Surgical Planning: A Machine Learning and Optimization Approach

Hongyou Zhou[1] David Back[2] Marc Toussaint[1]

***Abstract*— Effective bone fracture reconstruction is crucial for patient outcomes, yet existing methods struggle with spatial misalignment and predictive precision. This work leverages machine learning for surgical planning, applying ideas and methods of image in-painting to the problem of bone reconstruction. We employ Variational Autoencoders (VAE) to generate estimated healthy bone structures from fractured CT scans.**

Our key contributions include: i) a novel adaptation of Spatial Transformer Network (STN) for 3D spatial registration, enabling robust alignment of CT fragments, ii) a comparative evaluation of autoencoder architectures for modeling bone CTs, and iii) an extension of these techniques to masked CT images, allowing predictive reconstruction of healthy bone structures.

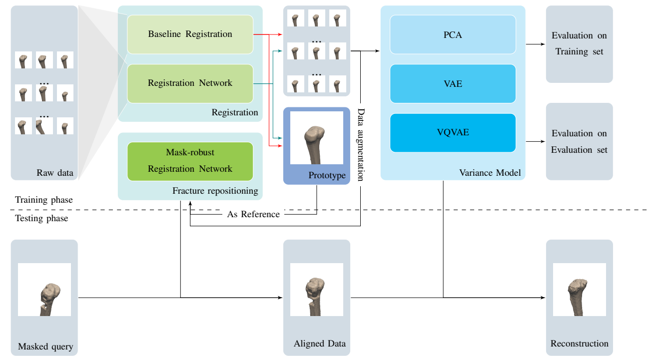


Fig. 1: **Overview of the proposed methods**. On the far left is the dataset of unregistered CT images. Subsequently, we apply two different registration methods. The first involves optimizing the affine transformation parameters for each data pair, while the second focuses on optimizing the neural network parameters. After this step, we obtain (1) the registered CT images and (2) a reference CT image, which we refer to as the **prototype**. Using the registered data, combined with random masking, we train multiple Auto-en/decoder. Finally, the registered data, along with the previously obtained **prototype**, is used to train a model for registering masked CT images.

I. Introduction

Bone fractures are relatively common, and the effectiveness of reconstructive surgery plays a critical role in determining a patient's long-term quality of life, given the limited regenerative capacity of bone tissue [15], [8], [12]. In this project, we explore the use of machine learning to aid in bone reconstruction surgery.

To determine the optimal repositioning of the bone fragments, we first need to generate an estimated healthy variant of the fractures. This problem can be considered related to image in-painting in the field of computer science. A commonly adopted approach, as mentioned in Masked Autoencoder (MAE) [6], involves using Transformers [14], [4] or Variational Autoencoders (VAE) [9] as encoder-decoder architectures. By randomly masking parts of the training data, the model is trained to enhance its ability to reconstruct missing information. Additionally, more efficient feature extraction modules, such as Fast Fourier Convolution (FFC) [3], have been demonstrated in Resolution-Robust Large Mask Inpainting (LAMA) [13] to enhance the performance of image inpainting.

However, since mainstream image models are primarily based on Convolutional Neural Networks (CNN) [10] and CNNs only possess translational invariance, naively applying existing encoder-decoder architectures fails to map data transformed by special Euclidean group transformations SE(3) onto the same feature vectors. To address this issue, a common approach in the medical field is to perform data registration, which eliminates the special Euclidean group SE(3) transformations present in the data, allowing subsequent computations to be conducted within a standardized coordinate system. Common methods for point-cloud registration include Iterative Closest Point (ICP) [1] and Random Sample Consensus (RANSAC) [5]. However, since the CT data we use are significantly denser than point clouds, employing deep learning-based registration methods designed for tensors is more appropriate. VoxelMorph [2] leverages Spatial Transformer Networks (STN) [7], a network capable of computing the partial derivatives of affine transformations for each voxel. By integrating STN into a U-Net [11] architecture, VoxelMorph generates a registration vector for each voxel in the tensor, eventually producing a registration field.

Our current research focuses on the development of autoencoders to assist in surgical planning. More specifically, our goal is to develop a system capable of processing raw CT fragment data, registering it to a standardized coordinate system, and predicting its healthy variant.

Our novel contributions include: i) the modified STN to predict the global spatial registration of a 3D image tensor, ii) a comparison of alternative AE architectures to model bone

[1]Technical University Berlin, Learning and Intelligent Systems, Berlin, Germany hongyou.zhou@tu-berlin.de toussaint@tu-berlin.de

[2]Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Center for Musculoskeletal Surgery, Berlin, Germany david.back@charite.de

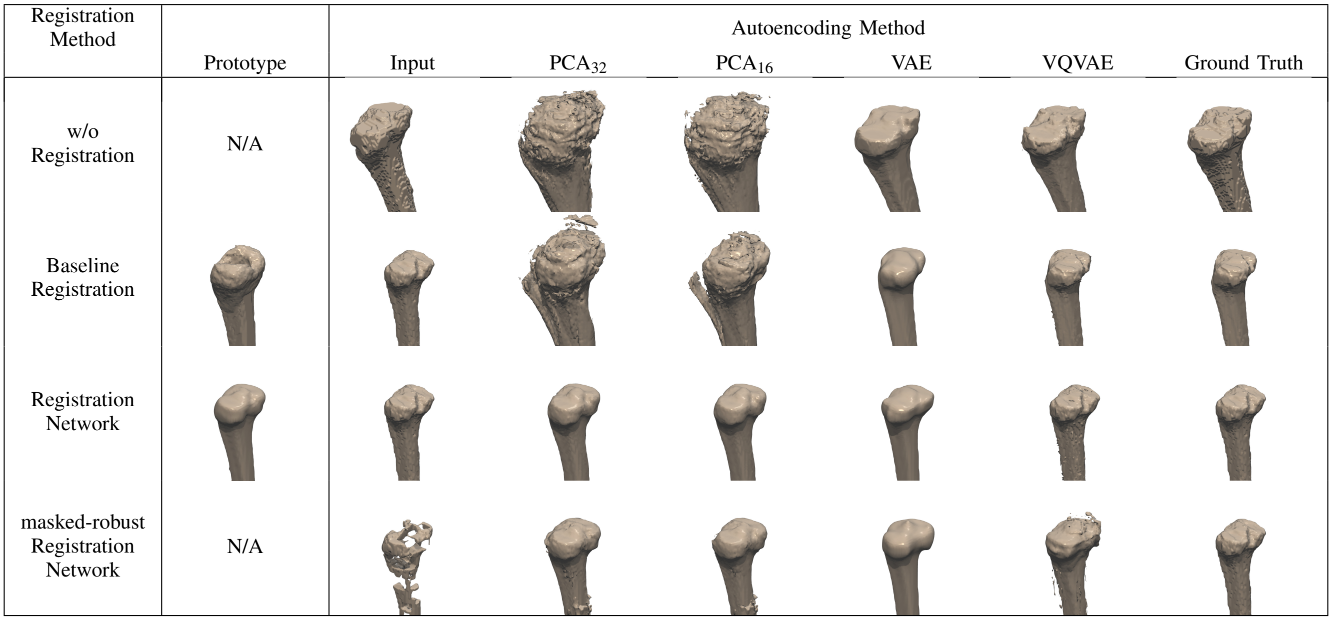


TABLE I: Qualitative comparison of reconstructed Tibia for investigated Registration and Autoencoding methods. 3D models were generated using a multi-threshold Marching Cubes methods.

CTs, and iii) an extension of both of the above to masked CT images, enabling the predictive decoding to a healthy bone in standard coordinates.

II. Method

A. Registration for intact Tibia

As shown in 1, our pipeline takes an unnormalized CT as input, followed by registration modules.

In our experiments, we tested three alternative registration methods: directly optimizing the registration parameters, training a Registration Network [7] to predict the registration parameters, and optimizing deep network parameters with the input data masked. The data obtained from these three methods were then used to train the encoder-decoder model.

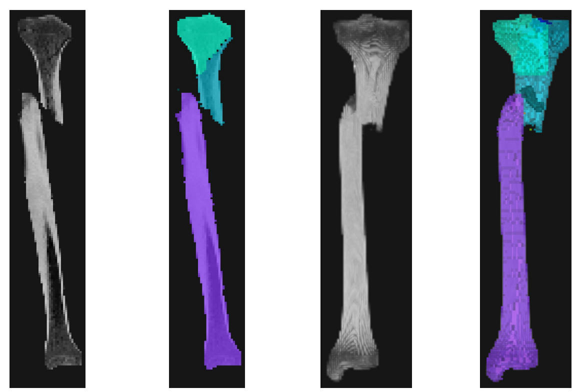
We further tested three different encoder-decoder models: traditional PCA, VAE, and the increasingly popular VQVAE.

Tab. I shows qualitative results for all combinations of the registration and encoder-decoder methods. We find that VQVAE performs the best; however, qualitative analysis indicates that VQVAE tends to overfit the data, leading to artifacts in the generated results, as shown in the Tab. I.

The results indicate that using the trained Registration Network outperforms the other registration methods, while also maintaining acceptable error levels when applied to masked data.

III. Conclusion

In this study, we developed a registration method to eliminate SE(3) transformations present in the dataset, thus aligning the entire data set with a unified coordinate system. Subsequently, we compared the impact of two different registration methods on the performance of an encoder-decoder model and found that training a registration method based on a STN yielded better results. Finally, we extend the registration method to alignment of masked information, enabling the trained model to be applied to real-world data.



(a) (b) (c) (d)

Fig. 2: Illustrations of a fractured tibia in 2D and 3D. (a) and (c) show the 2D and 3D representations, respectively. (b) and (d) show the corresponding segmentations, with different colors representing different segmentation.

In the future, based on the results of this study, we will perform registration on real skeletal fragments and use the trained encoder-decoder model to reconstruct the healthy target variant. As shown in Fig. 2, given the large number of fragments, our next step will be to extend the problem to address a combinatorial optimization problem in a continuous 3D space.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.

[2] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[3] L. Chi, B. Jiang, and Y. Mu, "Fast fourier convolution," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4479–4488, 2020.

[4] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[5] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[6] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009.

[7] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," 2016. [Online]. Available: https://arxiv.org/abs/1506.02025

[8] M. S. Jones and B. Waterson, "Principles of management of long bone fractures and fracture healing," *Surgery (Oxford)*, vol. 38, no. 2, pp. 91–99, 2020.

[9] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[12] M. L. Schenker, R. L. Mauck, J. Ahn, and S. Mehta, "Pathogenesis and prevention of posttraumatic osteoarthritis after intra-articular fracture," *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, vol. 22, no. 1, pp. 20–28, 2014.

[13] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky, "Resolution-robust large mask inpainting with fourier convolutions," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 2149–2159.

[14] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[15] A.-M. Wu, C. Bisignano, S. L. James, G. G. Abady, A. Abedi, E. Abu-Gharbieh, R. K. Alhassan, V. Alipour, J. Arabloo, M. Asaad, *et al.*, "Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the global burden of disease study 2019," *The Lancet Healthy Longevity*, vol. 2, no. 9, pp. e580–e592, 2021.