MADiMa 2023

A Comparative Analysis of Sensor-, Geometry-, and Neural-Based Methods for Food Volume Estimation

Lubnaa Abdur Rahman, *Ioannis Papathanail, Lorenzo Brigato, and Stavroula Mougiakakou*

ARTORG Center for Biomedical Engineering Research

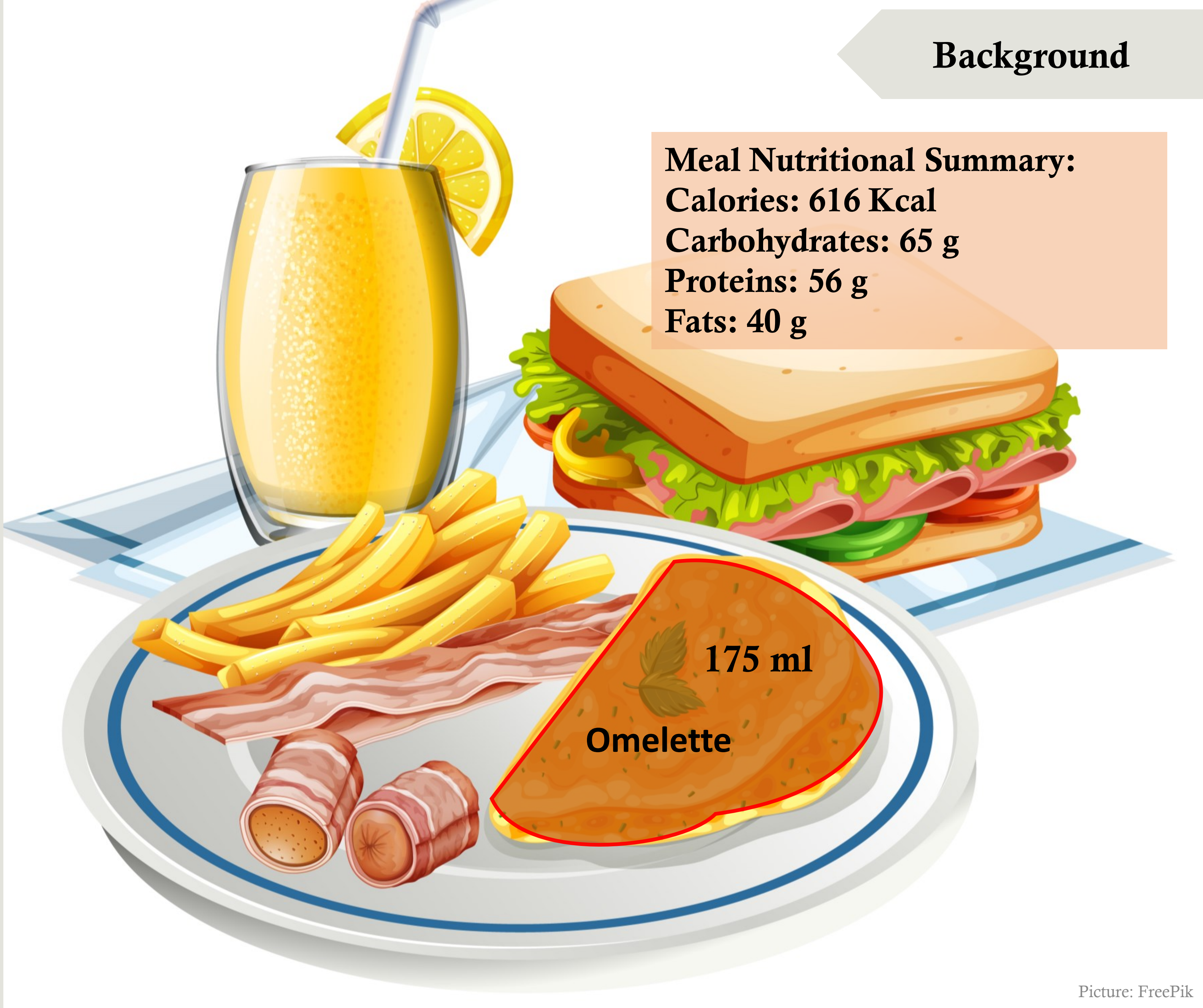
University of Bern

MELISSA

UNIVERSITÄT BERN

Automatic Dietary Assessment

Background



Picture: FreePik

Automatic food volume estimation remains a challenge

Error rate as high as:

85 %[1]

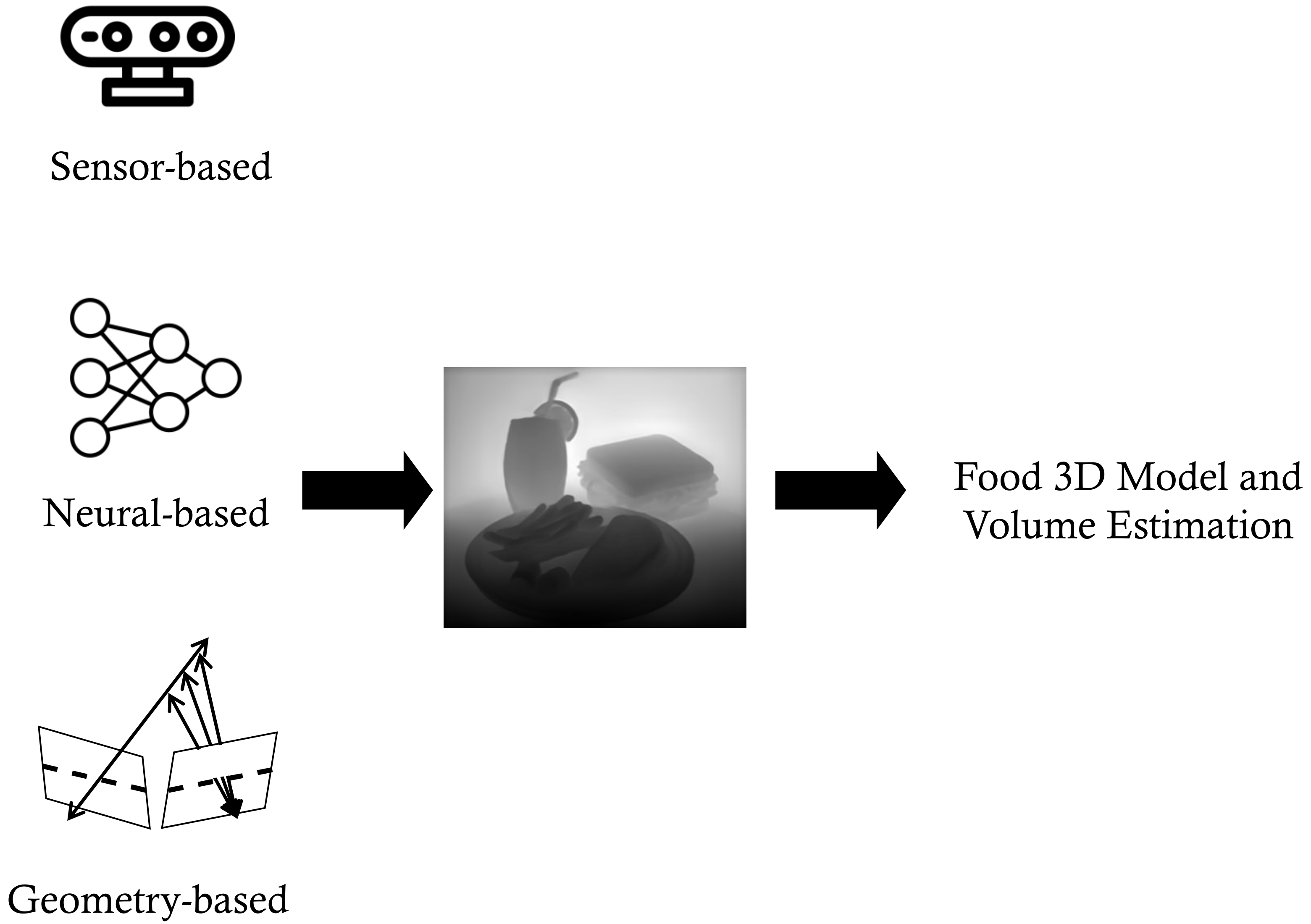
Traditional Methods

[1] Amugongo, L.M., Kriebitz, A., Boch, A. and Lütge, C., 2022, December. Mobile computer vision-based applications for food recognition and volume and calorific estimation: A systematic review. In Healthcare (Vol. 11, No. 1, p. 59). MDPI.

Aim

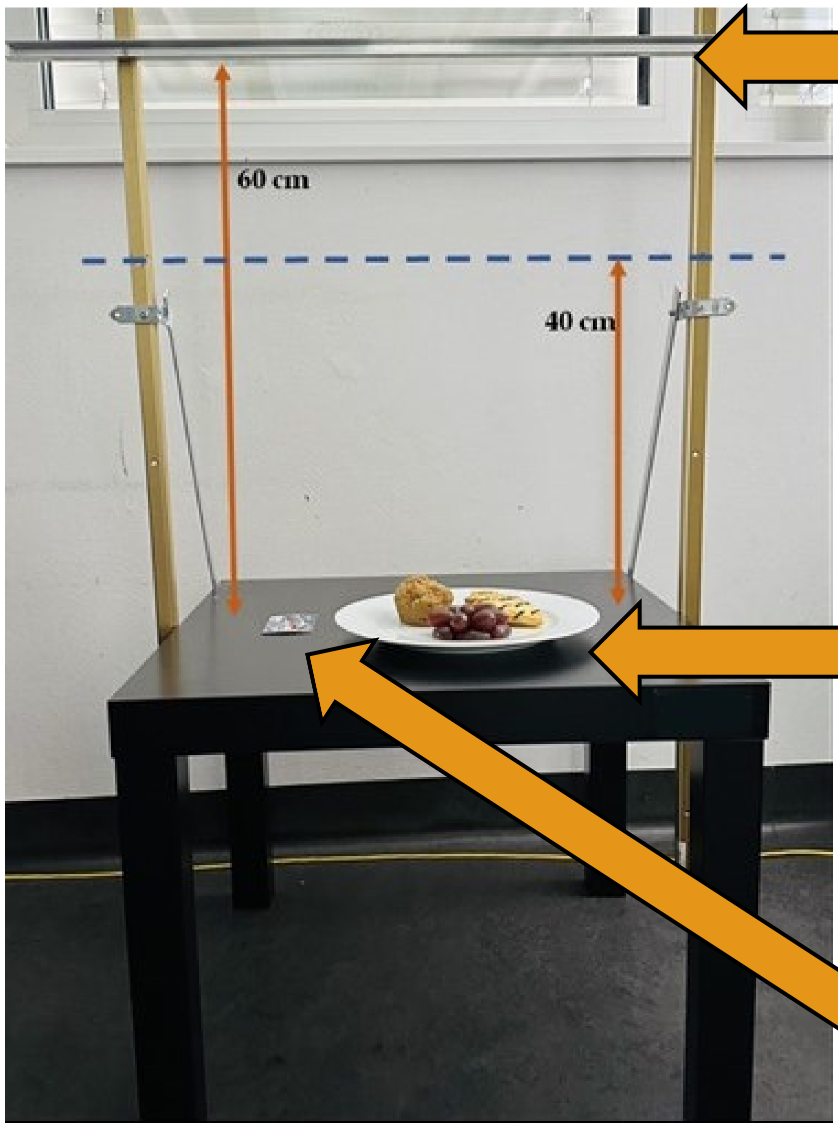
Automating via image analysis

- Translating the 2D Food into 3D
- Relies on the presence of a depth map
- Aim to assess the accuracy and practicality of these various methods in different scenarios.



Setup

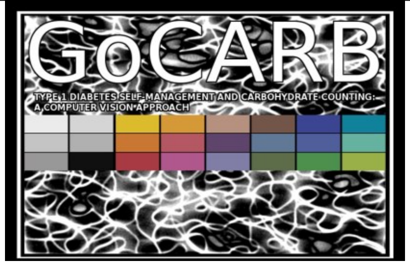
- Constant distance: 40 cm and 60 cm
- Constant lighting condition
- Top view: 90° (+75° for geometry-based)
- Reference card for further scaling



Capture device

Plated meal

Reference Card



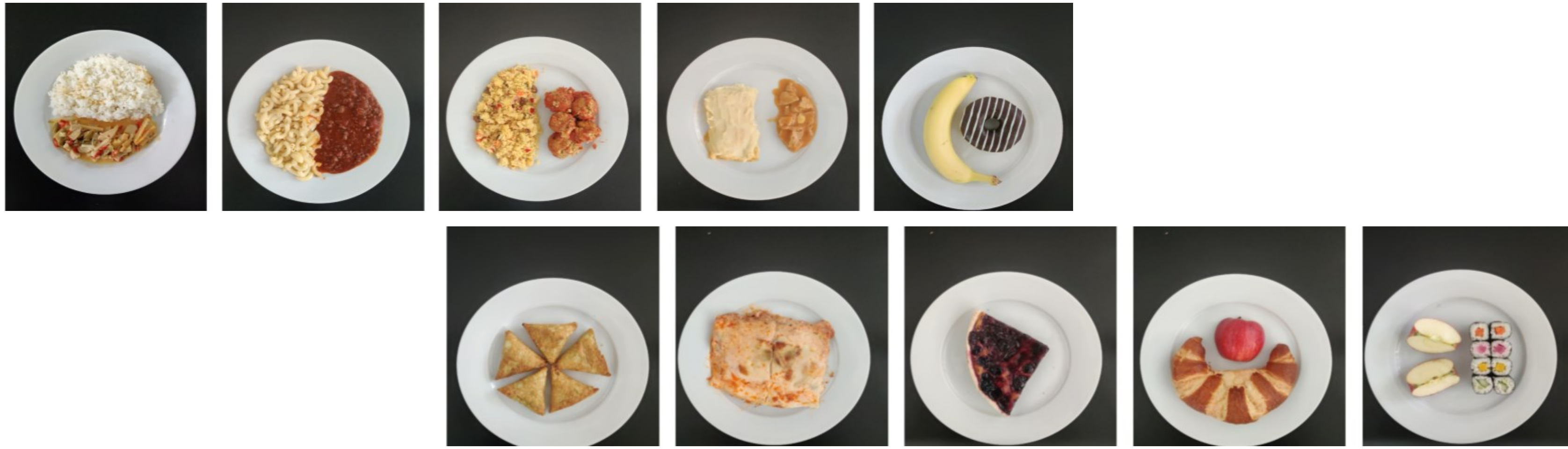
Methods

Data

- 20 Meal Images
- Images at 40 cm and 60 cm
- Captured with 3 different devices
- Depth from 2 different sensors

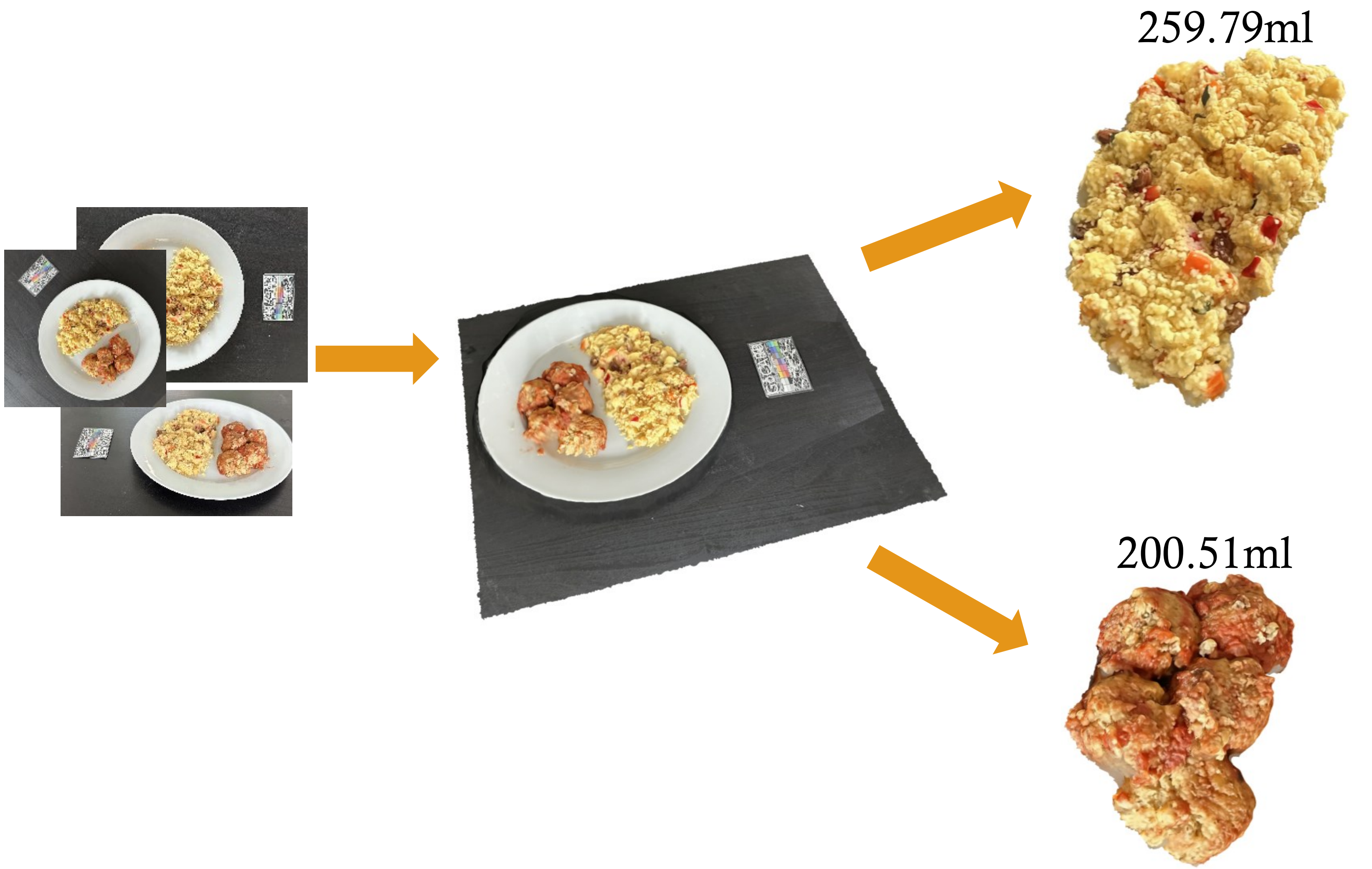
10 Plastic Food Meals

10 Real Food Meals



Ground Truth

- ≈25 images
- 360 view of meals
- Scaling
- Splitting of 3D meal into separate food items
- Volume computation



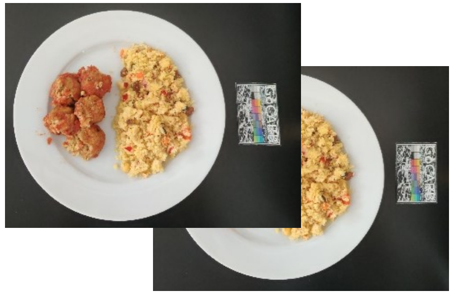
Capture

- RGB-D captures with two different sensors
- For geometric approach: 2 stereo RGB images
- For neural approach: single RGB

Intel D455

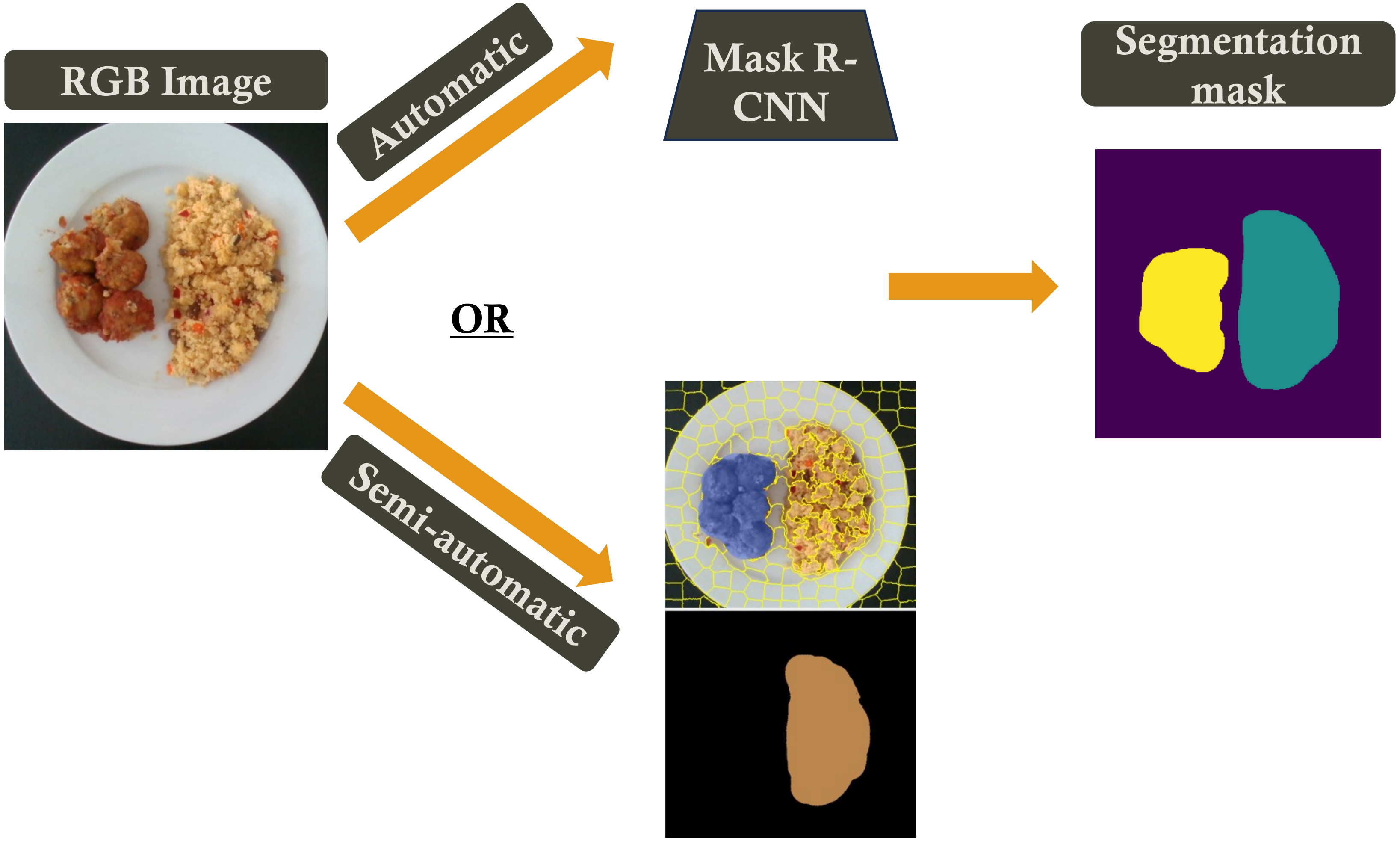
iPhone 14 Pro

goFOOD™ app on OnePlus 7 Pro



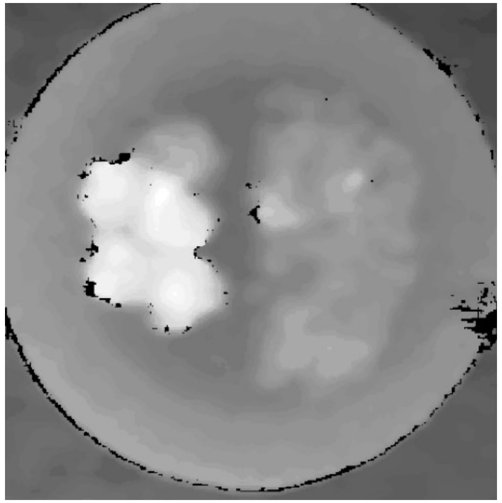
Segmentation

- Automatic food segmentation using Mask R-CNN
- If mask was unsatisfactory, semi-automatic segmentation



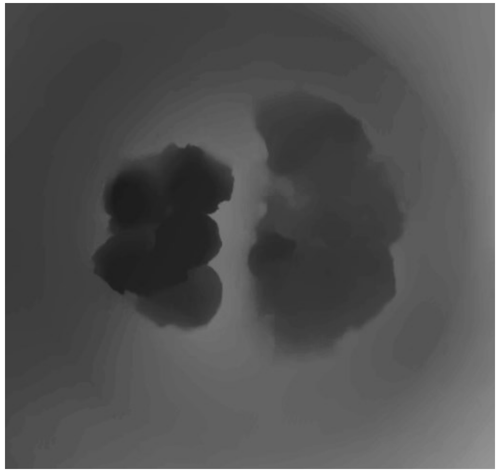
Sensor-based Depth: Intel RealSense D455

- Stereoscopic depth sensor
- Absolute depth values
- Depth filtering for extreme values
- Depth values are misestimated at 60 cm



Sensor-based Depth: LiDAR

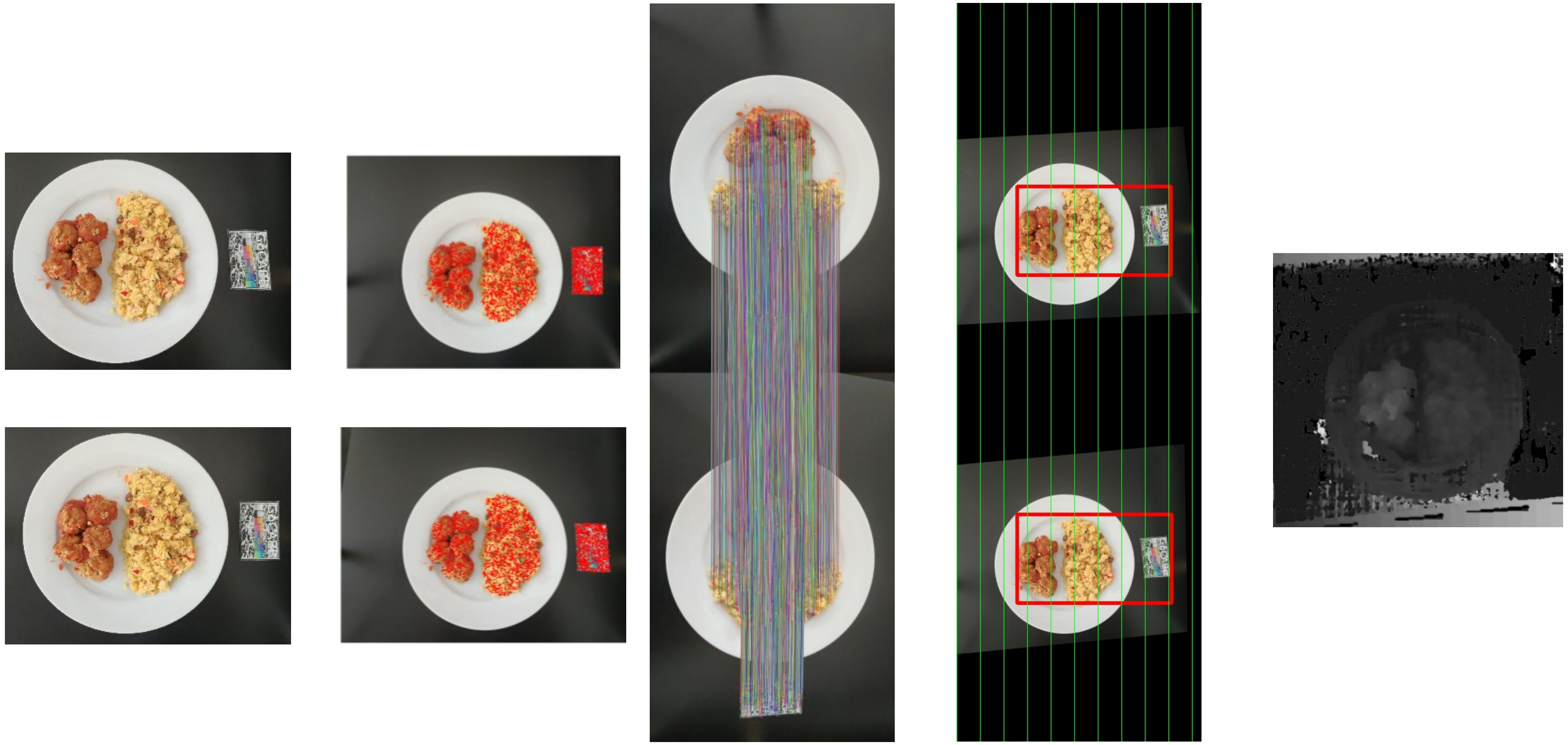
- Remote sensing method that uses emitted light to record depth
- Direct RGB-D Capture
- Depth filtering for extreme values



Geometry-based Depth: Stereo Matching

- Stereo Pair of RGB Images [1]
- Detected key points filtering
- Rectification
- SGBM based disparity map
- Converted to Depth map
- Scaling using reference card

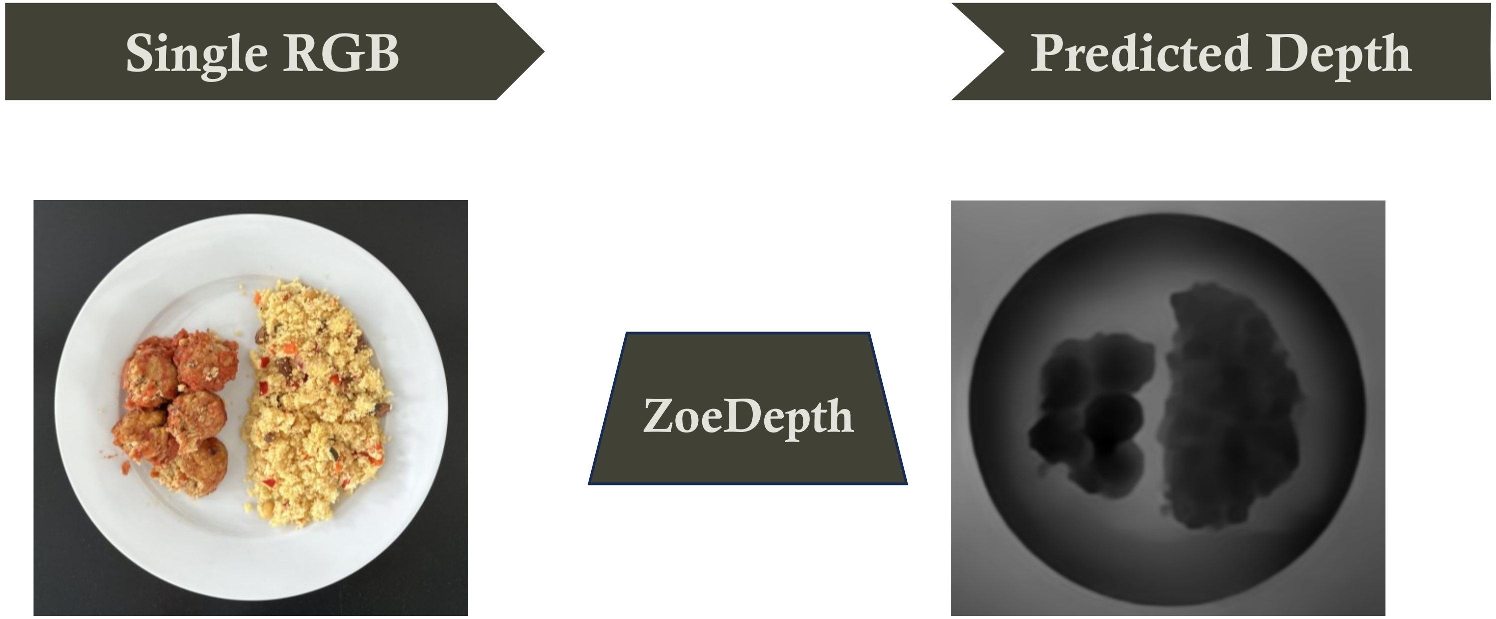
Stereo Pair → Key point detection, filtering, and matching → Rectification → Inferred Depth



[1] Lu, Y., Stathopoulou, T., Vasiloglou, M.F., Pinault, L.F., Kiley, C., Spanakis, E.K. and Mougiakakou, S., 2020. goFOODTM: an artificial intelligence system for dietary assessment. *Sensors*, *20*(15), p.4283.

Neural-based Depth: ZoeDepth

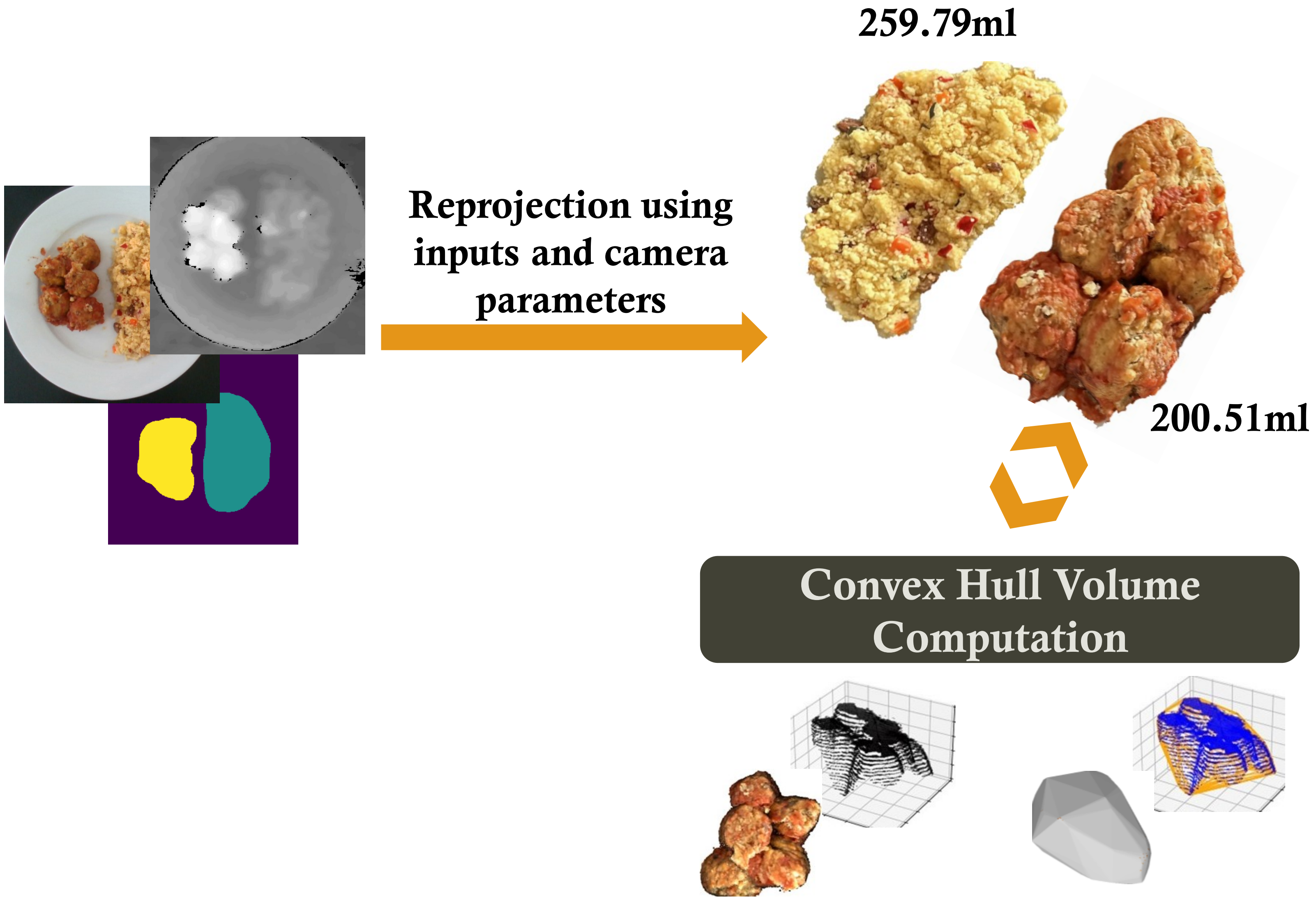
- CNN-based model integrating both relative [1] and absolute depth [2]
- Single RGB image
- Depth values without units
- Scaling using reference card



[1] Lasinger, K., Ranftl, R., Schindler, K. and Koltun, V., 2019. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *arXiv preprint arXiv:1907.01341*.
[2] Bhat, S.F., Birkl, R., Wofk, D., Wonka, P. and Müller, M., 2023. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*.

Reprojection to 3D & Volume Computation

- Reprojection to 3D point clouds (PCDs)
- Outlier removal using nearest neighbor
- Enclosed food items into polygon (convex hull)
- Volume computed



Overall

- LiDAR lowest error
- Intel RealSense D455 achieved second-best results at 40 cm, followed by neural- and geometry-based approaches
- Geometry-based method performed better at 60cm

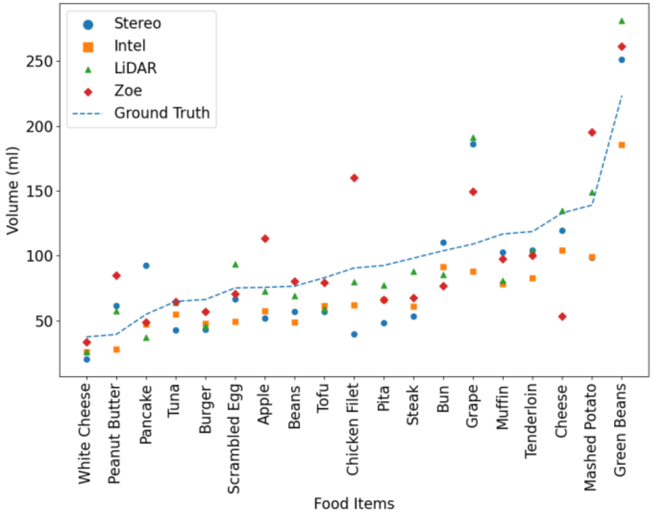
Mean absolute percentage error

Method	Plastic		Real	
	40 cm	60 cm	40 cm	60 cm
Intel RealSense D455 sensor	26.15	36.41	25.06	41.07
LiDAR sensor	**21.32**	**22.76**	**17.45**	**16.40**
Geometry-based	30.54	29.99	27.21	23.57
Neural-based	30.40	35.61	26.41	30.25

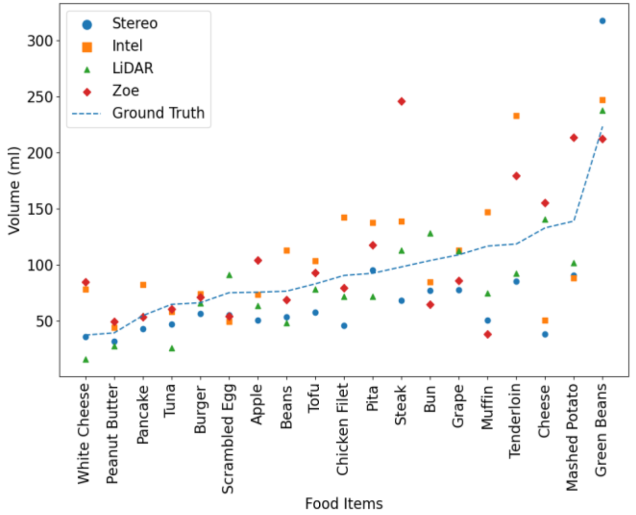
Estimated vs Ground Truth

- Less errors for real food. Plastic foods are reflective
- Neural-based with most variation in results.

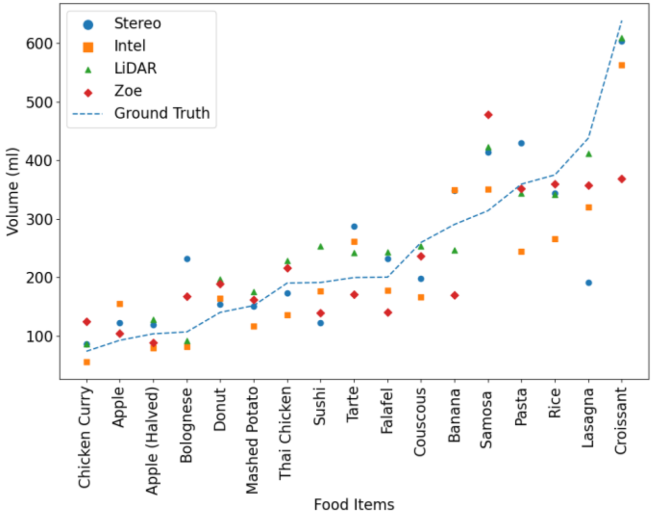
Plastic Food 40 cm



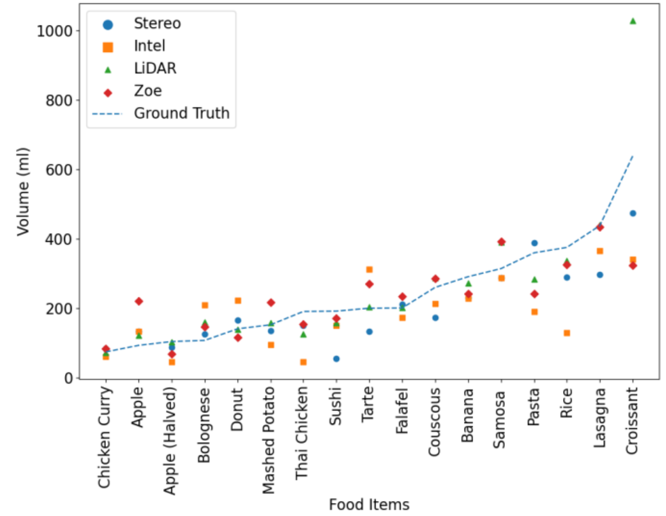
Plastic Food 60 cm



Real Food 40 cm



Real Food 60 cm



Depth based automatic food volume estimation

Sensor-based: Intel D455

- Effective in controlled environments
- Works best for shorter distances
- May face limitations in unstructured settings

Sensor-based: LiDAR

- Accurate, reliable, and flexible
- Can be used on the go
- Limited to hardware availability

Geometry-based: Stereo Matching

- Widely applicable
- Balance between accuracy and hardware availability
- Less user friendly

Neural-based: ZoeDepth

- Adequate accuracy with single image
- Further fine tuning required
- Not limited by specialized hardware

- Assess performance and application of diverse methods
- Plan to release dataset containing 20 meals captured at 40 cm and 60 cm publicly
- LiDAR demonstrates superior performance
- Future directions: fine-tuning the depth model for similar food items and conducting additional experiments while expanding the dataset.

Questions