Automatic segmentation of kidney and liver tumors in CT images

Dina B. Efremova[1], Dmitry A. Konovalov[1], Thanongchai Siriapisith[2], Worapan Kusakunniran[3], Peter Haddawy[4]

[1] College of Science and Engineering, James Cook University, Townsville, 4811, Australia, `dmitry.konovalov@jcu.edu.au`
[2] Department Radiology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, 10700, Thailand
[3] Faculty of Information and Communication Technology, Mahidol University, Nakhonpathom, 73170, Thailand
[4] Bremen Spatial Cognition Center, University of Bremen, Germany

Abstract. Automatic segmentation of hepatic lesions in computed tomography (CT) images is a challenging task to perform due to heterogeneous, diffusive shape of tumors and complex background. To address the problem more and more researchers rely on assistance of deep convolutional neural networks (CNN) with 2D or 3D type architecture that have proven to be effective in a wide range of computer vision tasks, including medical image processing. In this technical report, we carry out research focused on more careful approach to the process of learning rather than on complex architecture of the CNN. We have chosen MICCAI 2017 LiTS dataset for training process and the public 3DIRCADb dataset for validation of our method. The proposed algorithm reached DICE score 78.8% on the 3DIRCADb dataset. The described method was then applied to the 2019 Kidney Tumor Segmentation (KiTS-2019) challenge, where our single submission achieved 96.38% for kidney and 67.38% for tumor Dice scores.

Keywords: kidney · liver · tumor · segmentation

1   Introduction

Hepatic tumor may cause a serious threat to human health and lives. To prevent and monitor liver diseases it is important to provide accurate segmentation of abnormal tissues in the organ. Although liver segmenting task has achieved good results thanks to CNN, localization of liver tumors is still a demanding problem and has some room for improvement.

CT is generally used image modality by radiologists and oncologists for liver tumor evaluation, but sometimes CT scans have noise in them due to reduction of the CT radiation dose, which is always a trade-off between image quality and health risks for patients. Other main issues in the segmentation task are the large scale of spatial and structural variability, low contrast between liver and tumor

tissues, high variation in size, shape and number of lesions, even the similarity of nearby organs.

To date, segmentation of biomedical and medical images is an active research area. In this work we adopted the delineation between methods based on how autonomous they were actually detecting liver and/or liver tumor. According to this, algorithms for segmentation fall into groups of semi-automatic and fully-automatic techniques. The methods for segmentation of CT images are reviewed next.

1.1   Related semi-automatic methods

The related studies based on semi-automatic methods are reviewed here. In all semi-automatic methods, a qualified radiologist must first locate liver and/or liver tumors manually by selecting a bounding box or other area selection.

In early attempts to perform segmentation might be used such technique as graph cut [1], when computer vision approach could not operate very deep and extensive networks, so it highly relied on complex mathematical base.

In 2005, Liu et al. [2] developed a method for segmentation of the liver contour, where a Canny edge detector was used together with a snake algorithm and a gradient vector flow (GVF) field as its external force. The method achieved a median 5.3% error by segmentation volume on 551 2D liver 512×512 images. This category of methods based on the local pixel intensity and/or gradients were actively explored [19] and demonstrated reasonable results in liver tumor segmentation even in low contrast CT images.

Siriapisith et al. have proposed a 2D segmentation method [3] that applies the concept of variable neighborhood search by iteratively alternating search through intensity and gradient spaces. They have claimed to achieve the segmentation performance with a DSC of $84.48 \pm 5.84\%$ and $76.93 \pm 8.24\%$ for large and small liver tumor respectively.

A texture-specific BoVW method [4] for the retrieval of focal liver lesions has been introduced by Xu et al. The bag of visual words (BoVW) model is meant for feature representation that can integrate various handcrafted features like intensity, texture, and spatial information and thus is able to effectively characterize various liver tumors.

Zheng et al. [5] have presented the method which combines hybrid algorithms such as a unified level set method (LSM) coupled with hidden Markov random field and expectation-maximization (HMRF-EM). The proposed LSM approach incorporates both region information and edge information to evolve the contour and it is more resistant to edge leakage than the single-information driven LSMs.

Interaction is required for the mentioned above liver and liver tumor segmentation methods, that fact restricts using those frameworks with well-qualified specialists.

1.2   Related fully-automatic methods

The major recent breakthrough in the field of semantic segmentation is widely attributed to the general-domain fully convolutional neural networks (FCNs) of [6], [7] and biomedical-domain U-net of [8]. The term *semantic* means that each pixel is assigned a label or a class of objects during training and prediction phases. For example, in this context, each image pixel could be assigned one of three labels: *liver*, *tumor*, *other*. Therefore the semantic segmentation could be conveniently defined as a per-pixel classification problem.

Since the CNN methods have been developing at accelerating rate, it is revealing to note their publication years.

In 2015, for liver tumor segmentation, Li et al. [9] achieved the Dice similarity coefficient [10] (DICE, see Eq. 3) of 80% and the area under the Precision-Recall curve of 0.9556, where the results were the averages of 30 leave-one-out cross validations on 30 CT images. Most notably it was clearly demonstrated that their CNNs outperformed other traditional machine learning methods such as AdaBoost, random forest (RF), and support vector machine (SVM). Gaussian smoothing filter was used as prepossessing, and the images were downsized by a factor of 2. Four different CNN architectures (with 6 or 7 layers) were tested, where their input shapes ranged between 13×13 and 19×19 grayscale image patches.

The CNN approach in [9] did not follow the currently most widely used segmentation CNN architectures [6], [7], [8]. The fully convolutional neural networks (FCNs): in VGG-FCN [6] and U-Net[8] CNNs, there are two distinct sections: *encoder* and *decoder*. Encoder layers or even a complete classification CNN (for example, VGG16 [11] in [6]) downsize the input image by up to 32 times while the image features are extracted. Then the decoder layers reconstruct the original input shape with the required number of segmentation layers or channels. In [9], only the encoder-type CNN was used by centering it at each input pixel.

In 2016, Dou et al. [12] developed a 3D version of the VGG-FCN [6] architecture with deep supervision to hidden layers, so-called 3D deeply supervised network (3D DSN), which could accelerate the optimization convergence rate and improve the prediction accuracy. Additionally, 3D DSN generated the high-quality score map that helped to make contour refinement with a fully connected conditional random field (CRF) to obtain refined segmentation results.

Lu et al. [13] proposed method for liver segmentation consisted of two steps: first, using 3D CNNs to detect liver and make probabilistic segmentation, and second, to refine accuracy of initial segmentation with graph cut and the previously learned probability map. The suggested approach was validated on 3DIRCADb dataset and reached 9.36% and 0.97% for volume overlap error (VOE, see Eq. 2) and relative volume difference (RVD, see Eq. 2) respectively.

In 2017, Christ et al. [14] trained two cascaded UNet-type [8] FCNs. The first FCN segmented a liver out of the rest of the inner body tissues. Then the second FCN segmented lesions from the output ROIs (regions-of-interest) of the first FCN. Dense 3D CRF was used as the post-processing to refine the FCN predictions. DICE over 94% was achieved on the 15 hepatic tumor volumes from

the abdominal CT dataset 3DIRCADb [15] for liver segmentation and 56% for lesions.

Sun et al. [16] used a segmentation CNN conceptually similar to the FCN architecture of [6], where the AlexNet [17] CNN was used as the encoder. On the 3DIRCADb dataset, Sun et al. reported the VOE of $15.6 \pm 4.3\%$. In addition to the publicly available 3DIRCADb dataset, the private JDRD dataset was labeled by two radiologists at The First Hospital of Jilin University. The unique feature of the JDRD dataset was its three CECT (contrast-enhanced computed tomography) per-slice images taken at three blood flow phases: arterial (ART), portal venous (PV), and delayed (DL) phase at the same lesion locations. When all three grayscale phase-specific images where combined into three-channel images, the studied multi-channel segmentation CNN (MC-FCN) improved VOE = $8.1 \pm 4.5\%$.

In 2018, Li et al. [18] developed a hybrid densely connected UNet (H-DenseUNet), which consists of a 2D DenseUNet and a 3D counterpart. H-DenseUNet worked in an end-to-end manner, where the intra-slice representations and inter-slice features can be jointly optimized through a hybrid feature fusion (HFF) layer for accurate liver and lesion segmentation. H-DenseUNet was trained on MICCAI 2017 Liver Tumor Segmentation (LiTS) dataset [20] and validated on the 3DIRCADb dataset achieving the liver DICE = 98.2% and tumor DICE = 93.7%. Worth noting that they also conducted experiments exclusively on 3DIRCADb dataset through cross-validation and achieved the liver segmentation DICE = 94.7% and tumor DICE = 65%.

Some recent papers have been dedicated to attention mechanism, like [21] and [22].

In 2018, Jin et al. [21] used a 3D hybrid residual attention-aware segmentation method, called RA-UNet, in their experiments. Attention modules were stacked so that the attention-aware features could change adaptively as the network went "very deep" due to residual learning. The model was trained on MICCAI 2017 LiTS dataset and validated on 3DIRCADb with DICE 97.7% and 83% for liver and lesion segmentation respectively.

In 2019, Jiang et. al [22] proposed a 3D FCN structure, composed of multiple Attention Hybrid Connection Blocks (AHCBlocks) densely connected with both long and short skip connections and soft self-attention modules. Same training process with LiTS and 3DIRCADb datasets estimated DICE 95.9% and 73.4% for liver and tumor segmentation accordingly.

At the moment a large number of solutions have been proposed for liver tumor segmentation from CT images. The fully-automatic methods have received major attention recent years, because it is meant to lift burden of segmentation from human experts and exclude human bias and mistakes.
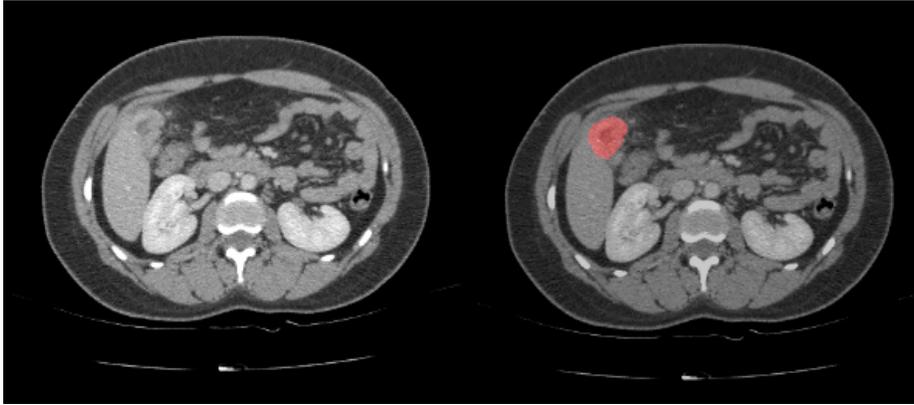
Fig. 1. An example of contrast-enhanced CT scan showing the difficult case of lesion segmentation when edges of the tumor barely distinguishable in the original image (on the left). The red region denotes the ground truth mask of liver lesion (on the right).

2   Methods

2.1   Datasets

Liver For this research two datasets have been used: MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge [20] and 3DIRCADb [15] (3D Image Reconstruction for Comparison of Algorithm Database).

The LiTS dataset contains of 201 contrast-enhanced 3D abdominal CT volumes with different types of tumor contrast levels, abnormalities in tissues size and varying amount of lesions. We have used this dataset for training our model.

The 3DIRCAD dataset includes 20 venous phase enhanced CT volumes from various European hospitals with different CT scanners involving 120 liver tumors of different sizes. We have evaluated our method on this dataset.

Expert radiologists have manually outlined liver tumor contours for all images on a slice-by-slice basis in order to determine the ground truth. The 3Dircadb dataset is segmented by a single radiologist, while the LiTS dataset is created in collaboration with seven hospitals and research institutions and manually reviewed by independent three radiologists.

Because of imbalanced classes, liver tumor areas are significantly less than background, we have applied data augmentation to the training dataset. Such techniques as elastic transformation, shifting, scaling, and rotating have been used.

Kidney The 2019 Kidney Tumor Segmentation (KiTS) Challenge [23] training dataset contained 210 different patients. The KiTS challenge required automatic segmentation of 90 test patients for which the ground truth segmentations were not released before the submission due date (29th of July, 2019).
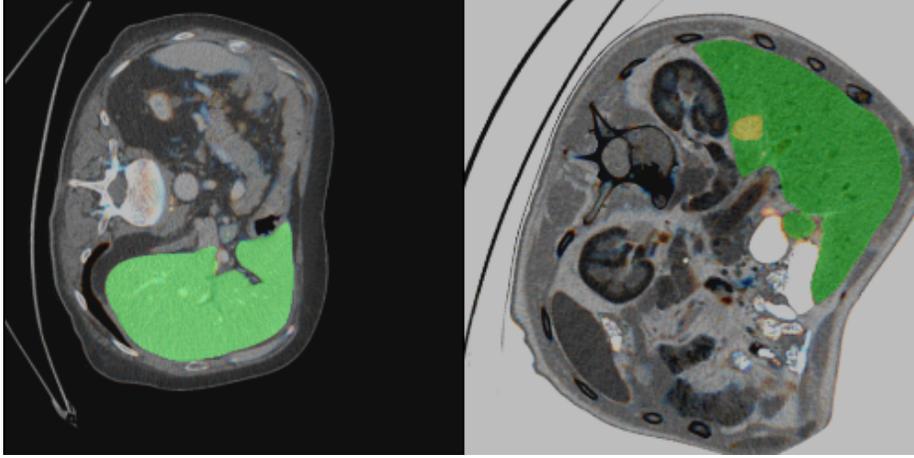
2.2   Semantic Segmentation of Images

Fig. 2. An example of the method segmentation: liver (green) and liver tumor (yellow) segmentation.

For the purpose of this paper we have selected a variation of U-Net [8], LinkNet-34 [24], where ResNet-34 [25] has been used as the feature encoder and PyTorch implementation was from [26]. LinkNet-34 has a reasonable number of parameters and a good balance between running time and accuracy.

One of the problems of deep learning with the CNN is that the learning phase, where the network undergoes training process from scratch, can be very time-consuming and may need a very large set of images. A simple yet effective transfer learning strategy is introduced to overcome several problems at once. First of all, pre-trained weights are already learnt to recognize patterns in images, the network needs less time to converge to a new solution, usually a better solution than in case of training from scratch. Transfer learning prevents or at least mitigates over-fitting problem.

In our method, we have reused the ImageNet-trained ResNet-34 encoder without freezing the weights during training process due to more advanced segmentation CNN (compared to FCN-8s). However, to assist in more effective way of using the pre-trained network, the learning rate has been reduced by factor of 10 when applied to the encoder, whereas to the randomly initialized LinkNet-34 decoder layers, the learning rate without any change has been applied.

The input layer of the LinkNet-34 model has been modified, instead of original 3-channel RGB colour we have conducted experiments with single channel and so-called 2.5D architecture [27]. Single channel is applicable due to that all CT scans are grayscale and have only one colour channel. 2.5D architecture proposed by Han [27] is a 2D deep CNN which takes a stack of adjacent

slices from the volumetric images as input and produces the segmentation map corresponding to the center slice.

As for the loss function, the binary cross entropy with negative DICE coefficient (see Eq. 1) has been used,

$$loss(y, \hat{y}) = bc(y, \hat{y}) - log(dice(y, \hat{y})), \tag{1}$$

where y is a target mask, $\hat{y}$ is the corresponding LinkNet-34 output, $bc(y, \hat{y})$ is the binary cross entropy, $dice(y, \hat{y})$ is the DICE coefficient.

Training process has been performed on the LiTS dataset consisted of 131 patients with 58,638 image-mask pairs, whereas validation on 3DIRCADb dataset has been done with two different approaches: tested on tumors larger than 100-pixel area and on tumors of any size.

3   Results

Table 1. Comparison of liver and liver tumor segmentation results on 3DIRCAB dataset.

Method	VOE(%)	RVD(%)	DICE(%)	Type
Li et al. [18]	3.57 ± 1.66	0.01 ± 0.02	98.2 ± 1	liver
H-DenseUNet	11.68 ± 4.33	-0.01 ± 0.05	93.7 ± 2	tumor
Deng et al. [29]				
3D CNN	26.93 ± 8.51	6.55 ± 14.91	85 ± 6	tumor
Jin et al. [21]	4.5	-0.1	97.7	liver
RA-UNet			83	tumor
Huang et al. [30]				
semi-automatic	27.05 ± 9.19	4.23 ± 19.28	84 ± 7	tumor
Jiang et al. [22]			95.9	liver
AHCNet	1.35	0.13	73.4	tumor
our			96.2	liver
LinkNet-34 [24]			78.8	tumor

3.1   Metrics

To evaluate performance of the segmentation task different metrics are applied, although we have focused on widely used ones that are usually utilized for liver and liver tumor segmentation such as the mean ratios of volume overlap error (VOE), relative volume difference (RVD),

$$\text{VOE} = \left[1 - \frac{|A \cap B|}{|A \cup B|} \right], \quad \text{RVD} = \frac{|A| - |B|}{|B|}, \tag{2}$$

and Dice similarity coefficient (DICE) [9], [16], [28].

$$\text{DICE} = \frac{2 \times |A \cap B|}{|A| + |B|}, \tag{3}$$

where A it the segmentation result and B is the ground truth.

3.2   Comparison to other methods

As mentioned before, we have trained the model on the LiTS dataset and evaluated it on the 3DIRCADb dataset. For comparison, we have chosen papers with the similar approach, see Table 1.

Some techniques have been more useful than others for the segmentation task. For example, data augmentation has had positive contribution to the accuracy of the method, while utilizing widely used 2.5D approach has not improved any DICE metrics. The approach with different learning rates has allowed to converge the model more quickly and to a better solution.

4   Conclusion

Our goal was to research a different approach to the segmentation task and show that the training pipeline could matter as much as the CNN architecture. While using of advanced CNN models may be constrained by hardware and the complexity of their implementation, a customized training pipeline could achieve competitive baseline results with relatively simple CNNs in fraction of time what would normally required for more complex CNNs. We deliberately selected an off-the-shelf CNN (LinkNet-34), which was not the state-of-the-art network. Consistently applying different kinds of techniques, we have reached competitive results and outperformed at least one compound CNN [22] for liver and liver tumore segmentations..

The proposed method was applied to the 2019 Kidney Tumor Segmentation Challenge [23], and the corresponding results were submitted for evaluation achieving the 38th place out of 106 submissions, where our Dice scores were 0.9638 (kidney), 0.6738 (tumor), and 0.8188 (composite, i.e. mean of kidney and tumor scores).

References

1. Boykov, Yuri, Olga Veksler and Ramin Zabih. Fast approximate energy minimization via graph cuts. Proceedings of the Seventh IEEE International Conference on Computer Vision 1 (1999): 377-384 vol.1.
2. Liu, Fan Shuo, Binsheng Zhao, Peter Klaus Kijewski, Liang Wang and Lawrence H. Schwartz. Liver segmentation for CT images using GVF snake. Medical physics 32 12 (2005): 3699-706 .

3. Siriapisith, Thanongchai, Worapan Kusakunniran and Peter Haddawy. A General Approach to Segmentation in CT Grayscale Images using Variable Neighborhood Search. 2018 Digital Image Computing: Techniques and Applications (DICTA) (2018): 1-7.
4. Xu, Yingying, Lanfen Lin, Hongjie Hu, Dan Wang, Wenchao Zhu, J. J. Wang, Xianhua Han and Yen-Wei Chen. Texture-specific bag of visual words model and spatial cone matching-based method for the retrieval of focal liver lesions using multiphase contrast-enhanced CT images. International Journal of Computer Assisted Radiology and Surgery 13 (2017): 151-164.
5. Zheng, Zhou, Xuechang Zhang, Huafei Xu, Wang Liang, Siming Zheng and Yueding Shi. A Unified Level Set Framework Combining Hybrid Algorithms for Liver and Liver Tumor Segmentation in CT Images. BioMed research international (2018).
6. Long, Jonathan, Evan Shelhamer and Trevor Darrell. Fully convolutional networks for semantic segmentation. CVPR (2015).
7. Shelhamer, Evan, Jonathan Long and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014): 3431-3440.
8. Ronneberger, Olaf, Philipp Fischer and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv abs/1505.04597 (2015): n. pag.
9. Li, Wen Jung, Fucang Jia and Qingmao Hu. Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. (2015).
10. Dice, Lee Raymond. Measures of the Amount of Ecologic Association Between Species. (1945).
11. Simonyan, Karen and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR abs/1409.1556 (2014): n. pag.
12. Dou, Qi, Hao Chen, Yueming Jin, Lequan Yu, Jing Qin and Pheng Ann Heng. 3D Deeply Supervised Network for Automatic Liver Segmentation from CT Volumes. MICCAI (2016).
13. Lu, Fang, Fa Wu, Peijun Hu, Zhiyi Peng and Dexing Kong. Automatic 3D liver location and segmentation via convolutional neural network and graph cut. International Journal of Computer Assisted Radiology and Surgery 12 (2016): 171-182.
14. Christ, Patrick Ferdinand, Florian Ettlinger, Felix Grn, Mohamed Ezzeldin A. Elshaer, Jana Lipkov, Sebastian Schlecht, Freba Ahmaddy, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Felix Hofmann, Melvin D'Anastasi, Seyed-Ahmad Ahmadi, Georgios A Kaissis, Julian Holch, Wieland H. Sommer, Rickmer F Braren, Volker Heinemann and Bjoern H. Menze. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. ArXiv abs/1702.05970 (2017): n. pag.
15. Soler, L. and others. 3D Image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database, http://www.ircad.fr/softwares/3Dircadb/3Dircadb.php?lng=en
16. Sun, Changjian, Shuxu Guo, Huimao Zhang, Jing Li, Meimei Chen, Shuzhi Ma, Lanyi Jin, Xiaoming Liu, Xueyan Li and Xiaohua Qian. Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs. Artificial intelligence in medicine 83 (2017): 58-66 ..
17. Krizhevsky, Alex, Ilya Sutskever and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. Commun. ACM 60 (2012): 84-90.
18. Li, Xiaomeng, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu and Pheng Ann Heng. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. IEEE Transactions on Medical Imaging 37 (2017): 2663-2674.

19. Li, Bing Nan, Chee Kong Chui, Stephen K. Y. Chang and Sim Heng Ong. A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images. Expert Syst. Appl. 39 (2012): 9661-9668.
20. LiTS—Liver   Tumor   Segmentation   Challenge   (2017), https://competitions.codalab.org/competitions/17094
21. Jin, Qiangguo, Zhao-Peng Meng, Changming Sun, Leyi Wei and Ran Su. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. ArXiv abs/1811.01328 (2018): n. pag.
22. Jiang, Huiyan, Tianyu Shi, Zhiqi Bai and Liangliang Huang. AHCNet: An Application of Attention Mechanism and Hybrid Connection for Liver Tumor Segmentation in CT Volumes. IEEE Access 7 (2019): 24898-24909.
23. Nicholas Heller and Niranjan Sathianathen and Arveen Kalapara and Edward Walczak and Keenan Moore and Heather Kaluzniak and Joel Rosenberg and Paul Blake and Zachary Rengel and Makinna Oestreich and Joshua Dean and Michael Tradewell and Aneri Shah and Resha Tejpaul and Zachary Edgerton and Matthew Peterson and Shaneabbas Raza and Subodh Regmi and Nikolaos Papanikolopoulos and Christopher Weight. The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes ArXiv abs/1904.00445 (2019): n. pag.
24. Chaurasia, Abhishek and Eugenio Culurciello. LinkNet: Exploiting encoder representations for efficient semantic segmentation. 2017 IEEE Visual Communications and Image Processing (VCIP) (2017): 1-4.
25. He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 770-778.
26. Shvets, Alexey A., Alexander Rakhlin, Alexandr A. Kalinin and Vladimir I. Iglovikov. Automatic Instrument Segmentation in Robot-Assisted Surgery using Deep Learning. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (2018): 624-628.
27. Han, Xiguang. Automatic Liver Lesion Segmentation Using A Deep Convolutional Neural Network Method. ArXiv abs/1704.07239 (2017): n. pag.
28. Siriapisith, Thanongchai, Worapan Kusakunniran and Peter Haddawy. Outer Wall Segmentation of Abdominal Aortic Aneurysm by Variable Neighborhood Search Through Intensity and Gradient Spaces. Journal of Digital Imaging 31 (2018): 490-504.
29. Deng, Zhuofu, Qingzhe Guo and Zhiliang Zhu. Dynamic Regulation of Level Set Parameters Using 3D Convolutional Neural Network for Liver Tumor Segmentation. Journal of healthcare engineering (2019).
30. Huang, Qing, Hui Ding, Xiaodong Wang and Guangzhi Wang. Robust extraction for low-contrast liver tumors using modified adaptive likelihood estimation. International Journal of Computer Assisted Radiology and Surgery 13 (2018): 1565-1578.