Deep Learning-Based Head and Neck Radiotherapy Planning Dose Prediction via Beam-Wise Dose Decomposition

Bin Wang[1,2], Lin Teng [3,1], Lanzhuju Mei[1], Zhiming Cui[1,4], Xuanang Xu[5], Qianjin Feng[3(✉)], and Dinggang Shen[1,6(✉)]

[1] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
dgshen@shanghaitech.edu.cn
[2] School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[3] School of Biomedical Engineering, Southern Medical University, Guangzhou, China
fengqj99@smu.edu.cn
[4] Department of Computer Science, The University of Hong Kong, Hong Kong, China
[5] Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, USA
[6] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China[3,1]

Abstract. Accurate dose map prediction is key to external radiotherapy. Previous methods have achieved promising results; however, most of these methods learn the dose map as a black box without considering the beam-shaped radiation for treatment delivery in clinical practice. The accuracy is usually limited, especially on beam paths. To address this problem, this paper describes a novel "disassembling-then-assembling" strategy to consider the dose prediction task from the nature of radiotherapy. Specifically, a global-to-beam network is designed to first predict dose values of the whole image space and then utilize the proposed innovative beam masks to decompose the dose map into multiple beam-based sub-fractions in a beam-wise manner. This can disassemble the difficult task to a few easy-to-learn tasks. Furthermore, to better capture the dose distribution in region-of-interest (ROI), we introduce two novel value-based and criteria-based dose volume histogram (DVH) losses to supervise the framework. Experimental results on the public OpenKBP challenge dataset show that our method outperforms the state-of-the-art methods, especially on beam paths, creating a trustable and interpretable AI solution for radiotherapy treatment planning. Our code is available at https://github.com/ukaukaaaa/BeamDosePrediction.

Keywords: Head and neck cancer · Radiation therapy · Dose prediction · Beam mask · Dose volume histogram

B. Wang and L. Teng—Equal Contribution

L. Wang (Eds.): MICCAI 2022, LNCS 13437, pp. 575–584, 2022.
https://doi.org/10.1007/978-3-031-16449-1_55

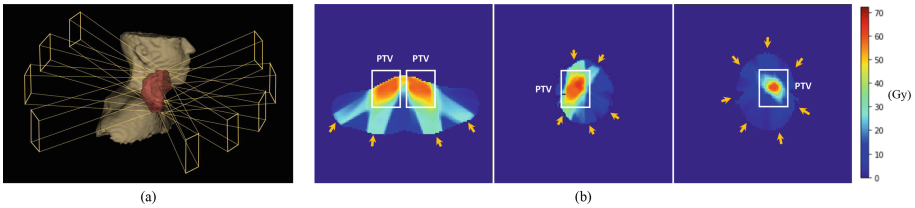

Fig. 1. (a) Demonstration of external radiotherapy with beam-shaped radiation. (b) Different slices of dose map with clear beam paths.

1 Introduction

External radiotherapy is a mainstream therapy widely used for head and neck cancer treatment. Its efficacy highly relies on high-quality treatment plans, in which a dose volume is elaborately designed to deliver a prescribed dose of radiation to the tumor while minimizing the irradiation received by organs-at-risk (OARs). In clinical workflow, this procedure is often accomplished by physicians manually adjusting numerous planning parameters and weights in a trial-and-error manner [2,4,7], which is not only time-consuming but also requires a great level of expertise. Hence, it is greatly demanded to develop an automatic method to predict accurate dose map for cancer treatment planning.

In recent years, due to the explosive development of machine learning techniques [11,13], many deep learning-based methods have been proposed to handle this challenging task. The prior efforts can be generally categorized into three groups. The first group of methods focuses on designing variant neural network architectures. For example, Liu et al. [6] have designed a cascaded 3D U-Net model, incorporating global and local anatomical features. The second group uses novel loss functions. Ngyuen et al. [10] have demonstrated that more accurate dose map can be generated when DVH loss is included. The last line methods propose to exploit additional prior knowledge and integrate into the network learning, including the gradient information [12] and distance information [15].

In external radiotherapy, treatment is achieved by delivering the radiation from several different directions (Fig. 1). Each direction of radiation will result in a beam-shaped dose volume. Due to the beam-wise delivery manner, the resulting dose volume often exhibits sharp edges near the beam boundary. The dose intensities inside the beam regions are much higher than those outside the beam regions. However, such critical prior knowledge is hardly considered in the previous methods, which often leads to some unsatisfying dose distributions on the beam paths and finally affects the prediction performance. This is because the input CT images do not contain any knowledge of the radiation beams. Thus, it is difficult for the deep network to infer the dose distribution accurately without extra prior knowledge of the beam shapes.

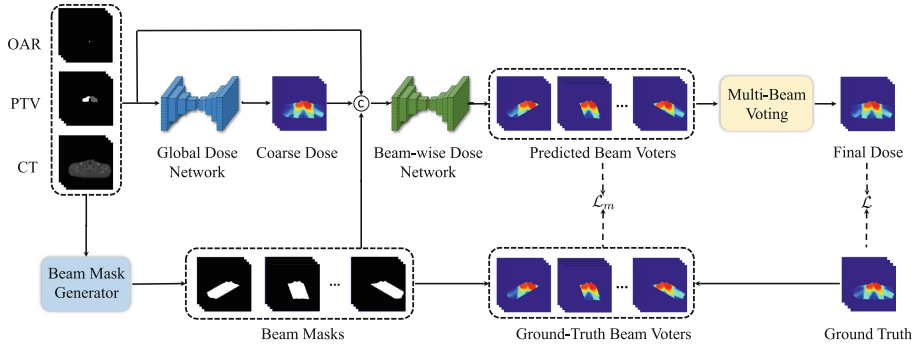

Fig. 2. An overview of our proposed method framework, including (a) global coarse dose prediction and (b) beam-wise dose prediction based on decomposition and multi-beam voting mechanism.

In this paper, to tackle the aforementioned challenges, we present a novel beam mask definition and take it as a prior knowledge for deep network training. To fully utilize this knowledge, a global-to-beam network embedded with multiple innovative strategies is used to conduct the dose prediction task from global to local, denoted as "disassembling-then-assembling" strategy. Specifically, in the first stage, we adopt Global Dose Network to coarsely predict the dose map, and simultaneously generate beam masks according to pre-defined angles and the planning target volume (PTV) masks. Then, in the second stage, we fine-tune the dose map through three novel strategies described as follows.

The main contributions of this work are three-fold. 1) The coarse dose map is guided by multiple beam masks from different angles, and decomposed into several sub-fractions of dose map (i.e., disassembling). Each sub-fraction is responsible for the prediction of corresponding beam path. 2) A multi-beam voting mechanism is proposed to reconstruct the final dose map, in which each voxel value is only determined by the sub-fractions containing that voxel (i.e., assembling). 3) We introduce a value-based DVH loss and a criteria-based DVH loss to focus on ROI regions more accurately and efficiently. To validate our proposed method, extensive experiments have been conducted and show that our method outperforms the state-of-art approaches both qualitatively and quantitatively.

2 Method

Although previous methods have achieved promising performance, we still find many regions with inaccurately-predicted dose values, especially along the beam paths. To solve this problem, we introduce a novel beam mask generator and a global-to-beam network by first predicting dose values of the whole image space and then decomposing the dose map into multiple beam voters by utilizing the proposed beam masks to conduct beam-wise dose prediction. The framework of our proposed method is illustrated in Fig. 2 with details described in the following sub-sections.

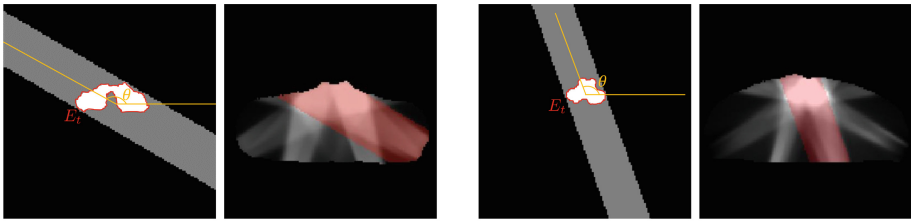

Fig. 3. Demonstration of beam mask generation.

2.1 Global Coarse Dose Prediction

In our framework, the input includes CT image, PTV, and OAR masks. Specifically, CT provides anatomical information. PTV indicates cancer region information (requiring high dose radiation), and OAR masks refer to normal tissues. As shown in Fig. 2, we first employ Global Dose Network to coarsely predict dose values of the whole image space. More importantly,before Beam-wised Dose Network at second-staged, we introduce additional prior knowledge (i.e., beam masks) that is helpful for dose map refinement.

2.2 Beam-Wise Dose Prediction

The output of the first stage is a coarse dose map, in which the basic shape of the PTV region and dose values are predicted roughly, especially for the results on the beam paths. To solve this problem, we generate beam masks according to the pre-defined angles and PTV masks. Then, Beam-wise Dose Network at the second stage refines the dose map in the manner of decomposition and voting strategies. The critical components in beam-wise dose prediction model are elaborated as below.

Beam Mask Generator. Introducing the beam information as a prior knowledge can provide a more targeted manner for predicting dose values on the beam paths. However, the beam information is not provided in most datasets. Thus, the main problem is how to acquire and represent such important information. To address this challenge, we propose a beam mask generator (Fig. 3). The beam path is mainly related to the location of PTV and the angle of the beam. Since the PTV location is different on each slice, we build the beam mask slice-by-slice to make it more suitable for the beam path, which can be defined as:

$$B(t) = F(\theta, E_t), \tag{1}$$

where θ refers to the angle and E_t includes the coordinates of PTV edges on the t-th slice. The detailed process of $F(\cdot)$ is to first calculate the slope of the beam and then utilize the points on edge boundary to find the intercept, resulting in two lines at the end. The region between these two lines represents the respective beam mask.

Decomposition of Dose Map. In each slice of dose map, there are usually more than one beam paths, whose angles and dose value distributions are different, causing the difficulty to jointly fine-tune all beam path regions. Different from the existing dose prediction techniques, our proposed method predicts multiple sub-fractions of the dose map, named beam voter. Each beam voter is responsible for the dose prediction on one beam mask. By decomposing the dose prediction of a whole image space into a set of beam-based prediction tasks, it is relatively easier for the refinement network to learn features of each beam-mask region. In the training process, we use a multi-beam MAE loss to supervise the prediction on each beam-mask region independently, which can be defined as:

$$\mathcal{L}_m = \frac{1}{N_i} \sum_{i=1}^{N_i} |P_i - G_i| \,, \tag{2}$$

where P_i and G_i refer to the prediction and the ground truth of the i-th beam voter. N_i denotes the maximum number of the beam voters.

Multi-beam Voting. The output of the refinement network in the second stage is multiple beam voters, each of which represents one sub-region of the final dose map. To merge these beam voters, we propose a Multi-Beam Voting mechanism, in which the dose value of each voxel is finally predicted and voted by the beams passing this voxel. Note that if it is only passed by one single beam, its dose value will be set by the same location on the corresponding beam voter. On the other hand, if a voxel is passed by multiple beams, all beam voters are responsible for this voxel to generate the final dose value by an average voting operation. After voting, we use MAE loss $\mathcal{L}_r$ to supervise the final prediction.

2.3 Training Objective

Value-Based DVH Loss. As a commonly used metric in radiotherapy, DVH has been considered as an important tool to enhance the dose prediction quality in ROI. Although many existing methods have proposed DVH loss [10] to calculate the volume difference between prediction and ground-truth DVH curves, it requires repeated image-based computations since a processing of the whole 3D image is needed for each dose value threshold. If we reduce the computational consumption by increasing the threshold interval, it would lose the accuracy to fit the ground-truth DVH curve. Hence, we proposed a value-based DVH loss to balance the efficiency and accuracy, which is defined as follows:

$$\mathcal{L}_{vDVH} = \frac{\sum_{s=1} \sum_{n=1}^{N_s} \left| R(\hat{Y} \cdot M_s)_n - R(Y \cdot M_s)_n \right|}{\sum_{s=1} N_s}, \tag{3}$$

where M_s denotes the ROI masks. $R(\cdot)$ is a sort operation. $\hat{Y}$ and Y refer to the prediction and ground-truth dose maps, respectively.

Specifically, we first use the ROI mask to extract the dose values in ROI region and obtain the DVH curve by conducting the sort operation. Then, we compute the difference between the prediction and the ground-truth DVH to supervise the network. In this way, the volume information can be represented by the rank of sorted dose values, with no hyper-parameter set in this loss function. Note that only one round computation is needed on the whole image, and the processing is directly employed on the dose values, which is more efficient and accurate.

Criteria-Based DVH Loss. In clinical treatment planning, the quality of the planning dose is evaluated by checking a set of critical points on the DVH curve, such as $C_1^{ptv}, C_{95}^{ptv}, C_{99}^{ptv}, C_{0.1cc}^{oar}$, and C_{mean}^{oar}, which indicate the dose values received by the top-ranked 1%, 95%, 99% volume of PTV, the dose value received by the top-ranked 0.1cc volume of OAR, and the mean dose received by OAR, respectively. We represent these criteria by calculating the sorted dose values ranked at the 99th, 5th, and 1st percentile inside the PTVs, and the maximum and mean of sorted dose values inside the OARs, respectively. Then, we define criteria-based DVH loss as

$$\mathcal{L}_{cDVH} = D_1^{ptv} + D_{95}^{ptv} + D_{99}^{ptv} + D_{0.1cc}^{oar} + D_{mean}^{oar}, \tag{4}$$

where $D_1^{ptv}, D_{95}^{ptv}, D_{99}^{ptv}, D_{0.1cc}^{oar}, D_{mean}^{oar}$ are the difference between prediction and the ground truth calculated at the corresponding criteria.

Finally, the total loss function will be elaborated as below:

$$\mathcal{L} = \mathcal{L}_r + \alpha\mathcal{L}_{vDVH} + \beta\mathcal{L}_{cDVH} \tag{5}$$

where α and β are hyper-parameters to balance the two DVH loss terms.

3 Experiment

We evaluate our proposed method on the public OpenKBP dataset [1] from 2020 AAPM Grand Challenge, which includes 340 head and neck cancer patients (200 for training, 40 for validation, and 100 for testing). For each patient, paired CT scan, OAR mask, PTV mask, possible dose mask, and ground-truth dose map are provided with the same size of $128 \times 128 \times 128$. The model performance is tracked with two metrics used in the Challenge above, i.e., Dose score and DVH score. Moreover, we introduce the DVH curve as another tool to evaluate the accuracy of dose prediction in ROI.

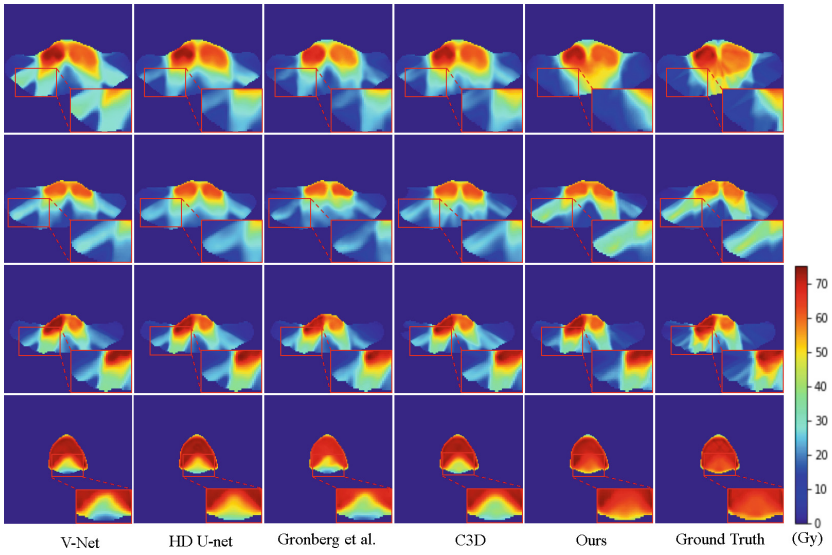

Fig. 4. Visualization of our proposed method and SOTA methods. An obvious improvement on beam paths can be seen in the red enlarged boxes. (Color figure online)

For the first stage of coarse prediction, random rotation ranged from -20° to 20° and random flip along Z-axis are performed to augment the data. Note that the PTV masks and OAR masks are merged in one mask, where the different mask was assigned with a different label. The beam masks are generated based on Eq. (1), where θ is set according to the pre-defined angles in the dataset [1]. Both Global Dose Net and Beam-wise Dose Net have a U-Net structure. To supervise the networks, the Adam optimizer with an initial learning rate of 1e−4 is used during the training process. And each epoch takes about two minutes on a GPU of NVIDIA Tesla M40 24 GB.

Table 1. Quantitative comparison with state-of-the-art methods (*: no released code; †: p-value < 0.05).

Method	Dose score [Gy]	DVH score [Gy]
V-Net [8]	$2.922 \pm 1.166^{\dagger}$	$1.545 \pm 1.178^{\dagger}$
Xu et al. [14]	2.753^{*}	1.559^{*}
Zimmermann et al. [16]	$2.620 \pm 1.100^{*}$	$1.520 \pm 1.060^{*}$
HD U-net [9]	$2.592 \pm 1.048^{\dagger}$	$1.643 \pm 1.123^{\dagger}$
Gronberg et al. [3]	$2.563 \pm 1.143^{\dagger}$	$1.704 \pm 1.096^{\dagger}$
C3D [6]	$2.429 \pm 1.031^{\dagger}$	$1.478 \pm 1.182^{\dagger}$
Lin et al. [5]	2.357^{*}	1.465^{*}
Ours	$\mathbf{2.276 \pm 1.013}$	$\mathbf{1.257 \pm 1.163}$

3.1 Comparison with State-of-the-Art Methods

To validate the advantage of our proposed method, we compare with state-of-the-art methods, including V-Net [8], Xu et al. [14], Zimmermann et al. [16], HD U-net [9], Gronberg et al. [3] and C3D [6], which won the first and second place on the 2020 AAPM challenge learderboard [1], respectively, and Lin et al. [5].

The quantitative results are shown in Table 1. Our method significantly outperforms existing methods, in terms of dose score and DVH score. Moreover, we can see from the visualization results of dose maps in Fig. 4 that our prediction matches better with the ground truth, especially on the beam paths. We also provide the DVH curves of the prediction and the ground truth in Fig. 5 showing that our DVH curves are closer to the ground-truth DVH curves.

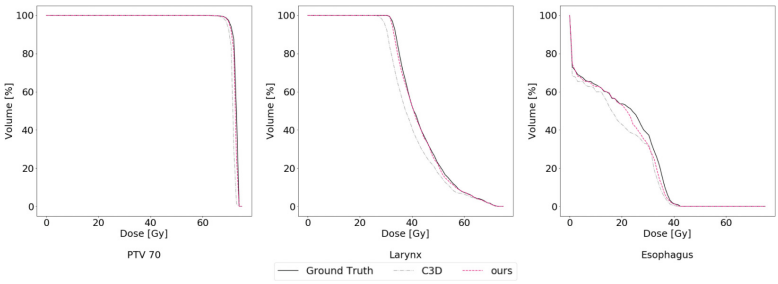

Fig. 5. Visualization of DVH curves by our method and SOTA methods, including DVH curves of PTV70, Larynx, and Esophagus.

3.2 Ablation Study

In this study, we take a cascaded 3D U-Net as the Baseline model and evaluate the effectiveness of two key components of our proposed method: (1) beam-wise dose prediction (BDP) including the decomposition of dose map and multi-beam voting scheme, and (2) valued-based DVH loss (L_{vDVH}) and criteria-based DVH loss (L_{cDVH}). The quantitative results are presented in Table 2. It can be seen that BDP improves the prediction accuracy (46.1% improvements in terms of dose score), indicating that the introduction of the beam masks facilitates the network to learn more features from beam paths region. Then, adding L_{vDVH} shows 10.8% improvement of DVH score and also makes prediction of dose map more efficient, i.e., reducing the computational time from 6 min an epoch by previous DVH loss [10] to 2 min an epoch. Additionally, due to the fact that L_{cDVH} incorporates key criteria of treatment plannings in clinics, it helps promote dose score by 12.2% and dvh score by 20.2%, making our prediction the best result in this dose prediction task.

Table 2. Ablation study of our method, evaluated with dose score and DVH score.

Baseline	BDP	L_{vDVH}	L_{cDVH}	Dose score [Gy]	DVH score [Gy]
✓				2.862 ± 1.049	1.586 ± 1.146
✓	✓			2.401 ± 1.033	1.567 ± 1.179
✓	✓	✓		2.398 ± 1.011	1.459 ± 1.212
✓	✓	✓	✓	$\mathbf{2.276 \pm 1.014}$	$\mathbf{1.257 \pm 1.163}$

4 Conclusion

We describe a novel "disassembling-then-assembling" strategy and propose a global-to-beam framework to accurately conduct the dose prediction task. The proposed method first learns the whole image space of the dose map and then decomposes it into beam-based sub-fractions by proposed beam masks. Moreover, we get the final dose map by utilizing multi-beam voting strategy. Besides, we propose two novel value-based and criteria-based DVH loss to focus on ROI region efficiently. Experimental results demonstrate our method can more precisely predict dose map compared with state-of-the-art methods. The predicted dose is very close to the physically deliverable one and thus can be used as a good starting point in treatment planning, substantially reducing the time and inter-observer variations in clinical workflow. We aim to leverage this method to create trustable and interpretable AI solution for radiotherapy.

References

1. Babier, A., et al.: OpenKBP: the open-access knowledge-based planning grand challenge and dataset. Med. Phys. **48**(9), 5549–5561 (2021)
2. Dias, J., Rocha, H., Ferreira, B., do Carmo Lopes, M.: Simulated annealing applied to IMRT beam angle optimization: a computational study. Physica Medica **31**(7), 747–756 (2015)
3. Gronberg, M.P., Gay, S.S., Netherton, T.J., Rhee, D.J., Court, L.E., Cardenas, C.E.: Dose prediction for head and neck radiotherapy using a three-dimensional dense dilated u-net architecture. Med. Phys. **48**(9), 5567–5573 (2021)
4. Kearney, V., Chan, J.W., Valdes, G., Solberg, T.D., Yom, S.S.: The application of artificial intelligence in the IMRT planning process for head and neck cancer. Oral Oncol. **87**, 111–116 (2018)
5. Lin, Y., Liu, Y., Liu, J., Liu, G., Ma, K., Zheng, Y.: LE-NAS: learning-based ensemble with NAS for dose prediction. arXiv preprint arXiv:2106.06733 (2021)
6. Liu, S., Zhang, J., Li, T., Yan, H., Liu, J.: A cascade 3D U-Net for dose prediction in radiotherapy. Med. Phys. **48**(9), 5574–5582 (2021)
7. Men, C., Romeijn, H.E., Taşkın, Z.C., Dempsey, J.F.: An exact approach to direct aperture optimization in IMRT treatment planning. Phys. Med. Biol. **52**(24), 7333 (2007)
8. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)

9. Nguyen, D., et al.: 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture. Phys. Med. Biol. **64**(6), 065020 (2019)
10. Nguyen, D., et al.: Incorporating human and learned domain knowledge into training deep neural networks: a differentiable dose-volume histogram and adversarial inspired framework for generating pareto optimal dose distributions in radiation therapy. Med. Phys. **47**(3), 837–849 (2020)
11. Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D.: Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 455–463. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_52
12. Tan, S., et al.: Incorporating isodose lines and gradient information via multi-task learning for dose prediction in radiotherapy. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12907, pp. 753–763. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87234-2_71
13. Xiang, L., et al.: Deep embedding convolutional neural network for synthesizing CT image from t1-weighted MR image. Med. Image Anal. **47**, 31–44 (2018)
14. Xu, X., et al.: Prediction of optimal dosimetry for intensity-modulated radiotherapy with a cascaded auto-content deep learning model. Int. J. Radiat. Oncol. Biol. Phys. **111**(3), e113 (2021)
15. Zhang, J., Liu, S., Yan, H., Li, T., Mao, R., Liu, J.: Predicting voxel-level dose distributions for esophageal radiotherapy using densely connected network with dilated convolutions. Phys. Med. Biol. **65**(20), 205013 (2020)
16. Zimmermann, L., Faustmann, E., Ramsl, C., Georg, D., Heilemann, G.: Dose prediction for radiation therapy using feature-based losses and one cycle learning. Med. Phys. **48**(9), 5562–5566 (2021)