Thai Food Detection and Classification

Napat Jirataranon
Faculty of ICT, Mahidol University, Nakhon Pathom, Thailand 73170
napat.jir@student.mahiol.ac.th

Natthaphat Pintip
Faculty of ICT, Mahidol University, Nakhon Pathom, Thailand 73170
natthaphat .pin@student.mahidol.ac.th

Sittikorn Maneewong
Faculty of ICT, Mahidol University, Nakhon Pathom, Thailand 73170
sittikorn.man@student.mahidol.ac.t h

***Abstract*—This project introduces an intelligent Thai food detection and classification system aimed at assisting health-conscious individuals in estimating their calorie intake from images. By integrating YOLOv11 for object detection and ResNet34 for image classification, the system can identify multiple Thai dishes from an uploaded image, classify them, and provide corresponding calorie estimates. This solution is designed for real-time application and enhances user convenience by automating food tracking.**

Keywords—Thai food detection, YOLOv11, ResNet34, calorie estimation, object detection, image classification, deep learning.

I. Introduction

In recent years, personal health monitoring has gained increasing attention, especially in diet and calorie tracking. Manual meal logging can be time-consuming and inaccurate. This project proposes a visual-based solution that allows users to upload an image of a meal and automatically receive food item classifications and estimated calorie counts. Key contributions include:

- An end-to-end pipeline that detects and classifies Thai food from images.
- Integration of YOLOv11 for object detection and ResNet34 for dish classification.
- Calorie estimation based on THFOOD-50 dataset labels.

II. Methodology

A. Dataset

Two primary datasets were used in this project:

- ***Platefood Dataset (Roboflow)**: Annotated images of Thai food plates for YOLOv11 object detection.*
- ***THFOOD-50 Dataset**: Images labeled with 50 Thai dish categories for training the ResNet34 classifier.*

Images were cleaned using a custom datasetcleaner.py script to eliminate corrupted files.

B. Model Development

- ***Object Detection (YOLOv11)***
  - *Variants: YOLOv11 (Models L, M, N)*
  - *Fine-tuning on Platefood dataset*
  - *Optimizer: AdamW, custom learning rate*
  - *Loss functions: Box loss, classification loss, DFL loss*
  - *Adjustments for underrepresented classes (e.g., plates)*
- ***Classification***
  - *Fully connected layer replaced with Dropout + Linear classifier*
  - *Data augmentation: Horizontal flips, random color jitter, random resized crop, Gaussian blur, normalization (ImageNet mean/std)*

III. Results

A. Experimental Setup

All models were trained for 100 epochs with early stopping. Evaluated on GPU environment with precision-recall and F1-Confidence curves.

B. Performance Metrics

The following metrics were used to evaluate both object detection and classification:

- *mAP@0.5 and mAP@0.5:0.95: For measuring localization accuracy.*
- *Precision, Recall, F1 Score: For measuring classification quality.*
- *Confusion Matrix (Normalized): To visualize per-class performance.*

C. Model Comparison

- *Detection Models (YOLOv11)*

Metric / Feature	Model L	Model M	Model N
F1 Score (All Classes)	0.60 @ 0.076	0.58 @ 0.069	0.68 @ 0.262
Precision (All Classes)	1.00 @ 0.918	1.00 @ 0.947	1.00 @ 1.000
Recall (All Classes)	0.94 @ 0.000	0.94 @ 0.000	0.91 @ 0.000
mAP@0.5	0.720	0.714	0.708
mAP@0.5:0.95	0.60	0.58	0.60
Confusion Matrix	Misclassifications (dish1, plate)	Higher confusion (dish2, plate)	Best separation
Validation Loss Trend	Stable with fluctuations	Slight fluctuation	Smooth decline
F1-Confidence Stability	Moderate	Moderate	Most consistent
Training Convergence	100 epochs	100 epochs	100 epochs
Overfitting	Slight overfitting	Mild underfitting	Best generalization balance

Chosen Model: Model L was selected due to its optimal balance between high precision and recall, particularly its superior mAP@0.5 and stable validation loss trend, making it most suitable for accurate plate and dish detection in practical use.

- *Classification Models (ResNet)*

Model	Learning rate	Batch size	Epoch	Dropout	Train loss	Valid Loss	Train-Valid diff	Result	Train acc	Test acc
ResNet50	1e-4	64	20	0	0.1338	0.3865	0.2527	Overfit	97.0%	88.1%
ResNet50	1e-4	128	20	0.3	0.1239	0.3273	0.2034	Overfit	96.6%	89.1%
ResNet50	1e-4	128	20	0.5	0.1311	0.3768	0.2457	Overfit	96.1%	88.4%
ResNet18	1e-4	128	20	0.3	0.2332	0.4776	0.2444	Overfit	91.8%	86.3%
ResNet34	1e-4	16	20	0.3	0.2695	0.4134	0.1439	Overfit	92.3%	89.4%
ResNet34*	1e-4	32	20	0.3	0.3154	0.484	0.1686	Overfit	90.0%	86.9%
ResNet34*	1e-4	32	8	0.3	0.4233	0.4466	0.0233	Just right	87.0%	87.5%

*Chosen Model: ResNet34 (*with weight decay and transformer adjustments) was selected for its balanced accuracy, lowest train-validation loss difference, and stable performance, indicating strong generalization capabilities.*

IV. CONCLUSIONS

Successfully integrated YOLOv11 and ResNet34 to create an effective Thai food detection and calorie estimation system. Plate detection requires further improvements due to dataset imbalance. Future work includes employing focal loss, threshold tuning, and targeted data augmentation.

REFERENCES

[1] Chakkrit Termritthikun. (2021). THFOOD-50: A dataset for Thai food image classification.Available online: https://github.com/chakkritte/THFOOD-50

[2] Pathirana, S. (2023). Plate Food Dataset. Roboflow Universe Dataset Repository. Available online: https://universe.roboflow.com/subhash-pathirana-cdryb/plate_food

[3] Ultralytics. (2024). YOLOv11: You Only Look Once Object Detection and Classification. Ultralytics YOLO. Available online: https://docs.ultralytics.com

[4] PyTorch. (2024). PyTorch: An open-source machine learning framework. Meta AI. Available online: https://pytorch.org

[5] Gradio. (2024). Gradio: Build & Share Delightful Machine Learning Apps. Hugging Face. Available online: https://www.gradio.app