

Received November 11, 2021, accepted November 24, 2021, date of publication November 29, 2021, date of current version December 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131216

# OP-convNet: A Patch Classification-Based Framework for CT Vertebrae Segmentation

SYED FURQAN QADRI<sup>1,2</sup>, LINLIN SHEN<sup>1,2</sup>, (Member, IEEE), MUBASHIR AHMAD<sup>3</sup>,  
SALMAN QADRI<sup>4</sup>, SYEDA SHAMAILA ZAREEN<sup>5</sup>, AND SALABAT KHAN<sup>1</sup>

<sup>1</sup>College of Computer Science and Software Engineering, Computer Vision Institute, Shenzhen University, Shenzhen, Guangdong Province 518060, China

<sup>2</sup>AI Research Center for Medical Image Analysis and Diagnosis, Shenzhen University, Shenzhen, Guangdong Province 518060, China

<sup>3</sup>Department of Computer Science and IT, The University of Lahore (Sargodha Campus), Sargodha 40100, Pakistan

<sup>4</sup>Computer Science Department, MNS-University of Agriculture, Multan 60650, Pakistan

<sup>5</sup>Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Corresponding author: Linlin Shen (llshen@szu.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 91959108, and in part by the Science Foundation of Shenzhen under Grant JSGG20180508152022006.

**ABSTRACT** Accurate vertebrae segmentation from medical images plays an important role in clinical tasks of surgical planning, diagnosis, kyphosis, scoliosis, degenerative disc disease, spondylolisthesis, and post-operative assessment. Although the structures of bone have high contrast in medical images, vertebrae segmentation is a challenging task due to its complex structure, abnormal spine curves, and unclear boundaries. In recent years, deep learning has been widely applied in the segmentation of vertebrae images. In this paper, towards a robust and automatic segmentation system, we present an overlapping patch-based convNet (OP-convNet) model for automatic vertebrae CT images segmentation. Due to the greater memory and processing costs associated with 3D convolutional neural networks, as well as the risk of over-fitting, we employ overlapping patches in segmentation tasks using 2D convNet. In the proposed vertebrae segmentation method, OP-convNet effectively keeps the local information contained in CT images. We divide CT image slices into equal-sized square overlapping patches and applied the RUS-function on these patches for class balancing to minimize computational requirements. Then, these patches are input into the model along with their corresponding ground truth patches. This method has been evaluated on publicly available CT images from the MICCAI CSI workshop challenge. The results indicate that OP-convNet has precision (PRE) of 90.1%, specificity (SPE) of 99.4%, accuracy (ACC) of 98.8%, F-score of 90.1% in terms of the patch-based classification accuracy, and BF-score of 90.2%, sensitivity (SEN) of 90.3%, Jaccard index (JAC) of 82.3%, dice similarity score (DSC) of 89.9% in terms of the segmentation accuracy that outperform previous methods across all metrics.

**INDEX TERMS** Vertebrae segmentation, deep learning, convolutional neural networks, CT image, OP-convNet.

## I. INTRODUCTION

Image segmentation is a process that converts raw medical image data into meaningful, spatially structured information, which is necessary for scientific discovery [1]. Vertebrae segmentation is a prerequisite step for automatic spine analysis. Vertebral fracture detection, spine deformities assessment, and computer-assisted surgery of the spine [2] are facilitated

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

by accurate vertebrae segmentation. Automated analysis of spine faces positive diversity among different tomographic scans, including dedicated spine scanning and the abdomen, chest, and neck scans. Therefore, it needs a generic vertebrae segmentation robust to a range of image resolutions and spine coverage. This requires that the vertebrae are visible clearly with their anatomical structure and show which spine section they belong to.

The goal of vertebra segmentation in a computed tomography (CT) volume is to diagnose, quantitatively analyze,

and plan surgical/therapy for various spine diseases and disorders such as spine trauma, scoliosis, and other spinal pathologies. Like many other image-guided diagnoses in medical images, the vertebrae segmentation procedure can be accelerated and enhanced by computer-aided diagnosis (CAD) systems because manual segmenting of vertebrae in 3D medical images can be time-consuming and laborious, with the unpredictability of personal subjectivity. In the process of diagnostic radiology, computer-aided diagnosis (CAD) plays a significant role. Patients' images are processed faster because CAD systems shorten the time it takes for routine tasks to be completed, resulting in fewer mistakes caused by fatigue physicians. This is because CAD is relatively sensitive, as well as repeated and robust. Low computing time is required for healthcare purposes [3]. One of the biggest challenges of CAD application is the segmentation and detection of vertebrae from computed tomography (CT) images. CT images allow the evaluation of bone in details because it has high contrast and spatial resolution. Its availability is higher, and the overall low cost of the patient examination makes it the most widespread imaging technique.

It isn't easy to perform vertebrae segmentation because of anatomical variations. Such variations go to spinal variable coverage and structure that guides anatomy cues for vertebrae segmentation, especially the sacrum and ribs. Furthermore, neighboring vertebra often has similarity in appearance and shape too. Many methods focused on such challenges for automatic segmentation have been published in [4]–[7]. The model-fitting has remained a challenge in the vertebrae segmentation technique, which has often relied on statistical shape modeling, as well as shape constrained deformable and active shape modeling [8]–[10]. Some approaches proposed are based on atlas [11], active contours [12], and level sets with shape prior [13]. Due to the registration process, these atlas-based approaches have a high computational cost. While the accuracy of most of these deformable models is promising, they are sensitive to initialization, performed manually or automatically. Intervertebral discs bounding boxes were found using an interactive variant of marginal space learning used for segmentation of vertebrae based on graph cut and Markov random field in [14]. The work in [15] applied the Adaboost-based object detection framework to find out the vertebral bodies bounding boxes that were used to segment each vertebral body by inflating a mesh from the center. The work in [5] described the vertebral bodies center detection by using random forest regression and this framework is utilized to highlight the ROIs where vertebrae were segmented. A method in [16] regress the distance to the vertebral body's nearest center by using multilayer perceptron (MLP) and then vertebral bodies are segmented to initialize an adaptive shape modeling by using detected locations. The work in [17] used probabilistic boosting-tree classifier to detect the vertebral boundary by combining shape modeling and machine learning, that applied a surface mesh adaptation to a vertebra with combination of statistical shape

model (SSM). This SSM was used for mesh initialization to impose shape constraints.

Deep learning is a gateway in the development of medical image analysis in various clinical and research fields, e.g., brain tumors detection [18], supervised segmentation methods on volumetric medical images by 3D fully convolutional network [19], liver segmentation [20]–[22], prostate cancer segmentation from MRIs [23], and lung nodules segmentation using multi-scale 2D+3D features. Other methods based on CNN for CT lungs airways segmentation is used for leak detection [24], breast cancer classification [25], deep learning with level sets left were proposed in work [26] for heart ventricle segmentation and detection. Deep learning techniques are also applied in vertebrae segmentation and localization, e.g., vertebrae pathology grading [27], vertebrae localization for lumbar surgery in ultrasound images [28], and vertebrae segmentation [7], [29]–[32].

Deep learning is gaining popularity in CAD for vertebra segmentation, and several convolutional neural networks (CNNs)-based approaches have been proposed. Most recently published methods focus on vertebrae segmentation that replaced the explicit vertebral appearance and shapes modeling with convolutional neural networks (CNNs). A method in [33] used multiclass CNN for pixel labeling for lumbar vertebrae segmentation from 2D sagittal slices, and the lumbar region bounding box is estimated by a simple multilayer perceptron (MLP) for identifying the image's region of interest. In the subsequent work [34], a network for voxel classification was developed in the entire image from a 3D patch model, and a 2D frame was used to predict low resolution masks on the vertebral column that effectively handled the false-positive outside the vertebral region. A similar work was presented in [35] for spine segmentation. In that work, firstly, the lumbar region bounding box is estimated using regression CNN. Secondly, a classification CNN is used to do voxel labeling within the bounding box to achieve segmentation. A two-stage methodology [36] with interactive strategy is used to segment the vertebrae in downsampled images, and images were analyzed one after the other by a CNN for vertebrae segmentation. Fully-resolution images were analyzed in the second framework for refining the low resolution of vertebrae segmentation. Vertebrae localization and identification method in CT images are proposed in [37] by combining short and long-range contextual information in a supervised manner to develop a multi-task 3D FCN. A two-stage multi-class segmentation architecture based on a 3D graph convolutional segmentation network (GCSN) for 3D coarse segmentation and a 2D residual U-Net (ResUNet) for 2D segmentation refinement model is presented in [31], but several limitations exist in these works. Inferior segmentation is produced due to the blurry boundary, and it came at a high computational cost due to its complex network model.

While the aforementioned studies could segment the vertebrae CT image, they required high computational operations due to a large number of trainable parameters and were also low in segmentation accuracy. The consistent

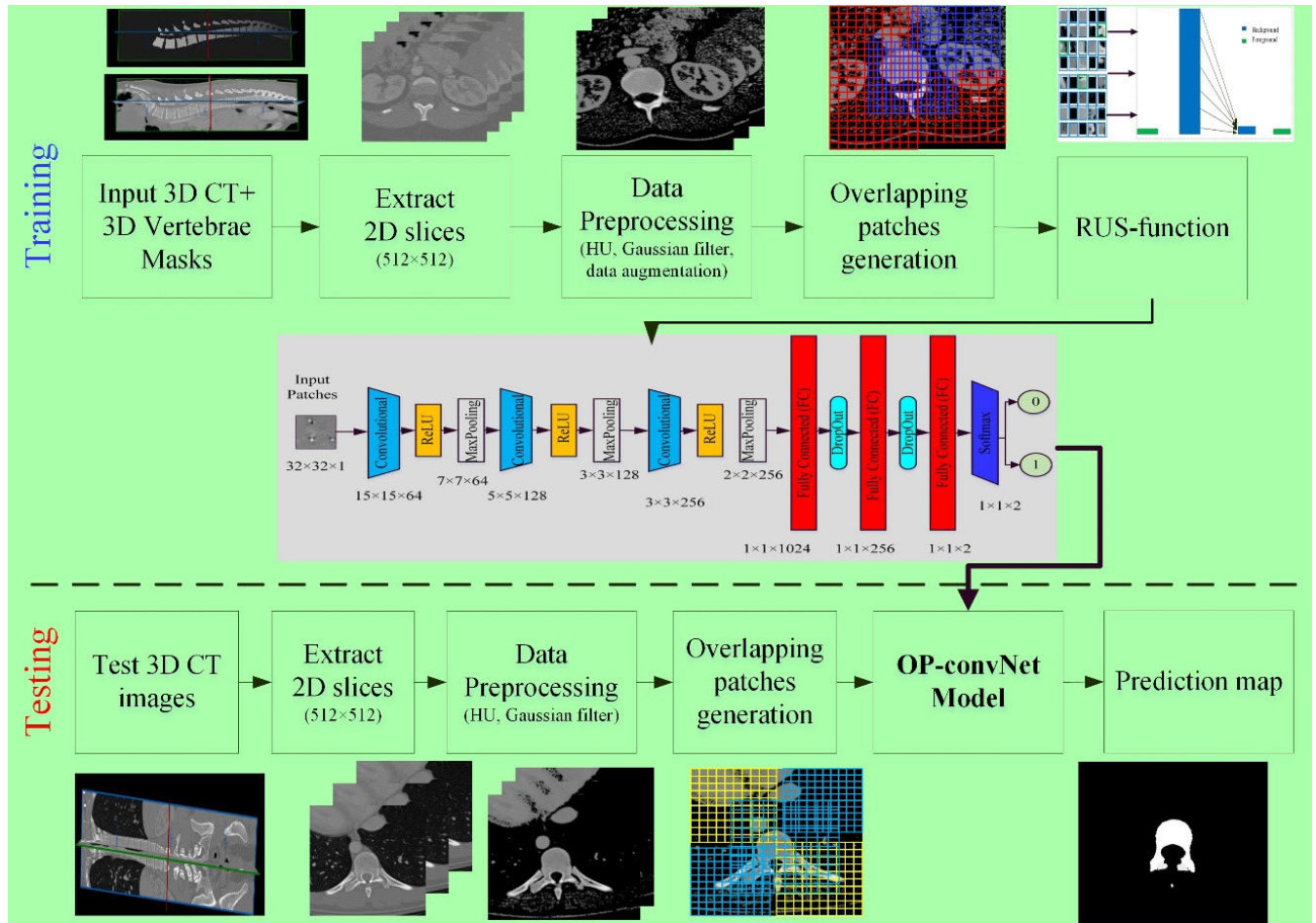


FIGURE 1. Flowchart of the proposed method for vertebrae segmentation.

problem of automatically, accurately, and rapidly segmenting the vertebrae persists. As a solution to these problems, we propose a vertebrae segmentation model from CT images that integrates patch-based deep learning with pre-and post-processing. As the spine CT images in the dataset are grayscale and the size of the vertebra is very small in image slices, a convolutional neural network [38], [39] is selected as an appropriate design to solve the two-class classification problem. The training and validation images are input as overlapping patches containing vertebrae and background. The proposed OP-convNet is a simple, patch-based architecture that requires low computational cost due to the small number of trainable parameters and gives high classification accuracy. Vertebrae image patches contain such gradients that can be exposed through convolutional neural networks.

The following are the main contributions of this paper.

- We propose an OP-convNet model to segment vertebrae CT images. The slices from CT images are divided into equal size square overlapping patches enhancing localization, so the trained network can emphasize local characteristics within each patch.
- The proposed patch-based convnet model avoids the difficulties associated with reduced feature map resolution and semantic feature loss.

- Class balancing of overlapping patches is achieved by the RUS-function.
- The results on publically available MICCAI CSI dataset indicate that OP-convNet achieved outstanding precision, specificity, accuracy, F-score in terms of the patch-based classification accuracy, and BF-score, sensitivity, Jaccard index, dice similarity score in terms of the segmentation accuracy without adding to the computational load.

## II. METHODOLOGY

Similar to other deep learning-based methods, our proposed method is also divided into training and testing phases. The proposed method further comprises the preprocessing (HU, Gaussian filter, data augmentation), overlapping patch generation, RUS-function for class balancing, OP-convNet framework and testing phase. Figure 1 shows the flowchart of the proposed methodology for vertebrae segmentation.

### A. DATA PREPROCESSING

The primary aim of the preprocessing phase is to identify bone pixels and improve differentiation between vertebrae and other tissues. We removed the noise artifacts from the entire CT images by applying a threshold technique in

this study. Then, outside the bone intensity range of 100HU (Hounsfield unit) and 1500HU, the intensity is set to 0 to minimize noise, imaging artifacts, and influence from the tissues surrounding the vertebral column. Our input spine CT data are volumetric data, so they need to be processed slice by slice. The applied threshold distinguishes the vertebrae from other soft tissues because they have higher pixel intensities in CT images than other tissues. But vertebrae have similar intensities to other bones like ribs, so we train our deep learning model to distinguish the vertebrae from the different bony structures in the CT images. We apply the Gaussian filter to control the smoothness of CT images with a fixed kernel size to enhance the spine’s segmentation accuracy and antique the noisy pixels’ effectness. Large deformation is correlated with a large value of Gaussian kernel, and the estimated transformation is used as a small kernel input of smoothness. Figure 2 shows  $512 \times 512$  pixels original CT slice images with respective ground truths from dataset.

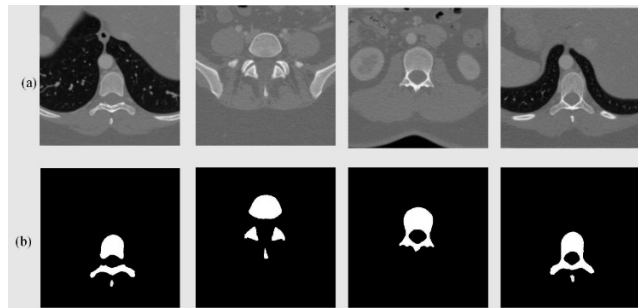


FIGURE 2. Examples of  $512 \times 512$  pixels images from dataset (a) Original CT slice image and (b) respective ground truth.

1) DATA AUGMENTATION

ConvNet’s accuracy increases with data augmentation, and overfitting is reduced [40]. Since the input images are rotationally invariant, we rotate each image ( $\theta$  degrees) to maximize the number of image samples. It should be noted that while rotating an image may degrade its high-frequency content slightly, it should have no effect on the background or foreground of the image. Following that, we randomly translate each image to increase the number of training samples for the network. The augmentation step based on image rotation is essential to the network’s performance [41] because of the limited number of images in the spine dataset. In the next step, the data are normalized to the range [0,1].

B. OVERLAPPING PATCH GENERATION

The input CT images are divided into overlapped patches with size of  $n \times n$ . We take certain stride pixels to create the overlapping patches as shown in figure 3.  $32 \times 32$  pixel size patch having a total of 1024 pixels, if the total pixels inside a patch is equal or more than 513, then the patch is labeled as 1 (vertebra or foreground patch) otherwise, it is labeled as 0 (non-vertebra or background). We used a certain pixel stride for the sliding window to create overlapping patches for training.

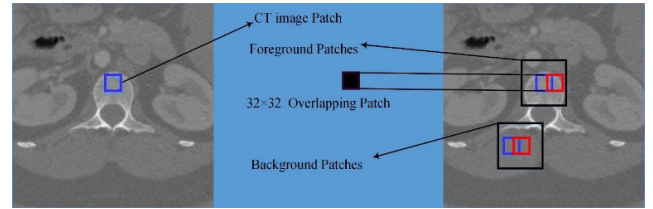


FIGURE 3. Overlapping patches of size  $32 \times 32$  pixels generation.

Testing images are also divided to generate  $n \times n$  pixel size overlapped patches to check the model’s accuracy and then segment the vertebrae from CT images. Figure 4 shows the randomly selected input patches with size of  $32 \times 32$  pixels.

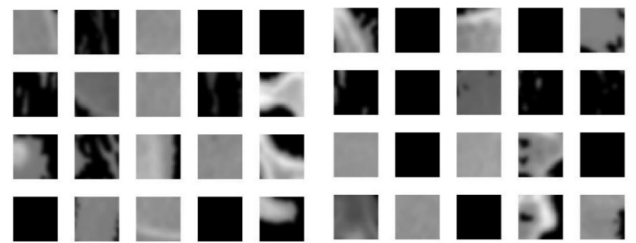


FIGURE 4. Randomly selected input patches with size of  $32 \times 32$  pixels.

C. RUS-FUNCTION

We used two classes for classification, and there was an imbalance size of numbers of training patches, as shown in the class distribution in figure 5. Because the vertebrae spatial area in the images is much smaller than the background area, most of training patches are labeled as 0, so the classifier can lead biasness in the background. From a medical perspective [42], a high recall rate (correct vertebra patch classification) is desirable, but a high false-negative rate (vertebra misclassified as background) is inappropriate on a practical level. The proportions of positive and negative training samples are to be balanced to address this dilemma [43]. We applied the random under-sampling function (RUS-function) for negative samples and generated a

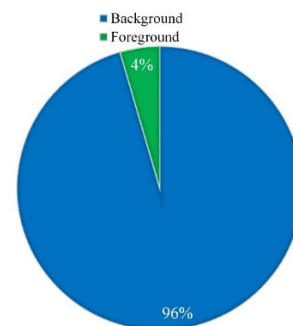
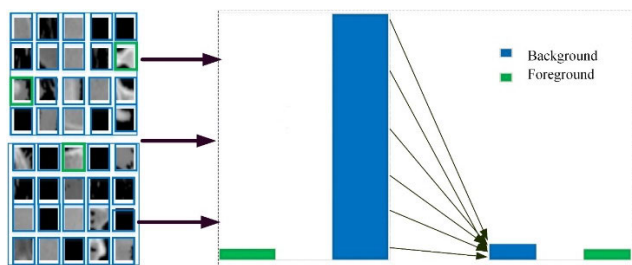


FIGURE 5. Class distribution in training overlapping patches, gray, background and green shows foreground (vertebrae) patches. This class distribution shows heavy imbalance samples.





**FIGURE 6.** RUS-function is applied to CT image patches for class balancing, to remove patches from the majority class (background).

balanced training set to deal with this issue. As a result, this increases the accuracy and convergence rate of the network during the training process [40], [41]. Figure 6 indicates the RUS-function is applied to CT image patches to remove majority class patches (background patches) for making balance classes before training stage.

#### D. OP-convNet FRAMEWORK

When combined with suitable selected local features, the overlapping patch-based technique appears to be an effective solution for the vertebrae segmentation problem. However, due to a lack of precise definitions for the vertebra to be classified, determining the appropriate features in the treated problem is challenging. Furthermore, choosing the most important and mutually independent features from a huge number of potentially available characteristics need a complicated statistical analysis based on large volumes of image data. This is why the OP-convNet is employed to implement the automatic feature extraction in this study. Figure 7 shows general proposed OP-convNet structure (overlapping patches as input, convolutions, non-linearities and pooling layers), followed by more convolutionary and fully connected layers. AlexNet [40], VGG net [44], GoogLeNet [45], ResNet [46], and other standard CNN architectures are designed to classify RGB images but all of these networks require an exact size of  $224 \times 224$  pixels of input, which is too large to meet our problem. Some strategies (zero-padding, re-sampling, etc.) can help overcome this drawback, but they require an excessive amount of processing effort. In addition, the above classification networks contain a large number of trainable parameters and require a large number of computational operations during a single forward pass. Another possibility was to use fully convolutional nets specifically designed for image segmentation, such as U-Net [47], SegNet [48], which can segment an entire two-dimensional image in a single forward pass. When using low-resolution feature maps for up-pooling, SegNet frequently loses neighboring information. Additionally, it focused more on central slices. Furthermore, when it comes to maintaining local features for image segmentation, U-net also has some drawbacks.

To address the aforementioned shortcomings, we propose a method for vertebrae CT segmentation that utilizes an overlapping patch-based convolutional neural network

architecture. It achieves a higher level of segmentation accuracy when applying overlapping patches, while maintaining the same degree of complexity as the standard CNN architecture. In this work, we proposed a specific convNet architecture for segmentation of vertebrae from square patch images. A convNet is a multi-stage deep learning framework (convolutions, non-linearities and pooling layers), followed by more convolutionary and fully connected layers. The raw pixel intensity image is given as input by convNet. Every class of data is represented by a unique neuron in the output layer (in our case, convNet is chosen as a suitable architecture to solve the two-class classification problem). The weights ( $W$ ) in the ConvNet are optimized using the backpropagation algorithm to minimize the classification error on the training.

#### 1) CONVOLUTIONAL LAYER

The convolutional layer (the first layer) takes square patches from the input image (using a stride value and padding, if necessary) or feature maps (for subsequent layers) and performs 2D convolution with a filter. A rectified linear unit (ReLU) is used to accelerate training by feeding the sum of the resulting convolutions into a non-linearity function [40]. The same filter is used for all feature maps in a given layer, but different filters are used for other feature maps. This filter sharing property of the conv. layer enables the detection of the same pattern in multiple locations across the feature map.

#### 2) POOLING LAYER

The feature map is down-sampled using the pooling layer, which frequently sums up the feature responses in each overlapping patch by computing the maximum activations (max-pooling). These results in features that are invariant to minor data translations.

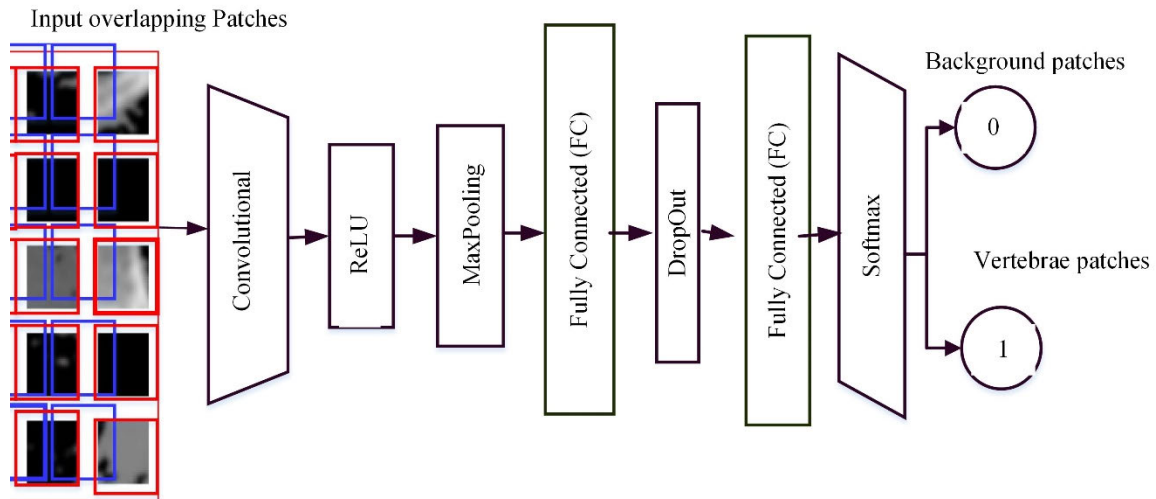
#### 3) FULLY CONNECTED LAYER

Convolution and pooling create small dimension feature maps than the input image, which are then processed through numerous fully connected (FC) layers. The initial few FC layers combine these feature maps to construct a feature vector. The final FC layer contains two neurons that use softmax regression to calculate the probability of classification for each class. To avoid overfitting, the fully connected layers are constrained using dropout [49].

#### E. TESTING

ConvNets' weights are initialized with Gaussian distribution. These weights are changed iteratively during training with the gradients of the loss function, which are computed using stochastic gradient descent (SGD) over a mini-batch of training data. After a given number of epochs, the learning rate slows down. To accelerate learning and reduce overfitting, momentum and weight decay are applied. After a set number of epochs, the training process is completed. The final model with the lowest validation loss value is selected.

The test spine CT image receives the same preprocessing. Test patches of  $32 \times 32$  pixels with a stride of 1 are fed to the



**FIGURE 7.** General proposed OP-convNet structure (overlapping patches as input, convolutions, non-linearities, and pooling layers), followed by more convolutionary and fully connected layers to solve the two-class classification problem.

trained model from unseen spine images. Each test patch will get a label through a trained model. In our case, OP-convNet predicts the two-class patch-classification problem (vertebra or background). A higher rate of false-positive (FP) was detected due to pixel-based segmentation. Because of most segmentation problems, background pixels are more than foreground pixels, so a systematic false negative (FN) error is more favorable than a systematic false positive (FP) error. For this reason, post-processing analysis was implemented. Then, the obtained binary predicted image is processed by applying simple morphological operations [50] to achieve fine segmentation.

### III. EVALUATION

#### A. DATASET

Our model's performance is evaluated using a publicly available MICCAI CSI dataset [51] gathered at a trauma hospital during routine medical examinations. The CT scan covers all thoracic and lumbar vertebrae which an in-plane resolution of 0.31mm to 0.45mm. These datasets were compiled from patients ranging in age from 16 to 35 years. Each slice is approximately  $512 \times 512$  pixels in size. There are between 520 and 600 slices in each case. This includes fifteen cases in total, ten cases for training, and five cases for testing. The dataset was received from the University of California, Irvine's medical center (Orange, CA, USA). Scan and reconstruction settings include a voltage of 120 kVp, intravenous contrast, and a 0.7 to 1.0 mm slice thickness. The scans were performed at a high spatial resolution as a continuous CT data set. Additionally, the challenge organizers provided reference segmentations that were generated semi-automatically and manually corrected.

#### B. EVALUATION MEASURES

Evaluation measures [52] used to evaluate the patch-based classification and vertebrae segmentation performance with

other existing methods. The evaluation measures are considered well-known and widely used in medical image analysis.

#### 1) CLASSIFICATION METRICS

The performance of patch-based classification is evaluated by precision (PRE), specificity (SPE), accuracy (ACC), and F-score for quantitative assessment. We draw the confusion matrix to perform the quantitative evaluation. Recall measures the ratio of positive patches in the ground truth that is also detected as positive by the proposed method. Specificity is used to measure the negative patches evaluated by the method that is negative in ground truth too. It identifies how sensitive the method is to detect the correct background patches. It is obtained by an equation as:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (1)$$

Precision is achieved as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

The overall accuracy is obtained as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

F-score is calculated as follows:

$$\text{F-score} = \frac{2 \times TP}{(2 \times TP) + FP + FN} \quad (4)$$

#### 2) SEGMENTATION METRICS

Vertebrae segmentation performance is evaluated by boundary F1 score (BF-score), sensitivity (SEN), Jaccard index (JAC) [53], and Dice similarity coefficient (DSC) [54] for quantitative assessment. BF-score determines how closely an object's predicted boundary corresponds to its ground truth boundary.

$$\text{BF-score} = \frac{2 \times (\text{PRE} \times \text{SEN})}{(\text{PRE} + \text{SEN})} \quad (5)$$

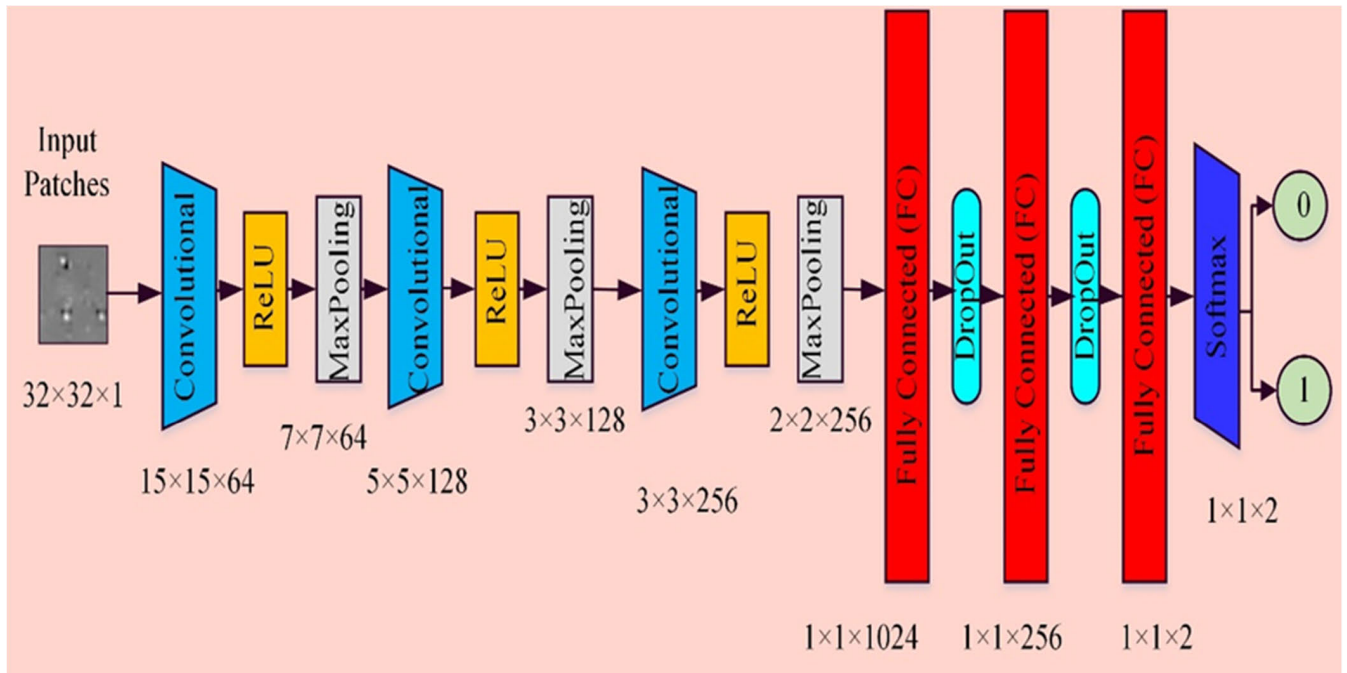


FIGURE 8. Architecture of OP-convNet.

Sensitivity (SEN) measures the proportion of correctly classified positive patterns.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

The Jaccard index JAC is used to measure the intersection spatial overlap divided by the union size of two label sets as:

$$\text{JAC} = \frac{|A \cap B|}{|A \cup B|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (7)$$

Dice similarity coefficient (DSC) measures the spatial overlapping between two binary images, its values range from zero (non-overlap) and one (perfect overlapped). The segmentation result and the ground truth, the DSC values are achieved by an equation as:

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

#### IV. EXPERIMENTS AND RESULTS

The proposed OP-convNet was trained by the techniques of mini-batch descent gradient and the momentum back-propagation. All biases were initialized to zero. The optimal convNet hyper-parameters (initial learning rate, momentum, weight decay, and mini-batch size) were found utilizing a pyramidal approach by doing a random search within the constrained space parametry [55]. Stochastic gradient descent (SGD) is used to train the network for 30 epochs with a mini-batch size of 128. The initial learning rate is 0.01 and, learning rate drop factor is 0.1 on each 10th epochs. The momentum and rate of weight decay L2 regularization are set to 0.9 and 0.0001, respectively. The fc4 and fc5 layers are constrained

using a dropout ratio of 0.5. The experiments are conducted using MATLAB 2018a on a 1.80 GHz i7 CPU, 32GB RAM, and an NVIDIA GeForce MX250 GPU.

##### A. PATCH-BASED CLASSIFICATION

Figure 8 illustrates the architecture of OP-convNet. Three convolutional-pooling layers were used in the convNet. For the three layers, the number of feature maps is 64, 128, and 256 based on study [38]. The filter sizes for all three layers are  $3 \times 3$ , whereas the max-pooling sizes are  $3 \times 3$ ,  $3 \times 3$ , and  $2 \times 2$  respectively. After preprocessing, input patches of  $32 \times 32$  pixels are fed to the model. The first 2D convolutional layer outputs a convolutional feature map with 64 neurons and has a kernel size of  $3 \times 3$ .

Zero paddings are applied to boundary pixels during convolution so that each layer's feature map is the same size as the input patch. This layer is followed by a rectified linear unit (ReLU) activation layer, which handles element-wise non-linearity. A  $3 \times 3$  maxpool layer subsamples the input patch image and output size is  $7 \times 7 \times 64$  at this stage. Subsequently, convolutional layers two and three have the same size of kernel ( $3 \times 3$ ), and ReLU layers extract 128 and 256 features maps, respectively. These layers are followed by  $3 \times 3$  and  $2 \times 2$  maxpooling, resulting in  $3 \times 3 \times 128$  and  $2 \times 2 \times 256$  size output. In OP-convNet, the three FC layers have 1024-256-2 neurons, respectively. The values of 1024 and 256 were chosen based on our empirical findings, and the number 2 was chosen to accommodate the number of object categories in our two-class (vertebrae/background) classification problem. The proposed architecture is most

TABLE 1. Parameters of the proposed OP-convNet model.

#	Name	Type	Activations	Learnables	Total learnables
1	Imageinput	Image Input	32×32×1	-	0
2	Conv_1	Convolution	15×15×64	Weights 3×3×1×64 Bias 1×1×64	640
3	Relu_1	ReLU	15×15×64	-	0
4	Maxpool_1	Max Pooling	7×7×64	-	0
5	Conv_2	Convolution	5×5×128	Weights 3×3×64×128 Bias 1×1×128	73856
6	Relu_2	ReLU	5×5×128	-	0
7	Maxpool_2	Max Pooling	3×3×128	-	0
8	Conv_3	Convolution	3×3×256	Weights 3×3×128×256 Bias 1×1×256	295168
9	Relu_3	ReLU	3×3×256	-	0
10	Maxpool_3	Max Pooling	2×2×256	-	0
11	fc_1	Fully Connected	1×1×1024	Weights 256×1024 Bias 1024×1	1049600
12	dropout_1	Dropout	1×1×1024	-	0
13	fc_2	Fully Connected	1×1×256	Weights 256×1024 Bias 256×1	262400
14	dropout_2	Dropout	1×1×256	-	0
15	fc_3	Fully Connected	1×1×2	Weights 2×256 Bias 2×1	514
16	softmax	Softmax	1×1×2	-	0
17	classoutput	Classification Output	-	-	0

suitable for optimizing performance of classification-based vertebrae segmentation.

The flattened output was sent into a fully connected layer (FC) with 1024 nodes and a dropout layer. With a probability of 0.5, the dropout layer removes different sets of nodes every time. Drop out layer maintained the regularization and, during training, avoids over-fitting of the convNet. In the first (FC) layer, a vector of 1024 features is created, while in the second (FC) layer, a vector of 256 features is produced, and dropout layers being applied. Finally, there is a softmax layer with two nodes that computes a probabilistic prediction for two classes. The softmax layer as output layer calculates L2 regularization cross-entropy loss. Table 1 listed the detailed configurations of OP-convNet model parameters. In total, 895019 patches were generated to train the model, in which 716019 patches (80%) were used for training and 179000 (20%) were used for validation.

Figure 9 illustrates the learned filters for convNet’s first, second, and third convolutional layers. These automatically trained filters compute gradients primarily with varying frequencies and orientations and color blobs necessary for classification. In addition to these learned filters, figure 10 shows the activations feature maps of pooling layers (maxpool1, maxpool2, and maxpool3). In the proposed method,

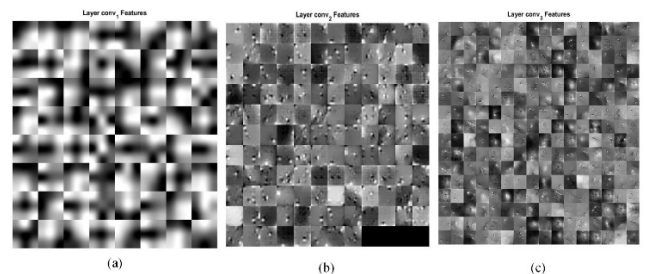
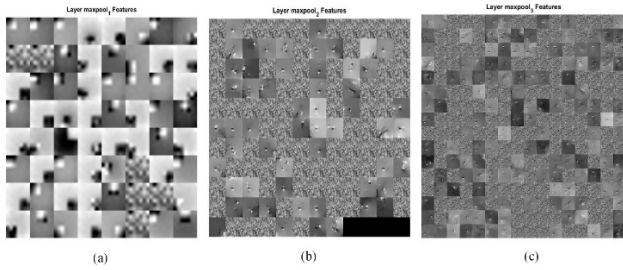


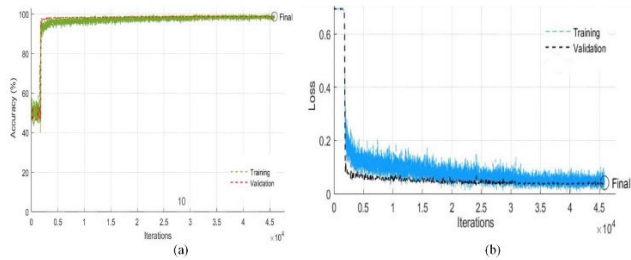
FIGURE 9. Filters size of 64, 128, and 256 visualizations with size of 3 × 3 × 1, 3 × 3 × 64 and 3 × 3 × 128 in the first, second, and third convolutional layers.

overlapping patches of the same size (a stride of 12 pixels) are used. The reason for the adoption of a 12-pixel stride is that when its size decreases, the number of patches grows, resulting in increased computational complexity. As a result, we determined a stride size of 12 pixels as the ideal value for our experimental study. Figure 11 depicts an OP-convNet fine-tuning process over 120 training epochs. As illustrated in the figure, the validation loss reaches its smallest value (0.106) after eight epochs, corresponding to a validation accuracy of 0.987. The training of OP-convNet requires 27 hours under our experimental setup.





**FIGURE 10.** Activations' visualization of three pooling layers: Pool1 (64@7 × 7), Pool2 (128@3 × 3), and Pool3 (256@2 × 2).

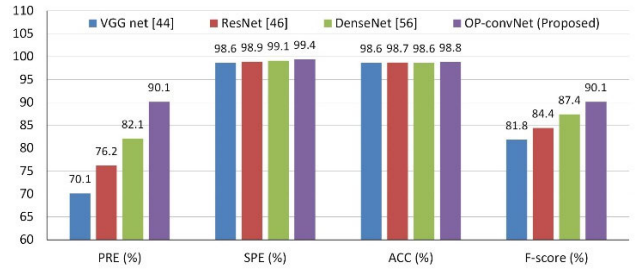


**FIGURE 11.** The validation graph shows (a) the average accuracy and (b) the average loss of each epoch (iterations).

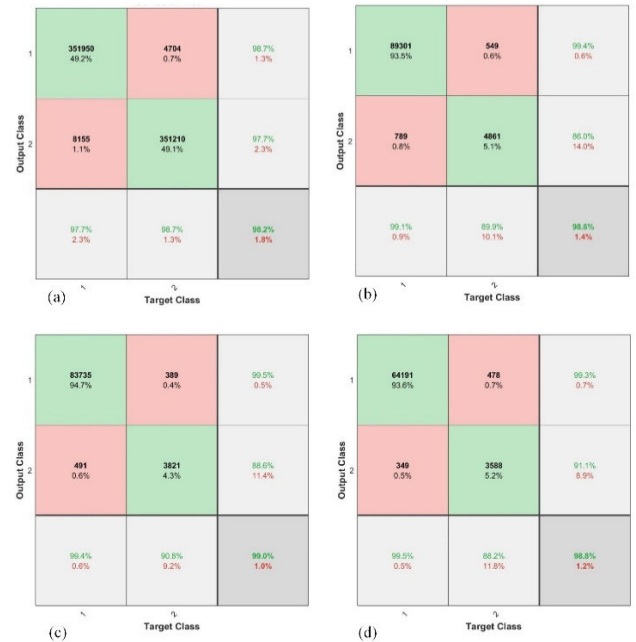
**TABLE 2.** Patch-based classification performance comparison against existing algorithms.

Models	PRE (%)	SPE (%)	ACC (%)	F-score (%)
VGG net [44]	70.1	98.6	98.6	81.8
ResNet [46]	76.2	98.9	98.7	84.4
DenseNet [56]	82.1	99.1	98.6	87.4
OP-convNet (Proposed)	<b>90.1</b>	<b>99.4</b>	<b>98.8</b>	<b>90.1</b>

The results in table 2 quantitatively verify the patch-based classification advantages of OP-convNet framework. Classical networks (like VGG net, ResNet, DenseNet) have been compared in order to show the benefits of our proposed method. From table 2, we can see that among the existing methods, our method has the best PRE (90.1%), SPE (99.4%), ACC (98.8%), and F-score (90.1%). OP-convNet outperforms on classical backbones in terms of PRE, SPE, ACC, and F-score. Compared with the VGG net [42], ResNet [44], DenseNet [52], our proposed method is significantly better by (20%, 0.8%, 0.2%, 8.3%), (13.9%, 0.5%, 0.1%, 5.7%), and (8%, 0.3%, 0.2%, 2.7%) average in (PRE, SPE, ACC, F-score) respectively. Therefore, it can be concluded that our proposed OP-convNet framework is superior in terms of its patch-based classification performance. Figure 12 shows patch-based classification performance comparison.



**FIGURE 12.** Patch-based classification performance comparison, chart shows OP-convNet achieved highest PRE, SPE, ACC, and F-score.



**FIGURE 13.** Confusion matrices of proposed OP-convNet model; (a) confusion matrix of balanced training data and (b–d) confusion matrices of randomly test cases 1–3.

### B. SEGMENTATION PERFORMANCE

After presenting the patch-based classification performance using training images, we now used five test cases containing 3418 slices to access the segmentation performance of our approach. Figure 13 shows the confusion matrices of training, and test cases separately. Figure 14 shows the segmentation results for the axial plane, where the original images and ground truth segmentation maps for each slice, are illustrated in first and second rows respectively. The third row shows the proposed approach's predicted segmentation map, and it can be seen that the proposed method produces well-segmented results. In comparison to related studies, table 3 shows the segmentation results outperforms previous methods in terms of BF-score, sensitivity (SEN), Jaccard index (JAC) and Dice similarity coefficient (DSC). OP-convNet achieved the best vertebrae segmentation performance, with BF-score of 90.2%, SEN of 90.3%, JAC of 82.3%, and DSC of 89.9%

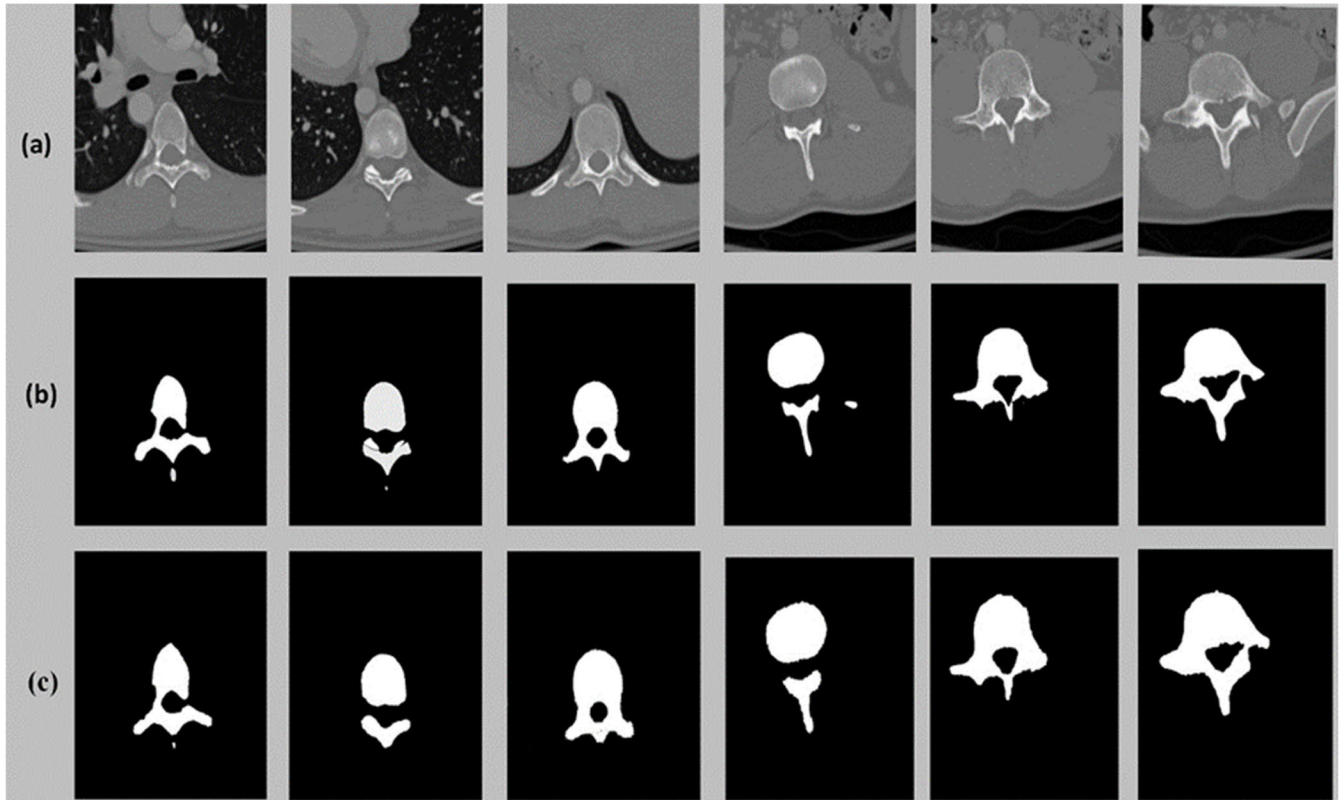


FIGURE 14. Qualitative results of segmentation (a) input images, (b) ground truth, and (c) predicted mask.

among other methods. In comparison to related studies, the reported results are considered to be excellent. It can be observed that OP-convNet outperformed both segmentation and classification-based segmentation networks. The experimental findings demonstrate the superiority of our method over other vertebra segmentation algorithms. Moreover, the promising results also show that proposed OP-convNet has exceptional generalization ability.

V. DISCUSSION

The results comparison in table 3 quantitatively verify the advantages of OP-convNet framework over existing methods. For example, U-Net [47] takes as input a whole image, whereas the proposed method utilizes overlapping patches based on the convNet architecture as input. The experimental results demonstrated that the U-Net did not perform well in segmentation due to a lack of local information and achieved 83.7% of Dice similarity coefficient (DSC), whereas the proposed OP-convNet performed significantly better (89.9% DSC). The local information may be kept more efficiently in the proposed method than the U-Net, where the slices are divided into overlapping patches, and the predictions are made independently for each patch. As a result, segmentation performance has improved. Another method D-TVNet [57] based on U-Net obtained a mean DSC of 86.17%. The experiment’s findings demonstrate that D-TVNet cannot identify the essential spots for measuring the angle of the spine curve

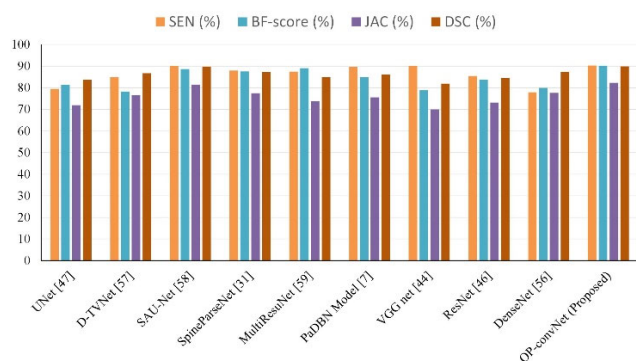
TABLE 3. Segmentation performance comparison against existing algorithms in term of SEN, BF-score, JAC, and DSC.

Networks	Models	SEN (%)	BF-score (%)	JAC (%)	DSC (%)
Segmentation Networks	UNet [47]	79.4	81.4	71.9	83.7
	D-TVNet [57]	84.9	78.2	76.5	86.7
	SAU-Net [58]	90.1	88.7	81.5	89.8
	SpineParseNet [31]	88.0	87.6	77.5	87.3
	MultiResuNet [59]	87.5	89.1	73.8	84.9
Classification-based Segmentation Networks	PaDBN Model [7]	89.7	84.9	75.6	86.1
	VGG net [44]	90.1	78.8	70.0	81.9
	ResNet [46]	85.4	83.7	73.2	84.5
	DenseNet [56]	77.9	80.0	77.6	87.4
	OP-convNet (Proposed)	<b>90.3</b>	<b>90.2</b>	<b>82.3</b>	<b>89.9</b>

when segmented bones are used. Furthermore, this method is unsuccessful at recognizing them when the bones are not sharp, and image noises are larger. Though D-TVNet method can to filter out some noise from images, in some cases, it can also eliminate essential bone pixels by mistake.

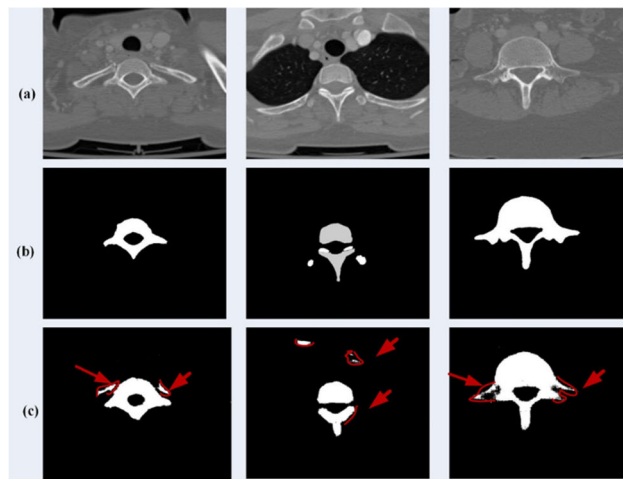
SAU-Net [58], a spatial attention-based densely connected U-Net, is proposed to illustrate the efficiency of

inter-slice information extraction for vertebrae segmentation. This approach makes a spine segmentation performance comparison with and without the interslice attention (ISA) module. However, the segmentation results (89.8% DSC) are not impressive, indicating that the ISA approach is likely to be misclassified. SpineParseNet [31] achieved a mean DSC of 87.3%, but several limitations exist in this work. First, it performed inferior segmentation due to the blurry boundary. Second, if the additional structures excluded from the spinal structures exist in the test data, the proposed approach failed to segment the additional structures. In this case, the output of the SpineParseNet should contain the additional structures, and more annotated data that contains the additional structures should be used to train the model. Third, the SpineParseNet failed to segment the top structures of the image due to the limited receptive field for the top of the image. Fourth, the dataset was collected from a single hospital. The robustness of the proposed method needs to be evaluated by multi-center datasets. While outperforming most U-Net variants, it can be observed that OP-convNet shows an improvement of 5% DSC on MultiuNet [59]. A classification based PaDBN segmentation model [7] is proposed for automatic CT vertebra segmentation and achieved an 86.1% DSC. This approach indicates that the initial thoracic vertebrae have a lower DSC due to the influence of the ribs and intervertebral discs; therefore, it wrongly segmented certain bones that were not visible in the label annotations, resulting in misclassification and a poor DSC. OP-convNet framework also outperformed other classical backbones in terms of DSC. Compared with the VGG net [44], ResNet [46], and DenseNet [56] proposed framework is significantly better by 8%, 5.4%, and 2.5% average in DSC respectively. Therefore, proposed model has strong predictive performance and application ability to assist planning and biomechanical analysis for vertebrae CT images segmentation. Figure 15 shows OP-convNet outperforms previous methods across all metrics in segmentation performance.



**FIGURE 15. Segmentation performance comparison, chart shows OP-convNet outperforms previous methods across all metrics in segmentation performance.**

There are several patterns of vertebrae, with certain patterns found at various spinal levels. For example, significant



**FIGURE 16. Poor segmentation results (a) input images, (b) ground truth, and (c) poor predicted mask.**

morphologic differences can be seen between two vertebrae with a large spatial separation within the spinal column, such as a lower lumbar vertebra and upper thoracic vertebra. Thus, proper segmentation of all vertebrae would be a difficult process. In order to achieve a robust segmentation, vertebra specific models require anatomical knowledge of the modeling process. The limitation of our proposed approach is found in some vertebra T6-T8, where poor segmentation has been observed due to the presence of rib structure. Also, vertebra T1 typically receives lower DICE scores than the other vertebrae. Figure 16 shows poor segmentation visualization results. This is probably because the T1 top vertebra is seen in certain datasets, whereas in other datasets, some cervical vertebrae are included. This sometimes causes the procedure to stretch the T1 vertebra to the C7 vertebra. This issue does not exist at the other end of the spinal column, specifically for the L5 vertebra, because all data sets include the sacrum, and L5 is not aligned with S5 in the sacrum. It is also possible to address these circumstances by obtaining more CNNs that are specifically trained to identify the C7/T1 transitions, as well as more 2D images or 3D image volumes datasets could be required. More datasets from more patients and other institutions will need to be corroborated in order to demonstrate the consistency, transferability, and robustness of the model.

## VI. CONCLUSION

This paper investigated the use of the deep learning OP-convNet framework for CT vertebrae segmentation. The results indicated that our proposed approach achieved outstanding precision, specificity, accuracy, F-score in terms of the patch-based classification accuracy, and BF-score, sensitivity (SEN), Jaccard index, dice similarity score (DSC) in terms of the segmentation accuracy on publically available MICCAI CSI dataset. Our research has demonstrated that by dividing the input slices into overlapping patches,



RUS-function for class balancing, and training the convNet framework on these overlapping patches, one can achieve superior segmentation performance over previous methods. The proposed OP-convNet architecture produces predictions for each input patch independently, retaining more local spatial information. Additionally, our framework outperforms prior methods in terms of all evaluation metrics with an overall DSC score of 89.9% for vertebral CT image segmentation. In a clinical application, segmented vertebrae can be used to prepare for subsequent automatic planning and positioning. The risk of deviation in a patient who undergoes an increased amount of bone biting can be minimized when pedicle screw implantation is employed. Additionally, single vertebra printing enables doctors to gain a better understanding of the tissue structure, which is especially beneficial for patients with scoliosis and enables surgeons to complete planning prior to surgery. Thus, our method improves the practicality and the accuracy of segmentation results in order to support clinical treatment without the complicated network architecture. Our future work will include a more extensive validation, improvement, and the possibility of implementing our proposed OP-convNet for additional tasks, such as multi-class classification and segmentation using 3D medical images. Additionally, future generations of GPUs with larger internal memory will enable the proposed model to examine different pathologies on a large dataset.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

- [1] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Dec. 2021.
- [2] D. Knez, B. Likar, F. Pernuš, and T. Vrtovec, "Computer-assisted screw size and insertion trajectory planning for pedicle screw placement surgery," *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1420–1430, Jun. 2016.
- [3] S. F. Qadri, S. A. Awan, M. Amjad, M. Anwar, and S. Shehzad, "Applications, challenges, security of wireless body area networks (WBANs) and functionality of IEEE 802.15.4/ZIGBEE," *Sci. Int.*, vol. 25, no. 4, pp. 697–702, 2013.
- [4] A. Suzani, A. Seitel, Y. Liu, S. Fels, R. N. Rohling, and P. Abolmaesumi, "Fast automatic vertebrae detection and localization in pathological ct scans—A deep learning approach," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, no. 7, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 678–686.
- [5] C. Chu, D. L. Belavý, G. Armbrecht, M. Bansmann, D. Felsenberg, and G. Zheng, "Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method," *PLoS One*, vol. 10, no. 11, pp. 1–22, 2015.
- [6] A. Sekuboyina, M. Rempfler, A. Valentinitich, B. H. Menze, and J. S. Kirschke, "Labeling vertebrae with two-dimensional reformations of multidetector CT images: An adversarial approach for incorporating prior knowledge of spine anatomy," *Radiol. Artif. Intell.*, vol. 2, no. 2, Mar. 2020, Art. no. e190074.
- [7] S. F. Qadri, D. Ai, G. Hu, M. Ahmad, Y. Huang, Y. Wang, and J. Yang, "Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in CT images," *Appl. Sci.*, vol. 9, no. 1, p. 69, 2019.
- [8] M. Pereanez, K. Lekadir, I. Castro-Mateos, J. M. Pozo, A. Lazary, and A. F. Frangi, "Accurate segmentation of vertebral bodies and processes using statistical shape decomposition and conditional models," *IEEE Trans. Med. Imag.*, vol. 34, no. 8, pp. 1627–1639, Aug. 2015.
- [9] B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "Shape representation for efficient landmark-based segmentation in 3-D," *IEEE Trans. Med. Imag.*, vol. 33, no. 4, pp. 861–874, Apr. 2014.
- [10] R. Korez, B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation," *IEEE Trans. Med. Imag.*, vol. 34, no. 8, pp. 1649–1662, Aug. 2015.
- [11] Y. Wang, J. Yao, H. R. Roth, J. E. Burns, and R. M. Summers, "Multi-atlas segmentation with joint label fusion of osteoporotic vertebral compression fractures on CT," in *Computational Methods and Clinical Applications for Spine Imaging* (Lecture Notes in Computer Science), vol. 9402. Cham, Switzerland: Springer, 2015, pp. 74–84.
- [12] J. S. Athertya and G. S. Kumar, "Automatic segmentation of vertebral contours from CT images using fuzzy corners," *Comput. Biol. Med.*, vol. 72, pp. 75–89, May 2016.
- [13] P. H. Lim, U. Bagci, and L. Bai, "A robust segmentation framework for spine trauma diagnosis," in *Computational Methods and Clinical Applications for Spine Imaging*. Nagoya, Japan: Springer, 2014, pp. 25–33.
- [14] B. Michael Kelm, M. Wels, S. Kevin Zhou, S. Seifert, M. Suehling, Y. Zheng, and D. Comaniciu, "Spine detection in CT and MR using iterated marginal space learning," *Med. Image Anal.*, vol. 17, no. 8, pp. 1283–1292, Dec. 2013.
- [15] D. Zukić, A. Vlasák, J. Egger, D. Hořinek, C. Nimsky, and A. Kolb, "Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images," *Comput. Graph. Forum*, vol. 33, no. 6, pp. 190–204, Sep. 2014.
- [16] A. Suzani, A. Rasoulalian, A. Seitel, S. Fels, and R. N. Rohling, "Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images," *Proc. SPIE*, vol. 9415, pp. 1–7, Mar. 2015.
- [17] H. Mirzaalian, M. Wels, T. Heimann, B. M. Kelm, and M. Suehling, "Fast and robust 3D vertebra segmentation using statistical shape models," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 3379–3382.
- [18] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
- [19] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, Oct. 2017.
- [20] 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.
- [21] M. Ahmad, Y. Ding, S. F. Qadri, and J. Yang, "Convolutional-neural-network-based feature extraction for liver segmentation from CT images," in *Proc. 11th Int. Conf. Digit. Image Process. (ICDIP)*, Aug. 2019, p. 159.
- [22] M. Ahmad, J. Yang, D. Ai, S. F. Qadri, and Y. Wang, "Deep-stacked auto encoder for liver segmentation," in *Proc. Chin. Conf. Image Graph. Technol.*, 2018, pp. 243–251.
- [23] X. Yang, C. Liu, Z. Wang, J. Yang, H. L. Min, L. Wang, and K.-T. T. Cheng, "Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI," *Med. Image Anal.*, vol. 42, pp. 212–227, Dec. 2017.
- [24] J.-P. Charbonnier, E. M. Van Rikxoort, A. A. Setio, C. M. Schaefer-Prokop, B. van Ginneken, and F. Ciompi, "Improving airway segmentation in computed tomography using leak detection with convolutional networks," *Med. Image Anal.*, vol. 36, pp. 52–60, Feb. 2017.
- [25] I. Hirra, M. Ahmad, A. Hussain, M. U. Ashraf, I. A. Saeed, S. F. Qadri, A. M. Alghamdi, and A. S. Alfakeeh, "Breast cancer classification from histopathological images using patch-based deep learning modeling," *IEEE Access*, vol. 9, pp. 24273–24287, 2021.
- [26] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, Jan. 2017.
- [27] A. Jamaludin, T. Kadir, and A. Zisserman, "SpineNet: Automated classification and evidence visualization in spinal MRIs," *Med. Image Anal.*, vol. 41, pp. 63–73, Oct. 2017.
- [28] N. Baka, S. Leenstra, and T. V. Walsum, "Ultrasound aided vertebral level localization for lumbar surgery," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2138–2147, Oct. 2017.



- [29] H.-J. Bae, H. Hyun, Y. Byeon, K. Shin, Y. Cho, Y. J. Song, S. Yi, S.-U. Kuh, J. S. Yeom, and N. Kim, "Fully automated 3D segmentation and separation of multiple cervical vertebrae in CT images using a 2D convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 184, Feb. 2020, Art. no. 105119.
- [30] S. F. Qadri, Z. Zhao, D. Ai, M. Ahmad, and Y. Wang, "Vertebrae segmentation via stacked sparse autoencoder from computed tomography images," in *Proc. 11th Int. Conf. Digit. Image Process. (ICDIP)*, Aug. 2019, p. 160.
- [31] S. Pang, C. Pang, L. Zhao, Y. Chen, Z. Su, Y. Zhou, M. Huang, W. Yang, H. Lu, and Q. Feng, "SpineParseNet: Spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 262–273, Jan. 2021.
- [32] S. F. Qadri, M. Ahmad, D. Ai, J. Yang, and Y. Wang, "Deep belief network based vertebra segmentation for CT images," in *Proc. Chin. Conf. Image Graph. Technol.*, vol. 757, 2018, pp. 536–545.
- [33] A. Sekuboyina, A. Valentinitich, J. S. Kirschke, and B. H. Menze, "A localisation-segmentation approach for multi-label annotation of lumbar vertebrae using deep nets," 2017, *arXiv:1703.04347*.
- [34] A. Sekuboyina, J. Kukačka, J. S. Kirschke, B. H. Menze, and A. Valentinitich, "Attention-driven deep learning for pathological spine segmentation," in *Computational Methods and Clinical Applications in Musculoskeletal Imaging* (Lecture Notes in Computer Science), vol. 10734. Cham, Switzerland: Springer, 2018, pp. 108–119.
- [35] R. Janssens, G. Zeng, and G. Zheng, "Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 893–897.
- [36] N. Lessmann, B. van Ginneken, and I. Išgum, "Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images," *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 1057408.
- [37] H. Liao, A. Mesfin, and J. Luo, "Joint vertebrae identification and localization in spinal CT images by combining short- and long-range contextual information," *IEEE Trans. Med. Imag.*, vol. 37, no. 5, pp. 1266–1275, May 2018.
- [38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [39] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
- [40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.
- [41] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
- [42] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," in *Proc. Nature Inspired Smart Inf. Syst. (NISIS), EU Co-Ordination Action*, Albufeira, Portugal, 2005, pp. 1–9. [Online]. Available: <http://www.nisis.de>
- [43] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
- [44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [45] 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.
- [46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Comput. Assist. Intervent.*, 2015, pp. 234–241.
- [48] 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.
- [49] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*.
- [50] Y. Kang, K. Engelke, and W. A. Kalender, "A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 586–598, May 2003.
- [51] J. Yao, J. E. Burns, D. Forsberg, A. Seitel, A. Rasoulian, P. Abolmaesumi, K. Hammernik, M. Urschler, B. Ibragimov, R. Korez, and T. Vrtovec, "A multi-center milestone study of clinical vertebral CT segmentation," *Comput. Med. Imag. Graph.*, vol. 49, pp. 16–28, Apr. 2016.
- [52] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, pp. 1–28, Dec. 2015.
- [53] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Vaudoise des Sci. Naturelles*, vol. 37, pp. 547–579, 1901.
- [54] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [55] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.
- [56] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.
- [57] J. Lyu, X. Bi, S. Banerjee, Z. Huang, F. H. F. Leung, T. T.-Y. Lee, D.-D. Yang, Y.-P. Zheng, and S. H. Ling, "Dual-task ultrasound spine transverse vertebrae segmentation network with contour regularization," *Comput. Med. Imag. Graph.*, vol. 89, Apr. 2021, Art. no. 101896.
- [58] Y. Zhang, L. Yuan, Y. Wang, and J. Zhang, "SAU-Net: Efficient 3D spine MRI segmentation using inter-slice attention," in *Medical Imaging With Deep Learning*, 2020, pp. 903–913.
- [59] 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.



**SYED FURQAN QADRI** received the Ph.D. degree in computer science and technology from BIT, Beijing, China, in 2019. He is currently working as a Postdoctoral Fellow at the Institute of Computer Vision, Shenzhen University. His research interests include computer vision, applied deep learning, image segmentation, classification, medical image processing, and pattern recognition. He has conducted extensive research and authored many peer-reviewed SCI- and EI-indexed scientific papers. He has received the Best Paper Awards from ICDIP-SPIE 2019 and IGTA-Springer 2017. He also serves as a Reviewer for IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, IEEE ACCESS, *Cancers*, and other journals from IEEE, MDPI, Wiley, SPIE, Taylor & Francis, and Tech Science Press (TSP).



**LINLIN SHEN** (Member, IEEE) received the B.Sc. and M.Eng. degrees from Shanghai Jiao Tong University, Shanghai, China, and the Ph.D. degree from the University of Nottingham, Nottingham, U.K. He was a Research Fellow at the University of Nottingham, U.K., working on MRI brain image processing. He is currently the "Pengcheng Scholar" Distinguished Professor with the School of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. He is also an Honorary Professor with the School of Computer Science, University of Nottingham. He serves as the Director of the Computer Vision Institute, the AI Research Center for Medical Image Analysis and Diagnosis, and the China-U.K. Joint Research Laboratory for Visual Information Processing. His research interests include deep learning, facial analysis, and medical image processing. He was listed as the Most Cited Chinese Researcher by Elsevier. He received the Most Cited Paper Award from the *Image and Vision Computing* journal. His cell classification algorithms were the winners of the International Contest on Pattern Recognition Techniques for Indirect Immunofluorescence Images held by ICIP 2013 and ICPR 2016. He serves as the Co-Editor-in-Chief for the *IET Cognitive Computation and Systems* journal.



**MUBASHIR AHMAD** received the B.S. degree (Hons.) in computer science from Hazara University Mansehra, Pakistan, the M.S. degree in computer science from IQRA University, Islamabad, Pakistan, in 2012, and the Ph.D. degree from the Beijing Institute of Technology, China, in 2019. Currently, he is an Assistant Professor with the Computer Science Department, The University of Lahore, Pakistan. His research interests include deep learning, pattern recognition, and medical imaging. He was a recipient of a Chinese government scholarship for his Ph.D. studies and the Excellent Research Paper Award of IGTA-Springer, Beijing, China, in 2017. He also serves as a Reviewer for IEEE Access.

**SYEDA SHAMAILA ZAREEN** received the bachelor's degree in software engineering from the University of Karachi, Pakistan, in 2015, and the master's degree in computer science from Mohammad Ali Jinnah University, Karachi, Pakistan, in 2018. She is currently a Doctoral Fellow at the Department of Information Technology, Beijing University of Technology, Beijing, China. Her research interests include machine learning, image processing, pattern recognition, medical image analysis, semantic web, and data mining.



**SALMAN QADRI** received the M.C.S. degree in computer science from BZU, in 2004, the M.S. degree in computer science from UAF, Pakistan, in 2008, and the Ph.D. degree from the Computer Science and IT Department, The Islamia University of Bahawalpur, Pakistan, in 2017. From 2006 to 2011, he worked as a Lecturer with the Virtual University of Pakistan and COMSATS University Islamabad (Vehari campus). He is currently working as an Associate Professor with the Computer Science Department, MNS-University of Agriculture, Multan, Pakistan. His research interests include machine learning, computer vision, bio informatics, and medical image analysis.



**SALABAT KHAN** received the Ph.D. degree from the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. He is a Postdoctoral Fellow at the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His current research interests include secure and transparent public-key infrastructure (PKI), transparent and secure key management in VANETs, distributed ledger technology, and blockchain.

...