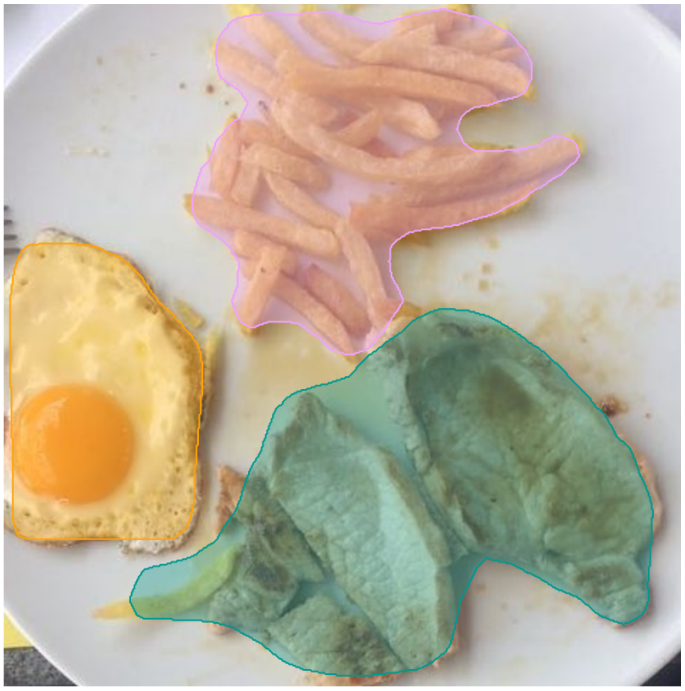
Deep Learning- Food Recognition Challenge

Samral Tahirli (samral.tahirli@studio.unibo.it)

March 2023



Content

1 Introduction …….. 1

2 Problem ……1

2.1 Description ……...1

2.2 Dataset ……1

3 Approaches ……...2

3.1 Detectron2 ……2

3.2 Model Zoo ……..3

3.2.1 Model R50-FPN ……..3

3.2.2 Model R101-FPN ...……3

3.3 Data Augmentation ……..4

4 Evaluation,Discussion and Results ……..4

5 Conclusion……..7

6 References ……7

1 Introduction

The automated visual analysis of food images has become a popular tool in recent years, with many applications, including dietary tracking and medical purposes. Using detectron2 and its pre-trained models, such as R50-FPN and R101-FPN, food recognition has become more accurate and efficient.Additionally, data augmentation techniques have further improved the models' accuracy, making it easier to promote healthy eating habits and improve medical diagnostics.

2 Problem

2.1 Description

The AICrowd Food Recognition Challenge entails the detection of individual food items in images, with two possible tasks: image segmentation and classification. The objective of segmentation is to accurately identify and localize the corresponding areas of objects in the input image at the pixel level, while classification entails assigning the input image to one of several predefined categories. Multiple detection and segmentation algorithms have been trained and assessed for their ability to tackle the challenges of semantic segmentation. The trained neural networks are evaluated by predicting test dataset outputs and comparing against established metrics.

2.1 Dataset

The data set for AICrowd Food Recognition Challenge contains 323 categories with their bounding boxes and segmentation masks.The data set is given in .json format and is divided into 3 parts-training,validation and test.

- This is the Training Set of 54,392 (as RGB images) food images, along with their corresponding 100,256 annotations from 323 food classes in MS-COCO format
- This is the suggested Validation Set of 946 (as RGB images) food images, along with their corresponding 1708 annotations from 323 food classes in MS-COCO format

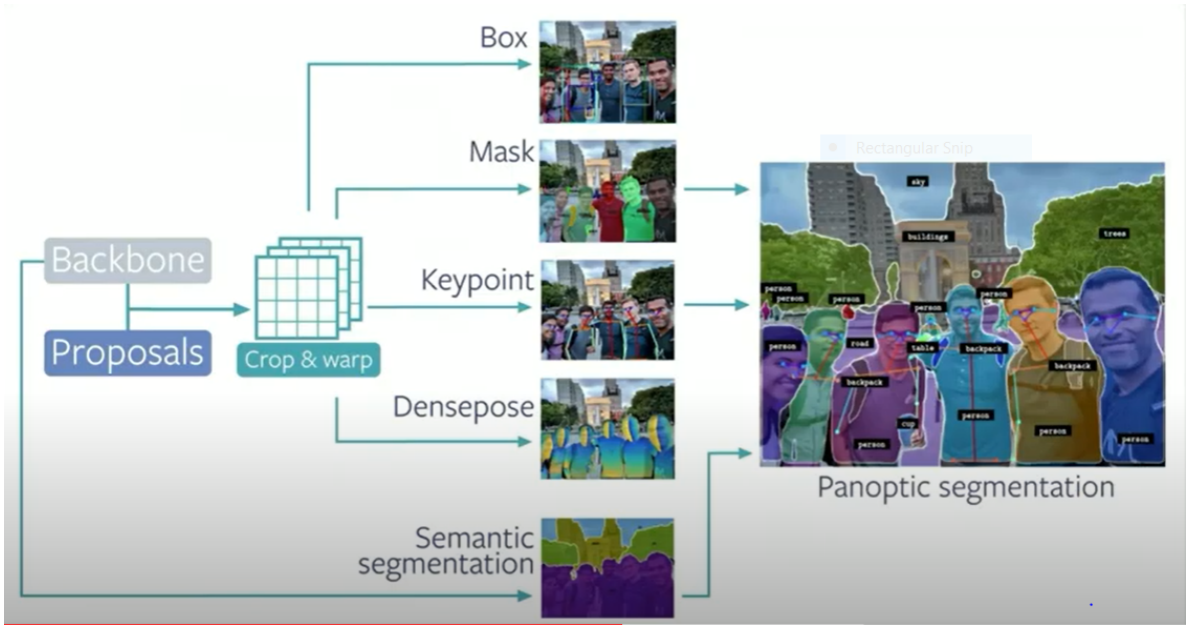
3 Approaches

3.1 Detectron2

Detectron2 is a library developed by Facebook AI Research (FAIR) in 2019, which provides algorithms for object detection and segmentation. It is a combination of Detectron and maskrcnn-benchmark and is written in Python, utilizing the PyTorch deep learning framework. The library simplifies the use of advanced object detection algorithms like Mask R-CNN and Faster R-CNN by offering pre-trained models and an intuitive API.

Detectron2 offers several significant features, including modularity, flexibility, performance, and ease of use. Its modular design allows users to customize object detection pipelines by mixing and matching different components, such as backbone networks, feature extractors, and object detection heads. The flexible architecture supports various object detection tasks, including instance segmentation, keypoint detection, and panoptic segmentation.

Detectron2 is designed to achieve state-of-the-art performance on computer vision benchmarks such as COCO, Cityscapes, and LVIS. Finally, it provides a user-friendly interface, allowing researchers and developers to quickly train and evaluate object detection models with minimal coding effort. Overall, Detectron2 is a powerful and flexible tool for object detection research and development and has gained popularity within the computer vision community.



3.2 Model Zoo

Detectron2's model zoo provides a collection of pre-trained models for various computer vision tasks such as object detection, semantic segmentation, and keypoint detection. The object detection model zoo contains multiple implementations of Faster R-CNN, RetinaNet + Faster R-CNN, and RetinaNet for each task.
Model Zoo is a term used to refer to a repository or collection of pre-trained models for various machine learning tasks, such as image classification, object detection, natural language processing, and more. It's a platform where researchers, developers, and data scientists can share their pre-trained models, making them easily accessible and reusable for others in the community. Model Zoo helps to reduce the computational cost and time required for training machine learning models by providing access to pre-trained models that can be fine-tuned or used for transfer learning on specific tasks.In this report, the pre-trained models are chosen as Model R50-FPN-3x and R101-FPN.

3.2.1 Model R50-FPN

The R50-FPN model is a pre-trained object detection model available in the Detectron2 model zoo. It is based on a ResNet-50 backbone and a Feature Pyramid Network (FPN) for multi-scale feature extraction. The FPN architecture consists of a bottom-up pathway, which is a standard ResNet architecture, and a top-down pathway, which combines high-resolution features from the bottom-up pathway with low-resolution features from higher levels of the FPN. The R50-FPN model is trained on the COCO dataset, which consists of images with object instances labeled with bounding boxes and segmentation masks.

3.2.2 Model R101-FPN

R101-FPN is a variant of the Faster R-CNN object detection model that uses a ResNet-101 backbone and Feature Pyramid Network (FPN) architecture. The ResNet-101 backbone is a deep residual neural network with 101 layers, which is pre-trained on the large-scale ImageNet dataset for image classification. The FPN architecture is designed to address the problem of detecting objects at different scales, by combining features at multiple scales using a top-down pathway and lateral connections.

3.3 Data Augmentation

Data augmentation is a crucial component in achieving optimal performance in a model. For this particular project, the augmentation techniques employed include ResizeShortestEdge, RandomFlip,RandomRotation and RandomCrop . With a 50% probability, RandomFlip horizontally flips the input image. On the other hand, RandomCrop crops a square patch with a fixed size of 640x640 from the input image. It's important to note that the crop size is not relative to the input image size, and the "absolute" parameter denotes this fixed size. Additionally, the project employs random contrast and brightness, which is defined within the range of 0.5 to 2.

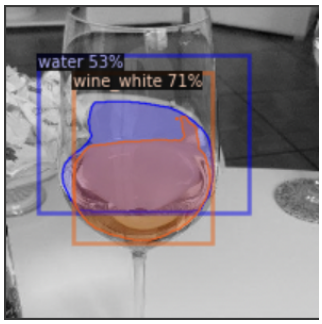
4 Evaluation and Discussion and Results

For this project, I employed two distinct pre-trained models, namely R50-FPN and R101-FPN, along with their augmented versions. R101-FPN is larger than R50-FPN, and therefore capable of producing superior performance results. However, I faced a significant challenge during the training phase, namely, the issue of out-of-memory. Initially, I set the number of epochs at 15,000, but due to the memory problem, I had to reduce it to 2,000, which resulted in subpar detection outcomes for R101-FPN with augmented versions. Additionally, I encountered another problem related to the insufficient number of samples in some categories. Consequently, some images did not have any detected objects in either of the models. Upon further investigation, I discovered that this was due to the lack of sufficient samples for those categories, which negatively impacted the model's overall performance. Finally, I compared the results of each model on the same images.

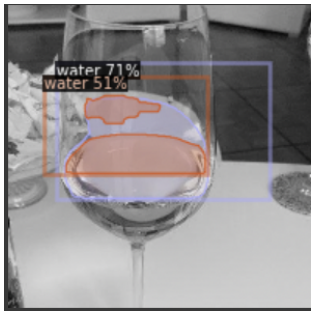
The table presented below displays all pertinent information for each model.

	Model-R50-FPN	Model-R50-FPN (with augmentation)	Model-R101-FPN	Model-R101-FPN (with augmentation)
Train Loss	1.852	1.531	1.537	1.885
AP (bbox)	3.548	2.995	7.143	1.431
AR (bbox)	0.085	0.077	0.139	0.039
AP (segm)	5.711	5.202	10.095	2.421
AR (segm)	0.126	0.122	0.187	0.063

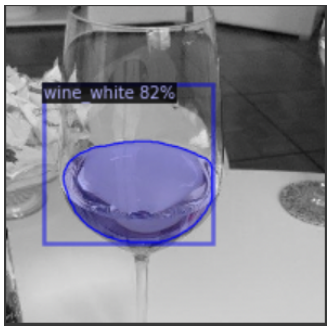
Model-R50-FPN



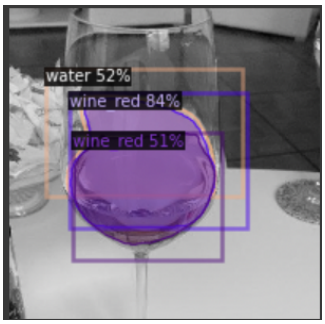
Model-R50-FPN(with augmentation)



Model-R101-FPN

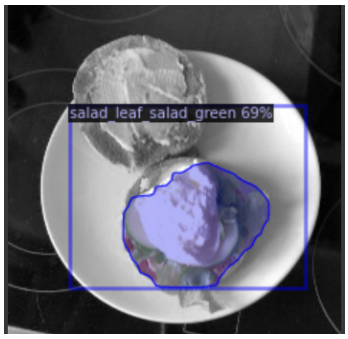


Model-R101-FPN(with augmentation)

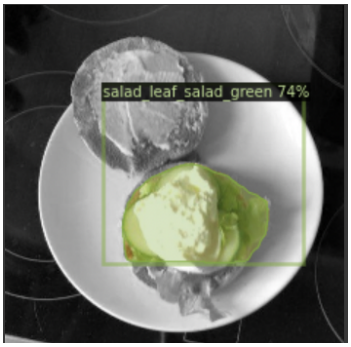


Model-R50-FPN

Model-R50-FPN(with augmentation)



Model-R101-FPN



Model-R101-FPN(with augmentation)

Model-R50-FPN

Model-R50-FPN(with augmentation)

Model-R101-FPN

Model-R101-FPN(with augmentation)

5 Conclusion

In conclusion, advancements in technology have enabled the automatic visual analysis of food images, providing a valuable tool for promoting healthy eating habits and improving medical diagnostics. The use of detectron2 and its pre-trained models, such as R50-FPN and R101-FPN, has played a critical role in improving the accuracy and efficiency of food recognition. Furthermore, data augmentation techniques have further enhanced the models' accuracy, providing a valuable tool for various applications, including dietary tracking and medical purposes. With the continued development of technology, automated food understanding will undoubtedly become more advanced, providing more comprehensive information about food and its impact on health.

6 References

AICrowd Food Recognition Challenge
https://www.aicrowd.com/challenges/food-recognition-benchmark-2022

Detectron2's documentation
https://detectron2.readthedocs.io/en/latest/