IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 4, 2020, accepted June 27, 2020, date of publication July 1, 2020, date of current version July 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006317

Using a Patch-Wise M-Net Convolutional Neural Network for Tissue Segmentation in Brain MRI Images

NAGARAJ YAMANAKKANAVAR, (Member, IEEE), AND BUMSHIK LEE, (Member, IEEE)
Department of Information and Communications Engineering, Chosun University, Gwangju 61452, South Korea

Corresponding author: Bumshik Lee (bslee@chosun.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government under Grant NRF-2019R1A4A1029769 and Grant NRF-2019R1I1A3A01058959.

ABSTRACT Accurate segmentation of brain tissues, such as gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), in magnetic resonance imaging (MRI) images, is helpful for the diagnosis of neurological disorders, such as schizophrenia, Alzheimer's disease, and dementia. Studies on MRI-based brain segmentation have received significant attention in recent years based on the non-invasive imaging and good soft-tissue contrast provided by MRI. A number of studies have used conventional machine learning strategies, as well as convolutional neural network approaches. In this paper, we propose a patch-wise M-net architecture for the automatic segmentation of brain MRI images. In the proposed brain segmentation method, slices from a brain MRI scan are divided into non-overlapping patches, which are then fed into an M-net model with corresponding ground-truth patches to train the network, which is composed of two encoder-decoder processes. Dilated convolutional kernels with different sizes are used in the encoder and decoder modules to derive abundant semantic features from brain MRI scans. The proposed patch-wise M-net overcomes the drawbacks of conventional methods and provides greater retention of fine details. The proposed M-net model was trained and tested on the open-access series of imaging studies dataset. The performance was measured quantitatively using the Dice similarity coefficient. Experimental results demonstrate that the proposed method achieves average segmentation accuracies of 94.81% for CSF, 95.44% for GM, and 96.33% for WM, meaning it outperforms state-of-the-art methods.

INDEX TERMS Brain MRI, convolutional neural network, M-net, tissue segmentation.

I. INTRODUCTION

Because neurologists investigate tissue abnormalities, such as cortical thickening, shrinkage, and ventricle expansion, for diagnosis using segmentation, accurate segmentation is one of the crucial factors for the correct diagnosis of neurological disorders. Therefore, segmentation of brain magnetic resonance imaging (MRI) tissue types, such as gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), into their basic cytoarchitectural tissue classes is useful for clinicians treating neurological disorders, such as epilepsy, schizophrenia, Alzheimer's disease, and dementia [1]. Furthermore, the automatic segmentation of brain tissue types (WM, GM, and CSF) is extremely important for neuroscientific studies, such as cortical surface extraction [2], [3], atrophy and volume measurement [4], [5], brain extraction [6], [7], and multiple sclerosis lesion segmentation [8]. However, manual segmentation by experts is time-consuming, prone to human error, and impractical for large datasets. Therefore, developing accurate methods for automated brain-tissue segmentation has become an active research area. Fig. 1 presents ground-truth tissue classes for a brain MRI scan. The segmentation map of the ground truth is categorized into the background, CSF, GM, and WM regions. The main objective of brain MRI segmentation is to classify image pixels into well-defined regions, where a set of pixels shares the same range of intensities, textures, or neighborhoods. This paper mainly focuses on the segmentation of CSF, GM, and WM regions in brain MRI scans using a deep learning architecture, namely a patch-wise M-net.

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

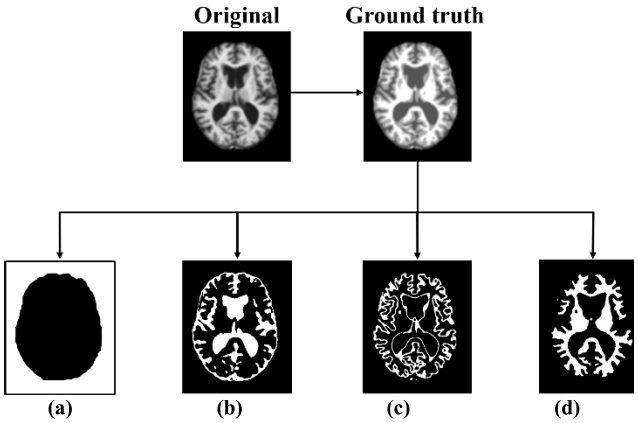

FIGURE 1. Tissue classes in a brain MRI original and ground-truth images: (a) background, (b) CSF, (c) GM, and (d) WM.

There are three main challenges in automated brain tissue segmentation. First, because there are large variations in brain anatomical structures with phenotypes, such as age, gender, and disease, it is difficult to generalize one specific segmentation method for all phenotypic categories. Second, it is difficult to process cytoarchitectural variations, such as gyral folds, sulci depths, thin tissue structures, and smooth boundaries between different tissues. This can result in confusing categorical labeling for distinct tissue classes. This is even challenging for human experts. Finally, MRI technology has limitations related to bias effects, signal-to-noise ratios, and motion artifacts. To overcome these challenges, many classical machine learning approaches using support vector machines [9], random forests,[10] neural networks [11], and population-specific atlases [12] have been developed for the segmentation of brain MRI scans. These methods require not only explicit spatial and intensity information but also the extraction of feature vectors from intensity information for accurate segmentation performance.

In recent years, deep-learning-based methods have received significant attention. State-of-the-art deep learning architectures for image segmentation employ computationally complex 3D models and require large amounts of training data. However, it is difficult to obtain large amounts of labeled training data for medical image analysis. Therefore, it is necessary to develop a method that can provide adequate performance with a relatively small amount of training data. The patch-based CNNs [15], also called slide-window-based CNNs, are useful in such a scenario because the model can efficiently be trained with a small amount of training data with multi-scale patches, whose sizes are different depending on different modalities such as T1- and T2- weighted images. However, the training and testing processes of the patch-based CNNs for segmentation take significant computation times because the model needs to run separately for each multi-sized patch. Jiong *et al*. [41] recently proposed the skip-connection U-net for WM hyperintensities segmentation from brain MRI, where an atlas-based method was also introduced in the pre-processing stage to remove non-brain tissues (namely skull-stripping), thus improving the segmentation accuracy. Bernal *et al*. [40] introduced the quantitative analysis of patch-based FCNN strategies for tissue segmentation on brain MRI. Furthermore, the author discussed various methods by fixing training and test sets, pre-processing pipeline, training with optimization schemes, and performance evaluation metrics.

Besides the deep learning-based approaches, several kinds of research using classical machine learning methods [42], [43], [47], [48] were recently performed. The fuzzy C-means framework to improve the temporal consistency of adults' brain tissue segmentation was proposed in [42], where the effectiveness of the results is demonstrated on the Baltimore longitudinal study of aging (BLSA) benchmark dataset. Chen *et al*. [43] proposed a new iterative linearly constrained minimum variance (LCMV) classification-based method, which was developed for hyperspectral classification and the different spatial filters in the classification of the brain tissues were investigated. Saif *et al*. [47] used a one-class support vector machine to segment a single tract from a whole-brain tractography and tract analysis used in human brain disease identification. Sun *et al.* [48] introduced the high-order feature learning framework for multi-atlas based label fusion, where high-order features of image patches are extracted and fused for segmenting ROIs of structural brain MRI. A means-covariance restricted Boltzmann machine (mcRBM) method is employed high order features of patches in brain MRI. Furthermore, given the computational resources and dataset size, the classical machine learning methods [42], [43], [47], [48] can produce good results because of easy tuning capability for hyperparameters, the flexibility of adding new features and it can utilize the prior anatomical structure of the brain provided by the datasets. Even though these models give better results, it requires a large amount of computation time. In [48], the limitation is that high dimensional features and the regularization terms (i.e., group Lasso and fused Lasso) will increase the computational burden for learning the voting weights. In addition, these classical methods are known not to work well on imbalanced and more massive datasets with a large number of features. Table 1 shows a summary of the related works for brain structure segmentation using deep learning. As shown in Table 1, most of the segmentation methods [13]–[15], [34], [35], [38] adopt the CNN architecture for the segmentation of brain MR images. Recently, researchers have explored both 2D and 3D local-neighborhood-based methods for the segmentation of brain MRI scans [13], [14]. The convolutional neural network (CNN)-based techniques have achieved excellent performances in various applications, including handwritten digit recognition, object detection, and semantic segmentation. A CNN was then applied for the segmentation of MRI scans. Nie *et al.* [17] used deep fully convolutional networks to train a model for multiple image modalities, such as T1, T2, and FA, and then combine layered feature maps in a final segmentation map output.

Bao and Chung [34] introduced the new technique for the segmentation of brain MRI based on multi-scale CNN

TABLE 1. Summary of related studies on brain structure segmentation using deep learning and machine learning methods.

Methods	Architectures	Datasets
Zhang *et al.* [13]	Patch-wise CNN	Private data
de Brebission *et al.* [14]	2D/3D patch-wise CNN	MICCAI 2012
Moeskops *et al.* [15]	Multi-scale patch-wise CNN	MICCAI 2012
Nie *et al.* [17]	FCN	Private data
Bao *et al.* [34]	2D Patch-wise CNN	IBSR data
Wachinger *et al.* [35]	2D Patch-wise CNN	MICCAI 2012
Wei *et al.* [38]	Multi-size Patch-based CNN	BrainWeb IBSR
Chen *et al.* [43]	2D Semantic-wise	BrainWeb
Pengcheng *et al.* [42]	3D Semantic-wise	BLSA

(MS-CNN), which gives differentiable features for each subcortical structure and generates a label probability map to a target image. However, there would be a lack of spatial constraints in the testing samples due to the irregular background in the brain images. Wachinger *et al.* [35] introduced the Neuroanatomy in T1-W MRI segmentation using the deep CNN. It is an end-to-end learning-based approach to brain segmentation that learns an abstract feature representation and a multi-class classification. Wei *et al.* [38] proposed a multi-model, multi-size, and multi-view deep neural network for the segmentation brain MRI on a slice-by-slice basis. Besides, it uses deep multi-size U-nets for large patch segmentation and a multi-size backpropagation neural networks for small patch classification. In addition to the applications of brain segmentation, deep learning approaches are also used for the segmentation of other anatomies such as the prostate, liver, and abdomen. The 3D adversarial pyramid anisotropic (3D APA-net) convolutional deep neural network is proposed for the prostate segmentation. The 3D APA-net architecture consists of 3D ResNet encoder, an anisotropic convolutional decoder, and multi-pyramid convolutional skip connections. Chen *et al.* [49] introduced an automated deep learning based abdominal multi-organ segmentation framework based on CNN. A multi-slice 2D neural network was developed to account for the correlative as well as complementary information between adjacent slices in the intrinsic 3D space while avoiding the heavy computation burden. Ahmad *et al.* [50] proposed an automatic liver segmentation based on the deep belief network (DBN). The DBN model is trained based on unsupervised pre-training and supervised fine-tuning. The unsupervised pre-training relates to a highly proficient learning technique that stacks restricted Boltzmann machines (RBMs), which are individually trained layer by layer. In supervised fine-tuning, all layers, along with backpropagation neural networks, are fine-tuned to perform the classification task. Milletari *et al.* [21] proposed using a Hough CNN based on Hough voting to facilitate fully automatic localization and segmentation of anatomies of interest. CNNs have some drawbacks in segmentation applications because, during the segmentation process, reconstruction should be performed using vectors, meaning one not only needs to convert a feature map into vectors but also needs to reconstruct brain images from vectors. Therefore, the SegNet [18] and U-net [19] architectures have been widely used for segmentation based on their advantages over CNNs in terms of reconstruction capabilities. SegNet [18] is a well-known architecture in computer vision for semantic segmentation but has rarely been used for brain MRI scan segmentation. It passes pooling indices to up-sampling layers, meaning it requires much fewer parameters, leading to faster training. In existing methods [13]–[15], [17], [34], [35], randomly selected portions and/or regions from a slice or MRI volume are used as patches for training the model. It is known that the performance of patch-based deep learning approaches relies on the size of patches [38]. The small size patches may not contain sufficient information to train the network architecture, thus resulting in relatively inaccurate segmentation results. In contrast, a large size of patches is more likely to contain slices from multiple classes, leading to the difficulty in patch classification. Motivated these problems regarding the size of patches in the existing methods, we propose the segmentation method with uniform-sized patches, which are fed into the M-net model for training. The complete information of the slices partitioned with the uniform size can be used as training data, thus resulting in robust segmentation performances with local detail information. As a result, more accurate segmentation results can be obtained. Furthermore, U-net [19] uses multi-scale information based on skip connections and captures both coarse- and fine-level information in deconvolutional layers. However, because it uses trainable up-sampling, U-net must learn many more parameters and is much slower to train than SegNet, which does not capture multi-scale information as effectively as U-net. Adiga *et al.* [20] proposed an end-to-end multi-class deep network called M-net with deep supervision functionality. However, the M-net architecture has issues related to limited memory for high-resolution input images because the number of feature channels increases over the resolution of the input images at each stage of down- and up-sampling, thus resulting in a more number of parameter values to store. Furthermore, the M-net is prone to losing local details because complete images are used as network inputs. To overcome these problems, we propose a patch-wise M-net architecture using multi-scale dilated convolutional kernels to perform automatic segmentation of brain MRI scans.

Specifically, we use individual non-overlapping patches extracted from input slices to train an M-net architecture. This patch-wise splitting of slices improves localization in MRI images and allows the trained network to focus on local details in each patch. Furthermore, dilated convolutional kernels with different sizes are used in encoder and decoder modules to extract abundant semantic features from brain MRI scans. The proposed method exhibits significant improvements over conventional methods for brain MRI segmentation in terms of several evaluation metrics. The main contributions of this paper can be summarized as follows.

- We propose a patch-wise M-net architecture to segment brain MRI scans. The use of patch-wise splitting of slices obtained from MRI scans enhances localization

because the trained network can focus on local details within each patch.

- The proposed patch-wise M-net architecture can avoid the problems of reduced resolution for feature maps and the loss of semantic features.
- Multi-scale dilated convolutional kernels are used to extract context information at different scales, thereby improving segmentation accuracy for brain MRI scans.
- We demonstrate that the proposed method achieves excellent segmentation accuracy in terms of the Dice similarity coefficient (DSC) and Jaccard index (JI), resulting in performance comparable to that of conventional methods on publicly available MRI datasets.

The remainder of this paper is organized as follows. In Section 2, we provide a detailed description of the proposed method. We analyze the performance of the proposed method on publicly available datasets and discuss experimental results in Section 3. This paper is concluded in Section 4.

II. PROPOSED METHOD

Before describing the proposed method, it is necessary to explain how U-net and SegNet work in greater detail because our proposed method is inspired by M-net, which is a variant of U-net. SegNet [18] and U-net [19] have been widely used in segmentation applications. The SegNet [18] architecture consists of an encoder and decoder. Inputs are down-sampled using multiple convolutions and max-pooling operations in the encoder. In the decoder, down-sampled feature maps are up-sampled using memorized max-pooling indices from corresponding encoder feature maps and convolution operations. Each pixel in the decoder output is classified independently using a Softmax classifier. SegNet-based segmentation yields poor performance compared to other existing methods because SegNet tends to lose neighborhood information when performing up-sampling using low-resolution feature maps. The main difference between SegNet and U-net lies in their up-sampling operations. For U-net, feature maps from the decoding path are concatenated with corresponding feature maps from the encoding path during up-sampling. The incorporation of concatenation improves localization. Therefore, U-net has been widely used in the segmentation of biomedical images, where improved localization is crucial for achieving better performance. Some known drawbacks of the U-net architecture are that it is unable to learn deep information and has weak generalization ability. To alleviate these problems effectively, the M-net architecture was proposed for segmenting brain MRI scans. The M-net [20] architecture has two side paths (left and right legs) and two main encoding and decoding paths, which helps it learn better features. Furthermore, the two side paths aid in learning fine-grained details from brain MRI scans. However, M-net also has some limitations when it takes a complete image as an input. Therefore, this model is prone to missing details in certain regions of images. To address these problems related to U-net and M-net for brain MRI scan segmentation, we propose a segmentation method that divides input images into non-overlapping patches and trains an M-net model using patches. Patches are beneficial for retaining local information in images. Furthermore, patches with small sizes are better for training compared to complete images because less memory is required for computation. The main challenge in segmenting brain structures in MRI scans is a multi-class problem. In this study, we performed multi-class segmentation to label every pixel in an MRI scan as one of the four classes: background, GM, WM, and CSF. To this end, the conventional M-net architecture [20] was modified such that the final layer produces a binary map for each of the four classes, rather than a single binary map representing foreground and background pixels. Input ground-truth segmentation maps are also converted into multi-channel binary segmentation maps for each class.

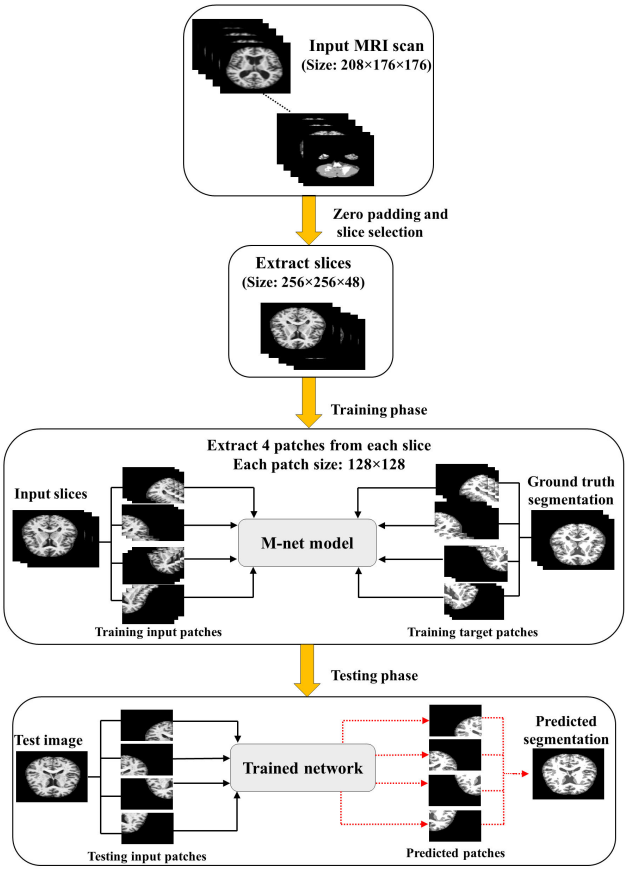

FIGURE 2. Block diagram of the proposed method.

Fig. 2 presents a block diagram of the proposed system. The dimension of each input axial scan in the OASIS dataset [33] is 208 × 176 × 176, and each scan consists of 176 slices. In this study, approximately 48 slices were extracted from each MRI scan and resized to dimensions of 256 × 256 × 48 by adding 24 pixels of zeros on the top and bottom of the image and 40 pixels of zeros on the left and right of the image. Therefore, each input scan consists of 48 slices with dimensions of 256 × 256. During the

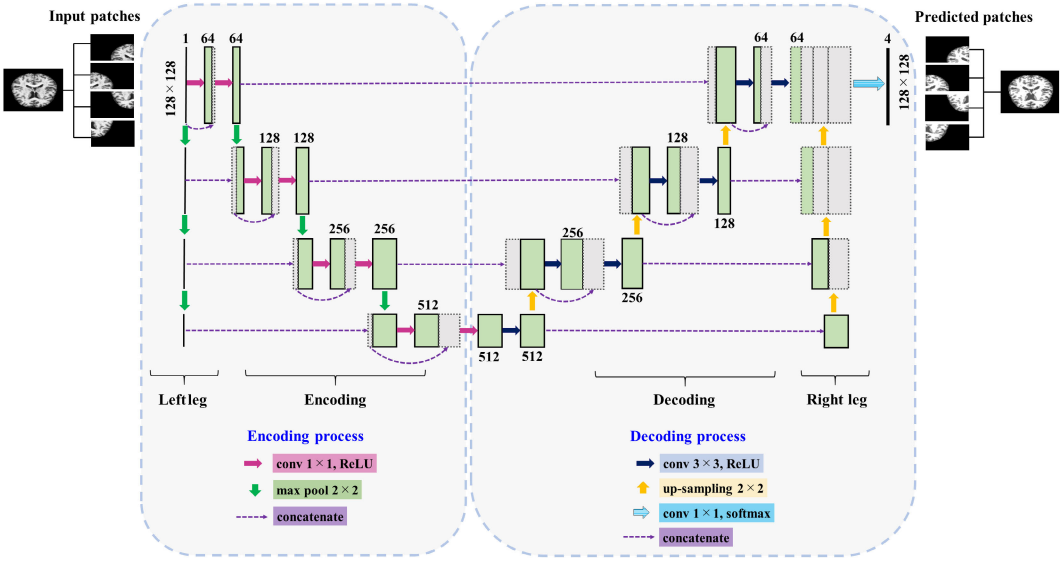

FIGURE 3. Schematic representation of the proposed method. Solid green boxes represent the outputs of convolution-ReLU blocks, and dashed boxes represent copied feature maps. The number of feature maps is denoted on top of each box.

training stage, slices of each MRI scan and their corresponding ground-truth segmentation maps are divided into different patches. The dimensions of an input slice are 256× 256, and each slice is split into four patches. Therefore, the dimensions of each partitioned patch are 128 × 128 in the proposed method. These patches are inputted into the M-net model for training and predicted segmentation results are obtained for test data.

The M-net architecture is an end-to-end multi-class deep network consisting of four parts. Similar to U-net, M-net consists of two side paths (left and right legs), as well as encoding and decoding paths. Inputs are down-sampled in the left-leg path and fed into the corresponding encoder layer. Similarly, decoding layer outputs are up-sampled to the input size in the right-leg path. The outputs obtained from the right-leg layer and the decoder layer are combined to produce a final output.

Fig. 3 shows a schematic representation of the proposed method. The proposed architecture consists of a convolutional layer, a rectified linear unit (ReLU), max-pooling layers, and up-convolutional layers. Each encoding layer consists of two consecutive blocks of 1 × 1 convolution and ReLU, where each convolution is followed by a ReLU activation function. The output of the convolution-ReLU layer is concatenated and down-sampled using 2 × 2 max-pooling operations with a stride of two. After every down-sampling step, the number of feature maps is doubled, as suggested in [19]. Each decoding layer also consists of two consecutive blocks of 3 × 3 convolutions and ReLU, where each convolution is followed by a ReLU activation function. The max-pooling operation is replaced by an up-convolution operation that helps reconstruct output images. The final layer is a 1 × 1 convolutional layer with a softmax activation function that outputs a reconstructed image.

The skip connections used in the patch-wise M-net are indicated by dotted arrows in Fig. 3. The skip connections adjacent to convolutional filters enable the network to learn useful features [26], and the input-to-encoder (left leg), encoder-to-decoder, and decoder-to-output (right leg) skip connections ensure that the network has sufficient information to derive fine-grained details from brain MRI scans. Furthermore, we utilize dilated convolutional kernels with different sizes for the encoder and decoder modules. It is well known that smaller convolutional kernels are more sensitive to small targets compared to large kernels [27]. However, increasing the size of the receptive field can make use of context information from a larger image region [28] and accelerate model convergence [29]. In the proposed method, we use 1 × 1 convolutional kernels in the encoder module to derive abundant semantic features from brain MRI scans. In contrast, 3 × 3 convolutional kernels are used in the decoder. The semantic features obtained by each convolutional block in the encoder module are transferred to a corresponding convolutional block with the same resolution in the decoder module. Dilated convolutions can expand the receptive field without losing resolution or coverage, meaning they can aggregate multi-scale contextual information to improve segmentation accuracy [30, 31].

III. EXPERIMENTAL RESULTS

A. DATASET AND EXPERIMENTAL SETUPS

The proposed method is evaluated on an Open Access Series of Imaging Studies (OASIS) [33] and Internet Brain Segmentation Repository (IBSR) [37] datasets. Table 2 shows the details of the OASIS and IBSR datasets. The OASIS

TABLE 2. Information for oasis and IBSR datasets.

	No. of subject	
	OASIS	IBSR
Males	160	14
Females	256	4
Total	**416**	**18**

dataset [33] was created by Washington University, which has an Alzheimer's disease research center that tracks information from both non-demented and demented subjects. The dataset consists of both longitudinal and cross-sectional MRI brain images. The longitudinal dataset contains multiple scans of each subject over a period of time, and the cross-sectional category contains details for 416 subjects with ages between 18 and 96 years.

We selected the first 50 subjects (ID OAS1_0001_MR1 to OAS1_0054_MR1) from the OASIS dataset. From the selected data, the first 30 subjects were used for training, and the model was tested on the remaining 20 subjects. The axial, sagittal, and coronal planes of MRI slices are used for training and testing for the segmentation of brain MRI in our experiments. The dimension of the axial scan in the OASIS dataset is $208 \times 176 \times 176$ (height $\times$ width $\times$ slices), and each axial scan consists of 176 slices in total. For the experiment, the original axial scan is resized to a dimension of $256 \times 256 \times 176$ by padding 24 pixels of zero to the top and bottom of the image and 40 pixels of zeros to left and right of the image. Similarly, the original dimensions of the sagittal ($176 \times 208 \times 176$) and coronal ($176 \times 176 \times 208$) scans are resized to dimensions of $256 \times 256 \times 176$ and $256 \times 256 \times 208$, respectively. We also performed the experiments on the IBSR dataset, which comprises of 18 T1-weighted MRI images of 4 healthy females and 14 healthy males with age ranging from 7 to 71 years. The MRIs in the IBSR is provided after pre-processing such as skull-stripping, normalization, and bias field correction. The ground truth is made with manual segmentation by experts with tissue labels as 0, 1, 2, and 3 for background, CSF, GM, and WM, respectively. In our experiments, the first 12 subjects were used for training, while the model was tested on the remaining six subjects. The original axial scans ($256 \times 128 \times 256$) in the dataset are resized to a dimension of $256 \times 256 \times 256$ by zero-padding with 64 pixels to top and bottom of the image to use the patches in our proposed method efficiently. In a similar way, the original dimensions of the sagittal ($128 \times 256 \times 256$) and coronal ($256 \times 256 \times 128$) are also resized to dimensions of $256 \times 256 \times 256$ for the experiments.

Our proposed method is compared with SegNet [18], U-net [19], and M-net [20], which are tested with the parameters shown in Table 3. A total of 50 subjects are used for experiments, where 30 subjects for training and 20 subjects for testing are used, respectively. The dimension of all planes in the datasets are resized to 256×256 by padding zeros in the brain MRI. The SegNet [18] passes pooling indices to the up-sampling layers from the encoder to the decoder. For U-net [19] and M-net [20], the decoder section mirrors the encoder with up-sampling layers concatenated with feature maps from the encoder at the same feature map level. The final output is computed using 1×1 convolution with the Softmax activation function.

B. RESULT AND DISCUSSION

The network was trained using a batch size of one for a maximum of 10 epochs. In our experiments, we observed that the loss function of the network converged to the lowest value within 10 epochs and tended to exhibit overfitting above 10 epochs. The models were optimized according to a categorical cross-entropy loss function using the stochastic gradient descent optimization method with an initial learning rate of 0.0001 and a high momentum rate of 0.99. For initializing weights, a normalization technique [22] was adopted. Experiments were performed using the Keras framework on an Nvidia 1080Ti GPU.

The axial, sagittal, and coronal planes of MRI slices were used for the segmentation of brain MRI in our experiments. For training, we extracted a total of 48 slices starting from the 10th slice with a uniform interval of three slices in all planes, regardless of the positions of slices. In other words, we considered both central slices (i.e., slices with more information) and non-central slices (i.e., slice with less information) for training.

Fig. 4 and Fig. 5 shows the segmentation results for axial, coronal, and sagittal planes of the OASIS and IBSR datasets, respectively. It can be observed in Fig. 4 and Fig. 5 that the proposed method achieves well-segmented results for GM, WM, and CSF on both datasets.

To verify the proposed method, the DSC, JI, precision, recall, and specificity are used to evaluate segmentation quality. The DSC [23] and JI [24] metrics are the most widely used metrics for evaluating the performance of segmentation methods. All evaluation metrics were used to compute the similarity between two sample sets for segmentation. These metrics indicate how closely a predicted segmentation map matches the corresponding ground-truth segmentation map. The evaluation metrics for brain tissue segmentation are defined as follows:

$$DSC = \frac{2.TP}{2.TP + FP + FN} \tag{1}$$

$$JI = \frac{TP}{TP + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Furthermore, we analyzed the mean squared error (MSE) for each method. MSE is an average squared difference between

TABLE 3. Experimental setups.

Parameters	SegNet [18]	U-net [19]	M-net [20]	Proposed method
Input image size	256×256	256×256	256×256	128×128
Training set	30 subjects	30 subjects	30 subjects	30 subjects
Testing set	20 subjects	20 subjects	20 subjects	20 subjects
Number of Epochs	10	10	10	10
Learning rate	0.0001	0.0001	0.0001	0.0001
Number of patches (for 1 slice)	1	1	1	4
Convolution kernels at encoder	3×3, stride of 1	3×3, stride of 1	3×3, stride of 1	1×1, stride of 1
Convolution kernels at decoder	3×3, stride of 1	3×3, stride of 1	3×3, stride of 1	3×3, stride of 1
Max pooling Operation	2×2, stride of 2	2×2, stride of 2	2×2, stride of 2	2×2, stride of 2
Activation function	ReLU	ReLU	ReLU	ReLU
Output function	Softmax	Softmax	Softmax	Softmax

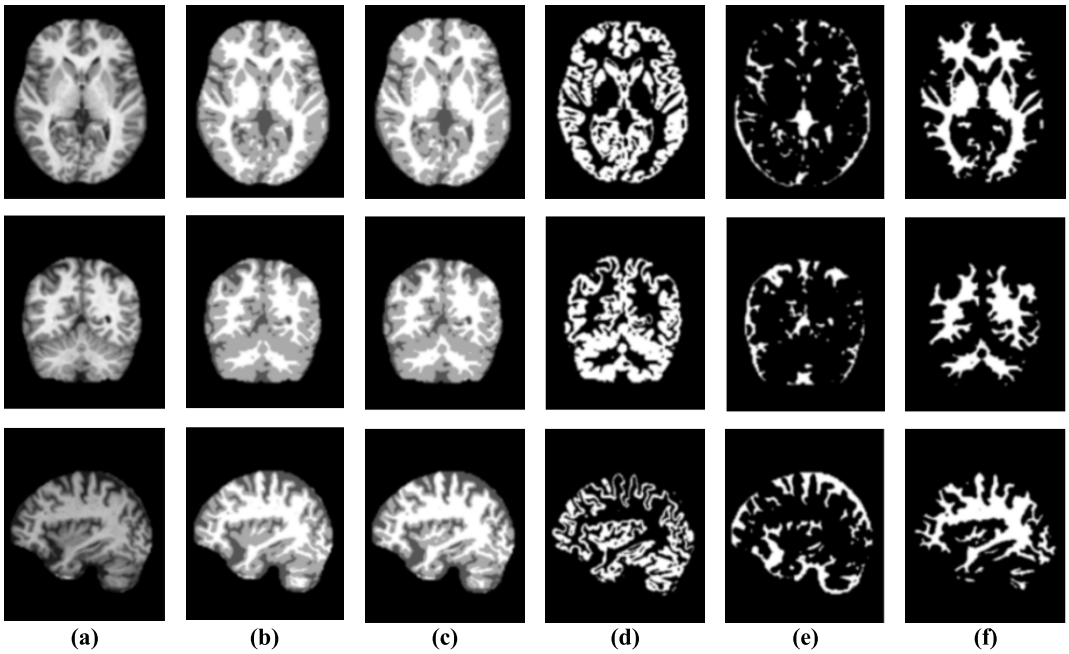

FIGURE 4. Illustration of segmentation results obtained for our proposed method for axial, coronal and sagittal (top to bottom) using OASIS dataset:(a) Original input images, (b) ground truth segmentation map, (c) predicted segmentation map, (d) predicted GM(binary map), (e) predicted CSF (binary map), and (f) predicted WM (binary map).

a ground-truth segmentation map S and predicted segmentation map S'. MSE is defined as

$$MSE = \frac{1}{RC} \sum_{i=1}^{R} \sum_{j=1}^{C} (S_{ij} - S'_{ij})^2 \tag{6}$$

where R and C are the height and width of the image, i and j are pixel indices, respectively. To validate the efficacy of our proposed patch-wise M-net, we also performed experiments using the SegNet, U-net, and M-net architectures to benchmark their performances on the OASIS dataset. Fig. 6 and Fig. 7 presents the segmentation results for GM, CSF, and WM for existing methods and the proposed method. The first column (a) in Fig. 6 and Fig. 7 presents the original images from MRI scans. The second column (b) presents the ground-truth segmentation maps for the original images. In the third column (c), the segmentation maps generated using SegNet, U-net, M-net, and the proposed method are presented (top to bottom). The fourth column (d), fifth column (e), and sixth column (f), present binary GM, CSF, and WM maps, respectively, generated by SegNet, U-net, M-net, and the proposed method (top to bottom).

As shown in Fig. 6 and Fig. 7, the proposed method produces the best segmentation results. Compared to the results of the other segmentation methods, the quality of the segmentation map generated by the proposed method is clearly superior. One can observe that the segmentation results of the U-net and SegNet architectures lack fine details compared to those of the proposed method, as indicated by the red squares. In particular, SegNet tends to miss many fine details because it loses neighboring information when performing up-sampling from low-resolution feature maps. Similarly, U-net has difficulty capturing detailed textures, particularly

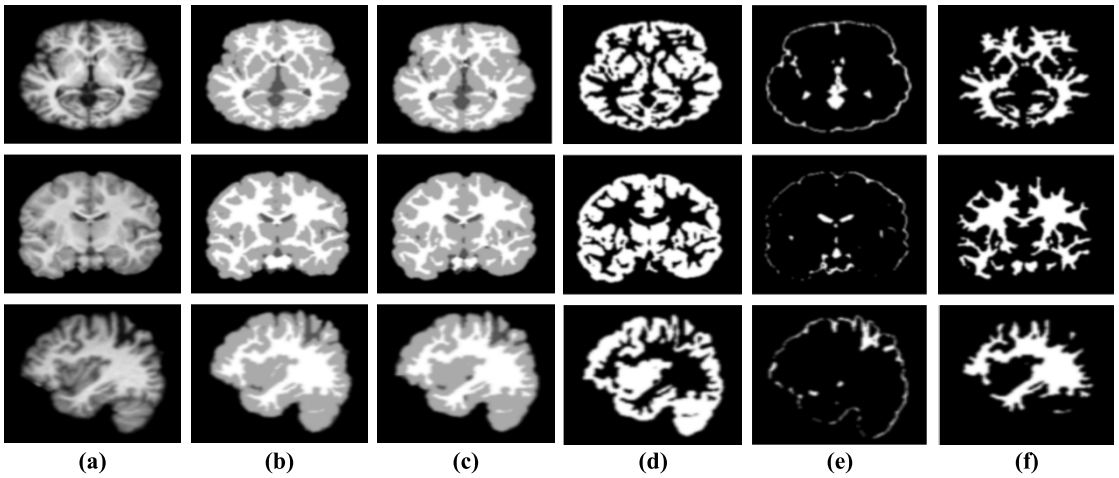

FIGURE 5. Illustration of segmentation results obtained for our proposed method for axial, coronal and sagittal (top to bottom) using IBSR dataset:(a) Original input image, (b) ground truth segmentation map, (c) predicted segmentation map, (d) predicted GM (binary map), (e) predicted CSF (binary map), and (f) predicted WM (binary map).

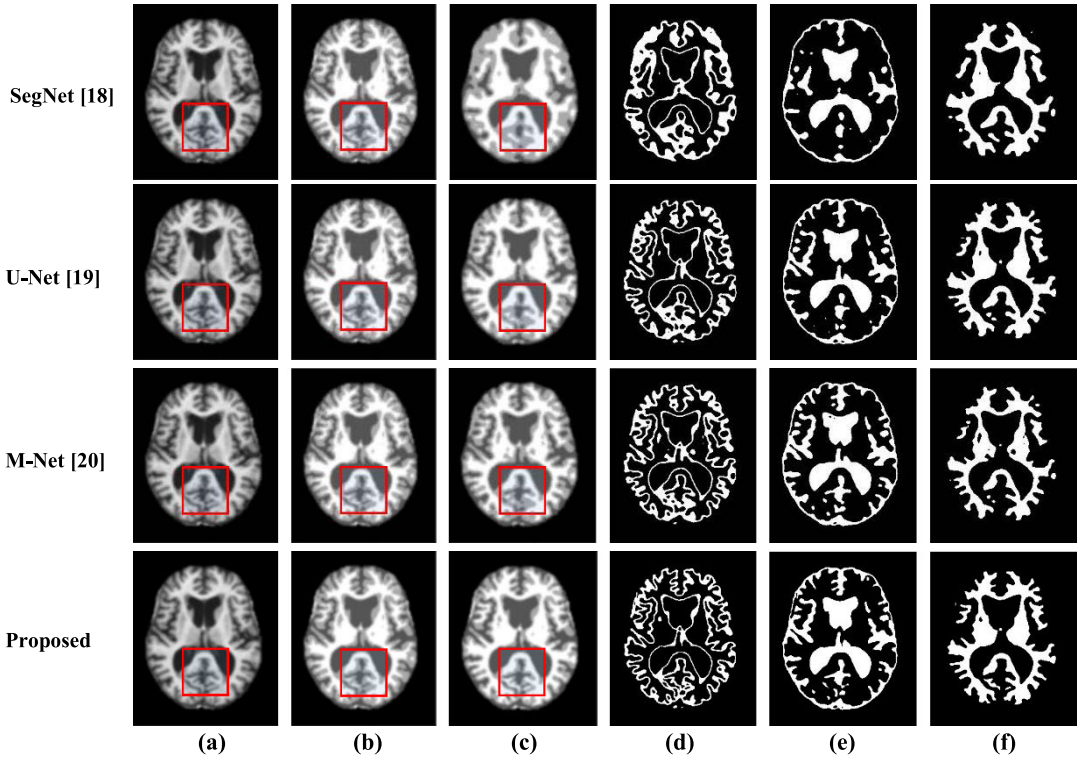

FIGURE 6. Segmentation results for GM, CSF, and WM for existing methods and the proposed method on the OASIS dataset: (a) original input image; (b) ground-truth segmentation map; (c) segmentation results generated by SegNet, U-net, M-net, and the proposed method (top to bottom); (d) GM maps generated by SegNet, U-net, M-net, and the proposed method (top to bottom); (e) CSF maps generated by SegNet, U-net, M-net, and the proposed method (top to bottom); and (f) WM maps generated by SegNet, U-net, M-net, and the proposed method (top to bottom).

at the boundaries between WM and GM. The feature maps obtained by the encoder include low-level features and are concatenated with the same level of decoder features via skip connections. Additionally, low-level features are fused with high-level features from lower levels of decoder, leading to mismatches between low- and high-level features. The fusion of these two arguably incompatible sets of features can lead to inconsistencies throughout learning, thereby adversely affecting the prediction of boundaries between WM and GM. As a result, potential semantic gaps between concatenated low- and high-level features can be overserved in the results of U-net [32]. For M-net, it has been reported that the side paths

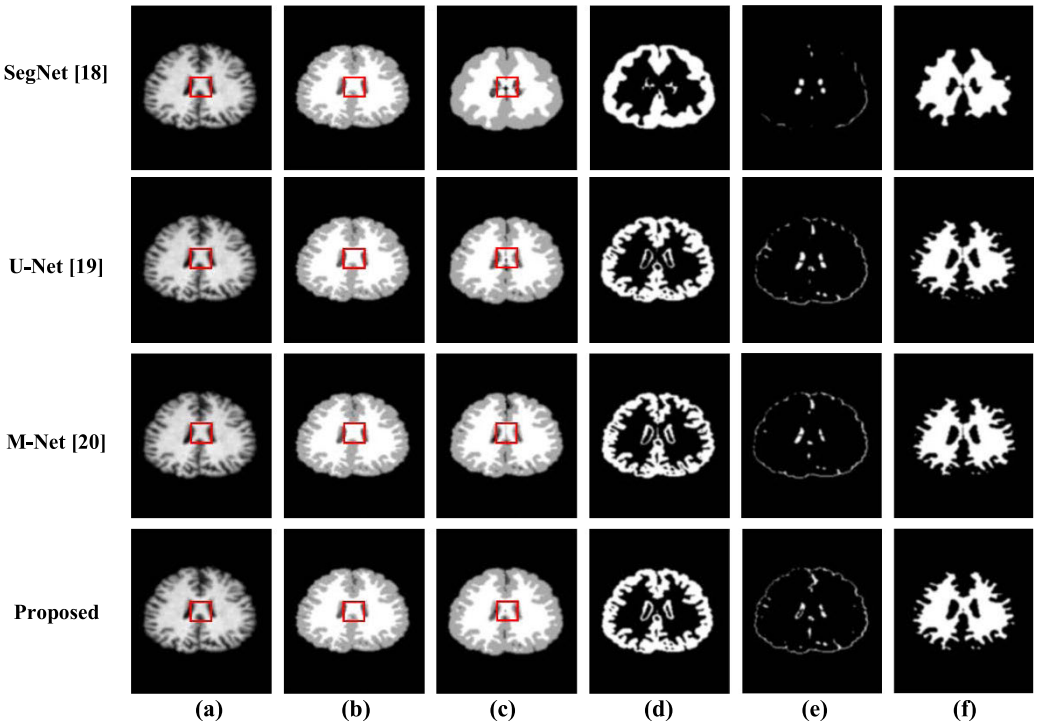

FIGURE 7. Segmentation results for GM, CSF, and WM for existing methods and the proposed method using IBSR dataset: (a) original input image; (b) ground-truth segmentation map; (c) segmentation results generated by SegNet, U-net, M-net, and the proposed method (top to bottom); (d) GM maps generated by SegNet, U-net, M-net, and the proposed method (top to bottom); (e) CSF maps generated by SegNet, U-net, M-net, and the proposed method (top to bottom); and (f) WM maps generated by SegNet, U-net, M-net, and the proposed method (top to bottom).

aid in capturing additional information compared to U-net, but M-net still fails to preserve accurate boundaries [20].

To overcome the drawbacks of the conventional methods discussed above, our proposed method adopts a non-overlapping patch-wise M-net architecture for brain MRI scan segmentation. By utilizing non-overlapping patches for each slice, the predictions for each patch can be performed separately, and local information can be preserved, resulting in enhanced segmentation performance with a similar degree of complexity compared to the original M-net architecture. Furthermore, the skip connections between pairs of convolutions in the proposed patch-wise M-net architecture aid in learning better features, and the side skip connections aid in preserving fine details.

Table 4 compares the proposed method to the conventional methods. One can observe that the segmentation accuracy of the proposed method significantly outperforms the conventional methods in terms of DSC, JI, precision, recall, specificity, and MSE for both the datasets. This can be attributed to the fact that compared to the conventional methods, the proposed method uses a more optimal up-sampling method. Because slices are divided into patches and predictions are made separately for each patch in the proposed method, fine local details can be preserved, resulting in better segmentation performance compared to the conventional methods, which use complete slices as inputs. It is noted that since the original ground-truth annotations in the IBSR dataset do not contain sulcal parts of CSF tissue unlike GM [44], the mean DSC values of CSF show relatively lower segmentation performance as compared to OASIS dataset. The results on the IBSR dataset with labeled sulcal CSF (SCSF) voxels as GM tissue to minimize the differences between segmentation masks and ground-truth labels are reported in several works [45], [46]. However, for fair comparisons, the experiments on the original IBSR dataset without additional annotation were performed for all methods in this paper.

We also investigated the influence of patch size on the segmentation performance of the proposed method. Experiments were performed with different patch sizes of 128×128 and 64×64. The results are listed in Table 5. In Table 5, one can see that when the patch size is 128×128, the proposed method required only 4.6 min of processing per subject. This value increases to 6.25 min for 64×64 patches. The greater the number of patches, the more resources required to train the network successfully and the more parameters that must be optimized during training. However, smaller patch sizes result in better performance in terms of the DSC. This is because smaller patch sizes can produce more training data for the network. Additionally, local regions can be reconstructed more precisely. It was determined that the 128×128 patch size represents an acceptable tradeoff between the DSC and the time required to process a single subject.

To evaluate the effectiveness of non-overlapping patches for brain MRI scan segmentation, we compared the

TABLE 4. Performance comparisons between the proposed method and conventional methods using oasis and IBSR datasets.

Parameter	WM				GM				CSF			
	SegNet [18]	U-net [19]	M-net [20]	Proposed method	SegNet [18]	U-net [19]	M-net [20]	Proposed method	SegNet [18]	U-net [19]	M-net [20]	Proposed method
OASIS Dataset												
Axial plane												
DSC (%)	74.36	92.18	93.36	**96.33**	75.93	90.32	92.05	**95.44**	71.67	89.82	91.64	**94.81**
JI (%)	59.18	85.50	87.56	**92.93**	61.20	82.35	85.27	**91.28**	55.85	81.53	84.57	**90.14**
Precision (%)	93.48	94.52	94.34	**97.59**	71.01	90.94	91.42	**94.62**	69.79	86.30	91.95	**94.88**
Recall (%)	81.73	89.96	92.41	**95.10**	81.58	89.71	92.68	**96.27**	73.65	91.64	93.33	**94.74**
Specificity (%)	97.45	99.70	99.68	**99.86**	97.01	99.19	99.22	**99.49**	98.65	98.39	99.66	**99.80**
Coronal plane												
DSC (%)	82.12	93.14	94.35	**95.68**	78.21	92.25	93.53	**94.96**	74.03	89.53	91.46	**93.34**
JI (%)	69.23	87.20	89.63	**91.35**	64.16	85.45	87.85	**88.24**	61.26	82.36	84.25	**87.40**
Precision (%)	91.87	94.12	95.43	**96.19**	92.61	93.16	94.19	**95.14**	83.34	86.08	91.65	**93.48**
Recall (%)	87.46	92.26	93.26	**94.34**	85.28	91.28	92.39	**93.06**	81.53	92.14	93.24	**94.74**
Specificity (%)	98.20	99.45	99.71	**99.79**	96.95	98.41	99.55	**99.61**	96.26	98.22	99.39	**99.67**
Sagittal Plane												
DSC (%)	82.42	92.44	93.95	**95.18**	80.23	91.13	92.56	**94.45**	77.45	88.63	92.15	**93.17**
JI (%)	69.54	86.34	87.65	**90.78**	67.57	83.26	85.69	**87.96**	63.51	81.26	85.36	**87.11**
Precision (%)	93.11	93.97	94.79	**95.89**	92.14	92.36	93.43	**95.05**	84.78	85.89	92.32	**93.09**
Recall (%)	87.72	91.37	92.21	**93.98**	86.61	91.08	91.27	**92.89**	82.14	91.67	93.63	**94.15**
Specificity (%)	98.29	99.16	99.43	**99.65**	97.13	98.12	99.23	**99.28**	97.43	98.11	99.48	**99.55**
IBSR Dataset												
Axial plane												
DSC (%)	72.86	89.45	90.35	**91.42**	75.63	91.53	92.20	**93.21**	68.42	84.34	84.95	**85.35**
JI (%)	65.34	81.56	82.49	**83.15**	67.32	85.41	86.15	**87.22**	59.32	75.85	75.98	**76.16**
Precision (%)	89.89	91.26	92.69	**94.16**	85.31	94.52	94.98	**95.59**	82.56	89.76	90.43	**87.89**
Recall (%)	84.67	90.34	91.53	**93.41**	89.27	93.65	93.87	**94.37**	87.78	91.38	92.39	**89.67**
Specificity (%)	97.11	99.09	99.45	**99.72**	98.43	99.49	99.47	**99.83**	96.43	99.65	99.46	**99.52**
Coronal plane												
DSC (%)	70.15	88.45	89.36	**90.53**	73.65	90.23	91.45	**92.33**	66.54	83.65	84.15	**84.72**
JI (%)	62.35	79.38	80.14	**81.45**	65.42	83.56	84.16	**85.23**	57.32	76.86	76.94	**77.21**
Precision (%)	88.19	90.75	91.42	**92.65**	85.63	88.49	89.39	**91.76**	84.42	88.75	91.40	**92.86**
Recall (%)	83.37	89.95	90.36	**91.48**	88.21	87.17	88.18	**90.13**	81.79	86.47	90.74	**91.31**
Specificity (%)	97.56	99.34	99.45	**99.69**	98.89	98.46	99.10	**99.28**	96.75	97.76	98.25	**99.18**
Sagittal Plane												
DSC (%)	71.53	86.84	87.32	**89.68**	74.62	89.46	90.44	**91.89**	65.49	80.75	81.19	**83.57**
JI (%)	63.41	78.63	79.62	**81.34**	66.85	81.53	82.15	**83.65**	54.86	73.96	74.10	**75.32**
Precision (%)	86.59	91.44	91.96	**92.88**	88.59	90.35	91.49	**93.74**	82.49	88.46	89.92	**91.45**
Recall (%)	84.31	88.28	89.76	**91.10**	87.31	88.49	90.24	**91.62**	80.75	87.29	87.99	**90.62**
Specificity (%)	97.40	98.34	98.69	**99.01**	97.64	98.94	99.03	**99.43**	96.13	98.82	99.04	**99.16**

MSE

	SegNet [18]	U-net [19]	M-net [20]	Proposed method
OASIS	0.021	0.008	0.006	**0.003**
IBSR	0.013	0.009	0.007	**0.005**

TABLE 5. Effects of different patch sizes on segmentation performance in terms of the DSC, JI, and runtime performance for processing a test subject.

Models	DSC			JI			Time (min)
	WM	GM	CSF	WM	GM	CSF	
Patch size: 128×128	96.33	95.44	94.81	92.93	91.28	90.14	4.67
Patch size: 64×64	96.35	95.48	94.89	92.95	91.37	90.28	6.25

TABLE 6. Parameter comparisons for automatic segmentation of brain MRI images based on M-net models.

No.	Parameters	M-net [20]	Overlapping patch-wise M-net	Proposed method
1	Input size	256×256	128×128	128×128
2	Training set	30 subjects	30 subjects	30 subjects
3	Testing set	20 subjects	20 subjects	20 subjects
4	Number of patches	1	32 (slide of 8 pixels)	4
5	Number of epochs	10	10	10
6	Computation time	4.5 h	40 h	5.8 h
7	DSC	0.92	0.96	0.95
8	JI	0.86	0.91	0.90

performance of the proposed method to those of a conventional M-net and overlapping patch-wise M-net. As discussed previously, complete slices are used for training and testing inputs in the conventional M-net architecture, whereas partitioned patches are used for the overlapping and proposed non-overlapping patch-wise M-net architectures. Table 6 lists

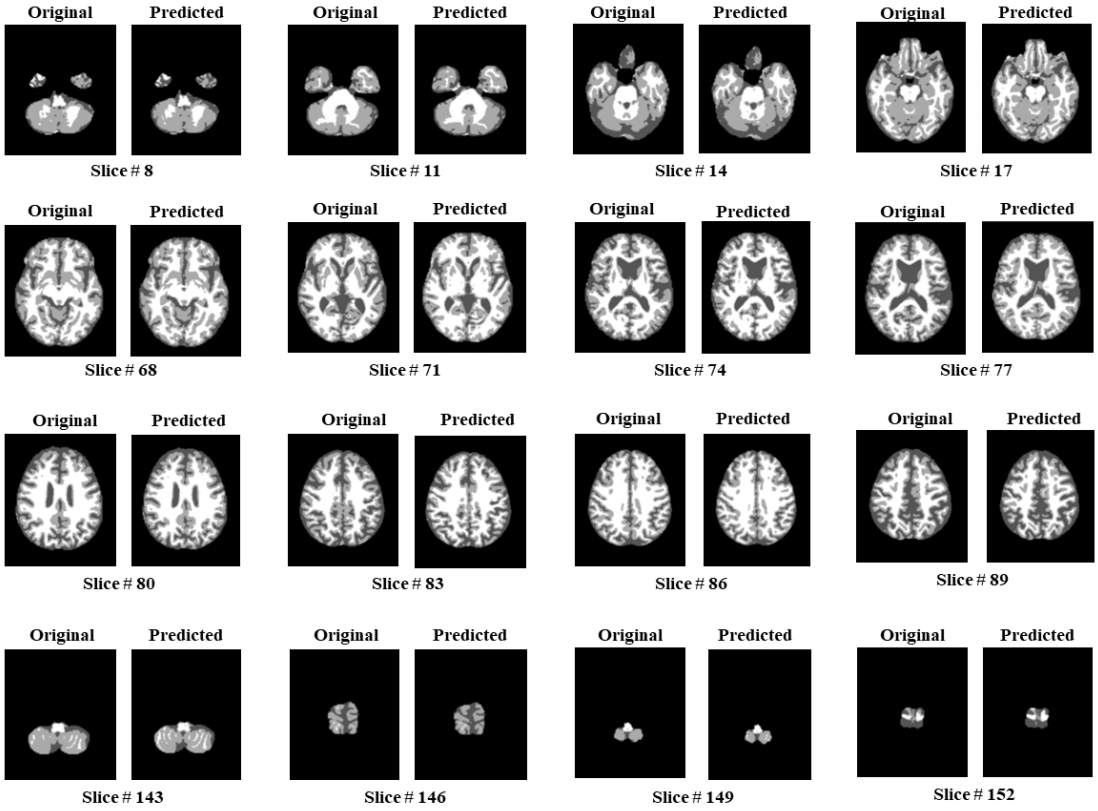

FIGURE 8. Segmentation images generated using central and non-central slices.

the experimental setups and results for the three models under consideration. For the overlapping M-net, overlapping patches of the same size as those used in the proposed method (but with a slide of eight pixels) were considered. Although the classification accuracies show almost identical results compared to the other methods with 0.96 for the DSC and 0.91 for the JI, the output predicted images were not reconstructed accurately. This is because multiple convolution operations were performed over the same pixel elements. Additionally, the overlapping patch-wise M-net requires significantly more computation time because the network must be trained separately for each overlapping patch. The overlapping patch-wise M-net requires 40 h of training using the images considered in our experimental setup, whereas the proposed method only requires 5.8 h of training.

Fig. 8 illustrates the effectiveness of central and non-central slices for segmentation using the proposed method. One can observe that the first and last rows contain non-central slices, which provide less information compared to the central slices in the middle two rows. Although non-central slices contain less information, the proposed method is still capable of accurately segmenting images when using non-central slices.

The influence of different multi-scale convolutional kernels and different patch sizes for the segmentation of brain MRI is investigated. We tested five sets of convolutional kernels for the encoder, and decoder {(1 × 1, 1×1), (3 × 3, 3 × 3), (3 × 3, 6 × 6), (6×6, 12 × 12), (1 × 1, 3 × 3)}, which are applied to three different patch sizes: 32 × 32, 64 × 64, and 128 × 128 on the OASIS dataset. Table 7 shows that convolutional kernels with smaller (1 × 1, 1 × 1) and larger (6 × 6, 12 × 12) sizes achieve relatively lower performances. It is due to the fact that the features extracted by the smaller kernel can be highly local-adaptive and may not have more general characteristics of the image. In contrast, larger kernels tend to have a lower capability to share information among neighboring pixels, and a large number of weights requires significantly higher computational time. In addition, since the features extracted with the larger kernel size are prone to be generic and spread across the image, this might fail to capture smaller, complex features in the image. To overcome these problems, we use multi-scale kernels with an optimal size of 1 × 1 convolutional kernel to extract finer feature maps in the encoder, and 3 × 3 convolutional kernels to provide larger receptive fields in the decoder, respectively. It can be observed in Table 7 that the smaller patch sizes (32 × 32) with multi-scale kernels ((1 × 1, 3×3)) show better performances in terms of the DSC scores. This is because the smaller patch size can produce more training data for the network to train. Moreover, the local regions will be reconstructed more precisely. Besides, when identical multi-scale kernels are used, computational times for the patch size 32 × 32 takes three times more than one of the patch size 128 × 128. Therefore,

TABLE 7. DSC scores with respect to different patch sizes with multi-scale convolutional kernels for oasis dataset.

Different kernel sizes (encoder, decoder)	WM		GM		CSF		MSE	Computational complexity
	DSC (%)	JI (%)	DSC (%)	JI (%)	DSC (%)	JI (%)		
Patch size: 32×32								
Patch-wise M-net (1×1, 1×1)	95.17	90.79	93.72	88.18	92.39	85.87	0.005	11h, 10m
Patch-wise M-net (3×3, 3×3)	96.19	92.66	95.14	90.73	94.60	89.75	0.004	22h, 17m
Patch-wise M-net (3×3, 6×6)	95.53	91.44	94.56	89.68	94.42	89.43	0.004	44h, 21m
Patch-wise M-net (6×6,12×12)	95.85	92.03	94.85	90.21	93.78	88.30	0.004	110h, 40m
Patch-wise M-net (1×1, 3×3)	**96.41**	**92.97**	**95.93**	**91.65**	**94.98**	**90.51**	**0.003**	17h, 11m
Patch size: 64×64								
Patch-wise M-net (1×1, 1×1)	94.93	90.36	93.30	87.45	92.50	86.05	0.005	10h, 5m
Patch-wise M-net (3×3, 3×3)	96.15	92.30	94.87	90.25	94.33	89.28	0.004	15h, 17m
Patch-wise M-net (3×3, 6×6)	95.35	91.15	94.43	90.07	93.19	89.08	0.005	18h, 20m
Patch-wise M-net (6×6, 12×12)	94.53	89.14	92.62	85.58	90.41	81.57	0.007	75h, 35m
Patch-wise M-net (1×1, 3×3)	**96.35**	**92.95**	**95.48**	**91.37**	**94.89**	**90.28**	**0.003**	12hh, 12m
Patch size: 128×128								
Patch-wise M-net (1×1, 1×1)	94.99	90.45	94.30	89.22	93.95	88.59	0.004	9h, 7m
Patch-wise M-net (3×3, 3×3)	96.01	92.34	94.85	90.21	93.82	88.37	0.004	12h,11m
Patch-wise M-net (3×3, 6×6)	93.75	88.23	90.81	83.18	89.46	80.93	0.007	6h, 5m
Patch-wise M-net (6×6, 12×12)	93.64	88.04	91.82	84.88	89.36	80.77	0.007	18h, 9m
Patch-wise M-net (1×1, 3×3)	**96.33**	**92.93**	**95.44**	**91.28**	**94.81**	**90.14**	**0.003**	5h, 5m

h: hours, m: minutes

it is concluded that the multi-scale kernels (1 × 1, 3×3) with 128 × 128 patch size shows a decent tradeoff between the DSC score and computational time in the proposed method.

IV. CONCLUSION

The automatic segmentation of brain MRI scans is important for the diagnosis of neurological disorders. In this paper, we presented a patch-wise M-net architecture with multi-scale dilated convolutional kernels that can achieve better performance compared to conventional methods. The partitioning with individual uniform-sized patches for a slice can better reflect local details by predicting the information for each patch. Besides, the model can be trained better with the local information in the uniform patch-wise method. High segmentation accuracy can be obtained by using the uniform-partitioned patches. Eventually, complete information of the slices can be used as training data, thus resulting in robust segmentation performances with detailed local information. Furthermore, the proposed model has the ability to make predictions for multi-class segmentation, unlike conventional methods, which were developed to handle binary segmentation problems. The proposed method performs well under various imaging conditions and yields good results depending on image resolution. Our method demonstrates significant improvement in terms of popular metrics, such as the DSC and JI, for the segmentation of brain MRI scans into CSF, GM, and WM regions, exhibiting average DSC and JI values of 0.95 and 0.90. These values represent improvements of approximately 3% (DSC), and 6% (JI) compared to the state-of-the-art methods.

REFERENCES

[1] R. Smith-Bindman, D. L. Miglioretti, E. Johnson, C. Lee, H. S. Feigelson, M. Flynn, R. T. Greenlee, R. L. Kruger, M. C. Hornbrook, D. Roblin, L. I. Solberg, N. Vanneman, S. Weinmann, and A. E. Williams, "Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010," *JAMA*, vol. 307, no. 22, pp. 2400–2409, Jun. 2012.

[2] C. Xu, D. L. Pham, M. E. Rettmann, D. N. Yu, and J. L. Prince, "Reconstruction of the human cerebral cortex from magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, no. 6, pp. 467–480, Jun. 1999.

[3] D. MacDonald, N. Kabani, D. Avis, and A. C. Evans, "Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI," *NeuroImage*, vol. 12, no. 3, pp. 340–356, Sep. 2000.

[4] J. de Bresser, M. P. Portegies, A. Leemans, G. J. Biessels, L. J. Kappelle, and M. A. Viergever, "A comparison of MR based segmentation methods for measuring brain atrophy progression," *NeuroImage*, vol. 54, no. 2, pp. 760–768, Jan. 2011.

[5] S. Valverde, A. Oliver, and X. Llado, "A white matter lesion-filling approach to improve brain tissue volume measurements," *NeuroImage, Clin.*, vol. 6, pp. 86–92, Aug. 2014.

[6] N. Kovacevic, N. J. Lobaugh, M. J. Bronskill, B. Levine, A. Feinstein, and S. E. Black, "A robust method for extraction and automatic segmentation of brain images," *NeuroImage*, vol. 17, no. 3, pp. 1087–1100, Nov. 2002.

[7] E. Roura, A. Oliver, M. Cabezas, J. C. Vilanova, À. Rovira, L. Ramió-Torrentà, and X. Lladó, "MARGA: Multispectral adaptive region growing algorithm for brain extraction on axial MRI," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 655–673, Feb. 2014.

[8] M. Cabezas, A. Oliver, E. Roura, J. Freixenet, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, and X. Lladó, "Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding," *Comput. Methods Programs Biomed.*, vol. 115, no. 3, pp. 147–161, Jul. 2014.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sep. 1995.

[10] L. Breiman and A. Cutler. (2007). *Random Forests-Classification Description: Random Forests*. [Online]. Available: http://www.berkeley.edu/users/breiman/RandomForests/cc_home.htm

[11] P. J. Werbos and P. John, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Cambridge, MA, USA: Harvard Univ., 1974.

[12] P. M. Thompson, M. S. Mega, and A. W. Toga, "Disease-specific brain atlases," in *Brain Mapping: The Disorders*. New York, NY, USA: Academic, 2000, pp. 131–177.

[13] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, Mar. 2015.

[14] A. Brébisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2015, pp. 20–28.

[15] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Isgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.

[16] Z. Kong, J. Luo, S. Xu, and T. Li, "Automatical and accurate segmentation of cerebral tissues in fMRI dataset with combination of image processing and deep learning," *Proc. SPIE*, vol. 10485, Feb. 2018, Art. no. 104850A.

[17] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1342–1345.
[18] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Springer, 2015, pp. 234–241.
[20] V. Adiga and J. Sivaswamy, "FPD-M-net: Fingerprint image denoising and inpainting using M-net based convolutional neural networks," in *Inpainting and Denoising Challenges*. Cham, Switzerland: Springer, 2019.
[21] F. Milletari, S. Ahmadi, C. Kroll, A. Plate, V. Rozanski, and J. Maiostre, "HoughCNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound," *Comput. Vis. Image Understand.*, vol. 164, pp. 92–102, 2017.
[22] 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, 1998.
[23] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
[24] P. Jaccard, "The distribution of the flora in the alpine zone.1," *New Phytologist*, vol. 11, no. 2, pp. 37–50, Feb. 1912.
[25] B. Lee, W. Ellahi, and J. Y. Choi, "Using deep CNN with data permutation scheme for classification of Alzheimer's disease in structural magnetic resonance imaging (sMRI)," *IEICE Trans. Inf. Syst.*, vol. E102.D, no. 7, pp. 1384–1395, Jul. 2019.
[26] R. Kumar Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015, *arXiv:1505.00387*. [Online]. Available: http://arxiv.org/abs/1505.00387
[27] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587
[28] F. Zhang, N. Cai, G. Cen, F. Li, H. Wang, and X. Chen, "Image super-resolution via a novel cascaded convolutional neural network framework," *Signal Process., Image Commun.*, vol. 63, pp. 9–18, Apr. 2018.
[29] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2881–2890.
[30] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122
[31] X. Fu, N. Cai, K. Huang, H. Wang, P. Wang, C. Liu, and H. Wang, "M-net: A novel U-net with multi-stream feature fusion and multi-scale dilated convolutions for bile ducts and hepatolith segmentation," *IEEE Access*, vol. 7, pp. 148645–148657, 2019.
[32] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.
[33] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *J. Cognit. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, Sep. 2007.
[34] S. Bao and A. Chung, "Multi-scale structured CNN with label consistency for brain MR image segmentation," *Comput. Methods Biomechanics Biomed. Eng., Imag. Vis.*, vol. 6, pp. 113–117, Jan. 2016.
[35] C. Wachinger, M. Reuter, and T. Klein, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, Apr. 2018.
[36] S. Piao and J. Liu, "Accuracy improvement of UNet based on dilated convolution," *J. Phys., Conf. Ser.*, vol. 1345, Nov. 2019, Art. no. 052066.
[37] (2012). *IBSR Dataset*. [Online]. Available: https://www.nitrc.org/projects/ibsr
[38] J. Wei, Y. Xia, and Y. Zhang, "M3Net: A multi-model, multi-size, and multi-view deep neural network for brain magnetic resonance image segmentation," *Pattern Recognit.*, vol. 91, pp. 366–378, Jul. 2019.
[39] H. Jia, Y. Xia, Y. Song, D. Zhang, H. Huang, Y. Zhang, and W. Cai, "3D APA-net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 447–457, Feb. 2020.
[40] J. Bernal, K. Kushibar, M. Cabezas, S. Valverde, A. Oliver, and X. Llado, "Quantitative analysis of patch-based fully convolutional neural networks for tissue segmentation on brain magnetic resonance imaging," *IEEE Access*, vol. 7, pp. 89986–90002, 2019.
[41] J. Wu, Y. Zhang, K. Wang, and X. Tang, "Skip connection U-Net for white matter hyperintensities segmentation from MRI," *IEEE Access*, vol. 7, pp. 155194–155202, 2019.
[42] P. Li, Y. Zhao, Y. Liu, Q. Chen, F. Liu, and C. Gao, "Temporally consistent segmentation of brain tissue from longitudinal MR data," *IEEE Access*, vol. 8, pp. 3285–3293, 2020.
[43] C.-C.-C. Chen, J.-W. Chai, H.-C. Chen, H. C. Wang, Y.-C. Chang, Y.-Y. Wu, W.-H. Chen, H.-M. Chen, S.-K. Lee, and C.-I. Chang, "An iterative mixed pixel classification for brain tissues and white matter hyperintensity in magnetic resonance imaging," *IEEE Access*, vol. 7, pp. 124674–124687, 2019.
[44] S. Valverde, A. Oliver, M. Cabezas, E. Roura, and X. Lladó, "Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations," *J. Magn. Reson. Imag.*, vol. 41, no. 1, pp. 93–101, Jan. 2015.
[45] P.-L. Bazin and D. L. Pham, "Topology-preserving tissue classification of magnetic resonance brain images," *IEEE Trans. Med. Imag.*, vol. 26, no. 4, pp. 487–496, Apr. 2007.
[46] S. Roy, A. Carass, P.-L. Bazin, S. Resnick, and J. L. Prince, "Consistent segmentation using a rician classifier," *Med. Image Anal.*, vol. 16, no. 2, pp. 524–535, Feb. 2012.
[47] S. Saif, S. Rahman, and N. Sharmin, "One-class SVM for human brain segmentation," in *Proc. 22nd Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dhaka, Bangladesh, 2019, pp. 1–7.
[48] L. Sun, W. Shao, M. Wang, D. Zhang, and M. Liu, "High-order feature learning for multi-atlas based label fusion: Application to brain segmentation with MRI," *IEEE Trans. Image Process.*, vol. 29, pp. 2702–2713, 2020.
[49] Y. Chen, D. Ruan, J. Xiao, L. Wang, B. Sun, R. Saouaf, W. Yang, D. Li, and Z. Fan, "Fully automated multi-organ segmentation in abdominal magnetic resonance imaging with deep neural networks," 2019, *arXiv:1912.11000*. [Online]. Available: http://arxiv.org/abs/1912.11000
[50] M. Ahmad, D. Ai, G. Xie, S. F. Qadri, H. Song, Y. Huang, Y. Wang, and J. Yang, "Deep belief network modeling for automatic liver segmentation," *IEEE Access*, vol. 7, pp. 20585–20595, 2019.

NAGARAJ YAMANAKKANAVAR (Member, IEEE) received the B.E. degree from the Sri Jayachamarajendra College of Engineering, Mysore, India, the M.Tech. degree from the Sri Bhagawan Mahaveer Jain College of Engineering, Bengaluru, India, and the Ph.D. degree from the National Institute of Technology Karnataka, Surathkal, India. He was an Assistant Professor with Dayanand Sagar University, Bengaluru, from 2018 to 2019. He was a Postdoctoral Researcher with Chosun University, from 2019 to 2020. He is currently working as an Assistant Professor with the Department of Information and Communication Engineering, Chosun University, South Korea. His research interest includes medical image processing.

BUMSHIK LEE (Member, IEEE) received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea and the M.S. and Ph.D. degrees in information and communications engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. He was a Research Professor with KAIST, South Korea, in 2014, and a Postdoctoral Scholar with the University of California, San Diego (UCSD), CA, USA, from 2012 to 2013. He was a Principal Engineer with the Advanced Standard Research and Development Laboratory, LG Electronics, Seoul. He is currently an Assistant Professor with the Department of Information and Communication Engineering, Chosun University, South Korea. His research interests include medical image processing, video processing, video compression, and video security.