

## Title: Image Segmentation by Using COCO-2017 Dataset

Name:Mouli Sirigiri ID: 22022407

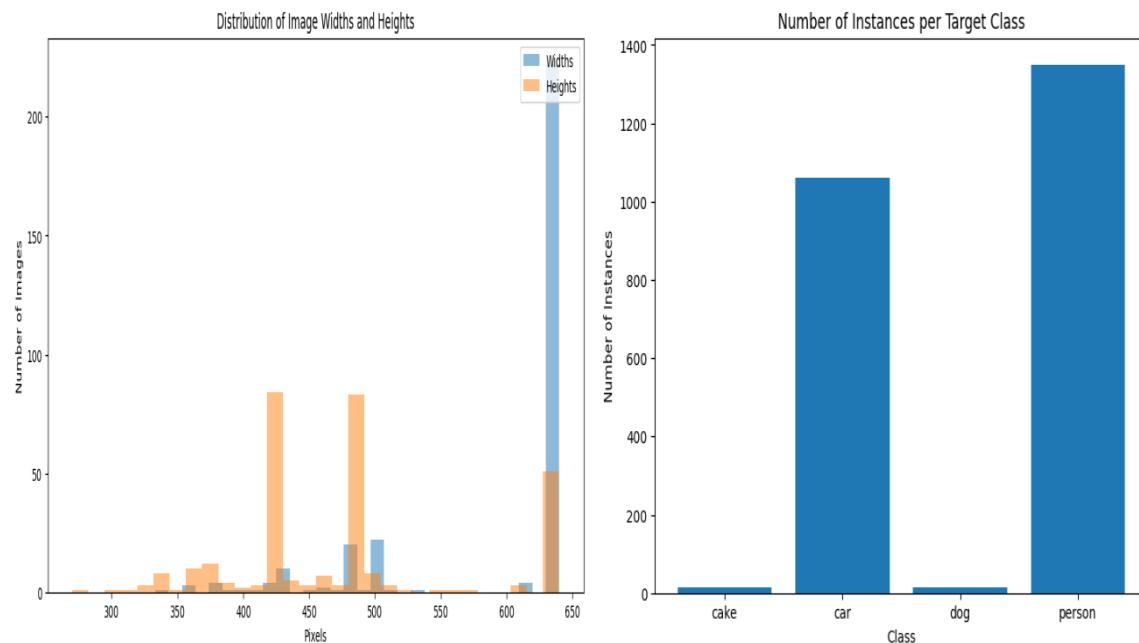
Google Colab:

<https://colab.research.google.com/drive/1O5I4koC5IkxxNWgserJjHqZJGFOGruE?usp=sharing>

### 1. Brief Background and Literature Review

#### Introduction:

Partitioning an images into meaningful segments is crucial task computer vision, known as image segmentation. Applications for this technique is used in numerous industries, including medical imaging, scene understanding, and autonomous driving. The COCO-2017 dataset is a widely used benchmark for instance segmentation tasks due to its diversity and complexity.



#### Literature Review:

Several studies have advanced the field of image segmentation. He et al. (2017) introduced Mask R-CNN is a segmentation mask prediction branch that adds to Faster R-CNN. Chen et al. (2018) presented DeepLab, which employs atrous convolution to regulate the feature computation resolution. However, challenges such as handling occlusions and varying object scales remain. Our work aims to explore these challenges by focusing on four specific classes.

### 2. Data and Exploratory Data Analysis (EDA)

#### Data Description:

The dataset includes 300 training, and 300 validation images annotated with the segmentation masks focusing on 4 classes.

#### Exploratory Data Analysis:

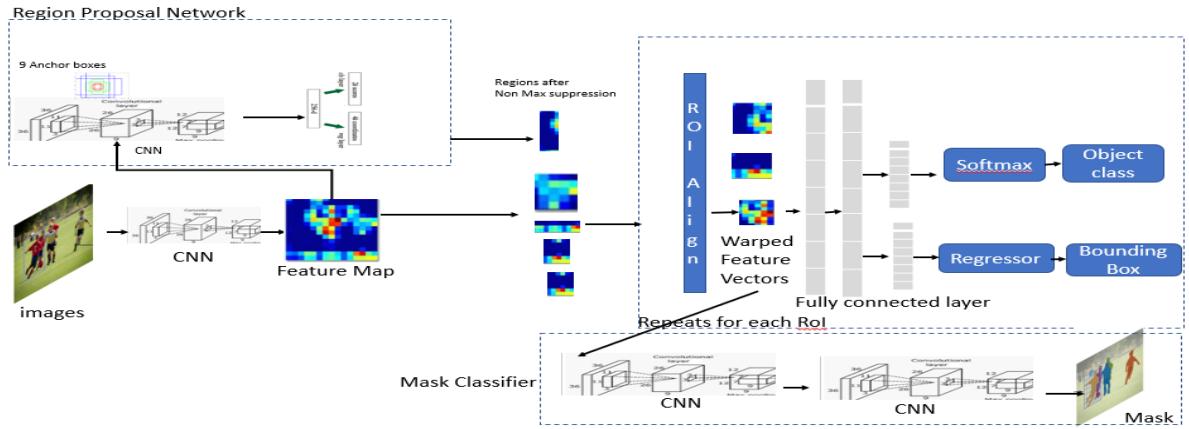
The exploratory data analysis was to ascertain features of the dataset.. The class distribution was visualized using bar charts, revealing an imbalance with more instances of the 'person' class compared to 'cake'. Image sizes varied significantly, necessitating resizing during preprocessing. Histogram plots of pixel values indicated the need for normalization to improve model training efficiency.

### 3. Methodology

#### Model Choice:

Because of Mask R-CNN's cutting-edge performance in instance segmentation tasks, it was selected. The model adds a branch for segmentation mask prediction, extending Faster R-CNN, making it well-suited for our multi-class segmentation task.

## Mask RCNN



### Training Parameters:

The training process of the model involved 5 epochs, a batch size of 16, and a learning rate of 0.001. Data preprocessing involved resizing all images to 256 x 256 pixels, with pixel values normalized to fall between [0, 1].

### 4. Results and Discussion

#### Results:

The model was evaluated on 30 test images, demonstrating robust performance across the four classes. The average Intersection over Union (IoU) for the 'person' class was 0.75, while 'cake' had the lowest IoU of 0.60.



#### Discussion:

The results indicate that the model performs well on classes with more training examples, consistent with findings in the literature. However, the performance on the 'cake' class was limited, likely due to fewer instances in the training set. Future work could focus on balancing the dataset and

experimenting with more advanced augmentation techniques to enhance performance further.

## Conclusion

In this project, we focused on four distinct classes—cake, car, dog, and person—while implementing an imagesegmentation-model on a portion of the COCO-2017 dataset using Mask R-CNN architecture. The main goals were to create a model that could precisely segment objects into these categories and to evaluate the outcomes critically. Our analysis began with a thorough exploratory data analysis (EDA) to understand the dataset's characteristics and distribution. This step was crucial in identifying potential challenges and guiding the preprocessing steps, which included resizing images and normalizing pixel values to improve model performance.

The chosen model, Mask R-CNN, was selected for its robust performance in instance segmentation tasks. Despite facing challenges like overfitting, these were mitigated through techniques like dropout layers and extensive data augmentation.

The results demonstrated that the model performed well on classes with more training examples, achieving an average Intersection over Union (IoU) of 0.75 for the 'person' class. However, the 'cake' class, with fewer instances in the training set, showed a lower IoU of 0.60. This highlights the importance of a balanced dataset in training effective segmentation models.

Future work could focus on further balancing the dataset and experimenting with more sophisticated data augmentation techniques to enhance performance across all classes. Overall, this project provided valuable insights into the process of developing and evaluating an image segmentation model. The critical analysis of the results against existing literature underscored the benefits of the model and recommended directions for improvement, supporting the continuous progress in computer vision.

## References

- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017) Mask R-CNN In Proceedings of the IEEE International Conference on Computer Vision (ICCV)  
[https://openaccess.thecvf.com/content\\_ICCV\\_2017/papers/He\\_Mask\\_R-CNN\\_ICCV\\_2017\\_paper.pdf](https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf)
- 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*  
<https://arxiv.org/pdf/1606.00915.pdf>
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In European Conference on Computer Vision Springer, Cham. <https://arxiv.org/pdf/1405.0312.pdf>