Elmer Jeto Gomes Ataide*, Shubham Agrawal, Aishwarya Jauhari, Axel Boese, Alfredo Illanes, Simone Schenke, Michael C. Kreissl and Michael Friebe

Comparison of Deep Learning Algorithms for Semantic Segmentation of Ultrasound Thyroid Nodules

Abstract: Ultrasound (US) imaging is used as a preliminary diagnostic tool for the detection, risk-stratification and classification of thyroid nodules. In order to perform the risk stratification of nodules in US images physicians first need to effectively detect the nodules. This process is affected due to the presence of inter-observer and intra-observer variability and subjectivity. Computer Aided Diagnostic tools prove to be a step in the right direction towards reducing the issue of subjectivity and observer variability. Several segmentation techniques have been proposed, from these Deep Learning techniques have yielded promising results. This work presents a comparison between four state of the art (SOTA) Deep Learning segmentation algorithms (UNet, SUMNet, ResUNet and Attention UNet). Each network was trained on the same dataset and the results are compared using performance metrics such as accuracy, dice coefficient and Intersection over Union (IoU) to determine the most effective in terms of thyroid nodule segmentation in US images. It was found that ResUNet performed the best with an accuracy, dice coefficient and IoU of 89.2%, 0.857, 0.767. The aim is to use the trained algorithm in the development of a Computer Aided Diagnostic system for the detection, risk-stratification and classification of thyroid nodules using US

images to reduce subjectivity and observer variability.

Keywords: Ultrasound Imaging, Thyroid Nodules, Deep Learning, Image Processing, Computer Aided Diagnosis

<https://doi.org/10.1515/cdbme-2021-2224>

1 Introduction

Thyroid nodules are solid or cystic lumps in the thyroid gland, which can either be benign or malignant, and they are one of the most commonly diagnosed nodular lesions in the adult population [1]. Ultrasound (US) imaging is used as a preliminary diagnostic tool for the detection, risk-stratification and classification of thyroid nodules. It is used because of its availability, affordability and lack of ionizing radiation. In order to efficiently provide a risk-stratification and classification of a thyroid nodule, a physician first needs to detect the nodule. The current detection process is highly subjective in nature and there exists a high rate of inter-observer and intra-observer variability. This stems from the varied experience levels and visual perceptions of the physicians when it comes to detecting the nodules in US thyroid images.

Computer Aided Diagnosis is a solution that addresses the issue of inter-observer and intra-observer variability and subjectivity. The first step in this process is the detection of the thyroid nodule. This is also referred to as the segmentation of the nodule. Automatic segmentation of ultrasound images is a challenging task due to the heavy speckle noise, low contrast, and shadowing effects of ultrasound images.

Several segmentation techniques have been used like graph-based methods, deformable models, and classical methods. Kollorz et al. [1] presented an approach to semi-automatically segment 2D slices of collected 3D US volumes using power watersheds independent of the type of the nodule. Keramidas et al. [2] presented a Thyroid Boundaries Detection algorithm for detection of the thyroid area and later applied KNN to classify the small patches out from the sliding window based on the Local Binary Patterns texture feature. Even though it was an efficient technique, it could only output rough shapes and locations where the nodules were present. Lakovidis et al. [3] applied active contour methodologies for fine detection of thyroid nodules. Chan

*Corresponding author: (Elmer Jeto) (Gomes Ataide): INKA-Healthtech Innovation Laboratory & Department of Radiology and Nuclear Medicine, Otto-von-Guericke University Magdeburg, Medical Faculty, Leipziger Str. 44, Magdeburg, Germany, e-mail: elmer.gomesataide@ovgu.de

2nd Author (Shubham) (Agrawal), 3rd Author (Aishwarya) (Jauhari): Faculty of Information Technology, Otto-von-Guericke University, Magdeburg, Germany

4th Author (Axel) (Boese), 5th Author (Alfredol) (Illanes): INKA-Healthtech Innovation Laboratory, Otto-von-Guericke University Magdeburg, Medical Faculty, Magdeburg, Germany

6th Author (Simone) (Schenke), 7th Author (Michael C.) (Kreissl): Department of Radiology and Nuclear Medicine, Division of Nuclear Medicine, Otto-von-Guericke University Magdeburg, Medical Faculty, Magdeburg, Germany

8th Author (Michael) (Friebe): INKA-Healthtech Innovation Laboratory, Otto-von-Guericke University Magdeburg, Medical Faculty, Magdeburg, Germany

and Vesse [4] suggested Active Contour without Edges technique based on the level-set method. Deep learning models have also proven to be efficient in segmentation problems and have achieved noteworthy results. Some deep learning-based segmentation methods have also been used for medical images and achieved exceptional improvements by using deep convolutional neural networks (DCNN)[5], fully convolutional neural networks (FCN)[6], and its variants[7].

This paper presents a comparison between four state-of-the-art (SOTA) algorithms developed for the purpose of image segmentation. Each algorithm architecture was adapted for thyroid US images. The results obtained are compared and the best in terms of its overall performance is highlighted. The purpose of doing this is to determine which of these methods can be used most effectively for the development of a complete Computer Aided Diagnostic system for the detection, risk-stratification and classification of thyroid nodules using US images.

2 Methods

This study used US image data of thyroid nodules collected at the Department of Radiology and Nuclear Medicine, Division of Nuclear Medicine, Otto-von-Guericke University Magdeburg, Medical Faculty. A total of 1011 images were used in this study. Each image consisted of a minimum of one thyroid nodule scanned in both transverse and longitudinal planes.

2.1 Data Pre-processing and Augmentation

The images were first pre-processed in order to remove unnecessary background information and noise and then resized and cropped. This was followed by dilation using a kernel of size 15x15. The contours were identified using CHAI_APPROX_SIMPLE algorithm on the processed output image from the previous step and reshaped the image based on the coordinates of the bounding rectangle of the contour. The non-square images were then transformed into square images of dimensions 256*256. The pre-processed image was then centred on its mean and normalized. These processed images were then used for training four SOTA segmentation algorithms. Data augmentation in terms of horizontal and vertical shifts, rotation at angles in the range of 5 to 15 degrees, different affine transformations, and grey value variations were employed. Additionally, random elastic deformations used in the training samples delivered better generalization results. The final augmented number of

images was 6066. A train-test split of 75%-25% was used during the experiments.

2.2 Experimental Setup

Four SOTA semantic segmentation algorithms were selected to be tested and compared in this study. These were UNet [6], SUMNet [8], ResUNet [9] and Attention UNet [10]. The architectures for these networks were adapted to fit the data specifically. The details are given below.

2.2.1 UNet

The network architecture was designed as per the specifications given in [6]. In order to make the model more robust, Gaussian noise was applied in the input along with histogram equalization to make the network contrast independent. Stochastic Gradient Descent was used as an optimizer along with Binary cross entropy as a loss function. The network was trained for 350 epochs with a batch size of 16 and early stoppage.

2.2.2 SUMNet

The network architecture was designed as per the specifications given in [8]. Adam optimizer was used along with a weighted combination of Dice loss and binary cross-entropy as loss functions. Additionally, a learning rate of 1e-4 and a beta1=0.99 was set for the network. The network was trained for 350 epochs with a batch size of 16 and early stoppage.

2.2.3 ResUNet

The architecture of the network was designed according to the specifications given in [9]. Similarly to UNet, Gaussian noise was applied in the input, along with histogram equalization so as to make the network contrast independent and linearly scaled to obtain the Gaussian distribution. The network was trained using an Adam optimizer and a learning rate of 1e-4 and beta1= 0.99. Binary cross-entropy was used as a loss function. The network was trained for 350 epochs with a batch size of 32 and early stoppage.

2.2.4 Attention UNet

The architecture of the network was designed according to the specifications given in [10]. The model was trained using the Adam optimizer, along with the momentum=0.99, batch-normalization after each convolutional layer, and deep supervision. The model used binary cross-entropy and Dice loss as loss functions. Gating parameters were initialized so that attention gates can pass through every feature vector at all spatial locations. The network was trained for 350 epochs with a batch size of 32 and early stoppage.

2.3 Evaluation

The evaluation of the output obtained from each algorithm was done using the following three techniques. These are highlighted as follows:

Accuracy: Can be defined as the percentage of true positive outcomes against total outcomes.

Dice Coefficient: Dice coefficient can be defined as $(2 * \text{area of overlap between predicted output and ground truth}) / \text{total number of pixels in both the images}$. Is positively correlated and its value also ranges from 0 to 1.

Intersection over Union (IoU): Intersection-Over-Union (IoU), also called the Jaccard Index, is the most commonly used metric to measure the performance of semantic segmentation. It is defined as the $(\text{area of overlap between the predicted output and the ground truth}) / (\text{area of union between the predicted output and the ground truth})$.

3 Results & Discussion

Table 1 shows the comparison of the results for each of the selected evaluation metrics for US thyroid nodule segmentation. Out of all the proposed architectures, ResUNet yields the highest value for all the metrics, followed by Attention UNet, SUMNet, and UNet. Figure 1 shows the training and validation loss curves for ResUNet. The training curve depicts a slow descent to a reduced loss value. This ensured proper training of the network while preventing over-fitting. The validation curve depicts a rapid descent to a low loss value and is maintained through the remaining epochs.

Table 1: Comparison of results obtained from all four algorithms for the selected evaluation metrics for the semantic segmentation of US thyroid nodules

Model/Metrics	Accuracy (%)	Dice Coefficient	IoU
UNet	79.7	0.766	0.635
SUMNet	83.1	0.795	0.672
ResUNet	89.2	0.857	0.767
Attention UNet	85.4	0.821	0.720

Figure 1: Training (A) and validation (B) loss curves for the ResUNet

Figure 2 shows the segmentation results from two different nodules as examples. The segmentation results are given for each of the algorithms used along with the original images and ground truths. It can be seen that the segmentation outputs reflect the results calculated by the metrics. Among the four selected algorithms, ResUNet results in a

segmentation that is closest to the ground truth compared to the remaining three algorithms.

4 Conclusion

The segmentation of thyroid nodules is a crucial task as it acts as the first step towards the overall classification. This work presents the evaluation results for the segmentation of thyroid US images using four different architectures. From the results, it can be observed that ResUNet architecture outperforms other architectures with a high value for all the metrics used for the evaluation of the model. One of the

gradients to the initial layers using skip connections and this helps in training the network fast in the early stages. The next steps would be testing the developed ResUNet architecture with more data and its subsequent integration in a Computer Aided Diagnostic system for the detection, risk-stratification and classification of thyroid nodules that can be used by the physicians to reduce overall diagnostic subjectivity.

Author Statement

The authors state no conflict of interest. Informed consent was taken. Ethical approval was obtained (RAD362-16/19) at the Otto-von-Guericke University Hospital, Magdeburg.

References

- [1] E. Kollorz, E. Angelopoulou, M. Beck, D. Schmidt, and T. Kuwert, "Using Power Watersheds to Segment Benign Thyroid Nodules in Ultrasound Image Data," *Bildverarbeitung für die Medizin* 2011, pp. 124–128, 2011.
- [2] Keramidas EG, Iakovidis DK, Maroulis D, Karkanis S (2007) Efficient and Effective Ultrasound Image Analysis Scheme for Thyroid Nodule Detection[C]. International Conference on Image Analysis and Recognition. Springer-Verlag, 4633: 1052-1060.
- [3] Iakovidis DK, Savelonas MA, Karkanis SA, Maroulis DE(2006) Segmentation of Medical Images with Regional Inhomogeneities[C]. International Conference on Pattern Recognition. IEEE Computer Society, 3: 976-979.
- [4] Chan TF, Vese LA (2001) Active Contours Without Edges. *IEEE Trans.Image Processing*. 7: 266-277.
- [5] J. Ma, F. Wu, T. Jiang, Q. Zhao, and D. Kong, "Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 11, pp. 1895–1910, 2017.
- [6] O. Ronneberger, "Invited Talk: U-Net Convolutional Networks for Biomedical Image Segmentation," *Informatik aktuell*, pp. 3–3, 2017.
- [7] Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 2016 Fourth International Conference on 3D Vision (3DV), 2016.
- [8] S. Nandamuri, D. China, P. Mitra, and D. Sheet, "SUMNet: Fully Convolutional Model For Fast Segmentation of Anatomical Structures in Ultrasound Volumes," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019.
- [9] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.
- [10] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, Daniel Rueckert: Attention U-Net: Learning Where to Look for the Pancreas. 2018 Computer Vision and Pattern Recognition.

Figure 2: Segmentation outputs from each of the algorithms along with the original image and ground truths

major reasons for superior results obtained using ResUNet is the presence of residual blocks of convolutional layers that enables consistent training of the network as the depth of the network increases. Residual blocks also enable easy flow of