

# AnatomyGen: Deep Anatomy Generation From Dense Representation With Applications in Mandible Synthesis

Amir H. Abdi<sup>1</sup>

Heather Borgard<sup>1</sup>

Purang Abolmaesumi<sup>1</sup>

Sidney Fels<sup>1</sup>

AMIRABDI@ECE.UBC.CA

HEATHER.BORGARD@UBC.CA

PURANG@ECE.UBC.CA

SSFELS@ECE.UBC.CA

<sup>1</sup> *Electrical and Computer Engineering Department, University of British Columbia, Canada*

## Abstract

This work is an effort in human anatomy synthesis using deep models. Here, we introduce a deterministic deep convolutional architecture to generate human anatomies represented as 3D binarized occupancy maps (voxel-grids). The shape generation process is constrained by the 3D coordinates of a small set of landmarks selected on the surface of the anatomy. The proposed learning framework is empirically tested on the mandible bone where it was able to reconstruct the anatomies from landmark coordinates with the average landmark-to-surface error of 1.42 mm. Moreover, the model was able to linearly interpolate in the  $\mathbb{Z}$ -space and smoothly morph a given 3D anatomy to another. The proposed approach can potentially be used in semi-automated segmentation with manual landmark selection as well as biomechanical modeling. Our main contribution is to demonstrate that deep convolutional architectures can generate high fidelity complex human anatomies from abstract representations.

**Keywords:** Deep generative model, 3D convolutional neural network, Shape generation, Geometric morphometrics, Shape interpolation

## 1. Introduction

Underlying dynamics of musculoskeletal systems are often studied by means of computational modeling. These models provide insights into system properties which are otherwise impossible or difficult to measure directly such as muscle fiber excitations, joint forces, and internal stresses (Hicks et al., 2015). Advancements in numerical modeling, computer graphics, and biomedical imaging, as well as increased interdisciplinary collaborations, have turned computational modeling into a blooming field of research (Andersen et al., 2017).

To study the biomechanics of a particular subject or patient, researchers create generic templates of anatomies from image-driven measurements or cadaver data. These templates are then morphed or rescaled to form subject-specific models that represent characteristics of the individual. However, this is a challenging and labour-intensive process, contingent on the availability of segmentations and the quality of imaging.

Statistical shape models (SSM) have been extensively investigated in the context of large-scale analysis of anatomical shapes and image segmentation (Zhang and Golland, 2016). This family of algorithms almost universally utilize some form of principal component analysis (PCA) and mixture models to represent variabilities in the population. While these methods have the potential to morph a statistical template to an unseen sample, they are limited or regularized by the principal

modes of variation and their respective variances. SSM approaches depend on the correspondence or coregistration of the training shapes as well as the quality of mapping to the unseen data.

With recent advances in deep learning architectures, some methods have been established in synthesizing photorealistic 2D images (Goodfellow et al., 2014; Rosca et al., 2017). With the extension of deep learning models to the realm of 3D geometries, inspiring efforts are being made to learn deterministic or probabilistic dense representations of 3D shapes (Girdhar et al., 2016; Smith and Meger, 2017; Liu et al., 2017; Tatarchenko et al., 2017) and synthesize unseen objects from learned distributions (Wu et al., 2016) or from 2D images (Choy et al., 2016). The closest work to our current approach is that of Brock et al., which uses a low-fidelity variational autoencoder for voxel-based shape modeling and a 3D convolutional classifier for object detection (Brock et al., 2016).

While synthesizing realistic shapes and images with deceptive overall perceptions is an intriguing objective in computer graphics and computer vision, the healthcare and biomedical engineering communities are more concerned about the clinical implications of the generated data. These concerns go beyond 3D modeling and are often raised in response to any generative model, including adversarial learning paradigms used to generate MR, CT (Wolterink et al., 2017), and PET data (Pan et al., 2018), or vision solutions for synthesizing realistic skin lesions (Baur et al., 2018). In other efforts, the image information is encoded in a dense latent space to regularize segmentation and super-resolution tasks (Oktay et al., 2018). Regardless, an association between the generated data and their clinical value is currently missing from the generative models in healthcare (Kazemini et al., 2018).

To this end, we propose a deterministic generative convolutional approach to synthesize 3D anatomies conditioned on clinically relevant dense representations. The performance of this method is demonstrated on the human mandible bone because of its complexities due to thin plates, and multiple concavities and processes. The trained deep model is evaluated on its ability to generate new mandible samples from unseen landmark coordinates. Moreover, we demonstrate the model’s potential in smoothly interpolating in-between shapes in the dense space. The goal here is neither to morph a template average shape, nor to segment anatomy, but to map representations of the 3D shape from an abstract dense  $\mathbb{Z}$ -space to the 3D voxel-grid space.

## 2. Method

### 2.1. Network Architecture

The generative model proposed here, which we formally refer to as  $g(z)$ , is a deep neural network which deterministically maps a dense vectorized representation,  $z$ , to a cubic tensor,  $V$ . This function can be summarized as  $g : \mathbb{Z}^l \rightarrow \mathbb{V}$ , where  $\mathbb{V}$  represents the 3D binarized voxel-grid where voxels belonging to the object are assigned the value of 1.

The highest performing network architecture that we experimented with is depicted in Figure 1. The network starts with three dense layers, with no non-linear activations, which linearly combine the elements of the input dense vector of size  $|\mathbb{Z}|$  and form a bigger feature vector of size  $64 \times |\mathbb{Z}|$ , which is subsequently reshaped into a 4D tensor of shape  $|\mathbb{Z}| \times 4 \times 4 \times 4$ .

The dense layers are followed by several transposed strided 3D convolutions (TConv3D), which are generally referred to as deconvolutions and implemented as gradient of equivalent convolutional layers. Except for the first TConv3D layer with a kernel size of 1 and stride of 1, all other TConv3D layers have a kernel size of 4 and a stride of 2 across all the 3 axes. All layers are followed by Contin-

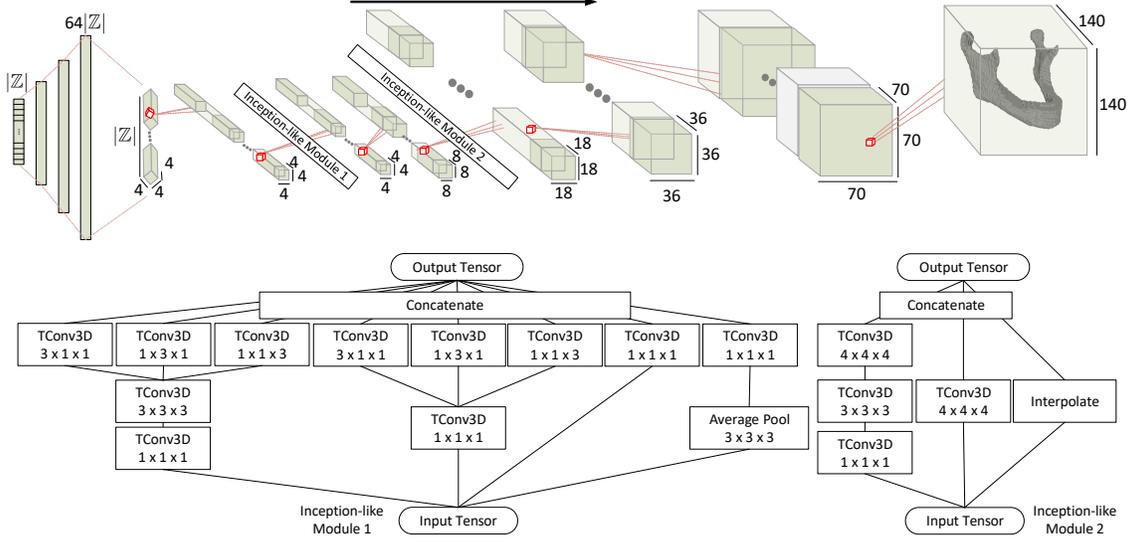


Figure 1: Architecture of the deep model for 3D shape generation from landmarks.

ously Differentiable Exponential Linear Units ( $\text{CELU}$ ) with  $\alpha = 1$ , except for the last layer which is followed by a sigmoid function. Padding is applied when necessary to achieve the target voxel-grid cube of size  $140^3$ . The last TConv3D layer only has one kernel and its input is normalized using 3D batch normalization ( $\text{BNorm3D}$ ).

Inspired by the success of the inception networks in designing deeper architectures for 2D image recognition tasks (Szegedy et al., 2016), 3D deconvolutional inception modules were designed and integrated into the AnatomyGen architecture. The inception-like blocks are depicted in Figure 1.

## 2.2. Loss Function

The generative model,  $g$ , is trained towards minimizing an approximated version of the Dice loss between the generated and expected voxel values, defined as follows

$$L_{dice}(V, \hat{V}) = \frac{2\langle V, \hat{V} \rangle_F}{\sum_{ijk} [V + \hat{V}]_{ijk}}, \quad (1)$$

where  $\langle \cdot \rangle_F$  is the Frobenius inner product of the two tensors.

## 2.3. Dense Representation

In general, the input space,  $\mathbb{Z}$ , can be any dense representation. In computer graphics, and in the context of 3D shape generation, a variety of choices have been investigated, including single or multiple 2D projections of the shape, shape silhouette, or vectors of classes or styles (Soltani et al., 2017; Firman et al., 2016).

In the context of 3D biomedical modeling, generating realistic shapes within the physiological variations of the population is valuable, yet, it is not necessarily coupled with clinical applications.

Description	Left No.	Right No.
Anterior point of the condylar process	1	16
Medial point of the condylar process	2	19
Posterior point of the condylar process	3	18
Lateral point of the condylar process	4	17
Deepest point of the mandibular notch	5	15
Peak of the coronoid process	6	14
Anterior end of the coronoid process	7	13
Mandibular foramen	8	12
Retromolar Fossa	9	11
Posterior concavity of the ramus	20	26
Mental tubercle	22	24
Angle of mandible	25	21
Anterior end of ramus	27	28
Genial tubercle on sagittal plane		10
Mental protuberance on sagittal plane		23

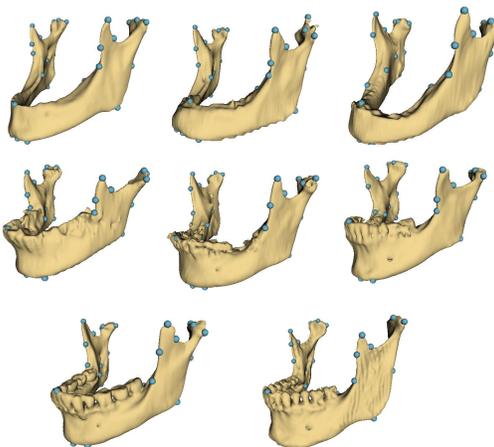


Figure 2: Some sample mandibles demonstrating the degree of variability across the data. The landmarks which constitute the input shape representation, along with their descriptions, are also visualized.

Generally speaking, there is a notion of clinical value and relevancy associated with certain regions of the anatomy. For instance, in biomechanical modeling, some regions are considered landmarks of substantial importance such as muscle and tendon insertion sites and collision surfaces. These areas play vital roles in force transmission and stress analysis; consequently, they are crucial in the validity and exactness of the resultant biomechanical model.

Based on the above intuition, and with guidance from an otolaryngology surgeon, 28 landmarks were selected on the surfaces of all samples. These landmarks were chosen based on two criteria: clinical relevancy, and reproducibility. Most of the landmarks corresponded to anatomical structures and muscle attachment sites (Figure 2). The validity of the landmarks was later checked by another dentist.

The 3D coordinates of the selected landmarks were used to form the input dense space for the generative model. As a result, the size of the input dense vector was  $|\mathbb{Z}| = 3n$ , where  $n$  is the total number of landmarks selected for each sample, in this case 28.

### 3. Data and Training

#### 3.1. Data

Our dataset includes a total of 103 mandibular meshes segmented from CT images. We collected the data from four publicly available data sources. The first subset of data comprises of 48 samples collected from the Vancouver General Hospital, Canada. Among the samples, 24 of them have teeth, while 24 mandibles are missing all or most of their dentition. The age range of the individuals was from 42 to 87. The second subset is published by Wallner et al. in figshare and consists of 10 mandibular samples (6 males, 4 females), all of whom lacked dentition (Wallner and Egger, 2018). The rest of the data were collected from the MICCAI Auto-segmentation challenge 2015 and The

Cancer Imaging Archive (TCIA) (Raudaschl et al., 2017a; Zuley et al., 2016). These mandibles were segmented as not to include the teeth and with a flat alveolar process.

### 3.2. Preprocessing

All 3D meshes were roughly aligned using the group-wise student’s-t mixture model rigid registration algorithm based on their point clouds (Ravikumar et al., 2016). We used 50 mixture components for this alignment. Each closed surface mesh was then placed in the center of a 3D cubic voxel grid of size  $140^3$ . Each voxel was determined to be either inside or outside the mesh using ray tests and assigned a value of one or zero. This resulted in a discretized voxel-based representation of the mesh with isotropic voxels of size 1 mm, imitating a segmentation mask of the mandible obtained from an isotropic CT scan. The size of the matrix was set based on the maximum facial width reported in the comprehensive dataset of the FaceBase project (Brinkley et al., 2016).

### 3.3. Training

Adam optimizer, with default momentum parameters and  $\ell_2$  regularization of  $1e-5$ , was used for training. The learning rate was initialized as  $1e-3$ , with an exponential decay rate of 0.99 applied after every epoch. Weights were uniformly initialized from a symmetric interval defined adversely proportional to the number of input channels and kernel size to keep the output of the layer bounded within reasonable limits (He et al., 2015).

Dataset was randomly partitioned into 80% training-validation set, and 20% test set. Following common practices of data augmentation in convolutional networks, each training sample was randomly mirrored, shifted, and rotated. The probability of mirroring was 50% and was limited to the sagittal plane. The rotation was limited around the vertical axis. Each sample was augmented independently and the perturbed copies were generated on-the-fly during training. We used fixed random seeds for all the experiments to mitigate the inter-experimental variance.

The proposed network architecture was implemented using the open source PyTorch library. The implementation of the learning model and the training framework is publicly shared here: <https://github.com/amir-abdi/LandmarksToShape>. To enable reproducibility of the experiments, the preprocessed voxel-based representation of the data and their associated landmarks’ coordinates accompanies the code according to each dataset’s respective license and data sharing agreements.

## 4. Evaluation and Results

As opposed to the segmentation problem where the generated masks are compared with ground-truth manual annotations, there is no single ground-truth for a given input dense representation. In other words, since the generated shape is not constrained to comply with the edges of an image, the problem is ill-posed. Therefore, there is no easy way to evaluate the performance of the shape generation process.

In this work, the input abstract representation was formed by the 3D coordinates of surface landmarks. Thus, the average distance of these landmarks to the reconstructed surface (landmark-to-surface distance; L2S) is the most relevant metric for evaluation. For a model parameterized by  $\theta$ , the L2S metric for a set of landmarks  $z$  is calculated as

$$L2S(z, \theta) = \frac{1}{n} \sum_i^n d(z_i, g_\theta(z)), \quad (2)$$

Table 1: The performance of the trained model is evaluated based on the distance of the landmarks to the generated shape (L2S). The generated shapes were also compared with the test shapes, from which the landmarks were selected, based on the Dice Coefficient (DSC), Hausdorff at 95th percentile (HD95), and Surface Mean Distance (SMD) metrics. The results are compared with the average mandible shape and the best performing mandible segmentation models in the MICCAI challenge (Raudaschl et al., 2017b).

	L2S (mm)	HD95 (mm)	SMD (mm)	DSC (%)
MICCAI (Range)	—	2.5 - 10.5	0.5 - 2.8	78 - 93
Average Shape	$4.47 \pm 3.71$	$7.09 \pm 1.83$	$2.01 \pm 0.74$	$56.43 \pm 0.67$
<b>AnatomyGen</b>	<b><math>1.42 \pm 1.05</math></b>	<b><math>3.79 \pm 0.96</math></b>	<b><math>1.19 \pm 0.29</math></b>	<b><math>74.35 \pm 7.45</math></b>

where  $d(p, s)$  is the Euclidean distance of 3D point  $p$  to the surface of shape  $s$ . The L2S error for the landmarks selected on the samples of the test set was measured at  $1.42 \pm 1.05$  mm. Except for the landmarks on the peak of the coronoid processes (Figure 2, landmarks 6 & 14) with an error of close to 5 mm, the L2S error for almost all landmarks were below 2 mm ranging from 0.65 mm to 2.21 mm, with an average of  $1.16 \pm 0.46$  mm.

The test sample from which the landmarks were selected is one of the many possible solutions for the shape generation problem. Therefore, the generated samples were also compared with their corresponding reference test sample based on the Dice Coefficient (DSC), Surface Mean Distance (SMD) and Hausdorff at the 95th percentile (HD95). Finally, to make sure that the AnatomyGen model is taking the landmark coordinates into account, all the above metrics were also calculated for the average mandible shape. The results of the mentioned analyses are reported in Table 1. Some mandibles generated from unseen input dense vectors are visualized in Figure 3 for qualitative evaluation.

We should highlight that segmentation was not among the main objectives of the proposed method; however, it can still be used for semi-automated segmentation with manual selection of landmarks. Therefore, results from the best performing models of the MICCAI 2015 challenge in mandible segmentation are included in Table 1 only to demonstrate that the shapes generated by the proposed method are within reasonable standards.

To test the generalizability of the model, we ran experiments where we smoothly morphed one 3D shape into another by linearly interpolating in the  $\mathbb{Z}$ -space. Here, a set of landmarks selected on the surface of a test sample were incrementally updated towards landmarks of another test sample. At each increment, AnatomyGen was used to generate the mandible corresponding to the landmarks' coordinates. In Figure 4, the two test samples (left and right) and results of their linear interpolation are visualized.

## 5. Discussion and Conclusion

In this work, a deep 3D convolutional approach is proposed to generate anatomical shapes from dense vectorized representations. In our experiments, the 3D coordinates of a selected set of landmarks on the surface of the shape were used as the abstract representation of the shape. Thanks to the randomized and regularized training framework, the deep model was not only able to gen-



Figure 3: Mandibles generated using the AnatomyGen model from unseen vectors.

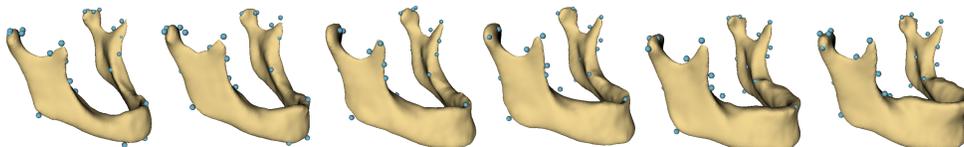


Figure 4: Smooth linear interpolation between two extreme cases of the test set.

eralize to unseen test samples and generate realistically looking samples but also demonstrated its potential in interpolating in-between shapes. The average distances between input landmarks and the generated surfaces (L2S) were close to 1 mm, which is similar to the resolution of the voxel grid.

The only exceptions to the high-performing L2S metric were the landmarks on the peak of the coronoid processes (Figure 2). Our hypothesis is that the coronoid processes were so thin that the errors in their reconstruction did not contribute much to the Dice loss. As a result, the model was not highly penalized for inaccurate reconstruction of this thin structure. Including the L2S metric in the loss function would have mitigated this error; however, it would have limited the generalizability of the proposed method as we had aimed for the input to be any abstract representation and not be limited to landmarks.

The dataset used in this study were formed by merging four data sources; thus, they were diversified with respect to their anatomical variations, the presence of teeth, and the amount of bone loss.

Moreover, the segmentations were carried out by various experts on images of different imaging systems which resulted in the occasional presence of holes and gaps in a subset of samples due to their imprecise segmentations.

No information beyond the input dense vector is provided to the AnatomyGen model. As a result, the proposed shape generation process is expected to be a one-to-many mapping between the encoded dense representation and the 3D voxel-based shape. Consequently, if the trained model generates a shape from a set of landmarks selected on the surface of a given shape, the two anatomies will not necessarily be identical. However, during model evaluation, the samples of the test set, from which the landmarks originated, were considered as one of the many acceptable potential reconstructions. With this assumption in mind, we compared the generated mandibles with their corresponding sample of the test set (Table 1).

A limitation of this study lies in its fully supervised training without encoding the posterior distribution; thus, there is no principled way for this model to generate random samples. However, AnatomyGen demonstrated acceptable interpolation potential in incrementally morphing a given shape to another (Figure 4). As a result, a possible solution for random shape synthesis is to model the statistical coordinates of the landmarks as a multivariate distribution and sample from to generate random mandibles.

A potential application of the AnatomyGen model is in semi-automated segmentation. In this approach, the expert manually selects a small number of predefined landmarks on the 3D medical image and the deep model generates the corresponding segmentation map. This approach is modality agnostic as no image features are included in the process.

The approach introduced here differs from statistical shape models as it does not morph a model across modes of variation, but generates an anatomy from an abstract representation. It is also different from segmentation methods as the generative model is blind to the image features. Basically, the proposed method is a step towards generating biomedical models from abstract representations.

Our next step is to investigate shape generation and reconstruction from other representative forms, such as partial shapes, where the missing parts of the anatomy are completed using a deep generative model. This approach is helpful in surgical planning where the normal pre-morbid mandibular form is unknown.

## Acknowledgments

We would like to thank Dr. Eitan Prisman from the Vancouver General Hospital for his support throughout this research. This research was undertaken, in part, thanks to the funding from the Vanier Scholar award of the Natural Sciences and Engineering Research Council of Canada (NSERC) to the first author, Amir H. Abdi.

## References

- Michael Skipper Andersen, Mark De Zee, Michael Damsgaard, Daniel Nolte, and John Rasmussen. Introduction to Force-Dependent Kinematics: Theory and Application to Mandible Modeling. *Journal of Biomechanical Engineering*, 2017.
- Christoph Baur, Shadi Albarqouni, and Nassir Navab. Generating highly realistic images of skin lesions with GANs. In *Lecture Notes in Computer Science*, pages 260–267. Springer International Publishing, 2018.

- James F Brinkley, Shannon Fisher, Matthew P Harris, Greg Holmes, Joan E Hooper, Ethylin Wang Jabs, Kenneth L Jones, Carl Kesselman, Ophir D Klein, Richard L Maas, Mary L Marazita, Licia Selleri, Richard A Spritz, Harm van Bakel, Axel Visel, Trevor J Williams, Joanna Wysocka, FaceBase Consortium, and Yang Chai. The FaceBase Consortium: a comprehensive resource for craniofacial researchers. *Development*, 143(14):2677–88, 2016. ISSN 1477-9129. URL <https://www.facebase.org>.
- Andrew Brock, Theodore Lim, J. M. Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. In *Neural Information Processing Conference*, 2016.
- Christopher B. Choy, Danfei Xu, Jun Young Gwak, Kevin Chen, and Silvio Savarese. 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. In *Lecture Notes in Computer Science*, volume 9912 LNCS, pages 628–644, 2016.
- Michael Firman, Oisín Mac Aodha, Simon Julier, and Gabriel J. Brostow. Structured Prediction of Unobserved Voxels from a Single Depth Image. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5431–5440, 2016.
- Rohit Girdhar, David F Fouhey, Mikel Rodriguez, and Abhinav Gupta. Learning a predictable and generative vector representation for objects. In *Lecture Notes in Computer Science*, volume 9910 LNCS, pages 484–499, 2016.
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2015.
- Jennifer L. Hicks, Thomas K. Uchida, Ajay Seth, Apoorva Rajagopal, and Scott L. Delp. Is My Model Good Enough? Best Practices for Verification and Validation of Musculoskeletal Models and Simulations of Movement. *Journal of Biomechanical Engineering*, 137(2):020905, 2015.
- Salome Kazemini, Christoph Baur, Arjan Kuijper, Bram van Ginneken, Nassir Navab, Shadi Albarqouni, and Anirban Mukhopadhyay. Gans for medical image analysis. *Computing Research Repository (CoRR)*, abs/1809.06222, 2018.
- Jerry Liu, Fisher Yu, and Thomas Funkhouser. Interactive 3D Modeling with a Generative Adversarial Network. In *International Conference on 3D Vision (3DV 2017)*, pages 126–134, 2017.
- Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A. Cook, Antonio de Marvao, Timothy Dawes, Declan P. Regan, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2): 384–395, 2018.

- Yongsheng Pan, Mingxia Liu, Chunfeng Lian, Tao Zhou, Yong Xia, and Dinggang Shen. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer’s disease diagnosis. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2018)*, volume 11072 LNCS, pages 455–463. Springer, Cham, 2018.
- Patrik F. Raudaschl, Paolo Zaffino, Gregory C. Sharp, Maria Francesca Spadea, Antong Chen, Benoit M. Dawant, Thomas Albrecht, Tobias Gass, Christoph Langguth, Marcel Lüthi, Florian Jung, Oliver Knapp, Stefan Wesarg, Richard Mannion-Haworth, Mike Bowes, Annaliese Ashman, Gwenael Guillard, Alan Brett, Graham Vincent, Mauricio Orbes-Arteaga, David Cárdenas-Peña, German Castellanos-Dominguez, Nava Aghdasi, Yangming Li, Angelique Berens, Kris Moe, Blake Hannaford, Rainer Schubert, and Karl D. Fritscher. Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. *Medical Physics*, 44(5): 2020–2036, 2017a.
- Patrik F. Raudaschl, Paolo Zaffino, Gregory C. Sharp, Maria Francesca Spadea, Antong Chen, Benoit M. Dawant, Thomas Albrecht, Tobias Gass, Christoph Langguth, Marcel Lüthi, Florian Jung, Oliver Knapp, Stefan Wesarg, Richard Mannion-Haworth, Mike Bowes, Annaliese Ashman, Gwenael Guillard, Alan Brett, Graham Vincent, Mauricio Orbes-Arteaga, David Cárdenas-Peña, German Castellanos-Dominguez, Nava Aghdasi, Yangming Li, Angelique Berens, Kris Moe, Blake Hannaford, Rainer Schubert, and Karl D. Fritscher. Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. *Medical Physics*, 44(5): 2020–2036, 2017b. ISSN 00942405.
- Nishant Ravikumar, Ali Gooya, Serkan Çimen, Alejandro F. Frangi, and Zeike A. Taylor. A multi-resolution T-mixture model approach to robust group-wise alignment of shapes. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016)*, volume 9902 LNCS, pages 142–149. Springer, Cham, 2016.
- Mihaela Rosca, Balaji Lakshminarayanan, David Warde-Farley, and Shakir Mohamed. Variational approaches for auto-encoding generative adversarial networks. *Computing Research Repository (CoRR)*, abs/1706.04987, 2017.
- Edward J. Smith and David Meger. Improved adversarial systems for 3d object generation and reconstruction. In Sergey Levine, Vincent Vanhoucke, and Ken Goldberg, editors, *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pages 87–96. PMLR, 2017.
- Amir Arsalan Soltani, Haibin Huang, Jiajun Wu, Tejas D. Kulkarni, and Joshua B. Tenenbaum. Synthesizing 3D Shapes via Modeling Multi-view Depth Maps and Silhouettes with Deep Generative Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2511–2519. IEEE, 2017.
- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.
- Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE, October 2017.

- Jürgen Wallner and Jan Egger. Mandibular CT dataset collection. figshare, 2018.
- Jelmer M Wolterink, Peter R Seevinck, and Anna M Dinkla. MR-to-CT Synthesis using Cycle-Consistent Generative Adversarial Networks. In *Med-NIPS*, 2017.
- Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenenbaum. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. *Graphical Models and Image Processing (CVGIP)*, 53(2):157–185, 2016.
- Miaomiao Zhang and Polina Golland. Statistical shape analysis: From landmarks to diffeomorphisms. *Medical Image Analysis*, 33:155–158, 2016.
- Margarita L. Zuley, Rose Jarosz, Shanah Kirk, Yueh Lee, Rivka Colen, Kimberly Garcia, Dominique Delbeke, Michelle Pham, Paul Nagy, Gorkem Sevinc, Marla Goldsmith, Subair Khan, Jose M. Net, Fabiano R. Lucchesi, and Natalia D. Aredes. Radiology data from the cancer genome atlas head-neck squamous cell carcinoma collection. The Cancer Imaging Archive, 2016.