UCSG-NET - Unsupervised Discovering of Constructive Solid Geometry Tree

Wrocław University of Science and Technology

Kacper Kania[1], Maciej Zieba[1,2], Tomasz Kajdanowicz[1]

[1]Wroclaw University of Science and Technology [2]Tooploox

kacp.kania@gmail.com

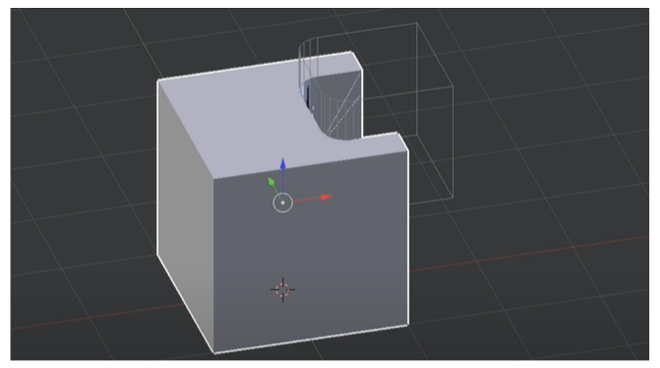
TOOPLOOX

Contributions

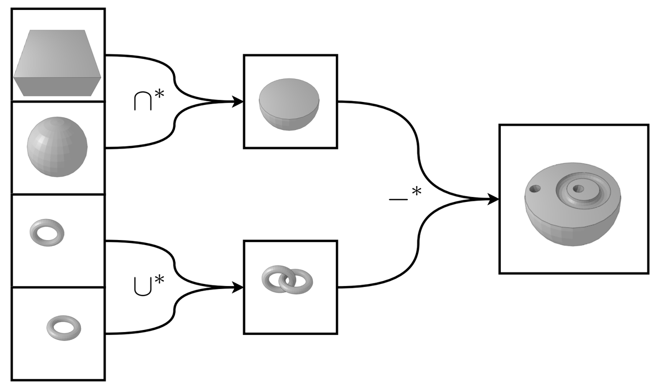
- The first method that learns CSG operations in **unsupervised manner**.
- Predicted CSG trees are fully **interpretable** and **controllable**. They can aid design of 3D objects.
- In terms of reconstruction quality, our model is **on par** with existing methods that aim for interpretable 3D object reconstruction.

Motivation

CSG is a common tool that can be in 3D graphics software for modeling objects with complex topology.

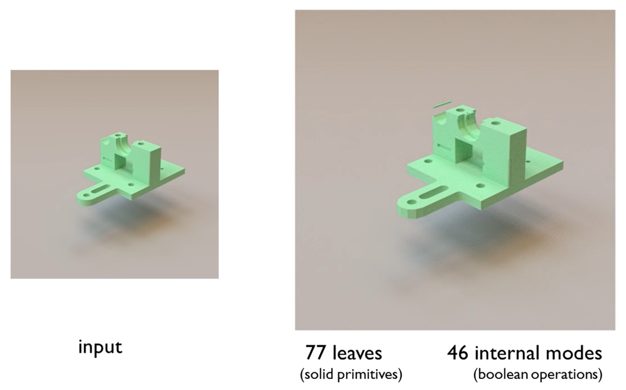


We aim to automate the process by predicting CSG trees that create these objects.



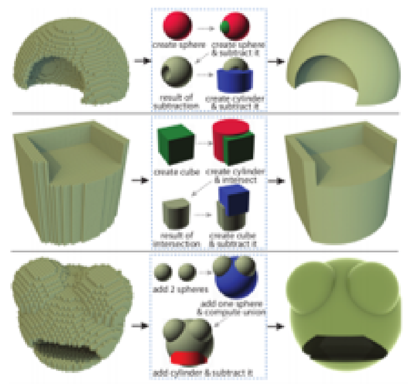
Existing approaches

Time consuming inference



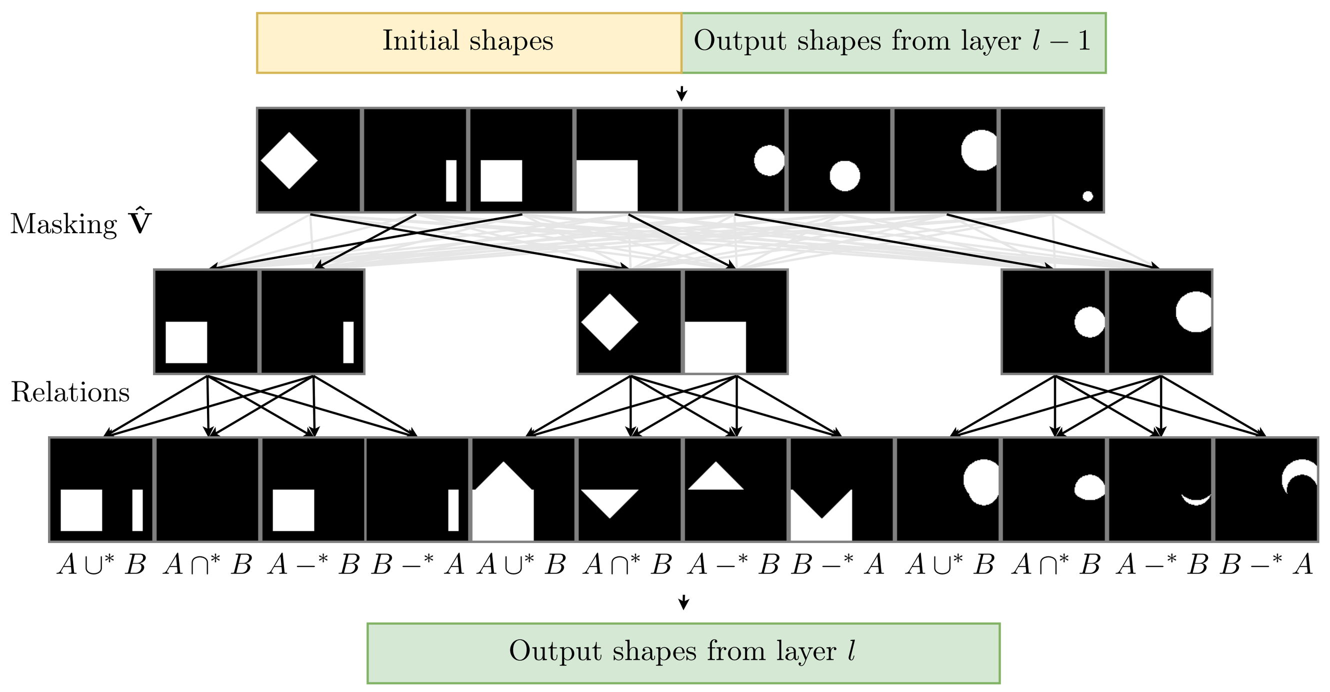
InverseCSG, Du et al., SIGGRAPH 2018

Require supervision of CSG operations



CSG-Net, Sharma et a., CVPR 2018

Learnable CSG Layer



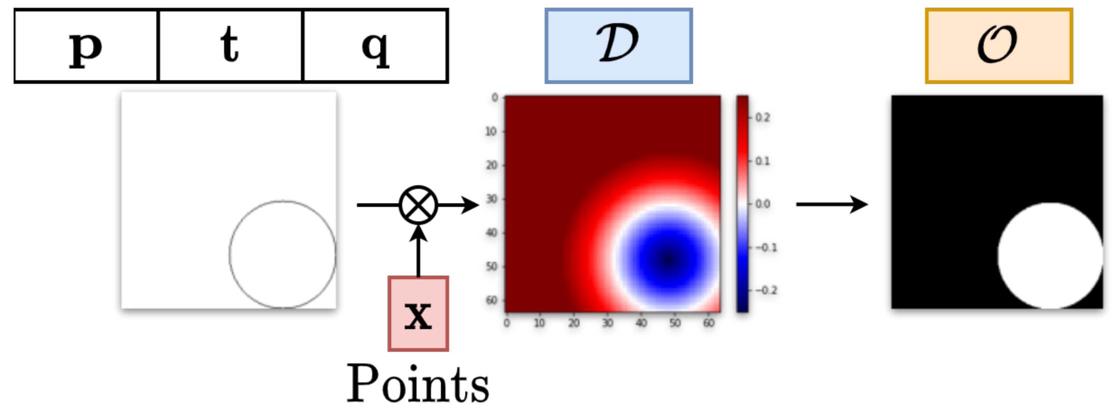
- Masks (separate for left and right operands of CSG operations) select pair of elements:

$$\mathbf{V}_{\text{left}} = \text{softmax}(\mathbf{K}_{\text{left}}\mathbf{z}) \qquad \mathbf{V}_{\text{right}} = \text{softmax}(\mathbf{K}_{\text{right}}\mathbf{z})$$

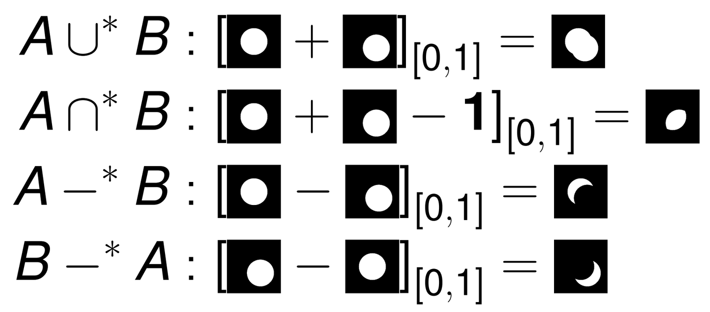
- Masked shapes $\hat{V}_{\text{side},i}$ are obtained by sampling Gumbel-Softmax.
- Operands for l-th layer are obtained as:

$$A = \mathcal{O}_{\text{left}} = \sum_{i=1}^{M} \mathcal{O}_i \hat{V}_{\text{left},i} \qquad B = \mathcal{O}_{\text{right}} = \sum_{i=1}^{M} \mathcal{O}_i \hat{V}_{\text{right},i}$$

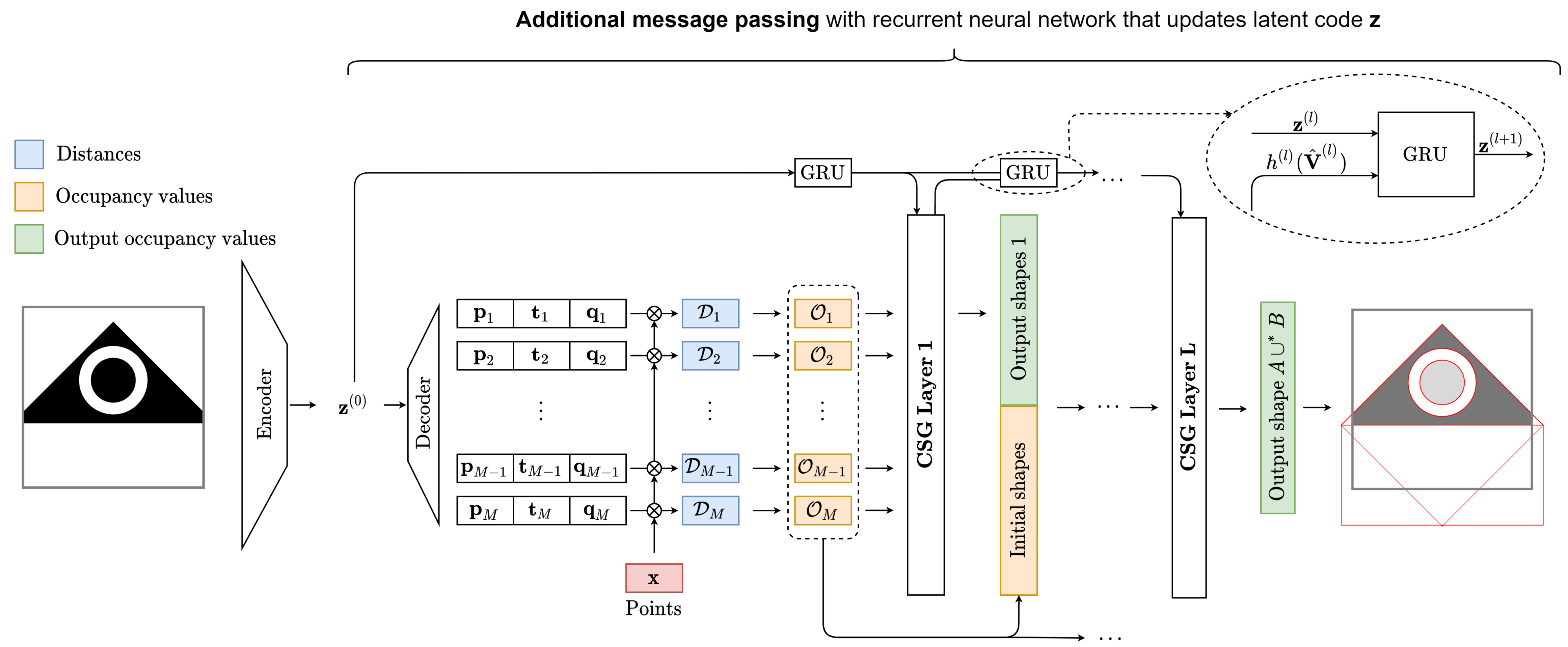
Representing a single shape and CSG operations



CSG operations for occupancies



UCSG-NET



Encoder transforms the input (images or voxels) into latent code **z**

Decoder predicts parameters of M primitives (ex. spheres and cuboids)
p - shape parameters
t - translation
q - rotation

Converter transforms signed distances into occupancy values

CSG Layers select pairs from input shapes and applies all possible continuous CSG operations

In **the last layer**, only union is selected to output a single shape

Training UCSG-NET

First stage

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{P}} + \lambda_{\text{T}}\mathcal{L}_{\text{T}} + \lambda_{\alpha}|\alpha|$$

$\mathcal{L}_{\text{MSE}} = \mathbb{E}_{\mathbf{x}\in\mathbf{X}}[(\mathcal{O}^{(L)} - \mathcal{O}^*)^2]$ — Reconstruction error

$\mathcal{L}_{\text{P}} = \sum_{i=1}^{M}\sum_{p_i\in\mathbf{p}_i}\max(-p_i, 0)$ — Maintain feasible parameters of primitives

$\mathcal{L}_{\text{T}} = \sum_{i=1}^{M}\max(||\mathbf{t}_i||^2, 0.5)$ — Minimizes length of translation vector

$\lambda_{\alpha}|\alpha|$ — Pushes occupancy values towards {0, 1}

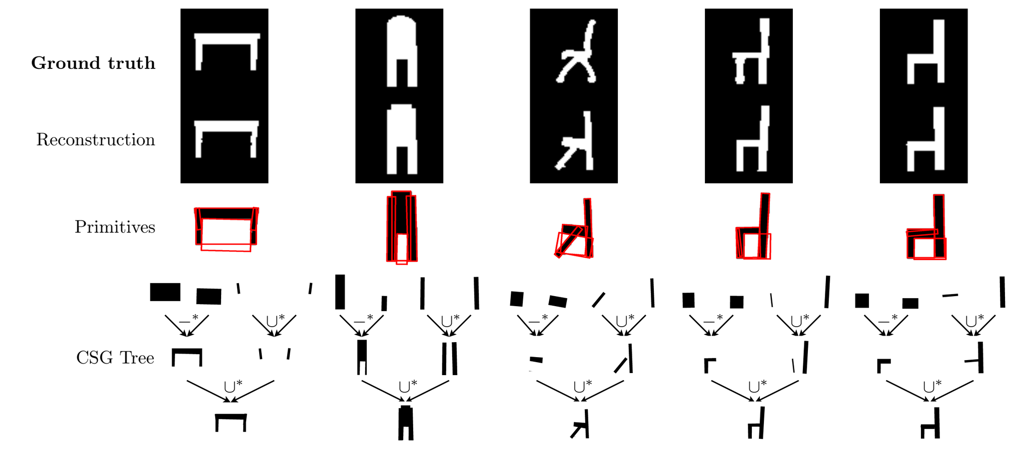
Second stage - when $\alpha \leq 0.05$

$$\mathcal{L}^*_{\text{total}} = \mathcal{L}_{\text{total}} + \lambda_{\tau}\sum_{l=1}^{L}|\tau^{(l)}|$$

Pushes CSG operations towards one-hot vectors

Results - 2D CAD Dataset

Method	Mode	k	$i=0$	$i=\infty$
CSG-NETSTACK	Supervised	1	3.98	2.25
CSG-NETSTACK	Supervised	10	1.38	0.39
CSG-NETSTACK	RL	1	1.27	0.57
CSG-NETSTACK	RL	10	1.02	**0.34**
Our	Unsupervised	1	**0.32**	-



Results - 3D ShapeNet

CD - Chamfer Distance $\times 10^3$

	High interpretability			Low interpretability	
	Ours	VP	SQ	BAE	BSP-Net
CD	2.085	2.259	1.656	1.592	**0.446**