Anatomy-guided Pathology Segmentation

Alexander Jaus[1,2], Constantin Seibold[3,4], Simon Reiß[1], Lukas Heine[3], Anton Schily[3], Moon Kim[3], Fin Hendrik Bahnsen[3], Ken Herrmann[4,5], Rainer Stiefelhagen*[1], and Jens Kleesiek*[3]

[1] Karlsruhe Institute of Technology, Karlsruhe, Germany
{firstname.lastname}@kit.edu
[2] HIDSS4Health - Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany
[3] Institute for AI in Medicine, University Hospital Essen, Essen, Germany
{firstname.lastname}@uk-essen.de
[4] Department of Nuclear Medicine, University of Duisburg-Essen
{firstname.lastname}@uk-essen.de
[5] German Cancer Consortium (DKTK)- University Hospital Essen, Essen, Germany

Abstract. Pathological structures in medical images are typically deviations from the expected anatomy of a patient. While clinicians consider this interplay between anatomy and pathology, recent deep learning algorithms specialize in recognizing either one of the two, rarely considering the patient's body from such a joint perspective.
In this paper, we develop a generalist segmentation model that combines anatomical and pathological information, aiming to enhance the segmentation accuracy of pathological features. Our Anatomy-Pathology Exchange (APEx) training utilizes a query-based segmentation transformer which decodes a joint feature space into query-representations for human anatomy and interleaves them via a mixing strategy into the pathology-decoder for anatomy-informed pathology predictions.
In doing so, we are able to report the best results across the board on FDG-PET-CT and Chest X-Ray pathology segmentation tasks with a margin of up to 3.3% as compared to strong baseline methods. Code and models are available at github.com/alexanderjaus/APEx.

Keywords: Pathology Segmentation · Anatomical Guidance · PET-CT

1 Introduction

Throughout their extensive training, radiologists acquaint themselves with human biology and physiology, enabling them to discern typical patterns in the anatomy of both healthy individuals and those presenting health concerns. Years of clinical practice empower doctors to use this underlying knowledge about the body to associate very nuanced visual anatomy abnormalities with specific diseases correctly. This holistic approach of doctors, considering both anatomy and

* shared last author

pathology in the tissue is contrasted by the vast amount of current automatic pathology segmentation models that specialize in narrow disease types and fall short of an overall understanding of body structures [23,16]. These models are generally end-to-end semantic segmentation learners [22,21], and resemble models designed for the natural image domain and as such could be applied interchangeably in both domains, from pathology- to street-scene- [29] and everyday object segmentation [6,25]. Conversely, the medical imaging field has an obvious, yet often disregarded continuity which is – of course – the context is always the human body with the patient's anatomy.

While patients' anatomical features vary, the medical biases that associate anatomy with pathology for radiological assessment remain constant, such as simple observations, that a fracture has to be associated with a bone structure or that tumor locations often correspond to anatomical regions. When identifying a pathology, current segmentation models might or might not pick up anatomy-pathology correlations during training, which is the reverse direction to using anatomical priors for pathology identification. In the spirit of a doctor's workflow, we ask: Can explicitly learned human anatomy improve a model's capability to predict pathological structures?

Within this work, we explore different strategies to incorporate anatomical knowledge which we model as anatomical labels to improve upon pathology predictions. Inspired by the training of medical professionals, we propose a joint training procedure in which our network learns to predict both: anatomy and pathology via our proposed APEx architecture.

We summarize our contributions as follows: (1) We ablate multiple strategies on how to incorporate learned anatomical knowledge into pathology segmentation models. (2) We introduce a query-based, joint anatomy and pathology instance segmentation model APEx, which is capable of enriching pathology predictions by integrating anatomy knowledge via shared embeddings and query mixing. (3) We validate the performance of APEx on two different medical datasets covering whole-body FDG-PET-CT and chest X-Ray with as diverse anatomical structures as bones, organs and vessels for an improved joint anatomy-pathology recognition of +2.0%, +3.3% respectively.

Related Work: Leveraging anatomical knowledge as a prior has been addressed in some previous works in the form of shapes [27,20,18], expected textures [11] or atlas-based segmentation models [10] to aid segmentation. While these works mostly aim to segment anatomical structures, some works have started to leverage anatomical priors to improve upon pathology detection in X-Ray [17] or Colorectal Cancer segmentation in CT [28]. These studies highlight the beneficial effects of using anatomical priors but mainly utilize manually designed features tailored towards a specific use-case (e.g. the intestinal wall is important to detect colorectal cancer).

Within this work, instead of following the idea of manually designed features, we opt for a more data-driven and deep-learning-inspired approach: We investigate the usefulness of learned anatomical features to aid the segmenta-

tion of pathology. This approach breaks free of hand-crafted prior limitations and allows us to capture the knowledge available thanks to recently available holistic anatomical datasets [26,12,24]. We hypothesize that utilizing anatomical features helps identify pathologies as deviations from the expected anatomy.

2 Methodology

In this section, we first present the learning setup for anatomy and pathology segmentation and walk through our ablations to incorporate anatomical knowledge into the model training. Finally, we derive our so called **A**natomy-**P**athology **Ex**change (**APEx**) strategy to jointly learn both anatomy and pathology.

2.1 Preliminaries

Our formulation of the anatomy and pathology segmentation task depends on a training dataset:

$$\mathcal{D} = \{(x_i, a_i, p_i)\}_{i=0}^{N} \ , \tag{1}$$

with $x_i \in \mathbb{R}^{3\times H\times W}$ referring to one of the N images in the dataset, while $a_i \in [0,\dots,A]^{H\times W}$ is the associated anatomy with A classes and $p_i \in [0,\dots,P]^{H\times W}$ the pathology mask with P classes within the image. The task of a trained model is to predict, for new unseen test images x_t for each pixel in the image the correct anatomy categories a_t as well as the correct pathology classes p_t.

If the dataset provides instance-level annotations, we extend the approach to an instance-aware regime. Each anatomical mask a_i and pathological mask p_i then includes not only class- but instance-aware targets.

To investigate whether anatomical knowledge aids in identifying deviations from expected anatomy, we will examine two different tasks in two distinct domains: semantic segmentation of cancer in PET-CT images and instance-aware segmentation of thoracic abnormalities in chest X-Rays.

To accommodate these varied requirements, we opt for a 2D model due to the constraints of the X-Ray domain and model the 3D PET-CT images as sliced 2D images. To address the differing demands of semantic and instance-aware segmentation, we align with recent advancements in segmentation literature [5,14,2,30] which intertwine both semantic- and instance segmentation through the design choice of predicting high-dimensional query vectors, which combined with pixel-wise embeddings, encode instance-wise segments in an image. These queries are then employed to classify each segment, encapsulating information about both the segment's class and its shape. As a starting point for the experiments, we choose a Mask2Former [5] architecture. Our chosen setup is flexible in the choice of image modalities and in the choice of segmentation tasks.

2.2 Incorporating Learned Anatomical Knowlege: A roadmap

To investigate how to incorporate anatomical knowledge into the model training, we perform several ablations in a five-fold cross-validation setting in the domain of PET-CT. The baseline comparison model is a Mask2Former [5] model trained only on pathological labels. We report the 5-fold Validation IoU scores of naive anatomy incooperation techniques in Tab. 1 (left).

Table 1. Val scores on the 5-fold CV PET-CT splits. A. Cond, A. Pred and γ denote anatomy conditioning, auxillary anatomy learning and a weight factor respectively.

Naive Anatomy Incooperation				
Method	A. Cond	A. Pred	γ	IoU
Baseline	–	–	–	54.34 ± 1.46
Pretrain	✓	–	–	56.64 ± 3.06
Multitask	–	✓	1	56.10 ± 3.36
Multitask	–	✓	10	57.12 ± 4.17
Multitask	–	✓	142	55.89 ± 3.03
Ana In	✓	–	–	57.23 ± 2.71
Ana In	✓	✓	1	56.52 ± 4.14

Architecture Ablations	
Method	IoU
Baseline	54.34 ± 1.46
+Shared BB	54.44 ± 4.14
+Shared PD	58.69 ± 3.63
⌞Query Sum	59.56 ± 3.64
⌞Query Sum 2-ways	59.35 ± 3.18
⌞Query Mean	59.78 ± 3.23
⌞Cross Attention (CA)	59.42 ± 2.42
⌞CA per feature level	58.48 ± 2.52

First, we investigate the effect of pretraining on anatomy. This leads to an improvement of about 2.3%.

Multitask Prediction: Next, we compare to jointly learned features using a multi-task setting approach. We paste the pathological labels atop the anatomical labels predicting an additional class. Despite being suboptimal, since PET-CT pixels could be interpreted as both, anatomy and pathology, depending on the context, this leads to a similar improvement as pretraining. However, treating pathology as just another class underestimates its significance. To address this, we apply a weighted loss with weight γ, amplifying the pathology class's importance by 10-fold and 142-fold to equate it with the 142 anatomical labels. The 10-fold increase yields positive results, whereas the 142-fold adjustment demonstrates the challenge of selecting an appropriate weight factor.

Anatomy as an Auxiliary Input: Inspired by atlas-based segmentation methods, we input anatomical labels along with the PET-CT image mimicking an optimal anatomy atlas. Using this procedure, we receive similar results as with the previous approach.

Architecture Ablations: The last section's analysis underscores that while anatomical knowledge can enhance pathology prediction, its effective utilization is complex. Thus, in our second experimental series, we postulate that due to the inherent overlap between anatomical and pathological labels, a two-head prediction approach is optimal.

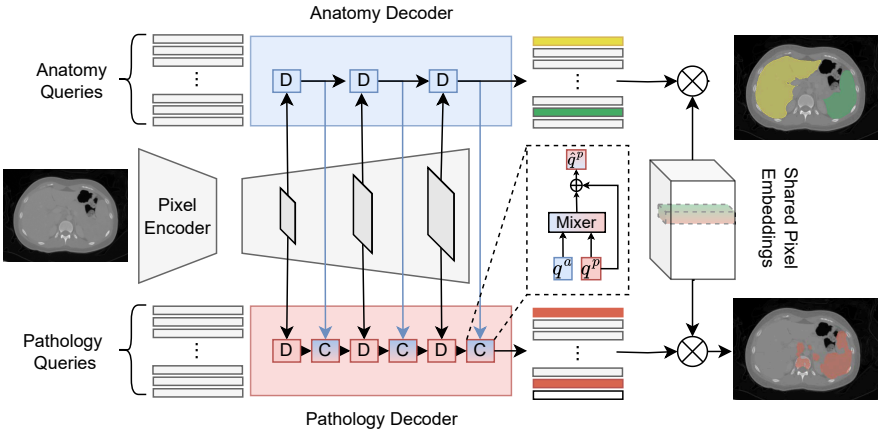

Fig. 1. Overview of the proposed APEx Method, leveraging a shared pixel encoder, shared pixel embedding space, separate decoders and a query-mixing module.

Initially, only the ResNet50 backbone is shared between the two prediction heads, resulting in no major improvement. A critical adjustment involves the sharing of a PixelDecoder across both anatomical and pathological prediction tasks. This integration significantly boosts the performance, evidenced by a notable increase of over 4% in IoU. This enhancement underscores the PixelDecoder's role in generating pixel embeddings rich in anatomical and pathological information, marking it as a crucial element in our design reflecting the dual role of each pixel in this task.

Query Mixing Strategies: Ultimately, as we employ distinct transformer decoders for anatomical and pathological predictions, we probe the efficacy of information exchange mechanisms via query exchange. This reflects the possibility of a direct exchange of queries representing anatomical and pathological segments. We explore various strategies, including nonparametric mixing and more flexible communication strategies such as cross-attention. While almost all strategies lead to a positive effect, none of them shines out as a clear winner.

We conclude this section with the insight that a two-head prediction one for the anatomy and one for pathologies leveraging shared pixel-embedding is a crucial design choice. On top, enabling communication between the different decoders leads to a further performance boost. The best-ablated model performs about 5.44% better than the naive baseline model.

2.3 Proposed Approach: APEx

APEx is based on a query-based segmentation approach leveraging anatomical and pathological information. It incorporates anatomical context via the exchange of information between two decoders: One tasked to segment the anatomy and one tasked to segment the pathology. We show the overall method in Fig. 1.

Shared Embedding Architecture: Starting with a standard 2D image $x \in \mathbb{R}^{3\times H\times W}$ we encode the image using a feature extractor f^{extr} (parameterized by a ResNet50 [9]), which maps x to a set of feature maps at different scales $f^{extr}(x) = \{F_i\}_{i=0}^{n}$ with $F_i \in \mathbb{R}^{H_i\times W_i}$, such that $H_i > H_{i+1}$ and $W_i > W_{i+1}$ hold, i.e., feature maps successively get smaller in spatial extent. These feature maps are then decoded using an arbitrary pixel-decoder. We choose to use the deformable DETR [30] model as a pixel-decoder producing a set of enriched pixel embeddings $\{J_i\}_{i=0}^{n}$, with $J_i \in \mathbb{R}^{d\times H_i\times W_i}$.

Anatomy and Pathology Decoders: Our architecture is motivated in computing separate query vectors for anatomy and pathology classes and let the anatomy queries influence the pathology queries while limiting the reverse influence only to a shared embedding space.

Each enriched pixel encoding map J_i is accessed by two decoding functions $f_i^{ana}(\cdot)$ and $f_i^{path}(\cdot)$ from the function sets $\{f_i^{ana}(\cdot)\}_{i=n}^{1}$ and $\{f_i^{path}(\cdot)\}_{i=n}^{1}$ which either decode the anatomy or the pathology from it.

Randomly initialized, but learnable parameter-queries q_0^{ana} and q_0^{path} are transformed via

$$q_{i+1}^{ana} = f_i^{ana}(q_i^{ana}, J_i) \text{ and} \tag{2}$$

$$q_{i+1}^{path} = f_i^{path}(q_i^{path}, J_i) \ , \tag{3}$$

and optimized during training. The decoders $f_i(\cdot)$ follow a standard masked transformer setup, i.e. queries are transformed through a cross-attention layer that attends to the joint embeddings of the respective scale i, followed by a self-attention and feed-forward layer. For the pathology branch $\{f_i^{path}(\cdot)\}_{i=n}^{1}$ to explicitly adhere to the learned anatomical queries an anatomy-to-pathology communication strategy is designed next.

Anatomy to Pathology Communication Strategy: Medical personnel have access to a large amount of knowledge regarding the human body, which current pathology segmentation models do not have. Besides the implicit information exchange via the shared pixel embedding, we propose to integrate a communication step $f_i^{mix}(\cdot)$ after each pathology-decoder step $f_i^{path}(\cdot)$. There the queries q_i^{path} resulting from the scale i pathology-decoder are enriched with the anatomy queries q_i^{ana} from the anatomy-decoder as follows:

$$\hat{q}_i^{path} = f_i^{mix}(q_i^{ana}, q_i^{path}) \tag{4}$$

Here, $\hat{q}_i^{path}$ is the anatomy-enriched pathology query which, through a mixing strategy is capable of capturing anatomical information. We did not find a superior mixing strategy and thus would either recommend averaging the queries as a nonparametric approach or using a cross-attention mixing module.

In this asymmetric architectural setup, anatomical information influences the pathology-specific queries while the anatomy branch stays agnostic to any

Table 2. Comparison of APEx against multiple SOTA methods in the PET-CT domain (left). We highlight the **best** and the second best performance.

Method	PET-CT VAL			PET-CT TEST		
	IoU	Dice	BIoU	IoU	Dice	BIoU
DLV3+[3]	55.00 ± 3.5	70.91 ± 3.0	54.78 ± 3.6	53.60 ± 5.4	69.65 ± 4.8	53.07 ± 5.4
M2F[5]	54.34 ± 1.4	70.41 ± 1.22	54.16 ± 1.6	55.48 ± 1.1	71.36 ± 1.0	55.02 ± 1.1
UNET[22]	57.62 ± 3.2	73.07 ± 2.6	57.38 ± 3.3	56.43 ± 1.5	72.14 ± 1.3	55.86 ± 1.4
Ours (CA)	$\mathbf{59.43 \pm 2.4}$	$\mathbf{74.52 \pm 1.9}$	$\mathbf{59.21 \pm 2.6}$	$\mathbf{57.5 \pm 0.9}$	$\mathbf{73.01 \pm 0.7}$	$\mathbf{57.04 \pm 0.9}$

pathology and simply reflects the patient-specific anatomy details serving as a useful foundational prior in pathology assessment. This design is ablated against an inferior design in which the anatomy branch is updated by the pathology as well (cf. Tab. 1: Query Sum 2-ways).

Joint Anatomy and Pathology Segmentation: Bringing the whole architecture and processing steps together into our Anatomy and Pathology Exchange (APEx) pipeline, we predict the anatomy and pathology segments through the following dot product:

$$out^{ana} = J_0 \cdot q_{n-1}^{ana} \text{ and} \tag{5}$$

$$out^{path} = J_0 \cdot \hat{q}_{n-1}^{path} \ , \tag{6}$$

Query vectors are passed through a simple classifier to associate anatomy or pathology classes to the predicted segments. The parameters of all components, namely f^{extr}, f^{ana}, f^{path}, and f^{mix} are optimized via weighted cross-entropy and binary mask losses enforced on each anatomy and pathology prediction out^{ana} and out^{path}.

3 Experiments and Results

Datasets: To assess our method's broad applicability, we performed experiments across two vastly different medical imaging domains: FDG-PET-CT, and

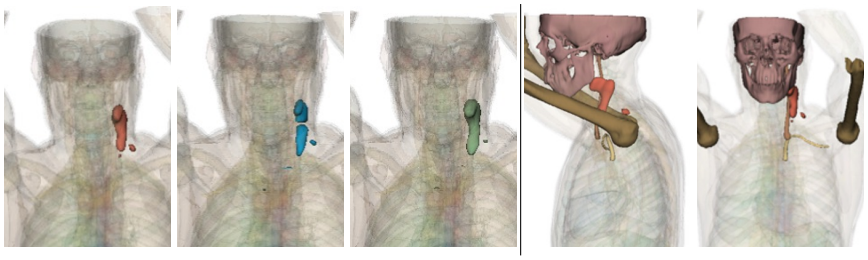

Fig. 2. Stacked 2D tumor predictions next to top-5 attended anatomical structures.

Chest X-Ray. For FDG-PET-CT, due to the absence of a comprehensive dataset with both anatomical and lesion annotations, we merged two distinct datasets: autoPET [7], which provides lesion annotations, and the Atlas dataset [12], offering anatomical details. We exclude patients without pathologies, motivated by the high accuracy ($\geq 95\%$) of binary classifiers for cancer detection in PET images. Our study utilized 185 3D volumes for five-fold cross-validation and an additional test set of 125 cancer patients from the remaining dataset. To adapt images to the selected 2D setting, we slice them axially and stack CT and PET images channel-wise, leaving the third channel empty.

In the X-Ray domain, we evaluate the properties of our method on the ChestXDet[15] dataset containing 13 pathology classes. To train anatomy segmentation, we predict anatomy pseudo-labels onto this dataset using a model trained on the PaxRay++ dataset [24]. We evaluate the different methods using five-fold cross-validation on the training set. During training, we omit images with no pathologies.

Baselines and Methods: When evaluating models across different domains, we determine the best performing candidate based on the performance on the individual validation sets. We use either the official test splits, if they exist, or a test set that we reserved beforehand. We benchmark APEx on PET-CT against established 2D segmentation baselines such as UNet [22], DeeplabV3+ [3] and Mask2Former [5]. In all experiments, we ensure models are trained using identical data and learning pipelines to isolate the effect of incorporating anatomical knowledge. For chest X-Ray, we compare on instance segmentation against PointRend [13], MaskDino [14] and Mask2Former[5]. Regarding the specific APEx architecture, we choose the Cross-Attention Query Mixer, as it offered a competitive performance with the lowest standard deviation during our initial ablations (cf. Tab. 1).

3.1 Semantic- and Instance Segmentation Results

PET-CT Results: In Tab. 2 we report the Dice, IoU and Boundary IoU [4] (BIoU) performances of the previously mentioned baseline segmentation models against our method. All models have been initialized with LVM-MED weights [19] to provide a fair comparison. The results indicate that our method is capable of outperforming multiple strong competitors on our five-fold validation splits and the holdout testset. Fig. 2 shows qualitative results as well as the most attended anatomical structures during the cross-attention query mixing step.

ChestXDet Results: In Tab. 3.1, we show the performance of different state-of-the-art instance segmentation methods trained using the same backbone. We see that our method improves over the Mask2Former-baseline by $\sim$3.75% mAP. Across 12 of 13 pathologies, our method achieves the best, or second-best performance, improving over recent transformer architectures as well as established CNN models. Detailed results and qualitative examples are in the Appendix.

Table 3. ChestXDet[15] results. We highlight the best performance in **bold** and the second best by underlining. Detailed results in Appendix.

Pathology	MRCNN[8]	CascCRCNN [1]	PointRend[13]	MskDino[14]	M2F[5]	Ours (CA)
mAP (Val)	13.98 ± 0.40	14.64 ± 0.44	15.33 ± 0.82	16.44 ± 1.43	16.57 ± 0.67	$\mathbf{17.16 \pm 0.63}$
mAP (Test)	13.72 ± 0.41	13.86 ± 0.70	15.14 ± 0.44	14.38 ± 0.74	13.87 ± 0.53	$\mathbf{17.20 \pm 0.33}$

4 Conclusion

We proposed a novel way of leveraging anatomical information to improve pathology segmentation and showed the efficacy of the general concept of anatomy-guidance in two different domains covering diverse anatomical structures and pathologies. Besides improved performance, our method APEx encourages the exchange of anatomical information to ensure pathology segments are informed by the patient's anatomy, aligning more with the workflow of doctors that developed over decades.

Acknowledgments. The present contribution is supported by the Helmholtz Association under the joint research school "HIDSS4Health – Helmholtz Information and Data Science School for Health and was supported by funding from the pilot program Core-Informatics of the Helmholtz Association (HGF). This work was performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6154–6162 (2018)
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020)
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
4. Cheng, B., Girshick, R., Dollár, P., Berg, A.C., Kirillov, A.: Boundary iou: Improving object-centric image segmentation evaluation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15334–15342 (2021)
5. Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation (2022)
6. Gamal, M., Siam, M., Abdel-Razek, M.: Shuffleseg: Real-time semantic segmentation network. arXiv preprint arXiv:1803.03816 (2018)

7. Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., Rubin, D.: A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. Scientific Data **9**(1), 601 (2022)
8. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Huang, H., Zheng, H., Lin, L., Cai, M., Hu, H., Zhang, Q., Chen, Q., Iwamoto, Y., Han, X., Chen, Y.W., et al.: Medical image segmentation with deep atlas prior. IEEE Transactions on Medical Imaging **40**(12), 3519–3530 (2021)
11. Ibragimov, B., Toesca, D., Chang, D., Koong, A., Xing, L.: Combining deep learning with anatomical analysis for segmentation of the portal vein for liver sbrt planning. Physics in Medicine & Biology **62**(23), 8943 (2017)
12. Jaus, A., Seibold, C., Hermann, K., Walter, A., Giske, K., Haubold, J., Kleesiek, J., Stiefelhagen, R.: Towards unifying anatomy segmentation: automated generation of a full-body ct dataset via knowledge aggregation and anatomical guidelines. arXiv preprint arXiv:2307.13375 (2023)
13. Kirillov, A., Wu, Y., He, K., Girshick, R.: Pointrend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9799–9808 (2020)
14. Li, F., Zhang, H., Xu, H., Liu, S., Zhang, L., Ni, L.M., Shum, H.Y.: Mask dino: Towards a unified transformer-based framework for object detection and segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3041–3050 (2023)
15. Liu, J., Lian, J., Yu, Y.: Chestx-det10: Chest x-ray dataset on detection of thoracic abnormalities (2020)
16. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
17. Müller, P., Meissen, F., Brandt, J., Kaissis, G., Rueckert, D.: Anatomy-driven pathology detection on chest x-rays. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 57–66. Springer (2023)
18. Navarro, F., Shit, S., Ezhov, I., Paetzold, J., Gafita, A., Peeken, J.C., Combs, S.E., Menze, B.H.: Shape-aware complementary-task learning for multi-organ segmentation. In: Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10. pp. 620–627. Springer (2019)
19. Nguyen, D.M., Nguyen, H., Diep, N.T., Pham, T.N., Cao, T., Nguyen, B.T., Swoboda, P., Ho, N., Albarqouni, S., Xie, P., et al.: Lvm-med: Learning large-scale self-supervised vision models for medical imaging via second-order graph matching. arXiv preprint arXiv:2306.11925 (2023)
20. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., De Marvao, A., Dawes, T., O'Regan, D.P., et al.: Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. IEEE transactions on medical imaging **37**(2), 384–395 (2017)
21. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)

22. Ronneberger, O., P.Fischer, Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241. Springer (2015)
23. Schultheiss, M., Schmette, P., Bodden, J., Aichele, J., Müller-Leisse, C., Gassert, F.G., Gassert, F.T., Gawlitza, J.F., Hofmann, F.C., Sasse, D., et al.: Lung nodule detection in chest x-rays using synthetic ground-truth data comparing cnn-based diagnosis to human performance. Scientific Reports **11**(1), 15857 (2021)
24. Seibold, C., Jaus, A., Fink, M.A., Kim, M., Reiß, S., Herrmann, K., Kleesiek, J., Stiefelhagen, R.: Accurate fine-grained segmentation of human anatomy in radiographs via volumetric pseudo-labeling. arXiv preprint arXiv:2306.03934 (2023)
25. Wang, W., Neumann, U.: Depth-aware cnn for rgb-d segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 135–150 (2018)
26. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5) (2023)
27. Yao, J., Cai, J., Yang, D., Xu, D., Huang, J.: Integrating 3d geometry of organ for improving medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22. pp. 318–326 (2019)
28. Zhang, R., Bai, Z., Yu, R., Pang, W., Wang, L., Zhu, L., Zhang, X., Zhang, H., Hu, W.: Ag-crc: Anatomy-guided colorectal cancer segmentation in ct with imperfect anatomical knowledge. arXiv preprint arXiv:2310.04677 (2023)
29. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters **15**(5), 749–753 (2018)
30. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159 (2020)