# Unsupervised Shape and Pose Disentanglement for 3D Meshes

Mentor: Mr. Ashish Tiwari

Supervisor: Prof. Shanmuganathan Raman

Haikoo Khandor and Jinay Dagli

## INTRODUCTION

Parametric models have been used widely since a long time for a variety of important tasks, such as animation and image reconstruction. The major feature of these models is the breakdown of the image surface into shape and pose components. In [1], the authors make use of self-consistency, cross-consistency and ARAP (As-Rigid-As-Possible) [2] blocks to learn shape and pose (disentangled from the image) spaces from meshes of these images. While doing this, they make use of spiral convolution. We look to have an alternative approach by using Mesh Pooling [3] and GCNs [4], in its place.

## Self-Consistency and Cross-Consistency Constraints

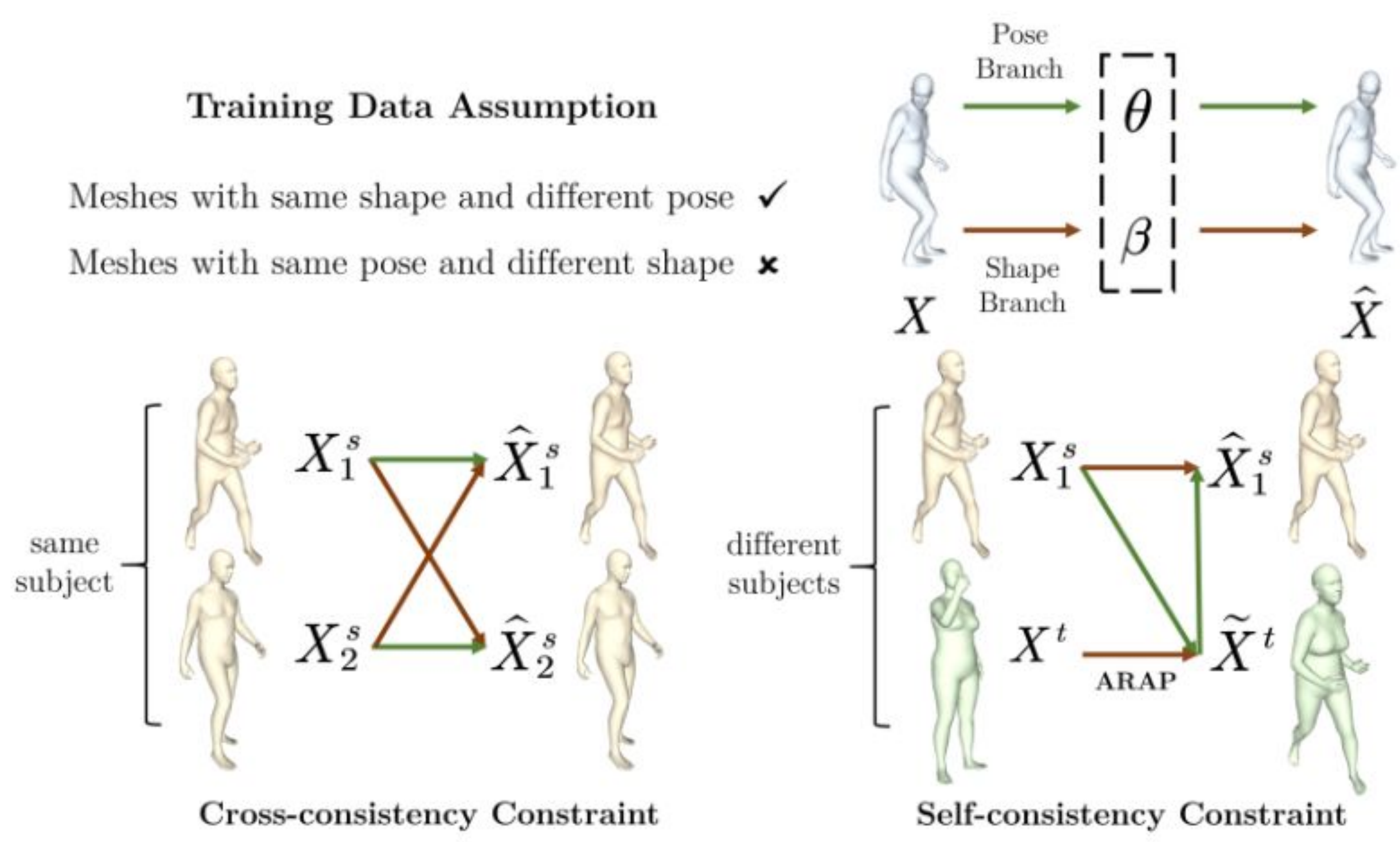


Fig. 1: Schematic Overview of Shape and Pose disentangling mesh auto-encoder [1]

## Spiral Convolution and Proposed Alternatives

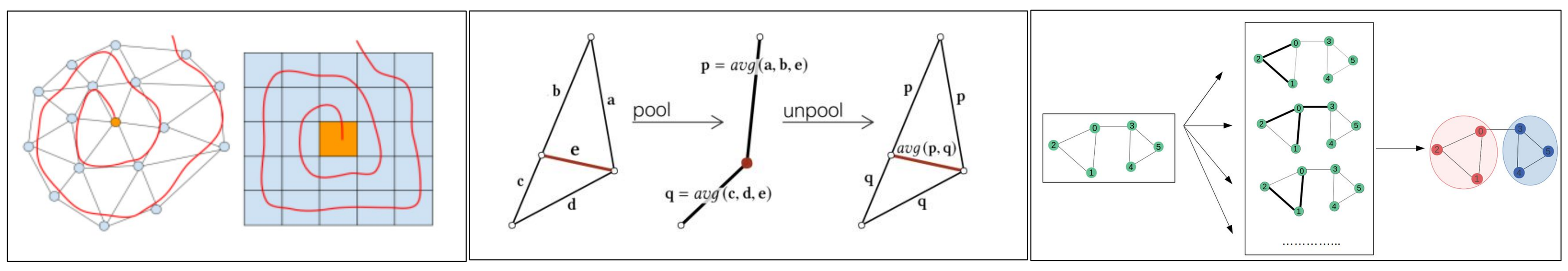


Fig. 2: (a) Spiral Convolution [5]; (b) Mesh Pooling [3] and (c) Graph Convolutional Neural Networks [6]

Mesh CNN's operates on the edges of the mesh which contain more geometric information than vertices or faces. Pooling collapses certain edges resulting in smoothing and complexity reduction. It learns which edges to collapse by the Mesh CNN network.

Graph CNN's are the generalization of CNN's wrt the distance metric. While CNN's are built for ordered data and Euclidean distances i.e. the rectangular grid of pixels, Graph CNN's are for data where the number of neighbours vary for each node and they are unordered too.

## Experimental Setup

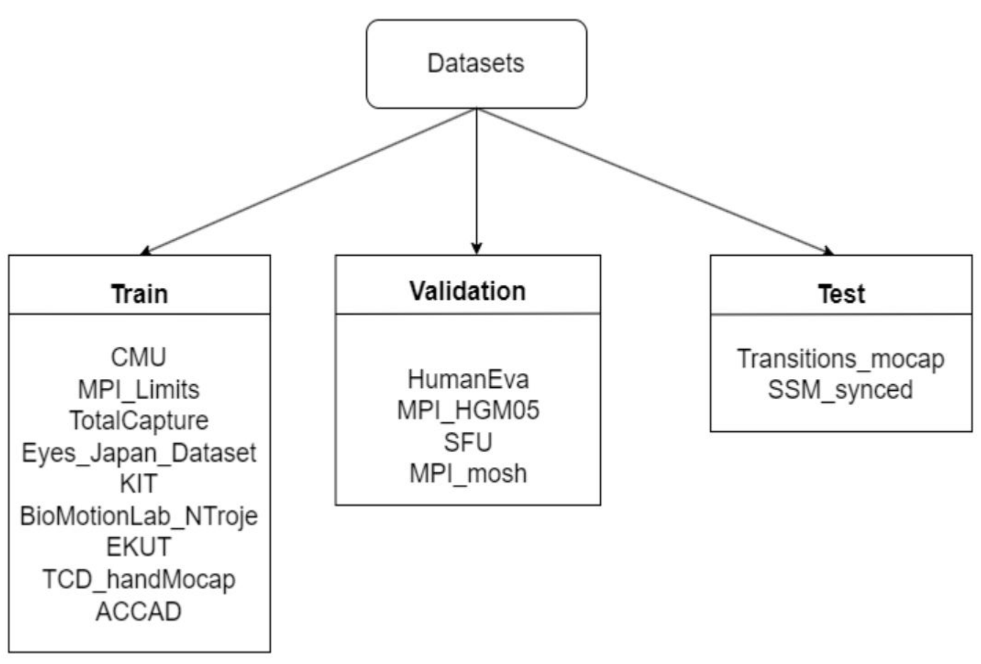


Fig. 2: Distribution of the Datasets Used [1]

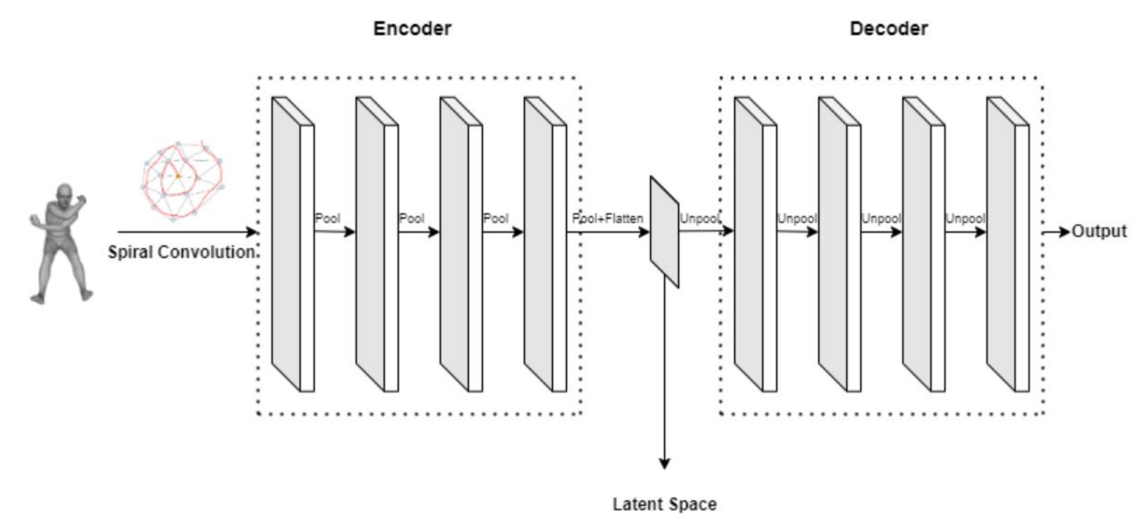


Fig. 3: Proposed Architecture

## Results and Conclusion

The authors have reported a mean error of **19.43**, when using ARAP solver in the implementation, as compared to 54.44 for the SOTA Geometric Disentanglement Variational Autoencoder (GDVAE) [7].

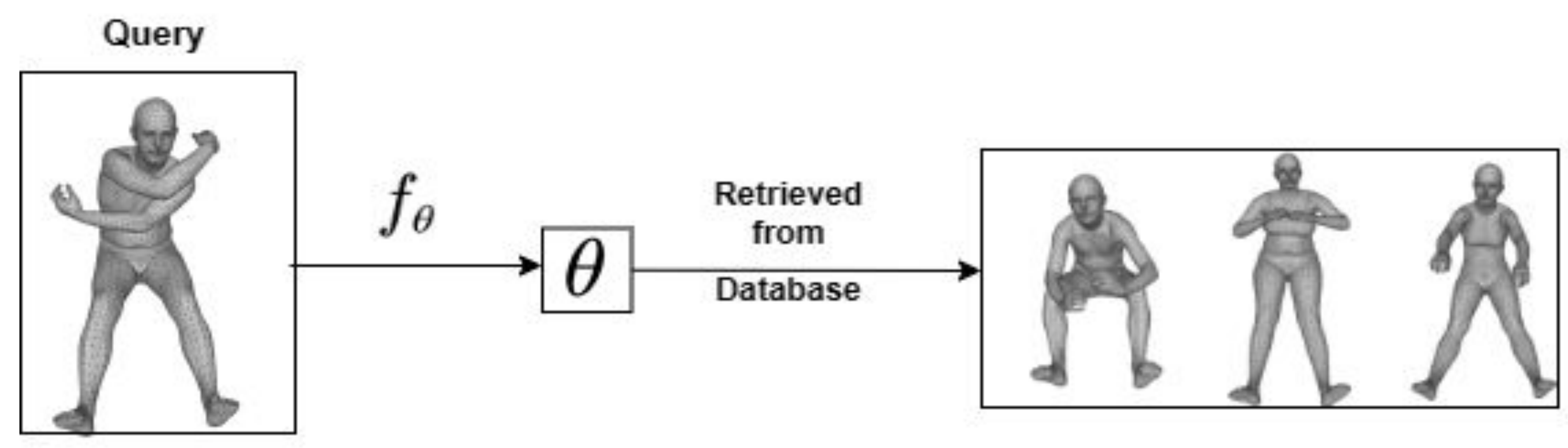


Fig. 4: Pose Retrieval with the Proposed Model: The image on right represents images of *different subjects* most similar to the query in the pose code. [1]

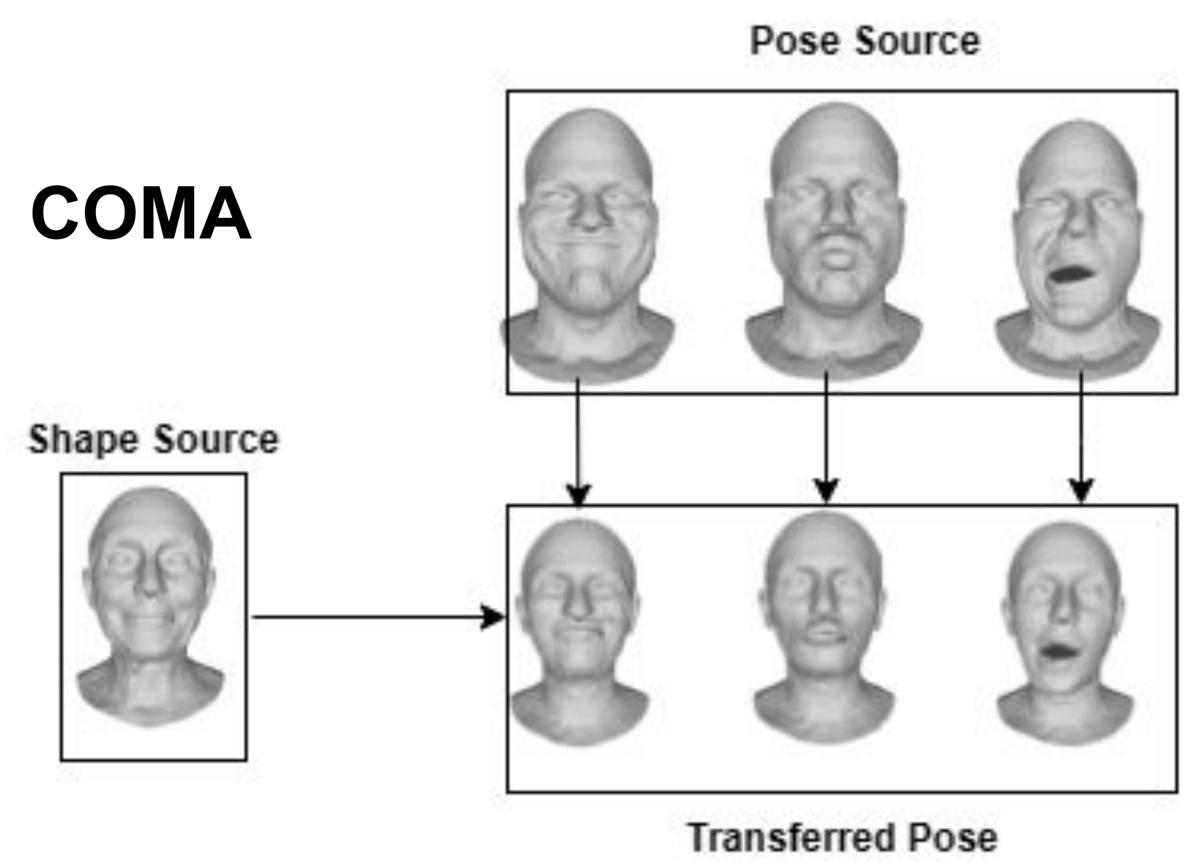


Fig. 5: Transferred Pose from pose sources to shape sources on COMA dataset [1]

## Future Work

Currently, we have implemented pose transfer by using spiral convolution as the operator for finding the nearest pose for a given shape source. Using the shape source and the pose source, we implemented the transfer of pose to different subjects.

In future work, we plan to change this spiral convolution with Mesh Pooling, used in MeshCNNs; and Graph CNNs (efficient variants of CNNs) to observe the performance of the proposed architecture.

## REFERENCES

1. AKeyang Zhou and Bharat Lal Bhatnagar and Gerard Pons-Moll; Unsupervised Shape and Pose Disentanglement for 3D Meshes in European Conference on Computer Vision (ECCV), August 2020
2. Sorkine, O., Alexa, M.: As-rigid-as-possible surface modeling. In: Symposium on Geometry processing. vol. 4, pp. 109–116 (2007).
3. Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Sachar Fleishman, and Daniel Cohen-or; MeshCNN: A network with an Edge in Computing Research Repository (CoRR), 2018.
4. Thomas M. Kipf and Max Welling; Semi-Supervised Classification with Graph Neural Networks in Computing Research Repository (CoRR), 2016.
5. Giorgos Bouritsas, Sergiy Bokhnyak, Neural 3D Morphable Models; Neural 3D Morphable Models: Spiral Convolutional Networks for 3D Shape Representation Learning and Generation in Computing Research Repository (CoRR), 2019.
6. Graph Convolutional Neural Networks; Understanding Graph Convolutional Networks for Node Classification in Towards Data Science.
7. Aumentado-Armstrong, T., Tsogkas, S., Jepson, A., Dickinson, S.; Geometric disentanglement for generative latent shape models. in Proceedings of the IEEE International Conference on Computer Vision. pp. 8181–8190 (2019)