Relative Loss Balancing for High-Fidelity Synthetic CT Generation in the Abdomen

Michael de Francesco[1,2 [0009−0004−7392−0753]], Bowen Xin[1 [0000−0002−4545−9574]], Jason Dowling[2 [0000−0001−9349−2275]], Michaela Hoesl[3 [0000−0002−9177−1470]], and Xiuying Wang[1 [0000−0001−7160−5929]]

[1] School of Computer Science, Faculty of Engineering, The University of Sydney, Sydney, NSW, Australia
[2] Australian E-health Research Centre, CSIRO, Brisbane, Queensland, Australia
[3] MR Research Collaborations, Siemens Healthcare Pty Ltd, Melbourne, Victoria, Australia

Abstract. Accurate synthetic Computed Tomography (sCT) generation from magnetic resonance (MR) imaging is critical for downstream clinical applications such as radiotherapy dose calculation. Optimising multi-objective loss functions remains a key challenge in medical image translation. While the current state-of-the-art `nnsyn` algorithm demonstrates highly robust performance in sCT generation, its statically weighted multi-objective loss function lacks the adaptability required to fully resolve gradient interference in anatomically complex regions such as the abdomen. To address this, we present a dynamic loss weighting framework utilising Random Loss Balancing with Random Lookback (ReLoBRaLo). By dynamically balancing the multi-objective loss function as the model trains, our approach systematically avoids gradient conflict by initially prioritising global Mean Squared Error before shifting focus to Anatomical Feature Perception Loss. We evaluated the framework on 175 abdomen MR-CT patient pairs from the SynthRAD2025 Grand Challenge dataset. The ReLoBRaLo implementation yielded statistically significant improvements ($p < 0.05$) over the static `nnsyn` state-of-the-art baseline, SoftAdapt relative balancing and Gradient Normalisation (GradNorm) implementation across all evaluated metrics. It achieved superior voxel-wise similarity (MAE, PSNR, MS-SSIM) and geometric consistency (mDice, HD95). Furthermore, this structural fidelity is achieved with a negligible 0.3% computational overhead per epoch. Our framework provides a highly scalable, mathematically principled approach to loss balancing, successfully bridging the gap between global image structure and high-fidelity anatomical detail required for robust medical image translation.

Keywords: Synthetic CT · MR-to-CT Translation · Loss Function Balancing · Dynamic Weighting.

1 Introduction

Radiotherapy planning traditionally relies on Computed Tomography (CT) for dosimetric calculation and Magnetic Resonance Imaging (MRI) for superior soft-tissue target delineation [10]. However, co-registering these modalities inherently introduces anatomical mismatches. These errors are severely compounded in the abdomen and lungs due to intrinsic organ motion (e.g., respiratory and cardiac motion), leading to 2-5mm misalignments that can compromise dose distribution and risk healthy tissue [3, 6, 8].

An MR-only workflow resolves this by utilising deep learning to generate synthetic CTs (sCT) directly from MRI, eliminating registration errors and reducing unnecessary ionising radiation. The primary challenge in MR-to-CT translation is generating an sCT with both high voxel-wise image similarity and strict geometric boundary consistency. As highlighted by the SynthRAD2025 challenge report [9], achieving this accuracy in the abdomen and lungs is exceptionally difficult; the higher anatomical complexity, larger field of view, and massive inter-patient heterogeneity significantly degrade sCT fidelity compared to less mobile regions like the head and neck.

To address the limitations of existing generative approaches, this work builds upon the `nnsyn` framework [12]. This architecture was selected for its demonstrated state-of-the-art performance in the SynthRAD2025 challenge [11], where it secured overall first place among sCT generation methods. However, while the framework's robust ensembled ResUNet achieved high accuracy, its reliance on a 50/50 statically weighted composite loss function between Mean Squared Error (MSE) and Anatomical Feature Perception (AFP) [7] introduces a critical optimisation bottleneck. Static weights fail to adapt to evolving training dynamics, inevitably causing gradient interference between global image similarity and fine-grained anatomical perception. By adopting the `nnsyn` backbone, this research isolates this static limitation, allowing for a focused investigation into how dynamic balancing can overcome the flaws of static composite loss functions to further improve high-fidelity sCT generation.

Our main contributions are summarised as follows:

- We identify the inherent gradient interference between global image similarity and local geometric consistency caused by static loss formulations in medical image synthesis. To resolve this, we propose an adaptation of Relative Loss Balancing with Random Look-back (ReLoBRaLo) [1], introducing a dynamic loss weighting framework specifically tailored to mitigate gradient conflict in the MRI-to-CT translation task.
- We successfully integrate this balancing mechanism into the state-of-the-art `nnsyn` architecture, a state-of-the-art baseline model. In doing so, we demonstrate that dynamic objective balancing eliminates the manual tuning bottleneck of the original static implementation, elevating an already highly robust ensembled ResUNet backbone.
- Through a comprehensive ablation study, we provide empirical validation that our dynamically weighted framework yields statistically significant improvements over the original static baseline, as well as competing loss-balancing

algorithms including Gradient Normalisation (GradNorm) [2] and SoftAdapt [4]. By isolating the loss formulation, we demonstrate that our approach achieves superior performance across both voxel-wise intensity metrics (MAE, PSNR, MS-SSIM) and geometric boundary metrics (mDice, HD95), all while introducing negligible computational overhead to the training pipeline.

2 Methodology

2.1 Pre & post-processing

To ensure a fair and consistent comparison with the static baseline, all MR and CT volumes underwent the standardised pre-processing and post-processing pipelines established by the `nnsyn` framework [12]. Briefly, this included registering the CT to the MRI using elastix [5] and utilising intensity clipping with z-score normalisation. The network outputs were subsequently reverted to their original physical spacing and Hounsfield Unit scales prior to all quantitative evaluations.

2.2 Model Architecture

The core `nnsyn` model utilises the 3D ResUnet-L architecture. The network comprises 6 stages, containing [1, 3, 4, 6, 6, 6] blocks respectively, with corresponding feature map dimensions of [32, 64, 128, 256, 320, 320]. Each block consists of two Conv3d layers, two Instance Normalisation layers, and two LeakyReLU activations. Skip connections are employed to enhance synthesis quality.

The model was trained using a composite Masked Anatomical Perception loss, which is a combination of AFP loss [7] for geometric consistency and a masked MSE loss for image similarity. The baseline `nnsyn` model statically weighted each loss term λ_{mse} and λ_{afp} equally at 0.5. The final loss was defined as:

$$\begin{aligned} L_{MAP} =& \lambda_{mse} L_{MSE}(sCT \cdot m, CT \cdot m) + \\ & \lambda_{afp} L_{AFP}(f(sCT \cdot m - (1-m)), f(CT \cdot m - (1-m))) \end{aligned} \tag{1}$$

Where m is the body contour mask, sCT is the predicted CT, CT is the ground truth CT, and f is the student TotalSegmentator. Figure 1 shows the high-level architecture of the baseline state-of-the-art `nnsyn` model [12].

2.3 Dynamic Loss Framework

In standard generative frameworks, the composite loss is defined as a statically weighted sum of objective terms. However, enforcing static weights creates inherent gradient interference during MRI-to-CT translation. In the optimisation process, the network must update its weights to learn the macroscopic bulk density distribution (guided by the MSE loss). Conversely, localised anatomical boundary preservation (guided by the Masked AFP loss) is highly disruptive if

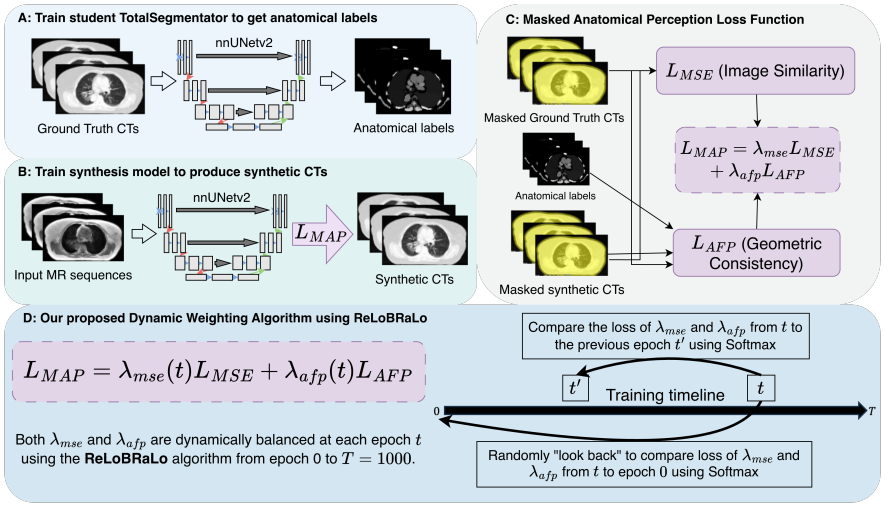

Fig. 1. Model architecture of `nnsyn` [12]. **A** shows how the segmented label maps of each CT are produced using TotalSegmentator and `nnUNetv2`. **B** demonstrates how the synthetic CTs are produced using `nnUNetv2`. **C** highlights the MAP Loss function, masking the synthetic and ground truth CT with the body contours provided. **D** details our proposed MAP loss function, which uses the ReLoBRaLo algorithm to dynamically balance the MSE and AFP loss (See eqs. (1) to (3) for the full balancing algorithm).

enforced before the global image similarity stabilises, often resulting in anatomical artefacts and training instability.

To autonomously resolve this objective conflict, we adapt the ReLoBRaLo algorithm [1] for the MR-to-CT synthesis task. ReLoBRaLo replaces static coefficients with dynamic, epoch-dependent weights that continuously adapt based on the relative convergence speed of each specific loss term.

The total dynamically weighted loss at epoch t is defined as:

$$L_{MAP}(t) = \lambda_{mse}(t) L_{MSE} + \lambda_{afp}(t) L_{AFP} \tag{2}$$

Unlike magnitude-based balancers, such as Gradient Normalisation (GradNorm) [2], ReLoBRaLo computes the dynamic weights $\lambda_i(t)$ for each objective $i \in \{mse, afp\}$ by evaluating the historical relative loss improvement (See section D of Figure 1). The algorithm for this loss balancing technique is defined as:

$$\begin{aligned} \lambda_i^{bal}(t, t') &= n \cdot \frac{\exp\left(\frac{\mathcal{L}_i(t)}{\mathcal{T}\mathcal{L}_i(t')}\right)}{\sum_{j=1}^{n} \exp\left(\frac{\mathcal{L}_j(t)}{\mathcal{T}\mathcal{L}_j(t')}\right)}, \quad i \in \{1, \ldots, n\} \\ \lambda_i^{hist}(t) &= \rho \lambda_i(t-1) + (1-\rho)\lambda_i^{bal}(t, 0) \\ \lambda_i(t) &= \alpha \lambda_i^{hist} + (1-\alpha)\lambda_i^{bal}(t, t-1) \end{aligned} \tag{3}$$

The default parameters of this algorithm were used as per the ReLoBRaLo official implementation [1]. The exponential decay rate $\alpha = 0.999$, the Bernoulli random variable $\rho = 0.99$, giving an $\mathbb{E}[\rho]$ close to 1. The value n is the total number of loss terms. The intermediate step $\lambda_i^{bal}(t, t')$ calculates the scalings based on the relative improvements of each term between time steps t' and t. The following step $\lambda_i^{hist}(t)$ defines whether the scalings calculated in the previous time step (ρ evaluates to 1) or the relative improvements since the beginning of training (ρ evaluates to 0) should be carried forward [1].

To isolate the efficacy of the proposed ReLoBRaLo framework, an ablation study was conducted against a static weighting baseline, as well as two established dynamic weighting algorithms: GradNorm and SoftAdapt. All models were trained under the default `nnsyn` hyperparameters, differing only in how λ_{mse} and λ_{afp} were calculated at each epoch.

3 Experiments & Results

3.1 Dataset

For this MR-to-CT translation task, the abdomen dataset was used from the SynthRAD2025 Grand Challenge dataset. The data were collected from 5 different centres, including three hospitals in the Netherlands and two hospitals in Germany [11]. In total, there were 175 MR-CT image pairs provided, as well as their associated body contour masks. During the algorithm development, a 5-fold cross-validation was embedded in the nnUNetv2 framework, with 140 pairs used for training and 35 for validation.

3.2 Quantitative Evaluation

Metric Evaluation All 175 sCTs produced by the 5-fold cross-validation models were evaluated against the ground truth CTs from the SynthRAD2025 Challenge dataset within the associated body contour masks. All metrics presented are derived from the internal validation split, as access to the official test set is restricted. The evaluation uses both image similarity metrics and geometric consistency metrics. The image similarity metrics included Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Multi-Scale Structural Similarity Index Measure (MS-SSIM). Geometric consistency evaluation included multi-class DICE coefficient (mDice), and Hausdorff Distance 95th percentile (HD95). They were used in local validation (the code for the metrics is available at: https://github.com/SynthRAD2025/metrics).

As detailed in Table 1, the integration of ReLoBRaLo yielded statistically significant improvements ($p < 0.05$) across all intensity metrics compared to the statically weighted baseline, except MS-SSIM and PSNR, which were approximately equal to the baseline. ReLoBRaLo achieved an average decrease of 0.0707 for MAE. Beyond global image similarity, the ReLoBRaLo implementation improved both the geometric consistency metrics, with an average increase

Table 1. Quantitative comparison of synthetic CT generation models. Results are reported as Mean ± Standard Deviation across 175 test patients. Best results are in **bold**. * denotes a p-value less than 0.05 and *** denotes a p-value less than 0.001, highlighting statistically significant improvement over the baseline `nnsyn` using a Wilcoxon signed-rank test.

Model	MAE (HU) ↓	PSNR (dB) ↑	MS-SSIM ↑	mDice ↑	HD95 (mm) ↓
Baseline `nnsyn`	35.28 ± 7.87	35.59 ± 1.93	0.968 ± 0.040	0.864 ± 0.044	4.27 ± 1.81
`nnsyn` + GradNorm	41.52 ± 9.20	33.81 ± 1.85	0.955 ± 0.050	0.829 ± 0.058	5.18 ± 2.48
`nnsyn` + SoftAdapt	**35.05 ± 7.81*****	**35.65 ± 1.92*****	**0.968 ± 0.038*****	0.867 ± 0.043***	4.22 ± 1.76
`nnsyn` + **ReLoBRaLo**	35.20 ± 7.83*	35.55 ± 1.93	0.968 ± 0.040	**0.868 ± 0.042*****	**4.16 ± 1.78*****

of 0.0040 for mDice score and a decrease of 0.1123 for HD95. The reduction in standard deviation for metrics in the `nnsyn` + ReLoBRaLo model demonstrate that the relative-improvement balancing not only shifts the median performance higher than the `nnsyn` baseline but also reduces the variance of the predictions.

Ablation Study As seen in Table 1, the ReLoBRaLo implementation outperforms SoftAdapt in geometric consistency (mDice, HD95), another relative loss-balancing algorithm, whist also achieving comparable image similarity metrics (MAE, PSNR, MS-SSIM). Validating our hypothesis that dynamically weighting composite loss algorithms produces sCTs with greater accuracy. Conversely, the GradNorm implementation suffered severe performance degradation. This implementation was statistically significantly worse in every metric when compared to the baseline `nnsyn`, validating our hypothesis that magnitude-based scaling is highly susceptible to gradient interference during MR-to-CT translation.

To contextualise the quantitative gains, Figure 2 provides a qualitative comparison of the lowest MAE patient case generated by the proposed model and baselines. As highlighted by the zoomed-in patches, the proposed ReLoBRaLo framework still preserves complex anatomical spinal boundaries compared to both the static baseline and competing dynamic methods, but still exhibits a lower MAE. Conversely, Figure 3 shows the worst-performing patient case by MAE. In this instance, registration mismatches drive elevated global MAE across all evaluated models; however the proposed ReLoBRaLo model still achieves slightly better anatomical perception of the spinal region. Thus showing this model does not introduce novel instabilities or geometric hallucinations when stressed, successfully preserving the foundational robustness of the architecture.

Dynamic Weighting & Training Efficiency To understand the mechanism driving these performance improvements, we analysed the trajectory of the dynamic loss weights (λ_{mse} and λ_{afp}) across the training epochs in fold 0 (see Figure 4). The empirical trajectory directly reflects our hypothesis regarding gradient interference. During the initial phase of training, ReLoBRaLo autonomously prioritises the macroscopic MSE objective, forcing the network to establish accurate bulk density mapping. As the bulk structure stabilises and the relative

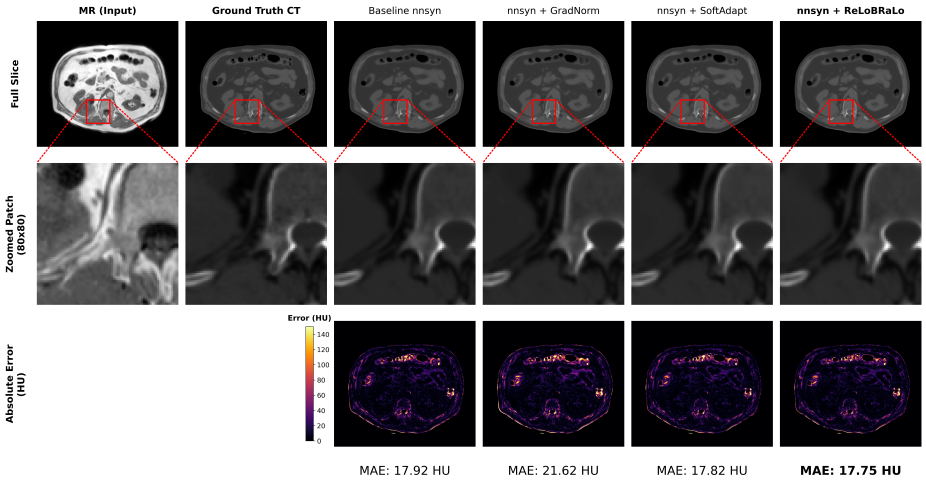

Fig. 2. Qualitative comparison of the best-performing case by MAE. A visualisation of the middle axial slice of Patient 1ABA104. The error maps show the HU difference between each model's sCT and the ground truth CT.

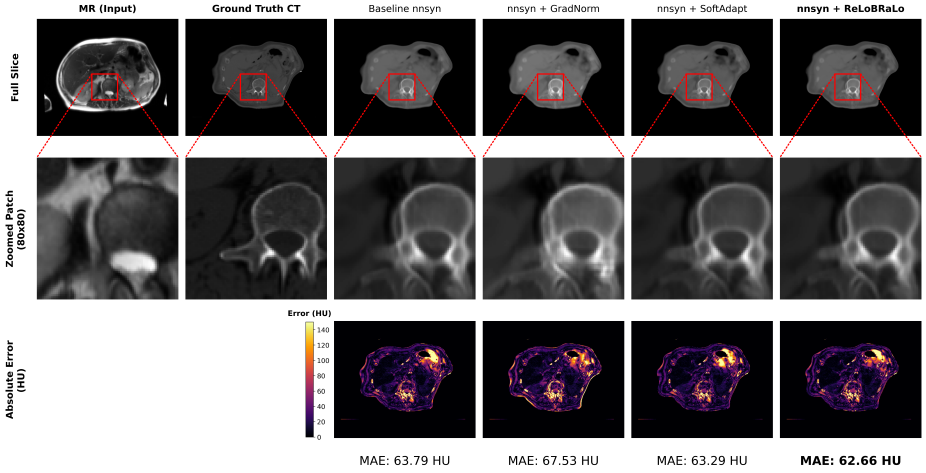

Fig. 3. Qualitative comparison of the worst-performing case by MAE. A visualisation of the middle axial slice of Patient 1ABC118. The error maps show the HU difference between each model's sCT and the ground truth CT.

improvement of the MSE loss plateaus, the framework reallocates gradient focus toward the localised AFP objectives.

Empirical logging of the training pipeline reveals an average epoch time of 69.81 seconds for the static `nnsyn` baseline, compared to 70.03 seconds for the ReLoBRaLo model. This represents an increase of only 0.22 seconds per epoch (a $\approx 0.3\%$ overhead). Because the algorithm strictly relies on loss statistics and a Softmax operation, it achieves optimal objective sequencing without compromising the scalability required for this medical translation task.

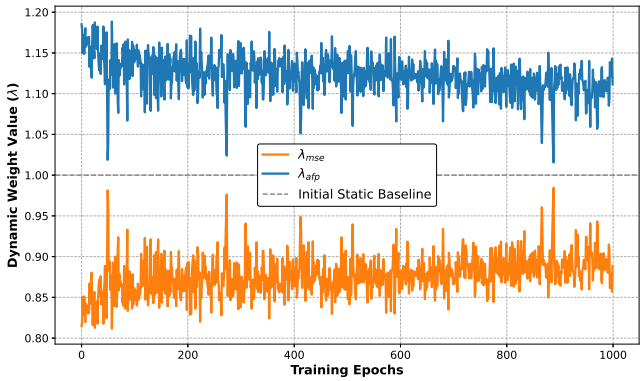

Fig. 4. The weight of λ_{mse} and λ_{afp} across 1000 epochs within fold 0 in the `nnsyn` model using ReLoBRaLo.

4 Discussion & Conclusion

In this work, we presented a dynamic loss weighting framework utilising ReLoBRaLo to optimise the synthesis of CT images from MR sequences. Our results demonstrate that an already accurate statically weighted baseline model, `nnsyn`, can be improved upon with a dynamic loss framework. By autonomously balancing the optimisation trajectory, our framework systematically avoids gradient interference. This dynamic reallocation yields synthetic CTs that not only achieve superior voxel-wise intensity metrics but also exhibit enhanced geometric consistency (mDice, HD95).

The clinical implications of this structural fidelity are substantial. Accurate preservation of soft tissue interfaces and skeletal boundaries is a strict prerequisite for downstream tasks such as synthetic CT-based dose calculation in radiotherapy. Furthermore, the integration of ReLoBRaLo introduces a negligible computational overhead of approximately 0.3% per training epoch. By relying on historical loss trajectories rather than complex auxiliary networks or second-order derivatives, the framework maintains the high-throughput scalability required for volumetric clinical datasets.

Future work will explore extending the dynamic weighting mechanism to other model architectures and further medical image translation tasks, as well as applying this proposed model to further patient data and clinical robustness tests that include dosimetry metrics. In conclusion, the integration of ReLoBRaLo provides a highly efficient, mathematically principled approach to loss balancing, successfully bridging the gap between global image structure and high-fidelity anatomical detail in MR-to-CT synthesis.

Acknowledgments. The authors acknowledge the financial support provided by the CSIRO Industry PhD program scholarship and Siemens Healthineers. The authors also gratefully acknowledge the use of computational resources and support provided by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) High-Performance Computing (HPC) systems, which were instrumental in training the models presented in this work.

Disclosure of Interests. The authors declare the following financial interests and potential competing interests: This research was financially supported by funding from Siemens Healthineers and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Bischof, R., Kraus, M.A.: Multi-objective loss balancing for physics-informed deep learning. Computer Methods in Applied Mechanics and Engineering **439**, 117914 (2025)
2. Chen, Z., Badrinarayanan, V., Lee, C.Y., Rabinovich, A.: Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In: International conference on machine learning. pp. 794–803. PMLR (2018)
3. Edmund, J.M., Nyholm, T.: A review of substitute ct generation for mri-only radiation therapy. Radiat Oncol **12**(1), 28 (2017). https://doi.org/10.1186/s13014-016-0747-y, https://www.ncbi.nlm.nih.gov/pubmed/28126030, edmund, Jens M Nyholm, Tufve eng Review England 2017/01/28 Radiat Oncol. 2017 Jan 26;12(1):28. doi: 10.1186/s13014-016-0747-y.
4. Heydari, A.A., Thompson, C.A., Mehmood, A.: Softadapt: Techniques for adaptive loss weighting of neural networks with multi-part loss functions. CoRR **abs/1912.12355** (2019), http://arxiv.org/abs/1912.12355
5. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging **29**(1), 196–205 (2010). https://doi.org/10.1109/tmi.2009.2035616, 1558-254x Klein, Stefan Staring, Marius Murphy, Keelin Viergever, Max A Pluim, Josien P W Journal Article Research Support, Non-U.S. Gov't United States 2009/11/20 IEEE Trans Med Imaging. 2010 Jan;29(1):196-205. doi: 10.1109/TMI.2009.2035616. Epub 2009 Nov 17.
6. Liu, Y., Lei, Y., Wang, Y., Wang, T., Ren, L., Lin, L., McDonald, M., Curran, W.J., Liu, T., Zhou, J., Yang, X.: Mri-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic ct generation method. Phys Med Biol **64**(14), 145015 (2019). https://doi.org/10.1088/1361-6560/ab25bc, 1361-6560 Liu, Yingzi Lei, Yang Wang, Yinan Wang, Tonghe Ren, Lei Lin, Liyong McDonald, Mark Curran, Walter J Liu, Tian Zhou, Jun Yang, Xiaofeng R01 CA184173/CA/NCI NIH HHS/United States R01 CA215718/CA/NCI NIH HHS/United States Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England 2019/05/31 Phys Med Biol. 2019 Jul 16;64(14):145015. doi: 10.1088/1361-6560/ab25bc.
7. Longuefosse, A., Denis de Senneville, B., Dournes, G., Benlala, I., Baldacci, F., Desbarats, P.: Anatomical feature-prioritized loss for enhanced mr to ct translation. Phys Med Biol **70**(14) (2025). https://doi.org/10.1088/1361-6560/adea07, https://www.ncbi.nlm.nih.gov/pubmed/40588008, longuefosse, Arthur Denis de Senneville, Baudouin Dournes, Gael Benlala, Ilyes Baldacci, Fabien Desbarats, Pascal eng England 2025/07/01 Phys Med Biol. 2025 Jul 14;70(14). doi: 10.1088/1361-6560/adea07.
8. Osman, M.M., Cohade, C., Nakamoto, Y., Wahl, R.L.: Respiratory motion artifacts on pet emission images obtained using ct attenuation correction on pet-ct. Eur J Nucl Med Mol Imaging **30**(4), 603–6 (2003). https://doi.org/10.1007/s00259-002-1024-x, https://www.ncbi.nlm.nih.gov/pubmed/12536242, osman, Medhat M Cohade, Christian Nakamoto, Yuji Wahl, Richard L eng Clinical Trial Germany 2003/01/22 Eur J Nucl Med Mol Imaging. 2003 Apr;30(4):603-6. doi: 10.1007/s00259-002-1024-x. Epub 2003 Jan 21.
9. Rogowski, V., Terpstra, M.L., Wahl, N., Kamp, F., van der Bijl, E., Galapon Jr., A., et al.: Generating synthetic computed tomography for radiotherapy: Synthrad2025 challenge report. arXiv preprint arXiv:2605.13555 (2026), https://arxiv.org/abs/2605.13555

10. Schmidt, M.A., Payne, G.S.: Radiotherapy planning using mri. Physics in Medicine & Biology **60**(22), R323 (2015). https://doi.org/10.1088/0031-9155/60/22/R323, https://doi.org/10.1088/0031-9155/60/22/R323
11. Thummerer, A., van der Bijl, E., Galapon, A.J., Kamp, F., Savenije, M., Muijs, C., Aluwini, S., Steenbakkers, R.J., Beuel, S., Intven, M.P., et al.: Synthrad2025 grand challenge dataset: Generating synthetic cts for radiotherapy from head to abdomen. Medical physics **52**(7), e17981 (2025)
12. Xin, B., Sun, Z., Min, H., Belous, G., Dowling, J.: Team koalai: Ensembled resunet with masked anatomical perception loss (2025), synthRAD2025 Challenge