Object Wake-up: 3D Object Rigging from a Single Image

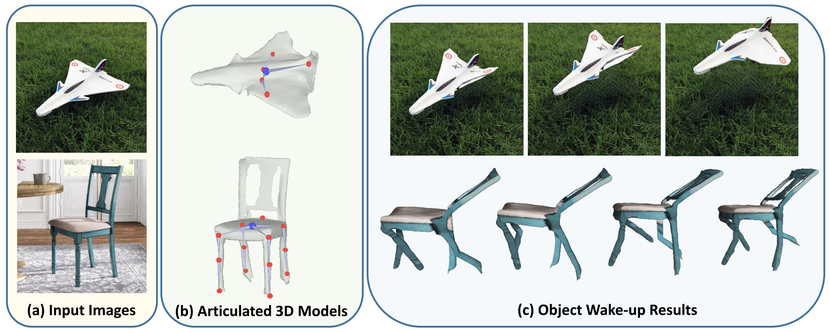
Ji Yang* Xinxin Zuo* Sen Wang Zhenbo Yu Xingyu Li Bingbing Ni Minglun Gong Li Cheng
University of Alberta Shanghai Jiao Tong University
Huawei Hisilicon University of Guelph

ECCV TEL AVIV 2022

Code & Model
https://kulbear.github.io/object-wakeup/

Problem Definition and Challenges

Goal: Given a single chair image, we wake it up by reconstructing its 3D shape and skeleton, as well as animating its plausible articulations and motions.



Motivations & Challenges:

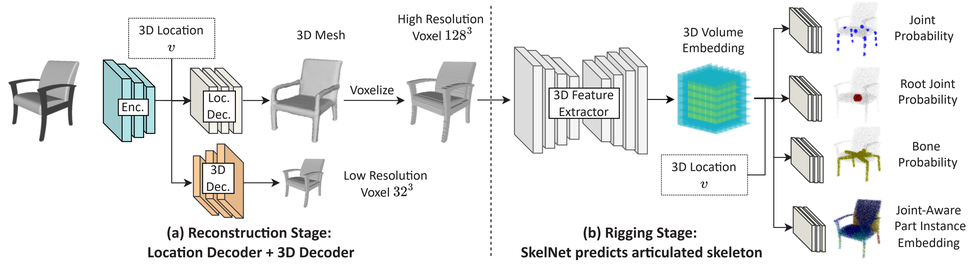
- Previous works typically perform 3D reconstruction without considering the structure of the object [1], or perform object and character rigging separately [2].
- This work could give rise to numerous downstream augmented and virtual reality applications.
- Existing dataset is limited in both feasibility and diversity.

Overview

Object Wake-up is the first to build an automated approach to tackle the entire process of reconstruct such generic 3D objects, rigging and animation, all from single images.
Multihead Implicit Function is a novel and effective skeleton prediction approach. It is the first work in using the deep implicit functions for modelling 3D object skeleton structure.
Novel in-house datasets ShapeRR and SSkel of general 3D objects are constructed, containing annotated 3D skeletal joints and photo-realistic re-rendered images, respectively.

Method

Network Architecture:

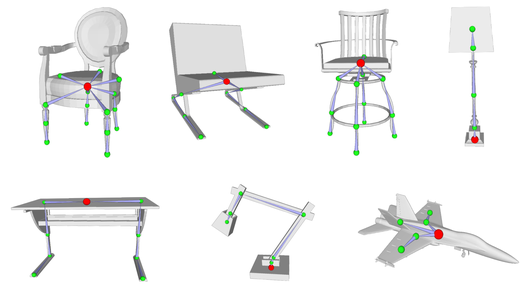


Our overall pipeline. It starts with the proposed Transformer-based model for 3D reconstruction consisting of a DeiT image encoder, an auxiliary 3D CNN voxel prediction branch and the occupancy decoder. The proposed SkelNet accepts the high resolution voxelized input from the reconstructed 3D mesh, and predicts articulated skeleton with a multi-head architecture.

ShapeRR and SSkel Dataset

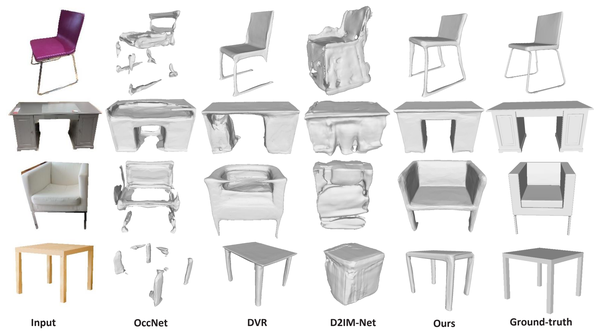
Dataset Image Quality Comparison:

Sampled Annotation in our SSkel Dataset:

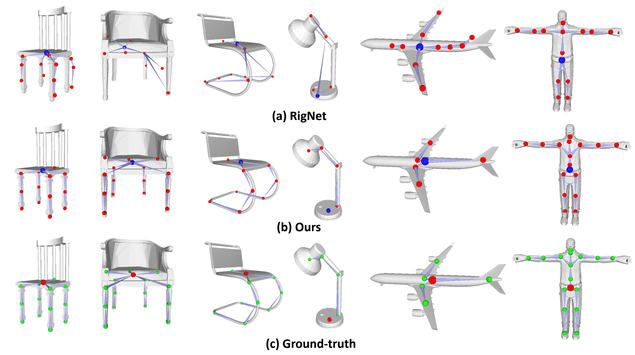


Experiments & Results

Reconstruction Qualitative Results:



Rigging Qualitative Results:



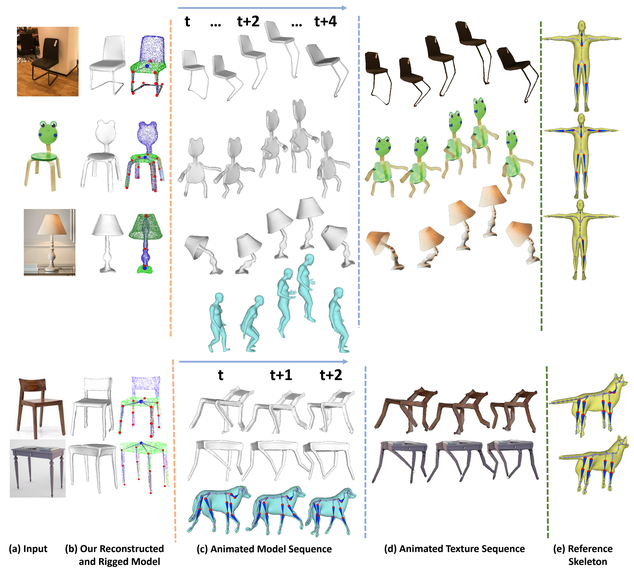
Reconstruction: Quantitative Results on ShapeRR Dataset

	Chamfer Distance (↓)					Volumetric IoU (↑)				
ShapeNet	Chair	Table	Lamp	Airplane	Avg.	Chair	Table	Lamp	Airplane	Avg.
OccNet [21]	1.9347	1.9903	4.5224	1.3922	2.3498	0.5067	0.4909	0.3261	0.5900	0.4918
DVR [24]	1.9188	2.0351	4.7426	1.3814	2.5312	0.4794	0.5439	0.3504	0.5741	0.5029
D^{2}IM-Net [17]	**1.8847**	1.9491	4.1492	1.4457	2.0346	**0.5487**	0.5332	0.3755	0.6123	0.5231
Ours	1.8904	**1.7392**	**3.9712**	**1.2309**	**1.9301**	0.5436	**0.5541**	**0.3864**	**0.6320**	**0.5339**
Pix3D	Table	Chair	Desk	Sofa	Avg.	Table	Chair	Desk	Sofa	Avg.
OccNet [21]	7.425	9.399	15.726	14.126	11.625	0.215	0.201	0.143	0.152	0.190
DVR [24]	8.782	6.452	12.826	11.543	9.901	0.187	0.237	0.165	0.187	0.185
D^{2}IM-Net [17]	8.038	7.592	11.310	9.291	9.057	0.205	0.244	0.183	0.207	0.215
Ours	**6.449**	**6.028**	**8.452**	**8.201**	**7.282**	**0.239**	**0.277**	**0.219**	**0.241**	**0.242**

Rigging: Quantitative Results on SSkel Dataset:

	Chair			Table			Lamp			Airplane			Average		
metrics	J2J	J2B	B2B	J2J	J2B	B2B	J2J	J2B	B2B	J2J	J2B	B2B	J2J	J2B	B2B
RigNet-GT	0.052	0.042	0.035	0.061	0.049	0.040	0.132	0.110	0.098	0.096	0.081	0.073	0.061	0.046	0.041
Ours-GT	**0.030**	**0.023**	**0.021**	**0.044**	**0.032**	**0.028**	**0.097**	**0.071**	**0.063**	**0.075**	**0.062**	**0.056**	**0.047**	**0.038**	**0.033**
RigNet-rec	0.048	0.035	0.033	0.060	0.046	0.038	0.143	0.116	0.102	0.103	0.084	0.076	0.063	0.047	0.042
Ours-rec	**0.036**	**0.024**	**0.022**	**0.047**	**0.033**	**0.029**	**0.101**	**0.073**	**0.065**	**0.081**	**0.065**	**0.059**	**0.051**	**0.041**	**0.036**

Animation Demonstration:



References:

[1] Mescheder, Lars, et al. "Occupancy Networks: Learning 3D Reconstruction in Function Space". CVPR (2019).
[2] Lin, Angela S., et al. "RigNet: Neural Rigging for Articulated Characters" SIGGRAPH (2020): Volume 39.