

# ROI Segmentation in Thyroid Thermography via Deep Learning

Alessandro C. B. da Silveira<sup>1</sup>, Sílvia C. D. Pinto<sup>1</sup>

<sup>1</sup> Computational Modeling Department, Polytechnic Institute (IPRJ)  
Rio de Janeiro State University (UERJ), Brazil

alessandro.silveira@grad.iprj.uerj.br, silvia.dias@iprj.uerj.br

**Abstract.** *Thyroid cancer is the most frequent endocrine tumor in the world, and its traditional diagnosis depends on invasive methods such as ultrasonography and fine-needle aspiration biopsy (FNAB), which are subject to limitations. Infrared thermography emerges as a non-invasive and accessible alternative examination that records the thermal increase caused by tumor hypermetabolism. Thus, this study proposes an automatic approach through the adaptation of U-Net to obtain segmentation of the region of interest in thermographic images of the thyroid from the Visual Lab Thyroid Database. It was concluded that the proposed model achieved good performance (Dice: 0.9066; IoU: 0.8312; Accuracy: 0.8913), but still faces challenges in the delineation of fine contours due to the reduction of convolutional blocks in the encoder and loss of resolution. The results highlight the potential of the solution to automate initial screening, suggesting as next steps the use of K-fold cross-validation and morphological operations to refine the obtained masks.*

## 1. Introduction

Thyroid cancer is the most common neoplasm of the endocrine system, with an incidence that has been increasing globally in recent decades. The clinical investigation of these nodules is fundamental, since, although the majority are benign, the exclusion of malignancy is mandatory to ensure successful treatment and increase survival rates [González et al. 2017, Sebastiao and Ulbricht 2020, Damião et al. 2020, Júnior et al. 2021].

Among the conventional diagnostic methods are: ultrasonography, which despite being the first choice, frequently presents nonspecific results and high false-positive rates (reaching 50% in some studies), while fine-needle aspiration biopsy (FNAB), although it is the gold standard, is an invasive procedure that produces indeterminate results in up to 30% of cases [Sebastiao and Ulbricht 2020], and computed tomography, which can expose the patient to ionizing radiation, representing an additional potential health risk [Moran et al. 2018a, Damião et al. 2020].

Considering these limitations, infrared thermography emerges as an alternative and attractive tool for being a non-invasive, painless, extremely low-cost, and fast method, which requires no contact with the patient nor uses ionizing radiation. The physiological justification for its use lies in the fact that malignant tumors promote angiogenesis (formation of new blood vessels) and have high metabolic rates, which generates an increase in local temperature compared to

adjacent healthy tissues [González et al. 2017, Moran et al. 2018b, Moran et al. 2019, Sebastiao and Ulbricht 2020, Damião et al. 2020].

Since the thyroid is a superficial gland, located a few centimeters from the skin, the heat emitted by metabolic processes characteristic of tumors can be precisely captured by infrared cameras and transformed into thermal maps [Moran et al. 2018a]. The recent development of methods based on Artificial Intelligence (AI), specifically Convolutional Neural Networks, is justified by achieving accuracy rates of up to 97% in nodule detection, positioning thermography as a promising complementary technology for medical screening and diagnostic support [González et al. 2017, Damião et al. 2020, Júnior et al. 2021].

Regarding the types of thermographic examinations, there are two main acquisition approaches: static and dynamic infrared thermography. While static thermography captures thermal data at a single moment, conveying instantaneous information on surface temperatures for analysis of parameters such as minimum, maximum, and mean, dynamic thermography continuously monitors thermal variations after applying a stress (usually cold), recording a sequence of images to analyze the tissue response to rewarming, which is especially useful in thyroid studies [González et al. 2017, Vicnesh et al. 2024].

The advantages of dynamic thermography include greater thermal contrast between normal and altered tissues, detailed analysis of the physiological stress response, better differentiation between malignant and benign nodules, greater reliability by standardizing responses with less influence from external factors, and the ability to enable early and accurate detection of metabolic changes.

Studies conducted on medical images have demonstrated that infrared thermography, especially when combined with AI techniques and advanced statistical analysis, shows robust performance in the classification and detection of thyroid anomalies [Moran et al. 2018a]. The evolution of studies indicates that the method has surpassed traditional screening techniques in sensitivity and has achieved accuracy levels above 95% in computational models over the last decade [Alves and Gabarra 2016, Moran et al. 2019, Júnior et al. 2021].

Furthermore, methods that previously perform segmentation of the region of interest (ROI) of the image, the neck in this study, have achieved more accurate classification of thyroid tumors. Since the thyroid is a superficial gland [González et al. 2017, Moran et al. 2018a], isolating this area helps reduce noise and irrelevant data from other body parts, making the final accuracy dependent on the success of the segmentation. As an example, models using ResNet achieved more than 92% after isolating asymmetric thermal areas [Moran et al. 2019, Damião et al. 2020], while [Júnior et al. 2021] achieved 97% accuracy using prior manual cropping before processing cascaded CNNs.

Segmentation allows for the extraction of morphological and local statistical metrics [Alves and Gabarra 2016, Moran et al. 2019]. It also reduces false positives by filtering out naturally warm regions, such as the clavicle. Therefore, the objective of this work is to present an automatic and efficient approach for ROI segmentation, which we consider to be the essential initial stage for AI algorithms to focus on the real thermal patterns for diagnosis.

## 2. Materials and Methods

The following sections present, in detail, the steps developed in this work to obtain the segmentation of the region of interest (ROI) in thyroid thermography images, as presented in the pipeline of Figure 1.

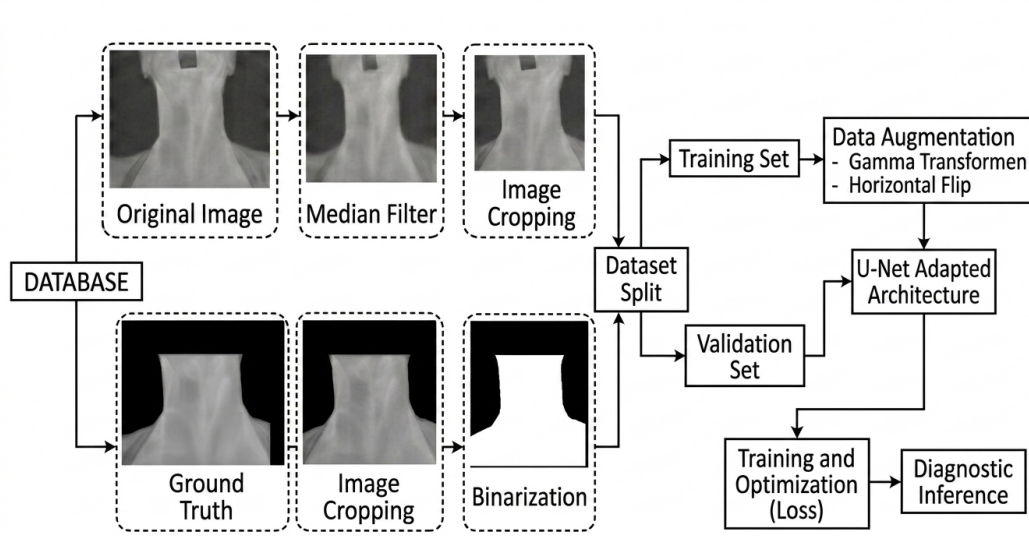


Figure 1. Pipeline of the proposed ROI segmentation model.

### 2.1. Database

In this work, we chose to work with the public repository of thermographic images from the Visual Lab of Universidade Federal Fluminense (UFF), known as the Visual Lab Thyroid Database. The images are captured by the FLIR SC620 camera (640x480 pixels, thermal sensitivity  $< 0.04^{\circ}C$ ) and follow a dynamic acquisition protocol with thermal stress, generating time series of 20 or 21 images per patient (one acquisition every 15 seconds) [González et al. 2017, González et al. 2021]. Figure 2 illustrates an example of an image obtained during the thermographic examination.

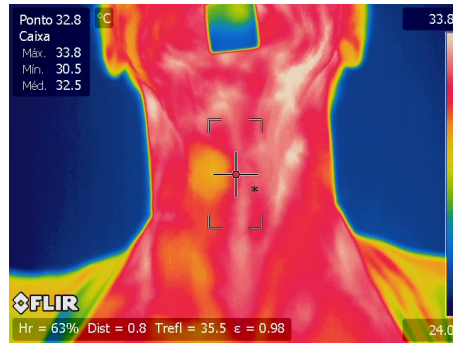


Figure 2. Image obtained through thyroid thermography examination.

This database consists of 2231 grayscale thermographic images (as well as their corresponding color images) of the region comprising the neck, chin, and upper chest of a total of 111 patients. It also includes the region-of-interest masks, called ground

truth (GT), of 2173 images obtained by an edge detection-based algorithm, as well as the manual segmentation, performed by specialists, of nodules from 16 patients. In this case, each image in the database is additionally accompanied by a file containing the annotation made by the specialist indicating the location of the nodule, and a file with the corresponding temperature matrix, which records the thermal value of each pixel in the image.

The use of this database acquired through the dynamic protocol provides a larger number of images compared to the static protocol, and this is beneficial for the segmentation approach proposed in this work, since the training step requires a large number of samples for efficient learning. In this case, it is worth noting that we will not consider the temporal correlation between samples.

## **2.2. Image Preprocessing**

This step aims to process the images in order to reduce acquisition noise, standardize the dimensions of the original images, and prepare the data for the supervised training step of the neural network.

The first step consists of applying a median filter, which is a non-linear smoothing technique widely used in medical image processing, especially effective at suppressing salt-and-pepper impulsive noise. The definition of the median filter window size was conducted through qualitative evaluation, comparing the  $3\times 3$ ,  $5\times 5$ , and  $7\times 7$  pixel configurations on representative samples of the dataset. The  $3\times 3$  window proved insufficient for adequate reduction of the impulsive noise characteristic of infrared thermographic images, while the  $7\times 7$  window introduced blurring in the edges of the anatomical structures of interest, with perceptible loss of sharpness in thermal transition regions relevant to the segmentation task. The  $5\times 5$  configuration was then chosen as it provided the best qualitative trade-off between noise attenuation and preservation of the spatial resolution of clinical structures.

The second step consisted of a centered crop applied to both the original images and the GT images provided by the database, both filtered. Thus, the images, originally with a resolution of  $640\times 480$  pixels, were reduced to  $480\times 480$  pixels by means of a symmetric cut relative to the geometric center pixel of the image. This operation eliminated the peripheral regions, which have low clinical relevance and do not contribute to the segmentation task, while simultaneously reducing the number of pixels processed per sample, decreasing the memory demand and computational cost associated with training the neural network.

The third step corresponded to the binarization of the GT images that will serve as reference for the automatic generation of ROI masks. The conversion was performed using a fixed threshold, producing a binary mask that precisely delineates the annotated structures. The threshold was obtained from a prior analysis of a subset of images using the Simple Global Thresholding method [Gonzalez and Woods 2018]. Once this threshold was chosen, during the loading of masks at the training stage, thresholding was performed with a threshold of 0.5, with values above this threshold converted to 1 and the remaining values to 0. This step ensured strict binariness of the masks, correcting any intermediate values introduced by interpolation or numeric type conversion operations during processing.

Finally, the fourth step consisted of normalizing the pixel values of the input images. The original values, in the integer range  $[0, 255]$ , were converted to the continuous range  $[0.0, 1.0]$  by dividing by 255. This normalization is a standard practice in deep neural network training, as it ensures that the input values are on a scale compatible with the activation functions used, stabilizes the gradient optimization process, and contributes to faster and more stable convergence of the model.

### 2.3. Data Augmentation

Considering the filtered and resized images, to expand the diversity of the training set and mitigate the risk of overfitting, we decided to apply Data Augmentation techniques in an offline manner, that is, the image transformations were applied and stored prior to training, resulting in a static and reproducible dataset [Shorten and Khoshgoftaar 2019]. The transformations used were: (i) Horizontal Flipping, applied simultaneously to the images and their respective binary GT masks, preserving the spatial correspondence between both; and (ii) Gamma Correction with parameter  $\gamma = 1.5$ , applied exclusively to the input images of the neural network according to the transformation:

$$I' = 255 \cdot (I/255)^\gamma$$

The value of  $\gamma$  was selected empirically after evaluating multiple values; values below produced insufficient enhancement of low-contrast structures, while values above resulted in excessive saturation, compromising the discrimination between adjacent tissues. The adopted value promoted moderate enhancement of brightness and contrast, simulating lighting variations recurrent in real clinical environments [Gonzalez and Woods 2018].

Thus, the combination of these operations generated four versions per sample: the original filtered and resized image, its horizontally flipped version, and both versions subjected to gamma correction, each accompanied by its respective binary GT mask. As a result, the training set expanded from 2173 to 8692 samples, representing a 4-fold expansion factor. This procedure is aligned with established practices in the medical image segmentation literature, in which offline data augmentation has demonstrated significant contribution to the generalization of deep learning models on limited-size datasets [Ronneberger et al. 2015].

## 3. Segmentation

### 3.1. Network Architecture Model: Adapted U-Net

The architecture adopted in this work is based on the U-Net originally proposed by [Ronneberger et al. 2015] for biomedical image segmentation, with three targeted modifications justified by the characteristics of the dataset and the binary segmentation task.

The first modification consists of reducing the number of convolutional blocks in the encoder from four to three, totaling four processing levels including the bottleneck block. This adaptation reduces the model's abstraction capacity in exchange for lower computational cost and lower risk of overfitting given the limited size of the available dataset. The second modification is the inclusion of Batch Normalization after each convolutional layer, a practice established in the literature after the original U-Net that stabilizes training and allows the use of higher learning rates [Ioffe and Szegedy 2015].

He Normal initialization [He et al. 2015] was equally adopted in the convolutional layers, being the standard recommended choice for layers with ReLU activation as it preserves the gradient variance throughout the network. The third modification is the addition of a Dropout layer with a rate of 50% in the bottleneck block [Srivastava et al. 2014], functioning as a regularizer to mitigate the overfitting expected in scenarios with reduced medical datasets.

The remaining elements faithfully follow the original architecture, that is, each convolutional block consists of two 3×3 convolutional layers with ReLU activation function, the number of filters doubles at each encoder level (64, 128, and 256, respectively) with a bottleneck block of 512 filters, downsampling operations use 2×2 MaxPooling, and the decoder performs progressive reconstruction of spatial resolution through 2×2 transposed convolutions (Conv2DTranspose) combined with symmetric skip connections via concatenation of feature maps, with convolutional blocks of 256, 128, and 64 filters, respectively. The output layer consists of a 1×1 convolution with sigmoid activation function in replacement of the original softmax, suitable for the binary segmentation task. Training was conducted with batches of size 8, with shuffling of the training set at each epoch, using the TensorFlow/Keras framework.

### 3.2. Loss Function and Evaluation Metrics

The adopted Loss function is a combination of Binary Cross-Entropy (BCE) and Dice Loss functions. This strategy is widely used in medical segmentation under class imbalance [Sudre et al. 2017], and is defined as:

$$L = \text{BCE}(y, \hat{y}) + (1 - \text{Dice}(y, \hat{y}))$$

where the Dice coefficient is given by:

$$\text{Dice}(y, \hat{y}) = \frac{(2 \cdot \sum(y \cdot \hat{y}) + \varepsilon)}{(\sum y + \sum \hat{y} + \varepsilon)}$$

where  $y$  represents the GT segmentation mask and  $\hat{y}$  represents the mask predicted by the model, and  $\varepsilon = 10^{-6}$  for numerical stability. This combination aims to balance the pixel-by-pixel penalization offered by BCE — which tends to be dominated by the majority class (background) — with the sensitivity to class imbalance characteristic of Dice loss.

Regarding the model evaluation metrics for segmentation were used pixel-by-pixel Accuracy, the Dice Coefficient, and the Jaccard Index, known as the Intersection over Union (IoU) metric.

### 3.3. Training

The training stage was performed using the Adam optimizer [Kingma and Ba 2014] with an initial learning rate of  $10^{-4}$ , determined empirically after comparative evaluation of different values. Up to 50 training epochs were defined, with the following regularization and control strategies: ModelCheckpoint, to save the model with best performance on the validation set according to the Dice Coefficient; EarlyStopping with patience of 10 epochs, monitoring validation loss; ReduceLROnPlateau with a reduction factor of 0.2 and patience of 5 epochs, with a minimum rate of  $10^{-6}$ , allowing dynamic adjustment of

the learning rate according to the evolution of validation loss; and continuous per-epoch metric logging via CSVLogger, enabling training resumption in case of interruption.

The training set consisted of 8132 image–mask pairs, corresponding to the four versions generated by data augmentation, that is, the original image, its horizontal flipping, the gamma correction, and the combination of both transformations, applied to the 2033 original samples designated for training. The validation set consisted of the 140 original images from seven acquisition sessions (seven patients), ensuring that the model evaluation reflects performance on real unmanipulated data. Training was performed on an AMD Radeon RX 7600 GPU with ROCm platform support.

## 4. Results

### 4.1. Quantitative Performance

The model was evaluated on the 140 image–mask pairs reserved for validation, as described in Section 3. Training was terminated by EarlyStopping at the end of epoch 10, after identifying the absence of improvement in validation loss from the minimum recorded at epoch 0. The results obtained on the validation set are presented in Table 1.

**Table 1. Performance metrics on the validation set.**

| Epoch     | Dice          | IoU           | Accuracy      | Loss          |
|-----------|---------------|---------------|---------------|---------------|
| <b>0</b>  | 0.8873        | 0.8024        | 0.8729        | 0.5473        |
| <b>1</b>  | 0.9025        | 0.8248        | 0.8898        | <b>0.4707</b> |
| <b>5</b>  | <b>0.9066</b> | <b>0.8312</b> | <b>0.8913</b> | 0.5887        |
| <b>9</b>  | 0.9039        | 0.8265        | 0.8899        | 0.6639        |
| <b>10</b> | 0.8914        | 0.8057        | 0.8788        | 0.7771        |

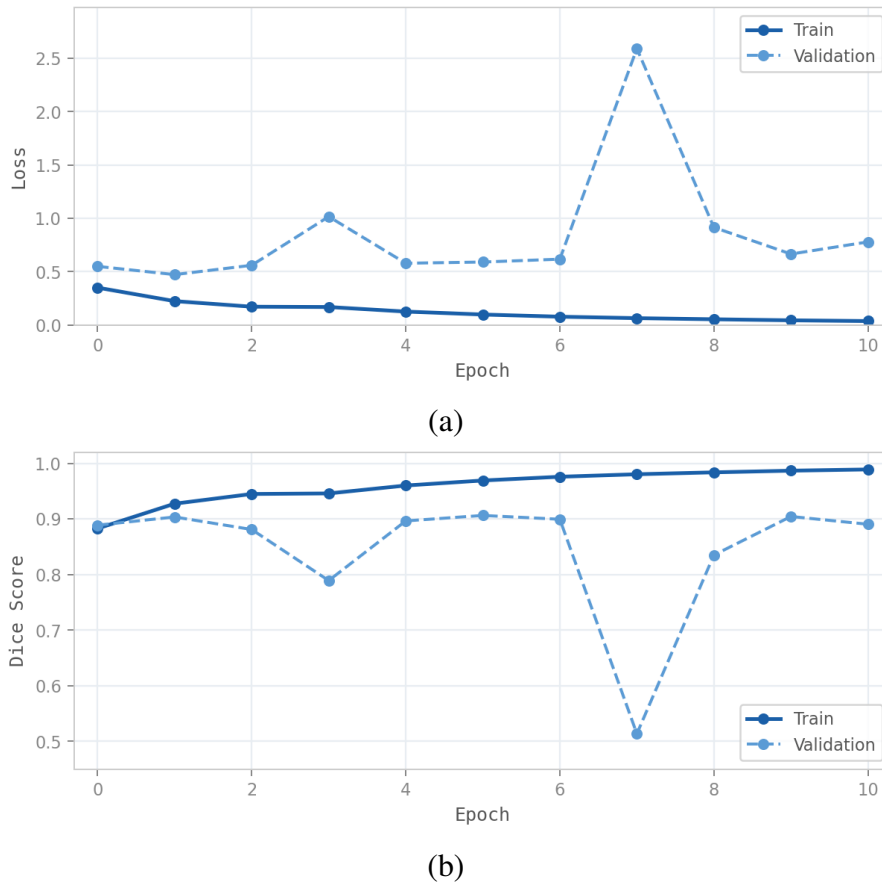
The best validation Dice Coefficient, used as the selection criterion by ModelCheckpoint, was 0.9066, obtained at epoch 5, with a corresponding IoU of 0.8312. These results indicate high overlap between the predicted masks and the GT, considering the thermographic nature of the images, in which the absence of sharp anatomical contrast between tissues represents an intrinsic challenge for the segmentation task.

The rapid convergence observed with a validation Dice greater than 0.88 already at epoch 0 is consistent with the capacity of the adopted architecture, which, by incorporating an additional level of abstraction relative to shallower configurations, allows the capture of hierarchical temperature patterns relevant to the delineation of the regions of interest. In a preliminary experiment with a shallower architecture (two encoder blocks, filters 64, 128, bottleneck with 256 filters and 30% Dropout in the encoder), the maximum validation Dice obtained over 50 epochs was 0.8716, evidencing the contribution of architectural depth to the model’s final performance.

### 4.2. Analysis of the Training Process

The Loss and Dice Coefficient curves over the epochs are presented in Figure 3. Training exhibited consistent progression of training Loss, with a reduction from 0.3488 at epoch 0 to 0.0354 at epoch 10, and a corresponding increase in training Dice from 0.8827 to 0.9888. In contrast, the validation metrics exhibited less regular behavior, with more

pronounced oscillations between consecutive epochs, a phenomenon expected given the moderate size of the validation set. The low cardinality of this set is a determining factor for the observed instability, that is, with few examples, the estimate of any metric has high statistical variance, such that an incorrect prediction on a relatively small number of samples produces proportionally significant variations in the Dice Coefficient between consecutive epochs [Litjens et al. 2017]. This behavior is amplified by the fact that the 140 samples come from only 7 acquisition sessions using the same equipment and protocol, which reduces the statistical diversity of the set and makes it sensitive to specific patterns of a restricted subgroup of patients.



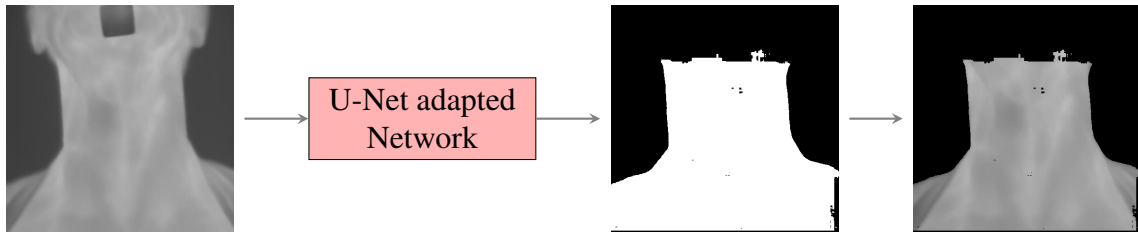
**Figure 3. Evolution of Loss metrics in (a) and Dice Coefficient in (b) over the training and validation epochs.**

The difference between training Dice (0.9888) and validation Dice (0.8914) at the last recorded epoch, of approximately 9.7 percentage points, indicates the presence of moderate overfitting. This behavior is expected in scenarios with a reduced number of real patients (111 patients), even with training set expansion, since the four versions generated per sample are deterministic transformations of the original images and do not introduce genuine acquisition variability. The adopted Data Augmentation strategy contributed to attenuating overfitting relative to the condition without data expansion, but does not eliminate it completely, which is consistent with findings in the literature on limited-size medical image datasets [Ronneberger et al. 2015, Litjens et al. 2017].

### 4.3. Qualitative Analysis

The task performed by the model consists of generating binary masks that delineate the region of interest (ROI) corresponding to the thyroid area in the thermographic image. When applied to the original image, this mask isolates the ROI, eliminating peripheral regions with no clinical relevance. The filtered and resized thermographic image is submitted to the adapted U-Net network, which produces the binary mask corresponding to the predicted ROI; this mask is then applied to the input image, resulting in the segmented ROI image of the thyroid region, as presented in the pipeline of Figure 4.

The visual analysis of the predicted masks relative to the GT reveals two recurring limitation patterns. The first refers to imprecision in the delineation of ROI edges. The majority of masks generated by the model exhibit contours that do not adhere precisely to the GT, manifesting as local expansions or retractions relative to the reference contour. This behavior can be attributed to the nature of the combined BCE + Dice loss function adopted during training, that is, both terms operate on the totality of the image pixels and, given that the edge region represents a narrow strip of pixels at the boundary between the two classes, their proportional contribution to the optimization gradient is reduced relative to the internal regions of the mask, where the concentration of positive pixels is substantially higher. As a consequence, the training signal from the contour pixels is diluted in the global computation of the loss function, limiting the model's ability to refine precision in the delineation of ROI edges. As a result, the model learns to correctly predict the general shape of the ROI with relative ease, but does not receive a sufficiently strong training signal to refine contour precision [Ronneberger et al. 2015].



**Figure 4. Processing pipeline: filtered and resized input image, segmentation via adapted U-Net neural network, result of the generated binary mask, and application of the mask to the input image.**

The second pattern refers to the model's tendency to discard small-dimension components belonging to the ROI. In certain images, regions of interest with a small area that are spatially isolated from the main body of the ROI are omitted from the predicted mask, resulting in incomplete delineation. This behavior stems from the combination of the low number of pixels these sub-areas represent in the total image and the progressive pooling strategy of the adapted U-Net architecture, which, by reducing spatial resolution along the encoder, may lose reference to small structures before they are encoded with sufficient representation in the feature space [Ronneberger et al. 2015]. Although Skip Connections partially mitigate this loss by reintroducing high-resolution spatial information in the decoder, components of very small dimensions may not leave a sufficiently representative trace to be recovered during mask reconstruction.

## 5. Discussion and Conclusions

The present work proposed and evaluated an automatic approach for Region of Interest segmentation in thyroid thermographic images, based on an adaptation of U-Net with three convolutional blocks in the encoder, Batch Normalization, Dropout in the bottleneck, and combined BCE + Dice loss function. The results obtained on the validation set with a Dice Coefficient of 0.9066, IoU of 0.8312, and accuracy of 0.8913, referring to the model saved at epoch 5 by ModelCheckpoint, demonstrate the viability of the proposed methodology to automate the ROI delineation stage, eliminating the dependency on manual segmentation and making the diagnostic workflow more reproducible and scalable.

It is worth noting that the early stopping criterion (EarlyStopping) monitored the validation loss, while ModelCheckpoint independently monitored the validation Dice Coefficient. Due to this separation, the best validation loss was found at an epoch prior to the epoch of best Dice (epoch 5), which enabled training termination when the patience of 10 epochs without improvement in loss was exhausted, even though the validation Dice reached its maximum at epoch 5 and deteriorated in subsequent epochs.

The comparison with the shallower architectural configuration corroborates the relevance of network depth for capturing hierarchical temperature patterns present in thermographic images. The Data Augmentation protocol contributed to containing overfitting, although the difference of approximately 9.7 percentage points between training and validation Dice at the last recorded epoch indicates that genuine acquisition variability between patients, not reproducible by deterministic transformations, represents the main limiting factor for model generalization.

### 5.1. Limitations and Future Perspectives

Given the constraint on the number of patients, the adoption of K-fold Cross-Validation constitutes a direct improvement prospect, as it produces more robust performance estimates that are less sensitive to the specific data split, mitigating the bias introduced by the fixed partition adopted in the present work. In contexts of deep learning applied to medical images with reduced datasets, this approach has demonstrated consistent contribution to more reliable evaluations of model generalization capacity [Litjens et al. 2017].

Regarding improving the quality of the generated masks, there is the possibility of applying mathematical morphology operations as a post-processing step. Morphological closing operations allow filling gaps and internal discontinuities in the predicted regions, while morphological opening suppresses small spurious regions outside the real ROI [Gonzalez and Woods 2018]. These operations act directly on binary masks without modifying the underlying model and represent a low computational cost strategy to specifically mitigate the two limitation patterns identified in the qualitative analysis of the results.

Finally, future investigations could benefit from systematic comparison with alternative semantic segmentation architectures such as U-Net++ and DeepLabV3+, whose structures have demonstrated competitive results in medical image segmentation tasks [Litjens et al. 2017]. The comparative evaluation of these architectures against the adapted U-Net [Ronneberger et al. 2015] would allow identifying which structure best adapts to the particularities of the dataset in question.

## Acknowledgements

The authors acknowledge the use of the generative artificial intelligence tool Claude Sonnet 4.6 (Chatly) to assist in text translation. All conceptual contributions, experimental design, analyses, and scientific conclusions presented in this work are entirely the responsibility of the authors.

## References

- Alves, M. L. D. and Gabarra, M. H. C. (2016). Comparison of power doppler and thermography for the selection of thyroid nodules in which fine-needle aspiration biopsy is indicated. *Radiologia Brasileira*, 49(5):311–315.
- Damião, C., Montero, J., Moran, M., Da, R., Cruz Filho, R., Fontes, C., Lima, G., and Conci, A. (2020). On the possibility of using temperature to aid in thyroid nodule investigation. *Scientific Reports*, 10.
- Gonzalez, R. C. and Woods, R. E. (2018). *Digital Image Processing*. Pearson, New York, 4 edition.
- González, J., Damião, C., and Conci, A. (2017). An infrared thermal images database and a new technique for thyroid nodules analysis. *Studies in health technology and informatics*, 245:384–387.
- González, J. R., Damião, C., Moran, M., Pantaleão, C. A., Cruz, R. A., Balarini, G. A., and Conci, A. (2021). A computational study on the role of parameters for identification of thyroid nodules by infrared images (and comparison with real data). *Sensors*, 21(13).
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *IEEE International Conference on Computer Vision (ICCV 2015)*, 1502.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift.
- Júnior, R. J. A., Silva, I., Silva, A., and Paiva, A. (2021). Detecção de nódulos da tireoide em exames de termografia utilizando redes neurais convolucionais em cascata. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde - SBCAS*, pages 269–280, Porto Alegre, RS, Brasil. SBC.
- Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88.
- Moran, M., Conci, A., and Araújo, A. (2019). Evaluation of quantitative features and convolutional neural networks for nodule identification in thyroid thermographies. pages 747–751.
- Moran, M., Conci, A., Gonzalez, J., Araújo, A., Fiirst, W., Damiao, C., Lima, G., and Cruz Filho, R. (2018a). Identification of thyroid nodules in infrared images by convolutional neural networks. pages 1–7.

- Moran, M., Conci, A., Gonzalez, J., Saade, D., Fiirst, W., Araújo, A., Damiao, C., Lima, G., Cruz Filho, R., and Resmini, R. (2018b). Monomodal image registration by tensor analysis: Going beyond the brightness constancy assumption. pages 1–8.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing.
- Sebastiao, C. and Ulbricht, L. (2020). Termografia como método para auxilio diagnóstico do câncer de tireoide. *Pan American Journal of Medical Thermology*, 5:19.
- Shorten, C. and Khoshgoftaar, T. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958.
- Sudre, C., Li, W., Vercauteren, T., Ourselin, S., and Cardoso, M. J. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. volume 2017, pages 240–248.
- Vicnesh, J., Salvi, M., Hagiwara, Y., Hah, Y., Mir, H., Barua, P. D., Chakraborty, S., Molinari, F., and Acharya, U. (2024). Application of infrared thermography and artificial intelligence in healthcare: A systematic review of over a decade (2013-2024). *IEEE Access*, PP:1–1.