Leveraging Semantic Asymmetry for Accurate Gross Tumor Volume Segmentation of Nasopharyngeal Carcinoma in Planning CT

Zi Li*[1,2,4](✉), Ying Chen*[3], Zeli Chen[1,4], Yanzhou Su[1,4], Tai Ma[1], Tony C. W. Mok[1,4], Yan-Jie Zhou[1,4], Yunhao Bai[1], Zhilin Zheng[1,4], Le Lu[1], Yirui Wang[1], Jia Ge[3], Senxiang Yan[3], Xianghua Ye[3](✉), and Dakai Jin[1]

[1] DAMO Academy, Alibaba Group
[2] The University of Hong Kong, Hong Kong
[3] The First Affiliated Hospital, Zhejiang University, China
[4] Hupan Lab, 310023, Hangzhou, China
alisonbrielee@gmail.com;hye1982@zju.edu.cn

Abstract. In the radiation therapy of nasopharyngeal carcinoma (NPC), clinicians typically delineate the gross tumor volume (GTV) using non-contrast planning computed tomography to ensure accurate radiation dose delivery. However, the low contrast between tumors and adjacent normal tissues requires radiation oncologists to delineate the tumors with additional reference from MRI images manually. In this study, we propose a novel approach to directly segment NPC gross tumors on non-contrast planning CT images, circumventing potential registration errors when aligning MRI or MRI-derived tumor masks to planning CT. To address the low contrast issues between tumors and adjacent normal structures in planning CT, we introduce a 3D Semantic Asymmetry Tumor Segmentation (SATS) method. Specifically, we posit that a healthy nasopharyngeal region is characteristically bilaterally symmetric, whereas the presence of nasopharyngeal carcinoma disrupts this symmetry. Then, we propose a Siamese contrastive learning segmentation framework that minimizes the voxel-wise distance between original and flipped areas without tumor and encourages a larger distance between original and flipped areas with tumor. Thus, our approach enhances the sensitivity of deep features to semantic asymmetries. Extensive experiments demonstrate that the proposed SATS achieves the leading NPC GTV segmentation performance in both internal and external testing.

1 Introduction

Nasopharyngeal carcinoma (NPC) ranks among the most prevalent head & neck malignancies affecting the nasopharyngeal region, with patient prognosis substantially enhanced through early diagnosis and intervention [7]. A significant proportion of NPC patients can achieve complete remission following radiation

* Equal contribution. ✉ Corresponding author.

therapy (RT) [6]. Notably, this type of cancer exhibits a remarkable sensitivity to radiation therapy, wherein a pivotal component of this therapeutic intervention is the accurate delineation of the gross tumor volume (GTV). In clinical practice, magnetic resonance imaging (MRI) has emerged as the predominant imaging modality for NPC, owing to its superior resolution in visualizing soft tissues. Subsequently, cross-modality registration is conducted between MRI and non-contrast planning computed tomography (pCT) to transfer tumor delineations from MRI to pCT scans for treatment planning [31]. However, cross-modality registration is non-trivial due to substantial modality gaps and variations in scanning ranges. Alternatively, physicians may integrate pCT and MRI mentally to assist in delineating the GTV. Nevertheless, this approach is time-consuming, often taking 1-2 hours per case, and is prone to potential inaccuracies.

Recent advances in learning-based methods have shown success in segmenting NPC tumors from MRI scans [12, 17, 21, 22, 25, 26]. However, MRI does not provide direct electron density measurements, which are critical for radiotherapy planning. Tumor masks derived from MRI must be transformed into pCT using image registration, a process prone to alignment errors. Approaches [4, 28] combine CT and MRI for tumor segmentation, although misalignment between the two modalities can reduce performance compared to single-modality approaches. Other researchers [2, 20, 29, 30] focus on contrast-enhanced CT-based segmentation. Still, these methods often achieve low performance (e.g., Dice scores below 70%) due to tumor infiltration into adjacent tissues and limited contrast in pCT, particularly for soft tissues such as mucous membranes, muscles, and nerves.

This study aims to segment the NPC gross tumor volume (GTV) in non-contrast planning CT (pCT), avoiding registration errors associated with aligning MRI or MRI-derived tumor masks to pCT. Direct segmentation of NPC GTV in non-contrast pCT is challenging due to indistinct boundaries between tumors and adjacent soft tissues [19], such as membranes, muscles, and vessels. Meanwhile, we observe that medical image analysis benefits from the bilateral symmetry of human anatomy, evident in structures like the head, brain, breasts, lungs, and pelvis. Research [3, 13, 14, 23, 24, 34] highlights the utility of symmetry-based approaches in enhancing early detection capabilities. Therefore, we propose a tumor segmentation method for NPC, which leverages the observation that a healthy nasopharyngeal region is bilaterally symmetric, but the presence of a tumor disrupts this symmetry. While prior work has explored symmetry in medical imaging, our approach differs significantly in how symmetric features are utilized. For instance, [13] employs symmetric position encoding for brain structures using an autoencoder, without explicit constraints on symmetric or asymmetric regions (e.g., via custom losses or modules). [34] leverages pelvic symmetry to detect fractures but focuses primarily on symmetric anatomy. In contrast, our method emphasizes the contrast between asymmetric lesion areas.

The main contributions of this work are: 1) We introduce a 3D *semantic asymmetry tumor segmentation (SATS)* method for NPC GTV in non-contrast pCT, which is the most common imaging modality in RT planning. To the best of our knowledge, this is the first work to tackle the NPC GTV segmentation in

non-contrast CT scans and employ the symmetry cue for the GTV segmentation. 2) We develop a Siamese contrastive learning segmentation framework with an asymmetrical region selection approach, which facilitates the learning of asymmetric tumor features effectively. 3) We demonstrate that our proposed SATS achieves *state-of-the-art* performance in NPC GTV segmentation, outperforming the leading methods in internal and external testing datasets.

2 Method

We propose a 3D semantic asymmetry tumor segmentation (SATS) method based on the semantic asymmetry property of the gross tumor in the nasopharyngeal area, to enable accurate NPC GTV segmentation. Given one CT scan, as shown in Figure 1 (a), we utilize a shared encoder-decoder module to process both the original image $I \in \mathbb{R}^{D\times H\times W}$, where D, H, W are CT image spatial dimensions, and its flipped image I', thereby encoding them into a symmetric representation. Subsequently, we introduce a non-linear projection module and a distance metric learning strategy to refine the resulting feature maps. We intend to maximize the dissimilarity between E and E_f at corresponding anatomical locations on the abnormalities and normalities. The distance metric learning paradigm is illustrated in Figure 1 (b).

2.1 Asymmetrical Abnormal Region Selection

We focus on asymmetrical lesion areas relative to the central sagittal axis, i.e., region B of Figure 1 (b). To this end, we perform: 1) head-neck position normalization (bilateral symmetry along the central sagittal axis) of the overall head-neck region by utilizing rigid registration (rotation and translation). 2) The asymmetrical abnormal region is obtained by subtracting symmetrical lesion regions from the original mask.

To be specific, considering that image asymmetry may originate from pathological or non-pathological sources, such as changes in imaging angles and patient postures, we pre-process the CT scans using [33] to ensure that the scans are symmetric along the central sagittal axis. We manually select a patient CT image with bilateral symmetry along the central sagittal plane, which serves as an atlas, and then align other patient CT images to the atlas space through affine registration. This step helps to alleviate the influence of other asymmetric anatomical structures in the head & neck that may mislead the model.

Then, we detect asymmetric abnormal regions using the available tumor annotation. The semantic segmentation mask of I is denoted as $s \in \{0,1\}^{D\times H\times W}$, where 0 represents the background and 1 represents the foreground of tumors. Through the flip operation, we can obtain the flipped semantic mask s' of I'. Subsequently, an asymmetrical mask $\boldsymbol{m}$ is defined to locate asymmetrical regions in the image I, as $\boldsymbol{m} = \boldsymbol{s} - \boldsymbol{s} \cap \boldsymbol{s'}$, where $\boldsymbol{m} \in \{0,1\}^{D\times H\times W}$. Note that 1 and 0 represent the asymmetrical and symmetrical regions in I, respectively.

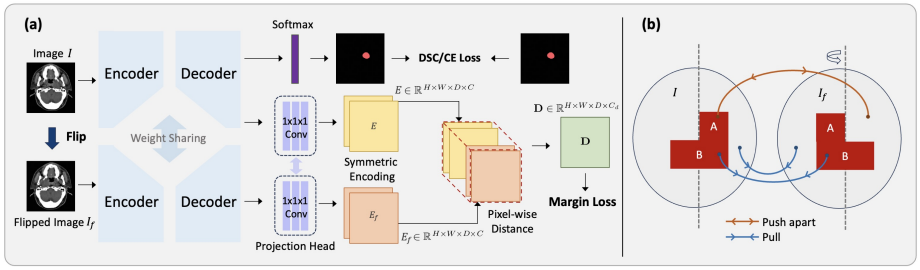

Fig. 1: **(a)** Our SATS model begins with the encoder-decoder module, which uses shared weights to process two input signals and encode them into a discriminative representation. This representation is then further processed through a non-linear projection module and a distance metric learning module to produce feature maps. **(b)** A graphical representation of our metric learning strategy. Circles indicate individual CT images, I, while red squares highlight the tumors. The tumors are composed of A and B, representing symmetrical and asymmetrical lesions relative to the central sagittal axis of symmetry, respectively.

2.2 Asymmetrical Learning Strategy

Our segmentation loss function is comprised of two components: a combination of Dice and entropy loss for the conventional segmentation purpose, and a voxel-wise margin loss specifically designed for asymmetric abnormal regions.

Metric-based margin loss. In the asymmetric anomaly region, we aim to minimize the similarity between the features of any point and its corresponding point on the central sagittal axis. To achieve this, we employ pixel-level margin loss. Based on above asymmetrical abnormal region $\boldsymbol{m}$, the margin loss between features $E \in \mathbb{R}^{H\times W\times D\times C}$, where C is the number of output features, and flipped E' after a non-linear projection is as:

$$\begin{aligned} \boldsymbol{l}_{margin} = \textstyle\sum_{i,j,k}^{D,H,W} [\mathbf{1}_{(m(i,j,k)=1)} ||E(i,j,k) - E'(i,j,k)||^2 + \\ \mathbf{1}_{(m(i,j,k)\neq 1)} \max(t - ||E(i,j,k) - E'(i,j,k)||^2, 0)] \end{aligned} \tag{1}$$

where $\mathbf{1}$ is the indicator function, and t defines a margin that regulates the degree of dissimilarity in semantic asymmetries.

Overall loss function. We approach tumor segmentation as a binary segmentation task, utilizing the Dice loss, binary cross-entropy loss, and margin loss as our objective function. The overall loss function is formulated as: $l = l_{dice} + l_{ce} + \beta l_{margin}$, where β is the weight balancing the different losses.

2.3 Siamese Segmentation Architecture

Our SATS architecture comprises the encoder-decoder module and the projection head. While both components are engaged during the training process, only the encoder-decoder module is required during inference.

Siamese encoder-decoder. The backbone is a shared U-shaped encoder-decoder architecture, as shown in Fig. 1. The encoder employs repeated applications of 3D residual blocks, with each block comprising two convolutional layers with $3 \times 3 \times 3$ kernels. Each convolutional layer is succeeded by InstanceNorm normalization and LeakyReLU activation. For downsampling, a convolutional operation with a stride of 2 is utilized to halve the resolution of the input feature maps. The initial number of filters is 32 and doubles after each downsampling step to maintain constant time complexity except for the last layer. In total, the encoder performs four downsampling operations.

Projection head. We utilize a non-linear projection g to transform the features before calculating the distance in margin loss, which aims to enhance the quality of the learned features. It consists of three $1\times1\times1$ convolution layers with 16 channels followed by a unit-normalization layer. The first two layers in the projection head use the ReLU activation function. We hypothesize that directly applying metric learning to segmentation features might lead to information loss and diminish the model's effectiveness. For example, some asymmetries in CT images are non-pathological and may stem from variations in the patient's head positioning and posing, yet they are beneficial for segmentation. Utilizing a non-linear projection may filter out such irrelevant information from the metric learning process, ensuring it is preserved in the features used for segmentation.

3 Experiments

3.1 Data Preparation and Implementation Details

We collected an *in-house dataset* from the hospital for the model development, which consisted of 163 NPC patients with pCT, contrast-enhanced diagnostic CT, and diagnostic MRIs of T1 & T2 phases. Diagnostic CT and MRI were registered as, initially, a rigid transformation [1] was applied to the MRI images to approximately align with the CT images. Then, deformable registration algorithm, deeds [11], was utilized to achieve precise alignment. The contrast-enhanced CT and MRIs are used to guide radiation oncologists to generate ground-truth GTV in pCT. Also, we collected a public dataset, SegRap2023 [5],

[5] https://segrap2023.grand-challenge.org/dataset/

Dataset	Train	Test
In-house	130	33
External	0	117

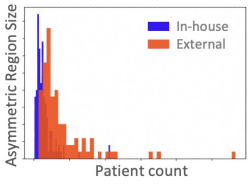

Fig. 2: Left: data partitioning situation. Right: the size of the asymmetric regions of different individuals, with the x-axis displaying each test object.

as *external testing dataset*, containing 118 non-contrast pCT and enhanced CT. *Annotations of all datasets were examined and edited by two experienced radiation oncologists following the international GTV delineation consensus guideline [18].* For evaluation, 20 % of the in-house dataset was randomly selected as the internal testing set, and the entire curated SegRap2023 was used as the external testing dataset. As shown in Figure 2, the asymmetric regions in the external data are larger than those of in-house, making the task more challenging. **Implementation.** The model training is divided into two stages. In the first stage, only the Siamese encoder-decoder is trained for 800 epochs with a learning rate of $1e-2$ and decayed via a polynomial schedule. Then, the projection head is trained for 200 epochs, with a learning rate of $1e-2$ for the projection head and $1e-5$ for the encoder-decoder, both with decayed via a polynomial schedule. The patch size is $56 \times 192 \times 192$ and the batch size is 2. For the voxel-wise contrastive loss, we use a margin hyperparameter $t = 20$ and $\beta = 1$.

3.2 Comparing to State-of-the-art Methods

Comparison methods. We conducted a comprehensive comparison of our method with **ten** cutting-edge approaches, encompassing prominent CNN-based, Transformer-based and Mamba-based methods, to evaluate its performance. CNN-based methods include STU-Net S [15], STU-Net B [15], STU-Net L [15], MedNeXt [32] and nnUNet [16]. Transformer-based methods include UNETR [9], TransUNet [5], SwinUNETR [8] and its variant SwinUNETR-v2 [10]. Mamba-based methods include UMambaBot [27]. To maintain a fair comparison, we trained all competing models for an equal number of epochs, 1000. **Evaluation metrics.** We evaluate the performance using the Dice similarity coefficient, DSC (%), and the 95th percentile of the Hausdorff distance (HD95, mm) and average surface distance (ASD, mm) across all cases.

In-house dataset performance. Table. 1 summarizes the quantitative segmentation performance and model parameters. Under a relatively small number of parameters, the proposed SATS demonstrates an improvement over previous approaches. For example, SATS outperforms the transformer-based SWinUNETR-V2 in DSC, and HD95 by 0.81% and 3.6%, respectively. Figure 3 presents the segmentation results of the top four performing methods (SATS, SwinUNETR-V2, SwinUNETR, and nnUNet) on a sample from the in-house dataset. It can be observed that our SATS method exhibits higher accuracy in boundary segmentation (e.g., the nasal septum). **Robustness.** Large primary tumors can cause asymmetrical anatomical changes. NPC patients often show lymphatic involvement, significantly affecting the integrity and symmetry of nearby structures. Figure 5 highlights cases of lymphatic invasion, demonstrating our robustness in handling lymph nodes while accurately segmenting the primary tumor.

Performance in external evaluation. Table 1 and Figure 4 summarize the external testing results. Several conclusions can be drawn. First, the proposed SATS achieves the best performance as compared to all other leading methods in external evaluation. As compared to an increase of 0.92% DSC over the 2nd best-performing method (nnUNet) in internal testing, SATS exhibits a

Table 1: Quantitative results on NPC GTV segmentation task. In-house$_{train}$ ⇒ In-house$_{test}$ represents training on scans from the In-house dataset and segmenting images in the test set of the In-house dataset. ↑: Higher values are better. ↓: Lower values are better. The last column presents the number of model parameters (in millions). The best-performing results are shown in bold while the second-best results are indicated by underlining. ‡: Statistical significant with $P < 0.05$ in comparison to our SATS.

Method	In-house$_{train}$ ⇒ In-house$_{test}$		In-house$_{train}$ ⇒ External$_{test}$		Para. (M)
	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓	
UMambaBot	79.27 ± 7.77‡	4.66 ± 3.93	63.08 ± 12.02‡	9.22 ± 7.52	64.76
UNETR	75.75 ± 8.92‡	5.41 ± 4.07	62.56 ± 12.50‡	9.27 ± 7.46	93.01
TransUNet	78.95 ± 8.28‡	6.42 ± 12.89	62.96 ± 13.49‡	9.52 ± 8.16	119.37
SwinUNETR	80.01 ± 8.04	4.52 ± 2.77	62.90 ± 11.90‡	9.11 ± 7.41	62.19
SwinUNETR-V2	<u>80.41 ± 7.80</u>	4.17 ± 2.40	63.81 ± 12.11‡	8.90 ± 7.32	72.89
MedNeXt	76.15 ± 9.83‡	5.09 ± 3.93	<u>64.77 ± 12.05</u>‡	9.01 ± 7.50	61.80
STU-Net S	79.04 ± 7.30	4.95 ± 4.08	63.50 ± 11.96‡	9.07 ± 7.33	14.60
STU-Net B	78.86 ± 7.38	4.91 ± 3.98	63.54 ± 12.05‡	9.14 ± 7.46	58.26
STU-Net L	79.24 ± 7.23	4.64 ± 3.80	63.50 ± 11.91‡	9.09 ± 7.25	440.30
nnUNet	79.30 ± 9.77	<u>4.07 ± 2.77</u>	64.40 ± 11.82‡	<u>8.84 ± 7.40</u>	30.70
SATS (Ours)	**81.22 ± 8.33**	**4.02 ± 2.74**	**66.80 ± 12.02**	**8.51 ± 7.84**	30.70

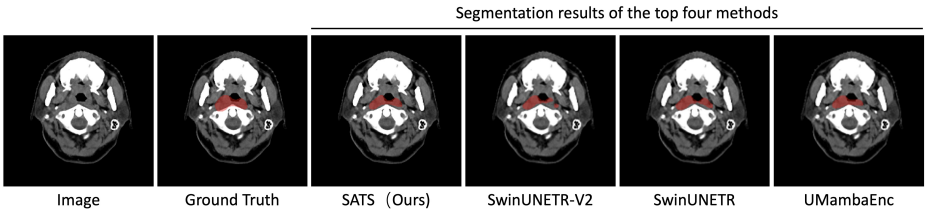

Fig. 3: Example CT slices with tumor segmentation overlays (red color) using different methods on the *In-house dataset*.

substantial improvement of 4.4% DSC over nnUNet in external evaluation. This demonstrates the better generalizability of the proposed semantic asymmetry learning in NPC GTV segmentation. Third, the proposed SATS consistently outperforms other leading methods in terms of HD95 (>3.7% error reduction). Lastly, although SwinUNETR-V2 performs 2nd best in internal testing with 1.11% DSC improvement over nnUNet, nnUnet outperforms SwinUNETR-V2 in external testing by 0.61% DSC. This result indicates the strong performance of CNN-based nnUNet over transformer-based segmentation models.

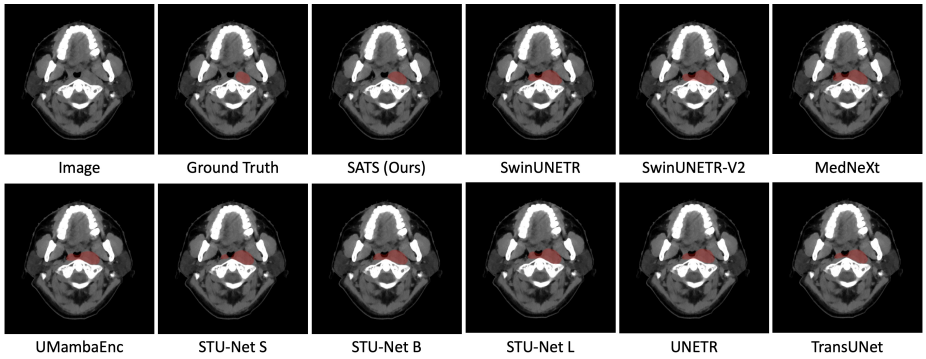

Fig. 4: Example CT slices with tumor segmentation overlays (red color) using different methods on the *External dataset*.

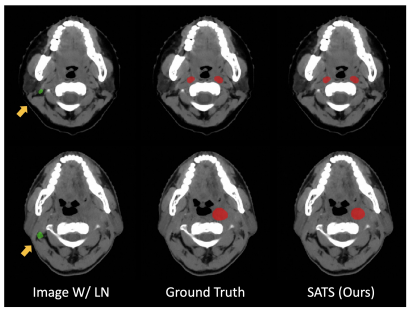

Fig. 5: Segmentation results for two nodal-involved (green) patients.

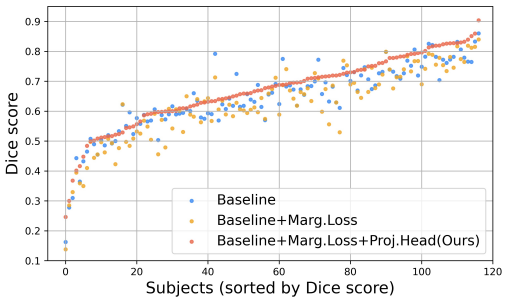

Fig. 6: Dice score compared to the baselines shown for each test subject.

3.3 Ablation Studies

Effect of projection head and margin loss. Table 2 demonstrates performance metrics for different segmentation models on the external data (In-house$_{train}$ $\Rightarrow$ External$_{test}$). There is a significant performance boost (+4.98% DSC, $-0.60mm$ ASD and $-1.15mm$ HD95) when both the projection head module and margin loss are taken into account. **Effect of semantic asymmetry learning.** In Figure 6, we present a comparative analysis of our method against the baseline configurations that exclude the projection head module and/or employ margin loss baselines. As depicted, our method demonstrates consistent superiority over all baseline models across the majority of the 117 test scans. **Failure cases analysis.** Our method performs poorly on extreme outliers in Figure 6, misclassifying symmetric lesions as asymmetric and achieving lower DSC than the nnUNet baseline. A complementary framework combining our approach with nnUNet could enhance clinical robustness.

Table 2: Influence of the effect of the projection head and margin loss.

Proj. Head	Marg. Loss	DSC (%)	ASD (mm)	HD95 (mm)
✗	✗	63.44 ± 10.54	2.97 ± 1.37	7.22 ± 3.34
✗	✓	61.50 ± 10.02	3.20 ± 1.39	7.73 ± 3.58
✓	✓	66.32 ± 10.48	2.60 ± 1.36	6.58 ± 3.50

4 Conclusion

We propose a novel semantic asymmetry learning method that leverages the inherent asymmetrical properties of tumors in the nasopharyngeal region. Our method demonstrates a significant improvement in NPC GTV segmentation by effectively utilizing semantic symmetry inherent in anatomical structures, achieving superior performance compared to state-of-the-art methods, as validated on both an internal test set and an independent external dataset.

Acknowledgments. This research was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant No.2024-KYI-00I-I05 and Zhejiang Provincial Spearhead & Pathfinder + X R&D Breakthrough Program under Grant No.2024C03043.

Disclosure of Interests. The authors have no competing interests.

References

1. Bai, X., Bai, F., Huo, X., et al.: Matching in the wild: Learning anatomical embeddings for multi-modality images. CoRR **abs/2307.03535** (2023) 5
2. Bai, X., Hu, Y., Gong, G., Yin, Y., Xia, Y.: A deep learning approach to segmentation of nasopharyngeal carcinoma using computed tomography. Biomedical Signal Processing and Control **64**, 102246 (2021) 2
3. Chen, H., Wang, Y., Zheng, K., Li, W., Chang, C., Harrison, A.P., et al.: Anatomy-aware siamese network: Exploiting semantic asymmetry for accurate pelvic fracture detection in x-ray images. In: ECCV. vol. 12368, pp. 239–255 (2020) 2
4. Chen, H., Qi, Y., Yin, Y., et al.: Mmfnet: A multi-modality MRI fusion network for segmentation of nasopharyngeal carcinoma. Neurocomputing **394**, 27–40 (2020) 2
5. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis p. 103280 (2024) 6
6. Chen, Y.P., Chan, A.T., Le, Q.T., Blanchard, P., Sun, Y., Ma, J.: Nasopharyngeal carcinoma. The Lancet **394**(10192), 64–80 (2019) 2
7. Chua, M.L., Wee, J.T., Hui, E.P., Chan, A.T.: Nasopharyngeal carcinoma. The Lancet **387**(10022), 1012–1024 (2016) 1
8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. vol. 12962, pp. 272–284 (2021) 6

9. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B.A., et al.: UNETR: transformers for 3d medical image segmentation. In: IEEE Winter Conference on Applications of Computer Vision. pp. 1748–1758 (2022) 6
10. He, Y., Nath, V., Yang, D., Tang, Y., Myronenko, A., Xu, D.: Swinunetr-v2: Stronger swin transformers with stagewise convolutions for 3d medical image segmentation. In: MICCAI. vol. 14223, pp. 416–426 (2023) 6
11. Heinrich, M.P., Jenkinson, M., Brady, S.M., Schnabel, J.A.: Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In: MICCAI. pp. 115–122 (2012) 5
12. Huang, J.b., Zhuo, E., Li, H., Liu, L., Cai, H., Ou, Y.: Achieving accurate segmentation of nasopharyngeal carcinoma in mr images through recurrent attention. In: MICCAI. pp. 494–502 (2019) 2
13. Huang, J., Li, H., Li, G., Wan, X.: Attentive symmetric autoencoder for brain MRI segmentation. In: MICCAI. vol. 13435, pp. 203–213 (2022) 2
14. Huang, W., Liu, W., Zhang, X., Yin, X., Han, X., Li, C., Gao, Y., et al.: Lidia: Precise liver tumor diagnosis on multi-phase contrast-enhanced ct via iterative fusion and asymmetric contrastive learning. In: MICCAI. pp. 394–404 (2024) 2
15. Huang, Z., Wang, H., Deng, Z., Ye, J., Su, Y., et al.: Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training. CoRR **abs/2304.06716** (2023) 6
16. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: MICCAI. pp. 488–498 (2024) 6
17. Ke, L., Deng, Y., Xia, W., Qiang, M., et al.: Development of a self-constrained 3d densenet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images. Oral Oncology **110**, 104862 (2020) 2
18. Lee, A.W., Ng, W.T., Pan, J.J., Poh, S.S., Ahn, Y.C., AlHussain, H.o.: International guideline for the delineation of the clinical target volumes (ctv) for nasopharyngeal carcinoma. Radiotherapy and Oncology **126**(1), 25–36 (2018) 6
19. Li, C., Zhang, X., Gao, Y., Yin, X., Lu, L., et al.: Improved esophageal varices assessment from non-contrast ct scans. In: MICCAI. pp. 349–359 (2024) 2
20. Li, S., Xiao, J., He, L., Peng, X., Yuan, X.: The tumor target segmentation of nasopharyngeal cancer in ct images based on deep learning methods. Technology in cancer research & treatment **18**, 153–160 (2019) 2
21. Li, Y., Dan, T., Li, H., Chen, J., Peng, H., Liu, L., Cai, H.: Npcnet: Jointly segment primary nasopharyngeal carcinoma tumors and metastatic lymph nodes in mr images. IEEE Transactions on Medical Imaging **41**(7), 1639–1650 (2022) 2
22. Liao, W., He, J., Luo, X., Wu, M., Shen, Y., et al.: Automatic delineation of gross tumor volume based on magnetic resonance imaging by performing a novel semisupervised learning framework in nasopharyngeal carcinoma. International Journal of Radiation Oncology* Biology* Physics **113**(4), 893–902 (2022) 2
23. Liu, C.F., Padhy, S., Ramachandran, S., et al.: Using deep siamese neural networks for detection of brain asymmetries associated with alzheimer's disease and mild cognitive impairment. Magnetic resonance imaging **64**, 190–199 (2019) 2
24. Liu, Y., Zhou, Z., Zhang, S., Luo, L., Zhang, Q., Zhang, F., et al.: From unilateral to bilateral learning: Detecting mammogram masses with contrasted bilateral network. In: MICCAI. vol. 11769, pp. 477–485 (2019) 2
25. Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., et al.: Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: MICCAI. pp. 318–329 (2021) 2

26. Luo, X., Liao, W., He, Y., Tang, F., Wu, M., Shen, Y., et al.: Deep learning-based accurate delineation of primary gross tumor volume of nasopharyngeal carcinoma on heterogeneous magnetic resonance imaging: a large-scale and multi-center study. Radiotherapy and Oncology p. 109480 (2023) 2
27. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. CoRR **abs/2401.04722** (2024) 6
28. Ma, Z., Zhou, S., Wu, X., Zhang, H., Yan, W., et al.: Nasopharyngeal carcinoma segmentation based on enhanced convolutional neural networks using multi-modal metric learning. Physics in Medicine & Biology **64**(2), 025005 (2019) 2
29. Mei, H., Lei, W., Gu, R., Ye, S., Sun, Z., Zhang, S., et al.: Automatic segmentation of gross target volume of nasopharynx cancer using ensemble of multiscale deep neural networks with spatial attention. Neurocomputing **438**, 211–222 (2021) 2
30. Men, K., Chen, X., Zhang, Y., Zhang, T., Dai, J., Yi, J., Li, Y.: Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. Frontiers in oncology **7**, 315 (2017) 2
31. Razek, A.A.K.A., King, A.: Mri and ct of nasopharyngeal carcinoma. American Journal of Roentgenology **198**(1), 11–18 (2012) 2
32. Roy, S., Köhler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., et al.: Mednext: Transformer-driven scaling of convnets for medical image segmentation. In: MICCAI. vol. 14223, pp. 405–415 (2023) 6
33. Tian, L., Li, Z., Liu, F., Bai, X., Ge, J., Lu, L., et al.: Same++: A self-supervised anatomical embeddings enhanced medical image registration framework using stable sampling and regularized transformation. ArXiv **abs/2311.14986** (2023) 3
34. Zeng, B., Wang, H., Xu, J., Tu, P., Joskowicz, L., Chen, X.: Two-stage structure-focused contrastive learning for automatic identification and localization of complex pelvic fractures. IEEE Transactions on Medical Imaging **42**(9), 2751–2762 (2023) 2