International Journal of Industrial Engineering & Production Research September 2022 Vol. 33, No. 3: 1-16
DOI: 10.22068/ijiepr.33.3.1

RESEARCH PAPER

Prostate Segmentation and Lesions Classification In CT Images Using Mask R-CNN

AmirHossein Masoumi [1], Rouzbeh Ghousi [2]*& Ahmad Makui [3]

Received 27 July 2021; Revised 30 April 2022; Accepted 23 May 2022;
© Iran University of Science and Technology 2022

ABSTRACT

Non-cancerous prostate lesions such as calcification, prostate enlargement, and inflammation cause too many problems for men's health. This research proposes a combination of image processing techniques and deep learning methods for classification and segmentation of the prostate in CT-scan images by considering the experienced physicians' reports. Due to the various symptoms and nature of these lesions, a three-phases innovative approach has been implemented. In the first phase, using Mask R-CNN, in the second phase, considering the age of each patient and comparison with the standard size of the prostate gland, and finally, using the morphology features, the presence of three common non-cancerous lesions in the prostate gland has investigated. A hierarchical multitask approach is introduced and the final amount of classification, localization, and segmentation loss is 1%, 1%, and 7%, respectively. Eventually, the overall loss ratio of the model is about 9%. In this study, a medical assistant approach is introduced to increase diagnosis process accuracy and reduce error using a real dataset of abdominal and pelvics' CT scans and the physicians' reports for each image. A multi-tasks convolutional neural network; also presented to perform localization, classification, and segmentation of the prostate gland in CT scans at the same time.

KEYWORDS: *Computer-aided diagnosis; Deep learning; Mask R-CNN; Prostate lesion; Image processing.*

1. Introduction

In developed countries, in men, prostate cancer (after skin cancer) is the second most common cancer and the second deadliest cancer (after lung cancer), and one in six men develops this cancer [1]. The structure of this gland is in such a way that it has an anterior and peripheral zone, as well as central and mid zone, so that the anterior shell and the central zone are structurally different, which is why diseases of the external and internal zones of the prostate are different [2]. The three most common forms of prostate disease are prostatitis, BPH , and prostate calcification. A man may experience one or more of these conditions, and in some cases, their damages will cause serious health threats when they occur. These lesions cause enlargement of the prostate gland, constricting the urethra and causing various urinary symptoms [3]. Studies show that between 30 and 50 percent of men over 50 have prostate lesions; the diseases' cause is still unknown. However, clinical studies have attributed it to age, genetics, race, hormonal factors, chemicals, and nutrition [4]. Gordon et al. used a group Markov model, estimated the health system costs for prostate cancer treatment from 2016 to 2025. In this study, total prostate disease costs in Australia in 2016 were estimated at 9.27 million US dollars, increasing to 3.843 million US dollars in 2025; i.e., average patient costs in 2025, with a 42 percent increase, expected to meet 34,941 US dollar [5].

Therefore, to control unwanted costs over the next decade, it is vital to develop systematic strategies for timely diagnosis of suspected patients before the diseases find a way to spread. Even though prostate stones are common, yet in many cases, they remain unknown until adulthood. Since many stones are small and

* Corresponding author: ***Rouzbeh Ghousi***
ghousi@iust.ac.ir

1. *School of Industrial engineering, Iran University of Science & Technology, Tehran, 13114-16846, Iran.*
2. *School of Industrial engineering, Iran University of Science & Technology, Tehran, 13114-16846, Iran.*
3. *School of Industrial engineering, Iran University of Science & Technology, Tehran, 13114-16846, Iran.*

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

asymptomatic, they are challenging to diagnose and often have nonspecific symptoms. Tang et al. performed a study on subjects who underwent a comprehensive medical checkup at the West Hospital of China between 2014 and 2015. During this period, 70,546 male subjects underwent a transabdominal ultrasonography examination for screening. They used Logistic regression analysis to analyze the association between risk factors and prostatic calcification. Finally, after the all-patient analysis group, age and anterior-posterior diameter were more potent risk factors of prostatic calcification than the other factors [6].

After explaining the research goals, a background on medical information on the prostate disease and examining the research basics have been discussed in the section 2. Thus, the basic concepts of image processing, neural networks, and deep learning to image classification and segmentation and modeling have been investigated thoroughly in this section. Section 3 describes the proposed method for diagnosing the prostate gland, semantic segmentation, and classification of abdominal and pelvic images and detecting lesions in CT scan images; furthermore, the mean of implementing this method has been described. The outcome of the implementation of the model and its analysis have been presented in the section 5.

2. Related Works

Today, deep learning and image-processing techniques are widely used for diagnosis and treatment aims. Since time is an essential factor in treatment procedures, timely and accurate diagnosis of lesions through medical images is highly important. Prostate segmentation is one of the most challenging medical radiology topics due to low soft-tissue contrast in CT images [7]. Therefore, prostate segmentation on CT images has been classified into three different strategies: Contour and shape-based, Region-based, and supervised and unsupervised classification [8].

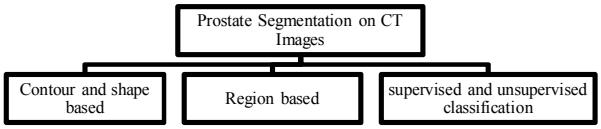


Fig. 1. The strategies for prostate segmentation

Contour and shape-based methods exploit prostate boundary and edge information for segmentation. In this category, Active shape models are used more than the other methods in prostate CT images. Tang et al. use ASM to segment the prostate, bladder, and rectum simultaneously [9]. Contrary to them, Feng et al. proposed an ASM based on a weighted combination of gradient and probability distribution function features to segment only the prostate [10].

In region-based methods, local intensity and statistical tools like mean and standard deviation is used in an energy minimization framework to achieve segmentation. This category is divided into several techniques such as level sets, graph partition, and atlas-based. However, the most popular technique is the atlas-based one. Acosta et al. evaluated the different strategies to simultaneously segment prostate, bladder, and rectum from CT images by selecting the most similar atlases from a prebuilt 24-atlas subset [11]. In another paper, Acosta et al. used non-rigid demons registration to build a multi-atlas-based segmentation of the prostatic urethra to extend the number of atlases for reducing the errors. They used leave-one-out cross-validation for scoring their method [12].

The classifiers-based segmentation methods consider the prostate as a learning problem. Since 2014, much research has focused on image semantic segmentation, and the results promise significant advances. The main reason for their success is due to the use of different Convolutional Neural Networks that can counter pixel-level estimates using pre-trained networks on large-scale datasets. Semantic segmentation differs from image-level classification and objects recognition; it also requires output masks with the two-dimensional spatial distribution. Traditional semantic image processing methods were based on spectral clustering theory. The image, depending on the relationship of weights between different pixels and the threshold value given to the model, is divided into two parts. The first semantic image segmentation algorithm was introduced in 2000 by Shi et al., named N-Cut. This algorithm showed the least difference between segments and the maximum difference between pixels in each segment. In addition to the low accuracy, this image segmentation method requires user intervention; therefore, it is

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

not suitable for rapid segmentation [13]. Before introducing deep neural networks, feature classification was the most commonly used method of image processing. Exactly similar to the concept of feature in machine learning and pattern recognition, in computer vision and image processing, the feature refers to the information used to solve the original model. Liu et al. introduced the essential feature extraction methods used in the semantic segmentation of images. For instance, methods such as pixel colour, the histogram of oriented gradients, scale-invariant feature transform, local binary pattern, Harris Corners, subpixel corners, Features from Accelerated Segment Test, Susan test, Fast-ER, multiple AGAST, detector, and the Textons test, can be named [14].

Most of the semantic image processing traditional methods were based on Low-order visual information, could not deliver accurate results. With the advent of machinery vision since 2007, various fields of science, including image segmentation, have experienced a significant leap. Long et al. presented a Fully Convolutional Networks (FCN) to provide the algorithm's desired input size. In traditional methods, image segmentation used to be done in three time-consuming stages [15]. Noh et al. presented a non-convolutional neural network consisting of deconvolution and decapitalization layers. In a manner that non-convolutional layers classify each pixel and segment the image pooling layers as well. The algorithm then continues the learning process with different input images' resolutions to provide the image's best semantic segmentation [16]. Zhao et al. introduced a network called PSP[2] , attempting to optimize the entire network in the middle layers of the network through computing an auxiliary error based on the general image information and architecture of the ResNet [3] [17]. Guo et al. presented a deep feature-learning model using sparse patch matching to segment the prostate in MRI images. They used a stacked sparse auto-encoder to extract the latent features instead of handcrafted ones due to obtaining more accuracy [18]. Yan et al. suggested using an autoregressive neural network as a classifier to detect prostate cancer areas in MRI images. They introduced the Energy Minimization Procedure method to identify the feature map by considering the relationship between adjacent pixels [19]. They also proposed another model for prostate segmentation based on features extracted using convolutional neural networks; in his method, the Alexnet architecture was used to extract features from candidate sets of prostate images [20]. Milletari et al. trained an end-to-end convolutional neural network on prostate MRI images. He also introduced a new objective function based on the Dice coefficient to optimize the training phase parameters to solve the contrast between the Region of Interest (ROI) and the background in 3D images [21]. Cheng et al. used Holistically-Nested Networks (HNNs) inspired by an end-to-end convolutional network to segment the prostate [22].

Algorithms-based convolutional neural networks have led to significant advances in solving problems in the field of computer vision, including object and location detection, semantic segmentation, and instance segmentation. Therefore, applying convolutional neural network-based approaches to solve medical image analysis problems has become widespread [23]. As a result, using convolutional neural networks has led to promising results in various medical fields such: auto-diagnosis of diabetic retinopathy, auto-diagnosis of melanoma, accurate measurement of internal cardiovascular haemorrhage, three-dimensional segmentation of liver and lung, mammography images' segmentation, and Knee cartilage Three-dimensionalization has been documented [24].

Recently, there can be seen significant improvements in convolutional neural networks based architectures; Feature Pyramid Networks, Autoregressive, Refine Net, Dilated Nets, and Retina-nets are some examples to be named [25]. Besides, single-shot models such as YOLO and SSD have made object recognition in images 100 to 1000 times faster than regional-based algorithms [26]. In 2015, U-Net architecture was used by ISBI[4] to segment images of the brain and spinal cord; it was also used to split neural structures. Since then, U-net architecture has been used for a wide range of medical image analysis topics, including volumetric segmentation of 3D structures and tomographic photo reconstruction [27]. Region-based overlapping convolutional neural networks were developed that improve fast region-based CNN for semantic segmentation, object recognition, and sample image segmentation. Region-based overlapping convolutional neural networks are often used to identify large-scale objects, semantic, and sample segmentation and improve texts [28]. Yu et al. proposed a volumetric convolutional neural network with mixed,

[2] Pyramid Scene Parsing
[3] Residual Network
[4] International Society for Biomedical Imaging

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

stacked connections to segment 3D prostate images. In the presented model, a fully convolutional model with residual blocks was introduced to enable volume-to-volume prediction. Also, an auxiliary variable was added to increase the matching speed [29]. Kazemifar et al. used the U-Net structure for building a deep CNN model. They improve automatic delineation of the boundaries between target and surrounding normal soft tissues on prostate CT scan images [30]. Anjanli et al. presented a fully automated 2D-3D U-Net architecture for male pelvic CT image segmentation. The architecture consists of a 2D organ volume localization network followed by a 3D segmentation network for volumetric segmentation of prostate, bladder, rectum, and femoral heads [31]. Polymeri et al. proposed an algorithm based on CNN to measure prostate volume in cancer patients [32]. Ma et al. suggested using an image patch-based deep convolutional network on the model to segment the prostate from regions of interest in CT scans. Finally, with a multi-atlas label, they manage to do image segmentation [33].

Considering the physicians' reports about the images can significantly improve image classification accuracy [34]. Kisilev et al. introduced a new multi-task CNN method for detection and semantic description of lesions in mammography images [35]. Ghavami et al. compared six CNNs to segment prostate glands in MRIs. In this research, a set of 232 MRIs with labels that experienced physicians provided was used [36].

The current research's main intention is to provide a systematic method leading to accurate detection of prostate and the mentioned lesions through CT-scan processing. Therefore, it has been attempted to present a solution providing automatic and systematic detection of prostate lesions in CT images through using the science of image processing and deep neural networks by considering the age and morphological features.

3. Materials and Methods

3.1. Data set

A series of abdominal and pelvic CT scans of 130 patients recorded in three different time intervals at Milad Hospital (located in Tehran, Iran) has been collected in the first step. CT scans were recorded with four cuts' types of 1.5, 3, 5, and 7 mm in both injection and non-injection modes. Due to the full bladder, CT scans with drug injection could be more accurate in showing different body parts. In this study, images with a 5 mm slice were used as the final data, taking approximately 100 images of each patient at each abdominal and pelvis imaging time. 595 images of the prostate gland were extracted from a dataset obtained from Milad Hospital in Tehran.

Fig. 2. Healthy prostate and non-cancerous prostate lesions

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

3.2. Data preparation

3.2.1. Image separation

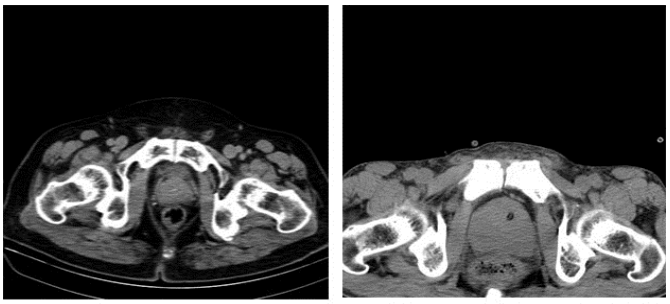
Only ~15% of the CT slices acquired for a pelvic CT scan for a patient contains Prostate. One of the main Difficulties in this study was to separate prostate CT scans from other images in each image set. The prostate gland remains relatively small throughout childhood; then, it begins to grow at the time of adolescence due to testosterone secretion. The prostate gland reaches approximately a regular size at about age 20 and remains the same (about 30 cc) until 40-50. The prostate's average weight in a healthy adult is about 20 to 25 grams, and its measures are about 3*2*4 cm [2]. As a result, the prostate can finally be seen in four CT scans images with a 5 mm cut. At this stage, referring to the physician's report, the state of prostate health and the types of the lesions were determined.

3.2.2. Noise data

In this research, noise data has been divided into two categories (
Fig. 3):

- Images of patients with a catheter,
- The images became blurred due to the patient's motion during the shooting.

The first category of noise data was isolated, and 15 images were excluded from the dataset. Because of the similarity of catheter cross-sectional view to calcification disease, they cause errors in the model learning process. For the second category, the image resolution was improved via image enhancement techniques; therefore, the images became usable.



blured image Image with a catheter

Fig. 3. Noise data

3.2.3. Annotation

The annotation at the image level, the tags are generally attributed to the whole image; therefore, it does not specify which part of the image each tag belongs to[37]. However, the annotation at the region level and annotating images, each tag's relationship to that image's region is also specified. Thus, in each tag's training step, annotation at the regional level focuses only on the relevant region; therefore, it makes the learning process faster by ignoring the unrelated parts[38]. The selected region of each image, that is, the prostate gland, was determined and colored. In this research, images were first classified into two groups.

Edge detection is an essential method for image processing, pattern recognition, and computer vision. The outcome of using these methods in an image, ideally, should result in a set of curves showing the boundaries of the objects and the curves that are inconsistent with the surface orientation. Besides, applying its algorithms to an image can decrease the processing time and increase accuracy [39]. So in this research, portions of the image that were not part of the selected region, considering contour lines in the image and coding in Python software, were removed from all images.

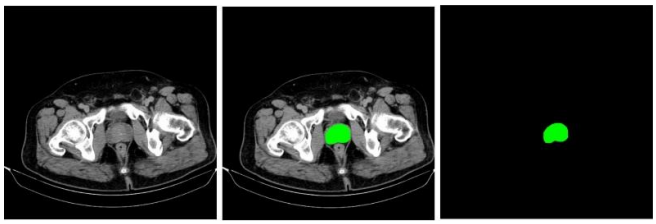


Fig. 4. Annotating and eliminating the background of the images

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

3.2.4. Data augmentation

Here, in the process done, it has been attempted to artificially increase the network generalization power over increasing the variety and the volume of data. The portion of images with calcification was about one-ninth of the whole dataset and declining the learning accuracy. It also caused model overfitting; therefore, data augmentation techniques were used to increase patients' datasets with calcification. Therefore, images being rotated and zoomed in, new images were created, and the image ratio changed to one-third. Finally, the total number of images increased to 700.

3.2. Three-Phases diagnosing process

We should present a method for automatically detecting the prostate gland in CT images due to prostate gland tissue similarity to other abdomen and pelvis organs. More than that, the nature and symptoms of these three lesions were different. For example, the diagnosis of BPH is related to the age and size of the prostate gland, while prostatitis is diagnosed by considering its shape. Additionally, two lesions might have been seen in the same image simultaneously. These limitations lead us to present a novel three-Phases diagnosing process and detect each lesion in one phase.

Fig. 5 illustrates the three-Phases diagnosing process, which is presented in this research:

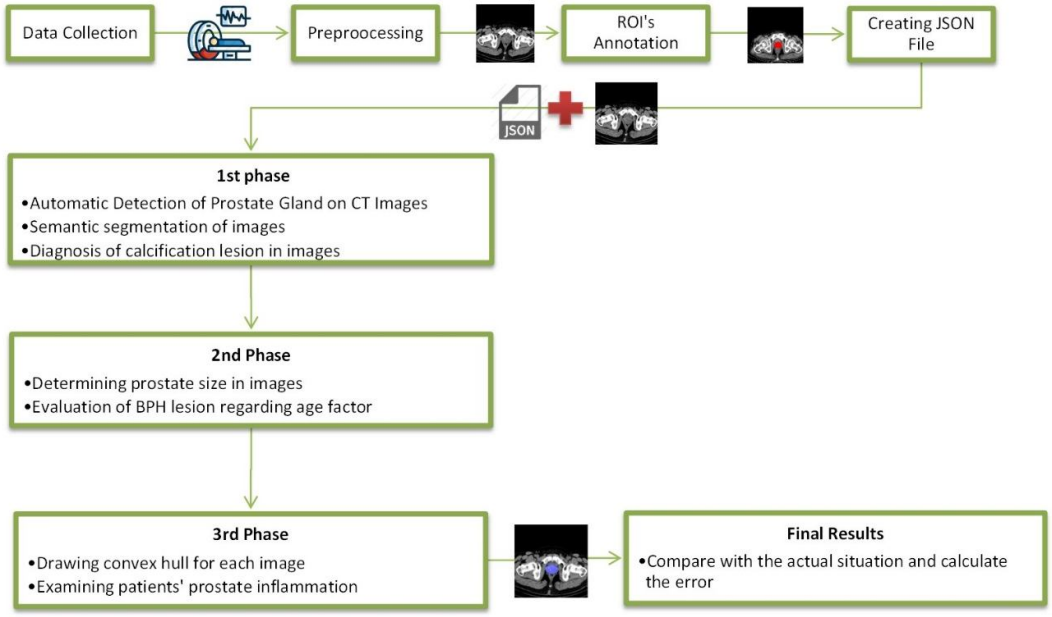
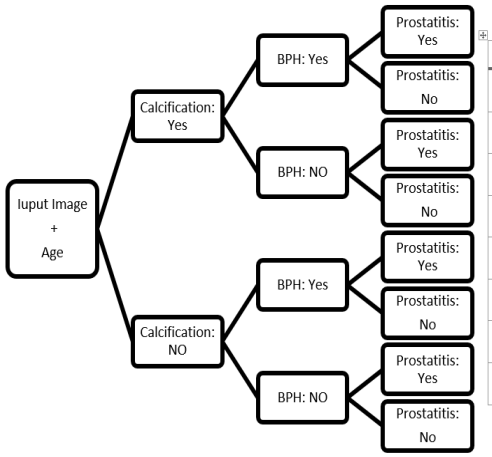


Fig. 5. Three-step diagnosing process

Finally, Fig. 6 shows eight different classes that may be predicted for images.



Class	Calcification	BPH	Prostatitis	Final Label
G_1	No	No	No	healthy
G_2	No	No	Yes	Patient
G_3	No	Yes	No	Patient
G_4	No	Yes	Yes	Patient
G_5	Yes	No	No	Patient
G_6	Yes	No	Yes	Patient
G_7	Yes	Yes	No	Patient
G_8	Yes	Yes	Yes	Patient

Fig. 6. Decision tree for prediction of the state of each Image

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

3.3.1. The first phase

This multi-tasks deep network aims to find the best parameters to convert input images into output information. As a result, it is necessary to enter all the expected data for this network's output. The network uses its backpropagation pattern and learns the optimal combination of parameters and weights to convert input images to output knowledge. The transfer learning method is used to extract features from abdominal and pelvic CT scans. As stated before, the purpose of the Mask-RCNN learning network is to obtain three types of information for each of the target items in the image: the object's overlay area, the bounding box over the object, and the corresponding label. Fig. **7** illustrates the Mask R-CNN framework, which is used in this research by changing the classifier layer to improve the network's performance.

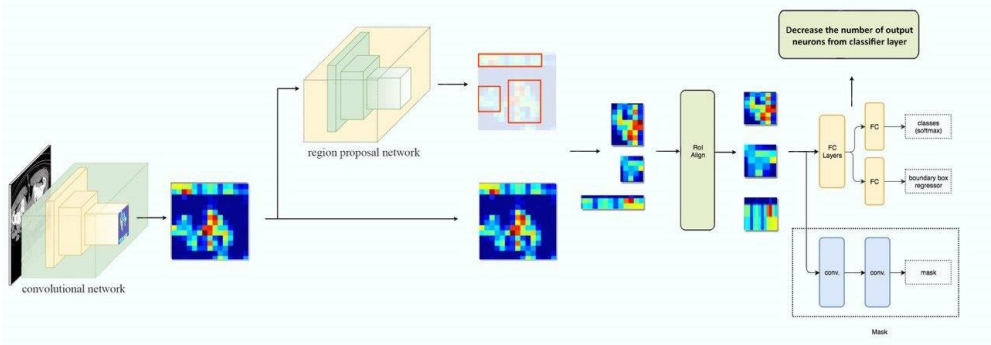


Fig. 7. The mask R-CNN framework

Fig. **7** shows the schematic diagram of the segmentation method, which consists of feature extraction, detection, and segmentation modules. The first step in a deep learning network is to provide its input. In this research, input data consisted of CT scans and all expected output data for each image. Then, the RPN generates the candidate region box in the feature map, and the ROIAlign's outputs are fixed-size feature images. Following that, the detection module is used for target box classification. A CNN in the segmentation module completes the process, and the segmented CT scan is used as output.

- **Image Sharpening**

The first step in a deep learning network is to provide its input. In this research, input data consisted of CT scans and all expected output data for each image. First, the colored region perimeter is extracted from the image using the contour lines concept for reaching the ideal output. Contour is a list of points representing a curve in the image; this curve can point around an object. According to the obtained images, the contours in this image can be obtained by calculating the image gradient and identifying points in the image where the gradient has changed; therefore, the points with the change in their gradient direction represent the edges, and the input boundary shows the colored region. Thus, the image's vertical and horizontal gradients should be calculated; then, boundary regions should be calculated while combining these two gradients. If A is the input image, then the size and angle of the image gradient can be calculated as equations (1)-(4):

$$B_v(j,k) = A(j,k+1) - A(j,k-1) \quad (1)$$

$$B_h(j,k) = A(j+1,k) - A(j-1,k) \quad (2)$$

$$B(j,k) = \sqrt{B_h(j,k)^2 + B_v(j,k)^2} \quad (3)$$

$$\theta = \arctan\left(\frac{B_v(j,k)}{B_h(j,k)}\right) \quad (4)$$

The well-known Soble filter for image processing can also obtain gradients in horizontal and vertical directions. The following image shows the gradient size and angles in a full prostate image and how the edges and points of resonance are represented. After obtaining the prostate boundary points in the image, the bounding box should also be calculated. If the set P shows all boundaries of the prostate, the bounding box is calculated as Equations (5) and (6):

$$(X_{\min}, Y_{\min}) = \left(\{\min(x) \mid x \in p\}, \{\min(y) \mid y \in p\}\right) \quad (5)$$

$$(X_{\max}, Y_{\max}) = \left(\{\max(x) \mid x \in p\}, \{\max(y) \mid y \in p\}\right) \quad (6)$$

The category where the target object is located is the latest information needed from each image. Since drawing a comparison with RGB color

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

codes for non-patient images, RGB color codes for patient's images can be distinguished; the relevant category can also be easily identified due to the differentiation of colors at the image marking stage. At last, the results of the above calculations for the prostate are shown in Fig. 8.

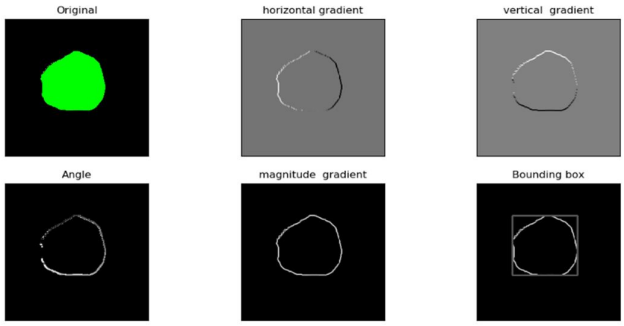


Fig. 8. Steps to calculate the image gradient

The information extracted at this stage was stored in a JSON file, and along with the images, they were entered into the model (**Fig. 9**).

{**"licence"**:1,**"file_name"**:"1.6.2r.jpg",**"coco_url"**:"C:/Users/TEN/Documents/Mask/SAMPLES/annotator/1.6.2r.jpg" ,**"height"**: 512 ,**"width"**: 512 , **"date_captured"**: "2019-04-10 17:02:52" ,**"flickr_url"**: " C:/Users/TEN/Documents/Mask/SAMPLES/annotator/1.6.2r.jpg " ,"id": 582191 },{**"segmentation"**: [[266.0, 343.5, 282.0, 335.5, 291.0, 327.5, 297.5, 324.0, 302.5, 315.0, 301.5, 303.0, 289.0, 290.5, 271.0, 283.5, 254.0, 282.5, 234.0, 287.5, 223.5, 298.0, 220.5, 305.0, 220.5, 314.0, 226.5, 327.0, 235.0, 336.5, 266.0, 343.5]] ,**"area"**: 3739.5 ,**"iscrowd"**: 0 ,**"image_id"**: 582191 ,**"bbox"**: [220.5, 282.5, 82.0, 61.0] ,**"category_id"**: 1 ,"id": 100520005265 }

Fig. 9. JSON file format

- **RPN network**

As this framework's approach is semantic segmentation, the prediction of image foreground and background is essential. RPN predicts them by bounding boxes with different multiples and proportions of length and width. As mentioned before, the prostate is located almost in the center of each image; by removing the portions of the image that were not part of the selected region, the bounding box in the predicted feature image is delimited to decrease the generation time of candidate regions.

- **ROIAlign layer**

The ROIAlign layer recognizes the generated candidate ROI by using the pooling operation. The structure is consists of the pooling operations, which transfer the feature maps with different sizes through the ROIAlign layer and maps them into fixed-size feature maps.

- **Loss function**

The multi-task loss function of Mask R-CNN combines the loss of classification, localization, and segmentation mask as equation (7):

$$L_{total} = L_{class} + L_{box} + L_{mask} \tag{7}$$

Before describing each of them, the parameters needed to define the triple loss functions was introduced in Tab**.** :

Tab. 1. Parameters used in error calculation

Parameters	Description
Pi	anticipated probability of an object being placed in the anchor i
Pi*	real label (binary) indicating whether anchor i is an object or not
ti	coordinates of the four points predicted for the object's boundary box
ti*	actual coordinates of the boundary box
Ncls	normalization-class parameter
Nbox	normalization parameter of anchors
$\hat{y}_{ij}^{k}$	the predicted value for pixel (i, j) in the mask learned from the k label in the network
yij	pixel label (i, j) in a real mask with m×m measurements

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

Masks produced in this network have m×m measurements for each part of the selected region and each category. Therefore, the network's overall output size in the mask area is equal to K×m^2. Since the model attempts to learn the correct mask for each class, there is no competition between classes to produce the mask. L_{mask} is the mean cross-entropy loss function that is calculated for each k class as equation (8):

$$L_{mask} = -\frac{1}{m^2}\sum_{1\le i,j\le m}\left[y_{ij}\log\hat{y}_{ij}^{k} + \left(1-y_{ij}\right)\log\left(1-\log\hat{y}_{ij}^{k}\right)\right] \quad (8)$$

The loss function for the classification is also calculated as equation (9):

$$L_{class} = -\frac{1}{N_{cls}}\sum_{i}\left[p_i^{*}\log p_i + \left(1-p_i^{*}\right)\log\left(1-p_i\right)\right] \quad (9)$$

At last, the localization loss function, also, has been calculated based on the difference between the coordinates t_i and t_i^*, plus taking the L_1^{smooth} function into account, which is resistant to noises and points located away as equation (10),(11)

$$L_1^{smooth}(x) = \begin{cases} 0.5x^2, if\ |x|<1 \\ |x|-0.5, otherwise \end{cases} \quad (10)$$

$$L_{box} = -\frac{1}{N_{box}}\sum_{i} p_i^{*}.L_1^{smoth}\left(t_i - t_i^{*}\right) \quad (11)$$

Therefore, the ultimate goal of the deep learning network introduced in this research is to minimize L_{total} to obtain the best and most accurate masks and boundary boxes around the objects and identify each object's correct class.

- **Determining the optimal number of learning and evaluation data**

In a dataset classification, the aim is to reach the highest possible accuracy in classification and identifying different classes. Therefore, to select the best ratio, the confusion matrix was formed in three different data ratios in learning and evaluating (Tab. -4). As can be seen, the accuracy index in the ratio of 80% and 85% are approximately equal; but since it is costly in medical matters if a patient is placed in the healthy group, according to the value of the Recall index, eventually the learning and classifying data were divided into 85% to 15%.

Tab. 2. Confusion matrix in 85%-15% ratio

The Proportion of Training Data to Evaluation Data: 85%-15%		Real state: Normal	Real state: Calcification	Accuracy Indexes
Prediction	Normal	69	1	$c = 99.0\%$
	Calcification	0	35	$recall = 100\%$ $precision = 98.5\%$

Tab. 3. Confusion matrix in 80%-20% ratio

The Proportion of Training Data to Evaluation Data: 80%-20%		Real state: Normal	Real state: Calcification	Accuracy Indexes
Prediction	Normal	45	1	$ac = 99.2\%$
	Calcification	2	22	$recall = 98.9\%$ $Precision = 98.9$

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

Tab. 4. Confusion matrix in 90%-10% ratio

The Proportion of Training Data to Evaluation Data: 90%-10%		Real state		Accuracy Indexes
		Normal	Calcification	
Prediction	Normal	93	0	$ac = 95.0\%$
	Calcification	1	45	$recall = 95.74\%$ $Precision = 97.8\%$

3.3.2. The second phase

After CT scans segmentation, considering patients' age, each patient's prostate gland's size would be calculated. Using a classifier network, CT scans would be classified into two groups; with and without prostate enlargement. Tab. shows the mean width, height, length of the prostate in each age group.

Tab. 5. The measures of the prostate in different age groups [2]

Age	Group 1 (40–49 years)	Group 2 (50–59 years)	Increase from Group 1 to Group 2 (%)	Group 3 (60–70 years)	Increase from Group 2 to Group 3 (%)	P
Width (mm)	44.45±4.67	45.53±4.70	2.4	46.69±5.10	2.5	0.0134
Height (mm)	31.66±4.38	32.34±4.60	2.1	33.50±5.91	3.6	0.0471
Length (mm)	37.55±4.27	38.89±4.64	3.5	41.13±6.18	5.7	0.0002
TPV (mm)	28.17±8.75	30.83±9.64	9.4	35.03±17.41	13.6	0.0023
TZW (mm)	30.25±4.84	31.92±5.15	5.5	34.21±7.08	7.2	0.0001
TZH (mm)	21.64±3.73	23.24±4.58	7.4	25.02±6.07	7.7	0.0001
TZL (mm)	25.24±4.03	27.20±4.46	7.8	29.43±6.24	8.2	<0.0001
TZV (mm)	8.95±3.88	11.15±5.95	24.6	14.92±13.49	33.8	0.0001

TPV: total prostate volume, TZH: transition zone height, TZL: transition zone length, TZV: transition zone volume, TZW: transition zone width

In this study, a two-step algorithm was used to calculate both the prostate gland's area and dimensions. In the first step, finding the images' contour lines, the masked boundary regions' coordinates in the first step were determined. In the second step, the shoelace algorithm was used to calculate the area and calculate the dimensions.

3.3.3. The third phase

In this phase, the sum of the pixel differences between the original image and the convex hull would be calculated by mapping the original image onto its convex hull. In the last step, using the following multipath function, images would be classified. In other words, since the inflamed prostate has a stellar shape and loses its spherical shape, the difference between the original image and its convex hull will be significant. Therefore, by counting this difference in the number of pixels, the images can be classified into two classes: one with inflammation and the other with no inflammation sign. All the details is mentioned in Tab. .

Tab. 6. Third phase classification function

Function Name	Function formula	Description
Counter function	$PS_i = Ch_i - Or_i$ (12)	Ch_i : The sum of the pixels of the original image Or_i : The sum of the pixels inside the convex cover space
Classification function	$C_i = \begin{cases} 0, & Ps \le 1200 \pm 50 \\ 1, & Ps \ge 1200 \pm 50 \end{cases}$ (13)	$C_i = 0$:No Prostatitis $C_i = 1$:Prostatitis

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

4. Results

Fig. 10 shows the error functions. The TensorBoard interface is used to illustrate the graphs. During the learning process, alteration in classification, localization, and segmentation loss functions was descending. The final amount of classification and localization loss was less than 1%, while the segmentation loss was about 7%. One of the most influential factors in segmentation loss function is the high similarity of the prostate gland to the surrounding tissue. Eventually, the overall loss ratio of the model was about 9%. Each epoch of the training process had 1000 iteration, and the least loss function was obtained in the 90th epoch. The horizontal axis in all of the diagrams in Fig. 10 is the number of epochs multiplied by 100.

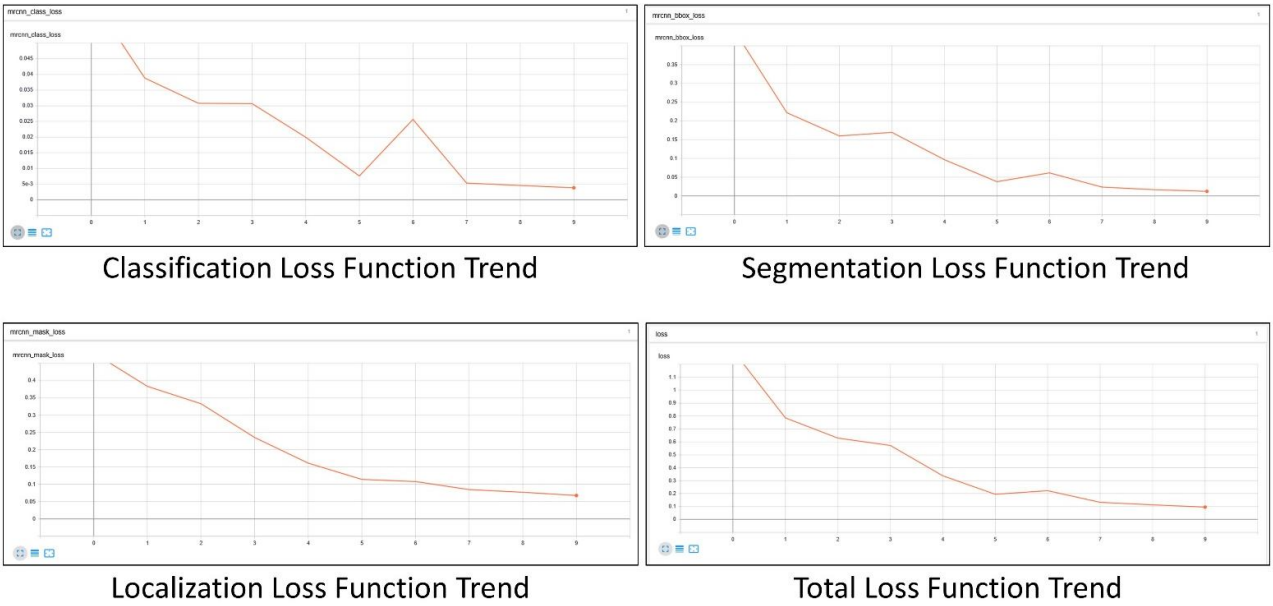


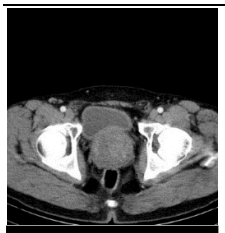
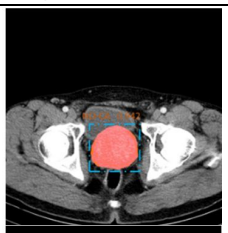
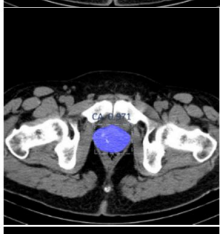
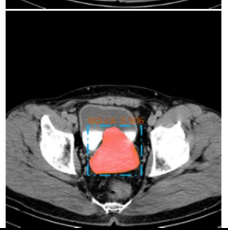
Fig. 10. Loss function trends

The final Prediction of the six sample images is summarized in Tab. 7, which $C_i, i = \{1,2,3\}$ stands for the predicted class in each phase, and P_c shows the classification accuracy in the first phase. The variables Width and Height, respectively, measured the prostate gland's height and width in the second phase. The variable P_s illustrates the difference in the number of convex hull pixels with the selected region.

Tab. 7. Final results for six sample images

Input picture	Image segmentation network output	Predicted classes in each phase	physician's opinion	Prediction & Class
		$\begin{cases} C_1 = 0 \\ P_C = 0.955 \end{cases}$ $\begin{cases} C_2 = 0 \\ Width = 37.94mm \\ Height = 33.35mm \end{cases}$ $\begin{cases} C_3 = 0 \\ Ps = 102 \end{cases}$	Without lesion	Without lesion (G1)

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

 	$\begin{cases} C_1 = 0 \\ P_C = 0.942 \end{cases}$ $\begin{cases} C_2 = 1 \\ Width = 61.5mm \\ Height = 66.7mm \end{cases}$ $\begin{cases} C_3 = 0 \\ Ps = 179 \end{cases}$	BPH	BPH (G3)
	$\begin{cases} C_1 = 1 \\ P_C = 0.971 \end{cases}$ $\begin{cases} C_2 = 1 \\ Width = 54.5mm \\ Height = 40.7mm \end{cases}$ $\begin{cases} C_3 = 0 \\ Ps = 88 \end{cases}$	Calcification	Calcification (G5)
	$\begin{cases} C_1 = 0 \\ P_C = 0.893 \end{cases}$ $\begin{cases} C_2 = 1 \\ Width = 85.71mm \\ Height = 56.26mm \end{cases}$ $\begin{cases} C_3 = 0 \\ Ps = 2644 \end{cases}$	Prostatitis & BPH	Prostatitis & BPH (G4)
	$\begin{cases} C_1 = 1 \\ P_C = 0.936 \end{cases}$ $\begin{cases} C_2 = 0 \\ Width = 43.18mm \\ Height = 37.94mm \end{cases}$ $\begin{cases} C_3 = 0 \\ Ps = 91 \end{cases}$	Calcification	Calcification (G5)
	$\begin{cases} C_1 = 0 \\ P_C = 0.906 \end{cases}$ $\begin{cases} C_2 = 1 \\ Width = 73.28mm \\ Height = 68.70mm \end{cases}$ $\begin{cases} C_3 = 0 \\ Ps = 567 \end{cases}$	BPH	BPH (G3)

Tab. 8 illustrates the absolute and relative frequency of each class. It is worth mentioning that the proportion of patient images was 27% of the whole dataset. More than that, almost 82% of images that were labeled as the patient, had only one lesion, and 18% of them consisted of two lesions.

Tab. 8. The frequency of each class

Class	Calcification	BPH	Prostatitis	Final Label	Absolute Frequency	Relative Frequency
G_1	No	No	No	healthy	٤٣٥	0.73
G_2	No	No	Yes	Patient	١٧	0.03
G_3	No	Yes	No	Patient	٥٦	0.09
G_4	No	Yes	Yes	Patient	٢٤	0.04
G_5	Yes	No	No	Patient	٥٣	0.09
G_6	Yes	No	Yes	Patient	٠	0.00
G_7	Yes	Yes	No	Patient	٤	0.01
G_8	Yes	Yes	Yes	Patient	٠	0.00

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

5. Discussion

Diagnostic imaging has become an indispensable procedure in medical science. Prostate segmentation is still a challenging issue, and with the progression of technology for the diagnosis and treatment of prostate lesions, new prerequisites must be considered. Automatic segmentation of the prostate gland has many applications in prostate disease therapy**.** Prostate cancer has been one of the most commonly investigated in recent years; however, the prostate's non-cancerous lesions also caused many problems for patients and can be a sign of prostate cancer.

Although MR images of the prostate have better soft-tissue contrast than CT images, an endorectal coil should be done to enhance contrast, which is painful and may have after-effects. A prostate MRI generally takes approximately 30–45 minutes, while the abdominal and pelvic CT scan takes about 10 to 15 minutes.

With regards to using a deep learning framework for prostate segmentation, Zhiyong Liu et al. recently has used mask-rcnn for prostate segmentation and cancer detection in ultrasound images [40]. Wang et al. segmented MRI prostate images for detecting dominant intraprostatic lesions [41]. They also used Dual Mask R-CNN in another research for prostate and tumor segmentation on CT images [42].

Contrary to previous research, which focused on prostate cancer and MRI, this paper presents an innovative method for classification and segmentation of the prostate gland in CT-scan images. Furthermore, our method can detect three prostate lesions simultaneously by considering the age of each person. We collaborated with physicians at one of Iran's most famous hospitals to collect data and annotate them by considering their professional reports to reach a reliable result. As nature and structural form of lesions are different, we proposed a stepwise approach in three phases for lesion detection.

In the first phase, a region convolutional neural network has been modified to segment the images. In the data preparation step, sections of the images that did not include the prostate gland were removed. This action helped the network recognize the prostate gland's location with more than 99% accuracy. As well in this phase, the diagnosis of calcification lesion in images was investigated. In the second and third phases, by considering each patient's age and the morphology features, the presence of BPH and prostatitis diseases was assessed. Finally, the patients' final status was compared with the physician's opinion to evaluate the model's final accuracy. The total loss function of this method was about 9%.

Based on the experiences and the results of implementing the proposed approach, some points worth to be considered in future research:

- Classification and segmentation of other parts of the abdominal and pelvic region to diagnose other diseases such as Bladder wall thickening.
- Using 3D imaging techniques for volumetric prostate reconstruction from 2D CT scans.
- Using deep generator networks to generate synthetic data due to data shortage in medical issues.

Declarations

Funding Not applicable

Conflicts of Interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Availability of data and material Not applicable

Code availability Not applicable

Ethics approval We are in compliance with ethical standards

Consent to participate Informed consent was obtained from legal guardians.

Consent for publication The participant has consented to the submission of the case report to the journal.

References

[1] He, B.-M., et al., *Prostate cancer risk prediction models in Eastern Asian populations: current status, racial difference, and future directions*. Asian Journal of Andrology, Vol. 22, No. 2, (2020), pp. 158-161.

[2] Zhang, S.-J., et al., *Relationship between age and prostate size*. Asian journal of andrology, Vol. 15, No. 1, (2013), pp. 116-120.

[3] Aslam, H.M., et al., *Spectrum of prostatic lesions*. International archives of medicine,. Vol. 6, No. 1, (2013), pp. 36-36.

[4] Kong, D.-P., et al., *Prevalence and clinical application of TMPRSS2-ERG fusion in Asian prostate cancer patients: a large-sample study in Chinese people and a systematic review*. Asian Journal of Andrology, Vol. 22, No. 2, (2020), pp. 200-207.

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

[5] Gordon, L.G., et al. *Estimating the healthcare costs of treating prostate cancer in Australia: A Markov modelling analysis*. in *Urologic Oncology: Seminars and Original Investigations*. (2018), Elsevier.

[6] Tang, Z., et al., *The prevalence and risk factors of prostatic calcification: an analysis of 68 705 subjects*. Asian journal of andrology, Vol. 20, No. 4, (2018), p. 417.

[7] Comelli, A., et al., *Deep Learning-Based Methods for Prostate Segmentation in Magnetic Resonance Imaging.* Applied Sciences,. Vol. 11, No. 2, (2021), p. 782.

[8] Ghose, S., et al., *A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images.* Computer methods and programs in biomedicine, Vol. 108, No. 1, (2012), pp. 262-287.

[9] Tang, X., et al. *Geometric-model-based segmentation of the prostate and surrounding structures for image-guided radiotherapy*. in *Visual Communications and Image Processing 2004*. International Society for Optics and Photonics, (2004).

[10] Feng, Q., et al., *Segmenting CT prostate images using population and patient□ specific statistics for radiotherapy.* Medical physics, Vol. 37, No. 8, (2010), pp. 4121-4132.

[11] Acosta, O., et al. *Evaluation of multi-atlas-based segmentation of CT scans in prostate cancer radiotherapy*. in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. (2011), IEEE.

[12] Acosta, O., et al., *Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy.* Radiotherapy and Oncology, Vol. 125, No. 3, (2017), pp. 492-499.

[13] Shi, J. and J. Malik, *Normalized cuts and image segmentation*. IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, No. 8, (2000), pp. 888-905.

[14] Liu, X., Z. Deng, and Y. Yang, *Recent progress in semantic image segmentation.* Artificial Intelligence Review, Vol. 52, No. 2, (2019), pp. 1089-1106.

[15] Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2015).

[16] Noh, H., S. Hong, and B. Han. *Learning deconvolution network for semantic segmentation*. in *Proceedings of the IEEE international conference on computer vision*. (2015).

[17] Zhao, H., et al. *Pyramid scene parsing network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017).

[18] Guo, Y., Y. Gao, and D. Shen, *Deformable MR prostate segmentation via deep feature learning and sparse patch matching.* IEEE transactions on medical imaging, Vol. 35, No. 4, (2015), pp. 1077-1089.

[19] Yan, K., et al. *Automatic prostate segmentation on MR images with deep network and graph model*. in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (2016), IEEE.

[20] Yan, K., et al. *Comprehensive autoencoder for prostate recognition on MR images*. in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. (2016), IEEE.

[21] Milletari, F., N. Navab, and S.-A. Ahmadi. *V-net: Fully convolutional neural networks for volumetric medical image segmentation.* in *2016 Fourth International Conference on 3D Vision (3DV)*. (2016), IEEE.

[22] Cheng, R., et al. *Automatic MR prostate segmentation by deep learning with*

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

holistically-nested networks. in *Medical Imaging 2017: Image Processing*. International Society for Optics and Photonics, (2017).

[23] Zagoruyko, S., et al., *A multipath network for object detection*. arXiv preprint arXiv:1604.02135, (2016).

[24] Esteva, A., et al., *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, Vol. 542, No. 7639, (2017), pp. 115-118.

[25] Lin, G., et al. *Refinenet: Multi-path refinement networks for high-resolution semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017).

[26] Redmon, J. and A. Farhadi, *Yolov3: An incremental improvement*. arXiv preprint arXiv:1804.02767, (2018).

[27] Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. (2015), Springer.

[28] He, K., et al. *Mask r-cnn*. in *Proceedings of the IEEE international conference on computer vision*. (2017).

[29] Yu, L., et al. *Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images*. in *Thirty-first AAAI conference on artificial intelligence*. (2017).

[30] Kazemifar, S., et al., *Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning*. Biomedical Physics & Engineering Express, Vol. 4, No. 5, (2018), p. 055003.

[31] Balagopal, A., et al., *Fully automated organ segmentation in male pelvic CT images*. Physics in Medicine & Biology, Vol. 63, No. 24, (2018), p. 245015.

[32] Polymeri, E., et al., *Deep learning-based quantification of PET/CT prostate gland uptake: association with overall survival*. Clinical Physiology and Functional Imaging, Vol. 40, No. 2, (2020), pp. 106-113.

[33] Ma, L., et al. *Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion*. in *Medical Imaging 2017: Image Processing*. International Society for Optics and Photonics, (2017).

[34] Litjens, G., et al., *A survey on deep learning in medical image analysis*. Medical image analysis, Vol. 42, (2017), pp. 60-88.

[35] Kisilev, P., et al., *Medical image description using multi-task-loss CNN*, in *Deep Learning and Data Labeling for Medical Applications*. Springer. (2016), pp. 121-129.

[36] Ghavami, N., et al., *Automatic segmentation of prostate MRI using convolutional neural networks: Investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration*. Medical image analysis, Vol. 58, (2019), p. 101558.

[37] Varma, D.R., *Managing DICOM images: Tips and tricks for the radiologist*. The Indian journal of radiology & imaging, Vol. 22, No. 1, (2012), p. 4.

[38] Yang, C., M. Dong, and J. Hua. *Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning*. in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, (2006), IEEE.

[39] Wang, Y., et al., *Combining global, regional and contextual features for automatic image annotation*. Pattern Recognition, Vol. 42, No. 2, (2009), pp. 259-266.

[40] Liu, Z., et al., *Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-*

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]

aided diagnosis of prostate cancer. Future Generation Computer Systems, Vol. 114: (2021), pp. 358-367.

[41] Wang, T., et al. *MRI-based prostate and dominant lesion segmentation using deep neural network*. in *Medical Imaging 2021: Computer-Aided Diagnosis*. International Society for Optics and Photonics, (2021).

[42] Wang, T., et al., *Prostate and tumor segmentation on PET/CT using Dual Mask R-CNN*. SPIE Medical Imaging. Vol. 11600. (2021), SPIE.

Follow This Article at The Following Site:

Masoumi A, Ghousi R, Makui A. Prostate Segmentation and Lesions Classification In CT Images Using Mask R-CNN. IJIEPR. 2022; 33 (3) :1-16
URL: http://ijiepr.iust.ac.ir/article-1-1302-en.html

[Downloaded from ijiepr.iust.ac.ir on 2026-11-10]