Towards More Physically Plausible Generative Models

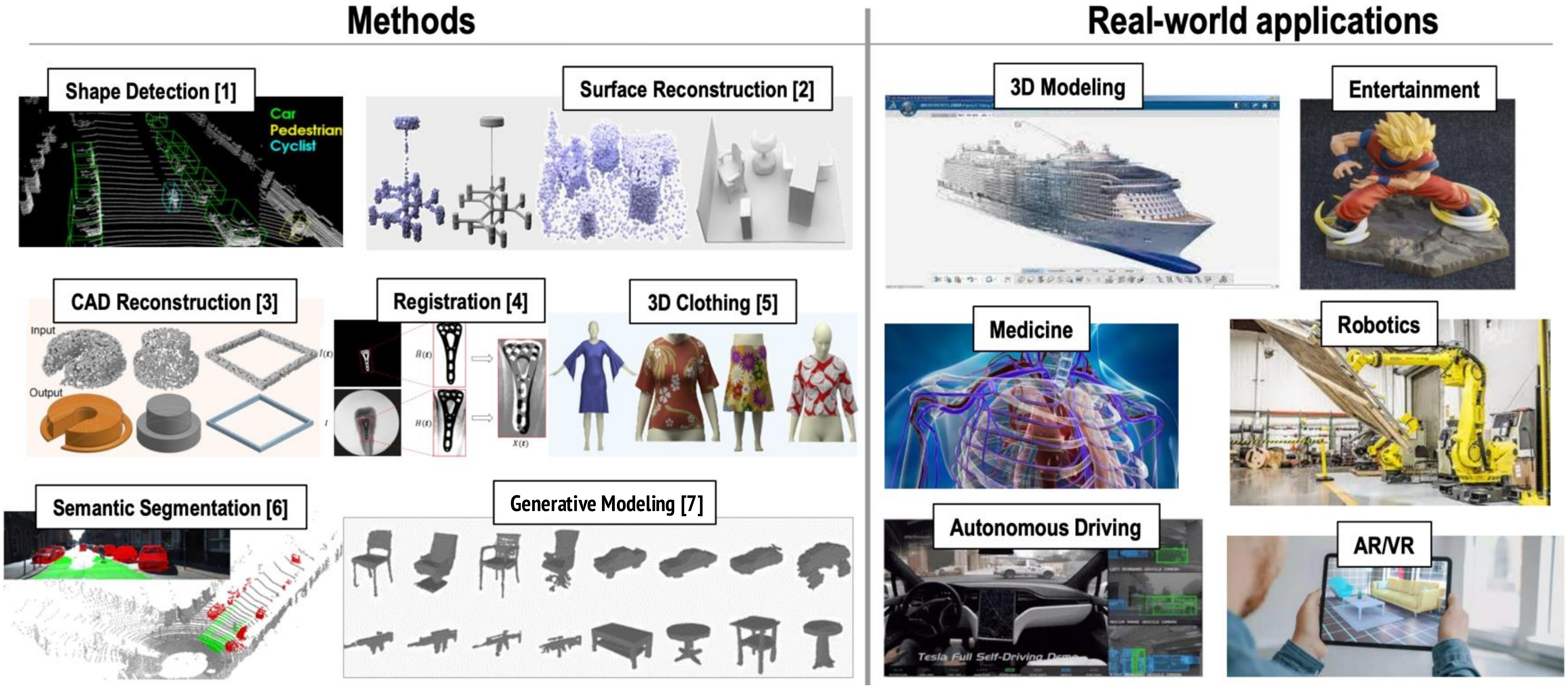
CVPR 2025 Workshop:
Ind3D: *Enforcing Inductive Bias in 3D Generation from Geometric, Physical, Topological, and Functional Perspectives*

Maks Ovsjanikov

Joint work with: *M. Mezghanni,, M. Boulkenafed, L. Maillard, T. Durand, …*

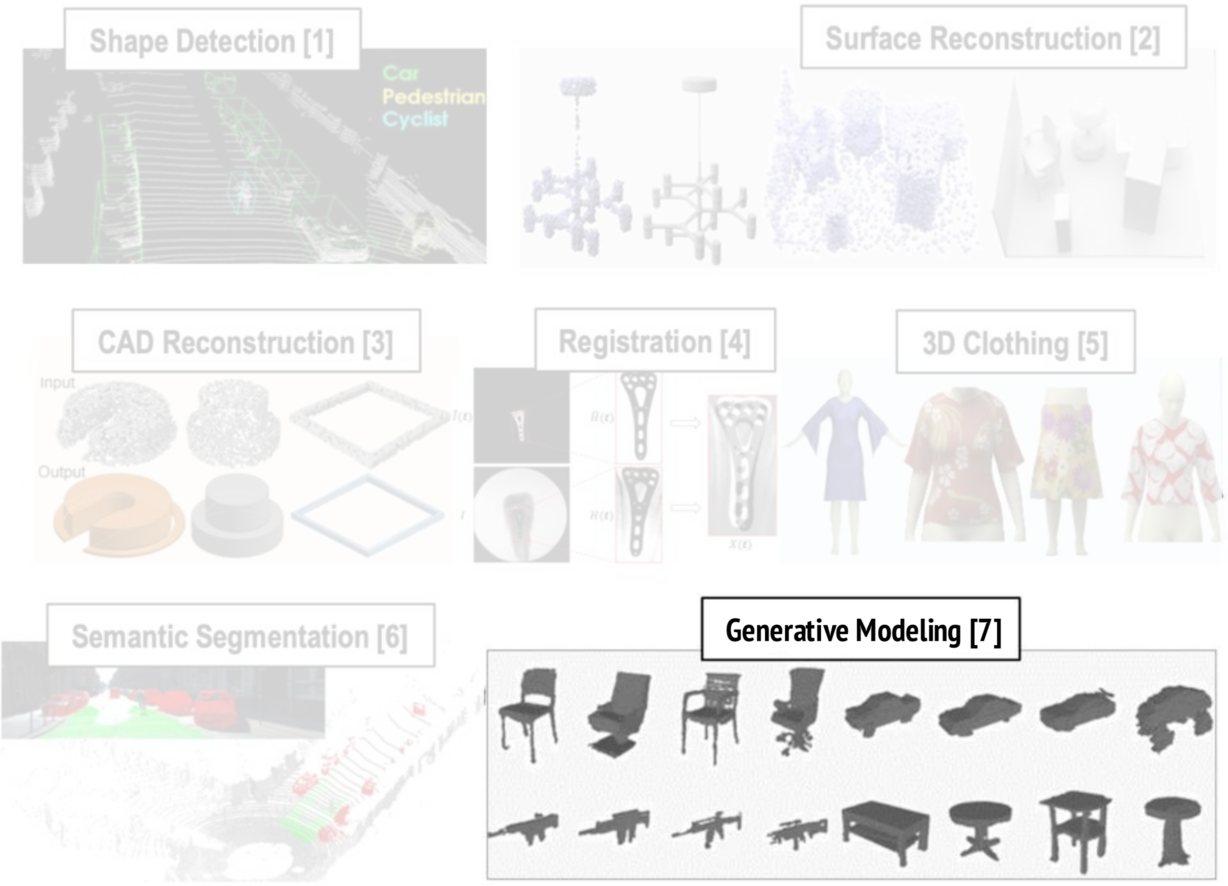
ÉCOLE POLYTECHNIQUE
INSTITUT POLYTECHNIQUE DE PARIS
Google DeepMind

Motivation: Gap Between Methodology and Applications

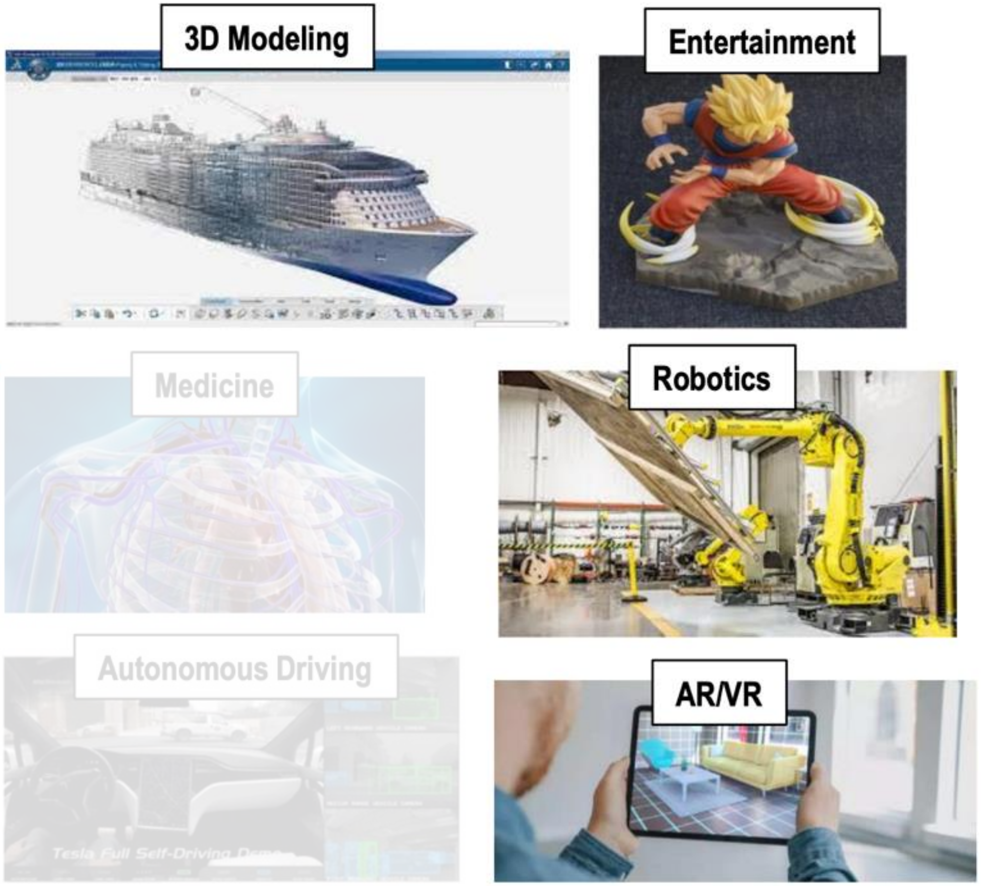


Motivation: Gap Between Methodology and Applications

Methods



Real-world applications



3D Modeling is Expensive

Image source: The Next Leap: How A.I. will change the 3D industry - Andrew Price

3D Generative Modeling Goals

Realism



Image source: Evermotion

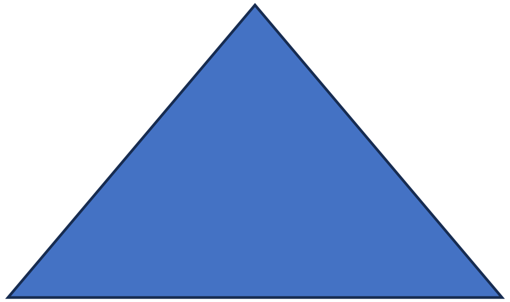
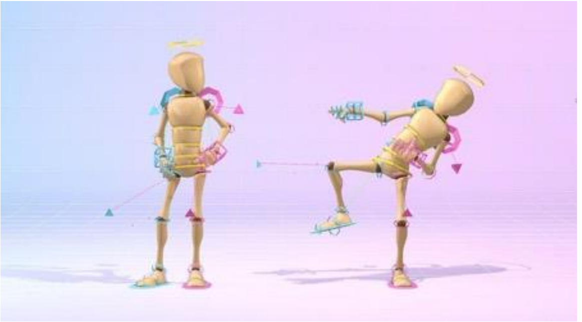


Image source: Objaverse-XL

Data Efficiency



Source: Ugur Ulvi Yetiskin

Controllability

Plan for Today

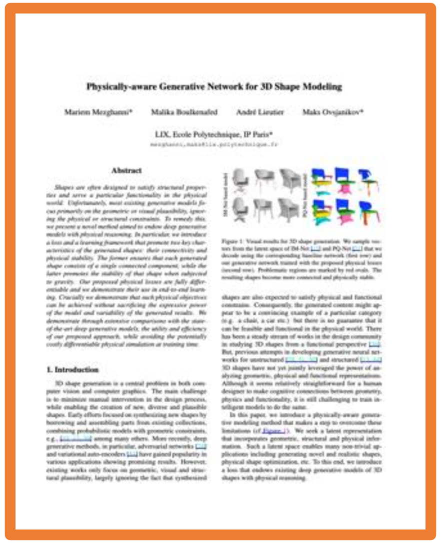
Two major topics:

1. Generative Modeling for 3D Objects

2. Generative Modeling for indoor 3D Scenes

Mariem Mezghanni

Léopold Maillard



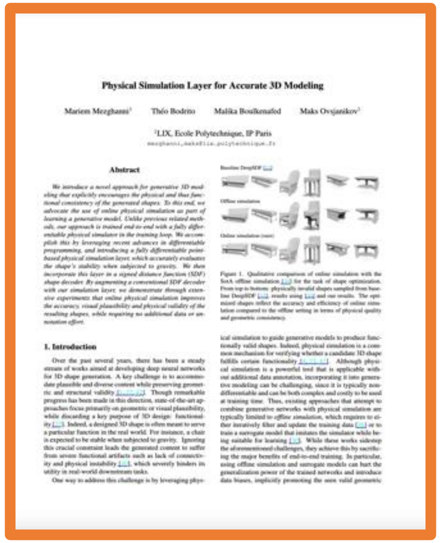
Physically-aware Generative Network for 3D Shape Modeling

Mariem Mezghanni* Malika Boulkenafed André Lieutier Maks Ovsjanikov*

LIX, Ecole Polytechnique, IP Paris*

Abstract

1. Introduction



Physical Simulation Layer for Accurate 3D Modeling

Mariem Mezghanni[1] Théo Bodrito Malika Boulkenafed Maks Ovsjanikov[1]

[1]LIX, Ecole Polytechnique, IP Paris

Abstract

1. Introduction

DeBaRA: Denoising-Based 3D Room Arrangement Generation

Léopold Maillard[1,2] Nicolas Sereyjol-Garros[2] Tom Durand[2] Maks Ovsjanikov[1]

Abstract

1 Introduction

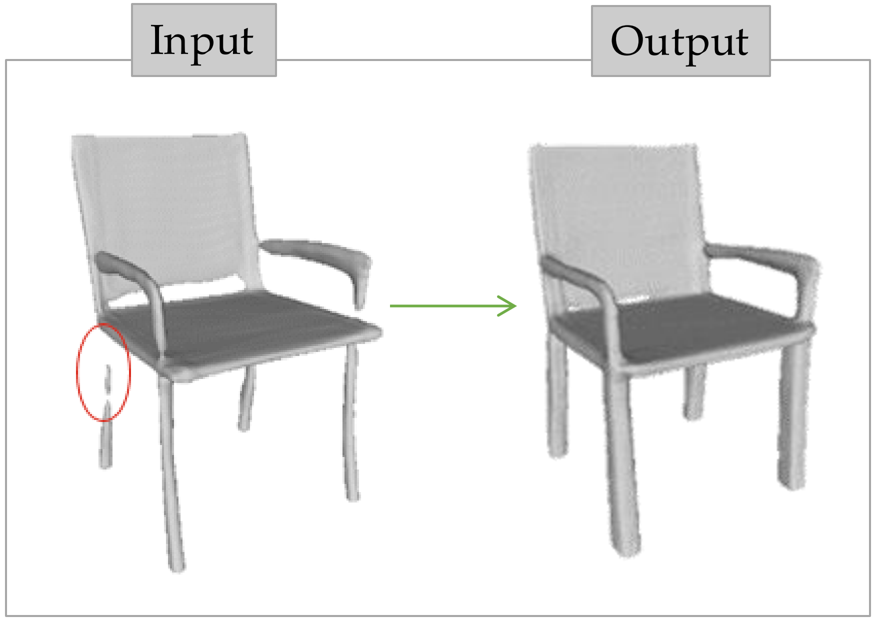
CVPR 2021

CVPR 2022 (Oral)

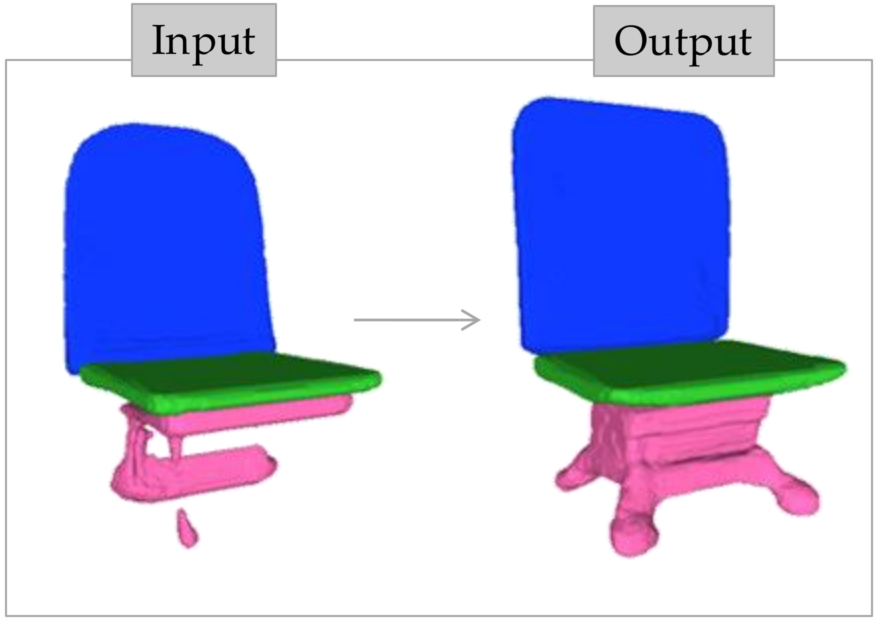
NeurIPS 2024

Overall Goal

Endow generative networks biases to promote **connectivity** and **physical stability.**



Connectivity



Physical stability

M. Mezghanni, et al. "Physically-aware generative network for 3d shape modelling," CVPR 2021.

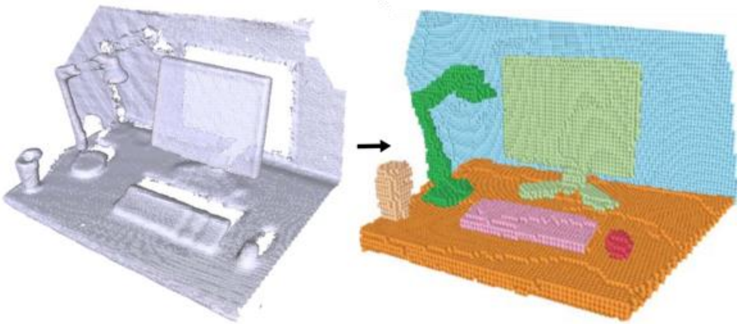
M. Mezghanni, et al. "Physical simulation layer for accurate 3d modeling," CVPR 2022.

Why connectivity and physical stability ?

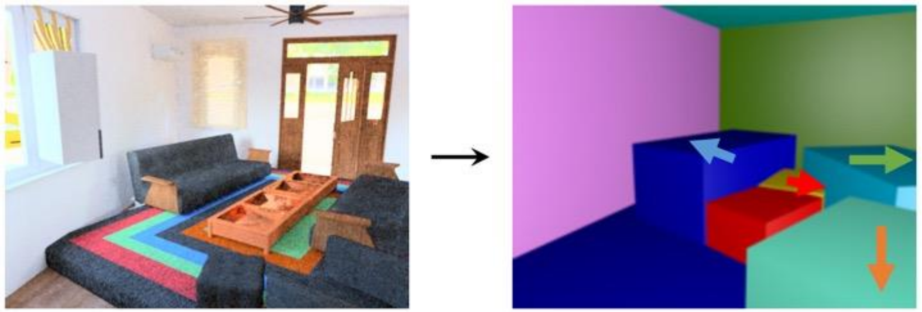
- Frequent cause of failures
- Represent a shared functional requirement across different shape categories
- Physical stability has proved beneficial for boosting many computer vision and graphics tasks.

Example functional failures

3D printing [1]



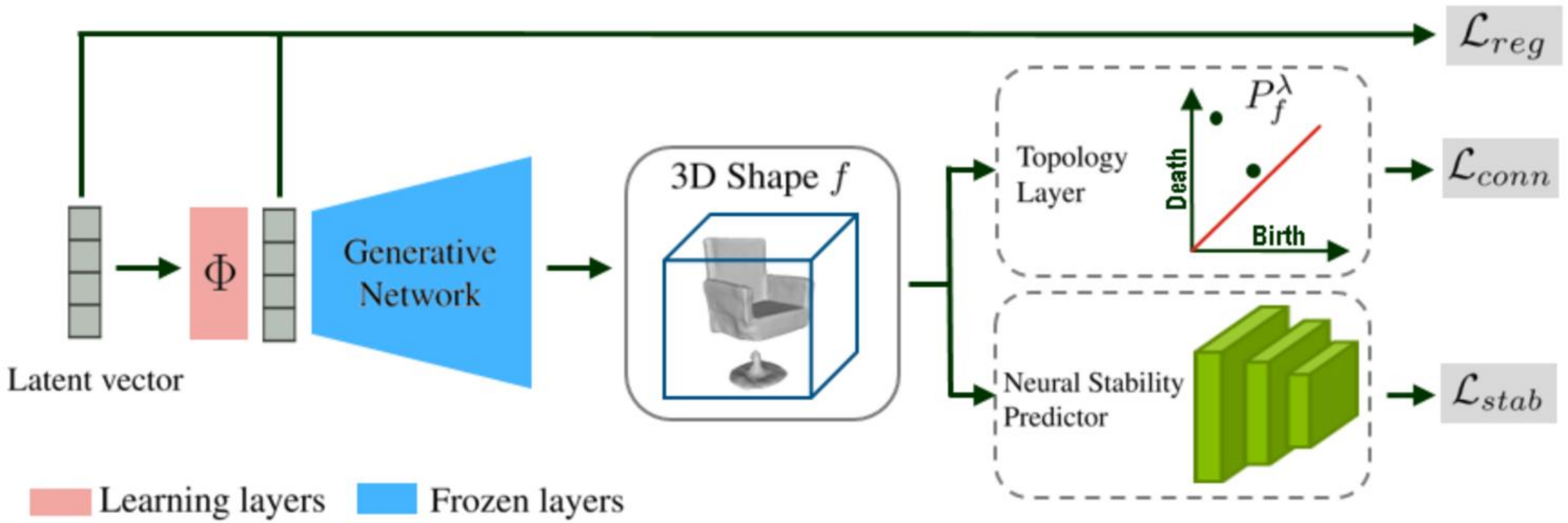
Scene segmentation [2]



3D reconstruction [3]

[1] Make It Stand: Balancing shapes for 3D fabrication, Prévost et al., ACM SIGGRAPH, 2013.
[2] Beyond point clouds: Scene understanding by reasoning geometry and physics? Zheng et al., CVPR, 2013
[3] Learning to exploit stability for 3d scene parsing.. Du et al., NeurIPS, 2018

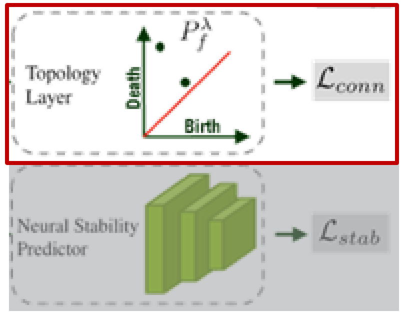
Method – Overview

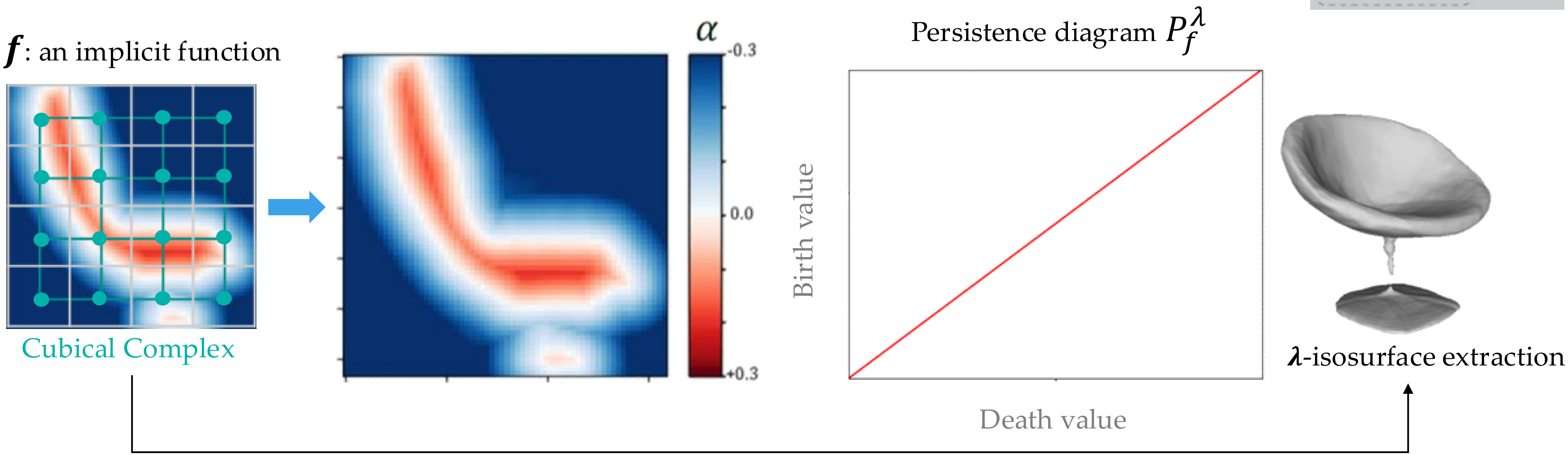


M. Mezghanni, et al. "Physically-aware generative network for 3d shape modelling," CVPR 2021.

Method

Differentiable Connectivity Loss via Persistent Homology

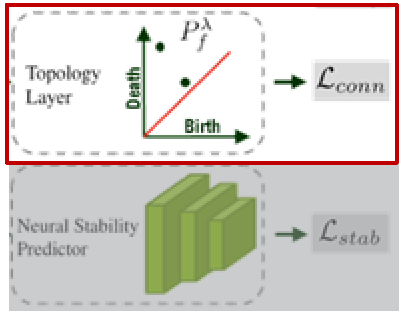


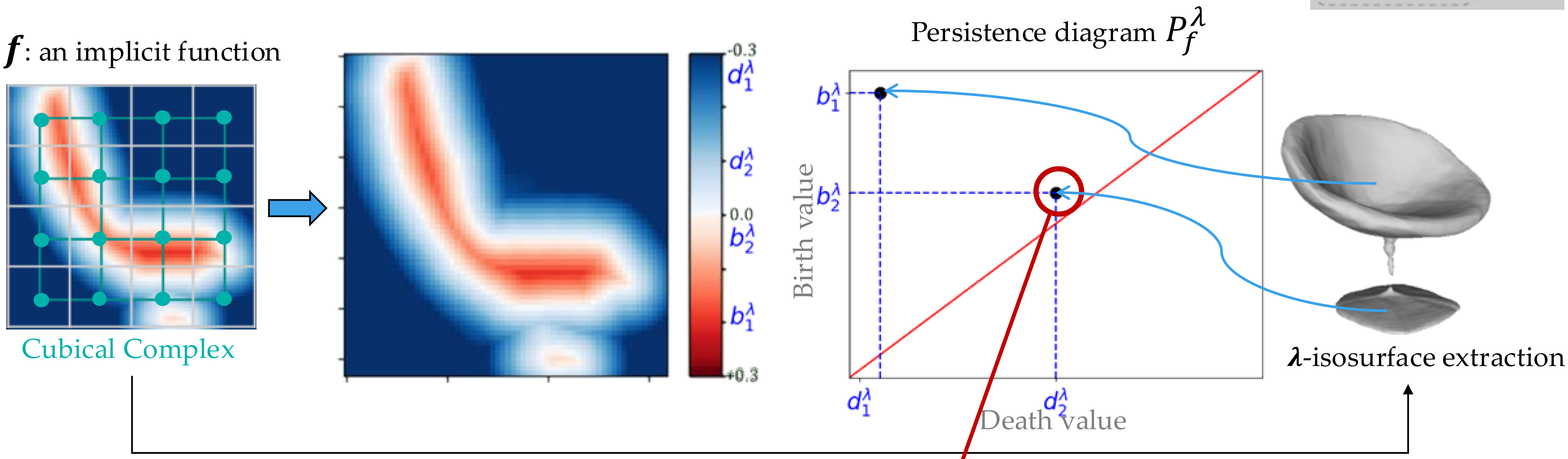


[1] A topology layer for machine learning, Gabrielsson et al., PMLR, 2020
[2] Topological Function Optimization for Continuous Shape Matching, Poulenard et al., CGF, 2018

Method

Differentiable Connectivity Loss via Persistent Homology



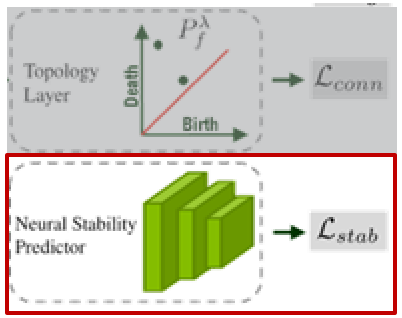


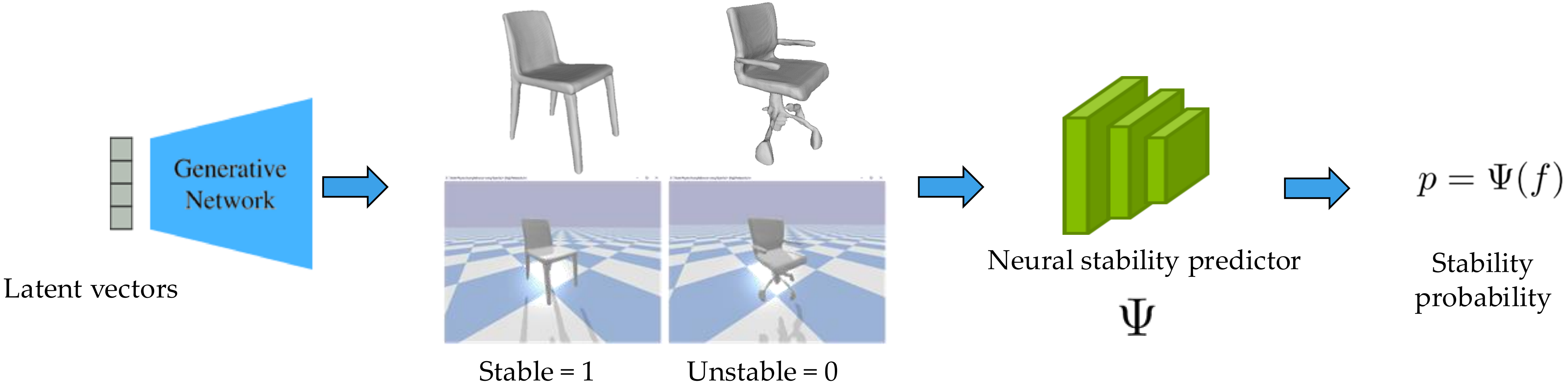
$$\mathcal{L}_{conn} = \sum_{(b_i^\lambda, d_i^\lambda) \in P_f^\lambda; \geq 2} (b_i^\lambda - d_i^\lambda)$$

[1] A topology layer for machine learning, Gabrielsson et al., PMLR, 2020
[2] Topological Function Optimization for Continuous Shape Matching, Poulenard et al., CGF, 2018

Method

Physical Stability Loss via a Surrogate Model

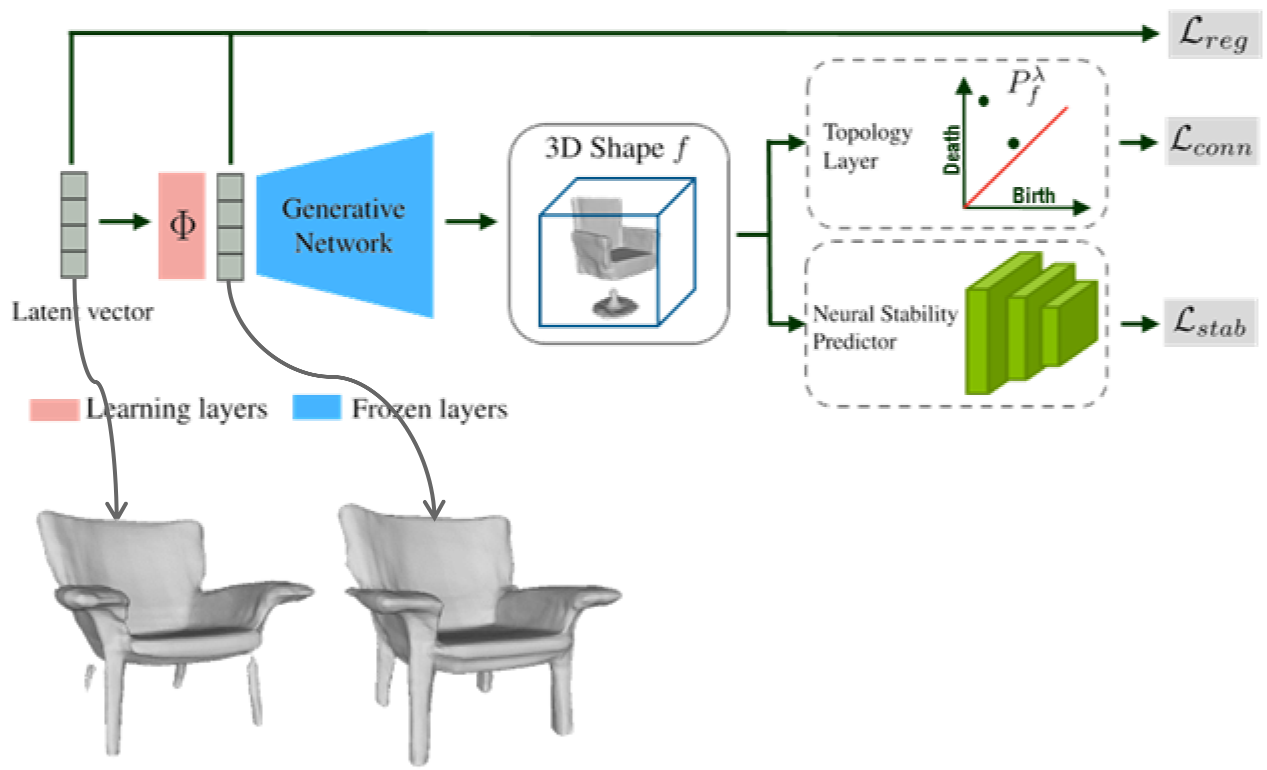




$$\mathcal{L}_{stab} = max(1 - \Psi(f), \alpha); \alpha = 0.5$$

M. Mezghanni, et al. "Physically-aware generative network for 3d shape modelling," CVPR 2021.

Learning Framework



Training stages:

1. Train a generative network G
2. Freeze G and train a **mapping network** Φ

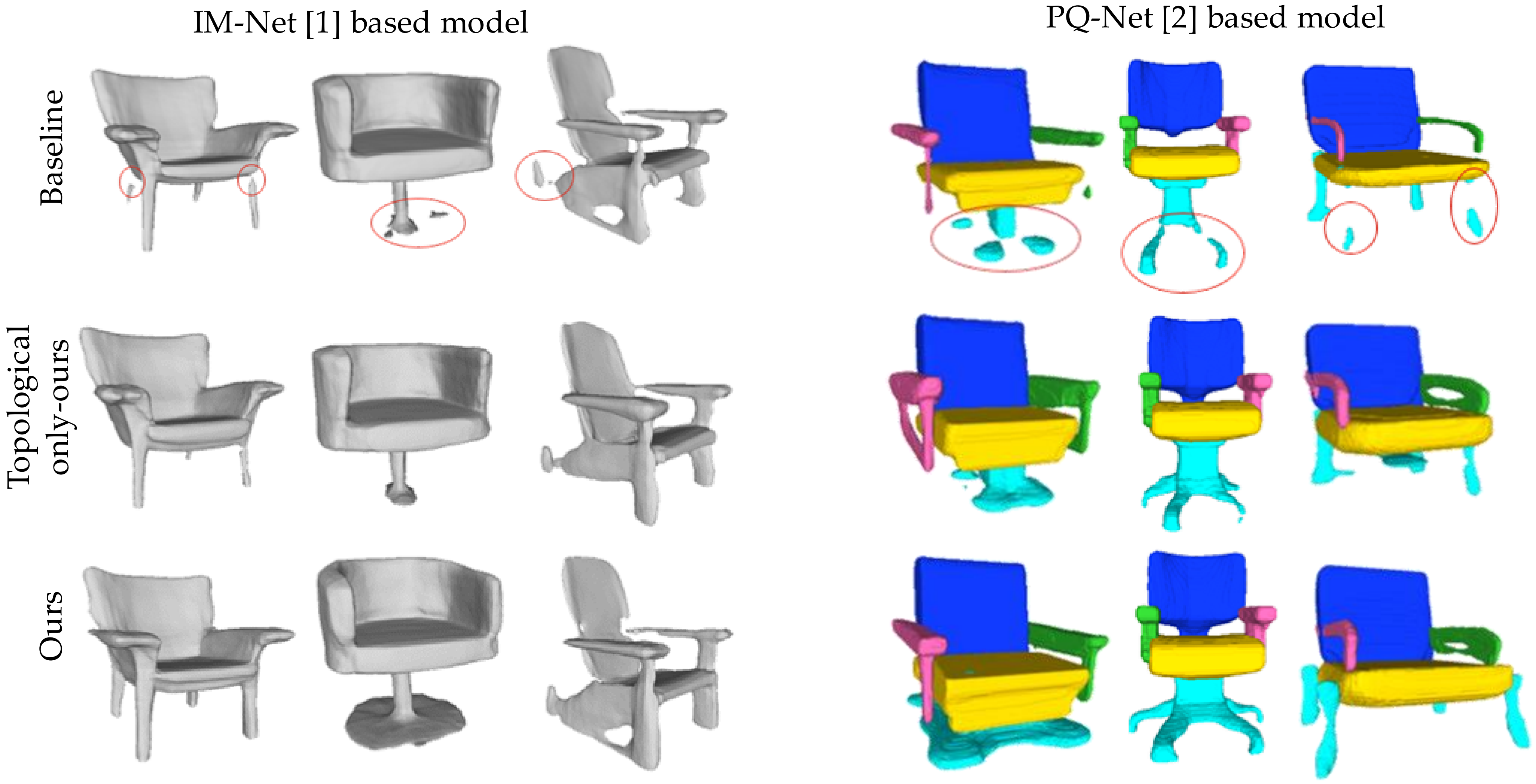
Motivation: **preserve the diversity and quality of the generated content** since the latent space of objects is unchanged.

$$\mathcal{L}_{reg}(z) = \|z - \Phi(z)\|_2; \quad \mathcal{L}_{total} = \mathbb{E}_{z \in \mathcal{V}}[\mathcal{L}_{reg} + \alpha_c \mathcal{L}_{conn} + \alpha_s \mathcal{L}_{stab}]$$

Learned latent space of shapes

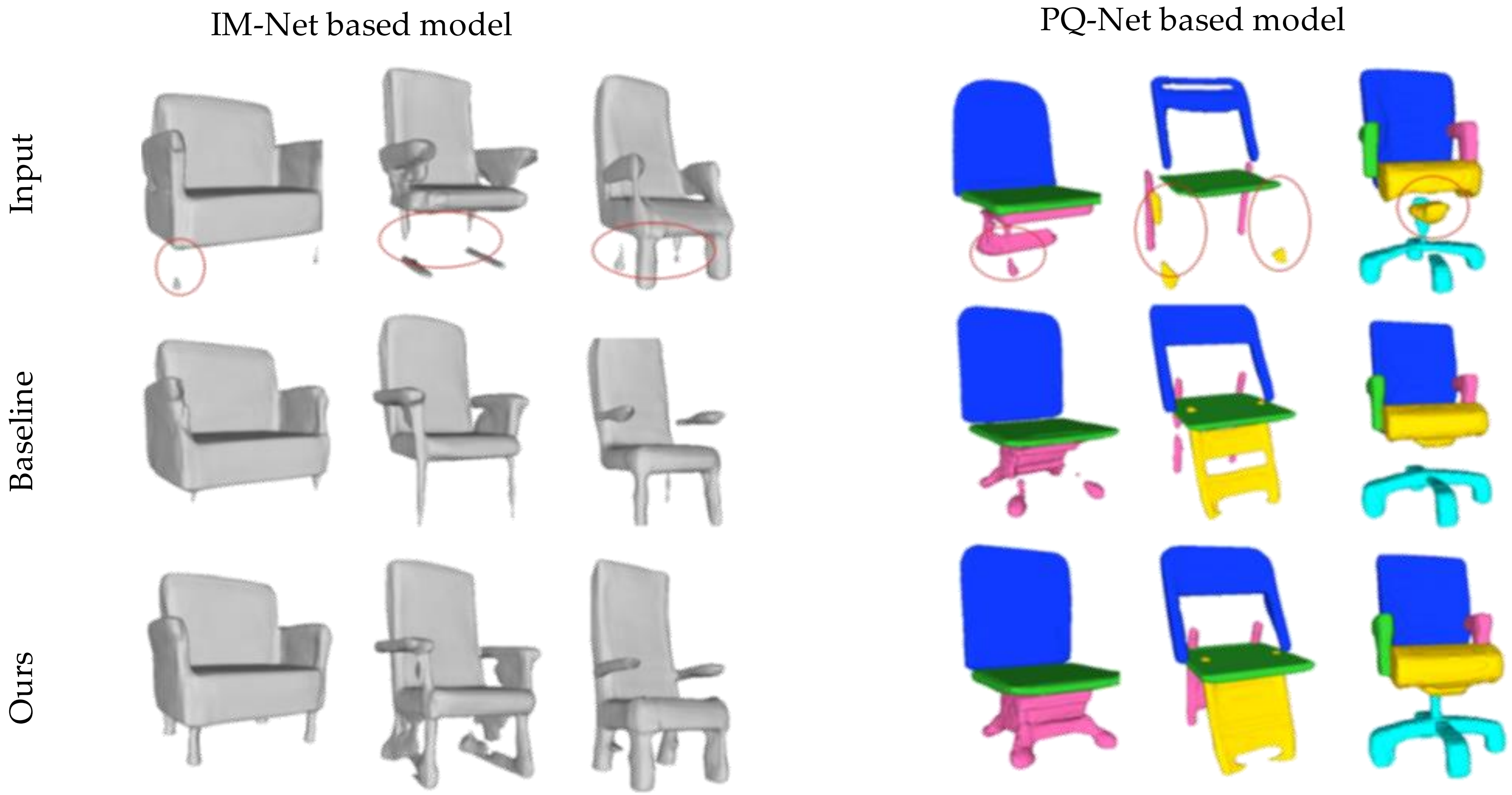
M. Mezghanni, et al. "Physically-aware generative network for 3d shape modelling," CVPR 2021.

Results: Shape Generation



[1] IM-NET: Learning implicit fields for generative shape modeling. Chen et al., CVPR, 2019
[2] PQ-NET: A generative part Seq2Seq network for 3D shapes. Wu et al., CVPR, 2020

Results: Shape Correction

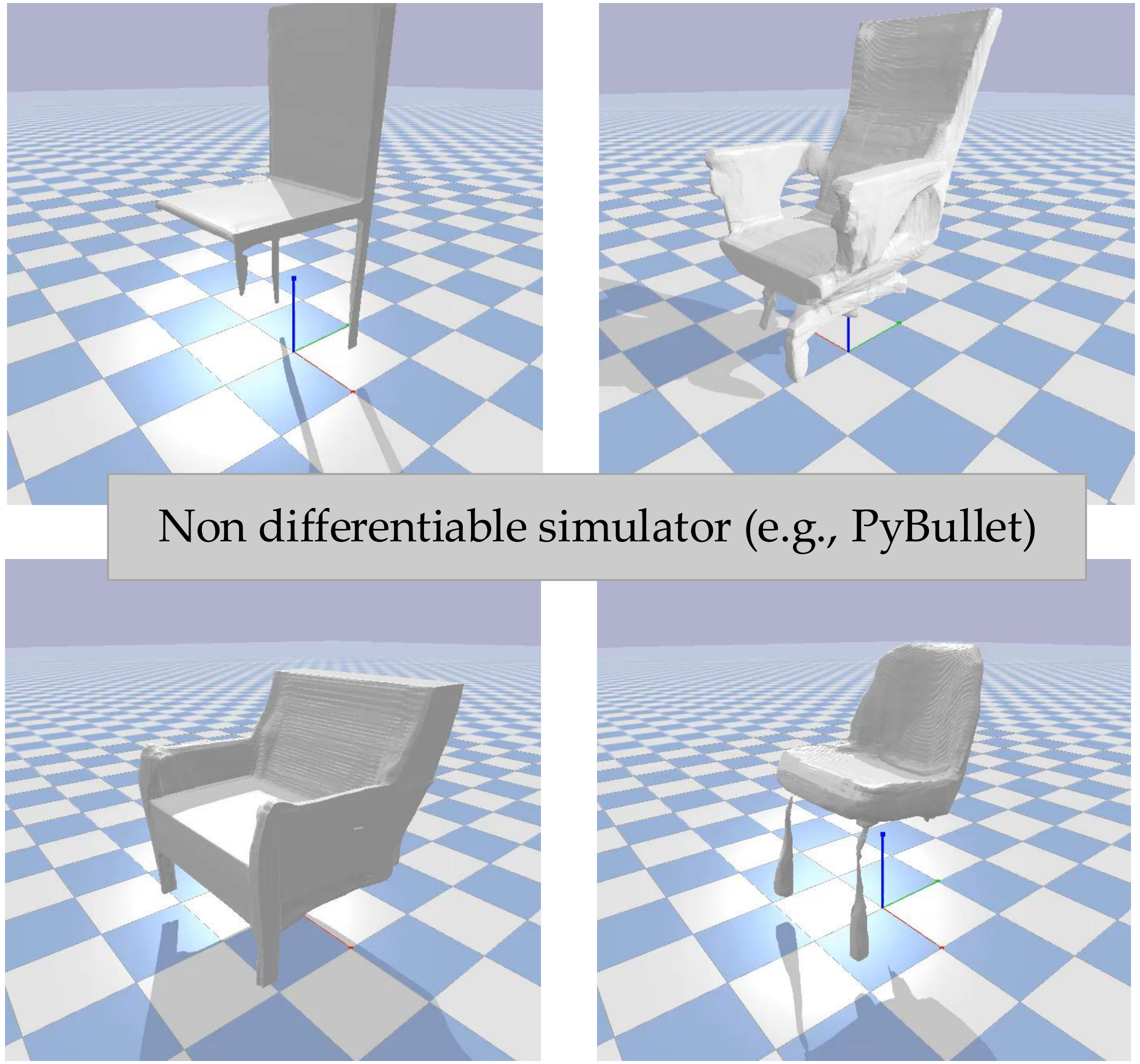


M. Mezghanni, et al. "Physically-aware generative network for 3d shape modelling," CVPR 2021.

Offline vs Online Simulation

Offline simulators

- **Easy-to-use and mature**
- **Non-differentiable**: need to be combined with gradient approximation methods (instability of numerical gradients).



Non differentiable simulator (e.g., PyBullet)

Our contribution:

- Build a **differentiable** point-based physical simulator
- Learn generative network DeepSDF [1] with **online physical simulation**.

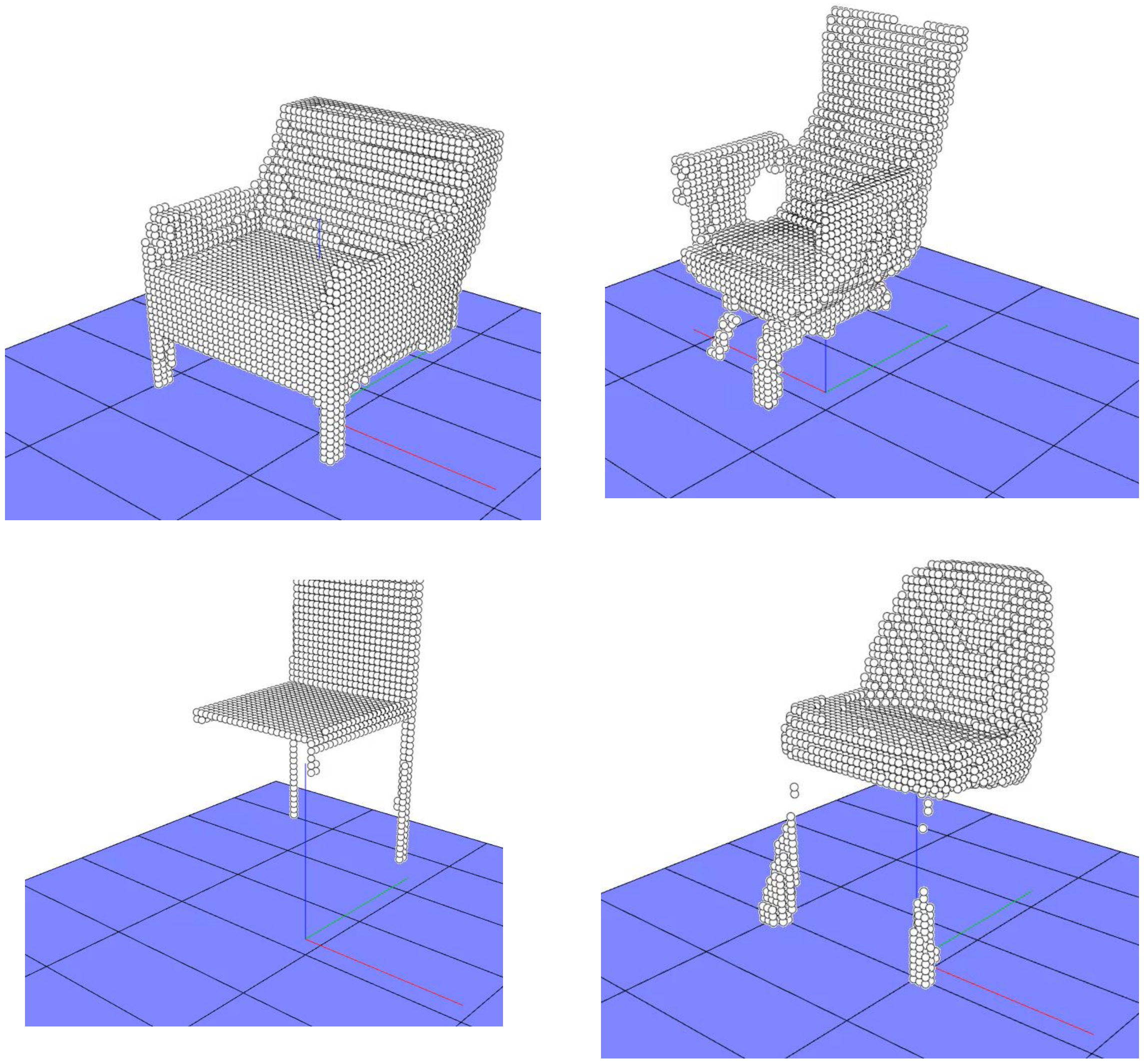
M. Mezghanni, et al. "Physical simulation layer for accurate 3d modeling," CVPR 2022.

Differentiable simulator Ψ

We build a **differentiable simulator Ψ** using the **DiffTaichi [1]** framework :

$$\Psi(\boldsymbol{\mathcal{C}}) = \{(\mathrm{p}_t, \mathrm{r}_t);\, t \in [1, T]\}$$

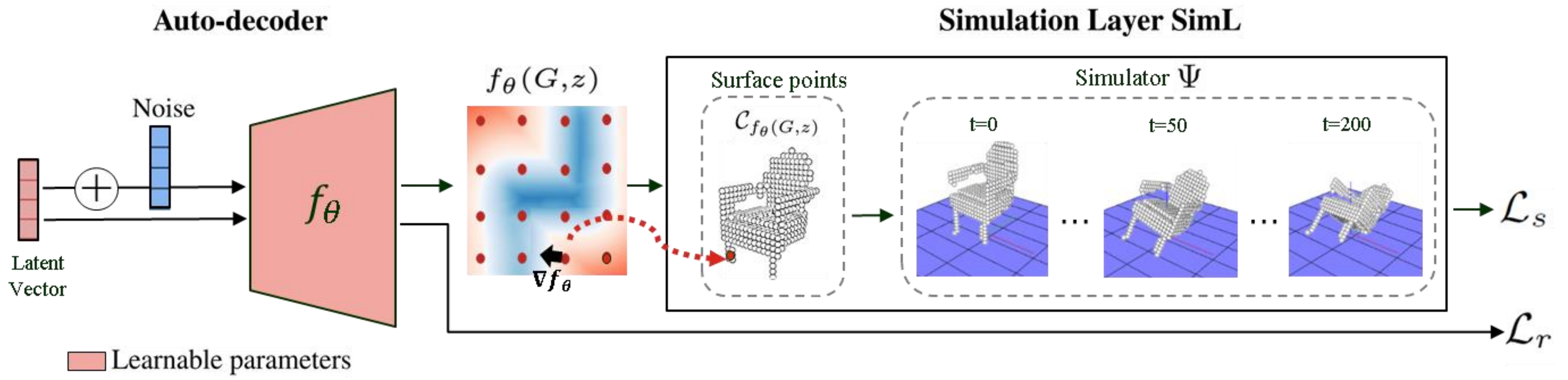
DiffTaichi naturally supports simulation of a **point cloud $\boldsymbol{\mathcal{C}}$** . We simulate $\boldsymbol{p_t}$ and $\boldsymbol{r_t}$: the *position* and the *rotation* of $\boldsymbol{\mathcal{C}}$ center of mass during simulation.



[1] DiffTaichi: Differentiable Programming for Physical Simulation. Hu et al., ICLR, 2020

M. Mezghanni, et al. "Physical simulation layer for accurate 3d modeling," CVPR 2022.

Learning Framework

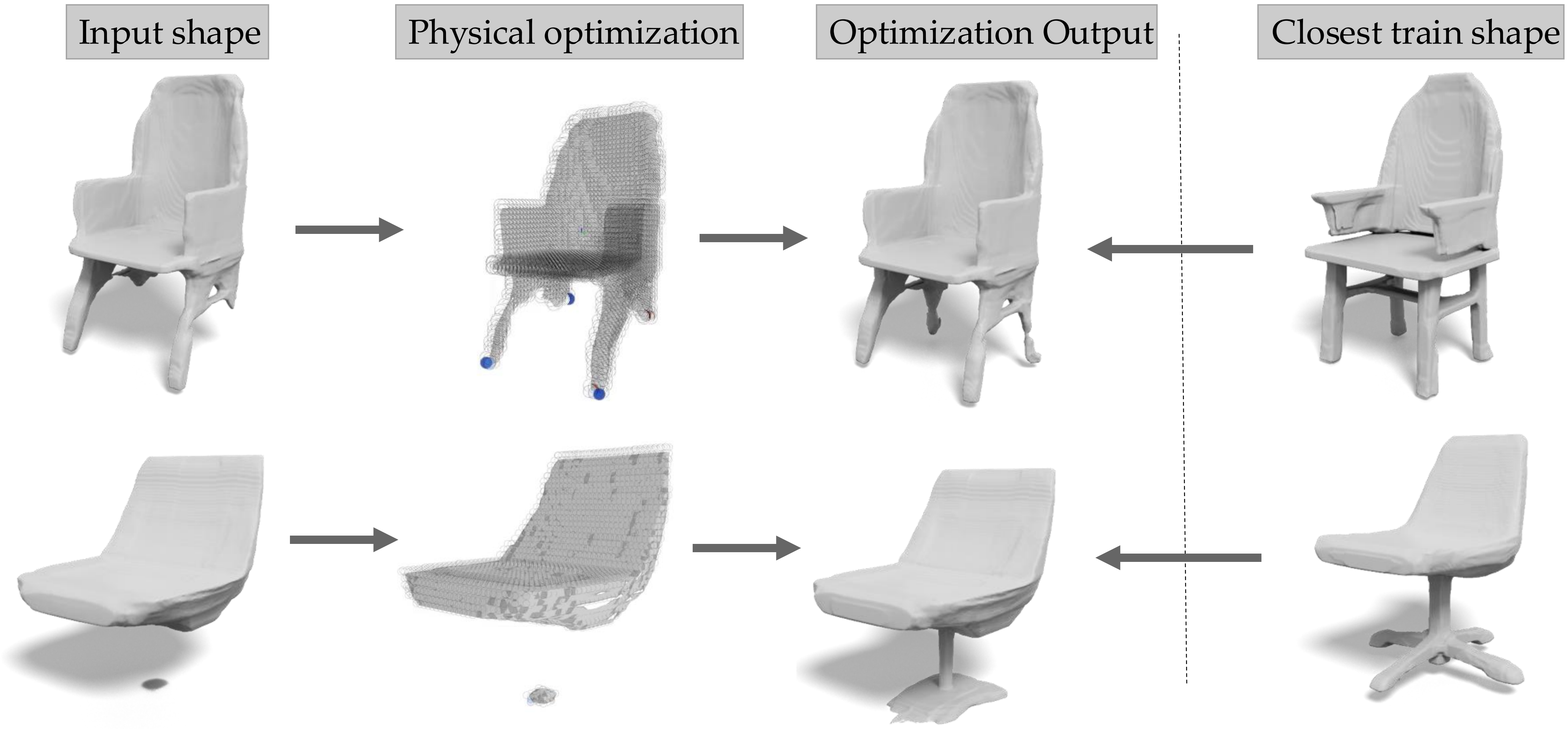


We train the auto-decoder by jointly optimizing a reconstruction and stability-based losses:

$$\mathcal{L} = \mathcal{L}_r + \alpha_s \mathcal{L}_s$$

M. Mezghanni, et al. "Physical simulation layer for accurate 3d modeling," CVPR 2022.

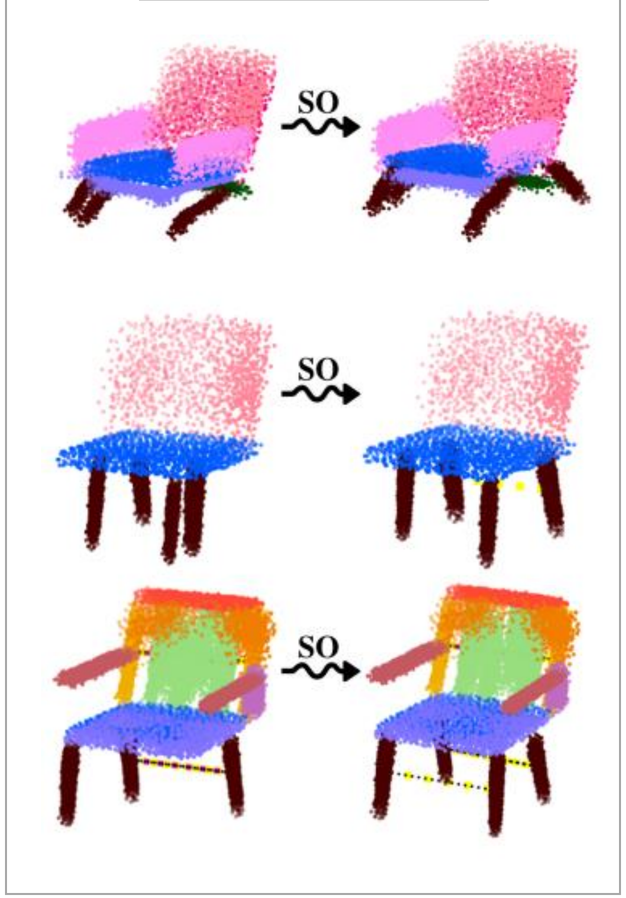
Results: Shape Optimization



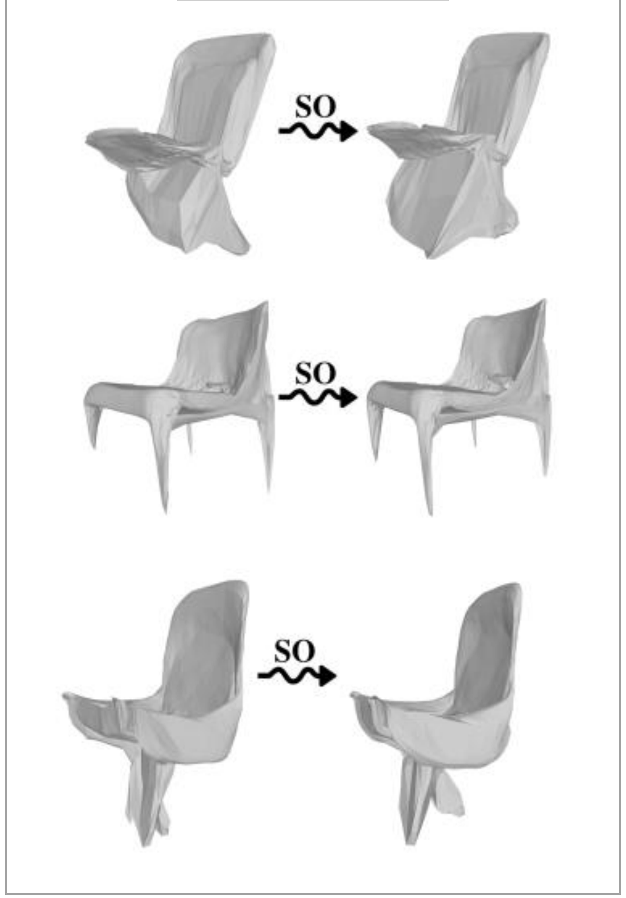
M. Mezghanni, et al. "Physical simulation layer for accurate 3d modeling," CVPR 2022.

Shape Optimization

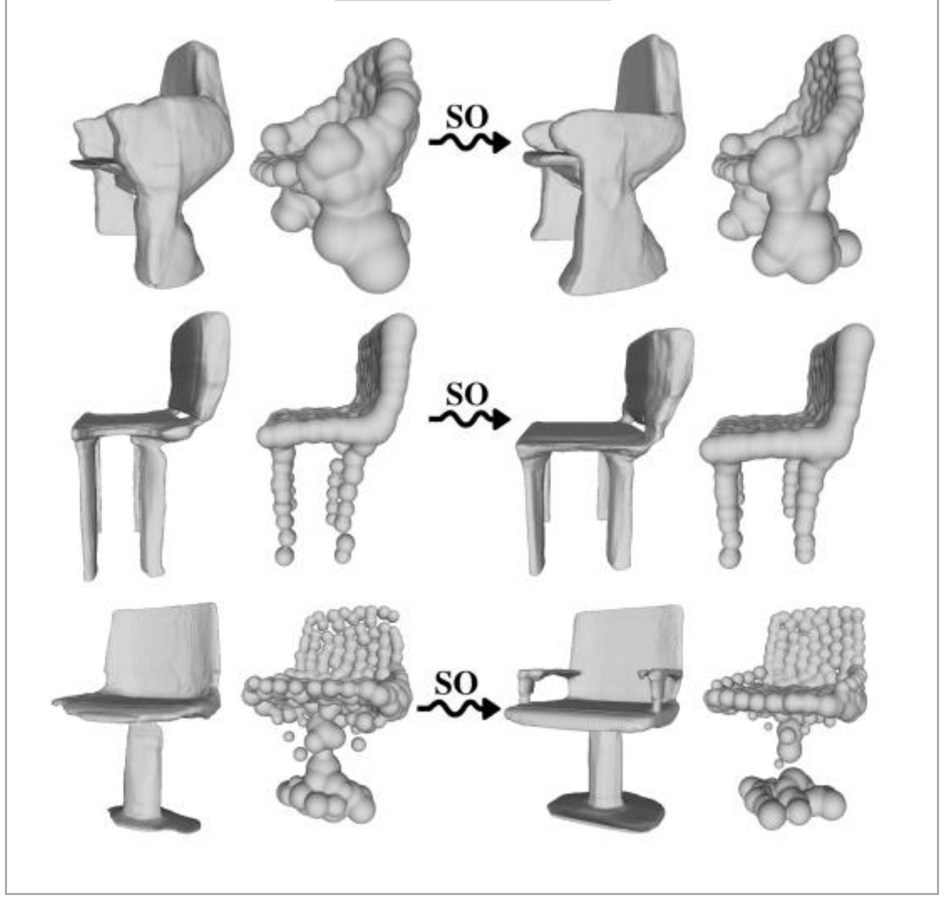
StructureNet [1]



AtlasNet [2]



DualSDF [3]



[1] StructureNet: Hierarchical graph networks for 3d shape generation. Mo et al., SIGGRAPH Asia, 2019
[2] AtlasNet: A Papier-Mache approach to Learning 3D Surface Generation. Groueix et al., CVPR, 2018
[3] Dualsdf: Semantic shape manipulation using a two-level representation. Hao et al., CVPR, 2020

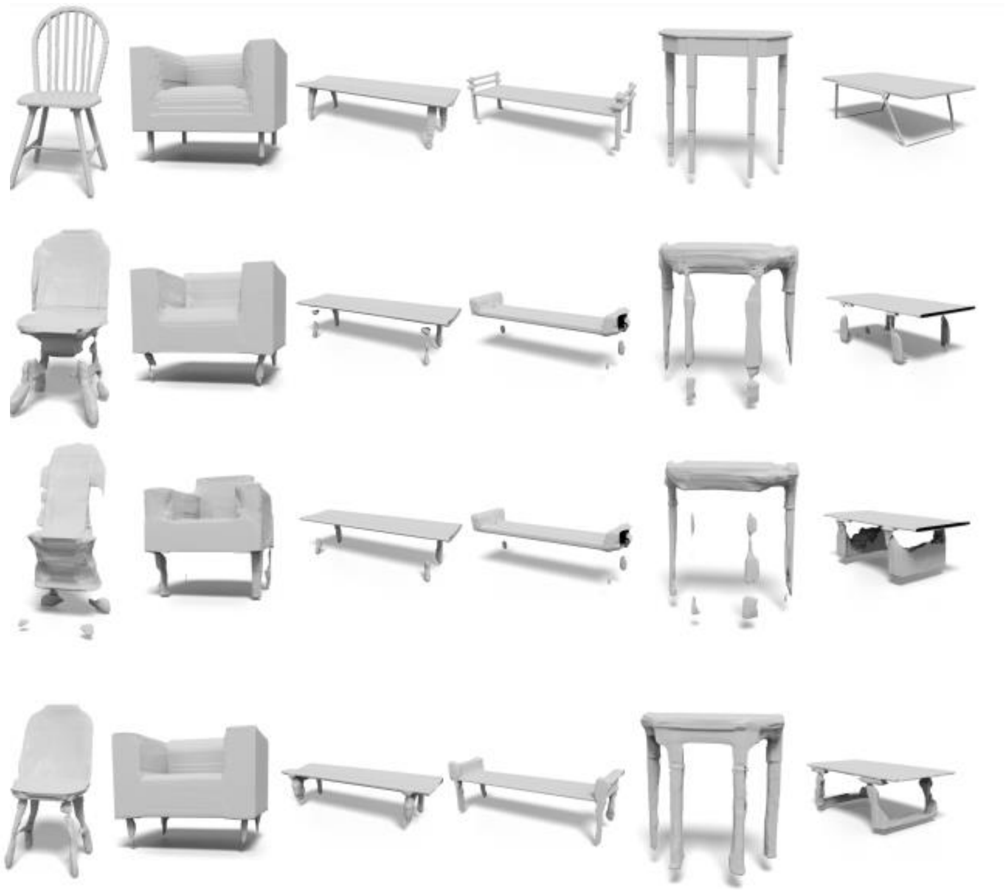
Shape Reconstruction

Ground truth

DeepSDF

DeepSDF +
Offline Simulation

Phys-DeepSDF
(ours)



M. Mezghanni, et al. "Physical simulation layer for accurate 3d modeling," CVPR 2022.

DeBaRA: Denoising-Based 3D Room Arrangement Generation

Léopold Maillard[1,2], Nicolas Sereyjol-Garros, Tom Durand[2], Maks Ovsjanikov[1]

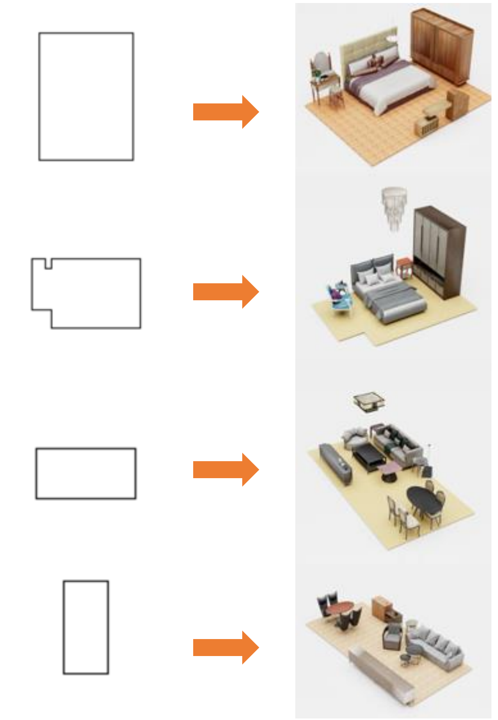
[1]LIX, École Polytechnique, IP Paris [2]Dassault Systèmes

NeurIPS 2024

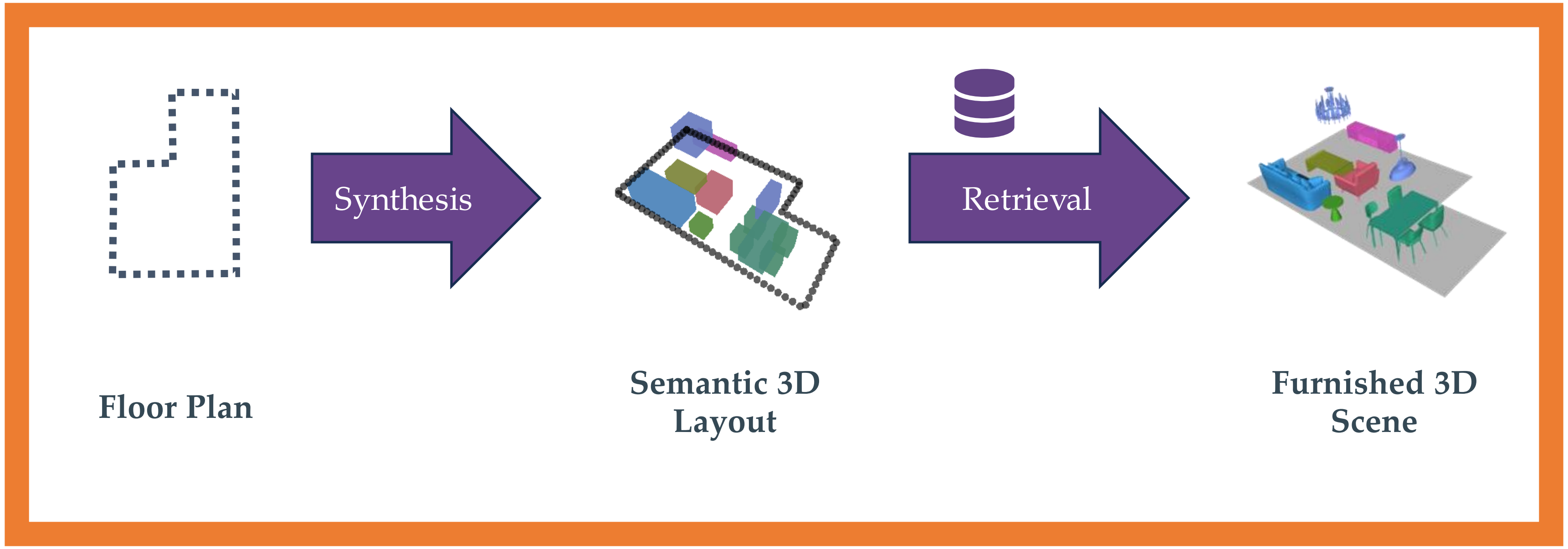
Task: Controllable 3D Indoor Scene Synthesis

Goal

Floor plan	Furnished scene



Basic Pipeline



L. Maillard, et al. "DeBaRA: Denoising-Based 3D Room Arrangement Generation," NeurIPS 2024.

Motivation: Controllable 3D Indoor Scene Synthesis

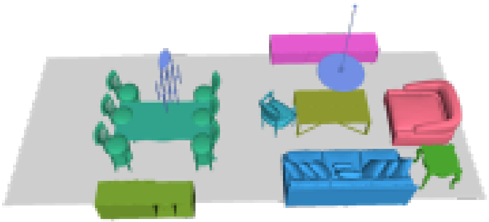
Challenges

- Inherent complexity of object **interactions**.
- Requirement to fulfill **spatial**, **ergonomic** and **functional** constraints.
- Limited amount of **training data**.

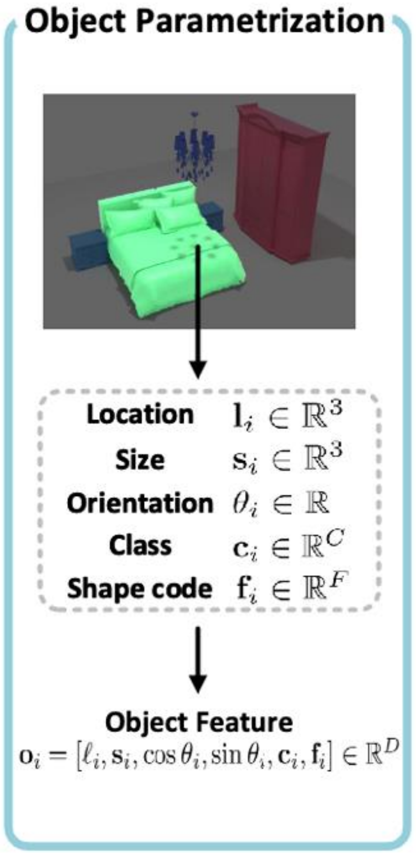
Background

- Existing methods are either **autoregressive** or use **diffusion models** for all object attributes jointly

Autoregressive



Off-the-shelf Diffusion



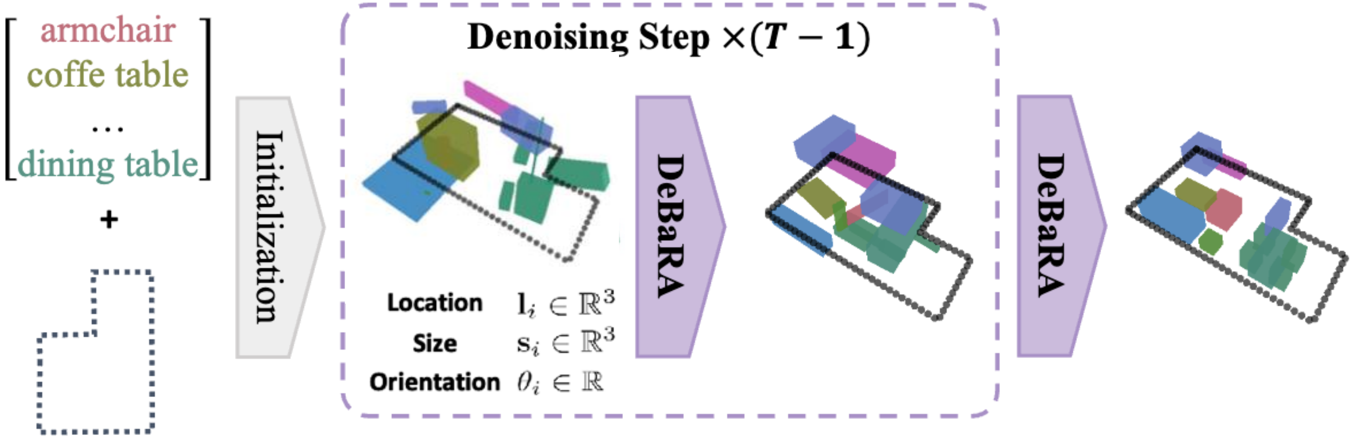
Denoising in a high-dimensional space

Mixing spatial and semantic features

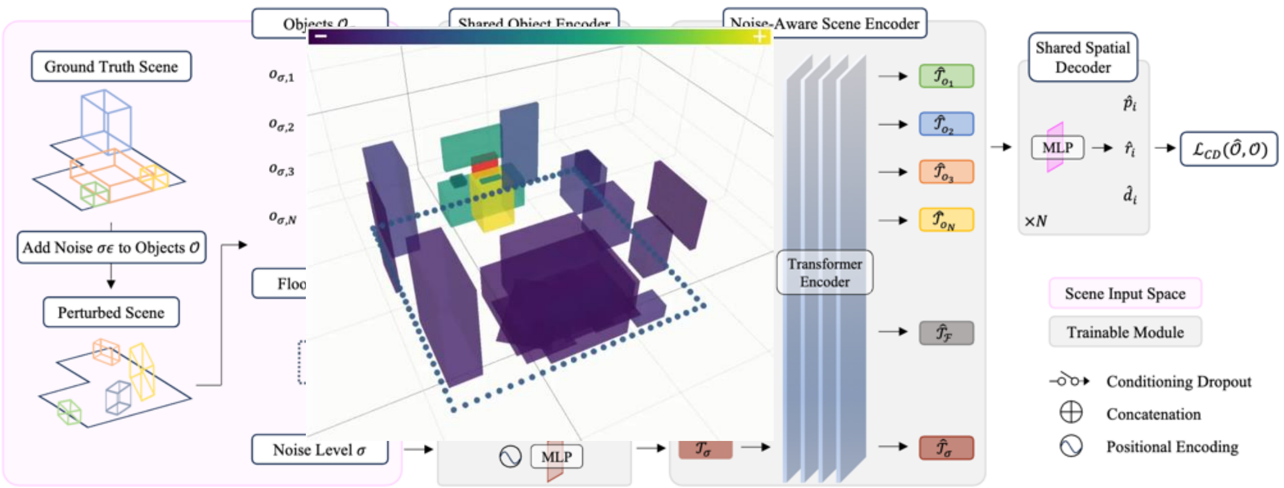
Tang et al. *DiffuScene: Denoising Diffusion Models for Generative Indoor Scene Synthesis*, in CVPR 2024

Our Approach: Separating Geometry and Semantics

Key idea:
Only denoise the spatial features. Treat the semantic features (object categories) as conditioning.



Architecture

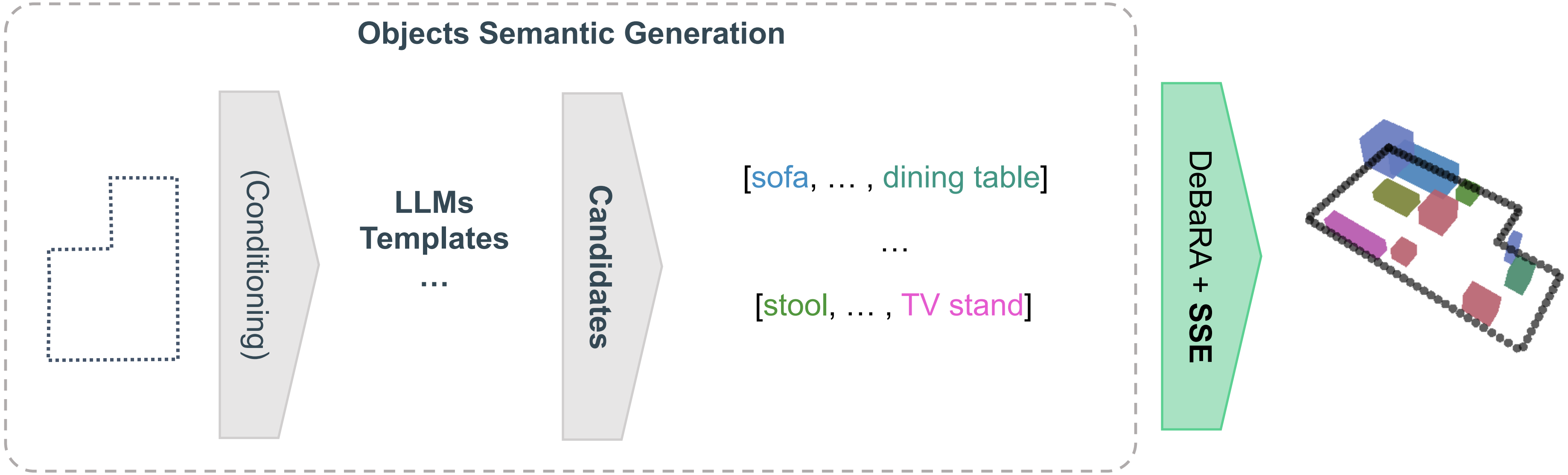


L. Maillard, et al. "DeBaRA: Denoising-Based 3D Room Arrangement Generation," NeurIPS 2024.

How to obtain the conditioning signal?

Input set of object categories can be *provided* by external sources such as a LLM [3].

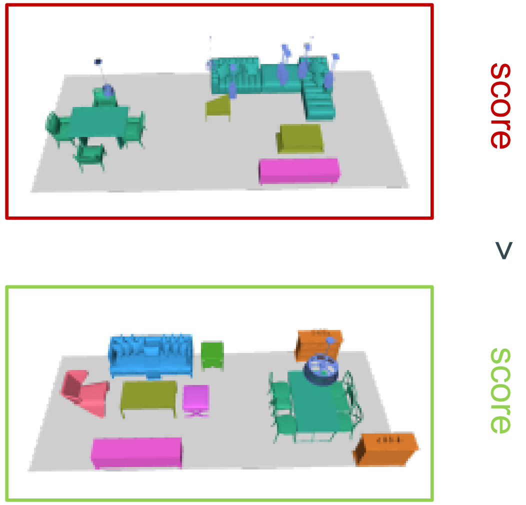
Alternatively, we propose a Self Score Evaluation (SSE) to select the sets that lead to the most realistic scenes. SSE uses density estimation with a trained model.



[3] Feng et al. *LayoutGPT: Compositional Visual Planning and Generation with Large Language Models*, in NeurIPS 2023

How to obtain the conditioning signal?

Candidate sets of object categories can be automatically generated by a LLM, and using SSE, further selected to generate a plausible 3D layout, or automatically **discarded**.

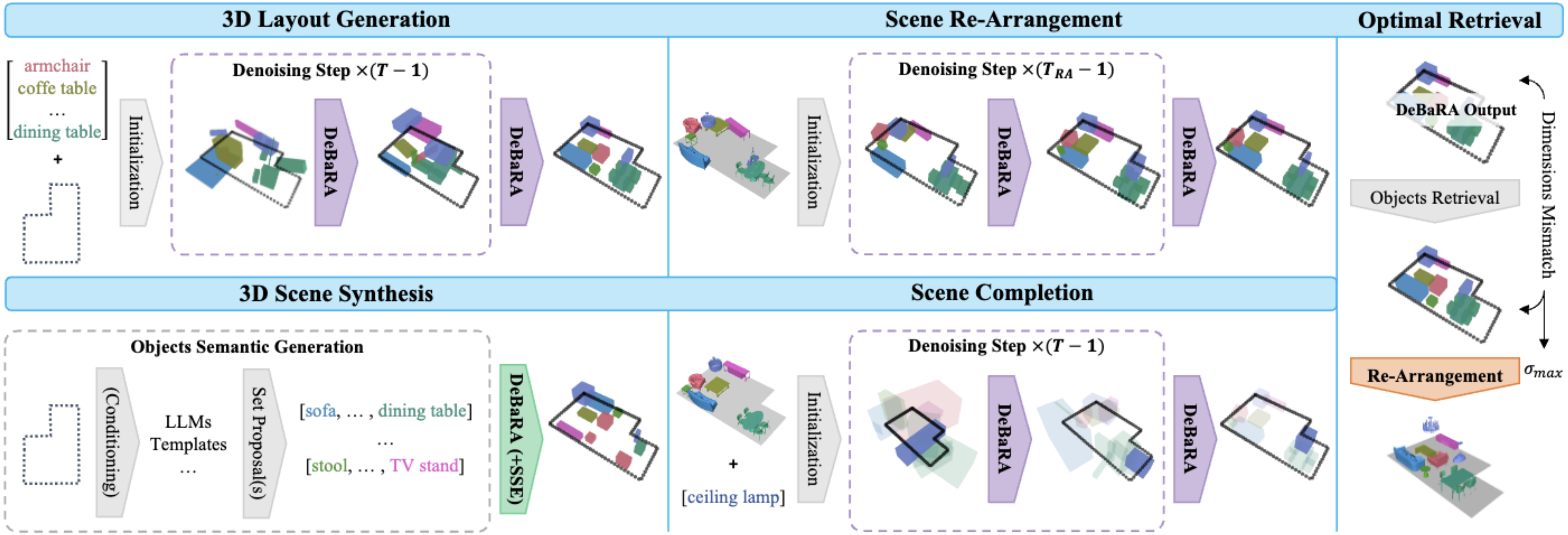


Top-down view of scenes generated by DeBaRA from LLM-generated candidates and their associated SSE scores.

L. Maillard, et al. “DeBaRA: Denoising-Based 3D Room Arrangement Generation,” NeurIPS 2024.

Many Possible Applications

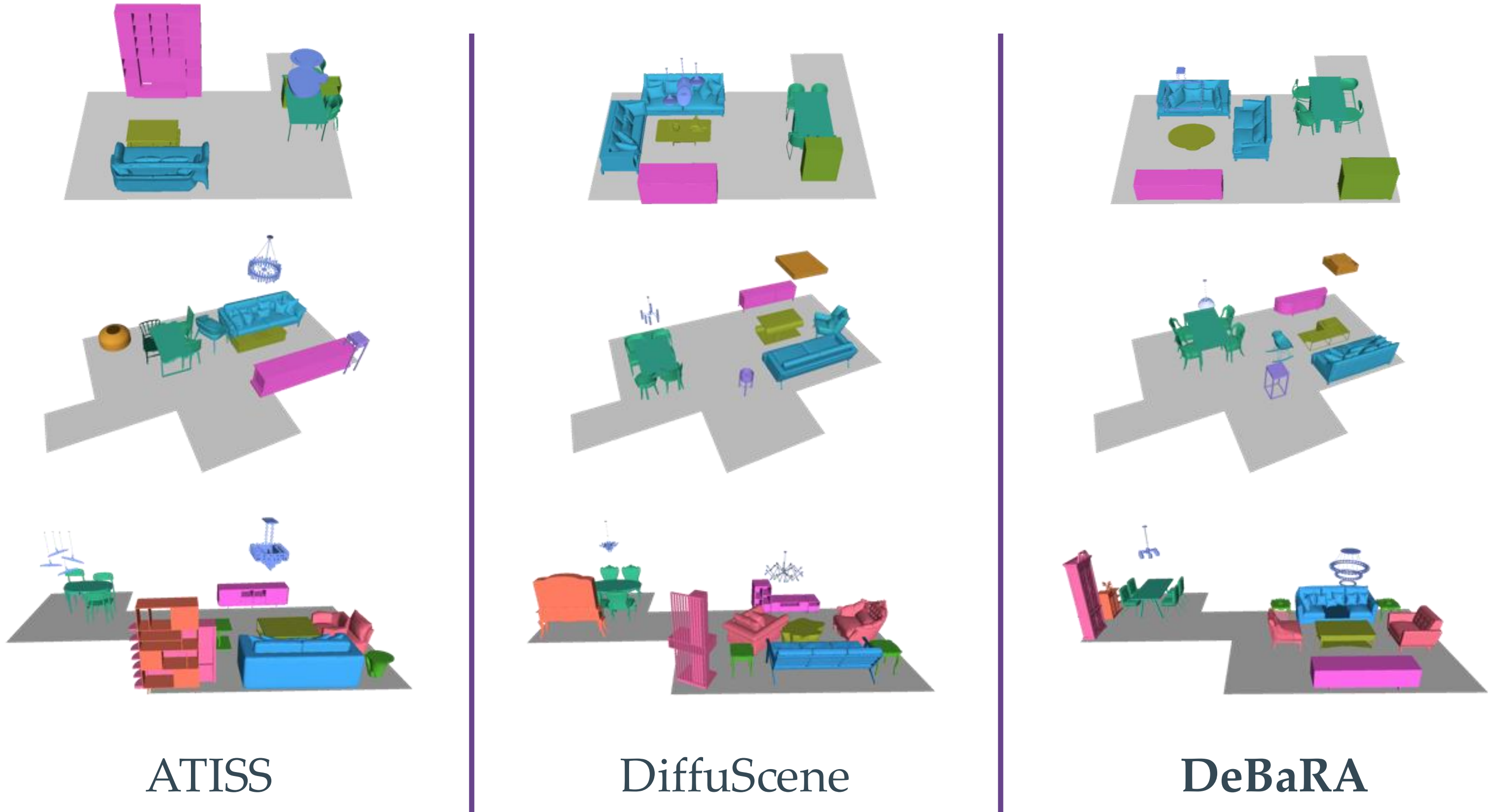
A single pre-trained model can be used for several downstream applications.



L. Maillard, et al. “DeBaRA: Denoising-Based 3D Room Arrangement Generation,” NeurIPS 2024.

Results – 3D Layout Generation, Synthesis, and Re-arrangement

Improved Accuracy in 3D Layout Generation, Scene Synthesis, and Re-arrangement.



Methods	Living Rooms			
	FID (↓)	KID (↓)	SCA (%)	OBA (↓)
LayoutGPT [9]	35.53	13.69	72.8	2913.6
ATISS [38]	25.67	8.91	71.8	857.3
DiffuScene [51]	21.54	6.40	69.7	341.1
DeBaRA (ours)	**18.89**	**3.57**	**68.3**	**167.8**

Methods	Dining Rooms			
	FID (↓)	KID (↓)	SCA (%)	OBA (↓)
LayoutGPT [9]	32.80	8.99	67.6	2447.4
ATISS [38]	28.05	9.26	63.2	702.4
DiffuScene [51]	23.06	5.35	57.7	266.4
DeBaRA (ours)	**22.04**	**4.41**	**52.4**	**132.8**

Especially strong improvement in physical consistency.

L. Maillard, et al. "DeBaRA: Denoising-Based 3D Room Arrangement Generation," NeurIPS 2024.

Thank You

Questions?

Acknowledgements:

M. Mezghanni, L. Maillard, M. Boulkenafed, …

Work supported by the ERC Starting Grant StG-2017-758800 (EXPROTEA), ERC Consolidator VEGA and the ANR AI Chair AIGRETTE.