3D U-JAPA-Net: Mixture of Convolutional Networks for Abdominal Multi-organ CT Segmentation

Hideki Kakeya[1(✉)], Toshiyuki Okada[1], and Yukio Oshiro[2]

[1] Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan
kake@iit.tsukuba.ac.jp
[2] Ibaraki Medical Center, Tokyo Medical University, Ami, Japan

Abstract. This paper introduces a new type of deep learning scheme for fully-automated abdominal multi-organ CT segmentation using transfer learning. Convolutional neural network with 3D U-net is a strong tool to achieve volumetric image segmentation. The drawback of 3D U-net is that its judgement is based only on the local volumetric data, which leads to errors in categorization. To overcome this problem we propose 3D U-JAPA-net, which uses not only the raw CT data but also the probabilistic atlas of organs to reflect the information on organ locations. In the first phase of training, a 3D U-net is trained based on the conventional method. In the second phase, expert 3D U-nets for each organ are trained intensely around the locations of the organs, where the initial weights are transferred from the 3D U-net obtained in the first phase. Segmentation in the proposed method consists of three phases. First rough locations of organs are estimated by probabilistic atlas. Second, the trained expert 3D U-nets are applied in the focused locations. Post-process to remove debris is applied in the final phase. We test the performance of the proposed method with 47 CT data and it achieves higher DICE scores than the conventional 2D U-net and 3D U-net. Also, a positive effect of transfer learning is confirmed by comparing the proposed method with that without transfer learning.

Keywords: Convolutional neural networks · Deep learning · Transfer learning U-net · 3D U-net · Multi-organ segmentation · Mixture of experts

1 Introduction

Multilayer neural networks attracted great attention in the 1980s and in the early 1990s. The most influential work was the invention of error back-propagation learning [1], which is still used in current deep neural networks. During those years several types of network architectures were tried, such as Neocognitron [2] and mixture of experts [3]. After the long ice age of neural networks from the late 1990s to around 2010, the idea of deep convolutional neural networks (CNNs), which inherited some features of Neocognitron, was proposed [4] and realized unprecedented performance in the area of image recognition. Owing to the rapid progress of graphical processor units (GPUs),

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11073, pp. 426–433, 2018.
https://doi.org/10.1007/978-3-030-00937-3_49

fast training of deep convolutional neural networks has been enabled with a low-cost PC, which leads to the current boom of deep learning.

Deep learning using a CNN can be a strong tool in the area of medical imaging also. Since the proposal of U-net [5], which is based on fully convolutional network (FCN) [6], deep CNNs have been applied to various biomedical image segmentation tasks and have outperformed the conventional algorithms. To apply deep CNNs to 3D volume data, 3D U-net has been proposed [7], where 3-dimensional convolutions are applied to attain volumetric segmentation. 3D U-net is easily applied to multi-organ CT segmentation, which is an important pre-process for computer-aided diagnosis and therapy.

The drawback of 3D U-net is that its judgement is based only on the local volumetric data, which often leads to errors in multi-organ segmentation. Some modifications of learning have been tried to overcome this problem. For example, Roth et al. proposed a hierarchical 3D FCN that takes a coarse-to-fine approach, where the network is trained to delineate the organ of interest roughly in the first stage and is trained for detailed segmentation in the second stage [8]. A probabilistic approach can be merged with FCNs [9], but the performance is not improved significantly.

Before the rise of deep learning, several approaches to multiple organ segmentation from 3D medical images had been proposed. These approaches commonly utilize a number of radiological images with manual tracing of organs, called atlases, as training data, and can be classified into multi-atlas label fusion, machine learning, and statistical atlas approaches.

Statistical atlas approaches have been most commonly applied to abdominal organ segmentation. Explicit prior models constructed from atlases, such as the probabilistic atlas (PA) [10, 11] and statistical shape models [12, 13] are used in these approaches.

Okada et al. proposed abdominal multi-organ segmentation method using conditional shape-location and unsupervised intensity priors (S-CSL-UI), assuming that variation of shape and location of abdominal organs were constrained by the organs whose segmentation was stable and relatively accurate [14]. The method using hierarchical modeling interrelation of organs improved the accuracy and stability of segmentation, and it demonstrated effective reduction of the search space. These methods, however, have been outperformed by CNNs.

In this paper, we propose a new 3D U-net learning scheme, which we name 3D U-JAPA-net (Judgement Assisted by PA). The proposed scheme utilizes not only CNNs but also PA information to overcome the drawbacks of the conventional 2D U-net and 3D U-net. Also, the proposed method comprises transfer learning [15] and mixture of experts to make effective use of PA information so that more accurate multi-organ segmentation may be attained.

2 Methods

The goal in this paper is to realize fully-automated segmentation of 8 abdominal organs: liver; spleen; left and right kidneys; pancreas; gallbladder (GB); aorta; and inferior vena cava (IVC). For this purpose, we compare the performances of the

following 5 methods: S-CSL-UI; 2D U-net; 3D U-net; Mixture of 3D U-nets; and 3D U-JAPA-net, which we propose in this paper.

A U-net consists of a contracting path and an expansive path. The contracting path follows the typical architecture of a convolutional network. It consists of repeated application of two 3×3 convolutions, each followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation. At each down-sampling step, the number of feature channels is doubled. Every step in the expansive path consists of an up-sampling of the feature map followed by a 2×2 up-convolution that halves the number of feature channels, a concatenation with the feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. Cropping is needed due to the loss of border pixels in every convolution. At the final layer, a 1×1 convolution is used to map each 64-component feature vector to the desired number of classes.

3D U-net is a simple expansion of 2D U-net, where both convolution and max pooling operates in 3 dimensions like $3 \times 3 \times 3$ or $2 \times 2 \times 2$. In [7], batch normalization (BN) [16] is introduced before each ReLU. Though 3D U-net can reflect the 3D structure of CT data, size of calculation becomes huge when the input data size is large.

3D U-JAPA-net, which we introduce here, is an expansion of 3D U-net. The learning scheme of 3D U-JAPA-net is shown in Fig. 1.

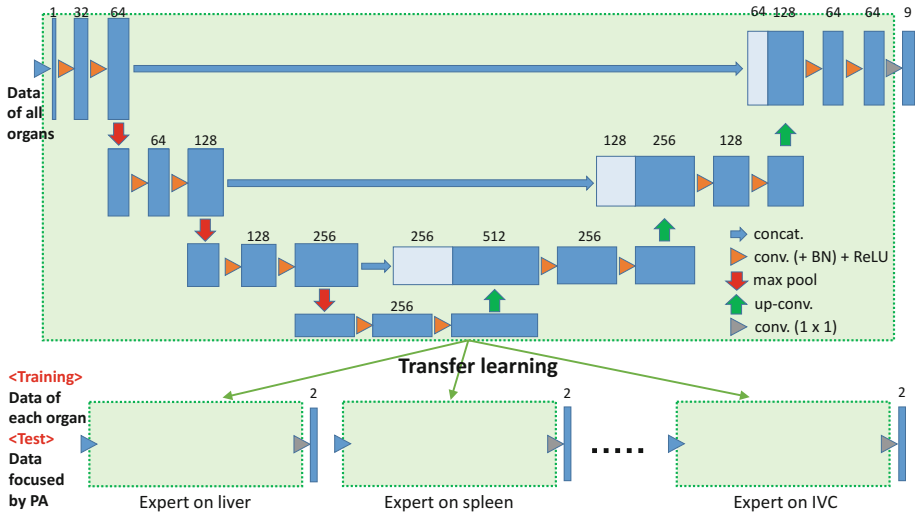


Fig. 1. Learning scheme of 3D U-JAPA-net. Blue boxes represent feature maps and the numbers of feature maps are denoted on top of each box. (Color figure online)

In the first learning phase, a 3D U-net with 9 output layers corresponding each class (8 organs and background) is trained using the whole data inside the bounding boxes of all organs. Also, we prepare PA for each organ based on the training data with the method in [14]. After the first training converges, the weights of this network are transferred to 8 expert 3D U-nets for each organ, each of which has 2 output layers

(the organ and the background). Therefore the initial weights of 8 networks are the same except for those connected to the final output layer. In the second learning phase, each expert 3D U-net specialized for each organ accepts volumetric data including the corresponding organ, and the weights are modified by the gradient descent method.

In the test phase, the trained network specialized for each organ accepts data including the voxels whose PA values of that organ are non-zero. If the output "organ" is larger than the output "background" in the final layer, that voxel is labeled as part of that organ.

To see the effect of transfer learning in the above scheme, we also test the system where the first learning phase is removed from 3D U-JAPA-net, which means that each expert network starts from random weights before training.

Since the judgement is given by voxel unit in the U-net based systems, debris emerges in the result. We apply largest component selection as the post-process to remove debris for all the U-net based systems.

3 Experiments and Results

We compared the performances of the following 5 methods: S-CSL-UI; 2D U-net; 3D U-net; Mixture of 3D U-nets for each organ without transfer learning (3D M-U-nets); and 3D U-JAPA-net.

Each method was tested to segment 8 abdominal organs: liver; spleen; left and right kidneys; pancreas; GB; aorta; and IVC. We used 47 CT data from 47 patients with normal organs obtained in the late arterial phase at the same hospital and applied two-fold cross-validations to evaluate the performance of each method. The resolution of each CT slice image was 512×512 pixels. Among 47 CT data, 9 data had 159 slices and the voxel size was $0.625 \times 0.625 \times 1.25$ [mm^3]. The voxel size of other 37 data was $0.781 \times 0.781 \times 0.625$ [mm^3] and the numbers of slices were between 305 and 409. The last one, consisting of 345 slices, had $0.674 \times 0.674 \times 0.625$ [mm^3] voxels.

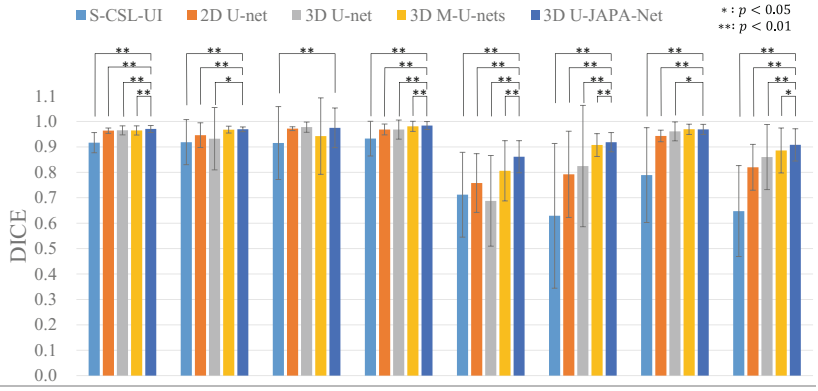
For 2D U-net, the slice images were first down-sampled to 256×256 pixels. Then the same algorithm in [5] was used, where 3×3 convolutions were applied twice in each layer and the max pooling and up-conversion were applied 4 times. For 3D U-net, the input to the network was a $132 \times 132 \times 116$ voxel tile of the image with 1 channel. After that, the same algorithm in [7] was used, where $3 \times 3 \times 3$ convolutions were applied twice in each layer, while the max pooling and up-conversion were applied 3 times. The output of the final layer becomes $44 \times 44 \times 28$ voxels due to the repeated truncation in every convolution.

We applied dropout of connections in the bottom layer to avoid over-fitting both in 2D U-net and 3D U-net. Data augmentation was not applied in this experiment for simplicity. Also, the training data and the test data are made so that the output voxels may not overlap in order to reduce the calculation time.

As for 3D U-JAPA-net, the above 3D U-net was used both in the first and the second stages of training. When the weights were transferred from the trained network to the expert network for each organ, the connections between the last layer and the second last layer were randomized, for the numbers of output layers were different between the 3D U-net in the first stage of training and the expert 3D U-nets for each organ.

All the U-net components were implemented with TensorFlow framework [17]. The PC we used was composed of Intel Core i7-8700 K CPU, 32 GB main memory, and NVIDIA GeForce GTX 1080 Ti GPU with 11 GB video memory. The detail of training was as follows: training epochs = 30; learning rate = 1.0×10^{-4}; batch size = 3. It took 2.5 h to train 2D-U-net with this PC. As for 3D U-net, it took 16 h to train all organs and it took between 40 and 130 min to train each organ respectively except for the liver, which took 9.5 h to train because of its large size.

Figure 2 shows the DICE scores given by the 5 segmentation methods. The results of paired t-tests between the proposed method and the other methods are indicated in the figure. As the figure shows, the proposed method attains notably better performance than the conventional methods. The performance given by the mixture of experts without transfer learning is poorer, which shows that transfer learning is effective to attain high DICE scores.



		liver	spleen	r-kidney	l-kidney	pancreas	GB	aorta	IVC
DICE	S-CSL-UI	0.917	0.919	0.915	0.933	0.712	0.629	0.789	0.647
	2D U-net	0.964	0.946	0.972	0.969	0.758	0.792	0.943	0.820
	3D U-net	0.965	0.932	**0.978**	0.968	0.688	0.825	0.961	0.860
	3D M-U-nets	0.965	0.968	0.943	0.981	0.806	0.908	**0.969**	0.886
	3D U-JAPA-Net	**0.971**	**0.969**	0.975	**0.984**	**0.861**	**0.918**	**0.969**	**0.908**

		liver	spleen	r-kidney	l-kidney	pancreas	GB	aorta	IVC
Recall	S-CSL-UI	0.977	0.958	0.958	0.964	**0.831**	0.794	0.851	0.720
	2D U-net	0.964	0.946	**0.970**	0.965	0.689	0.734	0.941	0.806
	3D U-net	**0.983**	0.904	0.969	0.948	0.564	0.840	0.942	0.815
	3D M-U-nets	0.977	0.963	0.929	0.977	0.776	0.891	**0.968**	0.858
	3D U-JAPA-Net	0.978	**0.966**	0.969	**0.981**	0.816	**0.906**	0.964	**0.903**

		liver	spleen	r-kidney	l-kidney	pancreas	GB	aorta	IVC
Precision	S-CSL-UI	0.865	0.895	0.881	0.909	0.635	0.547	0.753	0.601
	2D U-net	**0.964**	0.952	0.973	0.973	0.880	0.917	0.948	0.842
	3D U-net	0.948	**0.982**	0.987	**0.992**	**0.968**	0.832	**0.984**	**0.936**
	3D M-U-nets	0.953	0.973	0.988	0.985	0.867	0.932	0.972	0.928
	3D U-JAPA-Net	**0.964**	0.973	**0.990**	0.987	0.922	**0.937**	0.975	0.920

Fig. 2. Results of eight abdominal organ segmentation by five methods. Bold numbers are the highest DICE/recall/precision rates for each organ.

Figure 3 shows an example of segmentation results, which effectively demonstrate usefulness of the proposed method. 3D U-JAPA-net can recall part of pancreas and GB that other U-net based systems miss. Also, the leakage, which stands out in S-CSL-UI, is not apparent in the other methods including 3D U-JAPA-net.

The effect of increase in training data was tested by comparing 2-fold and 4-fold cross-validations of 3D U-JAPA-net. The result is shown in Table 1. The DICE score is improved significantly in the segmentation of pancreas ($p = 0.012$) by applying 4-fold cross-validation.

The performances of the proposed method and the prior method [8], which is also a modified version of 3D U-net, are compared in Table 2. DICE scores obtained by the proposed method are distinctively higher, which indicates the excellence of the proposed method.

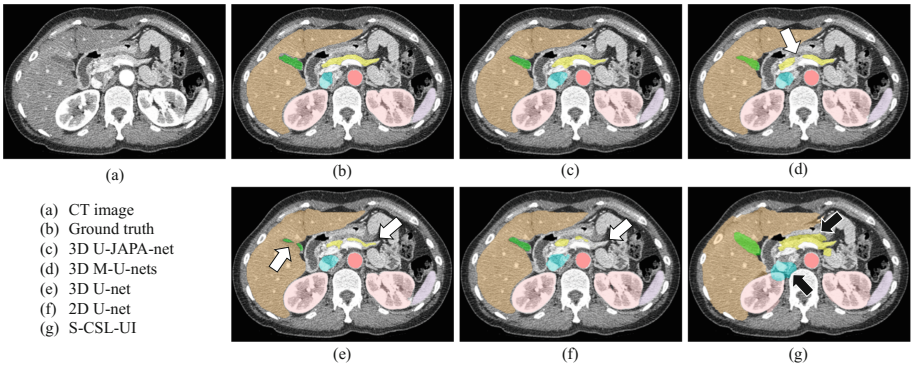


Fig. 3. An illustrative segmentation results obtained by five methods. White arrows show the failed regions by the conventional priors. Black arrows show the leakages by the conventional priors. (Color figure online)

Table 1. Comparison of 3D U-JAPA-net DICE scores obtained by two-fold and four-fold cross-validations.

	Liver	Spleen	r-kidney	l-kidney	Pancreas	GB	Aorta	IVC
2-fold	0.971	0.969	0.975	0.984	0.861]*	0.918	0.969	0.908
4-fold	0.971	0.969	0.986	0.985	0.882]	0.915	0.966	0.907

Table 2. Comparison of two modified versions of 3D U-net.

		Roth et al. [8]			3D U-JAPA-net		
		Liver	Spleen	Pancreas	Liver	Spleen	Pancreas
DICE	Mean	0.954	0.928	0.822	0.971	0.969	0.882
	Std	0.020	0.080	0.102	0.014	0.014	0.070
	Median	0.960	0.954	0.845	0.974	0.973	0.901
Subjects		150 (testing)			47 (4-fold cross validation)		

4 Discussion

When 2D U-net and 3D U-net are compared, 2D U-net is good at segmenting larger organs, while 3D U-net is adept at segmenting smaller organs. Since 2D U-net covers larger areas in a single slice, it can grasp wider areas with a single shot, which leads to the above characteristic of performance.

3D U-JAPA-net overcomes the drawback of 3D U-net, which covers a smaller area in each slice, with the help of PA and outperforms both 2D U-net and 3D U-net in segmentation of almost all organs. Improvement by 3D U-JAPA net is especially significant in segmentation of pancreas, GB, and IVC, which have been difficult to segment properly for the conventional methods. The effect of transfer learning is significant in these organs, which shows the validity of the proposed method.

The number of data used here is limited and further study with a larger data size is needed to increase the reliability of the proposed method. In general, however, deep neural networks can attain better performance when the number of training data increases. A higher DICE score may be obtained if we use a larger data set for training with the proposed method.

In this paper PA has been used to see where the value is non-zero or not, for the number of CT samples is small and the values are discrete. When the number of samples is increased and the probabilities become more reliable, arithmetic usage of the probability values can raise DICE scores.

5 Conclusion

In this paper, we have proposed 3D U-JAPA-net, which uses not only the raw CT data but also the probabilistic atlas of organs to reflect the information on organ locations to realize fully-automated abdominal multi-organ CT segmentation. As a result of the 2-fold cross-validation with 47 CT data from 47 patients, the proposed method has marked significantly higher DICE scores than the conventional 2D U-net and 3D U-net in the segmentation of most organs.

The proposed method can be easily implemented for those who can use TensorFlow or similar deep learning tools, for all needed to be done in the proposed method is to make a probabilistic atlas, train a 3D U-net, copy the trained weights to the mixture of 3D U-nets, and train those 3D U-nets. The method described here is worth a trial for those who want to make a reliable fully-automated multi-organ segmentation system with little effort.

Acknowledgements. This research is partially supported by the Grant-in-Aid for Scientific Research, JSPS, Japan, Grant number: 17H00750.

References

1. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986)
2. Fukushima, K., Miyake, S., Ito, T.: Neocognitron: a neural network model for a mechanism of visual pattern recognition. IEEE Trans. Syst. Man Cybern. **SMC-13**(3), 826–834 (1983)
3. Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E.: Adaptive mixtures of local experts. Neural Comput. **3**(1), 79–87 (1991)
4. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **1**, 1097–1105 (2012)
5. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
6. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. arXiv:1411.4038 (2014)
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds.) MICCAI 2016, LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
8. Roth, H., et al.: Hierarchical 3D fully convolutional networks for multi-organ segmentation. arXiv:1704.06382 (2017)
9. Yang, Y., Oda, M., Roth, H., Kitasaka, T., Misawa, K., Mori, K.: Study on utilization of 3D fully convolutional networks with fully connected conditional random field for automated multi-organ segmentation form CT volume. J. JSCAS **19**(4), 268–269 (2017)
10. Park, H., Bland, P.H., Meyer, C.R.: Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Trans. Med. Imag. **22**(4), 483–492 (2003)
11. Zhou, X., et al.: Constructing a probabilistic model for automated liver region segmentation using non-contrast Xray torso CT images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, LNCS, vol. 4191, pp. 856–863. Springer, Berlin (2006). https://doi.org/10.1007/11866763_105
12. Heimann, T., Meinzer, H.-P.: Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal. **13**(4), 543–563 (2009)
13. Lamecker, H., Lange, T., Seebaß, M.: Segmentation of the liver using a 3D statistical shape model. Technical report. Zuse Institute, Berlin (2004)
14. Okada, T., Linguraru, M.G., Hori, M., Summers, R.M., Tomiyama, N., Sato, Y.: Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. Med. Image Anal. **26**(1), 1–18 (2015)
15. Pratt, L.Y.: Discriminability-based transfer between neural networks. NIPS Conf.: Adv. Neural Inf. Process. Syst. **5**, 204–211 (1993)
16. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 (2015)
17. Abadi, M., Agarwal, A., Barham, P., et al.: TensorFlow: large-scale machine learning on heterogeneous systems. arXiv:1603.04467 (2016)