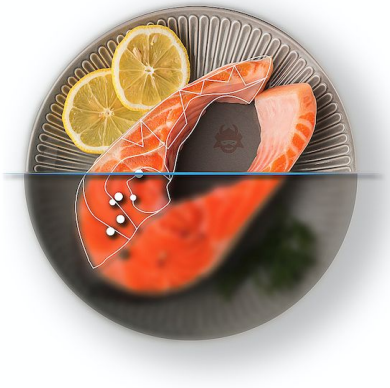
The Food Recognition Benchmark: Using Deep Learning to Recognize Food on Images



M.Sc. Gaurav Singhal
Oviva AG

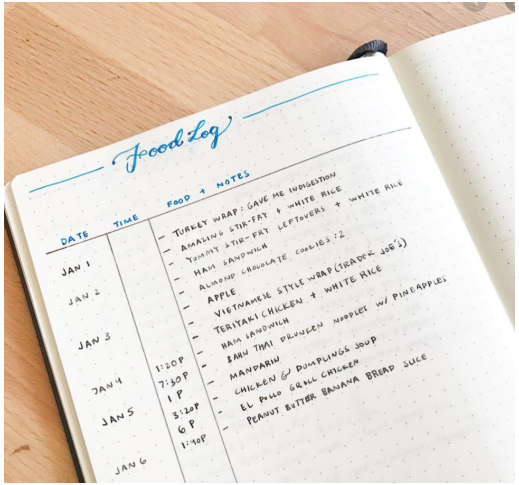
Motivation

- Unhealthy diets can lead to excess weight and obesity
- Unhealthy diets pose a greater risk to morbidity and mortality than unsafe sex, and alcohol, drug, and tobacco use combined
- There is substantial individual variability in glycemic response to identical meals, suggesting limited use of generic diet recommendation (Personalization)
- 20th century saw major changes in daily diets as a result human beings' average weight is increasing
- Overweight person may develop various diseases such as:
  - Musculoskeletal disorder
  - Hypertension
  - Type-2 diabetes

Motivation

Food Tracking is a challenge. Especially the written notes and questionnaires !!!

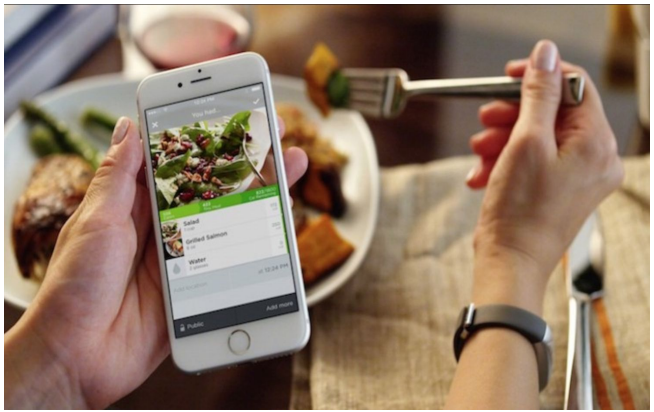
- Tedious
- Difficulty to summarize
- Time consuming
- Lack of information (how much?)



Motivation

Identifying food items from an image is a first step in simplifying the task of providing feedback to a patient regarding his diet and eating habits.

- Easier to log
- Easy reporting



Previous Work

Literature	Method	Dataset	Evaluation
Pishva et al. 2001	Shape-based segmentation. Textural, colour, cluster analysis to determine the bread type.	Private dataset. Contains 73 bread type (categories)	95% Top-1 accuracy
Ciocca et al. 2016	JSEG - for segmentation; SVM, KNN for classification	UNIMIB2016	Accuracy: KNN - 73.8%, SVM - 78.9%
Aguilar et al. 2018	Food Segmentation and Detection. YOLOv2, FCN, DarkNet- 19	UNIMIB2016	Precision - 0.841, Recall - 0.922

Previous Work

Literature	Method	Dataset	Evaluation
Ye et al. 2019	Mask R-CNN using MobileNet and ResNet, Multi SVM	10 food categories from MS COCO dataset	Accuracy: Multi SVM- 88.86%, Mask R-CNN (ResNet) - 92.35%
Freitas et al. 2020	Mask R-CNN, DeepLab V3, SegNet, ENet, FCN	Proprietary dataset: Handpicked 9 classes from 50	Mask R-CNN - mAP=0.87, Other method- mAP<0.79
Okamoto et al. 2021	Faster R-CNN, Deeplab V3	UEC-FoodPix Complete	mIoU=0.555 Accuracy=0.668

Dataset

- The data used in this study was made available by the MyFoodRepo app users between July 7th 2018 and June 8th 2020
- The dataset consists of 24,119 images containing a total of 39,325 segmented polygons
- The food images are categorized in 273 different classes with at least 35 annotations per class


water
wine-red
egg
apple
dark-chocolate
pasta-spaghetti
chips-french-fries
pizza-margherita-baked

Public dataset released in 3 phase:

- Top-40 classes
- Top-61 classes
- Top-273 classes.

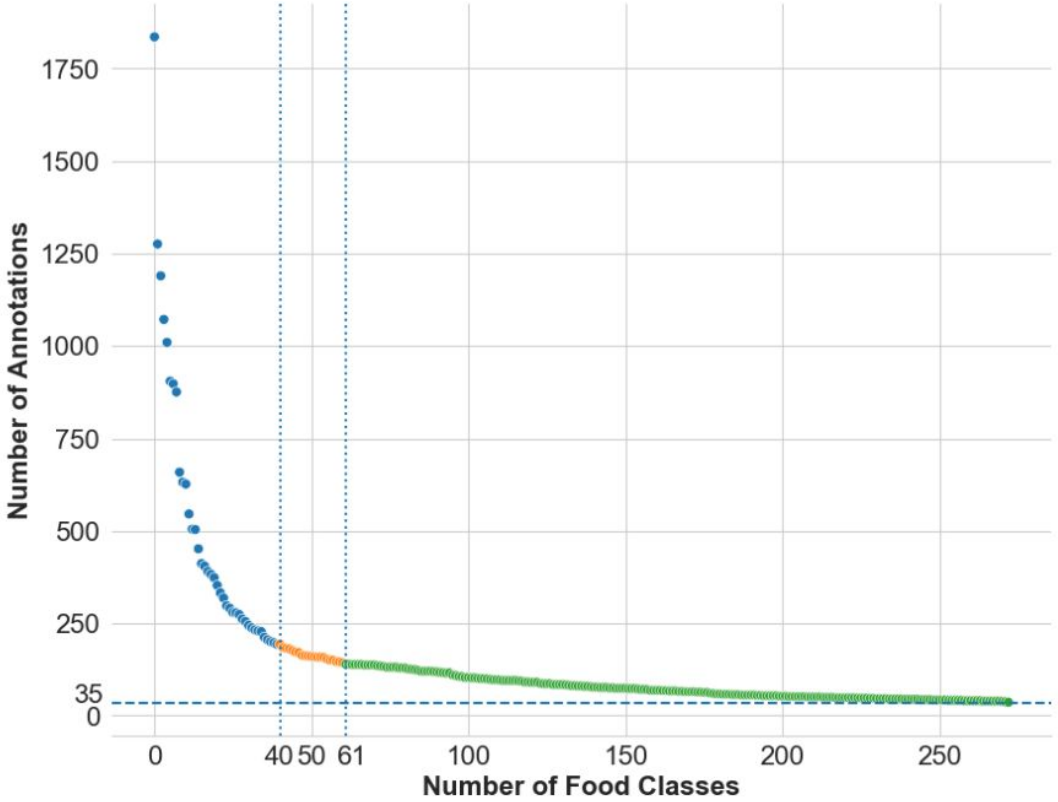


Fig: Distribution of number of annotations available across the food classes in the dataset

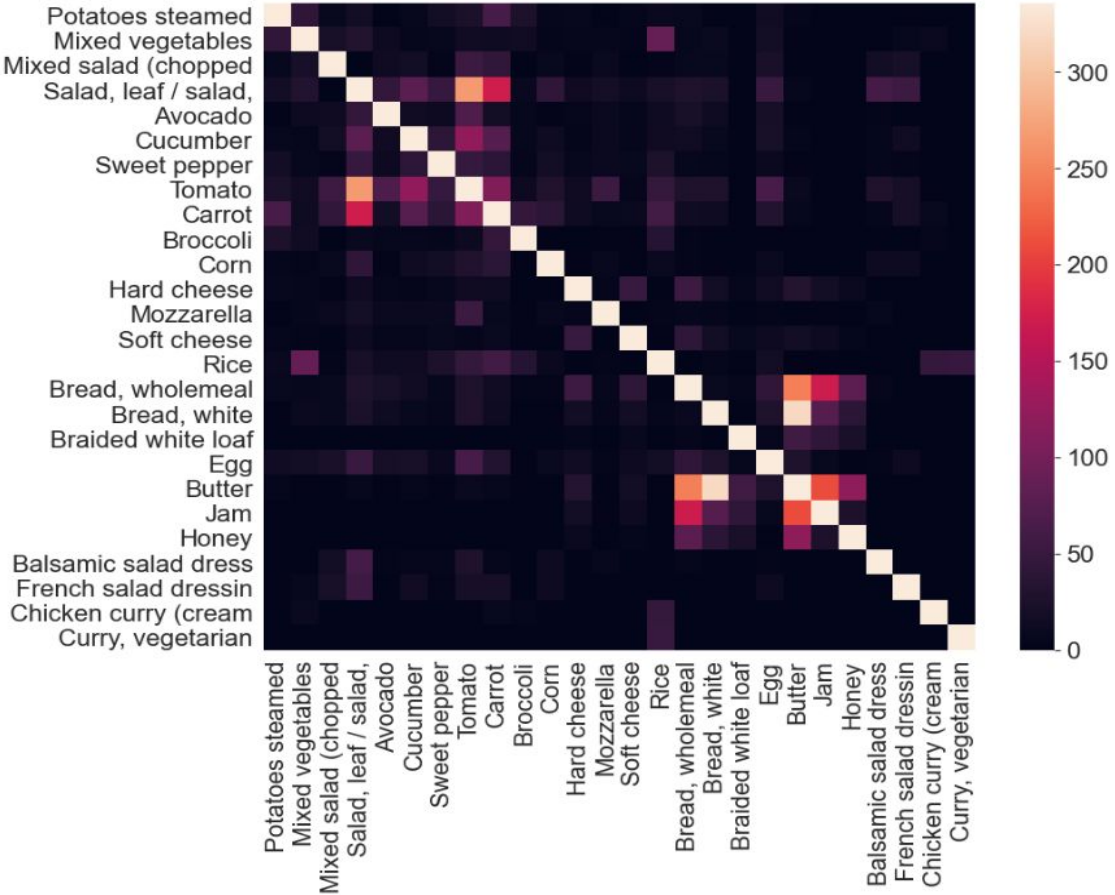
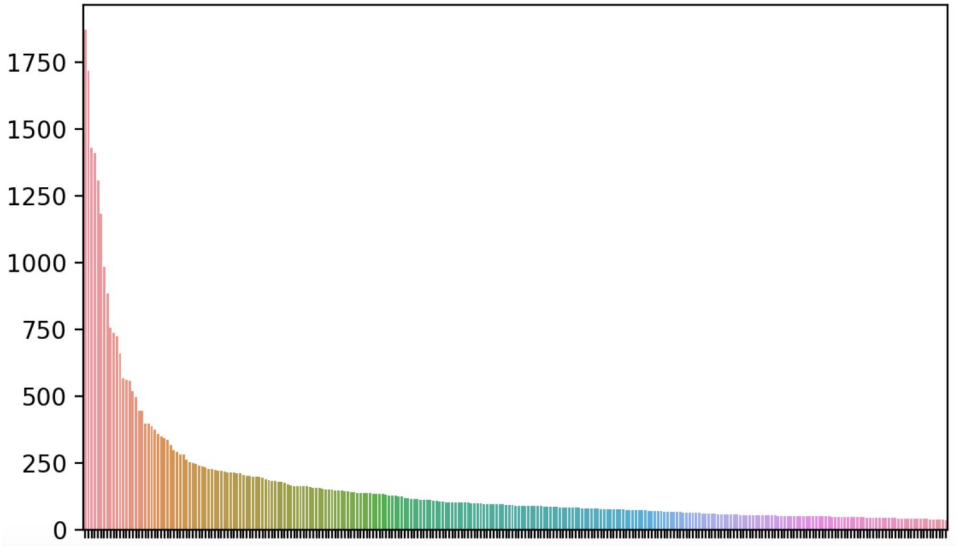


Fig: Co-occurence matrix of 26 food classes

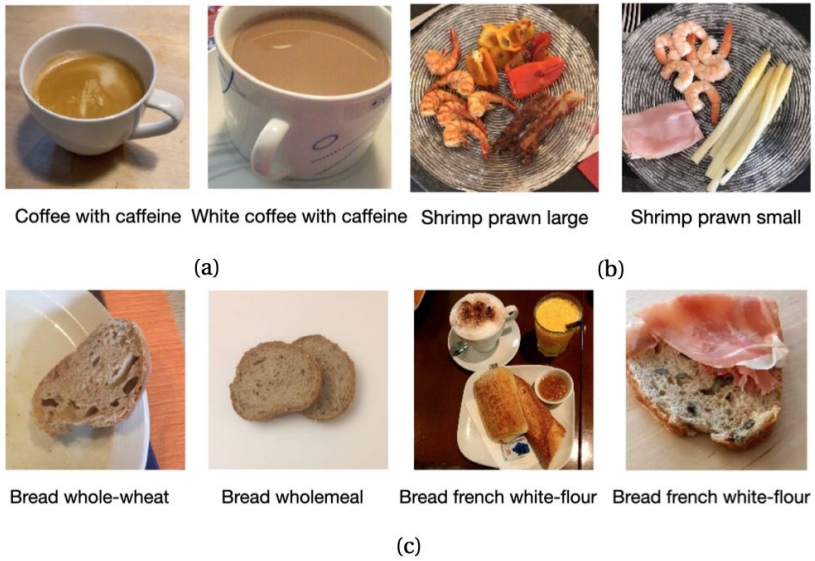
Challenges

1. **Class Imbalance**
   Top 15 categories cover 35% of the annotations.



Challenges

2. **Similar Foods**
   Many food items have similar visual properties but have different categories. Pictures with such categories create confusion for the model



Challenges

3. **Overlapping (Occlusion)**
   Many a time people put food on top of another food, maybe not by choice but some food is consumed in this fashion.

Wholemeal bread, Jam

Fresh cheese, wholemeal bread

Pasta, Cheese

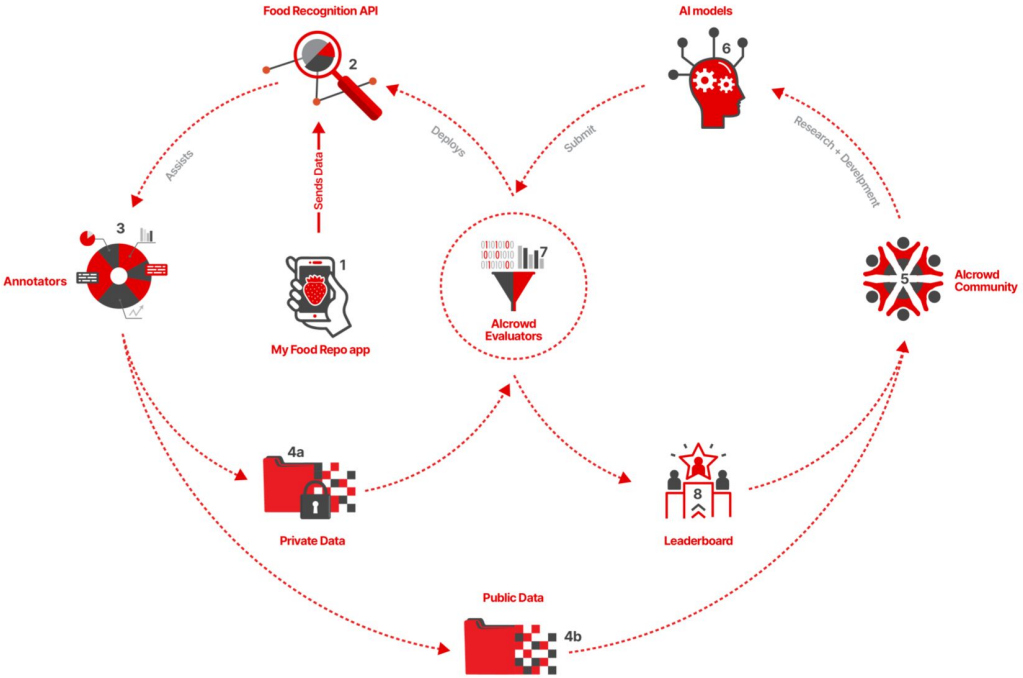
Challenges

4. **Partial Annotations**
   The pictures in the dataset are annotated by the people and therefore some of the images are partially annotated.

Mineral water, Artichoke, Quinoa, Sweet pepper, Savoury puff pastry, Shrimp / prawn (large), Chickpeas, Zucchini, Rice

Mayonnaise, Sushi, French beans, Semi-hard cheese, Soft cheese, Egg, Raw ham, French salad dressing

Ecosystem


Food Recognition API
2
AI models
6
Deploys
Submit
Research + Development
Assists
Sends Data
3
Annotators
1
My Food Repo app
7
Alcrowd
Evaluators
Alcrowd
Community
4a
Private Data
8
Leaderboard
Public Data
4b

Methods

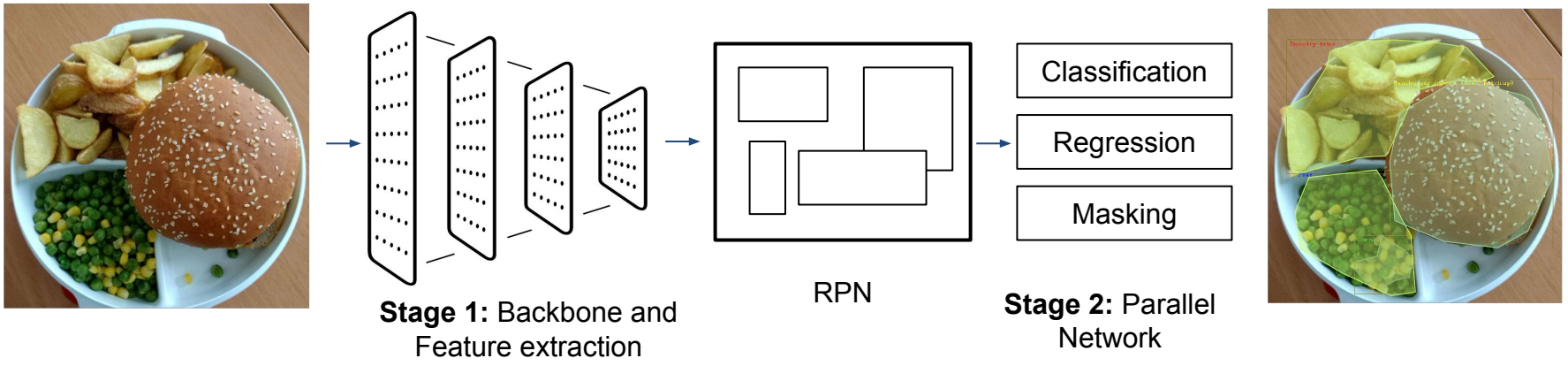


Fig: Anatomical architecture of CNN-based Instance Segmentation

Methods

Top performing methods used by the participants

- Mask R-CNN (Among the first few methods for Instance Segmentation)
  Comprises - FPN Backbone, RPN, Heads
- Cascade R-CNN
  Comprises - FPN Backbone, RPN, cascaded heads (3)
- Hybrid Task Cascade (HTC)
  Comprises - FPN Backbone, RPN, cascaded heads (3) with reinforcing the flow of information between masking heads
- DetectoRS
  Comprises - RFP Backbone, RPN, cascaded heads (3) with reinforcing the flow of information between masking heads

All these methods are CNN based and anatomically work in two-stages

What Didn’t Worked?

- Additional classification model to condition the predictions
- Hierarchical filtering based on food ontology
- Co-occurrence matrix to filter unlikely predictions
- Putting weights to the classes

Evaluation: Improving With Time

First benchmark for food image recognition was conducted in 4 rounds.

	Number of Food Categories	Number of Training Images	Number of Training Annotations	mean Average Precision IoU >0.5	mean Average Recall IoU >0.5
Round 1	40	5545	7735	0.573	0.831
Round 2	61	7949	11468	0.633	0.886
Round 3	273	24119	39325	0.551	0.884
Round 4	273	24119	39325	**0.568**	0.767

Ablation Study

	Backbone	Experiment	mean Average Precision$_{IoU>0.5}$	mean Average Recall$_{IoU>0.5}$
Mask RCNN	ResNet50	Baseline	0.473	0.707
		MultiScale Training	0.482	0.732
		Weighted Loss	0.479	0.713
		Train Time Augmentation	0.487	0.741
		Combined	0.506	0.809
	ResNet101	Baseline	0.485	0.706
		Combined	0.523	0.817
	ResNeXt101	Baseline	0.474	0.710
		Combined	0.535	0.825
Hybrid Task Cascade (HTC)	ResNet50	Baseline	0.484	0.800
		Combined	0.525	0.861
	ResNet101	Baseline	0.491	0.798
		Combined	**0.539**	0.867

Conclusion

- Introduce a novel instance segmentation dataset with 24,119 real world images from 273 food classes
- Dataset comprises of visually unbiased sample of food images as we used personalized health cohort of generally healthy subjects
- Improved models will reduce human-time to annotate data
- The benchmark attracted 1065 participants (as of 17th April, 2021) from 71 countries

Our Contribution

- The dataset is licensed under Creative Commons CC-BY-4.0. It can be downloaded from here: https://www.aicrowd.com/challenges/food- recognition-challenge/dataset_files.
- Source code to re-produce the results can be found here: https://gitlab.aicrowd.com/aicrowd/research/myfoodrepo- experiments.

Thank You