Food Identification with Segmentation

1. Introduction

Food image segmentation involves identifying and classifying ingredients on a plate. This is currently utilized in nutrition tracking, dietary monitoring, and automated calorie estimation. The challenge with Food image segmentation is that food items often have irregular shapes, overlapping boundaries, and visually similar appearances that make accurate segmentation difficult.

To support this growing research area, Wu et al. introduced **FoodSeg103**[1], a large-scale dataset containing over 7,000 food images labelled with 103 food and ingredient categories. Designed specifically for segmentation, this dataset offers high-quality pixel-level annotations that present a challenging but realistic benchmark for segmentation algorithms (see Figure 1: FoodSeg103 Sample Training Images and Masks).

In this project, a semantic segmentation model based on **DeepLabV3** with a **ResNet-101** backbone was used. The model was fine-tuned on the FoodSeg103 dataset to evaluate its ability to distinguish between various food items.

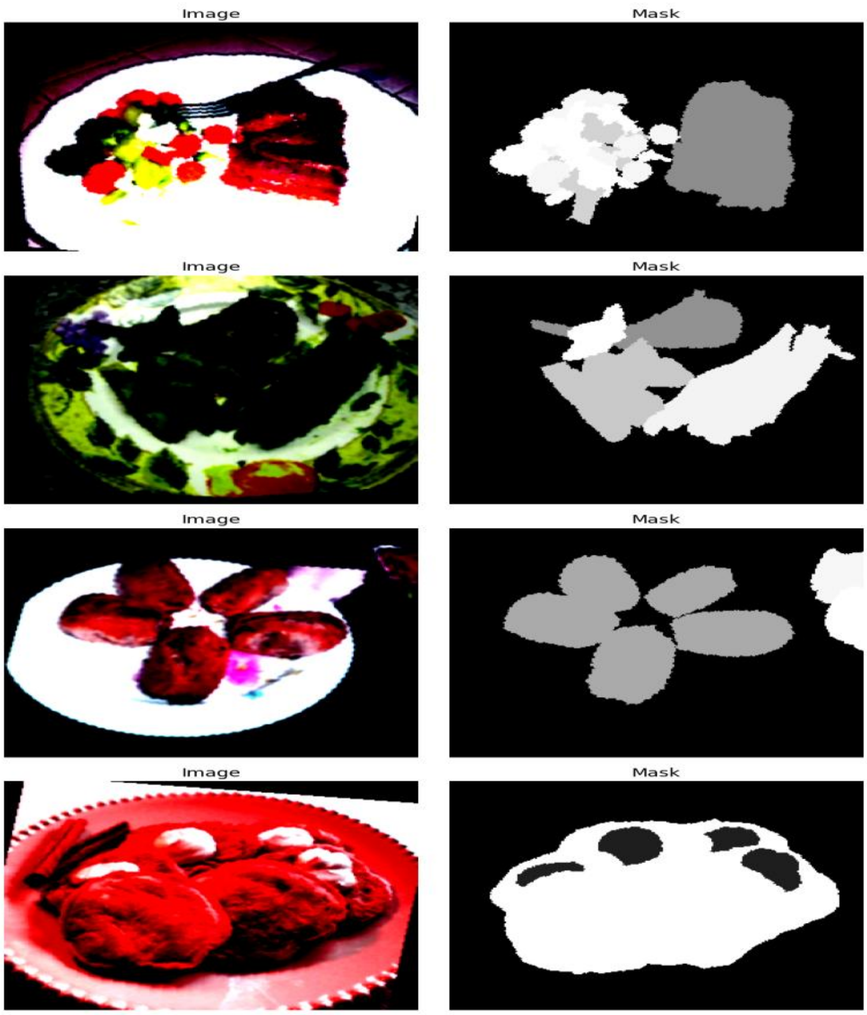

Figure 1: FoodSeg103 Sample Training Images and Masks

2. Algorithm and Implementation

The model used for this particular is DeepLabV3 with a backbone of ResNet101, which is known for its strong performance in semantic segmentation tasks.

For this specific pipeline (see Figure 2: Pipeline Flow)**:**

- A **pre-trained DeepLabV3-ResNet101 model**, originally trained on the COCO dataset, was adapted for food segmentation.
- Only the **last block of the ResNet-101** backbone was fine-tuned, allowing the model to specialize in food-specific features while retaining general visual knowledge from its original training using the augmented **FoodSeg103** images and masks.
- Standard preprocessing steps such as resizing, normalization, and augmentation (e.g., flipping) were applied to the images and their corresponding masks. Other augmentations like adjusting the brightness, hue, saturation etc., were applied to the images alone.

Flow of the Pipeline:

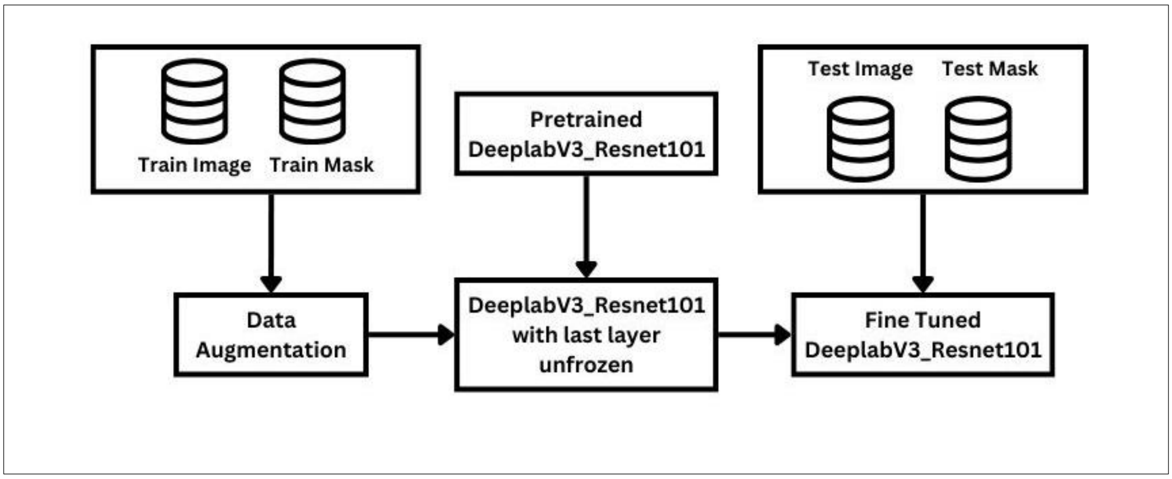

Figure 2: Pipeline Flow

3. Results

The model was evaluated using **mean Intersection over Union (mIoU)**, a standard metric for segmentation tasks. After training *(Epoch 50/50, Loss: 0.1331, Mean IoU: 0.5450),* the model achieved a respectable average mIoU of **Mean IoU: 41.69%** across classes with the test image. This score reflects the difficulty of the dataset and the subtle visual differences between many of the food categories.

Visual results (see Figure 3: Test Result with MioU of 41.69) showed that the model could clearly segment distinct food items in images with little clutter or occlusion. In more complex scenes, particularly where multiple similar-looking ingredients were mixed together, the model could not perform that well.

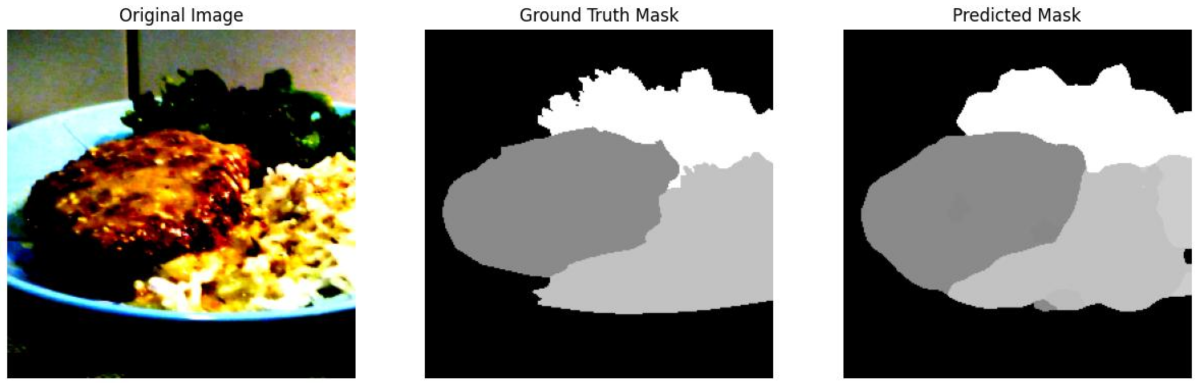

Figure 3 : Test Result with MIoU of 41.69

4. Key Observations and Discussion

- **Transfer learning worked well**: The decision to fine-tune only the final ResNet101 block allowed the model to adapt without losing its foundational visual understanding.
- **Data augmentation helped generalize**: Even simple augmentations improved robustness across different image styles.
- **Class confusion was common**: Visually similar items like different sauces or diced ingredients were often misclassified. This suggests that colour and texture alone may not be sufficient and highlights the potential benefit of incorporating additional context like ingredient lists or recipes.

5. Conclusion

This work demonstrates that DeepLabV3 with a ResNet-101 backbone can be effectively adapted for food segmentation using the FoodSeg103 dataset. While the model shows solid results, especially on clear and simple food scenes, performance still lags for more visually ambiguous items. In future iterations, combining image data with auxiliary information—such as recipe context (using ReLeM), as mentioned in the paper [1] or edge detection (Sobel [2])—could help push the performance further. The FoodSeg103 dataset remains a valuable benchmark for this evolving research area.

Reference

[1] W. Wu, Y. Wang, X. Li, et al., "FoodSeg103: A New Large-scale Benchmark for Food Segmentation," [Online]. Available: https://xiongweiwu.github.io/foodseg103.html

[2] "Edge Detection Using OpenCV," *LearnOpenCV*, Aug. 2, 2021. [Online]. Available: https://learnopencv.com/edge-detection-using-opencv/