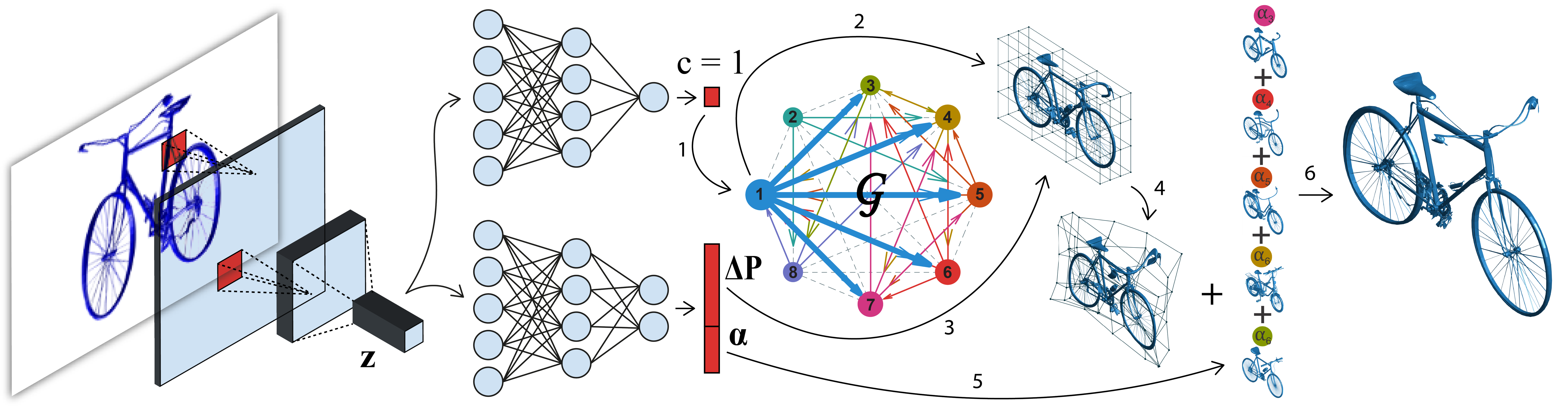
Image2Mesh: A Learning Framework for Single Image 3D Reconstruction

Jhony K. Pontes†, Chen Kong⋆, Sridha Sridharan†,
Simon Lucey⋆, Anders Eriksson†, Clinton Fookes†
Queensland University of Technology† Carnegie Mellon University⋆

Abstract

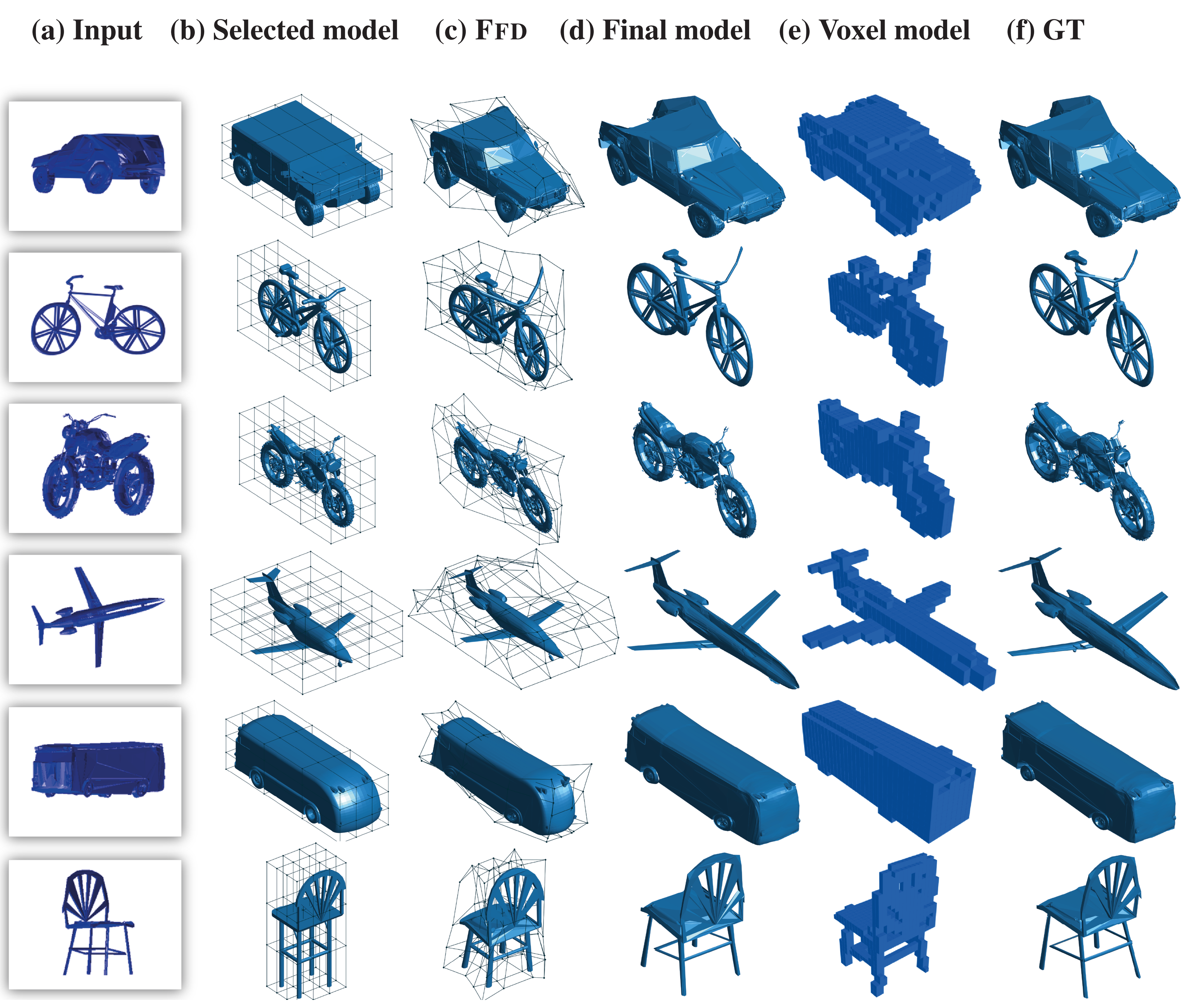
A challenge that remains open in 3D deep learning is how to efficiently represent 3D data to feed deep neural networks. Recent works have been relying on volumetric or point cloud representations, but such approaches suffer from a number of issues such as computational complexity, unordered data, and lack of finer geometry. An efficient way to represent a 3D shape is through a polygon mesh as it encodes both shape's geometric and topological information. However, the mesh's data structure is an irregular graph (collection of vertices connected by edges to form polygonal faces) and it is not straightforward to integrate it into learning frameworks since every mesh is likely to have a different structure. Here we address this drawback by efficiently converting an unstructured 3D mesh into a regular and compact shape parametrization that is ready for machine learning applications. We developed a simple and lightweight learning framework able to reconstruct high-quality 3D meshes from a single image by using a compact representation that encodes a mesh using free-form deformation and sparse linear combination in a small dictionary of 3D models. In contrast to prior work, we do not rely on classical silhouette and landmark registration techniques to perform the 3D reconstruction. We extensively evaluated our method on synthetic and real-world datasets and found that it can efficiently and compactly reconstruct 3D objects while preserving its important geometrical aspects.

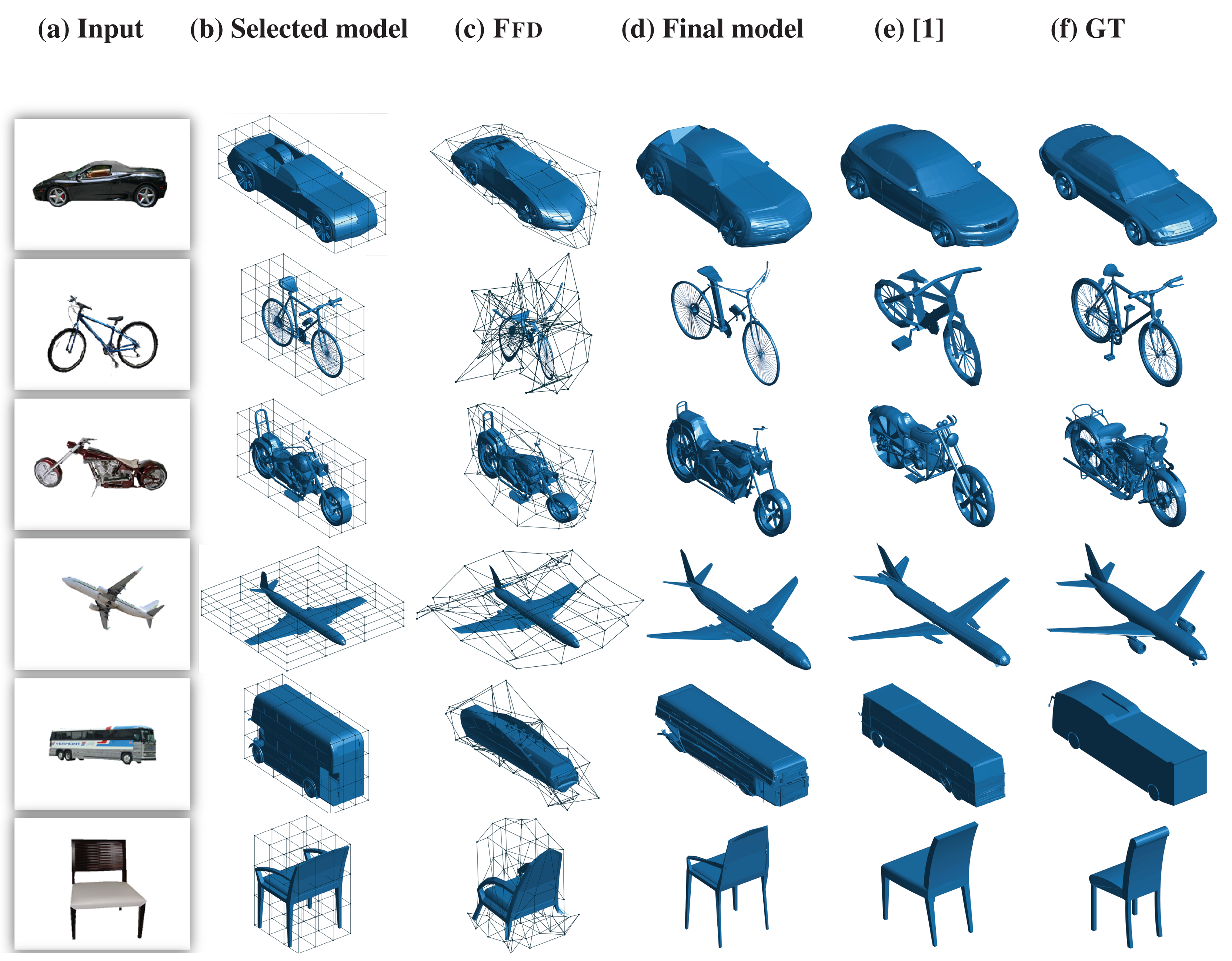
Proposed Method



Given a single image, our framework employs a convolutional autoencoder to extract the image's latent space $\mathbf{z}$ to be classified into an index c and regressed to a shape parametrization ($\mathbf{\Delta P}$, α). We use a graph embedding $\mathcal{G}$ to compactly represent 3D meshes. The estimated index c selects from $\mathcal{G}$ the closest 3D model to the image. The selected model is then deformed with the estimated parameters - free-form deformation (FFD) displacements $\mathbf{\Delta P}$ and sparse linear combination weights α. For instance, model 1 is selected (arrows 1 and 2), FFD is then applied (arrows 3 and 4), and finally the linear combination with the nodes 3, 4, 5, 6, and 7 (blue arrows on the graph that indicates the models in dense correspondence with node 1) are performed (arrow 5) to reconstruct the final 3D mesh (arrow 6).

Experiments on Synthetic and Real Datasets





Left: Visual results on synthetic data. (a) shows the input image; (b) the selected model from the graph; (c) the selected model deformed by FFD. The final 3D model deformed by linear combination is shown in (d). The voxelized final model is shown in (e) and the ground truth in (f).
Right: Visual results on real-world data. (a) shows the input image; (b) the selected model; (c) the selected model deformed by FFD. The final 3D model reconstructed by linear combination is shown in (d). We compare with [1] in (e) and the ground truth is shown in (f).

References

[1] J. K. Pontes, C. Kong, A. Eriksson, C. Fookes, S. Sridharan, S. Lucey, Compact Model Representation for 3D Reconstruction, in *3DV*, 2017

Project Webpage