Transfer Learning for Fine-Grained Semantic Segmentation of Fast Food Components Group 30

Jiyuan Luo
jxl220083@utdallas.edu

Saketh Madishetty
sxm230092@utdallas.edu

Omar Mukhtar Shaik
oxs230000@utdallas.edu

Daniel Vincent
drv180000@utdallas.edu

University of Texas at Dallas
800 W Campbell Rd, Richardson, TX 75080

1. Problem Statement

This project implements fine-grained semantic segmentation for precise identification of Chick-Fil-A meal components. Utilizing transfer learning techniques with OpenMMLab's MMSegmentation [1] framework, we develop a custom-tailored model capable of accurately delineating individual food items—specifically sandwiches, fries, and drinks—within composite meals. The methodology encompasses data collection, annotation, model selection, and optimization to achieve robust segmentation performance for these specific fast-food classification targets. This segmentation capability enables potential downstream applications including automated calorie estimation, portion control analysis, and nutritional tracking systems that could benefit both consumers and the food service industry in making more informed dietary decisions.

2. Approach

Our research will leverage transfer learning for fine-grained semantic segmentation through OpenMMLab's MMSegmentation [1]. We will first annotate collected data, then employ pre-trained models—specifically Knet [4] for its high accuracy and FastSCNN [2] for its computational efficiency. We will evaluate both models' performance characteristics during training and may adjust our model selection based on computational constraints. Using MMSegmentation's framework, we will conduct comprehensive training, evaluation, and prediction phases. To meet the computational demands of model training, we will utilize cloud-based GPU resources, ensuring sufficient processing power without hardware limitations. If time permitting, we will deploy the optimized model for real-time inference via webcam to perform semantic segmentation on Chick-Fil-A meal combinations, demonstrating practical application of our approach in a consumer food context.

For this application-oriented project, we adapt established semantic segmentation techniques previously developed for urban scene understanding to the novel domain of food item identification. By transferring knowledge from pre-trained models, we recalibrate their capabilities to recognize specific food categories (sandwiches, fries, and drinks) instead of urban elements. We maintain the architectural foundation while modifying classification heads and semantic categories. This approach leverages robust feature extraction capabilities from diverse visual scenes while requiring only targeted fine-tuning on our custom food dataset, significantly reducing computational demands while achieving specialized domain expertise.

3. Data

This project utilizes a hybrid dataset comprising both internet-sourced and personally captured images of Chick-Fil-A meals. The dataset undergoes manual annotation using LabelMe [3] software, where each image is meticulously segmented to identify three primary food categories: sandwiches, fries, and drinks. This annotation process creates pixel-level semantic masks that serve as ground truth data, enabling the model to learn precise boundaries between different meal components during the training phase.

4. Evaluation

We will evaluate semantic segmentation performance using multiple metrics (IoU, Accuracy, Dice, F-score, Precision, and Recall) calculated per-class and globally. Visualization outputs include confusion matrices and segmentation overlays to assess boundary precision. This framework enables comparison between Knet [4] and FastSCNN [2] architectures to determine which model best balances accuracy and computational efficiency for food segmentation.

References

[1] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox

and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. 1

[2] Rudra PK Poudel, Stephan Liwicki, and Roberto Cipolla. Fast-scnn: Fast semantic segmentation network. *arXiv preprint arXiv:1902.04502*, 2019. 1

[3] Kentaro Wada. Labelme: Image Polygonal Annotation with Python. https://github.com/wkentaro/labelme. 1

[4] Wenwei Zhang, Jiangmiao Pang, Kai Chen, and Chen Change Loy. K-Net: Towards unified image segmentation. In *NeurIPS*, 2021. 1