Perception (Vision)

Romain Cazorla

Lab-STICC
Segula Technologies
ENIB
IMT - Atlantique

romain.cazorla@segula.fr

25/10/2022

Outlines

Introduction

- Overview of perception tasks.
- Presentation of general workflow for detection and segmentation.
- Presentation of deep learning applied to 3D point cloud.

Sensor	Output	Range	Strength	Weakness
Camera	Image	$<$30m	Precise, easily available	Lack distance information
RGB-D Camera	RGB-D Image	$<$3m	Cheap	Short range
LiDAR	Point cloud	$>$5m	Precise, High resolution	Cost

Table: Sensors comparison[3]

Methods allowing transformation between different output exists (ex: image to point cloud). Beware, transformation often induces a lack of information.

Sensors and data

Data

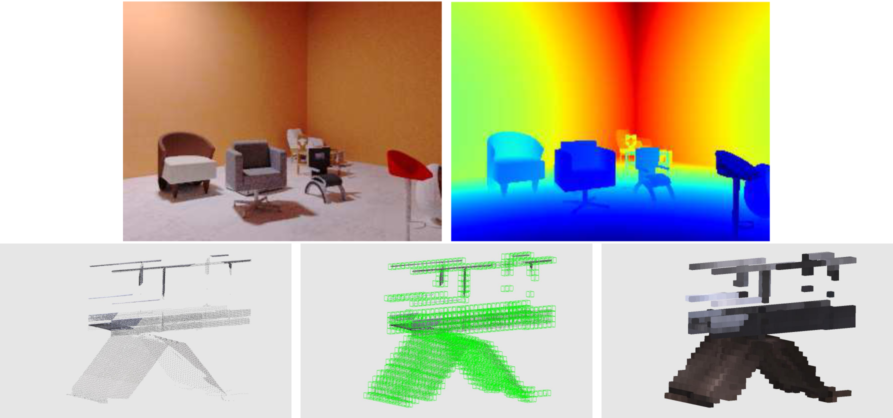


Figure: First line : Pictures and associated depth canal [5] Second line : Point cloud, voxelising, voxels (voxel size : 0.1m)

Deep learning is a subset of machine learning, focused on neural network "with more than two layers" with the following characteristics :

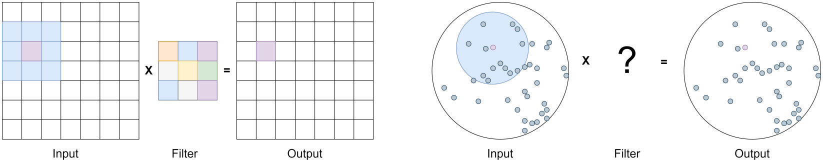
- More neurons than previous networks
- More complex ways of connecting layers and neurons
- Higher need for computational power
- Automatic feature extraction

Deep Learning A Practitioner's Approach, Patterson J. and Gibson A. [7]

Sensors and data

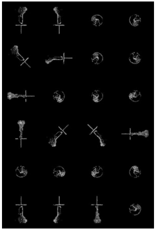
Point cloud specificity

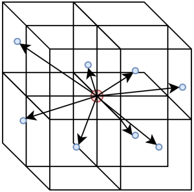
- Automatic feature extraction as a pillar of deep learning.
- Convolution as a feature extraction method.



How to extract features from point clouds, an irregular, unordered data format ?

- Point Cloud Feature Extraction
  - Projection-based
    - 2D Projections[2][12]
    - Voxels[6]
  - Point-based
    - MLP[8][9]
    - Convolution[10]
    - Graph[11]





Point cloud semantic segmentation methods taxonomy[4]

Perception tasks

- Detection : Finding object(s) of interest in perceived data.
- Segmentation : Dividing perceived data in its different component.

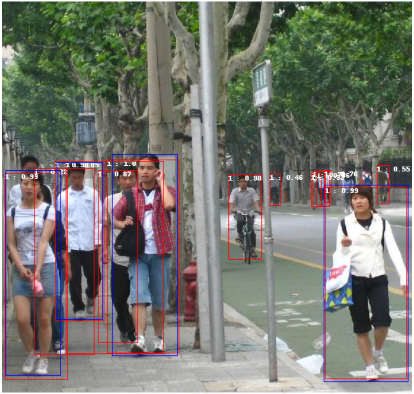
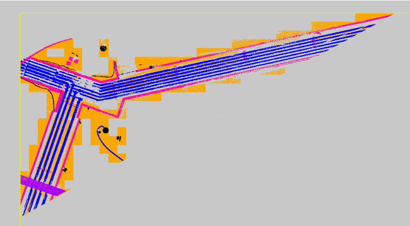



Figure: Detection and segmentation examples.

Perception techniques

Metrics

- True Positive (TP).
- True Negative (TN).
- False Positive (FP).
- False Negative (FN).

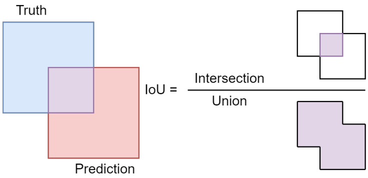
$$Precision = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{TP + TN}$$

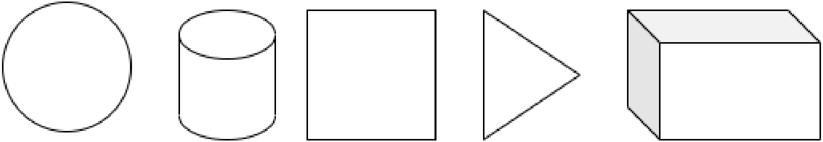
$$Average\ Precision = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Intersection\ over\ Union = \frac{TP}{TP + FP + FN}$$

For detection, others metrics exist and should be studied on a case by case basis depending on the dataset used.



- Detection : Finding object(s) of interest in perceived data.
- What to do in the case of simple shapes ?



- Model-fitting : detecting a known object.
- Work well in controlled environment.
- Use a generalised Hough transform or RANSAC.

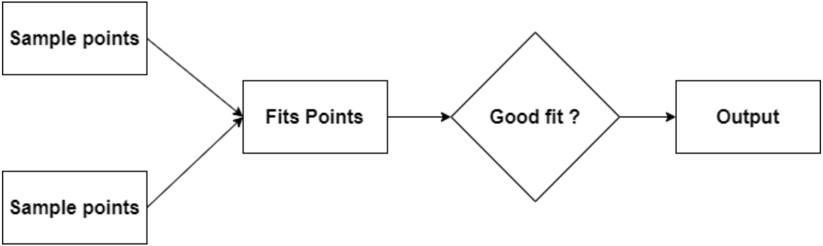


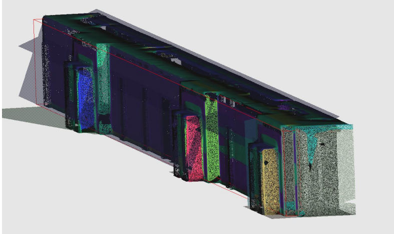
Figure: Model-fitting workflow

Primitive, such as plane or cylinder, can also be detected with classical methods, such as RANSAC.

Raw cloud

$\longrightarrow$

Fitted primitive



Detection

Deep learning methods [3]

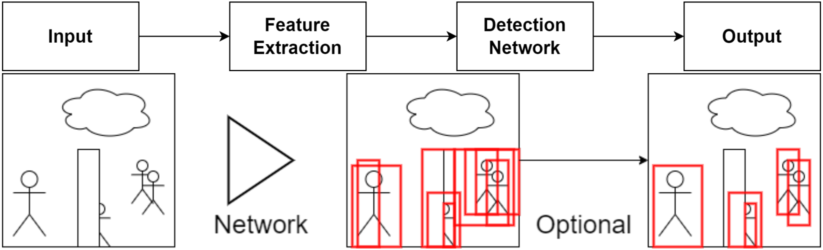
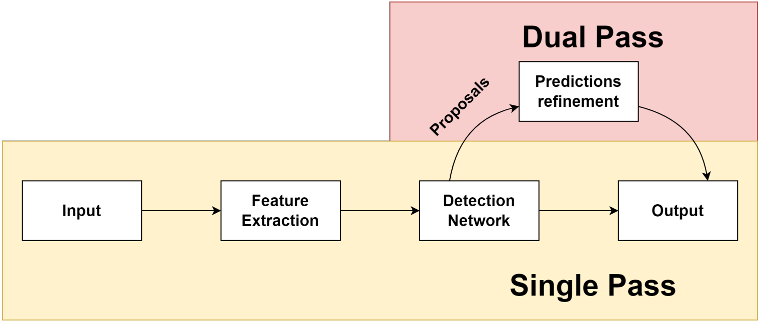


Figure: General detection workflow.

Fast
Single Pass ⟷ **Dual Pass**
Precise



→ **Features**

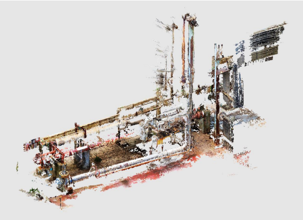
↓

←

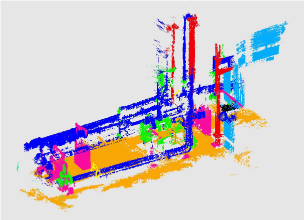
Segmentation

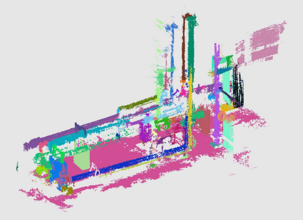
Types of segmentation

Raw point cloud

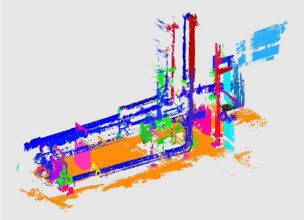


Semantic Segmentation





Instance Segmentation

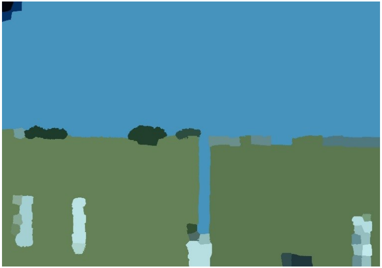


Panoptic Segmentation

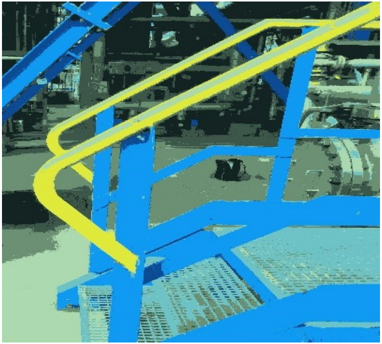
1. State of the art use deep learning.
2. Simple cases of instance segmentation can use traditional methods [1].

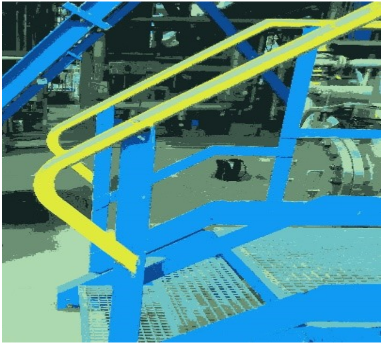
Segmentation

Traditional Methods-Graph Cut Normalised



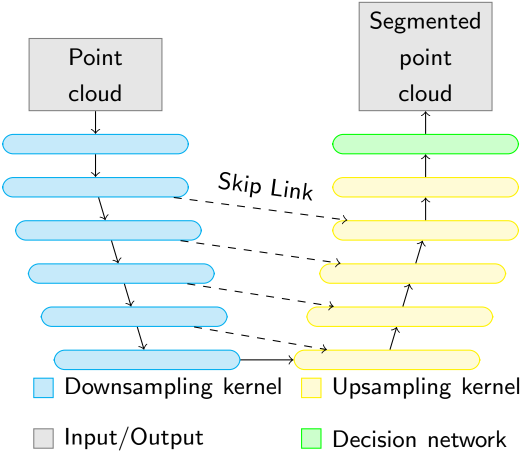
Sur-segmentation → Graph construction → Graph Cut





- Optimisation method.
- Difficulty in fixing the number of classes.

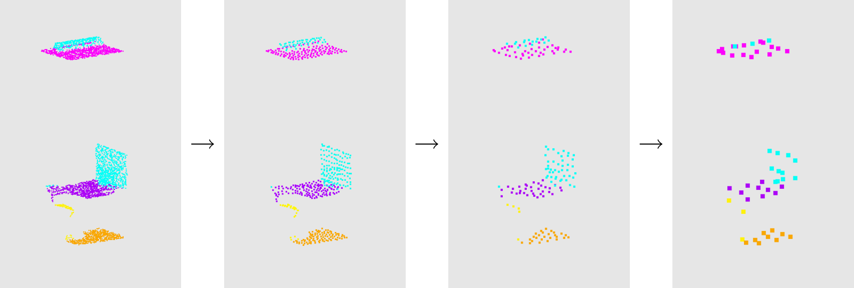
- Optimisation method.
- Consider image as a topological surface.
- Sensible to initial conditions.



- Negative log likelihood loss is used for training.
- Downsampling can use a mix of convolutional and pooling layer.
- Upsampling can use a mix of convolutional and deconvolutional layer.
- Skip link preserve high resolution information.

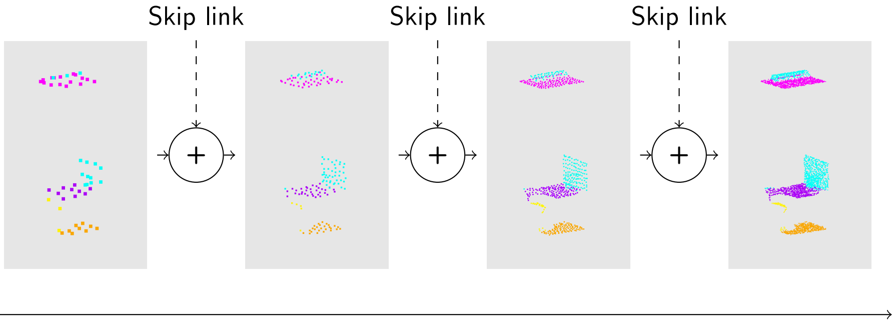
Downsampling

Increasing number of features
Decreasing data resolution



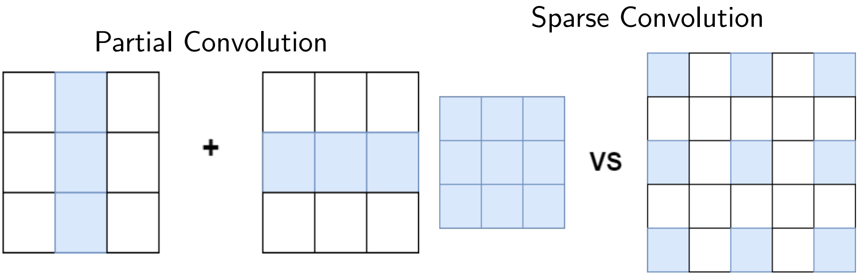
Upsampling

Injection of data lost during downsampling



Increasing data resolution
Decreasing number of features

Real time perception



- Real time application need to make decision in less than 0.1s.
- Simple data structure is prevalent in the 3D case.
- Tracking is often used to improve performance along several frame.

Sensor Fusion

- Different data bring different kind of information.
- Sensor fusion try to take advantage of information provided by different sensors.

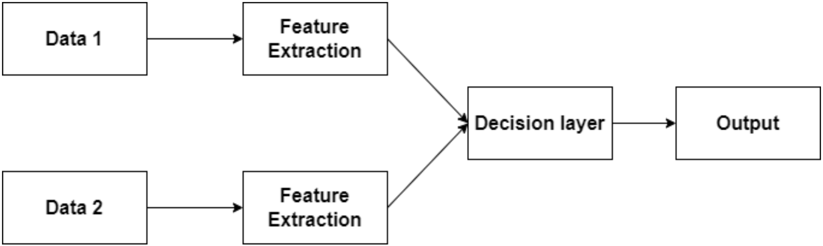


Figure: Common sensor fusion workflow

References I

[1] Dhouha Attia. "Segmentation d'images par combinaison adaptative couleur-texture et classification de pixels. : Applications à la caractérisation de l'environnement de réception de signaux GNSS". PhD thesis. Université de Technologie de Belfort-Montbeliard, Oct. 3, 2013.

[2] Alexandre Boulch et al. "SnapNet: 3D Point Cloud Semantic Labeling with 2D Deep Segmentation Networks". In: *Computers and Graphics* 71 (2018), pp. 189–198. DOI: 10.1016/j.cag.2017.11.010.

[3] Duarte Fernandes et al. "Point-Cloud Based 3D Object Detection and Classification Methods for Self-Driving Applications: A Survey and Taxonomy". In: *Information Fusion* 68 (Apr. 1, 2021), pp. 161–191. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2020.11.002.

[4] Yulan Guo et al. "Deep Learning for 3D Point Clouds: A Survey". In: *IEEE TPAMI 2020* (2020). arXiv: 1912.12033v1.

[5] John McCormac et al. "SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-training on Indoor Segmentation?" In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, Oct. 2017, pp. 2697–2706. ISBN: 978-1-5386-1032-9. DOI: 10.1109/ICCV.2017.292.

[6] Hsien-Yu Meng et al. "VV-Net: Voxel VAE Net with Group Convolutions for Point Cloud Segmentation". In: *IEEE/CVF International Conference on Computer Vision (ICCV)*. International Conference on Computer Vision (ICCV). Oct. 2019, pp. 8499–8507. DOI: 10.1109/ICCV.2019.00859. arXiv: 1811.04337.

[7] Josh Patterson and Adam Gibson. *Deep Learning A Practitioner's Approach*. O'Reilly. Aug. 2017. ISBN: 978-1-4919-1425-0.

[8] Charles Ruizhongtai Qi et al. "PointNet : Deep Learning on Point Sets for 3D Classification and Segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 652–660.

[9] Charles Ruizhongtai Qi et al. "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space". In: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. Neural information processing systems foundation, June 7, 2017, pp. 5105–5114. arXiv: 1706.02413.

References II

[10] Hugues Thomas et al. "KPConv: Flexible and Deformable Convolution for Point Clouds". In: *Proceedings of the IEEE International Conference on Computer Vision*. Vol. 2019-Octob. Institute of Electrical and Electronics Engineers Inc., Oct. 1, 2019, pp. 6410–6419. ISBN: 978-1-72814-803-8. DOI: 10.1109/ICCV.2019.00651. arXiv: 1904.08889.

[11] Yue Wang et al. "Dynamic Graph Cnn for Learning on Point Clouds". In: *ACM Transactions on Graphics*. Vol. 38. 5. Association for Computing Machinery, Oct. 1, 2019. DOI: 10.1145/3326362. arXiv: 1801.07829.

[12] Bichen Wu et al. "SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud". In: Proceedings - IEEE International Conference on Robotics and Automation. Institute of Electrical and Electronics Engineers Inc., Sept. 10, 2018, pp. 1887–1893. ISBN: 978-1-5386-3081-5. DOI: 10.1109/ICRA.2018.8462926. arXiv: 1710.07368.