SynthRAD Challenge Algorithm Summary for Team FGH_365

Yubo Fan and Han Liu

Department of Computer Science, Vanderbilt University, Nashville, TN 37235, USA
yubo.fan@vanderbilt.edu
han.liu@vanderbilt.edu

1 Method

1.1 Overview

In the report, we introduce our method for both task 1 and task 2 of the SynthRAD challenge. Because the challenge dataset [1, 2] consists of multi-modality (MR and CBCT), multi-region (brain and pelvis), and multi-site (sites A, B, and C) images, we design our algorithms to be modality/anatomy/site (MAS)-specific to avoid the potential issue of domain shift. We first develop two independent MAS-specific solutions (we name them "solution #1" and "solution #2" for simplicity) in parallel to generate synthetic CTs (sCTs). At inference time, the algorithm is only provided with the information of the modality (MR or CBCT) and the anatomical region (brain or pelvis) but not the site (A, B, or C) of the image. We design an uncertainty-based algorithm to predict the source site of the input image, then we use the model belonging to the predicted site to generate the sCT. The final output is the ensemble (average) of the sCT predictions from the two independent MAS-specific solutions. In the sections below, we describe the two MAS-specific solutions and how we integrate them using our site-prediction algorithm as our final solution.

1.2 MAS-specific solution #1

In this solution, we train one CT synthesis deep learning (DL) model per MAS, resulting in 11 models in total (pelvis MR data only have two sites: A and C). We only use data from the corresponding MAS to train each model. We adopt a 3D version of the pix2pix conditional GAN (cGAN) [3] as our network architecture. It consists of a generator and a discriminator. The generator is a 3D encoder-decoder network with nine residual blocks in the bottleneck, and the discriminator D is a five-layer 3D CNN based on PatchGAN [3]. At inference time, we use the DL model that corresponds to the specific MAS to predict the sCT (the site information is predicted by the algorithm described in Section 1.4).

Multiple loss terms are used to train the network. The discriminator is trained to classify fake or real CTs using a cross-entropy loss, and the adversarial loss for the

* Y. Fan and H. Liu—Equal contribution

generator is denoted as L_{adv}^{G}. To further regularize the generator to generate accurate sCTs, as suggested in [3], we use the L1 distance between the real CT (rCT) image and the sCT image as the reconstruction loss, which is denoted as L_{L1}^{G}. We also use an edge-aware loss L_{edge}^{G} to help improve the quality of the sCT [4, 5]. It is calculated as the L1 distance of edge maps (extracted by a 3D Sobel edge detector) between the rCT and sCT. Furthermore, we add two additional loss terms to make the loss focus more on the regions inside the body mask (provided by the challenge) and the bone mask (obtained by thresholding the rCT image and then by morphological operations), denoted by L_{body}^{G} and L_{bone}^{G}. We calculate these two losses by repeating the L1 reconstruction loss and the edge-aware loss inside the masks. To summarize, the total loss for the generator can then be expressed as:

$$L_{total}^{G} = \lambda_1 L_{adv}^{G} + \lambda_2 L_{L1}^{G} + \lambda_3 L_{edge}^{G} + \lambda_4 L_{body}^{G} + \lambda_5 L_{bone}^{G}, \quad (1)$$

where,

$$L_{body}^{G} = \lambda_2 L_{L1_{\text{body}}}^{G} + \lambda_3 L_{edge_{mask}}^{G}, \quad (2)$$

and

$$L_{bone}^{G} = \lambda_2 L_{L1_{\text{bone}}}^{G} + \lambda_3 L_{edge_{bone}}^{G}, \quad (3)$$

λ_1, λ_2, λ_3, λ_4, and λ_5 are empirically set as 1, 10, 1, 10, 1, respectively.

1.3 MAS-specific solution #2

In this solution, we train one unified DL model with dynamic convolution (DC) layers [6–8] that are conditioned on the MAS. Different from solution #1, the model is trained with all the data from different modalities, regions, and sites. The network architecture is the same 3D pix2pix as described in solution #1, but the first two and last two convolutional layers are replaced with DC layers. The parameters of the DC layers are generated by a multilayer perceptron (MLP) and conditioned on the binary code that corresponds to the specific MAS. At inference time, we provide the binary code to the CT synthesis network so that the parameters of the network can be adaptive to the MAS. We use the same loss function configuration as the one introduced in solution #1.

1.4 Site-prediction algorithm and final output of our method

At inference time, only the modality (MR in task 1 and CBCT in task 2) and the anatomical region (brain and pelvis) are provided by the organizers as additional information to the image. Our MAS-specific solutions #1 and #2 cannot work optimally without the correct site information. We thus design a simple uncertainty-based algorithm to predict the site of the input image. To demonstrate our algorithm, we use an input image that comes from site A as an example. We assume that this image (from site A) is considered to be the out-of-distribution (OOD) data for the models that are trained/designed exclusively for site B or site C, thus the uncertainty of model from site B or site C will be higher than the uncertainty of the model from site A. We measure the uncertainty by calculating the mean absolute error (MAE) between the

MAS-specific solutions #1 and #2 introduced above, and we select the site with the lowest MAE as the predicted site. This algorithm is shown in Fig. 1.

We then use the MAS-specific models (from solutions #1 and #2) that correspond to the predicted site to generate sCT. Since we find that solutions #1 and #2 have similar performance in the validation phase of the challenge, we use the ensemble (average) of sCT generated by the two independent solutions as the final output of our method.

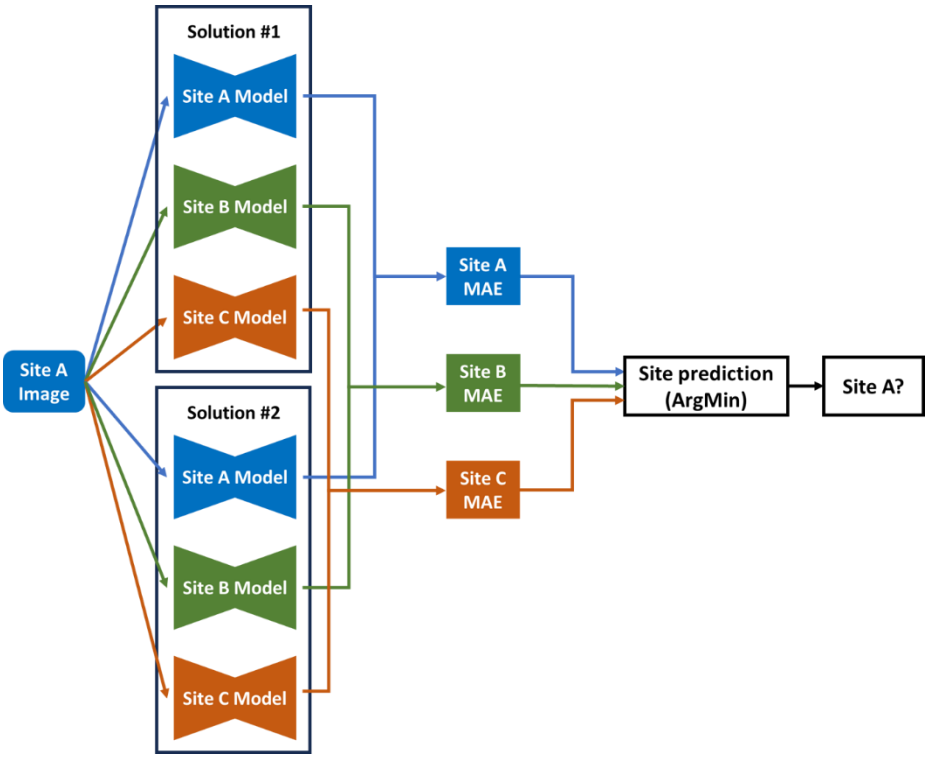

Fig. 1. An illustration of the site-prediction algorithm. We use an input image from site A as an example.

1.5 Training details

We implement our method using PyTorch and MONAI. For all the models in solutions #1 and #2, we use the same learning rate of 2e-4. We split the provided dataset into 90% for training and 10% for model validation. The best epoch for each model is selected based on the MAE on the validation set. Data augmentation used in training includes random affine transformation, random left-right flipping, random Gaussian noise, random bias field (for MR images), and random contrast adjustment. We do not change the original voxel size of each image. Due to the GPU memory size, the networks can only take cropped patches as input. During training, all the brain images in

solution #1 are randomly cropped into 128×128×128, and all the pelvis images are randomly cropped into 256×128×64. All the images in solution #2 are randomly cropped into 192×192×128.

Acknowledgements. We would like to thank Dr. Benoit Dawant for his advice and helpful discussions during this challenge.

References

1. Thummerer, A., van der Bijl, E., Galapon Jr, A., Verhoeff, J.J.C., Langendijk, J.A., Both, S., van den Berg, C. (Nico) A.T., Maspero, M.: SynthRAD2023 Grand Challenge dataset: Generating synthetic CT for radiotherapy. Medical Physics. 50, 4664–4674 (2023). https://doi.org/10.1002/mp.16529.
2. Thummerer, A., Huijben, E., Terpstra, M., Gurney-Champion, O., Afonso, M., Pai, S., Koopmans, P., Eijnatten, M. van, Perko, Z., Maspero, M.: SynthRAD2023 Challenge design. (2023). https://doi.org/10.5281/zenodo.7746020.
3. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-To-Image Translation With Conditional Adversarial Networks. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017).
4. Luo, Y., Nie, D., Zhan, B., Li, Z., Wu, X., Zhou, J., Wang, Y., Shen, D.: Edge-preserving MRI image synthesis via adversarial network with iterative multi-scale fusion. Neurocomputing. 452, 63–77 (2021). https://doi.org/10.1016/j.neucom.2021.04.060.
5. Fan, Y., Khan, M.M.R., Liu, H., Noble, J.H., Labadie, R.F., Dawant, B.M.: Temporal bone CT synthesis for MR-only cochlear implant preoperative planning. In: Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling. pp. 358–363. SPIE (2023). https://doi.org/10.1117/12.2647443.
6. Hu, S., Liao, Z., Zhang, J., Xia, Y.: Domain and Content Adaptive Convolution Based Multi-Source Domain Generalization for Medical Image Segmentation. IEEE Transactions on Medical Imaging. 42, 233–244 (2023). https://doi.org/10.1109/TMI.2022.3210133.
7. Zhang, J., Xie, Y., Xia, Y., Shen, C.: DoDNet: Learning to Segment Multi-Organ and Tumors from Multiple Partially Labeled Datasets. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1195–1204. IEEE, Nashville, TN, USA (2021). https://doi.org/10.1109/CVPR46437.2021.00125.
8. Liu, H., Fan, Y., Li, H., Wang, J., Hu, D., Cui, C., Lee, H.H., Zhang, H., Oguz, I.: Moddrop++: A dynamic filter network with intra-subject co-training for multiple sclerosis lesion segmentation with missing modalities. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 444–453. Springer Nature Switzerland Cham (2022).