Training β-VAE by Aggregating a Learned Gaussian Posterior with a Decoupled Decoder

Jianning Li[1,2], Jana Fragemann[2], Seyed-Ahmad Ahmadi[3], Jens Kleesiek[2] , and Jan Egger[1,2]
1. IKIM, University Medicine Essen, Germany
2. ICG, Graz University of Technology, Austria
3. NVIDIA, Munich, Germany
Jianning.Li@uk-essen.de

MAD MICCAI Workshop 2022

MICCAI2022 Singapore

Problem

The balance of the reconstruction loss and the kullback-leibler divergence loss in β-VAE

$$\mathcal{L}(\theta, \phi; \mathbf{x}, \mathbf{z}, \beta) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \beta\, D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$

Large beta: continuous, Gaussian and more likely disentangled latent space.
Small beta: good reconstruction.

- A good generative model shall meet both requirements.
- Given different datasets, beta usually needs to be specifically tuned (tedious).
- It is not necessarily guaranteed that a proper beta can be found for a specific dataset, as the two losses are inherently antagonistic and at times cannot be optimized jointly.

Dataset

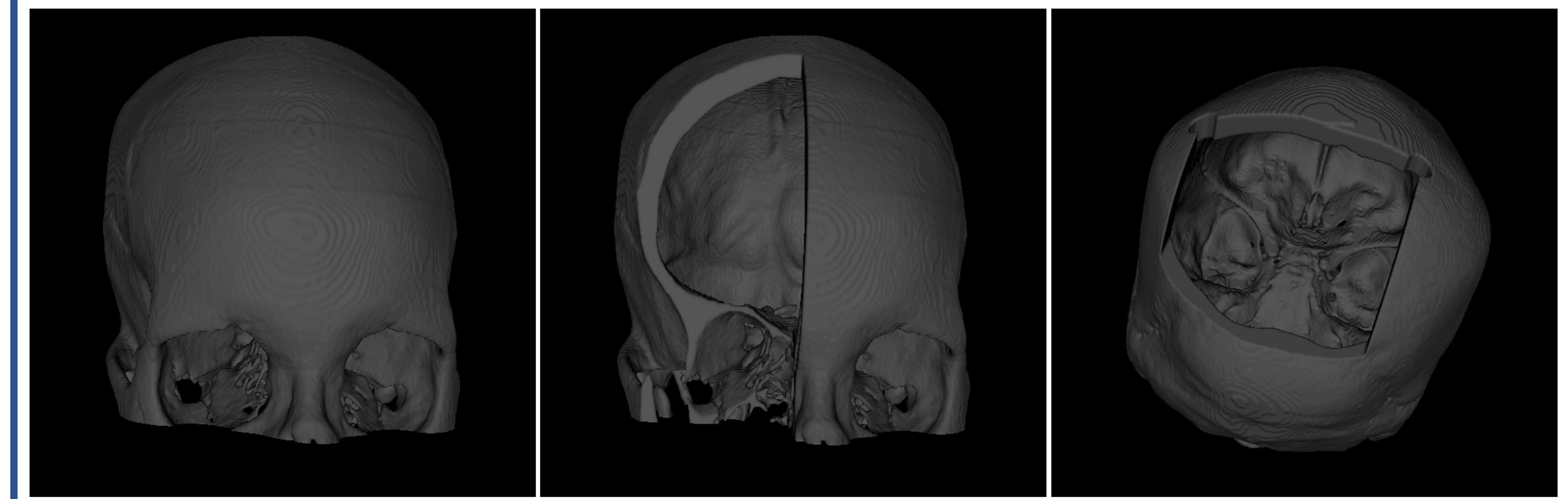

Fig.1. The dataset contains 100 samples for each of the three types of skulls - complete skulls (left), skulls with facial defects (middle) and skulls with cranial defects (right).

Method

Train a β-vae in two stages

Stage 1: train a β-VAE using a very large beta.
Stage 2: train a decoupled decoder using only the reconstruction loss.
Full VAE: aggregating the trained encoder, the sampling process from Stage 1 and the trained decoder from Stage 2.

$$z \sim q_\theta(z|x), x \sim p_\phi(x|z)$$

- The two losses are antagonistic and might not be optimized jointly for some datasets.

The expectation of the KLD term is a upper bound of the **mutual information** between the data and the latent variables.

$$E_{\sim p_{(x)}}[D_{KL}(q_\theta(z|x)||p(z))] \geqslant I_{\sim q_\theta(z|x)}(x, z)$$

Larger beta leads to a wider Gaussian distribution, and therefore higher uncertainty in the sampling phase, which makes it more difficult for the decoder to learn a proper reconstruction.

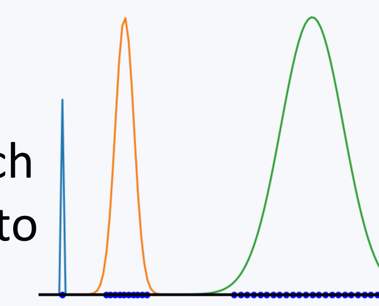

Fig.2. 1D Gaussian distribution with increasing standard deviation.

- Decouple the reconstruction loss from the kullback-leibler divergence loss, and meet the two requirements on the latent space and reconstruction in two separate stages. **Stage 1 aims for a continuous and Gaussian latent space, and the Stage 2 aims for an optimal reconstruction.**

Experiments & Results

Training Curves

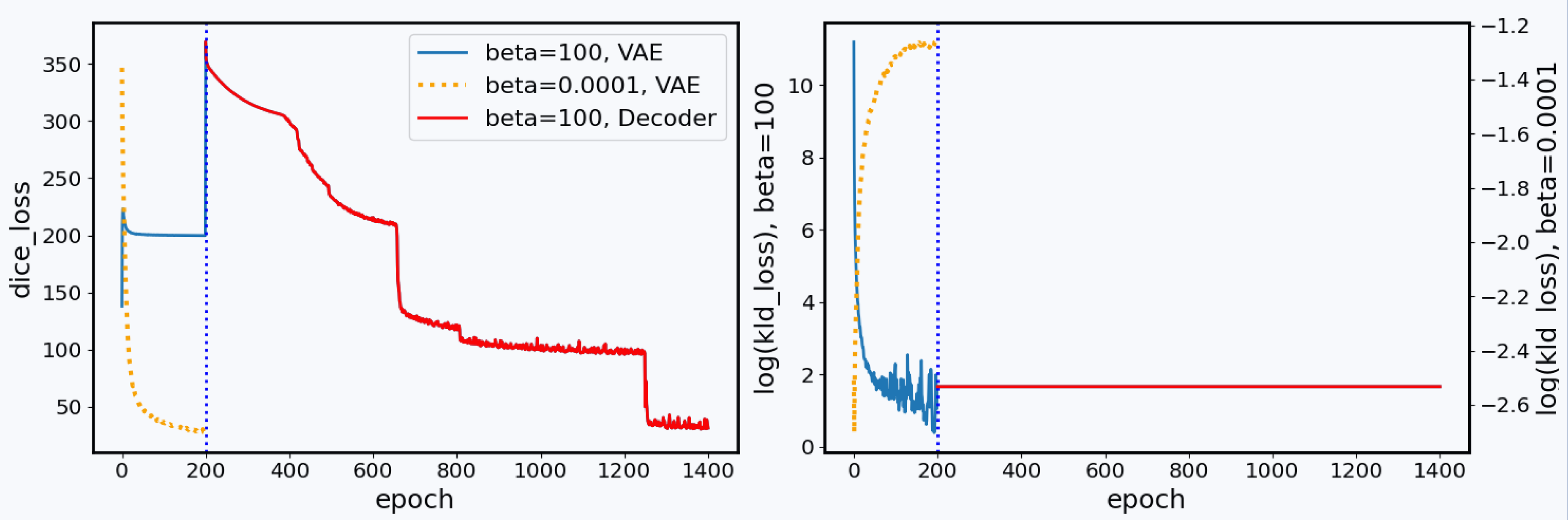

Fig. 3. training curve of the VAE and decoder regarding Dice (left) and KLD (right) loss.

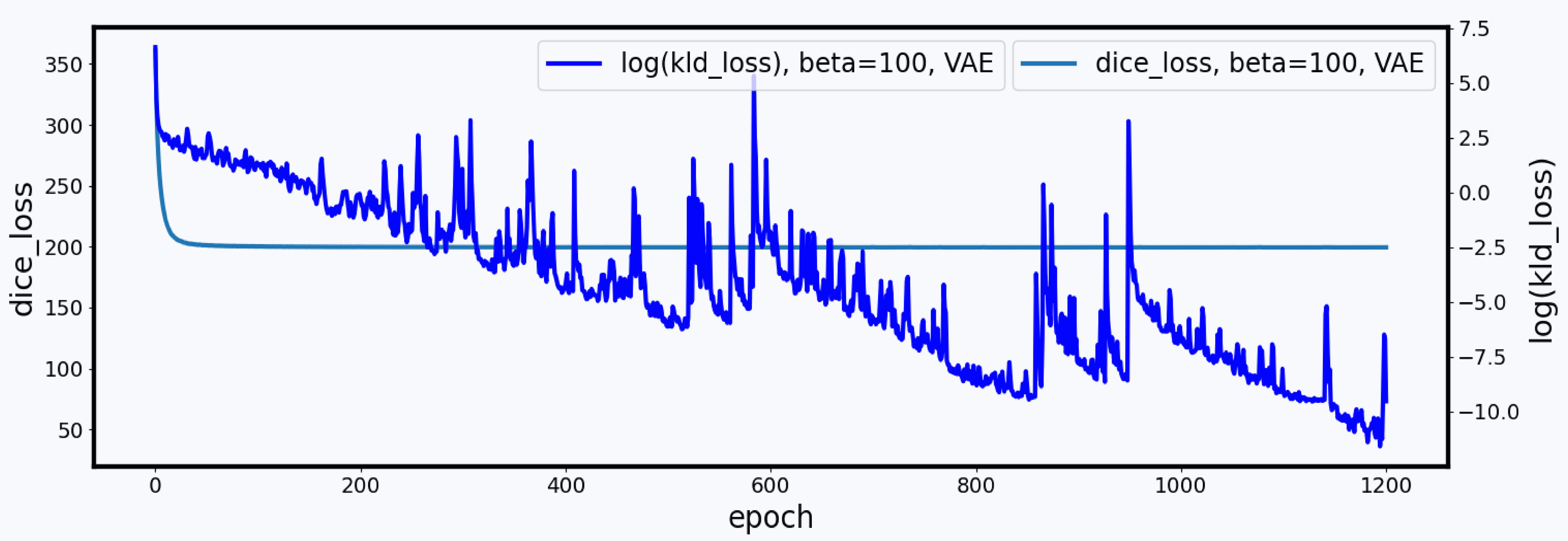

Fig.4. Training a VAE (the same VAE used in Fig.3.) for 1200 epochs under β = 100. The curve shows the Dice and KLD loss in the entire training process.

With γ = 0, we expect that the resulting latent variable would be decoded to the original defective sample (i.e., skull reconstruction). With γ = 1, we expect that decoding the latent variable yields a complete skull (i.e., skull shape completion)

$$z^{cr\rightarrow co} = \mathbf{z}_{ts} + \gamma DEV_{cr}$$
$$z^{fa\rightarrow co} = \mathbf{z}_{ts} + \gamma DEV_{fa}$$

Latent Distributions

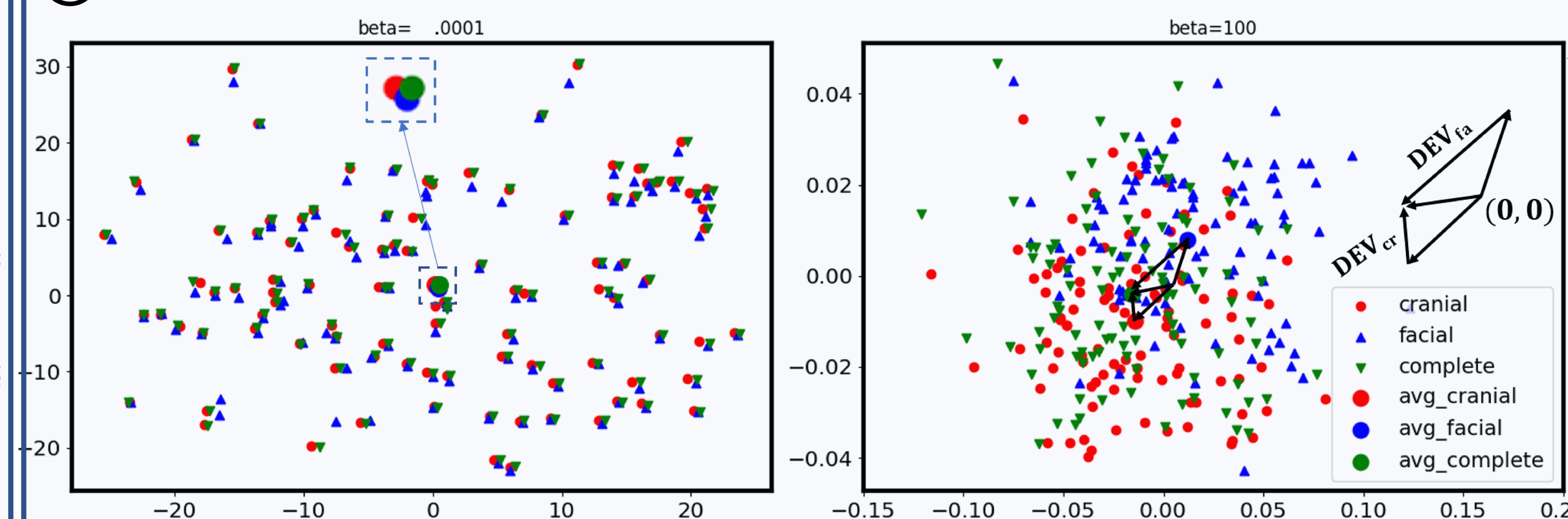

Fig. 5. The distribution of the latent variables given β = 0.0001 (left) and β = 100 (right). The large filled circles represent the centroids of the respective skull classes. The black arrows on the RHS of the plot point from the origin (0, 0) to the centroids, and from the two defective centroids (red and blue) to the complete center (green).

$$DEV_{cr} = \frac{1}{N}\sum_i^N(\mathbf{z}_i^{co} - \mathbf{z}_i^{cr}) = \frac{1}{N}\sum_i^N \mathbf{z}_i^{co} - \frac{1}{N}\sum_i^N \mathbf{z}_i^{cr}$$

$$DEV_{fa} = \frac{1}{N}\sum_i^N(\mathbf{z}_i^{co} - \mathbf{z}_i^{fa}) = \frac{1}{N}\sum_i^N \mathbf{z}_i^{co} - \frac{1}{N}\sum_i^N \mathbf{z}_i^{fa}$$

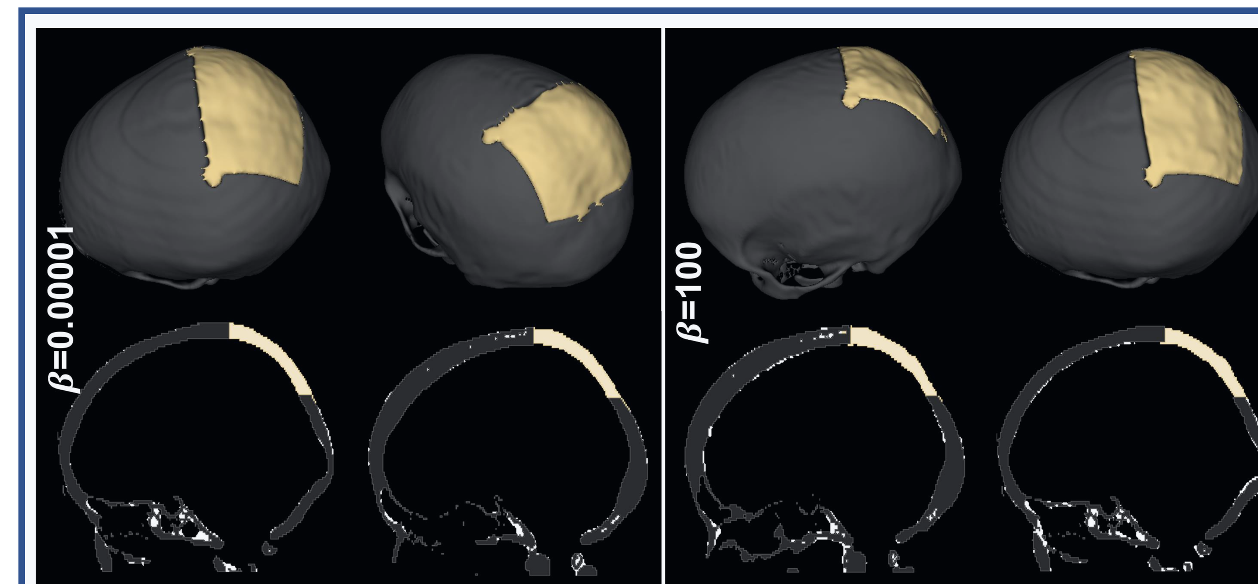

Fig.7. Cranial shape completion results given β = 0.0001 and β = 100. The implant shown in yellow corresponds to the deviation vectors DEV_{cr}.

Skull Reconstruction & Skull Shape Completion

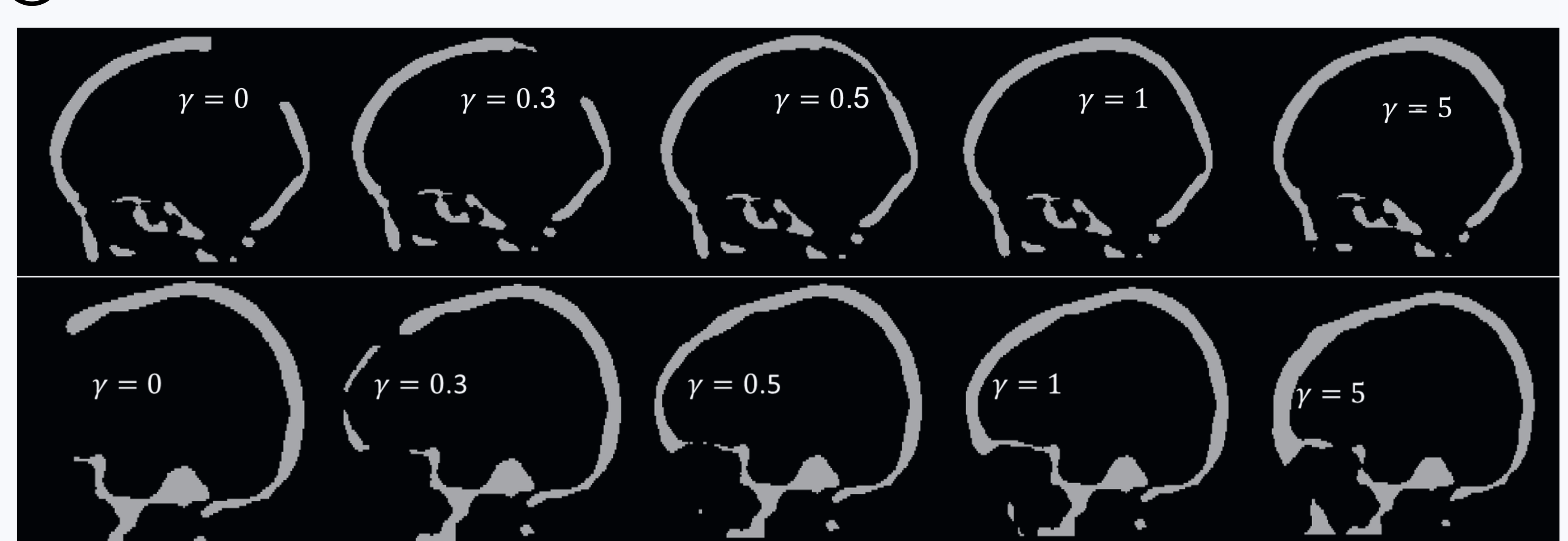

Fig.6. Skull shape completion given β = 0.0001 and different γ. The first and second row shows the shape completion results given a cranial and facial defect, respectively.

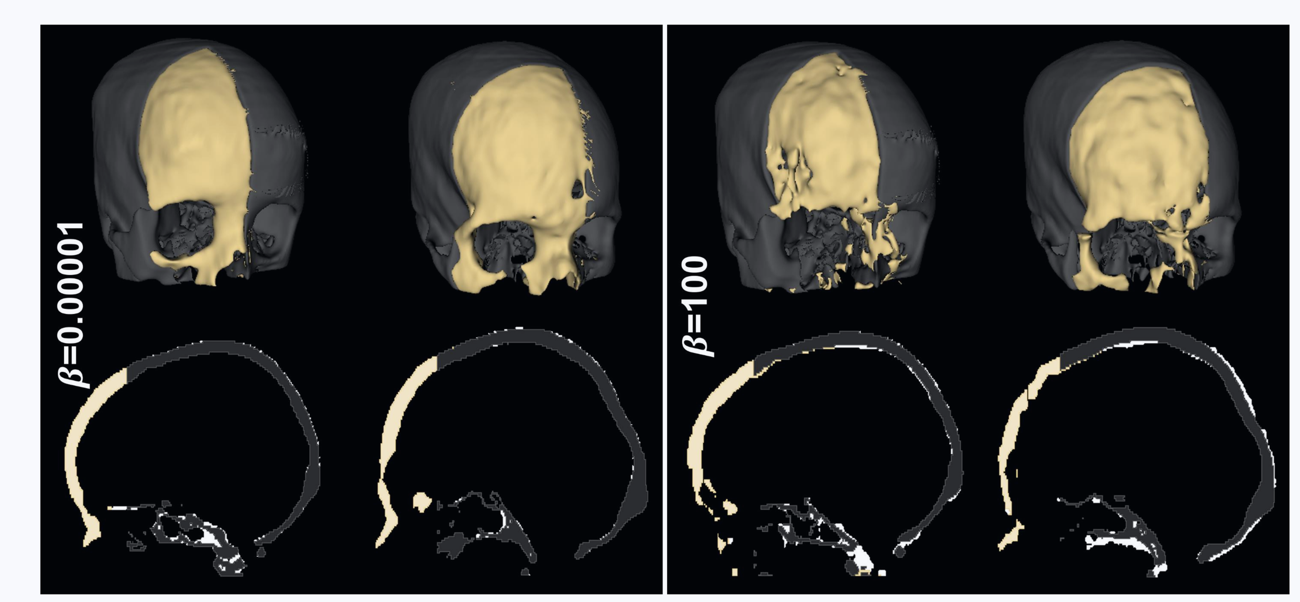

Fig.8. Facial shape completion results given β = 0.0001 and β = 100. The implant shown in yellow corresponds to the deviation vectors DEV_{fa}.

★ **Acknowledgement:** This work was supported by the REACT-EU project KITE (Plattform für KI Translation Essen).

Powered by MONAI TensorFlow GitHub