**Image processing model to estimate nutritional values in raw and cooked vegetables**

Tan Jo Yen   
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

Dr. Sivakumar  
line 2: *dept. name of organization*  
*(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

**Abstract**

The presence of various foods with high calories and low nutritional values in the modern era has led to increasing worldwide chronic diseases patients. People are advised to constantly monitor their dietary behaviours in their daily lives to practice healthy diets. Clinical dietary assessment methods and mobile calorie-tracking apps are mainly utilized to record daily food consumptions yet are less user-friendly that resists people to consistently practise for the long term, while image-based assessment models are developed to recognise the foods and estimate the food nutritional values directly from food images. However, the existing models do not consider the nature of foods which their nutritional values change after being cooked. As such, an InceptionV3-based multioutput convolutional neural network (CNN) model, *VegeNet* that estimates the nutritional values of cooked and uncooked vegetables is developed. This deep learning model successfully classifies the food images at 97% accuracy and estimate the nutritional values at 15.30% mean relative error, which contributes as an added visual-based food assessment solution to help users save time from having to weigh and entering the information for calorie tracking, meanwhile avoid underreporting problems.

**Keywords**

Image-based dietary assessment model, CNN, InceptionV3, cooked and uncooked vegetables.

1. **Introduction**

Food is one of the main parts in a human’s daily life; the nutrient contents in foods are the energy source for living things to sustain lives and perform various activities. Nowadays, people are easily exposed to heavily processed foods that are more flavourful and addicting yet are less healthy. The overconsumption of these foods leads to chronic diseases like obesity, diabetes, and cardiovascular diseases that have become younger and more common. According to the World Health Organization (WHO) (2021), worldwide obesity has been increasing by three times since the year of 1975; more than 650 million adults were obese in year 2016 while 38 million children at ages under 5 were overweight or obese in year 2019. As such, WHO advises the community to practise healthy diets to overcome the mentioned problems (World Health Organization, 2020).

To better understand a person’s dietary behaviour for health improvement, dietary assessment methods like real-time recording, 24-hour dietary recall, dietary history, and food frequency questionnaire (Shim et al., 2014) are utilized by dietitians to assist obese patients whereas calorie tracking apps have been developed for smart phone users to record the foods eaten on meal-basis, which allow users to calculate and track the food consumption patterns, meanwhile provide real-time personalized feedbacks and suggest diet goals to help the users to better achieve their dieting objectives. Despite being more systematic and data-centric, the calorie-tracking apps are found to be uneasy to use especially for outside-eaters and are subject to underreporting problems, as they mostly require the users to enter the foods eaten manually by the correct names and volumes, then choose the best options from the suggestion lists. The users either need to find the closest items or weigh each ingredient to obtain the most accurate calorie and nutritional values, they may also tend to select the suggestions with lower calories, which leads to underreporting and underestimation of nutrient intakes (Kipnis et al., 2001; Pettitt et al., 2016).

To alleviate the mentioned problems, this paper introduces an image-based dietary assessment model that estimates nutritional values of cooked and uncooked vegetables directly from input image, through deep-learning technique to perform food type and cooking method classifications, as well as food weight estimation, whereby the predicted outputs are taken to estimate the nutritional values based on official food data. This project focuses on image analysis to differentiate between raw and cooked vegetables, because the weights, water contents, and other nutritional values of natural products change when cooked (Filipic, 2016). This model adds value to the existing visual-based food assessment technology by considering the nature of vegetables in both raw and cooked conditions, which can be deployed in mobile applications for more efficient and accurate diet monitoring. Besides, the more convenient dietary assessment method encourages people to consistently track the nutrients consumed, helping them to better understand own dieting behaviours and choose healthier foods or cooking methods to maintain body health.

The following parts of this paper is organized as follows: section 2 reviews the related works on image-based food recognition and calorie estimation, section 3 explains the proposed solution, section 4 describes the datasets prepared, section 5 discusses the implementation steps and evaluation results, chapter 6 concludes this paper with project limitation and suggestions for future works.

1. **Related work**

Image-based dietary assessment models have been developed by data scientists and researchers using machine learning and deep learning approaches, of which the model functions can be categorized into 2 types: (1) food recognition, and (2) food nutritional values estimation.

***2.1 Food recognition***

Prior to the popular adoption of CNN deep learning algorithm on image classification tasks, Support Vector Machine (SVM) algorithm had been a promising machine learning approach utilized in the image data analysis field to perform image classification and predictions. Villalobos et al. (2012) used SVM as the classifier to identify food and fruits for calorie estimations with the colour, size, and shape properties from the images while Pouladzadeh et al. (2014) improved the SVM model by adding Gabor filter for texture segmentation to be taken as input feature for more accurate food recognition. However, SVM performs well only when the dataset is small (Bhargava et al., 2020) whereas the deep-learning approach – CNN outstands SVM for its ability to select features automatically (Hasan et al., 2019) and analyse large image dataset as observed in (Kagaya, Aizawa and Ogawa, 2014). Since then, CNN has been famously used for image classification tasks.

A new CNN model for image recognition requires very large dataset for training which consumes very long time (up to months), and the model is highly subject to overfitting issue (Hassannejad et al., 2016; Szegedy et al., 2015). These problems can be solved by using pretrained networks like *AlexNet*, *ZFNet*, *VGGNet*, *GoogleNet* and *ResNet* (Howard et al., 2017) as the feature learning layer, followed by classification layers customized for the specific image recognition tasks., or fine-tuning the pretrained model for both feature extraction and classification tasks. Yanai and Kawano (2015) showed that a fine-tuned pretrained DCNN model along with SVM classifier performed better than conventional SVM model and DCNN model without fine-tuning, Ciocca, Napoletano and Schettini (2018) fine-tuned *ResNet-50* for food classification on Food-475 dataset, and Hassannejad et al. (2016) fine-tuned classification layers of the pretrained network introduced by Google (Szegedy et al., 2016) on ETH Food-101, UEC-Food 100, and UEC-Food 256 image datasets. Moreover, Phiphiphatphaisit and Surinta (2020) fine-tuned the MobileNet architecture originally developed by Howard et al. (2017) by replacing the average pooling layer and fully connected layer with global average pooling and batch normalization layers to resolve overfitting problem in MobileNet.

To overcome the limitations of the above-mentioned models that only recognizes single food in an image, Pouladzadeh and Shirmohammadi (2017) utilized Selective Search algorithm in Map Reduce to segment food regions by pixels and classify the pixels into various groups before being processed with CNN. Ege and Yanai, (2018b) developed a CNN model which could estimate bounding boxes around each food in the images and estimate calories for respective foods, by adding pseudo-bounding boxes to the calorie annotated image dataset while training the model. Distinctively, Mezgec et al. (2019) developed a model that performs pixel-level classification rather than real-life image classification with fully convolutional networks (FCN) to identify multiple foods in an image, yet this model was trained with fake food image dataset which is less representative of real food images.

Rather than interclass food recognition, Zhang, Lu and Zhang (2016) developed a CNN model that recognizes intraclass dishes through the ingredients and cooking methods classifications. Martinel, Foresti and Micheloni (2018) also proposed WISER – a deep neural network with 2 branches that can learn more features from vertical layers of food images to recognize variances within food classes, while Zheng, Zou and Wang (2018) integrated superpixel-based mid-level feature extraction approach and DCNN to extract image features, then adopted SVM as the classifier for intraclass food recognition.

***2.2 Food nutritional values estimation***

Using food recognition task as the first point, the ability to estimate food nutritional values improves the practicability of visual-based dietary assessment model. Ege and Yanai, (2017, 2018a) trained CNN models using food image datasets obtained from school lunch blog and online recipes which the calorie values are provided. This method assumes that the serving sizes of each labelled photos are for one person, which is seen to be less generalized and is subject to over or underestimation as the portion of foods eaten differ among individuals. As such, food volume estimation is critical for accurate food nutritional value estimation, but this can be very challenging. Some researchers used reference objects as calibrations to estimate food volumes. Pouladzadeh, Shirmohammadi and Al-Maghrabi (2014) took the user’s thumb for calibration to compute food volumes based on area and depth information collected. Okamoto and Yanai (2016) utilized preregistered known-size reference object while Ege, Shimoda and Yanai (2019) used estimated rice image pixels as reference, to obtain the food size by comparing the number of pixels between foods and reference object. Hassannejad et al. (2017) used a 5 x 5 PVC checkerboard card as the size reference to estimate food volume. However, this method requires reference object to be always available whenever a food image needs to be captured and analysed.

To avoid using reference objects, segmentation techniques are adopted to segment food regions. Okamoto and Yanai (2016) used k-means segmentation technique and GrabCut algorithms to segment “food”, “dish”, and “background”, whereas Dehais, Anthimopoulos and Mougiakakou (2016) developed a CNN model for food region segmentation to produce border maps with pixel values ranging from 0 to 1, whereby values closer to 1 were known as pixels closer to the border. Sari et al. (2020) demonstrated image thresholding and K-means++ clustering as segmentation techniques to extract the food regions from the images. Dissimilarly, Ruenin, Bootkrajang and Chawachat (2020) directly estimates the food weights and the nutritional values, with the help of divided food tray provided by hospitals. The divided food tray allows the algorithm to easily identify the food area and hence estimate food weights.

***2.3 Summary***

CNN has been widely adopted to perform food classification and calorie estimations on a range of food image datasets and have been achieving promising results. However, it is observed that most papers did not consider the cooked status of the foods identified, whereby this gap is filled up in this paper with the recognitions of raw and cooked vegetables.

1. **Proposed solution**

To identify raw and cooked vegetables and estimate their nutritional values., this paper proposes a multioutput CNN model, *VegeNet* which classifies the food type, the cooking methods, and estimates the weights of the foods in the images. Later, the predicted values are taken to calculate the estimated nutritional values based on the U.S. Department of Agriculture (USDA) nutrition information data.

***3.1 Overview of CNN***

CNN is a type of deep neural network that takes inputs from the input layer, followed by a convolutional layer as the first hidden layer in the network for high level images feature extractions, and other hidden layers like activation function layers, pooling layers, and fully connected layers to produce predictions (see **Figure 1**). These hidden layers are involved in CNN to serve different purposes:

1. Convolutional layer – computes the convolution of an original image (input signal) through a set of kernel filters to generate an activation map

2. Activation function – defines how the output of neuron is ranged, the rectified linear unit (ReLU) has been commonly used in deep neural networks for its power to reduce model training time

3. Pooling layer – also known as “sub-sampling layer” which reduces the dimensions of output feature map from previous layers to avoid duplicated computations

4. Fully connected layer – to merge the flattened feature maps received from convolutional and pooling layers into a “fixed-size category” for classification.

***3.2 Proposed network***



Figure 1: Architecture of convolutional neural network (CNN) (Prabhu, 2018).

Like most papers, transfer learning technique is adopted in this paper to develop a multioutput CNN model, to avoid overfitting issue when training small image dataset with 7754 images. *VegeNet* utilizes pretrained *InceptionV3* as the feature learning layer, whereby the outputs of *InceptionV3* are taken as the inputs for the 3 branches to predict the food type, cook-method, and weight from food images (see **Figure 2**). Generally, each of the 3 branches is made up of one flattening layer, one dense layer (128 neurons, RELU activation), one batch normalization layer, and a dense layer as the final output layer. However, dropout layers are included in the food-type and weight branch before the output dense layer, but not in the cook-method branch, because cook-method classification is the most challenging task among the 3. This approach is taken to regularize the model meanwhile maintain high prediction accuracies. In the output dense layers, ‘softmax’ activation is used for food type and cook-method classifications while ‘linear’ activation is used for the weight estimation.

Diagram

Description automatically generated

Figure 2: Network architecture of *VegeNet*.

These predicted food types, cook-methods, and weights values produced by CNN model are concatenated and linked to the nutrition information dataset to estimate the calories and macronutrients of vegetables. As the nutrition information data in the USDA dataset are in 100g of food weights, the nutritional values are calculated using 100g as the base weight and multiply by the estimated food weight as illustrated in theequation below.

where

*Ni* = estimated nutritional value for the captured vegetable (calories and macronutrients)

*Wi* = estimated weight of vegetables in the analysed image

*NUSDA* = nutritional value of vegetable in 100g base weight found in USDA dataset

Diagram

Description automatically generated

Figure 3: Process flow of developed GUI.

Graphical user interface, application, Word

Description automatically generated

Figure 4: GUI demonstration.

***3.3 Graphical user interface (GUI)***

To demonstrate the model practicability, a user interface is developed which allows users to select the images to be predicted and prints out the estimated nutritional values by loading the CNN model as the predictor and the nutrition information database for nutritional values retrieval. **Figure 3** shows the process flow of the nutritional values estimating application and **Figure 4** demonstrates the functions of this GUI.

1. **Dataset**

In this project, the food image dataset and nutrition information dataset are required. The food image dataset is made up of 8 vegetable categories and 1 non-vegetable category. The vegetable images are collected primarily by preparing 8 types of vegetables (broccoli, red and green cabbages, carrot, cauliflower, corn, cucumber, lettuce) which are cut in multiple shapes and cooked in 5 methods (uncooked, steamed, roasted without oil, roasted with oil, stir-fried with oil). The vegetables are prepared in different appearances, weighed, and captured from multiple angles and heights, which are then pre-processed through a 3-fold Canny edge-detection and image cropping technique. Besides, to help the model recognize the non-vegetable foods, a non-vegetable category is added to the image dataset by sample images from the hospital food images utilized in (Ruenin et al., 2020). The weights of all foods are recorded as an attribute for model training. On top of that, data augmentation is done using left and right rotations, left and right shearing, image flipping, random brightness, and image skewing techniques with the python Augmentor library to randomly generate 1000 images. An image dataset comprising 7754 images is prepared after the image acquisition and pre-processing steps. Meanwhile, a duplicated set of food images are prepared by adding unsharp-mask filtering to study the effectiveness of added image filtering on the model classification accuracy.

Table 1: Sample nutrition information data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Food\_ type** | **Cook\_ method** | **Cal (kcal)** | **Prot**  **(g)** | **Carb (g)** | **Fat (g)** |
| Lettuce | uncooked | 15 | 1.36 | 2.87 | 0.15 |
| Lettuce | Cooked\_  nofat | 15.47 | 1.4 | 2.96 | 0.15 |
| Lettuce | Cooked\_  wfat | 43 | 1.3 | 3.33 | 3.15 |
| Lettuce | Roasted\_  nofat | 21.43 | 1.94 | 4.1 | 0.21 |
| Lettuce | Roasted\_  wfat | 36.82 | 1.89 | 3.98 | 2.91 |
| Broccoli | uncooked | 34 | 2.82 | 6.64 | 0.37 |
| Cauliflwr | Uncooked | 25 | 1.92 | 4.97 | 0.28 |
| Cabbage\_  green | Cooked\_  nofat | 26 | 1.33 | 6.02 | 0.1 |
| Cabbage\_  red | uncooked | 31 | 1.43 | 7.37 | 0.16 |
| Corn | uncooked | 86 | 3.27 | 18.7 | 1.35 |
| Carrot | uncooked | 41 | 0.93 | 9.58 | 0.24 |
| Cucumber | uncooked | 15 | 0.65 | 3.63 | 0.11 |
| Non\_vege | other | 0 | 0 | 0 | 0 |

The nutrition information dataset is obtained from the USDA data portal (<https://fdc.nal.usda.gov/>), of which the nutrition information of the vegetables involved are extracted from the FNDDS dataset. Data pre-processing like removing unwanted attributes and adding cook-method classes through average ratio calculations are done to match the food classes to the image dataset. The final USDA nutrition information dataset consists of 5 attributes (food\_type, cook\_method, energy (kcal), protein (g), carbohydrate (g), and fat(g)) and 54 instances that represent different food categories and cook-methods (**Table 1**).

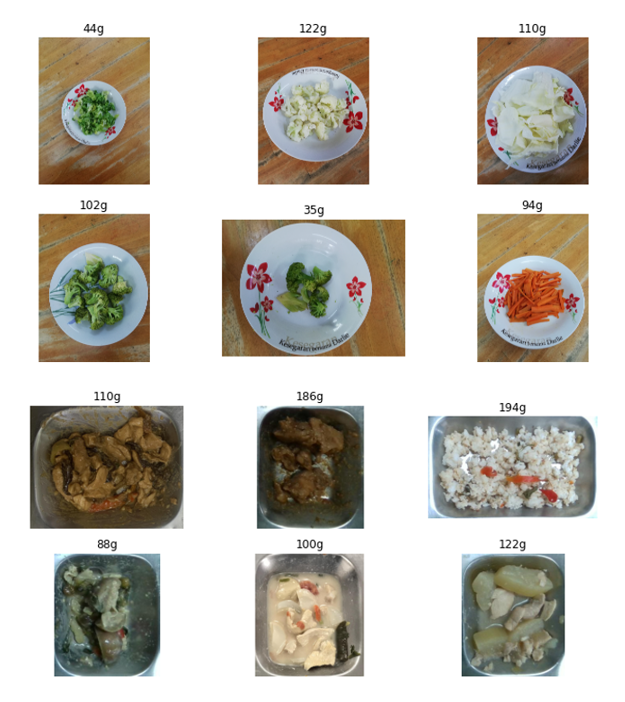


Figure 5: Sample of food images collected.

1. **Experiment & Result**

***5.1 Implementation steps***

The images are resized to 299x299 in image data generator and fed into the *InceptionV3* multioutput CNN model for model training and predictions. Other than the original images, the duplicated dataset with added unsharp-mask filtering is also used for model training and testing. The model training and validation processes are done in Keras environment. The parameter settings for model training are provided in **Table 2**, and the Keras EarlyStopping function is called to stop the model training process when the models do not improve for 10 epochs.

Table 2: Parameters to train *VegeNet*

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Optimizer | Adam (learning rate = 0.0001, decay = 0.000001) |
| Learning rate | 0.0001 |
| Epochs | 100 |
| Batch size, validation batch size | 64 or 32 |
| Image data generator batch size | 32 or 16 |
| Training loss | Food type: categorical\_crossentropy  Cook-method: categorical\_crossentropy  Weight: mean squared error (MSE) |
| Loss weights | Food type: 1  Cook-method: 1.5  Weight: 4 |
| Metrics | Food type: accuracy  Cook-method: accuracy  Weight: mean absolute error (MAE) |

Table 3: List of model evaluation metrics

|  |  |
| --- | --- |
| Precision |  |
| Recall |  |
| Accuracy |  |
| Abs error |  |
| Rel error |  |
| MAE |  |
| MRE |  |

***5.2 Model assessment***

Since the proposed model involves 3 predictions – food type classification, cooking method classification, and weight estimation, whereby these outputs are taken to estimate the food nutritional values, the model validation and testing stage involves both evaluation metrics for classification and estimation performances, which are listed in **Table 3**.

The model assessment results are shown in **Figure 6** and **Tables 4** and **5**.

Table 4: Model classification performance

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **VegeNet** | **VegeNet \_USM filtered** |
| Food type classifi-cation | Accuracy | 1.0 | 1.0 |
| Precision | 1.0 | 1.0 |
| Recall | 1.0 | 1.0 |
| Cook-method classifi-cation | Accuracy | 0.97 | 0.97 |
| Precision | 0.97 | 0.97 |
| Recall | 0.97 | 0.97 |

The model is trained with 6754 images and tested with 1000 images of the image dataset. The evaluation results show that *VegeNet* achieves 100% and 97% accuracies for food type and cook-method classifications. Besides, the model estimates the food weights and food nutritional values at around 16% of mean relative errors. In general, *VegeNet* that is trained and tested with unsharp-masked filtered images fails to outstand the model trained with unfiltered images, indicating that the additional filter does not help to improve the model performance.

Table 5: Weight and nutritional values estimation performance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | **VegeNet** | **VegeNet \_USM filtered** |
| Weight | MAE (g) | | 13.86 | 15.80 |
| MRE (%) | | 15.60 | 18.11 |
| Nutri values | MAE (g) | Cal | 4.73 | 5.61 |
| Carb | 0.80 | 0.95 |
| Prot | 0.19 | 0.22 |
| Fat | 0.21 | 0.25 |
| MRE (%) | | 15.30 | 17.76 |

Chart, line chart

Description automatically generated

Figure 6: Model training output

While comparing with the related works, **Table 6** shows that *VegeNet* achieves outstanding results in both classification and nutritional values estimation tasks, of which the classification performance is the highest at 97% and nutritional values estimation is the second most accurate at only 15.30% mean relative errors. Nevertheless, all models are trained and tested with similar but different datasets, these values are only the general guidelines to assess the model performance in this domain.

The high classification result is expected due to small category size, as the images are classified by food type and cook-method separately via 2 branches in the CNN model, whereby the 32 categories are split into 9 food types and 6 cook-methods, thus reduces the classification complexity. As shown in **Table 4**, the model can 100% accurately predict the food types and 97% accurately predict the cook-methods. The high achievement in food type classification substantially advances the model performance in classification tasks over the total 32 categories. On the other hand, the nutritional values estimation is done by weight prediction through the CNN model using ‘linear’ activation layer in the weight branch. **Table** 5 shows that *VegeNet* achieves 13.86g of mean absolute error and 15.60% mean relative error in weight prediction. As the nutritional values are calculated using the USDA nutrition information at 100g base weight, low error in weight prediction leads to low error in nutritional values estimation. It is also expected that the outstanding performance of *VegeNet* is mainly due to the adoption of *InceptionV3* pretrained network as the base convolutional neural network which substantially enhance the feature learning ability of the model which performs better than *VGG16* and *ResNet-50* utilized by Ege and Yanai (2017) and Ruenin et al. (2020).

Table 6: Model performances of proposed model and related works\*

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Image dataset** | **Classification**  **Accuracy / Average precision** | **Nutritional value estimation**  **Mean relative error** |
| VegeNet  (This project) | Self-captured vegetable images + Hospital food images  (32 categories, 7,754 images) | 97.00 % | 15.30% |
| Faster R-CNN + ResNet50  (Ruenin et al., 2020) | Hospital food images  (40 categories, 20,084 images) | 73.35 % | 16.97% |
| Multi-task CNN  (Ege & Yanai, 2018a) | Recipe dataset  (36 categories, 7,361 images) | 81.20%; 54.4% | 27.40% |
| Faster R-CNN  (Ege & Yanai, 2017) | School lunch dataset  (21 categories, 4,877 images) | 90.7% | 21.40% |
| SVM + Gabor filter  (Pouladzadeh et al., 2014) | Self-captured images  (15 categories; 3,000 images) | 90.41% | 14.00% |

\*The values of other models are directly taken from the papers; each model is trained with different dataset.

1. **Conclusion**

***6.1 Summary***

In the recent decade, conventional machine learning and deep CNN models have been developed to recognise the foods through image classification and estimate the food calories through diverse techniques, yet these models mainly focus on prediction for cooked foods and did not consider the nutritional values difference between cooked and uncooked foods. This paper proposes an InceptionV3-based multioutput CNN model which estimates the nutritional values of cooked and uncooked vegetables to address this issue.

***6.2 Contribution***

To the best of our knowledge, this is the first visual-based dietary assessment model that distinguishes the nutrient contents in cooked and uncooked vegetables. This work is essential as vegetable is one the main nutrient sources for healthy diet, yet the nutritional values differ when vegetable is cooked in disparate ways. Other than calorie values, the model also estimates the values of the macronutrients in vegetables – carbohydrate, protein, and fat. This is because calorie value is just the total energy consumed yet the macronutrient is more important information to maintain balance diet. To estimate food nutritional values, this proposed model estimates the food portion by weight predictions, hence an image dataset that consists of 32 food categories and 7754 vegetable and non-vegetable images with known weights are prepared for model training and testing.

***6.3 Limitation and future work suggestion***

Despite high model accuracy obtained in this project, several limitations are seen in this model:

1. The limitation of datasets. To estimate the nutritional values via food weight prediction, the image dataset with known food weights is required, yet this type of dataset is not easily obtained, especially the vegetable images. Thus, the primarily collected images are very limited to the vegetable types and cooking methods, as well as the container, location, background, capturing device, and the surrounding lightings which make the model prediction less generalized. Besides, model is subject to overfitting due to small image dataset utilized. Moreover, the nutrition information extracted from the USDA dataset is ambiguous, for example “cooked with added oil” does not clearly specify the amount of oil added also lead to the high deviation between predicted values and actual values. To overcome this limitation, the collection of large image dataset with known food weights as well as the correct food nutritional values shall be done as a future work, so that the models developed are more applicable in real life diet control.

2. It is known that in real life, the images inputted for predictions are captured under disparate environments like lightings and would affect the image quality hence influence the model prediction accuracy. Other than unsharp mask filtering, there are many conventional image pre-processing techniques like variational denoising methods and transform techniques for image denoising. Moreover, deep CNN-based image denoising models are found to advance the image pre-processing performance. The addition of deep learning model in image pre-processing step shall be done in the future work.

3. This model only considers the energy and macronutrient contents of vegetables. The accurate visual-based dietary assessment application that involves more types of foods and micronutrient information shall be built with more professional nutrition knowledge and research to consider the differences in nutritional values of cooked and uncooked foods.

1. **Reference**

Bhargava, D., Vyas, S., & Bansal, A. (2020). 7 - Comparative analysis of classification techniques for brain magnetic resonance imaging images. In D. Koundal & S. B. T.- A. in C. T. for B. I. A. Gupta (Eds.), *Advances in Computational Techniques for Biomedical Image Analysis: Methods and Applications* (pp. 133–144). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-820024-7.00007-4>

Ciocca, G., Napoletano, P., & Schettini, R. (2018). CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding, 176–177*, 70–77. <https://doi.org/https://doi.org/10.1016/j.cviu.2018.09.001>

Dehais, J., Anthimopoulos, M., & Mougiakakou, S. (2016). Food Image Segmentation for Dietary Assessment. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management,* 23–28. <https://doi.org/10.1145/2986035.2986047>

Ege, T., Shimoda, W., & Yanai, K. (2019). A New Large-Scale Food Image Segmentation Dataset and Its Application to Food Calorie Estimation Based on Grains of Rice. *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management*, 82–87. <https://doi.org/10.1145/3347448.3357162>

Ege, T., & Yanai, K. (2017). Estimating Food Calories for Multiple-Dish Food Photos. *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, 646–651. <https://doi.org/10.1109/ACPR.2017.145>

Ege, T., & Yanai, K. (2018a). Image-Based Food Calorie Estimation Using Recipe Information. *IEICE Transactions on Information and Systems*, *E101.D*(5), 1333– 1341. <https://doi.org/10.1587/transinf.2017MVP0027>

Ege, T., & Yanai, K. (2018b). Multi-Task Learning of Dish Detection and Calorie Estimation. *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*, 53–58. <https://doi.org/10.1145/3230519.3230594>

Filipic, M. (2016). *Steam, roast vegetables to retain nutrients*. Chow Line - The Ohio State University. <https://u.osu.edu/chowline/2016/10/28/steam-roast-vegetables-to-retain-nutrients>

Hasan, H., Shafri, H. Z. M., & Habshi, M. (2019). A Comparison Between Support Vector Machine (SVM) and Convolutional Neural Network (CNN) Models For Hyperspectral Image Classification. *IOP Conference Series: Earth and Environmental Science*, 357, 12035. [https://doi.org/10.1088/1755- 1315/357/1/012035](https://doi.org/10.1088/1755-%201315/357/1/012035)

Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., & Cagnoni, S. (2016). Food Image Recognition Using Very Deep Convolutional Networks. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, 41–49. <https://doi.org/10.1145/2986035.2986042>

Hassannejad, H., Matrella, G., Ciampolini, P., Munari, I. D., Mordonini, M., & Cagnoni, S. (2017). A New Approach to Image-Based Estimation of Food Volume. *Algorithms*, *10*(2: 66), 1–14. <https://doi.org/10.3390/a10020066>

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *ArXiv:1704.04861*.

Kagaya, H., Aizawa, K., & Ogawa, M. (2014). Food Detection and Recognition Using Convolutional Neural Network. *Proceedings of the 22nd ACM International Conference on Multimedia*, 1085–1088. <https://doi.org/10.1145/2647868.2654970>

Kipnis, V., Midthune, D., Freedman, L. S., Bingham, S., Schatzkin, A., Subar, A., & Carroll, R. J. (2001). Empirical Evidence of Correlated Biases in Dietary Assessment Instruments and Its Implications. *American Journal of Epidemiology*, *153*(4), 394–403. <https://doi.org/10.1093/aje/153.4.394>

Martinel, N., Foresti, G. L., & Micheloni, C. (2018). Wide-Slice Residual Networks for Food Recognition. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 567–576. <https://doi.org/10.1109/WACV.2018.00068>

Mezgec, S., Eftimov, T., Bucher, T., & Koroušić Seljak, B. (2019). Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. *Public Health Nutrition*, *22*(7), 1193–1202. https://doi.org/DOI: 10.1017/S1368980018000708

Okamoto, K., & Yanai, K. (2016). An Automatic Calorie Estimation System of Food Images on a Smartphone. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, 63–70. <https://doi.org/10.1145/2986035.2986040>

Pettitt, C., Liu, J., Kwasnicki, R. M., Yang, G.-Z., Preston, T., & Frost, G. (2016). A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate. *British Journal of Nutrition*, *115*(1), 160–167. https://doi.org/DOI: 10.1017/S0007114515004262

Phiphiphatphaisit, S., & Surinta, O. (2020). Food Image Classification with Improved MobileNet Architecture and Data Augmentation. *Proceedings of the 2020 The 3rd International Conference on Information Science and System*, 51–56. <https://doi.org/10.1145/3388176.3388179>

Pouladzadeh, P., & Shirmohammadi, S. (2017). Mobile Multi-Food Recognition Using Deep Learning. *ACM Trans. Multimedia Comput. Commun. Appl.*, *13*(3s), 36:1- 36:21. <https://doi.org/10.1145/3063592>

Pouladzadeh, P., Shirmohammadi, S., & Al-Maghrabi, R. (2014). Measuring Calorie and Nutrition From Food Image. *IEEE Transactions on Instrumentation and Measurement*, *63*(8), 1947–1956. <https://doi.org/10.1109/TIM.2014.2303533>

Prabhu (2018). *Understanding of Convolutional Neural Network (CNN) — Deep Learning*. Medium. <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>

Ruenin, P., Bootkrajang, J., & Chawachat, J. (2020). A System to Estimate the Amount and Calories of Food That Elderly People in the Hospital Consume. *Proceedings of the 11th International Conference on Advances in Information Technology*. <https://doi.org/10.1145/3406601.3406613>

Sari, Y. A., Saputra, V. W., Agustina, A., Wani, Y. A., & Bihanda, Y. G. (2020). Comparison of Image Thresholding and Clustering Segmentation Methods for Understanding Nutritional Content of Food Images. *Proceedings of the 5th International Conference on Sustainable Information Engineering and Technology*, 124–129. <https://doi.org/10.1145/3427423.3427441>

Shim, J.-S., Oh, K., & Kim, H. C. (2014). Dietary assessment methods in epidemiologic studies. *Epidemiol Health*, *36*, e2014009. <https://doi.org/10.4178/epih/e2014009>

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>

Villalobos, G., Almaghrabi, R., Pouladzadeh, P., & Shirmohammadi, S. (2012). An image procesing approach for calorie intake measurement. *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, 1–5. <https://doi.org/10.1109/MeMeA.2012.6226636>

World Health Organization. (2020). *Healthy diet*. WHO Newsroom. <https://www.who.int/news-room/fact-sheets/detail/healthy-diet>

World Health Organization. (2021). *Obesity and overweight*. WHO Newsroom. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

Yanai, K., & Kawano, Y. (2015). Food image recognition using deep convolutional network with pre-training and fine-tuning. *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 1–6. <https://doi.org/10.1109/ICMEW.2015.7169816>

Zhang, X., Lu, Y., & Zhang, S. (2016). Multi-Task Learning for Food Identification and Analysis with Deep Convolutional Neural Networks. *Journal of Computer Science and Technology*, *31*(3), 489–500. [https://doi.org/http://dx.doi.org/10.1007/s11390- 016-1642-6](https://doi.org/http://dx.doi.org/10.1007/s11390-%20016-1642-6)

Zheng, J., Zou, L., & Wang, Z. J. (2018). Mid-level deep Food Part mining for food image recognition. *IET Computer Vision*, *12*(3), 298–304. <https://doi.org/https://doi.org/10.1049/iet-cvi.2016.0335>