

Automatic Cranial Defect Reconstruction with Self-Supervised Deep Deformable Masked Autoencoders

Marek Wodzinski^{1,2}, Daria Hemmerling¹, Mateusz Daniol¹

Abstract—Thousands of people suffer from cranial injuries every year. They require personalized implants that need to be designed and manufactured before the reconstruction surgery. The manual design is expensive and time-consuming leading to searching for algorithms whose goal is to automatize the process. The problem can be formulated as volumetric shape completion and solved by deep neural networks dedicated to supervised image segmentation. However, such an approach requires annotating the ground-truth defects which is costly and time-consuming. Usually, the process is replaced with synthetic defect generation. However, even the synthetic ground-truth generation is time-consuming and limits the data heterogeneity, thus the deep models' generalizability. In our work, we propose an alternative and simple approach to use a self-supervised masked autoencoder to solve the problem. This approach by design increases the heterogeneity of the training set and can be seen as a form of data augmentation. We compare the proposed method with several state-of-the-art deep neural networks and show both the quantitative and qualitative improvement on the SkullBreak and SkullFix datasets. The proposed method can be used to efficiently reconstruct the cranial defects in real time.

Index Terms—Deep Learning, Cranial Implant, Shape Completion, Skull Defect, Masked Autoencoder, Self-Supervised Learning

I. INTRODUCTION

Cranial injuries are a common result of neurosurgery, traffic accidents, or warfare. Thousands of people suffer from such injuries every year and require personalized implants [1]. The process of modeling and manufacturing such implants requires expertise, is time-consuming, and leads to substantial costs and waiting time. However, the process can be at least partially automated by deep learning algorithms [2], [3].

Nowadays, the most common approach to automatic cranial defect reconstruction is to treat it as a volumetric segmentation that can be solved by deep neural networks [2], [3]. Numerous

The project was funded by The National Centre for Research and Development, Poland under Lider Grant no: LIDER13/0038/2022 (DeepImplant). We gratefully acknowledge Polish HPC infrastructure PLGrid support within computational grant no. PLG/2023/016239. The research was partially supported by the program "Excellence Initiative - Research University" for AGH University.

¹Marek Wodzinski, Daria Hemmerling, and Mateusz Daniol are with the AGH University of Krakow, Department of Measurement and Electronics, Krakow, Poland.

²Marek Wodzinski is also with the University of Applied Sciences Western Switzerland (HES-SO Valais), Institute of Informatics, Sierre, Switzerland.

© 2024 IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

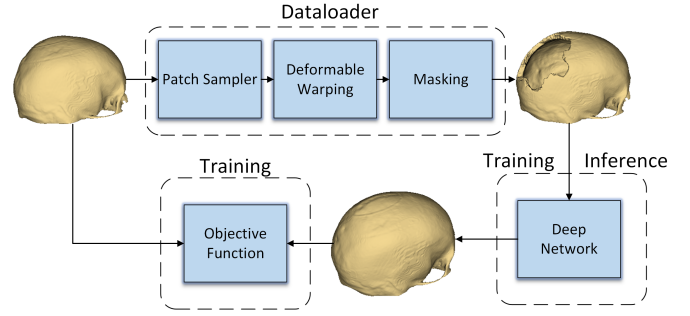


Fig. 1. The pipeline of the proposed deep deformable masked autoencoder. Firstly, The healthy skull is loaded, then masking patches are randomly selected, deformably transformed, and used for masking to produce the defective skull. Then, a dedicated encoder-decoder network is used to perform the reconstruction.

works present the usefulness of segmentation networks in this context, varying from the use of simple convolutional architectures [4]–[6], to more advanced methods involving image registration-based or generative augmentation [7]–[9]. Some of the works attempt to reformulate the problem and to solve it in other domains, e.g. as a point-cloud completion task [10], [11]. Nevertheless, the major problem that all the approaches attempt to solve is to increase the heterogeneity of the training set leading to the improvement of the model generalizability. The challenge comes from the fact that it is extremely costly and time-consuming to acquire and annotate cases with real cranial defects. Therefore, the broadly used open datasets like SkullBreak, SkullFix, or MUG500 consist mostly of real skulls with synthetic defects suffering from relatively large homogeneity and difficulties with generalization into real cases [12], [13]. Therefore, it would be beneficial to propose a method that automatically generates heterogeneous ground-truth defects. The masked autoencoders (MAEs) seem to be perfect for this task.

The idea behind the masked autoencoders is to randomly delete part of the input and then train an encoder-decoder network to reconstruct the missing data. The task is considerably more difficult than training the classical autoencoders to just recover its input. The self-supervised pretrained masked autoencoders are then useful for other downstream tasks because they learn both general and detailed features associated with the data. Since masked autoencoders do not require the ground-truth annotations, they can be seen as a

powerful self-supervised pretraining tool [14]–[16].

Contribution: In this work, we propose an alternative approach to the automatic cranial defect reconstruction by using a deep deformable masked autoencoder. We prove its usability and compare it to the state-of-the-art deep architectures dedicated to volumetric segmentation. We confirm that the implicit heterogeneity introduced by the masked autoencoder training improves the automatic cranial defect reconstruction.

II. METHODS

A. Deformable Masked Autoencoder

In this work, we propose an approach based on a deep masked autoencoder. During training, for each case, we generate a random number of patches with variable sizes that are going to mask out part of the healthy skulls, thus generating defective ones. The generated patches are randomly transformed by deformable elastic deformation. It results in smooth shapes that resemble real cranial defects. Such patches are then used to generate the ground truth for the self-supervision.

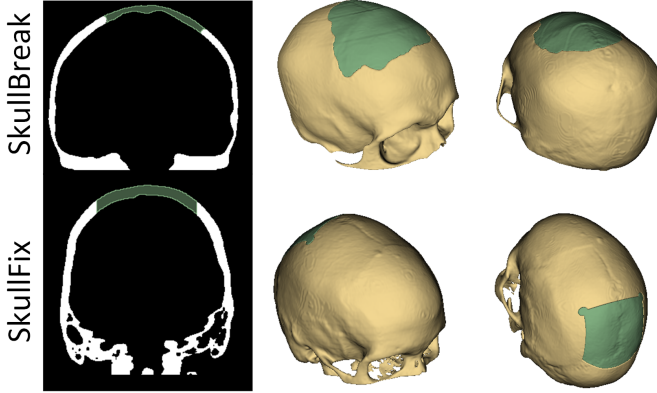


Fig. 2. Exemplary cases from the SkullFix and SkullBreak datasets. Note the sharp boundaries in the SkullFix dataset, and the smooth, irregular ones in the SkullBreak dataset.

The defective inputs are then passed to the masked autoencoder network. The goal of the network is to reconstruct the input shape from the defective ones. This way, we slightly reformulate the problem from the missing shape segmentation into learning the overall shape of skulls. Since the patches are generated, inserted, and deformed randomly, this approach strongly increases the data heterogeneity, thus leading to improvements in the model’s generalizability. In this work, we use a Residual 3-D UNet as the autoencoder backbone [6]. The processing pipeline is shown in Figure 1.

Since the processed data is binary, we reconstruct the patches using the Soft Dice Score as the loss function. The initial ablations confirmed that such an approach is more stable and converges faster than experiments using mean absolute or mean squared differences. Other widely used objective functions to train masked autoencoders in the computer vision domain are not useful in the discussed context since we are processing directly binary data.

B. Datasets and Experimental Setup

We train the proposed method on the healthy skulls available in the SkullFix and SkullBreak datasets [12]. The training part of the SkullFix and SkullBreak datasets consists of 100 and 114 skulls respectively, resulting in 214 training cases. The volumetric segmentation networks used for comparison are trained using the combined datasets with known ground truth, resulting in a training set consisting of 570 defective cases. Exemplary cases from the SkullFix and SkullBreak datasets are shown in Figure 2.

The proposed method, as well as the methods used for comparison, are evaluated using the SkullBreak and Skull Fix test sets consisting of 100 and 110 skulls with defects respectively. We compare the proposed method to several state-of-the-art segmentation architectures: (i) the Residual UNet [17], [18], (ii) the UNETR [19], (iii) the SwinUNETR [20], (iv) the Attention UNet [21], (v) the SegResNet [22], all openly available in the MONAI library [23]. We also perform ablation studies related to the influence of random deformable transformations applied during the input masking.

All the experiments are implemented in PyTorch with the support of TorchIO library [24]. The input cases are centered, cropped to the skull, and resampled to 256x256x256 voxels. The evaluation metrics were calculated on the skull defects using original resolution after upsampling the inference output, without further postprocessing. The skull defects were calculated by morphological operations. We evaluate the methods using the Dice Coefficient (DSC), Boundary Dice Coefficient (BDSC), and 95th percentile of Hausdorff distance (HD95), following the conventions introduced by the authors of the SkullFix and SkullBreak datasets during the AutoImplant challenge [2], [3]. All the networks were trained until convergence using a computing cluster with NVIDIA A100 40GB GPUs. The Soft Dice Loss was used as the objective function, the AdamW as the optimizer with learning rate/weight decay equal to 0.001 and 0.01 respectively, and with decreasing the learning rate by exponentially decaying scheduler with the ratio equal to 0.995.

III. RESULTS AND DISCUSSION

The visual comparison of the methods on a randomly chosen case from the test set is shown in Figure 3. The comparison presenting the performance concerning DSC, BDSC, and HD95 are presented in Figure 4. The results are summarized in Table I. The figure compare the proposed method to the state-of-the-art medical segmentation architectures.

The results confirm that the proposed method outperforms the state-of-the-art solutions by a considerable margin. The trained model improves the reconstruction quality with respect to all the quantitative metrics. Moreover, the results show that the random elastic deformations of the generated patches are crucial to improve the performance on the SkullBreak dataset. In case of sharp patches, the network is unable to learn smooth boundaries that are crucial for an accurate defect reconstruction. On the other hand, the effect is less significant

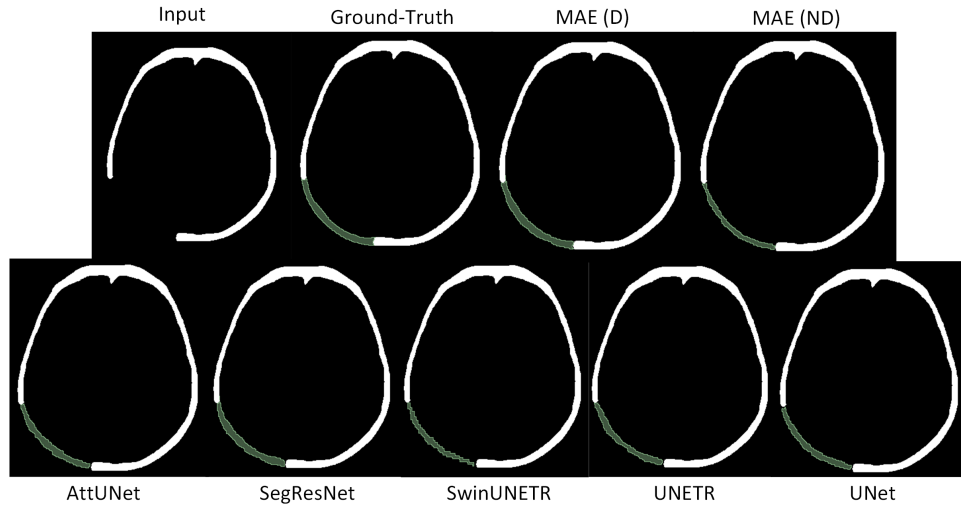


Fig. 3. Exemplary visual comparison of the proposed deformable MAE to the volumetric segmentation methods.

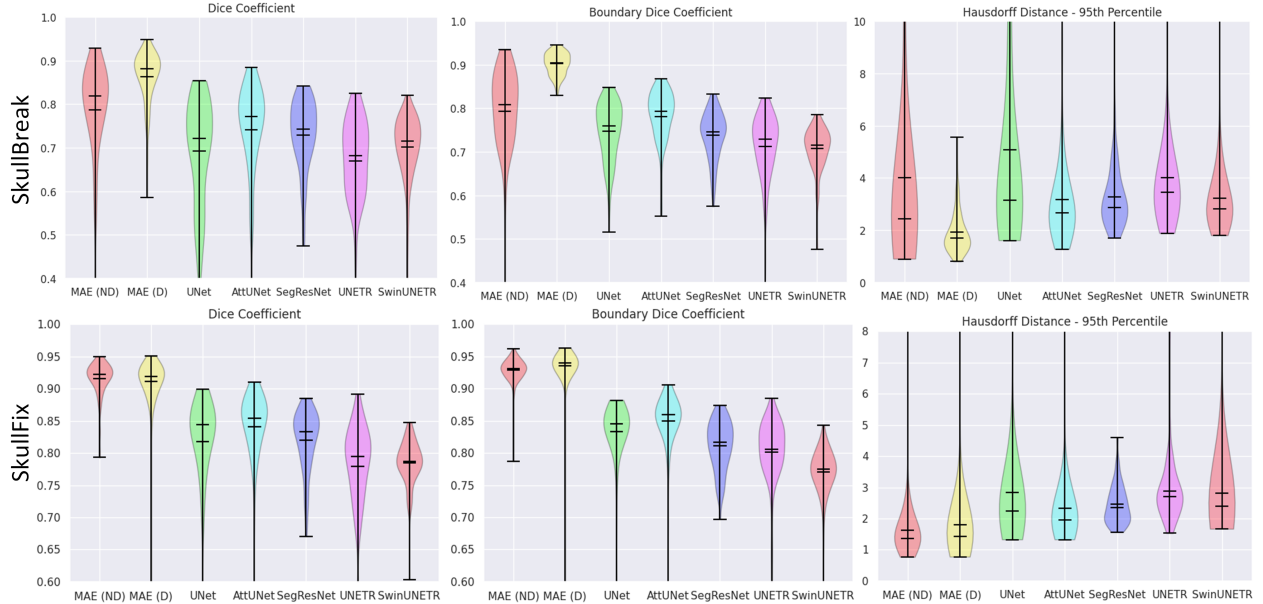


Fig. 4. The qualitative results of the proposed approach compared to the state-of-the-art segmentation networks. The "D" denotes the proposed approach with the deformable elastic deformations and "ND" is the masking with sharp edges. Note the significant improvement of the proposed approach for the SkullBreak dataset.

TABLE I

TABLE PRESENTING THE MEAN AND MEDIAN VALUE FOR DSC, BDSC, AND HD95 FOR THE PROPOSED MAE-BASED APPROACH, AND THE STATE-OF-THE-ART NETWORKS IN VOLUMETRIC SEGMENTATION. THE "D" DENOTES THE PROPOSED APPROACH WITH THE DEFORMABLE ELASTIC DEFORMATIONS AND "ND" MASKING WITH SHARP EDGES.

Method	SkullBreak						SkullFix					
	DSC \uparrow		BDSC \uparrow		HD95 [mm] \downarrow		DSC \uparrow		BDSC \uparrow		HD95 [mm] \downarrow	
	Avg	Med	Avg	Med	Avg	Med	Avg	Med	Avg	Med	Avg	Med
MAE (ND)	0.787	0.820	0.793	0.808	4.02	2.43	0.916	0.922	0.929	0.930	1.63	1.36
MAE (D)	0.863	0.882	0.902	0.905	1.92	1.70	0.911	0.919	0.935	0.939	1.80	1.43
UNet	0.692	0.721	0.746	0.758	5.10	3.15	0.817	0.844	0.832	0.845	2.83	2.24
AttUNet	0.741	0.772	0.781	0.793	3.17	2.65	0.841	0.854	0.850	0.859	2.33	1.94
SegResNet	0.729	0.743	0.737	0.746	3.28	2.87	0.819	0.833	0.811	0.816	2.45	2.34
UNETR	0.670	0.681	0.712	0.728	4.01	3.44	0.779	0.794	0.801	0.806	2.88	2.71
SwinUNETR	0.702	0.715	0.708	0.715	3.21	2.82	0.785	0.786	0.770	0.775	2.82	2.39

on the SkullFix dataset where all the defects are sharp and the deformable improvement is not crucial.

Another advantage of the method is connected with easy extendability into new datasets. The method can be applied to all datasets containing healthy skulls, without the necessity of manual preprocessing and defect synthesis. Therefore, in future work, we plan to acquire and combine more open datasets consisting of healthy skulls from computed tomography.

The disadvantage of the proposed method is a longer training time. Training of the method on just 214 cases required about 1200 epochs to converge, in contrast to the supervised segmentation architectures which all converged before reaching 500 epochs. It is connected with the fact that the proposed method needs to learn the variability of the whole skull, including facial regions and significantly larger defects, not present in the ground truth of the SkullFix/SkullBreak datasets. Nevertheless, there is no significant difference in the inference time between the approaches, allowing one to perform the inference in less than 100ms using a modern workstation equipped with e.g. NVIDIA RTX 3090 GPU.

The method could be further improved by augmenting the healthy skulls, before performing the patch generation and masking. This could lead to a further increase in the dataset heterogeneity and the model generalizability.

IV. CONCLUSIONS

In this work, we proposed an alternative method to automatic cranial defect reconstruction based on the self-supervised deep masked autoencoders, enhanced by random elastic deformations of the masked input patches. We compared the proposed method to several state-of-the-art solutions and confirmed that it improves the reconstruction quality by more than 0.1 and 1.0 mm in terms of Dice score and Hausdorff distance respectively. In further work, we plan to extend the proposed approach by providing augmentation to the healthy skulls themselves to further increase the dataset heterogeneity. Moreover, we plan to extend the comparison using other datasets and confirm the usability in real clinical settings.

REFERENCES

- [1] Michael C. Dewan et al., "Global neurosurgery: the current capacity and deficit in the provision of essential neurosurgical care. Executive Summary of the Global Neurosurgery Initiative at the Program in Global Surgery and Social Change," *Journal of Neurosurgery*, vol. 130, no. 4, pp. 1055–1064, 2019.
- [2] Jianning Li et al., "AutoImplant 2020-First MICCAI Challenge on Automatic Cranial Implant Design," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2329–2342, 2021.
- [3] Jianning Li et al., "Towards clinical applicability and computational efficiency in automatic cranial implant design: An overview of the autoimplant 2021 cranial implant design challenge," *Medical Image Analysis*, vol. 88, pp. 102865, 2023.
- [4] B. Yang, K. Fang, and X. Li, "Cranial Implant Prediction by Learning an Ensemble of Slice-Based Skull Completion Networks," *MICCAI 2021, Cranial Implant Design Challenge*, vol. 13123 LNCS, pp. 95–104, 2021.
- [5] H. Mahdi et al., "A U-Net Based System for Cranial Implant Design with Pre-processing and Learned Implant Filtering," *MICCAI 2021, Cranial Implant Design Challenge*, vol. 13123 LNCS, pp. 63–79, 2021.
- [6] M. Wodzinski, M. Daniol, and D. Hemmerling, "Improving the Automatic Cranial Implant Design in Cranioplasty by Linking Different Datasets," *MICCAI 2021, Cranial Implant Design Challenge*, vol. 13123 LNCS, pp. 29–44, 2021.
- [7] David G. Ellis and Michele R. Aizenberg, "Deep Learning Using Augmentation via Registration: 1st Place Solution to the AutoImplant 2020 Challenge," *MICCAI 2020, Cranial Implant Design Challenge*, pp. 47–55, 2020.
- [8] K. Kwarciak and M. Wodzinski, "Deep Generative Networks for Heterogeneous Augmentation of Cranial Defects," *Proceedings of the IEEE/CVR International Conference on Computer Vision - LIMIT Workshop*, pp. 1066–1074, 2023.
- [9] M. Wodzinski et al., "Deep learning-based framework for automatic cranial defect reconstruction and implant modeling," *Computer Methods and Programs in Biomedicine*, vol. 226, pp. 1–13, 2022.
- [10] M. Wodzinski et al., "High-Resolution Cranial Defect Reconstruction by Iterative, Low-Resolution, Point Cloud Completion Transformers," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 333–343, 2023.
- [11] P. Friedrich et al., "Point Cloud Diffusion Models for Automatic Implant Generation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 112–122, 2023.
- [12] O. Kodym et al., "SkullBreak / SkullFix – Dataset for automatic cranial implant design and a benchmark for volumetric shape learning tasks," *Data in Brief*, vol. 35, 2021.
- [13] J. Li et al., "MUG500+: Database of 500 high-resolution healthy human skulls and 29 craniotomy skulls and implants," *Data in Brief*, vol. 39, 2021.
- [14] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16000–16009.
- [15] Christoph Feichtenhofer, Yanghao Li, Kaiming He, et al., "Masked autoencoders as spatiotemporal learners," *Advances in neural information processing systems*, vol. 35, pp. 35946–35958, 2022.
- [16] Chaoning Zhang, Chenshuang Zhang, Junha Song, John Seon Keun Yi, Kang Zhang, and In So Kweon, "A survey on masked autoencoder for self-supervised learning in vision and beyond," *arXiv preprint arXiv:2208.00173*, 2022.
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [18] Eric Kerfoot, James Clough, Ilkay Oksuz, Jack Lee, Andrew P. King, and Julia A. Schnabel, "Left-ventricle quantification using residual u-net," in *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9*. Springer, 2019, pp. 371–380.
- [19] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R. Roth, and Daguang Xu, "Unetr: Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 574–584.
- [20] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R. Roth, and Daguang Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI Brainlesion Workshop*. Springer, 2021, pp. 272–284.
- [21] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, et al., "Attention U-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
- [22] Andriy Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*. Springer, 2019, pp. 311–320.
- [23] M. Jorge Cardoso et al., "MONAI: An open-source framework for deep learning in healthcare," Nov. 2022.
- [24] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin, "TorchIO: A Python library for efficient loading, preprocessing, augmen-

tation and patch-based sampling of medical images in deep learning,”
Computer Methods and Programs in Biomedicine, vol. 208, pp. 106236,
2021.