Technical Report for Segmentation of Organs-at-Risk and Gross Tumor Volume of NPC (SegRap2023)

Zhaohu Xing[1], Hongqiu Wang[1] (✉) and Lei Zhu [1]

[1] The Hong Kong University of Science and Technology (Guangzhou)
hwang007@connect.hkust-gz.edu.cn

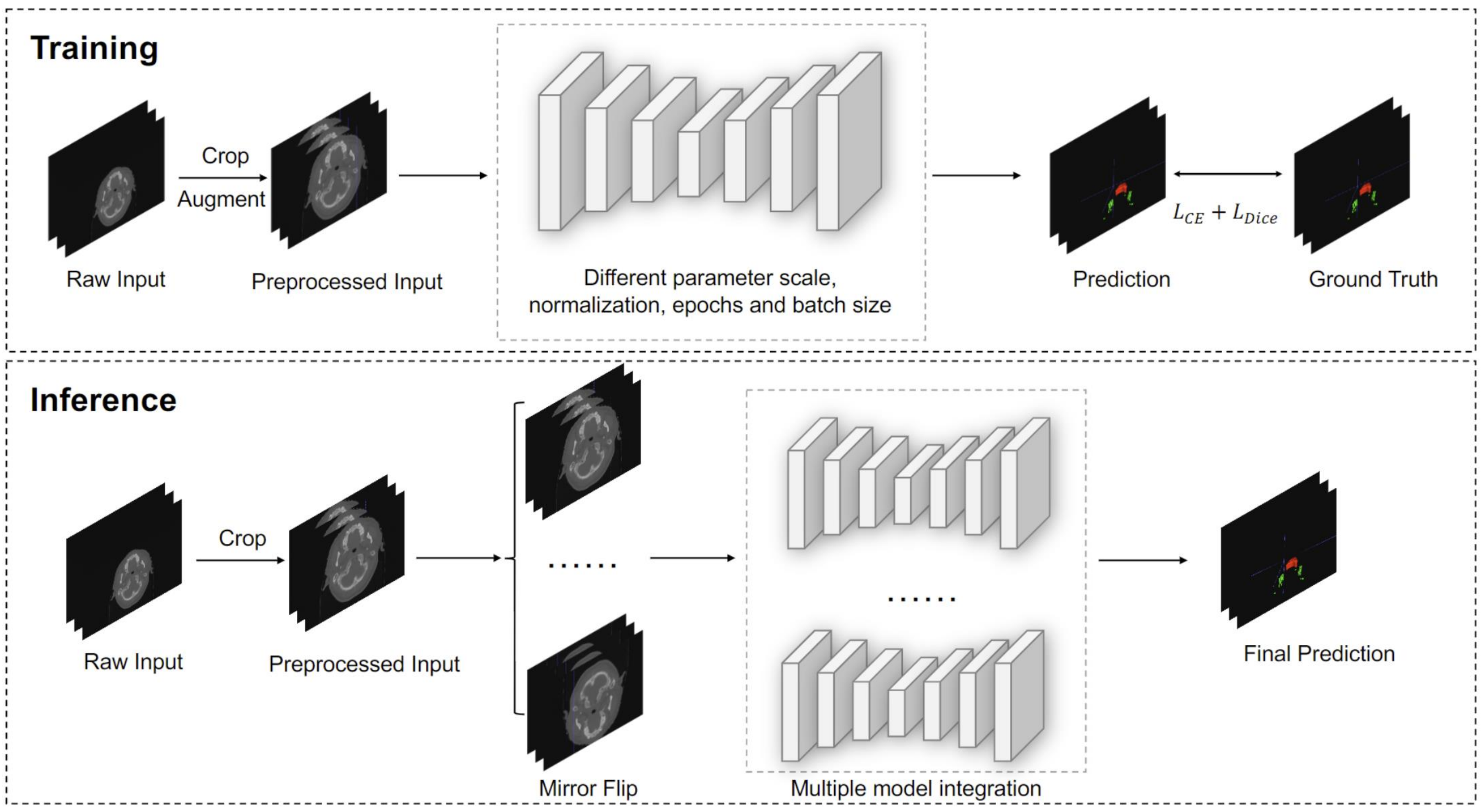
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Introduction

- **Background**: Nasopharyngeal carcinoma (NPC) is a common and severe malignancy that primarily affects the head and neck region. Accurate delineation of gross tumor volume (GTV) and organs at risk (OARs) is a crucial and challenging task for effective radiotherapy of NPC.
- **Contributions**: We design a crop strategy to reserve the core area of input data to improve training speed and explore multiple models and hype-parameter settings for brain organs and tumor segmentation. And then, we also apply the testing time augmentation strategies, such as mirror prediction and overlapped window inference. Finally, we integrate three good models in testing phase to obtain a more robust prediction.

Method

- **Pre-processing:** First, we need to crop the useless area of a 3D input image due to the whole size is so large. We reserve the non-zero area to reduce the size of input data, which can significantly save training time. Second, in the analysis phase, we read all the training data and compute the mean and std value of all the foreground pixels, which can normalize the CT image. Then we normalize every training and validation CT image and re-sample every input CT image to this spacing to maintain the data consistency.
- **Data Augmentation Strategies:** In addition, to increase the diversity of the training data and improve the robustness of the model, we use an extensive data augmentation operations. In the spatial dimension, we use random rotation, random scale and random flip to execute spatial transform, and in the pixel dimension, we use gaussian noise, gaussian blur, brightness, contrast enhancement and gamma transform to add additional noise for input 3D image.In the test phase, we also use the test time augmentation to improve the robustness of the prediction.
- **Proposed Methodology:** We choose UNet structure as the basic network. According to the input 3D image shape, we choose [64, 256, 256] as the input patch size. We also try to select a smaller patch size (e.g. [64, 128, 128]) to segment the tiny region. We also explore various settings for the UNet model according to this task, such as training epochs, normalization, model size, patch size, batch size and so on. We find the combination of larger training epochs, more parameters and larger batch size has the positive effect to the segmentation result, detailed in experiments section.



- **Training and testing strategies:** All experiments were carried out using PyTorch and were trained on an NVIDIA A100 GPU with 40 GB of memory. To ensure consistency during training, all CT images were resized to a same spacing, which is [3.0, 0.54199219, 0.54199219]. All models were trained with 2000 epochs. The initial learning rate is 0.01 with a poly learning rate schedule. This allows the model to be more stable in the later stages of training. And all experiments use deep supervision strategy to improve the feature representation ability.

Method

- **Experimental Results**: Quantitative results obtained from various models for task 02 are displayed in the table to the right. In our pursuit of robust predictions, we have carefully chosen the top three models.

Model	Batch size	epoch	Parameter scale	Normalization	Mean metric (val)
UNet	2	2000	Base	Instance	0.757
UNet	4	2000	Base	Instance	0.761
UNet	4	2000	Large	Instance	0.764
UNet	8	2000	Large	Instance	0.759
UNet	8	2000	Large	Batch	0.759