CBCTLiTS: A Synthetic, Paired CBCT/CT Dataset For Segmentation And Style Transfer

M. E. Tschuchnig[1,2] and P. Steininger[3] and M. Gadermayr[1]

[1]Department of Information Technology and Digitalisation, Salzburg University of Applied Sciences, Austria
[2]Department of Artificial Intelligence and Human Interfaces, University of Salzburg, Austria
[3]MedPhoton GmbH, Austria

Abstract

Medical imaging is vital in computer assisted intervention. Particularly cone beam computed tomography (CBCT) with defacto real time and mobility capabilities plays an important role. However, CBCT images often suffer from artifacts, which pose challenges for accurate interpretation, motivating research in advanced algorithms for more effective use in clinical practice. In this work we present CBCTLiTS, a synthetically generated, labelled CBCT dataset for segmentation with paired and aligned, high quality computed tomography data. The CBCT data is provided in five levels of quality, reaching from a large number of projections with high visual quality and mild artifacts to a small number of projections with severe artifacts. This allows thorough investigations with the quality as a degree of freedom. We also provide baselines for several possible research scenarios like uni- and multimodal segmentation, multitask learning and style transfer followed by segmentation of relatively simple liver to complex liver tumor segmentation. CBCTLiTS is accesssible via https://www.kaggle.com/datasets/
maximiliantschuchnig/cbct-liver-and-liver-tumor-segmentation-train-data.

CCS Concepts

• Computing methodologies → 3D imaging; Image segmentation; Reconstruction; Matching;

1. Introduction

In the realm of computer-assisted interventions (CAI), precise and reliable imaging, especially intraoperative imaging, is paramount. Cone-Beam Computed Tomography (CBCT) is frequently employed to facilitate intraoperative interventions [RSC*06, RFF*23] by providing detailed, three-dimensional representations of a patient's anatomy utilizing a cone-shaped X-ray beam and a flat-panel detector, integrated into a mobile system [JSWM02]. However, rapid intraoperative imaging often comes with the disadvantage of significantly lower image quality [WAR*24] (compared to preoperative imaging), which can negatively impact the performance of downstream tasks such as segmentation. Typical applications for CBCT include intraoperative assessments and radiation therapy.

To assess the impact of reduced image quality on downstream tasks, the availability of image datasets with ground truth annotations is essential. We identified a scarcity of publicly available CBCT datasets with ground truth annotations limiting research in automated CBCT image analysis. On the dataset sharing platform Kaggle, e.g. the search term *CT Segmentation* returns 79 datasets while the term *CBCT Segmentation* returns only one. Due to this lack of publicly available datasets there is comparably little research in the field of CBCT image analysis, especially in the intraoperative setting. A comparison of Pubmed searches in the years 2014 - 2024 with the search strings *(CT) AND (Segmentation) AND (intraoperative)* and *(CBCT) AND (Segmentation) AND (intraoperative)* reveals, that research regarding intraoperative CBCT segmentation is only 5.8% as common as intraoperative computed tomography (CT) segmentation research.

As stated by Bilc et al. [BCL*23], the liver, as the largest solid organ in the human body, plays a crucial role in metabolism and digestion. It is also a common site for both primary and metastatic tumors, making it essential for comprehensive tumor staging and treatment planning. Primary liver cancer is the second leading cause of cancer-related deaths globally, and CT imaging is widely used to assess liver morphology, texture, and focal lesions. The Liver Tumor Segmentation Benchmark (LiTS) [BCL*23] dataset consists of overall 201 abdominal volumes (with 194 containing liver lesions), separated into a labelled training and an unlabelled testing dataset. We chose this well studied dataset (1000 citations) as a basis for our CBCT synthesised dataset due to the large amount of volumes, the availability of masks and the availability of two targets corresponding to different difficulty levels. These targets include the segmentation of relatively simple, regular and large liver areas and complex, small and irregular liver tumor areas.

Contributions: To enable thorough research in the field of CBCT analysis, a robust dataset with ground truth annotations
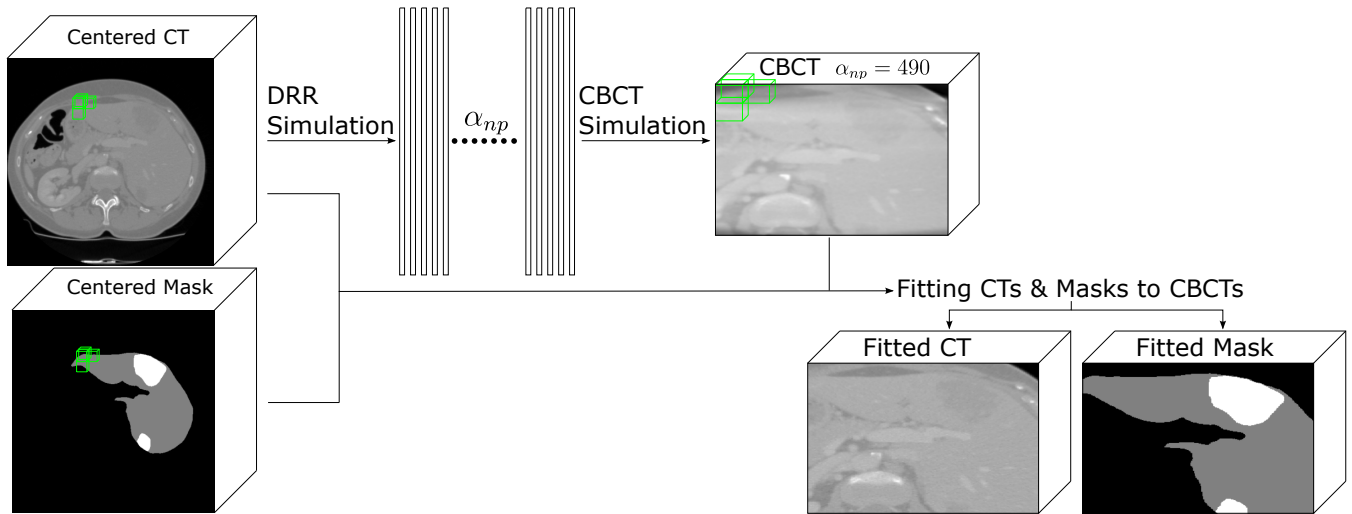
Figure 1: *Process of CBCTLiTS generation: after centering the original CT volumes around the liver (using the liver segmentations), DRRs were simulated from these centered CT. Then, CBCT were simulated and aligned with the original CT and masks.*

is needed. We present CBCTLiTS, an adaptation of the LiTS dataset [BCL*23], providing ground truth segmentation masks and paired CT samples. Our contributions include

- Dataset Creation: Development of CBCTLiTS, including CBCT with varying volume quality (five quality levels), ground truth segmentations and paired CT samples, facilitating various research scenarios such as uni- and multimodal segmentation, multitask learning, and style transfer. The dataset consists of 201 samples, 131 in the training dataset with segmentations and 70 in the testing dataset.
- Baseline Results: We provide baseline results for potential research directions, facilitated by CBCTLiTS, including uni- and multimodal segmentation, multitask learning and style transfer.

2. Dataset

To generate CBCTLiTS the steps shown in Figure 1, were taken. Initially, CT volumes were centered around either the liver (for training data with available liver masks) or the volumes middle (for test data). Following this, digitally reconstructed radiographs (DRR) were generated, and utilized to synthesise CBCT scans. The CBCT volume quality was varied by adjusting the number of DRRs (undersampling) used in reconstruction. Finally, the original CT volumes and corresponding masks were fitted to the synthesised CBCT to match the same field-of-view.

As a basis for generating synthetic CBCT scans, real, diagnostic CT volumes are needed. CT represents an imaging technology corresponding with a high visual quality with relatively few artifacts and improved image quality parameters like intensity homogeneity (compared to CBCT imaging). The LiTS CT dataset [BCL*23] was chosen as a basis to generate the synthetic CBCTLiTS data set. LiTS comprises 131 abdominal CT scans in the training set and 70 test volumes. The 131 training volumes include segmentations

of both the liver and liver tumors. This dataset contains data from seven institutions with a diverse array of liver tumor pathologies, including primary and secondary liver tumors with varying lesion-to-background ratios. It also features a mix of pre- and post-therapy CT scans. The CT images were acquired using different scanners and acquisition protocols. We refer to Bilic et al. [BCL*23] for further LiTS details. LiTS was chosen due to its large number of samples (201), label balance (ranging from relatively simple, regular and large liver area to complex, small and irregular liver tumor area) and prominence. LiTS is also part of The Medical Segmentation Decathlon dataset [ARB*22].

In order to simulate CBCT from the CT volumes, DRRs were generated using Siddon ray tracing [Sid85] under the assumption of a monochromatic X-ray spectrum. Given that standard CBCT with a centered detector typically exhibit small fields-of-view (approximately $25cm^3$), only specific areas of the original CT were extracted. To make effective use of the provided ground truth masks, the liver area was chosen as the center for CBCTLiTS training data. This center was calculated by taking the center of gravity of the ground truth liver masks. The cone-beam projection geometry for DRR simulation was derived from the Loop-X (TM) [KKSB22, KSB*22, RFF*23, WAR*24], a mobile CBCT scanner used, e.g., in navigated spine surgery and radiation therapy. These projections were then used in a CBCT reconstruction algorithm, implementing filtered back-projection, specifically the Feldkamp-Davis-Kress (FDK) [FDK84] algorithm. The reconstruction size of the CBCT was set to $252mm \times 246mm \times 250mm$ around the center with a voxel size of $0.688 \times 1.032 \times 0.688$. Unlike in the case of real-world CBCT scans, X-ray pre-processing steps such as beam hardening correction or Compton scatter correction were omitted, since idealized DRRs do not suffer from such effects but show primary signals only.

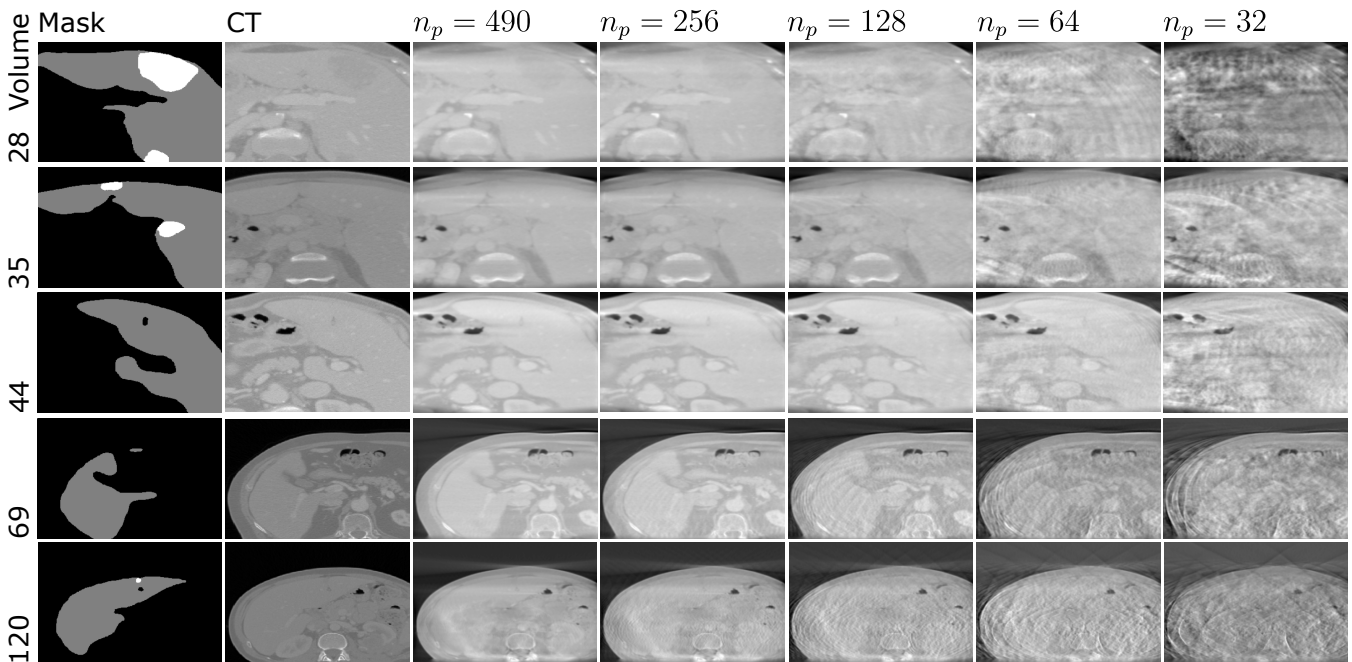To simulate CBCT of different qualities, CBCT reconstruction

Figure 2: *Different volumes of CBCTLiTS data (vertically). Horizontally, the presented data consists of the aligned mask, the aligned CT and the five different qualities of synthesised CBCT. n_p represents the number of projections used for CBCT synthesis.*

was undersampled (by varying numbers of DRRs used for reconstruction). For different numbers of projections n_p, the angular gap between the generated DRRs was equidistantly distributed over a 360 degrees scan range of simulated full-fan acquisitions. The number of projections n_p was varied between $\{490, 256, 128, 64, 32\}$ to generate scans with high visual quality similar to CT ($n_p = 490$) as well as scans showing clear degradations ($n_p < 128$) and scans in between (qualitative evaluation). Tasks like segmentation and style transfer are facilitated by aligning the fields-of-view, rotation, and spacing of the CT and masks with the synthesized volumes. Linear 3D transformation, combined with resampling is applied to align the mask and CT with the simulated CBCT.

3. Experimental Setup

To provide baselines for several possible research directions enabled by CBCTLiTS, this section introduces a segmentation baseline, based on unet [RFB15] and model adaptations investigating multitask learning (Subsect. 3.1), multimodal learning (Subsect. 3.2) and style transfer (Subsect. 3.3). All evaluated segmentation models are based on 3D unet [ÇAL*16]. This 3D unet was adapted by applying batch norm after each layer in the double convolutional blocks.

All models were trained utilizing a sum of binary cross-entropy and Dice similarity. Training was performed on an Ubuntu server using NVIDIA RTX A6000 graphics cards. To binarize the masks, a threshold of 0.5 was applied to each channel of the unet (segmentation) output. All experiments were trained and evaluated four

times to facilitate stable results with the same random splits. Adam was used as an optimizer with a learning rate of 0.005. For training, the converted CBCTLiTS training data was separated into training-validation-testing splits with ratios of 0.7 (training), 0.2 (validation), and 0.1 (testing), respectively. In experiments where the whole volume was applied directly (holistic) the volumes had to be downscaled (isotropic) by the factor of two due to the large size of the volumes and memory restrictions (48 GB VRAM).

3.1. Combining segmentation and image reconstruction using multitask learning

The aim of this adaptation is to investigate if the additional target of image reconstruction added to a segmentation model is beneficial for training the segmentation model [TCMSG24]. To accomplish this, a multistream architecture is utilized with two layers in the end, with the first layer establishing baseline segmentation. The second stream leads into a 3D convolutional layer with a linear activation function to facilitate image reconstruction. Therefore the updated loss contains a segmentation loss as well as an image reconstruction loss (l2). Two cases of this additional image reconstruction setting are investigated. The first investigated case is the reconstruction of the input CBCT (Table 1: multitask-c), i.e. the second stream is similar to an autoencoder. It aims to add morphological regularization to model training. The second case investigated how the reconstruction of the highest possible quality CBCT effects segmentation performance (Table 1: multitask-b) adding additional denoising effects. For more information we refer to Tschuchnig et al. [TCMSG24].

Table 1: *Results based on CBCTLiTS. Results are presented for different CBCT qualities and liver as well as liver tumor segmentation. The first results show baseline segmentation scores using CBCT, both holistically and patched. Then, results of multitask learning approaches multitask-b and multitask-c, followed by multimodal learning approaches with varying degrees of misalignment are presented. Finally, the basic 3D unet was applied to CBCT data, style transferred into the CT domain. Data presented in bold was superior to its corresponding baseline (holistic or patched). Baseline CT (aligned) segmentation reached a mean Dice of 0.930 for liver and 0.303 for tumor segmentation.*

			Liver Segmentation					Liver Tumor Segmentation				
			490	256	128	64	32	490	256	128	64	32
	base CBCT	[TCMSG24]	0.884	0.884	0.859	0.817	0.784	0.165	0.162	0.093	0.061	0.029
	base patch CBCT	[TCMSG24]	0.811	0.814	0.768	0.693	0.650	0.122	0.122	0.060	0.042	0.028
3.1	multitask-c	[TCMSG24]	0.877	0.879	**0.874**	**0.826**	0.771	0.149	0.153	0.092	0.058	**0.049**
	multitask-b	[TCMSG24]	0.877	0.884	**0.871**	**0.836**	0.775	0.149	0.147	0.078	0.051	0.028
	patch multitask-c	[TCMSG24]	**0.847**	**0.821**	**0.786**	**0.720**	0.609	**0.171**	**0.127**	**0.112**	0.036	**0.031**
	patch multitask-b	[TCMSG24]	**0.847**	**0.831**	**0.771**	**0.720**	0.591	**0.171**	**0.154**	**0.079**	0.028	0.021
3.2	perfect alignment	[TSG24]	**0.933**	**0.931**	**0.931**	**0.933**	**0.932**	**0.330**	**0.322**	**0.298**	**0.325**	**0.297**
	affine-s1	[TSG24]	**0.906**	**0.897**	**0.879**	**0.857**	**0.806**	0.154	0.124	0.057	0.023	**0.051**
	affine-s0.5	[TSG24]	**0.895**	**0.891**	**0.875**	**0.863**	**0.851**	0.096	0.098	0.077	0.056	**0.107**
	affine-s0.25	[TSG24]	**0.904**	**0.906**	**0.893**	**0.883**	**0.883**	**0.181**	0.160	**0.163**	**0.169**	**0.171**
	affine-s0.125	[TSG24]	**0.908**	**0.900**	**0.889**	**0.884**	**0.887**	**0.187**	**0.166**	**0.166**	**0.152**	**0.185**
	elastic-s1	[TSG24]	**0.907**	**0.897**	**0.880**	**0.857**	**0.801**	0.153	0.122	0.057	0.022	**0.049**
	elastic-s0.5	[TSG24]	**0.895**	**0.891**	**0.870**	**0.845**	**0.800**	0.098	0.103	0.085	0.029	**0.049**
	elastic-s0.25	[TSG24]	**0.893**	**0.898**	**0.869**	**0.853**	**0.846**	0.154	0.142	**0.136**	**0.129**	**0.122**
	elastic-s0.125	[TSG24]	**0.910**	**0.901**	**0.891**	**0.885**	**0.887**	**0.186**	**0.168**	**0.163**	**0.151**	**0.168**
3.3	sty-transfer CBCT		**0.901**	**0.888**	**0.867**	**0.835**	**0.787**	0.104	0.097	0.065	0.029	**0.040**

3.2. Multimodal learning, combining preoperative CT and intraopertative CBCT

One approach to improve intraoperative imaging is to combine low quality intraopterative volumes with potentially misaligned, high quality, preoperative volumes [TSG24]. This multimodal learning approach was investigated by performing early fusion of CT, as high quality, preoperative data and low quality CBCT, as intraoperative data. Different forms of misalignment with affine (random scaling, rotation and translation, Table 1: affine-s) and affine followed by elastic misalignment (Table 1: elastic-s) were performed, simulating real world scenarios with different degrees of alignment. To keep the number of parameters minimal, an alignment parameter α_a was introduced, controlling the maximum strength of all misalignment methods. Misalignment was performed using TorchIO RandomAffine and RandomElasticDeformation. For more information we refer to Tschuchnig et al. [TSG24].

3.3. Converting CBCT to CT using style transfer

Since the synthesised CBCT volumes are of lower visual quality then the CT volumes, a style transfer from the CBCT domain to the CT domain has the potential of segmentation performance increases. By performing this image-to-image style transfer, structureal CBCT artifacts and noise can theoretically be removed. This theory was experimentally investigated by using the CBCTLiTS test dataset to train a 3D Pix2Pix model, available at `https://github.com/MaxTschuchnig/TFVox2Vox`, for style transfer from the CBCT to the CT domain. Then, this model was applied to the CBCT training dataset to convert the CBCT scans of different data qualities to CT scans. The transformed data was subsequently used for liver and liver tumor segmentation using the baseline segmentation unet (Table 1: sty-transfer CBCT).

4. Results

Figure 2 shows example results of the data generation process with several CBCTLiTS samples. It displays, from left to right random samples of the masks and CT, aligned and in the same field of view as the generated CBCT followed by different visual quality levels of the CBCT. Liver lesions are labelled with the value 2 (white) and the liver with the value 1 (grey). The background voxels are labeled as 0 (black).

Table 1 shows results of different segmentation setups with the goals of liver and liver tumor segmentation. All presented results are based on CBCTLiTS, with the first results showing segmentation scores using CBCT, both holistically and patched. Further, results of three different adaptations, multitask learning, multimodal learning and style transfer, using the unique components of CBCTLiTS are presented. All CBCT results are reported for the different CBCT qualities (labeled using the amount of projections n_p) and for both liver and liver tumor segmentation. Baseline segmentation using the aligned, high quality CT reached a mean Dice of 0.930 for liver and 0.303 for liver tumor segmentation. For comparison, one of the best performing models of the original LiTS challenge reported scores of 0.962 for liver and 0.739 for liver tumor segmentation [IJK*21, BCL*23].

5. Discussion

The presented CBCTLiTS dataset is a unique dataset with paired CBCT and CT data. The dataset is also labelled, making it suitable

for segmentation. In its most basic form, the CBCT dataset can be used to train segmentation models to segment relatively simple, regular and large liver areas and complex, small and irregular liver tumor areas. However, the different CBCT volume qualities enable a range of further research and evaluation possibilities, from simulating low dose CBCT to facilitating multitask learning (multitask-b). By far the most significant investigated segmentation performance increase resulted from the paired CBCT and CT data, used in the different multimodal learning approaches. Using multimodal learning, assuming perfect CT and CBCT alignment, even the baseline CT model could be beat in several cases. Here, the differing CBCT qualities again allow for important insights, showing a negative correlation between reducing volume quality and improvements attributable to multimodal learning. Multimodal learning could further be enriched by preoperative masks, potentially further increasing segmentation accuracy. Finally, the pairing of CBCT and CT data allows for experimentation of image-to-image translation approaches. Here, further experimentation using e.g. diffusion models is enabled by CBCTLiTS. CBCTLiTS is a unique dataset useful for a multitude of research scenarios, especially for studies in the field of intraoperative medical imaging.

6. Conclusion

We have introduced CBCTLiTS, a novel synthetic, paired CBCT/CT dataset designed to facilitate research in medical image segmentation and style transfer. CBCTLiTS addresses the scarcity of high-quality, publicly available CBCT datasets with annotated segmentations by providing ground truth segmentation masks and paired CT data. This dataset is available in five different CBCT quality levels, ranging from visual quality comparable to the original CT to low visual quality with a significant amount of artifacts.

Additionally to presenting the dataset and the process of dataset generation, several possible research scenarios, like uni and multimodal segmentation are introduced and baseline values for segmentation model adaptations are given, including holistic as well as patched approaches, These adaptations include multitask learning, multimodal learning and style transfer. Additional experiments show that CBCTLiTS is an especially capable resource to investigate intraoperative models. To further evaluate CBCTLiTS, a highly interesting future research direction, would be to evaluate the presented results with real CBCT/CT paired datasets, like Pancreatic-CT-CBCT-SEG [HRC*21].

Acknowledgments

References

[ARB*22] ANTONELLI M., REINKE A., BAKAS S., FARAHANI K., KOPP-SCHNEIDER A., LANDMAN B. A., OTHERS, CARDOSO J.: The medical segmentation decathlon. *Nature communications 13*, 1 (2022), 4128. 2

[BCL*23] BILIC P., CHRIST P., LI H., VORONTSOV E., BEN-COHEN A., KAISSIS G., OTHERS, MENZE B.: The liver tumor segmentation benchmark (lits). *Medical Image Analysis 84* (Feb. 2023). doi:10.1016/j.media.2022.102680. 1, 2, 4

[ÇAL*16] ÇIÇEK Ö., ABDULKADIR A., LIENKAMP S. S., BROX T., RONNEBERGER O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19* (2016), pp. 424–432. 3

[FDK84] FELDKAMP L. A., DAVIS L. C., KRESS J. W.: Practical cone-beam algorithm. *Josa a 1*, 6 (1984), 612–619. 2

[HRC*21] HONG J., REYNGOLD M., CRANE C., CUARON J., HAJJ C., MANN J., ZINOVOY M., YORKE E., LOCASTRO E., APTE A., ET AL.: Breath-hold ct and cone-beam ct images with expert manual organ-at-risk segmentations from radiation treatments of locally advanced pancreatic cancer [data set]. the cancer imaging archive. *The Cancer Imaging Archive https://doi. org/10.7937/TCIA. ESHQ-4D90* (2021). 5

[IJK*21] ISENSEE F., JAEGER P. F., KOHL S. A., PETERSEN J., MAIER-HEIN K. H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods 18*, 2 (2021), 203–211. 4

[JSWM02] JAFFRAY D. A., SIEWERDSEN J. H., WONG J. W., MARTINEZ A. A.: Flat-panel cone-beam computed tomography for image-guided radiation therapy. *International Journal of Radiation Oncology* Biology* Physics 53*, 5 (2002), 1337–1349. 1

[KKSB22] KARIUS A., KAROLCZAK M., STRNAD V., BERT C.: Technical evaluation of the cone-beam computed tomography imaging performance of a novel, mobile, gantry-based x-ray system for brachytherapy. *Journal of Applied Clinical Medical Physics 23*, 2 (2022), e13501. 2

[KSB*22] KARIUS A., SZKITSAK J., BORONIKOLAS V., FIETKAU R., BERT C.: Quality assurance and long-term stability of a novel 3-in-1 x-ray system for brachytherapy. *Journal of Applied Clinical Medical Physics 23*, 9 (2022), e13727. 2

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (2015), pp. 234–241. 3

[RFF*23] ROEDER F., FASTNER G., FUSSL C., SEDLMAYER F., STANA M., BERCHTOLD J., OTHERS, GAISBERGER C.: First clinical application of image-guided intraoperative electron radiation therapy with real time intraoperative dose calculation in recurrent rectal cancer: technical procedure. *Radiation Oncology 18*, 1 (2023), 186. 1, 2

[RSC*06] RAFFERTY M. A., SIEWERDSEN J. H., CHAN Y., DALY M. J., MOSELEY D. J., JAFFRAY D. A., IRISH J. C.: Intraoperative cone-beam ct for guidance of temporal bone surgery. *Otolaryngology—Head and Neck Surgery 134*, 5 (2006), 801–808. 1

[Sid85] SIDDON R. L.: Fast calculation of the exact radiological path for a three-dimensional ct array. *Medical physics 12*, 2 (1985), 252–255. 2

[TCMSG24] TSCHUCHNIG M. E., COSTE-MARIN J., STEININGER P., GADERMAYR M.: Multi-task learning to improve semantic segmentation of cbct scans using image reconstruction. In *BVM Workshop* (2024), pp. 243–248. 3, 4

[TSG24] TSCHUCHNIG M. E., STEININGER P., GADERMAYR M.: Multimodal learning to improve segmentation with intraoperative cbct & preoperative ct. *arXiv preprint arXiv:2406.11650* (2024). 4

[WAR*24] WEI C., ALBRECHT J., RIT S., LAURENDEAU M., THUMMERER A., CORRADINI S., OTHERS, LANDRY G.: Reduction of cone-beam ct artifacts in a robotic cbct device using saddle trajectories with integrated infrared tracking. *Medical Physics* (2024). 1, 2