IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 17, 2021, accepted April 4, 2021, date of publication April 8, 2021, date of current version April 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071754

PMED-Net: Pyramid Based Multi-Scale Encoder-Decoder Network for Medical Image Segmentation

ABBAS KHAN[1,2], HYONGSUK KIM[1,2], (Senior Member, IEEE), AND LEON CHUA[3], (Life Fellow, IEEE)

[1]Division of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea
[2]Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju 54896, Republic of Korea
[3]Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720, USA

Corresponding author: Hyongsuk Kim (hskim@jbnu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government under Grant NRF 2019R1A2C1011297 and Grant NRF-2019R1A6A1A09031717, and in part by the U.S. Air Force Office of Scientific Research under Grant FA9550-18-1-0016.

ABSTRACT A pyramidical multi-scale encoder-decoder network, namely PMED-Net, is proposed for medical image segmentation. Different variants of encoder-decoder networks are in practice for segmenting the medical images and U-Net is the most widely used one. However, the existing architectures for segmenting medical images have millions of parameters that require enormous computations which results in memory and cost-inefficiency. To overcome such limitations, we come up with the idea of training small networks in a cascaded form for coarse-to-fine prediction. The proposed adaptive network is extended up to six pyramid levels, and at each level, features are extracted at different scales of the input image. Each lightweight encoder-decoder network is trained independently to minimize loss, where succeeding level networks further refine the prior predictions. Evaluation and comparison of our architecture were performed on four different publicly available medical image segmentation datasets: International Skin Imaging Collaboration (ISIC) challenge 2018 dataset, brain tumor dataset, nuclei dataset, and X-ray dataset. The experimental results of the PMED-Net are either better or on par with other state-of-the-art networks in terms of IoU, F1-Score, and sensitivity metrics. Moreover, PMED-Net is efficient in terms of parameterized complexity as it has 1/21.3, 1/21.1, 1/14.0, 1/11.6, 1/11.2, 1/6.64, and 1/4.95 times fewer parameters than SegNet, U-Net, BCDU-Net, CU-Net, FCN-8s, ORED-Net, and MultiResUNet respectively. The pre-trained models, datasets information, and implementation details are available at *https://github.com/kabbas570/Pyramid-Based-Encoder-Decoder.*

INDEX TERMS Convolutional neural networks, encoder-decoder architecture, medical image processing, semantic segmentation.

I. INTRODUCTION

Medical image processing is one of the core areas that is investigated using deep learning [1]. With the advent of Artificial Neural Networks (ANNs), deep learning is providing state-of-the-art performance for Computer-Aided Diagnosis (CAD) systems [2]–[4] due to its robustness and generalizability. The goal of medical image analysis is to provide doctors with a precise interpretation of medical images, as this is important for diagnosing any disease in the early stages. Manual analysis of medical images is a cumbersome and tedious task, so there is a dire need for developing computer algorithms to automate the diagnosis process [5]. In medical image processing, researchers need to solve various problems such as classification [6], tracking [7], detection [8], and segmentation [9] to analyze the pathologies of a disease within the candidate medical images.

After the emergence of AlexNet [10], convolutional neural networks (CNNs) have emerged as a standard for solving these problems. In CNNs, the aim is to train algorithms to visualize and recognize patterns in images with minimum human intervention. This capability of CNNs

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

has resulted in their application to all fields of computer vision, from self-driving cars [11] to facial recognition [12], bioinformatics [13], [14], stereo vision [15], 3D scene reconstruction [16], and healthcare [17] with no exception. In medical image processing, different imaging technologies like magnetic resonance imaging (MRI), microscopy ultrasound, dermoscopy, X-ray, and computer tomography (CT) are used to capture images of the human body [18]. The goal of CAD system is to analyze these images, and produce accurate and quick diagnostic reports for medical specialists so that patients can receive immediate and effective treatment.

During recent few years, deep neural networks (DNNs) have replaced all classical and hand-engineered features based recognition and segmentation methods [19]. However, supervised deep learning-based models are data-hungry that require an extensive amount of training data (with well-defined ground truths) [20]. Procuring large-amounts of training data is often impractical and infeasible (especially for the rarely occurring diseases) [21]. Furthermore, obtaining medical data faces challenges related to logistics approvals regarding patient privacy, storage problem, getting data from proprietary ancestral raw files, and ground truth generation. Data augmentation strategies can provide an alternative approach to meet this data requirement. However, it results in a compromised training performance due to presence of similar textures, shapes, and correlated features [22].

Image segmentation refers to the process of identifying images at pixel level [23]. For medical images, segmentation is very crucial in many applications for extracting the region of interest (ROI). It can divide an image into different ROIs to give a clear interpretation of a diseased organ, tissues, or cells [24]. For illustration, Figure 1 shows examples of four publicly available medical image segmentation datasets used for the experiments conducted in this paper.

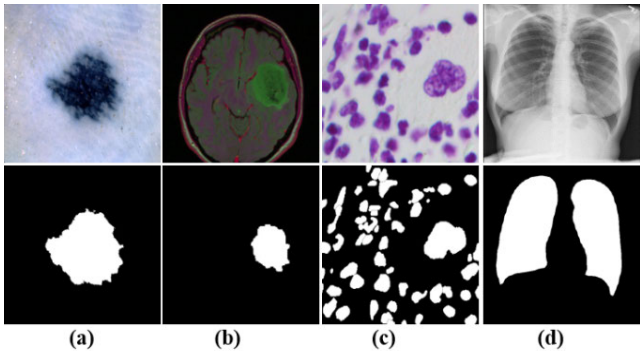

FIGURE 1. Medical image segmentation: the first row is the inputs and the second is the ground truths for (a) ISIC, (b) Brain tumor, (c) Nuclei, and (d) X-ray datasets, respectively.

In this study, we propose a small and efficient pyramid based multi-scale encoder-decoder network called PMED-Net for medical image segmentation. The main contributions of this work can be summarized as follows.

- An architecture that employs small pyramid based encoder-decoder networks in a cascaded fashion is proposed for extracting complex lesions and biomarkers contained within medical images by leveraging their multi-scale feature representations.
- We address the adaptive techniques of network size to achieve an optimal trade-off between performance and computations.
- Features of different scales are extracted with the use of pyramid-based encoder decoder networks.
- In terms of model parameters, the proposed architecture is 95.30% smaller than SegNet [25], 95.27% than U-Net [26], 92.90% than BCDU-Net [27], 91.42% than CU-Net [28], 91.11% smaller than FCN-8s [29], 84.94% than ORED-Net [30], and 79.81% smaller than MultiResUNet [31].

II. RELATED WORK

Medical image segmentation had been investigated even before the advent of deep learning. The graph-cut method [32], thresholding based on histograms [33], and edge-region based techniques [34] were one of the popular schemes. To extract coherent regions, clustering algorithms were implemented [35], and for some cases in which images had an irregular pattern and boundaries, the fuzzy c-means algorithm (FCM) was introduced [36]. However, these clustered based methods were limited in their application due to their dependence on prior information about the number of clusters. A region growing based method was proposed in [37], which grouped the pixels with the same intensities in one region. However, the method is semi-automated, requiring human supervision for selecting the initial seed region.

In deep learning, most of the networks used for segmentation are encoder-decoder based topology [38]. All these networks follow the same strategy of increasing the depth and decreasing the spatial dimension of the feature maps in the encoder, while in the decoder, their mission is vice versa [39]. Fully convolutional network (FCN) [29], on the other hand, was the first model to extend the power of contemporary classification networks such as AlexNet [10], VGG [40], and GoogleNet [41], for segmentation task and performed much better than patch-based methods [42]. Furthermore, FCN offers variable stride rates to generate coarser-to-finer predictions (FCN-8s, FCN-16s, and FCN-32s). The encoder part is same for all FCN versions while the decoder differs in terms of the up-sampling stride. In FCN-8s and FCN-16s the predictions are added with previously pooled layers to make finer final predictions. SegNet is another popular encoder-decoder based architecture and it is widely used for semantic-segmentation [25]. The decoding part up-samples low-resolution feature maps using the pooling indices from the encoder to create sparse feature maps. One of the most famous networks for the segmentation of medical images is U-Net [26]. The network is similar to an encoder-decoder architecture, with skip connections from encoder to decoder side. In the encoder part, after two consecutive convolutions, a 2×2 max-pooling is performed to reduce the

TABLE 1. Comparison of the proposed architecture with other segmentation methods.

Methods	Strengths	Weaknesses
FCN-8s [29]	No fully connected layer, can be used for the input images of any size.	The decoder part is not effective as it does not utilizes the information from all pooling or subsampling layers.
SegNet [25]	The use of max-poling indices from encoder to decoder part significantly improves the segmentation performance.	Requires memory to store the indices of max-pooling operation. Decoder generates sparse feature maps of higher resolutions which may not be efficient for sparse datasets.
U-Net [26]	An end-to-end training, few training samples, and preservation of the full context of the input images.	Slow training in the middle layers of deeper models.
BCDU-Net [27]	Skip connections with Bi-directional ConvLSTM and densely connected convolutions in the encoder part provide better feature reuse.	Rigorous training for the Bi-directional ConvLSTM, input data is processed in both forward and backward paths.
CU-Net [28]	Configurations of additional skip connections among two U-Nets help to transmit high resolution information from shallow to deeper layers and loss weighted sampling scheme for class imbalance problem.	Addition of auxiliary supervision, branch supervision, and using two U-Nets make the overall architecture very large and slow.
ORED-Net [30]	The outer residual skip paths minimize the information loss and training time.	The model requires to be trained rigorously.
MultiResUNet [31]	Inception-like blocks [39] iteratively reuse spatial features across various scales and multi-resolution analysis.	Limited generality and reduced performance for the datasets with less instances.
PMED-Net (proposed)	Less number of parameters, and adaptive techniques of network size achieve the optimal trade-off between performance and computations.	Reuse of input and pre-processing of data at different scales.

feature map size. In decoder, it uses the up-sampling with a stride $= 2$, to recover the resolution [26].

A variety of modifications to the basic structure of U-Net have been proposed with the goal of improving its performance. By introducing a cascaded deep framework for brain tumor segmentation, CU-Net [28] could outperform the original U-Net architecture. However, with the addition of auxiliary supervision, branch supervision, and using two cascaded U-Net the overall architecture of CU-Net becomes very large and slow. A deep neural network called Bi-directional ConvLSTM U-Net with Densely connected convolutions (BCDU-Net) was proposed by [27] to utilize the Bi-directional ConvLSTM (BConvLSTM) and dense convolutions [43] with U-Net. BConvLSTM (replaced the skip connections of U-Net) and densely connected convolutions, in the encoding path were implemented for better feature reuse.

The ORED-Net architecture was proposed in [30], to segment eye regions in multiple classes. The network is based on SegNet [25], with non-identity residual connections from encoder to decoder side to reduce information loss. Ibtehaz and Rahman [30] designed an enhanced version of U-Net named MultiResUNet. Each pair of convolutional layers of U-Net were replaced with Inception-like blocks [39]. The authors claim that this strategy iteratively reuse spatial features across various scales. For multi-resolution analysis, 3×3, 5×5, and 7×7 kernels are used in parallel but this results in increasing memory requirements. To address this, they factorized the larger convolution filters into a series of 3×3 convolutions.

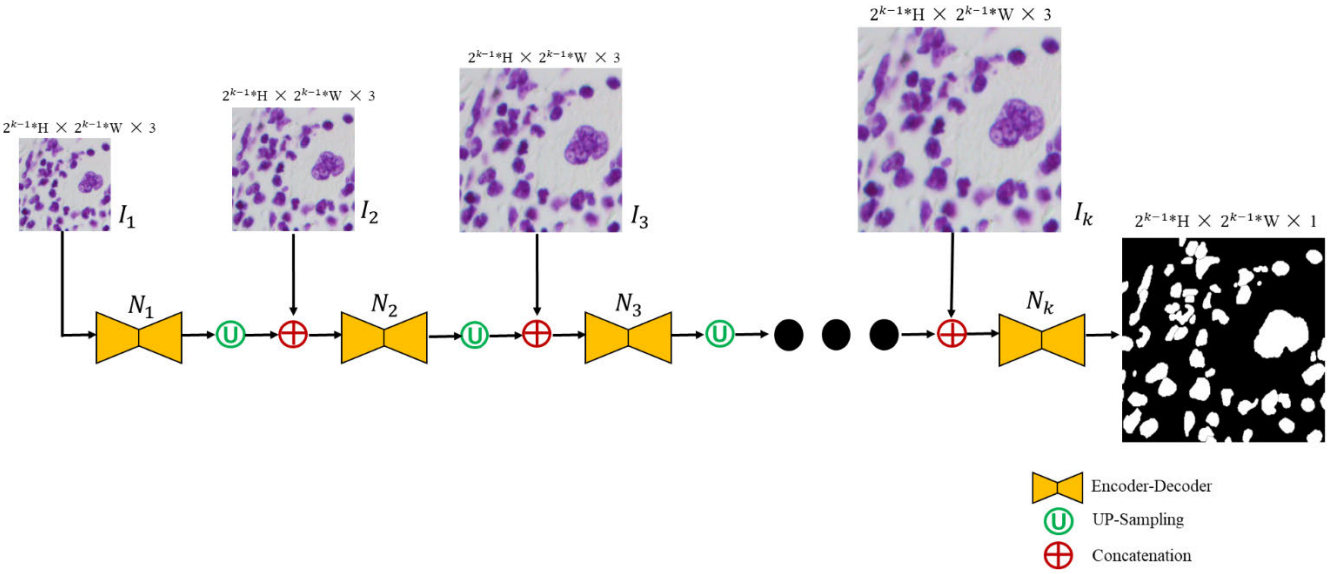

FIGURE 2. The proposed Pyramid Based Multi-Scale Encoder-Decoder Network (PMED-Net). Each encoder-decoder network or number of pyramid level N_k, k = {1, 2, 3, 4, 5, 6} computes a coarse segmentation map which is further refined by the next level network and so on. PMED-Net is adaptively extended up to six pyramid levels to extract the input image features at different scales.

Keeping in mind the loopholes and an excessive number of parameters in all these networks, we developed PMED-Net for medical image segmentation. The pyramid architecture enables the network to extract features at different scales, and cascaded models are employed for a coarse-to-fine prediction. Furthermore, we achieved superior performance compared to the other state-of-the-art models in terms of intersection over union (IoU), F1 scores, and sensitivity metrics on four publicly available medical image segmentation datasets.

The rest of the paper is organized as follows: Section III discusses the proposed framework, and evaluation metrics are enlisted in Section IV. The dataset details and ablation studies are included in Section V and VI, respectively. Finally, Section VII showcases the evaluation results, followed by concluding remarks in Section VIII.

III. PROPOSED ARCHITECTURE

The PMED-Net architecture shown in Figure 2, consists of six small encoder-decoder networks, where each network generates coarse predictions that are further refined at the next level. Predictions made by k^{th} level encoder-decoder network (N_k), are up-sampled with stride 2, concatenated with input image, and used as an input for N_{k+1} network. The proposed cascaded methodology enables the network to reuse the information iteratively and extract the features at different resolutions.

A. PYRAMID LEVELS

The proposed PMED-Net architecture, has six pyramid levels, which enables the model to extract the input image details at different scales. If the input image size is H × W then the corresponding input and ground truth sizes for the six pyramid levels, (level-1, level-2, level-3, level-4, level-5, and level-6) will be 2^{k-1}*H × 2^{k-1}*W, where k = (1, 2, 3, 4, 5, 6) for each corresponding level. We used bilinear interpolation for up-sampling, to match the dimensions. The intuitive strategy of these pyramid levels increases the network's ability to extract the details of smaller regions of interest at different scales from the images.

B. NETWORK STRUCTURE

At each pyramid level k, a small encoder-decoder network is trained independently to reduce the loss function. The predictions of this network are then up-sampled using bi-linear interpolation to match the dimensions of the next level pyramid because the next network input size is double compared to the preceding one. The up-sampled predictions are concatenated with the input image and further used by the next level network. The exceptional case is for the level-1 network, where the coarse estimation is not available, and the network uses only the images as the input. The reuse of input images at different scales improves the flow of information and finer details while generating the latent feature representations.

C. ENCODER-DECODER NETWORK

At each level k within the proposed scheme, a three-stage encoder-decoder network is trained independently to estimate the segmentation map. The detailed architecture of a single light-weighted encoder-decoder network is shown in Figure 3. The number of feature maps in the three encoder stages are increased as 16, 32, and 64. At each stage, we used two consecutive 3 × 3 convolutions with Rectified Linear Unit (ReLU) activation function [44] which is followed by max-pooling with stride = 2 and window size = 2 × 2 to decrease the spatial dimension. Starting from 16, after each

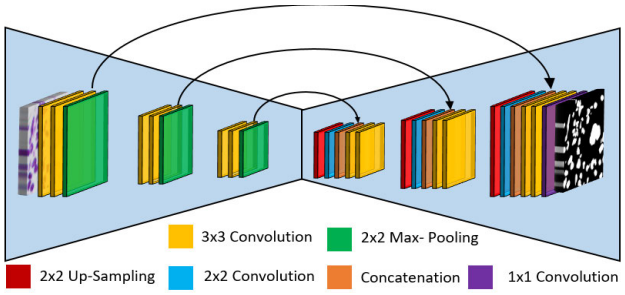

FIGURE 3. The architecture of a single encoder-decoder network.

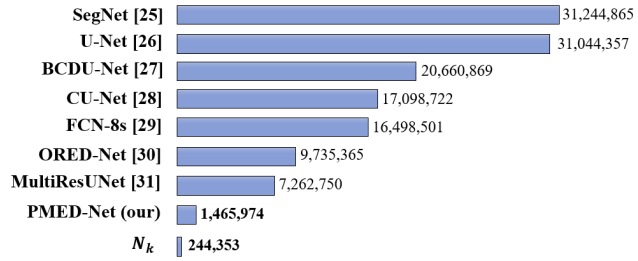

FIGURE 4. Different Models size in terms of parameters.

stage the number of feature maps is doubled and the maximum number of feature maps was limited to 64 in the encoder part to minimize the number of trainable parameters for each encoder-decoder network.

After third stage, the feature maps are up-sampled with stride 2 and directly fed to the decoder part. Before concatenating with the corresponding encoder feature maps, a 2×2 convolution halves the number of feature maps. Then two, 3×3 convolutions with ReLU activation function [43] are applied in decoder part.

This sequence of 2×2 up-sampling, 2×2 convolution, concatenation, and two, 3×3 convolutions are stacked together to match the dimensions with the respective encoder end (at each stage). The final segmentation map is generated by using a 1×1 convolution operation. The last convolution layer uses sigmoid activation function.

The proposed encoder-decoder is lightweight and relatively shallower network (compared to a conventional encoder-decoder) in terms of the feature maps and computational depth. Thus, employing it alone would produce coarser segmentation results. To overcome such problem, we stacked it in cascaded (as shown in Figure 2). The predictions (obtained from the proceeding model instance) are refined by concatenating the finer-scale feature representations resulting in superior performance compared to the existing framework while drastically reducing the computational requirements.

Overall, the PMED-Net architecture is quite small and has far less parameterized complexity as compared to other segmentation networks, shown in Figure 4. Each instance of the proposed network has only three stages with a much fewer number of feature maps, therefore the level-1 model (*i.e.*N_1) has 244,209 parameters, and the remaining each one N_k (k = 2, 3, 4, 5, 6) has 244,353 parameters. This slight increase in parameters happens because the networks at these levels also use the previous coarse predictions to refine it further in addition to the input images. In total, the proposed architecture comprises 1,465,974 parameters for its six pyramid level training.

D. NETWORK TRAINING

We trained each of the encoder-decoder networks (N_k) independently and compute the coarse prediction p_k, for the given input $I_k \oplus P_k$,

$$p_k = N_k\left(I_k \oplus P_k\right), \tag{1}$$

where P_k is the up-sampled prediction, I_k is the input image and the symbol$\oplus$ represents the concatenation. N_k represents the encoder-decoder network for k = (1, 2, 3, 4, 5, 6). Each N_k aims to reduce the dice-loss at different scales of the input. The network N_1 (shown in Figure 2) was trained with 48×48 images and for each subsequent pyramid level we doubled the resolution. This process was iterated until the level-6.

We trained each network using the prediction of the previous network as an initialization. All the networks were trained using Adam optimization [45] with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate was set to 1e-4 with a batch size of 2. The loss function is defined in term of the dice coefficient [46] as follow,

$$\text{Loss} = 1 - \left[\frac{2 * (Target \cap Prediction)}{(Target + Prediction)}\right]. \tag{2}$$

IV. EVALUATION METRICS

We used different evaluation metrics to evaluate and compare the performance of the PMED-Net architecture. First of all, we computed the confusion matrix between prediction and ground truth by calculating the number of true positives (TP), true negatives (TN), false positives (FP), and false-negatives (FN). These variables are used to measure the performance of the network in terms of intersection over union (IoU), F1-Score, and recall/sensitivity. IoU is the ratio of the area of overlap to the area of union between prediction and the ground truth. In terms of the variables of the confusion matrix, it is defined as,

$$IoU = \left(\frac{TP}{TP + FN + FP}\right). \tag{3}$$

Precision defines the ability of the model to locate relevant objects only, and recall evaluates true positive detections relative to all ground truths. In terms of the confusion matrix's variables, precision and recall are defined as;

$$Precision = \left(\frac{TP}{TP + FP}\right), \tag{4}$$

and

$$Recall = \left(\frac{TP}{TP + FN}\right). \tag{5}$$

F1-Score is the harmonic-mean between precision and recall and is expressed as,

$$F1\_Score = \left(\frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \right). \quad (6)$$

V. DATASETS

We used four different publicly available medical image segmentation datasets for our experiments conducted in the proposed study. For each dataset, a pixel-wise prediction was performed. The details for each dataset and the distribution of data for training, validation, and testing are described in this section.

A. ISIC 2018 (SKIN LESION ANALYSIS TOWARDS MELANOMA DETECTION) DATASET

This dataset was released by International Skin Imaging Collaboration (ISIC) in 2018 [47], [48]. It contains 2594 dermoscopy images that are available at: *https://challenge2018.isic-archive.com/.* The dataset consist of different challenging tasks like boundary segmentation, attribute detection, and disease classification. For all the experiments conducted in this paper, we used 1816 images for training, 258 for validation, and 520 for testing taken from task-1 of boundary segmentation.

B. BRAIN TUMOR DATASET

This dataset was obtained from The Cancer Imaging Archive (TCIA) which contained 110 cases of lower-grade glioma patients. The data has MR images along with FLAIR abnormality segmented masks. For the proposed experiments, we deleted the images without label pixels, and after data filtering, we left with 880 images along with their ground truths. These images were split into training (600), validation (100), and holdout test images (180). The dataset is available at the following link: *https://www.kaggle.com/mateuszbuda/lgg-mri-segmentation/version/1*.

C. X-RAY DATASET

The X-ray dataset used in this paper is composed of four different datasets, namely the Montgomery County chest X-ray set, Japanese Society of Radiological Technology (JSRT) dataset [49], the Shenzhen chest X-ray set [50]–[52], and the National Institutes of Health (NIH) Chest X-ray Dataset [53].

The Montgomery County X-ray dataset was obtained from the Department of Health and Human Services of Montgomery County, MD, USA. It contains 138 posterior-anterior X-rays from their tuberculosis control program. The set has 80 normal and 58 abnormal scans together with their corresponding ground truth masks available at: *http://openi.nlm.nih.gov/imgs/collections/NLM-MontgomeryCXRSet.zip*.

The JSRT dataset was created by JSRT and the Japanese Radiological Society (JRS) for different tasks such as computer-aided diagnosis, image compression, and picture archiving. It consists of 247 images having 154 with and 93 without lung nodule. A pixel-wise lung annotation masks of 246 images are also provided for segmentation tasks at the following link: *http://db.jsrt.or.jp/eng.php.*

The Shenzhen dataset contains 662 X-ray images, among which 326 are normal and 336 X-rays have symptoms of Tuberculosis. Pixel-wise annotation masks of 566 instances are available at: *https://www.kaggle.com/yoctoman/shcxr-lung-mask* .

Overall, by combining the above three datasets, we had total of 950 images, divided into training (850) and validation set (100). For testing purposes, we used a different dataset named NIH dataset. One hundred samples have been taken from the NIH Chest X-ray dataset and annotated manually by [54] having various lung diseases. These images are available at: *https://nihcc.app.box.com/s/r8kf5xcthjvvvf6r7l1an99e1nj4080m.* This NIH dataset includes several severities of lung diseases that can evaluate the network performance and generalization capability more effectively.

D. NUCLEI DATASET

This dataset contains 670 segmented nuclei images and is provided by Data Science Bowl 2018 Segmentation Challenge available at: https://www.kaggle.com/gangadhar/nuclei-segmentation-in-microscope-cell-images. The images were captured under different conditions, magnification, and modalities (brightfield vs. fluorescence) and provided with a mask for each nucleus. As a pre-processing step, all the nuclei of single input image were combined together in one ground truth. Images were randomly assigned into a training set (510), validation set (60), and a testing set (100).

VI. ABLATION STUDIES

The effectiveness of the proposed PMED-Net architecture was also evaluated by comparing it with different ablated variants of it. We investigated two versions of PMED-Net in our ablation study: (1) Rather than using pyramids of different scales, we used the same size images in all six levels of the network (2) We increased or decreased the number of pyramid levels or encoder-decoder networks in architecture. This strategy is used for each dataset to experimentally determine the optimal tradeoff between performance and the computations.

For the case (1), we used NIH X-ray segmentation dataset in the ablation study for which the optimal performance is obtained at the fourth level of PMED-Net. By using images of the same sizes in all four levels, the network is unable to extract features at different scales. So, the performance of PMED-Net is lower as compared to using the pyramid of different scales in all four levels. The quantitative results of this experiment are listed in Table 2. In the implementation of *the* 'without pyramid', method all images are of the same size (384×384).

In the proposed method, we employed six pyramid levels to develop PMED-Net. The six levels are determined empirically. Although for some datasets the optimal performance is

TABLE 2. Performance comparison using different scales of images and same size images in four levels of the proposed architecture for NIH X-ray dataset segmentation.

Method	F1-Score	IoU	Sensitivity
With Pyramids	**0.8414**	**0.9139**	**0.9057**
Without Pyramids	0.7326	0.8456	0.7501

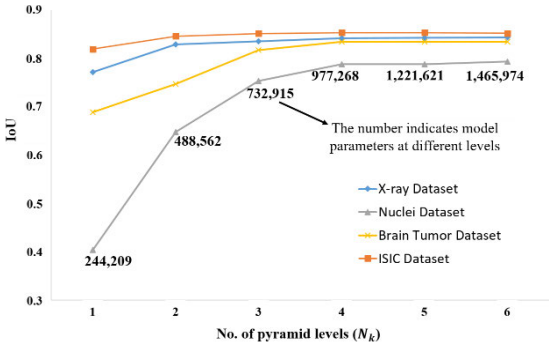

FIGURE 5. IoU plotted as a function of number of pyramid level or number of encoder-decoder networks N_k, set in cascade for each dataset.

obtained at the fourth or fifth level (as shown in Figure 5) and for other datasets performance improves up to the sixth level. We also extended the pyramid levels beyond six levels (*i.e.* up to seven and eight levels), but the performance gain was statistically insignificant. Accordingly, in this study, we set the maximum level to six and the minimum levels are dependent on the dataset itself.

We analyzed the performance of PMED-Net by changing the number of encoder-decoder networks in the architecture. A different number of encoder-decoder networks were cascaded, ranging from one to six for all four datasets. As the number of levels increased, improvement in the performance could be observed as shown in Figure 5.

The optimal number of levels depends upon the complexity of the dataset, and the boost in performance for the six pyramid levels is different for each dataset. For all four datasets, there was a significant improvement in IoU from level-1 to level-4, and by further increasing the number of levels, increment in IoU is quite slow. Thus, by considering the complexity of the dataset and a tradeoff between performance and computations we can adaptively change the network size. However, using more number of levels required longer training and testing time.

The PMED-Net architecture performs a coarse-to-fine prediction in a cascaded manner, as shown in Figure 6. At level-1, the network can identify the area of interest to be segmented. However, still it cannot distinguish between different nuclei, a higher level networks further refine these coarse predictions and segment each nucleus more clearly. For the sake of visualization, we scaled all predictions in Figure 6 to the same size.

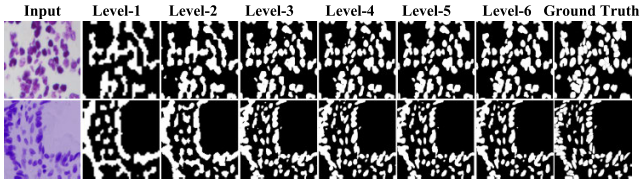

FIGURE 6. Predictions of the proposed architecture (from left to right in a coarse-to-fine way).

VII. RESULTS AND DISCUSSIONS

As introduced in Section V, we used four publically available medical image datasets, to evaluate and compare the performance of the proposed PMED-Net. All the experiments of the proposed study were conducted using a PC equipped with an NVIDIA Titan XP GPU and a Keras framework with Tensorflow backend.

A. ISIC SEGMENTATION

The quantitative analysis for the ISIC dataset between PMED-Net and the other comparing networks are listed in Table 3. For each evaluation index, the proposed PMED-Net outperforms other networks. While the most reliable results, in comparison with our network, were produced by CU-Net. However, they are 3%, 1.82%, and 1.2% less accurate than the proposed network in terms of the IoU, F1-Score, and sensitivity metric, respectively. The results of FCN-8s were the lowest, and this results from under segmentation of the area of interest.

TABLE 3. Segmentation results measured by IoU, F1-Score, and sensitivity metric for the ISIC dataset.

Network	IoU	F1-Score	Sensitivity
U-Net [26]	0.8154	0.8983	0.8693
SegNet [25]	0.8087	0.8942	0.8442
FCN-8s [29]	0.6929	0.8186	0.7107
BCDU-Net [27]	0.8128	0.8967	0.8577
CU-Net [28]	0.8207	0.9015	0.8904
ORED-Net [30]	0.8181	0.8999	0.8527
MultiResUNet [31]	0.7593	0.8631	0.7775
PMED-Net (proposed)	**0.8510**	**0.9197**	**0.9028**

For visualization purposes, the qualitative results are shown in Figure 7. The first column is the input, the second one is ground truth, and the proceeding columns are the segmentation maps generated by U-Net, FCN-8s, SegNet, BCDU-Net, CU-Net, ORED-Net, MultiResUNet and PMED-Net, respectively.

B. NUCLEI SEGMENTATION

The quantitative results for the nuclei segmentation task are listed in Table 4. The performance of the proposed architecture was comparatively better than that of SegNet, FCN-8s, CU-Net, ORED-Net, MultiResUNet and U-Net in terms of the IoU and F1-score. PMED-Net performs on par with BCDU-Net.

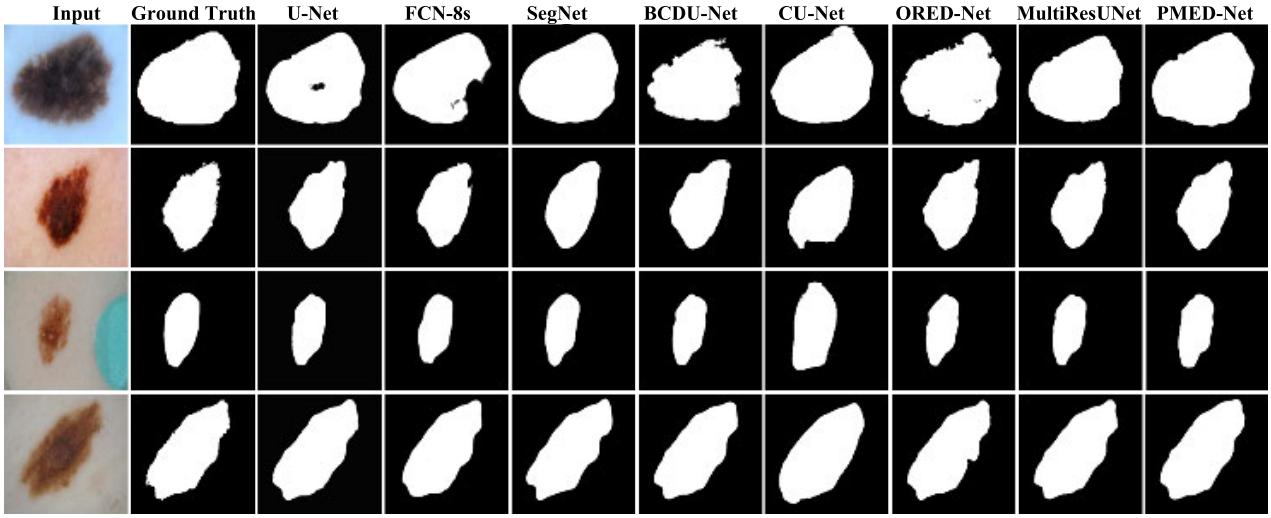

FIGURE 7. Experimental outputs of the ISIC dataset for different networks.

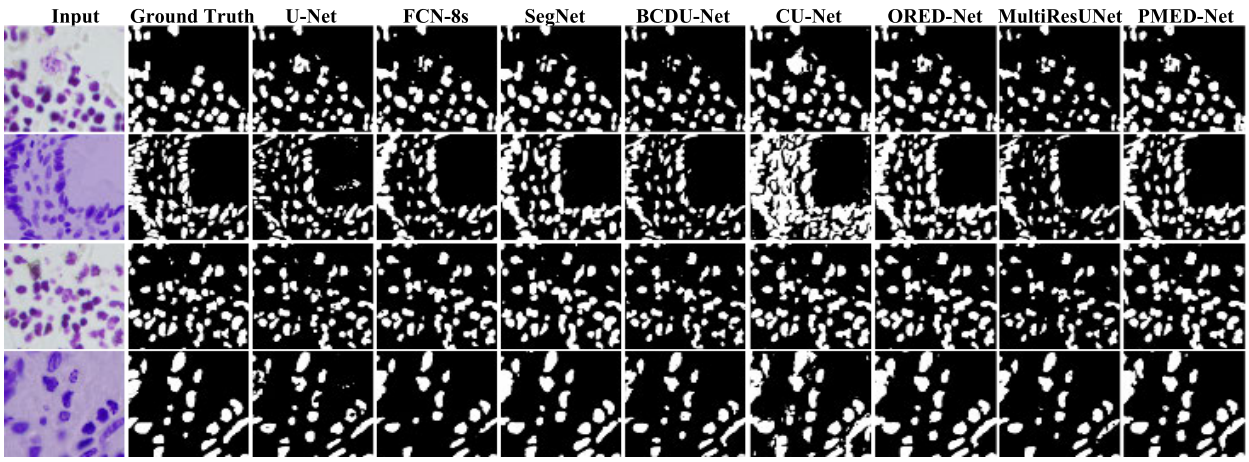

FIGURE 8. Experimental results of different networks for the Nuclei dataset.

TABLE 4. Evolution of PMED-Net architecture compared with other networks for nuclei segmentation.

Network	IoU	F1-Score	Sensitivity
U-Net [26]	0.7859	0.8801	0.9307
SegNet [25]	0.7509	0.8577	0.8922
FCN-8s [29]	0.7727	0.8718	0.8452
BCDU-Net [27]	**0.7952**	**0.8859**	**0.9374**
CU-Net [28]	0.7281	0.8427	0.9089
ORED-Net [30]	0.7896	0.8824	0.9308
MultiResUNet [31]	0.7397	0.8503	0.7811
PMED-Net (proposed)	0.7931	0.8846	0.9242

BCDU-Net performs marginally (0.21%, 0.13%, and 1.3% in terms of IoU, F1-Score, and sensitivity, respectively) better than PMED-Net utilizing 14 times more parameters. The PMED-Net architecture was extended to six pyramid levels for this dataset, and the performance improvement contributed by each level is shown in Figure 5. As can be seen, each extra stage in the pyramid level network further refined the previous predictions. The PMED-Net compromised only around 1.3 million parameters as compared to 20.66 million parameters of BCDU-Net.

The visual results of PMED-Net and the other comparing models are shown in Figure 8. PMED-Net gave satisfactory performance in segmenting the small nuclei and clearly distinguishing the boundaries of each nucleus when compared to U-Net, FCN-8s, SegNet, ORED-Net, MultiResUNet and CU-Net which were unable to distinctively differentiate the region of interest.

C. BRAIN TUMOR SEGMENTATION

Table 5 illustrates the quantitative performance of the proposed architecture for the brain tumor dataset as compared to the other networks. For this dataset, PMED-Net was extended to four pyramid levels, and after fourth level there was insignificant improvement in the performance, as shown in Figure 5. PMED-Net outperforms SegNet, FCN-8s, CU-Net, ORED-Net, and MultiResUNet in terms of IoU and F1-score whereas slightly underperforms compared to U-Net, BCDU-Net, CU-Net and ORED-Net in terms of sensitivity.

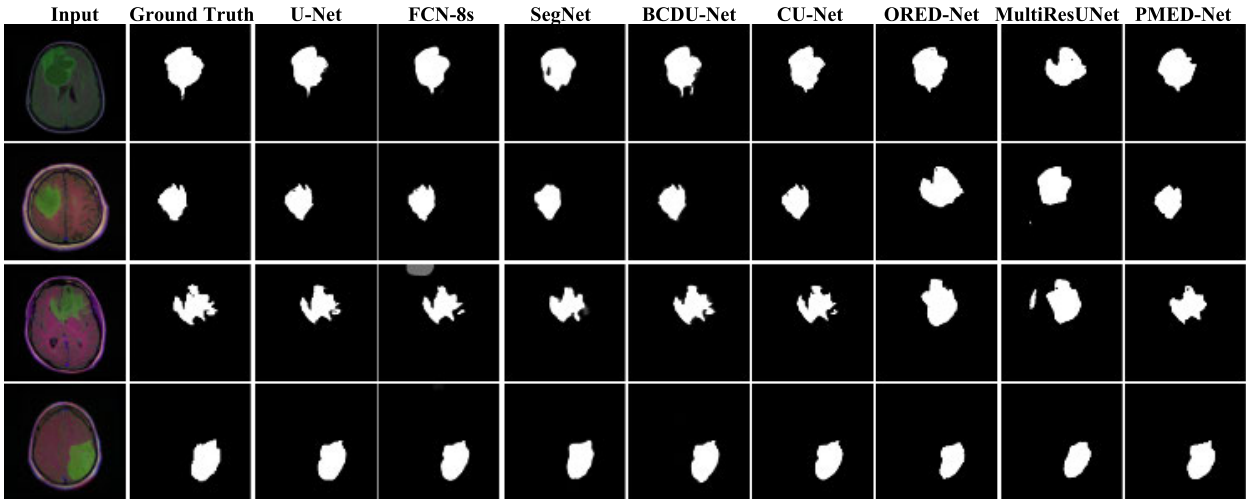

FIGURE 9. Qualitative comparison of results of segmentation for brain tumor dataset.

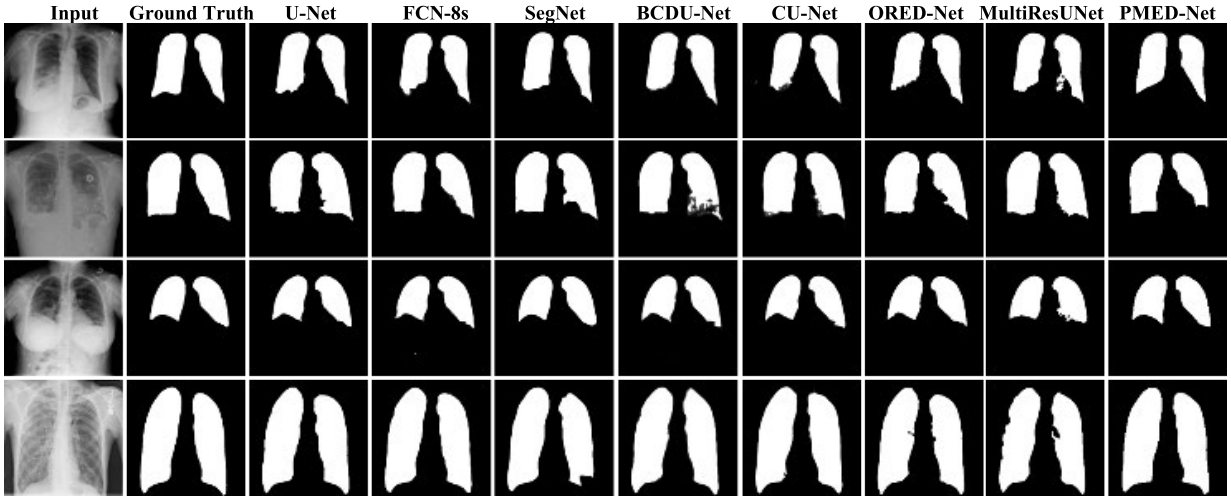

FIGURE 10. Visual results for NIH X-ray dataset segmentation.

TABLE 5. Comparison of evaluation metrics for brain tumor dataset segmentation.

Network	IoU	F1-Score	Sensitivity
U-Net [26]	**0.8564**	**0.9226**	0.9398
SegNet [25]	0.8114	0.8959	0.9125
FCN-8s [29]	0.81032	0.8952	0.9036
BCDU-Net [27]	0.8562	0.9225	**0.9497**
CU-Net [28]	0.8039	0.8913	0.9420
ORED-Net [30]	0.7786	0.8755	0.9372
MultiResUNet [31]	0.7141	0.8332	0.8280
PMED-Net (proposed)	0.8339	0.9093	0.9253

The visual results for the brain tumor dataset are shown in Figure 9. The PMED-Net architecture for this dataset had fewer than one million parameters (977,268) and was capable of producing on par or better results as compared to the other comparative methods.

D. X-RAY SEGMENTATION

The quantitative and qualitative analysis, for X-ray segmentation task, was performed on the NIH database. Table 6 summarizes the segmentation performance of PMED-Net architecture against each evaluation metric and all other networks. For each evaluation index, the proposed network performance is significantly better than all other networks in terms of the IoU, F1-Score, and sensitivity metric.

The qualitative results of the X-ray segmentation are shown in Figure 10. PMED-Net performance is good in terms of segmenting small regions and boundaries, which is evident from row 2 of Figure 10. Such optimal performance is obtained with four level PMED-Net architecture where the six level PMED-Net enhance the performance by only 0.83%, but cost 1.5 times more parameters.

Moreover, in this study, we also included few bad segmentation examples of the PMED-Net, shown in Figure 11, where

TABLE 6. Experimental results of PMED-Net for X-ray segmentation and comparison against other networks.

Network	IoU	F1-Score	Sensitivity
U-Net [26]	0.7993	0.8884	0.83052
SegNet [25]	0.7942	0.8853	0.8149
FCN-8s [29]	0.8159	0.8986	0.8374
BCDU-Net [27]	0.8062	0.8927	0.8419
CU-Net [28]	0.7848	0.8794	0.8081
ORED-Net [30]	0.7772	0.8746	0.8093
MultiResUNet [31]	0.8065	0.8929	0.8258
PMED-Net (proposed)	**0.8414**	**0.9139**	**0.9057**

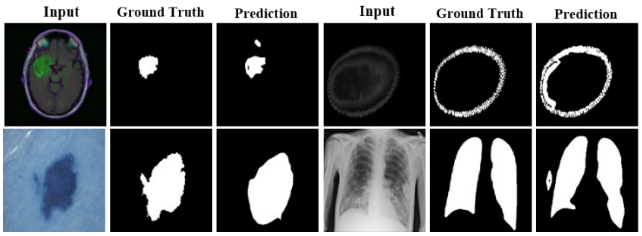

FIGURE 11. Examples of bad segmentation by PMED-Net.

the proposed network performance is reduced as it either over-segments or under-segments the region of interest.

VIII. CONCLUSION

In summary, we have presented a pyramid based multi-scale encoder-decoder, PMED-Net, for medical image segmentation. The proposed PMED-Net has quite less number of parameters and training as well as inference time, making it more efficient and applicable for embedded applications in healthcare. The PMED-Net architecture uses a coarse-to-fine prediction approach at each pyramid level to extract features with different scales using small encoder-decoder networks. We have extended the architecture up to six pyramid levels (where the optimal number of levels determined empirically). At each level, a light-weighted encoder-decoder network is trained independently, and then its predictions are up-sampled, concatenated with the next pyramid level images, and used as input for the next level encoder-decoder network. We have evaluated and compared PMED-Net on four different publicly available medical image datasets. The results show that the proposed PMED-Net significantly improves the computer aided diagnosis of medical images compared to the other state-of-the-art networks with much lower parameterized complexity.

REFERENCES

[1] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019.
[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[3] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[4] T. Hassan, M. U. Akram, B. Hassan, A. M. Syed, and S. A. Bazaz, "Automated segmentation of subretinal layers for the detection of macular edema," *Appl. Opt.*, vol. 55, no. 3, pp. 454–461, Jan. 2016.
[5] B. Hassan, G. Raja, T. Hassan, and M. U. Akram, "Structure tensor based automated detection of macular edema and central serous retinopathy using optical coherence tomography images," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 33, no. 4, pp. 455–463, 2016.
[6] M. U. Rehman, S. H. Khan, S. M. D. Rizvi, Z. Abbas, and A. Zafar, "Classification of skin lesion by interference of segmentation and convolotion neural network," in *Proc. 2nd Int. Conf. Eng. Innov. (ICEI)*, Jul. 2018, pp. 81–85.
[7] T. Fechter and D. Baltas, "One-shot learning for deformable medical image registration and periodic motion tracking," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2506–2517, Jul. 2020.
[8] Z. Li, M. Dong, S. Wen, X. Hu, P. Zhou, and Z. Zeng, "CLU-CNNs: Object detection for medical images," *Neurocomputing*, vol. 350, pp. 53–59, Jul. 2019.
[9] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, p. 311.
[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[11] Q. Rao and J. Frtunikj, "Deep learning for self-driving cars: Chances and challenges," in *Proc. 1st Int. Workshop Softw. Eng. for AI Auton. Syst.*, May 2018, pp. 35–38.
[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. BMVC*, vol. 1, no. 3, Sep. 2015, p. 112.
[13] S. D. Ali, W. Alam, H. Tayara, and K. Chong, "Identification of functional piRNAs using a convolutional neural network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Oct. 29, 2020, doi: 10.1109/TCBB.2020.3034313.
[14] A. Wahab, S. D. Ali, H. Tayara, and K. To Chong, "IIM-CNN: Intelligent identifier of 6 mA sites on different species by using convolution neural network," *IEEE Access*, vol. 7, pp. 178577–178583, 2019.
[15] W. Luo, A. G. Schwing, and R. Urtasun, "Efficient deep learning for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5695–5703.
[16] C. Hane, C. Zach, A. Cohen, R. Angst, and M. Pollefeys, "Joint 3D scene reconstruction and class segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 97–104.
[17] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, 2017.
[18] L. Mansi, R. Grassi, V. Cuccurullo, and A. Rotondo, "Diagnostic imaging techniques: Lessons learned," in *Advanced Imaging Techniques in Clinical Pathology*. New York, NY, USA: Humana Press, 2016, pp. 159–160.
[19] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, 1993.
[20] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng., Annu. Rev.*, vol. 2, no. 1, pp. 315–337, 2000.
[21] D. Ciregan, 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.
[22] G. Marcus, "Deep learning: A critical appraisal," 2018, *arXiv:1801.00631*. [Online]. Available: https://arxiv.org/abs/1801.00631
[23] S. G. Langer, "Challenges for data storage in medical imaging research," *J. Digit. Imag.*, vol. 24, no. 2, pp. 203–207, 2011.
[24] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, 2019.
[25] 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.
[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[27] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 406–415.
[28] H. Liu, X. Shen, F. Shang, F. Ge, and F. Wang, "CU-Net: Cascaded U-Net with loss weighted sampling for brain tumor segmentation," in *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*. Cham, Switzerland: Springer, 2019, Art. no. 102111.

[29] 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.
[30] 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.
[31] R. A. Naqvi, D. Hussain, and W.-K. Loh, "Artificial intelligence-based semantic segmentation of ocular regions for biometrics and healthcare applications," *Comput., Mater. Continua*, vol. 66, no. 1, pp. 715–732, 2021.
[32] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, Jul. 2001, pp. 105–112.
[33] N. Ramesh, J.-H. Yoo, and I. Sethi, "Thresholding based on histogram approximation," *IEE Proc. Vis., Image Signal Process.*, vol. 142, no. 5, pp. 271–279, 1995.
[34] N. Sharma and A. K. Ray, "Computer aided segmentation of medical images based on hybridized approach of edge and region based techniques," in *Proc. Int. Conf. Math. Biol.*, 2015, pp. 150–155.
[35] B. C. Patel and G. R. Sinha, "An adaptive K-means clustering algorithm for breast image segmentation," *Int. J. Comput. Appl.*, vol. 10, no. 4, pp. 35–38, Nov. 2010.
[36] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193–199, Mar. 2002.
[37] M. Mancas, B. Gosselin, and B. Macq, "Segmentation using a region-growing thresholding," *Proc. SPIE*, vol. 5672, pp. 388–398, Mar. 2005.
[38] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*. [Online]. Available: https://arxiv.org/abs/1606.08921
[39] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," 2020, *arXiv:2001.05566*. [Online]. Available: https://arxiv.org/abs/2001.05566
[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[42] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, Jan. 2011.
[43] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[44] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: https://arxiv.org/abs/1803.08375
[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[46] J. Bertels, T. Eelbode, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimizing the Dice score and Jaccard index for medical image segmentation: Theory and practice," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 92–100.
[47] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.
[48] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, Aug. 2018, Art. no. 180161.
[49] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *Amer. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, Jan. 2000.
[50] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y.-X. Wang, P.-X. Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 233–245, Feb. 2014.
[51] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 577–590, Nov. 2013.
[52] S. Stirenko, Y. Kochura, O. Alienin, O. Rokovyi, Y. Gordienko, P. Gang, and W. Zeng, "Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation," in *Proc. IEEE 38th Int. Conf. Electron. Nanotechnol. (ELNANO)*, Apr. 2018, pp. 422–428.
[53] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.
[54] Y. Tang, Y. Tang, J. Xiao, and R. M. Summers, "XLSor: A robust and accurate lung segmentor on chest X-rays using criss-cross attention and customized radiorealistic abnormalities generation," 2019, *arXiv:1904.09229*. [Online]. Available: https://arxiv.org/abs/1904.09229

ABBAS KHAN received the B.S. degree in electrical engineering from Bahria University, Islamabad, Pakistan, in 2018. He is currently pursuing the M.S. degree in electronics and information engineering with Jeonbuk National University, Jeonju, South Korea. His research interests include medical image processing, computer vision, precision agriculture, optical flow, and depth estimation.

HYONGSUK KIM (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Missouri, Columbia, MO, USA, in 1992. From 2000 to 2002 and 2009 to 2010, he was a Visiting Scholar with the Nonlinear Electronics Laboratory, Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA, USA. Since 1993, he has been a Professor with the Division of Electronics Engineering, Jeonbuk National University, Jeonju, South Korea. His current research interests include memristors and its application to the implementation of neural networks.

LEON CHUA (Life Fellow, IEEE) is widely known for his invention of the memristor and the Chua's circuit. His research has been recognized internationally through numerous major awards, including 17 honorary doctorates from major Universities in Europe and Japan, and seven U.S. patents. He was elected as a Foreign Member of the European Academy of Sciences (Academia Europea), in 1997, the Hungarian Academy of Sciences, in 2007, and an Honorary Fellow of the Institute of Advanced Study at the Technical University of Munich, Germany, in 2012. He was a recipient of many major prizes, including the Frederick Emmons Award, in 1974, the IEEE Neural Networks Pioneer Award, in 2000, the First IEEE Gustav Kirchhoff Award, in 2005, the International Francqui Chair, Belgium, in 2006, the Guggenheim Fellow Award, in 2010, the Leverhulme Professor Award, U.K., from 2010 to 2011, and the EU Marie Curie Fellow Award, in 2013. In 2002, he was also a recipient of the Top 15 Most Cited Authors Award from all fields of engineering, from the Current Contents (ISI) database, published during the ten-year period, from 1991 to 2001.