CTSpine1K: A Large-Scale Dataset for Spinal Vertebrae Segmentation in Computed Tomography

Yang Deng [1,2,3], Ce Wang [1,2], Yuan Hui [1,2], Qian Li [1,2], Jun Li [1], Shiwei Luo [4], Mengke Sun [1], Quan Quan [1], Shuxin Yang [1], You Hao [1,2], Pengbo Liu [1], Honghu Xiao [5], Chunpeng Zhao [5], Xinbao Wu [5], S. Kevin Zhou [1,2,3]

1 Key Lab of Intelligent Information Processing of Chinese Academy of Sciences, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China
2 Suzhou Institute of Intelligent Computing Technology, Chinese Academy of Sciences, Suzhou, 215028, China
3 School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, and also with the Center for Medical Imaging, Robotics, Analytic Computing & Learning (MIRACLE), Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, Jiangsu, China
4 Department of Radiology, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, 510006, China
5 Beijing Jishuitan Hospital, Beijing, 100035, China

Abstract

Spine-related diseases have high morbidity and cause a huge burden of social cost. Spine imaging is an essential tool for noninvasively visualizing and assessing spinal pathology. Segmenting vertebrae in computed tomography (CT) images has always been the base of quantitative medical image analysis for clinical diagnosis and surgery planning of spine diseases. Current publicly available annotated datasets on spinal vertebrae are small in size. Due to the lack of a large-scale annotated spine image dataset, the mainstream deep learning-based segmentation methods, which are data-driven, are heavily restricted. In this paper, we introduce a large-scale spine CT dataset called CTSpine1K, curated from multiple sources for vertebra segmentation, which contains 1,005 CT volumes with over 500,000 labeled vertebrae slices and 11,172 vertebrae belonging to different spinal conditions. Based on this dataset, we conducted several spinal vertebrae segmentation experiments to set the first benchmark. We believe that this large-scale dataset will facilitate further research in many spine-related image analysis tasks, including but not limited to vertebrae segmentation, labeling, 3D spine reconstruction from biplanar radiographs, and image superresolution and enhancement. Our dataset are publically available at `https://xnat.health-ri.nl/data/archive/projects/africai_miccai2024_ctspine1k` and `https://github.com/MIRACLE-Center/CTSpine1K`.

Keywords

Spine Dataset, Vertebrae Segmentation, Computed Tomography (CT), Medical Imaging

1. Introduction

The *spine* is an important part of the musculoskeletal system, sustaining body mobility and protecting the spinal cord, the most important neural pathway in the body (Cai et al., 2020; Sekuboyina et al., 2020). From top to bottom along the trunk of the human body, the spine consists of 7 cervical vertebrae (C1-C7), 12 thoracic vertebrae (T1-T12), 5 lumbar vertebrae (L1-L5), 1 sacral vertebra, and 1 caudal vertebra. Note that people are containing L6 (resulting from sacral lumbarization) or losing L5 (resulting from lumbar sacralization) with a rare occurrence in a population.

Each vertebra is at risk of disease due to bearing the

A. Cervical and thoracic vertebrae

B. Thoracic and lumbar vertebrae

C. Metal artifact

D. Lumbar sacralization (red arrow)

E. Sacral lumbarization (yellow arrow)

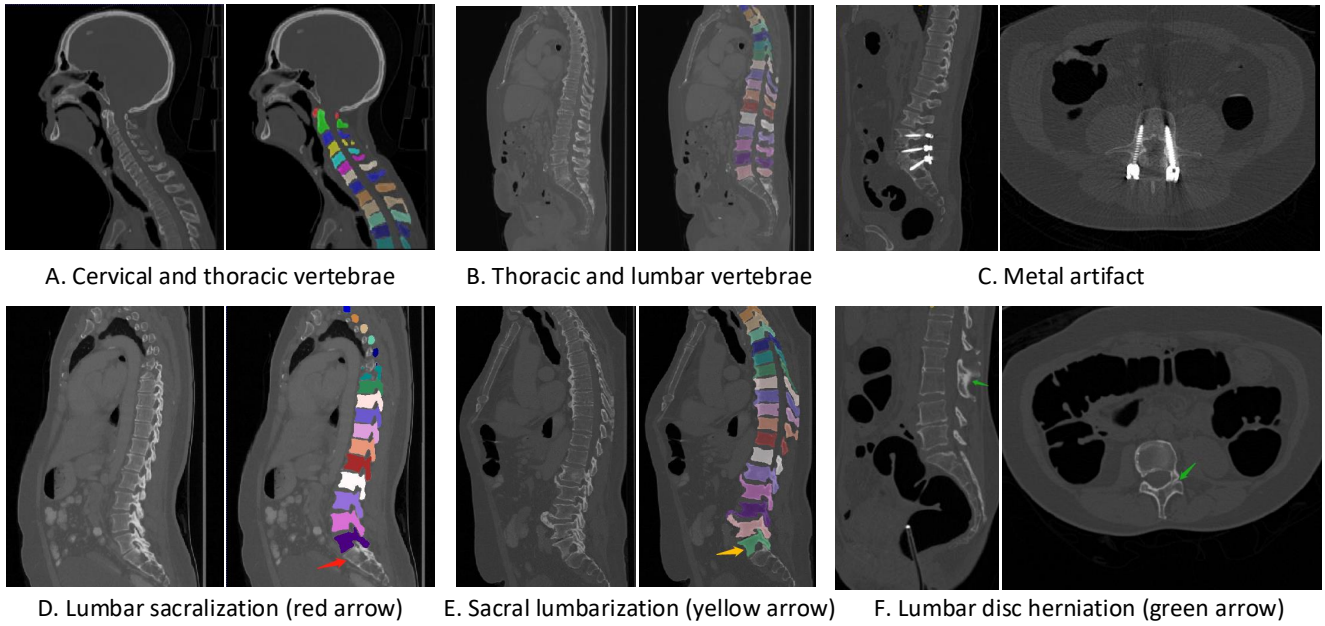F. Lumbar disc herniation (green arrow)

Figure 1: Spine CT image examples with various conditions.

load of the human body (Faruqi et al., 2018). Spine-related diseases, including degenerative changes, spinal inflammation, spinal tumors, spinal tuberculosis, and spinal infection, have a high morbidity and cause a huge burden of social cost. In clinical practice, common spinal diseases include mainly degenerative diseases of the spine and disc herniation (Balériaux and Neugroschl, 2004). Degenerative diseases of the spine common in the elderly include cervical spondylosis, lumbar spondylosis, and thoracic spinal stenosis. Disc herniation includes cervical disc herniation that involves the cervical spine, thoracic disc herniation that involves the thoracic spine, and common lumbar disc herniation, among which lumbar disc herniation is the most common. Lumbar degenerative diseases are also the most common in the clinic, with lumbar spinal stenosis, lumbar spondylolisthesis, and lumbar instability. Treatment varies with the disease entity, and the clinical scenario can be nonspecific (Li et al., 2015).

Spinal imaging via computed tomography (CT), magnetic resonance imaging (MRI), radiography, ultrasound, positron emission tomography (PET), and other radiologic imaging modalities is essential for noninvasively visualizing and assessing spinal pathology. Computational methods support and enhance a physician's ability to utilize these imaging techniques for diagnosis, noninvasive treatment, and intervention in clinical practice. Analysis algorithms developed in the field of computer vision, computer graphics, signal processing, and machine learning have been adapted to analyze spinal images (Li et al., 2015). Conventionally, CT is preferred to study the spine

due to a high bone-soft tissue contrast. There are diverse image appearance variations due to differences in vertebral position, metal artifacts and spinal diseases, etc., challenging the analysis algorithms. Fig. 1 gives some examples of these various conditions.

Spinal or vertebral image segmentation in CT is crucial in all applications regarding automated quantification of spinal morphology and pathology. Over the recent years, deep learning has achieved remarkable success in various medical imaging applications (Zhou et al., 2021, 2019) and many automated spine image segmentation approaches have been proposed (Lessmann et al., 2019; Payer et al., 2020). However, all these approaches are data-dependant and have either been validated on private datasets or small public datasets. Considering SpineWeb [1], a popular archive for multi-modal spine data, it lists only two CT datasets: CSI2014 (Yao et al., 2012) and xVertSeg (Korez et al., 2015), both of which only contain dozens of CT scans. Therefore, those approaches are heavily restricted. To address the concerns of large-scale data availability, Sekuboyina et al. (Sekuboyina et al., 2020) organized the Large Scale Vertebrae Segmentation Benchmark(VerSe) as a challenge in conjunction with the International Conference on Medical Image Computing and Computed Assisted Intervention (MICCAI) 2019 and MICCAI 2020. With VerSe'19 [2], they released into the public domain a diverse dataset of 160 spine multi-detector CT scans with 1,735 vertebrae

---

1. http://spineweb.digitalimaginggroup.ca/
2. https://verse2019.grand-challenge.org/

(120 seen CT scans and 40 hidden CT scans). For VerSe'20 [3], the upgraded version of VerSe'19, they released 300 CT scans, the largest public spine CT dataset to date. These two datasets provide the ground truth of each vertebra and are currently the most used dataset for vertebrae segmentation. Nonetheless, these datasets are still small. Further, most CT scans from VerSe'19 and VerSe'20 datasets are 'cropped', only containing a small area of the spine and abandoning other information about surrounding organs.

To advance the research in spinal image analysis, we hereby present a large-scale and comprehensive dataset: CTSpine1K. We collect and annotate a large-scale spinal vertebrae CT dataset from multiple domains and different manufacturers, totaling 1,005 CT volumes (over 500,000 labeled slices and over 11,000 vertebrae) of diverse appearance variations. We carefully design an exquisite and unified annotation pipeline to ensure the quality of annotations. To the best of our knowledge, our CTSpine1K dataset is the largest publicly available annotated spine CT dataset. We evaluate the dataset's quality by conducting benchmark experiments for vertebrae segmentation.

2. The CTSpine1K Dataset

2.1 Data Collection

To build a comprehensive spine image dataset that replicates practical appearance variations, we curate a large-scale CT dataset of spinal vertebrae from the following four open sources.

COLONOG. This sub-dataset comes from COLONOGRAPHY dataset related to a CT colonography trial (Johnson et al., 2008). We randomly select one of the two positions (we will open source the codes used for selection), which have similar information for each patient in our dataset. There are 825 CT scans, and they are in a Digital Imaging and Communication in Medicine (DICOM) format.

HNSCC-3DCT-RT. This sub-dataset contains three-dimensional (3D) high-resolution fan-beam CT scans collected during pre-, mid-, and post-treatment using a Siemens 16-slice CT scanner with a standard clinical protocol for 31 head-and-neck squamous cell carcinoma (HNSCC) patients (Bejarano et al., 2008). These images are in a DICOM format.

MSD T10. This sub-dataset comes from the 10th Medical Segmentation Decathlon Simpson et al. (2019). To attain more slices containing the spine, we select task03_liver dataset consisting of 201 cases. These images are in a Neuroimaging Informatics Technology Initiative (NIfTI) format (https://nifti.nimh.nih.gov/nifti-1).

COVID-19. This sub-dataset consists of non-enhanced chest CTs from 632 patients with COVID-19 infections. The images were acquired at the point of care in an outbreak setting from patients with Reverse Transcription Polymerase Chain Reaction (RT-PCR) confirmation for the presence of SARS-CoV-2 (Harmon et al., 2020). We pick 40 scans with the images stored in a NIfTI format.

We reformat all DICOM images to NIfTI to simplify data processing and de-identify images, meeting contributing sites' institutional review board (IRB) policies. More details for those sub-datasets can be found in (Johnson et al., 2008; Bejarano et al., 2008; Simpson et al., 2019; Harmon et al., 2020). All existing sub-datasets are under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0); we will keep the license unchanged. It should be noted that for sub-dataset task03_liver and sub-dataset COVID-19, we only choose a part of cases from them, and in all these data sources, we exclude those cases of very low quality. The overview of our dataset and the thorough comparison with the VerSe Challenge dataset can be seen in Table 1.

2.2 Data Annotation

We design a unified and rigorous labeling standard and pipeline before annotation because medical image annotation is a highly time-consuming and subjective task. The annotation pipeline is shown in Fig. 2.

To reduce the annotation workload, we first use the public dataset from VerSe'19 and VerSe'20 Challenges to train a segmentation network using the nnUnet algorithm (Isensee et al., 2021). As mentioned earlier, most of the VerSe Challenge samples are 'cropped', abandoning the information about surrounding structures such as organs. Therefore, we selected cases with complete CT images (without cropping) and the same image spacing between images and their corresponding ground truth, totaling 41 cases that could be used. Then, for an image to be annotated, we invoke the trained segmentation model to predict segmentation masks and invite some junior annotators to refine the labels based on the prediction results. All these refined labels by junior annotators are checked by two senior annotators for further refinement. If the senior annotator finds it challenging to determine the annotations, these data will be sent to one of the trained spine surgeons, whose image-reading experience averages 12 years. Finally, all of these annotated labels undergo a random double-check by coordinators to ensure the final quality of annotations.

If there exist any wrong cases in double-check, they are corrected by annotators. The human-corrected annotations and their corresponding images are then added to

---

3. https://verse2020.grand-challenge.org/

1. https://verse2019.grand-challenge.org/
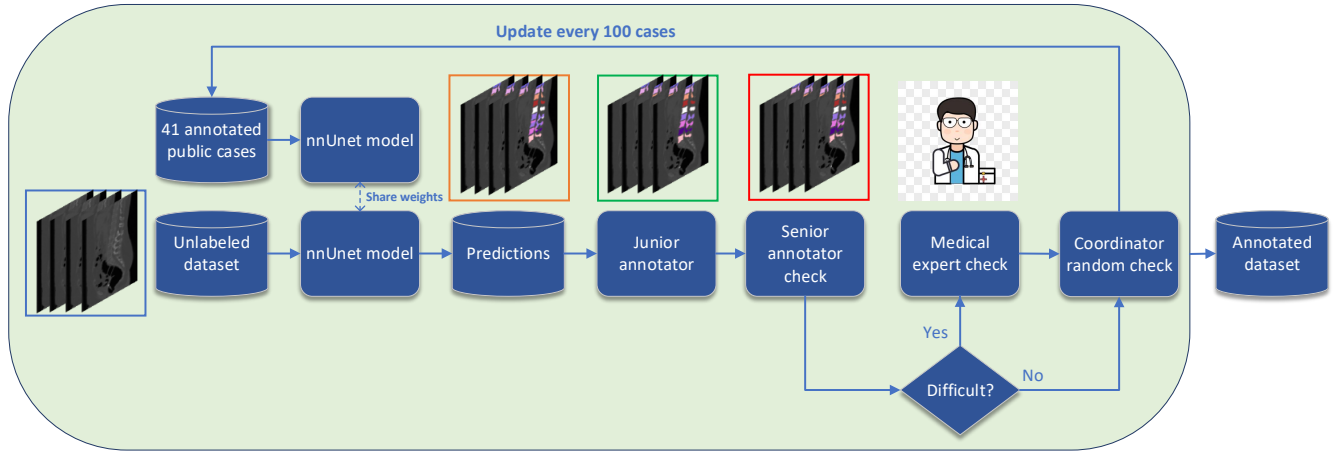2. https://verse2020.grand-challenge.org/

Figure 2: The proposed annotation pipeline.

Table 1: Overview of our large-scale CTSpine1K dataset. Ticks[✓] in the table refer to the fact that this dataset contains some special cases with sacral lumbarization or lumbar sacralization, and we list their IDs in the open source. 'Tr\Ts_pu\Ts_pr' donates training\test_public\test_private. We exclude the metal-artifact cases due to the difficulty of labeling them.

. Dataset name	VerSe'19,20 [✓]	COLONOG [✓]	HNSCC-3DCT-RT	MSD T10 [✓]	COVID-19	CTSpine1K [✓]
# of patients(n)	41	784	31	150	40	1005
# of labeled vertebrae(n)	511	8022	443	2105	602	11172
# of Tr\Ts_pu\Ts_pr(n)	-	480\152\152	20\5\6	90\30\30	20\10\10	610\197\198
# of cervical vertebrae(n)	17	4	217	132	66	419
# of thoracic vertebrae(n)	285	4039	226	1198	479	5942
# of lumbar vertebrae(n)	194	3911	0	744	57	4712
# of sacral lumbarization	8	16	0	1	0	17
# of lumbar sacralization	3	9	0	2	0	11
Mean spacing(mm)	(0.81±.1, 0.81±.1, 0.88±.16)	(0.75±.08, 0.75±.08, 0.81±.16)	(1.09±.14, 1.09±.14, 2.0)	(0.77±.1, 0.77±.1, 1.55±1.22)	(0.79±.09, 0.79±.09, 4.5±1.34)	(0.76±.11, 0.76±.11, 1.10±.94)
Mean size	(512, 512, 640±197)	(512, 512, 542±58)	(512, 512, 202±19)	(512, 512, 478±264)	(512, 512, 93±76)	(512,512,504±155)
Age	56.2 ± 17.6	56.7 ± 6.1 years	64.3 ± 12.8 years	Unknown	Unknown	-
Manufacturer	Siemens,GE,Philips,Toshiba	Siemens,GE,Philips,Toshiba	Siemens	Unknown	Unknown	-
Source and Year	URL[1,2] 2019,2020	(Johnson et al., 2008)	(Bejarano et al., 2008)	(Simpson et al., 2019)	(Harmon et al., 2020)	-

the training data to retrain a more powerful model. To speed up the annotation process, we update the database every 100 cases and retrain the deep learning model. The process is iterative until the annotation task is finished. The whole annotation process is operated with ITK-SNAP (Yushkevich et al., 2006a) software. Segmentation masks are also saved in a NIfTI format. In total, we have annotations for 1005 CT volumes.

2.3 Usage Notes

The data within this work is licensed under the CC BY-NC-SA 4.0 International License. The data can be freely downloaded and used for your own research purposes, but we kindly ask investigators to cite this paper in their publications. The data are suitable for visualization in a variety of software, including 3D Slicer (Yushkevich et al., 2006b) and ITK-SNAP (Yushkevich et al., 2006a).

2.4 Code availability

We provide the Python scripts and our trained model on GitHub [1], which can serve as a starting point for the community to march on future development based on our CTSpine1K dataset. All the datasets were uploaded to the XNAT [2] platform following MICCAI 2024 AFRICAI Imaging Repository White Paper (Starmans and Tsirikoglou, 2024).

3. Benchmarking Experiment

Based on CTSpine1K, we use a fully supervised method to train a deep network for spinal vertebrae segmentation to establish a benchmark. In recent years, the nnUnet model has achieved better results than other methods in many medical image segmentation tasks and has become the acknowledged baseline in medical image segmentation Isensee et al. (2021). nnUnet is essentially a U-Net Ronneberger

---

1. https://github.com/MIRACLE-Center/CTSpine1K
2. https://xnat.bmia.nl/data/archive/projects/africai_miccai2024_ctspine1k

et al. (2015), but with specific network architecture, design parameters, and training parameters self-adapted to the dataset's characteristics, together with powerful data augmentation. Therefore, we choose nnUnet as the benchmarking model for vertebrae segmentation. Due to our dataset's huge amount of high-resolution 3D images, we use the 3D full-resolution U-net architecture. More details for the nnUnet model can be seen in Isensee et al. (2021).

3.1 Experimental Setup

Data Split. Our dataset contains 1005 3D CT volumes (on average, each scan has 504 slices and 11 labeled vertebrae) with over 500,000 labeled vertebrae slices (size of 512x512). The CTSpine1K dataset is separated into a training dataset (610 subjects), a public test dataset (197 subjects), and a private test dataset (198 subjects). More details about the data split can be seen in Table 1. Aiming to compare the domain differences between our annotated dataset and the VerSe Challenge datasets, we use the 41 public cases from the VerSe Challenges as Test_VerSe.

Evaluation Metrics. Every vertebra (from C1 to L6) is labeled with integer values from 1 to 25. Here, we use two ubiquitous metrics prevalent in the medical image segmentation domain: (i) Dice Coefficient (DSC). DSC is computed per label: $2|A*B|/(|A|+|B|)$, where A is the set of foreground voxels of a certain label, and B is the predicted mask. (ii) 95th percentile Hausdorff distance (HD95) in mm. HD measures the local maximum distance between the two surfaces constructed from the ground truth and predicted segmentation map at the 95th percentile.

Implementation Details. We train the nnUnet model for 1,000 epochs, keeping the training configuration, such as the learning rate and data augmentation, etc., the same as the original settings in reference (Isensee et al., 2021). Due to the limit of computing sources, we use *all* folds rather than run five-fold cross-validation while training. The experiments are implemented using Pytorch on an RTX 3090 GPU. The training time for the last round is around 15 days.

3.2 Results

3.2.1 Quantitative Results.

We calculate the two metrics of each vertebra, and the results are reported in Table 2. On the one hand, our experimental results are close to those reported in reference (Sekuboyina et al., 2020) with the same model (nnUnet), verifying the high quality of our annotations. On the other hand, Table 2 shows it is difficult to segment the diseased vertebrae (the DSC of L6 is almost 0). Specifically, the existence of L6 confuses the model, resulting in prediction dislocations (see the last row in Fig. 4). Thus, our labeled

dataset, which contains many L6 cases, is very valuable for the diseased vertebrae segmentation (we have stated those cases that are hard for annotation in the readme.txt file). Table 2 illustrates that the model trained with our annotations can achieve good performance on our CTSpine1K dataset but a much worse performance on the VerSe Challenge datasets, which explains there is an obvious domain gap between our annotated dataset and the public dataset. We infer the reason is that the COlONOG dataset is based on an empty stomach and colon, confusing the deep learning model by the changes of air content in the abdomen (see Fig. 3). Therefore, our annotations are a good complement to the existing datasets.

3.2.2 Qualitative results.

Some visualization results are presented in Fig. 4, where we can observe that the baseline model can achieve excellent segmentation results. Nevertheless, some failed predictions occur when spinal diseases exist, especially sacral lumbarization and lumbar sacralization. Besides, the image's resolution of Z direction is closely related to the results, and a lower resolution leads to worse results. Maintaining a reasonable performance for a low resolution is a research challenge. Image superresolution (Peng et al., 2020) might be worth exploring.

4. Conclusion

We collect and annotate a large-scale spine CT dataset, including 1,005 CT scans with over 11,000 vertebrae. Furthermore, we validate our dataset using several benchmarking segmentation experiments. We will have more experiments in the future. We believe this work will help stimulate further research on spine-related conditions, including vertebrae segmentation, labeling, and 3D spine reconstruction from biplanar radiographs.

Author contributions statement

Y.D., C.W., and S.K.Z wrote the manuscript. The data was annotated by Y.D., C.W., Y.H., Q.L., J.L., S.W.L., M.K.S., Q.Q., S.X.Y., P.B.L, and Y.H.. Y.D. developed the algorithm. H.F.X., C.P.Z., and X.B.W. provided medical guidance to the annotators and reviewed the annotations to ensure their accuracy. All authors reviewed the manuscript. S.K.Z is responsible for the design of the entire work.

Acknowledgments

Table 2: The DSC and HD95 (in mm) results. The Mean of Vertebrae indicates the average of each vertebra (multi-class), and the Mean of Spine indicates the average of the whole spine (binary classification).

Label	Test_public DSC/HD95	Test_private DSC/HD95	Test_VerSe DSC/HD95	Label	Test_public DSC/HD95	Test_private DSC/HD95	Test_VerSe DSC/HD95
C1	.951/1.55	.957/1.34	.446/1.69	T7	.881/4.43	.764/5.98	.400/8.67
C2	.974/1.15	.975/1.22	.923/1.20	T8	.856/6.96	.772/6.23	.395/8.74
C3	.934/8.73	.945/7.55	.605/1.32	T9	.949/1.76	.925/3.18	.597/6.33
C4	.946/1.90	.608/3.65	.613/0.95	T10	.952/1.59	.951/1.85	.813/3.83
C5	.573/2.43	.726/3.13	.231/0.95	T11	.959/1.49	.961/1.81	.887/3.36
C6	.802/4.51	.824/2.38	.121/1.07	T12	.931/1.49	.965/1.73	.889/1.00
C7	.930/2.18	.899/2.16	.228/4.04	L1	.965/1.27	.883/1.76	.925/2.34
T1	.952/1.60	.927/2.36	.623/2.53	L2	.965/1.50	.913/1.91	.902/4.83
T2	.960/1.43	.932/2.59	.883/2.36	L3	.963/1.61	.952/2.05	.830/7.17
T3	.955/1.64	.798/7.47	.817/4.14	L4	.965/2.32	.945/2.61	.795/8.94
T4	.899/3.83	.831/3.64	.723/6.44	L5	.969/1.49	.943/2.45	.723/10.31
T5	.681/6.67	.830/4.90	.614/9.05	L6	0/66.21	0/80.33	0/45.16
T6	.814/5.20	.762/5.99	.483/9.21				
Mean of Vertebrae	.869/5.40	.840/6.41	.619/6.23	Mean of Spine	.985/1	.984/1	.929/3.08

| VerSe' | COLONOG | VerSe' | COLONOG | VerSe' | COLONOG |

Figure 3: The difference between the COLONGO dataset and VerSe'dataset.

Ethical Standards

This work was conducted in accordance with the Declaration of Helsinki. The CTSpine1K dataset presented in this study is a derivative data resource constructed by curating and standardizing data from four publicly available datasets: COLONOG, HNSCC-3DCT-RT, MSD T10, and COVID-19. We confirm that the original studies involving human participants for these constituent datasets were approved by their respective Institutional Review Boards (IRB) or Ethics Committees. Specifically: The COLONOG and HNSCC-3DCT-RT datasets were obtained from The Cancer Imaging Archive (TCIA), where all data are de-identified and collected with necessary ethical approvals and informed consent (or waiver of consent) as detailed in their original publications (Johnson et al., 2008; Bejarano et al., 2008). The MSD T10 and COVID-19 datasets were accessed via their respective public challenges or open repositories, consistent with the ethical standards described by their original curators (Simpson et al., 2019; Harmon et al., 2020). All data utilized in the creation of CTSpine1K were fully anonymized and de-identified prior to our access.

We strictly complied with the original data usage licenses and terms of use for each source. No new human subjects were recruited for this study, and no attempt was made to re-identify any individuals.

Conflicts of Interest

The authors have no conflicts of interest.

Data availability

The dataset is publicly available at `https://xnat.health-ri.nl/data/archive/projects/africai_miccai2024_ctspine1k` and `https://github.com/MIRACLE-Center/CTSpine1K`.

References

Danielle L Balériaux and Carine Neugroschl. Spinal and spinal cord infection. *European Radiology Supplements*, 14(3):E72–E83, 2004.

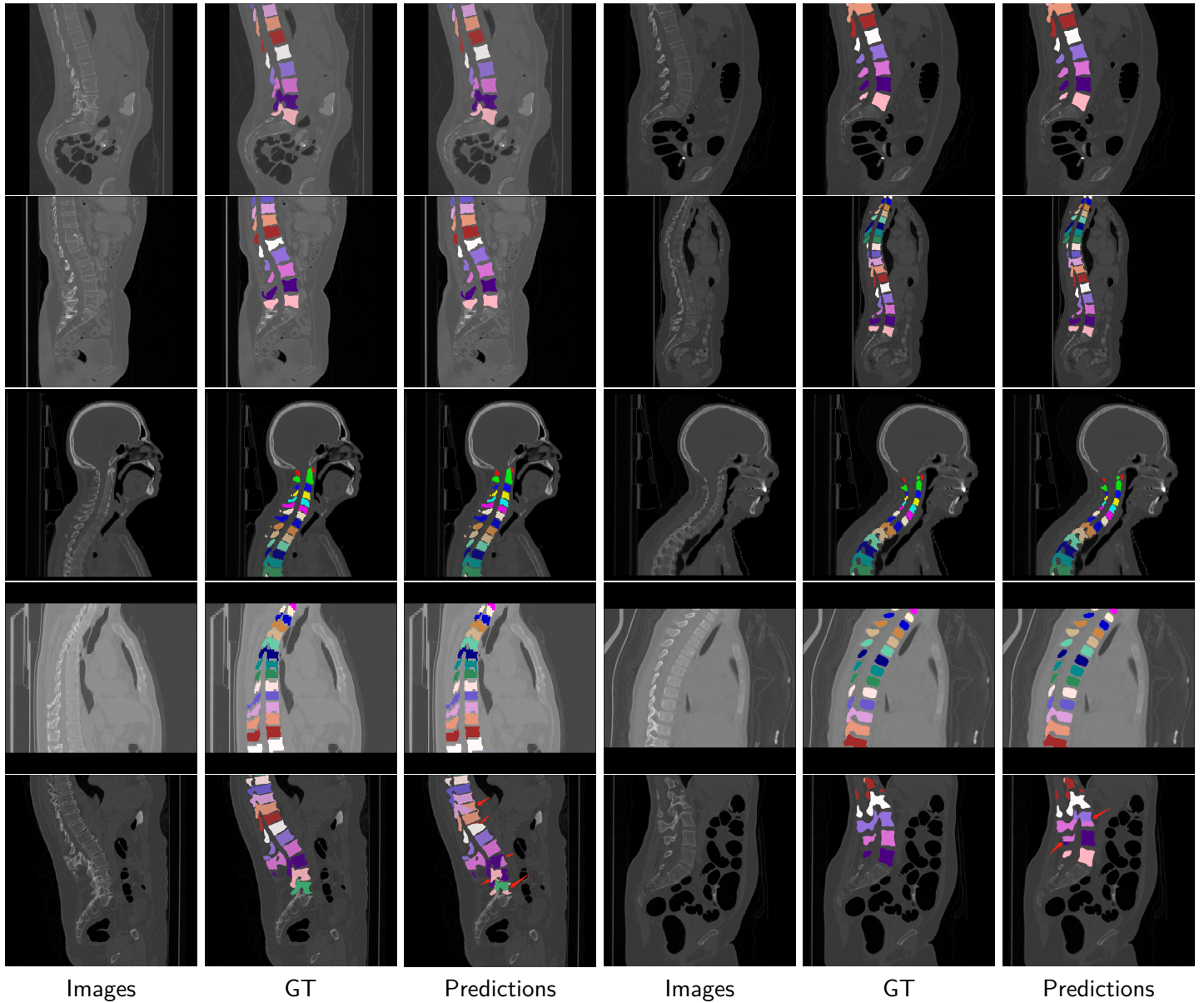| Images | GT | Predictions | Images | GT | Predictions |

Figure 4: The visualization results on different sub-dataset. The first row and the fourth row, respectively, represent the sub-dataset COLONOG, MSD T10, HNSCC-3DCT-RT, and COVID-19. The last row indicates some failed predictions resulting from spinal diseases: sacral lumbarization and lumbar sacralization.

Tatiana Bejarano, M De Ornelas Couto, and I Mihaylov. Head-and-neck squamous cell carcinoma patients with ct taken during pre-treatment, mid-treatment, and post-treatment dataset. the cancer imaging archive, 2008.

Yunliang Cai, Liansheng Wang, Michel Audette, Guoyan Zheng, and Shuo Li. *Computational Methods and Clinical Applications for Spine Imaging*. Springer, 2020.

Salman Faruqi, Chia-Lin Tseng, Cari Whyne, Majed Alghamdi, Jefferson Wilson, Sten Myrehaug, Hany Soliman, Young Lee, Pejman Maralani, Victor Yang, et al. Vertebral compression fracture after spine stereotactic body radiation therapy: a review of the pathophysiology and risk factors. *Neurosurgery*, 83(3):314–322, 2018.

Stephanie A Harmon, Thomas H Sanford, Sheng Xu, Evrim B Turkbey, Holger Roth, Ziyue Xu, Dong Yang, Andriy Myronenko, Victoria Anderson, Amel Amalou, et al. Artificial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets. *Nature Communications*, 11(1):1–7, 2020.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021.

C Daniel Johnson, Mei-Hsiu Chen, Alicia Y Toledano, Jay P Heiken, Abraham Dachman, Mark D Kuo, Christine O Menias, Betina Siewert, Jugesh I Cheema, Richard G Obregon, et al. Accuracy of ct colonography for detection of large adenomas and cancers. *New England Journal of Medicine*, 359(12):1207–1217, 2008.

Robert Korez, Bulat Ibragimov, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec. A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation. *IEEE Transactions on Medical Imaging*, 34(8):1649–1662, 2015.

Nikolas Lessmann, Bram Van Ginneken, Pim A De Jong, and Ivana Išgum. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Medical Image Analysis*, 53:142–155, 2019.

Shuo Li, Jianhua Yao, et al. *Spinal Imaging and Image Analysis*. Springer, 2015.

Christian Payer, Darko Stern, Horst Bischof, and Martin Urschler. Coarse to fine vertebrae localization and segmentation with spatialconfiguration-net and u-net. In *VISIGRAPP (5: VISAPP)*, pages 124–133, 2020.

Cheng Peng, Wei-An Lin, Haofu Liao, Rama Chellappa, and S Kevin Zhou. Saint: spatially aware interpolation network for medical slice synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7750–7759, 2020.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 234–241. Springer, 2015.

Anjany Sekuboyina, Amirhossein Bayat, Malek E Husseini, Maximilian Löffler, Markus Rempfler, Jan Kukačka, Giles Tetteh, Alexander Valentinitsch, Christian Payer, Martin Urschler, et al. Verse: a vertebrae labelling and segmentation benchmark. *arXiv preprint arXiv:2001.09193*, 2020.

Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.

Martijn Pieter Anton Starmans and Apostolia Tsirikoglou. Miccai 2024 africai imaging repository white paper, March 2024. URL `https://doi.org/10.5281/zenodo.10816769`.

Jianhua Yao, Joseph E Burns, Hector Munoz, and Ronald M Summers. Detection of vertebral body fractures based on cortical shell unwrapping. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 509–516. Springer, 2012.

Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006a.

Paul A Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C Gee, and Guido Gerig. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006b.

S Kevin Zhou, Daniel Rueckert, and Gabor Fichtinger. *Handbook of medical image computing and computer assisted intervention*. Academic Press, 2019.

S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M

Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 2021.