MedShapeNet - A Large-Scale Dataset of 3D Medical Shapes for Computer Vision


Abstract—We present MedShapeNet, a large collection of anatomical shapes (e.g., bones, organs, vessels) and 3D surgical instrument models. Prior to the deep learning era, the broad application of statistical shape models (SSMs) in medical image analysis is evidence that shapes have been commonly used to describe medical data. Nowadays, however, state-of-the-art (SOTA) deep learning algorithms in medical imaging are predominantly voxel-based. In computer vision, on the contrary, shapes (including, voxel occupancy grids, meshes, point clouds and implicit surface models) are preferred data representations in 3D, as seen from the numerous shape-related publications in premier vision conferences, such as the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), as well as the increasing popularity of ShapeNet (about 51,300 models) and Princeton ModelNet (127,915 models) in computer vision research. MedShapeNet is created as an alternative to these commonly used shape benchmarks to facilitate the translation of data-driven vision algorithms to medical applications, and it extends the opportunities to adapt SOTA vision algorithms to solve critical medical problems. Besides, the majority of the medical shapes in MedShapeNet are modeled directly on the imaging data of real patients, and therefore it complements well existing shape benchmarks consisting of computer-aided design (CAD) models. MedShapeNet currently includes more than 100,000 medical shapes, and provides annotations in the form of paired data. It is therefore also a freely available repository of 3D models for extended reality (virtual reality - VR, augmented reality - AR, mixed reality - MR) and medical 3D printing. This white paper describes in detail the motivations behind MedShapeNet, the shape acquisition procedures, the use cases, as well as the usage of the online shape search portal: [https://medshapenet.ikim.nrw/]

Index Terms—3D Medical Shapes, ShapeNet, Benchmark, Anatomy Education, Shapeomics, Deep learning, Augmented Reality, Virtual Reality, Mixed Reality, Extended Reality, Diminished Reality, Medical Visualization, 3D Printing, Stereolithography, Face Reconstruction, Medical Data Sharing, Data Privacy

1 INTRODUCTION

The success of deep learning in so many fields of applications [1], [2], [3] is in not small part due to the availability of large, high-quality datasets [4], such as ImageNet [5], CIFAR [6], and a2d2 [7]. In computer vision,
Common 3D shape representations include point clouds, voxel grids, meshes and implicit surface models (signed distance functions), which follow different data structures, cater for different algorithms but are convertible to each other [17].

- Maximilian Löffler is with the Universitätsklinikum Freiburg, Hugstetter Strasse 55, 79106 Freiburg, Germany.
- Hans Liebl is with the Department of Neuroradiology, Klinikum Rechts der Isar, Ismaninger Str. 22, 81675 Munich, Germany.
- P. F. Christ, H. B. Li and B. Menze are with the Department of Quantitative Biomedicine, University of Zurich, Winterrhuterstrasse 190, 8057, Zurich, Switzerland.
- F. Bolelli, C. Grana and L. Lumetti are with the University of Modena and Reggio Emilia, Department of Engineering “Enzo Ferrari”, Via Vivarelli 10, 41125, Modena, Italy.
- J. Ma is with the Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5S 1A8 Canada; Peter Munk Cardiac Centre, University Health Network, 585 University Ave, Toronto, ON M5G 2N2, Canada; Vector Institute, 661 University Ave Suite 710, Toronto, ON M5G 1M1, Canada.
- J. Staal is with the Shanghai AI Laboratory, Yunjin Road, Shanghai, 200032, People’s Republic of China.
- R. Charleghi and S. Beter are with the School of Mechanical and Manufacturing Engineering, UNSW, Sydney, 2052, NSW, Australia.
- A. Soomra is with the School of Computer Science and Engineering, UNSW, Sydney, 2052, NSW, Australia.
- R. Souza is with the Advanced Imaging and Artificial Intelligence Lab, Electrical and Software Engineering Department, and the Hotchkiss Brain Institute, University of Calgary, Calgary, Canada.
- L. Rittner is with the Medical Image Computing Lab, School of Electrical and Computer Engineering (FEEC), University of Campinas, Campinas, Brazil.
- R. Frayne is with the Radiology and Clinical Neurosciences Departments, the Hotchkiss Brain Institute, University of Calgary, Calgary, Canada, and the Seaman Family MR Research Centre, Foothills Medical Center, Calgary, Canada.
- T.-C. Chang is with Merck, Rahway, NJ 07065, USA.
- S.-A. Ahmadi is with the NVIDIA GmbH, Bavaria Towers - Blue Tower, Einsteinstrasse 172, 81677 Munich, Germany.
- Y. Ji and P. Luo are with the University of Hongkong, Pok Fu Lam, Hong Kong, People’s Republic of China.
- E. Verbeeck and T. V. Leeuwen are with the Department of Development & Regeneration, KU Leuven Campus Kulak, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium.
- M.-S. Kim and F. Nensa are with the Institute of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen (AöR), Hüfelandstraße 55, 45147 Essen, Germany.
- J. Gehrke is with the German Cancer Consortium (DKTK), Partner Site Essen, Hüfelandstraße 55, 45147 Essen, Germany, and the Department of Physics, TU Dortmund University, August-Schmidt-Str. 4, 44227 Dortmund, Germany.
- J. Krebs is with the Department of Cellular and Physiological Sciences, Life Sciences Centre, 1544 - 2350 Health Sciences Mall, University of British Columbia, Vancouver, British Columbia, V6T 1Z3 Canada.
- M. Reyes is with the Institute for Neuroinformatics, Technical University of Munich, Ismaninger Str. 22, Munich, Germany.
- M. S. Kamp and A. Ahourayya are with the Institute for Neuroinformatics, Technical University of Munich, Ismaninger Str. 22, Munich, Germany.
- M. Kamp is also with the M. Kamp and A. Ahourayya are with the Institute for Neuroinformatics, Technical University of Munich, Ismaninger Str. 22, Munich, Germany.
- J. Gehrke is with the German Cancer Consortium (DKTK), Partner Site Essen, Hüfelandstraße 55, 45147 Essen, Germany, and the Department of Physics, TU Dortmund University, August-Schmidt-Str. 4, 44227 Dortmund, Germany.
- C. Krebs is with the Department of Cellular and Physiological Sciences, Life Sciences Centre, 1544 - 2350 Health Sciences Mall, University of British Columbia, Vancouver, British Columbia, V6T 1Z3 Canada.
- E. Verbeeck and T. V. Leeuwen are with the Department of Development & Regeneration, KU Leuven Campus Kulak, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium.
- M.-S. Kim is with the Institute of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen (AöR), Hüfelandstraße 55, 45147 Essen, Germany.
- J. Gehrke is with the German Cancer Consortium (DKTK), Partner Site Essen, Hüfelandstraße 55, 45147 Essen, Germany, and the Department of Physics, TU Dortmund University, August-Schmidt-Str. 4, 44227 Dortmund, Germany.
- M. Kamp is with the Institute for Neuroinformatics, Ruhr University Bochum, Germany, and the Department of Data Science & AI, Monash University, Australia.
the medical imaging data (computed tomography, magnetic resonance imaging, positron emission tomography, ultra sound, X-ray) commonly used in clinical research. As a result, the transferability of state-of-the-art (SOTA) vision algorithms to medical/clinical problems is limited, since vision methods developed on general 3D shapes are not directly transferable to volumetric, gray-scale medical data. Therefore, the community needs a large, high-quality shape database for medical imaging. With MedShapeNet, we provide a large-scale dataset of 3D medical shapes, i.e., voxel occupancy grid (VOR), mesh and point representations of human anatomies (e.g., liver, heart, lung, kidney, vertebrae, rib) - formats that advanced vision algorithms are compatible with but are underrepresented in current medical imaging research. While ShapeNet is comprised of 3D computer-aided design (CAD) models of real-world objects (e.g., plane, car, chair, desk), the medical shapes from MedShapeNet are directly extracted from the imaging data of real patients (e.g., Figure 1). MedShapeNet by itself is therefore not only a unique dataset for medical imaging but also an ideal alternative and complement to the common shape benchmarks, like ShapeNet.

- C. Gsauer, F. Hötzle and B. Puladi are with the Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany.
- V. Badel is with the Institute of Fundamentals and Theory in Electrical Engineering, Graz University of Technology, 8010 Graz, Austria.
- K. Krieger, M. Gunzer and J. Chen are with the Leibniz-Institut für Analytische Wissenschaften-ISAS e.V., 44139 Dortmund, Germany.
- M. Gunzer is with the Institute for Experimental Immunology, University Hospital, University Duisburg-Essen, Hufelandstrasse 55, 45147 Essen, Germany.
- R. Röhrig and B. Puladi are with the Institute for Medical Informatics, University Hospital RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany.
- A. J. V. Loomis is with the Computer Vision for Human-Computer Interaction Lab, Department of Informatics, Karlsruhe Institute of Technology, Am Fasanengarten 5, 76131 Karlsruhe, Germany.
- L. E. Podleska is with the Department of Tumor Orthopedics and Sarcoma Surgery, University Hospital Essen (AöR), Hufelandstrasse 55, 45147 Essen, Germany.
- M. A. Fink is with the Clinic for Diagnostic and Interventional Radiology, University Hospital Heidelberg, Im Neuenheimer Feld 420, 69120 Heidelberg, Germany.
- K. Tserpes is with the Department of Informatics and Telematics, Harokopio University of Athens, 9 Omira, 177 78, Tiareos, Greece.
- E. A. Garza-Villarreal, T. Balducci and D. Angeles-Valdez are with the Institute of Neurobiolohy, Universidad Nacional Autónoma de México campus Juriquilla, Boulevard Juriquilla 3001, Juriquilla, Queretaro, 76230, México. D. Angeles-Valdez is also with the Department of Biomedical Sciences of Cells and Systems, Cognitive Neuroscience Center, University Medical Center Groningen, University of Groningen, P.O. Box 3000, 9700 AD Groningen, The Netherlands.
- F. Dubost is with Google LLC, California, USA.
- S. Schreiber, H. Mattern and O. Speck are with the German Centre for Medical Image Computing and Analytics, Potsdam, Germany.
- Y. Jin is with the Center for Biomedical Imaging Computing and Analytics, Potsdam, Germany.
- O. Basa is with the Department of Pediatrics III, University Hospital Essen, University Medicine Essen, Hufelandstrasse 55, 45147 Essen, Germany.
- T. Maal is with the 3D Imaging Lab, Radboud University Nijmegen Medical Center, Geert Grooteplein 10, 6525 GA, Nijmegen, The Netherlands.
- M. J. H. Witjes is with the Department of Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, P.O. Box 3000, 9700 AD Groningen, The Netherlands.
- O. Basa and J. Egger are with the Center for Virtual and Extended Reality in Medicine (ZeRM), University Hospital Essen, University Medicine Essen, Hufelandstrasse 55, 45147 Essen, Germany.
- J. Li, A. Pepe, C. Gsauer, Y. Jin, G. Luijten, N. Solak and J. Egger are with the Institute of Computer Graphics and Visualization (IVC), Graz University of Technology, Inffeldgasse 16c, 8010 Graz, Austria.
- J. Li, A. Pepe, C. Gsauer, Y. Jin, G. Luijten, N. Solak and J. Egger are with the Institute of Computer Graphics and Visualization (IVC), Graz University of Technology, Inffeldgasse 16c, 8010 Graz, Austria.
- X. Chen is with the Center for Biomedical Manufacturing and Life Quality Engineering, State Key Laboratory of Mechanical System and Vibrations, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, People's Republic of China.
- V. D. Vu is with the Department of Diagnostic and Interventional Radiology, University Hospital Giessen, Justus-Liebig-University Giessen, Ismaningerstr. 22, 85748 Garching bei München, Germany.
- A. R. Memon is with the Department of Mechanical Engineering, Mehran University of Engineering and Technology, Jamshoro 76062, Sindh, Pakistan.
- A. R. Memon and X. Chen are with the Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, People's Republic of China.
- J. S. Kirschke is with the Geschäftsführende Oberarzt Abteilung für Interventionselle und Diagnostische Neuroradiologie, Universitätsklinikum der Technischen Universität München, Ismaningerstr. 22, 81675 München, Germany.
- Ezequiel de la Rosa is with imec, Kingdom Begaustria 1b, 3012 Leuven, Belgium, and the Department of Informatics, Technical University of Munich, Boltzmannstraße 3, 85748 Garching bei München, Germany.
MedShapeNet makes an effort to bridge the gap between the medical imaging and computer vision community, and to promote the translation of vision algorithms to medical applications. The benefits are reciprocal: it makes it easier for vision researchers to work on medical applications and encourages medical researchers to revisit and adopt shape-based methods from computer vision for medical problems. The MICCAI society, a leading professional organization in medical image computing and computer assisted intervention, has initiated a special interest group in Shape in Medical Imaging (ShapeMI, https://shapemi.github.io/), suggesting the significance of the role shape-based methods play in this field. Table 1 provides a non-inclusive list of organizations/events that focus on promoting shape methods for medical applications.

MedShapeNet includes diverse anatomical shapes and can facilitate the development and evaluation of data-driven, shape-based methods for a variety of medical as well as vision problems. On the one hand, numerous existing medical problems can be solved using shape-based methods. A typical example is cranial implant design [20], [21], [22], [23], [24], [25], [26], which is commonly formulated as a shape completion problem and solved using well established completion methods from computer vision [27], [25], [27], [28], [29], [31]. The same concept can be conveniently extended to the design of other bone grafts (e.g., ribs, spine) and even artificial organs (e.g., liver, heart, kidney) for 3D bio-printing. Another representative example is statistical shape modeling (SSM), which has long been employed for medical image segmentation [32], [33] and anatomy modeling [34], [35], [36], [37], [38], [39], [40], [41] by the community. Shape priors and/or geometric constraints of various anatomies (e.g., aorta, skull) can also be derived from MedShapeNet for downstream segmentation and reconstructive tasks [12], [13], [41], [45], [46], [47]. Last but not least, MedShapeNet offers opportunities to explore shape-based methods for problems that are traditionally solved based on gray-scale medical images, such as disease diagnosis. Switching to medical shapes allows one to exploit more computationally efficient and geometry-oriented methods, such as sparse convolutional neural networks [48], for the medical diagnostic problems. On the other hand, anatomical shapes are also commonly used for general computer vision research aimed at (primarily) non-medical applications, such as facial modeling [49], [50] and internal anatomy (e.g., skeleton, organs) inference [51], [52].

MedShapeNet also contains pathological anatomies, such as tumors, brains, kidneys and livers (Figure 3), as well as brains from patients with cognitive impairment (e.g., Alzheimer’s disease) or substance use disorder (e.g., alcohol use disorder - AUD, cocaine use disorder - CUD). Machine learning models can be trained for automatic abnormality detection using such shape data. Through statistical analysis and comparison, geometric differences between normal and pathological anatomies can be quantified, which facilitates automatic diagnostics and the discovery of geometric biomarkers [53], [54]. MedShapeNet can also be used for anatomy education, as it provides the 3D models of a variety of human anatomies, both normal and pathological, that can be 3D printed or used digitally in an extend reality, such as augmented reality (AR), environment [55], [56]. MedShapeNet also benefits researchers who want to study the shape variations of a certain anatomy, but do not have access to the 3D scans and lack the resources to create the segmentations manually.

The manuscript is organized as follows. Section 2 discusses the shape and voxel features in medical imaging, and the motivation of this project. Section 3 introduces the different sources from which the shape data in MedShapeNet are derived. Section 4 presents several interesting use cases of MedShapeNet, and demonstrates how MedShapeNet can be used in real-world applications in computer vision, medical imaging and augmented reality. Section 5 introduces the online interface of MedShapeNet and how to use it. Section 6 concludes the manuscript and discusses the future work.

2 SHAPE AND VOLUME FEATURES

Shapes describe objects’ geometries, provide a foundation for computer vision, and serve as a computationally efficient
way to represent images despite not capturing voxel features like gray-scale medical images.

Even though the main motivation behind MedShapeNet is to emphasize the importance of shape characteristics, such as jaggedness, volume, elongation, etc., over voxel features, and to show that voxel features are redundant for certain tasks, learning algorithms might require additional (voxel) information to construct a decision boundary in some situations. For example, liver and brain tumors can have a noticeable impact on the morphology and/or volume of the corresponding organ (Figure 3), so that learning algorithms can easily distinguish between healthy and tumorous organs based on these shape features alone. However, for pathologies that do not induce (obvious) morphological changes, such as neurodegenerative diseases (e.g., mild cognitive impairment or Alzheimer’s disease), shape-related features might not be discriminative enough for learning algorithms to converge during training. In the latter case, adding additional voxel features is beneficial. Refer to Section 4.2.4 for preliminary experimental evidence of these assumptions.

Another example where voxel features are essential is when accurate spatial location is necessary, such as during precision tumor therapy. In [57], the authors show that spatial predictive maps that indicate areas of early tumor (glioblastoma) recurrence and infiltration can be derived from preoperative MRIs, and used for targeted radiotherapy [58]. The predictive maps are generated via a voxel-wise classification of the gray-scale tumor voxels. As shown in Figure 3, the predictive map shows the spatial pseudo-probability of tumor infiltration. Areas with high probability have higher risks of tumor recurrence after resection.

How to optimally combine voxel features with shapes is an interesting topic requiring further investigation. With MedShapeNet, one can investigate (1) to what degree a pathological condition, such as tumor, Alzheimer’s disease (AD) and substance use disorder (SUD) can be captured by the shape features of the organs affected (e.g., the brain), determined by the convergence of a classifier when trained on shape features alone; (2) what shape features are the most discriminative of a pathology and how to calculate them [59], (3) how to effectively integrate voxel features into shapes when shape features alone are not sufficient; and (4) whether there are associations between voxel and shape features. In the example of [57], one can ask whether the high infiltration voxels induce morphological changes to the tumor (boundaries) correspondingly.

To answer these questions and support future research on this endeavor, MedShapeNet links the ‘source of shapes’ i.e., the original medical images with its shape collections, so that the voxel information of a specific shape can be retrieved whenever needed. The following section describes the ‘source of shapes’ in detail.

3 SOURCES OF SHAPES

The anatomical shapes in MedShapeNet are converted from binary segmentation masks (voxel occupancy grids) of organs, bones, vessels, muscles, etc., using Marching Cubes [105]. We collect the segmentation masks from different sources, where the segmentation masks are either generated automatically by a segmentation network (e.g., in the case of TotalSegmentator) or manually, as those of the ground truth in the training set of a public medical image segmentation challenge [106], [107], [108]. Some of the masks are from our own datasets. Table 2 summarizes the data sources, such as TotalSegmentator [60], MUG500+ [61], the Human Connectome Projects (HCP) [62] and the aortic vessel tree (AVT) dataset [64]. Miscellaneous sources include the Skull-striped MRI Glioblastoma Multiforme (GBM) Dataset [65] and the Medical Augmented Reality Facial Data Collection [63], as shown in Figure 1. Note that different sources could contain the same anatomy. For example, both the TotalSegmentator and VerSe datasets include vertebrae. The anatomical shapes in MedShapeNet are provided as meshes (stl), points and voxel occupancy grids to cater for different vision algorithms.

Privacy and ethics considerations: The MedShapeNet database is created exclusively for research and educational purposes. The majority of the source datasets are Creative Commons (CC)- or CC BY 4.0-licensed (Refer to Table 2 for data licenses). Publicly sharing medical data is encouraged but regulated at the same time due to potential privacy concerns [109], [110]. MedShapeNet does not include gray-scale medical images, which contain patient-specific information, such as racial identity, that can be inferred using an identity recognition network [111], [112]. Training on shape data encourages a machine learning model to focus on learning discriminative geometric features rather than learning irrelevant patients’ identities, which may undermine the robustness and trustworthiness of the

<table>
<thead>
<tr>
<th>Sources (link)</th>
<th>Description</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zise Institute Berlin (ZIB)</td>
<td>shape-informed medical image segmentation and shape priors in medical imaging</td>
<td>research group</td>
</tr>
<tr>
<td>ShapeML</td>
<td>shape modeling and analysis in medical imaging</td>
<td>MICCAI workshop</td>
</tr>
<tr>
<td>SIG</td>
<td>shape processing/analysis/learning in medical imaging</td>
<td>MICCAI special interest group (SIG)</td>
</tr>
<tr>
<td>Autoplant I, II</td>
<td>skull shape reconstruction and completion</td>
<td>professional organization</td>
</tr>
<tr>
<td>Wish</td>
<td>women in Shape Analysis, shape modeling</td>
<td>MICCAI workshop</td>
</tr>
<tr>
<td>STACOM</td>
<td>statistical atlases and computational models of the heart</td>
<td>research group</td>
</tr>
<tr>
<td>SAMIA</td>
<td>shape analysis in medical image analysis</td>
<td>MICCAI workshop</td>
</tr>
<tr>
<td>CIBC</td>
<td>image and geometric analysis</td>
<td>journal special issue</td>
</tr>
<tr>
<td>GeoMedIA</td>
<td>geometric deep learning in medical image analysis</td>
<td>proceedings</td>
</tr>
<tr>
<td>IEEE TMI</td>
<td>geometric deep learning in medical imaging</td>
<td>book</td>
</tr>
<tr>
<td>PMLU</td>
<td>geometric deep learning in medical image analysis</td>
<td>proceedings</td>
</tr>
<tr>
<td>Elsevier</td>
<td>Riemannian geometric statistics in medical image analysis</td>
<td>proceedings</td>
</tr>
<tr>
<td>Springer</td>
<td>geometric methods in bio-medical image processing</td>
<td>proceedings</td>
</tr>
</tbody>
</table>

TABLE 1
A Non-inclusive List of Organizations/Events Featuring Shape Methods for Medical Applications
The 1200 Subjects Data Release in Human Connectome Projects (HCP) includes 1113 structural 3T head MRI scans of healthy young adults. Each MRI scan, the segmentation masks of the skull and the brain are extracted using the Cortical Surface Extraction script provided by BrainSuite [http://brainsuite.org/]. Due to the highly complex brain geometries, the size of a brain mesh converted from a segmentation mask exceeds one Gigabyte. Considering the limited space for storing the shape data, we downsized the brain masks by a factor of 1.6 before converting them to meshes. This course of action reduces the size of each brain shape to 200 MB - 500 MB at the cost of reduced shape quality. A example of such brain shape is shown in Figure 1.

3.1 TotalSegmentator
The TotalSegmentator dataset from Wassermann et al. [60] includes over 1000 CT scans and the corresponding segmentations of 104 anatomical structures covering the whole body, which are generated automatically by an nnUNet-based segmentation network [114] and have been used, for example, to improve disease diagnosis by correlating organ volumes with disease occurrences in humans [115].

3.2 Human Connectome Projects (HCP)

The 1200 Subjects Data Release from the Human Connectome Projects (HCP) includes 1113 structural 3T head MRI scans of healthy young adults. From each MRI scan, the segmentation masks of the skull and the brain are extracted using the Cortical Surface Extraction script provided by BrainSuite [http://brainsuite.org/]. Due to the highly complex brain geometries, the size of a brain mesh converted from a segmentation mask exceeds one Gigabyte. Considering the limited space for storing the shape data, we downsized the brain masks by a factor of 1.6 before converting them to meshes. This course of action reduces the size of each brain shape to 200 MB - 500 MB at the cost of reduced shape quality. An example of such brain shape is shown in Figure 1.

3.3 MUG500+
The MUG500+ dataset contains the binary segmentation masks and meshes of 500 healthy human skulls and 29 cranioectomy skulls with surgical defects [61]. The skull masks are segmented from head CT scans by thresholding.

3.4 SkullBreak/SkullFix
The SkullBreak/SkullFix dataset includes the binary segmentation masks of healthy human skulls and the corresponding skulls with artificial defects. The binary skull masks are segmented from head CT scans from the CQ500 dataset (http://headctstudy.qure.ai/dataset), using thresholding, similar to MUG500+ [61].

3.5 AVT

The aortic vessel tree (AVT) dataset [64] contains 56 computed tomography angiography (CTA) scans of healthy aorta as well as the segmentation masks of the corresponding aortic vessel trees, including the aorta, aortic arch, branch and iliac arteries, as shown in Figure 1.
Fig. 3. Example pathological shapes in MedShapeNet, including tumorous kidney (paired), brain (with real and synthetic tumors), liver and head & neck, as well as diseased coronary arteries. For illustration purpose, the opacity of some shapes is reduced to reveal the underlying tumors. We can study the effects of tumors on the morphological changes of an anatomy (e.g., brain) using such pathological data.

3.6 VerSe
The large scale vertebrae segmentation (VerSe) challenge [67], [116] provides the segmentation masks of vertebrae from around 210 subjects [117], [118]. 2745 vertebra shapes are generated from the challenge dataset.

3.7 ASOCA
The automated segmentation of coronary arteries (ASOCA) challenge provides the manual segmentations of 20 normal and 20 diseased coronary arteries [94].

3.8 3DTeethSeg
Automated teeth localization, segmentation, and labeling from intra-oral 3D scans are crucial tasks in modern dentistry, significantly improving dental diagnostics, treatment planning, and population-based studies on oral health. Before initiating any orthodontic or restorative treatment planning, it is essential for a CAD system to accurately segment and label each instance of teeth in the 3D dental scan. This eliminates the need of time-consuming manual adjustments by the dentist. To address this need, the 3D Teeth Scan Segmentation and Labeling Challenge (3DTeethSeg) [72], [73] was organized in conjunction with the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) in 2022. This challenge provides the upper and lower intra-oral 3D scans of 900 subjects, along with the corresponding manual annotations for teeth segmentation and labeling tasks. The data annotation was performed in collaboration with clinical evaluators with more than 10 years of expertise in orthodontist, dental surgery, and endodontics. A preliminary benchmark of state-of-the-art methods for the challenge can be found in [72].

3.9 LNDb
The data from the automatic lung cancer patient management (LNDb) challenge [98], [99] comprises lung nodule segmentations performed by five radiologists on low-dose computed tomography images within the scope of lung cancer screening. A total of 861 lung nodule segmentation masks are publicly available, corresponding to 625 individual nodules segmented on 204 CTs. Radiologists were asked to independently screen each CT and identify all pulmonary nodules and segment those with an in-plane dimension larger than or equal to 3mm. No consensus or review between radiologists was performed, meaning that there is a variable number of segmentations per nodule (between 1 and 3).

3.10 EMIDEC
The Emidec (automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI) dataset is composed of 150 exams with delayed enhancement-MRI (or DE-MRI) images in short axis orientation covering the left ventricle from normal cases or patients with myocardial infarction, with the contouring of the myocardium and diseased areas (if present) from experts in the domains [101], [102]. The database is composed of the imaging
exam and the associated clinical information. The targeted cohort is any patient admitted in a cardiac emergency department with symptoms of a heart attack. Indeed, DE-MRI is a method of choice to evaluate the extent of myocardial infarction, and by extension, to assess viable tissues after an injury. The images are acquired roughly 10 minutes after the injection of a gadolinium-based contrast agent, and then the fibrotic area appears bright in T1-weighted DE-MRI whereas normal tissue appears dark. There is an unbalanced distribution between normal (1/3) and pathological (2/3) cases, corresponding roughly to real life in an MRI department. This dataset was available as part of the Emidec challenge organized in conjunction with the STACOM workshop during the MICCAI conference in 2020 [57]. Even if the data are freely available for research topic, the owner stays the University Hospital of Dijon (France).

3.11 ToothFairy
Dental implant placement within the jawbone is a routinely executed surgical procedure, which can become complex due to the local presence of the Inferior Alveolar Nerve (IAN) crossing the homonymous osseous structure (the Inferior Alveolar Canal, IAC in short). In particular, the nerve is in close relation to the roots of molars, and its position must thus be carefully detailed before the surgical removal. As avoiding contact with the IAN is a primary concern during these operations, segmentation plays a key role in surgical preparations. With the goal of pushing the development of deep learning frameworks to automatically segment the IAC, the ToothFairy dataset has been released by “ToothFairy: A Cone-beam Computed Tomography Segmentation Challenge” [92] organized within MICCAI 2023. ToothFairy extends the previously released Maxillo dataset [91], [119], [120], and it comprises 443 dental scans, captured using the NewTom/NTVGiMK4 CBCT scanner, operating at 3 mA and 110 kV, with a voxel size of 0.3 mm³. The scans have been acquired with an intra-slice distance of 0.3 mm, yielding volumes with shapes ranging from (148, 265, 312) to (169, 342, 370) across the Z, Y, and X axes, respectively. The voxel values, represented in Hounsfield Units (HU), span from −1000 to 5264. The dataset includes 2D sparse annotations for all 443 volumes, while only a subset of 153 volumes features detailed 3D voxel-level annotations of the IAC. The ground-truth annotations of the IAC have been produced by a team of five experienced maxillofacial surgeons using an ad-hoc developed tool that leverages different computer vision techniques to assist the user during the annotation [121], [122]. An additional test-set of 50 CBCT volumes has been acquired using a standard CBCT scanning protocol (i-CAT, 3D Imaging System, Imaging Sciences International Inc, Hatfield, PA, USA) in “Extended Field” modus (FOV: 16cm diameter/22 cm height; scan time: 2 × 20s; voxel size: 0.4 mm). These data represent the ToothFairy challenge evaluation dataset and, in this case, only the ground-truth annotations are made available.

3.12 HECKTOR
The training set of the HEad and neCK TumOR segmentation and outcome prediction (HECKTOR) challenge [74], [75] comprises 524 3D FDG-PET/CT images from seven hospitals with manual primary tumor and metastatic lymph nodes contours. The data originates from FDG-PET and low-dose non-contrast-enhanced CT images (acquired with combined PET/CT scanners) of the H&N region of patients with oropharyngeal H&N cancer. The training set of the HECKTOR challenge is used for MedShapeNet.

3.13 autoPET
Similar to TotalSegmentor, whole-body segmentations are extracted from the PET/CT dataset provided by the AutoPet challenge [84], [86], using an semi-supervised segmentation network [83], [123], [124]. The autoPET dataset itself comes from cancer patients and also includes the manual segmentations of whole-body tumor lesions. It should be noted that the morphologies of some of the anatomies might be affected due to the existence of tumors.

3.14 Calgary-Campinas
The Calgary-Campinas (CC) dataset [95] consists of T1 magnetic resonance imaging (MRI) volumes acquired in 359 presumed healthy subjects on scanners from three different vendors (GE, Philips, and Siemens) and at two magnetic field strengths (1.5 T and 3 T). Data were obtained using T1-weighted 3D imaging sequences (3D MP-RAGE (Philips, Siemens), and a comparable T1-weighted spoiled gradient echo sequence (GE)) designed to produce high-quality anatomical data with 1 mm³ voxels. Age and gender for all subjects were known (176 M: 183 F, 53.5 +/- 7.8 years, min:18 years, max: 80 years), however information about subject ethnicity was not available. Probabilistic brain masks were obtained by combining the output of eight automated brain segmentation algorithms [125], [126], [127], [128], [129], [130], [131], [132] using the Simultaneous Truth and Performance Level Estimation (STAPLE) algorithms [133]. The quality of the brain masks was validated against T2 manual brain segmentations obtained in a stratified manner across vendor, magnetic field, and subject sex combinations. The CC dataset has been used to investigate brain extraction models [134], [135], domain shift and adaptation in brain MRI [136], [137], as well as MRI reconstruction [138], [139].

3.15 AMOS
The AMOS dataset [97], both diverse and robust, includes 500 CT and 100 MRI images gathered from a variety of scanners and locations. It covers 15 distinct categories of abdominal organs: the spleen, right kidney, left kidney, gallbladder, esophagus, liver, stomach, aorta, inferior vena cava, pancreas, right adrenal gland, left adrenal gland, duodenum, bladder, and prostate/uterus. The images were predominantly collected from patient examinations involving abdominal tumors or other abnormalities.

3.16 AbdomenCT-1K and FLARE
The AbdomenCT-1K dataset includes the manual segmentations of the liver, kidney, spleen, and pancreas from over 1000 CT scans [67]. A subset of the dataset was used in the fast and low-resource semi-supervised abdominal
organ segmentation (FLARE) challenge, which provides the manual segmentation of 13 abdomen organs, including the liver, spleen, pancreas, right kidney, left kidney, stomach, gallbladder, esophagus, aorta, inferior vena cava, right adrenal gland, left adrenal gland, and duodenum [88], [89]. Note that some of the CT scans are acquired from cancer patients. Tumors can affect the morphologies of these organs.

3.17 ISLES
The ischemic stroke lesion segmentation (ISLES) challenge [79] provides a dataset comprising of 250 brain MRIs along with binary masks depicting stroke infarctions. The dataset encompasses diverse brain lesions in terms of volume, location, and stroke pattern. The manually delineated segmentation masks are derived by refining pre-segmentations obtained using a 3D UNet [140].

3.18 Synthetic Anatomical Shapes
Generative adversarial networks (GANs) are capable of generating realistic 3D data [141]. Besides real anatomical shapes, MedShapeNet also includes synthetic shapes generated by GANs, which can be used for augmenting the dataset in deep learning-based tasks. In MedShapeNet, we use GANs to generate synthetic tumors for 27390 real brains, as shown in Figure 3. These synthetic brain masks can be used in combination with the original tumor labels to train a tumor segmentation network.

3.19 Medical Instrument
Besides anatomical shapes, MedShapeNet also contains the 3D models of medical instruments used primarily in oral and cranio-maxillofacial surgeries, such as the drill bits, scalpel and chisel, as shown in Figure 4. The 3D instrument models are obtained by scanning the corresponding instruments manually using two structured-light-based 3D scanners, namely, Autoscan Inspec (Shining 3D Corporation, Hangzhou, Zhejiang, China) and Artec Leo (Artec3D, Senningerberg, Canton Luxembourg, Luxembourg). The initial scans are post-processed (e.g., noise removal) using proprietary software, Ultrascan version 2.0.0.7 and Artec Studio 17 Professional, before they are incorporated into the database. These instrument models can be used for surgical tool tracking (detection, classification) in augmented reality (AR) and mixed reality (MR) [55] for medical education and research. They can also be used in virtual reality (VR) applications.

3.20 Pathological Shapes
To increase the variability of the shape collections, MedShapeNet contains not only normal/healthy anatomical shapes, such as the kidneys from TotalSegmentor and the brains from HCP, but also pathological ones, which are derived from patients diagnosed with a specific pathological condition, such as tumor (liver, kidney, etc) and CUD (SUDMEX CONN, Table 2). Figure 3 shows the tumorous kidneys, brains, livers and head & neck, as well as diseased coronary arteries from different sources. We also use generative adversarial networks (GANs) to generate synthetic brain tumors, as shown in Figure 3.

4 Annotation and Use Cases
MedShapeNet provides annotations in the form of paired data. Large, high-quality, paired data are valuable assets in computer vision research [51], [142], as they facilitate supervised training of machine learning models and
promising SOTA results. For example, Yu, J. et al. [142] curated a dataset, CelebV-Text, containing facial text-video pairs, which can be used for text-driven generation of face-centric videos. Similarly, Xing, J. et al. [143] used BIWI [144] and VOCASET [145], datasets containing paired audio (e.g., speech)-visual (e.g., facial expressions/motions) sequences, for speech-driven 3D facial animation. Keller, M., et al. [51] constructed a dataset containing body surface-skeleton pairs extracted from 2000 X-ray absorptiometry (DXA) scans. A regressor was trained to infer the inside skeleton given the outside body surface of humans in various shapes and poses. In these examples, the input are the texts, audios and body surfaces, while the ground truth, a.k.a. annotations, are the corresponding videos, 3D facial models and skeletons. In [52], the authors constructed a paired pose-organ dataset and trained a deep model on it to infer the deformation of internal organs from patients’ poses. The pose parameters were derived from whole-body skin segmentations of the CT dataset, while the organ deformations were calculated from the 3D models of the corresponding internal organs. In MedShapeNet, pairedness is defined as having two composites (anatomical shapes and/or meta information) coming from the same subject, and one of them is used as input and the other is used as the ground truth. The most basic paired data in MedShapeNet consist of the shapes and their corresponding anatomical categories, such as ‘liver’, ‘heart’, ‘kidney’, and ‘lung’, which can be used to train a classifier for anatomical shape categorization. Synthetic shapes are marked with '_synthetic', to distinguish them from shapes obtained from real imaging data.

4.1 Benchmarks Derived from MedShapeNet

Benchmark datasets for various interesting shape-based applications can be derived from MedShapeNet in the form of paired data, which facilitate supervised learning of a mapping relationship, i.e., paired data can be used as input and ground truth for training a deep neural network. Based on their direct applications, we roughly group all potential benchmarks into three categories: discriminative, reconstructive and variational. The following discusses the three categories of benchmarks in detail. Table 3 shows a non-inclusive list of benchmarks (paired data) that can be derived from MedShapeNet. In Section 4.2, we present in specific four of the benchmarks and their corresponding use cases.

4.1.1 Discriminative Benchmarks

The paired data are comprised of the patients’ meta information, such as pathologies, medical histories, and the corresponding anatomical shapes. An example of such paired data would be the liver shapes from healthy subjects and patients diagnosed with liver cancer. The health status (i.e., healthy, cancerous) is extracted from the patients’ meta information, while the liver shapes are derived from the corresponding segmentation masks. These benchmarks are mainly used for diagnostic tasks, in which a classifier is trained to discriminate cancerous livers from healthy ones based on liver shapes. Diagnosis (screening) of a pathological condition, such as cancer, is usually based on gray-scale medical images. Nevertheless, with the Discriminative benchmarks, one can investigate the possibility of discriminating between pathological and healthy subjects using only the shape of the affected
organ(s). Furthermore, analogous to 3D shape classification for shape retrieval, a classifier can be trained to classify the shapes into different anatomical categories.

4.1.2 Reconstructive Benchmarks

The paired data are comprised of different anatomical shapes derived from the whole-body segmentations of a patient. These benchmarks are usually used in reconstructive tasks, where the 3D shapes of an anatomy need to be reconstructed under the geometric constraint of existing ones. Numerous novel applications can be developed using such paired data. For example, given paired skull-face shapes (Figure 5), we can train a regressor to reconstruct human faces from the skeletal remains, specifically the skulls, to automate forensic facial reconstruction [146], which is considered a tedious, expensive and highly subjective procedure in archaeological research and criminal investigation; given paired skin-fat shapes derived from whole-body segmentations (Figure 5), a machine learning model can be trained to predict the spatial distribution of body fat, an important health risk indicator, from body surfaces (i.e., skins) [147]. Similarly, we can infer other internal body compositions (e.g., skeletons, organs) from a person’s body surface and vice versa, or infer the 3D shape a missing internal organ given its surrounding anatomies. New reconstructions are expected to be naturally aligned with given anatomies (i.e., the input). Such a naturalness criterion is automatically enforced by training on the paired data derived from the same subject. Therefore, these benchmarks are also potentially useful for applications where realism is desired e.g., animation.

4.1.3 Variational Benchmarks

Variational benchmarks are usually used for conditional reconstruction of 3D anatomical shapes. Besides the geometric constraints and the naturalness criterion mentioned above, new reconstructions are expected to have an additional attribute, such as age, gender and pathology, which can be extracted from the patients’ meta information as in the Discriminative Benchmarks. For example, it is possible to reconstruct multiple faces of different ages from the same skull, by including the meta information age as a supervising factor during training. Similarly, it is also possible to impose a pathological condition, such as tumor, on healthy anatomies or model the morphological changes of an anatomy during disease progression [148]. Variational auto-encoder (VAE) [149] and GANs are commonly used for such conditional reconstructive tasks.

4.2 Use Cases of MedShapeNet

In this section, we describe five real-world use cases of MedShapeNet, including (1) a forensic facial reconstructor, which reconstructs soft facial structures from the underlying skull; (2) an anatomy completer [150], which reconstructs the 3D shapes of anatomies that are missing in the input; (3) a skull reconstructer, which reconstruct the full skull structures when the skull is damaged, e.g., when (part of) the cranium or facial bones are missing; (4) a brain shape classifier that detects tumorous brains and (5) anatomy education in AR/MR. We show that problems (1-3) can be solved under a shape completion/inpainting framework, an active area of research in computer vision [151], [152], [153], [154], [155], where the 3D head models, the complete set of anatomies and the full skulls are regarded as the ground truth, while the skulls, the incomplete anatomy set (in which one or several anatomies are missing) and the damaged skulls are the input, respectively. Convolutional neural networks are trained to learn the respective mappings. We derived such paired skull-face and anatomy datasets from whole body segmentations as described in Section 4. Damaged skulls can be generated by removing part of the bone structures from full skulls [24], [26]. Note that this section only aims at demonstrating how MedShapeNet can be used to solve vision/medical problems, rather than presenting SOTA results for each problem. To build upon the preliminary investigation, please refer to the codes and pretrained models that are publicly released at https://github.com/Jianningli/medshapenet-feedback.

4.2.1 Forensic Facial Reconstruction

Forensic facial reconstruction refers to the process of restoring a persons facial features from the underlying skull. It is a common practice in archaeological research and criminal investigation, where the identity of an ancient person or victim needs to be determined from the remains [146]. Forensic facial reconstruction is usually carried out...
Fig. 6. Benchmarks for various vision applications can be derived from MedShapeNet, such as (A) forensic facial reconstruction, (B) anatomical shape reconstruction, and (C) skull reconstruction.

manually by a designer or sculptor, which is highly time-consuming and subjective. To automate this process, a facial reconstructor can be trained using paired skull-face data derived from the whole-body PET-CT segmentations in MedShapeNet, as seen in Figure 5. Figure 6 (A) shows how the paired skull-face data can be extracted from the whole-body segmentations in MedShapeNet. An input skull that is not included in training, the prediction from the facial reconstructor and the ground truth are also illustrated. We can see that the prediction and the ground truth bear sufficient resemblance for identification purposes.

4.2.2 Multi-class Anatomy Completion

An anatomy completer learns the spatial and geometric relationship among different anatomies of the same person. Given a set of anatomies, the anatomy completer detects and then reconstructs the ones that are missing. Twelve organs are derived from the whole-body segmentations of TotalSegmentor, including the lung, heart, spleen, stomach, pancreas, spine, rib cage, liver, kidney, aorta, a pair of autochthon muscles, and the pulmonary artery. Random anatomies are removed from them to create multiple incomplete anatomy sets, as shown in Figure 6 (B). A convolutional denoising auto-encoder is trained to learn a many-to-one mapping between the incomplete sets and the 12 anatomies. Figure 6 (B) also illustrates an input and the corresponding prediction in 3D and 2D coronary views. The completer reconstructs the 3D shapes of the missing anatomies in different classes, which geometrically and spatially fit existing ones. The multi-class anatomy completer is potentially helpful in creating pseudo labels for whole-body segmentation, where it generates initial segmentation masks for the anatomies that have not been annotated in a whole-body CT scan. Refer to [150] for implementation details of the anatomy completer.

4.2.3 Skull Reconstruction

The task aims to reconstruct a full skull when the skull is damaged on the facial area, as seen in Figure 6 (C). Damaged skulls can be generated by erasing (part of) the facial voxels from full skulls, and a machine learning model can be trained on such paired skulls i.e., damaged and the corresponding full skull, to restore the erased voxels. Refer to [156] for implementation details of the skull reconstruction model. Damaged skulls can also be generated by erasing voxels around the cranium, and the same model can be trained for automatic cranial implant design [24], [26].

4.2.4 Screening and Classification of Brain Tumors

Conventional data-driven methods for the screening and classification of brain tumors are usually based on gray-scale MRIs [157], [158], [159]. The input of the classifier can be either the whole or skull-stripped MRI scans [160]. In this use case, we train a convolutional neural network (CNN)-based classifier using instead only the brain shapes represented as binary voxel grids, to discriminate between tumorous and healthy brains. The classifier has shown good convergence and generalizability, achieving over 80% accuracy on the training and test set, respectively. The experiment demonstrates that the existence of tumors are reflected on the brain morphologies that can be captured by a standard CNN-based classifier, and that voxel features from gray-scale MRIs are redundant for the tumor detection task. Similar results are observed when the classifier is trained to distinguish between male and female brain shapes. It is shown that the volume differences between tumorous versus non-tumorous, and male versus female brains are statistically significant (t-test) - a shape-related feature that could have been learnt by the classifier to make decisions. It remains to be investigated whether the conclusion holds true for the stratification of different tumor subtypes.

Nevertheless, the classifier cannot converge properly when trained to discriminate brain shapes extracted from healthy subjects and CUD or AD patients, indicating that these brain pathologies are not well reflected on shape features. As discussed in Section 2, how to extract more
discriminative shape features or incorporate voxel features into the training process when shape features alone are insufficient require future investigation.

4.2.5 Anatomy Education in Extend Reality (XR) & 3D Printing

Besides data-driven research, MedShapeNet can also benefit a variety of AR/MR/VR applications that require 3D anatomical models [161]. A typical use case is AR-based anatomy education, which, different from conventional teaching methods, relies on virtual anatomical models [162]. In MedShapeNet, these 3D models are freely available to users and can be conveniently obtained using the online interface of MedShapeNet (to be discussed in Section 5).

In Figure 7 (A), a whole-body model is displayed using the Microsoft HoloLens AR glasses. The whole-body model can be dissembled into individual anatomies, which can be moved, zoomed in/out and rotated and in the virtual environment, allowing students to learn the shape and relative position of an anatomical structure. In this regard, the models may also be interesting for the upcoming Apple Vision Pro [163]. The shapes could even be used for Diminished Reality (DR) [164], e.g., for anatomy education [165]. Wherever necessary, these virtual models can also be converted into physical models via 3D printing. Figure 7 (B) and Figure 7 (C) show the manipulation of the heart and the kidney in the first-person and third-person views, from the perspective of a teacher. In this regard, the models may also be interesting for the upcoming Apple Vision Pro [163]. The shapes could even be used for Diminished Reality (DR) [164], e.g., for anatomy education [165]. Wherever necessary, these virtual models can also be converted into physical models via 3D printing. Figure 7 (D) and Figure 7 (E) show a 3D-printed facial phantom and a virtual skull model registered to the phantom. The virtual tumor models are also displayed on top of the registered models to show their relative position inside the skull of the patient.

4.3 Potential Negative Impact

To avoid potentially harmful societal impact, computer vision research involving human-derived data should be conducted with care. Since MedShapeNet is designed specifically for research at the junction of computer vision and medicine, proper ethics guidelines should be followed throughout methodology development and experimental design. For example, publicly sharing neuroimaging data bears high privacy risks and should be regulated, since they contain patients’ facial profiles [166], [167]. A study shows that participants who are anonymously involved in a clinical trial can be identified by matching the faces reconstructed from their head MRI scans with photographs on social media, with the help of a face recognition software [113]. Therefore, besides removing patients’ meta information before releasing neuroimaging data, defacing is also commonly practiced [168], [169]. Nevertheless, as demonstrated by the forensic facial reconstruction example described in Section 4.2.3, the facial profiles of the patients can still be reconstructed from skulls, when the entire facial structures are absent. Further removing the facial bones from skulls cannot completely resolve the issue either, as we have shown in our previous study that a machine learning model can reconstruct the original skulls even when the skulls are damaged (e.g., part of the bones on a skull are missing) [156], as seen in Figure 6 (C) and Section 4.2.3. Facial profiles can still be restored by first repairing the damaged skull using a skull reconstruction model discussed in Section 4.2.3 and then applying the facial reconstructor to the reconstructed skull, according to Figure 6 (C, A). MedShapeNet facilitates the training of face/skull reconstruction models for anyone with a basic command of machine learning, but at the same time makes it more difficult to protect patients’ privacy when it comes to sharing neuroimaging data.

Another double-edged use case of MedShapeNet is to train a machine learning model to identify drug or alcohol consumption/addiction based on facial features. Users can easily retrieve the facial models of SUD and normal cohorts from MedShapeNet and train a binary classifier on them. The application benefits early detection and intervention of SUD, but may be abused for discrimination in unauthorized situations. Furthermore, since MedShapeNet preserves the correspondence between the shapes and the source datasets, patients’ meta information, such as age, race, gender,
Fig. 8. Main panels of the MedShapeNet web interface. A, C: choosing an anatomy category 'liver'. B: selecting an anatomy instance 's12273_liver.nii.g_1.stl' and displaying it in an interactive 3D viewer. D: downloading the entire MedShapeNet database. E: an overview of currently available medical shapes, their categories and download links.

medical history, etc., if available in the source datasets, can be mapped to each shape model, which facilitates the learning of some controversial mapping relationships. For example, the ethnic identity or medical history could potentially be predicted based on a person’s skull or facial profiles by training a classifier.

It is therefore the responsibility of the researchers to weigh the social benefits against the potential negative societal impacts while developing models using MedShapeNet.

5 A WEB INTERFACE FOR MedShapeNet

A user-friendly, easy-to-use web API facilitates convenient access to the shape data within MedShapeNet, and makes it easier for researchers to use the database in their research. Inspired by the web API of the well-known ShapeNet (https://shapenet.org/), we developed a web-based interface for MedShapeNet, which can be visited at https://medshapenet-ikim.streamlit.app/. Users can search, download and inspect in a 3D viewer an individual shape, or batch download an entire category of anatomies. Desired shapes can be retrieved by the corresponding anatomy classes, such as 'heart', 'brain', 'hip', 'liver', as shown in Figure 8(A, C). The names of the shapes matching the search query will be displayed in a drop-down menu.

The corresponding shape will be displayed in a 3D viewer underneath the search box after clicking on one of the search results (Figure 8(B)). An overview of currently available medical shapes, their categories and download links is also shown on the main page of the interface (Figure 8(E)). The size of the overall database amounts to several terabytes (TB), which substantially exceeds the free space quota of most server providers, including Streamlit. We solve the problem by separating the shape storage (schiebo) from the Streamlit server running the web interface, to reduce the cost of storing large quantities of data on servers.

5.1 Search Queries

The MedShapeNet web interface returns shapes of choice by matching users’ queries with the anatomy classes provided in the names of the shape files. Table 4 shows a list of possible queries that will return at least one result in the MedShapeNet web interface. The search query for anatomies whose names contain multiple words is composed of the individual words and underscores that connect the words (e.g., atrium_left, gluteus_medius_left, inferior_vena_cava, lung_upper_lobe). No results will be displayed if the search query does not match any existing file names. Users can search by anatomy (e.g., liver) or pathology (e.g., tumor).
5.2 User Feedback

We use GitHub to manage the communication among users, developers, and contributors of MedShapeNet. It provides a mechanism for researchers to contribute shapes, provide feedbacks (e.g., report corrupted shapes, suggest improvements) and showcase their own research/applications utilizing MedShapeNet. As an incentive, shape contributors can be credited as collaborators of the MedShapeNet project and their research can be featured on the GitHub page upon request. Detailed contribution guidelines is available at https://github.com/Jianningli/medshapenet-feedback. A quality check will be performed before incorporating new shapes into MedShapeNet, to avoid introducing corrupted data and discrepancies. Since the shape data come from different sources, a consistency check will also be conducted to ensure that the shape data with the same class annotations correspond to the exact same anatomical part.

6 DISCUSSION AND CONCLUSION

High-quality, annotated datasets are valuable assets for data-driven research. We created MedShapeNet, with the firm belief that, in the near future, it will become a commonly referenced resource in the computer vision and medical imaging community. The construction of MedShapeNet is an ongoing effort and requires continuous contributions from the community, since the majority of its shape collections are acquired from data sources not owned by us. MedShapeNet also relies on the community to refine its shape collection and define more interesting use cases at the junction of computer vision and medical imaging (refer to Section 5.2).

In this white paper, we have introduced the initial efforts we have taken to construct MedShapeNet, most importantly by (1) bringing together the community for data contribution (most of the co-authors have contributed a source dataset for the shape collection); (2) deriving benchmark datasets for several interesting applications (Section 4.2), and open-sourcing them to support future research on the respective directions; (3) constructing an online interface to facilitate searching and downloading shapes of choice (Section 3); and (4) bringing up several interesting shape-related research topics that are worthy of future investigation (Section 2) and discussing the precautions that should be taken to comply with the ethics guidelines (Section 4.3). Furthermore, compared to vision datasets, large medical datasets are much more difficult to curate due to the sensitive, distributed and scarce nature of medical images. As a result, the medical imaging community has only recently started catching up with the development of vision algorithms that can exploit large datasets, with more and more medical researchers becoming open to data-sharing in recent years. Thus, MedShapeNet has the potential to bridge the gap between the vision and medical imaging community, by providing a versatile dataset that both vision and medical researchers are accustomed to. Last but not least, MedShapeNet is a freely available 3D repository for extended reality research and applications. For future development of MedShapeNet, we will primarily focus on the following aspects:

- **Increase the size and diversity of the shape collection**: we will collect more shapes, especially pathological ones (e.g., glioblastoma, aorta with aneurysm) to further enrich MedShapeNet, and engage more researchers from the community to join the initiative.
- **Promote MedShapeNet**: we will disseminate MedShapeNet more actively in the research community of computer vision and medical imaging, by presenting it in conferences, symposia, seminars and classrooms (teaching), and organizing hackweeks/workshops/challenges.
- **Define new benchmarks and establish more use cases**: we, together with the community, will derive more benchmark datasets from MedShapeNet and explore interesting use cases based on them.
- **Improve the shape search portal**: we will improve the online portal of MedShapeNet by iteratively refining the shape search functionality and improving the user interface for a better user experience.
- **Provide more shape annotations**: we will extract more meta information from the source datasets and incorporate them into the corresponding shape data as annotations.
- **Redesign the naming convention of the shapes**: we will design a more inclusive and compact naming convention for the shapes, from which essential information, such as anatomy categories, source datasets, pathologies, etc., can be deduced.

ACKNOWLEDGMENTS

This work was supported by the REACT-EU project KITE (Plattform für KI-Translation Essen, EFRE-0801977, https://kite.ikim.nrw/), FWF enFaced 2.0 (KLI 1044, https://enfaced2.ikim.nrw/), AutoImplant (https://autoimplant.ikim.nrw/) and „NUM 2.0“ (FKZ: 01KK2121). Behruz Puladi was funded by the Medical Faculty of RWTH Aachen University as part of the Clinician Scientist
Program. In addition, we acknowledge the National Natural Science Foundation of China (81971709; M-0019; 820111530141). The work of J. Chen was supported by the Bundesministerium für Bildung und Forschung (BMBF, 161101272). The work of ISAS was supported by the “Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen” and “Der Regierende Bürgermeister von Berlin, Senatskanzlei Wissenschaft und Forschung”. Furthermore, we acknowledge the Center for Virtual and Extended Reality in Medicine (ZvRM, https://zvrm.ume.de/) of the University Hospital Essen. The CT-ORG dataset was obtained from the Cancer Imaging Archive (TCIA). CT-ORG was supported in part by grants from the National Cancer Institute, U101CA190214 and U101CA187947. We thank all those who have contributed to the MedShapeNet collection (directly or indirectly).

REFERENCES


