User-Aided Modeling:
Assisting Convolutional Neural Networks for 3D Reconstruction Tasks

Colton Bishop

Advisor: Dr. Jia Deng

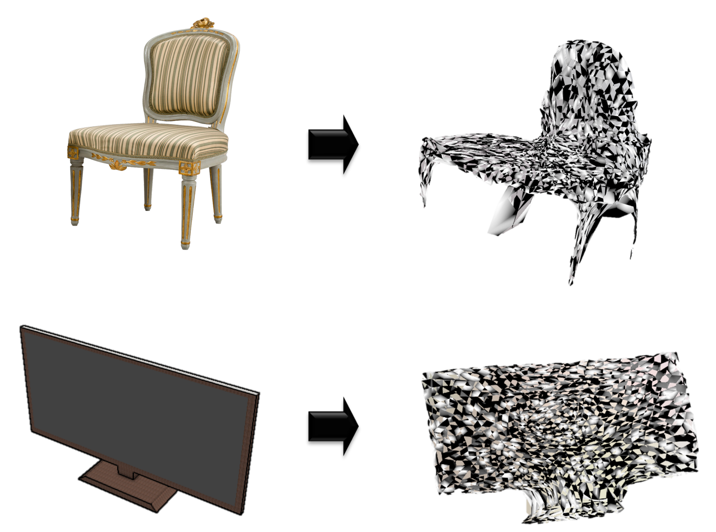
Teaching Assistant: Hei Law

OBJECTIVE

My project proposes and evaluates a topology and symmetry-aware deep learning architecture that incorporates user analysis to improve the performance of a state of the art mesh modeling implementation: Pixel2Mesh. In particular, this new system seeks to generate more accurate and appealing mesh models from single RGB images by considering planes of symmetry, complexity demarcations, and domain shape specifications input by the user. Experimentation shows that these additional forms of human analysis can, in some instances, effectively inform the model during and after processing to produce more accurate 3D shapes. Specifically, this paper explores how these user inputs work to help Pixel2Mesh overcome it's object complexity limitations, perspective bias, and difficulties with back estimation.

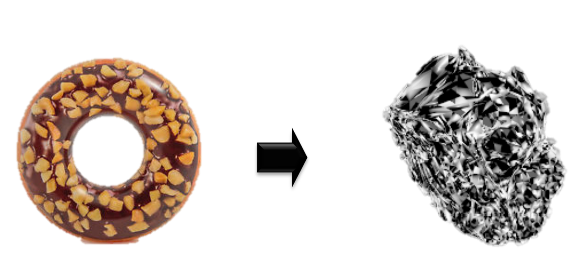
CURRENT LIMITATIONS

Overcome Perspective Bias and Back Estimation Difficulties



The current state of the art system has difficulty recognizing symmetries that obvious to a human observe. It is often confused by different perspectives that make the objects appear skewed.

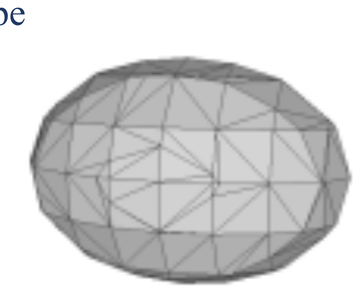
Object Complexity Limitations



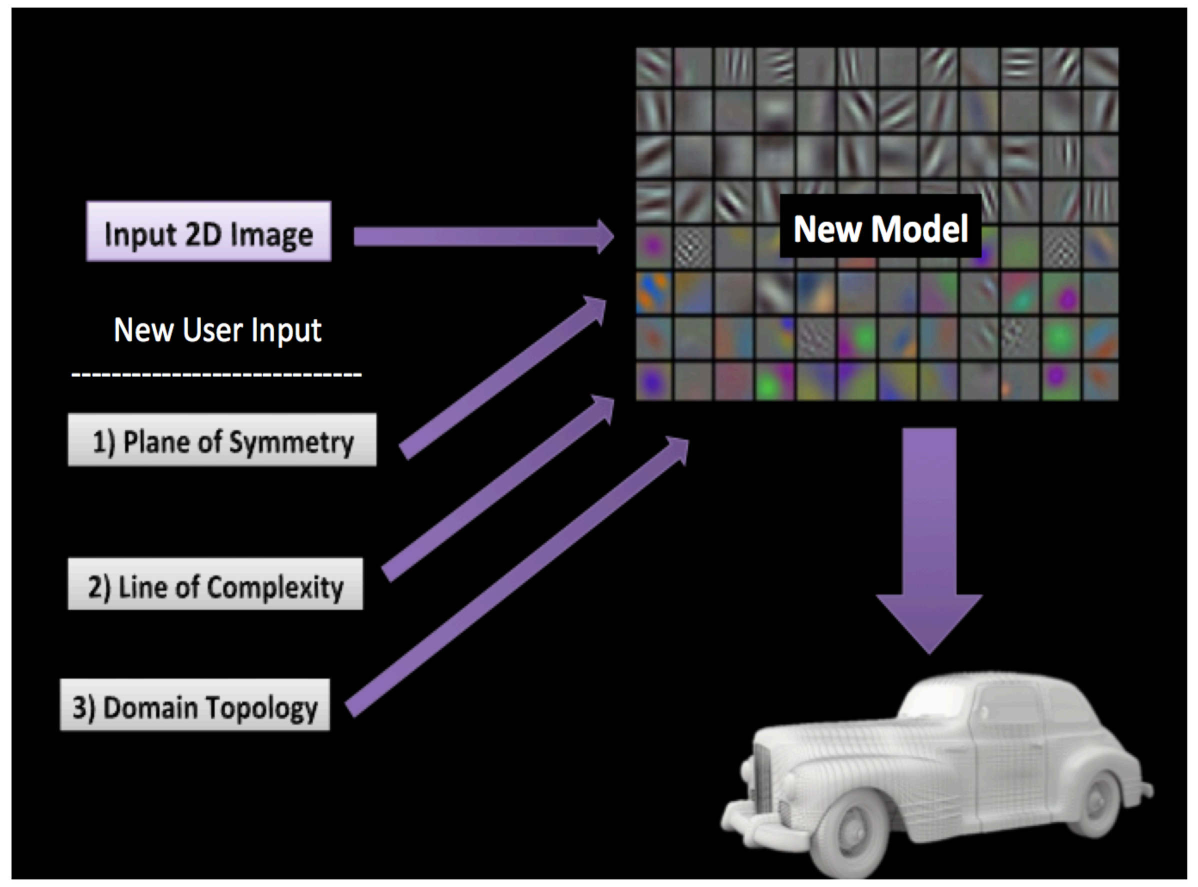
The current system is only capable of modeling simple objects of genus 0 (or non-complex objects with no holes).

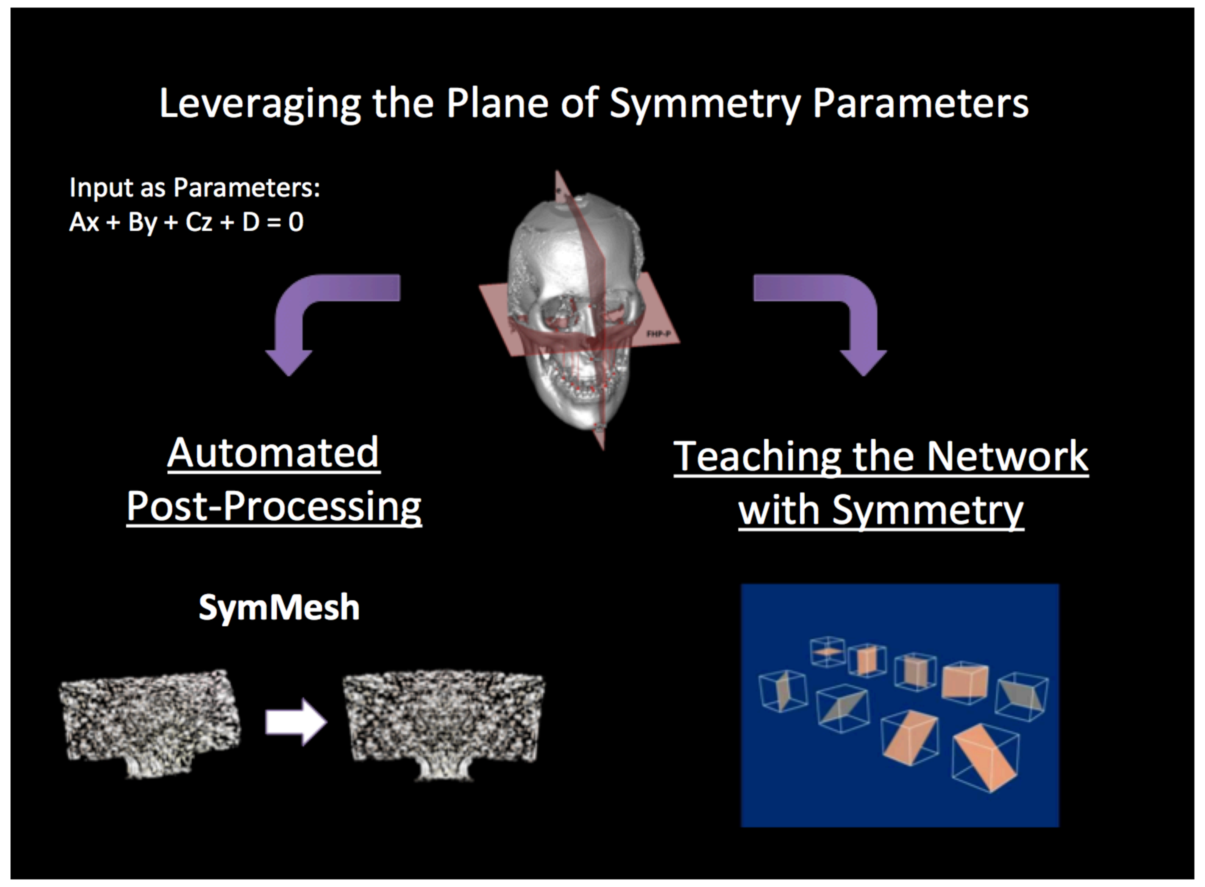
Topological Assumptions and Generalizations

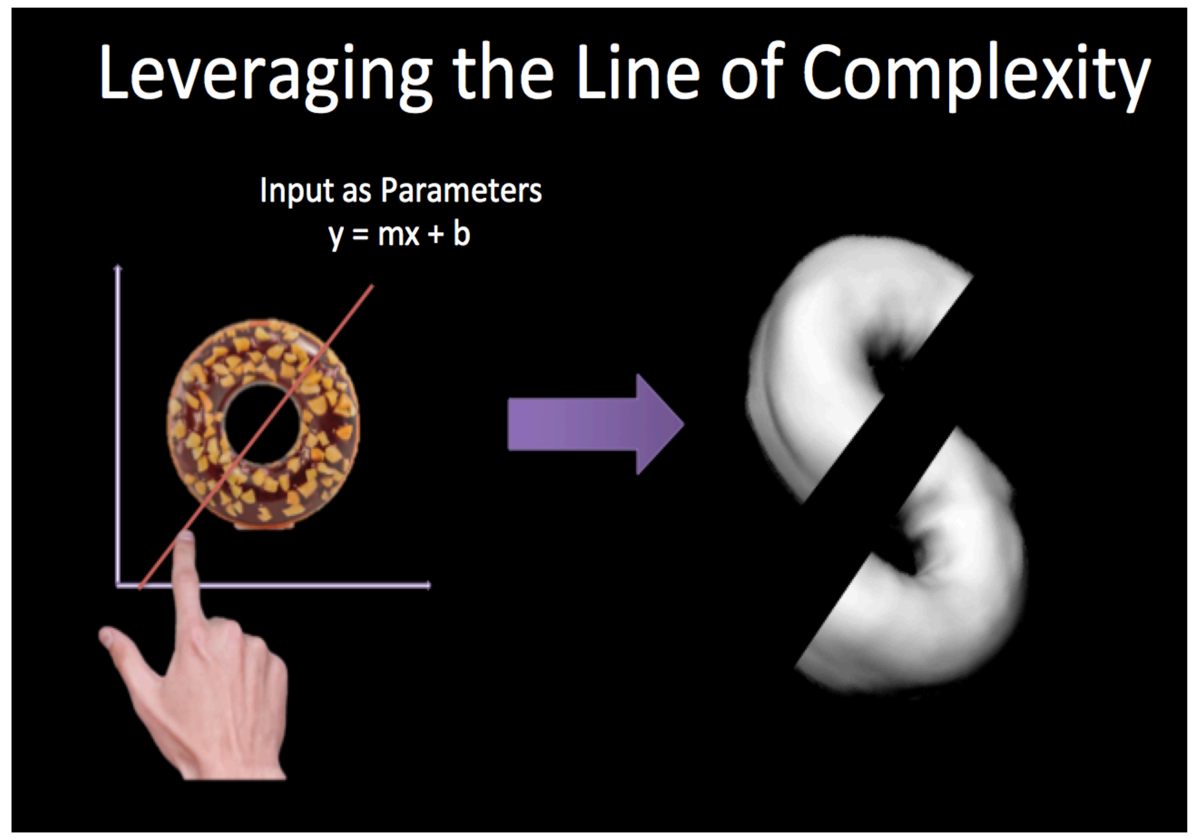
The system works through progressive deformations of an initial base ellipsoid. The use of a generalized ellipsoid shape for this base mesh wastes the potential for initial specification of form and topology and leaves no space to encode prior knowledge of the shape into the base mesh.

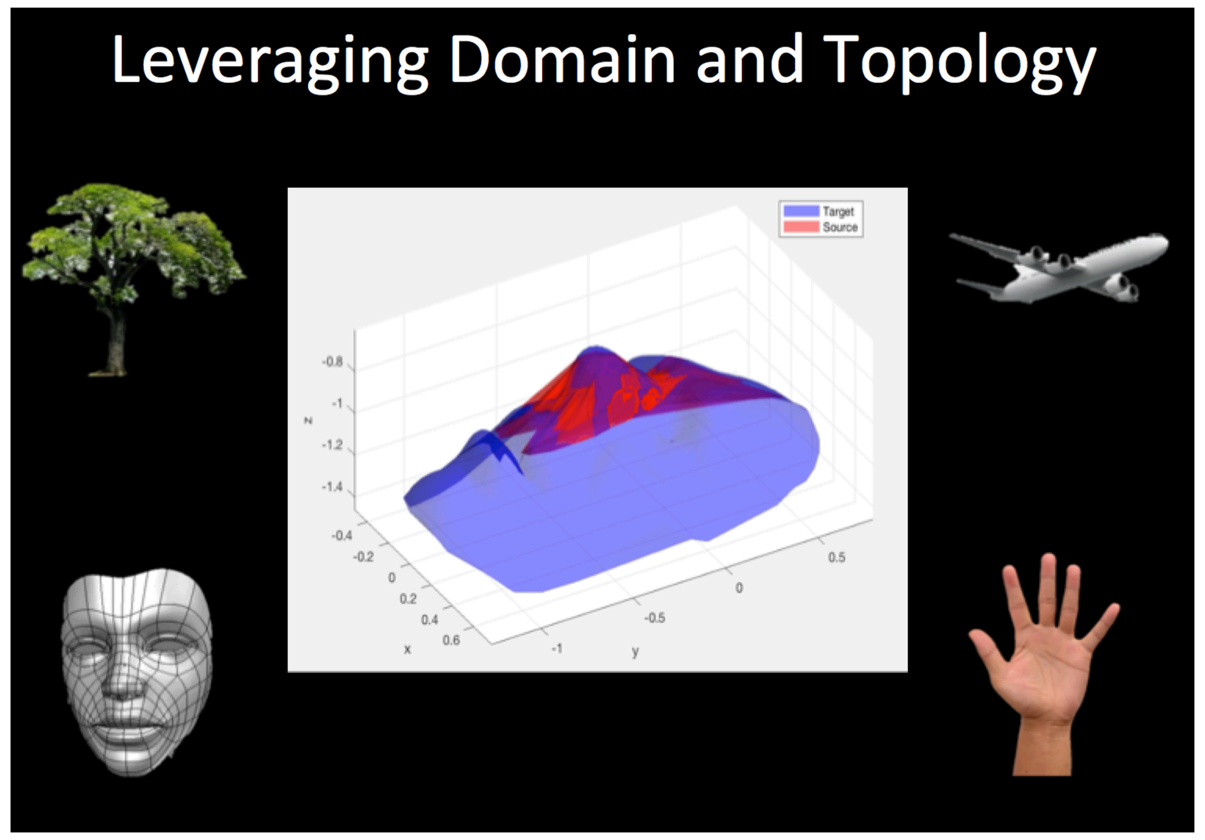


APPROACH AND IMPLEMENTATION

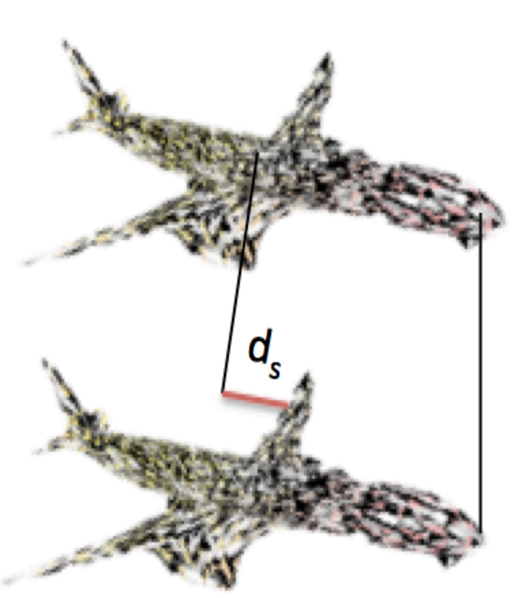








RESULTS AND EVALUATION

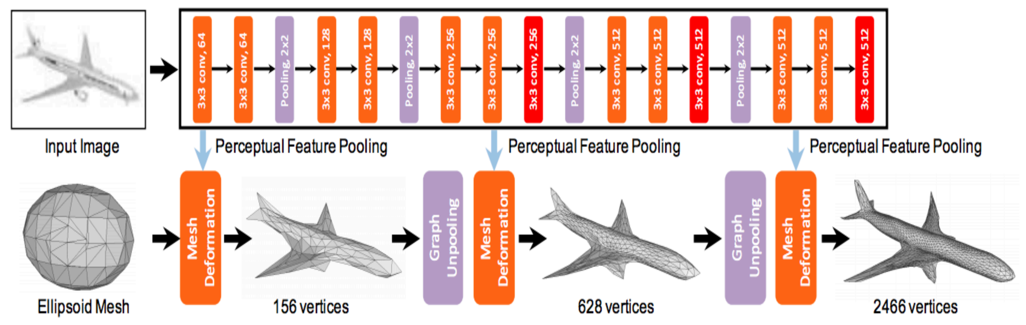


$$\text{Mesh Difference} = \sum_{s \in S} d_s$$

$$\tau = \sum_{s \in S} d_s + \sum_{g \in G} d_g$$

Isolated Component	Evaluation Dataset	Quantitative Evaluation	Qualitative Evaluation
Symmetric Post-Processing	Symmetrical image-model pairs from the ShapeNet database	Baseline τ: 2.65 τ: 2.41	+ Symmetric Structure
Symmetry-Aware Architecture	Symmetrical image-model pairs selected from the Free3D and ShapeNet databases	Baseline τ: 4.22 τ: 4.76	+ More Symmetric Structure
Complexity Analysis	Complex image-model pairs selected from the Free3D and ShapeNet databases	Baseline τ: 3.51 τ: 3.36	+ Complex Topology
Topological Base Change	Human face models selected from the Turbosquid and Free3D databases	Baseline τ: 2.402 τ: 2.02	+ Presence of Key Facial Features

Pixel2Mesh

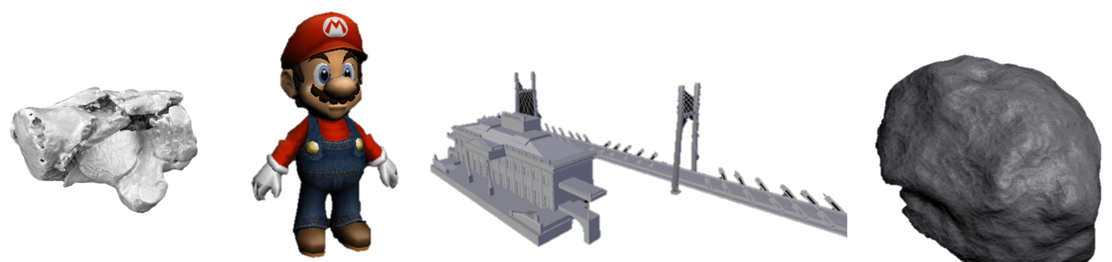


Pixel2Mesh is an end-to-end deep learning architecture proposed by Wang et. al. that represents a 3D mesh model in a graph-based convolutional neural network and "produces correct geometry by progressively deforming an ellipsoid, leveraging perceptual features extracted from the input image." The Pixel2Mesh architecture is depicted above.

Extensive experimentation provided by Wang et. al. reveals the Pixel2Mesh implementation to be both quantitatively superior (in terms of shape estimation) and qualitatively superior (in terms of surface details like smoothness and continuity) when compared to other previously mentioned state of the art methods such as the 3D Recurrent Reconstruction Neural Network and the Point Set Generation Network.

Significance of Contributions

From an economic vantage point, if the time and resources needed to manually analyze properties of the object before modeling would be less than the time and resources that would need to go into post-processing of poorly modeled objects, this approach could effectively increase the efficiency and reduce the costs for any business that has applications for 3D reconstruction, or constructing 3D models from 2D images.



Analysis of Medical Imaging — Video Game Design — Architectural Modeling — Analysis of Astronomical Data

Secondly, from an academic vantage point beyond potential industrial applications, this project seeks to research and analyze how successful different forms of human input are at improving the performance of convolutional neural networks for 3D modeling.

CONCLUSIONS

In defense of 3D reconstruction as a user-centered, mesh-based task.

Our results suggest that the task of 3D reconstruction, like many other deep learning and computer vision tasks, may best be accomplished by assisting neural networks with additional human analysis. As has been established, mesh-based 3D reconstruction is one of the most promising new contenders in the field of 3D modeling. The lightweight nature of mesh objects, along with their easy deformability and their capacity to be encoded with higher-level information, provide many opportunities that other 3D model representations do not. The approaches explored by this paper take advantage of the unique properties of waveform meshes; in fact, these techniques are dependent on the properties of the mesh representation. Thus, the results of this paper serve not only as a defense of user input-centered 3D reconstruction, but also as a defense of the mesh-based approach to 3D reconstruction.