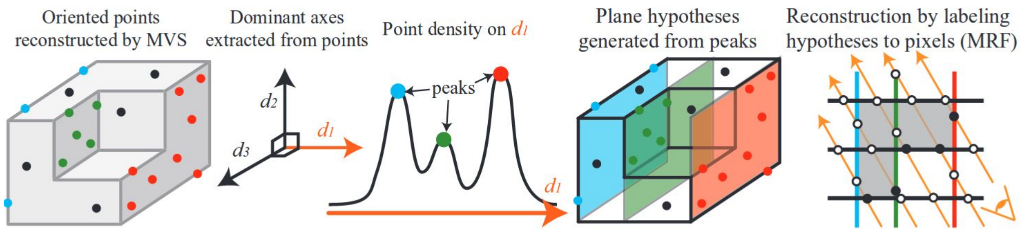
Deep Learning for 3D
Toward surface generation

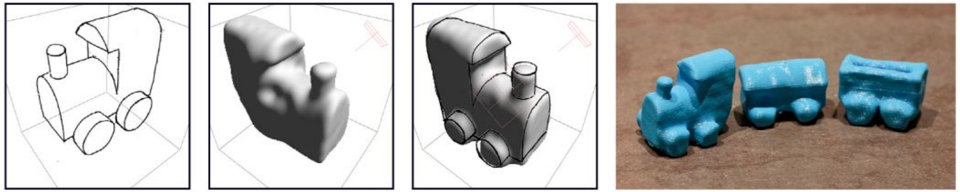
Thibault GROUEIX, Pierre-Alain LANGLOIS

Why learn ?

1. Get rid of hand crafted priors - Manhattan world assumption **[Furukawa2009]**



2. Discover complex prior from data itself - Discovering 3D from sketch **[Delanoy2017]**



a) Initial drawing | b) 3D prediction seen from another viewpoint | c) New drawing and updated prediction | d) 3D printed objects

Thibault Groueix, Pierre-Alain Langlois, 2018

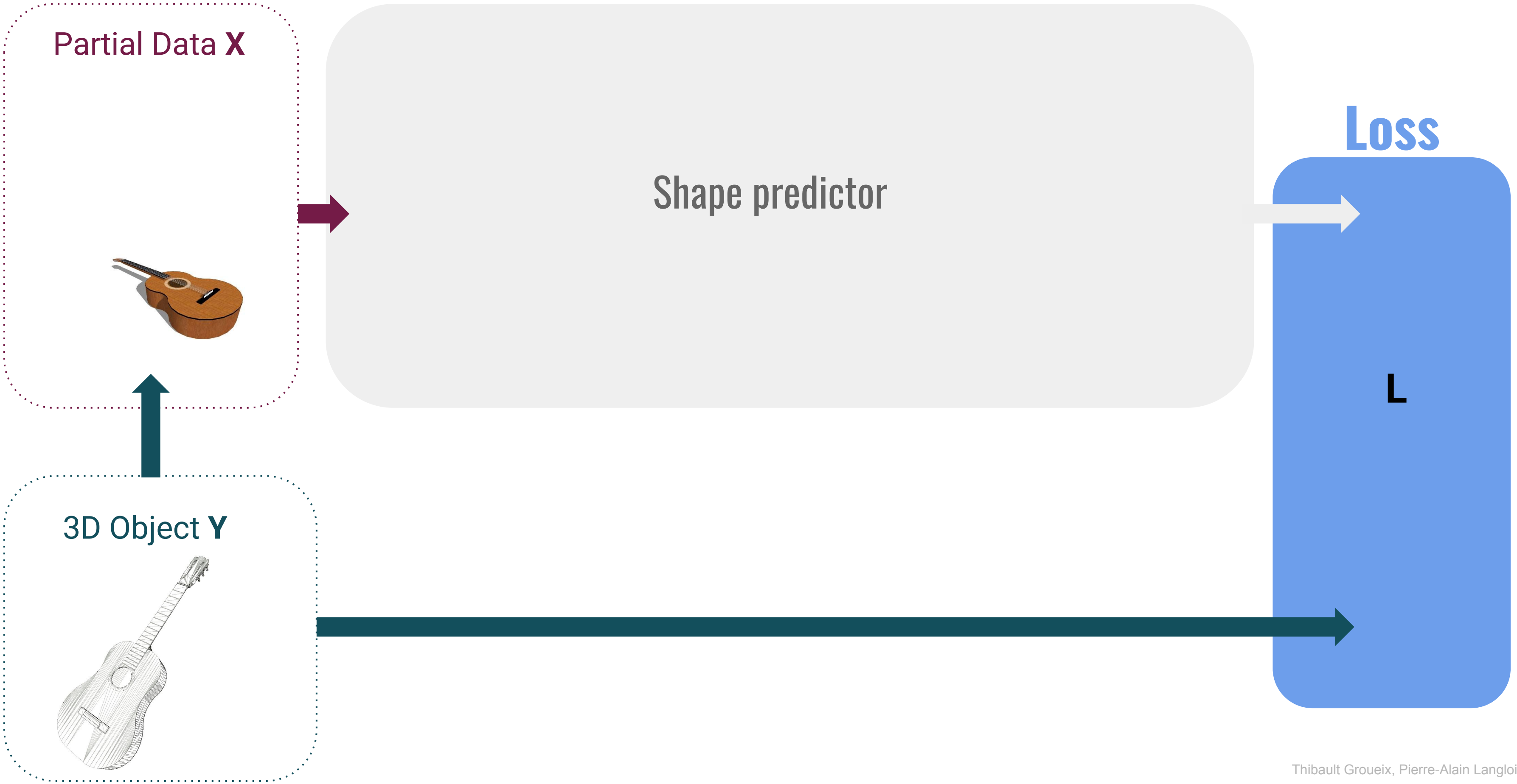
Data types

RGB Image(s)

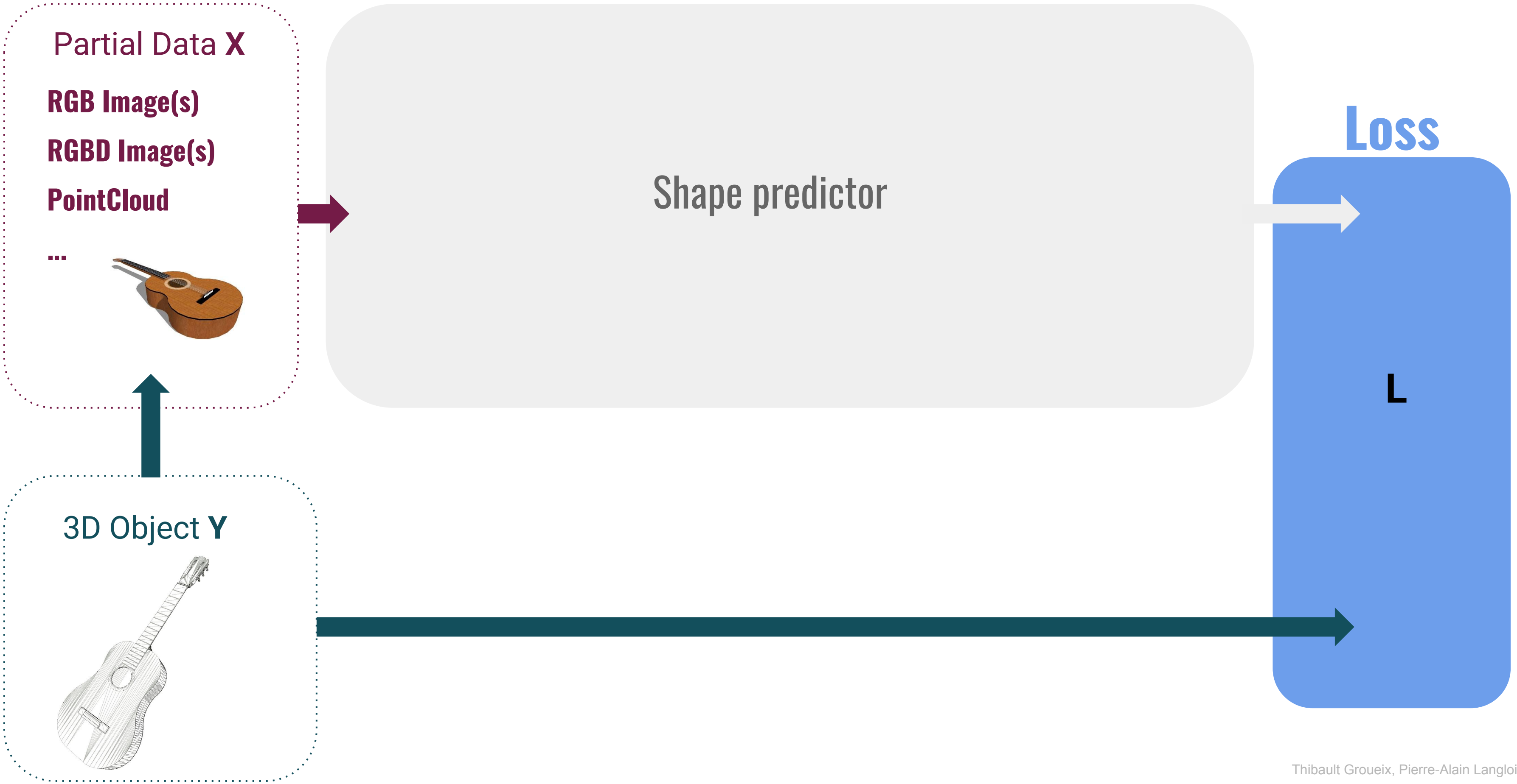
RGBD Image(s)

PointCloud

Typical learning framework based on synthetic data

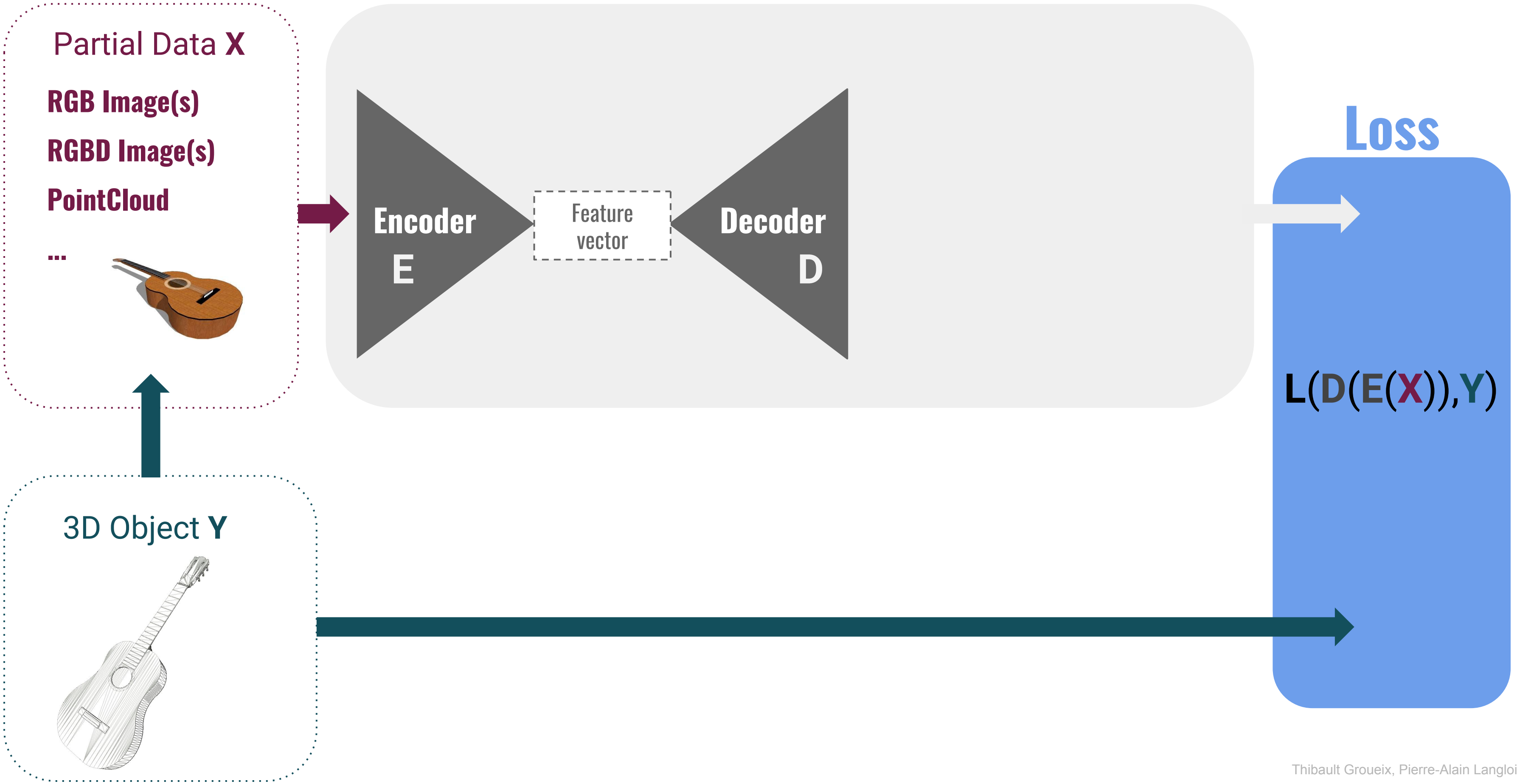


Training setup for 3D reconstruction

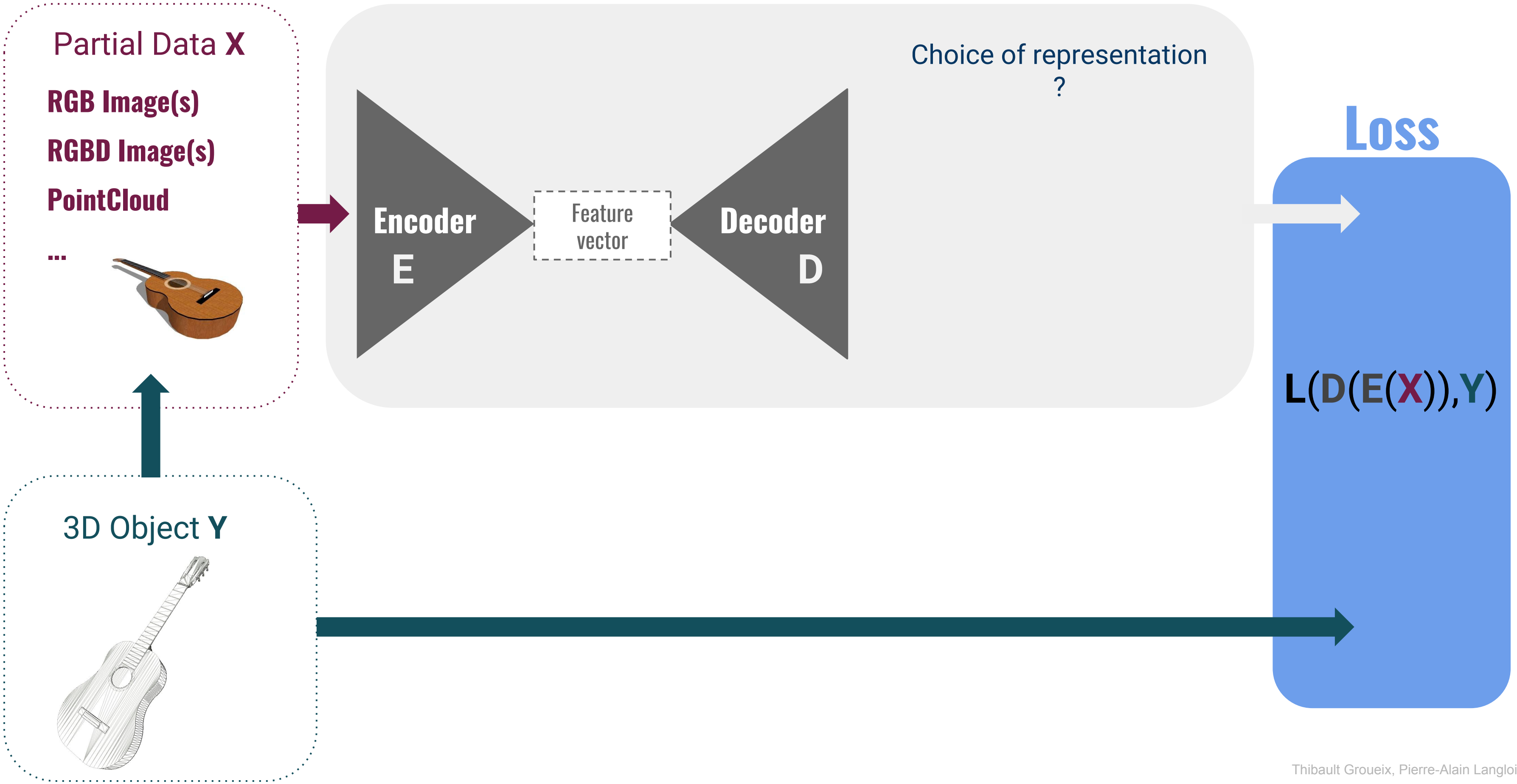


Thibault Groueix, Pierre-Alain Langlois, 2018

Training setup for 3D reconstruction



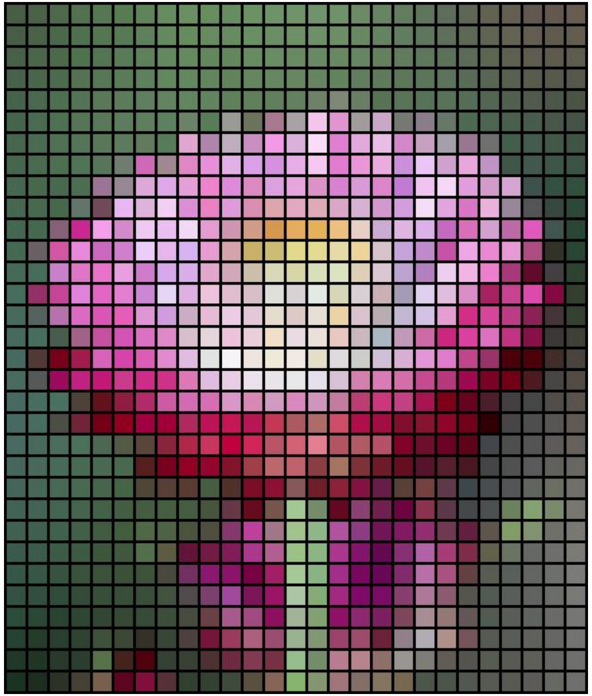
Training setup for 3D reconstruction



Thibault Groueix, Pierre-Alain Langlois, 2018

Representations

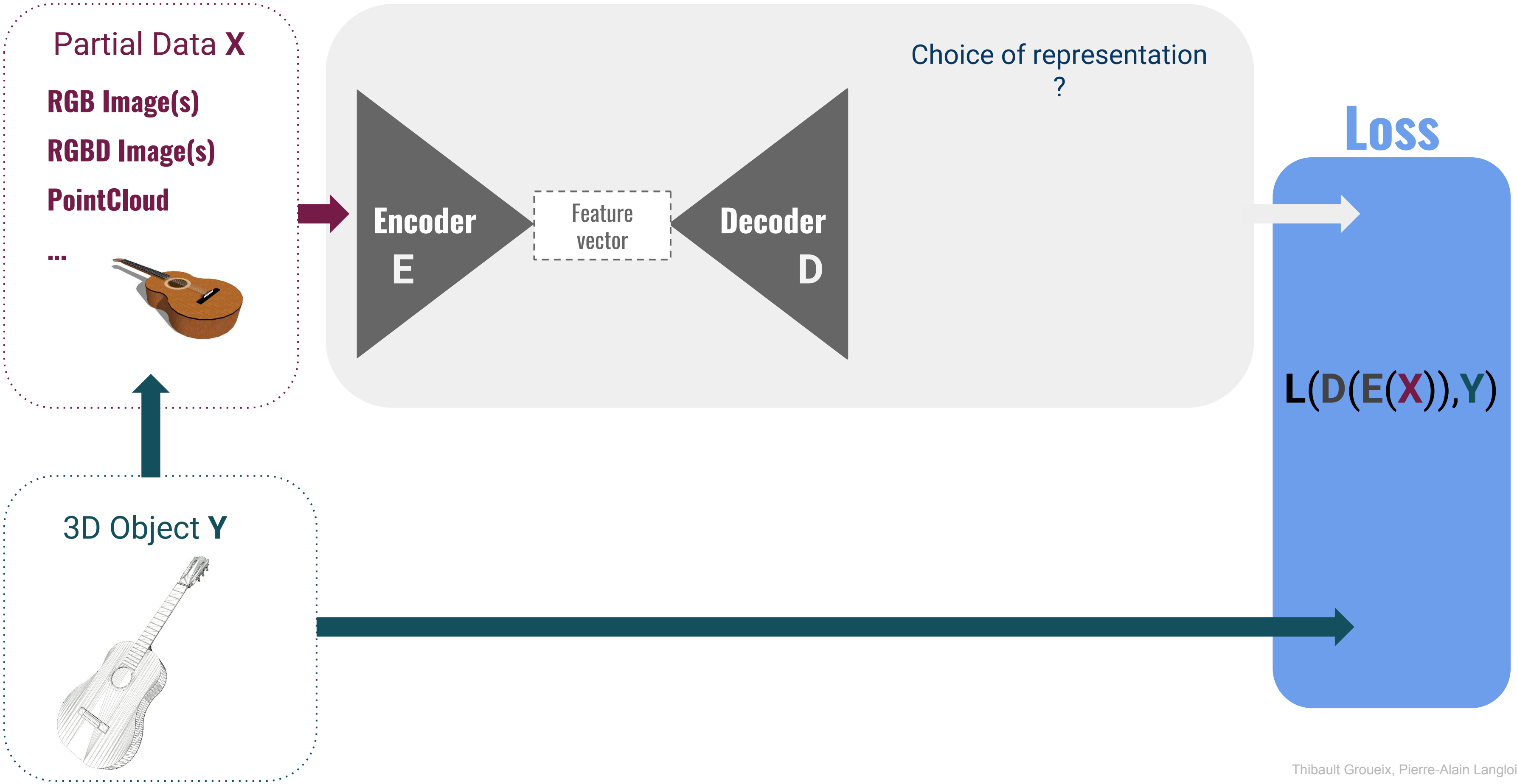
Obvious in 2D...



Not so obvious in 3D !

Thibault Groueix, Pierre-Alain Langlois, 2018

Training setup for 3D reconstruction



Thibault Groueix, Pierre-Alain Langlois, 2018

Training setup for 3D reconstruction

Partial Data **X**

- **RGB Image(s)**
- **RGBD Image(s)**
- **PointCloud**
- **...**

Encoder E

Feature vector

Decoder D

Choice of representation

1. **Volumetric (OctNet)**
2. **PointClouds (PointSetGen)**
3. **Surfaces (AtlasNet)**
4. **Geometric Primitives**

Loss

L(D(E(X)),Y)

3D Object **Y**

Training setup for 3D reconstruction

Partial Data **X**

RGB Image(s)

RGBD Image(s)

PointCloud

...

Encoder
E

Feature vector

Decoder
D

Choice of representation

1. **Volumetric (OctNet)**
2. **PointClouds (PointSetGen)**
3. **Surfaces (AtlasNet)**
4. **Geometric Primitives**

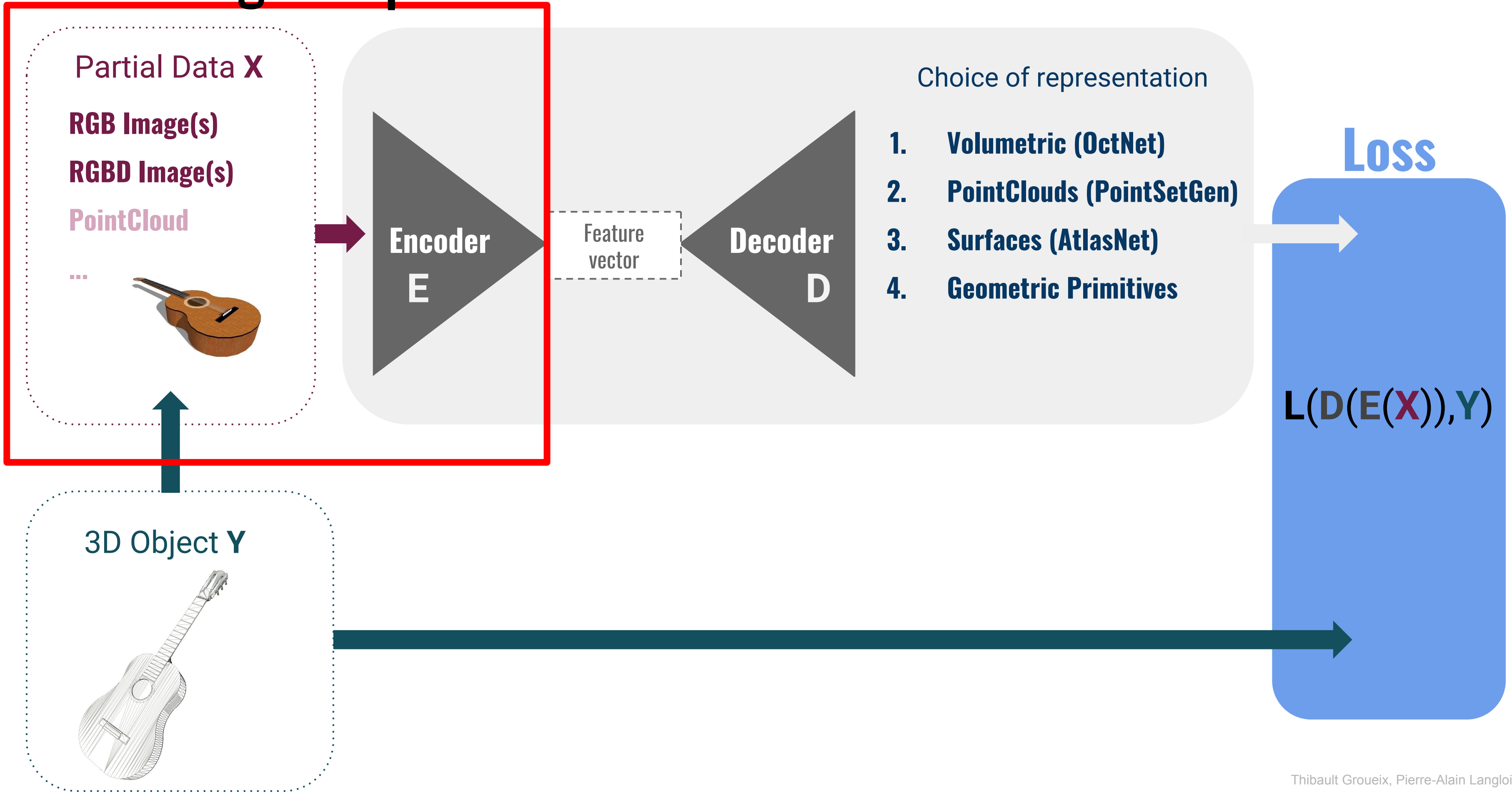
3D Object **Y**

Loss

L(D(E(X)),Y)

Thibault Groueix, Pierre-Alain Langlois, 2018

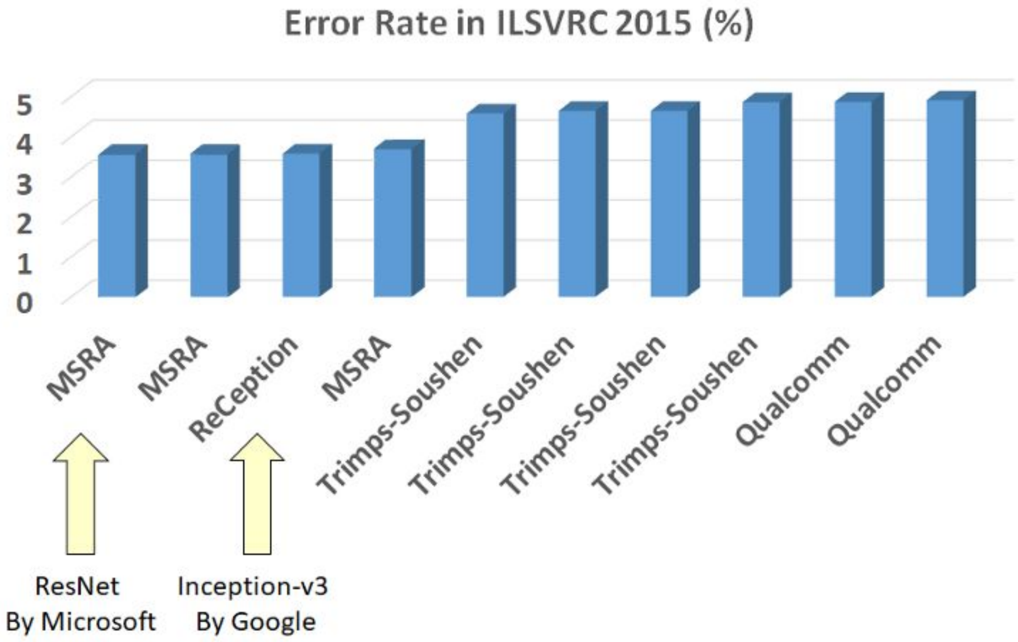
Training setup for 3D reconstruction



Encoders for RGB & RGBD images

Encoder
E

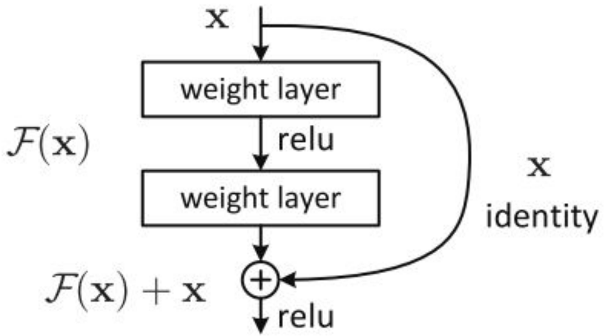
Do not reinvent the wheel :
Use State-of-the-art 2D networks



Encoders for RGB & RGBD images

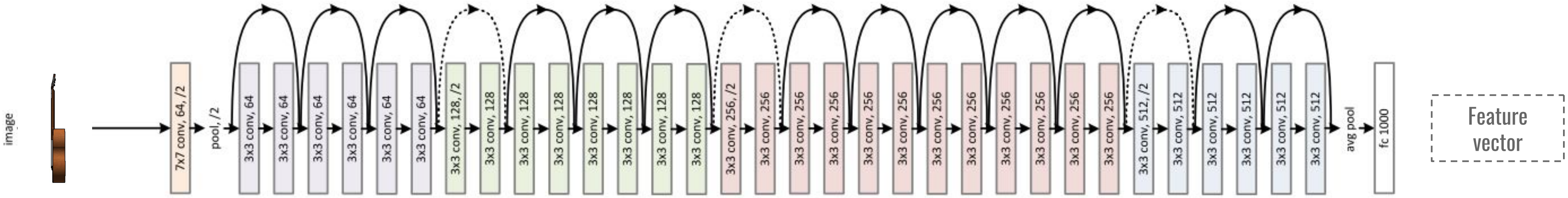
Do not reinvent the wheel :
Use State-of-the-art 2D networks

- Resnet [He2015] -> Skip connections
- BatchNorm [Ioffe2015]

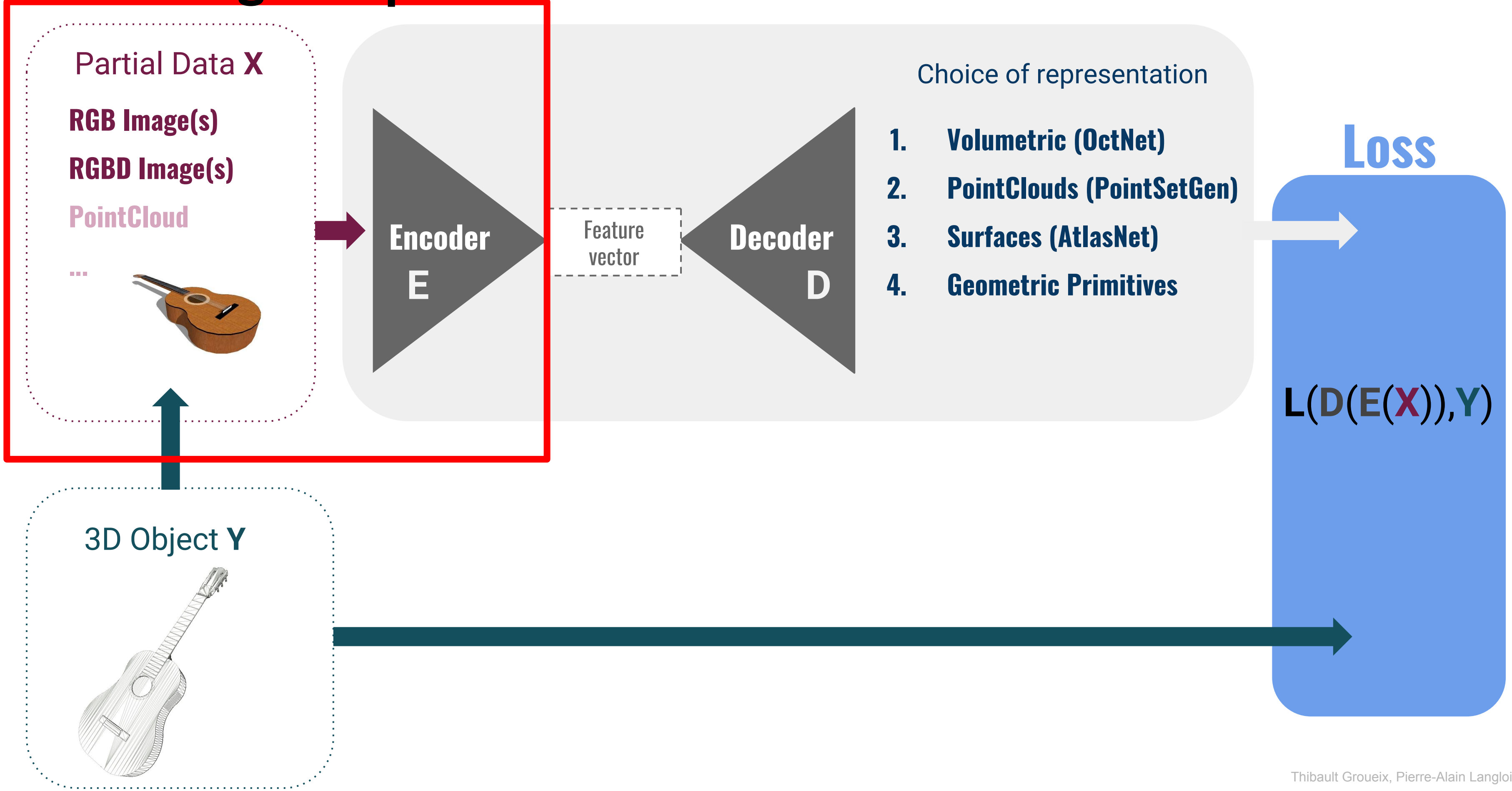


Encoder
E

Resnet 34 [He2015]

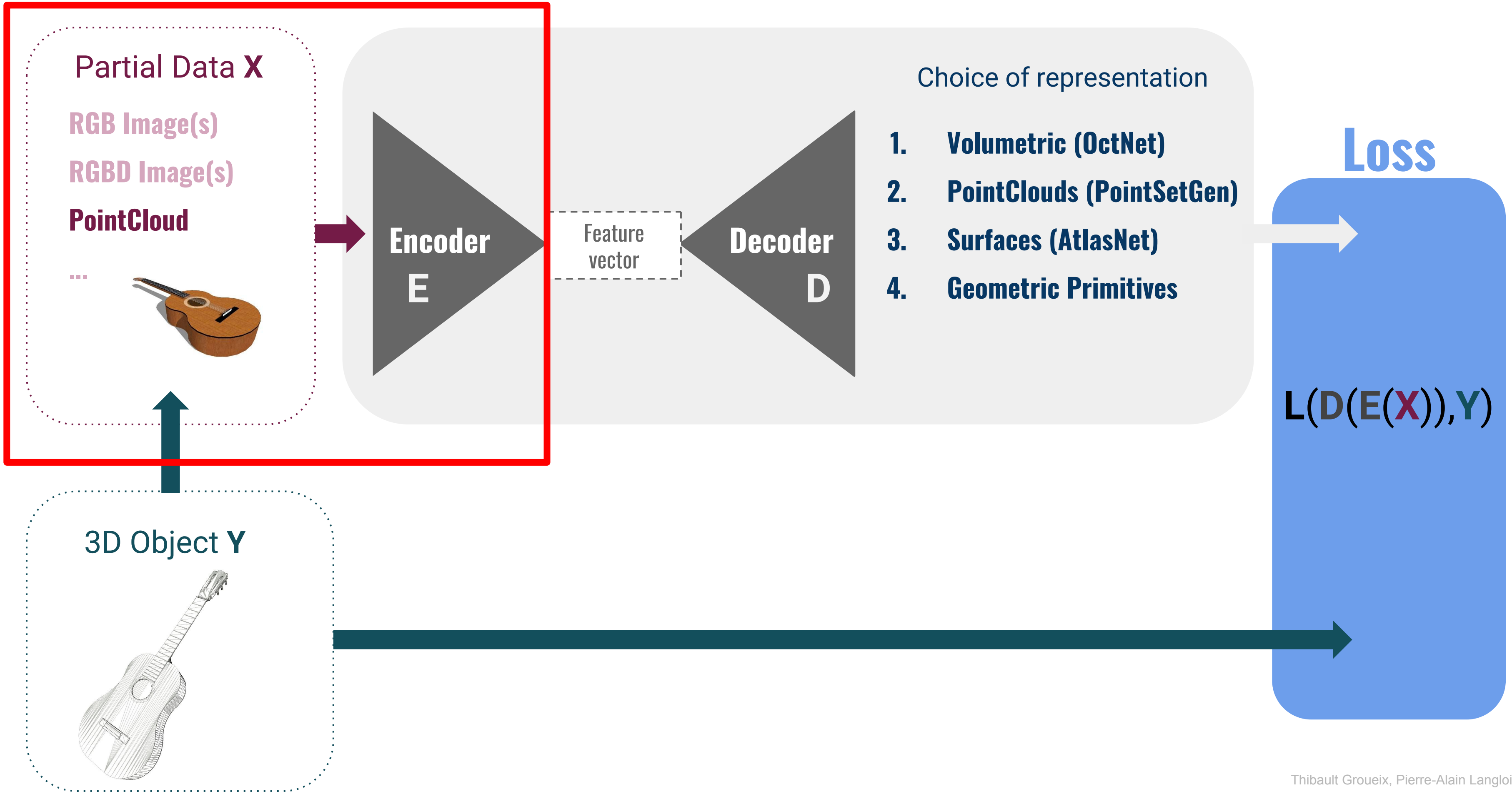


Training setup for 3D reconstruction

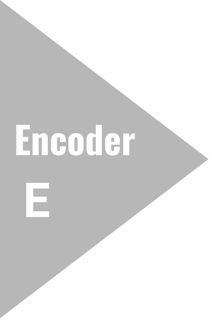


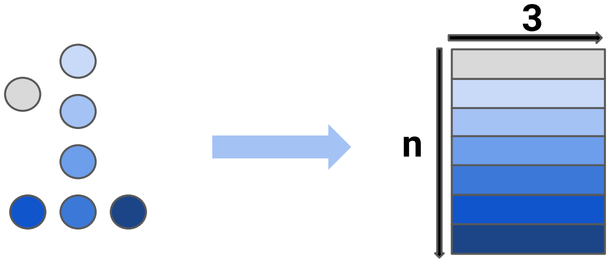
Thibault Groueix, Pierre-Alain Langlois, 2018

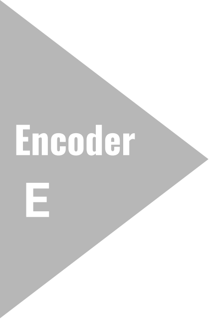
Training setup for 3D reconstruction

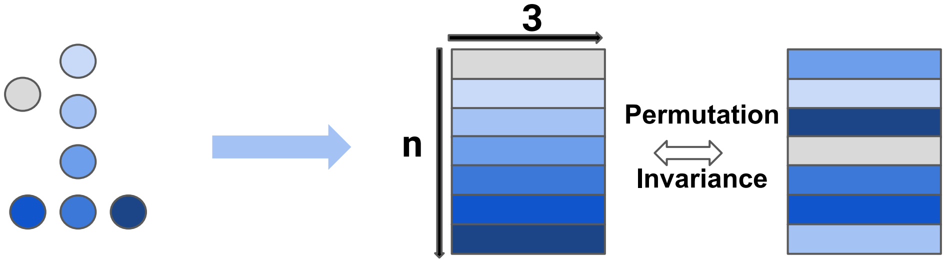


Thibault Groueix, Pierre-Alain Langlois, 2018


Encoder
E

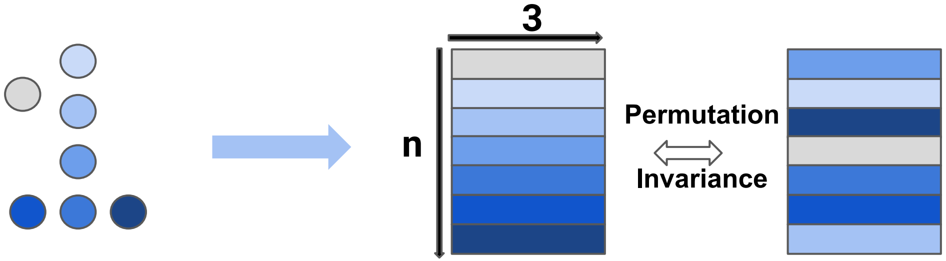

3
n



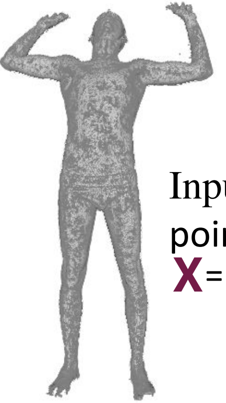


PointNet [Qi2017]

Encoder
E

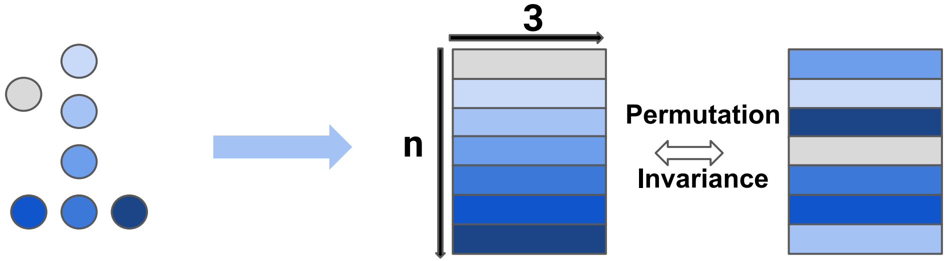


PointNet [Qi2017]

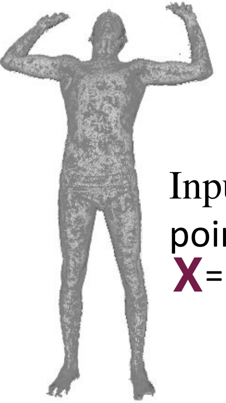


Input
pointcloud
$\mathbf{X} = (x_1, x_2, \ldots, x_n)$

Encoder
E



PointNet [Qi2017]



Input
pointcloud
$\mathbf{X} = (x_1, x_2, \ldots, x_n)$

$x_1 = (1,2,3) \longrightarrow$

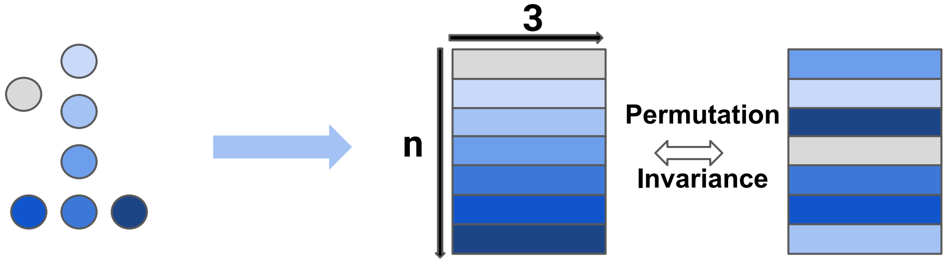
$x_2 = (1,1,1) \longrightarrow$

$x_{\ldots} = (2,3,2) \longrightarrow$

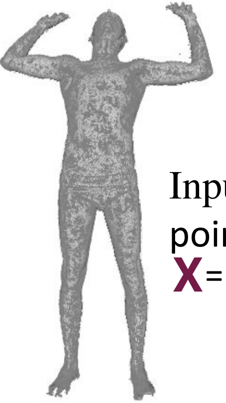
$x_n = (2,3,4) \longrightarrow$

$$\mathbf{E}((x_1, x_2, \ldots, x_n)) = \quad x_1 \; , \ldots , \quad x_n$$

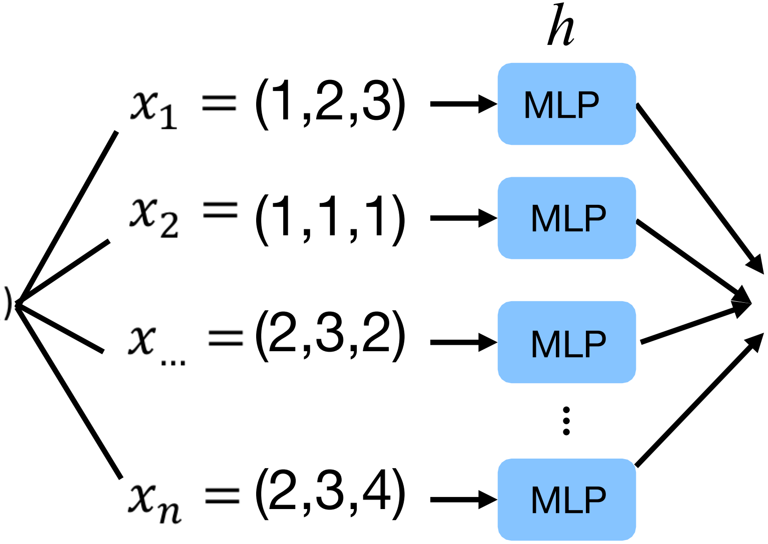
Encoder
E



PointNet [Qi2017]

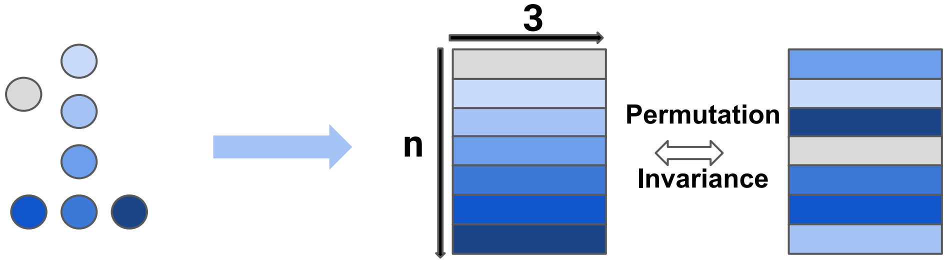


Input
pointcloud
$\mathbf{X} = (x_1, x_2, \ldots, x_n)$

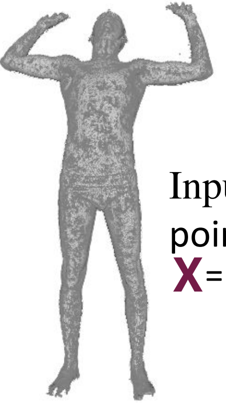


$$\mathbf{E}((x_1, x_2, \ldots, x_n)) = \quad h(x_1), \ldots, h(x_n)$$

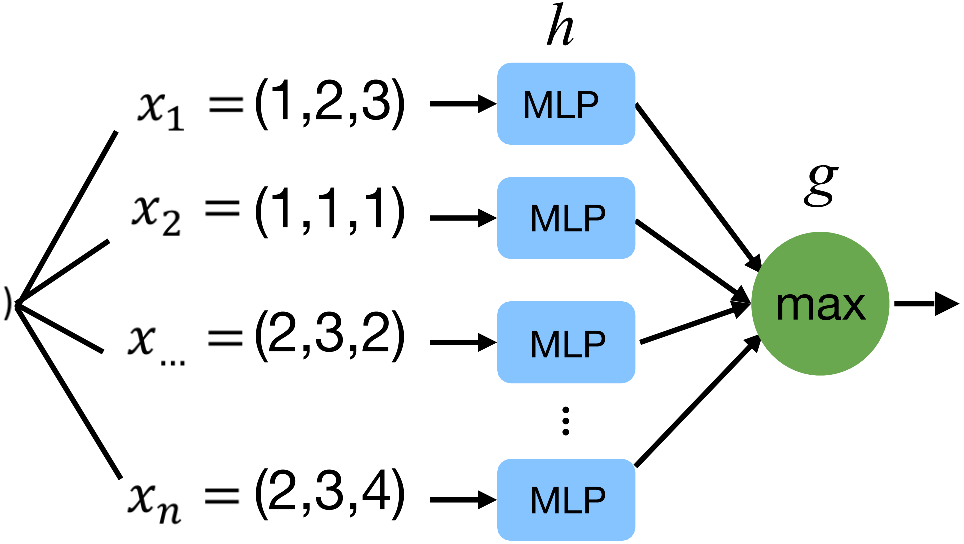
Encoder
E



PointNet [Qi2017]



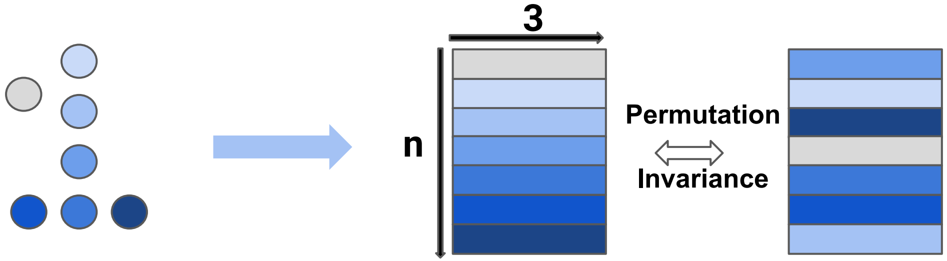
Input pointcloud
$\mathbf{X} = (x_1, x_2, \dots, x_n)$

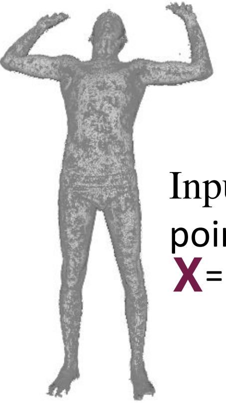


$$\mathbf{E}((x_1, x_2, \dots, x_n)) = \quad g(\, h(\, x_1 \,), \dots, \; h(\, x_n \,)\,)$$

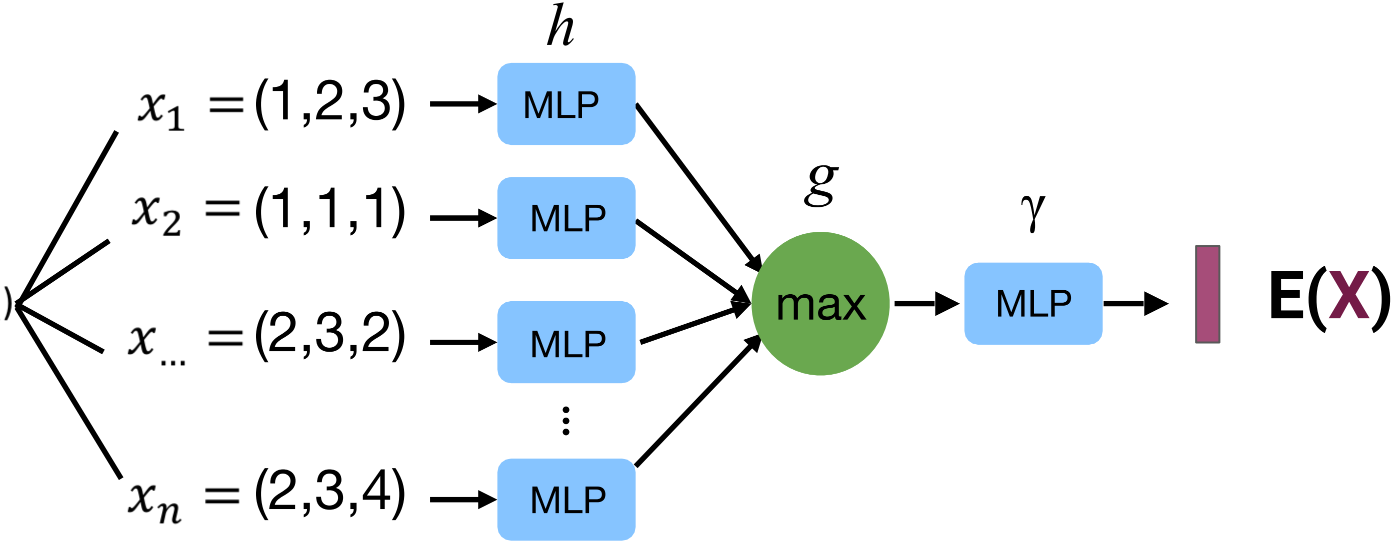
Encoder
E

PointNet [Qi2017]

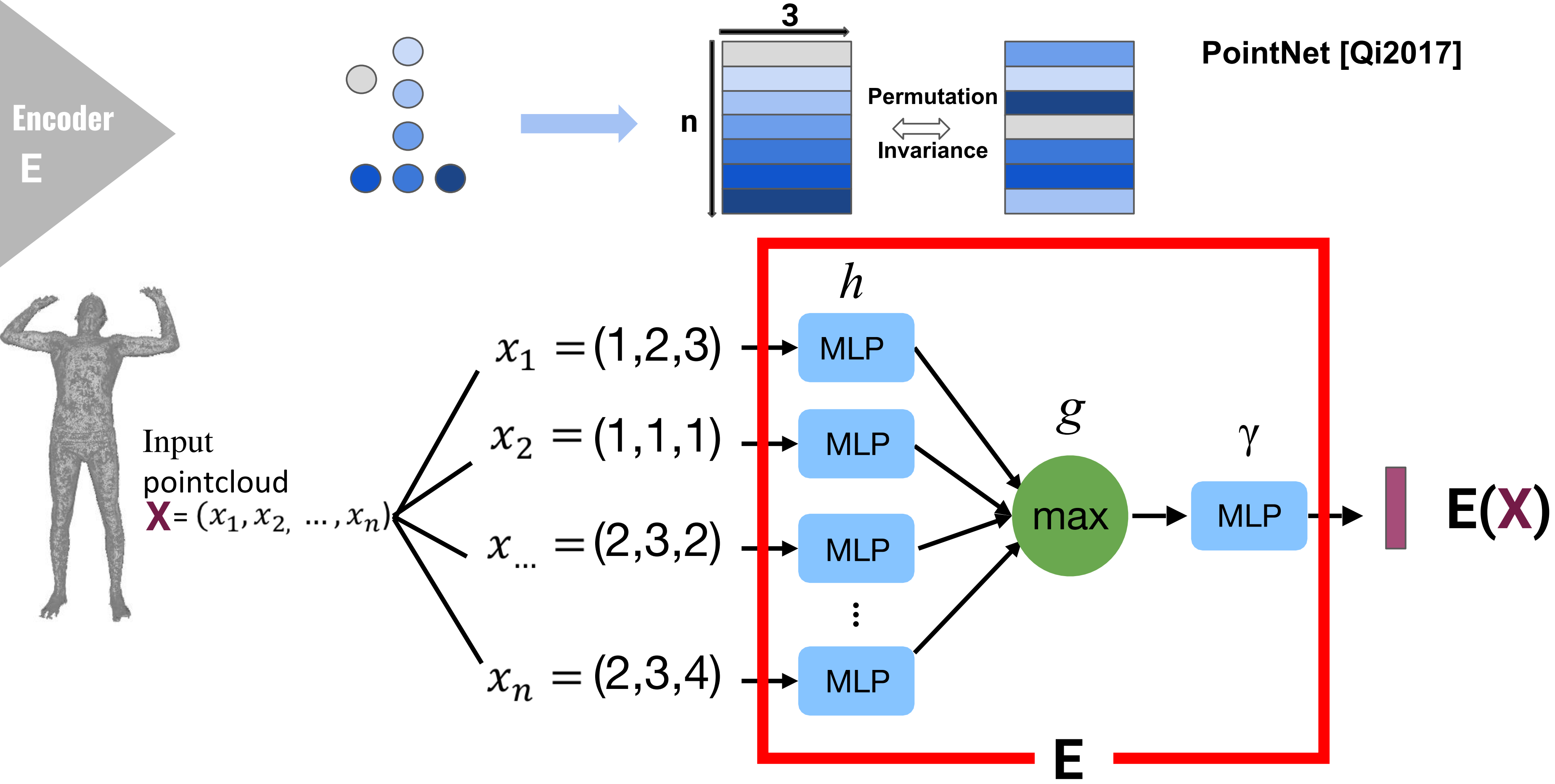




Input
pointcloud
X = $(x_1, x_2, \ldots, x_n)$



$$\mathbf{E}((x_1, x_2, \ldots, x_n)) = \gamma(\, g(\, h(x_1), \ldots, h(x_n)\,)\,)$$

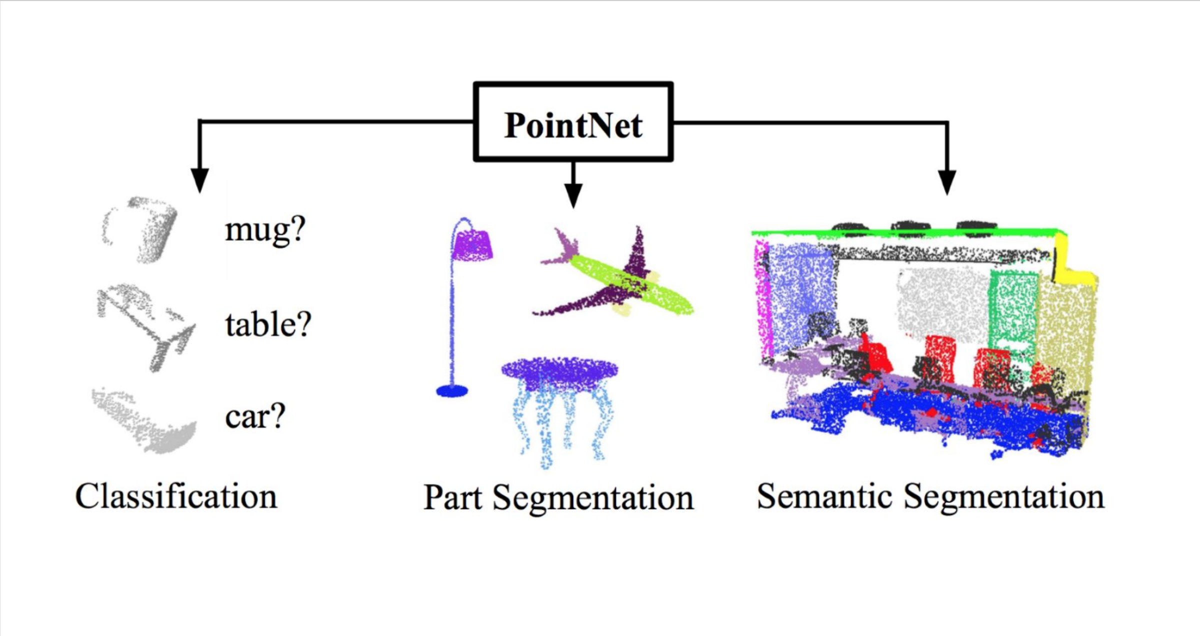


$$\mathbf{E}((x_1, x_2, \ldots, x_n)) = \gamma(\, g(\, h(\, x_1\,), \; \ldots, \; h(\, x_n\,)\,)\,)$$

Encoder

E

Results : Unified framework for various tasks



Credit [**Qi2017**]

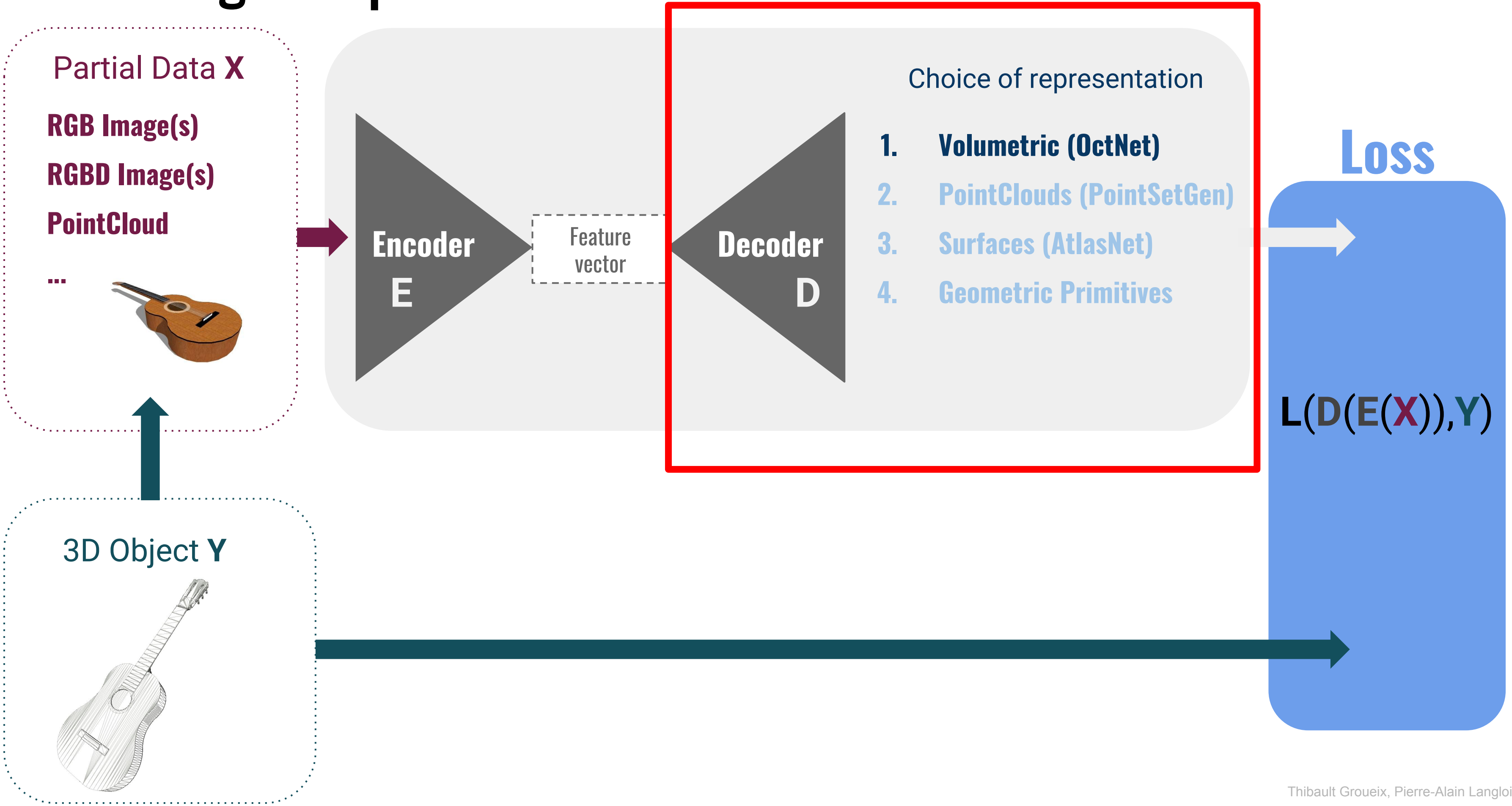
Encoder

E

A number of alternatives exists

- ➔ PointNet++ [**Qi2017b**]:
- ➔ KD-Trees : **[Klokov2017]**
- ➔ PCPNet [**Guerrero2017**]
- ➔ Large-scale PointClouds : SuperPointGraph [**Landrieu2018**]
- ➔ Build a graph on top and apply graph neural networks : SyncSpecNet **[Yi2016]**
- ➔ Projection on enclosing sphere and equivariant convolutions from SO(3) [**Esteves2018**, **Cohen2018**]

Training setup for 3D reconstruction



Thibault Groueix, Pierre-Alain Langlois, 2018

Encoder
E

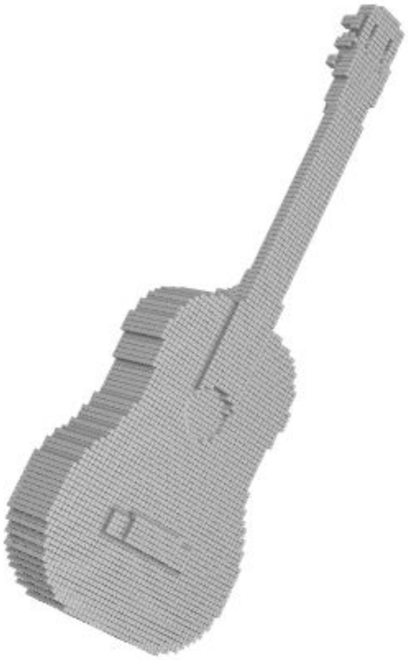
Voxels 3d-r2n2 **[Choy2016]**, Voxnet **[Maturana2015]**, **[Qi2016]**, **[Wu2015]**

Decoder
D

➔ A 3D regular grid which subdivides a bounding box in the 3D space

➔ Allows direct generalization of the 2D methods (convolutions, pooling)

➔ Subject to the **curse of dimensionality** : memory inefficient

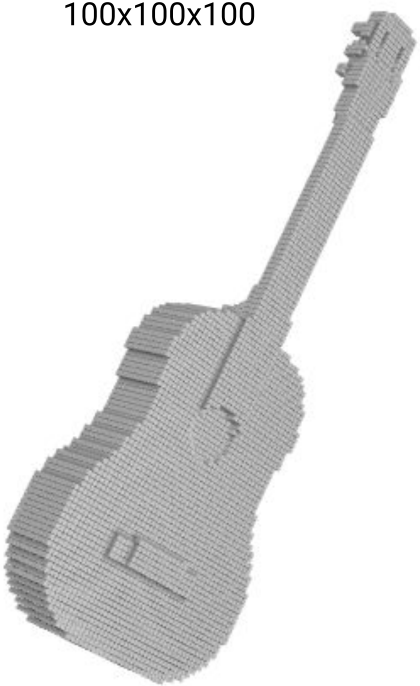


Encoder
E

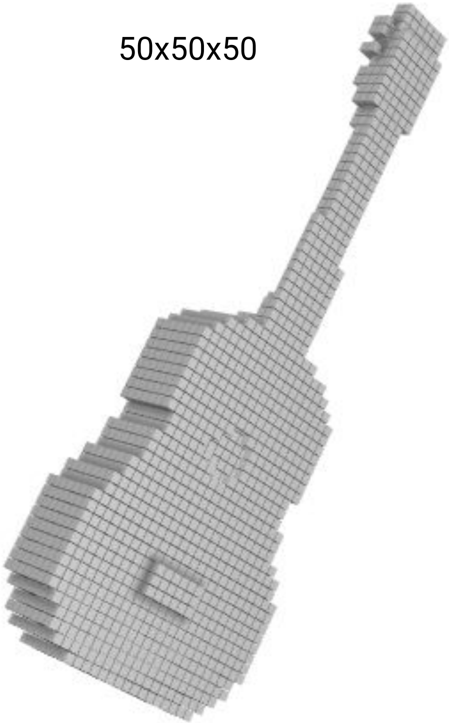
Volumetric representations

Decoder
D

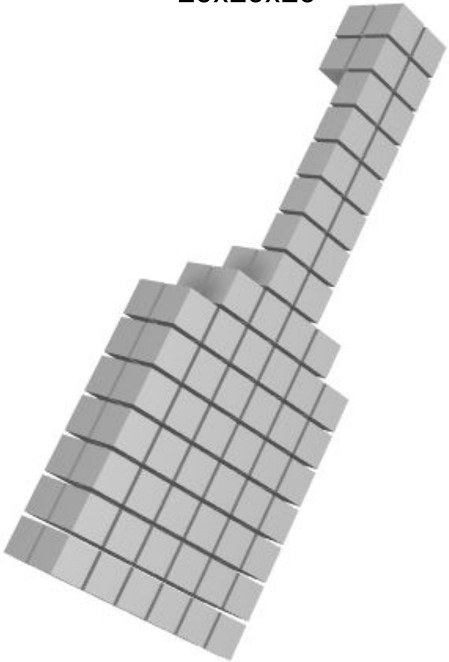
100x100x100



50x50x50



20x20x20



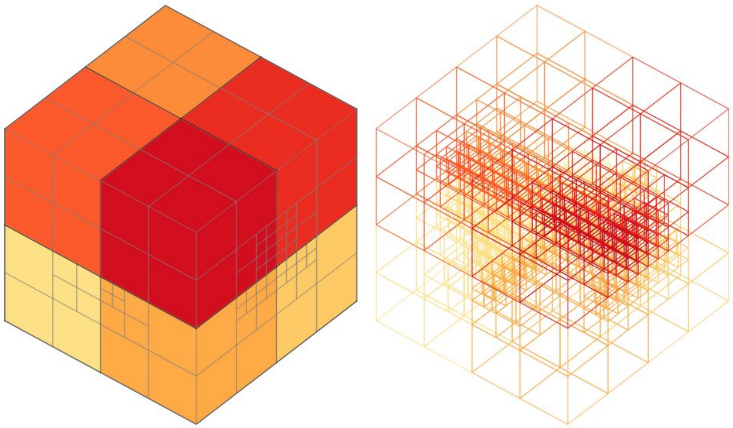
Hybrid Grid-Octree Data Structure

Encoder
E

Decoder
D

Octnet [**Riegler2017**], OGN [**Tatarchenko2017**]

- ➔ Grid of octrees with fixed small depth : typically 3
- ➔ Computationally more effective
- ➔ Good compression rate



Thibault Groueix, Pierre-Alain Langlois, 2018

OctNet input

- ➔ If a cell contains data from the mesh, it takes value 1 and it is subdivided
- ➔ Otherwise, it takes the value 0
- ➔ Easy to compare with the L2 distance over voxels

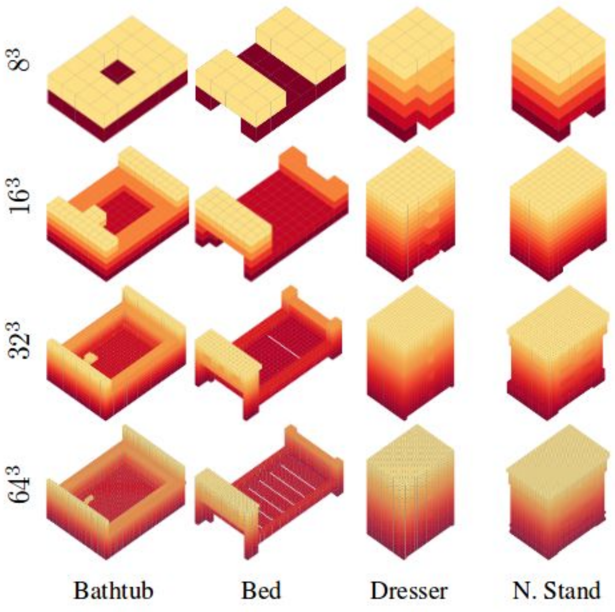


Figure 8: **Voxelized 3D Shapes from ModelNet10.**

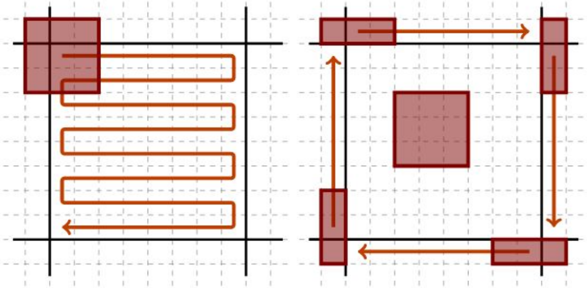
Convolutions on Grid-Octree Data Structure

Encoder
E

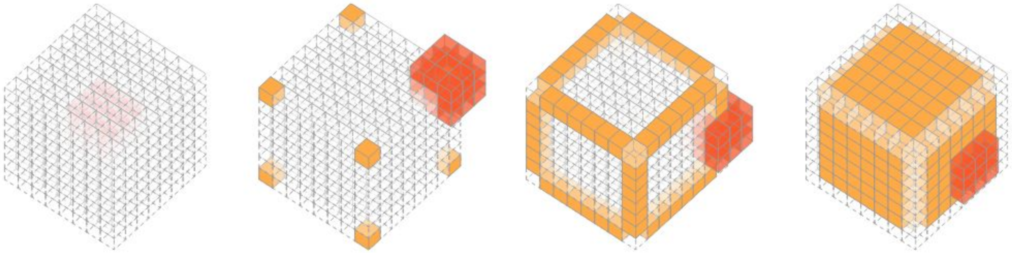
Decoder
D

➔ Improvement : Inside a given cell the convolution result is the same. We can compute it once.

➔ Convolution is computed on the boundaries



(a) Standard Convolution (b) Efficient Convolution



(a) Constant (b) Corners (c) Edges (d) Faces

Figure 14: **Efficient Convolution.**

Thibault Groueix, Pierre-Alain Langlois, 2018

Pooling on Grid-Octree Data Structure

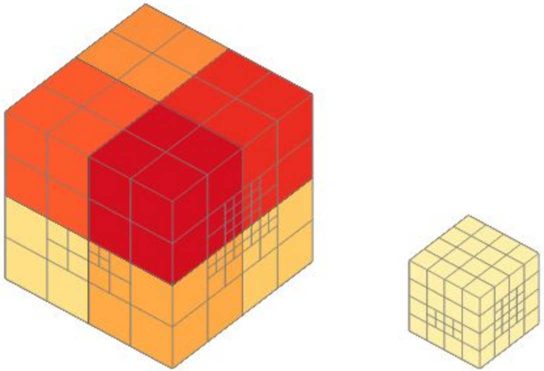
Encoder

E

➔ Combines 8 neighbouring shallow octrees(a) into one shallow octree (b)

➔ Voxels at level higher than *depth* are halved in size

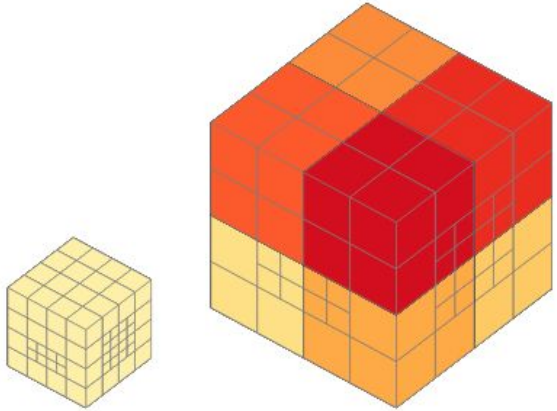
➔ Voxels at level *depth* in an octree are pulled



(a) Input (b) Output

Unpooling on Grid-Octree Data Structure

- ➔ Nearest neighbour interpolation
- ➔ Nodes at depth 0 spawn a new shallow tree (grid size is multiplied by 8)
- ➔ All other nodes double their sizes
- ➔ Need to know whether terminal voxels can be splitted in 8 to capture finer details **[Tatarchenko2017]**



(a) Input (b) Output

Octree generating networks - results

Decoder
D

OGN **[Tatarchenko2017]**

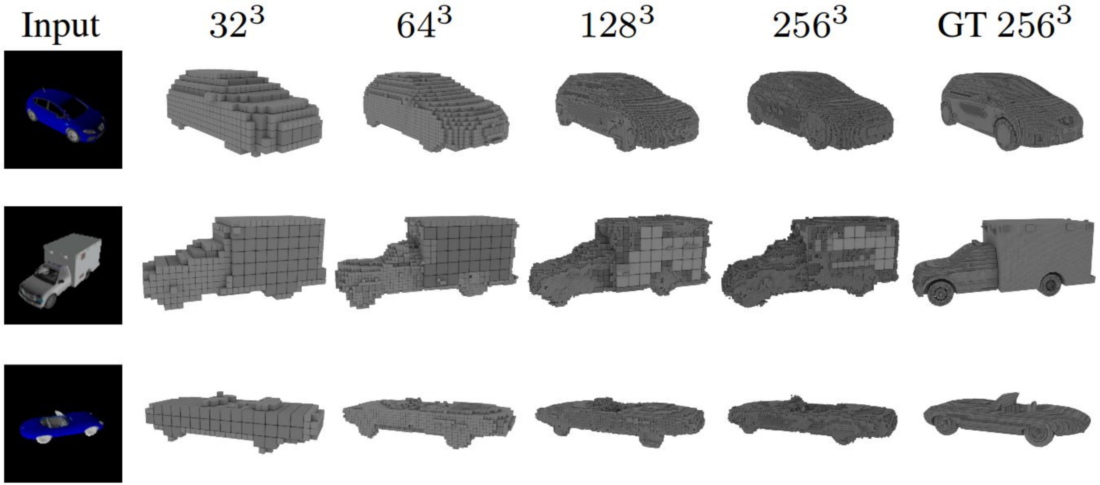


Figure 8. Single-image 3D reconstruction on the ShapeNet-cars dataset using OGN in different resolutions.

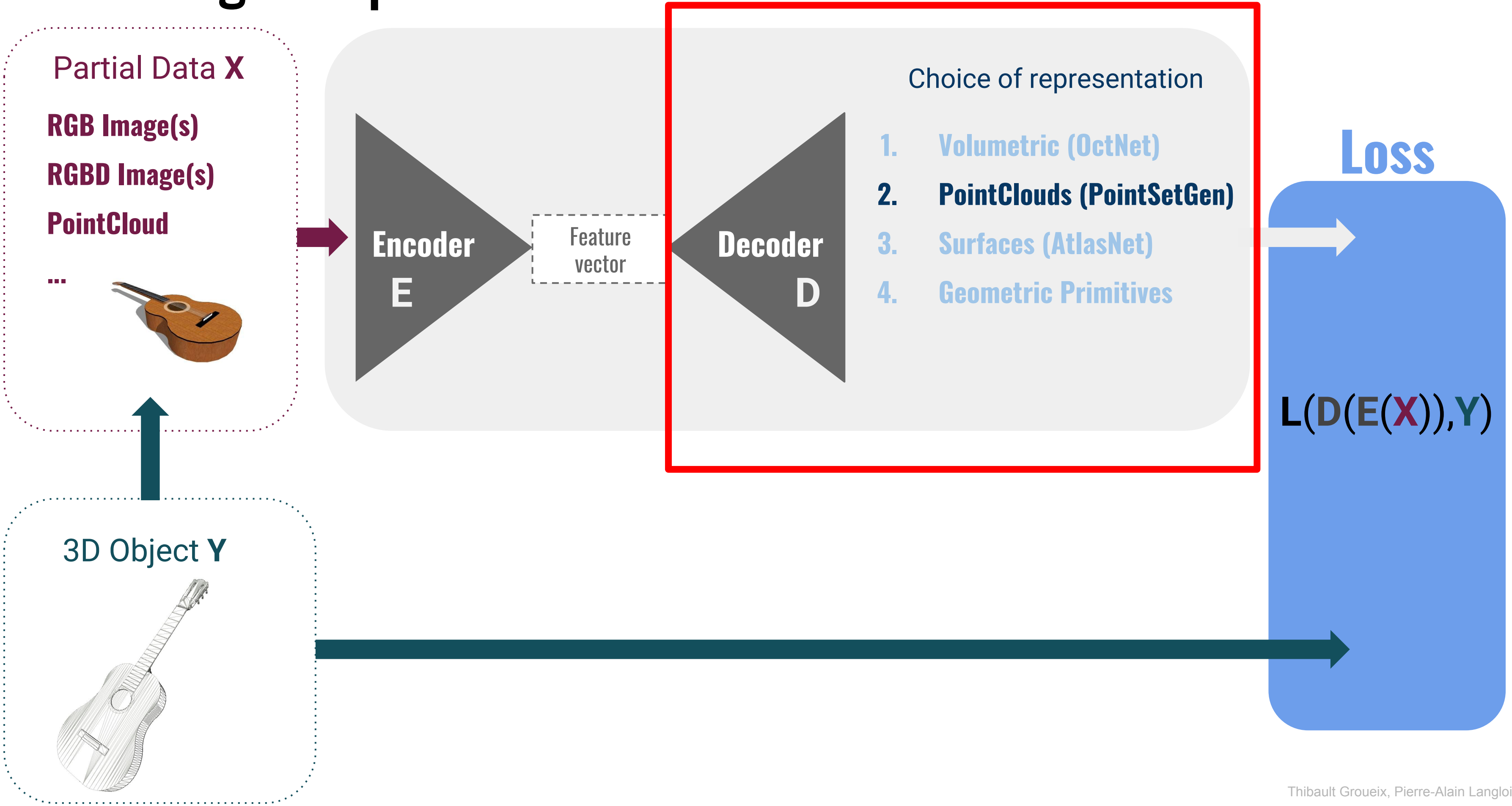
Octree-based reconstruction

Encoder
E

Decoder
D

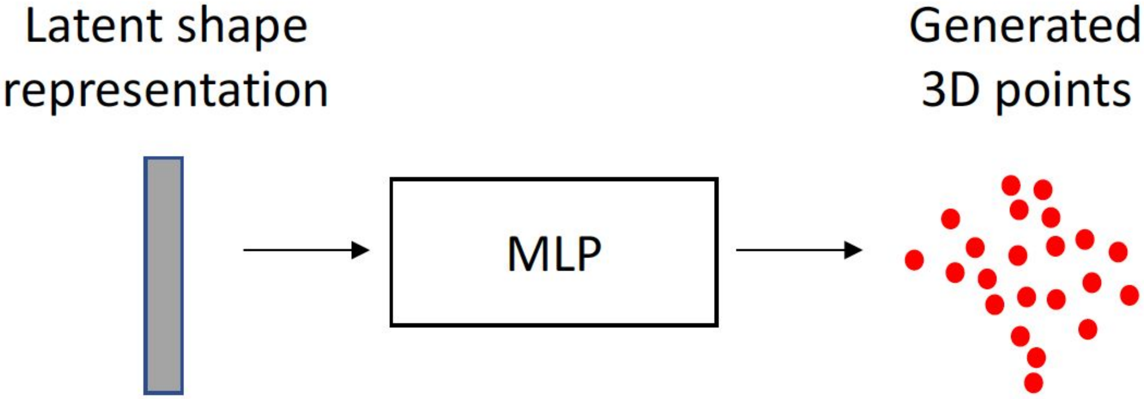
- ➔ Gives insights regarding the extension of network operations to 3D data structures
- ➔ Important improvement in the fight again the curse of dimensionality
- ➔ Gives quantitative results regarding the **need for higher resolutions**

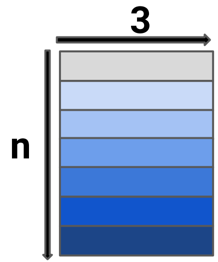
Training setup for 3D reconstruction



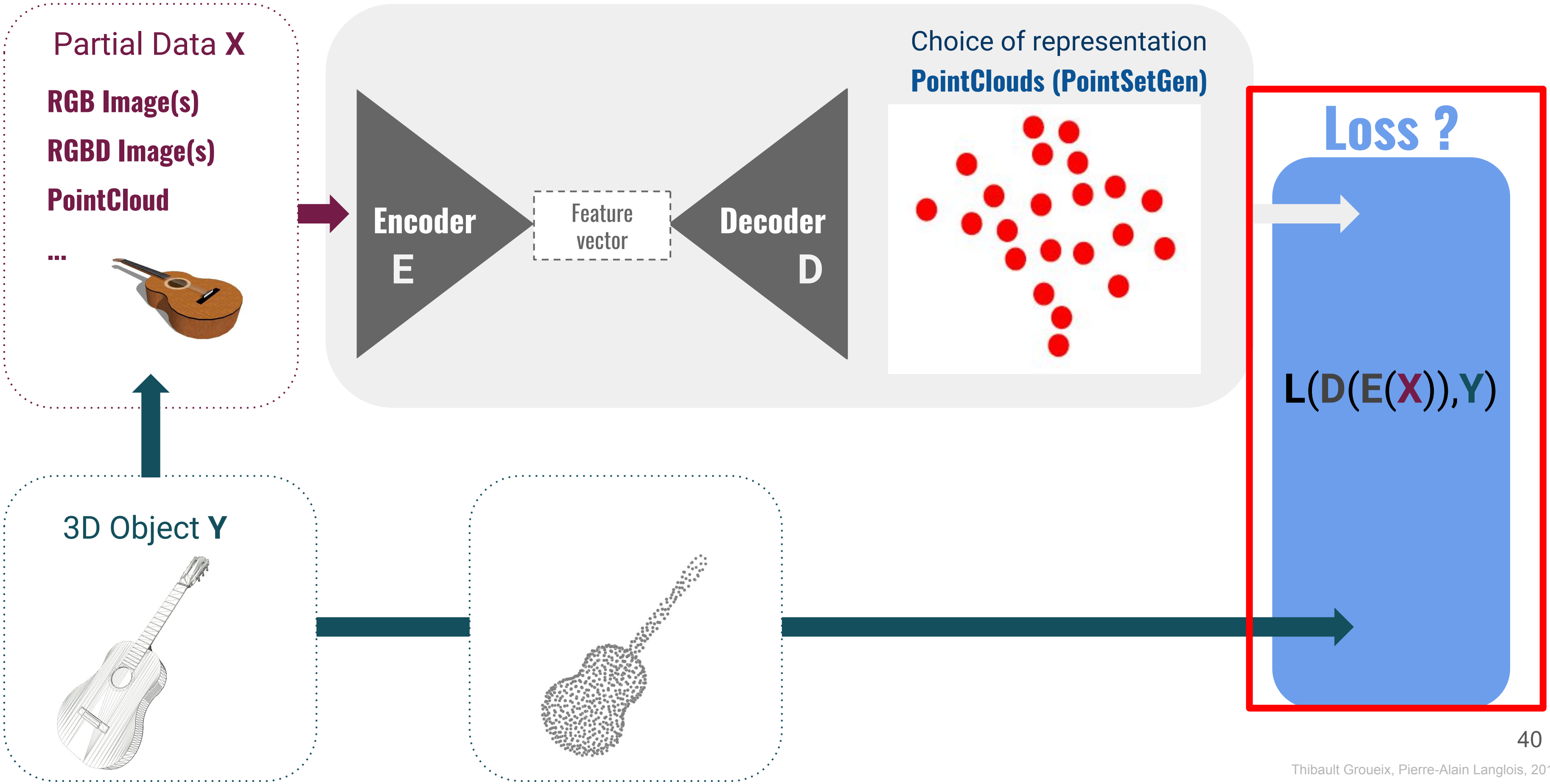
Thibault Groueix, Pierre-Alain Langlois, 2018

Generating points PointSetGen[**Fan2017**]



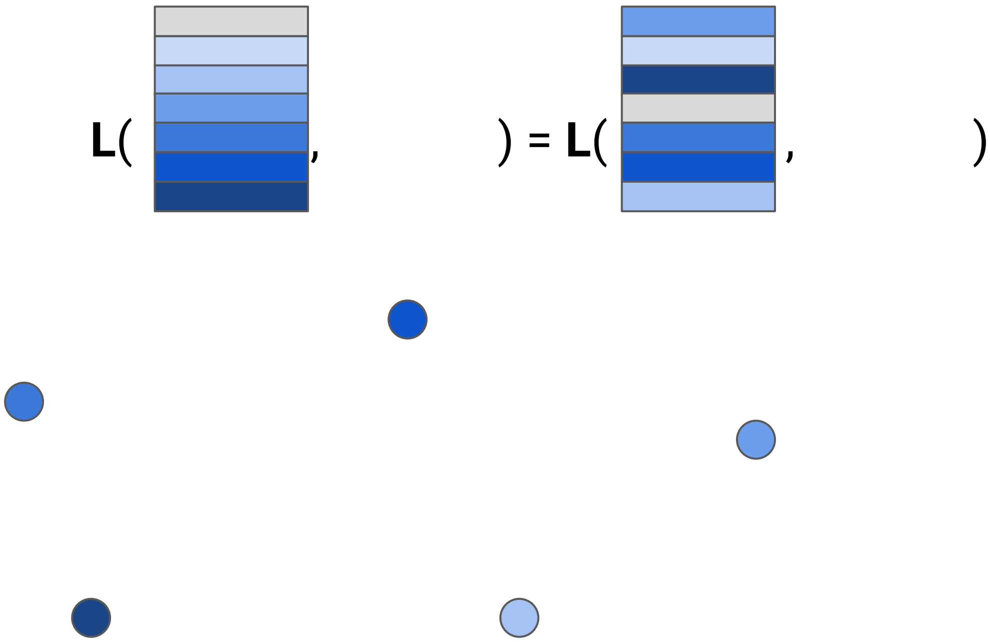


Training setup for 3D reconstruction



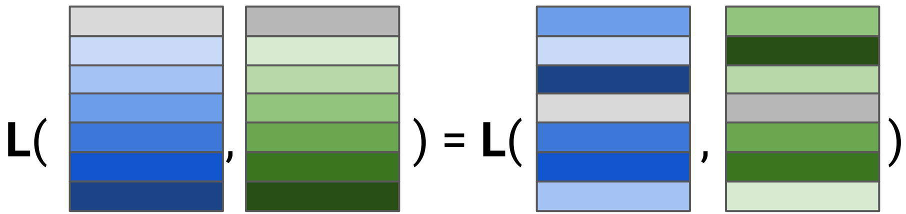
Thibault Groueix, Pierre-Alain Langlois, 2018

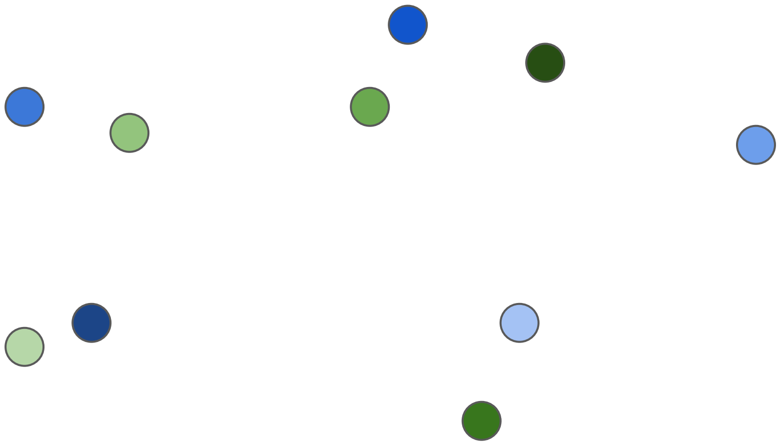
Loss on pointclouds



	Complexity
EMD	n^3
Chamfer	n^2

Loss on pointclouds



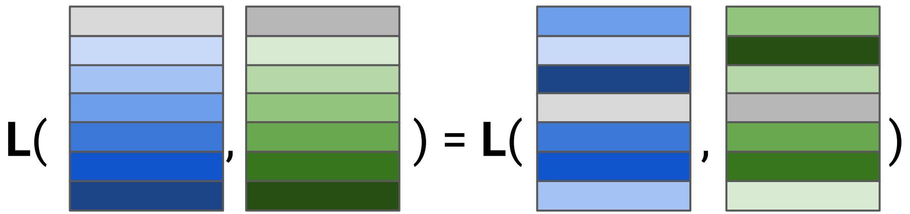


	Complexity
EMD	n^3
Chamfer	n^2

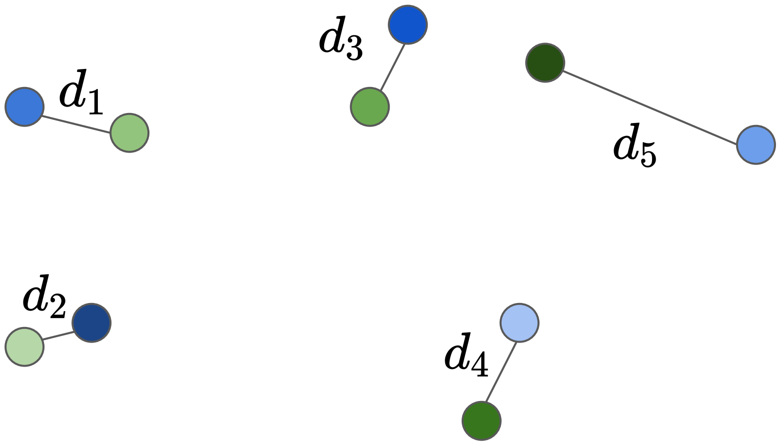
Loss on pointclouds

Find the **optimal assignement** and compute **Earth Mover Distance** (**EMD**)

- ➔ Hungarian Algorithm **[Kuhn1955]** ~O(n^3)
- ➔ Simplex based solver through LP formulation ~O(Hungarian)
- ➔ Sinkhorn regularization [**Cuturi2013**] in near linear time [**Altschuler2017**]
- ➔ (1+ε) approximation [**Bertsekas1988**] in ~O(n^3)



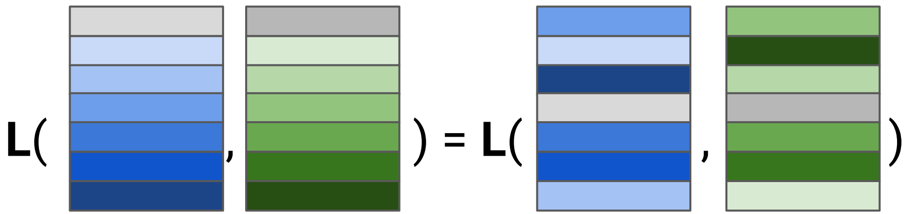
$$= \frac{1}{5} \cdot (d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$



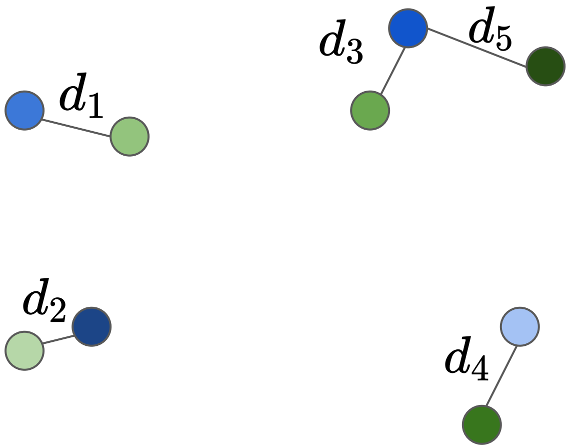
	Complexity
EMD	**n^3**
Chamfer	**n^2**

Loss on pointclouds

Find the **nearest neighbours** and compute
Chamfer Distance (CD) = L(,) +



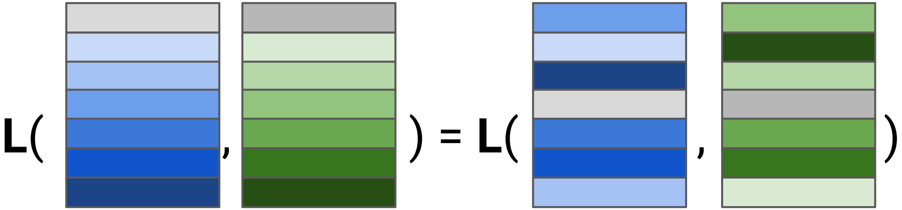
$$= \frac{1}{5} \cdot (d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$



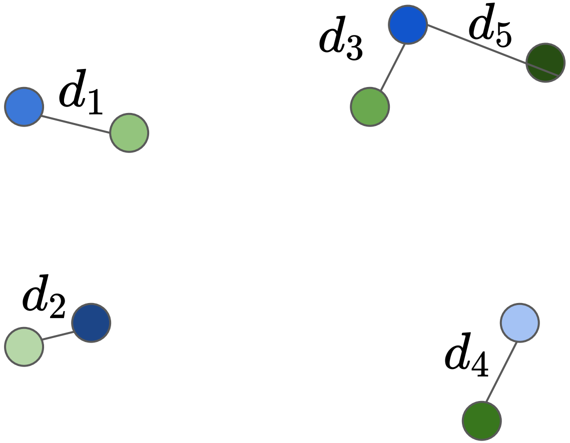
	Complexity
EMD	**n³**
Chamfer	**n²**

Loss on pointclouds

Find the **nearest neighbours** and compute
Chamfer Distance (CD) = L(,) +



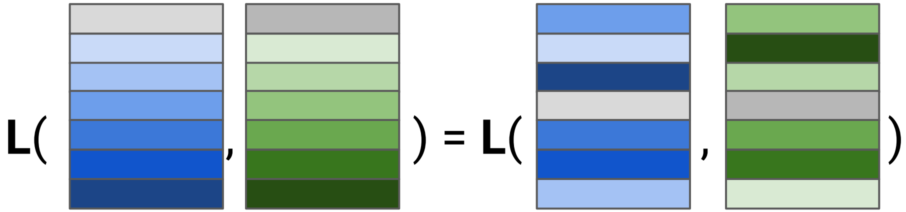
$$= \frac{1}{5} \cdot (d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$



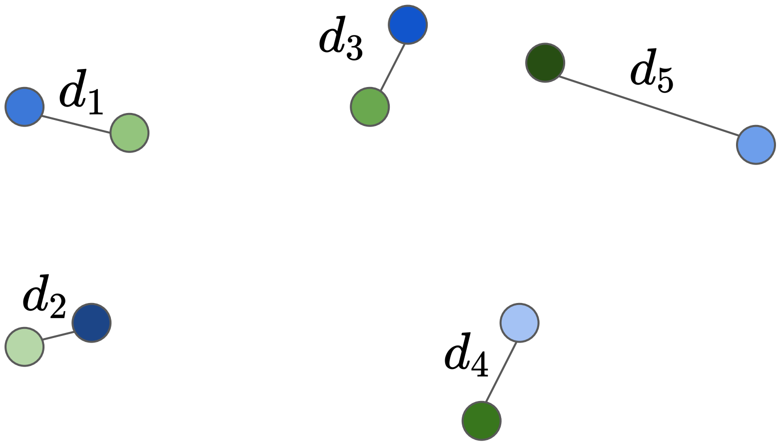
	Complexity
EMD	**n³**
Chamfer	**n²**

Loss on pointclouds

Find the **nearest neighbours** and compute
Chamfer Distance (CD) = L(,) + L(,)



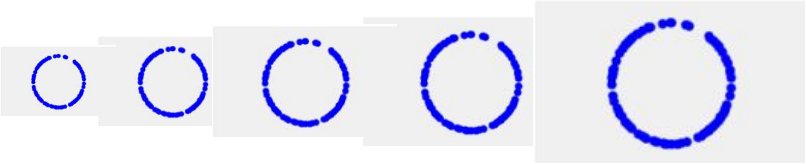
$$= \frac{1}{5} \cdot (d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)$$



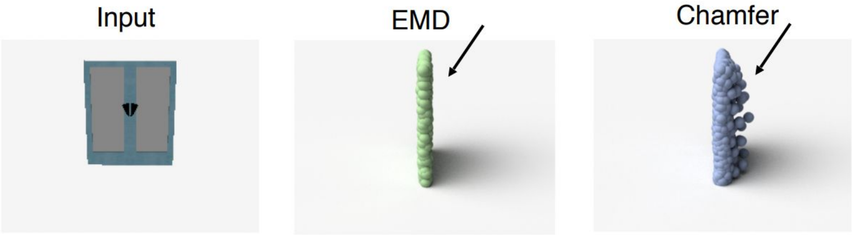
	Complexity
EMD	n^3
Chamfer	n^2

Loss on pointclouds : the mean shape carries characteristics of the distance metric

Distribution $\mathbb{S}$ of pointclouds of varying radius

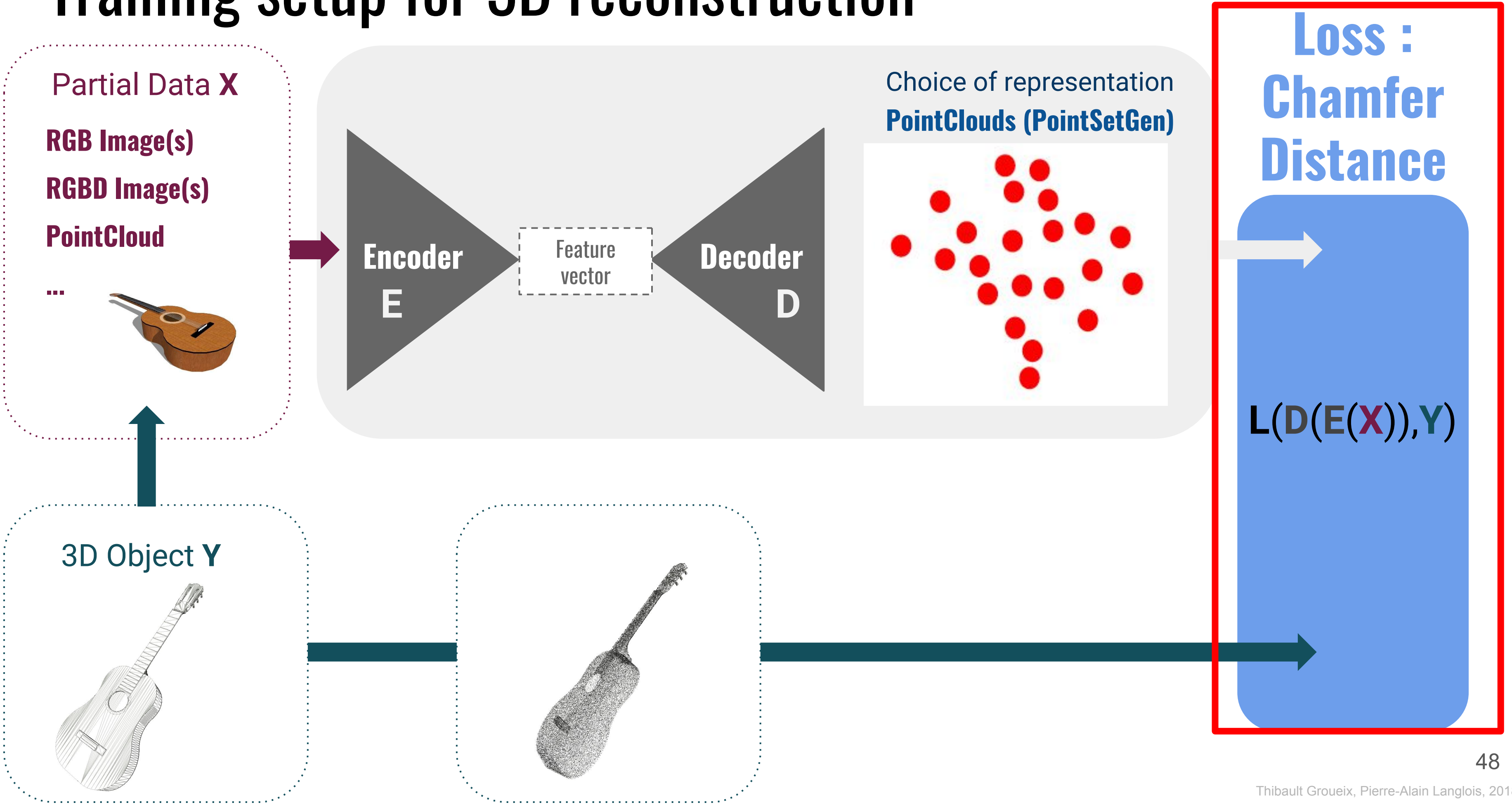


$$\bar{x} = \underset{x}{\operatorname{argmin}}\ \mathbb{E}_{s \sim \mathbb{S}}[d(x, s)]$$



Credit : [**Fan2016**]

Training setup for 3D reconstruction



Thibault Groueix, Pierre-Alain Langlois, 2018

Generating points

Encoder
E

Decoder
D

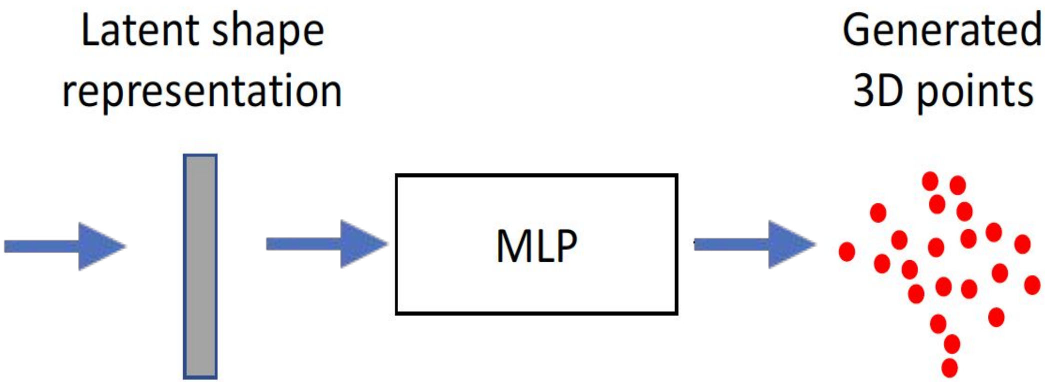
Test Shape

Generating points

Encoder
E

Decoder
D

Test Shape

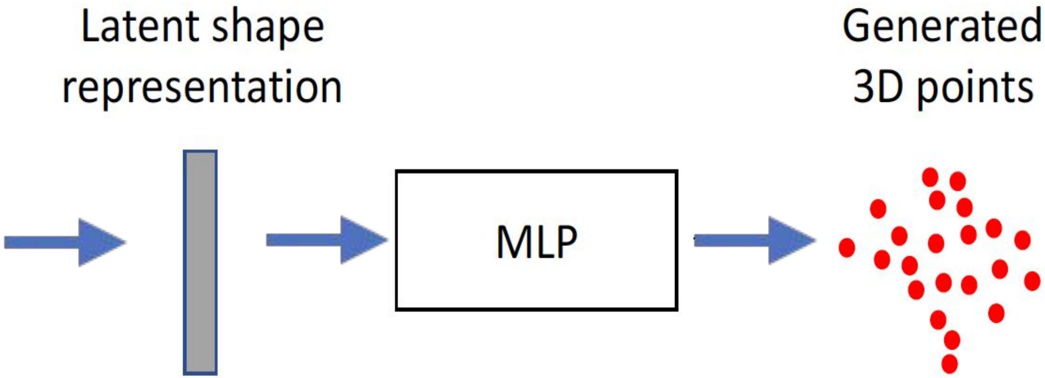


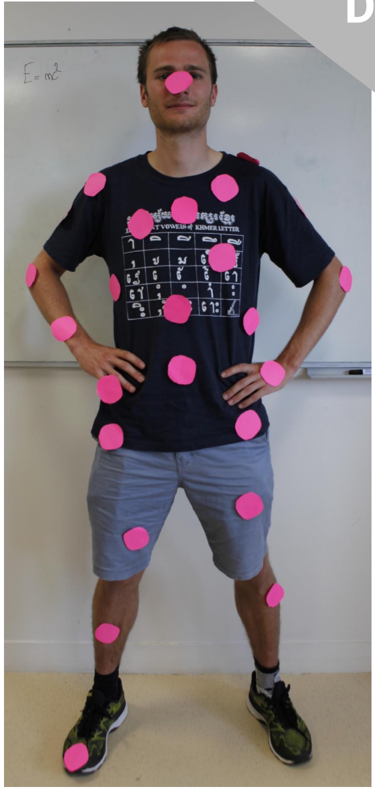
Generating points

Encoder
E

Decoder
D

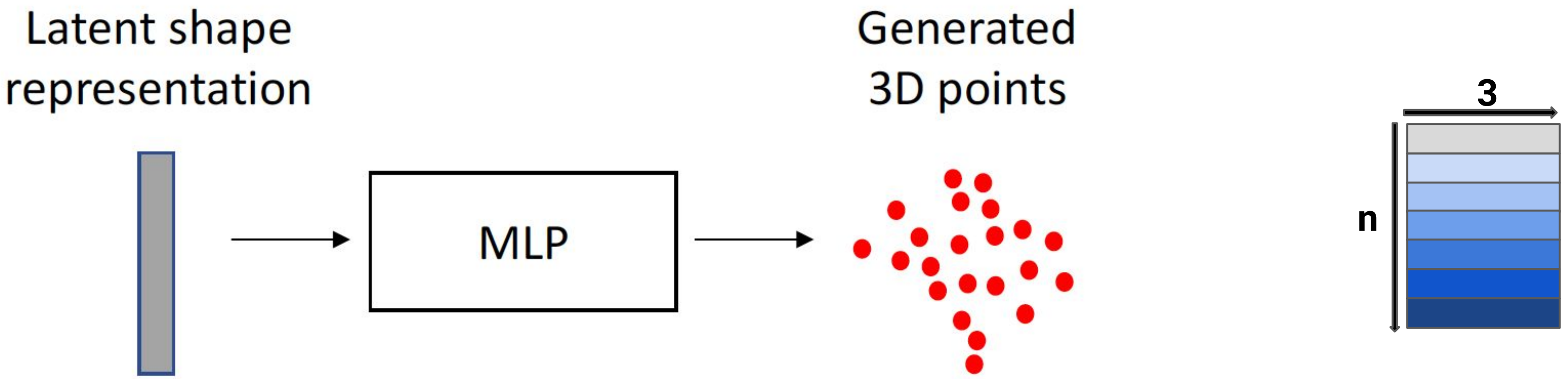
Test Shape





Limitation of PointSetGen [Fan2017]

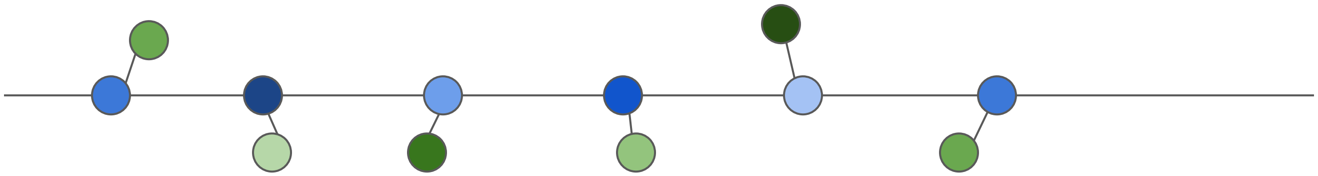
- ➔ **Generate a fixed number of points**
- ➔ Points connectivity is missing
- ➔ Generated points are not correlated enough to belong to an implicit surface



Limitation of PointSetGen [Fan2017]

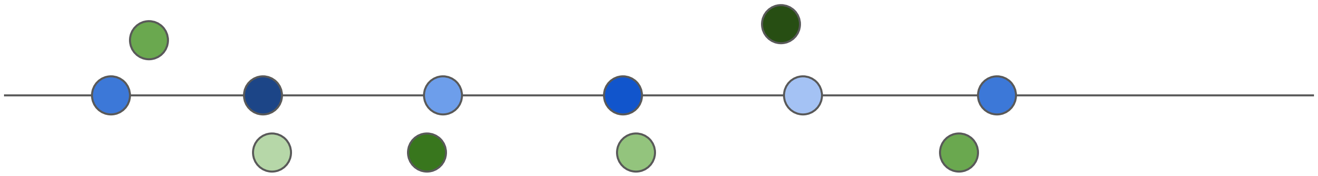
Limitation of PointSetGen [Fan2017]

- ➔ Generate a fixed number of points
- ➔ **Points connectivity is missing**
- ➔ Generated points are not correlated enough to belong to an implicit surface



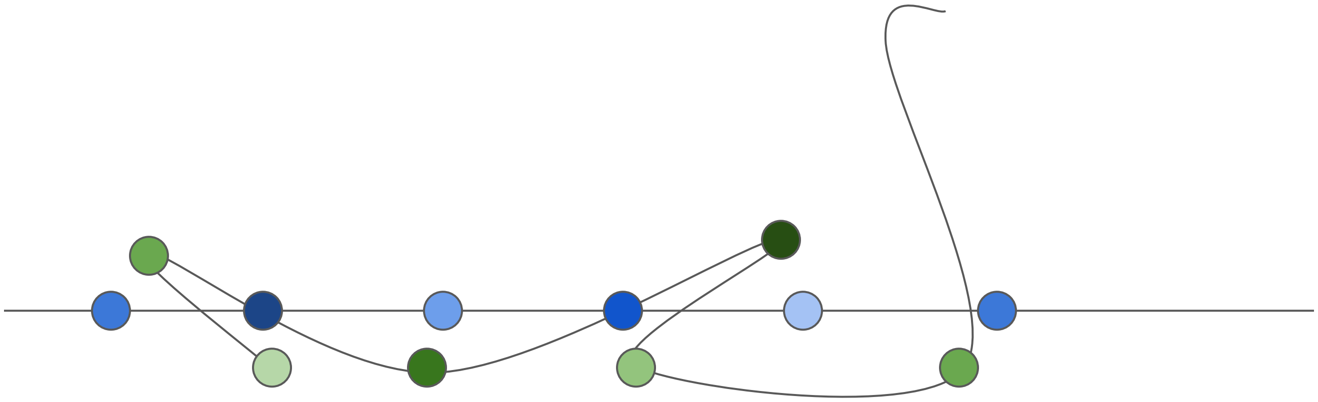
Limitation of PointSetGen [Fan2017]

- ➔ Generate a fixed number of points
- ➔ **Points connectivity is missing**
- ➔ Generated points are not correlated enough to belong to an implicit surface



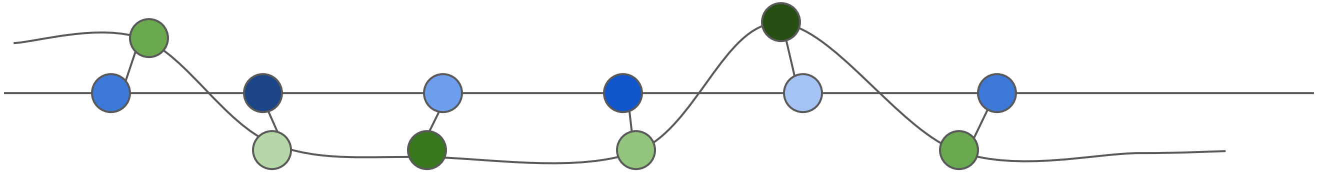
Limitation of PointSetGen [Fan2017]

- ➔ Generate a fixed number of points
- ➔ **Points connectivity is missing**
- ➔ Generated points are not correlated enough to belong to an implicit surface



Limitation of PointSetGen [Fan2017]

- ➔ Generate a fixed number of points
- ➔ Points connectivity is missing
- ➔ **Generated points are not correlated enough to belong to an implicit surface**



Reconstructing the mesh from a pointcloud :
Poisson Surface Reconstruction [**Kazhdan2013**]

Training setup for 3D reconstruction

Partial Data **X**

RGB Image(s)

RGBD Image(s)

PointCloud

...

3D Object **Y**

Encoder
E

Feature vector

Decoder
D

Choice of representation

1. **Volumetric (OctNet)**
2. **PointClouds (PointSetGen)**
3. **Surfaces (AtlasNet)**
4. **Geometric Primitives**

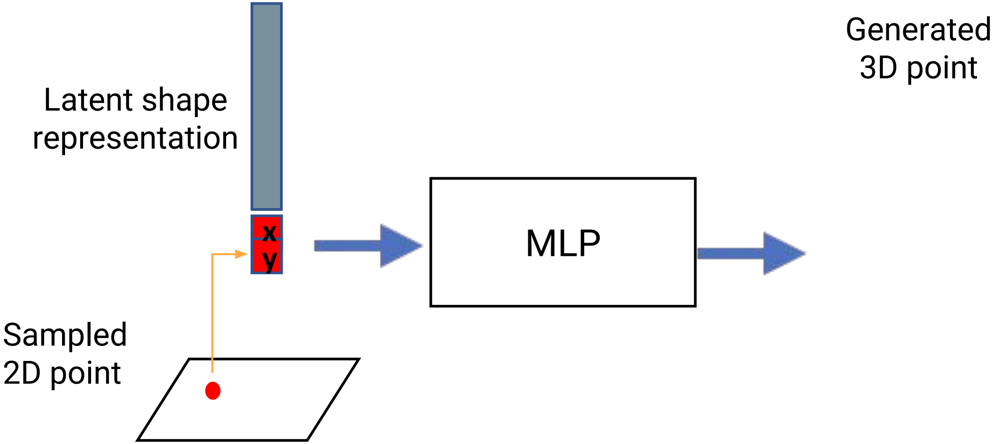
Loss

L(D(E(X)),Y)

Thibault Groueix, Pierre-Alain Langlois, 2018

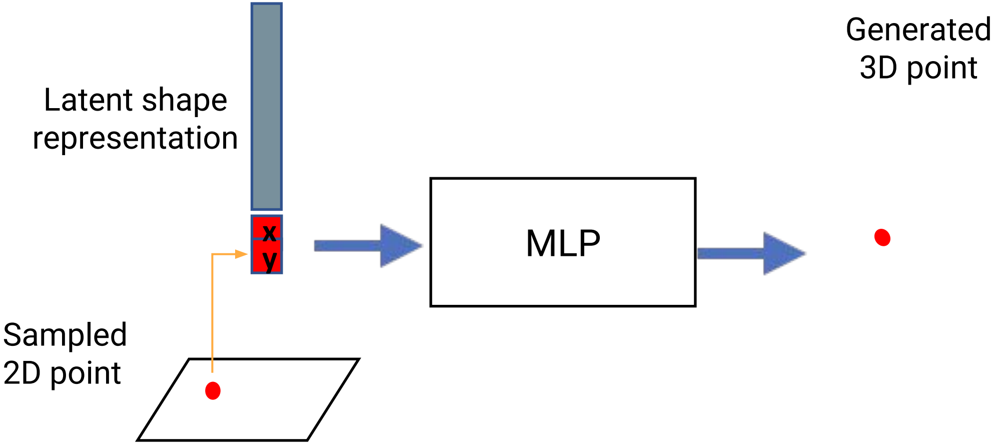
Deform a surface [Groueix2018]

Decoder
D



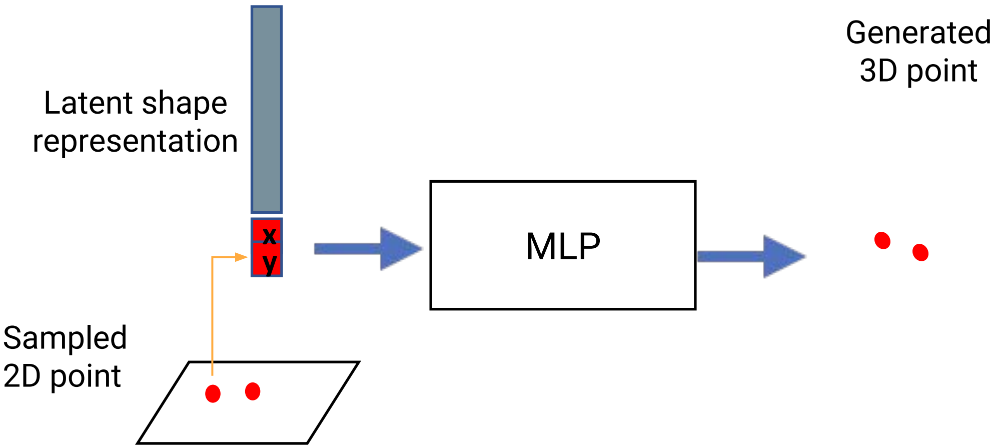
Deform a surface [Groueix2018]

Decoder D



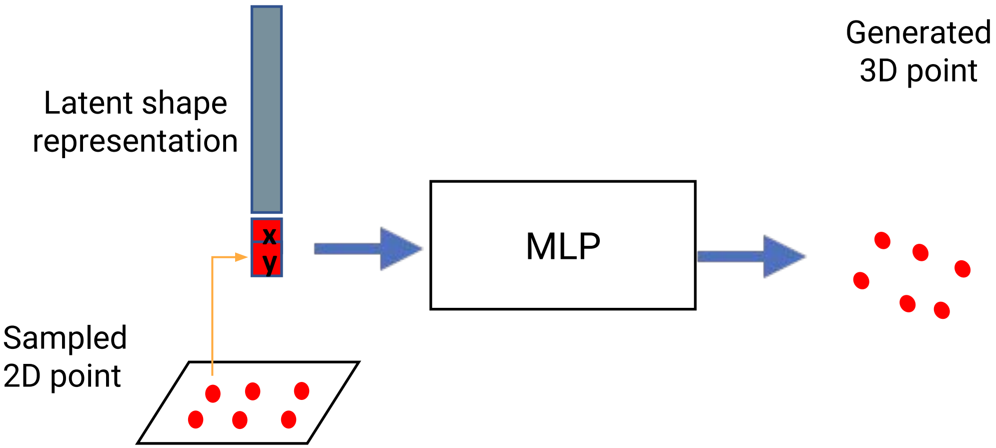
Deform a surface [Groueix2018]

Decoder D

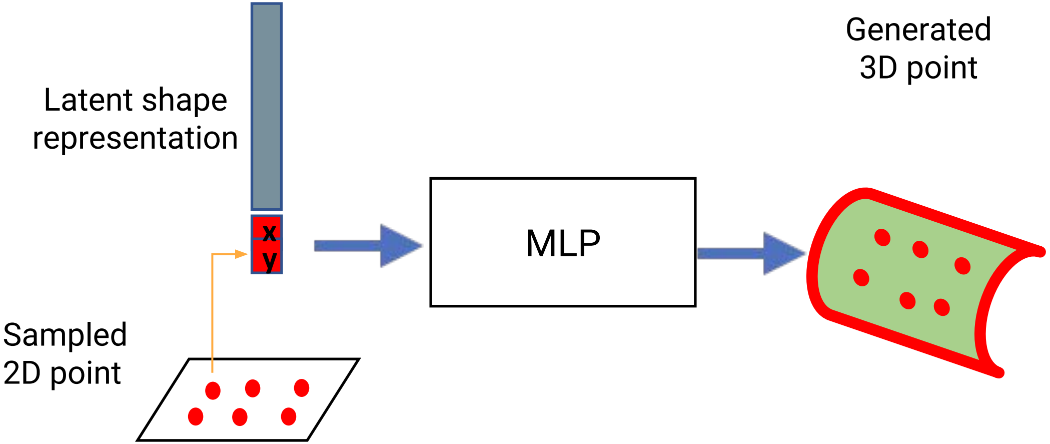


Deform a surface [Groueix2018]

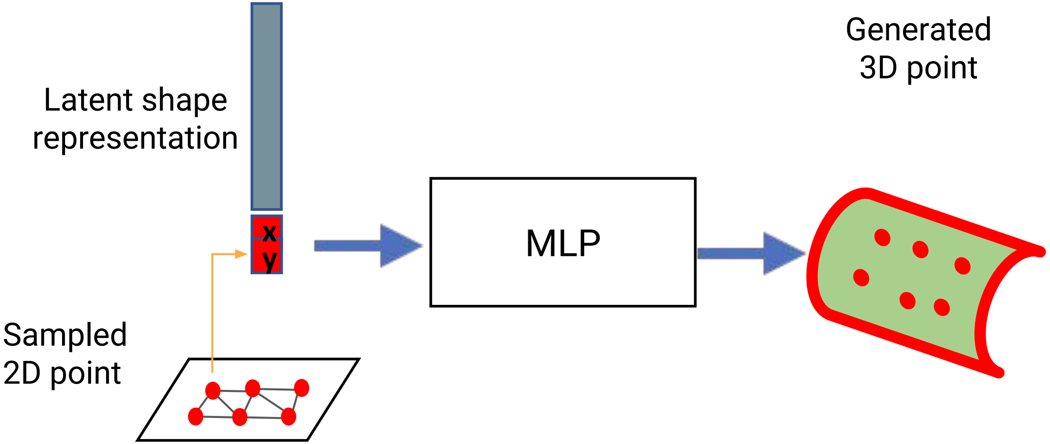
Decoder
D



Deform a surface [Groueix2018]

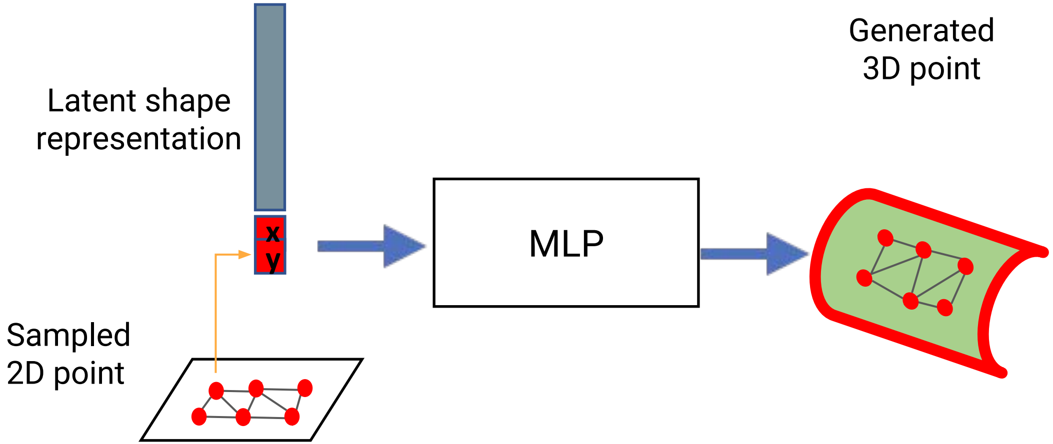


Deform a surface [Groueix2018]



Deform a surface [Groueix2018]

Decoder D

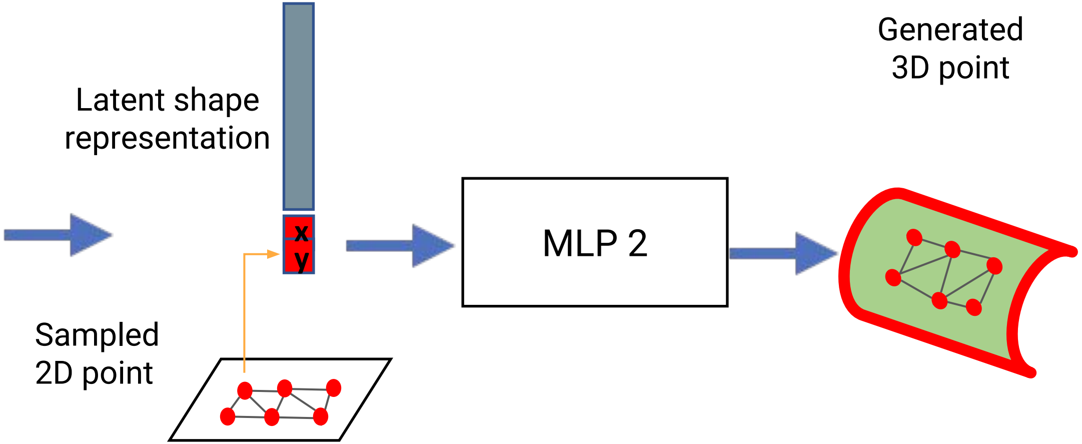


Deform a surface [Groueix2018]

Encoder
E

Decoder
D

Test Shape



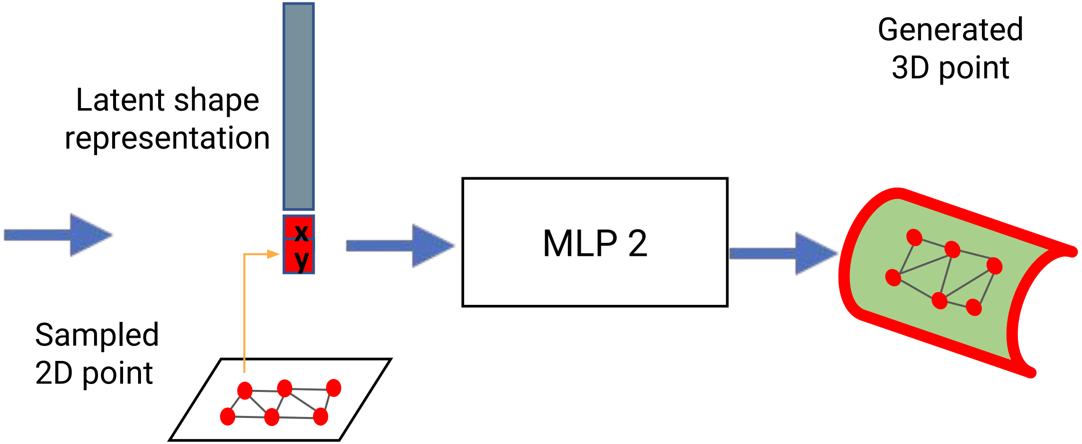
Thibault Groueix, Pierre-Alain Langlois, 2018

Deform a surface [Groueix2018]

Encoder
E

Decoder
D

Test Shape

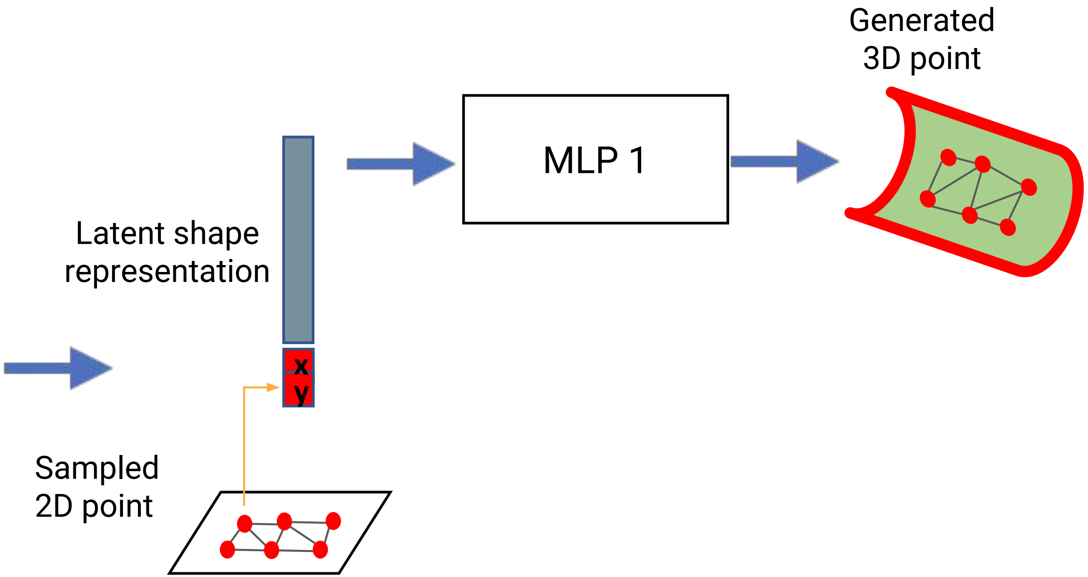


Deform a surface [Groueix2018]

Encoder
E

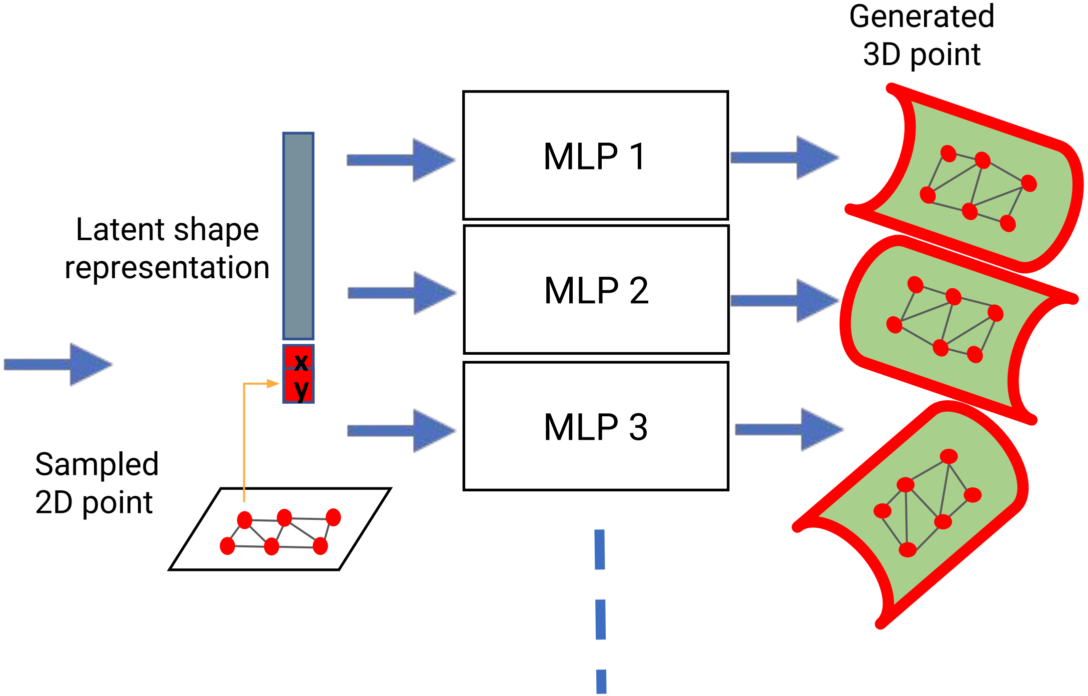
Decoder
D

Test Shape



Deform a surface [Groueix2018]

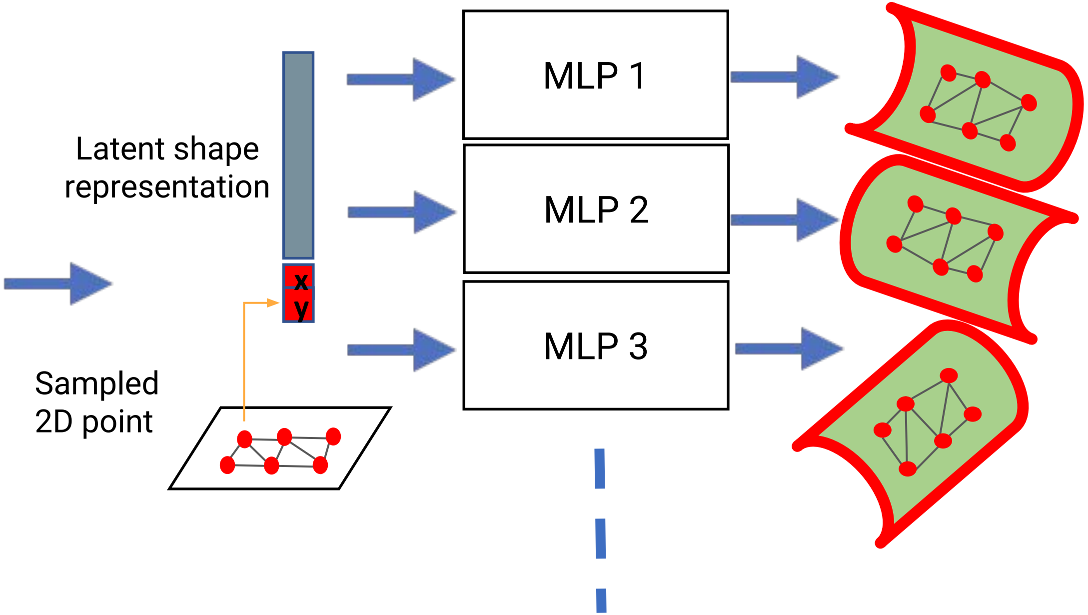
Test Shape



Deform a surface [Groueix2018]

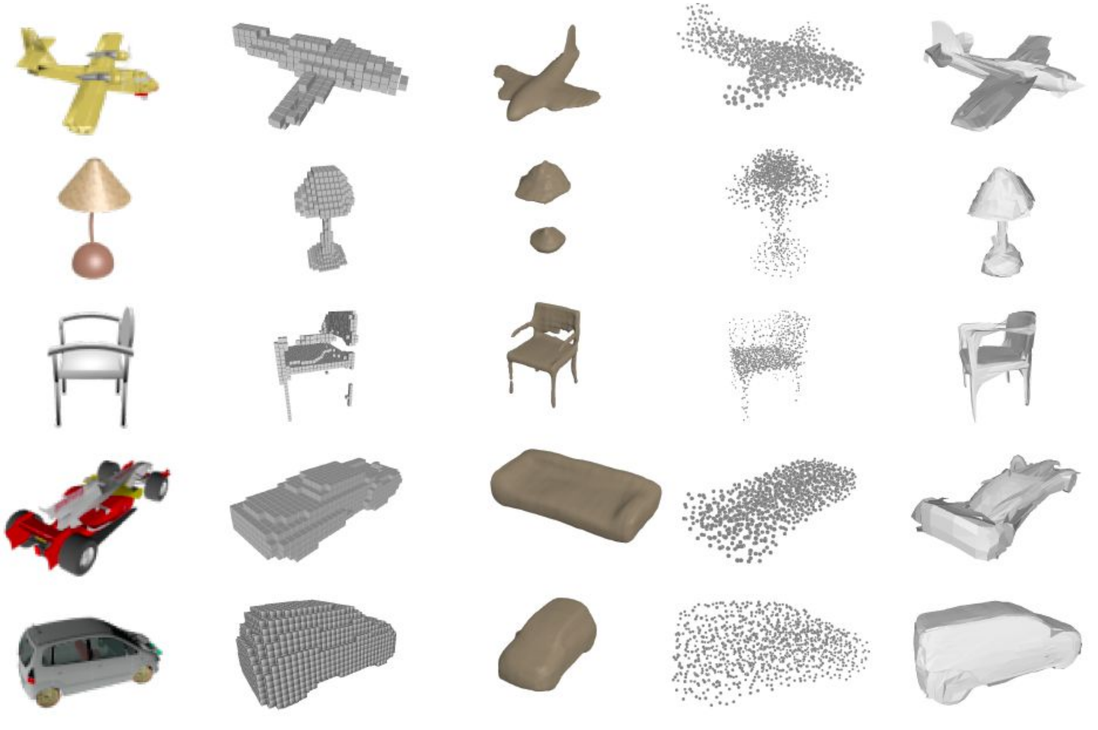
Encoder
E

Test Shape



Decoder
D

Results : Single View Reconstruction



(a) Input (b) 3D-R2N2 (c) HSP (d) PSG (e) Ours

Direct application : mesh parametrization

State-of-the-art correspondences of FAUST [Groueix2018b]

Limitations of learned approaches

- ➔ Hard to add geometric constraints in the design of a neural net architecture e.g. Watertight reconstruction. cf http://imagine.enpc.fr/~groueixt/atlasnet/viewer-svr/

- ➔ Hard to scale to large scenes and/or very high level of details.

- ➔ Biased by data
- ➔ ...

What was not covered today

Traditional methods : Shape from X

Graph Based methods : Spectral and spatial methods

Equivariant methods : SphericalCNNs

Other Point Based Methods : PointNet++, PCPNet

Differential rendering for inverse graphics : Neural renderer, rendernet

Geometric primitives : Shape Abstraction, Supervised Fitting of Geometric Primitives to 3D Point Clouds

Making it work on real sensor data : domain adaptation, data augmentation

Multiple sources fusion through TSDF : **[Riegler2017]**

The choice of representation of 3D data is critical

We journeyed from **Volumes**...,

... through **Pointclouds**...,

to **Surfaces**.

Thank you

Bibliography

- **Resnet [He2015]** : He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- **PointNet [Qi2017] :** Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." Proc. Computer Vision and Pattern Recognition (CVPR), IEEE 1.2 (2017): 4.
- **3d-r2n2 [Choy2016]** : C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In Proc. of the European Conf. on Computer Vision (ECCV), 2016.
- **Voxnet [Maturana2015]** : D. Maturana and S. Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS), 2015.
- **[Qi2016]** : C. R. Qi, H. Su, M. Nießner, A. Dai, M. Yan, and L. Guibas. Volumetric and multi-view cnns for object classification on 3d data. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016.
- **[Wu2015]** : Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2015.
- **Octnet [Riegler2017] :** Riegler, Gernot, Ali Osman Ulusoy, and Andreas Geiger. "Octnet: Learning deep 3d representations at high resolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Vol. 3. 2017.
- **OGN [Tatarchenko2017]** : Tatarchenko, Maxim, Alexey Dosovitskiy, and Thomas Brox. "Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs." Proc. of the IEEE International Conf. on Computer Vision. Vol. 2. 2017.
- **PointSetGen [Fan2017]** : Fan, Haoqiang, Hao Su, and Leonidas J. Guibas. "A Point Set Generation Network for 3D Object Reconstruction from a Single Image." CVPR. Vol. 2. No. 4. 2017.
- **[Kuhn1955]** : Kuhn, Harold W. "The Hungarian method for the assignment problem." 50 Years of Integer Programming 1958-2008. Springer, Berlin, Heidelberg, 2010. 29-47.
- **[Cuturi2013]** : Cuturi, Marco. "Sinkhorn distances: Lightspeed computation of optimal transport." Advances in neural information processing systems. 2013.
- **[Altschuler2017]** : Altschuler, Jason, Jonathan Weed, and Philippe Rigollet. "Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration." Advances in Neural Information Processing Systems. 2017.
- **[Bertsekas1988]** : Bertsekas, Dimitri P. "The auction algorithm: A distributed relaxation method for the assignment problem." Annals of operations research 14.1 (1988): 105-123.
- **[Delanoy2017]** : Delanoy, Johanna, et al. "What you sketch is what you get: 3D sketching using multi-view deep volumetric prediction." arXiv preprint arXiv:1707.08390 (2017).

Bibliography

- **[Groueix2018]** : Groueix, T., Fisher, M., Kim, V., Russell, B., and Aubry, M. (2018, June). AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In CVPR 2018.
- **[Furukawa2009]** : Furukawa, Yasutaka, et al. "Manhattan-world stereo." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
- [**Klokov2017**] : Klokov, Roman, and Victor Lempitsky. "Escape from cells: Deep kd-networks for the recognition of 3d point cloud models." *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017.
- [**Qi2017b**] : Qi, Charles Ruizhongtai, et al. "Pointnet++: Deep hierarchical feature learning on point sets in a metric space." *Advances in Neural Information Processing Systems*. 2017.
- [**Guerrero2017**]: Guerrero, Paul, et al. "PCPNet Learning Local Shape Properties from Raw Point Clouds." *Computer Graphics Forum*. Vol. 37. No. 2. 2018.
- [**Landrieu2018**] : Landrieu, Loic, and Martin Simonovsky. "Large-scale point cloud semantic segmentation with superpoint graphs." *arXiv preprint arXiv:1711.09869* (2017).
- [**Yi2016**] : Yi, Li, et al. "SyncSpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation." *CVPR*. 2017.
- [**Esteves2018**] : Esteves, Carlos, et al. "Learning so (3) equivariant representations with spherical cnns." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
- [**Cohen2018**] Cohen, Taco S., et al. "Spherical CNNs." *arXiv preprint arXiv:1801.10130* (2018).
- [**Groueix2018b**] Groueix, Thibault, et al. "3D-CODED: 3D Correspondences by Deep Deformation." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- **[Riegler2017]**Riegler, G., Ulusoy, A. O., Bischof, H., & Geiger, A. (2017, October). Octnetfusion: Learning depth fusion from data. In 3D Vision (3DV), 2017 International Conference on (pp. 57-66). IEEE.

Additional Material

PointClouds Analysis Motivation

Source : http://graphics.stanford.edu/courses/cs468-17-spring/schedule.html

- **Robot Perception**

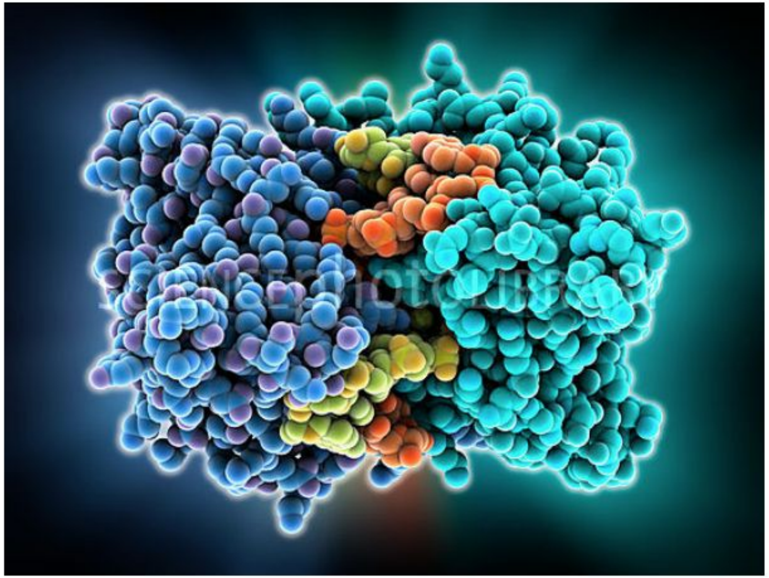
What and where are the objects in a LiDAR scanned scene?

https://3dprint.com/116569/self-driving-cars-privacy/

PointClouds Analysis Motivation

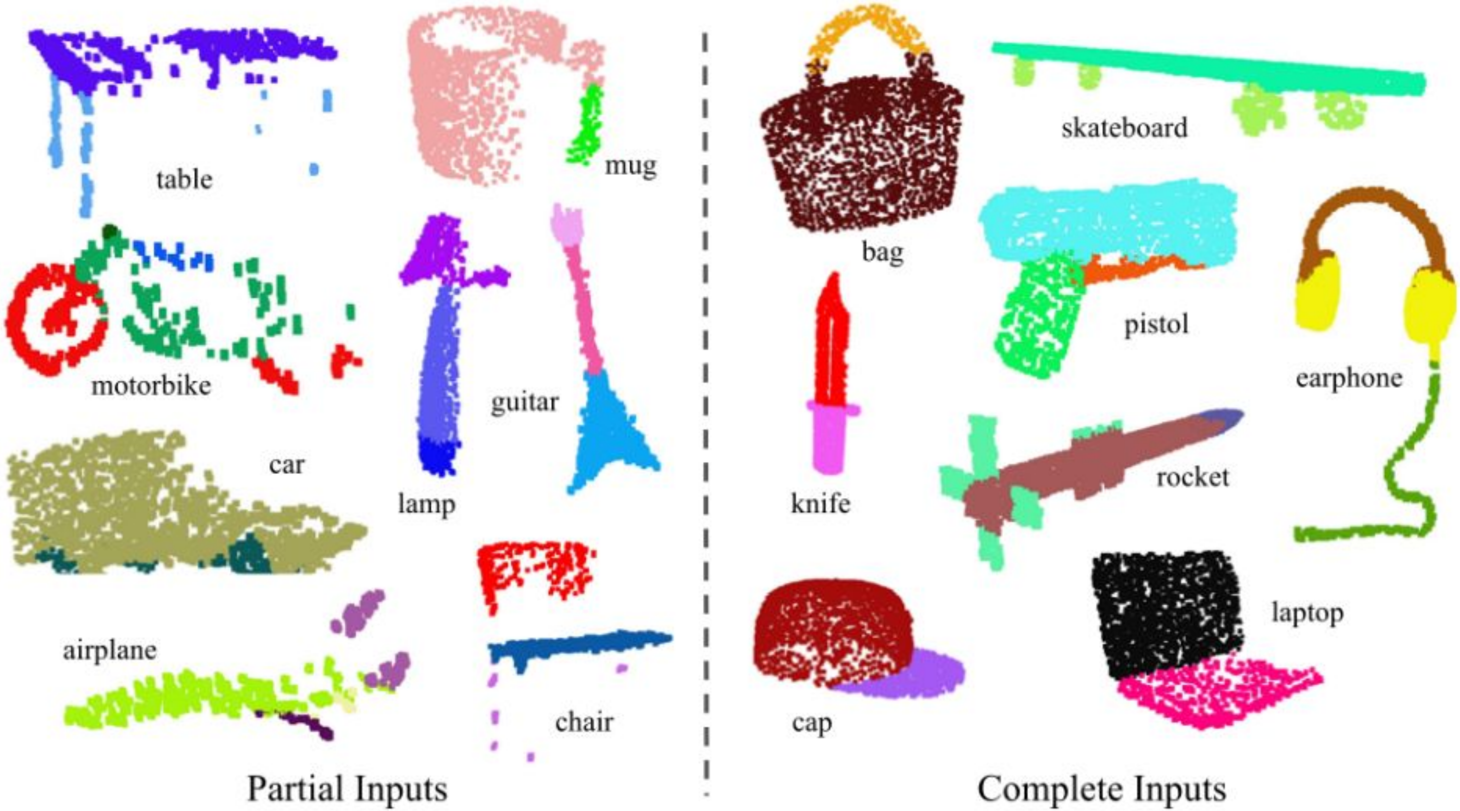
- **Molecular Biology**

Can we infer an enzyme's category (reactions they catalyze) from its structure?



EcoRV restriction enzyme molecule, LAGUNA DESIGN/SCIENCE PHOTO LIBRARY

PointNet Results : object semantic segmentation
Credit [Qi2017]

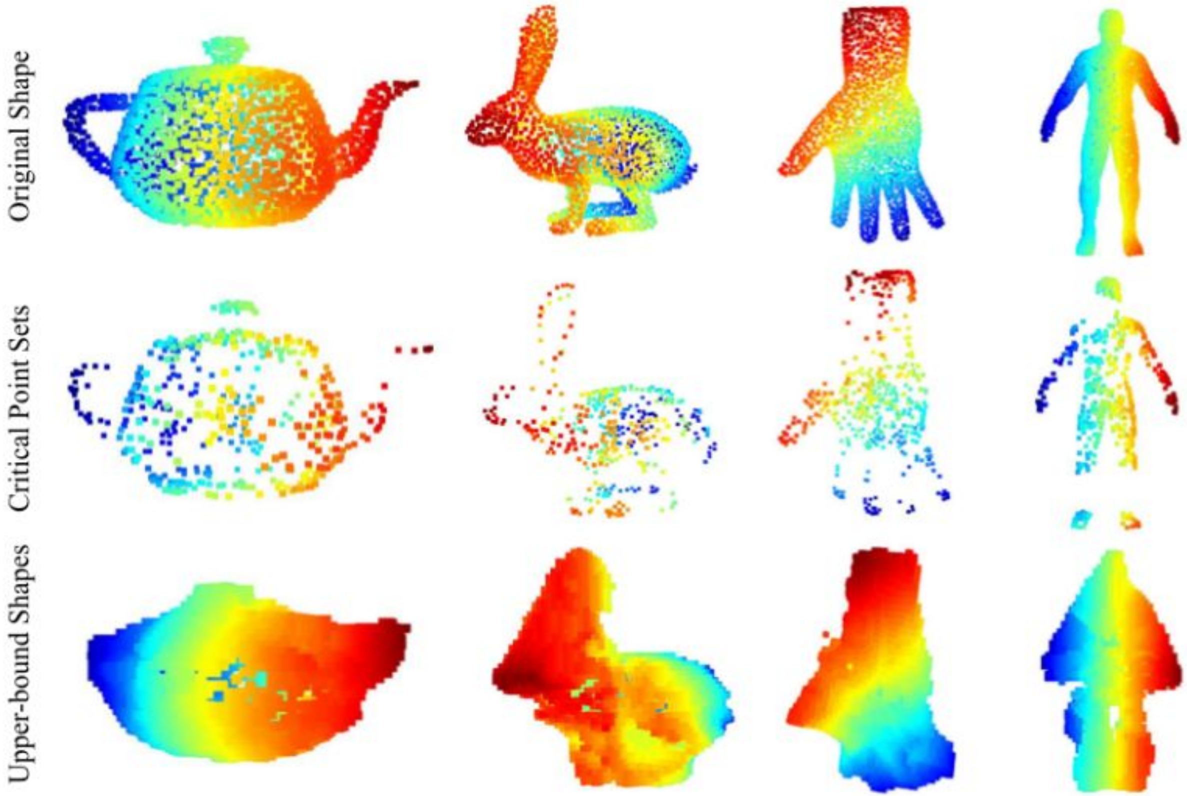


PointNet Results : scene semantic segmentation Credit [Qi2017]

)18

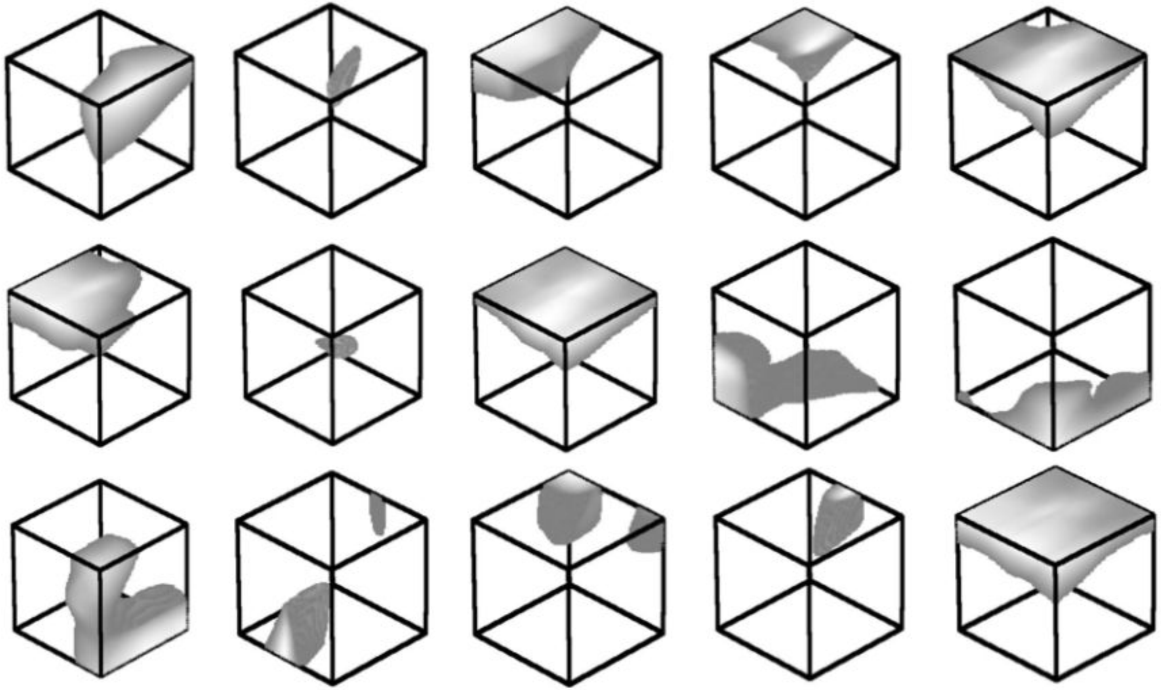
PointNet Analysis : Critical and Upper bound Point Set

Credit [**Qi2017**]



PointNet Analysis : features activation Credit [**Qi2017**]

➔ Find the top-K points in a dense volumetric grid that activates neuron X.

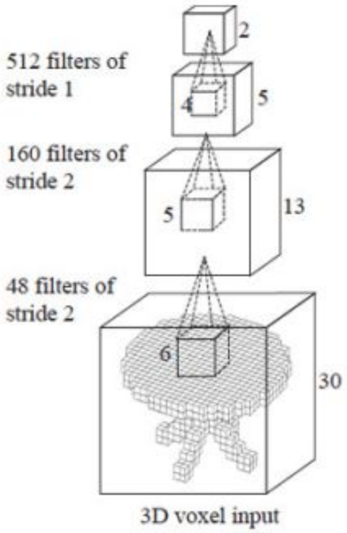


PointNet Limitations

Credit [Qi2017]

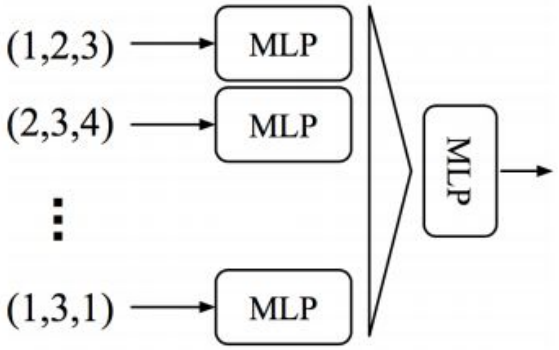
- Hierarchical Feature Learning
- Increasing receptive field

Global Feature Learning
Receptive field:
one point OR all points



3D CNN (Wu et al.)

V.S.



PointNet (vanilla) (Qi et al.)

	Analyse	Generation
RGBD	CNNs (resnet)	CNNs (resnet)
Mesh	SyncSpecNet Graph CNNs	AtlasNet Neural renderer RenderNet
Image-based Methods	Dosovitsky et al, ECCV 2016	SurfNet
Voxel Based Methods	3D-r2n2 OctNet Hierarchical Surface Prediction	3D-r2n2 Hierarchical Surface Prediction Octree Generative Networks
Point Based Methods	PointNet, PointNet++ SuperPoints Graph (large scale) PCPNet	PointSetGen,
Primitives	Shape Abstraction	Shape Abstraction Supervised Fitting of Geometric Primitives to 3D Point Clouds