Newari Food DetectoChef: A Recipe and Description Generator for Authentic Newari Cuisine

Poojan Koju
Department of Computer Engineering
Khwopa College of Engieering,TU
Bhaktapur, Nepal
KCE076BCT023@khwopa.edu.np

Rocky Suwal
Department of Computer Engineering
Khwopa College of Engieering,TU
Bhaktapur, Nepal
KCE076BCT031@khwopa.edu.np

Romik Gosai
Department of Computer Engineering
Khwopa College of Engieering,TU
Bhaktapur, Nepal
KCE076BCT032@khwopa.edu.np

Sagar Suwal
Department of Computer Engineering
Khwopa College of Engieering,TU
Bhaktapur, Nepal
KCE076BCT034@khwopa.edu.np

***Abstract*—In recent years, deep learning has made impressive progress in computer vision, revolutionizing areas like image recognition, image segmentation, and object detection. Among these applications, object detection has gained much attention and is widely used in real-world situations. Notably, object detection has been increasingly applied to the domain of food recognition. In this paper, introduce a real-time system capable of recognizing Newari food items and generating corresponding recipes. However, due to the extensive variety of food types, recognizing food items through image recognition is generally challenging. Additionally, the unavailability of a comprehensive Newari food dataset further complicates the task. To address these challenges, we compiled a comprehensive Newari food dataset consisting of 40 frequently occurring food classes in traditional Newari meals. We employed transfer learning in conjunction with the YOLOv5m object detector model to achieve accurate and efficient food item recognition. For accurate and efficient training of our model, we require a large dataset containing a wide variety of food item images. However, due to the lack of such a dataset, we applied data augmentation techniques to enhance model performance and address data scarcity. Through augmentation, we expanded the dataset, leading to improved accuracy and generalization in detecting Newari food items. After performing necessary augmentation and hyperparameter tuning, we were able to obtain precision, recall, F1 score, and mAP-0.5 scores of approximately 70.58%, 71.56%, 71.07%, and 75.74%, respectively. These outcomes highlight the efficacy of our approach in recognizing Newari food items and providing valuable insights into this culinary tradition.**

***Index Terms*—Deep Learning, Object Detection, Newari Food Detection,**

I. INTRODUCTION

NEWARI cuisine is renowned for its rich flavors, cultural significance, and diverse array of dishes, making it a cherished culinary experience for both food enthusiasts and connoisseurs. Food holds a vital role in Newari culture, embodying the essence of ancient traditions dating back to the history of the Kathmandu valley. Newari cuisine encompasses a unique fusion of non-vegetarian and vegetarian items, alongside alcoholic and non-alcoholic beverages, appealing not only to Newars but also to non-Newars and tourists alike. However, non-Newars and tourists may encounter difficulties in recognizing Newari food items due to their unfamiliarity with the diverse range of dishes, ingredients, and presentation styles unique to Newari cuisine. The complexity and richness of Newari food can pose challenges in identifying and appreciating these delicacies without prior knowledge or guidance. Deep learning has revolutionized computer vision, elevating food detection to remarkable levels of accuracy and efficiency. Thus, this paper presents an innovative system that can bridge this gap and offer a comprehensive solution for food recognition and exploration.

Tourism serves as one of the primary drivers of Nepal's economy, making a substantial contribution to the country's Gross Domestic Product (GDP). According to the survey report by the ”Ministry of Culture, Tourism, and Civil Aviation”, traditional Newari cuisine holds immense appeal not only for locals but also for tourists visiting Nepal. However, despite the remarkable advancements in technology and machine learning, which have facilitated smarter marketing strategies, personalized travel recommendations, and seamless online booking experiences, the food industry focused on traditional Newari cuisine remains relatively unexplored.

The Newari Khaja Set is a delightful combination of various food dishes, including beaten rice, bara/wo: (newari pancake), choila (spiced grilled meat), and kwati (mixed bean soup), making it a rich and culturally significant gastronomic experience. However, accurately recognizing these dishes through image recognition encounters inherent challenges due to the wide diversity and intricate nature of Newari food items. Compounding the difficulty is the unavailability of a comprehensive Newari food dataset, further hindering the development of machine learning models for this specific cuisine.

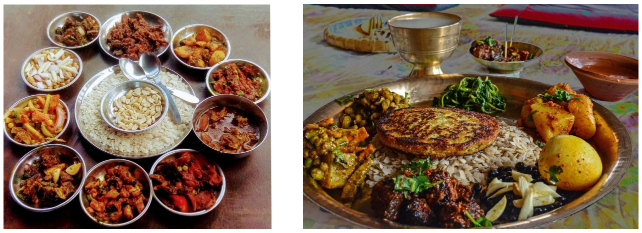


Fig. 1. Newari Khaja Set -Samay Baji

In order to address these challenges, we diligently curated a comprehensive Newari food dataset, encompassing 40 frequently found food classes in traditional Newari meals. Each food item in the dataset is carefully labeled, streamlining the training process for our AI model. However, these food dishes can be served either together on the same plate without distinct boundaries or in separate plates and bowls with distinct boundaries, making single-label image classification models ineffective. To tackle this issue, we adopted multi-label object detection model-YOLO, capable of detecting and localizing multiple items in an image using bounding boxes. By leveraging transfer learning techniques and implementing the YOLOv5 object detector model [1] , we successfully achieved precise and efficient recognition of Newari food items in real-time.

Our system not only provides instant identification of Newari dishes but also goes beyond by generating detailed recipes for each recognized food item. Leveraging an extensive repository of curated and authentic Newari recipes, our system offers users valuable insights into the culinary methods, ingredients, and cultural significance associated with each dish. This immersive experience allows users to explore the unique flavors and historical heritage embedded within Newari cuisine. Moreover, the system serves as an invaluable tool for preserving and sharing the rich traditions of Newari culinary artistry with a global audience. By combining the power of deep learning, transfer learning, and a carefully crafted dataset, our system marks a significant step forward in promoting Newari culture through the enticing world of food recognition and recipe generation.

In summary, this paper has the following contributions:

- We introduce a labeled image dataset for Newari food recognition tasks, covering 40 commonly found food classes in traditional Newari meals.
- We employing the YOLO model for real-time food item detection, to achieve fast processing, optimizing the utility and effectiveness of the limited available data to generate authentic recipes for the detected Newari food items.

In the following sections, we provide a comprehensive overview of our work on ”newari Food DetectoChef”. We begin with a discussion on the existing literature in section II (LITERATURE SURVEY), identifying gaps in the current research landscape. In section III (Methodology), we provide a detailed explanation of our methodology, including the construction of the Newari food dataset and the YOLO object detector model’s architecture. Subsequently, in section IV (Results and Discussion), we present the outcomes of our experiments and performance evaluations, followed by an in-depth discussion on the system’s potential and avenues for future enhancements in section V (Conclusion and Futute Enhancement).

II. LITERATURE SURVEY

In 2022, Pandey. D., et al. presented a paper that addresses the challenge of recognizing traditional Indian food dishes in images. The authors propose a solution by introducing two labeled datasets, IndianFood10 and IndianFood20 [1]. Utilizing transfer learning with the YOLOv4 object detector model on IndianFood10, they achieve a high mAP score of 91.8% for object detection. The paper emphasizes the significance of multi-label object detection models, which enable the localization of multiple dishes in a single image, making it practical for various applications in the Indian food domain.

In 2019, by Sun, J., et al. proposed a food-specialized detection deep learning architecture that leverages knowledge transferred from a pretrained food/non-food classification model [2]. Unlike existing approaches that treat object detection and image classification as separate tasks, their work bridges the gap between the two by utilizing transferred features. The experiments demonstrate that initializing the network with transferred features improves generalization, leading to significant precision improvement of over 10% compared to plain networks.

In 2017,Chen, X., et al. introduce ”ChineseFoodNet,” a large-scale food image dataset for recognizing Chinese dishes. The dataset contains over 180,000 food photos across 208 categories, including web recipe, menu pictures, and real dishes [3]. The paper outlines the dataset construction process, utilizing machine learning methods to reduce manual labeling efforts. The authors benchmark several deep convolutional neural networks on ChineseFoodNet and propose a two-step data fusion approach called ”TastyNet,” achieving high accuracy rates 81.43% and 81.55% on the validation and test sets respectively.

Another paper by Fakhrou. A., et al. at 2021, introduces a smartphone-based system to aid children with visual impairments in recognizing food dishes and fruit varieties [4]. The system uses a trained deep CNN model and a new deep CNN model developed through ensemble learning on a customized food recognition dataset. The ensemble model achieves 95.55% accuracy on the customized dataset, outperforming state-of-the-art CNN models.

In 2017, Termritthikun. C., et al. proposed a new network called NU-InNet, inspired by the Inception module used in GoogLeNet, to reduce processing time and model size [5]. The NU-InNet is tested on the THFOOD-50 database, containing 50 kinds of famous Thai food. Results show that NU-InNet reduces processing time by a factor of 2 and model size by a factor of 10 compared to GoogLeNet, while maintaining the same recognition precision. This significant reduction allows for efficient Thai-food recognition applications on smartphones.

In 2018, Subh. M.A., et al. proposed a new deep convolutional neural network (CNN) configuration for detecting and recognizing local food images [6]. The authors introduce a new dataset of popular Malaysian food items collected from publicly available Internet sources. CNN achieved higher accuracy compared to traditional approaches with manually extracted features for food recognition. Convolution masks revealed that food color features dominate the feature map. Additionally, CNN showed superior accuracy in food detection compared to conventional methods.

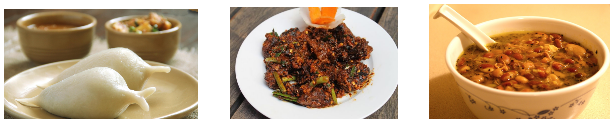
In 2020, Phiphiphatphais. S., et al. presented a paper that paper tackles the difficult task of classifying real-world food images, considering variations in perspectives and the presence of other objects [7]. The authors propose a modified MobileNet architecture with various enhancements to improve accuracy and prevent overfitting. Experimental results show that the proposed MobileNet version achieves significantly higher accuracies compared to the original architecture. When combined with data augmentation techniques, the proposed MobileNet outperforms other architectures in food image recognition.

III. METHODOLOGY

The primary goal of our project is to create an accurate deep learning-based model for detecting Newari food items and generating corresponding recipes. Our aim is to achieve a high level of precision in recognizing and categorizing the diverse array of Newari dishes. To accomplish this, we have carefully collected a specialized Newari food dataset manually, encompassing a wide variety of frequently appearing food classes in traditional Newari meals. Unlike relying on existing datasets, we have devoted efforts to create a unique collection specifically tailored for Newari cuisine. Utilizing the power of transfer learning with the YOLOv5 object detector model, we train our system on Google Colab, an efficient and cloud-based platform offered by Google. This ensures a convenient and powerful environment for training our AI model, enabling real-time Newari food item detection and recipe generation.

A. Dataset Description

The dataset used for training the object detection model comprises 903, encompassing 40 different Newari food item classes, comprising a fusion of non-vegetarian, vegetarian items, non-alcoholic and alcoholic beverages [8]. It has a total size of approximately 525 MB, with images varying in dimensions from 2448 x 3264 pixels to 385 x 332 pixels. The dataset was collected from Google and through manual photography at Newari food shops and home-cooked meals. In addition to the images, the dataset also includes a short description of each food item and corresponding recipes.



Yomari Choila Kwati(Soup)

Fig. 2. Sample images

B. System Workflow

1) Data Preparation: At the beginning of our project, we faced the challenge of not having an existing dataset for this particular task. Hence, the primary and essential step was to collect data. To create a comprehensive dataset, we thoroughly gathered images from various sources. We captured our own photos while visiting Newari food shops and preparing meals at home. Furthermore, to focus on the most relevant food items, we carefully selected 40 frequently or popularly consumed Newari food items. Additionally, we enriched the dataset by including images from public sources like Google and Instagram.

2) Annotation: We utilized an open-source software called makesense.ai [9] for the annotation of images in our dataset. Each image was carefully examined, and we manually created bounding boxes around each food item of interest, labeling each box with the corresponding food class. The annotations were saved in YOLO format, with a separate text file generated for each image, containing the information about the coordinates of the bounding boxes and their respective food class numbers. This comprehensive annotation process enabled us to prepare the dataset for training our YOLOv5-based object detection model effectively.

3) Data Processing: Before training our YOLOv5 model, we conducted data processing tasks to optimize the dataset. We resized all images to a uniform size of 640 pixels for consistency. Additionally, data augmentation techniques were applied, including default YOLOv5 augmentations along with rotation, translation, shearing, perspective transformation, and vertical flipping. These measures aimed to enhance model robustness and improve its ability to handle variations in real-world scenarios. Furthermore, we performed a split ratio of 80:20 on the processed dataset, dividing it into an 80% training set and a 20% validation set to fine-tune the model and evaluate its performance on unseen data.

4) Model Training: We trained the YOLOv5 model on the carefully prepared dataset, which involved setting up the necessary framework and importing essential libraries to facilitate the training process. During the training phase, the model iteratively learned from the dataset, adjusting its parameters to optimize the detection of Newari food items. The YOLOv5 architecture's efficiency and speed enabled us to train the model efficiently, resulting in a highly accurate and robust food recognition system. By leveraging the power of deep learning and transfer learning techniques, we successfully equipped the YOLOv5 model with the ability to recognize and localize a diverse range of traditional Newari dishes.

The optimizer used in our model is Stochastic Gradient Descent (SGD), a classic and widely-used optimization algorithm in deep learning. Its main principle involves updating model parameters in the direction of the negative gradient of the loss function with respect to the training data. This iterative process is performed on mini-batches of data during training, introducing randomness to the gradient estimation, which aids in avoiding local minima and achieving faster convergence. SGD remains a fundamental and effective choice for optimizing neural network models.

In the YOLOv5 model, the compound loss is calculated based on the objectness score, class probability score, and bounding box regression score. For the class probability and object score loss calculation, the Binary Cross-Entropy with Logits Loss function from PyTorch is employed. This loss function effectively handles the binary classification nature of the objectness score and class probability score. Additionally, the bounding box regression score is computed using the Mean Squared Error (MSE) loss function, which measures the discrepancy between the predicted bounding box coordinates and the ground-truth coordinates. By optimizing this combination of loss components, the model aims to achieve accurate and precise detection of Newari food items, enhancing the overall performance of the object detection.

Mathematical expression of Binary cross-entropy:

$$BCE(y_{\mathrm{t}}, y_{\mathrm{p}}) = -\frac{1}{N}\sum_{i=1}^{N}(y_{\mathrm{t}}\log(y_{\mathrm{p}}) + (1 - y_{\mathrm{t}})\log(1 - y_{\mathrm{p}}))$$

where, y_{t} represents actaul output,
y_{p} represents the predicted output of model,
N represents number of samples in batch

Mathematical expression of Mean Squared Error:

$$MSE(y_{\mathrm{o}}, y_{\mathrm{p}}) = \frac{1}{N}\sum_{i=1}^{N}(y_0 - y_{\mathrm{p}})^2$$

where, y_{t} observed value,
y_{p} predicted values,
N number of data points

Furthermore, learning rate and other parameter such as last learning rate , weigh decay and so on are also specified . Then, the fit function is utilized to train the model using the training and validation datasets. After training, the model is saved for future use. To assess its performance, various loss graphs are plotted, and precision, recall, and F1 score are computed using a confusion matrix which is described on (section IV).

5) Recipe Generation: After accurately identifying the Newari food items in the images in real-time, our system utilizes a vast repository of curated and authentic Newari recipes. For each recognized food item, the system generates detailed recipes, including ingredients, cooking instructions, and cultural significance. This feature provides users with instant access to traditional Newari recipes, enhancing their culinary experience and allowing them to explore and savor the rich heritage of Newari cuisine.

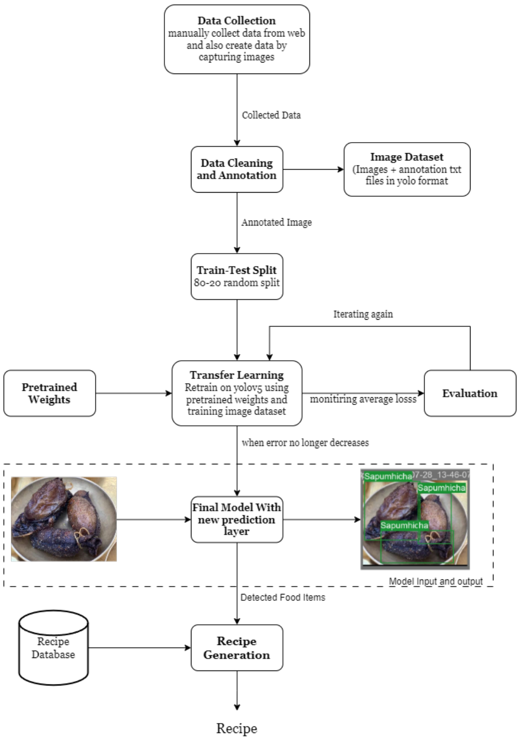


Fig. 3. System Workflow

C. YOLO V5

YOLOv5, an evolution of the YOLO architecture, is a highly efficient and accurate object detection model used extensively in computer vision tasks. It employs a CSPDarknet53 backbone network to extract intricate features from input images and multiple detection heads to predict bounding boxes, class probabilities, and confidence scores for objects. YOLOv5 integrates the focal loss to address class imbalance, assigning greater importance to challenging objects during training. One of its key advantages lies in its real-time performance, allowing fast processing of high-resolution images. The model's adaptability and scalability make it a versatile choice for various applications. Its easy implementation and well-documented code base make YOLOv5 accessible to researchers and developers for customization and fine-tuning to specific datasets and tasks. In our newari food detection project, we utilized YOLOv5 due to its combination of speed, accuracy, and flexibility, enabling precise recognition of Newari food items in real-time applications.

IV. RESULT AND DISCUSSION

We employed various evaluation methods to assess the performance of our YOLOv5 object detection model. During the model training process, different YOLO loss values were computed. The objective of the training was to minimize the loss function, which represents the dissimilarity between the predicted output and the ground truth. This minimization process helps improve the accuracy of the model's predictions for Newari food item detection.

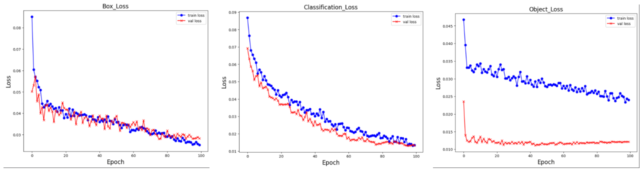


Fig. 4. Different Loss Plot Of YOLOv5 Model

The final loss values can be observed in the table given below.

Type/Loss	Train	validation
Box	0.0262	0.0281
Object	0.0236	0.0120
Classification	0.0175	0.0147

To enhance our model's accuracy, we conducted various experiments, exploring the effects of different design choices and hyper parameter adjustments. Some of the hyperparameter changes made during our investigation include: epoches = 100, rotation = 10 degrees, learning rate = 0.01, image translation = 0.2, shear = 0.2 , flipud =0.3 , mix-up = 0.2.

During our evaluation process, we employed standard metrics for object detection models, including Precision, Recall, and F1-score. To calculate these metrics, we plotted a confusion matrix by determining the number of True Positive (TP), False Positive (FP), and False Negative (FN) instances. Using the values from the confusion matrix, we further computed evaluation metrics like Precision (Positive Predictive Value), Recall (Sensitivity)and F1-score. This comprehensive analysis of the model's performance allows us to assess its precision in correctly identifying positive cases, recall in capturing all positive cases, and overall effectiveness in terms of accuracy and predictive power.

A. *Precision*

Precision measures of the accuracy of positive predictions or decisions made by a model or a system. Mathematically, it can be expressed as:

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

B. *Recall*

Recall measures how well the system is able to identify all instances of the positive class. Mathematically, it can be expressed as:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

C. *F1 Score*

The F1 score is a statistical metric used to evaluate the performance of a classification or detection system. Mathematically, it can be expressed as:

$$\text{F1 Score} = \frac{2*(precision*recall)}{(precision+recall)}$$

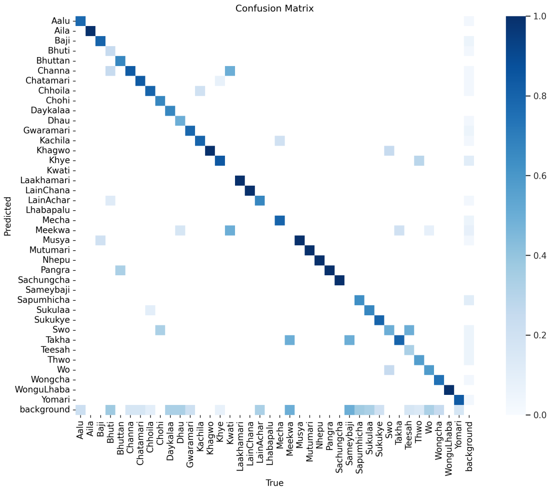


Fig. 5. confusion Matrix

Furthermore, we employed Intersection over Union (IoU) thresholds of 0.5 and 0.5 to 0.95 for the evaluation of our model. The scores of various metrics for different models were tabulated as follows:

Metrics/Models	**YOLOv5s**	**YOLOv5m**
Without hyperparameter tuning		
Precision	50.47%	72.88%
Recall	66.67%	65.76%
F1 score	57.45	69.17%
mAP_0.5	62.49%	72.80
mAP_0.5-0.95	30.20%	48.13%
With hyperparameter tuning		
Precision	55.54%	70.58%
Recall	55.57%	71.56%
F1 score	55.55%	71.07%
mAP_0.5	59.17%	75.74%
mAP_0.5-0.95	30.17%	48.38%

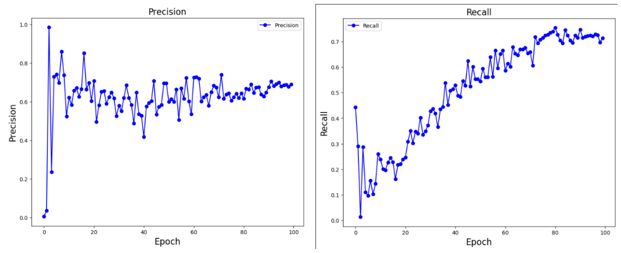


Fig. 6. Precision And Recall plot Of YOLOv5

The implementation of various augmentation techniques and hyperparameter tuning improve the performance of our YOLOv5 model. As a result, our system achieved a significant boost in precision (70.58%), recall (71.56%), and F1-score (71.07%). The mean average precision (mAP) also showed promising results, reaching 75.74% at an IoU threshold of 0.5 and 48.38% at an IoU threshold of 0.5 to 0.95. These enhancements showcased the potential of YOLOv5 in accurately detecting Newari food items.

However, we faced challenges related to the limited availability of a comprehensive and diverse dataset specifically for Newari food detection. To overcome this, we employed data augmentation techniques to prepared dataset artificially. While the augmentation improved performance, we acknowledge the importance of a richer and more authentic dataset to further enhance the model's generalization capabilities. Addressing the data scarcity and refining the model with real-world data remain essential considerations for future advancements in our system. With continued refinement, our YOLOv5-based approach has the potential to excel in identifying Newari food items and enriching users' understanding of this unique culinary tradition.

V. Conclusion and Future Enhancement

Newari Food DetectoChef has revolutionized the culinary world by employing the YOLO (You Only Look Once) model to generate authentic newari cuisine recipes and descriptions. This powerful combination ensures an unparalleled experience, providing users with accurate and detailed instructions to recreate traditional newari dishes. With YOLO's real-time object detection capabilities, DetectoChef enhances the visual representation of each dish, making the cooking process even more accessible and enjoyable. As a result, DetectoChef has become a cutting-edge tool, preserving the cultural heritage of newari cuisine while offering users a delightful gastronomic journey. As for Future Enhancements to our project, we could implement these:

- Introduce multilingual support for broader accessibility.
- Include video tutorials with step-by-step cooking instructions.
- Provide nutritional information for each recipe.
- Implement user ratings and reviews for feedback and community engagement.
- AI-generated ingredient substitutions for unavailable items.
- Seasonal and regional dish recommendations.
- Develop an interactive virtual cooking assistant.
- Collaborate with newari restaurants for signature dishes promotion.
- Provide detailed information about the historical significance and cultural context of each dish.

Acknowledgement

We extend our sincere gratitude to the ICTCEEL-2023 organizers for their support and our institution, Khwopa College of Engineering, for providing resources and an academic environment conducive to our research. Special thanks to the Newari restaurants for generously sharing their datasets, enabling the development of our machine learning model. We also appreciate the invaluable contributions of all individuals involved in this project. Your support has been instrumental in our exploration of Newari food detection.

References

[1] D. Pandey, P. Parmar, G. Toshniwal, M. Goel, V. Agrawal, S. Dhiman, L. Gupta, and G. Bagler, "Object detection in indian food platters using transfer learning with yolov4," in *2022 IEEE 38th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, 2022, pp. 101–106.

[2] J. Sun, K. Radecka, and Z. Zilic, "Exploring better food detection via transfer learning," in *2019 16th International Conference on Machine Vision Applications (MVA)*. IEEE, 2019, pp. 1–6.

[3] X. Chen, Y. Zhu, H. Zhou, L. Diao, and D. Wang, "Chinesefoodnet: A large-scale image dataset for chinese food recognition," *arXiv preprint arXiv:1705.02743*, 2017.

[4] A. Fakhrou, J. Kunhoth, and S. Al Maadeed, "Smartphone-based food recognition system using multiple deep cnn models," *Multimedia Tools and Applications*, vol. 80, no. 21-23, pp. 33 011–33 032, 2021.

[5] C. Termritthikun, P. Muneesawang, and S. Kanprachar, "Nu-innet: Thai food image recognition using convolutional neural networks on smartphone," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, no. 2-6, pp. 63–67, 2017.

[6] M. A. Subhi and S. M. Ali, "A deep convolutional neural network for food detection and recognition," in *2018 IEEE-EMBS conference on biomedical engineering and sciences (IECBES)*. IEEE, 2018, pp. 284–287.

[7] S. Phiphiphatphaisit and O. Surinta, "Food image classification with improved mobilenet architecture and data augmentation," in *Proceedings of the 3rd International Conference on Information Science and Systems*, 2020, pp. 51–56.

[8] P. Koju, R. Suwal, R. Gosai, and S. Suwal. [Online]. Available: https://www.kaggle.com/datasets/pujancozu/newarifooddetection

[9] P. Skalski, "Make Sense," https://github.com/SkalskiP/make-sense/, 2019.