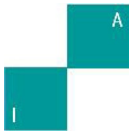
Inteligencia Artificial 27(74), 22-36
doi: 10.4114/intartif.vol27iss74pp22-36



INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

A Novel Deep Learning Model for Pancreas Segmentation: Pascal U-Net

Ender Kurnaz[1,A], Rahime Ceylan[1], Mustafa Alper Bozkurt[2], Hakan Cebeci[2], Mustafa Koplay[2]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering and Nature Sciences, Konya Technical University, Konya, Turkey
[A]ekurnaz@ktun.edu.tr
[2]Department of Radiology, Faculty of Medicine, Selcuk University, Konya, Turkey

Abstract A robust and reliable automated organ segmentation from abdomen images is a crucial problem in both quantitative imaging analysis and computer-aided diagnosis. In particular, automatic pancreas segmentation from abdomen CT images is the most challenging task based on two main aspects (1) high variability in anatomy (like as shape, size, etc.) and location across different patients and (2) low contrast with neighbouring tissues. Due to these reasons, the achievement of high accuracies in pancreas segmentation is a hard image segmentation problem. In this paper, we propose a novel deep learning model which is a convolutional neural network-based model called Pascal U-Net for pancreas segmentation. The performance of the proposed model is evaluated on The Cancer Imaging Archive (TCIA) Pancreas CT database and abdomen CT dataset which is taken from Selcuk University Medicine Faculty Radiology Department. During the experimental studies, the k-fold cross-validation method is used. Furthermore, the results of the proposed model are compared with the results of traditional U-Net. If results obtained by Pascal U-Net and traditional U-Net for different batch sizes and fold number is compared, it can be seen that experiments on both datasets validate the effectiveness of the Pascal U-Net model for pancreas segmentation.

Resumen La segmentación automatizada, robusta y fiable de órganos a partir de imágenes abdominales es un problema crucial tanto en el análisis cuantitativo de imágenes como en el diagnóstico asistido por ordenador. En particular, la segmentación automática del páncreas a partir de imágenes de TC abdominales es la tarea más difícil y se basa en dos aspectos principales: (1) la gran variabilidad de la anatomía (como la forma, el tamaño, etc.) y la ubicación en diferentes pacientes (2) el bajo contraste con los tejidos vecinos. Debido a estas razones, el logro de altas precisiones en la segmentación del páncreas es un problema difícil de segmentación de imágenes. En este artículo, proponemos un nuevo modelo de aprendizaje profundo basado en una red neuronal convolucional llamada Pascal U-Net para la segmentación del páncreas. El rendimiento del modelo propuesto se evalúa en la base de datos de tomografía computarizada de páncreas de The Cancer Imaging Archive (TCIA) y en el conjunto de datos de tomografía computarizada de abdomen del Departamento de Radiología de la Facultad de Medicina de la Universidad de Selcuk. Durante los estudios experimentales se utiliza el método de validación cruzada k-fold. Además, los resultados del modelo propuesto se comparan con los de la U-Net tradicional. Si se comparan los resultados obtenidos por la Pascal U-Net y la U-net tradicional para diferentes tamaños de lote y número de pliegues, se observa que los experimentos realizados en ambos conjuntos de datos validan la eficacia del modelo Pascal U-Net para la segmentación del páncreas.

Keywords: Pancreas Segmentation, Deep Learning, Pascal U-Net, U-Net.
Palabras clave: Segmentación del páncreas, Deep Learning, Pascal U-Net, U-Net.

ISSN: 1137-3601 (print), 1988-3064 (on-line)

1 Introduction

The Pancreas, which is one of the organs located in the abdominal region of the human body, consists of two basic tissues, endocrine and exocrine. These tissues play an important role in digestion and glucose metabolism. The pancreas is one of the rare organs that performs both internal and external secretion and secretes the enzymes necessary for the digestion of the three basic nutrients carbohydrates, proteins, and fats. In the detection of pancreas-related diseases, blood and urine tests are first requested from the patient. If abnormalities are detected in these tests, imaging methods can be used. Magnetic Resonance Cholangiopancreatography Imaging, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Endoscopic Ultrasound methods are used as imaging methods. Pancreas segmentation plays an important role as input to many computer-aided diagnosis systems that can provide quantitative analysis for diabetic patients. Accurate segmentation may also be required for other computer-aided detection methods to detect pancreatic cancer. The pancreas can vary in shape, position, and size in each person due to its structure. As a result of these differences, it can be quite challenging image segmentation problem to detect the pancreas. Some studies try to overcome this problem in the literature.

Various studies in the literature show that deep learning methods are gaining more and more importance in pancreas segmentation. Roth et al. [12] created probability maps by extracting superpixels from abdomen CT images using the Simple Linear Recurrent Clustering method. Pancreas segmentation was performed using probability maps and Convolutional Neural Network (CNN) structure, and the result of the Dice Similarity Coefficient (DSC) is 0.68. Farag et al. [3] obtained the CT images they used in the introduction as maps consisting of two parts, the superpixel map and the Random Forest probability map. In their study, the most remarkable result was obtained as 0.70 DSC by Convolutional Neural Network architecture. Zhou et al. [21] divided 3D CT images into three angles axial, sagittal, and coronal in their study. In the first stage, they created a rough segmentation map of the pancreas by using the 2D Fully Convolutional Network method separately at each angle. They combined these segmentation maps into 3D and then cropped them to be the input images of the second stage. They reconstructed these images into three angles and created precise pancreatic segmentation maps. In their study, DSC was obtained as 82.4%. Roth et al. [15] used a 3D U-Net model to perform multiple segmentation of organs in the abdominal region. This 3D U-Net model consists of two stages, and they used the segmentation results obtained in the first stage as the input to the second stage. In the first stage, they roughly determined the regions where organs and vessels could be, and so they facilitated the classification process to be made in the second stage. In their study, they obtained 0.82 DSC for the pancreas. Liu et al. [8] developed a two-stage, ensemble-based fully convolutional network model and performed pancreatic segmentation. In the first stage, they determined the candidate region from 3D CT images using superpixels and created 2.5D images. Then, in the second stage, they trained the network with 5 U-Net models containing different loss functions and obtained these five different results as segmentation output. They found the DSC as 0.84 in the study with 4-fold cross-validation. Yang et al. [18] performed the segmentation of the pancreas with inter- and intra-sectional information using a cascade neural network structure consisting of Fully Convolutional Network (FCN) and Recurrent Neural Network (RNN) structures. They used the FCN structure to extract intra-slice information for pancreatic segmentation and the RNN structure to extract inter-slice information. This study, which was carried out with the 4-fold cross-validation method, reached the DSC of 0.87. Man et al. [10] proposed a Deep Q Network method for the pancreas segmentation problem. The axial, coronal, and sagittal sections of the 3D pancreas images were determined separately by the Deep Q Network, and then they obtained the 3D segmentation of the pancreas with the U-Net model. They combined the segmentation maps obtained from 3 different angles into a single result using the Majority Vote method. As a result of their study, they found the average DSC as 0.86.

In this study, a U-Net model based on CNN has been developed and the proposed method is called Pascal U-Net. Pancreas segmentation was performed by Pascal U-Net cross-validation. The Pascal U-Net model was inspired by the Pascal number triangle and U-Net model. Separate studies were carried out for the batch sizes from 1 to 10, respectively for 2-, 4-, and 6-fold cross-validation and the results were obtained by taking the average of the results obtained from 10 repetitions for training and testing for each batch size.

The rest of this paper is organized as follows. The datasets used and the proposed methodology is discussed in Section 2. In Section 3, the experimental setup and the results are discussed. Finally, this paper concludes with a summarization of the findings and potential future research directions in Section 4.

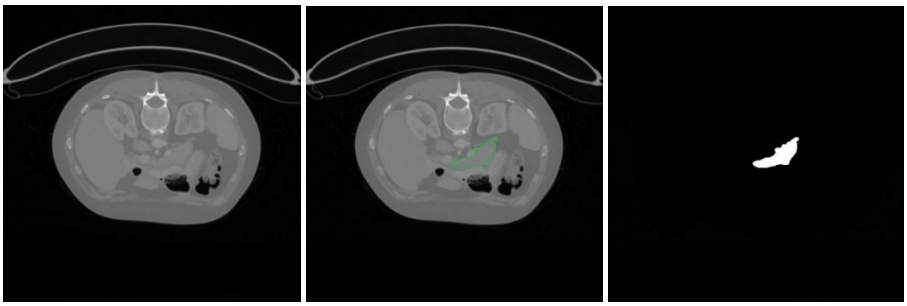
2 Materials and Methods

The Pascal U-Net model proposed for pancreas segmentation and the features of the dataset used are presented in detail in this section. The performance of the proposed Pascal U-Net model was tested on two different datasets and the results were compared with the traditional U-Net model.

2.1 Features of Datasets

In this study, two different datasets consisting of abdominal CT images were used. The first of these is the publicly available Pancreas CT dataset taken from The Cancer Imaging Archive (TCIA) database. The other is a dataset consisting of abdominal CT images taken from the Radiology Department of Selcuk University Medical Faculty Hospital (SUMFH). Due to the limited graphics card of the computer used in the study, datasets for deep learning models were created by selecting a slice in which the pancreas could be segmented from each patient in both datasets.

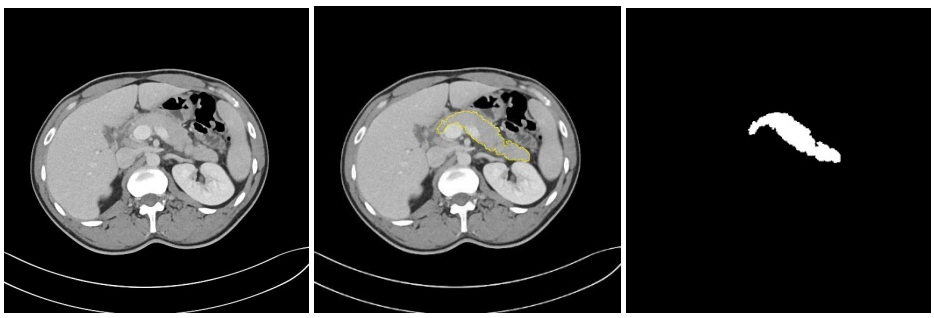
The first dataset used in the study is the Pancreas CT dataset taken from the TCIA database [13]. In this dataset taken from the National Institutes of Health Clinical Center in the United States, there are 82 contrast-enhanced 512 x 512 pixels CT images taken from 53 men and 27 women aged 18 to 76 years. An original CT image in the TCIA dataset, the labeled version of this image, and the mask of the desired segmentation can be seen in Figure 1.



(a) (b) (c)

Figure 1. A CT image from the TCIA database (512x512 pixels): (a) Original image, (b) Labeled image, (c) Mask

The images used as the second dataset in the study were created from abdominal CT images taken from the Department of Radiology at Selcuk University Medical Faculty Hospital (SUMFH). There are 58 CT images taken from 30 men and 28 women in the SUMFH dataset. All patients were examined on a 128-row dual-source CT system (Somatom Definition Flash; Siemens Healthcare, Forchheim, Germany), on which 2 x-ray tubes were arranged at a 95-degree angle to each other and contained 2 detector sets. The examination protocol was as follows: 120 kVp, 512 x512 matrix, 64 x 0.6-mm collimation. The images in this dataset have been labeled by a radiologist who is an expert in his field in the Radiology Department of SUMFH. Figure 2 shows an original CT image in the SUMFH dataset, the labeled version of this image, and the mask of the labeled image.



(a) (b) (c)

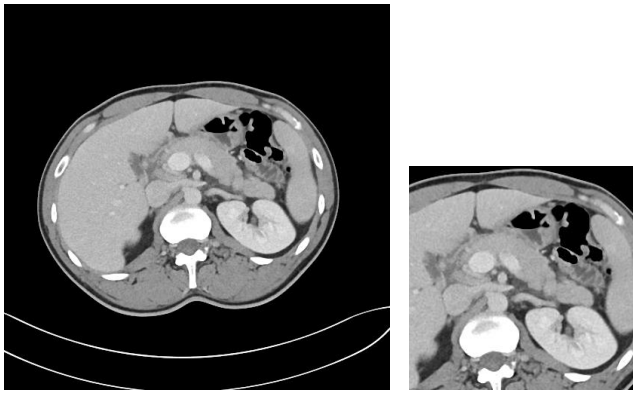
Figure 2. A CT image from SUMFH database (512x512 pixels): (a) Original image, (b) Labeled image, (c) Mask

2.2 Preprocessing

2.2.1 Determination of ROI

Before the images in each dataset are given to the deep learning network, preprocessing processes which consist of several steps are applied. As can be seen in Figure 3, the regions of interest (ROI) where the pancreas may be detected were determined, and the images with 512 × 512 pixels dimensions were cropped. In this way, areas where there is no pancreas were not scanned, and the size of the images in the entire dataset was reduced to 256 × 256

pixels. So, areas where there is no pancreas are not affected by segmentation results and this is to enable faster training and processing of more images at the same time during training.

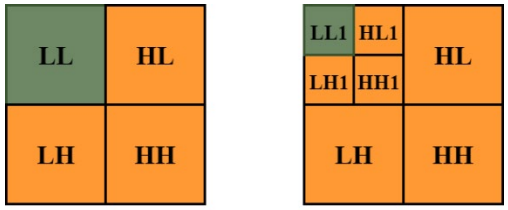


(a) (b)

Figure 3. A sample of ROI process: (a) Original image (512 × 512 pixels), (b) Cropped ROI image (256 × 256 pixels)

2.2.2 Feature Extraction

After the determination of ROI, by applying a two-level 2D Discrete Wavelet Transform to the 256 × 256 pixels images, the dimensions of the images were converted to 64 × 64 pixels by preserving the low-frequency information. The basic idea of Discrete Wavelet Transform (DWT) is to provide time-frequency representation. 2D DWT represents an image in terms of a set of shifted and expanded wavelet functions and scaling functions [4]. When one level 2D DWT is applied to an image, LL subsampling of the image is half the size of the raw image. When two-level 2D DWT is applied, subsamples of this LL subsampling are obtained as LL1, LH1, HL1, and HH1. One-level and two-level 2D DWT analyses of an image can be seen in Figure 4 (a) and, (b).



(a) (b)

Figure 4. 2D Discrete wavelet transform analyses of an image: (a) One level (b) Two level

The dimensions of the ROI images were converted to 64 x 64 pixels by applying two-level 2D DWT with Daubechies (db2) wavelet preserving the low frequency (LL) information. Figure 5 (a) and (b) show the one-level and two-level 2D DWT results of an ROI image. Figure 5 (c) and (d) give the one-level and two-level 2D DWT results of the mask of this ROI image.

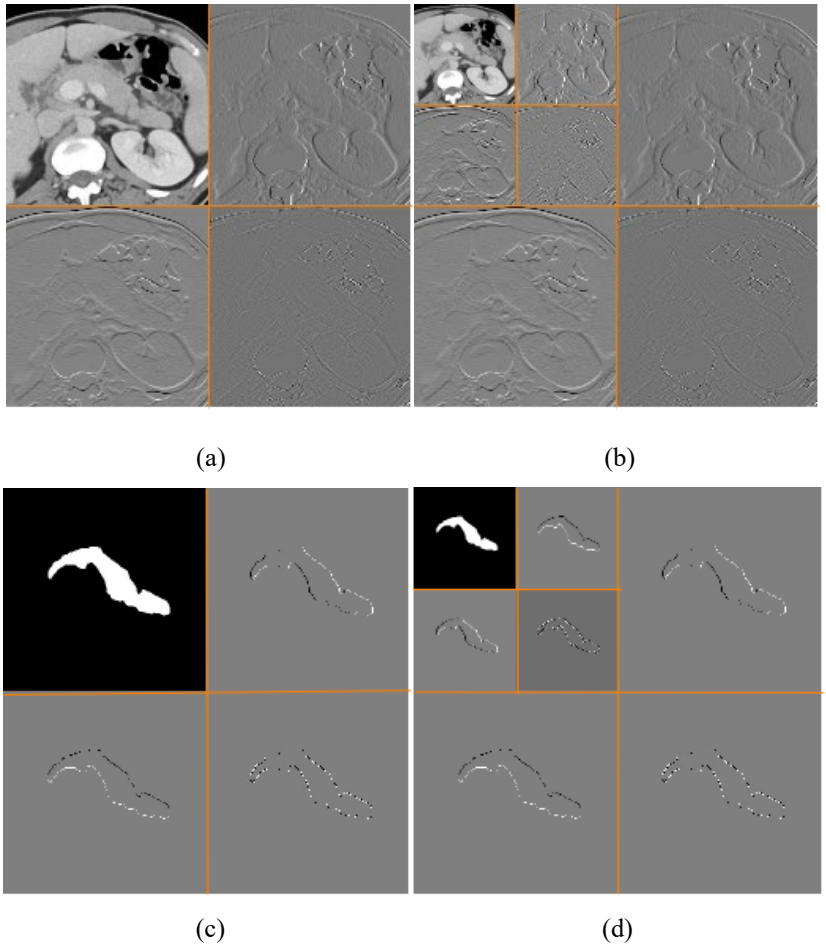


Figure 5. An ROI image (a) One-level 2D DWT result of the image (b) Two-level 2D DWT result of the image (c) One-level 2D DWT result of the mask (d) Two-level 2D DWT result of the mask

2.3 Pascal U-Net

Studies in the biomedical field have gained momentum with the development of deep learning models. In particular, the development of CNN-based structures has gained great importance in this area. Ronneberger et al. [11] achieved great success in the segmentation competition with the U-Net model which was developed in 2015 and stated that they obtained a model that could come up with a solution to the field of biomedical image processing.

The U-Net model basically consists of two parts, the encoder and the decoder. The encoder consists of three layers: The convolutional layer, Batch normalization layers, and Pooling layer. The decoder includes convolutional layers and deconvolution layers. The U-Net encoder is downsampled in a total of 4 times. Symmetrically, the decoder also performs corresponding upsampling 4 times to restore the feature map obtained by the encoder to a picture of the same size as the original picture. When the right-side decoder implements upsampling, the feature maps on the encoder side of the corresponding level are concatenated by matrix cascading. This operation is beneficial to supplement the missing pixel position information during the convolution process, thereby improving the segmentation accuracy [19].

In this study, a novel model called Pascal U-Net was developed, inspired by the U-Net model and Pascal's triangle. The U-Net++ model used by Zhou et al. [22] in their study gave an idea for the development of the Pascal U-Net model. The most important feature that makes the Pascal U-Net model different from U-Net is that a connection is created by adding convolution layers instead of just a direct connection between the encoder and decoder parts. These convolution layers between the two sections consist of convolution blocks created by the number sequence in Pascal's triangle. The structure of the proposed Pascal U-Net model is shown in Figure 6.

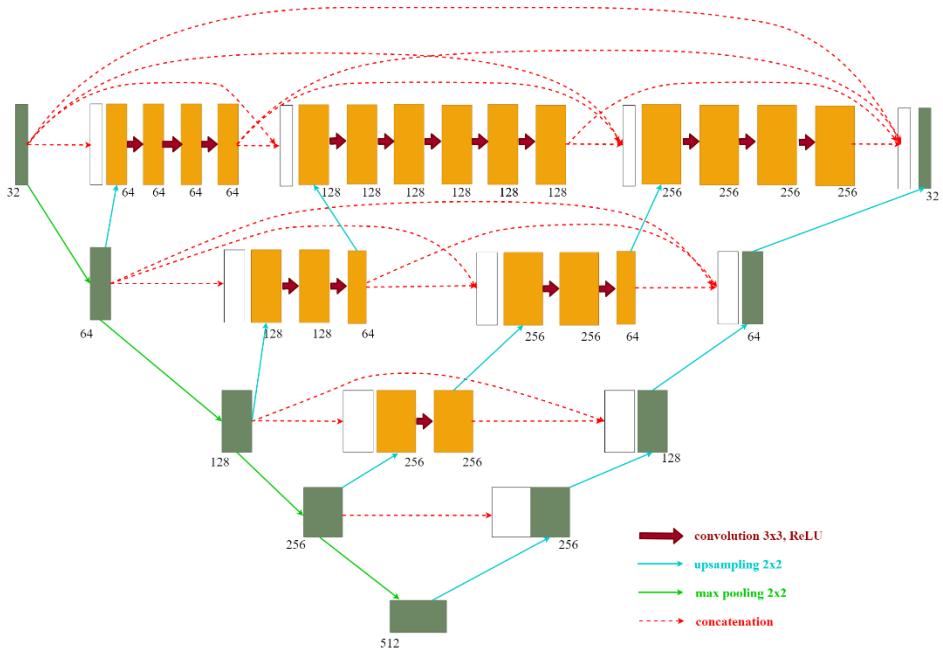


Figure 6. Pascal U-Net model

As seen in Figure 6, all layers of the Pascal U-Net model consist of convolution blocks. The image given to the input of the network is divided into 3 points: 1) Once the convolution is applied, the obtained feature maps are transferred to the convolution blocks in the internal structure of the model. 2) It is transferred directly to the decoder part. 3) Feature maps are extracted with the pooling layer.

With the pooling layer, the feature map is transferred to 3 different layers as in the input image. In addition, the feature map is transferred to the internal structure of the model by oversampling and the feature map is combined with the extracted image.

Images from the encoder and feature maps are combined in the layers in the internal structure of the model and convolution is applied to these images as much as the numbers in Pascal's triangle. Blocks that contain as many convolutions as the numbers in Pascal's triangle are called Pascal blocks. Images taken from the previous Pascal block are transferred to the forward Pascal block in the internal structure.

In the decoder part, as in the known U-Net, the images are combined with the images taken from the encoder part after upsampling. In addition to the U-Net, the images obtained from Pascal blocks in the Pascal U-Net are also transferred to the decoder part. While the number of parameters in the traditional U-Net model is 7 million, it is 10 million in our proposed Pascal U-Net model. Due to the high number of convolution blocks and parameters used in the Pascal U-Net model compared to the U-Net model, training times can be longer in Pascal U-Net.

3 Results

In this study, pancreas segmentation was performed with both the U-Net model and the Pascal U-Net model proposed in the study on abdomen CT images of the abdomen taken from TCIA and SUMFH, and the results were presented comparatively.

In both models, the batch sizes were changed from 1 to 10, respectively, and the results were obtained from 10 repetitions of training and testing for each batch size. The results were recorded by averaging the results repeated ten times for each batch value. In the study, 2-, 4- and 6-fold cross-validation method was used. All runs were performed with 500 iterations, $2x10^{-4}$ learning rate, and Adam optimizer, and no data augmentation was used.

3.1 Performance Evaluation Metrics

Dice similarity coefficient (DSC) and Jaccard metrics were used as performance evaluation metrics for the segmentation results obtained. These performance evaluation metrics are metrics formulated with values taken from a confusion matrix. The confusion matrix, also called the error matrix, is the matrix used to evaluate the classification performance [16]. In a confusion matrix, TP represents true positive, TN true negative, FP false

positive, and FN false negative. In this study, TP is the number of pixels labeled as the pancreas and determined as pancreas by the model, TN is the number of pixels labeled as background and determined as background by the model, FN is the number of pixels labeled as the pancreas and determined as background by the model and FP is the number of pixels labeled as background and determined as pancreas by the model as a result of segmentation. The DSC and Jaccard metrics are expressed in Equation 1-2, respectively.

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

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

In addition to DSC and Jaccard, Hausdorff Distance metrics were used in this paper. As illustrated in Figure 7, the Hausdorff Distance is a measure of the spatial distance between two sets of points [2]. The Hausdorff Distance (HD) is defined in Equation 3.

$$HD(P_1, P_2) = max(hd(P_1, P_2), hd(P_2, P_1)), \tag{3}$$

where,

$$hd(P_1, P_2) = \underset{x \in P_1}{max}\ \underset{y \in P_2}{min}\ ||x - y||. \tag{4}$$

The function $hd(P_1,P_2)$ is referred to as the directed Hausdorff Distance from P_1 to P_2. It ranks each point of P_1 according to its distance to the nearest point of P_2.

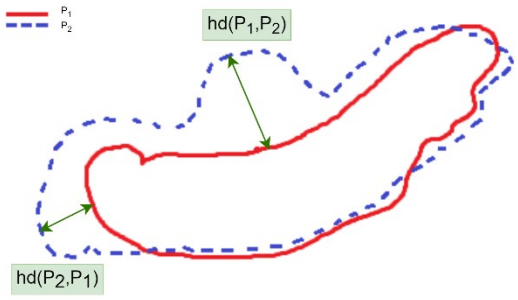


Figure 7. A schematic showing the Hausdorff distance between two sets of points P_1 and P_2

In addition to these evaluation metrics, Bland-Altman analysis was used in this paper as statistical analysis. In 1983 Altman and Bland proposed an analysis in medicine [1], based on the quantification of the agreement between two quantitative measurements by considering the mean difference and constructing limits of agreement. The Bland-Altman plot analysis is a simple way to assess a bias between the mean differences and to estimate an agreement interval of the differences between the two different methods [5]. In this study, the superiority of the models over each other is evaluated according to Bland-Altman analysis in terms of DSC scores.

3.2 Pancreas Segmentation Results

In this study, pancreas segmentation was performed on two different datasets consisting of abdominal CT images. For the segmentation of the pancreas, results were obtained with both the U-Net model and the Pascal U-Net model proposed in this study. The results obtained with different fold cross-validation methods in different batch sizes are presented in detail in this section.

3.2.1 Segmentation Results Obtained on the TCIA Dataset

In the study, firstly, pancreas segmentation was performed from abdominal CT images in the TCIA dataset, which is a dataset frequently used in the literature. The ROI was obtained from abdomen CT images and size reduction was performed with a 2-level 2D discrete wavelet transform. Pancreas segmentation was performed by the U-Net and Pascal U-Net models. Segmentation results were evaluated using 2 different metrics. The results of the U-Net and Pascal U-Net models are presented in Table 1 in detail for both deep learning models. As can be seen from the

tables, the performance of the deep learning models implemented was evaluated using DSC, Jaccard, and Hausdorff Distance metrics. In addition, the average running time of the algorithm (sum of training and testing time) is presented when cross-validation is used for each case in these tables.

Table 1: Pancreas segmentation results obtained by the U-Net and Pascal U-Net models on the TCIA dataset for K-Fold Cross-Validation (CV)

		U-Net				Pascal U-Net			
K Fold CV	Batch Size	DSC (%)	Jaccard (%)	Hausdorff Distance (mm)	Time (h.m.s)	DSC (%)	Jaccard (%)	Hausdorff Distance (mm)	Time (h.m.s)
K=2	1	**65.20**	**48.69**	2.3163	0.17.31	30.00	17.85	4.1864	0.43.59
	2	64.37	47.59	2.2512	0.09.02	65.35	48.19	2.8759	0.30.36
	3	58.89	41.96	2.605	0.05.46	**68.74**	**52.31**	2.3956	0.31.01
	4	56.81	40.11	2.5569	0.05.21	67.92	51.41	2.5983	0.30.42
	5	55.81	39.31	**2.2437**	0.06.49	67.63	51.33	**2.332**	0.29.57
	6	52.45	36.02	2.6316	**0.04.02**	65.47	48.64	2.4486	0.26.43
	7	52.14	35.61	2.5017	0.04.07	64.13	47.12	2.9037	0.25.41
	8	54.89	37.81	2.6485	0.04.30	63.97	47.26	2.5834	0.25.18
	9	50.54	34.13	2.8964	0.04.12	63.97	46.74	2.6177	0.25.39
	10	53.63	37.00	2.3424	0.04.28	62.44	45.69	2.5058	**0.24.39**
K=4	1	**68.36**	**52.12**	**2.0263**	0.29.53	29.25	17.85	3.8626	1.58.19
	2	65.30	48.82	2.2166	0.18.13	67.43	51.04	2.5766	1.16.57
	3	62.47	45.97	2.4653	0.13.22	69.73	53.93	2.1941	1.18.12
	4	61.07	44.51	2.2769	0.11.46	69.80	53.78	2.0984	1.14.33
	5	60.17	43.50	2.4338	0.10.29	**70.27**	**54.31**	2.1335	1.11.40
	6	58.34	41.73	2.4785	0.09.48	68.96	53.08	2.2075	1.11.07
	7	56.65	40.11	2.435	0.09.09	68.53	52.46	2.1594	1.10.23
	8	55.90	39.43	2.5094	0.09.05	68.29	52.03	**2.0168**	1.08.26
	9	55.00	38.64	2.5008	0.08.35	66.76	50.35	2.1404	1.08.31
	10	57.24	40.69	2.1285	**0.08.29**	67.12	50.89	2.2538	**1.06.06**
K=6	1	**70.24**	**53.99**	**1.9459**	0.49.10	29.79	18.99	3.7938	3.13.33
	2	65.73	49.21	2.2106	0.31.47	68.39	52.40	2.391	2.03.37
	3	63.92	47.35	2.2409	0.22.33	70.73	55.12	**1.8976**	2.08.15
	4	62.31	45.65	2.4507	0.19.14	70.36	54.55	2.1056	2.02.10
	5	60.70	44.24	2.3295	0.17.05	**71.35**	**55.73**	2.0251	1.58.49
	6	59.92	43.26	2.2926	0.16.10	70.66	54.96	2.0001	1.55.37
	7	57.71	41.13	2.4704	0.15.35	69.80	53.96	1.9922	1.54.39
	8	57.65	40.96	2.2691	0.14.42	69.49	53.51	2.2283	1.52.55
	9	57.09	40.40	2.399	0.14.07	69.22	53.30	2.0492	1.51.32
	10	55.52	39.12	2.4322	**0.13.31**	68.48	52.46	2.0562	**1.49.30**

When the results for 2-, 4- and 6-fold cross-validation in Table 1 are analyzed for each metric and batch size; it is seen that the best results of DSC, Jaccard, and Hausdorff Distance for U-Net were 70.24%, 53.99%, and 1.9459 mm when the batch size is 1 and cross-validation fold number is 6. In addition to this, the best results of DSC and Jaccard for Pascal U-Net were found at 71.35% and 55.73% when batch size is 5 and cross-validation fold number is 6. The best Hausdorff Distance result for Pascal U-Net is obtained as 1.8976 when batch size is 3 and cross-validation fold number is 6. Average running times according to Table 1; the shortest time is found when batch size is 6 in 2-fold cross-validation for U-Net and batch size is 10 in 2-fold cross validation for Pascal U-Net. The longest time is found when batch size is 1 in 6-fold cross validation for both models.

In the performed study, a visual comparison is presented in Figure 8 for the results with the highest values of DSC obtained using the U-Net and Pascal U-Net models on the TCIA dataset for 2-, 4- and 6-fold cross-validation.

As shown in Figure 8, the proposed method, Pascal U-Net, represents a high overlap in segmentation between automatically found results and manually labeled boundaries.

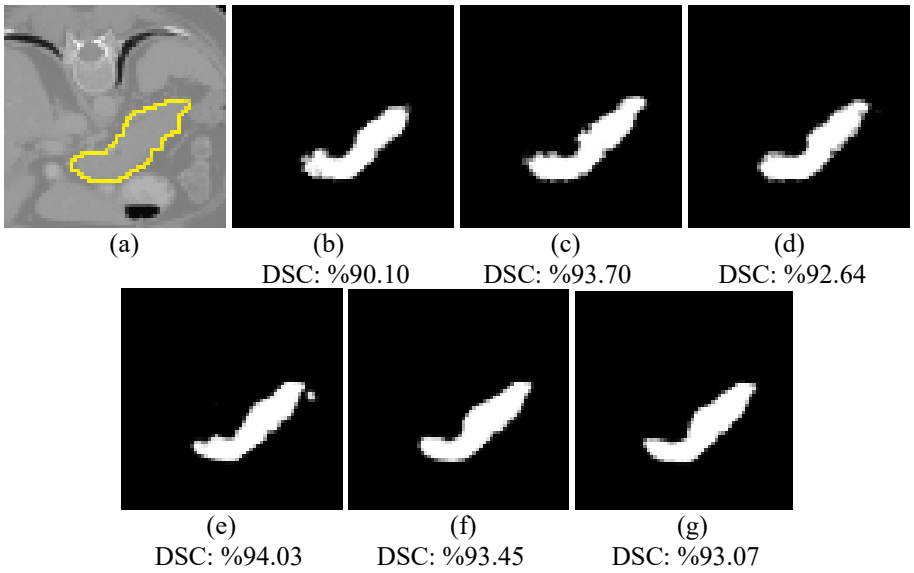


Figure 8. Visual comparison for segmentation results with the highest DSC value (a) Target image, (b) 2-fold CV result for U-Net, (c) 4-fold CV result for U-Net, (d) 6-fold CV result for U-Net, (e) 2-fold CV result for Pascal U-Net, (f) 4-fold CV result for Pascal U-Net, (g) 6-fold CV result for Pascal U-Net

3.2.2 Segmentation Results Obtained on the SUMFH Dataset

In this study, pancreas segmentation was performed on the dataset created by abdominal CT images taken from SUMFH. In this dataset, first of all, the ROI was obtained from CT images of the abdomen, and size reduction was performed with 2-level 2D DWT. Pancreas segmentation was first performed by the U-Net model on the dataset consisting of images obtained after the ROI acquisition and size reduction processes. Segmentation results were evaluated according to 3 metrics (DSC, Jaccard, and Hausdorff Distance). Then, the same segmentation process on the same dataset was performed by the Pascal U-Net model proposed in the study, and the results are presented in Table 2 in detail for both deep learning models. Segmentation results are presented in Table 2 for 2-, 4- and 6-fold cross-validation.

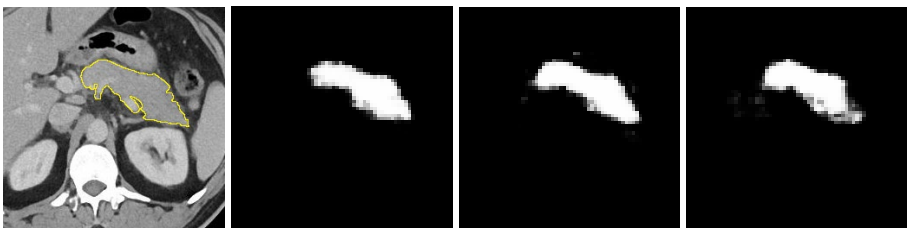
Table 2: Pancreas segmentation results obtained by the U-Net and Pascal U-Net models on the SUMFH dataset for K-Fold Cross-Validation (CV)

		U-Net				Pascal U-Net			
K Fold CV	Batch Size	DSC (%)	Jaccard (%)	Hausdorff Distance (mm)	Time (h.m.s)	DSC (%)	Jaccard (%)	Hausdorff Distance (mm)	Time (h.m.s)
K=2	1	**63.51**	**47.14**	**2.565**	0.08.09	40.58	34.24	5.361	0.32.01
	2	60.35	44.26	2.8045	0.06.38	61.47	45.32	3.2532	0.23.26
	3	58.51	41.75	2.816	0.04.05	64.53	**48.18**	3.2173	0.20.39
	4	54.23	38.22	3.1103	0.03.36	**64.67**	47.78	3.2045	**0.18.23**
	5	53.43	37.46	3.1479	0.03.27	63.74	47.19	**2.625**	0.19.03
	6	53.12	37.11	3.2153	0.03.25	63.83	46.73	3.0418	0.18.40
	7	52.53	36.71	2.7355	0.03.36	62.56	45.03	3.1449	0.18.25
	8	51.92	36.22	3.2968	0.03.27	62.82	45.34	3.1573	0.18.34
	9	54.08	38.20	2.9613	0.03.30	61.76	44.33	3.1759	0.18.26
	10	51.27	35.20	3.2379	**0.03.21**	60.49	43.48	3.2	0.18.42
K=4	1	**66.96**	**50.89**	2.5263	0.23.07	46.66	37.48	4.9295	1.24.12

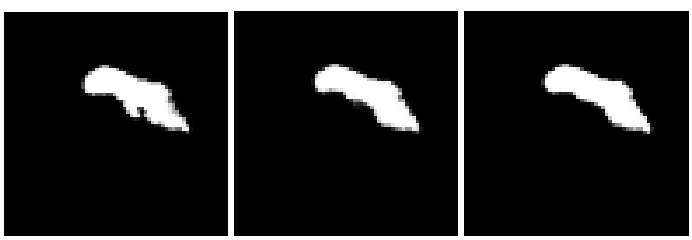
	2	65.59	49.62	2.4967	0.15.23	64.62	48.96	2.7526	0.58.56
	3	62.15	46.24	2.671	0.10.07	66.52	50.56	**2.4227**	0.56.29
	4	60.40	44.38	2.8439	0.09.08	67.75	51.27	2.6712	0.54.38
	5	59.49	43.59	2.9303	0.08.32	**68.23**	**52.17**	3.3013	0.52.18
	6	57.37	41.75	2.7772	0.08.09	66.74	49.85	2.7911	0.50.34
	7	56.69	40.88	2.9819	0.08.09	67.72	50.91	2.4564	0.50.15
	8	57.84	41.82	2.9646	0.07.50	67.19	50.48	2.4671	0.50.17
	9	55.71	39.97	3.0524	**0.07.44**	66.15	49.08	3.5471	0.49.29
	10	57.02	40.88	**2.4106**	0.07.49	65.05	48.40	2.6547	**0.48.23**
K=6	1	**68.05**	**52.01**	2.4814	0.35.20	49.41	40.32	4.7017	2.14.34
	2	66.22	50.08	2.5267	0.24.20	65.60	50.19	2.8202	1.34.21
	3	64.28	48.36	2.6497	0.16.39	66.94	51.59	2.5349	1.33.41
	4	62.49	46.62	**2.3473**	0.15.01	68.64	**52.90**	2.4498	1.29.58
	5	60.97	45.12	2.7388	0.13.37	68.36	52.22	2.609	1.26.39
	6	59.25	43.45	2.7723	0.13.00	**68.92**	52.50	2.4755	1.24.02
	7	57.48	41.89	3.0049	**0.12.12**	68.34	51.74	**2.3938**	1.23.32
	8	57.48	41.45	2.9712	0.12.32	67.23	50.83	3.2494	1.22.35
	9	58.66	42.96	2.7662	0.12.39	67.60	51.00	2.6482	1.21.39
	10	55.98	40.21	2.7695	0.12.17	66.93	49.88	2.7788	**1.20.07**

In Table 2, the best values obtained in the pancreas segmentation results performed by U-Net and Pascal U-Net models for different batch sizes and cross-validation fold numbers on the SUMFH dataset are indicated in bold. When the results for 2-, 4- and 6-fold cross-validation in Table 2 are analyzed for each metric and batch size; it is seen that the best results of DSC and Jaccard for U-Net are 68.05%, 52.01%, respectively when the cross-validation fold number is 4. Besides, the best Hausdorff Distance for U-Net is 2.3473 mm when the cross-validation number is 6. For the Pascal U-Net, the best result of DSC is 68.92% when 6-fold cross-validation and batch size is 6. The best result of Jaccard for Pascal U-Net is 52.90% when the cross-validation fold number is 6 and batch size is 4 and the best result of Hausdorff Distance is 2.3938 mm when 6-fold cross-validation and batch size is 7. Considering average running times for U-Net according to Table 2; the shortest time is found when batch size is 10 in 2-fold cross-validation as 3 minutes and the longest time is found when batch size is 1 in 6-fold cross-validation as 35 minutes. On the other hand, the shortest time for Pascal U-Net is 18 minutes when batch size is 4 in 2-fold cross-validation, and the longest time is 2 hours and 14 minutes when batch size is 1 in 6-fold cross-validation.

The visual comparison for the results with the highest values of DSC obtained by U-Net and Pascal U-Net models and 2-, 4- and 6-fold cross validation on the SUMFH dataset is shown in Figure 9. As shown in Figure 9; Pascal U-Net has significant improvement in segmentation according to U-Net for all visual results.



(a) (b) DSC: %90.09 (c) DSC: %89.22 (d) DSC: %88.80



(e) (f) (g)

DSC: %92.94 DSC: %93.51 DSC: %92.96

Figure 9. Visual comparison for segmentation results with the highest DSC value (a) Target image, (b) 2-fold CV result for U-Net, (c) 4-fold CV result for U-Net, (d) 6-fold CV result for U-Net, (e) 2-fold CV result for Pascal U-Net, (f) 4-fold CV result for Pascal U-Net, (g) 6-fold CV result for Pascal U-Net

4 DISCUSSION

In this paper, a new deep-learning model has been proposed for pancreas segmentation, which is an important problem in biomedical image segmentation. The performance of the proposed deep learning model was evaluated on two datasets including Pancreas CT images taken from the TCIA database and abdomen CT images obtained from the SUMFH Radiology Department. In both datasets, a slice in which the pancreas can be seen was selected for each patient. In these selected images, in both datasets, the areas where the pancreas is not present in the CT image are cropped and the ROI is obtained. During this process, the sizes of the images are reduced in half. Then, by applying two-level 2D DWT to the images whose ROI was obtained, both the features of the images were preserved and their dimensions were reduced. Thanks to these preprocessing methods, the raw images, which are 512 x 512 pixels in both datasets, are reduced to 64 x 64 pixels by preserving their properties to a great extent. In this way, the TCIA dataset was created to contain 82 images, and the SUMFH dataset to contain 56 images.

The segmentation results performed on both datasets by the proposed Pascal U-Net model, which is inspired by both the U-Net model and the Pascal triangle are analyzed in detail. In pancreas segmentation carried out in this paper with deep learning models; 2-, 4- and 6-fold cross-validation methods were used. The batch sizes were taken from 1 to 10 and each run was repeated 10 times. The averages of the results obtained from these 10 repetitions are presented. The calculation of all results was obtained using 3 basic performance evaluation metrics (DSC, Jaccard, and Hausdorff Distance).

The analysis of the optimal pancreas segmentation results obtained for the U-Net and Pascal U-Net models on the TCIA and SUMFH datasets is given in Table 3.

Table 3: Optimum DSC results obtained with both models on both datasets.

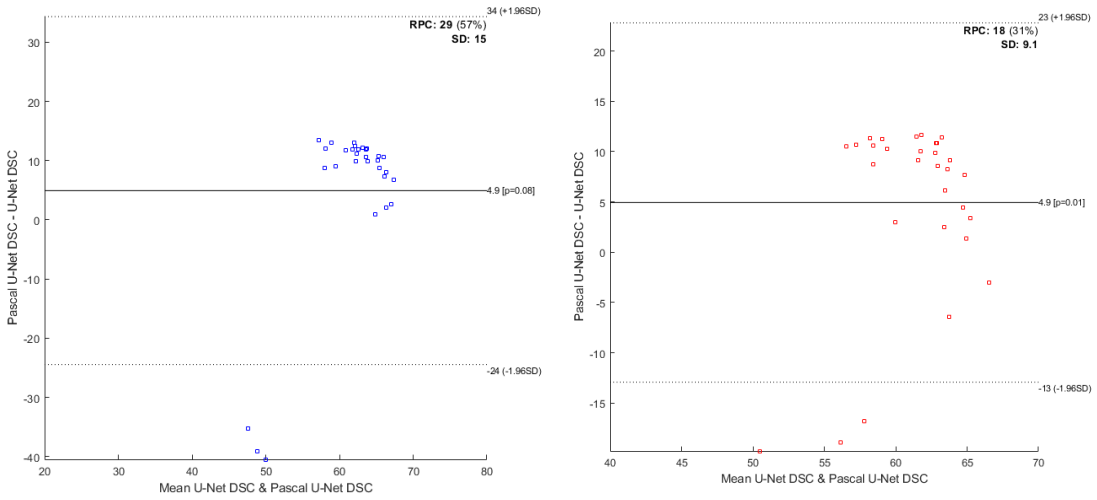
Dataset	Model	Cross-Validation	DSC (%)	Jaccard (%)	Hausdorff Distance (mm)	Batch Size	Time (h.m.s)
TCIA	U-Net	2-fold	65.20	48.69	2.3163	1	0.17.31
		4-fold	68.36	52.12	2.0263	1	0.29.53
		6-fold	70.24	53.99	1.9459	1	0.49.10
	Pascal U-Net	2-fold	68.74	52.31	2.3956	3	0.31.01
		4-fold	70.27	54.31	2.1335	5	1.11.40
		6-fold	71.35	55.73	2.0251	5	1.58.49
SUMFH	U-Net	2-fold	63.51	47.14	2.565	1	0.08.09
		4-fold	66.96	50.89	2.5263	1	0.23.07
		6-fold	68.05	52.01	2.4814	1	0.35.20
	Pascal U-Net	2-fold	64.67	47.78	3.2045	4	0.18.23
		4-fold	68.23	52.17	3.3013	5	0.52.18
		6-fold	68.92	52.50	2.4755	6	1.24.02

As shown in Table 3, on TCIA dataset, the best DSC results are found as 70.24% in 49 minutes with U-Net and 71.35% in 1 hour and 58 minutes for Pascal U-Net. Considering the time, even though U-Net has better performance than Pascal U-Net, Pascal U-Net provides better overlap as 1.11%.

When the results on the SUMFH dataset are examined in Table 3, like in the results of the TCIA dataset, it can be seen that the best DSC is obtained on 6-fold cross-validation. The DSC difference between the results of U-Net and Pascal U-Net is 0.87%. But, considering the visual results presented in Fig. 9, it can be said that Pascal U-Net has a higher overlap than U-Net.

In addition to this, according to Table 3, the Hausdorff Distance for U-Net is better than Pascal U-Net on the TCIA dataset. Besides, the Hausdorff Distance on the SUMFH dataset is 2.4755 mm for Pascal U-Net, while the result is 2.4814 mm for U-Net.

Furthermore, the Bland-Altman analysis is used to determine the superiority of the U-Net and Pascal U-Net models over each other. Since the mean of the differences should be spread around zero according to the Bland-Altman analysis, it would be correct to examine the agreement of these two methods with the Bland-Altman analysis. The Bland-Altman plot of DSC results on the U-Net and Pascal U-Net models for TCIA and SUMFH datasets is shown in Figure 10 (a) and (b), respectively. In Figure 10, the blue squares are the DSC differences of the U-Net and Pascal U-Net methods used for the TCIA dataset, while the red squares are the DSC differences of the two methods used for the SUMFH dataset. It can be summarized the lack of agreement by calculating the bias, estimated by the mean difference and the standard deviation of the differences. As can be seen in the figure, DSC spreads for the TCIA dataset gave better results than the SUMFH dataset. While the reproducibility coefficient is 29 for the TCIA dataset, it is 18 for the SUMFH dataset. When the differences found are normally distributed, it is expected that the differences will be randomly distributed around zero and 95% of them will be between the limits of agreement. In our problem, as seen in Figure 10, 3 out of 30 difference points were observed to be out of limits for both datasets. According to the results of the analysis, it can be said that there is an agreement between the two models, and they are alternatives to each other.



(a) (b)

Figure 10. Bland-Altman Plot of DSC on both models for both datasets (a) TCIA dataset (b) SUMFH dataset.

The best pancreas segmentation results obtained in this paper are compared with the studies in which pancreas segmentation was performed with deep learning models in Table 4.

Table 4: A comparison of the pancreas segmentation results obtained in this paper with the literature.

Author	Method	Number of Data	DSC	Jaccard	Hausdorff Distance
Roth et al. [12]	Superpixel + CNN	82	0.68	-	-
Farag et al. [3]	AlexNet + Superpixel	80	0.71	0.58	-
Zhou et al. [30]	CNN	240	-	0.45	-
Roth et al. [14]	Holistically-nested CNN	82	0.81	0.69	-
Roth et al. [15]	Fully Convolutional Network	150	0.82	-	-
Gibson et al. [6]	Dense V-Net	90	0.78	-	-
Yang et al. [18]	Cascade Neural Network	82	**0.88**	-	-

Man et al. [10]	Deep Q Network	82	0.87	-	-
Lu et al. [9]	Ringed Residual U-Net	82	**0.88**	-	-
Liu et al. [8]	Superpixel + 5 Different U-Net models	82	0.84	0.73	-
Wang et al. [17]	Dual-input v-mesh FCN	82	0.87	-	18.41
Li et al. [7]	MAD-UNet	82	0.86	**0.76**	4.40
This paper	Pascal U-Net	56	0.69	0.53	2.48
This paper	Pascal U-Net	82	0.71	0.56	**2.03**

The comparison of the best segmentation results obtained in this paper with the literature in terms of 2 different metrics is presented in Table 4 and the best values are indicated in bold. According to Table 4; in terms of DSC metric, it is seen that the best result was obtained by Lu et al. [9] with 88%. In terms of Jaccard, were obtained by Li et al. [7], with 76%. Considering the Hausdorff Distance, the best result is 2.03 mm in this paper. The main reason why the values obtained for DSC and Jaccard metrics in pancreas segmentation studies performed with other deep learning methods in the literature are higher than this paper is due to the processing limit of the graphics card in the workstation used.

5 CONCLUSION

This paper proposes an innovative deep-learning model called Pascal U-Net. This model is improved by inspiring the traditional U-Net and Pascal triangle. It was observed that the Pascal U-Net model shows high sensitivity and specificity and high overlap in segmentation than the traditional U-Net model in both datasets used in this study. Two different datasets of abdominal CT images were used. The first of these is the publicly available Pancreas CT dataset taken from the TCIA database. In this dataset, there are 82 CT images. The best DSC results found using the U-Net model in the TCIA dataset are 70.24% in 49 minutes when the batch size is 1 in 6-fold cross-validation. The best DSC results obtained using the Pascal U-Net model on the TCIA dataset are 71.35% in 1 hour 58 minutes when batch size is 5 in 6-fold cross-validation. The other dataset consists of abdominal CT images taken from the Radiology Department of Selcuk University Medical Faculty Hospital. There are 58 CT images in this dataset. The best DSC results obtained by using the U-Net model on the SUMFH dataset are 68.05% in 35 minutes with 6-fold cross-validation, when the batch size is 1. It is seen that the best DSC results obtained on the SUMFH dataset using the Pascal U-Net model is 68.92% in 1 hour and 24 minutes when the batch size is 6 in 6-fold cross-validation.

When compared with the literature, it was observed that the results obtained were lower than the literature in terms of DSC and Jaccard performance metrics. In the proposed Pascal U-Net model, if dense blocks are used instead of convolution blocks between the encoder and decoder sections, the number of parameters of the model increases. Thus, it can give better results in a longer time. In this study, pancreas segmentation was performed by taking 2,4 and 6-fold cross validation methods and images in batch size values from 1 to 10. In the future, results can be obtained with more multiple cross validation methods and Pascal U-Net model with larger batch sizes. Since the graphics card capacity of the used workstation is limited, the number of images in the datasets used in this paper has been adjusted to be a slice for each patient and dense blocks could not be used. Results can be obtained with more data numbers that can be obtained using data augmentation methods or Generative Adversarial Networks (GANs) on workstations with better graphics cards.

Acknowledgements

We are thankful to Selcuk University Faculty of Medicine for their assistance in acquiring the CT images used in this study.

References

[1] Altman, D. G., & Bland, J. M. (1983). Measurement in Medicine: The Analysis of Method Comparison Studies. Journal of the Royal Statistical Society. Series D (The Statistician), 32(3), 307–317. doi: 10.2307/2987937

[2] Beauchemin, M., Thomson, K. P., & Edwards, G. (1998). On the Hausdorff distance used for the evaluation of segmentation results. Canadian journal of remote sensing, 24(1), 3-8. doi: 10.1080/07038992.1998.10874685

[3] Farag, A., Lu, L., Roth, H. R., Liu, J., Turkbey, E., & Summers, R. M. (2016). A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. IEEE Transactions on Image Processing, 26(1), 386-399. doi: 10.1109/TIP.2016.2624198

[4] Ghazali, K. H., Mansor, M. F., Mustafa, M. M., & Hussain, A. (2007, December). Feature extraction technique using discrete wavelet transform for image classification. In 2007 5th Student Conference on Research and Development (pp. 1-4). IEEE. doi: 10.1109/SCORED.2007.4451366

[5] Giavarina, D. (2015). Understanding bland altman analysis. Biochemia medica, 25(2), 141-151. doi: 10.11613/BM.2015.015

[6] Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., ... & Barratt, D. C. (2018). Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE transactions on medical imaging, 37(8), 1822-1834. doi: 10.1109/TMI.2018.2806309

[7] Li, W., Qin, S., Li, F. and Wang, L. (2021), MAD-UNet: A deep U-shaped network combined with an attention mechanism for pancreas segmentation in CT images. Med. Phys., 48: 329-341. doi: 10.1002/mp.14617

[8] Liu, S., Yuan, X., Hu, R., Liang, S., Feng, S., Ai, Y., & Zhang, Y. (2019). Automatic pancreas segmentation via coarse location and ensemble learning. IEEE Access, 8, 2906-2914. doi: 10.1109/ACCESS.2019.2961125

[9] Lu, L., Jian, L., Luo, J., & Xiao, B. (2019). Pancreatic segmentation via ringed residual U-Net. IEEE Access, 7, 172871-172878. doi: 10.1109/ACCESS.2019.2956550

[10] Man, Y., Huang, Y., Feng, J., Li, X., & Wu, F. (2019). Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net. IEEE transactions on medical imaging, 38(8), 1971-1980. doi: 10.1109/TMI.2019.2911588

[11] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham. doi: 10.1007/978-3-319-24574-4_28

[12] Roth, H. R., Farag, A., Lu, L., Turkbey, E. B., & Summers, R. M. (2015, March). Deep convolutional networks for pancreas segmentation in CT imaging. In Medical Imaging 2015: Image Processing (Vol. 9413, pp. 378-385). SPIE. doi: 10.1117/12.2081420

[13] Roth, H. R., Farag, A., Turkbey, E. B., Lu, L., Liu, J., & Summers, R. M. (2016). Data From Pancreas-CT. The Cancer Imaging Archive. doi: 10.7937/K9/TCIA.2016.tNB1kqBU

[14] Roth, H. R., Lu, L., Lay, N., Harrison, A. P., Farag, A., Sohn, A., & Summers, R. M. (2018). Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. Medical image analysis, 45, 94-107. doi: 10.1016/j.media.2018.01.006

[15] Roth, H. R., Oda, H., Zhou, X., Shimizu, N., Yang, Y., Hayashi, Y., Oda, M., Fujiwara, M., Misawa, K., & Mori, K. (2018). An application of cascaded 3D fully convolutional networks for medical image segmentation. Computerized Medical Imaging and Graphics, 66, 90-99. doi: 10.1016/j.compmedimag.2018.03.001

[16] Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment, 62(1), 77-89. doi: 10.1016/S0034-4257(97)00083-7

[17] Wang, Y., Gong, G., Kong, D., Li, Q., Dai, J., Zhang, H., ... & Xue, J. (2021). Pancreas segmentation using a dual-input V-mesh network. Medical Image Analysis, 69, 101958. doi: 10.1016/j.media.2021.101958

[18] Yang, Z., Zhang, L., Zhang, M., Feng, J., Wu, Z., Ren, F., & Lv, Y. (2019, July). Pancreas segmentation in abdominal CT scans using inter-/intra-slice contextual information with a cascade neural network. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 5937-5940). IEEE. doi: 10.1109/EMBC.2019.8856774

[19] Zhang, J., Du, J., Liu, H., Hou, X., Zhao, Y., & Ding, M. (2019). LU-NET: An improved U-Net for ventricular segmentation. IEEE Access, 7, 92539-92546. doi: 10.1109/ACCESS.2019.2925060

[20] Zhou, X., Ito, T., Takayama, R., Wang, S., Hara, T., & Fujita, H. (2016). Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting. In Deep Learning and Data Labeling for Medical Applications (pp. 111-120). Springer, Cham. doi: 10.1007/978-3-319-46976-8_12

[21] Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E. K., & Yuille, A. L. (2017, September). A fixed-point model for pancreas segmentation in abdominal CT scans. In International conference on medical image computing and computer-assisted intervention (pp. 693-701). Springer, Cham. doi: 10.1007/978-3-319-66182-7_79

[22] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham. doi: 10.1007/978-3-030-00889-5_1