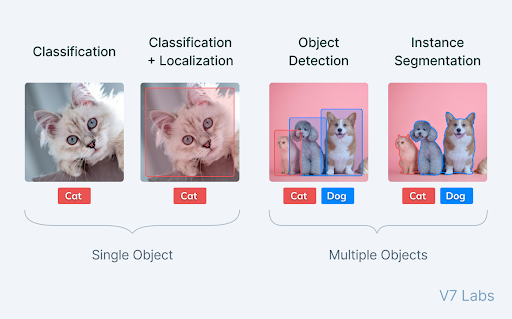
**Image segmentation**

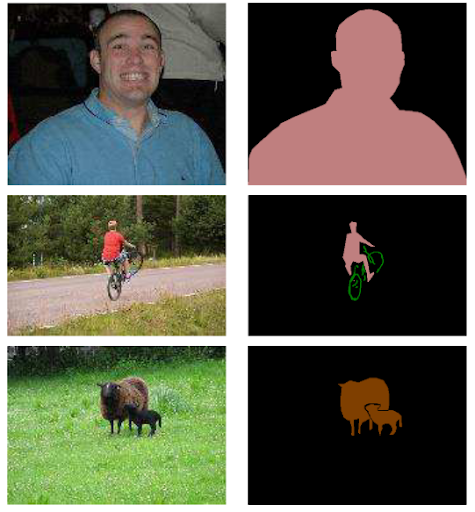
1. What is Image Segmentation
2. Annotation for Image Segmentation
3. Types of Image Segmentation tasks
4. Traditional Image Segmentation techniques
5. Deep Learning-based methods
6. Applications of Image Segmentation

**What is Image Segmentation?**

* A sub-domain of computer vision and digital image processing which aims at grouping similar regions or segments of an image under their respective class labels.
* A superset of image classification with the model pinpointing where a corresponding object is present by outlining the object’s boundary.



* Skill levels: Classification < Detection < Segmentation
* In CV, most image segmentation models consist of an encoder-decoder network as compared to a single encoder network in classifiers.
* The encoder encodes a latent space representation of the input which the decoder decodes to form segment maps, or in other words maps outlining each object’s location in the image.



**A typical segment map**

**Annotation for Image Segmentation**

1. **Polygon Annotation**

Polygon Annotation allows us to annotate segment masks maps by setting up waypoints throughout the boundaries of objects which the model has to segment.

**Strength** - These boundaries help us to form a polygonal region that we can treat as the segment map for a particular object.

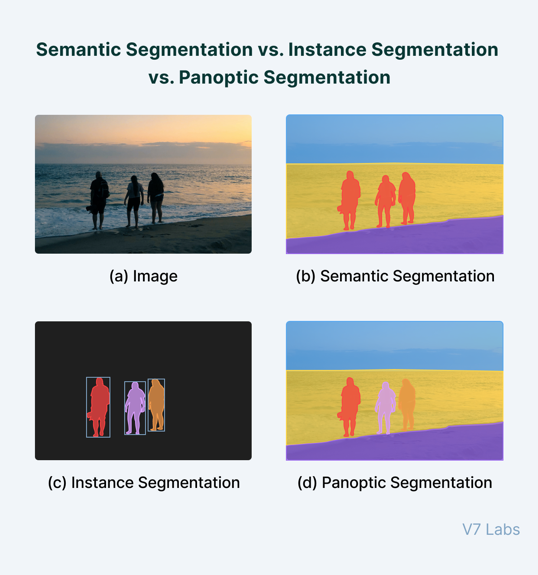
**Weakness** – This form of annotation lacks precision and can be done where the objects are mostly polygonal, or high precision is not of the utmost need.

1. **Auto-annotate**

Auto-annotate can help create high-precision segment maps very quickly for delicate and important use cases like self-driving cars and medical imaging.

* The type of annotation required and the precision needed varies according to model use cases and segmentation maps.
* Annotated datasets for tasks like Semantic segmentation are easy to build while annotations for Instance and Panoptic segmentation are harder as they require to consider overlaps between objects.

**Types of Image Segmentation tasks**

1. **Semantic segmentation**

It segments out a broad boundary of objects belonging to a particular class.

**Weakness** – It is a poorly defined problem statement when there are closely grouped multiple instances of the same class in the image. It provides very little in-depth detail or information on the image.

1. **Instance segmentation**

It provides a segment map for each object it views in the image, without any idea of the class the object belongs to. Thus, these models classify pixels into categories on the basis of “instances” rather than classes.

**Weakness** – This algorithm has no idea of the class a classified region belongs to

**Strength** – It can segregate overlapping or very similar object regions on the basis of their boundaries.

1. **Panoptic segmentation**

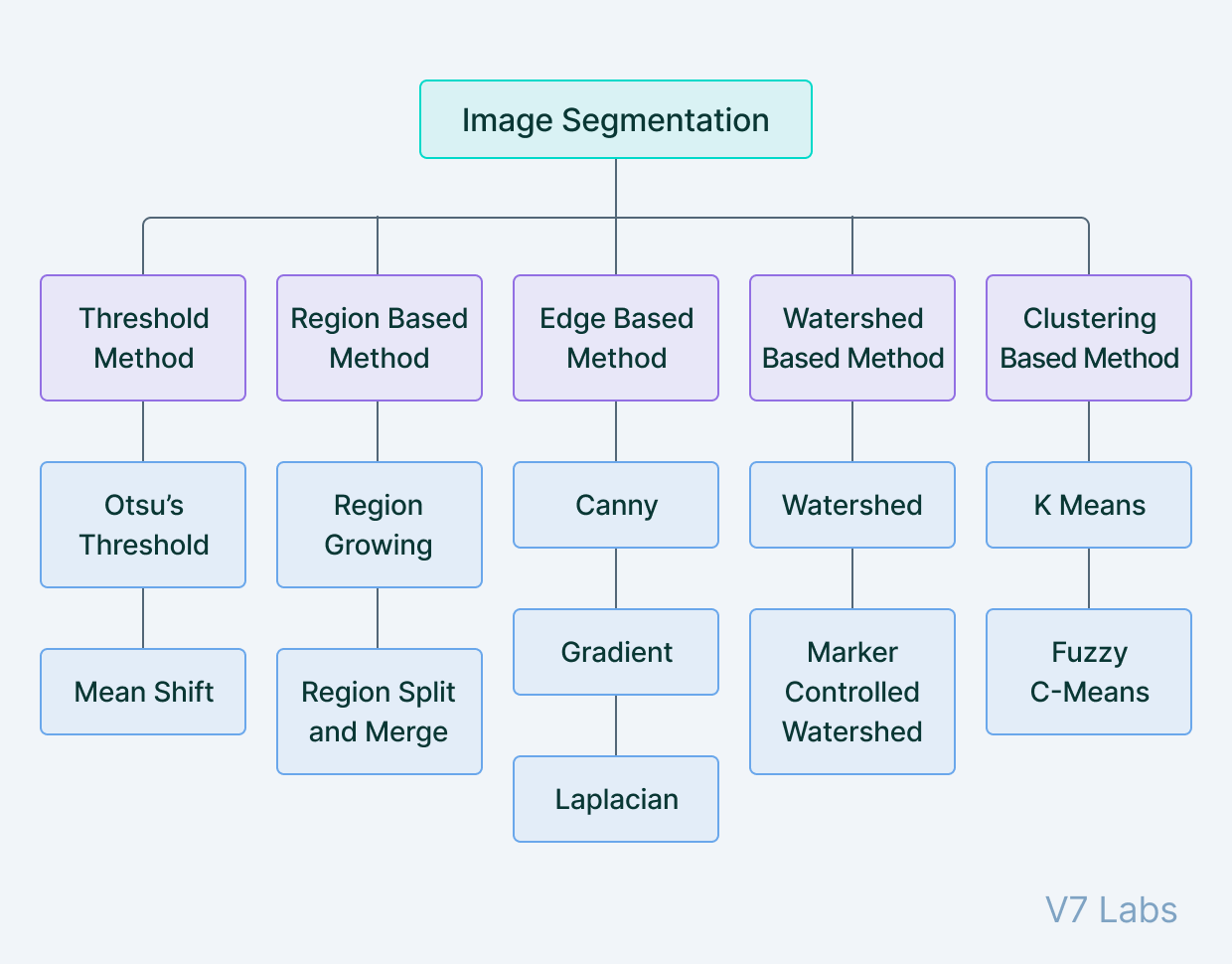
It is by far the most informative, being the conjugation of instance and semantic segmentation tasks.

Panoptic segmentation gives us the segment maps of all the objects of any particular class present in the image.

**Strength** – These algorithms find large-scale applicability in popular tasks like self-driving cars where a huge amount of information about immediate surroundings must be captured with the help of a stream of images.

**Traditional Image Segmentation techniques**

* In traditionally, image segmentation originally started from Digital Image Processing coupled with optimization algorithms. And these primitive algorithms took a local view of the features in an image and focused on local differences and gradients in pixels.
* Algorithms that took a global view of the input image came much later on with methods like **adaptive thresholding**, **Otsu’s algorithm**, and **clustering algorithms** being proposed amongst classical image processing methods.



1. **Thresholding**

A threshold is set for dividing pixels into two classes.

Set to “1” : Pixels value > Threshold value

Set to “0” : Pixels value < Threshold value

The image is thus converted into a binary map, resulting in the process often termed binarization.

**How can I set up the threshold?**

* + Histogram of cumulative distribution
  + Otsu Algorithm (“A threshold selection method from gray-level histograms”)
  + Globalizing or Localizing the input images

**Where can I use this method?**

* + The difference in pixel values between the two target classes is very high.
  + It is easy to choose an average value as the threshold.

Thresholding is often used for image binarization so that further algorithms like contour detection and identification that work only on binary images can be used.

A picture containing snow, tree, outdoor, skiing

Description automatically generatedA picture containing tree, snow, outdoor, snowboarding

Description automatically generated

1. **Region-Based Segmentation**

Looking for similarities between adjacent pixels and grouping them under a common class.

1. Start with some pixels set as seed pixels
2. Detecting the immediate boundaries of the seed pixels
3. Classifying them as similar or dissimilar
4. The immediate neighbors are then treated as seeds and the steps are repeated till the entire image is segmented

**Methods of region-based segmentation**

* + Region Growing, Region Merging, Region Splitting, Split and Merge, Watershed …

**Watershed algorithm**: Working by starting from the local maxima of the Euclidean distance map and grows under the constraint that no two seeds can be classified as belonging to the same region or segment map.

1. **Edge Segmentation (edge detection)**

The task of detecting edges in images

**Strength** – Not that complex and the edge feature is important when tasking with isolating image regions.

**Weakness** – Accuracy is reduced when the input is mixed with noise or partially shaded.

From a segmentation-based viewpoint, edge detection corresponds to classifying which pixels in an image are edge pixels and singling out those edge pixels under a separate class correspondingly.

Edge detection is generally performed by using special filters that give us edges of the image upon convolution. These filters are calculated by dedicated algorithms that work on estimating image gradients in the x and y coordinates of the spatial plane.

**How can I separate regions?**

* + Edge detection – Gradient, Laplacian, LoG, Canny filter etc…
  + Edge connection – Local processing, Global processing(Hough transform)

**Canny edge detection algorithm**

To detect “Optimal Edge”, there three features have to be better.

* + Noise smoothing
  + Edge enhancement
  + Edge localization

Canny edge detection was developed to maximize these 3 features, as such this algorithm has 3 steps

* + After “Gaussian smoothing”, Get the size and direction of Gradient
  + The edge detected through “Non-maximum suppression” is thinned to a size of 1 pixel
  + Remain just meaningful edge by using “Edge Linking” and “Hysteresis thresholding”

A person holding a flower

Description automatically generated with medium confidence

1. **Clustering-based Segmentation**

Modern segmentation procedures that depend on image processing techniques generally make use of clustering algorithms for segmentation.

Clustering algorithms perform better than their counterparts and can provide reasonably good segments in a small amount of time.

**K-means clustering** : Taking all the pixels into consideration and clustering them into “k” classes. Differing from region-growing methods, clustering-based methods do not need a seed point to start segmenting from.

**Deep Learning-based methods**

* Semantic segmentation models provide segment maps as outputs corresponding to the inputs they are fed.
* These segments maps are often n-channeled with n being the number of classes the model is supposed to segment. Each of these n-channels is binary in nature with object locations being “filled” with ones and empty regions consisting of zeros.
* The ground truth map is a single channel integer array the same size as the input and has a range of “n”, with each segment “filled” with the index value of the corresponding classes. (classes are indexed from 0 to n-1)

A picture containing diagram

Description automatically generatedThe model output in an “n-channel” binary format is also known as a two-dimensional one-hot encoded representation of the predictions.

Neural networks that perform segmentation typically use an encoder-decoder structure where the encoder is followed by a bottleneck and a decoder or upsampling layers directly from the bottleneck.

**Convolutional Encoder-Decoder Architecture**

Encoder decoder architectures for semantic segmentation became popular with the onset of works like SegNet in 2015.

**SegNet**

This model proposes the use of a combination of convolutional and downsampling blocks to squeeze information into a bottleneck and form a representation of the input.

The decoder then reconstructs input information(bottleneck) to form a segment map highlighting regions on the input and grouping them under their classes.

Finally, the decoder has a sigmoid activation at the end that squeezes the output in the range(0,1).

Diagram

Description automatically generated

* SegNet was accompanied by the release of another independent segmentation work at the same time,

U-Net, which first introduced skip connections in Deep Learning as a solution for the loss of information observed in downsampling layers of typical encoder-decoder networks.

* Skip connections are connections that go from the encoder directly to the decoder without passing through the bottleneck.
* In other words, feature maps at various levels of encoded representations are captured and concatenated to feature maps in the decoder. This helps to reduce data loss by aggressive pooling and downsampling as done in the encoder blocks of an encoder-decoder architecture.
* Skip Connections were a big hit, specifically in the domain of medical imaging, with U-Net providing state-of-the-art results in cell segmentation for the diagnosis of disease.

Chart

Description automatically generated

* **Atrous Convolutions** (made by DeepLab, Facebook) :
  + Replacing simple pooling operations
  + Preventing significant information loss while downsampling
* **Multi-scale feature extraction** with the help of Atrous Spatial Pyramid Pooling (introduced by DeepLab)
  + To help the network segment objects regardless of their sizes
* **CRFs** (Fully Connected Conditional Random Filelds)
  + To recover boundary information which is one of the important parts of Semantic and Instance segmentation

Table

Description automatically generatedCoupling the fine-grained localization accuracy of CRFs, the recognition capacity of CNNs led DeepLab to provide highly accurate segment maps, beating methods like FCNs and SegNet by a wide margin.

**References**

[1] <https://www.v7labs.com/blog/image-segmentation-guide>

[2] <https://m.blog.naver.com/PostView.naver?blogId=laonple&logNo=220873446440&navType=by>