MADiMa 2018

A Multi-Task Learning Approach for Meal Assessment

Ya Lu, Dario Allegra, Mario Anthimopoulos, Filippo Stanco, Giovanni Maria Farinella, Stavroula Mougiakakou

UNIVERSITÄT BERN
ARTORG CENTER
BIOMEDICAL ENGINEERING RESEARCH

UNIVERSITÀ degli STUDI di CATANIA

Contents

- Introduction
- Method
- Experimental results
- Conclusion

Introduction - Motivation

Type 1 Diabetes

Diet-related chronic diseases

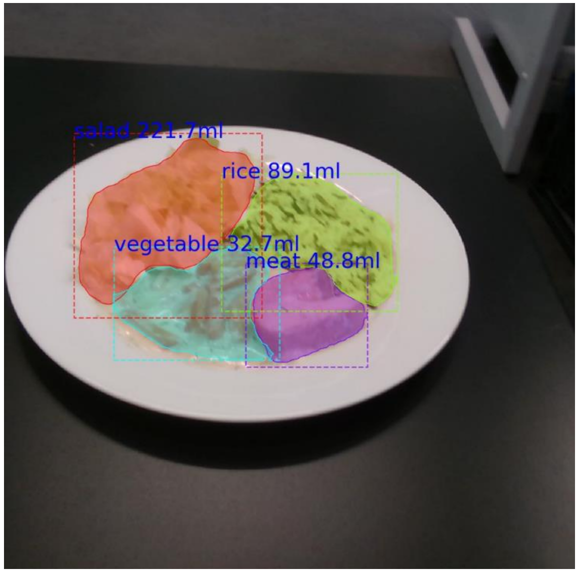
Obesity

Malnutrition

Introduction - Goal

- Propose a multi-task learning approach to realize food recognition, segmentation and volume estimation through one network.

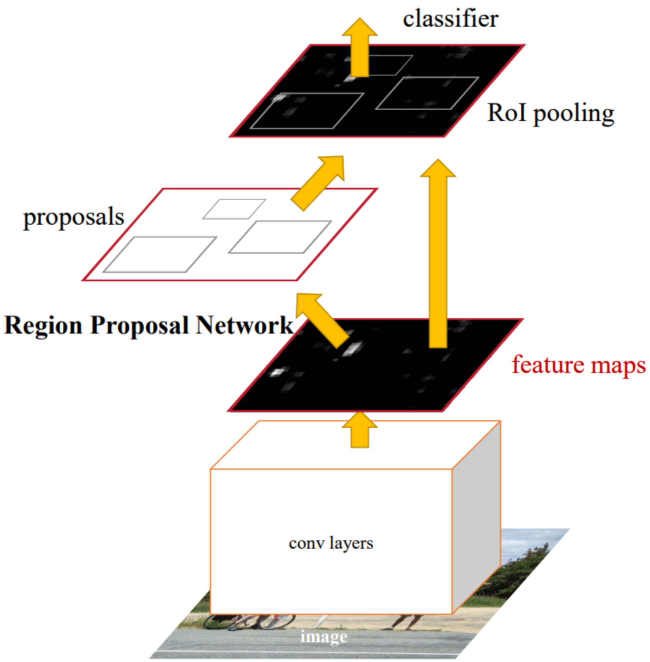
Single RGB image input



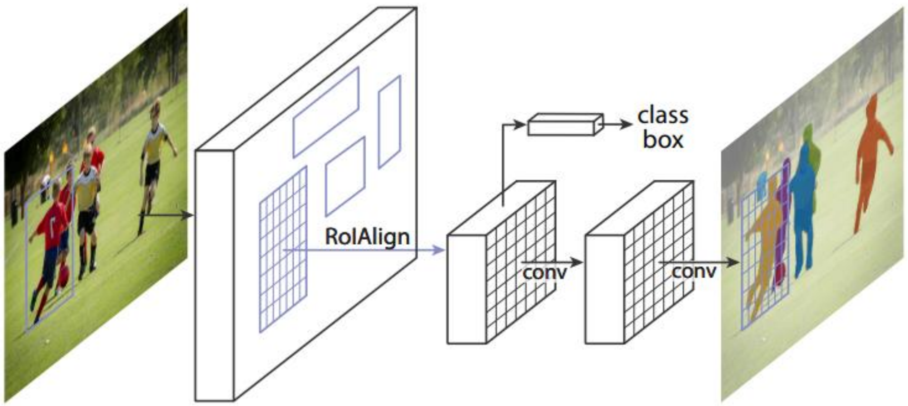
Output

Method - Network architecture (1/4)

Introduction of MaskR-CNN:



Architecture of Faster R-CNN [1]

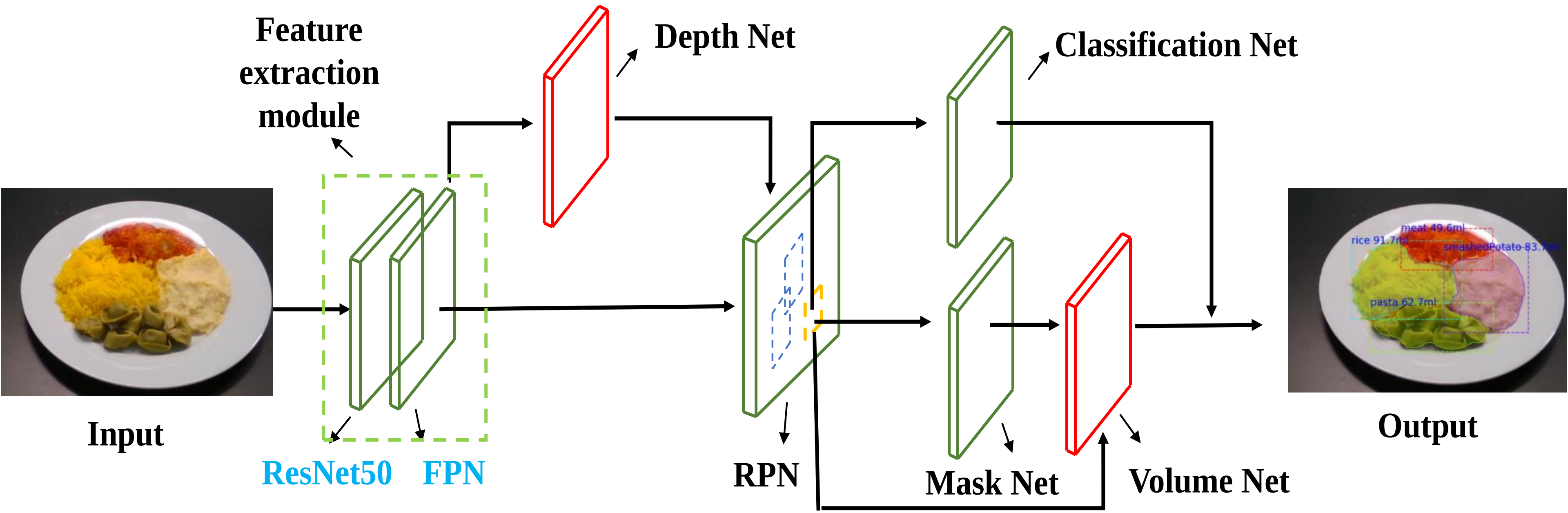


Architecture of MaskR-CNN [2]

[1] Shaoqing Ren, et al., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016
[2] Kaiming He, et al., MaskR-CNN, 2017

Method - Network architecture (2/4)

Multi-task network architecture:

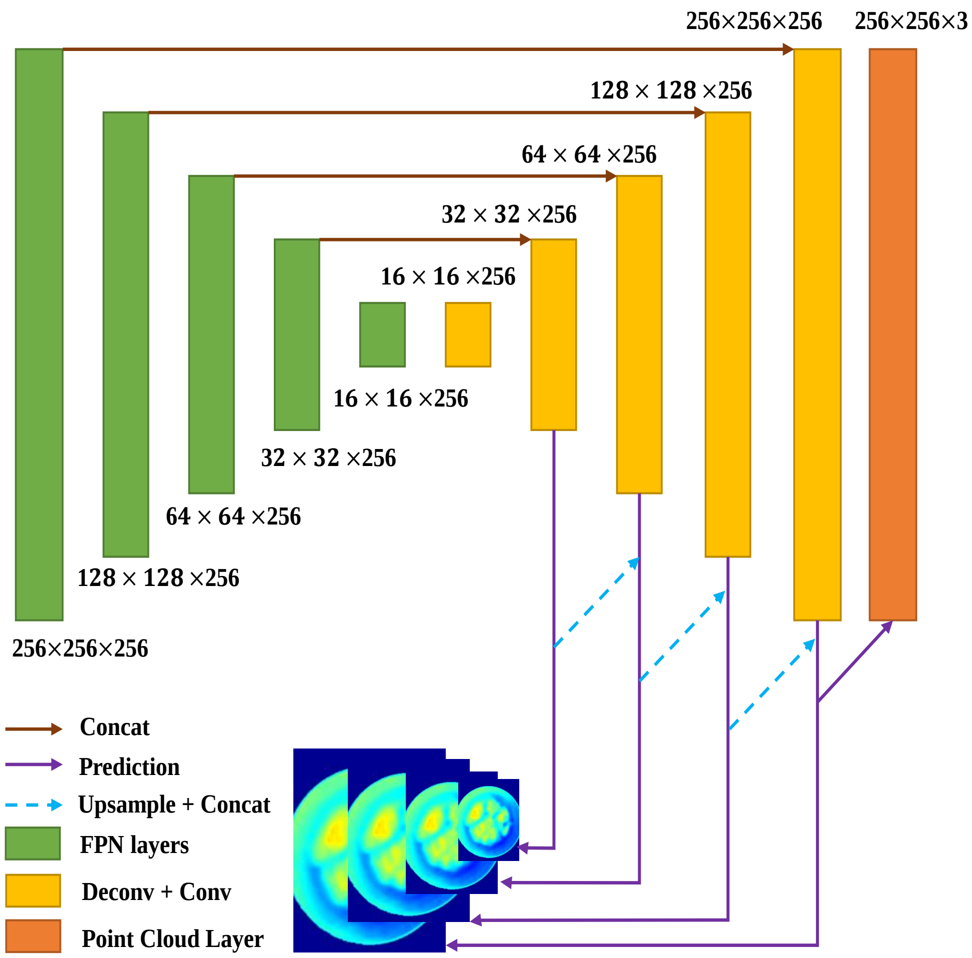


Same with MaskRCNN

Proposed by this paper

Method - Network architecture (3/4)

Depth Net



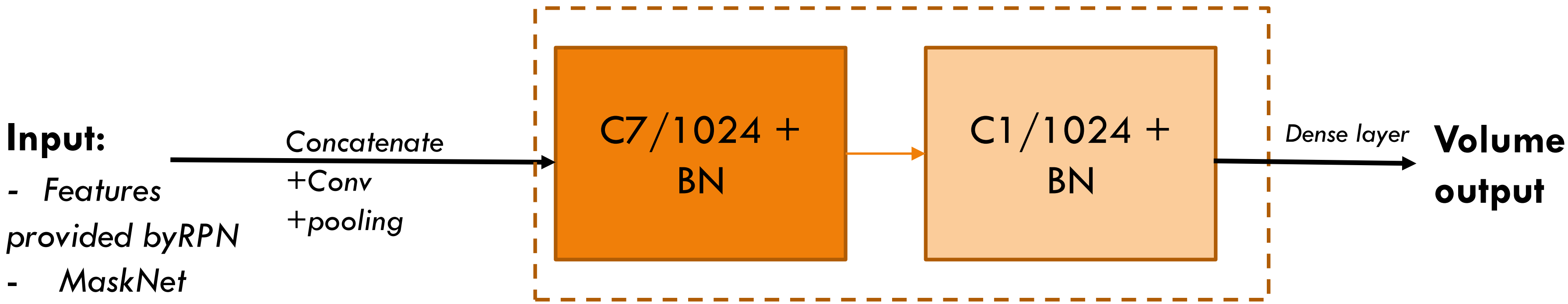
Convert depth image to point cloud:

$$X_I^i = \begin{bmatrix} x_I^i \\ y_I^i \\ z_I^i \end{bmatrix} = K^{-1} \begin{bmatrix} u_I^i \\ v_I^i \\ d_I^i \end{bmatrix}$$

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

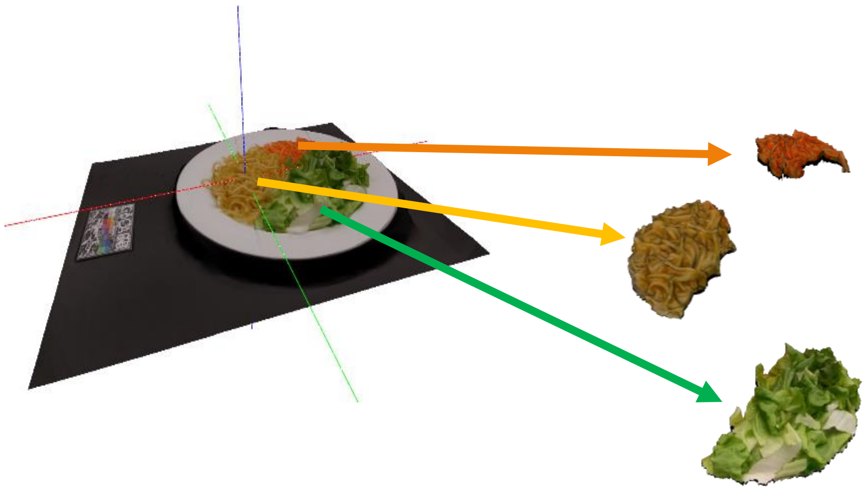
Method - Network architecture (4/4)

- Volume Net



Experimental results (1/8)

- Dataset – Madima17 database
  - 80 central-European meals, 2-4 food items per meal
  - RGB-D image pairs captured at different angle of view and distance for each meal
  - Food categories, segmentation map and volume are annotated

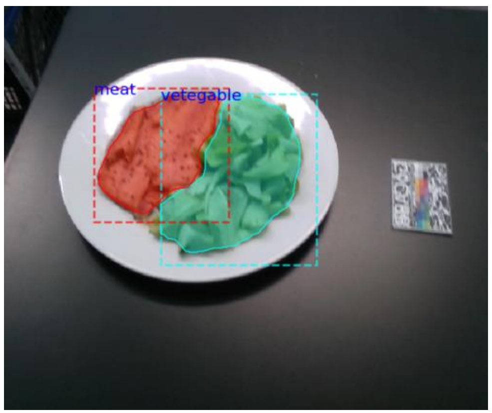


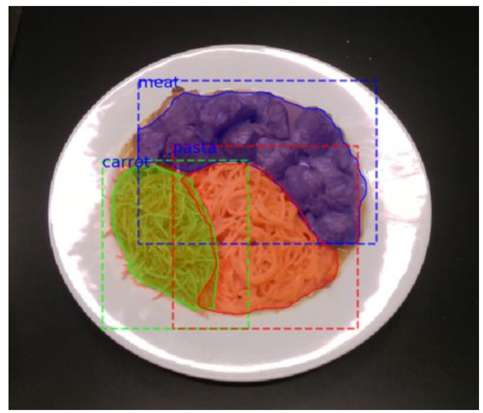
The experiments are trained with full set, while tested on different datasets.

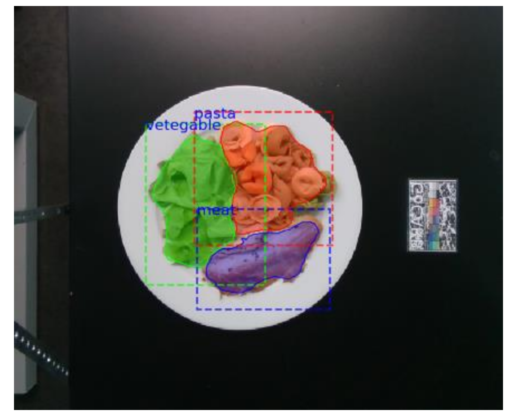
- Fixed set
  - 90°, 40cm; 80 images
- Free set
  - Random angle and distance; 160 images
- Full set
  - 90°, 60°, 40cm, 60cm + free set; 480 images

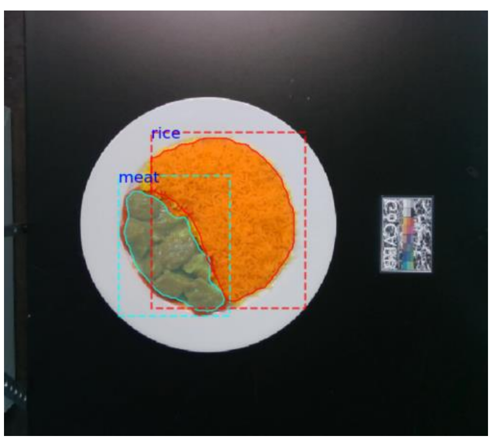
Experimental results (2/8)

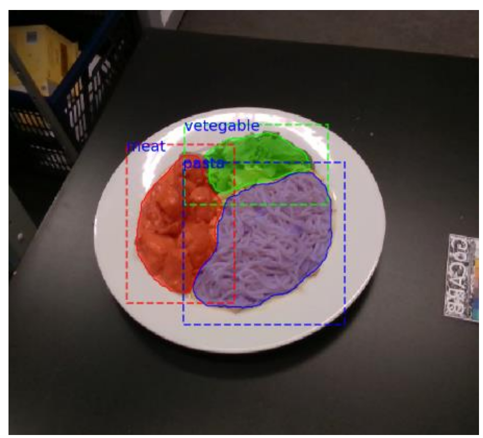
- Food segmentation and recognition

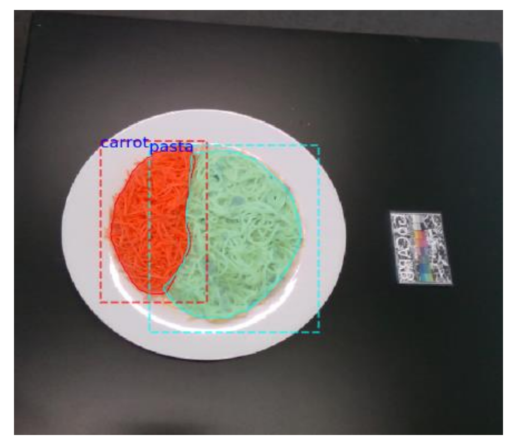












Experimental results (3/8)

- Food segmentation & recognition
  - Evaluation metrics
    - F-value

$$NI_{\min}(T \to S) = Min_i \left(\frac{Max_j(|S_i \cap T_j|)}{|S_i|} \right)$$

$$NI_{sum}(T \to S) = \frac{\sum_i Max_j(|S_i \cap T_j|)}{\sum_i |S_i|}$$

$$F_x = \frac{2 \times NI_x(T \to S) \times NI_x(S \to T)}{NI_x(T \to S) + NI_x(S \to T)}, x = \min \textit{ or } sum$$

    - AP

$$mAP = \frac{1}{10} \sum_{IoU} AP_{IoU}, \; IoU \in [0.5:0.05:0.95]$$

Experimental results (4/8)

- Food segmentation & recognition

Comparison of Segmentation method

	Fixed set		Full set	
Method	F_{sum}(%)	F_{min}(%)	F_{sum}(%)	F_{min}(%)
Proposed	**94.36**	**83.90**	**94.10**	**78.18**
Method in [3]	93.69	74.26	-	-
Method in [4]	92.47	73.36	91.83	75.33

[3] *D. Allegra, et al., A Multimedia Database for Automatic Meal Assessment Systems. Madima Workshop, 2017.*
[4] J.Dehais, et al., Dish Detection and Segmentation for Dietary Assessment on Smartphones. *Madima Workshop, 2015.*

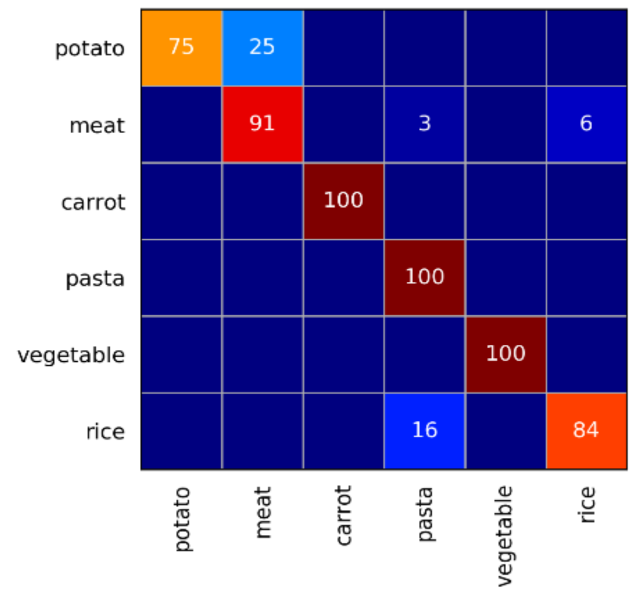
Experimental results (5/8)

- Food segmentation & recognition

Quantitative results using AP measures

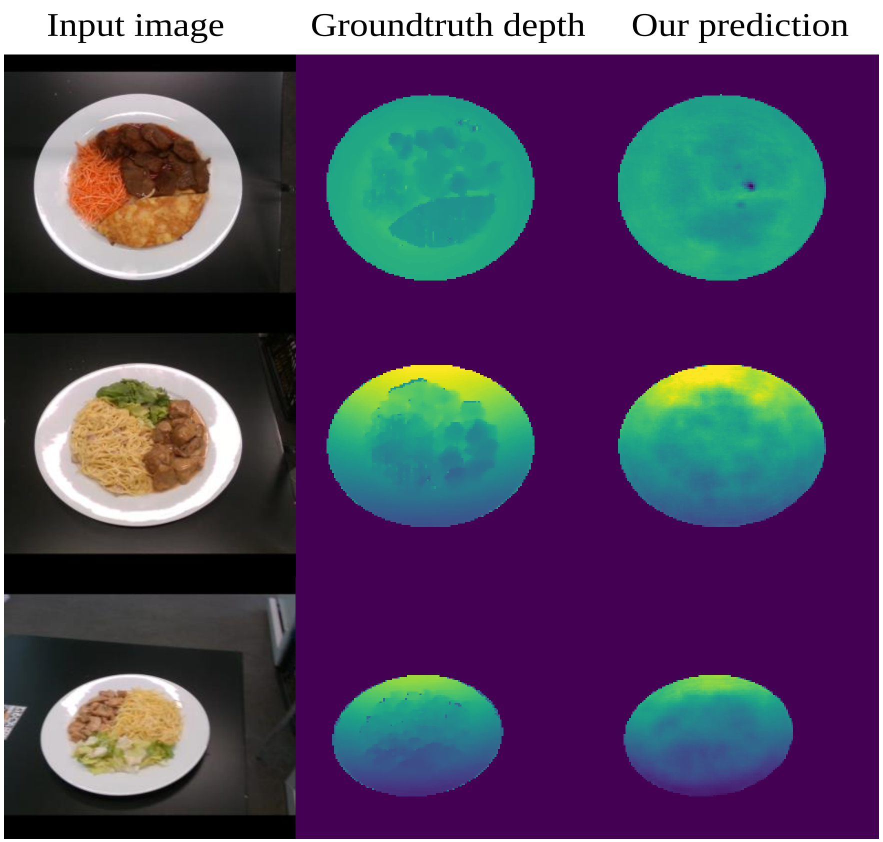
Dataset	mAP (%)	AP_{50} (%)	AP_{75} (%)
Fixed	69.4	90.4	85.7
Free	63.2	83.7	79.6
Full	64.7	85.1	79.1

Confusion matrix on Full set



Experimental results (6/8)

- Depth estimation



	Free set		Full set	
Method	MAD (mm)	ARD (%)	MAD (mm)	ARD (%)
Proposed	6.75	1.25	5.71	1.13
Method in [3]	8.64	1.76	6.03	1.25

Experimental results (7/8)

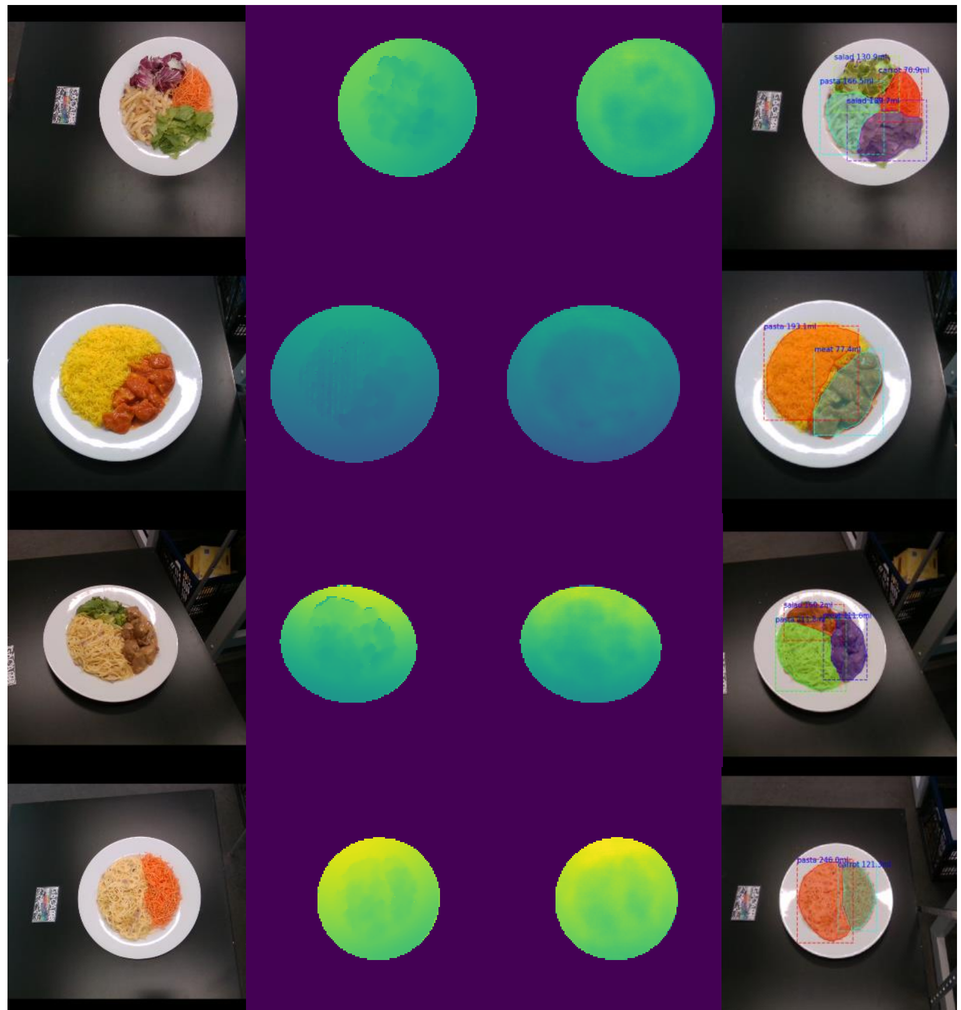
- Volume estimation

	Food item's average percentage error			
Method	Fixed (%)	Free (%)	Full (%)	Process time (s)
Proposed	**17.5**	**19.1**	**19.0**	**<0.2**
3D Reconstruction [3]	22.6	36.1	33.1	5.5

Experimental results (8/8)

- Some result samples of the whole pipeline:

Input image | Depth groundtruth | Depth Prediction | Meal assessment result



Conclusion

- A multi-task learning approach is proposed for meal assessment, which only needs one RGB image as input.
- The proposed method achieved superior performance compared with state-of-art methods.
- Future work includes the extension of the methods to images with multiple dishes and database with higher diversities.

Thank you for the attention!

Questions?