{G}\max_{D} L_{cGAN}(D,G) + \lambda L_{total} \quad (7)$$

where $L_{cGAN}(D,G)$ is:

$$L_{cGAN}(D,G) = E_{X\sim P_{data}(X)}[\log D(x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z))] \quad (8)$$

where λ is a regularisation parameter between the generator and discriminator if the $\lambda > 0$ allows the loss function to be reduced. This study is based on the work of Meni *et al.* [52], where the lambda value starts from 0.001 and is lightly increased to a more considerable value to see how it affects the model [52].

IV. Experimental Setup

A. Training Parameters

The proposed model was built using the TensorFlow open-source framework (v2.10.1) and implemented in Python (version 3.10) on a Windows 10 operating system using a hardware system equipped with an NVIDIA GeForce RTX3050 GPU. The training process, involving 200 iterations, was accelerated using the Compute Unified Device Architecture. The Adam optimisation function was chosen as the optimiser, with a learning rate of 2e−4 and a batch size of 16. These settings were selected to enhance the model's training efficiency and achieve optimal results.

B. Evaluation Metrics

The primary metrics used to evaluate the quality of the newly generated masks include "Accuracy" (Acc.), "Specificity" (Spec.), "Precision" (Prec.), Recall, "Dice Coefficient" (Dice), and F1-Score. These metrics are commonly used in medical image segmentation evaluations [23] to provide objective measures that assess the rendering quality of different models. Most of these metrics were computed using a confusion matrix, which involves four fundamental components [53–55]: True Positive (TP) rate, True Negative (TN) rate, False Positive (FP) rate and False Negative (FN) rate, as shown through Eqs. (9)–(13).

$$Acc. = \frac{Tp + TN}{TP + TN + FP + FN} \quad (9)$$

$$Spec. = \frac{TN}{TN + FP} \quad (10)$$

$$Prec. = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$Dice = \frac{2TP}{(FP + FN + 2 \times TP)} \quad (13)$$

$$F1\text{-}score = (1 + \beta)\,\frac{(Prec. \times Recall)}{\beta^2(Prec. + Recall)} \quad (14)$$

where $\beta = 1$.

V. Results and Discussion

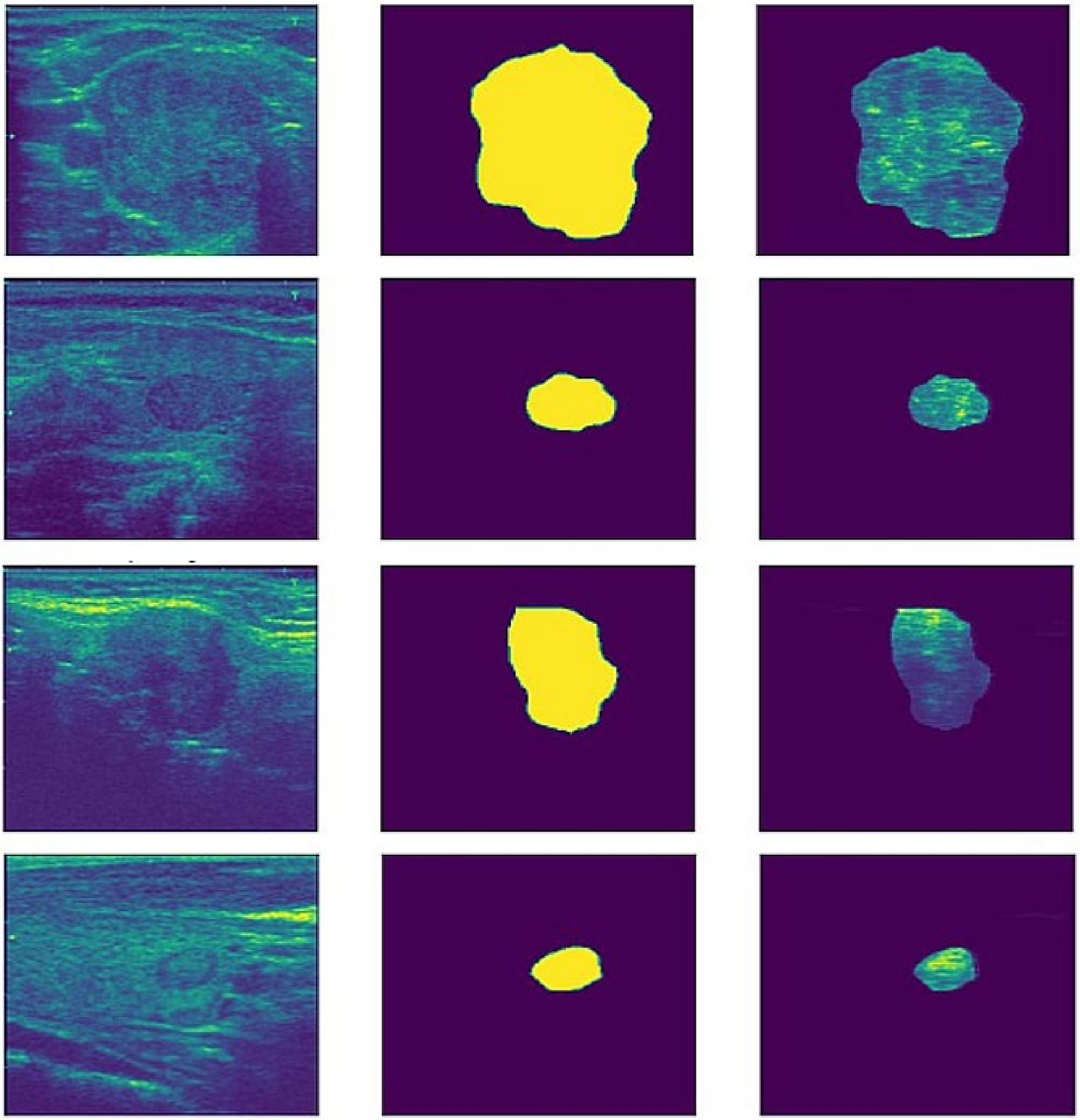
A pix2pix network was developed and trained to segment thyroid cancer from ultrasound images. First, the model generator was processed and stabilised. Subsequently, the model was used with deeper layers to obtain good segmentation. Fig. 5 shows the input images, their true mask and the prediction masks. Fig. 5(a) shows the input image from the ultrasound dataset, whereas Fig. 5(b) shows the true mask of thyroid nodules for the same ultrasound images. Fig. 5(c) displays the prediction masks that resulted from the stabilised model, which is very close to the true mask found in Fig. 5(b). In addition, one of the critical findings of this study indicates that incorporating loss functions into GANs effectively alleviate the mode collapse problem.

Fig. 6 demonstrates the loss function behaviour for the generator and the discriminator. In Fig. 6(a), the soft dice loss function was used for the generator to take the weight of the similarity region between the truth and predicted masks without assigning weight to the unbalanced data. In Fig. 6(b), the Jaccard loss function was used to enhance the convergence between the generator and discriminator, but it is obvious from the oscillating curve in both figures that the model is not stable.

A combination of the soft dice and Jaccard loss functions was used during model training to benefit from both function properties and stabilise the model.

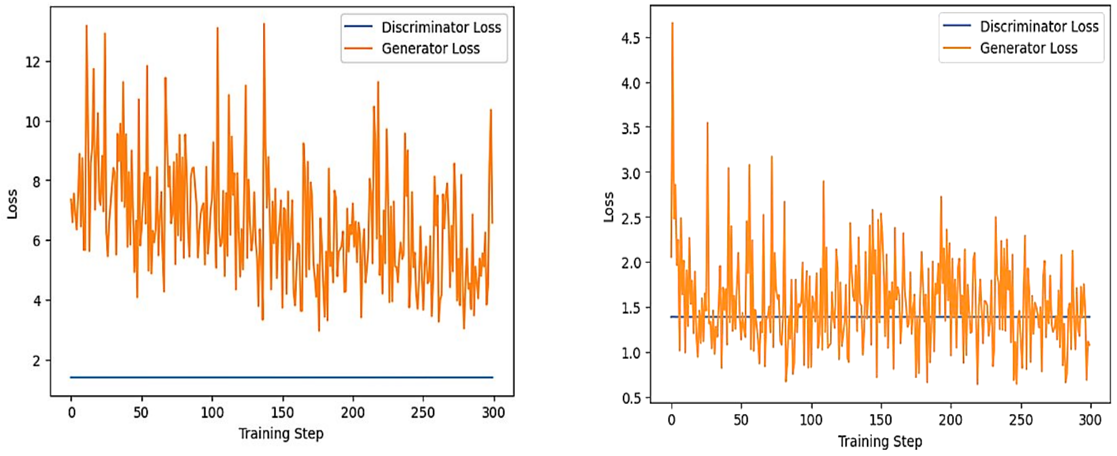
As shown in Fig. 6(c), the value of the generator and discriminator oscillates; the generator loss function curve appeared unstable as it increased and decreased sequentially, even though different hyperparameter (λ) values were used. In addition, the values of the loss function for the discriminator were small; however, the downwards trend was not continuous and the values were not consistent with the generator values.

After combining three loss functions, namely, soft dice, Jaccard and binary cross-entropy (the pix2pix standard functions), the optimal loss decreased, achieving a loss function value of 0.1, whereby the curve continuously decreased when using the same hyperparameter values, as shown in Fig. 6(d).

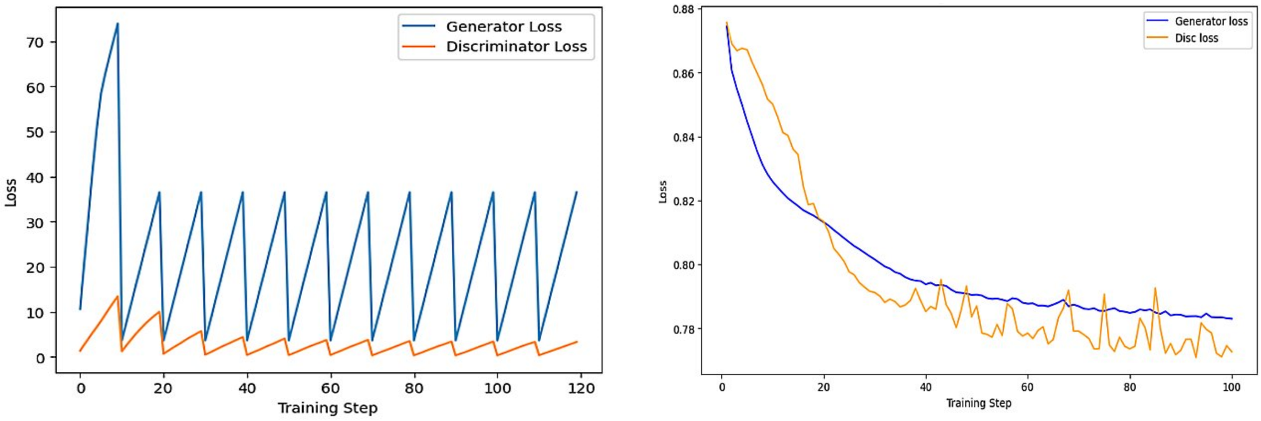


(a) (b) (c)

Fig. 5. Images of thyroid nodules and masks (a) input images based on the original dataset, (b) the true mask originally from the dataset, and (c) the predicted masks that resulted from the proposed model.



(a) (b)



(c) (d)

Fig. 6. Generator and discriminator loss function behaviour. (a) Using soft dice loss function; (b) Using Jaccard loss function; (c) Using two loss functions; (d) Generator and discriminator loss function after stabilization.

The stabilisation of the model means it is possible to control the learning speed between the generator and discriminator by taking advantage of each loss function, as reported in [49].

The stability of the generator also led to a stable discriminator, as the loss value began to decrease and match the LG values until it was completely stable. The difference between the collapse in the accuracy of the generator in the model with unsupervised and supervised functions affected the results during training. The influence of the supervised functions had a more significant effect on the generator's behaviour, affecting the model weights and its updates. However, the effect was less significant after 200 iterations, as the model adapted to the functions. The model adaptation improved the unsupervised function and its effect on the weights.

This is identical to the results reported in [10]. Table I presents the evaluation metrics (Acc., Spec., Prec., F1-score and p-values) for the model performance and the Confidence Interval (CI) ranges for these metrics. CIs have been calculated for each metric in the table, representing a range of values containing lower and upper bounds. These values mean that the true accuracy of the model on the test set lies between them [56]. Table I shows that all the evaluations metrics are between the lower and upper bounds of the CI and that the p-value was 0.029, which means the model's performance was statistically significant because it was less than 0.05.

TABLE I. RESULTS OF THE PIX2PIX MODEL AFTER STABILISATION USING A GENERATOR AND DISCRIMINATOR

Model	Acc.	Spec.	Prec.	F1-Score	*p*-value
Stable pix2pix	97%	94%	93%	92%	0.029
CI	(93.683%, 97.371%)	(65.61%, 93.8%)	(73.438%, 93.188%)	(73.438%, 92.154%)	(60.044%, 97.025%)

Another output of the proposed model is the identification of cancerous areas in the ultrasound images and their subsequent segmentation based on the generated mask. Fig. 7 shows the result of the test ultrasound image, where Fig. 7(a) shows the original input image with the cancer region, whereas Fig. 7(b) shows the radiologist's manually segmented cancer region image. Fig. 7(c) shows the predicted mask from the proposed pix2pix model in this work, whereas Fig. 7(d) presents the mask that was cropped from the original image using the model. The last column in Fig. 7 (e) shows the parts of the images that the model focused on for segmentation using Grad-CAM. Based on Fig. 7, the location of the nodules of the predicted mask (Fig. 7(c)) is similar as the location in the GRAD-CAM (Fig. 7(e)). The model was also tested with clinical ultrasound images from Hospital Sultan Abdul Aziz Shah in Malaysia, using manual segmentation by an experienced radiologist with more than 10 years of experience.

Based on the results, combining more than one loss function led to good model performance in predicting the cancer area, which was almost identical to the radiologist's image contours. Evidently, the proposed pix2pix model was more accurate with the stabilised generator and discriminator, where many λ values were used to determine the best value, as shown in Table II.

The best λ value essentially balanced the combination of loss functions with the standard loss function of the model. The use of small λ values led to converged accuracy in the model before and after stabilisation, achieving good prediction accuracy, although it is essential to note that this does not prevent failure. When higher λ values were used, the results were precise in terms of the difference between the stable and unstable models. As such, the stabilised model recorded an excellent accuracy of up to 97% using λ = 25. The outstanding accuracy supports the proposed model for medical applications to facilitate cancer detection and segmentation of the cancer region.

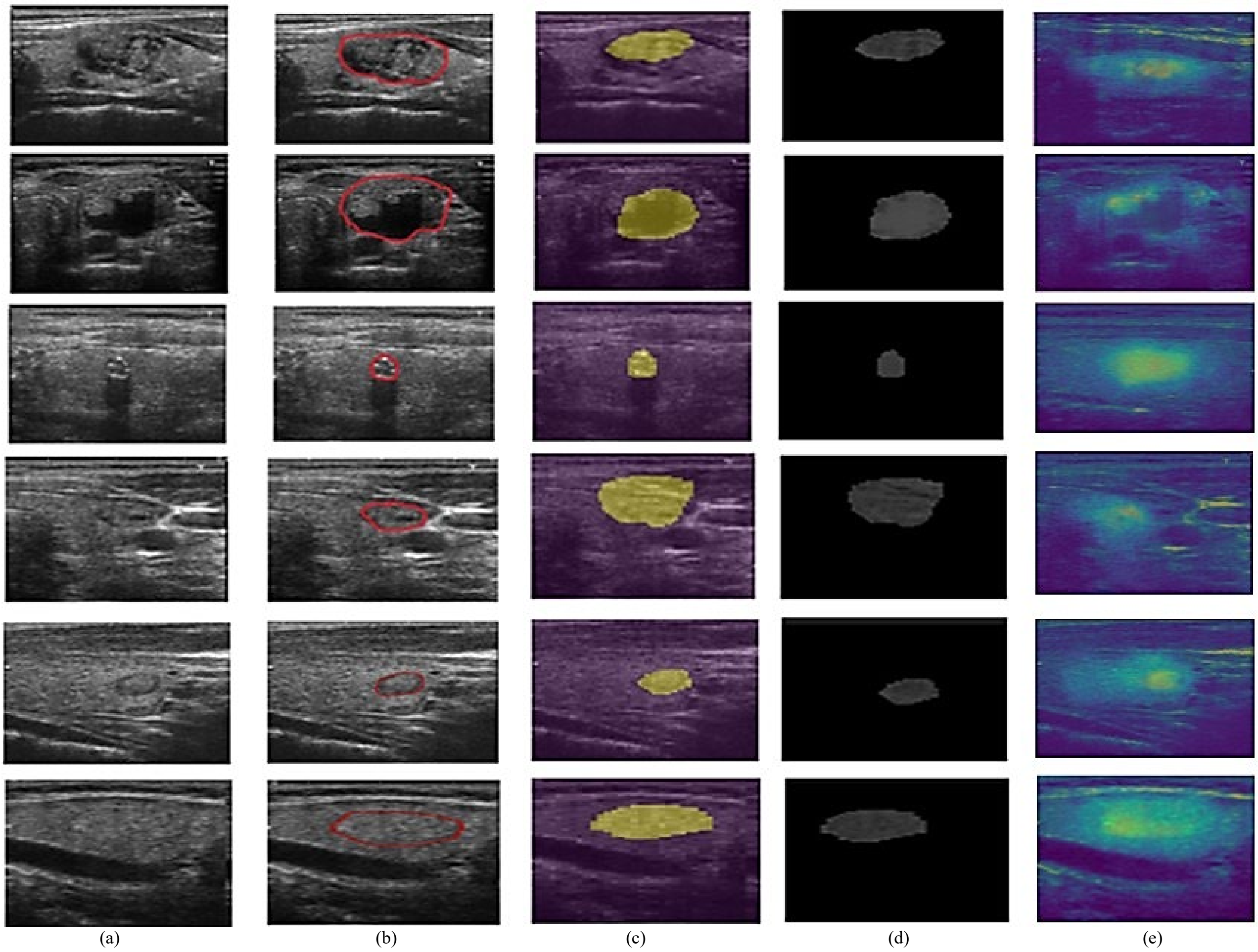


Fig. 7. Comparison masks' results from the proposed model and the radiologist. (a) Input image; (b) Radiologist segmentation; (c) Predicted mask; (d) Cropped mask; (e) Grad-CAM.

TABLE II. EVALUATION MEASUREMENTS OF BOTH MODELS WITH DIFFERENT

Lambda Values	Model	Acc.	Spec.	Prec.	F1-score
λ = 0.001	Unstable pix2pix	88%	88%	89%	88%
	Stable pix2pix	**92%**	**92%**	**95%**	**93%**
λ = 0.01	Unstable pix2pix	**89%**	**89%**	**91%**	**89%**
	Stable pix2pix	**89%**	**89%**	**91%**	**89%**
λ = 0.9	Unstable pix2pix	**92%**	**92%**	**95%**	**93%**
	Stable pix2pix	89%	89%	91%	89%
λ = 5	Unstable pix2pix	72%	71%	81%	71%
	Stable pix2pix	**92**	**92%**	**95%**	**93%**
λ = 15	Unstable pix2pix	77%	77%	75%	75%
	Stable pix2pix	**88%**	**88%**	**89%**	**88%**
λ = 25	Unstable pix2pix	**92%**	**92%**	**95%**	**93%**
	Stable pix2pix	**97%**	**94%**	**93%**	**92%**
λ = 35	Unstable pix2pix	72%	72%	81%	71%
	Stable pix2pix	**92%**	**92%**	**95%**	**92%**

Note: Acc., accuracy; Spec., specificity; Prec., precision; and F1-score

For assessing the model's performance, the 10-fold cross-validation method was employed; it was repeated ten times to make all data represented in training and validation sets. Table III shows accuracies and standard deviations in 10-folds obtained by cross-validation; the average accuracy is about 95%. The high accuracy and relatively low standard deviation of the proposed model demonstrate the stability of the model [57].

TABLE III RESULTS OF TENFOLD CROSS-VALIDATION

10-fold cross-validation	Accuracy	Standard deviation
Fold1	88.78%	0.0001
Fold2	90.90%	1.0600
Fold3	95.80%	2.9399
Fold4	95.80%	3.0728
Fold5	96.58%	3.1330
Fold6	97.11%	3.1493
Fold7	98.08%	3.2220
Fold8	96.81%	3.0921
Fold9	97.44%	3.0158
Fold10	97.97%	2.9747
Average	**95.53%**	**2.9747**

VI. COMPARISON WITH STATE-OF-THE-ART METHODS

In Table IV, recent techniques, namely, TRFE+, FCG-Net and GLFNet, show good performance in medical image segmentation. Using the adaptive region to enhance the performance of nodule segmentation, TRFE+ achieved a high accuracy of 92% and an F1-score of 72%. This is because the adaptive region prior guidance module was used to take full advantage of the thyroid region features. FCG-Net can extract multiresolution features while reducing the number of parameters by using a full-scale skip connection, resulting in an accuracy and F1-score of

95% and 82%, respectively. In contrast, the improvement of the model was small in the segmentation of the ultrasound image because the model requires more clinical verification. GLFNet combined local and global features that were extracted from the image using the self-attention convolution fusion block. Even though the model accuracy was high, the proposed model in this study outperforms this and other state-of-the-art segmentation networks. The accuracy and F1-score of the proposed model were 97% and 92%, respectively. Furthermore, the visual results that were assessed by the experienced specialist proved that the present model is the most suitable segmentation technique for thyroid nodules.

TABLE IV. COMPARISONS BETWEEN THE PROPOSED MODEL AND THE STATE-OF-THE-ART SEGMENTATION MODELS

Model	Ref.	Year	Acc.	F1-Score	Dice
TRFE+	[58]	2022	92%	72%	75.37%
FCG-Net	[59]	2023	95%	82%	80.42%
GLFNet	[60]	2024	96%	74%	74.62%
Proposed model			**97%**	**92%**	**87.97%**

The proposed model performed well because efficiently took advantage of GAN architecture for learning the model using a small number of labelled images. The other reason is that using three loss functions to stabilise the model, the model took the similarity region between the truth and predicted masks to benefit from the convergence between the generator and discriminator. This step led to better segmentation for the contour of the thyroid nodules, similar to the manual segmentation of the radiologist.

The proposed model was compared with the standard pix2pix model with a U-Net generator and one loss function, as in Table V. Notably, a 97% accuracy was achieved in contrast with the standard model, in which the accuracy was achieved at 91%; this is because the stable pix2pix had more than one loss function that led to stabilising and enhancing the generator behaviour.

TABLE V. PERFORMANCE COMPARISONS FOR THYROID NODULE SEGMENTATION MODEL

Model	Acc.	Spec.	Prec.	F1-Score	*p*-value	CI	Effect sizes
Standard Model	91%	91%	90%	90%	0.018	98%	0.06
Stable pix2pix	97%	94%	93%	92%	0.029	97%	0.35

Note: Acc., accuracy; Spec., specificity; Prec., precision; and CI, confidence intervals

VII. CONCLUSION

This study successfully introduced an improved novel algorithm based on the pix2pix model to enhance the segmentation of thyroid nodules in ultrasound images and reduce the manual labelling workload associated with medical image segmentation. The proposed method focused on accurately delineating cancerous regions by employing a deeper version based on the pix2pix model. The use of three types of loss functions enabled the generator and discriminator to reach a stable equilibrium, whereas extensive experiments carried out to validate the proposed model's effectiveness and robustness demonstrated its superiority both visually and statistically. The findings also recorded enhanced nodule segmentation by incorporating unsupervised loss functions to prevent generator collapse. By leveraging this technology, a favourable balance between performance enhancement and reduction in annotation costs was achieved, leading to a fully automated computer-assisted segmentation system for thyroid ultrasound images without human intervention. Future efforts should aim to enhance the model's capability to segment and classify diseased organ parts simultaneously. The model's accuracy can also be improved by augmenting the dataset size.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Huda F. AL-Shahad: Writing—original draft, Visualization, Validation, Investigation. Razali Yaakob: Writing—review & editing, Validation. Nurfadhlina Mohd Sharef: Writing—review. Hazlina Hamdan: Writing—review. Hasyma Abu Hassan: Data Collection, Validation. All authors approved the final version.

REFERENCE

[1] Z. Li *et al.*, "A weakly supervised deep active contour model for nodule segmentation in thyroid ultrasound images," *Pattern Recognit. Letters*, vol. 165, pp. 128–137, Jan. 2023. doi: 10.1016/j.patrec.2022.12.015

[2] M. Zhou *et al.*, "Automatic malignant thyroid nodule recognition in ultrasound images based on deep learning," in *Proc. E3S Web Conferences*, Sep. 2020, vol. 185, 03021. doi: 10.1051/e3sconf/202018503021

[3] J. Song *et al.*, "Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules," *Medicine (Baltimore).*, vol. 98, no. 15, e15133, Apr. 2019. doi: 10.1097/MD.0000000000015133

[4] M. B. Gulame, V. V Dixit, and M. Suresh, "Materials today: Proceedings Thyroid nodules segmentation methods in clinical ultrasound images : A review," *Mater. Today Proceeding*, vol. 45, pp. 2270–2276, 2021. doi: 10.1016/j.matpr.2020.10.259

[5] F. N. Tessler, W. D. Middleton, and E. G. Grant, "Thyroid Imaging Reporting and Data System (TI-RADS): A user's guide," *Radiology*, vol. 287, no. 1, pp. 29–36, Apr. 2018. doi: 10.1148/radiol.2017171240

[6] Z. Akkus *et al.*, "Deep learning for brain MRI segmentation: State of the art and future directions," *Journal of digital imaging*, vol. 30, no. 4, pp. 449–459, Aug. 2017. doi: 10.1007/s10278-017-9983-4

[7] Y. Sharifi *et al.*, "Deep learning on ultrasound images of thyroid nodules," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 636–655, Apr. 2021. doi: 10.1016/j.bbe.2021.02.008

[8] J. Chen, H. You, and K. Li, "A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images," *Computer methods and programs in biomedicine*, vol. 185, 105329, Mar. 2020. doi: 10.1016/j.cmpb.2020.105329

[9] K. K. D. Ramesh *et al.*, "A review of medical image segmentation algorithms," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 7, no. 27, 169184, Jul. 2018. doi: 10.4108/eai.12-4-2021.169184

[10] A. Kunapinun *et al.*, "Improving GAN learning dynamics for thyroid nodule segmentation," *Ultrasound in Medicine and Biology*, vol. 49, no. 2, pp. 416–430, Feb. 2023. doi: 10.1016/j.ultrasmedbio.2022.09.010

[11] C. H. Sudre *et al.*, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," *Springer*, vol. 10553, pp. 1–8, 2017. doi: 10.1007/978-3-319-67558-9_28

[12] I. Goodfellow *et al.*, "Generative adversarial networks," *Communications of the Acm*, vol. 63, no. 11, pp. 139–144, 2020. doi: 10.1145/3422622

[13] D. Duque-Arias *et al.*, "On power jaccard losses for semantic segmentation," in *Proc. the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, SCITEPRESS—Science and Technology Publications, 2021, pp. 561–568. doi: 10.5220/0010304005610568

[14] S. W. Kwon *et al.*, "Ultrasonographic thyroid nodule classification using a deep convolutional neural network with surgical pathology," *Journal of digital imaging*, vol. 33, no. 5, pp. 1202–1208, 2020. doi: 10.1007/s10278-020-00362-w

[15] E. Papini *et al.*, "Minimally-invasive treatments for benign thyroid nodules: A delphi-based consensus statement from the Italian Minimally-Invasive Treatments of the Thyroid (MITT) group," *International Journal of Hyperthermia*, vol. 36, no. 1, pp. 375–381, Jan. 2019. doi: 10.1080/02656736.2019.1575482

[16] B. R. Haugen *et al.*, "2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer," *Thyroid*, vol. 26, no. 1, pp. 1–133, Jan. 2016. doi: 10.1089/thy.2015.0020

[17] Q. Yang *et al.*, "Biomedical signal processing and control DMU-Net : Dual-route mirroring U-Net with mutual learning for malignant thyroid nodule segmentation," *Biomedical Signal Processing and Control*, vol. 77, 103805, Apr. 2022. doi: 10.1016/j.bspc.2022.103805

[18] Q. Kang *et al.*, "Thyroid nodule segmentation and classification in ultrasound images through intra- and inter-task consistent learning," *Medical Image Analysis*, vol. 79, 102443, Jul. 2022. doi: 10.1016/j.media.2022.102443

[19] C. Chu, J. Zheng, and Y. Zhou, "Ultrasonic thyroid nodule detection method based on U-Net network," *Computer Methods and Programs in Biomedicine*, vol. 199, 105906, Feb. 2021. doi: 10.1016/j.cmpb.2020.105906

[20] J. Wu *et al.*, "Ultrasound image segmentation of thyroid nodules based on joint up-sampling," *Journal of Physics: Conference Series*, vol. 1651, no. 1, 012157, Nov. 2020. doi: 10.1088/1742-6596/1651/1/012157

[21] M. Liu *et al.*, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *Neuroimage*, vol. 208, 116459, Mar. 2020. doi: 0.1016/j.neuroimage.2019.116459

[22] M. Wang *et al.*, "Automatic segmentation and classification of thyroid nodules in ultrasound images with convolutional neural networks," *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data MICCAI*, Springer, vol. 12587, pp. 109–115, 2021. doi: 10.1007/978-3-030-71827-5_14

[23] A. Iqbal *et al.*, "Generative adversarial networks and its applications in the biomedical image segmentation : A comprehensive survey," *International Journal of Multimedia Information Retrieval*, vol. 11, no. 3, pp. 333–368, 2022. doi: 10.1007/s13735-022-00240-x

[24] S. Xun *et al.*," Generative adversarial networks in medical image segmentation: A review," *Computers in Biology and Medicine*, vol. 140, 105063, Jan. 2022. doi: 10.1016/j.compbiomed.2021.105063

[25] S. Altun and M. Fatih, "Biomedical signal processing and control brain MRI high-resolution image creation and segmentation with the new GAN method," *Biomedical Signal Processing and Control*, vol. 80, no. P1, 104246, 2023. doi: 10.1016/j.bspc.2022.104246

[26] L. Zhu *et al.*, "DualMMP-GAN: Dual-scale multi-modality perceptual generative adversarial network for medical image segmentation," *Computers in Biology and Medicine*, vol. 144, 105387, 2022. doi:10.1016/j.compbiomed.2022.105387

[27] J. S. Chen *et al.*, "Deepfakes in ophthalmology applications and realism of synthetic retinal images from generative adversarial networks," *Ophthalmology Science*, vol. 1, no. 4, 100079, 2021. doi: 10.1016/j.xops.2021.100079

[28] J. Zhang *et al.*, "Dense GAN and multi-layer attention based lesion segmentation method for COVID-19 CT images," *Biomedical Signal Processing and Control*, vol. 69, 102901, Aug. 2021. doi: 10.1016/j.bspc.2021.102901

[29] S. Nema *et al.*, "RescueNet: An unpaired GAN for brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 55, 101641, 2020. doi: 10.1016/j.bspc.2019.101641

[30] G. Cheng, H. Ji, and L. He, "Correcting and reweighting false label masks in brain tumor segmentation," *Medical Physics*, vol. 48, no. 1, pp. 169–177, 2021. doi: 10.1002/mp.14480

[31] Y. Li, Y. Chen, and Y. Shi, "Brain tumor segmentation using 3D generative adversarial networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 35, no. 4, 2021. doi: 10.1142/S0218001421570020

[32] Y. Xue *et al.*, "SegAN: Adversarial network with multi-scale L 1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, no. 3–4, pp. 383–392, 2018. doi: 10.1007/s12021-018-9377-x

[33] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018. doi: 10.1109/TMI.2017.2708987

[34] T. Kim *et al.*, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th International Conference on Machine Learning* Mar. 2017, vol. 4, pp. 2941–2949,.

[35] Y. Chen *et al.,* "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," *Stain Technology*, vol. 6, no. 2, pp. 47–49, Mar. 2018. https://doi.org/10.48550/arXiv.1803.01417

[36] M. K. Kar, D. R. Neog, and M. K. Nath, "Retinal vessel segmentation using multi-scale residual convolutional neural network (MSR-Net) combined with generative adversarial networks," *Circuits, Systems, and Signal Processing*, vol. 42, no. 2, pp. 1206–1235, 2023. doi: 10.1007/s00034-022-02190-5

[37] Kaggle. [Online]. Available: https://www.kaggle.com/datasets/eiraoi/thyroidultrasound

[38] M. Khalaf and B. N. Dhannoon, "Skin lesion segmentation based on U-shaped network," *Karbala International Journal of Modern Science,* vol. 8, no. 3, pp. 493–502, Aug. 2022. doi: 10.33640/2405-609X.3248

[39] W. Weng and X. Zhu, "INet: Convolutional networks for biomedical image segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021. doi: 10.1109/ACCESS.2021.3053408

[40] X. Tong *et al.*, "ASCU- Net : Attention gate, spatial and channel attention U-net or skin lesion segmentation," *Diagnostics,* vol. 11, no. 3, 2021. doi: 10.3390/diagnostics11030501

[41] D. Popescu, M. Deaconu, and L. Ichim, "Retinal blood vessel segmentation using Pix2Pix GAN," in *Proc. 29th Mediterranean Conference on Control and Automation (MED)*, 2021, pp. 1173–1178. doi: 10.1109/MED51440.2021.9480169

[42] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976. doi: 10.1109/CVPR.2017.632

[43] V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Medical Image Analysis*, vol. 54, pp. 280–296, 2019. doi: 10.1016/j.media.2019.03.009

[44] H. Bhatia, "Generalized loss functions for generative adversarial networks," M.S. thesis, Dept. of Mathematics and Statistics, Queen's Univ., Kingston, Ontario, Canada, 2020.

[45] H. Li *et al.* , "An improved pix2pix model based on Gabor filter for robust color image rendering," *Mathematical Biosciences and Engineering: MBE*, vol. 19, no. 1, pp. 86–101, 2022. doi: 10.3934/mbe.2022004

[46] R. Toda *et al.*, "Lung cancer CT image generation from a free-form sketch using style-based pix2pix for data augmentation," *Scientific Reports*, pp. 1–10, 2022. doi: 10.1038/s41598-022-16861-5

[47] T. Salimans *et al.*, "Improved techniques for training GANs," in *Proc. 30th Conference on Neural Information Processing Systems, Barcelona, Spain*, Jun. 2016, pp. 2234–2242.

[48] T. E. J. Vallejo *et al.*, "Towards lane detection using a generative adversarial network," *Nova Scientia*, vol. 15, no. 31, pp. 1–11, Nov. 2023. doi: 10.21640/ns.v15i31.3094

[49] Q. Du Nguyen and H. T. Thai, "Crack segmentation of imbalanced data: The role of loss functions," *Engineering Structures*, vol. 297, 116988, Oct. 2023. doi: 10.1016/j.engstruct.2023.116988

[50] J. Ma *et al.*, "Loss odyssey in medical image segmentation," *Medical Image Analysis*, vol. 71, 2021. doi: 10.1016/j.media.2021.102035

[51] L. E. Christovam, M. H. Shimabukuro, and M. D. L. B. T. Galo, "Pix2pix conditional generative adversarial network with MLP loss

function for cloud removal in a cropland time series," *Remote Sensing*, pp. 1–25, 2022. https:// doi.org/10.3390/rs14010144

[52] M. J. Meni *et al.*, "Entropy-based guidance of deep neural networks for accelerated convergence and improved performance," *Information Sciences*, vol. 681, 121239, Sep. 2024. doi: 10.1016/j.ins.2024.121239

[53] T. Anbalagan *et al.*, "Analysis of various techniques for ECG signal in healthcare, past, present, and future," *Biomedical Engineering Advances*, vol. 6, 100089, Jan. 2023. doi: 10.1016/j.bea.2023.100089

[54] J. Zhou *et al.*, "Prediction of lncRNA-disease associations via an embedding learning HOPE in heterogeneous information networks," *Molecular Therapy. Nucleic Acids*, vol. 23, pp. 277–285, Mar. 2021. https://doi.org/10.1016/j.omtn.2020.10.040

[55] P. Geng *et al.*, "STCNet: Alternating CNN and improved transformer network for COVID-19 CT image segmentation," *Biomedical Signal Processing and Control*, vol. 93, 106205, Oct. 2024. doi: 10.1016/j.bspc.2024.106205

[56] D. Carter *et al.*, "Convolutional neural network deep learning model accurately detects rectal cancer in endoanal ultrasounds," *Techniques in Coloproctology*, vol. 28, no. 1, 2024. doi: 10.1007/s10151-024-02917-3

[57] G. James *et al.*, *An Introduction to Statistical Learning*, Springer in Statistics, vol. 103, New York, NY: Springer New York, 2013.

[58] H. Gong *et al.*, "Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules," *Computers in Biology and Medicine,* vol. 155, 2023. doi: 10.1016/j.compbiomed.2022.106389

[59] J. Shao *et al.*, "FCG-Net: An innovative full-scale connected network for thyroid nodule segmentation in ultrasound images," *Biomedical Signal Processing and Control*, vol. 86, 105048, 2023. doi: 10.1016/j.bspc.2023.105048

[60] S. Sun *et al.*, "GLFNet: Global-local fusion network for the segmentation in ultrasound images," *Computers in Biology and Medicine*, vol. 171, 108103, Feb. 2024. doi: 10.1016/j.compbiomed.2024.108103