Food Image Recognition and Calorie Prediction

Narayana Darapaneni
Director – AIML
Great Learning/Northwestern
University Illinois, USA
darapaneni@gmail.com

Subhav Kataria
Student – AIML
Great Learning
Delhi, India
kataria217@gmail.com

Anwesh Reddy Paduri
Data Scientist - AIML
Great Learning
Delhi, India
anwesh@greatlearning.in

Vishal Singh
Student – AIML
Great Learning
Delhi, India
vishalsngh1992@gmail.com

Nayana Bansal
Student – AIML
Great Learning
Delhi, India
naybansa09@gmail.com

Yasharth Singh Tarkar
Student – AIML
Great Learning
Delhi, India
yasharthsinghrajput@gmail.com

Abhijeet Kharade
Mentor – AIML
Great Learning
Delhi, India
Kharadeabhi@gmail.com

Abstract -- **With the increasing number of health issues reported due to obesity and overeating, people have become cautious about their diet intake to prevent themselves from the diseases such as hypertension, diabetes, and other heart related problem which are caused due to obesity. As per the data shared by WHO, at least 2.8 million people are dying each year because of being overweight or obese. The important part of any healthy diet plan is its calories intake. Hence, we propose a deep learning-based technique to calculate the calories of the food items present in the image captured by the user. We used a layer-based approach to predict calorie in the food item which include Image Acquisition, Food item classification, Surface area detection and calorie prediction.**

Keywords – Mask R-CNN, ROI (Region of Interest), IOU (Intersection Over Union), Resnet 50

I. INTRODUCTION

It is very important in today's time that people should be aware of what they are consuming and what will be its impact on the body. So, a system that can help individuals to maintain their calories intake is very important. Most of the world's population live in countries where overweight and obesity kills more people than any other health disease. The problem here is not about having enough food, it is about the people not knowing what is in their diet. If people could estimate their calorie intake during a day, they can easily decide on the number of calories they want to consume. However, managing calorie intake is a very cumbersome task which involves the people to manually keep a track of food item they have consumed throughout the day and they must determine the calories they have consumed. This process is not only manual but also inaccurate as the calorie estimation not only depends on what you are eating but also depends on how much are having.

With the advancement in the field of image processing techniques, the image recognition models are in demand. Researchers are aggressively deploying image recognition model for various uses such as self-driving cars, cancer detection, video frame analysis etc. Researchers have also shown keen interest in predicting the calories present in the food item with the help of the image. Researchers have used various machine learning and deep learning techniques to perform the task of calories estimation with the help of supplied images.

Manal Chokr and Shady Elbassuoni (2017) [8] proposed a solution in which they used the supervised learning technique to perform the single food classification and its calorie prediction. They used the model which takes an image of the food item as input and provides the calorie as output. They developed a Mathwork Image Process tool which was used to extract features from the image. Extracted features were then compressed and fed to a classifier and regressor to identify the food type and determine the size of the food item respectively. Finally, the output of both classifier and regressor is fed to another regressor which provides the calories as the output.

Parisa Pouladzadeh1, Abdulsalam Yassine1, Shervin Shirmohammadi (2015) [9] developed a deep learning-based model which takes a food image clicked from the mobile camera which is capable of estimating calorie for the mixed portion of the food item as well. The dataset used contained 3000 images clicked under different condition with different camera model and then the clicked image is given as input. They used color segmentation, k-mean clustering, and texture segmentation tools. They employed Cloud SVM and deep neural network to increase the performance of the image identification model. For calorie prediction they used the reference object approach wherein they mandated the presence of the thumb in the image so that their model can use the thumb present in the mage as reference for the size estimation of food item present in the image which helped in calorie estimation.

In earlier researches, the researchers have collected 101 different food images [10] such as Chicken Wings, Tacos,

Bread pudding etc. The dataset has 101 food categories with 101000 images where most of the images contains mixed food items. They used 750 images for model training purpose and 250 images for testing purpose. They used model based on Random Forests to mine discriminative visual components and efficient classification of the images. However, they used the model for the image classification only

Nowadays, there are many android and web-applications available which allows users to track the food intake to help them manage their calorie. However, these applications rely on the users to manually select the food items they had and the size and dimension of the food items they ate to track record of the diet and to further estimated the calorie consumption. This process is not accurate as it is almost impossible for the user to input the accurate food size they ate.

Hence, we propose a solution to this problem. The proposed model can predict the calories user consumed with the help of the image of the food portion user had. The model basically makes use of layered approach wherein each layer performs different tasks. The first layer takes the image as the input from the user and prepares it to be fed to the second layer. The second layer employs the deep learning method known as Mask R-CNN which performs the food identification task along with the bounding box and mask generation. The third layer calculated the surface area covered by the food items which in turn is fed to the last layer. The fourth layer predicts the calories consumed by the user based on the surface area covered by the food item present in the supplied image.

The model provides satisfactory results in estimating the calories intake by the user with the help of image. This model uses the dataset which contains 638 images of food item which belongs to 6 different categories (Apple, Oranges, Omelet, Pizza toast, Banana and Idli). These images are collected from internet, hand-picked from the available dataset such as Food 101[4], UNIMIB2016[5] and UEC FOOD-100[6]. The images are preprocessed as per the model requirement and used to train our model.

In this paper, the model employs Mask R-CNN deep learning technique for the mask and bounding box determination for the image containing different food items which helps the model to identify the food item, estimate the surface area occupied by the food item and to determine the calorie associated with it with the help of defined mathematical formula. The model provides the satisfactory results in determining the calorie in the food portion present in the image.

II. MATERIALS AND METHOS

A. Overview

Our model takes Image of the food items as input and give calories of the food item as output. In this there are number of middle process which are done to achieved to this. Firstly, food item whose calorie need to be predicted is identified in the captured image. After the food item identification its size and volume are determined and at last food calorie is estimated to 128*128 pixels before they are used in the model. Mask R-CNN Algorithm is used for image recognition and calorie prediction is done using approximate proportion approach format.

B. Dataset

Our dataset has six different food items. Dataset is custom generated by selecting food item image from internet and by selecting images from the existing dataset such as Food 101. Food Item are Banana, Pizza Toast, Orange, Idle, Hot Dog, Omelet.

Food Item	Image Count
Banana	113
Pizza Toast	104
Orange	101
Idli	113
Hot Dog	102
Omelet	105

Fig.1.: Dataset details

As images are gathered from different sources it is important to scale them, so they are scaled to 128*128 pixels before they are used in model Also images are manually annotated using Pixel Annotation tool and masks are created for model training purpose.

C. Food-Item Identification

We use instance segmentation to create a pixel wise mask of object in the image. This technique gives us more granular understanding of the food data in the image. Here we use Mask R-CNN for image segmentation which used ROI and IOU to generate bounding boxes, provide labels and mask.

As in figure 2. Bounding box is created across the identified food item and label omelet is assigned to the food item.

Fig.2: Mask output of omelet

Mask R-CNN is a deep neural network aimed to solve instance segmentation in machine learning. It separates different objects in an image or a video and gives out the objects bounding boxes, classes and masks. There are two stages in mask R-CNN first is to generate a proposal about the reason where there might be an object based on the input image second is to predict the class of the object, refines the boundary boxes and generate mask in pixel level. Both stages are connected to the backbone structure which helps in feature extraction. Here we are using Resnet 101 as backbone. Mask R-CNN Model is initialized with preloaded weights.
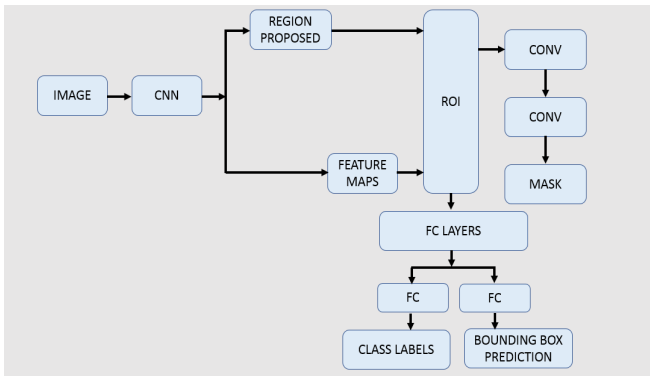
Fig.3: Mask R-CNN Model

To train a Mask R-CNN based image recognition model, we need many images to train deep learning models. Since the dataset we collected is not large enough to train the model, we used transfer learning approach from matterport repository [7].

With the help of **transfer learning**, instead of training a model from scratch, we started with a weights file that is been trained on the COCO dataset. COCO dataset contains lot of images (~120K), so the trained model weights have already learned a lot of the features common in natural images, which really helps.

D. Food Calorie Prediction

As the same food can be taken at different depths to generate different picture sizes, we need a method to calculate calorie or estimate the size of the food in a real-world scenario. After we get the desired food items detected along with their masks, we need the real object sizes, which is not possible through a pin-hole camera image alone. So, we take a referencing approach that references the food-objects to the size of the pre-known object to extract the actual size of the food contained in that specific image.

For food calorie prediction, method used is approximation of proportions. In this approach calorie per mask of the food class is taken as reference for predicting calorie of the input image. A spread sheet has been prepared containing "Class", "Calorie Per Mask", "Minimum Calories", "Maximum calorie".

Food	Calorie_per_mask	Minimum_Calorie	Maximum_Calorie
Banana	0.02739726	72	135
Hot Dog	0.0375	100	290
Omelette	0.01725	50	80
Orange	0.0385	130	320
Pizza Toast	0.051555556	200	242
Idli	0.009428571	25	40

Fig.4: Calorie details used in predicting food calorie

In Proportion Approximation Approach each image has been scaled to the size of (128*128) and segmented area of the food class is calculated and further multiplied with "Calorie per Mask ".

III. EXPERIMENTAL RESULTS

In this section we present the experimental result of all the sections like Food identification which generate bounding boxes, provide labels and mask. Dataset is divided in train and test (Validate) set in the ration of 80:20.

A. Dataset Inspection

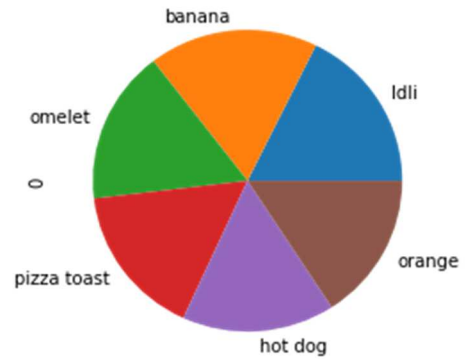Our dataset has six food items named banana, Pizza Toast, omelet, orange and idle.

Fig.5: Food classes in dataset

Images are of different lengths, so we converted them to 128 * 128 and loaded them. Similarly, mask images had also converted them to 128 * 128. Initial dataset was class imbalanced, so we added more images and did data augmentation.

During Visual dataset inspection it was observed classes were initially imbalanced, which were balanced by adding new images. Also, Image size is changed to 128*128. Minimum class image is 101 and maximum class image is 113.
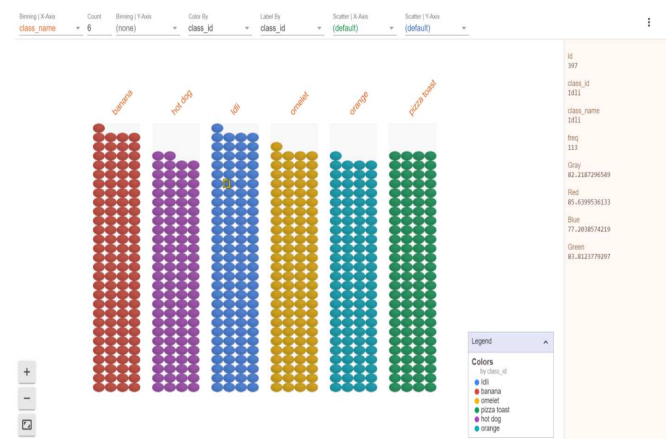

Fig.6: Initial Visual representation of dataset

B. Food Item Identification

Our goal is to identify the type of food item after extracting the features from the image. Here Mask R-CNN is used to generate bounding box, assign labels and mask to the image.

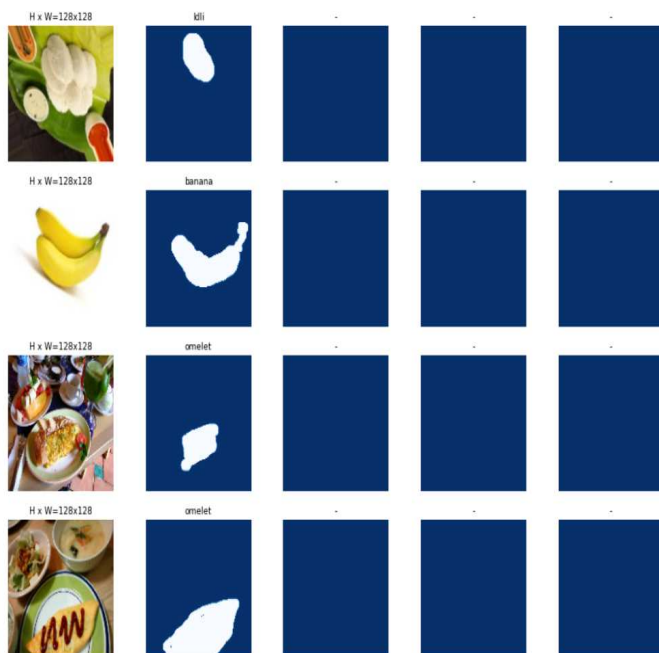Analyzing images and mask after preprocessing, converting to 128 * 128 dimensions.

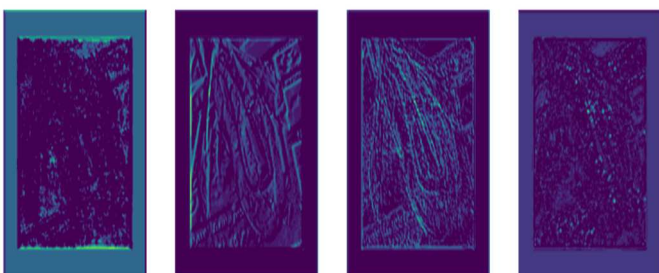Fig.7: Analyzing images and mask after preprocessing

Fig.8: Analyzing activation of hidden layers

C. Calorie Prediction

A referencing approach that references the food-objects to the size of the pre-known object to extract the actual size of the food contained in that specific image. Refer figure 4

Fig.9: Actual Image- Image before Mask R-CNN and calorie prediction

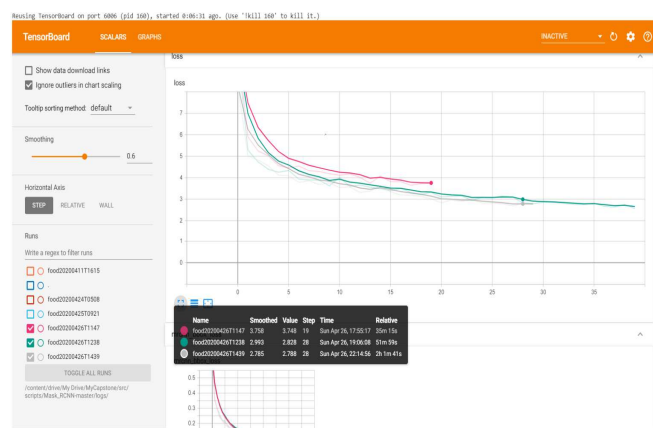Fig.10: shows 242 calories for pizza toast - Image after Mask R-CNN and calorie prediction

D. Visualization Of Model Accuracy & Loss

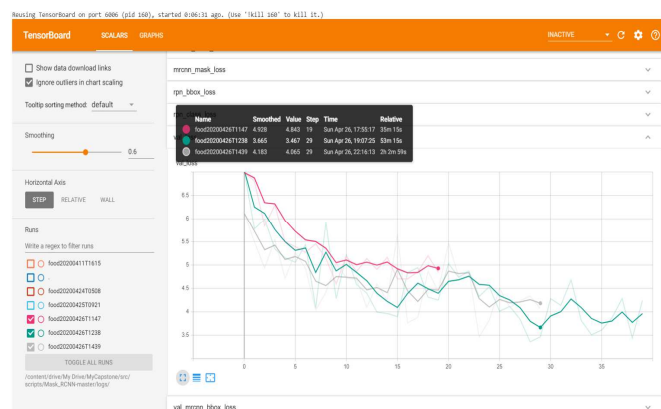Fig.11: Overall loss while training the Train dataset

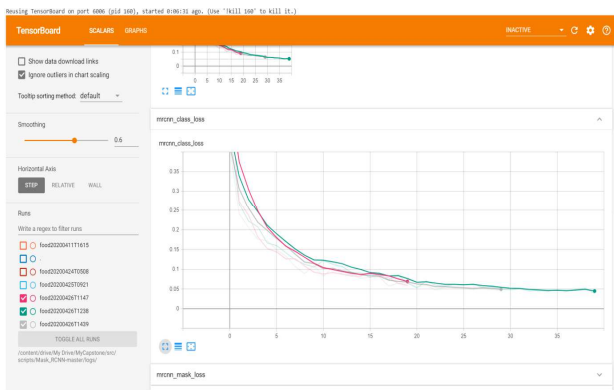Fig.12: Overall loss while Validating/testing the Test dataset

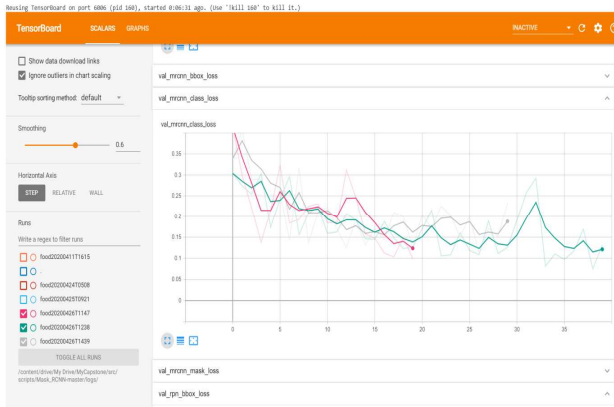Fig.13: Classification loss while training the Train dataset

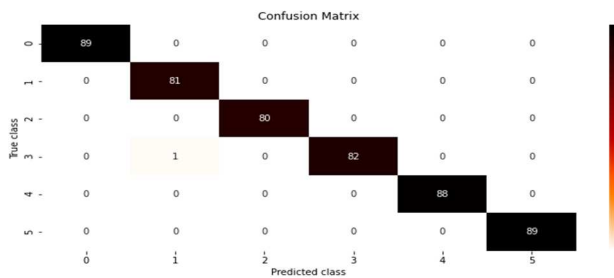Fig.14: Classification loss while Validating/testing the Test dataset

Fig.15: Confusion matrix while Training the Train data (0=BG ,1=Orange, 2=hot dog, 3= omelet ,4=banana, 5=pizza toast , 6=idli)
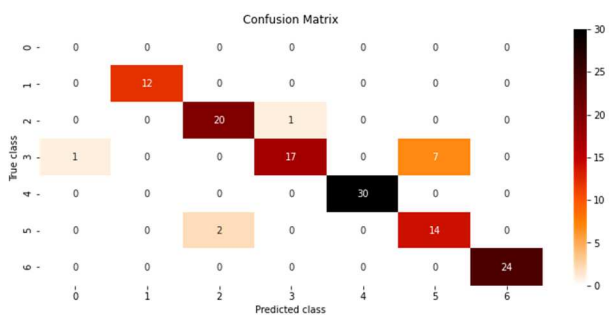

Fig.16: Confusion matrix while Validating/testing the Test data (0=Background ,1=Orange, 2=hot dog, 3= omelet ,4=banana, 5=pizza toast , 6=idli)

IV. CONCLUSION

In this paper, we have used the deep learning-based model to predict the total calories of the food item present in the image. To develop this solution, we used Mask R-CNN technique to create mask and bounding boxes. This in turn helped the model to calculate the surface area occupied by the different food items in the image which further facilitated the model with the ability to satisfactorily predict the calories associated with each food item. Calorie are estimated with the help of mathematical formulas which compares the proportion of the image occupied by each food items and determines the calories associated with it.

In future work, we plan to extend the scope of model by increasing the ability of the model to identify a greater number of food items instead of 6 food items, which we used in our current dataset. The dataset we used contain 638 images of food item with 6 different categories, which will be extended in our next model. Finally, we would like to calculate the calories of the food item based on their volume with the help of 3D images. As the technology is advancing, it would be interesting to work on the model development which can handle 3D image as input and predict the calories of the food item with better results.

V. REFERENCES

[1] World Health Orgamisation
https://www.who.int/features/factfiles/obesity/en/

[2] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2980-2988.

[3] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[4] Bossard L., Guillaumin M., Van Gool L. (2014) Food-101 – Mining Discriminative Components with Random Forests. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham

[5] Ciocca, Gianluigi & Napoletano, Paolo & Schettini, Raimondo. (2016). Food Recognition: A New Dataset, Experiments, and Results. IEEE Journal of Biomedical and Health Informatics. PP. 1-1. 10.1109/JBHI.2016.2636441.

[6] Ge, ZongYuan & Mccool, Chris & Sanderson, Conrad & Corke, Peter. (2015). Modelling local deep convolutional neural network features to improve fine-grained image classification. 4112-4116. 10.1109/ICIP.2015.7351579.

[7] Cao, Yuanzhouhan, et al. "Exploiting Depth From Single Monocular Images for Object Detection and Semantic Segmentation." IEEE Transactions on Image Processing, vol. 26, no. 2, Feb. 2017, pp. 836–46. DOI.org (Crossref), doi:10.1109/TIP.2016.2621673.

[8] Chokr, Manal and Shady Elbassuoni. "Calories Prediction from Food Images." AAAI (2017).

[9] Pouladzadeh, Parisa & Yassine, Abdulsalam & Shirmohammadi, Shervin. (2015). FooDD: Food Detection Dataset for Calorie Measurement Using Food Images. Lecture Notes in Computer Science. 9281. 10.1007/978-3-319-23222-5_54

[10] Food recognition: a new dataset, experiments and results (Gianluigi Ciocca, Paolo Napoletano, Raimondo Schettini) In IEEE Journal of Biomedical and Health Informatics, volume 21, number 3, pp. 588-598, IEEE, 2017

[11] He Y, Xu C, Khanna N, Boushey CJ, Delp EJ. FOOD IMAGE ANALYSIS: SEGMENTATION, IDENTIFICATION AND WEIGHT ESTIMATION. Proc (IEEE Int Conf Multimed Expo). 2013;2013:10.1109/ICME.2013.6607548. doi:10.1109/ICME.2013.6607548

[12] Estrada, F.J. & Jepson, A.D. Int J Comput Vis (2009) 85: 167. https://doi.org/10.1007/s11263-009-0251-z

[13] Liang, Yanchao & Li, Jianhua. (2017). Computer vision-based food calorie estimation: dataset, method, and experiment. https://arxiv.org/pdf/1705.07632.pdf.

[14] P. Pouladzadeh, S. Shirmohammadi and R. Almaghrabi, "Measuring Calorie and Nutrition from Food Image", *IEEE Transactions on Instrumentation & Measurement*, vol. 63, no. 8, pp. 1947-1956, August2014.

[15] *World Health Statistics 2012*, 2012, [online] Available: http://www.who.int/gho/publications/world_health_statistics/2012/en/index.html.

[16] *Obesity Study*, October 2011, [online] Available: http://www.who.int/mediacentre/factsheets/fs311/en/index.html.