

Article

Deep Learning-Based Methods for Prostate Segmentation in Magnetic Resonance Imaging

Albert Comelli ^{1,2} , Navdeep Dahiya ³ , Alessandro Stefano ^{2,*} , Federica Vernuccio ⁴, Marzia Portoghese ⁴, Giuseppe Cutaia ⁴, Alberto Bruno ⁴, Giuseppe Salvaggio ⁴  and Anthony Yezzi ³

¹ Ri.MED Foundation, Via Bandiera, 11, 90133 Palermo, Italy; acomelli@fondazionerimed.com

² Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), 90015 Cefalù, Italy

³ Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA; ndahiya3@gatech.edu (N.D.); anthony.yezzi@ece.gatech.edu (A.Y.)

⁴ Dipartimento di Biomedicina, Neuroscienze e Diagnostica avanzata (BIND), University of Palermo, 90127 Palermo, Italy; federica.vernuccio@unipa.it (F.V.); marzia.portoghese@unipa.it (M.P.); giuseppe.cutaia@community.unipa.it (G.C.); bruno-alberto@hotmail.it (A.B.); p.salvaggio@libero.it (G.S.)

* Correspondence: alessandro.stefano@ibfm.cnr.it; Tel.: +390921920149

Featured Application: The study demonstrates that high-speed deep learning networks could perform accurate prostate delineation facilitating the adoption of novel imaging parameters, through radiomics analyses, for prostatic oncologic diseases.

Abstract: Magnetic Resonance Imaging-based prostate segmentation is an essential task for adaptive radiotherapy and for radiomics studies whose purpose is to identify associations between imaging features and patient outcomes. Because manual delineation is a time-consuming task, we present three deep-learning (DL) approaches, namely UNet, efficient neural network (ENet), and efficient residual factorized convNet (ERFNet), whose aim is to tackle the fully-automated, real-time, and 3D delineation process of the prostate gland on T2-weighted MRI. While UNet is used in many biomedical image delineation applications, ENet and ERFNet are mainly applied in self-driving cars to compensate for limited hardware availability while still achieving accurate segmentation. We apply these models to a limited set of 85 manual prostate segmentations using the k-fold validation strategy and the Tversky loss function and we compare their results. We find that ENet and UNet are more accurate than ERFNet, with ENet much faster than UNet. Specifically, ENet obtains a dice similarity coefficient of 90.89% and a segmentation time of about 6 s using central processing unit (CPU) hardware to simulate real clinical conditions where graphics processing unit (GPU) is not always available. In conclusion, ENet could be efficiently applied for prostate delineation even in small image training datasets with potential benefit for patient management personalization.

Keywords: deep learning; segmentation; prostate; MRI; ENet; UNet; ERFNet; radiomics



Citation: Comelli, A.; Dahiya, N.; Stefano, A.; Vernuccio, F.; Portoghese, M.; Cutaia, G.; Bruno, A.; Salvaggio, G.; Yezzi, A. Deep Learning-Based Methods for Prostate Segmentation in Magnetic Resonance Imaging. *Appl. Sci.* **2021**, *11*, 782. <https://doi.org/10.3390/app11020782>

Received: 17 December 2020

Accepted: 12 January 2021

Published: 15 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the biomedical imaging field, target delineation is routinely used as the first step in any automatized disease diagnosis system (i.e., radiotherapy system) and, in the last few years, in radiomics studies [1,2] to obtain a multitude of quantitative parameters from biomedical images [3,4]. These parameters are then used as imaging biomarkers to identify any possible associations with patient outcome. The first task of a radiomics analysis is the automatic and user-independent target (e.g., tumor or organ) delineation to avoid any distortion during the feature extraction process [5]. Manual segmentation might seem like the simplest solution to obtain target boundaries, but it is a time-consuming and user-dependent process that affects the radiomics signature [6]. For this reason, an automatic and operator-independent target delineation method is mandatory. Nevertheless, the segmentation process remains a challenging field of research. Over the years many different

types of segmentation techniques have been developed, for example, [7–9]. Some of the previous techniques include thresholding [10], k-means clustering [11], watersheds [12], followed by more advanced algorithms such as active contour methods [8,13], graph cuts [14], random walks [15], conditional and Markov random fields [16] to name a few. In recent years, particularly the last decade, the field of Machine Learning (ML) and Deep Learning (DL) has seen exponential growth and has produced models that have shown remarkable performance across many benchmark datasets and many different problem domains [17,18]. In general, an artificial intelligence method learns from examples and makes predictions without prior specific programming [19]. In the case of DL, these models implement networked structures to mimic the human brain transforming imaging data in feature vectors. Briefly, between the input and output, a variable number of hidden layers is implemented and the various nodes are connected to others with different weights.

The initial development of DL models was towards image classification problems, followed by object detection and finally, image segmentation, which is seen as a pixel level classification problem where each pixel is classified with one of many possible label classes. For example, in tumor segmentation, every voxel can be classified as either belonging to the class label of the object of interest (target) or the background. Since it is a very common task across many different problem domains, hundreds of different DL based models have been presented for the delineation task over the past several years: fully convolutional [20], encoder-decoder [21], multi-scale and pyramid [22–24], attention [25], recurrent neural [26], generative and adversarial training [27,28] based networks. Even during the current pandemic, DL networks have been widely used to help clinicians diagnose COVID-19 [29,30]. It is beyond the scope of this paper to discuss and describe all these different types of models. Interested readers are directed to recent comprehensive reviews [31,32] of DL based methods/models for image segmentation.

In this study, we deal with the issue of prostate region delineation on magnetic resonance imaging (MRI) studies. Prostatic volume extraction helps in the planning of biopsies, surgeries, focal ablative, radiation, and minimally invasive (e.g., intensity focused ultrasound [33]) treatments. In addition, benign prostatic hyperplasia, also called prostate enlargement, is one of the most common conditions affecting men [34]. A correlation between prostatic volume, and the incidence of prostate cancer, where early tumor identification is crucial to reduce mortality, has been shown in [35]. Since only part of prostate cancer is clinically significant, risk stratification is mandatory to avoid over-diagnosis and over-treatment [36,37]. For this reason, radiomics in MRI has acquired a crucial role in the risk stratification process [19,36]. MRI allows calculating prostatic volume considering the prostate as an ellipsoid. Unfortunately, the shape of the prostate varies and the determination of its volume based on the ellipsoid formula is often incorrect [38]. The presence of prostate cancer may alter the prostate volume as reported, for example, in the study of [39]: the authors reported that shape differences in the prostate gland were consistently observed between patients with or without prostate cancer maybe as the result of cancer localized in the peripheral zone. For this reason, the manual delineation is more accurate than the previously described method but takes time, requires experience, and is highly operator-dependent as noted above. Consequently, several automatic algorithms have been proposed, for example, [40–42]. Due to the lack of large amounts of labeled data for the training process, DL is still far from a widespread application in the biomedical environment. So, there is a need to develop DL networks to obtain accurate delineations with fewer training examples. Then, we explore the efficacy of Efficient Neural Network (ENet) [43] and Efficient Residual Factorized ConvNet (ERFNet) [44] that are mainly applied in self-driving cars to compensate for limited hardware availability while still achieving accurate segmentation, and UNet that is used in many biomedical image delineation applications [45]. Using a limited set of 85 manual prostate segmentation training data, we show that ENet model can be used to obtain accurate, fast and clinically acceptable prostate segmentations.

2. Materials and Methods

2.1. Experimental Setup

To test DL based methods for prostate segmentation, we used prostate studies of patients who underwent MRI examinations using the Achieva scanner (Philips Healthcare, Best, The Netherlands) with a pelvic phased-array coil (8 channel HD Torso XL). Specifically, from September 2019 to May 2020, 202 consecutive patients were referred to our Radiology Department to perform a prostate MRI examination. We excluded patients from the study for (a) incomplete MRI examination due to intolerance, discomfort, or claustrophobia ($n = 11$); (b) patients with radical prostatectomy ($n = 18$), subjected to transurethral resection of the prostate (TURP) ($n = 20$), or radiotherapy ($n = 17$); (c) lack of median lobe enlargement defined as intra-vesical prostatic protrusion characterized by overgrowth of the prostatic median lobe into the bladder for at least 1 cm ($n = 51$). So, our final study population consisted of 85 patients (age range 43–75 years, mean age 59 ± 8.4 years) with median lobe enlargement. By reviewing radiological reports, no pathological MRI findings were found in 35 patients (except for median lobe enlargement), while 50 prostate lesions (42 in peripheral zone and 8 in transitional zone) suspected for prostate cancer classified using the Prostate imaging reporting and data system (PI-RADS) 2.1 [46] were found: 18 PI-RADS 3 score, 28 PI-RADS 4 score, and 4 PI-RADS 5 score lesions with size ranged between 0.6 and 1.9 cm (mean 1.052 ± 0.28). In addition, in our study population, by evaluating capsular involvement, 18 patients had capsular abutment and 3 patients had capsular irregularity. It means that the presence of suspected prostate cancer lesions, in our study population, can at least distend the gland boundaries. Consequently, the determination of prostate volume using the above mentioned ellipsoid formula [38] is not suitable, while manual and automatic segmentations are not (or less) affected by this issue.

In this study, axial T2-weighted images with parameters shown in Table 1 were used. However, due to MRI protocol routine update during the study time, datasets had different resolution (2 studies with a matrix resolution of 720×720 ; 45 studies with a matrix resolution of 672×672 ; 23 studies with a matrix resolution of 576×576 ; 15 studies with a matrix resolution of 320×320). Consequently, the datasets had different resolutions and sizes. Since DL networks require inputs of the same size for the training process, MRI images were resampled to the isotropic voxel size of $1 \times 1 \times 1 \text{ mm}^3$ with a matrix resolution of 512×512 (matrix resolution in the middle between 720 and 320) using linear interpolation. A set of trained clinical experts (FV, MP, GC, and GS authors) hand segmented the prostate region. The simultaneous ground truth estimation STAPLE tool [47] was used to combine the different segmentations from the clinical experts in a consolidated reference. Finally, manual delineations were resampled using nearest neighbor interpolation and converted to masks with 0 for the background and 1 for the prostate area.

Table 1. Parameter of MRI protocol.

Parameter	Repetition Time (ms)	Echo Time (ms)	Flip Angle (Degrees)	Signal Averages	Signal-to-Noise Ratio
T2w TSE	3091	100	90	3	1

2.2. Deep Learning Models

Three different deep learning models including UNet [45], ENet [43], and ERFNet [44] were investigated to account for accurate prostate segmentation, fast training time, low hardware requirements for inference, and low training data requirements. Specifically, UNet was modified to improve segmentation accuracy, as reported in [48,49]. Briefly, (i) 3×3 convolutions were replaced by 5×5 convolutions, (ii) zero padding was used to ensure that the size of the output feature maps was the same as the input size, and (iii) an input size of 512×512 with 32 filters was used on the first contraction path layer, with doubling of feature maps after each max pool and stopping at 256 feature maps and 2D size of 64×64 .

Concerning ENet and ERFNet (see Table 1 in [43] and Table 2 in [44] for the description of their architecture), they were mainly applied in self-driving cars to compensate for limited hardware availability maintaining high accuracy and successfully used in two biomedical segmentation issues [48,49], that is, in the segmentation of high resolution computed tomography (HRCT) images characterized by a slice thickness much lower than that of the T2 weighted images of the prostate studies. This means that the number of slices of each patients' study was much greater than in this study.

Table 2. The model parameters and shape output after the first hidden layer in the ENet model for a given provided input image (Patient #7 slice #20).

Layer (Type)	Output Shape	Parameters Number
input_1 (InputLayer)	(None, 512, 512, 1)	0
conv2d_1 (Conv2D)	(None, 256, 256, 15)	150

2.3. Training Methodology

Due to a limited amount of data, the k-fold cross-validation strategy was applied by randomly dividing the dataset into k sub-datasets of equal size (17 patients, k = 5). For each network, we trained k models by combining k-1 folds into the training set and keeping the remaining fold as a holdout test set. Despite the fact that 2D models were considered, slices from the same study were never used for both training and testing purposes. So, there was no cross-contamination between training and test sets.

Moreover, the data augmentation technique was applied in six different modalities to increase the statistic. Additionally, data standardization and normalization were adopted to prevent the weights from becoming too large, to make the model converge faster, and to avoid numerical instability. Regarding loss function, prostate segmentation suffers from the imbalanced data problem because there are very few examples of the positive class compared to the background or negative class. In terms of image segmentation, the target (i.e., the prostatic region) is small compared to the background, which may be composed of many different organs or types of tissue exhibiting a wide range of intensity values. Some slices may have a very small target area compared to the background. This makes it hard for the DL to learn a reliable feature representation of the foreground class. In such cases, the networks tend to simply predict most voxels as belonging to the background class. To deal with this problem, various loss functions have been proposed over the years. These loss functions typically aim to solve the class imbalance problem by providing a larger weight to foreground voxels. This translates to a higher penalty in the loss function for foreground voxels that are misclassified by the network leading to the network being able to learn the foreground object representation more effectively. One such loss function which the authors of this paper have experimentally determined to be better suited for the biomedical image delineation process is the Tversky loss function [50]. Specifically, the Dice similarity coefficient (DSC) between P and G is defined as:

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \quad (1)$$

where P and G are the predicted and ground truth labels. DSC measures the overlap between P and G and is used as a loss function in many DL approaches. Nevertheless, DSC is the harmonic mean of false positives and false negatives and weighs both equally. To modify their weights, the Tversky index [51] was proposed as:

$$S(P, G; \alpha \beta) = \frac{|P \cap G|}{|P \cap G| + \alpha|P \setminus G| + \beta|G \setminus P|} \quad (2)$$

α and β control the penalty magnitude of false positives and false negatives. Using this index, the Tversky loss [50] is defined as:

$$T(\alpha \beta) = \frac{\sum_{i=1}^N p_{0i} g_{0i}}{\sum_{i=1}^N p_{0i} g_{0i} + \alpha \sum_{i=1}^N p_{0i} g_{1i} + \beta \sum_{i=1}^N p_{1i} g_{0i}} \tag{3}$$

Additional information about the study design is shown in Figure 1.

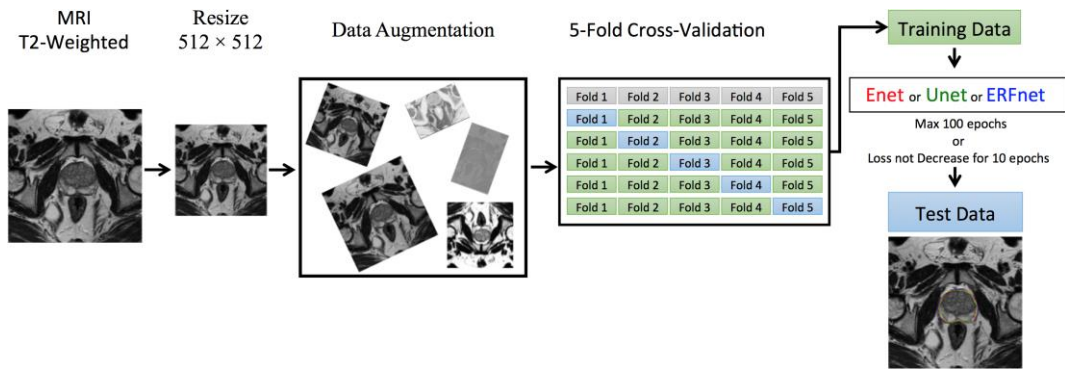


Figure 1. Workflow of the proposed segmentation method.

Starting from 16 patients, the best learning rates for each network were determined experimentally. We used a learning rate of 0.0001 for ENet, 0.00001 for ERFNet, and UNet with Adam optimizer [52]. A batch size of eight slices, α and β of 0.3 and 0.7, respectively for the Tversky loss function, were identified. All the models were allowed to train for a maximum of 100 epochs with an automatic stopping criteria of ending training when the loss did not decrease for 10 epochs continuously. The GEFORCE RTX 2080 Ti with 11GB of RAM (NVIDIA) was used to train DL models and run inference. Table 2 and Figure 2 show the feature representation learned from the first hidden layer in the ENet model.

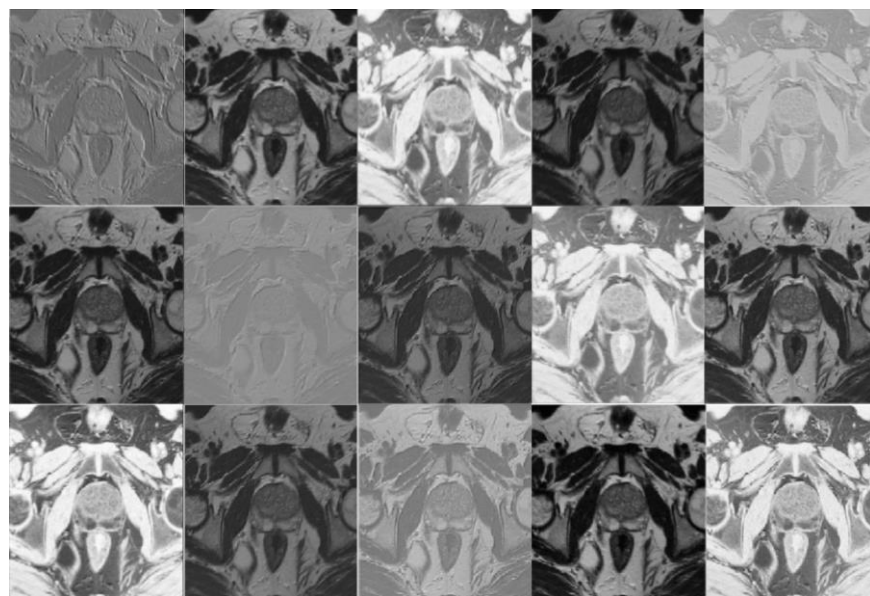


Figure 2. Feature maps (None, 256, 256, 15) extracted from the first hidden layer in the ENet Model for Patient #7 slice #20.

3. Results

Sensitivity, positive predictive value (PPV), DSC, volume overlap error (VOE), and volumetric difference (VD) were used for performance evaluation:

$$\text{Sensitivity} = \frac{TP}{TN + FN} \quad (4)$$

$$\text{PPV} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{DSC} = \frac{2TP}{2TP + FP + FN} \quad (6)$$

$$\text{VOE} = 1 - \frac{TP}{TP + FP + FN} \quad (7)$$

$$\text{DSC} = \frac{2TP}{2TP + FP + FN} \quad \text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \quad (8)$$

Table 3 shows the performance obtained using ENet, UNet, and ERFNet methods. In particular, ENet showed a mean DSC of $90.89 \pm 3.87\%$, UNet of $90.14 \pm 4.69\%$, and ERFNet of $87.18 \pm 6.44\%$.

Table 3. Performance segmentation using the ENet, UNet, and ERFNet methods.

	Sensitivity	PPV	DSC	VOE	VD
ENet					
Mean	93.06%	89.25%	90.89%	16.50%	4.53%
±std	6.37%	3.94%	3.87%	5.86%	9.43%
±CI (95%)	1.36%	0.84%	0.82%	1.24%	2.00%
UNet					
Mean	88.89%	91.89%	90.14%	17.66%	3.16%
±std	7.61%	3.31%	4.69%	6.91%	9.36%
±CI (95%)	1.62%	0.70%	1.00%	1.47%	1.99%
ERFNet					
Mean	89.93%	85.44%	87.18%	22.18%	5.70%
±std	10.92%	5.43%	6.44%	9.61%	14.72%
±CI (95%)	2.32%	1.16%	1.37%	2.04%	3.13%

Analysis of variance (ANOVA) based on DSC was calculated to test statistical differences (a p -value < 0.05 indicates a significant difference) between methods considering all patients ($n = 85$). Table 4 shows how though ENet and UNet minimized the difference between manual and automated segmentation.

Table 4. ANOVA on the DSC showed statistical differences between segmentation methods.

ANOVA	F Value	F Critic Value	p -Value
ENet vs. ERFNet	20.70407668	3.897407169	0.000010236
ERFNet vs. UNet	11.69135829	3.897407169	0.000788084
ENet vs. UNet	1.301554482	3.897407169	0.255553164

Despite the fact that they were statistically identical, they were computationally different. ENet is much faster than UNet. Specifically, Table 5 shows the comparison of computational complexity and performance of the three models. As both ENet and

ERFNet were developed for real-time applications, these are relatively smaller and faster than UNet. As shown in the table, the ENet model has an order of magnitude with fewer parameters than both ERFNet, and UNet while ERFNet has less than half the number of parameters compared to UNet. Consequently, the size of trained ENet is only 6 MB compared to 65 MB for the UNet model. To estimate the delineation time, we considered one of the trained models during the k-fold strategy for all three architectures and then computed the average. Using a fairly advanced GPU device (GEFORCE RTX 2080 Ti, 11 GB VRAM, 4352 CUDA Cores, NVIDIA), it takes only 1 s for ENet and about 1.5 s for UNet to generate segmentation on a 3D dataset (average 40 slices of 512×512). However, when GPU hardware is not available then computation needs to be done on the CPU. In such a scenario, the size of a model can play a big role. On an AMD Ryzen 2950x processor, ENet only takes about 6 s while UNet takes about 40 s to delineate a study. Soon, this computational advantage of ENet may make it possible to use this model to segment cases on simple hardware like iPads or smartphones for faster clinical workflow. Finally, only the ENet model makes use of batch normalization layers, which have some parameters which are not trained, that is, gradients are not back-propagated during the training process.

Table 5. Computational complexity of the three models.

Model Name	Number of Parameters		Size on Disk	Inference Times/Dataset	
	Trainable	Non-Trainable		CPU	GPU
ENet	362,992	8352	5.8 MB	6.17 s	1.07 s
ERFNet	2,056,440	0	25.3 MB	8.59 s	1.03 s
UNet	5,403,874	0	65.0 MB	42.02 s	1.57 s

In Figure 3, we plot the training DSC and Tversky loss function for each DL network for one fold. DSC and Tversky loss plots indicate that the ENet model converges much faster than both ERFNet and UNet. ENet model reaches a DSC = 0.85 in less than 15 epochs. Consequently, it is much faster to train a new ENet model compared to the other two if more training data become available in the future. Another noticeable feature is that the UNet training loss is much less compared to ENet and ERFNet, indicating the presence of overfitting. It can be concluded that even though ENet and UNet models are not statistically different, it may be advantageous to prefer ENet over UNet. Finally, 2D and 3D segmentation examples of three patient studies are shown in Figures 4 and 5, respectively.

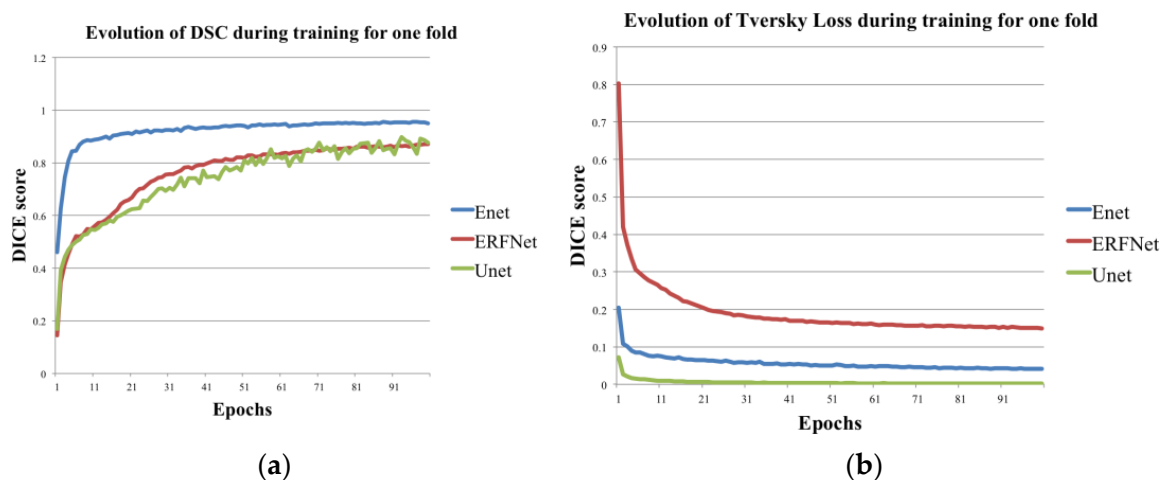


Figure 3. (a) Training DSC and (b) loss function Tversky loss plots for each of the three models for one fold.

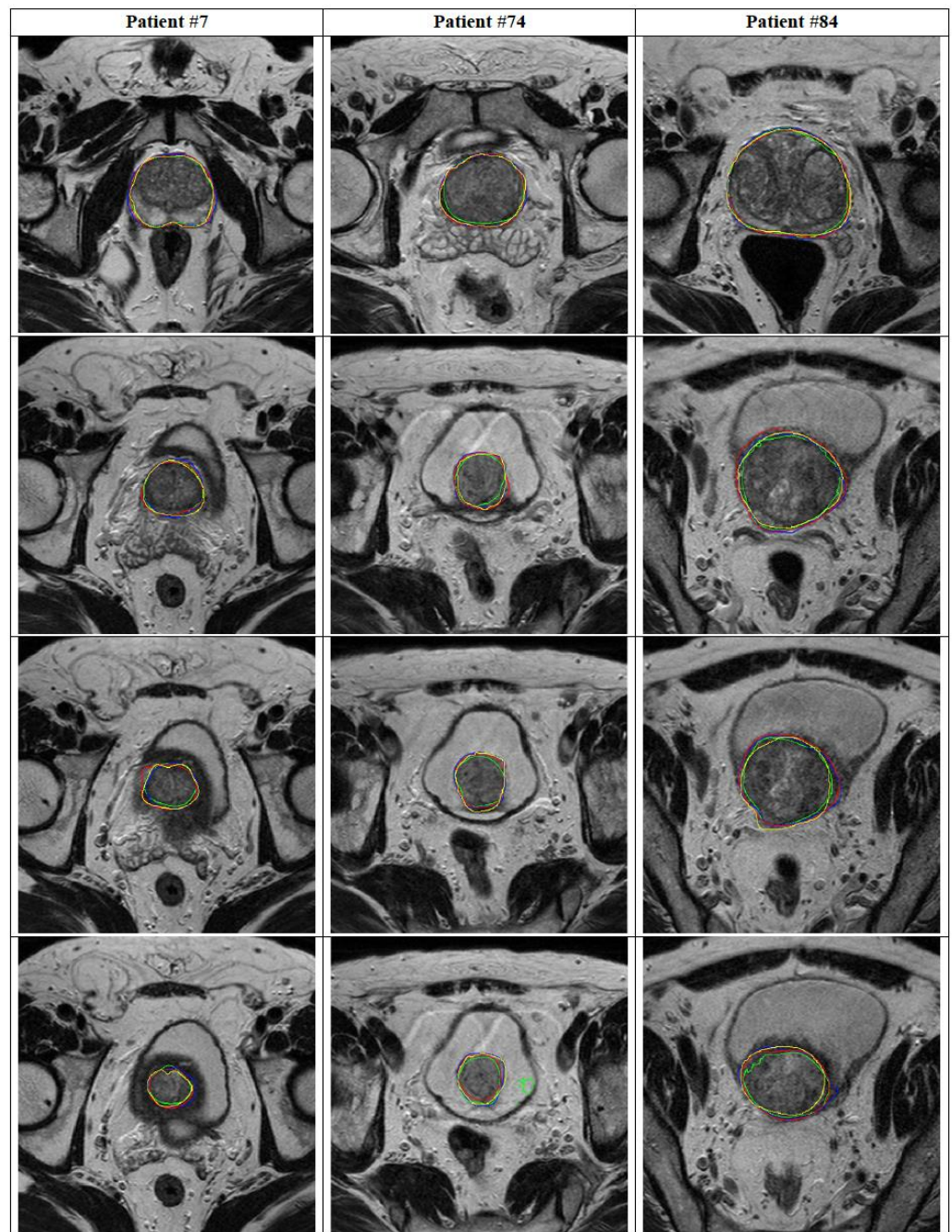


Figure 4. Comparison of segmentation performance for the three Net architectures in #7, #74, and #84 patients (four different slices for each patient). The manual segmentation (yellow), ENet (red), ERFNet (blue), and U-Net (green) are superimposed.

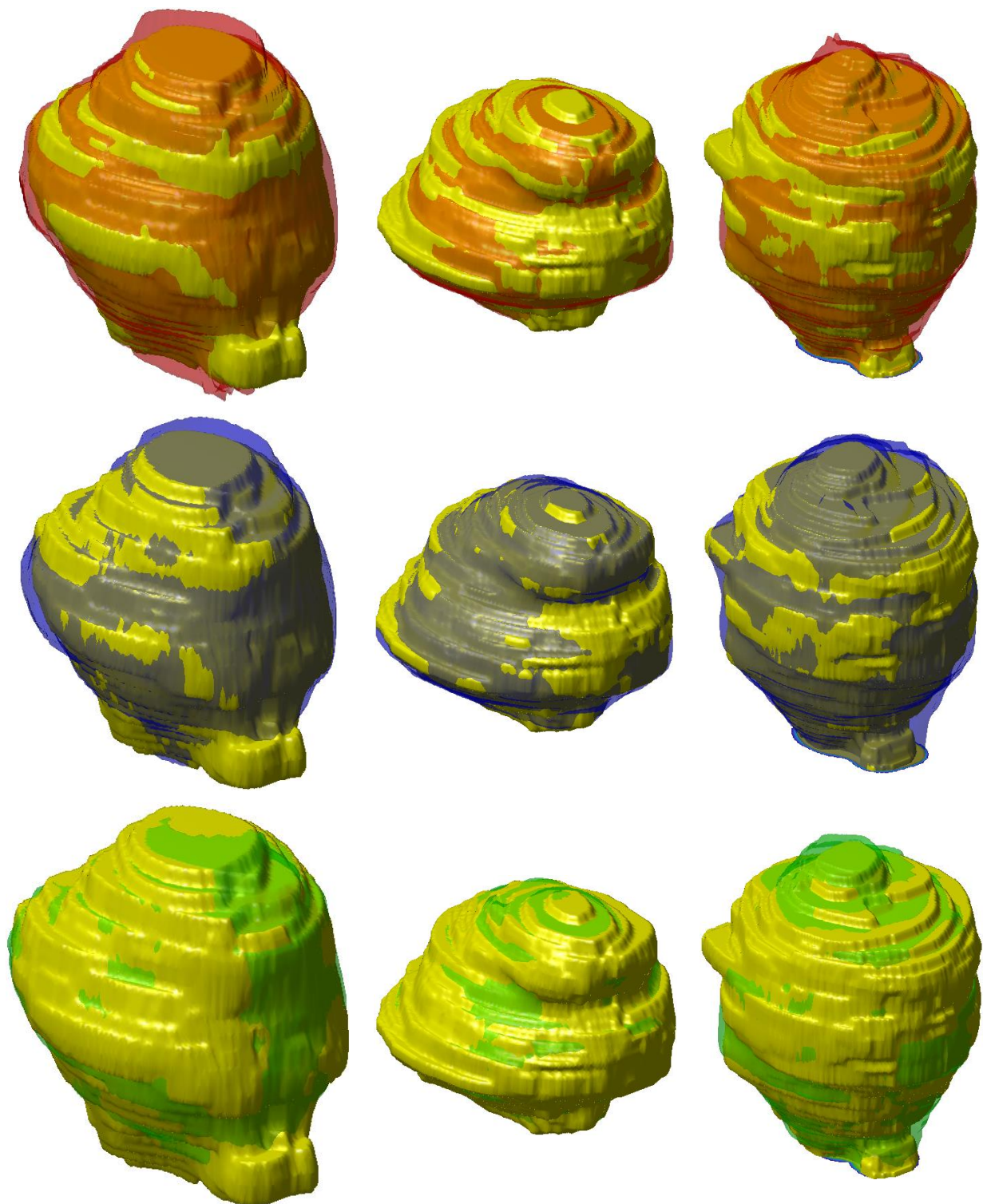


Figure 5. Comparison of 3D segmentation of prostate (patients #7, #74, and #84) for each column using the three Net architectures. The manual segmentation (yellow), ENet (red), ERFNet (blue), and U-Net (green) are superimposed.

4. Discussion

In this paper, we investigate the prostatic region segmentation in MRI studies using three different DL networks (namely UNet, ENet, and ERFNet). The aim is to reduce patient mortality being only a part of prostate cancer that is clinically significant. An accurate and

operator-independent segmentation process is needed to obtain a relevant texture-based prediction model. So, the aim of this work was not only just to test the segmentation results of the proposed models, but to evaluate if these models can yield a practical benefit in obtaining accurate and reproducible results. The inclusion of DL models in radiomics analyses will be reserved for a forthcoming paper. The first model considered in this study was UNet, which has been adopted in several image delineation processes [45]. ENet [43], and ERFNet [44] have been implemented for the segmentation process in self-driving cars, and successfully used in lung and aorta segmentation tasks [48,49]. Specifically, they were used for the segmentation of HRCT images characterized by a very high number of slices for each study (about 600 and 450 slices for the lung and aorta studies, respectively). Authors used 32 patients' studies for the parenchyma extraction process [49], and 72 studies for the aorta segmentation process [48]. In this study, only 85 studies were used considering that each patients' image dataset consists of about 40 slices. In addition, to our knowledge, these DL models have never been applied to prostate segmentation before.

In general, a DL approach requires a multitude of labeled data for training and validation purposes. For this reason, DL models are not widely used in clinical practice. As already reported in the Introduction section, there is a need to develop DL networks capable of obtaining accurate segmentations with few training examples. This issue is addressed in some studies, that is, the one-shot learning approach [53], which eliminates the need for iterative sample selection and annotation and the contrastive learning method [29] for the automated diagnosis of COVID-19 with few samples for training. In our study, we applied all three DL models to a small dataset of 85 studies provided with manual prostate segmentation adopting (i) a data augmentation strategy to reduce overfitting, (ii) a data standardization and normalization to prevent too large weights, to make the model converge faster, and to avoid numerical instability, (iii) the five-fold cross-validation strategy to obtain good results despite the few training examples, and (iv) the Tversky loss function [50] to avoid to predict most voxels as belonging to the background class. In the last case, starting from the consideration that DL methods suffer from the imbalanced data problem because the target (i.e., the prostate) is very small compared to the background, we provided a larger weight to target voxels to learn the foreground object representation more effectively. Finally, we compared the obtained performances showing that accurate and clinically acceptable prostate segmentations with few training examples were obtained using indifferently the three DL models (DSC > 87%).

Specifically, results showed that ENet and UNet had better performance in minimizing the difference between automated and manual segmentations than ERFNet. Substantially, ENet and UNet were statistically identical but computationally different; ENet was much faster than UNet (see Figure 3). Also, the training Tversky loss of the UNet was much less compared to ENet. For these reasons, though UNet and ENet were not statistically different, ENet seems to be the best solution. This could justify the time required to include DL networks in radiomics analyses by removing the user-dependence and achieving accurate prostate segmentations (DSC = 90.89%) using a few training examples. In this way, our model can be used to improve prognosis evaluation and prediction of patient outcomes, allowing the personalization of patient management. However, the results presented in this study derive from the performance of DL networks on proprietary imaging datasets; for routine clinical application, it should be mandatory to test and validate the proposed methods in multicenter studies and/or on a large set of publicly released representative training data, such as PROMISE12 [42]. Moreover, in the present study, we test DL networks for the whole prostate gland segmentation, with ENet demonstrating the best performance; however, a main clinical goal is the segmentation of related prostatic structures or substructures such as the prostatic zones (transition, central and peripheral), neurovascular bundles or seminal vesicles. The performance of DL networks, especially ENet, on this topic should play an essential role in many medical imaging and image analysis tasks such as cancer detection, patient management, and treatment planning including surgical planning, and should be analyzed in future works. Automatic segmentation of

the whole prostate gland and prostatic zone (transition, central and peripheral) without inter-user variability will lead, in the future, to a correct localization of prostate cancer. This result will increase the reliability of computer-aided design (CAD) algorithms which will help automatically create PI-RADS zone maps to reduce inter-user variability among clinicians when interpreting prostate MRI images. In this scenario, radiomics analysis should be performed automatically providing information that can lead clinicians on the management of patients with prostate cancer.

5. Conclusions

Our study demonstrates that faster and less computationally expensive DL networks can perform accurate prostate delineation and could facilitate the adoption of novel imaging parameters, through radiomics analyses, for prostatic oncologic diseases. Specifically, we assessed the performance of three DL networks using data augmentation, standardization, and normalization, and the five-fold cross-validation strategies, and the Tversky loss function in a small dataset of 85 studies. All DL networks achieved accurate prostate segmentations with a DSC > 87%. Nevertheless, differences related to training time and data requirements were highlighted. ENet and ERFNet, developed for self-driving cars, were much faster than UNet. In addition, ENet had better performance (DSC = 90.89%) than ERFNet (DSC = 87.18%). Future studies with more patients could improve the results.

Author Contributions: Conceptualization, A.C. and G.S.; Data curation, F.V. and G.S.; Formal analysis, A.C.; Funding acquisition, A.S. and A.C.; Investigation, A.S.; Methodology, A.S. and A.C.; Project administration, A.Y.; Resources, F.V., M.P., G.C., A.B., G.S.; Software, A.C. and N.D.; Supervision, A.Y.; Validation, F.V.; Visualization, A.C. and A.S. Writing—original draft, A.C., N.D., A.S., and F.V.; Writing—review and editing, A.C. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: This retrospective study was approved by the Institutional Review Board of our Hospital. Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: This work was partially supported by grant W911NF-18-1-0281, funded by the USA Army Research Office (ARO): “Extending Accelerated Optimization into the PDE Framework”, and by grant R01-HL-143350 funded by the National Institute of Health (NIH) “Quantification of myocardial blood flow using Dynamic PET/CTA fused imagery to determine physiological significance of specific coronary lesions”.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Stefano, A.; Comelli, A.; Bravata, V.; Barone, S.; Daskalovski, I.; Savoca, G.; Sabini, M.G.; Ippolito, M.; Russo, G. A preliminary PET radiomics study of brain metastases using a fully automatic segmentation method. *BMC Bioinform.* **2020**, *21*, 325. [[CrossRef](#)]
2. Cuocolo, R.; Stanzione, A.; Ponsiglione, A.; Romeo, V.; Verde, F.; Creta, M.; La Rocca, R.; Longo, N.; Pace, L.; Imbriaco, M. Clinically significant prostate cancer detection on MRI: A radiomic shape features study. *Eur. J. Radiol.* **2019**, *116*, 144–149. [[CrossRef](#)]
3. Baeßler, B.; Weiss, K.; Santos, D.P. Dos Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study. *Investig. Radiol.* **2019**, *54*, 221–228.
4. Gallivanone, F.; Interlenghi, M.; D’Ambrosio, D.; Trifirò, G.; Castiglioni, I. Parameters influencing PET imaging features: A phantom study with irregular and heterogeneous synthetic lesions. *Contrast Media Mol. Imaging* **2018**, *2018*, 12. [[CrossRef](#)]
5. Comelli, A.; Stefano, A.; Coronello, C.; Russo, G.; Vernuccio, F.; Cannella, R.; Salvaggio, G.; Lagalla, R.; Barone, S. *Radiomics: A New Biomedical Workflow to Create a Predictive Model*; Springer: Cham, Switzerland, 2020; pp. 280–293.
6. Comelli, A.; Stefano, A.; Russo, G.; Sabini, M.G.; Ippolito, M.; Bignardi, S.; Petrucci, G.; Yezzi, A. A smart and operator independent system to delineate tumours in Positron Emission Tomography scans. *Comput. Biol. Med.* **2018**, *102*, 1–15. [[CrossRef](#)]
7. Dahiya, N.; Yezzi, A.; Piccinelli, M.; Garcia, E. Integrated 3D anatomical model for automatic myocardial segmentation in cardiac CT imagery. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2019**, *7*, 690–706. [[CrossRef](#)]

8. Comelli, A. Fully 3D Active Surface with Machine Learning for PET Image Segmentation. *J. Imaging* **2020**, *6*, 113. [[CrossRef](#)]
9. Foster, B.; Bagci, U.; Mansoor, A.; Xu, Z.; Mollura, D.J. A review on segmentation of positron emission tomography images. *Comput. Biol. Med.* **2014**, *50*, 76–96. [[CrossRef](#)] [[PubMed](#)]
10. Bi, L.; Kim, J.; Wen, L.; Feng, D.; Fulham, M. Automated thresholded region classification using a robust feature selection method for PET-CT. In Proceedings of the International Symposium on Biomedical Imaging, Brooklyn, NY, USA, 16–19 April 2015; Volume 2015-July.
11. Dhanachandra, N.; Mangle, K.; Chanu, Y.J. Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm. In Proceedings of the Procedia Computer Science, Algiers, Algeria, 18–19 October 2015.
12. Chevrefils, C.; Chérier, F.; Grimard, G.; Aubin, C.E. Watershed segmentation of intervertebral disk and spinal canal from MRI images. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Redondo Beach, CA, USA, 8–11 July 2007.
13. Comelli, A.; Bignardi, S.; Stefano, A.; Russo, G.; Sabini, M.G.; Ippolito, M.; Yezzi, A. Development of a new fully three-dimensional methodology for tumours delineation in functional images. *Comput. Biol. Med.* **2020**, *120*, 103701. [[CrossRef](#)]
14. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *Pattern Anal. Mach. Intell. IEEE Trans.* **2001**, *23*, 1222–1239. [[CrossRef](#)]
15. Stefano, A.; Vitabile, S.; Russo, G.; Ippolito, M.; Marletta, F.; D'arrigo, C.; D'urso, D.; Gambino, O.; Pirrone, R.; Ardizzone, E.; et al. A fully automatic method for biological target volume segmentation of brain metastases. *Int. J. Imaging Syst. Technol.* **2016**, *26*, 29–37. [[CrossRef](#)]
16. Plath, N.; Toussaint, M.; Nakajima, S. Multi-class image segmentation using conditional random fields and global classification. In Proceedings of the 26th International Conference on Machine Learning, ICML 2009, Montreal, QC, Canada, 14–18 June 2009.
17. Dey, D.; Slomka, P.J.; Leeson, P.; Comaniciu, D.; Shrestha, S.; Sengupta, P.P.; Marwick, T.H. Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. *J. Am. Coll. Cardiol.* **2019**, *73*, 1317–1335. [[CrossRef](#)] [[PubMed](#)]
18. Zhou, T.; Fu, H.; Zhang, Y.; Zhang, C.; Lu, X.; Shen, J.; Shao, L. M2Net: Multi-modal Multi-channel Network for overall survival time prediction of brain tumor patients. *arXiv* **2020**, arXiv:2006.10135.
19. Cuocolo, R.; Cipullo, M.B.; Stanzione, A.; Uggia, L.; Romeo, V.; Radice, L.; Brunetti, A.; Imbriaco, M. Machine learning applications in prostate cancer magnetic resonance imaging. *Eur. Radiol. Exp.* **2019**, *3*, 35. [[CrossRef](#)] [[PubMed](#)]
20. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
21. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked denoising autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
22. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017.
23. Ghiasi, G.; Fowlkes, C.C. Laplacian pyramid reconstruction and refinement for semantic segmentation. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Amsterdam, The Netherlands, 8–16 October 2016.
24. He, J.; Deng, Z.; Qiao, Y. Dynamic multi-scale filters for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019.
25. Chen, L.C.; Yang, Y.; Wang, J.; Xu, W.; Yuille, A.L. Attention to Scale: Scale-Aware Semantic Image Segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
26. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
27. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016—Conference Track Proceedings, San Juan, Puerto Rico, 2–4 May 2016.
28. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets Mehdi. *arXiv* **2018**, arXiv:1411.1784.
29. Chena, X.; Yao, L.; Zhou, T.; Dong, J.; Zhang, Y. Momentum Contrastive Learning for Few-Shot COVID-19 Diagnosis from Chest CT Images. *arXiv* **2020**, arXiv:2006.13276.
30. Chen, X.; Yao, L.; Zhang, Y. Residual attention U-net for automated multi-class segmentation of COVID-19 chest CT images. *arXiv* **2020**, arXiv:2004.05645.
31. Garcia-Garcia, A.; Orts-Escobedo, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput. J.* **2018**, *70*, 41–65. [[CrossRef](#)]
32. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
33. Borasi, G.; Russo, G.; Alongi, F.; Nahum, A.; Candiano, G.C.; Stefano, A.; Gilardi, M.C.; Messa, C. High-intensity focused ultrasound plus concomitant radiotherapy: A new weapon in oncology? *J. Ther. Ultrasound* **2013**, *1*, 6. [[CrossRef](#)] [[PubMed](#)]
34. Langan, R.C. Benign Prostatic Hyperplasia. *Prim. Care Clin. Off. Pract.* **2019**, *361*, 1359–1367. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, L.; Wang, Y.; Qin, Z.; Gao, X.; Xing, Q.; Li, R.; Wang, W.; Song, N.; Zhang, W. Correlation between prostatitis, benign prostatic hyperplasia and prostate cancer: A systematic review and meta-analysis. *J. Cancer* **2020**, *11*, 177–189. [[CrossRef](#)]

36. Giambelluca, D.; Cannella, R.; Vernuccio, F.; Comelli, A.; Pavone, A.; Salvaggio, L.; Galia, M.; Midiri, M.; Lagalla, R.; Salvaggio, G. PI-RADS 3 Lesions: Role of Prostate MRI Texture Analysis in the Identification of Prostate Cancer. *Curr. Probl. Diagn. Radiol.* **2019**. In press. [[CrossRef](#)]
37. Cuocolo, R.; Stanzione, A.; Ponsiglione, A.; Verde, F.; Ventimiglia, A.; Romeo, V.; Petretta, M.; Imbriaco, M. Prostate MRI technical parameters standardization: A systematic review on adherence to PI-RADSV2 acquisition protocol. *Eur. J. Radiol.* **2019**, *120*, 108662. [[CrossRef](#)]
38. Turkbey, B.; Fotin, S.V.; Huang, R.J.; Yin, Y.; Daar, D.; Aras, O.; Bernardo, M.; Garvey, B.E.; Weaver, J.; Haldankar, H.; et al. Fully automated prostate segmentation on MRI: Comparison with manual segmentation methods and specimen volumes. *Am. J. Roentgenol.* **2013**, *201*, W720–W729. [[CrossRef](#)]
39. Rusu, M.; Purysko, A.S.; Verma, S.; Kiechle, J.; Gollamudi, J.; Ghose, S.; Herrmann, K.; Gulani, V.; Paspulati, R.; Ponsky, L.; et al. Computational imaging reveals shape differences between normal and malignant prostates on MRI. *Sci. Rep.* **2017**, *7*, 1–11. [[CrossRef](#)]
40. Comelli, A.; Terranova, M.C.; Scopelliti, L.; Salerno, S.; Midiri, F.; Lo Re, G.; Petrucci, G.; Vitabile, S. A kernel support vector machine based technique for Crohn's disease classification in human patients. In Proceedings of the Advances in Intelligent Systems and Computing, Torino, Italy, 10–12 July 2017; Springer: Cham, Switzerland, 2018; Volume 611, pp. 262–273.
41. Ghose, S.; Oliver, A.; Martí, R.; Lladó, X.; Vilanova, J.C.; Freixenet, J.; Mitra, J.; Sidibé, D.; Meriaudeau, F. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. *Comput. Methods Programs Biomed.* **2012**, *108*, 262–287. [[CrossRef](#)]
42. Litjens, G.; Toth, R.; van de Ven, W.; Hoeks, C.; Kerkstra, S.; van Ginneken, B.; Vincent, G.; Guillard, G.; Birbeck, N.; Zhang, J.; et al. Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Med. Image Anal.* **2014**, *18*, 359–373. [[CrossRef](#)]
43. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. *arXiv* **2016**, arXiv:1606.02147.
44. Romera, E.; Alvarez, J.M.; Bergasa, L.M.; Arroyo, R. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 263–272. [[CrossRef](#)]
45. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Fuzhou, China, 13–15 November 2015.
46. Prostate Imaging Reporting & Data System (PI-RADS). Available online: <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS> (accessed on 17 February 2019).
47. Warfield, S.K.; Zou, K.H.; Wells, W.M. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* **2004**, *23*, 903–921. [[CrossRef](#)] [[PubMed](#)]
48. Comelli, A.; Dahiya, N.; Stefano, A.; Benfante, V.; Gentile, G.; Agnese, V.; Raffa, G.M.; Pilato, M.; Yezzi, A.; Petrucci, G.; et al. Deep learning approach for the segmentation of aneurysmal ascending aorta. *Biomed. Eng. Lett.* **2020**, 1–10. [[CrossRef](#)]
49. Comelli, A.; Coronello, C.; Dahiya, N.; Benfante, V.; Palmucci, S.; Basile, A.; Vancheri, C.; Russo, G.; Yezzi, A.; Stefano, A. Lung Segmentation on High-Resolution Computerized Tomography Images Using Deep Learning: A Preliminary Step for Radiomics Studies. *J. Imaging* **2020**, *6*, 125. [[CrossRef](#)]
50. Salehi, S.S.M.; Erdogmus, D.; Gholipour, A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Quebec City, QC, Canada, 10 September 2017; pp. 379–387.
51. Tversky, A. Features of similarity. *Psychol. Rev.* **1977**, *84*, 327. [[CrossRef](#)]
52. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
53. Zheng, H.; Yang, L.; Chen, J.; Han, J.; Zhang, Y.; Liang, P.; Zhao, Z.; Wang, C.; Chen, D.Z. Biomedical image segmentation via representative annotation. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, HI, USA, 27 January–1 February 2019; pp. 5901–5908.