*sustainability*

MDPI

*Article*

# Consumer Usability Test of Mobile Food Safety Inquiry Platform Based on Image Recognition

**Jun-Woo Park [1], Young-Hee Cho [2], Mi-Kyung Park [2,*] and Young-Duk Kim [1,*]**

[1] Division of Automotive Research, Daegu Gyeongbuk Institute of Science & Technology (DGIST), Daegu 42988, Republic of Korea; qkrwnsdn233@dgist.ac.kr
[2] School of Food Science and Biotechnology, Kyungpook National University, Daegu 41566, Republic of Korea; aprilyhee@naver.com
* Correspondence: parkmik@knu.ac.kr (M.-K.P.); ydkim@dgist.ac.kr (Y.-D.K.); Tel.: +82-53-785-4641 (Y.-D.K.)

**Abstract:** Recently, as the types of imported food and the design of their packaging become more complex and diverse, digital recognition technologies such as barcodes, QR (quick response) codes, and OCR (optical character recognition) are attracting attention in order to quickly and easily check safety information (e.g., food ingredient information and recalls). However, consumers are still exposed to inaccurate and inconvenient situations because legacy technologies require dedicated terminals or include information other than safety information. In this paper, we propose a deep learning-based packaging recognition system which can easily and accurately determine food safety information with a single image captured through a smartphone camera. The detection algorithm learned a total of 100 kinds of product images and optimized YOLOv7 to secure an accuracy of over 95%. In addition, a new SUS (system usability scale)-based questionnaire was designed and conducted on 71 consumers to evaluate the usability of the system from the individual consumer's perspective. The questionnaire consisted of three categories, namely convenience, accuracy, and usefulness, and each received a score of at least 77, which confirms that the proposed system has excellent overall usability. Moreover, in terms of task completion rate and task completion time, the proposed system is superior when it compared to existing QR code- or Internet-based recognition systems. These results demonstrate that the proposed system provides consumers with more convenient and accurate information while also confirming the sustainability of smart food consumption.

**Keywords:** food safety; image recognition; artificial intelligence; usability test

check for updates

**Citation:** Park, J.-W.; Cho, Y.-H.; Park, M.-K.; Kim, Y.-D. Consumer Usability Test of Mobile Food Safety Inquiry Platform Based on Image Recognition. *Sustainability* **2024**, *16*, 9538. https://doi.org/10.3390/su16219538

Academic Editor: Giuliana Vinci

Received: 30 September 2024
Revised: 26 October 2024
Accepted: 30 October 2024
Published: 1 November 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

As free trade agreements (FTAs) between countries are expanded worldwide, the amount of food trade is on the rise every year and the kinds of imported food products that consumers can purchase in the market are also increasing. The increase in imported food has the advantage of broadening the range of choices according to tastes, nutrition, and the health conditions of consumers, but the indiscriminate consumption of imported food of which the safety of has not been verified may be fatal to health. For example, certain products can contain unsuitable ingredients or raw materials regulated by each country's government. In addition, due to the complex distribution process, food products may suffer from deterioration, contamination, damage, and expiration. Moreover, it is still inconvenient for consumers, as they should personally check before purchasing or consuming. Firstly, since most imported food products are distributed with food safety labels (e.g., nutrition, ingredients, allergies, expiration date, etc.) printed or designed in the local language of each manufacturing country, it is difficult for consumers to obtain accurate information if they do not speak the language of each country. Secondly, although the safety information is indicated by attaching an adhesive printed label (or sticky label) in the local language, it is still difficult for general consumers to grasp it intuitively due

to the small font size and lack of readability. Thirdly, even if the product name of the imported food purchased by the consumer is identified by simply reading the product label, it is still difficult to confirm in real time whether the corresponding food is unsuitable for consumption or subject to recall. Currently, it is impossible to confirm a recall by directly looking at the product, but there is a way to confirm this by accessing a specific website (e.g., the FDA [1] in the U.S.A., RASFF [2] in the EU, KFDA [3] in Korea) which provides it. However, consumers still have to make an effort to input the website address and product name themselves, and there is also the problem of similar product names being searched at the same time, which is still inconvenient. In summary, in order for consumers to safely purchase and consume food, it is essential to check three requirements as follows:

(1) Fundamental information (e.g., name or ID of food, nutrition facts, etc.) on purchased products.
(2) Safety information on whether the food is unsuitable or subject to recall.
(3) The above information should be accurately checked in real time.

In order to solve the above issues by breaking away from the legacy analog labeling method, many studies have been conducted. As a representative example, traceability systems [4–6] are explored to track product information and location in the distribution process. Recently, one of the most popular traceability methods is two dimensional codes such as barcodes and QR (quick response) codes [7]. RFID (radio frequency identification) [8] and NFC (near-field communication) [9] are also alternative solutions. Although the above technologies have a great advantage in that the ID of the product is accurately recognized in real time using a specific light source or communication chip, there is a significant disadvantage in that it is difficult for innocent consumers to use in daily life due to the fact that a specific dedicated optical reader and communication device are required. In addition, QR codes and barcodes also have a precondition that a separate label (e.g., sticker) must be attached to the existing wrapping package or inserted into the package in advance when the product is manufactured. In addition to providing food safety information, QR codes are also used for a variety of purposes including marketing, surveys, and other commercial events. That is, if QR codes are used for unclear purposes, consumers will be confused in finding the safety information they want.

Meanwhile, as the spread of mobile terminals such as smart phones has recently expanded and AI (artificial intelligence)-based object recognition technology and services have been widely provided to general consumers, new research is being conducted in the field of food identification and recognition in conjunction with the subject of food safety. As one of the most representative technologies for food recognition, the OCR (optical character recognition) system [10] allows for the accurate and fast recognition of food product information. In general, OCR is a method of extracting and capturing characters printed on a product using a vision camera and outputting them as digitized characters or worlds, thus enabling access to and editing of the original text data. However, since it is affected not only by the size, color, and the font style of the characters but also by the background color of the paper on which the characters are printed on the actual product, the recognition rate is not constant. In addition, when the illumination intensity and the camera shot angle are unstable, the accuracy becomes more vulnerable. To explore these problems, many studies have been conducted recently to improve recognition performance by adopting various image processing and deep learning schemes. Representatively, the authors of [11] designed a text detection network which can be learned in an end-to-end method based on deep learning, and through this, they proposed accurate text location prediction and a method for distinguishing it from the background. In the field of food service, the authors of [12] also propose MenuAI, which recognizes the product name (text) in a restaurant menu with OCR and then provides nutritional information such as the calories and protein content for each food. The principle of text recognition for menu recognition is the adoption of a CRNN (convolutional recurrent neural network) [13] for the feature extraction and recognition of text. The authors of [14] proposed a direct recognition technology for the text of the expiration date (or use-by date) to confirm safety information.

In particular, the recognition accuracy was improved by simultaneously applying an FCN [15] and CRNN [16] deep learning network. Although such deep learning-based OCR research is widely conducted, when various words and sentences are mixed in a product, it is still difficult for consumers to confirm whether the recognized character string is the information they originally want. That is, consumers may have the inconvenience of photographing printed characters several times in order to identify desired information.

Another representative field in which deep learning-based image analysis technology for food recognition is used is in the detection of foreign bodies in food and quality inspection. The authors of [17] introduced an efficient detection system which automatically learned foreign bodies in almond and green onion flakes and then analyzed the detection performance using the U-Net [18] network. The authors of [19] proposed a technique to detect and classify worms in stored grains using Faster R-CNN [20] and improved the detection performance through network adjustment and optimization. Although the above studies of food inspection and evaluation easily identify the type of food based on the captured image, they do not directly provide identification and safety information for products with unique packaging designs. However, to the best of our knowledge, there has been no research case on directly recognizing food packaging using deep learning with mobile devices.

Accordingly, this study proposes a system that provides safety information in real time by directly recognizing images of packaged food while satisfying the three requirements mentioned above. The system introduces a novel mobile recognition algorithm which allows users to easily capture product packaging images with a smartphone and simultaneously check product names (identification) and related safety information through deep learning technology. In addition, a novel usability test in terms of recognition rate, recognition speed, and a SUS-based questionnaire is also performed for an accurate performance evaluation from the individual point of view. These two outcomes of technological excellence and consumer satisfaction promote the consumption of safe food and also improve the sustainability of individual health.

The rest of this paper is organized as follows: Section 2 describes the architecture of the image-based food recognition system on smartphones. Section 3 analyzes and verifies the system performance via experiments. Section 4 presents a new SUS (system usability scale) questionnaire design for usability testing and analyzed the results in detail. Section 5 discusses the limitations of the proposed system, solutions to overcome them, and the sustainability of smart food consumption. Finally, Section 6 concludes the paper with a brief summary and future work.

## 2. Image-Based Food Recognition System

### *2.1. System Architecture*

The package recognition system for commercial processed food consists of a mobile phone (Samsung Inc., GalaxyA23, Suwon-si, Republic of Korea), with an Android OS 8.0 (Google Inc., Los Angeles, CA, USA) a gateway server, and an AI server; the overall architecture is shown in Figure 1. The mobile phone includes an AI-based application (app) which is designed to actually recognize the target food product by photographing, transmitting, and then receiving information. In general, since most smartphones provide camera modules, the user can easily take pictures using them without additional devices. After capturing the image of a food package, the image is transmitted to the gateway server through a wireless network (e.g., Wi-Fi, LTE [21]) and the gateway server verifies that the received image has no abnormalities before it forwards it to the AI server for classification and recognition. The AI server is equipped with an actual deep learning neural network (convolution layer) and detection algorithm. Before being inputted into the neural network, the image data received by the AI server are resized to 640 $\times$ 360 pixels. This is because cameras mounted on mobile phones generally have a 16:9 or similar aspect ratio. As the resized image passes through the deep learning network, features are extracted and learned. Then, if the recognition result is derived through the deep neural network, the ID of the

product is returned to the gateway server as a query response. The gateway server searches the separately built database (DB) in order to extract the matching product name and it finally transmits to the smartphone the detailed information (e.g., ingredient list, nutrition facts, recall reports, etc.).

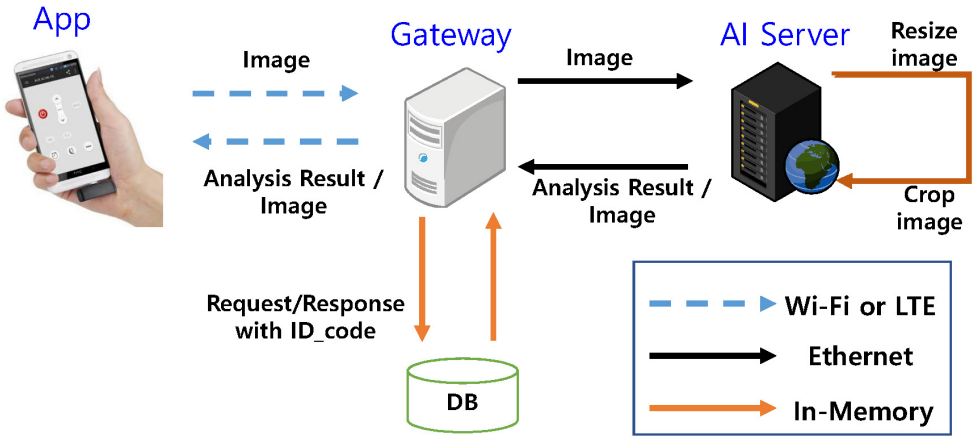



**Figure 1.** The architecture of image recognition.

### *2.2. Mobile App Architecture*

The main purpose of the mobile app is to help consumers take pictures of desired products and to receive safety information about those products as a result. The configuration of the app GUI design is shown in Figure 2, and the sequence of operations from the user's point of view is as follows: First, by clicking the camera symbol button indicated by the red square box in the figure, the user enters the camera mode for taking pictures. The app utilizes a guideline to highlight the objects by removing unnecessary background images, as shown in the second image in Figure 2. Afterwards, if the user takes a picture of the front of the food packaging, only the guided area is extracted and transmitted to the AI server through the gateway server. The app shows the product information received by the server and whether it is safe to consume. In addition, when users click on the representative image, they can also check nutritional and ingredient information provided by the manufacturer.

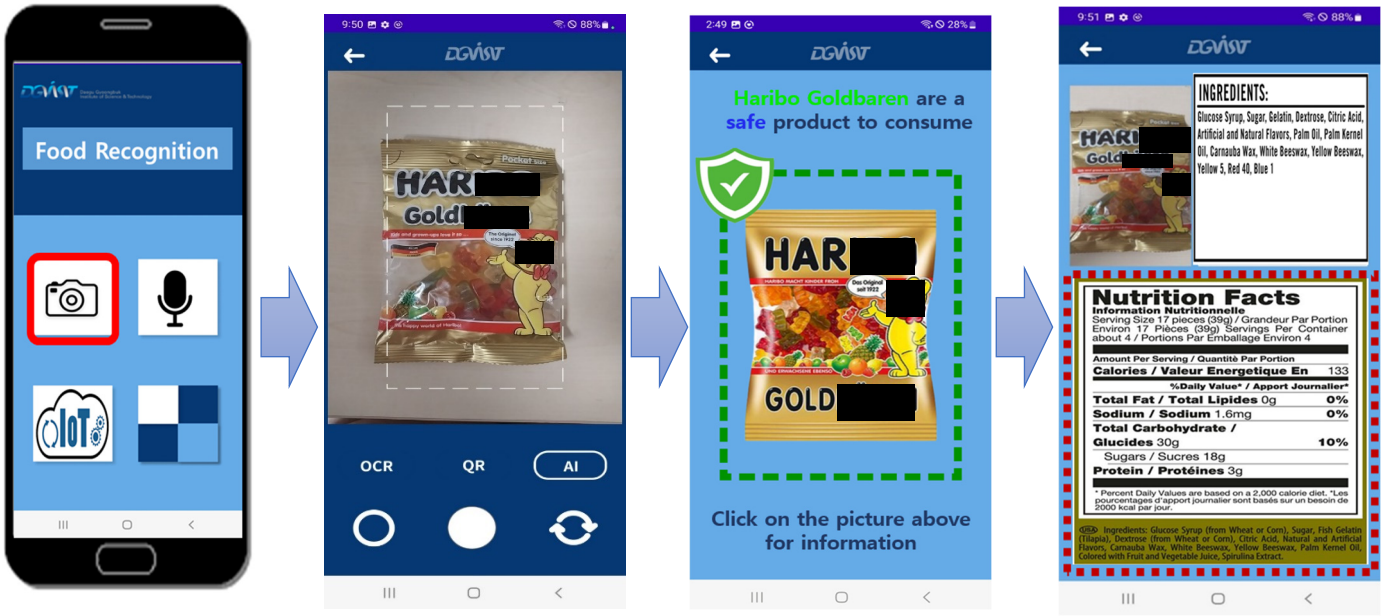



**Figure 2.** Query process via mobile app.

If the queried food is unsafe (e.g., unsuitable or subject to recall), the user is notified of the risk through a red-colored message, as shown in Figure 3. Although unsuitability or recall information is confirmed in real time on the FDA website, the proposed system assumes that the above information has been secured in advance and registered on the AI server.




**Figure 3.** Query result for unsafe food.

### *2.3. Data Collection and Detection Alogirthm*

A total of 100 kinds of food beverages and dietary supplements were selected for package recognition, and the selection criteria were determined by the products which were most imported and consumed in Korea from 2021 to 2023. Table 1 summarizes the brand names and countries of origin for all selected products. In order for the selected products to be classified through deep learning, 800 photos were taken for each product using various smartphone cameras (e.g., iPhone, Galaxy, Xiaomi, etc.), and a total of 80,000 datasets were created for 100 kinds of products.

**Table 1.** List of food products for recognition.

| Brand Name | Country of Origin | Brand Name | Country of Origin |
|---|---|---|---|
| Yellow Tail Shiraz | Australia | Kuri Yoshiwara Youkan | Japan |
| Kirkland Signature Dried Plums | USA | Tok Jelly Strawberry | China |
| Sunview Organic Raisins | USA | Mini Mentos New Rainbow | Vietnam |
| Chaokoh UHT Coconut Milk | Tai | Haribo Worms Sour | Turkey |
| Aroy-D Coconut Milk | Tai | Trolli Sour Glowworms | Germany |
| Crispconut Coconut Chips | Tai | Trolli Berries | Germany |
| Hunt's Tomato Paste | USA | Trolli Tarantula Glowworms | Germany |
| Fitted Black Olive | Spain | Haribo Happy Grapes | Turkey |
| Fitted Green Olive | Spain | Snickers Minis | USA |
| Starbucks Caffe Verona | USA | M&M's Fun Size | USA |

**Table 1.** *Cont.*

| Brand Name | Country of Origin | Brand Name | Country of Origin |
|---|---|---|---|
| Starbucks Breakfast Blend | USA | Lotus Original Belgian Waffle with Chocolate | Belgium |
| Nespresso Ispirazione Firenze Arpeggio | Swiss | Kinder Bueno | Italy |
| Dukes Blend | Australia | Twix | Netherlands |
| Pringles Original | Malaysia | Twix Minis | Germany |
| Oatmeal Mini Bite | Vietnam | Stella Artois 5% | Germany |
| Gemez Enaak Extra | Indonesia | Ferrero Rocher | Italy |
| Purple Sweet Potato Chip | Vietnam | Heineken Original | Netherlands |
| Pringles Potato Crisps Sour Cream and Onion | Malaysia | Kirkland Signature Protein Bar Variety Pack | USA |
| Original Premium Saltine Crackers | Canada | Tsingtao Beer | China |
| Sweetened Banana Chips | Philippines | Guinness Draught | Island |
| Savi Chip | Bulgaria | San Miguel Pale Pilsen Beer | Hong Kong |
| Lotus Biscoff | Belgium | Hoegaarden | Belgium |
| Ricos Round Nacho Chips | USA | Hoegaarden Rose 3% | Belgium |
| Potato Chip Sour Cream and Onion | Malaysia | Heineken Non-Alcoholic | Netherlands |
| St Michel Palets | France | Mogu Mogu Peach Flavored Drink | Tai |
| Pringles Potato Crisps Hot and Spicy | Malaysia | Power O2 Orange Lemon Flavor | Germany |
| Kid-o Creamy Butter Cracker Sandwich | Tai | Mogu Mogu Yogurt Flavored Drink | Tai |
| Shultz Pretzels Mini Twists | USA | Sponsor | Tai |
| Kirkland Signature Ruby Red Grapefruit Cocktail Gery Cheese Crackers | USA | Kirkland Signature Organic Unsweetended Almond Non-Dairy Beverage Vanilla | USA |
| Oishi Prawn Crackers Spicy Flavor | Philippines | Mogu Mogu Lychee Juice 25% | Tai |
| Uncle Pop Stick Barley Snack | China | Gery Cheese Crackers | Indonesia |
| Lago Vanilla Wafers | Italy | Tasco Young Coconut Juice with Pulp | Tai |
| Lago Cacao Wafers | Italy | Mogu Mogu Pineapple | Tai |
| Vanilla Wafer Roll | Indonesia | Mi Goreng Stir Fry Noodles | Indonesia |
| Choco Wafer Roll | Indonesia | Hao Hao Instant Noodles | Vietnam |
| Cheese Wafer Roll | Indonesia | Tonkotsu Ramen | Vietnam |
| Goldbaren | Germany | Mi Lau Thai | Vietnam |
| Mello Big Marshmallow | Philippines | Instant Noodles Shrimp Creamy Tom Yum Flavor | Tai |
| Starmix | Germany | Nescafe Dolce Gusto Cold Brew | England |
| I Alpha Candy C | China | Alfa Cafe Nutra Signature Blend | USA |
| Fruity-bussi | Germany | Coffee G7-3in1 | Vietnam |
| Gingerbon | Indonesia | Bios Life C Plus | USA |

**Table 1.** *Cont.*

| Brand Name | Country of Origin | Brand Name | Country of Origin |
|---|---|---|---|
| Belgian Coffee Sweets | Belgium | Men's Energy Pack | USA |
| Tok Jelly Peach | China | Healthpak | USA |
| Tok Jelly Green Grape | China | Fitline Activize | Germany |
| Happy Cola | Germany | Cal Mag D | USA |
| Jelly Straws | Taiwan | Double X Refill | USA |
| Fruittella Yogurt | China | Yeast B | USA |
| Tootsie Pops Miniatures | USA | Nutrikids Protein (Berry) | USA |
| Hitokuchi Neri Youkan | Japan | Alive! Once Daily | USA |

When constructing a dataset for training, it is important that it consists of photos taken under various external conditions. Although the app provides guidelines indicating the area to include objects (first photo in Figure 3), the quality of the image varies significantly depending on the camera angle, type of lighting, the area of space occupied by the object in the image, etc. Since these conditions may result in performance degradations of neural networks, the dataset was carefully constructed to reflect the above shooting conditions, as shown in Figure 4. Figure 4a depicts the image acquisition method according to two positions of lighting sources. First, when the light is placed directly above the product, the light is obscured by the camera or user's hand, and then a shadow is formed in the photo. In the second case, there is no shadow in the photo because the product is facing the wall. Figure 4b shows cases where the lighting color is white, yellow, or there is no lighting, and the photos were taken using a smartphone flash without external lighting. Figure 4c shows cases where the smartphone was held horizontally (0 degrees) to the subject or positioned at an angle of approximately 0 to $\pm$30 degrees. Figure 4d shows three cases of subject shooting areas. The first case is when the subject is located exactly within the guideline, the second is when the subject is slightly smaller than the guideline area, and the third is when the subject is taken larger than the guideline area. Figure 5 shows examples of raw images captured according to the above combination of external conditions. These datasets ensure the accuracy of recognition regardless of conditions of the consumer's shooting manner.

Using the collected dataset, a classification model was created by fine-tuning the YOLOv7 network [22]. The YOLO network consists of several layers with different structures and features. Among them, the YOLOv7 proposes a trainable bag-of-freebies method which improves accuracy without increasing information cost during training. In the training process of deep learning models, it is necessary to mitigate the phenomenon of overconfidence in the predicted results by obtaining soft labels. The previous YOLO version often involved having two separate heads for independent calculations. However, in YOLOv7, a feature is introduced where both heads are guided together. This multi-task mechanism achieves higher accuracy without affecting inference time in a single training session.

While the legacy CNN-based model was sufficient to train the data, there was a problem of decreasing parameter utilization when stacked infinitely close. YOLOv7 addresses this issue by introducing the E-ELAN (extended efficient layer aggregation network) architecture, which allows for the flexible stacking and efficient control (e.g., expand, shuffle, merge cardinality) of convolution layers. However, due to concatenation, the E-ELAN does not separate width and depth effectively. To further optimize and reduce computational complexity, a compound scaling method is proposed, as depicted in Figure 6. The compound model scaling approach demonstrates further optimization potential and confirms a structure where width and depth are adjusted according to the consistency of the concatenation-based model. With these features, YOLOv7 is well suited for the study and fine-tuning is performed on the E6E model, which is the deepest model in the YOLOv7.

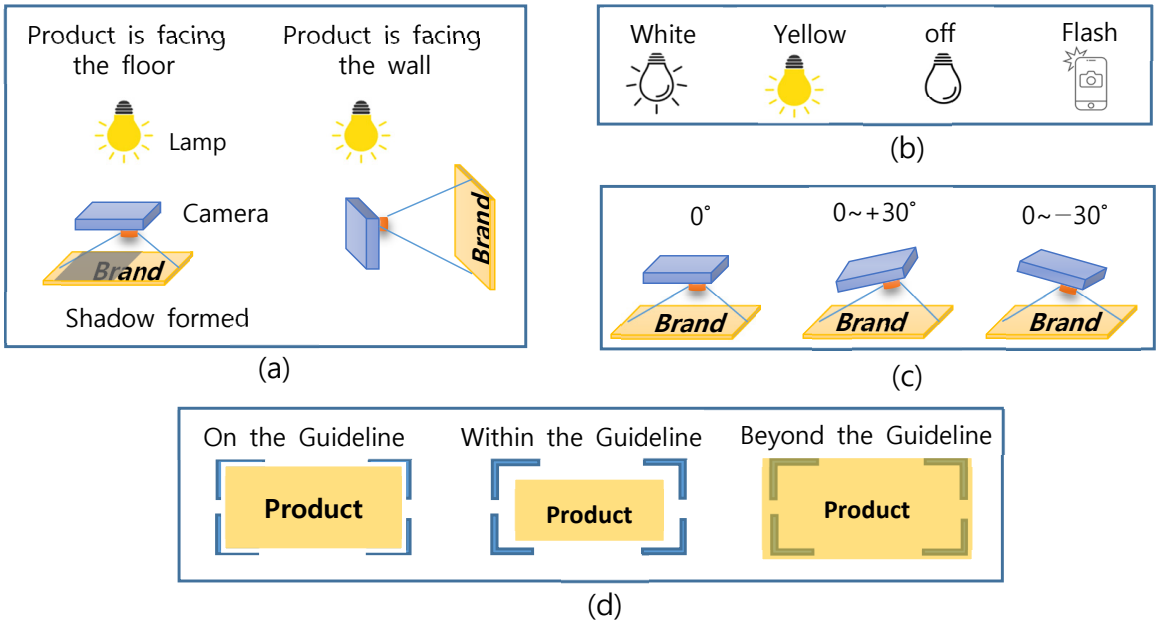



**Figure 4.** Various shooting methods: (**a**) directions of light; (**b**) types of lamps; (**c**) angles of shooting; (**d**) subject locations.



**Figure 5.** Examples of taken photos.

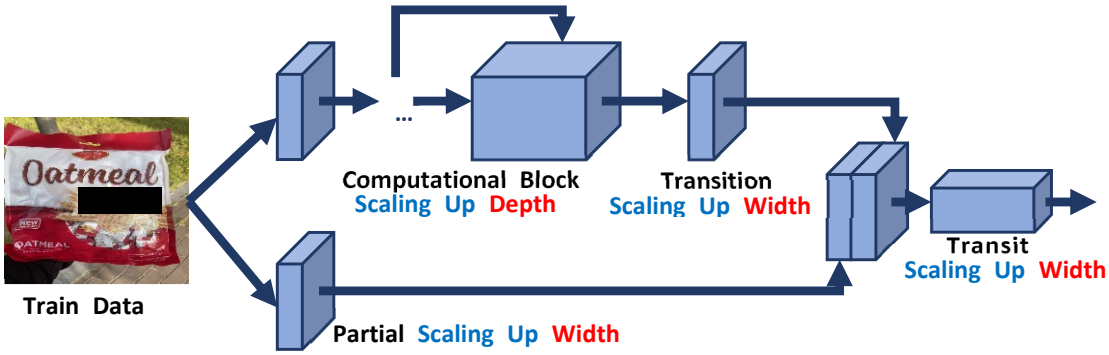



**Figure 6.** Compound scaling up depth and width for concatenation-based model [23].

The processing method for food package detection utilizing the model created by fine-tuning the E6E model is described in Figure 7, depicting when the system receives an image through the app from multiple users. For data preprocessing, the image size is converted to 640 × 360 pixels before passing through the model. The resized image data then passes through the YOLOv7 network generated using the collected dataset and the resulting output is obtained. The YOLO operation consists of three main parts, namely

the backbone, neck, and head. The backbone is a convolutional neural network which gathers image pixels to form features at various scales, and it is a pretrained model with the collected dataset. The FPN (feature pyramid network) becomes the neck of the network, where it combines and mixes the ConvNet layer representations before the data are passed to the prediction head. The head is part of the network that performs bounding box and class predictions with YOLO loss functions such as cross-entropy loss for classes, L1 loss for boxes, and objectness loss for objectness.

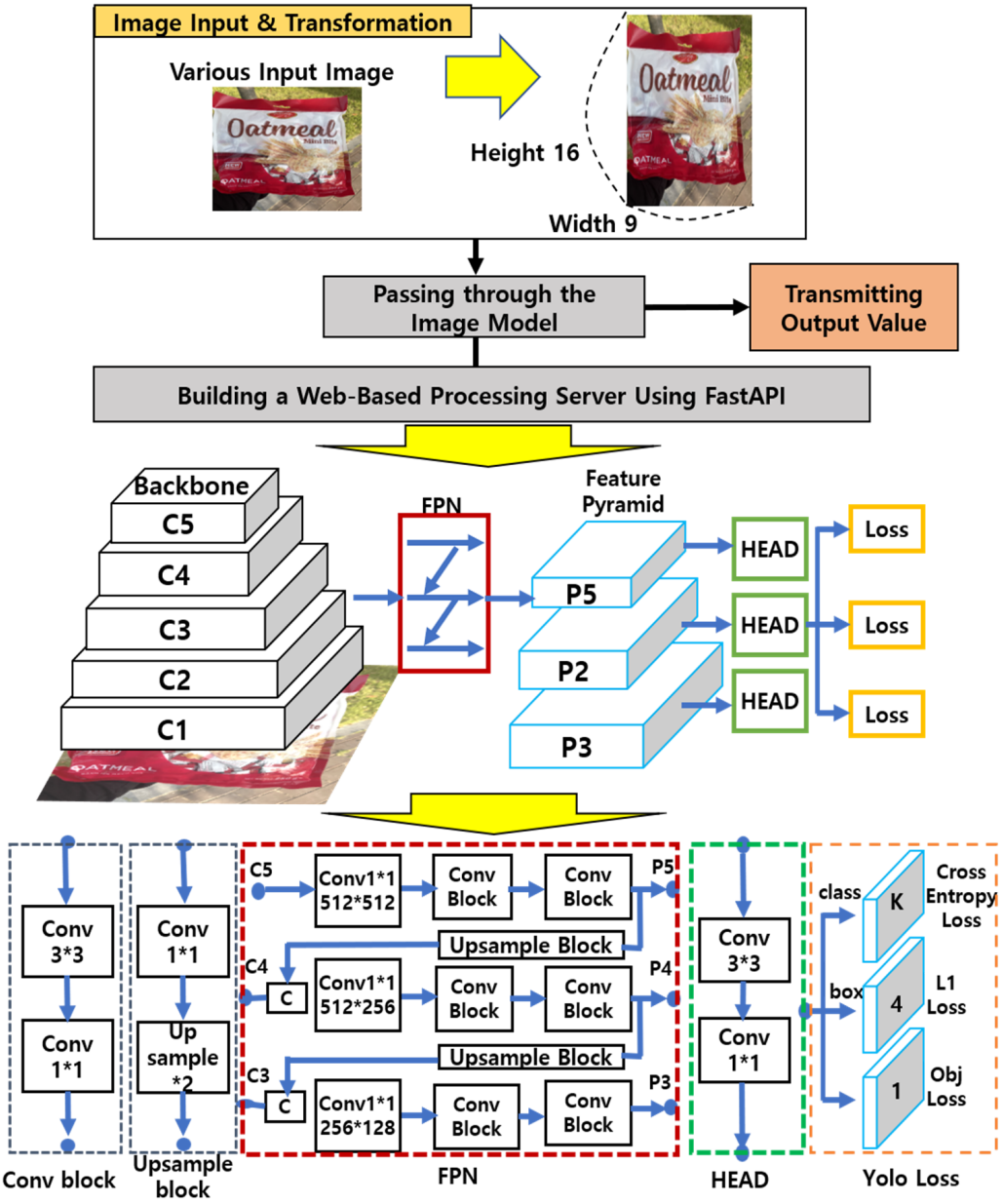



**Figure 7.** Real-time food classification processing system using YOLOv7.

Since the proposed system needs to support a large number of users, we also propose a web-based server architecture. For this purpose, we developed a processing server using FastAPI [24] and HTTP, as depicted in Figure 8. This server generates JSON-formatted dynamic data transmitted through various apps accessed by multiple users. The designed system prioritizes real time processing by placing the interaction with the learning models at the topmost layer, simplifying communication and interaction.

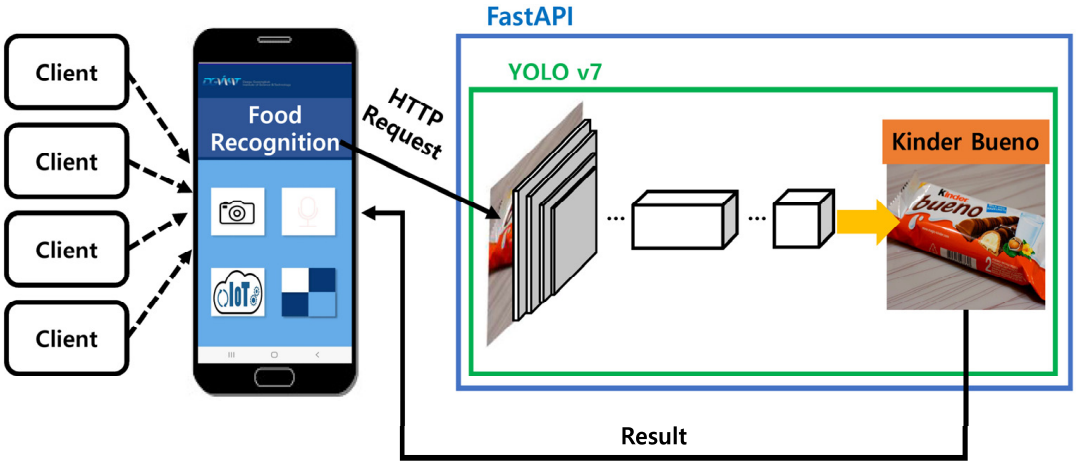



**Figure 8.** FastAPI-based web server interface structure.

## 3. Performance Evaluation

The evaluation of the trained model was conducted using the mAP (mean average precision), where AP represents the performance of recognition algorithms as a single value by combining the values of precision and recall. The calculation methods for precision and recall are expressed in Equations (1) and (2) below, where TP, FN, and FP represent true positive, false negative, and false positive values in the confusion matrix of the model. Precision, as shown in Equation (1), refers to the ratio of true positives among the items predicted as true by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{TP}}{\text{All Detections}} \tag{1}$$

Recall, as shown in Equation (2), refers to the ratio of items predicted as true by the model among those that are actually true.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{All Ground Truths}} \tag{2}$$

The values of TP and FP used for precision and recall can only be determined by calculating the IoU (intersection over union), which measures the intersection between the predicted result and the ground truth, as illustrated in Figure 9.

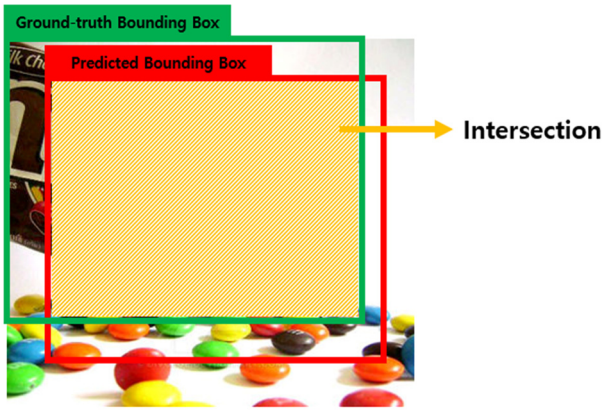



**Figure 9.** IoU definition.

The IoU of images passing through the trained model is calculated as shown in Equation (3). It measures the area of overlap between the predicted bounding box and the ground truth bounding box and then divides this area by the area of their union. In this

equation, $B$ represents the bounding box, $gt$ denotes the ground truth, and $p$ denotes the predicted box.

$$\text{IoU} = \frac{\text{Area}\left(\text{B}_{\text{gt}} \cap \text{B}_{\text{p}}\right)}{\text{Area}\left(\text{B}_{\text{gt}} \cup \text{B}_{\text{p}}\right)} \tag{3}$$

We trained a model using the YOLOv7 architecture, specifically the E6E variant, with a dataset containing 100 kinds of food items. The gathered dataset of 80,000 images is divided into three categories, as shown in Figure 10. The ratios of training, validation, and test data are 70%, 25%, and 5% and consist of 56,000, 20,000, and 4000 photos, respectively.

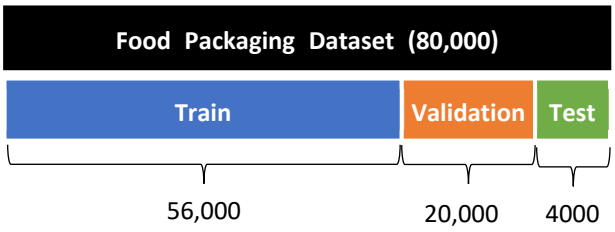



**Figure 10.** Data split for training, validation, and testing.

The model was trained with parameters set as follows: image size of 640 × 640, batch size of 32, starting learning rate of 0.01, ending learning rate of 0.2, two iterations, momentum of 0.937, and training for 300 epochs. Table 2 provides detailed descriptions of these parameter values, which dynamically affect the overall performance of deep neural networks. However, the settings below are not always suitable for all cases and may vary depending on various external factors such as the type of food, capturing conditions, and image quality.

**Table 2.** List of parameters with descriptions.

| Parameter | Value | Description |
|---|---|---|
| Epoch | 300 | Although a large number of epochs helps the model to learn detailed patterns, too high of an epoch value can cause overfitting. In general, 300 epochs is considered as optimal, balancing learning and overfitting prevention [25,26]. Figure 11 also shows that performance becomes stable as epochs approach 300. |
| Learning Rate | 0.01 ~0.2 | Training starts with a low learning rate (0.01) for stable initial learning and to prevent drastic changes in the model. As training progresses, the learning rate gradually increases (0.2) to accelerate convergence. This approach, known as "Learning Rate Scheduling" or "Learning Rate Warm-up" [27], prevents early instability and promotes faster convergence. |
| Batch Size | 32 | A batch size of 32 is set to balance memory usage and training speed. Although small batch sizes reduce memory usage and allow the model to update weights more frequently, too small of a value can make the network unstable. In our repeated experiments with 2 RTX 4090 D6X 24GB GPUs, a batch size of 32 is optimal. |
| Iteration | 2 | The value of iterations refers to the process where one batch is input into the model and a single weight update occurs. Since the batch size is set to 32 and the dataset is sufficiently large, we set the number of iterations low for fast learning. That is, we made the model learn faster while also learning the characteristics of the data appropriately. |
| Image Size | 640 | The 640 × 640 resolution is recommended by YOLOv7, as it can include sufficient detail while maintaining computational efficiency. If a larger resolution is used, more details can be learned. However, the computational cost also increases. Thus, training with 640 × 640 images achieves an optimal balance of memory and computational costs for the proposed system. |
| Momentum | 0.937 | Momentum accelerates the speed of gradient descent by reflecting the speed of the previous learning step while helping to converge steadily along the gradient. The optimal momentum simultaneously improves the learning speed and performance of the model.<br>The set value of 0.937 is slightly higher than the commonly used 0.9, which provides an appropriate balance to reduce fluctuations during training and enables faster convergence. This means that information from previous learning steps has a strong influence on the weight update of the current feature. |

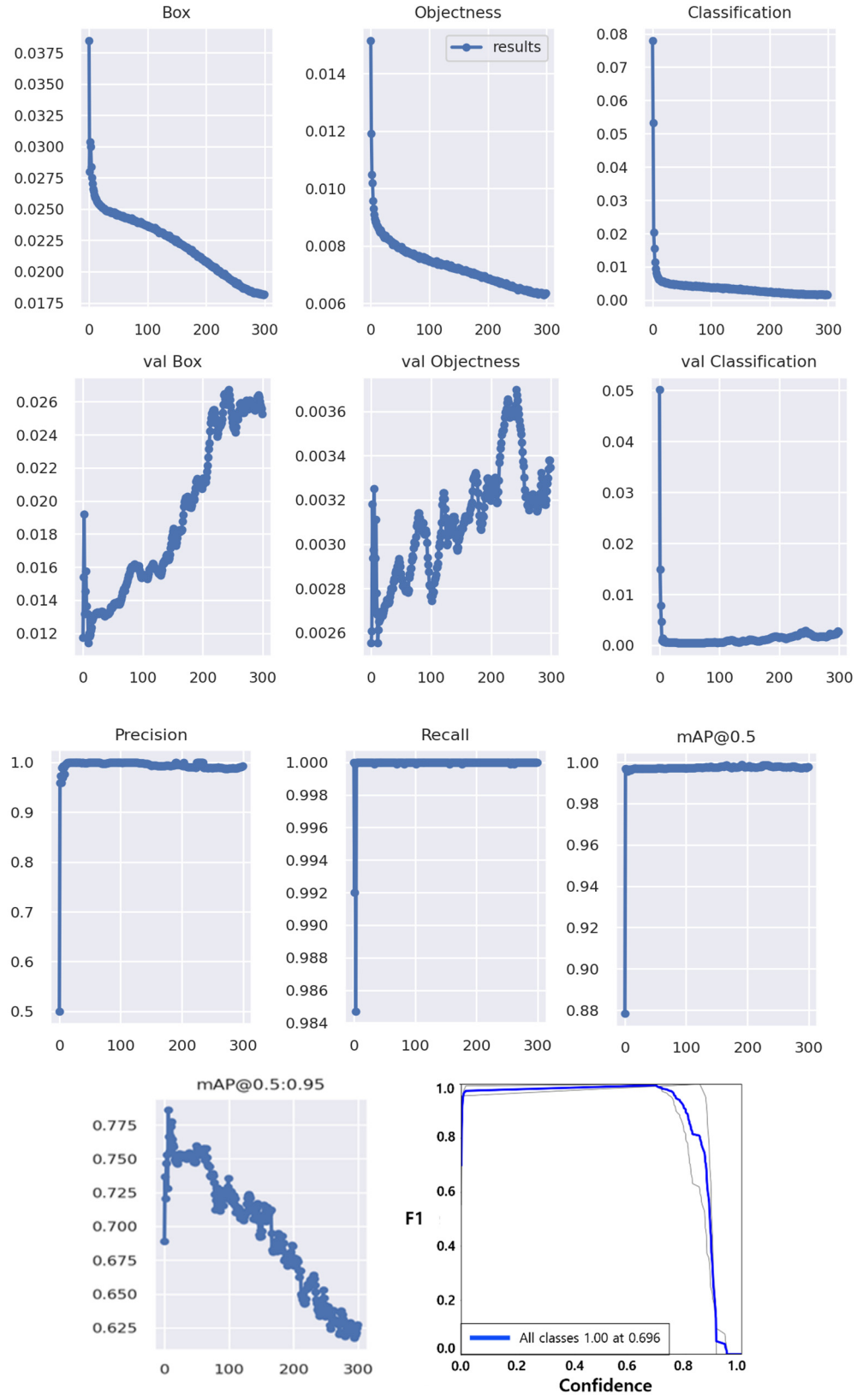



**Figure 11.** Results of training.

The testing of the trained model is depicted in Figure 11. The X-axis of the overall graph represents the number of epochs, while the Y-axis indicates the loss values. The test results of the trained model showed that both precision and recall were close to one,

confirming its very high performance. The box loss indicates the error in bounding box predictions and decreases as training progresses, suggesting that the model can predict the object's location more accurately. The objectness loss represents the performance of object detection, and it also decreases, indicating that the model can detect objects more accurately. The classification loss indicates the error in object classification, and it gradually decreased during training, reflecting an improvement in the model's classification performance. The mAP at 0.5 refers to the mean average precision at a 0.5 IoU threshold, and it maintained a high value during testing. On the other hand, the mAP at 0.5:0.95 value measures performance across various IoU thresholds and shows a slight decrease in the latter part, suggesting the possibility of overfitting. Therefore, we generated the model with the best performance among those trained via early stopping. The last graph shown in Figure 11 displays the model's F1 score. As the confidence increases, the reliability of predictions improves, but this does not guarantee an enhancement in quality. In addition, the closer the average of precision and recall is to one, the better the model's performance. The blue curve plotted on the graph indicates that when the confidence is 0.696, the F1 score reaches one, signifying optimal model performance. And we can see that after an initial rise, it reaches a peak and then sharply declines at excessively high confidence levels. The black curve provides a performance comparison, showing a tendency for the F1 score to decrease more rapidly as confidence increases. Consequently, it is confirmed that the model's performance is best at a confidence level of 0.696, while a confidence level above 0.8 is associated with a significant performance degradation.

The performance of the best-trained model was measured and presented as percentages in Table 3. Evaluated on 20,000 images across 100 classes, the results showed a precision of 99.23%, recall of 100%, and an F1 score of 99.46%. Additionally, the mAP at 0.5 was 99.62% and the mAP at 0.5:0.95 was confirmed to be 78.59%, and thus the optimal model was obtained.

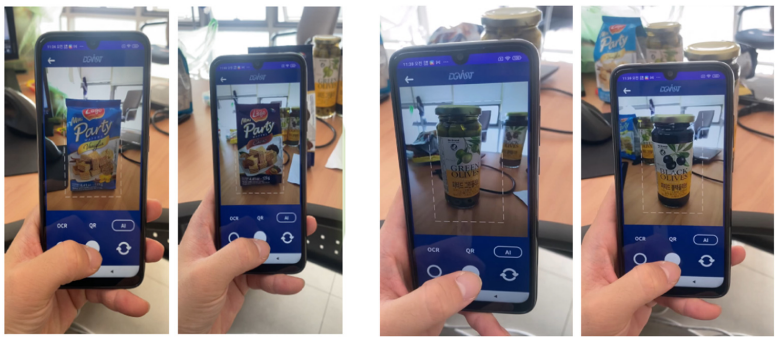
In addition to performance verification using the gathered dataset, we also conducted tests on images captured in real time using a smartphone. The detection accuracy was measured by capturing images of each product 20 times. As a result, the system correctly identified 1960 of a total of 2000 tests, which confirmed that the images were accurately classified with 98% accuracy even though they were taken from various angles and backgrounds. Figure 12 below illustrates the results derived from the model after receiving the image data, showcasing the detection outcomes of three randomly selected products out of the total 100 food items. According to the detection results, crispy coconut chips were recognized with a 90% confidence score, Mi Goreng ramen with 87%, and cocoa waffles with 80%. Moreover, the proposed system also has excellent classification performance for products with similar shapes and colors. Figure 13a successfully distinguishes between Lago's vanilla waffles and cocoa waffles and Figure 13b distinguishes between Fitted's green olives and black olives. This will also make it possible to determine the authenticity of a product.




**Figure 12.** Random image test.

**Table 3.** Evaluation of the optimal performance model.

| Precision | Recall | F1 Score | mAP@0.5 | mAP@0.5:0.9 |
|---|---|---|---|---|
| 99.23% | 100% | 99.46% | 99.62% | 78.59% |



(a) (b)

**Figure 13.** Similar image test: (**a**) vanilla wafers vs. cacao wafers; (**b**) green olives vs. black olives.

## 4. Consumer Usability Test

### *4.1. Test Participants*

Evaluation participants were recruited from college students who were familiar with using smart phones and mobile applications and consumers who were interested in safety information about international foods. A total of 71 volunteers participated, including 32 men and 39 women. The age composition was as follows: 40 people (56.3%) in their 20s, 13 people (18.3%) in their 30s, three people (4.2%) in their 40s, eight people (11.3%) in their 50s, and seven people (9.9%) were over 60 years old. The final educational level was as follows: 11 people (15.5%) had a high school (HS) diploma or less, 34 people (47.9%) were either college students or had a bachelor's (BS) degree, and 26 people (36.6%) were either graduate students or had master's (MS)/Ph.D degrees.

### *4.2. Test Methods*

Although the efficiency of performance was demonstrated in terms of detection accuracy and recall in Section 3, this section conducts a subjective evaluation of the proposed system from the consumer's perspective. The following two metrics were defined to verify how quickly and exactly users obtain the desired information:

- Completion rate: The number of correctly identified food products was compared to the total number of attempted trials using two tools (i.e., Internet/QR codes, smartphone), which is expressed by Equation (4). There are two types of tasks, checking the food information (name, ingredients, etc.) and checking for recall. First, the "Check food information" task involved randomly selecting 50 out of 100 food products and searching the information via each tool (Internet/QR codes, proposed app). Since 71 people attempted 50 times, the total number of trials for each method was 1350. The second task, "Check for recall" also involved randomly selecting 50 out of 100 products and searching recall information via each tool. The total number of attempts for this task was 1350 for each tool, which is same as the first task.

$$\text{Completion Rate} = \frac{\text{Number of completed trials}}{\text{Total number of attempted trials}} \times 100 \quad (4)$$

- Task time: The total time required for a consumer to receive query results using three methods (Internet/QR codes, proposed app) was determined. This included system response time (e.g., app launch time, photo capture time, wireless communication delay, etc.) and user confirmation time. Although system response time was measured accurately by a system clock, user confirmation time included the time to read the

displayed text information and determine whether the information was what the user wanted. This delay varied for each individual, but it was measured objectively using a stopwatch. This experiment was also conducted with two tasks, which is same as the completion rate. And the number of attempts for each task was also the same, at 1350.

Among the tasks to check for products, the method via the app was performed as shown in Figures 2 and 3. As a comparison group, the tasks via the Internet/QR codes were performed using portal sites (e.g., Google, Bing, Yahoo, etc.) or by directly accessing the websites of manufacturers or the FDA. All food products were placed on the table, as shown in Figure 14a, and each participant compared the performance of the proposed app and the legacy web browser using a smartphone, as shown in Figure 13b.



(a) (b)

**Figure 14.** Conducting usability tests: (**a**) samples of food products; (**b**) test via smartphone.

In addition to the above two metrics, the SUS (system usability scale) [28] model was adopted for in-depth qualitative evaluation of the usability of the proposed system. Although the original SUS model consisted of 10 questions, in this study, the scope of questions was redesigned and expanded to 16 questions based on convenience, accuracy, and usefulness, as shown in Table 4. The score of the question was given one to five points based on the Likert scale [29], and each score indicated "Strongly disagree", "Disagree", "Neutral", "Agree", and "Strongly agree", respectively. For all questions, odd numbers were positive questions and even numbers were negative questions. To normalize the obtained scores to 0 to 100, we subtracted one from the positive question score and subtract five from the negative question score, and then all scores were added together. Finally, the sum was multiplied by 2.5 to obtain the normalized score. According to the previous work [30], the average SUS score is 68 points, and a score of 70 or higher is generally considered to have good usability.

**Table 4.** Proposed SUS questionnaire with three categories.

| Category | Question | Score * |
|---|---|---|
| Convenience | 1. I thought this system was convenient to use. | 1 2 3 4 5 |
| | 2. I found this system took too much time to get response. | 1 2 3 4 5 |
| | 3. I thought this system was easy to use. | 1 2 3 4 5 |
| | 4. I think I would need assistance to be able to use this system. | 1 2 3 4 5 |
| Accuracy | 5. I found this system had good integrity. | 1 2 3 4 5 |
| | 6. I found it was difficult to recover from failure in operations. | 1 2 3 4 5 |
| | 7. I found this system provided consistent contents. | 1 2 3 4 5 |
| | 8. I found there were too many errors in this system. | 1 2 3 4 5 |
| Usefulness | 9. I think most people would quickly learn how to use this system. | 1 2 3 4 5 |
| | 10. I found this system provided insufficient contents. | 1 2 3 4 5 |
| | 11. I thought this system could be used without a manual. | 1 2 3 4 5 |
| | 12. I thought this system was unnecessary for daily life. | 1 2 3 4 5 |
| | 13. I thought this system was helpful for purchasing imported foods. | 1 2 3 4 5 |
| | 14. I thought this system was of no use to purchase safe foods. | 1 2 3 4 5 |
| | 15. I think I would like to use this system frequently. | 1 2 3 4 5 |
| | 16. I think I would discourage others from using this system. | 1 2 3 4 5 |

* 1: strongly disagree; 2: disagree; 3: neutral; 4: agree; 5: strongly agree.

### *4.3. Evaluation Results*

Firstly, the task completion rate of the legacy Internet (including QR codes) access method and the proposed app was compared in Table 5. The proposed system showed better performance in not only checking general food information (e.g., product name, ingredients) but also checking for recalls. In particular, with regard to checking the recalls, the completion rate of the proposed system was 96%, but the Internet access method was only 68%. The legacy Internet access method requires people to access the FDA website, find the recall information menu, and then manually input the product name and identify duplicate items themselves, as shown in Figure 15. This complicated process led to frequent task failures. Moreover, the product names of some imported foods could not be identified because they were written in foreign languages rather than the participants' native language or English. This makes it impossible to input the brand name to check for recalls on the FDA website. In addition, some products had QR codes which are inserted into the package, but most of them were linked to the manufacturer's website, so consumers still have to search for information on their own. On the other hand, the proposed system showed a high completion rate by automatically deriving relevant information when consumers merely took a picture of the product using the app.

**Table 5.** Comparison of task completion rate.

| Task | Via Internet or QR Code | Via Proposed |
|---|---|---|
| Check food information (50 trials) | 78% | 97% |
| Check for recalls (50 trials) | 68% | 96% |

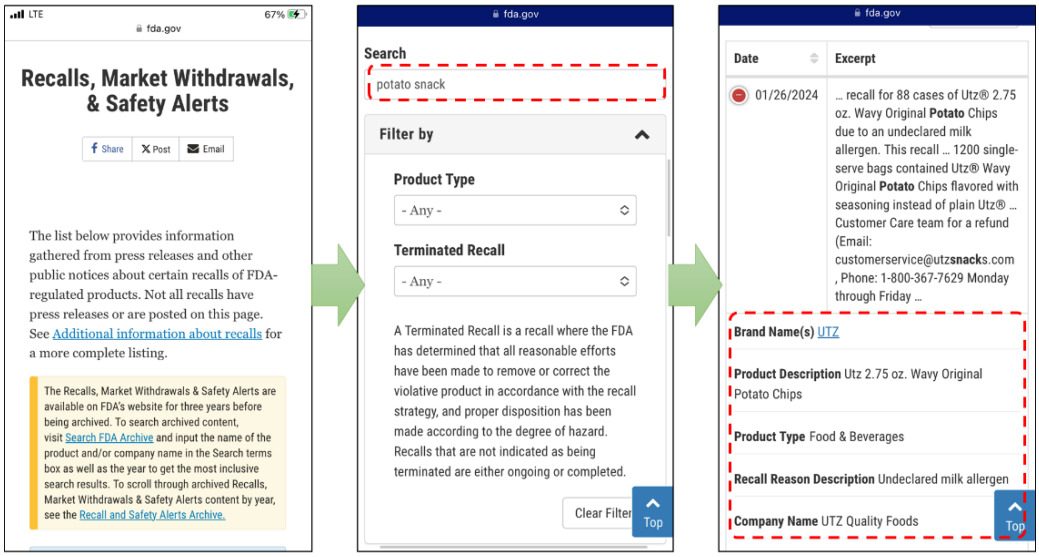



**Figure 15.** Checking for recalls via the FDA web pages.

Next, a comparison of the average task completion time for the tasks is shown in Table 6. Regarding checking general information and recall information, the legacy method took 16.4 s and 19.5 s, respectively, but the proposed system showed 5.1 s and 4.2 s, respectively. The proposed system mainly includes Wi-Fi connection delay and processing time within the server, but most of them are as fast as 5 s. The proposed system shows a relatively short time, even including Wi-Fi connection delay and server processing time, but the traditional method takes a significantly longer time because it goes through several steps such as launching the browser, accessing the related website (e.g., portal, FDA, manufacturer), and then typing the full product name. Meanwhile, the reason why recall confirmation via the proposed app is faster than checking general information is because the recall status is shown first and additional information is provided later, as shown in

Figure 3. Finally, it was revealed that the proposed system was excellent in both metrics of task completion rate and time.

**Table 6.** Comparison of Average task completion time.

| Task | Via Internet or QR Code | Via Proposed |
|---|---|---|
| Check food information (50 trials) | 16.4 s | 5.1 s |
| Check for recalls (50 trials) | 19.5 s | 4.2 s |

In addition to the evaluation of task completion, the extended SUS survey results are summarized in Table 7. The average score for the three sub-types of SUS questions (i.e., convenience, accuracy, and usefulness) were 4.15, 3.85, and 4.31 points, and their normalized scores were 83, 77, and 86 points, respectively. All three categories were judged to have good usability with respect to the average SUS score, exceeding 70 points. According to individual interviews after the questionnaire, it was revealed that the usability score was high by the fact that the system was simple and it was easy to check the country of origin, ingredients, and recall status of imported foods with a smartphone. However, it was found that satisfaction with accuracy was somewhat low due to network delays and interruptions that occurred, especially when transmitting large amounts of image data.

**Table 7.** Results of SUS questionnaire.

| Category | Question | Scores (Frequency $n$ = 71) | | | | | Avg | Dev | SUS Score(Norm) |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | | | |
| Convenience | 1 | 0 | 4 | 7 | 30 | 30 | 4.2 | 0.8 | 83 |
| | 2 | 0 | 15 | 24 | 16 | 30 | 3.4 | 1.1 | |
| | 3 | 0 | 2 | 3 | 24 | 42 | 4.5 | 0.7 | |
| | 4 | 0 | 0 | 7 | 18 | 46 | 4.5 | 0.7 | |
| | | | | Subtotal | | | 4.15 | 0.82 | |
| Accuracy | 5 | 3 | 8 | 18 | 24 | 18 | 3.6 | 1.1 | 77 |
| | 6 | 1 | 4 | 24 | 23 | 19 | 3.8 | 1.0 | |
| | 7 | 4 | 2 | 12 | 28 | 25 | 4.0 | 1.1 | |
| | 8 | 1 | 5 | 13 | 25 | 27 | 4.0 | 1.0 | |
| | | | | Subtotal | | | 3.85 | 1.05 | |
| Usefulness | 9 | 0 | 8 | 14 | 15 | 24 | 3.9 | 1.0 | 86 |
| | 10 | 0 | 0 | 5 | 19 | 47 | 4.6 | 0.6 | |
| | 11 | 0 | 0 | 4 | 20 | 47 | 4.6 | 0.6 | |
| | 12 | 0 | 2 | 9 | 30 | 30 | 4.2 | 0.8 | |
| | 13 | 0 | 2 | 8 | 23 | 38 | 4.4 | 0.8 | |
| | 14 | 0 | 3 | 5 | 26 | 37 | 4.4 | 0.8 | |
| | 15 | 0 | 4 | 10 | 27 | 30 | 4.2 | 0.9 | |
| | 16 | 0 | 3 | 10 | 26 | 32 | 4.2 | 0.8 | |
| | | | | Subtotal | | | 4.31 | 0.78 | |

Meanwhile, in order to analyze the differences in individual tendencies of the participants, we also conducted an additional survey by classifying the participants in terms of age and education level. Regarding age, we divided the participants into the younger generation (20 s and 30 s) and the older generation (over 50s). In terms of educational level, we classified them into three categories, i.e., HS degree, BS degree, and graduate degrees (MS, Ph.D degree), respectively. The SUS results are shown in Figure 16. In Figure 16a, the convenience scores were the same for both age groups, but the junior group's accuracy score was 78 points, which was higher than the senior group's score of 75 points. This is because the junior group, who are skilled in operating smartphones, can solve problems on their own even if an unexpected error occurs, while the senior group is less familiar with the new digital interface and usage. However, in terms of system usability, the senior

group scored 89 points, which was higher than the junior group's score of 85 points. This is because young people can obtain safety information through alternative methods (e.g., Internet surfing) without using the proposed system. Figure 16b shows that the convenience, accuracy, and usefulness scores all decrease as educational level increases. This is because highly educated people tend to carefully analyze the limitations and vulnerabilities of the system and scored them based on their experienced knowledge. For example, in addition to the basic function of checking safety information, they requested improvements in limitations such as font size, resolution, and user experience (UX) scenarios and then gave low scores for these deficiencies. Nevertheless, the minimum SUS score exceeded 70 points, indicating that the system was still useful.

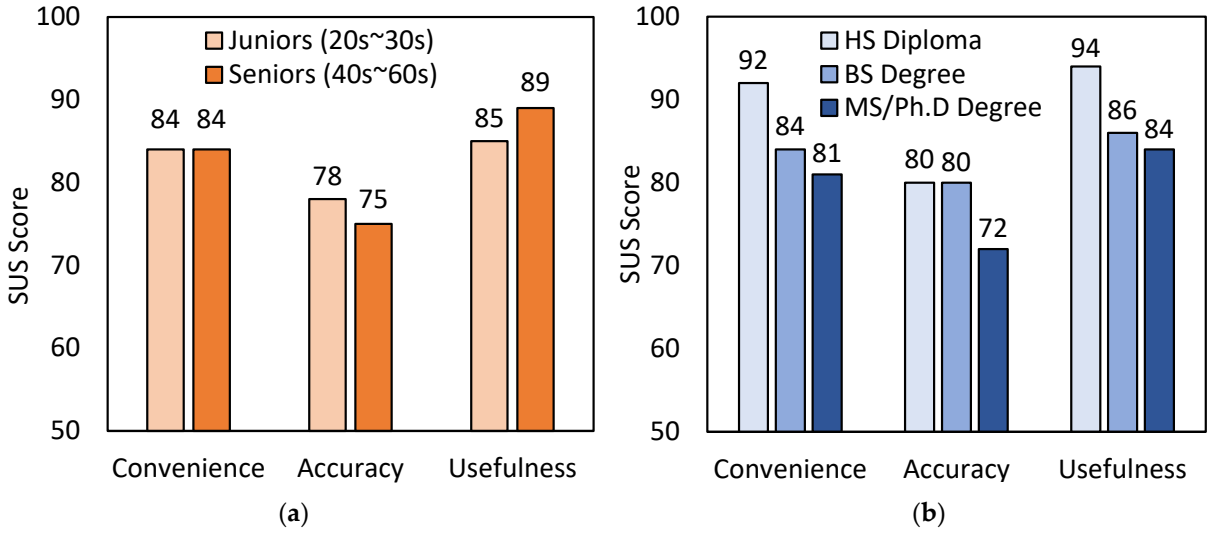



**Figure 16.** SUS scores with different ages and academic levels. (**a**) comparison between different ages; (**b**) comparison between different academic levels.

## 5. Limitations and Discussions

Despite the extensive usability test, it is important to note that this study also had some limitations as follows: first, this study involved only 71 participants, which somewhat limits generalizability. Second, the number of products tested in this study is only about 100, which is not enough to represent all kinds of imported foods. Accordingly, for future work, we will secure a dataset of more diverse types of foods and conduct long-term research targeting a larger number of general consumers. In this study, we gathered a total of 80,000 images, but most of the capturing and annotation tasks were performed manually, which took a lot of time. One way to secure a sufficient dataset is to obtain images directly from manufacturers when they pass through customs. This requires establishing standardized guidelines for image datasets, but it will ensure a consistent high-quality dataset in the long-term perspective.

The third issue is how many image datasets are needed for deep learning-based product recognition. In general, the performance of a deep learning network increases as the amount of training data increases. However, it is still difficult to generalize the performance correlation because there is a specific case that performance actually decreases when there is a lot of data containing errors or low-quality data. Nonetheless, in this study, we identified the minimum amount of image data, which guarantees performance based on the gathered dataset. Table 8 shows the change in accuracy performance according to the amount of image data trained by the deep learning model. Although these are the performance results before optimizing hyper-parameters, it is confirmed that at least 800 images were required to ensure a reasonable performance of over 90%.

**Table 8.** Accuracy results according to the amount of trained data.

| Metric | Number of Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **100** | **200** | **300** | **400** | **500** | **600** | **700** | **800** |
| Accuracy | 25% | 55% | 58% | 62% | 68% | 78% | 80% | 90% |

Fourth, we trained volunteers on how to use the app before conducting usability tests, but some people who were not familiar with the app failed to recognize the images due to poor shooting. Figure 17 shows successful and failed product recognition cases, respectively. Figure 17a shows a successful case where the object is positioned as close as possible to the white dashed guidelines provided by the app, while Figure 17b shows a failure case where the object is partially covered by a finger or captured too small. Similarly, if the ambient lighting is too low or the photo is blurred due to shadows from obstacles, it can also cause recognition failures. To overcome this, additional techniques such as noise-filtering and normalization will be required. Alternatively, there is a method to build an additional dataset that reflects these unstable conditions by the fact that the performance can be improved by learning more diverse images.

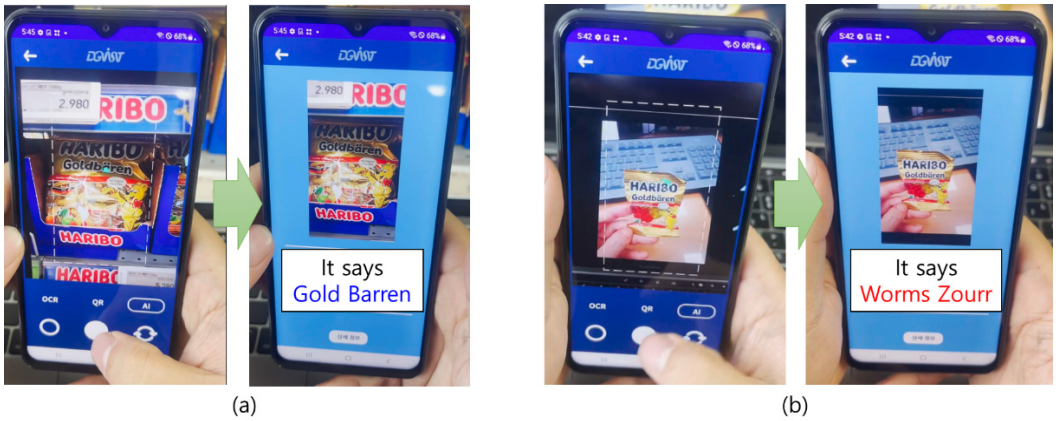



**Figure 17.** Comparison of recognition results: (**a**) success; (**b**) failure.

Finally, we discuss how the proposed system enables sustainable food consumption. Throughout the paper, we have explained consumers' sustainability based on convenience, accuracy, and usability. Despite some limitations, if the proposed system is established in the field, the following effects can be achieved:

First, from the perspective of food manufacturers and customs agencies, it can significantly strengthen the food supply chain by promoting sustainable purchase and consumption by consumers. This can improve the sustainability of the overall economic system and encourage further quality improvement and investment by companies.

Second, from the perspective of individual consumers, the proposed system can help ensure the safety of the foods they care about, thus contributing to sustainable healthier consumption. In addition, even if the ingredients of the food do not match their health conditions, accurate information provision increases trust in the food manufacturers.

Third, it helps people live more convenient and sustainable lives by improving conventional (e.g., barcodes or Internet searches) methods and leveraging new artificial intelligence technologies. That is, conventional manual inquiries can be replaced with a one-stop service, which ensures the sustainability of a smart life. This also helps food manufacturers provide more convenient and sustainable services to consumers and continuously develop high-end technologies.

In summary, the proposed system sustainably maintains the health of individual consumers and makes the ecosystem consisting of food production, distribution, and consumption more sustainable.

## 6. Conclusions

It is important for consumers to check whether food is safe before purchasing it. Although consumers can read the packaging label directly or use an Internet-based inquiry service, the inquiry process is still complex and difficult. Thus, firstly, we proposed a mobile food safety inquiry system which automatically recognizes food packaging images through a smartphone camera and transmits detailed information to consumers. The system has been optimized for YOLOv7 and can recognize 100 types of food by learning from a total of 80,000 images, which we personally captured and collected. We conducted a total of 2000 tests to verify the performance, showing a recognition accuracy of 98%.

Secondly, for more qualitative validation, we also designed and conducted a new SUS-based usability test by interviewing consumers of various age groups. A total of 71 volunteers were recruited, and evaluation was conducted by categorizing them by gender, education, and age. We expanded the existing SUS into three questions consisting of convenience, accuracy, and usefulness, and as a result, we confirmed that the proposed system was excellent overall, with usefulness scoring 86 points, convenience scoring 83 points, and accuracy scoring 77 points. Thus, if this system is distributed to the public, consumers will be able to easily check which foods are safe, which leads more sustainable and healthy lives.

For future work, we plan to expand the system to provide more than 100 kinds of product classification and develop the technique of improving the recognition performance even when food packaging is in poor condition, such as damaged or wrinkled packaging.

**Author Contributions:** Y.-D.K. and M.-K.P. carried out the main conceptual design of the proposed system and writing of the manuscript; J.-W.P. and Y.-H.C. mainly dealt with data analysis and experiments (i.e., configuration and realization of proposed system). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a grant (no. 21163MFDS518) from the Ministry of Food and Drug Safety of Korea.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the In-stitutional Review Board of Kyungpook National University (Protocol code: 2024-0330, Date of approval: 19 July 2024).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The dataset that support the findings of this study are openly available in Kaggle community at https://www.kaggle.com/datasets/parjunwoo/fooddatasert (accessed on 29 September 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. U.S Food & Drug Administration, Recalls, Market Withdrawals, & Safety Alerts. Available online: https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts (accessed on 1 April 2024).
2. The European Rapid Alert System for Food and Feed (RASFF). Available online: https://webgate.ec.europa.eu/rasff-window/screen/search (accessed on 1 April 2024).
3. Korea Food & Drug Safety Administration. Available online: https://www.foodsafetykorea.go.kr/main.do (accessed on 1 April 2024).
4. Rejeb, A.; Keogh, J.G.; Zailani, S.; Treiblmaier, H.; Rejeb, K. Blockchain technology in the food industry: A review of potentials, challenges and future research directions. *Logistics* **2020**, *4*, 27. [CrossRef]
5. Yu, Z.; Jung, D.; Park, S.; Hu, Y.; Huang, K.; Rasco, B.A.; Wang, S.; Ronholm, J.; Lu, X.; Chen, J. Smart traceability for food safety. *Crit. Rev. Food Sci. Nutr.* **2020**, *62*, 905–916. [CrossRef] [PubMed]
6. Kravenkit, S.; So-In, C. Blockchain-Based Traceability System for Product Recall. *IEEE Access* **2022**, *10*, 95132–95150. [CrossRef]
7. Konstantinos, R.; Aggeliki, K.; Dimitris, F.; Thomas, F.; Leonidas, H.; Christina, B. Evaluating the Use of QR Codes on Food Products. *Sustainability* **2022**, *14*, 4437. [CrossRef]
8. Duroc, Y. From Identification to Sensing: RFID Is One of the Key Technologies in the IoT Field. *Sensors* **2022**, *22*, 7523. [CrossRef]
9. Vedat, C.; Busra, O.; Kerem, O. The Survey on Near Field Communication. *Sensors* **2015**, *15*, 13348–13405. [CrossRef]

10. Tang, Q.; Lee, Y.; Jung, H. The Industrial Application of Artificial Intelligence-Based Optical Character Recognition in Modern Manufacturing Innovations. *Sustainability* **2024**, *16*, 2161. [CrossRef]
11. Peng, H.; Yu, J.; Nie, Y. Efficient Neural Network for Text Recognition in Natural Scenes Based on End-to-End Multi-Scale Attention Mechanism. *Electronics* **2023**, *12*, 1395. [CrossRef]
12. Ju, X.; Lo, F.P.W.; Qiu, J.; Shi, P.; Peng, J.; Lo, B. MenuAI: Restaurant Food Recommendation System via a Transformer-based Deep Learning Model. *arXiv* **2022**, arXiv:2210.08266a.
13. Keren, G.; Schuller, B. Convolutional RNN: An Enhanced Model for Extracting Features from Sequential Data. *arXiv* **2016**, arXiv:1602.05875.
14. Ahmet, S.; Sangchul, A. A generalized framework for recognition of expiration dates on product packages using fully convolutional networks. *Expert Syst. Appl.* **2022**, *203*, 117310.
15. Jonathan, L.; Evan, S.; Trevor, D. Fully Convolutional Networks for Semantic Segmentation. *arXiv* **2014**, arXiv:1411.4038.
16. Baoguang, S.; Xiang, B.; Cong, Y. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. *arXiv* **2015**, arXiv:1507.05717.
17. Son, G.-J.; Kwak, D.-H.; Park, M.-K.; Kim, Y.-D.; Jung, H.-C. U-Net-Based Foreign Object Detection Method Using Effective Image Acquisition System: A Case of Almond and Green Onion Flake Food Process. *Sustainability* **2021**, *13*, 13834. [CrossRef]
18. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention MICCAI, Munich, Germany, 5–9 October 2015.
19. Shen, Y.; Zhou, H.; Li, J.; Jian, F.; Jayas, D.S. Detection of storedgrain insects using deep learning. *Comput. Electron. Agricult.* **2018**, *145*, 319–325. [CrossRef]
20. Shaoqing, R.; Kaiming, H.; Ross, G.; Jian, S. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497.
21. Aymen, Z.; Samer, A. Performance Modeling and Analysis of LTE/Wi-Fi Coexistence. *Electronics* **2022**, *11*, 1035. [CrossRef]
22. Chienyao, W.; Alexey, B.; Hongyuan, L. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
23. Neelum, N.; Sellapan, P.; Abdul, Q.; Iftikhar, A. A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor. *IEEE Access* **2020**, *8*, 55135–55144.
24. Chen, J. Model Algorithm Research based on Python Fast API. *Front. Sci. Eng.* **2023**, *3*, 7–10. [CrossRef]
25. Raniah, Z.; Humera, S. A study of the optimization algorithms in deep learning. In Proceedings of the 2019 Third International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 10–11 January 2019; IEEE: New York, NY, USA, 2019; pp. 536–539.
26. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 16133–16142.
27. Kalra, D.S.; Barkeshli, M. Why Warmup the Learning Rate? Underlying Mechanisms and Improvements. *arXiv* **2024**, arXiv:2406.09405.
28. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum. Comput. Interact.* **2008**, *24*, 574–594. [CrossRef]
29. Gail, S.; Anthony, A. Analyzing and Interpreting Data From Likert-Type Scales. *J. Grad. Med. Educ.* **2013**, *5*, 541–542.
30. Ahmad, R.; Muhammad, H. Evaluation of a Chair-Mounted Passive Trunk Orthosis: A Pilot Study on Able-Bodied Subjects. *Sensors* **2021**, *21*, 8366. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.