# Image Segmentation

ICT4201:DIP

# What is Segmentation?

- Image segmentation is a fundamental technique in digital image processing and computer vision.
- It involves partitioning a digital image into multiple segments (regions or objects) to simplify and analyze an image by separating it into meaningful components, Which makes the image processing more efficient by focusing on specific regions of interest.
- Steps of segmentation
  1. Groups pixels in an image based on shared characteristics like colour, intensity, or texture.
  2. Assigns a label to each pixel, indicating its belonging to a specific segment or object.
  3. The resulting output is a segmented image, often visualized as a mask or overlay highlighting the different segments.

# Why do we need Image Segmentation?

- Image segmentation is crucial in computer vision tasks because it breaks down complex images into manageable pieces.
  - It's like separating ingredients in a dish.
- By isolating objects (things) and backgrounds (stuff), image analysis becomes more efficient and accurate.
  - This is essential for tasks like self-driving cars identifying objects or medical imaging analyzing tumours.
- Understanding the image's content at this granular level unlocks a wider range of applications in computer vision.

# Semantic Classes in Image Segmentation

- In semantic image segmentation, we categorize image pixels based on their semantic meaning, not just their visual properties.
- This classification system often uses two main categories: Things and Stuff.
  - **Things**: Things refer, to countable objects or distinct entities in an image with clear boundaries, like people, flowers, cars, animals etc. So, the segmentation of "Things" aims to label individual pixels in the image to specific classes by delineating the boundaries of individual objects within the image
  - **Stuff:** Stuff refers to specific regions or areas in an image different elements in an image like background or repeating patterns of similar materials which can not be counted like road, sky and grass which may not have clear boundaries but play a crucial role in understanding the overall context in an image. The segmentation of "Stuff" involves grouping of pixels in an image into clearly identifiable regions based on the common properties like colour, texture or context.

# Types of Image Segmentation

- The high level categorization of image segmentation techniques are based on the nature of the segmentation. The main types of Image Segmentation are:
  - Semantic Segmentation
  - Instance Segmentation
  - Panoptic Segmentation

# Semantic segmentation

- Semantic Segmentation is one of the different types of image segmentation where a class label is assigned to image pixels using deep learning (DL) algorithm.
- In Semantic Segmentation, collections of pixels in an image are identified and classified by assigning a class label based on their characteristics such as colour, texture and shape. This provides a pixel-wise map of an image (segmentation map) to enable more detailed and accurate image analysis.
  - For example, all pixels related to a ‘tree’ would be labelled the same object name without distinguishing between individual trees.
  - Another example would be, group of people in an image would be labelled as single object as 'persons', instead of identifying individual people.

# Workflow of Semantic Segmentation

1. **Data Analysis**: Analyze labeled training data to understand object classes and segmentation patterns.

2. **Network Design**: Create a semantic segmentation network with convolutional layers for feature extraction, contextual information integration, and upsampling layers for dense classification.

3. **Training**: Train the network using the annotated dataset to learn pixel-wise classification and optimize segmentation accuracy using loss functions like cross-entropy or Dice loss.

4. **Inference**: Deploy the trained model to process unseen images and generate segmentation masks by classifying each pixel into specific semantic categories.

Some of the Semantic Segmentation techniques are U-Net, FCN (Fully Convolutional Networks), DeepLab, PSPNet (Pyramid Scene Parsing Network) and SegNet.

# Applications of Semantic Segmentation

- **Scene Understanding**: Semantic segmentation aids in understanding the content and context of complex scenes by identifying and categorizing various objects and regions within an image.
- **Autonomous Driving**: In autonomous vehicles, semantic segmentation enables scene perception by detecting and classifying objects like roads, pedestrians, vehicles, and obstacles to navigate safely.
- **Medical Image Analysis**: Semantic segmentation is crucial in medical imaging for identifying and segmenting anatomical structures or abnormalities, assisting in diagnosis and treatment planning.
- **Video Surveillance**: In video analytics systems, semantic segmentation facilitates object detection and tracking by segmenting and analyzing the motion and behavior of objects over time.
- **Image Editing and Augmentation**: Semantic segmentation powers advanced image editing and augmentation techniques by enabling precise selection and manipulation of specific objects or regions in the image.

# Instance segmentation

- In image segmentation of computer vision task is a more sophisticated feature which involves identifying and delineating each individual object within an image. So instance segmentation goes beyond just identifying objects in an image, but also delineate the exact boundaries of each individual instance of that object.
- So, the key focus of instance segmentation is to differentiate between separate objects of the same class. for example, if there are many cats in a image, instance segmentation would identify and outline each specific cat. The segmentation map is created for each individual pixel and separate labels are assigned to specific object instances by creating different coloured labels which will represent different 'cat' in the group of cats in an image.
  - Instance segmentation is useful in autonomous vehicles to identify individual objects like pedestrians, other vehicles and any objects along the navigation route.
  - In medical imaging, analysing scan images for detection of specific abnormalities are useful for early detection of cancer and other organ conditions.

# Workflow of Instance Segmentation

1. **Object Detection**: The algorithm processes the input image and identifies potential objects by predicting bounding boxes and object classifications.

2. **Bounding Box Refinement**: Post-processing techniques may be employed to refine the predicted bounding boxes, ensuring accurate localization of object instances.

3. **Semantic Segmentation**: Within each refined bounding box, a semantic segmentation model segments the pixels to differentiate the object instance from its background, producing a segmentation mask for each object.

4. **Instance Labeling**: Finally, each segmented object instance is assigned a unique label, and the corresponding segmentation masks are combined to generate a comprehensive instance segmentation map for the entire image.

Some of the instance based segmentation techniques are Mask R-CNN, Faster R-CNN with Mask Branch, Cascade Mask R-CNN, SOLO (Segmenting Objects by Locations) and YOLACT (You Only Look At CoefficienTs).

# Applications of instance segmentation

- **Object Detection and Recognition**: Instance segmentation facilitates accurate object detection, recognition, and classification in complex scenes with multiple overlapping objects.
- **Scene Understanding**: By providing detailed object-level segmentation, instance segmentation enhances scene understanding and context-aware image analysis.
- **Medical Imaging**: Instance segmentation aids in identifying and delineating specific anatomical structures or abnormalities in medical images for diagnosis and treatment planning.
- **Robotics and Autonomous Systems**: Instance segmentation is crucial for robotic vision systems and autonomous vehicles to perceive and interact with the surrounding environment effectively.

# Panoptic segmentation

- Panoptic segmentation combines the strengths of instance segmentation and semantic segmentation to provide a holistic view of the visual scene.
- It assigns a unique label to every pixel in the image, where each label encodes both the semantic category and the instance identity. This means that it not only identifies what each pixel represents (semantic information) but also which specific object (instance) it belongs to. As a result, panoptic segmentation provides a complete and detailed understanding of the visual scene.

# Importance of Panoptic Segmentation

- **Rich scene understanding:** It goes beyond just identifying objects (like semantic segmentation) or just giving bounding boxes (like object detection). Panoptic segmentation provides a complete picture at the pixel level, understanding both what something is (a car, a person) and how many instances there are (that specific car, that particular person).
- **Real-world applications:** This detailed understanding is crucial for tasks like self-driving cars. The car needs to know not only that there's a person there, but how many people and exactly where they are. Panoptic segmentation helps with this by providing both class labels (pedestrian) and instance IDs (individual person).
- **Beyond self-driving cars:** Panoptic segmentation has applications in medical imaging (analyzing cell structures), AR/VR (creating more realistic simulations), and even smart cities (tracking objects and events for better management).

# How Panoptic Segmentation Works

- Panoptic segmentation typically involves a combination of two neural networks: one for semantic segmentation and one for instance segmentation. These networks work together to produce a single, coherent output.

- Network Architecture
  1. Backbone Network: A backbone network, often a convolutional neural network (CNN), extracts features from the input image.
  2. Semantic Segmentation Branch: This branch processes the features to generate a dense, pixel-wise classification map, labeling each pixel with a semantic category.
  3. Instance Segmentation Branch: This branch generates bounding boxes and masks for each instance, distinguishing between different objects of the same category.
  4. Fusion Module: The outputs from the semantic and instance segmentation branches are combined to produce the final panoptic segmentation map.

# How Panoptic Segmentation Works

- Loss Functions
  - To train a panoptic segmentation model, a combination of loss functions is used:
    - **Semantic Loss:** Measures the accuracy of pixel-wise classification.
    - **Instance Loss:** Measures the accuracy of instance identification, including bounding box regression and mask prediction.
    - **Panoptic Loss:** Ensures the final output is a coherent combination of both semantic and instance segmentation results.

# How Panoptic Segmentation Works

- EfficientPS Architecture
  - EfficientPS overcomes the limitations of earlier panoptic segmentation by adding innovation that integrates instances and semantic segmentation more effectively.
- Step-by-step working of EfficientPS is provided below:
  - Step 1: Shared Backbone
    - EfficientPS starts with a shared backbone, which serves as the foundation for both instance and semantic segmentation tasks. This shared backbone extracts essential features from the input images, providing a common basis for subsequent processing.
  - Step 2: Two-Way Feature Pyramid Network (FPN)
    - EfficientPS incorporates a two-way FPN that facilitates communication between the shared backbone and the instance and semantic heads. This bidirectional FPN ensures that relevant features are propagated efficiently across different network layers, enhancing the model's ability to capture fine details and spatial information.
  - Step 3: Instance and Semantic Heads
    - EfficientPS utilizes separate instance and semantic heads, each comprising three modules designed to capture fine features and improve segmentation accuracy. These specialized heads focus on refining the extracted features and generating precise masks for individual object instances and semantic categories.
  - Step 4: Panoptic Fusion Module
    - The final step in the EfficientPS architecture is the panoptic fusion module, which combines the outputs from the instance and semantic heads to produce the panoptic segmentation result. This fusion process ensures a seamless integration of instance and semantic information, resulting in a more coherent and accurate scene understanding.

# Applications of Panoptic Segmentation

- Autonomous Driving
  - Enhanced Perception in Driverless Cars

- Robotics
  - Enhanced Object Manipulation and Scene Understanding

- Surveillance and Security
  - Enhanced Surveillance in Public Spaces

- Augmented Reality (AR) and Virtual Reality (VR)
  - Immersive Experiences in Gaming and Training Simulations

- Medical Imaging
  - Enhanced Diagnosis and Treatment Planning

# Challenges in Panoptic Segmentation

- **Class Imbalance**
  - **Issue**: Sideline parity in the numbers of occurrences across various category of objects can lead to biased training or incorrect segmentation.
  - **Solution**: Methods including class re-balancing during training or the use of weighted loss functions are some of the considerations for this obstacle.
- **Instance Confusion**
  - **Issue:** An example of the version of this instance class which are in close proximity or overlap cannot be properly differentiated, causing confusion in instance segmentation.
  - **Solution:** Instance segmentation algorithms with better boundary lines and overall delineation methods by clustering might be helpful in resolving such problems.
- **Semantic Context Understanding**
  - **Issue**: Underlying the contextual meaning of those objects within a scene is as important as accurate segmentation, which however, can be quite challenging, especially in densely packed or perceptually ambiguous scenes.
  - **Solution**: The figure of context information, for instance, scene parsing or global context modeling, will broaden the perspective of the model and effectively interpret semantic relations.

# Challenges in Panoptic Segmentation

- **Computational Complexity**
  - Issue: Instance-level semantic segmentation poses heavy demand for processing of large amounts of data at both levels of semantics and object instances, hence requiring excessive number of computational resources.
  - **Solution**: By optimizing algorithms, exploiting parallel processing, and making use of accelerated hardware (e.g. GPUs), means of computing complexity can be handled.
- **Data Annotation**
  - **Issue**: Annotating panoptic datasets demands the definition of both semantic classes and specific instances and, thereby, it is a laborious and time-consuming task.
  - **Solution**: Automated or semiautomated annotation tools, crowdsourcing procedures, and data augmentation schemes can definitely simplify the generation of annotated datasets.

# Techniques of Image Segmentation

- The traditional image segmentation uses thresholding, edge detection, Region-Based Segmentation, clustering algorithms and Watershed Segmentation.
- These techniques are more reliant on principle of image processing, mathematical operation and heuristics to separate an image into meaningful regions. They are:
  - **Thresholding:** This method involves selecting a threshold value and classifying image pixels between foreground and background based on intensity values
  - **Edge Detection:** Edge detection method identify abrupt change in intensity or discontinuation in the image. It uses algorithms like Sobel, Canny or Laplacian edge detectors.
  - **Region-based segmentation:** This method segments the image into smaller regions and iteratively merges them based on predefined attributes in colour, intensity and texture to handle noise and irregularities in the image.
  - **Clustering Algorithm:** This method uses algorithms like K-means or Gaussian models to group object pixels in an image into clusters based on similar features like colour or texture.
  - **Watershed Segmentation:** The watershed segmentation treats the image like a topographical map where the watershed lines are identifies based on pixel intensity and connectivity like water flowing down different valleys.

# Deep learning image segmentation models

- **U-Net:** This model uses U-Shaped network to efficiently segment medical images. This model is very efficient in working with small amount of data and provide precise segmentation.
- **Fully Convolutional Network (FCN):**This model has the ability to process image of any size and output spatial maps. This is achieved by replacing fully connected layers in a conventional CNN with convolutional layers. This helps in segmenting an entire image pixel by pixel.
- **SegNet:** This model includes an encoder-decoder network, used for tasks like scene understanding and object recognition. The encoder here captures the context in the image and the decoder performs the precise localization and segmentation objects by using the context.

# Deep learning image segmentation models

- **DeepLab:** The key feature of DeepLab is the use of atrous convolutions used to capture multi-scale context with multiple parallel filters.
- **Mask R-CNN:** This model extents the Faster R-CNN object detection framework, by adding a branch for predicting segmentation masks alongside bounding box regression.
- **Vision Transformer (ViT):** A new model that applies transformers to image segmentation. The image is divided into patches and processes them sequentially to understand the global context of the image.

# Image Segmentation Approaches

- Image segmentation involves partitioning an image into multiple segments to simplify its representation and make it more meaningful and easier to analyze.
- Two primary approaches dominate the field of image segmentation:
  1. Similarity Approach
  2. Discontinuity Approach
- Each approach has its methods and applications, tailored to different types of images and objectives.

# Similarity Approach

- The similarity approach in image segmentation groups pixels or regions based on their similar properties. This method assumes that regions with similar characteristics should be grouped together.
- Common techniques in the similarity approach include:
  - Thresholding
  - Region Growing
  - Clustering (e.g., K-means Clustering, Mean Shift Clustering)
  - Graph-Based Segmentation (e.g., Normalized Cuts, Min-Cut/Max-Flow)

# Discontinuity Approach

- The discontinuity approach focuses on detecting and exploiting abrupt changes in intensity or color to identify boundaries between different regions. This approach is useful for images where regions are defined by clear edges.
- Common techniques in the discontinuity approach include:
  - Edge Detection (e.g., Sobel Operator, Canny Edge Detector)
  - Line Detection (e.g., Hough Transform)
  - Corner Detection (e.g., Harris Corner Detector)
- These approaches and techniques provide the foundation for effectively segmenting images, making them crucial for various applications in computer vision and image processing.

# Segmentation Algorithm

Mathematically, we can define the problem of segmentation as follows. Let R represent the entire spatial region occupied by an image. Image segmentation tries to divide the region R into sub-regions $R_1$ ,$R_2$ , .... $R_n$ , such that:$\bigcup_{i=1}^{n} R_i = R$

Ri is a connected set for i =1,2,....,n.

$R_i \bigcap R_j = \phi$for all i and j.

Q(Ri) = TRUE for i = 1,2,...,n.

Q(Ri U Rj) = FALSE for any adjacent regions Ri and Rj.

- $\bigcup_{i=1}^{n} R_i = R$
- Ri is a connected set for i =1,2,....,n.
  $R_i \bigcap R_j = \phi$
- for all i and j.
- Q(Ri) = TRUE for i = 1,2,...,n.
- Q(Ri U Rj) = FALSE for any adjacent regions Ri and Rj.

Here, Q($R_i$) is a logical predicate defined over the regions in the set $R_i$, and \phi represents the null set.

# Threshold-Based Segmentation

- Thresholding is one of the segmentation techniques that generates a binary image from a given grayscale image by separating it into two regions based on a threshold value.

- Hence pixels having intensity values greater than the said threshold will be treated as white or 1 in the output image and the others will be black or 0.
    - $g(x,y) = \begin{cases} 1 & if\ |Z(x,y| > T \\ 0 & \text{otherwise} \end{cases}$; T is the threshold value

- If the threshold T is constant in processing over the entire image region, it is said to be **global thresholding**.

- If T varies over the image region, we say it is **variable thresholding**.

# Global Thresholding

Original Image

Thresholded and segmented Image

The algorithm is explained below.

1. Select an initial estimate of the threshold T.
2. Segment the image using T to form two groups $G_1$ and $G_2$: $G_1$ consists of all pixels with intensity values > T, and $G_2$ consists of all pixels with intensity values ≤ T.
3. Compute the average intensity values $m_1$ and $m_2$ for groups $G_1$ and $G_2$.σ
4. Compute the new value of the threshold T as $T = (m_1 + m_2)/2$
5. Repeat steps 2 through 4 until the difference in the subsequent value of T is smaller than a pre-defined value δ.
6. Segment the image as $g(x,y) = 1$ if $f(x,y) > T$ and $g(x,y) = 0$ if $f(x,y) \leq T$.

# Variable Thresholding

- In this approach, compute a variable threshold at each point from the neighborhood pixel properties.

- Let us say that we have a neighborhood $S_{xy}$ of a pixel having coordinates $(x,y)$. If the mean and standard deviation of pixel intensities in this neighborhood be $m_{xy}$ and $\sigma_{xy}$, then the threshold at each point can be computed as:
  - $T_{xy} = a_{xy} + bm_{xy}$
  - where $a$ and $b$ are arbitrary constants. The above definition of the variable threshold is just an example. Other definitions can also be used according to the need.

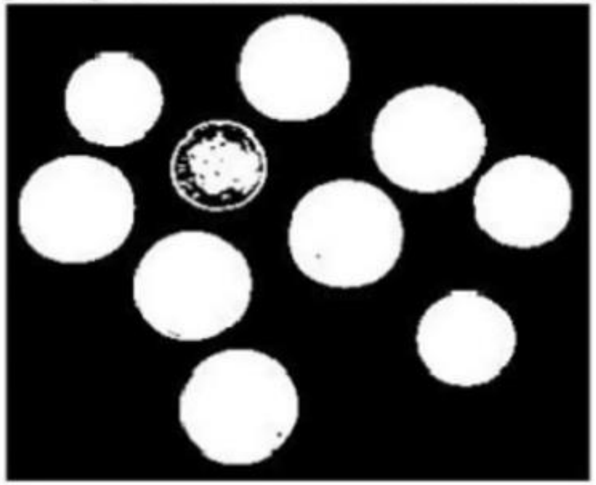
- The segmented image is computed as:
  - $g(x,y) = \begin{cases} 1 \; if \; |Z(x,y)| > T_{xy} \\ 0 \;\; if |Z(x,y)| \le T_{xy} \end{cases}$
  - Moving averages can also be used as thresholds. This technique of image thresholding is the most general one and can be applied to widely different cases.

# Otsu's Method

- Otsu's method is an automatic thresholding technique that determines the optimal threshold value by minimizing the intra-class variance of the pixel intensity distribution. It is widely used in scenarios where the histogram of the image intensity is bimodal, making it a popular choice for medical imaging and document analysis.

Original image.

Image segmentation with Otsu thresholding.



# Edge-Based Image Segmentation

- Edge-based segmentation focuses on identifying the boundaries between different regions in an image. This technique detects significant changes in intensity or color, which typically indicate the presence of edges.
  - Sobel Operator: The Sobel operator is a gradient-based edge detection method that uses convolution with Sobel kernels to approximate the gradient of the image intensity. It highlights regions with high spatial frequency, effectively detecting edges. The Sobel operator is particularly useful for detecting horizontal and vertical edges.
  - Canny Edge Detector :The Canny edge detector is a multi-stage algorithm that includes noise reduction, gradient calculation, non-maximum suppression, and edge tracking by hysteresis. It is known for its ability to detect a wide range of edges while minimizing false positives. The Canny edge detector is widely used in applications requiring precise edge detection, such as object recognition and medical imaging.
  - Laplacian of Gaussian (LoG):The Laplacian of Gaussian (LoG) method combines Gaussian smoothing with the Laplacian operator to detect edges. It involves smoothing the image to reduce noise and then applying the Laplacian operator to highlight regions of rapid intensity change. The LoG method is effective in detecting edges with high accuracy and is commonly used in image enhancement and feature extraction.

# Region-Based Image Segmentation

- Region Growing: Region growing starts with a seed point and expands the region by adding neighboring pixels that have similar properties. The process continues until no more pixels can be added. Region growing is simple and intuitive, producing connected regions, but it is sensitive to noise and requires careful selection of seed points.
- Region Splitting and Merging: Region splitting and merging is a hierarchical method that involves dividing the image into smaller regions and then merging adjacent regions with similar properties. Initially, the entire image is considered as a single region. The region is then recursively split until the resulting regions are homogeneous. Adjacent regions with similar properties are then merged. This technique is effective in handling complex images with varying intensity levels.
- Watershed Segmentation: Watershed segmentation treats the image as a topographic surface, where pixel values represent the elevation. It identifies the catchment basins and ridge lines, segmenting the image into distinct regions. The watershed algorithm is particularly useful for separating overlapping objects in an image, making it popular in medical imaging and object detection.

# Clustering-Based Image Segmentation

- K-means Clustering
  - K-means clustering partitions pixels into K clusters based on their features. It iteratively assigns each pixel to the nearest cluster center and updates the cluster centers to minimize the sum of squared distances between pixels and their corresponding centers. K-means clustering is simple and efficient, making it suitable for various applications, including image compression and color quantization.
- Mean Shift Clustering
  - Mean shift clustering identifies clusters by shifting a window towards regions of higher density, effectively finding the modes of the data distribution. Unlike K-means, mean shift does not require the number of clusters to be specified in advance, making it a flexible and adaptive technique. It is particularly effective in segmenting images with complex distributions of pixel values.
- Fuzzy C-means Clustering
  - Fuzzy C-means clustering extends the K-means algorithm by allowing each pixel to belong to multiple clusters with varying degrees of membership. This approach is beneficial in handling images with ambiguous or overlapping regions, providing a more robust segmentation result. Fuzzy C-means clustering is commonly used in medical imaging and remote sensing.

# Artificial Neural Network-Based Segmentation

- Convolutional Neural Networks (CNNs) are widely used for image segmentation tasks due to their ability to learn spatial hierarchies of features. CNN-based models, such as U-Net and SegNet, have shown remarkable performance in medical imaging, autonomous driving, and satellite image analysis.

  - U-Net : U-Net is a fully convolutional network designed for biomedical image segmentation. It consists of a contracting path to capture context and a symmetric expanding path to enable precise localization. U-Net's architecture allows for the efficient segmentation of high-resolution images with limited training data.

  - SegNet : SegNet is another popular CNN-based model for semantic segmentation. It employs an encoder-decoder architecture, where the encoder captures spatial features, and the decoder performs upsampling to generate pixel-wise segmentations. SegNet is effective in applications requiring detailed segmentation, such as urban scene understanding and object detection.

# Souce

- https://www.geeksforgeeks.org/explain-image-segmentation-techniques-and-applications/
- https://www.geeksforgeeks.org/image-segmentation-approaches-and-techniques-in-computer-vision/