Computers in Biology and Medicine 123 (2020) 103886

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Skull shape reconstruction using cascaded convolutional networks

Oldřich Kodym *, Michal Španěl, Adam Herout

Department of Computer Graphics and Multimedia, Brno University of Technology, Božetěchova 2, 612 66 Brno, Czech Republic

ARTICLE INFO

Keywords:
Cranial implant design
Anatomical reconstruction
3D shape completion
Convolutional neural networks
Generative adversarial networks

ABSTRACT

Designing a cranial implant to restore the protective and aesthetic function of the patient's skull is a challenging process that requires a substantial amount of manual work, even for an experienced clinician. While computer-assisted approaches with various levels of required user interaction exist to aid this process, they are usually only validated on either a single type of simple synthetic defect or a very limited sample of real defects. The work presented in this paper aims to address two challenges: (i) design a fully automatic 3D shape reconstruction method that can address diverse shapes of real skull defects in various stages of healing and (ii) to provide an open dataset for optimization and validation of anatomical reconstruction methods on a set of synthetically broken skull shapes.

We propose an application of the multi-scale cascade architecture of convolutional neural networks to the reconstruction task. Such an architecture is able to tackle the issue of trade-off between the output resolution and the receptive field of the model imposed by GPU memory limitations. Furthermore, we experiment with both generative and discriminative models and study their behavior during the task of anatomical reconstruction.

The proposed method achieves an average surface error of 0.59 mm for our synthetic test dataset with as low as 0.48 mm for unilateral defects of parietal and temporal bone, matching state-of-the-art performance while being completely automatic. We also show that the model trained on our synthetic dataset is able to reconstruct real patient defects.

1. Introduction

Patient-specific implants (PSIs) are often used for the treatment of cranio-facial defects. Especially in cases of larger defects caused by trauma, tumour resection or decompressive craniectomy, it is usually required to reconstruct the original skull shape for aesthetic purposes and protection of intracranial structures against mechanical impact [1,2].

Current state-of-the-art methods usually comprise of using a patient CT scan to design the implant pre-operatively and then 3D printing of the result using bio-compatible materials such as titanium, porous polyethylene or polyether ether ketone [3,4]. Alternatively, implants can be cast in a 3D-printed mold from bone cement which can be loaded with antibiotics to decrease the risk of infection [5]. Such approaches lead to a reduction of operative time and improved patient results [6].

Provided that precise enough tissue segmentation is obtained from the CT data, the process of computer-aided design (CAD) of PSIs remains the most important step that affects the final quality and reproducibility of PSIs [7]. This presents a challenging and tedious task for the clinician or engineer designing the PSI. To ensure correct healing and prevent complications, the PSI must fit precisely to the defect border without any steps and with gaps of less than 0.8 mm between the implant and remaining tissue [8]. Furthermore, smoothness and symmetry of the anatomy should be preserved to ensure a correct aesthetic result [9]. The first step of the implant design is correct reconstruction of the missing shape of the skull anatomy from which the PSI is then derived.

1.1. Related work

A considerable number of CAD systems for skull reconstruction make heavy use of the natural facial symmetry by identifying the best symmetry plane and then mirroring the healthy part of the skull onto the defect area [10]. Recent automation of parts of this process led to an efficient and user-friendly way to provide an aesthetically correct result. Such methods, however, cannot account for bilateral defects reaching into both sides of the skull. Also, because the symmetry is usually not perfect in real cases, manual corrections are often needed to produce a correct reconstruction. Another group of methods is based on surface interpolation under the assumption of a nearly spherical shape

* Corresponding author.
E-mail address: ikodym@fit.vutbr.cz (O. Kodym).

https://doi.org/10.1016/j.compbiomed.2020.103886
Received 16 April 2020; Received in revised form 18 June 2020; Accepted 25 June 2020
Available online 27 June 2020

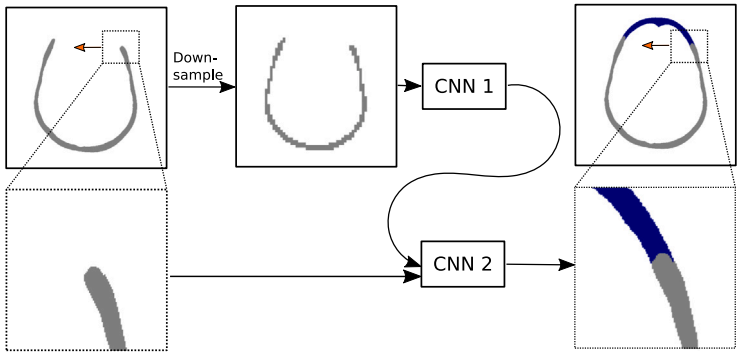

Fig. 1. The proposed skull reconstruction method fully automatically produces a binary shape of the missing part of the skull using sliding-window approach with coarse-resolution middle step.

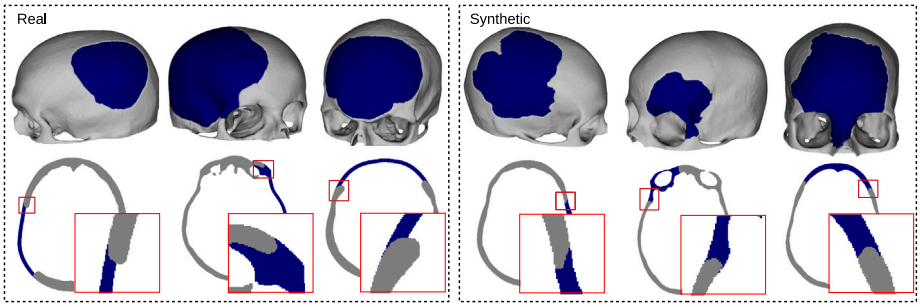

Fig. 2. Examples of 3D models and slices through defective skulls from real patients (left) and synthetically generated defective skulls (right). The real patient samples include defects reconstructed by an experienced clinician.

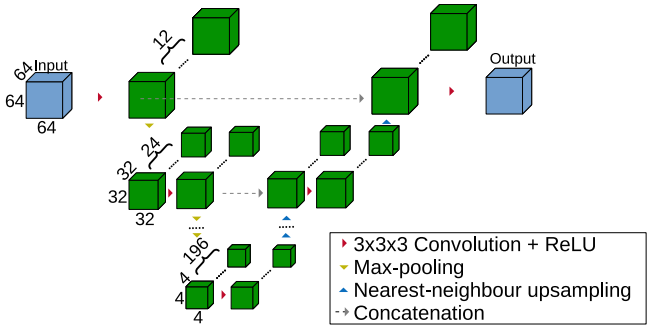

Fig. 3. Overview of the 3D CNN backbone architecture example for an input of size 64^3. Note that for an input of size 128^3, the CNN is deeper and the bottleneck tensor has 384 channels.

of the cranium [11,12]. Interpolation-based methods can guarantee desired continuity on the tissue-implant interface and also make it possible to modify the resulting fit by adjusting the parameters. Although these methods work well on smaller defects, they often struggle with larger defects because of a lack of constraints and they cannot model more complex anatomy shapes such as orbitals.

Current state-of-the-art methods usually exploit some kind of deformable models [13]. Statistical shape models combined with geometric morphometrics have been studied extensively in the context of skull reconstruction [14–16], achieving an average surface error of 0.47 mm for defects of the parietal and temporal area and 0.75 mm for small mid-facial defects, as measured on simple synthetically created defects against the original bone shape. These methods, while providing a good reconstruction result, rely on a clean, well-defined defect border, which is rarely the case in real patient cases with complex fractures in various stages of healing and bone resorption, as can be seen in the example slice in Fig. 1.

Another interesting approach has recently made use of convolutional denoising auto-encoders in the first attempt to use a deep learning-based shape completion for skull reconstruction [17], although only operating on a very coarse resolution. More details on the topic of automatic skull shape reconstruction can be found in the work of Buonamici et al. [18].

Using deep learning for a general 3D shape reconstruction (also shape inpainting, shape completion) is a well studied research topic in the literature. The basic approach is to represent the incomplete input shape as a binary voxel grid and train a 3D convolutional neural network (CNN) with a deterministic denoising auto-encoder architecture to output the completed binary shape [19]. These approaches cannot be utilized for bigger volumes due to GPU memory limitations. To tackle this, some authors exploit different data representations such as graphs or point clouds [20,21]. Another group of authors use the 3D CNN only for coarse shape estimation, refining the result in the post-processing step [22–24].

An orthogonal research direction in this area led to substituting the purely discriminative CNN models with generative models such as generative adversarial networks (GANs) or variational auto-encoders [20,23,24], suggesting that the shape completion task actually has multiple correct solutions conditioned on a single input. This issue of one-to-many mapping has also been raised by authors in the context of anatomical shape reconstruction [25,26] where inter-expert variability of the resulting shape is also taken into account. However, the argument that the variability of the output should be enforced at the cost of precision measured against the original shape is in a conflict with the current literature on skull reconstruction where the original shape is considered to be the ground-truth.

1.2. Contributions

In this work, we design a cascaded CNN architecture for the estimation of a high-resolution 3D anatomy shape conditioned on the input defective skull. Although symmetry is used in the proposed

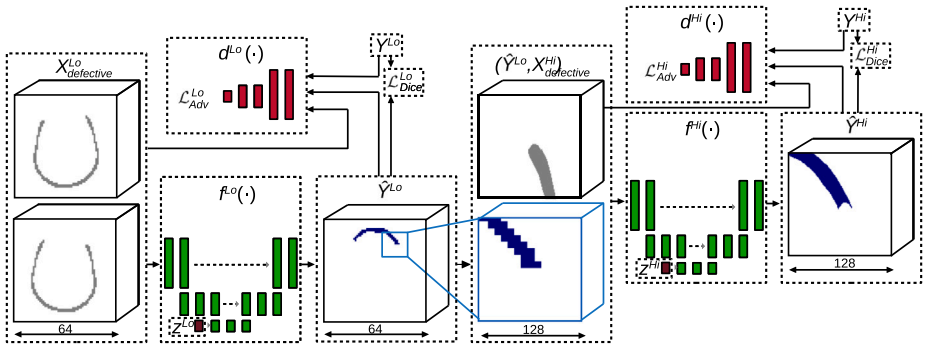

Fig. 4. Overview of the proposed 3D CNN cascade. The symmetrized low-resolution input X^{Lo} is fed into the first model f^{Lo} to produce the missing shape estimation $\hat{Y}^{Lo}$. Then, it is concatenated to the high-resolution input X^{Hi} and fed into the second model f^{Hi} to produce the final high-resolution missing shape $\hat{Y}^{Hi}$. Discriminator CNNs $d(\cdot)$ and latent vectors z are only used in the generative model.

method as an additional guiding signal, the method can successfully reconstruct defects reaching into both sides of the skull as well as into more complex anatomical regions such as orbitals. To the authors' best knowledge, this is the first deep learning-based method of 3D shape reconstruction that reaches a high enough resolution to be clinically viable for the skull reconstruction task. To address the issue of multiple possibly correct solutions, we also experiment with a probabilistic generative version of the proposed model. Finally, in order to improve the reproducibility of research in the area of automatic skull reconstruction, we introduce an open dataset of skull shapes with synthetic defects. The dataset mimics the variability in shape, position and bone resorption present among real patients and we show that a CNN model trained on this synthetic dataset also performs well on challenging real patient cases without any further pre-processing.

2. Materials

For training and validation of the proposed method, we used a public CQ500 dataset [27] as a source of head CT patient data. A total of 189 of the scans were rigidly aligned and the skulls were segmented and saved as 3D binary arrays of 512^3 voxels. Finally, 5 different defects were created on each segmented skull with an emphasis on simulating the variability in real defective patient skulls. We simulate the defects by subtracting randomly deformed combinations of spheres followed by morphologically rounding the defect edges to account for various genesis and healing processes of real defects. To allow for structured validation, the defects were categorized into *unilateral parietal*, *unilateral frontal* (the orbital area) and *bilateral* groups. In addition to these three groups, two more random defects were generated on each skull. The resulting dataset of 945 defective skulls with ground-truth original shapes along with further details is publicly available as the SkullBreak dataset[1] and we refer interested readers there for further information regarding the details of defective skulls generation process.

The synthetic dataset was split into 179 training and 10 testing skulls, resulting in 895 training and 50 testing defect shapes in total. To evaluate the ability of our approach to generalize, we also utilized an internal dataset of 9 real defective patients. For these patients, ground-truth skull reconstructions made by a clinician experienced in cranial implant design were available. Several samples from both datasets are shown in Fig. 2.

3. Methods

We formulate the skull reconstruction task as finding the missing part of the anatomy represented by binary volume $Y = X_{healthy} - X_{defective}$. Thus, we look for the function $f(\cdot)$ with parameters θ that maps the defective skull to an estimated shape $\hat{Y} = f_\theta(X_{defective})$ from distribution $P(\hat{Y}|X_{defective})$ of shapes that correctly complete it.

[1] https://www.fit.vut.cz/person/ikodym/skullbreak.

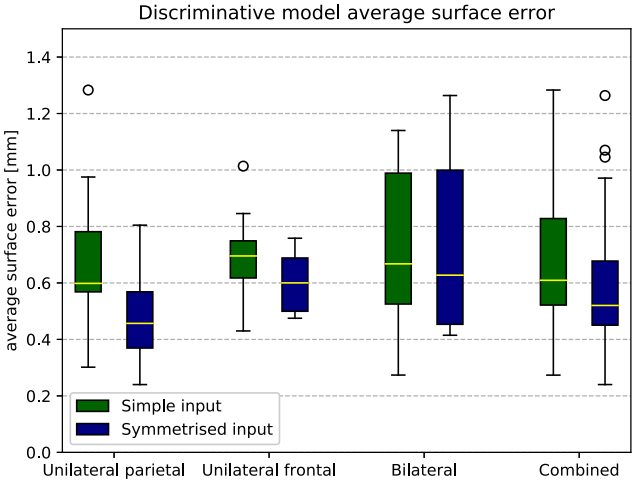

Fig. 5. The overall performance of the discriminative model on different groups of testing defective skulls. Average surface error [mm] for a simple input (green) and symmetrized input (blue).

3.1. Reconstruction model architecture

We use a combination of two CNN models with a 3D U-net [28] backbone to approximate the function f, with parameters θ being the trainable weights of the CNN. The individual models differ from the original 3D U-net in several ways. Instead of up-convolutions, we use nearest-neighbor up-sampling followed by regular convolution in the decoder part of the model, as this has been shown to improve the model training process and performance in some cases [29]. The number of down-sampling and up-sampling layers is such that the bottleneck tensor has spatial dimensions of 4^3 as shown in Fig. 3. This ensures that the output neurons of the CNN have a sufficient receptive field to correctly model the shape of missing anatomy in the case of defects with a large surface area.

Each of the models operates on a different resolution. The first model, denoted f^{Lo}, takes an input volume down-sampled to 64^3 voxels and is trained to output an estimate of missing anatomy $\hat{Y}^{Lo} = f^{Lo}(X^{Lo}_{defective})$ on an equivalent resolution of 3.2 mm per voxel. While this resolution is too low to model anatomy with enough precision, it can provide an initial estimate of the missing shape. The second model, denoted f^{Hi}, then takes a $128 \times 128 \times 128$ crop of the input data at the original high resolution concatenated to an up-sampled output of the first model. This model is trained to output the corresponding patch of the final missing anatomy estimate $\hat{Y}^{Hi} = f^{Hi}(\hat{Y}^{Lo}, X^{Hi}_{defective})$, which can be viewed as a super-resolution of the

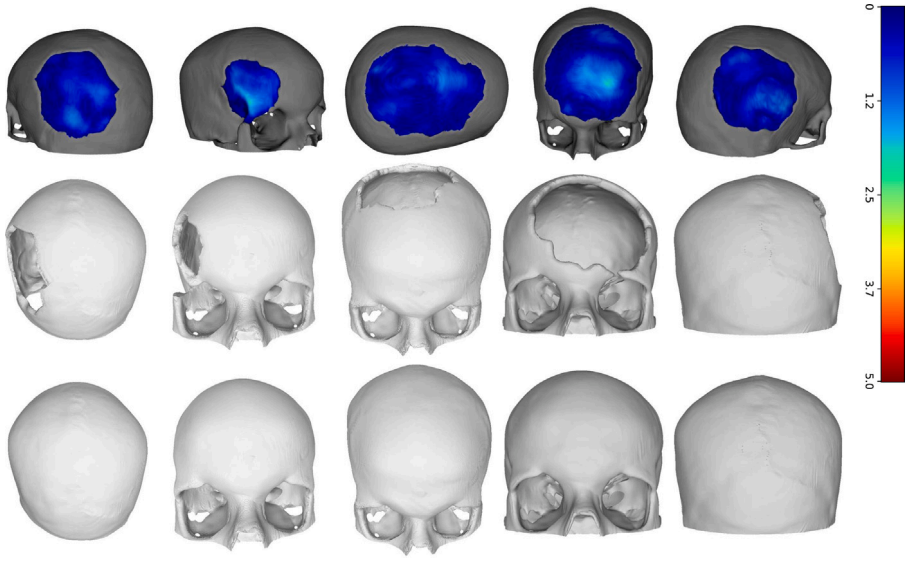

Fig. 6. Examples of results of the discriminative model reconstructions for parietal, frontal, bilateral and random defects, respectively. From top to bottom: Surface error maps, input synthetically broken skulls and reconstructed skull shapes. Note that the majority of the reconstructed surface reaches errors of less than one millimeter in all cases.

initial missing anatomy estimate conditioned on the remaining part of the skull at full resolution.

During inference, the first model provides enough contextual information about the overall shape of the defective skull while the second model can ensure precise contact at the defect border. The final estimate can be inferred by first computing the coarse estimate $\hat{Y}^{Lo}$ and then computing the final estimate $\hat{Y}^{Hi}$ using the sliding window approach, substantially reducing the memory footprint.

Symmetrized input. The chosen U-net architecture in the low-resolution CNN is in fact not well suited for transferring information from one side of the volume to the opposing side as this transfer can only happen in the deeper layers of the model where the shape information is already compressed. However, the ability to preserve the anatomical symmetry is a critical part of the method. To this end, we concatenated a sagitally flipped copy of the volume to the low-resolution CNN input. This makes it easier to propagate the symmetry information using convolutional kernels and skip connections of the U-net architecture. The effect of symmetrizing input is demonstrated in Section 4.

3.2. Optimization

We optimize the CNNs using training batches of size 2, which fully utilize the available GPU. An Adam optimizer is used as it is currently one of the most widely used optimization algorithms suitable for most deep learning applications [30]. Although we train both CNNs with their respective loss functions $\mathcal{L}^{Lo}$ and $\mathcal{L}^{Hi}$, we train the cascade in an end-to-end manner. The training samples for the first model $(Y^{Lo}, X^{Lo}_{defective})$ and the random training crops for the second model $(Y^{Hi}, X^{Hi}_{defective})$ are always sampled from the same skull volume.

Discriminative model. We first assume that the skull reconstruction task has a single correct ground-truth solution given by the original missing anatomy shape Y. This allows us to use a reconstruction loss similar to a segmentation task. We chose the soft Dice loss [31] due to its good performance in dealing with class imbalance. The two losses are defined as

$$\mathcal{L}^{Lo}_{Dice} = Dice(Y^{Lo}, f^{Lo}(X^{Lo}_{defective})), \quad (1)$$

$$\mathcal{L}^{Hi}_{Dice} = Dice\left(Y^{Hi}, f^{Hi}\left(f^{Lo}(X^{Lo}_{defective}), X^{Hi}_{defective}\right)\right), \quad (2)$$

and we optimize them iteratively for 300 000 training steps. While it is possible to optimize the whole cascade using only the $\mathcal{L}^{Hi}$ loss, we found that using the auxiliary loss $\mathcal{L}^{Lo}$ is necessary for correct model behavior.

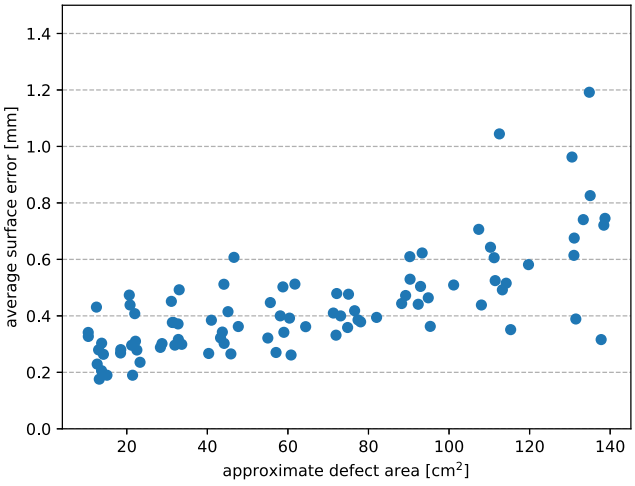

Fig. 7. The performance of the discriminative model in the context of a reconstructed surface area. While the variability of the model output errors increases with larger defects, the results are within an acceptable range even for a majority of the larger defects.

Generative model. To make the described reconstruction model generative, we make two modifications well known from GAN literature [32] to both CNNs. Namely, we add the adversarial loss function $\mathcal{L}_{Adv}$ in the form of a discriminator CNN $d(\cdot)$, which allows the model to learn the distribution $P(\hat{Y}|X_{defective})$, and inject a random latent vector into the reconstruction CNNs, which allows them to randomly sample from this distribution. We concatenate the random latent vector with the bottleneck tensor of both CNNs as shown in Fig. 4. The discriminator CNNs have the same architecture as the encoder part of the reconstruction CNNs with additional dense layers that output the discriminator scores. We use the improved Wasserstein GAN formulation with gradient penalty [33] during the training. Given a combination of the defective skull shape and the missing anatomy shape, the discriminator is trained to assign a low score $d(Y, X_{defective})$ to the ground-truth missing shape and a high score to the reconstructed missing shape $d(\hat{Y}, X_{defective})$ at both a low and high resolution, using the low-resolution discriminator d^{Lo} and high-resolution discriminator d^{Hi}. To optimize the reconstruction CNNs in this case, we use a combination of the reconstruction

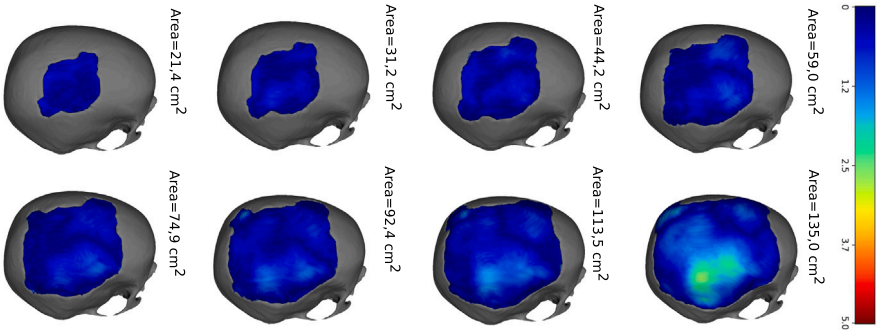

Fig. 8. Examples of results of the discriminative model reconstructions for a set of defects with different scales on a single test skull. An area where the model output deviates from the original shape by more than 2 mm can be observed in the last case.

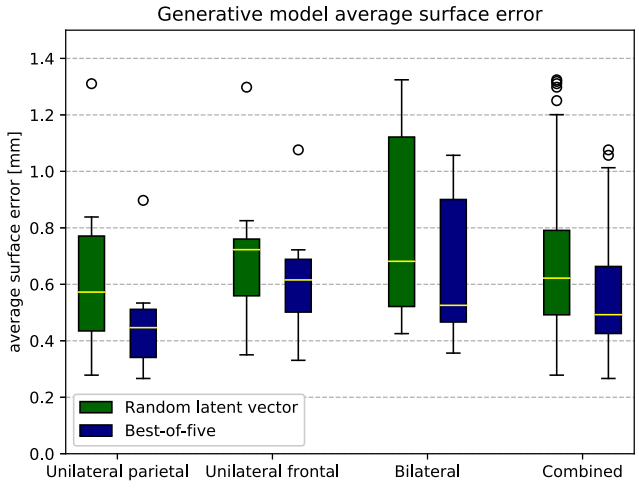

Fig. 9. The overall performance of the generative model on different groups of testing defective skulls. Average surface error [mm] for reconstruction with random latent vector (green) and for best-of-five reconstructions (blue).

and adversarial loss, similarly to Wang et al. [24]. The losses of the reconstruction CNNs in this case are defined as

$$\mathcal{L}^{Lo} = \mathcal{L}^{Lo}_{Dice} + \alpha \mathcal{L}^{Lo}_{Adv}, \tag{3}$$

$$\mathcal{L}^{Hi} = \mathcal{L}^{Hi}_{Dice} + \alpha \mathcal{L}^{Hi}_{Adv}, \tag{4}$$

where α is set to 10^{-2}. We again optimize d^{Lo}, f^{Lo}, d^{Hi} and f^{Hi} iteratively for 300 000 training steps. For an overview of our method and both discriminative and generative models, see Fig. 4.

4. Experimental results

The experiments discussed in this section were run on a system with 11 GB Titan Xp GPU and a quad-core i5 processor with 24 GB RAM. The complete training of the models took approximately 8 days. After the model is trained, the method is able to fully reconstruct each skull in under 5 s, which is important for its efficient use in clinical practice. This is achieved by first inferring the low-resolution model on the full down-sampled volume and then sequentially inferring the high-resolution model on positions where the low-resolution model predicted a defect until the whole estimated defect area is processed. For visualization, the voxel grid was converted into a polygonal mesh which was then smoothed using a two-step smoothing algorithm [34].

We measured the precision of each method as the average symmetric unsigned distance between the surface voxels of the output reconstruction and the original anatomy shape which we considered to be ground-truth. We only measured the error on the outer surface of the skull because the inner surface is not relevant for cranial implant design in the clinical practice. To get more insight into performance of our method, we divided the evaluation into four groups. The *unilateral-parietal*, *unilateral-frontal* and *bilateral* defect groups are described in Section 2 and the *combined* group includes all the defects, including random ones. Table 1 contains the average surface error for all models tested on each defect group.

Table 1
Average surface error [mm] for individual defect groups.

Method	Synthetic defects				Real defects
	UP	UF	Bi	Total	
Statistical shape models [15]	0.47	–	–	–	–
Discriminative + simple input	0.69	0.69	0.78	0.68	–
Discriminative + symmetrized input	0.48	**0.60**	0.73	0.59	0.80
Generative + symmetrized input (random)	0.63	0.71	0.81	0.68	–
Generative + symmetrized input (best of 5)	**0.46**	0.62	**0.65**	**0.56**	**0.69**

We first evaluated the performance of the discriminative model and the effect of the symmetrized input on the error distribution in the testing set. The model with a simple input was able to reconstruct each testing skull successfully. However, we noticed that the errors in unilateral groups reached similar values as the bilateral group. This is in conflict with the expectation that while bilateral defects could allow for some variability in correctly completed shapes, the unilateral defects should be more directly constrained by the condition of symmetry and thus yield lower surface reconstruction errors. The effect of symmetrizing input as described in Section 3 was that the average measured error of the reconstructed unilateral defects dropped from 0.69 mm to 0.48 mm for parietal and from 0.69 mm to 0.60 mm for frontal defects. As expected, the bilateral defects group was less affected by the symmetrized input, although the error still slightly decreased since some bilateral defects are in fact partly constrained by the symmetry. The overall performance of the discriminative model for both simple and symmetrized input is shown in Fig. 5. The overall average surface error of the discriminative model with a symmetrized input for the whole testing set was 0.59 ± 0.21 mm. Several examples of the discriminative model reconstructions are shown in Fig. 6.

In order to explore the relation between the discriminative model performance and the area of the reconstructed defects, we created an extra set of nine cranial defects in each of the ten designated test skulls. The defects were created by subtracting the same shape with different scales from each skull (see Fig. 8 for their illustration). The surface area of the resulting skull defects ranged from 10 to 140 cm^2. The resulting surface errors of the discriminative model outputs are shown in Fig. 7 in the form of a scatter plot. While there is an apparent correlation between the measured surface error and the reconstructed surface area, the average surface error was under 0.7 mm for all defects up to an area of 100 cm^2. For even larger defects, the average surface error exceeded 1 mm in several cases. However, for majority of the cases, the surface errors of the results were still well under this value.

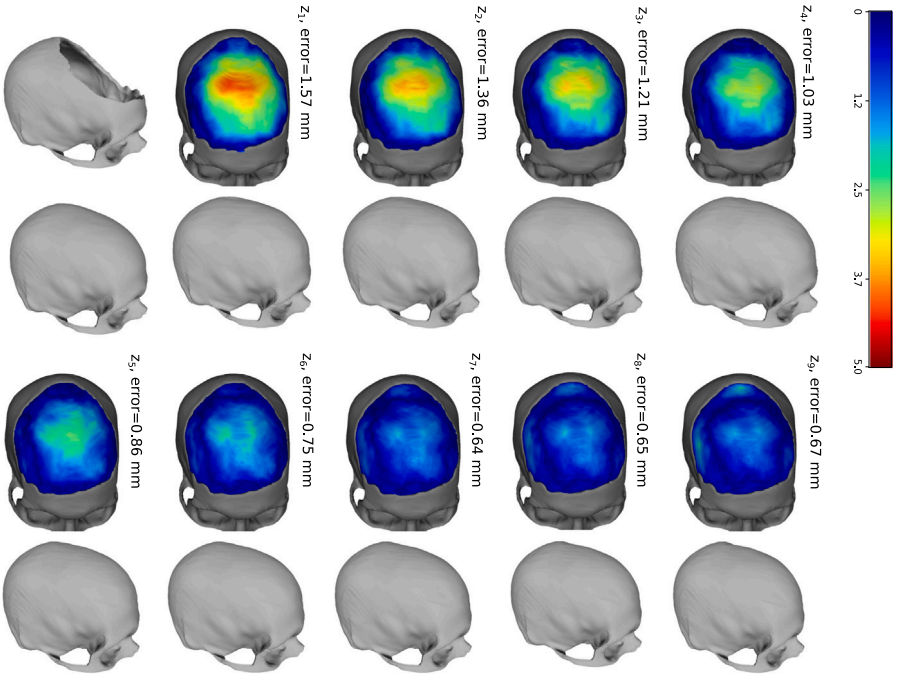

Fig. 10. The input defective skull (first row, left) and original skull shape (second row, left) and examples of output reconstructions resulting from linear interpolation in the latent space superimposed onto the input defective skull. Surface distance from the original shape can be seen decreasing in the middle part of the defect, however, it increases in areas near the defect border where errors are unacceptable for aesthetic reasons.

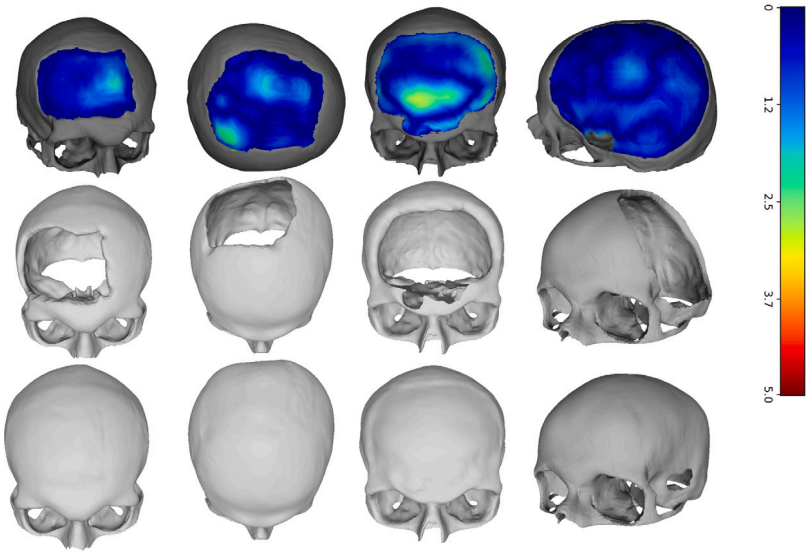

Fig. 11. Example outputs of the discriminative model for real patient data. Although some reconstruction faults can be seen in the last two cases, suggesting that real training data of target population should be added to the model in the future, the reconstruction is usually correct. The surface distance to the original shape is well below one millimeter on average.

Next, we evaluated the performance of the generative model with symmetrized input volumes and random input latent vectors z. The overall average surface error was 0.68 ± 0.28 mm. For each defect group, the error of the generative model was higher than that of the discriminative model. However, it should be noted that since now we consider multiple correct reconstructions for a single skull defect, the error measured against the ground-truth shape might not be a good indicator of the method's performance. The generative model allows us to sample multiple different outputs for a single input defective skull by changing the input latent vectors. Therefore, we also experimented with generating multiple reconstructions and measuring the best achieved result. The overall average surface error when measuring the best-of-five sampled reconstructions for each testing skull was 0.56 ± 0.21 mm. The results for individual defect groups, as seen in Fig. 9, were similar to the discriminative model in this case. However, a reduction of the error can be noticed in the *bilateral* group, with the error reduced from 0.73 mm to 0.65 mm when compared to the discriminative model. This might once again be explained by the fact that due to weaker symmetry constraints in this group, the variability of acceptable reconstructions is greater. Therefore, generating subsequent different samples constrained on the same input increases the probability of generating at least one sample close to the original ground-truth shape.

To further illustrate the behavior of the generative model, we also conducted an experiment with latent space interpolation for one bilateral defect. We set both latent vectors for low-resolution z^{Lo} and high-resolution z^{Hi} to only contain constant values $c/10$ and we generated samples for $c = 1, 2 \dots 9$. Our experiments showed that the generative model responds to these changes in total latent vector energy the most and we leave investigation into the limits of achievable anatomical variability in the output for future work. The resulting reconstructions of the generative model along with the ground-truth original shape are shown in Fig. 10. We also reported the measured surface errors against the ground-truth shape for each sample. It can be seen that the model is able to sample from the learned manifold of solutions, allowing for manipulating the reconstructed shape while still keeping a seamless connection to the original bone.

Finally, in order to evaluate the ability of our approach to generalize, we also tested the performance of the models trained exclusively on our synthetic dataset on an internal dataset of real defective patients without any fine-tuning of the model. Both models reconstructed the real defects mostly successfully. However, there was an expected increase in the surface error in both the discriminative model output and the best-of-five generative model output. In some cases, there were also visible faults such as slight depressions or even holes as seen in Fig. 11.

This could be partly attributed to the fact that the real testing patients come from a different geographic location, in which the anatomical variability of the skull is different [35]. Specifically, the differences in average shapes of the two datasets aligned using the same alignment method are illustrated in Fig. 13. The fact that these basic shape characteristics are learned by the low-resolution reconstruction model may lead to wrong estimation of the cranial volume in frontal part of the skull and even holes in parts which extend significantly beyond the anatomical variability observed in the training dataset. Overall, the outcome of this experiment is encouraging, although real defective patient scans from the target population should be added to the training process before evaluating the method performance in a real clinical setting.

5. Discussion

For deployment of reconstruction methods into the clinical workflow, several conditions must be met. First, symmetry of the skull should be preserved as well as possible, including in cases where the patient's skull itself is partly asymmetric and where the defect reaches partly into both sides of the skull. Second, the automatic reconstruction should fit very precisely to the defect borders. Although the models presented in this work will occasionally produce slightly asymmetric result or fail to avoid some depressions around the defect border, our results show that the proposed method can achieve an overall satisfactory performance in this regard, as illustrated by example reconstructions in Figs. 6 and 11. The measured average surface errors shown in Figs. 5 and 7 also show how the performance is affected by different shapes and sizes of the defects, including bilateral defects, orbital area reconstructions and defects with surface area of over $100\,\mathrm{cm}^2$. The implications of these results for the future implementation of the method into clinical practice should now be assessed by clinicians with experience in this area.

In the context of the current state of the art in the area of skull reconstruction, our approach differs from conventional mirroring-based and interpolation-based methods by its ability to reconstruct an arbitrary part of the skull present in the described dataset without requiring any parameter adjusting. Its ability to generalize to unseen skulls is, however, fully dependent on the variability of the training dataset used for model optimization. Fig. 12 demonstrates how using the model on a population where shapes of the skulls come from different distribution causes occasional faults and a slight increase in the average surface error of the reconstructions. Nevertheless, this issue will be mitigated

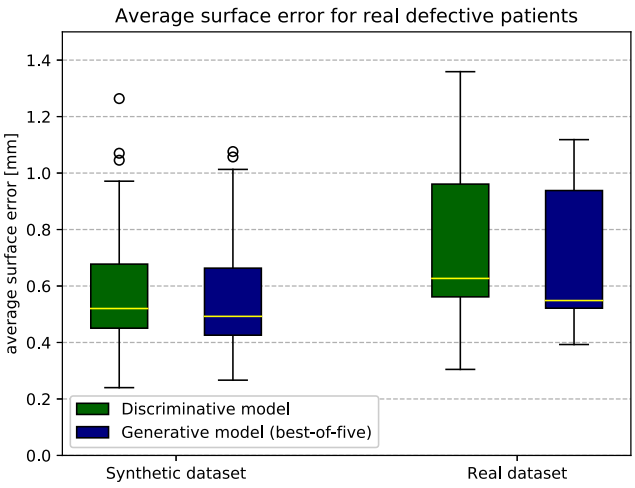

Fig. 12. Comparison of performance of the models' average surface error [mm] for reconstruction of synthetic defects and real defects. Results shown for discriminative model (green) and for best-of-five outputs of generative model (blue).

by introducing cases from the target population into the dataset and retraining the reconstruction model in the near future.

Methods based on statistical shape models also possess this dependency on training dataset variability and the potential ability to model any part of the skull. This makes them very similar to the approach proposed in this work in terms of possible target use cases. Fuessinger et al.[15] achieved an average surface error of $0.47\,\mathrm{mm}$ when reconstructing unilateral spherical defects of the cranial area with a radius of $5\,\mathrm{cm}$. This could be compared to the performance of our discriminative and generative models reaching 0.48 and $0.46\,\mathrm{mm}$ average surface error, respectively, on the unilateral parietal defect group. In contrast, our method does not require any manual cleaning of the defect border as the seamless fit of the reconstructed part to the rest of the skull is handled by the CNN model. It would be interesting to see the performance of the statistical shape model on more challenging parts of the introduced dataset such as defects of the orbital area and larger bilateral defects.

A more general comparison is currently limited by the lack of standardized datasets and methodology to evaluate the anatomical reconstruction methods. Especially in the case of bilateral defects in which symmetry cannot be used to uniquely define the correct output, we argue that although the absolute distance from the ground-truth shape might give an adequate estimate of how well a method performs, it should not be used as the single criterion of correct reconstruction. In addition to variability in cranium shape, modeling structures such as skull protuberances, sutures, or uneven surface is unnecessary for means of PSI design. Therefore, the most relevant metric to measure the reconstruction method performance would be the amount of time required by the operating expert to design clinically acceptable PSI from the initial reconstruction. However, this is infeasible without the method being deployed into clinical practice.

Since the discriminative model outputs reach lower average surface error than the randomly sampled outputs of the generative model, it can be concluded that it is more suitable for a completely automatic setting. However, the generative model could alternatively be used in a semi-automatic setting. In case the initial reconstruction is not satisfactory for further processing, several subsequent samples from the generative model could be offered to the expert to increase the chance of avoiding falling back to a less efficient conventional workflow.

Finally, the reconstruction method is not limited to skull reconstruction task or anatomical reconstruction in general. The method can potentially be applied to any shape completion task where both global contextual information as well as fine structural details need to be taken into account during the data volume reconstruction.

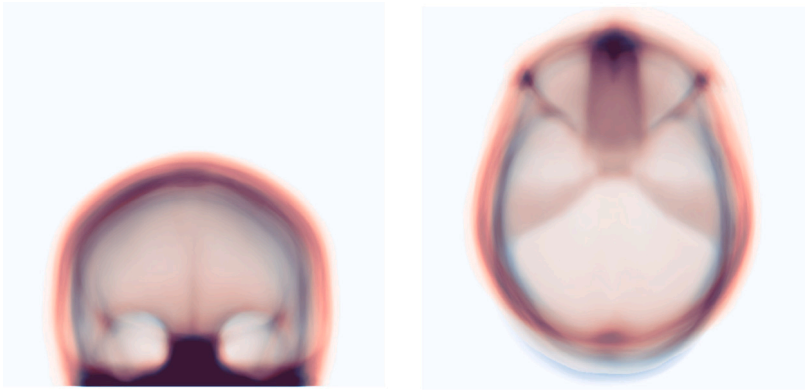

Fig. 13. Superimposed frontal (left) and axial (right) projections of the segmented skulls. The 10 testing cases of the synthetic dataset rendered in blue and the 7 testing cases of the internal dataset in red. A difference in several shape characteristics of the skulls can be observed.

6. Conclusions

This work presented a multi-scale cascaded CNN architecture for general shape completion applied to the reconstruction of missing skull anatomy in a fully automatic manner. We also showed that symmetrized input can increase the performance in this task and that both discriminative and generative models can be used successfully. The proposed method reaches enough precision and robustness to be considered in clinical practice. Validation was done on a synthetic dataset which closely mimics real patient cases and this dataset was made public.

The model trained exclusively on synthetic data also performs well on real defective patient cases, but adding samples from the real target population to the training should be considered in order to improve the results. Further testing with more patient data in clinical setting is now required to fully confirm its efficacy and identify any limitations.

Currently, the method is constrained to the cranial and orbital area. We plan to extend the method and the dataset to include maxilla and zygomatic bones as well in the future. Adapting the method for alternative data representations, such as point clouds or graphs, could also be explored as a way to improve processing speed and precision.

Declaration of competing interest

Michal Španěl is also a CEO of Tescan 3DIM company at the time of writing this paper.

Acknowledgments

This work was supported by The Ministry of Education, Youth and Sports of Czech Republic from the Large Infrastructures for Research, Experimental Development and Innovations project “IT4Innovations”. Part of the dataset was provided by TESCAN Medical and TESCAN 3DIM companies in the form of anonymized CT scans. We also gratefully acknowledge the support of the NVIDIA Corporation with the donation of one NVIDIA TITAN Xp GPU for this research.

References

[1] D.B. Kurland, A. Khaladj-Ghom, J.A. Stokum, B. Carusillo, J.K. Karimy, V. Gerzanich, J. Sahuquillo, J.M. Simard, Complications associated with decompressive craniectomy: A systematic review, Neurocrit. Care 23 (2) (2015) 292–304, http://dx.doi.org/10.1007/s12028-015-0144-7.

[2] M.-Y. Lee, C.-C. Chang, C.-C. Lin, L.-J. Lo, Y.-R. Chen, Custom implant design for patients with cranial defects, IEEE Eng. Med. Biol. Mag. 21 (2) (2002) 38–44, http://dx.doi.org/10.1109/MEMB.2002.1000184.

[3] A.L. Jardini, M.A. Larosa, R.M. Filho, C.A. de Carvalho Zavaglia, L.F. Bernardes, C.S. Lambert, D.R. Calderoni, P. Kharmandayan, Cranial reconstruction: 3D biomodel and custom-built implant created using additive manufacturing, J. Cranio-Maxillofac. Surg. 42 (8) (2014) 1877–1884, http://dx.doi.org/10.1016/j.jcms.2014.07.006.

[4] J. hyeon Oh, Recent advances in the reconstruction of cranio-maxillofacial defects using computer-aided design/computer-aided manufacturing, Maxillofac. Plast. Reconstr. Surg. 40 (1) (2018) http://dx.doi.org/10.1186/s40902-018-0141-9.

[5] H. van de Belt, D. Neut, W. Schenk, J.R. van Horn, H.C. van der Mei, H.J. Busscher, Infection of orthopedic implants and the use of antibiotic-loaded bone cements: A review, Acta Orthop. Scand. 72 (6) (2001) 557–571, http://dx.doi.org/10.1080/000164701317268978.

[6] K. Rudman, C. Hoekzema, J. Rhee, Computer-assisted innovations in craniofacial surgery, Facial Plast. Surg. 27 (04) (2011) 358–365, http://dx.doi.org/10.1055/s-0031-1283054.

[7] M. van Eijnatten, R. van Dijk, J. Dobbe, G. Streekstra, J. Koivisto, J. Wolff, CT image segmentation methods for bone used in medical additive manufacturing, Med. Eng. Phys. 51 (2018) 6–16, http://dx.doi.org/10.1016/j.medengphy.2017.10.008.

[8] R. Marsell, T.A. Einhorn, The biology of fracture healing, Injury 42 (6) (2011) 551–555, http://dx.doi.org/10.1016/j.injury.2011.03.031, Bone Regeneration in the 21st Century.

[9] M. Bashour, History and current concepts in the analysis of facial attractiveness, Plast. Reconstr. Surg. 118 (3) (2006) 741–756, http://dx.doi.org/10.1097/01.prs.0000233051.61512.65.

[10] X. Chen, L. Xu, X. Li, J. Egger, Computer-aided implant design for the restoration of cranial defects, Sci. Rep. 7 (1) (2017) http://dx.doi.org/10.1038/s41598-017-04454-6.

[11] Y.-W. Chen, C.-T. Shih, C.-Y. Cheng, Y.-C. Lin, The development of skull prosthesis through active contour model, J. Med. Syst. 41 (10) (2017) http://dx.doi.org/10.1007/s10916-017-0808-2.

[12] Y. Volpe, R. Furferi, L. Governi, F. Uccheddu, M. Carfagni, F. Mussa, M. Scagnet, L. Genitori, Surgery of complex craniofacial defects: A single-step AM-based methodology, Comput. Methods Programs Biomed. 165 (2018) 225–233, http://dx.doi.org/10.1016/j.cmpb.2018.09.002.

[13] A. Marzola, M. Servi, Y. Volpe, A reliable procedure for the construction of a statistical shape model of the cranial vault, in: C. Rizzi, A.O. Andrisano, F. Leali, F. Gherardini, F. Pini, A. Vergnano (Eds.), Design Tools and Methods in Industrial Engineering, Springer International Publishing, Cham, 2020, pp. 788–800.

[14] W. Semper-Hogg, M.A. Fuessinger, S. Schwarz, E. Ellis, C.-P. Cornelius, F. Probst, M.C. Metzger, S. Schlager, Virtual reconstruction of midface defects using statistical shape models, J. Cranio-Maxillofac. Surg. 45 (4) (2017) 461–466, http://dx.doi.org/10.1016/j.jcms.2016.12.020.

[15] M.A. Fuessinger, S. Schwarz, C.-P. Cornelius, M.C. Metzger, E. Ellis, F. Probst, W. Semper-Hogg, M. Gass, S. Schlager, Planning of skull reconstruction based on a statistical shape model combined with geometric morphometrics, Int. J. Comput. Assist. Radiol. Surg. 13 (4) (2017) 519–529, http://dx.doi.org/10.1007/s11548-017-1674-6.

[16] M.A. Fuessinger, S. Schwarz, J. Neubauer, C.-P. Cornelius, M. Gass, P. Poxleitner, R. Zimmerer, M.C. Metzger, S. Schlager, Virtual reconstruction of bilateral midfacial defects by using statistical shape modeling, J. Cranio-Maxillofac. Surg. 47 (7) (2019) 1054–1059, http://dx.doi.org/10.1016/j.jcms.2019.03.027.

[17] A. Morais, J. Egger, V. Alves, Automated computer-aided design of cranial implants using a deep volumetric convolutional denoising autoencoder, in: Advances in Intelligent Systems and Computing, Springer International Publishing, 2019, pp. 151–160, http://dx.doi.org/10.1007/978-3-030-16187-3_15.

[18] F. Buonamici, R. Furferi, L. Genitori, L. Governi, A. Marzola, F. Mussa, Y. Volpe, Reverse engineering techniques for virtual reconstruction of defective skulls: an overview of existing approaches, in: Proceedings of CAD2018, CAD Solutions LLC, 2018, http://dx.doi.org/10.14733/cadconfp.2018.6-10.

[19] A. Sharma, O. Grau, M. Fritz, VConv-DAE: Deep volumetric shape learning without object labels, in: G. Hua, H. Jégou (Eds.), Computer Vision – ECCV 2016 Workshops, Springer International Publishing, Cham, 2016, pp. 236–250.

[20] D. Stutz, A. Geiger, Learning 3D shape completion from laser scan data with weak supervision, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, http://dx.doi.org/10.1109/cvpr.2018.00209.

[21] W. Wang, D. Ceylan, R. Mech, U. Neumann, 3DN: 3D deformation network, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2019, http://dx.doi.org/10.1109/cvpr.2019.00113.

[22] A. Dai, C.R. Qi, M. NieBner, Shape completion using 3D-encoder-predictor CNNs and shape synthesis, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2017, http://dx.doi.org/10.1109/cvpr.2017.693.

[23] O. Litany, A. Bronstein, M. Bronstein, A. Makadia, Deformable shape completion with graph convolutional autoencoders, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, http://dx.doi.org/10.1109/cvpr.2018.00202.

[24] W. Wang, Q. Huang, S. You, C. Yang, U. Neumann, Shape inpainting using 3D generative adversarial network and recurrent convolutional networks, in: 2017 IEEE International Conference on Computer Vision, ICCV, IEEE, 2017, http://dx.doi.org/10.1109/iccv.2017.252.

[25] A.H. Abdi, H. Borgard, P. Abolmaesumi, S. Fels, AnatomyGen: Deep anatomy generation from dense representation with applications in mandible synthesis, in: International Conference on Medical Imaging with Deep Learning, London, United Kingdom, 2019.

[26] A.H. Abdi, M. Pesteie, E. Prisman, P. Abolmaesumi, S. Fels, Variational Shape Completion for Virtual Planning of Jaw Reconstructive Surgery, in: Lecture Notes in Computer Science, Springer International Publishing, 2019, pp. 227–235, http://dx.doi.org/10.1007/978-3-030-32254-0_26.

[27] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N.G. Campeau, V.K. Venugopal, V. Mahajan, P. Rao, P. Warier, Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study, Lancet 392 (10162) (2018) 2388–2396, http://dx.doi.org/10.1016/s0140-6736(18)31645-3.

[28] Özgün. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention, MICCAI 2016, Springer International Publishing, 2016, pp. 424–432, http://dx.doi.org/10.1007/978-3-319-46723-8_49.

[29] M. Kolarik, R. Burget, K. Riha, Upsampling algorithms for autoencoder segmentation neural networks: A comparison study, in: 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT, IEEE, 2019, http://dx.doi.org/10.1109/icumt48472.2019.8970918.

[30] S. Sun, Z. Cao, H. Zhu, J. Zhao, A survey of optimization methods from a machine learning perspective, IEEE Trans. Cybern. (2020) 1–14, http://dx.doi.org/10.1109/tcyb.2019.2950779.

[31] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, IEEE, 2016, http://dx.doi.org/10.1109/3dv.2016.79.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 27, Curran Associates, Inc., 2014, pp. 2672–2680.

[33] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of Wasserstein GANs, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017, pp. 5767–5777.

[34] A. Belyaev, Y. Ohtake, A comparison of mesh smoothing methods, in: Israel-Korea Bi-National Conference on Geometric Modeling and Computer Graphics, Tel Aviv University, 2003, pp. 83–87.

[35] Y.S.N. Jayaratne, R.A. Zwahlen, A systematic review of interethnic variability in facial dimensions, Plast. Reconstr. Surg. 129 (1) (2012) 164e–165e, http://dx.doi.org/10.1097/prs.0b013e3182362e3f.