International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

AI for Healthy Diet

Ashok Veda[1], Shikha Kathuria[2]

[1]CEO, Rubixe, Bengaluru, India
ashok[at]rubixe. com

[2]AI Intern, Rubixe, Bengaluru, India
shikha[at]rubixe. com

Abstract: ***Food object recognition and nutritional value determination-an application of Deep Learning. "The doctor of the future will give no medicine, but will instruct his patients in care of the human frame, in diet, and in the cause and prevention of disease." – Thomas Edison. Today, doctors and nutritionists take a holistic approach to healthcare where diet, exercise, lifestyle, pre-existing conditions and a multitude of other things are considered in a bid to avoid and treat diseases. This article proposes a solution to track one's diet, calories and nutrients before consumption so that users can make an educated choice of what and how much they want to consume in their quest for healthy life.***

Keywords: AI, Healthy Diet, Calorie detection, Deep Learning, Visual Recognition, Micronutrient detection, Macronutrient detection, Mask RCNN, Image Detection, Food Volume, Food Area, Food Weight, Food Height, Food Density

1.Introduction

AI is rapidly revolutionizing healthcare by accelerating medical research, detecting malignant tumors and predicting chances of acquiring a disease. AI is helping deliver affordable healthcare to disadvantaged sections of society by reimaging how healthcare is delivered. As it's said, prevention is better than cure-avoiding a disease even before it happens is economical and goes a long way to achieve overall wellbeing of society. This necessitates tools which can seamlessly monitor diets and track calories and nutrients intake.

A traditional approach to monitoring diets is maintaining a food journal, where one notes down the food items every time they eat. While motivated individuals can sustain this habit for a while, many might give it up soon as it's inherently tedious and inefficient. Also, analyzing data to obtain actionable insights is cumbersome. The mechanism also assumes that individual is proficient in ingredients that constitute the food items.

Studies have shown that accurate dietary assessment is very important for assessing the effectiveness of weight loss interventions. Identifying food and its contents, just by clicking an image, would be win-win proposition and go a long way to combat obesity epidemic. With such applications in mind, we seek to exploit the advances in machine learning and deep learning to train models that identify food items from digital photos.

With the growing reliance on smart and self-tracking devices, there has been an increasing demand for innovative technologies to ease the tracking of food consumption behavior of individuals. One of the most prevalent examples is the usage of smart-phones and fitness bands to monitor physical activity (number of steps, distance travelled, exercise, etc.) Following this trend, we propose to exploit the convenience of smart-phones and capabilities of deep-learning based food-image recognition system, to build a smart food logging system.

1.1 Confirm the Opportunity

Individuals today are spoilt for choices of applications when it comes to tracking fitness, diet and medical needs. In the diet tracking space, a multitude of apps ask users to key-in individual food items that they are consuming. Some advanced apps detect food through phone camera but aren't able to estimate the weight of the food items and hence the corresponding calories and nutrients. Our app simplifies the entire process of food logging by identifying food items and estimating calories directly from the food image clicked. It determines the area and weight by using an easily available reference object. It goes a step further to provide information on all micro and macronutrients instead of just calories.

Impact of our nutrition app

The WHO states that good nutritional status is an important determinant of quality of life. Leading a healthy lifestyle has become a rage today and has given rise to a multi-billion dollar healthcare industry. The tech-savvy young population is very conscious of what they eat, especially due to sedentary lifestyle. Their life mantra is "If it's not posted on their social media account, it hasn't happened". You have to be very cautious of what you post on your social media as this comes with certain stigma about what you eat and your physical appearance. This necessitates tools that give you easy access to nutritional information.

Also in the older generation, many of the diseases from which they suffer are associated with the natural aging process (such as loss of appetite and a decline in intake) but are compounded by dietary factors. This in turn leads to a decreased immune function and magnifies risks of chronic diseases.

Given the importance of understanding dietary patterns and the challenges with current food-intake methods, we propose an intelligent nutritional assessment system that can help make healthier food choices. Our app would monitor dietary-patterns and provide nutritional analysis through the application of computer vision and machine learning.

2. Characterize the Problem and Profile the Data

Problem definition?

1. Poor eating habits include under-or over-eating, not having enough of the healthy foods we need each day, or consuming too many types of food and drinks, which are low in fiber or high in fat, salt and/or sugar.
2. Unhealthy diet is one of the major causes of obesity and malnutrition.
3. These unhealthy eating habits can affect our nutrient intake, including, protein, carbohydrates, essential fatty acids, vitamins and minerals as well as fiber and fluid.

The benefits of digital innovation in the food industry influence not only those needing nutritious food, but also fitness enthusiasts and medical professionals.

It is not only under-nutrition that threatens global health, over-nutrition is also a primary concern, especially in developed countries. In fact, obesity is the number one cause of preventable death in the USA, where nearly 40 percent of the population suffers from the condition, according to the Center for Disease Control and Prevention (CDC).

Our digital innovation will help people to make healthier choices by addressing their food and nutrition problem and by educating them about what they are consuming.

Who is facing this problem?

1. Women, infants, children and adolescents are at particular risk of obesity and malnutrition.
2. In 2014, about 13% of the world's adult population (11% of men and 15% of women) were obese.

Where is the problem occurring?

Every country in the world is affected by one or more forms of obesity and malnutrition. Combating malnutrition in all its forms is one of the greatest global health challenges.

Why is this problem occurring?

1. Teenagers are notorious for bad nutritional choices. Peer influence, easy access to fast food, addictive behaviours, and lack of knowledge about proper nutrition are the major causes of obesity.
2. Working adults often miss meals and choose the quick pick-me-up provided by nutritionally deficient snacks and junk foods, which leads to obesity.
3. Limited access to the nutrition rich diet.
4. Lack of Resources to gain the knowledge about the Nutrition, and many more.

2.1 Profiling the data: Constructing Food Image Dataset

We started off by building a robust dataset of several food images. We first defined 8 "super categories" representing generic food items including burger, rice, pizza, samosa to name a few. For each of these super categories, we visually inspected and classified 700+ food images into one of the 8 food categories. We also added a category for non-food items, for which we randomly sampled about 10, 000 different images from MS-COCO dataset. The prediction model would be trained to make predictions on these visual food categories (or classes).

In each of these 8 categories, (from Indian, Italian, American cuisines) we ensured that we had at least 60 images per category. The images were collected by crawling the web using Google, Bing, Instagram, Flickr and other social media tools. These were manually vetted to confirm whether each image crawled indeed belonged to the food class being searched for. We then split the dataset-into train, validation, and test data-for training and evaluating the model.

Figure 1: Food Image Dataset

Further, we used **ImgLab** as an image annotation tool to annotate images and extract xml files for each of the images.

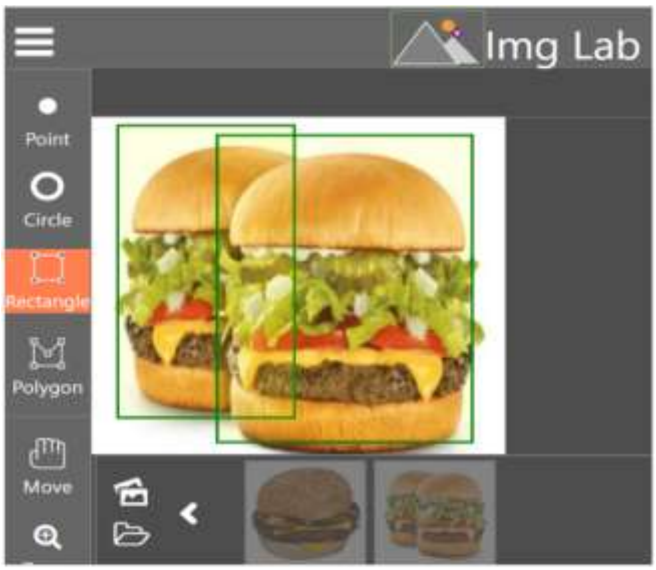


Figure 2: Image Annotation

3.Architect and Deploy the Solution

The project was divided into 5 phases, each representing a distinct set of activities. Refer figure below for details of each phase. We will cover major phases in detail in the following sections.

Figure 3: Process Flow

3.1 Model Building

3.1.1 Object detection system

Object detection is the process of finding and classifying objects in an image. Below is high level flow for object detection

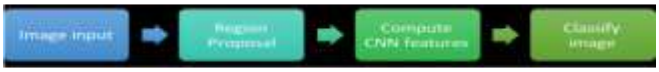


Figure 4: Object Detection Flow

R-CNN (Region Based Convolutional Neural Networks) is a deep learning approach that combines rectangular region proposals with convolutional neural network features. It is a two-stage detection algorithm.

1. The first stage identifies a subset of regions in an image that might contain an object. The R-CNN detector generates region proposals using an algorithm such as Edge Boxes. The proposal regions are cropped out of the image and resized. Further, the region proposal bounding boxes are refined by a support vector machine (SVM) that is trained using CNN features.
2. The second stage classifies the object in each cropped and resized region.

Algorithm evolution

Problems with classic RCNN

RCNN was the state-of-the-art method when the reports were first published in 2014. However, it had many drawbacks. For instance, it used disk space to store intermediate results, consuming I/O operations on disk. Since the initial proposal has a large number of region proposals and each region needs to be propagated in CNN, it is computationally costly. The whole recognition process was dependent on the initial region proposals, which were not learnt but estimated using traditional image processing methods.

Figure 2: RCNN object detection steps

Move towards Fast-RCNN

The authors of RCNN came up with a new idea to optimize the whole training process into a single-stage task. This improved version of RCNN was reported as Fast-RCNN. The training process is a single stage-task where all the convolutional layers get updated. The main bottleneck of disk I/O was completely removed from the pipeline, leading to much faster training and testing as compared to the previous RCNN model. Experiments revealed that this model was 10 times faster to train as compared to RCNN. Furthermore, it was also discovered that computing region proposals also took less than a second.

Algorithm used-Mask R-CNN

Mask R-CNN presents a conceptually simple, flexible, and general framework for object instance segmentation. The approach efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance. The method Mask R-CNN, extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. It is simple to train and adds only a small overhead to Faster R-CNN. Moreover, Mask R-CNN is easy to generalize to other tasks, e. g. allowing us to estimate human thumb in the same framework. It has shown top results in all tracks of the COCO suite of challenges, including instance segmentation, bounding-box object detection, and person keypoint detection.

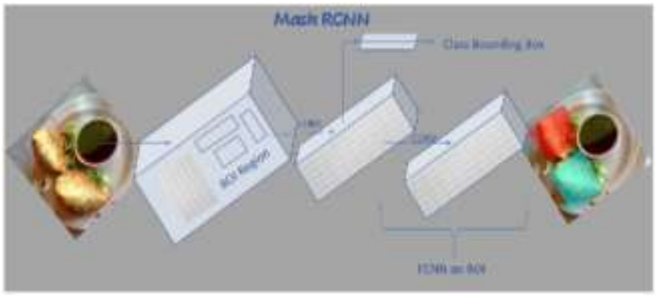


Figure 6: Mask RCNN object detection steps

3.1.2 Implementation steps

Importing necessary libraries

In this application we have used TensorFlow (1. x) with GPU and Keras.

Image resizing

To support training multiple images per batch we resized all images to the same size. For example, 1024x1024px on MS COCO Dataset.

Bounding boxes

Standard available datasets provide bounding boxes and masks. To support training on multiple datasets we opted to ignore the bounding boxes that come with the dataset and generate them on the fly instead. We picked the smallest box that encapsulates all the pixels of the mask as the bounding box. This simplifies the implementation and also makes it easy to apply image augmentations that would otherwise be harder to apply to bounding boxes, such as image rotation.

Learning rate

Here we are using a learning rate of 0.001. Previously used high learning rates caused the weights to explode, especially when using a smaller batch size.

3.2 Model Training

With the growing success of deep learning for visual recognition, we used a deep convolution network as the model to recognize food images.

Transferable features for image recognition

In recent years, it was observed that using networks pre-trained on the MS-COCO dataset and transferring those to other datasets provided a significant boost to performance than training new models from scratch. This was due to the ability of deep convolution networks to learn general features applicable to several computer vision tasks. Following this success, we use pre-trained MS-COCO models and fine-tuned them on our food dataset. During the course of this project, we tried several models and updated our application as new state-of-the-art models got invented. In this paper, we focus our attention on more recent models, ResNet101, and report their performance. During training, we followed the standard approaches for data augmentation such as rotation, random crop, random contrast, etc.

4.Calories and Nutrients Determination

To come up with a most accurate approach for calories and nutrients calculation, we analysed many past studies and research papers and augmented this with our testing. Here we will present various approaches evaluated and the final solution used for calories and nutrients calculation.

4.1 Mask Area Calculation

We will first look at how to calculate mask area of any object as it's the building block for calories and nutrients calculation.
Extracting the masked pixels for each category and taking thumb (considering standard thumb dimensions as 5cm * 2.3 cm) as a reference, we calculated area (in square inch) of a food in an image. We used below parameters for calculation.

TA	Standard thumb area
TI	Area of Masked thumb from the image with reference to standard thumb area
MT	Pixels of masked thumb from the image
MF	Area of masked food
RFA	Real Food Area

Calculations steps

Area of a standard thumb (TA) = 11.5 cm² = 1.782504 inch²

TI = Area of MT / TA (in inches)
RFA = Area of MF / TI

4.2 Calories & Nutrients Calculation Approaches

Once we have mask area calculated, we will now look at two broad methods to calculate calories and nutrients in food items.

4.2.1 Weight method

In this option, the approach is to **estimate** the height and density of the food categories and store in a local database. We get masked surface area of the food from the Mask RCNN algorithm. Then, we calculate the volume of the food by multiplying the masked surface area with the height. Finally, the weight is derived by multiplying the volume with the density.

Pros
1. Better estimation of calories and nutrients as height and density of food items doesn't vary much.
2. Following the global standards for many API integrations, we can map calories, macro and micro nutrients per gram for the food categories. This will also help to scale faster with newer food categories.
3. Easy maintenance.

Cons
1. Manual maintenance of average heights and densities of the food categories.

4.2.2 Mask surface area method

We get mask surface area (in sq. Inch) by the mask RCNN algorithm. The approach is to create a local database for calories, macro and micro nutrients by estimating calories and all nutrients at **per square inch level** for each food category. For example-One 12 inch plate of pizza has 1200 calories and plate size is 113 square inch, then we calculate calorie per square inch by 1200/113. The same exercise needs to be done to calculate per square inch micro and macro nutrients. For example-One 12 inch plate of pizza has 10 grams of proteins, then protein per square inch would be 10gm/113.

Pros
1. No approximation of any dimension.

Cons
1. We will have to estimate per square inch calories, all macro and micro nutrients for all the food classes by calculating calories and nutrients per area basis.
2. Difficult to calculate per square inch calories and all other nutrients for non-flat foods such as Samosa and Burger (We will have to calculate calories and nutrients considering one full plate of food item).

3. Not a global standard to convert everything to per square inch as all the available API's provide values in per gram.

After analysing both the above options, we decided to go ahead with Option 1 –**Weight Method** as it follows global standards and is easy to scale.

4.3 Calories and Nutrients Calculation

For Weight method (as discussed in Section 4.2.1 above), we need to estimate the height, volume, density and weight of the food items. Once the weight is determined, we can use standard APIs to get calories and nutrients information.

4.3.1 Height estimation table for our food categories

This table is populated with research from google and by manually taking the heights of many food items and then taking an average.

estimated_food_height_inch={'cheeseburger_475cal': 2, 'fries_350cal': 1.3, 'pizza': .7, 'rice': 1, 'samosa': 1.5, 'sandwich': 1.3, 'apples': 2.5, 'coconuts': 4.5}

4.3.2 Density estimation table for our food categories

This table is populated with research from Google and from the website www.aqua-calc.com

food_density_gm_per_cm3={'cheeseburger_475cal': 0.69, 'fries_350cal': 0.54, 'pizza': 0.65, 'rice': .72, 'samosa': .61, 'sandwich': .45, 'apples': 0.46, 'coconuts': .35}

4.3.3 Volume and weight calculations

1. Volume (inch cube) = Surface Area (inch square) * **estimated_food_height_inch.**

2. Volume (cm cube) = Volume (inch cube) * 16.3871.

3. Weight (grams) = Volume (cm cube) * **food_density_gm_per_cm3.**

5. Model Evaluation

Here we present the results of our application “AI for Healthy Diet” in the training phase, where we evaluate the performance on test data of the original dataset.

5.1 Evaluation of Object Detection Model

While several models have been explored in this paper, we present the performance of ResNet-101 (101-layer ResNet) and Mask-RCNN trained with ResNeXt-101.

Figure 7: Model tuning parameters

Among the models, the best train accuracy of 80.86% and test accuracy of 85.01% was obtained by a combination of Mask-RCNN with ResNeXt-101. ResNet-50 performance wasn’t in the acceptable range and was discarded (possibly due to convergence challenges). We also looked at the inference speed of the models and we were able to make predictions at the rate of 80-100 images per second for the 101-layer models, or 1 image in 0.01 seconds. This is fairly fast, however the end-to-end inference results depends on the round-trip latency of transferring the image and result to-and-fro from server. The models occupy close to 250MB for the 101 layer models. Since this model is going to be stored only on the server, the model does not cause a memory constraint.

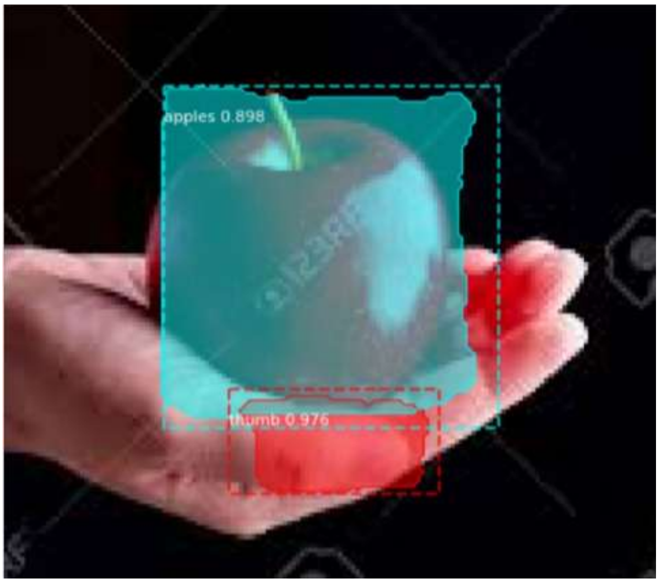


Figure 8: Object Detection from the model

5.2 Evaluation of Weight Calculation

We checked the weight accuracy by conducting a simple experiment-

1. Measured the weight of an apple on weighing scale as 179 grams

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Figure 9: Manual weight measurement

2. Ran below image through our application and model estimated the apple weight to be 210 grams.

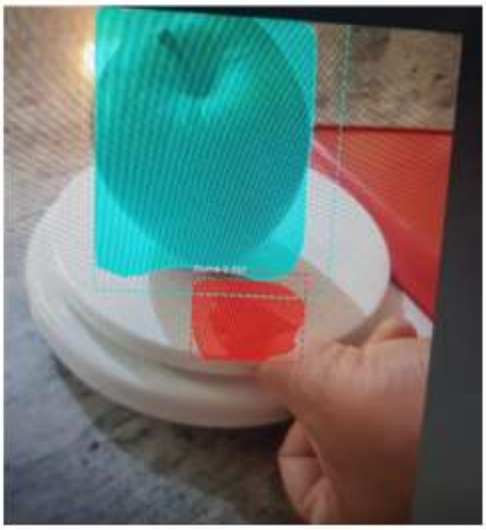
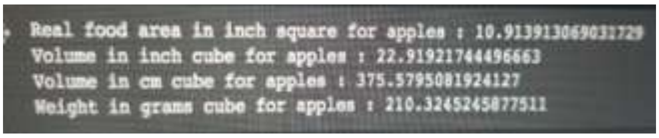


Figure 10: Apple and thumb detected by model



```
Real food area in inch square for apples : 10.913913069031729
Volume in inch cube for apples : 22.91921744496663
Volume in cm cube for apples : 375.5795081924127
Weight in grams cube for apples : 210.3245245877511
```

Figure 11: Weight calculated by the model

3. Accuracy of the weight calculation can be calculated as-weight of the apple on weighing scale / weight of apple predicted from the model = 179/210 * 100 = 85%

Results look promising with model accuracy around **85%**

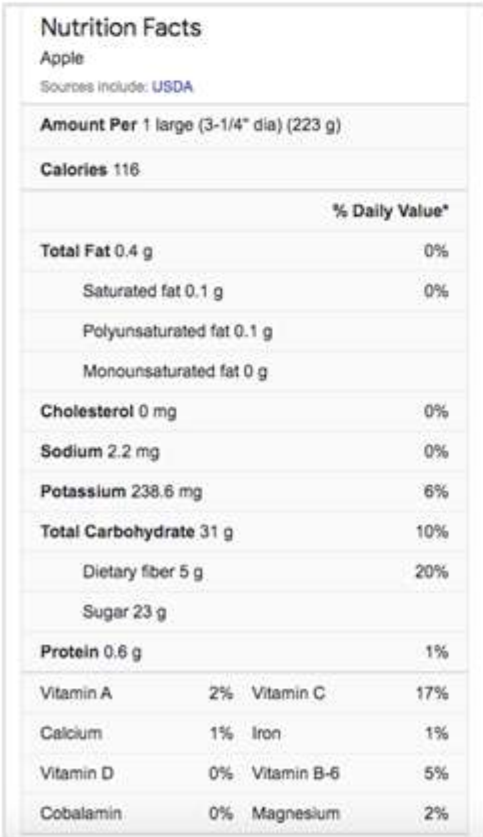
5.3 Evaluation of API Interface

An Application Programming Interface, or API, is a set of commands, functions, protocols and objects that allow for programmatic access to a system. We have integrated **Edamam** API in our application using Flask framework. The API takes weight as an input from our model and returns calories and all the macro and micro nutrients in a JSON file format.

We compared nutritional results from our application to the information available in public domain. We found that the results from API interface are similar to the actual nutrient values.

Web results indicate 116 calories in 223 grams of apple which translates to 110 calories in 212 grams of apple. The API integrated in our model has also estimated 110 calories for the same weight of the apple.

Calculation on other nutrients also yields similar results and hence this API is shortlisted to be used with the application.



Nutrition Facts
Apple
Sources include: USDA

Amount Per 1 large (3-1/4" dia) (223 g)			
Calories 116			
			% Daily Value*
Total Fat 0.4 g			0%
Saturated fat 0.1 g			0%
Polyunsaturated fat 0.1 g			
Monounsaturated fat 0 g			
Cholesterol 0 mg			0%
Sodium 2.2 mg			0%
Potassium 238.6 mg			6%
Total Carbohydrate 31 g			10%
Dietary fiber 5 g			20%
Sugar 23 g			
Protein 0.6 g			1%
Vitamin A	2%	Vitamin C	17%
Calcium	1%	Iron	1%
Vitamin D	0%	Vitamin B-6	5%
Cobalamin	0%	Magnesium	2%

Figure 13: Nutrient Values from Internet

5.4 Factors that can affect the model accuracy and performance

1. Poor quality of images

Poor quality images like upside-down, images in very dark lightning and half-eaten foods would result in a poor performance of the model.

2. Images without the reference object thumb

The images without the reference object thumb can affect the model accuracy and performance depending on the number of pixels, image scaling and distance from which the image has been clicked.

3. Class imbalance in the dataset

A specific problem we faced in the our dataset was of imbalanced data, i. e. there is slight variance in the number of instances per class. This imbalance would induce a bias in the model favouring a better performance to those classes with more data even if other food items are relatively easy to classify.

6.Business Value

In this white paper, we propose a mobile application for food calories, micronutrients and macronutrients estimation from a single image of the food without user entering any additional information. There are already many different apps and products available to do the same, but most of these apps assume that the user will provide the name of the food item or the ingredients, as well as the size of the food items they are consuming.

Some apps on the other hand needs user to input 2 set of images (Top & Side) and a fixed size reference object like coin to calculate the size of the food. These apps then run it against a database of food items to calculate the calories present in them. In our application, we propose to alleviate the user from the burden of entering the above information and from providing 2 food images. This is particularly beneficial when such information is difficult to obtain and time consuming to enter.

In our app, we are using thumb as a reference object which is easy to include in the food images rather than the fixed size reference object coin. Also since this application estimates surface area and weight of the food objects, the results are much more accurate as compared to the other applications, which give generalized numbers depending on the food items.

We hope our simple and effective approach will serve as a solid baseline and help ease future research in AI for heath domain.

7.Conclusion

Accurate estimation of dietary intake is important for assessing the effectiveness of weight loss interventions. The key innovation in this paper is the simple method to calculate approximate weight of the food and estimate calories and nutrients. Our proposed algorithms are based on Mask R-Convolutional Neural Network (CNN). In the future, we plan to improve performance of the algorithms and enhance the accuracy of current measurements of dietary caloric and nutrients intake.

8.Scale up the POC

1. Tune and optimize

Since our dataset is relatively small, we can expand our dataset on more images of the existing food categories. More real-world images for existing food categories would be added with different viewing angles, distance, noise and obstructions to make the model perform better. We would also work towards removing imbalance of the dataset to make predictions even better.

2. Additional food categories

Our plan is to train the model on more food categories to expand the reach of our application. We plan to make it more generic by including all common food categories.

9.Other Applications

1. Nutrition Tracking

For healthy lifestyle maintenance, there are many useful nutrients that should be present in a person's dietary intake. The application can log the user's full day's nutrition, which would be derived from the images that user would provide during the day. Nutrition tracking feature would then suggest important nutrients that were lacking from user's daily diet. User can then take required actions to incorporate the nutrients in the diet.

2. Diet Recommendations

The application can be extended by adding a recommendation section to propose alternate diet options depending on if the user has chosen Weight Maintenance or Weight Reduction category. The Application would provide other similar low calorie dishes or substitute of calorie rich food items in the existing dish.

3. Food Weighing with Smart Phone

In Supermarkets, before adding food items to their cart, users can weigh the items they wish to purchase and they can increase or decrease the quantity. This will save their time when standing in queues at the billing counter and will help to make purchasing process efficient.

4. Prediction of Diseases

The daily data collected from our application can be used to derive health trends alerting the user about the calorie and nutrition excess or deficiencies and the possible threats of other major ailments over coming years like diabetes, blood pressure or heart attack towards which the user may be slowly moving.

5. Integration with Smart Gadgets

We can integrate the dietary intake data with the age, weight, height and the calories burnt via physical activity as measured on fitness applications. This can help to derive more meaningful and accurate insights into one's fitness needs.

Acknowledgment

I have completed this project under the supervision of Mr. Ashok Veda (CEO), Rubixe, and Mr. Deepak D. (Program Manager), Rubixe. I would like to express my gratitude to both the mentors for allowing me to work as a Data Science Intern at Rubixe and encouraging me to write this paper. I would also like to thank Garima Jain, Madhuri Bodle, Syed Fazal, Ankur Sharma (AI Interns), Rubixe for the assistance.

References

[1] https://aaltodoc.aalto.fi/bitstream/handle/123456789/32513/master_Koirala_Janaki_2018.pdf?sequence=1&isAllowed=y
[2] https://arxiv.org/pdf/1909.11946v1.pdf
[3] https://arxiv.org/pdf/1703.06870.pdf
[4] https://www.researchgate. net/post/Can_anyone_help_me_in_conversion_of_Area_from_Sqpixels_to_Sqmm
[5] https://www.traditionaloven. com/tutorials/surface-area/convert-mm-diam-circle-to-sq-mm-square-millimeter.html