IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 June 2024, accepted 2 July 2024, date of publication 8 July 2024, date of current version 18 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3425171

RESEARCH ARTICLE

Automatic Implant Generation for Cranioplasty via Occupancy Networks

STEFANO MAZZOCCHETTI[1], MIRKO BEVINI[2], GIOVANNI BADIALI[2,3], GIUSEPPE LISANTI[4], LUIGI DI STEFANO[4], (Member, IEEE), AND SAMUELE SALTI[4]

[1]eDIMES Lab—Laboratory of Bioengineering, Department of Medical and Surgical Sciences, University of Bologna, 40138 Bologna, Italy
[2]Oral and Maxillo-Facial Surgery Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna, Italy
[3]Department of Biomedical and Neuromotoric Science (DIBINEM), University of Bologna, 40138 Bologna, Italy
[4]Department of Computer Science and Engineering (DISI), University of Bologna, 40126 Bologna, Italy

Corresponding author: Stefano Mazzocchetti (stefano.mazzocchett5@unibo.it)

This work was supported by the Alma Idea 2022 grant—Alma Mater Studiorum—University of Bologna, funded by the European Union—NextGenerationEU program (Italian D.M. 737/2021) under Grant CUP J45F21002000001.

ABSTRACT The design of patient-specific implants for cranioplasty surgery is time-consuming and challenging. Hence, the 2021 AutoImplant II challenge, consisting of the SkullBreak and SkullFix datasets, was organized to foster research on computer vision techniques pursuing automation of the cranial implant design task. Data-driven methods working on Computed Tomography (CT) emerged as a promising procedure for the realization of such a task. The best performing approaches turned out to rely on ensembles of Convolutional Neural Networks (CNN) architectures that either process each CT slice separately or the entire voxelized volume through computationally demanding three-dimensional convolutions. More recently, few methods were designed to deal with different data representations, for instance point clouds, to perform skull completion. Similarly, we investigate a novel solution for implant generation that deploys a conditioned occupancy network. Starting from the partial point cloud, we directly reconstruct the completed voxel grid by evaluating the learned occupancy function in the given space resolution. Our approach can generate high-quality implants achieving qualitative and quantitative results comparable to state-of-the-art methods on the SkullBreak and SkullFix datasets while requiring significantly less computational resources. The model trained on the SkullBreak dataset successfully generalize to real craniotomies provided in the MUG500+ dataset.

INDEX TERMS 3D deep learning, autoimplant challenge, cranioplasty surgery, shape completion, personalized medicine, surgery planning, implicit neural representation.

I. INTRODUCTION

Cranioplasty is the surgical procedure involving the repair or replacement of a cranial bone defect arising from traumatic injuries, tumor resections, and other neurosurgical interventions. In cranioplasty, it is crucial to restore both the aesthetic and functional aspects of the cranial skeleton. During the past years, this medical procedure has witnessed significant advances thanks to joint interdisciplinary efforts between neurosurgeons, biomaterial scientists, radiologists, and bioengineers. Hence, surgical planning relies nowadays on manual manipulation of three-dimensional medical imaging data carried out by specialized professionals through sophisticated and expensive computer aided design (CAD) software. Nonetheless, the process is time-consuming, as it requires from days to weeks to generate a cranial implant. Moreover, the resulting implant can vary significantly across clinicians, as shown in [6]. To overcome these limitations as well as reach better accuracy and faster surgical planning, the first (2020) and second (2021) editions of the AutoImplant challenge were organized in order to foster research on computer vision techniques pursuing automation of the cranial implant design task. Two datasets have been released in the challenge: SkullBreak and SkullFix. They consist of binary voxel grid triplets extracted from Computed Tomography (CT) scans: the complete skull, the defective

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea Bottino.

skull and the implant. While in SkullFix the synthetic implant does not vary much, in SkullBreak the defect is injected in a random position with a random shape. A major finding from the challenge deals with data-driven methods working on CTs emerging as a promising approach to achieve automation of implant design. In particular, the task can be addressed by Computer Vision techniques aimed at shape completion, which pertains to predicting a complete shape given a partial one. Most of the proposed solutions rely on CNNs (Convolutional Neural Networks) that process either 3D voxel data [12], [16], [19], [29] or 2D CT slices [25], [33]. Even if CNN-based solutions that work on the above regular, grid-like data structures yield promising results and, nowadays, achieve state-of-the-art performance in the AutoImplant challenge, they require careful ad-hoc pre-processing of the input voxel grid to reduce its size without compromising the output quality, a compromise attainable on the particular data released for the challenge but not possible in general. In real cases where such pre-preprocessing is not feasible, their deployment is likely to be hindered by high-memory consumption and slow training and inference time. Recently, few works [7], [27] have tried to overcome the above-mentioned limitations by performing the shape completion task on memory-efficient data structures such as point clouds and then applying a voxelization step to match the output format required by the challenge. In particular, the latest work in the field [7] proposes the first investigation on the use of Diffusion Models [9], [36], arguably the most successful AI tool for visual content generation, to address the implant generation task.

In this work, we explore the use of another family of modern and effective deep learning tools, i.e. implicit neural representations of 3D shapes [4] (also more broadly referred to as neural fields [32]). In particular, we propose a novel approach to automatic cranial implant generation based on representing the completed skull as a continuous occupancy function parameterized as a neural network [21]. The occupancy function is conditioned by the encoding of the partial point cloud representing the surface of the defective skull. The complete skull is generated by evaluating the occupancy function on the whole voxel grid at its original spatial resolution and voxel size, seamlessly possible thanks to the continuous nature of the learned implicit representation, and the implant is then extracted by the reconstruct and subtract technique.

Our approach generates high-quality implants with a low memory footprint, a fast generation time and an end-to-end training protocol. Quantitative evaluation on the SkullBreak and SkullFix datasets shows that such limited consumption of resources does not affect output quality, which is comparable with respect to current state-of-the-art methods. Eventually, it is worth highlighting how, unlike any previous proposals, ours learns a continuous function representing the whole skull and, therefore, is conducive to seamlessly generating implants at any desired resolution.

II. RELATED WORK

The use of 3D printed patient-specific implants is nowadays common practice for cranioplasty. Current approaches for designing the implant rely on the manipulation of three-dimensional medical imaging data. This method can require days to weeks to create an implant and necessitate highly trained professionals. Other approaches are based on computer-aided design (CAD) like: mirroring-based methods [1], [5]; surface interpolation methods [2], [28] or template deformation methods with statistical shape models [8], [24]. Although CAD approaches are effective, they lack generalizability depending, for example, on the defect position or the dataset used to select the template.

Instead, deep learning methods arise as a promising solution for the automatic implant generation task. The first (2020) and second (2021) editions of the AutoImplant challenge allowed investigation of the effectiveness of deep learning approaches on the skull completion task for synthetic cranial defects.

Promising techniques rely on 3D CNN architectures [12], [19], [29]. The authors of [29] trained a cascade of two 3D U-Nets directly on the voxel grid. The first network aims to predict the implant shape which is concatenated with the defective input and fed into the second U-Net to predict the complete skull. Even if they reached high-quality results, the training time and GPU memory consumption represent a drawback, as highlighted in [13]. Since the majority of the space is occupied by the background [11], Li et al. [16] proposed a coarse-to-fine solution to deal with the memory consumption problem. The authors identified that the difference between coarse and high-resolution voxel grids is the arrangement pattern of valid voxels. For this reason, after a coarse voxel grid is generated by a 3D CNN architecture, the high-resolution shape is obtained by rearranging valid voxels. Another solution to avoid wasting memory and time to process the empty background relies on sparse 3D convolutions, like those implemented in [12], [14]. Despite the lower memory consumption, all these solutions obtain lower performance compared to [29].

Slice-based methods [25], [33] tried to address the automatic implant generation task by using 2D convolution separately on each slice of the CT scan. The authors of [25] proposed to deal with this task following an inpainting approach. The inpainting is performed separately for each slice exploiting CNN. They trained three different CNN networks to predict directly the implant slices. Each network processes slices taken along different axes of the volume (i.e., multiaxial CNN). A post-processing stage synthetizes the estimated 2D slices to obtain the final volumetric implant. Yang et al. [33] proposed to use an ensemble of three RNN and three CNN multiaxial models to compensate for the loss of global skull features observed in the previous solution [25]. The output of the CNN multiaxial slice networks and RNNs is then fused to obtain the final implant estimation. While slice-based methods have lower computational requirements

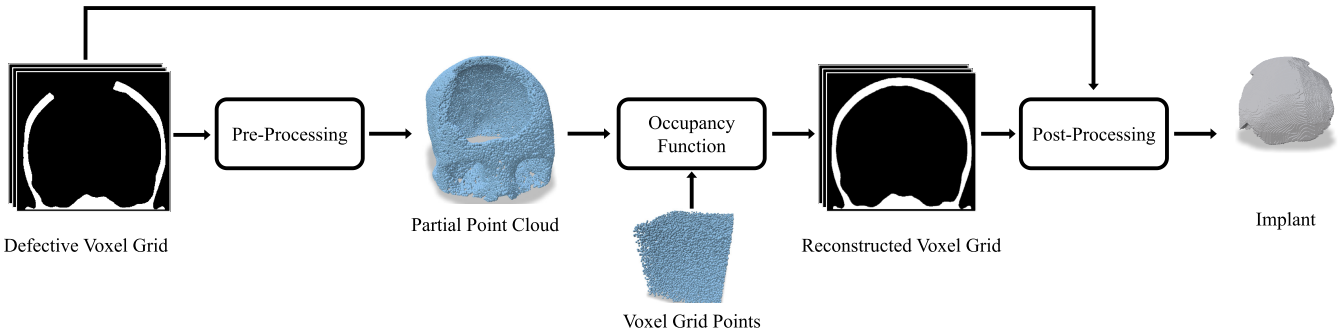

FIGURE 1. Complete pipeline for implant generation. The defective binary voxel grid is pre-processed to generate a partial point cloud that acts as a condition for the occupancy function. The implicit function computes an occupancy value for each point of the voxel grid generating a completed skull. The final implant is obtained with a post-processing stage.

compared to methods that process the whole 3D volume [12], [29], the overall quality is lower as the generated surfaces are not as smooth as the one obtained with such methods and they exhibit a loss of details due to missing global information.

To deal with both the loss of global information and the high memory consumption, a few recent works [7], [27], [30] proposed to tackle the skull implant generation problem by using Point Clouds as data representation. The authors of [27] employ a Generative Adversarial Network (GAN) to generate the complete shape of the skull. However, this solution processes low-resolution point clouds which results in a poor mesh reconstruction. A solution to the low resolution problem is proposed in [30], where multiple random samples of a high-resolution skull are completed through a point cloud completion method based on transformers [34], and the results are merged to generate the implant. However, completion performance is always lower than that attained by methods using 3D CNNs. Instead, [7] proposed to use Diffusion Probabilistic Models [36] to generate the implant' points cloud. The output of the diffusion model is further processed by a voxelization network [20] to reconstruct the complete voxel grid. Thanks to the powerful generation ability of the diffusion model, they can ensemble different generations for each defective skull. Even if they can achieve results comparable to the state of the art, the diffusion network and the voxelization network cannot be trained end-to-end and the probabilistic diffusion generation requires 1000 denoising steps. Overall, the implant generation process can take up to 1200s.

Differently from the solutions described above we propose to deal with the cranial implant design problem by learning an occupancy function conditioned on the partial point cloud of the defective skull. The learnable implicit function can be trained end-to-end with few pre-processing and post-processing steps. The proposed solution has a fast inference time and a small GPU memory consumption and can achieve comparable qualitative and quantitative results with respect to state-of-the-art methods.

III. CRANIAL IMPLANT GENERATION

We address cranial implant generation as a shape completion task. As depicted in Figure 1 our proposed solution consists

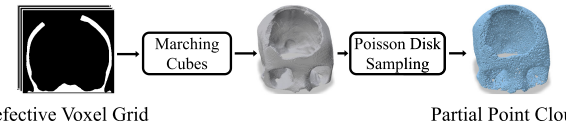

FIGURE 2. Pre-processing pipeline defined to sample the defective skull partial point cloud.

of three main stages. First, the CT scan of the partial, i.e. defective, skull is converted into a point cloud with a pre-processing step. Then, the partial point cloud is provided as input to a learned deep occupancy network which, for each 3D query point, predicts the probability of belonging to the complete shape. By computing and thresholding the occupancy function at each center point of a chosen voxel grid, a binary voxel grid representing the completed skull is obtained. Finally, the implant is extracted through a post-processing step.

A. PRE-PROCESSING

Starting from the binary volume extracted from the CT scan, the goal is to obtain a point cloud that represents the surface of the partial skull. To this end, the triangular mesh is extracted using the Marching Cubes algorithm [18]. Then, in order to have a uniform partial point cloud of fixed size, $n_{partial}$ vertices are sampled using the Poisson Disk Sampling algorithm [35] as shown in Figure 2. We set $n_{partial} = 35000$ in our experiments.

B. OCCUPANCY NETWORK

Differently from [7] which exploits diffusion models to perform the shape completion starting from a partial point cloud, we propose to represent the skull using a learned implicit function, i.e. the occupancy function, parametrized as a neural network. Peculiarly, our training process does not mandate discretizing the 3D space and shapes are represented in a continuous way. In more detail, we learn a neural network that predicts an occupancy value for each 3D query point, conditioned by the partial point cloud extracted from the defective skull. To this end, we employ the Convolution Occupancy Network proposed in [21].

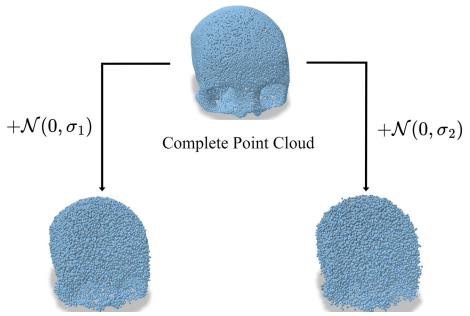

FIGURE 3. Query points generation starting from the complete point cloud sampled from the ground truth shape.

The input point cloud $x_{partial} \in \mathbb{R}^{n_{partial} \times 3}$ is first encoded into a 3D feature grid using a shallow PointNet [22] with local pooling. All the features that fall in the same voxel are processed through average pooling to encapsulate local geometry information. This process outputs a 3D feature volume, x_c, of dimension $r \times r \times r \times d$, where r is the resolution of the voxel grid and d is the feature dimensionality. This feature volume is then processed through a 3D convolutional U-Net architecture [23]. This step is important to integrate both local and global shape information. The convolution occupancy network exploited in our model works on a volume with $r = 64$. In order to extend the receptive field of our network and avoid generating incomplete implants caused by the sparseness of the volume, we modified the 3D U-Net used in [21] by adding one 3D convolutional layer.

The goal of the occupancy function is to estimate the probability of an input query point $q \in \mathbb{R}^3$ to belong to the conditioning 3D shape.

Thus, along with the spatial coordinates of the query point, the predictor (aka implicit decoder) takes as input the trilinear interpolation of the 3D features associated with the vertices of the voxels in x_c where the query point falls. The overall learnable occupancy function can be expressed as:

$$f_\theta(q, \phi(q, x_c)) \rightarrow [0, 1], \tag{1}$$

where f_θ is a neural network composed of multiple ResNet blocks, θ are the weights of the network, q is the query point for which the occupancy is predicted, x_c is the 3D feature volume grid and $\phi(q, x_c)$ returns the query point feature.

At training time, $60k$ query points are chosen from the ground truth point cloud by extracting samples from the complete volume. As illustrated in Figure 3, this involves two distinct sampling phases. In the first one, a set of sampled points is slightly displaced by adding random Gaussian noise with a standard deviation of σ_1 to the coordinates. The second set of sampled points undergoes a similar procedure, but this time with a different standard deviation, σ_2, where σ_2 is higher than σ_1. In addition, random points are uniformly sampled from the whole bounding volume. The selection of uniform points in the whole volume is necessary because more than 90% of the volume is empty [11].

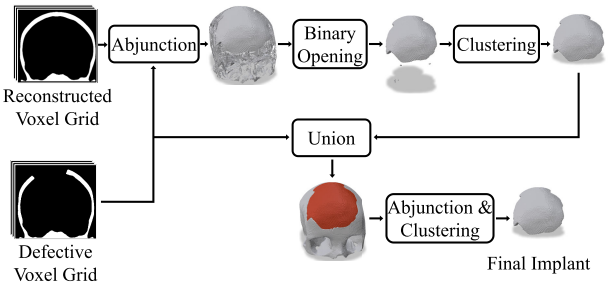

FIGURE 4. Post-processing pipeline defined to generate the final implant.

FIGURE 5. SkullBreak samples for each defect type. The ground truth implant is highlighted in red.

Differently from [21], we employ the Focal Loss [17] between the predicted $\hat{o}_q$ and the true occupancy values o_q as the objective function:

$$L(\hat{o}_q, o_q) = -[\alpha\, o_q\, (1 - \hat{o}_q)^\gamma\, log(\hat{o}_q) + \tag{2}$$

$$+ (1 - \alpha)\, (1 - o_q)\, \hat{o}_q^\gamma\, log(1 - \hat{o}_q)] \tag{3}$$

Indeed, due to 90% of the binary volume being empty, the focal loss enables to weigh less the many easy-to-classify samples (e.g., points in the empty part of the volume) while giving more importance to harder ones (e.g., points that belong to the skull or lay close to the surface). At inference time, each point within a voxel grid of any desired resolution can be used as a query.

C. POST-PROCESSING

Compared to existing solutions that generate implants at a smaller resolution than the input one, and require slow and ad-hoc postprocessing to upsample it to the desired resolution, thanks to its implicit formulation our method can directly generate the complete skull voxel grid at arbitrary resolutions. Therefore, the aim of our postprocessing is only implant extraction. The overall post-processing pipeline is shown in Figure 4. The first step performs an abjunction between the generated skull voxel grid and the defective one in order to select only the voxels of interest for the implant. Since some noise emerges during this process, a 3D binary opening operation is performed. After this step, the implant can be recognized as the largest connected volume. Therefore, a clustering algorithm is exploited to extract the implant. Finally, to enhance the implant's border shape, the extracted implant is added back to the defective skull voxel grid and re-processed by performing a binary closing, an abjunction operation, and a second step of clustering.

TABLE 1. Comparison of the average DCS, bDCS and HD95 obtained on the SkullBreak test set. For [7] we also report the result obtained considering their ensemble strategy.

Method	DCS ↑	bDCS ↑	HD95 ↓	Resolution	Memory[GB]	Time[s]
3D U-Net [29]	0.87	0.91	2.32	$240 \times 200 \times 240$	5.7	75
3D U-Net (sparse) [12]	0.71	0.80	4.60	$512 \times 512 \times 277$	18.4	3
2D U-Net [33]	0.87	0.89	2.13	512×512	2.7	270
PCDM [7]	0.86	0.88	2.51	$512 \times 512 \times 512$	12.9	1002
PCDM (n=5) [7]	0.87	0.89	2.45	$512 \times 512 \times 512$	12.9	1071
Ours	0.85	0.89	2.39	$512 \times 512 \times 512$	4.1	27

TABLE 2. Comparison of the average DCS, bDCS and HD95 obtained on the SkullFix test set. For [7] we also report the result obtained considering their ensemble strategy.

Method	DCS ↑	bDCS ↑	HD95 ↓	Res	Memory[GB]	Time[s]
3D U-Net [29]	0.91	0.95	1.79	$240 \times 200 \times 240$	5.7	75
3D U-Net (sparse) [12]	0.81	0.87	3.04	$512 \times 512 \times 151$	7.6	48
2D U-Net [33]	0.89	0.92	1.98	512×512	2.7	317
PCDM [7]	0.90	0.92	1.73	$512 \times 512 \times 512$	12.9	1071
PCDM (n=5) [7]	0.90	0.93	1.69	$512 \times 512 \times 512$	12.9	1210
Ours	0.90	0.93	1.86	$512 \times 512 \times 512$	4.1	28

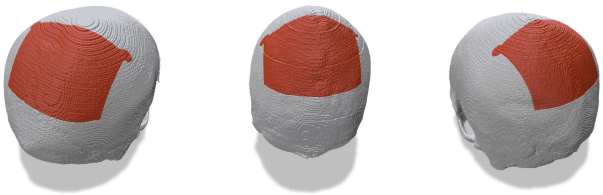

FIGURE 6. Some training samples of SkullFix with the highlighted ground truth implant in red.

IV. EXPERIMENTS

In this section, we report on a series of experiments performed to assess the validity of the proposed approach. A brief description of the implementation settings and the datasets employed is first provided. Then, an analysis of the quantitative performance is conducted, also in comparison with state-of-the-art models, such as the cascaded 3D U-Net architecture [29], the method based on sparse volumetric convolutions [12], the ensemble of multi-axial slice-based 2D network [33] and the latest proposal in the field, namely the approach relying on a diffusion probabilistic model of point clouds followed by a voxelization network [7] (PCDM). Finally, qualitative samples of the implant meshes reconstructed using the proposed solution are presented, along with some concluding remarks.

A. IMPLEMENTATION DETAILS AND SETTINGS

Our networks have been trained for 1000 epochs using Adam [10] as the optimizer and a learning rate of $10e-4$, for both datasets. The Focal Loss has been used as the training objective function while, at validation time, the Intersection over Union (IoU) over the entire binary volume is used to select the best checkpoint. We set $\gamma = 2$ and $\alpha = 0.7$ for the Focal Loss. All the experiments have been performed on a NVIDIA RTX3090 Ti GPU with 24GB of memory. All the point coordinates are normalized in the interval $[0, 1]$. As described in Section III-B, a total number of $60k$ points are used as query points for each partial shape. These query points are sampled from the complete point cloud of the skull as follows: $20k$ points are randomly sampled and modified by adding a Gaussian noise with $\sigma_1 = 0.005$; other $20k$ points are sampled and modified by adding a Gaussian noise with $\sigma_2 = 0.01$ and the remaining $20k$ points are sampled uniformly from the unit cube. In the post-processing, the binary opening is performed using a three-dimensional kernel of size 3, while the binary closing has a kernel size of 5.

B. DATASETS

The data used in the experiments comprise the training sets of the SkullBreak and SkullFix [11] datasets from the AutoImplant II Challenge. They are composed of binary CT scan slices where synthetic defects have been injected. Both are subsets of the publicly available CQ500 [3] dataset, which was collected in various hospitals across India.

The SkullBreak training set has 114 samples with 5 different synthetic defects for each skull with multiple positions and shapes. The aim is to emulate the variability in cranioplasty by placing the defect in specific positions, such as bilateral, fronto-orbital, and parieto-temporal, but also randomly. Some examples are shown in Figure 5. In this dataset, the defect's borders are not perfectly cropped to reflect the bone resorption at different stages in cranioplasty. The skulls in the dataset have an isotropic voxel size of 0.4mm with a voxel grid of $512 \times 512 \times 512$.

The SkullFix training set has 100 skulls with one implant per sample. Each skull has a variable voxel size with a variable voxel grid of $512 \times 512 \times Z$. For this reason, following [7], the CT scans have been resized with an

TABLE 3. Different measures for the results obtained by the proposed solution on the SkullBreak and SkullFix datasets considering different defects.

		SkullBreak						SkullFix
		Bilateral	Fronto-Orbital	Parieto-Temporal	Random 1	Random 2	Overall	Overall
DSC ↑	Average	0.86	0.86	0.86	0.83	0.84	0.85	0.90
	Min	0.69	0.70	0.65	0.57	0.61	0.57	0.79
	Max	0.94	0.93	0.94	0.94	0.92	0.94	0.98
	Std	0.06	0.05	0.07	0.09	0.07	0.07	0.05
bDSC ↑	Average	0.90	0.88	0.90	0.89	0.89	0.89	0.93
	Min	0.80	0.74	0.83	0.76	0.77	0.74	0.88
	Max	0.95	0.93	0.95	0.95	0.94	0.95	0.98
	Std	0.03	0.04	0.03	0.05	0.04	0.04	0.02
HD95[mm] ↓	Average	2.62	1.92	2.08	2.68	2.65	2.39	1.86
	Min	0.98	1.13	0.89	0.89	0.89	0.89	0.45
	Max	5.20	2.83	4.08	5.20	4.88	5.20	3.78
	Std	1.09	0.46	0.79	1.13	0.93	0.96	0.83

isotropic voxel of 0.45mm which results in a cubic voxel grid of $512 \times 512 \times 512$. Some samples are shown in Figure 6. Differently from SkullBreak, the defect shape and position are similar in the whole SkullFix dataset. Therefore, the shape completion task is easier when compared with SkullBreak.

Since the challenge is no longer active, to assess the performance of all models we used the train/test splits created from the original training sets in [7] and kindly shared with us by the authors. The new training sets have been additionally split into train/validation sets to validate the network during training and perform early stopping. More in detail, for SkullBreak we use 78 samples for train, 8 for validation and 28 for test. While for SkullFix we consider 68 samples for training, 7 for validation, and 25 for the test set.

Finally, an experiment on a dataset not used for training has been performed to verify the generalization abilities of the proposed approach. In particular, the subset of the MUG500+ dataset [15] with real craniotomies and the ground truth implants designed by expert clinicians has been used. This dataset was constructed based on the head CT scans acquired from the Medical University of Graz (MUG) in clinical routines. The partial skulls in this dataset are characterized by a voxel grid size of $512 \times 512 \times Z$ where $Z \in [147, 291]$. Moreover, the voxel size dx, dy, dz is variable. In particular $dx \in [0.35, 0.5]$, $dy \in [0.35, 0.5]$ and $dz \in [0.5, 1.0]$. As pointed out in [31], completion networks are trained to inpaint the missing shape of the skull. In contrast, the optimal shape of the implant depends on the geometric and mechanical properties of the 3D printing material which the networks cannot be aware of, e.g. the implants designed by clinicians in the MUG500+ dataset are thinner than the actual skull bone [6]. Hence, quantitative comparisons between the ground-truth and the network predictions are not meaningful, and we provide qualitative results to assess the performance of the model on this dataset.

C. QUANTITATIVE EVALUATION

In this section, we provide a detailed quantitative evaluation in terms of output quality, resource consumption and inference speed.

FIGURE 7. Qualitative results for the SkullFix dataset.

1) OUTPUT QUALITY

Tables 1 and 2 summarize the average results on both datasets in comparison with existing state-of-the-art solutions. As reported in the table, due to the large memory footprint required by 3D convolutions, the experiments conducted in [7] employed different resolutions for each method, following the settings used in the original solutions [7], [12], [29], [33]. Here, we adopt the same resolutions. In particular, a voxel grid resolution of $240 \times 200 \times 240$ was used for 3D U-Net [29], $512 \times 512 \times 277$ for 3D U-Net (sparse) [12], $512 \times 512 \times 512$ for PCDM [7] and an image of 512×512 for 2D U-Net [33]. It is worth noting that to compare results with the ground-truth, the outputs of the competitors' models require a post-processing step by using decropping, padding and spline interpolation to bring back each voxel grid to the original resolution and voxel size, which are then used to compute the metrics. Notably, this slow and ad-hoc post-processing step is not required in our approach, which can seamlessly produce outputs at arbitrary resolution and voxel size thanks to its continuous formulation.

Results are reported in terms of Dice Similarity Score (DSC), boundary Dice Similarity Score (bDSC) and Hausdorff distance at the 95th percentile (HD95), as proposed in the AutoImplant II challenge. More in detail: the Dice Similarity Score (DSC) measures the area of overlap between the predicted implant and the ground truth implant; the boundary Dice Similarity Score (bDSC) quantifies the DSC

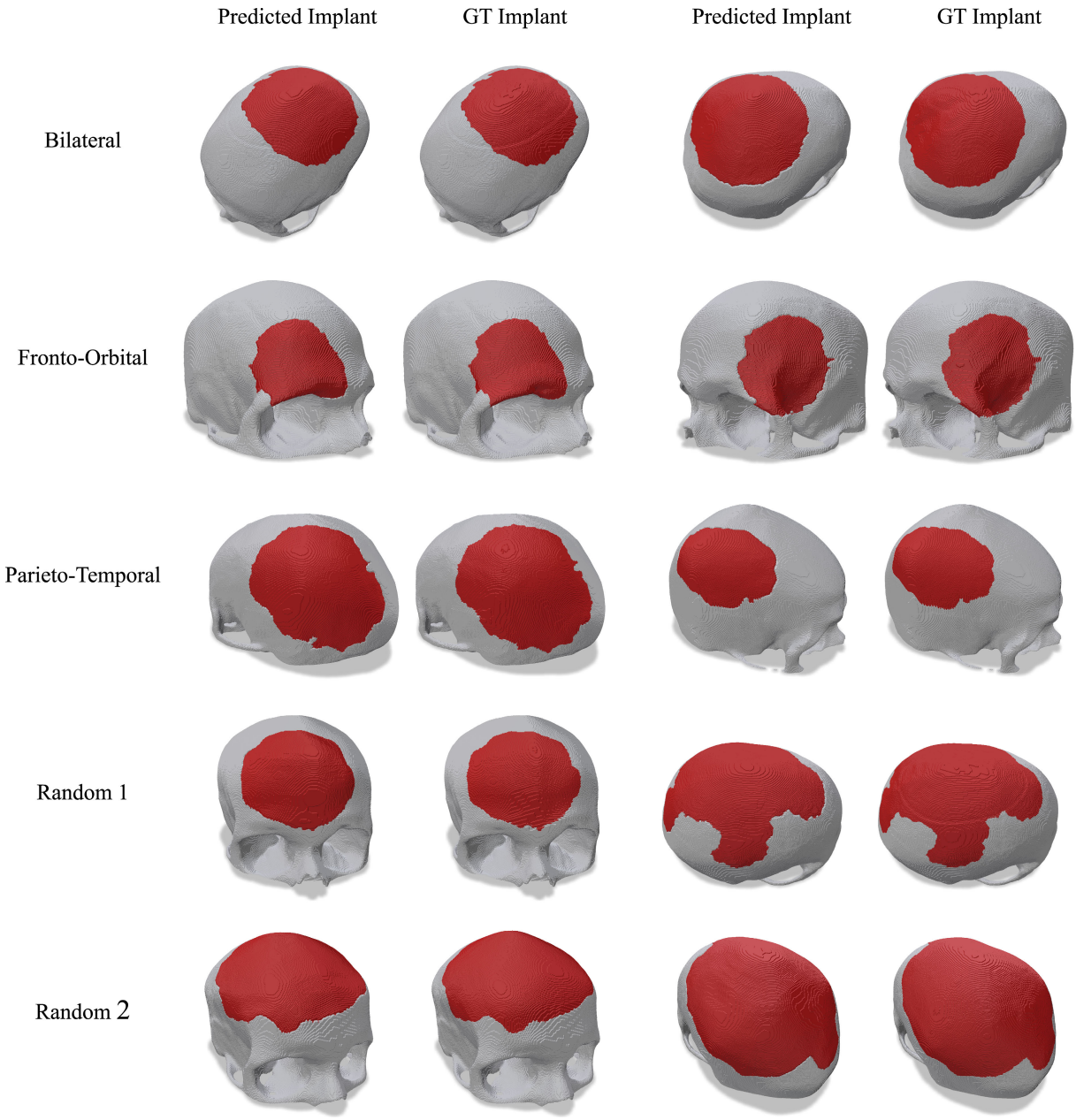

FIGURE 8. Qualitative results for the SkullBreak dataset.

at the border of the implant, while the Hausdorff Distance 95% (HD95) measures how far the contours of the predicted implant are from the ground truth at the 95th percentile.

From Tables 1 and 2 we can appreciate how there is one method that is under-performing, i.e. 3D U-Net (sparse) [12]. The remaining methods including our solution attain similar performance, e.g. our solution obtains average bDSC values of 0.89 and 0.93 on the SkullBreak and SkullFix datasets, respectively, which is the second best result in both datasets. As shown in [13], the bDSC is the most important metric due to its correlation to the doctor's evaluation.

Table 3 provides a more detailed evaluation of the different types of defects. Overall, performance is consistent across defect types and the overall variance across samples is low, especially for the important bDSC metric. Some large variations are mainly due to synthetic defects which produce unrealistic cranioplasty scenarios, e.g. large frontal defects in the eye regions (see also Figure 11). In these cases, the eye orbits can be considered as a part of the defect leading to large errors, e.g. low minimum DSC values on the Random 1 and Random 2 defect types. Indeed, on the SkullFix dataset, where real defects are considered and only minor variations in shape and position appear, such outliers are absent and the standard deviation is lower. Detailed results for SkullFix are reported in the last column of Table 3.

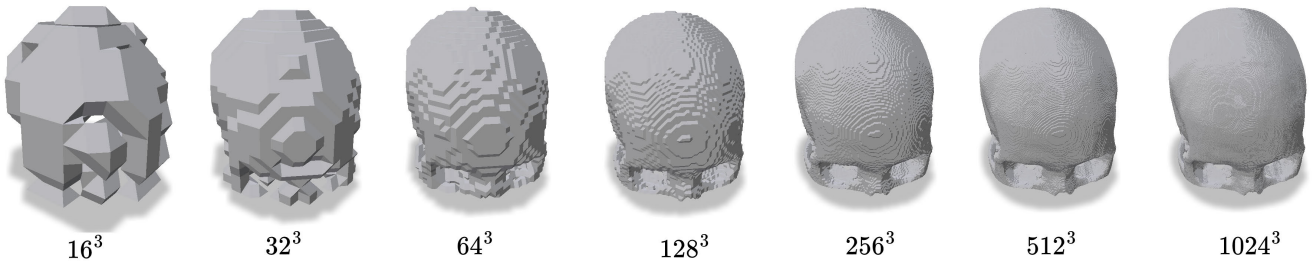

FIGURE 9. Surface reconstruction for different voxel grid resolutions.

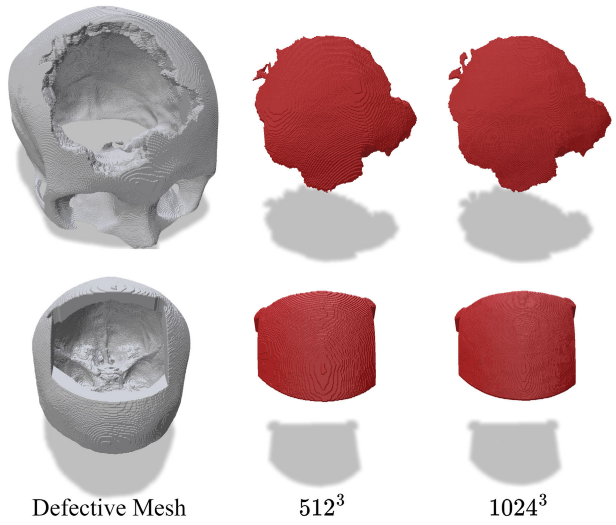

FIGURE 10. Comparison between implants generated with a 512^3 and 1024^3 voxel grid resolution.

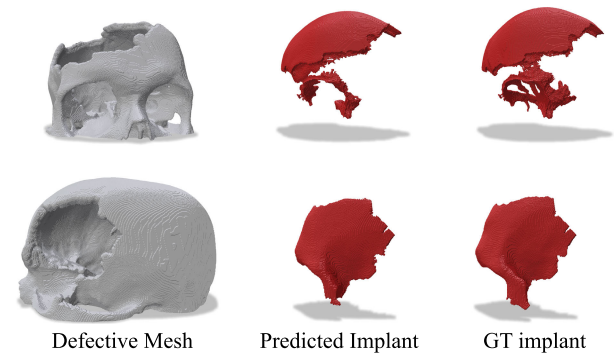

FIGURE 11. Defective skull mesh, reconstructed implant and GT implant for two error cases in the SkullBreak test set.

2) RESOURCE CONSUMPTION AND INFERENCE SPEED

In terms of efficiency, among the well-performing methods ours is by far the fastest, as reported in Tables 1 and 2. In particular, our solution is about 40x faster than the method based on diffusion models [7] which requires more than 15 minutes to generate an implant. Our solution is also faster than methods processing directly voxel grids like 3D U-Net [29] since handling voxel grids with 3D convolutions is highly expensive. For this reason, they have to process an input voxel grid with a lower resolution of 240 × 200 × 240 which, however, as detailed above, requires a costly and ad-hoc post-processing operation to undo the effects of cropping, padding and spline interpolation that increases the overall time to run such solutions. Moreover, such ad-hoc reduction of input size is possible only because of the peculiar arrangement of the input voxel grids in the considered dataset. Indeed, in these datasets there is a large amount of empty voxels around the skulls along the axes, which can be cropped, thereby reducing the size of the voxel grid, without incurring information loss. This is however not the case in general, and had 3D U-Net be required to process the input volumes at the original resolution, it would have required a massive amount of memory and would have been much slower. Indeed, we were not able to run it on the original voxel grids even on a GPU with 40 GBs of V-RAM.

D. QUALITATIVE EVALUATION

In this section, we qualitatively analyze the performance of the proposed approach by showing the mesh generated from the binary voxel grid. Qualitative results for the SkullFix test samples are presented in Figure 7 while in Figures 8 we show two generated implants alongside ground truth for each type of defect in SkullBreak. We can observe that the proposed method generalizes well to unseen defects while still being able to generate high-quality implants featuring a smooth surface. In particular, generalization abilities on SkullBreak are particularly remarkable since it contains 5 different synthetic defects that have been generated with a random shape in different locations of the skull. Nonetheless, our approach is able to produce an implant that can perfectly fit these varying defects, providing experimental evidence of its ability to generalize to diverse skull defects and patient anatomies.

The learned deep occupancy function can effectively represent the overall shape of the skull. To highlight this phenomenon, we evaluate the occupancy function at different voxel grid resolutions and then visualize in Figure 9 the corresponding meshes extracted by Marching Cubes. It can be observed how the surface becomes smoother as the resolution increases. Theoretically, our approach is not bound to an upper resolution limit. Thus, we performed an additional experiment by evaluating the occupancy function on a 1024^3 resolution voxel grid. The resulting mesh is also

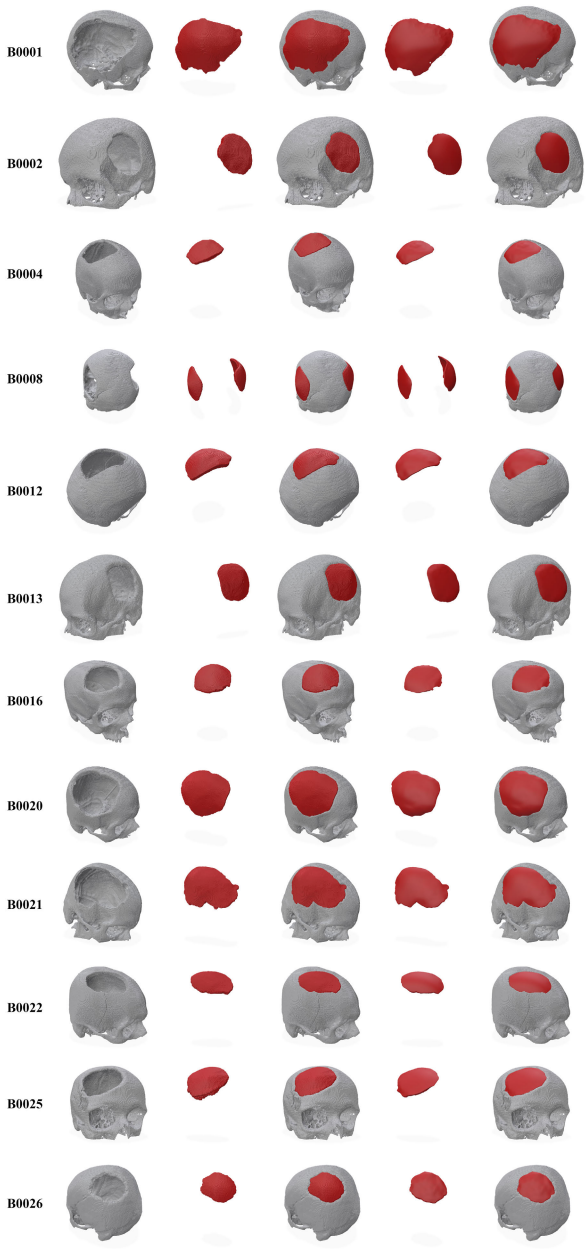

FIGURE 12. Qualitative results on the real craniotomy skull of MUG500+ [15]. Starting from the left column we report: the ID; the defective skull; the isolated predicted implant; the predicted implant with the partial skull; the implant designed by the clinician; and the implant designed by the clinician with the partial skull (best in colors and zoomed).

shown in Figure 9. Moreover, to extract the implant, the defective binary mask can be up-sampled via interpolation to 1024^3 resolution. In Figure 10 we show the corresponding triangular meshes for two examples. It can be noticed that the higher-resolution implant has a smoother surface and exhibits more fine-grained details. Such detailed, high-quality results are possible only with our implicit, continuous formulation, which can generate implants at arbitrary resolutions, as methods based on U-Nets or diffusion models are bound to work at a fixed output resolution.

Although the proposed approach can effectively reconstruct the defective region of the skull, some errors have been observed, as shown in Figure 11. The first row displays a large, randomly-shaped defect encompassing the parietal, frontal, and sphenoid bones of the skull. Such a defect is unrealistic in a cranioplasty scenario, particularly due to the need to incorporate the eye orbit into the implant design. We can notice how the predicted and ground truth implants are quite different, which is reasonable as there can be many possible plausible reconstructions for large defect areas. The second row of Figure 11 shows, instead, the predicted and ground truth implants for a Fronto-Orbital defect type. It can be seen that the reconstructed zygomatic bone is thinner compared to the ground truth and some fine-grained details are missing.

Finally, the generalization capability of our solution has been evaluated by running the proposed model on the subset of the MUG500+ dataset [15] with real craniotomies. Given the diversity of positions, sizes, and contours of the defects, the model pre-trained on the SkullBreak dataset has been employed. It is worth pointing out that the quality of the segmentation of the CT scans in this dataset is significantly lower than in SkullBreak and SkullFix. Nonetheless, as shown by the qualitative results reported in Figure 12, our approach can effectively generalize to a new dataset and generate plausible implants that can be used as a starting point for clinicians to speed up the implant design workflow. Even in the presence of two defects in the *B*0008 case, our model can produce a good reconstruction from which we extracted two implants by selecting two clusters in the post-processing stage.

V. CONCLUSION

We have introduced a novel approach for automatic cranial implant generation. Ours is the first solution to explore the use of neural fields to tackle this problem. The learned deep occupancy function is conditioned on the point cloud of the partial defective skull. This allows for limiting the memory footprint of the method while being able to generate high-quality implants with fast inference time. We achieve comparable scores with respect to SOTA methods that in contrast need longer generation time, higher memory consumption or costly ad-hoc post-processing. Peculiarly, our solution based on neural fields enables generating seamlessly high-resolution patient-specific implants due to the continuous nature of the learned occupancy function. This leads to high-resolution triangular meshes that can result in higher quality implants. Moreover, the model can generalize on real clinical cases coming from a different dataset exhibiting robustness against possible biases from the training set. We hope that the insights and experimental findings offered by our work will foster further research on the deployment of neural fields for automated and patient-specific design of surgical implants. As future work, we plan to investigate automatic implant design with a focus on the splanchnocranium portion of the skull, a region far

richer in fine-grained details (e.g. in the jaw and teeth) which may require the use of periodic activation functions [26] in both the conditioning encoder and the implicit decoder to more accurately reconstruct high-frequency details.

Finally, since this paper presents a study on the feasibility and effectiveness of implicit representations for the task of implant design and does not present a clinical study of a real deployment of the system, it has not addressed the ethical implications of such deployment. Yet, we deem it worth discussing such implications to conclude our study. In particular, during the deployment of deep-learning-based Software as Medical Devices (DL-based SaMD) one major concern is the lack of transparency of deep learning models. Therefore, it would be important that such models are deployed in a real clinical setting only as an aid tool for clinicians to speed up the implant design process and not to take decisions in place of them. Having a clinician to always review the output of the model and take the final decision on the implant shape would certainly help to mitigate the potential ethical concerns of its deployment. In addition, the implant predicted by our solution is only an initial guess that should then be properly adapted by the clinician for thickness, rim width, surface area, and volume [6]. The second major issue in the deployment of an AI/ML-based SaMD is the potential bias inherited from the training datasets. Therefore, before clinical deployment, the biases and limitations of the trained models should be thoroughly characterized. Moreover, to prevent biases, a large and diverse training dataset should be collected, encompassing CT scans across various genders, ethnicities, and age groups. We note that our model shows good generalization capabilities on the MUG500+ dataset, collected in Austria, despite being trained solely on scans from the SkullBreak dataset, which comprises CT scans from hospitals in India. This suggests that it already naturally exhibits some robustness to ethnicity biases of the training set.

ACKNOWLEDGMENT

The work carried out when Stefano Mazzocchetti was with the Department of Computer Science and Engineering (DISI), University of Bologna, 40126 Bologna, Italy.

REFERENCES

[1] X. Chen, L. Xu, X. Li, and J. Egger, "Computer-aided implant design for the restoration of cranial defects," *Sci. Rep.*, vol. 7, no. 1, p. 4199, Jun. 2017.

[2] Y.-W. Chen, C.-T. Shih, C.-Y. Cheng, and Y.-C. Lin, "The development of skull prosthesis through active contour model," *J. Med. Syst.*, vol. 41, no. 10, pp. 1–10, Oct. 2017.

[3] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier, "Development and validation of deep learning algorithms for detection of critical findings in head CT scans," 2018, *arXiv:1803.05854*.

[4] L. D. Luigi, A. Cardace, R. Spezialetti, P. Z. Ramirez, S. Salti, and L. D. Stefano, "Deep learning on implicit neural representations of shapes," in *Proc. ICLR*, 2023, pp. 1–20.

[5] J. Egger, M. Gall, A. Tax, M. Ücal, U. Zefferer, X. Li, G. von Campe, U. Schäfer, D. Schmalstieg, and X. Chen, "Interactive reconstructions of cranial 3D implants under MeVisLab as an alternative to commercial planning software," *PLoS One*, vol. 12, no. 3, Mar. 2017, Art. no. e0172694.

[6] Z. Fishman, J. G. Mainprize, G. Edwards, O. Antonyshyn, M. Hardisty, and C. M. Whyne, "Thickness and design features of clinical cranial implants—What should automated methods strive to replicate?" *Int. J. Comput. Assist. Radiol. Surgery*, vol. 1, no. 1, pp. 1–10, 2024.

[7] P. Friedrich, J. Wolleb, F. Bieder, F. M. Thieringer, and P. C. Cattin, "Point cloud diffusion models for automatic implant generation," 2023, *arXiv:2303.08061*.

[8] M. A. Fuessinger, S. Schwarz, C.-P. Cornelius, M. C. Metzger, E. Ellis III, F. Probst, W. Semper-Hogg, M. Gass and S. Schlager, "Planning of skull reconstruction based on a statistical shape model combined with geometric morphometrics," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 13, no. 4, pp. 519–529, Apr. 2018.

[9] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. NIPS*, vol. 33, 2020, pp. 6840–6851.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[11] O. Kodym, J. Li, A. Pepe, C. Gsaxner, S. Chilamkurthy, J. Egge, and M. Španěl, "SkullBreak/SkullFix—Dataset for automatic cranial implant design and a benchmark for volumetric shape learning tasks," *Data Brief*, vol. 35, Apr. 2021, Art. no. 106902.

[12] A. Kroviakov, J. Li, and J. Egger, "Sparse convolutional neural network for skull reconstruction," in *Proc. Towards Automatization Cranial Implant Design Cranioplasty*. Cham, Switzerland: Springer, 2021, pp. 80–94.

[13] J. Li et al., "Towards clinical applicability and computational efficiency in automatic cranial implant design: An overview of the AutoImplant 2021 cranial implant design challenge," *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102865.

[14] J. Li, C. Gsaxner, A. Pepe, D. Schmalstieg, J. Kleesiek, and J. Egger, "Sparse convolutional neural network for high-resolution skull shape completion and shape super-resolution," *Sci. Rep.*, vol. 13, no. 1, p. 20229, Nov. 2023.

[15] J. Li et al., "MUG500+: Database of 500 high-resolution healthy human skulls and 29 craniotomy skulls and implants," *Data Brief*, vol. 39, Dec. 2021, Art. no. 107524.

[16] J. Li, G. von Campe, A. Pepe, C. Gsaxner, E. Wang, X. Chen, U. Zefferer, M. Tödtling, M. Krall, H. Deutschmann, U. Schäfer, D. Schmalstieg, and J. Egger, "Automatic skull defect restoration and cranial implant generation for cranioplasty," *Med. Image Anal.*, vol. 73, Oct. 2021, Art. no. 102171.

[17] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[18] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Seminal Graphics*. New York, NY, USA: ACM, 1998, pp. 347–353.

[19] H. Mahdi, A. Clement, E. Kim, Z. Fishman, C. M. Whyne, J. G. Mainprize, and M. R. Hardisty, "A u-net based system for cranial implant design with pre-processing and learned implant filtering," in *Towards Automatization Cranial Implant Design Cranioplasty II*. Cham, Switzerland: Springer, 2021.

[20] S. Peng, C. Jiang, Y. Liao, M. Niemeyer, M. Pollefeys, and A. Geiger, "Shape as points: A differentiable Poisson solver," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 13032–13044.

[21] S. Peng, M. Niemeyer, L. Mescheder, M. Pollefeys, and A. Geiger, "Convolutional occupancy networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2020, pp. 523–540.

[22] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[24] W. Semper-Hogg et al., "Virtual reconstruction of midface defects using statistical shape models," *J. Cranio-Maxillofacial Surgery*, vol. 45, no. 4, pp. 461–466, Apr. 2017.

[25] H. Shi and X. Chen, "Cranial implant design through multiaxial slice inpainting using deep learning," in *Towards the Automatization of Cranial Implant Design in Cranioplasty*. Cham, Switzerland: Springer, 2020, pp. 28–36.

[26] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proc. NIPS*, 2020, pp. 7462–7473.

[27] H. Sulakhe, J. Li, J. Egger, and P. Goyal, "CranGAN: Adversarial point cloud reconstruction for patient-specific cranial implant design," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 603–608.

[28] Y. Volpe, R. Furferi, L. Governi, F. Uccheddu, M. Carfagni, F. Mussa, M. Scagnet, and L. Genitori, "Surgery of complex craniofacial defects: A single-step AM-based methodology," *Comput. Methods Programs Biomed.*, vol. 165, pp. 225–233, Oct. 2018.
[29] M. Wodzinski, M. Daniol, and D. Hemmerling, "Improving the automatic cranial implant design in cranioplasty by linking different datasets," in *Towards the Automatization of Cranial Implant Design in Cranioplasty II*. Cham, Switzerland: Springer, 2021, pp. 29–44.
[30] M. Wodzinski, M. Daniol, D. Hemmerling, and M. Socha, "High-resolution cranial defect reconstruction by iterative, low-resolution, point cloud completion transformers," in *Proc. Int. Conf. Med. Image Comput. Computer-Assisted Intervent*, 2023, pp. 333–343.
[31] M. Wodzinski, M. Daniol, M. Socha, D. Hemmerling, M. Stanuch, and A. Skalski, "Deep learning-based framework for automatic cranial defect reconstruction and implant modeling," *Comput. methods programs Biomed.*, vol. 226, Jun. 2022, Art. no. 107173.
[32] Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, and S. Sridhar, "Neural fields in visual computing and beyond," *Comput. Graph. Forum*, vol. 41, no. 2, pp. 641–676, May 2022.
[33] B. Yang, K. Fang, and X. Li, "Cranial implant prediction by learning an ensemble of slice-based skull completion networks," in *Towards the Automatization of Cranial Implant Design in Cranioplasty II*. Cham, Switzerland: Springer, 2021, pp. 95–104.
[34] X. Yu, Y. Rao, Z. Wang, Z. Liu, J. Lu, and J. Zhou, "PoinTr: Diverse point cloud completion with geometry-aware transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 12498–12507.
[35] C. Yuksel, "Sample elimination for generating Poisson disk sample sets," *Comput. Graph. Forum*, vol. 34, no. 2, pp. 25–32, May 2015.
[36] L. Zhou, Y. Du, and J. Wu, "3D shape generation and completion through point-voxel diffusion," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5806–5815.

STEFANO MAZZOCCHETTI received the B.S. and M.S. degrees in automation engineering from the University of Bologna, Italy. He is currently pursuing the Ph.D. degree with the eDIMES Lab—Laboratory of Bioengineering, Department of Medical and Surgical Sciences, University of Bologna.

From 2022 to 2023, he was a Research Fellow with the Department of Computer Science and Engineering (DISI), University of Bologna. His research interests include medical imaging analysis, 3D computer vision, and augmented reality in minimally invasive surgeries.

MIRKO BEVINI received the degree in medicine and surgery from the University of Bologna, in 2019. He is currently working at the Oral and Maxillo-Facial Surgery unit of S.Orsola-Malpighi University Hospital, Bologna, Italy. He is a Maxillofacial Surgery Resident at the University of Parma, Italy. Formerly a Research Fellow at the same institution, his research topics focus on the application of machine learning and CAD-CAM technology to maxillofacial surgical procedures.

GIOVANNI BADIALI carries out his research activity in the field of maxillofacial surgery, with particular interest for surgery of congenital and acquired malformations of the face (adult/child/unfavorable outcomes of trauma) and skeletal therapy of sleep apneas. More specifically, he is interested in computer-assisted diagnosis and planning of surgical treatment, 3D-printing for surgery, computerized biomodels of normal and pathological anatomy of the face, surgical navigation, and augmented reality as innovative navigation technology for the head-neck district (European Project VOSTARS—HORIZON 2020).

GIUSEPPE LISANTI is currently an Associate Professor with the Department of Computer Science and Engineering, University of Bologna. He has co-authored more than 40 papers in the most prestigious journals and conferences in computer vision, multimedia analysis, pattern recognition, and machine learning. He is involved in research collaborations with other research centers, both national and international. He has actively been participating in various roles in several research projects and a number of industry-funded projects. He received the Best Paper Award from the IEEE Computer Society Workshop on Biometrics, in 2017.

LUIGI DI STEFANO (Member, IEEE) received the Ph.D. degree in electronic engineering and computer science from the University of Bologna, in 1994. He was a Scientific Consultant for major companies, in the fields of computer vision and machine learning. He is currently a Full Professor with the Department of Computer Science and Engineering, University of Bologna, where he founded and leads the Computer Vision Laboratory (CVLab). He is the author of more than 150 papers and several patents. His research interests include image processing, computer vision, and machine/deep learning. He is a member of the IEEE Computer Society and IAPR-IC.

SAMUELE SALTI is currently an Associate Professor with the Department of Computer Science and Engineering (DISI), University of Bologna, Italy. In 2020, he co-founded the start-up eyecan.ai. He has co-authored more than 60 publications and seven international patents. His main research interest includes computer vision, in particular 3D computer vision, and machine/deep learning applied to computer vision problems. He was awarded the Best Paper Award Runner-Up at 3DIMPVT 2011 and 3DV 2021. He was also nominated several times outstanding reviewer at top vision and machine learning conferences, such as CVPR, NeurIPS, ICCV, and ICML.

Open Access funding provided by 'Alma Mater Studiorum - Università di Bologna' within the CRUI CARE Agreement