Panoptic Segmentation: A comprehensive pathway to A Real-World AI Vision

[1]Dr Nellutla Sasikala, [2]Vemuri Pravalika
[1]Professor, [2]UG Scholar
[1,2] Department of Electronics& Communication Engineering, Karimnagar, INDIA

***Abstract*— For computer vision tasks like object detection, recognition, and classification, relay on feature extraction, labelling and segmentation of captured videos or images. Applications like smart city, health care, geoscience and remote sensing are based on video analysis. Image segmentation for video analysis plays a vital role. One of the novel segmentation strategies which has been recently developed is panoptic segmentation. Panoptic segmentation is a fusion of semantic and instance segmentation. In self autonomous driving, medical image analysis, crowd counting, etc. have complicated background components, the high variability of object appearances, numerous overlapping objects and ambiguous object boundaries makes the task challenging. For such applications panoptic segmentation is used which provides several state-of-art methods and robust learning.**

***Index Terms*— panoptic segmentation, instance segmentation, semantic segmentation, data annotation, computer vision, medical image segmentation, object detection.**

I. INTRODUCTION

Now-a-days sensors are used for capturing data and are deployed in smart cities to enable the data collection from multiple sources in real time. For security purpose, they are installed in public and residential areas. Devices with video capturing capabilities are significantly in usage. This leads to opportunities for analysis and interface through computer vision technology. Smart city applications such as public security using video surveillance, motion tracking, pedestrian behavior analysis, health care services, medical analysis and autonomous driving, etc. can be developed from the videos and images captured which contain useful information. Machine learning and big data analytic tools play an essential role in the field and on the other hand computer vision takes rely on feature extraction, labelling and image segmentation. Therefore, there is a strong need for proper labelling the data in the AI learning process where information can be extracted from images. Automatic labeling[1] of subject of interest or object is achieved from Bounding box labelling and image segmentation.

Countable entities or objects which have similar texture or structure contained in homogenous region are termed as things. The uncountable regions such as water, roads and sky, etc. are termed as stuff. For identifying things and stuff clearly many visual algorithms are used. One of the segmentation techniques which is used for identification of things and stuff is semantic segmentation. In contrary, another technique is developed which is used for processing of only things which is instance segmentation. In instance segmentation things in the image/video in process where an object is deleted and isolated with a bounding box, or a segmentation mask. Therefore, instance segmentation and semantic segmentation are traditional approaches to current trends in segmentation that is used for masking or highlighting each specific contents in an image.

Scientific segmentation[2] to an operation of labelling things and stuff by denoting the things of the same class with same colors. On the other hand, in instance segmentation things are labelled by different colors, and the stuff is ignored i.e.; the background is ignored. In this line, objects can be further classified

into things and stuff [where stuff would be sky, water, etc.; things would be persons, cars, animals, etc.]. derived from the semantic and instance is panoptic segmentation.

Panoptic segmentation is a complex computer vision task that solves both instance segmentation and semantic segmentation problems together, enabling a more detailed understanding of a given scene. Panoptic segmentation provides detailed contents in the image, generates much more information for analysis, enables computationally efficient operation using AI models by separating both things and stuff of the same class or type using different colors. Unified CNN[2] based methods or merging instance and semantic segmentation results can be used for panoptic segmentation. Panoptic segmentation has great perspective because of its efficiency and also used in wide range of applications. Even though the process of this segmentation is greater, many challenges slow down the improvement. These challenges include scale of objects in the science, the cultured scenes, the weather changes, the quality of used datasets, and computational cost while using a large-scale datasets.

II. BACKGROUND:

Image segmentation improves object detecting methods in which these can be classified as semantic segmentation and instance segmentation. Pixel classification is done in semantic segmentation whereas object classification from a same is done in instance segmentation.

SEMANTIC SEGMENTATION:

Dense prediction is carried out by pixel level segmentation of the scene in semantic segmentation. Semantic segmentation is the operation of labelling each pixel in the image with the corresponding class that represents the category of the pixels i.e.; semantic segmentation classifies different regions in the images belonging to the same category of things or stuff. End to end segmentation of natural images was done with CNN's only after 2014, even though semantic segmentation was proposed in 2007.

The labelling of each pixel in image regions is carried out by the primary process which is spatial analysis. Basic architectures are CNN based methods such as U-Net, Seg Net, Fully Connected Networks (FCN) and Decom Net[3] which are used for segmenting region with acceptable accuracy and quality. These applications are based on the scenes of image and panoptic segmentation allows its deep understanding which helps in the analysis of the scene.

Acquiring accurate data sets is very important in order to deploy a successful system that utilizes the vast advantage of panoptic segmentation. There are public data sets for machine learning which can be easily accessible and can be implemented instead of any system. COCO, Common Object in Context, provides image annotations for 1.5 million common object instances. There is no necessity to manually annotate the objects that appear frequently. Another dataset is city scrapes which is used to life scenes. It includes 10 things categories and 20 stuff categories which includes pedestrian to build things. Pastin is an excellent database set for application of AI.

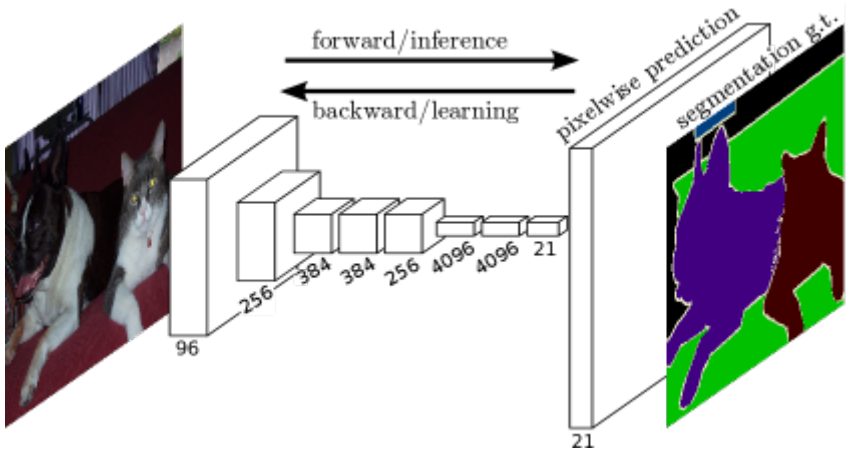

Image Courtesy : From Google

III. TECHNIQUES OF PANOPTIC SEGMENTATION:

A new direction in image segmentation is panoptic segmentation which has been a breakthrough in computer vision and enables a combined view of things and stuff.

Sometimes instance segmentation and panoptic segmentation techniques are used separately before combining the aggregate results of panoptic segmentation.

III.I. RGB IMAGE DATA

The primary data source for panoptic segmentationare RGB images data, because they are widely used in video cameras, image scanners, digital cameras, computer and mobile phone displays. For example, panoptic fusion which is one of the models of panoptic segmentation is used in online volumetric semantic mapping system[5] that combines both stuff and things. The prediction of class labels of background regions(stuff) and foreground objects(things) relies on first predicting pixel wise panoptic labels for entire RGB frames by using semantic and instance segmentation o/ps.

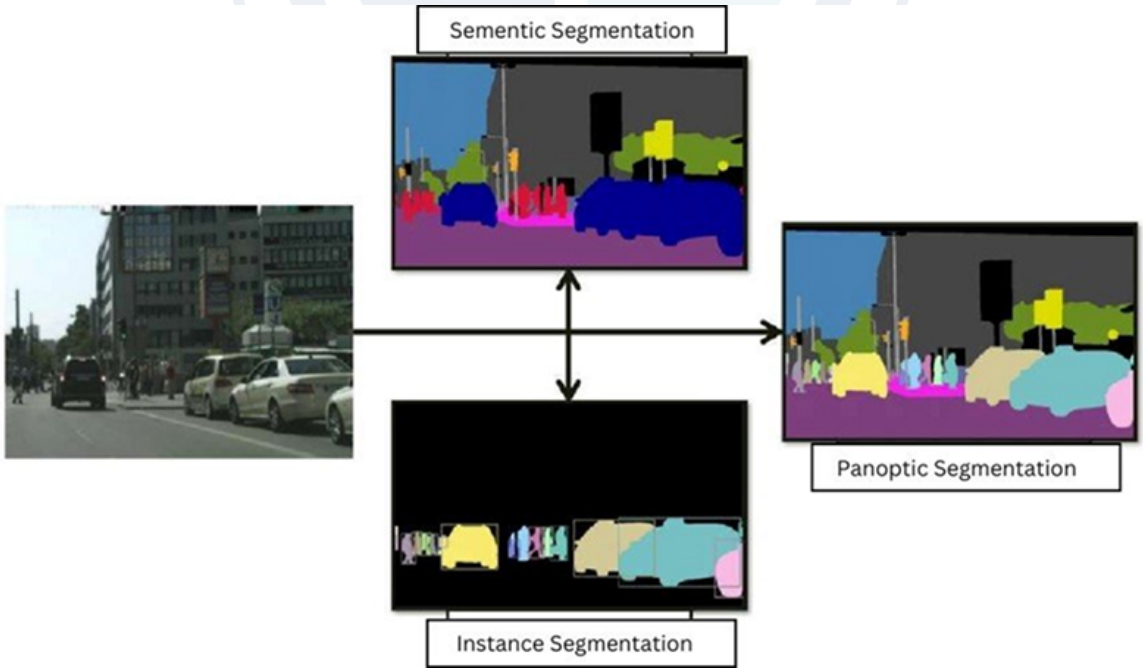

Image Courtesy : From Google

III.II. MEDICAL IMAGES :

Medical image segmentation plays an essential role in computer aided diagnosis systems as medical imaging being one of the most valued applications of computer vision. Different kinds of images are used

for both diagnosis and therapeutic purposes such as X-Rays, Computed Tomography(CT) scan, Magnetic Resonance Imaging(MRI), ultrasound, Nuclear Medicine Imaging and Positron Emission Tomography(PET) scan. In medical images, instance segmentation is used for assigning class values to each pixel and separating objects within the same class. An unique ID is assigned to every single object. Images morphology spatial location and distribute objects helps in analyzing the biological behavior. Instance segmentation has its own limitations as cell-R-CNN which is a panoptic architecture proposed. Typically, the encoder of the instance segmentation model is used to learn the global semantic label features accomplished by jointly training a semantic segmentation model.

III.III. LIDAR DATA :

High resolution digital elevation models with unideal accuracy (10cm can be created) by LIDAR technology which is similar to RADAR. Lidar data are highly accurate and robust so they are preferred for object detectionusing panoptic segmentation of LIDAR space. KITTI dataset contains annotated LIDAR scenes with different environments and scenarios / has extensively used those while an explicit approach has been emphasized. A distinct and contrary approach is adopted to cluster the object segments using CNN architecture. Since the clustering doesn't need computation time and energy as that of CNN's, the model thus adopted can be deployed even with a CPU.

Applications:

Because of the increasing performance of panoptic segmentation systems have developed which are helpful for various tasks and applications.

a) Object detection:

Object detection plays a vital role in computer vision and image processing. Object detection process has become more manageable and accurate due to the panoptic segmentation. It refers to detecting instances of semantic objects of a particular class in digital images and videos. Panoptic segmentation has received significant attention for novel and robust object detection schemes.

b) Medical Image Analysis ;

One of the significant applications of segmentation is medical image analysis. Depending on the segmentation objects, different techniques are used for analyzing and segmenting, with the advent of panoptic segmentation, a wide range of techniques have developed in the medical field. An example like segmentation of overlapped nuclei is taken and a bending loss regularization network was proposed for nuclei segmentation. Bending loss is generated from penalties of large and small curvatures, where high penalties are reversed for contour with large curvatures and small penalties are reserved for contours with small curvatures. This has helped minimize the bending loss and avoided the general contours which were surrounded by multiple nuclei. The MoNuSeg dataset was used to validate this framework using different metrics including the aggregated Jaccard (AJI), Dice, KQ and PQ. With the invert of Panoptic segmentation features fusion network, which is an instance segmentation process to analyze biological and biomedical images, it has become easy for nuclei segmentation. TCGA-tumor data set has been employed to validate the Panoptic segmentation where TCGA-tumor data set contains 30 histopathology images in size 1000x1000 obtained from the Cancer Genome Atlas (TCGA) at 40x magnification. Image comprised of seven organs including breast, bladder, colon, kidney, liver, prostate and stomach.

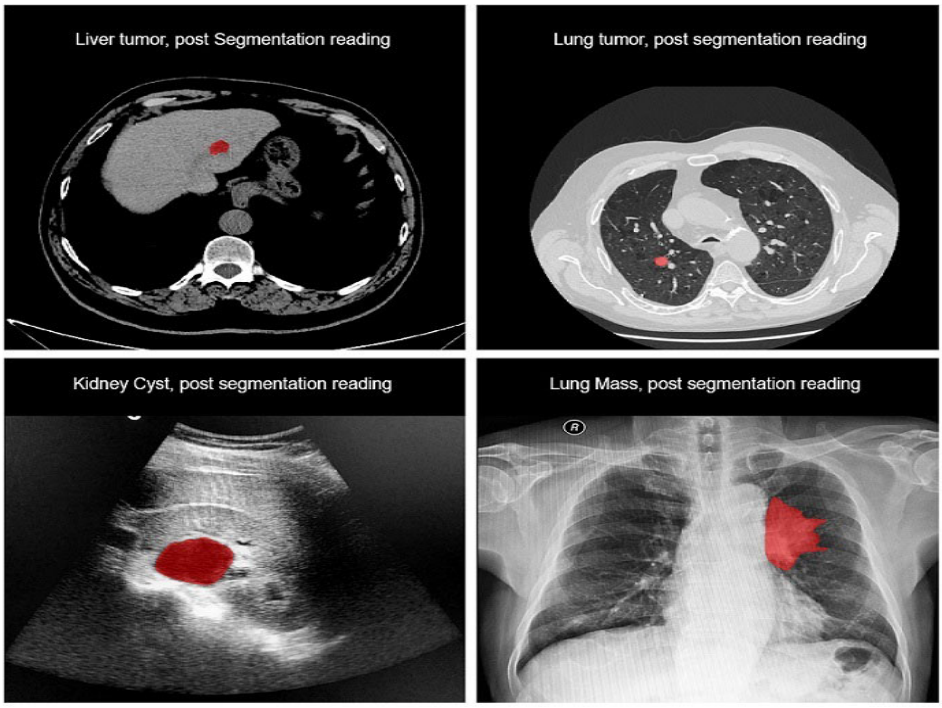

Image Courtesy : From Google

c) Autonomous Self-driving:

One of the main and critical application of Panoptic segmentation is autonomous self-driving. Autonomous self-driving system relies on the deeper level understanding of the scene and better perceptive of the scene. Self-driving Cars system can be builton the data collected from the hardware sensors such as LIDAR, cameras, RADARs and Sonardata (which has been widely used with the latest advances in DL andcomputer vision). Panoptic Segmentation can help in semantic context where pixel represent cars Vs pedestrians vs drivable space and overall architecture context where pixel represent the same car vs other car objects. Pixel level semantic and instance segmentation of camera images based on a single, multitask learning DNN is possible by NVIDIA while is an efficient Scheme. The precise wise and as a whole, understanding of a scene is possible by this method which has enabled the training of a panoptic segmentation base DNN.

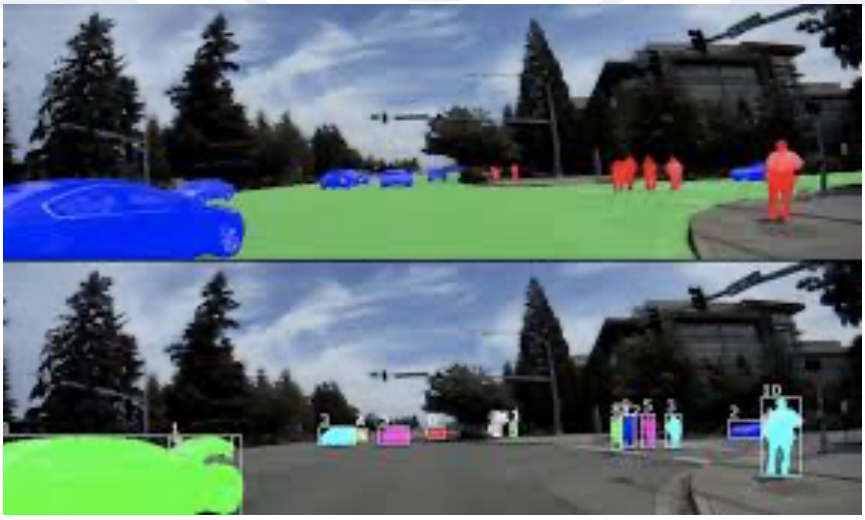

Image Courtesy : From Google

d) UAV remote Sensing:

Road conditions monitoring and urban planning are implemented by UAV remote sensing where panoptic segmentation is the essential method. More comprehensive information to provided by panoptic segmentation than semantic segmentation technology. The panoptic segmentation algorithm identifies UAV helps to solve some problems like the large target scene and small target of UAV, resulting in the back of the foreground targets of the segmentation results and the poor quality of the segmentation mask.

Improvement in the network feature extraction is introduced which has the deformable convolution as one of the network features. To enhance the overall quality of foreground target mask, Mask IoU module is developed and integrated into the instance segmentation branch. UAV-ODC panoptic Segmentation data set is used to test and validate the panoptic segmentation model.

e) Dataset annotations:

Images Categorizing and labelling data for implementing segmentation algorithms or other AI based solution is what can be also employed to perform data set annotations. Panoptic segmentation can be also employed to perform a dataset annotations[7]. In order to annotate a dataset, panoptic segmentation is used which helps to conduct image annotation using a collaborator(human) and automated assistant(based on panoptic segmentation) or both. Instance Segmentation, semantic segmentation and annotating data sets are combinedly obtained from a weaklysupervised panoptic segmentation. An industrial application of panoptic segmentation for annotating datasets are studied where a 3D model is used to generate model of industrial buildings which can improve the inventories performed remotely, where a precise estimation of objects can be performed. For example, in a nuclear power site, equipment positions can be first analyzed using panoptic segmentation of collected panoramic images before going on site, which reduces the cost and time of maintenance.This is a huge break point to advances in automation of large scale industries using panoptic segmentation. A comprehensive virtual aerial image data set, named VALID is proposed, that contains of 6690 high resolution images that are annotated with panoptic segmentation and classified into 30 categories.

f) Data augmentation:

One of the promising application of panoptic segmentation is for data augmentation. Data augmentation schemes that operate exclusively in pixel space and requires no additional data or training and which are computationally inexpensive to implement is possible by using panoptic segmentation. Panoptic data Segmentation method ,PanDA is retraining of existing models of different PanDA augmented data sets, where a high performance gains are achieved in instance segmentation and panoptic segmentation.

Other Applications:

Panoptic segmentations[8][1] finds its application in other research fields namely biology and agriculture for analyzing and segmentation of images. One of the recent research in biology include the research on pigs. Panoptic segmentation has efficiently segmented individual pig through a neural network (for semantic segmentation) using different network heads and post process methods. Even with dirty lens and occlusion, there is a 95% accuracy of segmentation of pigs with panoptic.

IV. PUBLIC DATASETS :

The most important part of the panoptic segmentation is the datasets, which are enabled by the growth of ML and DL algorithms. The datasets usually contain batches thousands or millions of images of ground truth frames, which can be utilized by the models for the implementation of different algorithms. For example, ImageNet helps to evaluate visual recognition algorithms. VGGFace2 helps to validate face recognition methods. The most famous data sets used for image segmentation is Cityscape, Synthia and Mapillary.

Dataset 1. Mapillary Vistas:

It is a traffic related data set with large scale collection of segmented images which is for instance or semantic or panoptic segmentation. This data set is comprised into training, validation and test sets. Size of each set is 18,000, 2000 and 5000 images respectively. Total number of classes is 65, where 28 refers to stuff and 37 for things. It includes different image sizes ranged from 1024x768 to 4000x6000.

Dataset 2 KITTI:

This data set consists of image captures from various places of the metropolis of Karlsruhe, Germany which includes highways and rural region. KITTI is composed of 12K images, 5 different classes, where no. of things and stuff are not specified. Each image of data set contains around 3D pedestrian and 15 vehicles.

Dataset 3 Semantic KITTI:

3D point cloud version of the objects in KITTI data set is the Semantic KITTI. The data is captured with LIDAR from a field of view of 360°. It consists of 43,000 scans of with 28 classes.

Dataset 4 Middlebury Stereo:

It has 7 different object poses which has a depth images and RGB-D image captured from different field-of-views. The images are gleaned with various resolutions including 640x480 and 1280x1024.

Dataset 5 CityScrapes:

This data set is formed from 50 city images, different scenes captured and has large-scale-size. This is the most used data set for instance and Semantic segmentations. 20k labelled images with coarse annotation are present in city scrapes data set where the number of object classes is about 30.

Image Courtesy : From Google

Dataset 6 COCO Panoptic:

Popularly used data set for image segmentation and recognition. Microsoft COCO contains more than 2 million images while 32,800 images are labelled for instance segmentation with labelled images including stuff (various objects like no animals, people, etc.) and things such as reads, sky, etc.. For validating panoptic segmentation, anovel version of COCO data set assigned to instance and semantic labels

of each pixel of any image which different colors are suitable. 123k images are labelled and divided into 172 classes in which 91 are stuff and 80 are things.

CONCLUSION:

Panoptic segmentation is a critical next step in computer vision, allowing us to move beyond perceiving isolated semantic or instance-based information, to a holistic understanding of visual scenes. By clearly separating countable things and uncountable stuff, it allows for a better perception of the scene that is indispensable for further high-level tasks, no matter in what domains. By enabling higher degrees of accuracy, automation, and decision-making, panoptic segmentation can benefit applications ranging from smart city surveillance to autonomous driving, from healthcare diagnostics to industrial inspection. However, with the advancements of the very deep learning models and deep learning datasets, which includes the large-scale annotated datasets, this is no longer the case; despite a series of challenges such as different scales of objects, complex background, and computation consumption.

REFERENCES:

[1] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2961–2969, 2017.

[2] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9404–9413, 2019.

[3] T.-Y. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: Common Objects in Context," *European Conference on Computer Vision (ECCV)*, pp. 740–755, 2014.

[4] M. Cordts, M. Omran, S. Ramos, et al., "The Cityscapes Dataset for Semantic Urban Scene Understanding," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3213–3223, 2016.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[7] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354–3361, 2012.

[8] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, 2017