2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA) | 978-1-6654-1914-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/MeMeA52024.2021.9478765

# Application of Hybrid Network of UNet and Feature Pyramid Network in Spine Segmentation

Xingxing Liu
Department of Electrical and Computer Engineering
Iowa Technology Institute
The University of Iowa
Iowa City, Iowa, USA
xingxing-liu@uiowa.edu

Wenxiang Deng
Department of Electrical and Computer Engineering
Iowa Technology Institute
The University of Iowa
Iowa City, Iowa, USA
dengwenxiang@gmail.com

Yang Liu
Department of Electrical and Computer Engineering
Iowa Technology Institute
The University of Iowa
Iowa City, Iowa, USA
yang-liu-ece@uiowa.edu

***Abstract*—Spine segmentation is a common task for spinal imaging and spinal surgical navigation. Spine segmentation provides valuable information for the diagnosis, and the segmentation output can also serve as an input for downstream surgical navigation. Unfortunately, spine segmentation is a labor-intensive task. In this study, we applied a deep network combining feature pyramid network (FPN) and UNet to the segmentation of vertebral bodies (VBs), referring as Res50_UNet. Compared with the original UNet, Res50_UNet has the following enhancements: 1) five consecutive spine MRI slices and two coordinate maps are concatenated as the input; 2) the convolutional block from ResNet are used; 3) an FPN architecture is applied to extracting rich multi-scale features and obtaining segmentation output. Experiments were conducted on an annotated T2-weighted MRIs of the lower spine dataset. We have benchmarked Res50_UNet against UNet and other UNet based network structures. It was found that Res50_UNet needs the lowest number of epochs (~1000 epochs) to achieve steady-state performance. The accuracy (AC) of Res50_UNet is higher than 99.5% with only 1000 epochs, which is very impressive. This study demonstrated the feasibility of applying Res50_UNet in spine segmentation. The network integrates the characteristics of FPN and UNet. These results have shown the potential for Res50_UNet in spine MRI segmentation, especially when a low number of epochs is desirable.**

***Keywords*—Spine segmentation, deep learning, medical image processing, computer vision**

## I. INTRODUCTION

Spine segmentation is a common task for spinal imaging and spinal surgical navigation. Historically, spine segmentation of medical images, such as nuclear magnetic resonance imaging (MRI) and x-ray computed tomography (CT), is performed manually. Spine segmentation provides valuable information for the diagnosis, and the segmentation output can also serve as an input for downstream surgical navigation. For example, [1] firstly utilizes a convolutional neural network (CNN) to get the segmentation of the vertebrae from x-ray images, then measures the Cobb angle [2] to assess the spine curvature. Unfortunately, spine segmentation is a labor-intensive task.

Image segmentation is a common task in computer vision and image processing. In recent years, deep learning has become a powerful tool. In [3], Long, et al. proposed a fully convolutional network (FCN) for image semantic segmentation. An FCN takes an image with arbitrary size as input, and through a sequence of convolutional layers, it outputs a high-resolution segmentation mask with the same size as the input image. Feature pyramid network (FPN), proposed by Lin et al. [4], was initially developed for object detection; owing to its character of multi-scale analysis, FPN could also be applied to image segmentation tasks [5-6].

Compared with natural images, medical images usually have lower contrast. For example, the boundary between vertebral bodies (VBs) and discs are not always clear, which makes it difficult for segmentation task. UNet [7] is a fully convolutional neuron network with a symmetric encoder-decoder model and skip connections between the encoder and decoder blocks. UNet provides high-resolution feature maps to the decoder block and overcomes the trade-off between localization and the use of context information, thereby achieving accurate segmentation for medical images.

In this study, we adopted and modified a network that combines UNet and FPN for spine segmentation [8]. This type of network has been used for image segmentation of natural images, but not yet in spine segmentation. Compared with the original UNet, our UNet/FPN hybrid has three main enhancements. First, we increase the number of channels of the input from one to seven. The input includes five spine MRI slices and two coordinate maps sharing the same size as the MRI slices. This modification not only incorporates more context

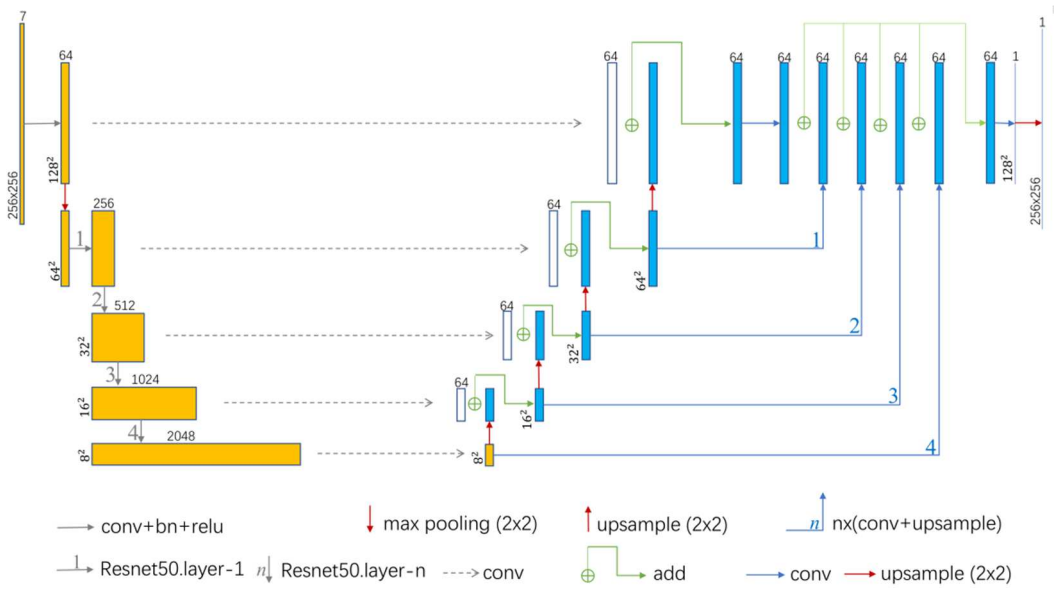


**Fig. 1**. The architecture of Res50_UNet.

information to the input, but also adds two coordinate maps, which provide additional information of the spine orientation to the network. Second, we replace the original encoder with the convolutional part of ResNet [9]. More specifically, we use an implemented network architecture "resnet50_32x4d" in PyTorch. This replacement substantially increases the depth of the network. Third, an FPN architecture is applied to extracting rich multi-scale features and obtaining segmentation output. In this manuscript, the implemented network is referred as Res50_UNet.

## II. Methods

### *A. Network*

UNet has a symmetric encoder-decoder architecture, including a contracting path and an expansive path. The contracting path consists of a sequence of convolution layers, each followed by a rectified linear unit (ReLU), and a max-pooling operation for downsampling. On the expansive path, through a sequence of upsampling of the feature map followed by a convolution layer ("up-convolution"), the scale of the feature map expands as the depth of the network increases. In addition, with skip connections, the output feature maps on the contracting path, which retain more context information, are concatenated with corresponding upsampled feature maps on the expansive path.

The Res50_UNet also has an overall "U" structure, but the overall architecture has three main modifications compared with the original UNet. First, on the contracting path, the original encoder blocks are replaced with that of ResNet. The downsampling operation of the feature maps on the contracting path is achieved by a sequence of convolution with stride 2 instead of max-pooling. Second, the concatenation operation connecting the output feature maps of the encoder blocks and the upsampled feature maps on the expansive path is removed; instead, the convoluted feature maps on the contracting path are added to the upsampled feature maps on the expansive path. Third, all the output feature maps of each decoder block on the expansive path, which have different scales, are combined with an FPN architecture to form the final output segmentation result. Fig. 1 shows the architecture of Res50_UNet.

For our application, the input of Res50_UNet is a concatenation of five consecutive MRI slices and two coordinate maps. Five consecutive MRI slices provide more context information to the network, while two coordinate maps inform the network additional information of the orientation of the segmentation targets. The pixel value of coordinate map x in the vertical direction is constant, while it increases linearly from zero to one as the coordinate increases in the horizontal direction. For coordinate map y, the pixel value of coordinate map y in the horizontal direction is constant, while it increases linearly from zero to one as the coordinate increase in the vertical direction. Fig. 2 displays one sample of the input.

For comparison, we have implemented UNet and another two UNet based networks: UNet_BN [1] and UNet_Dense [10]. In [1], the authors modified the original UNet by adding batch normalize (BN) layer to each convolution + ReLU block. UNet_Dense combines UNet and dense block [11]. Fig 3, Fig. 4 and Fig. 5

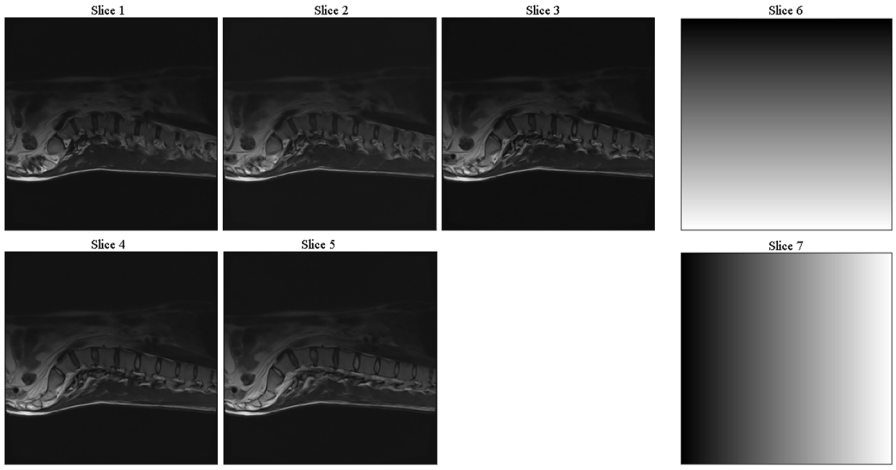


**Fig. 2**. One sample of the input of the network: five consecutive spine MRI slices (slice 1 - 5) and two coordinate maps (slice 6 - 7) of the same size. Slice 6 is coordinate map y, and slice 7 is coordinate map x.

illustrate the architectures of our implemented UNet, UNet_BN, UNet_Dense, respectively.

### *B. Dataset*

We use a spine dataset [12], which consists of T2-weighted turbo spin-echo MR spine images of 23 patients, to train the network and test the performance of the trained modules. The spine MRIs of each patient contains at least 7 VBs of the lower spine (T11 – L5). The ground truth segmentation of each spine MRI is a binary mask manually segmented. The size of each spine MRI and its corresponding segmentation mask is 305 × 305, and they are resized to 256 × 256 for training the network. Fig. 6 shows a spine image, its ground truth segmentation mask, and the superimposed image (segmentation is pseudocolored in red).

## III. EXPERIMENTS

To characterize the performance of networks, we used the negative of the natural logarithm of the Dice similarity coefficient (DSC) [13] between the ground truth segmentation mask (GT) and the predicted segmentation mask (SR) of the network as the loss function. The following two equations represent the definition of DSC and loss function, respectively. To accommodate for scenarios where the denominator of the DSC equation become zero, we added a constant, Smooth, to both numerator and denominator in the DSC equation. The value of constant Smooth was set as 1.

$$dice\ similarity\ coefficient\ (DSC) = \frac{2|GT\ \cap\ SR|}{|GT|\ +\ |SR|}$$

$$loss = -\ln\left(\frac{2|(GT\ \cap\ SR) + Smooth|}{|GT|\ +\ |SR| + Smooth}\right)$$

We randomly assigned the dataset into a training set and a testing set with the ratio of 4:1 on patient level. More specifically, images from 19 patient MRIs were assigned into the training set and images from 4 patient MRIs were assigned into the testing set. Given 39 MRI slices per patient, there are 741 spine MRI slices in the training set and 156 spine MRI slices in the testing set. We use Adam to optimize the training process and ReduceLROnPlateau to reduce the learning rate. Because Res50_UNet incorporates some layers of "resnet50_32x4d" in PyTorch, so pretrained layers on ImageNet are available. In this paper, we provide the segmentation results output by both pretrained Res50_UNet and non-pretrained Res50_UNet. When training Res50_UNet, UNet_BN and UNet_Dense, the learning rate is set as 0.005. But setting learning rate as 0.005 caused loss explosion when training UNet, so the learning rate for training UNet is 0.00001. Batchsize is set as 16. The training epochs are 10,000 in total. The experiment was conducted on Google Colab. Training Res50_UNet on the training set takes approximately 6h.

## IV. RESULTS AND DISCUSSIONS

### *A. Qualitative Display of the Segmentation Results*

We have implemented the original UNet and three other networks based on UNet, including Res50_UNet, UNet_BN, and UNet_Dense. The performance of these networks on segmenting the spine MRIs were investigated and compared. Fig. 7 displays three different spine MRIs randomly selected from three different patients from the testing set, respectively, with ground truth segmentation and output segmentation results of four aforementioned networks. After 7000 epochs of training, the loss has converged. As showed in Fig. 7, Res50_UNet achieves impressive segmentation performance. For example, in the second row of Fig. 7, Res50_UNet can detect the VB, which is not labeled in the segmentation ground truth. For UNet, UNet_BN and UNet_Dense, the segmentation results are comparable to Res50_UNet.

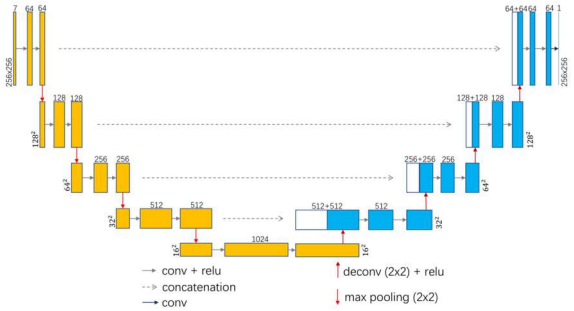


**Fig. 3**. The architecture of UNet [7].

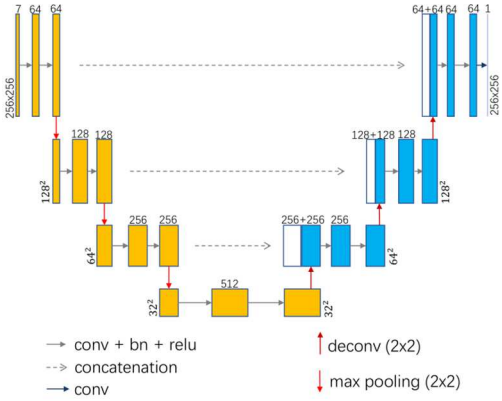


**Fig. 4**. The architecture of UNet_BN [1].

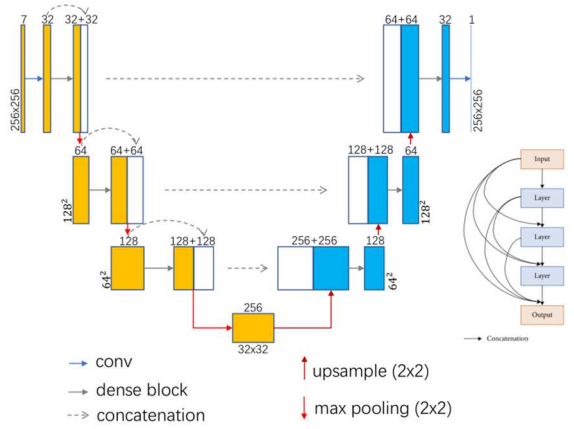


**Fig. 5**. The architecture of UNet_Dense (left side) [10] and a dense block (right side) [11].

### *B. Quantitative Evaluation of the Segmentation Results*

To quantitatively characterize the performance of four networks, we use six metrics, including accuracy (AC), sensitivity (SE), specificity (SP), Dice similarity coefficient (DSC), Jaccard similarity (JS) [14], and mean square error (MSE). Their definitions are represented by the following equations, where TP, TN, FP, FN denote the number of true positive, true negative, false positive, false negative segmented pixels, respectively, while GT, SR represent the ground truth segmentation mask and predicted segmentation mask output by the network, respectively. We calculate the mean values of the six

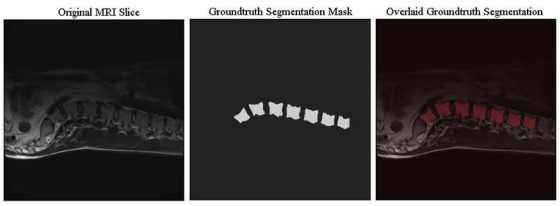


**Fig. 6**. One sample of spine MRI from the dataset and its corresponding ground truth segmentation. Column 1: original MRI slice; column 2: groundtruth segmentation mask; column 3: overlaid groundtruth segmentation.

metrics on the testing set as standard to quantitatively evaluate the performance of these networks. Given that some of these metrics calculated on images containing small part of or even no VB are really low, only images whose groundtruth segmentation mask containing more than 1500 VB pixels will be taken into consideration. Therefore, 79 MRI slices on the testing set were chosen to calculate these metrics.

$$accuracy\ (AC) = \frac{TP + TN}{TP + TN + FP + FN}$$

$$sensitivity\ (SE) = \frac{TP}{TP\ +\ FN}$$

$$specificity\ (SP) = \frac{TN}{TN +\ FP}$$

$$Dice\ similarity\ coefficient\ (DSC) = \frac{2|GT\ \cap\ SR|}{|GT|\ +\ |SR|}$$

$$Jaccard\ similarity\ (JS) = \frac{|GT\ \cap\ SR|}{|GT\ \cup\ SR|}$$

$$mean\ square\ error(MSE) = \frac{1}{N}\sum_{i=1}^{N}(GT_i - SR_i)^2$$

We trained four networks and saved the module every 1000 epochs, then calculated the mean values of these metrics of each module on the chosen MRI slices on the testing set to see how the performance of networks changes as epoch increases. It was found that Res50_UNet needs the lowest number of epochs (~1000 epochs) to achieve steady-state performance (Fig. 8). The accuracy (AC) of Res50_Unet is higher than 99.5% with only 1000 epochs, which is very impressive. At steady state (> 6000 epochs), all four networks have comparable performance. These results have shown the potential for Res50_UNet in spine MRI segmentation when a low number of epochs is desirable. Also, it is noted that pretrained Res50_UNet (labeled with pink circle) outperforms non-pretrained Res50_UNet (labeled with red circle).

### *C. Limitations and Future Work*

This study established the initial feasibility of applying Res50_UNet to spine segmentation. In this study, we investigated VB segmentation on T2 MRI data. In the future, we plan to optimize the network structure further and investigate the application of Res50_UNet in other domains of spine segmentation.

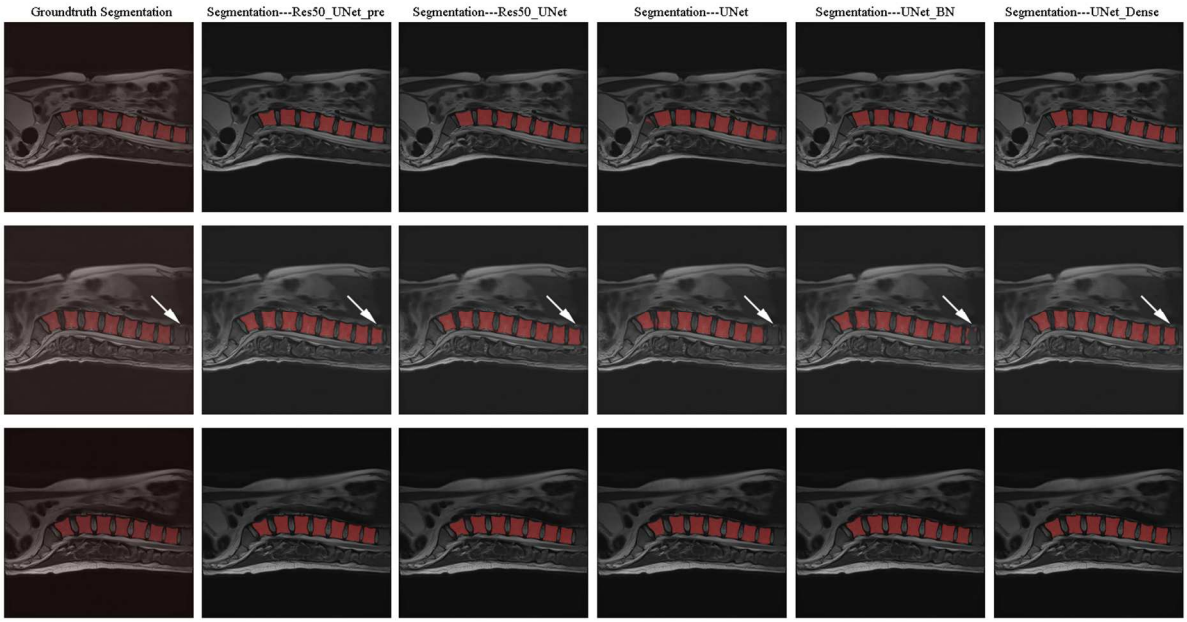


**Fig. 7**. Segmentation results of three spine MRIs from four networks. Column 1: groundtruth segmentation; column 2: segmentation of Res50_UNet based on pretrained module; column 3: segmentation of Res50_UNet trained on T2-weighted spine MRIs only; column 4: segmentation of UNet; column 5: segmentation of UNet_BN; column 6: segmentation of UNet_Dense. Three rows correspond to three different spine MRIs from three different patients. White arrows in the second row indicate the VB not marked on ground truth but segmented by networks.

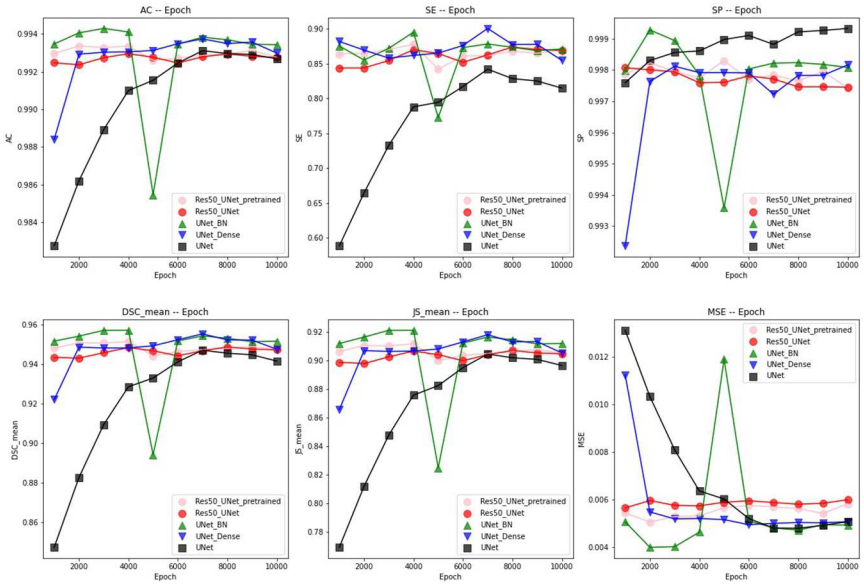


**Fig. 8**. Metrics evaluating the segmentation performance of four networks. Each small figure corresponds to one metric, and we use different colored marks to represent four networks. Pink circle: Res50_UNet based on pretrained module; red circle: Res50_UNet trained on T2-weighted spine MRIs only; green upward triangle: UNet_BN; blue downward triangle: UNet_Dense, black square: UNet.

## V. Conclusions

This study demonstrated the feasibility of applying Res50_UNet in spine segmentation. The network integrates the characteristics of FPN and UNet. High accuracy was achieved with a low number of epochs. These results have shown the potential for Res50_UNet in spine MRI segmentation, especially when a low number of epochs is desirable.

## Acknowledgment

This project is supported in part by University of Iowa Startup Funds and Cottrell Foundation Research Grant.

## REFERENCES


[1] M. Horng *et al.*, "Cobb Angle Measurement of Spine from X-Ray Images Using Convolutional Neural Network," *Computational and Mathematical Methods in Medicine,* vol. 2019, 2019, 2019.

[2] J. Cobb, *Outline for the Study of Scoliosis*, vol. 5, Instructional Course Lectures, 1948.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *2015 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*, pp. 3431-3440, 2015, 2015.

[4] T. Lin *et al.*, "Feature Pyramid Networks for Object Detection," *30th Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017)*, pp. 936-944, 2017, 2017.

[5] H. Zhao *et al.*, "Pyramid Scene Parsing Network," *30th Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017)*, pp. 6230-6239, 2017, 2017.

[6] G. Ghiasi *et al.*, "Laplacian Pyramid Reconstruction and Refinement for Semantic Segmentation," *Computer Vision - Eccv 2016, Pt Iii,* vol. 9907, pp. 519-534, 2016, 2016.

[7] O. Ronneberger *et al.*, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention, Pt Iii,* vol. 9351, pp. 234-241, 2015, 2015.

[8] A. Kirillov *et al.*, "Panoptic Feature Pyramid Networks," *2019 Ieee/cvf Conference on Computer Vision and Pattern Recognition (Cvpr 2019)*, pp. 6392-6401, 2019, 2019.

[9] K. He *et al.*, "Deep Residual Learning for Image Recognition," *2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*, pp. 770-778, 2016, 2016.

[10] S. Jegou *et al.*, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation," *2017 Ieee Conference on Computer Vision and Pattern Recognition Workshops (Cvprw)*, pp. 1175-1183, 2017, 2017.

[11] G. Huang *et al.*, "Densely Connected Convolutional Networks." pp. 2261-2269.

[12] C. Chu *et al.*, "Fully Automatic Localization and Segmentation of 3D Vertebral Bodies from CT/MR Images via a Learning-Based Method," *Plos One,* vol. 10, no. 11, NOV 23 2015, 2015.

[13] L. DICE, "MEASURES OF THE AMOUNT OF ECOLOGIC ASSOCIATION BETWEEN SPECIES," *Ecology,* vol. 26, no. 3, pp. 297-302, 1945, 1945.

[14] P. Jaccard, "The Distribution of the Flora in the Alpine Zone," *New Phytologist,* vol. 11, no. 2, pp. 37-50, 1912.