

DeepNutriNet: A Dual-Stream Learning Framework for Automated Food Recognition and Caloric Assessment

Prof. Rani Prakash¹, Suman Biradar²

¹Professor, Master of Computer Application VTU's CPGS, Kalaburagi, India

²Student, Master of Computer Application VTU's CPGS, Kalaburagi, India

Abstract -Recognizing food and estimating calories is part of fostering healthier habits and avoiding diet-related diseases such as obesity and diabetes. Calorie tracking methods of the past are still prone to error and inconvenient, justifying the push for automated approaches. We introduce DeepNutriNet, a dual-stream deep learning framework for food recognition and calorie estimation from images. The framework employs a CNN-based detection model that identifies a set of food items that may or may not be overlapping with a classification stream for resolving classification uncertainty caused by visual similarity. We leverage a nutrition knowledge base as a means for converting the end product of recognition into an estimated calorie total at each item and at the meal level. Experiments conducted on publicly available food image datasets with images featuring varying numbers of food items demonstrated high recognition accuracy and favorable calorie estimates. Collectively, these results speak to the ability of this framework to be a viable means of dietary monitoring and precision nutrition intervention.

Key Words: Food recognition, Calorie estimation, Deep learning, CNN, Dual-stream framework, Nutritional analysis, Precision nutrition, Dietary monitoring.

1. INTRODUCTION

Food is essential to human survival and well-being, as it provides the energy and nutrients necessary to support physical performance and mental well-being. However, modern dietary habits that involve eating high calorie and processed foods have contributed to the global epidemic of diseases related to lifestyle factors including obesity, cardiovascular disease, and type-2 diabetes [1]. According to the World Health Organization, poor dietary habits are still one of the main risk factors for premature death worldwide. [2]. As a result, there has been an increase interest in intelligent systems that can track food consumption, analyze nutritional value, and promote healthier eating habits. Traditional methods for tracking calories, pocket food diaries, nutritional databases, or recording portion sizes, are often tedious, inaccurate, and dependent upon user compliance. [3]. These limitations highlight the need for automated, reliable, and non-intrusive solutions. Through advances in computer vision and deep learning, it is possible to accurately identify foods from images which provides the opportunity for dietary monitoring applications [4]. Still, food recognition

is still difficult because of intra-class variability, dish similarities, inconsistent portion sizes, and various presentations/styles across dishes from various cuisines [5]. Deep learning approaches, specifically convolutional neural networks (CNNs) have demonstrated great success in image classification tasks; object detection systems, such as YOLO (You Only Look Once) and Faster R-CNN, have become standard for detecting multiple items in a single photograph [6][7]. This paper describes a dual-stream framework, DeepNutriNet, that combines detection and classification with a knowledge base that supports nutrition information allowing for automation of food recognition/ Calorie estimation with greater accuracy and practicality of use.

2. PROBLEM STATEMENT

Accurate dietary intake monitoring is still a poor area of healthcare and nutrition management. Conventional approaches using manual food diaries and apps often suffer from user-specific bias, underreporting, and incorrect estimates that lead to faulty calorie information. Image-based food recognition has some good ideas, but it faces many practical challenges, including variation across food types, overlapping items, inconsistent portion size, and visual similarities between dishes. While deep learning approaches offer strong performance for image recognition, they can often take either a detection approach or classification approach in isolation and limit their application to spectrum of daily use, including potentially simultaneous detection and calorie estimation.

3. OBJECTIVES

The main goal of this work is to build an intelligent structured framework for automatic food recognition and caloric estimation using deep learning techniques. The framework will use a YOLO-based object detection model to localize and differentiate multiple food items and a CNN-based classification model, including transfer learning, for more accurate recognition. A pizza food-image dataset from on-line Kaggle will be used for training and validation that will transfer well to food with a somewhat wider variety of cuisines and portion sizes. Furthermore, the study's purpose is to build a user-friendly application that allows the user to take a picture of food, detect food in real-time, obtain a caloric estimate,

and get customized recommendations to aid in dietary management.

4. METHODOLOGY USED

1) Data Collection: The data set was assembled from available food image collections on Kaggle that featured many classes of Indian and international foods. Each class contained a large group of images, accounted for different lighting, viewing angle, and backgrounds, to allow for variance. Furthermore, the data set included samples of single-item and multi-item foods to help in the tasks of object detection and classification. Each of these aspects made for suitable data set based on the range of visuals available to train the framework.

2) Data preprocessing: Images were resized to a fixed size for compatibility with the CNN and YOLO architecture. Appropriately, data augmentation tasks, notably through rotation, flipping, scaling, and adjustment of brightness, were utilized to generalize the data. Pixel values were normalized to the scaling [0,1] to improve convergence speed. The data set was split into training, validation, and testing subsets for evaluating effectiveness.

3) Feature Extraction: CNNs were used to automatically extract features from the raw images. Initial layers extracted low-level features of edges and textures, and further layers extracted higher-level features in a food "physical form". Additionally, Transfer learning techniques through pre-trained layers of feature extraction, improved detected feature and improved object recognition across presentation of food. The approach reduced the need for manual engineering to recognise food features and patterns for object detection.

4) Model Selection: A dual-stream framework was utilized with YOLO object detection and CNN classifier for improved recognition. YOLO was selected for its accuracy when detecting multiple objects in real-time; the classification stream incorporated transfer learning with models such as VGG19 and InceptionResNet to improve performance. This framework guaranteed strong performance in multi-class and multi-item situations.

5) Model Training: Detection and classification models were trained separately with labeled food images with supervised learning. Hyperparameters were tuned for optimal performance; specifically, learning rate, batch size, and epochs. The use of GPU provided additional training speed for effectiveness, while early stopping and checkpointing prevented overfitting and saved the best models for replicating the experiments.

6) Model Evaluation: Performance was evaluated using accuracy, precision, recall, F1-score, and mean Average Precision (mAP) for detection. Calorie estimation accuracy was validated against nutrition standards. Additionally, confusion matrices were used to understand strengths and

weaknesses in classification. Results indicated better effectiveness when using the dual-stream framework compared to single-stream models for both better recognition and calorie estimation.

7) Integration with Flask: A simple web-based application was developed using Flask to provide a user-friendly interface for real-time food identification and calorie assessment. Users were able to upload images or stream live input for the food identification service. The backend incorporated trained YOLO and CNN models to return bounding boxes, predicted classes, calorie values, and dietary recommendations. The application included other features including user authentication and result storage to facilitate practical use within a nutritional monitoring context.

5. LITERATURE SURVEY

Article [1] 'Classification of Food Objects Using Deep Convolutional Neural Network Using Transfer Learning' is authored by Dipta Gomes in 2024 focusing on a fine-tuned Xception CNN food classification method achieving 95.20% accuracy. The paper emphasizes how food ingredient classification is improved through transfer learning, opening up possibilities for dietary habit recognition and food waste reduction. Food-101 dataset is used in the research which offers 101 classes of food images along with an extensive exploration of deep learning methods to maximize accuracy. The framework aims to provide a method for identifying food types by using advanced architectures of a convolutional neural network.

Article [2] 'Recognition of food images based on transfer learning and ensemble learning' is authored by Le Bu, Caiping Hu, and Xiuliang Zhang in 2024 discussing food image recognition challenge in complex backgrounds and inter-class similarity by proposing a dual approach of transfer learning and ensemble learning for the recognition process. For the recognition, the researchers used four pre-trained CNN models (VGG19, ResNet50, MobileNet V2, AlexNet) but were fine-tuned for the recognition process based on food datasets. An ensemble model was used in this paper to achieve 96.88% accuracy on the Food-11 dataset and this result was shown to achieve superior output than a single model. The methods also included weight-ensemble and vote-ensemble approaches that increased recognition reliability in food and nutrition monitoring

Article [3]: "Visual nutrition analysis: leveraging segmentation and regression for food nutrient estimation" by Yaping Zhao, Ping Zhu, Yizhang Jiang and Kaijian Xia (2024): This article creates a fully-fledged nutritional content prediction model that combines UNet segmentation and regression techniques for estimating nutrients directly from food images. This research utilized the Nutrition5k dataset which contains intricate

nutritional labels to determine the shortfalls of the volume-based estimated nutrient methods. The model produced an average percentage mean absolute error (PMAE) of 17.06% in estimating the nutrients weight, calories, fats, carbohydrates, and protein. The study adds Squeeze-and-Excitation structures to enhance the food analysis feature representation in visual nutrition analysis. Article [4]: "Fine-grained food image classification and recipe generation using deep learning" by RSA Kareem (2024): This article creates an integrated framework that takes artificial neural networks and natural language processing and combines them into a food image classification system with recipe generation capabilities. The study addresses challenges in fine-grained classification in food recognition and extends functionality by providing contextual recipe information. Within the methodology, the authors used a combination of multi-modal learning approaches to better understand food beyond classification. The work provides practical applications in dietary monitoring and culinary recommendation systems using advanced deep learning architectures.

Article[5] 'Machine Learning Based Approach on Food Recognition and Calorie Estimation' by Z Shen in 2020: This research presented a wide-ranging machine learning mechanism to obtain automatic food image classification and associated calorie estimation while incorporating certain practical aspects of deployment. The objective of the study was to address some of the aspects of dietary monitoring in the real-world by using some degree of food recognition and nutrition monitoring. The research included applications of multiple machine learning methods to provide more consistent identification of food products and predicted energy content of the food group. The method shows significant advances that can be added to automated dietary assessment monitoring systems as part of a health monitoring project.

Article [6] 'Improved food image recognition by leveraging deep learning and computer vision techniques' by A Karabay in 2024: This study focused on improving food recognition accuracy by applying multi-faceted computer vision and deep learning methods relative to food computing applications. The study addressed aspects of quality control in food computing applications by demonstrating specific CNN bases recognition systems. This study provided adequate demonstration of deep learning methods that can potentially prevent poor quality food products from being sold on the market through automated quality interventions. This study also included aspects of smartphone and social media application to interface with technology applications that can be utilized in real world food monitoring systems.

Article [7] is titled "Computer vision-based food calorie estimation: dataset, method, and experiment" authored by Yanchao Liang and Jianhua Li in 2019. In this investigation, the authors describe a new food image dataset containing food items with detailed volume and

mass information along with typical deep learning methods aimed towards complete estimates of calorie counts. This research tackles important challenges to automated dietary assessment using computer vision based methods. The approach also details additional methods for food detection that use integration with nutritional databases for accurate estimates of calories in food images. An experimental validation in this paper demonstrates effectiveness of the described methods in real-world scenarios with practical consideration for monitoring calorie intake for healthy living.

Article [8] is titled "Nutritional Content Detection Using Vision Transformers - An Intelligent Approach" authored by the research team in 2024. In this contribution to the literature, these researchers present completely new approach to farming data types in vision transformers on photographs of food to assess nutritional content. In a significant advancement to dietary monitoring technology, this is another step away from standard CNN models. Researchers adopt this mode by using self-attention and global dependency modeling in their study that overcome limitations in using traditional CNNs. The proposed process incorporates multimodal fusion methods of visual food object analysis and text from actual recipes for complete nutritional analysis of foods. Results in this study document superior predictability of both macronutrients and micronutrient tasks through transformer-based architecture modeling.

6. SYSTEM DESIGN

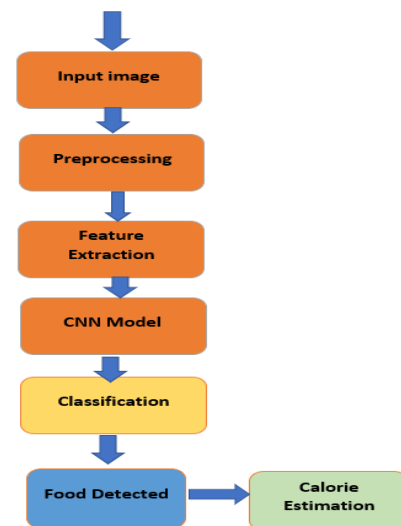


Figure 1: System Architecture of Food Detection

The architecture of the proposed food recognition and calorie estimation system operates in a multi-staged process which is intended to handle an input image towards a final output. That output summarizes the entire workflow of the system, beginning with a food image that is then preprocessed to address quality-related issues and

remove noise, along with resizing, reshaping and reformatting for processing. Next, the image is put through a feature extraction process to gather visual features deemed essential to food recognition systems. These features are then sent into a CNN model that learns hierarchical representations of the visual features to facilitate food classification. The classification module will generate the food label which will be chosen from the pre-studied classes of food items. Finally, the output will go through a calorie estimation step where mapping between the value of known nutrition in the identified item will be calculated to finalize the calorie estimate. This process completes a sustainable workflow where calorie estimation and recognition are bundled together in one pipeline.

7. SCREENSHOTS

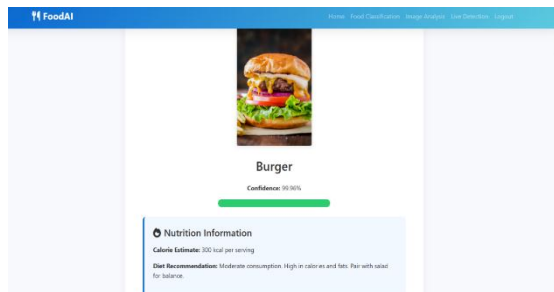


Figure 2 : Home page

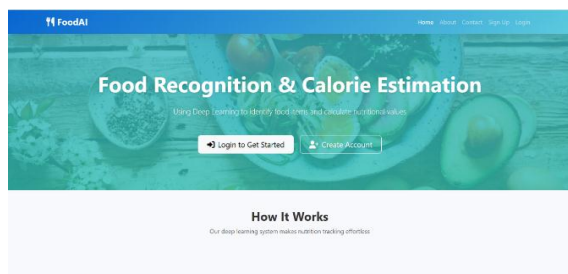


Figure 3 : Predicated Result

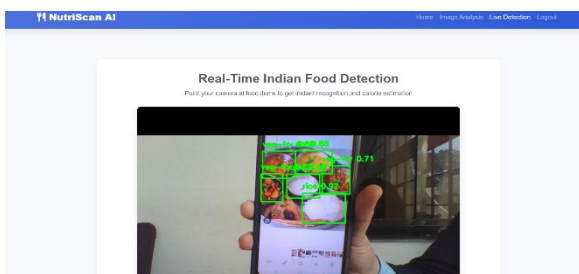


Figure 4 : Real-Time Food Detection

8. CONCLUSION & FUTURE SCOPE

This study presented an intelligent framework for automated food detection, recognition, and calories estimation by merging the use of CNN-based VGG19 and YOLO architecture. VGG19 provided strong feature extraction for classification, whereas YOLO provided advanced real-world object detection, accuracy, and real-time classification of food categories. The study also deployed the framework using a user-friendly Flask web application, allowing users to upload images and receive recognition and calorie information to support enhanced dietary monitoring and health management. The smart food detection system proved to be more automated, precise, provided real-time feedback, and relied less on human effort and error than traditional food logging methods in the study. Future work could expand the framework with larger and diverse datasets that allow for further generalization in real-world applications, including diverse cuisines, preparation styles, and serving sizes. Advanced architectures such as EfficientNet, Vision Transformers, or very large (ensemble) models could explore performance improvements and minimize computational costs. Moreover, the system could be extended to mobile applications with real-time capture features, multilingual and voice-enabled interactions, and integrated with wearable or fitness track device for ongoing engagement and continuous food detection.

REFERENCES

[1] D. Gomes, "Classification of Food Items Utilizing Deep Convolutional Neural Network with Transfer Learning," International Journal of Engineering and Management Research, vol. 14, no. 2, 2024.

[2] L. Bu, C. Hu, and X. Zhang, "Identification of Food Images Using Transfer Learning and Ensemble Techniques," BMC Public Health, vol. 24, article 280, 2024.

[3] Y. Zhao, P. Zhu, Y. Jiang, and K. Xia, "Visual Nutrition Analysis: Using Segmentation and Regression Techniques to Estimate Food Nutrients," Frontiers in Nutrition, vol. 11, article 1469878, Dec. 2024.

[4] R.S.A. Kareem, "Fine-grained food image classification and recipe generation using deep learning," Computers in Biology and Medicine, vol. 177, article 108612, 2024.

[5] Z. Shen, "Machine Learning based Approach on Food Recognition and Calorie Estimation," Procedia Computer Science, vol. 176, pp. 685–693, 2020.

[6] A. Karabay, "Enhanced food image recognition with deep learning and computer vision approaches," Scientific Reports, vol. 15, article 95770, 2024.

[7] Y. Liang and J. Li, "Food Calorie Estimation Based on Computer Vision: Dataset, Method and Experiment," Pattern Recognition, vol. 96, 2019.

[8] Research Group, "Nutritional Content Detection Using VisionTransformers -An Intelligent Method," International Journal of Innovative Research in Engineering and Management, vol. 11, no. 6, Dec. 2024.

[9] Research Group, "Indian Food Recognition and Calorie Estimation Using YOLOv8," International Journal of Computer Research and Technology, vol. 8, no. 6, 2023.