Evaluation of Transfer Learning with a U-Net Architectures for Kidney Segmentation

Caio Eduardo Falcão Matos[1(✉)], João Guilherme Araújo do Vale[3], Marcos Melo Ferreira[1], Geraldo Braz Júnior[2], and João Dallyson Sousa de Almeida[2]

[1] Federal Institute of Education, Science and Technology of Maranhão, São Luís, Brazil
caio.matos@ifma.edu.br, marcos.ferreira@discente.ufma.br

[2] Applied Computer Group, Federal University of Maranhão, São Luís, Brazil
{geraldo,jdallyson}@nca.ufma.br

[3] Federal University of Maranhão, São Luís, Brazil
jga.vale@discente.ufma.br

Abstract. Kidney cancer emerges as one of the primary causes of mortality due to neoplasms on a global scale. Early detection and diagnosis of this disease often allow for more treatment options, contributing to the reduction of death rates. In this way, a correct delimitation of kidneys and renal tumor areas provides better analysis and diagnosis of suspicious lesions, contributing to treatment planning. This task is usually performed manually, making the process susceptible to fatigue (physical and visual) and distraction. Therefore, computational techniques, such as deep neural networks, are presented with great prominence as alternatives to improve segmentation precision and contribute to the early diagnosis of kidney cancer. In this work, we propose a methodology for kidney segmentation in computed tomography images by transfer learning to the U-Net network architecture. The KiTS19 dataset was used to evaluate the proposed methodology and obtained the best result for kidney segmentation of 96.0% of average Dice coefficient and average Jaccard index of 94.4%, using a pre-trained EfficentNet as an encoder for a U-Net.

Keywords: Kidney Segmentation · EfficentNet · U-Net · Transfer Learning

The authors thank the Higher Education Personnel Improvement Coordination (CAPES) - Finance Code 001, National Council for Scientific and Technological Development (CNPq) and the Foundation for Research and Scientific and Technological Development of the State of Maranhão (FAPEMA) for the financial support for the development of this work.

A. Cunha et al. (Eds.): MobiHealth 2023, LNICST 578, pp. 110–121, 2024.
https://doi.org/10.1007/978-3-031-60665-6_8

1 Introduction

Cancer is an ailment characterized by the uncontrolled growth and spread of certain cells within the body. It has the potential to initiate in nearly any part of the human body, comprised of trillions of cells [25,26]. According to the World Health Organization (WHO) report, cancer is the second leading cause of death worldwide, accounting for about 9.6 million deaths in 2018 alone [22].

Kidney cancer is also called renal cancer, is among the 14 most common types of cancer in the world, with over 400,000 diagnoses and over 170,000 deaths in 2020 alone [14]. An estimated 79,000 new cases of kidney cancer will be diagnosed this year 2022 in the United States, where 13 thousand of these will result in deaths [1].

Individuals identified in the early stages of the disease, where there is no spread, exhibit approximately a 93% relative survival rate of around 93% over a five-year span. Conversely, if the diagnosis occurs at an advanced stage with disease metastasis, the rate drops to 12%, underscoring the significance of early detection [2]. In this context, the early diagnosis assumes a pivotal role as an indispensable tool for prognosis and treatment, significantly augmenting the prospects of curative outcomes for the patient.

In recent years, medicine in general has made great strides in prevention, detection and treatment. Imaging tests, such as computed tomography (CT), represent a non-invasive alternative to obtain additional information to assist the physician. In this way, advances in computer-aided detection and diagnosis (CAD) methodologies have contributed to the early detection of kidney cancer, indicating suspicious areas and accurately diagnosing abnormalities. Renal area segmentation is an essential step for computer-assisted diagnosis or urological treatment. Therefore, several approaches based on deep learning have been widely applied in this task. Convolutional neural networks, especially the U-Net [24] networks and their various variations, have achieved excellent results in the semantic segmentation task focused on medical images [16,21,27,33].

This work aims to explore the transfer learning approach using a pre-trained EfficientNet to improve the coding step of a U-Net architecture and segment the kidney regions in a CT image. Furthermore, the contributions of this work include (1) the adaptation and integration of architectural models based on CNNs (Inception, ResNet, VGG, MobileNet, and EfficientNet) as the encoders for the U-Net architecture, (2) the evaluation of previous models with transfer learning for renal segmentation in computed tomography images.

2 Related Work

The task of segmenting kidneys and kidney tumors on computed tomography images aims to assist specialists in diagnosing kidney cancer. For both tasks, computational techniques can be applied to reduce the risks of radiation exposure (radiological examinations) and invasive procedures such as renal biopsy. Thus, the application and development of techniques aimed at image processing and deep learning are investigated by several works to analyze and segment kidneys and suspicious tumor regions.

Mu et al. [19] introduced a multi-resolution strategy combined with V-Net networks to automatic segment kidneys and renal tumors in CT images, training two V-Nets using images with different resolutions to determine the region of interest (kidneys and tumors).

Approaches that apply successive semantic segmentation networks to locate objects in images are called cascade architecture. The works of Zhang et al. [32] and Hou et al. [10] develop this approach for targeting kidneys and renal tumors. In the first one presents a two-stage approach utilizing a 3D UNet model. In the first stage, a rudimentary model is employed to identify the kidney's location and crop sub-volume patches encompassing the entire organ. Subsequently, the second stage involves a multitask 3D UNet model, which concurrently segments both the kidney and tumors based on the patches obtained from the initial stage. Conversely, the second work proposes a three-stage semantic segmentation pipeline based on 3D U-Nets, where the first two projects are based on nnUNet and are responsible for segmenting kidneys and the third stage for segmenting kidney tumors.

Isensee and Maier-Hein [15] propose an approach that integrates three networks, namely Plain 3D U-Net, Residual 3D U-Net, and Pre-activation Residual 3D U-Net, based on the U-Net architecture for the segmentation of kidneys and tumors. Türk, Lüy and Barışçı [30] demonstrate a hybrid model using the superior features of existing V-Nets models is presented. This model presents modifications in the encoding and decoding phases in the standard architecture.

Cruz et al. [6] presents um método baseado em três estágios, Scope Reduction, Segmentação dos rins e redução de falsos positivos. A primeira utiliza a rede AlexNet para efetuar slices classification. The second utiliza a rede U-Net para segmentação dos rins. At last, reduction of false positives by removing small segmented fragments.

Adaptations aimed at the U-Net network were applied using the methods proposed by [28] and [5] for kidney segmentation in the Kits19 dataset. The first method proposes the 2.5D MFFAU-Net (Multi-level Feature Fusion Attention U-Net), which uses a 2.5D model to learn to combine representations of 2D slices applied in a ResConv architecture in MFFAU-Net for kidney segmentation. The other method applies a deep learning model, which is an ensemble of U-Net models developed after testing various model variations.

A modification of the convolution blocks of the U-Net network architecture is proposed by Hou et al. [11]. According to this work, dilated convolution blocks (DCB) are applied to replace the traditional pooling operations of the U-Net network. Such changes aim to retain semantic information better. This approach additionally employs three networks (reduced resolution, complete resolution, and refinement) for the segmentation of renal structures and tumors.

When reviewing the relevant literature, it becomes evident that variations in the architecture of the convolutional neural network U-Net [24] have been consistently employed across all proposed methodologies. This underscores the inherent potential of encoder-decoder-based convolutional neural networks. Consequently, the approach proposed in this study aims to harness the full capabilities of the U-Net convolutional network by leveraging a diverse set of pretrained architectures as encoders for the task of segmenting kidneys in computed tomography images.

3 Materials and Methods

This section presents the proposed methodology for segmenting the kidneys in computed tomography images. For the development of the experiments, CT images were used from the KiTS19 database [9]. Then, we applied a preprocessing step for each volume to improve the contrast of the organs. In the next step, we applied pre-trained CNN architectures as an encoder of the U-Net to evaluate how the transfer of learning can positively influence the segmentation of the kidneys in CT images. We use a U-Net network because it could be easily customizable, changing the encode (and adapting the decode) and improving the skip connections. We hypothesize that transfer learning techniques could generate more reliable and precise results. Figure 1 illustrates the steps of the proposed methodology. In the following subsections, each of these steps is detailed.

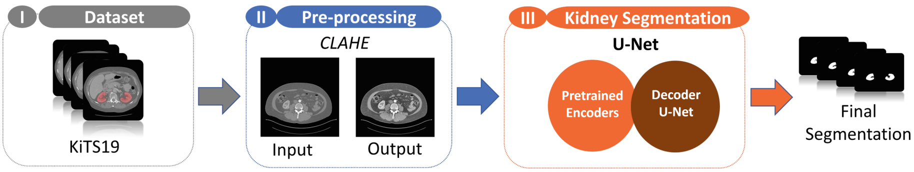

Fig. 1. Steps of the proposed method.

3.1 Dataset

In this work, we used a database with CT images aimed at the challenge of segmenting kidney tumors called KiTS19.

This dataset comes from patients undergoing nephrectomy for renal tumors at the University of Minnesota Medical Center (USA) between 2010 and 2018. A total of 300 exams were collected in this survey. Of which, 210 were made available for training and validation and 90 to test according to the challenge. In addition to imaging exams, a manual segmentation, performed by medical students under the supervision of Dr. Christopher Weight, is made available representing the kidneys and tumors for each patient [9]. CT images and masks (Ground Truth), available in NIFTI format for each volume, are converted into slices, each image being grayscale with a resolution of 512×512, where the number of slices varies between patients (Fig. 2).

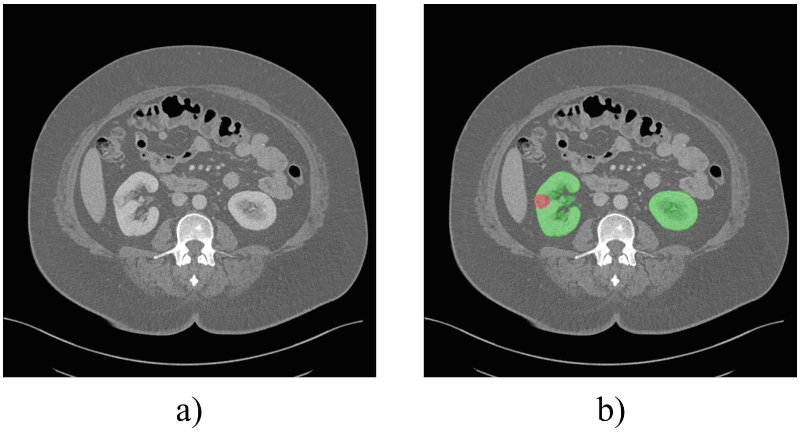

Fig. 2. Example from the KiTS19 dataset. (a) CT scan slice and (b) provided kidney and tumor annotations.

3.2 Image Pre-processing

The objective of this preprocessing stage is to augment the contrast of the objects of interest, namely, kidneys and tumors, in comparison to the other organs present in the examination. This enhancement aims to provide superior characteristics for the subsequent phases of the methodology.

In this step, we initially performed the resizing of the slices and ground truth to 256×256 pixels due to the limitations of the computational resources. The preprocessing technique known as CLAHE [23], which stands for Contrast Limited Adaptive Histogram Equalization, represents a variation of adaptive histogram equalization (AHE) aimed at mitigating excessive contrast amplification. This method functions within localized image regions referred to as "blocks" rather than operating on the entire image. In this approach, each pixel of the original image resides at the center of its respective contextual region. The process entails clipping the histogram of the original image and redistributing the pixels across the various gray levels. Then, the CLAHE was applied to the KiTS19 base images using a grid size of 8×8 and a clip contrast limit 2.0. The application of this technique aims to increase the contrast of the object of interest, the kidneys, in relation to the other organs present in the abdominal region. Figure 3 depicts the technique's application on the KiTS19 dataset.

3.3 Kidney Segmentation

CNN architectures focused on semantic segmentation typically incorporate encoding and decoding networks such as the U-Net network. This property aims to reduce and restore the image resolution to capture the most relevant details in their respective encoding and decoding phases. Therefore, we apply the

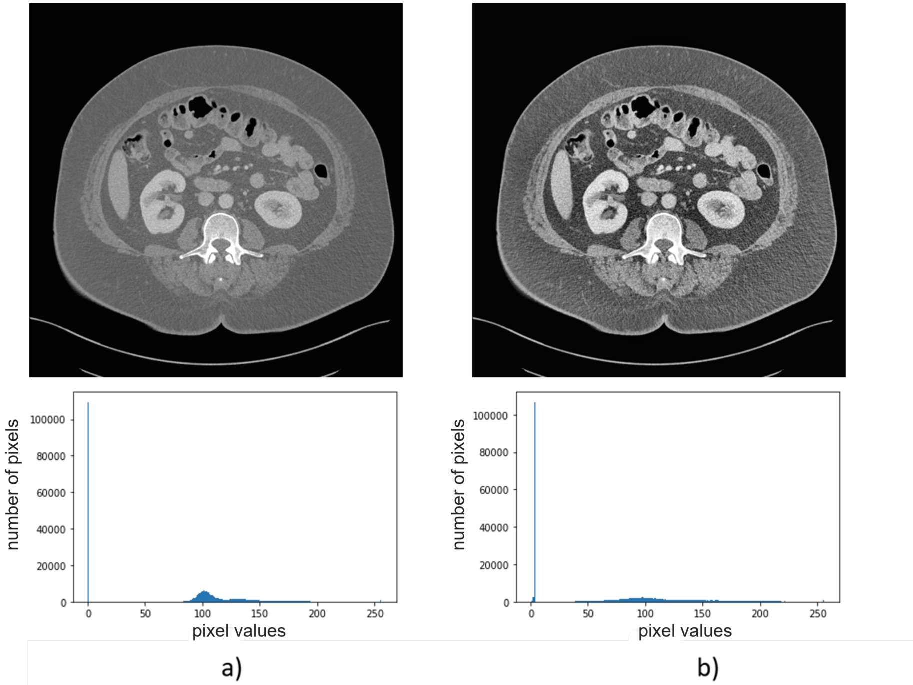

Fig. 3. Result of applying CLAHE with the respective histograms. (a) original image and (b) preprocessed image.

CNN-based architecture of the EfficientNet network as an alternative to replace the U-Net network's encoder for segmentation of the kidneys.

EfficientUNet Architecture. EfficientNet's architecture is based on convolutional neural networks and has a method called Compound Scaling that uniformly scales all dimensions of depth/width/resolution using a composite coefficient. The architecture of this network has 8 variations from B0 to B7, where each model number refers to variants with a more significant number of [29] parameters. The inverted residual block called MBConv developed in the MobileNetV2 [12] architecture became the basis for constructing EfficientNet networks, where each of the 8 models (B0–B7) share standard blocks with subtle complexities in their architectures. Thus, in this work, we use the EfficientNet-B3 architecture composed of 7 MBConv blocks, where each of the blocks presents the size of the applied filter (3×3 or 5×5) and the default activation function used, which can be, respectively, ReLU [20] and ReLU6 [13] for blocks MBConv1 and MBConv6. An overview of the architecture can be seen in Fig. 4.

Figure 5 presents the architecture evaluated in this work. This approach modifies the standard U-Net network by replacing the encoder phase with a CNN-based architecture with pre-trained weights. We adopted an EfficientNet-B3

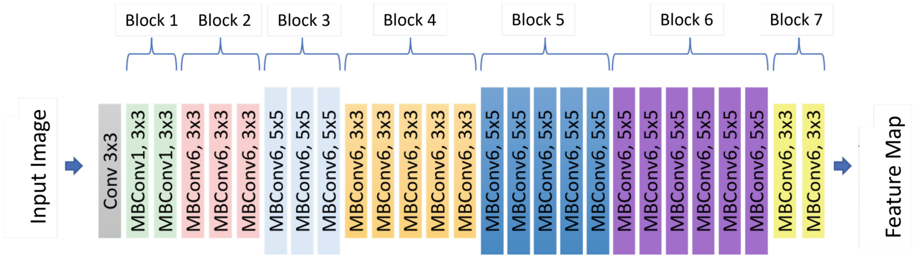

Fig. 4. Architecture of EfficientNet-B3 with MBConv as basic building blocks.

network, in its standard architecture of MBConv blocks, with pre-trained weights in the public database ImageNet [7] as the encoder of the UNet network for kidney segmentation. On the other hand, we use the standard structure of the U-Net [24] architecture for the decoder phase. This phase enables the precise location of elements in the image as it combines high-level features and spatial information by a sequence of up-convolution and concatenation with feature maps through skip connections operations.

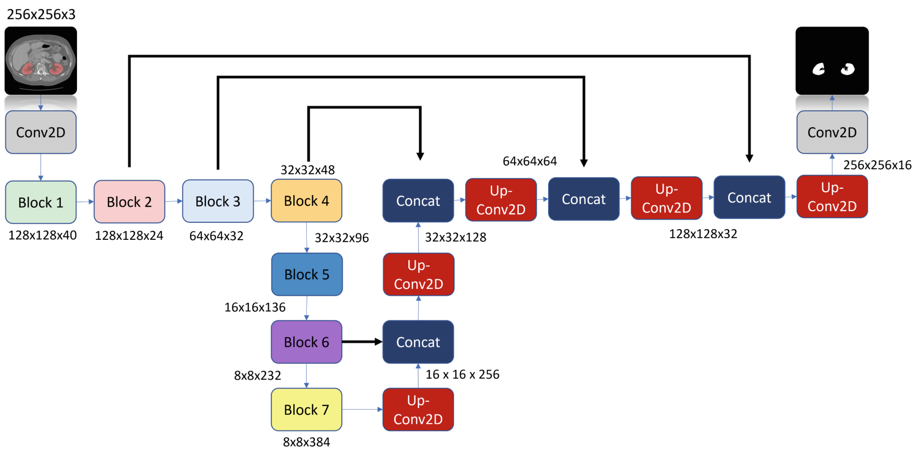

Fig. 5. Proposed architecture of EfficientNet-B3 as encoder in U-Net.

4 Results

In this section, we present the results of the experiments carried out in evaluating the proposed methodology. The utilized metrics encompassed the Sørensen-Dice index, commonly referred to as the Dice coefficient (DSC) [8] and Jaccard index (JCD) [17].

We divided the 210 exams provided by the KiTS19 dataset into three datasets respecting the proportions 70% (147 exams) for training, 20% (42 exams) for validation, and 10% (21 exams) for the test. A random separation of the data sets was conducted based on patients (exams) to ensure the integrity of the slices in the training, testing, and validation sets.

To study the effect of different encoders, beyond EfficientNet, we perform experiments with VGG16, ResNet50, InceptionV3, SE-Resnet-152, MobileNetv2 and EfficientNetB3. The architecture formed by each of the encoders combined with the UNet decoder was trained for 50 epochs, using a *batch* size equal to 16, Adam optimizer with a learning rate of 0.0001 and DiceLoss + BinaryFocalLoss as a loss function. This set of parameters was defined after several experiments, and this set produced the best results. Table 1 shows that the EfficientNet-B3, when used as an encoder for a U-Net obtained the best results among other tested encoders.

Table 1. Results of the combination of pre-trained models with U-Net for kidney segmentation.

Method	**DSC**	**JCD**	**Error**
U-Net + InceptionV3	89.19%	85.82%	0.1209
U-Net + Resnet-50	91.32%	89.15%	0.0965
U-Net + SE-Resnet-152	91.33%	88.95%	0.0954
U-Net + VGG-16	91.76%	89.59%	0.0906
U-Net + MobileNetV2	93.93%	92.20%	0.0658
U-Net + EfficientNetB3	**96.04%**	**94.47%**	**0.0445**

Table 1 shows that the MobileNet and EfficientNet networks outperformed the other applied networks. The results of 89.1% Dice and 85.8% Jaccard obtained by the Inception network demonstrate limitations in kidney identification and segmentation, as it performed the worst among the networks. The ResNet, SE-ResNet, and VGG networks achieved relatively close results in terms of Dice, Jaccard, and error rates. Among these three networks, VGG achieved the highest scores with Dice, Jaccard, and Error rates of 91.7%, 89.5%, and 0.09, respectively.

However, EfficientNet and MobileNet networks focus on the foreground and produce better results for kidney segmentation, as reflected in their Dice, Jaccard, and Error values. When comparing these two networks, it is observed that EfficientNet outperformed MobileNet in all metrics. Considering only the Dice coefficient, the EfficientNet network scored 2% points higher than MobileNet, reaching 93.9%.

Figure 6 provide examples of kidney segmentation carried out by the MobileNet and EfficientNet networks, respectively. It is observed that both networks exhibit good performance in distinguishing the kidneys from other abdominal organs and structures.

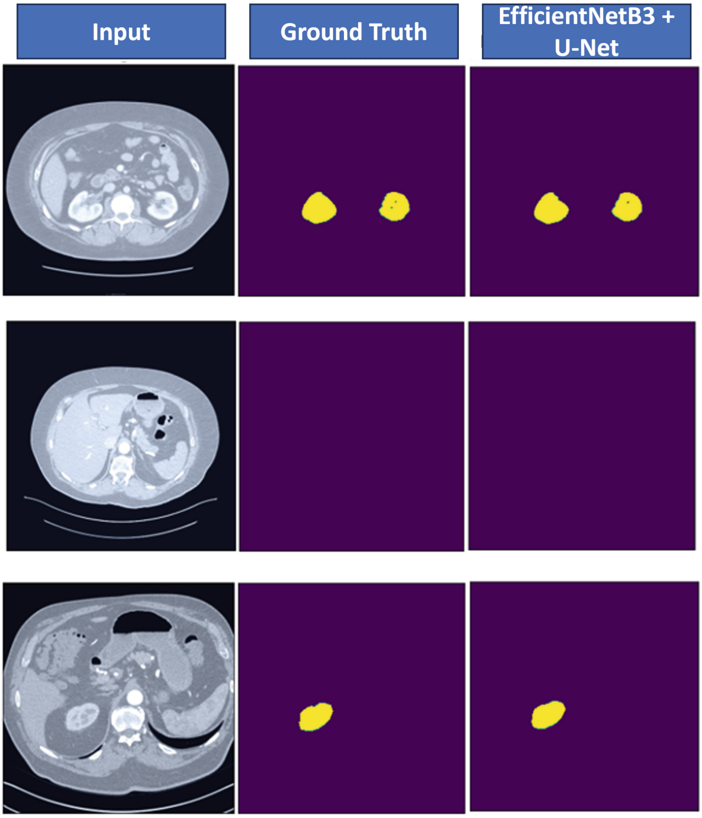

Fig. 6. Examples of kidney segmentation using the EfficientNetB3 + U-Net network. From left to right: input slice, expert annotation (ground truth), and segmentation performed by the network.

In Table 2, we compare the performance of the proposed methodology with the works that represent the state of the art in the KiTS19 database. It is important to note that the methods cited do not provide information about the exact procedures for splitting the training, testing, and validation datasets, nor do they specify how slices are divided within the datasets. As a result, it is not possible to make an exact comparison of the results.

The EfficientNet-B3 has proven to be the best encoder when used in conjunction with the U-Net network. This combination yields results comparable to the ensemble methods employed by most of the studies, as well as the hybrid approach of the V-Net network, which is considered the benchmark for the best-performing method in kidney segmentation. The proposed methodology achieved Dice and Jaccard scores of 96% and 94.4%, respectively, with only a 1.7% point difference from the method proposed by [30]. Furthermore, it offers significantly fewer hyperparameter choices and a customizable architecture for supplementary tasks.

Table 2. Comparison of the proposed method with works aimed at kidney segmentation using the KiTS19 dataset.

Work	Technique	DSC	JCD
Ma (2019) [18]	Ensemble Networks	97.3%	-
Zhang et al. (2019) [32]	Cascade	97.4%	-
Mu et al. (2019) [19]	MR-VB-Net	97.4%	-
Hou et al. (2019) [10]	Cascade	96.7%	-
Isensee et al. (2019) [15]	Ensemble Deep Net	97.4%	-
Trk et al. (2020) [30]	Hybrid V-Net	**97.7%**	-
Cruz et al. (2020) [6]	U-Net 2D	96.3%	93.0%
Hou et al. (2020) [11]	U-Net Modified	96.7%	-
Xie et al. (2020) [31]	SE-ResNeXT	96.7%	-
Jason et al. (2021) [5]	Ensemble of U-Net	94.9%	-
Peng et al. (2023) [28]	2.5D MFFAU-Net	92.4%	-
Proposed Method	**EfficientNetB3 + U-Net**	**96.0%**	**94.4%**

5 Conclusion

The development of methodologies and architectures for semantic segmentation aimed at identifying organs and tumor lesions in computed tomography scans presents itself as a challenging task, especially in abdominal CT scans, which encompass a wide variety of organs such as kidneys, spleen, bladder, among others. The methodology proposed in this study sought to evaluate the application of some of the leading fully convolutional networks (FCNs), particularly the EfficientNet network, as an encoder and feature extractor for the U-Net network, applied to the segmentation of kidneys and renal tumors in CT scans.

The methodology proposed in this work sought to explore and evaluate the learning transfer of CNN-based architectures using the EfficientNet network as an encoder and feature extractor of the U-Net network for segmenting kidneys in CT scans.

The results obtained for the kidney segmentation task demonstrate that the EfficientNet network applied as an encoder of the U-Net network produces an efficient approach for identifying kidneys in CT images. In addition, the proposed methodology obtained results comparable to the literature for kidney segmentation. Thus, the results for this task indicate that architectures with different encoders and decoders emerge as a research area where they allow the construction of methods with the integration of different networks aimed at kidney segmentation. We also observe that the application of the standard U-Net decoder can reconstruct features at each stage of both backbones, thus obtaining information oriented towards both local context and global information.

Despite promising results, the proposed method could be improved in some ways. As future works, the adaptation of the pre-trained EfficientNet-B3 model applied to variations of the U-Net network such as Attention U-Net [21], U-Net++ [33], BCDU-Net [4], LSTM-UNet [3] could contribute to a better identification of the kidneys in CT scans.

References

1. ACS: American cancer society - key statistics about kidney cancer (2022). https://www.cancer.org/cancer/kidney-cancer/about/key-statistics.html
2. ACS: American cancer society - survival rates for kidney cancer (2022). https://www.cancer.org/cancer/kidney-cancer/detection-diagnosis-staging/survival-rates.html
3. Arbelle, A., Raviv, T.R.: Microscopy cell segmentation via convolutional LSTM networks. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1008–1012. IEEE, Hilton Molino Stucky, Venice (2019)
4. Azad, R., Asadi-Aghbolaghi, M., Fathy, M., Escalera, S.: Bi-directional ConvLSTM U-net with densley connected convolutions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. IEEE, Seoul (2019)
5. Causey, J., et al.: An ensemble of U-net models for kidney tumor segmentation with CT images. IEEE/ACM Trans. Comput. Biol. Bioinf. **19**(3), 1387–1392 (2021)
6. da Cruz, L.B., et al.: Kidney segmentation from computed tomography images using deep neural network. Comput. Biol. Med. **123**, 103906 (2020)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE, Miami (2009)
8. Fleiss, J.L., Levin, B., Paik, M.C., et al.: The measurement of interrater agreement. Stat. Methods Rates Proport. **2**(212–236), 22–23 (1981)
9. Heller, N., et al.: The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019)
10. Hou, X., Xie, C., Li, F., Nan, Y.: Cascaded semantic segmentation for kidney and tumor. Submissions to the (2019)
11. Hou, X., et al.: A triple-stage self-guided network for kidney tumor segmentation. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 341–344. IEEE, Iowa City (2020)
12. Howard, A., Zhmoginov, A., Chen, L.C., Sandler, M., Zhu, M.: Inverted residuals and linear bottlenecks: mobile networks for classification, detection and segmentation (2018)
13. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
14. Internacional, W.: World cancer research fund international - kidney cancer statistics (2022). https://www.wcrf.org/cancer-trends/kidney-cancer-statistics/
15. Isensee, F., Maier-Hein, K.H.: An attempt at beating the 3D U-net. arXiv preprint arXiv:1908.02182 (2019)
16. Isensee, F., et al.: nnU-Net: self-adapting framework for U-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
17. Jaccard, P.: The distribution of the flora in the alpine zone. 1. New Phytol. **11**(2), 37–50 (1912)
18. Ma, J.: Solution to the kidney tumor segmentation challenge 2019 (2019)

19. Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y.: Segmentation of kidney tumor by multi-resolution VB-nets (2019)
20. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: ICML. Omnipress, Madison (2010)
21. Oktay, O., et al.: Attention U-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
22. OPAS/OMS: Pan-American health organization - cancer. 2022 (2022). https://www.paho.org/pt/topicos/cancer
23. Pizer, S.M., et al.: Adaptive histogram equalization and its variations. Comput. Vision Graph. Image Process. **39**(3), 355–368 (1987)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
25. Roy, N.K., Bordoloi, D., Monisha, J., Anip, A., Padmavathi, G., Kunnumakkara, A.B.: Cancer-an overview and molecular alterations in cancer. Fusion Genes Cancer 1–15 (2017)
26. Sprouffske, K., Athena Aktipis, C., Radich, J.P., Carroll, M., Nedelcu, A.M., Maley, C.C.: An evolutionary explanation for the presence of cancer nonstem cells in neoplasms. Evol. Appl. **6**(1), 92–101 (2013)
27. Sun, J., Darbehani, F., Zaidi, M., Wang, B.: SAUNet: shape attentive U-net for interpretable medical image segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 797–806. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_77
28. Sun, P., et al.: 2.5 d MFFAU-Net: a convolutional neural network for kidney segmentation. BMC Med. Inform. Decis. Making **23**(1), 1–11 (2023)
29. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR, Long Beach (2019)
30. Türk, F., Lüy, M., Barışçı, N.: Kidney and renal tumor segmentation using a hybrid V-Net-based model. Mathematics **8**(10), 1772 (2020)
31. Xie, X., Li, L., Lian, S., Chen, S., Luo, Z.: SERU: a cascaded se-ResNeXT U-Net for kidney and tumor segmentation. Concurr. Comput.: Pract. Exp. **32**(14), e5738 (2020)
32. Zhang, Y., et al.: Cascaded volumetric convolutional network for kidney tumor segmentation from CT volumes. arXiv preprint arXiv:1910.02235 (2019)
33. Zhou, Zongwei, Rahman Siddiquee, Md Mahfuzur, Tajbakhsh, Nima, Liang, Jianming: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1