CPP-UNet: Combined Pyramid Pooling Modules in the U-Net Network for Kidney, Tumor and Cyst Segmentation

Caio Eduardo Falcão Matos, Geraldo Braz Junior, João D. S. Almeida, and Anselmo C. Paiva

***Abstract*—Renal carcinoma stands prominently as a significant contributor to global cancer-related mortality rates, highlighting the critical importance of early detection and diagnosis in the management of this ailment. Moreover, the rising incidence of kidney tumors poses a challenge in differentiating between malignant and benign lesions using radiographic methods. Therefore, we present CPP-UNet, an innovative convolutional neural network-based architecture designed for the segmentation of renal structures, including the kidneys themselves and renal masses (cysts and tumors), in a computed tomography (CT) scan. Particularly, we investigate the fusion of the Pyramid Pooling Module (PPM) and Atrous Spatial Pyramid Pooling (ASPP) for improving the UNet network by integrating contextual information across multiple scales. Our proposed method yielded promising outcomes in the Kidney and Kidney Tumor Segmentation challenge (KiTS21 and KiTS23) datasets, exhibiting Dice indices of 93.51% and 92.84% for Kidneys and Masses, 90.33% and 92.08% for Renal Masses, and 85.69% and 88.17% for Tumors, respectively.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/8866

***Index Terms*—Pyramid Pooling Module, Segmentation Renal Diseases, Kidney Cancer, U-Net.**

I. Introduction

Renal function or structure can be altered due to chronic kidney diseases (CKD). This type of condition is characterized by progressive and gradual development. Moreover, the condition significantly elevates the risk of complications and mortality, particularly concerning cardiovascular health [1], [2]. Renal cell carcinoma (RCC), alternatively referred to as renal cell adenocarcinoma, stands as the prevailing variant of kidney malignancy, comprising roughly 90% of all instances of renal cancer [3]. In this type of cancer, the renal tumor may manifest as a single structure or multiple instances, potentially affecting one or both kidneys [4].

Kidney cancer exhibits alarming global statistics, ranking as the 13th most prevalent cancer. Projections for this disease suggest over 430,000 new cases and 179,000 deaths in 2020. Emerging reports highlight the escalating incidence and mortality rates of renal cancer in Latin America, warranting attention and further investigation [5]. In terms of statistics in Latin America, kidney cancer had a notable mortality rate of 4.28 deaths per 100,000 inhabitants, according to research carried out in 2017 [6]. At the same time, several studies in the region have highlighted a consistent increase in the incidence of kidney cancer over the last decade, highlighting the need for a deeper understanding of this trend [6].

In 2024, approximately 81,610 new cases are projected to be diagnosed, with 52,380 cases in men and 29,230 cases in women, according to a study conducted by the American Cancer Society. Moreover, the estimated mortality rate is approximately 14,390 individuals, including 9,450 men and 4,940 women [7].

Therefore, early diagnosis of this condition becomes an essential tool for the prognosis and treatment of the disease, thus increasing the patient's chances of recovery. Patients diagnosed in the early stages of the disease, with no disease dissemination, exhibit a survival rate of approximately 93% over five years. Conversely, when the diagnosis occurs at an advanced stage (with disease metastasis), this rate drops to 12%, underscoring the significance of early diagnosis [8].

Imaging tests are crucial in treating illnesses, offering various tools and techniques for detection, staging, and care. Diagnostic imaging offers detailed information about structural or illness-related changes. Characteristics such as low cost, high-quality imaging, prompt execution, and applicability to different diagnoses make Computed Tomography (CT) one of the most accessible radiological examinations for patients [9]. On the other hand, the interpretation of this type of examination is done manually through visual analysis, requiring maximum attention from the specialist physician. This process can lead to both visual and physical fatigue, which can have a negative impact on the identification of suspicious structures in the examination.

Renal tumors exhibit a complex internal structure, with variations in texture, position, and size. Relying solely on radiological images can result in inaccurate characterization of certain tumors, even by experienced urologists, necessitating the use of pathological reports for a definitive diagnosis. In this context, the automatic segmentation of kidneys, cysts, and renal tumors enables a more precise characterization of lesions, aiding the medical specialist.

Therefore, the accurate segmentation of the kidneys, cysts, and renal tumors enables the specialist physician to conduct more precise analysis and diagnosis of suspicious lesions, contributing to treatment planning. Consequently, this study introduces the CPP-UNet, a new architecture derived from the U-Net that incorporates different Pyramid Pooling blocks.

C. E. F. Matos, G. Braz, J. D. S. Almeida, and A. C. Pailva are with Applied Computing Center, Federal University of Maranhão, São Luís, Brazil (e-mails: caioefalcao@nca.ufma.br, geraldo@nca.ufma.br, jdallyson@nca.ufma.br, and paiva@nca.ufma.br).

Specifically, the Pyramid Pooling Module (PPM) and Atrous Spatial Pyramid Pooling (ASPP) were used to extend the U-Net as part of the encoder, aiming to provide local context information at various resolutions.

In general, the following points can be highlighted as contributions of the proposed method:

- Proposal of a novel convolutional network architecture capable of segmenting kidneys, cysts, and renal tumors in CT images with high precision. Adaptation and incorporation of architectural models (PPM + ASPP + U-Net) not covered in the literature to build a convolutional network capable of multiresolution analysis with local and global contexts.
- Developing a high-precision method for the task of segmenting kidneys, cysts, and renal tumors in CT images. Thus, it is believed that this approach provides effective segmentation that can be integrated into CAD (Computer-aided diagnosis) systems, aiming to increase productivity and improve the diagnosis of renal cancer.

II. Related Works

The first evidence concerning kidney tumors is typically derived from the patient's computed tomography scan by the specialist physician. Therefore, the segmentation of kidneys, cysts, and renal tumors aims the purpose of assisting specialists in accurately diagnosing kidney cancer. Various studies investigate the application and development of image processing techniques and deep learning for the analysis and segmentation of kidneys and suspected tumor regions. Below are presented works in the literature related to the theme of this research.

Approaches that employ successive semantic segmentation networks for object localization in images are called cascade or ensemble architectures. The works [10]–[12] produce a method for segmenting kidneys and renal masses (cysts and tumors). The method proposed in the first study is structured into two stages: general and specific segmentation, which apply a modified version of the U-Net architecture, called nnUNet, to perform semantic segmentation. This work was validated on the KiTS21 image database, achieving best results of 97.52% for kidney and mass segmentation, 88.51% for renal masses, and 86.93% for renal tumors.

On the other hand, in [11], a cascade based on *3D U-Net* networks is implemented in two stages, which seek low and high-resolution spatial contextual information, respectively. The first stage is applied to low-resolution input data for kidney and renal tumor segmentation. In contrast, the second is applied to high-resolution input data to segment kidneys and renal tumors. This approach was applied to the KiTS21 dataset and achieved the best results of 97.15%, 87.90%, and 86.38% Dice coefficients for kidney and mass segmentation, renal masses, and renal tumors, respectively.

The methodology presented by [12] also relies on ensemble techniques, applying the neural network model with 3D U-Net architecture. The network is designed to follow a two-stage cascade network approach. The KiTS21 dataset was used to evaluate the method, which achieved the best results for kidney and mass segmentation, renal masses, and renal tumors, respectively, with scores of 97.60%, 87.60%, and 83.10%.

In [13], an approach is proposed for segmenting kidneys, cysts, and renal tumors in computed tomography (CT) scans using transfer learning. They employ the Pre-trained Residual 3D U-Net with pre-trained weights and the Atrous Spatial Pyramid Pooling module associated with the decoder. As a result, this methodology achieved Dice coefficients of 97.30%, 87.40%, and 82.20% for kidney and mass segmentation, renal masses, and renal tumors, respectively.

Adaptations of the standard U-Net network architecture are widely explored for segmentation in medical images. The works [14], [15] propose variations of this network using attention mechanisms for the segmentation task of kidneys, cysts, and renal tumors. The first study is based on a variation of the U-Net network called DenseU-Net, which was adapted with the attention gate mechanism as skip connections. On the other hand, the second study proposes using another mechanism called Contrast Attention to perform skip connections between the encoder and decoder of the network.

Training and post-processing strategies for the 3D U-Net are explored in [16] aiming to accurately segment kidneys, cysts, and kidney tumors in CT images. The model was validated on the KiTS23 dataset with dice scores for kidney and masses, masses, and tumors: 97.9%, 85.7%, and 82.6%. Strategies include segmentation at different resolutions, resampling the original images, utilizing two independent 3D U-Nets, and multi-scale post-processing.

A methodology employed by [17] is based on Auto3DSeg, an automated 3D medical image segmentation solution. For the segmentation of kidneys, tumors, and cysts, the SegResNet and SwinUNETR networks were applied together using an architecture search mechanism called DiNTS. As a result, this approach achieved 95.6%, 79.2%, and 75.8% Dice for kidneys and masses, masses, and tumors, respectively, when applied to the KiTS23 dataset.

The approach proposed by [18] utilized initial preprocessing with histogram matching associated with the 3D U-Net network and multiple data augmentation techniques. As a result, in the KiTS23 database, it achieved 94.7%, 76.0%, and 71.3% for kidney and masses, masses, and tumors segmentation, respectively.

The method proposed in [19] uses a multi-scale supervised adversarial learning UNet (MSALDS-UNet). This approach applies multiple discriminators along the decoding path of the segmentation network to implement multi-scale adversarial learning and increase the network's segmentation accuracy.

The proposed approaches primarily employ architectures based on the convolutional U-Net network, making modifications to the standard architectural scheme, which demonstrates the efficiency of encoder-decoder networks. However, the traditional architecture presents limitations in relating different features within the global context of the image due to the type of convolution used. Therefore, the CPP-UNet architectural model is proposed to enhance the U-Net convolutional network's ability to identify features at different scales, varying across distinct image sub-regions, by combining PPM and ASPP pyramids.

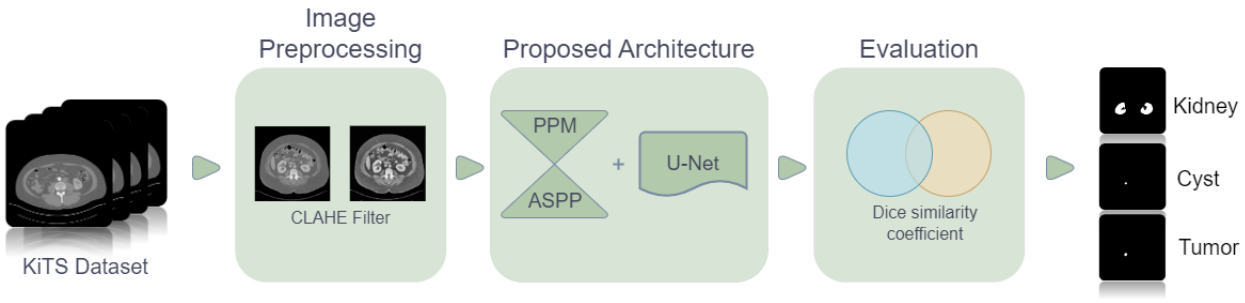


Fig. 1. Steps of the proposed method.

III. Proposed Method

This section presents the procedures to develop the proposed method for segmenting kidneys, renal cysts, and tumors in computed tomography (CT) images. Fig. 1 outlines the four steps developed in this study.

Initially, the KiTS21 and KiTS23 versions [20] of the public image database provided by the segmentation challenge were acquired. Subsequently, preprocessing was performed on all CT volumes, including resizing and enhancing the images. Following that, the architectural model CPP-UNet was proposed, which integrates a Pyramid Pooling Module (PPM) along with Atrous Spatial Pyramid Pooling (ASPP) blocks as part of the encoder of the U-Net convolutional network. This model was applied to the preprocessed image database, and the results obtained were evaluated in the final stage of the methodology.

A. Image Acquisition

This research utilized the 2021 and 2023 editions of the Kidney Tumor Segmentation Challenge (KiTS) image dataset. The primary goal of this challenge is to advance kidney tumor segmentation research and improve the diagnosis and treatment of patients with renal cancer. Both datasets versions are crafted to promote the advancement of automated segmentation systems for kidneys, renal tumors, and cysts. Additionally, each volume provides computed tomography images and ground truth semantic segmentations in NIFTI (Neuroimaging Informatics Technology Initiative) format [21]. This study used slices of the volumes as grayscale images with 512x512 pixels.

The databases comprise 300 and 599 cases for KiTS21 and KiTS23, respectively. Each case on the dataset has a CT exam (relating to a patient). The computed tomography exam is a volume with a variable number of slices and, therefore, a variable number of images. Thus, in this work, all slices of each CT exam were extracted to construct the input dataset for the proposed method. Fig. 2 illustrates a slice and its respective marking provided by the KiTS21 and KiTS23 databases. In this illustration, the colors green, red, and blue correspond to the markings of the kidneys, tumors, and renal cysts, respectively.

In this work, both datasets were used in a 2D context, considering each patient's slices as the input images for the proposed method, in accordance with the scope of each CT exam.

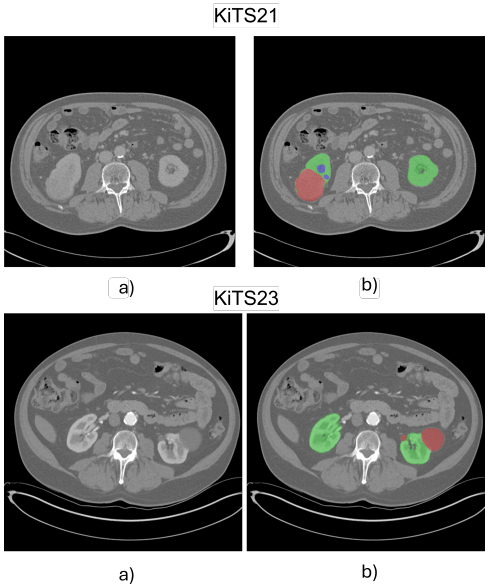


Fig. 2. Example of some images of the dataset. (a) Slice of the exam and (b) Ground truth semantic segmentations.

B. Image Preprocessing

This preprocessing stage aims to reduce the dimensionality of the slices and increase the contrast of the objects of interest (kidneys, cysts, and tumors) about other organs present in the scan. This enables better features for the subsequent stages of the methodology.

Therefore, due to computational limitations, we reduced the slices from (512x512) to ($256x256$) to ensure the application of the proposed architectural model in this study. Next, the CLAHE preprocessing technique [22] was applied using parameters of 2.0 and $8x8$ for clip limit and grid size, respectively. These parameters were chosen empirically after some experiments. This technique works on small parts of the image, with each pixel of the original image located at the center of the contextual region. As a result of this process, the histogram generated by the new image exhibits improved contrasts.

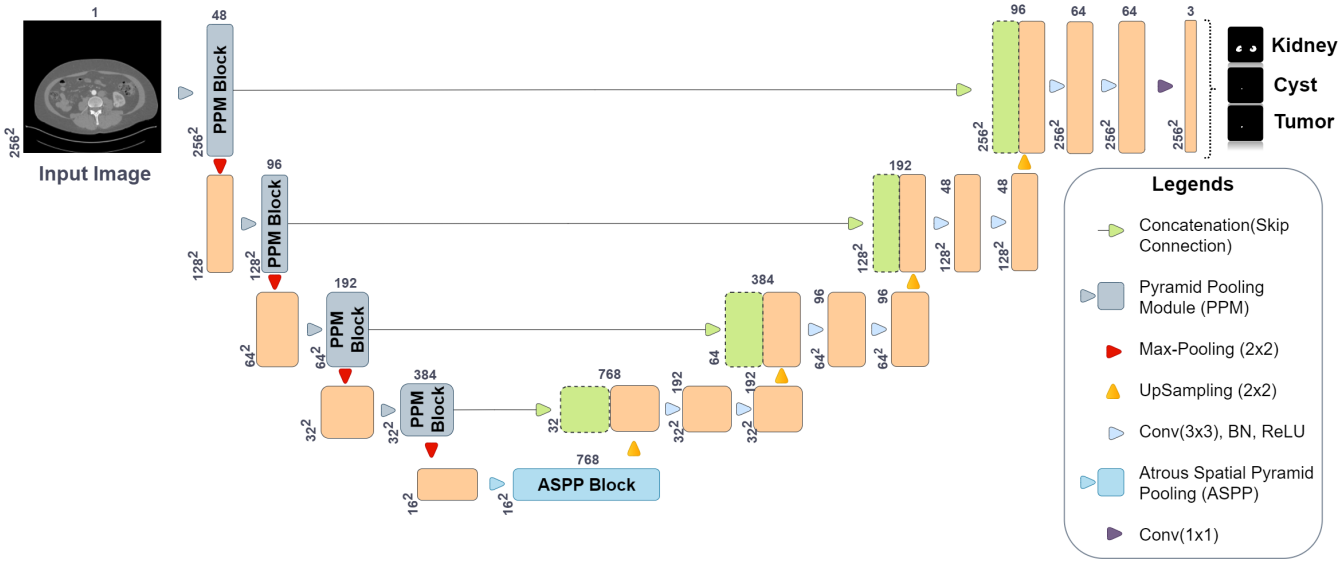


Fig. 3. CPP-UNet architecture.

C. Proposed Architecture - CPP-UNet

Similarly to the traditional UNet network, the proposed network architecture consists of a contraction pathway (left side) and an expansive pathway (right side), as shown in Fig. 3.

Named CPP-UNet, this architecture is based on adapting the standard UNet network structure. The Pyramid Pooling combination replaces the feature extractor network in the standard UNet Encoder. The encoder follows the conventional configuration of a convolutional network. The procedure involves successively applying the PPM block at each encoder level, followed by a 2x2 kernel max-pooling with a stride of 2 for dimensionality reduction. The number of feature maps is doubled at each downsampling step. Fig. 4(i) demonstrates an overview of this block that employs four simultaneous convolutions on the input feature map, with kernels of size 1x1, 3x3, 5x5, and 7x7, each followed by ReLU and BatchNorm. After execution, the block ends by concatenating all four previous convolutions.

After executing the encoder, the resulting feature map is directed to the ASPP module (Atrous Spatial Pyramid Pooling). An overview of the ASPP block can be seen in Fig. 4(ii). This module applies a clustering pyramid using dilated convolutions (atrous convolutions), allowing you to capture features and observe objects at various scales. The ASPP module is specially designed to identify objects in global contexts in images [23]. Subsequently, in the decoding phase, there is the recurrent application of upsampling on the feature map, followed by concatenation with the corresponding cropped feature map from the encoding stage, and two 3x3 convolutions, each followed by ReLU and BatchNorm.

This combination, in turn, is developed by applying the Pyramid Pooling Module (PPM) and Atrous Spatial Pyramid Pooling (ASPP) together. This step of the CPP-UNet architecture aims to transform the input CT images into a more abstract and compact feature space, where patterns such as kidneys, cysts, and renal tumors can be efficiently identified and encoded.

IV. Results

This section details the experiment results used to validate the proposed method. We introduce the metrics used for performance validation, outline and discuss the results obtained in each method step, and provide a comparative analysis with other applied architectural models and related works.

We emphasize that the proposed approach is based on the composition of pooling pyramids as feature extractors in the UNet network for segmenting kidneys, renal cysts, and tumors.

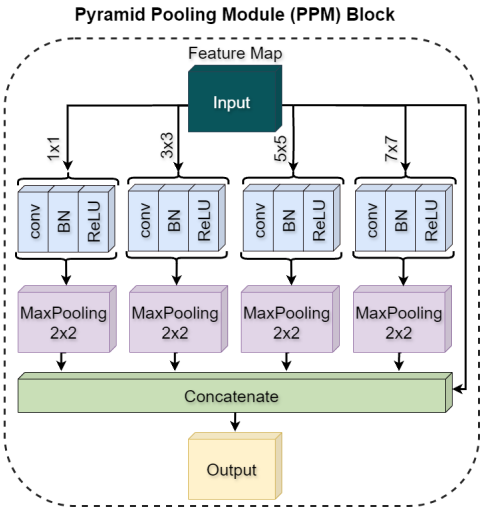
A. Evaluation Metrics

The accuracy of the proposed model was evaluated using metrics commonly employed in medical image segmentation methods in CAD/CADx systems. The metrics comprise the Dice similarity coefficient and the Jaccard index [24]–[26].

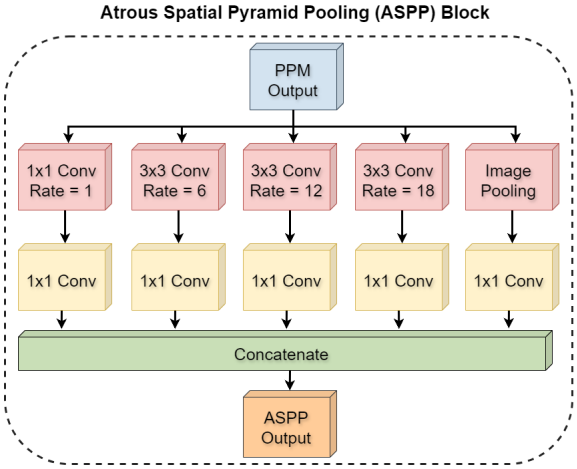
B. Dataset Preparation

Our experiments used the KiTS21 and KiTS23 datasets with their respective quantities of cases: 300 and 599. Thus, for the first dataset, the 300 provided cases were divided into three data sets respecting the proportions of 70% (210 cases) for training, 20% (60 cases) for testing, and 10% (30 cases) for validation. On the other hand, out of the 599 cases in the KiTS23 dataset made available by the challenge, only 489 are provided for model training with annotated regions. On the other hand, the KiTS23 dataset specifies that 489 cases are allocated for training and 110 cases for testing and evaluation. However, the test set is used exclusively to classify the proposed methods submitted to the challenge. Therefore, in this study, we subdivided the 489 cases, following the criteria of 70% (342 cases), 20% (97 cases), and 10% (48 cases) for training, testing, and validation, respectively.

The dataset was divided into training, validation, and testing sets by referencing the individual cases to ensure the absence

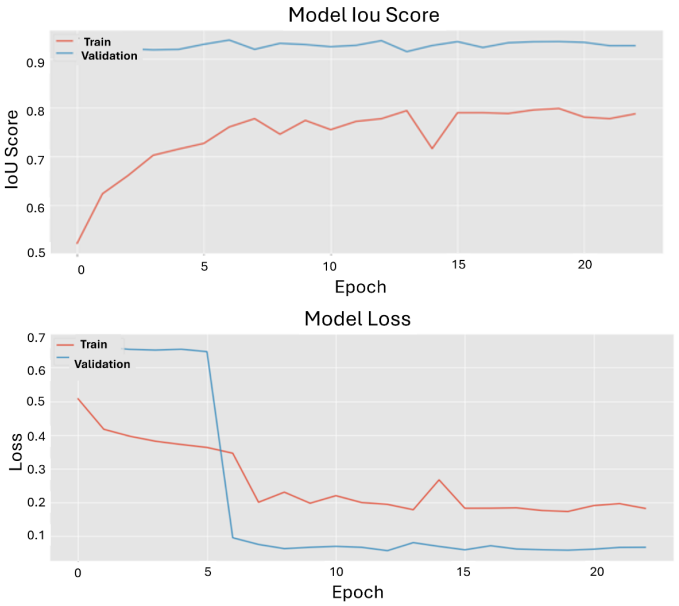


((i)) Pyramid Pooling Module block: convolution, batch normalization, and ReLU operations (Blue Blocks) with kernel sizes 1x1, 3x3, 5x5, and 7x7, followed by 2x2 max pooling (Purple Blocks). The output concatenates the pooling operations and the input feature map.

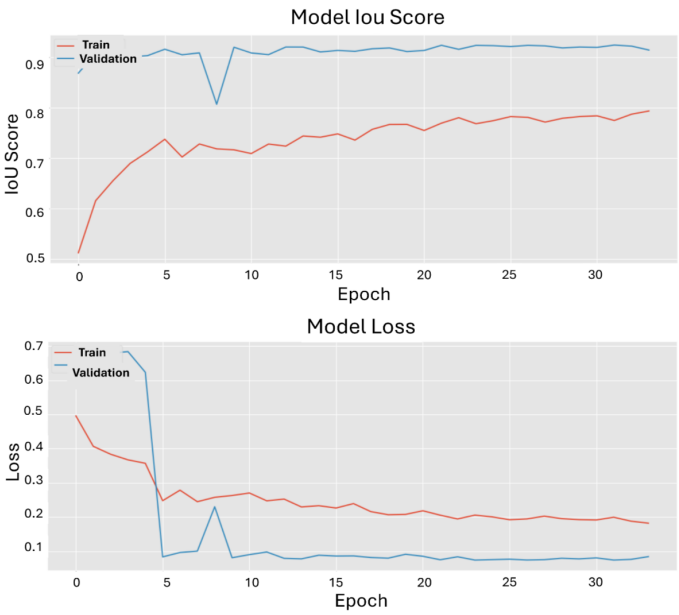


((ii)) Atrous Spatial Pyramid Pooling block: 4 Atrous Convolutions (rates 1, 6, 12, and 16) and image pooling (Red Blocks), followed by 1x1 convolutions (Yellow Blocks). All blocks are concatenated to produce the output.

Fig. 4. Example of blocks PPM and ASPP architecture.



((i)) Model applied to the KiTS21



((ii)) Model applied to the KiTS23

Fig. 5. CPP-UNet model training curves with accuracy (IoU Score) and error function (Loss) information.

of bias in the methodology. This approach ensured that slices from the same case were present exclusively in one of the sets—either training, validation, or testing—without overlapping across multiple sets.

The KiTS challenge dataset demonstrates a notable class imbalance, with kidneys, tumors, and cysts represented unequally. For instance, the KiTS21 dataset contains 65,164 slices, but only 22,997 are labeled as kidneys, 8,341 as tumors, and 2,869 as cysts. Similarly, the KiTS23 dataset comprises 95,221 slices, with 32,375 labeled as kidneys, 12,182 as tumors, and 4,677 as cysts. This imbalance significantly hampers the development of accurate segmentation techniques for these categories, highlighting the need for further research and attention.

C. Segmentation

As described in Section III-C, kidney, renal cyst, and renal tumor segmentation is achieved using the proposed CPP-UNet architecture. This architecture was trained for 100 epochs, using a batch size of 8, Adam optimizer with a learning rate 0.0001, and DiceLoss + CategoricalFocalLoss as the loss function. Parameters and techniques used in the training phase, it was made through a vast set of previous experiments, with this particular set producing the best results. Fig. 5 shows the graph of the error function and IoU during training.

Additionally, we applied other convolutional networks based

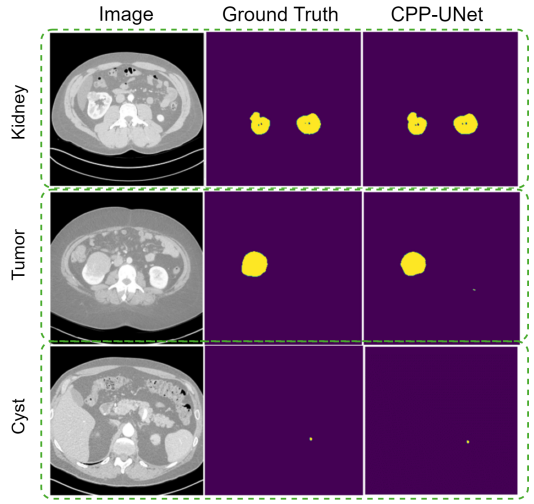
TABLE I
RESULTS OBTAINED IN MODELS CONSTRUCTION STEP

KiTS21 dataset							
Model	Dice - Kidney	Dice - Tumor	Dice - Cyst	Jaccard - Kidney	Jaccard - Tumor	Jaccard - Cyst	Loss
UNet (Standard)	92,60%	83,59%	97,14%	90,74%	83,59%	**97,14%**	0.069711
PPM-Deeplabv3+	93,31%	85,11%	94,69%	90,95%	84,73%	94,67%	0.071176
CPP-UNet (Our)	**93,45%**	**85,69%**	94,97%	**91,39%**	**85,06%**	94,97%	**0.06714**
KiTS23 dataset							
Model	Dice - Kidney	Dice - Tumor	Dice - Cyst	Jaccard - Kidney	Jaccard - Tumor	Jaccard - Cyst	Loss
UNet (Standard)	94,21%	88,08%	95,35%	92,34%	87,36%	95,35%	0.058371
PPM-Deeplabv3+	90,81%	83,59%	94,19%	88,28%	82,89%	94,19%	0.084479
CPP-UNet (Our)	**94,36%**	**88,17%**	**95,99%**	**92,52%**	**87,38%**	**95,99%**	**0.05618**

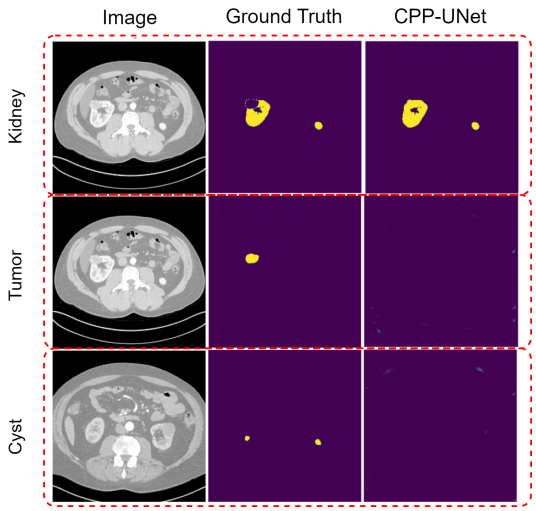
on PPM and ASPP to compare the results obtained. Table I compares the results obtained by the UNet (Standard), PPM-Deeplabv3+ [27], and CPP-UNet models developed in this study. As a result, CPP-UNet outperforms Standard U-Net and PPM-UNet when applied to the KiTS21 and KiTS23 datasets. With this proposed architecture, we achieved a tumor Dice result of 85.69% on the first dataset and 88.17% on the second dataset. The results demonstrate that the proposed architecture CPP-UNet achieved the best performance in segmenting kidneys, cysts, and kidney tumors compared to the standard UNet architecture in both the KiTS21 and KiTS23 databases. The PPM-Deeplabv3+ network, which combines the PPM and ASPP pyramid modules, was also used as a comparison criterion. The results indicate that the proposed CPP-UNet model performs better in the KiTS21 and KiTS23 databases.

Examples of kidney, cyst, and renal tumor segmentation performed by CPP-UNet are available in Figs. 6 and 7 for the KiTS21 and KiTS23 datasets, respectively. Observing in detail the segmentation results in cases of error, we noticed that the features of the tumor regions and the kidneys have similar aspects in some cases, thus making it difficult to differentiate them using the proposed method, causing incorrect segmentation. Another observation about these cases is the complexity of segmenting the cysts, as they comprise a small portion of the kidneys, thus requiring a method with high sensitivity in identifying such structures. It is observed that, in both examples, the proposed model effectively segmented the kidneys, cysts, and renal tumors compared to expert annotations. It is worth noting that the proposed architecture achieved accurate segmentation for slices containing renal tumor segmentation, even with samples of tumors of varying sizes.

These results suggest that combining the PPM and ASPP pyramid modules as feature extractors (encoder) of the UNet network resulted in a robust encoder-decoder architecture with high performance in segmenting kidneys, cysts, and kidney tumors.



((i)) Success Cases: Examples of accurate segmentation of kidneys, cysts, and tumors performed by CPP-UNet.



((ii)) Error Case: Example of incorrect segmentation of kidneys, cysts, and renal tumors performed by CPP-UNet.

Fig. 6. Examples of segmentation cases CPP-UNet architecture applied to the KiTS21 dataset

D. Comparison with Related Works

Table II compares the proposed methodology with the benchmark studies in the KiTS21 and KiTS23 databases. As per the challenge, the results for the classes of kidneys, cysts, and tumors must be combined to compute a metric for the entire set. Hence, the macro-classes are defined as Kidneys and Masses (Kidney + Tumor + Cyst), Masses (Tumor + Cyst), and Tumor.

Comparatively, the method proposed in this study achieved results comparable to the state-of-the-art method for segmenting kidneys, cysts, and renal tumors. Furthermore, excellent

TABLE II
COMPARISON OF METHODOLOGY RESULTS WITH RELATED WORKS

Work	Method	Dice - Kidney and Masses	Dice - Masses	Dice - Tumor	Average Dice
		KiTS21			
Zhao et al. 2022 [10]	Cascaded nnU-Net	**97.52%**	**88.51%**	86.93%	**90.99%**
Golts et al. 2022 [11]	Ensemble of 3D U-Net	97.15%	87.90%	86.38%	90.48%
George 2022 [12]	Cascaded 3D U-Net	97.60%	87.60%	83.10%	89.43%
Zheng et al. 2024 [19]	Multi-scale adversarial UNet	96.10%	84.49%	**87.46%**	89.35%
Yang et al. 2022 [13]	Transfer Learning U-Net	97.30%	87.40%	82.20%	88.97%
Sun et al. 2022 [14]	Attention gate DenseU-Net	97.10%	81.20%	81.50%	88.70%
Wu and Liu 2022 [15]	CAU-UNet	97.00%	86.30%	81.10%	88.10%
Proposed Method	**CPP-UNet**	**93.51%**	**90.33%**	**85.69%**	**89.85%**
		KiTS23			
Uhm et al. 2023 [16]	**3D U-Net with post-processing**	**97.9%**	**85.7%**	**82.6%**	**88.7**
Myronenko et al. 2023 [17]	SwinUNETR and SegResNet	95.6%	79.2%	75.8%	83.5
Stoica et al. 2023 [18]	3D U-Net with data augmentation	94.7%	76.0%	71.3%	80.7
Proposed Method	**CPP-UNet**	**92,84%**	**92,08%**	**88,17%**	**91,03%**

results were obtained for renal tumor segmentation, achieving high Dice rates in both datasets (KiTS21 and KiTS23).

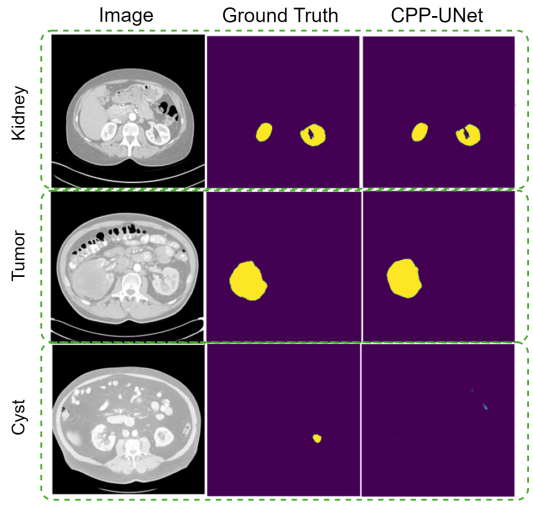
In the KiTS21 dataset, the model achieved a tumor Dice of 85.69%, corresponding to a difference of only 1.24 p.p. compared to the method proposed by Zhao et al. (2022), the challenge winner. When compared with Zheng et al. (2024), we observe that the proposed network achieved comparable results, with a difference of only 1.77 percentage points, positioning it close to the state-of-the-art in renal tumour segmentation. Compared to the current state-of-the-art KiTS23 dataset, the method of this study achieved a Dice of 88.17%, being approximately 5% higher than the work of Uhm et al. (2023), which is part of the top 5 in the challenge.

These results demonstrate the potential of using PPM and ASPP integrated into the UNet network, enabling the construction of a robust architectural model based on extracting local and global features at various resolutions for renal cancer segmentation.

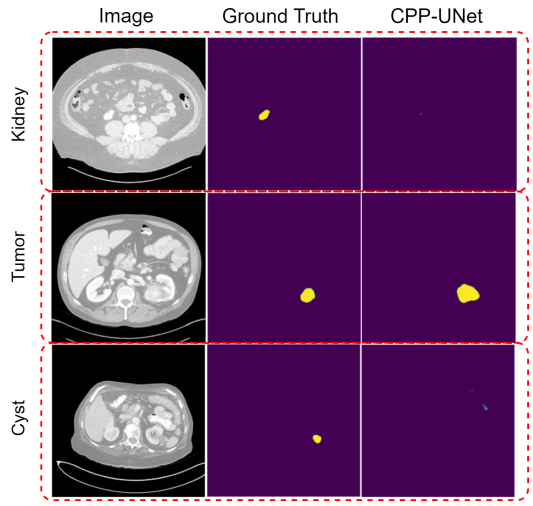
V. CONCLUSIONS

The semantic segmentation of organs and neoplasms in computed tomography presents itself as a challenge when considering the complexity of differentiation among the structures contained in the abdominal region.

CPP-UNet, a newly elaborated architectural model, emerges as a focal element in developing the proposed methodology, which is based on widely applied convolutional neural networks. To accomplish this, the pyramid pooling modules PPM and ASPP were incorporated as the encoder and feature extractor of the UNet network for segmenting kidneys, cysts, and renal tumors in CT images. The proposed method achieved a Dice of 93.51% and 92.84% for Kidneys and Masses, 90.33% and 92.08% for Renal Masses, and 85.69% and 88.17% for Tumors, standing out among the promising methods found in the literature in both databases. Furthermore, the combined use of these pooling pyramids integrated into the UNet encoder emerges as an area of research that allows for developing new models based on medical image segmentation. The proposed method has some limitations, including a high number of adjustable parameters, class imbalance in the database, and



((i)) Success Cases: Examples of accurate segmentation of kidneys, cysts, and tumors performed by CPP-UNet.



((ii)) Error Case: Example of incorrect segmentation of kidneys, cysts, and renal tumors performed by CPP-UNet.

Fig. 7. Examples of segmentation cases CPP-UNet architecture applied to the KiTS23 dataset

the need for mechanisms to fine-tune the PPM block output to better differentiate cysts and tumors.

In future research, we suggest the use of advanced hyperparameter optimization techniques, such as Hyperopt, Scikit Optimize and Optuna, to explore parameters related to the number of blocks used, the depth of the pooling pyramid and the resolutions employed in both blocks (PPM and ASPP). We suggest applying data augmentation techniques to address the class imbalance in the dataset. Additionally, incorporating an attention mechanism to filter the PPM block output could improve the sensitivity of the proposed method in differentiating between kidneys, cysts, and renal tumors with similar characteristics. Overall, the outcomes indicate that the proposed model, CPP-UNet, holds promise for integration into a Computer-Aided Diagnosis (CAD) system, with the goal of aiding in the diagnosis of kidney cancer.

ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001, Fundação de Amparo à Pesquisa Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) (Brazil), and Empresa Brasileira de Serviços Hospitalares (Ebserh) Brazil (Grant number 409593/2021-4). We appreciate the financial support.

REFERENCES

[1] A. L. Ammirati, "Chronic kidney disease," *Revista da Associação Médica Brasileira*, vol. 66, pp. s03–s09, 2020. DOI:https://doi.org/10.1590/1806-9282.66.S1.3.

[2] A. Chang, A. Finelli, J. S. Berns, and M. Rosner, "Chronic kidney disease in patients with renal cell carcinoma," *Advances in chronic kidney disease*, vol. 21, no. 1, pp. 91–95, 2014. DOI:https://doi.org/10.1053/j.ackd.2013.09.003.

[3] T. Gansler, S. Fedewa, M. B. Amin, C. C. Lin, and A. Jemal, "Trends in reporting histological subtyping of renal cell carcinoma: association with cancer center type," *Human pathology*, vol. 74, pp. 99–108, 2018. DOI:https://doi.org/10.1016/j.humpath.2018.01.010.

[4] R. Siegel, J. Ma, Z. Zou, and A. Jemal, "Cancer statistics, 2014.," *CA: a cancer journal for clinicians*, vol. 64, no. 1, pp. 9–29, 2014. DOI:https://doi.org/10.3322/caac.21208.

[5] Q. Cai, Y. Chen, X. Qi, D. Zhang, J. Pan, Z. Xie, C. Xu, S. Li, X. Zhang, Y. Gao, *et al.*, "Temporal trends of kidney cancer incidence and mortality from 1990 to 2016 and projections to 2030," *Translational Andrology and Urology*, vol. 9, no. 2, p. 166, 2020. DOI:https://doi.org/10.21037/tau.2020.02.23.

[6] X. Bai, M. Yi, B. Dong, X. Zheng, and K. Wu, "The global, regional, and national burden of kidney cancer and attributable risk factor analysis from 1990 to 2017," *Experimental Hematology & Oncology*, vol. 9, pp. 1–15, 2020. DOI:https://doi.org/10.1186/s40164-020-00181-3.

[7] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA: A Cancer Journal for Clinicians*, 2024. DOI:https://doi.org/10.3322/caac.21820.

[8] American Cancer Society, "Kidney cancer early detection, diagnosis, and staging." Available online: https://www.cancer.org/content/dam/CRC/PDF/Public/8661.00.pdf (accessed on 06 February 2024), February 2020.

[9] E. Bercovich and M. C. Javitt, "Medical imaging: from roentgen to the digital revolution, and beyond," *Rambam Maimonides medical journal*, vol. 9, no. 4, 2018. DOI:https://doi.org/10.5041/RMMJ.10355.

[10] Z. Zhao, H. Chen, and L. Wang, "A coarse-to-fine framework for the 2021 kidney and kidney tumor segmentation challenge," in *Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings*, pp. 53–58, Springer, 2022. DOI:https://doi.org/10.1007/978-3-030-98385-7_8.

[11] A. Golts, D. Khapun, D. Shats, Y. Shoshan, and F. Gilboa-Solomon, "An ensemble of 3d u-net based models for segmentation of kidney and masses in ct scans," in *Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings*, pp. 103–115, Springer, 2022. DOI:https://doi.org/10.1007/978-3-030-98385-7_14.

[12] Y. George, "A coarse-to-fine 3d u-net network for semantic segmentation of kidney ct scans," in *Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings*, pp. 137–142, Springer, 2022. DOI:https://doi.org/10.1007/978-3-030-98385-7_18.

[13] X. Yang, J. Zhang, J. Zhang, and Y. Xia, "Transfer learning for kits21 challenge," in *Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings*, pp. 158–163, Springer, 2022. DOI:https://doi.org/10.1007/978-3-030-98385-7_21.

[14] P. Sun, Z. Mo, F. Hu, X. Song, T. Mo, B. Yu, Y. Zhang, and Z. Chen, "Segmentation of kidney mass using agdenseu-net 2.5 d model," *Computers in Biology and Medicine*, vol. 150, p. 106223, 2022. DOI:https://doi.org/10.1016/j.compbiomed.2022.106223.

[15] M. Wu and Z. Liu, "Less is more: Contrast attention assisted u-net for kidney, tumor and cyst segmentations," in *Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings*, pp. 46–52, Springer, 2022. DOI:https://doi.org/10.1007/978-3-030-98385-7_7.

[16] K.-H. Uhm, H. Cho, Z. Xu, S. Lim, S.-W. Jung, S.-H. Hong, and S.-J. Ko, "Exploring 3d u-net training configurations and post-processing strategies for the miccai 2023 kidney and tumor segmentation challenge," *arXiv preprint arXiv:2312.05528*, 2023. DOI:https://doi.org/10.48550/arXiv.2312.05528.

[17] A. Myronenko, D. Yang, Y. He, and D. Xu, "Automated 3d segmentation of kidneys and tumors in miccai kits 2023 challenge," *arXiv preprint arXiv:2310.04110*, 2023. DOI:https://doi.org/10.48550/arXiv.2310.04110.

[18] G. Stoica, M. Breaban, and V. Barbu, "Analyzing domain shift when using additional data for the miccai kits23 challenge," *arXiv preprint arXiv:2309.02001*, 2023. DOI:https://doi.org/10.48550/arXiv.2309.02001.

[19] S. Zheng, Q. Sun, X. Ye, W. Li, L. Yu, and C. Yang, "Multi-scale adversarial learning with difficult region supervision learning models for primary tumor segmentation," *Physics in Medicine & Biology*, vol. 69, no. 8, p. 085009, 2024. DOI:https://doi.org/10.1088/1361-6560/ad3321.

[20] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, *et al.*, "The 2021 kidney and kidney tumor segmentation challenge." Available online: https://kits21.kits-challenge.org/ (accessed on 06 January 2024), 2021.

[21] S. Samuel, M. Moore, B. Patterson, H. Sheridan, and C. Sorensen, "Neuroimaging data primer: A resource for curating digital imaging and communications in medicine (dicom) and neuroimaging informatics technology initiative (nifti) files," 2020. DOI:https://doi.org/10.7302/7995.

[22] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987. DOI:https://doi.org/10.1016/S0734-189X(87)80186-X.

[23] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018. DOI:https://doi.org/10.48550/arXiv.1802.02611.

[24] M. Bland, *An Introduction to Medical Statistics*. OUP Oxford, 2015. ISBN: 9780191002991.

[25] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, pp. 1–28, 2015. DOI:https://doi.org/10.1186/s12880-015-0068-x.

[26] T. Eelbode, J. Bertels, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3679–3690, 2020. DOI:https://doi.org/10.1109/TMI.2020.3002417.

[27] C. E. F. Matos, M. V. S. L. Oliveira, J. O. B. Diniz, A. G. S. Fernandes, G. B. Junior, and A. C. de Paiva, "Ppm-deeplab: Módulo de pirâmide

de pooling como codificador da rede deeplabv3+ para segmentaçao de rins, cistos e tumores renais," in *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*, pp. 210–221, SBC, 2023. DOI:https://doi.org/10.5753/sbcas.2023.229611.

Caio Falcão is a PhD student at Federal University of Maranhão (UFMA) since 2020. He received his Master's Degree in Computer Science from Federal University of Maranhão (UFMA). He is currently a Professor of Basic, Technical, and Technological Education at the Federal Institute of Education, Science and Technology of Maranhão (IFMA), Barreirinhas Campus. He served as a Substitute Professor at the Federal University of Maranhão (UFMA) and as a researcher at the Applied Computing Center (NCA/UFMA), developing mainly research/extension projects focused on the following topics: image processing, pattern recognition, machine learning, and medical images.

Geraldo Braz Junior receive PhD in Electrical Engineering from Federal University of Maranhão. He is currently Associate Professor I at the Federal University of Maranhão a permanent member of the Postgraduate Programs of the Master's in Computer Science (PPGCC/UFMA) and the Doctorate in Computer Science Association UFMA-UFPI (DCCMAPI). Has experience in computer vision, machine learning, deep learning, and medical image processing.

João Almeida received a DSc. Degree in Electric Engineering from the Federal University of Maranhão (UFMA), Brazil, in 2013. He is currently Associate Professor I at the Federal University of Maranhão. He coordinates the VipLab-UFMA Vision and Image Processing Laboratory. He has experience in Computer Science, working mainly on the following topics: image processing, pattern recognition, machine learning, and ophthalmological medical images.

Anselmo Paiva received a DSc. Degree in Computing from Pontifical Catholic University of Rio de Janeiro. He is currently a Full Professor at the Federal University of Maranhão. Coordinator of the NCA-UFMA Applied Computing Center. He has experience in Computer Science, with an emphasis on Graphics Processing, working mainly on the following topics: Virtual and Augmented Reality, Computer Graphics, GIS, Medical Image Processing, and Volumetric Visualization.