



How do robots see?

An introduction to semantic segmentation in images

Alano Martins Pinto

Computer vision research leader

email: alano_pinto@atlantico.com.br

fone: +5585999509376

Paulo Bruno Serafim

Software Development Engineer

email: paulo.bruno@hp.com

phone: +5585999972923



Atlântico

1.

Segmentation





Segmentation

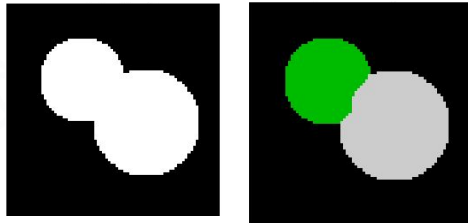
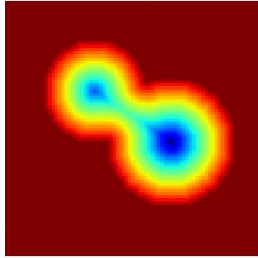
Segmentation is an important part of image analysis. It refers to the process of partitioning an image into multiple segments.

More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

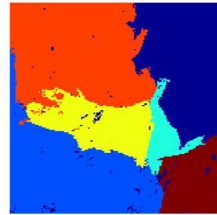
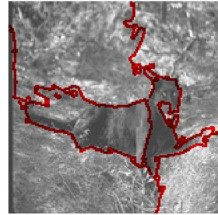
The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse.



Methods



Topology
(Watershed)



Clustering
(C-Means)



Threshold - Histogram
(Otsu)

2.

Autoencoders

Neural Networks



Convolution Kernel

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

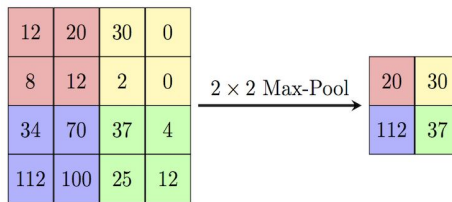


CNN

During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot product between the entries of the filter and the input and producing a 2-dimensional activation map of that filter.

Neural Network

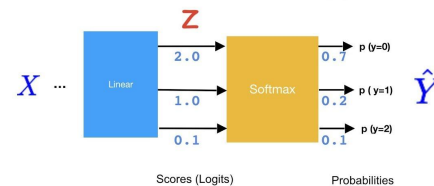
Brief introduction



Max-Pooling

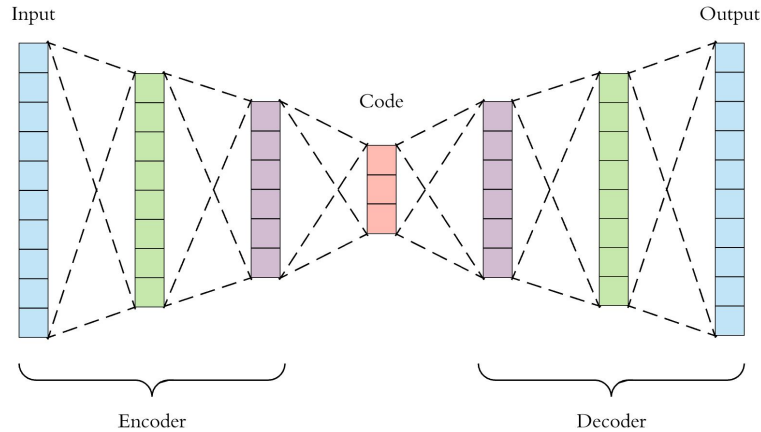
Max pooling is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.

Meet Softmax $\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$ for $j = 1, \dots, K$.



Softmax

Softmax function, also known as softargmax or normalized exponential function, is a function that takes as input a vector of K real numbers, and normalizes it into a probability distribution consisting of K probabilities proportional to the exponentials of the input numbers.



Autoencoders

An autoencoder is a neural network that is trained to attempt to copy its input to its output. Internally, it has a hidden layer h that describes a code used to represent the input.

The network may be viewed as consisting of two parts: an encoder function: $\mathbf{h} = \mathbf{f}(\mathbf{x})$ and a decoder that produces a reconstruction: $\mathbf{r} = \mathbf{g}(\mathbf{h})$





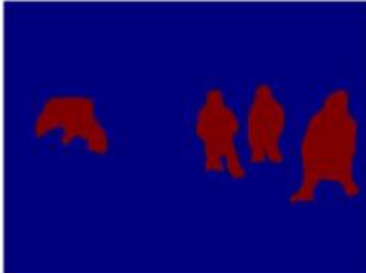
CODE



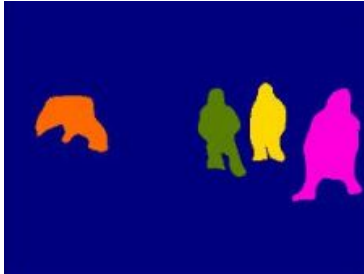
Segmentation types



Original



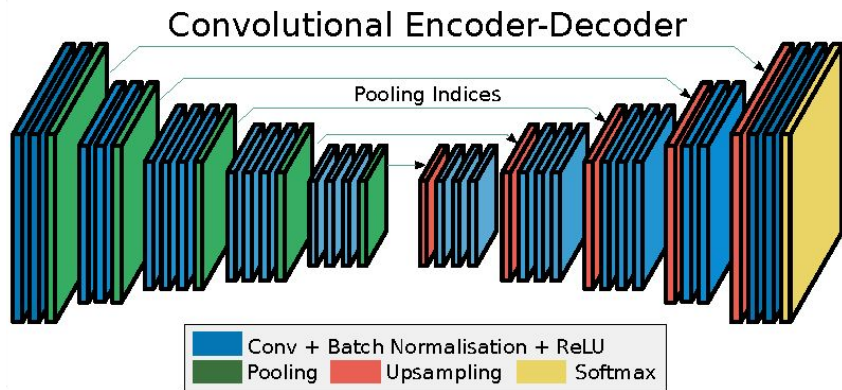
Class based



Instance based



Context based



SegNet

SegNet, is designed to be an efficient architecture for pixel-wise semantic segmentation. It is primarily motivated by road scene understanding applications which require the ability to model appearance (road, building), shape (cars, pedestrians) and understand the spatial-relationship (context) between different classes such as road and sidewalk





CODE





Atlântico
Antecipar o futuro. Esse é o nosso negócio.

Thank you!

SEDE FORTALEZA

Salinas Shopping
Av. Washington Soares, 909, Lojas
42, 43, 44 e 45 - Edson Queiroz
CEP: 60811-341 Fortaleza - CE

Fone: +55 (85) 3216-7800
Fax: +55 (85) 3216-7864

UNIDADE SÃO PAULO

Alphaville Comercial
Alameda Araguaia, 933
Alphaville
CEP: 04752-005 - São Paulo - SP

Fone/Fax: +55 (11) 3506-3700



Atlântico
Antecipar o futuro. Esse é o nosso negócio.