Campus Cafeteria Recognition and Billing System - Based on YOLOv7 and Regression Models

Zi-Yun Lai
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
lai-zi-yun@gapp.nthu.edu.tw

Jhen-Wei Huang
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
abc2985556655@gapp.nthu.edu.ew

Cheng-En Ho
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
leoho1040088@gmail.com

Chuan-Jhong Li
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
redblackgreen@gapp.nthu.edu.tw

Yu-Sheng Ma
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
s110062131@m110.nthu.edu.tw

Wei-Syuan Liao
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan y
flo511jb@gmail.com

***Abstract*—This research investigates the feasibility of integrating YOLOv7 and regression models into the pricing system of the self-service cafeteria at our university, referred to as "Jin-Zhan." We developed a system that utilizes YOLOv7 for dish recognition on the tray and employs a regression model to learn the underlying pricing strategy. The data for training the model was collected in-house. Ultimately, our system demonstrated reliable performance with YOLOv7, and while there is room for improvement in the regression model, we believe that with further refinement, this approach is viable.**

Keywords—YOLOv7, Meal Recognition, Billing System, Regression Model

I. Introduction

Jin-Zhan, a self-service cafeteria at National Tsing Hua University, charges for meals based on the variety and quantity of dishes, with additional price adjustments made according to the portion sizes. However, these pricing standards are quite ambiguous and have frequently sparked debates online. In response, we propose an automated checkout system utilizing a regression model to learn the cafeteria's implicit pricing mechanism, aiming to make the self-service meal payment process more equitable. We aim to achieve 100% accuracy under the condition of a +/-5-dollar error margin.

Overall, we employ a YOLOv7 model trained on our collected data for dish recognition and a regression model for price estimation. In practice, image data is initially processed by YOLOv7, followed by the regression model using the identified information for price estimation.

This paper will first review related research, highlighting the advantages and distinctions of our study. We will then detail our methods and results, focusing on the regression model and YOLOv7, followed by a discussion based on our findings.

II. Related Work

To the best of our knowledge, there is a limited body of research focused on recognition and pricing systems for self-service cafeterias within Taiwan. A study of note that aligns with our research is the one by Andy Wu [1], which utilizes the YOLOv3 model for the automated recognition and billing of cafeteria dishes. This research is similar in scope to our own. However, a key difference lies in the dataset size; Wu's study collected approximately 500 data points, whereas our study boasts a dataset of over 1000 entries. Furthermore, Wu's system was only capable of identifying 12 dish types. In contrast, our approach allows the system to accurately obtain pricing information for more than 70 different types of dishes. Additionally, Wu's study was limited to situations where each dish has a fixed quantity and price. Our study, on the other hand, attempts to expand the application to scenarios where pricing is based on the quantity of the food.

III. Methodology

This section outlines the methodology employed in our research, focusing on the integration of the YOLOv7 model for dish recognition and the regression model for price estimation in the self-service cafeteria system.

A. Data Collection and Proccesing

This study commenced with an analysis of the pricing criteria used in the cafeteria. The dishes were primarily categorized into three types: rice, side dishes (mostly vegetables, tofu, etc.), and main dishes (such as pork, chicken, and fish). The cafeteria's standard pricing was identified as: 50 for a set of one rice and three side dishes, with main dishes being priced separately.

However, the pricing for other combinations, such as two dishes with one rice, was not explicitly stated. Through extended analysis, we were able to preliminarily deduce the pricing standard for each type of dish: 10 for a side dish, 20 for rice, and main dishes priced as marked. We used these findings as a benchmark, defining the 'fair price' as the cost calculated based on this formula.

The data collection phase was executed during the month of November in the year 2023, resulting in the acquisition of a dataset comprising 1017 images. This meticulous process unfolded during the designated lunch and dinner hours, spanning Monday through Friday, facilitated by a well-structured rotating schedule. Two dedicated team members were stationed in close proximity to the cafeteria's checkout counter, serving as the point of initiation for data collection.

Each of the two team members meticulously captured approximately twenty images during each predefined time slot. The photographic process was conducted with utmost

precision, ensuring that each dish on the participants' plates was vividly and comprehensively depicted, devoid of any occlusion or overlap by adjacent items.

Subsequently, the initial set of photographs underwent thorough preprocessing by dedicated members of the data team. Their responsibilities encompassed the critical tasks of filtering out suboptimal images and meticulously eliminating extraneous artifacts such as background tables or inadvertent inclusion of customers' hands. This rigorous preprocessing ensured the dataset's fidelity and integrity.

To enrich and diversify the dataset, the data collection process incorporated various data augmentation techniques. These techniques included image flipping, rotation, controlled blurring, adjustment of brightness levels, and the introduction of controlled pixel noise. Each augmentation method was meticulously applied to enhance the robustness of the dataset and mitigate potential biases.

We employed Roboflow for the labeling of our dataset. Initially, we established two primary container labels, “plate” and “box”, with the aim of facilitating subsequent regression models in predicting meal prices based on area. Rice items were labeled as “brown rice”, “purple rice”, and “white rice”, while side dishes were uniformly labeled as “side_dish”. For the main course, distinct labels were initially assigned for each main course type. However, midway through the project, our modeling team observed that the numerous main course labels resulted in an insufficient number of samples for each class. Consequently, we adjusted the main course labels based on price, resulting in “main_dish25”, “main_dish30”, and “main_dish40”, where the number means the price of the dishes. The sample image data is illustrated in Fig. 1.

Fig. 1. The sample image data

Finally, the training dataset consisted of 1788 data points, the validation dataset comprised 283 data points, and the test set contained 137 data points.

In addition to food images, we systematically collected the prices associated with each image and calculated the “fair price” for every picture. In essence, each data entry encompasses details about the types and positions of dishes, food image, the fair price estimation, and the actual price.

B. *YOLOv7 Model for Dish Recognition*

The YOLOv7 model was chosen for object recognition due to its robust performance. As a one-step model, YOLOv7 efficiently utilizes resources while maintaining high accuracy. The training was conducted in the Google Colab environment, utilizing T4 RAM to accelerate the training process. A batch size of 8 was employed for training, and the model underwent training for 100 epochs, which amounted to approximately four hours of training time.

C. *Regression Model for Price Estimation*

We initiated the discussion by exploring the model's parameters. Our regression model is comprised of five crucial parameters: the area of the container, the area dedicated to rice, the combined area for main dishes, the total area encompassing side dishes, and the numerical count of side dishes. It is imperative to underscore that our dependent variable does not represent the actual price. Instead, we leverage the price difference between the actual price and a reference point known as the “fair price” within our model. Further details on the concept of fair prices will be expounded upon in subsequent discussions.

An elucidation on how the model's outcomes are employed for price prediction is imperative (“Fig. 2”). Upon obtaining the price difference, simultaneous computations are performed based on the YOLOv7 results, employing a fair formula provided by Jinzhan (The same formula is instrumental in computing the fair price before the training phase of our model). This culminates in the determination of the “fair price”, which is then combined with the price difference to yield a floating-point predicted price. Subsequently, this figure is rounded to the nearest integer multiple of five, thereby finalizing our prediction.

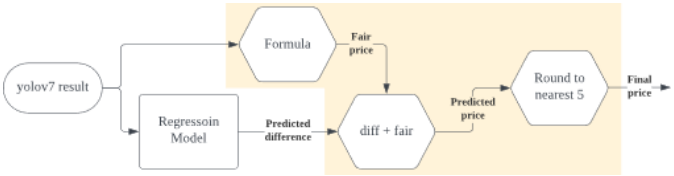

Fig. 2. Regression model prediction flow

IV. Result

In this section, we present the results of our experiments and evaluate the performance of our proposed system. We provide a comprehensive analysis of the accuracy and effectiveness of our model in recognizing and pricing cafeteria meals.

A. *Data analysis*

We computed the distribution of prices within our dataset and summarized it in the form of a pie chart (“Fig. 3”).

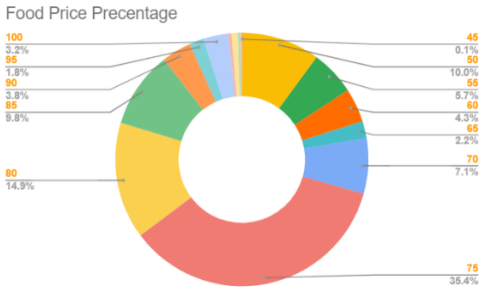

Fig. 3. Food price distribution in the dataset

As observed, when utilizing a naive strategy, we can achieve a minimum accuracy of 35.4%. By expanding the error margin to ± 5 dollars, the accuracy improves to approximately 60.0%.

We also compiled statistics on the discrepancies between the 'fair price' and the ground truth, as depicted in the accompanying figure. (“Fig. 4”).

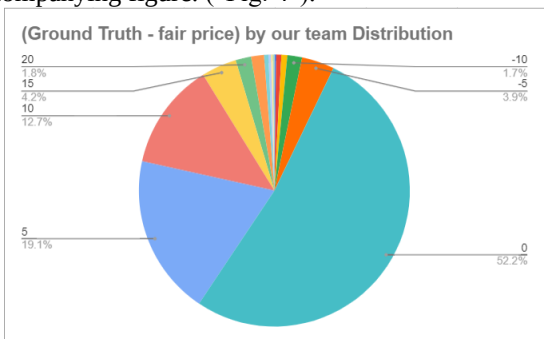

Fig. 4. Error distribution between fair price and ground truth

As observed, when utilizing the “fair price” strategy, we can achieve a minimum accuracy of 52.2%. By expanding the error margin to ± 5 yuan, the accuracy improves to approximately 75.2%.

B. Recognition Accuracy

The comparison between the training results and ground truth is shown in Fig. 5 and Fig. 6.

Fig. 5. The labeled image

Fig. 6. The recognized results generated from the model.

As observed, the trained YOLOv7 model demonstrates a high level of precision in accurately recognizing the types and positions of dishes.

Fig. 7 illustrates our confusion matrix. The values along the diagonal line exceed 0.9, indicating a confidence level of over 90% in the correctness of our predictions. However, the class “main_dish_40”exhibits lower performance, attributed to the limited amount of collected data for this category, hindering optimal model training.

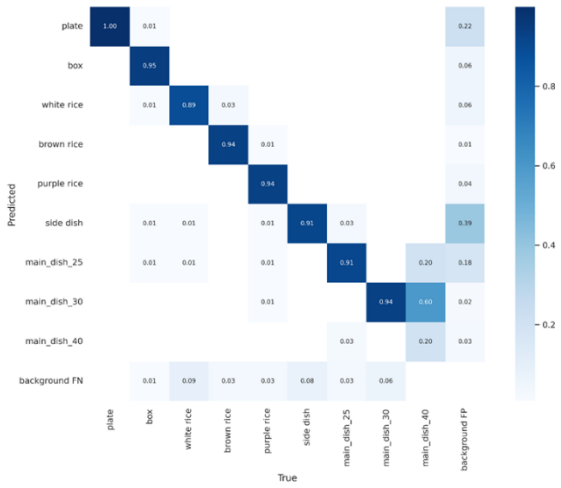

Fig. 7. Confusion matrix

Fig. 8 and Fig. 9 depict the precision and recall curves, respectively, generated by the YOLOv7 model. Precision is calculated using the formula (1):

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (1)$$

Recall is calculated using the formula(2):

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (2)$$

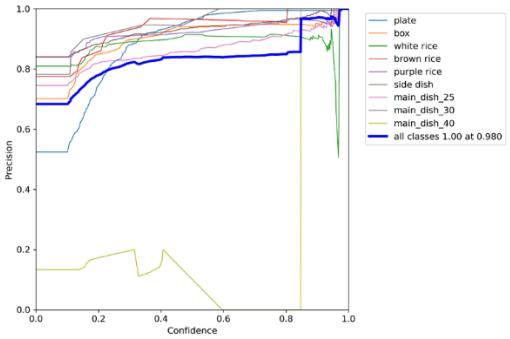

Fig. 8. Precision curves

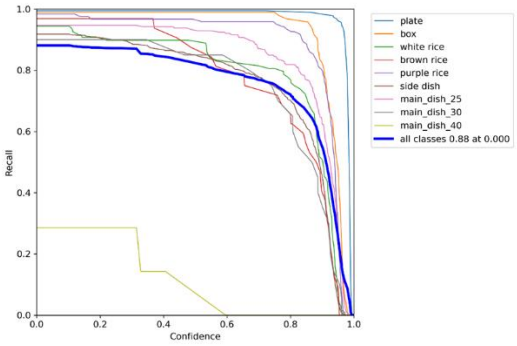

Fig. 9. Recall curves

Additionally, the F1-score, representing the balance between precision and recall, is calculated using the formula(3):

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (3)$$

Fig. 10 is introduced to illustrate the F1-score, providing a comprehensive overview of the model's performance across precision and recall.

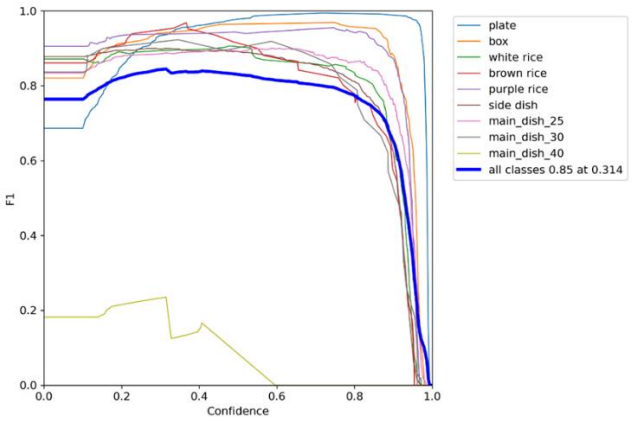

Fig. 10. F1-score curves

C. Regression Model for Price Estimation

Transitioning to the outcomes of our regression model, our evaluation reveals noteworthy insights. The model's performance metrics underscore its efficacy, with a 7% Mean Absolute Percentage Error (MAPE) on the validation dataset and a 6.5% MAPE on the test dataset ("Fig. 11"). These metrics serve as a testament to the model's accuracy in predicting price differentials. Moreover, specific enhancements implemented in the model, such as reducing the variety of main dishes and applying a simplified assumption to the count of side dishes, have demonstrated a tangible impact, resulting in a relative decrease in MAPE and an improvement in accuracy. This synthesis of methodological intricacies and tangible results underscores the robustness of our regression model in the context of price prediction within our project.

MAPE		***Accuracy*[a]**
Valid dataset	***Test dataset***	
7.09%	6.55%	86.6%

[a]Accuracy within a range of ± 5 dollars on Test data

Fig. 11. Regression model performance

In the table below ("Fig. 12"), we compare three methods: naive guess, fair price formula, and regression model, listing their accuracy on the test set.

Accuracy Comparison	**Accuracy**	
	Accuracy	***±5 dollars accuracy***
Naive guess	34.3%	60.0%
Fairprice formula	54.7%	75.2%
Regression model	34.3%	86.6%

Fig. 12. Accuracy comparison table

With an error margin of ± 5 dollars, our model exhibits improved performance.

D. Throughput

Finally, we integrated both models and assessed the overall accuracy.

Model Accuracy	**Accuracy**	
	Accuracy	***±5 dollars accuracy***
Our system	30.9%	80.2%

Fig. 13. Overall system accuracy

In the overall system testing, we conducted tests exclusively using the test dataset due to time constraints. While we couldn't perform comprehensive on-site testing, some observed issues include price prediction fluctuations when using a webcam for streaming input. This issue can be mitigated by using a fixed mounting setup to stabilize the input angle. Additionally, due to the limited diversity in the training data, there was a tendency for the model to occasionally misclassify non-food objects as food items.

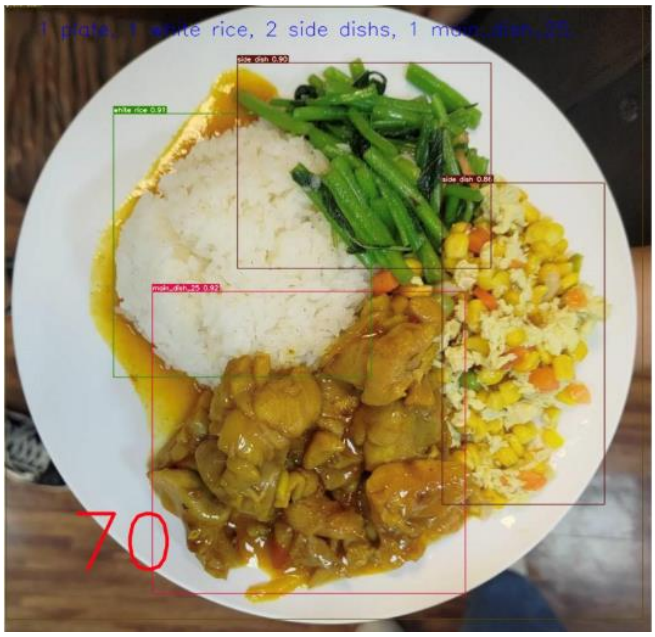

Fig. 14. The generated result from our system

V. Discussion

In this section, we will discuss the performance of the Regression model and YOLOv7 model, as well as explore potential avenues for improvement.

A. YOLOv7 Model

We examined the model's labeling results and identified errors in recognition, such as instances where two overlapping green vegetables could not be successfully distinguished by the model. Some of these errors were attributed to inaccuracies in the data labeling process.

Furthermore, in the early stages of our research, we standardized the format of the images to simplify training. However, this led to a reduction in the model's overall versatility. To enhance the model's practicality, we propose the inclusion of images captured from different angles, lighting conditions, and varying plate sizes. Additionally, introducing non-food objects into the scenes can contribute to a more robust model.

The primary factor contributing to the accuracy disparity between the overall system and the regression model was the recognition errors made by YOLOv7. By implementing the measures, we believe that we can significantly improve the overall accuracy of the system.

B. Regression Model

Based on the results, we find the performance of the Regression model to be unsatisfactory. However, it's important to note that we cannot conclusively determine the existence of hidden pricing strategies.

We believe there are several avenues for improving the accuracy of the regression model. For instance, in image processing, we can incorporate semantic segmentation to enhance the precision of dish area estimation. Additionally, using structured lighting can provide height information, leading to better volume estimation. Training separate models for different checkout personnel and considering consumer-related information like gender, in addition to food parameters, are also potential strategies for improvement.

However, it's worth mentioning that as the number of parameters increases, the required amount of training data also rises. In our specific case, the dataset was already limited, and incorporating additional variables would pose significant challenges. Given our current dataset, expanding the range of variables under consideration would indeed make the research more complex.

VI. Conclusion

The establishment of a self-service meal recognition and pricing model using YOLOv7 and Regression model has proven to be a feasible solution. In terms of recognition, high-quality results can be achieved with a dataset of just over 1000 images, and the labeling methodology developed in this study allows the system to adapt to changes in menu items without the need for retraining. As for the Regression model, we employed a basic approach, leaving room for the exploration of various advanced techniques. Therefore, in the context of the pricing model, we consider it a promising avenue for further development.

Data And Code Availability

If you would like to know more details about this study, please refer to this website, https://github.com/lazumo/Jinzhan-Cafeteria-AI which includes training data and related code for the model.

Contribution

Z.-Y. L (20%): Task assignment, progress planning, meeting minutes, proposal writing, data collection, report integration, providing consultation, team leadership.

C.-E. H (17%): Data collection, data filtering and labeling, assistance in report writing, model improvement support.

J.-W. H (17%): Data collection, data filtering and labeling, assistance in report writing, model improvement support.

C.-J. L (12%): Data collection, assistance in report organization.

Y.-S. M (17%): Regression model setup, model testing, model training, oral reporting, assistance in report writing.

W.-S. L (17%): Training YOLOv7 model, assistance in report writing, oral reporting, proposal writing.

References

[1] A. Wu, "採用 YOLOv3 模型的自助餐菜色自動辨識結帳系統," AndyWu's Notes, Feb. 02, 2020. [Online]. Available: https://notes.andywu.tw/2020/%E6%8E%A1%E7%94%A8yolov3%E6%A8%A1%E5%9E%8B%E7%9A%84%E8%87%AA%E5%8A%A9%E9%A4%90%E8%8F%9C%E8%89%B2%E8%BE%A8%E8%AD%98%E7%B5%90%E5%B8%B3%E7%B3%BB%E7%B5%B1/.

[2] Kin-Yiu Wong. YOLOv7 github. Available: https://github.com/WongKinYiu/yolov7 , Accessed on: Oct. 2023.

[3] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M., "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023*, pp. 7464-7475.

[4] Alberto Rizzoli, "The Ultimate Guide to Object Detection," V7 Labs, https://www.v7labs.com/blog/object-detection-guide , Jub. 2021 (accessed Oct. 2023).

[5] James Skelton, "Step-by-step instruction for training YOLOv7 on a Custom Dataset," Paperspace, https://blog.paperspace.com/train-yolov7-custom-data/ , Jan. 2023 (accessed Oct. 2023).