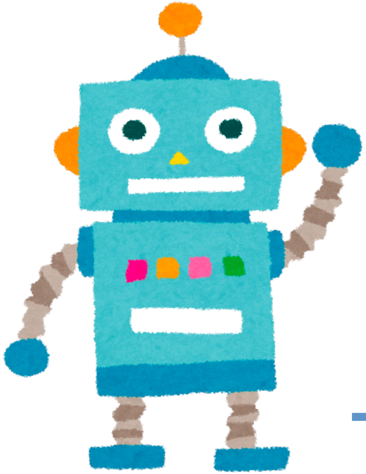
Exploring Generative 3D Shapes
Using Autoencoder Networks

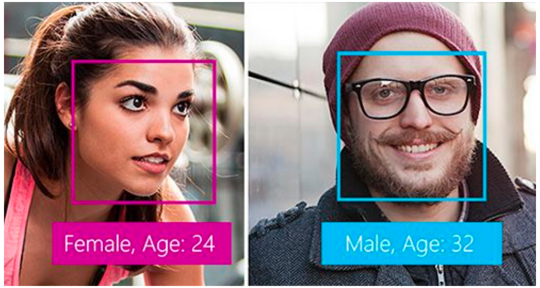
Nobuyuki Umetani

AUTODESK RESEARCH

Motivation

- Machine learning is booming





https://docs.microsoft.com/en-us/azure/cognitive-services/face/overview

Face recognition

http://theoatmeal.com/blog/google_self_driving_car

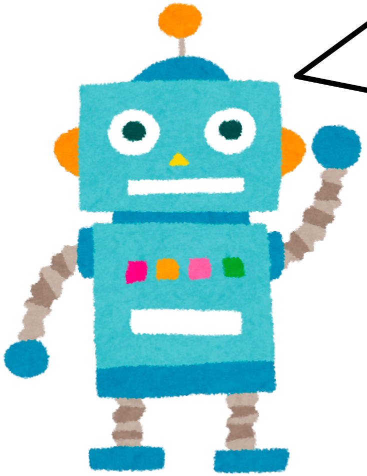
Self-driving car

https://www.engadget.com/2016/03/13/google-alphago-loses-to-human-in-one-match/

Alpha Go

Motivation

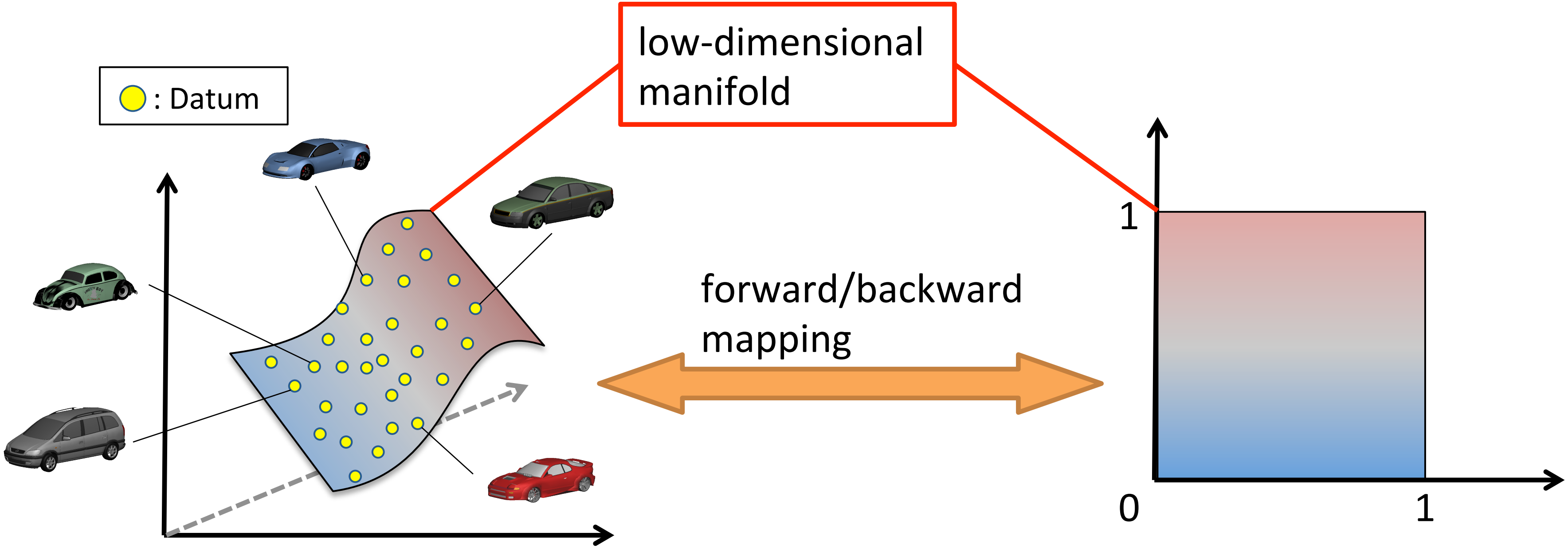
- Machine learning for 3D shape
  - Understanding a class of shapes



In my understanding, shapes of cars look like..

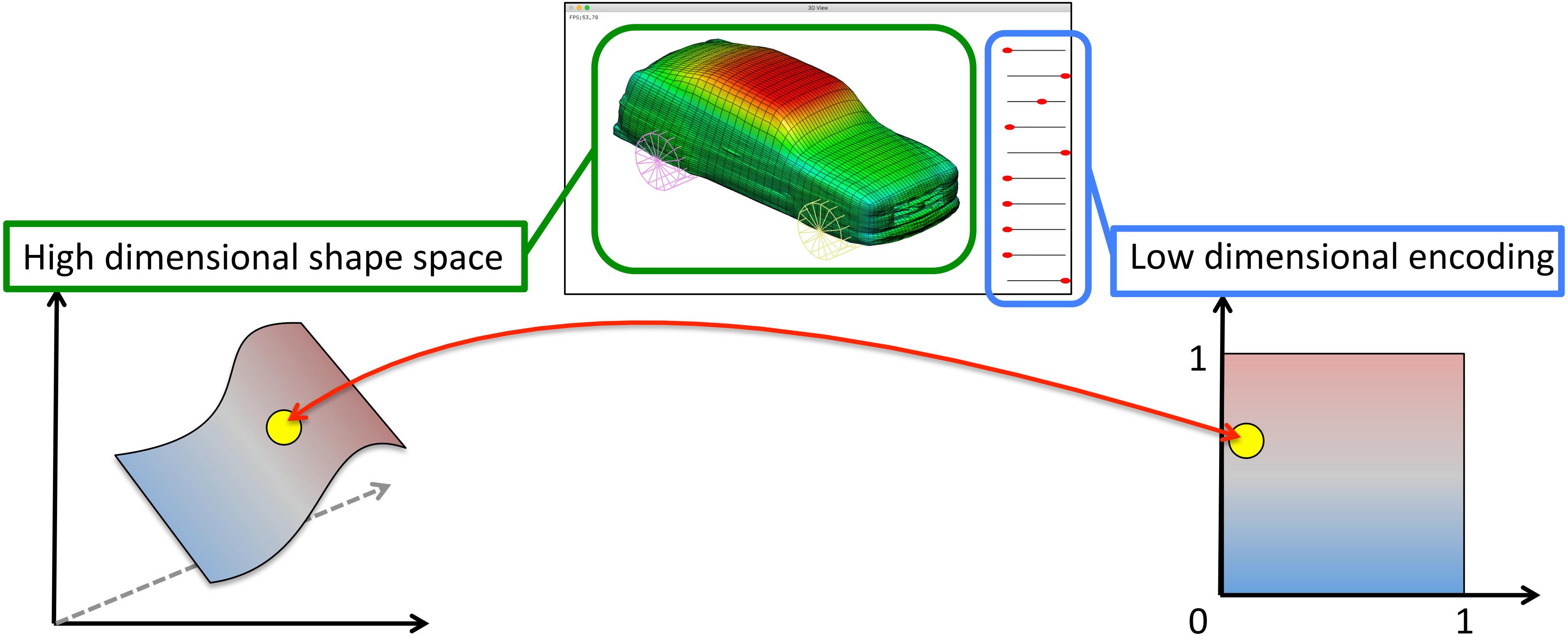
Motivation

- Machine learning for 3D shapes
  - Finding low dimensional manifold in space

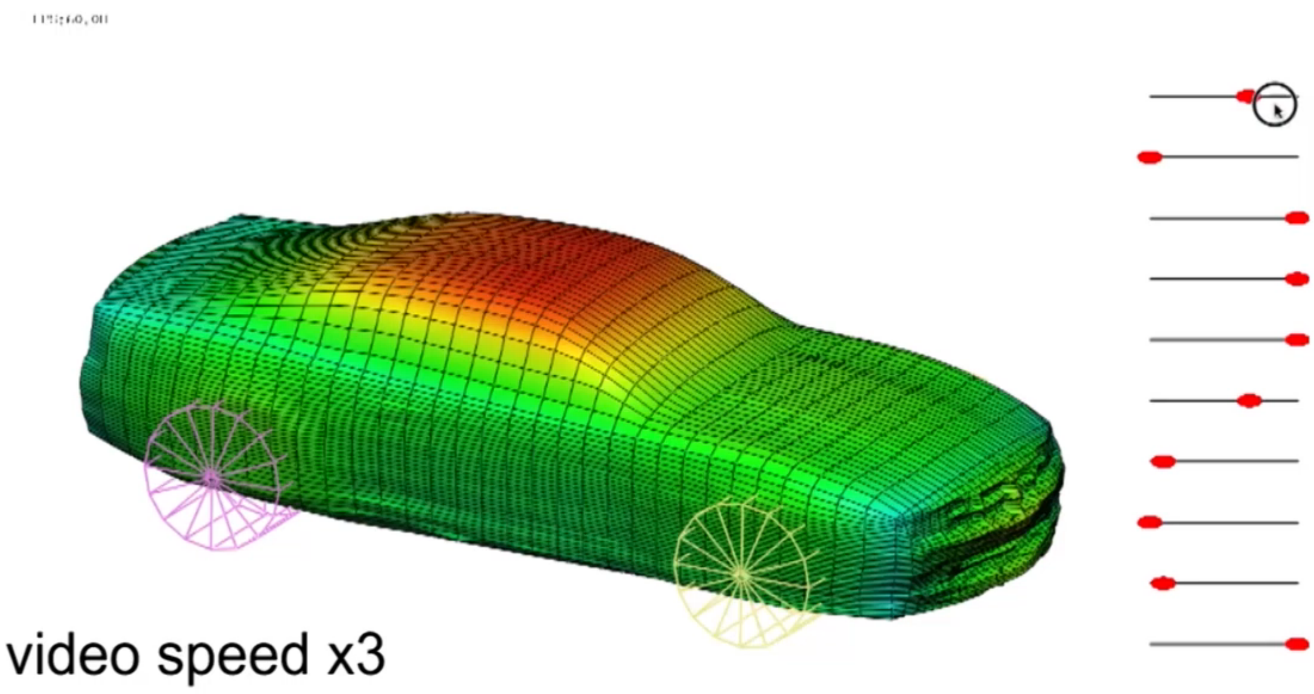


Synthesizing New Shapes

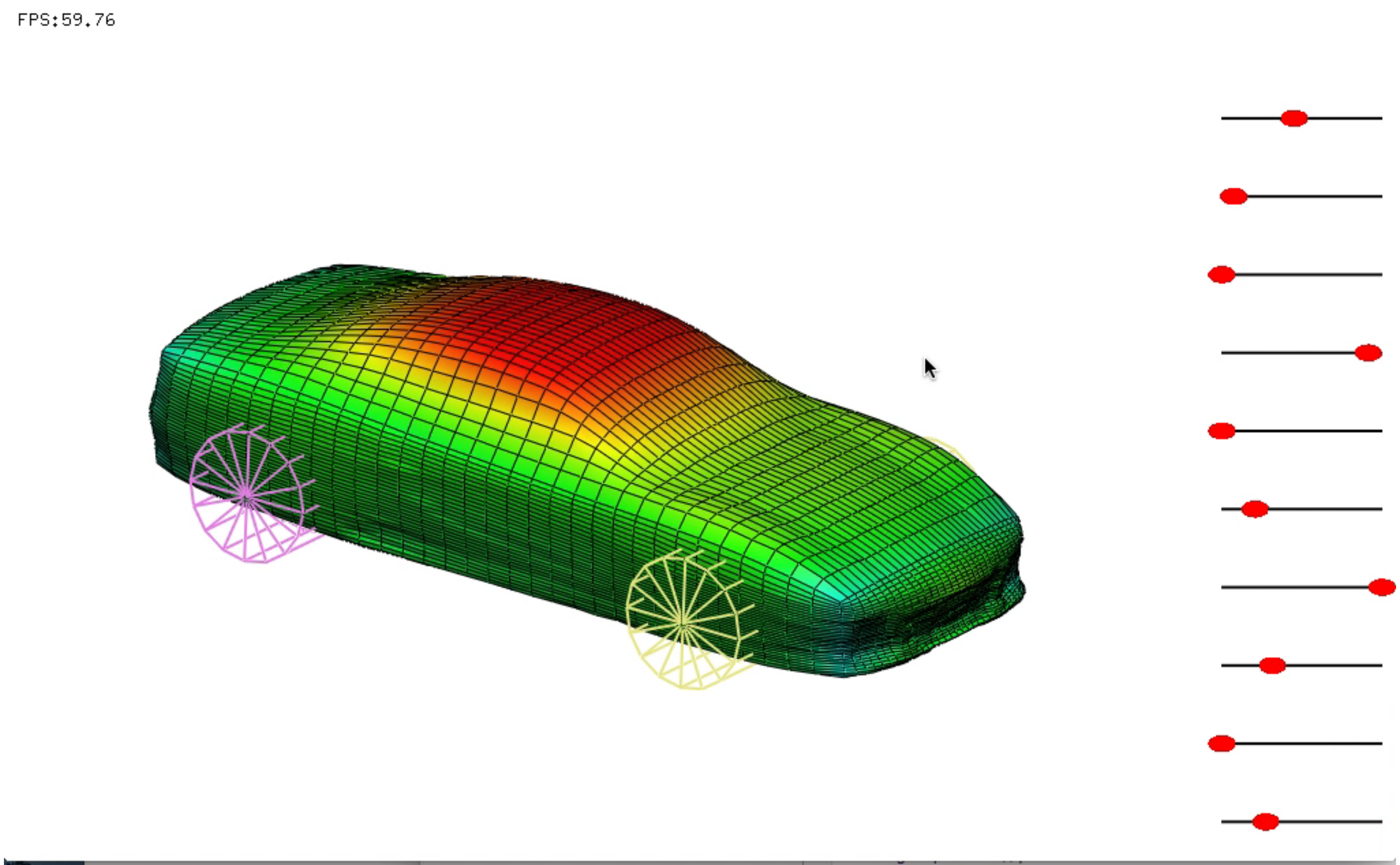
- Mapping low dimensional manifold to the shapes



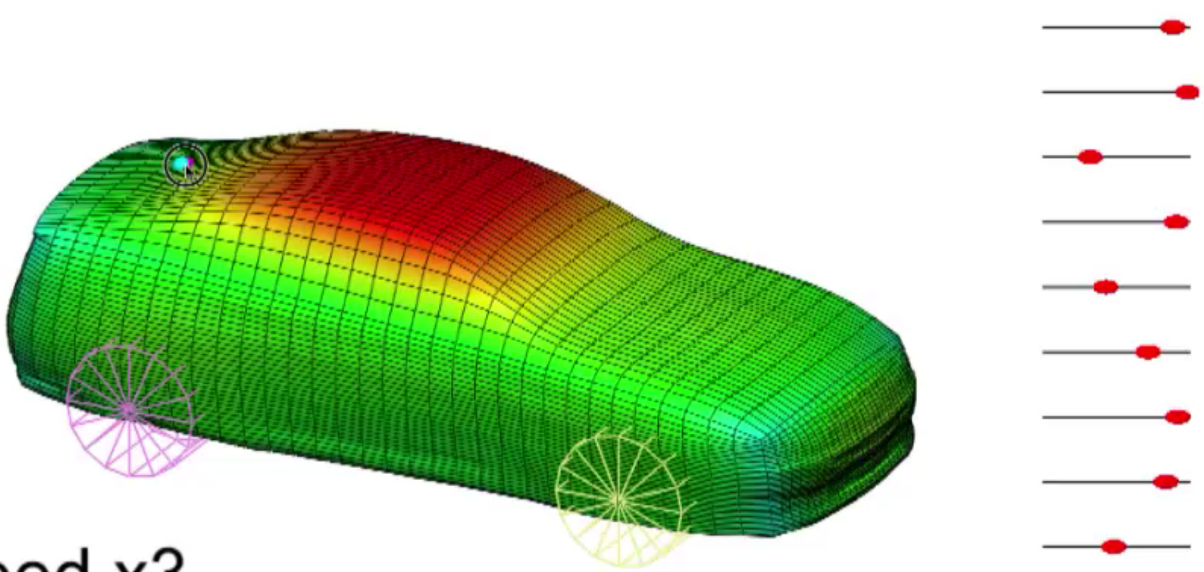
Preview of the Result: Slider bar



Preview of the Result: Random



Interactive Exploration

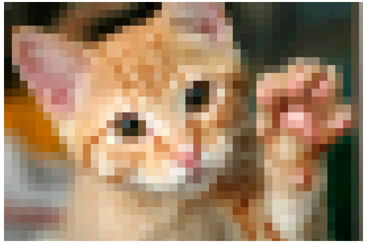


video speed x3

Parameterization problem

- Shape need to be represented by fixed dimensional vector/tensor

2D Image



$$\begin{pmatrix} 53.3 \\ \vdots \\ 236.1 \\ \vdots \\ 67.2 \end{pmatrix}$$

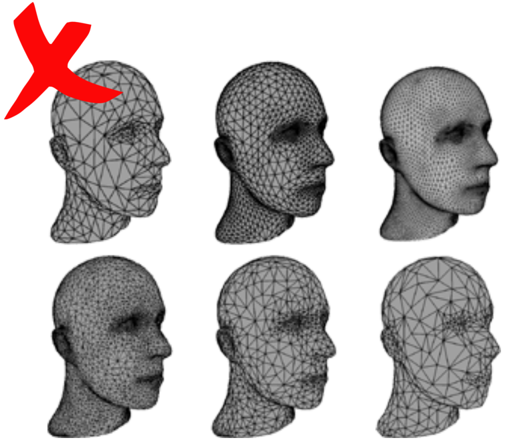
3D shape

?

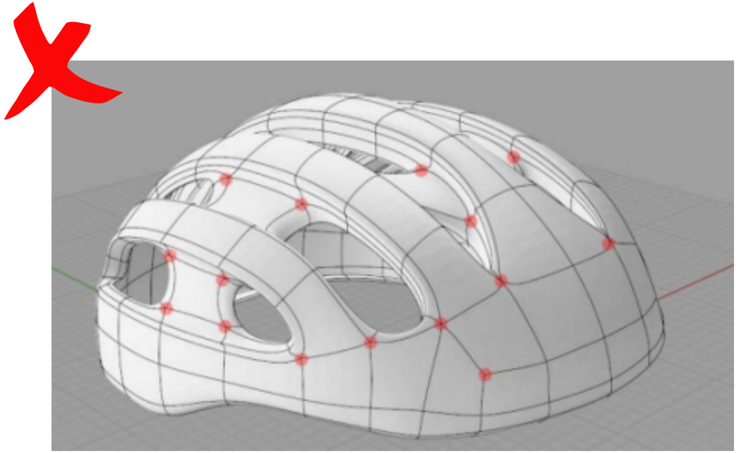
$$\begin{pmatrix} 53.3 \\ \vdots \\ 236.1 \\ \vdots \\ 67.2 \end{pmatrix}$$

Parameterization problem

- Triangle mesh / NURBS are not suitable for ML
  - Topology / #points are not constant



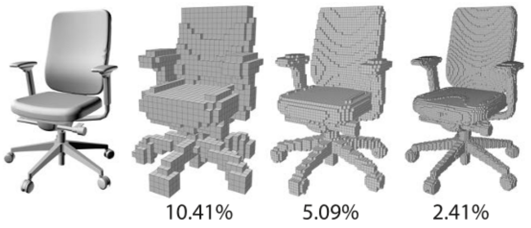
Triangle mesh



NURBS

Related Work: Voxel model

- Difficult to synthesize detailed 3D shape
  - Expensive memory cost, noise…



[FPNN, Li et al., 2017]

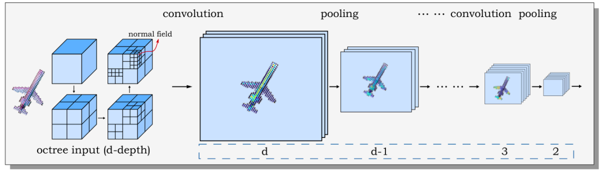
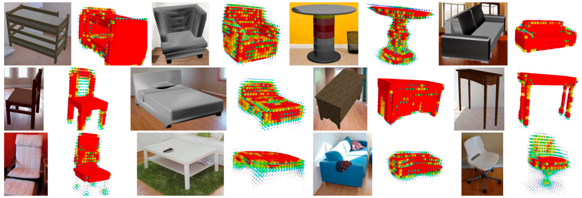
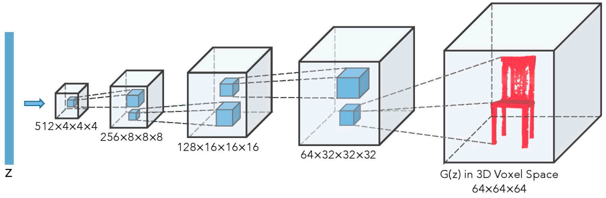


Fig. 1. An illustration of our octree-based convolutional neural network (O-CNN). Our method represents the input shape with an octree and feeds the averaged normal vectors stored in the finest leaf octants to the CNN as input. All the CNN operations are efficiently executed on the GPU and the resulting features are stored in the octree structure. Numbers inside the blue dashed square denote the depth of the octants involved in computation.

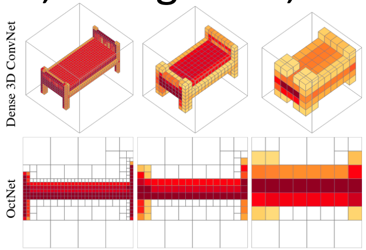
[O-CNN, Wang et al., 2017]



[Girdhar et al., 2016]



[3D GAN, 2016]



(a) Layer 1: 32^3 (b) Layer 2: 16^3 (c) Layer 3: 8^3

[OctNet, Riegler et al., 2017]

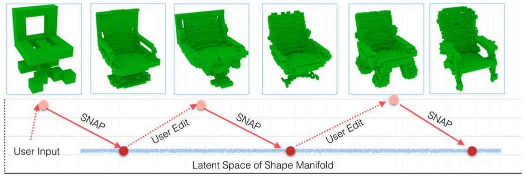
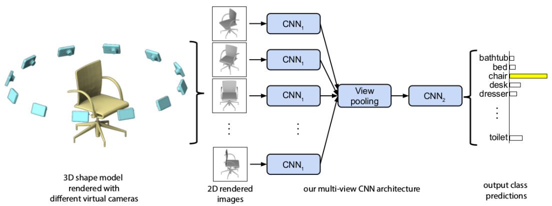


Fig. 2. A typical editing sequence. The user alternates between painting voxels (dotted arrows) and executing SNAP commands (solid arrows). For each SNAP, the system projects the current shape into a shape manifold learned with a GAN (depicted in blue) and synthesizes a new shape with a generator network.

[liu et al., 2017]

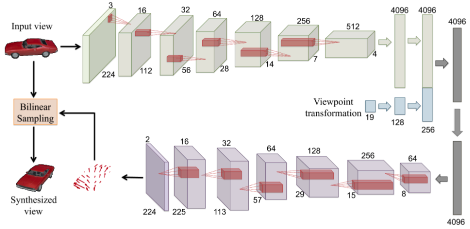
Related Work: Multi-view model

- Difficult to synthesize 3D data

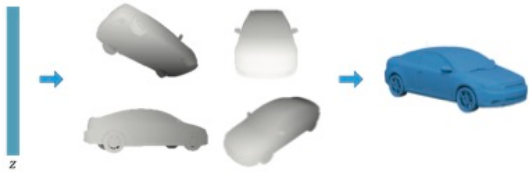


[Su et al., 2017]

[Tulsiani et al., 2017]

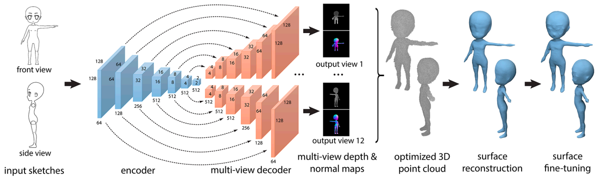


[Zhou et al., 2016]



[Soltani et al. 2017]

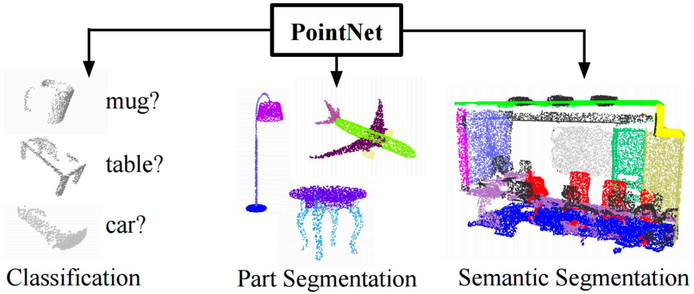
[Park et al., 2017]



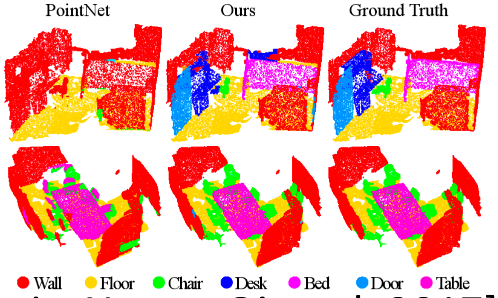
[Lun et al. 2017]

Related Work: Point-based model

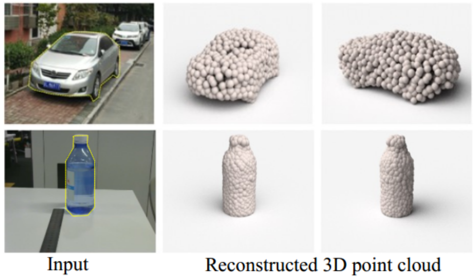
- Difficult to synthesize detailed 3D shape
  - Hard to define surface



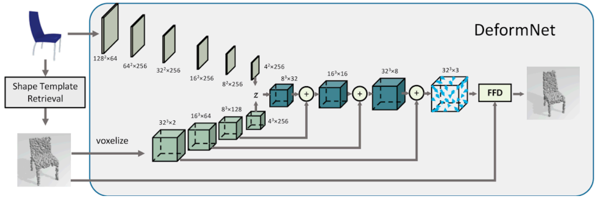
[PointNet, Qi et al, 2017]



[PointNet++, Qi et al, 2017]



[Fan et al, 2017]

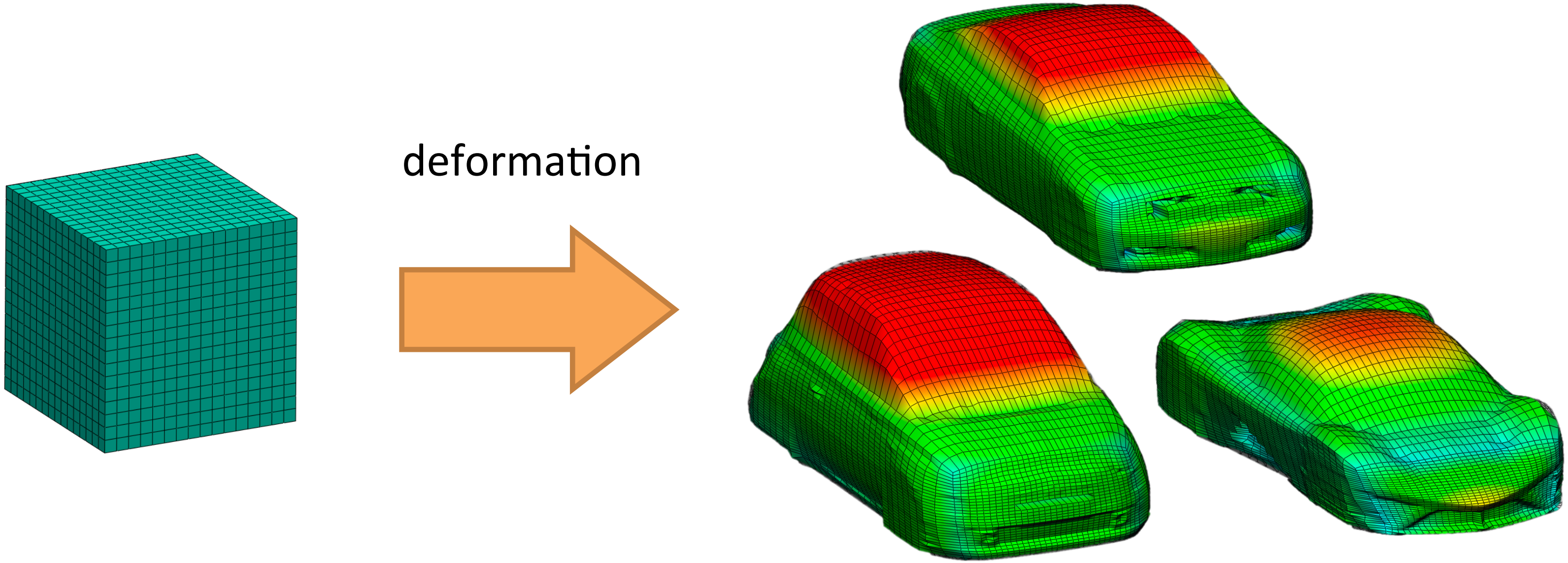


[DeformNet, Kurenkov et.al, 2017]

[Gadelha et al.,2017

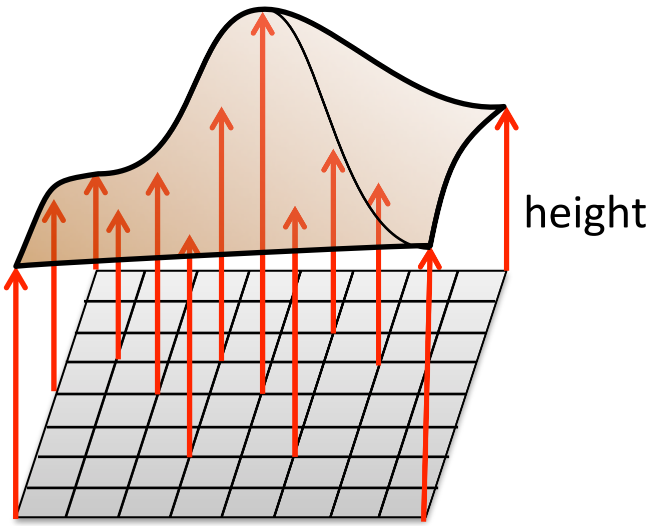
Our Approach: Mesh Representation

- Quad mesh with constant topology
- Deforming a template mesh into input shape

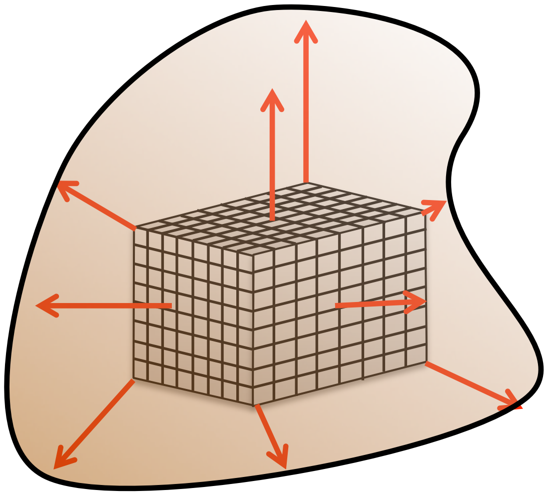


Our Approach: 3D Shape as Height Field

- Storing XYZ coordinates is redundant
- Height field from a cube in its normal



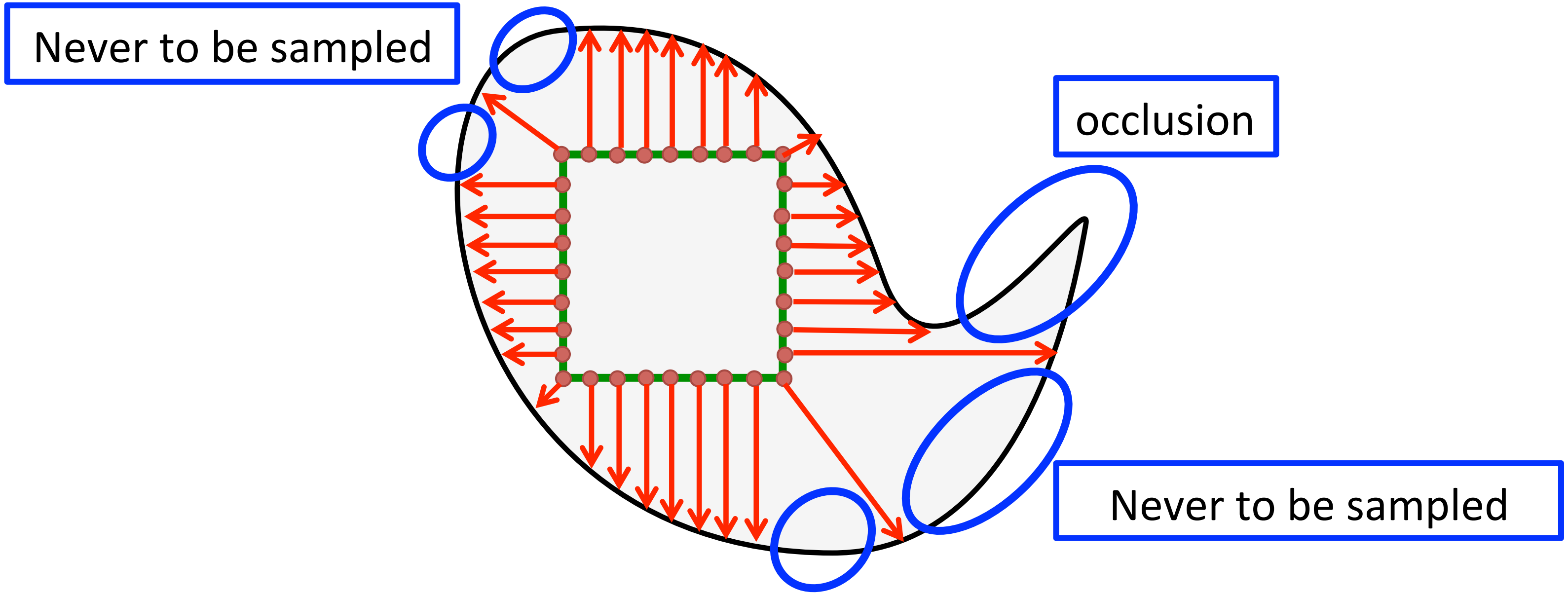
2.5 dimensional



3 dimensional

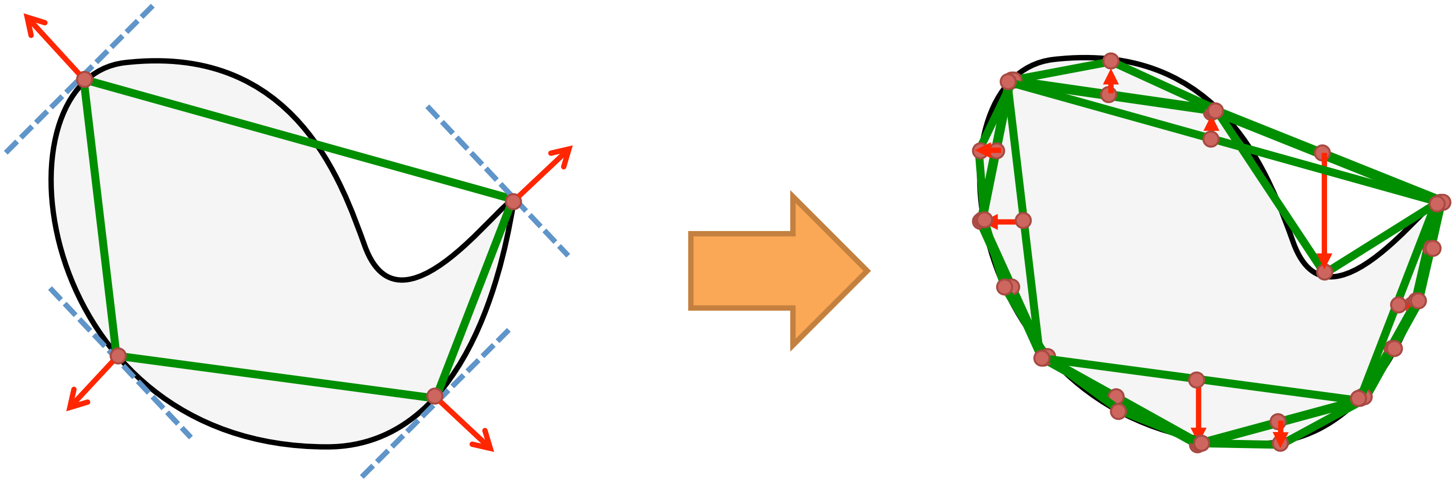
Naïve Projection Leads to Distortion

- Distortion means biased sampling density
- Not robust to the occlusion



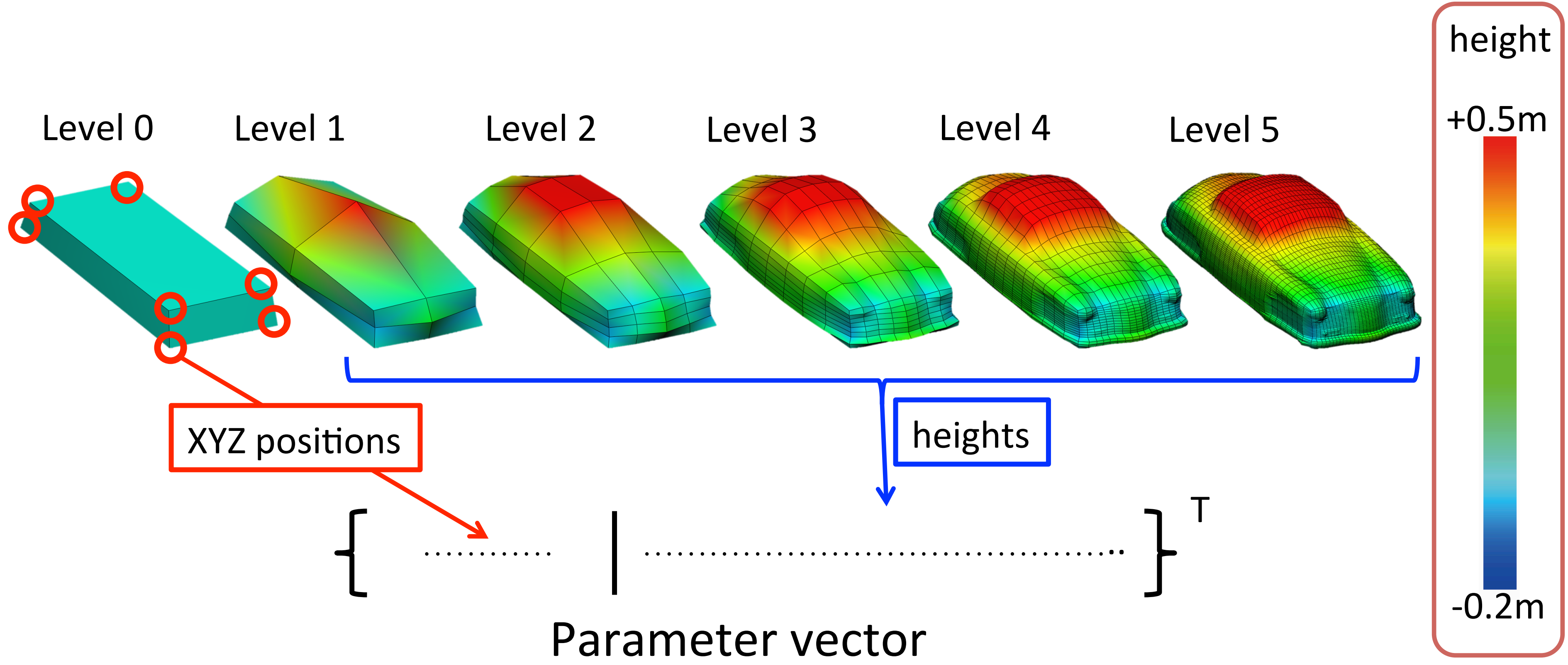
Hierarchical Projection

- We repeat subdivision and projection
- Key observation: concave shape is locally convex



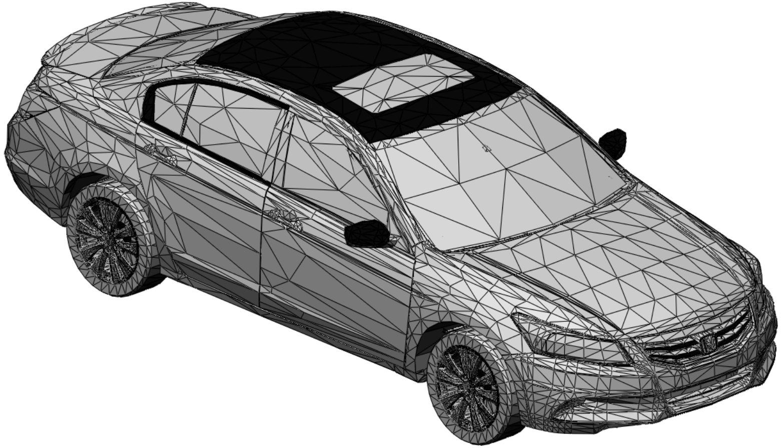
initialization subdivision & projection

Hierarchical Projection in 3D

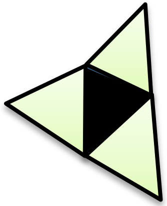


Challenge

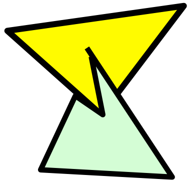
- Projecting points into polygon soup



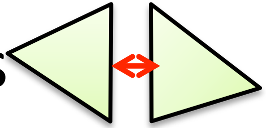
- Inverted triangle



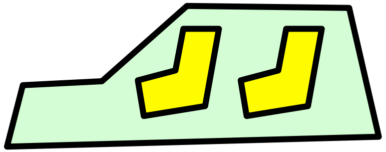
- Intersections



- Gaps and holes

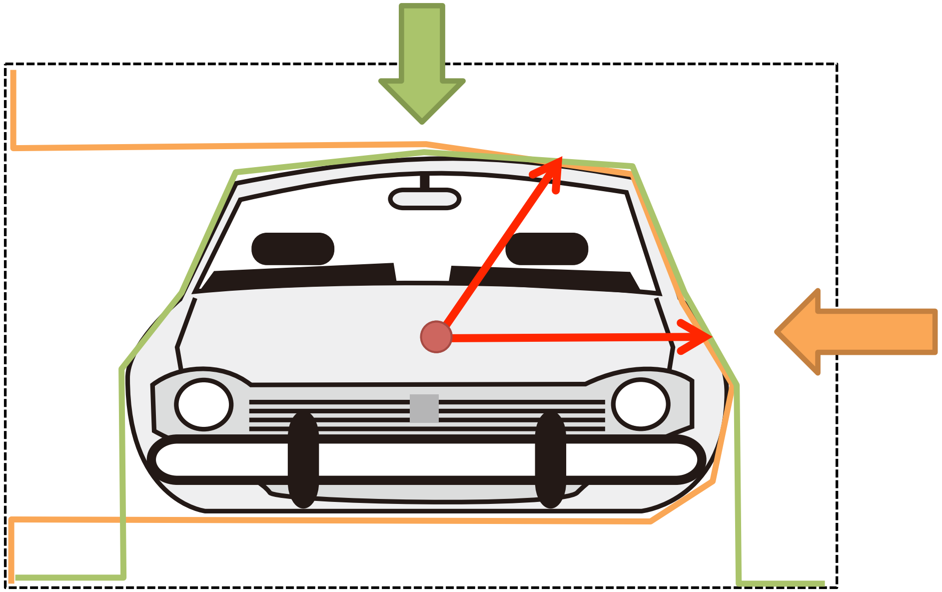


- Internal structures

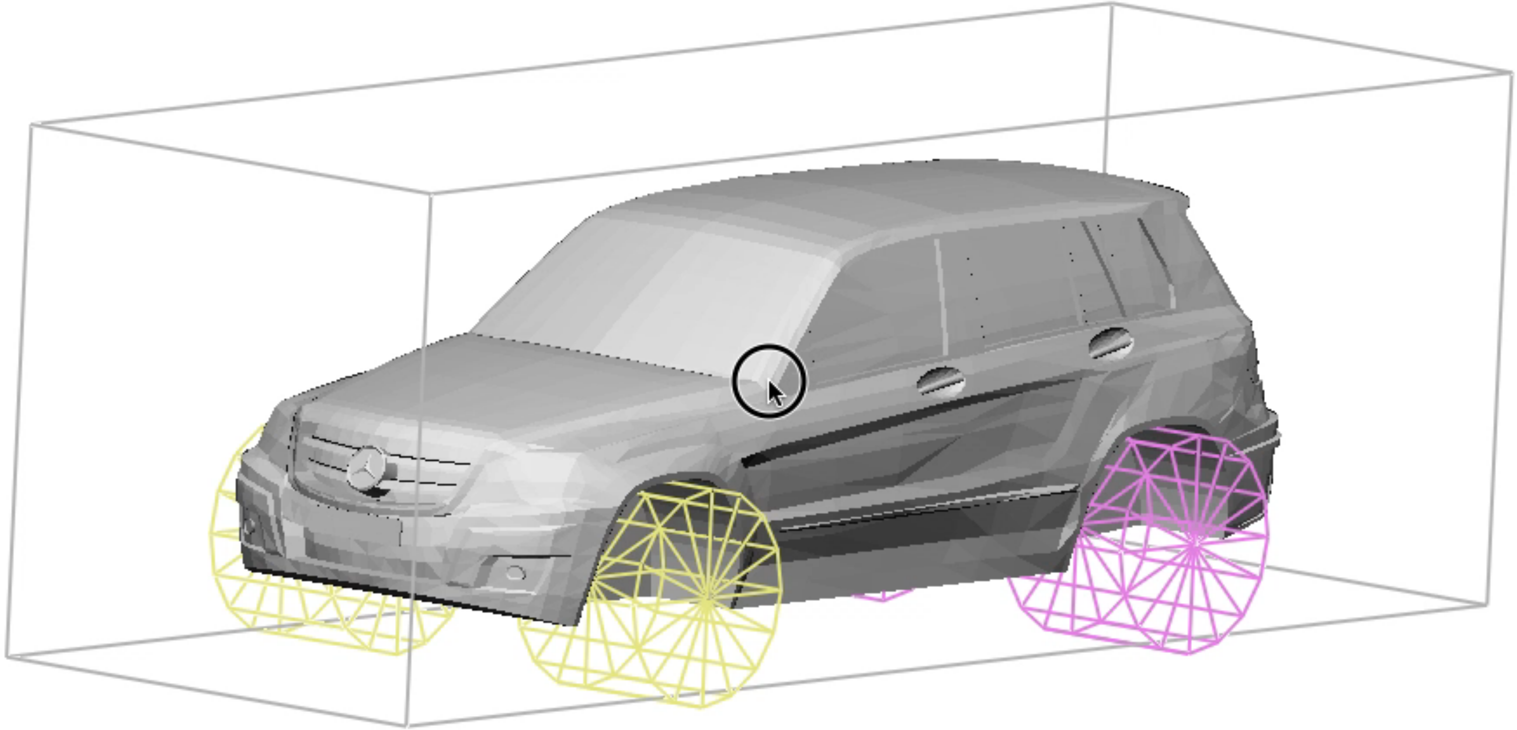


Depth Field Techniques

- Depth field can be efficiently computed with GPU
- More orthogonal projection is more trusted



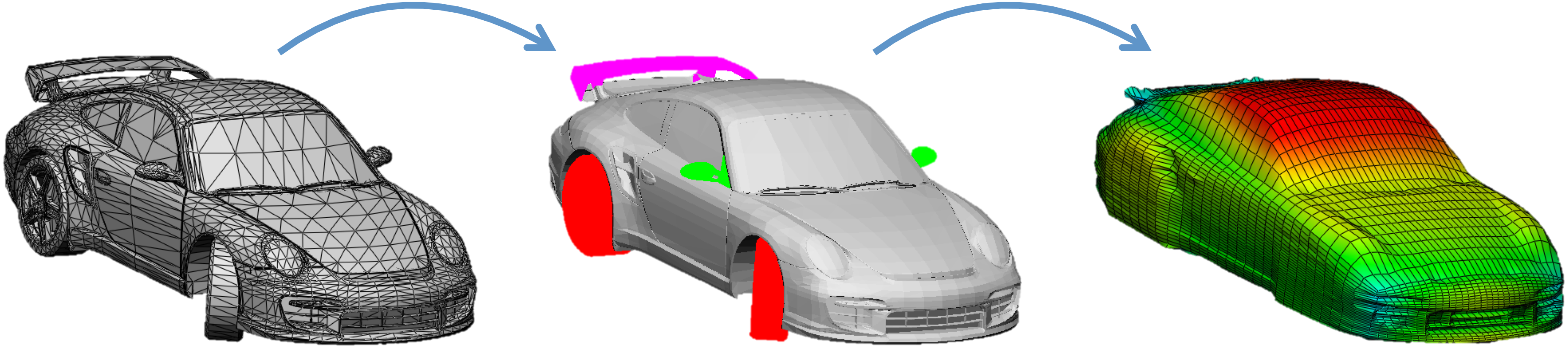
Parameterization



Over 1,200 Car Shapes from ShapeNet

[chang et al. 2015]

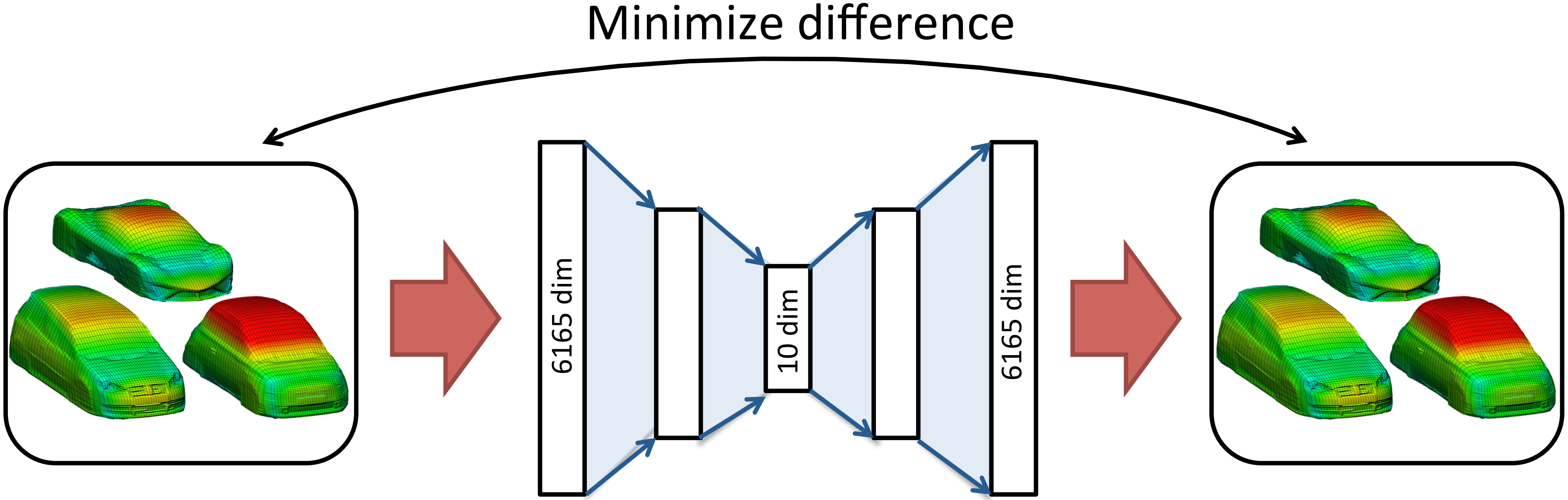
Manual annotation (5min) Parameterization(0.3sec)



Data is available on my web page: http://nobuyuki-umetani.com/

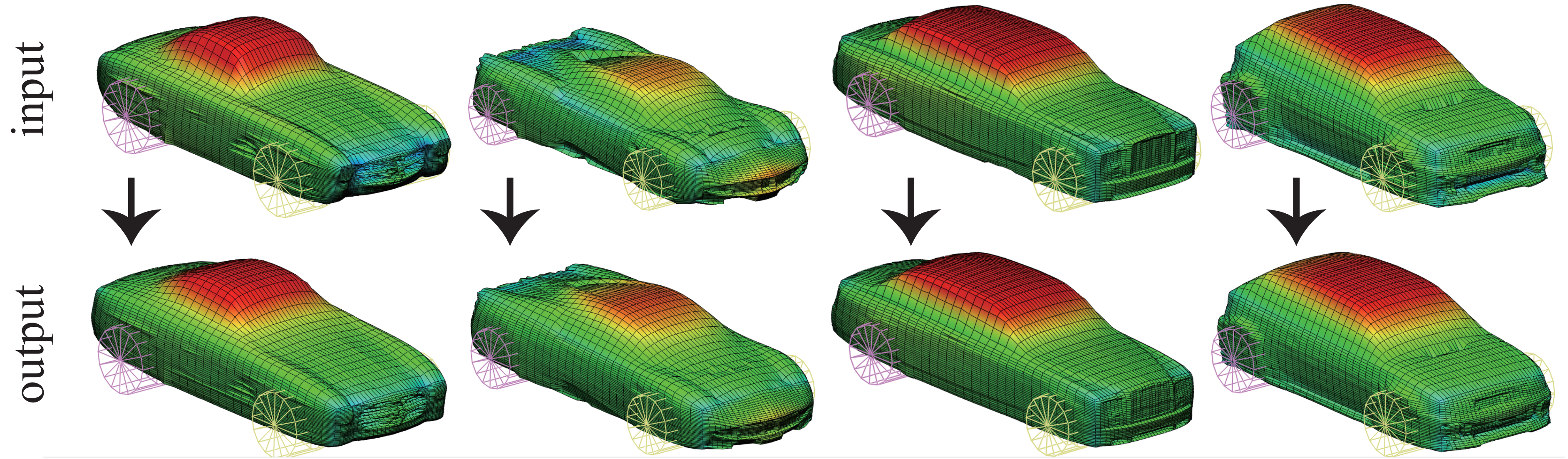
Autoencoder Network

- Input and output of network is as same as possible



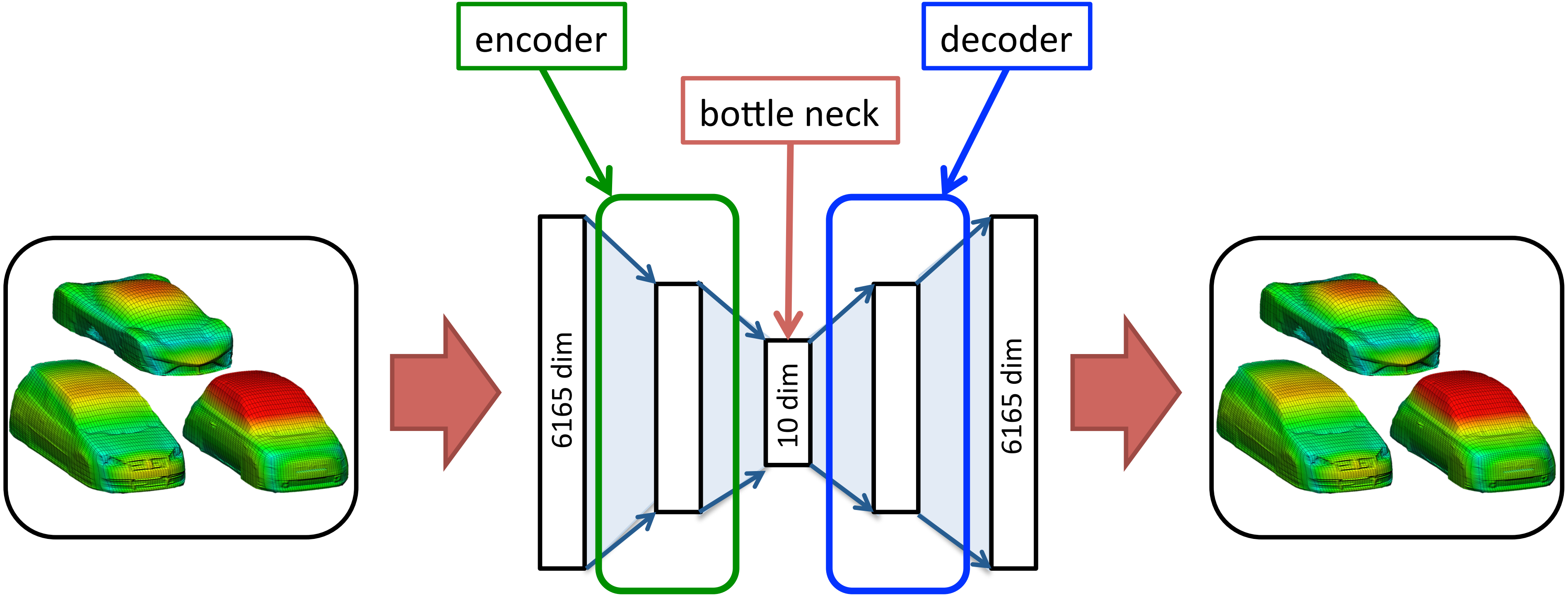
Autoencoder Network

- We could train the network preserving shapes



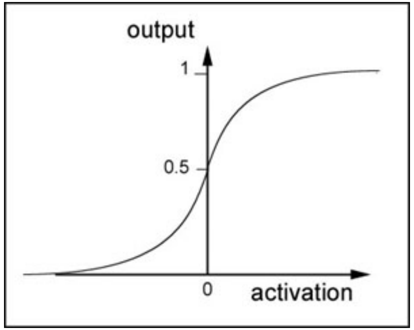
Autoencoder Network

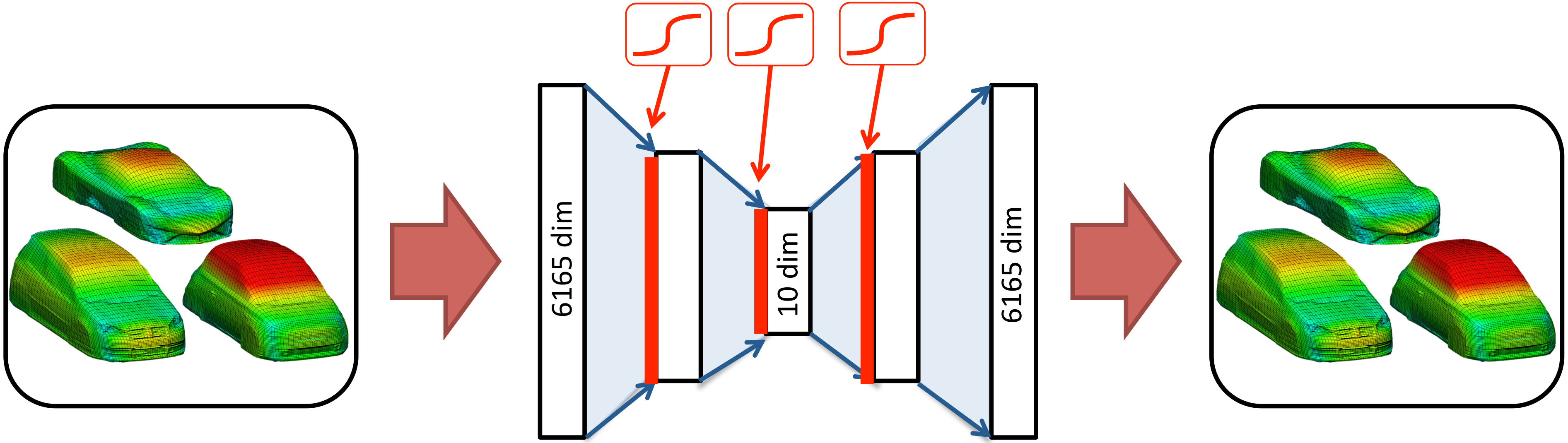
- Nonlinear dimensional reduction



Autoencoder Network

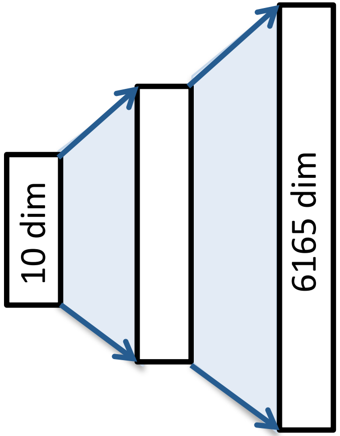
- We use Sigmoid function for activation
- Decoder takes value between [0,1]

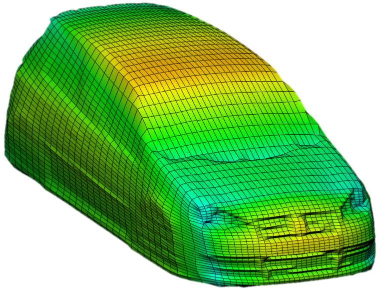




Autoencoder Network

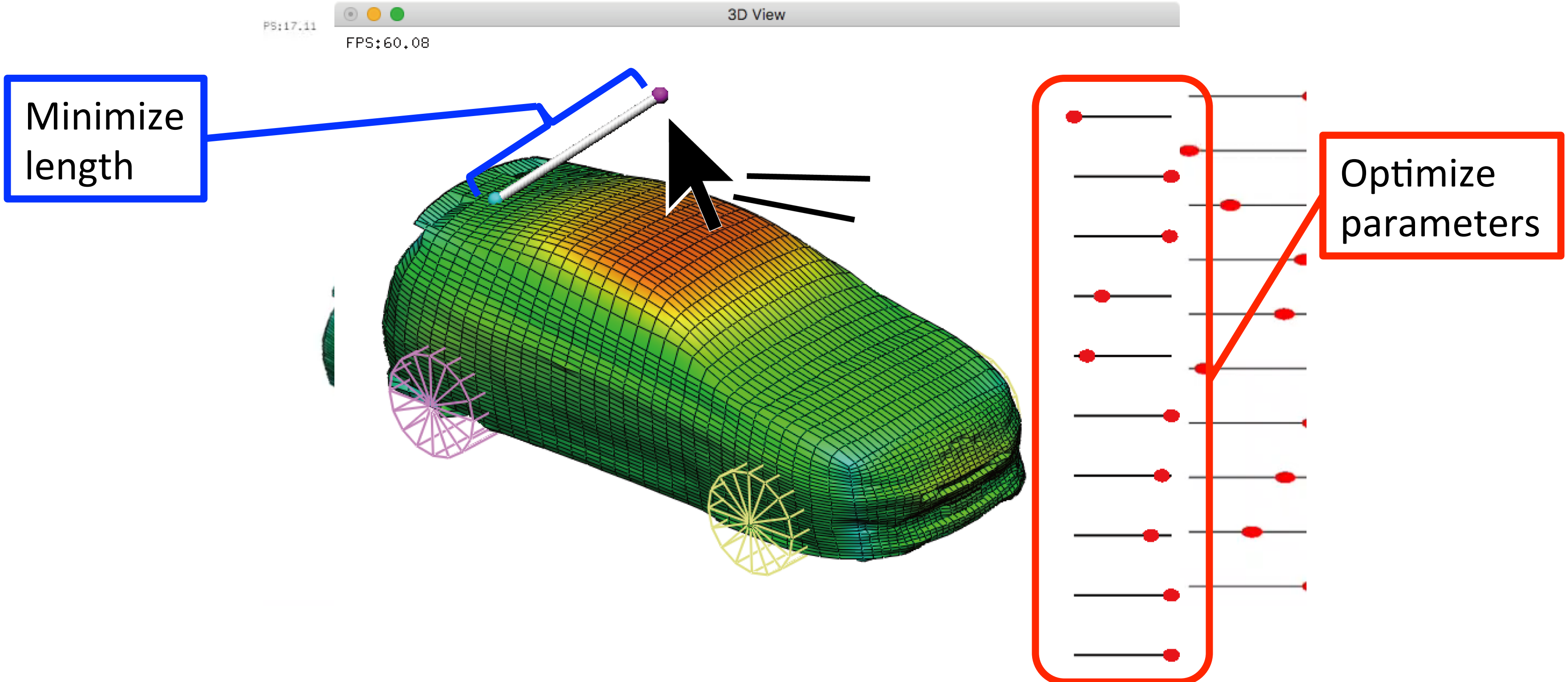
$$\{0.3 \ldots 0.1, \ldots 0.9\}^T$$





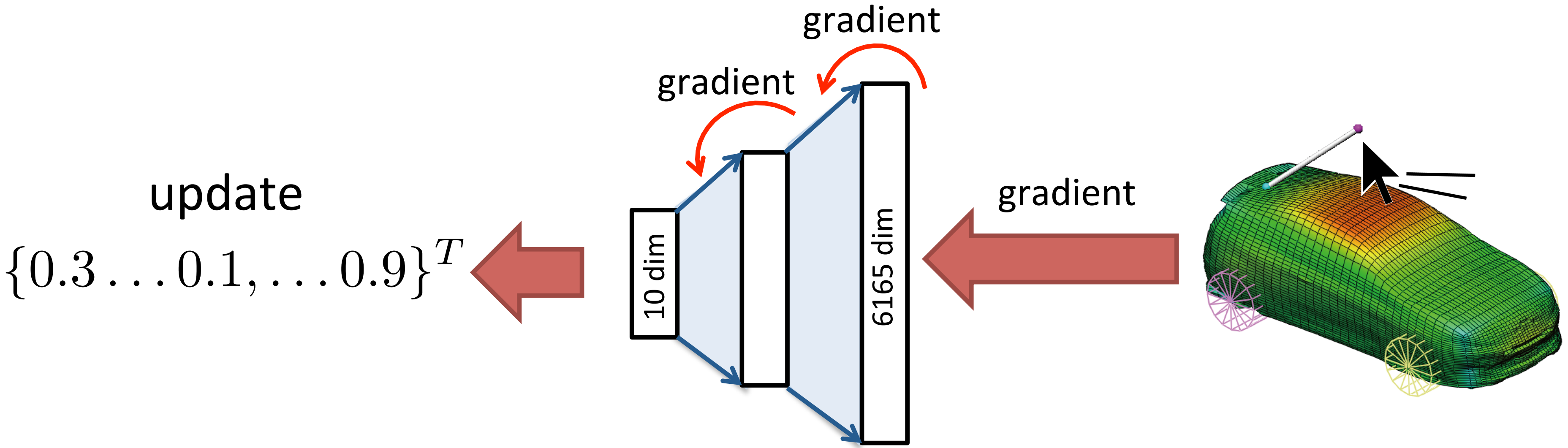
Interactive Exploration

- Pulling operation to guide the synthesis



Interactive Exploration

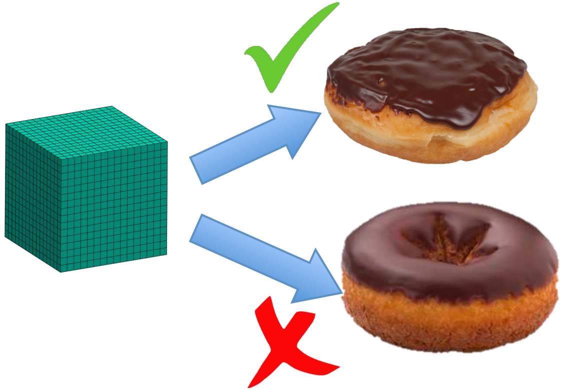
- Our parameterization is close form
  - Easily differentiable for grad-based optimization



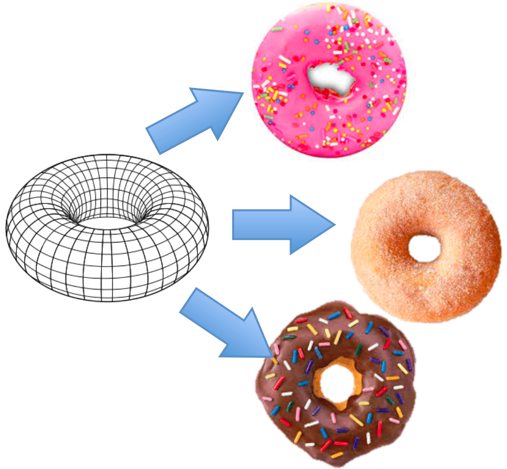
Live Demo!

Limitation

- Only handles one-class of shape
  - Difficult to handle different topologies



- Only handles nearly convex shapes
  - Starting from a template mesh other than cube
  - combinations of several meshes

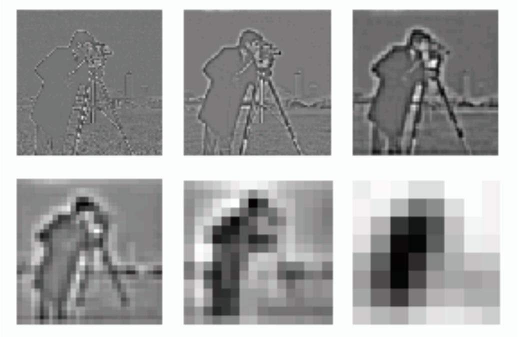


Future work

- Advanced generation framework
  - GAN, VAE

https://www.slideshare.net/vsevolodrodionov/itsubbotnik-rodionov-talk-neural-networks-in-js-js-itsubbotnik-2016

- Filter operations on the subdivision
  - Convolution operation
  - Gaussian Pyramid / Laplacian Pyramid



http://library.wolfram.com/infocenter/Demos/4532/

Acknowledgement

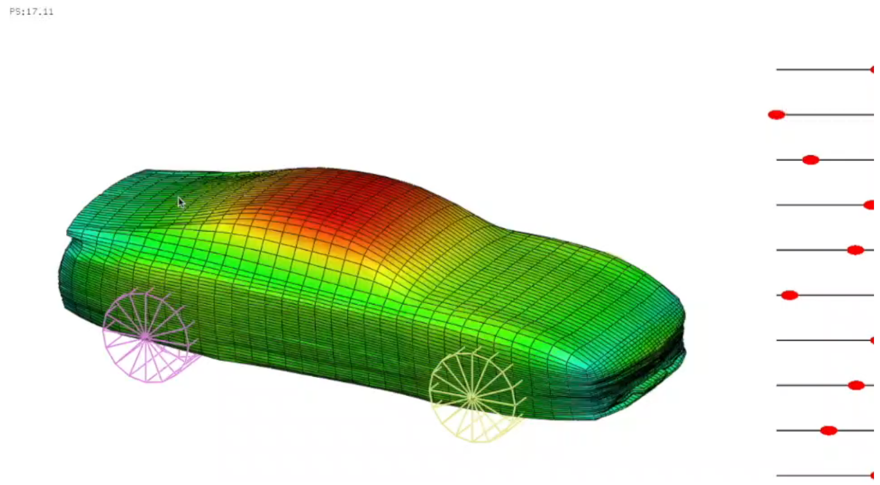
- Anonymous reviewers

- Ryan Schmidt and Alec Jacobson for many advice

dgp

dynamic graphics project

Exploring Generative 3D Shapes Using Autoencoder Networks



Nobuyuki Umetani

AUTODESK RESEARCH