valeo.ai

Nuages de Points et Modélisation 3D

5 - Machine learning II

Overview

Machine learning courses

- Surface reconstruction
- Descriptors and machine learning
- Image based processing
- Geometric deep learning

} Today

- Convolutional and Transformer based architectures

} ML course 3

- Tasks and corresponding architectures

} ML course 4

I - Image-based approaches

Idea

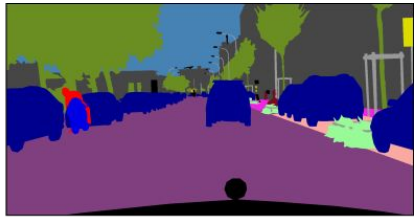
Image-based approaches

Image processing is a well studied problem

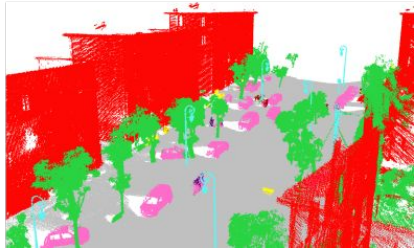
Idea

Image-based approaches

Image processing

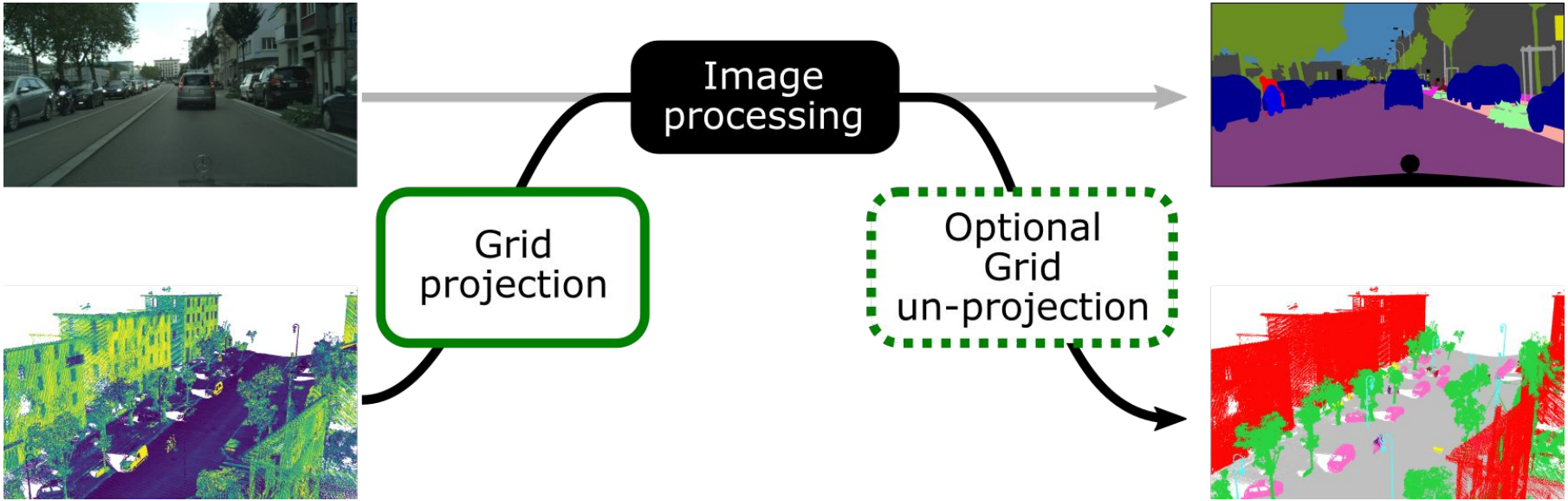


?



Idea

Image-based approaches



Regular grid projections

Image-based approaches

valeo.ai

Images are pixels arrays

Implicit neighborhoods

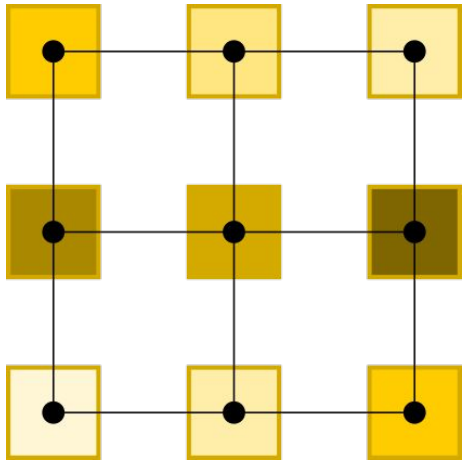
Information is in the color and relative position of the pixels

Thanks to this grid structure

Optimized network architectures

Fast (hardware optimization)

Relatively low memory cost



2D projections

Image-based approaches

2D convolution for an image patch centered on pixel n:

$$\mathbf{h}[n] = \sum_{f \in \{1,\dots,C\}} \sum_{m \in \{-M/2,\dots,M/2\}^2} \mathbf{K}_f[m]\, \mathbf{f}_f[n+m]$$

With **f**: input features and **K**: convolution kernel

And new architectures for images:

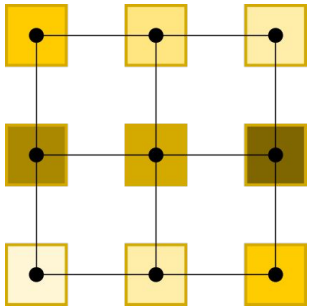
Vision transformers, MLP Mixers, …

Regular grid projections

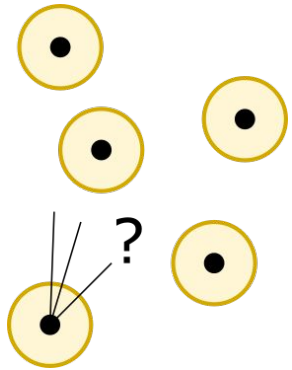
Image-based approaches

Images are pixels arrays
Implicit neighborhoods
Information is in the color and relative position of the pixels

Thanks to this grid structure
Optimized network architectures
Fast (hardware optimization)
Relatively low memory cost



Point clouds:
~~Implicit neighborhoods~~
~~Information is in the color and relative position of the pixels~~



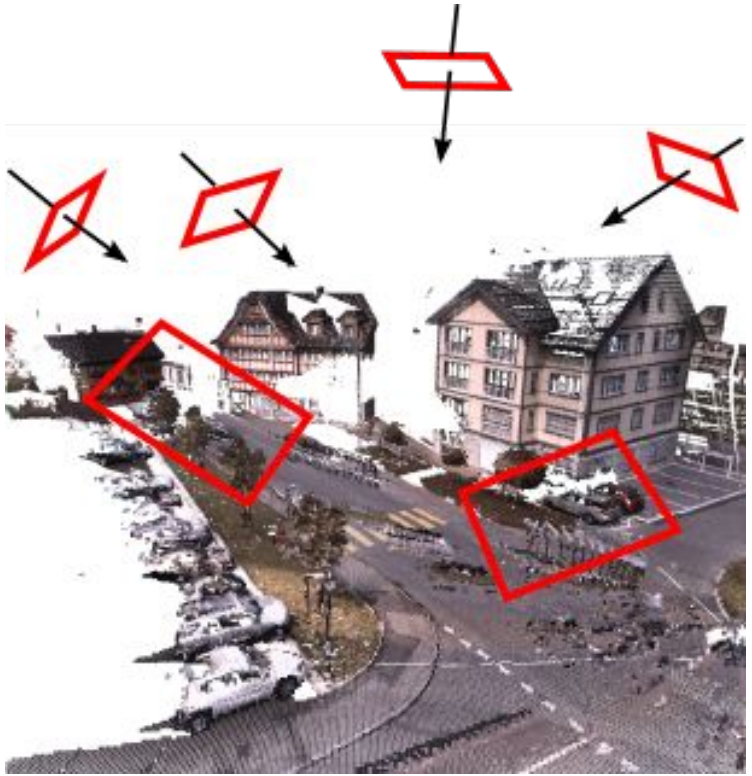
Idea
Find a way to create grid data from point cloud

Image projections projections

Image-based approaches

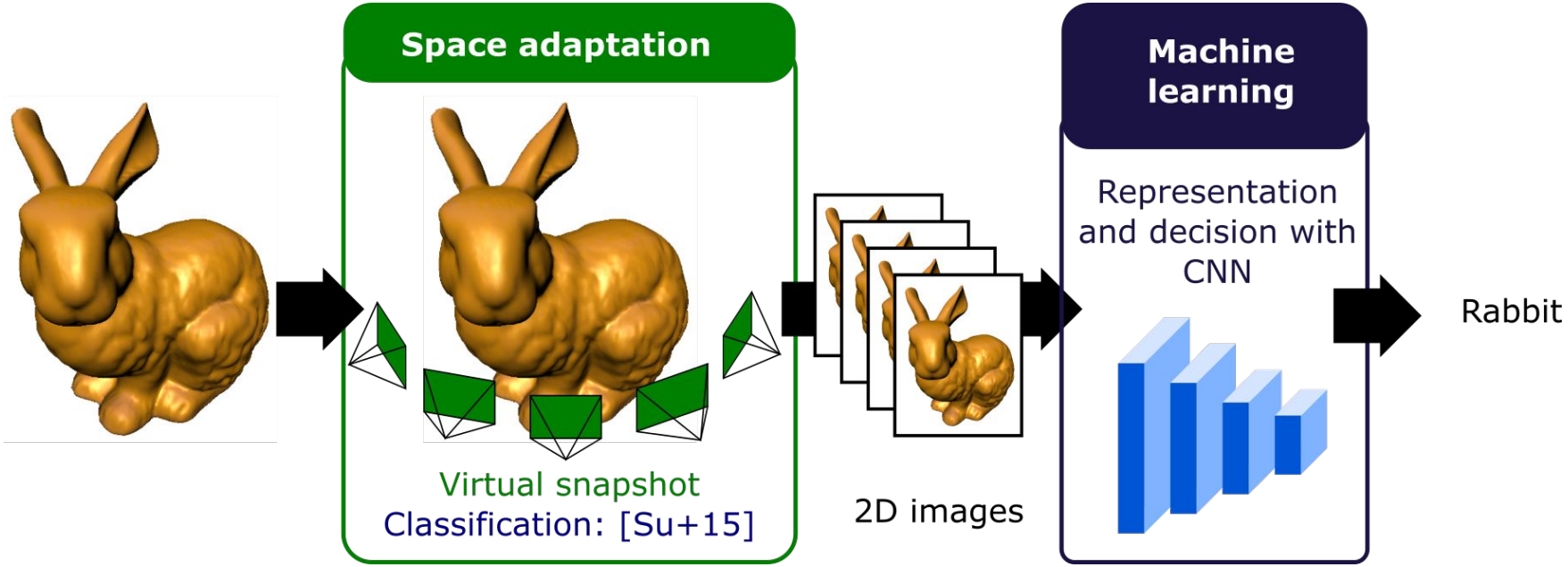
Generate images representing the scene

- Use a 3D renderer
- Take virtual snapshots of the scene
- Work in the image



Classification pipeline

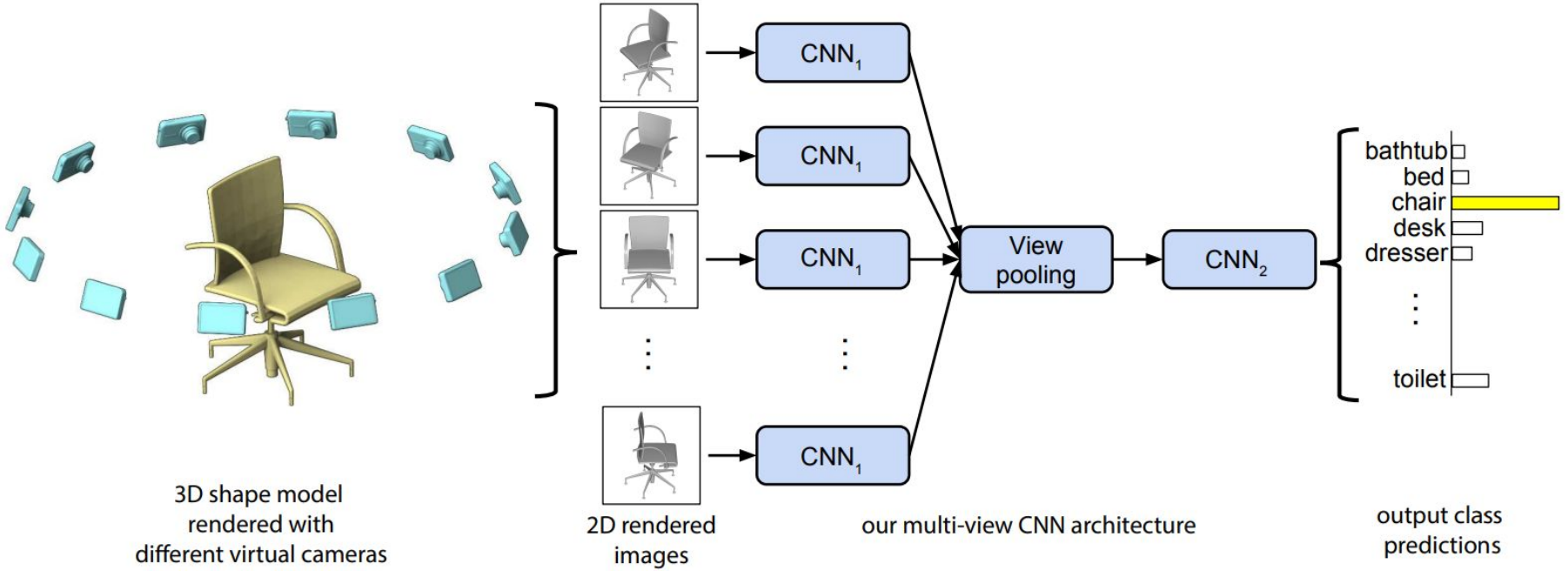
Image-based approaches



Hang Su et al. “Multi-view convolutional neural networks for 3D shape recognition”. In: Proceedings of the IEEE international conference on computer vision. 2015, pp. 945–953

Classification pipeline

Image-based approaches



Hang Su et al. “Multi-view convolutional neural networks for 3D shape recognition”. In: Proceedings of the IEEE international conference on computer vision. 2015, pp. 945–953

Classification

valeo.ai

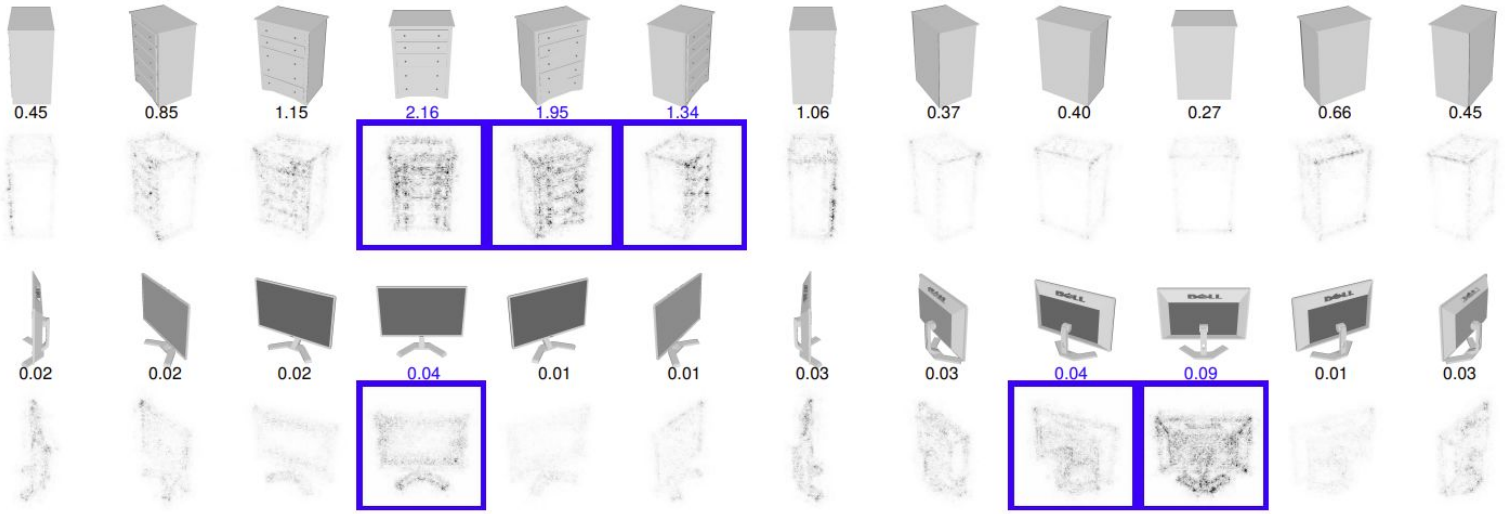
Image-based approaches

Method	Training Config.			Test Config.	Classification (Accuracy)	Retrieval (mAP)
	Pre-train	Fine-tune	#Views	#Views		
(1) SPH [16]	-	-	-	-	68.2%	33.3%
(2) LFD [5]	-	-	-	-	75.5%	40.9%
(3) 3D ShapeNets [37]	ModelNet40	ModelNet40	-	-	77.3%	49.2%
(4) FV	-	ModelNet40	12	1	78.8%	37.5%
(5) FV, 12×	-	ModelNet40	12	12	84.8%	43.9%
(6) CNN	ImageNet1K	-	-	1	83.0%	44.1%
(7) CNN, f.t.	ImageNet1K	ModelNet40	12	1	85.1%	61.7%
(8) CNN, 12×	ImageNet1K	-	-	12	87.5%	49.6%
(9) CNN, f.t.,12×	ImageNet1K	ModelNet40	12	12	88.6%	62.8%
(10) MVCNN, 12×	ImageNet1K	-	-	12	88.1%	49.4%
(11) MVCNN, f.t., 12×	ImageNet1K	ModelNet40	12	12	89.9%	70.1%
(12) MVCNN, f.t.+metric, 12×	ImageNet1K	ModelNet40	12	12	89.5%	**80.2%**
(13) MVCNN, 80×	ImageNet1K	-	80	80	84.3%	36.8%
(14) MVCNN, f.t., 80×	ImageNet1K	ModelNet40	80	80	**90.1%**	70.4%
(15) MVCNN, f.t.+metric, 80×	ImageNet1K	ModelNet40	80	80	**90.1%**	79.5%

* f.t.=fine-tuning, metric=low-rank Mahalanobis metric learning

Classification

Image-based approaches



Saliency map (derivative of output w.r.t. the pixel input)

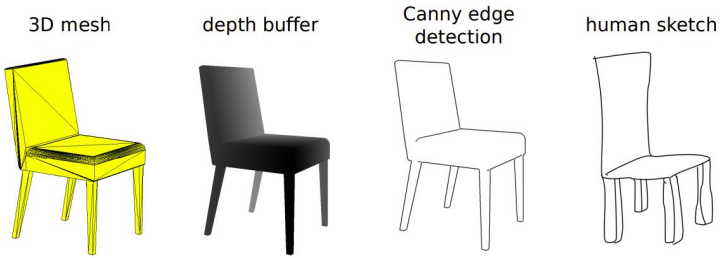
Most discriminative features

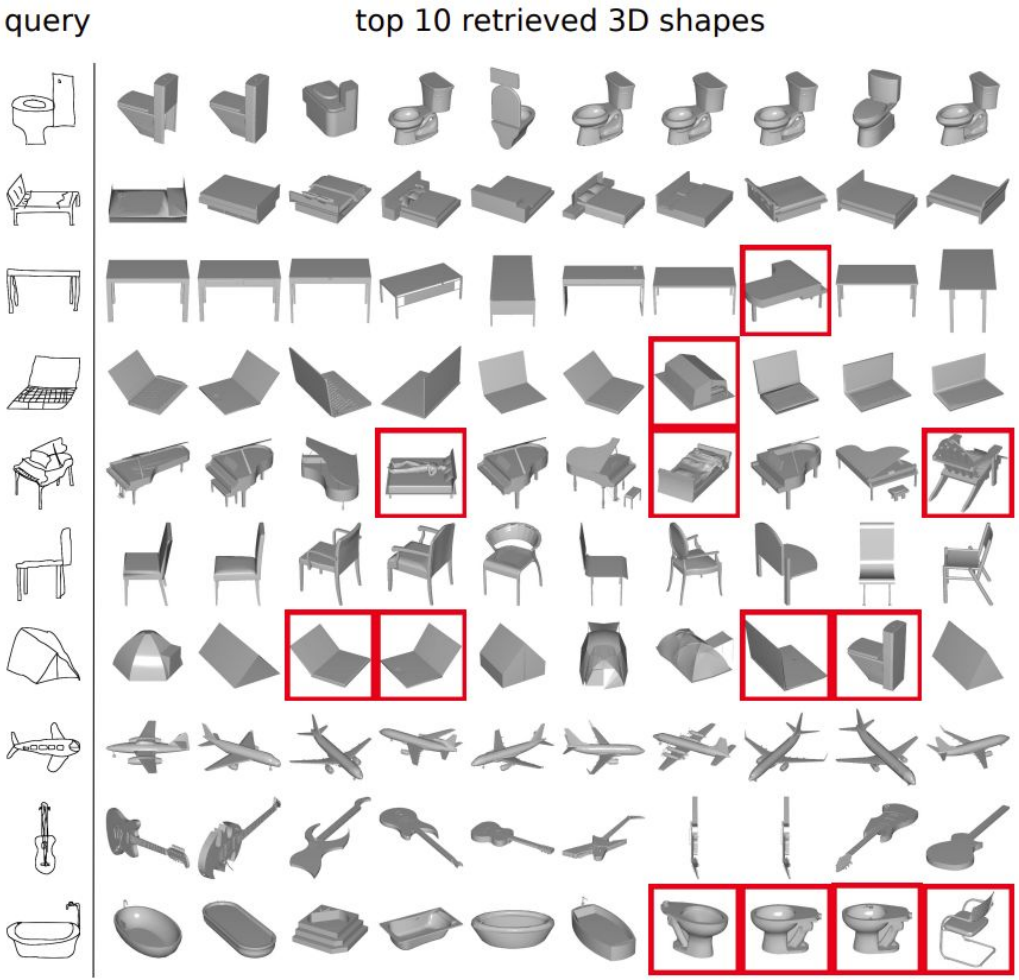
Hang Su et al. “Multi-view convolutional neural networks for 3D shape recognition”. In: Proceedings of the IEEE international conference on computer vision. 2015, pp. 945–953

Shape retrieval

Image-based approaches

Training with different input





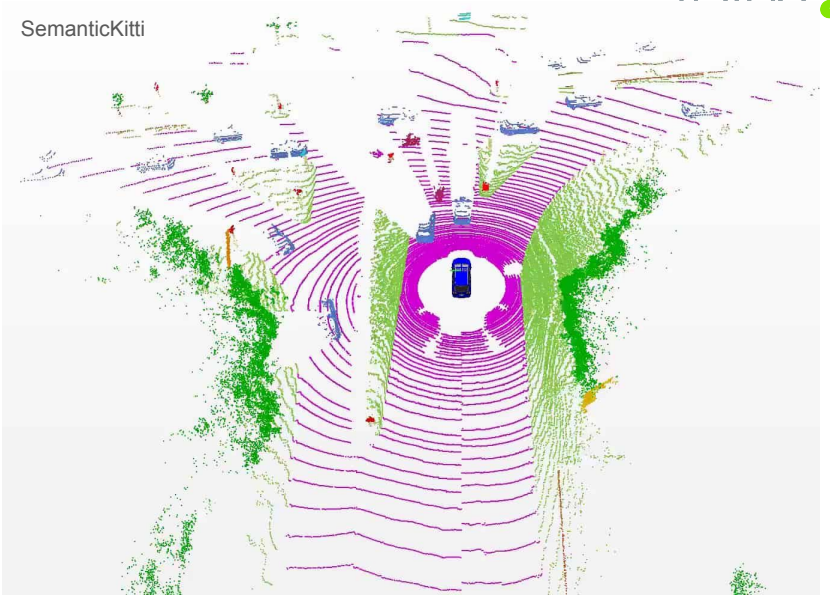
Semantic segmentation

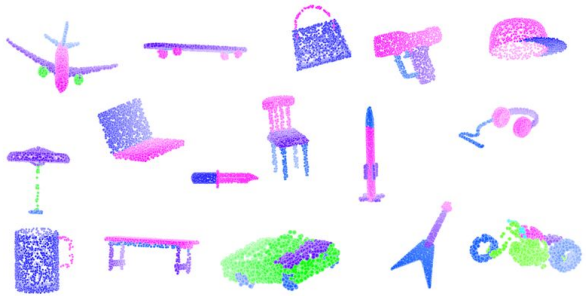
Image-based approaches

One label per point

Scannet

SemanticKitti

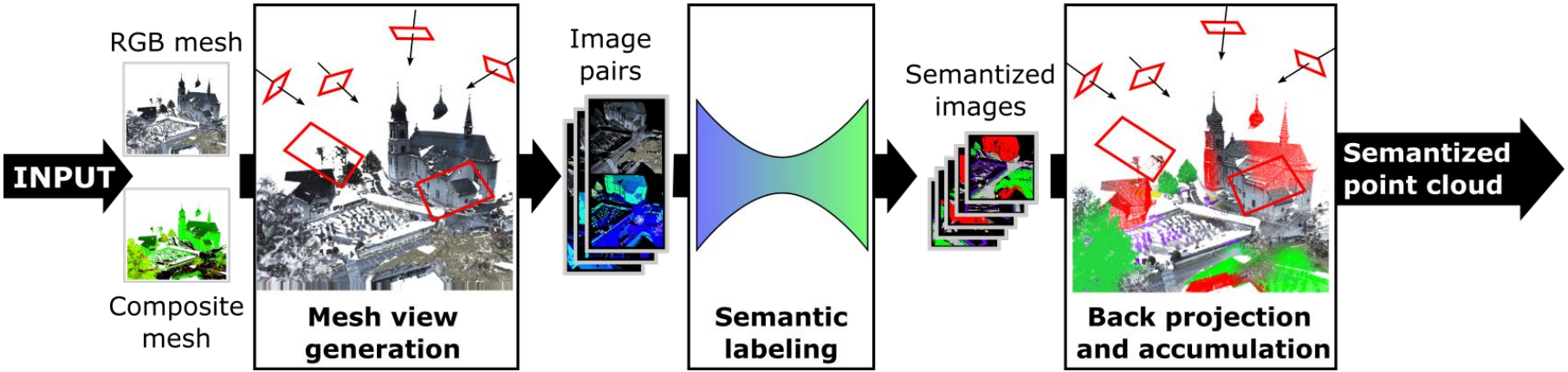




ShapeNet part seg

Semantic segmentation pipeline

Image-based approaches



Alexandre Boulch et al. “SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks”. In: Computers & Graphics (2017)

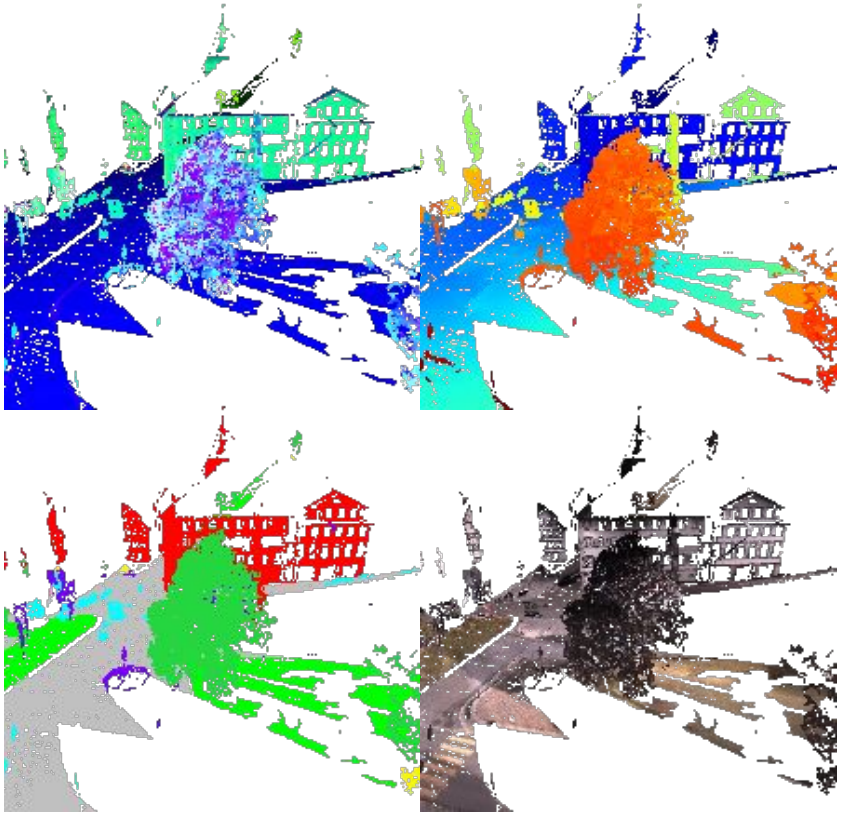
SnapNet

Image-based approaches

Reprojection trick

Generate a snapshot of the scene with fake colors corresponding to point ids

- Allow to generate different snapshots (w / wo colors, geometric features, ground truth at training...)
- Easy reprojection of the results on the original points



Alexandre Boulch et al. “SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks”. In: Computers & Graphics (2017)

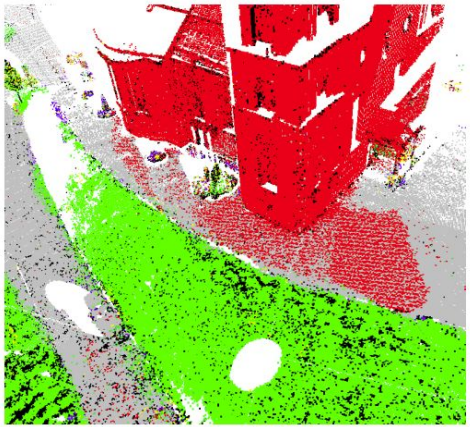
SnapNet

Image-based approaches

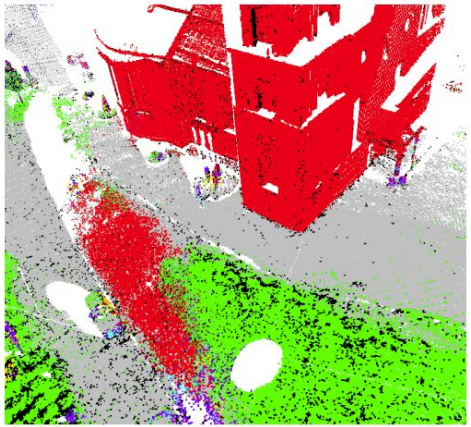
valeo.ai

Each modality carries different features

→ need for RGB + Geometry



RGB



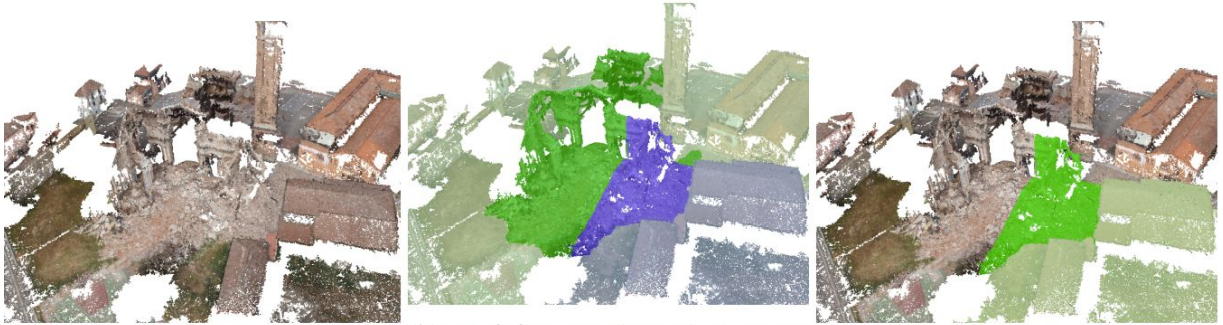
Composite

SnapNet

Image-based approaches

(a) RGB colors. (b) Depth composite texture. (c) Predictions.

Figure 10: Semantic labeling of photogrammetric data.



(a) RGB colors (b) Training set (green), test set (blue). (c) Rubble prediction.

SnapNet - advantages and limitations

valeo.ai

Image-based approaches

Pros

- Benefit from architectures from image processing
- Use of pre-trained models from large image datasets
- Unlimited number of snapshot a given scene (straightforward data augmentation)

Cons

- Good snapshot strategy, which can vary from a dataset to another
- Requires a mesh

Alexandre Boulch et al. “SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks”. In: Computers & Graphics (2017)

Range projection

Image-based approaches

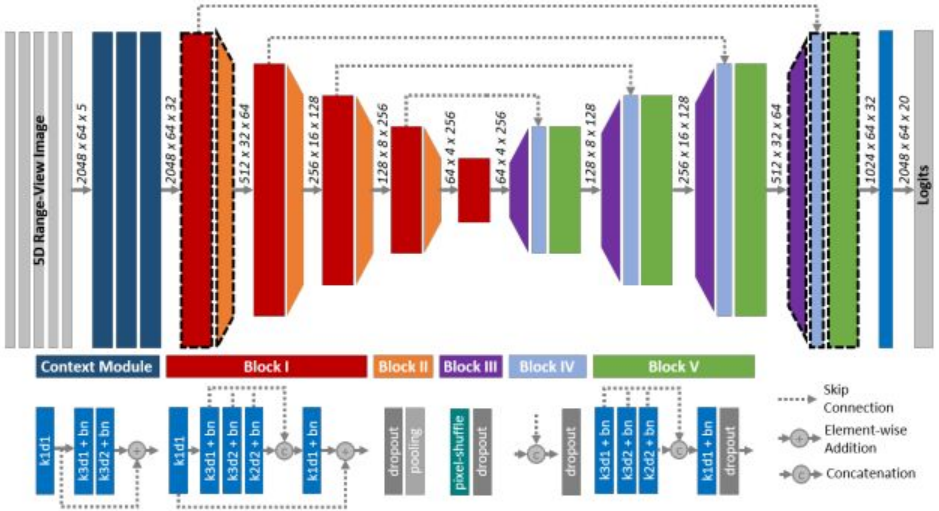
Exploit sensor information to produce images

SalsaNext

Image-based approaches

valeo.ai

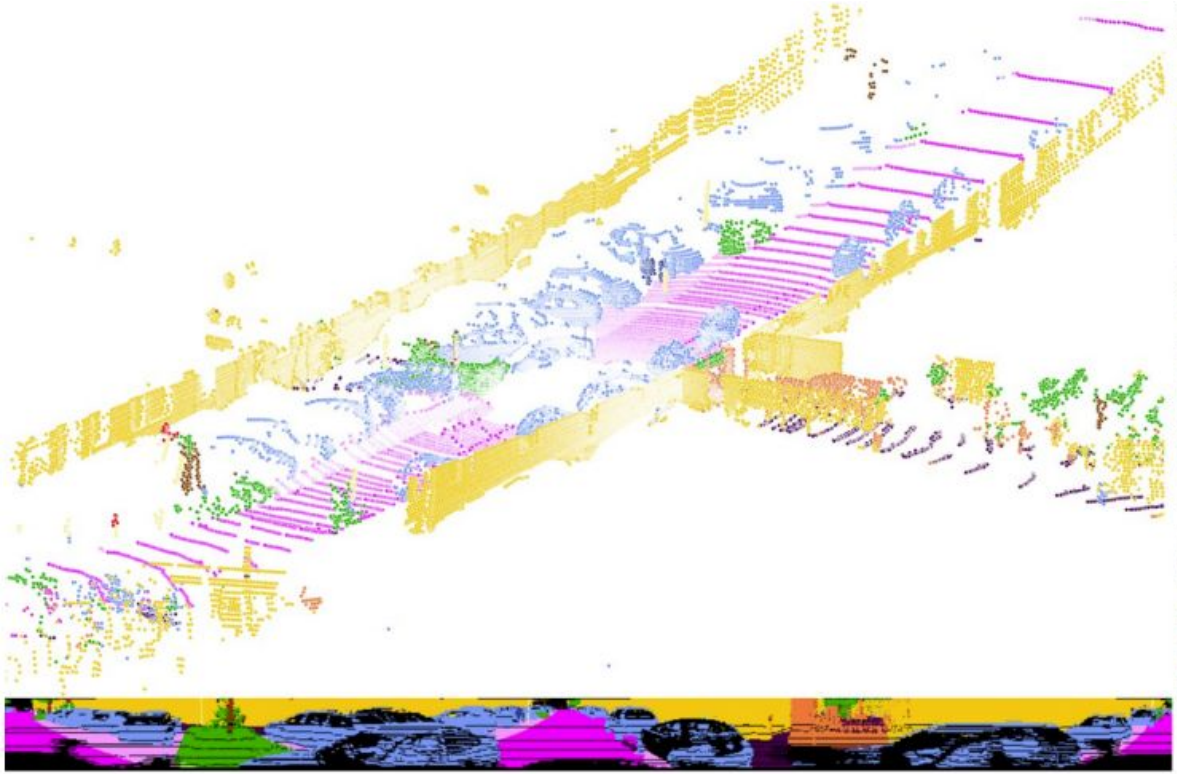
Use image backbone (U-Net) for semantic segmentation



Cortinhal, Tiago et al. "Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds." ISVC 2020.

SalsaNext

Image-based approaches

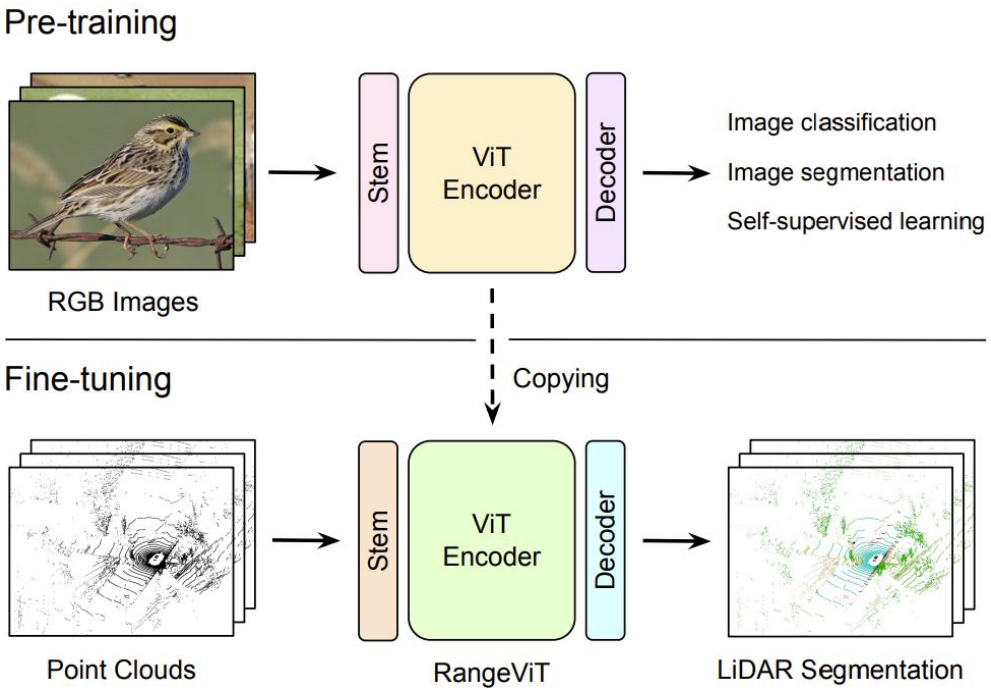


RangeViT

Image-based approaches

LiDAR segmentation based on range images & vision transformers (ViTs)

- Unify architectures in LiDAR and image domain
- Leverage image pre-trained ViTs for LiDAR segmentation



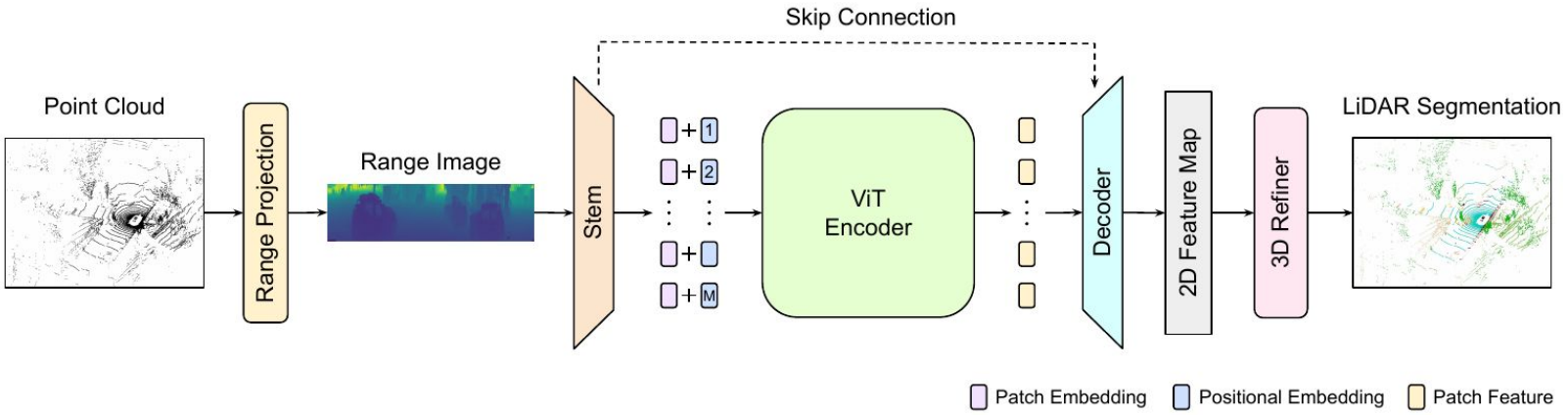
Angelika Ando et al., “RangeViT: Towards Vision Transformers for 3D Semantic Segmentation in Autonomous Driving” CVPR 2023

RangeViT

Image-based approaches

Detail of the architecture

Stem + skip connection / 3D refiner



Angelika Ando et al., “RangeViT: Towards Vision Transformers for 3D Semantic Segmentation in Autonomous Driving” CVPR 2023

RangeViT

Image-based approaches

valeo.ai

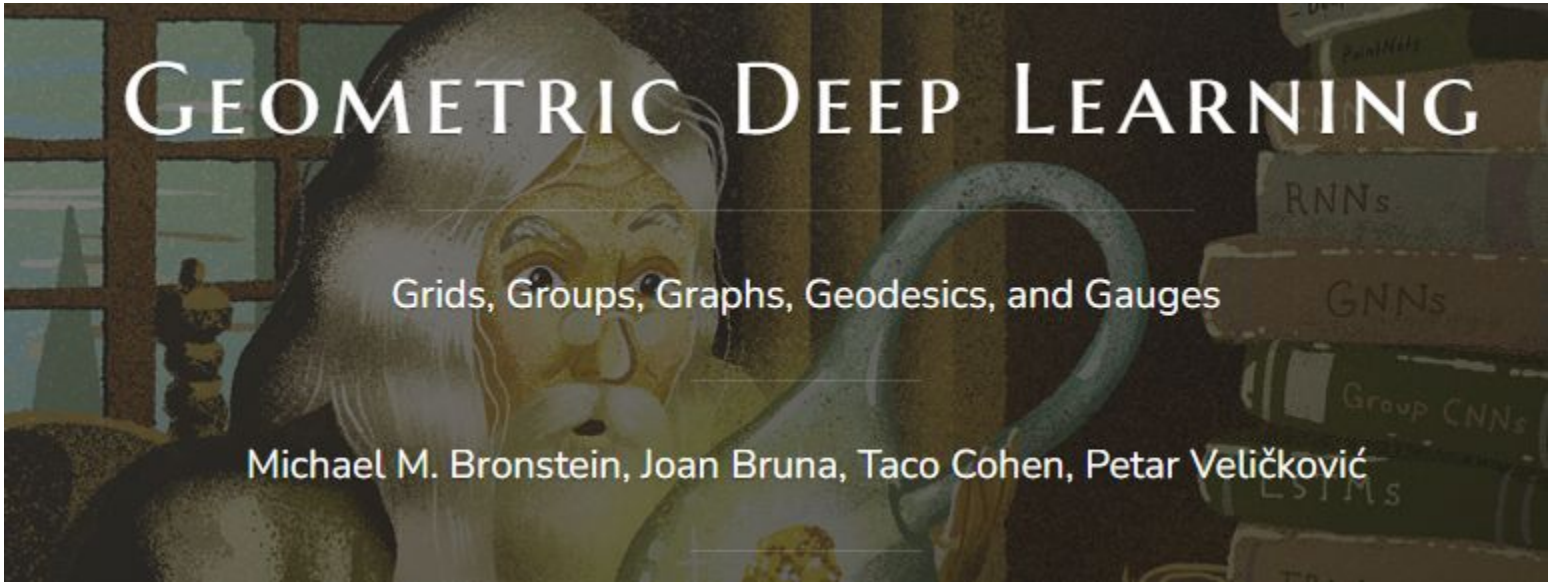
Use vision transformers

Various initializations

Pre-training	Rand	DINO	IN21k	CS
mIoU	72.37	73.33	74.77	**75.21**

Method	mIoU (%)
Voxel-based	
Cylinder3D [69]	**76.1**
2D Projection-based	
RangeNet++ [36]	65.5
PolarNet [66]	71.0
SalsaNext [11]	72.2
RangeViT-IN21k (ours)	74.8
RangeViT-CS (ours)	**75.2**

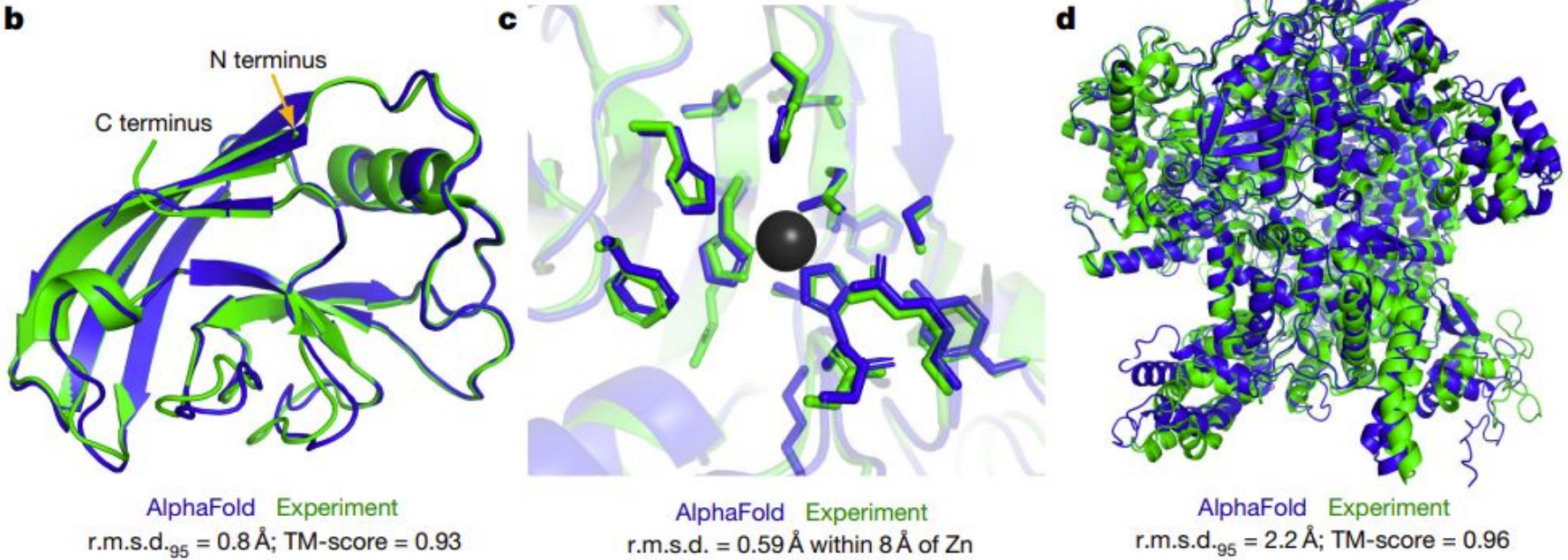
II - Graph deep learning (a ridiculously small part)



https://geometricdeeplearning.com/

AlphaFold

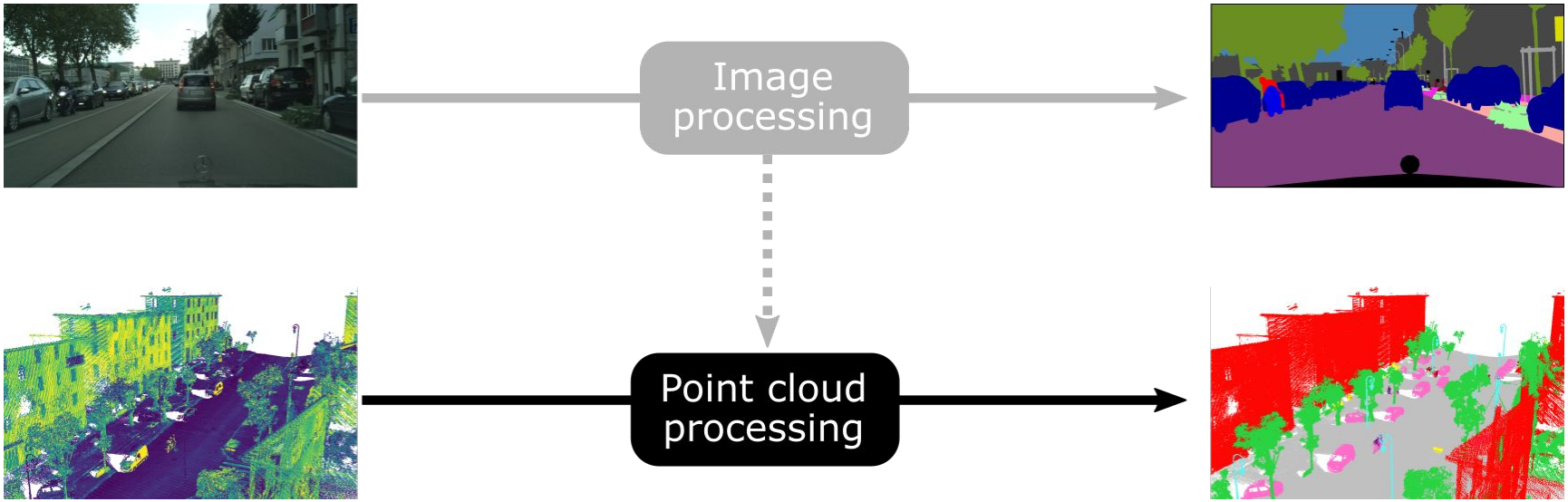
valeo.ai



Jumper et al. Highly accurate protein structure prediction with AlphaFold, Nature, 2021

Geometric deep learning

Geometric deep learning



Main problem

Geometric deep learning

valeo.ai

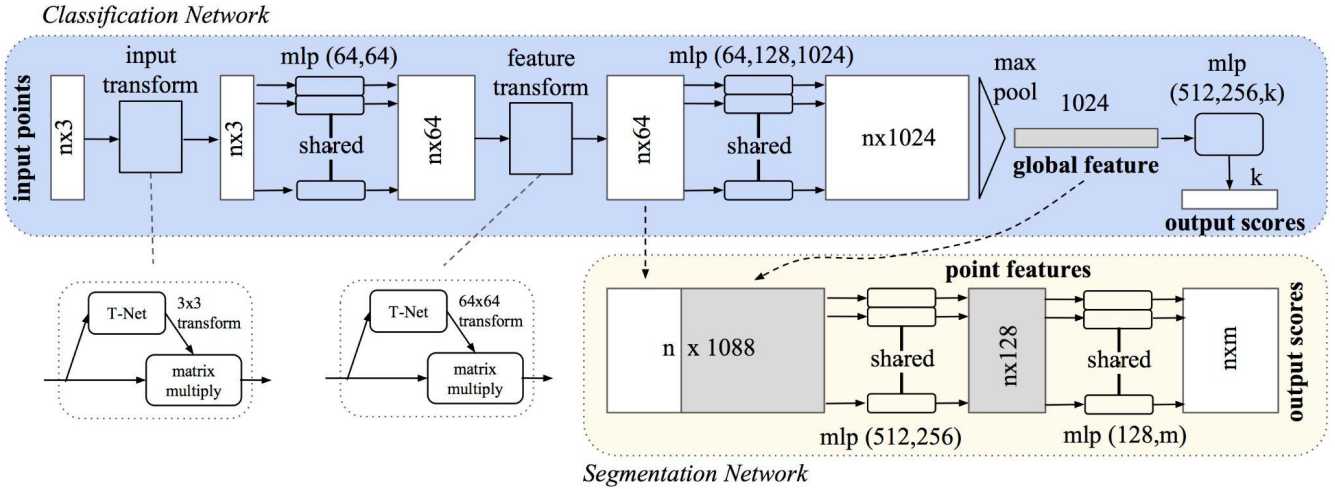
Point cloud are permutation invariant

$$P = [p_1, p_2, \ldots, p_i, \ldots, p_j, \ldots, p_n] \iff P = [p_1, p_2, \ldots, p_j, \ldots, p_i, \ldots, p_n]$$

Neural networks (as usual) are not permutation invariant: swapping features has an influence on the output (e.g., shuffling the pixels of an image)

PointNet

Geometric deep learning



- Use point-wise operation: MLP on feature vector (e.g., coordinates of the points)
- Aggregate with a permutation invariant operation: global pooling (max or average)

Charles R Qi et al. “PointNet: Deep learning on point sets for 3D classification and segmentation”. In: Proc. Computer Vision and Pattern Recognition (CVPR), IEEE 1.2 (2017), p. 4

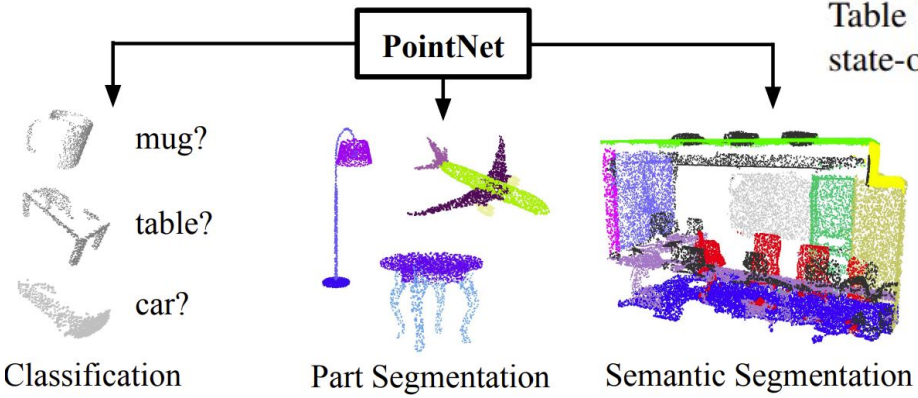
PointNet

Geometric deep learning

valeo.ai

Multiple applications

- Classification
- Segmentation
- …



	input	#views	accuracy avg. class	accuracy overall
SPH [11]	mesh	-	68.2	-
3DShapeNets [28]	volume	1	77.3	84.7
VoxNet [17]	volume	12	83.0	85.9
Subvolume [18]	volume	20	86.0	**89.2**
LFD [28]	image	10	75.5	-
MVCNN [23]	image	80	**90.1**	-
Ours baseline	point	-	72.6	77.4
Ours PointNet	point	1	86.2	**89.2**

Table 1. **Classification results on ModelNet40.** Our net achieves state-of-the-art among deep nets on 3D input.

PointNet++

Geometric deep learning

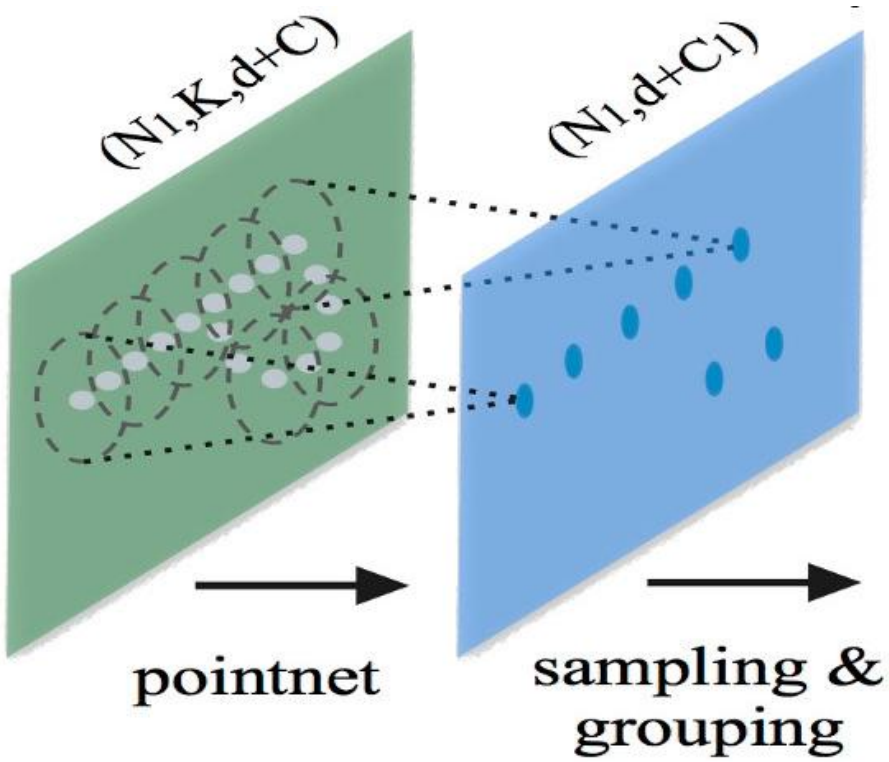
valeo.ai

PointNet limitation

Global operations induce a loss in detail analysis

Solution

- Apply local PointNets
- Aggregate results on support points
- Iterate



PointNet++

Geometric deep learning

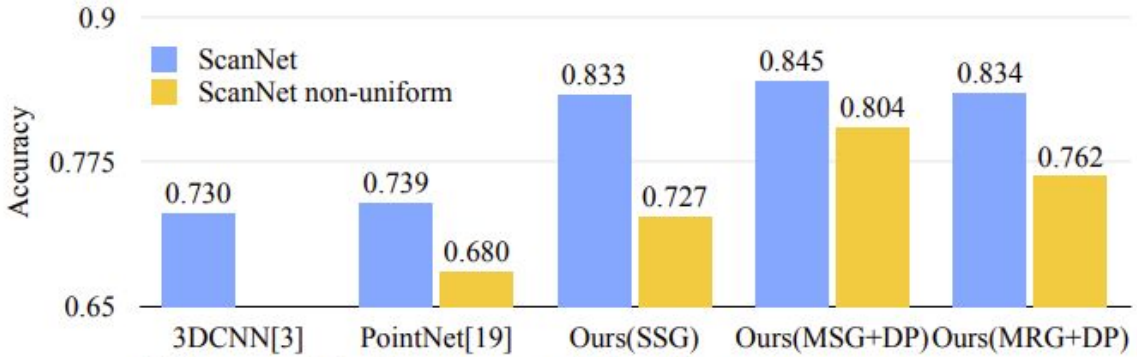


Figure 5: Scannet labeling accuracy.

Big boost, particularly, in semantic segmentation.

II - Geometric deep learning

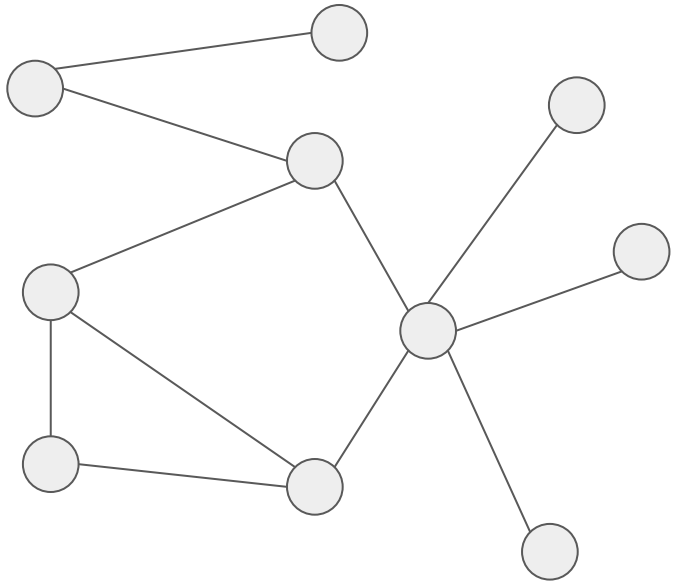
A - Graph convolution

Graph

Graph message passing

Points + edges

Undirected

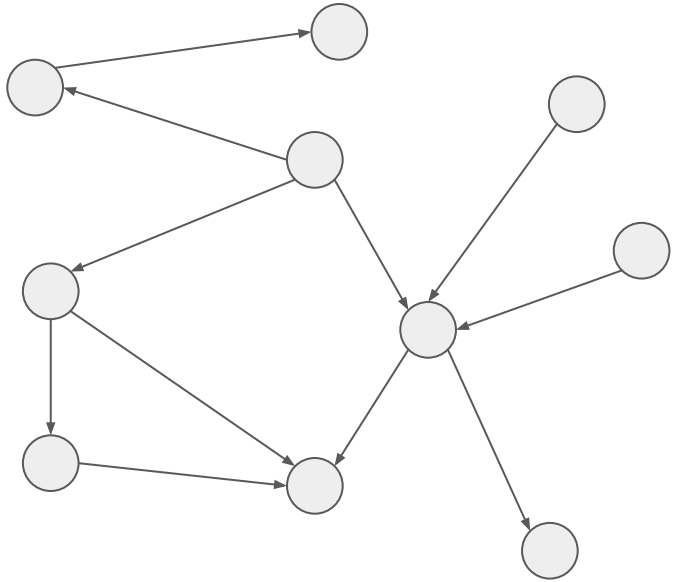


Graph

Graph message passing

Points + edges

Directed



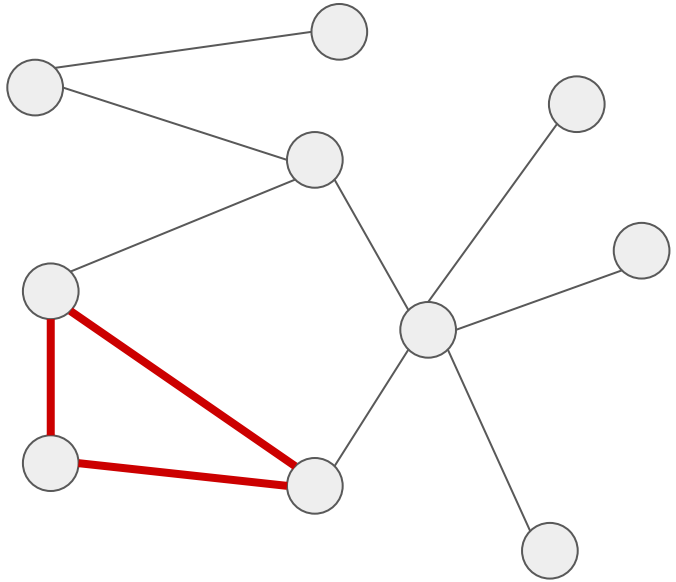
Graph

Graph message passing

Points + edges

Undirected

Cycles



Graph

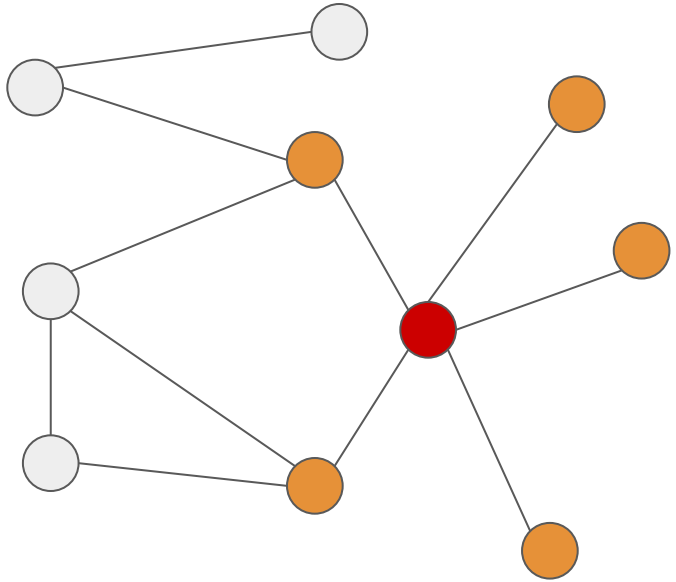
Graph message passing

Points + edges

Undirected

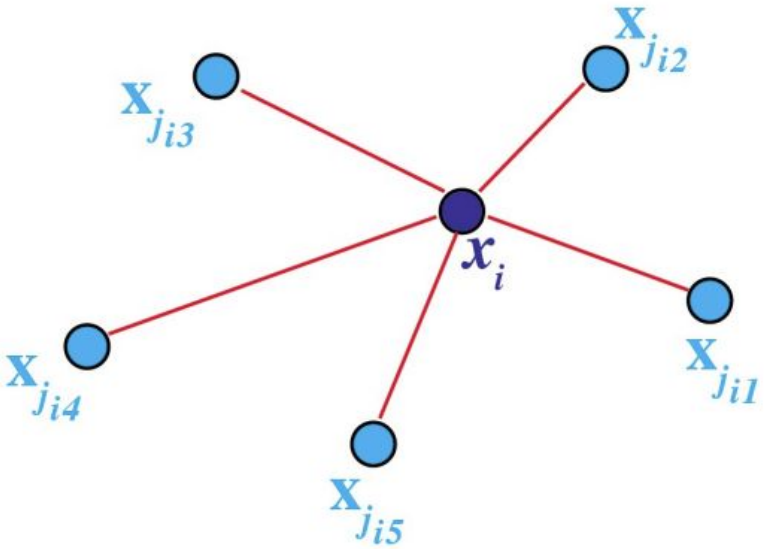
Cycles

Neighbors



DGCNN

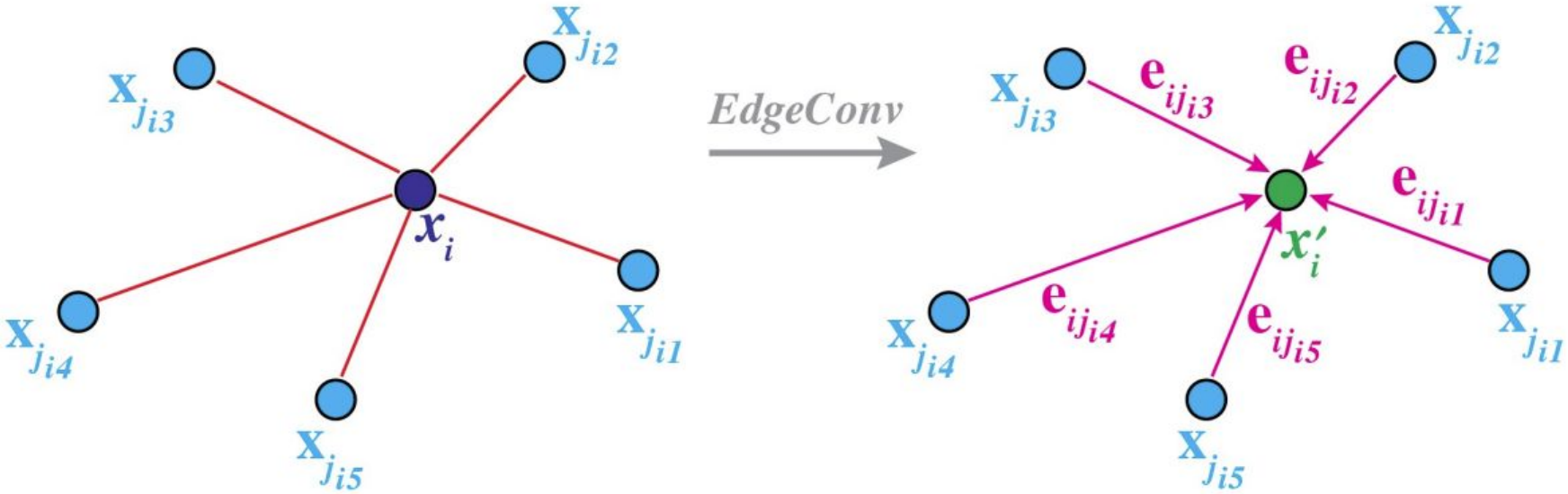
Graph convolution



Wang et al. Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics, 2019

DGCNN

Graph convolution



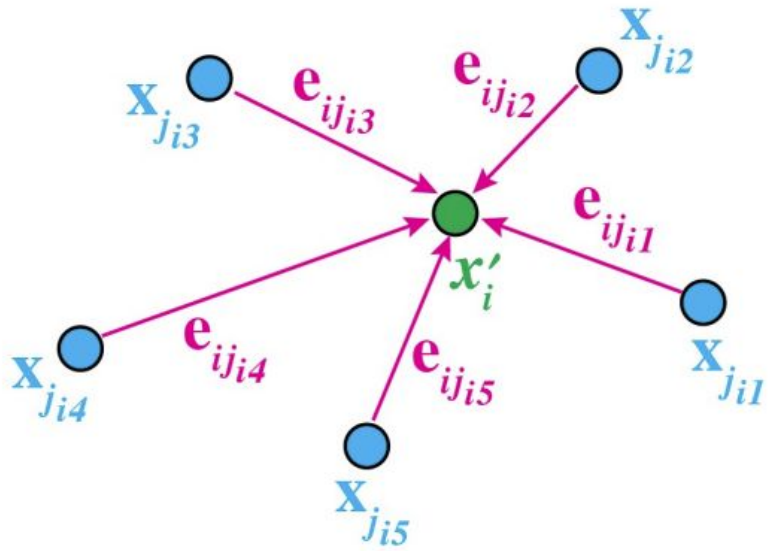
Wang et al. Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics, 2019

DGCNN

Graph convolution

Similar to images that look at pixel neighborhoods, edge convolution looks at graph neighborhoods.

$$\mathbf{x}'_i = \square_{j:(i,j)\in\mathcal{E}} h_\Theta(\mathbf{x}_i, \mathbf{x}_j).$$



Wang et al. Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics, 2019

PointNet / PointNet++

Graph convolution

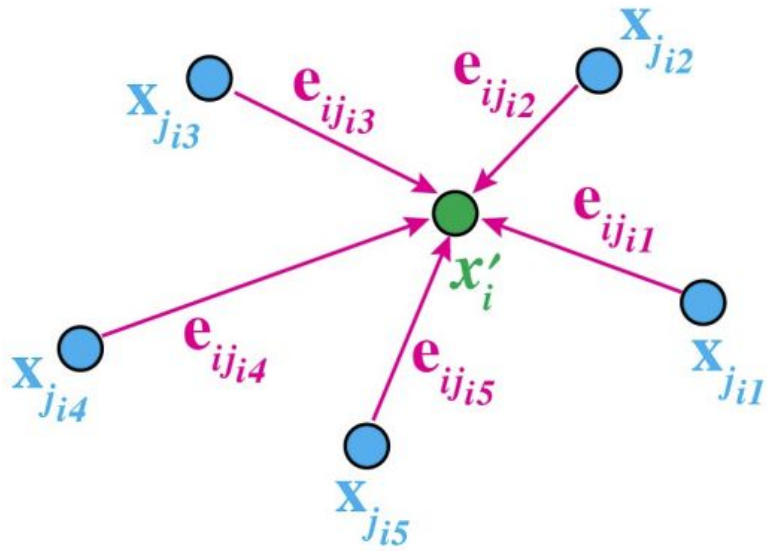
PointNet / PointNet++

$$h_\Theta(\mathbf{x}_i, \mathbf{x}_j) = h_\Theta(\mathbf{x}_i),$$

And $\square_{j:(i,j)\in\mathcal{E}}$ is a max pooling

for PointNet++ (Null for PointNet)

(The last aggregation is max pooling for all)



Wang et al. Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics, 2019

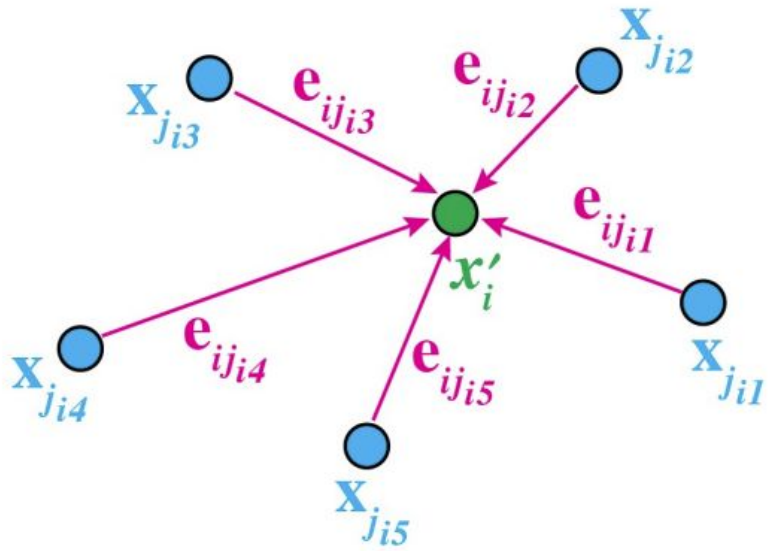
PCNN

Graph convolution

PCNN uses Gaussian kernel an weighting of the sum of the neighbors

$$h_\Theta(\mathbf{x}_i, \mathbf{x}_j) = h_\Theta(\mathbf{x}_j)$$

$$x'_{im} = \sum_{j \in \mathcal{V}} (h_{\theta(\mathbf{x}_j)}) g(u(\mathbf{x}_i, \mathbf{x}_j)),$$



Atzmon et al. Point Convolutional Neural Networks by Extension Operators, 2018

DGCNN

Graph convolution

DGCNN

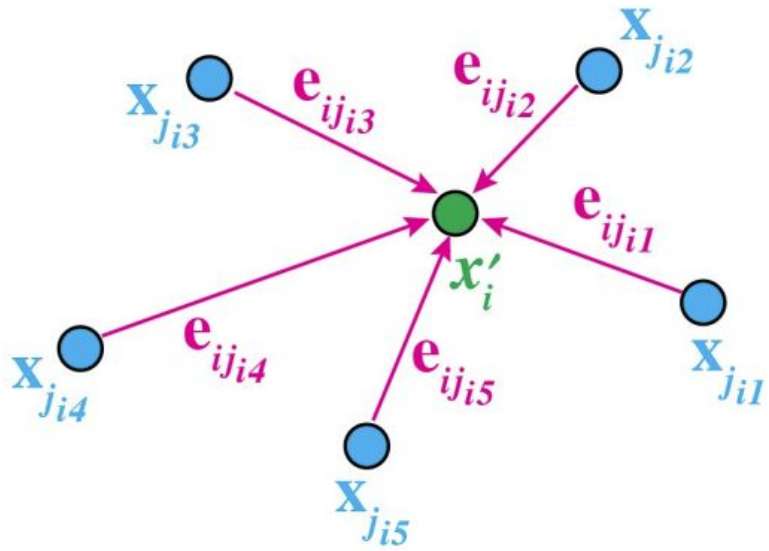
$$h_\Theta(\mathbf{x}_i, \mathbf{x}_j) = \bar{h}_\Theta(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i).$$

Then (in practice we also put a batchnorm)

$$e'_{ijm} = \mathrm{ReLU}(\theta_m \cdot (\mathbf{x}_j - \mathbf{x}_i) + \phi_m \cdot \mathbf{x}_i),$$

and

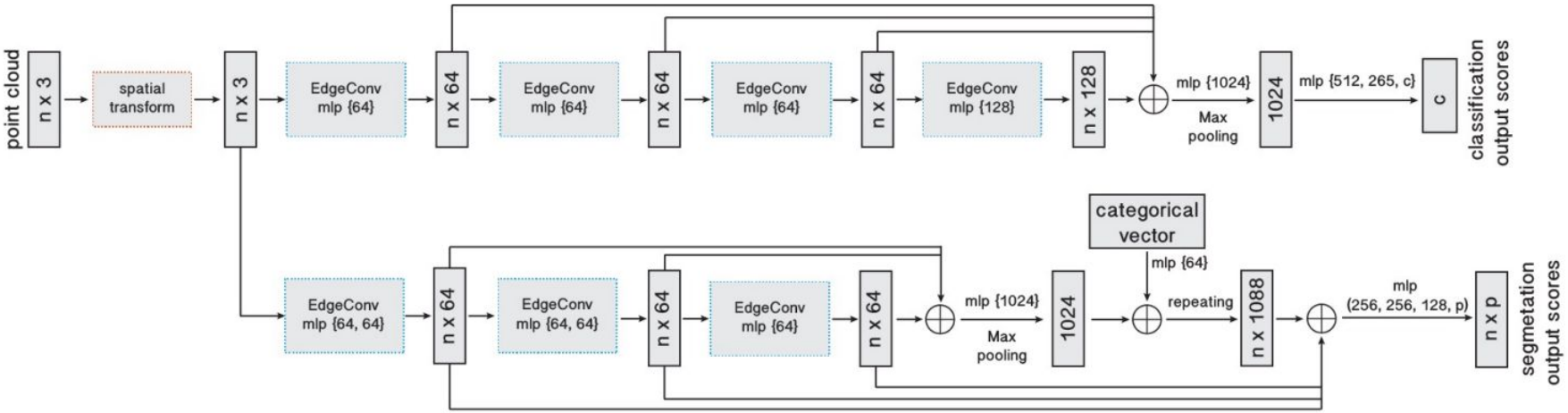
$$x'_{im} = \max_{j:(i,j)\in\mathcal{E}} e'_{ijm},$$



Wang et al. Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics, 2019

Design CNN

Graph convolution



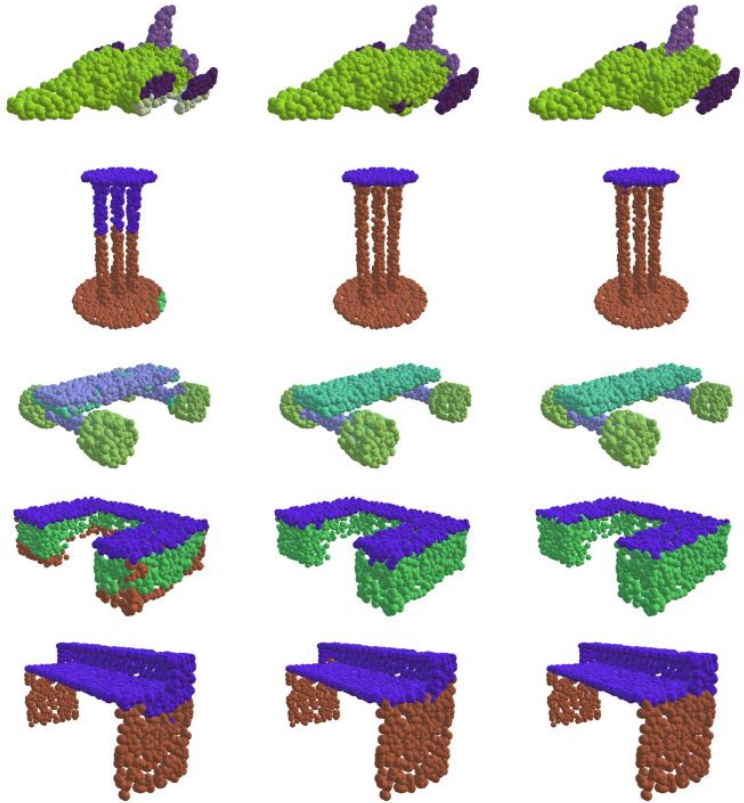
Wang et al. Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics, 2019

Design CNN

Graph convolution

Local features are better recognized than with only global pooling (PointNet)

Better performances in segmentation (classification also).

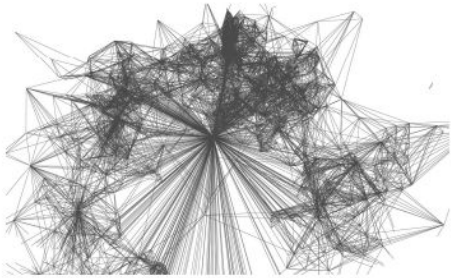


Wang et al. Dynamic Graph CNN for Learning on Point Clouds, ACM Transactions on Graphics, 2019

Super Point Graph

(a) RGB point cloud

(b) Geometric partition

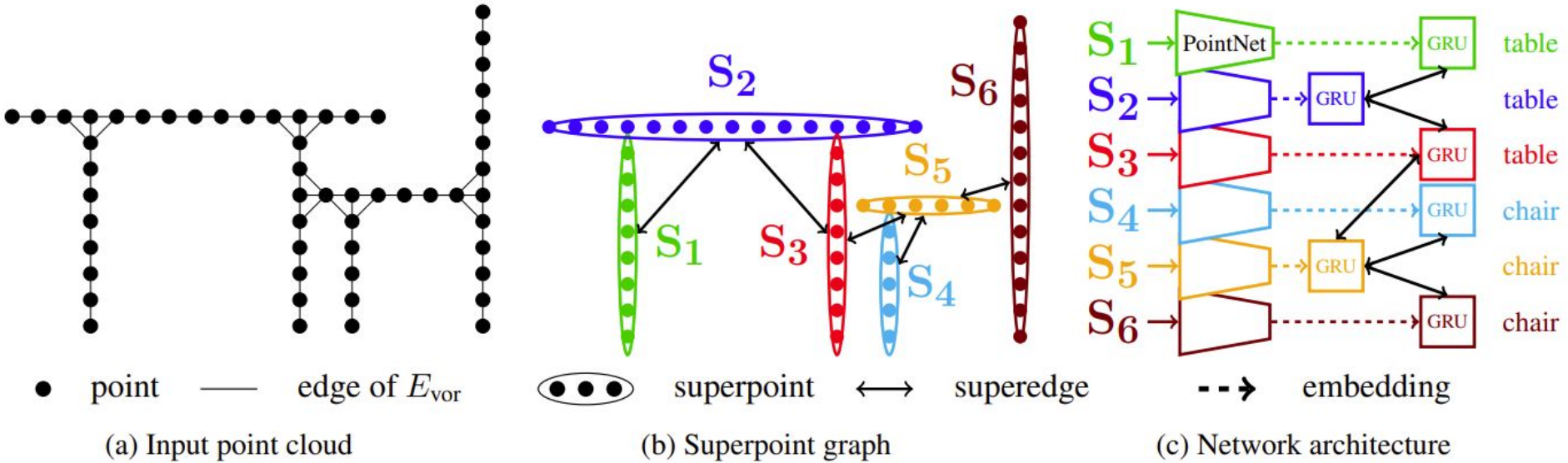


(c) Superpoint graph

(d) Semantic segmentation

Landrieu et al. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs, CVPR 2018

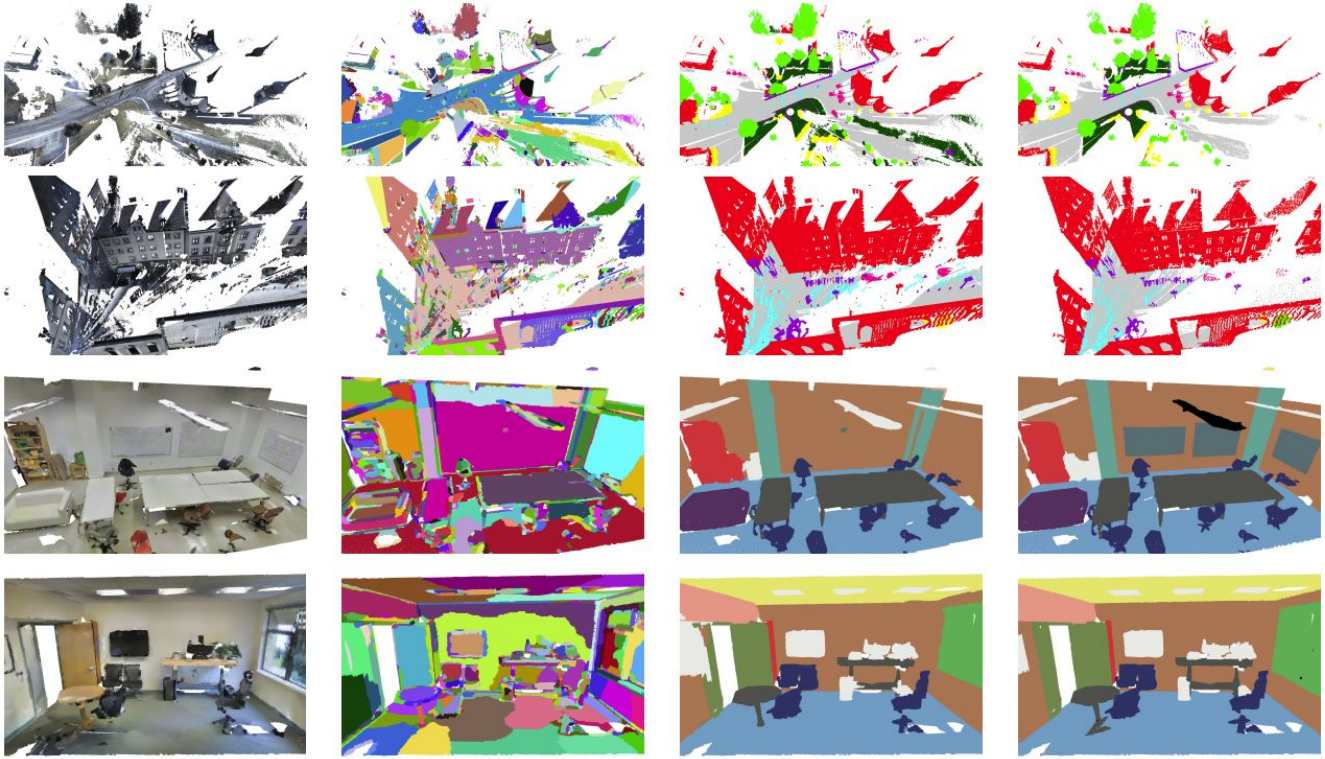
Super Point Graph



Landrieu et al. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs, CVPR 2018

Super Point Graph

valeo.ai



(a) RGB point cloud (b) Geometric partitioning (c) Prediction (d) Ground truth

Landrieu et al. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs, CVPR 2018

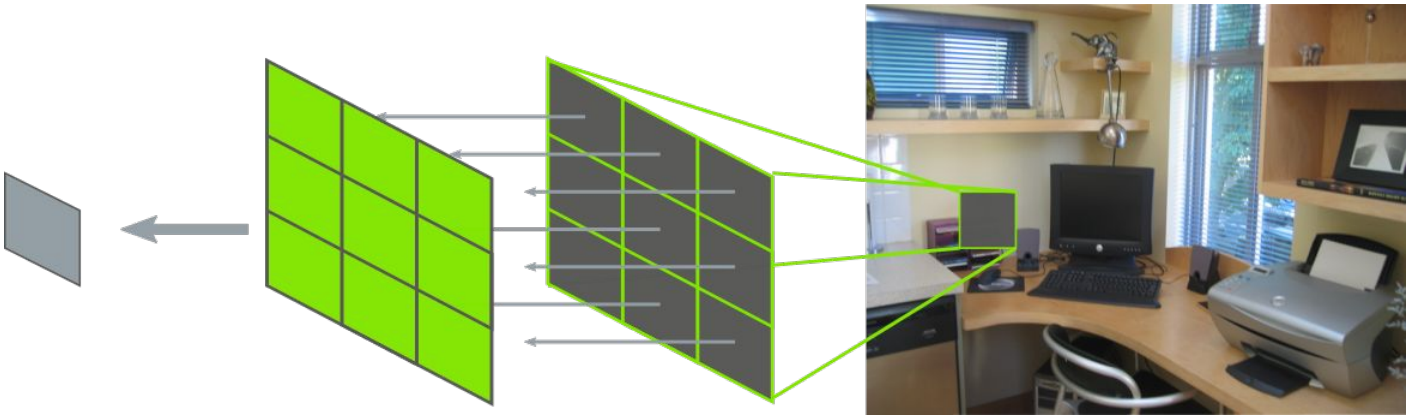
II - Geometric deep learning

A - Mimicking convolution

Convolution on images

Convolution for image processing

$$\mathbf{h}[n] = \sum_{f \in \{1,\ldots,C\}} \sum_{m \in \{-M/2,\ldots,M/2\}^d} \mathbf{K}_f[m]\, \mathbf{f}_f[n+m]$$

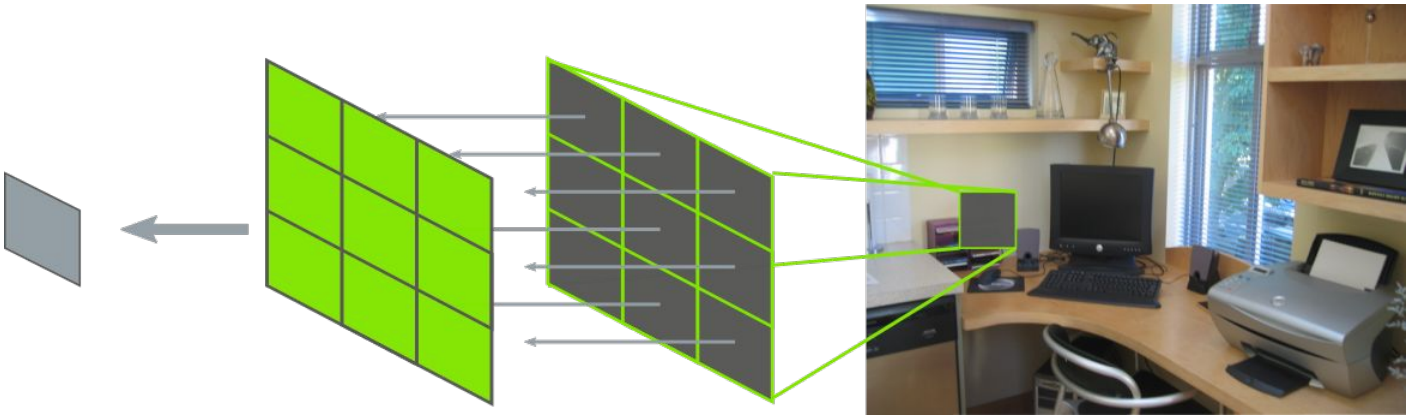


Convolution on images

valeo.ai

Convolution for image processing

$$\mathbf{h}[n] = \sum_{f \in \{1,\ldots,C\}} \mathbf{K}_f^\top \, \mathbf{f}_f(n)$$

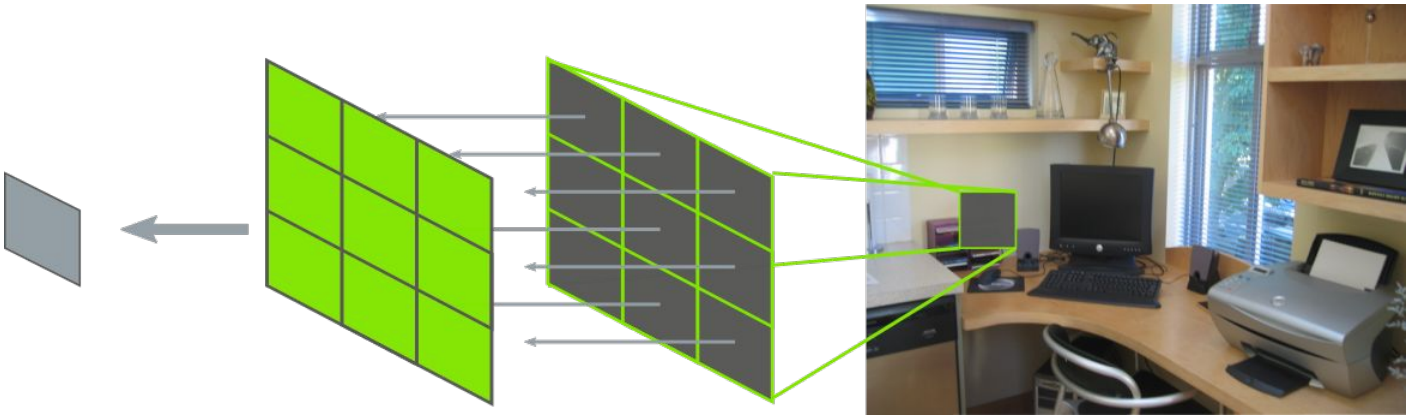


Convolution on images

valeo.ai

Convolution for image processing

$$\mathbf{h}[n] = \sum_{f \in \{1,\dots,C\}} \underbrace{\mathbf{K}_f^\top}_{\text{Kernel space}} \underbrace{\mathbf{f}_f(n)}_{\text{Feature space}}$$



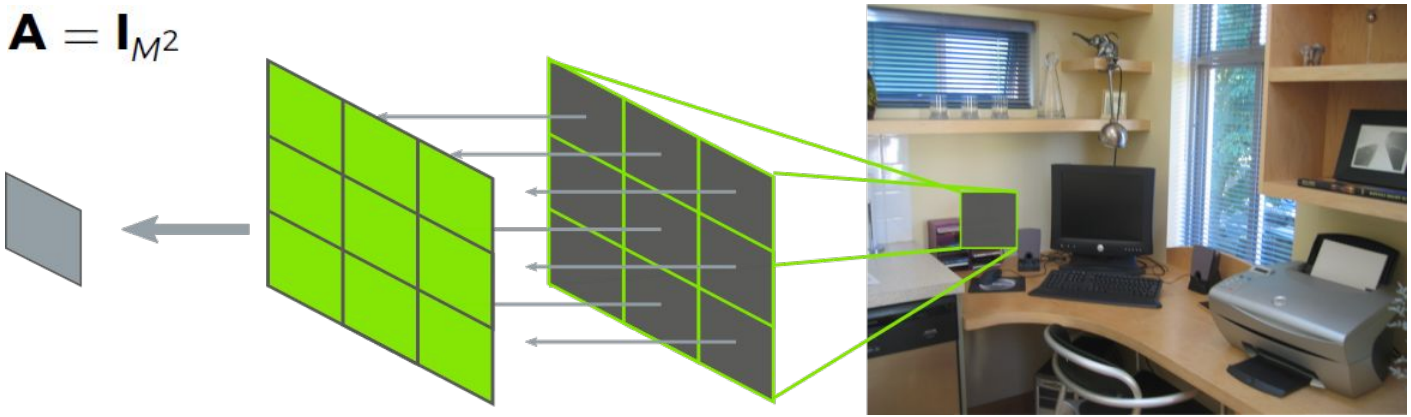
Convolution on images

Convolution for image processing

$$\mathbf{h}[n] = \sum_{f \in \{1,\dots,C\}} \underbrace{\mathbf{K}_f^\top}_{\text{Kernel space}} \quad \mathbf{A} \quad \underbrace{\mathbf{f}_f(n)}_{\text{Feature space}}$$

With **A** the alignment matrix

Image processing: $\mathbf{A} = \mathbf{I}_{M^2}$

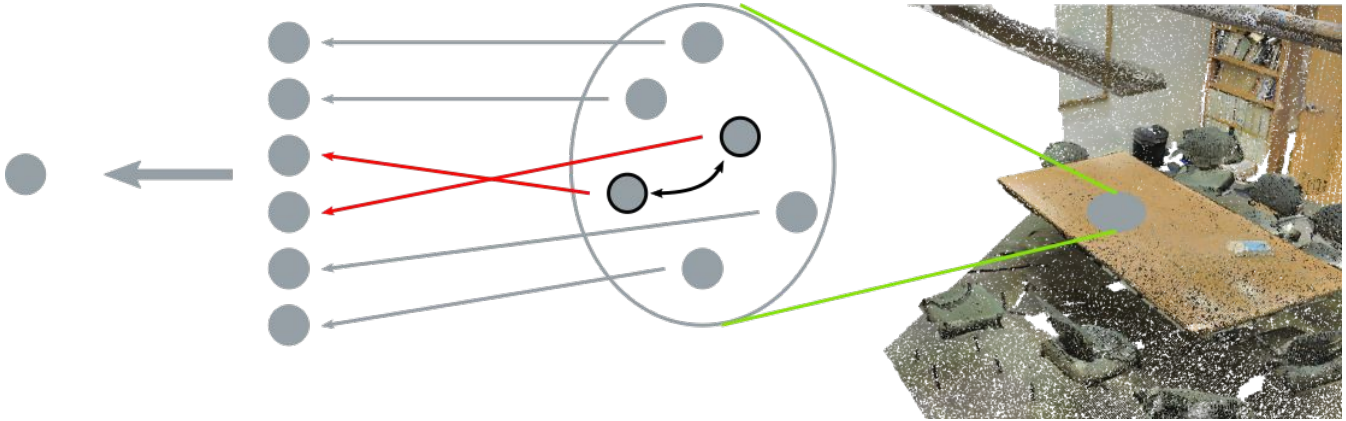


Convolution for points

Apply the same formula on a small set of points:

$$\mathbf{h}[n] = \sum_{f \in \{1,\dots,C\}} \underbrace{\mathbf{K}_f^\top}_{\text{Kernel space}} \; \mathbf{A} \; \underbrace{\mathbf{f}_f(n)}_{\text{Feature space}}$$

Problem: **A** is not permutation invariant



Convolution on points

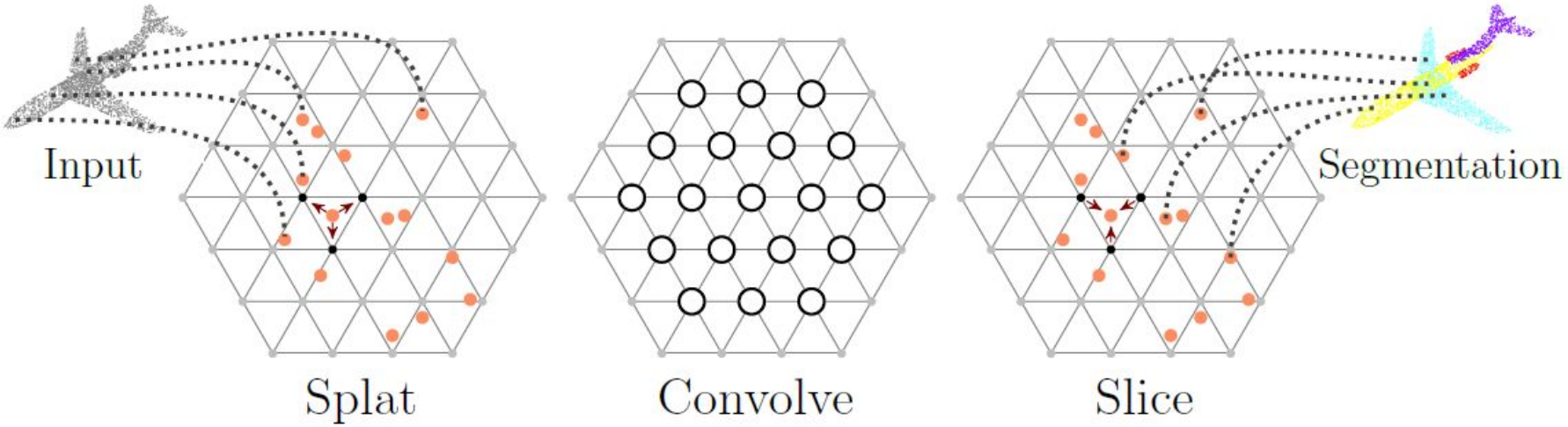
A must be estimated from the neighborhood N of n:

$$\mathbf{h}[n] = \sum_{f \in \{1,\dots,C\}} \underbrace{\mathbf{K}_f^\top}_{\text{Kernel space}} \mathbf{A}(\mathcal{N}) \underbrace{\mathbf{f}_f(n)}_{\text{Feature space}}$$

SplatNet

valeo.ai

Estimation of A: Interpolation of the features on a regular grid (barycentric



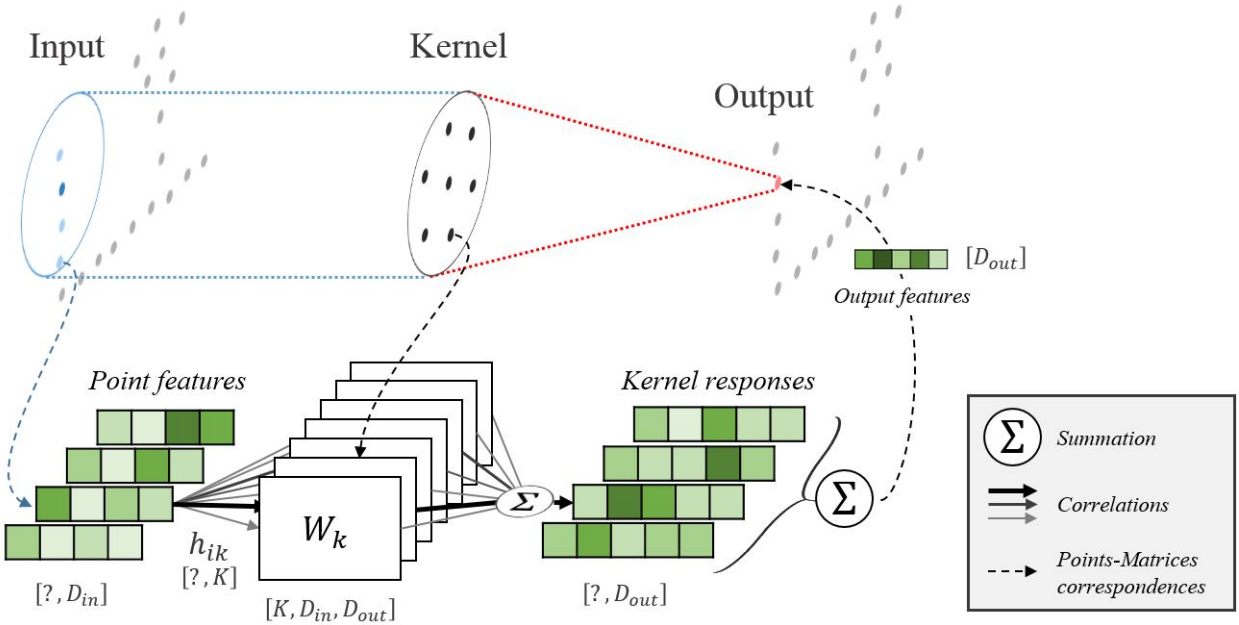
Hang Su et al. “SPLATNet: Sparse Lattice Networks for Point Cloud Processing”. In: arXiv preprint arXiv:1802.08275 (2018)

KPConv

KPConv

$$\alpha_{i,j} = \alpha(y_i, \hat{x}_j) = \max\left(0, 1 - \frac{\|y_i - \hat{x}_j\|}{\sigma}\right)$$

Estimation of A: Create kernel locations in space, weighted interpolation to all kernel
location based on distance.



Hugues Thomas et al. “Kpconv: Flexible and deformable convolution for point clouds”. In: ICCV 2029

ConvPoint

Estimation of A: Create kernel locations in space, weighted interpolation learned with MLP.

$$a_{i,j} = a(y_i, \hat{x}_j) = \mathrm{MLP}(y_i - \hat{x}_j)$$

Optimization of both MLP weights and kernel point positions.

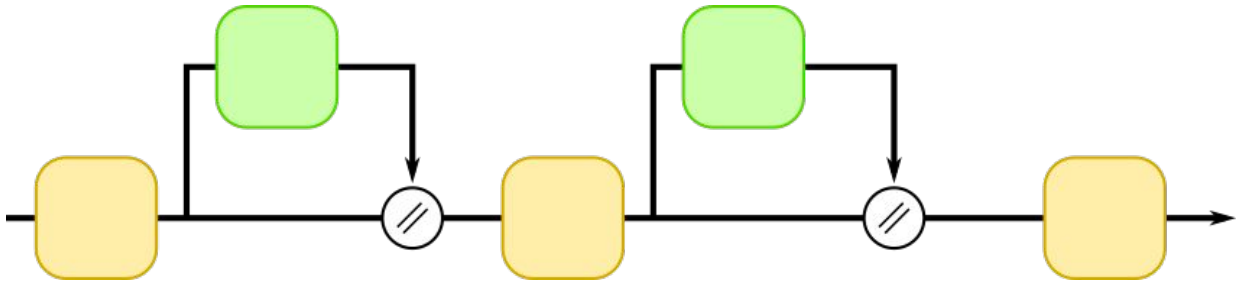
“Generalizing discrete convolutions for unstructured point clouds”. In: Computer and Graphics 2020

FKAConv

valeo.ai

Estimation of A: Direct estimation of A using a mini-PointNet.

$$a_{i,j} = a_i(\hat{x}_j) = \text{MLP}_i(\hat{x}_j, \{\hat{x}_k\}_k) \approx \text{PointNet}(\{\hat{x}_k\}_k)$$



Neighborhood search

valeo.ai

Convolution is a local operation.

- K-nearest neighbors search
- Ball search

K-nearest neighbors search

valeo.ai

Let q be the support point (center of the neighborhood):

$$\text{argtop-K}_{\mathbf{p}\in P}\{-||\mathbf{p}-\mathbf{q}||\}$$

Pros:

- All neighborhoods have the same cardinal
- Relatively fast

Cons:

- Neighborhoods scales vary

Ball search

valeo.ai

Let q be the support point (center of the neighborhood):

$$\{\mathbf{p} \in P, s.t. ||\mathbf{p} - \mathbf{q}|| < r\}$$

with r the ball radius.

Pros:

- All neighborhoods have the same scale

Cons:

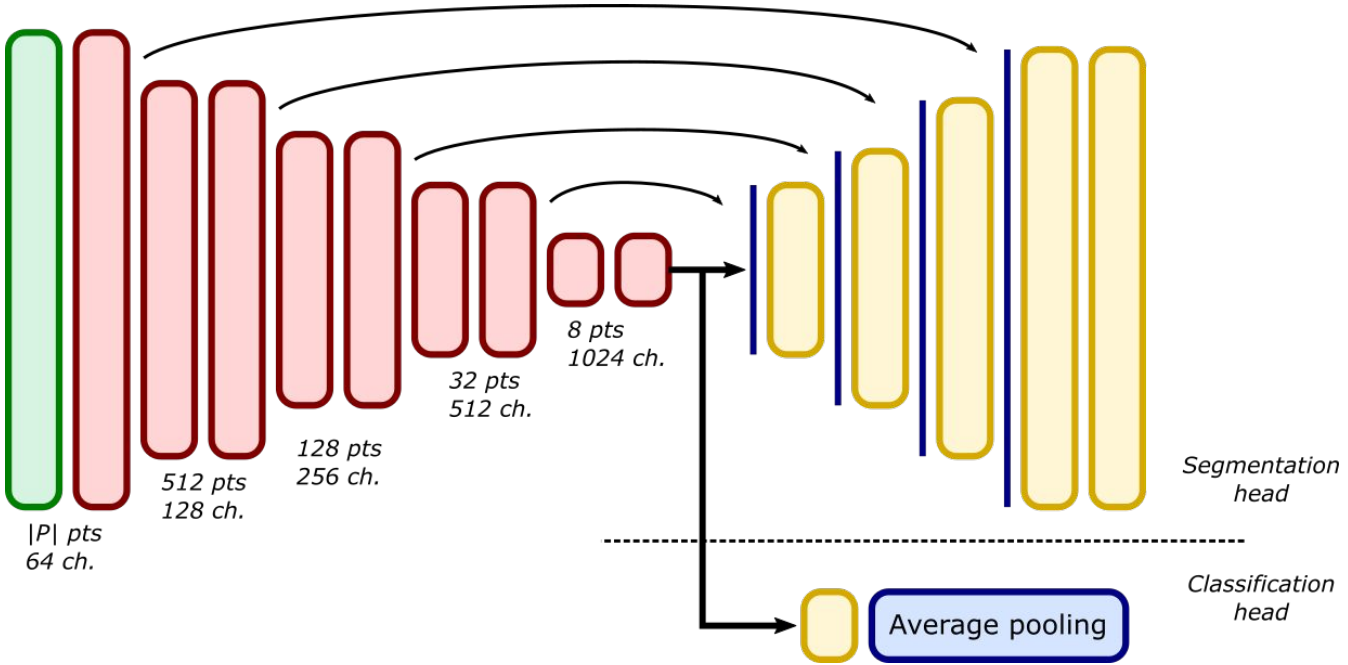
- Neighborhoods cardinals (number of points) vary
- Usually slower than K-nn

II - Geometric deep learning

B - Sampling

Progressive dimension reduction

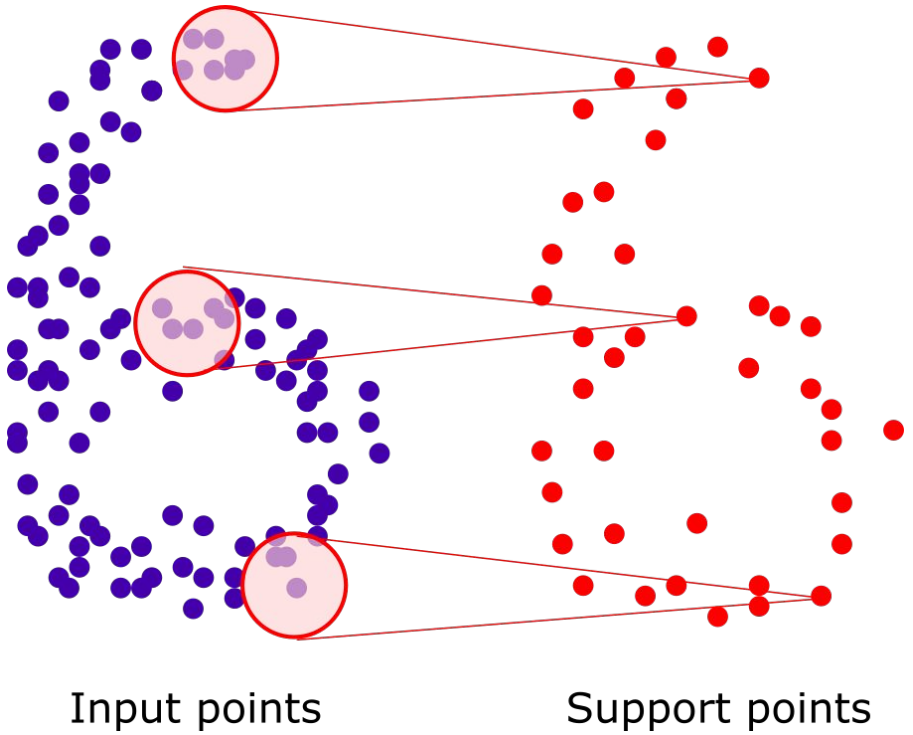
What is the equivalent of stride for convolution on points ?



Support point sampling

Q (Support points), points used as neighborhood centers for the convolution operation.

Usually Q is a subset of P



Random sampling

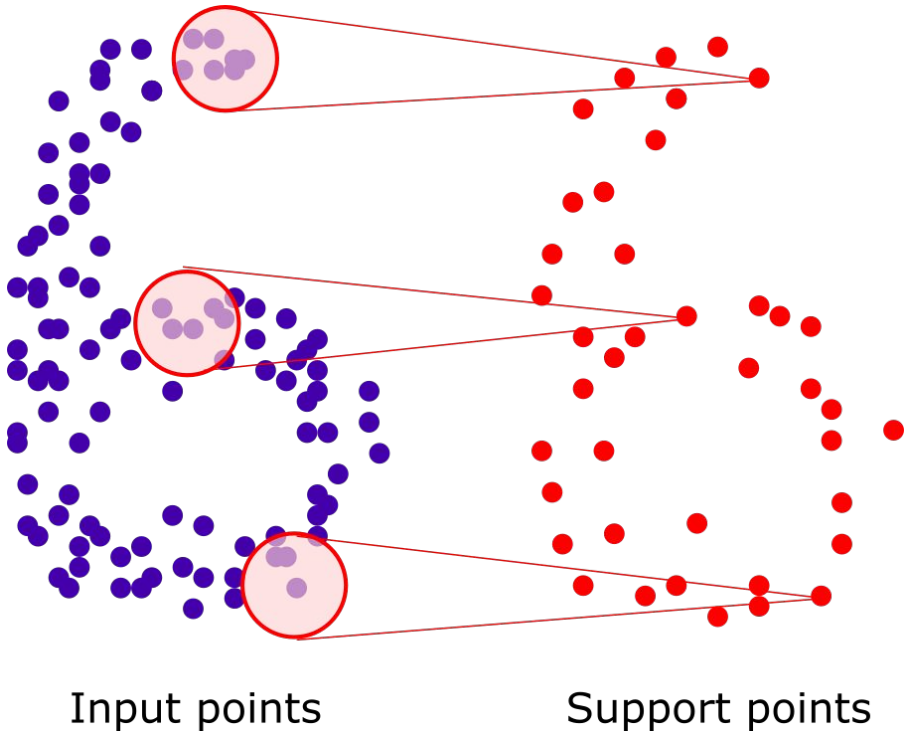
Uniform selection of the input points.

Pros:

- simple and fast.

Cons:

- loss of geometric information on area with low density or extreme points.

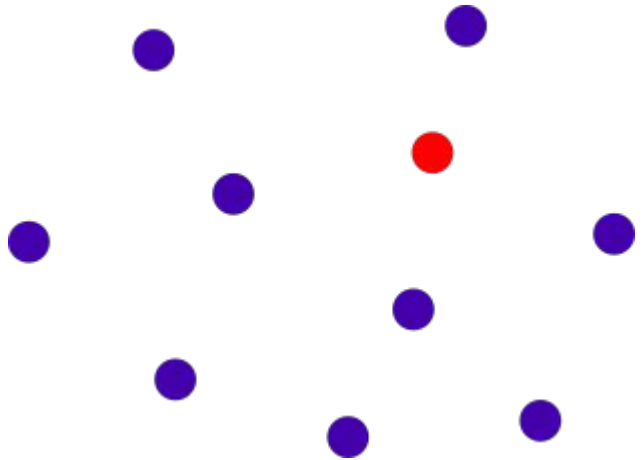


Furthest point sampling

Introduced in PointNet++: iteratively select the further point from the previously selected.

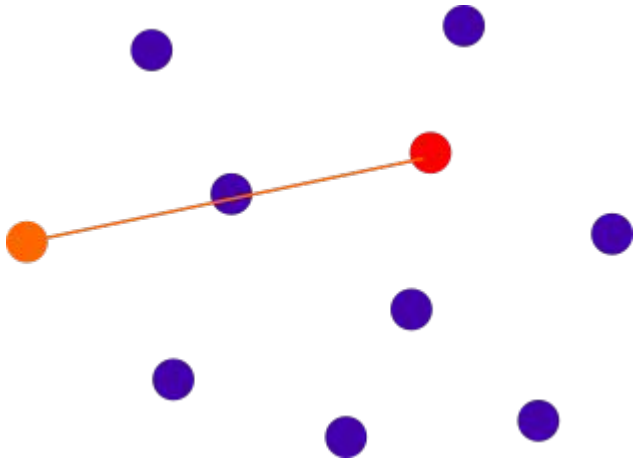
Furthest point sampling

Introduced in PointNet++: iteratively select the further point from the previously selected.



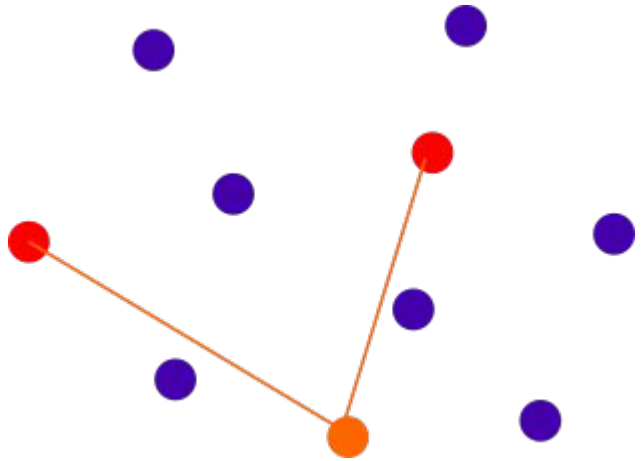
Furthest point sampling

Introduced in PointNet++: iteratively select the further point from the previously selected.



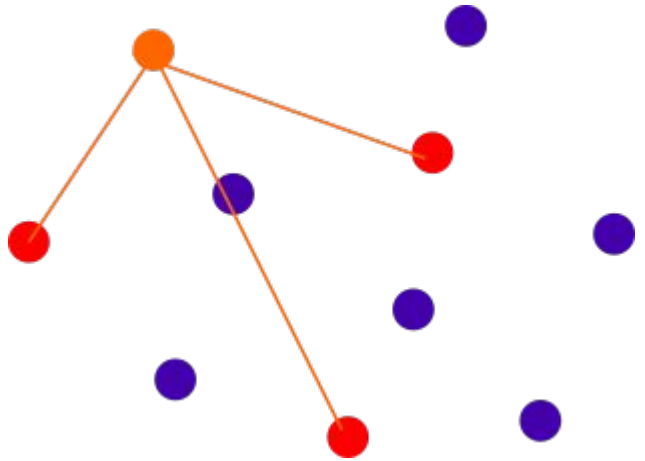
Furthest point sampling

Introduced in PointNet++: iteratively select the further point from the previously selected.



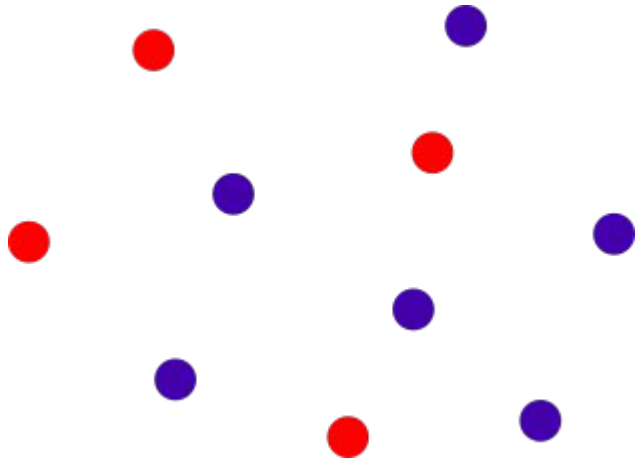
Furthest point sampling

Introduced in PointNet++: iteratively select the further point from the previously selected.



Furthest point sampling

Introduced in PointNet++: iteratively select the further point from the previously selected.



Voxel-grid sampling

valeo.ai

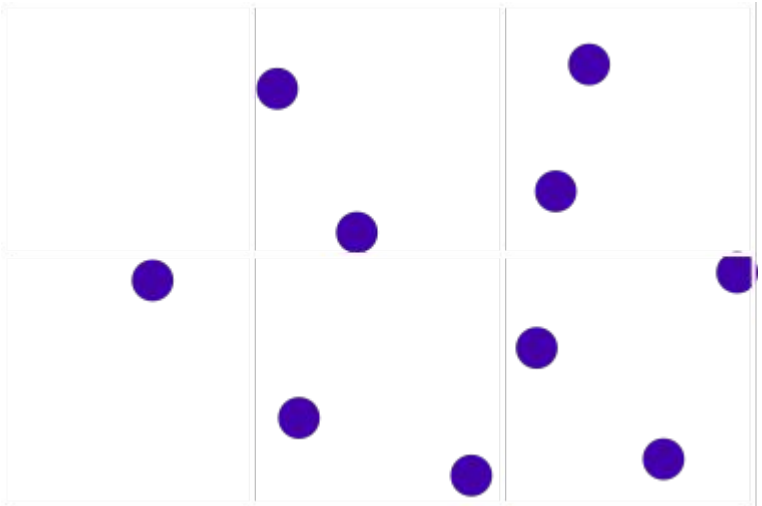
Apply a voxel pooling: select a point in each voxel.

Pros:

- fast

Cons:

- voxel size is extra parameter, may lead to variable number of points



Voxel-grid sampling

valeo.ai

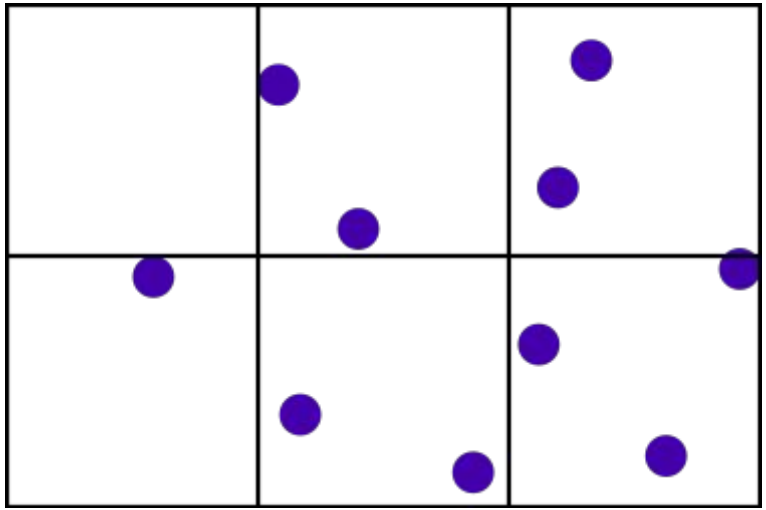
Apply a voxel pooling: select a point in each voxel.

Pros:

- fast

Cons:

- voxel size is extra parameter, may lead to variable number of points



Voxel-grid sampling

valeo.ai

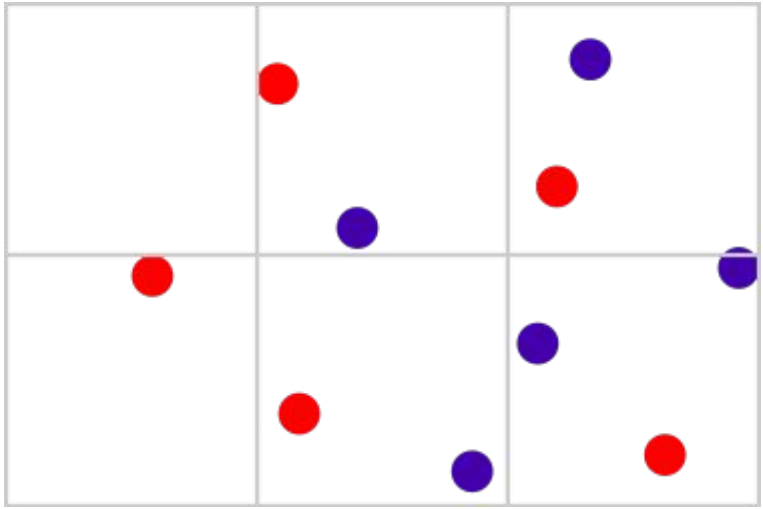
Apply a voxel pooling: select a point in each voxel.

Pros:

- fast

Cons:

- voxel size is extra parameter, may lead to variable number of points



Voxel-grid sampling

valeo.ai

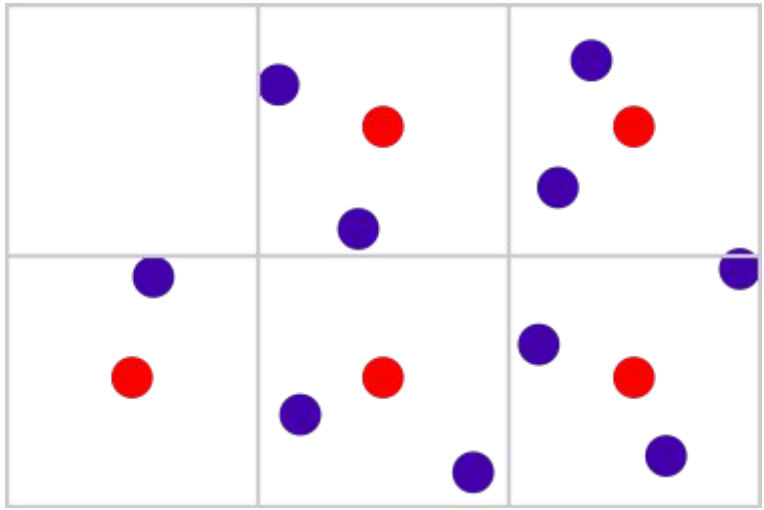
Apply a voxel pooling: select a point in each voxel.

Pros:

- fast

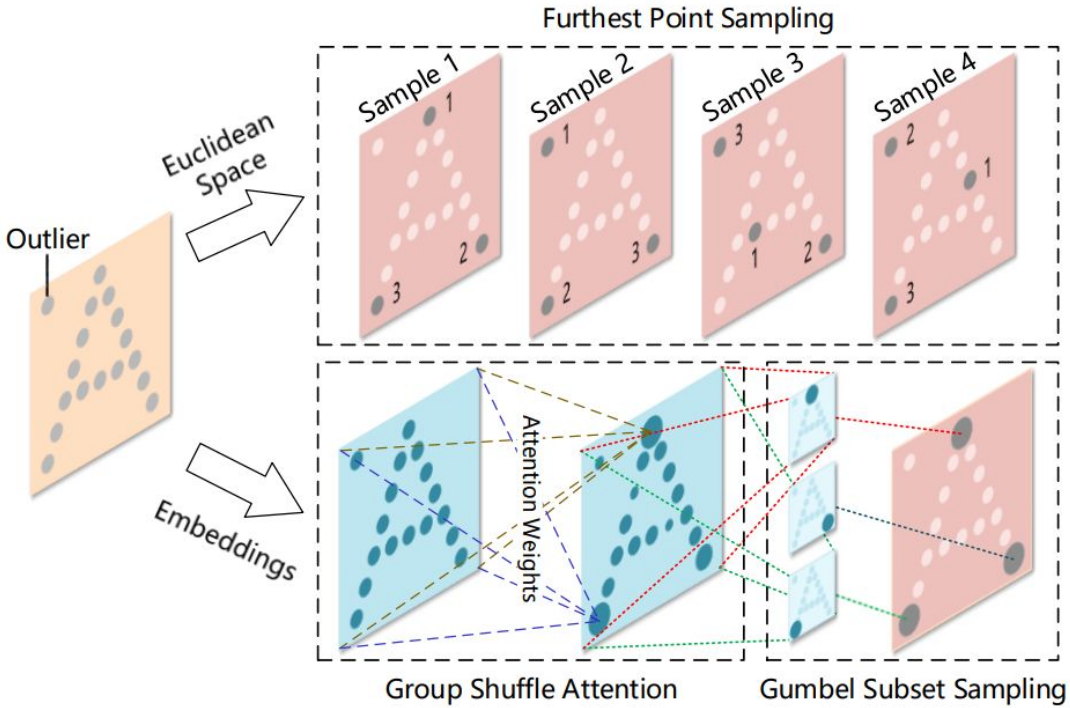
Cons:

- voxel size is extra parameter, may lead to variable number of points



Attention pooling

Learned attention on the points for outlier robustness.



Jiancheng Yang et al. “Modeling point clouds with self-attention and gumbel subset sampling”. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp. 3323–3332

Conclusion

What we didn’t talk about

valeo.ai

Image structure

→ ++ efficient / benefit from image knowledge

→ ++ if an easy image rendering

→ - - rendering can be hard

Pure graph structure

→ ++ very high performances

→ ++ evolution toward graph attention (AlphaFold)

→ - - when the graph structure is not obvious, additional computing

Practical session

valeo.ai

First neural network training

Classification on ModelNet40

→ PointNet (light)

→ DGCNN (light)

We still run them CPU !