Introduction

- The rapid rise in obesity rates and malnutrition in modern society has created a health crisis.
- To aid in improving the evaluation of meal nutrition, I propose a deep-learning network for food recognition and localization using the YOLOv2 algorithm.
- The YOLO algorithm takes as input an image and returns class and bounding box predictions as output.
- To improve the accuracy of the network, the backend detection model was pretrained on the ETHZFOOD101 dataset.

Data and Features

- ETHZFOOD101 contains 101 food classes and has ~1000 images per class, but does not contain bounding box data, thus making it appropriate to use as the training data for the backend object recognition model for DeepMeal. [1]
- The UECFOOD100 and UECFOOD256 are similar but separate data-sets that contain 100 and 256 food classes respectively with ~100 images per class, and localization information for each sample, thus making it appropriate to use as the training/validation data for the full YOLOv2 algorithm for DeepMeal. [3][2]

![Dataset examples. The UEC data-sets include images and bounding box localizations. The ETHZ data-set includes just images organized by class.](image)

Methods and Models

- The YOLO algorithm splits the input image into a S x S grid and for each grid generates bounding boxes, confidence, and class probabilities. [4]
- The YOLOv2 algorithm uses an object recognition network as a backend model. A modified GoogLeNet (Inception V1) model was used for the baseline DeepMeal network. Instead of using inception modules, the extraction model relies on 3x3 and 9x9 convolutional layers. [4]
- Baseline uses a backend model pretrained on ImageNet data, while the proposed approach uses a backend model trained on ETHZFOOD101 data.

![Model for YOLO algorithm. Divides image into grid and for each grid cell predicts bounding boxes, confidence, and class probabilities.](image)

![Loss function for YOLO algorithm. Adds localization, confidence, and classification loss together.](image)

Results and Discussion

- Table: Results for baseline and ETHZ backend-trained model.

![Architecture for the Extraction CNN model.](image)

- Although the ETHZ trained backend model fails to classify well due to undertraining, it has some early success.
- Baseline gives good results because its backend was fully trained on ImageNet data.

Conclusion and Future Work

- The performance improvements of DeepMeal could allow for a more accurate and high-speed platform for tracking of nutritional habits.
- Future work includes fully training the backend ETHZ model, expanding the data in the ETHZFOOD101 dataset for training the backend model and trying different CNN network architectures such as DenseNets and ResNets.

References