



ARTICLE

From Imperfection to Perfection: Advanced 3D Facial Reconstruction Using MICA Models and Self-Supervision Learning

Thinh D. Le, Duong Q. Nguyen, Phuong D. Nguyen and H. Nguyen-Xuan*

CIRTECH Institute, HUTECH University, Ho Chi Minh City, 72308, Viet Nam

*Corresponding Author: H. Nguyen-Xuan. Email: ngx.hung@hutech.edu.vn

Received: 30 July 2024 Accepted: 03 September 2024 Published: 27 January 2025

ABSTRACT

Research on reconstructing imperfect faces is a challenging task. In this study, we explore a data-driven approach using a pre-trained MICA (MetrIC fAce) model combined with 3D printing to address this challenge. We propose a training strategy that utilizes the pre-trained MICA model and self-supervised learning techniques to improve accuracy and reduce the time needed for 3D facial structure reconstruction. Our results demonstrate high accuracy, evaluated by the geometric loss function and various statistical measures. To showcase the effectiveness of the approach, we used 3D printing to create a model that covers facial wounds. The findings indicate that our method produces a model that fits well and achieves comprehensive 3D facial reconstruction. This technique has the potential to aid doctors in treating patients with facial injuries.

KEYWORDS

3D face reconstruction; self-supervised learning; face defect; 3D printed prototypes

1 Introduction

Currently, people who are injured due to traffic accidents, occupational accidents, birth defects, and diseases have lost a part of their body. As illustrated, many parts of the human body are injured and lose their natural function as shown in Fig. 1. Herein, the head area, which is soft and most vulnerable to external impacts, has the highest injury rate and probability of death. Reconstructing the wound and creating a product to cover it is crucial for patients to reduce the risk of re-injury, improve aesthetics, and increase confidence [1,2]. This research introduces an alternative way to reduce the risk of re-injury and avoid impact when moving. It also helps to increase aesthetics and confidence for people with injured faces. In addition, with the advancements in 3D printing and scanning technologies in rapid prototyping, the reconstruction process of organs has become faster, more accurate, and safer [3,4]. However, recreating the functions and the true shapes of the facial features, which have been damaged by injury, remains highly challenging and requires multiple trials in the treatment process.

The reconstruction of human body parts is divided into three main research directions including [5]: the reconstruction of the body [6,7], the reconstruction of the extremities [8,9], and the reconstruction of the head as shown in Fig. 1. The reconstruction of extremities has garnered significant research interest as compared to other areas of the human body such as the reconstruction of bones



in the hands [10–12] and the legs [13]. The reconstruction of the body is often divided into two parts: the reconstruction of the spine [14–16] and the reconstruction of the ribs [17–19]. Previous studies have been focusing on the reconstruction of the body and the extremities as shown in Fig. 2 because the main reason is that these areas are rather less susceptible to injury compared to other areas of the human body and they are also relatively easy to implement, easy to assemble, and less risky for patients in the treatment process. The upper body can be divided into two main groups, namely, the reconstruction of the face [20,21] and the reconstruction of the skull [22]. Due to the high fatality rate associated with injuries sustained in this region, it is considered the most perilous area for humans. Consequently, research on this region of the human body is relatively scarce compared to other areas.

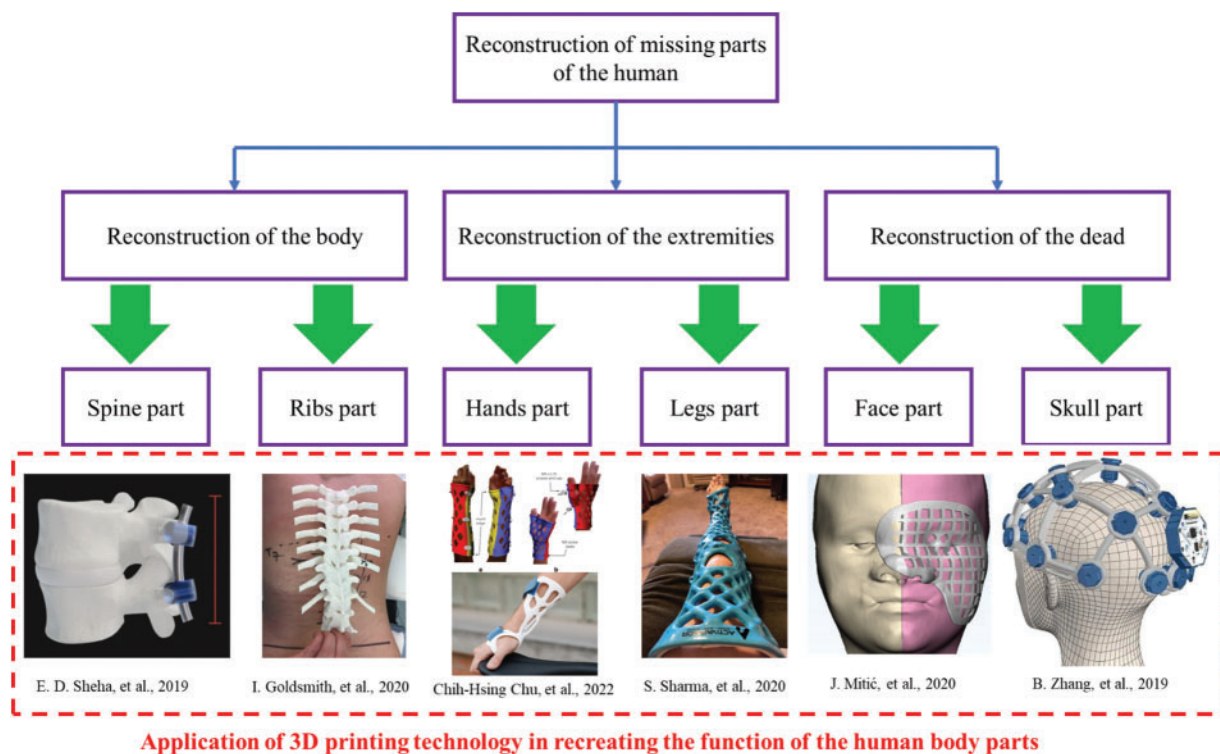


Figure 1: Replacing some parts of the human body with 3D printing technology [6,10,12,13,20,22]

Having analyzed the unique risks associated with the patient's head area, we recognize the importance of conducting thorough studies to enhance patient safety and optimize outcomes in this critical area. A study on patients with head trauma treated at the Walter Reed National Military Medical Center was conducted, as described by Farber et al. [23–25]. The application of microsurgery techniques to treat severe head wounds in a war zone was implemented for the first time. This study explored various methods to heal these injuries to achieve the best possible outcome for patients. In addition, Zhou et al. [26,27] proposed an extended forehead flap method to reconstruct the deformities of the nose or the central points of the face. The method was performed on 22 patients, including 13 men and 9 women. The outcome was that 17 people had burns and 5 people had other injuries. This method of using forehead-expanded flaps was shown to be capable of reconstructing the deformities of the nose and mouth in patients with only one extensive treatment. This flap can be flexibly adjusted to fit the size of the deformity right in the middle of the face. Wang et al. [28] used a combination of virtual surgical planning and 3D printing in reconstructing facial deformities on the jawbone. Study

data were collected from the jaw structure of patients in the period from 2013 to 2020. The patients were divided into two groups, in which group 1 (20 people) used the above technique and group 2 (14 people) used conventional manual surgery. The results indicate that the present approach offers numerous advantages to patients receiving the treatment. The positive outcomes of the aforementioned studies have established a new trend in providing care for those who have been injured in hazardous areas. The proposed method generates great potential for regenerating and rejuvenating various body parts. However, it is important to note that the study is only limited to an experimental setting and has not been implemented practically yet.

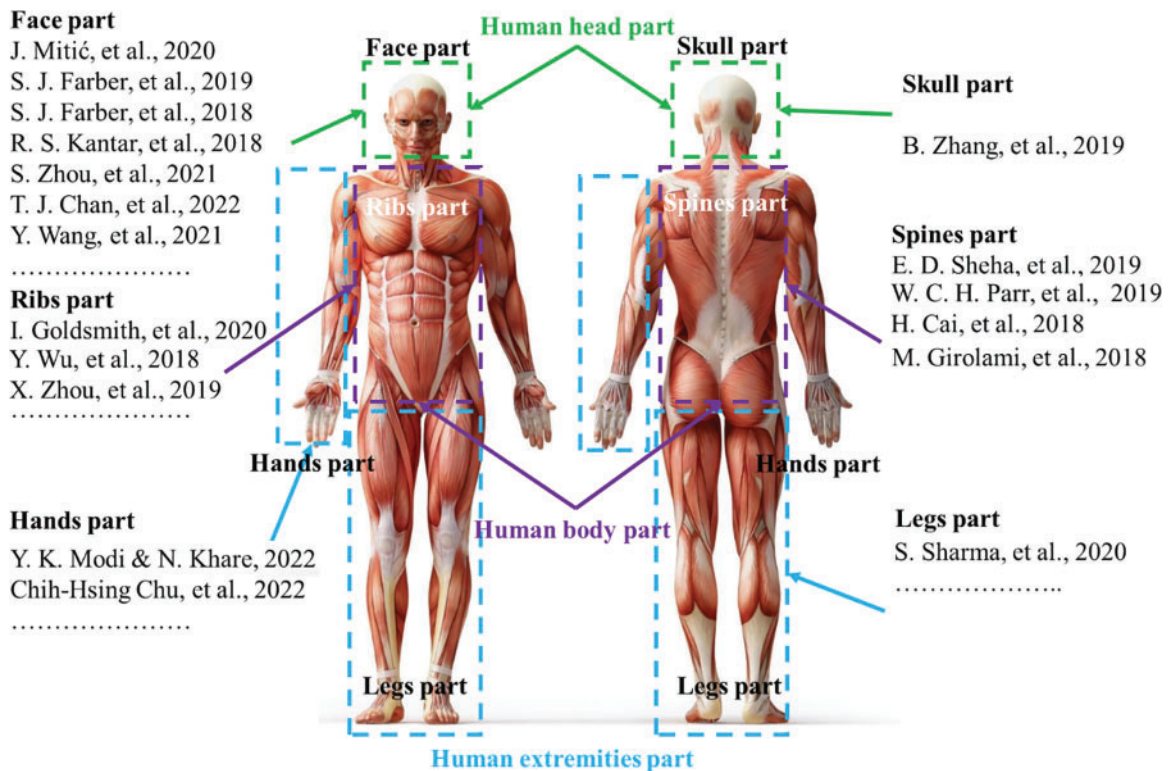


Figure 2: Some studies on reconstruction of parts of the human body [10–20,22–28]

In the previous study [29], we obtained promising results in 3D reconstructing face defects. We also identified the limitations of the published methods for regenerating new body parts to replace damaged ones and proposed an effective solution. Specifically, we implemented a process using 3D printing technology and deep learning to extract the filling to aid in masking for patients with imperfect facial defects. A general framework of the wound covering model was depicted in Fig. 3. This approach can assist patients with facial trauma regain their confidence in daily life while also protecting and cleaning the wound from external elements. Additionally, we introduced a dataset that is capable of covering multiple facial trauma cases. However, the deep learning model used in the previous work required significant computational resources for training due to its complexity, resulting in training times of several days, and presenting a challenge to conducting research in this field.

In the present study, we contribute the following:

- Using the MICA model [30] for faster reconstruction of imperfect 3D facial data;

- Proposing a new training strategy by integrating the pre-trained MICA model with self-supervised learning to improve 3D facial reconstruction;
- Assessing the effectiveness of the proposed methods and discussing the challenges of filling in missing parts for patients with facial trauma.

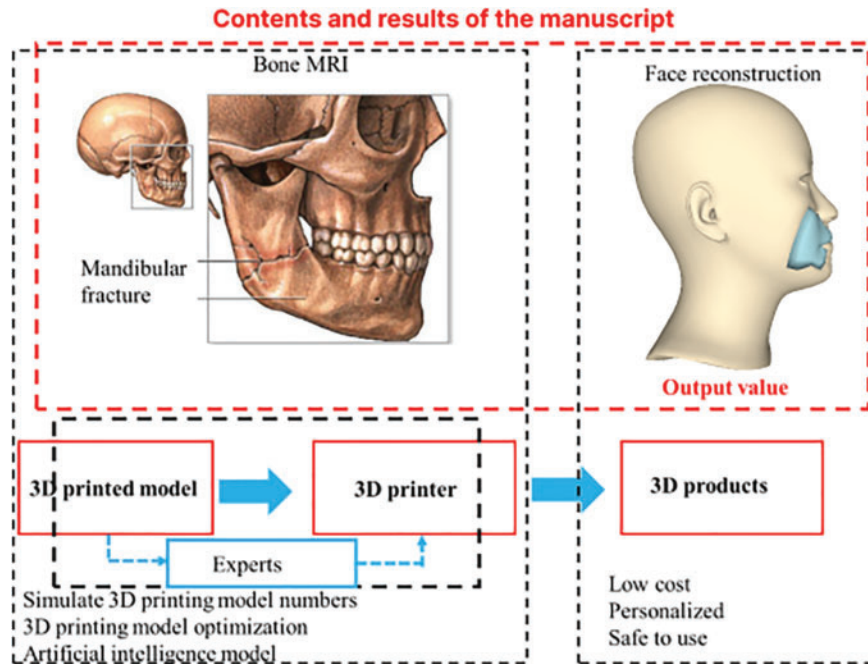


Figure 3: The proposed strategy and the results obtained from this study

2 Application of the MICA Model for Imperfect 3D Facial Reconstruction

In this section, we present the MICA model [30] for imperfect 3D facial reconstruction. Additionally, the geometric loss function is presented to evaluate the error of the reconstruction model. The imperfect 3D facial reconstruction process using the MICA model is detailed in Fig. 4.

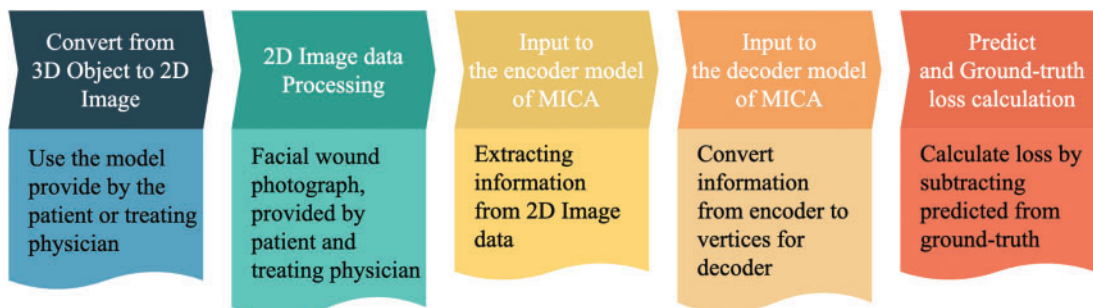


Figure 4: The MICA model's imperfect 3D facial reconstruction process

Among 3D face reconstruction models, the MICA model stands out as effective due to its ability to analyze facial features and to reconstruct finer details compared with the previous methods.

Additionally, it performs well on evaluation datasets with high reliability. The MICA model is applied with 2D image input, which is highly suitable as it requires fewer computing resources and is more cost-effective than 3D input formats. The MICA model includes two main components: Identity Encoder (IE) and Geometry Decoder (GD) as shown in Fig. 5. We modify the inputs and outputs in the process of reconstructing the patient's injured head area as shown in Fig. 6. The input is a 2D image, which shows the actual patient's face with the wound. The output is a 3D mesh restored to its pre-injury state.

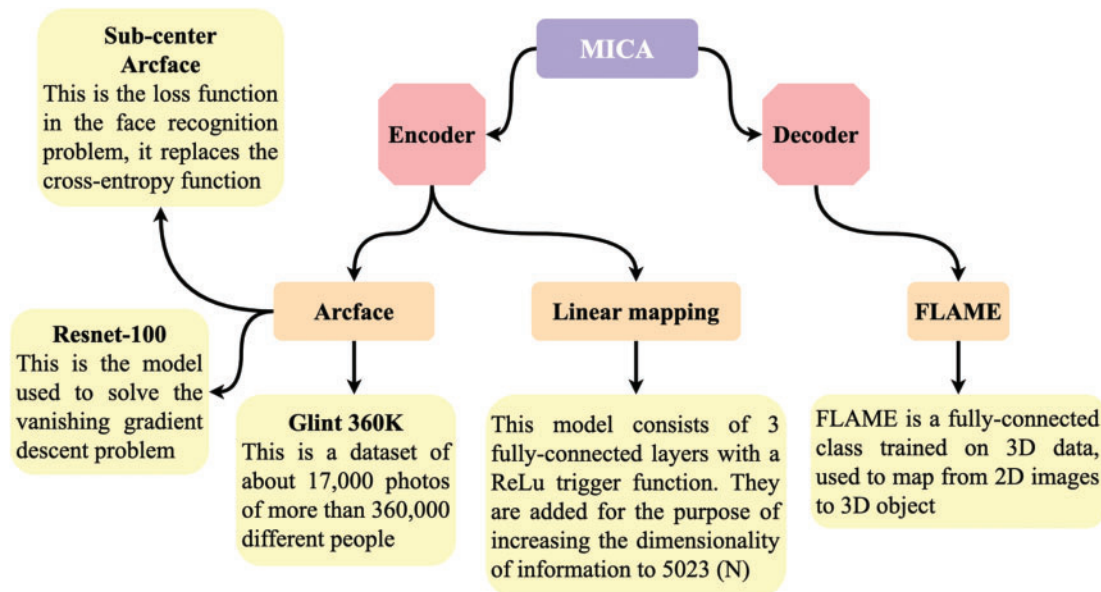


Figure 5: A framework of the MICA model

2.1 Identity Encoder

IE is formed by two main components: the ArcFace model and the Linear Mapping model. The goal of Encoder is to extract identity-specific information from an input 2D image using two components: ArcFace and Linear Mapping. First, the ArcFace model determines the level of similarity between the input image and the reference image of the same identity and then generates a feature vector to capture identity-specific information. Second, the Linear Mapping model maps this feature vector to a higher dimensional space to improve the discriminative power of the feature representation.

- ArcFace: The ArcFace architecture (a model developed based on ResNet-100) [31–33] uses a shortcut to map input from previous layers to following layers, thereby avoiding the phenomenon of vanishing gradients [34]. This is important for updating model weights and enabling the model to learn the best features. At the same time, This approach is widely recognized for its effectiveness, as demonstrated by its application in several studies [33,34]. These works highlight the superiority of additive angular margin loss in achieving more discriminative features compared to traditional methods. Moreover, additional research [35,36] has further validated its advantages, particularly in improving recognition accuracy under challenging conditions. This is necessary when the input images have only one color, which makes it extremely difficult to identify the identity differences among the people in the photos. This model is trained with Glint 360k data [37] consisting of about 170 million images of 360

thousand people from 4 different continents. The study uses the knowledge learned from this model to enrich the results of the problem.

- Linear mapping (\mathcal{M}): At the end of the model, an additional block of three fully connected layers with the ReLU activation function is used to process the information and adjust its size to fit the Decoder components when the information is passed on to them. The pre-trained model will also be used to train the data in the study further.

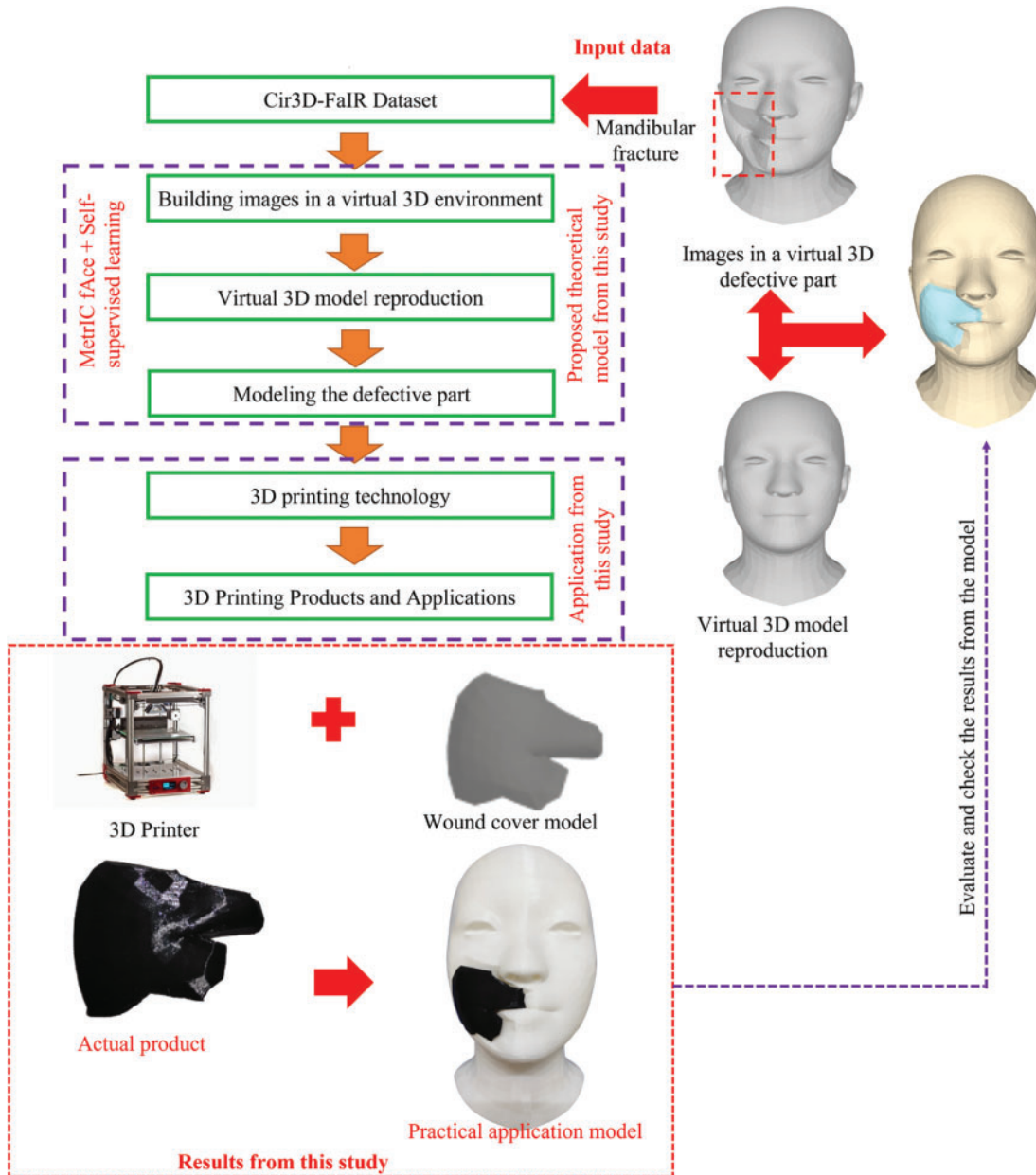


Figure 6: The proposed approach

2.2 Geometry Decoder

In the GD, the MICA model utilizes FLAME [38–40], which is the well-known model for transferring feature information from 2D data vectors to anthropometric morphology of faces. This is a model that has been trained on 33,000 precisely aligned 3D faces. In MICA, this model is further trained on many other datasets including LYHM [41], and Stirling [42]. This extensive training allows the Decoder component to reconstruct important features such as nose shape, size, and face thickness. As previously discussed in the preceding section, the FLAME model is a well-known model for transferring features, and it is employed within the Decoder part of the MICA model. By incorporating incomplete facial data into the transfer learning process, we can capitalize on the robust and well-optimized pre-trained MICA model. This approach is not only suitable for our study but also demonstrates its efficacy in harnessing the full potential of the pre-existing MICA model, thus enhancing the overall performance of the 3D face reconstruction process.

In [30], ArcFace architecture is expanded by a small mapping network \mathcal{M} that maps the ArcFace features to their latent space, which can then be interpreted by GD:

$$\mathbf{z} = \mathcal{M}(\text{ArcFace}(I)), \quad (1)$$

where $\mathbf{z} \in \mathbb{R}^{300}$ and I is the RGB input image. Research by Zielonka et al. [30] used FLAME as a GD, which consists of a single linear layer:

$$\mathcal{G}_{3DMM}(\mathbf{z}) = \mathbf{B} \times \mathbf{z} + \mathbf{A}, \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{3N}$ is the geometry of the average human face and $\mathbf{B} \in \mathbb{R}^{3N \times 300}$ contains the principal components of the 3DMM and N is the number of vertices of the output. Then, the loss function of the MICA model is defined as follows:

$$\mathcal{L} = \sum_{(I, \mathcal{G}) \in \mathcal{D}} |k_{\text{mask}}(\mathcal{G}_{3DMM}(\mathcal{M}(\text{ArcFace}(I))) - \mathcal{G})|, \quad (3)$$

where \mathcal{G} is the ground truth mesh, D is the data set and k_{mask} is a region dependent weight.

2.3 Geometric Loss Function

The geometric loss function is a critical metric in 3D modeling and reconstruction tasks, as it measures the discrepancy between predicted and ground truth models in three-dimensional space. It is commonly defined using the Euclidean distance, which provides a straightforward and intuitive metric for evaluating differences between corresponding points in 3D space [43]. For two points p and q in 3D space, the Euclidean distance is given by:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}. \quad (4)$$

In this study, we use the mean of all distances of vertices to compute the geometric loss. Let P and Q be the sets of vertices for two 3D models with n vertices each. The geometric loss function is then computed as:

$$L = \frac{1}{n} \sum_{i=1}^n d(p_i, q_i), \quad (5)$$

where p_i and q_i are corresponding vertices in the sets P and Q , respectively. This approach allows us to quantify the overall difference between the two 3D models by averaging the distances between corresponding vertices.

3 Recommendations from the Study

In this section, we propose incorporating self-supervised learning techniques into a pre-trained MICA model to enhance the results of imperfect 3D face reconstruction.

3.1 Self-Supervised Learning

The self-supervised learning algorithm was adopted to enhance the learning ability of the pre-trained Encoder component in our model. The central idea is that training on a vast dataset, followed by introducing a small amount of new data, can significantly alter model weights, resulting in poor performance. The self-supervised learning algorithm, named Swapping Assignments Between Views (SwAV), was developed by Caron et al. [44] and is derived from the contrastive instance learning algorithm [45]. These methods are based on traditional clustering techniques [45,46]. They typically operate offline, alternating between assigning steps to each model cluster. In this approach, image features from the entire dataset are initially clustered, and the assigned clusters are used to predict different views of the image during training. However, clustering-based methods are unsuitable for online learning models due to the extensive steps required to compute the necessary image features for clustering. This self-supervised learning technique can be conceptualized as comparing different views of the same image and the clusters to which their features are assigned. Specifically, the research determines the code from an augmented version of an image sample and predicts this code from other augmented versions of the same image. Given two image features \mathbf{z}_t and \mathbf{z}_s from two distinct augmented versions of the same image, their code \mathbf{q}_t and \mathbf{q}_s is computed by matching these attributes with a set of K prototypes $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K\}$. Subsequently, the study constructs the *swapped* prediction step using the loss function as follows:

$$L(\mathbf{z}_t, \mathbf{z}_s) = \ell(\mathbf{z}_t, \mathbf{q}_s) + \ell(\mathbf{z}_s, \mathbf{q}_t), \quad (6)$$

where the function $L(\mathbf{z}_t, \mathbf{z}_s)$ measures the fit between the attribute \mathbf{z}_t and \mathbf{q}_s the code.

3.1.1 Online Clustering

Online clustering is a technique used in self-supervised learning of visual representations to group similar image patches or features into clusters. This approach is based on the idea that similar patches or features will have analogous representations in the learned visual space. In online clustering, the visual features of numerous unlabeled images are extracted and clustered incrementally as new images become available. The process begins with an empty set of clusters. As new features are extracted from an image, they are either assigned to one of the existing clusters or used to create a new cluster. Each image \mathbf{x}_n is transformed into an augmented view \mathbf{x}_{nt} by applying a transformation t from the set \mathcal{T} of transformations. The augmented view is then mapped to a vector using a non-linear mapping function f_θ resulting in \mathbf{x}_{nt} . This feature is then projected onto the unit sphere, i.e., $\mathbf{z}_{nt} = f_\theta(\mathbf{x}_{nt})/\|f_\theta(\mathbf{x}_{nt})\|_2$. Subsequently, the code \mathbf{q}_{nt} is computed from this attribute by mapping \mathbf{q}_{nt} to a set of K trainable prototype vectors $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K\}$. The study represents these prototypes by a matrix \mathbf{C} with columns $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K\}$.

3.1.2 Swapped Prediction Problem

The loss function in Eq. (6) has two terms that establish a “swapped” prediction problem by predicting the code \mathbf{q}_t from the feature \mathbf{z}_s and \mathbf{q}_s from \mathbf{z}_t . Each term represents the cross-entropy loss function between the code and the probability obtained from using the softmax function that contains

the dot product between and all the prototypes in, i.e.,

$$\ell(\mathbf{z}_t, \mathbf{q}_s) = - \sum_k \mathbf{q}_s^{(k)} \log \mathbf{p}_t^{(k)}, \quad \text{where} \quad \mathbf{p}_t^{(k)} = \frac{\exp\left(\frac{1}{\tau} \mathbf{z}_t^\top \mathbf{c}_k\right)}{\sum_{k'} \exp\left(\frac{1}{\tau} \mathbf{z}_t^\top \mathbf{c}_{k'}\right)}, \quad (7)$$

where τ is the temperature parameter. Using this loss function for all images and pairs of augmentations yields a general loss function for the swapped prediction problem:

$$-\frac{1}{N} \sum_{n=1}^N \sum_{s,t \sim \mathcal{I}} \left[\frac{1}{\tau} \mathbf{z}_{nt}^\top \mathbf{C} \mathbf{q}_{ns} + \frac{1}{\tau} \mathbf{z}_{ns}^\top \mathbf{C} \mathbf{q}_{nt} - \log \sum_{k=1}^K \exp\left(\frac{\mathbf{z}_{nt}^\top \mathbf{c}_k}{\tau}\right) - \log \sum_{k=1}^K \exp\left(\frac{\mathbf{z}_{ns}^\top \mathbf{c}_k}{\tau}\right) \right]. \quad (8)$$

This loss function is minimized based on two \mathbf{C} prototype parameters and the parameter of the image Encoder f_θ , which is used to output the $(\mathbf{z}_{nt})_{n,t}$ attribute.

3.2 Computing Codes Online

To enable the study to function online as well as offline, codes $(\mathbf{q}_t, \mathbf{q}_s)$ are computed using only the image attributes within a batch. This approach allows the application of \mathbf{C} prototypes across different batches. SwAV employs a clustering method to group multiple versions of the prototypes. Codes are calculated using the \mathbf{C} prototypes in such a way that all samples in a specific image batch are evenly distributed among the prototypes. This even distribution ensures each code for different images in the batch is unique, avoiding trivial solutions where every image has the same code. Given the feature vector \mathbf{B} consisting of $\mathbf{Z} = \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_B$, this study aims to map them to the prototypes $\mathbf{C} = \mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K$. The mappings are represented as $\mathbf{Q} = \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_K$ and \mathbf{Q} is optimized to maximize the similarity between the properties and the prototypes:

$$\max_{\mathbf{Q} \in \mathcal{Q}} \text{Tr}(\mathbf{Q}^\top \mathbf{C}^\top \mathbf{Z}) + \varepsilon H(\mathbf{Q}), \quad (9)$$

where H is the entropy function, $H(\mathbf{Q}) = - \sum_{ij} \mathbf{Q}_{ij} \log \mathbf{Q}_{ij}$, and ε is the parameter that controls the smoothness of the mapping. It is observed that strong entropy regularization (i.e., using high ε) can often lead to a trivial solution, where all templates collapse into a unique representation and all prototypes have the same property. Therefore, ε should be fixed at a small value. To ensure equal partitioning, Asano et al. [46] constrained the matrix that belonged to the transportation polytope. To address this issue, we propose an adaptation of the solution proposed by [44], which involves constraining the transportation polytope to the minibatch when working with mini-batches as follows:

$$\mathcal{Q} = \left\{ \mathbf{Q} \in \mathbb{R}_+^{K \times B} \mid \mathbf{Q} \mathbf{1}_B = \frac{1}{K} \mathbf{1}_K, \mathbf{Q}^\top \mathbf{1}_K = \frac{1}{B} \mathbf{1}_B \right\}, \quad (10)$$

where $\mathbf{1}_K$ is a vector of ones with K dimensions. These constraints ensure that on average each prototype is selected at least B/K times in a batch. When a continuous solution for \mathbf{Q}^* in Eq. (10) is found, the discrete code can be calculated using the rounding process. In an online setting, we used mini-batches with discrete codes that performed worse than continuous codes. These soft codes \mathbf{Q}^* are the solution of Eq. (9) over the set \mathcal{Q} and can be written as:

$$\mathbf{Q}^* = \text{Diag}(\mathbf{u}) \exp\left(\frac{\mathbf{C}^\top \mathbf{Z}}{\varepsilon}\right) \text{Diag}(\mathbf{v}), \quad (11)$$

where \mathbf{u} and \mathbf{v} are the renormalization vectors in \mathbb{R}^K and \mathbb{R}^B , respectively. The vector renormalization is calculated using a small number of matrix multiplications and the iterative Sinkhorn-Knopp algorithm [47].

3.3 Post-Processing: Filling Extraction

To achieve completion of the wound healing process and effectively aid in concealing facial imperfections, we have employed the filling extraction technique, which was initially developed and discussed in one of our prior publications [29]. This approach is specifically tailored to analyze the wound-affected area on the face, isolate it, and extract it separately for subsequent utilization in 3D printing applications.

By employing this technique, our final printable product is an amalgamation of the original, imperfect face and the separately extracted wound portion. Notably, the extracted part is displayed in a distinct color to differentiate it from the rest of the facial structure, as demonstrated in Fig. 7. The combination of these elements allows for a more comprehensive understanding of the wound area and its impact on the overall facial reconstruction process.

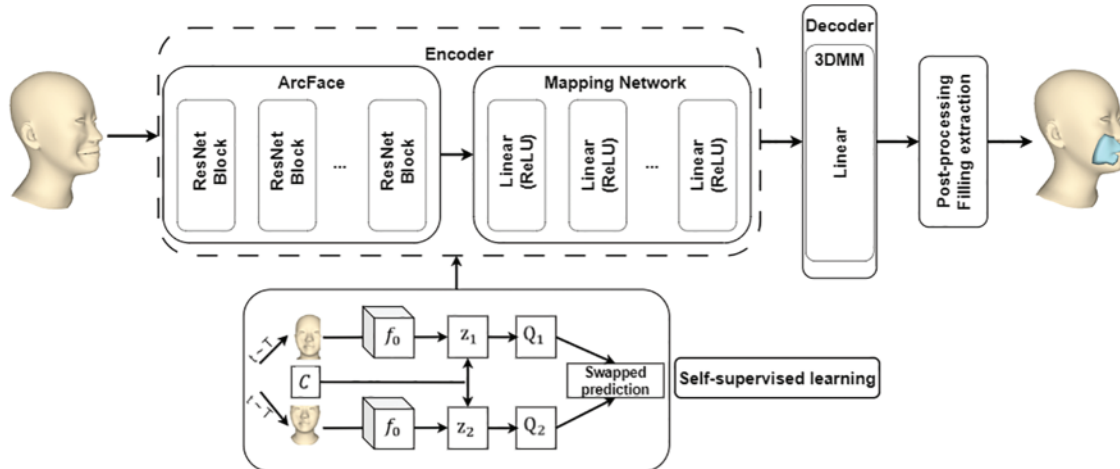


Figure 7: A model for supervised learning

4 Experiments

4.1 Building Dataset

To effectively implement the transfer learning and self-supervised learning processes within the MICA model, we have undertaken a comprehensive data preprocessing approach, which consists of two primary steps. Each step has been carefully designed to ensure the optimization of our model's performance and applicability in the context of reconstructing incomplete facial structures.

The first step focuses on preparing the input data in a 2D format. This decision is grounded in two main considerations. Firstly, utilizing a 2D input format is consistent with the MICA model's architectural requirements, thus allowing for seamless integration of the data. Secondly, the adoption of 2D input data presents an alternative method for facial reconstruction, which significantly simplifies the process for medical professionals. As a result, physicians are no longer required to provide 3D MRI (Magnetic Resonance Imaging) scans. They can rely on more readily available 2D images. This adaptation seeks to create a more conducive environment that enables healthcare practitioners to

efficiently engage with our model, ultimately delivering enhanced value and outcomes for patients. Additionally, patients can preview their facial appearance after treatment, providing them with a clear understanding of the expected results. The second step of the data pre-processing procedure involves employing augmentation techniques within the mathematical space to standardize the model's output. This standardization serves as the Ground Truth for the MICA model to calculate the loss value, ensuring that the reconstructed facial structures are accurate and reliable. By integrating these augmentation techniques, we aim to improve the model's overall performance and facilitate a more precise reconstruction of facial defects. The specific details of the input and output parameters are described in greater detail below:

- The 2D input format is obtained by capturing images from all 3D data with varying lighting directions. The 2D images have a resolution of 1024×1024 pixels and are captured from 3687 3D samples within our dataset, ensuring that the angle of capture reveals any scarring present.
- The MICA model's output, with 3D format, is modified to match with the Ground Truth. By default, MICA generates eyeballs for the output mesh but the Ground Truth mesh of the Cir3D-FaIR dataset has no eyeballs, so we have conducted eyeball removal procedures for the entire set of MICA model's outputs. Another problem is that 3D printing needs a watertight mesh to have a good processing step, but the output mesh of the MICA model has a hole that makes the mesh non-watertight, therefore the hole-fixing procedures need to be used. The computational techniques of eyeball removal and hole-fixing procedures introduced in our previous research [29] and employed in this space have been encapsulated within the Cirmesh library and made available in Python developed by our group, with the easy installation using the command: `pip install Cirmesh`. After that, the MICA model's output and the ground truth mesh could be compared with each other.

To prepare for the training process, the research divides the data into a 6:2:2 ratio for the training, validation, and test sets, respectively. During the data splitting procedure, the study ensures that images of different individuals are allocated to each dataset, preventing the exchange of facial information among three sets.

4.2 Determining the Set of Parameters

Runtime: The experiments were conducted on Google Colab Pro. The proposed model requires computational cost and runtime. The first epoch takes about 30 min for the training and validation process, while other epochs take only about 90 s to complete as the data is already stored.

Hyperparameters: In all experiments, this research fixed the version of hyperparameters. The optimal algorithm is Adam [47] whose initial learning rate value is 1×10^{-3} , which is kept constant throughout the training. At the same time, a batch size of 50 images is used for each epoch. Finally, the experimental process is conducted for 50 epochs for each case as shown in Fig. 8.

Fig. 8 indicates the change in the loss function value for each epoch in two cases: 20% data and 100% data. Both cases share a similar pattern of loss function change for the first 30 epochs. The case with 100% data shows faster convergence in the following 20 epochs. This suggests that the present model is highly stable.

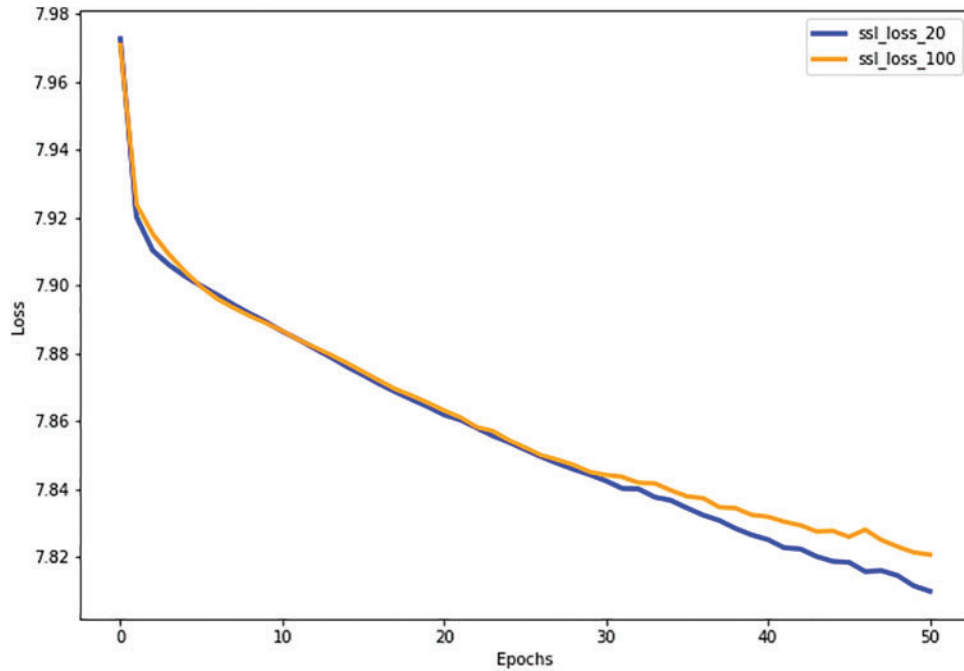


Figure 8: Loss function value corresponding to each epoch

4.3 Experiment Settings

We used a self-supervised algorithm for model training: We conducted a total of 5 experiments. In the first experiment, the study used the pre-trained model of MICA [30] with its enriched data. In the remaining cases, we utilized the self-supervised learning technique as mentioned in Section 3 to train the Encoder component before proceeding to the supervised learning process. The present method conducted 50 epochs with self-supervised learning in two main cases. As discussed earlier, our study initially used a sample of 20% of the training data and then used a larger sample to evaluate the loss of the proposed models. Fig. 9 shows the loss of self-supervised algorithm training when sampling 20% and 100% of the data. When approaching the 50th epoch, the loss function of the method with 20% of the data sample performed better than that with 100% of the data sample. This can be explained by the fact that when dealing with larger amounts of data, self-study may have some difficulties in adapting to the new data. The strategy of training the MICA model combined with self-supervised learning is described in Algorithm 1.

Algorithm 1: Self-supervised learning for MICA model

Input : $data_ssl \leftarrow$ wounded dataset to train self-supervised learning model

$model_MICA \leftarrow$ Best weight of MICA model trained on unwounded face data

Output : The loss function values of the self-supervised learning model training at each ratio

$n \leftarrow \text{len}(data_ssl)$ \triangleright number of data to train for the self-supervised learning model

$mean_loss_ratio \leftarrow \text{List}[]$ \triangleright List containing the losses of the self-supervised learning model training by ratio

$ratio \leftarrow [0.2, 0.4, 0.6, 0.8, 1]$

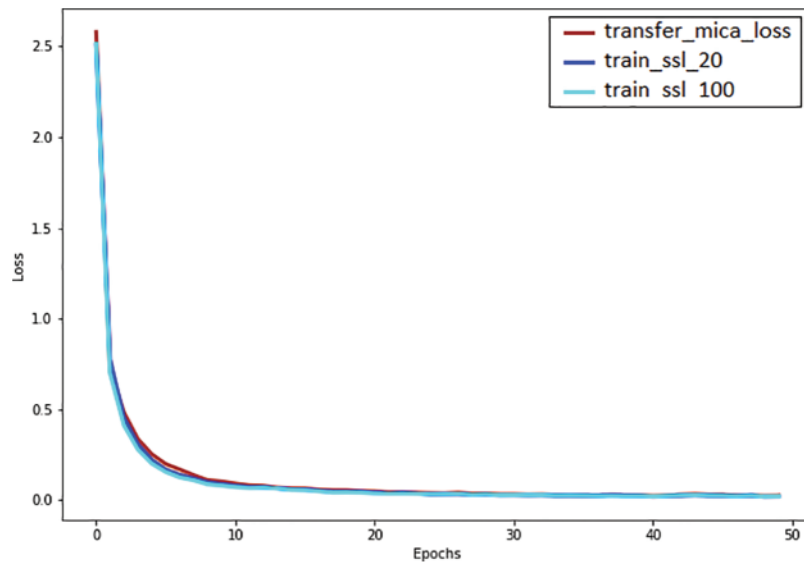
(Continued)

Algorithm 1 (continued)

```

for  $i \leftarrow 0$  to 4 do
     $rate \leftarrow ratio[i]$                                 ▷ Divide the dataset with wounds according to the
                                                            ratios
     $data\_ssl\_train \leftarrow data\_ssl[(rate - 0.2) \times n : rate \times n]$ 
     $loss\_ratio \leftarrow List[]$ 
    for  $epoch \leftarrow 1$  to 50 do
         $model\_ssl \leftarrow$  Training a self-supervised learning model on  $data\_ssl\_train$  from  $model\_MICA$ 
         $loss\_epoch \leftarrow$  The loss function of over each training epoch
         $loss\_ratio.add(loss\_epoch)$ 
    end for
     $mean\_loss \leftarrow \frac{1}{50} \times \sum loss\_ratio$ 
     $mean\_loss\_ratio.add(mean\_loss)$ 
end for
return  $mean\_loss\_ratio$ 

```

**Figure 9:** Loss of training set

During the training process, Fig. 10 shows that the loss function path of the training set of all three cases is stable and all values go below 0.5 at the 10th epoch. After that, the loss line decreased slightly and approached near zero at the 50th epoch. In all cases, the loss lines are not significantly different from each other. For the validation set, the loss lines seem to be volatile at the 10th epoch, and then become stable again as they move toward the 50th epoch.

In our recent research [29], the outcomes of the developed model were assessed using the geometric loss function. So, in this study, we explored the performance of a self-supervised learning model on a unique dataset, comparing it to the well-known transfer learning approach, specifically the Mica model. Using the geometric loss function for assessment, the results demonstrate the effectiveness of our approach in enhancing model performance. As shown in Table 1, the transfer learning method

yielded a loss of 0.36, indicating a reasonable level of accuracy in reconstructing facial injuries. However, when applying our self-supervised learning model to just 20% of the training data, we observed a reduction in loss to 0.3, showing an improvement over the transfer learning method. Upon expanding the dataset to 100% of the available data, our self-supervised learning model continued to improve, reducing the loss marginally further. This consistent improvement, even with increased data, showcases the potential of our model in handling larger datasets efficiently.

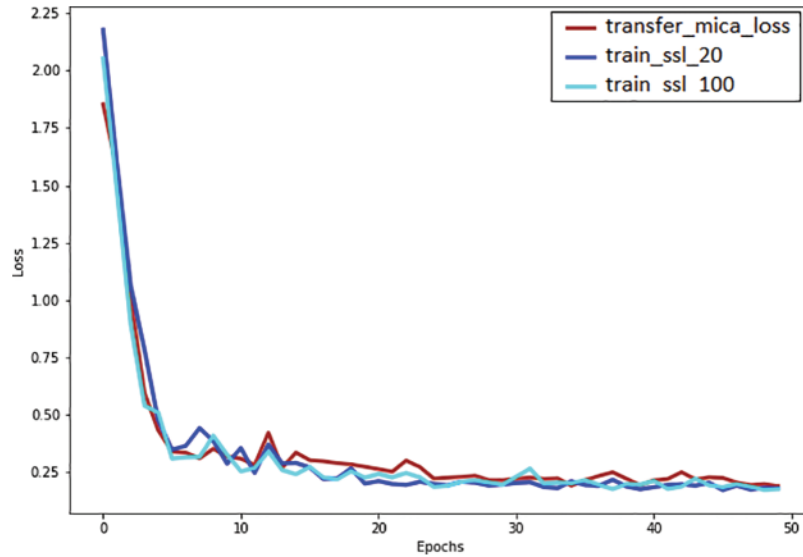


Figure 10: Loss of validation set

Table 1: Geometric distance function values of the outcomes

Method	Mean distance
Transfer learning	0.3642
Self-supervised learning with 20% data	0.3353
Self-supervised learning with 100% data	0.3247

These findings suggest that employing a limited amount of data in the pre-trained model using the self-supervised learning technique is a feasible approach that offers superior efficiency compared to the transfer learning method. Notably, our datasets comprise only 3687 images, accounting for less than one-tenth of the data in the purely pre-trained MICA model. As a result, our approach effectively leverages the capabilities of the trained MICA model while remaining applicable to the current research context.

Fig. 11 shows some examples illustrating the model's ability to reconstruct the structure of human facial wounds and the deviation between the output and the reality, with scar types divided into 4 main groups including The top of the head, Forehead, Chin and Nose. The Deviation column describes the difference between the Output column and the Ground Truth column with the unit of measure calculated in units of coordinates, specifically brighter regions signify larger errors, which can be identified as wounds. As shown in Fig. 11, wounds at the forehead position are the most elusive, producing the highest errors, and wounds around the nose area give extremely good results. For a

more detailed analysis of the results, we perform a statistic with the output of a total of 291 object files as input including 53 object files are The top of the head, 71 object files are Forehead, 65 object files are Chin and 102 object files are Nose as shown in Table 2. Specifically, the deviations in the coordinate units are converted to % and averaged by the number of each location, with the smallest mean error being the Nose region with 1.2% and the largest average error being the Forehead region with 7.8%, the other two regions are almost similar. This result satisfied the requirements of the Hospital cooperating in the field of plastic surgery, it supports the doctor in most stages of facial regeneration and the doctor’s intervention is only to bring completeness.

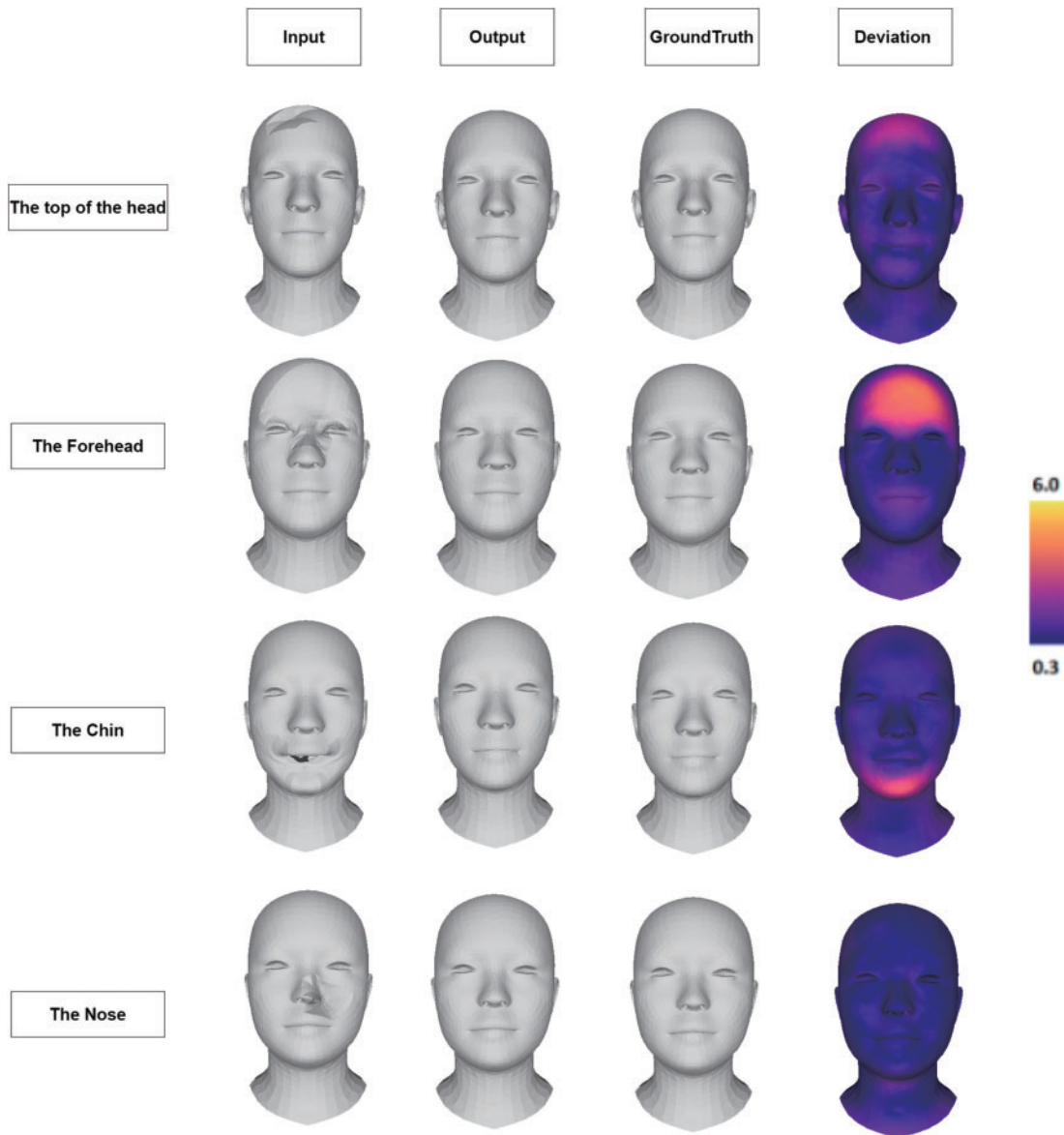


Figure 11: Deviation between obtained results and referred ones

Table 2: Geometric distance error for each scar location

Scar location	Quantity	Mean distance error %
The top of the head	53	6.5
Forehead	71	7.8
Chin	65	4.7
Nose	102	1.2

Although some errors will inevitably occur while reconstructing injured areas, the proposed model successfully reconstructed the majority of the face, leaving virtually no unaffected regions. These findings suggest that the model is proficient in addressing the challenges of reconstructing defects stemming from wounds, even when solely relying on 2D input. The results of this study have been applied to the creation of experimental products using 3D printing technology, as shown in Fig. 12, to be utilized in laboratory experiments. The plastic material used is PLA (Polylactic Acid), a versatile biopolymer. PLA is easily synthesized from abundant renewable resources and is biodegradable. It has demonstrated significant potential as a biomaterial in various healthcare applications, including tissue engineering and regenerative medicine [48]. However, it is important to note that PLA is not suitable for implantation into human skin. Our 3D printing results are intended to demonstrate its efficiency and to provide an example of its potential real-world applications. Due to the high costs associated with bioprinting technologies that can create implantable materials for humans, we hope to conduct experiments shortly using materials that are compatible with human skin grafting. The 3D printer we use features an enclosed chamber to maintain a stable temperature during the printing process. The printing procedure always begins with cleaning the hot end. We print at a lower speed of about 60 mm/s, and the PLA printing temperature is set at a high 205°C. The flow rate is increased by 5% above normal. The 3D printer is placed on a stable surface, and all components are checked to ensure they do not cause vibrations during printing. Additionally, we use an ADXL345 3-axis accelerometer to compensate for any vibrations and reduce issues with poor interlayer adhesion and rough surfaces. We monitor print quality by visually inspecting every 3D printed sample. If there is significant poor interlayer adhesion or surface roughness, the samples are reprinted. The results of this study reveal significant potential for practical applications, especially in creating models that mask flaws in the human face.

4.4 Limitation

The approach in this study provides relatively high performance in the task of imperfect 3D facial reconstruction. However, the current results may not fully capture the complexities of real-world scenarios due to the significant differences in individual faces. Creating a real-world dataset for this purpose is both challenging and costly. Therefore, in this study, we have adopted an open-solution approach. Our method has been tested on synthetic data, with the expectation that it can be applied to real-world data once available, thereby enhancing the reliability and generalizability of our conclusions.

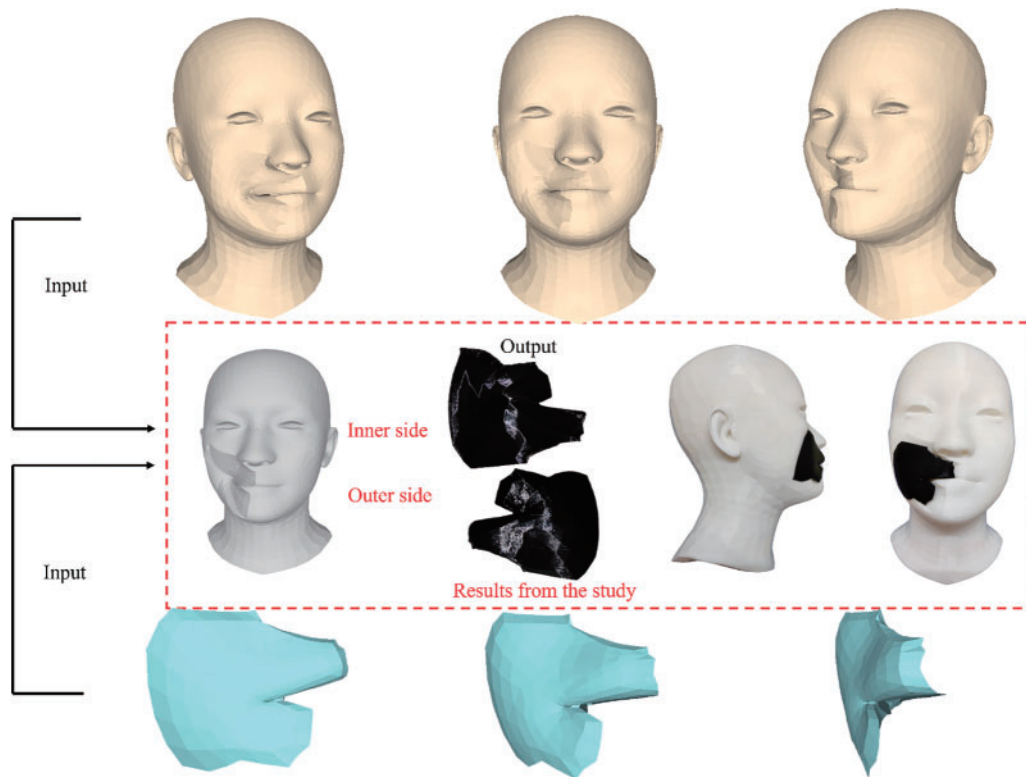


Figure 12: 3D printed prototypes

5 Conclusions

This paper has demonstrated the advantages of integrating a pre-trained model relying on a self-supervised learning method with defective 3D face reconstruction. The proposed approach significantly benefits the underlying problem by capitalizing on the capabilities of the existing models. The noteworthy findings can be pointed out as follows:

- In this research, we re-introduce the Cir3D-FaIR dataset, meticulously designed to address a distinct patient-specific issue, imperfect faces. This dataset is unique in its composition and application, marking its inaugural use in the training of AI models. Given the novelty of the CirFair3D dataset and the specificity of the problem it aims to solve, identifying an existing neural network model for direct comparison posed a significant challenge. As a consequence, conventional benchmarks for performance comparison were not readily applicable.
- In reason of this, we adopted a self-supervised learning approach tailored to the unique characteristics of the CirFair3D dataset. The absence of pre-existing models directly comparable to our scenario underscores the innovative aspect of our methodology. This study, therefore, not only introduces a new dataset but also pioneers a self-supervised learning approach tailored to leverage the full potential of this novel dataset.
- By strategically integrating pre-trained models and self-supervised learning, it is possible to achieve an efficient system where new data can be incorporated without requiring a complete retraining process. This approach yields reliable results without demanding an extensive amount

of training data. Additionally, it enhances the system's performance without adding complexity to the model. These results demonstrate the feasibility of reconstructing incomplete faces using only 2D images, providing surgeons and orthopedic trauma specialists with additional data and improving procedures to better support their patients.

- The training time significantly reduced from a duration of several days to a matter of mere hours, facilitating the extraction of facial features from incomplete faces. Moreover, this reduction allows for the implementation of the training process on platforms like Google Colab without necessitating a high-performance computer.
- Although the obtained results exhibit a relatively wide range of errors, from 0.3 to 6 as shown in Fig. 11, which is larger than the error margins observed in our previous research, the average error remains at a manageable level of 0.32, indicating that the overall effect is not significantly compromised. This can be easily understood, as the method attempts to extract information from a 2D format, which inherently lacks depth information compared to the 3D format.

Acknowledgement: Authors would like to thank Hoang Nguyen and Associate Professor Nguyen Thanh Binh for their recommendation for his assistance in this study. We would like to thank the Vietnam Institute for Advanced Study in Mathematics (VIASM) for the hospitality during our visit in 2023 when we started to work on this paper.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Thinh D. Le, Duong Q. Nguyen, H. Nguyen-Xuan, Phuong D. Nguyen; data collection: Phuong D. Nguyen; analysis and interpretation of results: Thinh D. Le, Duong Q. Nguyen, Phuong D. Nguyen; draft manuscript preparation: Thinh D. Le, Duong Q. Nguyen, H. Nguyen-Xuan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Our source code and data can be accessed at <https://github.com/SIMOGROUP/ImperfectionFacialReconstruction> (accessed on 01 May 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Meisel EM, Daw PA, Xu X, Patel R. Digital technologies for splints manufacturing. In: Proceedings of the 38th International MATADOR Conference, 2022; p. 475–88.
2. Asanovic I, Millward H, Lewis A. Development of a 3D scan posture-correction procedure to facilitate the direct-digital splinting approach. *Virt Phys Prototyp*. 2019;14(1):92–103. doi:10.1080/17452759.2018.1500862.
3. Cheng K, Liu Y, Wang R, Zhang J, Jiang X, Dong X, et al. Topological optimization of 3D printed bone analog with PEKK for surgical mandibular reconstruction. *J Mech Behav Biomed Mater*. 2020 Jul;107:103758.
4. Popescu D, Zapciu A, Tarba C, Laptoiu D. Fast production of customized three-dimensional-printed hand splints. *Rapid Prototyp J*. 2020;26(1):134–44. doi:10.1108/RPJ-01-2019-0009.

5. Yin C, Zhang T, Wei Q, Cai H, Cheng Y, Tian Y, et al. Tailored surface treatment of 3D printed porous Ti6Al4V by microarc oxidation for enhanced Osseointegration via optimized bone in-growth patterns and interlocked bone/implant Interface. *Bioact Mater.* 2022 Jan;7:26–38. doi:10.1021/acsami.6b05893.
6. Goldsmith I MBBS, MD FRCSE, RCPS FRCS, CTh FRCS. Chest wall reconstruction with 3D printing: anatomical and functional considerations. *Innovat: Technol Techniq Cardiothor Vascul Surg.* 2022 Jun;17(3):191–200. doi:10.1177/15569845221102138.
7. Alvarez AG, Evans PL, Dvoglanski L, Goldsmith I. Design, additive manufacture and clinical application of a patient-specific titanium implant to anatomically reconstruct a large chest wall defect. *Rapid Prototyp J.* 2021 Jan;27(2):304–10. doi:10.1108/RPJ-08-2019-0208.
8. Poier PH, Weigert MC, Rosenmann GC, Carvalho MGR, Ulbricht L, Foggianto JA. The development of low-cost wrist, hand, and finger orthosis for children with cerebral palsy using additive manufacturing. *José Aguiomar Foggianto.* 2021;73(3):445–53. doi:10.1007/s42600-021-00157-0.
9. Li J, Tanaka H. Feasibility study applying a parametric model as the design generator for 3D-printed orthosis for fracture immobilization. *3D Print Med.* 2018;4(1):1–15. doi:10.1186/s41205-017-0024-1.
10. Sheha ED, Gandhi SD, Colman MW. 3D printing in spine surgery. *Ann Transl Med.* 2019;7(5):164. doi:10.21037/atm.2019.08.88.
11. Modi YK, Khare N. Patient-specific polyamide wrist splint using reverse engineering and selective laser sintering. *Mat Technol.* 2022;37(2):71–8. doi:10.1080/10667857.2020.1810926.
12. Chih-Hsing Chu IJW, Sun JR, Liu CH. Customized designs of short thumb orthoses using 3D hand parametric models. *Assist Technol.* 2022;34(1):104–11. doi:10.1080/10400435.2019.1709917.
13. Sharma S, Singh J, Kumar H, Sharma A, Aggarwal V, Amoljit Singh Gill NJ, et al. Utilization of rapid prototyping technology for the fabrication of an orthopedic shoe inserts for foot pain relieve using thermo-softening viscoelastic polymers: a novel experimental approach. *Meas Control.* 2020;53(3–4):519–30. doi:10.1177/0020294019887194.
14. Girolami M, Boriani S, Bandiera S, Barbanti-Bródano G, Ghermandi R, Terzi S, et al. Biomimetic 3D-printed custom-made prosthesis for anterior column reconstruction in the thoracolumbar spine: a tailored option following en bloc resection for spinal tumors. *Europ Spine J.* 2018;27:3073–83. doi:10.1007/s00586-018-5708-8.
15. Parr WCH, Burnard JL, Wilson PJ, Mobbs RJ. 3D printed anatomical (bio)models in spine surgery: clinical benefits and value to health care providers. *J Spine Surg.* 2019;5(4):549–60. doi:10.21037/jss.2019.12.07.
16. Cai H, Liu Z, Wei F, Yu M, Xu N, Li Z. 3D printing in spine surgery. In: *Advances in experimental medicine and biology.* Singapore: Springer; 2018. p. 1–27. doi:10.1007/978-981-13-1396-7_27.
17. Goldsmith I, Evans PL, Goodrum H, Warbrick-Smith J, Bragg T. Chest wall reconstruction with an anatomically designed 3D printed titanium ribs and hemi-sternum implant. *3D Print Med.* 2020;6(1):1–26. doi:10.1186/s41205-020-00079-0.
18. Wu Y, Chen N, Xu Z, Zhang X, Liu L, Wu C, et al. Application of 3D printing technology to thoracic wall tumor resection and thoracic wall reconstruction. *J Thorac Dis.* 2018;10(12):6880–90. doi:10.21037/jtd.2018.11.109.
19. Zhou XT, Zhang DS, Yang Y, Zhang GI, Xie ZX, Chen MH, et al. Analysis of the advantages of 3D printing in the surgical treatment of multiple rib fractures: 5 cases report. *J Cardiothorac Surg.* 2019;14(105):1–7. doi:10.1186/s13019-019-0930-y.
20. Mitić J, Vitković N, Manić M, Trajanović M. Reconstruction of the missing part in the human mandible. In: *In European Conference on Computer Vision, 2020;* p. 202–5.
21. Cai M, Zhang S, Xiao G, Fan S. 3D face reconstruction and dense alignment with a new generated dataset. *Displays.* 2021;70:102094. doi:10.1016/j.displa.2021.102094.
22. Zhang B, Sun H, Wu L, Ma L, Xing F, Kong Q, et al. 3D printing of calcium phosphate bioceramic with tailored biodegradation rate for skull bone tissue reconstruction. *Bio-Desi Manufact.* 2019;2:161–71. doi:10.1007/s42242-019-00046-7.

23. Farber SJ, Latham KP, Kantar RS, Perkins JN, Rodriguez ED. Reconstructing the face of war. *Mil Med.* 2019 Jul;184(7–8):236–46. doi:10.1093/milmed/usz103.
24. Farber SJ, Kantar RS, Diaz-Siso JR, Rodriguez ED. Face transplantation: an update for the united states trauma system. *J Craniofac Surg.* 2018 Jun;29(4):832–8. doi:10.1097/SCS.0000000000004615.
25. Kantar RS, Rifkin WJ, Cammarata MJ, Maliha SG, Diaz-Siso JR, Farber SJ, et al. Single-stage primary cleft lip and palate repair: a review of the literature. *Ann Plast Surg.* 2018 Nov;81(5):619–23. doi:10.1097/SAP.0000000000001543.
26. Zhou SB, Gao BW, Tan PC, Zhang HZ, Li QF, Xie F. A strategy for integrative reconstruction of midface defects using an extended forehead flap. *Facial Plast Surg Aesth Med.* 2021 Nov;23(6):430–6. doi:10.1089/fpsam.2020.0484.
27. Chan TJ, Long C, Wang E, Prisman E. The state of virtual surgical planning in maxillary reconstruction: a systematic review. *Oral Oncol.* 2022 Oct;133:106058. doi:10.1016/j.oraloncology.2022.106058.
28. Wang Y, Qu X, Jiang J, Sun J, Zhang C, He Y. Aesthetical and accuracy outcomes of reconstruction of maxillary defect by 3D virtual surgical planning. *Front Oncol.* 2021 Oct;11:718946. doi:10.3389/fonc.2021.718946.
29. Nguyen PD, Le TD, Nguyen DQ, Nguyen TQ, Chou LW, Nguyen-Xuan H. 3D Facial imperfection regeneration: deep learning approach and 3D printing prototypes. arXiv:2303.14381. 2023.
30. Zielonka W, Bolkart T, Thies J. Towards metrical reconstruction of human faces. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. *Computer Vision–ECCV 2022.* Cham: Springer Nature Switzerland; 2022. p. 250–69. doi:10.1007/978-3-031-19778-9_15.
31. Deng J, Guo J, Liu T, Gong M, Zafeiriou S. Sub-center ArcFace: Boosting face recognition by large-scale noisy web faces. In: *Computer Vision–ECCV 2020.* Cham: Springer; 2020. p. 741–57. doi:10.1007/978-3-030-58621-8_43.
32. Zhao J, Yan S, Feng J. Towards age-invariant face recognition. *IEEE Trans Pattern Anal Mach Intell.* 2020 Jul;44(1):474–87. doi:10.1109/TPAMI.2020.3011426.
33. Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: additive angular margin loss for deep face recognition. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019; Long Beach, CA, USA.* doi:10.1109/CVPR.2019.00482.
34. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; Las Vegas, NV, USA.* doi:10.1109/CVPR.2016.90.
35. Kajla NI, Missen MMS, Luqman MM, Coustaty M, Mehmood A. Additive angular margin loss in deep graph neural network classifier for learning graph edit distance. *IEEE Access.* 2020 Nov;9:201752–61. doi:10.1109/ACCESS.2020.3035886.
36. Wang X, Wang S, Chi C, Zhang S, Mei T. Loss function search for face recognition. In: *Proceedings of the 37th International Conference on Machine Learning, 2020; PMLR.*
37. An X, Zhu X, Gao Y, Xiao Y, Zhao Y, Feng Z, et al. Partial FC: training 10 million identities on a single machine. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2021; Montreal, QC, Canada.*
38. Sanyal S, Bolkart T, Feng H, Black MJ. Learning to regress 3D face shape and expression from an image without 3D supervision. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019; Long Beach, CA, USA.*
39. Cudeiro D, Bolkart T, Laidlaw C, Ranjan A, Black MJ. Capture, learning, and synthesis of 3D speaking styles. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019; Long Beach, CA, USA.*
40. Li T, Bolkart T, Black MJ, Li H, Romero J. Learning a model of facial shape and expression from 4D scans. *ACM Trans Graph.* 2017;36(6):194–201. doi:10.1145/3130800.3130813.

41. Dai H, Pears N, Smith W, Duncan C. Statistical modeling of craniofacial shape and texture. *Int J Comput Vis.* 2020;128(2):547–71. doi:10.1007/s11263-019-01260-7.
42. Feng ZH, Huber P, Kittler J, Hancock P, Wu XJ, Zhao Q, et al. Evaluation of dense 3D reconstruction from 2D face images in the wild. In: *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018; Xi'an, China.
43. Zhou Y, Wu C, Li Z, Cao C, Ye Y, Saragih J, et al. Fully convolutional mesh autoencoder using efficient spatially varying kernels. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020; Red Hook, NY, USA: Curran Associates Inc.; p. 9251–62.
44. Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. In: *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020; Vancouver, BC, Canada.
45. Wu Z, Xiong Y, Yu SX, Lin D. Unsupervised feature learning via non-parametric instance discrimination. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018; Munich, Germany.
46. Asano YM, Rupprecht C, Vedaldi A. Self-labelling via simultaneous clustering and representation learning. arXiv:1911.05371. 2019.
47. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv:1711.05101. 2019.
48. DeStefano V, Khan S, Tabada A. Applications of PLA in modern medicine. *Eng Regenerat.* 2020 Jan;1(5):76–87. doi:10.1016/j.engreg.2020.08.002.