applied sciences

MDPI

Article

An Improved V-Net Model for Thyroid Nodule Segmentation

Büşra Yetginler [1,2,*] and İsmail Atacak [3]

[1] Department of Computer Engineering, Graduate School of Natural and Applied Sciences, Gazi University, Ankara 06500, Turkey
[2] Department of Computer Engineering, Faculty of Engineering and Architecture, Kırşehir Ahi Evran University, Kırşehir 40100, Turkey
[3] Department of Computer Engineering, Faculty of Technology, Gazi University, Ankara 06560, Turkey; iatacak@gazi.edu.tr
* Correspondence: busra.yetginler@gazi.edu.tr

Abstract: Early diagnosis of increasingly common thyroid nodules is crucial for effectively and accurately managing the disease's monitoring and treatment process. In practice, manual segmentation methods based on ultrasound images are widely used; however, owing to the limitations arising from the imaging sources and differences based on radiologist opinions, their standalone use may not be sufficient for thyroid nodule segmentation. Therefore, there is a growing focus on developing automatic diagnostic approaches to assist radiologists in nodule diagnosis. Although current approaches have yielded successful results, more research is needed for nodule detection because of the complexity of the thyroid region, irregular tissues, and blurred boundaries. This study proposes an improved V-Net segmentation model based on fully convolutional neural networks (V-Net) and squeeze-and-excitation (SE) mechanisms for detecting thyroid nodules in two-dimensional image data. In addition to the strengths of the V-Net approach in the proposed model, a squeeze-and-excitation (SE) mechanism was used to emphasize important features and suppress irrelevant features by assigning weights to the significant features of the model. Experimental studies utilized the Digital Database Thyroid Image (DDTI) and Thyroid Nodule 3493 (TN3K) datasets, and the improved V-Net-based model was validated using the V-Net, fusion V-Net, and SEV-Net methods. The results obtained from the experimental studies demonstrate that the proposed model outperforms the V-Net, fusion V-Net, and SEV-Net models, with a Dice score of 84.51% and an IoU score of 76.27% for the DDTI dataset. Similarly, on the TN3K dataset, it achieved superior performance compared to all benchmarked models, with Dice and IoU scores of 83.88% and 75.50%, respectively. When considering the results in the context of the literature, the proposed model demonstrated the best performance among all models, achieving an average score of 80.39% for the DDTI dataset and 79.69% for the TN3K dataset, according to both Dice and IoU metrics. The model, with a Dice score of 84.51%, competes at a competitive level with Ska-Net, which exhibits the best performance in this metric with a score of 84.98% on the DDTI dataset, whereas it achieved the best performance among existing models with an IoU score of 75.5% on the TN3K dataset. The achievement of the proposed model may make it an effective tool that radiologists can use for thyroid nodule detection.

Keywords: medical image segmentation; thyroid nodules; ultrasound; deep learning; improved V-Net

Academic Editors: Antonio Fernández-Caballero, Haitao Zhao and Meng Wang

Received: 21 February 2025
Revised: 24 March 2025
Accepted: 27 March 2025
Published: 1 April 2025

Citation: Yetginler, B.; Atacak, İ. An Improved V-Net Model for Thyroid Nodule Segmentation. *Appl. Sci.* **2025**, *15*, 3873. https://doi.org/10.3390/app15073873

1. Introduction

The thyroid is a gland located at the front of the neck. This butterfly-shaped endocrine gland produces hormones that regulate body temperature, blood pressure, heart rate, and metabolism [1,2]. Any anomalies that affect the thyroid can disrupt its operations and cause many diseases [3]. Thyroid nodules are commonly seen within the endocrine system and arise from abnormal growth of thyroid cells due to factors such as radiation exposure and iodine intake [4–6]. These nodules are quite prevalent in the general population, occurring in approximately 19–68% of individuals, and are frequently detected incidentally during routine neck scans [1]. Although these nodules, which are commonly seen in elderly individuals and women, are benign, a definitive and consistent evaluation process is needed to determine whether they are malignant [3]. Benign nodules can often be managed through surgical procedures or medications, whereas malignant ones encompass metastatic and primary thyroid cancers that necessitate surgical treatment [4]. Recently, there has been a noticeable rise in thyroid nodules, which are attributed to irregular eating habits and inconsistent work–rest cycles [7]. As per 2020 statistics, approximately 590,000 individuals were diagnosed with thyroid cancer globally, ranking it as the 10th most prevalent type of cancer overall. Even though patients in the early stages might not show obvious symptoms, those in the advanced stages can face serious issues, such as hoarseness, trouble breathing, and difficulties with swallowing [8]. Early diagnosis and intervention are important to enhance survival rates and to better understand disease progression.

Ultrasound is a popular technique for thyroid examination because it offers high-resolution images quickly and does not involve radiation or invasive procedures. However, it has some drawbacks, such as interference from natural noise, low specificity, and unwanted signals [6,9]. In addition to the limitations of ultrasound techniques and the fact that thyroid nodules often have blurry and irregular borders, having an uneven shape can lead doctors to make incorrect decisions [6]. Consequently, there is a need for supplementary applications along with traditional diagnostic methods for accurate assessment and early identification. Techniques for segmenting thyroid nodules and glands fall into three common categories: machine learning (ML) and deep learning (DL), contour and shape, and region-based methods [10]. Despite being commonly employed due to their effective segmentation capabilities, contour and shape-based methods [11,12] have primarily been evaluated on images from small datasets. These techniques offer significant benefits, including prevention of boundary leakage and reduced sensitivity to nodule echogenicity and noise. Consequently, they can yield accurate and dependable results even in noisy images where echogenicity differences are minimal. Nonetheless, contour and shape-based approaches are sensitive to image quality, require manual initial contouring and parameter tuning, and have high computational costs, which can lower their performance on datasets with low-resolution images. Furthermore, these constraints make applying these methods effectively on datasets with high-dimensional images challenging. Region-based methods [13] manage the segmentation process by using gray intensity information from pixels within a designated region. Variations between high-gray-intensity and low-gray-intensity pixels determine the boundaries of regions to be segmented. These methods yield effective results for images with homogeneous and smooth structures. The ease of application and fast segmentation processes are among their significant advantages. However, these methods may struggle to define segmentation boundaries in images that contain pixels with similar gray intensities [10].

In the ML and DL methods [14,15], the segmentation process relies on classifying image pixels. This process is carried out by ML methods, applying it to features obtained from classical techniques, whereas DL methods implement it internally and automatically derive their features. Most of the methods in this category provide important advantages, such

as insensitivity to nodule echogenicity, automatic segmentation, and high segmentation accuracy and efficiency. However, they also encounter challenges, such as requiring extensive labeled datasets and prolonged training times [10]. Recent studies have focused on developing and using DL methods for segmenting nodule regions from thyroid images, leading to segmentation models that outperform the existing methods. By leveraging their advantages in image classification, numerous studies have proposed innovative models utilizing the U-Net framework for segmentation tasks [14,16,17]. The U-Net architecture is commonly utilized in this area because of its straightforward structural design, facilitation of generalization through symmetric training, and capability to learn features across various levels, enabled by skip connections [14]. Nonetheless, this method struggles to deal with the unclear and blurry edges of nodule regions in the segmentation process and tends to be insensitive to small nodules [17].

This study addressed the complexity of tissue boundaries surrounding the thyroid by proposing an improved V-Net model for two-dimensional image segmentation tasks, considering the limitations of ultrasound imaging techniques and the U-Net method. The proposed model represents an architecture based on V-Net and SE mechanisms, aiming to be applied to various data. The distinguishing features of the improved V-Net model compared to similar studies in the literature can be summarized in the following three points:

- Strengthening the encoder layers of the V-net through triple fusion.
- Enhancing the capability to select more significant features by employing SE mechanisms in the encoder layers.
- Enabling rapid and accurate detection of important features with SE mechanisms applied to the final layer before the output in the decoder layers.

The remainder of this paper is organized as follows: Section 2 covers similar studies in the literature. Section 3 discusses the materials and methods used in this study. Section 4 presents experimental studies, their results, and comparisons with the literature. The Section 5 provides an overall evaluation of this study and discusses future research directions.

2. Literature Review

Like many other diseases, early detection of thyroid nodules plays a crucial role in preventing the condition from worsening to a point where it could significantly impair a patient's quality of life or pose a threat to their life. This is especially true for malignant nodules, for which timely intervention can greatly contribute to effective treatment. Over the years, extensive research has focused on contour-shape analysis, ML, and DL methodologies. Recently, attention has shifted towards DL-based approaches. Many of these studies have emphasized integrating DL techniques with alternative methods or enhancing them with functional modules to achieve successful segmentation outcomes. Below is a summary of the literature on DL applications in thyroid nodule segmentation.

Chu Chen et al. [18] proposed a marker-guided U-Net segmentation (MGU-Net) approach based on manually determined information for the nodule diagnosis process. They used a non-public dataset containing 2246 thyroid nodules collected. The experimental results indicated that their method outperformed the other approaches, achieving a Dice score of 95.76% and an IoU success of 91.46%. They also evaluated their approach in terms of computational cost and found it to be more efficient than the methods they compared.

Sun Jiawei et al. [5] developed a dual-path neural network called TNSNet based on the DeepLabV3+ backbone, which uses soft shape supervision to improve the detection of nodule boundaries and segmentation performance. They evaluated the developed network using a dataset of 3786 thyroid nodule ultrasound images. Their experimental studies achieved an accuracy of 95.81% and a Dice score of 85.33% using the TNSNet model.

Yu Mei et al. [19] developed a weakly supervised semantic segmentation model that uses only image-level classification labels to extract the semantic features of benign and malignant nodules and creates class activation maps. They employed a dual-branch soft erase module (DSEM), a scale feature adaptation module (SFAM), and an edge self-attention module (ESAM). The proposed model achieved Dice scores of 60.2% for benign nodules and 71.3% for malignant nodules on a non-public dataset and Jaccard scores of 46.2% for benign nodules and 56.1% for malignant nodules. The model was validated through ablation studies and compared with the state-of-the-art (SOTA) methods. These results indicate that the model is a cost-effective approach for thyroid nodule segmentation.

Tao Zhen et al. [20] proposed the LCA-Net approach, which successfully uses a local feature information and global context information structure to segment nodule boundaries. In order to verify the model's success in addressing the segmentation challenges posed by weak edges in ultrasound images and the complexity of the thyroid tissue structure, they utilized the TN-SCUI2020 and TN3K datasets, which have different data sources and dataset sizes. Based on the experimental results, they reported that the LCA-Net model achieved a Dice score of 90.26%, 90.68% precision, and 91.84% recall for the TN-SCUI2020 dataset and a Dice score of 82.08%, 80.55% precision, and 85.34% recall for the TN3K dataset. They achieved a Jaccard score of 82.65% for the TN-SCUI2020 dataset and a Jaccard score of 71.18% for the TN3K dataset.

Nguyen Tien Dat et al. [21] developed an approach incorporating nested and attention-based networks to leverage the advantages of these structures for improving segmentation performance. They used a suggestion network (SN) and an enhancement network (EN) to construct a nested network. They noted that segmenting the thyroid region was challenging when the lesions were small or when significant lesions were present. They achieved more successful results than the other methods, with a Dice score of 61.2% for the DDTI dataset and 83.7% for the 3DThyroid dataset. In the processing time analysis, they mentioned that the proposed method required a longer processing time than the other segmentation networks.

Chen Hongyu et al. [22] designed a frequency-domain enhancement network (FDE-Net), a U-Net-based network structure for accurately and rapidly detecting thyroid nodules. In the model, they preferred structures that enhanced the image contrast and reduced factors such as noise, negatively impacting the segmentation performance. The authors noted that their approach had a weak effect on the segmentation of multiple nodules. However, it has a good diagnostic effect for nodules of different sizes and possesses robust generalization capabilities compared with other models. They achieved a Dice score of 83.54% in thyroid nodule segmentation using FDE-Net.

Li Geng et al. [23] presented a computer-aided diagnostic model called transformer fusing CNN network (TCNet) for segmenting malignant thyroid nodules. They combined large-kernel CNNs and enhanced transformer architectures with a multi-scale fusion module (MFM) designed to create high-level feature maps. In experimental studies, they showed that their models achieved Dice scores of 82.65% and 86.42% for the thyroid datasets MTNS and TN-SCUI2020, respectively, and a Dice score of 84.89% for the GLAS colon histology dataset. They obtained IoU scores of 74.38%, 78.56%, and 80.02% for the MNTS, TN-SCUI2020, and GLAS datasets, respectively.

Shao Jiajun et al. [8] proposed a new thyroid nodule segmentation model (FCG-Net) developed from UNet3+ to create low computational cost feature maps. In the proposed model, they used a ghost module, ghost bottleneck, and squeeze-and-excitation (SE) blocks. The authors utilized two datasets for evaluation: the DDTI dataset and a combined comprehensive dataset formed by merging the DDTI with hospital data. They achieved highly successful results, with 95.40% accuracy and a Dice score of 80.42%

on the DDTI dataset and 94.88% accuracy and a Dice score of 86.70% on the combined comprehensive dataset. They reported that their approach not only increased the accuracy of ultrasonic diagnosis but also reduced hardware and computational requirements.

Liu Weihua et al. [24] developed a shape-margin knowledge-augmented network (SkaNet) as an effective model for thyroid nodule segmentation and diagnosis. They evaluated the model using a CNN subnetwork and a transformer subnetwork on both an open-access dataset and a non-public dataset to obtain shared features. Using the open-access dataset, they achieved a Dice score of 84.98% for thyroid nodule segmentation and an accuracy of 98.06% for thyroid nodule diagnosis. In the non-public dataset, they reported a Dice score of 86.01% and an accuracy of 94.61%. An IoU of 73.88% for the DDTI dataset and an IoU of 75.45% for the non-public dataset were achieved.

Radhachandran Ashwath et al. [16] proposed a multi-task approach using an anomaly detection (AD) module to automatically detect and segment thyroid nodules from ultrasound images. They evaluated the impact of the AD module on various state-of-the-art nodule segmentation architectures using the UCLA, DDTI, and Stanford CINE datasets. Unlike previous studies, their study also included images without nodules during the evaluation. Among the evaluated models, SResUNET-AD, a U-Net with an ImageNet pre-trained encoder and AD achieved the highest F1 score of 83.9% on the test set and an across-the-image-widths Dice score of 80.8%. Utilizing the same dataset, they recorded their best results on positive images with MSUNet-AD, reaching a Dice score of 65.5% and an IoU of 56.5%. SResUNet-AD also demonstrated superior performance on both the DDTI and Stanford CINE datasets, with a Dice score of 40.2% and an IoU of 33.3% and a Dice score of 59.2% and an IoU of 51% for positive images.

Ma Xiaoxuan et al. [25] presented a study based on adversarial networks with multi-scale joint loss for thyroid nodule segmentation (TNSeg), which utilizes a segmentation block, a discriminative block, and adversarial training. They stated that accurate nodule segmentation from ultrasound images is necessary for effective planning of diagnosis and treatment and that existing approaches struggle due to intra-nodule variability. TNSeg includes a segmentation block and a discriminative block. The UperNet framework and Swin-Unet were combined in the segmentation block. In addition, the study proposed a new multiscale loss function that outperforms traditional loss functions in performance optimization. The authors achieved a Dice score of 85.71%, an Hd95 of 12.93, and a Jaccard score of 73.18% on the TN3K dataset; a Dice score of 74.93%, an Hd95 of 24.51, and a Jaccard score of 61.11% on the DDTI dataset; and a Dice score of 92.06%, a Jaccard score of 90.02%, and an Hd95 of 13.35 on the TNUD dataset.

Xiang Zhou et al. [26] proposed a multi-attention-guided U-Net (MAUNeT) approach, which includes a multi-scale cross-attention (MSCA) module and a dual-attention (DA) module that enhances encoder features, to address the issue of significant data discrepancies among data from different centers. They evaluated the effectiveness of the developed model on multi-center ultrasound images of thyroid nodules obtained from three different sources involving 17 hospitals. They utilized a federated learning approach to protect privacy during training. Using the proposed model, they achieved Dice scores of 90.8%, 91.2%, and 88.7% for datasets from three different centers, D1, D2 and D3, respectively.

Wang Shidan et al. [27] developed YOLO-Thyroid, based on the YOLOv8 architecture, to tackle issues stemming from ultrasound image limitations in thyroid nodule detection. Their methodology incorporates SIoU for accurate boundary regression and employs a class-weighted binary cross-entropy loss function to alleviate class imbalance effects. Additionally, they presented a C2fA module featuring coordinate attention (CA) to improve feature extraction capabilities. In assessments using the DDTI dataset, YOLO-Thyroid achieved an impressive 43.6% mAP0.5 (mean average precision at an IoU threshold of 0.5),

outperforming leading YOLO models. They recorded a weight average recall rate of 58.2% when identifying nodules with at least one characteristic suggestive of a malignancy.

Table 1 provides an overview of the literature review summarizing DL-based studies on thyroid nodule segmentation. The presentation of these studies in the table is organized based on the content methodology, data sources, and experimental results.

Table 1. Overview of the literature review encompassing DL-based studies for thyroid nodule segmentation.

Author	Model	Dataset	Performance Results
Chu et al. [18]	MGU-Net	Non-public	Dice: 95.76% IoU: 91.46%
Sun et al. [5]	TNSNet	Non-public	Dice: 85.33% Acc: 95.81%
Yu et al. [19]	SSE-WSSN	Non-public	Dice: 60.2% (benign) Jaccard (IoU): 46.2% (benign) Dice: 71.3% (malignant) Jaccard (IoU): 56.1% (malignant)
Tao et al. [20]	LCA-Net	TN-SCUI2020	Dice: 90.26% Precision: 90.68% Jaccard (IoU): 82.65% Recall: 91.84%
		TN3K	Dice: 82.08% Precision: 80.55% Jaccard (IoU): 71.18% Recall: 85.34%
Nguyen et al. [21]	Method Based on Suggestion and Enhancement Networks	DDTI	Dice: 61.2%
		3DThyroid	Dice: 83.7%
Chen et al. [22]	FDE-Net	Non-public	Dice: 83.54%
Li et al. [23]	TCNet	MNTS	Dice: 82.65% IoU: 74.38%
		TN-SCUI2020	Dice: 86.42% IoU: 78.56%
		GLAS (Colon histology)	Dice: 84.89% IoU: 80.02%
Shao et al. [8]	FCG-Net	DDTI	Dice: 80.42% Acc: 95.40%
		DDTI + Hospital dataset	Dice: 86.70% Acc: 94.88%
Liu et al. [24]	Ska-Net	DDTI	Dice: 84.98% Acc: 98.06% IoU: 73.88%
		Non-public	Dice: 86.01% Acc: 94.61% IoU: 75.45%
Radhachandran et al. [16]	Anomaly Detection (AD) Module	UCLA (Image-wide) SResUnet-AD	Dice: 80.8% F1 Score: 83.9%
		UCLA (Average Across Positive Images-AAPI) MSUNet-AD	Dice: 65.5% IoU: 56.5%
		DDTI (AAPI) SResUnet-AD	Dice: 40.2% IoU: 33.3%
		Stanford CINE (AAPI) SResUnet-AD	Dice: 59.2% IoU: 51%
Ma et al. [25]	TNSeg	DDTI	Dice: 74.93% Hd95: 24.51 Jaccard (IoU): 61.11%
		TN3K	Dice: 85.71% Hd95: 12.93 Jaccard (IoU): 73.18%
		Non-public (TNUD)	Dice: 92.06% Hd95: 13.35 Jaccard (IoU): 90.02%
Xiang et al. [26]	MAUNet	D1 (Non-public)	Dice: 90.8%
		D2 (Non-public)	Dice: 91.2%
		D3 (Non-public)	Dice: 88.7%
Wang et al. [27]	YOLO-Thyroid	DDTI	mAP0.5: 43.6% Recall: 58.2%

Although the studies presented here yielded successful results in performance, there is a pressing need for further investigation because of the complexities in the thyroid area,

indistinct boundaries, irregular tissues, and issues related to multiple nodules. This study proposes a novel DL-based method to enhance thyroid nodule segmentation performance for two-dimensional datasets by strengthening V-Net's encoder layers through a triple fusion structure. An SE mechanism was integrated into this structure to improve its performance further.

3. Materials and Methods

This section outlines the materials and methods required to conduct experimental studies on all models for thyroid nodule segmentation. Initially, it introduces the image datasets utilized in this study—Digital Database Thyroid Image (DDTI) and Thyroid Nodule 3493 (TN3K)—along with descriptions of their preprocessing process. Subsequently, it describes the improved V-Net, V-Net, and fusion V-Net models used for comparison and the SE mechanism configuration included in the improved V-Net and SEV-Net models through their structural diagrams. Finally, at the end of the section, it presents the metrics used to measure the performance of the models employed in this study: Dice and IoU.

This study utilized the DDTI and TN3K datasets made available as open-access resources. After preprocessing and data augmentation of these datasets, we trained separate models, including V-Net, fusion V-Net, SEV-Net, and Improved V-Net. Once the training and validation were completed, the models were evaluated using the test data. Nodule segmentation was executed on thyroid ultrasound images; the results were qualitatively analyzed by comparing ultrasound images with ground truth masks and quantitatively predicted masks through metrics, such as the Dice coefficient and IoU. A schematic representation of this workflow is shown in Figure 1.

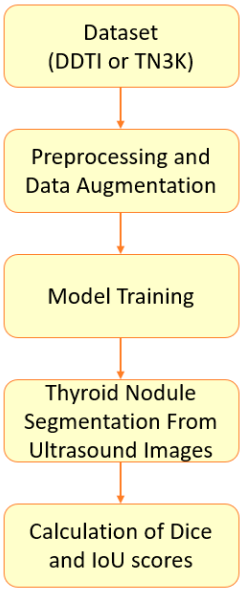

Figure 1. A schematic representation of the workflow.

3.1. Image Datasets and Preprocessing

This study utilized two datasets: DDTI and TN3K. DDTI is an open-access dataset made available by Pedraza et al. [28] to develop algorithms for analyzing thyroid nodules. TN3K, also available in open access, is a thyroid nodule segmentation dataset created by Gong et al. [29] to progress research on computer-aided diagnostic systems. The DDTI dataset comprised 390 XML files containing detailed descriptions and diagnoses of thyroid lesions created by radiologists, featuring one or more ultrasound images per patient at a resolution of 560 × 360 pixels. Preprocessing was performed to remove images lacking

damaged mask information and to eliminate corrupted XML files. Dual images were separated, whereas those with mask information were chosen for use. The masks corresponding to these images were generated by reviewing the XML files. Ultimately, we obtained a dataset comprising 635 images and masks, each sized at 560 × 360 pixels. The TN3K dataset was obtained from various ultrasound imaging systems, including GE Logiq E9, ARIETTA 850, and RESONA 70 B, at Zhujiang Hospital of Southern Medical University. In selecting samples for the dataset, several criteria were considered: at least one thyroid nodule must be present, no blood signals should be included, and among images taken from the same area or perspective of a patient, the nodule closest to the center should be retained. Based on these criteria, the resulting dataset comprises 3493 ultrasound images from 2421 patients, all annotated with pixel-level labels [29].

The technical information present in the image surroundings of the DDTI dataset was cleaned during preprocessing, as it could adversely affect model training. The image data provided in the TN3K dataset were combined into separate folders for images and masks, which were initially provided as two separate folders. Modifications were made to both image and mask names in the TN3K dataset. For the experimental studies, the DDTI and TN3K datasets were initially split into 80% for training and 20% for testing, with an additional allocation of 20% from the training set reserved for validation. Consequently, this resulted in the DDTI dataset containing 406 samples for training, 127 for testing, and 102 for validation. The TN3K dataset comprised 2235 samples for training, 699 for testing, and 559 for validation. In order to decrease the GPU workload, the data from both datasets were resized to 256 × 256 pixels, followed by normalization of the pixel values to a range between 0 and 1. No padding or cropping was performed during this process. The resulting data loss from this adjustment was considered negligible.

3.2. V-Net

V-Net is a method similar to U-Net that employs volumetric convolutions, making it suitable for segmenting computed tomography scans where organ and tissue segmentation pose difficulties [30]. Figure 2 illustrates the schematic representation of the V-Net developed for segmenting nodules in two-dimensional thyroid images.

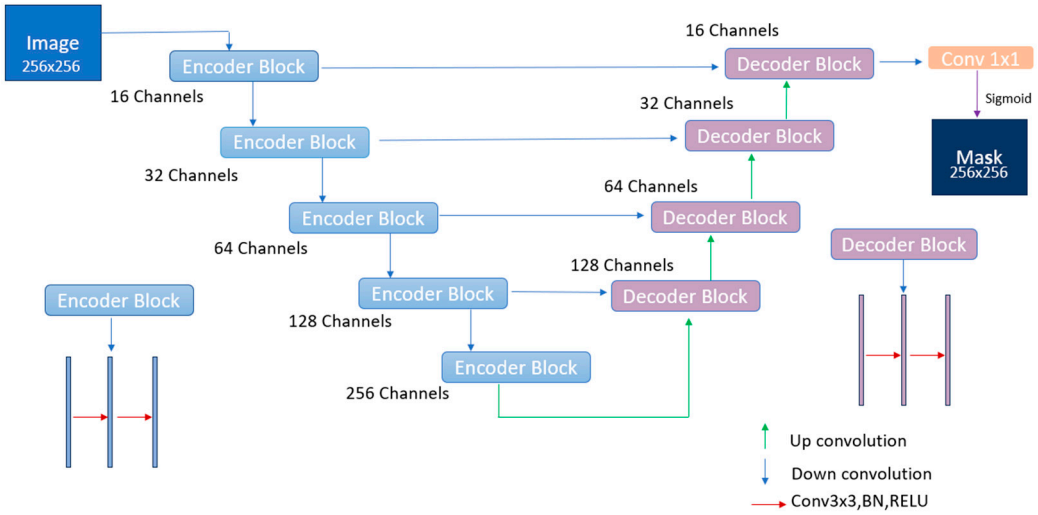

Figure 2. The schematic representation of the V-Net developed for segmenting nodules in two-dimensional thyroid images.

The V-Net model employs convolutional layers to reduce the resolution and extract features. The model combines two basic parts: the encoder and decoder. The encoder part consists of a compression path that operates at different resolutions at various stages, while

the decoder part expands the dimensions until it reaches the original size [30]. Each layer in the encoder computes a feature set twice as large as the previous layer. The decoder part aims for segmentation by providing the necessary information using feature maps [31]. Similar to U-Net, features from the left side of the network are conveyed to the right side to enhance segmentation quality by collecting fine details.

3.3. Fusion V-Net

The fusion V-Net is employed to enhance feature extraction in scenarios where existing DL models lack sufficient data. This is achieved by integrating the complementary features from the same image. Consequently, it can yield favorable outcomes even with smaller datasets by capturing diverse characteristics [32]. Fusion architecture aims to achieve more features by incorporating the input image multiple times at different stages [33]. Figure 3 illustrates the schematic representation of the fusion V-Net model for thyroid nodule segmentation, where the encoder part comprises fusion blocks containing data from different channels, whereas the decoder part is composed of a V-Net designed for two-dimensional data.

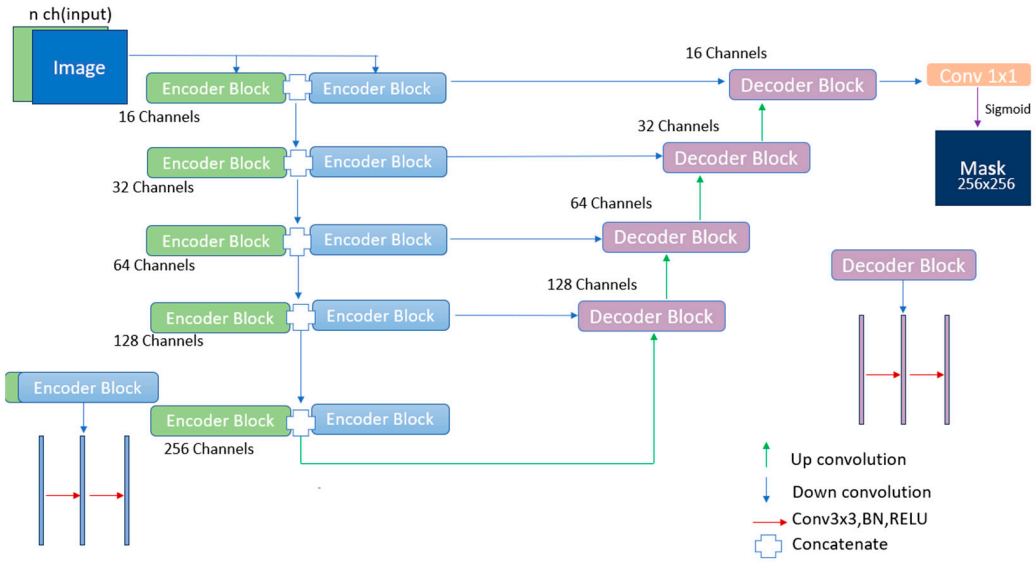

Figure 3. The schematic representation of the fusion V-Net for thyroid nodule segmentation.

3.4. Squeeze-and-Excitation (SE) Mechanism

The SE mechanism facilitates the restructuring of features by highlighting significant ones while diminishing those deemed less important, with an emphasis on inter-channel relationships. The SE mechanism becomes more specialized as the layers progress, enabling class-specific interpretations of the features [34]. They are employed to capture channel dependencies and enhance the convolutional characteristics of architecture [35]. Figure 4 illustrates the configuration of the SE mechanism.

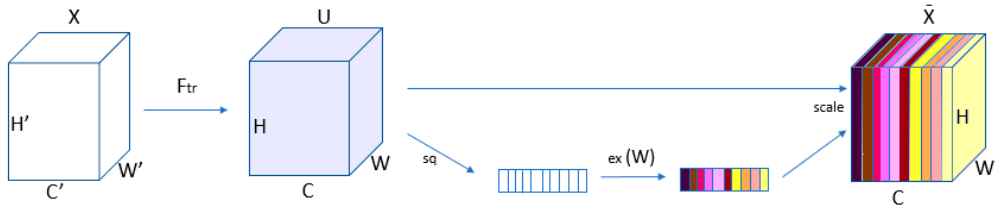

Figure 4. The configuration of the SE mechanism.

In order to emphasize the key features, attention weights indicating the significance of each channel are computed following the squeeze process. The input data are multiplied by the attention weights to learn the importance of each channel and highlight the crucial ones. From the symbols illustrated in Figure 4, X denotes the input data, and F_{tr} represents a layer that extracts features from the input data. U, referred to as feature maps, undergoes a compression process in which information is condensed to extract details from each channel. In the excitation stage, the importance of the channels is dynamically determined using information from the squeeze stage, and adjustments are made accordingly. $\overline{X}$ signifies a feature map arranged based on channel significance [34].

3.5. SEV-Net

SEV-Net is a model created by incorporating SE mechanisms into the existing V-Net architecture. Figure 5 shows a schematic representation of SEV-Net for segmenting nodules in two-dimensional thyroid images.

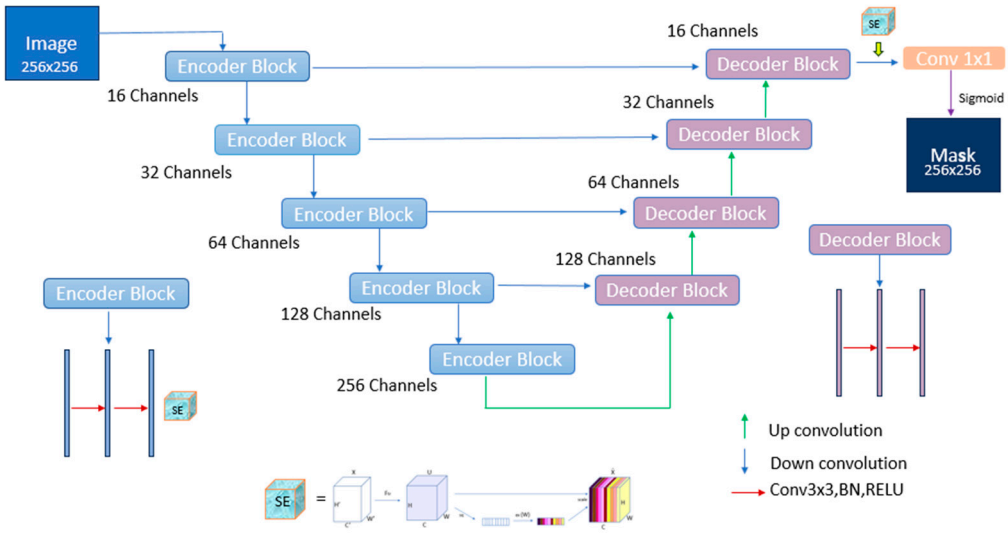

Figure 5. The schematic representation of SEV-Net for segmenting nodules in two-dimensional thyroid images.

SE mechanisms have been employed in numerous medical image segmentation studies using the V-Net model across various stages and functional modules [36,37]. As in the improved V-Net method, this architecture integrates the SE mechanism into all encoder blocks, as well as the last decoder block prior to the output. By incorporating the SE mechanism exclusively in the last decoder block, significant features are emphasized, and insignificant ones are suppressed quickly and accurately.

Figure 6 illustrates the integration of the SE mechanism into the encoder blocks containing three convolution operations.

The SEV-Net architecture comprises five encoder layers and four decoder layers. It utilizes two convolutional operations in the first, second, eighth, and ninth layers while employing three convolutional operations in the third through seventh layers. Integration of the SE mechanism occurs within encoder blocks that feature two convolutional operations in the same manner. By integrating the SE mechanism into each encoder block, the network effectively learns critical features. In the SE mechanism, the squeeze operation is first performed by applying global average pooling to generate scalar values that contain channel-wise information. Subsequently, the excitation phase commences with two fully connected layers. The ReLU activation function was implemented in the first fully connected layer, while the second utilized a sigmoid activation function to yield

normalized weights ranging from 0 to 1 through feature-dimension loading. To emphasize the key features, normalized weights were then adjusted for each channel characteristic and transformed into outputs.

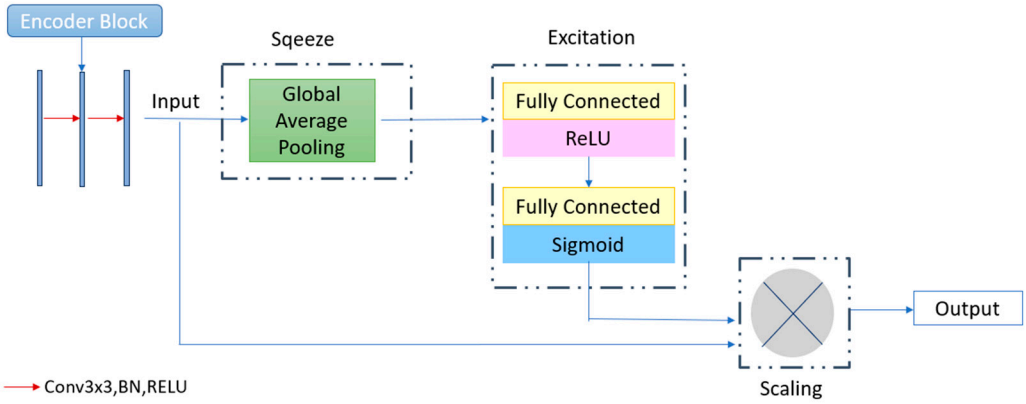

Figure 6. Integration of the SE mechanism into the encoder blocks containing three convolution operations.

3.6. Improved V-Net

The improved V-Net represents a new approach incorporating SE mechanisms into the V-Net based on triple fusion to achieve high accuracy in thyroid nodule segmentation. Figure 7 depicts the schematic representation of the proposed improved V-Net method for segmenting the thyroid nodules.

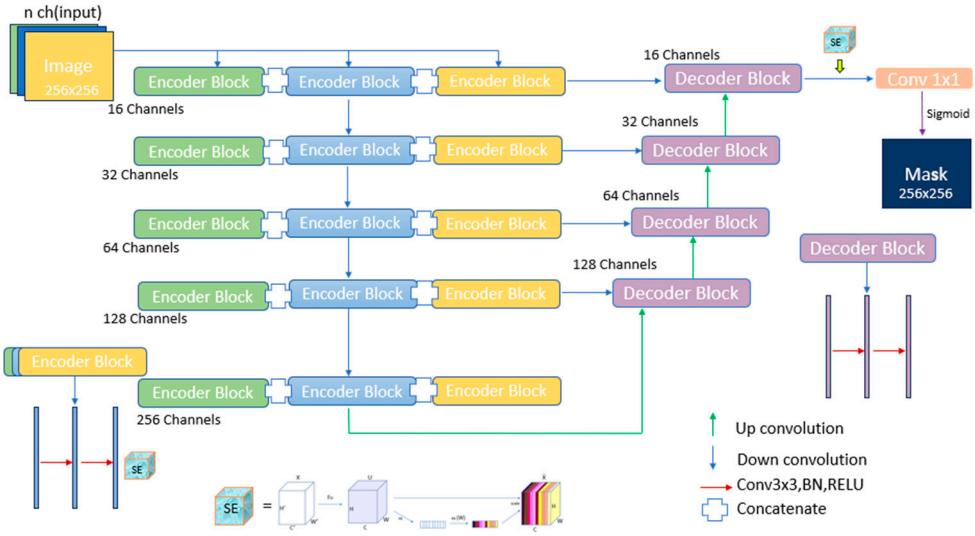

Figure 7. The schematic representation of the improved V-Net for thyroid nodule segmentation.

In order to achieve better feature extraction throughout the model layers, two convolution operations are applied in the encoder blocks of the first and second layers, while three convolutions are utilized in those of the third, fourth, and fifth layers. Two convolutions were implemented in the eighth and ninth layers of the decoder section, with three convolutions applied in both the sixth and seventh layers. Following these convolutions, batch normalization is applied for rapid and stable training alongside ReLU activation, introducing a nonlinear transformation based on the input values. Dropout is also used to mitigate overfitting while maintaining model generalizability. Furthermore, in this method, after extracting features, SE blocks are utilized to determine the significance of each channel, aiming to achieve highly accurate results by highlighting key features. The SE mechanism is

integrated into each layer of the encoder section to facilitate extracting of key features. The decoder section is incorporated solely into the last block to ensure that significant features are obtained quickly and accurately before the output. It is vital that the SE mechanism is applied at appropriate stages within the method, as its inclusion across all decoder layers could hinder the network speed. Furthermore, residual connections are employed in this method to enhance the feature learning efficiency and boost the overall performance.

3.7. Performance Assessment

The assessment of model performance in medical segmentation involves measuring how closely their predicted outputs align with actual accuracy masks defined by experts. Commonly employed metrics for evaluating model accuracy in segmentation issues include precision, sensitivity, Dice, and IoU. This study utilized Dice and IoU metrics to evaluate the performance of the methods used for comparison with the improved V-Net approach.

Dice Similarity Coefficient (Dice): The Dice metric quantifies the extent to which spatial overlap exists between actual (ground truth) masks and predicted masks [38]. Ranging from 0 to 1, this metric is commonly utilized in scenarios where pixel discrepancies arise within the segmentation classes [39,40]. Figure 8 shows an area graph of the Dice similarity coefficient.

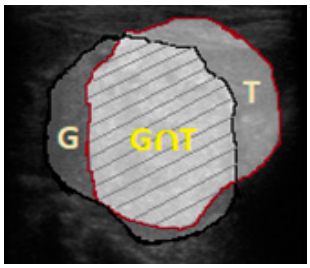

Figure 8. The area graph of the Dice similarity coefficient.

To compute this coefficient—which assesses the similarity between two sets—refer to the equation below:

$$\text{Dice} = \frac{2\times|G \cap T|}{|G|+|T|} \quad (1)$$

where G denotes pixels from the actual segmentation mask, and T represents those from the predicted segmentation mask. Intersection G∩T represents the overlapping pixels found in both masks [38].

Intersection Over Union (IoU): The IoU metric refers to the similarity between the predicted and actual regions of an object in the image. This is defined as the ratio of the intersection of the predicted and actual regions to the union [41]. Equation (2) presents the correlation that illustrates the definition of IoU.

$$\text{IoU} = \frac{|G \cap T|}{|G \cup T|} \quad (2)$$

where G represents the actual segmentation mask pixels, and T defines the predicted segmentation mask pixels. The IoU metric, also called the Jaccard index, compares the similarity between these two regions to assess the predicted pixels [42].

4. Experimental Results and Discussion

Experimental studies of the improved V-Net-based method proposed for thyroid nodule segmentation and the V-Net, fusion V-Net, and SEV-Net methods used to verify its performance in this study were carried out on a local computer with an Intel(R) Core (TM)

i7-12700KF CPU and NVIDIA Geforce GTX 3090 GPU (NVIDIA, Santa Clara, CA, USA). Models for these methods were developed using Python (version 3.10.14) programming alongside TensorFlow and Keras libraries. The configuration of the models utilized in the experimental research has been outlined as follows: The V-Net consists of a 5-layer encoder and a 4-layer decoder. The fusion V-Net includes two 5-layer encoders along with one 4-layer decoder. SEV-Net features a 5-layer encoder, a 4-layer decoder, and SE mechanisms. The improved V-Net comprises three 5-layer encoders, one 4-layer decoder, and SE mechanisms. Each block of these models contains either two or three convolutional layers. In each block of the encoder layers, 3×3 convolutional filters were used, and batch normalization was applied to their outputs. The results were passed through the ReLU activation function, followed by dropout, to prevent overfitting. A convolutional layer was implemented to prevent losing high-level features within the encoder layers instead of the maximum pooling layers typical in U-Net structures. Each block within the decoder layer effectively employs deconvolution to double the feature-map dimensions. Similar to the process in the encoder layer blocks, these layer blocks apply batch normalization, pass through ReLU activation functions, and sequentially undergo dropout on their 3×3 convolution outputs. Residual connections were used in the models to facilitate learning information from the previous blocks. At the end of the decoder blocks, a convolutional layer with a 1×1 size and a sigmoid activation function was employed for detecting masks and backgrounds. In the improved V-Net model, the squeeze ratio of the SE mechanisms was set to 4. These blocks were implemented in each block during the encoder phase, whereas in the decoder phase, they were added only to the last decoder block before the output. The definition and pre-processing steps used to perform the experiments on the DDTI and TN3K datasets are presented in the Section 3; initially, a split ratio of 0.8 was applied to create training (80%) and testing (20%) image collections. Following this, an additional split ratio of 0.2 was utilized on the existing training image set to obtain validation and training image sets for the experiments. Data augmentation techniques were employed in the final training set to strengthen the generalization ability of the model.

This made the DDTI dataset six times more prominent and the TN3K dataset four times larger. The data augmentation process included applying random rotations, shifts, horizontal flips, zooming in/out, and angled shifts to the training samples. Consequently, the amounts of training, testing, and validation data used in the experiments were 2436, 127, and 102 for the DDTI dataset and 8940, 699, and 559 for the TN3K dataset, respectively. The other parameter settings were as follows: epoch number = 100, batch size = 4, learning rate = 10^{-4}, and dropout = 0.3.

This study evaluated the performance of the segmentation models using quantitative and qualitative metrics. For quantitative assessment, changes in loss values across epochs and variations in Dice scores were analyzed during the training process, while standard measures, including Dice and IoU, were employed during the testing process. For qualitative assessment, the images predicted by the models were visually compared with the ground truth masks.

Figure 9a,b show graphs of the changes in training and validation losses and Dice scores by epoch for the proposed method on the DDTI and TN3K datasets. Despite experiencing a sudden increase in validation loss and a sudden drop in validation Dice score toward later epochs of training on the DDTI dataset, the model quickly corrected this state to an acceptable level. The validation loss and Dice score results for the TN3K dataset remained at an acceptable level.

Table 2 compares three V-Net-based models with the improved V-Net model based on their Dice and IoU scores, as well as training times on the DDTI dataset. The scores highlighted in bold indicate that the model in the corresponding row produced the most

successful result in terms of the metric in that column. Upon reviewing these results, it can be noted that although it demonstrates better performance than the V-Net, fusion V-Net, and SEV-Net models, with Dice and IoU metric values of 0.8451 and 0.7627, respectively, it incurs higher resource usage because of its training time of 7286.6776 s compared to others. Fusion V-Net obtained the lowest Dice score, with a value of 0.8103, whereas V-Net reached the lowest IoU performance, with a value of 0.7261. The Dice performance of the V-Net is also very close to that of the lowest-performing fusion V-Net in terms of this metric. However, V-Net had the lowest resource consumption, with a training time of 2680.5712 s.

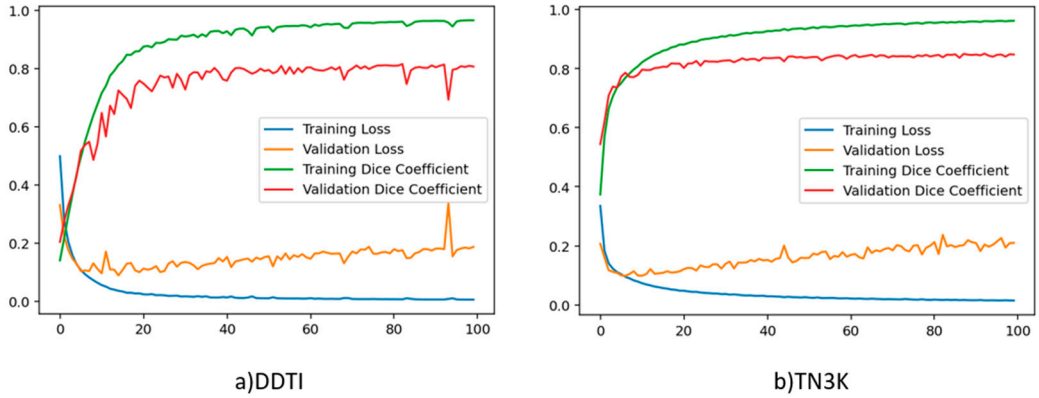

Figure 9. The graphs of the changes in training and validation losses and Dice scores by epoch for the proposed method on the DDTI and TN3K datasets.

Table 2. The comparison of three V-Net-based models with the improved V-Net model based on Dice score, IoU score, and training times on the DDTI dataset.

Model	Dice	IoU	Training Times (s)
V-Net	0.8136	0.7261	**2680.5712**
Fusion V-Net	0.8103	0.7401	4828.9201
SEV-Net	0.8295	0.7433	4839.4166
Improved V-Net (Proposed model)	**0.8451**	**0.7627**	7286.6776

Table 3 presents the Dice and IoU scores achieved by the three V-Net-based models, along with the improved V-Net model on the TN3K dataset, including their respective training times. When the performance results were evaluated, similar to those from the DDTI dataset, it was evident that the proposed model outperformed the others, with Dice and IoU metrics scoring 0.8388 and 0.7550, respectively, while also having the highest resource consumption, with a training time of 26957.3628 s. In this dataset, SEV-Net achieved the lowest performance with a Dice score of 0.8212 and an IoU score of 0.7389, whereas V-Net had the lowest resource consumption with a training time of only 9752.1226 s.

Table 3. The comparison of three V-Net-based models with the improved V-Net model based on Dice score, IoU score, and training times on the TN3K dataset.

Model	Dice	IoU	Training Times (s)
V-Net	0.8235	0.7460	**9752.1226**
Fusion V-Net	0.8290	0.7497	18002.7762
SEV-Net	0.8212	0.7389	10514.5439
Improved V-Net (Proposed model)	**0.8388**	**0.7550**	26957.3628

The ability of DL methods to automatically extract features from raw image data, learn complex patterns, and generalize patterns in image data has given them a critical role in medical imaging and diagnosis, particularly for thyroid nodule segmentation. Advanced methods such as U-Net and architectures that integrate multiscale features and attention mechanisms have recently gained popularity in this area. Table 4 compares the performance of DL-based studies in the literature that used the DDTI dataset for thyroid nodule segmentation with those of the improved V-Net method.

Table 4. Comparison of the performance of DL-based studies utilizing the DDTI dataset for thyroid nodule segmentation in the literature with that of the improved V-Net approach.

Author	Model	Dice (%)	IoU (%)	(Dice + IoU)/2 (%)
Gong et al. [29]	TRFE+	75.37 ± 2.14	60.47 ± 1.08	67.92 ± 1.61
Ma et al. [43]	AMSeg	74.81 ± 2.07	60.89 ± 0.97	67.85 ± 1.52
Sun et al. [44]	GLFNet	74.62	**79.58**	77.10
Xu et al. [45]	MEF-UNet	69	57.62	63.1
Wu et al. [46]	MFMSNet	79.96	70.20	75.08
Shao et al. [8]	FCG-Net	80.42	-	-
Liu et al. [24]	Ska-Net	**84.98**	73.88	79.43
Ma et al. [25]	TNSeg	74.93	61.11	68.02
Radhachandran et al. [16]	Anomaly Detection (AD) Module	40.2	33.3	36.75
Nguyen et al. [21]	Method based on Suggestion and Enhancement Networks	61.2	-	-
Xie et al. [47]	US-Net	81.90 ± 1.68	70.39 ± 1.35	76.15 ± 1.52
Yetginler and Atacak	**Improved V-Net (Proposed model)**	84.51	76.27	**80.39**

The thyroid region prior-guided feature-enhancement network (TRFE+) [29], a multi-tasking learning framework, simultaneously determines the nodule size, nodule and gland position while minimizing errors in classifying non-thyroid regions as thyroid nodules. While it performs well in scenarios involving multiple nodules and thyroid gland segmentation, it recorded a Dice score of 75.37 ± 2.14% for thyroid nodule segmentation using the DDTI dataset. The global–local fusion network (GLFNet) [44] enhances segmentation by merging global semantic insights with local details; however, it underperforms compared to most methods in the table in terms of Dice score despite achieving an IoU score of 79.58%. MEF-UNet [45], an end-to-end multi-scale feature-extraction and fusion network, captures morphological and lesion characteristics, but it is the second-lowest-performing method among those presented, with an IoU score of 57.62%. The multifrequency and multiscale interactive CNN-transformer hybrid network (MFMSNet) [46] guarantees a more accurate segmentation performance by providing a low computational cost with octave convolutions while effectively utilizing the transformer structure to capture long-range dependencies. The findings from two publicly accessible breast ultrasound datasets (BUSI and BUI), along with one thyroid ultrasound dataset (DDTI), indicate that MFMSNet achieved a Dice score of 79.96% and an IoU score of 70.20% for the DDTI dataset, demonstrating acceptable performance in segmenting thyroid nodules. FCG-Net [8], developed to create feature maps with less computation, achieved more successful qualitative results for large nodules than the improved V-Net; however, it achieved a Dice metric score of 80.42%, falling short by 4.09% against the recommended model's performance. The multi-scale adversarial strategy AMSeg [43], based on Swin-Unet, effectively detects well-defined edges while

enhancing edge pixel density but shows limited efficacy on smaller datasets. Ska-Net [24] is an effective network approach that enhances shape and margin knowledge augmented network for thyroid nodule segmentation and diagnosis. It attained the highest Dice score of 84.98% on the DDTI dataset among all the methods presented here, although its IoU score was recorded at 73.88%. TNSeg [25] introduces an innovative strategy that leverages adversarial networks with multi-scale joint losses and employs a discriminative block and a segmentation block, achieving a Dice score of 74.93% and an IoU score of 61.11%, reflecting the average performance compared to existing methods. The AD module-based method [16] presents a novel multitask approach for detecting and segmenting nodules from ultrasound images. However, when evaluated against the other methods presented here, the performance results on the DDTI dataset were notably low for both metrics. Specifically, this method recorded Dice and IoU scores of 40.2% and 33.3%, respectively. The method, built on a suggestion and enhancement network architecture [21], incorporates a structure comprising nested attention mechanism networks. With a Dice score of 61.2% on the DDTI dataset, this method recorded the lowest performance for this metric following the AD module-based approach. US-Net [47] was successfully applied to the TN3K and DDTI datasets, achieving an impressive Dice score of 81.90 $\pm$ 1.68%. In addition, it attained an IoU score of 70.39 $\pm$ 1.35%. The proposed method produced a value close to that of Ska-Net, which achieved the highest score, with a Dice score of 84.51%. The IoU metric achieved a score of 76.27%, ranking it as the second-best performer after GLFNet. When assessing the results based on average performance across both Dice and IoU metrics, the improved V-Net method emerged as the top performer among those listed, with an average score of 80.39%. Ska-Net produced an impressive average performance score of 79.43%, which is closely aligned with that of the proposed method. GLFNet, which provided the highest performance in the IoU metric, fell behind both methods, with an average score of 77.10%.

In recent studies using the TN3K dataset, as in studies related to thyroid nodule segmentation with the DDTI dataset, the methodologies applied represent enhanced and innovative approaches based on DL. Table 5 compares the results of DL-based studies in the literature using the TN3K dataset with those of the improved V-Net method.

Table 5. Comparison of the performance of DL-based studies utilizing the TN3K dataset for thyroid nodule segmentation in the literature with that of the improved V-Net approach.

Author	Model	Dice (%)	IoU (%)	(Dice + IoU)/2 (%)
Özcan et al. [14]	Enhanced-TransUnet	82.92	70.87	76.90
Xie et al. [47]	US-Net	83.66 $\pm$ 0.46	72.92 $\pm$ 0.38	78.29 $\pm$ 0.42
Tao et al. [20]	LCA-Net	82.08	71.18	76.63
Bi et al. [17]	BPAT-UNET	83.64	71.87	77.76
Ma et al. [25]	TNSeg	**85.71**	73.18	79.45
Gong et al. [29]	TRFE+	83.30 $\pm$ 0.26	71.38 $\pm$ 0.43	77.34 $\pm$ 0.35
Ma et al. [43]	AMSeg	84.21 $\pm$ 0.21	72.48 $\pm$ 0.38	78.35 $\pm$ 0.30
Yetginler and Atacak	**Improved V-Net (Proposed model)**	83.88	**75.50**	**79.69**

Enhanced-TransUnet [14], developed to address the blurry boundaries of images and assist in detecting small nodules, improves segmentation performance by integrating a score matrix with skip connections and enhancing the properties of the skip connections. The acceptable Dice score of 82.92% and IoU score of 70.87% achieved by this method on the TN3K dataset, along with similar scores for these metrics from other datasets,

demonstrate that vision transformer (ViT) models are scalable when dealing with large volumes of data. US-Net [47], presented to improve tissue analysis, feature selection, and edge detection in the complex challenges of medical imaging, achieved a score very close to the improved V-Net, with an average Dice performance score of 83.66%. The score of 72.92% obtained in terms of the IoU metric can be considered a successful performance. LCA-Net [20], a local and context attention adaptive network for thyroid nodule segmentation, is designed to capture edge information through various modules, obtain more global relational information, enhance local features, and detect nodules of different locations and sizes. This method demonstrated more successful segmentation results for larger nodules. Despite achieving satisfactory outcomes, with a Dice score of 82.08% and an IoU score of 71.18%, it yielded the lowest Dice score among those studied within this dataset. BPAT-UNet [17], designed to enhance feature extraction for accurately identifying thyroid nodules, is a method capable of detecting small thyroid nodule objects and recognizing blurry boundaries. The method produced a Dice score of 83.64%, making it the model that yielded results closest to the US-Net. The obtained IoU score of 71.87% reflects an acceptable level of effectiveness for thyroid nodule segmentation. The TRFE+ method [29] achieved a Dice score of 83.30 $\pm$ 0.26% and an IoU score of 71.38 $\pm$ 0.43% from the TN3K dataset, notably higher than those recorded on the DDTI dataset. Its Dice performance indicates that it produced results quite close to those of the AMSeg and improved V-Net methods, which are noted for their good performance in this metric. Another model that leveraged both the DDTI and TN3K datasets, AMSeg [43], achieved outstanding results with a Dice score of 84.21 $\pm$ 0.21% and an IoU score of 72.48 $\pm$ 0.38% on the TN3K dataset. When the results were compared with the results of the model in the DDTI dataset, it was observed that they were quite high compared to DDTI dataset. Following TNSeg, which is known to have the highest Dice scores, it provided the second-highest performance in terms of this metric. TNSeg [25] is a method implemented on both datasets, much like TRFE+ and AMSeg, with its performance results on the TN3K dataset being considerably better than those on the DDTI dataset. The Dice score of 85.71% obtained from TN3K stands out as the highest among all listed methods. Additionally, with an IoU score of 73.18%, this method has one of the best average performances, following the improved V-Net model. The results in this table also reveal that the proposed improved V-Net method does not yield vastly different performance outcomes in terms of Dice and IoU scores from both datasets compared with the previously mentioned methods; instead, it shows more balanced and closely aligned results. While achieving a Dice score of 83.88% on the TN3K dataset places it behind TNSeg and AMSeg in this regard, its score is notably close to that of these methods, particularly AMSeg. Regarding the IoU metric, this method performed better than the other methods, scoring 75.50%. It achieved the best performance, with a score of 79.69% when evaluating the results based on the average performance across both metrics. Meanwhile, TNSeg produced an average performance result close to that of the improved V-Net, at 79.45%.

Figures 10 and 11 show qualitative comparisons of the nodule segmentation results from the methods applied to the DDTI and TN3K datasets, respectively. Compared with the manual segmentation results for nodule segmentation, the improved V-Net model produced results that were more closely aligned with the original segmentation masks. However, the proposed method faces challenges when segmenting larger nodule structures.

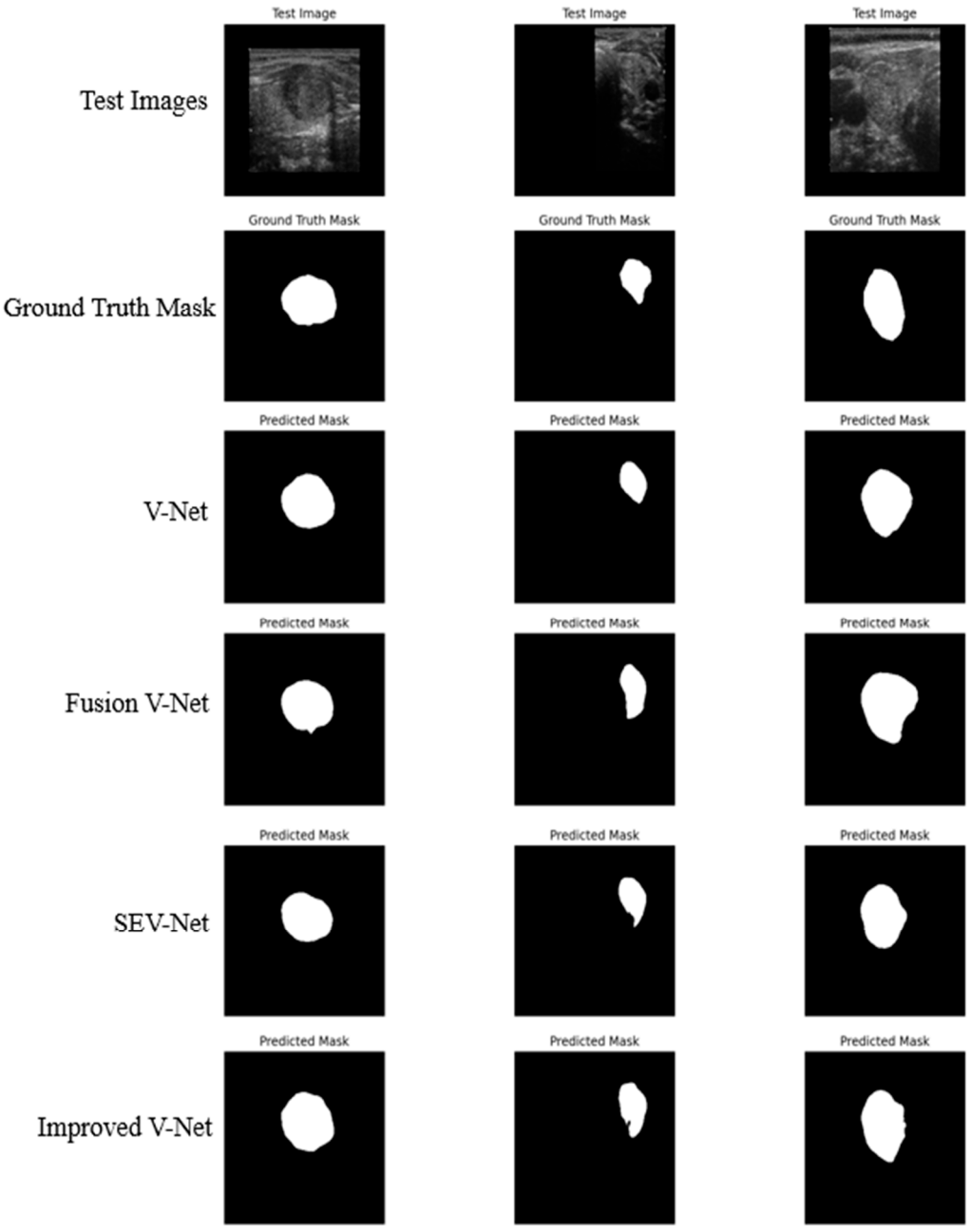

Figure 10. Qualitative comparison of nodule segmentation results of different methods using visualization for the DDTI dataset.

In this study, ablation experiments were conducted to test how an improved V-Net model with either two or three input sets in the encoder affects the performance, as well as the effect of using the SE mechanism across all decoder blocks versus only applying it to the last block. Table 6 presents a comparison of the performance obtained in terms of Dice, IoU, and training time by applying the model that receives three input sets in the encoder with the SE mechanism added to all decoder blocks and the improved V-Net model on the DDTI and TN3K datasets. As can be seen from the training time results for both datasets in Table 6, adding the SE mechanism to all blocks of the decoder layer increased the number of parameters, resulting in longer training times than the improved V-Net model. Furthermore, when assessing both models' performances based on Dice and IoU metrics, it is evident from the table results that including SE mechanisms across all decoder

blocks negatively affected the performance. Consequently, the proposed improved V-Net was developed using an SE mechanism that was added solely to the last decoder block.

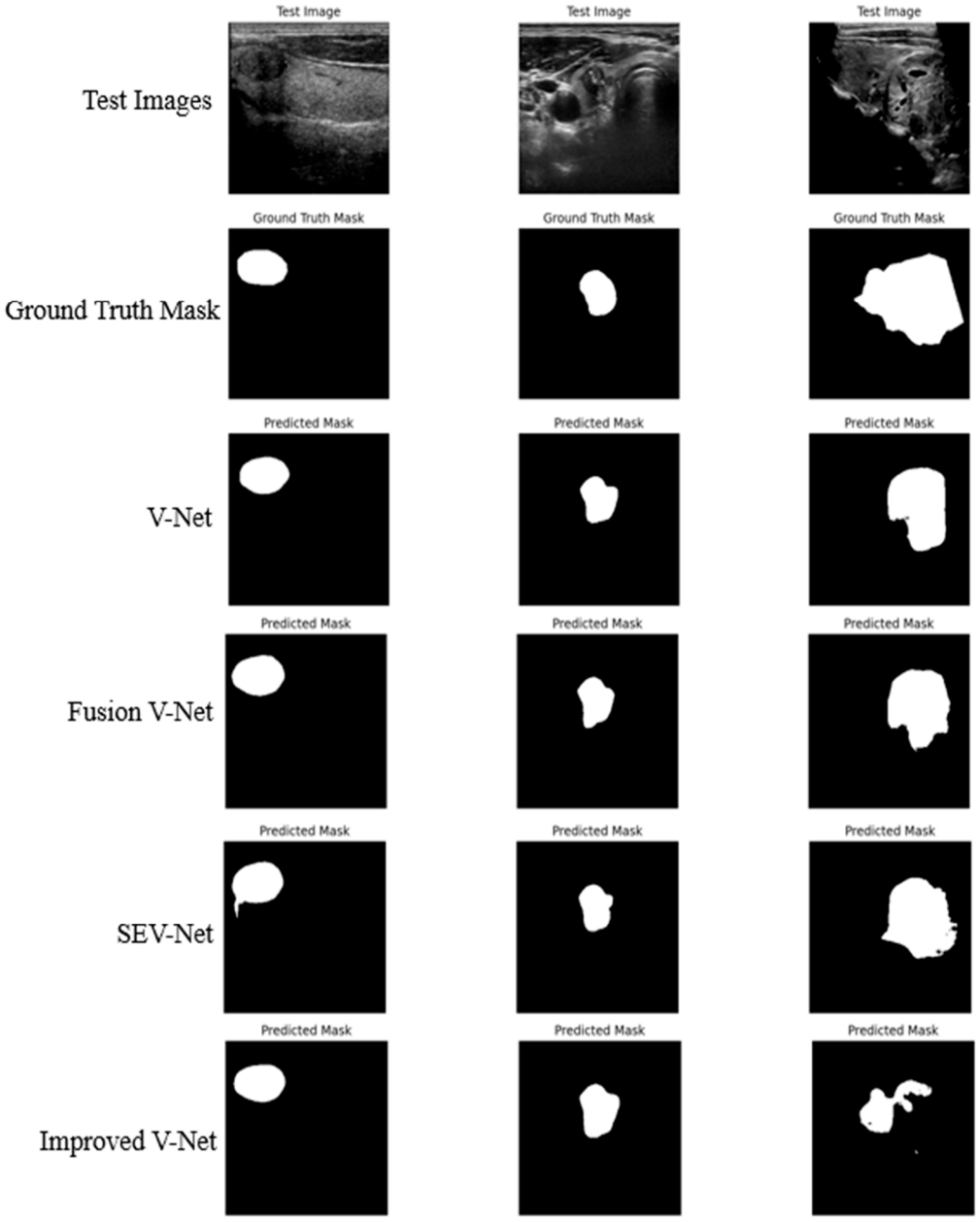

Figure 11. Qualitative comparison of nodule segmentation results of different methods using visualization for the TN3K dataset.

Table 6. Impact of integrating the SE mechanisms across all decoder blocks.

Model	DDTI			TN3K		
	Dice	IoU	Training Time (s)	Dice	IoU	Training Time (s)
With SE mechanisms in all Decoder layers	0.8326	0.7404	7445.8562	0.8355	0.7539	27,132.8646
With improved V-Net	0.8451	0.7627	7286.6776	0.8388	0.7550	26,957.3628

Table 7 illustrates the performance of the improved V-Net in terms of Dice, IoU, and training time metrics for the DDTI and TN3K datasets, depending on whether it receives two or three input sets in the encoder. The results from ablation experiments indicated that using a triple fusion encoder configuration led to improvements of 1.94% in Dice score and 2.45% in IoU for the DDTI dataset, along with enhancements of 0.56% in Dice score and 0.21% in IoU for the TN3K dataset, demonstrating that this model (improved V-Net) achieves superior segmentation results for thyroid nodules.

Table 7. Impact of the encoder fusion structures with two and three input sets.

Model	DDTI			TN3K		
	Dice	IoU	Training Time (s)	Dice	IoU	Training Time (s)
With two input sets in the encoder	0.8257	0.7382	5197.6772	0.8332	0.7529	19,320.5820
With three input sets in the encoder (improved V-Net)	0.8451	0.7627	7286.6776	0.8388	0.7550	26,957.3628

The findings from both quantitative and qualitative assessments derived from experimental studies, along with the literature comparison results, showed that the proposed improved V-Net approach, bolstered by a triple encoder fusion architecture and SE mechanisms, can effectively facilitate the successful segmentation of thyroid nodules. Therefore, this model can serve as a valuable tool for radiologists and healthcare professionals when integrated into actual workflows. To facilitate implementation, it is essential to convert the model into an accessible tool suitable for clinical environments. This conversion can be accomplished through a user-friendly interface that works with imaging devices via a database, making the results generated by the model comprehensible to the radiologists. Radiologists can readily interpret the segmentation results generated by the model through automatic markings that are directly applied to images via the interface. For the effective operation of this interface program within a clinical environment, it is essential to have a computer equipped with a GPU and high-performance hardware to ensure reliable functionality. Furthermore, optimizing the model using various images sourced from the database can significantly boost its performance. However, one of the primary challenges hindering the applicability of the model in this domain is the consent issues surrounding patient data privacy and ethics.

5. Conclusions

Thyroid disorders, especially malignant nodular thyroid cancers, are becoming a significant health issue worldwide. Early diagnosis and treatment are crucial for reducing complications and shaping the disease course. Consequently, there is an increasing need for tools and methods that can provide accurate and reliable segmentation results from images obtained using advanced imaging technologies. This study proposes an improved V-Net approach that utilizes a triple fusion architecture and a channel attention mechanism to segment thyroid nodules effectively. The triple fusion architecture was implemented within the encoder section of the model, whereas SE mechanisms were incorporated across all encoder blocks, as well as in the last decoder block. By merging features scaled differently at each encoder layer through triple fusion architecture, we aim to enhance model performance by creating a rich representation of features. The SE mechanisms facilitate the focus on significant features based on their interrelationships within the

encoder blocks. In addition, the SE mechanism employed in the last decoder block enables segmentation based on key features, thereby improving overall model effectiveness.

In this study, the performances of the methods employed for thyroid nodule segmentation were tested on two distinct image datasets: DDTI and TN3K. Ablation studies, quantitative assessments, and qualitative observations were conducted to develop and test the proposed model. The results from the ablation experiments indicate that implementing architecture with three input sets in the encoder (triple fusion architecture) enhances performance in terms of the Dice and IoU metrics. Furthermore, the experiments conducted in the decoder section indicated that using the SE mechanism only in the last decoder block contributed more to the model's performance in terms of both training time and relevant metrics than using it in all blocks. Consequently, the proposed method for thyroid nodule segmentation was structured as an improved V-Net model, with the encoder section using triple fusion architecture incorporating SE mechanisms at each block and the decoder section only including the SE mechanism in the last block. In the experiments on quantitative assessments, the proposed model and the V-Net, fusion V-Net, and SEV-Net models used for the performance comparison were applied to both datasets, and performance analyses were conducted by comparing the Dice, IoU, and training time metrics. The findings revealed that the proposed model outperformed the others in terms of the Dice and IoU scores across both datasets. Nevertheless, due to having the longest training time among all the models, it turned out to be the one with the highest computational expense. The results obtained from experiments conducted on qualitative observations indicated that while the proposed model successfully produced results in segmenting small- and medium-sized nodules, its success rate decreased in segmenting large nodules. The results of the literature comparison conducted on the DDTI and TN3K datasets show that the proposed model achieves a performance that surpasses that of existing models in terms of the average scores obtained for Dice and IoU.

Considering all these evaluations, we can conclude that the model can provide significant contributions to the early diagnosis of thyroid diseases and effective planning of the treatment process, especially because of its impressive performance in the segmentation of small- and medium-sized modules. However, the reduced success rate of large nodules highlights aspects that require further improvement. While integrating triple fusion architecture and SE mechanisms has enhanced performance, it has also led to longer training times and higher computational costs owing to an increase in parameters. This restricts its application where computational resources are limited. Thus, another aspect to enhance is the management of computational expenses. In light of this, we propose several strategies for future research aimed at optimizing the model architecture without compromising the performance by minimizing unnecessary calculations. The model architecture can be optimized to reduce unnecessary computations while maintaining performance. The generalization capability of the model can be enhanced by training it on various datasets.

Author Contributions: Conceptualization, B.Y.; methodology, B.Y.; investigation, B.Y. and İ.A.; software, B.Y.; formal analysis, B.Y. and İ.A.; resources, B.Y. and İ.A.; writing—original draft preparation, B.Y. and İ.A.; writing—reviewing and editing, B.Y. and İ.A.; visualization, B.Y.; supervision, İ.A.; project administration, İ.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in the DDTI repository at https://www.kaggle.com/datasets/dasmehdixtr/ddti-thyroid-ultrasound-images, accessed on 24 October 2024; reference number [28]. The data presented in this study are openly available in the TN3K at https://drive.google.com/file/d/1reHyY5eTZ5uePXMVMzFOq5j3eFOSp50F/view?usp=sharing, accessed on 24 October 2024; reference number [29].

Acknowledgments: The authors would like to thank Gazi University Academic Writing Application and Research Center for proofreading the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rehman, H.A.U.; Lin, C.Y.; Su, S.F. Deep learning based fast screening approach on ultrasound images for thyroid nodules diagnosis. *Diagnostics* **2021**, *11*, 2209. [CrossRef] [PubMed]
2. Mugasa, H.; Dua, S.; Koh, J.E.W.; Hagiwara, Y.; Lih, O.S.; Madla, C.; Kongmebhol, P.; Ng, K.H.; Acharya, U.R. An adaptive feature extraction model for classification of thyroid lesions in ultrasound images. *Pattern Recognit. Lett.* **2020**, *131*, 463–473. [CrossRef]
3. Swathi, G.; Altalbe, A.; Kumar, R.P. QuCNet: Quantum-Inspired Convolutional Neural Networks for Optimized Thyroid Nodule Classification. *IEEE Access* **2024**, *12*, 27829–27842. [CrossRef]
4. Yu, H.; Li, J.; Sun, J.; Zheng, J.; Wang, S.; Wang, G.; Ding, Y.; Zhao, J.; Zhang, J. Intelligent diagnosis algorithm for thyroid nodules based on deep learning and statistical features. *Biomed. Signal Process. Control* **2022**, *78*, 103924. [CrossRef]
5. Sun, J.; Li, C.; Lu, Z.; He, M.; Zhao, T.; Li, X.; Gao, L.; Xie, K.; Lin, T.; Sui, J.; et al. TNSNet: Thyroid nodule segmentation in ultrasound imaging using soft shape supervision. *Comput. Methods Programs Biomed.* **2022**, *215*, 106600. [CrossRef]
6. Li, Z.; Zhou, S.; Chang, C.; Wang, Y.; Guo, Y. A weakly supervised deep active contour model for nodule segmentation in thyroid ultrasound images. *Pattern Recognit. Lett.* **2023**, *165*, 128–137. [CrossRef]
7. Yang, T.Y.; Zhou, L.Q.; Li, D.; Han, X.H.; Piao, J.C. An improved CNN-based thyroid nodule screening algorithm in ultrasound images. *Biomed. Signal Process. Control* **2024**, *87*, 105371. [CrossRef]
8. Shao, J.; Pan, T.; Fan, L.; Li, Z.; Yang, J.; Zhang, S.; Zhang, J.; Chen, D.; Zhu, X.; Chen, H.; et al. FCG-Net: An innovative full-scale connected network for thyroid nodule segmentation in ultrasound images. *Biomed. Signal Process. Control* **2023**, *86*, 105048. [CrossRef]
9. Sun, J.; Wu, B.; Zhao, T.; Gao, L.; Xie, K.; Lin, T.; Sui, J.; Li, X.; Wu, X.; Ni, X. Classification for thyroid nodule using ViT with contrastive learning in ultrasound images. *Comput. Biol. Med.* **2023**, *152*, 106444. [CrossRef]
10. Chen, J.; You, H.; Li, K. A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. *Comput. Methods Programs Biomed.* **2020**, *185*, 105329. [CrossRef]
11. Savelonas, M.A.; Iakovidis, D.K.; Legakis, I.; Maroulis, D. Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 519–527. [CrossRef] [PubMed]
12. Koundal, D.; Gupta, S.; Singh, S. Automated delineation of thyroid nodules in ultrasound images using spatial neutrosophic clustering and level set. *Appl. Soft Comput.* **2016**, *40*, 86–97. [CrossRef]
13. Zhao, J.; Zheng, W.; Zhang, L.; Tian, H. Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. *Health Inf. Sci. Syst.* **2013**, *1*, 5. [CrossRef] [PubMed]
14. Özcan, A.; Tosun, Ö.; Dönmez, E.; Sanwal, M. Enhanced-TransUNet for ultrasound segmentation of thyroid nodules. *Biomed. Signal Process. Control* **2024**, *95*, 106472. [CrossRef]
15. Keramidas, E.G.; Maroulis, D.; Iakovidis, D.K. TND: A thyroid nodule detection system for analysis of ultrasound images and videos. *J. Med. Syst.* **2012**, *36*, 1271–1281. [CrossRef]
16. Radhachandran, A.; Kinzel, A.; Chen, J.; Sant, V.; Patel, M.; Masamed, R.; Arnold, C.W.; Speier, W. A multitask approach for automated detection and segmentation of thyroid nodules in ultrasound images. *Comput. Biol. Med.* **2024**, *170*, 107974. [CrossRef]
17. Bi, H.; Cai, C.; Sun, J.; Jiang, Y.; Lu, G.; Shu, H.; Ni, X. BPAT-UNet: Boundary preserving assembled transformer UNet for ultrasound thyroid nodule segmentation. *Comput. Methods Programs Biomed.* **2023**, *238*, 107614. [CrossRef]
18. Chu, C.; Zheng, J.; Zhou, Y. Ultrasonic thyroid nodule detection method based on U-Net network. *Comput. Methods Programs Biomed.* **2021**, *199*, 105906. [CrossRef]
19. Yu, M.; Han, M.; Li, X.; Wei, X.; Jiang, H.; Chen, H.; Yu, R. Adaptive soft erasure with edge self-attention for weakly supervised semantic segmentation: Thyroid ultrasound image case study. *Comput. Biol. Med.* **2022**, *144*, 105347. [CrossRef]
20. Tao, Z.; Dang, H.; Shi, Y.; Wang, W.; Wang, X.; Ren, S. Local and Context-Attention Adaptive LCA-Net for Thyroid Nodule Segmentation in Ultrasound Images. *Sensors* **2022**, *22*, 5984. [CrossRef]
21. Nguyen, D.T.; Choi, J.; Park, K.R. Thyroid Nodule Segmentation in Ultrasound Image Based on Information Fusion of Suggestion and Enhancement Networks. *Mathematics* **2022**, *10*, 3484. [CrossRef]

22. Chen, H.; Yu, M.-an.; Chen, C.; Zhou, K.; Qi, S.; Chen, Y.; Xiao, R. FDE-net: Frequency-domain enhancement network using dynamic-scale dilated convolution for thyroid nodule segmentation. *Comput. Biol. Med.* **2023**, *153*, 106514. [CrossRef] [PubMed]
23. Li, G.; Chen, R.; Zhang, J.; Liu, K.; Geng, C.; Lyu, L. Fusing enhanced Transformer and large kernel CNN for malignant thyroid nodule segmentation. *Biomed. Signal Process. Control* **2023**, *83*, 104636. [CrossRef]
24. Liu, W.; Lin, C.; Chen, D.; Niu, L.; Zhang, R.; Pi, Z. Shape-margin knowledge augmented network for thyroid nodule segmentation and diagnosis. *Comput. Methods Programs Biomed.* **2024**, *244*, 107999. [CrossRef]
25. Ma, X.; Sun, B.; Liu, W.; Sui, D.; Shan, S.; Chen, J.; Tian, Z. Tnseg: Adversarial networks with multi-scale joint loss for thyroid nodule segmentation. *J. Supercomput.* **2024**, *80*, 6093–6118. [CrossRef]
26. Xiang, Z.; Tian, X.; Liu, Y.; Chen, M.; Zhao, C.; Tang, L.N.; Xue, E.S.; Zhou, Q.; Shen, B.; Li, F.; et al. Federated learning via multi-attention guided UNet for thyroid nodule segmentation of ultrasound images. *Neural Netw.* **2025**, *181*, 106754. [CrossRef]
27. Wang, S.; Zhao, Z.A.; Chen, Y.; Mao, Y.J.; Cheung, J.C.W. Enhancing Thyroid Nodule Detection in Ultrasound Images: A Novel YOLOv8 Architecture with a C2fA Module and Optimized Loss Functions. *Technologies* **2025**, *13*, 28. [CrossRef]
28. Pedraza, L.; Vargas, C.; Narváez, F.; Durán, O.; Muñoz, E.; Romero, E. An open access thyroid ultrasound image database. In Proceedings of the 10th International Symposium on Medical İnformation Processing and Analysis, Cartagena de Indias, Colombia, 14–16 October 2014. [CrossRef]
29. Gong, H.; Chen, J.; Chen, G.; Li, H.; Li, G.; Chen, F. Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. *Comput. Biol. Med.* **2023**, *155*, 106389. [CrossRef]
30. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *arXiv* **2016**, arXiv:1606.04797v1. [CrossRef]
31. Türk, F.; Lüy, M.; Barışçı, N. Kidney and renal tumor segmentation using a hybrid v-net-based model. *Mathematics* **2020**, *8*, 1772. [CrossRef]
32. Zhang, Y.; Morel, O.; Blanchon, M.; Seulin, R.; Rastgoo, M.; Sidibé, D. Exploration of Deep Learning-based Multimodal Fusion for Semantic Road Scene Segmentation. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 25–27 February 2019; pp. 336–343. [CrossRef]
33. Türk, F.; Kökver, Y. Detection of Lung Opacity and Treatment Planning with Three-Channel Fusion CNN Model. *Arabian J. Sci. Eng.* **2024**, *49*, 2973–2985. [CrossRef]
34. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *arXiv* **2019**, arXiv:1709.01507v4. [CrossRef]
35. Türk, F.; Lüy, M.; Barışçı, N.; Yalçınkaya, F. Kidney Tumor Segmentation Using Two-Stage Bottleneck Block Architecture. *Intell. Autom. Soft Comput.* **2022**, *33*, 349–363. [CrossRef]
36. Liu, P.; Dou, Q.; Wang, Q.; Heng, P.A. An encoder-decoder neural network with 3D squeeze-and-excitation and deep supervision for brain tumor segmentation. *IEEE Access* **2020**, *8*, 34029–34037. [CrossRef]
37. Lu, S.; Han, J.; Li, J.; Zhu, L.; Jiang, J.; Tang, S. Three-dimensional Medical Image Segmentation with SE-VNet Neural Networks. In Proceedings of the 2021 3rd International Conference on Intelligent Medicine and Image Processing, New York, NY, USA, 23–26 April 2021; pp. 14–20. [CrossRef]
38. Hossain, M.I.; Amin, M.Z.; Anyimadu, D.T.; Suleiman, T.A. Comparative Study of Probabilistic Atlas and Deep Learning Approaches for Automatic Brain Tissue Segmentation from MRI Using N4 Bias Field Correction and Anisotropic Diffusion Pre-processing Techniques. *arXiv* **2024**, arXiv:2411.05456. [CrossRef]
39. Kato, S.; Hotta, K. Adaptive t-vMF dice loss: An effective expansion of dice loss for medical image segmentation. *Comput. Biol. Med.* **2024**, *168*, 107695. [CrossRef]
40. Türk, F. RNGU-NET: A novel efficient approach in Segmenting Tuberculosis using chest X-ray images. *PeerJ Comput. Sci.* **2024**, *10*, e1780. [CrossRef]
41. Rahman, M.; Wang, Y. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. In Proceedings of the Advances in Visual Computing, Las Vegas, NV, USA, 12–14 December 2016; Springer: Cham, Switzerland; pp. 234–244. [CrossRef]
42. Cho, Y.J. Weighted Intersection over Union (wIoU) for evaluating image segmentation. *Pattern Recognit. Lett.* **2024**, *185*, 101–107. [CrossRef]
43. Ma, X.; Sun, B.; Liu, W.; Sui, D.; Chen, J.; Tian, Z. AMSeg: A Novel Adversarial Architecture Based Multi-Scale Fusion Framework for Thyroid Nodule Segmentation. *IEEE Access* **2023**, *11*, 72911–72924. [CrossRef]
44. Sun, S.; Fu, C.; Xu, S.; Wen, Y.; Ma, T. GLFNet: Global-local fusion network for the segmentation in ultrasound images. *Comput. Biol. Med.* **2024**, *171*, 108103. [CrossRef]
45. Xu, M.; Ma, Q.; Zhang, H.; Kong, D.; Zeng, T. MEF-UNet: An end-to-end ultrasound image segmentation algorithm based on multi-scale feature extraction and fusion. *Comput. Med. Imaging Graph.* **2024**, *114*, 102370. [CrossRef] [PubMed]

46. Wu, R.; Lu, X.; Yao, Z.; Ma, Y. MFMSNet: A Multi-frequency and Multi-scale Interactive CNN-Transformer Hybrid Network for breast ultrasound image segmentation. *Comput. Biol. Med.* **2024**, *177*, 108616. [CrossRef]
47. Xie, X.; Liu, P.; Lang, Y.; Guo, Z.; Yang, Z.; Zhao, Y. US-Net: U-shaped network with Convolutional Attention Mechanism for ultrasound medical images. *Comput. Graph.* **2024**, *124*, 104054. [CrossRef]