

CHAPTER 5

Multi Organ Segmentation in Medical images

Beyza KAYHAN, Sait Ali UYMAZ

Multi Organ Segmentation in Medical images

Beyza KAYHAN

Konya Technical University

Sait Ali UYMAZ

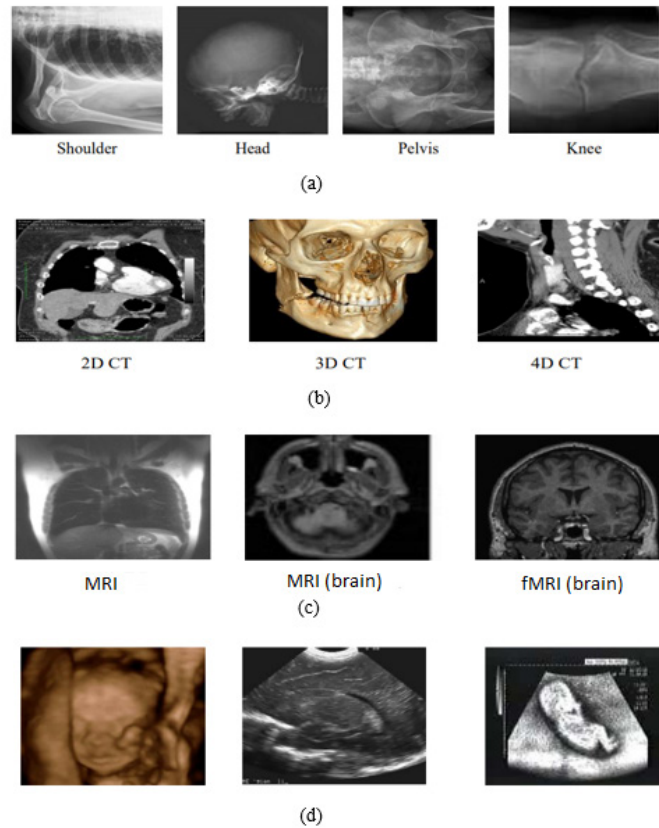
Konya Technical University

Introduction

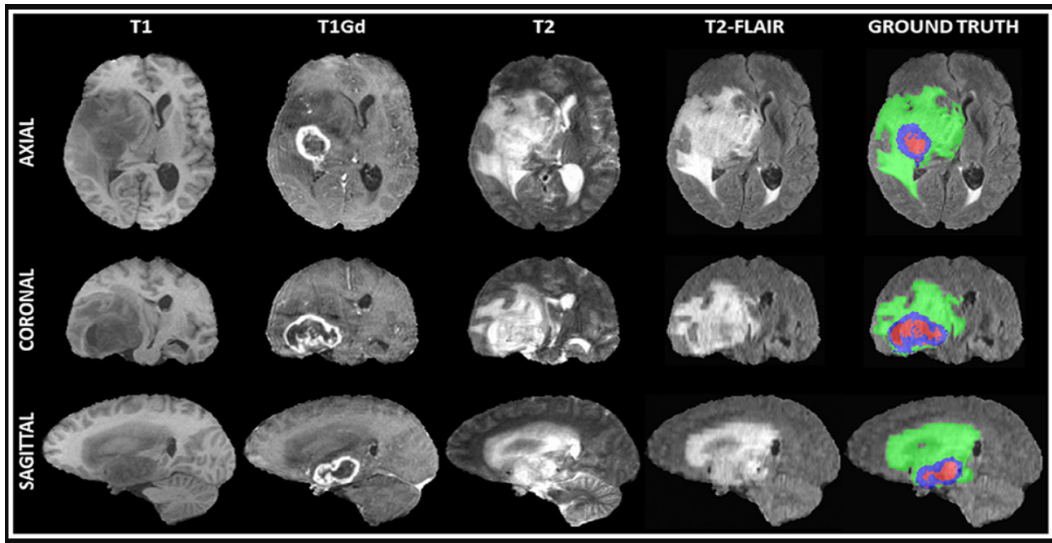
In recent years, research in computer vision, image processing and pattern recognition have increased significantly. Computer vision techniques are very effective in the fields of health and medical research. Medical imaging with computer vision has a significant impact on healthcare applications (Gao et al., 2018; Ulhaq et al., 2020). Medical imaging enables the structure of the human body to be viewed with different techniques to obtain information about the state of the organs and to determine the areas that need intervention. Medical imaging is important in diagnosing diseases, at the same time it provides a treatment plan by examining the progression and spread of the disease (Atlan & Pençe, 2021). X-ray, computed tomography (CT), magnetic resonance (MR), ultrasonography (US), which are medical imaging methods, are used in the diagnosis of Nuclear Medicine diseases. Examples of medical imaging methods are given in Figure 1. X-ray is the first medical imaging management with X-rays. The human body has a 3-dimensional (3D) structure, including width, height and depth. However, on the images of X-ray films, the structures in this depth are overlap. The CT imaging method uses X-rays, as in the X-ray method. This method provides a three-dimensional image by displaying the body in the form of cross-sections and eliminates the problem of overlapping depth information in images obtained by the X-ray method. CT Imaging method provides more detailed information than X-ray imaging. In the MR method, a three-dimensional image is obtained using radio waves and in the form of sections, as in CT images. In the US method, the imaging process is performed using high-frequency sound waves. In nuclear medicine, a chemical compound is injected into the patient through the mouth or breathing. As the compound radioactively decays in the body, the image is produced by the emitted X-rays or gamma rays (Ozekes, 2006).

Figure 1

Examples of (a) X-ray image, (b) CT image (c) MRI image (d) Ultrasound image (Kasban et al., 2015)



Analysis of images obtained by radiological methods helps physicians to understand internal anatomy and therefore plays an important role in medical application. Segmentation in medical images is an important issue. Image segmentation is the manual or automatic extraction of the region of interest in an image. Tissue and organ segmentation is performed on medical images (Norouzi et al., 2014). Figure 2 shows an example of segmentation on 3D MR images. Accurate segmentation of organs in medical images requires experience and takes a lot of time. The variability and complexity of anatomical structures in the human body has caused medical image segmentation to remain a difficult problem. Therefore, computer-aided systems that perform this process automatically have become necessary (Ma et al., 2021). Traditional methods and machine learning methods is used to automatic segmentation of medical images. Deep Learning ,the machine learning based methods, which have emerged as a subfield of machine learning, show the highest performance in medical image segmentation, especially in multiple organ segmentation (Fu et al., 2021). In this chapter, deep learning methods for segmentation in medical image are explored, especially based on segmentation of multi organ. Also, a study for multi organ segmentation is given.

Figure 2*Example of Segmentation on 3D MR Images (Reina et al., 2020)*

Deep learning in Multi-Organ Segmentation

Before deep learning models became popular, multi-organ segmentation used traditional and atlas-based methods. Traditional image segmentation uses mathematical models that manually extract features from the image and detect lines, edges (Fu et al., 2021). In atlas-based methods, manual labeling is done for segmentation. Based on these labels, an image record is created. Segmentation is performed by applying statistical methods to the image records. However, the differences in organs in the image records make segmentation difficult in terms of computation (Conze et al., 2021). To overcome this difficult situation, Deep Learning-based methods that do not require image record have been widely used for multiple organ segmentation in recent years and have achieved successful results (Lei et al., 2020).

Deep learning emerged as a sub-branch in machine learning. Performance in machine learning methods depends on the features extracted from the data. Processing large and complex data is very time consuming and difficult, which has led to the inadequacy of machine learning and the emergence of deep learning models. Deep learning models are based on artificial neural networks. Deep artificial neural networks consist of many layers by adding neural networks one after another. Deep learning is done by training these deep neural networks by extracting features from the data imported into the network thanks to nonlinear layers. The output obtained from each layer in the training phase is used in the next layer. After the training phase is complete, the unlabeled data is labeled using the extracted features (Dargan et al., 2020). There are several deep learning architectures used for multi-organ segmentation. These deep learning architectures are divided into pixel-based classification and end-to-end segmentation. In this chapter, Convolutional Neural Network (CNN), which performs pixel-based classification, Fully Convolutional Networks (FCN), U-Net and Region-based CNN (R-CNN), which perform end-to-end segmentation, are presented (Fu et al., 2021).

Convolutional Neural Network (CNN)

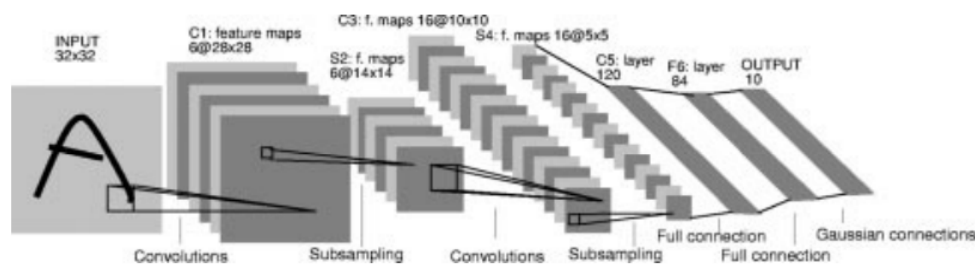
The first CNN architecture was designed by Yann LeCun in 1989 (Le Cun et al., 1989). This architecture, developed by 1998, was called LeNet (LeCun et al., 1998). LeNet

architecture is given in Figure 3. In the following years, more advanced CNN architectures have been proposed. Convolutional neural networks (CNN) are multilayer architectures consisting of convolution layers, usually nonlinear layers, batch normalization, pooling layers and fully connected layer. The convolution layer is the base layer of neural convolutional networks. Thanks to this layer, the feature map of the images is extracted, and there is more than one convolutional layer. New features are extracted from the input image by applying different convolution operations to the previously extracted feature maps (Guo et al., 2017). The pooling layer gradually decrease the size of the extracted feature vector of the given inputs. This reduction reduces the number of parameters to be computed by the model, thereby reducing computational costs and eliminating complexity (O'Shea & Nash, 2015). It consists of a nonlinear layer activation function in a CNN. The activation function helps to learn complex data. Choosing an appropriate activation function allows the model to learn faster and perform better. This layer converts an input into a nonlinear output (Khan et al., 2020). The batch normalization layer usually comes after the activation functions in deep neural networks and normalizes the inputs to a mean of zero and a fixed standard deviation (Bjorck et al., 2018). The last layer in a convolutional neural network is the fully connected layer. Fully connected layers resemble a classical neural network. It take the feature vector extracted from the previous layer, analyze it globally, and form a nonlinear combination for classification (Albawi et al., 2017; Khan et al., 2020).

While basic concepts such as edge, light and color are learned in the first layers of CNNs, the characteristics of the objects are learned in the next layers (Dargan et al., 2020). CNNs have sparse connections, so they have fewer parameters than a fully connected neural network with the same depth. Training in CNNs is performed using a gradient-based learning algorithm. Thanks to these algorithms, the generated weights are optimized by minimizing the loss (Alom et al., 2019).

Figure 3

LeNet Architecture (LeCun et al., 1998)



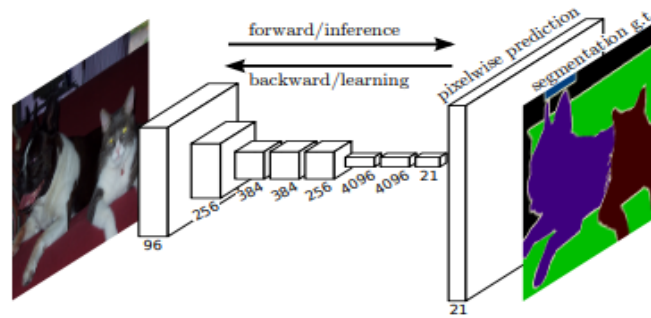
Fully Convolutional Networks (FCN)

The fully connected layers in classical convolutional neural networks cause the spatial information of the image to be lost. Since spatial information is important for image segmentation, classical convolutional networks are insufficient. To overcome this problem, fully connected neural networks have been proposed (Long et al., 2015). The FCN architecture given in Figure 4 is an architecture built on CNNs. FCN accepts images created in different sizes. Convolutional layers are used instead of fully connected layers in CNN architecture. FCN architectures have downsampling and upsampling parts. In the downsampling part, high-level semantic information is extracted from the images using convolutional layers. In the upsampling section, the size of the images is equalized to the input image by deconvolution of the feature maps created in the downsampling

section. This ensures that the spatial information of the original image is preserved. Finally, the segmentation output is obtained by classifying pixels from the obtained feature map. Based on the increase of upsampling, FCN architectures are divided into three types: FCN-32s, FCN-16s, and FCN-8s (Bi et al., 2018; Liu et al., 2021). In the FCN architecture, low-resolution feature maps are generated because the input is down-sampled by the convolutional layer and the pooling layer. To overcome this problem, skip connections are used in the FCN8 and FCN16 architectures. Thanks to the skip connections, high layer information and low layer information are combined to improve the pixel prediction result (Shelhamer et al., 2017).

Figure 4.

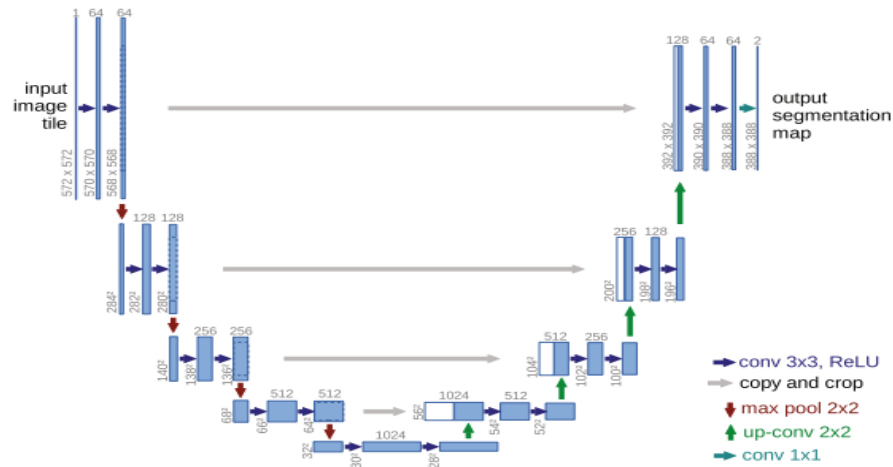
FCN Architecture (Long et al., 2015)



U-Net

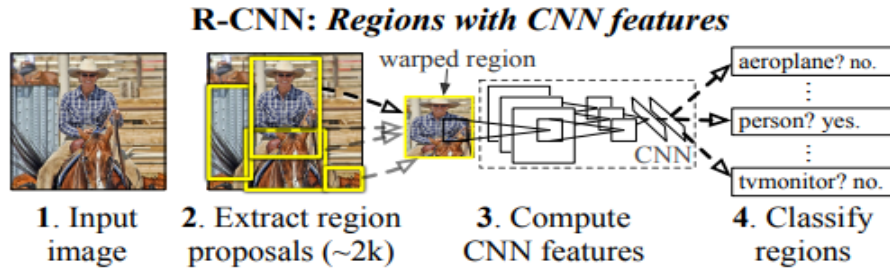
Ronneberger et al. (2015), developed the U-Net architecture, which is a popular variation of the FCN architecture. This architecture is designed for biomedical images and achieves high accuracies with a small number of images (Sorg, 2018).

As in the FCN architecture, it has fully convolutional layers and does not include a fully connected layer. There are two parts in the U-Net architecture, where encoding and decoding operations are performed. The encoder part consists of a typical CNN. It two consecutive 3x3 convolution layers are used at each level and ReLU is applied after convolution layers. Then, the number of steps is set to 2 and the downsampling is performed with the maximum pooling method of size 2x2. With each downsampling step, the size of the image decreases and the number of feature channels doubles. In the decoder part, the number of feature channels is halved using the 2x2 upsampling process. The feature vectors extracted in the encoder part of the mesh are combined with the corresponding feature vectors in the decoder part. Two consecutive 3x3 convolution operations are applied to these combined feature maps. After each convolution operation, the ReLU is applied. In the last layer, the images are matched to the desired number of classes by 1x1 convolution and the segmentation vector is extracted. In total, the U-net architecture has 23 convolutional layers (Ronneberger et al., 2015). In Figure 5, the general structure of the U-Net architecture is given.

Figure 5*U-Net Architecture (Ronneberger et al., 2015)*

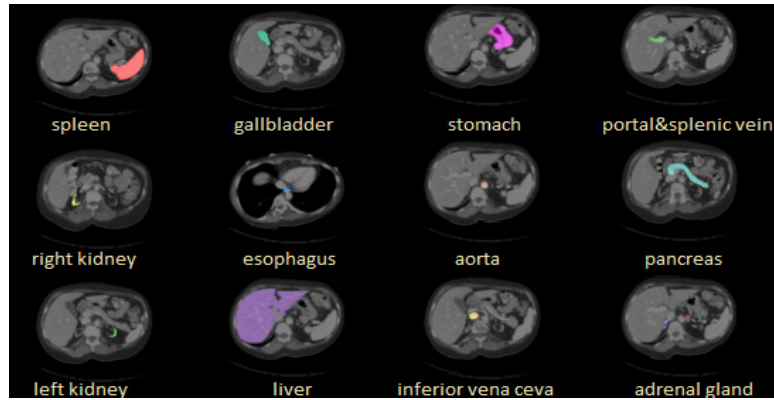
Region-based Convolutional Neural Network (R-CNN)

Object recognition requires localization of objects in the image as opposed to image classification. To solve the problem of object localization in images, a region-based CNN architecture is proposed. R-CNN architecture is shown in Figure 6. With this architecture, different objects in the images are drawn with the bounding box and classification (Girshick et al., 2014). In R-CNN method, 2000 candidate regions are extracted from the image to solve the problem of selecting a large number of regions. A selective search algorithm is used in this extraction of candidate regions (Uijlings et al., 2013). Each of the 2000 candidate regions extracted from the image is given as input to the CNN architecture and feature extraction is performed. The extracted feature maps are classified using the support vector machine (SVM) algorithm. Offset values are also estimated to determine the bounding boxes of the object. However, training the 2000 CNN architecture is very time consuming (Lei et al., 2020). To overcome this situation, the Fast-R-CNN architecture has been proposed (Girshick, 2015). In Fast R-CNN architecture, the feature vector is obtained using CNN architecture. From this feature map, candidate regions are extracted. A fixed size vector is generated using the ROI pooling layer for each candidate region. These vectors serve as input to fully connected layers. Two outputs are then obtained: the class of the object and its bounding box. The fast R-CNN architecture is faster than the R-CNN architecture because it uses a single CNN architecture (Girshick, 2015).

Figure 6*R-CNN Architecture (Girshick et al., 2014)*

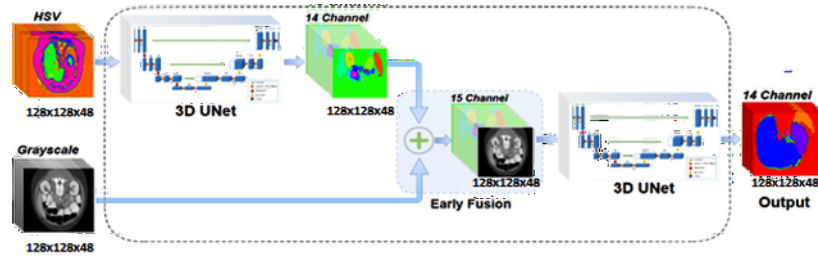
Case Study for Multi-Organ Segmentation

In this study, a computerized automatic diagnosis system was developed to assist radiologists in segmenting different organs simultaneously in abdominal computed tomography images. This study used the dataset provided by Vanderbilt University for the multiple organ segmentation competition. This dataset contains 30 images CT. The size of these images varies from 512x512x85 to 512x512x195. In addition, 13 organs are labeled in each CT image (Xu, 2015). Figure 7 shows the segmentation image of each organ. A Unet-based fusion model was proposed to perform the segmentation of these organs and this model is evaluated using the dice similarity coefficient. Before training the images of the dataset used in this study with the proposed model, their dimensions were equalized in the x, y, and z axes. image size is set to 128x128x48 and data augmentation was made by applying rotation and zooming operations to the CT images (Kayhan, 2022).

Figure 7*CT Images of Organs (Xu, 2015)*

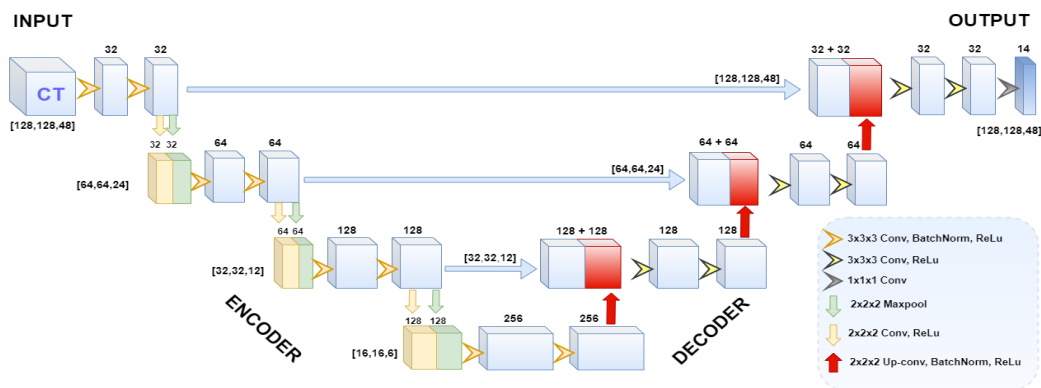
The multi-organ segmentation model proposed by Kayhan (2022) consists of two stages. In the first part, the images converted to HSV color space are trained with the 3D Unet-based model. The original images are combined with the segmentation map obtained as a result of this training, and the 3D Unet-based model in the second stage is taken as input and the predicted segmentation result is obtained. In figure 8, the general structure of the 3D U-Net based fusion model used in this study is given.

Figure 8
3D U-Net Based Fusion Model



In this study, the 3D U-Net model found in both stage is the same. This model has an encoder and a decoder section. In the encoder section, $3 \times 3 \times 3$ convolution operations are performed to extract the feature map from the input image. Batch normalization and ReLU activation functions are applied to the feature maps obtained after the convolution operations. When transitioning between levels, $2 \times 2 \times 2$ convolution with maximum pooling and 2 steps is applied. As a result of these processes, two outputs are obtained whose size is halved. The use of the maximum pooling layer is to prevent the model from remembering the features in the images and to prevent the loss of important information in the images by applying the convolution process. The outputs of these two layers are combined and used as input for the next layer. In the decoder section, the upsampling is performed until the size of the input image is reached. The feature vector extracted from each layer in the decoder section is combined with the feature vector from the encoder section. Then convolution operations are performed with a size of $3 \times 3 \times 3$. The segmentation image is created by implementing the $1 \times 1 \times 1$ convolution and the softmax to the feature map obtained in the last layer. The 3D U-Net based model used in this study is given in figure 9 (Kayhan, 2022).

Figure 9
3D U-Net Based Model



For the training of the 3D U-Net based model used in this study, the learning rate was set to $1e-3$, the batch size to 2, and the activation function to ReLU. This model is run in 300 steps on both raw and augmented data. Table 1 shows that data augmentation improves the segmentation outcome. After training with augmented data, the average segmentation result of all organs was found to be 79.6%, and the highest accuracy result was obtained for the liver organ, and the lowest accuracy rate was obtained for the left adrenal gland. Actual segmentation and estimated segmentation images are given in figure 10 and figure 11.

Organ segmentation is of great importance in medical image analysis. In this study, a computerised system was developed to automate the segmentation process

performed manually by radiologists. It has been shown that these systems can provide successful results and help radiologist.

Table 1

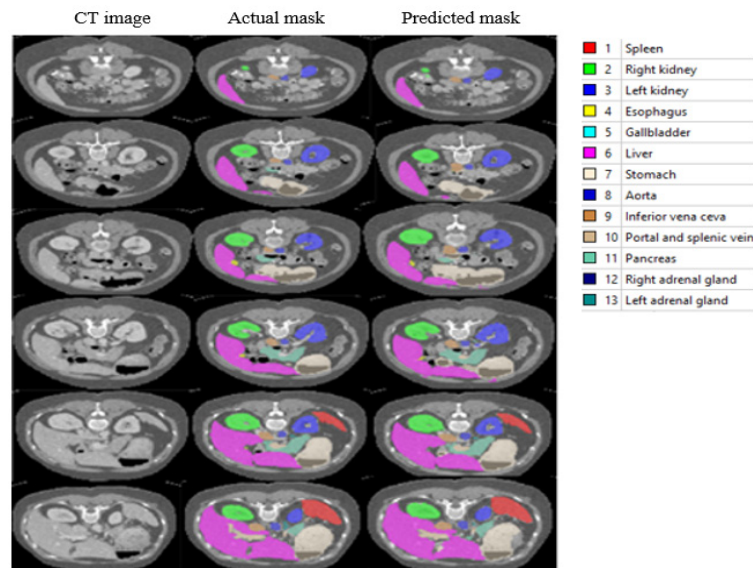
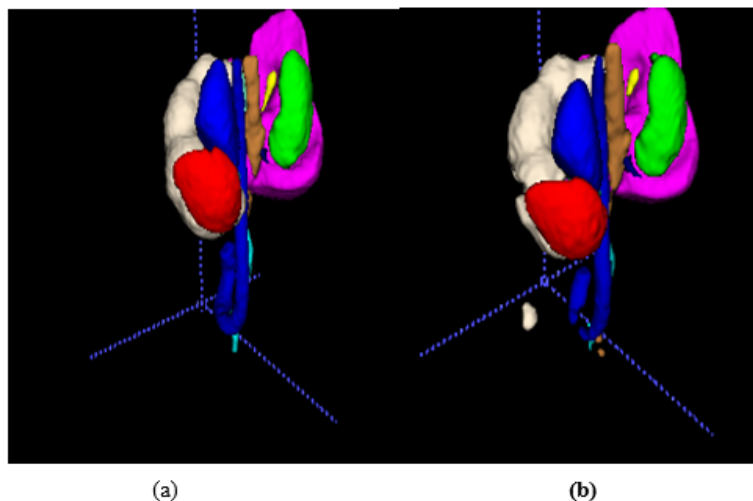
Segmentation Results Before and After Data Augmentation (Kayhan, 2022)

Organs	Data Augmentation	
	False	True
Spleen	0.929	0.94
Right kidney	0.918	0.934
Left kidney	0.92	0.937
Esophagus	0.704	0.698
Gallbladder	0.645	0.703
Liver	0.947	0.951
Stomach	0.766	0.847
Aorta	0.887	0.873
Inferior vena veva	0.798	0.816
Portal and splenic vein	0.638	0.698
Pancreas	0.663	0.774
Right adrenal gland	0.546	0.611
Left adrenal gland	0.517	0.558
Average	0.76	0.796

Table 2

Training and Testing Results of Augmented Data (Kayhan, 2022)

Organs	Data Augmentation	
	Train	Test
Spleen	0.951	0.94
Right kidney	0.937	0.934
Left kidney	0.939	0.937
Esophagus	0.887	0.698
Gallbladder	0.836	0.703
Liver	0.956	0.951
Stomach	0.917	0.847
Aorta	0.906	0.873
Inferior vena veva	0.879	0.816
Portal and splenic vein	0.811	0.698
Pancreas	0.841	0.774
Right adrenal gland	0.851	0.611
Left adrenal gland	0.834	0.558
Average	0.888	0.796

Figure 10*Actual Segmentation and Predicted Segmentation Images (Kayhan, 2022)***Figure 11***(a) Actual (b) Predictive 3D Segmentation Mask (Kayhan, 2022)*

References

- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6). IEEE.<https://doi.org/10.1109/ICEngTechnol.2017.8308186>
- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3), 292.<https://doi.org/10.3390/electronics8030292>
- Atlan, F., & Pençe, I. (2021). An overview of artificial intelligence and medical imaging technologies. *Acta Infologica*, 5(1), 207-230.<https://doi.org/10.26650/acin.813736>
- Bi, L., Feng, D., & Kim, J. (2018). Dual-path adversarial learning for fully convolutional network (FCN)-based medical image segmentation. *The Visual Computer*, 34,

- 1043-1052. <https://doi.org/10.1007/s00371-018-1519-5>
- Bjorck, J., Gomes, C., Selman, B., & Weinberger, K. Q. (2018). Understanding batch normalization. *arXiv preprint arXiv:1806.02375*.
- Conze, P.-H., Kavur, A. E., Cornec-Le Gall, E., Gezer, N. S., Le Meur, Y., Selver, M. A., & Rousseau, F. (2021). Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks. *Artificial Intelligence in Medicine*, 117, 102109. <https://doi.org/10.1016/j.artmed.2021.102109>
- Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27(4), 1071-1092. <https://doi.org/10.1007/s11831-019-09344-w>
- Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2021). A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica*, 85, 107-122. <https://doi.org/10.1016/j.ejmp.2021.05.003>
- Gao, J., Yang, Y., Lin, P., & Park, D. S. (2018). Computer vision in healthcare applications. *Journal of healthcare engineering*. <https://doi.org/10.1155/2018/5157020>
- Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448). <https://doi.org/10.1109/ICCV.2015.169>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
- Guo, T., Dong, J., Li, H., & Gao, Y. (2017). Simple convolutional neural network on image classification. . In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)* (pp. 721-724). IEEE. <https://doi.org/10.1109/ICBDA.2017.8078730>
- Kasban, H., El-Bendary, M., & Salama, D. (2015). A comparative study of medical imaging techniques. *International Journal of Information Science and Intelligent System*, 4(2), 37-58.
- Kayhan, B. (2022). *Deep learning based multi organ segmentation in computed tomography images* (Master's thesis, Konya Teknik Üniversitesi).
- Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8), 5455-5516. <https://doi.org/10.1007/s10462-018-9633-3>
- Le Cun, Y., Jackel, L. D., Boser, B., Denker, J. S., Graf, H. P., Guyon, I., & Hubbard, W. (1989). Handwritten digit recognition: Applications of neural network chips and automatic learning. *IEEE Communications Magazine*, 27(11), 41-46. <https://doi.org/10.1109/35.41400>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>
- Lei, Y., Fu, Y., Wang, T., Qiu, R. L., Curran, W. J., Liu, T., & Yang, X. (2020). Deep learning in multi-organ segmentation. *arXiv preprint arXiv:2001.10619*. <https://doi.org/10.48550/arXiv.2001.1061>

- Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3), 1224. <https://doi.org/10.3390/su13031224>
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
- Ma, J., Zhang, Y., Gu, S., Zhu, C., Ge, C., Zhang, Y., & Liu, X. (2021). Abdomenct-1k: Is abdominal organ segmentation a solved problem? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10), 6695-6714. <https://doi.org/10.1109/TPAMI.2021.3100536>
- Norouzi, A., Rahim, M. S. M., Altameem, A., Saba, T., Rad, A. E., Rehman, A., & Uddin, M. (2014). Medical image segmentation methods, algorithms, and applications. *IETE Technical Review*, 31(3), 199-213. <https://doi.org/10.1080/02564602.2014.906861>
- O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*. <https://doi.org/10.48550/arXiv.1511.08458>
- Özekes, S. (2006). *Computer aided detection in medical imaging* (Doctoral dissertation, Marmara Universitesi (Turkey)).
- Reina, G. A., Panchumorthy, R., Thakur, S. P., Bastidas, A., & Bakas, S. (2020). Systematic evaluation of image tiling adverse effects on deep learning semantic segmentation. *Frontiers in neuroscience*, 14, 65. <https://doi.org/10.3389/fnins.2020.00065>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28
- Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation [M]. *IEEE Computer Society*. <https://doi.org/10.1109/TPAMI.2016.2572683>
- Sorg, B. R. (2018). *Multi-Task Learning SegNet Architecture for Semantic Segmentation* (Doctoral dissertation, University of Dayton)
- Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International journal of computer vision*, 104, 154-171. <https://doi.org/10.1007/s11263-013-0620-5>
- Ulhaq, A., Khan, A., Gomes, D., & Paul, M. (2020). Computer vision for COVID-19 control: a survey. *arXiv preprint arXiv:2004.09420*. <https://doi.org/10.48550/arXiv.2004.09420>
- Xu, Z. (2015). *Multi-atlas labeling beyond the cranial vault - workshop and challenge*. Retrieved 6.3.2021 from <https://www.synapse.org/#!/Synapse:syn3193805/wiki/217760>

About The Authors

Beyza KAYHAN works as a research assistant in the department of computer engineering at Konya Technical University. She has a bachelor's degree in computer engineering from Selçuk University and a master's degree from Konya Technical University. She is currently doing her PhD at Konya Technical University. Her areas of interest are artificial intelligence, image processing and deep learning.

E-mail: bkayhan@ktun.edu.tr, **ORCID:** 0000-0002-7211-4360.

Sait Ali UYMAZ is an associate professor of computer engineering at Konya Technical University. He has a PhD in computer engineering from Selcuk University. His main areas of interest are artificial intelligence, data science and optimization.

E-mail: sauymaz@ktun.edu.tr, **ORCID:** 0000-0003-2748-8483

Similarity Index

The similarity index obtained from the iThenticate software for this book chapter is 8%.

To Cite This Chapter:

Kayhan, B. & Uymaz, S.A. (2023). Multi Organ Segmentation in Medical Images. In S.N. Guner, Y. Uzun & I. Cetin (Eds.), *Current Studies in Healthcare and Technology* (pp.59-72). ISRES Publishing.