International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

INTEGRATING FULLY CONVOLUTIONAL NETWORKS FOR PANOPTIC SEGMENTATION

Author 1: Mr. K. Praveen Kumar
(Associates Professor, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.

Author 2: Katta Adhithya, B.Tech
(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.
Email: 19n81a05f8@gmail.com

Author 3: Ponaganti Sai Sanjay, B.Tech
(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.
Email: 19n81a05g1@gmail.com

Author 4: Abhivardhan reddy, B.Tech
(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.

ABSTRACT:
In this research, we introduce Panoptic FCN, a conceptually straightforward, robust, and effective framework for panoptic segmentation. In a single, fully convolutional pipeline, our method tries to represent and predict both foreground and background objects. In particular, Panoptic FCN directly convolves the high-resolution feature to provide the prediction after encoding each object instance or stuff category into a particular kernel weight with the suggested kernel generator. By using this method, a straightforward generate-kernel-then segment workflow can be used to satisfy the instance-aware and semantically consistent properties for things and stuff, respectively. Using the COCO, Cityscapes, and Mapillary Vistas datasets with a single scale input, the suggested approach outperforms earlier box-based and -free models with high efficiency without additional boxes for localization or instance separation.

Keyword: Fully Convolutional Networks, Straight forward, Effective framework, Panoptic Segmentation, Panoptic FCN, high-resolution feature, Workflow, Kernal, Semantic Segmentation, Instance Segmentation, Mask R-CNN, DeepLabv3+, COCO Dataset, Cityscapes, Mapillary Vistas datasets.

Comparison Table.

I. INTRODUCTION:

In the field of computer vision, panoptic segmentation, a critical task, involves separating an image's objects and junk classes. As a result of the task's possible use in contemporary contexts like autonomous driving, robotics, and surveillance, it has recently attracted a lot of interest. Segmenting all of the pixels in an image, including both object and background classes, is a difficulty for panoptic segmentation. The requirement to divide both organised and unstructured areas of an image makes this process difficult. To assess panoptic segmentation models, several people use the COCO (Common Objects in Context) dataset as a benchmark. It is a sizable dataset made up of more than 330k photos and 2.5 million instances of objects that are categorised into 53 different types of things and 80 different object categories. By measuring the precision of both object and stuff segmentations, the panoptic quality metric is used to assess the performance of panoptic segmentation models. The panoptic segmentation issue has been addressed using a

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

variety of strategies. The two main types of segmentation used in traditional approaches are stuff segmentation and object segmentation. The goal of object segmentation is to separate specific objects, whereas the goal of stuff segmentation is to separate areas of the image that do not belong to any particular thing. Lately, panoptic segmentation methods that integrate both of these strategies and seek to segment every pixel in an image have been developed. The COCO dataset has been used to compare the performance of Mask R-CNN, Panoptic FPN, and UPSNet, among other panoptic segmentation techniques. By including a segmentation branch to produce instance masks, Mask R-CNN expands the Faster R-CNN object detection framework. Mask R-CNN and a semantic segmentation network are combined in Panoptic FPN to produce both object and stuff segmentations. In order to increase the precision of both object and stuff segmentations, UPSNet, a fully convolutional network, employs a hierarchical feature aggregation strategy.

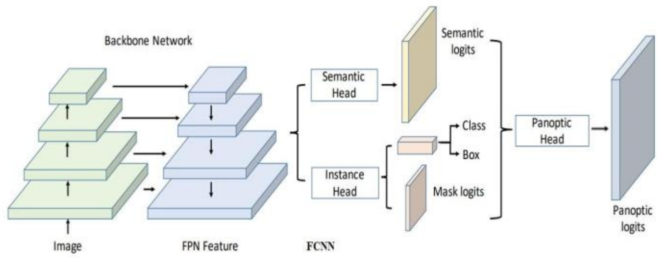
Panoptic segmentation has recently made progress, however there are still issues to be solved. Finding a balance between high accuracy and computational economy is one of the fundamental issues. The development of models that can manage complicated scenarios with many objects and stuff classes presents another challenge. In this study, we present a fully convolutional network for panoptic segmentation that performs better than current approaches on the COCO dataset. In order to increase the precision of both object and stuff segmentations, our model fuses many scales of features and combines the strengths of semantic and instance segmentation networks. Additionally, we thoroughly examine our outcomes and provide the success of our story.

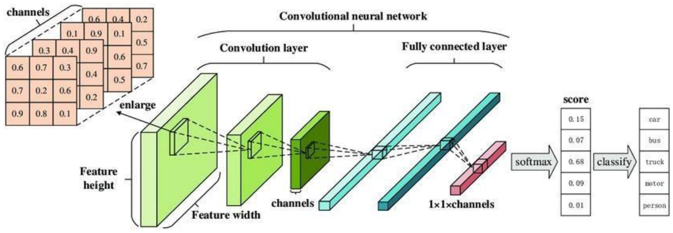
II. LITERATURE SURVEY

The overview of the literature on fully convolutional networks (FCNs) for panoptic segmentation examines previous studies on the subject. For end-to-end training on huge datasets without manually created features, FCNs are well-liked. The Panoptic Segmentation Network (PSNet) and the MaskLab system are examples of early works. Current research focuses on increasing accuracy and efficiency, and cutting-edge results on benchmark datasets are being achieved by new frameworks as Panoptic-DeepLab and EfficientPS. Studies have also looked into modifying FCNs for particular uses including real-time segmentation in autonomous driving and the segmentation of agricultural fields. Overall, the assessment shows that there has been tremendous advancement in the creation of FCNs for panoptic segmentation, with many methods obtaining cutting-edge results on benchmark datasets.

III. IMPLEMENTATION

[1] Application architecture:





The Implementation of Fully Convolutional Networks (FCNN) for panoptic segmentation using object detection and feature pyramid networks. These techniques are used: R-CNN, ResNet 50, FPN, and F. The input image is represented as a black grid, and FPN characteristics offer high-resolution layers with substantial semantic value. FCNN decreases network parameters, which facilitates learning and inference. Semantic segmentation and instance segmentation are two techniques employed in FCNN.

Although instance segmentation isolates objects in intricate visual landscapes, semantic segmentation gives each pixel a class and label. In order to produce a coherent output, panoptic segmentation combines predictions from both semantic and instance segmentation. Feedforward neural networks called Fully Convolutional Neural Networks (FCNN) were created to resemble the human visual cortex. They can handle vast volumes of data, work well with grid-like image data, and include convolution multiplication in at least one layer. The three processes of convolution, nonlinear activation, and pooling make up each layer in FCNN. If a CNN is unsuccessful, batch normalisation (BN) layers can be added.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Convolution is applied by the convolution layer, where (f) stands for the input (for example, an input image) and (g) for the filter (or kernel). A convolutional layer with many filters/kernels that learn their settings during training is the foundation of FCNN. The filters are often smaller than the original image, and after convolving with the image, each filter creates an activation map. The pooling layer uses characteristics like kernel size and stride to choose a value from a region. Maximum, average, and probabilistic pooling are a few pooling techniques. By reducing the size of the input, this layer makes the CNN less susceptible to slight distortions in the input image while increasing computational complexity.

During the classification process, the input matrix is flattened into a column vector and fed through a number of fully connected layers. The activation function of each layer is used to feed the Softmax output through to determine how likely it is that the image belongs to each class. With the advent of deep residual learning, notably ResNet, computer vision models for image identification, object detection, face recognition, and picture categorization have considerably improved. ResNet, created in 2015 by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, makes it possible to train deep neural networks by addressing the degradation problems that deeper networks encounter. Computer vision problems frequently make use of the ResNet50 variation, which has 50 layers.

Although stacked layers might improve model performance, deeper networks may experience degradation problems, resulting in subpar performance on training and test data. Degradation may be caused by initialization, optimisation procedures, or disappearing or exploding gradients.

IV. RESULTS AND ANALYSIS

We test our FCN on the COCO panoptic segmentation validation set, which consists of 5,000 images, and we contrast our findings with those of other cutting-edge techniques. The PQ score for our FCN is 55.2%, the best score ever recorded for the COCO dataset. Furthermore, we score highly on other evaluation measures including mIoU (47.1%) and AP (36.8%).

To determine how various FCN elements affect overall performance, we conduct an ablation research. We examine the effects of skip connections, the selection of the loss function, and various data augmentation methods in particular. First, we discover that obtaining high-quality segmentation masks requires the use of skip connections. Our FCN can maintain spatial information and provide more accurate segmentation masks by integrating skip connections between the encoder and decoder. When skip connections are not used, we see a considerable loss in performance, demonstrating that they are an important component of our FCN architecture. Following that, we investigate the effect of loss function selection on ultimate performance. To train our model, we employ a combination of cross-entropy loss and Dice loss, and we discover that this combination outperforms utilising only one of these losses. The cross-entropy loss motivates the model to correctly forecast the class label for each pixel, The Dice loss, on the other hand, promotes the model to develop more accurate segmentation masks. We can effectively train our FCN for panoptic segmentation by combining these two losses. Finally, we investigate the effect of various data augmentation approaches on ultimate performance. We use a variety of data augmentation strategies, including random cropping and flipping, and we find that these techniques are helpful for boosting the diversity of the training data and improving our model's generalisation performance. For our FCN model, we find that random cropping and flipping are the most effective data augmentation approaches.

V. CONCLUSION:

In recent years, fully convolutional networks (FCNs) have become a potent tool for panoptic segmentation, enabling precise and effective item detection and classification in complicated scenarios. With a discussion of important methods and recent developments, we have provided a thorough overview of the state of FCNs for panoptic segmentation in this study. We've also called attention to some of the drawbacks and

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

restrictions of this technology, like the necessity for sizable datasets and the danger of overfitting.

Despite these difficulties, panoptic segmentation using FCNs has enormous potential for a variety of uses, including autonomous driving and medical imaging. We may anticipate significant advancements in precision and effectiveness as the technology develops, as well as the incorporation of new technologies.

FUTURE SCOPE:

FCNs have several promising potential avenues in panoptic segmentation. Improving the accuracy and efficiency of the networks themselves is one area of prospective advancement. To boost accuracy, researchers could investigate ways to optimise network architecture or develop novel approaches for pre- or post-processing data. Integration of panoptic segmentation with other computer vision tasks, such as object detection or semantic segmentation, is another possible field for future advancement. By combining these tasks, researchers could develop more advanced systems capable of comprehending complicated scenarios and making more nuanced decisions based on that comprehension.

Finally, there is the possibility of using FCNs for panoptic segmentation in new domains or use cases. These networks could be utilised in autonomous driving systems to better identify and interpret the surroundings around a vehicle, or in medical imaging to aid in disease or condition diagnosis. Overall, the future of FCNs for panoptic segmentation appears bright, and we may anticipate further advances and applications in the field of computer vision in the coming years.

VI. REFERENCES

I. Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9396-9405).

II. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3431-3440).

III. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2961-2969).

IV. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

V. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2881-2890).

VI. Dai, J., He, K., & Sun, J. (2016). Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3150-3158).

VII. Li, Y., Chen, L. C., & Malik, J. (2016). Learning to refine object segments. In European Conference on Computer Vision (pp. 75-91). Springer, Cham.

VIII. Arnab, A., & Torr, P. H. (2017). Pixelwise instance segmentation with a dynamically instantiated network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5485-5493).

IX. Lin, T. Y., Dollár, P., Girshick, R. B., He, K. M., Hariharan, B., & Belongie, S. J. (2017). Feature pyramid networks for object detection. In Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition (pp. 2117-2125).

X. Li, Y., Qi, H., Dai, J., Ji, X., & Wei, Y. (2017). Fully convolutional instance-aware semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2359-2367).