Data-driven Shape Generative Model for Bones via End-to-end Learning with Self-supervised Registration and Variational Autoregression

Yuan Huang, Qing Zhou, Jisi Tang

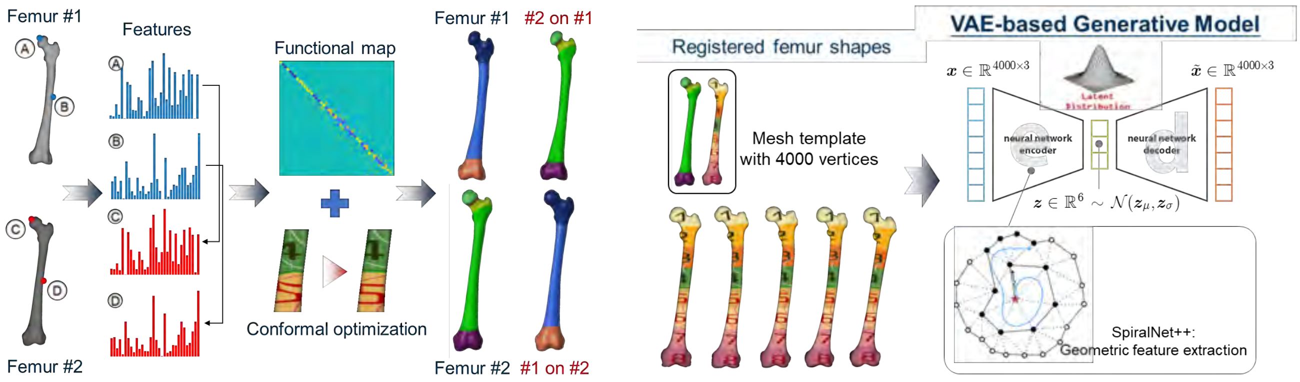
I. INTRODUCTION

The vehicle safety design and evaluation are experiencing a paradigm shift from physical testing to numerical simulation. Human body models (HBMs) are currently the most promising tool to simultaneously cope with the sophisticated human anatomy and diverse road accident scenarios. Conventional HBMs only represent the specific scanned individual. Statistical shape models are therefore brought up to quantitatively characterise population variation. However, existing shape models are mostly regression-based from a group of manually registered templates, which is time-/labour-expensive and cannot sufficiently capture the underlying prior distribution. In this study, we propose a data-driven shape generative model for bones, which autoregressively learns from raw samples and can generate high-fidelity new shapes in a fully unsupervised way.

II. METHODS

Self-supervised Non-rigid Registration

We intended to train a femur-shape generative model from the samples collected in [1]. However, mesh as the standard 3D-shape data structure is intrinsically discretisation-agnostic, so we first needed to unify the mesh pattern. A theoretically feasible solution would be to register the non-isometric samples onto a shared mesh template; however, the technical implementation is usually impractical with handcrafted features. To this end, a self-supervised learning-based registration was established (Fig. 1a). We built a Siamese network, i.e. the input is duplicated into two copies, randomly rotated and scaled respectively [2], and fed into two DiffusionNets [3] sharing the weight parameters. The spatial outputs are transformed into *spectral* through eigenfunctions of the Laplace-Beltrami operators. Thereafter, a combination of *spectral* contrastive, normalisation and regularisation losses are incorporated to facilitate the training. We leveraged a hybrid spatial-spectral framework to robustly infer pointwise correspondences from the learned features, as we did previously in [2]. Finally, the original samples in [1] were all registered onto a GHBMC femur-shaped template, formulating a new dataset with 90 femurs, all in the same mesh pattern (Fig. 1b) for subsequent training of the generative model.



(a) Self-supervised non-rigid registration [3]. (b) VAE-based generative model.

Fig. 1. The end-to-end autoregressive learning of the proposed generative model.

VAE-based Generative Model of Femur Shape

Variational autoencoder (VAE) [4] is a classical type of generative deep-learning model. It learns a compact latent representation of the high-dimensional but redundant realistic distribution, and subsequently generates valid samples from the low-dimensional latent space. In this study, we build a VAE network that autoregressively learns the shape distribution. It includes an encoder which embeds the registered 3D femur shapes (represented by 4000×3 array, nodes by *xyz*-coordinates) into a 6-dimensional latent space $\boldsymbol{z} = \{z_i\}, i \in \{1, 2, \dots, 6\}$; and a decoder which retrieves the original 3D shapes reversely (Fig. 1b).

J. Tang (e-mail: tangjs@cqu.edu.cn) is an Associate Professor with the Bioengineering College at Chongqing University, Chongqing, China. Y. Huang is a Research Investigator with the University of Michigan Medical School, Ann Arbor, MI, USA. Q. Zhou is a Professor with the School of Vehicle and Mobility at Tsinghua University, Beijing, China.

In practice, the 90 registered femur samples have the same mesh pattern (Fig. 1b), i.e. consistent vertex number and connectivity. The encoder relies on SpiralNet++ [5] to learn intrinsic geometric features on the template by fusing local structural information around the vertex of interest on the graph. Latent space $\boldsymbol{z}$ is formulated as a Gaussian distribution parameterised by its mean $\boldsymbol{z}_\mu$ and variance $\boldsymbol{z}_\sigma$. The decoder is a simple fully-connected network, which reconstructs the 4000×3 coordinate array from a latent vector sampled from predicted $\boldsymbol{z}$. The loss function is an ELBO-form (Evidence Lower BOund) [4] with a mean square error (MSE) term between the original and reconstructed coordinate array as well as a Kullback-Leibler divergence term evaluating the distance between z and the standard Gaussian distribution (***0***-mean and ***1***-variance).

III. INITIAL FINDINGS

After training of the VAE model, we can approximately characterise the prior distribution of the femur shapes with a standard 6-dimensional Gaussian distribution. That is, if we sample from the Gaussian distribution and pass the sampled latent vector through the trained decoder, a vertex coordinate array is generated, which can then be converted to a new valid femur shape with the inter-vertex connectivity from the template mesh. This is how the generative model works. Tuning dimensions of the latent vector continuously control different characteristics of the generated femur shape, respectively (Fig. 2), i.e. the overall size (z_6), slenderness (z_4), shaft curvature (z_2) and length (z_3), epiphysis/metaphysis style independent of diaphyisis (z_1), and some small-scale localised features (z_5). As shown in the figure, the generated shapes are considerably fidelic. No unreal artefacts are found even when the sampled latent vectors are quite far away from the mean value ($\pm 3\sigma$). It demonstrates that the Gaussian distribution essentially encodes the prior distribution of the femur shapes.

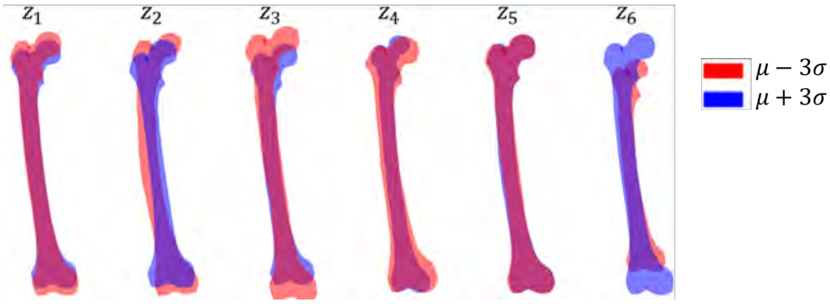


Fig. 2. Influence of each dimension of the latent vector on the generated femur bone shape. One dimension of the latent vector is changed from its $\mu - 3\sigma$ to $\mu + 3\sigma$ (i.e. -3 to +3 since the latent space approximately follows the Gaussian distribution) at a time with the other dimensions being μ (i.e. 0).

IV. DISCUSSION

We proposed a data-driven generative model of femur shapes in this study, which learns from a group of raw shapes segmented from CT scans. In general, our work has two key advantages. First, it is fully automatic, without having to recruit any low-consistency and error-prone human manipulations. Secondly, it generates higher-fidelity samples. Algorithms like principal component analysis (PCA) can also downsize the representation, but simply sampling from the PC-score space might not always result in geometrically/anatomically meaningful shapes. Researchers usually rely on regression to find a valid subspace therein, but regression discards part of the high-order details. Instead, our approach naturally captures the prior distribution by balancing between the reconstruction and regularisation (KL-divergence) terms in training. In future, quantitative evaluation of the sample's fidelity will be necessary, while incorporating advanced techniques, like diffusion model [6], might be beneficial to improving sample quality. It could also perform conditional generation by correlating the latent space with age, sex, height, and other anthropomorphic parameters to explicitly control the generated shapes.

V. ACKNOWLEDGEMENTS

This study was sponsored by the National Natural Science Foundation of China (52102476), and the State Key Laboratory of Intelligent Green Vehicle and Mobility under Project No. KFY2411.

VI. REFERENCES

[1] Du, W., *et al.*, IRCOBI, 2018.
[2] Tang, J., *et al.*, *IEEE Trans Med Imaging*, 2023.
[3] Sharp, N., *et al.*, *ACM Trans Graph*, 2022.
[4] Kingma, D., *et al.*, *ICLR*, 2014.
[5] Gong, S., *et al.*, *ICCVW*, 2019.
[6] Ho, J., *et al.*, *NeurIPS*, 2020.