



# OPEN Improved food image recognition by leveraging deep learning and data-driven methods with an application to Central Asian Food Scene

Aknur Karabay<sup>1</sup>, Huseyin Atakan Varol<sup>1</sup> & Mei Yen Chan<sup>2</sup>✉

The burden of diet-related diseases is high in Central Asia. In recent years, the field of food computing has gained prominence due to advancements in computer vision (CV) and the increasing use of smartphones and social media. These technologies provide promising potential in many applications by facilitating real-time information retrieval from food images for efficient digital food journaling, smart restaurants, and supermarkets etc. Yet, to develop a robust CV model for food information retrieval, a large-scale high quality dataset is required. Several food dataset have been developed covering Western, Mediterranean, Chinese etc. cuisines. These dataset solve the simpler problem of food classification with single food item per image, which is not practical for real-life scenarios, where meals typically consist of multiple food items. To address this gap, we developed a large-scale high-quality Central Asian Food Scenes Dataset for food localization and detection. The dataset contains 21,306 images across 239 food categories, 69,856 instances. To evaluate the dataset, we performed the parametric experiments with the object detection models, with the best results achieved using YOLOv8xl (mAP50 score of 0.677).

The rapid growth of food-related big data, driven by social media, the Internet of Things (IoT), and Artificial Intelligence (AI), has given rise to an interdisciplinary research field known as food computing<sup>1</sup>. Due to its significant implications for human health, diet, and disease, food computing has become a key area of focus in disciplines such as computer vision, multimedia, medicine, health informatics, agriculture, and bioengineering<sup>2</sup>. First introduced in 2015, the term “food computing” encompasses several major tasks, including food recognition, retrieval, and recommendation<sup>3</sup>. Among these, food recognition is foundational, involving the identification and classification of food items from visual data (e.g., images or videos), which is crucial for applications such as nutrition tracking, food authentication, smart restaurants and supermarkets, and waste management. Particularly in response to the rise in diet-related diseases, various tools are being developed to enable fast and accurate food logging for dietary monitoring<sup>3</sup>. Increasing people’s nutrition literacy and improving dietary habits start with increasing their awareness of current eating patterns. The importance of these solutions lies in the potential to foster healthy dietary patterns and serve as a preventive measure against chronic diseases including obesity, diabetes, and cardiovascular disease (CVD), which is mainly caused by the high intake of red meat and processed meats including obesity, diabetes, and cardiovascular disease (CVD), which is mainly caused by the high intake of red meat and processed meat<sup>4,5</sup>.

According to the Global Burden of Disease Studies, the incidence of diet-related CVD deaths and Disability-adjusted Life Years (DALYs) in 2019 amounted to 6.9 million and 153.2 million, respectively, indicating a substantial increase of 43.8% and 34.3% since 1990<sup>6,7</sup>. Many countries with high diet-related CVD deaths and DALYs deaths are situated in Central Asia, while the lowest death rates are in the high-income Asia Pacific region. Among the specific dietary patterns identified as having a major contribution to CVD deaths and DALYs are diets characterized by high salt intake, excessive empty calories, insufficient fruit, nuts, seeds, and vegetable consumption, and diets with low Omega-3 levels<sup>4,7,8</sup>.

<sup>1</sup>Institute of Smart Systems and Artificial Intelligence, Nazarbaev University, Astana 010000, Kazakhstan.

<sup>2</sup>Department of Biomedical Sciences, School of Medicine, Nazarbayev University, Astana 010000, Kazakhstan.

✉email: yen.chan@nu.edu.kz

Over the last decade, Artificial Intelligence (AI) has found a wide range of applications in the food industry and agriculture<sup>9</sup>. In combination with other components such as smart sensors, big data, and blockchain technologies, AI is being used in various stages of the food chains, such as food classification, production developments, quality testing and improvement, supply chain management, and food safety monitoring<sup>10</sup>. For example, in food safety, machine learning has greatly improved the detection of potential contamination sources during production<sup>11</sup>. Convolutional Neural Networks (CNN) can assist in automated quality control, preventing inferior products from entering the market by precisely identifying small defects or foreign objects on food items. Deep learning (DL) technologies like recurrent neural networks (RNN) and long short-term memory networks (LSTM) were used for real-time monitoring and detailed classification of product appearance production lines, enhancing overall product quality<sup>10</sup>. In agriculture, several AI-enabled surveillance systems have been developed to assist farmers with detecting pests and monitoring crops and soil issues to maximize the harvest yield<sup>12,13</sup>. Precision agriculture is getting interested, which uses AI systems in all steps from planting and watering to the final crop harvesting<sup>14–16</sup>. To automate the manual crop collection process several computer vision (CV) models have been developed for fruit and vegetable detection<sup>17–19</sup>. Another work has been done on the ripeness assessment of the fruits and vegetables, which is essential not only in the fields to automate the harvest collection process but also in the stocks and food chains for the products' freshness check<sup>20</sup>.

Besides this, the development of CV significantly increased the functionality of innovative tools, such as mobile applications, to offer convenient ways for individuals to automatically monitor and manage their nutritional intake<sup>21–23</sup>. In this work, we are focusing on the application of AI to facilitate nutrition literacy and dietary interventions. Empowering individuals to make well-informed decisions regarding their dietary selections can help reduce the burden of dietary-related diseases and enhance overall health outcomes. This starts with raising awareness of their current behaviors<sup>4,6</sup>. Capturing food images not only reduces the burden associated with maintaining traditional food diaries and benefits individuals lacking access to expert healthcare resources or consultation, but also boosts the social support to achieve collectively healthy eating objectives when shared on social media platforms<sup>21,24</sup>. Furthermore, food images contain precise contextual information that can be utilized by healthcare professionals to provide personalized diagnoses and treatment recommendations.

The development of automatic food logging using CV requires a high-quality dataset<sup>24</sup>. A major challenge is associated with the nature of the food domain such as intra-class and inter-class variability, visual characteristics, and high variability in shape. The problem of diet tracking using CV can be addressed in two approaches: image classification problem and object detection problem. In the first case, the entire image is classified by a single category or class. In this case, a single food item must be present on an image. Several web-crawled publicly available food classification datasets have been released, such as the Food-101 dataset<sup>25</sup>, which features 101 European food categories with 1,000 images per class, establishing itself as a benchmark for numerous recognition models<sup>6,26</sup>. Furthermore, the ISIA Food-500 dataset containing Asian, European, and African cuisines contains 500 food categories with over 400,000 images<sup>27</sup>. A DL-based food recognition system has been developed and trained on a dataset of 400,000 internet-scraped images, identifying 756 food classes predominantly consumed in Singapore<sup>28</sup>.

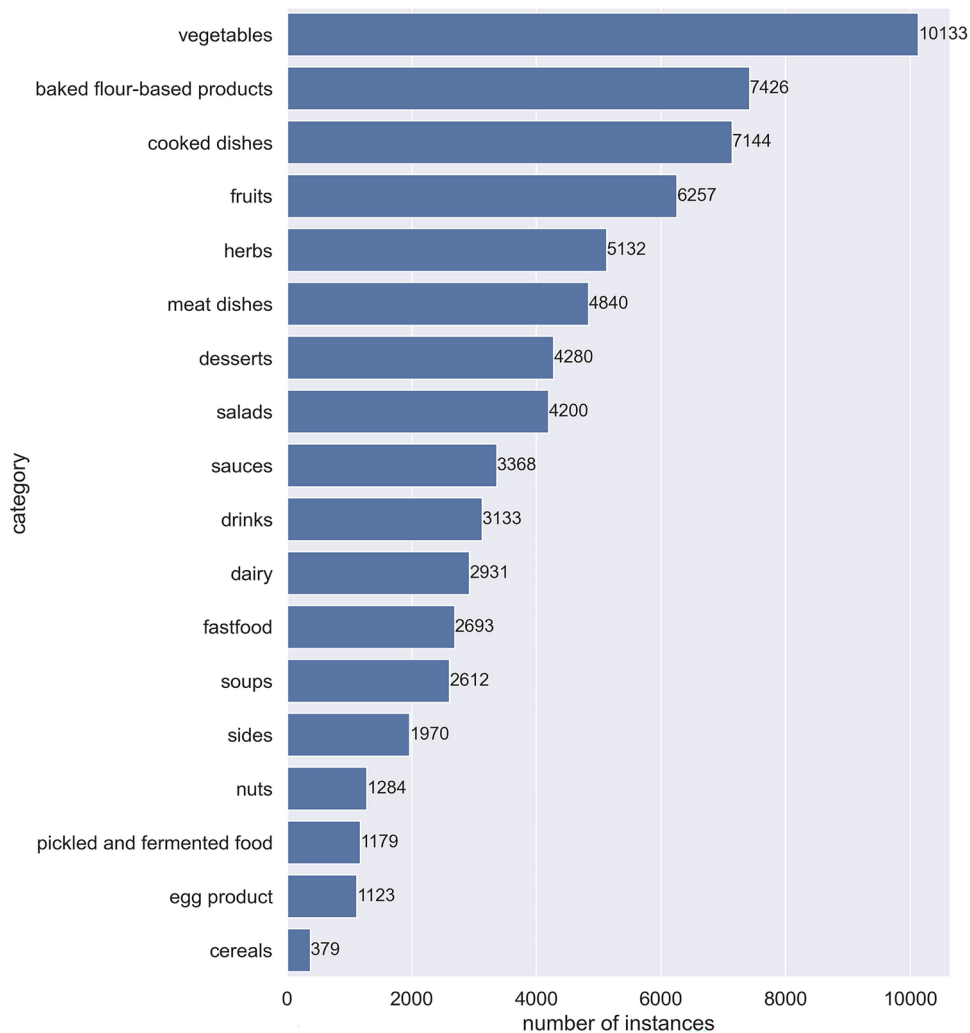
The second approach of food detection is a more complex task, since, it combines object classification and localization sub-tasks. In this case, food scenes containing multiple food items are considered. For this approach, very few publicly available datasets have been created. For example, the BTBUFood-60 dataset contains 60,000 images with 78,000 labeled instances across 60 food categories from Japanese cuisine<sup>29</sup>. Another publicly available dataset is UNIMIB2016, which contains 1027 images with a total of 3,616 instances spanning 73 food categories from Western cuisine<sup>26</sup>. Currently, most works on the application of CV for food recognition solve the classification problem. The main limitation of the classification datasets is that only one food item per image can be predicted, which may not be suitable in real life since food scenes comprising several food items are common. Therefore, for user convenience and fast food logging, food detection and localization datasets and models are required. Yet, no publicly available dataset covers Central Asian cuisine. This study aims to develop a food scene dataset that contains commonly consumed food items in Central Asia and to train a computer vision model for automatic food detection. The dataset includes annotations for each food item, labeled with rectangular bounding boxes and corresponding food class names.

In our previous work, we presented the first Central Asian Food Dataset (CAFD) for a classification task that contains 16,499 images of 42 Central Asian food items<sup>30</sup>. This work presents the first Central Asian Food Scenes Dataset (CAFSD), containing 21,306 images with 69,865 instances across 239 food classes. The dataset encompasses a wide array of food categories, including local Central Asian cuisine as well as Western, Mediterranean, Chinese, and others commonly consumed in Central Asia. This dataset presents a significant contribution to CV-assisted food and dietary tracking applications, aiming to enhance their utility by encompassing the diverse culinary landscape of Central Asia.

## Methods

### Dataset annotation

Developing and training an object detection model requires a high-quality dataset (Fig. 1). The food object detection task involves supervised learning, which required the input-output data pairs and model learns different food features and tries to localize the food item and then identify it. In this case, the input consists of images containing various food items, while the corresponding output is represented by label files. These label files encode both the food classes and the coordinates indicating the location of each food item within the image (i.e., the bounding box coordinates). The format of the label coordinates can be rectangular and polygonal and depending on the domain and type of the object to be detection the respective method can be selected. For the food object detection we have selected the rectangular bounding box format, which is preferred for its simplicity, efficiency, and consistency, making it faster and easier to implement, especially in large datasets. It is sufficient for



**Fig. 1.** Statistics of dataset instances across 18 coarse classes used in the annotation pipeline.

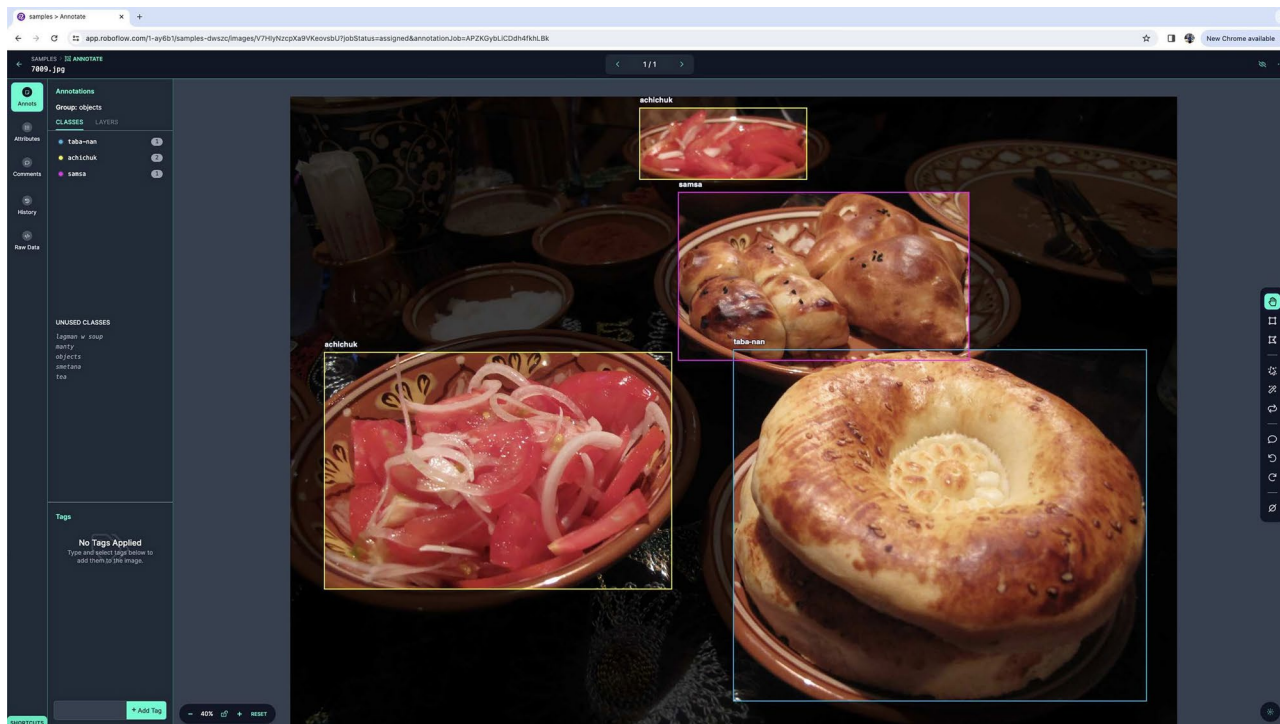
many object detection tasks and requires less computational power than polygon annotations, which are more precise but complex. Furthermore, many object detection models are optimized for the rectangular bounding box format. There are various image annotation platforms and software tools available to generate the label files<sup>31</sup>. To annotate the dataset with the bounding boxes, we used the Roboflow platform<sup>32</sup> due to its convenience in dataset statistics extraction and annotation management. Fig. 2 illustrates the sample annotations of the Central Asian traditional food scenes with the commonly consumed food items such as achichuk, taba-nan, and tandyr samsa.

Before starting the annotation process, we developed the protocol for an efficient annotation pipeline. Due to the large number of classes, we first created an ontology of food categories, and the annotation process was performed in two stages. First, all food items were labeled based on the 18 coarse classes: vegetables, baked flour-based products, cooked dishes, fruits, herbs, meat dishes, desserts, salads, sauces, drinks, dairy, fast-food, soups, sides, nuts and seeds, pickled and fermented foods, egg product, and cereals. In the second stage, the classes were split into finer labels resulting in 239 classes.

#### Dataset evaluation model

For the CAFSD evaluation and parametric experiment The You Only Look Once (YOLO) model was used, which is a state-of-the-art object detection algorithm renowned for its speed and accuracy<sup>33</sup>. Unlike traditional object detection methods, YOLO employs a single convolutional neural network (CNN) to predict bounding boxes and class probabilities simultaneously, making it fast compared to other two-step detection alternative models. YOLO divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell. Each bounding box is associated with a confidence score, indicating the model's confidence in the box containing an object and the accuracy of its localization. YOLO also employs non-maximum suppression to filter out redundant bounding boxes and improve detection accuracy<sup>33</sup>.

YOLOv8, an evolution of the YOLO architecture, enhances its predecessor by introducing various improvements. YOLOv8 utilizes a novel CSPDarknet53 backbone architecture, which improves feature



**Fig. 2.** Annotation sample of the Kazakh traditional food scenes on the Roboflow platform.

extraction efficiency while maintaining accuracy<sup>34</sup>. Additionally, it incorporates path-aggregation networks (PAN) and spatial pyramid pooling fast (SPPF) modules to capture multi-scale features effectively<sup>35,36</sup>. YOLOv8 also employs advanced data augmentation techniques and a progressive scaling strategy during training to further boost performance. These enhancements improve accuracy and robustness while retaining the real-time inference capability. Overall, the training results on the CAFSD suggest a higher capacity for generalizing, localizing, and predicting 239 food classes, showing a noticeable improvement compared to reported results on publicly available datasets. YOLO is known for. Overall, YOLOv8 represents a significant advancement in object detection, offering superior performance and efficiency for various computer vision tasks.

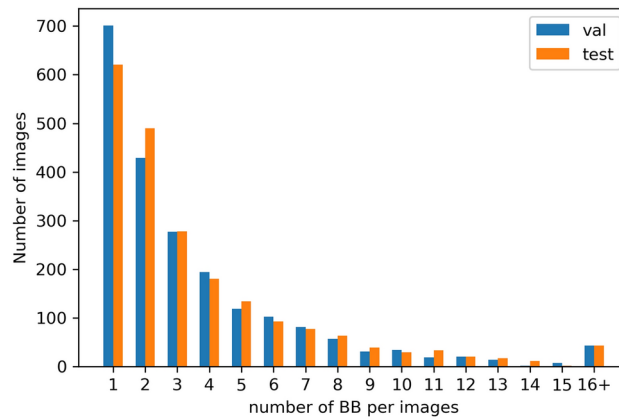
### Central Asian food scenes dataset

In this work, we propose the first Central Asia Food Scenes Dataset that contains 21,306 images with 69,856 instances across 239 food classes. To make sure that the dataset contains various food items, we took as a benchmark the ontology of the Global Individual Food Tool developed by the Food and Agriculture Organization (FAO) together with the World Health Organization (WHO)<sup>37</sup>. The dataset contains food items across 18 coarse classes: vegetables, baked flower-based products, cooked dishes, fruits, herbs, meat dishes, desserts, salads, sauces, drinks, dairy, fast food, soups, sides, nuts, pickled and fermented food, egg product, and cereals. Figure 1 illustrates the overall distribution of the class instances based on these categories.

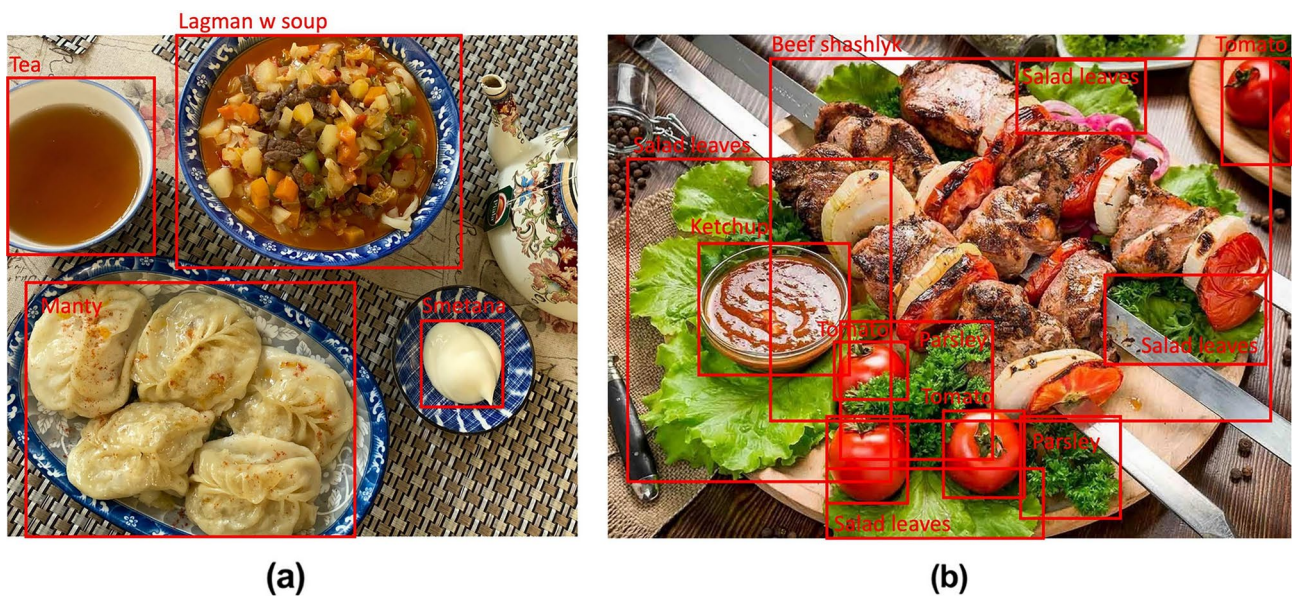
The dataset contains open-source web-scraped images from search engines (15,939 images) (i.e., Google, YouTube, and Yandex) and our own collected food images from everyday life (2324 images). To additionally extend the number of instances of the underrepresented classes, we have scraped open-source videos and extracted frames at a rate of one frame per second (3043 images). The dataset has been checked and cleaned for duplicates using the Python Hash Image library. Furthermore, we have also filtered out images less than 30 kB in size and replaced them by performing additional iterative data scraping and duplicate checks to ensure the high quality of the dataset.

The proposed Central Asian Food Scenes dataset is diverse in terms of image quality, lighting conditions, and device used (i.e., low-quality snips from mobile phones, high-resolution camera images, etc.). The dataset is unbalanced with a minimum of 40 instances per class and a maximum of up to 3050 instances per class. The Fig. 3 illustrates the distribution of the bounding boxes in the test and validation sets as the number of bounding boxes per image. It can be seen that the majority of the food scenes contain less than 15 bounding boxes per image.

Figure 4 illustrates annotated food scenes samples based on the annotation rules that we have followed to create the dataset: the liquid objects such as beverages and soups are annotated together with the dish itself (see Fig. 4a), solid food items are annotated without the plate (see Fig. 4a), in case one class is located on top of another class the annotations are made as shown on Fig. 4b; in case one class is obscured by another class and the rest of the background class is not visible we highlight only the visible part (see ‘Salad leaves’ class annotation on Fig. 4b).



**Fig. 3.** The distribution of a number of bounding boxes per image in the validation and test sets.



**Fig. 4.** Sample annotations based on the annotation protocol to illustrate overlapping and non-overlapping cases.

## Results

To evaluate the dataset we have performed parametric experiments using the YOLOv8 model, which is a state-of-the-art object detection algorithm<sup>33</sup>. For the model training the dataset was split into three sets: train set with 17,046 images (55,422 instances and ~80% of the dataset), validation set containing 2084 images (7062 instances and ~10% of the dataset) and test with 2176 images (7381 instances and ~10% of the dataset). While splitting the dataset, the condition of a minimum of five instances per class in each of the sets was followed. Furthermore, considering that the data was originally collected from scraped images and video frames, we split the videos into the training, validation, and test folds, and respective frames were included in each of the sets to avoid data leakage in the model training.

We performed transfer learning on the PyTorch<sup>38</sup> platform, using the pre-trained YOLOv8 models on the COCO (Common Objects in Context) dataset with 80 classes<sup>39</sup>. COCO is a large-scale, high-quality object detection dataset that contains over 330,000 images with more than 1.5 million object instances, which is widely used for training and evaluating machine learning models in object detection and segmentation tasks.

The training was performed on a single Tesla V100 GPU on an Nvidia DGX A100 server. Models were trained for 150 epochs with a learning rate of 0.001, batch size of 16, image size of 640 × 640 pixels, and automatic YOLOv8 augmentation. The trained models were assessed using two metrics: Mean Average Precision 50 (mAP50) evaluates the precision of detection results at an intersection-over-union (IoU) threshold of 0.5, and mAP95 follows a similar principle but considers varying IoU threshold from 0.5 to 0.95, providing a measure of precision under more stringent criteria. Both mAP50 and mAP95 offer comprehensive assessments of object detection performance across various IoU thresholds, with mAP50 focusing on typical object detection scenarios

and mAP95 emphasizing precise localization of objects<sup>40</sup>. These metrics together offer insights facilitating comprehensive performance evaluation.

Table 1 summarises the models' results on validation and test sets and illustrates the improvements in mAP50 and mAP95 metrics with the model size. The best performance metrics were obtained by training the YOLOv8xl model with more than 68 million parameters, for which the mAP50 and mAP95 scores for the test set are 0.677 and 0.601, respectively. Close mAP50 and mAP95 metrics in both validation and test sets across different model sizes ensure no overfitting occurred during the training. Besides the model accuracy, the model inference time is important as it determines how quickly the model can process data and provide results, improve user experience, and help ensure that AI systems are effective and practical in real-world scenarios. A shorter inference time means the model can analyze information faster, which is crucial for real-time tasks where prompt results are necessary, such as autonomous systems, medical diagnosis, or video and image processing. For comparison, we extracted the inference time using a Nvidia Tesla V100 GPU where the model was trained on a high-performance computer with a Nvidia RTX4090 GPU card. The results in Table 1 show that the inference on the Nvidia Tesla V100 GPU card for the largest YOLOv8xl and smallest YOLOv8n models are 5.1 ms and 0.7 ms, respectively. Whereas on the Nvidia RTX4090 GPU card, the inference time increases from 1.6 ms to 7.8 ms with the model size.

## Discussion

One of the main challenges of the CV in the food domain is the variability of food images. Environmental and technical factors like lighting conditions and camera angles, as well as differences in cooking styles and ingredients, make them more difficult than other image tasks. Moreover, fine-grained classification issues arise from intra-class variance, where foods within the same category exhibit diverse appearances due to factors like cooking styles and cultural influences. These challenges highlight the complexity of accurately detecting and classifying food items<sup>41,42</sup>. Fig. 5 illustrates the intra-class variation samples for the six classes of our CAFSD and their respective mAP50 scores.

Figure 6a illustrates the mAP50 scores variation across classes with different numbers of instances for both validation and test sets. It can be seen that due to the intra-class variation and complexity of the food domain, some classes might need fewer samples to achieve the same class mAP compared to other classes that have high intra-class variation. Furthermore, the quality of the images and the size of the bounding box can affect the output predictions, which can be observed from Fig. 6b illustrating the variation of the mAP50 score metrics for different bounding box sizes for both validation and test sets. Generally, larger bounding box sizes tend to provide more accurate predictions and higher mAP50 scores, likely due to the presence of finer features and details.

Besides the above-mentioned challenges such as intra-class variation, several other factors affect the model performance such as a cluttered background in the images and a large number of bounding boxes per image. Figure 6c shows that in general, the mAP50 score decreases with the number of bounding boxes per image in both validation and test sets.

Overall, the training results on the CAFSD suggest a higher capacity for generalizing, localizing, and predicting 239 food classes, showing a noticeable improvement compared to reported results on publicly available datasets. From the parametric experiments performed using our CAFSD, the best results were obtained using the YOLOv8lx model (i.e., mAP50 of 69.9% and 67.7% for validation and test sets, respectively). As for the other detection datasets, it has been reported that the macro average accuracy (MAA) on UNIMIB2016 is 56% which spans across 65 classes<sup>28</sup>. Experiments performed on the BTBUFood-60 dataset containing 60 food categories using the VGG16 model resulted in an mAP of 67.7%<sup>29</sup>. To compare, the parametric experiments on our first CAFD classification dataset with 42 classes showed Top-1 and Top-5 accuracy metrics of 88.70 and 98.59, respectively, using the ResNet152 model<sup>30</sup>. Historically, the Central Asian diet is known for the high consumption of meat dishes and dairy products due to the nomadic lifestyle<sup>43</sup>. Figures 7 and 8 show the distribution of different dishes within these categories.

Figure 7 illustrates that the dataset's meat-based classes with the highest number of instances include beef/lamb shashlik (11.9%), chicken shashlik (9.5%), sausages (8.8%), and fried beef/lamb (8.4%). Beef and lamb are grouped due to the difficulty in visually distinguishing these types of red meat. Notably, kazy-karta, a national dish based on horse meat, represents 6.9% of these instances. This distribution mirrors national statistics from the Bureau of National Statistics, indicating that beef is the most consumed meat in Kazakhstan, with an average of 24.68 kg per capita in 2023<sup>44</sup>. The per capita consumption of horse meat is 6.74 kg. Lamb dishes and chicken

Model	mAP50		mAP50-95		Number of parameters (M)	Inference time (ms)	
	Valid	Test	Valid	Test		Tesla V100	RTX4090
YOLOv8n	0.583	0.570	0.501	0.487	3.2	0.7	1.60
YOLOv8s	0.642	0.612	0.562	0.529	11.2	1.1	2.2
YOLOv8m	0.678	0.652	0.605	0.576	25.9	2.1	3.6
YOLOv8l	0.691	0.659	0.619	0.586	43.7	3.3	4.8
YOLOv8xl	<b>0.699</b>	<b>0.677</b>	<b>0.625</b>	<b>0.601</b>	68.2	5.1	7.8

**Table 1.** YOLOv8n model training results on the Central Asian Food Scenes dataset, model sizes, and inference times. Significant values are in bold.



**Fig. 5.** Samples of six different classes to illustrate the intra-class variation and their mAP50 scores.

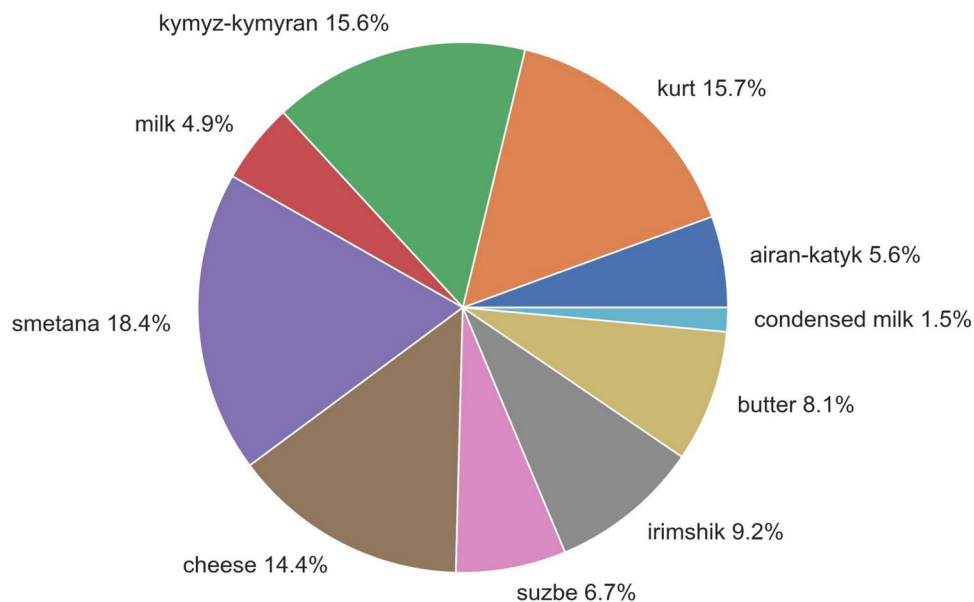
are also popular, with per capita consumption at 5.29 kg and 5.04 kg, respectively. Additionally, sausages are consumed at 2.12 kg per capita, while minced meat products have a per capita consumption of 7.07 kg.

High meat consumption in Central Asia is influenced by cultural preferences and local availability. The average meat consumption in Central Asia ranges from 50 to 70 kg per capita annually, with an average daily intake of 124.76 g/day, among the highest globally. According to the World Population Review, Kazakhstan has the highest per capita lamb consumption worldwide, averaging 8.5 kg per person annually. Recent studies suggest that shifting from a diet high in animal-based proteins to one higher in plant-based proteins may help to reduce risk factors associated with cardiovascular diseases and overall mortality<sup>43</sup>.

As for the dairy products, Fig. 8 depicts the distribution of image instances in the dataset by dairy product type. The most frequently represented dairy products are smetana (18.4%), kurt (15.7%), and kymyz-kymyryan (15.6%), followed by cheese (14.4%), irimshik (9.2%), butter (8.1%), suzbe (6.7%), airan-katyk (5.6%), milk (4.9%), and condensed milk (1.5%). According to the Bureau of National Statistics, per capita dairy product consumption in Kazakhstan was 227.2 kg in 2023<sup>44</sup>. This includes 14.889 liters of raw milk, 0.230 liters of concentrated milk without sugar, 13.166 liters of airan, 3.818 liters of smetana, 2.198 kg of irimshik, 0.581 kg of processed cheese, 4.224 kg of suzbe, and 1.084 kg of butter etc<sup>44</sup>.

Kazakhstan and Central Asia have a rich tradition of dairy consumption, with beverages like kumys and airan being particularly popular, especially in rural areas. Kumys, made from fermented mare's milk, is known for its distinct taste and slight alcoholic content, which results from lactic acid and alcoholic fermentation. It may also offer probiotic benefits. It is a nutritious, protein- and vitamin-rich drink with probiotic benefits<sup>45,46</sup>.





**Fig. 8.** Distribution of instances across dairy products.

Similarly, airan, a fermented milk drink made from mixed goat, cow, and sheep milk, is a daily staple known for its probiotics and easily digestible fatty acids and amino acids, which promote digestive health<sup>45,46</sup>.

Dairy products such as kurt, smetana, irimshik, and suzbe are also common in local cuisine. Kurt, a hard cheese made from dried sour milk, is a concentrated source of protein and calcium, is often consumed as a snack or with tea. Smetana, like sour cream, is widely used in various dishes, contributing fats and vitamins A and D. Irimshik, or dried curd, has a sweet, baked milk flavor and extended storage capacity, providing protein and minerals. Suzbe, a type of curd made by straining sour milk, is seasonally prepared and used in soups or consumed with milk or water, offering fats and proteins<sup>46</sup>.

These traditional dairy products are integral to the local diets, providing cultural significance and essential nutrients that support health and well-being. Their consumption reflects their importance in Central Asian culinary practices and their nutritional benefits<sup>45,46</sup>.

To conclude, CAFSD makes a valuable contribution to computer vision-assisted food and dietary tracking applications, which can also be used in various settings like smart restaurants and supermarkets. It could play a significant role in advancing local CV food-tracking applications, with the potential to enhance nutrition literacy, increase dietary awareness, and promote healthier food choices. These advancements have the potential to impact overall agriculture, the environment, and the food system in the region.

The performance outcomes of the trained food object detection models on our dataset demonstrate the effectiveness of the YOLOv8xl model as compared to smaller size models of the YOLOv8 family in accurately retrieving food-related information. As the next steps, we plan to integrate the model into a smartphone application and develop a dataset with macro- nutritional values and corresponding prediction models for more detailed guidance. A follow-up project will extend the dataset composition and the number of classes integrating regional food composition databases that encompass local foods, dishes, and their nutritional profiles. This step ensures our research is aligned with precise, region-specific data to support evidence-based decision-making for public health policies tailored to the target population. Furthermore, a comprehensive codebook to systematically link food images to their nutritional labels will be developed to visual data based on food type, portion size, and preparation method, with each image cross-referenced to nutritional data, including macronutrient and micronutrient profiles. This integration bridges the gap between visual representations of food and their corresponding nutritional values, making the dataset a valuable resource for dietary analysis and research, particularly in nutrition-related interventions.

### Data availability

The dataset is available in our GitHub repository: <https://github.com/IS2AI/Central-Asian-Food-Scenes-Data-set>.

Received: 6 July 2024; Accepted: 24 March 2025

Published online: 23 April 2025

### References

- Zahisham, Z., Lee, C. P. & Lim, K. M. Food recognition with resnet-50. In *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAET)*, 1–5, <https://doi.org/10.1109/IICAET49801.2020.9257825> (2020).
- Wróblewska, A. *et al.* Tasteset—recipe and food entities dataset, <https://doi.org/10.21227/11bb-v380> (2022).

3. Jiang, L., Qiu, B., Liu, X., Huang, C. & Lin, K. Deepfood: Food image analysis and dietary assessment via deep model. *IEEE Access* **8**, 47477–47489. <https://doi.org/10.1109/ACCESS.2020.2973625> (2020).
4. Meier, T., Grafe, K., Senn, F., et al. Cardiovascular mortality attributable to dietary risk factors in 51 countries in the WHO European Region from 1990 to 2016: a systematic analysis of the global burden of disease study. *Eur. J. Epidemiol.* **34**, 37–55. <https://doi.org/10.1007/s10654-018-0473-x> (2019).
5. Teng, C.-C. & Chih, C. Sustainable food literacy: A measure to promote sustainable diet practices. *Sustain. Prod. Consum.* **30**, 776–786. <https://doi.org/10.1016/j.spc.2022.01.008> (2022).
6. Wang, D. D. et al. Global improvement in dietary quality could lead to substantial reduction in premature death. *J. Nutr.* **149**, 1065–1074. <https://doi.org/10.1093/jn/nxz010> (2019).
7. Aringazina, A., Kuandikov, T. & Arkhipov, V. Burden of the cardiovascular diseases in Central Asia. *Cent. Asian J. Glob. Heal.* **7**, 321 (2018).
8. Miller, V. et al. Global dietary quality in 185 countries from 1990 to 2018 show wide differences by nation, age, education, and urbanicity. *Nat. Food* **3**, 694–702. <https://doi.org/10.1038/s43016-022-00594-9> (2022).
9. Ding, H. et al. The application of artificial intelligence and big data in the food industry. *Foods* **12** (2023).
10. Wakchaure, M., Patle, B. & Mahindrakar, A. Application of ai techniques and robotics in agriculture: A review. *Artif. Intell. Life Sci.* **3**, 100057. <https://doi.org/10.1016/j.ailsci.2023.100057> (2023).
11. Hassoun, A. et al. Food quality 4.0: From traditional approaches to digitalized automated analysis. *J. Food Eng.* **337**, 111216 (2023).
12. Javaid, M., Haleem, A., Singh, R. P. & Suman, R. Artificial intelligence applications for industry 4.0: A literature-based study. *J. Ind. Integr. Manag.* **7**, 83–111 (2022).
13. Jung, J. et al. The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. *Curr. Opin. Biotechnol.* **70**, 15–22 (2021).
14. Javaid, M., Haleem, A., Khan, I. H. & Suman, R. Understanding the potential applications of artificial intelligence in agriculture sector. *Adv. Agrochem* **2**, 15–30. <https://doi.org/10.1016/j.aac.2022.10.001> (2023).
15. Linaza, M. T. et al. Data-driven artificial intelligence applications for sustainable precision agriculture. *Agronomy* **11**, 1227 (2021).
16. Shadrin, D. et al. Enabling precision agriculture through embedded sensing with artificial intelligence. *IEEE Trans. Instrum. Meas.* **69**, 4103–4113 (2019).
17. Liu, G. et al. TomatoDet: Anchor-free detector for tomato detection. *Front. Plant Sci.* **13**, 942875 (2022).
18. Fu, L. et al. YOLO-Banana: A lightweight neural network for rapid detection of banana bunches and stalks in the natural environment. *Agronomy* **12**, 391 (2022).
19. Latha, R. et al. Fruits and vegetables recognition using YOLO. In *Proceedings of the International Conference on Computer Communication and Informatics (ICCCI)*, 1–6. <https://doi.org/10.1109/ICCCI54379.2022.9740820> (2022).
20. Xiao, B., Nguyen, M. & Yan, W. Q. Fruit ripeness identification using YOLOv8 model. *Multimed. Tools Appl.* **83**, 28039–28056. <https://doi.org/10.1007/s11042-023-16570-9> (2024).
21. Dalakleidi, K. V., Papadelli, M., Kapos, I. & Papadimitriou, K. Applying image-based food-recognition systems on dietary assessment: A systematic review. *Adv. Nutr.* **13**, 2590–2619. <https://doi.org/10.1093/advances/nmac078> (2022).
22. Guan, V. et al. A novel mobile app for personalized dietary advice leveraging persuasive technology, computer vision, and cloud computing: Development and usability study. *JMIR Form Res* <https://doi.org/10.2196/46839> (2023).
23. Nadeem, M., Shen, H., Choy, L. & Barakat, J. M. H. Smart diet diary: Real-time mobile application for food recognition. *Appl. Syst. Innov.* <https://doi.org/10.3390/asi6020053> (2023).
24. Kakani, V., Nguyen, V. H., Kumar, B. P., Kim, H. & Pasupuleti, V. R. A critical review on computer vision and artificial intelligence in food industry. *J. Agric. Food Res.* **2**, 100033. <https://doi.org/10.1016/j.jafr.2020.100033> (2020).
25. Bossard, L., Guillaumin, M. & Gool, L. V. Food-101—mining discriminative components with random forests. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 446–461. [https://doi.org/10.1007/978-3-319-10599-4\\_29](https://doi.org/10.1007/978-3-319-10599-4_29) (2014).
26. Ciocca, G., Napoletano, P. & Schettini, R. Food recognition: A new dataset, experiments, and results. *IEEE J. Biomed. Heal. Inform.* **21**, 588–598 (2016).
27. Min, W. et al. ISIA Food-500: A dataset for large-scale food recognition via stacked global-local attention network. *Proc. ACM Int. Conf. Multimed.* (2020).
28. Sahoo, D. et al. FoodAI. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/3292500.3330734> (2019).
29. Cai, Q., Li, J., Li, H. & Weng, Y. BTBUFood-60: Dataset for object detection in food field. In *Proceedings of the IEEE International Conference on Big Data and Smart Computing (BigComp)*, 1–4. <https://doi.org/10.1109/BIGCOMP.2019.8678916> (2019).
30. Karabay, A., Bolatov, A., Varol, H. A. & Chan, M.-Y. A Central Asian food dataset for personalized dietary interventions. *Nutrients* **15**. <https://doi.org/10.3390/nu15071728> (2023).
31. Ali, S. et al. An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-59413-5> (2020).
32. Roboflow. Roboflow: give your software the sense of sight. (2024). Accessed 25 March 2024.
33. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788 (2016).
34. Diwan, T., Anirudh, G. & Tembhurne, J. V. Object detection using yolo: Challenges, architectural successors, datasets and applications. *Multimed. Tools Appl.* **82**, 9243–9275 (2023).
35. Liu, S., Qi, L., Qin, H., Shi, J. & Jia, J. Path aggregation network for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8759–8768 (2018).
36. He, K., Zhang, X., Ren, S. & Sun, J. *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition*, 346–361 (2014).
37. FAO. FAO/WHO global individual food consumption data tool (GIFT): methodological document (2022).
38. Pytorch. Pytorch. (2024). Accessed: 25 March 2024.
39. Lin, T.-Y. et al. Microsoft COCO: Common objects in context. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 740–755 (Springer, 2014).
40. Zhu, H., Wei, H., Li, B., Yuan, X. & Kehtarnavaz, N. A review of video object detection: Datasets, metrics and methods. *Appl. Sci.* **10**. <https://doi.org/10.3390/app10217834> (2020).
41. Li, H., Guo, W., Lu, G. & Shi, Y. Augmentation method for high intra-class variation data in apple detection. *Sensors* **22**. <https://doi.org/10.3390/s22176325> (2022).
42. Gilal, N. U. et al. Evaluating machine learning technologies for food computing from a data set perspective. *Multimed. Tools Appl.* **83**, 32041–32068. <https://doi.org/10.1007/s11042-023-16513-4> (2024).
43. Auyeskhani, U. et al. Reducing meat consumption in Central Asia through 3D printing of plant-based protein—enhanced alternatives—a mini review. *Front. Nutr.* <https://doi.org/10.3389/fnut.2023.1308836> (2024).
44. QazStat. Food consumption in households of the Republic of Kazakhstan (2023). (2024). Accessed: 22 May 2024.
45. Martynushkin, A. et al. Modern trends and development problems of the milk and dairy products market in the Russian Federation. *77–84*. <https://doi.org/10.32743/kuz.agri.2020.77-84> (2020).
46. Konuspayeva, G., Baubekova, A., Akhmetadykova, S. & Faye, B. Traditional dairy fermented products in Central Asia. *Int. Dairy J.* <https://doi.org/10.1016/j.idairyj.2022.105514> (2023).

## Acknowledgements

We acknowledge Vasila Mirzamshova for assisting in dataset annotation. We also acknowledge Makpal Kairatkyzy and Indira 346 Imanaliyeva for their comments on the analysis of Central Asian diets.

## Author contributions

H.A.V., M.Y.C., and A.K. contributed to the study design. H.A.V., M.Y.C., and A.K. participated in data collection, analyzed the data, and produced the figures. H.A.V., M.Y.C., and A.K. wrote the final paper. All authors reviewed and approved the paper before submission.

## Funding

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23485288) as well as Nazarbayev University, under the Faculty Development Competitive Research Grant Program (Grant No. 201223FD2603).

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to M.Y.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025