Online Class-Incremental Learning For Real-World Food Image Classification

Siddeshwar Raghavan
raghav12@purdue.edu

Jiangpeng He*
he416@purdue.edu

Fengqing Zhu
zhu0@purdue.edu

School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana USA

Abstract

Food image classification is essential for monitoring health and tracking dietary in image-based dietary assessment methods. However, conventional systems often rely on static datasets with fixed classes and uniform distribution. In contrast, real-world food consumption patterns, shaped by cultural, economic, and personal influences, involve dynamic and evolving data. Thus, require the classification system to cope with continuously evolving data. Online Class Incremental Learning (OCIL) addresses the challenge of learning continuously from a single-pass data stream while adapting to the new knowledge and reducing catastrophic forgetting. Experience Replay (ER) based OCIL methods store a small portion of previous data and have shown encouraging performance. However, most existing OCIL works assume that the distribution of encountered data is perfectly balanced, which rarely happens in real-world scenarios. In this work, we explore OCIL for real-world food image classification by first introducing a probabilistic framework to simulate realistic food consumption scenarios. Subsequently, we present an attachable Dynamic Model Update (DMU) module designed for existing ER methods, which enables the selection of relevant images for model training, addressing challenges arising from data repetition and imbalanced sample occurrences inherent in realistic food consumption patterns within the OCIL framework. Our performance evaluation demonstrates significant enhancements compared to established ER methods, showing great potential for lifelong learning in real-world food image classification scenarios. The code of our method is publicly accessible at https://gitlab.com/viper-purdue/OCIL-real-world-food-image-classification

1. Introduction

Food image classification has shown great potential for improving food pattern tracking [16, 31, 36], nutritional and health analysis [35, 41], and dietary monitoring [12, 15, 30, 40]. Current food classification models have achieved remarkable performance on static datasets that do not change with time and have a fixed number of classes. However, food patterns change and food classes expand in the real world due to shifts in dietary styles and preferences [44, 49]. This makes it necessary for food image classification models to adapt and learn the new information from the changing data [19]. In order to accommodate evolving data, we propose to leverage Online Class Incremental Learning (OCIL) approaches (outlined in detail in [34]). These methods continuously learn from a growing stream of sequentially arriving data. A major challenge in OCIL methods is catastrophic forgetting [7, 13, 39] of old knowledge while learning new data, leading to performance degradation. In this work, we focus on Online Class Incremental Learning (OCIL) frameworks [17, 20, 21] for food image classification task. The OCIL setting is more realistic but challenging compared to offline, as the model must continually learn new classes from an online data stream encountering each incoming sample only once, while past observed data is not accessible.

*corresponding author

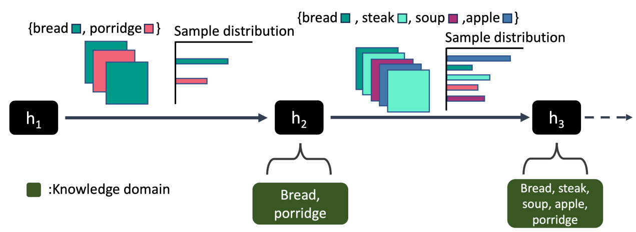

Figure 1. **R-OCIL for food image classification**. The model h learns new food class data sequentially from realistic consumption patterns without accessing previously learned data. These data streams may include repetitions and imbalanced sample distributions. The updated model should classify all food classes it has encountered.

Nevertheless, existing OCIL approaches operate under

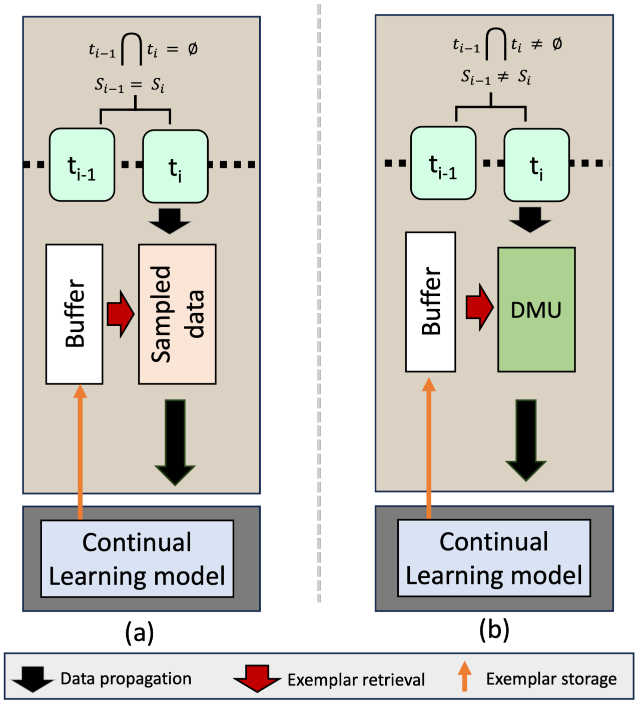

Figure 2. (a) Conventional ER OCIL strategy, (b) Our proposed realistic setting along with our DMU module. Existing OCIL ER studies [2, 3, 14, 29, 45, 48] used (a). Each ER technique employs a different exemplar update and retrieval strategy, as indicated by the corresponding arrows in the figure.

constraints that do not reflect real-world situations. For instance, certain constraints may impose equal task sizes, equal samples for each class in a task, and balanced training samples [2,3,14,48]. As a result, many existing methodologies assume that the training data is perfectly balanced, either explicitly or implicitly. However, real-world food consumption patterns seldom exhibit such fixed structure or balance. The most recent studies of Class-Incremental with Repetition (CIR) in offline scenarios [9,10,23,42] have gained significant attention as a promising direction in Continual Learning. These settings offer enhanced flexibility in task definition, facilitating the integration of new and previously encountered classes. However, the consideration of repetition in this context remains under-explored in the online setting. This poses a challenge, but it holds significant relevance in practical, resource-constrained scenarios like the classification of food images in applications such as diet management and food intake monitoring systems [1, 15].

In this work, we fill this gap and extend the OCIL setting to a more realistic setup by removing the constraints mentioned above, which we refer to as Realistic Online Class Incremental Learning (R-OCIL) as shown in Figure 1. To generate experimental benchmarks that closely mimic real-world food patterns influenced by different dietary routines, we first summarize real-world food consumption patterns into three main categories, including short-term, moderate-term, and long-term. We introduce a Realistic Data Distribution Module (RDDM) to simulate these food consumption patterns. These scenarios are formulated based on approaches for safe and sustainable weight loss and maintenance, as explained in Section 3.3. We propose a simple yet effective plug-and-play module for existing Experience Replay (ER) OCIL methods called Dynamic Model Update (DMU) as shown in Figure 2. The DMU optimizes the selection of the most representative samples for training, facilitating the natural accumulation of knowledge over time.

The main contributions of our work can be summarized as follows:

1. We propose a probabilistic framework called Realistic Data Distribution Module (RDDM) for Realistic Online Class Incremental Learning (R-OCIL) methods. To the best of our knowledge, this is the first such framework proposed for simulating various realistic food consumption patterns for image classification.

2. We introduce a dynamic plug-and-play module called Dynamic Model Update (DMU), which can be integrated with existing Experience Replay (ER) OCIL methods to improve learning accuracy and mitigate catastrophic forgetting in realistic scenarios.

3. We simulate three real-world food consumption patterns - short-term, moderate-term, and long-term on the challenging Food-101 [5] and VFN [18] datasets using our proposed RDDM framework. We show that incorporating our DMU module significantly improves the performance of ER methods across all three scenarios.

2. Related Work

2.1. Food Image Classification

Food image classification is a computer vision task that assigns a food category or label to an input food image. This task falls under the umbrella of image classification, wherein deep learning models are trained to recognize different food items based on their visual features as in [16, 31, 40, 41, 46, 47, 51]. The task typically involves assigning a single food category to an input image, assuming that the image contains only one type of food. Furthermore, recent research efforts have leveraged hierarchical structures [37] based on visual information to achieve further performance improvements. Nonetheless, all these methodologies rely on fixed food image datasets for training, lacking the capacity to learn from sequentially presented data. This limitation hampers their applicability in real-world scenarios where new foods are continuously encountered over time.

2.2. Online Class Incremental Learning

Class Incremental Learning (CIL) [4, 6, 24, 50] is a machine learning paradigm where a model is trained to progressively incorporate new classes or categories of data over time. This learning approach involves updating the model in such a way that it can accommodate additional classes without compromising its ability to recognize previously learned classes. In the context of CIL, there are two primary learning protocols: (i)*Offline* [10, 23, 38] (We refer to as CIL in this paper)(ii) *Online* [20, 34](We refer to as OCIL in this paper): In this protocol, the training data also arrives sequentially, similar to the offline scenario, but the model can only be trained on each sample once without accessing data from the previous tasks. Current OCIL methods can be taxonomized into two major categories: (i)Regularization-based [8, 25, 28, 32, 33, 52] and (ii) Experience Replay-based methods. Regularization-based techniques involve including an extra penalty term within the loss function to penalize updates to important parameters. Experience Replay-based (ER) techniques address the issue of catastrophic forgetting by storing a subset of learned task data as exemplars within a memory buffer. These exemplars are then employed for rehearsal during the process of continual learning. When compared to regularization-based methods, ER techniques have showcased superior efficiency. In this study, we center our attention on ER-based approaches for comparison. The MIR [2] method executes a virtual parameter update and retrieves data from the memory buffer that are primarily impacted, using the loss from the current mini-batch. Additionally, the buffer retrieval and storage procedures vary significantly between the"online" and "offline" settings due to the specific data access constraints.

In contrast, GSS [3] opts for exemplar selection based on gradient directions. Although initially designed for CIL(offline), iCaRL [34, 45] has exhibited effectiveness even in the OCIL(online) context, achieved through random exemplar selection coupled with the nearest-class-mean classifier. ASER [48], singles out the most efficient data for use as exemplars using adversarial shapley value. DVC [14] introduces a gradient-based exemplar selection process that primarily responds to incoming samples with high interference. This approach also includes a representation learner that maximizes mutual information, drawing inspiration from contrastive learning.

3. Problem Formulation

3.1. Problem setup

Online Class Incremental Learning involves learning a sequence of T tasks, denoted by t_1 to t_T, one task at a time. During the learning of any task t_i, no data from any other task other than t_i is accessed. The model is updated from h_1 to h_T. In each task t_i, there are K_i classes, and the distribution of samples and classes adheres to the three characteristics detailed in Section 3.2. For a specific task t_i, the training data for the model is denoted as S_i, which consists of n_i sets of training examples presented in the structure $(x_{i,1}, y_{i,1}), ...(x_{i,n_i}, y_{i,n_i})$. Here, x refers to an image, and y represents its associated label. The index i corresponds to the particular task t_i, where $i \in \{1, 2, ..., T\}$.

3.2. Characteristics of R-OCIL

We identify three main characteristics of R-OCIL and use them to propose a probabilistic formulation to simulate realistic food consumption patterns. Let $C = \{C_1, C_2, ..., C_N\}$ denote a set of N classes. The sample sizes of different classes appear in a task $t_i \in T$ are modeled as a random vector S_{t_i} where each element $s_{i,j}$ represents the sample size of class C_j in task t_i. We normalize the total samples of a class C_j by its occurrences across all tasks T.

Characteristic 1: The total number of classes in a given task t_i is not fixed. We denote the number of classes within that task as K_i, expressed as:

$$K_i = i \in C : 1 \leq \sum_j s_{i,j} < N \tag{1}$$

To uphold the concept of class incremental learning, we establish $K_i \geq K_{i-1}+1$, ensuring that each task t_i introduces at least one new class compared to the preceding task t_{i-1}. Additionally, by setting an upper limit below N (the overall number of classes), we prevent all food classes from being presented within a single task. This scenario, uncommon in reality, would lead to the transformation of the problem into Domain Incremental Learning [50]. In such cases, the model would need to process all data simultaneously.

Characteristic 2: The sample sizes of occurring classes within any given task t_i may not be equal.

$$\begin{gathered}\forall m, n \in N \mid m \neq n \\ P(s_{i,m} \neq s_{i,n}) > 0, \; \forall \; s_{i,m} \neq s_{i,n} \neq 0\end{gathered} \tag{2}$$

Characteristic 3: Classes appearing across different tasks could overlap. Let's consider two tasks t_i, t_j such that $i \neq j$

$$C_{t_i} \cap C_{t_j} \neq \emptyset \tag{3}$$

where C_{t_i} and C_{t_j} denote the classes occurring in tasks i and j.

3.3. Formulation of food consumption pattern categories

Food consumption patterns offer significant insights into the dietary behaviors of individuals or communities within

Table 1. **Food consumption categories**. The range of classes for each dataset experimented with and the corresponding α and β values used to produce patterns reflecting the number of represented classes.

Food Consumption Category	α	β	Food 101 101 Classes	VFN 74 Classes
Short-term	[9, 10, 11, 12]	[1, 2, 3]	20-40	10-30
Moderate-term	[5, 6, 7, 8]	[4, 5, 6]	40-80	30-50
Long-term	[1, 2, 3, 4]	[7, 8, 9]	80-101	50-74

a defined time frame. This encompasses details like the kinds and quantities of ingested food and beverages, eating routines, meal frequencies, and food preparation methods. Understanding these patterns is essential for delving into the relationship between diet and health consequences [12, 40, 41].

Due to the unavailability of datasets capturing food consumption patterns, we simulate them. Nonetheless, considering the multitude of potential permutations in food consumption patterns, replicating every single one proves to be a formidable task. To streamline this process, we classify these patterns into three primary types: *short-term*, *moderate-term*, and *long-term*. These classifications draw from dietary and nutritional recommendations [11, 26].

In reality, food consumption patterns often do not involve the consumption of every available food item. To address this, we introduce two hyperparameters labeled α and β. The hyperparameter α controls the repetition of previously observed classes and β controls the addition of new food classes. We must consider diverse dietary styles to determine suitable values for these hyperparameters corresponding to each food consumption category. For instance, certain diets like low-protein regimens, which restrict caloric intake, are typically followed for short duration, around six months, to facilitate healthy weight loss without negative consequences [26]. To simulate such a diet within the short-term food consumption category, we cap the number of accessible food classes at 40 for the Food-101 dataset and 30 for the VFN dataset, respectively. Thus, although α and β offer a broad spectrum of feasible values for encompassing various food consumption patterns, we have selected three sets of values as representatives of these dietary habits. These specific values are elaborated upon in Table 1.

4. Methodology

Our method comprises two primary components. Firstly, we introduce a probabilistic framework called the Realistic Data Distribution Module (RDDM). This module aims to simulate realistic food consumption patterns, as elaborated in Section 3.3, while aligning with the characteristics of Realistic OCIL as outlined in Section 3.2. Secondly, we propose a plug-and-play module named the Dynamic Model Update (DMU). This module is designed to seamlessly integrate with existing experience replay-based OCIL methods. The DMU operates by continuously assessing the model's current performance and dynamically selecting the most representative food images from the input image sequence within each task for training. This adaptive approach enhances overall learning and reduces forgetting.

4.1. Realistic Data Distribution Module (RDDM)

Simulating food consumption patterns can be challenging due to managing multiple variables and determining suitable repetition and probabilistic distributions. Additionally, it can be difficult to select which food classes to repeat, their order and distributions. In this paper, we introduce a framework for generating realistic food scenarios that can be customized to various dietary patterns by changing the input probability distribution and two hyperparameters. The proposed framework aims to simplify the process of generating realistic food consumption patterns, which can be helpful in various applications, such as food recommendation systems, continual health monitoring, and nutrition research.

We consider an RDDM formulation for realistic food consumption patterns following the three characteristics introduced in Section 3.2.

$$\begin{gathered} K_i^{repeat} = D(i, \alpha) \\ K_i^{new} = D(i, \beta) \\ K_i = K_i^{new} + K_i^{repeat} \end{gathered} \tag{4}$$

The task-specific distribution D determines the **number of classes** K_i present in a given task i. We split this equation into two parts. As described earlier, we are using the hyperparameter β to control the addition of new food classes and the hyperparameter α to control the repetition of previously observed classes. As a result, K_i is a stochastic variable that fulfills **Characteristic 1**. Different food consumption patterns can be simulated by selecting various data distributions D. For instance, in Section 5.3, we employ an exponential function as the data distribution D to generate a more significant level of imbalance. The degree of imbalance, represented by the imbalance factor (ρ), relies on the number of categories present in the dataset and the hyperparameters α and β as represented in Equation 5.

$$\rho = \frac{N - E}{\alpha} \tag{5}$$

where N represents total number of classes and E represent the number of encountered classes.

$$A_i = L(K_i, W_i^{'}) \tag{6}$$

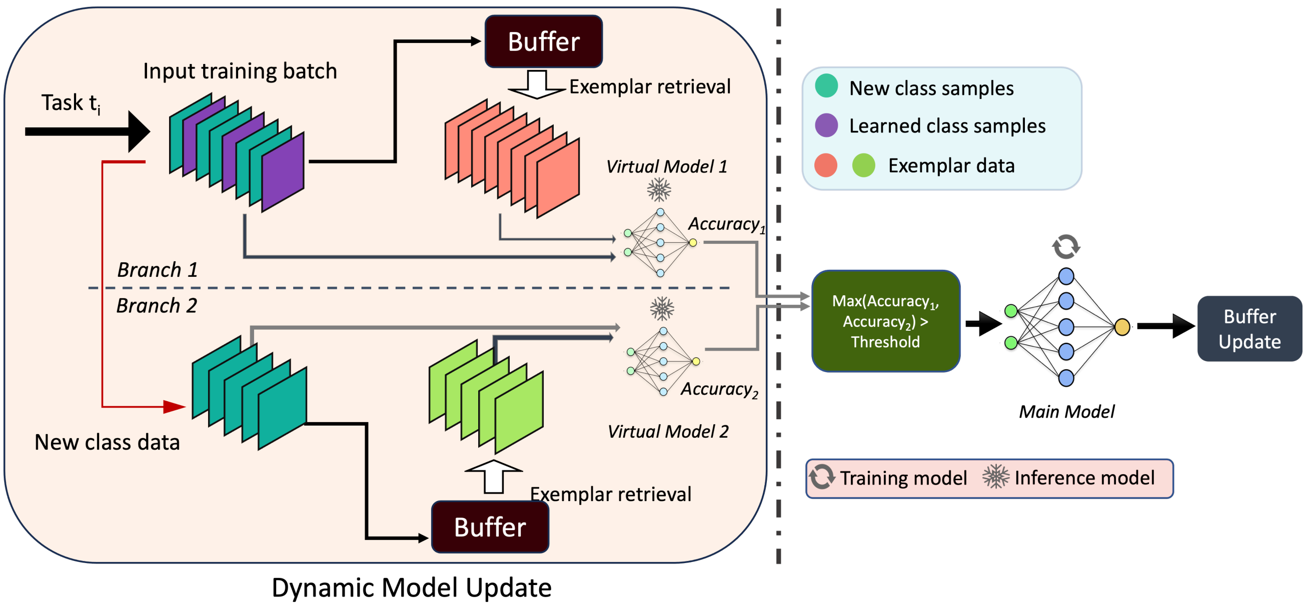

Figure 3. **Overview of proposed method**: The left section of the diagram illustrates the Dynamic Model Update (DMU) process, containing two concurrent data inference pathways. The first branch encompasses a virtual model denoted as VM_1, which receives the complete batch of data and the corresponding exemplars retrieved from the buffer. Simultaneously, the second branch involves another virtual model named VM_2, and receives only the non-repeated data (associated with new class samples) computed based on Equations 9 and 10 along with its corresponding exemplars. At any given task t_i, the virtual models correspond to the OCIL model trained until task t_{i-1}. The strategy for retrieving exemplars and updating the buffer depends on the specific ER method employed in the experiment. The accuracy of these two virtual models is subsequently compared against the accuracy of the initial task, serving as the threshold. Next, the input image sequence and exemplars from the virtual branch that exhibits superior performance are directed into the main OCIL model for training.

The binary indicator vector A_i denotes which classes appear in task t_i and is drawn from distribution $L(K_i, W_i^{'})$ shown in Equation 6, where the value "1" indicates the presence of a particular class. This distribution is dependent on the number of classes K_i and satisfies **Characteristic 3**. The weight vector $W_i^{'}$ determines the probability of a class appearing in task t_i, with higher weights indicating a greater likelihood of appearance. We set L as uniform distribution in Section 5.1 as each class has an equally likely chance of appearing in a task.

$$S_i = Q(A_i) \tag{7}$$

S_i represents the **sample sizes** of each class in A_i, and is drawn from the distribution $Q(A_i)$ as shown in Equation 7. To simplify the setup, we allocate the total number of samples for a class by dividing it by the total number of occurrences across all tasks and satisfies **Characteristic 2**. For example, if *"class 5"* has 500 samples and appears five times in ten tasks, each appearance of *"class 5"* in a task will have 100 samples. Any extra samples are allocated to the first task if this number is not divisible.

4.2. Dynamic Model Update (DMU)

In this section, we introduce the Dynamic Model Update (DMU), as depicted in Figure 3. Given the common occurrence of repeated foods in real-life food consumption patterns, which can act as natural replays and partially alleviate the impact of forgetting, we designed a straightforward and effective training module. The integration of DMU with any existing ER method can be seamlessly achieved without disrupting its workflow. Our primary objective is to enhance the model update pipeline to integrate seamlessly with the latest ER methods. Additionally, we showcase how our approach can be adapted to accommodate diverse buffer update strategies employed by various ER methods, resulting in enhanced performance for real-world food classification.

The key idea of our DMU is to avoid training the model without being aware of the input image data from the sequence. Moreover, since the input sequence can contain possible repetitions of previously encountered classes, the model could overfit if images from a specific class appear frequently. To address this, we have devised a strategy that dynamically selects whether to incorporate the entire training batch or solely the new classes based on the model's ongoing training performance. Upon encountering data from task t_i during the training phase, we establish two virtual models: VM_1 and VM_2. All layers are frozen in these virtual models, and gradient updates are not applied. These virtual models are clones of the primary model trained until the current task t, denoted as h_{i-1}. The first branch (VM_1) of these models is deduced using the complete input batch $[X, Y]$ and corresponding exemplars $[X_E, Y_E]$, as depicted in Equation 8.

For VM_2, the second branch of the models, the input comprises $[X_H, Y_H]$, alongside corresponding exemplars retrieved from the buffer, represented as X_{E_H}, Y_{E_H}. The selection of exemplars for the two virtual branches differs due to the distinct inputs of each branch. Each ER method employs a distinct exemplar retrieval strategy, ranging from random retrieval in iCaRL and GSS to constraint-based retrieval in MIR, ASER, and DVC. To acquire exemplars for the non-repeated (new sample) class images, the prior label distribution ($PD(t_i)$) and the current input $[X, Y]$ are compared. Here, $PD(t_i)$ denotes the stored values of all unique label sets that have emerged until task t_i. The computation of non-repeated data $[X_H, Y_H]$ can be accomplished using Equation 9, as illustrated in Figure 3.

$$\begin{aligned} \hat{y}^{VM_1} &= VM_1^{h_{i-1}}([X, X_E]) \\ prediction(x_i) &= argmax(\hat{y}_i^{VM_1}) \\ Accuracy^{VM_1} &= \frac{1}{N_1}\sum_{i=1}^{N_1}[\hat{y}^{VM_1} = y^{VM_1}] \end{aligned} \tag{8}$$

Where $\hat{y}^{VM_1} = [Y, Y_E]$ and N_1 is the corresponding batch size.

$$H([X, Y], PD(k)) = [X_H, Y_H] \tag{9}$$

The non-repeated labels are obtained by the function H, as illustrated in Equation 10, subsequently facilitating the extraction of the corresponding images.

$$[Y_H] = [Y] \cap PD(k) \tag{10}$$

Conversely, the second branch is inferred exclusively with the new class image $[X_H, Y_H]$ (i.e., non-repeated) samples and their corresponding exemplars $[X_{E\_H}, Y_{E\_H}]$, as shown in Equation 11

$$\begin{aligned} \hat{y}^{VM_2} &= VM_2^{h_{i-1}}([X_H, X_{E\_H}]) \\ prediction(x_i) &= argmax(\hat{y}_i^{VM_2}) \\ Accuracy^{VM_2} &= \frac{1}{N_2}\sum_{i=1}^{N_2}[\hat{y}^{VM_2} = y^{VM_2}] \end{aligned} \tag{11}$$

where $\hat{y}^{VM_2} = [Y_H, Y_{E\_H}]$ and N_2 is the corresponding batch size.

Following the initial task, the achieved accuracy is preserved as the threshold accuracy (TH), given that different food consumption patterns yield different accuracy and cannot be fixed. Subsequently, the DMU module employs this threshold accuracy to determine the branch with accuracy surpassing the threshold. The data from this branch is then directed into the main training pipeline. The primary model, trained until task t_i and represented as h_{i-1}, is subsequently updated via supervised backpropagation.

5. Experiments

5.1. Experimental setup

Datasets. We introduce two benchmarks for Realistic Online Class Incremental Learning (R-OCIL) in the realm of real-world food image classification: the Food-101 [5] dataset (101 classes) and the VFN [18] dataset (74 classes). We simulate food consumption patterns by utilizing the RDDM module, which categorizes food classes into short-term, moderate-term, and long-term, as discussed in Section 3.3. The Food-101 dataset comprises 101,000 images with 750 training images per class. The VFN dataset contains over 15,000 training images across 74 classes, representing commonly consumed food categories in the United States based on the WWEIA database [1]. Test sets are balanced, with 250 images per class for Food-101 and 50 images per class for VFN. Our experiment involves assessing 5, 10, and 20 tasks using a 5K memory buffer, facilitating performance evaluation of all considered methods.

Baselines. Our analysis will delve into the performance of five prevalent ER OCIL methods, namely iCaRL [45], MIR [2], GSS [3], ASER [48], DVC [14]. These methods are discussed in Section 2 and are used for evaluation within realistic food consumption scenarios simulated by RDDM. A comparison will be drawn between these methods and their enhanced versions, incorporating our proposed plug-and-play DMU. We incorporate a baseline method dubbed **Fine-tune** for reference. This method exclusively employs data from new classes and cross-entropy loss for continual learning without considering prior task performance. Hence, it serves as a performance lower bound. It is crucial to emphasize that our comparison does not encompass "offline" methods centered on repetition. This is due to the fundamental distinction between "online" and "offline" settings in terms of data encounter, training frameworks, and buffer strategies, as outlined in Section 1.

Evaluation Metrics. For a fair evaluation, we consider access to a held-out test set for each of the T tasks. Upon the model's completion of learning task t_i, we assess its test performance across all T tasks. This process allows us to construct the matrix $B \in \mathbb{R}^{T\times T}$, where $B_{i,j}$ represents the test classification accuracy of the model for task t_j after observing the final sample from task t_i [33]. Letting $\widehat{b}$ denote the vector of test accuracy for each task during random initialization, and we use the following metrics for performance evaluation: $Average\ Accuracy = \frac{1}{T}\sum_{j=1}^{T} B_{T,j}$ where $B_{i,j}$ is the accuracy on task j after the model has been trained from task 1 through i.

[1]https://data.nal.usda.gov/dataset/what-we-eat-america-wweia-database

Table 2. **Average Accuracy (%)** rate with task sizes of 5, 10, and 20 with a 5K memory buffer on the Food-101 and VFN dataset with exponential distribution (D). Results include existing ER methods and the enhanced version using our method (DMU) leveraging the RDDM framework for realistic food image sequences. The best accuracy results are highlighted in **boldface**.

Dataset	Food-101									VFN								
Task size	5			10			20			5			10			20		
Category	Short	Mod	Long	Short	Mod	Long	Short	Mod	Long	Short	Mod	Long	Short	Mod	Long	Short	Mod	Long
Finetune	53.00	51.00	48.65	45.23	37.50	35.10	22.00	23.67	21.11	33.10	25.50	26.90	23.67	20.80	19.17	14.62	11.84	10.80
iCaRL [CVPR '17]	75.02	69.34	57.75	67.0	56.6	49.65	57.24	51.69	47.13	59.04	52.01	38.39	50.78	46.64	36.07	44.09	47.96	34.81
iCaRL + DMU	**76.04**	**70.68**	**59.22**	**67.14**	**62.29**	**52.19**	**59.72**	**53.65**	**50.37**	**59.77**	**53.36**	**42.95**	**52.72**	**46.93**	**36.17**	**46.66**	**48.31**	**34.88**
Gain (Δ)	1.02	1.34	1.47	0.14	5.69	2.54	2.48	1.96	3.24	0.73	1.35	4.56	1.94	0.31	0.10	2.55	0.35	0.07
MIR [NeurIPS '19]	74.71	69.79	57.55	64.0	57.0	48.9	54.15	48.39	43.85	49.41	46.01	43.88	44.01	42.21	35.83	33.96	31.12	30.73
MIR + DMU	**82.14**	**75.69**	**67.97**	**72.73**	**71.17**	**67.72**	**65.50**	**63.03**	**62.04**	**54.33**	**48.06**	**44.90**	**50.78**	**45.64**	**36.50**	**38.81**	**34.95**	**34.81**
Gain (Δ)	7.43	5.9	10.42	8.73	14.17	18.82	11.35	14.64	18.19	4.92	1.96	1.02	6.77	3.43	0.67	4.92	3.83	4.08
GSS [NeurIPS '19]	72.17	69.77	58.54	51.5	50.5	47.8	40.33	42.8	38.51	53.41	45.64	37.19	44.68	44.13	34.49	33.74	34.75	29.82
GSS + DMU	**76.29**	**74.01**	**68.19**	**55.07**	**66.17**	**66.06**	**47.43**	**49.86**	**50.13**	**55.59**	**47.52**	**38.83**	**48.10**	**47.91**	**38.35**	**39.44**	**35.70**	**30.08**
Gain (Δ)	4.12	4.24	9.65	3.57	15.67	18.26	4.10	7.06	11.62	2.18	1.88	1.64	3.42	3.78	3.86	5.70	0.96	0.26
ASER [AAAI '21]	77.87	72.51	62.89	67.70	66.6	61.12	59.70	58.29	**56.35**	51.86	48.56	42.19	43.34	42.32	36.01	33.68	32.14	28.61
ASER + DMU	**78.33**	**73.96**	**64.06**	**72.27**	**68.17**	**64.06**	**60.59**	**59.74**	56.084	**52.78**	**51.15**	**46.08**	**43.59**	**44.47**	**40.36**	**33.74**	**32.75**	**30.07**
Gain (Δ)	0.46	1.45	1.17	4.57	2.1	2.94	0.89	1.45	–	0.92	2.59	3.89	0.25	2.15	4.35	0.06	0.61	1.46
DVC [CVPR '22]	77.06	72.86	64.43	73.46	66.89	61.05	58.74	60.51	59.90	48.55	39.88	38.57	41.64	29.96	30.53	30.93	34.56	27.66
DVC + DMU	**78.53**	**74.96**	**65.95**	**74.52**	**68.75**	**62.15**	**60.51**	**61.82**	**59.94**	**50.06**	**40.63**	**41.28**	**44.23**	**39.59**	**32.87**	**36.81**	**35.26**	**30.32**
Gain (Δ)	1.47	1.35	1.53	1.06	1.86	1.1	1.77	1.82	0.04	1.51	0.75	2.71	2.59	9.63	2.34	5.88	0.7	2.66

Implementation Details. We use ResNet-18 [22] pre-trained on ImageNet [27] as the backbone structure for all experiments in comparison, and our implementation is based on PyTorch [43]. Pre-trained networks simplify training and reduce training time, especially using complex datasets like Food-101 [5] and VFN [18]. The input image size for Food-101 and VFN is 224 $\times$ 224. For optimization, we use a stochastic gradient descent optimizer with a fixed learning rate of 0.001 and a weight decay of 10^{-4}. The training batch size is 16, while the testing batch size is 128. We set the distribution D (referring to Section 4.1) as exponential. Throughout the training process, the model encounters the incoming data, excluding exemplars, only once across all experimental iterations. In order to evaluate different experience replay methods, we establish a memory buffer size of 5K for storing exemplars during training for knowledge rehearsal. We conduct experiments by varying task sizes (5, 10, and 20) within all food consumption categories, as explained in Section 3.3.

5.2. Experimental Results

Table 2 presents the average accuracy across all task sizes and the three categories of food consumption patterns in both Food-101 and VFN datasets. It is evident that the continual learning performance varies considerably for different food consumption categories and task sizes. For instance, the model has fewer overall classes to learn in the short-term food consumption pattern with the fewest classes and the smallest task size. Consequently, even with fewer repetitions, the model achieves higher accuracy and less catastrophic forgetting. In contrast, we observe the lowest performance figures as the task sizes increase and accuracy is measured in the long-term food consumption category. This decrease in accuracy is because the model attempts to learn the most classes present across many tasks (as α reduces and β increases), leading to lower accuracy and higher forgetting, despite the increase in repeated classes. Specifically, we observe severe catastrophic forgetting problems when using Finetune due to the lack of training data for learned classes during the continual learning process. All the existing ER methods outperform Finetune since they store a small sample of learned training classes for knowledge rehearsal during the continual learning process. However, when we enhance the existing ER methods with our proposed DMU module, we observe a substantial increase in performance consistently across all categories and task sizes in both Food-101 and VFN datasets. Our DMU module selects the ideal data for training by comparing the performance of the entire training batch with the new class data, leading to higher learning and lower catastrophic forgetting.

We immediately observe that the performance on the VFN dataset is notably lower than that of the Food-101 dataset, as seen in Table 2. This difference in performance is due to the VFN dataset's smaller size, with only 15,000 training images across 74 categories, compared to the Food-101 dataset, which has 75,750 training images across 101 categories. Furthermore, the VFN dataset is an inherently imbalanced real-world dataset, which results in lower accuracy overall. Nevertheless, including the DMU module in the existing ER methodologies still significantly enhances performance across all experimented categories. On average, we demonstrate a 26% increase on the Food-101 dataset and a 12.5% increase on the VFN dataset, respectively.

Table 3. **Average Accuracy (%)** on Food-101 using DVC, and enhanced DVC with DMU by varying data distributions with a 5K buffer. The best results are marked in bold.

Method	Distribution	Accuracy
DVC	Exp	66.89
DVC+DMU	Exp	**68.75**
Gain (Δ)		**1.86**
DVC	Gauss	60.89
DVC+DMU	Gauss	**64.55**
Gain (Δ)		**3.75**

Table 4. **Average Accuracy (%)** on VFN using best performing methods on varied buffer sizes. The best results are marked in bold.

Dataset	VFN					
Buffer size	2K			0.5K		
Task size	20			20		
Category	short	med	long	short	med	long
MIR	30.50	28.93	28.22	27.81	24.72	26.24
MIR + DMU	**31.17**	**30.31**	**29.63**	**29.39**	**27.60**	**26.33**
Gain (Δ)	0.67	1.38	1.41	1.58	2.88	0.09
ASER	33.29	33.78	29.31	29.21	30.51	27.61
ASER+DMU	**33.86**	**34.70**	**29.48**	**29.78**	**30.92**	**28.26**
Gain (Δ)	0.57	0.92	0.17	0.57	0.41	0.65
DVC	21.66	20.02	19.00	20.95	19.73	18.80
DVC + DMU	**22.70**	**21.22**	**18.42**	**21.08**	**21.06**	**19.41**
Gain (Δ)	1.04	1.20	0.58	0.13	1.33	0.61
GSS	27.28	29.32	29.20	22.75	26.47	25.06
GSS + DMU	**28.20**	**32.29**	**29.22**	**24.49**	**30.85**	**25.98**
Gain (Δ)	0.92	2.97	0.02	1.74	4.38	0.92
iCaRL	42.50	34.00	33.60	36.28	25.60	26.22
iCaRL + DMU	**46.34**	**41.67**	**45.25**	**37.74**	**36.67**	**39.04**
Gain (Δ)	3.84	7.67	11.65	1.46	11.07	12.82

5.3. Ablation Study

In this section, we perform ablation studies to assess the effectiveness of our DMU module across varying memory buffers (1K and 0.5K) and different data distributions D simulated by RDDM (Section 4.1, Equation 4). Our focus for the data distribution is on a scenario exhibiting reduced imbalance. To achieve this, we configure D as Gaussian with a mean of zero and a standard deviation of 0.4, aiming to simulate a broader spectrum with decreased imbalance effects. We apply the DVC [14] method, utilizing a consistent exemplar size of 5K and a task size of 10 for the Food-101 dataset. To assess the impact of different buffer sizes on ER methods and the effectiveness of our proposed model across these sizes and present the results in Table 4.

The results presented in Table 3 highlight the significant enhancement our DMU module brings to the performance of existing ER methods on the Food-101 dataset, even in the presence of varying levels of imbalance (as demonstrated in Figure 4). These improvements are consistent across different buffer sizes, as highlighted in Table 4. Traditional ER

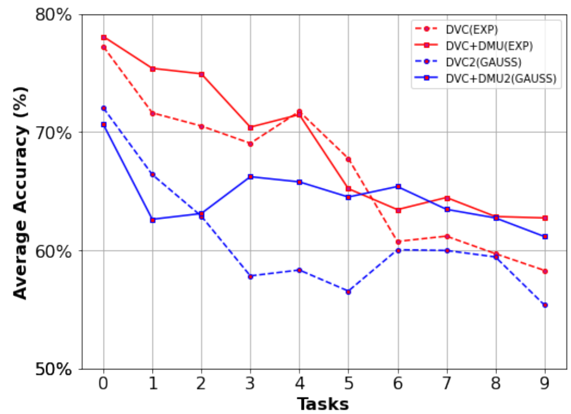

Figure 4. Accuracy of each incremental task in the Gaussian(GAUSS) and Exponential(EXP) distributions for the moderate food consumption category on DVC (with our proposed DMU module) using the Food-101 dataset with 10 tasks.

techniques, designed with balanced tasks and uniform sample distributions, are notably affected by the impact of input stream repetitions and imbalances, resulting in performance deterioration. These findings emphasize the effectiveness of targeted training with optimal samples compared to training with the complete set of input samples, irrespective of the specific ER method employed.

6. Conclusion

This paper focuses on Online Class-Incremental Learning for real-world food image classification. We introduce a novel Realistic Data Distribution Module (RDDM) to simulate real-world food consumption patterns in different scenarios along with a plug-and-play Dynamic Model Update (DMU) module, which is compatible with existing ER methods by independently targeting the training pipeline for improving performance. We show that our data distribution considers more realistic scenarios than existing OCIL systems, including a corresponding experimental benchmark. Furthermore, experimental results demonstrate our proposed method outperforms existing methods in more realistic food image classification settings on the challenging Food-101 and VFN datasets. Our future work will focus on improving the exemplar storage procedure by exploiting the most representative samples simulated by the RDDM framework. We also plan to evaluate our method on large supervised open-world datasets in the future.

References

[1] Ziad Ahmad, Marc Bosch, Nitin Khanna, Deborah A Kerr, Carol J Boushey, Fengqing Zhu, and Edward J Delp. A mobile food record for integrated dietary assessment. *Proceed-*

ings of the 2nd International Workshop on Multimedia Assisted Dietary Management, page 41–49, October 2016. 2

[2] Rahaf Aljundi, Lucas Caccia, Eugene Belilovsky, Massimo Caccia, Min Lin, Laurent Charlin, and Tinne Tuytelaars. *Online Continual Learning with Maximally Interfered Retrieval*. Curran Associates Inc., 2019. 2, 3, 6

[3] Rahaf Aljundi, Min Lin, Baptiste Goujaud, and Yoshua Bengio. Gradient based sample selection for online continual learning. *Advances in Neural Information Processing Systems*, 32, 2019. 2, 3, 6

[4] Eden Belouadah and Adrian Daniel Popescu. Il2m: Class incremental learning with dual memory. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 583–592, 2019. 3

[5] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 – mining discriminative components with random forests. *European Conference on Computer Vision*, 112(1):1–21, 2015. 2, 6, 7

[6] Francisco M. Castro, Manuel J. Marín-Jiménez, Nicolás Guil, Cordelia Schmid, and Karteek Alahari. End-to-end incremental learning. *Proceedings of the European Conference on Computer Vision*, pages 241–257, 2018. 3

[7] Arslan Chaudhry, Puneet K. Dokania, Thalaiyasingam Ajanthan, and Philip H. S. Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 1

[8] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-GEM. *International Conference on Learning Representations*, 2019. 3

[9] Aristotelis Chrysakis and Marie-Francine Moens. Online continual learning from imbalanced data. *Proceedings of the 37th International Conference on Machine Learning*, 119:1952–1961, 13–18 Jul 2020. 2

[10] Andrea Cossu, Gabriele Graffieti, Lorenzo Pellegrini, Davide Maltoni, Davide Bacciu, Antonio Carta, and Vincenzo Lomonaco. Is class-incremental enough for continual learning? *Frontiers in Artificial Intelligence*, 5, 2022. 2, 3

[11] Denise de Ridder, Floor Kroese, Catharine Evers, Marieke Adriaanse, and Marleen Gillebaart. Healthy diet: Health impact, prevalence, correlates, and interventions. *Psychol. Health*, 32(8):907–941, Aug. 2017. 4

[12] Heather A Eicher-Miller, Saul Gelfand, Youngha Hwang, Edward Delp, Anindya Bhadra, and Jiaqi Guo. Distance metrics optimized for clustering temporal dietary patterning among U.S. adults. *Appetite*, 144(104451):104451, Jan. 2020. 1, 4

[13] Ian Goodfellow, Mehdi Mirza, Xia Da, Aaron Courville, and Y. Bengio. An empirical investigation of catastrophic forgeting in gradient-based neural networks. 12 2013. 1

[14] Yanan Gu, Xu Yang, Kun Wei, and Cheng Deng. Not just selection, but exploration: Online class-incremental continual learning via dual view consistency. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7442–7451, June 2022. 2, 3, 6, 8

[15] Yue Han, Sri Yarlagadda, Tonmoy Ghosh, Fengqing Zhu, Edward Sazonov, and Edward Delp. Improving food detection for images from a wearable egocentric camera. 01 2023. 1, 2

[16] Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria De Munari, Monica Mordonini, and Stefano Cagnoni. Food image recognition using very deep convolutional networks. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, MADiMa '16, page 41–49, New York, NY, USA, 2016. Association for Computing Machinery. 1, 2

[17] Tyler Hayes, Nathan Cahill, and Christopher Kanan. Memory efficient experience replay for streaming learning. *IEEE Robotics and Automation Letters*, 4(10):9769–9776, 2019. 1

[18] Jiangpeng He, Luotao Lin, Heather Eicher-Miller, and Fengqing Zhu. Long-tailed food classification, 2022. 2, 6, 7

[19] Jiangpeng He, Luotao Lin, Jack Ma, Heather A Eicher-Miller, and Fengqing Zhu. Long-tailed continual learning for visual food recognition. *arXiv preprint arXiv:2307.00183*, 2023. 1

[20] Jiangpeng He, Runyu Mao, Zeman Shao, and Fengqing Zhu. Incremental learning in online scenario. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13923–13932, 06 2020. 1, 3

[21] Jiangpeng He and Fengqing Zhu. Online continual learning for visual food classification. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 2337–2346, October 2021. 1

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 7

[23] Hamed Hemati, Andrea Cossu, Antonio Carta, Julio Hurtado, Lorenzo Pellegrini, Davide Bacciu, Vincenzo Lomonaco, and Damian Borth. Class-incremental learning with repetition, 2023. 2, 3

[24] Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2019. 3

[25] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. 3

[26] Chrysi Koliaki, Theodoros Spinos, Marianna Spinou, Maria-Eugenia Brinia, Dimitra Mitsopoulou, and Nicholas Katsilambros. Defining the optimal dietary approach for safe, effective and sustainable weight loss in overweight and obese adults. *Healthcare (Basel)*, 6(3):73, June 2018. 4

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *ACM*, 60(6):84–90, may 2017. 7

[28] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018. 3

[29] Huiwei Lin, Baoquan Zhang, Shanshan Feng, Xutao Li, and Yunming Ye. Pcr: Proxy-based contrastive replay for online class-incremental continual learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2

[30] L. Lin, J. Guo, M. M. Aqeel, S. B. Gelfand, E. J. Delp, A. Bhadra, E. A. Richards, E. Hennessy, and H. A. Eicher-Miller. Joint temporal dietary and physical activity patterns: associations with health status indicators and chronic diseases. *Am J Clin Nutr*, 115(2):456–470, 02 2022. 1

[31] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In *International Conference on Smart Homes and Health Telematics*, 2016. 1, 2

[32] Xialei Liu, Marc Masana, Luis Herranz, Joost Van de Weijer, Antonio M. López, and Andrew D. Bagdanov. Rotate your networks: Better weight consolidation and less catastrophic forgetting. *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 2262–2268, 2018. 3

[33] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6470–6479, 2017. 3, 6

[34] Zheda Mai, Ruiwen Li, Jihwan Jeong, David Quispe, Hyunwoo Kim, and Scott Sanner. Online continual learning in image classification: An empirical survey. *Neurocomput.*, 469(C):28–51, jan 2022. 1, 3

[35] Runyu Mao, Jiangpeng He, Luotao Lin, Zeman Shao, Heather A. Eicher-Miller, and Fengqing Zhu. Improving dietary assessment via integrated hierarchy food classification. *2021 IEEE 23rd International Workshop on Multimedia Signal Processing*, pages 1–6, 2021. 1

[36] Runyu Mao, Jiangpeng He, Zeman Shao, Sri Kalyan Yarlagadda, and Fengqing Zhu. Visual aware hierarchy based food recognition. *Proceedings of the Pattern Recognition/ ICPR International Workshops and Challenges*, pages 571–598, 2021. 1

[37] Runyu Mao, Jiangpeng He, Zeman Shao, Sri Kalyan Yarlagadda, and Fengqing Zhu. Visual aware hierarchy based food recognition. *Proceedings of the Pattern Recognition. ICPR International Workshops and Challenges*, 12814:571–598, 2021. 2

[38] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer. Class-incremental learning: Survey and performance evaluation on image classification. *IEEE Transactions on Pattern Analysis and; Machine Intelligence*, 45(05):5513–5533, may 2023. 3

[39] Michael McCloskey and Neal J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. volume 24 of *Psychology of Learning and Motivation*, pages 109–165. Academic Press, 1989. 1

[40] Simon Mezgec, Tome Eftimov, Tamara Bucher, and Barbara Koroušić Seljak. Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. *Public Health Nutrition*, 22(7):1193–1202, 2019. 1, 2, 4

[41] Simon Mezgec and Barbara Koroušić Seljak. Nutrinet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients*, 9(7), 2017. 1, 2, 4

[42] Fei Mi, Lingjing Kong, Tao Lin, Kaicheng Yu, and Boi Faltings. Generalized class incremental learning. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 970–974, 2020. 2

[43] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. pages 8024–8035. Curran Associates, Inc., 2019. 7

[44] Barry M Popkin. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases2. *The American Journal of Clinical Nutrition*, 84(2):289–298, 2006. 1

[45] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5533–5542, 2017. 2, 3, 6

[46] Doyen Sahoo, Wang Hao, Shu Ke, Wu Xiongwei, Hung Le, Palakorn Achananuparp, Ee-Peng Lim, and Steven C. H. Hoi. Foodai: Food image recognition via deep learning for smart food logging. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '19, page 2260–2268, New York, NY, USA, 2019. Association for Computing Machinery. 2

[47] Zeman Shao, Shaobo Fang, Runyu Mao, Jiangpeng He, Janine Wright, Deborah Kerr, Carol Jo Boushey, and Fengqing Zhu. Towards learning food portion from monocular images with cross-domain feature adaptation, 2021. 2

[48] Dongsub Shim, Zheda Mai, Jihwan Jeong, Scott Sanner, Hyunwoo Kim, and Jongseong Jang. Online class-incremental continual learning with adversarial shapley value. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(11):9630–9638, 2021. 2, 3, 6

[49] Gudrun Sproesser, Matthew B Ruby, Naomi Arbit, Charity S Akotia, Marle dos Santos Alvarenga, Rachana Bhangaokar, Isato Furumitsu, Xiaomeng Hu, Sumio Imada, Gülbanu Kaptan, Martha Kaufer-Horwitz, Usha Menon, Claude Fischler, Paul Rozin, Harald T Schupp, and Britta Renner. Understanding traditional and modern eating: the TEP10 framework. *BMC Public Health*, 19(1):1606, Dec. 2019. 1

[50] Gido M van de Ven, Tinne Tuytelaars, and Andreas S Tolias. Three types of incremental learning. *Nat. Mach. Intell.*, 4(12):1185–1197, Dec. 2022. 3

[51] Hui Wu, Michele Merler, Rosario Uceda-Sosa, and John R. Smith. Learning to make better mistakes: Semantics-aware visual food recognition. *Proceedings of the 24th ACM International Conference on Multimedia*, 2016:172–176, 2016. 2

[52] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, page 3987–3995, 2017. 3