Medical Image Analysis (2024)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Generating Synthetic Computed Tomography for Radiotherapy: SynthRAD2023 Challenge Report - Supplementary Document B

Evi M. C. Huijben[a,1,2], Maarten L. Terpstra[b,c,1,2], Arthur Jr. Galapon[d,2], Suraj Pai[e,2], Adrian Thummerer[d,f,2], Peter Koopmans[g,2], Manya Afonso[h,2], Maureen van Eijnatten[a,2], Oliver Gurney-Champion[i,j,2], Zeli Chen[k], Yiwen Zhang[k], Kaiyi Zheng[k], Chuanpu Li[k], Haowen Pang[l], Chuyang Ye[l], Runqi Wang[m], Tao Song[n], Fuxin Fan[o], Jingna Qiu[o], Yixing Huang[o], Juhyung Ha[p], Jong Sung Park[p], Alexandra Alain-Beaudoin[q], Silvain Bériault[q], Pengxin Yu[r], Hongbin Guo[s], Zhanyao Huang[s], Gengwan Li[t], Xueru Zhang[t], Yubo Fan[u], Han Liu[u], Bowen Xin[v], Aaron Nicolson[v], Lujia Zhong[w], Zhiwei Deng[w], Gustav Müller-Franzes[x], Firas Khader[x], Xia Li[y], Ye Zhang[y], Cédric Hémon[z], Valentin Boussot[z], Zhihao Zhang[aa], Long Wang[aa], Lu Bai[ab], Shaobin Wang[ab], Derk Mus[ac], Bram Kooiman[ac], Chelsea A. H. Sargeant[ad], Edward G. A. Henderson[ad], Satoshi Kondo[ae], Satoshi Kasai[af], Reza Karimzadeh[ag], Bulat Ibragimov[ag], Thomas Helfer[ah], Jessica Dafflon[ai,aj], Zijie Chen[ak], Enpei Wang[ak], Zoltan Perko[al,2], Matteo Maspero[b,c,2,*]

[a] *Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
[b] *Radiotherapy Department, University Medical Center Utrecht, Utrecht, The Netherlands*
[c] *Computational Imaging Group for MR Diagnostics & Therapy, University Medical Center Utrecht, Utrecht, The Netherlands*
[d] *Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*
[e] *Department of Radiation Oncology (Maastro), GROW School for Oncology, Maastricht University Medical Centre, Maastricht, The Netherlands*
[f] *Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, Germany*
[g] *Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, The Netherlands*
[h] *Wageningen University & Research, Wageningen Plant Research, Wageningen, The Netherlands*
[i] *Department of Radiology and Nuclear Medicine, Amsterdam UMC, location University of Amsterdam, Amsterdam, The Netherlands*
[j] *Cancer Center Amsterdam, Imaging and Biomarkers, Amsterdam, The Netherlands*
[k] *School of Biomedical Engineering, Southern Medical University, Guangzhou, China*
[l] *School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China*
[m] *School of Biomedical Engineering, ShanghaiTech University, Shanghai, China*
[n] *Fudan University, Shanghai, China*
[o] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[p] *Indiana University, Bloomington, USA*
[q] *Advanced Development Engineering, Elekta Ltd, Montreal, Canada*
[r] *Infervision Medical Technology Co., Ltd. Beijing, China*
[s] *Department of Biomedical Engineering, Shantou University, China*
[t] *Independent researchers*
[u] *Department of Computer Science, Vanderbilt University, Nashville, USA*
[v] *Australian e-Health Research Centre, CSIRO, Herston, Queensland, Australia*
[w] *Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California (USC), Los Angeles, California, USA*
[x] *University Hospital Aachen, Aachen, Germany*
[y] *Center for Proton Therapy, Paul Scherrer Institut, Villigen, Switzerland; Department of Computer Science, ETH Zurich, Zurich, Switzerland*
[z] *University Rennes 1, CLCC Eugène Marquis, INSERM, LTSI, Rennes, France*
[aa] *Subtle Medical, Shanghai, China*
[ab] *MedMind Technology Co. Ltd., Beijing, China*
[ac] *MRI Guidance BV, Utrecht, The Netherlands*
[ad] *Division of Cancer Sciences, The University of Manchester, United Kingdom*
[ae] *Muroran Institute of Technology, Hokkaido, Japan*
[af] *Niigata University of Health and Welfare, Niigata, Japan*
[ag] *Image Analysis, Computational Modelling and Geometry, University of Copenhagen, Denmark*
[ah] *IACS, Stony Brook University, NY, USA*
[ai] *Data Science and Sharing Team, Functional Magnetic Resonance Imaging Facility, National Institute of Mental Health, Bethesda, USA*
[aj] *Machine Learning Team, Functional Magnetic Resonance Imaging Facility National Institute of Mental Health, Bethesda, USA*
[ak] *Shenying Medical Technology (Shenzhen) Co., Ltd., Shenzhen, Guangdong, China*
[al] *Delft University of Technology, Faculty of Applied Sciences, Department of Radiation Science and Technology, Delft, The Netherlands*

*Corresponding author: Heidelberglaan 100, 3508 GA, UMC Utrecht, P.O. Box 85500 Utrecht, The Netherlands, Tel.: +31-88 75 67492;
e-mail: m.maspero@umcutrecht.nl (Matteo Maspero)
[1] Equally contributing first authors
[2] Challenge Organizer

This document is a supplementary document to Huijben and Terpstra et al. "Generating Synthetic Computed Tomography for Radiotherapy: SynthRAD2023 Challenge Report".

1. Supplementary analyses and results

This document provides additional analyses and results from the SynthRAD2023 challenge, which defined two tasks: 1) magnetic resonance imaging (MRI) to computed tomography (CT) synthesis and 2) cone beam CT (CBCT) to CT synthesis. These additional analyses include comparing the performance differences between all teams for each evaluation metric and investigating their statistical significance. In addition, we analyze the runtime of each team's algorithm and examine the average performance per patient in the test set. Finally, we visually analyze the results for two low-performing patients.

1.1. Teams' performance and significance

To state the significance of a team outperforming another in terms of individual metrics, we employed the Wilcoxon signed-rank test (Wilcoxon, 1945) with Holm's adjustment for multiple testing (Holm, 1979) for each metric separately (Figures 1 and 2). The significance level for this test is set at $\alpha = 0.05$. Based on the image similarity metrics, high-ranking teams robustly outperform lower-ranked teams. Statistical significant improvements were observed when comparing all image metrics between a team and another team ranked at least seven places lower for task 1, or six placed lower for task 2. However, for the dose metrics, this relation is weaker. In task 1, no statistical significant differences were observed between the top fourteen teams regarding the photon dose metrics and top eleven teams regarding the proton dose metrics. In task 2, no statistically significant differences were observed between the top eight teams regarding the photon and proton dose metrics, except for the fifth team (Pengxin Yu), which significantly outperforms the sixth team (KoalAI) regarding the proton DVH metric. Furthermore, Figure 3 shows a detailed overview of the resource utilization per team per subtask.

1.2. Data influence

For task 1, we performed a more detailed analysis of the influence of magnetic field strength on sCT generation performance. However, the absence of variability in magnetic field strengths for centers B and C constrained this analysis to center A (Figure 4). For the brain, the only significant difference was observed for γ_{photon}, which decreased from 98.99 ± 1.43 for 1.5T to 97.33 ± 3.23 for 3T. In contrast, for the pelvis, a significant increase in performance was observed for 3T compared to 1.5T. Specifically, the SSIM increased from 0.83 ± 0.05 to 0.84 ± 0.05, γ_{photon} increased from 97.51 ± 3.45 to 98.75 ± 2.59, and γ_{proton} increased from 93.29 ± 4.05 to 95.64 ± 3.42.

In analyzing the synthetic CT (sCT) generation performance at the patient level for both tasks, we present the mean SSIM and mean photon gamma pass rate per patient in Figure 5. One pelvis patient in center A for task 2 is severely underperforming in terms of photon gamma pass rate, while multiple brain patients from center B are underperforming in task 1. A visual inspection of two of these patients (1BB183 and 2PA039) (Figure 6) reveals that for patient 1BB183 (Figure 6a), the image quality of sCTs is high. However, the table of the CT scanner is still visible in the ground truth CT but not in the MRI. Jetta_Pang was able to synthesize a table, increasing the gamma pass rate ($\gamma = 90.24\%$) while the other teams did not synthesize a table, which led to large dose errors ($\gamma = 75 - 79\%$). This leads to higher dose deposition within the target while omitting the dose deposition in the table, which received more than 10% of the target dose and is therefore included in the gamma pass rate analysis, causing low gamma pass rates for these patients. Further investigation shows this happened for multiple brain patients from center B in task 1. Furthermore, patient 2PA039 (Figure 6b) suffers from a low-quality CBCT with a central artifact combined with a small field-of-view, making image synthesis difficult.

References

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics , 65–70.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1, 80–83. doi:10.2307/3001968.

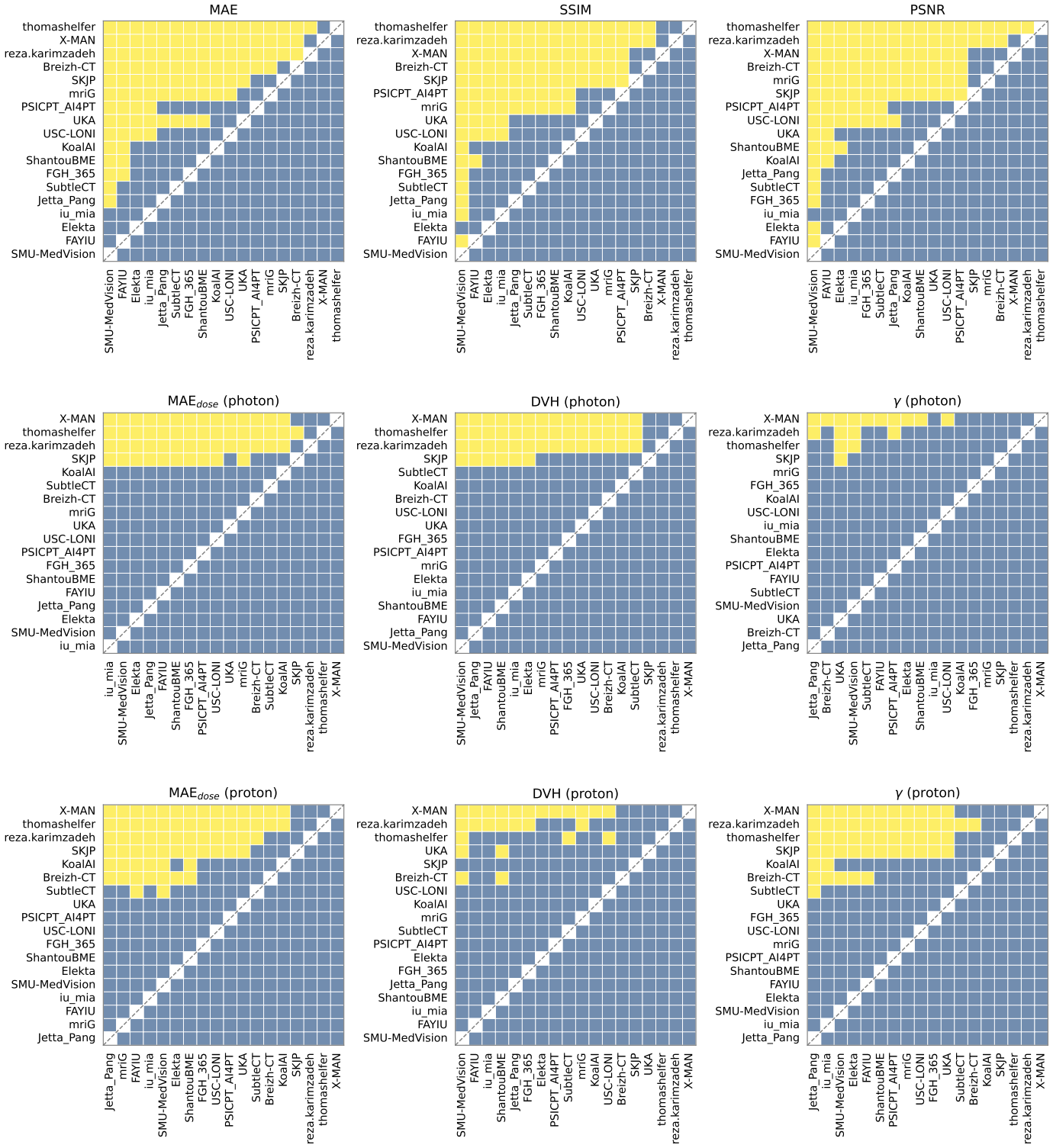


Fig. 1: Significance map of task 1. Visualizing pairwise comparisons of team performances for the individual metrics, where teams are sorted based on the performance for the respective metric. Yellow shading in the upper triangle indicates that the team on the x-axis performs significantly better than the y-axis, while blue indicates no significantly better performance. Blue in the lower triangle indicates that the team on the x-axis performs significantly worse than the team on the y-axis.

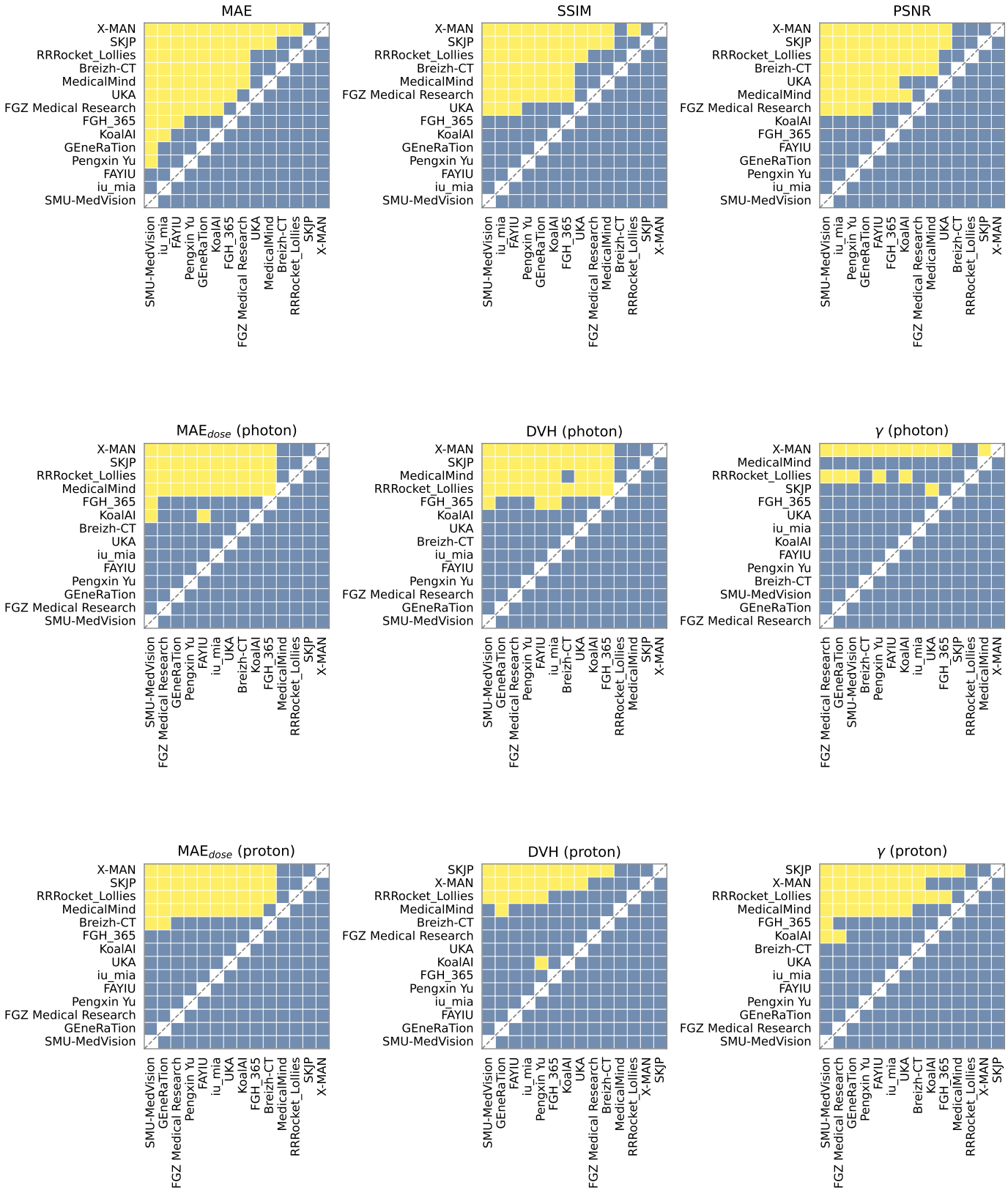


Fig. 2: Significance map of task 2. Visualizing pairwise comparisons of team performances for the individual metrics, where teams are sorted based on the performance for the respective metric. Yellow shading in the upper triangle indicates that the team on the x-axis performs significantly better than the y-axis, while blue indicates no significantly better performance. Blue in the lower triangle indicates that the team on the x-axis performs significantly worse than the team on the y-axis.

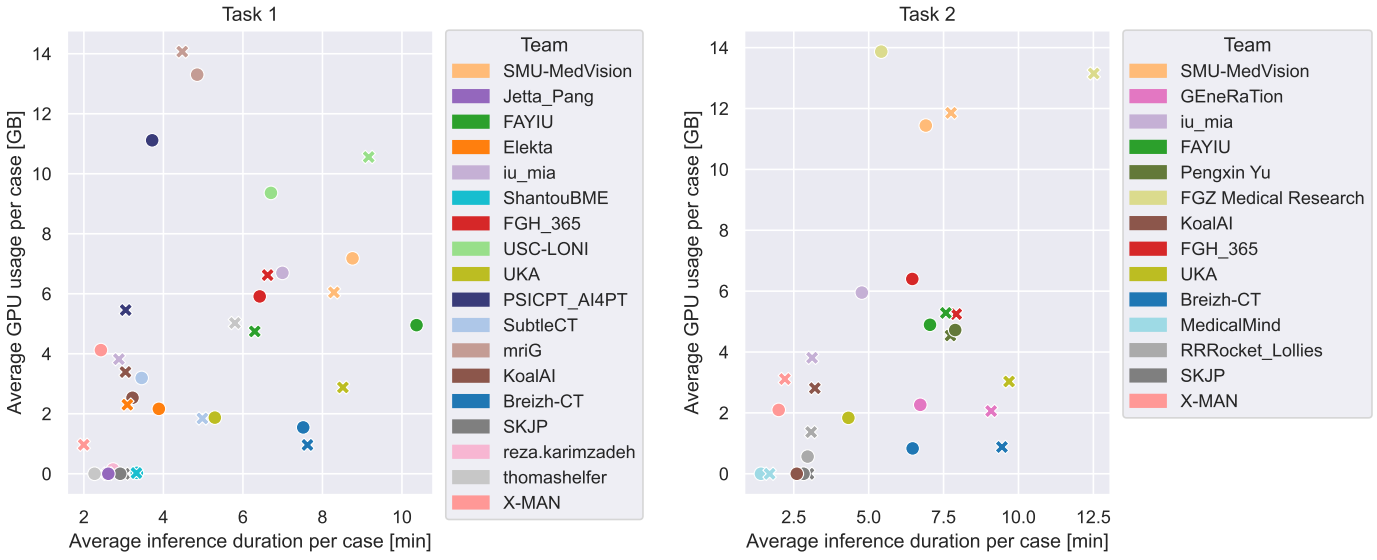


Fig. 3: Resource utilization during inference for tasks 1 (left) and 2 (right), represented by inference time and GPU usage. Values are averaged over 60 patients per subtask for each team, with different teams distinguished by colors. A cross represents brain averages, while a circle represents pelvis averages.

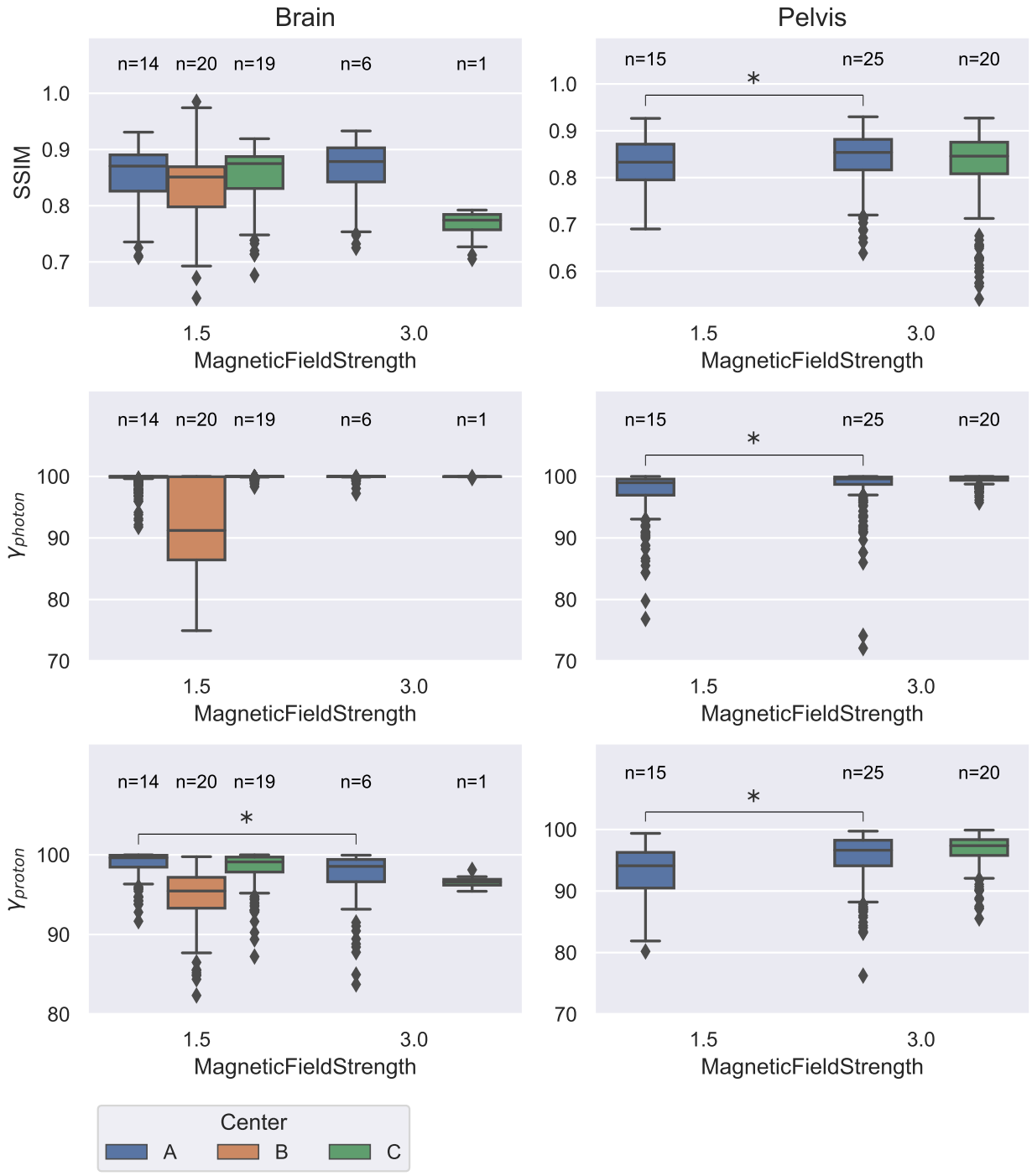


Fig. 4: Boxplots of the teams' performance for task 1 in terms of SSIM and gamma pass rates for photon and proton, grouped by different region, acquisition center and magnetic field strength. The number of test cases per subgroup is indicated by "n=x". Statistical significance (Mann-Whitney U-test, $\alpha = 0.01$) was calculated for center A between different field strengths of the same region and is indicated by an asterisk if significant.

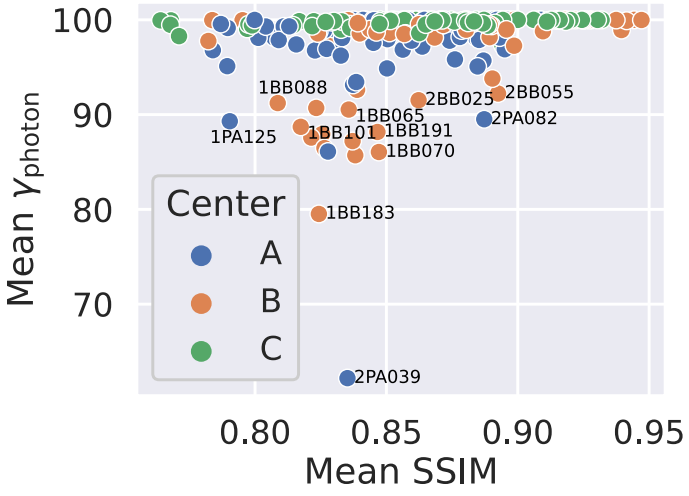
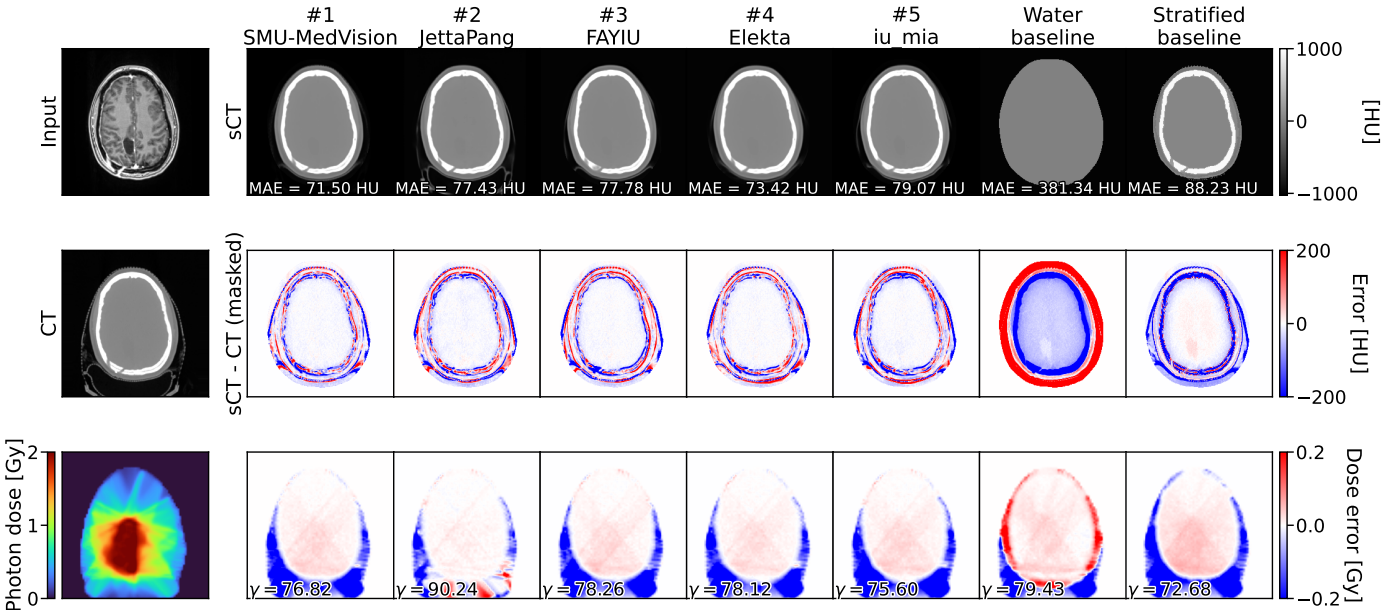
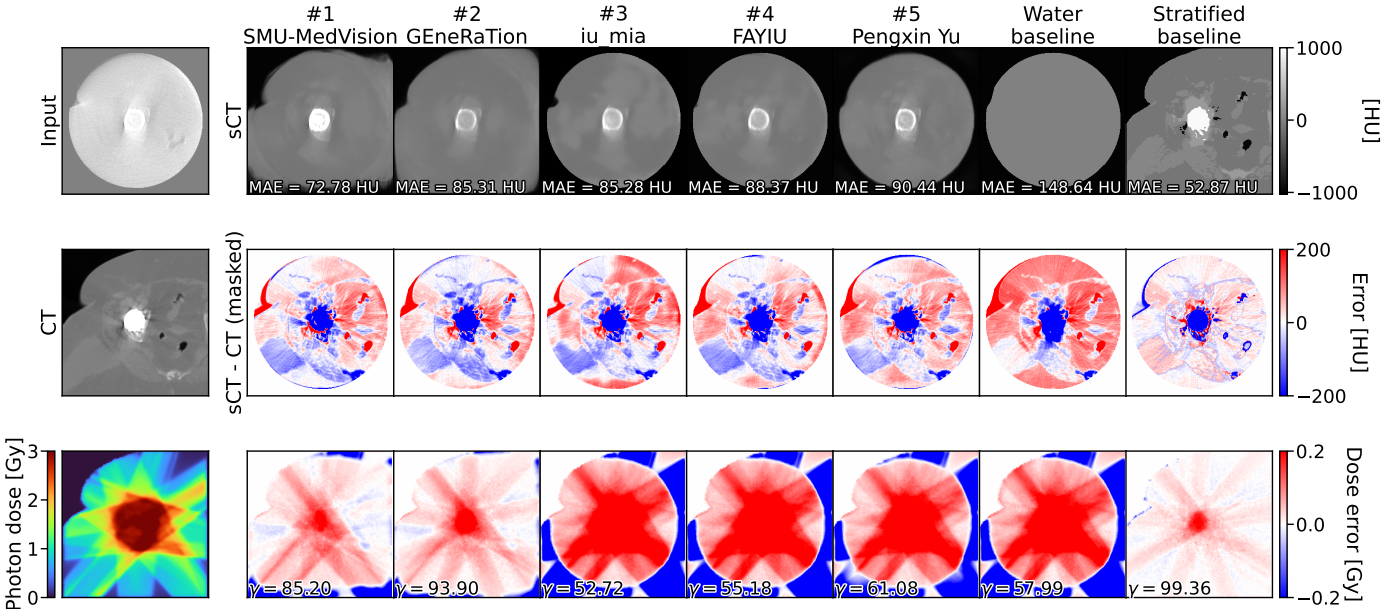


Fig. 5: Mean SSIM versus mean photon gamma pass rate per patient, averaged over all participants. The hue encodes the center. Two patients are clear outliers, while multiple patients of center B are underperforming.



(a) Example sCTs for outlier patient 1BB183. The image quality is high, but the CT still contains the table, which is difficult to synthesize if not present in the input. The image error map has been masked with the provided mask.



(b) Example sCTs for outlier patient 2PA039. The challenging anatomy and artifacts make accurate sCT generation difficult. The image error map has been masked with the provided mask.

Fig. 6: Examples of underperforming patients: patient 1BB183 for task 1 (MRI-to-CT; a) and patient 2PA039 for task 2 (CBCT-to-CT: b). The model input is shown in the upper left, and the ground truth is in the center-left. The sCT of the top five participants for task 1 and task 2 are shown in the top row. The difference from ground truth CT after masking with the provided mask is shown in the middle row. On the bottom left is the planned irradiation based on the CT for a photon (a) and proton (b) plan. The bottom row shows the dose difference when the treatment plan is applied to the sCT (CT dose - sCT dose).