MCLC-NET: Multimodal Continual Learning for Leaf Counting

Anonymous CVPR submission

Paper ID 16684

Abstract

Continual learning is an emerging area of research with applications across many domains. Leaf counting, and agriculture in general, provide natural application scenarios for continual learning. The plant images are only available over time, and there is a significant drift in data distribution. To the best of our knowledge, there is no work on continual learning for leaf counting applications. Researchers use multimodal data, including RGB, depth, and IR images, for agriculture data analysis. There are some existing continual learning approaches for RGB and text data; however, we could not find any continual learning work on multimodal data of RGB, IR, and depth images. In this work, we investigate multimodal continual learning for leaf counting. We propose a novel replay-based continual learning approach. Experimental results demonstrate that our model consistently outperforms single-modal with the mean squared error of 1.35 and existing multimodal continual learning baselines. In multimodal data analysis, we often have the problem of missing modalities. We propose a strategy to handle missing modalities efficiently using continual learning and obtain the lowest mean squared error of 1.51.

1. Introduction

Agriculture growth directly affects human health, making it a vital area of research. Plant growth monitoring is facilitated by plant phenotyping, where leaf counting is the key task [1, 9, 14, 16, 17, 20, 24, 30]. It enables an analysis of plant development, flowering pattern, expected yield prediction, etc. There are two primary categories of approaches for leaf counting; the first one involves direct regression, and the other one is segmentation followed by counting [10, 20, 44]. However, some studies have also explored object detection and localization [4] or density estimation [20].

Counting leaves in real-world settings is a complex task due to diversity in plant type, weather, lighting conditions, etc. Thus, training a deep model often requires large-scale datasets. Few works have explored multi-modality-based leaf counting, such as including depth and thermal images along with RGB images to overcome occlusion and low lighting conditions [9, 17]. However, it is assumed that all the modalities are available during training, and if there is a new dataset with a different set of modalities, then a new model is required to train. Additionally, even if we are given the same set of modalities, adapting the model on another dataset that has possible differences in domain, including plant type and environmental conditions, the performance is compromised on the former dataset, also known as catastrophic forgetting [26, 26, 31, 34]. To learn in non-stationary environment, Continual Learning (CL) has been a popular and emerging research field [7, 41]. It allows the addition of new knowledge to the model in a way that does not conflict with existing ones and hence avoids catastrophic forgetting of already learned concepts.

Inspired by the above gaps and the success of CL in data shift conditions, we propose a unified continual learning-based network for multi-modal leaf counting, namely Multi-modal Continual Learning for Leaf-counting (MCLC). We propose to store a few images from past datasets based on how closely they represent the actual data distribution, and then these samples are used to mimic the past domain while learning a new one. The proposed replay-based CL strategy enables knowledge transfer across datasets with different modalities.

The major contributions of this paper are as follows:

- MCLC facilitates continual learning for leaf counting datasets that possess various domain shifts, including plant type and weather conditions.
- MCLC works with any set of imaging modalities. To the best of our knowledge, this is the first work that considers missing modality conditions in leaf counting.
- We collect a real-world multi-modality sequential dataset containing six episodes or tasks for evaluation of multimodal continual learning frameworks in domain shift conditions for leaf counting. It includes RGB, thermal, and depth-based datasets for cucumber, zucchini, and capsicum plants.

The rest of the paper is organized as follows. Section

2 outlines our methodology of continual learning for multimodal leaf counting and how we handle missing modalities. Section 3 lists experimental results to demonstrate the efficacy of our method. Related works are discussed in Section 4. We list our conclusions in Section 5.

2. Methodology

The proposed Multimodal Continual Learning for Leaf Counting (MCLC) framework addresses the challenges of multimodal continual learning (CL) for agricultural image analysis, specifically leaf counting across varying crop types under different lighting conditions throughout the day and using different modalities such as RGB, depth, and thermal images. We develop two main components: Multimodal Feature Fusion with Continual Learning framework and Missing Modality Handling. We will discuss each element in detail.

2.1. Multimodal feature fusion with Continual Learning

In the MCLC framework, we handle each modality separately at first and then concatenate the obtained features later. Then, we prepare a single feature from the concatenated mean and variance feature maps. the steps are defined below:

Creating Individual Features for Each Modality. Each modality goes through its own convolutional layer to extract useful features. These features from RGB images are denoted as F_{RGB}, from depth images as F_{depth}, and from thermal images as F_{thermal}. These feature maps represent the unique information each data type provides about the image.

Combining the Features. To create a single, unified feature map that combines information from all three modalities, we use a statistical fusion approach. This approach calculates the average (mean) and average (variance) across the features from RGB, depth, and thermal images. Each element in the final feature map, F_{fused}, is the average of the corresponding elements from F_{RGB}, F_{depth}, and F_{thermal}:

$$F_{\text{fused}} = \frac{1}{3} \sum_{m} F_m$$

where $m \in \{\text{RGB}, \text{depth}, \text{thermal}\}$. In simple terms, each element in F_{fused} is a balanced blend of the RGB, depth, and thermal information, creating a feature map that captures the important features from each modality.

Using the Fused Features for Learning. This combined feature map, F_{fused}, is then passed into a continual learning framework. This allows the model to learn from one task after another without forgetting what it learned before. By feeding in F_{fused}, the model maintains a systematic view of the data, making it better at handling new tasks as they come.

The concatenated features are then used in the Cl technique. We used a replay buffer, which stores the samples from each previous task and is used for the training for the next task along with the data samples from current tasks. From the idea of experience replay [32], instead of random samples from the task-like replay approach, we store the samples that are the nearest in the distance. To ensure that similar and nearby samples are stored, we used the Euclidean distance to calculate the closeness of the samples.

2.2. Handling missing modality

Multimodal datasets are valuable and important, but they can be difficult to obtain. Sometimes, this is due to limited access to resources, cameras, or sensors, and other times, it's because of a lack of awareness about the benefits or existence of multimodal data. To ensure robustness under potential missing modality scenarios, we introduce a random modality omission mechanism. For each sample, a random integer $r \in \{1, 2, 3\}$ is generated, which determines one modality (RGB, depth, or thermal) to be considered missing by substituting zeros in place of the selected modality's feature map. This missing modality handling function is defined as:

$$F_m = \begin{cases} F_m, & \text{if } m \neq r \\ 0, & \text{if } m = r \end{cases}$$

where m is the modality and r represents the randomly selected missing modality index. This approach simulates practical scenarios and allows the model to generalize in cases where complete data from all modalities may not be available.

3. Experiments and Results

3.1. Dataset

This section presents the dataset used in this work, the experiment sequence, and the results obtained. First, we describe the public dataset used and the multimodal dataset specifically prepared for this study, detailing its composition and structure for continual learning tasks. Next, we outline the experimental setup, including the different configurations and methods tested and how each experiment was designed to evaluate the proposed model. Finally, we provide an analysis of the results, comparing performance across various setups and discussing key findings.

We used two different datasets; one is the Computer Vision Problems in Plant Phenotyping (CVPPP) dataset [3][27], and the other is multimodal leaf counting (MMLC), the self-collected real field dataset.

CVPPP Dataset. CVPPP consists of four subsets, named A1, A2, A3, A4, and features images of two plant

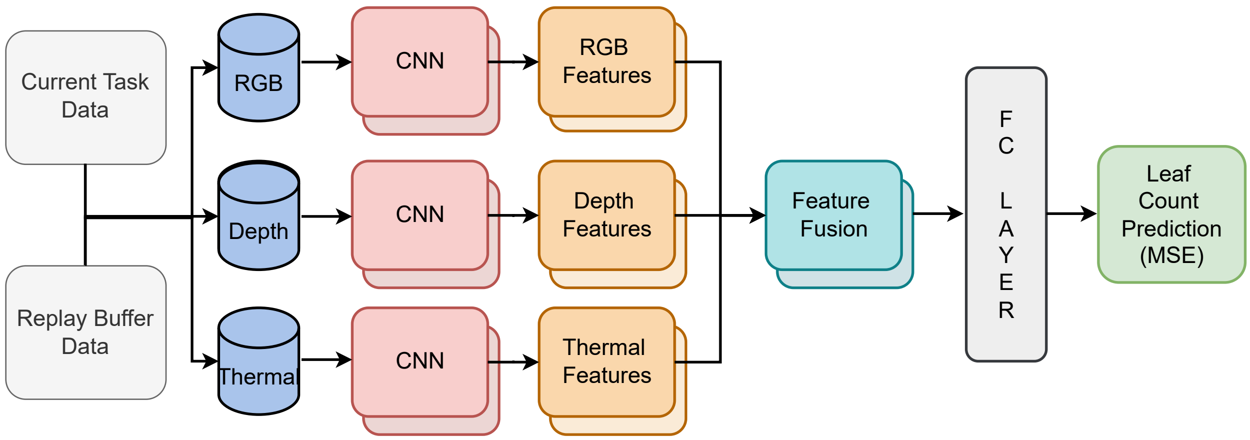


Figure 1. The figure shows the workflow of the multimodal leaf counting using the experience replay based continual learning.

Dataset Name	Modalities	Plants	No of Images	No of plant types	Annotation Type	Imaging Technique	Field Data?
CVPPP LCC	RGB	Arabidopsis, Tobacco	811	811	Binary masks, center masks, count data	High Res cameras	No
MSU-PID	Depth, IR, RGB, FMP	Arabidopsis, Bean	10000	16	pixel wise instance segmentation, leaf tips	Multiview Time series	No
Komatsuna	RGB, Depth	Brassica Rapa	1560	5	Pixel wise Instance Segmentation	Multiview Time Series	No
Our	**RGB, Depth, Thermal**	**Cucumber, Zucchini, Capsicum**	**1500**	**500**	**Leaf Count**	**High Res cameras**	**Yes**

Table 1. The table shows few of the available agricultural datasets on single modality (RGB) and multimodality dataset

species: Arabidopsis and Tobacco. This dataset includes RGB images of potted plants collected in a laboratory setting. Specifically, subsets A1 (128 images), A2 (31 images), and A4 (624 images) contain Arabidopsis plants, while subset A3 (27 images) includes images of Tobacco plants. The sample images of CVPPP are shown in Fig. 2.

MMLC dataset. MMLC is the dataset collected from multiple fields of various crops in their early stages in 3 modalities, namely RGB, depth, and thermal. The crops collected so far in all three modalities are cucumber, zucchini, and capsicum in 2nd week of its growth stage. The images of the crops are captured from the top view at a height within a meter. The images are taken in such a way that each plant is covered by each modality camera in the morning and evening on the same day. The RealSense depth sensor captures the pair of RGB and depth images. The thermal camera used is Fluke which captures the pair of RGB and thermal images at 4x resolution. The Google Pixel mobile phone captures the RGB images at higher resolution. We prepared the multimodal dataset by segregating the RGB, Thermal, and depth images in separate folders so that the corresponding images in each folder depict the same plant sample. The sample images from each crop are shown in Fig. 3, Fig. 4, and 5.

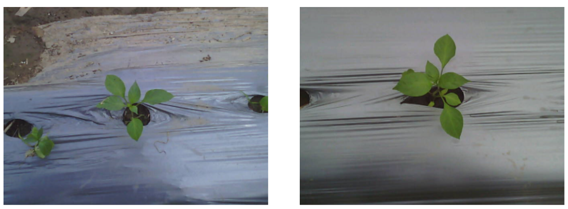
3.2. Experimental Setup

For the CVPPP dataset, we used RMSprop optimizer, mean absolute error, mean square error, and R squared value as the metrics. We use a learning rate of 0.0001 for 100 epochs to obtain the results for leaf counting methods, namely, regression, segmentation, density estimation, and object detection. For the continual learning approach, we used the

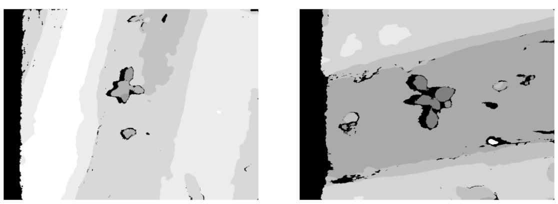
CVPR
#16684

A1 dataset A2 dataset A3 dataset A4 dataset

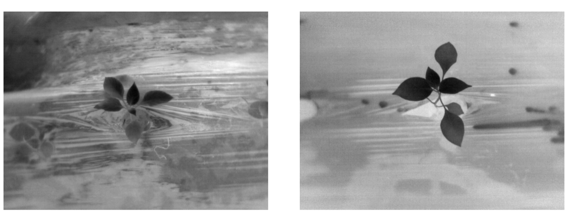
Figure 2. The figure shows the images of the subsets A1, A2, A3, A4 of the public dataset CVPPP.



(a) Capsicum RGB Morning-Evening images



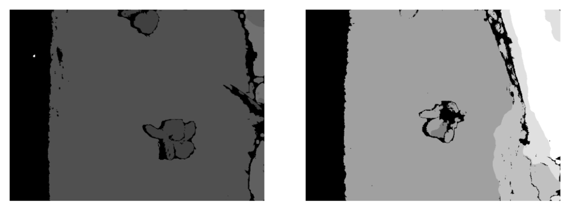
(b) Capsicum depth Morning-Evening images



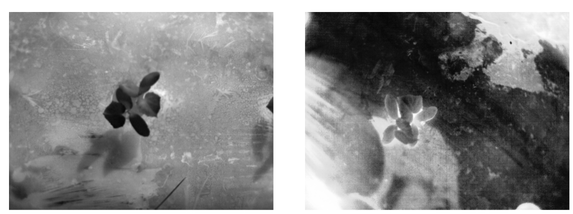
(c) Capsicum thermal Morning-Evening images

Figure 3. The figure shows the MMLC dataset images of capsicum plants for RGB, Depth, and Thermal modalities in the morning and evening (a) Capsicum RGB Morning-Evening images (b) Capsicum depth Morning-Evening images (b) Capsicum thermal Morning-Evening images.

(a) Cucumber RGB Morning-Evening images



(b) Cucumber depth Morning-Evening images



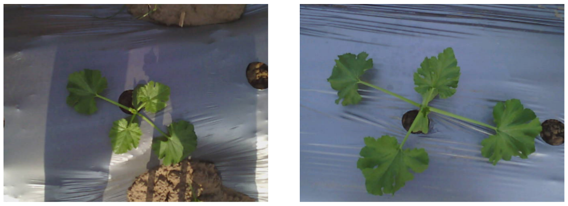
(c) Cucumber thermal Morning-Evening images

Figure 4. The figure shows the MMLC dataset images of cucumber plants for RGB, Depth, and Thermal modalities in the morning and evening (a) Cucumber RGB Morning-Evening images (b) Cucumber depth Morning-Evening images (b) Cucumber thermal Morning-Evening images.

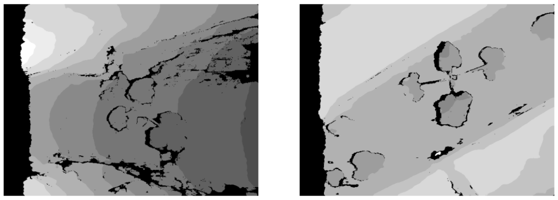
convolutional layer to pass the individual modalities to the base model and then used feature fusion to concatenate the separate features from individual modalities. We use mean squared error as the metric for all the experiments involving CL approaches.

3.3. Leaf counting using deep learning

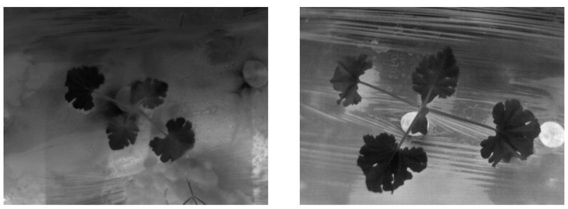
We performed standard leaf counting methods, which are mentioned in section 4 on CVPPP using the ResNet50 model. We used a public dataset, CVPPP, mentioned in section 3.1 for single modality (RGB images). Later, we used the MMLC dataset discussed in section 3.1 for multi-modal leaf counting.



(a) Zucchini RGB Morning-Evening images



(b) Zucchini depth Morning-Evening images



(c) Zucchini thermal Morning-Evening images

Figure 5. The figure shows the MMLC dataset images of zucchini plants for RGB, Depth, and Thermal modalities in the morning and evening (a) Zucchini RGB Morning-Evening images (b) Zucchini depth Morning-Evening images (b) Zucchini thermal Morning-Evening images.

We used the RGB images of the A1, A2, A3, and A4 subsets of CVPPP altogether for leaf counting using regression. For segmentation, we used the binary segmented images of RGB plant images of the whole data subset of CVPPP and fed them to the ResNet50 model. For the density estimation method, we used the images marked as the leaf centers for the RGB images of CVPPP. The results are shown in Table 3, 4, 5 and 6 respectively.

3.4. Leaf counting using continual learning

For the multi-modal leaf counting using the CL approach, we compared baselines of CL with our model MCLC framework on the MMLC dataset (capsicum and cucumber plants) on all three modalities, RGB, Depth, and Thermal.

Till now, the samples in the replay approach are stored randomly based on the memory size given. Further, for leaf counting using the CL approach, we stored the number of samples from each task according to the memory size specified but based on the importance of the samples. The importance of the samples is calculated using the nearest Euclidian distance. Once the distance of the sample is calculated from the cluster centroid, the samples will be stored in the memory for later use [33] [47].

3.5. Missing-modality in continual learning

Till now, the multimodality was handled in such a way that each task contained each modality. Further, for leaf counting using the CL approach, we plan to introduce the missing modality scenario [36]. For the first task, all 3 modalities will be included, but from task 2, the modalities may or may not be present. For example, task 1 contains the RGB, Depth, and Thermal images of the cucumber in the morning; task 2 contains the RGB and thermal images of the cucumber plant in the evening; task 3 contains the RGB and Depth images of the capsicum plant, task 4 contains the RGB images of zucchini plant, and so on.

3.6. Results

We performed the DL approach on CVPPP for leaf counting by using different leaf-counting approaches, as discussed in the literature. The result is given in Table 3, 4, 5 and 6. We can observe from Table 3 that the regression method for leaf counting gives the best performance on the CVPPP dataset with the lowest MSE value of 0.3442 and 1.0802 during training and testing, respectively, as compared to other methods. Therefore, we use the regression method for further experiments on multimodal leaf counting and continual learning leaf counting.

We also evaluate the regression method for the multimodal leaf counting on two datasets, namely MSU-PID and MMLC. The performance is shown in Table 2.

Dataset	Metrics	Training	Testing	Validation
MSU-PID	MSE	5.3351	3.3189	2.4031
	MAE	0.2031	0.3821	0.3273
	R2	0.9760	0.8737	0.9309
MMLC	MSE	0.1604	0.6469	0.5717
	MAE	0.3127	0.5867	0.5604
	R2	0.8518	0.3136	0.5306

Table 2. Performance evaluation of multimodal leaf counting on public multimodal dataset (MSU-PID) and our dataset (MMLC).

We analyzed the performance of ResNet50 on all three modalities of the dataset for the crops of capsicum, cucumber, and zucchini to understand the effect of multimodal data more precisely.

Metrics	Training	Testing	Validation
MSE	0.3442	1.0802	1.6191
MAE	0.3912	0.6994	0.6051
R2	0.9893	0.9656	0.9464

Table 3. Performance evaluation of Resnet50 on CVPPP for leaf counting using **regression** method.

Metrics	Training	Testing	Validation
MSE	0.6666	1.2125	1.5892
MAE	0.2614	0.6310	0.5107
R2	0.9792	0.9614	0.9473

Table 4. Performance evaluation of Resnet50 on CVPPP for leaf counting using **segmentation** method.

Metrics	Training	Testing	Validation
MSE	0.8539	2.0514	2.2643
MAE	0.6641	1.1014	1.1438
R2	0.9737	0.9244	0.9272

Table 5. Performance evaluation of Resnet50 on CVPPP for leaf counting using **density estimation** method.

Metrics	Training	Testing	Validation
MSE	16.9483	62.4298	16.4
MAE	3.7250	6.9737	3.4182
R2	0.0502	-1.8934	0.6034

Table 6. Performance evaluation of Resnet50 on CVPPP for leaf counting using **object detection** method.

Method	Avg MSE	Last row Avg MSE
Naive	8.4389	1.3875
Cumulative	1.2500	0.8648
Joint Training	0.8859	0.8859
Replay	3.2595	0.8664
GEM	3.1732	0.9071
A-GEM	3.2904	0.9103
EWC	4.8543	1.3549
SI	2.1511	1.4875
MCLC (ours)	1.3515	1.3271

Table 7. Performance evaluation of MCLC model and other baselines on **only RGB** images of MMLC dataset .

We have performed the experiments on the proposed model MCLC and compared it with various continual learning baselines, namely Replay, GEM, A-GEM, EWC, and SI on our MMCL dataset. The results are for RGB images, multimodal images, and for the scenario where some modalities were missing, shown in Table 7, 8 and 9.

4. Related works

Leaf counting is one plant phenotyping method by which the traits and growth of the plants can be observed. We will discuss the standard leaf counting methods, the agricultural dataset available, multimodal continual learning, and multimodal leaf counting.

Method	Avg MSE	Last row Avg MSE
Naive	2.0127	1.7349
Cumulative	0.9050	0.0199
Joint Training	0.6357	0.6357
Replay[32]	1.1499	0.6689
GEM	1.0762	0.4058
A-GEM	1.8768	0.4241
EWC	2.6659	1.7173
SI	1.9607	1.701
MCLC (ours)	1.4996	1.4221
Pheno[17]	3.3014	2.4139

Table 8. Performance evaluation of MCLC model and other baselines on **multimodal** images of MMLC dataset.

Method	Avg MSE	Last row Avg MSE
Naive	5832.2581	4.9819
Cumulative	26.0549	1.1355
Joint Training	1.1815	1.1815
Replay	3.789	0.9652
GEM	6.4625	1.0236
A-GEM	19.7548	1.081
EWC	1970.6864	1916.9084
SI	940.7727	1.7086
MCLC (ours)	1.5103	1.2164

Table 9. Performance evaluation of MCLC model and other baselines on multimodal images for **missing modality** cases of MMLC dataset.

4.1. Leaf counting using direct regression

Most leaf-counting approaches use direct regression techniques. For example, Andrei et al. trained their model on datasets with different sources and species, performed regression for leaf counting, and tested on the CVPPP dataset [3, 9, 27]. Similarly, Mario et al. used a deep learning model to predict leaf count on multimodal 2D images (including visible light, fluorescence, and near-infrared) [17]. In another work, Mario et al. performed regression on top-view images of rosette plants to count leaves [16]. Xinyan et al. used leaf skeletons with augmentation to address occlusion challenges, combining them with original images for better counting [44]. Andrei et al. also worked on multitasking, counting leaves, predicting leaf area, and classifying genotypes [11]. Guy et al. and Yotam et al. used multi-scale images to count leaves, using regression in cases with annotated datasets [12, 20]. Andrei et.al. in [10] used a layer-wise propagation approach in a deep neural network for the counting task. Mario et al. counted leaves in a target dataset using a model trained on a separate source dataset without direct access to the source data [38].

4.2. Leaf counting using segmentaion

Leaf segmentation can aid leaf counting tasks, with some methods performing segmentation before regression. Segmentation may be instance-based [22, 28, 46] or semantic-based [1]. Jean et al. used ML-based regression alongside leaf border prediction for occluded leaves [28], while Shubhra et al. and Mainak et al. combined SegNet-based segmentation with CNNs for improved counting [1, 8]. Lele et al. proposed a Mask-RCNN for instance-segmented images to support further counting [46].

4.3. Leaf counting using density estimation

Density estimation is used for leaf counting when leaf centers are identifiable. Some studies combine regression and density estimation for this purpose [2, 20]. For instance, Yotam et al. [20] developed methods using both regression and density estimation based on identified leaf centers. While density estimation specifically for leaf counting is rare, it has broader applications. Luca and Viet [13, 29] applied random forest regression for object density estimation, Alex et al. [21] used AlexNet for density estimation, and Victor et al. [23] created a map regression approach for this purpose. Furthermore, density estimation has also been used in medical science for counting tasks using dot annotations [43, 45].

4.4. Leaf counting using object detection and localization

In leaf counting, object detection helps by simultaneously classifying and locating leaves. For instance, Yi et al. [37] used YOLOv3 for detecting and counting leaves, and Michael et al. [4] employed Tiny-YOLOv3 for real-time leaf detection and counting. Similarly, Fuchun et al. [39] detected leaf tips with YOLO and counted based on tip numbers, while Shenglian et al. [25] used a deep learning network to detect and count leaves, addressing occlusions. Some methods, like Lele et al. [46], combined object detection and segmentation using Mask-RCNN. Although most studies use single-plant datasets in labs, Jan et al. [42] applied object detection on field plants, counting leaves by identifying plant-specific leaf points.

4.5. Multimodal Leaf counting

Multimodal learning is not new in machine learning models, and various modalities are often used in leaf counting to improve system performance and efficiency. For instance, Mario et al. in [17] used four modalities—visible light, fluorescence, depth, and near-infrared— within a deep learning architecture for leaf counting. However, resolution issues excluded the depth images from the counting task. Similarly, Andrei et al. in [9] collected data using fluorescence, infrared, RGB, and depth modalities, offering insights into different plant phenotyping applications.

4.6. Multimodal Continual learning

Leaf counting mostly relies on machine learning (ML), which is still unexplored for continual learning (CL). Although CL is common in classification tasks, it has rarely been applied to regression [18, 19]. Multimodal CL [5, 6, 15, 35, 36, 40, 48] has also been developed for diverse applications. For example, Fuchun et al. [36] proposed a method to update features and knowledge jointly, showing improvements in material recognition even with missing data types. Kai et al. [40] enhanced cross-modal image-text retrieval by separating training, indexing, and querying, while Tejas et al. [35] introduced a CL framework for vision and language tasks.

In this work, we are the first to introduce multimodal CL for leaf counting, using RGB, depth, and thermal data of different crops having variations in leaf structures to predict leaf count.

5. Conclusion

This work explores the application of continual learning for multimodal leaf-counting in plant images. To the best of our knowledge, this is the first attempt to apply continual learning for leaf counting. This is also the first attempt to apply continual learning on multimodal data of RGB, depth, and thermal modalities. We re-organize the leaf counting dataset for continual learning and apply various learning techniques. Experimental results show that the proposed replay approach performs best on RGB images. We improve the replay approach by employing a distance-based strategy to store frames in the buffer. The proposed strategy improves the results by 58.53%. Furthermore, we propose a framework to handle missing modalities with continual learning. The proposed method enables leaf counting without the availability of all modalities, with minimal increase in mean squared error. The current model is still memory and time-intensive. In our future work, we aim to optimize resource usage for multimodal continual learning.

References

[1] Shubhra Aich and Ian Stavness. Leaf counting with deep convolutional and deconvolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision workshops*, pages 2080–2089, 2017. 1, 7

[2] Shubhra Aich and Ian Stavness. Improving object counting with heatmap regulation. *arXiv preprint arXiv:1803.05494*, 2018. 7

[3] Jonathan Bell and Hannah M Dee. Aberystwyth leaf evaluation dataset. *URL: https://doi. org/10.5281/zenodo*, 168158 (17-36):2, 2016. 2, 6

[4] Michael Buzzy, Vaishnavi Thesma, Mohammadreza Davoodi, and Javad Mohammadpour Velni. Real-time plant leaf counting using deep object detection networks. *Sensors*, 20(23):6896, 2020. 1, 7

[5] Jie Cai, Xin Wang, Chaoyu Guan, Yateng Tang, Jin Xu, Bin Zhong, and Wenwu Zhu. Multimodal continual graph learning with neural architecture search. In *Proceedings of the ACM Web Conference 2022*, pages 1292–1300, 2022. 7

[6] Yuliang Cai and Mohammad Rostami. Dynamic transformer architecture for continual learning of multimodal tasks. *arXiv preprint arXiv:2401.15275*, 2024. 7

[7] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3366–3385, 2021. 1

[8] Mainak Deb, Krishna Gopal Dhal, Arunita Das, Abdelazim G Hussien, Laith Abualigah, and Arpan Garai. A cnn-based model to count the leaves of rosette plants (lc-net). *Scientific Reports*, 14(1):1496, 2024. 7

[9] Andrei Dobrescu, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Leveraging multiple datasets for deep leaf counting. In *Proceedings of the IEEE International Conference on Computer Vision workshops*, pages 2072–2079, 2017. 1, 6, 7

[10] Andrei Dobrescu, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Understanding deep neural networks for regression in leaf counting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1, 6

[11] Andrei Dobrescu, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Doing more with less: a multitask deep learning approach in plant phenotyping. *Frontiers in plant science*, 11:141, 2020. 6

[12] Guy Farjon, Yotam Itzhaky, Faina Khoroshevsky, and Aharon Bar-Hillel. Leaf counting: Fusing network components for improved accuracy. *Frontiers in plant science*, 12: 575751, 2021. 6

[13] Luca Fiaschi, Ullrich Köthe, Rahul Nair, and Fred A Hamprecht. Learning to count with regression forest and structured labels. In *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*, pages 2685–2688. IEEE, 2012. 7

[14] Fabio Fiorani and Ulrich Schurr. Future scenarios for plant phenotyping. *Annual review of plant biology*, 64(1):267–291, 2013. 1

[15] Sibo Gai, Zhengyu Chen, and Donglin Wang. Multi-modal meta continual learning. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021. 7

[16] Mario Valerio Giuffrida, Massimo Minervini, and Sotirios A Tsaftaris. Learning to count leaves in rosette plants. *In CVPPP workshop, page 13. British Machine Vision Association*, 2015. 1, 6

[17] Mario Valerio Giuffrida, Peter Doerner, and Sotirios A Tsaftaris. Pheno-deep counter: A unified and versatile deep learning architecture for leaf counting. *The Plant Journal*, 96(4):880–890, 2018. 1, 6, 7

[18] Wolfgang Grote-Ramm, David Lanuschny, Finn Lorenzen, Marcel Oliveira Brito, and Felix Schönig. Continual learning for neural regression networks to cope with concept drift in industrial processes using convex optimisation. *Engineering Applications of Artificial Intelligence*, 120:105927, 2023. 7

[19] Yujiang He and Bernhard Sick. Clear: An adaptive continual learning framework for regression tasks. *AI Perspectives*, 3 (1):2, 2021. 7

[20] Yotam Itzhaky, Guy Farjon, Faina Khoroshevsky, Alon Shpigler, and Aharon Bar-Hillel. Leaf counting: Multiple scale regression and detection using deep cnns. In *BMVC*. Newcastle, 2018. 1, 6, 7

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 7

[22] Dmitry Kuznichov, Alon Zvirin, Yaron Honen, and Ron Kimmel. Data augmentation for leaf segmentation and counting tasks in rosette plants. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 7

[23] Victor Lempitsky and Andrew Zisserman. Learning to count objects in images. *Advances in neural information processing systems*, 23, 2010. 7

[24] Lei Li, Qin Zhang, and Danfeng Huang. A review of imaging techniques for plant phenotyping. *Sensors*, 14(11):20078–20111, 2014. 1

[25] Shenglian Lu, Zhen Song, Wenkang Chen, Tingting Qian, Yingyu Zhang, Ming Chen, and Guo Li. Counting dense leaves under natural environments via an improved deep-learning-based object detection algorithm. *Agriculture*, 11 (10):1003, 2021. 7

[26] Michael McCloskey and Neal J Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, pages 109–165. Elsevier, 1989. 1

[27] Massimo Minervini, Andreas Fischbach, Hanno Scharr, and Sotirios A Tsaftaris. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern recognition letters*, 81:80–89, 2016. 2, 6

[28] Jean-Michel Pape and Christian Klukas. Utilizing machine learning approaches to improve the prediction of leaf counts and individual leaf segmentation of rosette plant images. *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*, 3:1–12, 2015. 7

[29] Viet-Quoc Pham, Tatsuo Kozakaya, Osamu Yamaguchi, and Ryuzo Okada. Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation. In *Proceedings of the IEEE international conference on computer vision*, pages 3253–3261, 2015. 7

[30] Roland Pieruschka and Uli Schurr. Plant phenotyping: past, present, and future. *Plant Phenomics*, 2019. 1

[31] Roger Ratcliff. Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychological review*, 97(2):285, 1990. 1

[32] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in neural information processing systems*, 32, 2019. 2, 6

[33] Ario Sadafi, Raheleh Salehi, Armin Gruber, Sayedali Shetab Boushehri, Pascal Giehr, Nassir Navab, and Carsten Marr. A continual learning approach for cross-domain white blood cell classification. In *MICCAI Workshop on Domain Adaptation and Representation Transfer*, pages 136–146. Springer, 2023. 5

[34] Chenze Shao and Yang Feng. Overcoming catastrophic forgetting beyond continual learning: Balanced training for neural machine translation. *arXiv preprint arXiv:2203.03910*, 2022. 1

[35] Tejas Srinivasan, Ting-Yun Chang, Leticia Pinto Alva, Georgios Chochlakis, Mohammad Rostami, and Jesse Thomason. Climb: A continual learning benchmark for vision-and-language tasks. *Advances in Neural Information Processing Systems*, 35:29440–29453, 2022. 7

[36] Fuchun Sun, Huaping Liu, Chao Yang, and Bin Fang. Multimodal continual learning using online dictionary updating. *IEEE Transactions on Cognitive and Developmental Systems*, 13(1):171–178, 2020. 5, 7

[37] Yi-Lin Tu, Wei-Yang Lin, and Yao-Cheng Lin. Automatic leaf counting using improved yolov3. In *2020 International Symposium on Computer, Consumer and Control (IS3C)*, pages 197–200. IEEE, 2020. 7

[38] Mario Valerio Giuffrida, Andrei Dobrescu, Peter Doerner, and Sotirios A Tsaftaris. Leaf counting without annotations using adversarial unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 6

[39] Mukesh Kumar Vishal, Biplab Banerjee, Rohit Saluja, Dhandapani Raju, Viswanathan Chinnusamy, Sudhir Kumar, Rabi Narayan Sahoo, and Jagarlapudi Adinarayana. Leaf counting in rice (oryza sativa l.) using object detection: A deep learning approach. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 5286–5289. IEEE, 2020. 7

[40] Kai Wang, Luis Herranz, and Joost van de Weijer. Continual learning in cross-modal retrieval. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3628–3638, 2021. 7

[41] Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: Theory, method and application. *arXiv preprint arXiv:2302.00487*, 2023. 1

[42] Jan Weyler, Andres Milioto, Tillmann Falck, Jens Behley, and Cyrill Stachniss. Joint plant instance detection and leaf count estimation for in-field plant phenotyping. *IEEE Robotics and Automation Letters*, 6(2):3599–3606, 2021. 7

[43] Weidi Xie, J Alison Noble, and Andrew Zisserman. Microscopy cell counting and detection with fully convolutional regression networks. *Computer methods in biomechanics and biomedical engineering: Imaging & Visualization*, 6(3): 283–292, 2018. 7

[44] Xinyan Xie, Yufeng Ge, Harkamal Walia, Jinliang Yang, and Hongfeng Yu. Leaf-counting in monocot plants using deep regression models. *Sensors*, 23(4):1890, 2023. 1, 6

[45] Yuanpu Xie, Fuyong Xing, Xiangfei Kong, Hai Su, and Lin Yang. Beyond classification: structured regression for robust cell detection using convolutional neural network. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 358–365. Springer, 2015. 7

[46] Lele Xu, Ye Li, Yuanyuan Sun, Lei Song, and Shan Jin. Leaf instance segmentation and counting based on deep object detection and segmentation networks. In *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, pages 180–185. IEEE, 2018. 7

[47] Rui Yang, Emmanuel Dellandrea, Matthieu Grard, and Liming Chen. Imbalanced data robust online continual learning based on evolving class aware memory selection and built-in contrastive representation learning. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 277–283. IEEE, 2024. 5

[48] Jingxin Zhang, Donghua Zhou, and Maoyin Chen. Monitoring multimode processes: A modified pca algorithm with continual learning ability. *Journal of Process Control*, 103: 76–86, 2021. 7