

A survey on deep learning techniques in image and video semantic segmentation (paper analysis)

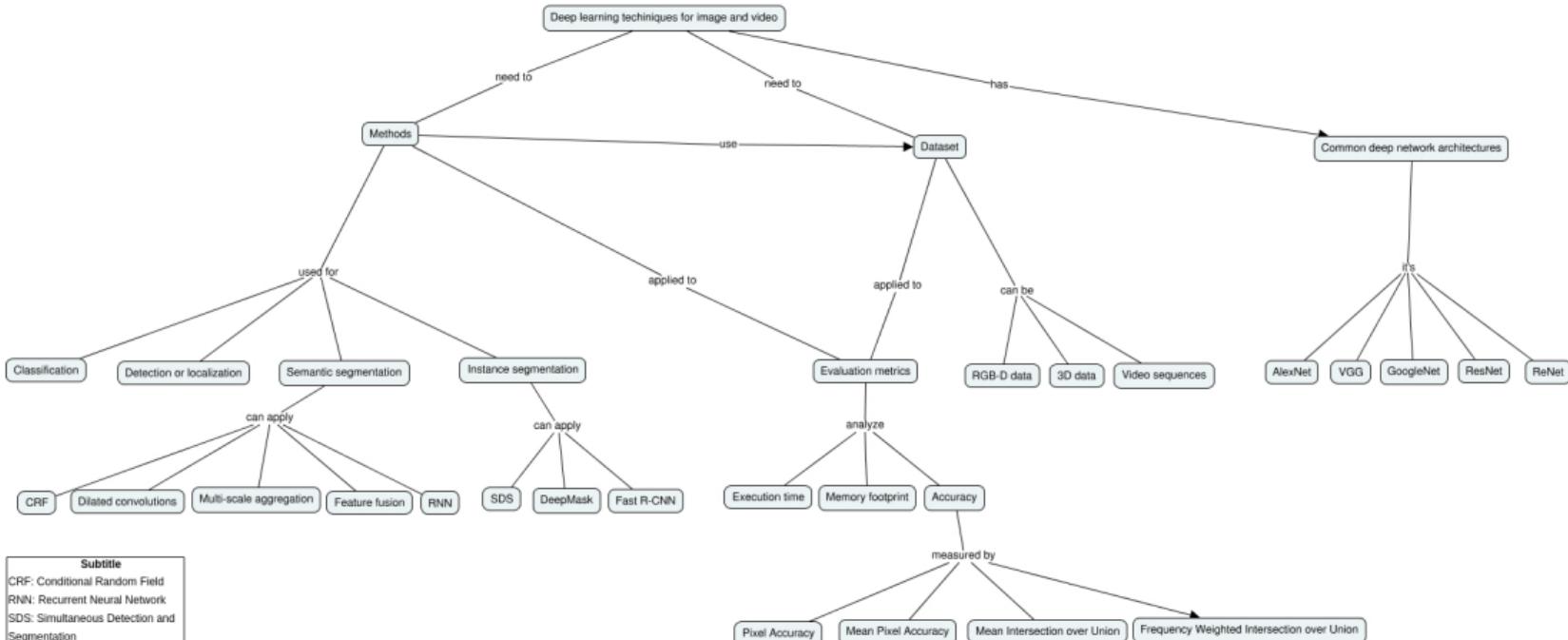
Jéssica Motta

SENAI CIMATEC

June, 2020



Concept Map



CNN- How it works?

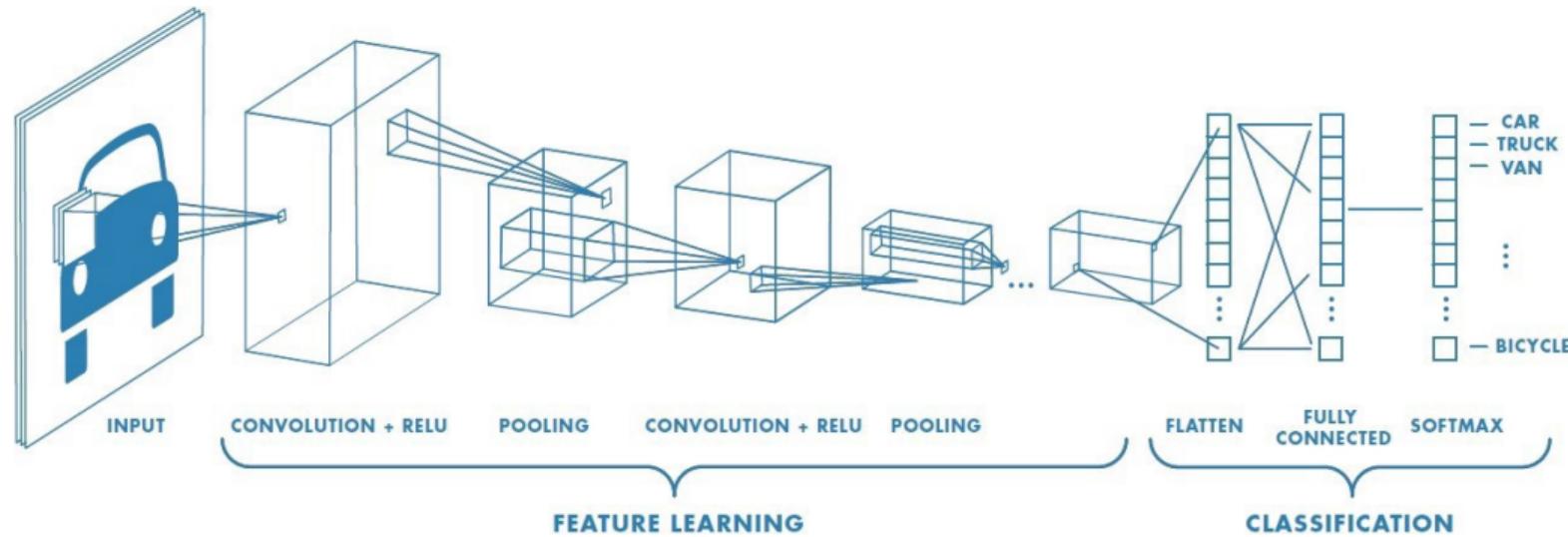


Figure: Convolutional Neural Network [Sha19]

Common deep networks architecture

COMPARATIVE FOR COMMON DEEP NETWORK ARCHITECTURES			
Network	Year champion ILSVRC*	Number of Layers	Accuracy
AlexNet	2012	3	84.6%
VGG	2013	16	92.7%
GoogleNet	2014	22	93.3%
ResNet	2016	152	96.4%

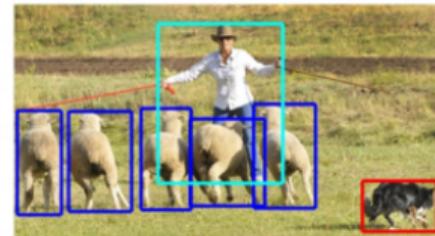
Table: Deep network architectures. [GGOEO⁺18]

*ILSVRC (ImageNet Large Scale Visual Recognition Challenge)

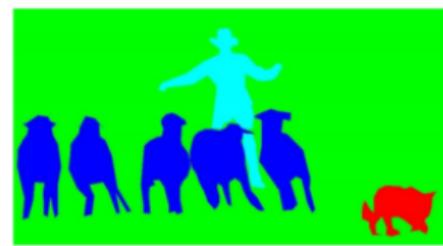
Methods to image analysis



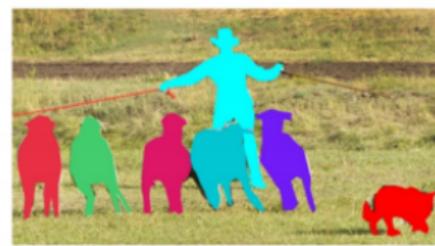
(a) Image classification



(b) Object localization



(c) Semantic segmentation



(d) Instance segmentation

Figure: Methods to image analysis. [LMB⁺14]

Semantic Segmentation applied in video



Semantic Segmentation [Zha17]

Semantic Segmentation

Use **ImageNet** pre-trained networks

Smaller dataset, even **worse** for RGB-D and 3D datasets

Use **synthetic datasets** extracted from commercial video games



Evaluation Metrics

Execution time

Memory footprint

Accuracy



Accuracy- IoU calculation

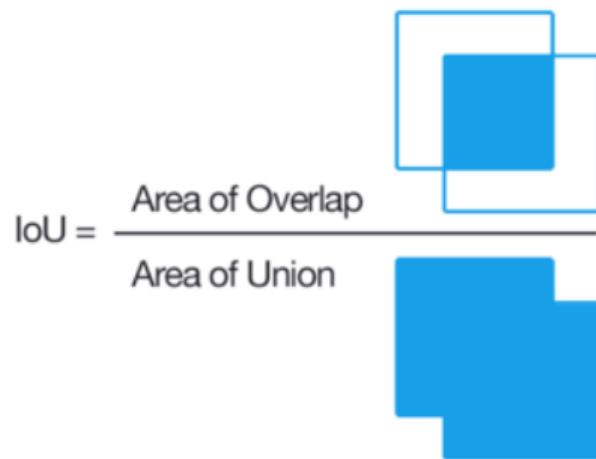


Figure: Accuracy evaluation. [Ros16]

Figure: IoU calculation visualized. [Tiu19]

Which cases doesn't apply deep learning?

For **high performance**, deep networks require **extremely large** datasets.

It's **expensive** to get data, computer power and hiring researchers.

Deep networks **aren't easily interpreted** as classical Machine Learning algorithms.



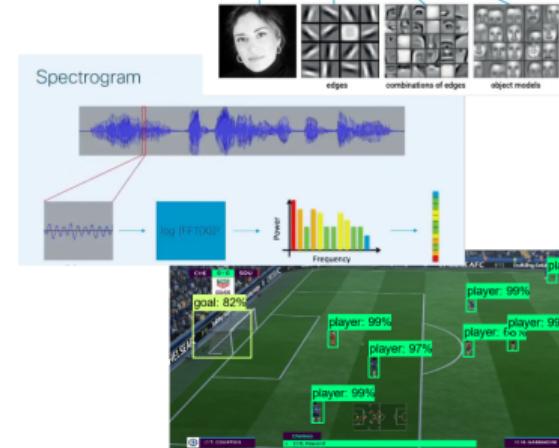
Advantages to use Deep learning against to Classical methods

Better performance (accurate results)

More data work with DP **but not** with CML

DP **isn't needed** for feature engineering

It's **adaptable** to different domains and applications



Advantages to use Classical methods against to Deep learning

Works better with **small dataset**

Low computational and financial cost

The algorithms are **easier** to understand and interpret

Conclusion

A survey to **help** researchers to choose the **best method and dataset** that are appropriate for their application

Semantic segmentation has been approached with many **success** stories but has some **problems** whose solution would be useful for real-world applications

And **deep learning** has proved to be extremely **powerful** to solve these problems

Future research directions

- 3D datasets
- Sequence datasets
- Point cloud segmentation using Graph Convolutional Networks (GCN)
- Real-time segmentation
- Memory
- Temporal coherency on sequences
- Multi-view integration

References I

-  Alberto Garcia-Garcia, Sergio Orts-Escalano, Sergiu Oprea, Victor Villena-Martinez, Pablo Martinez-Gonzalez, and Jose Garcia-Rodriguez, *A survey on deep learning techniques for image and video semantic segmentation*, Applied Soft Computing **70** (2018), 41–65.
-  Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, *Microsoft coco: Common objects in context*, European conference on computer vision, Springer, 2014, pp. 740–755.
-  Adrian Rosebrock, *Intersection over union (iou) for object detection*, 2016.
-  Shashikant, *Convolutional neural network: A step by step guide*, 2019.
-  Ekin Tiu, *Metrics to evaluate your semantic segmentation model*, 2019.

References II

-  Hengshuang Zhao, *Icnet for real-time semantic segmentation on high-resolution images*, 2017.