Head-and-Neck PET/CT Lesion Segmentation via SSIMH and SegResNet

Shuai Huang[1], Yuqi Cheng[1], Shilin Yu[1], and Hongtu Zhu[1]

University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
shuaishu@email.unc.edu, yuqic@unc.edu, shiliny@unc.edu,
htzhu@email.unc.edu

Abstract. We present a simple and effective pipeline for automatic detection and segmentation of primary tumors and lymph nodes in FDG-PET/CT for HECKTOR 2025 Task 1. The method starts with an anatomy-aware pre-crop of the head-and-neck region to suppress irrelevant context, followed by modality-specific intensity normalization with soft clamping. To mitigate cross-center domain shift, we apply single-subject, SSIM-guided spectrum swapping (SSIMH) on CT in the frequency domain without external references. For segmentation, we use a residual U-Net–style SegResNet with deep supervision and a combined Dice + Cross-Entropy loss. Training employs stratified five-fold cross-validation with foreground-centered sampling to emphasize small lesions. At inference, we use sliding-window tiling on the cropped volumes, lightweight post-processing to remove small isolated components, and a five-model ensemble by averaging per-voxel logits before softmax. On the official HECKTOR 2025 Task 1 test set, our approach achieves a GTVp Dice of **0.5779**, a GTVn aggregated Dice ($\mathrm{DSC}_{\mathrm{agg}}$) of **0.5280**, and a GTVn aggregated lesion-wise F1 of **0.3207**. The overall recipe is concise and reproducible, providing a strong and transparent baseline for multi-center head-and-neck PET/CT segmentation under domain shift. (Team name: BIGS2)

Keywords: HECKTOR 2025 · Head and neck cancer · PET/CT segmentation · Domain adaptation · SegResNet

1 Introduction and Task Motivation

Head and neck (H&N) cancers remain a major global burden and are prone to loco-regional failure despite modern therapy [1–3]. PET captures glycolytic activity while CT provides anatomy; their complementarity is crucial for accurate delineation of primary tumors and nodal disease [4]. HECKTOR 2025 extends prior editions with a larger, multi-center cohort (> 1200 patients across ≥ 11 centers) and refined metrics that emphasize both segmentation quality and *lesion detection* [5, 7]. **Task 1** targets fully automatic detection and segmentation of the primary tumor (GTVp) and metastatic lymph nodes (GTVn) on pre-treatment FDG-PET/CT, with evaluation combining instance-aware aggregated Dice ($\mathrm{DSC}_{\mathrm{agg}}$) and lesion-level F1 at IoU > 0.3 to reward methods that both *find* and *delineate* multiple lesions [6, 7]. Standardized nomenclature (AAPM TG-263) ensures consistency and clinical interoperability [7].

2 Data Description

The dataset comprises paired pre-treatment FDG-PET and low-dose, non-contrast-enhanced CT for each case, collected on combined PET/CT systems from multiple centers [4, 5]. Expert-drawn voxel-wise annotations provide a single label map with three classes: background (0), **GTVp** (1), and **GTVn** (2), following harmonized guidelines and centralized quality control; TG-263 naming ($GTVp$, $GTVn$) is adopted [7]. By design, imaging-defined targets may include contourable GTVn even when TNM reports N0, reflecting clinical practice in radiotherapy planning. The expanded 2025 cohort increases diversity and statistical power for benchmarking multimodal, multi-lesion segmentation under realistic domain shifts [5, 7]. The dataset builds upon previous HECKTOR editions [4, 5] and is now released as a unified multi-centric PET/CT resource [10].

3 Methods

3.1 Pre-processing

Data preparation. We first resample both CT and PET volumes to an isotropic resolution of **1×1×1 mm** and register them to a common voxel grid. To restrict the input to the clinically relevant head-and-neck (H&N) field-of-view while keeping the procedure fully automatic, we adopt a simple anatomy-aware cropping strategy: (i) detect the superior head boundary by applying a coarse PET threshold to locate metabolically active cranial tissue; (ii) estimate a 2D H&N center in the axial plane from the average foreground mask across superior slices; and (iii) extract a fixed-size **200×200×310 mm** bounding box centered at this point. In practice, this deterministic crop consistently covers the full H&N region across the training cohort while reducing the typical volume size from ∼500×500×900 to 200×200×310 voxels and eliminating irrelevant anatomy such as the thorax and couch.

Data normalization. After spatial standardization, we normalize CT and PET independently and then concatenate them into a 2-channel input $(\mathrm{CT}, \mathrm{PET})$:

- **CT:** intensities are linearly mapped from a predefined HU window to **[0,1]** and then passed through a **sigmoid** nonlinearity, which softly clamps outliers while preserving local contrast.
- **PET:** values are standardized to **zero mean and unit variance** across the cropped volume and similarly transformed by a **sigmoid** to dampen extreme uptake values.

This modality-specific normalization stabilizes optimization across centers and scanners while maintaining the relative CT/PET contrast that is crucial for joint tumor and lymph-node delineation.

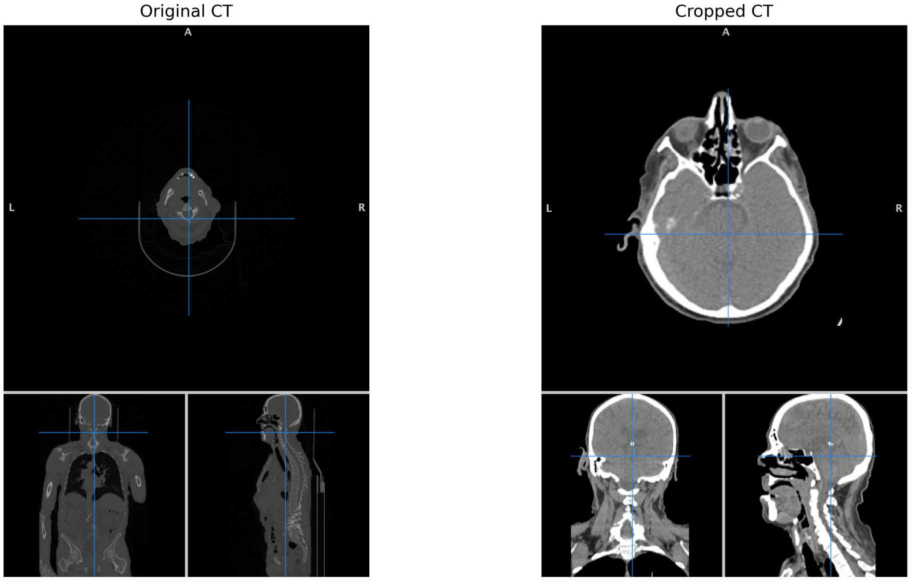

Fig. 1. Example of the pre-processing pipeline on a CHUM case. Left: original CT in scanner coordinates. Right: standardized head-and-neck crop after resampling to 1 mm isotropic spacing. The crop removes irrelevant anatomy while preserving the full tumor and nodal extent.

Domain adaptation via SSIMH. To mitigate inter-center appearance shifts, we adopt the **Spectrum Swapping-based Image-level Harmonization (SSIMH)** method [8, 9]. SSIMH operates in the frequency domain: given a source CT volume X_s and a reference (target) CT volume X_t, we compute their discrete cosine transforms (DCT), $F_s = \mathrm{DCT}(X_s)$ and $F_t = \mathrm{DCT}(X_t)$. Let $\Omega_{\mathrm{LF}}(\tau)$ denote the low-frequency band defined by a radius (threshold) τ; SSIMH replaces the source low-frequency coefficients with those of the reference,

$$F'_s[\Omega_{\mathrm{LF}}(\tau)] \leftarrow F_t[\Omega_{\mathrm{LF}}(\tau)], \quad X'_s = \mathrm{IDCT}(F'_s),$$

thereby transferring scanner/site-specific intensity/style while preserving high-frequency anatomical detail [8]. In practice, we apply SSIMH *slice-wise* with a 2D DCT on axial slices for efficiency and robustness [9]. Unless otherwise specified, we use the toolbox default $\tau = 3$ for the swapped low-frequency region. SSIMH is training-free and is executed after cropping but before normalizations; PET is left unchanged.

A qualitative example of SSIMH is shown in Fig. 2, demonstrating that low-frequency scanner/style statistics are aligned to the reference while anatomical high-frequency structures remain preserved.

Reference strategy. We fix a **single reference subject** from the training cohort and adapt each source CT to this reference (*one-to-one* harmonization). For

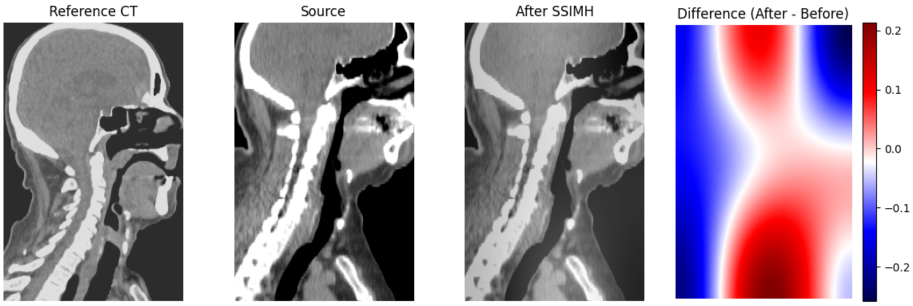

Fig. 2. SSIMH harmonization example. (CHUP-000) From left to right: (1) reference CT, (2) source CT before harmonization, (3) source CT after SSIMH, (4) intensity difference map (after – before). SSIMH selectively transfers low-frequency scanner/site appearance while preserving anatomical structures.

scenarios with multiple incoming scans per site, SSIMH also supports *batch* harmonization by adapting all source scans to a chosen site/template reference [9]. This yields a consistent intensity/style space across centers while leaving structural contrast largely intact.

Notes. (1) SSIMH is conceptually different from SSIM-guided histogram matching; our implementation follows frequency-domain spectrum swapping. (2) We clip X'_s to valid HU ranges and retain original voxel spacing/geometry so downstream resampling and segmentation are unaffected.

3.2 Network and Training

Architecture. We use **SegResNet** (residual U-Net) from MONAI as the backbone, with encoder–decoder residual blocks and deep supervision on decoder scales.

Patch sampling. From each pre-cropped H&N volume, we sample $\mathbf{192{\times}192{\times}192}$ voxel patches. Sampling is biased toward lesions to emphasize detection:

- probability **0.45** to center on **primary tumor**,
- probability **0.45** to center on **lymph nodes**,
- probability **0.10** to sample **background**.

Data augmentation. We apply **spatial** augmentations to both CT and PET: random affine transforms and random flips along all axes. For **CT** only, we apply **intensity** augmentations (random scale, shift, Gaussian noise, and blur). PET intensities are not perturbed.

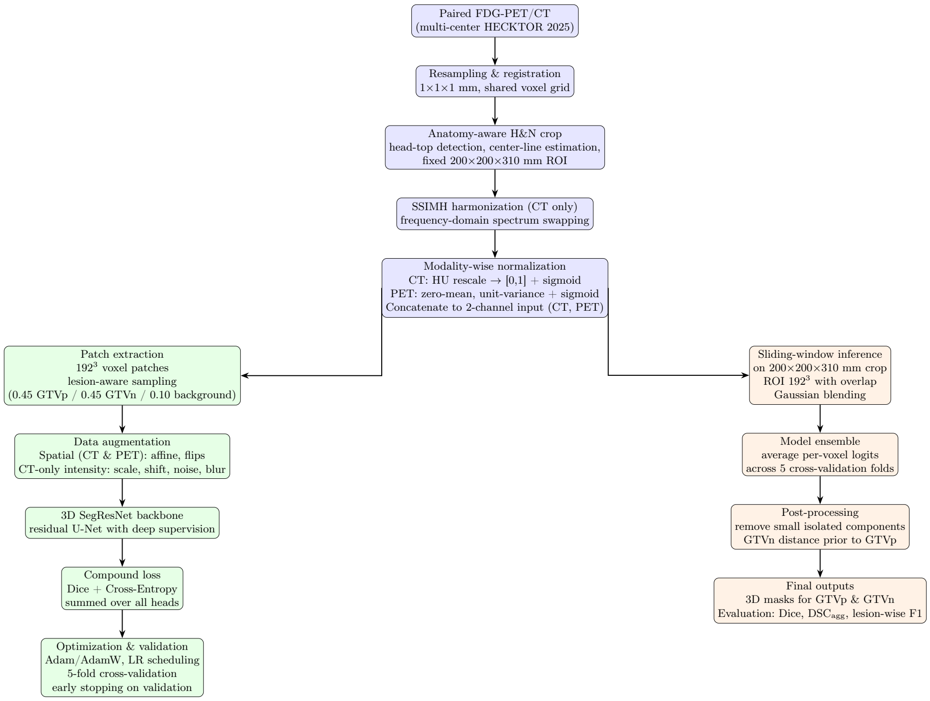
Paired FDG-PET/CT
(multi-center HECKTOR 2025)
Resampling & registration
1×1×1 mm, shared voxel grid
Anatomy-aware H&N crop
head-top detection, center-line estimation,
fixed 200×200×310 mm ROI
SSIMH harmonization (CT only)
frequency-domain spectrum swapping
Modality-wise normalization
CT: HU rescale → [0,1] + sigmoid
PET: zero-mean, unit-variance + sigmoid
Concatenate to 2-channel input (CT, PET)
Patch extraction
192^3 voxel patches
lesion-aware sampling
(0.45 GTVp / 0.45 GTVn / 0.10 background)
Data augmentation
Spatial (CT & PET): affine, flips
CT-only intensity: scale, shift, noise, blur
3D SegResNet backbone
residual U-Net with deep supervision
Compound loss
Dice + Cross-Entropy
summed over all heads
Optimization & validation
Adam/AdamW, LR scheduling
5-fold cross-validation
early stopping on validation
Sliding-window inference
on 200×200×310 mm crop
ROI 192^3 with overlap
Gaussian blending
Model ensemble
average per-voxel logits
across 5 cross-validation folds
Post-processing
remove small isolated components
GTVn distance prior to GTVp
Final outputs
3D masks for GTVp & GTVn
Evaluation: Dice, $\mathrm{DSC}_{\mathrm{agg}}$, lesion-wise F1

Fig. 3. Overview of the proposed pipeline for HECKTOR 2025 Task 1: shared pre-processing (blue), training pipeline (green), and inference/ensemble with post-processing (orange).

Objective and optimization. The training loss is **Dice + Cross-Entropy** computed at the main output and **summed over all deep-supervision heads**. Optimization uses standard stochastic gradient descent variants (e.g., Adam/AdamW) with cosine or step-decay scheduling; exact hyper-parameters are provided in code (omitted here for brevity).

Cross-validation protocol. We perform **5-fold cross-validation** with stratification by center and lesion presence where possible. Model selection is based on mean validation performance over folds. The final submission can be an ensemble (averaging logits) across the five fold-specific models.

4 Inference and Post-processing

At inference, we run **sliding-window** evaluation on the 200×200×310 mm crop using ROI size **192×192×192** with overlap and Gaussian blending. We take argmax over classes to obtain labels. We apply minimal post-processing: (i) connected-component "island" removal for both classes, discarding components smaller than a voxel-count threshold; and (ii) a distance-based prior on lymph nodes. Specifically, we compute a 3D Euclidean distance transform of the predicted GTVp and remove any GTVn component whose minimum surface-to-surface distance to GTVp exceeds a fixed threshold. This suppresses anatomically implausible, isolated node predictions far from the primary tumor while preserving contiguous or nearby nodal disease. No test-time augmentation is used unless specified.

5 Training and Validation Protocol

All experiments were conducted using the MONAI framework implemented in PyTorch. Prior to model training, CT and PET volumes were resampled to an isotropic resolution of 1 mm and cropped to a standardized head-and-neck field-of-view of $200 \times 200 \times 310$ mm using the official HECKTOR neck-cropping procedure. To mitigate inter-center variability, CT volumes were additionally harmonized with SSIM-based histogram matching, while both modalities were normalized using a sigmoidal intensity transform.

We adopted the 3D SegResNet architecture with deep supervision as the segmentation backbone. Training samples were generated by extracting 192^3 voxel patches according to a lesion-aware sampling strategy that oversamples tumor regions and challenging hard negatives (probabilities: 0.45/0.45/0.10). Data augmentation included spatial transformations applied to both CT and PET (random affine, elastic deformation, and flipping), supplemented with CT-specific intensity perturbations to reflect modality characteristics. Optimization was performed using a compound loss combining Dice and Cross-Entropy terms summed across all supervision scales.

Model development followed a 5-fold cross-validation scheme, with early stopping based on the validation loss to prevent overfitting. Consistent with the HECKTOR 2025 evaluation protocol, model performance was assessed using lesion-wise detection F1 score, Dice similarity coefficient for primary tumors and lymph nodes, surface-distance–based measures, and the aggregated Dice ($\mathrm{DSC}_{\mathrm{agg}}$) computed across all predicted lesions.

6 Results

On the internal validation split of HECKTOR 2025 Task 1, our method achieved: **GTVp Dice = 0.8393**, **GTVn aggregated Dice ($\mathbf{DSC}_{\mathbf{agg}}$) = 0.7657**, and **GTVn aggregated lesion-wise F1 = 0.5455**. Unless otherwise noted, these metrics were computed after sliding-window inference on the standardized $200 \times 200 \times 310$ mm head-and-neck crop with minimal post-processing and five-fold logit ensembling.

On the official HECKTOR 2025 Task 1 *test set*, our method obtained: **GTVp Dice = 0.5779**, **GTVn aggregated Dice ($\mathbf{DSC}_{\mathbf{agg}}$) = 0.5280**, and **GTVn aggregated lesion-wise F1 = 0.3207**. Test-set predictions were generated using the same pre-processing, SegResNet backbone, and ensemble configuration as in validation.

Table 1. Performance of the proposed method on the internal validation split and the official HECKTOR 2025 Task 1 test set.

Split	GTVp Dice ↑	GTVn DSC$_{\text{agg}}$ ↑	GTVn Lesion-F1 ↑
Validation (BIGS2)	0.8393	0.7657	0.5455
Test (BIGS2)	0.5779	0.5280	0.3207

Figure 4 summarizes the validation–test gap for the three challenge metrics, highlighting the impact of domain shift on lymph-node detection.

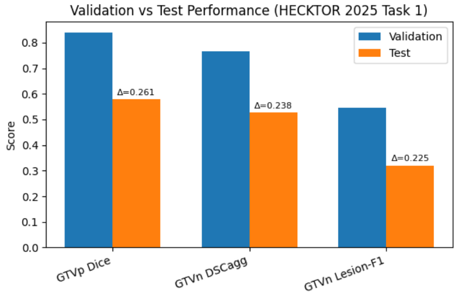

Fig. 4. Comparison of validation and test performance on HECKTOR 2025 Task 1.

In addition to the quantitative metrics, we provide a qualitative visualization of our model's prediction on a representative CHUM case in Fig. 5. The triplanar view shows that the model is able to localize both the primary tumor and involved lymph nodes, while most failure modes occur in small or low-uptake nodal regions, consistent with the lesion-wise F1 results.

7 Discussion and Interpretation

The primary tumor (GTVp) Dice of 0.8393 reflects accurate voxel-wise delineation in metabolically and morphologically conspicuous regions. For lymph nodes (GTVn), the aggregated Dice of 0.7657 indicates good overlap quality, while the aggregated lesion-wise F1 of 0.5455 highlights residual challenges in small or low-contrast nodal detection. This pattern—higher voxel-wise overlap than lesion-level detection—is consistent with the known difficulty of identifying small, scattered nodes in multi-center PET/CT. The anatomy-aware crop and foreground-biased sampling likely aided GTVp delineation; SSIMH was designed

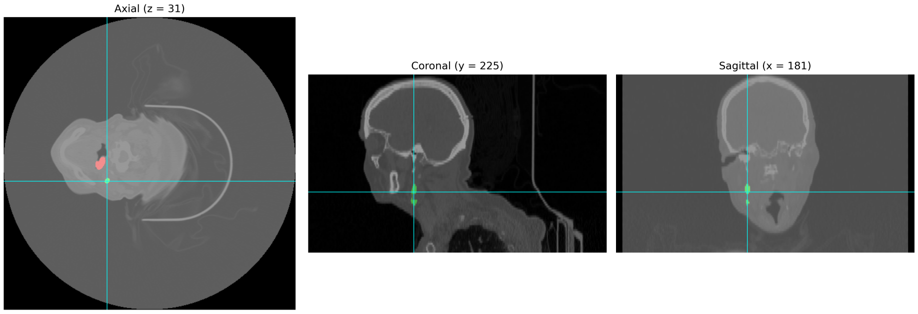

Fig. 5. Qualitative prediction example on case CHUM-001. The triplanar visualization overlays the predicted segmentation (red: GTVp, green: GTVn) onto the original CT volume. The model delineates the primary tumor clearly, while small, isolated nodal regions remain challenging.

to mitigate inter-center CT appearance shifts and stabilize nodal overlap, though lesion discovery remains the bottleneck reflected by F1.

8 Limitations and Future Work

First, fixed-size crops may be sub-optimal for extreme anatomies or cases where the primary tumor or nodal disease extends close to the superior or inferior bounds of the field-of-view; adaptive or anatomy-conditioned cropping could further reduce misses near volume boundaries.

Second, SSIMH currently uses a single reference subject for harmonization, which may not fully capture the diversity of acquisition protocols and scanner characteristics in all centers. Center-aware or multi-reference strategies could provide more robust harmonization without sacrificing structural fidelity.

Third, the pronounced drop in GTVn lesion-wise F1 from validation to test indicates that lesion-level detection for small nodes remains the main limitation. Future work will focus on improving nodal recall without inflating false positives, for example, via hard-negative mining, cascaded detection-and-segmentation schemes, or integrating uncertainty estimation into training and inference.

Finally, we have not yet explored test-time augmentation, more advanced ensembling strategies, or foundation-model initialization for PET/CT, all of which may further enhance generalization in multi-center settings.

9 Conclusion

We presented a compact, reproducible baseline for HECKTOR 2025 Task 1 using a MONAI SegResNet backbone combined with anatomy-aware pre-processing and SSIMH-based CT harmonization. The pipeline is intentionally simple—standardized resampling and head-and-neck cropping, frequency-domain CT harmonization, modality-specific normalization, lesion-aware patch sampling, and a five-model ensemble with minimal post-processing—yet achieves competitive performance on a challenging multi-center PET/CT segmentation task.

On the official HECKTOR 2025 test set, the method achieved a GTVp Dice of 0.5779, a GTVn aggregated Dice (DSC_{agg}) of 0.5280, and a GTVn aggregated lesion-wise F1 of 0.3207. These results indicate strong voxel-level segmentation for primary tumors and reasonable nodal overlap, while highlighting lesion-level detection for small nodes as a key area for improvement. We release this recipe as a transparent, easily reproducible baseline for future work on robust head-and-neck PET/CT segmentation and domain adaptation in large multi-center cohorts.

References

1. Parkin, D.M., Bray, F., Ferlay, J., Pisani, P.: Global cancer statistics, 2002. CA Cancer J. Clin. **55**(2), 74–108 (2005). https://doi.org/10.3322/canjclin.55.2.74
2. Bonner, J.A., Harari, P.M., Giralt, J., et al.: Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. Lancet Oncol. **11**(1), 21–28 (2010). https://doi.org/10.1016/S1470-2045(09)70311-0
3. Chajon, E., Lafond, C., Louvel, G., et al.: Salivary gland-sparing other than parotid-sparing in definitive head-and-neck intensity-modulated radiotherapy does not seem to jeopardize local control. Radiat. Oncol. **8**, 132 (2013). https://doi.org/10.1186/1748-717X-8-132
4. Oreiller, V., Andrearczyk, V., Jreige, M., et al.: Head and neck tumor segmentation in PET/CT: The HECKTOR Challenge. Med. Image Anal. **77**, 102336 (2022). https://doi.org/10.1016/j.media.2021.102336
5. Andrearczyk, V., Oreiller, V., Abobakr, M., et al.: Overview of the HECKTOR Challenge at MICCAI 2022: Automatic Head and Neck Tumor Segmentation and Outcome Prediction in PET/CT. In: Andrearczyk, V., Oreiller, V., Hatt, M., Depeursinge, A. (eds.) *Head and Neck Tumor Segmentation and Outcome Prediction*. HECKTOR 2022, LNCS, vol. 13626, pp. 1–30. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-27420-6_1
6. Kumar, N., Verma, R., Anand, D., et al.: A Multi-Organ Nucleus Segmentation Challenge. IEEE Trans. Med. Imaging **39**(5), 1380–1391 (2020). https://doi.org/10.1109/TMI.2019.2947628
7. HECKTOR Challenge Homepage, https://hecktor.grand-challenge.org/, last accessed 2025/09/12
8. Guan, H., Liu, S., Lin, W., Yap, P.-T., Liu, M.: Fast Image-Level MRI Harmonization via Spectrum Analysis. In: *Machine Learning in Medical Imaging (MLMI) 2022*, LNCS, vol. **13583**, pp. 201–209. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-21014-3_21
9. Guan, H., Liu, M.: Domain adaptation toolbox for medical data analysis. Neuroimage **268**, 119863 (2023). https://doi.org/10.1016/j.neuroimage.2023.119863
10. Saeed, N., Hassan, S., Hardan, S., Aly, A., Taratynova, D., Nawaz, U., Khan, U., Ridzuan, M., Andrearczyk, V., Depeursinge, A., Xie, Y., Eugene, T., Metz, R., Dore, M., Delpon, G., Papineni, V.R.K., Wahid, K., Dede, C., Ali, A.M.S., Sjogreen, C., Naser, M., Fuller, C.D., Oreiller, V., Jreige, M., Prior, J.O., Cheze

Le Rest, C., Tankyevych, O., Decazes, P., Ruan, S., Tanadini-Lang, S., Vallières, M., Elhalawani, H., Abgral, R., Floch, R., Kerleguer, K., Schick, U., Mauguen, M., Bourhis, D., Leclere, J.-C., Sambourg, A., Rahmim, A., Hatt, M., Yaqub, M.: *A Multimodal and Multi-centric Head and Neck Cancer Dataset for Segmentation, Diagnosis, and Outcome Prediction.* arXiv:2509.00367 (2025).