IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 August 2025, accepted 17 August 2025, date of publication 20 August 2025, date of current version 4 September 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3600672

RESEARCH ARTICLE

Size-Controllable Tumor Synthesis for Improved Detection of Small Bowel Carcinoid Tumors in CT Scans

SEUNG YEON SHIN[1], STEPHEN A. WANK[2], AND RONALD M. SUMMERS[3]

[1]Division of Electrical Engineering, Hanyang University ERICA, Ansan, Gyeonggi-do 15588, South Korea
[2]Digestive Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[3]Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA

Corresponding author: Seung Yeon Shin (seungyeons@hanyang.ac.kr)

This work was supported in part by the research fund of Hanyang University (HY-2023-202300000002960); and in part by the Intramural Research Program of the National Institutes of Health (NIH), Clinical Center and NIDDK.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the local Institutional Review Board.

ABSTRACT Carcinoid tumors in the small bowel are rare, making it challenging to collect a sufficiently large dataset of lesions with diverse sizes for training detection models using computed tomography (CT) scans. This scarcity particularly affects the detection performance on relatively small or large tumors. To address this limitation, we propose a novel image synthesis method that can selectively augment underrepresented tumor sizes to enhance detection performance. Our method enables size-controllable tumor generation by integrating a tumor segmentation model and a size-aware loss into the training process. Specifically, one dimension of the input noise vector is designated to control the size of the synthesized tumors. These tumors, generated at desired sizes, are implanted into CT scans to enrich the training data for a tumor detection network. Our method produces tumors with clearer size distinctions, while maintaining comparable visual realism compared to a baseline synthesis method. In a visual Turing test, human observers could not reliably distinguish synthetic tumors from real ones. Lesion-level evaluation using free-response receiver operating characteristic (FROC) curves demonstrated that detection performance improved when synthetic tumors were included during training (P=0.04). This method offers a promising direction for improving the detection of rare tumors such as small bowel carcinoid tumors.

INDEX TERMS Lesion detection, image synthesis, controllability, small bowel, carcinoid tumor, computed tomography.

I. INTRODUCTION

Recent advances in artificial intelligence (AI) have shown many successes in medical imaging tasks including lesion detection [1], [2], [3], [4]. Deep learning has had a huge role in these successes and data on which deep learning models are trained is its essential ingredient. However, not all problems have a sufficient amount of training data, and it is more difficult to train a model regarding such problems. The quality of the dataset, e.g., diversity and availability of relevant labels, as well as its scale affects the performance of the trained model.

Data augmentation is a common practice to tackle data scarcity; thus, to improve the generalizability of deep learning models [5]. In training a model for object detection and instance segmentation, it is important to have not only as many training images as possible but also a sufficient number of objects within each image [6], [7]. It is difficult for the model to generalize to objects of low occurrence. For example, small objects can be more difficult to detect because of their insufficient occurrence in training. In [6], a simple copy-paste augmentation method was used to improve the

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

detection performance on small objects. The augmented number of small objects increases their contribution to training. In [8], for the task of monocular 3D detection, a more elaborate object insertion method was developed to copy virtual objects and paste them into real scenes. Plausible physical properties of objects, such as locations, sizes, and appearances, are automatically identified for realistic insertion. Compared to natural images, medical images have distinctly different properties; therefore, the aforementioned methods cannot be easily adopted. For example, the simple copy-paste approach [6], [7] may be inappropriate for projectional radiography like X-ray since it introduces total occlusion at the pasted location.

There have been distinct efforts on augmenting the number of objects of interest, e.g., lesions, for the detection and segmentation in medical images [9], [10], [11]. Lung lesions and liver tumors on computed tomography (CT) scans were synthesized respectively in [10] and [11] using a sequence of image-processing operations, e.g., ellipse-based shape generation and texture generation using a Gaussian noise. Being label- and training-free, those methods require elaborate tweaks of each component based on clinical prior knowledge. Furthermore, each of them is not directly applicable to other types of lesion and imaging modality.

Generative models such as generative adversarial nets (GANs) [12] and diffusion models [13] have also been used to synthesize lesions of interest in past years [14], [15], [16], [17], [18]. In [14], dermatology images of skin lesions were synthesized by sampling from the learned distribution of the data. In [15], relatively abundant *normal* magnetic resonance (MR) images were translated to *abnormal* images that contain brain lesions, using CycleGAN [19]. While being simple, both methods have no explicit controllability on the synthesized lesion. In [16], lung cancer CT images were generated using conditional GANs [20]. The size and shape of synthesized lesions can be controlled by varying input sketches. However, their human-drawn sketch is not scalable. Within a similar framework, algorithmically generated lesion masks were instead used as input to synthesize CT images with intracranial hemorrhage in [17]. On the other hand, lesion masks themselves were also generated using another network for retinal fundus images in [18].

Carcinoid tumor, which is of our interest in this work, is a rare neoplasm (small bowel neoplasms including carcinoid tumors account for 0.5% of all cancers and about 8000 cases per year in the United States [21], [22]). They are often less than a centimeter in size and the small bowel is one of the most common sites for them (24%-44%) [23]. In [3], [4], and [24], the detection of small bowel carcinoid tumors was performed on CT scans. While the presented result is clinically relevant, as a rare disease, it was difficult to obtain a large-scale database; thus, it showed room for improvement in the performance. Especially, despite being overall small, relatively smaller or larger tumors take up only a small portion of the dataset, showing relatively inferior detection performance on them than on mid-volume tumors. This raises the question of whether augmenting these underrepresented tumor sizes could enhance their own detection accuracy and improve overall performance.

In this paper, we thus synthesize small bowel carcinoid tumors on CT scans to improve their detection. Specifically, we develop a size-controllable synthesis method so that we can focus more on augmenting deficient relatively smaller or larger tumors. Since we are interested in carcinoid tumors in the small bowel, our synthesis is performed on top of the identified small bowel in a patch-wise manner. A normal local image patch on the small bowel is translated to one containing a carcinoid tumor of desired size. It then can be implanted into the original location to augment the number of tumors in each CT scan. We believe the patch-wise image translation is a proper choice considering: (1) the complex shape of the small bowel [25], [26], [27], [28], [29], (2) small size of carcinoid tumors (carcinoid tumors take only very few voxels within the small bowel), and (3) limited size of the training data.

The main contributions of our work are as follows. (1) We propose a new image synthesis method where the size of a target object contained is adjustable. (2) We develop a fully automatic framework to augment CT scans with synthetic lesions of intended sizes, based on the size-controllable synthesis method. (3) We use scans augmented with synthetic tumors to improve the detection of small bowel carcinoid tumors, which are rare diseases. Each of the contributions, which correspond to steps of the proposed method, is depicted in Fig. 1.

II. DATASET

We used the same dataset used in [3]. Institutional review board approval was obtained for retrospective analysis of the dataset. The dataset is composed of CT scans of tumor-positive ($n = 33$) or negative ($n = 22$) patients (Table 1). Only one most relevant scan was used per patient. The scans are all intravenous and oral contrast-enhanced. VoLumen oral contrast was administered. Each scan covers the entire small bowel.

TABLE 1. Dataset composition. The number of scans, and the types of available grount-truth (GT) labels are presented for each subset. There is no patient overlap between subsets. Refer to the text for a description of each subset.

Subset	Tumor	# of scans (patients)	Available GT labels
Trainval	Positive	24	Tumor & small bowel segmentation
Test	Positive	9	Tumor segmentation
	Negative	22	Small bowel segmentation

The tumor-positive patients, who had at least one carcinoid tumor within the small bowel, all underwent surgery and their preoperative CT scans were used. Ground-truth (GT) segmentation of tumors was drawn for each CT scan by referring to an available ^{18}F-DOPA PET scan and the corresponding radiology report, resulting in 162 annotated tumors in total. Please refer to [3] for the distribution of real tumor volumes in our dataset. These CT scans from the

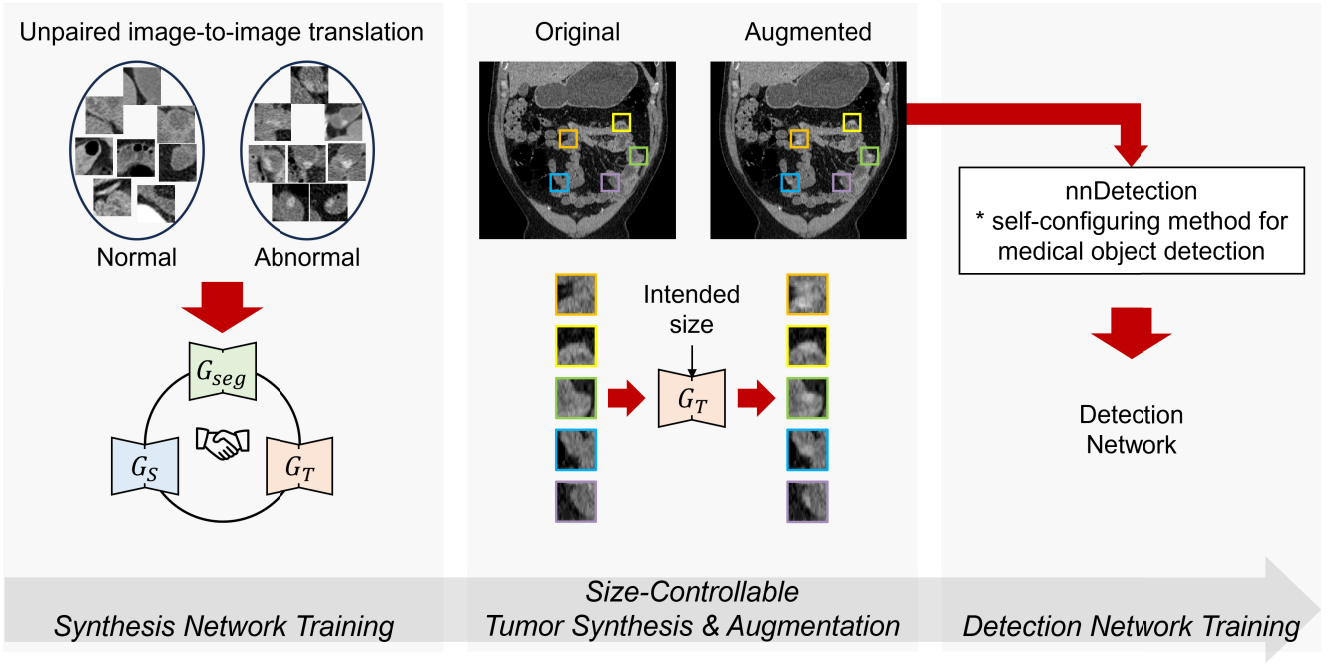


FIGURE 1. Flow of the proposed method. The component titled *Synthesis Network Training* corresponds to Sections III-A,III-B,III-C. The subsequent two components correspond to Sections III-D and III-E, respectively.

tumor-positive patients were split into two subsets, each of which is the *trainval set* ($n = 24$) and the *test positive set* ($n = 9$). For each of the scans in the *trainval set*, GT segmentation of the small bowel was also achieved. The *trainval set* is used for training and validating both the tumor synthesis network and the detection network.

The tumor-negative patients had no evidence of small bowel carcinoid tumors from both the CT and PET scans. Their CT scans compose the *test negative set* ($n = 22$). The CT scans in the *test negative set* have void tumor segmentation masks. GT segmentation of the small bowel was also achieved for the experiment presented in the Appendix.

All scans were resampled to have isotropic voxels of $1 \times 1\text{x}1\ \text{mm}^3$ before used. We note that creating a large dataset for rare diseases is difficult. Please refer to [3] for more details of the dataset, including the characteristics of the patients and scans, and the annotation process.

III. METHODS

A. MULTI-MODAL UNPAIRED IMAGE-TO-IMAGE TRANSLATION

We synthesize small bowel carcinoid tumors by using an unpaired image-to-image translation model. CycleGAN [19] provides a basic function for this, but it assumes one-to-one mappings by its cycle consistency, i.e., when translating from one domain to another and back again, it should arrive at where it started, so that it can generate only a single output image from each input image in a deterministic way. To generate multiple abnormal image patches with varying tumor sizes from a single normal image patch, we adopt ACL-GAN (adversarial-consistency loss GAN) [30], which supports multi-modal outputs, as the baseline synthesis method. It will be made size-controllable using the proposed method described in Section III-B.

Fig. 2 presents the overall network architecture of our method, incorporating ACL-GAN as a part. ACL-GAN is composed of two generators G_S, G_T and three discriminators D_S, D_T, $\hat{D}$. The generators $G_S : (x, z) \rightarrow x_S$ and $G_T : (x, z) \rightarrow x_T$ translate an image to domain S and T, respectively. Source and target domain images, $x_S \in X_S$ and $x_T \in X_T$, are normal and abnormal images, respectively, in this work. The set X denotes the union of X_S and X_T, i.e., $X = X_S \cup X_T$. The noise vector $z \in Z$ is sampled from the noise vector space Z. We use p_a and $p_{(a,b)}$ to denote the distribution of a and the joint distribution of pair (a, b).

The classical adversarial loss for the generator G_T and discriminator D_T is defined as:

$$L^T_{adv}(G_T, D_T, X_S, X_T) = \mathbb{E}_{x_T \sim p_T}[\log D_T(x_T)] + \mathbb{E}_{\bar{x}_T \sim p_{\{\bar{x}_T\}}}[\log(1 - D_T(\bar{x}_T))], \quad (1)$$

where $\bar{x}_T = G_T(x_S, z_1)$. The objective is $\min_{G_T} \max_{D_T} L^T_{adv}(G_T, D_T, X_S, X_T)$. Similarly, the loss function for the generator G_S and discriminator D_S is defined as:

$$L^S_{adv}(G_S, D_S, \{\bar{x}_T\}, X_S) = \mathbb{E}_{x_S \sim p_S}[\log D_S(x_S)] + (\mathbb{E}_{\hat{x}_S \sim p_{\{\hat{x}_S\}}}[\log(1 - D_S(\hat{x}_S))] + \mathbb{E}_{\tilde{x}_S \sim p_{\{\tilde{x}_S\}}}[\log(1 - D_S(\tilde{x}_S))])/2, \quad (2)$$

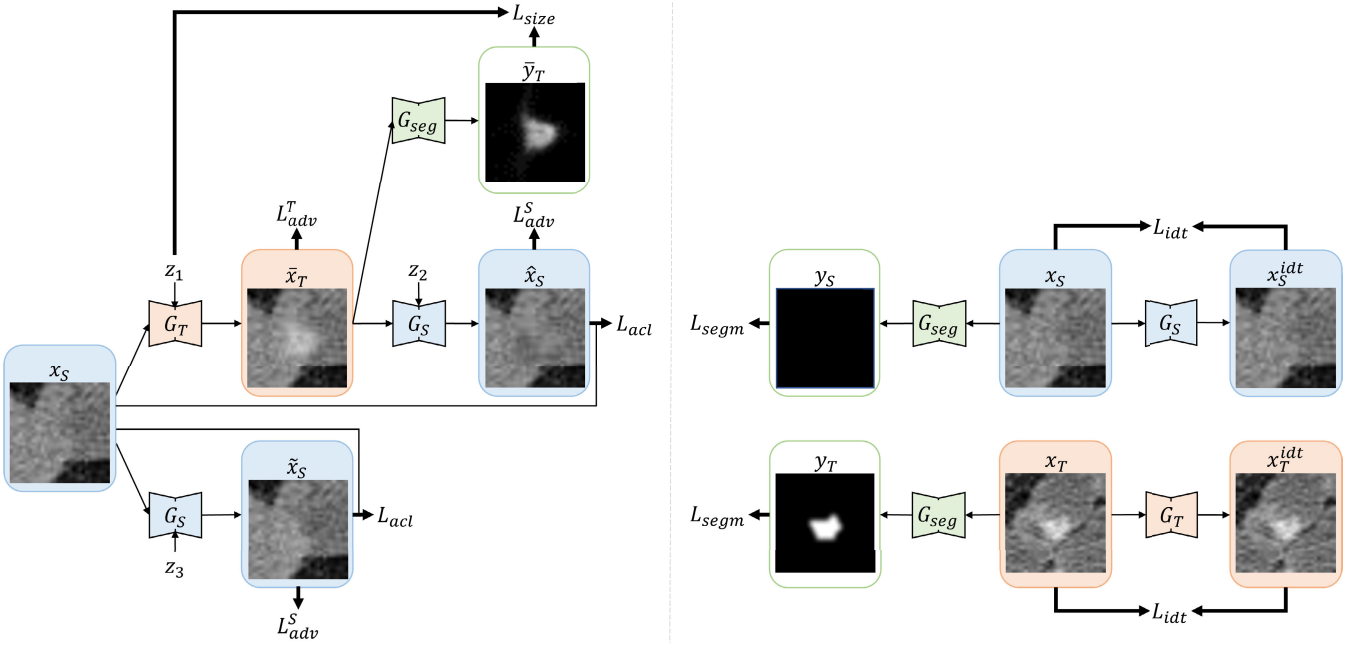


FIGURE 2. Networks in the proposed method, including two generators: $G_S : (X, Z) \rightarrow X_S$ and $G_T : (X, Z) \rightarrow X_T$, and three discriminators: D_S, D_T for L^S_{adv}, L^T_{adv} and $\hat{D}$ for L_{acl}. The discriminators are not shown here for brevity. They were adopted from ACL-GAN and the input noise vectors z_1, z_2, z_3 to the generators enable multi-modal outputs. The input noise vectors are sampled from $\mathcal{U}[-3, 3]$.L_{idt} is to encourage the generators to generate identical images when target domain images are given as input with no use of random noise vectors. To make the synthesis size-controllable, we incorporate a pretrained tumor segmenter G_{seg}, which is trained using L_{segm}, and the size loss L_{size} in training. Each color represents the same network or domain. Refer to the text for detailed explanations of each network and the associated loss functions.

where $\hat{x}_S = G_S(\bar{x}_T, z_2)$ and $\tilde{x}_S = G_S(x_S, z_3)$. Two types of generated images, $\hat{x}_S$ and $\tilde{x}_S$ are involved. The objective is $\min_{G_S} \max_{D_S} L^S_{adv}(G_S, D_S, \{\bar{x}_T\}, X_S)$.

Furthermore, the adversarial-consistency loss is used to minimize the distances between the joint distributions of $(x_S, \hat{x}_S)$ and $(x_S, \tilde{x}_S)$. It is defined as:

$$L_{acl} = \mathbb{E}_{(x_S,\hat{x}_S)\sim p_{(X_S,\{\hat{x}_S\})}}[\log \hat{D}(x_S, \hat{x}_S)] + \mathbb{E}_{(x_S,\tilde{x}_S)\sim p_{(X_S,\{\tilde{x}_S\})}}[\log(1 - \hat{D}(x_S, \tilde{x}_S))]. \quad (3)$$

This encourages the image translated back, $\hat{x}_S$, to retain important features of the source image, x_S, thereby promoting the preservation of those features in the 'intermediate' translated image, $\bar{x}_T$, as well.

In addition, the identity loss and the mask loss are also used. The identity loss, L_{idt}, is to encourage the generators to generate identical images when target domain images are given as input. It stabilizes the training process and further encourages feature preservation. It is defined as:

$$L_{idt} = \mathbb{E}_{x_S\sim p_S}[||x_S - x_S^{idt}||_1] + \mathbb{E}_{x_T\sim p_T}[||x_T - x_T^{idt}||_1], \quad (4)$$

where $x_S^{idt} = G_S(x_S, E^z_S(x_S))$, $x_T^{idt} = G_T(x_T, E^z_T(x_T))$, and E^z_S, E^z_T are the noise encoders of G_S and G_T, respectively. Each generator consists of two encoders, namely, an image encoder and a noise encoder, and one decoder. While the image encoder and decoder form an auto-encoder architecture, the noise encoder is similar to the style encoder in [31]. Given an input image x, the output of the image encoder and that of the noise encoder are forwarded to the decoder. When the outputs of both encoders are passed through without modification, the decoder is expected to reconstruct an image identical to the input. Then, it is used for the identity loss. If the output of the noise encoder is replaced by a random noise vector z, a new translated image is generated, which is then used for other losses.

The generators make two-channel outputs consisting of one image channel and the so-called bounded focus mask. The mask has values between 0 and 1, and the final translated image is obtained as $x_T = x'_T \odot x_m + x_S \odot (1 - x_m)$. Here, x_S denotes the source image, while x'_T and x_m represent the image and mask channels, respectively. The operator $\odot$ indicates element-wise multiplication. The mask loss implements a few constraints for the mask and it is defined as:

$$L_{mask} = \delta[(\max\{\sum_k x_m[k] - \delta_{\max} \times W, 0\})^2 + (\max\{\delta_{\min} \times W - \sum_k x_m[k], 0\})^2] + \sum_k \frac{1}{|x_m[k] - 0.5| + \epsilon}, \quad (5)$$

where $\delta_{\min}$ and $\delta_{\max}$ are hyper-parameters to set the minimum and maximum size of mask. The variable W denotes the number of voxels in an image, and k indexes the voxels. Therefore, the first two terms encourage the generator not only to make enough changes but also to preserve the background. The coefficient δ adjusts the relative influence of these two terms. The last term encourages the mask values

to be either 0 or 1. The constant ϵ is included to avoid division by zero.

When translating normal image patches extracted around the small bowel to abnormal, a synthesized tumor should be situated on the small bowel. In other words, only the small bowel region is encouraged to change during the translation within the patch. Otherwise, it can cause artificiality when placed back into the original location in a CT scan. We implement this constraint as a loss term and it is defined as:

$$L_{mask}^{sb} = -(1 - b_S)\log(1 - x_m), \tag{6}$$

where b_S is the GT small bowel segmentation of x_S. It is applied to x_S being translated to $\bar{x}_T$, and forces x_m to be 0 for the non-small-bowel region. L_{mask} and L_{mask}^{sb} are normalized by W in the end.

B. SIZE-CONTROLLABLE TUMOR SYNTHESIS

To make the synthesis size-controllable, we incorporate the pretrained tumor segmenter G_{seg} and the size loss L_{size} in training, as shown in Fig. 2. For the tumor segmenter, we use a network architecture, which is similar to that of the patch-based method in [3]. Since the network used in [3] is a double-branch network for joint classification and segmentation of tumor, we take only the segmentation branch. In summary, it has a U-Net structure [32]. Please refer to [3] for more details of the network architecture. Our tumor segmenter is pretrained using both normal (source, x_S) and abnormal (target, x_T) image patches, which are expected to present the empty and non-empty segmentations, respectively. The generalized Dice loss [33] is used as done in [3]. After the pretraining was done, an average Dice coefficient of 0.837 was achieved for abnormal image patches from the *trainval set*.

In training the whole network, while G_{seg} keeps being trained using x_S and x_T, it also predicts tumor segmentation, $\bar{y}_T$, for the synthesized abnormal image $\bar{x}_T$. The size (volume) of the synthesized tumor within $\bar{x}_T$ is measured as: $V_{\bar{y}_T} = \sum_k \bar{y}_T[k]$. We apply the size loss [34] to the computed size, which is defined as:

$$L_{size} = \begin{cases} (V_y - a)^2, & \text{if } V_y < a \\ (V_y - b)^2, & \text{if } V_y > b \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where a and b are lower and upper bounds on the size. L_{size} is normalized by W, the number of voxels in an image, in the end. We note that the size loss is meaningless with inaccurate measurement of the tumor size; therefore, we start the training of the whole network with the pretrained tumor segmenter G_{seg}.

When the synthetic abnormal image $\bar{x}_T$ is generated using the generator G_T, a random noise vector z_1, sampled from $\mathcal{U}[-3, 3]$, is inputted. It provides the generator with randomness, enabling multi-modal outputs, but there is no clear meaning of each dimension of the noise vector at the beginning of training. To use a particular dimension of the vector as a size-controller, i.e., the first dimension in this work, we linearly map its range $[-3, 3]$ into a desired size range of $[V_y^{min}, V_y^{max}]$. The value of the first dimension determines the desired size V_y^{target} of synthesized tumor. We use $V_y^{target} \times 0.9$ and $V_y^{target} \times 1.1$ for a and b in Eq. (7), respectively. In test time, a desired size is mapped back to the value in $[-3, 3]$, and then it composes the first dimension of z_1.

C. TRAINING OF SYNTHESIS NETWORKS

The total loss for training the proposed network is as follows:

$$\begin{aligned} L_{total} = {} & L_{adv}^T + L_{adv}^S + \lambda_{acl}L_{acl} + \lambda_{idt}L_{idt} + \lambda_{mask}L_{mask} \\ & + \lambda_{mask}^{sb}L_{mask}^{sb} + \lambda_{size}L_{size}, \end{aligned} \tag{8}$$

where λ_{acl}, λ_{idt}, λ_{mask}, λ_{mask}^{sb}, and λ_{size} control the influence of each loss term. Each term is introduced in Eqs. (1)–(7), respectively.

D. AUGMENTING CT SCANS WITH SYNTHETIC TUMORS

In the proposed method, G_T is trained in cooperation with G_S and G_{seg}. Once the training is done, G_T is used to synthesize abnormal image patches, which contain a tumor of a desired size, from normal patches. Normal image patches to be translated are sampled from scans in the *trainval set* with several criteria, which are: (1) they must be on the small bowel, (2) they must not overlap existing tumors, (3) they must not overlap gas bubbles trapped in the small bowel, and (4) they must not overlap each other. We utilize the GT tumor and small bowel segmentations in the *trainval set* to meet conditions (1) and (2). For condition (3), the gas bubbles are segmented on CT scans using a threshold value of ≤ -200.

In this work, we focus on synthesizing relatively more deficient smaller or larger tumors, for which the detector in our previous work [3] showed inferior performance. We used -3 or 3 for the first dimension of z_1 to generate the small or large tumors, respectively. The other dimensions remain random. Once translated to abnormal image patches, they are implanted into the original location in each CT scan. As a measure to ensure that a tumor is generated during the translation, the tumor segmenter G_{seg} is used to segment the synthesized image patches. Patches with no valid tumor, namely being smaller than V_y^{min}, are regarded as failure cases and discarded in the implantation. V_y^{min} is the lower bound of the synthesized tumor size in voxels, as described in Section III-B. The mentioned process is done fully automatically.

E. DETECTION NETWORK TRAINING

We use nnDetection [2] to train detectors as in our previous work [3]. Retina U-Net [1] is used as base network architecture. nnDetection is a self-configuring method for medical object detection, which automatically determines optimal parameters relating to training and inference. We believe it is suitable to highlight the difference between using a

purely real dataset and using a dataset augmented with synthetic tumors in terms of detection performance. We use the best-performing checkpoint during training.

F. EVALUATION DETAILS

We use the *trainval set* to train both the networks for tumor synthesis and detection. The *test set* is reserved solely for the evaluation of detection performance.

In training the synthesis network, we used a five-fold cross validation, where 24 scans were randomly divided into five subsets of 5/5/5/5/4 scans. In each fold training, one of the subsets was used for validation and the remaining were used for training. After the training, each trained model can be used to translate normal image patches to abnormal for scans in each respective validation subset.

Among the five trained models (folds), the one trained with 20 scans, is used to actually augment CT scans with synthetic tumors, as described in Section III-D. Finally, the detection network is trained using those 20 scans that now contain synthetic tumors. The corresponding original scans, which contain only real tumors, are not used in this training. This is to see the effect of the increased number of tumors while maintaining the number of scans used in training. The effect of increasing the number of training scans through augmentation of tumor-free scans with synthetic tumors is further investigated in the Appendix. The remaining 4 scans are used as a validation set in training the detection network.

For the synthesis network, a patch size of $32 \times 32 \times 32$ and a batch size of 3 were used. For Eq. (8), λ_{acl}, λ_{idt}, λ_{mask}, λ^{sb}_{mask}, and λ_{size} were set as 0.2, 1, 0.2, 0.8, and 0.001, respectively. We used 0.001 and 0.01 for δ and ϵ of Eq. (5). The variables $\delta_{\min}$ and $\delta_{\max}$ denote the minimum and maximum ratios of voxels that are allowed to change in an image patch during translation, namely, the ratios of foreground voxels; therefore, they were set by referring to the relevant statistics on the minimum and maximum tumor volumes in our dataset. We used 0.00025 and 0.1 for $\delta_{\min}$ and $\delta_{\max}$, respectively. Based on these values, we set V_y^{min} and V_y^{max} to $32 \times 32 \times 32 \times \delta_{\min}$ and $32 \times 32 \times 32 \times \delta_{\max}$, respectively.

The synthesis network was trained for 200000 iterations with an initial learning rate of 2×10^{-5} and a weight decay of 10^{-4}. The learning rate was reduced to 5×10^{-6} at the 100000^{th} iteration, while using the Adam optimizer. For data augmentation, image mirroring, rotation, and scaling were used. The hyper-parameters were chosen empirically through the grid search unless otherwise mentioned. We used the same hyper-parameters as those used in [3] when training the detection network.

The objective of this work is to have a size-controllability in synthesizing a tumor while maintaining its quality. To evaluate the quality of synthesized images, we first used Fréchet Inception Distance (FID) [35] and Kernel Inception Distance (KID) [36]. They calculate the distance between the real and the synthetic image distributions based on feature representations. The Inception network trained on natural image datasets is commonly used to extract the features of the real and synthetic images. To better match the medical imaging data, we instead used TotalSegmentator [37] trained for the segmentation of many organs, including the small bowel, on CT scans. In addition, we used the Precision and Recall for Distributions (PRD) [38] to measure both the quality and the diversity of synthesized images. The precision and recall respectively quantify the quality and the diversity, which are in a trade-off relationship. The area under the PRD curve (PRD-AUC) is reported.

We also performed a visual Turing test to qualitatively evaluate the quality of synthesized images. A radiologist and a gastroenterologist, both with 30+ years of experience, were provided with a set of randomly shuffled 20 real and 20 synthetic images, and were asked to classify them. To check the size-controllability, we plotted volume histograms of tumors, which had been generated using different values for the first dimension of z_1.

The detection performance was evaluated using the same method as done in [3]. Free-response receiver operating characteristic (FROC) curves and receiver operating characteristic (ROC) curves were used for lesion- and patient-level evaluations, respectively. For each ROC curve, the AUC and sensitivity/specificity at the Youden point were computed. We used the bootstrapping method introduced in [39] to perform a statistical analysis comparing the FROC and ROC curves.

We implemented the proposed synthesis network based on the code for ACL-GAN at https://github.com/hyperplane-lab/ACL-GAN using PyTorch 1.13.1. To train the detection network, the implementation of nnDetection at https://github.com/MIC-DKFZ/nnDetection was used as described in Section III-E. The experiments were performed in an environment with Intel Xeon Gold 6130 at 2.1GHz CPU, 32GB RAM, and NVIDIA Tesla V100 32GB GPU.

IV. RESULTS

A. RESULTS OF TUMOR SYNTHESIS

1) QUALITY OF SYNTHESIS

Table 2 presents the quantitative evaluation of tumor synthesis quality of different methods. While the compared methods showed almost no difference in FID and KID, the PRD-AUC value was increased (0.8118 $\rightarrow$ 0.8348) when the size loss L_{size} was incorporated to grant the size-controllability. The PRD curve and its summary metric, PRD-AUC, quantify the quality and diversity of tumor synthesis at the same time. The PRD-AUC value was further increased (0.8348 $\rightarrow$ 0.8382) when L^{sb}_{mask} was incorporated in training to encourage that changes are made only within the small bowel during the image translation.

Table 3 shows the result of visual Turing test. The answer, 'unsure' was included to roughly show the confidence of the observers in this test. When the synthetic and the real were considered positive and negative, respectively, the

TABLE 2. Quantitative evaluation of tumor synthesis quality. 'ACL-GAN' is the baseline synthesis method, and 'Proposed-S' is a variant of the proposed method, which is trained without L^{sb}_{mask}. While lower is better for FID and KID, higher is better for PRD-AUC. Refer to the text for an explanation of each metric.

Method	L^{sb}_{mask}	L_{size}	FID	KID (mean ± std)	PRD-AUC
CycleGAN [19]	-	-	1.144×10^{-3}	$7.195 \times 10^{-7} \pm 3.47 \times 10^{-8}$	0.7614
MUNIT [31]	-	-	1.097×10^{-3}	$6.922 \times 10^{-7} \pm 3.55 \times 10^{-8}$	0.8021
HIT [40]	-	-	1.086×10^{-3}	$6.891 \times 10^{-7} \pm 3.41 \times 10^{-8}$	0.8078
ACL-GAN [30]	-	-	1.060×10^{-3}	$6.525 \times 10^{-7} \pm 3.32 \times 10^{-8}$	0.8118
Proposed-S	-	✓	1.061×10^{-3}	$6.694 \times 10^{-7} \pm 3.58 \times 10^{-8}$	0.8348
Proposed	✓	✓	1.059×10^{-3}	$6.895 \times 10^{-7} \pm 4.08 \times 10^{-8}$	**0.8382**

radiologist achieved a sensitivity of 52.9% and a specificity of 55.0%, except for the unsure images. It represents the difficulty of distinguishing between real and synthetic tumors. The gastroenterologist, who has reviewed more than 500 CT scans of patients with proven small bowel carcinoid tumors, achieved better accuracy. We further compared the classifications by the two observers using Bowker's Test for Symmetry [41] ($P = 0.37$). This result indicates that their classifications are not significantly different and can be considered statistically consistent. Fig. 3 shows examples of the real and synthetic tumors, used in the visual Turing test.

TABLE 3. Result of visual Turing test. A senior radiologist and a senior gastroenterologist were given a set of randomly shuffled 20 real and 20 synthetic tumor images, and were requested to classify them. The synthetic tumors were generated within scans from the *validation set* (4 out of 24 scans in the *trainval set*). Refer to the text for more details of this experiment.

		prediction					
		radiologist			gastroenterologist		
		real	synthetic	unsure	real	synthetic	unsure
truth	real	11	9	0	13	7	0
	synthetic	8	9	3	6	13	1

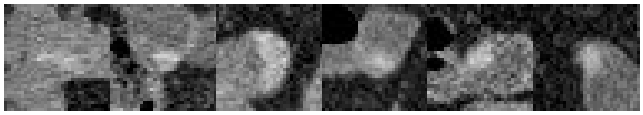


FIGURE 3. Examples of real (left three) and synthetic (right three) tumors, used in the visual Turing test.

2) SIZE-CONTROLLABILITY

As described in Section III-B, we used the first dimension of the input noise vector z_1 as the size-controller. The expected size of a synthesized tumor increases as the value for the size-controller dimension increases within the range $[-3, 3]$. Fig. 4 shows histograms of tumor volumes, which were generated using the baseline (ACL-GAN) and the proposed method. For each input normal patch, the synthesis was done 15 times using different noise vectors z_1. Each three take a different value in $\{-3, -1.5, 0, 1.5, 3\}$ for the size-controller dimension. Finally, randomly sampled 50 input normal patches were used, resulting in 150 syntheses per each of those five values. The pretrained tumor segmenter G_{seg} was used to segment tumors and thus to measure their volumes.

The baseline method showed no clear difference between the five histograms, each of which was obtained using a different value for the size-controller dimension. On the other hand, the proposed method showed higher dissimilarity between histograms, especially ones for the smallest (red) and the largest (blue) scales. They showed a chi-square distance of 83.2 while that of the baseline method is 3.5. Fig. 5 shows example synthetic tumors generated using the method described above.

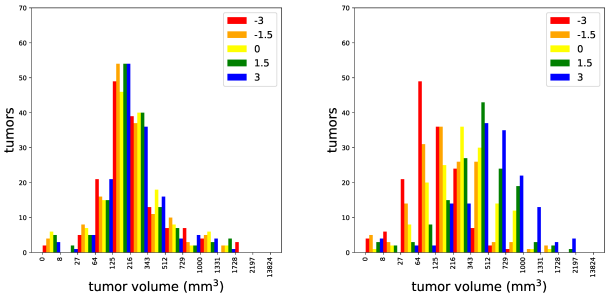


FIGURE 4. Histograms of tumor volumes generated using the baseline (left) and the proposed method (right). The size-controller dimension of the input noise vector z_1 was varied to one in $\{-3, -1.5, 0, 1.5, 3\}$ to generate from smaller tumors to larger ones. Please see the text for more details.

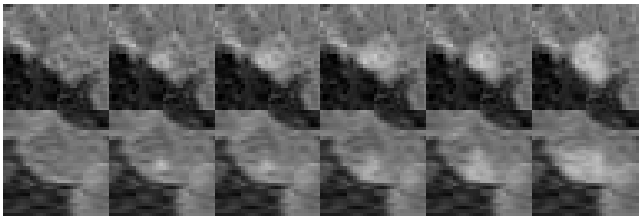


FIGURE 5. Example synthetic tumors of different sizes. The first column of each row represents a different input normal patch, and the followings are generated abnormal patches. They were intended to contain larger and larger tumors when moving on to the right. The same method described in the caption of Fig. 4 was used.

B. RESULTS OF TUMOR DETECTION

1) QUANTITATIVE RESULTS

Once training of the synthesis network is done, it is used to augment scans in the *training set* with synthetic tumors, as described in Section III-D. Fig. 6 compares FROC curves

of different methods, including the proposed method with varying numbers of synthetic tumors per scan. 'Proposed (5)' showed better performance not only compared to not using the augmentation ('No augmentation', P = 0.04), but also compared to the version with fewer synthetic tumors ('Proposed (2)', P = 0.13). However, increasing the number of synthetic tumors beyond this did not lead to further performance improvement. When 15 synthetic tumors were implanted per scan, 'Proposed (15)', it showed an even worse performance (P = 0.25). 'Proposed (5)' also showed a better performance (P = 0.08) than one based on ACL-GAN, which is not size-controllable.

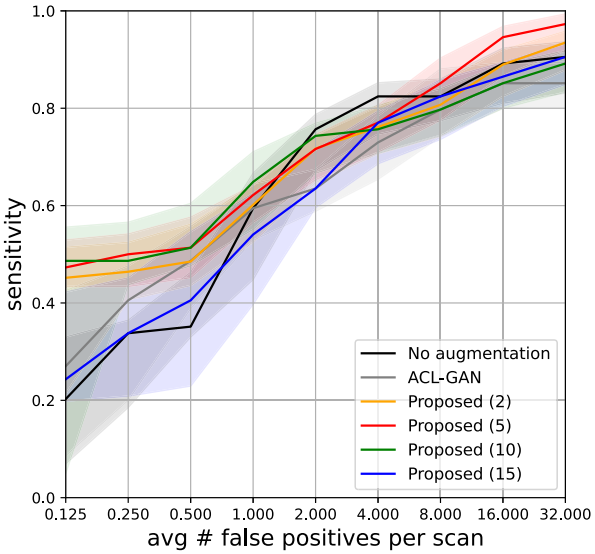


FIGURE 6. FROC curves of the detection network in the *test positive set*, with 95% confidence interval. For training, scans in the *training set* (20 out of 24 scans in the *trainval set*) were augmented with different numbers of synthetic tumors. For example, five synthetic tumors were implanted per each training scan for the model 'Proposed (5)'. Bootstrapping by randomly sampling test images with replacement is used to generate multiple FROC curves for each model. From these, 95% confidence intervals are computed.

Fig. 7 shows the detection performance with respect to tumor volume. The performance of 'No augmentation' and 'Proposed (5)' (in Fig. 6) were compared using FROC curves for each volume range. The effect of using synthetic tumors is maximized for the largest volume range (>=343 mm^3). Sensitivity values of (37.5%→50.0%, 50.0%→50.0%, 62.5%→62.5%, 62.5%→75.0%, 62.5%→87.5%, 87.5%→100.0%) were achieved at per-scan false positive rates of (0.5, 1, 2, 4, 8, 16), respectively. It aligns with the goal of this work, which is to focus on synthesizing relatively more deficient smaller or larger tumors and thus to improve the detection performance on them.

Fig. 9 shows ROC curves for the evaluation of patient-level detection. Nine positive and 22 negative patients in the *test set* were used. For each method, we additionally used the predicted small bowel segmentation to eliminate false positives outside the small bowel as in [3], which is known to enhance the patient-level detection. We found that the proposed method of utilizing synthetic tumors for training helps detect more true positives at the cost of increased false positives, resulting in a worse patient-level detection. The method that does not use the tumor synthesis but use the small bowel segmentation, namely 'No augmentation (w/ SB segm)', performed the best (P<0.01 against all the other methods).

2) QUALITATIVE RESULTS

Fig. 8 shows example detection results of ones without and with the augmentation, namely 'No augmentation' and 'Proposed (5)' in Fig. 6. The proposed method detects more true positives at the cost of increased false positives.

V. DISCUSSION

The goal of our tumor synthesis is to have size-controllability while maintaining quality. While the quality was evidenced quantitatively and qualitatively in Tables 2 and 3, respectively, the proposed method showed a better size-controllability than the baseline in Fig. 4. The proposed augmentation method, which increases the number of tumors in each scan using synthetic tumors, is fully automatic. No manual process, such as drawing a tumor mask, is involved in synthesizing and implanting tumors. Once a site for implantation is selected using the method described in Section III-D, an extracted image patch, together with a sampled noise vector, is fed into the network for translation. The value to the size-controller dimension constrains the size of a synthesized tumor. As the value increases, the expected size of a synthesized tumor increases. Although it allows the method to remain fully automatic, the value is chosen irrespective of the input image. In reality, applicable tumor sizes for synthesis depend on the content of an input patch, such as the bowel size and the presence of a gas bubble. For example, a large tumor should not be generated on a thin part of the small bowel to remain natural. It may hinder achieving a better size-controllability. Thus, we used a relaxed range for size in Eq. (7). Nevertheless, the proposed method showed higher size-controllability than the baseline.

In terms of detection, we found that: 1) implanting too many synthetic tumors did not help improve the performance when the number of real tumors for training is limited (Fig. 6). Although the quality of our tumor synthesis was favorably evaluated both quantitatively and qualitatively, there can be unnoticeable artifacts, which may result in overfitting when too many synthetic tumors are given. 2) The proposed method outperformed the baseline in the regime of low false positive while it performed on par otherwise (Fig. 6), and 3) the proposed method detects more true positives at the cost of increased false positives, resulting in a worse patient-level detection. Despite the tradeoff, the proposed method offers clinical value by assisting radiologists in CT examinations for detecting small tumors, specifically by increasing per-lesion sensitivity. Our analysis showed that false positives typically occur near enhanced regions, such as thickened bowel walls,

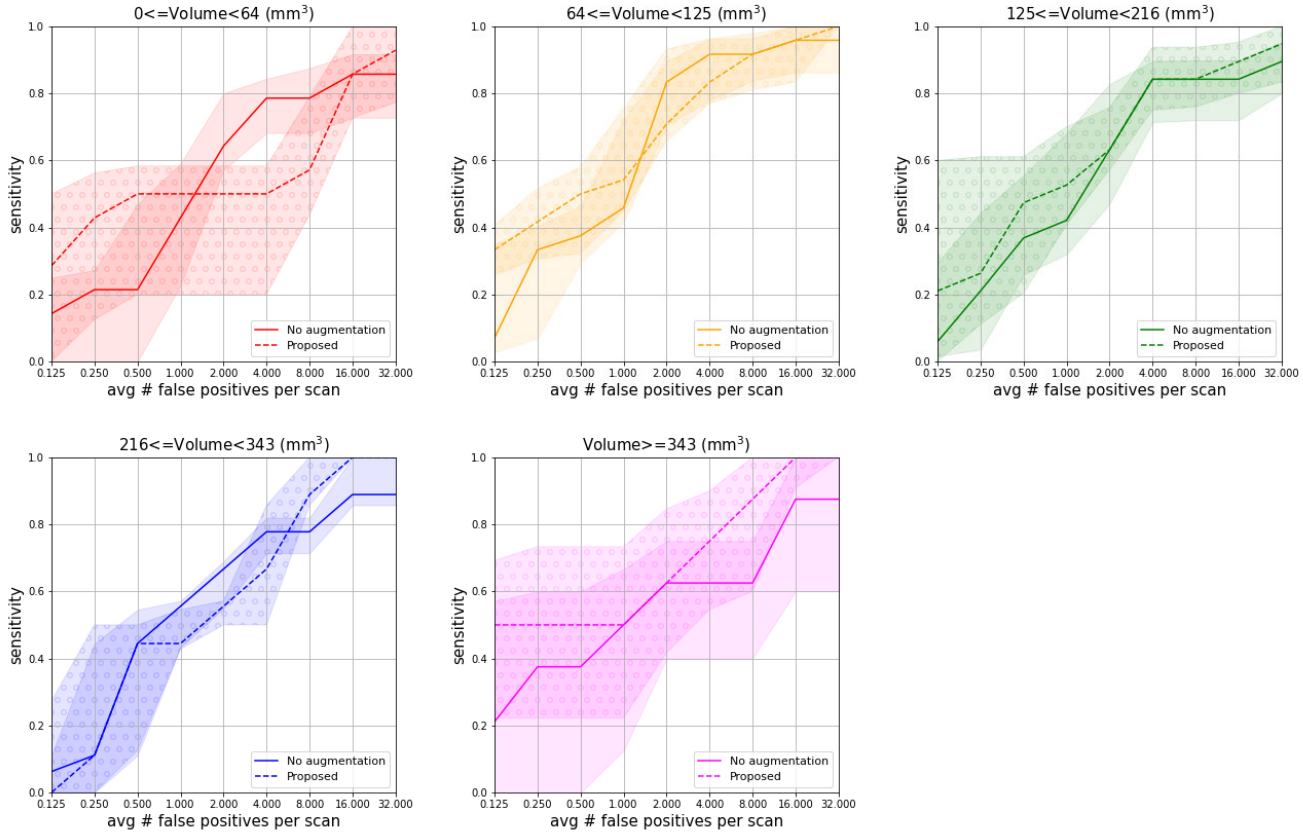


FIGURE 7. FROC curves in the *test positive set*, with respect to tumor volume. The results with and without the augmentation using synthetic tumors are compared for each volume range. Bootstrapping is used to generate multiple FROC curves from which 95% confidence intervals are computed. The confidence intervals for the proposed method are marked with the hatch pattern of circles for better differentiation.

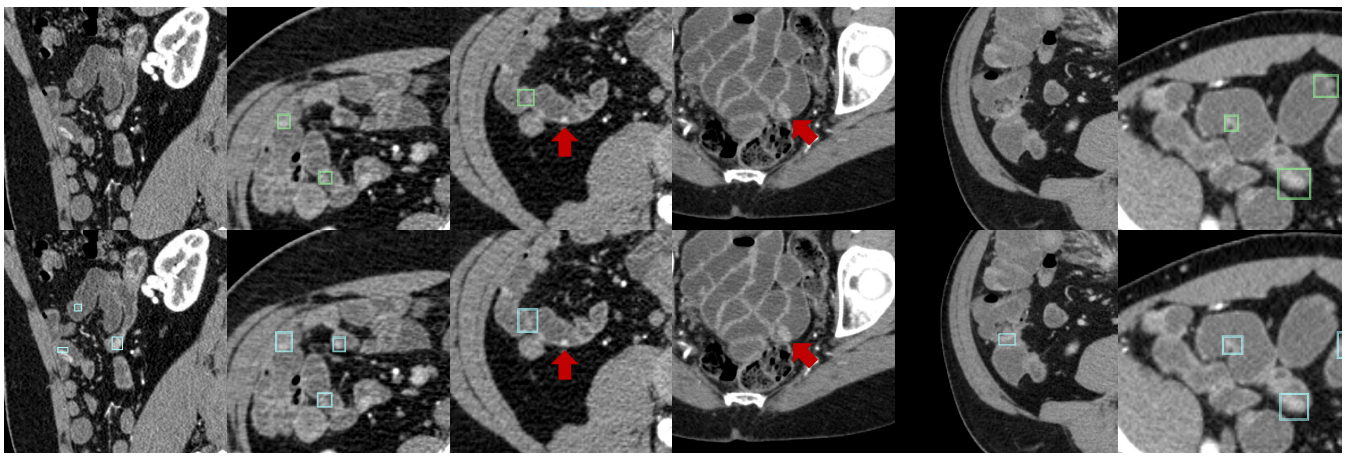


FIGURE 8. Example detection results are presented, with each row corresponding to a different method. The first row shows the results of the model without augmentation ('No augmentation' in Fig. 6), while the second row corresponds to the model with augmentation ('Proposed (5)'). From left to right, each group of two columns displays true positives, false negatives (with red arrows indicating the carcinoid tumors), and false positives detected by the proposed method, respectively.

rather than in trivial areas. The increase in false positives may be mitigated through the use of hard negative mining [42].

We used the same dataset used in [3]. Dealing with a rare disease, our dataset is of a limited size and from a single institution. While the dataset does not reflect the true clinical incidence of small bowel carcinoid tumors, it can be useful for predicting how the methods would perform in a general population [3]. Given its small size, we strategically split the dataset, and the *test set* was reserved solely for the evaluation of detection performance. We effectively trained the tumor synthesis network and the detection network within the *trainval set*.

In the current method, synthetic tumors are implanted throughout the entire small bowel without considering segment-specific predilections, even though carcinoid tumors most frequently occur in the ileum [21]. This may adversely

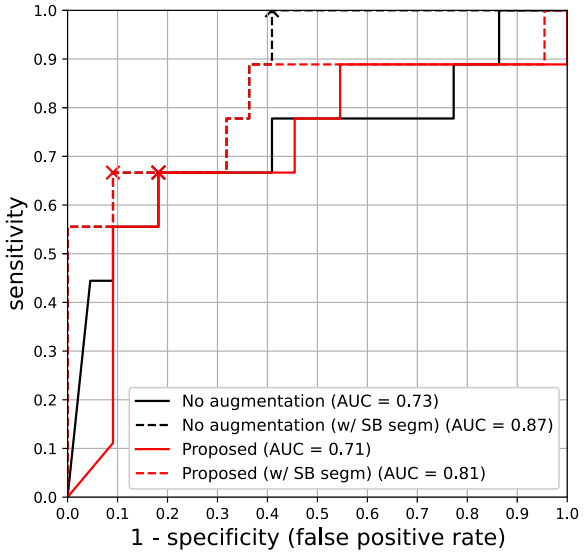


FIGURE 9. ROC curves of patient-level detection for the *test set*. Predicted small bowel segmentation is further used to eliminate false positives outside the small bowel for each method. The Youden point of each curve is marked with a cross.

affect the realism of the tumors when viewed in anatomical context. It is worth noting that the proposed method, including the selection of implantation sites, is fully automatic. A more realistic implantation strategy that reflects the varying tumor distribution along the bowel could potentially be achieved by incorporating an automatic path tracking method. However, the performance of existing path tracking approaches remains inadequate for reliable application in this context [29].

Although the proposed method built based on GANs already showed the high-quality synthesis result (synthetic tumors were not clearly distinguishable from real ones in Table 3), it can be considered to be combined with diffusion models [13] for better results. Also, it would be interesting to see if the use of synthetic tumors is effective also in *segmenting* tumors. We leave those as future work.

VI. CONCLUSION

In this work, we have proposed a new method for size-controllable tumor synthesis, with the application to the detection of small bowel carcinoid tumors in CT scans. The effectiveness of the proposed method was validated in terms of visual quality, size-controllability, as well as tumor detection. The method may be used for improved detection of other rare tumors.

APPENDIX
AUGMENTING TUMORFREE CT SCANS WITH SYNTHETIC TUMORS

Given the limited size of our dataset due to the rarity of the disease, we strategically split the data as described in Table 1 and Section II. The *test set*, consisting of both tumor-positive and tumor-negative scans, was reserved exclusively for evaluating detection performance. Meanwhile, the *trainval set* was used to train both the tumor synthesis and detection networks. In particular, after training the synthesis network, it was used to augment the *training set* (20 out of 24 scans in the *trainval set*) by implanting synthetic tumors alongside real ones. This strategy was designed to assess the impact of increasing the number of target instances during training without the need for additional scans. In this section, we further investigate whether augmenting tumor-free scans from the *test negative set* with synthetic tumors can enhance the detection network by increasing the number of training scans.

TABLE 4. Result of an additional visual Turing test. A senior radiologist and a senior gastroenterologist were given a set of randomly shuffled 20 real and 20 synthetic tumor images, and were requested to classify them. The same method as used for Table 3 was applied, except that the synthetic tumors were generated within scans from the *test negative set*.

		prediction					
		radiologist			gastroenterologist		
		real	synthetic	unsure	real	synthetic	unsure
truth	real	11	8	1	9	8	3
	synthetic	9	8	3	7	9	4

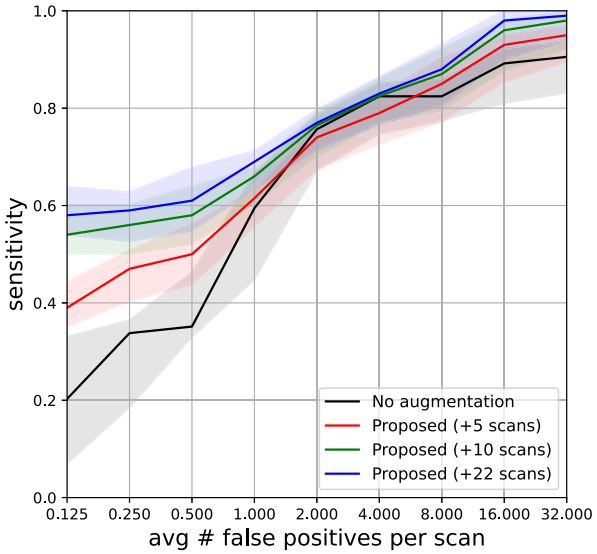


FIGURE 10. FROC curves in the *test positive set*, obtained by increasing the number of training scans through augmentation of tumor-free scans from the *test negative set* with synthetic tumors. For example, in the method 'Proposed (+5 scans)', five scans from the *test negative set* were used in addition to those in the *trainval set*. Each augmented scan was implanted with five synthetic tumors, which corresponds to the best-performing setting identified in Fig. 6. Bootstrapping is used to generate multiple FROC curves from which 95% confidence intervals are computed.

A. VISUAL TURING TEST

We first conduct an additional visual Turing test. The same method as used for Table 3 was applied, except that synthetic tumors generated within tumor-free scans from the *test negative set* were used. The results are presented in Table 4, once again demonstrating the difficulty in distinguishing

between real and synthetic tumors. The result of Bowker's Test for Symmetry [41] (P = 0.55) indicated that the classifications by the two observers were not significantly different and can be regarded statistically consistent.

B. RESULTS OF TUMOR DETECTION

Fig. 10 presents FROC curves obtained by increasing the number of training scans through augmentation of tumor-free scans from the *test negative set* with synthetic tumors. Using the maximum number of allowed scans ('Proposed (+22 scans)') resulted in improved performance compared not only to not using the augmentation ('No augmentation', P = 0.05), but also to the setting with fewer augmented scans ('Proposed (+5 scans)', P = 0.07).

CONFLICTS OF INTEREST

Potential financial interest: Author RMS receives royalties from iCAD, Philips, Scan Med, PingAn, MGB, and Translation Holdings and has received research support from Ping An (CRADA).

ACKNOWLEDGMENT

The contributions of NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this articles are those of the author(s) and do not necessarily reflect the views of NIH or the U.S. Department of Health and Human Services. The research used the high performance computing facilities of the NIH Biowulf cluster.

REFERENCES

[1] P. F. Jaeger, S. Köhl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H. Schlemmer, and K. Maier-Hein, "Retina U-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," in *Proc. Mach. Learn. Health NeurIPS Workshop*, Dec. 2020, pp. 171–183. [Online]. Available: https://proceedings.mlr.press/v116/jaeger20a.html

[2] M. Baumgartner, P. F. Jäger, F. Isensee, and K. H. Maier-Hein, "NnDetection: A self-configuring method for medical object detection," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds., Cham, Switzerland: Springer, 2021, pp. 530–539.

[3] S. Y. Shin, T. C. Shen, S. A. Wank, and R. M. Summers, "Fully-automated detection of small bowel carcinoid tumors in CT scans using deep learning," *Med. Phys.*, vol. 50, no. 12, pp. 7865–7878, Dec. 2023. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.100shin232/mp.16391

[4] S. Y. Shin, T. C. Shen, and R. M. Summers, "Improving segmentation and detection of lesions in CT scans using intensity distribution supervision," *Computerized Med. Imag. Graph.*, vol. 108, Sep. 2023, Art. no. 102259. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611123000770

[5] Q.-V. Dang and G.-S. Lee, "Scene text segmentation via multi-task cascade transformer with paired data synthesis," *IEEE Access*, vol. 11, pp. 67791–67805, 2023.

[6] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," 2019, *arXiv:1902.07296*.

[7] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2021, pp. 2917–2927, doi: 10.1109/CVPR46437.2021.00294.

[8] Y. Ge, H.-X. Yu, C. Zhao, Y. Guo, X. Huang, L. Ren, L. Itti, and J. Wu, "3D copy-paste: Physically plausible object insertion for monocular 3D detection," in *Proc. 37th Conf. Neural Inf. Process. Syst.*, 2023, pp. 17057–17071. [Online]. Available: https://openreview.net/forum?id=d86B6Mdweq

[9] K. H. Leung, W. Marashdeh, R. Wray, S. Ashrafinia, M. G. Pomper, A. Rahmim, and A. K. Jha, "A physics-guided modular deep-learning based automated framework for tumor segmentation in PET," *Phys. Med. Biol.*, vol. 65, no. 24, Dec. 2020, Art. no. 245032, doi: 10.1088/1361-6560/ab8535.

[10] Q. Yao, L. Xiao, P. Liu, and S. K. Zhou, "Label-free segmentation of COVID-19 lesions in lung CT," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2808–2819, Oct. 2021.

[11] Q. Hu, Y. Chen, J. Xiao, S. Sun, J. Chen, A. Yuille, and Z. Zhou, "Label-free liver tumor segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2023, pp. 7422–7432, doi: 10.1109/cvpr52729.2023.00717.

[12] R. Labaca-Castro, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 73–76. [Online]. Available: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

[13] J. Ho, A. N. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 6840–6851.

[14] S. I. Cho, C. Navarrete-Dechent, R. Daneshjou, H. S. Cho, S. E. Chang, S. H. Kim, J.-I. Na, and S. S. Han, "Generation of a melanoma and nevus data set from unstandardized clinical photographs on the internet," *JAMA Dermatol.*, vol. 159, no. 11, pp. 1223–1231, Nov. 2023, doi: 10.1001/jamadermatol.2023.3521.

[15] B. D. Basaran, M. Qiao, P. M. Matthews, and W. Bai, "Subject-specific lesion generation and pseudo-healthy synthesis for multiple sclerosis brain images," in *Simulation and Synthesis in Medical Imaging*, C. Zhao, D. Svoboda, J. M. Wolterink, and M. Escobar, Eds., Cham, Switzerland: Springer, 2022, pp. 1–11.

[16] R. Toda, A. Teramoto, M. Kondo, K. Imaizumi, K. Saito, and H. Fujita, "Lung cancer CT image generation from a free-form sketch using style-based pix2pix for data augmentation," *Sci. Rep.*, vol. 12, no. 1, p. 12867, Jul. 2022.

[17] G. Zhang, K. Chen, S. Xu, P. C. Cho, Y. Nan, X. Zhou, C. Lv, C. Li, and G. Xie, "Lesion synthesis to improve intracranial hemorrhage detection and classification for CT images," *Computerized Med. Imag. Graph.*, vol. 90, Jun. 2021, Art. no. 101929. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611121000781

[18] B. Hou, "High-fidelity diabetic retina fundus image synthesis from freestyle lesion maps," *Biomed. Opt. Exp.*, vol. 14, no. 2, pp. 533–549, Feb. 2023. [Online]. Available: https://opg.optica.org/boe/abstract.cfm?URI=boe-14-2-533

[19] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[21] R. Jasti and L. R. Carucci, "Small bowel neoplasms: A pictorial review," *RadioGraphics*, vol. 40, no. 4, pp. 1020–1038, Jul. 2020, doi: 10.1148/rg.2020200011.

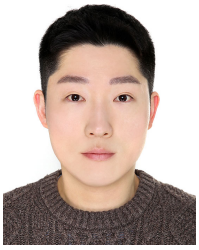
[22] (2018). *Key Statistics About Gastrointestinal Carcinoid Tumors*. [Online]. Available: https://www.cancer.org/content/dam/CRC/PDF/Public/8634.00.pdf

[23] M. S. Hughes, S. C. Azoury, Y. Assadipour, D. M. Straughan, A. N. Trivedi, R. M. Lim, G. Joy, M. T. Voellinger, D. M. Tang, A. M. Venkatesan, C. C. Chen, A. Louie, M. M. Quezado, J. Forbes, and S. A. Wank, "Prospective evaluation and treatment of familial carcinoid small intestine neuroendocrine tumors (SI-NETs)," *Surgery*, vol. 159, no. 1, pp. 350–357, Jan. 2016.

[24] S. Y. Shin, T. C. Shen, S. A. Wank, and R. M. Summers, "Improving small lesion segmentation in CT scans using intensity distribution supervision: Application to small bowel carcinoid tumor," *Proc. SPIE*, vol. 12465, Apr. 2023, Art. no. 124651S, doi: 10.1117/12.2651979.

[25] S. Y. Shin, S. Lee, D. Elton, J. L. Gulley, and R. M. Summers, "Deep small bowel segmentation with cylindrical topological constraints," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds., Cham, Switzerland: Springer, 2020, pp. 207–215.

[26] S. Y. Shin, S. Lee, and R. M. Summers, "Unsupervised domain adaptation for small bowel segmentation using disentangled representation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds., Cham, Switzerland: Springer, 2021, pp. 282–292.
[27] S. Y. Shin, S. W. Lee, and R. M. Summers, "A graph-theoretic algorithm for small bowel path tracking in CT scans," *Proc. SPIE*, vol. 12033, pp. 863–868, Apr. 2022, doi: 10.1117/12.2611878.
[28] S. Y. Shin, S. Lee, and R. M. Summers, "Graph-based small bowel path tracking with cylindrical constraints," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–5.
[29] S. Y. Shin and R. M. Summers, "Deep reinforcement learning for small bowel path tracking using different types of annotations," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds., Cham, Switzerland: Springer, 2022, pp. 549–559.
[30] Y. Zhao, R. Wu, and H. Dong, "Unpaired image-to-image translation using adversarial consistency loss," in *Computer Vision—ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., Cham, Switzerland: Springer, 2020, pp. 800–815.
[31] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Computer Vision—ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Cham, Switzerland: Springer, 2018, pp. 179–196.
[32] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds., Switzerland: Springer, 2016, pp. 424–432.
[33] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu, Eds., Cham, Switzerland: Springer, 2017, pp. 240–248.
[34] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, and I. Ben Ayed, "Constrained-CNN losses for weakly supervised segmentation," *Med. Image Anal.*, vol. 54, pp. 88–99, May 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841518306145
[35] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6629–6640.
[36] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in *Proc. Int. Conf. Learn. Represent.*, 2018. [Online]. Available: https://openreview.net/forum?id=r1lUOzWCW
[37] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach, and M. Segeroth, "TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images," *Radiol., Artif. Intell.*, vol. 5, no. 5, Sep. 2023, Art. no. 230024, doi: 10.1148/ryai.230024.
[38] M. S. M. Sajjadi, O. Bachem, M. Lucic, O. Bousquet, and S. Gelly, "Assessing generative models via precision and recall," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 5234–5243.
[39] F. Samuelson, N. Petrick, and S. Paquerault, "Advantages and examples of resampling for CAD evaluation," in *Proc. 4th IEEE Int. Symp. Biomed. Imag., From Nano Macro*, Apr. 2007, pp. 492–495.
[40] S. Qiao, R. Wang, S. Shan, and X. Chen, "Hierarchical image-to-image translation with nested distributions modeling," *Pattern Recognit.*, vol. 146, Feb. 2024, Art. no. 110058. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320323007550
[41] A. H. Bowker, "A test for symmetry in contingency tables," *J. Amer. Stat. Assoc.*, vol. 43, no. 244, pp. 572–574, Dec. 1948. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1948.10483284
[42] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2016, pp. 761–769, doi: 10.1109/CVPR.2016.89. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.89



SEUNG YEON SHIN received the B.S. degree in Electronic and Communication Engineering from Hanyang University ERICA, Ansan, South Korea, in 2011, and the Ph.D. degree in Electrical and Computer Engineering from Seoul National University, Seoul, South Korea, in 2019.

From 2019 to 2023, he was a Postdoctoral Fellow with the National Institutes of Health, Clinical Center, Bethesda, MD, USA. Since 2023, he has been an Assistant Professor with the Division of Electrical Engineering, Hanyang University ERICA. His research interest includes developing computational methods for solving problems relating to medical images. He was a recipient of the Fellows Award for Research Excellence, in 2023 at the National Institutes of Health, and the MICCAI Student Travel Award, in 2016.

STEPHEN A. WANK is a tenured Senior Investigator and Chief of the Digestive Diseases Branch, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD. He received his undergraduate and medical degrees from Duke University. He is Board Certified in Internal Medicine at Johns Hopkins Hospital, Gastroenterology at Stanford U. Medical Center and Endocrinology at NIH. His research interests center on the endocrine aspect of the gastrointestinal tract including the molecular and cellular biology of enteroendocrine cells and the origin, development and genetics of small bowel carcinoid tumors.

RONALD M. SUMMERS M.D., Ph.D., is a tenured Senior Investigator and Staff Radiologist in the Radiology and Imaging Sciences Department at the NIH Clinical Center in Bethesda, MD. He is a Fellowof the Society of Abdominal Radiologists, the American Institute for Medical and Biological Engineering and SPIE. His awards include the Presidential Early Career Award for Scientists and Engineers, the NIH Director's Award, and the NIH Ruth L. Kirschstein Mentoring Award. He is a member of the editorial boards of the *Journal of Medical Imaging* and *Academic Radiology* and a past member of the editorial boards of *Radiology* and *Radiology: Artificial Intelligence*. He has co-authored over 700 journal, reviewand conference proceedings articles and is a co-inventor on 17 patents. His research interests include thoracic and abdominal imaging, large radiology image databases, and artificial intelligence.