Flexible-C^m GAN: Towards Precise 3D Dose Prediction in Radiotherapy –Supplementary Material–

Riqiang Gao Bin Lou Zhoubing Xu Dorin Comaniciu Ali Kamen
Digital Technology and Innovation, Siemens Healthineers, Princeton NJ 08540
riqiang.gao@siemens-healthineers.com

A. Summary of Mathematical Symbols

G: the generator in our FCGAN.
$G(\cdot)$: the generated sample from G.
D: the discriminator in our FCGAN.
$\mathbf{C}$: the set of all conditions.
C^i: *i*-th condition in $\mathbf{C}$, which can be a scalar, matrix or others.
E^i: encoder that extracts features from the (predicted) dose map for *i*-th condition, which can be *Identity*.
$\mathbf{m}$: the miss-indicator of all conditions (vector).
m^i: the miss-indicator of *i*-th condition (scalar).
$\mathbf{Y}$: the 3D dose map for one patient; only used in derivation.
Y_j: the value of *j*-th voxel (scalar) in dose map $\mathbf{Y}$.
$\mathbf{Y}_i$: the 3D dose map for *i*-th patient.
$\hat{\mathbf{Y}}_i$: predicted 3D dose map for *i*-th patient.
$\mathbf{M}_t$: A binary matrix with the same dimension as $\mathbf{Y}$; only the voxel indexes with dose values in *t*-th bin are 1; only used in derivation.
$\mathbf{M}$: A binary matrix with the same dimension as $\mathbf{Y}$; voxel indexes in the target ROI are 1; only used in derivation.
M_j: the value of *j-th* voxel in mask $\mathbf{M}$. $M_j = 1$ if *j*-th voxel is in the ROI, otherwise $M_j = 0$.

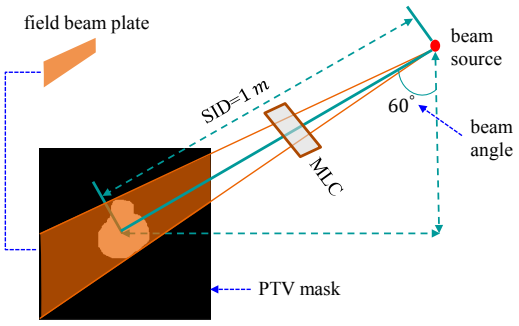


Figure 9. Field beam plate (single angle) illustration. We assume that MLC design ideally would limit the X-ray within PTV/source sector. Multi-leaf collimator (MLC) is a beam-limiting device filtering X-rays delivered to the patient.

$\mathbf{M}^s$: A binary matrix with the same dimension as $\mathbf{Y}$, in which voxel indexes of the *s*-th ROI are 1.
$\mathbf{M}_i^s$: A binary matrix with the same dimension as $\mathbf{Y}_i$, in which voxel indexes of the *s*-th ROI are 1 for *i*-th patient.

B. Data Preprocessing

We extract geometry information from the RT PLAN data with DICOM lookup [1]. For example, the angle list of the beam-static example (upper) in Figure 5 is `[22, 56, 111, 152, 180, 335]`. We use the OpenCV function `cv2.line` [3] to create the angle-plate (as Figure 5c) originating at `isocenter`. The illustration of creating the beam-plate is shown in Figure 9. The linear accelerator which generates beams has a consistent `source` to `isocenter` distance (SID) of 1000 mm. We can create the field beam plate (orange shadow in Figure 9) with one beam angle. The overall beam-plate (as in Figure 5d) is the average of field plates of all angles. Note that we only need the beam angles and PTV mask to create the angle/beam plates.

The region closer to the `isocenter` usually gets higher dose levels. To let the model focus on the important region,

[1] https://dicom.innolitics.com/ciods

we centralize the `isocenter` as the voxel center for input with two ranges in our experiments 96 mm $\times$ 224 mm $\times$ 224 mm and 144 mm $\times$ 256 mm $\times$ 256 mm. The axial range is smaller as the treatment beam passes orthogonal to the axial direction. Our major experiments are conducted with 96 mm $\times$ 224 mm $\times$ 224 mm with data size 32 mm $\times$ 192 mm $\times$ 192 mm. To separate the RT planning prediction from the dose scale (clinical-dependent) and remove potential clinical bias, we normalize the reference dose to 0-5 during training.

Seven organs at risk (OARs) have been included in the multiple channel input: bronchial, chest wall and rib, esophagus, heart, proximal bronchial, spinal cord, and lung. For those patients missing a particular organ mask, we follow [7] to impute the mask as zero volume.

C. Data Augmentation and Its Ablation Study

We apply data augmentation for all experiments in the main texts. We use the **Rotation**, **RandomResizedCrop**, and **Flip** for training, as shown in Figure 10. Those three operations are independent, and multiple input channels use the same operations for one patient, which avoids missing matches across multi-channel.

Rotation: The volume is rotated between $-30°$ to $30°$ with axial as the axis. This operation probability is 0.5.

Flip: The volume is flipped in the sagittal direction. The probability of this operation is 0.5.

Random Resized Crop (RRC): The extended volume with size $D\times H\times W$ is randomly cropped with Shape 1 (i.e., $r_1\cdot\hat{D}\times r_2\cdot\hat{H}\times r_3\cdot\hat{W}$), resize to Shape 2 ($\hat{D}\times\hat{H}\times\hat{W}$), then central cropped to Shape 3 (i.e.,$d\times h\times w$). The probability of this operation is 0.5. When the RRC is not executed, the volume is directly central cropped to Shape 3.

In majority of our experiments, we apply size $d\times h\times w$ as 32 $\times$ 192 $\times$ 192, with physical range 96 mm $\times$ 225 mm $\times$ 225 mm. The corresponding $D\times H\times W$ and $\hat{D}\times\hat{H}\times\hat{W}$ is 48 $\times$ 256 $\times$ 256 and 40 $\times$ 224 $\times$ 224, respectively.

In the ablation study (Table 9), we have size $d\times h\times w$ as 48 $\times$ 128 $\times$ 128, with physical range 144 mm $\times$ 256 mm $\times$ 256 mm. The corresponding $D\times H\times W$ and $\hat{D}\times\hat{H}\times\hat{W}$ is 60 $\times$ 150 $\times$ 150 and 54 $\times$ 139 $\times$ 139, respectively.

We conduct ablation experiments to validate the effectiveness of data augmentation, as in Table 6. Using our data augmentation, all the metrics have been improved.

Models	SDE (↓)	DDE(↓)	MAE(↓)	CEL (↓)
w/o aug.	6.29	1.73	3.03	0.10
w/ aug.	**5.80**	**1.48**	**2.64**	**0.05**

Table 6. Experiments comparing with and without data augmentation using the FCGAN.

D. Network Structure

One of our generator backbone DoseNet in FCGAN is shown in Figure 11. The image-level conditions (CT, PTV/OAR masks, and angle/beam plates) are fed as multi-channel input. The scalar condition (i.e., mode) and miss-mask are repeated to $\frac{d}{3}\times h\times w$. Together with random noise, scalar condition, and the miss-mask information form the gate volumes (e.g., orange in Figure 11), which have different sizes and have been added to multiple layers in the model. Another backbone we used in FCGAN is Probabilistic-UNet (PUNet) [6], which follows a similar way as DoseNet, and its probabilistic mechanism can be found in [6].

The discriminator D and encoder E use the same structure and relatively simpler than G. The structure contains five convolutional layers with the channel size `[32, 64, 128, 256, 256]`. After global adaptive average pooling, there is one fully-connected layer to convert features to class probabilities.

E. Supplements of SDV Loss

Here, we provide supplements about the derivation of shift-dose-volume (SDV) loss introduced in Sec. 3.3.2.

Definition 1. *The shift-dose-volume loss L^t_{sdv} of each bin $[D_t, D_t+\frac{D_T}{T})$ is the absolute error $\varepsilon(\cdot)$ of all voxels that contributed to $f(D_t)$ multiply the bin width w, i.e., expected error of $f(D_t)$ from voxel perspective. The L_{sdv} of each ROI is the sum of each bin L^t_{sdv} in its DVH.*

Given a ROI with mask $\mathbf{M}$, its L_{sdv} can be computed as:

$$\begin{aligned}\sum_j M_j\sum_{t=1}^{T}L^t_{sdv} &= \sum_{t=1}^{T}\sum_j \mathbb{1}(Y_j-D_t)\cdot\varepsilon(Y_j\cdot M_j)\cdot w\\ &= \sum_{t=1}^{T}\varepsilon(\mathbf{Y}\odot(\mathbf{M}_t+...+\mathbf{M}_T))\cdot w\\ &= \sum_{t=1}^{T} t\cdot\varepsilon(\mathbf{Y}\odot\mathbf{M}_t)\cdot w\\ &= \sum_{t=1}^{T}\frac{\cancel{T}D_t}{\cancel{D_T}}\cdot\varepsilon(\mathbf{Y}\odot\mathbf{M}_t)\cdot\frac{\cancel{D_T}}{\cancel{T}}\\ &= \sum_{t=1}^{T} D_t\cdot\varepsilon(\mathbf{Y}\odot\mathbf{M}_t)\end{aligned} \tag{11}$$

where $\odot$ is the Hadamard product, and $\mathbf{Y}$ is the referenced dose map. Assume the number of bin T is large enough to make every $Y_i\in[D_t, D_t+\frac{D_T}{T})$ indexed by $\mathbf{M}_t$ is close to D_t, i.e., $\lim_{T\to\infty} D_t\cdot\varepsilon(\mathbf{Y}\odot\mathbf{M}_t)=\forall_{j\in\mathbf{M}_t}Y_j\cdot\varepsilon(\mathbf{Y}\odot\mathbf{M}_t)$. [2] Since $\mathbf{M}=\sum_{i=1}^{T}\mathbf{M}_t$ and $\forall_{k\neq t}\mathbf{M}_t\odot\mathbf{M}_k=\mathbf{0}$, we have:

[2] $j\in\mathbf{M}_t$ denotes j-th pixel in $\mathbf{M}_t$ equals to 1 here.

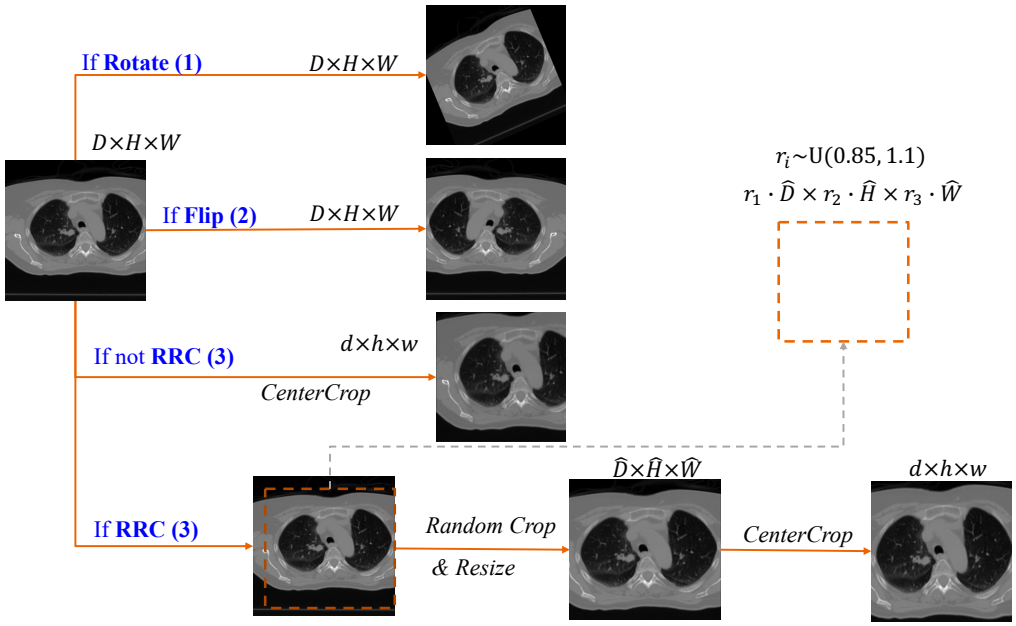


Figure 10. Three types of data augmentation used in training.

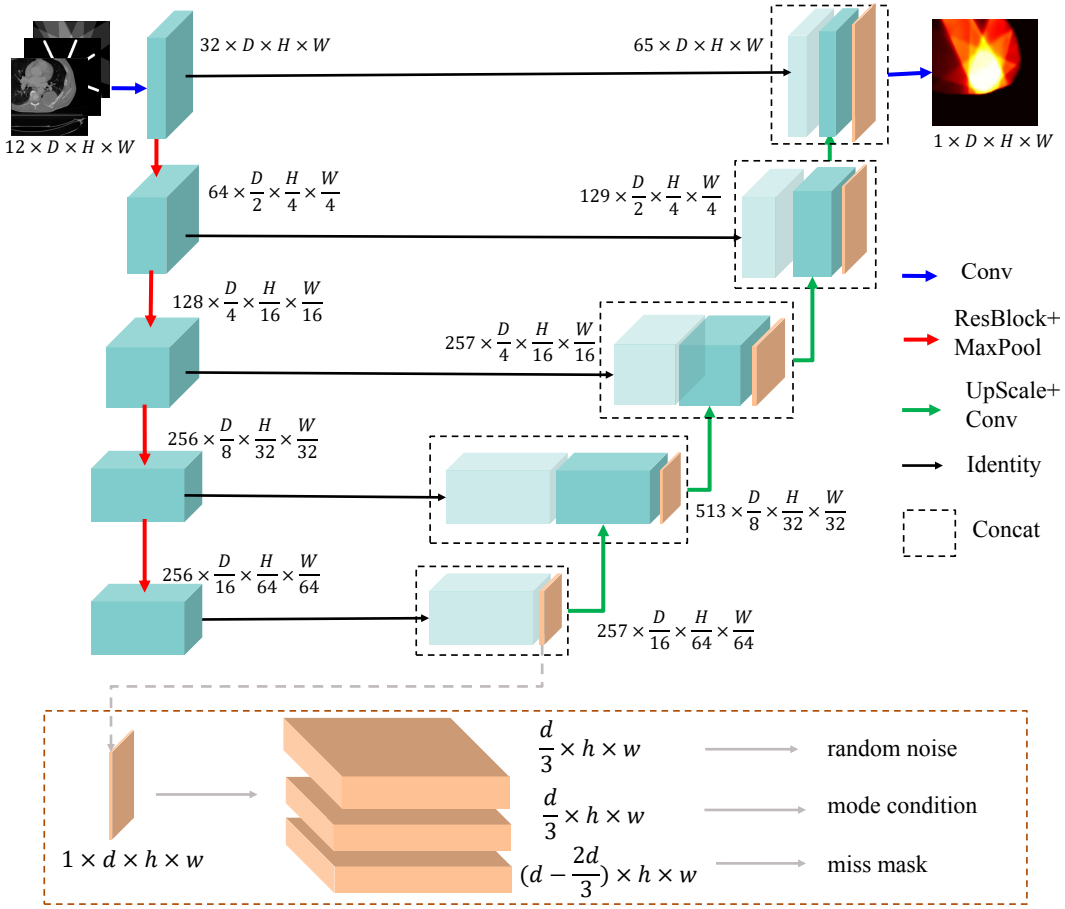


Figure 11. The network structure of DRUNet in our FCGAN framework (e.g., G in Figure 3). The lower part illustrates how to integrate the random noise, mode condition, and miss-mask into the network.

$$
\begin{aligned}
L_{sdv} &= \sum_{t=1}^{T} L_{sdv}^{t} = \frac{\sum_{t=1}^{T} \forall_{j \in \mathbf{M}_t} Y_j \cdot \varepsilon(\mathbf{Y} \odot \mathbf{M}_t)}{\sum_j M_j} \\
&= \frac{\sum_j Y_j (Y_j - \hat{Y}_j) M_j}{\sum_j M_j} \\
&= ||\mathbf{Y} \odot (\mathbf{Y} - \hat{\mathbf{Y}}) \odot \mathbf{M}||_1 .
\end{aligned} \tag{12}
$$

where $\hat{\mathbf{Y}}$ is predicted dose maps, j is the element index for dose map and mask, i.e., M_j and Y_j are j-th voxel in $\mathbf{M}$ and $\mathbf{Y}$ respectively, and $\varepsilon(\mathbf{Y} \odot \mathbf{M})$ is defined as $\sum_j (Y_j - \hat{Y}_j) M_j$. Note that here we use: $||\mathbf{A} \odot \mathbf{M}||_1 = \frac{\sum_j A_j M_j}{\sum_j M_j}$ when $\mathbf{M}$ is a binary matrix (mask), and $\mathbf{A}$ is any matrix has the same size with $\mathbf{M}$.

Eq. 12 is used for derivation and only for one patient and single ROI. When consider N patients with S ROIs, we have Eq. 10 in the main text, and as below:

$$
L_{sdv} = \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{S} \lambda_s ||\mathbf{Y}_i \odot (\mathbf{Y}_i - \hat{\mathbf{Y}}_i) \odot \mathbf{M}_i^s||_1 . \tag{13}
$$

Next, we provide justification of properties of L_{sdv} mentioned in Sec. 3.3.2.

Property 1: *Higher dose voxels contribute more to both DVH computation and our SDV loss.*

According to the definition of dose-volume histogram (DVH) for a region of interest (ROI) masked by $\mathbf{M}$:

$$
f(D_t) = \frac{\sum_j \mathbb{1}(Y_j - D_t) M_j}{\sum_j M_j} . \tag{14}
$$

The sum of fractional volume (y-axis in DVH) $f(D_t)$ of all T bins is:

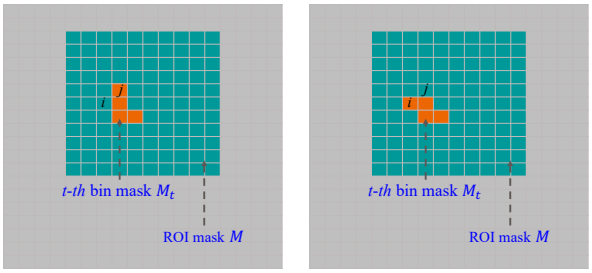
$$
F(D) = \sum_{t=1}^{T} \frac{\sum_j \mathbb{1}(Y_j - D_t) M_j}{\sum_j M_j} . \tag{15}
$$

The count of Y_j in $F(D)$ is $\lceil \frac{Y_j}{w} \rceil$, where $w = D_T / T$ is the bin width. Thus, the higher the dose value Y_j, the more count in DVH. Note that the bin width w should be multiplied to $f(D_t)$ to calculate the (expected) error or loss.

According to the definition of our SDV loss (e.g., Eq. 12), the dose map $\mathbf{Y}$ serves as a weighted mask when computing the voxel-wise errors $||(\mathbf{Y} - \hat{\mathbf{Y}}) \odot \mathbf{M}||_1$. Thus, higher-dose voxels have higher weights in L_{sdv} calculation. To sum up, we have justified Property 1.

Property 2: *Optimal L_{sdv} leads to an exact match of 3D dose maps in its ROI, while zero (i.e., optimal) DVH gaps theoretically can come from sub-optimal spatial mismatch.*

We use an example to justify this property. As shown in Figure 12, two voxels with two different dose values Y_i and Y_j belong to the same ROI. If their prediction $\hat{Y}_i$ and $\hat{Y}_j$ is switched compare with the reference values, i.e.:



(a) reference (b) prediction

Figure 12. Sub-optimal spatial dose pair with zero DVHs gap.

$$
\hat{Y}_i = Y_j, \hat{Y}_j = Y_i, \tag{16}
$$

according to Eq. 15, the predicted and referenced DVHs would be exactly the same for Y_i and Y_j, while spatial error exists. According to our L_{sdv} (i.e., Eq. 12 or Eq. 13), loss exists in those two voxels.

Contrastively, $L_{sdv} = 0$ if and only if the reference and predicted doses exactly match voxel-wise with the ROI. So, there is no spatial mismatch. This demonstrates the Property 2.

Summary. SDV loss is strictly derived from the widely-used dose-volume histogram, which enables spatially-unbiased deep learning optimization and evaluation motivated by the clinical interest.

F. Dataset Details and Discussion

The demographics of the two datasets are shown in Table 7. D1WA denotes Dataset 1 with the angle configuration available, which is from REQUITE [9] [3]. D2NA denotes Dataset 2 with no angle configuration, which comes from our collaborated clinical site. We removed non-coplanar plans due to their limited numbers.

It is noteworthy that RT planning data for research use is challenging. Our study is conducted on larger-scale data resources with more patients compared to previous studies, e.g., 1149 patients (ours) vs. 151 patients [4] vs. 141 patients [5] vs. 129 patients [2] vs. 340 patients [1, 7]. Aside from the availability of angle configuration, D1WA and D2NA have significantly discrepant populations, e.g., location, sex, smoking history, and tumor stage. Furthermore, D1WA is collected from multiple countries, which is heterogeneous in both population and planning mode. Thus, as shown in Table 3, the performance has dramatically dropped when a model is trained on one set and tested on another. Especially when trained on D2NA and tested on D1WA, the model 1) can not take advantage of the available

[3] The REQUITE lung dataset has been processed and filtered by Big Data Office of Siemens Healthineers. Some cases that can not been processed with rt tool https://github.com/qurit/rt-utils are not included in this study.

angle condition of D1WA since the training has no angles, and 2) has to deal with more heterogeneous testing (i.e., on D1WA) while trained on a relatively homogeneous population (D2NA).

Note for REQUITE data: We thank all the contributors to the REQUITE project, including the patients, clinicians and nurses. The core REQUITE consortium consists of David Azria, Erik Briers, Jenny Chang-Claude, Alison M. Dunning, Rebecca M. Elliott, Corinne Faivre-Finn, Sara Gutiérrez-Enríquez, Kerstie Johnson, Zoe Lingard, Tiziana Rancati, Tim Rattay, Barry S. Rosenstein, Dirk De Ruysscher, Petra Seibold, Elena Sperk, R. Paul Symonds, Hilary Stobart, Christopher Talbot, Ana Vega, Liv Veldeman, Tim Ward, Adam Webb and Catharine M.L. West.

REQUITE received funding from the European Union's Seventh Framework Program for research, technological development, and demonstration under grant agreement no. 601826. Funding for the five year REQUITE project ended on 30th September 2018. REQUITE does not benefit financially from supplying data and/or samples to researchers but does make a charge to cover its costs and support continued maintenance of the database and biobank beyond the ending of the funding period. To facilitate this continued access to researchers, the REQUITE Steering Committee approved a tiered cost recovery model for access to data and/ or samples. Contact REQUITE (requite@manchester.ac.uk) for more information on pricing. All authors had access to all the data reported in the study. The senior authors had full access to all the data in the study and had the final responsibility for the decision to submit for publication.

G. Discussion on Metrics

We have included four quantitative metrics (defined in Sec. 4.2.3) and qualitative results to evaluate the effectiveness of our method.

Mean absolute error (MAE) evaluates the overall match between the predicted and referenced maps. One drawback of MAE is that two clinically acceptable dose maps may have large MAE since the radiotherapy plan quality is in fact subjective. It also explains that some baseline models predict smoother Gaussian-like distribution of dose and even have a better MAE in some of our experiments (e.g., Table 2).

Compared with MAE, **shift-DV error (SDE)** and **discrete-DVH error (DDE)** focus on the match between the region of interests (i.e., PTV and OARs) in terms of DVH. The difference between DDE and SDE is that the former incorporates some discrete points and statistics of a DVH, and SDE measures error more comprehensively considering all bins.

Cross entropy loss (CEL) measures the discrepancy between the predicted dose mode and the referenced planning mode. Since the pretrained model for classifying is trained on real dose maps so that unrealistic doses will not have an accurate category matching, CEL can measure how realistic the predicted dose is to some degree. Ideally, if the pretrained model has an infinite capacity and is trained with infinite samples, no overfitting could exist in dose prediction so that CEL would be unbiased. However, overfitting can exist due to the limited availability of training data in radiotherapy planning. As disclosed in Sec. 4.2.3, the CEL metric can be somewhat biased but still helpful in evaluating the model in terms of mode match and generating realistic doses.

H. Additional Ablation Experiments

We also show more model comparison and qualitative results (e.g., Table 8, Figure 13) to support that our model generates doses that have better mode matches and are more realistic. Since no angle is specified here and due to the subjective nature of RT planning, both our predicted dose and reference dose can be considered acceptable plans even if the angles are slightly different. To specifically validate the adversarial training, we provide another ablation study by only removing the adversarial loss; the prediction becomes blurry (Fig. 14c), and the reconstructed error also increased (i.e., MAE from 2.64 to 2.70).

We also include another ROI strategy (larger region 144 mm $\times$ 256 mm $\times$ 256 mm, with voxel resolution 3 mm $\times$ 2 mm $\times$ 2 mm corresponding to axial/sagittal/coronal planes) to validate the robustness of our model, as in Table 9. Our model FCGAN mechanism, L_{mc}, and L_{sdv} achieve improvements independently in this different region/resolution, e.g., SDE has been reduced from 7.64 to 6.34 by FCGAN mechanism, and further reduced 9% by L_{sdv}, 5% by L_{mc}, respectively. Compared with Table 1 and 9, our method improves MAE more in the context of more focused (i.e., higher dose) regions with higher resolutions. It is reasonable since less interesting locations with lower doses in larger regions are usually easy to predict for all methods, which can reduce the percentage of our improvements.

In brief, our model is robust to different region/resolution, and have the potential to be adapted to different purposes.

I. Supplements of MC Loss

Clarification. We term Condition i as target condition, and Condition j as the support condition in Eq. 4. To simplify, Eq. 4 is only for one subject. The target condition i can be missing in training, e.g., plates condition in RT dose prediction, and the support condition j is from other observed conditions. $G(\cdot|m^i = 0, \cdot)$ is the generated sample when Condition i is missing, while Condition i for $G(\cdot)$ can be either given or missing.

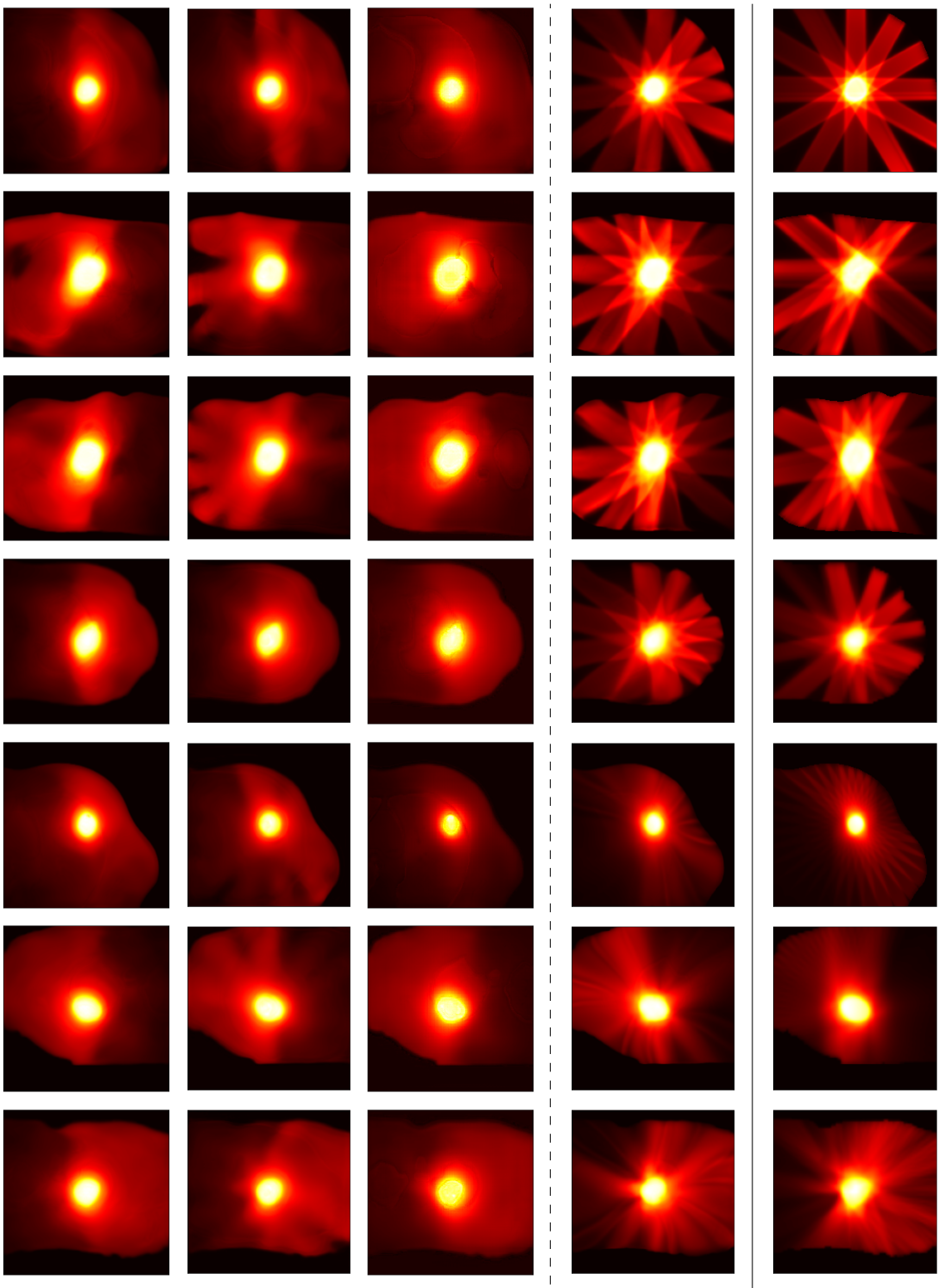


Figure 13. More qualitative results. Our model have realistic prediction, while baselines have been over-fitted to Gaussian-like distributions (make the following steps e.g., fluence map prediction harder). Since no angles has been specified during prediction, it is reasonable that the angles expanded slightly different between prediction and reference.

Characteristics	Values	D2NA	D1WA
Sex	Female/Male	52% / 48%	30% / 70%
Age at enrolment	Median (range), years	74 (31-97)	69 (39-91)
BMI	F: Mean (standard deviation)	27.4 (7.6)	25.8 (6.2)
	M: Mean (standard deviation)	27.2 (6.1)	26.6 (4.8)
Smoking status	Current	21.4%	40.3%
	Former	72.2%	54.9%
	Never	6.4%	4.8%
Family lung cancer history	positive percentage	-	18%
Tumor histology	Squamous	36%	33%
	Adenocarcinoma	38%	37%
	small cell	2%	4%
	others	24%	26%
Clinical tumor size stage	0	1%	2%
	1a, 1b	65%	34%
	2a, 2b	26%	26%
	3 or 4	9%	36%
Tumor overall stage	I-II	85%	44%
	III, IV	15%	56%

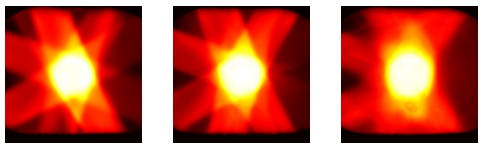
Table 7. More detailed information about the two datasets (D2NA and D1WA).

Models	SDE(↓)	DDE(↓)	MAE(↓)	CEL (↓)
UNet [8]	6.20	1.56	1.47	2.55
CUNet [7]	6.10	1.62	1.46	2.30
PUNet [6]	5.67	1.61	**1.40**	2.51
SwinUnetr [10]	6.34	1.61	1.45	2.61
DRUNet (ours)	5.74	1.55	**1.40**	2.53
FCGAN^{-} (ours)	**4.71**	**1.47**	1.49	**0.03**

Table 8. Experimental results on D2NA dataset with more baselines. FCGAN^{-} denotes no angle condition can be used.

Models	SDE (↓)	DDE(↓)	MAE(↓)	CEL (↓)
DRUNet	7.64	3.58	2.33	1.528
FCGAN$^{(-2)}$	6.34	1.66	**2.05**	0.013
FCGAN$^{(-1)}$	5.79	1.46	2.06	0.033
FCGAN	**5.51**	**1.39**	**2.05**	**0.008**

Table 9. Experiments on voxel resolution 3 mm × 2 mm × 2 mm of D1WA. Ablations on majority resolution. FCGAN$^{(-2)}$ has no L_{sdv} and no L_{mc}. FCGAN$^{(-1)}$ has no L_{mc}.



(a) reference (b) ours (c) w/o GAN

Figure 14. Ablation of GAN idea. Our model with adversarial training (b) generates more realistic samples. Without the adversarial loss, the prediction is more blurry as (c).

Justification. As mentioned in Sec. 3.2, the motivation of MC loss is to measure how predictions are consistent when one condition (from multiple conditions) is given versus missing. The consistency measurement is based on the observed conditions rather than the missing condition. Take face synthesis as an example (two conditions: w/ or w/o glasses as Condition j, sex as Condition i), the encoding of gender condition should be consistent whether the glass condition is given or missing, i.e., the sex prediction of a man with glasses and a man without glasses should both be male.

Summary. The MC loss is a regularization term and is self-explainable. It contributes to a robust prediction in the context of flexible multi-condition GAN, i.e., with potential missingness of conditions.

References

[1] Aaron Babier, Binghao Zhang, Rafid Mahmood, Kevin L. Moore, Thomas G. Purdie, Andrea L. McNiven, and Timothy C. Y. Chan. OpenKBP: The open-access knowledge-based planning grand challenge. *Medical Physics*, 2020. 4

[2] Ana María Barragán-Montero, Dan Nguyen, Weiguo Lu, Mu Han Lin, Roya Norouzi-Kandalan, Xavier Geets, Edmond Sterpin, and Steve Jiang. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. *Medical Physics*, 2019. 4

[3] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 1

[4] Vasant Kearney, Jason W. Chan, Samuel Haaf, Martina Descovich, and Timothy D. Solberg. DoseNet: a volumetric dose prediction algorithm using 3D fully-convolutional neural networks. *Physics in Medicine & Biology*, 2018. 4

[5] Vasant Kearney, Jason W. Chan, Tianqi Wang, Alan Perry, Martina Descovich, Olivier Morin, Sue S. Yom, and Timothy D. Solberg. DoseGAN: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation. *Scientific Reports*, 2020. 4

[6] Simon A.A. Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R. Ledsam, Klaus H. Maier-Hein, S. M. Ali Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A Probabilistic U-Net for Segmentation of Ambiguous Images. In *NeurIPS*, 2018. 2, 7

[7] Shuolin Liu, Jingjing Zhang, Teng Li, Hui Yan, and Jianfei Liu. Technical Note: A cascade 3D U-Net for dose prediction in radiotherapy. *Medical physics*, 2021. 2, 4, 7

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 7

[9] Petra Seibold, Adam Webb, Miguel E. Aguado-Barrera, David Azria, Celine Bourgier, Muriel Brengues, Erik Briers, Renée Bultijnck, Patricia Calvo-Crespo, Ana Carballo, Ananya Choudhury, Alessandro Cicchetti, Johannes Claßen, Elena Delmastro, Alison M. Dunning, Rebecca M. Elliott, Marie Pierre Farcy-Jacquet, Pietro Gabriele, Elisabetta Garibaldi, Antonio Gómez-Caamaño, Sara Gutiérrez-Enríquez, Daniel S. Higginson, Kerstie Johnson, Ramón Lobato-Busto, Meritxell Mollà, Anusha Müller, Debbie Payne, Paula Peleteiro, Giselle Post, Tiziana Rancati, Tim Rattay, Victoria Reyes, Barry S. Rosenstein, Dirk De Ruysscher, Maria Carmen De Santis, Jörg Schäfer, Thomas Schnabel, Elena Sperk, R. Paul Symonds, Hilary Stobart, Begoña Taboada-Valladares, Christopher J. Talbot, Riccardo Valdagni, Ana Vega, Liv Veldeman, Tim Ward, Christian Weißenberger, Catharine M.L. West, and Jenny Chang-Claude. REQUITE: A prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer. *Radiotherapy and Oncology*, 2019. 4

[10] Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. In *IEEE CVPR*, 2022. 7