CSC2457 Generative modelling in 3D

AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation

Date: 9th February 2021

Presenter: Varun Dharmendrakumar Pandya

Instructor: Animesh Garg

UNIVERSITY OF TORONTO

Motivation and Main Problem

Learning a representation for generating high resolution 3D shapes remains an open challenge. This paper introduces a method for learning to generate the surface of 3d shapes.

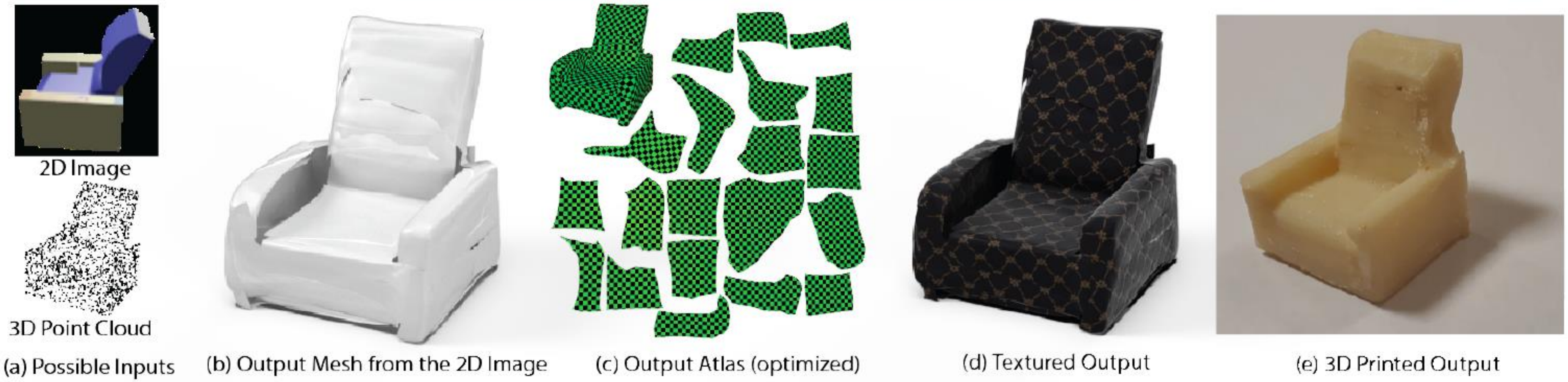
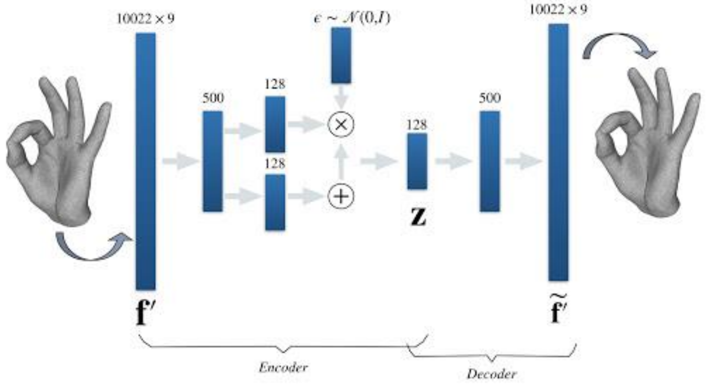


Figure 1. Given input as either a 2D image or a 3D point cloud (a), we automatically generate a corresponding 3D mesh (b) and its atlas parameterization (c). We can use the recovered mesh and atlas to apply texture to the output shape (d) as well as 3D print the results (e).

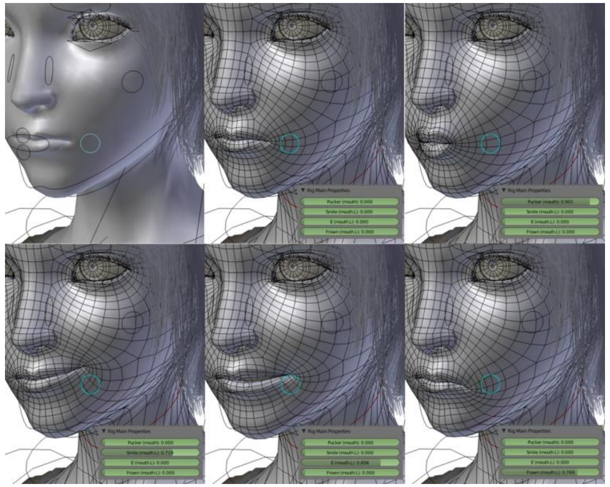
Motivation and Main Problem

<u>The Importance:</u> Forms the basis of many applications in 3D –

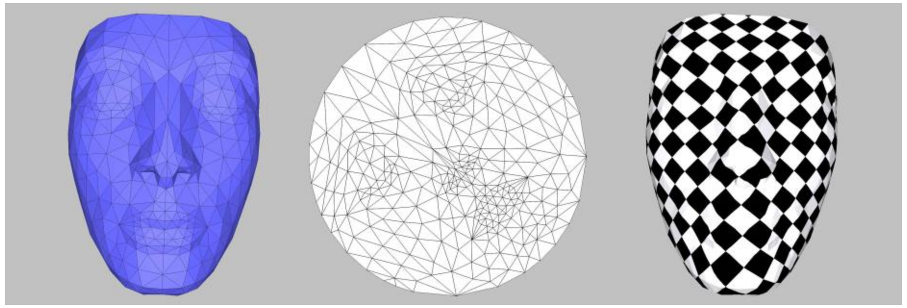


Autoencoding shapes

Single view reconstruction



Morphing



Parametrization

Motivation and Main Problem

The issues:

- memory issues with certain representations
- low resolution mesh formation
- parametrization of meshes and many more

Still an open problem?

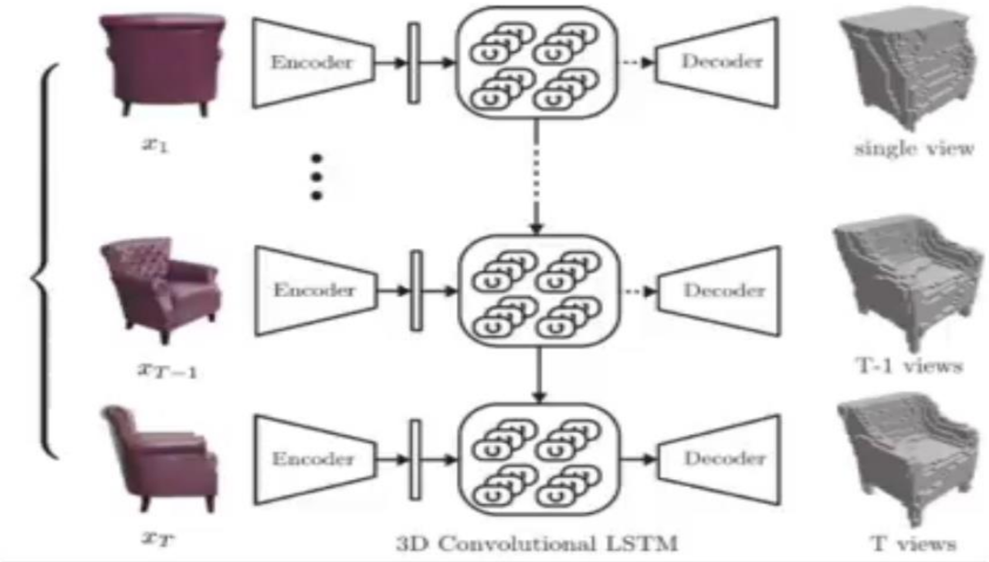
- Voxel grid based methods are memory intensive
- no surface connectivity(tesselation)
- automatically estimating correspondences from training shapes to the base meshes (gets increasingly hard for heterogeneous datasets).

Contributions

- **Problem:** Learning to generate the surface of 3D shapes
- **Importance and hardness:** Generating surface meshes with better precision for varied applications
- **Key issue of prior work:** Low precision with memory and tessellation issues
- **Key insight:** In the decoder, sample points from a 2D plane(unit square) to generate multiple patches in 3D
- **Result of this insight:** Generated surface mesh with better precision and without causing any memory issues compared to other related work

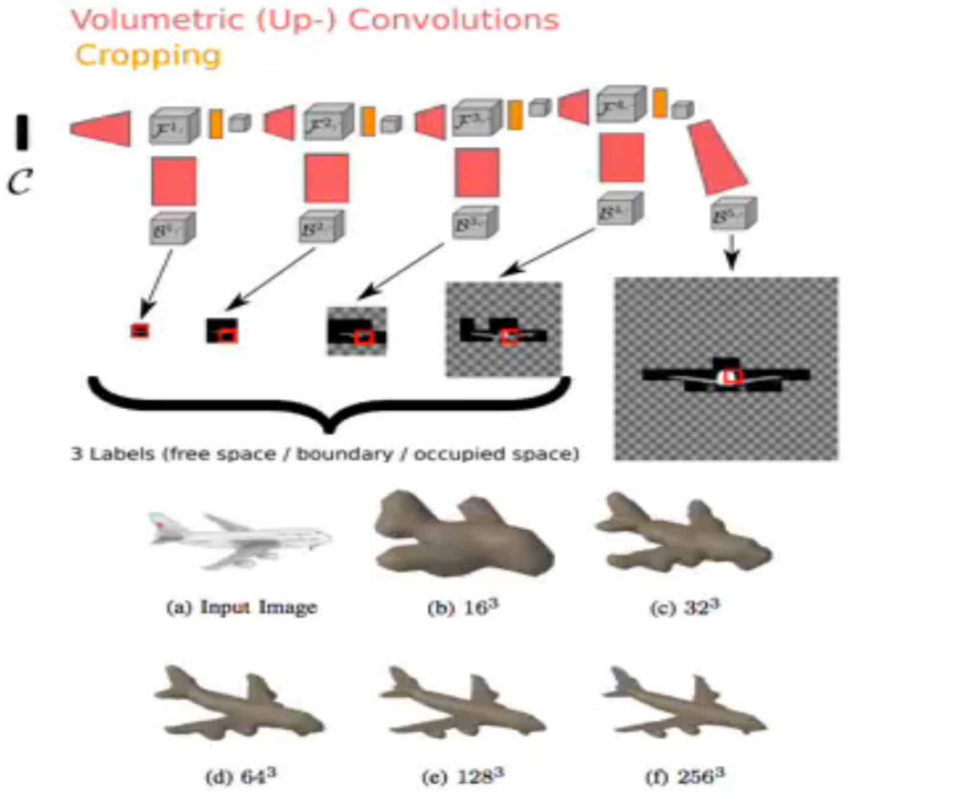
General Background

Generating voxels



Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S.
3D-R2N2: A unified approach for single and multi-view 3D object reconstruction, ECCV 2016

Natural, conceptually simple

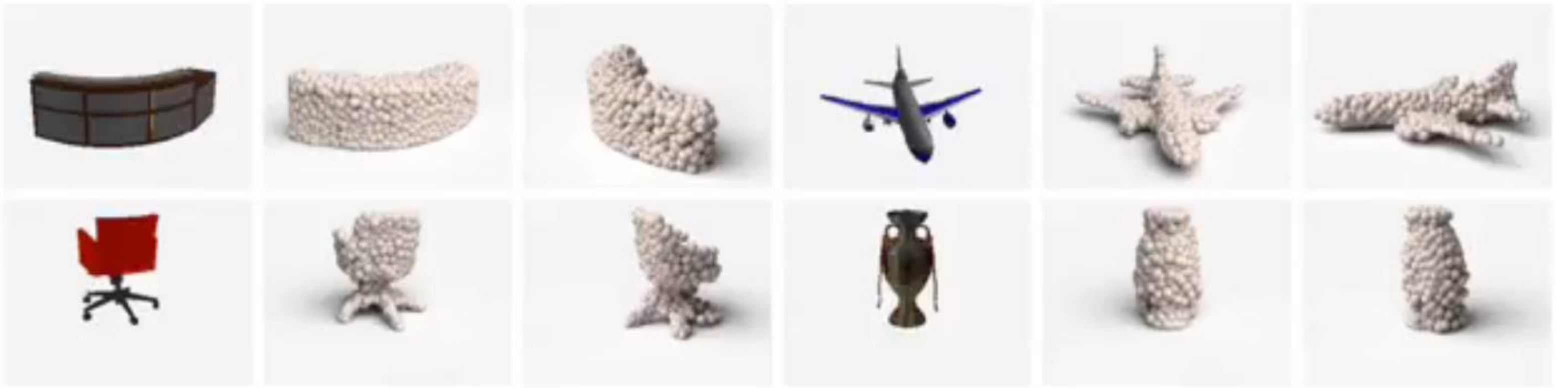


Häne, C., Tulsiani, S., & Malik, J.
Hierarchical surface prediction for 3D object reconstruction. 3DV 2017

Heavy or complicated

General Background

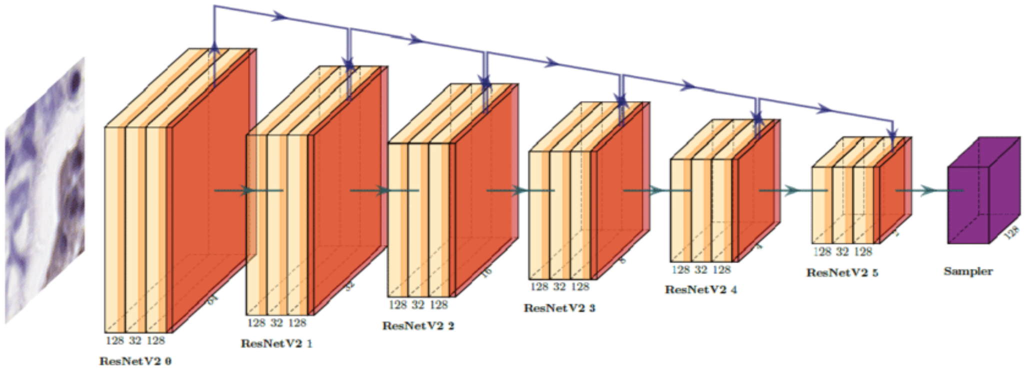
Generating points : PointSetGen

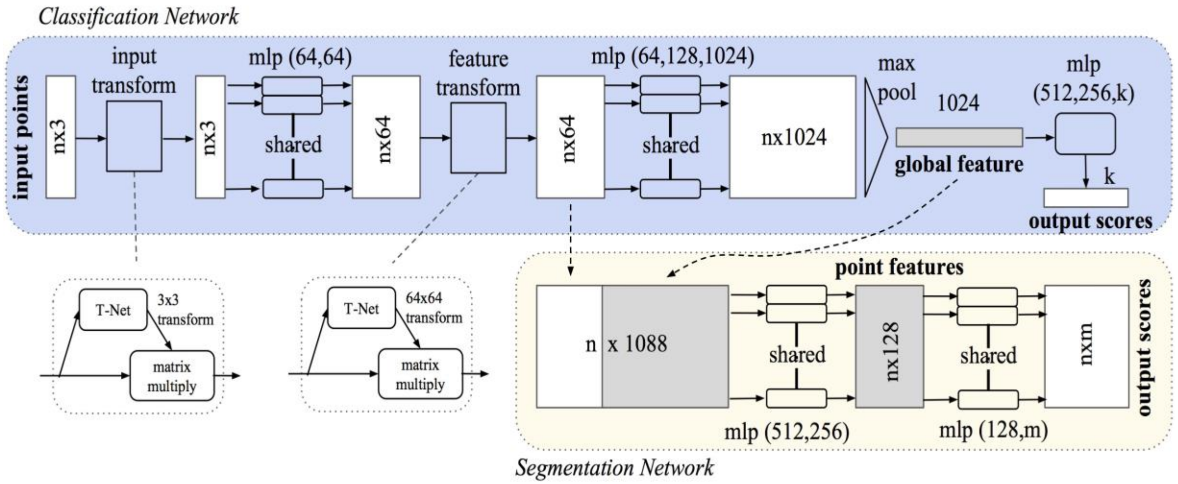


Fan, H., Su, H., & Guibas, L. A point set generation network for 3d object reconstruction from a single image. CVPR 2017

- Simple
- Unstructured point cloud

General Background





- **<u>Resnet encoder(for 2D image):</u>** Type of CNN made up of residual blocks with skip connections.

- **<u>PointNet encoders:</u>** Different architectures are available (usually contain a series of multi-layer perceptrons which are finally max or average pooled into one large vector)

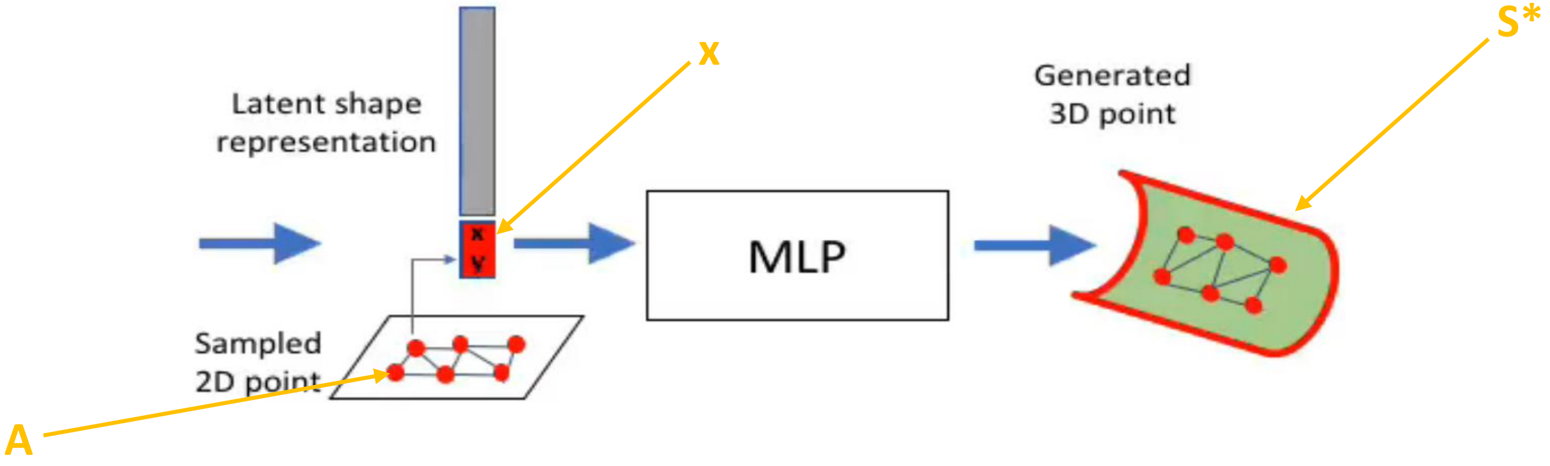
General Background

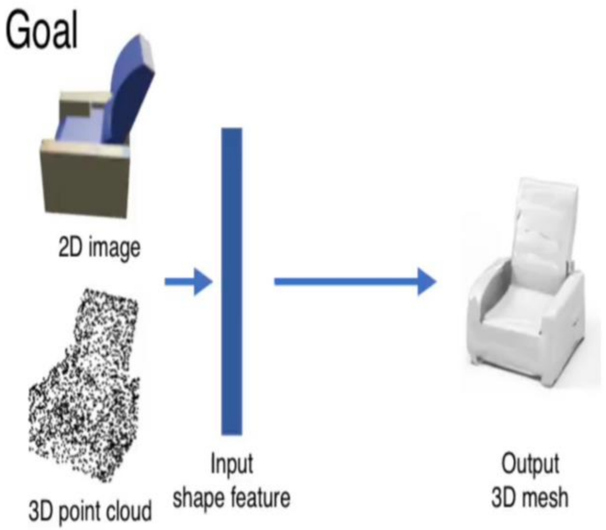
Chamfer distance We define the Chamfer distance between $S_1, S_2 \subseteq \mathbb{R}^3$ as:

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2$$

Problem Setting

- Here, an MLP with ReLUs $\boldsymbol{\varphi}_\theta$ with parameters θ can locally generate a surface by learning to map points in R^2 to surface points in R^3. To generate a given surface, we use several of these.
- Let A be a set of points sampled in the unit square $]0,1[^2$ and S^* a set of points sampled on the target surface. Next, we incorporate the shape feature x (latent vector) by simply concatenating them with the sampled point coordinates p ∈ A before passing them as input to the MLPs.

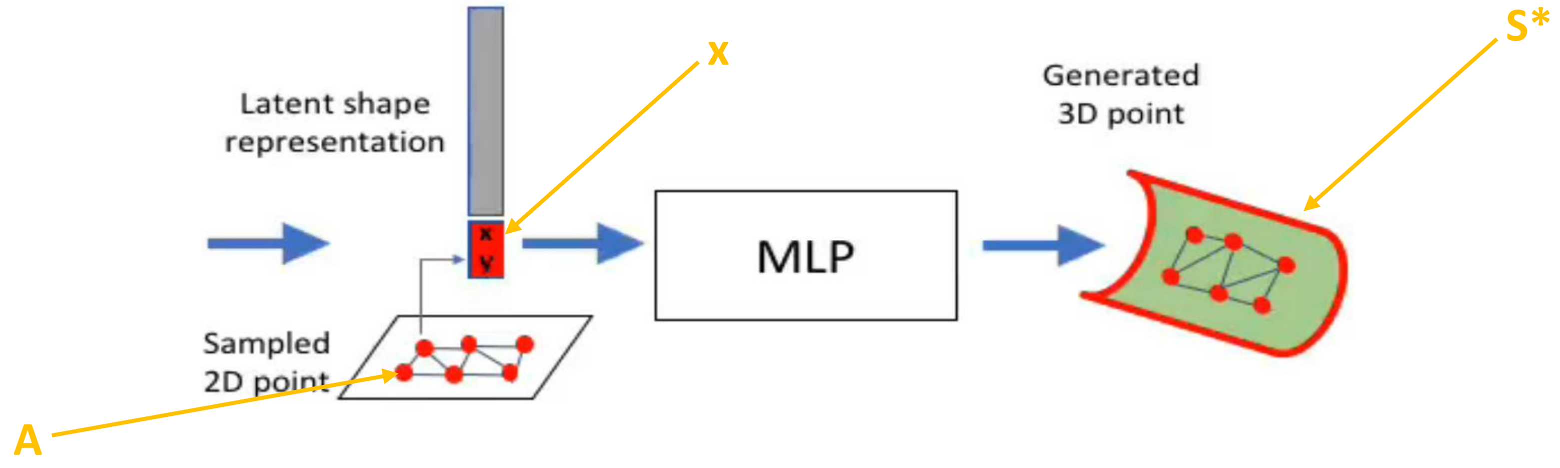




Problem Setting

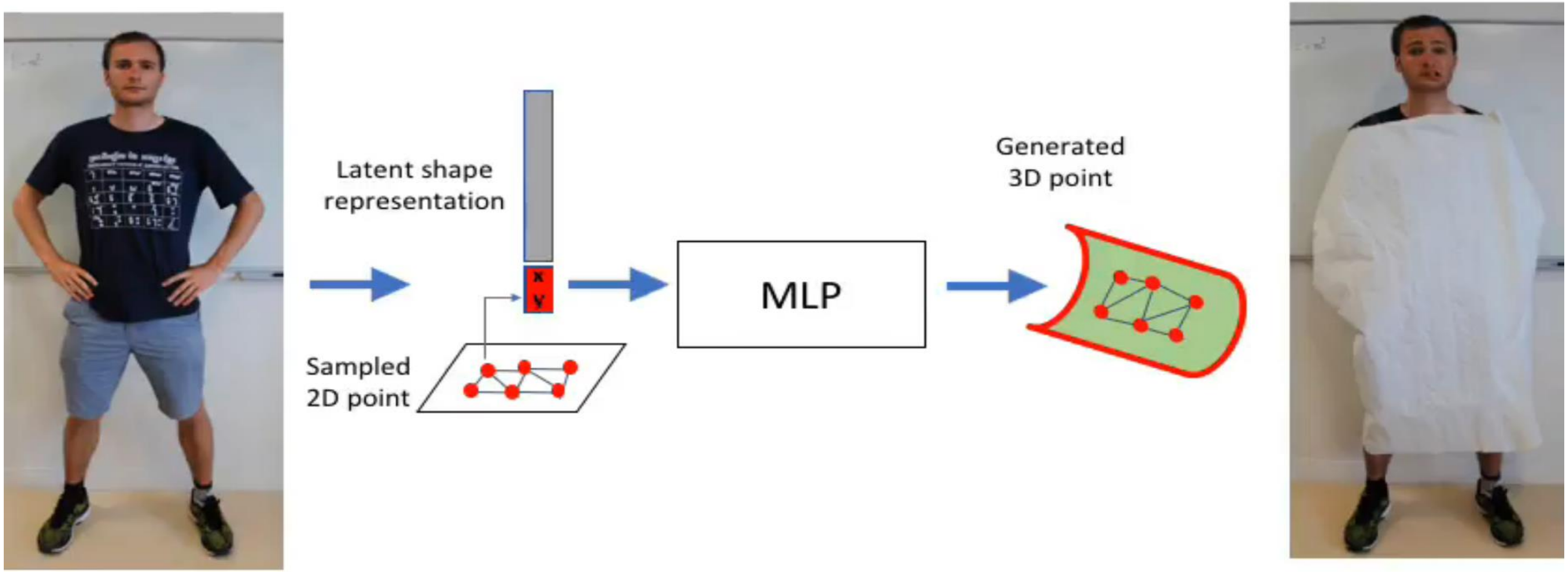
- We then minimize the Chamfer loss between the set of generated 3D points and S^*, **thus optimizing the loss based on the Chamfer distance metric**

$$L(\theta) = \sum_{p \in A} \sum_{i=1}^{N} \min_{q \in S^*} \left| \varphi_{\theta_i}(p; x) - q \right|^2 + \sum_{q \in S^*} \min_{i \in \{1,\ldots,N\}} \min_{p \in A} \left| \varphi_{\theta_i}(p; x) - q \right|^2$$



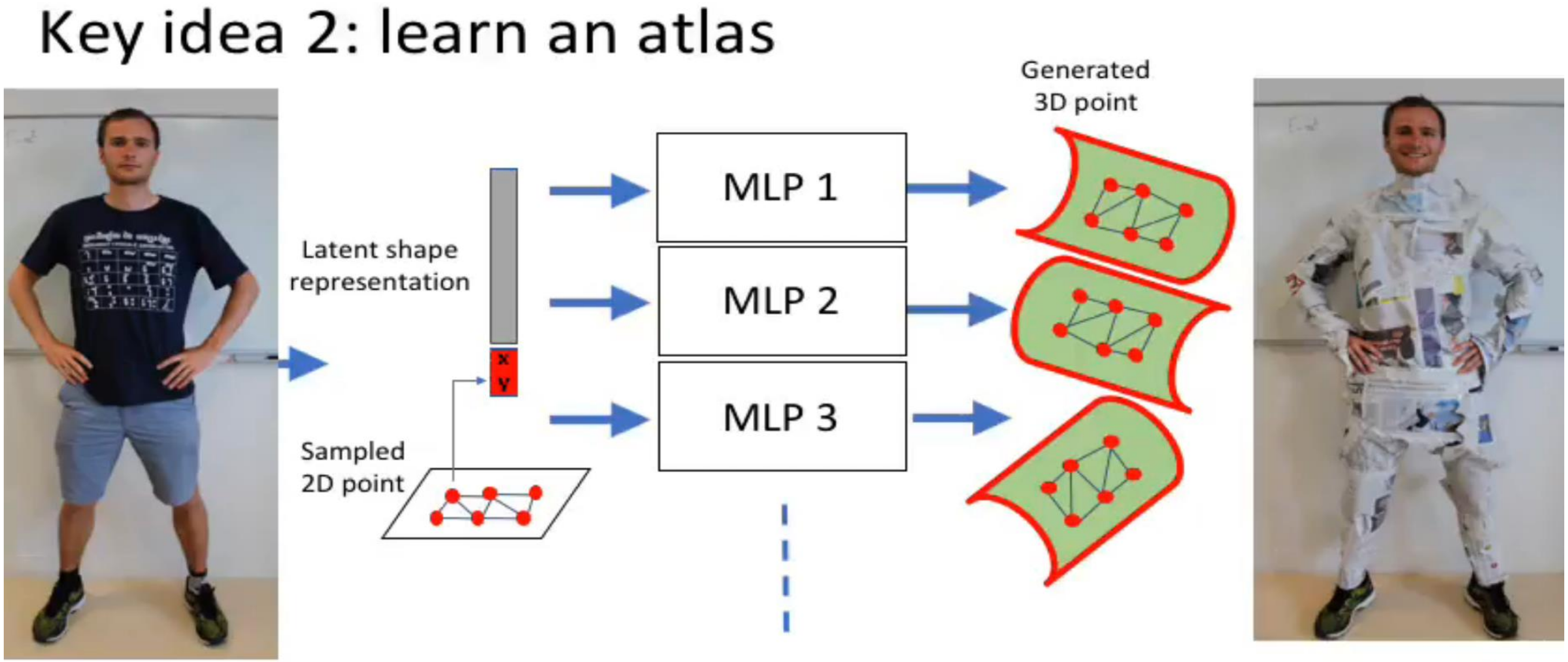
Approach

Key idea 1: deform a surface



Learnt simply by sampling many points and minimizing Chamfer distance

Approach

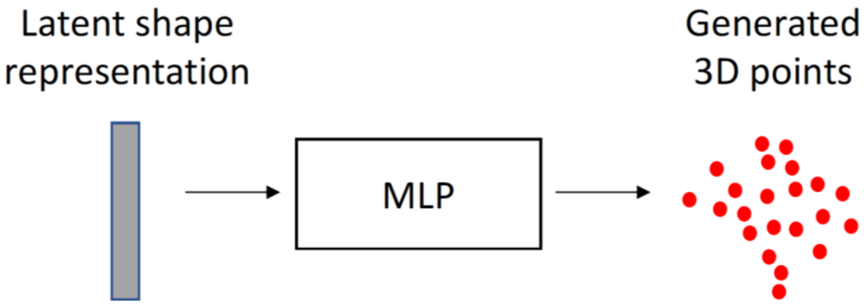


Learnt simply by sampling many points and minimizing Chamfer distance

Experimental Results

The proposed approach was evaluated on the standard ShapeNet Core dataset.

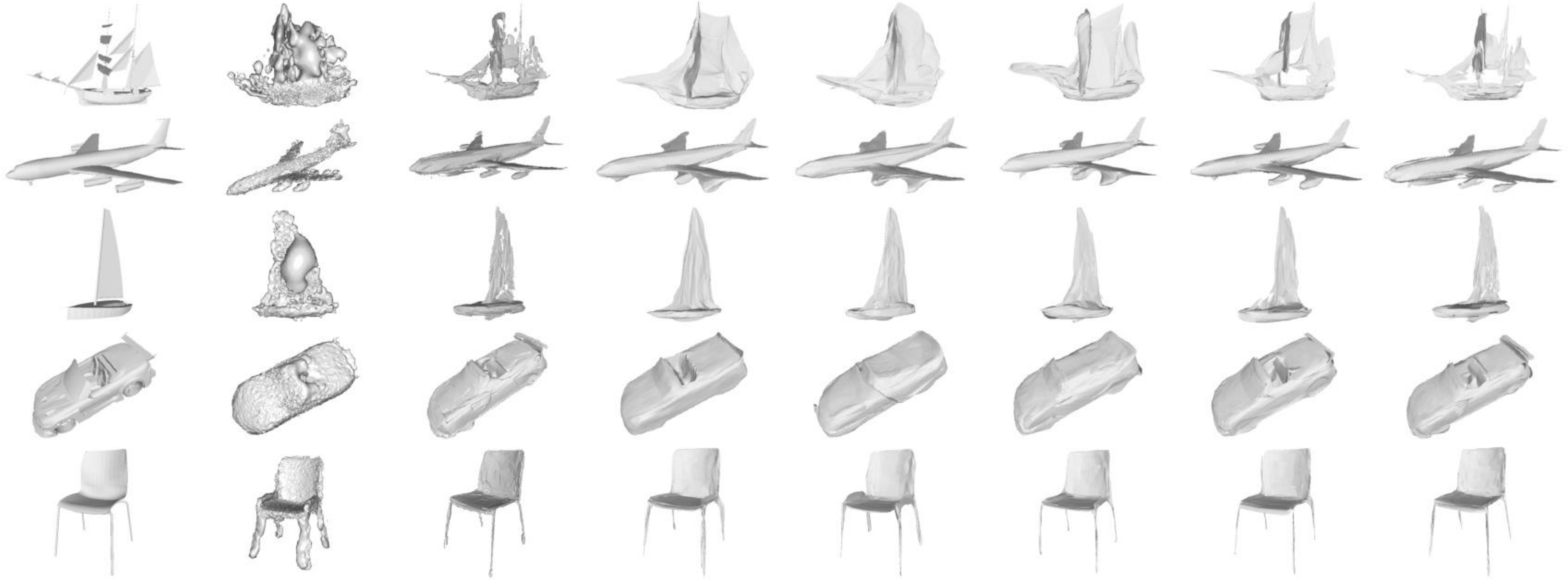
Chamfer distance and Metro criteria(to compare the output meshes with the ground truth for mesh connectivity) used. The proposed method outperforms the Points baseline method.



(a) Points baseline.

Method	CD	Metro
Oracle 2500 pts	0.85	1.56
Oracle 125K pts	-	1.26
Points baseline	1.91	-
Points baseline + normals	2.15	1.82 (PSR)
Ours - 1 patch	1.84	1.53
Ours - 1 sphere	1.72	1.52
Ours - 5 patches	1.57	1.48
Ours - 25 patches	1.56	1.47
Ours - 125 patches	**1.51**	**1.41**

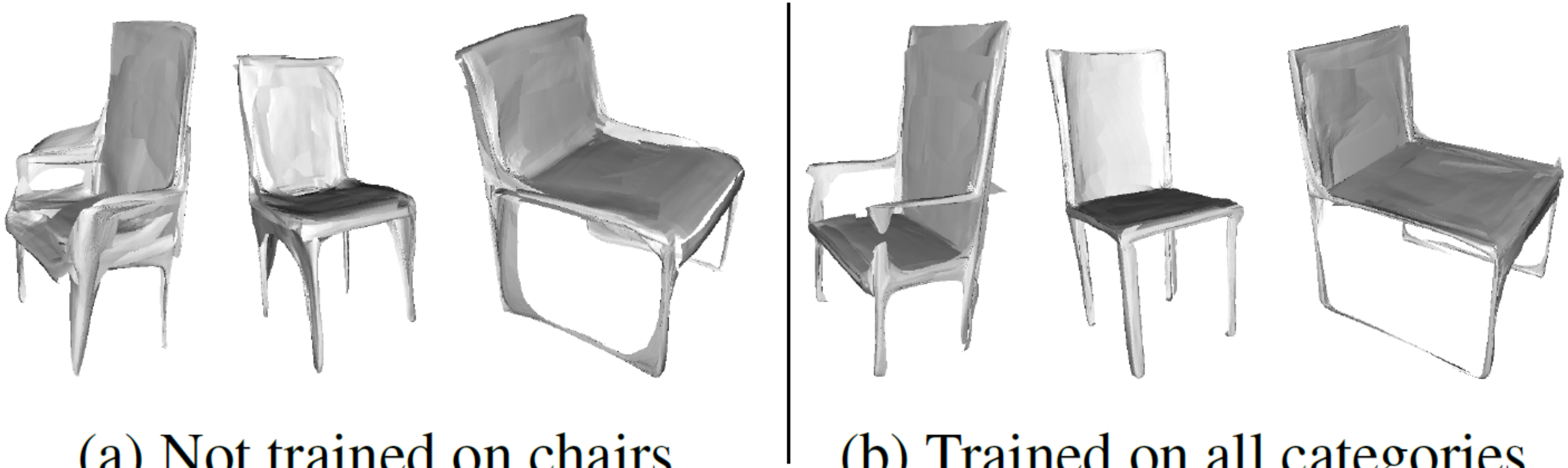
Experimental Results



(a) Ground truth (b) Pts baseline (c) PSR on ours (d) Ours sphere (e) Ours 1 (f) Ours 5 (g) Ours 25 (h) Ours 125

Figure 3. **Auto-encoder.** We compare the original meshes (a) to meshes obtained by running PSR on the point clouds generated by the baseline (b) and on the densely sampled point cloud from our generated mesh (c), and to our method generating a surface from a sphere (d), 1 (e), 5 (f), 25 (g), and 125(h) learnable parameterizations. Notice the fine details in (g) and (h) : e.g. the plane's engine and the jib of the ship.

Experimental Results



(a) Not trained on chairs | (b) Trained on all categories

Figure 4. **Generalization.** (a) Our method (25 patches) can generate surfaces close to a category never seen during training. It, however, has more artifacts than if it has seen the category during training (b), e.g., thin legs and armrests.

Experimental Results

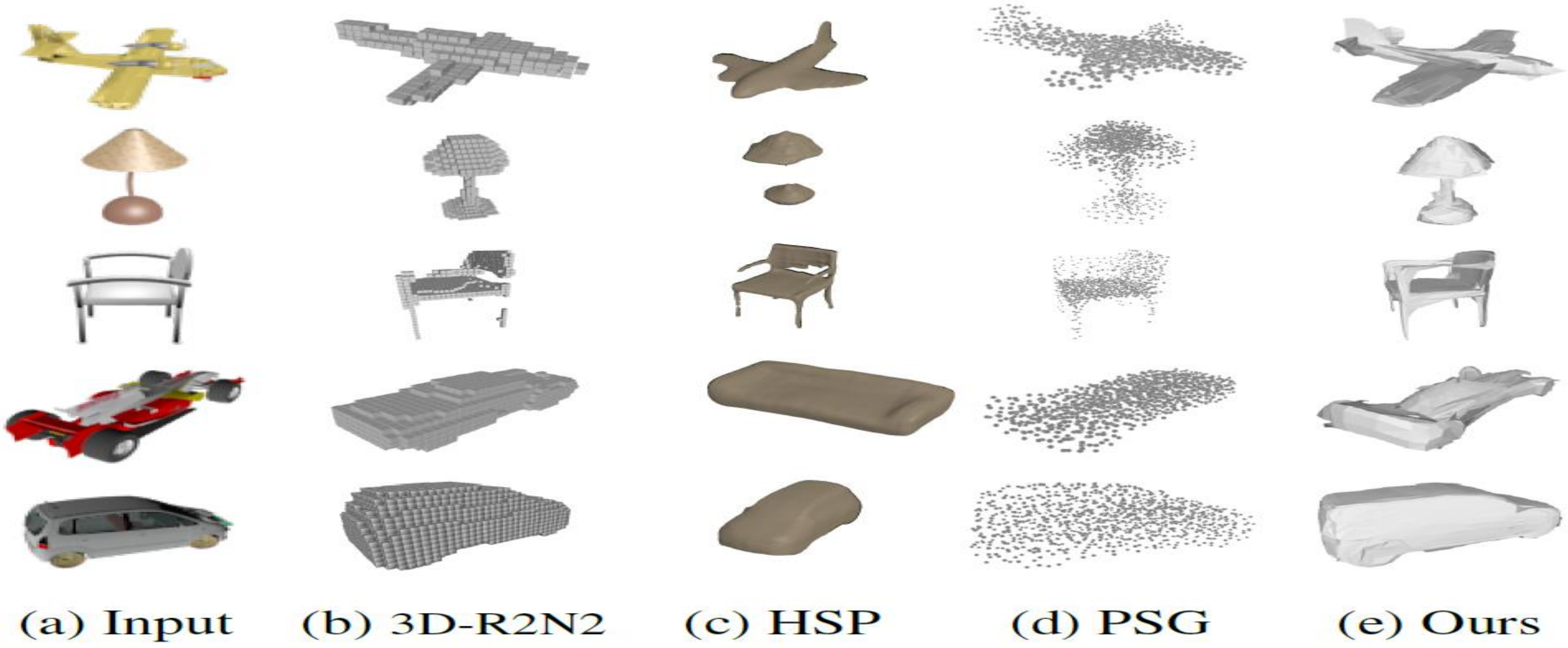
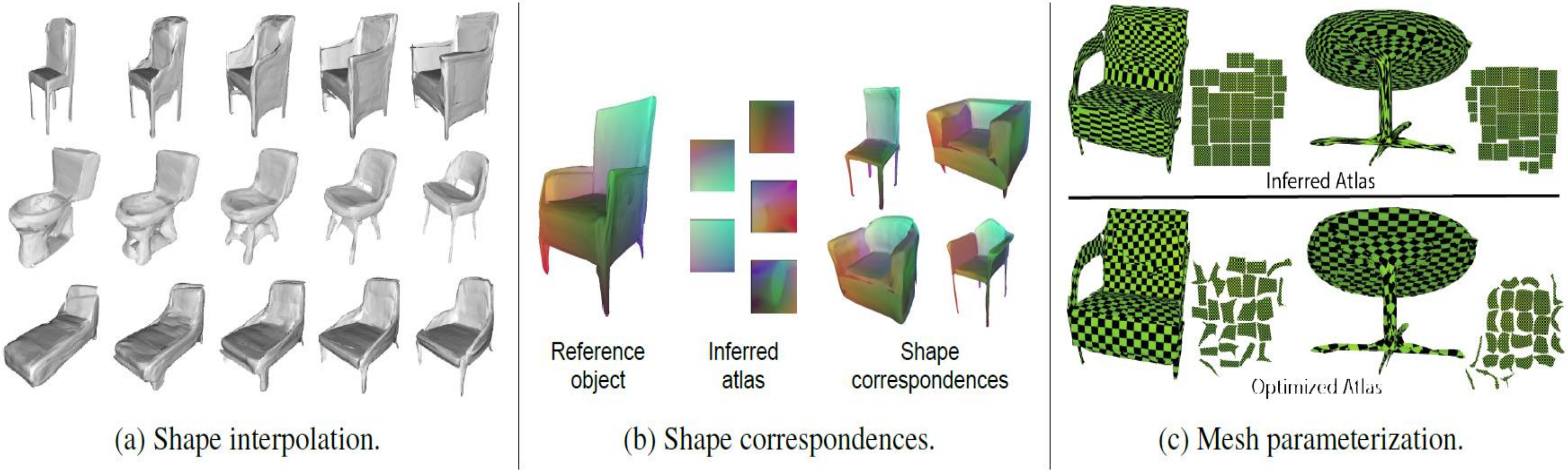


Figure 5. Single-view reconstruction comparison. From a 2D RGB image (a), 3D-R2N2 reconstructs a voxel-based 3D model (b), HSP reconstructs a octree-based 3D model (c), PointSet- Gen a point cloud based 3D model (d), and our AtlasNet a triangular mesh (e).

Experimental Results



(a) Shape interpolation. (b) Shape correspondences. (c) Mesh parameterization.

Figure 7. **Applications.** Results from three applications of our method.

Experimental Results

(a) Low-Res Input (b) High-Res reconstruction

Figure 8. **Super resolution.** Our approach can generate meshes at arbitrary resolutions, and the pointnet encoder [25] can take pointclouds of varying resolution as input. Given the same shape sampled at the training resolution of 2500, or 10 times less points, we generate high resolution meshes with 122500 vertices. This can be viewed as the 3D equivalent of super-resolution on 2D pixels.

Discussion of results

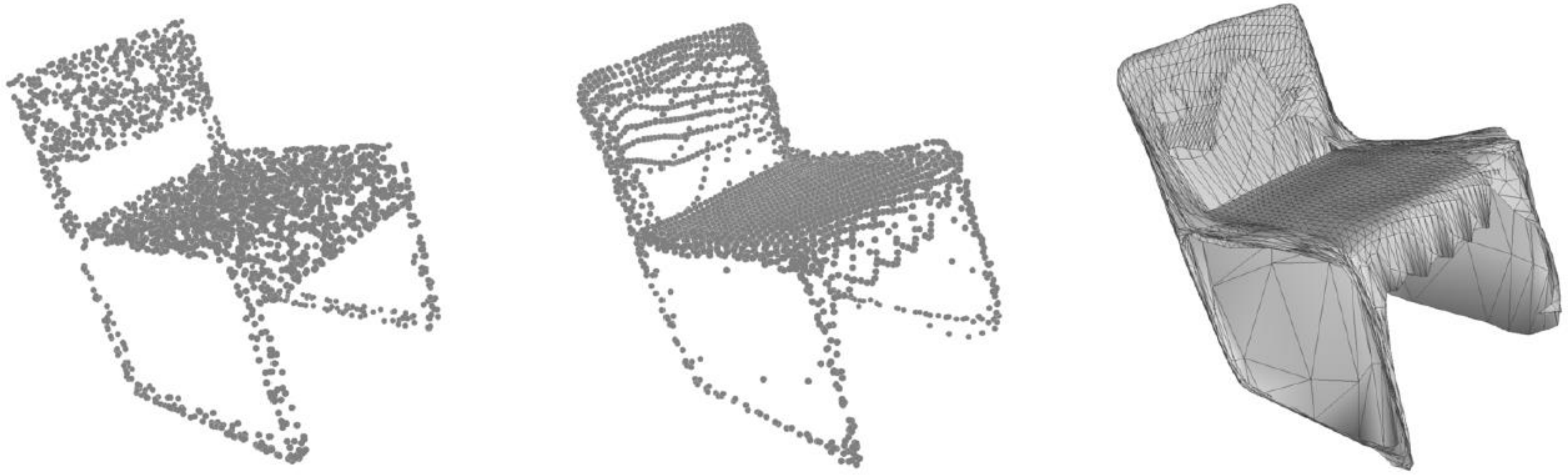
The following conclusions can be drawn from the results:

1. Not as memory intensive as other approaches like voxel grids
2. This general purpose approach finds applications in various other domains like shape interpolation, shape correspondence etc. that find their bases in auto-encoding and representation of surfaces.
3. The resolution and precision of the mesh can be easily controlled by adjusting the number of patches used in the decoder.
4. The proposed approach has good generalization capabilities (for unseen 3D shapes)

The given conclusions are fully supported by the results, as they clearly indicate both the qualitative (precision and resolution of meshes visually) and quantitative (comparing the Chamfer distance metric and Metro metric) measures.

Critique / Limitations / Open Issues

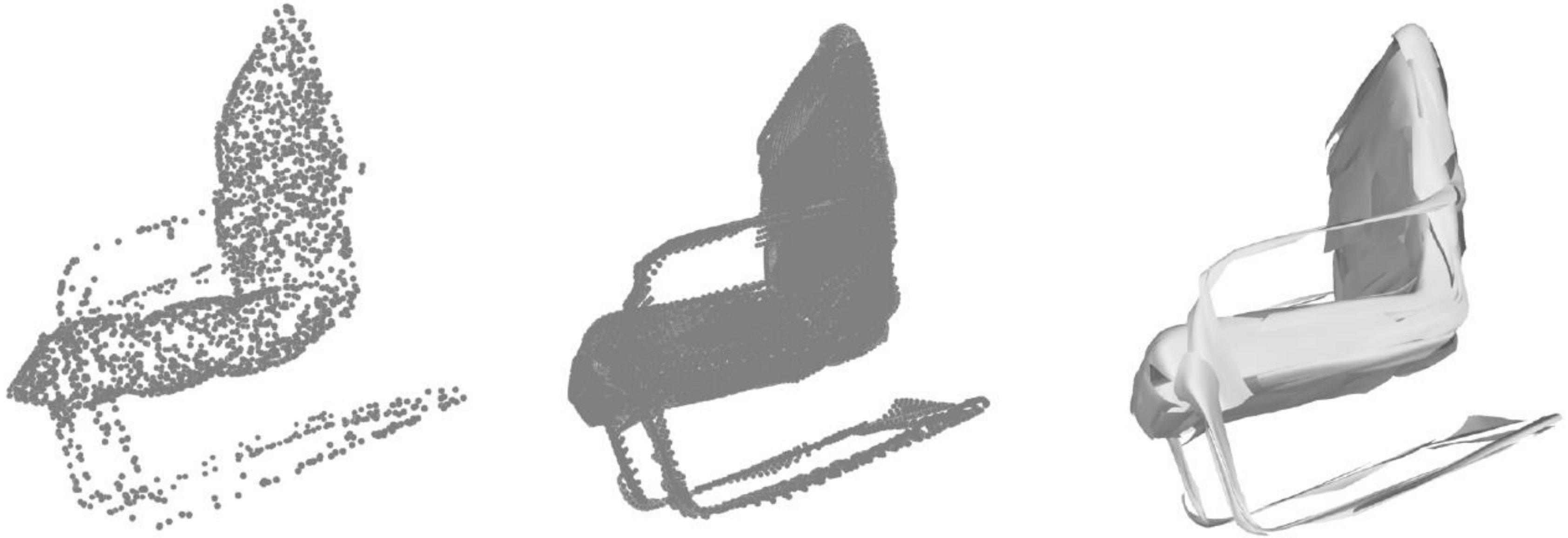
1. When a small number of learned parameterizations are used, the network has to distort them too much to recreate the object. This leads, when we try to recreate a mesh, to small triangles in the learned parameterization space being distorted and become large triangles in 3D covering undesired regions.



(a) **Excess of distortion.** Notice how, compared to the original point cloud (left), the generated pointcloud (middle) with 1 learned parameterization is valid, but the mapping from squares to surfaces enforces too much distortion leading to error when propagating the grid edges in 3D (right).

Critique / Limitations / Open Issues

2. As the number of learned parameterization increases, errors in the topology of the reconstructed mesh can be sometimes observed. In practice, it means that the reconstructed patches overlap, or are not stitched together.



(b) **Topological issues.** Notice how, compared to the original point cloud (left), the generated pointcloud (middle) with 125 learned parameterizations is valid, but the 125 generated surfaces overlap and are not stiched together (right).

Contributions (Recap)

- **Problem:** Learning to generate the surface of 3D shapes
- **Importance and hardness:** Generating surface meshes with better precision for varied applications
- **Key issue of prior work:** Low precision with memory and tessellation issues
- **Key insight:** In the decoder, sample points from a 2D plane(unit square) to generate multiple patches in 3D
- **Result of this insight:** Generated surface mesh with better better precision and without causing any memory issues compared to other related work