Convolutional Neural Networks for Prostate Magnetic Resonance Image Segmentation

TAHEREH HASSANZADEH [1], **LEONARD G. C. HAMEY** [1], **(Member, IEEE), AND KEVIN HO-SHON** [2]

[1]Department of Computing, Macquarie University, Sydney, NSW 2109, Australia
[2]Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW 2109, Australia

Corresponding author: Tahereh Hassanzadeh (tahereh.hassan-zadeh-koohi@hdr.mq.edu.au)

ABSTRACT One of the most accurate and non-invasive prostate imaging methods is magnetic resonance imaging (MRI). Segmentation is needed to find the boundary of the prostate, either automatically or semi-automatically. Recently, fully convolutional neural networks (FCNN) are being used for this purpose. In this paper, to improve the FCNN performance for prostate MRI segmentation, we analyze various structures of shortcut connections together with the size of a deep network and suggest eight different FCNNs-based deep 2D network structures for automatic MRI prostate segmentation. Our evaluations on the PROMISE12 dataset with ten-fold cross-validation indicate improved and competitive results. We analyze the results in detail, considering MRI slices, MRI volumes, test folds, and also the impact on prostate segmentation of using an EndoRectal Coil to capture the prostate MRI. Our best 2D network outperforms the state-of-the-art 3D FCNN-based methods for prostate MRI segmentation on publicly available data, without any further post-processing.

INDEX TERMS Automatic MRI segmentation, fully convolutional neural network, prostate MRI segmentation.

I. INTRODUCTION

The process of segmenting an image into several discrete and homogeneous regions is image segmentation. It aims to alter the representation of an image to find a region or regions of interest in an image. Medical image segmentation is one of the most significant and active areas of research in medical image processing. The purpose of medical image segmentation is using a precise method to find the boundary of a specific organ or tissue, and it is a fundamental step for clinical studies including: diagnosis of disease, monitoring of organs or particular tissues, and, more importantly, treatment planning. Medical image segmentation is a difficult task because in most cases a specific organ has different shapes and sizes in different people [1]. Also, in some studies, the intensity value of the Region of Interest (ROI) is the same as the adjacent organs that can make segmentation even more challenging [2].

Medical image segmentation is usually done in one of three ways: manually, semi-automatically, or automatically.

The associate editor coordinating the review of this manuscript and approving it for publication was Bohui Wang.

An expert radiologist can perform a manual segmentation of the ROI, but it is often time-consuming and tedious [3]. A further problem is that in some cases a radiologist may segment a specific image differently at various times or two radiologists may segment the same image dissimilarly [4]. However, when developing a semi-automatic or automatic segmentation method we almost always need ground truth images that should be created manually by expert radiologists. Even in the semi-automatic segmentation methods, an expert user is required to initialize or correct the segmentation. For example, the user can set a seed point or specify a region to start segmentation. In fully automatic segmentation, there is no human interaction during the segmentation of the image. In this type of segmentation, human knowledge is often employed to design an accurate method based on image processing and/or machine learning methods for image segmentation.

The prostate is a part of the male reproductive system that can suffer from many diseases but most importantly prostate cancer. Prostate cancer is one of the more common causes of death in developed countries [5]. According to Siegel *et al.* [6] there will be 164,690 new instances of

prostate cancer and 29,430 deaths because of prostate cancer in the United States in 2018.

In Australia, according to the Prostate Cancer Foundation of Australia more than 3,000 men die because of prostate cancer annually–more than the number of women death due to the breast cancer [7]. The Australian Institute of Health and Welfare estimates that 17,729 new prostate cancer cases will be diagnosed in 2018, which would be 23.8% of all new male cancer cases in Australia. Also, they estimate that 3,500 men will die from prostate cancer, representing 12.7% of all cancer deaths in Australia [8], [9]. In 2017, prostate cancer was the third cause of cancer in Australia after Breast and Colorectal cancers and also the third cause of death after Lung and Colorectal cancer [8], [9].

Finding the boundary of the prostate in the MRI image is fundamental to recognize possible disease in the prostate. Because of the considerable diversity in size, appearance, shape, and texture of the prostate and the lack of a clear prostate boundary, especially in malignant prostate tissues, prostate segmentation is a challenging problem even for expert radiologists [10], [11]. However, using EndoRectal Coil (ERC) during the MRI imaging process can help radiologist to observe the prostate tissue more accurately [12]. This device is placed into the rectum to obtain high-quality images during $1.5T$ MRI imaging, but, it creates spikes and bright regions in the MRI that can decrease the accuracy of automatic prostate image segmentation.

A Fully Convolutional Neural Network (FCNN) is a type of Convolutional Neural Network (CNN) that has been introduced for image segmentation [13]. The purpose of the FCNN is to create an output image analogous to the ground truth of the input image. U-Net [14] and DenseNet [15] are two FCNN-based networks for medical and natural image segmentation respectively. In this paper, we develop new segmentation methods for automatic 2D MRI prostate image segmentation based on U-Net and DenseNet structures. We suggest eight different structures with a particular focus on using various patterns of shortcut connections [16] as well as varying the size of the networks. We evaluate the performance of the following network structures: Straight, Bypass, Non-bypass Output from All, Output from All, Non-bypass Input to All, Input to All, Non-bypass Dense, and Dense models. After extensive experiments, and analyzing the results in detail, our best model (Non-bypass Dense) is found to outperform the state-of-the-art 3D FCNN-based prostate segmentation methods. Our results demonstrate that, shortcut connections may improve the results but can also have negative impact. Therefore, the starting and ending points of the shortcut connections are also critical. In addition, the results show that the quality of the training images has a significant effect on the final results.

II. BACKGROUND

Improvement in machine learning methods especially deep learning has convinced researchers to use deep learning in computer vision applications [17], [18]. A deep neural network is constructed from multiple layers of neurons such that each layer learns to transform its input data into a new more abstract. In particular, a Convolutional Neural Network (CNN) [19] is a kind of deep network that has been successfully applied for visual image processing. The leading operator of CNN is convolution, consisting of learnable filters or kernels that are convolved across the input image, computing the dot products between the filters and the receptive fields to produce feature maps [20]. In this section, we review papers that apply the combination of the CNN and other methods for MRI segmentation.

A successful hybrid method for prostate image segmentation uses an atlas as well as deep learning. Cheng *et al.* [21] proposed a hybrid method combining an atlas-based active appearance method along with a deep learning method to improve 3D MRI prostate image segmentation. They apply AAM for estimating the prostate boundary then use deep CNN to refine the boundary. In the first phase, they separate the atlas into various groups based on a similarity measure. Each image slice is assigned to the most similar atlas group and they then employ AAM training in each subgroup to find a boundary around the prostate. In the second phase, they extract 2D 64×64 image patches around the AAM predicted boundary. They use pre-trained AlexNet [22] to classify the patches into prostate and non-prostate to refine the boundary.

Cheng *et al.* [23] presented another work using both patch-based and holistic (image to image) deep learning methods for prostate image segmentation. In this paper, they employ a Holistically Nested Network (HNN) architecture for image-based segmentation. For training their network they first crop 25% of images from top, bottom, left, and right to find the prostate area. Then the Coherence Enhanced Diffusion (CED) filter is used to enhance the quality of the prostate boundaries. In the end, both the original MRI images and the CED-MRI images with their corresponding labels are used for training the HNN for the prostate image segmentation.

Recently Jia *et al.* [24] proposed a course-to-fine segmentation method using an atlas method and deep learning. In this paper, a registration-based segmentation is used to find the approximate boundary of the prostate; then they extract image patches around the prostate region to find the prostate boundary by applying deep network VGG-19 [25] and LeNet-5 [26]. In the paper, they fine-tune pre-trained VGG-19 for finding the prostate boundary. Also, to show the efficiency of utilizing pre-trained networks they train LeNet-5 from scratch using the extracted image patches. In the end, they conclude that using the pre-trained network is more precise than their separately trained network for prostate image segmentation. In related work, He *et al.* [27] proposed a three-level coarse-to-fine segmentation method. In the first level, the 3D volume of interest is extracted by employing 3D Haar features then an Adaptive Feature Learning Probability Boosting Tree (AFL-PBT) voxel classifier is used to classify pixels into three groups: near, interior, and exterior. Finally, CNN is used to refine the prostate boundary.

Proposal-based segmentation is another well-known model for natural image segmentation that has been applied for prostate image segmentation [28]. In this approach, an image is divided into several patches or proposals then the proposals that contain prostate are separated. For example, Yan *et al.* [28], first generate a set of prostate proposals by using the Geodesic Object Proposal (GOP) algorithm [29] for 3D segmentation of the prostates then a graph is used to select highly effective proposals. Finally, CNN is employed to detect highly effective features to refine the boundaries. Two other types of networks that apply for prostate image segmentation are Stacked Sparse Auto Encoder (SSAE) [11] and Independent Subspace Analysis (ISA) networks [30]. The SSAE uses a sparse patch matching method, and the ISA employs sparse label propagation method for feature extraction for prostate image segmentation.

Almost all of the papers discussed above use a combination of various image processing and machine learning methods for feature extraction, coarse segmentation and fine segmentation. However, recently some researchers employ only CNN for both feature detection and segmentation. A Fully Convolutional Neural Network (FCNN) is a version of CNN that is designed for image segmentation [13]. FCNN is constructed from two parts: down-sampling (encoding, convolution) and up-sampling (decoding, deconvolution). Each part includes blocks that contain the number of convolution layers and other possible operators such as Batch Normalization [31] and Dropout [32]. In the down-sampling part there is a pooling layer after each block to decrease the dimension of the feature maps [33] and in the up-sampling section there is a deconvolution [34] layer before the blocks to increase the size of the feature maps. In some networks, these parts are connected by a Bottleneck (Bridging) block. In the down-sampling section, the network tries to extract features as it goes from the higher resolution to lower resolution while the up-sampling part attempts to reconstruct the coarse-to-fine segmentation with transposed convolution [34]. FCNN utilizes an end-to-end (image-to-image or volume-to-volume) method for learning.

One of the first studies that used FCNN with 2D convolution for 2D medical semantic image segmentation is U-Net [14]. U-Net is constructed from three parts: the down-sampling and up-sampling sections such that each part includes four blocks, and a bottleneck block that connects the two other parts. All the blocks are constructed from two convolution layers followed by non-linearity. Finally, a 1×1 convolution is used in the last layer to output the segmentation. To improve the results in this structure, long connections are used for cropping and copying a part of the extracted feature maps from the down-sampling part and concatenating them with the obtained feature maps from the up-sampling section.

Also, a revised version of U-Net for prostate MRI image segmentation is introduced by Zhu *et al.* [10]. In this paper, they try to improve the accuracy of the network by adding a 1×1 convolution layer in each block. The 1×1 convolutions

make the network deeper as well as decreased the computational difficulty. Additionally, this paper uses dropout to overcome overfitting [35].

In 2016, the first version of 3D FCNN for segmentation of 3D volumes of prostate images was presented as V-Net [36]. In this network, there are four blocks in each part similar to U-Net, but the number of layers in each block is different. In each block, there is a residual shortcut connection for summing the input feature maps and the output feature maps of the block, element by element. Also, four long connections are used to concatenate the feature maps from the down-sampling to the up-sampling part, as in U-Net.

Another method for 3D MRI prostate segmentation is ConvNet [37]. The aim of this work is to analyze the effect of using short and long residual connections. In this network, each residual block is constructed from two convolution layers with kernel size 3×3. They use the short residual connection to sum the input of the block with the output of the second convolution layer element by element before applying the non-linearity. Moreover, they utilize long connections to sum the extracted feature maps from the down-sampling part of the network to the up-sampling part. They compare the network as variants with only short or long and with both residual connections. They show that using the combination of short and long residual connections is more effective for prostate image segmentation.

Finally, Mun *et al.* [38] proposed another FCNN-based network for 3D MRI prostate segmentation call the Baseline Convolutional Neural Network (BCNN). All the blocks contain three layers of convolution, but in the encoding blocks, there is a shortcut connection to sum the output of the first layer with the output of the second layer. A corresponding connection does not exist in the decoding part. Also, to reuse the extracted feature maps of the down-sampling part in the up-sampling part, they utilize long connections and element-wise sum. The primary purpose of the paper is the comparison of six different types of objective functions: the Jaccard Index, Hamming Distance, Euclidean Distance, Cosine Similarity, Dice Coefficient, and Cross Entropy. They show that Cosine Similarity is the best and the Dice Coefficient is the second best among the six objective functions to train the network for prostate image segmentation.

We use U-Net and DenseNet [15] to develop new network structures for prostate image segmentation. The Dense model was introduced by Huang *et al.* [15] for natural image classification and obtained promising results. In the Dense model, after each layer, there is a concatenation operator that can concatenate the output of all previous layers of the block. The fully convolutional Dense model was applied for color image segmentation by Jegou *et al.* [39] and obtained positive results for natural image segmentation.

III. PROPOSED MODELS
In this section, firstly, we introduce our FCNN-based network structure and then explain our proposed block structures for MRI prostate segmentation.
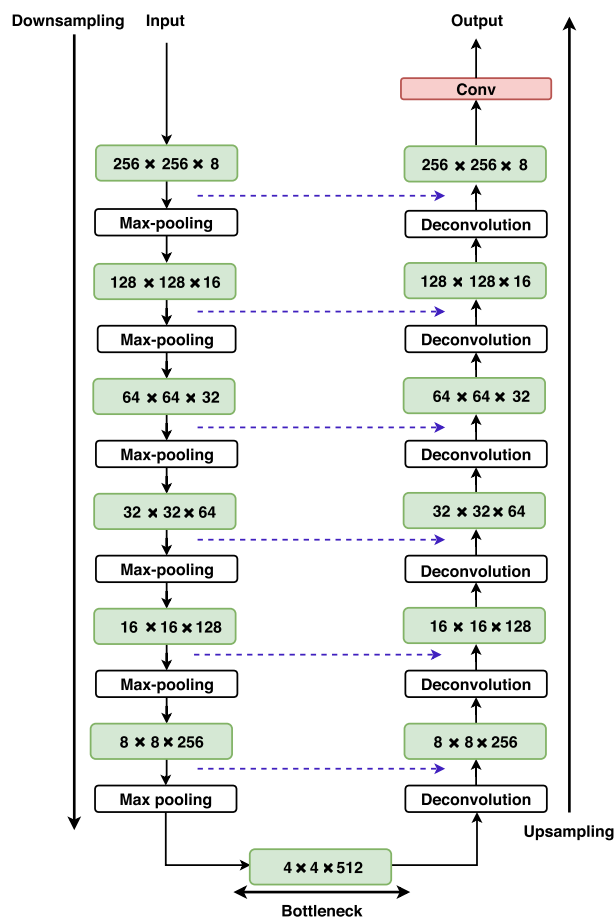
FIGURE 1. Proposed network architecture for prostate segmentation.

A. PROPOSED ARCHITECTURE

We suggest a relatively deep FCNN network structure based on U-Net [14]. The diagram of this architecture is shown in Figure 1. The network is constructed from three parts including down-sampling, bottleneck, and up-sampling. Six blocks can be seen in the down-sampling part and six blocks in the up-sampling section. In the bottleneck, there is another block to connect the two parts.

As shown in Figure 1, in the down-sampling part, the input image is directed into the first block, and the output feature maps (convolution output [20]) of the first block are fed to the next block as the input. This process is repeated several times. After each block in the down-sampling part, a max-pooling operator [33] halves the size of the feature maps. In the bottleneck, there is a block that connects the down-sampling section to the up-sampling part. In the up-sampling section, each block is preceded by a deconvolution layer (3 × 3 convolution with stride 2) to double the dimension of the feature maps. In Figure 1, the dimension is shown in each block are the resolution of the feature maps (the first two numbers) along with the output rate that is the number of feature maps output by each layer. In the down-sampling part the resolution decreases after each max-pooling operator to extract information and in the up-sampling section, the resolution increases to the original image size. The last layer is a 1 × 1 convolution layer with one output channel that produces the output segmentation image from the features maps output by the last block. The dimension of the input image and the output segmentation image is the same (256 × 256). Because of using multiple convolution layers and max-pooling in the down-sampling process a part of the spatial information is lost [16], [37]. Therefore, feature maps in the up-sampling part will have more information deficiency. To improve the quality of feature maps in the up-sampling section the extracted data from the down-sampling part can be reused by using *long* connections [40]. Our proposed architecture uses six long connections (shown with dash lines see Figure 1) to copy extracted feature maps to the up-sampling section to concatenate with feature maps from deconvolution.

In the following, we first explain the common operators that we utilize in each layer of the blocks including dropout, batch normalization, activation function, and the kernel size. We then describe each model separately based on the shortcut connections structure.

B. COMMON COMPONENTS

Since we have a deep network and our training images are limited, we should employ some approaches to control overfitting including dropout [32] and batch normalization [31] to create sparsity in our proposed models that can accelerate the training and improve the accuracy of the network.

Dropout can be an efficient method to exclude the complication of co-adaptations on the training data and also functions as a regularizer. Dropout randomly deletes a portion of the features by omitting hidden layers units with a specified probability [32].

Another feature in our proposed models is Batch Normalization (BN) [31] for data normalization during the training of the network. Data normalization is one of the most critical parts of training a network. It is common to normalize input data before training the network, but after applying the convolution operator and non-linearity, the distribution of the data will be changed. The purpose of BN is normalizing the output of network layers during training, and it is known that this normalization can accelerate the training of the network [31]. According to Ioffe and Szegedy [31], to calculate BN, each mini-batch should be normalized to zero mean and unit variance. BN starts with zero mean and unit variance normalization, but during the training, can learn other parameters that might be better for normalization.

One of the reasons for using BN is reducing the covariance shift [31] i.e. the changing distribution of the test data versus the training data [41]. If a network is trained with X as the input images and Y as the corresponding labels with a particular distribution, the network could learn the distribution of the training samples. If the network tested with new images from a different distribution, the results can be very poor. Moreover, internal covariance shift can happen during the backpropagation [31]. If the parameters of the first layer of
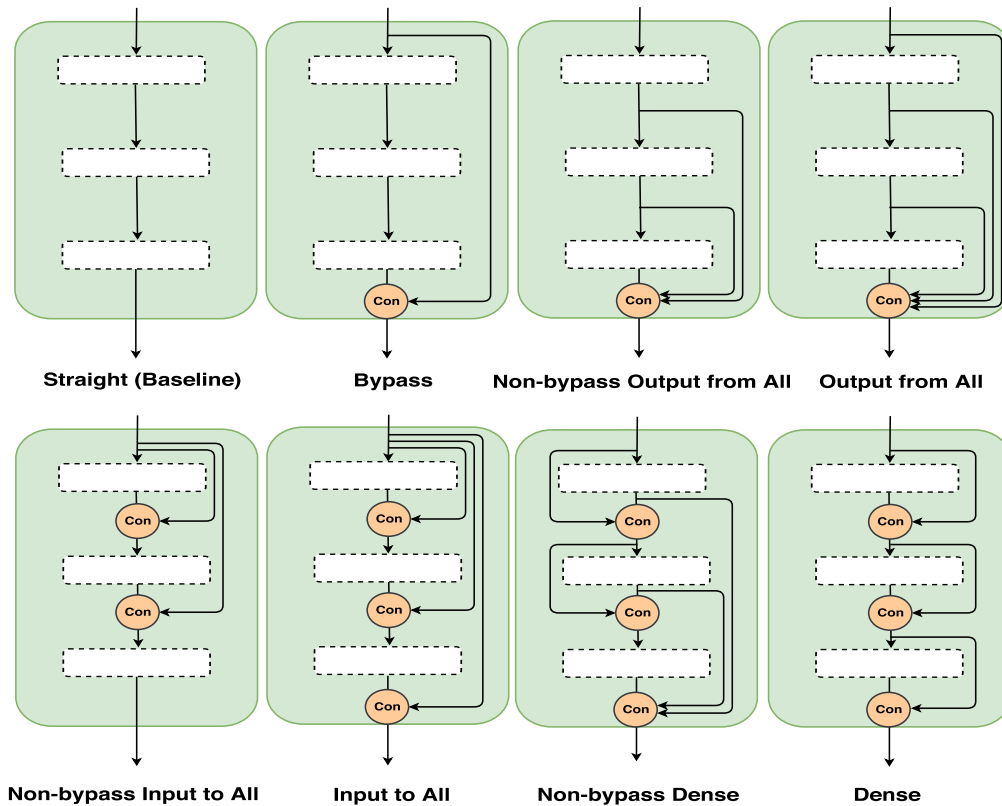
FIGURE 2. Proposed blocks architecture for prostate segmentation.

the network change, it could be the change the distribution of the second layer, and consequently, it changes the layer outputs as well [31]. Deep networks, such as our proposed models are more prone to the problem.

It is also known that BN can work as a regularizer similar to dropout [31]. Therefore, if BN is used in the network, dropout value selection should be made more carefully otherwise more information will be lost [42]. However, dropout cannot be replaced with BN because the effect of the BN on overfitting is less than dropout; therefore, a proper solution is using both of them simultaneously as we are doing in our proposed models.

As the activation function, our proposed models employ Rectified Linear Unit (ReLU) [43] for all layers within the blocks. ReLU was first used for training of deep networks in 2011 [44]. Activation function defines the output of each unit with regards to its input, and the ReLU defines the positive part of its argument ($f(x) = \max(0, x)$). ReLu can decrease the probability of the vanishing gradient [45] during the back propagation [44].

In the last layer of our proposed network, we apply the Sigmoid function [46] as the non-linearity. The sigmoid commonly employed for two-class classification. The equation of this function can be seen in the Equation 1.

$$f(x_i) = \frac{1}{1 + \exp^{-x_i}} \tag{1}$$

The convolution layers of our networks employ small 3×3 convolutional filters in the convolution layers. Using a stack of very small 3×3 receptive fields is more efficient than using bigger receptive fields like 5×5 or 7×7 [25]. Utilizing the small kernel the network will apply more non-linear layers, and it can decrease the number of model parameters [25].

In summary, in the proposed networks we use 3×3 kernel for convolution layers, apply ReLU as the activation function in the hidden layers after each convolution layer, and utilize BN after activation function for all six models. Dropout is also used to control overfitting after the last BN of the each block. In the next section, we discuss the key issue of shortcut connections.

C. PROPOSED BLOCKS

An issue that can be very significant in the final segmentation results is the architecture of the blocks. We proposed eight different structures for the blocks. Figure 2 shows the eight block models and they would appear with three layers and each layer includes a convolution layer with kernel size 3×3 follow by a Rectified Linear Unit (ReLU) activation function [43]. To improve the generalization of the network, batch normalization [31] and dropout [32] are also employed. In all proposed block structures, after each convolution layer, there is a batch normalization. Also, the dropout is only used after the last batch normalization of each block. As shown in Figure 2, we apply various structures of the shortcut

connections to investigate their effects on the prostate image segmentation. In our work, we use concatenation for both the shortcut and long connections in all of our proposed networks. Concatenation stacks the feature maps on top of each other.

As can be seen from Figure 2, our simplest model Straight is a baseline with no shortcut connections inside the blocks.

In the Bypass model, we employ a concatenation operator at the end of the block to combine the block's input feature maps with its output. This *bypass* connection directly provides the input of the current block to the next block.

In the Non-bypass Output from All model, we add a concatenation operator at the end of the block to collect the output feature maps from the all layers in the block (*gathering* connections) and feed them to the next block. Using this model, we can explore whether using feature maps with different levels of information increases the learning ability of the network. Also, in the Output from All model, the bypass connection is added to again forward the block's input to the next block.

In the Non-bypass Input to All model, the block's input feature maps are provided to each of the layers (*scatter* connections). We concatenate the output of each layer with the block's input before sending it to the next layer. In the Input to All model, the input is also concatenated with the final output. In other words, this model uses both bypass and scatter connections.

Our Dense model follows Huang *et al.* [15] Dense model structure. In the Dense model, after each layer there is a concatenation of the output of all previous layers with the original input of the block. In our work, we use the fully dense block for MRI prostate segmentation. Our implementation uses chain connections as shown in Figure 2 to implement the Full Dense block to deliver the same effect while decreasing the number of shortcut connections.

Lastly, the Non-bypass Dense model is equivalent to the Dense model except that there is no connection between the input and the output of the block. Again, chain connections are use in this network structure to connect the layers.

IV. EXPERIMENTS

A. DATASET AND PREPROCESSING

The PROMISE12 challenge dataset [47] is used for studying MRI prostate segmentation. The data was collected from four different hospitals, with two employing the EndoRectal Coil (ERC). The dataset includes 50 MRI volumes and their corresponding labels for training, and also 30 MRI volumes without ground truth images for testing. For the evaluation of our proposed networks, we apply ten-fold cross-validation on the 50 training volumes. For each cross-validation fold, the training data is separated into three categories: train, validation and test sets. The 50 volumes have 1377 image slices. Because of the limited available data, using data augmentation to increase the number of images is necessary. We use the combination of transformations: random rotation within a 10-degree range, horizontal flip, vertical flip, zoom,

horizontal and vertical translation, and elastic transformation [48] for augmenting the number of training data to 150000 slices. Also, each pixel is normalized to zero mean and unit variance method (Z-score) [49]. We calculate a mean and variance for each pixel position across all 1377 images. Then all images normalized using obtain parameters.

B. IMPLEMENTATION

Our proposed models are implemented based on the Keras Python library [50]. All experiments were conducted on the National Computational Infrastructure (NCI) using four Nvidia GPUs along with CPUs. To find appropriate optimizer and learning rate, firstly, we use ADAM with the learning rate of 0.01, 0.001, and 0.0001, as well as the SGD with the same learning rates including, 0.01, 0.001, and 0.0001, also, we set momentum as 0.9 and the weight decay as 1e-6. The preliminary experiments show that using ADAM optimizer with a learning rate of 0.001 is the best choice for our proposed models. Furthermore, to determine the appropriate batch size, we test four different sizes including; 8, 16, 32, and 64. We found that using 32 as the batch size yields better performance; therefore in all of the tests, we set batch size as 32 and training continued for 25 epochs (3.25 million image presentations). For the evaluation of our proposed models, we utilize the Dice Coefficient [52] (see Equation 2). Consider X as the predicted segmented image (set of pixels), $|X|$ as the cardinality of X, Y as the ground truth segmentation image, and $|Y|$ as the cardinality of Y.

$$DSC = \frac{2 \mid X \cap Y \mid}{\mid X \mid + \mid Y \mid} \quad (2)$$

Correspondingly, the continuous Dice Coefficient (cDSC) [53], [54] is employed as the loss function for the training of all networks (see Equation 3).

$$cDSC = -\frac{2 \sum (X_i Y_i) + 1.0}{\sum (X_i) + \sum (Y_i) + 1.0} \quad (3)$$

C. THE SIZE OF THE NETWORKS

The number of layers in each block is an important parameter of the network design. To find the appropriate block size for each proposed structure, we tested all structures with blocks ranging from two to five layers, heuristically expanding to larger block sizes (up to 9 layers) depending on the relative performance. The best block sizes discovered for our structures are as follows: two layers per block for the Input to All and Non-bypass Input to All models; three layers per block for the Straight, Bypass and Dense models; and seven layers per block for the Output from All, Non-bypass Output from All and Non-Bypass Dense models.

D. DROPOUT

Dropout is another important hyper-parameter of the CNN to combat overfitting and increasing the generalization of the network. The probability of dropout and its location(s) in the network can affect the network training capability considerably. To analyze the location and probability of dropout,

TABLE 1. Quantitative comparison of proposed models.

Method	Mean DSC	Median DSC
Straight	0.853	0.859
Bypass	0.858	0.863
Non–bypass Output from All	0.849	0.861
Output from All	0.865	0.88
Non–bypass Input to All	0.815	0.82
Input to All	0.819	0.846
Non–bypass Dense	**0.873**	**0.88**
Dense	0.834	0.852

TABLE 2. Statistical comparison of Non-bypass Dense model with other models using the Wilcoxon signed rank test.

Method	$p < 0.05$
Straight	0.0137
Bypass	0.0137
Non–bypass Output from All	0.0039
Output from All	0.0645
Non–bypass Input to All	0.0039
Input to All	0.0020
Dense	0.0039

we propose three different dropout configurations. In the first configuration, we locate the dropout after each layer, with the probability of 0.2. In the second configuration, we only employ one dropout after the last layer of each block, again with the probability of 0.2. The third configuration only uses dropout after each layer in the bottleneck block, with the probability of 0.5. In the preliminary experiments, to find the appropriate probability for each of the dropout configurations we examined 0.2, 0.5, and 0.8. Concerning the obtained results, the discussed dropout probabilities selected for each of the three configurations. Also, we analyze the proposed models without using dropout as well.

According to the obtained results, we found that the second strategy is more appropriate for our proposed models. It seems that using dropout after each layer can delete many units along with their connections and it can be the cause of losing more information. On the other hand, using dropout only in the bottleneck is not enough to increase the generalization of the network. For the main experiments, we use dropout at the end of each block with the probability of 0.2.

E. COMPARISON OF PROPOSED MODELS

Based on the results obtained (see Table 1), using ten-fold cross-validation, the Non-bypass Input to All model is the worst, and Non-bypass Dense model is the best model among all eight proposed networks for prostate MRI segmentation. The Non-bypass Dense model outperforms all other models with 0.873 mean DSC.

We use, the Wilcoxon signed rank test [55] to show the statistically significant differences among the proposed approaches. We compare the Non-bypass Dense model with all other models (see Table 2), and the statistical comparison of the mean DSC over all ten folds shows that the improvement of the Non-bypass Dense model is statistically significant ($p < 0.05$) in comparison with all other proposed networks except the Output from All model. The Output from All model is our the second-best proposed network.

The results of the Straight model show that it is possible for a network with no shortcut connections to outperform networks using different patterns of shortcut connections. For the Bypass model, the results demonstrate that using the bypass connection does not have a significant effect on the final segmentation and improve the result infinitesimally compare with the Straight model. In the Output from All model, the average mean DSC increased to 0.865 using bypass and gathering connections. It shows that using bypass along with gathering connections is a beneficial combination. However, for the Input to All model, the results indicate reduced segmentation performance. This result shows that the combination of the scatter and bypass connections is not an effective composition.

In the Dense model, each layer employs all possible features including the input feature maps of the block as well as the output of all previous layers in the block. However, this model could not even compete with the Straight model. Finally, the Non-bypass Dense model by omitting input to output connection and decreasing the number of feature maps in comparison with the Dense model improved the results to 0.873 which is the best result.

The results demonstrate that using a bypass connection in most cases except for the Output from All model does not have a significant effect and omitting it from the Dense network improved the results significantly.

As can be seen from Figure 2, the models that placed in the same column transfer an equal number of feature maps between blocks. In most of these pairs of corresponding networks, better performance was observed for the networks with the fewer internal shortcut connections. For example, the Straight model achieves better result compare with the Non-bypass Input to All model. The exception is that the Non-bypass Dense model outperforms the Non-bypass Output from All model.

Also, all the models in the first row (see Figure 2) use the same internal connections-their shortcuts connect only to the output. Among these networks the best results are obtained for the Output from All model which forward all computed and the input features to the next block. The results demonstrated that finding a precise structure is a trade-off between internal and external block structures.

Figure 3 shows histograms of the DSC performance across all test images, combining results from all ten folds. It compares the best (Non-bypass Dense) and worst (Non-bypass Input to All) models. As seen, some segmented images have low mean DSC between 0 to 0.1 in both networks. These images primarily contain small prostate regions that cannot be appropriately segmented by our proposed 2D networks. The worst model missed more small prostate regions.

Also, to show the performance of the best and second-best proposed models for the segmentation of the prostate with
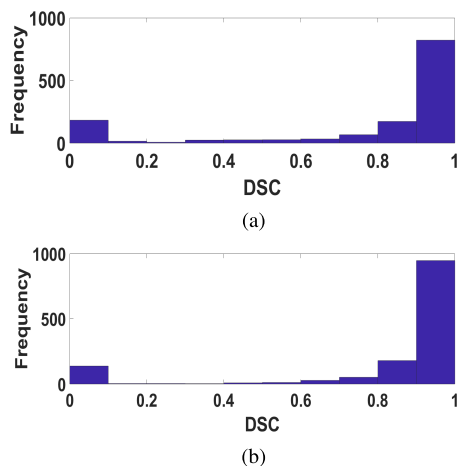
FIGURE 3. Comparison of the Non-bypass Input to All and Non-Bypass Dense models based on segmentation of all MRI slices. (a) The Non-bypass Input to All model. (The worst model). (b) The Non-bypass Dense model. (The best model).
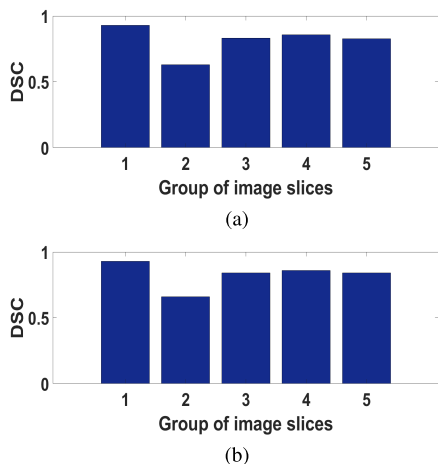

FIGURE 4. Comparison of the Output from All and Non-bypass Dense models based on the size of the prostate. (a) The Output from All model. (The second best model). (b) The Non-bypass Dense model. (The best model).

TABLE 3. The effect of number of layers on the final result of the Non-bypass Dense model.

Number of layers	Mean DSC
2	0.803
3	0.816
4	0.796
5	0.823
6	0.803
7	**0.873**
8	0.841
9	0.852

results than the Output from All model for segmentation of the image slices with small prostate.

The majority of segmentation results have a mean DSC between 0.8 and 1.0. Based on this range, the best model clearly achieves more high quality segmentation while the worst model has a significant number of segmentation with DSC between 0.5 and 0.8. The performance differences between the models reflect both differences in segmentation quality and missed smaller prostate regions.

In addition, to show the effect of the size of the network on the final results, the results of the Non-bypass Dense model as our best proposed model, using two to nine layers per block present in Table 3 as an example. As can be seen, using the different number of layers change the results considerably. The Non-bypass Dense model using seven layers per block obtain the better results.

F. ANALYSIS OF DATA FOLDS

We use the slices of the five MRI volumes for the test, five MRI volumes for the validation and the remaining 40 volumes for the training of the suggested models in each fold. The performance results show that the first, second, and the fifth folds are the most difficult folds for segmentation, such that in some cases their results change the ranking of the proposed models. However, for other folds, the segmentation results using different methods are much more similar that mainly contain the high-resolution test images.

The test set of the first fold contains five MRI volumes of which four were captured using the ERC and have bright regions around the ERC similar to Figures 5a and 5b. Also, some of the images have low contrast and wrap-around artifacts (see Figures 5a and 5b) that makes the prostate segmentation even more challenging. In the second fold, all the images in the test set were captured using the ERC have poor contrast resolution (everything is dark or bright), low spatial resolution (fuzzy images-see Figure 5c), and contain wrap-around artifacts in most of the images that make this fold the most challenging fold for segmentation. Furthermore, in the entire dataset, there is only one MRI volume that contains a very large prostate (see Figure 5d)-that volume appears in the fifth fold test set. Since the network for this fold has never seen such a large prostate during training, most of the models could not segment this prostate precisely despite the clear and
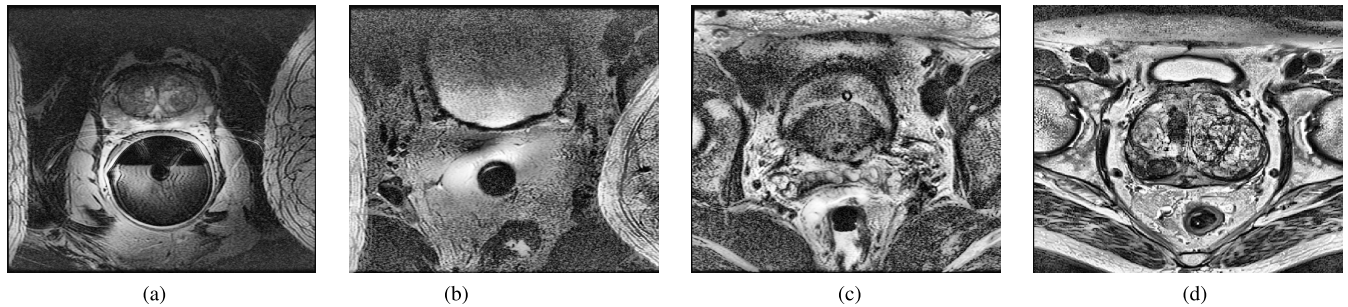
the different sizes we provide figure 4. Firstly, we analyze the size of the prostate on all 1377 image slices based on the number of pixels of the prostate. Among 1377 image slices only 788 image slices contain prostate. The smallest prostate has 418 pixels, the largest has 4625 pixels, and in average the prostate contains 2315 pixels in the dataset. We divide the images into five groups, the first group contain images without prostate (599 images), the second group contains images with 418 to 1366 prostate pixels (109 images), the third group contains images with 1367 to 2315 prostate pixels (242 images), the fourth group contains the images with 2316 to 3470 prostate pixels (352 images), and finally the fifth group contains the images that have 3471 to 4625 prostate pixels (75 images). As can be seen from figure 4, images with small prostate have lower DSC than other images. The Non-bypass proposed model obtain better

FIGURE 5. Four sample images to show the quality of images. (a) ERC+Artefacts. (b) ERC+Low res+Artefacts. (c) Fuzzy image. (d) Big prostate.

high-resolution image data. In contrast, the sixth fold yields the best segmentation performance because the test images are non-ERC and have good spatial and contrast resolutions. The test sets of the ninth and tenth folds are all captured using the ERC, but the images have high resolution and a large field of view that allows the network to use landmarks to find the prostate.

This analysis shows that for segmentation of MRI slices with good spatial and contrast resolutions there is no significant difference among well-structured models. The problematic folds highlight the actual differences among the proposed models.

G. QUALITATIVE COMPARISON

As a subjective evaluation of the Non-bypass Dense model as the best model, six images selected from the test set of the different folds and the segmentation results are presented in Figure 6 where the red border shows the ground truth and the green border indicates the predicted border. The first and second images (see Figure 6a and Figure 6b), were captured using ERC and the bright region can be seen around the ERC. However, our normalization method compensates the bright region and our proposed model segments the prostate properly. The Non-bypass Dense model segmented the prostate precisely in the third image (see Figure 6c), despite the wrap-around artifacts in the image. The fourth and fifth images (see Figures 6d and 6e), contain only small regions of the prostate, but our method still segments the prostate accurately. In the fourth image, the ERC effect is similar to the prostate, however, our model segment the prostate precisely. The last image (see Figure 6f) is the non-ERC image that is segmented well. Overall, the results show the capability of our best model for the segmentation of the prostate MRI.

H. COMPARISON WITH PRIOR WORKS

In prior work, only ten FCNN-based prostate segmentation methods have been published in conference proceedings or journals. Zhu *et al.* [10] utilized their unpublished dataset achieving 0.885 DSC. However, they excluded non-prostate slices which improves the DSC of their 2D network by reducing spurious detections. Also, Ji *et al.* [56]

TABLE 4. Quantitative comparison of the best proposed models with prior works.

Method	Mean DSC	Comparable
Zhu et al. [10]	0.885	No
Ji et al. [57]	0.91	No
Clark et al. [58]	0.86	No
Milletari et al. [36]	0.869	No
Chen et al. [59]	0.895	No
Liu et al. [60]	0.86	No
Drozdzal et al. [61]	0.874	No
Sun et al. [62]	0.898	No
Mun et al. [38]	0.853	Yes
Yu et al. [37]	0.869	Yes
Non–bypass Dense	**0.873**	––

and Clark *et al.* [57], excluded non-prostate slices from the PROMISE12 dataset and obtained 0.91 and 0.86 mean DSC using ten-fold and four-fold cross-validation respectively. Milletari *et al.* [36], Chen *et al.* [58], and Liu *et al.* [59] evaluated their networks on the test set of the PROMISE12 dataset for which ground truth is not publicly available and obtained 0.869, 0.895, and 0.86 mean DSC respectively. In addition, Drozdzal *et al.* [60], used two FCNN for segmentation. Firstly, they segmented the input image using FCNN and then used another residual based FCNN for boundary refinement. They applied their proposed framework for different organ MRI segmentation including prostate MRI segmentation and obtained 0.874 mean DSC on the test set of the PROMISE12 dataset. Finally, Sun *et al.* [61], proposed an interactive framework for medical image segmentation and a part of this framework is FCNN. They obtained 89.81 mean DSC on the test set of PROMISE12 dataset. Neither of these results are comparable with our results because they use different test conditions.

Yu *et al.* [37] evaluated their proposed model on both the PROMISE12 training set using cross-validation, achieving 0.869 DSC. They also report 0.894 DSC on the test set on the PROMISE12 dataset, which is not comparable with our models. Finally, Mun *et al.* [38] tested their 3D-FCNN (BCNN) network using ten-fold cross-validation on the training set of the PROMISE12 dataset and achieved 0.853 mean DSC. We compare our models with the cross-validation results in these two papers (see Table 4). Given that 3D methods
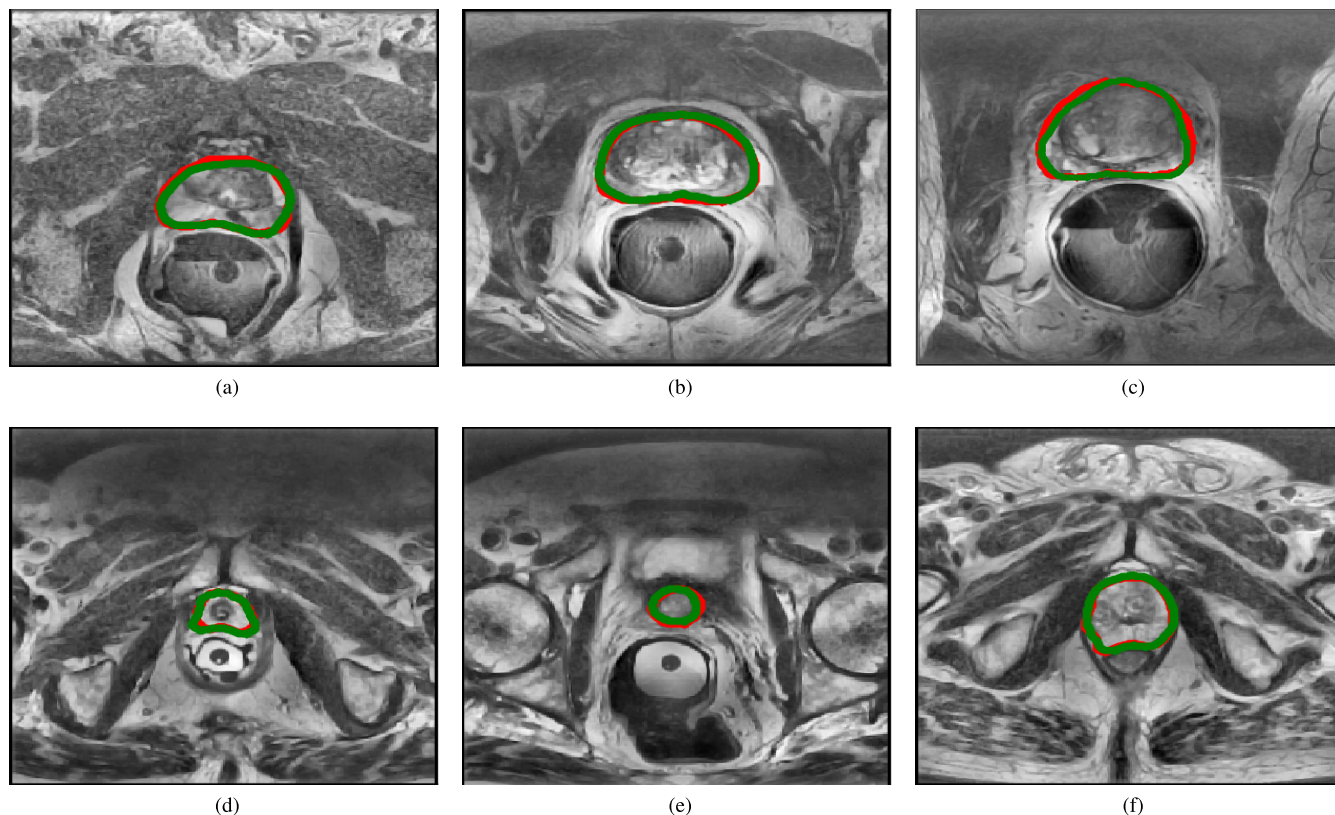
FIGURE 6. The six sample segmented images using Non-bypass Dense model. The red border is the ground truth and the green border is the predicted border.

segment the prostate as a 3D volume (they use information from adjacent slices), finding the prostate will be more straightforward and precise than the 2D models, especially in the lower and upper slices of the prostate volumes where the prostate is a small part of the image. However, as is shown in Table 4, the Non-bypass Dense 2D network outperforms both 3D methods for prostate image segmentation and achieves new state-of-the-art FCNN-based prostate segmentation results.

I. ANALYSING THE ENDORECTAL COIL EFFECT

The PROMISE12 dataset includes 50 MRI volumes for training, of which 24 are captured using the ERC. These volumes include 809 image slices while the 26 non-ERC volumes include only 568 slices. To show the effect of ERC on the final segmentation results, we evaluate the results of the Output from All and Non-bypass Dense models based on the obtained DSC per volume.

The obtained results of the Output from All model per volume is shown in Figure 7a where the red bars indicate the DSC of the ERC volumes and the blue bars show the DSC of the non-ERC volumes. In the Output from All model, the average mean DSC of ERC volumes is 0.8576 and the average DSC of non-ERC volumes is 0.8727. The average DSC in the non-ERC volumes is higher than ERC volumes but, in this model, the best individual result is achieved by

segmentation of an ERC volume, and segmentation of a non-ERC volume produces the worst individual outcome.

In Figure 7b the results of the Non-bypass Dense model per volume are presented. In this model, the average DSC of ERC volumes is 0.8698 and the average DSC of non-ERC volumes is 0.8749. The Non-bypass Dense model was more precise in the segmentation of both ERC and Non-ERC volumes, and more importantly we can see the smaller differences between average mean DSC of the ERC and Non-ERC volumes.

To test for statistically significant differences between the segmentation of the ERC and non-ERC volumes using the Output from All and Non-bypass Dense models, we evaluate the results using the Wilcoxon–Mann–Whitney test [62], which evaluate the differences between two independent sets. The comparison of the obtained results of the Output from All model, for segmentation of the ERC and non-ERC volumes using Wilcoxon–Mann–Whitney test is $p=0.26$, and for the Non-bypass Dense model is $p=0.90$. These results demonstrate that there are no significant differences between the segmentation of the ERC and Non-ERC volumes with either model.

These test results show that using ERC is not the only reason for a low mean DSC. The most important reasons are the low contrast and spatial resolutions and artifacts that can decrease the accuracy of the segmentation. Also, the low diversity of the training images. In both models, the lowest
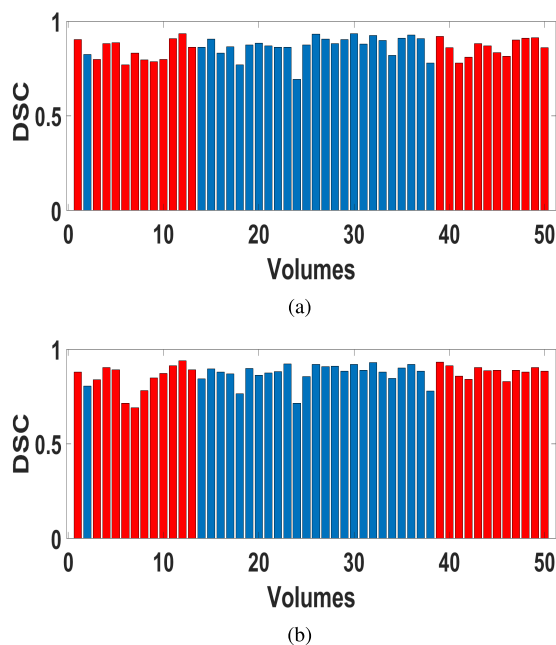
FIGURE 7. Comparison of the Output from All and Non-bypass Dense models based on obtain results per volume. Red bars, ERC-volumes; Blue bars, non-ERC volumes. (a) The Output from All model results per volume. (b) The Non-bypass Dense model results per volume.

mean DSC is for the 24th volume that is a non-ERC and contains the only example of the large prostate. Using zoom augmentation could address this problem somewhat, however increasing the zoom factor is not an appropriate solution because the features in the field of view around the prostate (muscle, fat, etc.) will be lost.

V. CONCLUSION

In this paper, we propose eight FCNN-based network structures for MRI prostate segmentation. Our proposed Non-bypass Dense model outperforms the comparable 3D FCNN-based segmentation methods when evaluated by cross-validation on the PROMISE12 training dataset, and achieves a new state-of-the-art for FCNN-based prostate segmentation on the PROMISE12 training dataset.

In our research, we have analyzed different parameters of FCNN with a particular focus on the structure of shortcut connections. The results of our novel structures show the benefits and advantages of reusing the extracted feature maps within and between the blocks, and also the impact of the network structure on the prostate MRI segmentation. Shortcut connections can help their network. However, our results show that using shortcut connections can also decrease the accuracy of the network; therefore, it is critical to use shortcut connections in the proper place in the network. Our experiments show that the bypass connection, which transfers the input feature maps to the output of the block is not beneficial for prostate segmentation in most cases and when it is used together with scattering connections it even significantly reduces performance. In contrast, gathering connections, which collect the output feature maps of the layers

can significantly improve performance in this application. Also, the results show that, among the models that transfer the equal number of feature maps between the layers, the models with the simpler architecture within the block mostly produce better results.

REFERENCES

[1] H. Xiaolei and G. Tsechpenakis, "Medical image segmentation," in *Information Discovery on Electronic Health Records* (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series), 10th ed. New York, NY, USA: Taylor & Francis, 2009, pp. 251–289.

[2] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *J. Med. Phys./Assoc. Med. Phys. India*, vol. 35, no. 1, p. 3, 2010.

[3] M. S. Fasihi and W. B. Mikhael, "Overview of current biomedical image segmentation methods," in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, 2016, pp. 803–808.

[4] D. R. R. White, A. S. Houston, W. F. Sampson, and G. P. Wilkins, "Intra- and interoperator variations in region-of-interest drawing and their effect on the measurement of glomerular filtration rates," *Clin. Nucl. Med.*, vol. 24, no. 3, pp. 177–181, 1999.

[5] G. P. Haas, N. Delongchamps, O. W. Brawley, C. Y. Wang, and G. de la Roza, "The worldwide epidemiology of prostate cancer: Perspectives from autopsy studies," *Can. J. Urol.*, vol. 15, no. 1, pp. 3866–3871, 2008.

[6] R. Siegel, M. Kimberly, and A. Jemal, "Cancer statistics, 2016," *CA, Cancer J. Clin.*, vol. 68, no. 7, 2018.

[7] *Prostate Cancer Foundation of Australia*. [Online]. Available: http://www.prostate.org.au/awareness/general-information/what-you-need-to-know-about-prostate-cancer/

[8] Australian Institute of Health and Welfare. (2017). *Australian Cancer Incidence and Mortality (ACIM)*. [Online]. Available: http://www.aihw.gov.au/acim-books

[9] *Australian Institute of Health and Welfare 2017*, Cancer Australia, 2017, vol. 101.

[10] Q. Zhu, B. Du, B. Turkbey, P. L. Choyke, and P. Yan, "Deeply-supervised CNN for prostate segmentation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2017, pp. 178–184.

[11] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1077–1089, Apr. 2016.

[12] M. Ghafoori, M. Alavi, M. Shakiba, and K. Hoseini, "The value of prostate MRI with endorectal coil in detecting seminal vesicle involvement in patients with prostate cancer," *Iranian J. Radiol.*, vol. 12, no. 1, 2015, Art. no. e14556.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, 2015, pp. 234–241.

[15] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4700–4708.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[17] H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, "Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2497–2506.

[18] H. I. Suk *et al.*, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Function*, vol. 220, no. 2, pp. 841–859, 2015.

[19] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, Contour and Grouping in Computer Vision*. Berlin, Germany: Springer, 1999, pp. 319–345.

[20] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

[21] R. Cheng *et al.*, "Active appearance model and deep learning for more accurate prostate segmentation on MRI," *Proc. SPIE*, vol. 9784, Mar. 2016, Art. no. 97842I.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[23] R. Cheng *et al.*, "Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks," *J. Med. Imag.*, vol. 4, no. 4, 2017, Art. no. 041302.

[24] H. Jia, Y. Xia, Y. Song, W. Cai, M. Fulham, and D. D. Feng, "Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging," *Neurocomputing*, vol. 275, pp. 1358–1369, Jan. 2018.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2014, pp. 1409–1556.

[26] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[27] B. He, D. Xiao, Q. Hu, and F. Jia, "Automatic magnetic resonance image prostate segmentation based on adaptive feature learning probability boosting tree initialization and CNN-ASM refinement," *IEEE Access*, vol. 6, pp. 2005–2015, 2018.

[28] K. Yan *et al.*, "Automatic prostate segmentation on MR images with deep network and graph model," in *Proc. Eng. Med. Biol. Soc. (EMBC), IEEE 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 635–638.

[29] P. Krähenbühl and V. Koltun, "Geodesic object proposals," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 725–739.

[30] S. Liao, Y. Gao, A. Oto, and D. Shen, "Representation learning: A unified deep learning framework for automatic prostate MR segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, 2013, pp. 254–261.

[31] S. Ioffe and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167

[32] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012.

[33] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3642–3649.

[34] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2528–2535.

[35] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 326–327, 1995.

[36] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.

[37] L. Yu, X. Yang, H. Chen, J. Qin, and P.-A. Heng, "Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images," in *Proc. Assoc. Adv. Artif. Intell.*, 2017, pp. 66–72.

[38] J. Mun, W.-D. Jang, D. J. Sung, and C.-S. Kim, "Comparison of objective functions in CNN-based prostate magnetic resonance image segmentation," in *Proc. Int. Conf. Image Process.*, 2017, pp. 3859–3863.

[39] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 11–19.

[40] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," in *Proc. Int. Conf. Mach. Learn.*, 2015.

[41] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *J. Statist. Planning Inference*, vol. 90, no. 2, pp. 227–244, 2000.

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[43] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature*, vol. 405, no. 6789, pp. 947–951, 2000.

[44] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

[45] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[46] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," Univ. California San Diego, San Diego, CA, USA, Tech. Rep., 1985.

[47] G. Litjens *et al.*, "Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge," *Med. Image Anal.*, vol. 18, no. 2, pp. 359–373, 2014.

[48] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. Int. Conf. Document Anal. Cognition*, vol. 3, 2003, pp. 958–962.

[49] E. Kreyszig, *Advanced Engineering Mathematics*. Hoboken, NJ, USA: Wiley, 2010.

[50] [Online]. Available: https://keras.io/

[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014.

[52] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[53] C. Shen *et al.*, "On the influence of dice loss function in multi-class organ segmentation of abdominal ct using 3D fully convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2018.

[54] R. R. Shamir, Y. Duchin, J. Kim, G. Sapiro, and N. Harel, "Continuous dice coefficient: A method for evaluating probabilistic segmentations," Tech. Rep., 2018.

[55] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.

[56] D. Ji, J. Yu, T. Kurihara, L. Xu, and S. Zhan, "Automatic prostate segmentation on MR images with deeply supervised network," in *Proc. Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, 2018, pp. 309–314.

[57] T. Clark, J. Zhang, S. Baig, A. Wong, M. A. Haider, and F. Khalvati, "Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks," *J. Med. Imag.*, vol. 4, no. 4, 2017, Art. no. 041307.

[58] W. Chen *et al.*, "W-Net: Bridged U-net for 2D medical image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018.

[59] Q. Liu, M. Fu, X. Gong, and H. Jiang. (2018). "Densely dilated spatial pooling convolutional network using benign loss functions for imbalanced volumetric prostate segmentation." [Online]. Available: https://arxiv.org/abs/1801.10517

[60] M. Drozdzal *et al.*, "Learning normalized inputs for iterative estimation in medical image segmentation," *Med. Image Anal.*, vol. 44, no. 1, Feb. 2018.

[61] J. Sun *et al.*, "Interactive medical image segmentation via point-based interaction and sequential patch learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018.

[62] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Ann. Math. Statist.*, vol. 18, no. 1, pp. 50–60, 1947.

TAHEREH HASSANZADEH received the B.Sc. degree in software engineering from Lahijan Azad University, Iran, and the M.Sc. degree in artificial intelligence from Qazvin Azad University, Iran. She is currently pursuing the master's degree with Macquarie University, Sydney, NSW, Australia. Her research interests include image processing, machine learning, and evolutionary algorithms. She published several papers in the areas of image processing, machine learning, and optimization. Also, she is a Reviewer of different conferences and journals.

LEONARD G. C. HAMEY (M'−) received the B.Sc. degree (Hons.) in statistics from Macquarie University, in 1982, and the Ph.D. degree in computer science from Carnegie Mellon University, in 1988. He is a Senior Lecturer and the Deputy Head of the Department of Computing, Macquarie University. His research interests include artificial neural networks and computer vision. He is a member of the Australian Pattern Recognition Society.

KEVIN HO-SHON received the joint MBBS and B.Sc. degrees (Hons.) in computer science for research in positron emission tomography from The University of Sydney. He continued this research in PET with the M.Sc. degree by research. He trained in radiology at Royal Prince Alfred Hospital (RPAH) and remained at RPAH as a Staff Specialist Interventional Radiologist, where he developed an interest in interventional oncology and liver intervention. He has been a part of the Liver Cancer Group and works closely with the Liver Transplant Unit. He also pursued his interest in applied mathematics and was awarded the Ph.D. degree in financial mathematics, in 2008. He then established one of the major radiology groups in Sydney and is a Founding Member of Macquarie Medical Imaging. He is currently the Medical Imaging Discipline Head of the Macquarie University Health Science Centre.

● ● ●