

A Unet-based Research on the Multi-Output Convolution Neural Network's Ability of Decreasing Mis-identification: Automatic Segmentation of Organs at Risk in Thorax

Jie Zhang

Zhejiang Cancer Hospital <https://orcid.org/0000-0002-0817-0464>

Yiwei Yang

Institute of Cancer and Medicine (ICBM), Chinese Academy of Sciences

Kainan Shao

Institute of Cancer and Medicine (ICBM), Chinese Academy of Sciences

Xue Bai

Institute of Cancer and Medicine (ICBM), Chinese Academy of Sciences

Min Fang

Institute of Cancer and Medicine (ICBM), Chinese Academy of Sciences

Guoping Shan

Institute of Cancer and Medicine (ICBM), Chinese Academy of Sciences

Ming Chen (✉ chenming@zjcc.org.cn)

Institute of Cancer and Medicine (ICBM), Chinese Academy of Sciences

Research

Keywords: multi-output convolution neural network, mis-identification, auto-segmentation, organs at risk, thoracic tumor

Posted Date: April 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-23945/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

To study a multi-output convolution neural network (CNN)'s capability of reducing mis-identification.

Material and Methods

To guarantee that the CNN's output number was the only experiment variable, we used Unet as research object. By modifying it into a multi-output (MO) one, we got a MO-Unet and the conventional single-output Unet (SO-Unet) as a comparing object. All images involved in this study were computed tomography (CT) scans coming from 105 patients with thoracic tumor. 3 organs at risk (OARs), i.e. lung, heart and spinal cord, were delineated by experienced radiation oncologists and were used as ground truth. The two models were both trained with 1240 CTs (856 images for learning and 384 images for monitor) and under the same learning settings. They were both tested on other 886 images. Dice and mis-identification pixels' number(n) were 2 metrics for evaluation.

Results

MO-Unet and SO-Unet achieved Dice of 0.9400 ± 0.0612 (average \pm standard deviation) and 0.9451 ± 0.0618 for lung, 0.9143 ± 0.1119 and 0.9160 ± 0.1071 for heart, 0.8988 ± 0.0657 and 0.9020 ± 0.0624 for spinal cord respectively. The two models' all average Dices were ≤ 0.005 . For the normalized number of cases with $n = 0$, MO-Unet and SO-Unet had 97.29% and 96.84% for spinal cord, 88.49% and 90.86% for heart, 81.26% and 77.09% for lung respectively. Compared to SO-Unet, the mis-identification cases of MO-Unet mainly felled in the range of small n .

Conclusions

The Dice results showed that the two models had comparable overlap. The n results suggested that the MO-Unet was better in decreasing mis-identification. Besides, a MO network is light-weighted to implement more delineation under the same computing source. Therefore, a MO network is promising in segmenting OARs and has the potential for a widespread application in China.

1. Background

Delineating organs at risk (OARs) in computed tomography (CT) images is an essential step for optimizing and quantitatively evaluating a radiation treatment plan[1–3]. Manual delineation is the main option in clinic, but time-consuming. Although the computer-assisted approaches (e.g. atlas-based algorithm[4–7]) have come to help lessen such burden, they can't guarantee the segmentation accuracy and reproducibility[8–10].

In recent years, with the development of convolution neural network (CNN) in the field of image, more and more networks[11–17] appear to segment OARs in CT images automatically and show good results. Feng X et al.[12] used three-dimension (3D) Unets to locate thoracic OARs and then segment them. They reported mean Dice of 0.89, 0.97, 0.93 and average 95% Hausdorff distance of 1.89 mm, 4 mm, 2.10 mm for spinal cord, lung and heart respectively. Tao He et al.[11] proposed a U-like network trained under a multi-task learning scheme. The major task was segmentation. The auxiliary task was global slice classification under the hypothesis of OARs concurrently appearing in similar slice orders for most patients. This U-like network reached heart Dice of 0.95. Besides, a 5-channel CNN with multiple images of highlighting different tissues achieved average heart, spinal cord and lung Dice of 0.91, 0.76 and 0.95[16]. A Unet-GAN[17] attained mean Dice of 0.85, 0.96 ~ 0.97 and 0.88 for heart, lung and spinal cord.

Among the above networks, Unet[18] is a classic one for good image segmentation. When we trained a two-dimension(2D) Unet to delineate thoracic OARs, a few exceptions with mis-identification as shown in Fig. 1 appeared. In this figure, some pixels are wrongly categorized into OARs. This phenomenon may be caused by the CNN principle. CNN conducts each pixel's classification as a separate task and only based on the gray distribution of a small-size image (i.e. receptive field). There is no prior knowledge, such as shape or identification of neighboring organ, involved in the classification[19]. Therefore, it may give a wrong classification when different organs, in a receptive field, shows similar gray value. As shown in Fig. 2, the lung pixels in the heart pixel's receptive field (r_{heart}) and the vitro air pixels in the arm pixel's receptive field (r_{arm}) both exhibit 0 grayscale. Besides, the heart pixels and the arm pixels show similar image intensities. Without knowing the 0 grayscales in r_{heart} and in r_{arm} belong to lung and vitro air respectively, it is a high likelihood that CNN wrongly classified the pixel B into heart.

To decrease the above mis-identification, a possible solution is to label various organs by using different numbers. Given that the common activation function of rectified linear unit (ReLU)[20] has no upper limit of its output and hence it may map an image intensity to any positive number, we can't use a preassigned number to represent an organ. Therefore, we built a multi-output (MO) network whose different outputs correspond to different organs. In this way, the network could learn an optimal number to represent a certain organ. The design of MO network is also known as multi-task[21, 22] (MT) or multi-label[12, 23] (ML) learning. They have been reported in some papers[12, 21, 23], but most of them focused on the overlap between model result and ground truth, less investigated their performances in reducing mis-identification.

To verify the above hypothesis, we modified a classic 2D Unet into a multi-output one (abbreviated as MO-Unet) and trained it for segmentation of three thoracic OARs (i.e. lung, heart and spinal cord). Then we compared its performances with a single output 2D Unet (abbreviated as SO-Unet). The rest of this paper is organized as follows. The results are shown and discussed in Sect. 2 and 3 respectively. Our conclusion is presented in Sect. 4. Section 5 gives the detailed architecture of MO-Unet and introduces our experiments.

2. Results

Table 1 summarized the statistics on the Dice of test set. The proposed MO-Unet achieved lung Dice of 0.9400 ± 0.0612 (average \pm standard deviation), heart Dice of 0.9143 ± 0.1119 and spinal cord Dice of 0.8988 ± 0.0657 . The SO-Unet segmented lung, heart and spinal cord with Dice of 0.9451 ± 0.0618 , 0.9160 ± 0.1071 and 0.9020 ± 0.0624 respectively. The 10th percentile suggested that, for MO-Unet, 90% of data reached lung Dice of ≥ 0.8841 , heart Dice of ≥ 0.8741 and spinal cord Dice of ≥ 0.8343 . For SO-Unet, 90% of data showed lung Dice of ≥ 0.9007 , heart Dice of ≥ 0.8654 and spinal cord Dice of ≥ 0.8431 . They suggested that the Dices achieved by MO-Unet were comparable to SO-Unet.

Table 1. Statistics on the Dice of test set*

	lung		heart		spinal cord	
	ours	SO-Unet	ours	SO-Unet	ours	SO-Unet
average	0.9400	0.9451	0.9143	0.9160	0.8988	0.9020
SD	0.0612	0.0618	0.1119	0.1071	0.0657	0.0624
5th percentile	0.8295	0.8374	0.7449	0.7682	0.7963	0.8026
10th percentile	0.8841	0.9007	0.8741	0.8654	0.8343	0.8431
20th percentile	0.9276	0.9360	0.9112	0.9102	0.8669	0.8758

*SD: standard deviation; α -th percentile: the value above which $(100-\alpha)\%$ of the data fall; the relatively better results were printed in bold.

Fig 3 plotted the cumulative distribution histogram of $n(T)$ for 3 OARs. For MO-Unet, there were 97.29%, 88.49% and 81.26% cases with $n(T)=0$ for spinal cord, heart and lung respectively. For SO-Unet, there were 96.84%, 90.86% and 77.09% cases with $n(T)=0$ for spinal cord, heart and lung respectively. When segmenting spinal cord (as shown in Fig 3A), MO-Unet had all cases with $n(5) \leq 22$; whereas SO-Unet had 98.87% cases with $n(5) \leq 22$. When segmenting heart (as shown in Fig 3B), in the $n(15)$ range of 53~401 which was denoted as yellow area, MO-Unet had a higher slope than SO-Unet. While in the $n(15)$ range of 401~948 which was denoted as blue area, the slope of SO-Unet is greater than MO-Unet. Fig 3C showed the result of lung delineation. In this figure, MO-Unet had a lower gradient when $n(20)$ ranged from 44 to 726.

3. Discussion

3.1 Discussion on the experimental results

It has attracted more and more interest to reduce duplicate clinical work by using CNN to segment OARs automatically. Along with its convenience, the security is also supposed to gain attention. Given that OAR delineation is used to optimize and evaluate a radiation treatment plan quantitatively, mis-identification is a non-negligible factor, especially for a serial organ, such as spinal cord. When any subunit of a serial

organ is irradiated by a dose of more than tolerance, the entire organ would fail. Thus, the maximum point dose received by a serial organ is what we try to lower when optimizing a radiation treatment plan. Under such circumstances, when delineating a spinal cord, if there is only one mis-identified pixel and the pixel falls inside planning target volume (PTV), it may result in an impossible treatment plan, because of the impossible tradeoff between lowering the spinal cord's maximum point dose and increasing PTV dose. Even that, the one mis-identified pixel has negligible influences on the metrics of Dice or 95% Hausdorff distance.

To the best of our knowledge, less papers reported the study of CNN's mis-identification. Given that the segmentation by CNN is actually a pixel-wise classification, some deviation-related metrics[24], such as 95% Hausdorff distance[15, 25], the average Hausdorff distance[26] and mean surface distance[15, 27–29], can't perform a comprehensive evaluation on the mis-identification. How many wrong classifications, i.e. n in this paper, conducted by a network can help us learn more about its performance.

In this work, we give a possible explanation for the wrong identification conducted by CNN, and propose a multi-output architecture as a potential solution. For validation, we adopted a classic segmentation network, namely Unet, as our research object.

The Dice statistics summarized in Table 1 suggest that MO-Unet and SO-Unet have comparable overlap. Their differences of average Dice for 3 OARs are ≤ 0.005 . Their bias of 20th percentiles of Dices for 3 OARs are < 0.009 . 3 points in Fig. 3 shows MO-Unet's preferable advantage in reducing mis-identification: (i) MO-Unet has a fewer mis-identification cases (i.e. $n(T) \neq 0$) than SO-Unet, except the heart delineation; (ii) in Fig. 3A, MO-Unet has 100% cases encompassing $n(5)$ of ≤ 22 , while some cases of SO-Unet shows $n(5)$ of > 22 ; (iii) in Fig. 3B and C, the higher slope of MO-Unet falls in the range of small $n(T)$ compared to the higher gradient of SO-Unet. It means that $n(T)$ of MO-Unet is generally smaller than SO-Unet. Above all, a multi-output architecture is a promising way to reduce mis-identification.

Additionally, a MO network can segment OARs simultaneously. Therefore, compared to a SO one, it lowers the computing source when performing the same amount of delineation work. Consequently, a MO one seems to be a better choice for a widespread application in China, since most hospitals in China can't afford a high configuration computing server.

3.2 Future work

This work focused on the potential of reducing mis-identification by using a MO architecture. Except for this method, a three-dimension (3D) network may contribute to it too, as it provides 3D features. In the future work, we will further investigate a 3D network's capability of decreasing mis-identification and compare it with a MO one.

4. Conclusion

In the study, we proposed a multi-output architecture as a potential solution to decrease mis-identification. We modified a classic CNN in image segmentation, i.e. Unet, into a multi-output one (MO-

Unet) and compare its performance with conventional Unet (SO-Unet) under the same learning settings. 2 metrics was adopted in our work: Dice and mis-identified pixels' number. The results showed that MO-Unet was able to achieve similar statistical Dice as SO-Unet and performed better in decreasing mis-identification. Besides, compared to SO-Unet, MO-Unet is a light-weighted network to implement the same segmentation workload. In conclusion, a multi-output network has the potential to segment OARs with high accuracy and low mis-identification. It is also a promising way for a broad application.

5. Method

5.1 Networks

The SO-Unet is an open-source network[30]. Its detailed architecture was shown in Fig. 4(a). Compared to the original Unet proposed by Olaf Ronneberger et al.[18], the filter number in each layer was reduced by half. It was constrained by the computational ability of hardware. In the expansive path, the transposed convolution[31] was adopted, instead of an upsampling of the feature map followed by a 2×2 convolution. All convolutions in this network were padded convolutions to guarantee the same size between input and output. We didn't adopt the overlap-tile strategy reported by Olaf Ronneberger et al.[18], because the input image in our work was different with theirs. In the work of Olaf Ronneberger et al., the input image was cell image whose border pixels had non-zero gray values. By using padded convolutions on these images, other zeros would be added to the input image border. It may change the grayscale distribution of the image border's receptive field and hence decreases the segmentation accuracy along the border. Differently, our input image was a thoracic CT whose border pixels usually equaled to zero (namely vitro air). Thus, the zero-padding operation wouldn't lead to the change of grayscale distribution in the receptive field of image border. The last convolution used activation function of sigmoid to achieve binary classification. All other convolutions used activation function of ReLU.

According to the possible explanation about mis-identification stated in the section of introduction, we added another two output branches to the last hidden layer of SO-Unet to constitute MO-Unet (as exhibited in Fig. 3(b)). That is, one MO-Unet produced three OARs' segmentations simultaneously. Three SO-Unets outputted three OARs' delineations respectively. In this way, the MO architecture is the only variable in our investigation. Consequently, the experiment results can help us assess whether the MO architecture can reduce mis-identification. In the meantime, we let the numbers representing different organs are learnable.

5.2 Experiments

5.2.1 Data acquisition and preprocessing

There were totally 105 patients with tumor in thoracic region enrolled in this experiment. All of them received CT scans by using a Light Speed (GE Healthcare, Chicago, America) or a Brilliance CT Big Bore system (Philips Healthcare, Best, the Netherlands). All CT images were reconstructed using a thickness of 5 mm and a matrix size of 512×512 with a resolution of approximately 1 mm. Among the acquired CT

images, only 2126 images that encompassed lung, heart and spinal cord were involved. All OARs were delineated by experienced radiation oncologists and were regarded as ground truth.

In each CT image, we linearly converted the stored pixel value ranging of -135 ~ 215 HU into image intensity of 0 ~ 255. To save computing sources, all images were cut into 512 × 256 to remove unnecessary external air information. It is achieved by using a threshold segmentation (threshold value = 0) and border following algorithm[32] to detect the foreground edge, and then cutting the image into 512 × 256 based on the foreground center.

5.2.2 Network training

Among all images, 856 images were in the training set and 384 images were in the validation set. The rest 886 images were used for test. During training, the images in the training set were used to optimize network. Those in the validation set were used to monitor the whole process of training. The test set were used for evaluation and comparison.

In our experiment, MO-Unet and SO-Unet were both trained using adaptive moment estimation[33](Adam) and the same parameters. The learning rate was 10^{-3} , batch size equaled to 16 and training epochs were 3000. The only difference was loss function (L). L for SO-Unet was one OAR Dice as shown in Eq. (2). L for MO-Unet was the average of three OARs' Dices as shown in Eq. (3).

$$Dice = \frac{X \cap Y}{2(X \cup Y)} \quad (1)$$

$$L_{SO-Unet} = \frac{1}{N} \sum_{i=1}^N Dice_i \quad (2)$$

$$L_{MO-Unet} = \frac{1}{N} \sum_{i=1}^N (Dice_i^{lung} + Dice_i^{heart} + Dice_i^{spinal\ cord}) \quad (3)$$

in which Dice was a metric to measure the similarity between model output X and ground truth Y . Dice ranged from 0 to 1. 1 represented the perfect similarity of X and Y . 0 meant that they had no overlay. $L_{SO-Unet}$ and $L_{MO-Unet}$ were the loss functions of SO-Unet and MO-Unet respectively. N was the number of CT images in training set. In Eq. (3), the 3 superscripts of Dice corresponded to 3 OARs respectively.

5.2.3 Evaluation

To compare the performances of MO-Unet and SO-Unet, 2 metrics were adopted: Dice and mis-identified number (n). They were measured in 2D.

Dice was a metric of measuring the spatial overlap between two sets of binary segmentations, as defined in Eq. (1).

n was the number of mis-identified pixels. A pixel ($x \in X$) was categorized as a mis-identified one when its minimum distance ($d(x)$) to any pixels in Y was greater than a threshold distance:

$$d(\mathbf{x}) = \min_{\mathbf{y} \in Y} \|\mathbf{x} - \mathbf{y}\| \quad (4)$$

$$n(T) = \#_{\mathbf{x} \in X} (d(\mathbf{x}) > T) \quad (5)$$

in which $\|\cdot\|$ meant the Euclidean distance. $\#$ represented number. T (unit: pixel) was the threshold distance. In our work, $T = 20$ for lung, $T = 15$ for heart and $T = 5$ for spinal cord. X and Y had the same denotations in Eq. (1).

T was set to avoid the influence of the non-pixel level delineation by human and the variance of inter-observer. More details about T can be found in the Supplementary file.

Abbreviations

OAR	organ at risk
CT	computed tomography
CNN	convolution neural network
3D	three-dimension
2D	two-dimension
ReLU	Rectified linear unit
MO	multi-output
SO	single-output
MT	multi-task
ML	multi-label
Adam	adaptive moment estimation

Declarations

- **Ethics approval and consent to participate**

The experimental protocol was established, according to the ethical guidelines of the Helsinki Declaration and was approved by the Human Ethics Committee of Zhejiang Cancer Hospital. Written informed consent was obtained from individual or guardian participants.

- **Consent for publication**

Not applicable

- **Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

- **Competing interests**

The authors declare that they have no competing interests.

- **Funding**

This work was supported by National Key Research and Development Program (2017YFC0113201), Youth Talent Foundation of Zhejiang Medical and Health Project (2019RC023), Zhejiang Key R&D Program (2019C03003), Zhejiang Natural Science Foundation (LQ20H180016) and Appropriate Technology Cultivation and Promotion of Zhejiang Medical and Health Project (2019ZH018).

- **Authors' contributions**

Jie Zhang contributed to the coding and wrote the manuscript. Yiwei Yang, Kainan Shao and Xue Bai performed the data collection. Min Fang delineated and confirmed the manual contour of organs at risk. Guoping Shan reviewed this manuscript. Ming Chen contributed to the conception of the study.

- **Acknowledgements**

Not applicable

References

1. Ezzell GA, Galvin JM, Low D, Palta JR, Rosen I, Sharpe MB, Xia P, Xiao Y, Xing L, Cedric XY: **Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee.** *Medical physics* 2003, **30**(8):2089-2115.
2. Mackie TR, Kapatoes J, Ruchala K, Lu W, Wu C, Olivera G, Forrest L, Tome W, Welsh J, Jeraj R: **Image guidance for precise conformal radiotherapy.** *International Journal of Radiation Oncology Biology Physics* 2003, **56**(1):89-105.
3. Mukesh M, Benson R, Jena R, Hoole A, Roques T, Scrase C, Martin C, Whitfield G, Gemmill J, Jefferies S: **Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help?** *The British journal of radiology* 2012, **85**(1016):e530-e536.
4. Fritscher KD, Peroni M, Zaffino P, Spadea MF, Schubert R, Sharp G: **Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours.** *Medical physics* 2014, **41**(5):051910.

5. Wachinger C, Fritscher K, Sharp G, Golland P: **Contour-driven atlas-based segmentation.** *IEEE transactions on medical imaging* 2015, **34**(12):2492-2505.
6. Ciardo D, Gerardi MA, Vigorito S, Morra A, Dell'Acqua V, Diaz FJ, Cattani F, Zaffino P, Ricotti R, Spadea MF: **Atlas-based segmentation in breast cancer radiotherapy: evaluation of specific and generic-purpose atlases.** *The Breast* 2017, **32**:44-52.
7. Van Elmpt W, Van der Stoep J, Van Soest J, Lustberg T, Gooding M, Dekker A: **EP-1225: Atlas-based segmentation reduces interobserver variation and delineation time for OAR in NSCLC.** *Radiotherapy and Oncology* 2017, **123**:S661.
8. Gomathi V, Karthikeyan S: **A Proposed Hybrid Medoid Shift with K-means (HMSK) segmentation algorithm to detect tumor and organs for effective radiotherapy.** In *Mining Intelligence and Knowledge Exploration*. Springer; 2013: 139-147.
9. Shahedi M, Halicek M, Guo R, Zhang G, Schuster DM, Fei B: **A semiautomatic segmentation method for prostate in CT images using local texture classification and statistical shape modeling.** *Medical physics* 2018, **45**(6):2527-2541.
10. Pekar V, McNutt TR, Kaus MR: **Automated model-based organ delineation for radiotherapy planning in prostatic region.** *International Journal of Radiation Oncology* Biology* Physics* 2004, **60**(3):973-980.
11. He T, Guo J, Wang J, Xu X, Yi Z: **Multi-task Learning for the Segmentation of Thoracic Organs at Risk in CT images.** In *Segmentation of Thoracic Organs at Risk in CT Images (SegTHOR) 2019; Venice, Italy*. Edited by Petitjean C, Ruan S, Lambert Z, Dubray B. 2019:1-4.
12. Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q: **Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images.** *Medical physics* 2019, **46**(5):2169-2180.
13. Trullo R, Petitjean C, Nie D, Shen D, Ruan S: **Joint segmentation of multiple thoracic organs in CT images with two collaborative deep architectures.** In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Volume 10553*. Edited by Cardoso MJ, Arbel T: Springer; 2017: 21-29.
14. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB: **Advances in Auto-Segmentation.** *Seminars in radiation oncology* 2019, **29**(3):185-197.
15. Yang J, Veeraraghavan H, Armato III SG, Farahani K, Kirby JS, Kalpathy-Kramer J, van Elmpt W, Dekker A, Han X, Feng X: **Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017.** *Medical physics* 2018, **45**(10):4568-4581.
16. Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L: **Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques.** *Acta Oncologica* 2019, **58**(2):257-264.
17. Lei Y, Liu Y, Dong X, Tian S, Wang T, Jiang X, Higgins K, Beitler J, Yu D, Liu T, et al: **Automatic multi-organ segmentation in thorax CT images using U-Net-GAN.** In *SPIE Medical Imaging. Volume 10950*. San Diego, California, United States: SPIE; 2019: 1095010

18. 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; October 5-9; Munich, Germany.* Springer; 2015:234-241.
19. Wollmann T, Gunkel M, Chung I, Erfle H, Rippe K, Rohr K: **GRUU-Net: Integrated convolutional and gated recurrent neural network for cell segmentation.** *Medical image analysis* 2019, **56**:68-79.
20. Nair V, Hinton GE: **Rectified linear units improve restricted boltzmann machines.** In *Proceedings of the 27th international conference on machine learning (ICML-10); June 21-24; Haifa, Israel.* 2010:807-814.
21. Moeskops P, Wolterink JM, van der Velden BH, Gilhuijs KG, Leiner T, Viergever MA, Išgum I: **Deep learning for multi-task medical image segmentation in multiple modalities.** In *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer; 2016:478-486.
22. Zhen X, Yin Y, Bhaduri M, Nachum IB, Laidley D, Li S: **Multi-task shape regression for medical image segmentation.** In *International Conference on Medical Image Computing and Computer-Assisted Intervention; October 17-21; Athens, Greece.* Springer; 2016:210-218.
23. van Harten LD, Wolterink JM, Verhoeff JJ, Išgum I: **Exploiting Clinically Available Delineations for CNN-based Segmentation in Radiotherapy Treatment Planning.** *arXiv preprint arXiv:191104967* 2019:1-7.
24. Chen H, Qi X, Yu L, Heng P-A: **DCAN: deep contour-aware networks for accurate gland segmentation.** In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition.* 2016:2487-2496.
25. Ren X, Xiang L, Nie D, Shao Y, Zhang H, Shen D, Wang Q: **Interleaved 3D-CNN s for joint segmentation of small-volume structures in head and neck CT images.** *Medical physics* 2018, **45**(5):2063-2075.
26. Kazemifar S, Balagopal A, Nguyen D, McGuire S, Hannan R, Jiang S, Owrangi A: **Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning.** *Biomedical Physics & Engineering Express* 2018, **4**(5):055003.
27. Tong N, Gou S, Yang S, Ruan D, Sheng K: **Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks.** *Medical physics* 2018, **45**(10):4558-4567.
28. Ambellan F, Tack A, Ehlke M, Zachow S: **Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative.** *Medical image analysis* 2019, **52**:109-118.
29. Kosmin M, Ledsam J, Romera-Paredes B, Mendes R, Moinuddin S, de Souza D, Gunn L, Kelly C, Hughes C, Karthikesalingam A: **Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer.** *Radiotherapy and Oncology* 2019, **135**:130-140.
30. <https://github.com/LynnChan706/keras-u-net>.
31. Dumoulin V, Visin F: **A guide to convolution arithmetic for deep learning.** *arXiv preprint arXiv:160307285* 2016.

32. Suzuki S: **Topological structural analysis of digitized binary images by border following.** *Computer vision, graphics, and image processing* 1985, **30**(1):32-46.
33. Kingma DP, Ba JL: **Adam: A method for stochastic optimization.** In *3rd International Conference on Learning Representations (ICLR)*. San Diego, USA; 2015: 1-15.

Figures

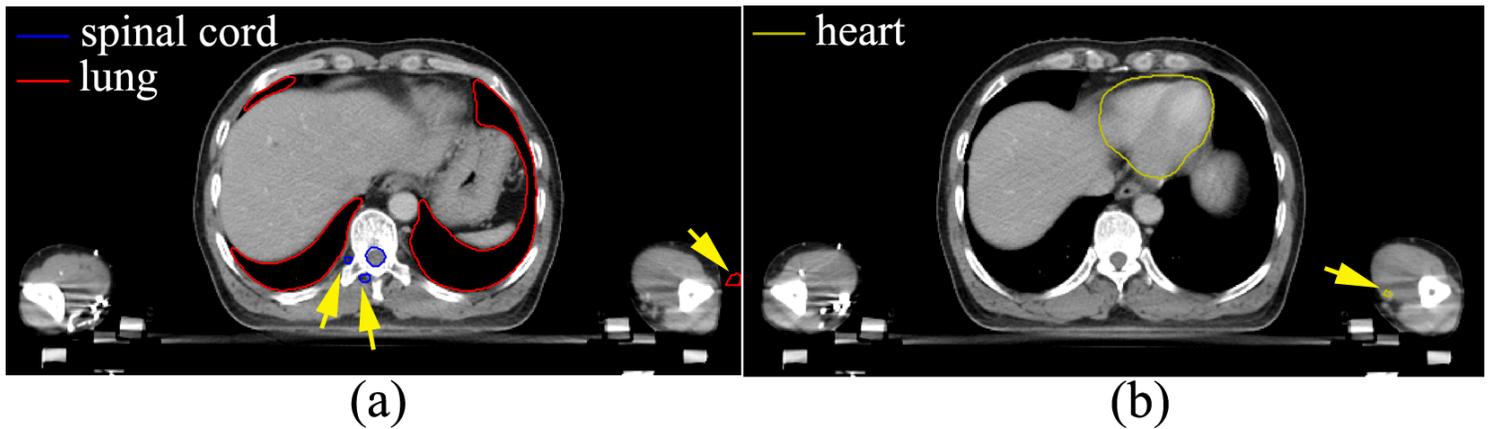


Figure 1

Illustration of mis-identification detected by Unet. Yellow arrows denote the mis-identification.

without knowing 0 grayscale in r_{heart} is lung and in r_{arm} is vitro air

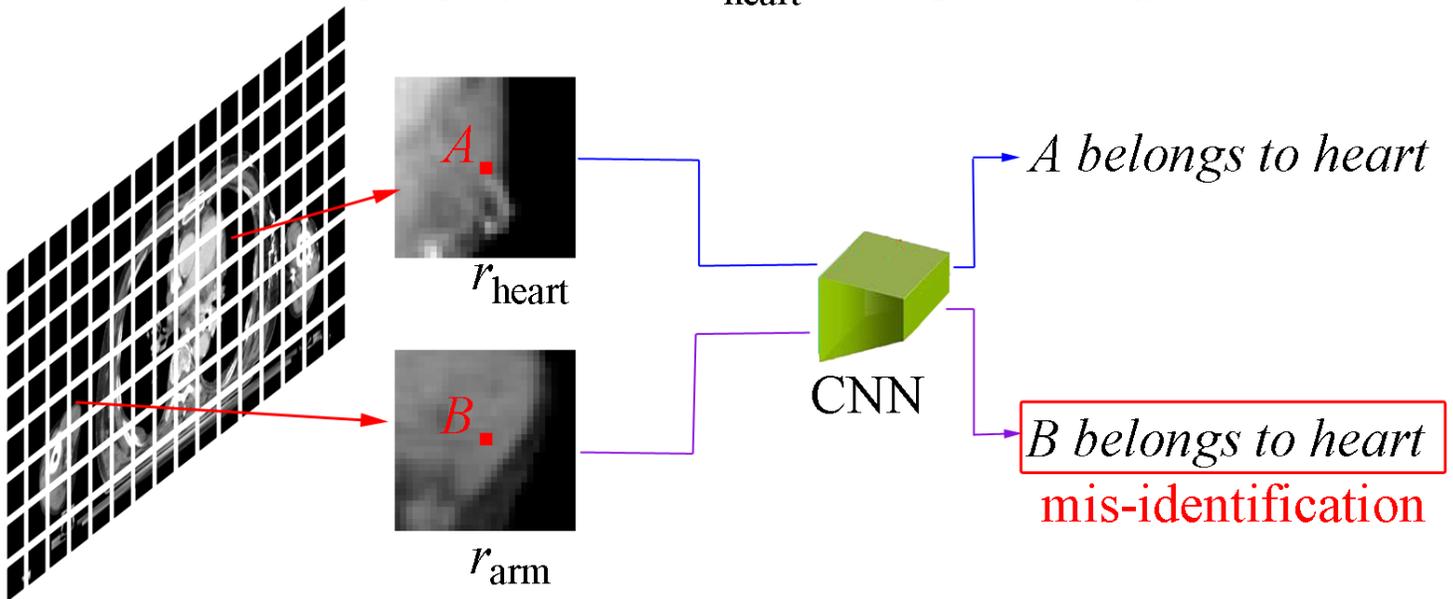


Figure 2

Illustration of convolution neural network (CNN) principle and possible reason of mis-identification. r_{heart} and r_{arm} denote two different organs' receptive fields.

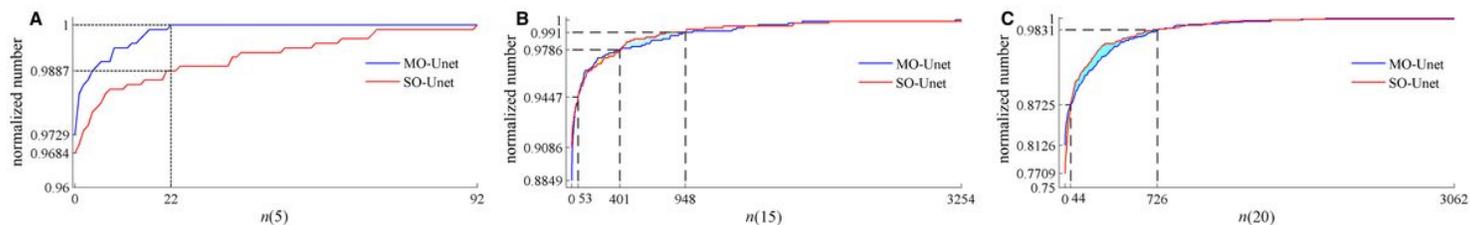
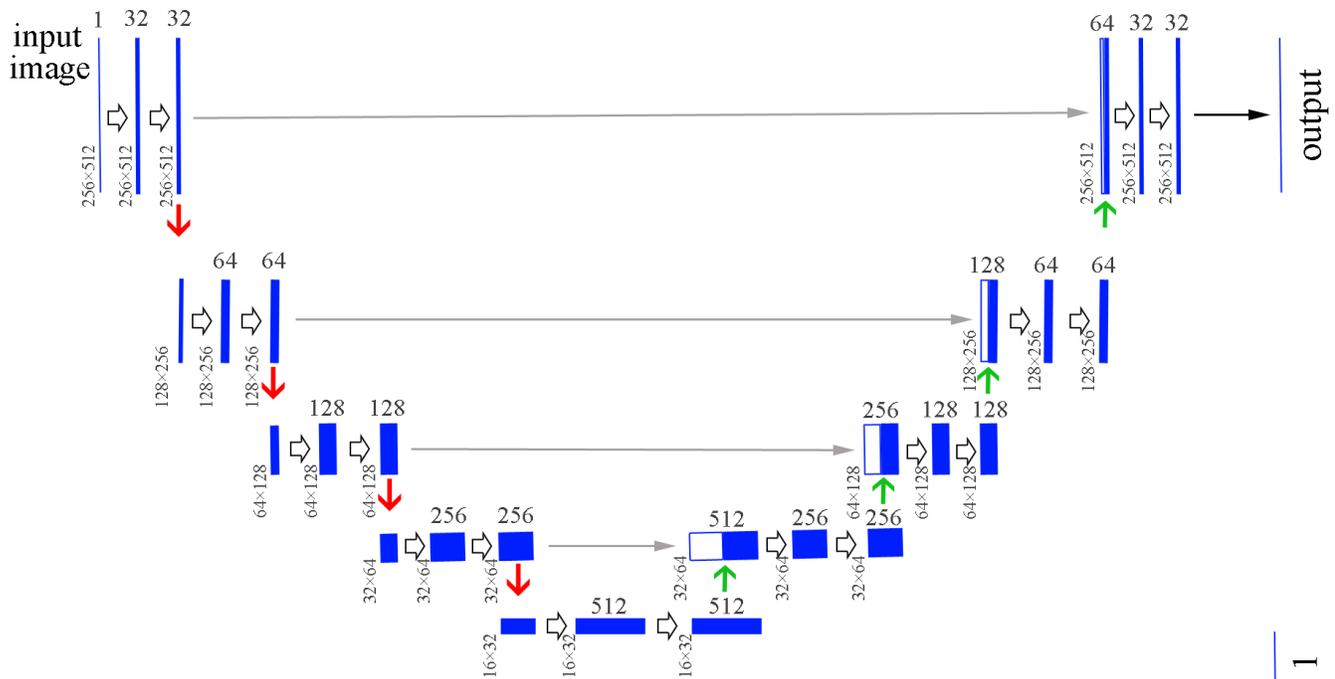


Figure 3

A. Cumulative distribution histogram of $n(5)$ for spinal cord delineation. B. Cumulative distribution histogram of $n(15)$ for heart delineation. The yellow area denotes the region where case growth rate of MO-Unet is greater than SO-Unet. The blue area denotes the region where case growth rate of SO-Unet is greater than MO-Unet. C. Cumulative distribution histogram of $n(20)$ for lung delineation. The yellow area denotes the region where case growth rate of MO-Unet is greater than SO-Unet. The blue area denotes the region where case growth rate of SO-Unet is greater than MO-Unet.

(a) SO-Unet



(b) MO-Unet

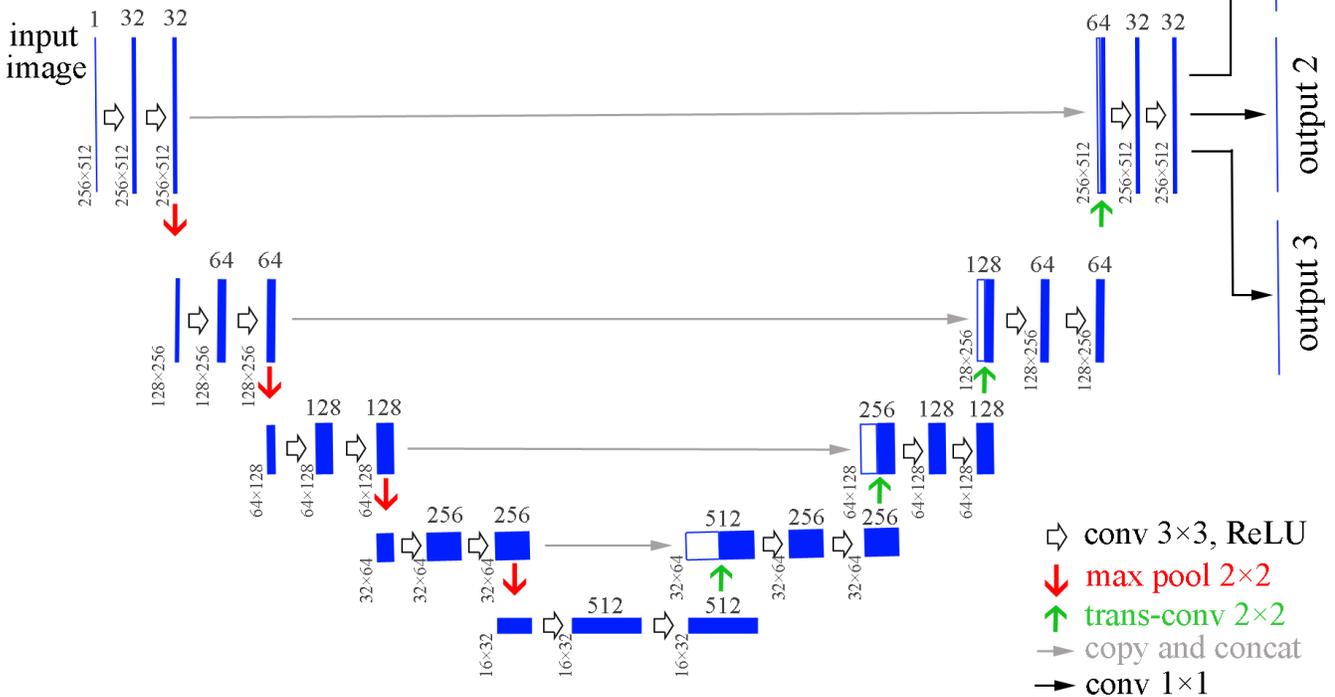


Figure 4

Detailed architecture of (a) SO-Unet and (b) MO-Unet. Each blue box is a multi-channel feature map. Its channel number is denoted on the box top and its x-y size is provided at the lower left edge of the box. White box represents the copied feature map. The arrows mean different operations. Abbreviation: conv $k \times k$ = convolution with kernel size of $k \times k$, ReLU = rectified linear unit, max pool $\alpha \times \alpha$ = max pooling with pooling size of $\alpha \times \alpha$, trans-conv $k \times k$ = transpose convolution with kernel size of $k \times k$, concat = concatenate.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.docx](#)