Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

Evaluation of deep learning model with Optimizing and Satisficing metrics for Lung Segmentation

Usma Niyaz, Abhishek Singh Sambyal, and Devanand

Central University of Jammu, J&K, INDIA
{usmabhatt, abhishek.sambyal, devanand}@gmail.com

Abstract. The segmentation in medical image analysis is a crucial and prerequisite process during the diagnosis of the diseases. The need for segmentation is important to attain the region of interest where the probability of occurrence of an abnormality such as a nodule in the lungs or tumor in the brain is high. In this paper, we have proposed a new architecture called FS-Net which is a convolutional neural network based model for the segmentation of lungs in CT scan images. It performs encoding of images into the feature maps and then decodes the feature maps into their respective lung masks. We have also trained the state-of-the-art U-Net on the same dataset and compared the results on the basis of optimizing and satisficing metrics. These metrics are useful for the selection of a better model with the maximum score at the satisfying condition. The FS-Net is computationally very efficient and achieves promising dice coefficient and loss score when compared with the U-Net taking one-third of the time.

Keywords: FS-Net, U-Net, data augmentation, lung segmentation, neural network, deep learning

1 Introduction

Deep Convolutional Neural Networks (DCNN) have certainly given the astonishing results in image analysis such as classification [1], segmentation [2], detection and localization [3]. The rise of the deep learning techniques in medical image analysis for diagnosis of patients has proven to be effectively beneficial to the society by reducing the efforts of radiologists and pathologists by giving the human level accuracy. Segmentation in the medical images is a complicated problem but it increases the efficiency of the model by performing analysis only in the region of interest and neglecting the irrelevant information. The segmentation is done in the preprocessing stage, so any mistake committed affects the other stages resulting in undesirable output. The manual segmentation is a slow process so there is a need for computer-aided systems that can do segmentation fast and accurate without human interaction. Automatic processing of medical images without human intervention reduces time, cost and the human error. The existing Computer Aided systems have produced appreciable results in segmentation but deep learning has outperformed the experts.

Traditional machine learning approaches for medical image segmentation including the graph-cut approach [4], amplitude segmentation based on histogram features [5] involves the long sequence of algorithms and the filters were manually chosen for the

Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

detection of the features such as lines, edges, and curves which was very time consuming and often fails when the images were of low contrast, blur, noisy which increases the efforts of experts. However, the deep learning approaches automatically chooses the best filters for an image and results in most prominent features for the better classification and segmentation.

In the big data domain, where the medical data is enormously increasing, it becomes necessary to utilize such a big amount for useful work. The utilization of deep neural networks to get trained on the large data speeds up the process of the diagnosis of the disease and performs comparatively better than human. However, in medical image segmentation, there is a limitation of less data availability and class imbalance. Therefore, augmentation techniques like rotation, shifting, translation, and scaling are applied to increase the data for the network to get trained well. Due to the ability to process and learn from a large amount of data, deep learning techniques segment the region of interest in an accurate manner. To measure the loss and similarity between the predicted and actual output, two techniques cross-entropy and dice similarity are used for the efficient training of classification and segmentation tasks [6].

In our work, we have proposed a CNN based architecture that is trained on lung CT scan images for the prediction of lung masks. The model consists of the encoder and decoder networks and the main idea is to get the features of CT scan images in the encoder network which includes the pooling layers that downsample the resolution of images by half and then use the upsampling layers in the decoder network to get the downsampled images into high-resolution images yielding lung masks. The network has all the vital layers of CNN like convolutional layer, max pooling, ReLU and dense layer, so this architecture is so called CNN based. Due to the less training examples in our dataset, we have used data augmentation to improve the training of the model. This allows the network to learn invariances such as spatial, transformational and rotational invariances. Dosovitskiy et al. [7] have shown in the unsupervised learning, how data augmentation helps in learning invariance.

Rest of the paper is organized as follows, Section 2 contains the related work of segmentation in medical images, Section 3 contains the methodology of the proposed architecture, Section 4 describes the results, and Section 5 concludes the research work with future directions.

2 Related Work

Many semi-automated and automated methods are used for the segmentation of medical images [8] such as thresholding, region growing, classifier, clustering, Markov random field model, artificial neural network (ANN), deformable models, and metamorphs model [9]. These methods are popular and still in use for small-scale analysis. For the segmentation of 2D biomedical images, the fully convolutional network-based architecture called U-Net is used [10]. Another network that is used for the segmentation in the medical image analysis is the Deep Contour Aware Networks (DCAN) [11]. This architecture is FCN with multilevel contextual features used for the segmentation of glands. The architecture has won 2015 MICCAI gland segmentation competition. Besides 2D image segmentation, 3D image segmentation model was also introduced for

Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

the volumetric segmentation of MR images of the prostate. The architecture is called V-Net [12].

The cascade fully convolutional neural network was introduced where three networks hierarchically segment substructures of the brain which is a whole tumor (W-Net), tumor core (T-Net) and sequentially enhancing tumor core (E-Net) [13]. The deep learning framework for interactive image segmentation of brain tumor was proposed. The model was trained on binary segmentation [14].

Many architectures proposed for the segmentation of medical images is inspired by the U-Net architecture. Brahim et al. [15] have used U-Net architecture for the segmentation of lungs in CT scan images and achieved the Dice coefficient index of about 0.9502. The U-Net based convolutional neural network was proposed for the segmentation of lungs using X-Rays with manually prepared lung masks [16].

The CNN-based method with 3D filters with some modifications over the existing U-Net architecture is used for the segmentation of brain tumor and bone in hands using MR images. Modifications made in the U-Net architecture were multiple segmented maps created at different scales and use of element-wise summation to forward feature maps [17].

3 Methodology

3.1 Preprocessing

The dataset used to train our model is from Kaggle Competition “Finding and Measuring lungs in CT data”. It consists of 267 CT scan images and lung masks of size 512x512 which is the standard size of the dicom images. The range of pixel values is varying for each CT scan image and the pixel values of their corresponding masks range from [0, 255]. Because of the varying range of pixel values, the CT scan images

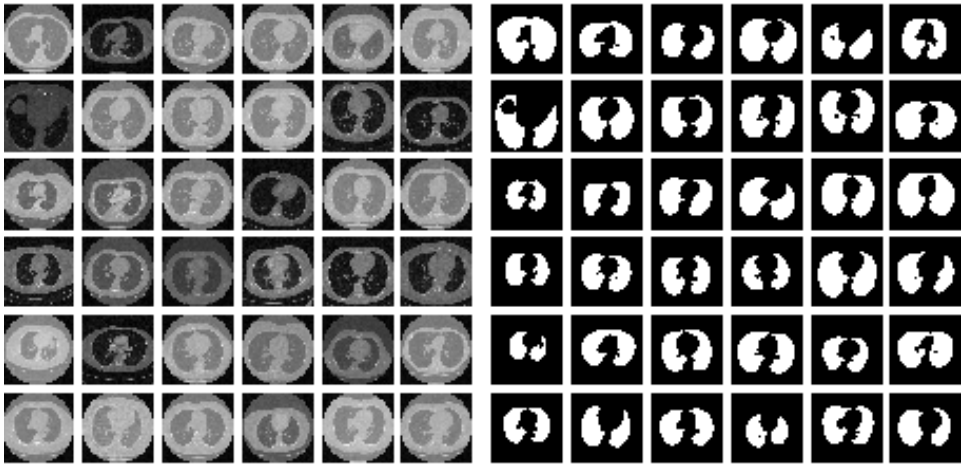

Fig. 1 Normalized CT scan slices and their corresponding masks.

Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

are normalized with their corresponding mask as shown in Fig. 1. Normalization is done to scale down the pixel values *f(x,y)*, of an image *f*, to a fixed range. The CT scan images and their masks are normalized to a range [0, 1] using Eq. (1).

$$f(x,y) = \frac{f(x,y) - min(f)}{max(f) - min(f)} \tag{1}$$

The dataset contains less training examples for the FS-Net to train well, so the images are augmented by shifting in width and height by 0.1, rotated by 45° and zoomed by 0.1, the same is done for their corresponding masks as shown in Fig. 2. Image augmentation helps the model to get generalized and robust by learning invariance, which also helps in reducing the overfitting. At every epoch, the images are augmented and given to our model for comprehensive learning which makes the model universal and performs well on highly varying input data.

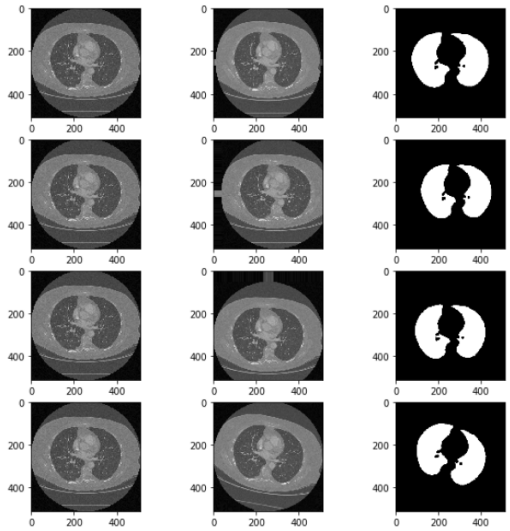

Fig. 2 Column 1 is the original CT scan images and Column 2 shows the augmented images that are zoomed, shifted in width, height by 0.1 and rotated by 45°. Column 3 shows their corresponding augmented masks.

3.2 FS-Net Architecture

The architecture of our proposed model consists of the encoder and a corresponding decoder network. The visualization of our proposed model is shown in Fig. 3 and described in detail in Table 1.

The encoder network consists of 10 convolutional layers with 3x3 kernel size. The convolutional layer is the convolution of two signals, element-wise multiplication and

Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

Table 1 Summary of the FS-Net.

Input	Output	Kernel	Activation	Parameters
Input	1 x 512 x 512	-	-	-
Conv1	32 x 512 x 512	3 x 3	ReLU	320
Conv2	32 x 512 x 512	3 x 3	ReLU	9248
Maxpool1	32 x 256 x 256	2 x 2	-	-
Conv3	64 x 256 x 256	3 x 3	ReLU	18496
Conv4	64 x 256 x 256	3 x 3	ReLU	36928
Maxpool2	64 x 128 x 128	2 x 2	-	-
Conv5	128 x 128 x 128	3 x 3	ReLU	73856
Conv6	128 x 128 x 128	3 x 3	ReLU	147584
Maxpool3	128 x 64 x 64	2 x 2	-	-
Conv7	256 x 64 x 64	3 x 3	ReLU	295168
Conv8	256 x 64 x 64	3 x 3	ReLU	590080
Maxpool4	128 x 32 x 32	2 x 2	-	-
Conv9	512 x 32 x 32	3 x 3	ReLU	1180160
Conv10	512 x 32 x 32	3 x 3	ReLU	2359808
Maxpool5	512 x 16 x 16	2 x 2	-	-
Upsample1	512 x 32 x 32	2 x 2	-	-
Conv11	128 x 32 x 32	3 x 3	ReLU	589952
Upsample2	128 x 64 x 64	2 x 2	-	-
Upsample3	128 x 128 x 128	2 x 2	-	-
Dense	128 x 128 x 128	-	-	16512
Conv12	1 x 128 x 128	1 x 1	Sigmoid	1153
Upsample4	1 x 256 x 256	2 x 2	-	-
Upsample5	1 x 512 x 512	2 x 2	-	-

sum of the image $g(i,j)$ and filter $f(i,j)$.

$$f[i,j] * g[i,j] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[i-n_1, j-n_2] \tag{2}$$

Each convolutional layer in the encoder network produces a set of feature maps which are given as input to the following layers that are Rectified Linear Unit (ReLU) and 2x2 max pooling layer. ReLU is the non-linearity activation function which is applied to each element of the feature maps and squashes the negative values to zero i.e. $f(x) = max(0,x)$. It converges faster and is computationally efficient [18].

To make the representations manageable, we downsample the feature maps into pooling features/ subsamples by a factor of 2 but doubles the feature channel at each downsampling step. Max pooling operates over each activation map independently and takes those pixel values forward in the network which are prominent. Several layers of max pooling make robust classification as it achieves more translational invariance and correspondingly there is the loss of spatial resolution of the feature maps. In the decoder network, the architecture consists of the upsampling layers followed by the convolutional layer. The decoder upsamples the features maps that are achieved by the encoder network and gives sparse feature maps as an output. The features maps are

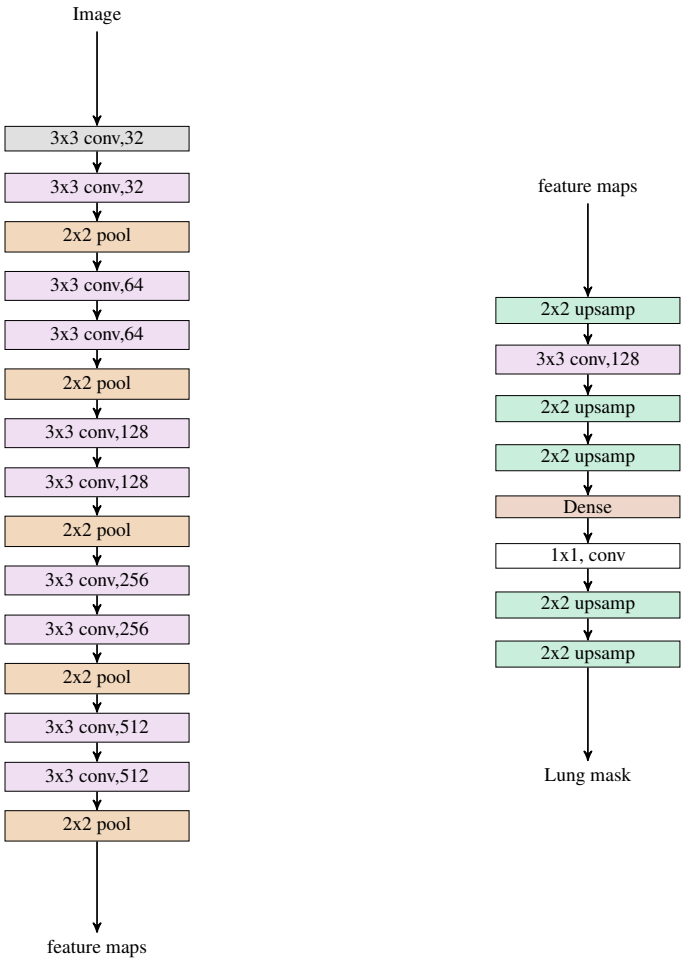

Fig. 3 Network Architecture of FS-Net.[1] **Left:** encoder network. **Right:** decoder network

connected to each and every activation unit of dense layers which are the convolved by the filter of a convolutional layer to output dense feature maps. The output of the last convolutional layer with kernel size 1x1 gives an output of 1x128x128 which is further upsampled to get the desired 512x512 lung mask.

The upsampling layers increase the resolution of images that are downsampled by the max-pooling layers. Therefore, the number of max-pooling layers and upsampling layers are kept equal. The features extracted in the encoder network are given to the upsampling layers, which process the given input and upscale its resolution by 2. The last convolutional layer of encoder network gives the output of shape 512x32x32 which

[1] Detailed Computational Graph of FS-Net is available at http://github.com/abhigoogol/FS-Net.

Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

is downsampled by max pooling to 512x16x16. In the decoder network, the first layer is an upsampling layer which upscales the output of 512x16x16 back to 512x32x32 and does the same for the rest of the layers. We have used a total of 5 upsampling layers equal to the number of max-pooling layers, to receive the size of the output at the last layer of this network, same as that of the input image. Unlike U-Net, FS-Net does not transfer the entire feature maps to the decoder so there is less memory consumption and there is no concatenation of these feature maps to the decoder feature maps of upsampling layers which in turn reduces the overall training time of the model. In the last convolutional layer of the model, the activation function used is sigmoid activation that squashes the value to the range [0, 1] and gives a probability output. The segmentation is predicted on the basis of maximum probability at each pixel.

3.3 Training, Loss and Dice coefficient

The dataset is split into 70% training set and 30% test set. The model learns from the training set and provides an unbiased evaluation of the model on the test set. The desired model has performed well on both the sets. The hyperparameters such as learning rate, batch size, and epochs are set to 0.001, 8, and 170 respectively. We have used Adam as an optimization function that converges the loss function of the model fast and gives better results by rectifying the problems like vanishing gradient, and high variance which are the causes of fluctuating loss function. The performance metrics taken for the proposed model is given by the Eq.(3) and Eq.(4).

$$\textit{Dice coefficient (Dice coeff)} = 2 * \frac{|X \cap Y|}{|X \oplus Y|} \tag{3}$$

where X is the lung area obtained by the segmentation based on our network and Y is the ground truth obtained by manual segmentation.

It is not easy to combine all the properties of the model into a single real number evaluation metric thus it is very useful to set up satisficing and optimizing metrics for the estimation of the performance of the model. Along with the optimizing metric-*Dice coeff*, we have also taken the satisficing metric -*training time* in consideration which gives us the better representation to choose the best model. For the satisficing metric we need a threshold value for the comparison and to get this value we first train the U-Net and calculate its training time. The training time taken by U-Net is considered as the threshold value and we have to compare the time taken by FS-Net for training with this value. The main aim is to maximize the Dice coeff with the subject that the training time for FS-Net should be less than the U-Net. The loss function used in this model measure the performance of the model whose output is a probability value between 0 and 1. The desired output of the function should be minimum i.e. smaller the loss more accurate is the prediction.

$$\textit{log loss} = -\frac{1}{N}\sum_{i=0}^{\infty}[y_i log(\hat{y_i}) + (1-y_i) log(1-\hat{y_i})] \tag{4}$$

where y_i is the actual output and $\hat{y_i}$ is the predicted output.

Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

4 Results

We have computed the results of the proposed FS-Net architecture and compared it with the state-of-the-art U-Net. The time taken by U-Net for training is 2 hr 26 min which is considered as the threshold value. The proposed model is considered better if it satisfies the following condition.

$$Model = Better\ \{if\ Dice\ coeff = max\ \&\ Time < threshold\}. \tag{5}$$

Table 2 Comparison results of the FS-Net and U-Net.

Architecture	Optimizer	Dice coeff	Loss	Epoch	Time	Parameters
FS-Net	Adam	0.9549	0.0498	170	51 min	7,846,081
U-Net	Adam	0.9629	0.0472	170	2 hr 26 min	5,319,265

Table 2 shows that the FS-Net and U-Net have almost the same dice coefficient and loss but FS-Net takes significantly less time to train. FS-Net is $3X^2$ times faster to train and requires less memory as compared to U-Net. The key feature to be noted from the results is that the proposed model gives the promising dice coefficient score and is computationally efficient when compared to U-Net so FS-Net is preferred as the better model with high dice score and low resource consumption.

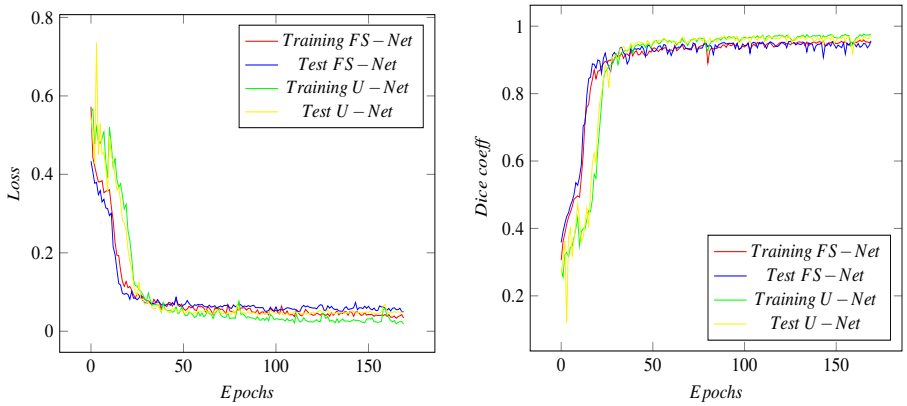

Fig. 4 FS-Net vs U-Net loss (**Left**) and Dice coeff (**Right**). The loss and dice score of the two models is almost same but time taken by FS-Net is less than U-Net.

[2] Actual training time of FS-Net is 2.863 times faster than U-Net. Model is trained on Nvidia K80.

Accepted in Soft Computing for Problem Solving (SocProS)[Proceedings in AISC], 2018

Fig. 5 shows the predictions of the lung mask $m(x,y)$ on test data and we performed pixel-wise multiplication of the predicted lungs masks with actual images $g(x,y)$ to segment the lung $s(x,y)$ in the CT data, $s(x,y) = m(x,y) * g(x,y)$.

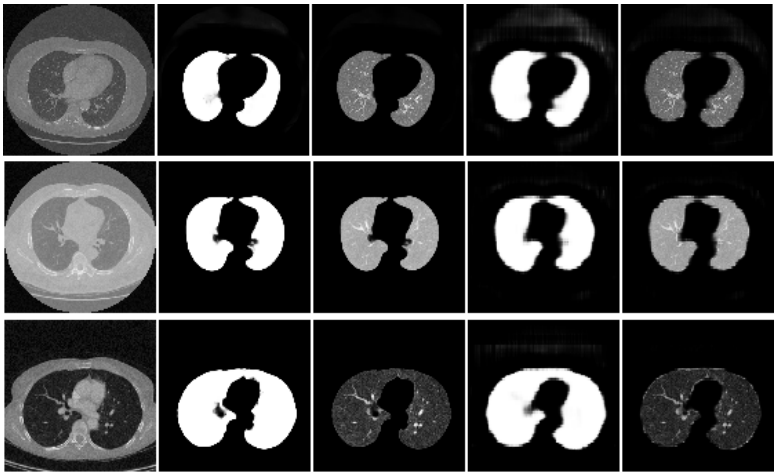

Fig. 5 Column 1 shows the original CT scans, Column 2 and 3 are the predicted lung masks and lung segments of U-Net. Column 4 and 5 shows the predicted lung masks and lung segments of FS-Net.

5 Conclusion

In our paper, we have presented a new architecture called FS-Net for the fast segmentation of lungs from CT scan images. It is a CNN based architecture which consists of the encoder network for the extraction of prominent features for the better segmentation and decoder network to retain the resolution of the images to output lung mask. FS-Net has shown a significant decrease in training time without sacrificing the dice coefficient and loss as compared to U-Net. The proposed model performs as good as the U-Net in almost one-third training time. In the future, our work is to compare our model with many other segmentation models like Segnet, FCN, Deeplabs nets, GANs and RefineNet. We will also implement ensemble learning for segmentation in the medical images where we can take different deep neural networks for segmentation and calculate their performance metrics to achieve better results.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. NIPS. pp. 1106-1114 (2012).doi:10.1145/3065386

2. Long, J., Shelhamer, E., and Darrell, T.: Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431- 3440 (2015). doi:10.1109/TPAMI.2016.2572683
3. Wang, N., Li, S., Gupta, A., Yeung, D.: Transferring rich feature hierarchies for robust visual tracking. arXiv preprint arXiv: 1501.04587 (2015).
4. Boykov, Y. Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. Computer Vision, ICCV 2001. Proceedings. Eighth IEEE International Conference on. Vol. 1. IEEE, 2001. doi:10.1109/ICCV.2001.937505.
5. Ramesh, N.,Yoo, J-H., Sethi, I. K.: Thresholding based on histogram approximation. IEE Proceedings-Vision, Image and Signal Processing 142.5 (1995): 271-279. doi:10.1049/ip-vis:19952007
6. Noviko, A. A., Lenis, D. , Major, D., Hladuvka, J. , Wimmer, M., Buhler, K.: Fully Convolutional Architectures for Multi-Class Segmentation in Chest Radiographs. IEEE Transactions on Medical Imaging. 37. 10.1109/TMI.2018.280608 (2018).
7. Dosovitskiy, A., Springenberg, J. T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. NIPS (2014)
8. Pham, D. L., Xu, C., and Prince, J. L.: Current methods in medical image segmentation. Annual Review of Biomedical Engineering, Vol.2, Issue.2000, pp.115-3, doi:10.1146/annurev.bioeng.2.1.315.
9. Huang, X.,Tsechpenakis, G.: Medical Image Segmentation. Information Discovery on Electronic Health Records, Chapter 10 (2009).
10. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convoutional Networks for Biomedical Image Segmentation. MICCAI 2015, pp. 234?241, (2015). doi:10.1007/978-3-319-24574-4_28.
11. Chen, H., Qi, X., Yu, l., Dou, Q., Qin, J., Heng, P. Deep contour-aware networks for accurate gland segmentation. Medical Image Analysis, pp.135-146, Vol.36, 2017. doi:10.1109/CVPR.2016.273.
12. Milletari, F., Navab, N., Ahmadi, S. A.: V-Net: Fully Convolutional Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision, pp.565-571 (2016). doi: 10.1109/3DV.2016.79
13. Wang G., Li, W., Ourselin, S., Vercauteren, T.: Automatic Brain Tumor Segmentation using Cascaded Anistropic Convolutional Neural Networks. BrainLes MICCAI (2017). doi:10.1007/978-3-319-75238-9_16
14. Wang, G., Li, W., Ourselin, S., Vercauteren, T., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J.: Interactive Medical Image Segmentation using Deep Learning with Image-specific Fine-tuning. IEEE Transactions on Medical Imaging(2018), vol.abs/1710.04043, 2017. doi:10.1109/TMI.2018.2791721
15. Skourt, B. A., Hassani, A., Majda, A.: Lung CT Image Segmentation Using Deep Neural Networks. The First International Conference on Intelligent Computing in Data Sciences, pp.109-113 (2018). doi: 10.1016/j.procs.2018.01.104
16. Kalinovshky, A., Kovalev, V.: Lung Image Segmentation using Deep Learning Methods and Convolutional Neural Networks. 13 International Conference on Pattern Recognition and Information Processing (2016), pp.21-24.
17. Kayalibay, B., Jensen, G., Smagt, P.: CNN-based Segmentation of Medical Imaging Data. Computing Research Repository (CoRR), vol.abs/1701.03056 (2017).
18. Nair, V., Hinton, G.E.: Rectified Linear Units Improve Restricted Boltzmann Machines. ICML'10 Proceedings of the 27th International Conference on International Conference on Machine Learning, pp. 807-814.