A Multi-Task Learning Approach for Meal Assessment

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

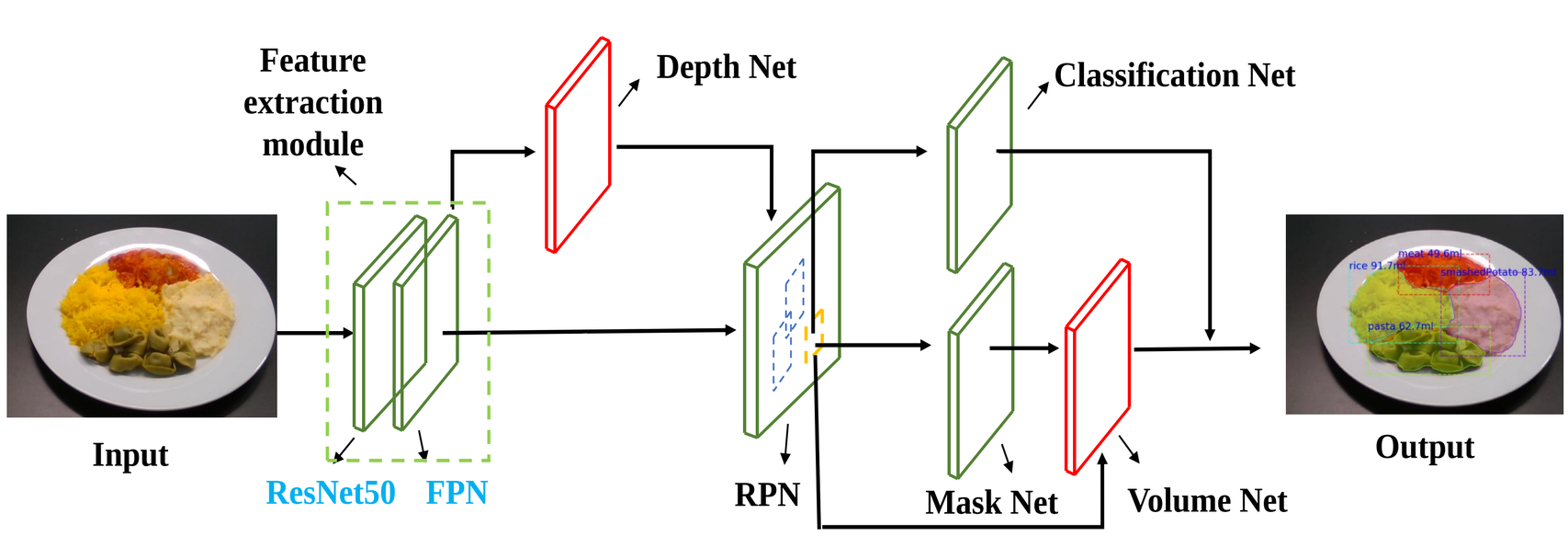
{ya.lu, marios.anthimopoulos, stavroula.mougiakakou}@artorg.unibe.ch; {allegra, fstanco, gfarinella}@dmi.unict.it;

Introduction

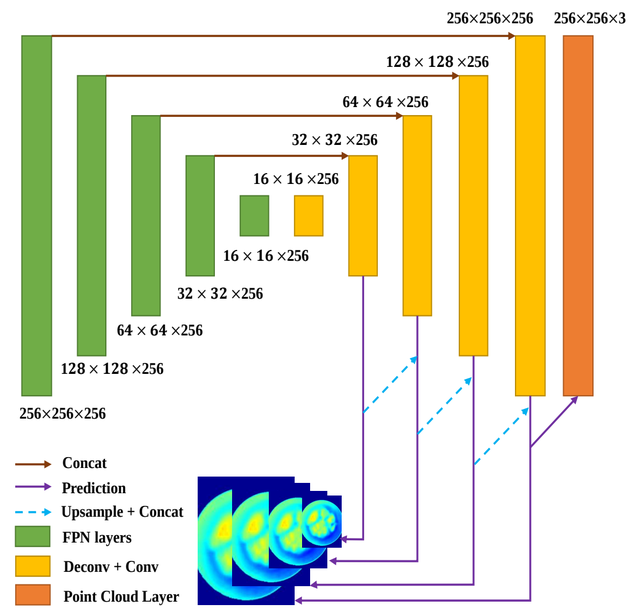
Key role in the prevention of diet-related chronic diseases plays the balanced nutrition together with a proper diet. The conventional dietary assessment methods are time-consuming, expensive and prone to errors. The advances in the fields of artificial intelligence and computer vision permitted the use of meal image to assess the nutrient content usually through three steps: food segmentation, recognition and volume estimation.

In this research work, we propose the use of one RGB meal image as input to a Multi-Task Learning based Convolutional Neural Network (MTL-CNN). The proposed approach achieved outstanding performance, while a comparison with state-of-the-art methods indicated that the proposed approach exhibits clear advantage in accuracy, along with a massive reduction of processing time.

Method

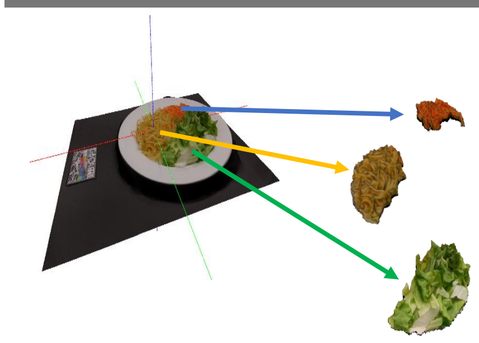


MTL-CNN architecture



Depth Net

Database

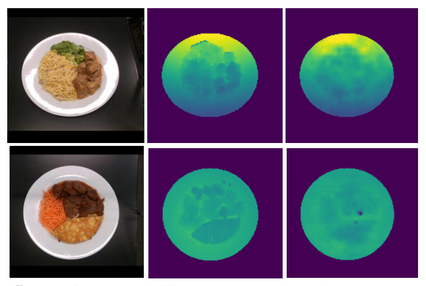


- 80 central-European meals
- 6 RGB-D image pairs at different angles and distances for each meal
- Food category, segmentation map and volume are annotated
- Three (3) datasets: Fixed distance and angle (FDA), free-hand (FH) and full (FULL)

Depth Estimation

Comparison of depth prediction method per dataset

	FH dataset		FULL dataset	
Method	***MAD (mm)***	***ARD (%)***	***MAD (mm)***	***ARD (%)***
Proposed	**6.75**	**1.25**	**5.71**	**1.13**
Method in [2]	8.64	1.76	6.03	1.25



Examples of depth estimations

Segmentation & Recognition

Comparison of segmentation method

	FDA dataset		FULL dataset	
Method	F_{sum}***(%)***	F_{min}***(%)***	F_{sum}***(%)***	F_{min}***(%)***
Proposed	**94.36**	**83.90**	**94.10**	**78.18**
Method in [2]	93.69	74.26	-	-
Method in [3]	92.47	73.36	91.83	75.33

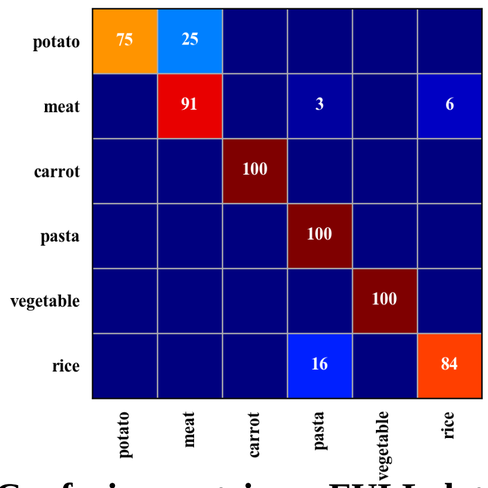
$$NI_{min}(T \to S) = Min_i\left(\frac{Max_j(|S_j \cap T_j|)}{|S_i|}\right) \qquad NI_{sum}(T \to S) = \frac{\sum_i Max_j(|S_i \cap T_j|)}{\sum_i |S_i|}$$

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

Quantitative results using Average Precision (AP)

Dataset	***mAP (%)***	AP_{50} ***(%)***	AP_{75} ***(%)***
FDA	69.4	90.4	85.7
FH	63.2	83.7	79.6
FULL	64.7	85.1	79.1

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

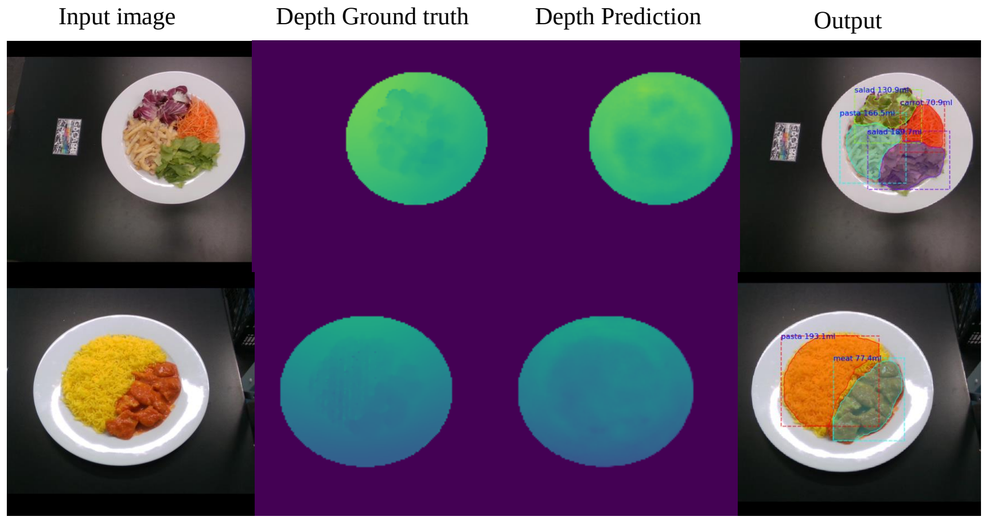


Confusion matrix on FULL dataset

Volume Estimation

Comparison of volume estimation

	Food item's average percentage error			
Method	***FDA (%)***	***FH (%)***	***FULL (%)***	***Process time (s)***
Proposed	**17.5**	**19.1**	**19.0**	**<0.2**
3D Reconstruction [2]	22.6	36.1	33.1	5.5



Some sample results

References

[1] K. He, et al., Mask R-CNN. *In arXiv:1703.06870*, 2017.

[2] D. Allegra, et al., A Multimedia Database for Automatic Meal Assessment Systems. *Madima Workshop*, 2017.

[3] J. Dehais, et al., Dish Detection and Segmentation for Dietary Assessment on Smartphones. *Madima Workshop*, 2015.

MADiMa 2018