Kidney tumor segmentation using an ensembling multi-stage deep learning approach. A contribution to the KiTS19 challenge

Gianmarco Santini[1], Noémie Moreau[1] and Mathieu Rubeaux[1]

Keosys Medical Imaging, Nantes, France

Abstract. Precise characterization of the kidney and kidney tumor characteristics is of outmost importance in the context of kidney cancer treatment, especially for nephron sparing surgery which requires a precise localization of the tissues to be removed. The need for accurate and automatic delineation tools is at the origin of the KiTS19 challenge. It aims at accelerating the research and development in this field to aid prognosis and treatment planning by providing a characterized dataset of 300 CT scans to be segmented. To address the challenge, we proposed an automatic, multi-stage, 2.5D deep learning-based segmentation approach based on Residual UNet framework. An ensembling operation is added at the end to combine prediction results from previous stages reducing the variance between single models. Our neural network segmentation algorithm reaches a mean Dice score of 0.96 and 0.74 for kidney and kidney tumors, respectively on 90 unseen test cases. The results obtained are promising and could be improved by incorporating prior knowledge about the benign cysts that regularly lower the tumor segmentation results.

Keywords: CNN, Kidney tumor, Segmentation, Deep Learning

1 Introduction

Kidney cancer represents 2.4% of the cancers worldwide with more than 400 000 new cases in 2018 [1], with large variations in incidence rates based on geography, ethnicity, gender and over the time [2]. In the last two decades, the growing use of medical imaging had two effects: the renal tumor size diagnosed has consistently decreased [3], and the number of localized renal masses found incidentally, on the other hand, has raised [4].

While the standard for renal cancer treatment has been traditionally radical nephrectomy (i.e. removal of both the tumor and damaged kidney), more preservative nephronsparing surgery (i.e. open or laparoscopic partial nephrectomy)

has more than quadrupled in the last 20 years [3] and is now a treatment of choice [5].

In this context, the information related to the kidney and tumor positions, shapes and sizes, is of outmost importance for the surgery evaluation and planning process. While imaging modalities such as CT allow precise detection of the tumors, there is still a lack of automatic delineation tools, and the evaluation continues to rely on the use of nephrometry scoring systems based on manual and relatively simple imaging features [6–8].

A series of methods have been proposed in literature to perform automatic kidney segmentation in various imaging modalities. When dealing with CT, traditional methods based on imaging features, deformable models, active shape models as well as atlases have been suggested [9]. More recently the global interest towards deep learning algorithms has led to an incredible variety of applications in the medical imaging field [10], and the area of kidney segmentation is no exception [11–14].

However, most of these achievements mainly focus on the automatic kidney segmentation, generally on healthy patients, without considering the specificity of cancerous tissue if present. Indeed, the renal tumor characterization can be challenging because of its variety in terms of position, extension and gray scale values. Moreover, its distinction with benign renal cysts is not always trivial, as illustrated by the IIF (Follow-up) categorization of the reference Bosniak classification [15], where the cysts have a 5-10% risk of being a kidney cancer.

A first attempt to identify and segment kidney cancer was proposed by [16], consisting in a computer-aided method able to distinguish healthy renal tissue from cancerous one in CT scans, by using a region growing algorithm and a gray level thresholding.

More recently, Zhou et al. [17] proposed another approach where a two-step segmentation strategy was developed by employing a single atlas model to isolate kidneys as first action, and then using supervoxels to identify possible cancerous areas from the previous segmentation. However, both methods require user interactions; in the first case to initialize some seeds for the region growing, whilst in the second work the user is demanded to manually select suspect regions from the super-voxel probability maps. Moreover in [17], only cases where an evident contrastographic difference between kidneys and tumor existed were considered and succeeded in the final segmentation.

In this context, the KiTS19 challenge [18] was proposed to stimulate the research and development of trustworthy kidney and kidney tumors segmentation methods, by providing a database of 300 annotated CT scans.

In this paper we propose an automatic segmentation method based on a multi-stage 2.5D deep learning approach to address the KiTS19 MICCAI challenge on tumor kidney segmentation. The rest of the paper is organized as follows. Section 2 presents a detailed overview of the data and methods employed. Section 3 describes the configuration used for training our model and gives the results obtained. Section 4 opens a discussion on the results and presents the conclusions.

2 Methods

2.1 Dataset

We used the KiTS19 Challenge database [18]. It consists of 300 contrast enhanced CT scans, acquired in the pre-operative arterial phase and selected from a cohort of subjects who underwent a partial or a radical nephrectomy between 2010 and 2018 at the university of Minnesota. The included volumes are characterized by different in plane resolutions ranging from 0.437 to 1.04 mm, with a slice thickness varying among cases from a minimum of 0.5 mm up to 5.0 mm.

The dataset provides also for each included case the ground truth mask of both tumor tissue and healthy kidney tissue. Ground truth labels have been manually created by a pool of medical students under the supervision of an expert clinician, by using only image axial projections. A detailed description of the ground truth segmentation strategy is described in [18].

2.2 Preprocessing

Before training our multi-stage model, data were standardized to take into account differences and reduce heterogeneities between the available scans. A re-slicing operation was first performed bringing all volumes to the same slice thickness of 3 mm. This value was chosen as a compromise between the most common axial resolutions observed in the dataset.

Secondly, the training data values were bracketed in the Hounsfield Unit (HU) range of -30 HU and 300 HU. Hence it was possible to easily remove from the segmentation the fat regions surrounding the kidney, that are characterized by values smaller than -30 HU [18]. On the other hand, the higher threshold was chosen to highlight the high intensity structures and borders that generally characterize the cancerous tissue. After that, data were standardized to a zero mean and unit variance distribution of the pixel values.

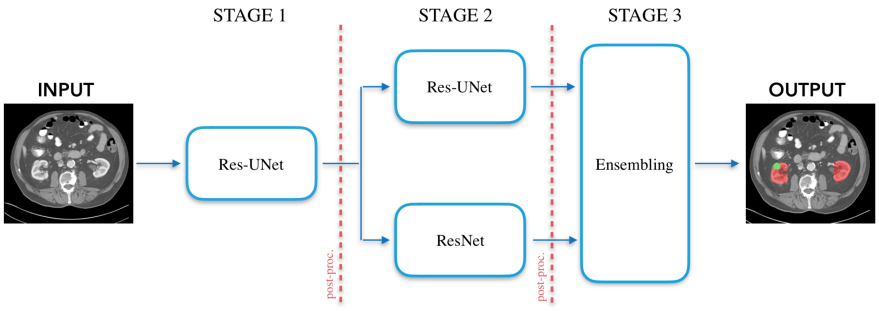

Fig. 1. Block diagram of the proposed segmentation model.

2.3 Multi-stage Deep learning approach

The proposed model is characterized by three different stages, as summarized in Figure 1. The first stage aims to roughly segment the region of interest (ROI) where to focalize the subsequent analysis, in this case the kidney region. In the second stage, the segmentation of the kidneys and cancerous tissue, is carried out by two different neural networks, which work on the image sub-portions extracted thanks to the use of the approximate kidney predictions from stage one. The results are finally combined in the last stage, where the final segmentation is obtained by using an ensembling operation. In the following, we provide details about the implementation of each stage.

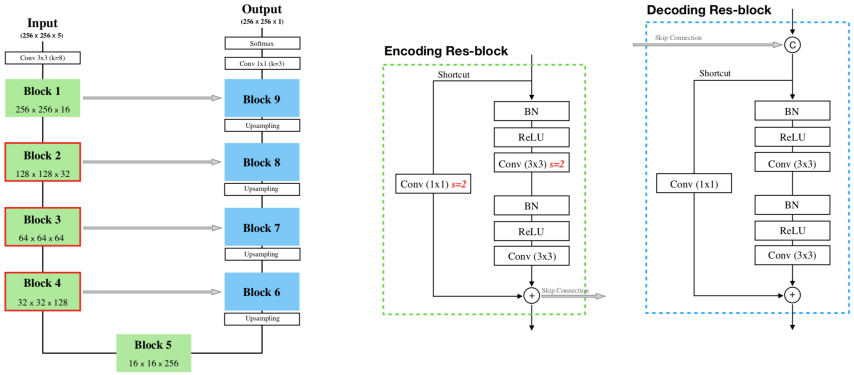

Fig. 2. Res-UNet architecture. On the picture left side, the residual block configurations. Marked in red the residual blocks where stride 2 convolutions are used. Upsampling layers precede the decoding res-blocks.

Stage 1: The initial kidney identification was performed using a Residual-UNet (or Res-UNet) designed with a standard encoder-decoder structure of a UNet [19]. Particularly, we built it with four encoding levels, as many to recreate the final prediction mask. For each level we used a pre-activated residual block [20] characterized by two convolutional layers and a shortcut connection to add the block input to its output, with the aim of speeding up the model convergence. In all the shortcut connections a convolutional layer was also added to fix some dimension differences occurring in the feature map processing. In all the residual blocks of both encoding and decoding paths, a 1×1 convolution was added to adjust the number of channels and make the addition operation possible, without any dimension mismatches. Moreover, in residual blocks two to four (highlighted in red in Figure 2) we used a 1×1 convolution with stride two to cope with the size reduction passing from a level to another. In all cases a linear activation function was adopted, in order to make the connection purpose as close as possible to the original one, proposed in [20].

To upsample the data and restore the original dimension a transposed convolution with stride two was applied before every residual block in the expanding path of the network. Skip connections were finally included to concatenate low level, but with higher resolution, feature maps from the encoder to the high level features in the decoding part. A detailed description can be seen in Figure 2.

As our purpose in the first stage was to roughly segment kidneys with cancerous tissue we tried to enhance the perception of global information rather than local ones by feeding the Res-UNet with subsampled versions of the original images, halved at 256×256 pixels in the x-y plane. Moreover, instead of a full 3D, we employed a 2.5D approach.

Using 2.5D inputs generally means to provide the network 2D slices coming from coronal, axial and sagittal planes [21–24]. In this case it consisted in passing to the model a series of adjacent axial slices stacked together along the channel dimension and make the network able to predict a mask corresponding only to the central slice.

The final prediction was carried out with a softmax function to assess whether pixels belonged to kidneys, tumor or background. Even if a three-labelled mask was produced, we merged the last two (kidneys and tumor labels) in a single meta-class and we used it in the subsequent stage to localize kidney region and extract two ROIs, i.e. one for each kidney, if present.

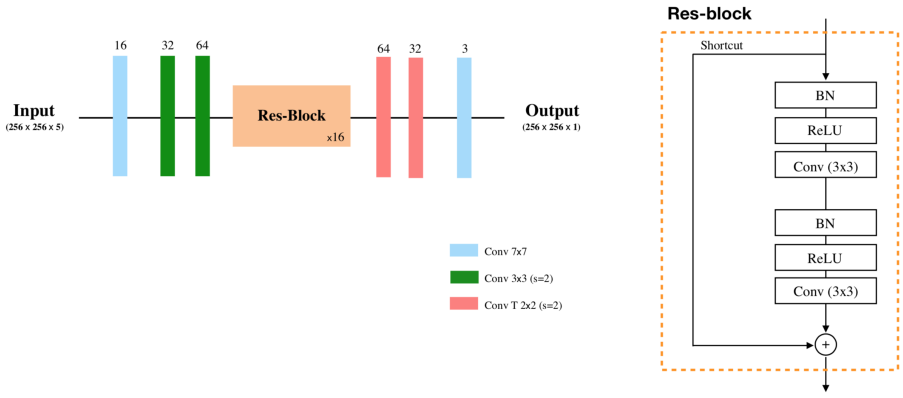

Fig. 3. Res-Net architecture. On top of each layer the number of kernels is specified.

Stage 2: As presented above, this stage was used to perform the actual segmentation of both kidney and cancerous tissue. We employed two different convolutional neural network (CNN) architectures to this aim: the first one is identical to the Res-UNet of stage one, while the second network resumes a model proposed in [25] and extended in another work [26] to solve a different segmentation task. (i.e. the SegTHOR challenge).

Compared to [26] we adopted the same network architecture (see Figure 3) but we simply modified it in order to work with 2.5D input images, that are bigger in size, and directly predict three classes with a softmax classifier. Dropout layer (p=0.5) and batch normalization with pre-activation were used throughout all the residual blocks.

The aforementioned networks were both trained trying to minimize a weighted categorical cross-entropy, with image sub-portions at the original full resolution along x-y and an interpolated slice thickness at 3 mm.

The ROIs to segment were extracted creating a bounding box to circumscribe the union of the kidney plus the tumor (or just the kidneys) by using the (x, y, z) coordinates from the first stage segmentations. After that, the bounding box was symmetrically expanded to reach the final size of 256×256 pixels. We experimentally verified that such dimension allowed to fully include the structures of interest inside the ROIs, even in case of extended tissue anomalies. Moreover, such input configuration allowed to reuse the first stage network configuration avoiding an interpolation process on the extracted images.

Stage 3: The proposed model ends with an ensembling stage, which combines prediction masks coming from the two networks of stage two. A higher segmentation accuracy is indeed demonstrated by combining more models with respect to the use of the single ones [27, 28].

2.4 Post-Processing

A simple post-processing operation was carried out at the end of both the first and second stages (red dashed lines in Figure 1). It mainly consisted in a labelling operation performed on the predicted segmentations, to identify and subsequently remove additional disconnected structures that could occur around kidneys or in other image positions. Such false positive structures were filtered away by counting the number of pixels belonging to every labeled object detected, leaving only the ones composed by more than 5000 pixels. We expected to find the kidneys with or without cancerous tissue as the largest connected structures after the two segmentation stages.

2.5 Evaluation metric

As proposed by the challenge organizers, the output image segmentation quality was assessed with the Dice score index (Eq. 1) computed on tumor and kidneys, considered as a single entity and on tumor as a standalone object. Both structures segmented with our method (S_{DL}) were compared with the ground truth segmentations (S_{GT}) provided.

$$Dice(S_{GT}, S_{DL}) = \frac{2 \cdot (S_{GT} \cap S_{DL})}{|S_{GT} + S_{DL}|} \tag{1}$$

3 Experimental settings and results

From 210 patients we received in the first round, we used 190 cases to train our model and the remaining 20 to validate it. The final 90 cases were provided by the organizers without the ground truth masks and were used to rank the methods proposed in the context of the challenge according to the average score obtained on the 90 test cases, combining the tumor and the kidney dice indexes.

3.1 Training settings

We designed our networks using Tensorflow [29] and trained them from scratch on single NVIDIA GTX 1080 of 11 GB, minimizing in all cases a weighted categorical cross-entropy, and speeding up the process using an Adam optimizer. An L2 regularization on kernel weights with 0.1 scale factor was also used.

250 epochs were fixed as upper bound limit for training each single network, but in all the conducted trials, the best model was always reached around 170 $\pm$ 30 epochs.

As anticipated above, all networks were trained with 2.5D input images resulting from the concatenation of two slices above and two slices below the current axial slice to segment.

In every training iteration, a balanced batch size of 32 samples was used. During stage one, the batch cases were balanced according to three image group types (roughly 33% each) characterized as follows. *Group B*: images in which neither kidneys nor tumors appear, *group K*: images with healthy kidney portions, *group KT*: images with kidney and tumor tissue. For stage two, 50% of cases belonged to *group K*, while the remaining 50% was filled with cases from *group KT*.

In order to prevent overfitting on the tumor segmentation, different data augmentation strategies such as axial rotation (angle $\in [-30°, 30°]$), horizontal flip, and central crop plus zoom were also employed on-the-fly on KT cases only, as they appear in the training dataset less frequently and with higher variability than the K cases.

The use of diverse combinations of initializations, training settings and network architectures generally help subsequent ensembling operations [27,28]. For this reason, we carried out different training configurations (especially for stage 2) in order to reduce the generalization error of the final prediction. Table 1 synthesizes different training configurations adopted for the networks employed.

3.2 Evaluation results

With the above configurations we obtained the results expressed in Table 2, on the 20 cases chosen randomly as validation set. We finally used the trained ensemble model to perform the prediction on the 90 test cases. We obtained a dice score of 0.96 and 0.74 for kidneys and tumor respectively. Segmentation image examples follow.

Table 1. Training configurations and data augmentation strategies employed for each network. Data augmentation operations refer only to input images presenting tumors (group KT). T.N. stands for Truncated Normal weights initialization.

	Data Augmentation			**Training Settings**		
	Rotation ($p = 1.0$)	H-Flip ($p = 0.5$)	Central crop ($p = 0.66$)	Weights Init	Loss weights [B, K, KT]	Learning rate
Res-Unet1	•			T.N. (*std* 0.1)	[0.3 1.0 3.0]	10^{-4}
Res-Unet2	•	•		T.N. (*std* 0.1)	[0.3 1.0 3.0]	10^{-4}
Res-Net	•	•	•	He uniform	[0.2 0.25 0.55]	10^{-3}

Table 2. Segmentation results for each stage designed, considering kidneys plus tumor (first column) and tumor alone (second column) on the validation set.

	Dice kidneys	**Dice tumor**
Res-Unet1	0.96 ± 0.02	0.52 ± 0.32
Res-Unet2	0.97 ± 0.02	0.70 ± 0.28
Res-Net	0.97 ± 0.02	0.72 ± 0.26
Ensembling	$\mathbf{0.98 \pm 0.01}$	$\mathbf{0.73 \pm 0.25}$

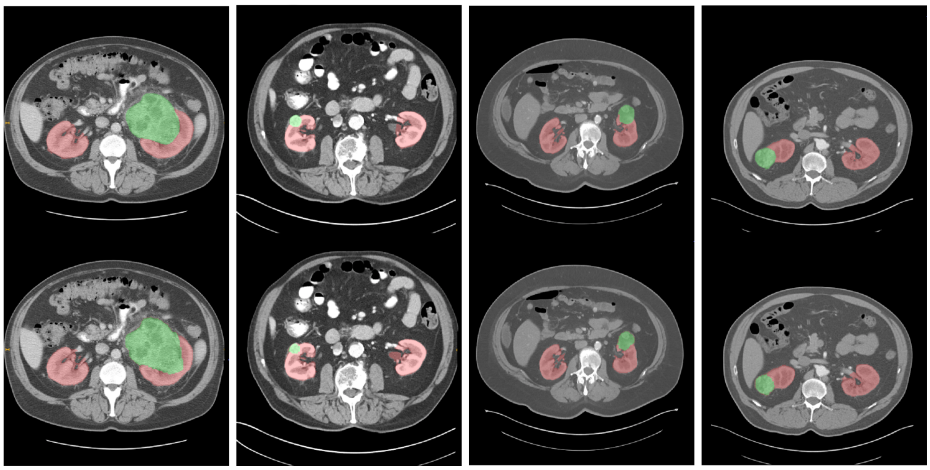

Fig. 4. Segmentation examples from four patients. On the top row the ground truth labels are reported (red for kidneys, green for tumors). The proposed model predictions are on the second row. Best viewed in color.

4 Discussion and conclusion

We presented an automatic method for semantic segmentation of kidneys and kidney cancerous tissue from contrastographic CT acquisitions.

As shown in Table 2, the use of a multi-step segmentation algorithm allowed us to obtain better results than using only a single segmenting step. Moreover, the ensembling operation of different CNN predictions proved to give slightly better results, compared to the individual network performances.

On the other hand, the use of an ensembling training strategy combined with a multi-stage segmentation process led to an increase in training time, due to the need to try more networks with different configuration settings. A single network took indeed around three days of training with our hardware configuration.

The use of a relatively large batch size allowed to better exploit batch normalization properties compared to smaller batches previously tried (8 or 16). Among the evaluations, the results obtained with a batch size of 32 samples were better. However, this choice penalized the use of state-of-the-art neural network architectures like Tiramisu [30] or DeepLab [31], which couldn't be trained with such input sizes, essentially due to memory constraints.

Finally, using 2.5D input tensor helped to provide some kind of volumetric information to the network resulting in a better segmentation level compared to simple 2D inputs.

Considering the final performance obtained from the 90 test patients, we can assert that the overall segmentation results, especially concerning the tumor identification remain promising, but are quite low as a consequence of false positive cases that were sometimes detected on kidney cysts, or false negative cases where the tumor lesion was not east to identify on the CT. For future improvements we plan to design a new training strategy, in which more of these specific cases could be passed to the model, in order to differentiate them from cases where cancerous tissues actually occur. Furthermore, the we would like to work on even more focused input data, by maybe adding an additional stage in order to focus on the analysis of even more detailed local features.

Acknowledgments

This work is financed by FEDER funds through "Programme opérationnel regional FEDER-FSE Pays de la Loire 2014-2020" n° PL0015129 (EPICURE).

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **68**(6) (2018) 394–424
2. Scelo, G., Larose, T.L.: Epidemiology and risk factors for kidney cancer. Journal of Clinical Oncology **36**(36) (2018) 3574–3581

3. Nguyen, M.M., Gill, I.S., Ellison, L.M.: The evolving presentation of renal carcinoma in the united states: trends from the surveillance, epidemiology, and end results program. The Journal of urology **176**(6) (2006) 2397–2400
4. Sun, M., Abdollah, F., Bianchi, M., Trinh, Q.D., Jeldres, C., Thuret, R., Tian, Z., Shariat, S.F., Montorsi, F., Perrotte, P., et al.: Treatment management of small renal masses in the 21st century: a paradigm shift. Annals of surgical oncology **19**(7) (2012) 2380–2387
5. Dominguez-Escrig, J.L., Vasdev, N., O'Riordon, A., Soomro, N.: Laparoscopic partial nephrectomy: Technical considerations and an update. Journal of minimal access surgery **7**(4) (2011) 205
6. Ficarra, V., Novara, G., Secco, S., Macchi, V., Porzionato, A., De Caro, R., Artibani, W.: Preoperative aspects and dimensions used for an anatomical (padua) classification of renal tumours in patients who are candidates for nephron-sparing surgery. European urology **56**(5) (2009) 786–793
7. Kutikov, A., Uzzo, R.G.: The renal nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. The Journal of urology **182**(3) (2009) 844–853
8. Simmons, M.N., Ching, C.B., Samplaski, M.K., Park, C.H., Gill, I.S.: Kidney tumor location measurement using the c index method. The Journal of urology **183**(5) (2010) 1708–1713
9. Torres, H.R., Queiros, S., Morais, P., Oliveira, B., Fonseca, J.C., Vilaca, J.L.: Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: A systematic review. Computer methods and programs in biomedicine **157** (2018) 49–67
10. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42** (2017) 60–88
11. Thong, W., Kadoury, S., Piché, N., Pal, C.J.: Convolutional networks for kidney segmentation in contrast-enhanced ct scans. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **6**(3) (2018) 277–282
12. Zheng, Y., Liu, D., Georgescu, B., Xu, D., Comaniciu, D.: Deep learning based automatic segmentation of pathological kidney in ct: local versus global image context. In: Deep Learning and Convolutional Neural Networks for Medical Image Computing. Springer (2017) 241–255
13. Jackson, P., Hardcastle, N., Dawe, N., Kron, T., Hofman, M., Hicks, R.J.: Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy. Frontiers in oncology **8** (2018) 215
14. Sharma, K., Rupprecht, C., Caroli, A., Aparicio, M.C., Remuzzi, A., Baust, M., Navab, N.: Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. Scientific reports **7**(1) (2017) 2049
15. Bosniak, M.A.: The current radiological approach to renal cysts. Radiology **158**(1) (1986) 1–10
16. Kim, D.Y., Park, J.W.: Computer-aided detection of kidney tumor on abdominal computed tomography scans. Acta radiologica **45**(7) (2004) 791–795
17. Zhou, B., Chen, L.: Atlas-based semi-automatic kidney tumor detection and segmentation in ct images. In: 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE (2016) 1397–1401

18. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer (2015) 234–241
20. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision, Springer (2016) 630–645
21. Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M.: A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In: International conference on medical image computing and computer-assisted intervention, Springer (2014) 520–527
22. Wolterink, J.M., Leiner, T., de Vos, B.D., van Hamersvelt, R.W., Viergever, M.A., Išgum, I.: Automatic coronary artery calcium scoring in cardiac ct angiography using paired convolutional neural networks. Medical image analysis **34** (2016) 123–136
23. Setio, A.A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., Van Riel, S.J., Wille, M.M.W., Naqibullah, M., Sánchez, C.I., van Ginneken, B.: Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. IEEE transactions on medical imaging **35**(5) (2016) 1160–1169
24. Roth, H.R., Wang, Y., Yao, J., Lu, L., Burns, J.E., Summers, R.M.: Deep convolutional networks for automated detection of posterior-element fractures on spine ct. In: Medical Imaging 2016: Computer-Aided Diagnosis. Volume 9785., International Society for Optics and Photonics (2016) 97850P
25. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision, Springer (2016) 694–711
26. van Harten, L., Noothout, J.M., Verhoeff, J., Wolterink, J.M., Isgum, I.: Automatic segmentation of organs at risk in thoracic ct scans by combining 2d and 3d convolutional neural networks. In: SegTHOR@ ISBI. (2019)
27. Lyksborg, M., Puonti, O., Agn, M., Larsen, R.: An ensemble of 2d convolutional neural networks for tumor segmentation. In: Scandinavian Conference on Image Analysis, Springer (2015) 201–211
28. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: International MICCAI Brainlesion Workshop, Springer (2017) 450–462
29. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). (2016) 265–283
30. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (2017) 11–19
31. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4) (2017) 834–848